Natural genetic variation in *Arabidopsis thaliana* photosynthesis

Pádraic J. Flood

Thesis committee

Promotor

Prof. Dr M. Koornneef

Personal chair at the Laboratory of Genetics

Director at the Max Planck Institute for Plant Breeding Research (MPIPZ) Cologne, Germany

Co-promotors

Dr M.G.M. Aarts

Associate professor, Laboratory of Genetics

Wageningen University

Dr J. Harbinson

Assistant professor, Horticulture and Product Physiology

Wageningen University

Other members

Prof. Dr P.C. Struik, Wageningen University

Prof. Dr J. de Meaux, University of Cologne, Germany

Dr E. Murchie, University of Nottingham, United Kingdom

Dr M. Malosetti, Wageningen University

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

Natural genetic variation in *Arabidopsis thaliana* photosynthesis

Pádraic J. Flood

Thesis

submitted in fulfilment 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 Friday 30 October 2015

at 4 p.m. in the Aula.

Pádraic J. Flood

Natural genetic variation in Arabidopsis thaliana photosynthesis

PhD thesis, Wageningen University, Wageningen, NL (2015)

With references, with summary in English

ISBN 978-94-6257-500-4

Contents

Chapter 1

General Introduction

-On writing a PhD thesis

I think perhaps the most important problem is that we are trying to understand the fundamental workings of the universe via a language devised for telling one another when the best fruit is.

-- (Terry Pratchett)

Where light becomes life

It is a little recognized fact that plants grow upside down. Yes, you read that right, I did just state that plants grow upside down. Before you go and throw this thesis in the rubbish bin, bear with me while I reveal the logic behind this apparent absurdity.

The first step towards elucidating the curious upside down nature of plant growth was taken by the Flemish chemist and physiologist Jan Baptist van Helmont. In 1643 he planted a willow tree weighing 5 pounds (2.2 kg) in a pot containing 200 pounds (90 kg) of soil previously dried in a furnace. For the next five years he watered this pot with either rainwater or distilled water after which he removed the willow and found that it weighed 169 pounds and 3 ounces (77 kg) whilst the soil had only lost 2 ounces (57 grams). From this he erroneously concluded that the tree obtained its mass from the water. It took about 150 years and the endeavors of many scientists, in which Joseph Priestley, Jan Ingen-Housz, Jean Senebier and Nicholas-Théodore de Saussure played prominent parts, to come to the correct conclusion that air, specifically carbon dioxide (CO_2), is responsible for most of the solid mass of plants (Hill, 2012). Arising from this, the key insight into the upside down nature of plants was garnered: plants are largely composed of air not soil, in this respect they are coalesced air which grows into the earth and not from it. This particular realization was one of the main motivations for me to pursue a PhD on the process which makes this remarkable phenomenon possible: photosynthesis.

Photosynthesis is the fundament upon which virtually all life on Earth depends[1]. It traps the light of the Sun in chemical bonds providing the energy necessary for life to maintain its improbable entropic state. In the case of oxygenic photosynthesis it also produces oxygen as a waste product which, via oxygenic respiration, can be used to extract much more of this stored sunlight than would otherwise be possible. Thus, with spectacular providence, oxygenic photosynthesis both injects more energy and matter into the biosphere and provides the necessary chemical tools for efficiently recovering this energy when needed.

Oxygenic photosynthesis is an ancient biological innovation dating back to over 3 billion years ago (Hohmann-Marriott and Blankenship, 2011); its evolution was a pivotal event in the history of life as it changed the chemistry of the planet from partially reducing to

[1] There are some communities of organisms which live on hydrothermal vents and are energetically independent of photosynthesis, however without the generation of the protective ozone layer which oxygenic photosynthesis gave rise to, it is doubtful there would be much water left even to sustain this life.

oxidizing. As already mentioned, it resulted in a massive increase in the amount of energy and matter flowing into the biosphere[2]. It also facilitated the development of the ozone layer which has protected the Earth from much of the ultraviolet radiation emitted by the Sun. In summary, oxygenic photosynthesis has been absolutely key in the development of a suitable environment on Earth within which life can thrive. Finally, unlike many key evolutionary innovations, such as multicellularity, eyes and intelligence, oxygenic photosynthesis only evolved once. The singularity of this event is striking and as such it is widely considered a key event in the history of life[3].

The unlikely nature of oxygenic photosynthesis is perhaps best explained by its sheer complexity; it is thought to be the most complex energy transducing process in biology (Shi et al., 2005). In plants this complexity is taken a step further by the separation of genes coding for photosynthetic proteins into two discrete genetic lineages, the nuclear genome and the chloroplast genome, or plastome. Both genomes encode subunits of large protein complexes, such as photosystem II (PSII), which must be expressed and manufactured in tight coordination and then fit together in a precise arrangement. One mistake can be fatal as photosynthetic energy transduction deals with high energy processes that generate reactive intermediates, which can, in particular, react with molecular oxygen. Small errors in the assembly and operation of these systems can lead to the creation of free radicals such as superoxide, which in turn can cause widespread and indiscriminate cellular damage. This high level of complexity and interdependence is thought to have constrained the evolution of photosynthesis by limiting the number of options upon which evolution can work (Shi et al., 2005).

The observation that core photosynthetic proteins are highly conserved, combined with little evidence that crops were limited by their photosynthetic capacity (usually referred to as source limitation) has led to a general lack of enthusiasm for improving photosynthesis. Since the turn of the millennium this has begun to change with numerous researchers thinking of ways to improve photosynthesis (Richards, 2000; Long et al., 2006). In general the improvement approaches touted have taken an engineering angle (Bar-Even et al., 2010; von Caemmerer and Evans, 2010; Raines, 2011; Furbank et al., 2015; Long et al., 2015; Ort et al., 2015) with little attention paid to the potential of natural genetic variation as a source for improving

[2] In this context when I say biosphere I am referring to the sum total of all living things on the planet not the specific region they inhabit, confusingly the term is used somewhat interchangeably and I could not find a more precise word.

[3] The only other evolutionary step which is also thought to be extremely unlikely is the evolution of eukaryotes, this also happened only once and has been hypothesized to be a key requirement for complex multicellular life to evolve (Lane and Martin, 2010).

photosynthesis (Flood et al., 2011; Lawson et al., 2012). In this thesis I aim to address this lack of attention by surveying natural genetic variation in photosynthesis in the model plant species *Arabidopsis thaliana,* hereafter referred to as Arabidopsis.

Arabidopsis is an unassuming plant, most people have probably either stepped on it, sprayed it with herbicide or pulled it from their flower bed without ever really seeing it. It is small, forms a rosette of leaves from which a thin wispy inflorescence grows and gives rise to a host of tiny, odorless white flowers. To the untrained eye it is perhaps the epitome of unimpressive. But beneath this banal exterior Arabidopsis holds a beguiling secret: it has become our gateway into the wondrous world of plants. Arabidopsis is the most unique crop there is; its fruit is the fruit of knowledge. Over the past half century this unassuming weed has been used to probe the inner workings of plants. In the year 2000 it became the first plant ever to have its whole genome sequenced (The Arabidopsis Genome Initiative, 2000) and it continues to be at the forefront of fundamental plant science research (Koornneef and Meinke, 2010).

Recently, there has been a growing interest in investigating natural genetic variation in Arabidopsis as such variation is often evolutionarily and ecologically informative (Koornneef et al., 2004; Alonso-Blanco et al., 2009). Prior to this recent interest in natural variation, most studies into Arabidopsis gene function worked with genetic variation created in the laboratory. Such manufactured variation is generated via mutagenesis which can be induced by chemical and ionizing radiation based techniques as well as random insertion of transposable elements and T-DNAs to disrupt gene function. This approach of using induced variation was very successful in identifying many important genes but failed to give any insight into how the processes which these genes control may vary in nature[4]. Understanding this is essential to understanding how plants adapt to different environments, which in turn is key to understanding how we may better tailor our crops for current and future environments.

In recent years, our ability to identify the underlying genetic causes of the variation we see around us has flourished. The seminal paper of Lander and Botstein (1989) initiated this through development of a statistical approach to link genetic polymorphisms[5] to phenotypes of

[4] Also these induced variation studies were conducted in a handful of genetic backgrounds, if these reference genomes happened to be natural gene knockouts then this approach will not reveal gene function. A classic example is the important flowering time gene *FRIGIDA* for which the Col-0 reference genotype is a natural null mutant, without additional genotypes the importance of this gene, and its downstream targets, would not have been recognised.

[5] In this context these polymorphisms are usually called genetic markers or molecular markers, as they identify (mark) a genetic position (locus).

interest. This approach localized or 'mapped' the genetic factors responsible for the observed phenotypic variation to a particular location in the genome, commonly known as a quantitative trait locus or QTL. These original studies were conducted in simple bi-parental mapping populations. We have since developed new types of populations (Kover et al., 2009; Huang et al., 2011) and increasingly sophisticated statistical methods (Segura et al., 2012), but the biggest game changer has been the explosion in the availability of cheap and rapid high density genotyping (Mardis, 2008; Metzker, 2010; Poland and Rife, 2012). The ready availability of huge numbers of genetic markers has expanded the range of possible quantitative genetic studies beyond those based on constructed mapping populations into much more diverse populations of genotypes collected directly from nature. The great benefit of such populations is their increased mapping resolution, which ironically was the reason they were not feasible prior to the availability of high density marker sets. This increased mapping resolution is a result of the historical recombination events which occurred in nature, reducing the extent of linkage disequilibrium[6] (Slatkin, 2008) and necessitating a tight physical linkage between marker polymorphisms and the causal loci. Hence the resulting 'high resolution' and the need for high density marker sets to identify such associations. Such 'natural' populations of genotypes scored with high density marker sets are called genome wide association (GWA) populations. The first GWA study was conducted in Arabidopsis in 2010 (Atwell et al., 2010) and was welcomed by the community with great fanfare. Since then much has been learned about the genetic architecture of complex traits, and although still a powerful technique, our expectations have been rightly tempered (Rockman, 2012). This is largely due to the huge number of loci of small effect accounting for a large proportion of the phenotypic variation, harking back to Fisher's infinitesimal model of quantitative traits (Fisher, 1918). Employing these advanced quantitative genetic techniques to identify the alleles underlying natural genetic variation in photosynthesis is an important research avenue. However, for this to be possible, one significant hurdle had to be overcome: high-throughput measurement of photosynthesis in living plants.

Photosynthesis is a rather elusive phenotype, in that the photosynthetic capacity of a plant cannot be determined without the aid of sophisticated measurement techniques. The gold standard of photosynthetic measurements is called infra-red gas analysis (IRGA) which measures the rate of CO_2 uptake by a leaf as a means to determine its photosynthetic rate. The

[6] Non-random association of genetic loci, usually due to physical proximity on the chromosome. The closer two loci are the less likely recombination will occur between them, thus they tend to be inherited together. It is this 'linkage' that is used to map QTLs, as the genetic polymorphisms (markers) which are closer to the causal mutation/allele will show tighter association (linkage) with the phenotype.

advantage of this technique is that it is a very direct measure of photosynthesis, as CO_2 uptake is what photosynthesis is all about. The downside of this technique is that it is quite slow and not at all suited for large scale screenings of thousands of plants. Such a phenotyping throughput is necessary if a sufficiently large number of plants are to be screened to permit the identification of the genetic loci responsible for the variation. To achieve such a throughput in a non-invasive manner another technique is called for, in this case chlorophyll fluorescence. Chlorophyll fluorescence allows non-invasive image-based phenotyping of plants for their photosynthetic rate by simply flashing them with specific wavelengths of light of known intensity (Baker, 2008). The main advantages of this technique over IRGA is the speed at which measurements can be made and the absence of cumbersome gas mixing systems and leaf chambers. Even so, at the outset of this doctoral project no chlorophyll fluorescence imaging system was capable of the measurement throughput we needed.

The mapping techniques described so far are focused exclusively on nuclear genetic variation and completely ignore variation in the organellar genomes, harbored by the mitochondria and the chloroplasts. As mentioned, these organelles are particularly important for photosynthesis (the chloroplast is the actual cellular location where photosynthesis takes place). The main reason that the quantitative genetic techniques described above ignore the organelle genomes is due to their non-mendelian uniparental inheritance pattern. In the case of Arabidopsis both mitochondria and chloroplasts are inherited from the mother, as is common in most angiosperms (Hagemann, 2004). This inheritance pattern precludes the possibility of the recombination and segregation which is essential for the success of genetic mapping. Generally the omission of these organellar genomes, collectively referred to as the plasmotype, is not given much thought as they contain a miniscule amount of genetic information when compared to the nuclear genome. This systematic neglect has been brought to the attention of the scientific community by several researchers (Budar and Roux, 2011; Pesole et al., 2012). The importance of natural intraspecific variation in the plasmotype for adaptation is becoming increasingly accepted (Bock et al., 2014; Dobler et al., 2014), further supporting the case for work on these diminutive, but crucially important, genomes. In Chapter 7 I address this discrepancy in research effort through the development of a novel population to study natural variation in Arabidopsis plasmotypes for a diverse array of phenotypes.

Finally, apart from the intellectual and scientific dividends, investigating natural genetic variation in plant photosynthesis ultimately has real world relevance which was, and is, a significant motivation behind this work. Currently, there are seven billion people on the Earth,

and most projections expect this number to rise to between nine and ten billion by 2050. While the range of social, environmental and economic challenges emerging from this increase are vast and of great importance, none of them matter if we do not have enough food. As such global agriculture is faced with an incredible challenge, it must increase yield whilst at the same time reducing both inputs and land usage. One obvious means to alleviate some of this pressure is to increase the primary productivity of the whole system. Since the Green Revolution of the 1950s and 60s, such increases in primary production have come about due to improved agronomic practices, increased input of fertilizers and pesticides, and breeding for genetically amenable traits such as harvest index, disease resistance, and leaf angle[7]. These approaches have been fantastically successful but they have also been, with the possible exception of agronomic practices, exhausted (Long et al., 2006). Harvest index is almost at its theoretical limit in elite crop cultivars: there is literally no room for improvement, and fertilizer and pesticide use ought to be reduced rather than increased. Out of all the traits bred for, photosynthesis is the last physiological frontier (Long et al., 2006). As discussed earlier, while there are numerous research groups working on this issues from many angles. In general, natural genetic variation has been sidelined. This is perhaps due to the main push for photosynthetic improvements coming from physiologists and not breeders. Physiologists intrinsically abhor black-box approaches such as genomic selection because they love to know how things work[8]. However, to redesign photosynthesis from first principles is a long term goal and unlikely to yield results in the near future. Natural genetic variation, on the other hand, offers a potential suite of pretested options (Masel, 2006). Variation which occurs in nature is mostly either adaptive or neutral as variation which significantly reduces fitness is quickly purged from the population. Thus, naturally occurring variation which affects the photosynthetic capacity of a plant will often be adaptive or beneficial in the environment that the plant evolved[9].

A final advantage of studying natural variation is that it is an unbiased approach. An implicit assumption in any engineering approach is that we understand the process well enough to consider redesigning it. A nice example is that most notorious of photosynthetic enzymes D-ribulose-1,5-bisphosphate carboxylase/oxygenase, more commonly referred to as Rubisco.

[7] Yield has always been bred for which is far from being 'genetically amenable', I wish to refer to some of the novel approaches of breeding in the green revolution. Selecting for yield, although important, is far from novel.
[8] Apologies for stereotyping, this is an opinion based on experience not evidence.
[9] Natural selection is not the only force at work in evolution, the other processes, such as drift, recombination and mutation, do not lead directly to adaptation and could conceivably lead to an allele of poor photosynthetic performance reaching a high frequency. However, even such a negative allele may be informative, flagging a role for its associated process in photosynthesis.

Rubisco is thought to be the most abundant protein on the planet and one of the least efficient enzymes in terms of size (it is big) and speed (it is slow). Not only is it slow but it is also rather promiscuous; its job is to catalyze a reaction between ribulose-1,5-bisphosphate (RuBP) and CO_2, which results in net carbon gain, but it often reacts RuBP with O_2 instead, which results in a net carbon loss[10]. This has been known for many years and improving Rubisco has been the dream of many photosynthetic researchers[11]. More recently, however, a group of researchers carefully analyzed the properties of many different Rubiscos and came to the conclusion that they may all be close to perfectly optimized (Tcherkez et al., 2006), which, if true, means that it is not a good target for improvement[12]. To exploit natural variation we do not need to pick a specific target but instead find a dimension of genetic variation that associates with a desirable phenotype and use this genetic knowledge in a breeding program. In principle, understanding the causal link between the genetic locus and phenotype is not necessary – all that matters is the correlation between genotype and phenotype, once this is achieved we can allow the best path to improvement to present itself to us. Having said all of the above I want to stress that I think both approaches, targeted and untargeted are warranted, just like in adapting populations variation is the key to survival. If we are to succeed at improving photosynthesis we should explore all possible approaches, including redesigning the whole pathway (Bar-Even et al., 2010).

The main goal of my PhD thesis research is to explore the natural variation in photosynthesis as a means to learn more about the process and to advertise it as a viable improvement approach. Hopefully the results presented in this thesis will encourage the photosynthetic research community to take more heed of natural genetic variation and, in turn, benefit us all through increased photosynthetic capacity and agricultural primary production.

[10] This is called photorespiration.
[11] This has been rather wryly summed up by Whitney et al. (2011) where they state that "the challenge of making a "better Rubisco" has exceeded the grasp and career of many scientists", although they still believe it is possible, stoic or stubborn, time will tell.
[12] Evolution solved this, not by fixing the enzyme but by placing it in a CO_2 rich O_2 poor environment, thus reducing unnecessary carbon loss due to photorespiration, this adaptation is called C_4 photosynthesis.

Layout of this thesis

The primary theme of this thesis is an investigation of natural genetic variation in plant photosynthesis[13]. In **Chapter 2** I review the literature concerning natural genetic variation in plant photosynthesis, aiming to bring together research from both agricultural and ecophysiological perspectives to learn from both approaches. This chapter is written in such a style to facilitate an easy introduction to the topic for photosynthetic physiologists and quantitative genetics, both of whom are unlikely to be familiar with the work in each other's fields. Since the publication of our review of the topic (Flood et al., 2011) there has been quite an increase in the amount of work done with numerous additional research papers published (Gu et al., 2012; Takai et al., 2013; Driever et al., 2014; Salas Fernandez et al., 2014) and even an additional review (Lawson et al., 2012).

Having outlined in Chapter 2 several key steps which needed to be taken to facilitate a quantitative genetic analysis of intraspecific variation in photosynthesis, I began with the development of a high-throughput phenotyping platform for photosynthesis. In **Chapter 3** I discuss key considerations for the development of such a platform and describe the system which was ultimately built. I also describe the phenomena of fluctuating heritability, a biological insight, which is a direct product of the high-throughput and non-invasive nature of the phenotyping platform.

In **Chapter 4** I use the phenotyping platform to screen three recombinant inbred line (RIL) populations for photosynthesis, growth and reflectance at specific wavelengths of light. With this data I further investigate the fluctuating heritability phenomena described in Chapter 3 and show that the QTLs underlying the trait heritabilities also shift through time. In the case of leaf angle they show a striking diurnal recurrence, while in the case of photosynthesis the influence of a QTL can extend for variable lengths of time, highlighting that the time-point at which a phenotype is measured is a crucial factor in determining which loci are identified.

This time dependency is further explored in **Chapter 5** where I use a GWA approach, which afforded us much greater mapping resolution. In this case I show that one highly significant locus, which encompasses the *accelerated cell death 6 (ACD6)* gene, is only statistically significant from 18 days after sowing. Again, illustrating just how important repeat phenotyping through time is, at most time-point this locus was invisible. In this chapter I also

[13] Specifically intraspecific variation, no C_3 C_4 comparisons here.

describe how the choice of population can affect the mapping results and describe the potential impact of a large effect organellar mutation on the GWA approach. The impact of this organellar mutation is followed up in Chapters 7 and 8.

In **Chapter 6** a different approach is taken to exploring photosynthetic variation. In this case due to constraints on the number of plants which could be screened, I conducted an explorative study into natural variation in phosphorylation of photosystem II (PSII) proteins at two light levels. This was the first study of its kind into this phenotype. Interestingly, I identified a correlation between PSII phosphorylation and temperature seasonality, suggesting that the variation observed may be of adaptive significance.

Chapter 7 was a direct result of the cytoplasmic noise identified in Chapter Five. In order to control for cytoplasmic variation I developed a method to swap the cytoplasm of Arabidopsis plants. This was made possible through the use of a genome elimination line, the *GFP-tailswap* haploid inducer (Ravi and Chan, 2010). Having constructed this population I then screened it for photosynthesis, growth, germination, and primary metabolism. Using these datasets the role that natural genetic variation in plasmotype has to play in plant performance could be explored.

In **Chapter 8**, I went into the field to study the distribution of a natural photosynthetic mutant, which had evolved in response to herbicide application. Initially the goal was to study the cost of photosynthetic impairment in the wild. In the end, however, it became a study documenting the first empirical evidence of an organelle mediated selective sweep distorting nuclear gene frequencies.

In **Chapter 9** I discuss all the results obtained and consider future research avenues which could be explored to take this research further.

A personal narrative

Finally I would like to finish this introduction by giving a personal narrative as to how the research progressed and how the thesis came to be in its current form. As scientists, our goal is objective knowledge but to deny our inevitably subjective nature would be both dishonest and misleading. I often find myself wondering about the situations which led a scientist to take a particular path. It would be naïve to believe that they were based purely on reason and logic, I think being open about the personal aspect of one's research trajectory could aid future readers in objectively weighing what is reported in one's writings[14].

I started this project on the 1st of July 2009 under the initial title of "Genetic analysis of plant plasticity of photosynthetic capacity and its tolerance of a fluctuating environment". The goal of this project was to assess the level of genetic variation present in Arabidopsis for photosynthetic traits under a range of environments. The original plan was to develop a phenotyper and then screen recombinant inbred lines (RILs). The development of the phenotyper took much longer than expected and by the time it was ready our plans had changed. We decided it made more sense to screen a GWA mapping population as we expected greater phenotypic variation and higher mapping resolution, plus it was a sexy new technique. So I set about screening the now famous (at least in the Labs of Genetics and Plant Physiology) HapMap population. Unfortunately I found the initial results less than encouraging, the literature was full of GWA studies where very strong associations were detected but in my data no such peaks were forthcoming[15]. To exacerbate my already faltering faith in GWA many of my fellow PhD students, who were also screening this HapMap population, also failed to detect strong associations. At the time the reason for this was far from obvious, it could be our phenotype data was too noisy, the statistical algorithms were incorrect[16], the population structure was not properly accounted for, or the genetic architecture of the trait was such that strong associations were not to be expected due to the absence of high frequency alleles of an appreciable effect size. To make matters worse I had identified one genotype with a very divergent photosynthetic phenotype only to discover that this phenotypic difference was due to a chloroplast mutation, which was not linked to any of my nuclear markers. Inclusion of this genotype drastically

[14] Of course I recognize that in most cases, i.e. scientific journals, such a detailed personal account would be impractical, however I think the thesis provides a special platform for a more diverse discourse on one's scientific research.

[15] In the end I did find *ACD6* as a strong peak, see Chapter 5, however back then it was impossible to map each time point, it took far too long, so I mapped 10, 14, 16, 18 and 19 days after sowing, in the HapMap the *ACD6* peak is not apparent until after day 20, bad luck I suppose.

[16] The algorithms and software were still in development stage, at this time it took about 24hrs to run one trait.

changed the GWA output resulting in many false positives (see Chapter 5). So I not only doubted the technique but also was very worried about the level of cytoplasmic noise caused by variation in either chloroplast or mitochondrial genomes, both of which contain genes known to be essential for photosynthesis. So, like many PhD students I felt like I had started down an experimental and intellectual cul-de-sac from which there was no return, I was yet another student to be sacrificed on the altar of science (excuse the drama but it did feel that way at times).

Thankfully I did not give up (as the very existence of this thesis can attest to), I decided that of the list of possible reasons for my 'disappointing' GWA results the easiest things to test were my level of phenotypic noise and the genetic architecture of the population. The screening of RIL populations, the original plan, would allow me to kill two birds with one stone, the much greater statistical power afforded by the within population replication of the two parental genomes should allow me to test whether our phenotyper was providing sufficient signal[17] and the simplified genetic architecture (i.e. no population structure or allelic heterogeneity) as well as the uniform cytoplasmic background should allow easier detection of this genetic signal. It worked, I got very significant QTLs which restored my faith in the whole approach (Chapter 4). This success gave me renewed enthusiasm, so much so that I decided to screen another GWA population, the Swedish RegMap. Due to a lower level of genetic diversity, greater linkage disequilibrium, and clearer population structure I hoped it would provide a middle ground of sorts between the HapMap and the RILs. The results of this were again far from clear but bolstered by my success with the RILs I ploughed on and decided to make use of my extensive phenotype datasets to best effect, employing a sort of bootstrapping of my GWA results by comparing the occurrence of SNPs at various time points and across experiments[18] (Chapters 4 and 5). Although my GWA doubts were partially alleviated by the work with the RILs I was still nagged by this persistent worry of variation in the organellar genomes giving rise to false associations. The solution came in a *eureka* moment during a conversation with a fellow PhD student, Erik Wijnker, who was working with a genome elimination mutant as part of his research. Erik had realized that the double haploids generated from this had the plasmotype of the mutant line and the nuclear genome of the male plant, thus we could make the plasmotypic background of the entire GWA panel uniform. It was only a small additional

[17] Due to this fixation on phenotype quality I decided to screen the core set of 164 lines for each population, instead of the full set of 300+. In retrospect this was a mistake as more genotypes would have given better mapping results, on the plus side, 8 reps per genotype meant I could map trait variance, which generated some nice additional loci.
[18] By this stage we could map a single trait in less than an hour and could set up the script to loop over many time points, facilitating analysis of an entire time series overnight.

step for me to propose the reciprocal via introgression of the causative mutation into different plasmotypic backgrounds, the results of these experiments are described in Chapter 7. The final additional research theme also resulted from this cytoplasmic mutation. The original accession had been collected from a railway line in England in 1988, the mutation afforded resistance to a herbicide, which was being sprayed at that time but discontinued in 1992. I happened to be passing by this station in August 2012 and decided to check and see if I could get some seed to test if the mutation persisted in the absence of the herbicide, lo and behold it was, this curiosity driven discovery led to my final and, in my opinion, most fun research chapter.

Chapter 6 was, in some ways, most in keeping with my original research goal, i.e. analysis of genetic variation of photosynthesis under fluctuating conditions, in this case light. This chapter was my first experience of collaborating with people outside my lab and with whom my only contact was via email. The collaboration was initiated by my promoter Maarten Koornneef who was approached at a conference in Japan by Cornelia Spetea. She had noticed that the backgrounds of her mutant lines differed substantially in their wild-type PSII phosphorylation levels. Natural variation in the phosphorylation of PSII proteins was an unexplored field and as such it was exciting research to be part of. Due to the difficulty in phenotyping it was not possible to screen anywhere near the number of lines necessary for genetic mapping, even so it was a very interesting experience as I learned much more about the regulation of photosynthesis and also, I gained experience of working in a long distance international collaboration.

I hope it is now clearer why a thesis on the "Genetic analysis of plant plasticity of photosynthetic capacity and its tolerance of a fluctuating environment" branched out to include cyto-nuclear interactions, PSII phosphorylation, and herbicide resistance. They are all linked to the more general theme of natural genetic variation in plant photosynthesis and this, with the Arabidopsis qualifier, is the final title of this work. Jan Baptist van Helmont's ground-breaking experiment in the 1600s was the first step on the long path to understand photosynthesis, a journey which is still not finished as much remains to be discovered. I hope this thesis will serve as another step on that path.

Chapter 2

Natural genetic variation in plant photosynthesis

Pádraic J. Flood[1,2], Jeremy Harbinson[2] & Mark G.M. Aarts[1]

[1]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

[2]*Horticulture and Production Physiology, Wageningen University, Wageningen, the Netherlands.*

Corresponding Author: Pádraic J. Flood (flood@mpipz.mpg.de)

Abstract

Natural genetic variation in plant photosynthesis is a largely unexplored and as a result an underused genetic resource for crop improvement. Numerous studies show genetic variation in photosynthetic traits in both crop and wild species, and there is an increasingly detailed knowledge base concerning the interaction of photosynthetic phenotypes with their environment. The genetic factors which cause this variation remain largely unknown. Investigations into natural genetic variation in photosynthesis will provide insights into the genetic regulation of this complex trait. Such insights can be used to understand evolutionary processes that affect primary production, allow a much greater understanding of the genetic regulation of photosynthesis and ultimately increase the productivity of our crops.

A new angle to photosynthesis research

Natural genetic variation (see Glossary) is the fuel of evolution; without it there can be no natural or artificial selection, no genetic drift and no adaptation of species to changing environments. Genetic variation consequently is a vital characteristic of any population which is required to adapt. Photosynthesis is one of the key physiological processes on Earth; in its absence, life as we know it could not exist. It is also the most complex energy transducing process (Shi et al., 2005), occurring throughout the illuminated biosphere, and as such exhibits great variability. As in all cellular processes, photosynthesis is genetically controlled and the genetic blueprint for photosynthesis varies both within and between species (Arntz and Delph, 2001; Hikosaka, 2010). The adaptive potential of photosynthesis to changing environments depends on the level of genetic variation for photosynthesis which is present within a population. Changing atmospheric levels of CO_2 and increasing demands of a growing world population mean that the photosynthetic adaptability of our crop and natural ecosystems is being tested.

The importance of photosynthesis arises from the role that it plays in the productivity of both natural and agricultural ecosystems. Because plant biomass is largely derived from photosynthetically captured carbon (Robson, 1973), variation in the efficiency or capacity of photosynthesis can lead to variation in growth rate and productivity, which are important factors in species competition and crop yield (**Box 1**). It is becoming more apparent that photosynthesis will have a key role to play in any sustainable society. To this end there have been several recent reviews both on the importance of photosynthesis and putative ways in which it might be improved in order to benefit human society (Peterhansel et al., 2008; Murchie et al., 2009; Sun et al., 2009; Zhu et al., 2010). By and large these reviews had a biotechnological focus, considering the ways in which photosynthesis could be engineered to suit our needs. The potential of genetic diversity in photosynthesis has been largely ignored in recent years, with the last review on this topic published over 20 years ago (Austin, 1989). In this review we will show how valuable a resource natural genetic variation in photosynthesis is, and that investigating natural variation not only highlights the diversity present, which can be of potential use in crop improvement, but it also can provide insights into the evolution and genetic regulation of this complex trait (Alonso-Blanco et al., 2009; Jung and Niyogi, 2009), and into species adaptation to varying environments (Zou et al., 2007; Mozdzer and Zieman, 2010).

Levels of photosynthetic variation

The nature and significance of variation in photosynthesis is dependent on the unit of measurement chosen. The photosynthesis of a community of plants such as a crop can be quantified as the photosynthetic rate per unit ground area, the rate per individual or the rate per unit area of the leaf. Each of these has particular uses and applicabilities and while it is relatively easy to translate rate per unit ground area to rate per plant and vice versa, the rate per unit leaf area is not easy to translate to either of the other two. Photosynthesis on a unit leaf area basis needs to be integrated to give the rate per plant or unit area of the crop, this requires knowledge of factors such as leaf area index (LAI), the variation of photosynthetic properties on an area basis between leaves, and the architecture of the canopy (**Box 1**). The standard way of describing photosynthesis is as a rate of CO_2 fixation per unit leaf area or similar, with 'better' photosynthesis being that with the highest rate per unit area under specified circumstances. When considering variation in the properties of photosynthesis we will focus on those properties expressed on a unit leaf area basis. We do this because this is how the activity of the photosynthetic process is best described and how it is most easily understood in physiological terms. In the other metrics of photosynthesis, such as the rate of carbon dioxide fixation per plant or per unit area of crop, the activity at the leaf level is combined with other factors, such as plant morphology and architecture, that are non-photosynthetic in character. Plant architecture and LAI are important traits but can be considered separately to the photosynthetic capacity per unit area of individual leaves. Also plant architecture and LAI have been brought close to optimal levels in the major crop plants by breeders and thus show limited potential for future crop improvement (Long et al., 2006; Parry et al., 2010). This in contrast to photosynthesis per unit leaf area, which has remained largely unexploited. Recently natural genetic variation in the leaf architecture of maize has been studied and used to describe the genetic nature of this trait and identify key genes (Tian et al., 2011) which illustrates the potential of natural genetic variation as a tool for understanding such complex traits.

An additional complication with measuring photosynthesis is that of sample size. Often for practical reasons photosynthetic rates are measured on only one leaf and at only one time. Photosynthetic rate however, varies between leaves and with time in individual leaves. The tendency to use the photosynthetic rate per unit leaf area of a few leaves, or even just a portion of one leaf, as a measure of the photosynthetic capacity of the plant as a whole, has led to a lack of correlation between photosynthetic capacity and plant yield (**Box 1**). The measurement of photosynthetic properties of leaves ideally requires a rigorous measuring regime per plant. This has made it difficult in the past to fully describe the photosynthetic properties of plants when large populations need to be

investigated. In order to accurately estimate the photosynthetic capacity of a plant care must be taken to adequately sample enough leaves, the necessary representative sample size will vary depending on factors such as canopy architecture, leaf age distribution etc. Since leaves are the basic unit of most photosynthetic measurements, particularly at the crop or ecophysiological level, the source of leaf level photosynthetic variation is of great interest. In this review we will focus on variation in the photosynthetic properties of plants at the leaf level and we will not consider those other factors such as plant architecture that also influence plant productivity.

What is the cause of photosynthetic variation?

An important factor of photosynthesis is its flexibility. All land plants use the C_3, or the allied C_4 or CAM, photosynthetic pathways. Considering, for the moment, just C_3 plants, it is clear that despite using the same basic physiology there is considerable variation in the photosynthetic properties of leaves both within and between species. This variation is well understood in terms of physiology and leaf anatomy and their interactions with the environment, but the genetic basis of these responses and how it has been shaped by artificial and natural selection is poorly understood. For plants using the C_4 or CAM systems, the same arguments apply: there is anatomical and physiological variation that has its origins in poorly understood genetic variation. In these cases the genetic changes that drove the evolution of these more specialized forms of photosynthesis from a C_3 precursor are of considerable interest.

Important questions regarding genetic variation in photosynthesis are:

- how do these genes give rise to this variation;
- which genes are involved;
- is photosynthetic variation caused by variation in a multitude of genes, or mainly caused by variation in only few genes?

The genetic blueprint or genomic DNA sequence tends to vary between individual plants. Often this has no discernible effects on the phenotype, but if it has, such variants can be subject to natural selection. In the case of photosynthesis, genes encoding the core metabolic processes e.g. for proteins in linear electron transport in Cyanobacteria (the ancestors of modern chloroplasts) are known to be highly conserved and therefore subject to limited levels of variation (Shi et al., 2005). This is due to the complex nature of photosynthesis requiring numerous protein–protein interactions, which has been shown to limit the rate of evolution (Shi et al., 2005). Since the core genes have

limited levels of variation, the heritable variation for photosynthesis observed in various crop and ecophysiological studies (**Table 1.1**) must come from variation in other genetic factors. It is worth noting that the genetic code of a plant is comprised of nuclear and cytoplasmic DNA. Whilst the focus of this review is on variation in nuclear DNA sequence which can be analyzed using modern techniques such as association mapping (**Box 2**), it is important to recognise that essential photosynthesis genes are encoded in the chloroplasts, which will not be analyzed using such techniques, but for which there is natural genetic variation (El-Lithy et al., 2005).

The causes of variation in photosynthetic properties per unit area of individual leaves can be divided into two main categories, morphological and physiological, which are both influenced by environment and genetics. The difference between sun and shade leaves of the same plant shows how plastic plant phenotypes can be within one genotype and it is a classic example of an environmental influence resulting in different leaf morphologies and photosynthetic properties (Boardman, 1977). Genetically based differences in leaf morphology are commonly encountered at the interspecific level, and often correlate with the growth habit of the species in question (Hikosaka, 2010; Donovan et al., 2011). For example, leaves of evergreen perennials generally last much longer than those of herbaceous annuals (Wright et al., 2004). This difference in longevity is associated with other important differences in leaf functionality that altogether give rise to two contrasting solutions to the question how best to optimize investment of resources (photosynthate, minerals) into leaves. Long-lived evergreen leaves need to be much more structurally durable and as such are often thicker, with a higher leaf mass per unit area, lower mesophyll conductance and as a result, lower photosynthetic rates (Flexas et al., 2008) and lower nitrogen use efficiency. Short-lived leaves of herbaceous plants usually invest less resources in durability and more in photosynthetic machinery, and thus have higher photosynthetic rates (Onoda et al., 2004; Wright et al., 2004) and higher nitrogen use efficiency. It is interesting to note that although the short-lived leaves may have higher rates of photosynthesis, the energy return per unit invested is often higher for long-lived leaves (Westoby et al., 2000; Wright et al., 2004).

In addition to variation in morphology there is also variation in the physiology of leaves, which also affects their photosynthetic properties per unit area. This physiological variation is unlikely to be due to major differences in the properties of structural proteins, and appears to be largely due to differences in (relative) protein abundances. Variation in the relative abundances of the core components will give rise to different photosynthetic phenotypes (Schöttler et al., 2004; Zhu et al., 2007).

Another important cause of photosynthetic variation can be found in the many associated processes which can influence the rate of photosynthesis. For example, the rate of sucrose export from the mesophyll cells and leaves is important in allowing sustained photosynthesis by leaves. A build-up of sucrose in the mesophyll cells can lead to a reduction in the rate of photosynthesis, due to negative feedback, largely mediated by phosphate sequestration (Stitt, 1991). Thus variation in the enzymes or plant structure involved in sucrose transport and use is likely to have an effect on photosynthesis. Photosynthetic nitrogen use efficiency has also been shown to be an essential physiological parameter influencing photosynthetic capacity and rate (Evans, 1989; Hikosaka, 2004). In essence this is due to the variation in allocation of nitrogen to the different cellular components and the efficiency of the morphological structure allowing for improved physiological performance under nitrogen limiting conditions(Hikosaka and Shigeno, 2009).

In recent years much work has been done to investigate the physiological and morphological causes of interspecific variation in photosynthesis per unit leaf area (Hikosaka, 2010). Taken together, these causes of variation in C_3 species are similar at the interspecific and intraspecific level. The research discussed so far illustrates the many levels, from morphology to physiology, which influence a plant's photosynthetic phenotype. This detailed understanding of the morphological and physiological influences on photosynthesis now requires in-depth understanding of the underlying genetics in order to complete the picture.

Investigations into natural genetic variation in photosynthesis

Past research into intraspecific genetic variation in photosynthesis per unit leaf area falls largely into two categories, those concerned with crop and yield related traits and those whose focus was more on ecophysiology and evolution. As outlined in a recent review, both agricultural and evolutionary studies provide valuable insights into the biology of many species and should be used in tandem(Moyle and Muir, 2010). This is particularly pertinent when studying photosynthesis: ecophysiologists show that natural selection occurs and agronomists show that artificial selection is possible.

Ecophysiological studies

The evolution of photosynthesis in natural populations is of great interest to ecologists and evolutionary biologists alike (Arntz and Delph, 2001), but in order for such evolution to occur genetic variation for the photosynthetic phenotype must be present within a species. A recent phylogenetic

study on the large subunits of Rubisco, which is the most abundant enzyme in nature catalyzing the first step in carbon fixation, showed that widespread positive selection has occurred at the interspecific level (Kapralov and Filatov, 2007). This is important as it shows variation and selection is occurring even in enzymes as crucial and highly conserved as Rubisco. Recently this has been investigated in more detail for a range of other metabolic reactions associated with growth and photosynthesis (Sulpice et al., 2010). In addition to this, a deletion-substitution event in the nuclear genes coding for the small subunits of the Rubisco complex has been shown in *Arabidopsis thaliana* (Schwarte and Tiedemann) indicating intraspecific variation for these proteins, although the functional impact of this genetic difference is unclear. It should also be made clear that while there appears to be structural and functional variation in key components such as Rubisco, there is also evidence to demonstrate that variation in the amounts of such proteins has a major impact on the photosynthetic rate (Pettigrew and Turley, 1998). However, since both the structural composition and relative abundance of photosynthetic components are genetically controlled the genetic factors responsible can be elucidated via a quantitative genetic approach due to its unbiased nature.

The photosynthetic rate per unit leaf area of individuals from several plant species has been measured (**Table 1**), with a variety of goals in mind. Most interestingly, several studies have shown that higher photosynthetic rates often correlate with higher fecundity and invasive potential (Arntz et al., 2000; Arntz and Delph, 2001; Zou et al., 2007; Mozdzer and Zieman, 2010). This can lead one to the conclusion that higher photosynthetic rates are desirable in nature, however this is likely not always the case; the studies mentioned above have largely looked at fast growing species in nutrient rich environments, and for slower growing species in more limiting environments such high rates of photosynthesis can be disadvantageous due to the higher demands it places on the plant in terms of nutrients and leaf structure. This ecophysiological research illustrates the increasingly detailed understanding we now have about the many different strategies plants employ to optimize their photosynthetic phenotype. Ecophysiological studies show that there is both variation and selection for photosynthetic traits in nature; this is very encouraging as it indicates a wealth of potential genetic resources for future improvement and adaptation of photosynthesis to changing environments.

Agronomic studies

Numerous studies have investigated photosynthetic variation in crop species (**Table 1.1**), and selection using this variation has been successfully carried out in the past (Crosbie et al., 1981) . However this approach to select photosynthesis directly was abandoned in crop breeding because it made little or no difference to yield (Crosbie and Pearce, 1982). The absence of any link between

improved photosynthesis per unit leaf area and yield is, at first glance, surprising. The resolution of this conflict lies with the method used to measure photosynthesis and thus provide the criterion for selection (see **Box 1** and **Box 3** for more details). For example, a study on maize (*Zea mays*) (Crosbie et al., 1981) showed that selection for photosynthetic rate per unit area is possible. However for technical reasons measurements were made on detached maize leaves transported to the lab, which is not an accurate indication of field photosynthetic performance. Also only maximum values of CO_2 fixation were determined, so no information on the photosynthesis per unit leaf area of these different genotypes under low, or fluctuating, light conditions, or additional environmental fluctuations, all of which will be encountered in the field, could be obtained. Finally the authors made no analysis of carbohydrate partitioning and leaf area index. So even though an increase in the carbon exchange rate was observed at the leaf level, other essential yield related traits were not co-selected, hence the likelihood of yield improvement was low. In another study, on soybean (*Glycine max*), improved photosynthetic rates were successfully selected for in test crosses. Since, however, the heritability of the trait was relatively low (Ojima, 1974), it was not straightforward to reliably and efficiently introgress the trait into elite cultivars due to the lack of molecular markers that currently facilitate selection of specific alleles. By the end of the 1980's most research aimed at breeding for 'better' photosynthesis had ceased. Recently though this has begun to change, particularly in cotton where an effort is being made to understand both the level of genetic variation present for leaf level photosynthesis and its main physiological causes (Pettigrew, 2004; Pettigrew and Gerik, 2007).

Table 1.1. Past investigations into genetic variation in plant photosynthesis

Quantitative genetic studies		
Species	**Traits investigated**	**Refs.**
Arabidopsis thaliana	Thermal dissipation, Rubisco small subunit diversity	(Jung and Niyogi, 2009; Schwarte and Tiedemann)
Rice (*Oryza sativa*)	Chlorophyll content, photosynthesis under water stress, leaf and flag leaf photosynthetic rate	(Teng et al., 2004; Yue et al., 2006; Hu et al., 2009; Takai et al., 2010)
Maize (*Zea mays*)	Cold tolerance of photosynthesis	(Fracheboud et al., 2002)
Wheat (*Triticum aestivum* L.)	High light stress, leaf chlorophyll content, chlorophyll fluorescence, dry matter accumulation	(Yang et al., 2007; Zhang et al., 2009; Li et al., 2010; Liang et al., 2010)
Soybean (*Glycine max*)	Chlorophyll fluorescence, RUBISCO activase	(Yin et al., 2010; Yin et al., 2010)
Barley (*Hordeum vulgare*)	Flag leaf morphology, chlorophyll content, chlorophyll fluorescence	(Guo et al., 2008; Xue et al., 2008)
Cabbage (*Brassica oleracea*)	Photosynthesis and water use efficiency	(Hall et al., 2005)
Cotton (*Gossypium ssp.*)	Effect of QTLs for productivity and drought on photosynthesis	(Levi et al., 2009)
Sunflower (*Helianthus annuus*)	Chlorophyll concentration, net photosynthesis, internal CO_2 concentration	(Hervé et al., 2001)
Ecophysiological studies		
Species	**Observations**	**Refs.**
Chenopodium album	Considerable genetic variation in photosynthesis found between populations, important for adaptation to new environments	(Haraguchi et al., 2009)
Sapium sebiferum	Invasive genotypes had higher photosynthesis than genotypes from place of origin.	(Zou et al., 2007)
Phragmites australis	Invasive genotypes had higher photosynthetic rates than native ones, in the same range.	(Mozdzer and Zieman, 2010)
Plantago major	Variation in growth rate and its relationship to photosynthesis	(Dijkstra and Lambers, 1989)
Pseudotsuga menziesii	Show genetic variation in carbon isotope discrimination and gas exchange in geographically diverse populations	(Zhang et al., 1993)
Carex aquatilis	Geographical variation in photosynthesis	(Chapin Iii and Oechel, 1983)
Metrosideros polymorpha	Photosynthetic variation along an altitudinal gradient	(Martin et al., 2007)

Robust and reproducible phenotyping

In order for any genetic analysis to proceed, the feasibility of phenotyping large, genetically coherent populations will need to be addressed. Such high-throughput phenotyping might prove to be the limiting factor in the exploitation of genetic variation in photosynthesis if steps are not taken to develop such technology. The three main considerations in the development of such phenotyping platforms are: (i) Reproducibility; any data generated should be indicative of genetic variation, and non-genetic variation caused by uncontrolled fluctuations in the environment should be minimised. This will require close attention to the homogeneity of both growth and measurement conditions. *(ii)* Speed; in order for genetic studies to be conducted, large numbers of plants (and leaves) need to be frequently and efficiently screened. There is great interest in developing high-throughput phenotyping platforms (Granier et al., 2006; Jansen et al., 2009)(see: http://www.plantphenomics.com) including developments by commercial companies. Especially when large plants are involved this requires the movement of the plant to the phenotyping centre. However, such movement should be avoided if possible as it can have unexpected physiological effects, especially on photosynthesis, as many plants will close their stomata when moved if not previously habituated to movement. Preferably phenotyping systems should be developed that do not involve the movement of plant but instead the equipment. This represents a considerable engineering challenge which needs to be overcome. (iii*)* Non-invasiveness; repeated measurements need to be made throughout the plant's life-cycle, and those which have a minimal effect on the plant are the most suited for the accurate estimation of plant phenotypes. Chlorophyll fluorescent imaging techniques allow such rapid non-contact measurement of photosynthesis (**Box 3**). Finally, phenotyping that has been done in controlled environments, in order to minimise noise, will need to be confirmed under field conditions in order to be of use in crop improvement. Note, however, that in itself the tolerance of photosynthesis to environmental fluctuation is a valuable trait.

How to investigate natural genetic variation in photosynthesis

So far we have shown that genetic variation for photosynthesis exists and selection of genetic variants is possible, but how can such variation be investigated? We propose that a quantitative genetic approach will be very useful in properly assessing both the genetic causes of photosynthetic variation at the leaf level and in designing future breeding programs. Initial insights from a detailed genetic study of natural variation in photosynthesis will, however, be fundamental in nature and not targeted to any particular aspect of photosynthesis, as quantitative genetic studies are essentially an unbiased

approach to unravel the genetic network underlying a trait (**Box 2**). Depending on which phenotypes will be scored, studies such as genome wide association mapping or quantitative trait locus (QTL) linkage mapping will provide novel information on genetic loci containing one or more genetic factors that affect photosynthesis in one way or another. Identified QTLs point in the direction where more detailed analysis has to be undertaken, such as detailed fine mapping, to indeed identify the genes or genetic factors underlying the QTL. Such analysis is not trivial for most plant species, such as crops, for which genome sequence and genomic tools are not readily available, compared to the few model species for which this is considerably easier such as *Arabidopsis* and rice (*Oryza sativa*) (Alonso-Blanco et al., 2009).

An important advantage of quantitative genetic studies is that many variant alleles can be found when genetically amenable populations are queried under the right conditions. In most cases this variation would not have been easily detected by mutant analysis because variant alleles often have only a minor effect on the phenotype of the plant (Alonso-Blanco et al., 2009). In the progeny of a biparental cross between homozygous parents, each allele will be present in an approximately 50:50 distribution. This high level of replication means that even minor effect allelic variants can be detected (**Box 2**). These variant alleles will likely not be found in the conserved core photosynthetic genes but rather in genes for auxiliary factors or modifiers of photosynthesis. Potentially these can be more interesting than core photosynthetic genes as such modifiers are likely to be novel, because they are unlikely to have been identified in previous mutant screens. Identification of modifiers of photosynthesis will expand our knowledge on the genetic regulatory network controlling photosynthesis in response to environment and differences in plant architecture, aimed at optimized photosynthesis for each cell. Variant alleles will also indicate those core components of the photosynthetic process for which genetic variation has occurred and for which allelic variants are subject to natural selection. Such loci are interesting targets for plant breeding, as there is likely to be genetic variation for them in crop species. This variation will reflect local adaptations, but can equally well indicate alleles that contribute to robustness of photosynthesis under different or varying environments. Especially rewarding in this respect will be photosynthetic phenotyping of plants grown under adverse or varying conditions, such as frequently found under natural or field, conditions: fluctuations in (daily) temperature, precipitation, light intensity, water supply and nutrient availability, amongst others.

Once this has been done for one species, for example a model species like *Arabidopsis* or rice, it will become increasingly clear which genes play a crucial role in fine-tuning photosynthesis to accommodate the fluctuating environments and provide an optimal phenotype of this species under

the given local circumstances. When combined with additional genomics technologies such as expression QTL or metabolite QTL analysis (Keurentjes et al., 2006), it will be possible to develop a list of candidate genes that can be used to genetically describe photosynthesis in terms of the genes involved; their interactions; and their regulations. Thus it will be possible to create and expand the genetic (regulatory) networks (Jansen et al., 2009; Hall et al., 2010) visualizing the interaction between different genes influencing photosynthesis. Together with physiological studies this will enable a more systematic view on photosynthesis, showing the interaction of photosynthesis with other physiological processes in the plant and indicating where variation is present and can be selected for in order to improve photosynthesis in crop species (Langridge and Fleury, 2011). It will also allow the unraveling of the genetic basis of the specific physiological adaptations found in plants displaying extreme photosynthetic phenotypes. A recent paper on natural genetic variation in enzyme activities and metabolite levels, shows the potential of the genetics and genomics approach to construct genetic regulatory networks (Sulpice et al., 2010).

A more detailed genetic understanding of photosynthesis will open doors to many avenues of application. If the genetic factors that give rise to specific photosynthetic properties can be identified, then the evolution of photosynthetic characteristics at the ecotype level can be understood. This will be of interest to population geneticists or ecologists wishing to understand how species respond to environmental change (Keurentjes et al., 2011). More practical implications of a detailed understanding of the genetic regulation of photosynthesis will aid modelers wishing to follow a systems biology approach to optimize crop management and make more accurate predictions on genotype-by-environment interactions (Yin and Struik, 2010). These benefits are likely to be exceeded by the potential for enormous improvement in crop breeding strategies aimed at quantitative traits such as photosynthesis (Langridge and Fleury, 2011), which is essential for future yield increases in both food and third generation biomass crops. Genetic dissection of the trait will allow for identification of alleles of key genes which can then be selected for and introgressed into elite lines (Langridge and Fleury, 2011). After almost a 20 year hiatus (Austin, 1989), breeding for photosynthesis is going through a renaissance with the aid of such genetic knowledge and know how.

Conclusions

Natural genetic variation in photosynthesis occurs in both crop and wild species under field conditions. Modern genetics and high throughput phenotyping allow a detailed genetic investigation of the natural genetic variation in photosynthesis per unit leaf area for a wide range of species. This will extend our understanding of the regulation of photosynthesis in plants, provide insight into the selective forces which act upon this trait in natural and cultivated populations, and provide the necessary knowledge for molecular breeding programs (Keurentjes et al., 2008; Langridge and Fleury, 2011). The incorporation of this knowledge into breeding programs should lead to an increase in crop photosynthesis. This increased photosynthesis per unit leaf area will ultimately increase plant yields provided other yield related variables such as harvest index do not deteriorate. Improving photosynthesis is a challenging task but will become more and more essential if the necessary yield increases are to be achieved. Our current yield barriers must be broken (Phillips, 2010) and photosynthesis is one key which cannot be left unused.

Acknowledgements

We would like to acknowledge Maarten Koornneef and Erik Wijnker, for their helpful comments and discussion during the preparation of this manuscript. In addition we acknowledge the Netherlands Organisation for Scientific Research (NWO) and the Technological Top Institute Green Genetics for their financial support.

Glossary

Gene: DNA sequence containing a protein coding region and all regulatory elements in the direct vicinity needed for transcription of this protein coding region.

Genetic factor: Any DNA sequence that contributes to the phenotype of an individual, as it comprises a gene or sequences contributing to the expression of a gene.

Genome Wide Association Studies (GWAS): is a quantitative genetic approach used to find genetic associations between genotype and phenotype in a population of individuals of unknown relatedness in order to identify genetic loci contributing to such phenotype.

Genotype: is the specific genetic constitution of an individual, determined by its nuclear and cytoplasmic DNA sequence.

Leaf area index (LAI): is a measure of the total leaf area per unit ground area.

Natural Genetic Variation: is the genetic variation that occurs both within (intraspecific) and between (interspecific) species. In this review we focus on spontaneously generated (natural) genetic variation found in crop as well as wild populations as opposed to variation generated in the lab via mutagenesis or transgenesis (Alonso-Blanco et al., 2009).

Photosynthesis (oxygenic): is the process by which plants, eukaryotic algae and Cyanobacteria use light in the approximate wavelength range of 400-700nm to split water, forming O_2 and reducing CO_2 to an organic form, thus converting some of the absorbed light into chemical energy.

Photosynthetic rate: the rate of CO_2 fixation per unit area of leaf, usually expressed as $\mu mol\ m^{-2}\ s^{-1}$; because photosynthesis occurs simultaneously with respiration, the rate of photosynthesis can be qualified as either a gross or net rate.

Phenotype: is the physical appearance of an individual as a consequence of its genotype and its environment.

Pmax: Also known as A_{max}, is the maximum photosynthetic rate of a leaf under light saturated conditions, this can also be thought of as the photosynthetic capacity of the plant.

Photosynthetic nitrogen use efficiency (PNUE): is the efficiency with which a plant can use the available nitrogen in its leaves or other photosynthetic tissues for photosynthetic processes.

Quantitative Trait Loci (QTL): are genetic loci each corresponding to a region of the genome containing genetic factors for which variation can be found between the studied genotypes, which have a statistically significant association with a particular phenotype, and which trait values are expressed in a continuous rather than discrete manner.

Single Nucleotide Polymorphisms (SNPs): are single base pair variants commonly used as genetic markers in quantitative genetic investigations.

Box 1. Photosynthesis and plant productivity

The relationship between photosynthesis and plant growth is not simple and has been hotly debated for many years (Demetriades-Shah et al., 1992; Monteith, 1994). The crux of the argument is, particularly in relation to this paper, will an improvement in photosynthesis correspond to an improvement in plant biomass accumulation? In order to answer this question it is important to be clear what is meant by photosynthesis. In this review we focus on the photosynthetic rate per unit leaf area, as this is the metric we wish to investigate using genetic variation, however when comparing the total photosynthetic capacity of one plant against another it is not the only consideration. The model of primary production per unit area and time (P_n kg m^{-2}) proposed by Monteith and Moss (1977) and summarized by Long et al. (2006) clearly shows the major factors associated with biomass accumulation in plants.

$$P_n = S_t \cdot \varepsilon_i \cdot \varepsilon_c / k$$

Where S_t represents the integrated irradiance over time (usually sunlight) per unit area (MJ m^{-2}), ε_i is the efficiency with which the radiation is intercepted, ε_c is the efficiency with which the intercepted radiation is converted to biomass and k is the energy content of the plant mass (MJ kg^{-1})(Long et al., 2006).

This formula shows the importance of both light capture (ε_i) and conversion (ε_c) to biomass production at the crop level. Both ε_i and ε_c are canopy level parameters that arise from the physiological and physical properties of leaves (ε_i and ε_c) and the arrangement of the leaves in the canopy (ε_c). While the physiological and related physical properties of leaves are important in determining biomass accumulation, account needs to be taken of factors such as canopy architecture that affect ε_i if productivity is to be fully understood. The photosynthetic properties of individual leaves are often different and vary with time, consequently, obtaining a comprehensive description of photosynthesis at the leaf level requires an extensive measurement campaign, which is often not feasible. Despite the complexity of the relationship between leaf level photosynthesis and productivity there are examples where a clear relationship has been shown. These examples and the potential benefits of improvements to photosynthesis on plant productivity have been discussed in detail elsewhere (Ojima, 1974; Zelitch, 1975; Monteith and Moss, 1977; Zelitch, 1982; Lambers, 1987; Peng et al., 1991; Evans, 1997; Horton, 2000; Richards, 2000; Sharkey et al., 2000; Nunes-Nesi et al., 2005; Kruger and Volin, 2006; Long et al., 2006; Parry et al., 2007; Sheehy et al., 2007; Peterhansel et al., 2008; Melis, 2009; Murchie et al., 2009; Sun et al., 2009; Amthor, 2010; Black et al., 2010; Parry et al., 2010; von Caemmerer and Evans, 2010; Zhu et al., 2010; Kirschbaum, 2011;

Raines, 2011). Also several studies on rice and wheat varieties released since the green revolution show that yield increases were dependent upon harvest index until the 1980's, after which the relative importance of biomass accumulation and photosynthesis began to increase. This indicates a transition from sink to source limitation in two of our major C_3 crops (Shearman et al., 2005; Hubbart et al., 2007), making the case for improvement of photosynthesis stronger. The rationale of one of the major scientific endeavors of our time, the IRRI C_4 rice project, is largely based on the proposition that improving leaf level photosynthesis will increase yield in rice (Sheehy et al., 2007; Hibberd et al., 2008). In summary, an increase in the rate of photosynthesis per unit leaf area will result in a corresponding increase in biomass accumulation should all other factors (leaf area, light interception etc.) remain constant.

Box 2. Quantitative Genetic Analysis

"Plant height", "number of seeds" and other traits for which phenotypes are expressed in continuous values rather than few discrete classes are known as quantitative traits. Often such traits are controlled by multiple genes or genetic factors, making them difficult to analyse. This complexity has many layers. First the genes involved generally have different alleles, occurring in different combinations in natural populations. Second, these genes interact with each other, affecting the individual contribution of alleles to the phenotype. Third, different environments will have a different impact on the phenotype of one genotype.

Quantitative traits can be genetically dissected to identify the genetic factors involved down to the DNA base pair level. Although genetic factors are often genes in the conventional sense, i.e. they encode a protein, they can also be sequences regulating epigenetic state or sequences encoding micro RNAs, having an indirect effect on the expression of a gene and often affecting gene expression from considerable distance of the target. The genetic analysis of quantitative traits (quantitative trait locus (QTL) analysis) is based on the statistically significant detection of association of a particular phenotype with the presence of a particular region of the genome. Genomic regions can be characterized by genetic markers, currently often single nucleotide polymorphisms (SNPs). QTL analysis can be conducted on genetically segregating populations with known ancestry, such as the F2, F3 or Fn progeny of a biparental cross or back cross (QTL linkage mapping), or on large collections of diverse natural accessions with unknown ancestry (Genome wide association studies (GWAS)) (Myles et al., 2009).

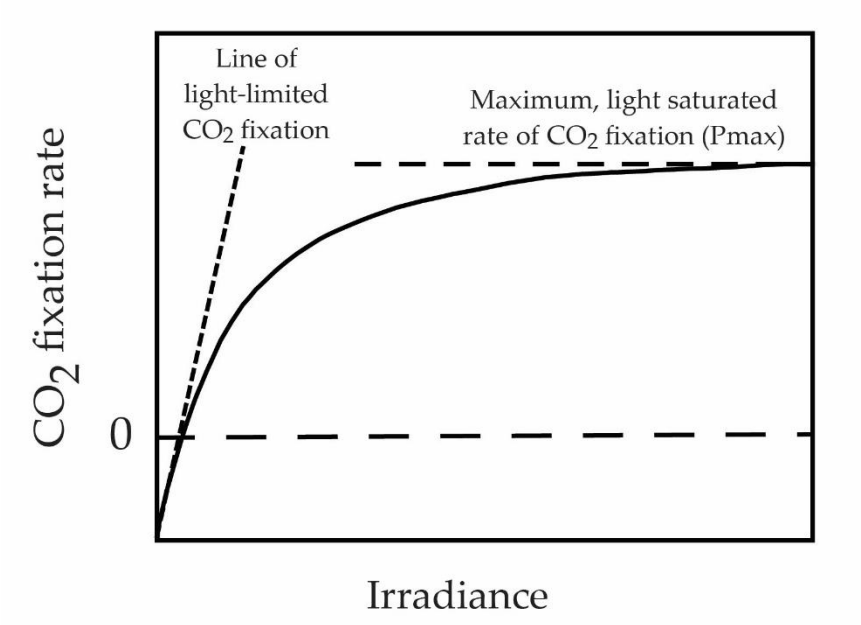
The analysis of quantitative traits using either QTL linkage mapping or GWAS, has been explained in detail in a recent review (MacKay et al., 2009). The generation of large numbers of genetic markers, such as SNPs, gave an important impulse to the development of GWAS. It is especially suitable for genetic analysis of large natural populations of organisms with limited offspring, such as humans, because it does not rely on a large genetically segregating population originating from few parental individuals (Weiss, 1998). The genetic resolution is largely determined by the number of markers and the number of genotypes, as with QTL linkage mapping, but also by the decay of linkage disequilibrium (LD), i.e. the distance from a marker at which genes cease to be physically (and genetically) associated with that marker. If these regions of LD are smaller, more markers are needed for genotyping but with these markers the genetic resolution will be higher. As high-throughput plant genotyping and phenotyping capabilities and the statistical analysis of association become increasingly sophisticated and available, GWAS is gaining in importance. For more information on these new developments in genomics-based population genetics we refer to an excellent recent review (Hamblin et al., 2011). For *Arabidopsis* a collection of close to 200 accessions, genotyped for 250,000 SNPs, has already been used to resolve a disease resistance locus down to the gene level (Atwell et al., 2010). Comparable association mapping populations are now also generated for rice (McNally et al., 2006) and maize (McMullen et al., 2009).

The possibility to resolve genetic variation at the gene level offers an important alternative to forward genetic mutant screens to identify genes involved in a certain process, with the additional advantage that new genes can be found that often would not be found in a mutant screen. Especially when they do not confer a strong deviant phenotype as a single (knock-out or activation tagged) mutant, the low frequencies in which they occur in mutagenized populations (typically less than 1 in 10,000), mean such mutants are generally not distinguished from wild types.

Box 3. Measuring and parameterising photosynthesis.

In order to conduct a genetic analysis of a complex trait like photosynthesis it is necessary to quantitatively describe the trait by means of parameters. These are quantities that summarise the capacity or extent of key functions or properties of the trait, for example the parameters of the Michaelis-Menten equation, V_{max} and K_m, describe the catalytic properties of an enzyme. Parameterisation of photosynthesis requires finding values that concisely summarise complex physiology, such as responses to irradiance, CO_2 concentration, temperature, nutrient availability, water supply etc. Given the complexity both of photosynthesis and its response to environmental

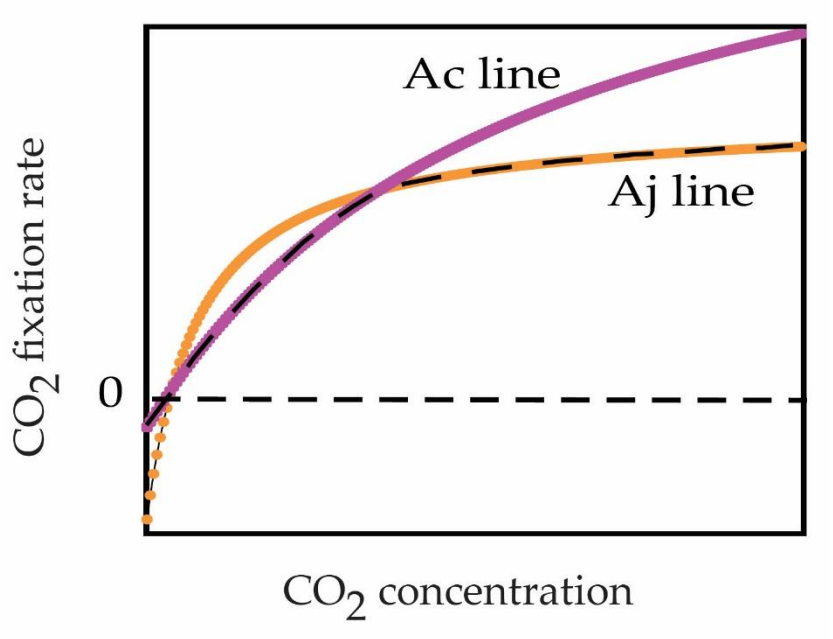
factors many parameters would be required to fully summarise the photosynthetic performance of a leaf or plant. While some of the parameters that have been developed to describe photosynthesis are widely used and standardized, others are improvised and non-standard and there is no complete, standard set of parameters that allows photosynthesis to be fully specified. In fact one of the challenges in the genetic analysis of photosynthesis will be establishing robust parameters to describe responses for which no generally accepted parameters exist – the better the parameter, the better will be the analysis.



Box 3, Figure 1. *A cartoon showing the characteristics of the responses of leaf assimilation to Ci (A/Ci) modelled using equations from Von Caemmerer (2000). The orange, Aj, line shows the response limited by the supply of ribulose-1,5-bisphosphate (rubp); the blue, Ac line shows the response limited by the activity of Rubisco. The actual assimilation rate will be the minimum of the two lines, shown as a dotted black line.*

The starting point for the parameterisation process is to have a measure of a photosynthetic response. The gold standard measure of photosynthesis is that of CO_2 fixation. The general characters of responses of CO_2 fixation to many environmental factors are known and some are well parameterised. A good example is the response of a leaf to changes in CO_2 concentration in the internal leaf air space, Ci, which are the result of changes in the CO_2 concentration in the air around the leaf. The response of CO_2 fixation (or assimilation) to Ci is graphically described in the so-called A/Ci curve (**Figure 1**). The A/Ci curve can be analysed by means of a basic model of photosynthetic responses to Ci proposed by Farquhar *et al.*(Farquhar et al., 1980). Depending on the assumptions that are made, these analyses yield a range of parameters, including the maximum CO_2 fixation rate under conditions of Rubisco limitation, the limitation by ribulose-1,5-bisphosphate (rubp) supply, the

point at which limitation of assimilation changes from Rubisco to rubp and, the limitation that prevails under ambient CO_2 concentrations(Sharkey et al., 2007). Depending on the measurements, and the assumptions made, a range of parameters can be used to describe A/Ci curves, of which V_{cmax} (maximum rate of assimilation when rubp supply is not limiting) and J_{max} (the maximum rate of assimilation when rubp supply is limiting) are the most essential. Another commonly measured response is that to incident irradiance (**Figure 2**), which is most often described using the light-limited quantum yield of CO_2 fixation and maximum rate of CO_2 fixation (Pmax) as parameters..



Box 3, Figure 2*. A sketch of a photosynthesis irradiance response showing the progression from the strictly light-limited region of the response at low irradiances through to light-saturation at high irradiances.*

Measurements of CO_2 fixation are, however, not easy to apply in mass phenotyping systems and they are often substituted by chlorophyll fluorescence measurements based on a non-contact method that can be used as a single point measurement or in an imaging mode. The basic principles of chlorophyll fluorescence, which arises predominantly from Photosystem II, thereby providing information about the operation and regulation of Photosystem II, have been recently reviewed (Baker et al., 2007; Baker, 2008) and will not be repeated in detail here. A wide range of physiologically useful measures can be estimated from fluorescence data (Baker et al., 2007; Baker, 2008), but possibly the most important is the estimate of the quantum yield for electron transport by Photosystem II (and thus for linear electron transport), known variously as $\Delta F/Fm$, Φ_{PSII} or Fq'/Fm'. This efficiency is an important measure of photosynthesis in its own right, but it can also be converted into a rate of electron transport and then to a rate of CO_2 fixation. Thus it is possible to substitute the

relatively easy to use fluorescence technology for the more labour intensive measurement of CO_2 fixation. In conclusion, both CO_2 fixation and chlorophyll fluorescence are means by which to measure the photosynthetic phenotype of a plant and as such are valuable screening tools in detecting genetic variation in plant photosynthesis.

Chapter 3

Phenomics for photosynthesis, growth and reflectance in *Arabidopsis thaliana* reveals circadian and long-term fluctuations in heritability

Pádraic J. Flood[1,2,5], Willem Kruijer[3], Sabine K. Schnabel[3], Rob van der Schoor[4,6], Henk Jalink[4,6], Jan F.H. Snel[4,7], Jeremy Harbinson[2], & Mark G.M. Aarts[1]

[1]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

[2]*Horticulture and Production Physiology, Wageningen University, Wageningen, the Netherlands.*

[3]*Biometris, Wageningen University and Research Centre, Wageningen, the Netherlands.*

[4]*Greenhouse Horticulture, Wageningen University and Research Centre, Wageningen, the Netherlands.*

[5]*current affiliation: Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany*

[6]*current affiliation: PhenoVation BV, Wageningen, the Netherlands.*

[7]*current affiliation: Adviesbureau JFH Snel, Wageningen*

Corresponding author: Mark Aarts (mark.aarts@wur.nl)

Abstract

Recent advances in genome sequencing technologies have shifted the research bottleneck in plant sciences from genotyping to phenotyping. This shift has driven the development of phenomics, high-throughput non-invasive phenotyping technologies. We describe an automated high-throughput phenotyping platform, the Phenovator, capable of screening 1440 Arabidopsis plants multiple times per day for photosynthesis, growth and spectral reflectance at eight wavelengths. Using this unprecedented phenotyping capacity, we have been able to detect significant genetic differences between Arabidopsis accessions for all traits measured, across both temporal and environmental scales. The high frequency of measurement allowed us to observe that heritability was not only trait specific, but for some traits was also time specific. Such continuous real-time non-destructive phenotyping will allow detailed genetic and physiological investigations of the kinetics of plant homeostasis and development. Ultimately such phenomic level technologies will provide more dynamic insights into plant physiology, and the necessary data for the omics revolution to reach its full potential.

Introduction

Photosynthesis is the primary entry point of energy into the biosphere and as such provides the foundation for life on earth. One prominent class of photosynthetic organisms are plants, which are responsible for the vast majority of the energy and biomass influx in the terrestrial biosphere. They are also the basis of our economy, providing the majority of calories necessary to sustain humanity. It is clear that plant photosynthesis is the keystone for our existence, but we know surprisingly little about the extent and basis of variation in this most fundamental of traits (Flood et al., 2011). The overarching reason for our lack of knowledge about intraspecific variation in photosynthesis is our inability to efficiently screen large numbers of plants. This epistemic Rubicon must be overcome for our survival as photosynthesis is the only major productivity-related trait which has yet to be improved (Long et al., 2006). To facilitate this, high-throughput phenotyping of photosynthesis must be developed.

Obtaining phenotypic data is the most time consuming and labour intensive step of many biological experiments (Houle, 2010). Despite this, the detail and extent of phenotypic data compares poorly with the increasingly complete genotype data now available (Finkel, 2009; Houle et al., 2010; Pieruschka and Poorter, 2012). This is not only due to the recent advances in genomics but also due to the complex multidimensional nature of phenotypes (Houle, 2010). The vast number of phenotypic states that a genotype can occupy can be visualised as its phenotypic space, which is often referred to as its phenome (Houle et al., 2010). In practice the phenome is a theoretical entity which can never be fully characterised. This was recognised by Houle *et al.* (2010) leading them to propose that phenomics may be understood as the "acquisition of high dimensional phenotypic data on an organism wide scale".

The phenotype is the result of the interplay between genetics and developmental, environmental and stochastic influences, where the intensity, frequency, order and interaction of these influences affect the outcome. Traditionally, due to its labour intensive nature, phenotyping was only feasible for a single time point on a subset of the traits which comprise the phenome. To reveal, however, the dynamic and variable nature of the phenome, requires numerous measurements across developmental and environmental gradients (Houle, 2010). Some phenomic (in the sense of Houle *et al*)

quality datasets for endophenotypes, i.e. transcriptomics, metabolomics, proteomics, ionomics, lipidomics, and even RNA directly undergoing translation (translatomics) have been produced (Joyce and Palsson, 2006; Kahlau and Bock, 2008). But as they rely on destructive measurements they only provide a snapshot of the endophenome at the time of measurement. These omics datasets not only lack dynamic insight but they also fail in another import aspect: they give no information about fluxes or growth. Yet, kinetic phenotypes or functional states, such as growth or photosynthesis, provide the most direct and integrative quantification of plant performance (Joyce and Palsson, 2006). They represent the combined effect of all other phenotypic levels, so the relevance of, for example, variation in gene expression can be assessed at higher organizational levels.

Both photosynthesis and plant growth are ideal traits to assess the functional relevance of endophenotypic omics datasets. Plant size reflects the integration of metabolic and developmental processes and is a good indicator of long term performance whilst photosynthesis and growth rate reflect more immediate physiological responses (Eberhard et al., 2008; Fatichi et al., 2014). Growth responses are most dynamic at the meristematic level (Körner, 2013) which are not amenable to rapid, frequent measurement, in contrast to plant size and photosynthesis, which together provide an ideal phenotypic window into genotype performance.

To this end we set about developing a high-throughput phenotyping platform which would allow us to continuously phenotype a large number of plants for photosynthesis and growth. This will result in phenomic data, though we recognise that the full characterisation of the phenome, namely all possible phenotypic outcomes, across all levels of organisational, developmental and environmental space, is beyond our current capabilities. Nevertheless this is a valuable step forward and will give high-dimensional phenotypic data which, in accordance with Houle *et al.* (2010) can be considered phenomics.

High intensity screening of a particular trait will allow for temporally detailed estimation of heritability. Broad sense heritability is a measure of how much of the phenotypic variance in a population can be attributed to genetic variation rather than other factors, such as a non-uniform environment (Visscher et al., 2008). It is often used to assess the potential responsiveness of traits to selection, whether natural or artificial

(Lynch and Walsh, 1998; Kruijer et al., 2015). Whilst it is well recognised that heritability is trait, population and environment specific, its variation with time is less well studied. If heritability shows significant time dependence then this will be of interest to breeders and evolutionary biologists, as the time point at which selection occurs will be crucial in determining the selection response. The ability to measure traits multiple times per day for prolonged periods in order to better understand the time-dependency of heritability was an important factor in the design of the phenotyping system we describe here.

System development

Design considerations

The overarching goal of the phenotyper system, which we named Phenovator, was accurate quantification of the phenotypic variation, so as to estimate the genetic variation, in natural populations of *Arabidopsis thaliana*, using photosynthesis and growth as phenotypic indicators of plant performance. This required that any noise, whether technical, environmental or otherwise, be minimized so that the genetic signal could be accurately assessed. This is particularly important when dealing with traits like photosynthesis, which are environmentally responsive and exhibit limited phenotypic diversity within a species (Flood et al., 2011). Thus the plant growth environment should be well controlled to minimize heterogeneity of the environment and allow high reproducibility. Key environmental variables which have a large effect on plant performance and often elicit a phenotypic response, are light, water, temperature and nutrient availability. To control these inputs the Phenovator was located in a climate-controlled growth chamber and equipped with an automated watering system.

To allow repeated measurements of the same plants, and minimize any measurement effects, the measurements must be non-invasive. We therefore developed an image-based phenotyping platform. A balance also had to be struck between the extensive (range) and intensive (detail) capabilities of the Phenovator. We chose to measure a restricted set of phenotypes that are important indicators for plant performance (photosynthetic activity, size, and colour) and to measure these with a high frequency, opting for intensity of measurement. Our optical measurement system was based on a camera. Other camera-based phenotyping systems have been developed and in many the

plants are moved to the phenotyping equipment (Jansen et al., 2009; Arvidsson et al., 2011; Skirycz et al., 2011; Tisné et al., 2013). This has the advantage that the number of plants that can be screened is only limited by the growing area, but the disadvantage is that the plants are not assayed under growth conditions and that the rate of throughput is decreased. In our system we opted to move the camera to the plants and as the camera can be moved at 6 ms^{-1} (much faster than a plant can be safely moved) we can image plants with a high-frequency, but the total growing area that can be imaged is limited by the camera movement system.

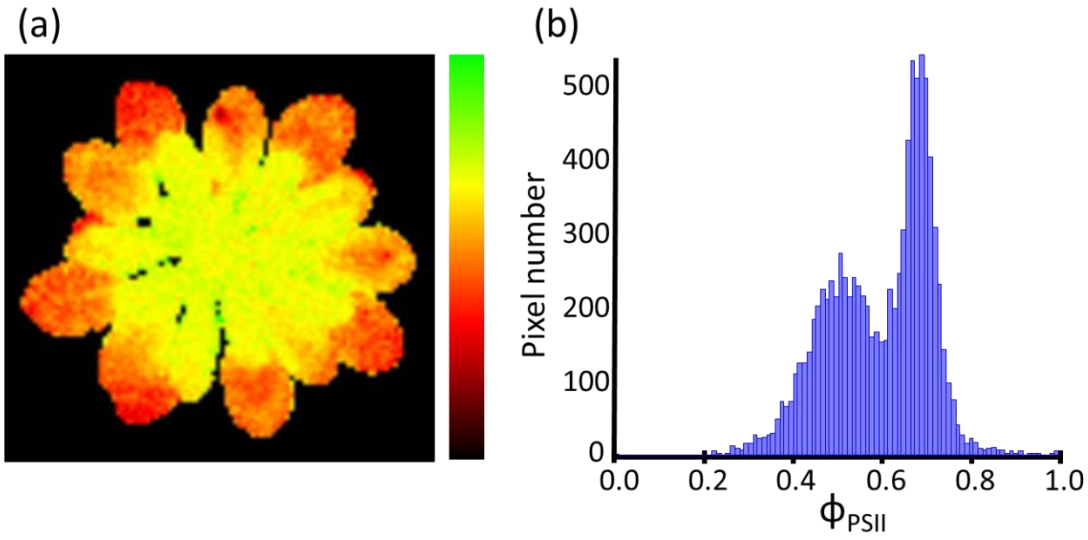


Figure 3.1. *Distribution of photosystem II efficiency (ΦPSII) in a phosphate deficient Arabidopsis. (a) False colour ΦPSII image of a phosphate deficient plant, the scale bar on the right shows ΦPSII values from 0 (black) to 1 (green). (b) Image (a) plotted as a histogram of pixels at specific ΦPSII values. The distribution is bimodal hence the mean value fails to aptly represent the plants phenotype.*

Since many phenotypes show spatial heterogeneity (see Figure 3.1 for an example), it was essential to image the entire above-ground part of the plant (roots are outside the scope of this phenotyper). Since our target species, *Arabidopsis thaliana* (Arabidopsis), forms a rosette, which until flowering is relatively flat, this could be achieved using a single camera. To be able to identify and characterise genetic variation we needed sufficient throughput to screen populations suited for genetic mapping, such as Recombinant Inbred Line (RIL) populations or genome wide association (GWA) panels. The latter populations usually consist of 300 or more genotypes (Korte and

Farlow, 2013) which with four replicates per genotype yields a minimum screening capacity of 1200 plants. To capture short-term changes in the phenotype it was decided that it should be possible to measure all plants within 60 minutes. Finally it was essential that the entire system was automated with control and data storage outside the growth room to minimize environmental fluctuations (particularly carbon dioxide) due to people entering the room.

System design

The Phenovator we designed consists of five main parts: a supporting frame, an ebb and flood hydroponic system, an XY camera movement system, a camera and a computer to control camera movement, imaging and data storage (Figure 3.2). The supporting frame was constructed from 100×100 mm^2 box-section aluminium beams (www.maytec.org) to support X-Y rails and the basins in which plants were grown. It also provides the rigid, stable camera platform necessary for imaging. To be able to image quickly requires a platform that is sufficiently stiff to eliminate vibration after the camera movement is complete. The camera movement system (www.elmekanic.nl) is capable of speeds of 6 m s^{-1} (though for safety reasons this is currently limited to only 1 m s^{-1}) and allows high reproducibility of camera positioning. We use a so-called "ebb and flood" hydroponic irrigation system to water and feed the plants growing in a rockwool (www.grodan.com) substrate (Figure 3.2a). Rockwool is a synthetic, relatively inert, fibrous substrate which allows manipulation of plant nutrition regimes (Gibeaut et al., 1997).

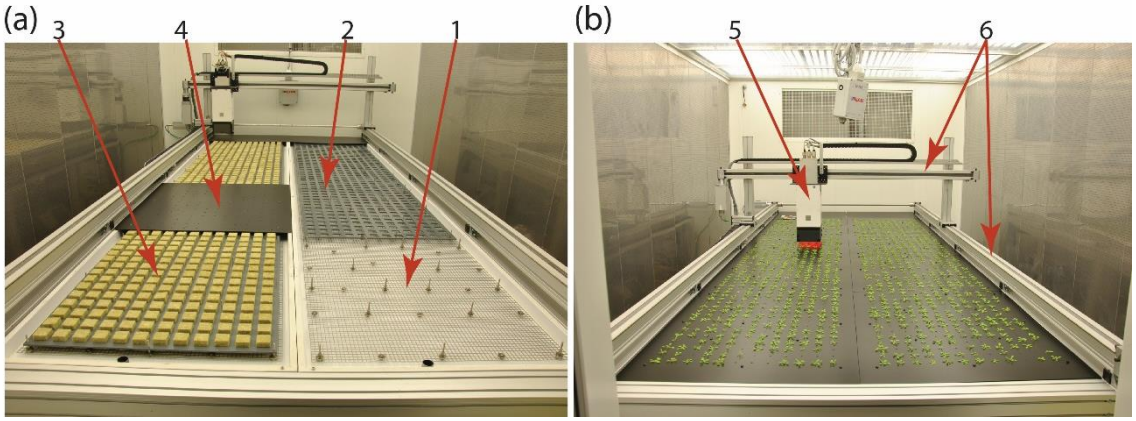


Figure 3.2. *The Phenovator. (a)The set-up of the growth system. 1: support grid for the rockwool blocks with support pins for the rockwool block spacing plate and the top plate; 2: the rockwool block spacing plate to position and hold the blocks. 3: this grid with*

rockwool blocks in place, and 4: the black PVC top plate. (b)The 'Phenovator' system in action: 5: the imaging head carrying the camera (the red light is the saturating pulse for ΦPSII measurement) 6: the X-Y rails used to position the imaging head.

The growth system (Figure 3.2) is comprised of two irrigation basins, each with their own irrigation tank and pump, which allows for two different irrigation and nutrient regimes to be applied in the same experiment. Each basin has space for 720 rockwool blocks 40×40×40mm in size giving a total capacity of 1440 rockwool blocks. The rockwool blocks are held 20 mm apart by a PVC grid that is attached to a rigid stainless steel grid upon which the blocks rest. The PVC grid prevents any sideways movement of the rockwool blocks which could cause the plants to shear, and is held 15 mm above the stainless steel base grid by spacers. The stainless steel grid provides structural integrity to the rockwool support system and is supported 5 mm above the bottom of the irrigation basin. The perforations in the grid allow for free circulation of nutrient solution, ensuring that all blocks receive irrigation for approximately the same amount of time. A spacing of less than 5 mm between the stainless steel base and the irrigation basin was found to sometimes cause problems of root death, possibly due to trapping of nutrient solution and anaerobiosis. On top of the rockwool blocks there is a black plastic non-reflective sheet of foamed PVC, 3 mm thick (Figure 3.2a, 3.4). In this sheet, 3-mm countersunk holes were drilled at distances of 60 mm and positioned above the centre of each rockwool block. All three layers are held in place using threaded stainless steel pins which were welded to the stainless steel grid. Four support studs fit into sockets drilled into the irrigation basin to hold this grid in a fixed position. All materials were tested for phytotoxicity and corrosion resistance, and were washed thoroughly before use. The black plastic cover ensures that there is no algal growth, restricts soil dwelling organisms such as the larvae of fungus gnats (*Bradysia spp*) and minimises background noise in the images, making automated image processing much easier.

Images are recorded using a monochrome camera (Pike; www.alliedvisiontec.com) mounted on the X-Y movement system. An 8-position filter wheel is mounted between the lens and the ccd chip of the camera to capture images in different wavelength bands. We measure reflectance at 480 nm, 532 nm, 550 nm, 570

nm, 660 nm, 700 nm, 750 nm and 790 nm with each filter having a full width at half maximum (FWHM) of 10 nm; these narrow spectral wavelength measurements allow for estimation of a range of plant pigments. The reflection bands at 480 nm, 570 nm and 660 nm are used to construct Red, Green and Blue (RGB) colour images. Chlorophyll content (Chl) is estimated from reflectance (R) at 700 and 790 nm after Gitelson et al. (2003) $Chl=(R700^{-1}-R790^{-1})\cdot R790$.

Projected leaf area (PLA) provides a good estimate of above ground biomass (Leister et al., 1999) and is estimated from near infrared (NIR) reflection at 790 nm; this wavelength was chosen so the plants could be measured both day and night without disturbing the day-night cycle. Four NIR light emitting diodes (LED) with a FWHM of 40 nm and a maximum radiant power of 1 W per LED provide the 790 nm radiation. NIR measurements are taken every three hours resulting in eight images per day.

We use chlorophyll fluorescence imaging to measure Φ_{PSII} (the light-use efficiency of PSII electron transport, also known as Fq'/Fm', or ΔF/Fm) (Genty et al., 1989; van Kooten and Snel, 1990; Baker, 2008) using a variation of the method of Genty and Meyer (1995). This method has the advantage of a good signal to noise ratio and has proved very suitable for our imaging conditions in which the unfiltered background irradiance is low owing to the shadowing effect of the imaging system. Measurements are made by illuminating the plants at the growth chamber actinic light level for 10 s followed by a 2 s saturating pulse of 5000 μmol m^{-2} s^{-1} with LEDs attached to the Phenovator camera head. At the end of the 10 s of actinic light and prior to the saturating light, 24 images are taken and averaged to generate the Fs image. During the saturating light pulse six images are taken of which that with the highest signal is used for the Fm' image. The LEDs are turned off after the saturating pulse and an additional 24 images are taken and averaged in order to generate a dark image to account for any background light from the fluorescent lamps in the growth chamber. A fluorescent target is imaged at the beginning of each measurement sequence in order to provide a factor to correct the Fs and Fm' measurements for the difference in light intensity used to produce the images (Genty and Meyer, 1995). Thus the Phenovator comprises only four moving parts, the X movement system, the Y movement system (these both comprise of motors, drive belts and bearings), the filter wheel and the camera focus. This simplicity is a strong advantage when long term experiments are undertaken.

Data processing

The growth platform containing the 1440 plants is divided up into 120 imaging positions (Figure S3.1) each of which contains 12 plants (3 x 4) thus each measurement cycle results in 120 images each containing 12 plants. Different measurement tasks (imaging Φ_{PSII}, NIR reflectance or spectral imaging) can be programmed in a daily schedule, which is used over the entire experiment. Analysis software has been developed to convert raw images from the imaging system to images of physiological parameters (e.g. Φ_{PSII}) or biochemical composition (e.g. chlorophyll content). Each image is matched to a table position, and the genotype planted at each position is provided via a comma separated (csv) file, thus enabling the image processing software to group images by genotype. Based on images containing 12 plants the analysis software (available upon request) calculates per replicate the parameters for each genotype. Each measurement protocol (e.g. measurement of Φ_{PSII}) produces its own parameters, which are calculated from a selected area within the image using a mask derived from the desired plant. A grid of vertical and horizontal reference lines (shown in Figure S3.1) is set by the user and provide the coordinates around which a box is drawn to select individual plants. A greyscale threshold (or mask), set by the user, is used to distinguish the plant from the background within this box. Twelve areas are defined and used to obtain a specific plant from the image. Only the pixels within the mask are used to estimate the phenotypic parameters.

All images (raw data and derived data) are stored, and the values of each phenotype are calculated per pixel. Both the pixel values and the averages over images are available to output in csv format. The spatial distribution of pixel data within any stored image can be shown (Figure 3.1 and Figure S3.1). Since our plants were grown for only four weeks under non-stressful conditions there was no spatial variation in any parameter so we will not discuss this further.

Materials and methods

Plant material and Cultivation

Unless otherwise stated all plants were grown as follows: seed was sown on wet filter paper and stratified for 6 days at 4°C. After stratification seed was sown directly on wet rockwool (www.grodan.com) which had been pre-soaked in a nutrient solution designed for Arabidopsis (see Table S3.1 for composition). One seed was sown per rockwool block (system described in "System design"). The growth conditions were as follows, 10/14 hr day/night, irradiance normally 200 µmol m^{-2} s^{-1} and 550 µmol m^{-2} s^{-1} in the high light experiment, 20/18˚C day/night temperature, 70% relative humidity, and ambient CO_2. Plants were irrigated daily with nutrient solution for 5 minutes. In total 57 genotypes were screened across four experiments, see Appendix S3.1 for details of genotype identity and number of replicates. The Φ_{PSII} estimates were compared with those of a MINI-PAM fluorometer (www.walz.com) to validate the measurements and no significant differences were found.

Measurement protocols

Φ_{PSII} was measured daily, one, four and seven hours into the photoperiod. This was considered sufficient to document any variations in the phenotype and allowed time for other measurements such as NIR, which was measured every three hours.

Statistical analysis

Variance components

The importance of several design factors was assessed by fitting the following mixed model for each trait and time-point using asreml-R (Butler et al., 2007):

Equation (1)

$$Y = \mu + C + G + Exp + Basin + x + y + TablePosition + x_{within} + y_{within} + G \times Exp + Exp \times (Basin + x + y + TablePosition + x_{within} + y_{within}) + Exp \times Basin \times G + Exp \times G \times (x_{within} + y_{within}) + R(Error)$$

Where µ is the overall mean, and G, Exp and Basin are the factors for respectively genotype, experiment and basin. The factor C represents check-genotypes that were not

included in subsequent analyses, but included in the mixed model in order to better estimate the variance components; it has one level for each check-genotype and one additional level representing all other genotypes. All terms except μ and C are defined as random effects. For traits and time-points that were only present in a single experiment, all terms involving Exp were dropped from the model. Spatial variability was modelled by the factors $x, y, TablePosition, x_{within}\ and\ y_{within}$ which represent respectively rows, columns, table (camera/image) position and within image rows and columns. While x and y model the coordinates across the whole platform, x_{within} and y_{within} model the spatial effects within images of 12 plants (3 x 4). A more detailed description of all design factors is given in Appendix S3.2.

Genotypic means

Genotypic means used in Figure 3.4 were calculated as the Best Linear Unbiased Estimators (BLUEs) for genotype, using a mixed model identical to equation (1) but with genotype as fixed effect.

Heritability estimates

Defining and estimating heritability in the context of a mixed model as defined by equation (1) is known to be difficult (Oakey et al., 2006), since not only the residual error contributes to the environmental variance (the generalized heritability proposed in Oakey *et al.* concerns line heritability and not the (plot level) heritability $\sigma_G^2/(\sigma_G^2 + \sigma_E^2)$, which is of interest here). To obtain more interpretable and commonly used heritability estimates we therefore performed classical analysis of variance (ANOVA) for the linear model with (fixed) effects for genotype, basin nested within experiment, and within image x_{within} and y_{within} coordinates. This included the most important main effects identified by the mixed model analysis described above; the fact that the interactions of design factors with genotype were small justifies the effects being fixed here. The genetic and environmental variance were estimated by respectively$(MS(G) - MS(E))/\overline{r}$ and MS(E), where MS(G) and MS(E) are the mean sums of squares for genotype and residual error (Lynch and Walsh (1998), Kruijer *et al*. 2015). Broad sense heritability was then estimated by the ratio of estimated genetic variance over the sum of estimated genetic and environmental variance. To facilitate direct comparison, heritability was estimated using 20 genotypes

which were screened under both light conditions (see Appendix S3.1 for details on the genotypes used).

Growth curve characterisation

PLA was measured throughout each experiment from NIR images and the masks generated from the Φ_{PSII} images, a total of 11 images per day. In order to summarize these data and estimate growth rates from repeated plant-size measurements, a flexible curve was fitted to the data for each plant. We used P-splines as a flexible semiparametric description of the curves (Eilers and Marx, 1996). P-splines are penalized B-splines resulting in smooth piecewise polynomial curves. For the implementation in the context of this paper we used the R package mgcv (Wood, 2011) with the function gam with its option for P-splines. Fitted curves and addition growth parameters are plotted in Figure 3.5 and supporting Figure 3.3. The (empirical) slope at all time points is calculated directly from the fitted values of the curve. Relative growth rates can be calculated based on the raw data series. However, for fluctuating time series growth rates are more reliable when a smooth curve is base of their calculation (Figure S3.3).

Results

System uniformity

The (spatial) uniformity and (temporal) reproducibility of the system were assessed by estimating the magnitude of several design factors using a mixed model (see Material & Methods and Appendix S3.1 for an overview the experiments), which included random effects for genotype, experiment, basin, and table position (Appendix S3.2, Data S3.1, Data S3.2). Using this model, genotypic means were calculated as the Best Linear Unbiased Estimators (BLUEs) for genotype. Spatial variability was modelled by row (x) and column (y) effects, as well as within image rows x_{within} and columns y_{within} . While x and y modelled the coordinates across the whole platform, x_{within} and y_{within} modelled the spatial effects within images of 12 plants (3 by 4). In addition to the main effects, second and third order interactions between design factors were included. A more detailed description of all design factors is given in Appendix S3.2.

For all traits and time points there is considerable genetic variation: the variance component for genotype is of a similar order of magnitude as the residual error variance,

which is consistent with the heritability estimates found below. Although the main effect of experiment was substantial, the genotype by experiment interaction was negligible for almost all traits. Only for the spectral measurements at 700 and 750 nm the genotype by experiment interaction was larger, but still small compared to the main genotypic variance (Data S3.1, Data S3.2). The phenotypic ranking of the genotypes can therefore be expected to be consistent across experiments. For Φ_{PSII} and spectral measurements, the position within the image showed a considerable main effect which is likely due to light gradients in the camera head. This effect of position within the image showed no interaction with genotype (Data S3.1, Data S3.2) and thus could be corrected for. In a few cases, there was some interaction between experiment and within image position, but never with genotype. Table position and the x and y coordinate across the whole platform showed a small main effect for some of the spectral measurements. Nevertheless, the very low variances of the interactions between genotype, experiment and the design-effects indicate that we can combine data from different experiments, allowing phenotyping of potentially thousands of accessions.

Phenotypic variation

The Phenovator has three main imaging protocols in routine use (Figure 3.3). The first, is used to measure photosynthetic efficiency via chlorophyll fluorescence (Φ_{PSII}), the second is used to measure pigment content via spectral imaging and the third measures PLA via near NIR imaging.

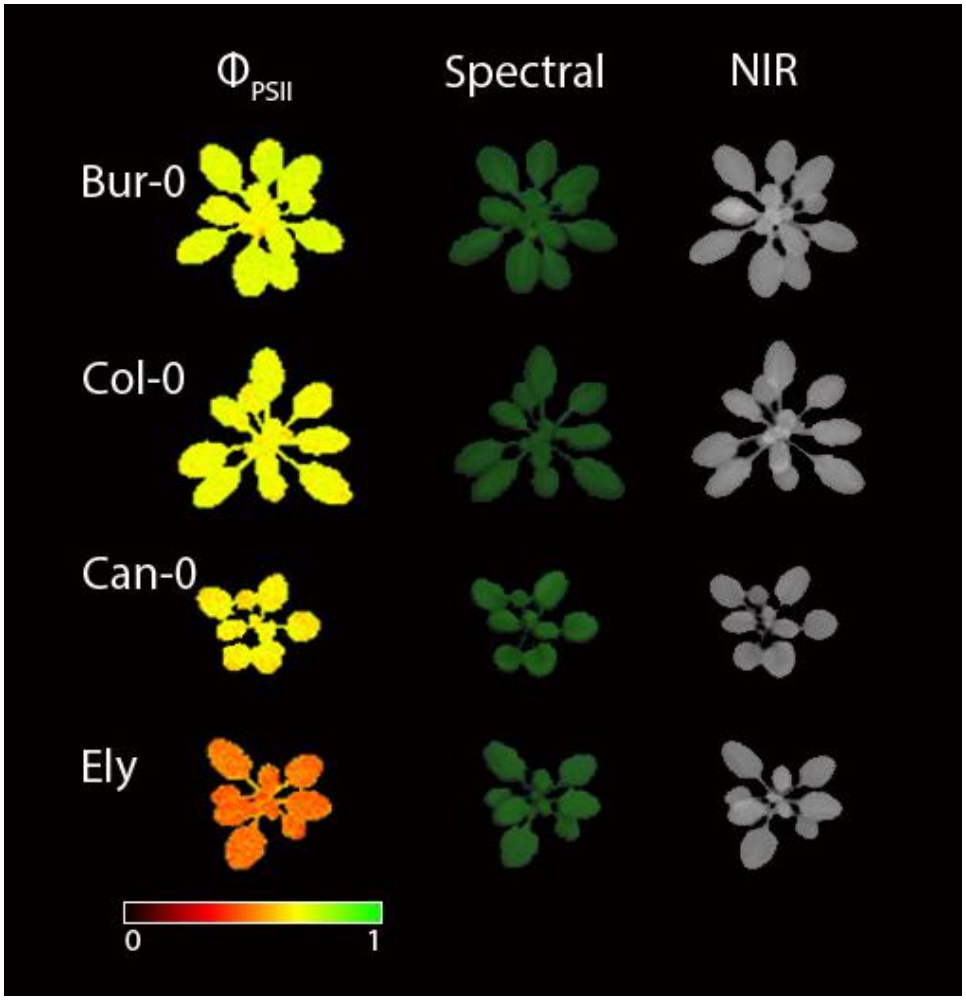


Figure 3.3. *Examples of images generated by the Phenovator. The first column shows false colour images of photosystem II efficiency (ΦPSII) running from 0 (black) to 1 (green). The second column shows the red-green-blue (RGB) output of the spectral measurements. The third column shows the images generated by near infra-red imaging (NIR) at 790 nm. The rows correspond to four different genotypes, Bur-0, Col-0, Can-0 and Ely. Ely is atrazine resistant, hence the much lower ΦPSII.*

Figure 3.4(a) and (b) show the total variation for Φ_{PSII} for 20 genotypes grown at 200 and 550 µmol m^{-2} s^{-1} light intensity. The two different light intensities were chosen both to test the flexibility of the system and to assess the response of the genotypes to these different conditions. Φ_{PSII} is influenced by both the light intensity and the genetic background of the plant measured. In addition to these differences, the high resolution

measurements allow the observation of both a daily fluctuations in Φ_{PSII} as well as a gradual upward trend through time at the higher light intensity.

The two genotypes with the lowest Φ_{PSII} are Ely, an atrazine resistant accession known to have a low light-use efficiency for PSII electron transport (El-Lithy et al., 2005) and RIL BC354 from the Bur-0 x Col-0 population (Simon et al., 2008), which is known to carry a mutant version of the *PDE237* gene affecting photosynthesis (Vlad et al., 2010). However, even without these unusual genotypes there is substantial variation for Φ_{PSII}. The variation from approx. 0.62 to 0.72 at 200 µmol m^{-2} s^{-1} and 0.54 to 0.63 at 55 µmol m^{-2} s^{-1} is about 12% (assuming an upper limit of 0.8 for Φ_{PSII}) for normal natural accessions, extending to almost 40% when the lines with unusually low light-use efficiency are included.

The chlorophyll reflectance index (Figure 3.4c,d) is a linear measure of chlorophyll content and, as expected (Anderson et al., 1995), decreases as the irradiance is increased from 200 to 550 µmol m^{-2} s^{-1}. PLA can be measured at short intervals, allowing the construction of growth curves (Figure 3.4e,f). Both genetic background and light intensity have a large effect on growth rates. Another interesting phenomenon is the undulating nature of the curves due to leaf movement. The percentage difference between images at neighbouring time points shows the movement more clearly (Figure 3.4g,h). The plant growth and leaf movement phenotypes are easily revealed and analysed because of the high imaging frequency. The fluctuation in PLA due to leaf movement can result in negative apparent growth rates, so we smoothed the curves before estimating growth rates (Figure 3.5).

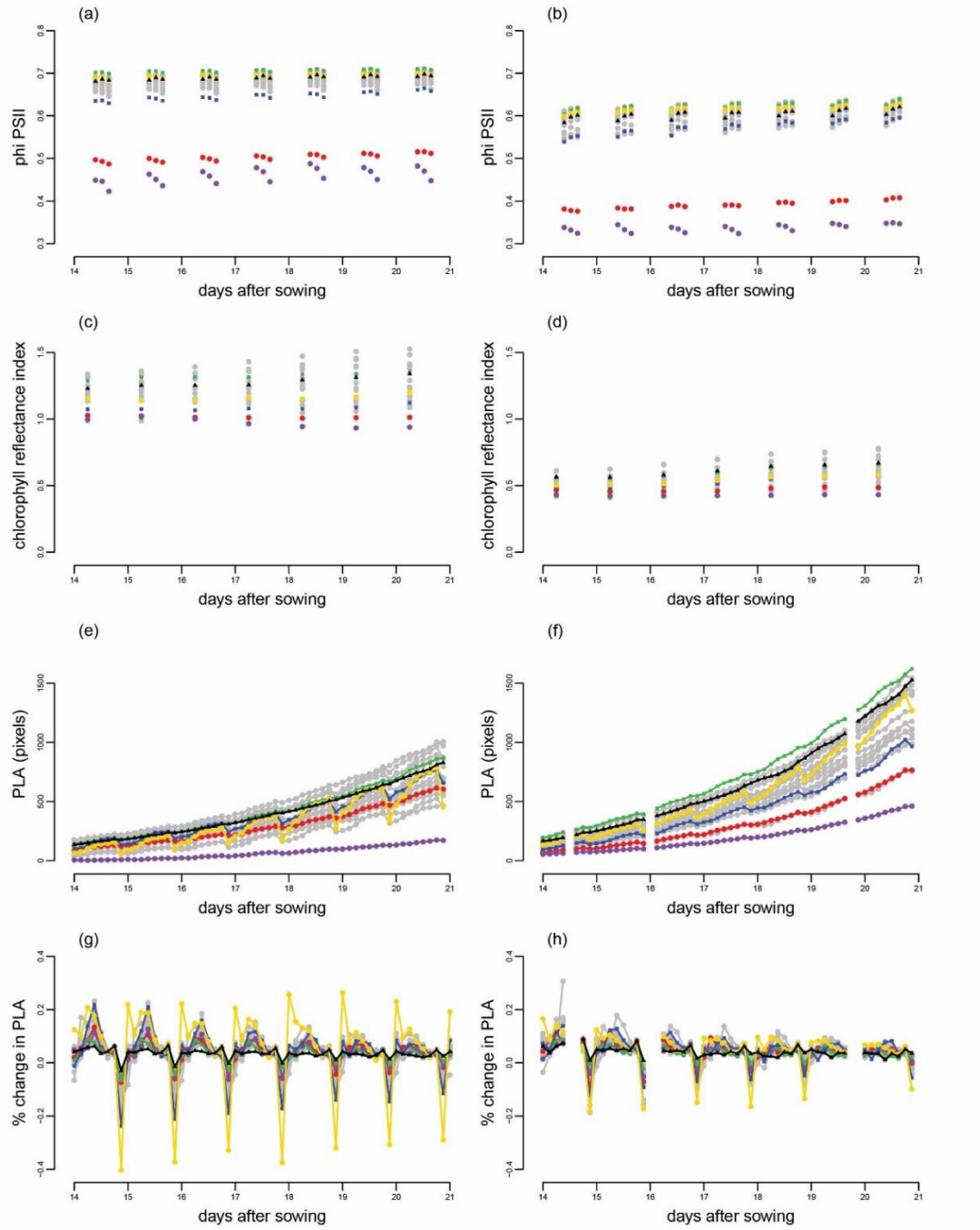


Figure 3.4. *Phenotypic diversity in twenty Arabidopsis genotypes grown at 200 µmol m-2 s-1 light intensity (a, c, e, g), and 550 µmol m-2 s-1 light intensity (b, d, f, h). Graphs (a) and (b) show Φ_{PSII} through time; (c) and (d) show chlorophyll reflectance index; (e) and (f) show projected leaf area (PLA). Finally, (g) and (h) show percentage change in PLA every 3hrs. All data points are genotypic means (BLUEs), combining observations on replicates from different experiments into one representative value for each genotype at each time point. Six genotypes, An-2 (yellow circles), BC354 (purple circles), Bur-0 (green squares), Col-0 (black triangles), Ely (red circles) and Ts-1 (blue squares) are indicated in colour. Error bars have been excluded for clarity, the significance of between genotype differences is apparent from the heritability estimates in Figure 3.6.*

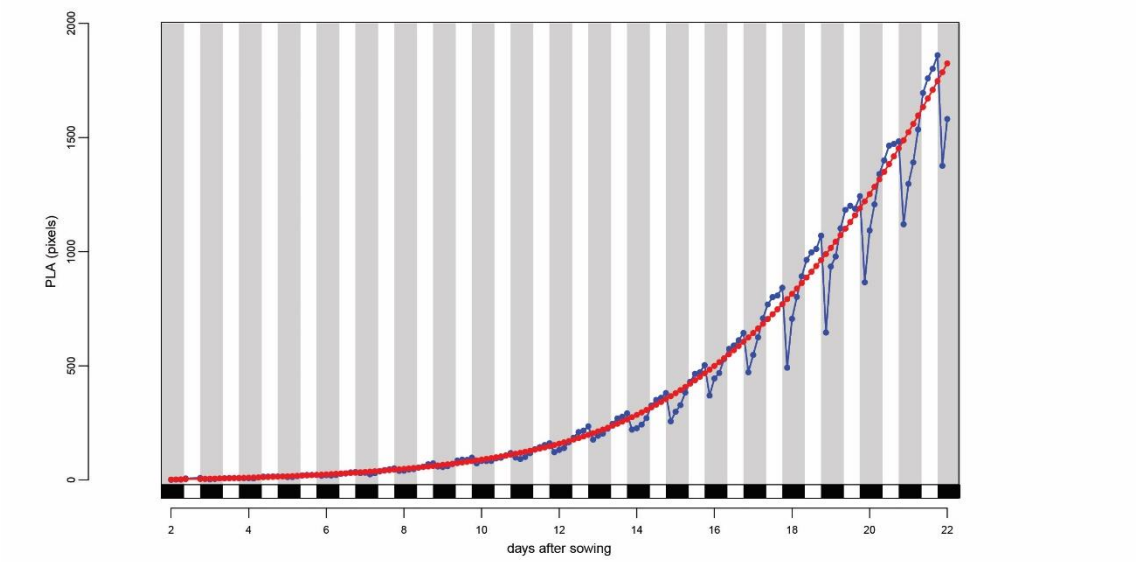


Figure 3.5. *Curve parameterisation of projected leaf area (PLA) for one replicate of genotype An-2. Blue line shows raw data and red line shows fitted values, white and grey bars indicate the day night cycle.*

Genetic variation

The heritability of a trait is a measure of the proportion of phenotypic variance explained by genetic effects (Visscher et al., 2008). Figure 3.6(a) shows the heritability through time for Φ_{PSII}. Addition of the photosynthetic extremes greatly inflates the genotypic variance which results in very high estimates of heritability. The heritability of Φ_{PSII} also shows a slight but recurrent daily rise, but is not affected by the difference in light intensity. The heritability of chlorophyll reflectance index and PLA show more gradual changes through time and are different depending of the light intensity (Figure 3.6b,c). Heritability of percentage change in PLA on the other hand is much more dynamic, with values shifting from 0.04 to 0.83 in the course of 6 hrs (Figure 3.6d), emphasizing the importance of frequent measurements. In general the heritability was slightly lower at higher light intensity, probably due to reduced overall leaf movement (Figure 3.4g,h). The most pronounced fluctuation is between day and night with heritability being much higher in the night than during the day. For spectral reflectance and growth curve traits the heritability also show variation through time but in a less dynamic fashion shifting over the course of several days (Figure S3.4 and Figure S3.5).

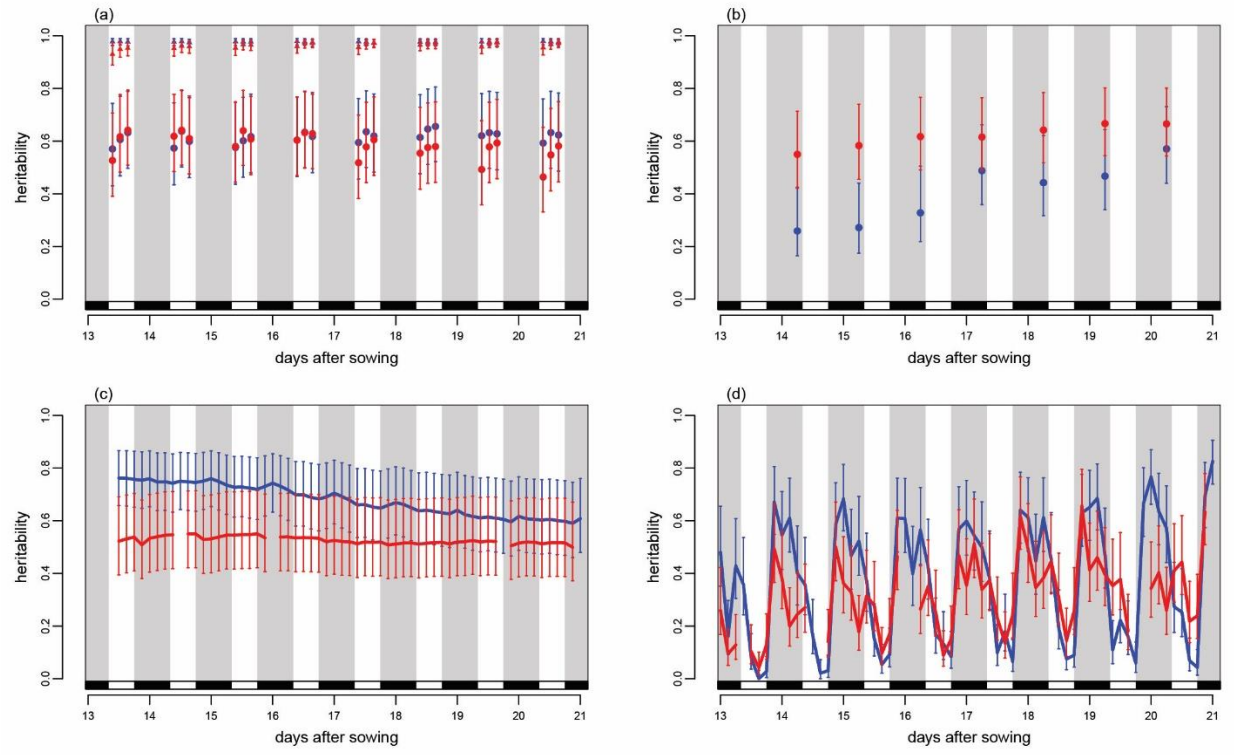


Figure 3.6 *Time course of heritability. (a) Heritabilities for ΦPSII at two light levels with and without extreme genotypes. Blue circles show heritabilities of ΦPSII at 200µmol m-2 s-1 without considering Ely and BC354. Blue triangles show heritability of ΦPSII at 200µmol m-2 s-1 including data for Ely and BC354. Red data point as for blue but where plants were grown at 550µmol m-2 s-1. (b) Heritability of Chlorophyll reflectance. (c) Heritability of projected leaf area (PLA). (d) Heritability of percentage change in projected leaf area. 200µmol m-2 s-1(blue) and 550µmol m-2 s-1(red), white and grey bars indicate the day night cycle. Error bars are 95% confidence intervals.*

Discussion

Uniformity and reproducibility

Uniformity and reproducibility of the Phenovator is essential if it is to be of any use. Although some design factors had a considerable main effect on the measurement (see Data S3.1, Data S3.2 and below), the interaction of these design factors with genotype was very small and can be corrected for. The effect of the position within the image can be attributed to light gradients in the camera head, while the effect of experiment may be due to small accumulated differences which are collectively significant but individually minor (Massonnet et al., 2010). The correction for design factors is achieved through the calculation of BLUEs for each genotype (Figure 3.4). The ability to correct design factors greatly increased the signal-to-noise ratio of the Phenovator, with signal being the genotypic effect and noise being unexplained phenotypic variation. This will be important when screening genetic mapping populations, as a larger contribution of the genotypic effect to the signal will increase the heritability of the measured trait. Finally, the estimation of the effect of experiment and its negligible interaction with genotype or other design factors allowed the combination of data from different experiments, greatly increasing the effective capacity of the system and the power of our statistical analysis.

Phenomic data

The only comparable system measuring photosynthetic and growth parameters is the GROWSCREEN FLUORO (Jansen et al., 2009), which can phenotype up to 60 plants hr^{-1} for growth and dark-adapted (maximum) PSII efficiency (Fv/Fm). Our system can measure the projected leaf area of 1440 plants in 20 minutes, and their light adapted PSII efficiency, or operating efficiency (Φ_{PSII}, Fq'/Fm') (Baker, 2008), in less than an hour. The operating efficiency of Φ_{PSII} directly relates to the rate of carbon fixation and ultimately growth and thus is physiologically more relevant than Fv/Fm when assessing genotype performance in a range of conditions (Genty et al., 1989). Using measurements of Φ_{PSII} we were able to determine differences in the rate of photosynthesis and estimate the genetic contribution to these differences for 20 genotypes of Arabidopsis grown at two light levels (Figure 3.4a,b and Figure 3.6). Of interest is the daily rise in Φ_{PSII} for all genotypes at 550 µmol m^{-2} s^{-1} with the exception of RIL BC354, which shows a daily

decline in Φ_{PSII}. The mutant allele of *PDE237* (At4g30720), normally encoding an oxidoreductase/electron carrier residing in the chloroplast stroma (Vlad et al., 2010), probably affects Φ_{PSII} due to accumulated PSII damage during the day.

Using NIR light allowed us to measure PLA throughout the day and night without disturbing the photoperiod. Since one NIR measurement of all 1440 plants takes only 20 minutes we could measure all plants 72 times per day. This frequency exceeds that required to capture growth or leaf movement in most cases, but it could be valuable to capture rapid responses such as those induced upon water stress or disease infection. For measurements of growth or leaf movement under non-stressed conditions, imaging once every three hours has proven to be sufficient (Figure 3.4g,h and Figure 3.5).

In addition to our priorities of measuring plant growth and photosynthesis we also measured the reflectance of individual plants at eight wavelengths of light. This made spectral imaging and estimating pigment content possible, which was also highly reproducible across experiments and genotypes. We were able to show a decline in chlorophyll reflectance when the plants were grown under high light conditions (Figure 3.4c,d) which is expected according to literature(Anderson et al., 1995).

Heritability through time

Using the phenotypic values for all traits across all genotypes we calculated the broad sense heritability of the different traits. As expected, heritability was trait specific, reflecting the genetic variation present for the trait. An unexpected finding was the amount this could vary through time. Daily fluctuations in heritability for some traits ranged from 0.04 to 0.83 (Figure 3.6). As far as we know it is the first time this has been described in such detail. The magnitude and frequency of this variation in heritability was much greater than expected and this has significant implications. In the case of percentage change in PLA every 3hrs, reflecting leaf angle at different time points (Figure 3.6b), the fluctuations in heritability show a diurnal pattern with a recurrent decline during the day under both light intensities. This may indicate the higher selection intensities present for leaf angle in light than in the dark, or alternatively a wider range of optima for leaf angle in the dark.

Two recent studies in Arabidopsis used high-throughput phenotyping to describe changing heritability through time. The first showed changes in the heritability for rosette compactness, which appears to increase linearly until the rosette has fully formed (Zhang et al., 2012). Two other traits, rosette area and circular area, showed fewer changes. The second study focused on root gravitropism (Moore et al., 2013); after being reoriented by 90˚, roots of seedlings were imaged every 2 minutes for 8 hours. The gravitropic response also showed a change in heritability through time. Interestingly some of the QTLs underlying this changing heritability were time specific and only detectable for short periods.

Our results show both large and dynamic fluctuations in heritability due to changes in the relative contribution of genetic diversity to the traits at different time points (Figure 3.6, Figure S3.4 and Figure S3.5). The implications of this dynamic variation in heritability are wide ranging. For traits with such strong fluctuations in heritability, the time they are measured at will have a considerable impact on the extent of variation found. For crop breeding programs this could result in the fixation of alleles which may not be optimal for trait improvement. Screening when heritability is low will reduce the ability to detect genetic variation and the response of the germplasm to selection is likely to be curtailed (Visscher et al., 2008). This can lead to a waste of resources in large-scale breeding experiments. Awareness of the fluctuations in heritability can also be used to inform the breeder when the variation in phenotypes is most relevant. As shown by Moore *et al.* (2013) the genetic loci responsible for the changing heritability can change through time. If fixation of a specific locus or set of loci is required, then identification of the time when they contribute most to phenotypic variance will result in more targeted breeding, and again, greater efficiency. Awareness of the extent and time dependency of variation in heritability will thus maximise the return on investment in trait selection (Brown et al., 2014).

From an ecological and evolutionary perspective, stronger selection often results in reduced heritability (Mousseau and Roff, 1987), thus if the intensity of selection varies with developmental time, traits which contribute to fitness when selection is greatest are likely to show a reduction in heritability. While this will require further validation it illustrates the value of high-throughput phenotyping for generating insights into the

genetic architecture of traits and the uses of such insights in the fields of breeding and evolutionary ecology.

Conclusion - where next?

The objective of our work has been to develop a high throughput phenotyping platform for photosynthesis (Φ_{PSII}) and growth. The rationale behind this is that phenotyping advances are essential for further rapid progress in plant genetics and breeding (Houle et al., 2010; Furbank and Tester, 2011; Pieruschka and Poorter, 2012). The choice of photosynthesis and growth was key, as they are both important traits with a complex polygenic architecture, and reliable high throughput phenotyping methods are needed if we are to mine natural variation or induced mutant libraries for these traits. Photosynthesis is of particular importance as it is the only major physiological trait not to have been directly bred for, and thus represents uncharted territory within which there may be considerable scope for crop improvement (Long et al., 2006; Flood et al., 2011). In nature photosynthesis has been shaped by selection in environments where many resources are limited but the supply of fixed carbon is not usually limiting factor for growth (Körner, 2013), while in agriculture resources are more abundant and the supply of fixed carbon is often limiting (Murchie and Niyogi, 2011). Adaptations which evolved to increase survival in the wild, but reduce yield in an agricultural context, may be selectively removed (Denison, 2012). For any such breeding program to be a success, there needs to be appropriate phenotyping (Cabrera-Bosquet et al., 2012). We have proved this is possible for Arabidopsis, though the system we describe would be suitable for any species which forms a flat rosette and for seedlings of most other species. Besides their importance for crop improvement, high throughput phenotypers are essential for quantitative genetic studies such as QTL or GWA mapping. High throughput screening will aid forward genetics approaches for the identification of QTLs and the genes responsible for the phenotypic differences in a population (Pieruschka and Poorter, 2012; Tisné et al., 2013). This is especially relevant when looking at natural accessions as such differences may represent adaptive alleles increasing fitness under specific environmental conditions (Tisné et al., 2013). Identification of such alleles is of interest for evolutionary biology and ecology, and to plant breeding as a source of genetic adaptations which can be used to tailor crop varieties to specific conditions.

The stability and design of the system allowed the combination of data from multiple experiments, increasing the effective capacity beyond the 1440 plants which can fit in a single screen. The design is such that a range of environmental variables, such as temperature, humidity and nutrient availability, can be controlled. To illustrate this flexibility we conducted one experiment at a higher light intensity. Such variation in the growth environment can be used to uncover hidden genetic variation not expressed under control conditions and identify genes important for adaptation to environmental fluctuations (Gibson and Dworkin, 2004).

Phenomic data is also essential for the advance of the omics revolution. To put all the current omics technologies into context, whole plant phenotyping of morphological and physiological traits is necessary. Without such phenomic data the relevance of variation in gene expression, metabolite or protein abundance to plant performance is much more difficult to assess. The integration of all levels of omics data from gene expression to growth rate will allow a systems biology approach to be undertaken which should greatly further our understanding of plant biology (Joyce and Palsson, 2006; Yuan et al., 2008; Chen et al., 2014). Our data show how informative phenomics data can be, revealing, for example, how a basic genetic parameter such as heritability can vary through time. This insight is a direct result of the expanded throughput, and particularly, intensity of measurements. The level of accuracy and throughput of our system shows it to be ideally suited for screening large populations of plants thus allowing future quantitative genetic studies of photosynthesis, growth, and the response of these traits to a range of environmental perturbations in Arabidopsis or any rosette species.

Acknowledgements

We thank Charles Neris Moreira for the use of his phosphorus deficient plant for Figure 1 and help with planting experiments. Eltje Groendijk and Gradus Leenders for technical assistance in construction. Gerrit Stunnenberg and Taede Stoker for help in running the experiments. Roxanne van Rooijen for her comments on the draft manuscript and Bas

Zwaan for insightful discussions surrounding heritability, its uses and significance. This work was financially supported by an ALW/TTI-Green Genetics grant from the Dutch Organization of Scientific Research (NWO).

Supporting Information

The following files are available in the compact disc at the back of the thesis.

Data S3.1. Variances of all design factors and interactions for each trait at 200 µmol m^{-2} s^{-1}. Separate excel sheets correspond to different traits measured. Raw data and graphs of variance components through time are shown on each excel sheet.

Data S3.2. Variances of all design factors and interactions for each trait at 550 µmol m^{-2} s^{-1}. Separate excel sheets correspond to different traits measured. Raw data and graphs of variance components through time are shown on each excel sheet.

Appendix S3.1. Overview of experiments conducted and genotypes used. A list of aliases and geographical origins is also included as well as the number of replicates sown and the experiments each genotype was included in.

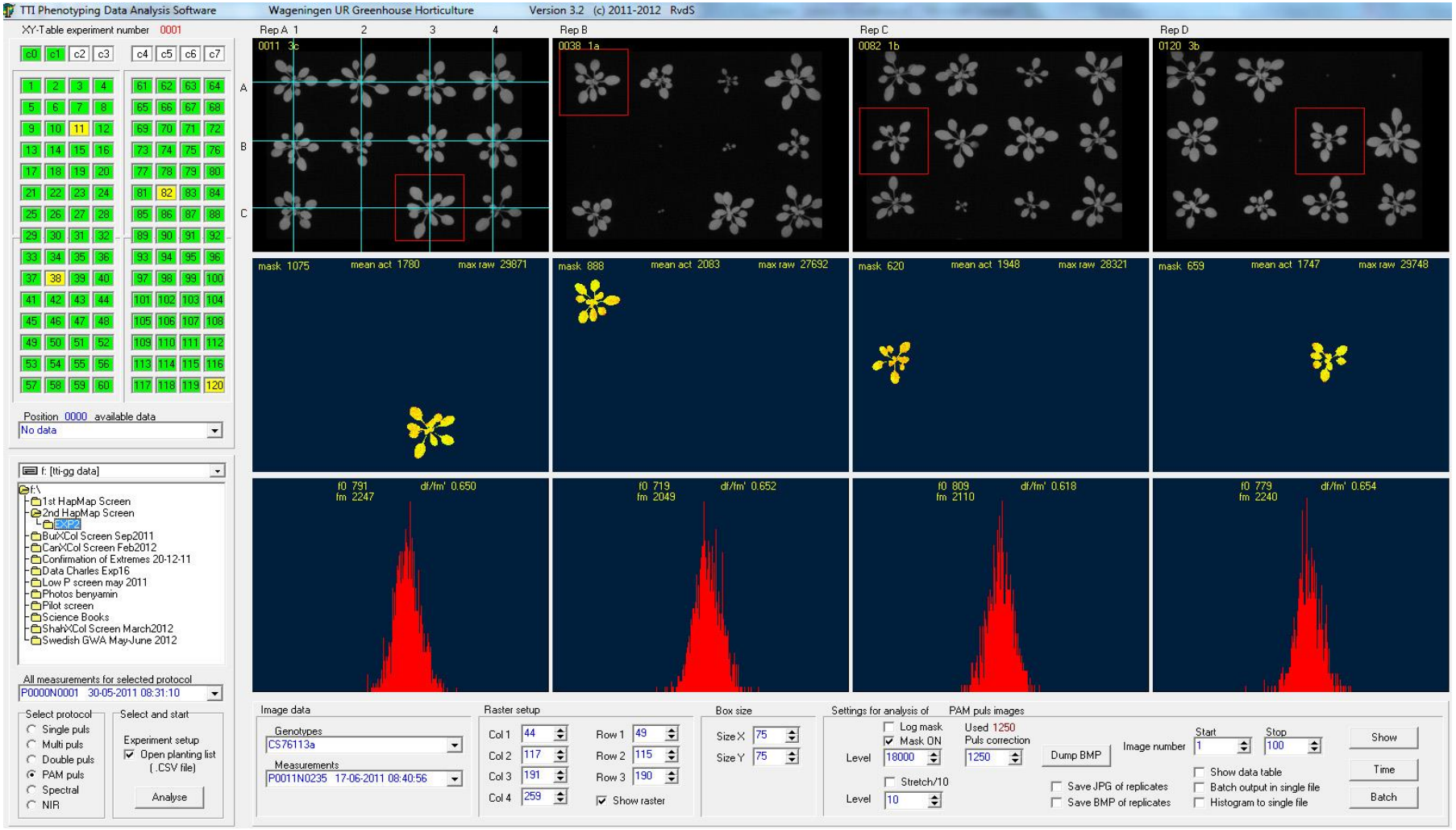


Figure S3.1. *Print screen showing analysis software. The top left panel shows all 120 imaging positions in green with the positions of the four replicates highlighted in yellow. These images of 12 plants are shown in the top row of pictures headed Rep A to Rep D. The plant which corresponds to the genotype being analysed is surrounded by a red box in each image. This plant is cut from the image using a mask level set in the control panel which is shown at the bottom of the image. The resulting image is thing shown by the middle row of pictures, not this shows a pixel map of ΦPSII distribution. This pixel map is then plotted as a histogram for each image in the last row of pictures*

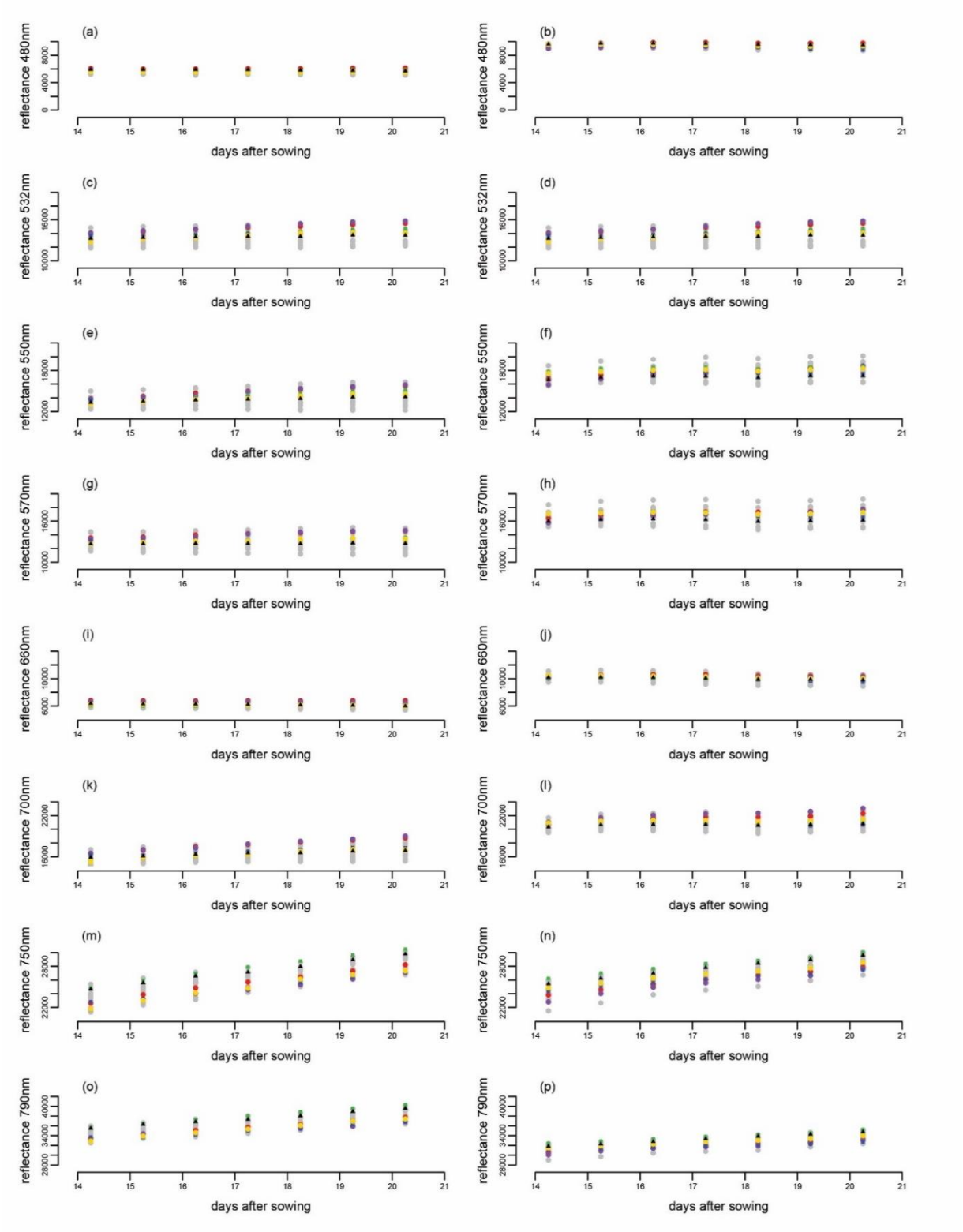


Figure S3.2. *Phenotypic variation in spectral reflectance at eight wavelengths. Phenotypic diversity in twenty Arabidopsis genotypes grown at 200 µmol m^{-2} s^{-1} light intensity (a, c, e, g, i, k, m, o), and 550 µmol m^{-2} s^{-1} light intensity (b, d, f, h, j, l, n, p). Wavelength assessed is indicated on the y axis. All data points are genotypic means (BLUEs), combining observations on replicates from different experiments into one representative value for each genotype at each time point. Six genotypes, An-2 (yellow circles), BC354 (purple circles), Bur-0 (green squares), Col-0 (black triangles), Ely (red circles) and Ts-1 (blue*

squares) are indicated in colour. Error bars have been excluded for clarity, the significance of between genotype differences is apparent from the heritability estimates in Figure S4.

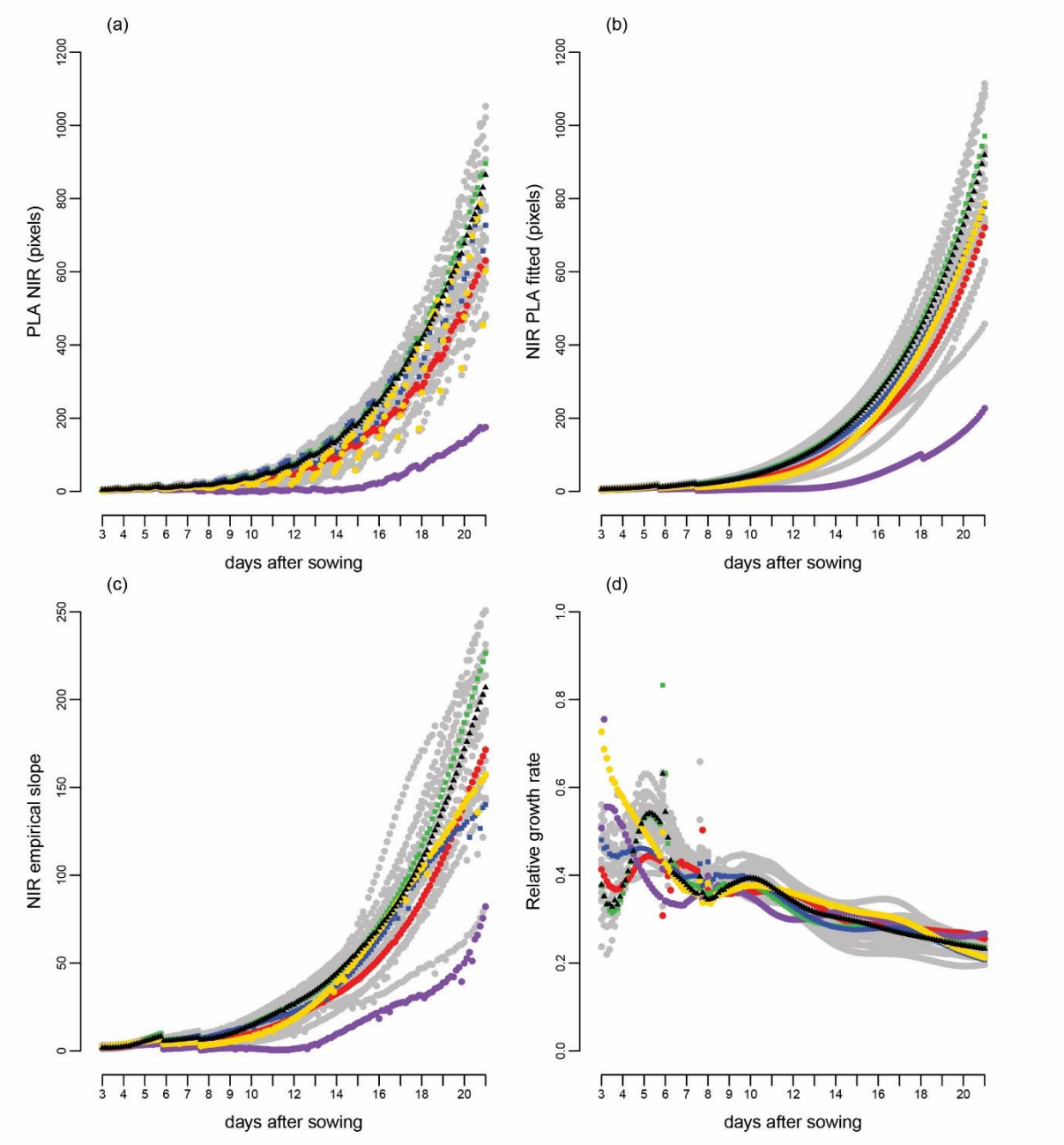


Figure S3.3. *Phenotypic variation in growth curve parameters. Phenotypic diversity in twenty Arabidopsis genotypes grown at 200 µmol m^{-2} s^{-1} light intensity. (a) Projected leaf area (PLA) from near infra-red (NIR) measurements. (b) Data from (a) fitted to a curve. (c) The empirical slope of the growth curve and (d) the relative growth rate. All data points are genotypic means (BLUEs), combining observations on replicates from different experiments into one representative value for each genotype at each time point. Six genotypes, An-2 (yellow circles), BC354 (purple circles), Bur-0 (green squares), Col-0 (black triangles), Ely (red circles) and Ts-1 (blue squares) are indicated in colour. Error bars have been excluded for clarity, the significance of between genotype differences is apparent from the heritability estimates in Figure S5.*

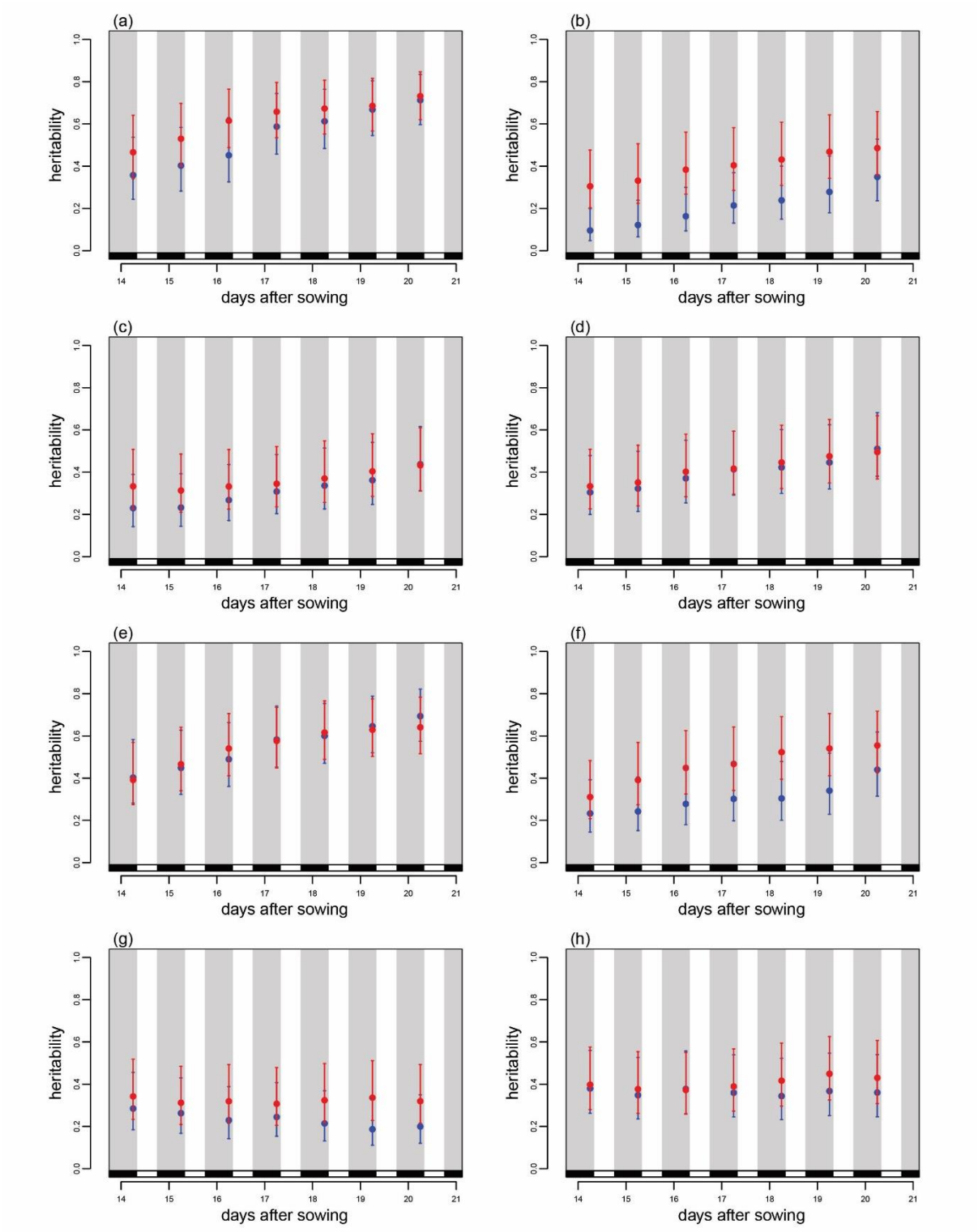


Figure S3.4. *Time course of Heritability for spectral reflectance at eight wavelengths. (a) 480nm, (b) 532nm, (c) 550nm, (d) 570nm, (e) 660nm, (f)700nm, (g) 750nm, and (h) 790nm. 200µmol m^{-2} s^{-1}(blue) and 550µmol m^{-2} s^{-1}(red), white and grey bars indicate the day night cycle. Error bars are 95% confidence intervals.*

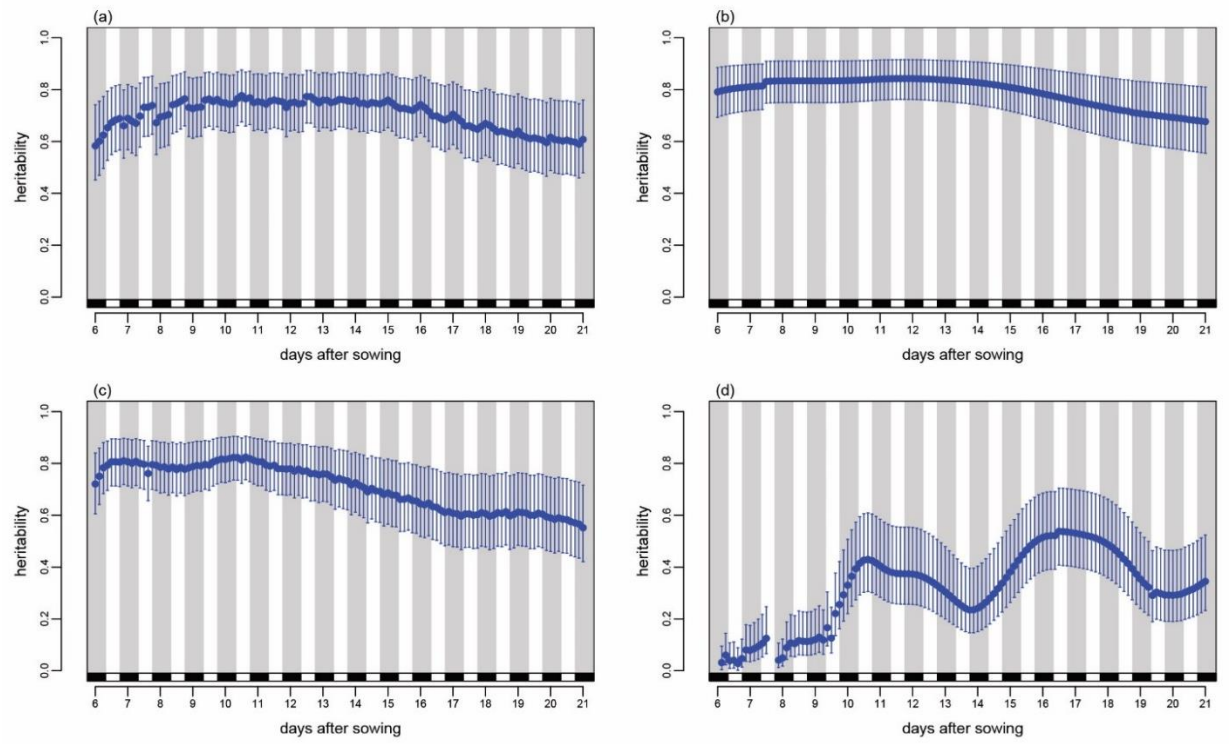


Figure S3.5. *Time course of heritability of growth curve parameters for plants grown at 200 µmol m^{-2} s^{-1} light intensity. (a) Projected leaf area (PLA) from near infra-red (NIR) measurements. (b) Data from (a) fitted to a curve. (c) The empirical slope of the growth curve and (d) the relative growth rate. White and grey bars indicate the day/night cycle. Error bars are 95% confidence intervals.*

Table S3.1: *Nutrient solution composition*

		Cations (mmol/l)		Anions (mmol/l)		Micronutrients (µmol/l)		Micronutrients (mmol/l)	
pH 6.1	6.1	NH_4	1.4	Cl	0.2	Fe	19	Si	< 0.01
EC (mS/cm)	1.5	K	5.7	S	3	Mn	11		
		Na	0.2	HCO_3	0.4	Zn	6.3		
		Ca	1.9	P	1.24	B	22		
		Mg	1.2			Cu	9.2		
						Mo	0.4		

Appendix S3.2: *Description of design factors described in Data S3.1 and Data S3.2.*

Calculated across three experiments at 200µmolm^{-2}s^{-1} and across a single experiment at 550µmolm^{-2}s^{-1}, see Appendix S3.1 for further details on experiments and number of replicates used.

Genotype Main effect of the genetic background on the trait

Exp Main effect of experiment on the trait.

genotype:Exp Interaction term describing the interaction, if any, between genotype and experiment

x_within_image Main effect of the x coordinate within the image of 12 plants taken by the camera

y_within_image Main effect of the y coordinate within the image of 12 plants taken by the camera

Table.pos Main effect of the imaging position the image of 12 plants was taken at in the table, there were 120 imaging positions.

X This is the x coordinate across the whole imaging platform

Y This is the y coordinate across the whole imaging platform

Basin Main effect indicating which of the two irrigation basin the plant was growing in.

Exp:Basin Interaction between basin and experiment

Exp:Table.pos Interaction between experiment and imaging position within table

Exp:x_within_image Interaction between experiment and the x coordinate within the image of 12 plants taken by the camera

Exp:y_within_image Interaction between experiment and the y coordinate within the image of 12 plants taken by the camera

Exp:x Interaction between experiment and the x coordinate across the whole imaging platform

Exp:y Interaction between experiment and the x coordinate across the whole imaging platform

Exp:Basin:genotype Interaction between genotype, experiment, and basin.

genotype:Exp:x_within_image Interaction between genotype, experiment, and the x coordinate within the image of 12 plants taken by the camera

genotype:Exp:y_within_image Interaction between genotype, experiment, and the y coordinate within the image of 12 plants taken by the camera

R residual error

Chapter 4

QTL analysis through time of *Arabidopsis thaliana* for photosynthesis, growth and chlorophyll reflectance reveals the dynamic nature of complex trait architecture

Pádraic J. Flood[1,2], Willem Kruijer[3], Hans Jansen[3], Sabine Schnabel[3], Fred van Eeuwijk[3], Jeremy Harbinson[2], & Mark G.M. Aarts[1]

[1]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

[2]*Horticulture and Production Physiology, Wageningen University, Wageningen, the Netherlands.*

[3]*Biometris, Wageningen University, Wageningen, the Netherlands.*

Introduction

A key attribute of all living things is their dynamism. It is an innate aspect of development and life history and an essential component of plasticity and adaption to environmental fluctuation. Despite this, the majority of quantitative genetic studies focus on a single time point or the simple comparison of control versus treatment, both of which ignore the dynamic nature of traits in response to time or environment (Wu and Lin, 2006). The recent expansion of high-throughput phenotyping techniques and the rapid advances in both computing power and genomic technologies have opened the possibility to perform quantitative trait locus (QTL) mapping not just at single time point or condition but across temporal and environmental transects. Traits which vary as a function of another continuous variable are termed function-valued traits. An example is the rate of carbon assimilation in response to changing light intensity (Stinchcombe and Kirkpatrick, 2012). Genetic mapping of these traits is referred to as functional mapping (Wu and Lin, 2006). Recently there has been an expansion in both methodological approaches and empirical studies focussing on function valued traits. From the empirical studies it has become increasingly clear that the relevant QTLs change throughout development (Yan et al., 1998; Busemeyer et al., 2013; Moore et al., 2013; Liu et al., 2014; Wurschum et al., 2014) and see Chapter 5). However with a few exceptions (Moore et al., 2013) the temporal resolution of the measurements is still low when compared to the response times biological systems are known to be capable of. Thus the research conducted to date can only provide a limited window of insight into how much the genetic variation of a trait changes along an environmental gradient or temporal progression. In other words, the question of how dynamic is the genetic architecture of a given trait in response to changing conditions, is only beginning to be answered.

In Chapter 3 of this thesis a new phenotyping platform, capable of screening 1440 Arabidopsis plants in less than an hour for photosynthetic efficiency (Φ_{PSII}), spectral reflectance, and projected leaf area (PLA) was introduced. The high throughput screening capacity of this phenotyping system enabled repeated measurements of recombinant inbred line (RIL) populations, with eight or more replicates per genotype. In this chapter, three such RIL populations, with Col-0 as a common parent, were screened for the traits just mentioned, the results of which were used to conduct both standard linkage mapping per RIL population, and joint QTL mapping using the combined populations as one set. The aim of this combined population approach was to increase both the statistical power and the mapping resolution to facilitate greater precision and confidence in the identification of QTLs (Bentsink et al., 2010;

Li et al., 2011). Joint QTL mapping has several potential advantages over traditional bi-parental mapping populations. Firstly, the inclusion of more than two parental genotypes enhances genetic diversity in the analysis and enables a more thorough exploration of the natural variation present for the trait of interest. Secondly, the use of a common parent in all populations facilitates the alignment of genetic maps as any alleles unique to the reference genotype will serve as common markers present in all populations. Thirdly, the common genetic background can help with the analysis of epistatic interactions. Fourthly, since joint QTL mapping involves several populations with a common parent the number of RIL genotypes is usually much larger than in a single bi-parental population. This greater number of genotypes increases the statistical power of the analysis therefore enhancing the detection of QTLs that might not pass the significance threshold in individual populations. Finally, each RIL line will have its own unique set of recombination events thus, the increased number of lines also means increased recombination which enhances the precision with which common QTLs can be mapped.

The aim of this chapter was to make use of the phenotyping facility described in Chapter 3 to combine functional and joint QTL mapping to investigate the QTL profile of Φ_{PSII}, chlorophyll reflectance and PLA through time.

Material and methods

Plant material

All RIL populations were obtained from the Versailles Arabidopsis Stock Centre (http://publiclines.versailles.inra.fr/) and are described by Simon et al. (2008). The parental lines were: Bur-0 an Irish accession; Can-0 from the Canary Islands; Shah collected in Tajikistan; and Col-0, the reference genotype. All populations were derived from a cross between one of the three wild accessions and the Col-0 reference genotype, which was used as a male in all crosses. Experiments were conducted using the F9 generation of the core set of 164 RILs for Bur x Col and Can x Col and162 for Shah x Col population as described in Simon et al. (2008). In total 490 RIL lines were screened in a minimum of 8 replicates each along with the four parental lines and one additional genotype, Ely, which due to its extreme Φ_{PSII} values was used to confirm that the genotype names corresponded to correct table positions and validate that the Φ_{PSII} measurements were accurate.

Plant growth and phenotyping

Each RIL population was screened in an independent experiment using the phenotyping platform described in Chapter 3. Plants were screened three time per day for Φ_{PSII} at one, four and seven hours into the photoperiod. The mask from these images was used to estimate PLA. The plants were also imaged using 790 nm illumination, (near infra-red or NIR imaging) every three hours starting at midnight of each day. These eight daily measurements were used to estimate the percentage change in PLA every three hours, diurnal fluctuations in which were used as a proxy for epinastic leaf movement. Finally, reflectance at eight wavelengths (480 nm, 532 nm, 550 nm, 570 nm, 660 nm, 700 nm, 750 nm and 790 nm) was measured one hour before the start of the photoperiod. Using the reflectance at 700 nm and 790 nm, chlorophyll reflectance (Chl) was estimated using the Gitelson et al. (2003) formula $Chl=(R700^{-1}-R790^{-1})\cdot R790$. This is a proxy for chlorophyll content and as such is a measure of greenness. In addition to each population being screened independently, a final experiment was conducted where 20 RILs from each population (19 in the case of ShaXCol), comprising of the photosynthetic extremes, were grown. Plants were grown in a climate controlled growth room set to a 10/14hr day/night cycle, 70% relative humidity, 20/18°C day/night temperatures and 200 µmol m^{-2} s^{-1} light. For this chapter a subset of phenotypes, Φ_{PSII}, Chl, PLA and percentage change were analysed.

Estimation of corrected means

Each RIL was planted in a minimum of eight replicates per experiment. To ensure that the replicates were evenly distributed the growth area of the phenotyper was divided up into four sections each containing 360 positions. Two replicates from each RIL were randomly sown in each of the four sections. The position of each replicate in the table was recorded as a design effect. The phenotyper imaged 12 plants at a time (an image of 3 x 4) plants, and the position of the plant within the image was recorded. The three genotypes (Col-0, Bur-0 and Ely) occurring across all experiments and the final experiment where 20 RIL lines from each population were grown alongside each other were used to estimate any bias introduced due to random variations between experiments. These design effects were included in a mixed model and used to calculate corrected genotypic means or best linear unbiased estimators (BLUEs). Further details on this procedure can be found in Chapter 3. The inclusion of design effects in the estimation of these means enabled the data from the different experiments to be analysed together which was essential for the joint QTL analysis.

Genetic mapping

For all traits, mapping was performed for each population separately and for all populations combined using the R/qtl package (Broman et al., 2003). For the combined analysis a common genetic map was constructed. As described previously (Simon et al., 2008) markers on the upper arm of Chromosome 3 in the Shah x Col population showed complete linkage, most likely due to the presence of a large inversion, which suppresses crossovers (Lysak and Mandáková, 2013). Since R/qtl will not work when more than one marker occupies the same genetic position, this complete linkage was overcome using the jittermap function in R/qtl, which moves the markers slightly apart. The cal.genoprob function was used to calculate the probability of having a particular genotype at a given position[19], this augments the genotype information by modelling multiple genotypes and estimating their probabilities as opposed to guessing just one likely genotype. This improves the mapping procedure by giving a more continuous score of genotype across the genome, resulting in smoother logarithm of odds (LOD) curves. A step of 1 centiMorgan (cM) was taken when estimating these augmented genotype profiles. All R/qtl analysis was done with a single QTL model (scanone) using a Haley-Knot regression (Broman et al., 2003), with Φ_{PSII} at one selected time-point per population analysed with a multiple QTL mapping (MQM) model (Arends et al., 2010) and

[19] Each RIL population was genotyped with approx. 80 markers so most of the genome is ungenotyped, when imputing the intermarker genotypes the cal.genoprob function improves the quality of the genetic imputation.

tested for epistatic and additive interactions using a two locus genome scan (scantwo) (Broman et al., 2003).

Functional mapping and data visualisation

Genetic mapping of phenotypes measured over time can be simply achieved by independently mapping at each time-point in a stepwise fashion and representing the results by plotting the LOD score against genetic position. One problem with this approach is that as the number of time-points and traits increase the number of plots becomes progressively larger and more difficult to handle. To overcome this, Kwak et al. (2014) developed the R package funqtl (R Core Team, 2014) which illustrates the plot of the LOD scores through time using colour intensity. This allows the numerous LOD plots to be combined into one plot greatly aiding the visualisation of LOD profiles through time. In addition to conducting a QTL analysis where each time point is treated as independent, Kwak et al. (2014) also developed two methods for integrating the information across all time-points to give a score of the overall evidence for a QTL. The MLOD statistic looks for the presence of a significant association at any time point and is useful for detecting QTLs of large effects which are present for only a brief interval of time. The SLOD statistic, on the other hand, is more powerful at detecting loci with effects which extend over a large period. For each analysis 5% significance thresholds were determined by a permutation test of 1000 permutations and plotted as dashed lines, blue for MLOD and red for SLOD.

The plots of LOD score through time and the MLOD and SLOD statistics generate a classical 2D static representation of the data which is suitable for printing. Function valued traits are, however, defined by the fact that they are not static but vary along some gradient, be it time, light or another variable. Therefore a more suitable representation of such traits, and the genetic mapping results which they generate, would allow the user to interact with the results. To address this, a new R package has recently been developed, called qtlcharts, which can create digital interactive plots (Broman, 2015). By their very nature they cannot be printed and are available for download on the compact disc at the back of this thesis. They need to be downloaded to your computer and viewed in a browser, no internet connection is required. These interactive plots allow a much more detailed and personalised exploration of the data, see Figure 4.1 for a print screen illustration of how to interpret them. The data is also represented in print should you prefer not to view these charts.

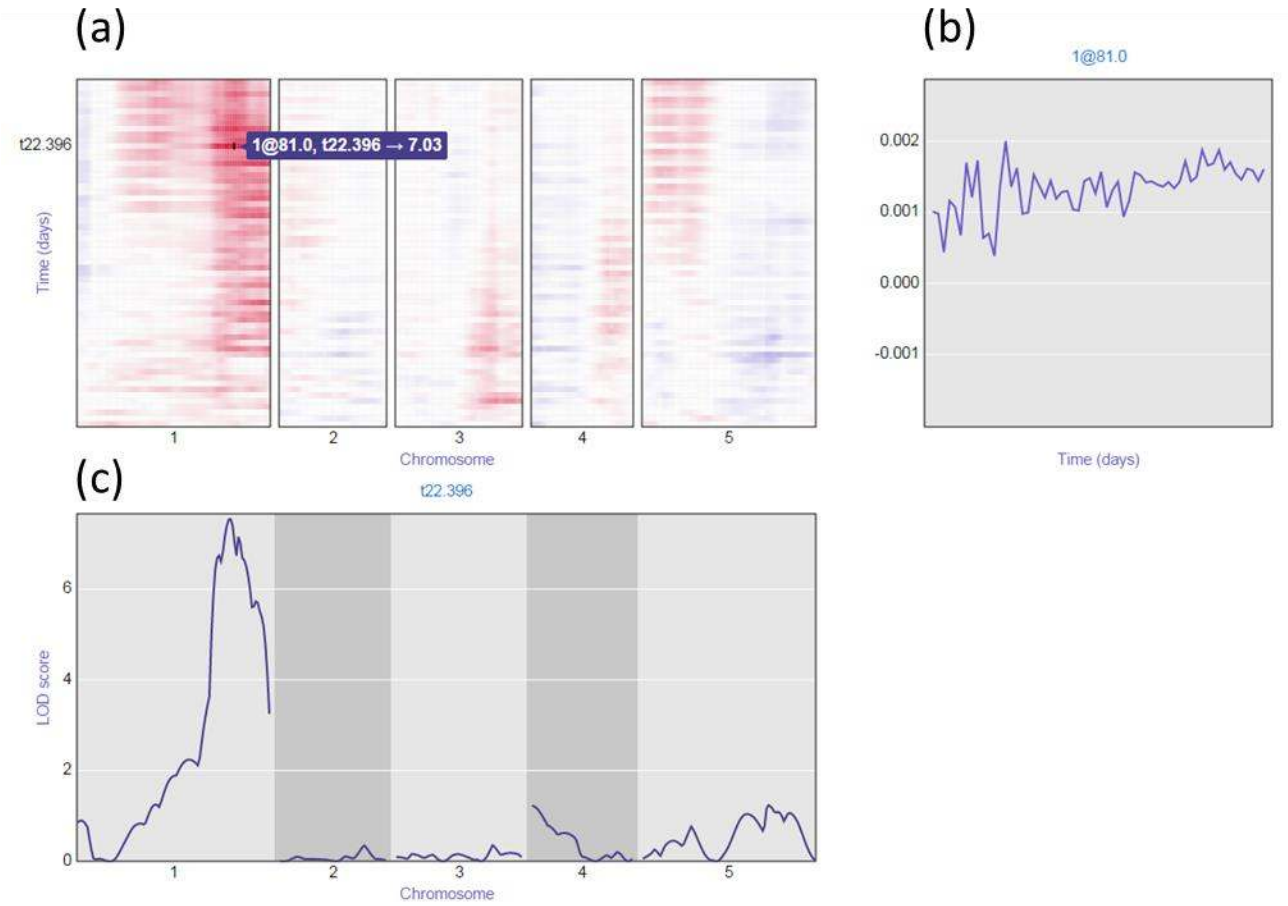


Figure 4.1. *Print screen of the interactive plot for Φ_{PSII} through time in the Bur x Col RIL population (interactive Figure 4.3). (a) A heatmap of the stepwise QTL analysis at each time-point, x-axis indicates genetic position and the y-axis shows the measurements through time. The dark blue box reading '1@81.0, t22.396 →7.03' indicates the location of the mouse cursor when the print screen was taken. The text in it is interpreted as, "Chromosome 1, the 81st measurement, 22.396 days after sowing, gives a LOD score of 7.03". The position at which the cursor is hovering is linked to panels (b) and (c) which indicate the estimated QTL effect on Φ_{PSII} through time (b), and the LOD curve at the particular time-point (c). In panel (a) red indicates the Bur-0 allele increases the phenotype value and blue indicates the Col-0 allele increases the phenotype value.*

Results

The distribution of Φ_{PSII} through time is shown in Figure 4.2 (see Supplemental Figures 4.1 to 4.3 for plots of the other phenotypes), the corrected means of the parents (BLUEs) are plotted in each panel to facilitate visual comparison. Perhaps the most striking result is that even though the populations all share Col-0 as a common parent the phenotypic distribution of the RILs is quite different. The Bur x Col population is shifted towards higher Φ_{PSII} values whilst the Can x Col shows the opposite, being shifted towards lower values. This bias of the phenotypic distribution per population mirrors the phenotypic positioning of the parents relative to one another, with Bur-0 showing the highest values and Can-0 the lowest. Even though Shah is phenotypically most similar to Col-0, the phenotypic distribution of this population is the widest, extending almost as far as the distribution of phenotypic values seen when all populations are plotted together.

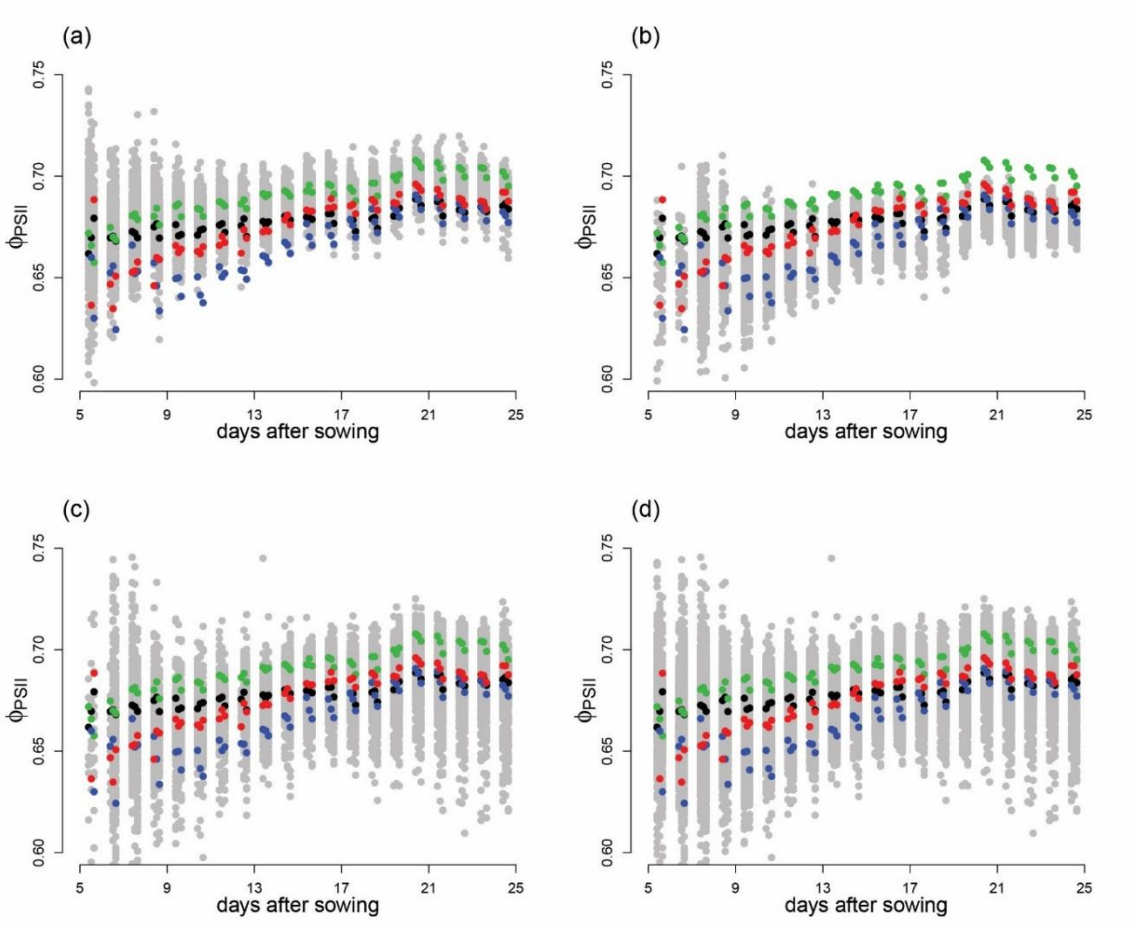


Figure 4.2. *Best linear unbiased estimators (BLUEs) of Φ_{PSII} for; (a) The Bur x Col population, (b) the Can x Col population, (c) the Shah x Col population and (d) all RILs combined. The BLUEs of the four parental genotypes across all experiments are included in each plot to facilitate comparison, Col-0 = black, Bur-0 =green, Can-0=blue and Sha=red. See supplemental figures 4.1-4.3 for plots of the other traits.*

Figures 4.3 to 4.6 shows the time dependency of the QTL profile for Φ_{PSII} through time for the each RIL population. From these figures it can be seen that the strengths of the QTLs vary over time with some not present at certain time-points. This can been seen more clearly interactive version of Figures 4.3-4.6 where the phenotypic effect of the different loci can be seen through time as well as LOD profiles for each time-point (see panels b and c of Figure 4.1). In the Bur x Col population (Figure 4.3) the only statistically significant QTL is found at the bottom of chromosome 1, named photosynthesis 1 (PS1). This QTL becomes more significant as time passes exhibiting its highest LOD scores 22 days after sowing, with the Bur-0 allele increasing Φ_{PSII} by as much as 0.002 (Figure 4.3a and interactive Figure 4.3). No other loci breach the significance threshold, but one suggestive QTL at the top of chromosome 5 is interesting in that it is only present in afternoon and evening measurements. In addition the plot of its effect through time shows a diurnal fluctuation (see interactive Figure 4.3).

In the Can x Col RIL population (Figure 4.4) a major QTL for Φ_{PSII} is found at the top of chromosome 4 (PS2) where the Can-0 allele causes a reduction in Φ_{PSII} values by up to 0.006. An initial candidate in this genomic region was the flowering time gene *FRIGIDA* (*FRI*) which is known to be highly pleiotropic and could conceivably affect Φ_{PSII}. To test this we phenotyped a late flowering Col-0 line with an active *FRI* allele introgressed from the Sf-2 genotype which complements Col-0 non-functional allele (Lee et al., 1993). We assayed 12 replicates plants of Col-0 *FRI* and found no difference when compared with Col-0 *fri* indicating that allelic variation at *FRIGIDA* is not responsible for the difference in Φ_{PSII} between Col-0 and Can-0. Further evidence supporting this conclusion is that both Bur-0 and Sha parental lines also harbour active *FRI* alleles (Brachi et al., 2010) and no QTL for Φ_{PSII} was detected near this locus in either population (Figures 4.3 and 4.6). Apart from the locus at the top of chromosome 4, two additional QTLs on chromosomes 3 (PS3) and 5 (PS4) were also detected in this population. In both cases the Can-0 allele caused a reduction in Φ_{PSII} (Figure 4.4). PS3 and PS4 are most significant and have the largest phenotypic effects at earlier stages than PS2 which peaks in significance around 16 days after sowing (DAS) whilst at this stage PS3 and PS4 are already beginning to dissipate (Best viewed in interactive Figure 4.4).

In the Shah x Col population (Figure 4.5) three significant QTLs were detected for Φ_{PSII}. The QTLs at the bottom of chromosomes 2 (PS5) and 4 (PS6) become increasingly significant as time passes, PS6 becoming significant from 16 DAS and PS5 from 20 DAS, in both cases the Shah allele increases Φ_{PSII} by up to 0.004. The QTL at the base of chromosome 5 (PS7)

shows the opposite pattern being significant earlier during the experiment, peaking at around 10 DAS, and in this case the Col-0 allele causes increased Φ_{PSII} values up to 0.003.

The joint QTL analysis for Φ_{PSII} (Figure 4.6) does not result in the detection of any new QTLs. PS1 which is found in the BurXCol population and is only a suggestive peak in the Can x Col and Shah x Col populations became significant in the combined population analysis. For this locus the Col-0 allele reduces Φ_{PSII} relative to Bur-0, Can-0 and Shah. The second QTL found in the combined analysis is PS4/7, this QTL was found in both Can x Col and Shah x Col populations (Figures 4.4 and 4.5) and was suggestive in the BurXCol population (Figure 4.3). In this case the Col-0 allele increases Φ_{PSII} relative to the other parental accessions. PS2 and PS3 also breach the significance threshold, both loci collocate with QTLs detected only in the Can x Col population (Figure 4.4), though in the case of PS3 there is a suggestive peak in the Shah x Col population (Figure 4.5).

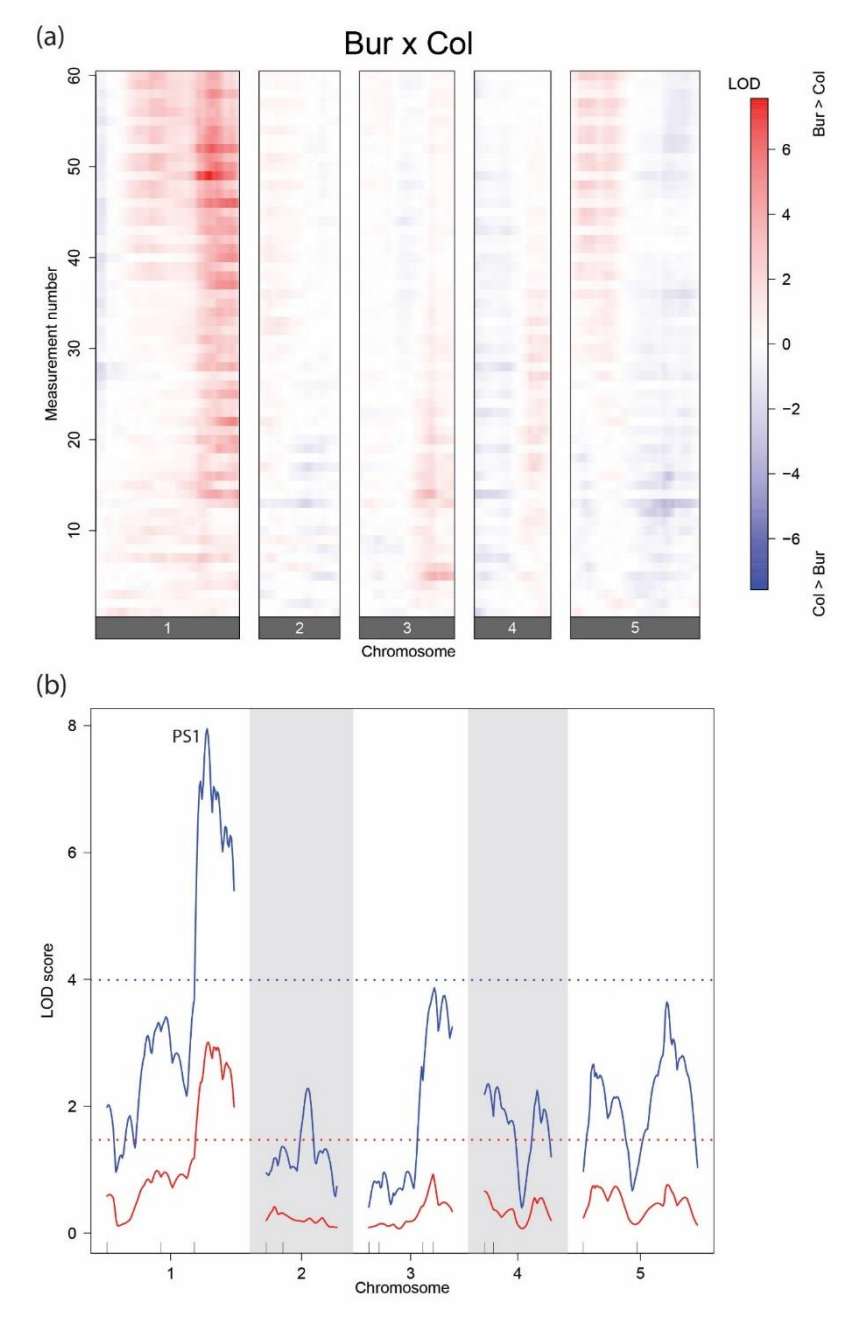


Figure 4.3. *QTL analysis of Φ_{PSII} for the Bur x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.3). Colours indicate the direction of the QTL effect with red indicating the Bur-0 allele increases Φ_{PSII} and blue indicating Col-0 increases Φ_{PSII}, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for Φ_{PSII} across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

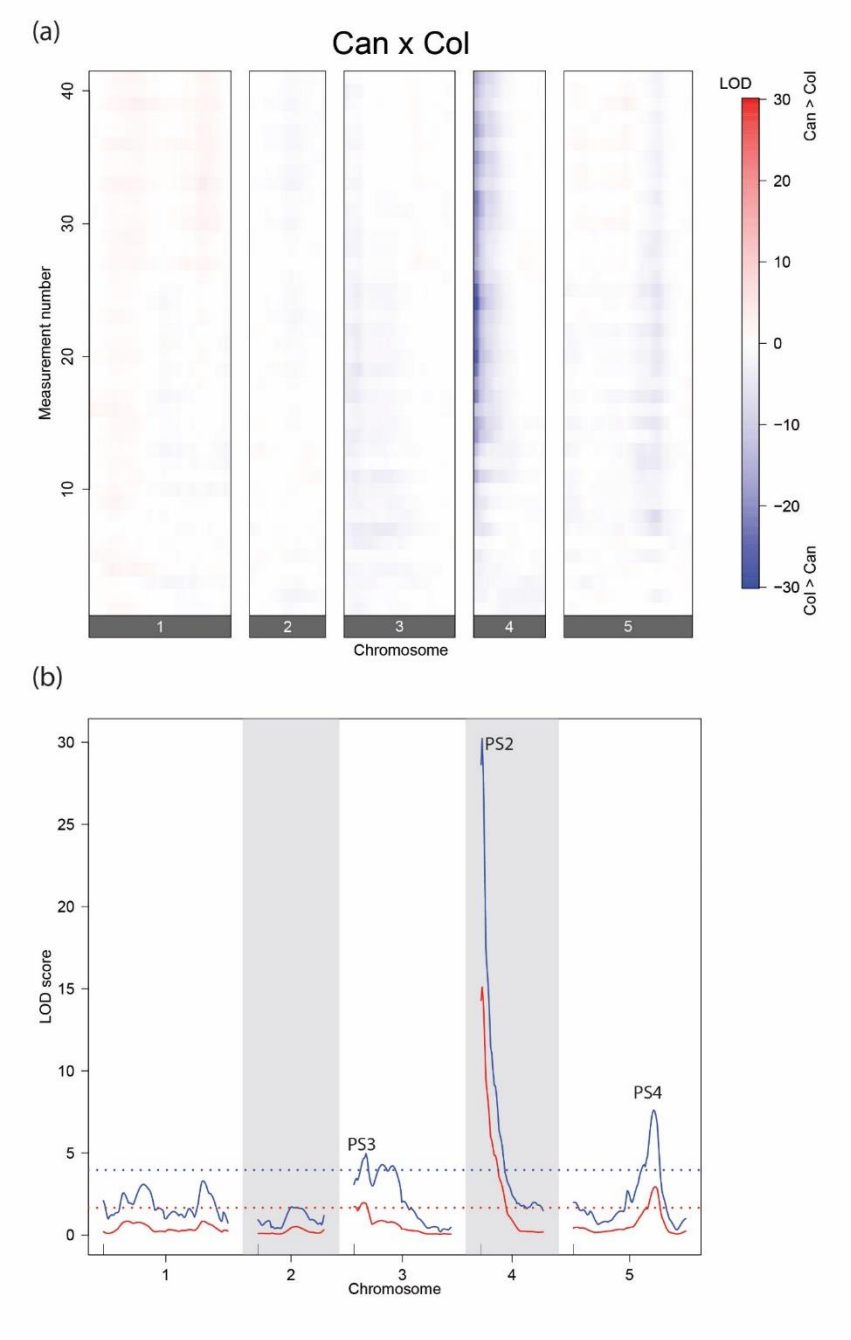


Figure 4.4. *QTL analysis of Φ_{PSII} for the Can x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.4). Colours indicate the direction of the QTL effect with red indicating the Can-0 allele increases Φ_{PSII} and blue indicating Col-0 increases Φ_{PSII}, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for Φ_{PSII} across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

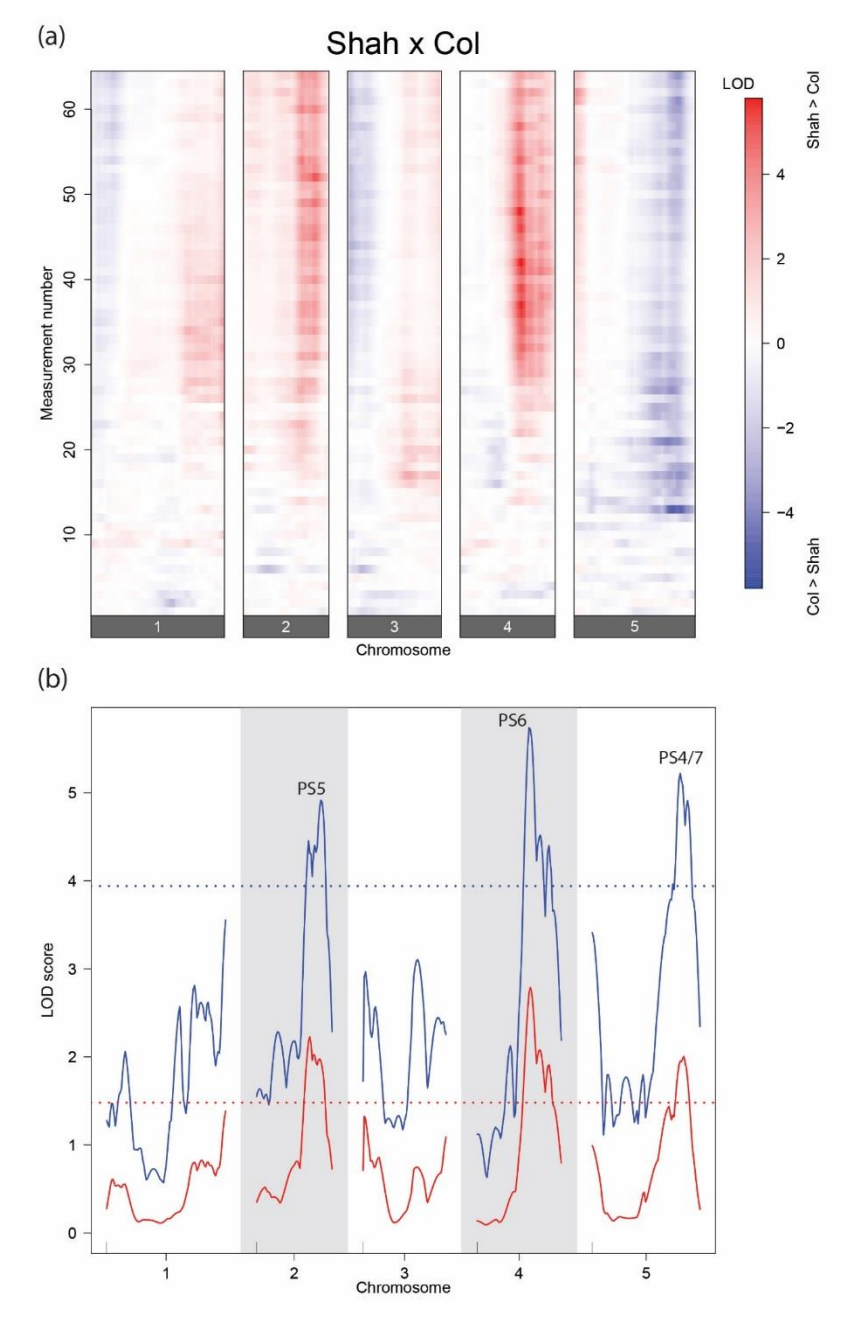


Figure 4.5. *QTL analysis of Φ_{PSII} for the Shah x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.5). Colours indicate the direction of the QTL effect with red indicating the Shah allele increases Φ_{PSII} and blue indicating Col-0 increases Φ_{PSII}, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for Φ_{PSII} across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

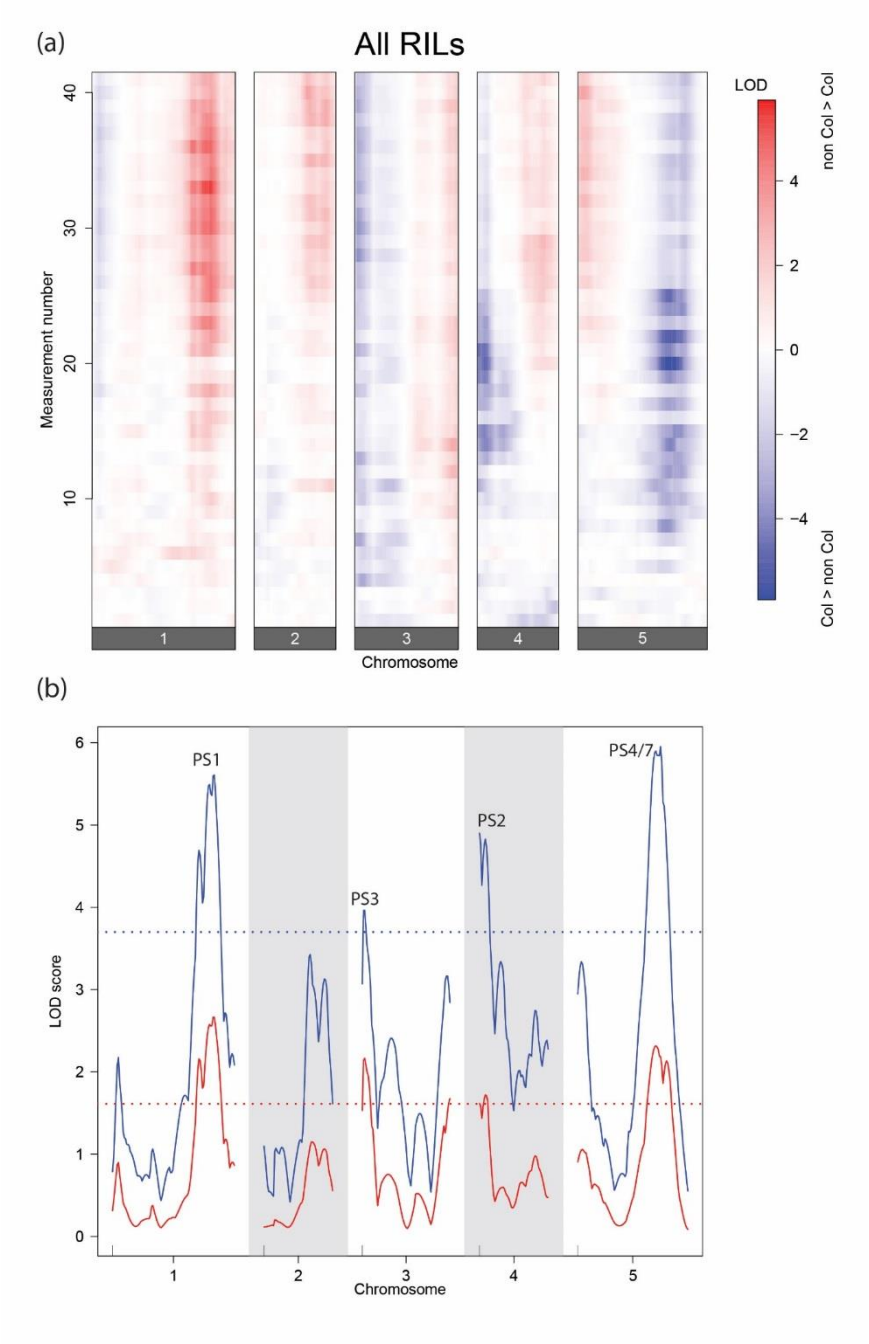


Figure 4.6. *Joint QTL analysis of Φ_{PSII} combining all RIL populations, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.6). Colours indicate the direction of the QTL effect with red indicating the Non Col-0 allele increases Φ_{PSII} and blue indicating Col-0 increases Φ_{PSII}, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for Φ_{PSII} across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

Several QTLs were detected for PLA, the general trend across all populations and the joint QTL analysis is that the Col-0 alleles result in reduced values (Figures 4.7 to 4.10 and interactive Figures 4.7 to 4.10). This is somewhat surprising since all parents show similar PLA values (Supplementary Figure 4.2). In the Bur x Col population a major QTL was detected at the top of chromosome 5 plant area 1 (PA1), this QTL became more significant through time with its largest effect at the later time-points (Figure 4.7). Two additional QTLs were also detected at the bottom of chromosomes 3 (PA2) and 4 (PA3). Three other loci, two at the upper end of chromosome 4 and one at the top of chromosome 1, are just below the significance threshold. PA1, PA2, PA3 and all three suggestive QTLs show the Bur-0 allele increasing PLA (Figure 4.7a). PA2 and PA3 are most significant early during the growth period and begin to dissipate around 10DAS. The suggestive locus at the top of chromosome 1 shows the opposite trend only beginning to become significant from 22DAS. The two suggestive loci at the upper end of chromosome 4 are weakly significant throughout the experiment as would be suggested by their relatively greater significance with the SLOD statistic than with the MLOD statistic (Figure 4.7b).

In the Can x Col population there is one significant locus at the bottom of chromosome 5 (PA4), again the Col-0 allele results in low PLA values (Figure 4.8). Another locus at the top of chromosome 3 (PA5) is only significant under the MLOD statistic and is the only one so far where the Col-0 allele results in an increase in PLA. This QTL is significant for the early measurements declining from 8DAS. There is also a suggestive QTL at the bottom of chromosome 1 just reaching the MLOD threshold, which is only significant from 23DAS onwards (Figure 4.8a).

In the Shah x Col population no QTL breached the SLOD threshold, although two QTLs on chromosome 5 breach the MLOD threshold (Figure 4.9). The locus at the lower end of chromosome 5 (PA6) becomes significant from 21DAS onwards and the Col-0 allele reduces PLA. The locus at the top of chromosome 5 (PA7) is unusual in that the Col-0 allele increases PLA up until 8DAS after which the locus become non-significant only to become significant again 24DAS but with the Col-0 allele now reducing PLA (Figure 4.9). This opposite effect of loci in the same genomic region at different stages during the experiment perhaps explains why this locus is detected as significant for the MLOD statistic but not for the SLOD statistic (Figure 4.9b).

The joint QTL analysis detected one new significant locus at the top of chromosome 4 (PA8) (Figure 4.10). This locus is most significant early in the growth period, dissipating after

9DAS, hence it is only significant for the MLOD statistic. The Col-0 allele reduces PLA for this locus. PA1 and PA2 were detected in the joint QTL analysis as well as a QTL at the base of chromosome 5 with collocates with PA4 and PA6 and is thus named PA4/6 (Figure 4.10).

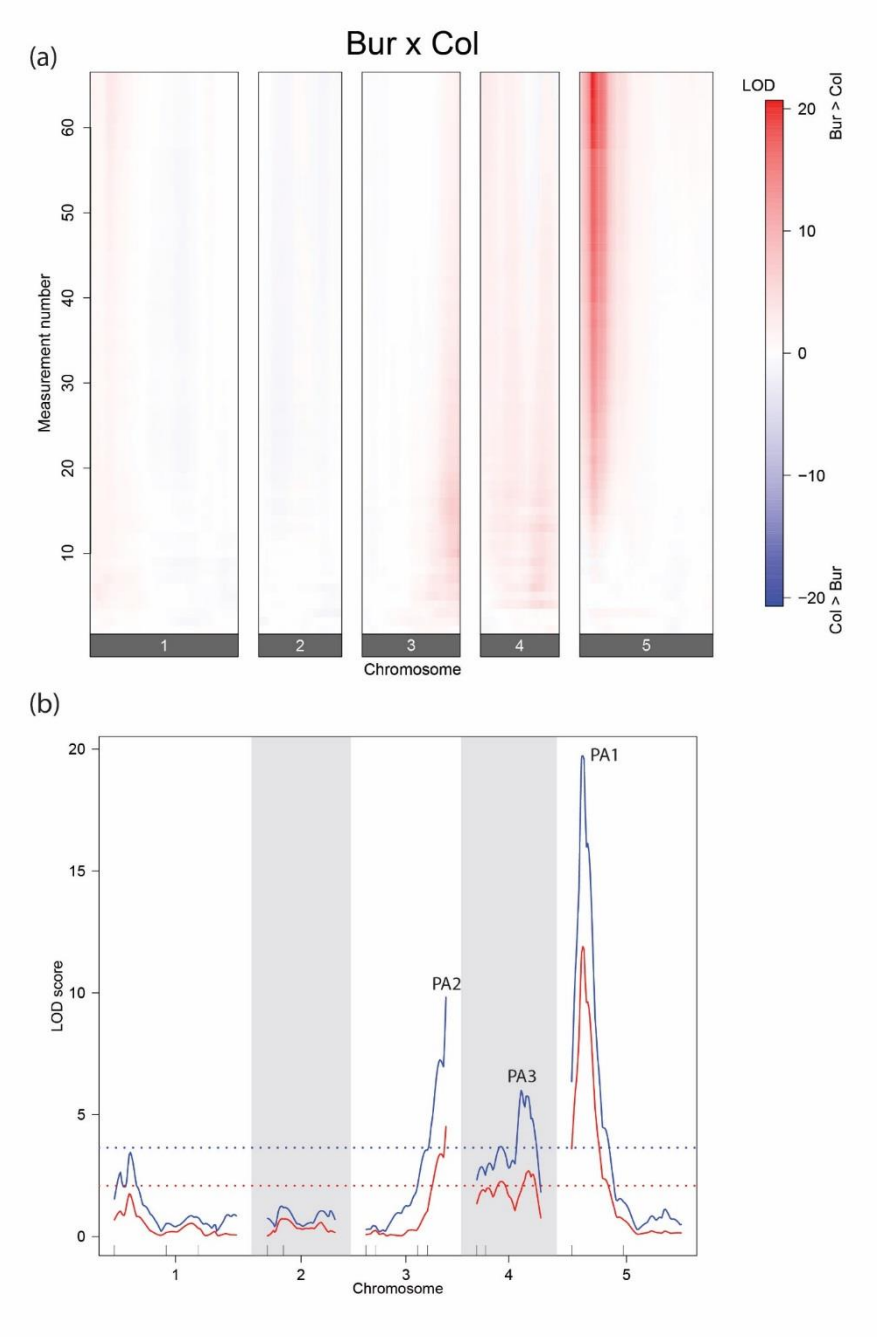


Figure 4.7. *QTL analysis of projected leaf area (PLA) for the Bur x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.7). Colours indicate the direction of the QTL effect with red indicating the Bur-0 allele increases PLA and blue indicating Col-0 increases PLA, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

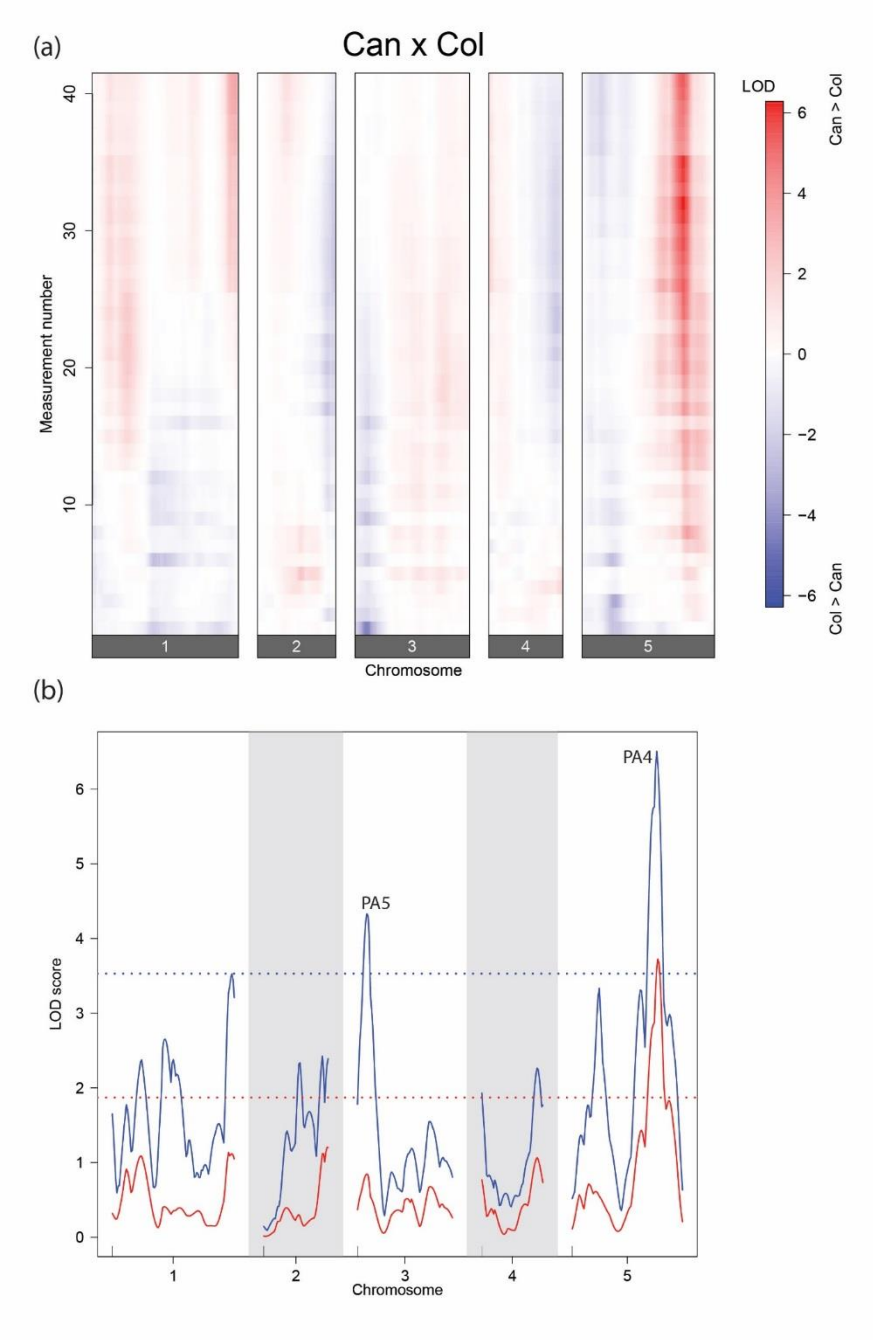


Figure 4.8. *QTL analysis of projected leaf area (PLA) for the Can x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.8). Colours indicate the direction of the QTL effect with red indicating the Can-0 allele increases PLA and blue indicating Col-0 increases PLA, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

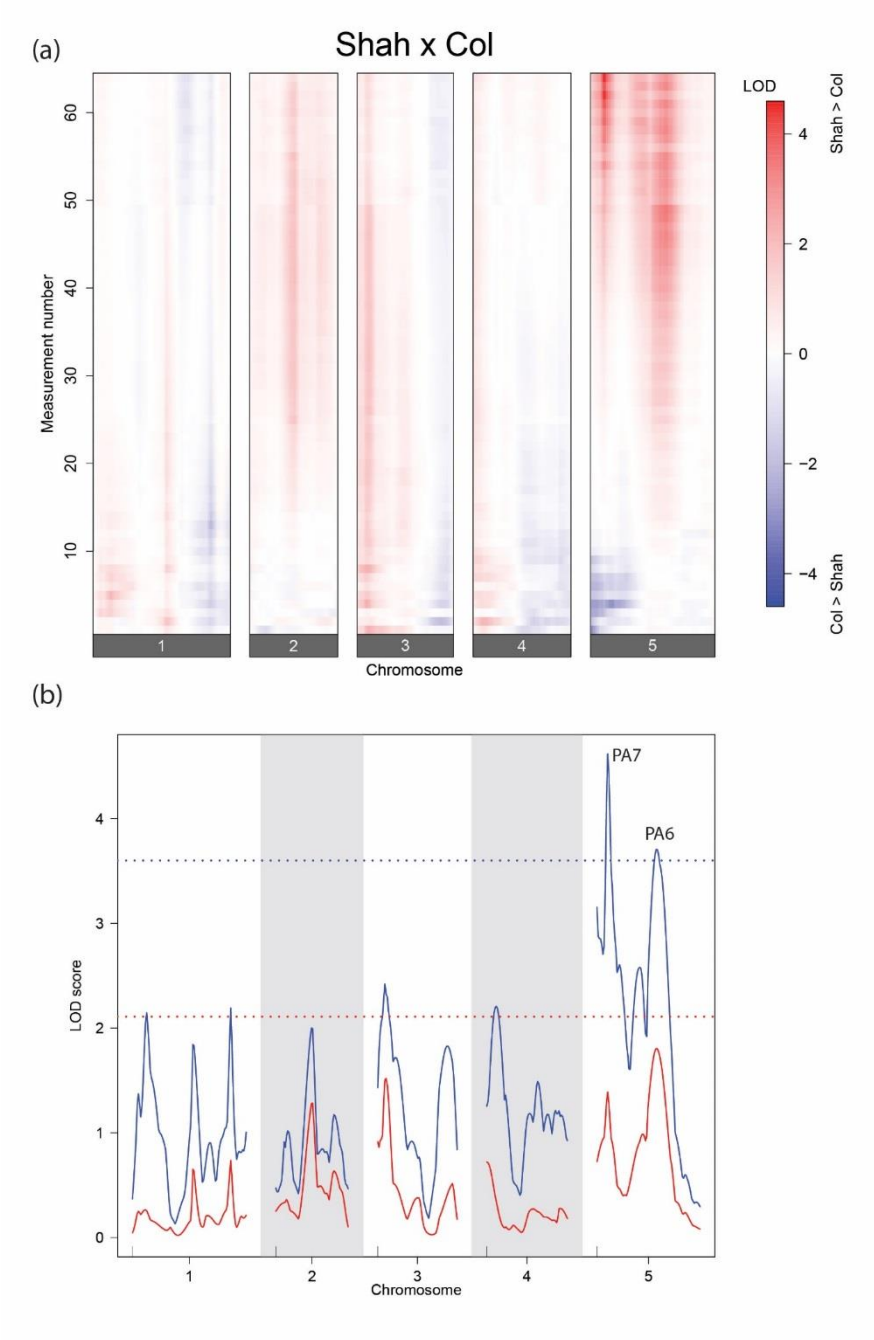


Figure 4.9. *QTL analysis of projected leaf area (PLA) for the Shah x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.9). Colours indicate the direction of the QTL effect with red indicating the Shah allele increases PLA and blue indicating Col-0 increases PLA, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

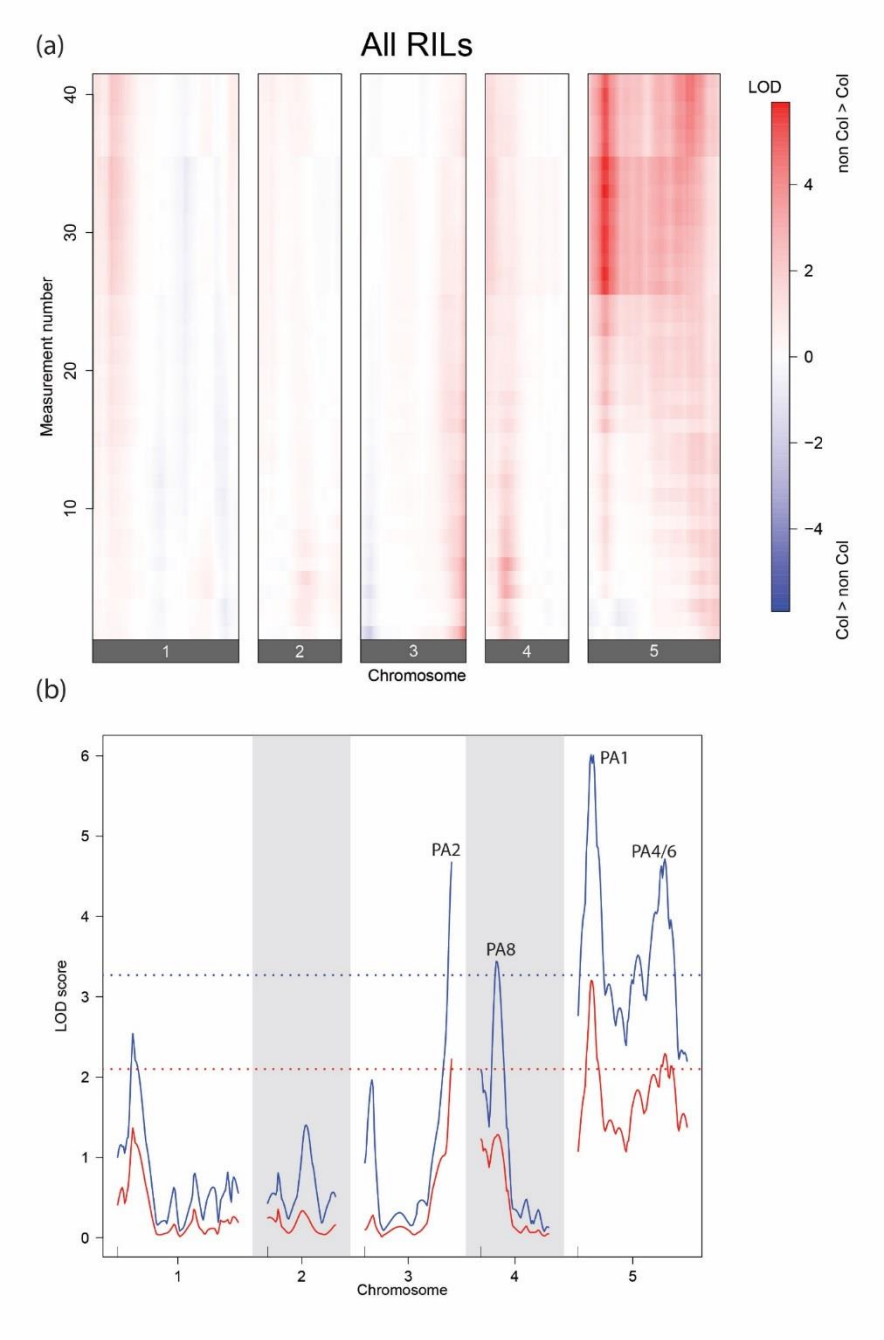


Figure 4.10. *Joint QTL analysis of projected leaf area (PLA) for all RIL populations combined, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.8). Colours indicate the direction of the QTL effect with red indicating the non Col-0 allele increases PLA and blue indicating Col-0 increases PLA, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

Chlorophyll reflectance resulted in the detection of a number of QTLs, in contrast to PLA the Col-0 alleles generally increased the trait value (Figures 4.11 to 4.14). The Bur x Col population contained one highly significant QTL at the top of chromosome 5 named chlorophyll

1 (CL1) (Figure 4.11). Two additional loci just breach the significance threshold at the tops of chromosomes 1 (CL2) and 4 (CL3). In all cases the Col-0 allele increases greenness (Figure 4.11a).

The Can x Col population has four strong QTLs for chlorophyll reflectance, one at the bottom of chromosome 1 (CL4), another two at the tops of chromosomes 3 (CL5) and 4 (CL6) and one more towards the end of chromosome 5 (CL7) (Figure 4.12). In all cases the Col-0 allele increases chlorophyll reflectance (Figure 4.12a) which is perhaps to be expected as Can-0 is the palest green of the four parental accessions (Supplemental Figure 4.1). CL6 is more significant earlier in the growth period peaking 16DAS and then declining to non-significance by 25DAS. In contrast CL4 and CL5 only begin to become significant after 16DAS and CL7 is significant at all time-points measured.

In the Shah x Col RIL population three QTLs were detected, two at opposite ends of chromosome 3 and with opposing effects (Figure 4.13). The Col-0 allele increases the trait value for the QTL at the top of chromosome 3 (CL8) and decreases it for the QTL at the bottom (CL9) with both loci becoming significant from 22DAS (Figure 4.13a and interactive Figure 4.13). The most significant locus detected in this population is a broad peak at the bottom of chromosome 5 (CL10) (Figure 4.13b).

In the combined QTL analysis no new loci were found to be significant (Figure 4.14). The QTL at the top of chromosome 3 collocated with CL5 and CL8 and was only detected above the MLOD threshold. The broad peak at the end of chromosome 5 collocated with CL7 and CL10, in both CL5/8 and CL7/10 the Col-0 allele increased the trait value (Figure 4.14a).

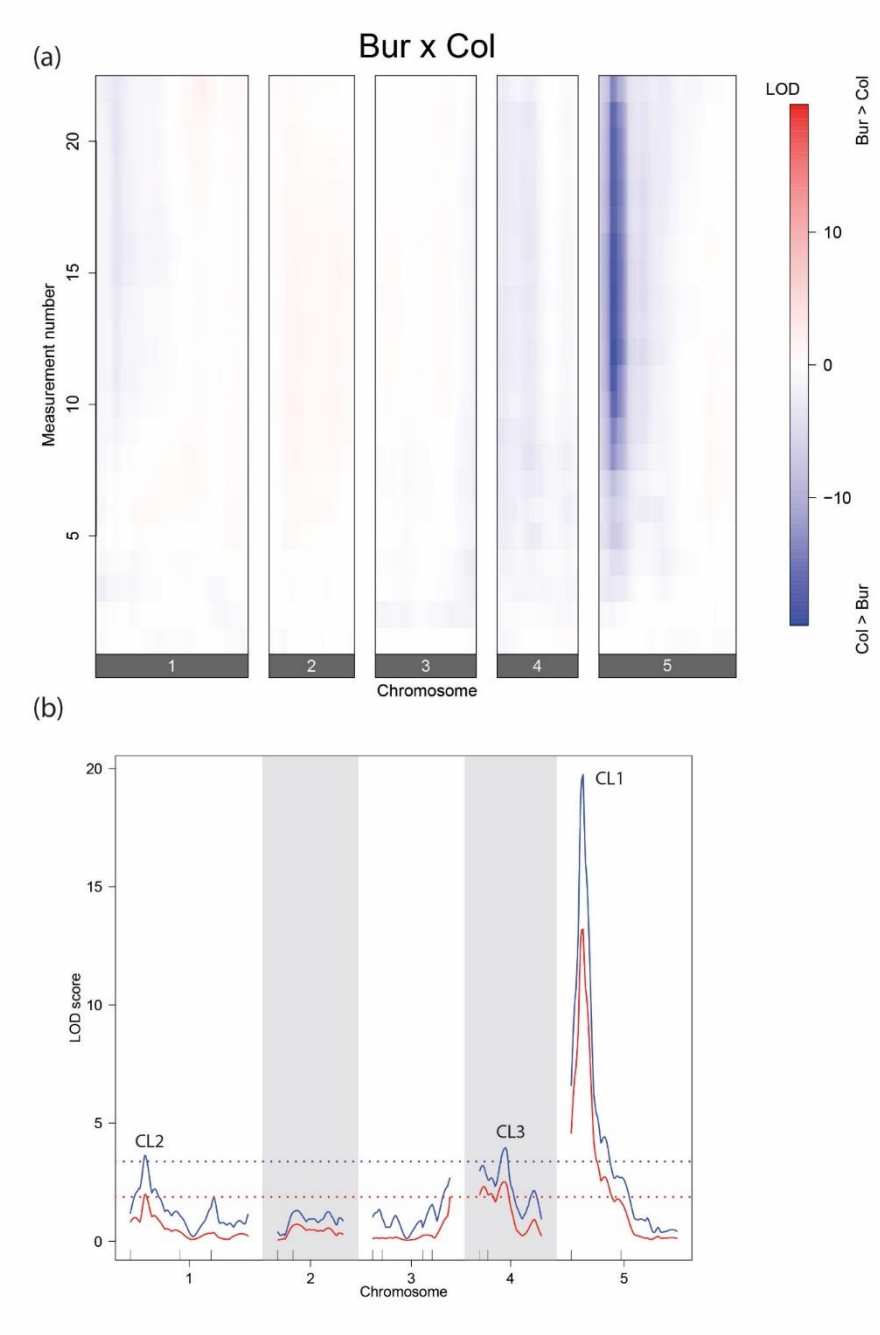


Figure 4.11. *QTL analysis of chlorophyll reflectance (CHL) for the Bur x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.11). Colours indicate the direction of the QTL effect with red indicating the Bur-0 allele increases CHL and blue indicating Col-0 increases CHL, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for CHL across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

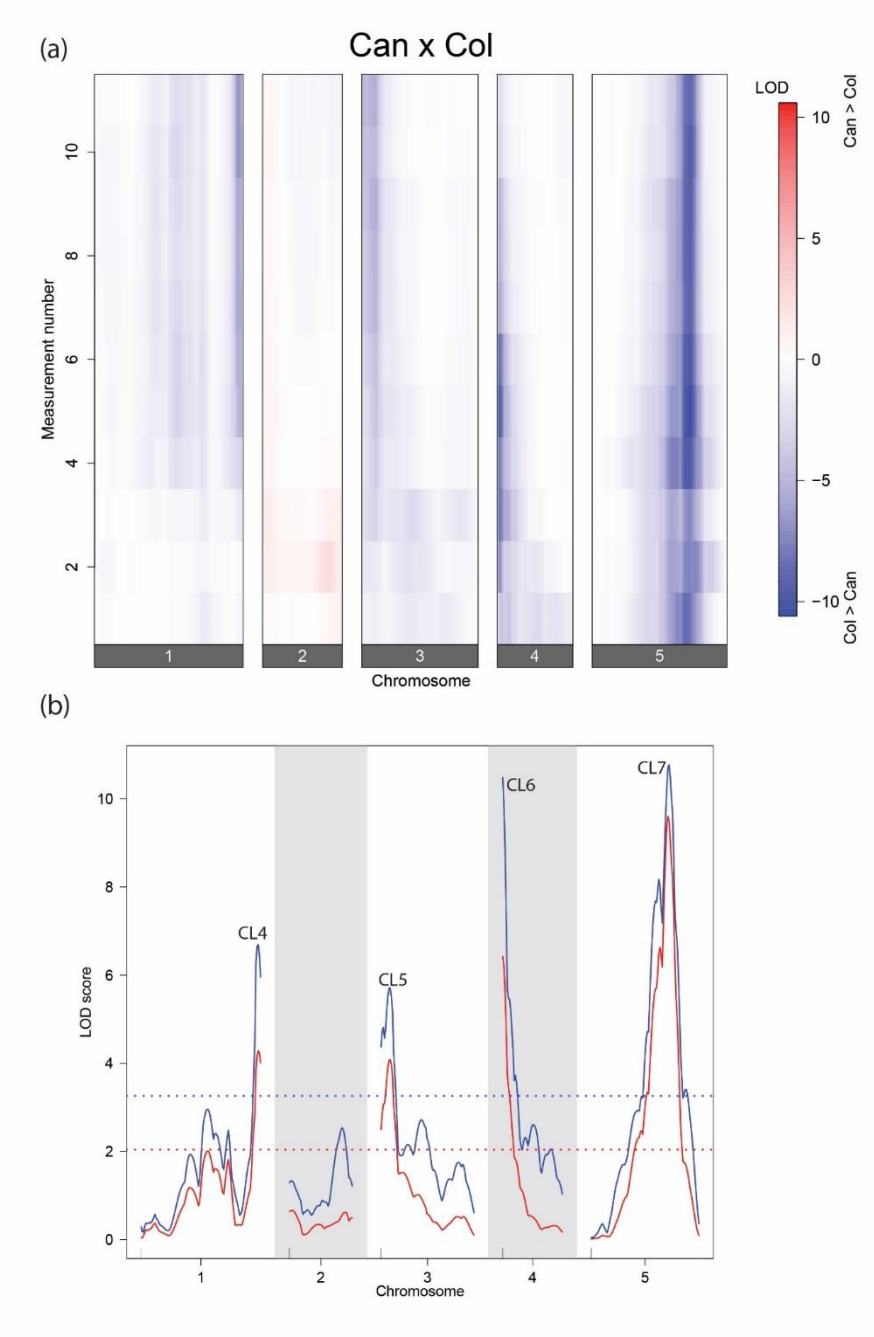


Figure 4.12. *QTL analysis of chlorophyll reflectance (CHL) for the Can x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.12). Colours indicate the direction of the QTL effect with red indicating the Can-0 allele increases CHL and blue indicating Col-0 increases CHL, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for CHL across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

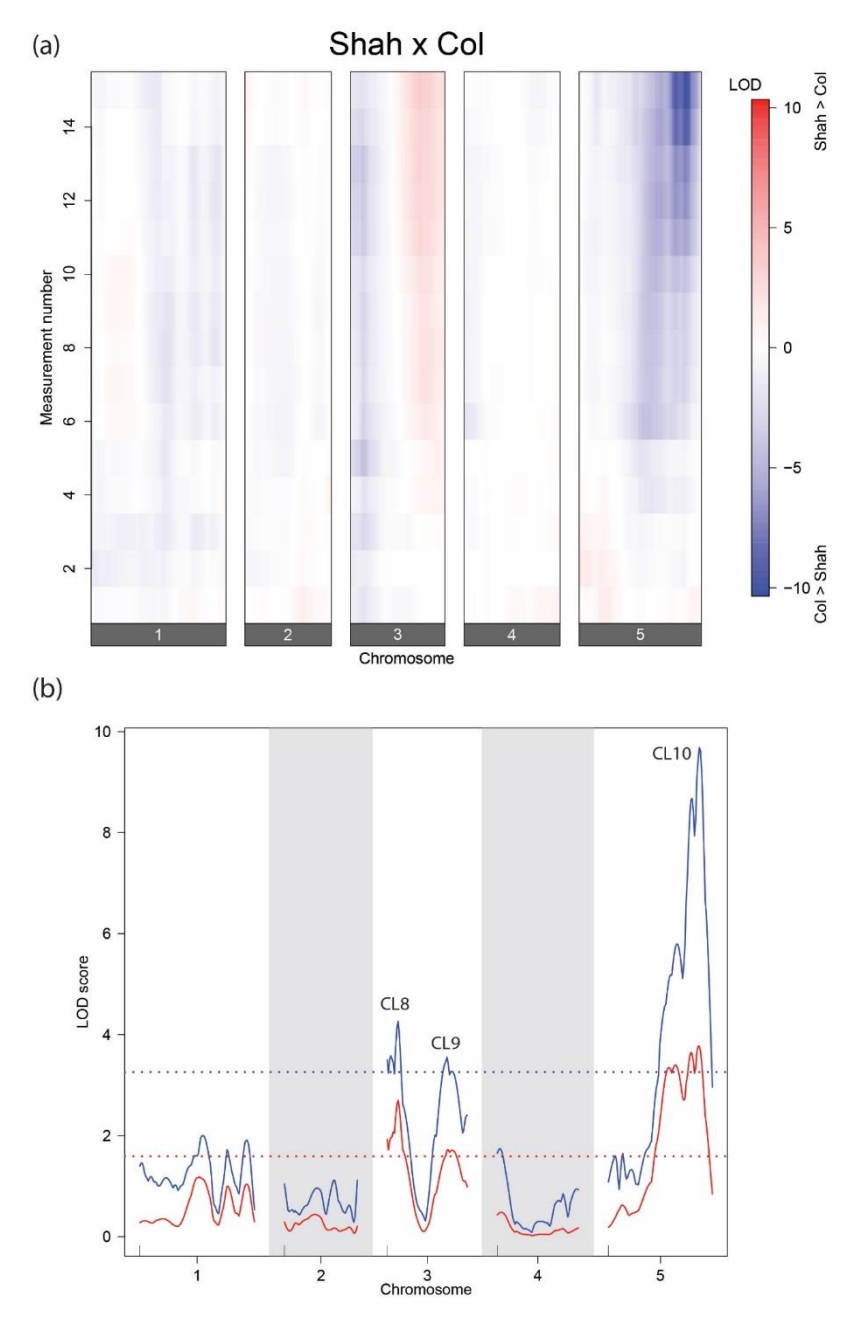


Figure 4.13. *QTL analysis of chlorophyll reflectance (CHL) for the Shah x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.13). Colours indicate the direction of the QTL effect with red indicating the Shah allele increases CHL and blue indicating Col-0 increases CHL, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for CHL across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

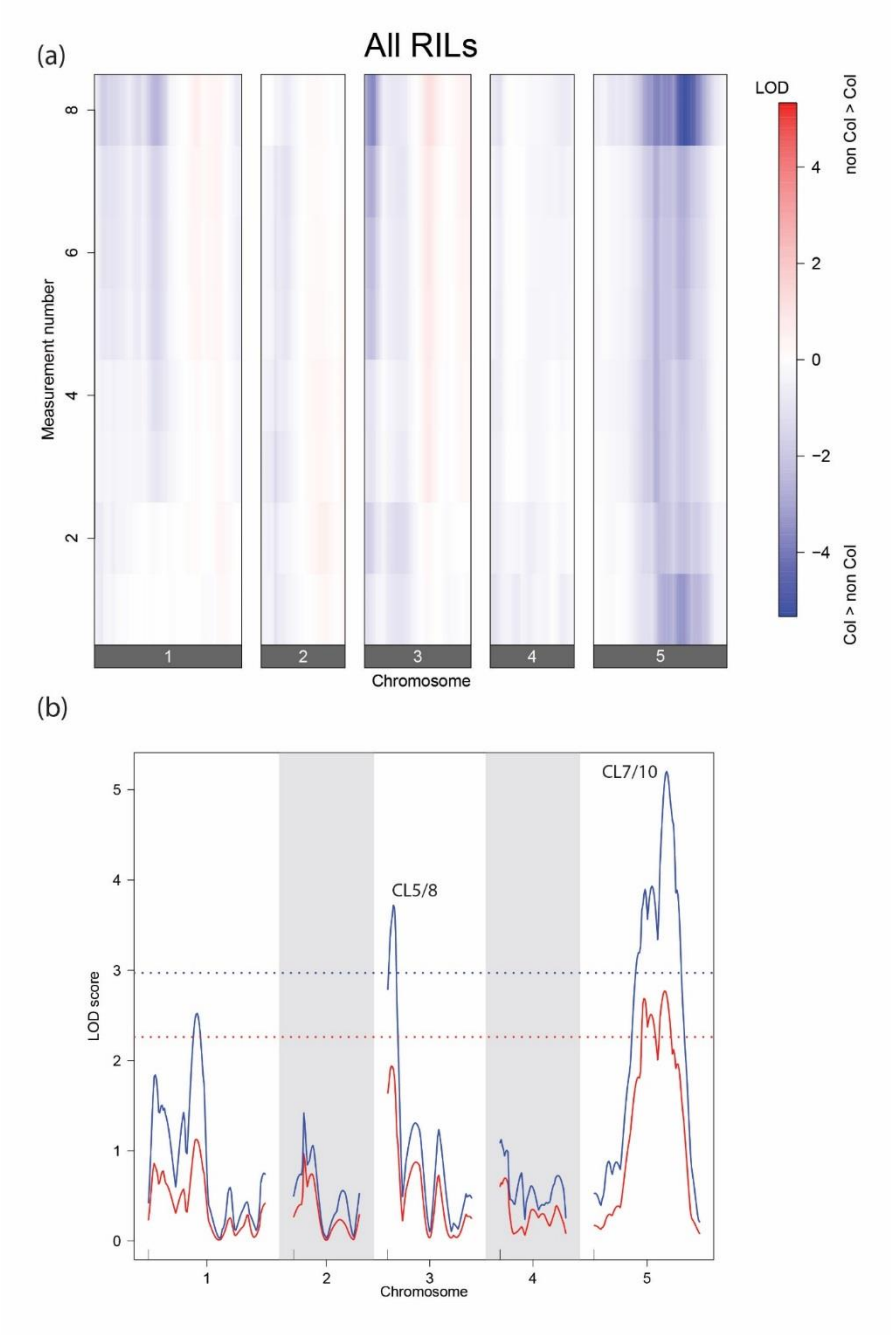


Figure 4.14. *Joint QTL analysis of chlorophyll reflectance (CHL) for all RIL populations combined, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.14). Colours indicate the direction of the QTL effect with red indicating the non Col-0 allele increases CHL and blue indicating Col-0 increases CHL, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for CHL across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

Percentage change in PLA every three hours, which relates to daily leaf movements, was the most dynamic trait in terms of the QTL LOD profiles detected (Figures 4.15 to 4.18). Many loci are significant resulting in very broad peaks and it becomes difficult to decide which loci represent separate QTLs (Figure 4.10)[20]. As a result the QTLs are not individually named and a more general description of the results is given. A QTL region on the upper portion of chromosome 1 is present in all populations and shows a very dynamic fluctuation in the allelic effect of the different parents, illustrated by the diurnal blue and red banding in Figures 4.15 to 4.18. A general trend is that the non Col allele increases trait values for a longer period of time (see the wider red bands) with all detected loci showing the same pattern.

The time-points at which the main QTL for Φ_{PSII} in each population were most significant were further analysed using multiple QTL mapping (MQM) to increase the power to disentangle linked loci (Jansen, 1993; Arends et al., 2010). These time-points were also scanned for genetic interactions using a two locus model (scantwo). Supplementary Figure 4.4 shows the results for the Bur x Col population 22DAS, panel (a) shows that there is little difference between the results of MQM and the single QTL model (scanone). In panel (b) a significant additive interaction was detected between QTL PS1 and the end of chromosome 5. There are also appears to be some epistasis between the top of chromosome 4 and the tops of chromosomes 3 and 5 and the bottom of chromosome 2 (Supplementary Figure 4.4b). Supplementary Figure 4.5 shows the results of Φ_{PSII} on 16DAS, again the scanone and MQM analysis detect the same loci, although the significance of QTL PS2 is greatly increased under the MQM model (panel a). PS2 shows significant additive interactions with QTLs PS3 and PS4 (panel b), there is also signs of epistasis between the proximal regions of chromosomes 1 and 3. In Supplementary Figure 4.6 the Shah x Col population again shows no new loci with the use of MQM but the significance of QTLs PS5, PS6 and PS7 are increased (panel a). All three QTLs detected show additive interactions although the interaction between PS5 and PS6 is most significant (panel b). PS7 appears to be in epistasis with the middle of chromosome 3.

[20] Unfortunately due to technical errors some time-points were missing in the Can x Col RIL reducing the temporal resolution considerably.

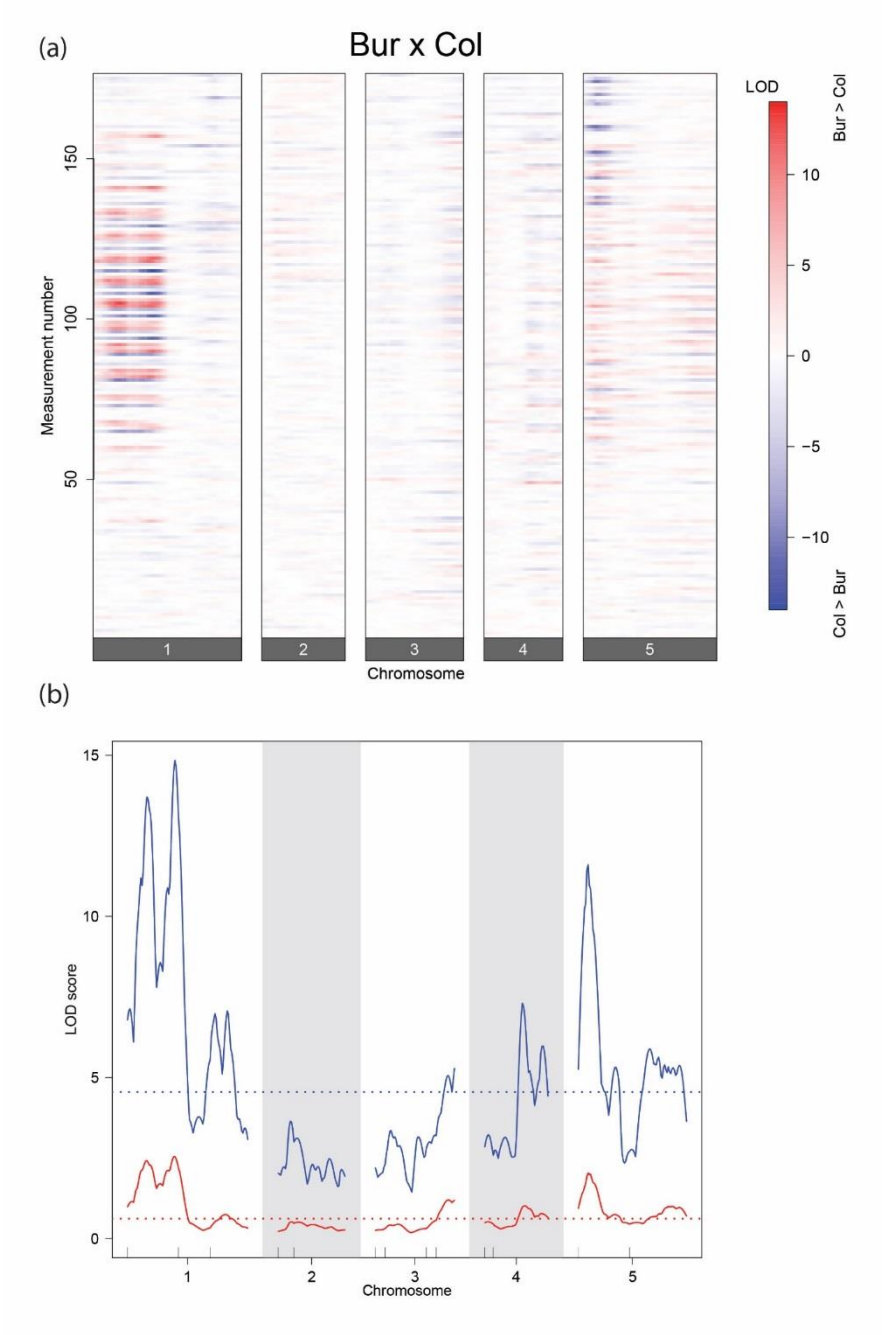


Figure 4.15. *QTL analysis of percentage change in PLA every 3hrs for the Bur x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.15). Colours indicate the direction of the QTL effect with red indicating the Bur-0 allele increases and blue indicating Col-0 increases, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for percent change in PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

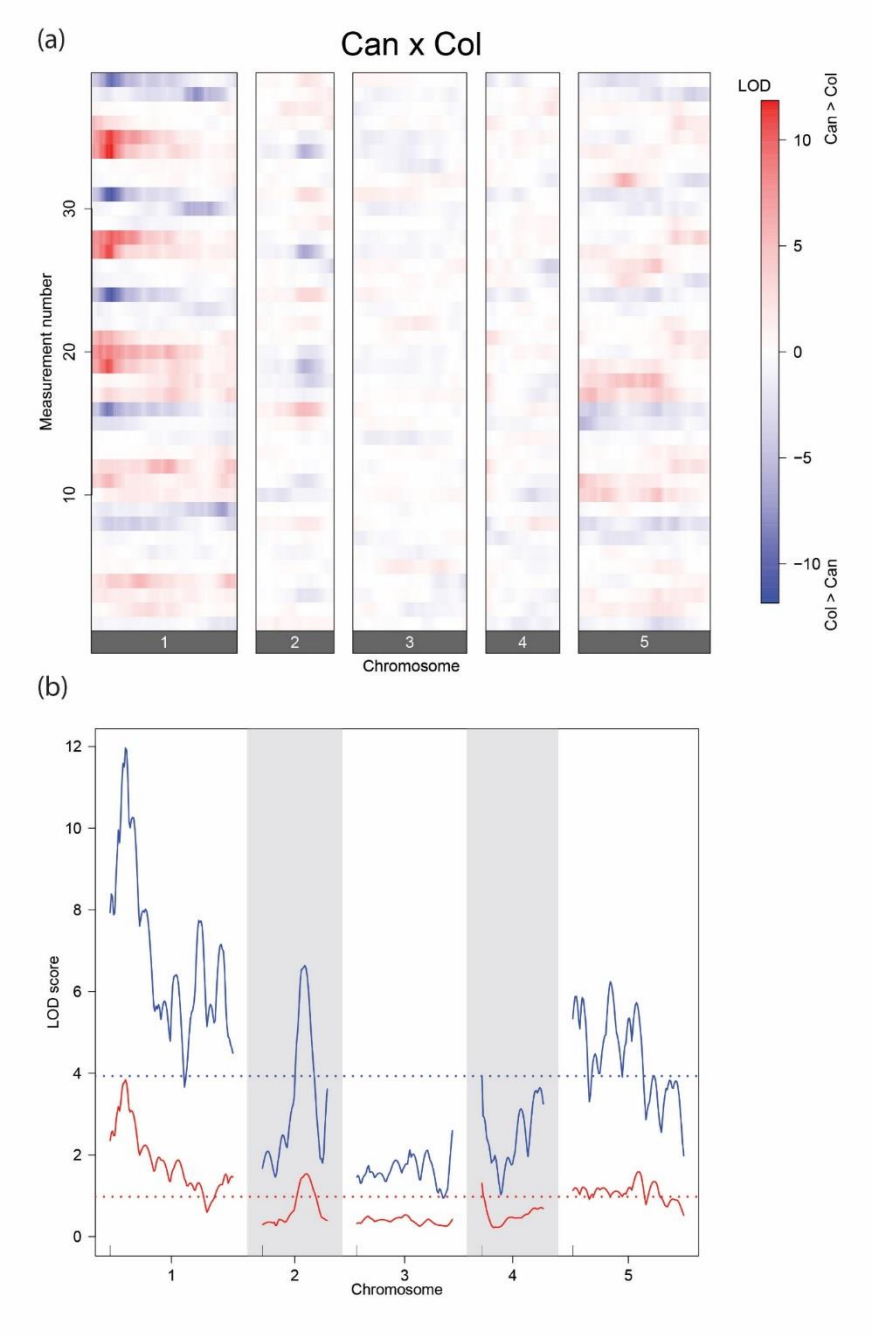


Figure 4.16. *QTL analysis of percentage change in PLA every 3hrs for the Can x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.16). Colours indicate the direction of the QTL effect with red indicating the Can-0 allele increases and blue indicating Col-0 increases, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for percent change in PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

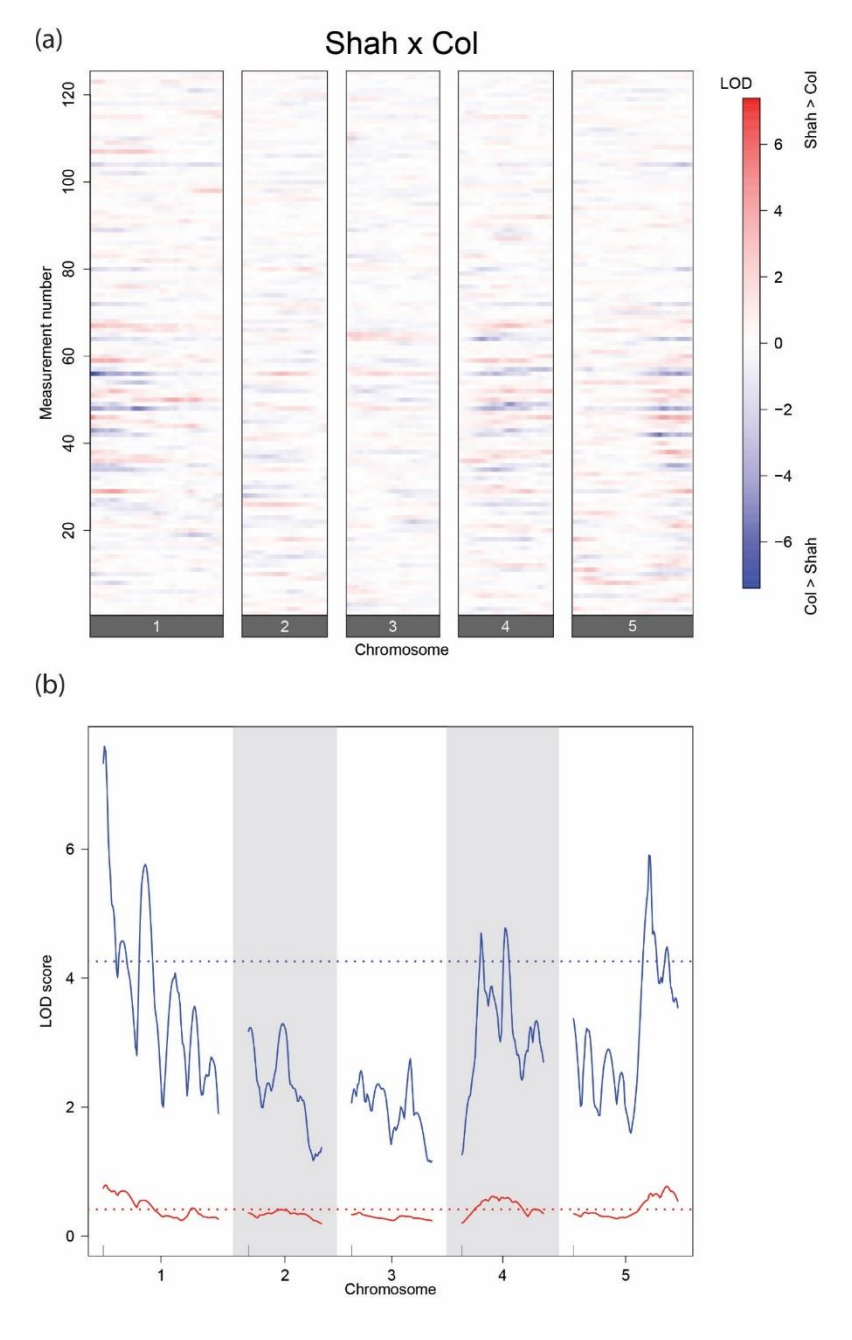


Figure 4.17. *QTL analysis of percentage change in PLA every 3hrs for the Shah x Col RIL population, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.17). Colours indicate the direction of the QTL effect with red indicating the Shah allele increases and blue indicating Col-0 increases, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for percent change in PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

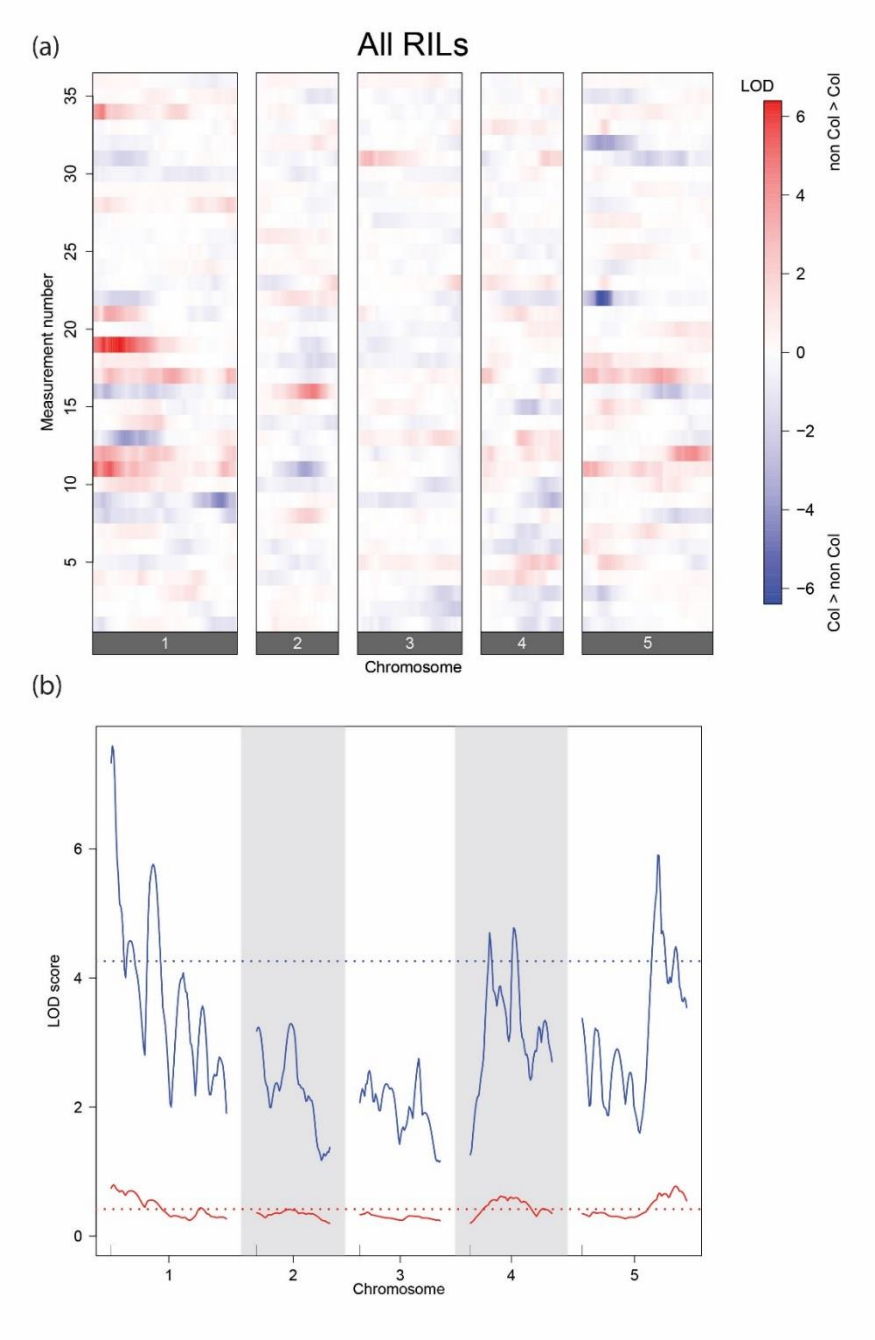


Figure 4.18. *Joint QTL analysis of percentage change in PLA every 3hrs for all RIL populations combined, (a) LOD scores through time, the genetic position is indicated by the x-axis and time is indicated on the y axis, this is listed as measurement number as the plotting software cannot handle missing time points (for labelled time-points please see interactive Figure 4.18). Colours indicate the direction of the QTL effect with red indicating the non Col-0 allele increases and blue indicating Col-0 increases, intensity represents LOD score as shown by the scale bar to the right of the figure. (b) Functional mapping results, the maximum LOD (MLOD, blue line) and the average LOD (SLOD, red line) for percent change in PLA across all time-points. Significance threshold of 0.05 based on 1000 permutations is indicated by the dashed horizontal lines (red for SLOD and blue for MLOD).*

Discussion

The results presented here show that the function valued traits analysed, Φ_{PSII}, PLA, chlorophyll reflectance and percentage change in PLA every three hours, all exhibit dynamic QTL profiles with respect to time. This illustrates how important phenotyping at multiple time-points is as single time-point analysis is likely to only identify a snapshot of the variation present.

The most significant QTL identified for Φ_{PSII} was PS2 located at the top of chromosome 4 in the Can x Col population, this locus shows a convex effect profile with the largest phenotypic impact around day 16 and a decreasing effect thereafter (Figure 4.4 and interactive Figure 4.4). Photosynthetic capacity is known to vary during the lifetime of both the plant and the leaf (Reich et al., 1991; Flood et al., 2011). Based on Figures 3 to 6 it is clear that this temporal variation in Φ_{PSII} (Figure 4.2) is underlain by different genetic loci at different time-points, indicating that the genetic variation controlling this process changes through time. The Can-0 allele of PS2 reduces Φ_{PSII} by 0.007 explaining about 45% of the additive genetic variance in this population. All other loci identified have smaller effect sizes with PS4/7, PS5 and PS6 having a maximum effect on Φ_{PSII} of 0.004, and PS1 and PS3 having an effect size of 0.002. Thus with the exception of the Can x Col population the loci identified explained a relatively small amount of the phenotypic variation observed in the RIL populations (Figure 4.2). This suggests that the remaining loci are numerous and of small effect, thus being below our detection threshold, and/or in epistatic interaction with one another (Rockman, 2012; Mackay, 2014). Supplemental Figures 4.4 to 4.6 show that there are indeed additive and epistatic interactions. The loci identified tend to be unique to one of the parents (Figure 4.6) suggesting that the alleles of medium to large effect are likely to be rare and that common alleles are either of small effect or in epistatic interactions or both. The size of the populations screened and the relatively small number of markers means it is unlikely that there is sufficient power or resolution to unmask physically linked epistatic interactions (Carlborg et al., 2006).

In the Can x Col RIL population the PS2, PS3, and PS4 QTLs for Φ_{PSII} collocate with the CL6, CL5, and CL7 QTLs for chlorophyll reflectance (Figures 4.4 and 4.12). To investigate this relationship a Pearson's product-moment correlation was performed on the BLUEs for both traits on day 16. The traits were found to be highly correlated ($r=0.68$, $p < 2.2 \times 10^{-16}$) (Supplementary Figure 4.7). This indicates that the causal loci for the two traits may be common in the Can x Col RIL population. Additionally CL7 collocates with CL10 Sha x Col population which also collocates with a Φ_{PSII} QTL (PS7) in this population (Figures 4.5 and 4.13). CL5

also collocates with CL8 but this locus shows no significant QTL for Φ_{PSII} in the Shah x Col population (Figure 4.5). The collocation of several Φ_{PSII} and chlorophyll reflectance QTLs in two of the three populations analysed indicates that there may be a common genetic origin. Chlorophyll and photosynthesis are known to be physiologically linked in that photosynthesis operates based on the light harvesting capacity of chlorophyll. In nature it has been shown that chlorophyll concentration per leaf area correlates positively with photosynthetic capacity (Murchie and Horton, 1997) a correlation which is also found in the growth chamber in this study (Supplemental Figures 4.7 and 4.8). This could be due to enhanced light absorption improving carbon fixation under the relatively low (200 $\mu molm^{-2}s^{-1}$ in the growth chamber compared with as much as 2000 $\mu molm^{-2}s^{-1}$ in the field) light levels that the plants were grown at. Can-0 is paler green and it has a lower Φ_{PSII} than the other three accessions (Supplementary Figure 4.1) Can-0 originates from the Canary Islands, the most southerly location sampled in this study, and thus may experience higher irradiances in its native habitat. This could ameliorate any reduced light harvesting capacity and may even represent a constitutive form of photoprotection, reducing light harvesting to limit photoinhibition (Ort et al., 2011).

The most significant QTL for PLA was PA1 located at the top of chromosome 5 in the Bur x Col RIL population (Figure 4.7). This locus was previously identified in this population (Vlad et al., 2010) and has been recently identified as a de novo mutation in the *SHOOT GROWTH 1 (SG1)* gene which occurred in the Col-0 parental line used for the initial cross (Coustham et al., 2014). This mutation results in hypermethylation of the genome and leads to complete degeneration of the plants after several generations. This epigenetic meltdown leads to poorer plant growth which amplifies during development and is reflected in the increasing significance of PA1 through time (Figure 4.7a). This QTL collocates with CL1 (Figures 4.7 and 4.11) which may confirm the observation of the authors of this study that the *sg1-1* mutants were also darker green (Coustham et al., 2014). In the case of PLA the Col-0 allele reduces plant size and in the case of CHL the Col-0 allele increases chlorophyll reflectance, which is the expected direction of effect of the *sg1-1* mutation (Figures 4.7 and 4.11). PLA QTL PA4/6 also collocates with PS4/7, and CL7/10, for these loci the Col-0 allele increases Φ_{PSII} and chlorophyll reflectance and decreases PLA (Figures 4.4d, 4.6d and 4.8d). This result combined with the *SG1* locus may explain the negative correlation of chlorophyll reflectance and PLA (Supplementary Figure 4.8) which is the opposite of what is reported in the literature (Brougham, 1960; Dodd et al., 2005).

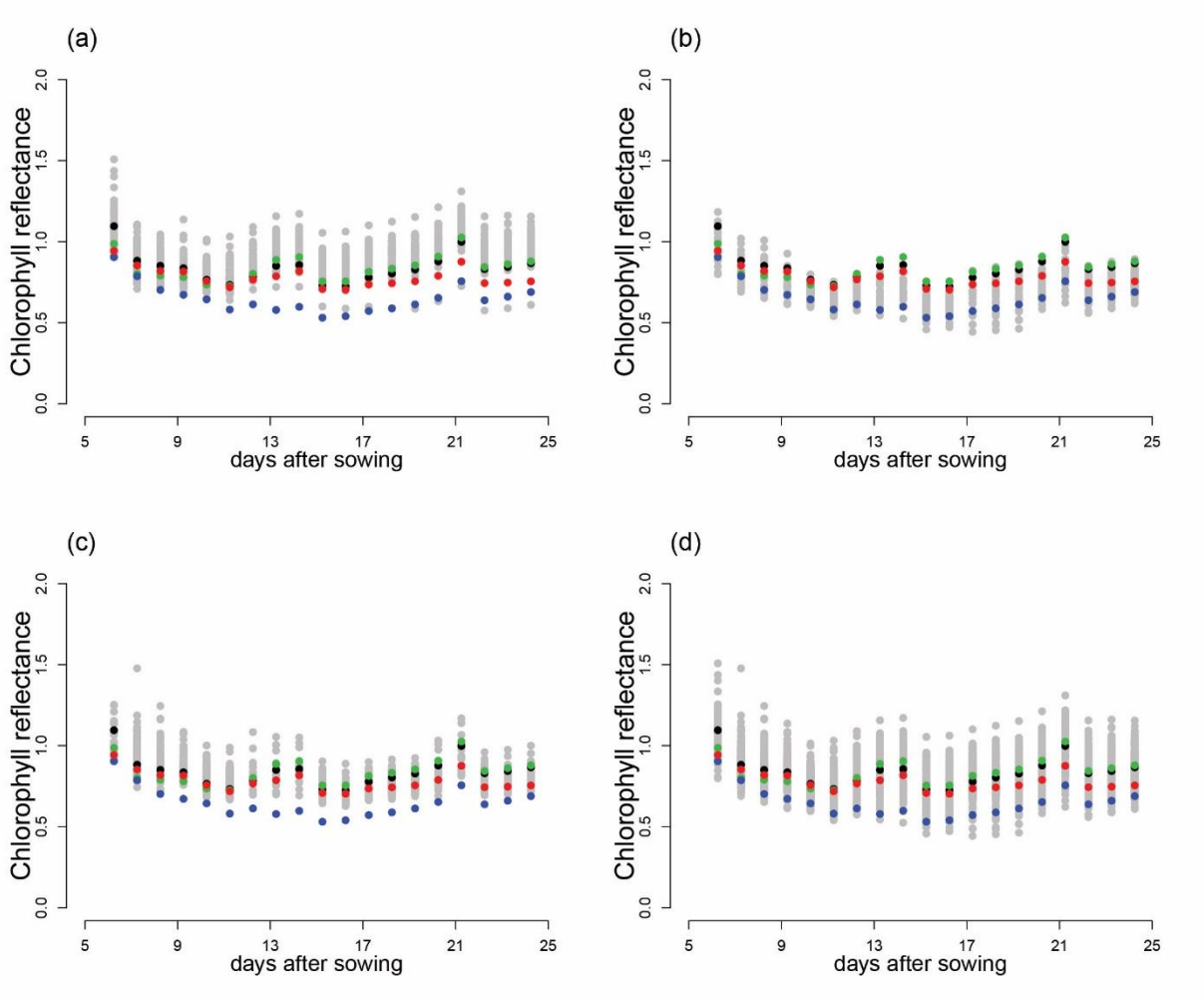
The final phenotype to be assessed was percentage change in PLA estimated from pixel masks generated using near infra-red (NIR) wavelengths. The NIR approach allowed measurements to be taken in the dark without perturbing the plants photoperiod (see Chapter 3). The percentage change in PLA was exhibited large diurnal fluctuations in heritability in Chapter 3, Figures 4.15 to 4.18 show some of the loci responsible for this fluctuation in heritability. There are striking alterations between the ranking of effect of alleles on chromosome one (note the blue and red banding in Figures 4.15 to 4.18). These plots are best assessed in the digital interactive plots where a much more detailed look at effects through time can be seen (interactive Figures 4.15 to 4.18). The very dynamic daily fluctuation in LOD scores is also overlain by a longer term fluctuations of up to 10 days where a locus at the top of chromosome 5 becomes increasingly significant through time. This collocates with PA1 and CL1 and may be a pleiotropic effect of the *sg1-1* mutation (Figures 4.7, 4.11, and 4.15) (Coustham et al., 2014). "Percentage change in PLA" QTLs are also identified at the top of chromosome 1 in the Can x Col and to a lesser extent in the Shah x Col population (Figure 4.16 and 4.17) which may indicate a common genetic basis.

Across all traits the joint QTL analysis only resulted in the detection of one new QTL for PLA (QTL PA8 in Figure 4.10). The QTLs which were found in both the individual populations and the combined analysis were not detected with greater significance or precision, thus for the traits and populations analysed here the joint QTL analysis was of little added value. This poor performance of the joint QTL approach could be due to a variety of reasons. It may be that the Bur-0, Can-0, and Shah are too divergent relative to one another and thus share little allelic overlap when compared to Col-0. It may also be that the common loci are of small effect size and are thus below detection threshold, even with the added power of the combined analysis. If this is the case expanding the number of individuals in a single RIL population to increase power (Keurentjes et al., 2007) as well as increasing the number of populations to increase the chance of some allelic overlap, is likely to result in an increase in the ability of the joint QTL analysis to add new insights. So far joint QTL analysis has been used to greatest effect in maize, where 4899 RILs were derived from 25 accessions crossed to one reference genotype (McMullen et al., 2009). It is likely that the success of this particular approach in Maize may indeed be due to its increase in population size (contrast 490 RILs derived from three accessions crossed to a reference genotype). In this case over a million genetic markers were known which meant that the researchers could perform a variation of the joint QTL approach called nested association mapping (NAM), which further improves the method by

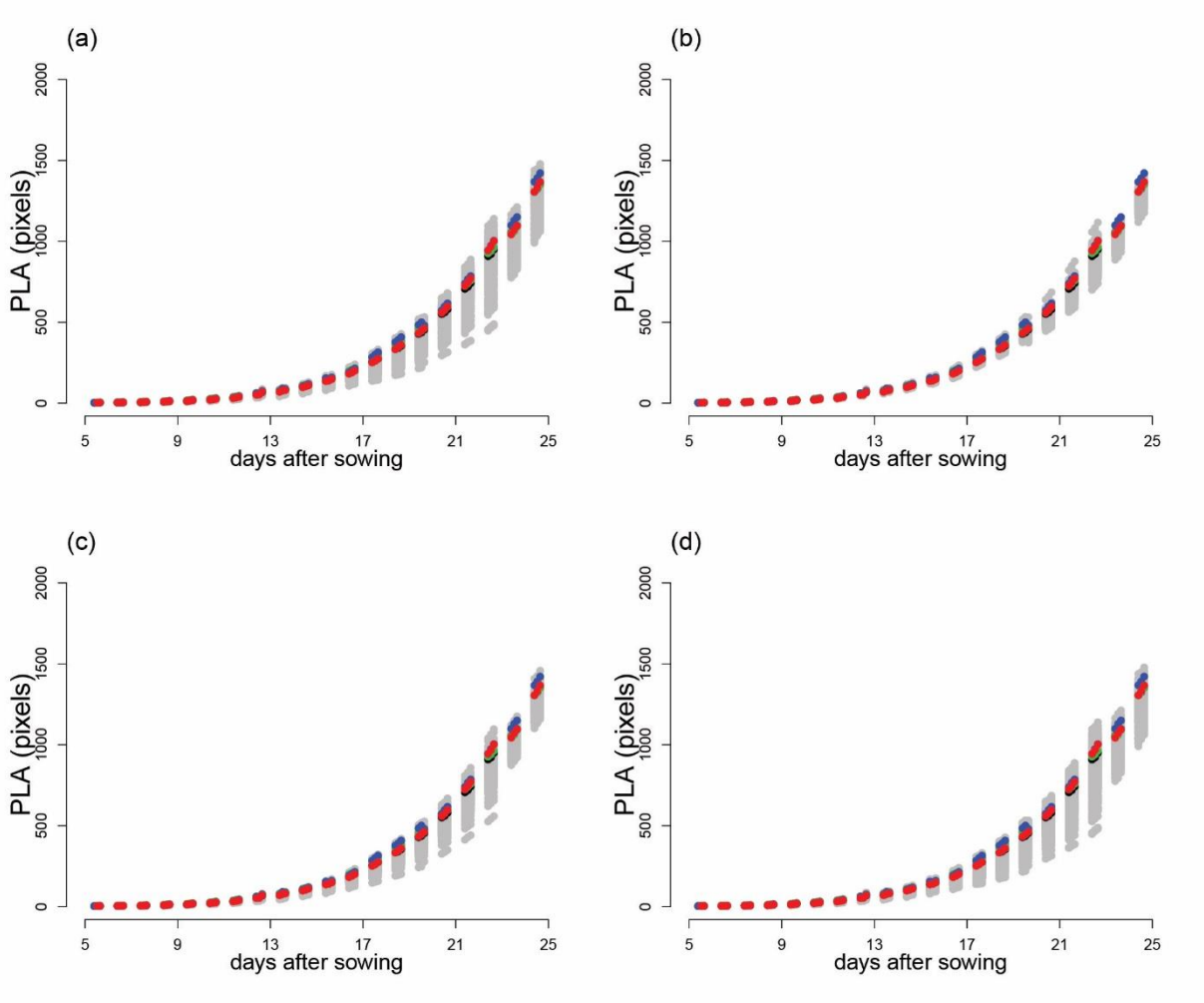
allowing the researcher to make use of historical recombination events (Buckler et al., 2009; Tian et al., 2011). In this case they found that indeed many of the loci where of very small effect sizes and the traits analysed, flowering time and leaf architecture, were characterised as highly polygenic (Buckler et al., 2009; Tian et al., 2011). Joint QTL analysis has been used successfully in Arabidopsis in the past (Bentsink et al., 2010; Li et al., 2011) although in neither study did the researchers detect new loci in the combined analysis. The analysis was considered a success due to loci being found in multiple populations therefore providing additional evidence that they are significant. Apart from the difference in the size of the populations analysed, the species studied, maize and Arabidopsis, have very different breeding systems. With maize being largely outbreeding and Arabidopsis mostly inbreeding. This difference is thought to have a profound impact on genome evolution (Charlesworth and Wright, 2001) and may explain the differences in both the level of allelic overlap and the architecture of the traits studied (Buckler et al., 2009). Outbreeding species will experience greater gene flow and reduced linkage disequilibrium, whilst inbreeding species are more likely to accumulate deleterious mutations due to the reduced efficiency of selection (Hough et al., 2013).

A final point to consider is that since the QTL responsible for trait variation show a time dependency, the moment of selection is likely to have a very large effect on the alleles selected. This has implications for both agriculture and evolutionary biology. In the case of breeding programs detailed knowledge of the key stages in plant performance can be used to select optimal QTLs at these time-points. However the ranking of the selected QTLs should be assessed at different stages in the plants life cycle in order to rule out potential antagonistic pleiotropy.

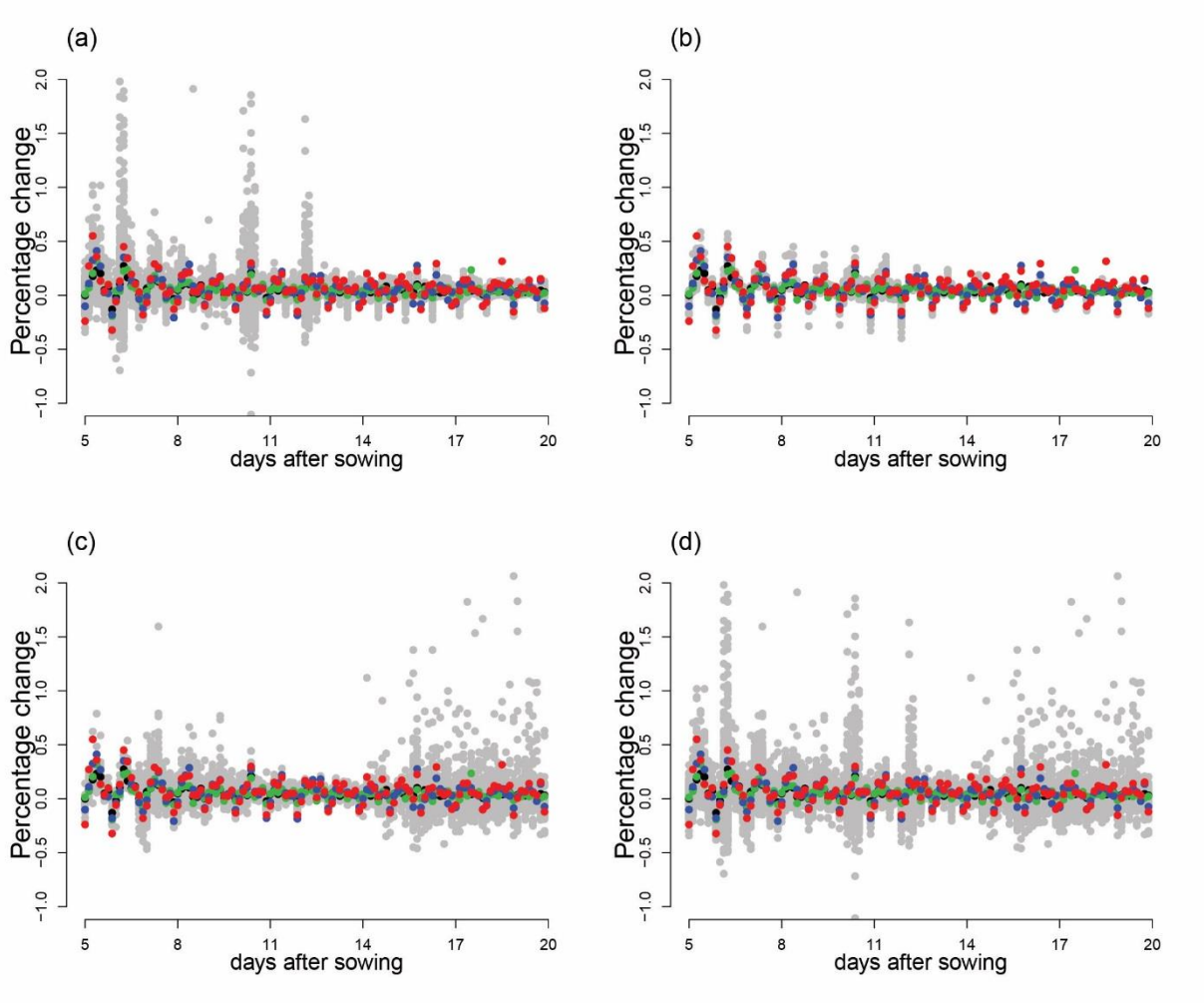
Supplemental material



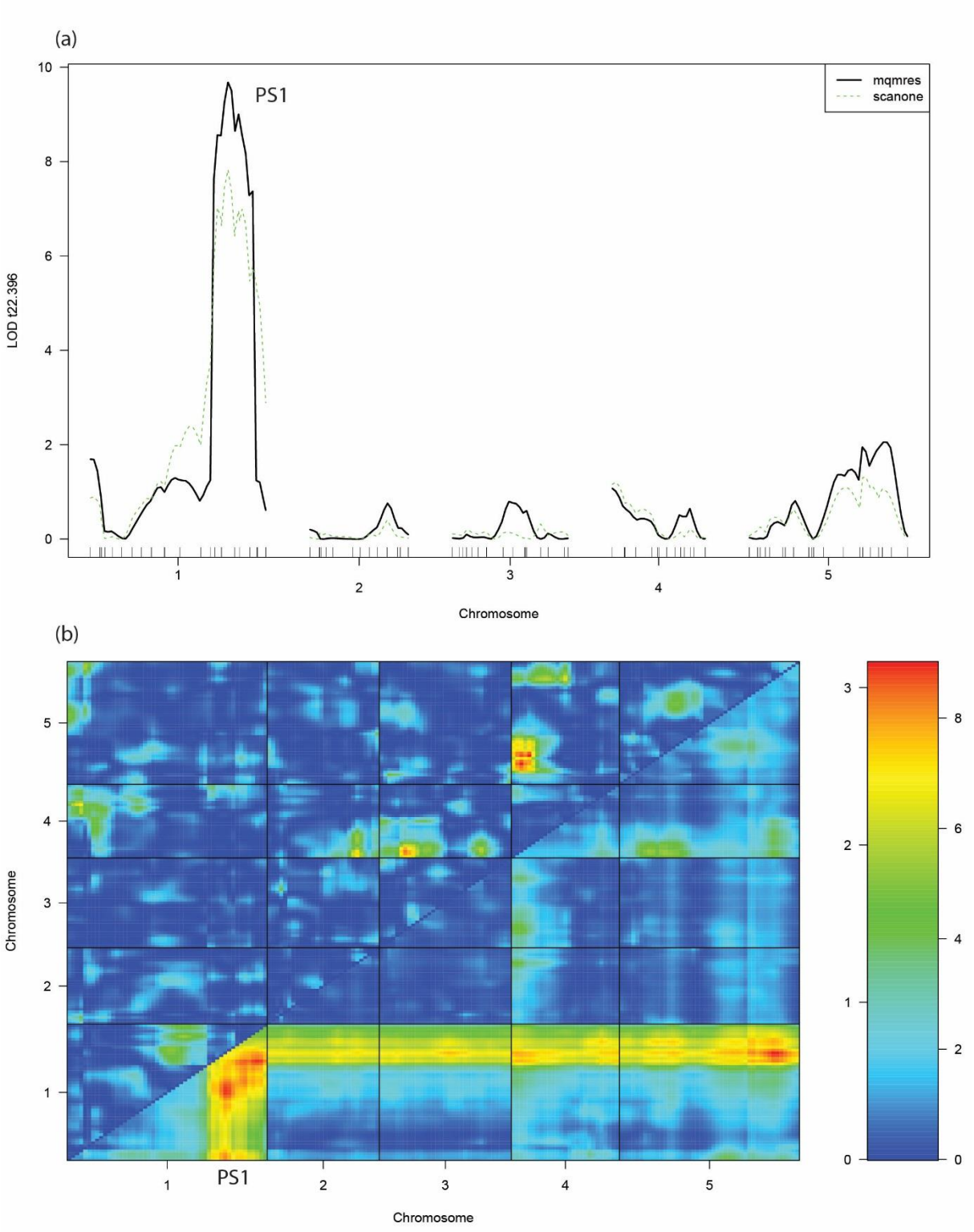
Supplementary Figure 4.1. *Best linear unbiased estimators of Chlorophyll reflectance. The Bur x Col population (a), the Can x Col population (b), the Shah x Col population (c), and all RILs combined (d). The four parental genotypes are included in each plot to facilitate comparison, Col-0 = black, Bur-0 =green, Can-0=blue and Sha=red.*



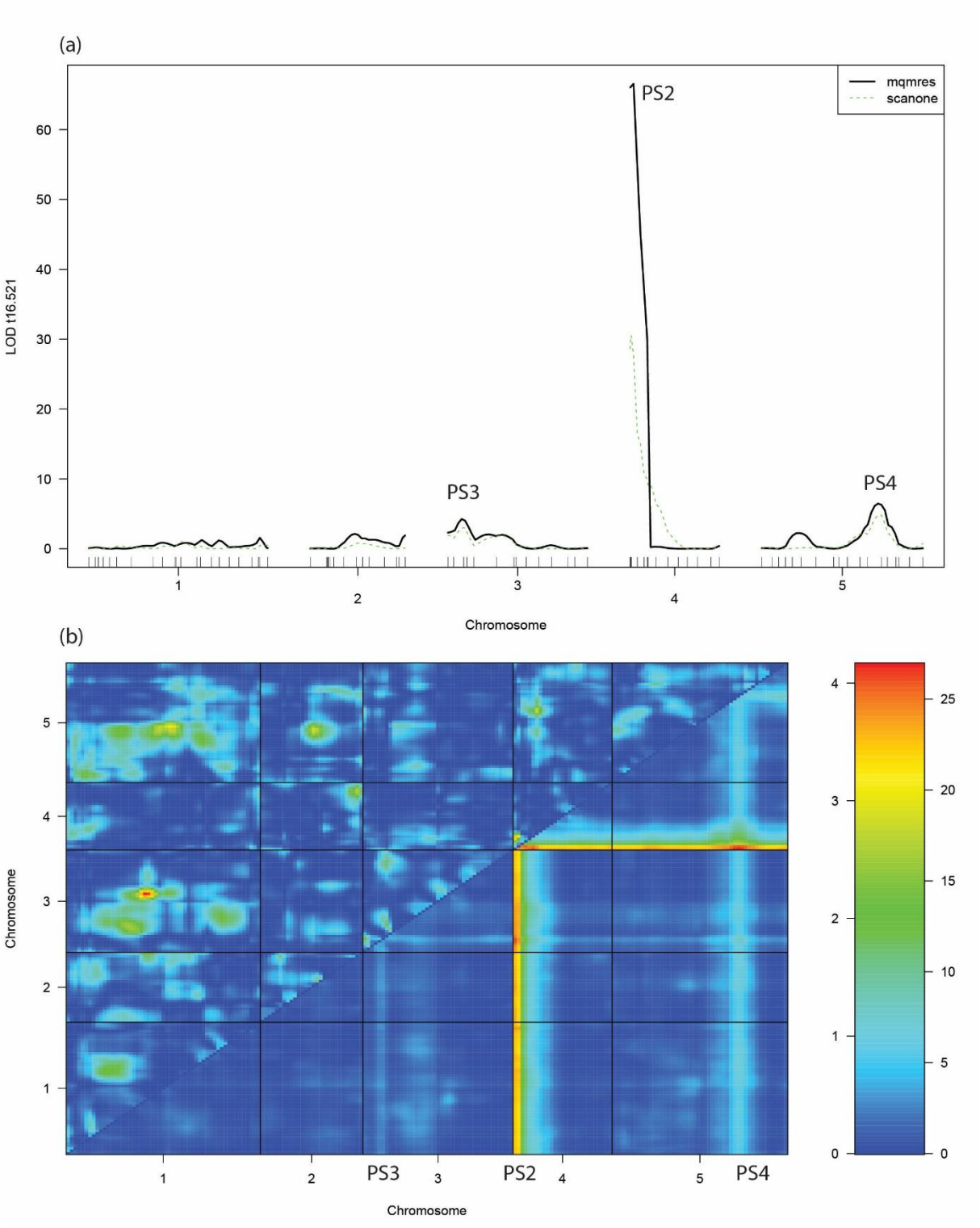
Supplementary Figure 4.2. *Best linear unbiased estimators of projected leaf area (PLA). The Bur x Col population (a), the Can x Col population (b), the Shah x Col population (c), and all RILs combined (d). The four parental genotypes are included in each plot to facilitate comparison, Col-0 = black, Bur-0 =green, Can-0=blue and Sha=red.*



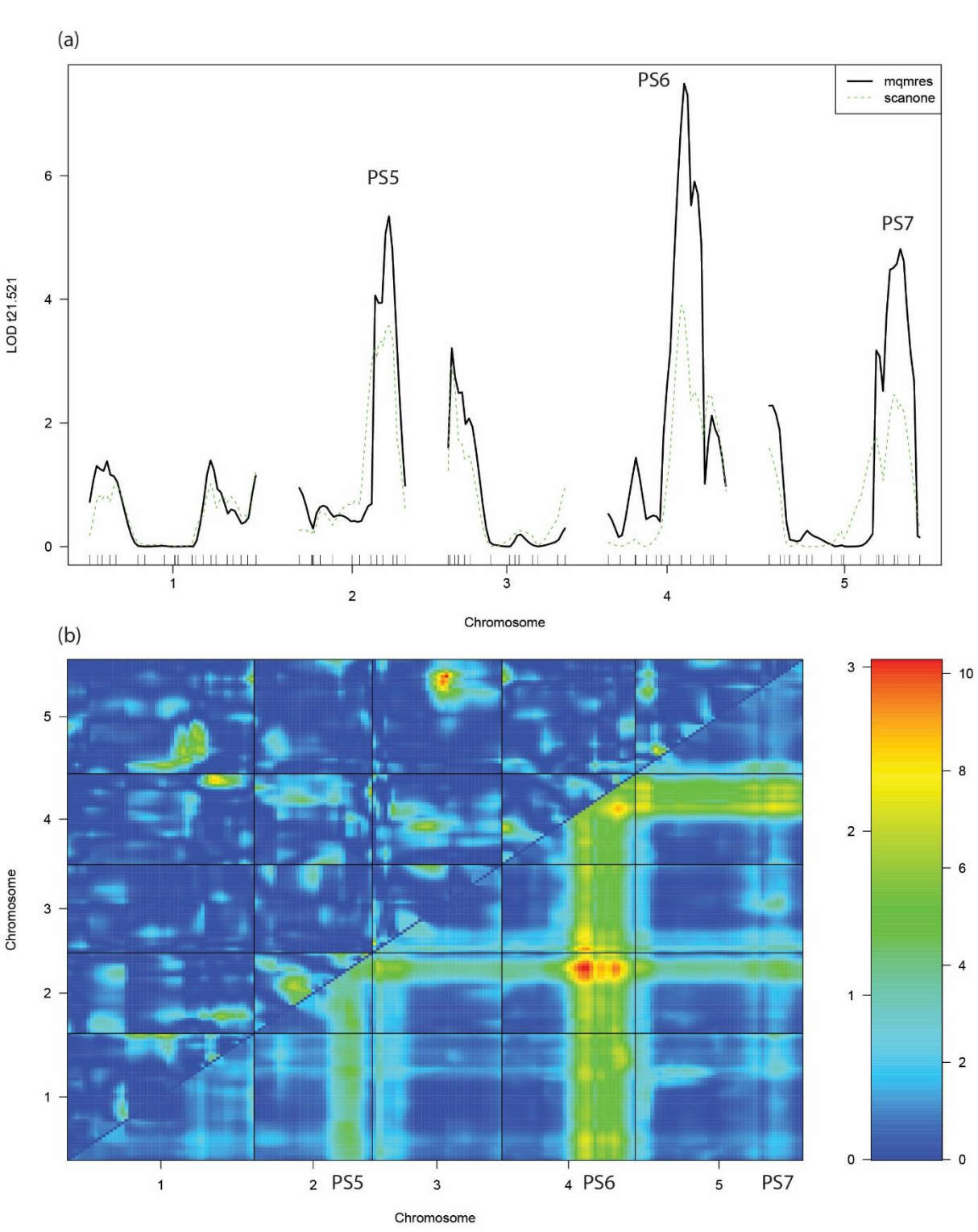
Supplementary Figure 4.3. *Best linear unbiased estimators of percentage change in PLA every three hours. The Bur x Col population (a), the Can x Col population (b), the Shah x Col population (c), and all RILs combined (d). The four parental genotypes are included in each plot to facilitate comparison, Col-0 = black, Bur-0 =green, Can-0=blue and Sha=red.*



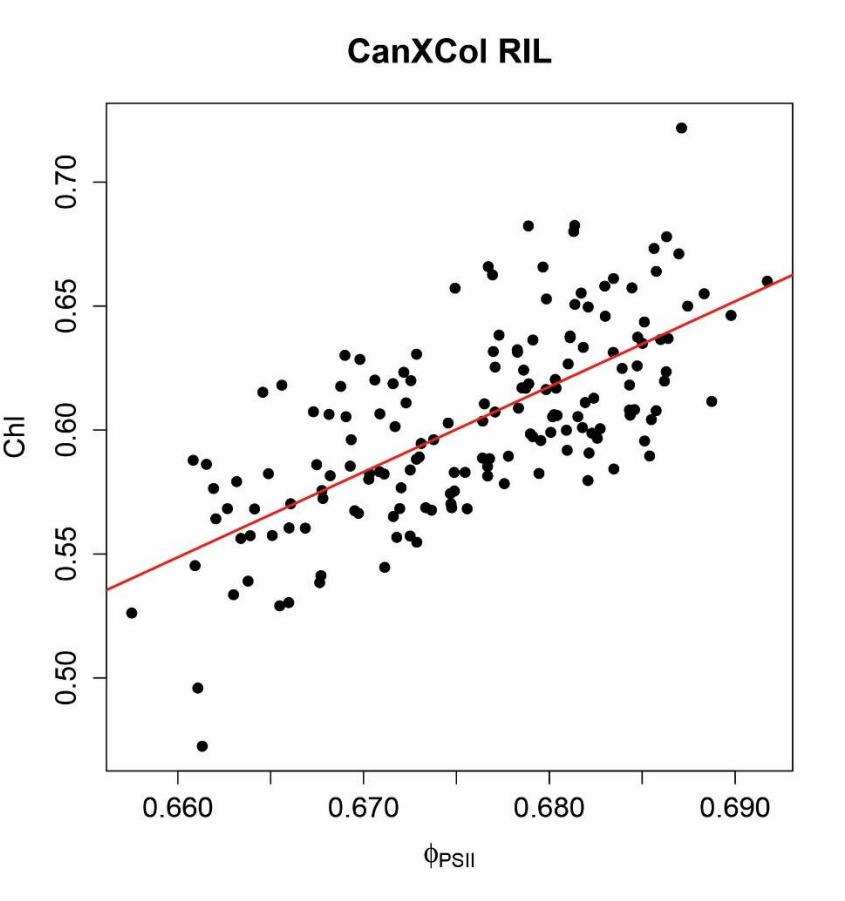
Supplementary Figure 4.4. *Φ_{PSII} in the Bur x Col population 22 days after sowing. (a) Comparison of single QTL (scanone,) green dotted line, and multiple QTL mapping (MQM), solid black line. (b) Results of a two locus scan (scantwo), upper triangle shows epistasis and the bottom triangle shows additive interactions, colour bar indicates LOD scores for epistasis (right) and additive interactions (left).*



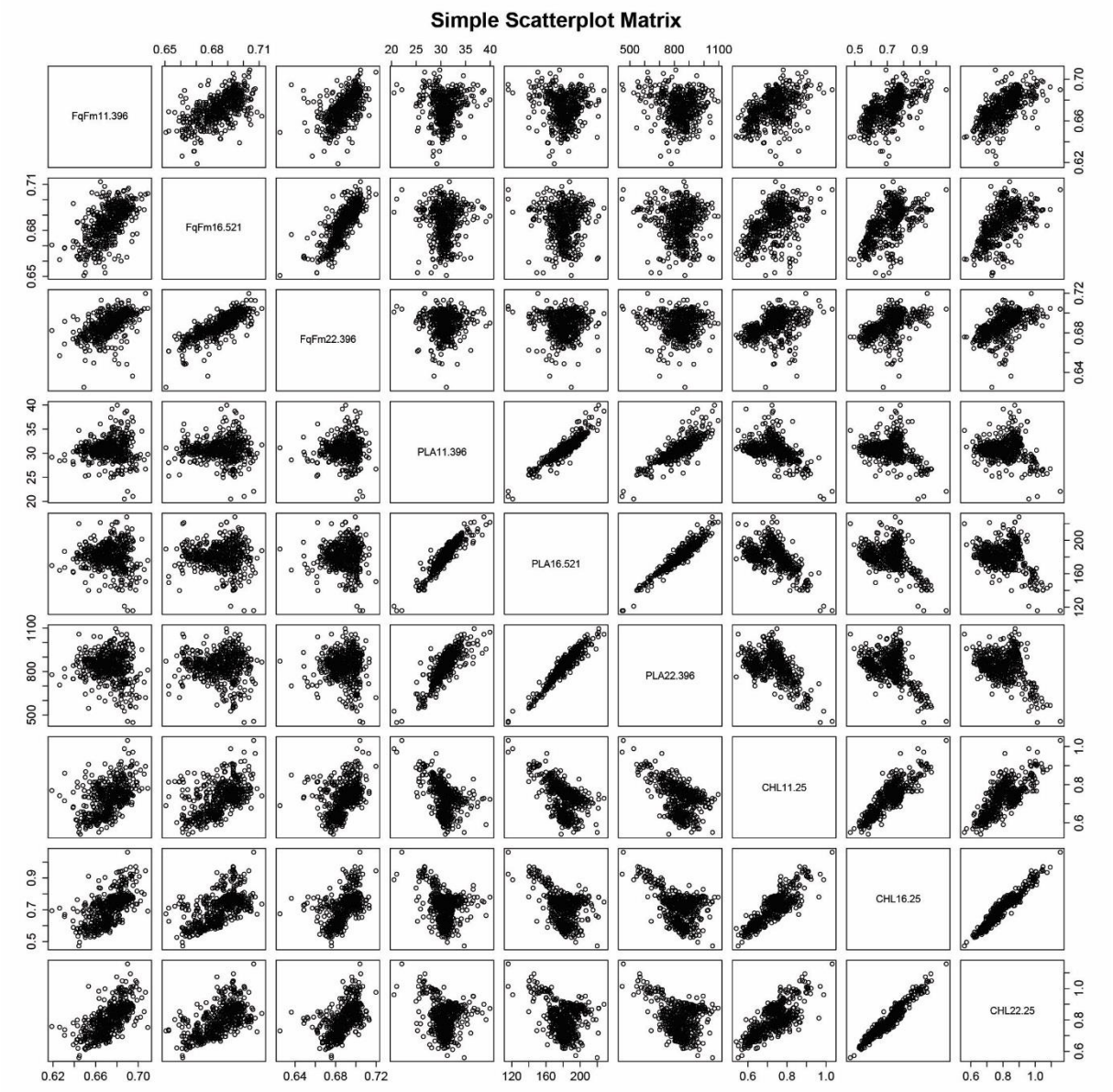
Supplementary Figure 4.5. *Φ_{PSII} in the Can x Col population 16 days after sowing. (a) Comparison of scanone (green dotted line) and multiple QTL mapping (MQM) results. (b) Results of a scantwo analysis, upper triangle shows epistasis and the bottom triangle shows additive interactions, colour bar indicates LOD scores for epistasis (right) and additive interactions (left).*



Supplementary Figure 4.6. *Φ_{PSII} in the Shah x Col population 21 days after sowing. (a) Comparison of scanone (green dotted line) and multiple QTL mapping (MQM) results. (b) Results of a scantwo analysis, upper triangle shows epistasis and the bottom triangle shows additive interactions, colour bar indicates LOD scores for epistasis (right) and additive interactions (left).*



Supplementary Figure 4.7. *Pearson's product-moment correlation in the Can x Col RIL population between* Φ_{PSII} *and Chl 16 days after sowing,* r =0.68, $p < 2.2 \times 10^{-16}$.



Supplementary Figure 4.8. *Matrix of scatterplots of Φ_{PSII} (FqFm11.396, FqFm16.521 and FqFm22.396), PLA (PLA11.396, PLA16.521 and PLA22.396) and chlorophyll reflectance (CHL11.396, CHL16.521 and CHL22.396), at three time-points per trait 11, 16 and 22 DAS.*

Chapter 5

Genome wide association mapping for photosynthesis in *Arabidopsis thaliana*

Pádraic J. Flood[1,2], Willem Kruijer[3], Sabine Schnabel[3], Daniel Oberste-Lehn[1,2], Frank Becker[1], Jeremy Harbinson[2] & Mark G.M. Aarts[1]

[1]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

[2]*Horticulture and Production Physiology, Wageningen University, Wageningen, the Netherlands.*

[3]*Biometris, Wageningen University, Wageningen, the Netherlands.*

Introduction

Photosynthesis is the fundament of almost all life on earth, it is the gateway of solar energy into the biosphere and the source of the oxygen necessary to release the full potential of this captured sunlight. Not only is it the life-force[21] of the planet but it also drives our economies be that from contemporary (food, biofuel) or fossil (oil, coal) sources. Despite its pivotal role at both ecological and economic levels, relatively little is known about both the level and nature of genetic variation for photosynthesis in natural populations (Flood et al., 2011). Understanding this is particularly pertinent as photosynthesis is the only major physiological trait yet to be actively improved in crops (Long et al., 2006). A lack of knowledge regarding the genetic architecture of photosynthesis, compounded by the lack of success in breeding programs has led many to argue that photosynthesis is already optimized and it would be more effective to focus our crop improvement efforts on other traits (Denison, 2012). This, however, may no longer be an option as the growing world population combined with an increasingly unstable climate means there is an urgent need to explore all possible avenues to increase our agricultural output whilst at the same time minimizing our ecological footprint. Improving primary production is a clear way in which these goals can be realized (Long et al., 2006).

Direct and sustained selection for increased photosynthetic capacity is unlikely to occur in nature as there are likely to be competing traits which lead to tradeoffs in natural ecosystems, for example higher nitrogen concentration in leaves leading to increased herbivory (Gassmann, 2005). In extreme cases, e.g. when CO_2 concentrations declined to below 200 ppm, many species have evolved mechanisms to increase their photosynthetic capacity by concentrating CO_2 around the active site of Rubisco resulting in increased primary production (Sage, 2004). This carbon concentrating photosynthesis is known as C4 photosynthesis and is found in some of our most productive crops, for example maize. The evolution of C4 photosynthesis on at least 66 separate occasions (Sage et al., 2012) shows that photosynthesis is subject to selection and has adapted to past changes in the physical environment. Interestingly, in the case of C4 photosynthesis the carbon concentrating mechanism is not without a cost and it is projected that C4 photosynthesis will lose its advantage over conventional C3 photosynthesis when the concentration of CO_2 exceeds 700ppm. In modern agricultural systems some of the tradeoffs found in nature are less limiting or can be overcome completely via good agronomic practices,

[21] There are organisms which live of geothermal energy, but they are a relatively small component of the biosphere, more importantly life as we know it, and most probably complex life, would not exist without oxygenic photosynthesis.

thus it should be possible to select for increased primary production in agriculture without incurring the negative tradeoffs experienced in nature.

The first step to initiating such a breeding program is to study the genetic architecture of photosynthesis in an effort to understand how much variation there is and over how many loci is it distributed. By understanding the genetic architecture of the intraspecific variation for photosynthesis more informed decisions can be made about possible avenues for improvement. If the majority of the phenotypic variation is explained by a few loci of large effect, then marker assisted selection or transgenesis are likely to be suitable approaches. If, however, the variation is distributed over numerous small effect loci then traditional breeding and/or genomic selection are likely to generate better results. Finally if, as is likely, there are a few loci of medium to large effect and many of small effect, a combination of the above approaches would be best.

We chose the model plant *Arabidopsis thaliana* (Arabidopsis) to perform a GWA for photosynthesis with. This species is well suited to such studies (Atwell et al., 2010) and has a large number of research resources, T-DNA insertion lines, transformation techniques etc. available, which aid in following up any associations detected. In this study we screened two sets of wild type Arabidopsis accessions for their photosynthetic efficiencies in standard growth chamber conditions. These phenotypic data were used to perform a genome wide association (GWA) to investigate the genetic architecture of the trait and to identify potential gene candidates responsible for the genetic variation in photosynthesis efficiency.

Materials and Methods

Plant material

In total 703 Arabidopsis genotypes were phenotyped, of which 633 had genotype data available both in the form of 214,051 SNPs (Atwell et al., 2010; Baxter et al., 2010; Li et al., 2010; Horton et al., 2012) and whole genome sequence data (Cao et al., 2011; Gan et al., 2011; Long et al., 2013). Genotypes were sampled from throughout the native (Eurasia and Africa) and invasive (North America and New Zealand) range of Arabidopsis in order to maximize potential genetic and phenotypic diversity (Figure 5.1). The 611 genotypes which were genotyped with 214K SNPs were divided into two overlapping mapping populations. The first consisted of 352 globally collected lines, this population was designed to maximize genetic diversity and minimize population structure and is called the HapMap population (Li et al., 2010). The second population consisted of 307 lines all of which come from Sweden, this population was designed as a regional mapping population where the combination of reduced allelic diversity and greater linkage disequilibrium is expected to increase the statistical power of the GWA mapping, in this thesis it is referred to as the Swedish RegMap (Horton et al., 2012). For the purpose of exploring the population size necessary to identify a strong peak, a subset of 68 northern accessions from the Swedish RegMap were also used as a separate mapping population in this thesis. The remaining 92 genotypes, 22 of which have been sequenced as part of the 1001 genomes project (Weigel and Mott, 2009), were included to increase the geographic coverage of our study and to screen for additional phenotypic extremes. An overview of the populations used is given in Table 5.1 and details of each accession is given in Supplemental Data 5.1.

T-DNA insertion lines were ordered from the Nottingham Arabidopsis Stock Centre (NASC) (http://arabidopsis.info/) (Scholl et al., 2000) after which they were propagated to increase seed availability and confirm homozygosity for the insert; an overview of the T-DNA insertion lines used is given in Supplemental Data 5.2.

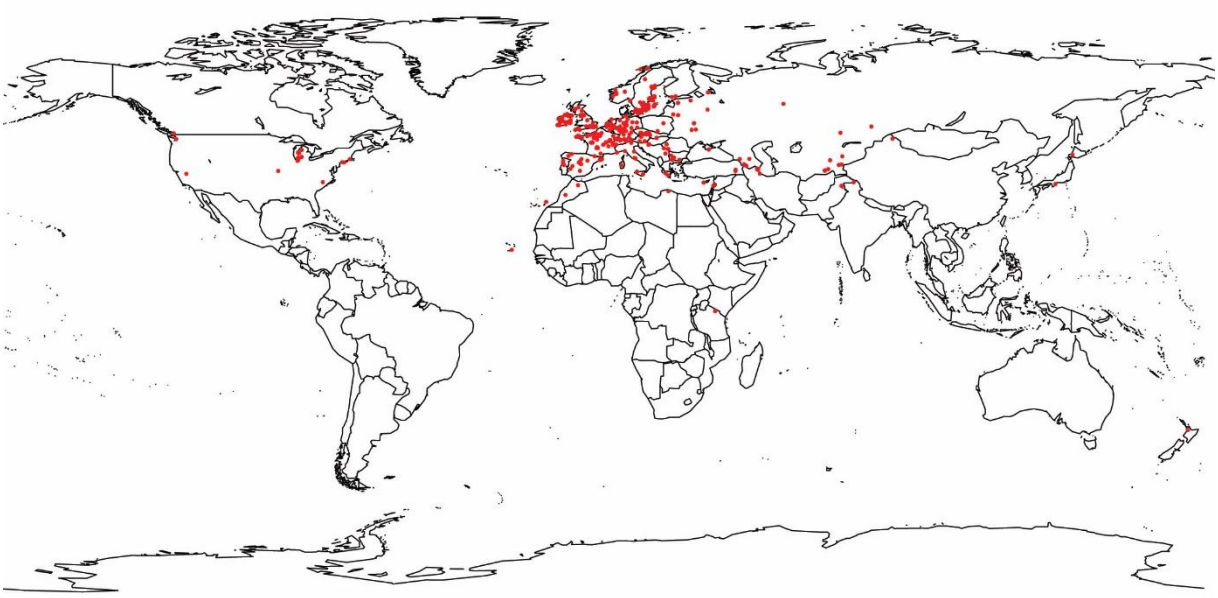


Figure 5.1. *Geographical distribution of the 703 Arabidopsis genotypes used in this study.*

Table 5.1. *Overview of populations screened, note the populations contain overlapping individuals so the sum of the numbers of individuals in the HapMap, Swedish RegMap and Northern Swedish populations is greater than ' All' .*

Population Name	**Number of individuals (genotyped)**
All	*703 (633 of which 611 250KSNP set)*
HapMap	*352 (352)*
Swedish RegMap	*307 (307)*
Northern Swedish	*68 (68)*

Plant growth and measurement

Plants were grown in a climate controlled growth chamber with the following settings: 10/14hr day night regime to suppress early flowering, 70% relative humidity, and 20/18°C day night temperature. Growth irradiance was 200 µmol $m^{-2}s^{-1}$ in all experiments. All genotypes were sown with a minimum of four replicates, though over 400 genotypes were sown with at least eight replicates and a further eight reference genotypes (Col-0, Bur-0, Ely, C24, Ws-4, Sha, Ler-1 and Can-0) were sown with more than 100 replicates each. In total six replicate experiments were conducted resulting in phenotypic measurements of 6,568 plants. Using chlorophyll fluorescence, plants were measured for the operating efficiency of photosystem II (Φ_{PSII}) three times per day; one, four and seven hours into the photoperiod over a 25-day period

(Baker, 2008). Projected leaf area (PLA) was calculated from these images and used to construct growth curves, from which fitted values and the slope of the curve was estimated. To minimize the influence of germination and plant establishment when comparing plant size and growth rates between genotypes, the PLA values were further processed by setting a threshold of 30 pixels as the t=0 (time zero) point. All plants were also screened once per day, one hour before the onset of the light period, for reflectance at eight wavelengths of light; 480 nm, 532 nm, 550 nm, 570 nm, 660 nm, 700 nm, 750 nm and 790 nm. Plants were fully randomized in each screen and design factors were recorded for each individual and used in the estimation of best linear unbiased estimators (BLUEs) of genotype (or 'genotypic effects'), as described in Chapter 3.

Genetic analysis

Additive genetic variance and narrow sense heritability were estimated as described by Kruijer et al. (2015). GWA was performed using an efficient mixed-model association (EMMA) procedure (Kang et al., 2010) implemented in R (R Core Team, 2014). For this we used 214,051 SNPs and mapped all time points, the minor allele frequency was set to 0.05. For key time points we used the GWA portal (http://gwas.gmi.oeaw.ac.at/), an updated version of the online GWA-app called GWAPP (Seren et al., 2012). This allowed the use of whole genome sequence data and expanded both the number of genotypes and markers available for analysis. In total, using the expanded dataset available through the GWA portal, GWA could be performed with 10,000,000 SNPs for 633 genotypes (Table 5.1).The proportion of phenotypic variance explained by a particular SNP was estimated by $4\ f\ (1\text{-}f)\ \beta^2$ divided by the total phenotypic variance, where F is the frequency of the SNP and β is the estimated SNP effect obtained from GWA (Kruijer et al., 2015). Genetic architecture was visualized by histograms of the estimate β for all SNPs above a significance threshold of 10^{-4} (i.e. $-\log_{10}(p) = 4$).

Results

Trait variance

The distribution of photosystem II efficiency (Φ_{PSII}) on a representative day is shown in Figure 5.2. Phenotypic values show a continuous distribution, which is typical for complex polygenic traits. The second peak in Figure 5.2 (a) is caused by a single genotype (Ely) which was identified in an earlier screen to be disturbed in photosynthesis (El-Lithy et al., 2005). This genotype contains a Ser-264-Gly mutation in its D1 protein which confers triazine resistance. A consequence of this resistance is significantly lower Φ_{PSII} efficiencies (see chapters 7 and 8). We used this genotype as a control in all experiments, resulting in 143 replicates responsible for the small secondary peak in Figure 5.2 (a). To best combine the data from multiple replicate experiments and to remove systematic variation caused by technical limitations, i.e. light gradients in the imaging head (Chapter 3 for further details) corrected genotypic means, (BLUEs) were estimated for each genotype. These BLUEs also show a normal distribution, with the exception of Ely (Figure 5.2 (b)).

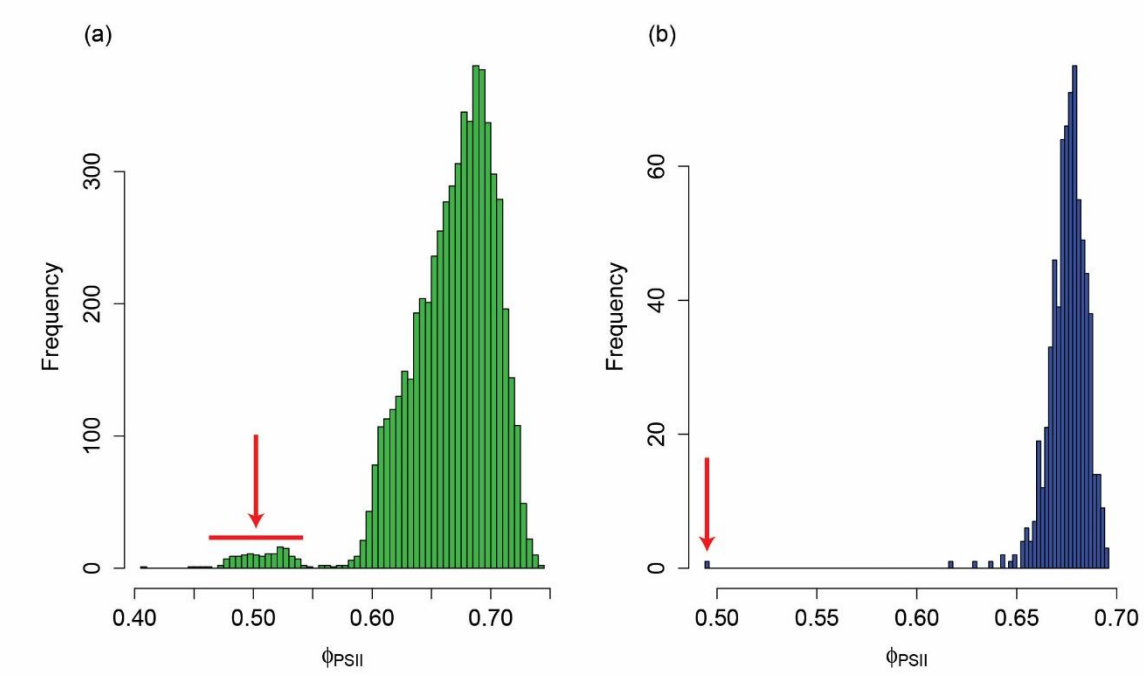


Figure 5.2. *Distribution of Φ_{PSII} values in the morning 18 days after sowing. (a) Raw values of all 5918 plants from six screens, (b) genotypic means (BLUEs) for all 702 genotypes. The red arrows indicate the position of Ely, a natural triazine resistant mutant with disturbed photosynthesis. Data in (a) includes 143 replicates of this genotype.*

Using the 5,918 phenotyped plants for which we also had genotype data (Table 5.1) we estimated broad sense heritability (H^2), and using kinship estimates from the 214,051 SNPs, partitioned the variance into additive and non-additive components to estimate narrow sense heritability (h^2) (Kruijer et al., 2015). This was done for all traits measured and a summary of

these values is shown in Figure 5.3 (a); see also Supplemental Figures 5.1 and 5.2 for individual plots per trait. From Figure 5.3 (a) it can be seen that the majority of the genetic variance appears to be additive.

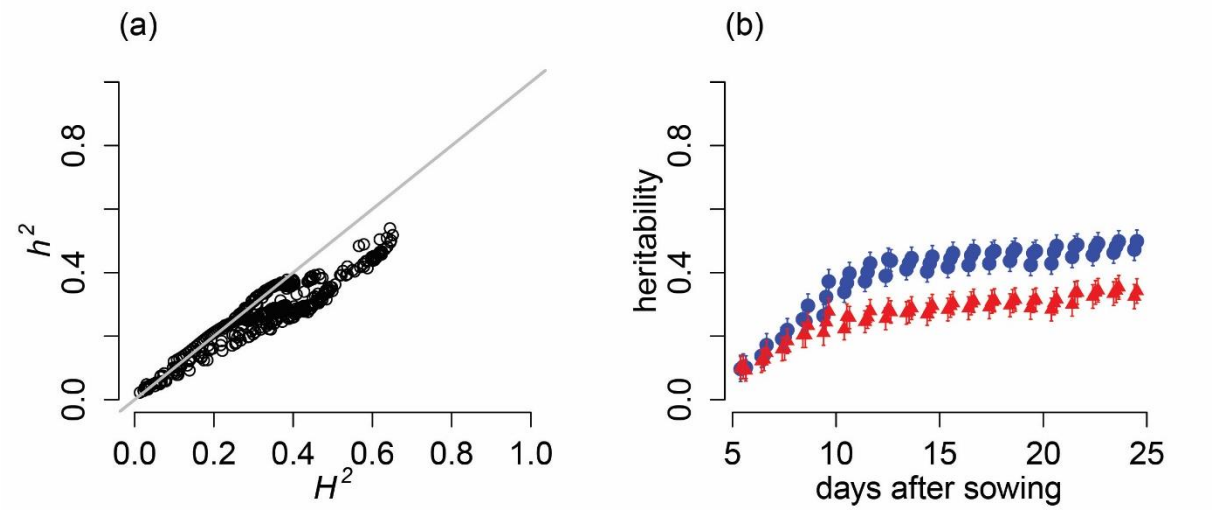


Figure 5.3. *Comparison of narrow sense heritability (h^2) and broad sense heritability (H^2). (a)Narrow sense heritability plotted against broad sense heritability for all traits, grey line indicating a one to one relationship is plotted as a visual guide. (b) Comparison through time of broad sense heritability (blue) to narrow sense heritability (red) for Φ_{PSII} (see Supplemental Figures 5.1 and 5.2 for all other traits).*

Genome wide association mapping

Using the BLUEs estimated for each genotype we performed genome wide association mapping (GWA) using an efficient mixed model approach (EMMA) (Kang et al., 2010; Segura et al., 2012). Since we have data for several traits and have multiple time points per trait it is no longer practical to show all the Manhattan plots displaying –log(p) values for all SNPs tested for associations (up to 1000 plots would be necessary). To overcome this we plotted all SNPs above -log(p) 5.5 through time (Figures 5.4, 5.5 and 5.6). From these results it can be seen that the strength of association (–log(p) value) between specific SNPs and a trait can fluctuate strongly with time – it has dynamic properties. Although some SNPs appear at only a single time point, indicating either extreme temporal dynamism or, more likely, false positives, many SNPs show a stable association for some time. These can be seen as arrays of significant SNPs at a specific locus for a period in time. This chapter will focus mainly on Φ_{PSII} and the results of the other traits will only be referred to where they provide additional insights into the genetic basis of this trait.

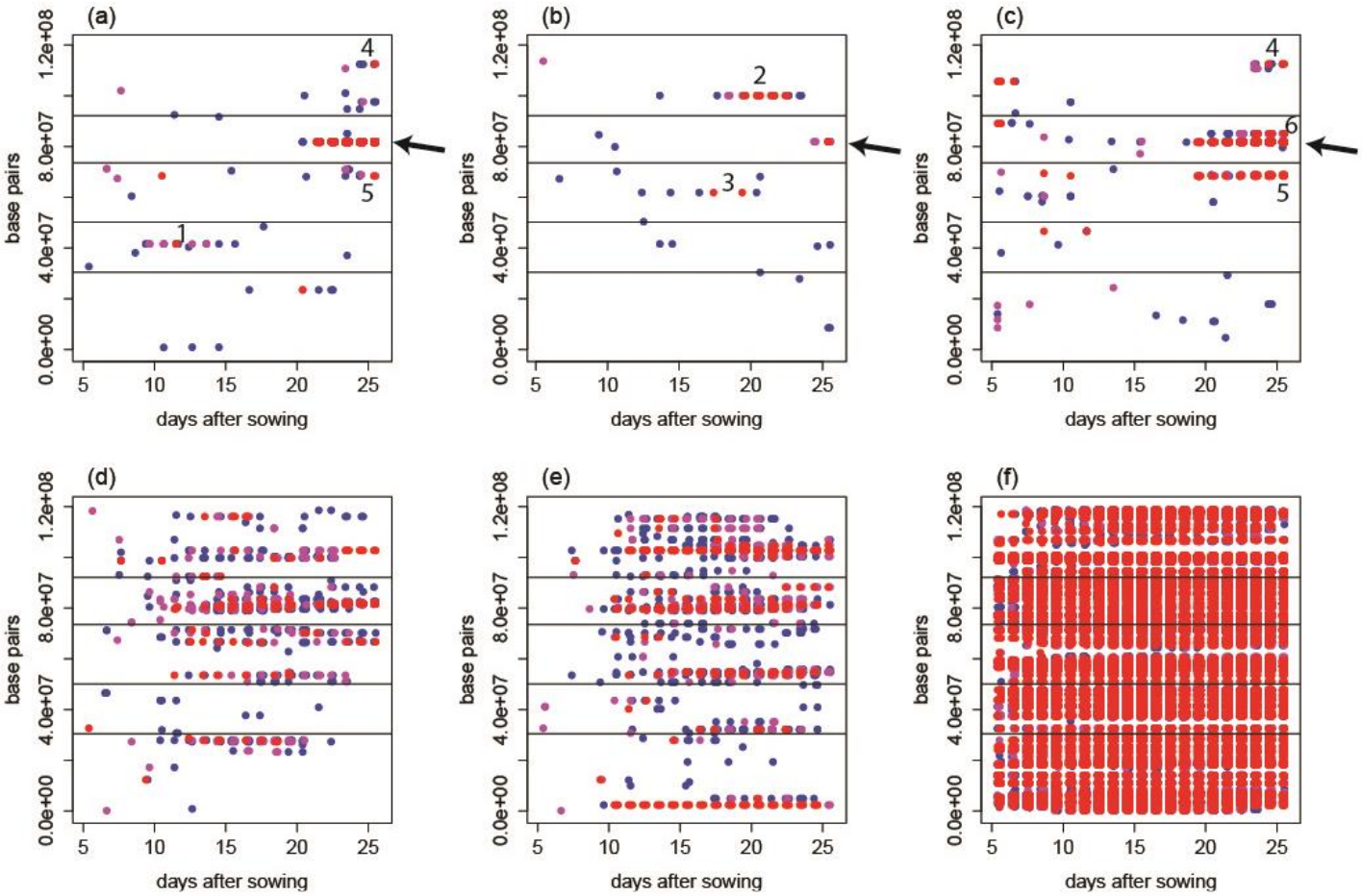


Figure 5.4. *Results of GWAS through time for Φ_{PSII}. Blue represents SNPs above* -log(p) *5.5, magenta represents SNPs above* -log(p) *6 and red represents SNPs above* -log(p) *6.5 (Bonferroni threshold). (a) All plants (n=611), (b) HapMap lines only (n=352), (c) Swedish RegMap only (n=307), (d) to (f) as in (a) to (c) with psbA mutant Ely included. The x axis shows the GWA results through time and the y axis indicates the genetic position. The horizontal lines indicate the chromosome boundaries going from chromosome 1 at the bottom to chromosome 5 at the top. The arrows indicate the position of ACD6, and the numbers indicate loci of interest.*

Figure 5.4 shows the results of three GWA analyses per day for Φ_{PSII} over a 20-day period. Many of the loci detected are unique to each population. When comparing the HapMap with the Swedish RegMap this is perhaps unsurprising since these two populations differ considerably in the genotypes used. What is perhaps surprising is the disappearance of some loci and the appearance of others when the two populations are combined and mapped as one. In Figure 5.5 PLA was mapped in the combined population using both raw data (Figure 5.5 a) and PLA estimates resulting from a fitted curve (Figure 5.5 b) (Chapter 3). In Figure 5.5 (c) the slope of the fitted curve was used as a mapping phenotype. Figure 5.5 (d) presents the same results except that instead of setting t=0 as the day the seed was sown, t=0 was set as the point at which a genotype breached a size threshold of 30 pixels. This was done to minimize the influence of variation in germination rate and seedling establishment, thus allowing the growth curves of the plants to be compared from a normalized starting point. This enhanced the genetic signal as the heritability of this trait increases when using a size threshold as a starting point as opposed to sowing date (Supplemental Figure 5.2 e and f). In addition to increasing the trait heritability it also leads to the detection of the locus on chromosome 4 which was detected when mapping Φ_{PSII}, indicated by a black arrow. Figure 5.6 shows the genetic associations detected for reflectance at eight wavelengths of light, in the combined population. The genetic associations detected for reflectance are distinct, with most of the overlapping peaks occurring between similar wavelengths of light. Perhaps in support of this, when these reflectance traits are plotted in a correlation matrix, closer wavelengths show tighter correlation (Supplemental Figure 5.3). At the two longest wavelengths, 750 nm and 790 nm, the peak on chromosome 4, which is visible in Figures 5.4 and 5.5, can be seen.

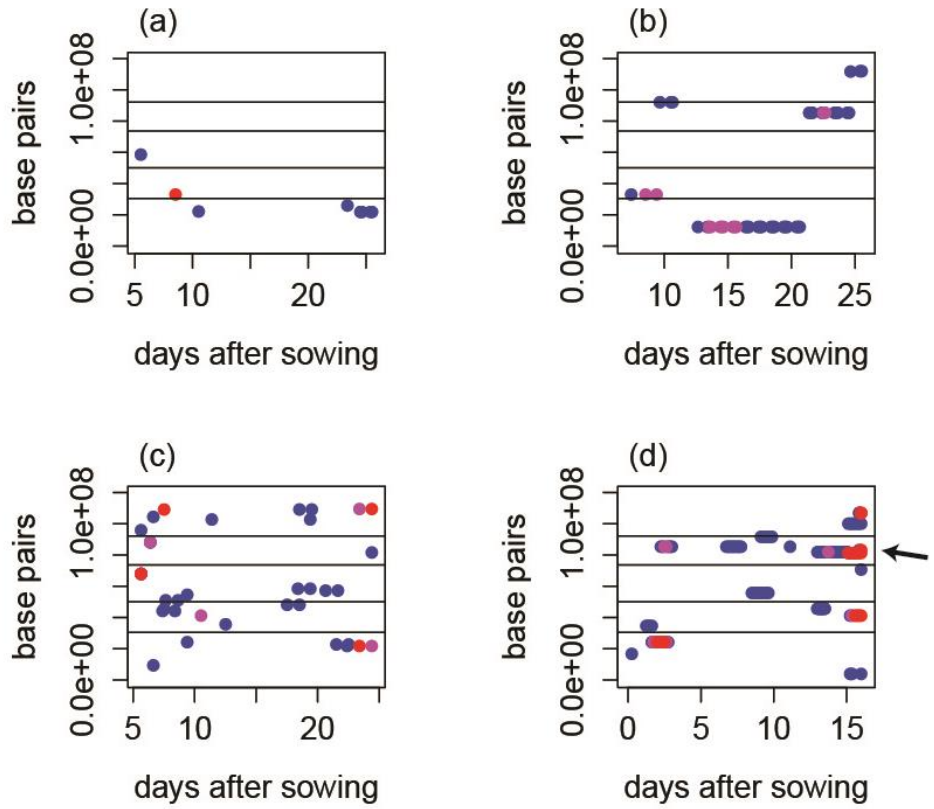


Figure 5.5. *GWA through time for projected leaf area (PLA). (a) Raw PLA data using a mask generated from Φ_{PSII} images. (b)When using a fitted PLA. (c) GWA for the values of the slope of the fitted curve. (d) Slope of fitted curve with t=0 determined by a 30 pixel threshold. Blue represents SNPs above* -log(p) *5.5, magenta represents SNPs above* -log(p) *6 and red represents SNPs above* -log(p) *6.5 (Bonferroni threshold). Arrow indicates the position of a co-locating peak overlying the ACD6 locus also identified in the Φ_{PSII} results.*

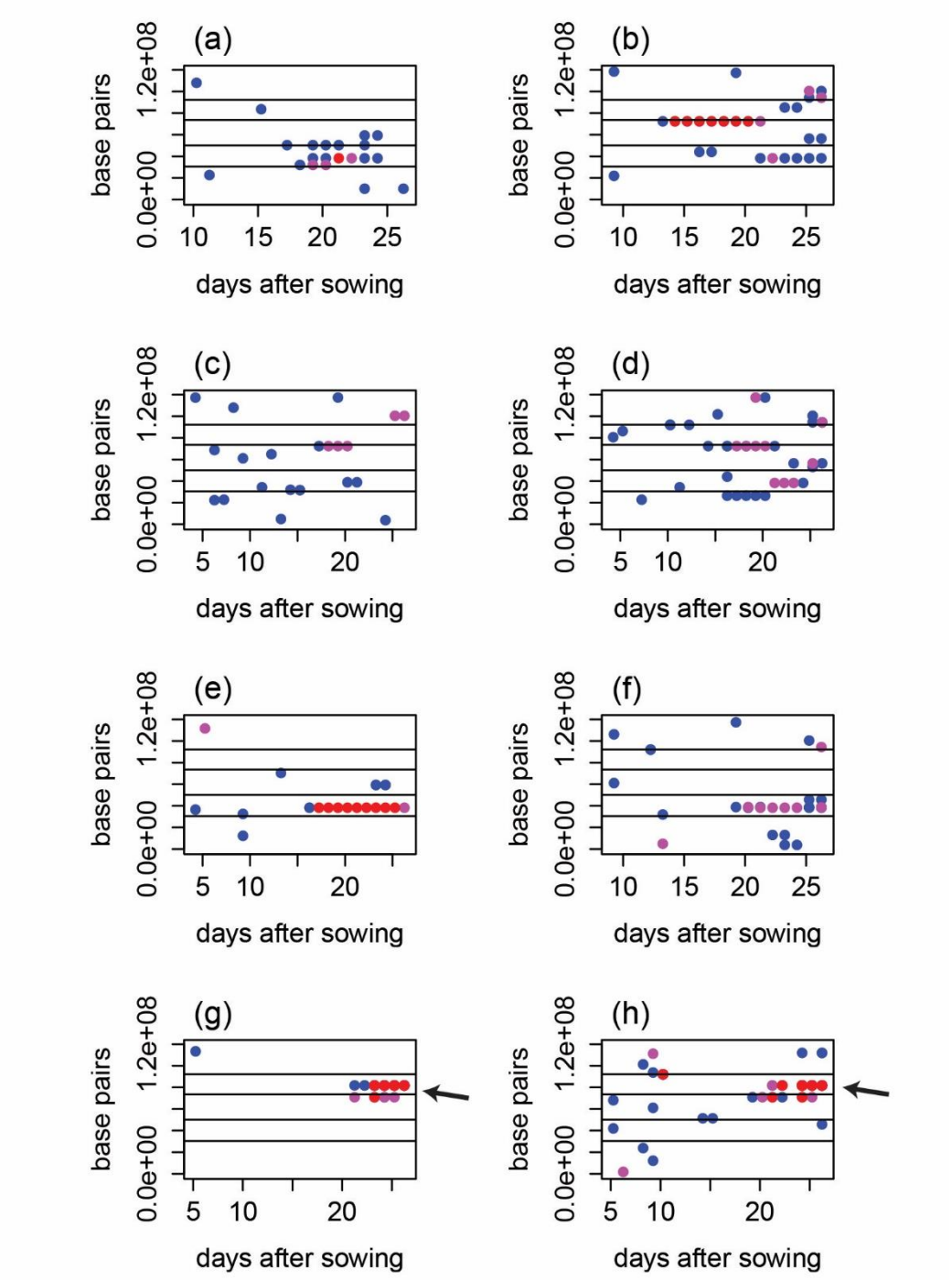


Figure 5.6. *Results of GWAS through time for reflectance of light at 480nm (a), 532nm (b), 550nm (c), 570nm (d), 660nm (e), 700nm (f), 750nm (g), and 790nm (h). Blue represents SNPs above* -log(p) *5.5, magenta represents SNPs above* -log(p) *6 and red represents SNPs above* -log(p) *6.5 (Bonferroni threshold). Arrows indicate the position of a co-locating peak overlying the ACD6 locus also identified in the* Φ_{PSII} *and PLA results.*

Candidate loci

The strong association of a region on chromosome 4 which was detected for Φ_{PSII}, slope of the growth curve and reflectance at 750 nm and 790 nm, contains about 50 genes of which 23 contained SNPs above the Bonferroni threshold (Supplementary Table 5.1). For all traits this quantitative trait locus (QTL) only begins to become significant from 20 days after sowing. More careful inspection of the Φ_{PSII} images from these time points showed that the older leaves of some genotypes exhibited a precipitous decline in Φ_{PSII} (Figure 5.7). Both the late onset of the QTL and the decline of Φ_{PSII} in older tissues suggested that necrosis may be responsible. One of the candidate genes was *ACCELERATED CELL DEATH 6* (*ACD6*; AT4G14400), alleles of which are known to cause necrosis (Lu et al., 2003; Todesco et al., 2010) which would result in a strong loss of Φ_{PSII}. When looking at the images of the plants in near infra-red wavelengths (NIR), the leaves which show a decline in Φ_{PSII} also show dark spots which resemble necrotic lesions (Figure 5.7 c). This is in agreement with the results in Figure 5.6 panels (g) and (h) where reduced reflectance at NIR wavelengths (750 nm and 790 nm) also detected a QTL at the *ACD6* locus. The co-occurrence of both the phenotypes (Figure 5.7) and the mapping results in Figures 5.4 and 5.6 suggests that the NIR reflectance and Φ_{PSII} phenotypes have a common genetic origin. In addition to causing necrosis, *ACD6* has also been associated with slowed rates of leaf initiation (Todesco et al., 2010). This fits with the results of Figure 5.5 (d) where the slope of the growth curve also maps to a peak over the *ACD6* locus. As a measure of the maximum rate of leaf initiation we obtained the maximum slope of the fitted growth curve for each genotype and mapped with this, again the *ACD6* locus was detected (Figure 5.8). Thus several lines of evidence, reduced Φ_{PSII} in older leaves, visible necrotic lesions in the same leaves at NIR wavelengths and the slope of the growth curve, all support the conclusion that the causal gene is *ACD6* (Figure 5.8). Furthermore many of the lines which exhibit these phenotypes in our datasets are also known to contain active Est-1 like *ACD6* alleles which are expected to promote the onset of necrosis (Todesco et al., 2010).

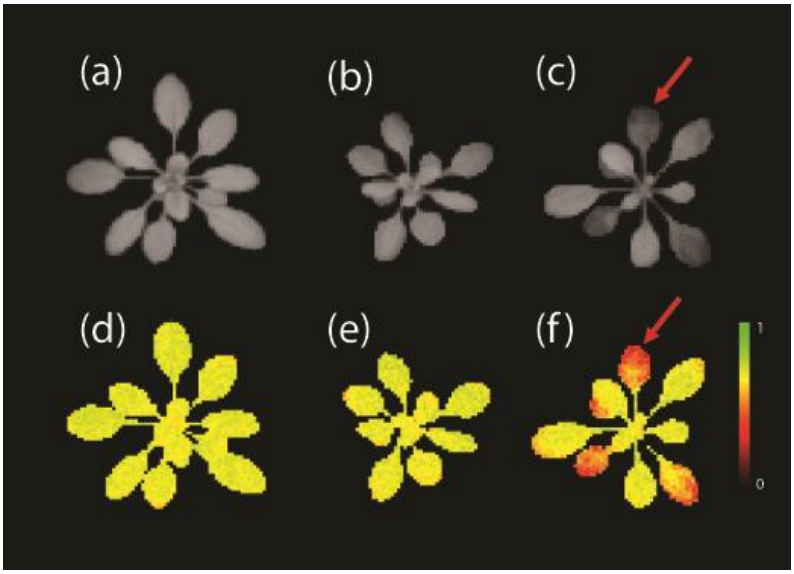


Figure 5.7. *Images of plants for three A. thaliana genotypes, 25 days after sowing. The top row, (a) to (c), shows plants at 790 nm reflectance (near infra-red, NIR) and the bottom row, (d) to (f), shows false colour Φ_{PSII} images (see colour scale bar). Pictures (a) and (d) are of Col-0, (b) and (e) are of Roed-17-319 and (c) and (f) are of TAA-14 (see Supplemental Data 1 for full list of references). The red arrows indicate a leaf showing signs of necrosis.*

Apart from the locus overlying *ACD6*, additional loci were detected (numbered on Figure 5.4 and shown in more detail in Supplemental Figure 5.4). Of these loci two, (4 and 5) appear to interact with the *ACD6* locus. When the highest SNP in the *ACD6* locus was set as a cofactor in the GWA mapping, the significance of QTLs 4 and 5 declined (Supplemental Figure 5.5). This indicates some form of genetic interaction, although it is possible that they may be synthetic associations brought about due to uncorrected population structure (Platt et al., 2010; Seren et al., 2012). It should be noted, however, that they do not completely disappear and contain plausible gene candidates (Supplemental Figure 5.5). QTL 4 contained AT3G49110 or *PEROXIDASE CA* (*PRXCA*) which has been shown to be important for generating an apoplastic oxidative burst upon pathogen infection. When this gene was knocked out the plants became more susceptible to *Pseudomonas syringae* (Daudi et al., 2012). The fact that this is an immune gene related to pathogen defence indicates that it may indeed be interacting with the *ACD6* locus perhaps somehow modulating the severity of the phenotype. Under QTL 5 AT5G49940 or NIFU-LIKE PROTEIN 2 (NFU2) was an interesting candidate as it has been shown to assemble chloroplastic iron-sulphur clusters (Gao et al., 2013) which are key components of several photosynthetic enzymes. QTL 1 shows a time of day dependency, generally only becoming significant in the afternoon and evening from 9 to 16 DAS (Figure 5.5 a). This QTL was right on top of AT2G26150 or *HEAT SHOCK TRANSCRITIONS FACTOR A2* (*HSFA2*),

which has recently been described as having a role to play in mediating response to high light stress (Jung et al., 2013). Based on gene annotation the most convincing candidate under QTL 2 is AT5G23440 or FERREDOXIN/THIOREDOXIN REDUCTASE SUBUNIT A (VARIABLE SUBUNIT) 1 (FTRA1). This gene is located in the chloroplast and is predicted to be involved in the light reactions (Zybailov et al., 2008). Although on the other side of this peak is another immune gene AT5G23400, which is a leucine rich repeat family protein, which could potentially play similar role to *ACD6* in affecting photosynthesis. QTL3 occurs next to yet another immune gene AT3G48090 or ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1) which has been shown to be an important regulator of the immune responses (Falk et al., 1999; Wittek et al., 2014). QTL 6 contains AT4G21680 or NITRATE TRANSPORTER 1.8 (NRT1.8) which is an interesting candidate as it is involved in nitrate transport from shoot to root under stress conditions, it appears to exhibit a trade-off in that it enhances stress tolerance but reduces growth under control conditions (Zhang et al., 2014) in this respect it is similar to *ACD6*.

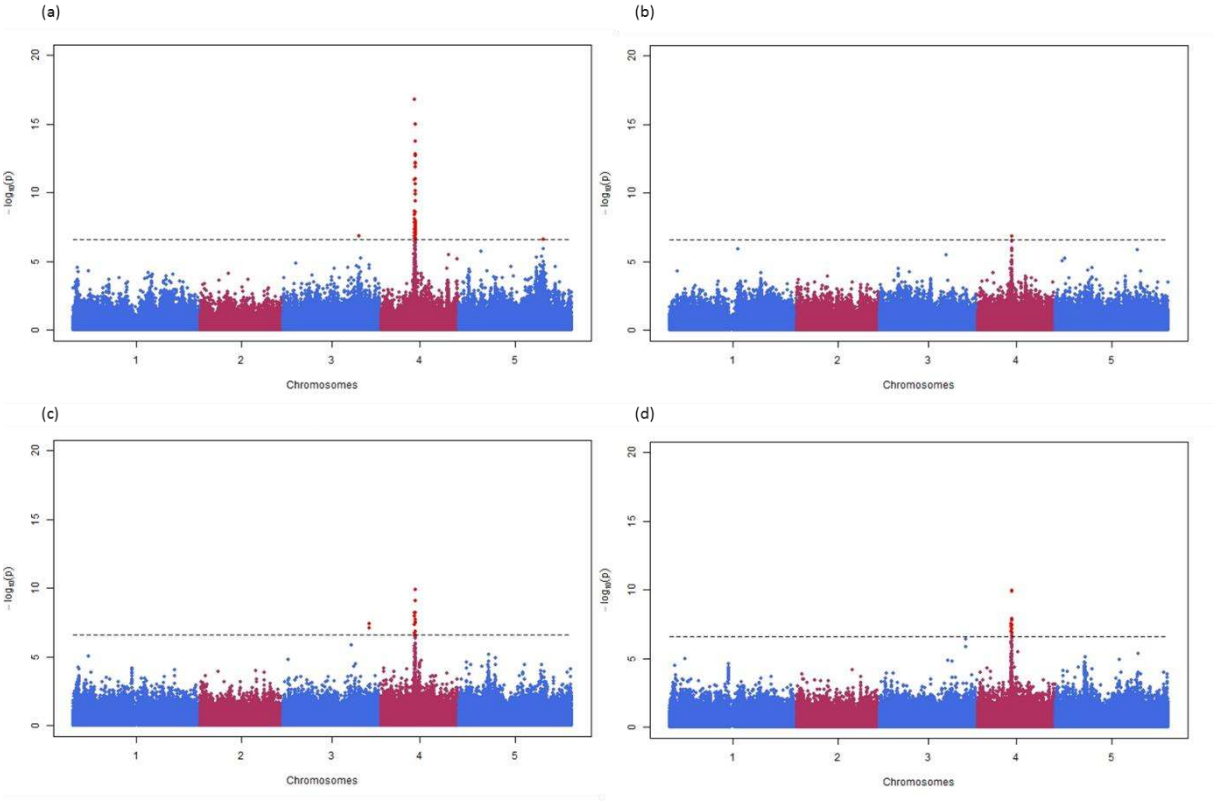


Figure 5.8. *Manhattan plots of traits identifying a QTL at the ACD6 locus. Genetic position is indicated on the x-axis and –Log(p)value is indicated on the y-axis, each marker is represented by a dot, the broken line indicates the Bonferroni significance threshold. (a) Φ_{PSII} on the morning of day 25. (b) The maximum slope from the growth curve calculated from Φ_{PSII} mask where t=0 was set to a size threshold of 30 pixels. (c) Reflectance at 790 nm and (d) 750 nm on the morning of day 24.*

T-DNA lines

Based on the results of the first four experiments and how often a particular SNP was detected above a -log(p) 4 threshold, a selection of additional candidate genes was made. T-DNA lines of these genes were ordered and tested to confirm homozygosity of the insert (Supplemental data 5.2). Of these a further selection was made based on availability of confirmed homozygous seed. These lines were included in a phenotypic screen using the same phenotyping platform and conditions as before. The results of this screen are presented in Supplemental data 5.3. T-DNA insertion lines of several genes resulted in clearly detrimental phenotypes (e.g. AT2G01320, AT2G30280, and AT3G59780). One particularly severe phenotype was for a T-DNA insertion in AT2G01290, the cytosolic ribose-5-phosphate isomerase gene, which was described previously as resulting in dysfunctional chloroplasts and reduced photosynthesis (Xiong et al., 2009). Apart from several T-DNA insertion lines exhibiting reduced Φ_{PSII}, three lines exhibited significantly higher Φ_{PSII} than the wild type (Figure 5.9, *P* values per time point reported in Supplemental data 5.3): AT3G45910 which is an unknown, mitochondrion targeted protein; AT3G63070 *HUA2 LIKE 3 (HULK3)* which is part of a family of proteins necessary for proper development and regulation of *FLOWERING LOCUS C FLC* (Jali et al., 2014); and AT5G25420 which is predicted to be a Xanthine permease involved in transmembrane transporter activity. Of these genes, the unknown mitochondrial protein and the permease had the most consistently higher Φ_{PSII} values throughout the duration of the experiment (Supplemental data 5.3). T-DNA insertion lines of the gene candidate underlying QTL 1 *HSFA2* along with double and triple knock-outs of two other A-type heat shock transcription factors *HSFA1D* and *HSFA3* were included in this screen (Jung et al., 2013). However no significant differences were found for Φ_{PSII} between the *HSF* knockouts and wild type. As discussed below, further experiments to confirm the gene knock-outs and the absence of additional T-DNA inserts are necessary, as well as cloning of the various alleles into the different knock-out lines that variation at these loci is indeed causal.

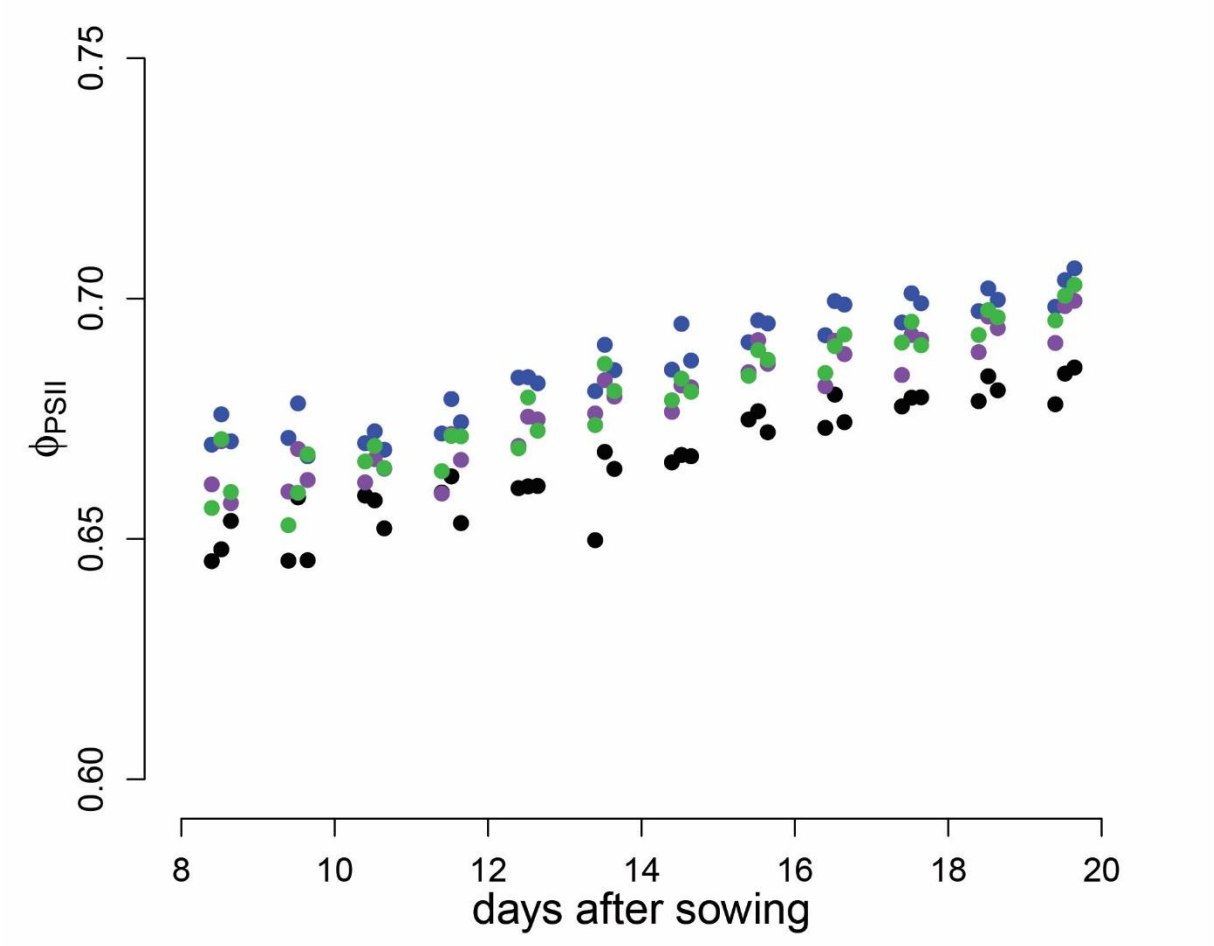


Figure 5.9. *Φ_{PSII} values of wild type (Col-0; black) and mutant T-DNA insertion lines for genes At3g45910 (blue), At3g63070 (purple), and At5g25420 (green) displaying higher Φ_{PSII} values than Col-0 in the course of 12 days.*

Genetic architecture

Figure 5.10 shows the distribution of the individual explained variance per SNP for all the SNPs above the -log(p) 4 threshold across all time points for Φ_{PSII}. As population size increases the average explained variance per SNP decreases. Initially this may make sense as the total variation should increase with increasing population size whilst the contribution of individual loci will not necessarily increase in parallel, hence a decrease in explained variance per SNP. However, one would not expect the effect size of a particular SNP to change with population size as the effect of the SNP should be independent. In Figure 5.11 we see that the distribution of effect sizes per SNP for Φ_{PSII} also shows the same trend. This phenomenon has been described previously and is known as the Beavis effect (Beavis, 1994). Taken together Figures 5.10 and 5.11 show the average variance explained and the average effect size per significant (above a -log(p) 4 threshold) SNP is quite small, which points to a highly polygenic genetic architecture where most of the variation is explained by loci of small effect.

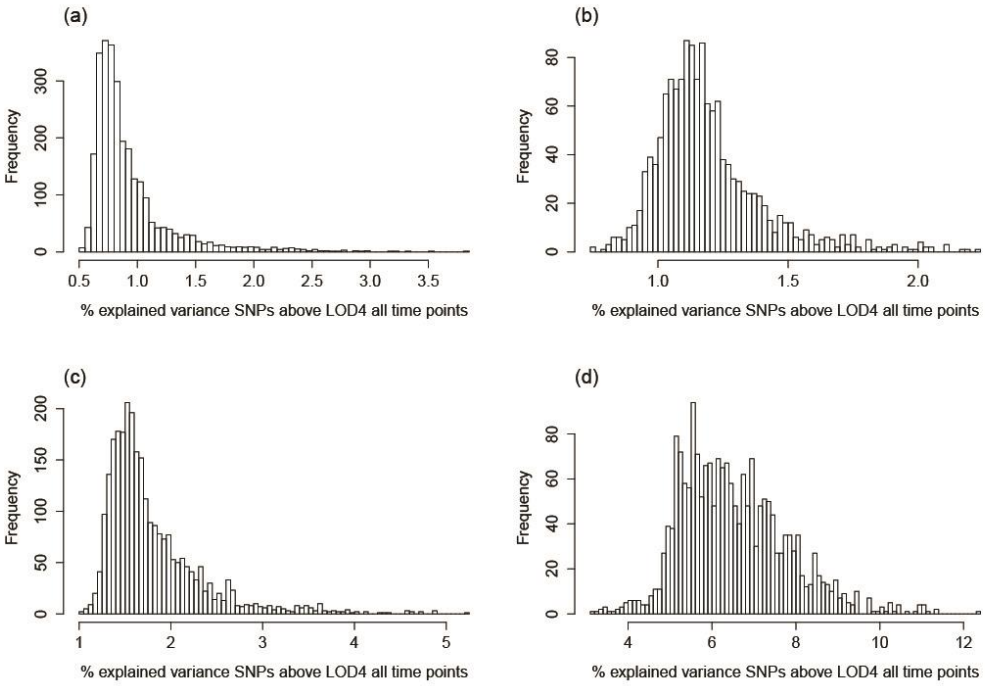


Figure 5.10. *Distribution of explained variance for SNPs above a -log(p) 4 threshold for Φ_{PSII}. (a) All genotypes, (b) HapMap panel, (c) Swedish RegMap panel, and (d) Northern Swedish genotypes .*

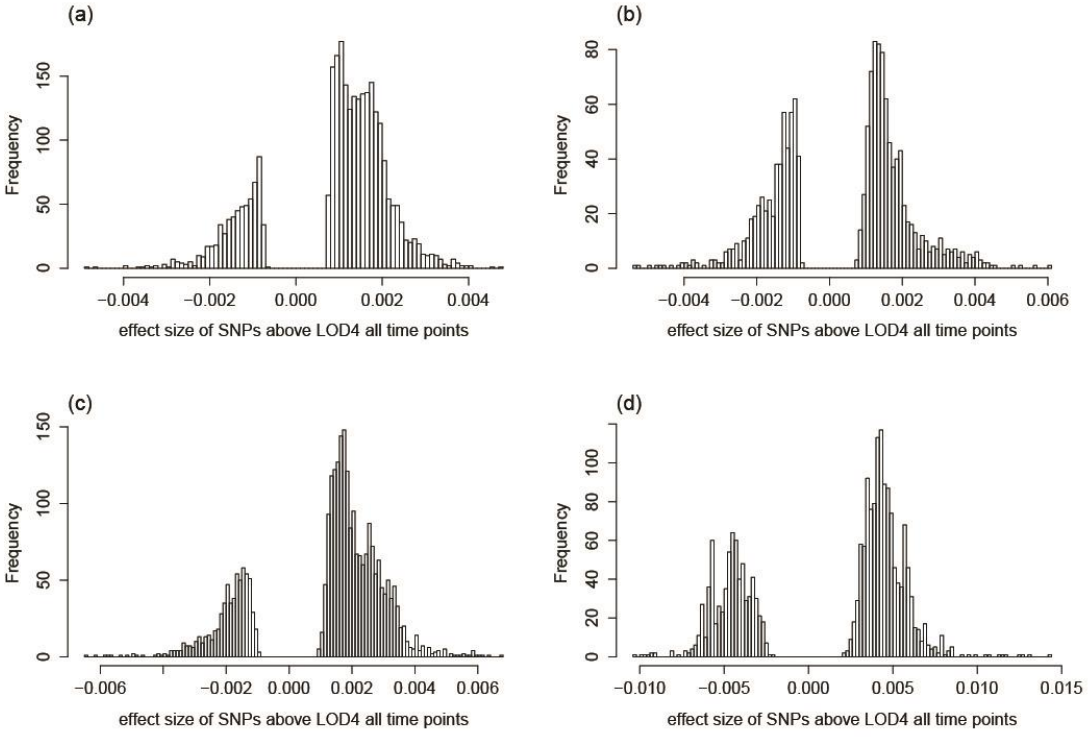


Figure 5.11: *Distribution of effect size for SNPs above a -log(p) 4 threshold for Φ_{PSII}. (a) All genotypes, (b) HapMap panel, (c) Swedish RegMap, and (d) Northern Swedish genotypes.*

Discussion

Photosynthesis is unlikely to be maximized under natural conditions due to competing interests or trade-offs. Our most strongly associated locus for Φ_{PSII} contains the *ACD6* gene. Allelic variation in the *ACD6* gene is responsible for a key fitness tradeoff between growth and resistance to pathogens. Alleles that give rise to reduced growth but enhanced disease resistance occur at a frequency of about 20% in natural Arabidopsis populations (Todesco et al., 2010). In recent years it has become increasingly clear that pathogen resistance is at the basis of numerous tradeoffs in natural populations (Todesco et al., 2010; Alcázar et al., 2014; Chae et al., 2014; Todesco et al., 2014). Interestingly many of the other significant QTLs detected also contain immune genes suggesting that there may be a more general link between disease resistance and photosynthetic performance. A gene ontology analysis using all genes containing markers above a -log(p) of 4, failed to find any enrichment. This could be due to linkage disequilibrium obscuring enrichments by adding an excess on non-causal genes. Nevertheless these results suggest that a potential target for future crop improvement programs should be the development of disease and pest resistance which minimizes the trade-offs usually incurred. Steps have already been taken for example; herbivore targeted chloroplast-located double stranded RNA (Zhang et al., 2015) which allows for the accumulation of these double stranded RNAs as the chloroplast does not degrade them thus negating the need for constitutive immune activation. Another recent development is the creation a system whereby the drought response of plants was controlled by exogenously applied agrochemicals (Park et al., 2015). Such a system could be adapted to create a pathogen targeted inducible immune response. The potential for controlled induction, i.e. after a *Phytophthora* warning for a potato crop, again removes the need for constitutive immune activity and therefore minimizing immune related tradeoffs.

When comparing the significance of the *ACD6* locus in the various populations we find considerable differences. Mapping with either all genotypes together or the Swedish RegMap gives very significant -log(p) scores (Supplemental Figure 5.6). The HapMap panel, which was designed to maximize genetic diversity, resulted in more precise detection of this locus, but with less statistical confidence (Supplemental Figure 5.6). This reduction in significance is probably due to genetic heterogeneity as there are several active alleles of the *ACD6* locus found globally (Todesco et al., 2014) whilst the allelic diversity in Sweden is likely to be much less. To test our ability to detect this locus we performed GWA on plants collected from Northern Sweden (n=68) and found similar power to find this locus when compared to the HapMap (Supplemental Figure 5.6). Importantly not only was the locus detected but the region was still

small, indicating that the level of recombination was sufficient to pinpoint this locus. This illustrates the potential advantages of performing GWAS in local populations. Due to reduced diversity such local mapping populations have the potential to combine the power of recombinant inbred lines whilst maintaining resolution similar to that found in natural populations. Detection of the *ACD6* locus was possible with only 68 genotypes, larger pools of local accessions are likely to perform even better with both increased power and resolution (Korte and Farlow, 2013).

Many of the loci detected were unique to each population (Figure 5.4). When comparing the HapMap with the Swedish RegMap this is perhaps unsurprising since the genetic diversity and structure of the two populations differ considerably. There are several potential explanations for the disappearance of some loci and the appearance of others in the combined population analysis. The first explanation would be that all loci detected are false positives representing random chance. This is in effect the null hypothesis. Whilst this hypothesis cannot be completely rejected, it is unlikely since many occur above the stringent Bonferroni threshold at multiple time-points (Figure 5.4). Second, genetic heterogeneity in the combined analysis reduced the power to detect associations. Genetic heterogeneity is where there are multiple causal alleles which results, either no association being detected or the generation of synthetic associations (Platt et al., 2010). Synthetic associations differ from false positives in that they are statistically valid but biologically false (Platt et al., 2010). This can happen when there are many alleles of the same gene segregating in a population. If a random non-causal locus happens to occur in a set of genotypes, which have different allelic variants of the gene, but which result in similar phenotypic score, then the GWA analysis will identify this as the most likely candidate locus, not the causal locus for which there are several allelic variants segregating. An example in Arabidopsis is the *FRIGIDA* locus, where there were multiple causal polymorphisms in this gene resulting in similar flowering time phenotypes. Not only did this interfere with the ability to find this known locus but is also gave rise to synthetic associations at distantly linked markers (Atwell et al., 2010). Genetic heterogeneity is more likely to be an issue in the more diverse HapMap, or in the combined analysis, than in the Swedish RegMap. Conversely the reduced allelic diversity present in regional mapping populations could mean that for some loci there is no genetic variation segregating and thus no variation to detect. Thus a third possible explanation for the differences between the HapMap and the Swedish RegMap is that some loci identified in the HapMap do not segregate in the Swedish RegMap. Finally, a fourth reason may be that the greater statistical power afforded by

the combined analysis, facilitated the detection of loci of small effect that were below the detection threshold when the populations were analyzed separately (Korte and Farlow, 2013). No matter the cause, a clear conclusion is that the choice of population is likely to have a very large impact on the loci detected, even when the populations overlap for many genotypes.

Another result illustrated in Figure 5.4 is the presence of synthetic associations induced by a mutation of large phenotypic effect. One genotype in the HapMap panel had a much lower Φ_{PSII} than any of the other accession studied (Figure 5.2). When this genotype was included in the mapping procedure it gave rise to many synthetic associations (Figure 5.4 d-f). The possibility of rare variants causing such synthetic associations has been both theoretically and empirically validated (Dickson et al., 2010), however, this case is, to our knowledge, unique, as the synthetic associations are due to a mutation in an organelle encoded protein. This mutation is in the chloroplast encoded *psbA* gene which results in the expression of a mutated D1 protein conferring triazine resistance (El-Lithy et al., 2005). It is found in the Ely genotype mentioned earlier and shown as an outlier in Figure 5.2. This genotype was collected from a railway line in England where triazines were sprayed and presumably evolved in response to this selective pressure (see Chapter 8). We termed this phenomena of organelle induced synthetic associations plasmotypic noise. This is ignored as a potential source of false positives in the GWAS literature, in general the role of organelle genomes in organismal performance has been poorly explored (Pesole et al., 2012) (Chapters 7 and 8). The number of synthetic associations was least in the combined analysis which is perhaps due to the larger population being less susceptible to distortion due to a single outlier. In the separate populations, the Swedish RegMap is affected to a much greater degree than in the HapMap, this is likely to be due to the fact that as well as being a phenotypic outlier it is also a genetic outlier as it is the only non-Swedish accession. This indicates that synthetic associations are more likely when a genotype is both a phenotypic and a genetic outlier. Thus care should be given when constructing a GWA population that there are no extreme genetic outliers as when they are phenotypically divergent they are likely to cause synthetic associations.

The small differences between narrow sense heritability and broad sense heritability suggests that most of the variation present for the traits measured is additive (Figure 5.3, Supplemental Figures 5.1 and 5.2). The generally additive nature of genetic variation is supported by recent studies (Hill et al., 2008; Polderman et al., 2015) and it has been argued that not only is additivity empirically validated but it is the most parsimonious explanation

(Robinson et al., 2014), but does Occam's razor really cut epistasis from the equation?[22] Many prominent quantitative geneticists have argued that due to the inherently non-linear nature of the molecular interaction underlying biological systems epistasis is almost certainly present (Mackay, 2014). It been also noted that epistatic variation can appear additive (Hansen, 2013; Mackay, 2014). The importance of epistasis is a hotly debated topic particularly with regard to GWA as it may account for the so-called missing heritability (Zuk et al., 2012).

While the *ACD6* locus could be detected with high confidence, other loci could not be identified with such certainty. This lack of power to detect loci is increasingly recognized as a potential shortcoming of GWA studies where only a few hundred individuals were used (Korte and Farlow, 2013). The likely causes of this failure to detect many loci are, in summary: An abundance of large effect size alleles occurring at low frequency, which GWA is not designed to detect; Allelic heterogeneity, which may explain the difference in power of RegMap with HapMap results for the *ACD6* locus; And perhaps most importantly an abundance of small effects size alleles which we are underpowered to detect (Gibson, 2012; Rockman, 2012).

If there are many rare alleles of large effect then instead of using GWA future genetic studies of photosynthesis should use constructed populations such as bi-parental or multi-parent mapping populations (Keurentjes et al., 2007; Kover et al., 2009; Huang et al., 2011). Any rare alleles which occur in the parents of such populations can be detected as all alleles are increased to more or less equal frequency during the construction of the population. For photosynthesis it is indeed possible that such rare alleles exist, the results of Chapter 3 of this thesis suggest that loci which have a detectable effect on Φ_{PSII} tend to be unique to the mapping population they are detected in. Also none of the QTLs mapped in Chapter 3 have been found in this GWA analysis, even though the parental lines of the RIL populations are in the HapMap panel. Such rare alleles are more likely to be deleterious (Gibson, 2012) and thus of less direct interest for improving photosynthesis.

The occurrence of rare alleles of significant effect does not preclude the possibility that there are also many common alleles of small effect. If there the common allelic variants affecting photosynthesis are generally of very small effect then for GWA to succeed in the future, for complex traits such as photosynthesis, much larger mapping panels will be necessary. In the world of human genetics recent studies using in excess of 250,000 individuals managed to identify loci at genome wide significance that collectively explained about 20% of the

[22] I ignore dominance effects as the results presented here come from homozygous lines.

additive genetic variance (Wood et al., 2014). Such massive studies are probably far beyond what will ever be financially feasible in the plant science community. However due to being able to conduct controlled experiments on homozygous individuals the numbers necessary to replicate similar results are probably much less. Nevertheless the pool of loci of sufficient effect size to be detected is likely to only account for a small proportion of the genetic variance present (Figure 5.10 and 5.11). Thus the rationale for using GWA as a gene finding study to describe the molecular basis of common variation in particulate detail becomes tenuous when effect sizes of alleles become increasingly small.

As the basis for investigating genetic architecture and initiating trait improvement strategies, GWA does have a role to play. Black box approaches such as genomic selection, where all markers are used to predict optimal combinations, are likely to lead to more rapid improvement of highly complex traits than characterization of the underlying variants. This is partially due to less stringent significance thresholds as there is less emphasis put on avoiding false positives and more focus is put on avoiding false negatives thus much more of the variation is likely to be harnessed for trait improvement. Based on the heritabilities estimated here (Figure 5.3), and its highly polygenic nature (Figures 5.10 and 5.11), it should be feasible to improve photosynthesis using such methods.

Another approach to identify the genetic basis of natural variation in photosynthesis could be to combine the construction mapping populations with trait selection. Generally, when choosing parents for a mapping population phenotypic extremes are chosen to maximize the potential variation assayed. If, however, genetically distinct plants from the same end of the phenotypic distribution are used instead, there is a greater chance of exploring more of the variation which is of interest, i.e. alleles leading to higher photosynthetic efficiencies. Positive transgressive segregation from a cross between two extreme lines, can be used to explore the phenotypic potential of the trait. Repeated screens of such F2 populations could be used to select transgressive segregants for use in a selection experiment. Such selection experiments can be used to explore the potential for phenotypic change. They often result in the exposure of cryptic genetic variation (Carlborg et al., 2006) which can then be selected upon, resulting in genetic assimilation of previously hidden phenotypes (Masel and Trotter, 2010). For most traits there is considerable room for improvement as it is highly unlikely that any trait in nature is ever truly optimized and even less likely that it is optimized to satisfy human (agricultural) needs. This has been shown experimentally in *Escherichia coli* where even after 50,000 generations in a stable environment there was a continued increase in fitness (Wiser et al.,

2013), and in maize where after 100 generations, oil and protein content continue to increase (Moose et al., 2004).

The results presented here represent the first GWA for photosynthesis in a wild plant species and indicate that photosynthesis is a highly polygenic trait with medium level heritabilities. The strongest association, *ACD6,* indicates that trade-offs, due to interaction with other traits i.e. disease resistance, may constrain its evolution in nature. These results taken together indicate the photosynthesis is a trait which is likely to show response to directional selection and is therefore amenable to improvement.

Acknowledgments

Thanks to Arthur Korte and Magnus Nordborg for seed of the Swedish RegMap population and to Hou-Sung Jung and Joanne Chory for seed of the *HSFA* double and triple mutants. To Rob van der Schoor for help with the phenotyper and to Gerrit Stunnenberg and Taede Stoker for help with maintaining the climate chamber in working order.

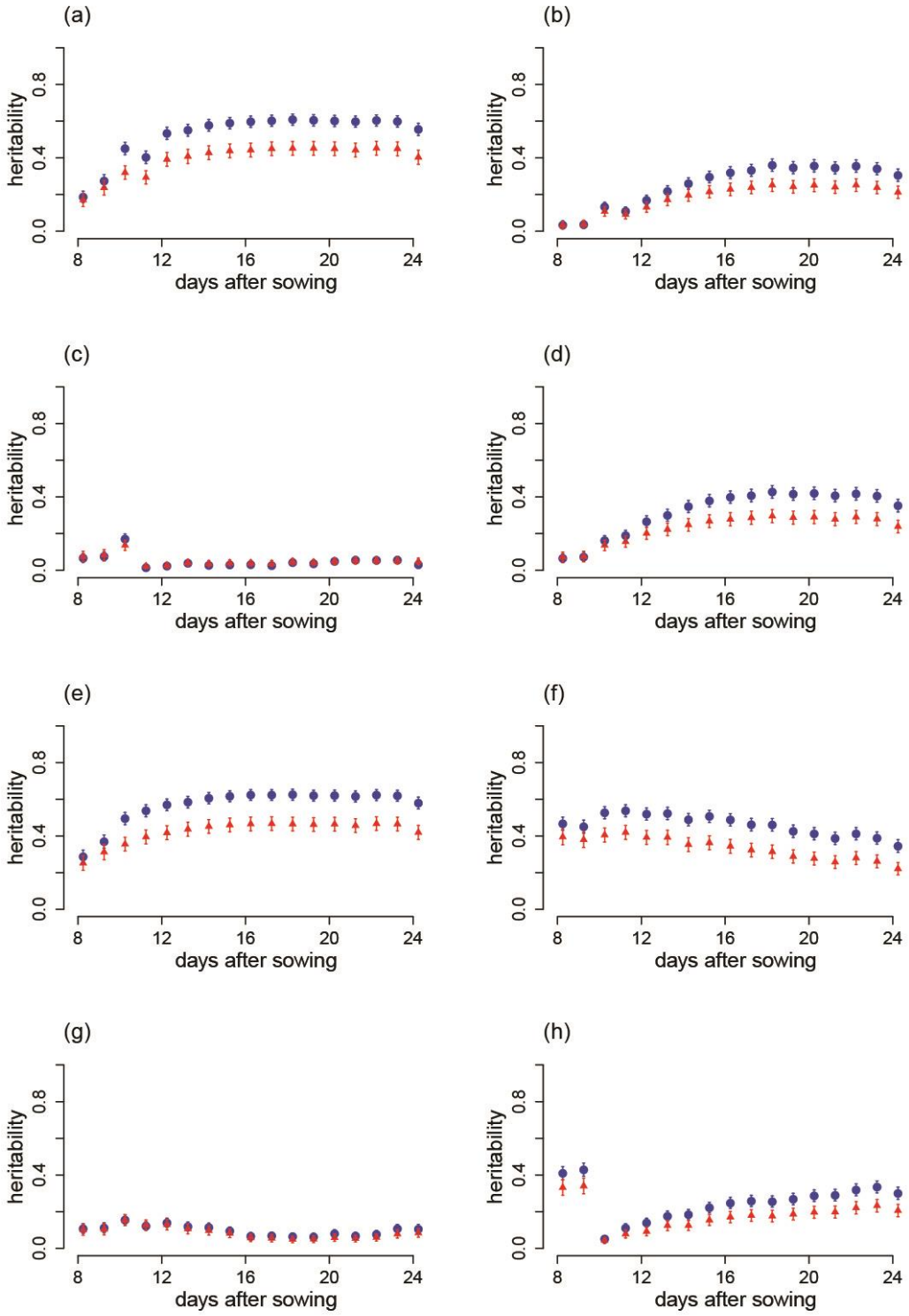
Supplemental data files, for files please email: flood@mpipz.mpg.de

Supplemental Data 5.1: List of all genotypes used, collection coordinates and aliases.

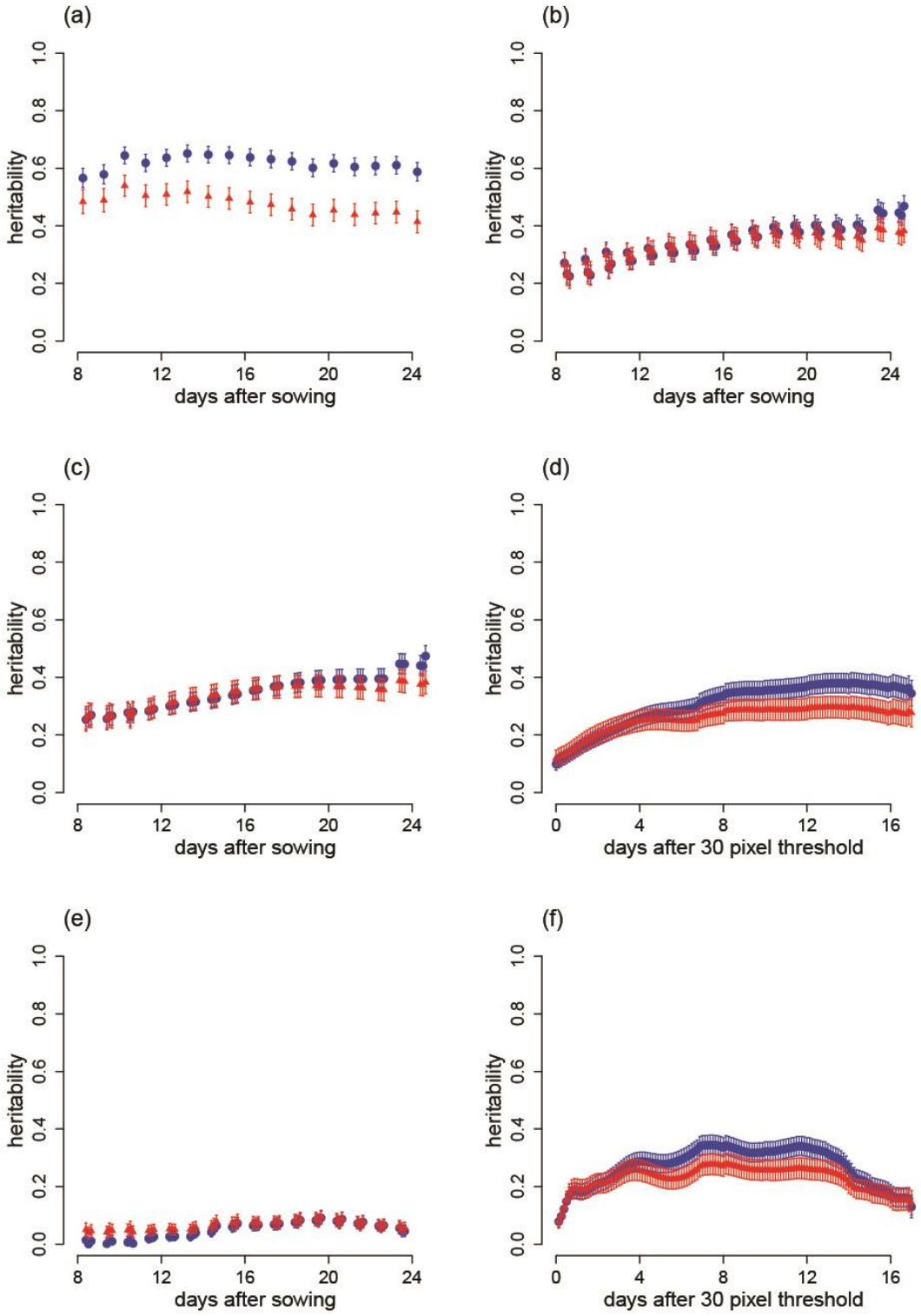
Supplemental Data 5.2: List of T-DNA insertion lines used in this study.

Supplemental Data 5.3: Phenotype data of all T-DNA insertion lines screened.

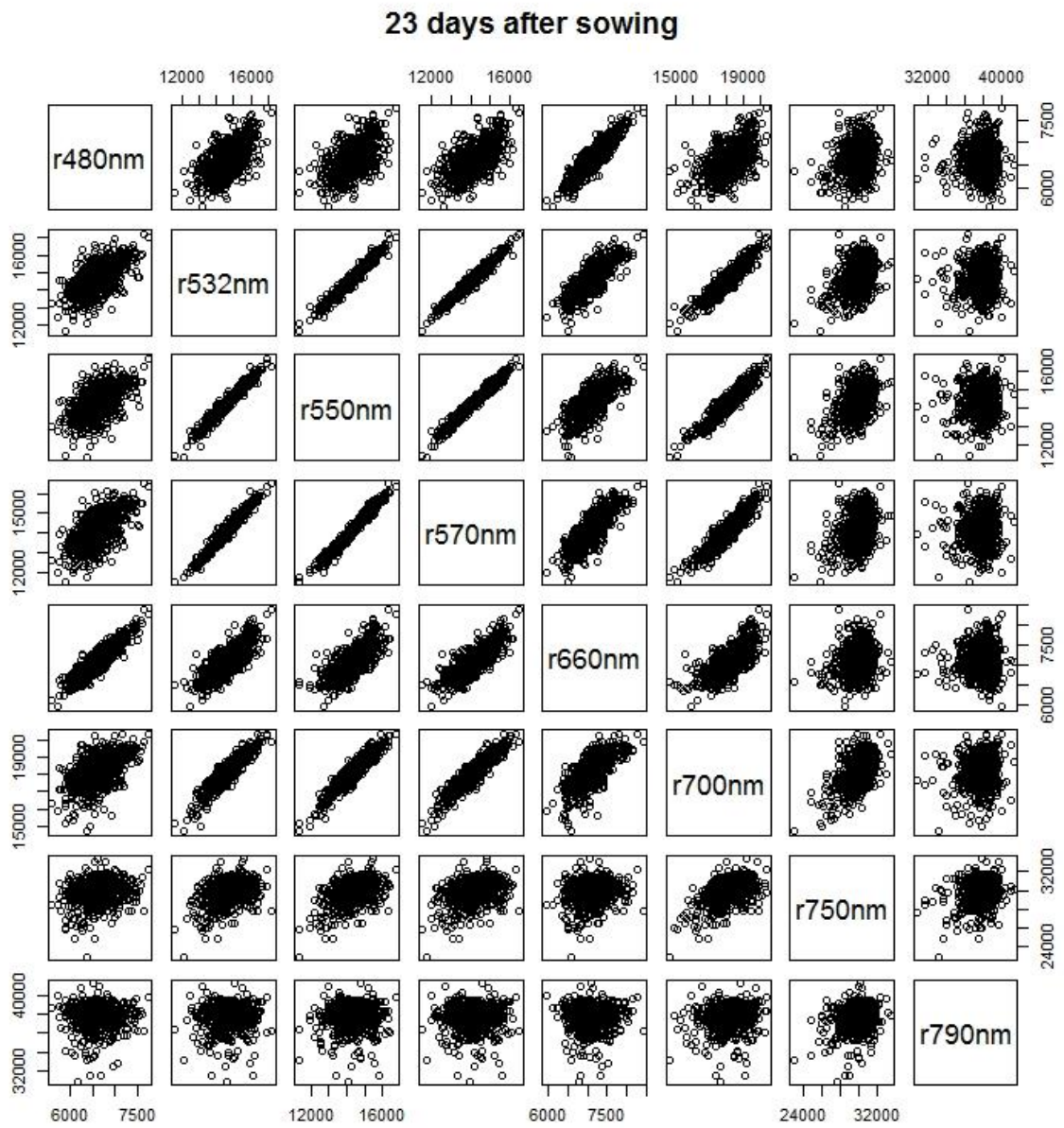
Supplemental Material



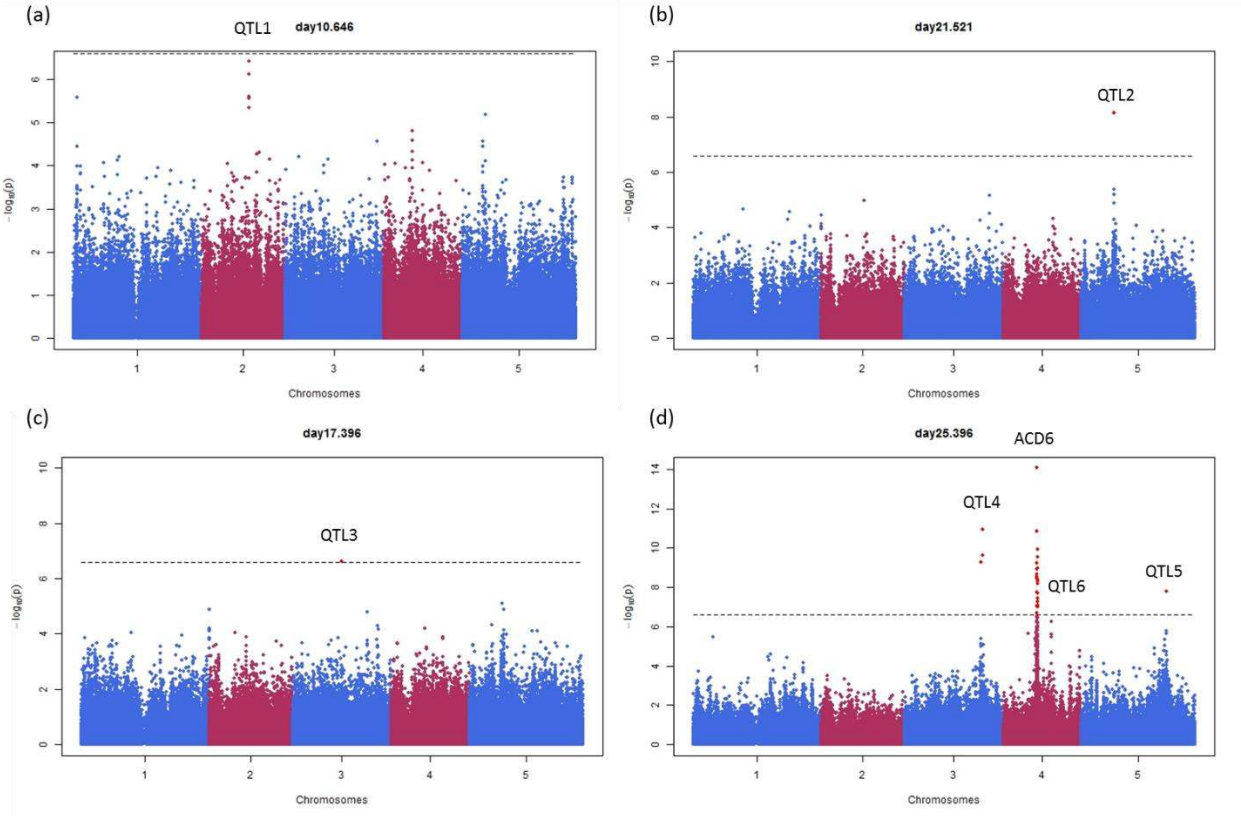
Supplemental Figure 5.1. *Comparison through time of broad sense heritability (blue) to narrow sense heritability (red) for reflectance at 480nm (a), 532nm (b), 550nm (c), 570nm (d), 660nm (e), 700nm (f), 750nm (g), and 790nm (h).*



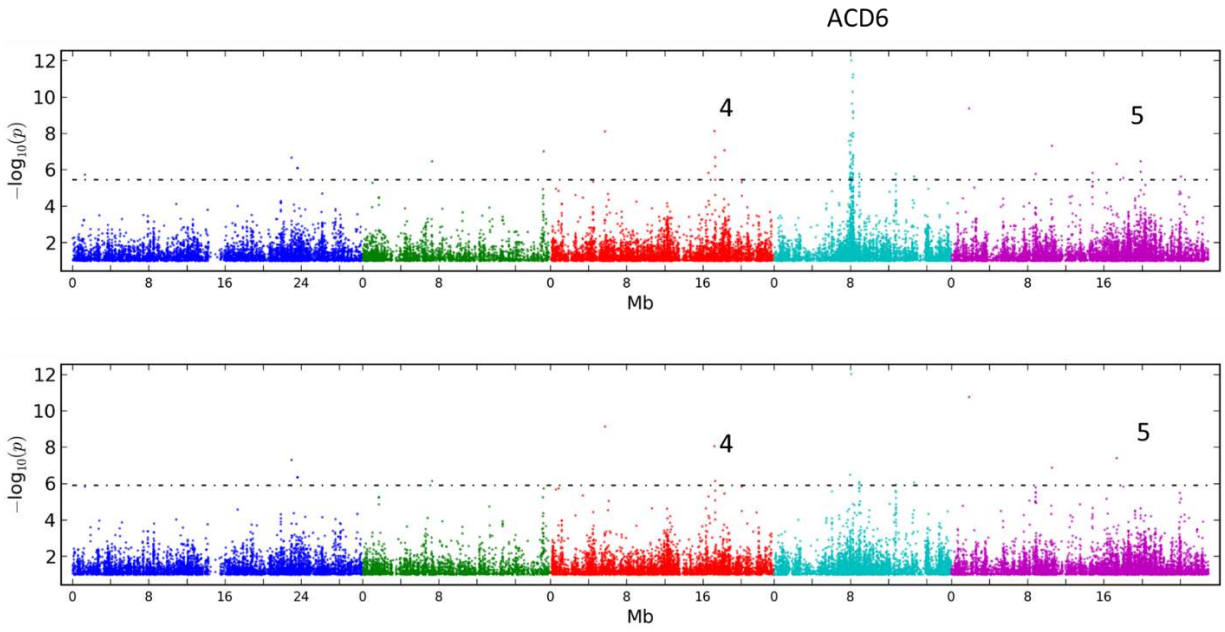
Supplemental Figure 5.2. *Comparison through time of broad sense heritability (blue) to narrow sense heritability (red) for chlorophyll reflectance (a), projected leaf area (PLA) (b), fitted PLA (c), fitted PLA where t=0 was set to a size threshold of 30 pixels (d), slope of growth curve estimated from fitted PLA values (e), slope of growth curve estimated from fitted PLA values where t=0 was set to a size threshold of 30 pixels (f).*



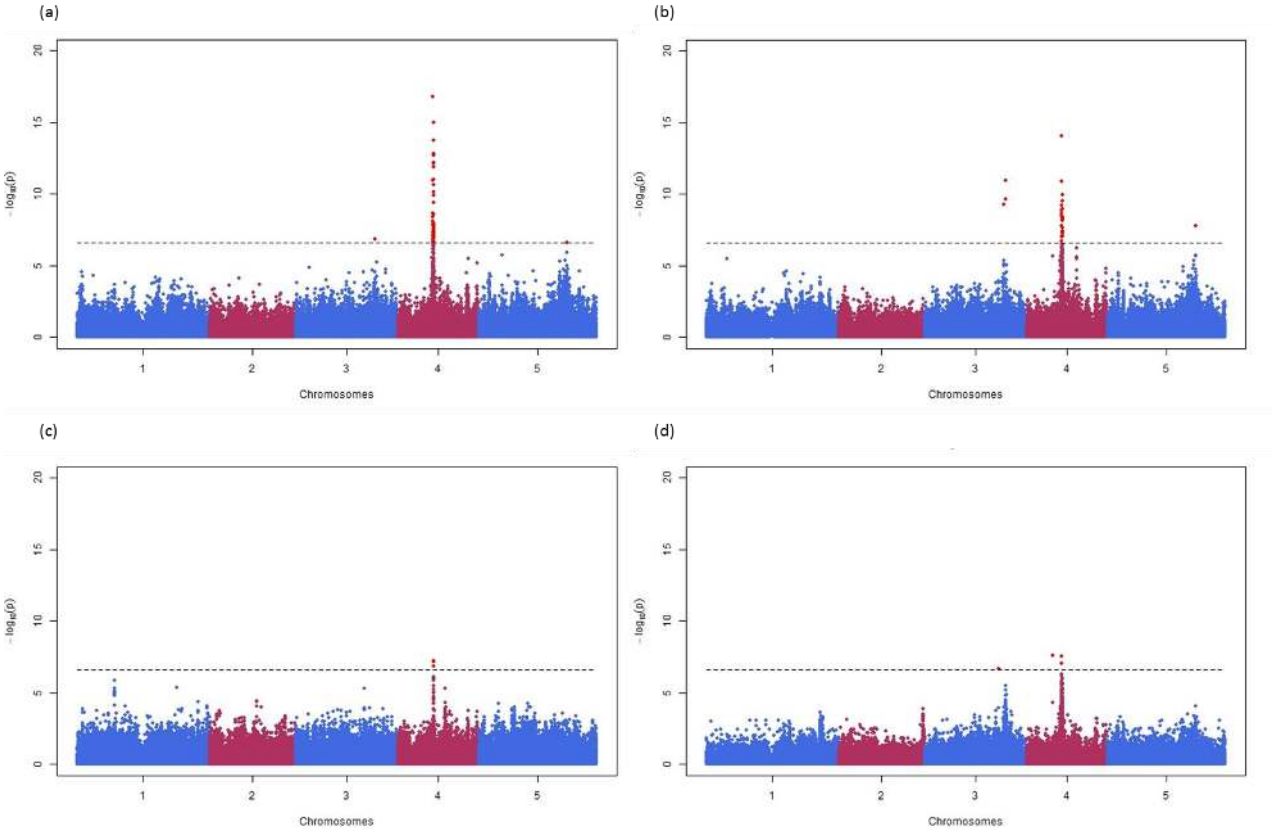
Supplemental Figure 5.3. *Correlation matrix of reflectance traits.*



Supplemental Figure 5.4. *Manhattan plots of additional QTL indicated in Figure 5.4. Genetic position is indicated on the x-axis and –Log(p)value is indicated on the y-axis, each marker is represented by a dot, the broken line indicates the Bonferroni significance threshold. Genotypes included in the analysis for each plot are as follows; all genotypes (a), HapMap genotypes (b) and (c), and Swedish RegMap (d). QTL are labeled and day of measurement is shown above each plot, MAF = 0.05.*



Supplemental Figure 5.5. *Results of accelerated mixed models analysis using GWAPP. Genetic position is indicated on the x-axis and –Log(p)value is indicated on the y-axis, each marker is represented by a dot, the broken line indicates the Bonferroni significance threshold. (a) Association for ΦPSII in the morning 25 DAS. (b) As in (a) with the most associated SNP at position 8094409 on chromosome 4 set as a cofactor. The numbers indicate the positions of QTL indicated on Figure 5.4, note the decline in significance when ACD6 is set as a cofactor. The MAF is set to 0 in this plot as GWAPP does not have MAF implemented for exporting plots but rather MAC which makes the output less comparable, see supplemental Figure 5.4 for comparison.*



Supplemental Figure 5.6. *Manhattan plots of Φ_{PSII} in the morning 25 DAS. Genetic position is indicated on the x-axis and –Log(p)value is indicated on the y-axis, each marker is represented by a dot, the broken line indicates the Bonferroni significance threshold. All genotypes (a), Swedish RegMap (b), HapMap (c), and northern Swedish genotypes (d).*

Supplemental Table 5.1. *List and TAIR annotation of the 23 gene containing significant SNPS in the significant locus on chromosome 4.*

Gene	Marker	Position in bp	*P*-value	Reference allele frequency	Effect size	Gene annotation
AT4G13730	m140440	7972249	7.56E-09	0.927152	0.002556	Ypt/Rab-GAP domain of gyp1p superfamily protein
AT4G13980	m140812	8076349	2.35E-09	0.849338	0.001868	HEAT SHOCK TRANSCRIPTION FACTOR A5
AT4G13990	m140843	8085573	8.54E-08	0.872517	0.001801	Exostosin family protein
AT4G13992	m140858	8089144	1.39E-08	0.740066	0.001414	Cysteine/Histidine-rich C1 domain family protein
AT4G14010	m140883	8093944	1.10E-11	0.811258	0.001982	RALF-LIKE 32
AT4G14020	m140886	8094409	1.69E-17	0.880795	0.003089	Rapid alkalinisation factor (RALF) family protein
AT4G14050	m140936	8104557	4.49E-08	0.769868	0.001442	Pentatricopeptide repeat (PPR) superfamily protein
AT4G14070	m140976	8115762	3.82E-09	0.832781	0.001771	ACYL-ACTIVATING ENZYME 15
AT4G14140	m141079	8150011	9.27E-12	0.877483	0.002462	DNA METHYLTRANSFERASE 2
AT4G14150	m141105	8160256	1.59E-13	0.889073	0.002832	PHRAGMOPLAST-ASSOCIATED KINESIN-RELATED PROTEIN 1
AT4G14160	m141119	8170580	1.70E-08	0.84106	0.001768	Sec23/Sec24 protein transport family protein
AT4G14165	m141133	8175186	1.15E-08	0.854305	0.001931	F-box family protein-related
AT4G14200	m141162	8186028	5.32E-08	0.839404	0.001721	Pentatricopeptide repeat (PPR) superfamily protein
AT4G14240	m141211	8207582	2.22E-07	0.692053	0.001247	Protein of unknown function DUF21
AT4G14340	m141299	8252182	1.86E-14	0.857616	0.002682	CASEIN KINASE I-LIKE 11
AT4G14342	m141306	8254521	1.01E-15	0.879139	0.002882	Splicing factor 3B subunit 5/RDS3 complex subunit 10
AT4G14368	m141380	8274507	6.41E-13	0.859272	0.002393	Regulator of chromosome condensation (RCC1) family protein
AT4G14370	m141418	8281977	1.21E-10	0.834437	0.00199	Disease resistance protein (TIR-NBS-LRR class) family
AT4G14400	m141496	8297892	1.38E-08	0.817881	0.001785	ACCELERATED CELL DEATH 6
AT4G14410	m141513	8301059	7.02E-13	0.716887	0.001784	BASIC HELIX-LOOP-HELIX 104
AT4G14420	m141532	8303999	2.16E-13	0.846026	0.002322	HR-like lesion-inducing protein-related
AT4G14440	m141541	8306682	2.61E-09	0.865894	0.001977	3-HYDROXYACYL-COA DEHYDRATASE 1
AT4G14530	m141623	8343754	1.05E-07	0.879139	0.001874	AGAMOUS-like 97

Chapter 6

Natural variation in phosphorylation of photosystem II proteins in *Arabidopsis thaliana* – is it caused by genetic variation in the STN kinases?

Pádraic J. Flood[1,2,*], Lan Yin[3,*], Andrei Herdean[3], Jeremy Harbinson[2], Mark G.M. Aarts[1] and Cornelia Spetea[3]

[1]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

[2]*Horticulture and Production Physiology, Wageningen University, Wageningen, the Netherlands.*

[3]*Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden.*

* These authors equally contributed to this work

Corresponding Author: Cornelia Spetea (cornelia.spetea.wiklund@bioenv.gu.se)

Abstract

Reversible phosphorylation of photosystem II (PSII) proteins is an important regulatory mechanism that can protect plants from changes in ambient light intensity and quality. We hypothesized that there is natural variation in this process in *Arabidopsis* (*Arabidopsis thaliana*), and that this results from genetic variation in the *STN7* and *STN8* kinase genes. To test this, *Arabidopsis* accessions of diverse geographic origins were exposed to two light regimes, and the levels of phospho-D1 and phospho-LHCII proteins were quantified by western blotting with anti-phosphothreonine antibodies. Accessions were classified as having high, moderate or low phosphorylation relative to Col-0. This variation could not be explained by the abundance of the substrates in thylakoid membranes. In genotypes with atrazine-resistant forms of the D1 protein, low D1 and LHCII protein phosphorylation was observed, which may be due to low PSII efficiency resulting in reduced activation of the STN kinases. In the remaining genotypes, phospho-D1 levels correlated with STN8 protein abundance in high light conditions. In growth light, D1 and LHCII phosphorylation correlated with longitude and in the case of LHCII phosphorylation also with temperature variability. This suggests a possible role of natural variation in PSII protein phosphorylation in the adaptation of *Arabidopsis* to diverse environments.

Keywords: *Arabidopsis thaliana*; natural variation; phosphorylation; photosystem II; STN kinase; temperature seasonality.

Introduction

Plants require light energy to drive photosynthesis. Their photosynthetic machinery is profoundly affected by changes in irradiance, which can be in both the intensity and in the spectral quality of light, and can occur across a range of temporal scales, from seconds (light flecks) to months (seasonal changes). Plants have developed diverse response mechanisms to adjust and protect their photosynthetic machinery in the face of such fluctuations (for recent review, see (Rochaix, 2013)). Short-term responses involve a dynamic reorganization of photosynthetic complexes, whereas long-term responses involve changes in the chloroplast and nuclear gene expression resulting in altered levels of the photosynthetic machinery to optimize and sustain photosynthesis. Both types of responses are induced by changes in the redox state of the photosynthetic electron transport chain, and are mediated through a complex network of reactions involving protein kinases and phosphatases (Rochaix, 2013). Plants which have a compromised ability to respond to both short and long term light fluctuations show fitness costs in nature (Külheim et al., 2002; Frenkel et al., 2007). In this study we aimed to quantify the natural genetic variation in these responses, both to document the extent of this variation, and to gain some initial insights into the selective forces, which may be acting on these processes.

Photosystem II (PSII) and PSI are connected in series through the electron transport chain, which includes the plastoquinone (PQ) pool, the cytochrome b_6f ($Cytb_6f$) complex and plastocyanin. Changes in the light environment may lead to more reduction or oxidation of the PQ pool initiating signaling processes that drive changes in the organization and composition of the photosynthetic machinery. The process of state transitions is used by algae and also plants, to correct the redox state of the PQ pool. In case of a reduced PQ pool, plastoquinol (PQH_2) docks to the Q_o site of $Cytb_6f$ (Vener et al., 1997). This event leads to the activation of a protein kinase that phosphorylates several proteins of the light-harvesting complex II (LHCII). Upon phosphorylation, the mobile part of LHCII is displaced from PSII to PSI, thus re-equilibrating the cross-sections of the antennae of PSII and PSI and their respective light excitation. The process is reversible, as over-excitation of PSI causes the oxidation of the PQ pool, deactivation of the kinase, dephosphorylation of LHCII proteins by a phosphatase and the return of LHCII to PSII. Using molecular genetic approaches the LHCII kinase was identified in *Arabidopsis* and named STN7 (Bellafiore et al., 2005).

Of the two photosystems, PSII is more susceptible to photoinactivation, and undergoes a repair cycle to replace its reaction centre D1 protein (Aro et al., 2005). In the plant thylakoid

membrane, PSII is mostly present as PSII-LHCII dimeric supercomplexes located in the appressed (grana) membranes. However, the repair of photoinactivated PSII complexes and the assembly of new ones occur through the monomeric form of PSII in the nonappressed (stroma) thylakoid membranes. PSII core protein phosphorylation in general, and D1 phosphorylation in particular, has been suggested to facilitate disassembly of photoinactivated PSII complexes and is thought to play a role in the regulation of PSII repair (Tikkanen et al., 2008). The kinase involved in PSII core protein phosphorylation was identified in *Arabidopsis* and named STN8 (Bonardi et al., 2005; Vainonen et al., 2005).

To elucidate the substrate specificity of STN7 and STN8 kinases, thylakoid protein phosphorylation patterns of wild-type *Arabidopsis* plants, *stn7* and *stn8* mutant lines have been monitored by different approaches, including Western blot analyses with different anti-phosphothreonine antibodies, and mass spectrometric analyses. The STN7 kinase is involved in phosphorylation of LHCII, CP29, CP26 and TSP9 proteins, whereas the STN8 kinase phosphorylates PSII core D1, D2, PsbH and to some extent CP43 proteins (for reviews, see (Rochaix, 2007; Pesaresi et al., 2011)). Besides the PSII core proteins, STN8 has additional targets, including the chloroplast calcium-sensing protein CAS (Vainonen et al., 2008), and a protein involved in cyclic electron transport (PGRL1) (Reiland et al., 2011), which is a part of the PSI-LHCI-LHCII supercomplex (DalCorso et al., 2008). Analysis of the protein phosphorylation profiles of the *stn7* and *stn8* mutants of *Arabidopsis* showed residual phosphorylation of the LHCII and PSII core proteins, respectively. However, this residual phosphorylation was undetectable in the *stn7* × *stn8* double mutant, indicating some substrate overlap between the STN7 and STN8 kinases (Vainonen et al., 2005; Fristedt et al., 2009).

The STN7 kinase appears to have a broader role than state transitions, and is also required for adaptation to light fluctuations (Tikkanen et al., 2010). For example, by subjecting *Arabidopsis* plants to alternative periods of low light and high light (HL), LHCII is phosphorylated during the low light and dephosphorylated during the high-light periods (Rintamaki et al., 1997). The loss of STN7 in plants subjected to this fluctuating light regime leads to a severe decrease in growth, indicating that STN7 has an important role in response to environmental changes (Bellafiore et al., 2005; Tikkanen et al., 2006). The loss of STN8 resulted in slower growth in rice (Nath et al., 2013) but not in *Arabidopsis* (Bonardi et al., 2005). In both species, the mutation leads to increased susceptibility of PSII to high light due to suppressed mobility of inactivated complexes during repair. A high level of PSII core protein phosphorylation is required for the adjustment of macroscopic folding of the thylakoid

membrane, which modulates protein mobility in this membrane (Fristedt et al., 2009). Significant enhancement in the thylakoid grana size in the *stn8* mutant slows down the movement of PSII from the grana to the stromal region during repair cycle, indicating that PSII core protein phosphorylation is involved in this process. Since the *stn8* mutant also displays reduced cyclic electron transport, the possibility has been raised that STN8 kinase activity may be important for fine-tuning of the photosynthetic machinery to fulfill the NADPH/ATP demands of chloroplast metabolism (Reiland et al., 2011).

Following the identification of the two kinases, their substrates and functions, an important remaining question concerns their mode of regulation. Previously, it was thought that light activates LHCII as a phosphorylation substrate by increasing the exposure of its N-terminal domain, containing the phosphorylation site, to the enzyme (Zer et al., 1999). Another report provided evidence in support of a light-induced exposure of the phosphorylation site of the CP43 subunit of the PSII core complex (Vink et al., 2000). The activation of the STN7 kinase was proposed to be strictly regulated by the redox state of PQ and the $Cytb_6f$ complex (Rintamaki et al., 2000; Bellafiore et al., 2005). In support of this, following a transfer from darkness to growth light (GL), an increase in the amount of the STN7 kinase was reported (Yin et al., 2012). The protein abundance of STN7 is regulated in a post-translational manner involving proteolysis and autophosphorylation (Willig et al., 2011). Most recently, it was shown that accumulation of the STN7 protein is controlled at the transcript level, in a light- and redox-dependent manner (Wunder et al., 2013). While low-light conditions increase STN7 kinase activity *in vivo*, HL levels bring about its inhibition. This inhibition appears to be in response to the increasing degree of reduction of the stroma and to be mediated by thioredoxin. This possible thioredoxin-linked inactivation of the kinase depends on, and is therefore subordinated to, the prior activation of the kinase by the redox state of PQ and the $Cytb_6f$ complex (Rintamaki et al., 2000).

Compared to what is known about the regulation of STN7, much less is known about the determinants of STN8 activity. Upon transfer from the dark to GL, an increase in D1 phosphorylation was reported (Yin et al., 2012). STN8 is even more active under HL conditions as the further increase in D1 phosphorylation in HL shows. The increased activity correlated with an increase of STN8 upon transfer from darkness to HL (Yin et al., 2012). Another report found no change in STN8 level upon transfer from low to high light conditions despite increase in D1 protein phosphorylation (Wunder et al., 2013), indicating that the regulation of kinase activity in response to changes in light intensity may be more important than its changing

abundance. No interdependence seems to exist between the STN kinases at protein levels (Wunder et al., 2013). Nevertheless, it has been proposed that the activity of STN8 may be regulated by STN7 in *Arabidopsis*, since in the green alga *Chlamydomonas reinhardtii* the activity of STN8 homologue Stl1 is regulated by its phosphorylation in an Stt7-dependent manner (Lemeille et al., 2010).

As outlined above, both the STN7 and STN8 kinases play essential roles in the response to changes in ambient light, by influencing LHCII distribution between PSII and PSI and facilitating protein repair, respectively. A recent review compares such responses across a wide phylogenetic spectrum (Grouneva et al., 2013), however very little is known about the intraspecific variation in these processes (Flood et al., 2011). Such within-species variation drives natural selection and may represent different adaptive strategies to photosynthetic regulation within a species. *Arabidopsis* has a wide geographic distribution, ranging from Tanzania to Norway and from Portugal to Japan. As such it occupies a wide range of light environments it may have adopted different strategies to cope with this environmental diversity. Here we screen 16 *Arabidopsis* genotypes comprised of 13 diverse accessions from throughout the natural range of *Arabidopsis*, as well as a backcross, the *stn8* mutant and a hybrid between two of these accessions. We have analyzed thylakoid protein extracts by western blotting with anti-phosphothreonine antibodies and we used the levels in the standard accession Col-0 as reference. Furthermore, we have investigated whether variation in the protein or transcript level of the respective kinase, as well as the geographic and climatic origin of the genotype correlate with the observed variation in PSII protein phosphorylation.

Material and Methods

(a) *Plant growth and light treatment*

Arabidopsis thaliana plants were grown for 34 days in a chamber at 100 µmol photons m^{-2} s^{-1} (GL) using a 10/14hr day/night cycle. After 14 h of darkness, plants were exposed for 3 h to GL and subsequently transferred for 3 h to high light (HL, 600 µmol photons m^{-2} s^{-1}). Leaf material corresponding to 2-3 g bulked from 5-6 plants was harvested, frozen immediately in liquid N_2 and stored at -80° C until thylakoid isolation.

A total of 16 *Arabidopsis* genotypes were included in this study (Table 6.1). We used 13 naturally occurring accessions from a range of geographic locations. In addition, a previously characterized *stn8* mutant line lacking a functional STN8 kinase (Fristedt et al.,

2009) was used. One F1 hybrid was included (Tsu-0 × Ws-4) to test for the presence of a dominant phenotype and because pilot experiments identified both parental accessions as extremes. The atrazine-resistant Ely accession was included because it has compromised PSII functioning due to a mutation in the chloroplast-encoded D1 protein in the Q_B-binding pocket (Ser264Gly) (El-Lithy et al., 2005). Cytoplasm from the Ely genotype was introduced into the L*er*-1 nuclear background by six rounds of back crossing (genotype (Ely × L*er*) × L*er* BC6 (ELB)). In addition to altered PSII efficiency, ELB allowed us some preliminary insights into the role of cytoplasmic variation versus nuclear variation on PSII protein phosphorylation.

Table 1. *The names, stock numbers, abbreviations, geographic origin and habitat of all genotypes used in this study (source: TAIR, www.arabidopsis.org). N/A, not applicable.*

Stock Number	Accession name	Abbreviated name	Country	Latitude (°)	Longitude (°)
CS76113	Columbia-0	Col-0	N/A	N/A	N/A
SALK060869	*stn8-1* (in Col-0)	*stn8*	N/A	N/A	N/A
CS76227	Shakdara	Sha	Tadjikistan	38.35	68.48
CS76106	C24	C24	Portugal	41.25	-8.45
CS76105	Burren-0	Bur-0	Ireland	52.9	-9
CS28595	Palermo	Pa-2	Italy	38.07	13.22
CS76192	Martuba	Mt-0	Libya	32.34	22.46
CS76210	Perm	Per-1	Russia	58.00	56.31
CS76100	Borky	Bor-4	Czech	49.40	16.23
CS76109	Canary Island	Can-0	Spain	29.21	-13.48
CS28780	Tsushima	Tsu-0	Japan	34.43	136.31
N/A	Tsu-0 × Ws-4	Tsu × Ws-4	N/A	N/A	N/A
CS5390	Wassilewskija-4	Ws-4	Belarus	52.3	30
CS76164	Landsberg *erecta*	L*er*-1	Poland	52.71	15.23
N/A	((Ely × L*er*) × L*er*)BC6	ELB	N/A	N/A	N/A
CS28631	PHW-31 (Ely)	Ely	England	52.39	0.26

(b) *Thylakoid isolation and protein analysis*

Thylakoid membranes were isolated from frozen leaves as previously described (Noren et al., 1999) with the modification that 10 mM NaF (a general phosphatase inhibitor) was included in all isolation buffers. Chlorophyll (Chl) was extracted in 80% (v/v) acetone and the concentration was determined according to (Porra et al., 1989). Thylakoid proteins were separated by SDS-PAGE in 14% (w/v) acrylamide gels containing 6 M urea followed by

electrotransfer and immunoblotting with various antibodies. The following antibodies were used: anti-D1, anti-Lhcb2, anti-STN8 and anti-STN7 from Agrisera (Umeå, Sweden), anti-phosphothreonine antibodies from Cell Signalling (New England BioLabs, U.K.) and Zymed (Invitrogen, Carlsbad, USA), and anti-CP43 from our laboratory. Western blots were analysed using a Fusion FX-7 imager (Vilbert Lourmat, France) and quantified using Multi Gauge software. Col-0 was used as internal standard, making it possible to compare different Western blots.

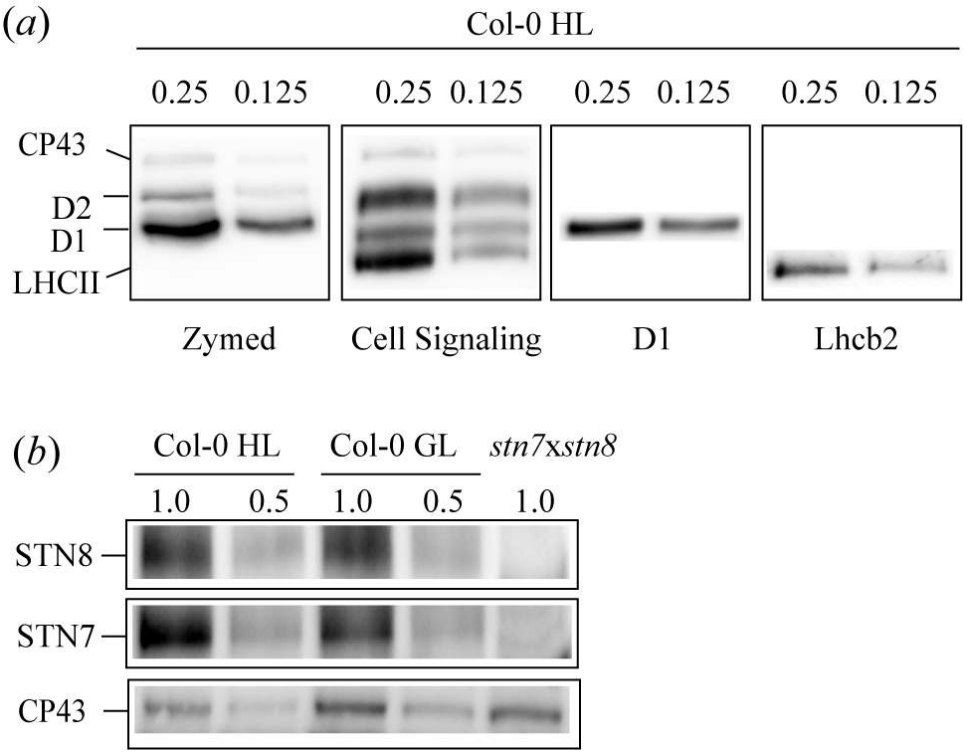


Figure 6.1. *Representative western blots of thylakoid proteins isolated from Col-0. The plants were illuminated for 3 h with growth light (GL, 150 µmol m^{-2} s^{-1}) and subsequently transferred for 3 h to high light (HL, 600 µmol m^{-2} s^{-1}). Thylakoid membranes were isolated in the presence of NaF and the proteins were separated by gel electrophoresis. (a) The phosphorylated PSII proteins were immunodetected with anti-phospho-Thr antibodies from Zymed and Cell Signalling. Control blots with anti-D1 and Lhcb2 antibodies are also shown. (b) The levels of STN8 and STN7 protein kinases are shown in parallel with control CP43 blots. Thylakoids isolated from the stn7×stn8 double mutant were used as a control for specificity of the anti-STN8 and STN7 antibodies. Two chlorophyll loadings (µg/lane) are shown to demonstrate the linearity of the immunodetected signal from each antibody used.*

As shown by representative blots for Col-0, both anti-phospho-Thr antibodies recognized PSII phosphoproteins, but with different affinities (Figure 6.1*a*). The Zymed antibody recognized the phospho-D1 protein best, whereas the Cell Signaling antibody was found most suitable for quantification of phospho-LHCII proteins. Two Chl loadings are shown to indicate the linearity of the immunodetected signal. Any possible variation in phosphorylation could be caused by variation in the amount of substrate, kinase involved or

other factors. Western blots with anti-D1 and anti-Lhcb2 antibodies indicated the level of corresponding substrate for the phosphorylation reaction (Figure 6.1*a*). The levels of the STN8 and STN7 kinases were assessed in samples from thylakoid preparations using specific antibodies, and control western blots with CP43 protein were used to correct the amount of protein loaded (Figure 6.1*b*). The *stn7* × *stn8* double mutant (Fristedt et al., 2009) was used to verify the identity of the corresponding cross-reacting bands with the STN8 and STN7 antibodies. Two Chl loadings are shown to indicate the linearity of the immunodetected signal for the three antibodies.

The selected accessions were analyzed by the assay described above using conditions optimized for Col-0. The levels of various immunodetected proteins were determined and expressed relative to those in Col-0 (Supplementary Tables S6.1 and S6.2). The genotypes were classified as displaying high (80-120%), moderate (40-80%) or low (<40%) levels of immunodetected proteins.

(c) ***RNA isolation and transcript analysis***

RNA was isolated from frozen leaves as described in (Onate-Sanchez and Vicente-Carbajosa, 2008). One fully expanded leaf was taken from three plants after 3 h of GL or HL treatment. The RNA concentration was measured using a NanoDrop 2000 and the volume adjusted with Millipore water to obtain a final concentration of 0.25 µg/µl. 1 µg of RNA was used for cDNA synthesis. Equal volumes of cDNA were used in all subsequent qPCRs. Eight reference genes (Czechowski et al., 2005; Dekkers et al., 2012) were tested on all samples (Supplementary Table S6.3) and seven of them were found stable enough for use in further analysis. Primer sequences are listed in Supplementary Table S6.3. A normalisation factor was calculated from the seven reference genes, which was used to calculate relative transcription levels.

(d) ***Statistical analysis***

Scatterplots were created in GraphPad Software (La Jola, CA). Best-fit lines were applied and the correlation coefficient (r^2) and its significance (two-tailed p value) were calculated with the same software. Correlations were considered significant at $p \leq 0.05$.

Results

(a) *D1 protein phosphorylation in growth and high light conditions*

Large differences in D1 protein phosphorylation were observed in GL in the 16 genotypes ranging between 4-100% of the phosphorylation level found in Col-0 (Figure 6.2*a*). D1 protein level in the studied accessions ranged between 70 and 120% of Col-0 (Supplementary Table S6.1), and cannot explain the large variation observed in phosphorylation. Notably, the relative level of the STN8 kinase varied largely among accessions (78-192%) (Figure 6.2*a*). Col-0 had the highest phosphorylation level, which was approx. 20% higher than that of the next highest genotype (Tsu-0; 79%, Supplementary Table S6.1). Both genotypes also displayed high levels of the STN8 kinase. Four genotypes (C24, Per-1, Ws-4 and Tsu × Ws-4) displayed moderate levels of phosphorylation but high levels of STN8 kinase, Five accessions displayed low D1 protein phosphorylation (Bur-0, Pa-2, Mt-0, Bor-4, Can-0, and Ely) although had high STN8 levels. Sha, L*er*-1, and ELB had very high STN8 levels (>120% of Col-0) but low D1 phosphorylation. Thus, there was no correlation between the STN8 protein level and the level of D1 phosphorylation, though *stn8,* which had the lowest level of phospho-D1, lacked any STN8 (Figure 6.2*b*). The Tsu × Ws-4 hybrid and the ELB backcross were moderate or lower than their parents in the D1 phosphorylation levels, despite having high(er) levels of the STN8 kinase. These data indicate that there must be another limiting/regulating factor for D1 protein phosphorylation under GL conditions than the substrate or the kinase levels.

D1 phosphorylation in HL ranged between 2-117% of Col-0 (Figure 6.2*b*). As in GL, there was little variation in D1 protein level (87-114%), whereas STN8 abundance showed greater variation (32-121%) among accessions (Supplementary Table S6.2). Under HL conditions, there was a much tighter grouping of genotypes than in GL with the exception of *stn8*, Ely and ELB showing low D1 phosphorylation (compare Figure 6.2*a* to Figure 6.2*b*). There was a general upward trend in D1 phosphorylation, and Col-0, whilst still high, was no longer an outlier. Six accessions displayed high phosphorylation levels and also high STN8 levels (Col-0, Bur-0, Mt-0, Per-1, Bor-4 and Tsu × Ws-4) (Supplementary Table S6.2 and Figure 6.2*b*). Other three displayed moderate phosphorylation levels and also moderate STN8 levels (Sha, C24 and Ws-4). The *stn8* mutant displayed residual D1 phosphorylation. The remaining genotypes displayed either high levels of phospho-D1 despite moderate STN8 kinase levels (Pa-2, Tsu-0 and L*er*-1) or moderate levels of phospho-D1 and low kinase levels (Can-0). Ely and ELB were found low phosphorylation accessions despite high STN8 protein levels,

suggesting that kinase abundance was not limiting. No significant correlation was obtained between phospho-D1 level and STN8 abundance if all accessions were included. Nevertheless, a weak but significant correlation was obtained if Ely and ELB were excluded from the analysis (Figure 6.2*b*).

When comparing GL to HL conditions, the phosphorylation of the D1 protein increased by 70%, whereas the STN8 level remained stable in Col-0 (Figure S6.1 *a, b*). The other accessions also displayed increased levels of phospho-D1, but the level of the STN8 protein either decreased (ELB, Pa-2, L*er*-1, Can-0, Tsu-0, Ws-4 and Tsu × Ws-4), remained quite stable (C24, Bur-0, Per-1, Bor-4 and Ely) or even increased (Mt-0) upon transfer from GL to HL. All genotypes showed reduced transcription of *STN8* in HL with the exception of ELB, C24, Bur-0, Pa-2, Sha and Mt-0, which either showed an increase or no difference in transcription (Figure S6.1*c*).

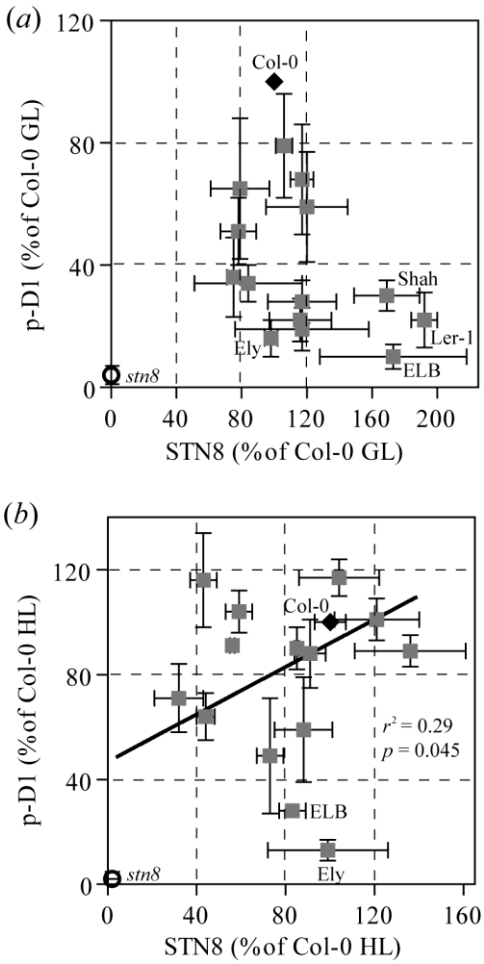


Figure 6.2. *Scatterplots comparing the levels of phospho-D1 (p-D1) relative to STN8 protein level under growth light (GL) (a) and high light conditions (HL) (b). The plotted data are expressed relative to Col-0 and are means of 2-3 technical replicates ±SD. ◆ represents Col-0, ○ represents the stn8 mutant, and ■ all other genotypes. Some extreme genotypes are also labelled in panel (a) with their names. No significant correlation was found in (a). A weak but significant correlation was found in (b) if Ely and ELB were excluded from the regression analysis. Broken lines delimit high, moderate and low D1 phosphorylation and STN8 protein levels.*

(b) ***LHCII protein phosphorylation in growth and high light conditions***

Phosphorylation of LHCII proteins in GL varied between 7-102 % of Col-0 (Figure 6.3*a*). With the exception of Ely and ELB, the level of Lhcb2 showed limited variation, whereas STN7 protein levels varied between 70-213% of Col-0 (Supplementary Table S6.1 and Figure 6.3*a*). There were seven high phosphorylation genotypes (Col-0, *stn8*, Sha, Mt-0, Per-1, Tsu-0, Tsu × Ws-4), seven moderate (C24, Bur-0, Pa-2, Bor-4, Can-0, Ws-4 and L*er*-1), and two low accessions (Ely and ELB). All high phosphorylation accessions displayed high levels of STN7 kinase, including Col-0 and the *stn8* mutant. Accessions displaying moderate levels of LHCII phosphorylation accumulated STN7 at either high or very high (>120%) levels, indicating that other factors limited the kinase activity. Tsu × Ws-4 displayed similar high phospho-LHCII as Tsu-0 although it had much higher levels of STN7 (213 versus 70%). One striking observation was that both genotypes with the atrazine-resistant cytoplasm, Ely and ELB, showed Lhcb2 levels of 143 and 126%, respectively, relative to Col-0, and yet showed extremely low levels of LHCII phosphorylation. The reduced phosphorylation levels in these two genotypes cannot be explained by the abundance of the kinase, which was 71 and 163% relative to Col-0, and indicate that the STN7 kinase was not properly activated.

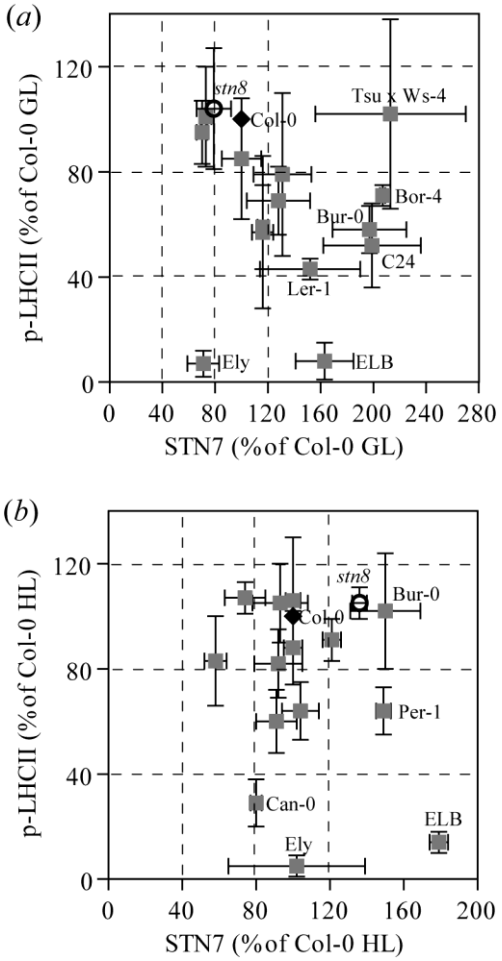


Figure 6.3. *Scatterplots comparing the levels of phospho-LHCII (p-LHCII) relative to STN7 protein level under growth light (GL) (a) and high light conditions (HL) (b). The plotted data are expressed relative to Col-0 and are means of 2-3 technical replicates ±SD. ◆ represents Col-0, ○ represents the stn8 mutant, and ■ all other genotypes. Some extreme genotypes are also labelled with their names. No significant correlation was found between LHCII phosphorylation and STN7 kinase abundance. Broken lines delimit high, moderate and low LHCII phosphorylation and STN7 protein levels.*

Phosphorylation of LHCII proteins in HL varied between 5-107% relative to Col-0, whereas Lhcb2 and STN7 protein levels varied between 89-123% and 58-179% relative to Col-0, respectively. Col-0, Sha, C24, Bur-0, Pa-2, Mt-0, Tsu-0, Tsu × Ws-4 and L*er*-1 were high phosphorylation accessions, Per-1, Bor-4 and Ws-4 were moderate, and Can-0, Ely and ELB were low phosphorylation accessions (Supplementary Table S6.2). As in GL conditions, the level of kinase did not appear to correlate with the level of LHCII phosphorylation (Figure 6.3*b*). Genotypes with high phosphorylation, including Col-0 and the *stn8* kinase, displayed high to very high levels of STN7 kinase. Also the moderate and the low phosphorylation accessions displayed high STN7 levels, indicating that factors related to activation rather than the amount of kinase may limit the phosphorylation reaction.

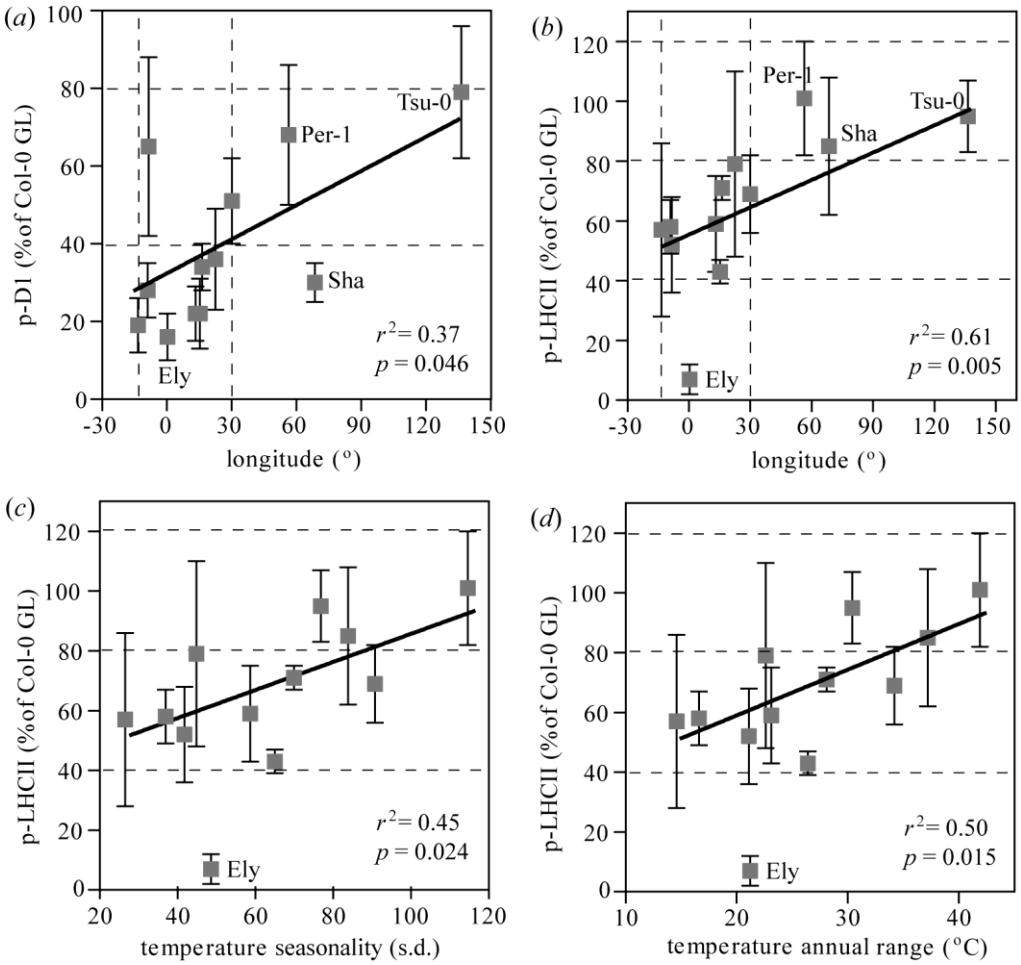


Figure 6.4. *Scatterplots comparing the levels of phospho-D1 (p-D1) (a) and phospho-LHCII (p-LHCII) (b-d) in growth light (GL) relative to geographic (a, b) and climatic factors (c, d). A significant correlation with longitude was found for both p-D1 (a) and p-LHCII levels (b). Vertical broken lines delimit longitude for European accessions (-14° to 30°). Per-1, Sha and Tsu-0 are labelled because they are non-European accessions. A significant correlation was found between the levels of p-LHCII and the temperature seasonality (c) and the temperature annual range (d). Correlation of p-D1 with the parameters in (c) and (d) was low but a trend was visible however not significant (r^2 = 0.23, p = 0.128 and r^2 = 0.24, p = 0.123, respectively). Temperature data were obtained from the WorldClim database (http://www.worldclim.org/). The phosphorylation data in all panels are expressed relative to Col-0 and are means of 2-3 technical replicates ±SD. Ely is labelled because it was excluded from regression analysis in all panels. Horizontal broken lines delimit high, moderate and low phosphorylation levels in all panels.*

When comparing GL to HL, LHCII phosphorylation decreased by 30% in Col-0, whereas STN7 abundance remained stable (Figure S6.2*a*, *b*). With few exceptions (Ler-1, C24 and Bur-0), phospho-LHCII also decreased in the other accessions, whereas STN7 abundance varied between the two light regimes. The *STN7* transcript level in GL was comparable among genotypes (Figure S6.2*c*). Upon transfer to HL, the *STN7* transcript abundance decreased in all genotypes except ELB and Mt-0. The extent of reduction in transcript abundance varied

considerably, with Col-0, C24, Per-1 and Ely showing a much more pronounced reduction than the other accessions. Interestingly, the *stn8* mutant did not show the same response as Col-0 wild type, retaining relatively higher expression of STN7 in HL.

Discussion

(a) *Variation in PSII protein phosphorylation and factors involved*

The reversible and differential phosphorylation of PSII proteins is dependent on the interplay between the STN7 and STN8 kinases. This process has been intensively studied in the standard lab accession Col-0 and *stn* mutants in the Col-0 background. A recent report compared phosphorylation levels in Col-0 with those in L*er*-0 and Ws-4, found that Ws-4 displayed 50% lower phospho-D1 and attributed this difference to 50% less STN8 kinase than in the other two accessions (Yin et al., 2012). Here we report on the occurrence of variation in D1 and LHCII protein phosphorylation ranging between approx. 5-120% in *Arabidopsis* accessions of diverse geographic origins, when expressed relative to Col-0. The large differences observed in GL in this set of accessions were not caused by variation in the amount of substrate or STN7 and STN8 protein levels. In HL, the levels of D1 phosphorylation correlated with the STN8 kinase levels, indicating that kinase abundance can be a limiting factor for phosphorylation under these conditions.

Among the genotypes we analyzed, some resembled the standard lab accession Col-0 in phosphorylation levels and were classified as high accessions. However, the other accessions displayed moderate or even low phosphorylation levels. Why would *Arabidopsis* accessions have variable phosphorylation of PSII proteins? Is this an adaptive mechanism facilitating survival and reproduction across the range of environmental conditions where *Arabidopsis* naturally occurs? In support of this hypothesis, we found a significant correlation between longitude and both D1 and LHCII phosphorylation in GL conditions (Figure 6.4 *a, b*). This suggests that there may be some form of selective pressure that correlates with longitude. In order to test this, climate data were obtained from the WorldClim database (Hijmans et al., 2005) (http://www.worldclim.org/). Bioclimatic variables 4 (temperature seasonality) and 7 (temperature annual range) correlated significantly with phospho-LHCII in GL conditions (Figure 6.4 *c, d*). Whilst there was some correlation between these climatic variables and phospho-D1, it was not found to be significant. This correlation between temperature variability and PSII protein phosphorylation in GL conditions is interesting and may be the outcome of a

photoprotective mechanism similar to that observed in evergreen trees which must maintain functioning leaves in very cold conditions (Verhoeven et al., 2009). Interestingly there was no correlation between protein phosphorylation and longitude in HL conditions, which could be due to a stronger, more geographically uniform, selective pressure in the HL response. This makes sense considering the damage an inappropriate response to HL can cause. To better understand the relationship between phosphorylation and the natural habitat, many more accessions will need to be investigated from a wide range of environments.

In line with published data (Rintamaki et al., 1997; Yin et al., 2012), we show that HL-treated plants contained more phospho-D1 whereas GL-treated plants displayed a higher extent of LHCII phosphorylation. The amount of kinase involved could be one mechanism to regulate enzyme activity, as indicated by the significant correlation between phospho-D1 levels and STN8 abundance in HL (Figure 6.2*b*). In the case of *Arabidopsis* STN7 or its *Chlamydomonas* homologue Stt7, it has been suggested that their amounts are regulated by the redox status of the electron transport chain, by phosphorylation and by transcript abundance (Willig et al., 2011; Wunder et al., 2013). In our panel we had two genotypes, Ely and ELB, that displayed reduced LHCII phosphorylation under both GL and HL conditions but high STN7 protein levels (Figure 6.3). These genotypes are atrazine resistant due to a deficient binding of quinones in the Q_B pocket, and as a result have reduced PSII efficiency (El-Lithy et al., 2005). Therefore, they are likely to have a more oxidized PQ pool especially at limiting irradiances. Although this requires experimentation, we use as support of our assumption the fact that at low, light-limiting irradiances Chl *b* deficient barley mutants displayed a more oxidized PQ pool than the wild type due to diminished PSII activity relative to PSI activity (Andrews et al., 1995). However, the unaffected abundance of STN7 protein relative to Col-0 in our study indicates that the redox state did not alter STN7 expression level. Therefore, the observed reduced level of LHCII phosphorylation is most likely due to reduced kinase activity.

(b) ***Regulation of D1 protein phosphorylation***

The factors regulating the amount of STN8 kinase in the membrane have not yet been investigated. D1 protein phosphorylation requires the presence of the STN8 kinase since phospho-D1 is hardly detected in the *stn8* mutant, and what remains is most likely due to a partial replacement by STN7 or other yet unknown kinases (Vainonen et al., 2005). This potential redundancy between STN7 and STN8 is illustrated by the much higher *STN7*

transcript and also protein levels in the *stn8* mutant in HL when compared to Col-0. The mechanism by which this difference is mediated is not clear and may be either direct or indirect. This may be due to the absence of functional STN8 protein stimulating additional *STN7* transcription and translation under HL conditions. However, the *stn8* mutant displayed unaltered LHCII phosphorylation levels, suggesting that the STN8 kinase does not play any role in this process. STN8 levels in Col-0 did not change upon transfer from GL to HL conditions in line with (Wunder et al., 2013).

All studied accessions displayed reduced D1 phosphorylation relative to Col-0 in GL despite high levels of the D1 substrate and STN8 kinase. One cause could be a poor activation of the STN8 kinase in GL, which could be related to the redox state of the PQ pool, as in the case of STN7. In support of this possibility is the low phosphorylation in Ely and ELB. As previously discussed with reference to STN7, due to low PSII efficiency, the PQ pool may be more oxidized which results in reduced activation of the kinase. ELB contains L*er*-1 nuclear DNA but the organellar DNA of Ely, and as such allows us to compare the phosphorylation level of PSII proteins in the same nuclear background. At both HL and GL ELB displayed half the phosphorylation level of D1 as that found in L*er*-1 (Figure 6.2), thus resembling Ely in the deficient activation of STN8. Since ELB is effectively identical to L*er*-1 as far as nuclear genome is concerned, the difference reflects the strong effect of the cytoplasm on the level of D1 phosphorylation. In all accessions except Ely and ELB, a weak but significant correlation was found in HL between phospho-D1 level and STN8 abundance. This indicates that under these conditions, the abundance of STN8 may be either limiting or plays a regulatory role in D1 phosphorylation.

(c) *Regulation of LHCII protein phosphorylation*

The phosphorylation of LHCII proteins enables the excitation and redox balance between PSII and PSI under low irradiance. This process requires the STN7 kinase, which is activated by a reduced state of the PQ pool under these light conditions (Rintamaki et al., 2000; Bellafiore et al., 2005). Upon exposure to high light, the kinase is inactivated by a thioredoxin-mediated reduction of disulfide bonds (Rintamaki et al., 2000). In our study, many accessions displayed moderate LHCII phosphorylation in GL, whereas the genotypes with atrazine-resistant forms of D1, Ely and ELB, displayed only residual phosphorylation levels, whilst at the same time showing much higher levels of Lhcb2 protein than Col-0 (Supplementary Table S6.1). There

may be a common cause with D1 phosphorylation, namely the inability to fully reduce the PQ pool, and thus to activate the kinase. ELB resembles Ely in low levels of LHCII protein phosphorylation, thus in the deficient activation of STN7. The reduction in PSII efficiency in these genotypes (El-Lithy et al., 2005) may result in both increased antennae size and reduced phosphorylation in order to increase PSII light-absorption relative to that of PSI under light-limiting conditions.

In our experimental conditions STN7 levels did not change upon shift from GL to HL in Col-0 despite decrease in LHCII phosphorylation levels (Supplementary Figure S6.2), indicating that STN7 is regulated at activity rather than protein level. This observation is in contrast with a recent report about down-regulation of STN7 at both protein and transcript levels (Wunder et al., 2013). The reason for this discrepancy could be the distinct light regimes used in this study or other yet unknown factors. However, the protein levels did change in other accessions (Supplementary Figure S6.2). Based on our results, the *STN7* transcript abundance decreased but did not correlate with the abundance of the STN7 protein (Supplementary Figure S6.2). This is in contrast to a recent study which showed that the accumulation of the STN7 protein was controlled at the level of transcript abundance (Wunder et al., 2013). However, that study was performed on Col-0 wild type and mutants in the Col-0 background, and based on our results it appears that Col-0 does not show a typical level of transcription for *STN7*.

The picture that emerges from these primary studies is that of highly diverse levels of PSII protein phosphorylation in nature, in which kinase activation may play a central role at least under GL conditions. It is likely that under HL conditions PSII core protein phosphorylation is in addition regulated by STN8 protein abundance in the thylakoid membrane. Longitude and temperature variability may be involved at least under GL conditions in variation of PSII protein phosphorylation. In conclusion, the significant variation found in both traits highlights our lack of understanding of the role these processes play in plant performance in nature. Using knockout mutants it has been shown that a complete absence of the STN7 kinase, and to a lesser extent the STN8 kinase, results in reduced fitness. This fitness cost is much more pronounced in the double mutant (Frenkel et al., 2007), once again illustrating that there is some degree of functional redundancy. A further conclusion from this work is that Col-0 appears to be an outlier accession. It operates at the phenotypic extreme for this trait and as such is most likely not representative of thylakoid protein phosphorylation in *Arabidopsis*. Based on an analysis of both *STN7* and *STN8* sequences in the many re-sequenced *Arabidopsis* accessions as found in the '1001 Genomes' website

(http://www.1001genomes.org/) it appears unlikely that the diversity of phosphorylation phenotypes observed is due to sequence variation in the kinase genes themselves. It is much more likely that the observed variation is due to variation elsewhere in the process, be that upstream signaling or downstream dephosphorylation rates. In order to identify the genes responsible for such variation, genetic mapping studies using either Recombinant Inbred Line (RIL) populations or genome wide association mapping panels could be undertaken (Bergelson and Roux, 2010). However currently the main limiting factor to such a study is not the availability of suitable genetic material but rather the ability to screen the necessary number of individuals (more than two hundred genotypes would be required), in sufficient replicates to allow for detection of the genetic loci involved, and even more to identify the causal sequence variation (Koornneef et al., 2004). Identification of such genes is likely to provide us with additional insights into the regulation of the photosynthetic process and the selective pressures acting upon this trait. Such knowledge will not only be of use to fundamental research but is likely to provide new avenues to crop improvement whereby the photoprotective processes can be optimized for different agricultural or climatic conditions (Flood et al., 2011; Murchie and Niyogi, 2011).

Acknowledgements

We thank Prof. Maarten Koornneef (Max-Planck-Institute for Plant Breeding, Cologne, Germany/Wageningen University) for initiating collaboration between the two laboratories. We thank Drs. Ross Alexander and Bas Dekkers (Wageningen University) for advice on qPCR analysis and Dr. Joost van Heerwaarden (Wageningen University) for help with climate data. This work was supported by funding from the Swedish Research Council and the Olle Enkvists Foundation (to C.S.), the Netherlands organisation of Scientific Research section Earth and Life Sciences (NWO-ALW), the Technological Top Institute Green Genetics, and the BioSolar Cells research programme (to P.J.F., J.H. and M.G.M.A.).

Electronic supplementary material is available

at http://dx.doi.org/10.1098/rstb.2013.0499 or via http://rstb.royalsocietypublishing.org.

Supplementary Material

Table S6.1. *Protein abundance and phosphorylation, gene expression, and chlorophyll a/b ratio in growth light (100 µmol m^{-2} s^{-1}).*

Genotype	D1	CP43	Lhcb2	p-D1	p-LHCII	STN8	STN7	Chl a/b	*STN7* transcript	*STN8* transcript
Col-0	100	100	100	100	100	100	100	3.41±0.00	3.71±0.29	1.23±0.15
stn8	92±14	81±17	100±14	4±3	104±23	0±0	79±13	3.39±0.02	2.57±0.39	0.16±0.02
Sha	100±0	82±29	77±18	30±5	85±23	169±20	100±15	3.45±0.03	2.59±0.76	1.58±0.30
C24	100±10	89±27	79±6	65±23	52±16	79±18	199±37	3.32±0.03	2.09±0.70	0.90±0.18
Bur-0	98±6	84±7	74±8	28±7	58±9	117±21	197±28	3.52±0.02	4.68±2.11	1.44±0.73
Pa-2	95±6	51±8	87±18	22±7	59±16	116±19	116±4	3.34±0.00	1.77±0.10	0.74±0.10
Mt-0	84±1	60±6	68±9	36±13	79±31	75±4	131±22	3.28±0.04	2.24±0.45	1.58±0.23
Per-1	69±20	53±4	77±4	68±18	101±19	117±7	73±4	3.44±0.00	1.00±0.39	1.11±0.27
Bor-4	103±22	84±19	100±35	34±6	71±4	84±33	207±2	3.28±0.03	3.24±0.37	2.17±0.25
Can-0	98±9	66±7	91±8	19±7	57±29	117±41	116±8	3.53±0.06	3.76±0.66	2.82±0.21
Tsu-0	96±7	69±1	78±4	79±17	95±12	106±5	70±4	3.41±0.05	2.62±0.31	2.25±0.23
Tsu × Ws-4	97±9	101±37	77±13	59±18	102±36	120±25	213±57	3.42±0.01	3.18±0.52	2.79±0.19
Ws-4	92±8	85±15	93±26	51±11	69±13	78±11	128±24	3.38±0.10	3.18±0.21	3.06±0.42
Ler-1	87±10	64±9	81±19	22±9	43±4	192±8	152±38	3.41±0.00	2.28±0.48	2.62±0.72
ELB	86±1	68±24	126±62	10±4	8±7	173±45	163±22	3.19±0.00	1.38±0.37	1.33±0.37
Ely	96±4	59±17	143±77	16±6	7±5	98±4	71±12	3.32±0.065	2.75±0.13	1.43±0.12

The data are expressed relative to Col-0 for D1, CP43, Lhcb2, p-D1, p-LHCII, STN8 and STN7. P-D1, p-LHCII are corrected with the levels of D1 and Lhcb2, respectively. The levels of STN8 and STN7 proteins are corrected with the levels of CP43 protein. The levels of *STN8* and *STN7* transcripts are corrected with a normalisation factor calculated from seven reference genes. The protein data are means of 2-3 technical replicates ±s.d., whereas the transcript data are means of 3 biological replicates ±s.e..

Table S6.2. *Protein abundance and phosphorylation, gene expression, chlorophyll a/b ratio in high light (600 µmol m^{-2} s^{-1}).*

Genotype	D1	CP43	Lhcb2	p-D1	p-LHCII	STN8	STN7	Chl a/b	*STN7* transcript	*STN8* transcript
Col-0	100	100	100	100	100	100	100	3.25±0.03	0.05±0.00	0.71±0.18
stn8	94±3	76±3	95±5	4±0	105±6	2±3	136±4	3.15±0.11	1.51±0.19	0.22±0.05
Sha	104±22	114±23	90±19	49±22	107±6	73±6	74±11	3.21±0.1	1.49±0.79	1.99±1.10
C24	99±22	88±1	89±16	59±20	105±15	88±14	93±15	3.27±0.12	0.13±0.06	0.92±0.35
Bur-0	114±15	115±6	101±4	101±8	102±22	121±19	150±19	3.34±0.05	1.00±0.50	1.31±0.18
Pa-2	98±11	87±9	108±4	104±8	88±14	59±6	100±5	3.23±0.01	1.17±0.61	0.81±0.20
Mt-0	100±31	104±13	121±15	117±7	91±8	104±18	121±5	3.19±0.1	1.98±0.85	1.96±0.31
Per-1	103±20	82±22	119±16	89±6	64±9	136±25	149±4	3.27±0.07	0.07±0.01	0.61±0.17
Bor-4	99±10	89±5	123±1	90±8	60±12	85±1	91±11	3.22±0.06	1.13±0.50	0.63±0.24
Can-0	86±19	104±35	103±10	71±13	29±9	32±11	80±3	3.41±0.1	1.02±0.53	0.71±0.15
Tsu-0	97±8	95±7	96±9	116±18	82±13	43±6	92±13	2.91±0.13	1.19±0.33	0.58±0.10
Tsu × Ws-4	87±4	105±8	82±6	88±13	106±24	91±8	100±4	3.42±0.02	0.66±0.15	0.84±0.08
Ws-4	105±6	105±0	105±7	64±9	64±11	44±4	104±10	3.42±0.02	1.05±0.30	0.84±0.12
Ler-1	94±14	95±24	92±2	91±1	83±17	56±1	58±6	3.04±0.09	1.31±0.23	1.72±0.68
ELB	90±7	75±14	98±20	28±1	14±4	83±6	179±5	3.13±0.02	1.66±0.51	2.07±0.79
Ely	93±4	64±15	109±12	13±4	5±4	99±27	102±37	3.32±0.03	0.08±0.01	0.58±0.11

The data are expressed relative to Col-0 for D1, CP43, Lhcb2, p-D1, p-LHCII, STN8 and STN7. P-D1, p-LHCII are corrected with the levels of D1 and Lhcb2, respectively. The levels of STN8 and STN7 proteins are corrected with the levels of CP43 protein. The levels of *STN8* and *STN7* transcripts are corrected with a normalisation factor calculated from seven reference genes. The protein data are means of 2-3 technical replicates ±s.d., whereas the transcript data are means of3 biological replicates ±s.e..

Table S6.3. *AGI codes, gene names, primer sequences and efficiencies for all genes tested.*

AGI Code	Gene name	Forward primer sequence	Reverse primer sequence	PCR efficiency	Reference
At5g01920	*STN8*	CGCCATTGACGCCTACCTTC	GGCATAGACGACGCCGAAAG	0.99	This study
At1g68830	*STN7*	CGATGGTCCGGTACAAAGCA	GGACAACGCGAGCAATCCTT	0.93	This study
At3g25800	*PDF1/ PP2AA2*	AATCGGTTGTGGAGAAGACG	GCGAAAAACCTGACATCAACAT	0.9	(Dekkers et al., 2012)
At5g12240 *	unknown	AGCGGCTGCTGAGAAGAAGT	TCTCGAAAGCCTTGCAAAATCT	0.85	(Czechowski et al., 2005)
At3g33520	*ARP6*	TAACAACTCAGGAGGACCCCA	CTACGACACCGAGCTGAT	0.939	(Dekkers et al., 2012)
At1g13440	*GAPDH*	TTGGTGACAACAGGTCAAGCA	AAACTTGTCGCTCAATGCAATC	0.93	(Czechowski et al., 2005)
At4g33380	unknown	TTGAAAATTGGAGTACCGTACCAA	TCCCTCGTATACATCTGGCCA	0.81	(Czechowski et al., 2005)
At5g25760	*PEX4*	TCCTGAGCCGGACAGTCCTC	CATAGCGGCGAGGCGTGTAT	0.98	This study
At2g28390	*SAND*	GTTGGGTCACACCAGATTTTG	GCTCCTTGCAAGAACACTTCA	0.83	(Czechowski et al., 2005)
At3g41768	*18S rRNA*	GGCTCTGGCTTGCTCTGATG	TGCCTTCCTTGGATGTGGTAG	0.99	(Rigola et al., 2006)

*Unstable amplification, excluded from subsequent analysis.

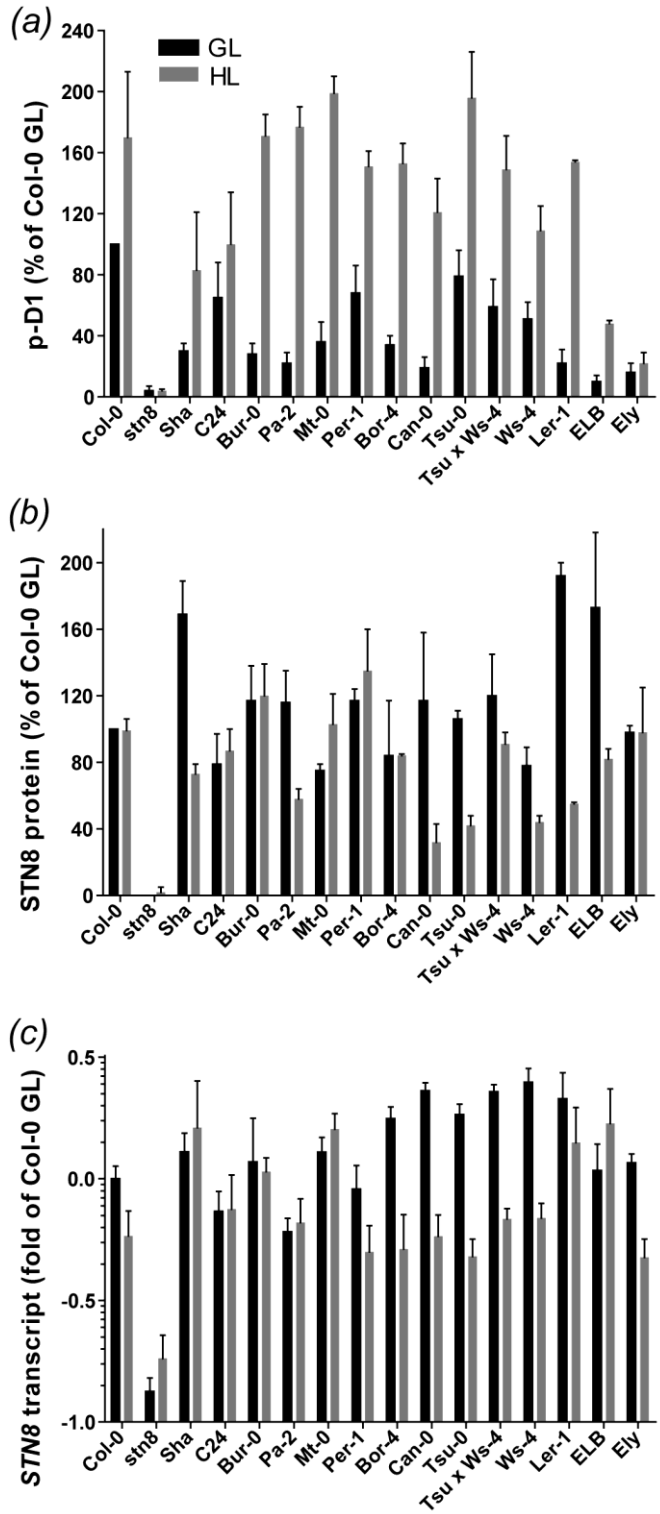


Figure S6.1. *Bar charts of D1 protein phosphorylation (a), of the STN8 protein (b) and of the STN8 transcript levels (c) in growth (GL) and high light (HL) expressed relative to Col-0 levels in GL conditions. Transcript levels are expressed on a log scale. The protein data are means of 2-3 technical replicates ±s.d., whereas the transcript data are means of 3 biological replicates ±s.e..*

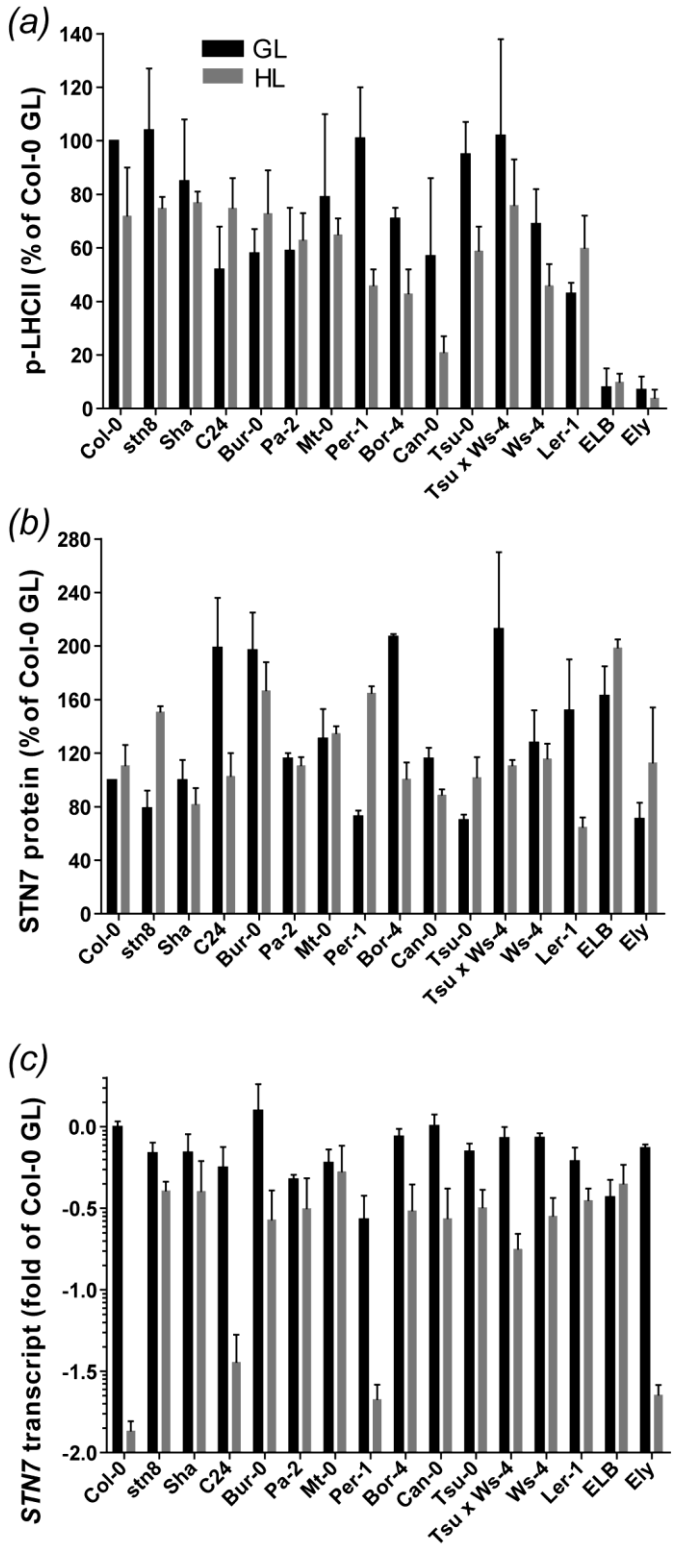


***Figure S6.2**. Bar charts of LHCII protein phosphorylation (a), of the STN7 protein (b) and of the STN7 transcript level (c) in growth (GL) and high light (HL) expressed relative to Col-0 levels in GL. Transcript levels are expressed on a log scale. The protein data are means of 2-3 technical replicates ±s.d., whereas the transcript data are means of 3 biological replicates ±s.e.*

Chapter 7

Construction and preliminary analysis of a nucleotype-plasmotype diallel in *Arabidopsis thaliana*

Pádraic J. Flood[1,2,3], Evangelos Kouklas[1], C. Bastiaan de Snoo[4], Jos A. Hageman[5], Frank Becker[1], Sabine Schnable[5], Leo Willems[6], Hetty Blankestijn[1], Jeroen van Arkel[7], Roland Mumm[7], Joost J.B. Keurentjes[1], Jeremy Harbinson[2], Wilco Ligterink[6], Fred van Eeuwijk[5], Mark G.M. Aarts[1], & Erik Wijnker[8].

[1]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

[2]*Horticulture and Production Physiology, Wageningen University, Wageningen, the Netherlands.*

[3]*Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany.*

[4]*Rijk Zwaan R&D Fijnaart, Fijnaart, the Netherlands.*

[5]*Biometris, Wageningen University, Wageningen, the Netherlands.*

[6]*Laboratory of Plant Physiology, Wageningen University, Wageningen, the Netherlands.*

[7]*PRI metabolomics, Wageningen University, Wageningen, the Netherlands.*

[8]*University of Hamburg, Department of Developmental Biology, Hamburg, Germany.*

Corresponding Author: Pádraic J. Flood (flood@mpipz.mpg.de)

Introduction

Plants and algae contain three distinct repositories of genetic information, the nuclear genome or nucleome, the mitochondrial genome or chondriome, and the plastid genome or plastome (Greiner and Bock, 2013). The nuclear genome of an individual, hereafter referred to as the nucleotype, contains the vast majority of the genetic information. As a result it is the main focus of most research, so much so that the roles of the chondriome and plastome, collectively referred to as the plasmotype, are often overlooked (Pesole et al., 2012). The plasmotype can, however, have a significant impact on a wide range of processes of both evolutionary and agricultural importance (Bock et al., 2014; Dobler et al., 2014). Examples in plants are the role of the plastome in photosynthetic adaptation to diverse environments (Iida et al., 2009; Greiner and Bock, 2013; Bock et al., 2014) and the role the chondriome plays in the development of gynodioecy via cytoplasmic male sterility (Caruso et al., 2012). The plants' plasmotype has further been shown to influence a plethora of traits ranging from biotic and abiotic stress tolerance such as disease resistance and cold tolerance to physiological and metabolic variation such as photosynthetic capacity and secondary metabolite abundance (Budar and Fujii, 2012; Greiner and Bock, 2013; Joseph et al., 2013).

The relative paucity of studies on the role of the plasmotype in plant adaptation is partially due to the difficulty of assessing its phenotypic effects independent of the nucleotype. This has occasionally been overcome using protoplast fusion, a tissue-culture technique where cells, usually from different species, are fused together resulting in a somatic hybrid. Such hybrids usually show cellular segregation for chloroplasts allowing for comparison of chloroplast function in different nuclear backgrounds. The mitochondria on the other hand often recombine with each other giving rise to a new mitochondrial lineage which is a chimera of the fused parents (Pelletier et al., 1983; Liu et al., 2005). Such tissue culture techniques are laborious, have a limited success rate, and are not applicable to the model species *Arabidopsis thaliana*. In this species, as in most angiosperms, organelles are maternally derived (Hagemann, 2004; Elsayed, 2011) with very low levels of paternal leakage (Azhagiri and Maliga, 2007). Potential nucleotype-plasmotype interactions have been studied via reciprocal F1 hybrids (Fujimoto et al., 2012), through recombinant inbred lines (RILs) in the special case that the RILs were derived from reciprocal hybrids (Joseph et al., 2013), and by introgression via repeated backcrossing (El-Lithy et al., 2005). All of these methods have specific drawbacks. In F1s, plasmotype effects may be confounded by maternal and paternal effects (Bashir et al., 2014) and the F1s nuclear heterozygosity may mask phenotypic effects of the plasmotype via

nuclear complementation (Stoll et al., 2015). In contrast, the nuclear genomes of RILs are homozygous, but prohibit direct comparisons of nuclear-cytoplasmic effects since all RILs have a unique recombinant nucleotype of the parental genotypes. Finally, introgression of a specific nucleotype into a new plasmotypic background requires repeated backcrosses which is very laborious, and even after many backcrosses, will not always result in complete nuclear replacement.

To compare variation in nucleotype-plasmotype compatibility and identify main and interaction effects, a complete reciprocal exchange of the nucleotype into a new plasmotypic background is ideally required. None of the techniques just described achieve this in a reliable fashion. A final approach to reciprocally exchange nucleotypes and plasmotypes is through the use of haploid-inducer lines. Haploid-inducer lines are plants, usually of a different species, whose genomes are (uniparentally) lost from the zygote after fertilization (Houben et al., 2011). This uniparental genome loss results in haploid offspring from which doubled haploids are derived. The mechanism underlying this is best understood for Barley (*Hordeum vulgare*) (Sanei et al., 2011). This species can be fertilized with pollen of *H. bulbosum*, which functions as a haploid-inducer. In the zygote, the Barley chromosomes load Barley CENTROMERE HISTONE 3 (CENH3) onto their centromeres when they replicate, but the *H. bulbosum* chromosomes fail to do so, leading to the gradual depletion of CENH3 (and centromere function) on the inducer-line derived chromosomes. After several mitotic cell-divisions, the *H. bulbosum* chromosomes are lost from the zygote and a haploid zygote remains. The underlying cause for the failure to load CENH3 on one parental set of chromosome centromeres remains unknown. Uniparental genome elimination mostly results when haploid-inducer lines are used as the male parent (Houben et al., 2011). Such male haploid-inducers are often used to create segregating haploid offspring. Subsequent doubling of the haploid genome, either by chance or through colchicine treatment, converts haploids into homozygous diploids (commonly known as doubled haploids).

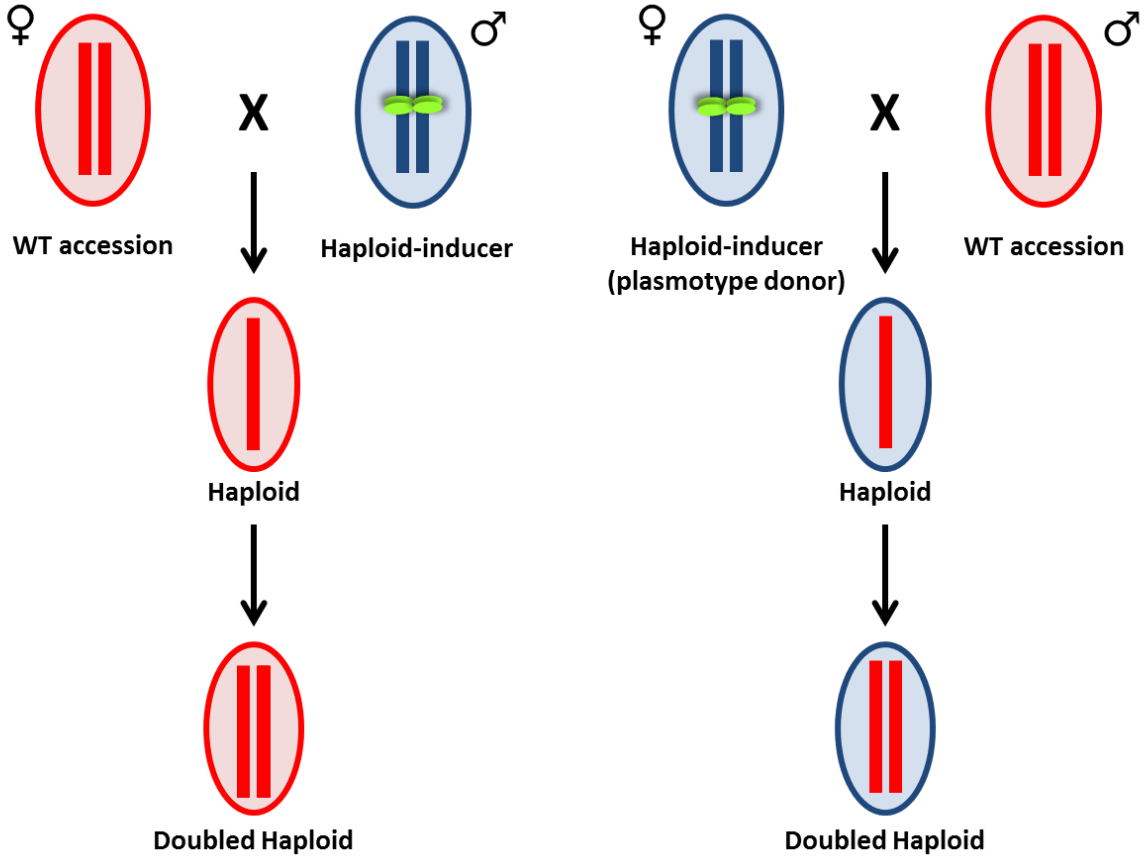


Figure 7.1. *Haploid-inducers can function as plasmotype donors. Vertical bars represent chromosomes, circles represent plasmotype (mitochondria and chloroplasts). Centromeres of the haploid-inducer are indicated in green to signify GFP-tagged CENH3 proteins that are present at the centromeres, and cause uniparental genome-elimination in Arabidopsis. Left panel shows the cross of a WT accession (female) with a haploid-inducer (male). The cross results in uniparental genome-elimination of the paternal genome, leading to a haploid (and later a doubled haploid) that is genetically identical to the paternal plant. Right panel shows a cross in which a haploid-inducer is used as a female parent. In this case, uniparental (female) genome-elimination leads to a haploid plant with the nucleotype of the father, but the plasmotype of the mother plant (the haploid-inducer).*

Male haploid-inducers make no genetic contribution to their haploid offspring, as their nucleus is eliminated and organelles are inherited from the maternal (non-inducer) parent. In some cases, however, the haploid-inducer can be used as female in crosses. In this case the inducer (but now maternal) nucleome is lost from the zygote, but its plasmotype is retained. The haploid inducer in this case acts like a plasmotype donor: its haploid offspring carry a paternally derived (non-inducer) nucleotype and a maternally (inducer line) derived plasmotype. These haploids have a new nucleotype plasmotype combination (sometimes referred to as cybrids) which can then be used to study main and interaction effects of the plasmotype (Figure 7.1).

Until the recently no such haploid-inducer system existed for *Arabidopsis thaliana.* This was overcome by complementing the normally lethal *htr12* mutation (also known as "*cenh3-1",* as it is mutated in the Arabidopsis *CENH3* gene), with a rescue construct encoding a GFP-HTR12 fusion protein. When this '*GFP-tailswap'* is crossed with a wild type (WT) accession, the *GFP-tailswap* derived chromosomes are lost from the zygote, and lead to uniparental genome-elimination (Ravi and Chan, 2010). This haploid-inducer represents a special case in which the inducer line is conspecific. Its genome-elimination mechanism is, as of yet, unclear, but like in *Hordeum* it is linked to the function of CENH3. This inducer line eliminates its nuclear genome when used as either the male or the female in a cross. When used as a female, it also acts as a plasmotype-donor.

To explore the effect of new nucleotype-plasmotype combinations (cybrids) a full test-panel consisting of all the 42 possible reciprocal cybrids of seven Arabidopsis accessions was generated using the *GFP-tailswap* line. Here we describe the creation of this cybrid diallel and present an initial exploratory analysis of the effect of plasmotype on organismal and metabolic phenotypes, with special attention to the impact of the triazine resistant Ely plasmotype.

Materials and Methods

Plant materials

Seven Arabidopsis accessions were chosen for construction of a nucleotype-plasmotype diallel. Ely (CS28631) is atrazine resistant (El-Lithy et al., 2005) and was found to have the lowest photosynthetic capacity from among a set of 703 accessions (*see Chapters 5, 6 and 8*).Ws-4 (CS5390) was included for its unusual photosystem II phosphorylation dynamics (Yin et al., 2012; Flood et al., 2014). Bur-0 (CS76105) was added for its extreme photosynthetic phenotype of having one of the highest Φ_{PSII} values out of a set of 703 accessions (*Chapter 5*). Shah (CS76227) was selected based on its capacity to induce cytoplasmic male sterility (CMS) in some crosses (Gobron et al., 2013). The set was completed by adding L*er*-1 (CS76164), Col-0 (CS76113), and C24 (CS76106). Both L*er*-1 and Col-0 show heterosis upon crossing with C24, an effect that is mildly enhanced when offspring carry the C24 cytoplasm (Fujimoto et al., 2012; Chen, 2013). The *GFP-tailswap* haploid-inducer (Ravi and Chan, 2010) is in a Col-0 background and was obtained from Simon Chan, Dept. of Plant Biology, University of California, Davis.

Phenotyping

Photosynthesis & Growth: All lines, with the exception of Ely nucleotype with Shah plasmotype (Ely^{Shah}: we will use this format, with the plasmotype as a superscript, to indicate which genome was derived from which parent), were screened in a high throughput phenotyping facility located in a climate controlled growth chamber. This phenotyping platform measured the plants for: photosynthetic efficiency (Φ_{PSII}) using chlorophyll fluorescence, reflectance at 480 nm, 532 nm, 550 nm, 570 nm, 660 nm, 700 nm, 750 nm and 790 nm, and projected leaf area (PLA) based on pixel counts of near infra-red (NIR) images (Chapter 3). The growth chamber was set to a 10 hr. day/14 hr. night, 20°C day/18°C night, 200 $\mu molm^{-2}s^{-1}$ light and 70% relative humidity. The plants were grown on a rockwool substrate and irrigated daily with nutrient solution (Chapter 3).

Germination: Seeds for the germination experiments were generated from two rounds of propagation. In the first round of propagation eight replicates per line were first sown in a growth chamber and moved after three weeks to an illuminated cold room at 4°C for six weeks for vernalization. After vernalization all plants were moved to a temperature-controlled greenhouse for flowering and seed ripening. Exceptions to this were Ler^{Ely}, $Ler^{Ws\text{-}4}$, and $Ely^{Ws\text{-}4}$ for which no doubled haploid seed was available at the beginning of the first propagation round. Ler^{Ely} and $Ler^{Ws\text{-}4}$ were sown later during the vernalization stage and flowered at the same time as the vernalized plants. $E^{Ws\text{-}4}$ produced haploid seed at a later stage and could not be included in the first propagation round. During flowering of the first propagation round there was a thrips infestation in the greenhouse which mainly affected lines with Ely nucleotype, resulting in reduced seed yield for these lines. To compensate for this in the second round of propagation, the six lines with Ely nucleotype which were sown at 8 replicates per line, all other lines were sown with 4 replicates per line. Plants were grown in a temperature-controlled greenhouse at 20°C, in this round only lines with Ely nucleotype were vernalized. All seeds were stratified for four days at 4°C before being assayed in the Germinator platform (Joosen et al., 2010) for seed size, germination rate and total germination.

Metabolomics: Plant material for metabolomic analysis was obtained from the photosynthetic phenotyping experiment just described. Plants were harvested 26 days after sowing, which due to the 10hr photoperiod was prior to bolting for all lines. Samples were frozen in liquid nitrogen, and samples of each genotype were subsequently combined into four pools each made up of material of approximately six replicates. Each pool was ground and homogenized before an aliquot was taken for further analysis. Reference samples for metabolomics analysis were

composed of material from all seven parents in equal amounts and then homogenized.The method used for the extraction of polar metabolites from Arabidopsis leaves was adapted from Lisec et al. (2006) as described by Carreno-Quintero et al. (2012). Specific adjustments for Arabidopsis samples were made as follows; the polar metabolite fractions were extracted from 100 mg (FW, with 5% deviation) of Arabidopsis leaf material. After the extraction procedure 100 µL aliquots of the polar phase were dried by vacuum centrifugation for 16 hours. The derivatization was performed on-line as described by Lisec et al. (2006), and the derivatized samples were analysed by a GC-TOF-MS system composed of an Optic 3 high-performance injector (ATAS^{TM}, GL Sciences, Eindhoven, The Netherlands) and an Agilent 6890 gas chromatograph (Agilent Technologies, Santa Clara, United States) coupled to a Pegasus III time-of-flight mass spectrometer (Leco Instruments, St. Joseph, United States). Two microliters of each sample was introduced in the injector at 70°C using split 20, the detector voltage optimal for the detection of the components was 1750 Volts.

All samples were analyzed in random sequence in four separate batches. The systematic variation that inadvertently is introduced by working in batches, was removed by analysis of covariance. In this model the batch number was used as a factor (four levels) and run number within a batch as a covariate since it is also expected that (some) variation will be introduced by the sample run order within each batch. Figure s7.1 shows two score plots from Principal Component Analysis of the before (left) and after correction (right). Before correction, the four batches are separated from each other and the quality control (QC) samples are spread considerably. After correction the QC samples cluster nicely together and the four batches can no longer be distinguished from each other. The corrected values were used in subsequent analysis of the - in total - 82 metabolites (Data s7.1).

Statistical analysis

Phenotype data were fit with a linear model using the statistical software R (R Core Team, 2014). The phenotype was set as the responsible variable and a null model was built including design factors such as the position in the imaging head or the harvest from which the seed came from (in the case of germination phenotypes). The null model varied depending on the phenotype, i.e. position in the imaging head had a significant impact on Φ_{PSII} but not on PLA, these null models were then compared with models containing nucleotype and plasmotype as additional terms (Appendix 7.1). Analysis of variance (ANOVA) was used to test if the addition of plasmotype or nucleotype was statistically significant ($p<0.05$). For traits where nucleotype or plasmotype had a significant effect a Tukey honest significant differences (Tukey HSD) test

was used to estimate the effect size and the significance level of specific nucleotypes and plasmotypes on the trait.

Results

Generation of plasmotype-donors

As was explained in Figure 7.1, *GFP-tailswap* can be used as a plasmotype donor when used as a female in crosses. The original Arabidopsis *GFP-tailswap* haploid-inducer is in a Col-0 background, and could therefore serve as a Col-0 plasmotype donor. To make new nucleotype-plasmotype combinations new plasmotype donors needed to be generated. To generate plasmotype donors for the other six plasmotypes (Bur-0, C24, Ely, L*er*-1, Shah, & Ws-4), parental accessions were pollinated with *GFP-tailswap* pollen (Figure 7.2). Since genome elimination does not occur with 100% efficiency, such *GFP-tailswap* crosses typically give rise to a mixture of offspring made up of: haploids that result from uniparental genome elimination, aneuploids due to incomplete uniparental genome elimination, and heterozygous F1s between *GFP-tailswap* and the WT accession in which no genome-elimination took place (Tan et al., 2015). These F1s have a non-Col (female derived) plasmotype (Figure 7.2). They were phenotypically selected based on self-fertility (seed-filled siliques), tested for nuclear heterozygosity using genetic markers (described in Wijnker et al. (2012)), and self-fertilized. New haploid-inducers/plasmotype donors (*htr12* -/- ; *GFP-tailswap* +/…) were selected in the F2 based on their distorted leaves (that result from a segmental duplication associated with the *GFP-tailswap* insertion; M. Ravi, pers. comm.) and semi-sterility phenotype due to incomplete rescue of centromere function by the *GFP-tailswap* (Ravi et al., 2011). These new plasmotype donors were genotyped to confirm homozygosity for *htr12* (Ravi and Chan, 2010; Ravi et al., 2011), and self-fertilized before use in further crosses.

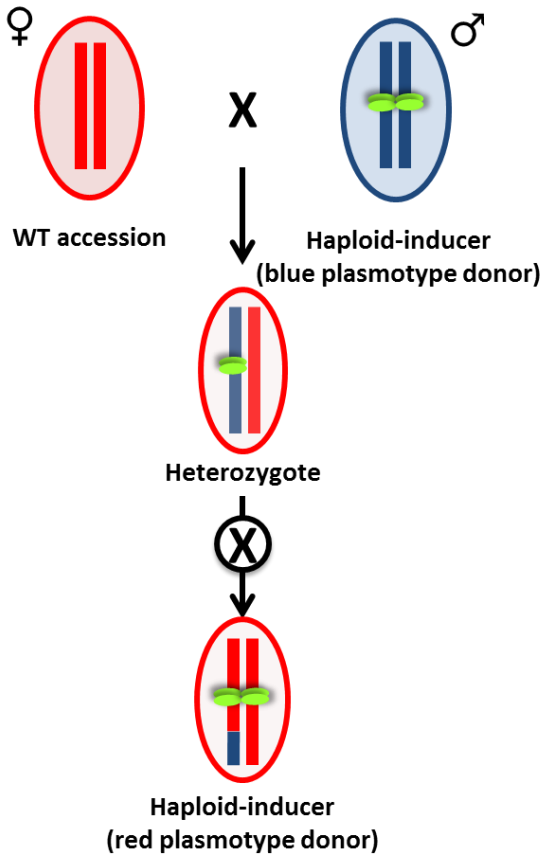


Figure 7.2. *the generation of new plasmotype donors. A WT accession was pollinated with GFP-tailswap pollen to give rise to a heterozygote (with the WT cytoplasm). The heterozygote was self-fertilized, to obtain a plant with a nuclear genotype that is homozygous mutant for hrt12, and contains the GFP-tailswap construct. Consequently, the selected new plasmotype donor (bottom) loads GFP-tagged CENH3 onto its centromeres. Note that the new plasmotype donor has a recombinant nucleotype. For a legend, please see Figure 7.1.*

Generation of a plasmotype-nucleotype diallel

To generate new nucleotype-plasmotype combinations, WT plants of all seven accessions (Col-0, Bur-0, C24, Ely, L*er*-1, Shah and Ws-4) were crossed as males to *GFP-tailswap* (the Col-0 plasmotype donor) as well as the six newly generated plasmotype-donors (Figure 7.3). The haploids arising from these 49 crosses were first identified based on their phenotype (as described in (Wijnker et al., 2014)) after which their haploidy was confirmed using 39 nuclear SNP markers. These haploid lines were self-fertilized to obtain doubled haploid offspring (Figure 7.1). The resulting 49 lines comprise a full diallel of 21 pairs of reciprocal nucleotype-plasmotype combinations as well as seven WT nucleotype-plasmotype combinations that have the same nucleotype-plasmotype combinations as their WT progenitors. These will hereafter be referred to as "self-cybrids" (Figure 7.3, diagonal). All doubled haploid cybrids and WT accessions were propagated for one generation before use in further experimentations with the

exception of Ely^{Shah} in which the original haploid died and had to be recreated thus limiting our analyses to 48 rather than 49 cybrids. While the self-cybrid nucleotype-plasmotype combinations are in theory genetically identical to the WT parental lines, these WT parental lines were included in the screens to test for possible effects of cybrid production (which involves a haploid growth stage). This brings the number of lines included in this study to a total of 56 (41 cybrids, 7 self-cybrids and 7 wild types).

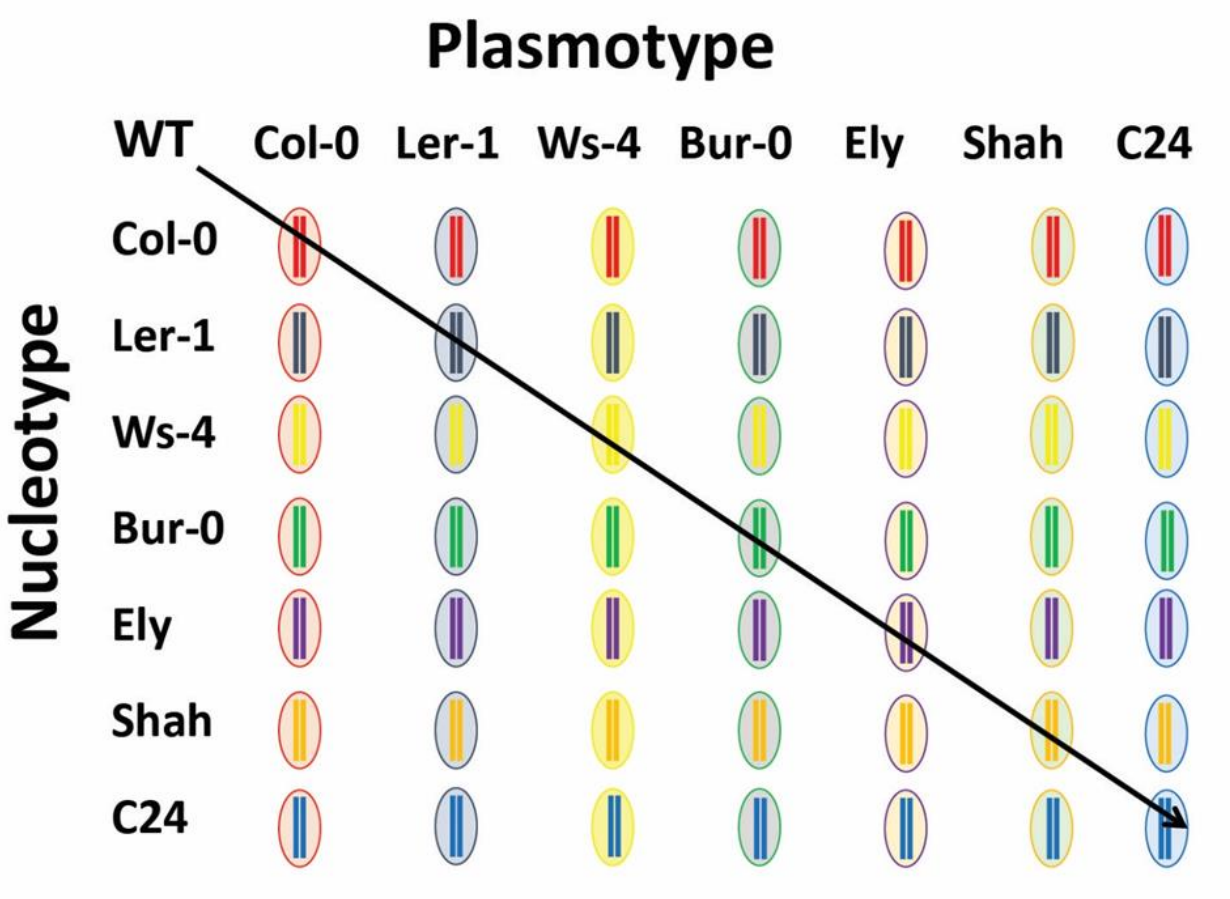


Figure 7.3. *Full diallel of all nuclear cytoplasmic combinations. The diagonal arrow highlights the WT nucleotype-plasmotype combinations, so called self-cybrid lines.*

Effect of cybrid generation

The self-cybrids and WT genotypes were compared to test for possible effects of cybrid-generation, a process encompassing uniparental genome-elimination and a haploid sporophytic growth phase. No significant differences were found between these self-cybrids and WT lines for plant size and Φ_{PSII}. The self-cybrids and WT accessions were therefore considered identical and pooled during further analyses.

Effect of plasmotype on organismal phenotypes

The main effects of the different nucleotypes and plasmotypes on the average Φ_{PSII} through time (10 to 21 days after sowing) are shown in Figure 7.4. Both Ely and Shah nucleotype cause a significant reduction in Φ_{PSII}, however this effect was much less than that of the triazine resistant Ely plasmotype. The reduction in Φ_{PSII} caused by the Ely plasmotype was expected and has been described previously (El-Lithy et al., 2005) (Chapters 3, 4, 5, and 8) however, this is the first time it has been transferred to a set of completely new nucleotypes and shows the high penetrance of this mutation. The Shah plasmotype also had a significant effect causing a small reduction in Φ_{PSII} (Figure 7.4b).

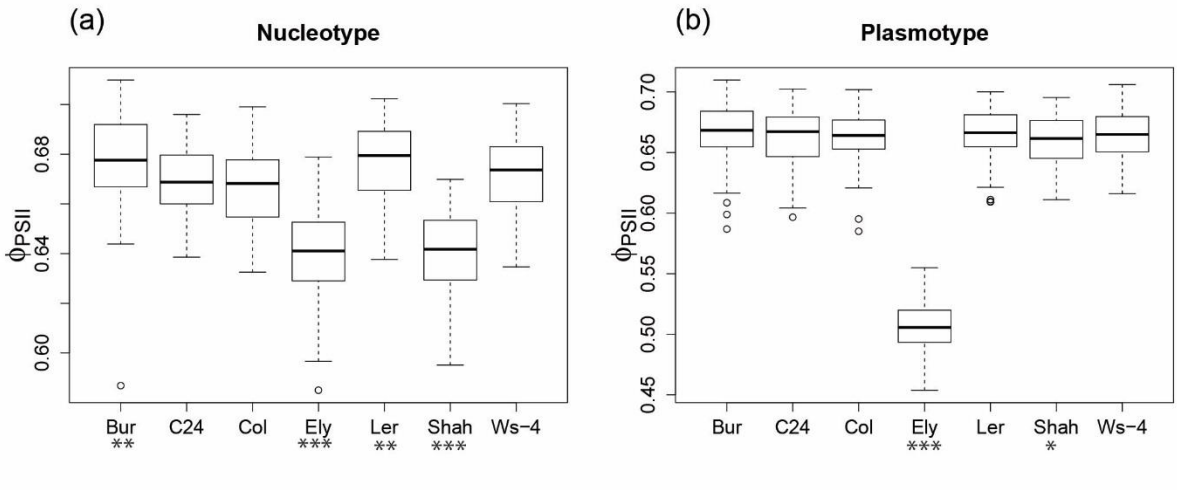


Figure 7.4. *Boxplot showing main effect of Φ_{PSII} per nucleotype (a) and plasmotype (b). The x-axis indicates the genotype and the y-axis indicates Φ_{PSII}, the black bar in the center of the box indicates the mean, * = P < 0.05, **= P < 0.01, *** = P < 0.001.*

The effect of plasmotype on growth, chlorophyll reflectance, seed size, and germination rate is shown in Figure 7.5. For all of these phenotypes the Ely plasmotype results in lower values. The reduction in growth, chlorophyll content, and seed size can be explained by the reduction in the efficiency of PSII. The cause of the reduction in t50, which indicates an increased rate of germination, is not so obvious.

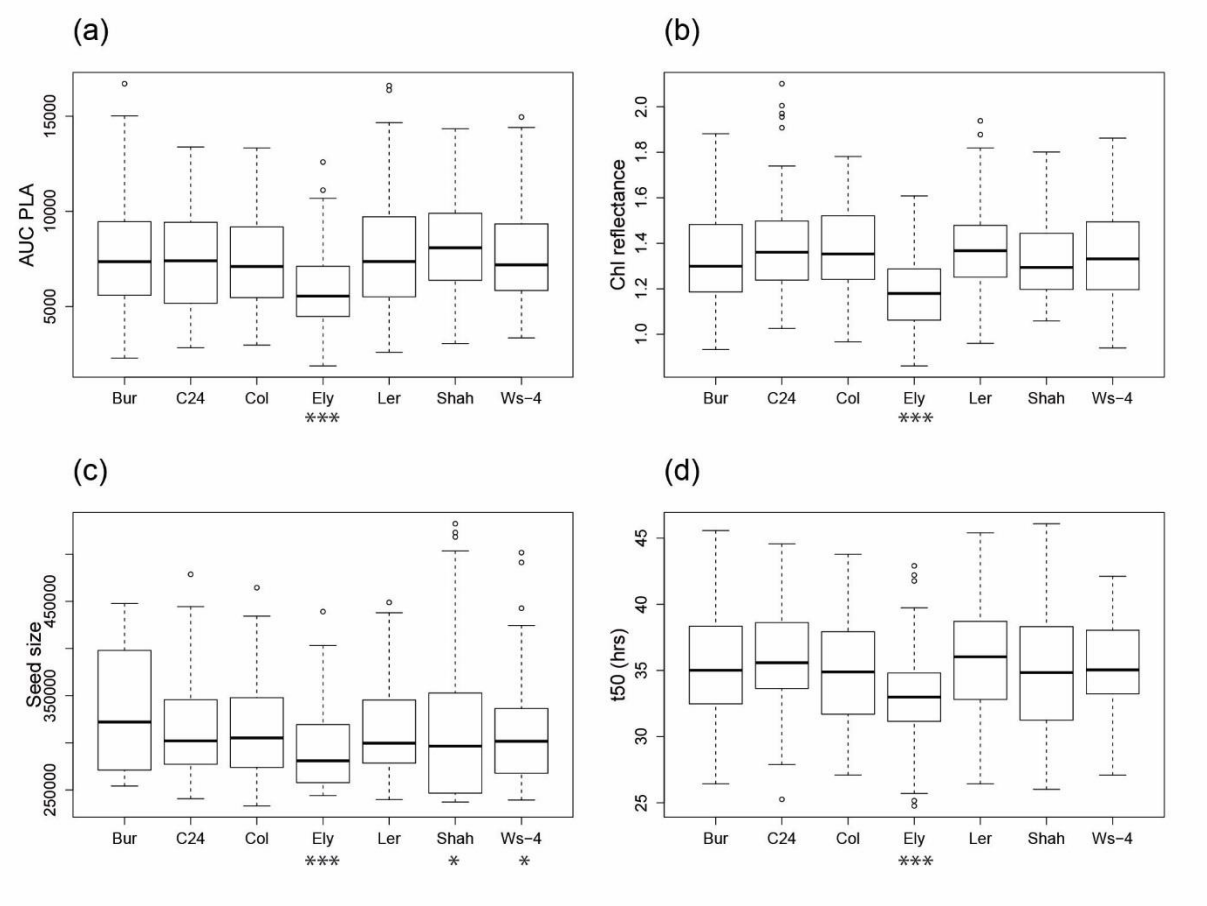


Figure 7.5. *Boxplots showing the main effect of plasmotype on (a) growth (AUC PLA = area under the curve calculated from projected leaf area, in pixels) (b) chlorophyll (Chl) reflectance (c) imbibed seed size (in μm^2) and (d) germination rate measured as t50, time it takes for 50% of the seeds to germinate, * = P < 0.05, *** = P < 0.001.*

Effect of plasmotype on primary metabolism

The abundance of 82 primary metabolites was measured to investigate plasmotype effects on primary metabolism. Most notable is a broad reduction of primary metabolites caused by the Ely plasmotype (Figure 7.6). A large number of the metabolites that reduced in abundance were sugars (Figure 7.7) which fits with the known impact of PSII efficiency on carbon assimilation (Heber et al., 1988).

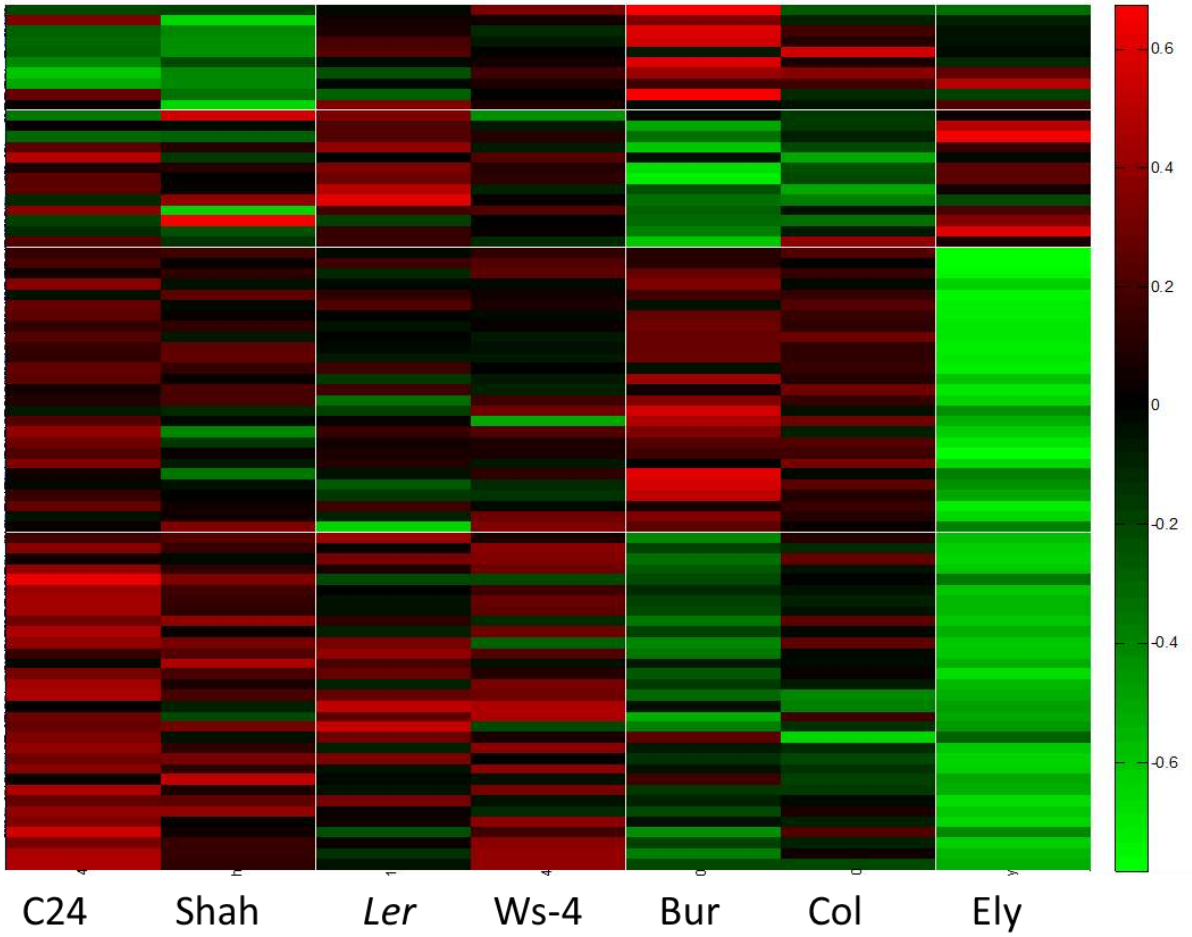


Figure 7.6. *Effect of plasmotype on primary metabolite profile. The 82 metabolites (Data s7.1) each have an individual row, red indicates the plasmotype increases abundance and green indicates the plasmotype reduces the metabolite abundance.*

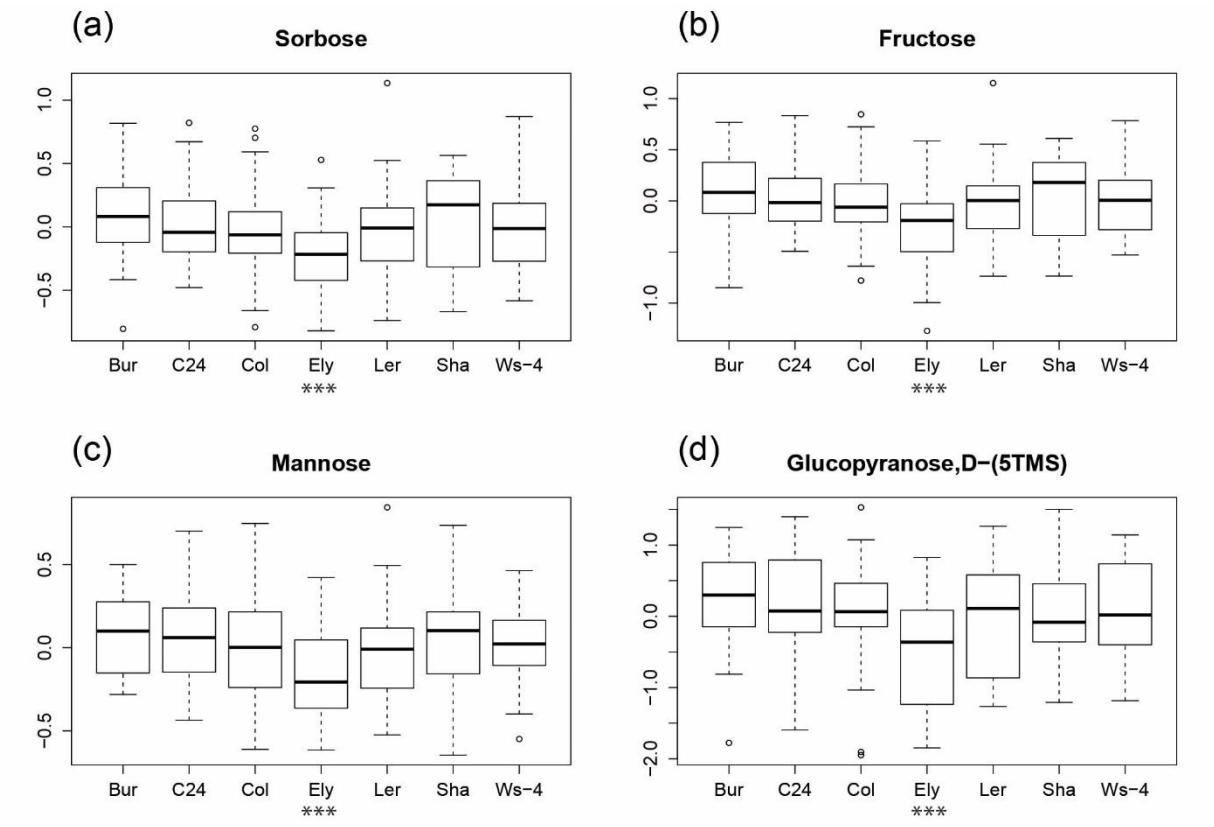


Figure 7.7. *Box plots showing main effect of plasmotype on selected primary metabolites. In all cases the Ely plasmotype reduces the metabolite abundance, *** = P < 0.001.*

Discussion

The goal of this chapter has been to introduce and describe the creation of a unique cybrid diallel in the model plant species *Arabidopsis thaliana*. The phenotype and metabolomic results are presented in an illustrative rather than comprehensive manner, but nevertheless there is sufficient evidence to suggest that this set of genotypes will be very useful for future research into nucleo-plasmotype interactions. In this discussion we will briefly look at some of the main results and will consider future experiments for which this population would be well suited. We will first consider the effect of the Ely plasmotype as its large impact on PSII efficiency is of relevance to the theme of this thesis.

The mutation in PsbA in the Ely plasmotype has a significant effect on Φ_{PSII}, and appears to have a number of pleiotropic effects. Figure 7.5a shows that the Ely plasmotype reduces plant growth, an effect that is likely related to the key role photosynthesis plays in growth (Long et al., 2006) although the exact relationship is still a matter of debate (Flood et

al., 2011). The reduced Φ_{PSII} and growth are also accompanied by reduced chlorophyll reflectance, a measure of plant greenness (Chapter 3), and seed size (Figure 7.5 b and c). These phenotypes likely reflect the reduction in photosynthetic efficiency resulting in a reduced plant performance which has also been borne out through numerous studies on the fitness costs of this particular mutation (Gronwald, 1994; Arntz et al., 2002; Vila-Aiub et al., 2009) (see Chapter 8 for further details on this). Concomitantly, the Ely plasmotype also had a significant main effect on the primary metabolite profile, generally reducing the abundance of primary metabolites (Figure 7.6). This reduction in primary metabolite abundance is congruent with reduced photosynthetic performance. Many of the metabolites which show clear reduction in abundance are sugars which makes sense from a physiological perspective, as if primary production is limited the plants are more likely to be source as opposed to sink limited, reduced metabolites contents would be an expected consequence of this (Figure 7.7).

What was surprising, however, was the highly significant main effect of Ely plasmotype on t50 (time until 50% of the seeds have germinated). The Ely plasmotype significantly reduced t50, thus increasing the rate of germination. Effects of triazine resistance on germination have been studied in foxtail millet (*Setaria italica*) (Darmency and Pernes, 1989) where no effect was found, in *Amaranthus* where it was found either to have no effect or reduce germination rates (Weaver and Thomas, 1986), and in Brassica, where resistance reduced rates of germination (Mapplebeck et al., 1982). Since none of the above mentioned examples caused an increase in the rate of germination it is possible that the triazine resistance is not causal but rather a separate mutation in either the plastome or chondriome is actually responsible. It is possible that this germination phenotype represents compensatory evolution, providing an adaptive advantage to the Ely accession in its native railway habitat and thus somewhat mitigating the negative impact of the *psbA* mutation. Previous studies have found a positive relationship between imbibed seed size and the rate of germination (Joosen et al., 2012) which is the opposite of what is found here, where Ely plasmotype reduces seed size but increases the rate of germination. This particular result will require follow up studies to elucidate the underlying cause, such as comparison of the Ely plasmotype sequence to that of the other accessions to identify potential candidate mutations. Ideally an independent Ser-264-Gly mutant could be used to generate a new triazine resistant plasmotype donor which would act as a control for the genetic background within which the mutation occurs.

Apart from the effects of the Ely plasmotype some of the other plasmotypes also had significant effects. The Shah plasmotype reduced Φ_{PSII} by as much as 0.008, which is larger

than the effect size of the largest QTL identified in Chapter 4, this is in line with a recent meta-analysis of the effects sizes of plasmotypic mutations generally being of moderate size (Dobler et al., 2014). The cause of this will require further study but it may be related to the capacity of Shah plasmotype to cause CMS in some genotypes (Gobron et al., 2013). The cause of this is thought to be a small mitochondrial peptide named *orf117Sha* the function of which is unknown, however apart from causing pollen death it may have other pleiotropic effects, of which reduced Φ_{PSII} could be one (Gobron et al., 2013). Both Ws-4 and Shah plasmotypes result in a significant increase in seed size when combined with the Bur-0 nucleotype, this is particularly interesting as seed size is a trait of key agronomic importance and the potential of cytoplasm as a breeding resources is rarely considered. One exception is a recent study of seed size in Soybean where 92 QTLs were detected pointing to plasmotype interactions, underlining the importance of plasmotype choice in a breeding program can be (Xu et al., 2011). The C24 plasmotype leads to a general increase in metabolite abundance and shows an opposite pattern to the Ely plasmotype (Figure 7.6), this fits well with previously noted parent of origin effect in C24 x Col hybrids where plants which have C24 as a mother are slightly larger than the reciprocal (Fujimoto et al., 2012).

The results presented here illustrate some of the effects of natural variation in the plasmotype. One of the advantages of the population described in thesis chapter is that it occurs in the model species Arabidopsis and thus the mechanistic causes for the plasmotype induced phenotypes can be more easily elucidated. Recent studies in Arabidopsis using reciprocal RIL populations have identified the plasmotype as a key epistatic hub for growth, secondary metabolites, and as a modulator of stochastic noise (Joseph et al., 2013; Joseph et al., 2013; Joseph et al., 2015). Understanding how the plasmotype can provide such a range of impacts is of relevance to both agriculture and evolutionary biology. There is an expanding body of research showing how important the plasmotype can be for plant adaptation (Budar and Fujii, 2012; Bock et al., 2014; Dobler et al., 2014). Examples such as adaptation to differing CO_2 availability (Iida et al., 2007) illustrate how relevant plasmotypic variation is for future adaptation to changing atmospheric CO_2. The role of plasmotype in key crop traits provides further motivation for this research (Weingartner et al., 2002; Dieckmann and Link, 2010; Xu et al., 2011). Our observations of the effects of the Shah plasmotype provide an interesting illustration: the Shah plasmotype reduces Φ_{PSII}, which –as a proxy for plant performance- is negative, but simultaneously leads to larger seeds, a desired trait. The creation of various new plasmotype donors during the development of the experimental panel provide a new way to

dissect these traits: the plasmotype donors can be used to generate DH populations in various plasmotypic backgrounds. Such populations which segregate for only two nucleotypes but multiple plasmotypes could be used to uncover unique cyto-nuclear interactions which represent hidden or cryptic genetic variation present in the nuclear genome (Gibson and Dworkin, 2004).

Future work on this genotypic set has already begun; firstly a more thorough analysis and integration of the data briefly skimmed in this chapter is underway as is the analysis of RNA-seq data for a subset of lines. The panel is also being screened under fluctuating conditions and more detailed physiological assays such as stomatal conductance are underway. This is only the tip of the iceberg as a recent publication shows that the plasmotype, and in particular the chloroplast, plays a central role in plant defence (de Torres Zabala et al., 2015). This panel will allow us to screen for variation in chloroplast immune functioning potentially opening up a new breeding target for plant pathogen resistance. A more detailed analysis of natural genetic variation and perhaps an expansion to a wider set of genotypes would be interesting as it will provide empirical evidence for the level and nature of evolutionary conservation of nucleo-plasmotype interactions within a species. Providing a new research method which can assess which nucleo-plasmotype interactions show divergent evolution and which are under evolutionary constraint. The panel described here is a uniquely powerful resource for the study of these neglected genomes (Pesole et al., 2012; Sloan, 2014) and it will provide the basis for many future research endeavors to characterize nucleotype-plasmotype interactions.

Acknowledgements

Duur Aanen is acknowledged for helpful discussions, and Jonas Klasen for help with statistical analysis. Rob Dirks and the Rijk Zwaan breeding company (Fijnaart, the Netherlands) are gratefully acknowledged for their help in genotyping.

Note: For Supplementary Data s1 please email flood@mpipz.mpg.de.

Supplemental material

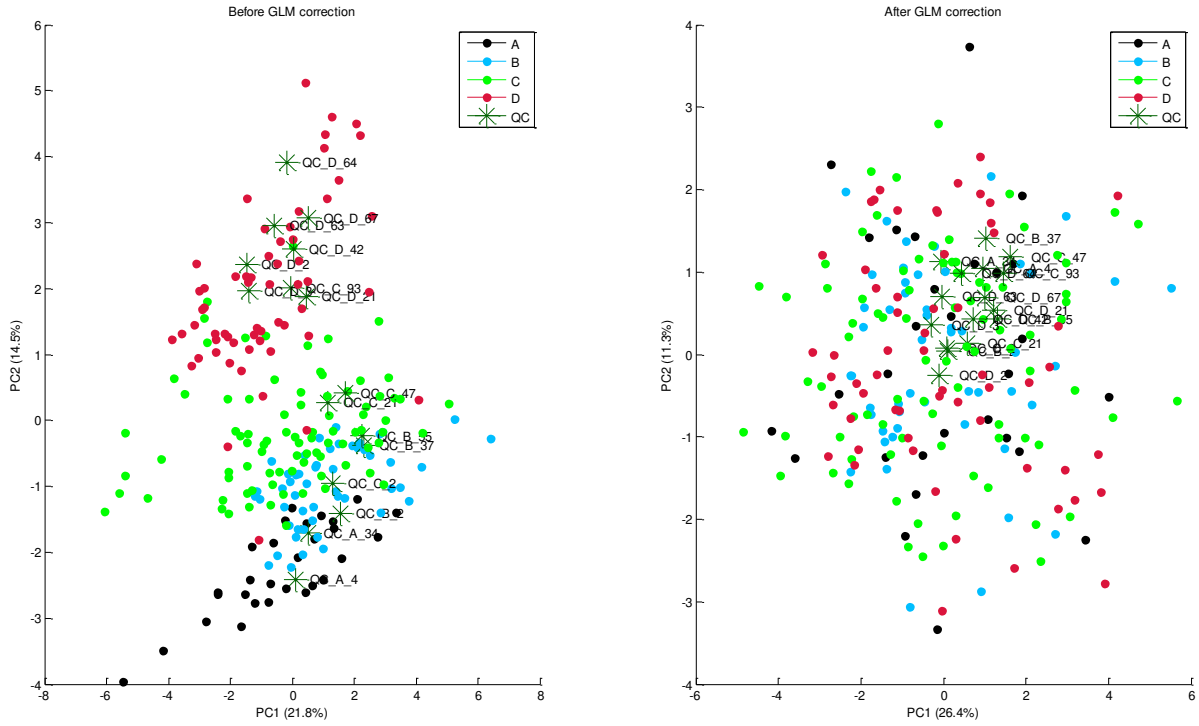


Figure s7.1: *Correction of metabolite batch effect using general linear models, colors represent different batches, stars indicate reference (quality control) samples.*

Appendix 7.1

Example of linear model comparisons for average Φ_{PSII} through time (FqFm_average), the generic format is: model name ← linear model of (responsible variable ~ 1 + factorA + factorB, data=dataset)

fit0 ← lm(FqFm_average ~ 1+Basin, data=b6) #this is the null model (reduced model)

fit1 ← lm(FqFm_average ~ 1+Basin+Table.pos, data=b6) # fit0 vs fit1 p = n.s. (not significant)

fit2 ← lm(FqFm_average ~ 1+Basin+Table.pos+Image.pos, data=b6) # fit0 vs fit2 p=0.000869 # fit1 vs fit2 p=0.000571

fit3 ← lm(FqFm_average ~ 1+Basin+Table.pos+Image.pos+x_within_image, data=b6) #fit2 vs fit3 p = n.s.

fit4 ← lm(FqFm_average ~ 1+Basin+Table.pos+Image.pos+x_within_image+y_within_image, data=b6) #Fit 3 vs Fit4 p = n.s.

fit5 ← lm(FqFm_average ~ 1+Basin+Table.pos+x_within_image, data=b6) #fit1 vs fit5 p = n.s.

fit6 ← lm(FqFm_average ~ 1+Basin+Table.pos+y_within_image, data=b6) #fit1 vs fit6 p=3.832e-05

fit7 ← lm(FqFm_average ~ 1+Basin+Table.pos+y_within_image+Nuclear, data=b6) #fit 6 vs fit7 p< 2.2e-16

fit8 ← lm(FqFm_average ~ 1+Basin+Table.pos+y_within_image+Nuclear+Cytoplasm, data=b6) # fit7 vs fit8 p< 2.2e-16

fit9 ← lm(FqFm_average ~ 1+Basin+Table.pos+y_within_image+Nuclear+Cytoplasm+Nuclear:Cytoplasm, data=b6) #fit8 vs fit9 p=7.651e-13

N.B. 'Nuclear:Cytoplasm' means test for nuclear and plasmotype interaction

Then run ANOVA(compare model1 to model2) which is written as:

anova(fit0, fit1)

Chapter 8

Whole genome hitchhiking on an organelle mutation

Pádraic J. Flood[1,2,3], Joost van Heerwaarden[4], Frank Becker[1], C. Bastiaan de Snoo[5], Jeremy Harbinson[2] & Mark G.M. Aarts[1]

[1]*Laboratory of Genetics, Wageningen University, Wageningen, the Netherlands.*

[2]*Horticulture and Production Physiology, Wageningen University, Wageningen, the Netherlands.*

[3]*Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany.*

[4]*Plant Production Systems, Wageningen University, Wageningen, the Netherlands.*

[5]*Rijk Zwaan R&D Fijnaart, Fijnaart, the Netherlands.*

Corresponding Author: Pádraic J. Flood (flood@mpipz.mpg.de)

Strong selection on beneficial mutations can increase the frequency of linked genomic regions (Andersen et al., 2012; Cutter and Payseur, 2013; Neher, 2013). In sexual organisms, the impact of this so-called genetic hitchhiking (Maynard-Smith and Haigh, 1974) is moderated by recombination, which breaks the linkage between selected loci and the rest of the genome. In self-fertilizing species this effect of recombination is reduced, potentially causing selection to extend to unlinked parts of the genome, even to the point that selection on organelles may affect nuclear diversity (Charlesworth and Wright, 2001). Although potentially of great evolutionary importance, such indirect selective sweeps have never been observed in nature. Here we show that strong selection on a chloroplast gene in the inbreeding plant species *Arabidopsis thaliana* has caused widespread and lasting hitchhiking of the whole nuclear genome. The selected allele confers herbicide tolerance and has spread more than 400 km along the British railway network, reshaping the genetic composition of local populations. This demonstrates that variation in organelle genomes is non-neutral (Bock et al., 2014; Dobler et al., 2014) and can strongly affect nuclear genetic diversity. We expect that selection on organelles affecting evolution of the nuclear genome is much more common than previously thought and needs to be considered when studying genome evolution.

In outcrossing species the extent to which selection on a locus in the nuclear genome can distort neighbouring gene frequencies is restricted by recombination (Cutter and Payseur, 2013; Neher, 2013; Slotte, 2014). In non-recombining genomes, such as those of mitochondria, chloroplasts, and bacteria, selection at a single locus can, however, result in genome-wide selective sweeps often resulting in clonal interference (Lang et al., 2013). There is increasing evidence that organellar genomes may be subject to both positive and background selection (Bazin et al., 2006; Bock et al., 2014; Dobler et al., 2014). While the potential consequences of selection on organellar genomes have been frequently discussed (Bock et al., 2014) there has been little recognition that in partially self-fertilizing species the effect of selection on organelles may extend to the nuclear genome as well (Wright et al., 2008).

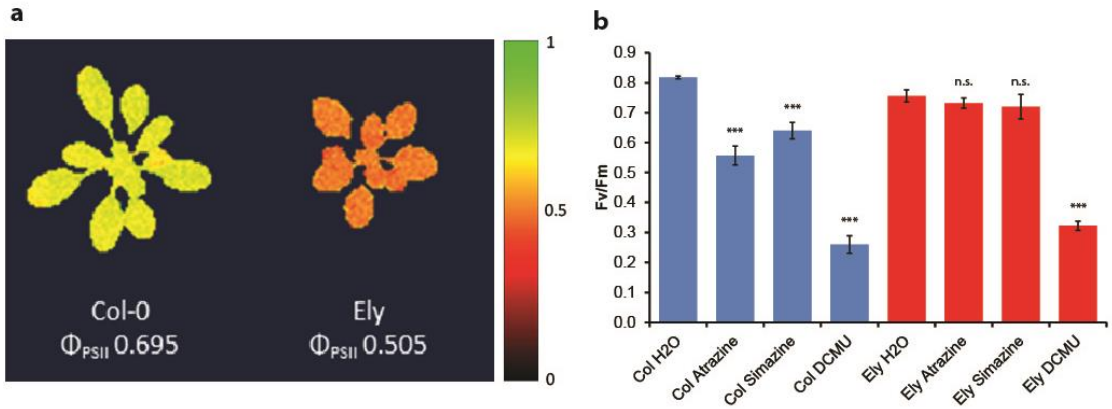


Figure 8.1. Photosynthetic performance of resistant and susceptible plants. *False colour chlorophyll fluorescence image of light-adapted photosystem II operating efficiency (Φ_{PSII}) (**a**),and dark adapted chlorophyll fluorescence (F_v/F_m) of susceptible (Col-0) and resistant (Ely) genotypes in response to incubation in H_2O, atrazine, simazine and DCMU (**b**) *** = $P < 0.0001$, n.s.= not significant.*

Here we report the natural distribution of a chloroplast mutation that is likely to have been under strong anthropogenic selection. The mutation confers resistance to triazine herbicides, which were widely used along British railways from 1957 until their discontinuation in 1992 over environmental concerns. Triazines inhibit photosynthesis by competing with plastoquinone for the Q_B binding site of the D1 protein of photosystem II (PSII), thereby inhibiting photosynthetic electron flow (Gronwald, 1994). The D1 protein is encoded by the chloroplastidic *psbA* gene, and a Ser-264-Gly amino acid substitution confers resistance via the removal of a hydrogen bonding site which prevents triazines from binding to D1 (Gronwald, 1994). This particular amino acid substitution has evolved at least 73 times across multiple species (Heap, 2015). In addition to conferring resistance to triazine herbicides this substitution also reduces the affinity of the Q_B site for plastoquinone, thereby reducing the efficiency of PSII (Gronwald, 1994) (Chapters 3, 5 and 7) (Figure 8.1). This reduction in PSII efficiency limits carbon assimilation under most environmental conditions (Darmency, 2013) and consequently results in a fitness cost (Warwick and Black, 1994; Arntz et al., 2000; Powles and Yu, 2010). Resistant genotypes are therefore expected to be removed from the population in the absence of triazine application (Warwick and Black, 1994; Arntz et al., 2000).

We assessed the dispersal and persistence of triazine resistance in several British populations of Arabidopsis, starting with a population at Ely railway station, Cambridgeshire, UK, in which triazine resistance was first detected in 1988 and after which the resistant genotype is named (El-Lithy et al., 2005). The cessation of triazine application, and their subsequent removal via active degradation (Shapir et al., 2007), is expected to result in a reduced relative fitness of resistant genotypes and their gradual loss from the environment.

To estimate the frequency of triazine resistant genotypes along British railways in the east and south west of England, we sampled populations on or next to railway tracks (railway populations) and at >500m from the nearest railway track (non-railway controls). Of the 573 plants assayed 51 carried the *psbA* mutation (Figure 8.2 and Supplemental Data 8.1). These resistant plants were found at 12 of the 47 sampled locations and were all located on or next to railway tracks. The mean frequency of resistance near railways was 0.12. None of the 24 non-railway populations yielded any resistant plants, confirming that the resistance is positively associated with railways ($p<0.0001$) and thus it is likely to be under human-mediated selection. Genotyping of the 51 resistant and 342 non-resistant plants for 30 polymorphic, nuclear SNPs (Supplemental Figure 8.1 and Supplemental Data 8.1) showed that all resistant plants belonged to one unique multi-locus haplotype (Supplemental Figure 8.2). The expected population frequency of such a haplotype by random association of SNPs is below 0.0001, strongly suggesting that these 51 identical haplotypes originated from a single ancestor. A comparison with different published genomic datasets failed to pinpoint a likely site of origin of the Ely genotype (Supplemental Figure 8.3), which is perhaps due to the generally flat population structure of British Arabidopsis (Platt et al., 2010). While the resistant haplotype was not the only one found at multiple locations, it was the most common and the only to be positively associated with railways. Additionally, from an existing dataset of 149 nuclear SNPs we identified four accessions from Liskeard, Cornwall, over 400 km from Ely (Figure 8.2 and Supplemental Data 8.1), that were identical to the Ely genotype for all 149 SNPs. Progeny of all these four accessions tested positive for the *psbA* mutation and exhibited atrazine resistance (Supplemental Data 8.2), providing further evidence that resistance is associated with a single nuclear haplotype.

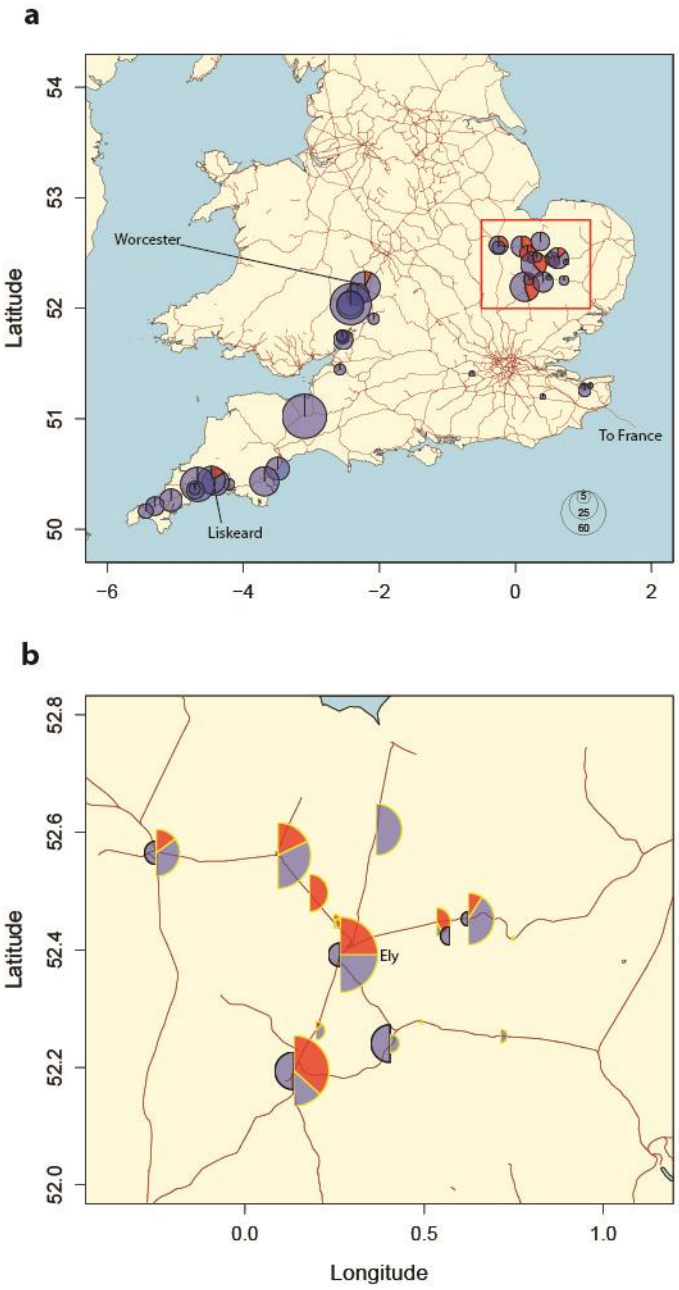


Figure 8.2. Map showing distribution of plants sampled. *Locations outside the red box where resistant plants were found are labelled (**a**). Plot of area surrounding the site of the original Ely accession (red box in (**a**)). Left semicircle with black circumference represents accessions from non-railway sites and right semicircle with yellow circumference represents accessions from railway sites (**b**). Blue represents susceptible plants and red represents resistant plants; railways are indicated with brown lines.*

Based on these results, we conclude that triazine resistance arose once among the tested populations and then spread in a single genetic background over a large area (Figure 8.2). The association of the triazine resistant genotype with railways likely facilitated its dispersal, as has been documented for other plant species (Blanchet et al., 2014). The current British rail network comprises 32,000 km of track and 30,000 ha of lineside vegetation (Network Rail, 2015). During the period of triazine application (at most 35 years), the habitat available to resistant Arabidopsis was large and probably free from competition as no other triazine resistant species has been reported on British railways (Heap, 2015).

The single origin of all resistant genotypes is striking, as the entire nuclear genome can be said to have hitchhiked along with this chloroplast mutation (Maynard-Smith and Haigh, 1974). This nuclear genetic hitchhiking was most likely facilitated by the strength of selection, the reduction in effective recombination due to the high level of inbreeding in Arabidopsis, and the haploid nature of the chloroplast genome which renders all mutations dominant. This study provides a particularly favourable case to detect an organelle-mediated selective sweep, as the occurrence of natural selection on a single organellar locus can be demonstrated by virtue of prior knowledge of the mutation, localized herbicide application, and the possibility of sampling systematically along the railways where the selection pressure was known to occur. We thereby overcome the usual limitations on inferring selection from population genetic data, in which observed patterns of diversity can often not be disentangled from those caused by demographic history (Bank et al., 2014). Nevertheless, none of the three conditions leading to nuclear genome hitchhiking - strong selection, inbreeding and haploid organelles - are unique to the case we describe. Inbreeding is common, organelles are always haploid and strong selection is common and expected to become more so in heavily human influenced environments (Palumbi, 2001; Andersen et al., 2012). Thus we expect that organelle-mediated selective sweeps resulting in nuclear genome hitchhiking are more common, though difficult to detect empirically.

A remaining question is why this population of photosynthetically-impaired plants persisted for 22 years after the cessation of triazine use. In crops, the *Ser-264-Gly* mutation reduces seed yield by 22 to 36% (Darmency, 2013). Although many crops exhibit photosynthetically limited growth (Flood et al., 2011; Darmency, 2013) this is not thought to be the case in natural ecosystems, where resource restriction and competition are more likely to be limiting (Fatichi et al., 2014). The picture that emerges for natural populations is that the fitness cost of triazine resistance in the absence of triazines is environment specific, being less at low temperatures and low light levels (Plowman et al., 1999) and greater in the presence of biotic stresses (Gassmann and Futuyma, 2005; Salzmann et al., 2008). A simple population genetic model shows that given the published estimates of migration of 1-2% per year (Falahati-Anbaran et al., 2014), and a selective cost of 0.22-0.36 it would take an estimated 10-19 generations to obtain frequencies equal to or higher than those observed for railway populations. Due to seed bank persistence and recruitment, Arabidopsis has an average generation time of 3-4 years (Lundemo et al., 2009). This would equate to 30 to 76 years to reach observed frequencies, which is fully compatible with the time since discontinuation of atrazine. In addition, the 35-year period of triazine use allowed the build-up of monogenic stands of

resistant individuals with a large seed bank to provide resistant populations with an additional buffer against extinction (Falahati-Anbaran et al., 2014).

The evolution and spread of this atrazine resistant genotype along British railways is a conspicuous example of anthropogenic disturbance greatly altering the population structure of a wild plant species (Banks et al., 2013). This organelle-mediated sweep is a striking example of linked selection, reaffirming the significant role that selection can play in genome evolution (Corbett-Detig et al., 2015) and how breeding systems and organelles have a large role to play in nuclear genome evolution.

Materials and Methods

Plant collection and cultivation. Seed was collected in August 2012 around Ely railway station and again at Ely and additional sites in Cambridgeshire and Suffolk during May and June 2013. A final round of sampling was conducted in May 2014 in the south-west of England, from Cornwall to Herefordshire. When collecting, both leaf and seed material were taken on or near railway-associated land and, where possible, additional samples were taken some distance away (> 500 m). This was done in order to sample a putatively local population not associated with the railway. GPS data were recorded for all plants sampled. Sampled leaf material was stored with silica gel to ensure rapid desiccation and optimal preservation of the DNA. Upon return to the laboratory, seed was sown on wet filter paper and stratified at 4° C for four days to break dormancy. The seed was then sown in a greenhouse on rockwool blocks (www.grodan.com). Depending on seed availability up to four seeds were sown per block with four blocks per genotype. Extra seedlings were removed after two weeks leaving one plant per block, and one leaf was taken from each accession for genotyping. Genotypes which showed no signs of flowering after five weeks where moved to a cold room for eight weeks of vernalization. Seed was harvested from all lines and stored (see Supplemental data 8.1 for full list).

Genotyping. A 350-bp region around the chloroplast *psbA* mutation was amplified using PCR. The PCR product was then digested overnight at 37°C using the restriction enzyme FspBI which only cleaves wild-type sequences, resulting in 210- and 140-bp fragments. Genotypes with the mutated sequence were not cleaved allowing easy identification of resistant genotypes (Supplemental data 8.3). 394 lines were further genotyped for 39 nuclear SNPs described previously (Wijnker et al., 2012). 9 SNPs were excluded from further analysis as they were

either monomorphic or had more than 10% missing data. The remaining 30 SNPs were used in all further analysis (Supplemental data 8.1 and Supplemental Figure 8.1).

Phenotyping. One resistant and one susceptible genotype (based on the PCR assay previously described) from each location sampled were sown on rockwool blocks. Three weeks after sowing leaves were cut and placed in a petri dish containing wet filter paper, soaked in either water or a 100µM solution of atrazine. The leaves were vacuum infiltrated as described in (El-Lithy et al., 2005), dark-adapted for 30 minutes, upon which the chlorophyll fluorescence parameters F_o and F_m were measured, and the F_v/F_m value calculated (Baker, 2008) using a FluorCam 800MF chlorophyll fluorescence imager (www.psi.cz). In susceptible genotypes treated with atrazine the fluorescence measuring beam closes some PSII reaction centres, resulting in a higher F_o (strictly no longer an F_o) and a decreased F_v/F_m. Resistant plants show no decline allowing easy assessment of atrazine tolerance. Cross tolerance to simazine was confirmed using Ely and Col-0 only (Figure 8.1). These genotypes were also assayed for resistance to an alternative PSII inhibitor DCMU for which the Ser-264-Gly mutation in D1 should provide no cross-tolerance (Figure 8.1 and Supplemental data 8.2).

Relatedness and geographic assignment. The relatedness matrix K was calculated as in Patterson et al. (2006) as: $K = \frac{1}{n}MM'$, where each element in the matrix M is given by $M_{ij} = \frac{C_{ij-\mu_j}}{\sqrt{p_j(1-p_j)}}$ with C_{ij} being the number of allele copies for individual *i* at locus *j* and μ_j the mean allele count and $p_j = \mu_j/2$ the population allele frequency for locus *j*. From the relatedness matrix calculated with the 30 SNPs, a neighbour-joining tree was constructed (Supplemental Figure 8.2). Continuous geographic patterns of multi-locus variation were modelled by spatial interpolation using 4377 regularly spaced grid points selected within a polygon encompassing the United Kingdom and a part of continental Europe. First, the 149 SNP marker matrix for 2567 geo-referenced individuals was reduced to 40 independent axes of variation (PCs) by principal component analysis (PCA). After excluding the Ely accession (also known under accession codes PHW31, CS28631 and 7502) from the matrix, PCs were averaged over 483 unique locations and a geospatial model was fitted to each PC using the function *Likfit* in the R package geoR. The predicted value for individual PCs at each grid point was then generated by model-based kriging, using the function *krige.conv* and using the fitted spatial models as an input. Finally the Euclidean distance along the 40 PCs between each grid point and Ely

genotype was calculated for every grid point, resulting in a map of estimated genetic distance to the Ely genotype (Supplemental Figure 8.3).

Testing for association with railways. Haplotype frequencies were calculated for the 393 individuals genotyped for 30 polymorphic nuclear SNPs, excluding the Ely reference genotype. The expected haplotype frequency under the assumption of independence between loci was calculated for each haplotype as the product of allele frequencies. At each location, haplotype frequencies were calculated for railway populations and non-railway controls. We tested for positive association with railways for the Ely haplotype and the twelve haplotypes occurring in more than one location, using a generalized linear model with a binomial link function.

Numerical simulation of persistence. Persistence of the mutation was simulated with a deterministic model assuming migration from an infinite wild-type population into a mutant population and subsequent selection against the mutant allele. Starting with a wild-type frequency, p, of zero, and assuming p=1 for the infinite wild-type population, each generation would see a change in wild-type frequency according to $p' = p(1-m) + m$ and $p'' = p'/(1+s(p'-1))$, with m and s being the proportion of immigrants and the selection coefficient respectively.

Supplemental files: The following supplemental files are to be found in the compact disc at the back of this thesis.

Supplemental Data 8.1. All lines collected, geographical coordinates, resistance and genotype data.

Supplemental Data 8.2. Chlorophyll fluorescence measurements of accessions with and without the Ser-264-Gly mutation in response to atrazine, simazine and DCMU.

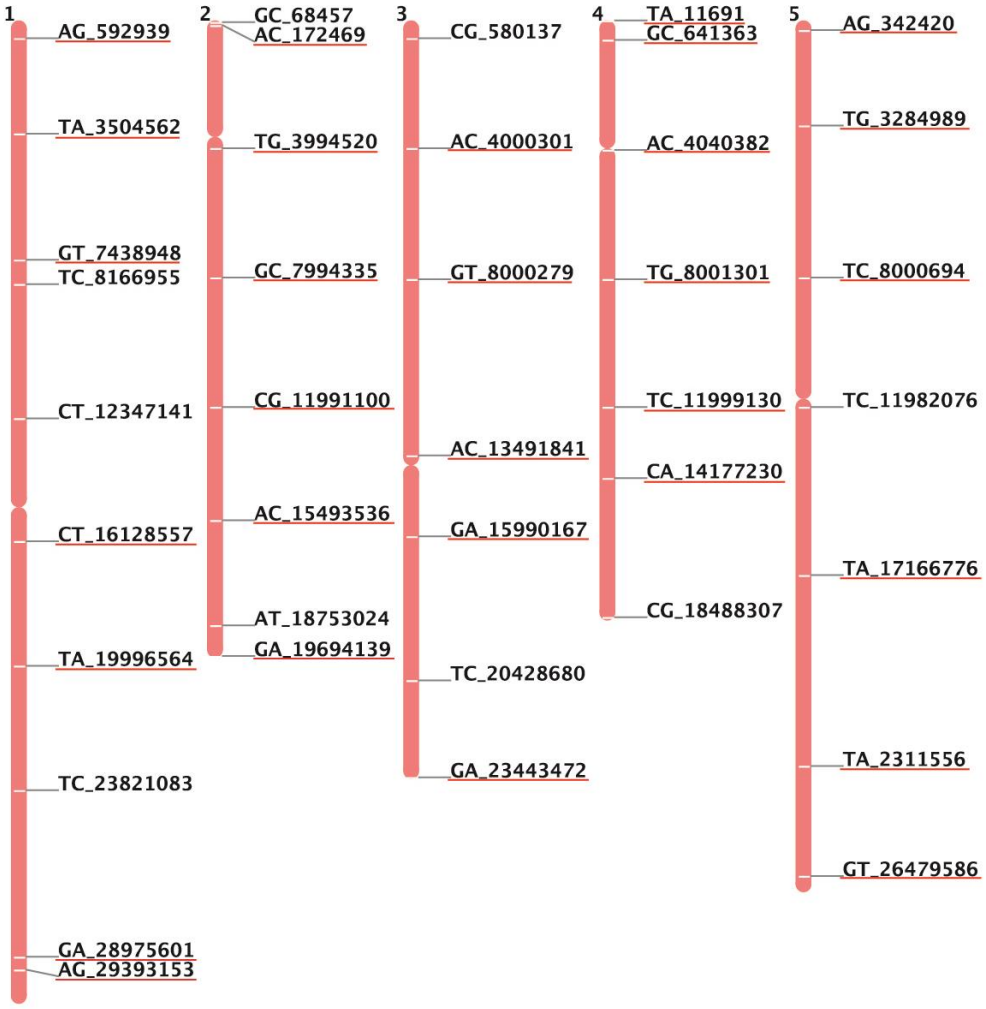
Acknowledgments

We thank P. Williams for details on the original collection site for Ely and E. Holub for details on lines collected from Liskeard. We thank J. Hoedemaker, S. Schoustra and N. Aarts for help with seed collection, J. Snel for advice on conducting the atrazine treatments, A. Wright and D. Hayward for information regarding current and past herbicide use on British railways, T. Morton, J. Bergelson, and M. Horton for seed of Ely relatives and details regarding their origin, T. Bisseling, J. Kirwan, M. Koornneef, E. Wijnker and B. Zwaan for comments on the manuscript.

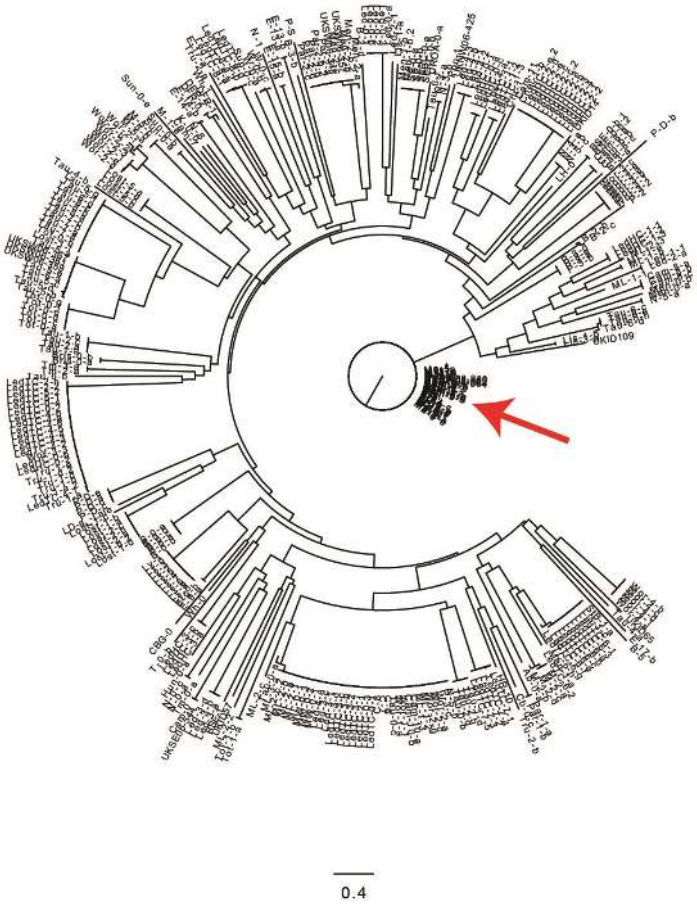
Author Contributions P.J.F., J.H., and M.G.M.A., conducted fieldwork. P.J.F., and F.B., propagated, genotyped and screened material. C.B.dS., conducted SNP genotyping. P.J.F., and J.vH., analysed data. P.J.F., J.H., M.G.M.A., and J.vH., designed the study and wrote the paper with comments from all other authors.

Author Information The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.J.F. (flood@mpipz.mpg.de).

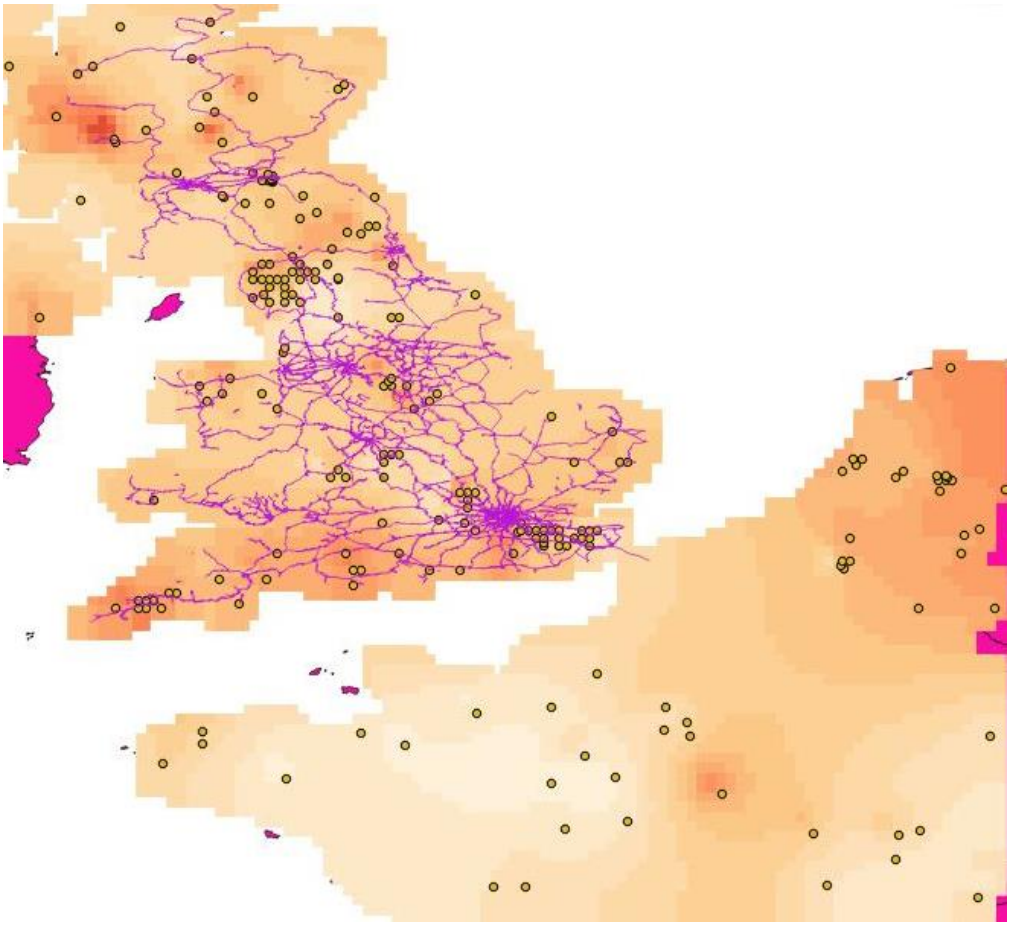
Supplementary Material



Supplemental Figure 8.1. Distribution of 39SNPs across the nuclear genome. *For each SNP the first letter indicates the Col-0 allele and the second the Ler-0 allele, the number indicates the position of the SNP in base pairs on each chromosome. The 30 polymorphic SNPs are underlined in red and each vertical line represents one of the five chromosomes, numbered one to five from left to right.*



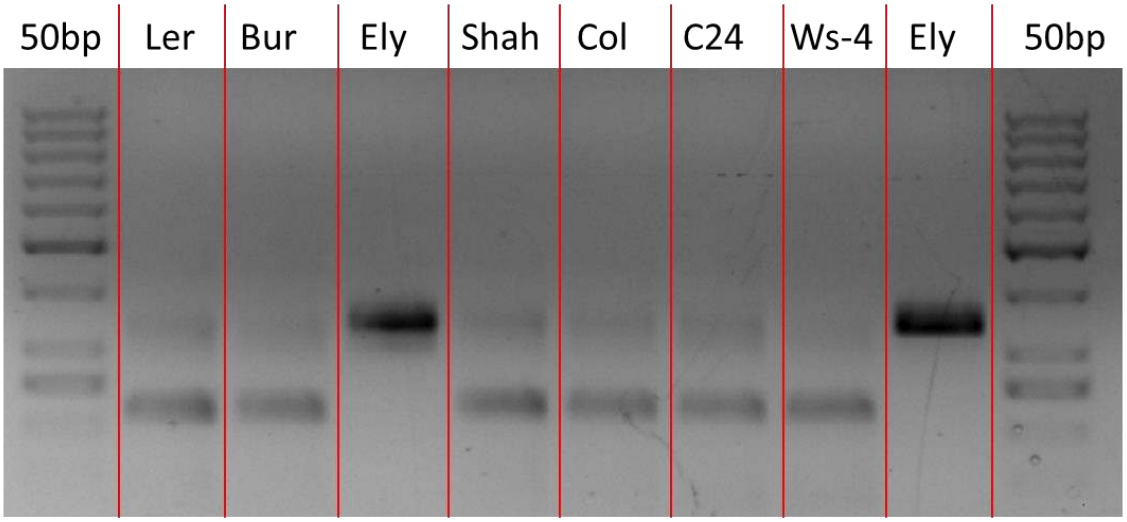
Supplemental Figure 8.2. *Neighbourhood joining tree of plants genotyped with 30SNPs, the cluster of genetically identical plants indicated with the red arrow represents all resistant plants.*



Supplemental Figure 8.3. *Genetic frequency map showing likelihood of origin of Ely genotype. The lighter the shade the higher the likelihood of the genotype originating from this area. Yellow circles show the origins of plants genotyped with 149 SNPs (Platt et al., 2010). Purple lines indicate the position of railway tracks.*

Supplemental Data 8.3. *Fragment size difference post FspBI digest for wild type and Ser-264-Gly mutant genotype. Wild type psbA gives 210- and 140-bp fragments and the resistant allele gives a 350-bp fragment. PCR products were loaded on 1% agarose gel with 2µl of 50bp GR as ladder.*

Gene	**Sequence**	**F/Rev**	**Length (bp)**	**Tm (ºC)**	**GC (%)**	**Product size (bp)**
psbAF	CTATGCATGGTTCCTTGGTAACTTC	For	25	54	44	350
psbAR	CGTTCATGCATAACTTCCATACCA	Rev	24	54	41.6	350



Chapter 9

General Discussion

"And so to conclude and to finish disputes"[23]

[23] From the traditional Irish song Arthur McBride, check out Paul Brady's 1977 rendition, it is on YouTube and it is wonderful.

Introduction

In this thesis I[24] have assessed natural genetic variation in plant photosynthesis with a diverse range of approaches. To begin I reviewed the existing literature (Chapter 2), based on this it was clear that a key limitation to the exploration of photosynthetic variation was the ability to screen large populations of plants. To overcome this I helped to develop a phenotyping platform, described in Chapter 3. Once this platform was fully functional I used two quantitative genetic approaches, recombinant inbred line mapping (Chapter 4) and genome wide association mapping (Chapter 5) to further characterize the variation present. Additionally, I explored natural variation in the phosphorylation of photosystem II (PSII) proteins in response to different light levels. This was not amenable to high-throughput screening and instead required more detailed physiological assays (Chapter 6). Results from Chapter 6 and Chapter 5 indicated that the plasmotype had a significant role to play in natural variation in photosynthesis. In Chapter 7, I developed an approach to disentangle such nucleo-plasmotype interactions. Finally, in Chapter 8, I explored one such photosynthetically impaired plasmotype in the field, this revealed a novel evolutionary insight that selection on organelles can result in nuclear genetic hitchhiking, a previously undescribed phenomenon that is likely to be quite common. In this discussion, I will aim to both distill and synthesize the main findings and insights from the previous chapters, and provide suggestions of some potential future research avenues, which have opened up as a direct result of the insights and resources developed.

Phenomics and the next bottleneck

"Arguably, empirical progress results mainly from better measurement." - David Houle et al. (2011)

In the recent aftermath of the genomics revolution it has almost become a cliché to say that the research bottleneck has moved from genotyping to phenotyping, however, as with many clichés it is based upon an element of truth[25]. In the case of photosynthesis, this is particularly pertinent as it is not a phenotype which lends itself to easy quantification. In fact, biology in general, seems to be limited more by measurement of phenotypes than anything else. The traits about which we know the most are those which are easiest to measure, for example height in humans, and flowering time in plants. This recent cliché of claiming a bottleneck is not a new

[24] I write 'I' in this discussion, but of course the work discussed was a collaborative effort.
[25] In Chapter 3 of this thesis, I make full and unashamed use of this cliché; such is the nature of scientific rhetoric.

phenomenon, it is just that the contrast has greatly increased. We now have a vast wealth of genomic information, which is of limited use without the corresponding phenomic information. To bridge this looming gap there have been numerous calls for the developments of phenomic technologies (Houle, 2010; Houle et al., 2010; Furbank and Tester, 2011; Pieruschka and Poorter, 2012; Dhondt et al., 2013). As a result, many new technological developments have been developed (Skirycz et al., 2011; Tisné et al., 2013; Dhondt et al., 2014) of which Chapter 3 of this thesis is one. The introduction and design consideration sections of this chapter reviews this literature, with a specific focus on phenomic technologies for photosynthesis. Instead of repeating this, in the discussion I will mainly focus on what we do with all of this data once it is created, a problem which is becoming more and more apparent.

"Memories are the key not to the past, but to the future." - Corrie Ten Boom

When I started research on this thesis (July 1st 2009) a one terabyte external hard drive was a thing to behold, it was hard to imagine why you would ever need so much storage space, yet by the end of my PhD I have filled four such hard drives. This is illustrative of the rapid and ever increasing rate of data generation. In the field of genomics, data storage and maintenance has already become a serious bottleneck, with several community level initiatives set up to address this issue (Kodama et al., 2012; Nakamura et al., 2013). In addition to data storage, data reduction is becoming increasingly important, for example deleting all sequence data, which is not polymorphic with the reference genotype (Christley et al., 2009; Hsi-Yang Fritz et al., 2011). In fact, recent developments suggest that the long-term data storage medium of the future will be DNA itself (Church et al., 2012; Goldman et al., 2013). So perhaps the genotype databases of the future will be the seed banks of today. However, as already mentioned in the preceding paragraph, generation of phenotype data will not be as simple as re-sequencing the genotype. It is unlikely that phenotyping large populations of plants, or other organisms for that matter, will ever be cheaper than storing the data. The time taken to phenotype, i.e. a growing season or longer (think trees) is also much greater than the time taken to obtain sequence data, and as the next generation sequencing era proceeds, this gap is set to widen (Stephens et al., 2015). Thus, although in the near future it may be cheaper to store genotype information in the seed, phenotype data still requires a more accessible storage format[26]. The nature of this storage format is non-trivial. For genomic data, annotation strategies and descriptive standards are in place so that almost the complete spectrum of genomic elements can be described in a database

[26] Long term storage of phenotype data for large germplasm collections will probably be done using DNA however this will be *in vitro* not *in vivo* (seed).

searchable format, for phenotypes however, this is far from the case (Deans et al., 2015). This is in a large part due to the nature of phenomic data, which occupies a highly multidimensional space for which we even lack linguistic capability of describing, never mind a standardized archival format. There are however recent developments to address this, both in the plant research community and the wider biological community (Deans et al., 2015; Oellrich et al., 2015). These efforts are timely indeed, as the availability of phenotype data is absolutely essential if the genomics revolution is to reach its full potential and, in the case of plant biology, provide the food needed to sustain an increasing world population (Zamir, 2013). In this particular case, success of the endeavor does not rest entirely on clever ideas and technological innovation but on community participation (Roche et al., 2014). Data storage will only work if researchers are willing to share what they have. This is becoming more widely accepted for genomic data but has quite some way to go for phenomic data, especially with regards to crop plants (Zamir, 2013).

Linking genotypes to phenotypes (and vice versa)

Even the most exquisite phenotype dataset, stored in a coherent and accessible way is for naught if the information within cannot be put in a biological context. The linking of phenotype with genotype, commonly referred to as the genotype-phenotype map, is arguably the key research endeavor of the field of genetics. In the last three decades this field has undergone a revolution coinciding with the development of techniques to link genetic polymorphisms with phenotype data (Lander and Botstein, 1989). These techniques allowed for the expansion of molecular genetic analysis and have culminated in the recent explosion of genome wide association studies (GWAS). In this thesis, I conducted both family-based mapping, using recombinant inbred lines (RILs), and genome wide association mapping using a diverse set of natural accessions. Here, I will discuss the relative merits and limits of such genetic mapping approaches and consider what next for photosynthesis.

In his landmark publication establishing the modern field of quantitative genetics, Fisher proposed an infinitesimal model, which distributes the genetic basis of a quantitative trait over a very large number of loci, each with very small effect (Fisher, 1918). This model served as the basis for most quantitative genetic theory until recent years. However, an implicit assumption of both QTL and GWA studies is that the allelic variation governing the trait of interest is not composed of just infinitesimals, but rather loci of significant magnitude and thus detectable (Marjoram et al., 2014). The success of identifying QTLs and even cloning some of

them has led many to think that this assumption is indeed valid (see Alonso-Blanco et al. (2009) for a review of QTLs identified in Arabidopsis). But are these loci of large effect representative of the allelic variation present in nature or do they only represent a particular subset of mutations? For example, a rare mutation particular to a parental line (El-Din El-Assal et al., 2001; Coustham et al., 2014), or a mutation under balancing selection (Todesco et al., 2010; Huard-Chauveau et al., 2013; Karasov et al., 2014). Rockman (2012) argues that such large effect mutations are not representative of the variation underlying complex traits, which is more likely to be in regulatory elements and exhibit non-Mendelian inheritance patterns, i.e. the phenotypic impact of the variant is largely determined by the genetic background within which it occurs. Such allelic variation could account for the phantom heritability described by Zuk et al. (2012), as well as explaining the relatively poor performance of GWA for complex traits such as photosynthesis (our most significant association contains *ACD6* that is known to be under balancing selection (Todesco et al., 2010), which is one of the few evolutionary forces that can actively maintain the presence of large effect mutations in a population). Photosynthesis is a trait which is likely to be under directional or stabilizing selection, in either case it can be reasonably assumed that most genotypes in existence will be 'close' to a phenotypic optimum[27], conditions under which Fisher's geometric model will apply and infinitesimals can be expected. Thus, with the exception of a few mutations of large effect, the majority of the variation is likely to be spread over a very large number of loci each contributing a very small amount to the variation present. An interesting question that now arises is, how small can the effect size of a mutation be before it becomes invisible to evolution?

The effective population size (N_e) of Arabidopsis is estimated to be somewhere in the range of 250,000 to 300,000 (Cao et al., 2011). Since natural selection has the capacity to select on mutations with an effect size of $1/N_e$ the effects sizes of evolutionarily relevant loci can in theory be as little as 3.3×10^{-6} which is far below our current detection thresholds. Based on this observation it is perhaps unsurprising that only a small number of loci were above the Bonferroni threshold, and, if as has been suggested (Rockman, 2012), these infinitesimals tend to be largely epistatic, then the chances of detection are further reduced. Nonetheless, this simple calculation does illustrate that the capacity for natural selection to select positive variants is large, and it is indeed possible for mutations of very small effect to be selected for (Olson-Manning et al., 2012). If we wish to 'see' all of the loci that natural selection can 'see' then we

[27] This optimum does not necessarily entail a high photosynthetic capacity but rather a capacity and flexibility suitable for the environment where the species occurs.

would need a population of a similar size. Whilst an experiment involving 250,000 Arabidopsis genotypes may initially seem beyond our wildest dreams, it is not impossible. In maize, nearly 1,000,000 plants were screened for flowering time in a recent experiment (Buckler et al., 2009), and in humans we have sequenced almost 1,000,000 genomes (Stephens et al., 2015). With 14 of the phenotyping platforms described in Chapter 3 a small team of researchers could easily screen 250,000 Arabidopsis genotypes in a year, also allowing for replicated genotypes to control for experimental variation. The resources are available, and the potential of Arabidopsis as a model for such studies is just beginning to be realized, but for this to happen, a significant proportion of the community will need to change mindset and follow the example of physics and medicine: it is time for big science on a small plant. Initiatives such as the 1001 genomes (Weigel and Mott, 2009) are a start, but we need something much bigger and much bolder if we are to achieve our stated research goals, namely, resolving the genotype-phenotype map to an almost particulate detail. As one of the few truly wild research organisms Arabidopsis is primely positioned, as the results are likely to be of more direct evolutionary and ecological relevance than similar studies in domesticated species such as maize and rice. Additionally, its naturally inbreeding nature means wild collected accessions are often fully homozygous, which facilitates easy replication of the same genotype across multiple environments. Also its small plentiful seeds allow for easy exchange of germplasm between research groups, something which is essential for such large scale research endeavors. Finally as John L. Harper famously noted *"Plants stand still and wait to be counted"*, meaning they are amenable to phenotyping approaches that are impractical or impossible in animals[28].

The brief overview of the literature and the current state of knowledge regarding the expected genetic architecture of polygenic traits, such as photosynthesis, fits with the results of Chapters 4 and 5. In Chapter 4 three RIL populations were screened for photosynthetic efficiency under growth chamber conditions. In each population QTLs were found, and with the exception of the Can x Col population, they explained relatively little of the variation present. In Chapter 5 a similar study was conducted using the GWA approach, again significant loci were detected and again they explained relatively little of the variation. The amount of variation these detected loci explain is probably overestimated due to the winner's curse (Beavis, 1994). Taken together this indicates that whilst the candidate loci identified are of interest, and worth confirming, it is questionable how much additional insight will be gained

[28] Also there are a lot less ethical issues to be overcome, people generally don't mind if you stress or kill a plant.

into either, the evolutionary forces which have sculpted photosynthesis, or routes by which photosynthesis may be improved (Barrett and Hoekstra, 2011).

Having considered how we might describe the basis of genetic variation in particulate detail I am now going to change tack and consider if we really need to and whether it is worth the effort. Travisano and Shaw (2013) recently published a commentary that was critical of the community wide embrace of genomics as a means to explain the world around us stating that *"Enthusiasm for the burgeoning of molecular detail has inflated expectations of its explanatory power, yet the more complex is the phenomenon demanding explanation, the more molecular details obscure understanding."* This critical outlook is somewhat tempered, as the authors themselves note, by the considerable insights that have been gained from the genomics revolution. But the question, as I see it, is not so much whether it is useful *per se* but rather, how useful, and would the community be wiser to invest more in alternative approaches? In my thesis I have used the quantitative genetic approach as a forward genetic gene finding tool, and as a means of getting a handle on the genetic architecture of photosynthesis. As a gene finding technique much follow up research will be needed to fully confirm what was found. The insights into the genetic architecture of photosynthesis are, however, very useful when contemplating ways of improving photosynthesis (Hansen, 2006). Firstly, the narrow sense heritability (h^2) for photosynthesis was repeatedly estimated to be 0.4, and secondly, as a trait photosynthesis is highly polygenic (Chapter 5). These two insights alone are useful for informing further steps. The heritability estimate can be plugged into the breeders equation $R=h^2.S$, where R is the expected phenotypic response, and S is the selection intensity (Lynch and Walsh, 1998). Thus, dues to its medium to high heritability (Geber and Griffen, 2003), with sufficient selective pressure we can expect photosynthesis to change at an appreciable rate[29]. The polygenic genetic architecture is also important as if there were just a few variants of large effect explaining most of the variation, a selection experiment would likely fix all positive loci and thus run out of variation to select upon (Roff, 2003; Le Rouzic and Carlborg, 2008). Thus, based on its heritability and polygenic architecture, I would argue that a selection experiment, as opposed to the massive GWAS I proposed previously, is likely to give the best return on investment. I would suggest combining some of the strengths of genomics (high density marker data) and select for photosynthesis in Arabidopsis using genomic selection (GS) as opposed to traditional phenotype based selection (Meuwissen et al., 2001). The reason I suggest this is twofold. First

[29] One potential problem is that photosynthesis does have a theoretical thermodynamic limit, thus there is a ceiling beyond which it cannot be improved, still, as almost the entire photosynthetic community repeatedly argues, there is room for improvement.

GS is a relatively new technique with great potential and is ideally suited to highly polygenic traits. An experiment using GS in Arabidopsis would provide several 'spin-off' insights into both the technique itself and the nature of the adaptive process, for example, molecular characterization of the genetic response and speed and intensity of the phenotypic response associated with these genetic changes[30]. Since Arabidopsis has a small genome it will be easier to characterize the spectrum of these genetic changes than with crops where GS is currently being carried out (Zhao et al., 2012; Spindel et al., 2015). Second, GS is likely to generate results more rapidly than phenotypic selection alone, as even with the phenomic platform developed in this thesis phenotyping will remain a bottleneck. This bottleneck can be mitigated through the use of genomic selection algorithms to expedite the creation of optimal genotypes (Heffner et al., 2009; Desta and Ortiz, 2014). The phenomic platform can be used periodically to recalibrate the model and improve the final outcome. Finally, the phenotype data generated in this thesis, should be sufficient to start such an experiment (Spindel et al., 2015), in other words, we can begin right away.

At this stage you may be wondering what would be the point, what would a selection experiment for photosynthesis in Arabidopsis add that your GWAS study has not already achieved?[31] Having been so critical of this approach you may be forgiven for thinking I am contradicting myself. If as I argue, there is little point in characterizing the genetic basis of polygenic variation in particulate detail then why bother doing such an experiment in Arabidopsis, why not go straight to crops? Well we should do it in crops, but as previously mentioned, in Arabidopsis the experiment has, via the results of this thesis, already been initiated. One would also have the full suite of resources of the premiere plant model available, greatly expanding the range of possibilities. In Arabidopsis the goal would not be to look purely at the sequence changes for insights into how to improve photosynthesis, since as previously discussed they are likely to be individually of very small effect, but rather assess the metabolic and physiological shifts. I would argue that the physiological basis of a phenotypic response for a complex trait such as photosynthesis, is likely to provide insights which are easier to apply to non-model species and much easier to characterize than the individual causal loci. It has been shown that even a snapshot sample of just 130 metabolites had a similar predictive capacity for complex traits as 56K SNPs in maize (Riedelsheimer et al., 2012). The physiological changes

[30] The relationship between genetic change and phenotypic change is itself an important research topic which is key to understanding the connectivity between the genotype-phenotype map, for example increasing canalisation or robustness will uncouple one from the other.

[31] Obviously improving the photosynthesis of Arabidopsis itself is not really that useful, despite my overtures in the introduction of this thesis, it is not much of a food crop.

that occur in the photosynthetically 'elite' Arabidopsis plants can be better used as targets for modification in crops than the few candidate loci already identified[32]. The big challenge will be to do the phenotyping under 'field' conditions, as the response to selection under our growth chamber conditions is unlikely to be relevant for increasing photosynthesis in the field[33].

Selection experiments using natural genetic variation have resulted in very impressive phenotypic responses, maize lines selected for high oil and protein content show no signs of a reduced selection response, even after more than 100 generations of selection (Moose et al., 2004). Even when the selection is not directly imposed adaptation continues after 50,000 generations (Wiser et al., 2013), going so far as to result in changes to the elemental stoichiometry of the organisms (Turner et al., 2015). Impressively, populations subjected to long-term selection often show no loss of additive genetic variation for the traits selected, counterintuitively the heritability of the selected trait can even increase. The source of this sustained genetic variation in the face of selection has been attributed to epistasis (Carlborg et al., 2006), exposure of cryptic genetic[34] variation (Le Rouzic and Carlborg, 2008; Paaby and Rockman, 2014) and, many loci of small effect i.e. infinitesimals (Laurie et al., 2004). The literature is somewhat divided over the importance for epistasis in influencing the response to selection, with Crow (2010) stating it was effectively irrelevant, while Mackay (2014) considers it of such importance that it could compromise the performance of genomic prediction. Nevertheless, the potential for photosynthetic improvement via genomic selection is unexplored territory, based on the results of this thesis it should be fertile ground from which a significant intellectual and esculent harvest can be reaped.

Nurturing neglected genomes

During the screening of diverse germplasm for the GWA study just discussed, one particular genotype stood out, this genotype had a PSII efficiency several standard deviations lower than the population average. Upon further investigation it transpired that this was due to a chloroplast encoded mutation in the D1 protein conferring triazine resistance (El-Lithy et al., 2005) (Chapter 8). Due to its strong phenotypic deviation this genotype generated false

[32] Just using the best lines we already have is an option, but lines selected specifically for photosynthesis are likely to give a clearer and stronger signal as they can be directly compared to the population from which they originated.
[33] With some modifications, i.e. waterproofing and a PAM fluorimeter, our system should be field ready.
[34] Le Rouzic and Carlborg (2008) proposed the term genetic charging and genetic discharged to describe the accumulation and release of cryptic genetic variation. I think this is nice intuitive terminology, unfortunately it doesn't appear to have caught on.

associations when included in the GWA analysis (Chapter 5). I termed this phenomenon 'plasmotypic noise'. For this particular genotype the solution was simple, remove it from the analysis, but it raised a general problem, namely how much does variation in the plasmotype effect GWA mapping? This was worrying as the GWA approach focuses exclusively on the nucleotype, and including the plasmotype is not straightforward. Due to uniparental inheritance of the plasmotype it does not exhibit the linkage disequilibrium upon which GWA depends. To address this, I developed a set of plasmotype donor lines and used them to generate a full cybrid diallel. Apart from helping to solve the problem of plasmotypic noise, this cybrid diallel represents an important new resource for the study of nucleotype-plasmotype interactions. As described in Chapter 7 these interactions were difficult to disentangle from background effects using prior research methods, the reciprocal design and complete exchange of genetic material overcomes such confounding factors.

In recent years the role of organelle genomes in adaptation and organism performance has been gaining progressively more recognition (Budar and Roux, 2011; Nunnari and Suomalainen, 2012; Pesole et al., 2012; Ballard and Pichaud, 2014). Organelles are increasingly accepted as central hubs of signal and information processing (Pogson et al., 2008; Horan et al., 2013) and are now known to have effects on diverse phenotypes ranging from plant immunity (de Torres Zabala et al., 2015) to animal personality (Løvlie et al., 2014). The results presented in this thesis provide further support for the important role of these organelles in plant performance. In general, the largest effects were found with the triazine resistant Ely plasmotype. This plasmotype has a compromised PSII, which reduces the efficiency of the light reactions and photosynthesis as a whole. Based on this, most of the resulting phenotypes can be easily explained i.e. reduced growth, smaller seeds, paler green, and a general reduction in metabolite abundance. One unexpected phenotype was an increase in the rate of germination. It remains to be seen, however, whether or not the compromised PSII efficiency is causally related to this germination phenotype. A rather fortuitous discovery at the end of my PhD, which may help in addressing this question, was the identification of an additional triazine resistant genotype. This genotype was collected in Huntley railway station in Scotland and genotyped for the 39 nuclear SNPs[35], described in Chapter 8. Based on this, the Huntley genotype was identified as having a divergent nuclear haplotype relative to Ely. This nuclear divergence combined with the geographic distance (over 800 km by train) has led me to conclude that this is an independent mutational event, although sequencing of its chloroplast

[35] Single nucleotide polymorphisms, not, in this context, to be confused with the Scottish Nationalist Party.

genome will be required to confirm this. Creation of a plasmotype donor line with this Huntley plasmotype is already underway. Once this donor line is generated the seven nucleotypes in the cybrid panel can be introduced into this plasmotypic background and tested for rate of germination. If their rate of germination also increases in this plasmotypic background, then that would provide further evidence to suggest a role for the D1 protein in germination rate. How this could be mediated is a matter of speculation, the lack of photosynthesizing tissue during early germination would suggest that the role of D1, if any, is not in its usual capacity as primary electron donor of the linear electron transport chain. If any link is found it would enhance our understanding of germination and illustrate how powerful a technique plasmotype exchange can be.

The Ely plasmotype was not the only one to have significant effects, both the Shah and Ws-4 plasmotypes increased seed size, the C24 plasmotype lead to a general increase in primary metabolite abundance, and the Shah plasmotype caused a reduction in PSII efficiency. The analysis of this data is still underway and these results only represent the tip of the iceberg. The cybrid panel has also been, or is being, screened for; growth and photosynthesis in fluctuating light, pollen viability, respiration rates, non-photochemical quenching, stomatal conductance, and global gene expression. The initial results are very encouraging. The gene expression (RNA seq) data are proving to be especially fascinating, providing a treasure-trove of new insights into natural variation in plasmotype-nucleotype interactions. In particular, highlighting the broad impact of altered PSII efficiency, when the Ely plasmotype is introduced. There is both overlap and divergence in how different nucleotypes respond to this, essentially exposing cryptic genetic variation. These preliminary results are in agreement with those of Chapter 6 where the impact of this plasmotype of PSII protein phosphorylation was shown to be strong but also somewhat modulated by the nucleotype.

One key step, yet to be taken, is to assay the effect of foreign plasmotype on plant performance under field conditions. This will be the acid test as to whether or not nucleotypes perform better with their native plasmotype. It would also allow for the testing of the adaptive significance of the variation in plasmotype and to what degree the associated nucleotype influences this adaptive potential, i.e. what is the penetrance of the plasmotype and are some nucleotypes more easily perturbed than others? In other words, is there variation in the robustness of some nucleotypes to plasmotypic perturbance and vice versa? Research on this cybrid panel has only just begun, but the range of possible experiments is limited only by the imagination of the researcher, I consider one potential experiment later in this discussion.

On a related note, interest in plasmotype exchange has garnered a lot of attention recently with the passing of legislation in the UK allowing mitochondrial substitution in humans (Gemmell and Wolff, 2015; Morrow et al., 2015). Whilst plants are clearly very different from humans with regards to their plasmotypic genome evolution (Lynch et al., 2006), some principles are likely to apply across kingdoms. It is now of direct medical relevance that we gain greater understanding of how the nucleotype and plasmotype interact and how to identify optimal plasmotypic donors. Panels such as the one described here can inform medical research as they are more experimentally tractable due to reduced methodological or ethical constraints. Such cybrid diallels can provide insights into fundamental aspects of the nucleo-plasmotype relationship that cannot be elucidated when studying humans or other outbreeding species, where full reciprocal exchange is difficult, if not impossible, to achieve.

Hitching a ride with a chloroplast ticket

"It is somewhat amazing that political policies can affect evolution, but such has been the case with herbicide resistance." – Jonathan Gressel

Genetic hitchhiking is where loci, which themselves have no direct fitness advantage, rise to high frequency due to physical linkage to a locus which does confer an advantage. This phenomenon was first suggested by Maynard-Smith and Haigh (1974) as a resolution to a paradox, which had been pointed out by Lewontin (1974), namely, that contrary to the predictions of the neutral theory of evolution, genetic diversity does not increase linearly with population size. Gillespie (2000) coined the term genetic draft to describe the genome wide phenomenon of reduced genetic variation due to genetic hitchhiking or linked selection. For many years the importance of genetic draft remained largely a theoretical debate, however, just this year Corbett-Detig et al. (2015) have, through the analyses of whole genome data for 40 species, provided solid empirical evidence that selection is indeed more efficient in species with large populations. Thus, due to linkage between selected loci and neighbouring genomic regions, genetic draft can account for the reduced diversity in these populations, Lewontin's paradox appears to be resolved. In this thesis, I describe a previously unconsidered form of genetic draft, namely that mediated by selection on the plasmotype.

Chapter 8 describes the first empirical observation of organellar mediated genetic draft and concludes that it is likely to be quite common. The reason it had not been observed is probably because nobody has really considered it. The only reference in the literature that I

could find was in a review by Charlesworth and Wright (2001) and even here it was an indirect remark:

"In addition, natural selection causes two forms of genetic hitchhiking in populations or genome regions with low recombination: 'selective sweeps' and 'background selection'. Both these processes predict low genetic diversity in non-recombining regions of genomes, and also within inbreeding populations. In asexual or highly inbreeding populations, the entire genome is affected, including the organelle genomes."

This lack of theoretical treatment meant there was effectively no hypothesis to test. When I started studying triazine resistant populations of Arabidopsis I was not looking for organelle mediated genetic hitchhiking. This was something I stumbled upon purely out of curiosity. The serendipitous manner of this discovery also illustrates another reason why this form of linked selection was not observed before, namely that it is difficult to detect. In this case I was lucky, I knew the selective agent (triazine herbicides), the habitat (railway tracks and stations), and the genetic basis of resistance (mutated *psbA* gene). With this prior knowledge it was clear that the selective pressure was acting on the plastid. All lines which contained the selected plastid had an identical nuclear haplotype which indicated a single origin and complete effective linkage. Thus the nucleotype had hitched a ride on this plastid mutation, organelle mediated genetic hitchhiking in action. The fortuitous nature of this discovery does not mean that this is a rare phenomenon, as outlined in Chapter 8, the conditions necessary to give rise to such genetic hitchhiking are basically some level of inbreeding[36] and selection on an organelle, both of which are by no means rare (Igic and Busch, 2013; Bock et al., 2014; Dobler et al., 2014).

This organelle mediated selection is not the only insight that can be gained from this work. The more general implication is that with increased inbreeding, strong selection can increase the frequency of a particular genotype at such a rate that there is insufficient time for much outbreeding to occur. Thus, physically unlinked parts of the genome become statistically linked and can hitch a ride with the selected locus, the reach of genetic draft is increased. I call such phenomena trans-genome sweeps[37], meaning that they are not constrained by physical

[36] I doubt that inbreeding is completely necessary, if the selection is strong enough then the nuclear genome associated with the organelle mutation will be favoured due to the generally uniparental inheritance pattern of organelles, again hitchhiking even in the face of outcrossing.

[37] It is in a way similar to clonal interference and Hill-Robertson interference, however the focus is less on the problem of fixing multiple beneficial mutations and more on how reduced effective recombination can interact with selection to distort allele frequencies.

linkage and occur across genomes i.e. from organelle to nucleus. How common such trans-genome sweeps occur is an area, which merits further study. Even in humans, strong sweeps on the Y chromosome may facilitate autosomal hitchhiking. There is even one instance of a Y chromosome sweep in my native country the impact of this on autosomal diversity remains to be investigated (Moore et al., 2006).

Finally, to assess how common organelle mediated sweeps are, I would suggest two approaches. The first would be to develop theoretical models, which could then be run using computer simulations. Assuming the model is well designed it should be possible to predict the relative importance of inbreeding versus selection strength, and at what levels do these two variables interact to results in organellar draft. Other parameters such as gene flow via pollen and seed dispersal (Ellstrand, 2014) could also be important to include, as if selection is strong and gene flow low then chances are the nucleotype within which the organellar mutation arose will be much less likely to have to 'share' the free ride. Thus, the impact on nuclear diversity in the selected population could be much more significant compared to when there is a constant supply of new nucleotypes introgressing into the favorable plasmotypic background. The second approach to assess how common organellar draft is would be to go back into the field and look for populations where it is likely to be occurring. A good starting point would be to go back to populations of plants which have evolved resistance to organelle-targeted herbicides. Many such populations are documented and resistance occurs across a wide range of species with differing life histories and breeding systems (Heap, 2015). Additionally, some of these populations will still experience selection pressure and others will not, thus, as with the case of Chapter 8, natural experiments have been set up and all that needs doing is some data collection[38].

Photosynthate use efficiency, a new research opportunity.

This thesis combines photosynthetic physiology with quantitative genetics and in the process of doing so has generated several new research avenues, one which is of particular interest is the ability to study natural genetic variation in photosynthate use efficiency. Primary energy capture in photosynthesis is mediated by the light reactions, downstream from the light reactions the so-called dark reactions convert this captured energy into sugars, which form the basis of all biochemistry. These sugars and all subsequent products made from them are known as photosynthate, namely a product of photosynthesis (Amthor, 2010). As with all biological

[38] Sounds wonderful, but it is never this easy is it?

processes the efficiency at which the initial photosynthate is converted into biomass is likely to vary between individuals, this variation can provide valuable insights into both the mechanics of the process and suggest avenues by which it may be adjusted for agricultural needs. However, developing an assay, which can distinguish efficiency in the light reactions from efficiency in the use of their products is not easy. The phenotyping approaches currently in use cannot independently assess the efficiency of the two processes and must always assess them as combined phenomena. However, the development of plasmotype donor lines (Chapter 7) and the existence of an atrazine resistant plasmotype (Chapter 8) can be used experimentally to split these two components of photosynthesis, allowing for an independent assessment.

As shown in Chapters 3, 5, 7 and 8, Arabidopsis plants containing an atrazine resistant cytoplasm have a reduced photosystem II efficiency (Φ_{PSII}), see also (El-Lithy et al., 2005). The particular mutation responsible for this resistance, *psbA* Ser-264-Gly, has arisen in at least 73 species (Heap, 2015) and has been intensively studied in many of them (Warwick and Black, 1994; Darmency, 2013). The general consensus being that it leads to a reduction in fitness in wild populations and a reduction in yield in agricultural populations. In both cases the poor performance of these plants has been attributed to a reduction in photosynthetic capacity (Warwick, 1991). This reduction in photosynthesis is due to a diminished efficiency of the light reactions, as the transfer of electrons from the Q_B site of the D1 protein to plastoquinone occurs at a slower rate (Gronwald, 1994). Thus, the supply of electrons into the linear electron transport chain is restricted and the efficiency of the light reactions as a whole is compromised. As a result of this, the supply of NADPH and ATP to the dark reactions is reduced and the production of primary photosynthates is curtailed. The clear decrease in sugars and other primary metabolites in the presence of the Ely plasmotype is confirmation of this (Chapter 7). Thus the Ser-264-Gly mutation specifically affects the efficiency of the light reactions, which in turn restricts the quantities of photosynthate available for the plant to grow.

The magnitude of the cost of this reduction in the efficiency of the light reactions has been shown to interact with both light intensity and temperature, with lower light intensities and temperatures generally diminishing the cost of the mutation (Plowman et al., 1999; Szigeti and Lehoczki, 2003; Darmency, 2013). As a pilot, the Ely nucleotype, with and without the resistant plasmotype, was grown at two light intensities, 100 $\mu molm^{-2}s^{-1}$ and 200 $\mu molm^{-2}s^{-1}$. In contrast to the published results, the effect of the Ser-264-Gly mutation was much greater at the lower light intensity (Figure 9.1). The discrepancy between the literature and this pilot experiment is perhaps due to the fact the temperature was kept constant in our conditions

(20°C), whilst in the literature the temperature and light intensity were both reduced. The severely restricted growth rate at the lower light intensities is perhaps because the plants are approaching their light compensation point[39] and the further reduction in the efficiency of the light reactions has a comparatively larger effect. It has been shown that electron transfer from the Q_B protein is limiting under low light (Heber et al., 1988). Based on this, one would expect the triazine resistant protein to result in a stronger phenotype at lower irradiance, which is what we see. Thus, the reported reduction in the cost of the Ser-264-Gly mutation at low light and low temperature is likely to be due to reduced temperature as opposed to reduced light[40]. This increased penetrance of the Ser-264-Gly mutation on the growth of the plants at the lower light intensity provides a window of opportunity. Using the plasmotype exchange approach described in Chapter 7 it is possible to give many Arabidopsis genotypes the Ser-264-Gly mutation. If genotypes with and without this mutation are grown at both a low (<100 $\mu molm^{-2}s^{-1}$) and high (>200 $\mu molm^{-2}s^{-1}$) light intensity the relative reduction in growth rates could be compared. This reduction in growth can be attributed to the differences in the efficiency at which the different genotypes convert the limited pool of photosynthate into biomass, and as a consequence a measure of photosynthate use efficiency can be obtained[41].

Of course this approach is not without complications. One potential problem would be the effect of allelic variation in genes such as *ACD6,* which have a large impact on growth, in this case due to altered immune functioning ((Todesco et al., 2010) and Chapter 5). Plants containing an active *ACD6* allele grow slower and may show a relatively high photosynthate use efficiency due to their small size in control conditions. This is a common problem in drought experiments where smaller plants often show greater drought tolerance due to their size as opposed to an intrinsically higher water use efficiency (Skirycz et al., 2011). This potential bias can be overcome via careful selection of germplasm to avoid genotypes containing such known large effect loci. Additionally critical assessment of the data with explicit consideration of such potential biases should be undertaken.

[39] The point where the incoming light energy is such that the rate of photosynthesis matches the rate of respiration, thus there can be no net carbon fixation.

[40] As further evidence to support this the Ser-264-Gly mutation was shown to have a wild type PSII efficiency at low temperatures in El-Lithy et al. (2005)

[41] In addition to measuring projected leaf area, which is the primary measure of plant growth used in this thesis, dry weight as well as fresh weight should be obtained to assess whether plants which maintain a higher growth rate have managed to do so by increasing the relative contribution of water to their biomass or by actually increasing the amount of carbon gained. Experiments assessing root shoot allometry should also be undertaken to confirm whether or not the gain in shoot biomass comes at the expense of the roots.

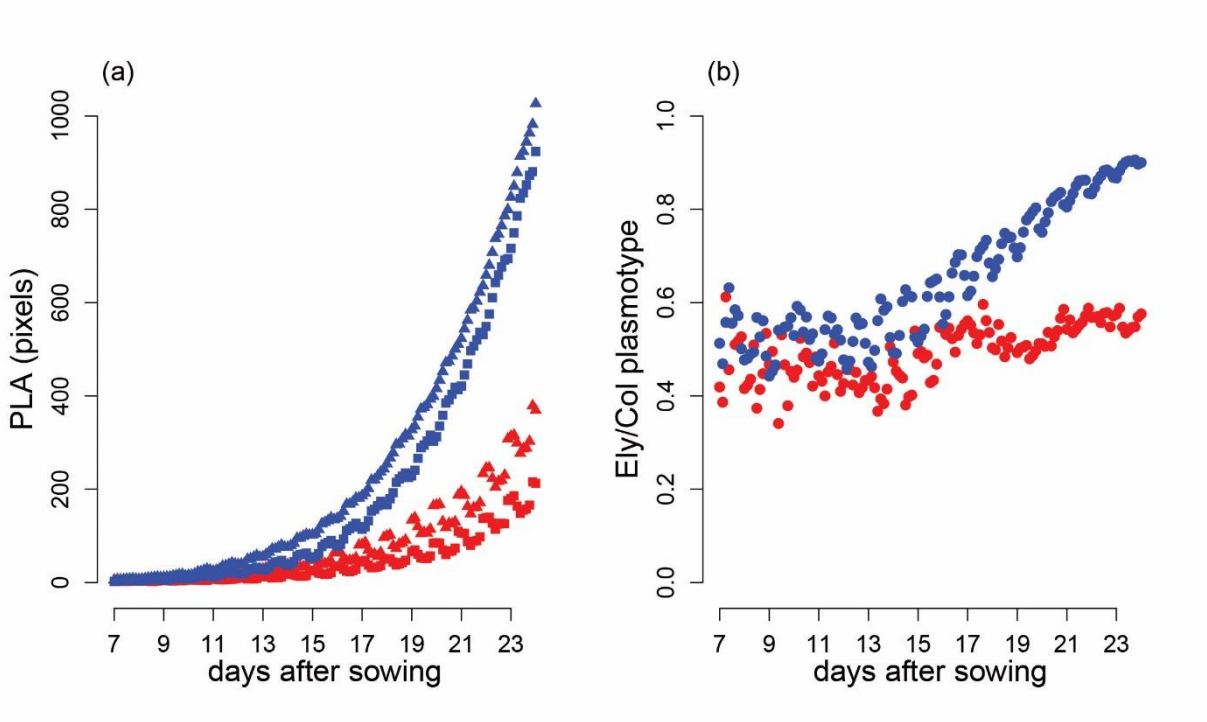


Figure 9.1. *Effect of reduced photosystem II efficiency on growth at two light levels. (a) Projected leaf area, triangles Ely^{Col}, squares Ely^{Ely}, blue 200 µmolm^{-2}s^{-1}, red 100 µmolm^{-2}s^{-1}. (b) Ratio of projected leaf area (PLA) of Col-0 (wild type) to Ely (triazine resistant) plasmotype, blue 200 µmolm^{-2}s^{-1}, red 100 µmolm^{-2}s^{-1}.*

Another complication is the potential for a large reduction in growth due to a cytoplasmic incompatibility between the Ely plasmotype and the introduced nucleotype being falsely attributed to the Ser-264-Gly mutation. The previously mentioned discovery of the Huntley genotype, a putatively independent Ser-264-Gly mutation will allow an additional control for any Ely specific nucleotype-plasmotype incompatibilities. It could be argued that any incompatibility will have a similar impact on growth at both light intensities, however, since the plasmotype is central to energy metabolism and homeostasis any incompatibility may be exacerbated due to the relative stress of such low light levels. In view of this, control plasmotypes without the Ser-264-Gly mutation should be used in all experiments. Although, as can be seen in Chapter 7, there appear to be high levels of nucleo-plasmotypic compatibility across wide genetic distances in Arabidopsis. This indicates a high level of conservation, which is perhaps due to the unique patterns of genome evolution which plant organelles exhibit (Lynch et al., 2006).

A final consideration is that, as a consequence of the reduction in the efficiency of the light reactions, the redox state of the chloroplast is likely to have been altered (Váradi et al., 2003). In Chapter 6, I showed that plants with the Ser-264-Gly mutation had reduced phosphorylation of PSII proteins, which may indicate an inability to fully reduce the

plastoquinone pool (Flood et al., 2014). The redox state of the plastoquinone pool is a regulatory hub for plant photosynthesis and alterations in its status are likely to have many pleiotropic effects (Foyer et al., 2012; Rochaix, 2013; Petrillo et al., 2014). Thus, in this case greater photosynthate use efficiency may be due to alternative redox sensitivities resulting in better maintenance of cellular homeostasis. The physiological basis of this will not necessarily result in improvements of photosynthate use efficiency under a normal redox status. Genetic variation in redox sensitivities is itself a fascinating research route, opening new possibilities to study genetic variation in cellular signaling. One simple approach could be to cross key photosynthetic or signaling mutants to the Col^{Ely} (Col-0 nucleotype with Ely plasmotype[42]) line as a means to study interaction between particular genes and both the redox status of the PQ pool as well as the impact of reduced PSII efficiency.

Having considered the potential shortcomings of this approach, it is fitting to consider its potential advantages. Recent research has shown that heterosis, one of the most important phenomena for agriculture, is partly due to improved resource use efficiency (Meyer et al., 2012). Additionally, work in the 1980s showed that that canola (*Brassica napus* L.) hybrids containing the Ser-264-Gly mutation still exhibited improved performance over their atrazine resistant mother (Grant and Beversdorf, 1985; Beversdorf et al., 1988). This improved growth and yield is likely to be due to improved photosynthate use efficiency as the Ser-264-Gly mutation precludes improvement of the light reactions. Triazine resistant cultivars of canola are still in use, with modern cultivars showing improved yield relative to initial introductions (Robertson et al., 2002; Darmency, 2013). Again, this improvement is likely to be due to improved resource use efficiency. The advantage of conducting this research in Arabidopsis is that any differences found can be characterized much more rapidly than in crops such as canola (an allopolyploid). In addition, the phenotyping platform described in Chapter 3 provides an excellent means of screening large numbers of plants for photosynthesis and growth at different light levels. The more tractable genetics of Arabidopsis combined with the resources developed in this thesis, phenomics, plasmotype exchange, and an additional Ser-264-Gly resistant plasmotype, allow a new and sophisticated approach to assess photosynthate use efficiency. One of the best characterized model systems for the study of heterosis is the Col-0 x C24 hybrid (Fujimoto et al., 2012). The effect of compromised PSII efficiency on heterosis can be very easily studied using the Ely plasmotype. Heterotic performance with and without a compromised PSII can be assessed relative to WT parents and parents harboring the resistant

[42] The choice of Col^{Ely} is simply because most gene knock-outs are available in the Col-0 nuclear background.

plasmotype. This will allow rapid and comprehensive assessment of whether the heterosis is mainly due to improved photosynthesis (Fujimoto et al., 2012) or improved resource use efficiency (Meyer et al., 2012). Since all the necessary lines for such an experiment are already created, this can be done very easily. A screen of a set of diverse genotypes for photosynthate use efficiency could be used to identify parents for the rapid creation of doubled haploid populations (Seymour et al., 2012). These doubled haploid populations can also be made to have multiple plasmotypes through the use of the plasmotype donor lines created in Chapter 7. Once such a doubled haploid population is made it could be screened for growth and photosynthesis at various light levels and QTLs responsible for differences in photosynthate use efficiency can be mapped[43]. Fine mapping and characterization of the causal loci will still be a significant undertaking but the potential insights into the regulation of cellular homeostasis and primary metabolism are huge.

Natural genetic variation in photosynthesis, fueling the future.

In the past six months alone three new reviews have been published outlining ways by which photosynthesis can be improved, without exception they have all disregarded natural genetic variation as a resource worth pursuing (Furbank et al., 2015; Long et al., 2015; Ort et al., 2015). One of these reviews was a perspective paper on improving photosynthesis and was authored by 25 prominent photosynthetic researchers (Ort et al., 2015). In this perspective they are disappointingly dismissive of natural genetic variation in photosynthesis, they state that: *"Fortunately, there are a wide range of potential avenues to improve photosynthetic efficiency along radically different paths than those dictated by evolutionary selection on extant and emergent natural variation."* And that *"Due to limited natural genetic variation in the enzymes and processes of plant photosynthesis, most of the limitations in photosynthetic efficiency will not likely be decisively overcome by conventional breeding approaches."* This exemplifies the fringe status natural genetic variation in photosynthesis appears to occupy, as outlined in my introduction and alluded to in the section on linking genotype with phenotype, this is likely due to the particular interest of the researchers and their particular rhetorical needs. In all cases the methods they argue for are focused on re-engineering and redesigning photosynthesis. Whilst these are valuable approaches they are perhaps somewhat premature considering our

[43] Naturally the genetic architecture of photosynthate use efficiency would also be a primary research goal, and as already argued, a selection experiment for photosynthesis would be very useful. Conducting such a selection experiment using genotypes with the Ser-264-Gly mutation would be a fascinating parallel experiment i.e. how would constraining PSII efficiency alter the evolutionary trajectory, what kind of compensatory mutations would evolve?

incomplete understanding of photosynthesis and they are likely to take quite some time before they yield edible results[44]. In the meantime we need a swift solution to the impending world food crisis, exploiting natural genetic variation using genomic selection is much more likely to yield such rapid results. One big advantage of using natural genetic variation and mining it via genomic selection is that the training population would be phenotyped under field conditions, thus the link from theoretical to practical is skipped[45]. The traits are optimized to perform under field conditions by the nature of how they were selected, while traits concocted in the mind of a researcher and built in the lab still need to be validated in the field. This combined with the high risk engineering approaches suggested further increases the time before such innovations can impact on crop yields. Contrary to their dismissal, I argue that the explosion of phenomic and genomic technologies, combined with the results presented in this thesis, mean that not only is natural genetic variation in photosynthesis an untapped resource, but that now is the moment to tap it! It is the proverbial low hanging fruit for improvement of crop primary production, we ignore it at our own peril.

[44] Not that we shouldn't try, I do think what they suggest are some awesome ideas and I would love to see it done.
[45] For most rapid returns this should be done directly in crop plants, although selection in Arabidopsis could be done in parallel.

References

Alcázar R, von Reth M, Bautor J, Chae E, Weigel D, Koornneef M, Parker JE (2014) Analysis of a Plant Complex Resistance Gene Locus Underlying Immune-Related Hybrid Incompatibility and Its Occurrence in Nature. PLoS Genet **10:** e1004848

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? Plant Cell **21:** 1877-1896

Amthor JS (2010) From sunlight to phytomass: On the potential efficiency of converting solar radiation to phyto-energy. New Phytologist **188:** 939-959

Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Felix M-A, Kruglyak L (2012) Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. Nat. Genet. **44:** 285-290

Anderson J, Chow W, Park Y-I (1995) The grand design of photosynthesis: Acclimation of the photosynthetic apparatus to environmental cues. Photosynth Res **46:** 129-139

Andrews J, Fryer M, Baker N (1995) Consequences of LHC II deficiency for photosynthetic regulation in chlorina mutants of barley. Photosynthesis Research **44:** 81-91

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

Arntz AM, DeLucia EH, Jordan N (2000) From Florescence to Fitness: Variation in photosynthetic rate affects fecundity and suvivorship. Ecology **81:** 2567-2576

Arntz M, Delph L (2001) Pattern and process: evidence for the evolution of photosynthetic traits in natural populations. Oecologia **127:** 455-467

Arntz MA, Vozar EM, Delph LF (2002) Serial Adjustments in Allocation to Reproduction: Effects of Photosynthetic Genotype. International Journal of Plant Sciences **163:** 591-597

Aro EM, Suorsa M, Rokka A, Allahverdiyeva Y, Paakkarinen V, Saleem A, Battchikova N, Rintamaki E (2005) Dynamics of photosystem II: a proteomic approach to thylakoid protein complexes. J Exp Bot **56:** 347-356

Arvidsson S, Pérez-Rodríguez P, Mueller-Roeber B (2011) A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. New Phytologist **191:** 895-907

Atwell S, Huang YS, Vilhjalmsson 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 JDG, 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. **465:** 627-631

Austin RB (1989) Genetic variation in photosynthesis. The Journal of Agricultural Science **112:** 287-294

Azhagiri AK, Maliga P (2007) Exceptional paternal inheritance of plastids in Arabidopsis suggests that low-frequency leakage of plastids via pollen may be universal in plants. The Plant Journal **52:** 817-823

Baker NR (2008) Chlorophyll fluorescence: A probe of photosynthesis in vivo. Annu. Rev. Plant Biol. **59:** 89-113

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

Ballard JWO, Pichaud N (2014) Mitochondrial DNA: more than an evolutionary bystander. Functional Ecology **28:** 218-231

Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD (2014) Thinking too positive? Revisiting current methods of population genetic selection inference. Trends in Genetics **30:** 540-546

Banks SC, Cary GJ, Smith AL, Davies ID, Driscoll DA, Gill AM, Lindenmayer DB, Peakall R (2013) How does ecological disturbance influence genetic diversity? Trends Ecol. Evol. **28:** 670-679

Bar-Even A, Noor E, Lewis NE, Milo R (2010) Design and analysis of synthetic carbon fixation pathways. Proceedings of the National Academy of Sciences **107:** 8889-8894

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet **12:** 767-780

Bashir T, Sailer C, Gerber F, Loganathan N, Bhoopalan H, Eichenberger C, Grossniklaus U, Baskar R (2014) Hybridization Alters Spontaneous Mutation Rates in a Parent-of-Origin-Dependent Fashion in Arabidopsis. Plant Physiology **165:** 424-437

Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, Vitek O, Salt DE (2010) A Coastal Cline in Sodium Accumulation in *Arabidopsis thaliana* Is Driven by Natural Variation of the Sodium Transporter AtHKT1;1. PLoS Genet **6:** e1001193

Bazin E, Glémin S, Galtier N (2006) Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals. Science **312:** 570-572

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. *In* Proceedings of the forty-ninth annual corn and sorghum industry research conference. Chicago, IL, pp 250-266

Bellafiore S, Barneche F, Peltier G, Rochaix JD (2005) State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. Nature **433:** 892-895

Bentsink L, Hanson J, Hanhart CJ, Blankestijn-de Vries H, Coltrane C, Keizer P, El-Lithy M, Alonso-Blanco C, de Andrés MT, Reymond M, van Eeuwijk F, Smeekens S, Koornneef M (2010) Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. Proceedings of the National Academy of Sciences **107:** 4264-4269

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. **11:** 867-879

Beversdorf WD, Hume DJ, Daonnelly-Vanderloo MJ (1988) Agronomic Performance of Trianzine-Resistant and Susceptible Reciprocal Spring Canola Hybrids. Crop Sci. **28:** 932-934

Black CC, Sung S-JS, Toderich K, Voronin PY (2010) Applying photosynthesis research to increase crop yields. Bull. Georgian Natl. Acad. Sci **4:** 101-110

Blanchet É, Penone C, Maurel N, Billot C, Rivallan R, Risterucci A-M, Maurice S, Justy F, Machon N, Noël F (2014) Multivariate analysis of polyploid data reveals the role of railways in the spread of the invasive South African Ragwort (*Senecio inaequidens*). Conserv. Genet.**:** 1-11

Boardman N (1977) Comparative photosynthesis of sun and shade plants. Annual review of plant physiology **28:** 355-377

Bock DG, Andrew RL, Rieseberg LH (2014) On the adaptive value of cytoplasmic genomes in plants. Mol. Ecol. **23:** 4899-4911

Bonardi V, Pesaresi P, Becker T, Schleiff E, Wagner R, Pfannschmidt T, Jahns P, Leister D (2005) Photosystem II core phosphorylation and photosynthetic acclimation require two different protein kinases. Nature **437:** 1179-1182

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and Association Mapping of *Arabidopsis thaliana* Flowering Time in Nature. PLoS Genet **6:** e1000940

Broman KW (2015) R/qtlcharts: Interactive Graphics for Quantitative Trait Locus Mapping. Genetics **199:** 359-361

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

Brougham RK (1960) The Relationship between the critical Leaf Area, Total Chlorophyll Content, and Maximum Growth-rate of some Pasture and Crop Planst. Annals of Botany **24:** 463-474

Brown TB, Cheng R, Sirault XRR, Rungrat T, Murray KD, Trtilek M, Furbank RT, Badger M, Pogson BJ, Borevitz JO (2014) TraitCapture: genomic and environment modelling of plant phenomic data. Current Opinion in Plant Biology **18:** 73-79

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Sofia da Silva H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The Genetic Architecture of Maize Flowering Time. Science **325:** 714-718

Budar F, Fujii S (2012) Cytonuclear Adaptation in Plants. Advances in Botanical Research **63:** 99-126

Budar F, Roux F (2011) The role of organelle genomes in plant adaptation. Plant Signal Behav **6:** 635-639

Busemeyer L, Ruckelshausen A, Moller K, Melchinger AE, Alheit KV, Maurer HP, Hahn V, Weissmann EA, Reif JC, Wurschum T (2013) Precision phenotyping of biomass accumulation in triticale reveals temporal genetic patterns of regulation. **3**

Butler D, Cullis BR, Gilmour A, Gogel B (2007) ASReml-R reference manual. Queensland Department of Primary Industries and Fisheries, Brisbane

Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Luis Araus J (2012) High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge. Journal of Integrative Plant Biology **54:** 312-320

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. **43:** 956-963

Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L (2006) Epistasis and the release of genetic variation during long-term selection. Nat Genet **38:** 418-420

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CWB, Mumm R, Bouwmeester H, Visser RGF, Keurentjes JJB (2012) Untargeted Metabolic Quantitative Trait Loci Analyses Reveal a Relationship between Primary Metabolism and Potato Tuber Quality. Plant Physiology **158:** 1306-1318

Caruso CM, Case AL, Bailey MF (2012) The evolutionary ecology of cytonuclear interactions in angiosperms. Trends in Plant Science **17:** 638-643

Chae E, Bomblies K, Kim S-T, Karelina D, Zaidem M, Ossowski S, Martín-Pizarro C, Laitinen Roosa AE, Rowan Beth A, Tenenboim H, Lechner S, Demar M, Habring-Müller A, Lanz C, Rätsch G, Weigel D (2014) Species-wide Genetic Incompatibility Analysis Identifies Immune Genes as Hot Spots of Deleterious Epistasis. Cell **159:** 1341-1351

Chapin Iii FS, Oechel WC (1983) Photosynthesis, respiration, and phosphate absorption by Carex aquatilis ecotypes along latitudinal and local environmental gradients. Ecology **64:** 743-751

Charlesworth D, Wright SI (2001) Breeding systems and genome evolution. Curr. Opin. Genet. Dev. **11:** 685-690

Chen M, Hofestädt R, Chen D, Altmann T, Klukas C (2014) Bridging Genomics and Phenomics. *In* Approaches in Integrative Bioinformatics. Springer Berlin Heidelberg, pp 299-333

Chen ZJ (2013) Genomic and epigenetic insights into the molecular bases of heterosis. Nat Rev Genet **14:** 471-482

Christley S, Lu Y, Li C, Xie X (2009) Human genomes as email attachments. Bioinformatics **25:** 274-275

Church GM, Gao Y, Kosuri S (2012) Next-Generation Digital Information Storage in DNA. Science **337:** 1628

Corbett-Detig RB, Hartl DL, Sackton TB (2015) Natural Selection Constrains Neutral Diversity across A Wide Range of Species. PLoS Biol **13:** e1002112

Coustham V, Vlad D, Deremetz A, Gy I, Cubillos FA, Kerdaffrec E, Loudet O, Bouché N (2014) SHOOT GROWTH1 Maintains Arabidopsis Epigenomes by Regulating *IBM1*. PLoS ONE **9:** e84687

Crosbie TM, Pearce RB (1982) Effects of Recurrent Phenotypic Selection for High and Low Photosynthesis on Agronomic Traits in Two Maize Populations1. Crop Sci. **22:** 809-813

Crosbie TM, Pearce RB, Mock JJ (1981) Recurrent Phenotypic Selection for High and Low Photosynthesis in Two Maize Populations1. Crop Sci. **21:** 736-740

Crow JF (2010) On epistasis: why it is unimportant in polygenic directional selection, Vol 365

Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. **14:** 262-274

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R (2005) Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. Plant Physiology **139:** 5-17

DalCorso G, Pesaresi P, Masiero S, Aseeva E, Schunemann D, Finazzi G, Joliot P, Barbato R, Leister D (2008) A complex containing PGRL1 and PGR5 is involved in the switch between linear and cyclic electron flow in Arabidopsis. Cell **132:** 273-285

Darmency H (2013) Pleiotropic effects of herbicide-resistance genes on crop yield: a review. Pest Management Science **69:** 897-904

Daudi A, Cheng Z, O'Brien JA, Mammarella N, Khan S, Ausubel FM, Bolwell GP (2012) The Apoplastic Oxidative Burst Peroxidase in Arabidopsis Is a Major Component of Pattern-Triggered Immunity. The Plant Cell **24:** 275-287

David Houle B, Christophe Pélabon, Günter P. Wagner, Thomas F. Hansen (2011) Measurement and Meaning in Biology. The Quarterly Review of Biology **86:** 3-34

de Torres Zabala M, Littlejohn G, Jayaraman S, Studholme D, Bailey T, Lawson T, Tillich M, Licht D, Bölter B, Delfino L, Truman W, Mansfield J, Smirnoff N, Grant M (2015) Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. Nature Plants **1**

Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B, Cooper LD, Courtot M, Csösz S, Cui H, Dahdul W, Das S, Dececchi TA, Dettai A, Diogo R, Druzinsky RE, Dumontier M, Franz NM, Friedrich F, Gkoutos GV, Haendel M, Harmon LJ, Hayamizu TF, He Y, Hines HM, Ibrahim N, Jackson LM, Jaiswal P, James-Zorn C, Köhler S, Lecointre G, Lapp H, Lawrence CJ, Le Novère N, Lundberg JG, Macklin J, Mast AR, Midford PE, Mikó I, Mungall CJ, Oellrich A, Osumi-Sutherland D, Parkinson H, Ramírez MJ, Richter S, Robinson PN, Ruttenberg A, Schulz KS, Segerdell E, Seltmann KC, Sharkey MJ, Smith AD, Smith B, Specht CD, Squires RB, Thacker RW, Thessen A, Fernandez-Triana J, Vihinen M, Vize PD, Vogt L, Wall CE, Walls RL, Westerfeld M, Wharton RA, Wirkner CS, Woolley JB, Yoder MJ, Zorn AM, Mabee P (2015) Finding Our Way through Phenotypes. PLoS Biol **13:** e1002033

Dekkers BJW, Willems L, Bassel GW, van Bolderen-Veldkamp RP, Ligterink W, Hilhorst HWM, Bentsink L (2012) Identification of Reference Genes for RT–qPCR Expression Analysis in Arabidopsis and Tomato Seeds. Plant and Cell Physiology **53:** 28-37

Demetriades-Shah TH, Fuchs M, Kanemasu ET, Flitcroft I (1992) A note of caution concerning the relationship between cumulated intercepted solar radiation and crop growth. Agricultural and Forest Meteorology **58:** 193-207

Denison RF (2012) Darwinian agriculture: How understanding evolution can improve agriculture. Princeton University Press

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends in Plant Science **19:** 592-601

Dhondt S, Gonzalez N, Blomme J, De Milde L, Van Daele T, Van Akoleyen D, Storme V, Coppens F, Beemster G, Inzé D (2014) High resolution time-resolved imaging of in vitro Arabidopsis rosette growth. The Plant Journal**:** n/a-n/a

Dhondt S, Wuyts N, Inzé D (2013) Cell to whole-plant phenotyping: the best is yet to come. Trends in Plant Science **18:** 428-439

Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010) Rare Variants Create Synthetic Genome-Wide Associations. PLoS Biol **8:** e1000294

Dijkstra P, Lambers H (1989) Analysis of specific leaf area and photosynthesis of two inbred lines of Plantago major differing in relative growth rate. New Phytologist **113:** 283-290

Dobler R, Rogell B, Budar F, Dowling DK (2014) A meta-analysis of the strength and nature of cytoplasmic genetic effects. J. Evolution Biol. **27:** 2021-2034

Dodd AN, Salathia N, Hall A, Kévei E, Tóth R, Nagy F, Hibberd JM, Millar AJ, Webb AAR (2005) Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. Science **309:** 630-633

Donovan LA, Maherali H, Caruso CM, Huber H, de Kroon H (2011) The evolution of the worldwide leaf economics spectrum. Trends in Ecology & Evolution **26:** 88-95

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

Eberhard S, Finazzi G, Wollman F-A (2008) The Dynamics of Photosynthesis. Annual Review of Genetics **42:** 463-515

Eilers PHC, Marx BD (1996) Flexible Smoothing with *B*-splines and Penalties. Statistical Science **11:** 89-102

El-Din El-Assal S, Alonso-Blanco C, Peeters AJM, Raz V, Koornneef M (2001) A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. Nat Genet **29:** 435-440

El-Lithy ME, Rodrigues GC, van Rensen JJS, Snel JFH, Dassen HJHA, Koornneef M, Jansen MAK, Aarts MGM, Vreugdenhil D (2005) Altered photosynthetic performance of a natural *Arabidopsis* accession is associated with atrazine resistance. J. Exp. Bot. **56:** 1625-1634

Ellstrand NC (2014) Is gene flow the most important evolutionary force in plants? American Journal of Botany **101:** 737-753

Elsayed A (2011) Analysis of paternal plastid inheritance in Arabidopsis thaliana and its impact on biosafety of gene-modified plants. lmu

Evans J (1989) Photosynthesis and nitrogen relationships in leaves of C3 plants. Oecologia **78:** 9-19

Evans LT (1997) Adapting and improving crops: The endless task. Philosophical Transactions of the Royal Society B: Biological Sciences **352:** 901-906

Falahati-Anbaran M, Lundemo S, Stenøien HK (2014) Seed dispersal in time can counteract the effect of gene flow between natural populations of *Arabidopsis thaliana*. New Phytol. **202:** 1043-1054

Falk A, Feys BJ, Frost LN, Jones JDG, Daniels MJ, Parker JE (1999) EDS1, an essential component of R gene-mediated disease resistance in Arabidopsis has homology to eukaryotic lipases. Proceedings of the National Academy of Sciences **96:** 3292-3297

Farquhar GD, von Caemmerer S, Berry JA (1980) A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta **149:** 78-90

Fatichi S, Leuzinger S, Körner C (2014) Moving beyond photosynthesis: from carbon source to sink-driven vegetation modeling. New Phytol. **201:** 1086-1095

Finkel E (2009) With 'Phenomics,' Plant Scientists Hope to Shift Breeding Into Overdrive. Science **325:** 380-381

Fisher RA (1918) The Correlation between Relatives on the Supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh **52:** 399-433

Flexas J, Ribas-CarbÓ M, Diaz-Espejo A, GalmÉS J, Medrano H (2008) Mesophyll conductance to CO2: current knowledge and future prospects. Plant, Cell & Environment **31:** 602-621

Flood PJ, Harbinson J, Aarts MGM (2011) Natural genetic variation in plant photosynthesis. Trends Plant Sci. **16:** 327-335

Flood PJ, Yin L, Herdean A, Harbinson J, Aarts MGM, Spetea C (2014) Natural variation in phosphorylation of photosystem II proteins in Arabidopsis thaliana: is it caused by genetic variation in the STN kinases? Philosophical Transactions of the Royal Society B: Biological Sciences **369**

Foyer CH, Neukermans J, Queval G, Noctor G, Harbinson J (2012) Photosynthetic control of electron transport and the regulation of gene expression. Journal of Experimental Botany **63:** 1637-1661

Fracheboud Y, Ribaut JM, Vargas M, Messmer R, Stamp P (2002) Identification of quantitative trait loci for cold-tolerance of photosynthesis in maize (Zea mays L.). J Exp Bot **53:** 1967-1977

Frenkel M, Bellafiore S, Rochaix J-D, Jansson S (2007) Hierarchy amongst photosynthetic acclimation responses for plant fitness. Physiologia Plantarum **129:** 455-459

Fristedt R, Carlberg I, Zygadlo A, Piippo M, Nurmi M, Aro EM, Scheller HV, Vener AV (2009) Intrinsically unstructured phosphoprotein TSP9 regulates light harvesting in Arabidopsis thaliana. Biochemistry **48:** 499-509

Fristedt R, Willig A, Granath P, Crevecoeur M, Rochaix JD, Vener AV (2009) Phosphorylation of photosystem II controls functional macroscopic folding of photosynthetic membranes in Arabidopsis. Plant Cell **21:** 3950-3964

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 of the National Academy of Sciences **109:** 7109-7114

Furbank RT, Quick WP, Sirault XRR (2015) Improving photosynthesis and yield potential in cereal crops by targeted genetic manipulation: Prospects, progress and challenges. Field Crops Research

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

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, Kahles A, Bohnert R, Jean G, Derwent P, Kersey P, Belfield EJ, Harberd NP, Kemen E, Toomajian C, Kover PX, Clark RM, Ratsch G, Mott R (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature **477:** 419-423

Gao H, Subramanian S, Couturier J, Naik SG, Kim S-K, Leustek T, Knaff DB, Wu H-C, Vignols F, Huynh BH (2013) Arabidopsis thaliana Nfu2 accommodates [2Fe-2S] or [4Fe-4S] clusters and is competent for in vitro maturation of chloroplast [2Fe-2S] and [4Fe-4S] cluster-containing proteins. Biochemistry **52:** 6633-6645

Gassmann A (2005) Resistance to herbicide and susceptibility to herbivores: environmental variation in the magnitude of an ecological trade-off. Oecologia **145:** 575-585

Gassmann AJ, Futuyma DJ (2005) Consequence of herbivory for the fitness cost of herbicide resistance: photosynthetic variation in the context of plant–herbivore interactions. J. Evolution Biol. **18:** 447-454

Geber Monica A, Griffen Lauren R (2003) Inheritance and Natural Selection on Functional Traits. International Journal of Plant Sciences **164:** S21-S42

Gemmell N, Wolff JN (2015) Mitochondrial replacement therapy: Cautiously replace the master manipulator. BioEssays **37:** 584-585

Genty B, Briantais J-M, Baker NR (1989) The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochimica et Biophysica Acta (BBA) - General Subjects **990:** 87-92

Genty B, Meyer S (1995) Quantitative Mapping of Leaf Photosynthesis Using Chlorophyll Fluorescence Imaging. Functional Plant Biology **22:** 277-284

Gibeaut DM, Hulett J, Cramer GR, Seemann JR (1997) Maximal Biomass of Arabidopsis thaliana Using a Simple, Low-Maintenance Hydroponic Method and Favorable Environmental Conditions. Plant Physiology **115:** 317-319

Gibson G (2012) Rare and common variants: twenty arguments. Nature Reviews Genetics **13:** 135-145

Gibson G, Dworkin I (2004) Uncovering cryptic genetic variation. Nature Reviews Genetics **5:** 681-690

Gillespie JH (2000) Genetic Drift in an Infinite Population: The Pseudohitchhiking Model. Genetics **155:** 909-919

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

Gobron N, Waszczak C, Simon M, Hiard S, Boivin S, Charif D, Ducamp A, Wenes E, Budar F (2013) A Cryptic Cytoplasmic Male Sterility Unveils a Possible Gynodioecious Past for *Arabidopsis thaliana*. PLoS ONE **8:** e62450

Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, Birney E (2013) Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature **494:** 77-80

Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, Thioux JJ, Rolland G, Bouchier-Combaud S, Lebaudy A, Muller B, Simonneau T, Tardieu F (2006) PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. New Phytologist **169:** 623-635

Grant I, Beversdorf WD (1985) Agronomic performance of triazine-resistant single-cross hybrid oilseed rape (Brassica napus L.). Canadian Journal of Plant Science **65:** 889-892

Greiner S, Bock R (2013) Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. BioEssays **35:** 354-365

Gronwald JW (1994) Resistance to photosystem II inhibiting herbicides. *In* SB Powles, JA Holtum, eds, Herbicide Resistance in Plants: Biology and Biochemistry. Lewis Publishers, pp 27-60

Grouneva I, Gollan P, Kangasjärvi S, Suorsa M, Tikkanen M, Aro E-M (2013) Phylogenetic viewpoints on regulation of light harvesting and electron transport in eukaryotic photosynthetic organisms. **237:** 399-412

Gu J, Yin X, Struik PC, Stomph TJ, 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:** 455-469

Guo P, Baum M, Varshney RK, Graner A, Grando S, Ceccarelli S (2008) QTLs for chlorophyll and chlorophyll fluorescence parameters in barley under post-flowering drought. Euphytica **163:** 203-214

Hagemann R (2004) The Sexual Inheritance of Plant Organelles. *In* H Daniell, C Chase, eds, Molecular Biology and Biotechnology of Plant Organelles. Springer Netherlands, pp 93-113

Hall D, Tegström C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. Briefings in Functional Genomics and Proteomics **9:** 157-165

Hall NM, Griffiths H, Corlett JA, Jones HG, Lynn J, King GJ (2005) Relationships between water-use traits and photosynthesis in Brassica oleracea resolved by quantitative genetic analysis. Plant Breeding **124:** 557-564

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

Hansen TF (2006) The Evolution of Genetic Architecture. Annual Review of Ecology, Evolution, and Systematics **37:** 123-157

Hansen TF (2013) Why epistasis is important for selection and adaptation. Evolution **67:** 3501-3511

Haraguchi A, Li B, Matsuki S, Nagata O, Suzuki JI, Hara T (2009) Variation and plasticity of photosynthesis and respiration in local populations of fat-hen Chenopodium album in northern Japan. Plant Species Biology **24:** 189-201

Heap I (2015) The International Survey of Herbicide Resistant Weeds. www.weedscience.com

Heber U, Neimanis S, Dietz KJ (1988) Fractional control of photosynthesis by the QB protein, the cytochrome f/b 6 complex and other components of the photosynthesic apparatus. Planta **173:** 267-274

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher. Crop Sci. **49:** 1-12

Hervé D, Fabre F, Flores Berrios E, Leroux N, Al Chaarani G, Planchon C, Sarrafi A, Gentzbittel L (2001) QTL analysis of photosynthesis and water status traits in sunflower (Helianthus annuus L.) under greenhouse conditions. Journal of Experimental Botany **52:** 1857-1864

Hibberd JM, Sheehy JE, Langdale JA (2008) Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. Current Opinion in Plant Biology **11:** 228-231

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25:** 1965-1978

Hikosaka K (2004) Interspecific difference in the photosynthesis–nitrogen relationship: patterns, physiological causes, and ecological importance. Journal of Plant Research **117:** 481-494

Hikosaka K (2010) Mechanisms underlying interspecific variation in photosynthetic capacity across wild plant species. Plant Biotechnology **27:** 223-229

Hikosaka K, Shigeno A (2009) The role of Rubisco and cell walls in the interspecific variation in photosynthetic capacity. Oecologia **160:** 443-451

Hill J (2012) Early Pioneers of Photosynthesis Research. *In* JJ Eaton-Rye, BC Tripathy, TD Sharkey, eds, Photosynthesis, Vol 34. Springer Netherlands, pp 771-800

Hill WG, Goddard ME, Visscher PM (2008) Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. PLoS Genet **4:** e1000008

Hohmann-Marriott MF, Blankenship RE (2011) Evolution of Photosynthesis. Annual Review of Plant Biology **62:** 515-548

Horan MP, Gemmell NJ, Wolff JN (2013) From evolutionary bystander to master manipulator: the emerging roles for the mitochondrial genome as a modulator of nuclear gene expression. European Journal of Human Genetics **21:** 1335-1337

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjálmsson BJ (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nature genetics **44:** 212-216

Horton P (2000) Prospects for crop improvement through the genetic manipulation of photosynthesis: Morphological and biochemical aspects of light capture. Journal of Experimental Botany **51:** 475-485

Houben A, Sanei M, Pickering R (2011) Barley doubled-haploid production by uniparental chromosome elimination. **104:** 321-327

Hough J, Williamson RJ, Wright SI (2013) Patterns of Selection in Plant Genomes. Annual Review of Ecology, Evolution, and Systematics **44:** 31-49

Houle D (2010) Numbering the hairs on our heads: The shared challenge and promise of phenomics. Proceedings of the National Academy of Sciences **107:** 1793-1799

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

Hsi-Yang Fritz M, Leinonen R, Cochrane G, Birney E (2011) Efficient storage of high throughput DNA sequencing data using reference-based compression. Genome Research **21:** 734-740

Hu SP, Zhou Y, Zhang L, Zhu XD, Li L, Luo LJ, Liu GL, Zhou QM (2009) Correlation and quantitative trait loci analyses of total chlorophyll content and photosynthetic rate of rice (Oryza sativa) under water stress and well-watered conditions. Journal of Integrative Plant Biology **51:** 879-888

Huang X, Paulo M-J, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proceedings of the National Academy of Sciences **108:** 4488-4493

Huard-Chauveau C, Perchepied L, Debieu M, Rivas S, Kroj T, Kars I, Bergelson J, Roux F, Roby D (2013) An Atypical Kinase under Balancing Selection Confers Broad-Spectrum Disease Resistance in Arabidopsis. PLoS Genet **9:** e1003766

Hubbart S, Peng S, Horton P, Chen Y, Murchie EH (2007) Trends in leaf photosynthesis in historical rice varieties developed in the Philippines since 1966. Journal of Experimental Botany **58:** 3429-3438

Igic B, Busch JW (2013) Is self-fertilization an evolutionary dead end? New Phytologist **198:** 386-397

Iida S, Miyagi A, Aoki S, Ito M, Kadono Y, Kosuge K (2009) Molecular Adaptation of *rbcL* in the Heterophyllous Aquatic Plant *Potamogeton*. PLoS ONE **4:** e4633

Iida S, Yamada A, Amano M, Ishii J, Kadono Y, Kosuge K (2007) Inherited maternal effects on the drought tolerance of a natural hybrid aquatic plant, Potamogeton anguillanus. **120:** 473-481

Jali SS, Rosloski SM, Janakirama P, Steffen JG, Zhurov V, Berleth T, Clark RM, Grbic V (2014) A plant-specific HUA2-LIKE (HULK) gene family in Arabidopsis thaliana is essential for development. The Plant Journal **80:** 242-254

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:** 902-914

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics **135:** 205-211

Jansen RC, Tesson BM, Fu J, Yang Y, McIntyre LM (2009) Defining gene and QTL networks. Current Opinion in Plant Biology **12:** 241-246

Joosen RVL, Arends D, Willems LAJ, Ligterink W, Jansen RC, Hilhorst HWM (2012) Visualizing the Genetic Landscape of Arabidopsis Seed Performance. Plant Physiology **158:** 570-589

Joosen RVL, Kodde J, Willems LAJ, Ligterink W, van der Plas LHW, Hilhorst HWM (2010) germinator: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. The Plant Journal **62:** 148-159

Joseph B, Corwin JA, Kliebenstein DJ (2015) Genetic Variation in the Nuclear and Organellar Genomes Modulates Stochastic Variation in the Metabolome, Growth, and Defense. PLoS Genet **11:** e1004779

Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ (2013) Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome, Vol 2

Joseph B, Corwin JA, Züst T, Li B, Iravani M, Schaepman-Strub G, Turnbull LA, Kliebenstein DJ (2013) Hierarchical Nuclear and Cytoplasmic Genetic Architectures for Plant Growth and Defense within Arabidopsis. The Plant Cell Online **25:** 1929-1945

Joyce AR, Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. Nature Reviews Molecular Cell Biology **7:** 198-210

Jung H-S, Crisp PA, Estavillo GM, Cole B, Hong F, Mockler TC, Pogson BJ, Chory J (2013) Subset of heat-shock transcription factors required for the early response of Arabidopsis to excess light. Proceedings of the National Academy of Sciences **110:** 14474-14479

Jung HS, Niyogi KK (2009) Quantitative Genetic Analysis of Thermal Dissipation in Arabidopsis. Plant Physiology **150:** 977-986

Kahlau S, Bock R (2008) Plastid Transcriptomics and Translatomics of Tomato Fruit Development and Chloroplast-to-Chromoplast Differentiation: Chromoplast Gene Expression Largely Serves the Production of a Single Protein. The Plant Cell Online **20:** 856-874

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet **42:** 348-354

Kapralov M, Filatov D (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evolutionary Biology **7:** 73

Karasov TL, Kniskern JM, Gao L, DeYoung BJ, Ding J, Dubiella U, Lastra RO, Nallu S, Roux F, Innes RW, Barrett LG, Hudson RR, Bergelson J (2014) The long-term maintenance of a resistance polymorphism through diffuse interactions. Nature **512:** 436-440

Keurentjes JJ, Koornneef M, Vreugdenhil D (2008) Quantitative genetics in the age of omics. Current Opinion in Plant Biology **11:** 123-128

Keurentjes JJB, Angenent GC, Dicke M, Santos VAPMD, Molenaar J, van der Putten WH, de Ruiter PC, Struik PC, Thomma BPHJ (2011) Redefining plant systems biology: From cell to ecosystem. Trends in Plant Science **16:** 183-190

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:** 891-905

Keurentjes JJB, Fu J, De Vos CHR, Lommen A, Hall RD, Bino RJ, Van Der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M (2006) The genetics of plant metabolism. Nature Genetics **38:** 842-849

Kirschbaum MUF (2011) Does enhanced photosynthesis enhance growth? Lessons learned from CO2 enrichment studies. Plant Physiology **155:** 117-124

Kodama Y, Shumway M, Leinonen R (2012) The sequence read archive: explosive growth of sequencing data. Nucleic Acids Research **40:** D54-D56

Koornneef M, Alonso-Blanco C, Vreugdenhil D (2004) Naturally occurring genetic variation in Arabidopsis thaliana. Annu Rev Plant Biol **55:** 141-172

Koornneef M, Meinke D (2010) The development of Arabidopsis as a model plant. Plant Journal **61:** 909-921

Körner C (2013) Growth controls photosynthesis–mostly. Nova Acta Leopoldina **114:** 273-283

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

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009) A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in *Arabidopsis thaliana*. PLoS Genet **5:** e1000551

Kruger EL, Volin JC (2006) Reexamining the empirical relation between plant growth and leaf photosynthesis. Functional Plant Biology **33:** 421-429

Kruijer W, Boer M, Malosetti M, Flood PJ, Engel B, Kooke R, Keurentjes J, Eeuwijk Fv (2015) Marker-Based Estimation of Heritability in Immortal Populations. Genetics **199:** 1-20

Külheim C, Ågren J, Jansson S (2002) Rapid Regulation of Light Harvesting and Plant Fitness in the Field. Science **297:** 91-93

Kwak I-Y, Moore CR, Spalding EP, Broman KW (2014) A Simple Regression-Based Method to Map Quantitative Trait Loci Underlying Function-Valued Phenotypes. Genetics **197:** 1409-1416

Lambers H (1987) Does variation in photosynthetic rate explain variation in growth rate and yield? Netherlands Journal of Agricultural Science **35:** 505-519

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185-199

Lane N, Martin W (2010) The energetics of genome complexity. Nature **467:** 929-934

Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature **500:** 571-574

Langridge P, Fleury D (2011) Making the most of 'omics' for crop breeding. Trends in Biotechnology **29:** 33-40

Laurie CC, Chasalow SD, LeDeaux JR, McCarroll R, Bush D, Hauge B, Lai C, Clark D, Rocheford TR, Dudley JW (2004) The Genetic Architecture of Response to Long-Term Artificial Selection for Oil Concentration in the Maize Kernel. Genetics **168:** 2141-2155

Lawson T, Kramer DM, Raines CA (2012) Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. Current Opinion in Biotechnology **23:** 215-220

Le Rouzic A, Carlborg Ö (2008) Evolutionary potential of hidden genetic variation. Trends in Ecology & Evolution **23:** 33-37

Lee I, Bleecker A, Amasino R (1993) Analysis of naturally occurring late flowering in Arabidopsis thaliana. Molecular and General Genetics MGG **237:** 171-176

Leister D, Varotto C, Pesaresi P, Niwergall A, Salamini F (1999) Large-scale evaluation of plant growth in Arabidopsis thaliana by non-invasive image analysis. Plant Physiology and Biochemistry **37:** 671-678

Lemeille S, Turkina MV, Vener AV, Rochaix J-D (2010) Stt7-dependent Phosphorylation during State Transitions in the Green Alga Chlamydomonas reinhardtii. Molecular & Cellular Proteomics **9:** 1281-1295

Levi A, Ovnat L, Paterson AH, Saranga Y (2009) Photosynthesis of cotton near-isogenic lines introgressed with QTLs for productivity and drought related traits. Plant Science **177:** 88-96

Lewontin RC (1974) The genetic basis of evolutionary change, Vol 560. Columbia University Press New York

Li H, Bradbury P, Ersoz E, Buckler ES, Wang J (2011) Joint QTL Linkage Mapping for Multiple-Cross Mating Design Sharing One Common Parent. PLoS ONE **6:** e17573

Li H, Tong Y, Li B, Jing R, Lu C, Li Z (2010) Genetic analysis of tolerance to photo-oxidative stress induced by high light in winter wheat (Triticum aestivum L.). Journal of Genetics and Genomics **37:** 399-412

Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO (2010) Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. Proceedings of the National Academy of Sciences **107:** 21199-21204

Liang Y, Zhang K, Zhao L, Liu B, Meng Q, Tian J, Zhao S (2010) Identification of chromosome regions conferring dry matter accumulation and photosynthesis in wheat (Triticum aestivum L.). Euphytica **171:** 145-156

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat. Protocols **1:** 387-396

Liu J, Xu X, Deng X (2005) Intergeneric somatic hybridization and its application to crop genetic improvement. **82:** 19-44

Liu W, Gowda M, Reif J, Hahn V, Ruckelshausen A, Weissmann E, Maurer H, Wurschum T (2014) Genetic dynamics underlying phenotypic development of biomass yield in triticale. BMC Genomics **15:** 458

Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, Zhang Q, Vilhjalmsson BJ, Korte A, Nizhynska V, Voronin V, Korte P, Sedman L, Mandakova T, Lysak MA, Seren U, Hellmann I, Nordborg M (2013) Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat Genet **45:** 884-890

Long Stephen P, Marshall-Colon A, Zhu X-G (2015) Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential. Cell **161:** 56-66

Long SP, Zhu X-G, Naidu SL, Ort DR (2006) Can improvement in photosynthesis increase crop yields? Plant, Cell & Environment **29:** 315-330

Løvlie H, Immonen E, Gustavsson E, Kazancioğlu E, Arnqvist G (2014) The influence of mitonuclear genetic variation on personality in seed beetles. Proceedings of the Royal Society B: Biological Sciences **281**

Lu H, Rate DN, Song JT, Greenberg JT (2003) ACD6, a Novel Ankyrin Protein, Is a Regulator and an Effector of Salicylic Acid Signaling in the Arabidopsis Defense Response. The Plant Cell Online **15:** 2408-2420

Lundemo S, Falahati-Anbaran M, StenØIen HK (2009) Seed banks cause elevated generation times and effective population sizes of *Arabidopsis thaliana* in northern Europe. Mol. Ecol. **18:** 2798-2811

Lynch M, Koskella B, Schaack S (2006) Mutation Pressure and the Evolution of Organelle Genomic Architecture. Science **311:** 1727-1730

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA

Lysak M, Mandáková T (2013) Analysis of Plant Meiotic Chromosomes by Chromosome Painting. *In* WP Pawlowski, M Grelon, S Armstrong, eds, Plant Meiosis, Vol 990. Humana Press, pp 13-24

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

MacKay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: Challenges and prospects. Nature Reviews Genetics **10:** 565-577

Mapplebeck LR, Machado VS, Grodzinski B (1982) Seed germination and seedingling growth characteristics of atrazine- susceptible and resistant biotypes of *Brassica campestris*. Canadian Journal of Plant Science **62:** 733-739

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends in Genetics **24:** 133-141

Marjoram P, Zubair A, Nuzhdin SV (2014) Post-GWAS: where next? More samples, more SNPs or more biology? **112:** 79-88

Martin RE, Asner GP, Sack L (2007) Genetic variation in leaf pigment, optical and photosynthetic function among diverse phenotypes of Metrosideros polymorpha grown in a common garden. Oecologia **151:** 387-400

Masel J (2006) Cryptic Genetic Variation Is Enriched for Potential Adaptations. Genetics **172:** 1985-1991

Masel J, Trotter MV (2010) Robustness and Evolvability. Trends in Genetics **26:** 406-414

Massonnet C, Vile D, Fabre J, Hannah MA, Caldana C, Lisec J, Beemster GTS, Meyer RC, Messerli G, Gronlund JT, Perkovic J, Wigmore E, May S, Bevan MW, Meyer C, Rubio-Díaz S, Weigel D, Micol JL, Buchanan-Wollaston V, Fiorani F, Walsh S, Rinn B, Gruissem W, Hilson P, Hennig L, Willmitzer L, Granier C (2010) Probing the Reproducibility of Leaf Growth and Molecular Phenotypes: A Comparison of Three Arabidopsis Accessions Cultivated in Ten Laboratories. Plant Physiology **152:** 2142-2157

Maynard-Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23-35

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Rosas MO, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic Properties of the Maize Nested Association Mapping Population. Science **325:** 737-740

McNally KL, Bruskiewich R, Mackill D, Buell CR, Leach JE, Leung H (2006) Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. Plant Physiology **141:** 26-31

Melis A (2009) Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. Plant Science **177:** 272-280

Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet **11:** 31-46

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics **157:** 1819-1829

Meyer RC, Witucka-Wall H, Becher M, Blacha A, Boudichevskaia A, Dörmann P, Fiehn O, Friedel S, von Korff M, Lisec J, Melzer M, Repsilber D, Schmidt R, Scholz M, Selbig J, Willmitzer L, Altmann T (2012) Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. The Plant Journal **71:** 669-683

Monteith JL (1994) Validity of the correlation between intercepted radiation and biomass. Agricultural and Forest Meteorology **68:** 213-220

Monteith JL, Moss CJ (1977) Climate and the Efficiency of Crop Production in Britain [and Discussion], Vol 281

Moore CR, Johnson LS, Kwak I-Y, Livny M, Broman KW, Spalding EP (2013) High-Throughput Computer Vision Introduces the Time Axis to a Quantitative Trait Map of a Plant Growth Response. Genetics **195:** 1077-1086

Moore LT, McEvoy B, Cape E, Simms K, Bradley DG (2006) A Y-Chromosome Signature of Hegemony in Gaelic Ireland. The American Journal of Human Genetics **78:** 334-338

Moose SP, Dudley JW, Rocheford TR (2004) Maize selection passes the century mark: a unique resource for 21st century genomics. Trends in Plant Science **9:** 358-364

Morrow EH, Reinhardt K, Wolff JN, Dowling DK (2015) Risks inherent to mitochondrial replacement, Vol 16

Mousseau TA, Roff DA (1987) Natural selection and the heritability of fitness components. Heredity **59:** 181-197

Moyle LC, Muir CD (2010) Reciprocal insights into adaptation from agricultural and evolutionary studies in tomato. Evolutionary Applications **3:** 409-421

Mozdzer TJ, Zieman JC (2010) Ecophysiological differences between genetic lineages facilitate the invasion of non-native Phragmites australis in North American Atlantic coast wetlands. Journal of Ecology **98:** 451-458

Murchie EH, Horton P (1997) Acclimation of photosynthesis to irradiance and spectral quality in British plant species: chlorophyll content, photosynthetic capacity and habitat preference. Plant, Cell & Environment **20:** 438-448

Murchie EH, Niyogi KK (2011) Manipulation of Photoprotection to Improve Plant Photosynthesis. Plant Physiology **155:** 86-92

Murchie EH, Pinto M, Horton P (2009) Agriculture and the new challenges for photosynthesis research. New Phytologist **181:** 532-552

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler E (2009) Association mapping: Critical considerations shift from genotyping to experimental design. Plant Cell **21:** 2194-2202

Nakamura Y, Cochrane G, Karsch-Mizrachi I (2013) The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research **41:** D21-D24

Nath K, Poudyal RS, Eom J-S, Park YS, Zulfugarov IS, Mishra SR, Tovuu A, Ryoo N, Yoon H-S, Nam HG, An G, Jeon J-S, Lee C-H (2013) Loss-of-function of OsSTN8 suppresses the photosystem II core protein phosphorylation and interferes with the photosystem II repair mechanism in rice (Oryza sativa). The Plant Journal **76:** 675-686

Neher RA (2013) Genetic Draft, Selective Interference, and Population Genetics of Rapid Adaptation. Annu. Rev. Ecol. Evol. S. **44:** 195-215

Network Rail (2015) http://www.networkrail.co.uk.

Noren H, Svensson P, Andersson B (1999) Auxiliary photosynthetic functions of Arabidopsis thaliana--studies in vitro and in vivo. Biosci Rep **19:** 499-509

Nunes-Nesi A, Carrari F, Lytovchenko A, Smith AMO, 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:** 611-622

Nunnari J, Suomalainen A (2012) Mitochondria: In Sickness and in Health. Cell **148:** 1145-1159

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modeling of additive and non-additive genetic line effects in single field trials. Theoretical and Applied Genetics **113:** 809-819

Oellrich A, Walls R, Cannon E, Cannon S, Cooper L, Gardiner J, Gkoutos G, Harper L, He M, Hoehndorf R, Jaiswal P, Kalberer S, Lloyd J, Meinke D, Menda N, Moore

L, Nelson R, Pujar A, Lawrence C, Huala E (2015) An ontology approach to comparative phenomics in plants. Plant Methods **11:** 10

Ojima M (1974) Improvement of photosynthetic capacity in soybean variety. Jpn. Agric. Res. Q **8:** 6-12

Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution: evaluating empirical support for theoretical predictions. Nat Rev Genet **13:** 867-877

Onate-Sanchez L, Vicente-Carbajosa J (2008) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. BMC Research Notes **1:** 93

Onoda Y, Hikosaka K, Hirose T (2004) Allocation of nitrogen to cell walls decreases photosynthetic nitrogen-use efficiency. Functional Ecology **18:** 419-425

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 MAJ, Peralta-Yahya PP, Prince RC, Redding KE, Spalding MH, van Wijk KJ, Vermaas WFJ, von Caemmerer S, Weber APM, Yeates TO, Yuan JS, Zhu XG (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proceedings of the National Academy of Sciences

Ort DR, Zhu X, Melis A (2011) Optimizing Antenna Size to Maximize Photosynthetic Efficiency. Plant Physiology **155:** 79-85

Paaby AB, Rockman MV (2014) Cryptic genetic variation: evolution's hidden substrate. Nat Rev Genet **15:** 247-258

Palumbi SR (2001) Humans as the World's Greatest Evolutionary Force. Science **293:** 1786-1790

Park S-Y, Peterson FC, Mosquna A, Yao J, Volkman BF, Cutler SR (2015) Agrochemical control of plant water use using engineered abscisic acid receptors. Nature **advance online publication**

Parry MAJ, Madgwick PJ, Carvalho JFC, Andralojc PJ (2007) Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. Journal of Agricultural Science **145:** 31-43

Parry MAJ, Reynolds M, Salvucci ME, Raines C, Andralojc PJ, Zhu X-G, Price GD, Condon AG, Furbank RT (2010) Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. Journal of Experimental Botany

Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. PLoS Genet **2:** e190

Pelletier G, Primard C, Vedel F, Chetrit P, Remy R, Rousselle, Renard M (1983) Intergeneric cytoplasmic hybridization in cruciferae by protoplast fusion. **191:** 244-250

Peng S, Krieg D, Girma F (1991) Leaf photosynthetic rate is correlated with biomass and grain production in grain sorghum lines. **28:** 1-7

Pesaresi P, Pribil M, Wunder T, Leister D (2011) Dynamics of reversible protein phosphorylation in thylakoids of flowering plants: the roles of STN7, STN8 and TAP38. Biochim Biophys Acta **1807:** 887-896

Pesole G, Allen JF, Lane N, Martin W, Rand DM, Schatz G, Saccone C (2012) The neglected genome. EMBO reports **13:** 473-474

Peterhansel C, Niessen M, Kebeish RM (2008) Metabolic Engineering Towards the Enhancement of Photosynthesis. Photochemistry and Photobiology **84:** 1317-1323

Petrillo E, Godoy Herz MA, Fuchs A, Reifer D, Fuller J, Yanovsky MJ, Simpson C, Brown JWS, Barta A, Kalyna M, Kornblihtt AR (2014) A Chloroplast Retrograde Signal Regulates Nuclear Alternative Splicing. Science **344:** 427-430

Pettigrew WT (2004) Cotton genotypic variation in the photosynthetic response to irradiance. Photosynthetica **42:** 567-571

Pettigrew WT, Gerik TJ (2007) Cotton Leaf Photosynthesis and Carbon Metabolism. *In* Advances in Agronomy, Vol 94, pp 209-236

Pettigrew WT, Turley RB (1998) Variation in photosynthetic components among photosynthetically diverse cotton genotypes. Photosynthesis Research **56:** 15-25

Phillips RL (2010) Mobilizing Science to Break Yield Barriers. Crop Sci. **50:** S-99-S-108

Pieruschka R, Poorter H (2012) Phenotyping plants: genes, phenes and machines. Functional Plant Biology **39:** 813-820

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Ågren J, Bossdorf O, Byers D, Donohue K, Dunning M, Holub EB, Hudson A, Le Corre V, Loudet O, Roux F, Warthmann N, Weigel D, Rivero L, Scholl R, Nordborg M, Bergelson J, Borevitz JO (2010) The Scale of Population Structure in *Arabidopsis thaliana*. PLoS Genet. **6:** e1000843

Platt A, Vilhjálmsson BJ, Nordborg M (2010) Conditions Under Which Genome-Wide Association Studies Will be Positively Misleading. Genetics **186:** 1045-1052

Plowman AB, Richards AJ, Tremayne MA (1999) Environmental effects on the fitness of triazine-resistant and triazine-susceptible *Brassica rapa* and *Chenopodium album* in the absence of herbicide. New Phytol. **141:** 471-485

Pogson BJ, Woo NS, Förster B, Small ID (2008) Plastid signalling to the nucleus and beyond. Trends in Plant Science **13:** 602-609

Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. The Plant Genome **5:** 92-102

Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet **47:** 702-709

Porra RJ, Thompson WA, Kriedemann PE (1989) Determination of Accurate Extinction Coefficients and Simultaneous-Equations for Assaying Chlorophyll-a and Chlorophyll-B Extracted with 4 Different Solvents - Verification of the Concentration of Chlorophyll Standards by Atomic-Absorption Spectroscopy. Biochimica Et Biophysica Acta **975:** 384-394

Powles SB, Yu Q (2010) Evolution in Action: Plants Resistant to Herbicides. Annual Review of Plant Biology **61:** 317-347

R Core Team (2014) R: A language and environment for statistical computing. *In*. R Foundation for Statistical Computing, Vienna, Austria

Raines CA (2011) Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: Current and future strategies. Plant Physiology **155:** 36-42

Ravi M, Chan SWL (2010) Haploid plants produced by centromere-mediated genome elimination. **464:** 615-618

Ravi M, Shibata F, Ramahi JS, Nagaki K, Chen C, Murata M, Chan SWL (2011) Meiosis-Specific Loading of the Centromere-Specific Histone CENH3 in *Arabidopsis thaliana*. PLoS Genet **7:** e1002121

Reich PB, Walters MB, Ellsworth DS (1991) Leaf age and season influence the relationships between leaf nitrogen, leaf mass per area and photosynthesis in maple and oak trees. Plant, Cell & Environment **14:** 251-259

Reiland S, Finazzi G, Endler A, Willig A, Baerenfaller K, Grossmann J, Gerrits B, Rutishauser D, Gruissem W, Rochaix JD, Baginsky S (2011) Comparative phosphoproteome profiling reveals a function of the STN8 kinase in fine-tuning of cyclic electron flow (CEF). Proc Natl Acad Sci U S A **108:** 12955-12960

Richards RA (2000) Selectable traits to increase crop photosynthesis and yield of grain crops. Journal of Experimental Botany **51:** 447-458

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet **44:** 217-220

Rigola D, Fiers M, Vurro E, Aarts MGM (2006) The heavy metal hyperaccumulator Thlaspi caerulescens expresses many species-specific genes, as identified by comparative expressed sequence tag analysis. New Phytologist **170:** 753-766

Rintamaki E, Martinsuo P, Pursiheimo S, Aro EM (2000) Cooperative regulation of light-harvesting complex II phosphorylation via the plastoquinol and ferredoxin-thioredoxin system in chloroplasts. Proc Natl Acad Sci U S A **97:** 11644-11649

Rintamaki E, Salonen M, Suoranta UM, Carlberg I, Andersson B, Aro EM (1997) Phosphorylation of light-harvesting complex II and photosystem II core proteins shows different irradiance-dependent regulation in vivo. Application of phosphothreonine antibodies to analysis of thylakoid phosphoproteins. J Biol Chem **272:** 30476-30482

Robertson MJ, Holland JF, Cawley S, Potter TD, Burton W, Walton GH, Thomas G (2002) Growth and yield differences between triazine-tolerant and non-triazine-tolerant cultivars of canola. Australian Journal of Agricultural Research **53:** 643-651

Robinson MR, Wray NR, Visscher PM (2014) Explaining additional genetic variation in complex traits. Trends in Genetics **30:** 124-132

Robson MJ (1973) The Growth and Development of Simulated Swards of Perennial Ryegrass: II. Carbon Assimilation and Respiration in a Seedling Sward. Annals of Botany **37:** 501-518

Rochaix JD (2007) Role of thylakoid protein kinases in photosynthetic acclimation. FEBS Lett **581:** 2768-2775

Rochaix JD (2013) Redox regulation of thylakoid protein kinases and photosynthetic gene expression. Antioxid Redox Signal **18:** 2184-2201

Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, Cain KE, Kokko H, Jennions MD, Kruuk LEB (2014) Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biol **12:** e1001779

Rockman MV (2012) The QTN program and the alleles that matter for evolution: All that's gold does not glitter. Evolution **66:** 1-17

Roff D (2003) Evolutionary quantitative genetics: Are we in danger of throwing out the baby with the bathwater? *In* Annales Zoologici Fennici, Vol 40. Helsinki: Suomen Biologian Seura Vanamo, 1964-, pp 315-320

Sage RF (2004) The evolution of C4 photosynthesis. New Phytologist **161:** 341-370

Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the Evolution of C4 Photosynthesis. Annual Review of Plant Biology **63:** 19-47

Salas Fernandez M, Strand K, Hamblin M, Westgate M, Heaton E, Kresovich S (2014) Genetic analysis and phenotypic characterization of leaf photosynthetic capacity in a sorghum (Sorghum spp.) diversity panel. Genetic Resources and Crop Evolution**:** 1-12

Salzmann D, Handley RJ, Müller-Schärer H (2008) Functional significance of triazine-herbicide resistance in defence of *Senecio vulgaris* against a rust fungus. Basic Appl Ecol **9:** 577-587

Sanei M, Pickering R, Kumke K, Nasuda S, Houben A (2011) Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proceedings of the National Academy of Sciences **108:** E498–E505

Scholl RL, May ST, Ware DH (2000) Seed and Molecular Resources for Arabidopsis. Plant Physiology **124:** 1477-1480

Schöttler MA, Kirchhoff H, Weis E (2004) The Role of Plastocyanin in the Adjustment of the Photosynthetic Electron Transport to the Carbon Metabolism in Tobacco. Plant Physiology **136:** 4265-4274

Schwarte S, Tiedemann R (2011) A gene duplication/loss event in the ribulose-1,5-bisphosphate-carboxylase/ oxygenase (Rubisco) small subunit gene family among accessions of Arabidopsis thaliana. Molecular Biology and Evolution **28:** 1861-1876

Segura V, Vilhjalmsson BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet **44:** 825-830

Seren Ü, Vilhjálmsson BJ, Horton MW, Meng D, Forai P, Huang YS, Long Q, Segura V, Nordborg M (2012) GWAPP: A Web Application for Genome-Wide Association Mapping in Arabidopsis. The Plant Cell Online **24:** 4793-4805

Seymour DK, Filiault DL, Henry IM, Monson-Miller J, Ravi M, Pang A, Comai L, Chan SWL, Maloof JN (2012) Rapid creation of Arabidopsis doubled haploid lines for quantitative trait locus mapping. Proceedings of the National Academy of Sciences **109:** 4227-4232

Shapir N, Mongodin EF, Sadowsky MJ, Daugherty SC, Nelson KE, Wackett LP (2007) Evolution of Catabolic Pathways: Genomic Insights into Microbial s-Triazine Metabolism. J. Bacteriol. **189:** 674-682

Sharkey TD, Bernacchi CJ, Farquhar GD, Singsaas EL (2007) Fitting photosynthetic carbon dioxide response curves for C3 leaves. Plant, Cell and Environment **30:** 1035-1040

Sharkey TD, Laporte MM, Kruger EL (2000) Will increased photosynthetic efficiency lead to increased yield in rice?1. *In* PLM J.E. Sheehy, B Hardy, eds, Studies in Plant Science, Vol Volume 7. Elsevier, pp 73-86

Shearman VJ, Sylvester-Bradley R, Scott RK, Foulkes MJ (2005) Physiological processes associated with wheat yield progress in the UK. Crop Science **45:** 175-185

Sheehy J, Ferrer A, Mitchell P, Elmido-Mabilangan A, Pablico P, Dionora M (2007) How the rice crop works and why it needs a new engine. *In* Charting new pathways to C, Vol 4, pp 3-26

Shi T, Bibby TS, Jiang L, Irwin AJ, Falkowski PG (2005) Protein Interactions Limit the Rate of Evolution of Photosynthetic Genes in Cyanobacteria. Molecular Biology and Evolution **22:** 2179-2189

Simon M, Loudet O, Durand S, Bérard A, Brunel D, Sennesal F-X, Durand-Tardif M, Pelletier G, Camilleri C (2008) Quantitative Trait Loci Mapping in Five New Large Recombinant Inbred Line Populations of *Arabidopsis thaliana* Genotyped With Consensus Single-Nucleotide Polymorphism Markers. Genetics **178:** 2253-2264

Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, Pucci A, Gonzalez N, Hoeberichts F, Tognetti VB, Galbiati M, Tonelli C, Van Breusegem F, Vuylsteke M, Inze D (2011) Survival and growth of Arabidopsis plants given limited water are not equal. Nature Biotechnology **29:** 212-214

Slatkin M (2008) Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. Nat Rev Genet **9:** 477-485

Sloan DB (2014) Using plants to elucidate the mechanisms of cytonuclear co-evolution. New Phytologist**:** n/a-n/a

Slotte T (2014) The impact of linked selection on plant genomic variation. Briefings in Functional Genomics **13:** 268-275

Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink J-L, McCouch SR (2015) Genomic Selection and Association Mapping in Rice (<italic>Oryza sativa</italic>): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. PLoS Genet **11:** e1004982

Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE (2015) Big Data: Astronomical or Genomical? PLoS Biol **13:** e1002195

Stinchcombe JR, Kirkpatrick M (2012) Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. Trends in Ecology & Evolution **27:** 637-647

Stitt M (1991) Rising CO2 levels and their potential significance for carbon flow in photosynthetic cells. Plant, Cell & Environment **14:** 741-762

Stoll K, Jonietz C, Binder S (2015) In Arabidopsis thaliana two co-adapted cyto-nuclear systems correlate with distinct ccmC transcript sizes. The Plant Journal **81:** 247-257

Sulpice R, Trenkamp S, Steinfath M, Usadel B, Gibon Y, Witucka-Wall H, Pyl ET, Tschoep H, Steinhauser MC, Guenther M, Hoehne M, Rohwer JM, Altmann T, Fernie AR, Stitt M (2010) Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. Plant Cell **22:** 2872-2893

Sun J, Yang L, Wang Y, Ort DR (2009) FACE-ing the global change: Opportunities for improvement in photosynthetic radiation use efficiency and crop yield. Plant Science **177:** 511-522

Szigeti Z, Lehoczki E (2003) A review of physiological and biochemical aspects of resistance to atrazine and paraquat in Hungarian weeds. Pest Management Science **59:** 451-458

Takai T, Adachi S, Taguchi-Shiobara F, Sanoh-Arai Y, Iwasawa N, Yoshinaga S, Hirose S, Taniguchi Y, Yamanouchi U, Wu J, Matsumoto T, Sugimoto K, Kondo K, Ikka T, Ando T, Kono I, Ito S, Shomura A, Ookawa T, Hirasawa T, Yano M, Kondo M, Yamamoto T (2013) A natural variant of NAL1, selected in high-yield rice breeding programs, pleiotropically increases photosynthesis rate. Sci. Rep. **3**

Takai T, Kondo M, Yano M, Yamamoto T (2010) A quantitative trait locus for chlorophyll content and its association with leaf photosynthesis in rice. Rice **3:** 172-180

Tan EH, Henry IM, Ravi M, Bradnam KR, Mandakova T, Marimuthu MPA, Korf I, Lysak MA, Comai L, Chan SWL (2015) Catastrophic chromosomal restructuring during genome elimination in plants,

Tcherkez GGB, Farquhar GD, Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proceedings of the National Academy of Sciences **103:** 7246-7251

Teng S, Qian Q, Zeng D, Kunihiro Y, Fujimoto K, Huang D, Zhu L (2004) QTL analysis of leaf photosynthetic rate and related physiological traits in rice (Oryza sativa L.). Euphytica **135:** 1-7

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408:** 796-815

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. **43:** 159-162

Tikkanen M, Grieco M, Kangasjarvi S, Aro EM (2010) Thylakoid Protein Phosphorylation in Higher Plant Chloroplasts Optimizes Electron Transfer under Fluctuating Light. Plant Physiol **152:** 723-735

Tikkanen M, Nurmi M, Kangasjarvi S, Aro EM (2008) Core protein phosphorylation facilitates the repair of photodamaged photosystem II at high light. Biochim Biophys Acta **1777:** 1432-1437

Tikkanen M, Piippo M, Suorsa M, Sirpio S, Mulo P, Vainonen J, Vener AV, Allahverdiyeva Y, Aro EM (2006) State transitions revisited-a buffering system for dynamic low light acclimation of Arabidopsis. Plant Mol Biol **62:** 779-793

Tisné S, Serrand Y, Bach L, Gilbault E, Ben Ameur R, Balasse H, Voisin R, Bouchez D, Durand-Tardif M, Guerche P, Chareyron G, Da Rugna J, Camilleri C, Loudet O (2013) Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. The Plant Journal **74:** 534-544

Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, Laitinen RAE, Huang Y, Chory J, Lipka V, Borevitz JO, Dangl JL, Bergelson J, Nordborg M, Weigel D (2010) Natural allelic variation underlying a major fitness trade-off in Arabidopsis thaliana. Nature **465:** 632-636

Todesco M, Kim S-T, Chae E, Bomblies K, Zaidem M, Smith LM, Weigel D, Laitinen RAE (2014) Activation of the *Arabidopsis thaliana* Immune System by Combinations of Common *ACD6* Alleles. PLoS Genet **10:** e1004459

Travisano M, Shaw RG (2013) Lost in the map. Evolution **67:** 305-314

Turner CB, Wade BD, Meyer JR, Lenski RE (2015) Evolution of organismal stoichiometry in a 50,000-generation experiment with Escherichia coli. http://biorxiv.org/biorxiv/early/2015/06/29/021360.full.pdf

Vainonen JP, Hansson M, Vener AV (2005) STN8 protein kinase in Arabidopsis thaliana is specific in phosphorylation of photosystem II core proteins. J Biol Chem **280:** 33679-33686

Vainonen JP, Sakuragi Y, Stael S, Tikkanen M, Allahverdiyeva Y, Paakkarinen V, Aro E, Suorsa M, Scheller HV, Vener AV, Aro EM (2008) Light regulation of CaS, a novel phosphoprotein in the thylakoid membrane of Arabidopsis thaliana. FEBS J **275:** 1767-1777

van Kooten O, Snel JH (1990) The use of chlorophyll fluorescence nomenclature in plant stress physiology. Photosynthesis Research **25:** 147-150

Váradi G, Polyánka H, Darkó É, Lehoczki E (2003) Atrazine resistance entails a limited xanthophyll cycle activity, a lower PSII efficiency and an altered pattern of excess excitation dissipation. Physiologia Plantarum **118:** 47-56

Vener AV, van Kan PJ, Rich PR, Ohad I, Andersson B (1997) Plastoquinol at the quinol oxidation site of reduced cytochrome bf mediates signal transduction between light and protein phosphorylation: thylakoid protein kinase deactivation by a single-turnover flash. Proc Natl Acad Sci U S A **94:** 1585-1590

Verhoeven A, Osmolak A, Morales P, Crow J (2009) Seasonal changes in abundance and phosphorylation status of photosynthetic proteins in eastern white pine and balsam fir. Tree Physiology **29:** 361-374

Vila-Aiub MM, Neve P, Powles SB (2009) Fitness costs associated with evolved herbicide resistance alleles in plants. New Phytologist **184:** 751-767

Vink M, Zer H, Herrmann RG, Andersson B, Ohad I (2000) Regulation of Photosystem II core protein phosphorylation at the substrate level: Light induces exposure of the CP43 chlorophyll a protein complex to thylakoid protein kinase(s). Photosynth Res **64:** 209-219

Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era - concepts and misconceptions. Nature Reviews Genetics **9:** 255-266

Vlad D, Rappaport F, Simon M, Loudet O (2010) Gene Transposition Causing Natural Variation for Growth in *Arabidopsis thaliana*. PLoS Genet **6:** e1000945

Von Caemmerer S (2000) Biochemical models of leaf photosynthesis. Csiro publishing

von Caemmerer S, Evans JR (2010) Enhancing C3 Photosynthesis. Plant Physiology **154:** 589-592

Warwick SI (1991) Herbicide Resistance in Weedy Plants: Physiology and Population Biology. Annual Review of Ecology and Systematics **22:** 95-114

Warwick SI, Black LD (1994) Relative fitnes of herbicide-resistant and susceptible biotypes of weeds. Phytoprotection **75:** 37-49

Weaver SE, Thomas AG (1986) Germination Responses to Temperature of Atrazine-Resistant and -Susceptible Biotypes of Two Pigweed (Amaranthus) Species. Weed Science **34:** 865-870

Weigel D, Mott R (2009) The 1001 Genomes Project for Arabidopsis thaliana. Genome Biology **10:** 1-5

Weiss KM (1998) In search of human variation. Genome Research **8:** 691-697

Westoby M, Warton D, Reich PB (2000) The Time Value of Leaf Area. The American Naturalist **155:** 649-656

Whitney SM, Houtz RL, Alonso H (2011) Advancing Our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco. Plant Physiology **155:** 27-35

Wijnker E, Deurhof L, van de Belt J, de Snoo CB, Blankestijn 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 K, de Snoo CB, Lelivelt CL, Keurentjes JJ, Naharudin NS, Ravi M, Chan SW, de Jong H, Dirks R (2012) Reverse breeding in *Arabidopsis thaliana* generates homozygous parental lines from a heterozygous plant. Nat. Genet. **44:** 467-470

Willig A, Shapiguzov A, Goldschmidt-Clermont M, Rochaix JD (2011) The phosphorylation status of the chloroplast protein kinase STN7 of Arabidopsis affects its turnover. Plant Physiol **157:** 2102-2107

Wiser MJ, Ribeck N, Lenski RE (2013) Long-Term Dynamics of Adaptation in Asexual Populations. Science **342:** 1364-1367

Wittek F, Hoffmann T, Kanawati B, Bichlmeier M, Knappe C, Wenig M, Schmitt-Kopplin P, Parker JE, Schwab W, Vlot AC (2014) Arabidopsis ENHANCED DISEASE SUSCEPTIBILITY1 promotes systemic acquired resistance via azelaic acid and its precursor 9-oxo nonanoic acid. Journal of Experimental Botany **65:** 5919-5931

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Ja, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Leach IM, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlov J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Bluher M, Bolton JL, Bottcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dorr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Graszler J, Gronberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson A, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindstrom J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Muller G, Muller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nothen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S,

Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundstrom J, Swertz MA, Syvanen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin M-R, Jousilahti P, Jula AM, Kaprio J, Kastelein JJP, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimaki T, Lupoli S, Madden PAF, Mannisto S, Manunta P, Marette A, Matise TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PEH, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif J-C, Tonjes A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, The Electronic Medical R, Genomics C, The MC, The PC, The LifeLines Cohort S, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, Melbye M, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Perusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Volzke H, Walker M, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJF, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet **46:** 1173-1186

Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73:** 3-36

Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, Cavender-Bares J, Chapin T, Cornelissen JHC, Diemer M, Flexas J, Garnier E, Groom PK, Gulias J, Hikosaka K, Lamont BB, Lee T, Lee W, Lusk C, Midgley JJ, Navas M-L,

Niinemets U, Oleksyn J, Osada N, Poorter H, Poot P, Prior L, Pyankov VI, Roumet C, Thomas SC, Tjoelker MG, Veneklaas EJ, Villar R (2004) The worldwide leaf economics spectrum. Nature **428:** 821-827

Wright SI, Nano N, Foxe JP, Dar V-UN (2008) Effective population size and tests of neutrality at cytoplasmic genes in *Arabidopsis*. Genet. Res. **90:** 119-128

Wu R, Lin M (2006) Functional mapping - how to map and study the genetic architecture of dynamic complex traits. Nature Reviews Genetics **7:** 229-237

Wunder T, Liu Q, Aseeva E, Bonardi V, Leister D, Pribil M (2013) Control of STN7 transcript abundance and transient STN7 dimerisation are involved in the regulation of STN7 activity. Planta **237:** 541-558

Wurschum T, Liu W, Busemeyer L, Tucker M, Reif J, Weissmann E, Hahn V, Ruckelshausen A, Maurer H (2014) Mapping dynamic QTL for plant height in triticale. BMC Genetics **15:** 59

Xiong Y, DeFraia C, Williams D, Zhang X, Mou Z (2009) Deficiency in a cytosolic ribose-5-phosphate isomerase causes chloroplast dysfunction, late flowering and premature cell death in Arabidopsis. Physiologia Plantarum **137:** 249-263

Xu Y, Li H-N, Li G-J, Wang X, Cheng L-G, Zhang Y-M (2011) Mapping quantitative trait loci for seed size traits in soybean (Glycine max L. Merr.). Theoretical and Applied Genetics **122:** 581-594

Xue DW, Chen MC, Zhou MX, Chen S, Mao Y, Zhang GP (2008) QTL analysis of flag leaf in barley (Hordeum vulgare L.) for morphological traits and chlorophyll content. Journal of Zhejiang University: Science B **9:** 938-943

Yan J, Zhu J, He C, Benmoussa M, Wu P (1998) Molecular Dissection of Developmental Behavior of Plant Height in Rice (Oryza sativa L.). Genetics **150:** 1257-1265

Yang D-L, Jing R-L, Chang X-P, Li W (2007) Quantitative Trait Loci Mapping for Chlorophyll Fluorescence and Associated Traits in Wheat (Triticum aestivum). Journal of Integrative Plant Biology **49:** 646-654

Yin L, Fristedt R, Herdean A, Solymosi K, Bertrand M, Andersson MX, Mamedov F, Vener AV, Schoefs B, Spetea C (2012) Photosystem II Function and Dynamics in Three Widely Used *Arabidopsis thaliana* Accessions. PLoS ONE **7:** e46206

Yin X, Struik PC (2010) Modelling the crop: From system dynamics to systems biology. Journal of Experimental Botany **61:** 2171-2183

Yin Z, Meng F, Song H, He X, Xu X, Yu D (2010) Mapping quantitative trait loci associated with chlorophyll a fluorescence parameters in soybean (Glycine max (L.) Merr.). **231:** 875-885

Yin Z, Meng F, Song H, Wang X, Xu X, Yu D (2010) Expression quantitative trait loci analysis of two genes encoding Rubisco activase in soybean. Plant Physiology **152:** 1625-1637

Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart Jr CN (2008) Plant systems biology comes of age. Trends in Plant Science **13:** 165-171

Yue B, Xue WY, Luo LJ, Xing YZ (2006) QTL Analysis for Flag Leaf Characteristics and Their Relationships with Yield and Yield Traits in Rice. Acta Genetica Sinica **33:** 824-832

Zamir D (2013) Where Have All the Crop Phenotypes Gone? PLoS Biol **11:** e1001595

Zelitch I (1975) Improving the efficiency of photosynthesis. Science **188:** 626-633

Zelitch I (1982) The Close Relationship between Net Photosynthesis and Crop Yield. BioScience **32:** 796-802

Zer H, Vink M, Keren N, Dilly-Hartwig HG, Paulsen H, Herrmann RG, Andersson B, Ohad I (1999) Regulation of thylakoid protein phosphorylation at the substrate level: reversible light-induced conformational changes expose the phosphorylation site of the light-harvesting complex II. Proc Natl Acad Sci U S A **96:** 8277-8282

Zhang G-B, Yi H-Y, Gong J-M (2014) The Arabidopsis Ethylene/Jasmonic Acid-NRT Signaling Module Coordinates Nitrate Reallocation and the Trade-Off between Growth and Environmental Adaptation. The Plant Cell **26:** 3984-3998

Zhang J, Khan SA, Hasse C, Ruf S, Heckel DG, Bock R (2015) Full crop protection from an insect pest by expression of long double-stranded RNAs in plastids. Science **347:** 991-994

Zhang J, Marshall JD, Jaquish BC (1993) Genetic differentiation in carbon isotope discrimination and gas exchange in Pseudotsuga menziesii - A common-garden experiment. Oecologia **93:** 80-87

Zhang K, Fang Z, Liang Y, Tian J (2009) Genetic dissection of chlorophyll content at different growth stages in common wheat. Journal of Genetics **88:** 183-189

Zhang X, Hause RJ, Borevitz JO (2012) Natural Genetic Variation for Growth and Development Revealed by High-Throughput Phenotyping in Arabidopsis thaliana. G3: Genes|Genomes|Genetics **2:** 29-34

Zhao Y, Gowda M, Liu W, Würschum T, Maurer H, Longin F, Ranc N, Reif J (2012) Accuracy of genomic selection in European maize elite breeding populations. Theoretical and Applied Genetics **124:** 769-776

Zhu X-G, 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:** 513-526

Zhu X-G, Long SP, Ort DR (2010) Improving Photosynthetic Efficiency for Greater Yield. Annual Review of Plant Biology **61:** 235-261

Zou J, Rogers WE, Siemann E (2007) Differences in morphological and physiological traits between native and invasive populations of Sapium sebiferum. Functional Ecology **21:** 721-730

Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences **109:** 1193-1198

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting Signals, N-Terminal Modifications and Abundance of the Chloroplast Proteome. PLoS ONE **3:** e1994

Summary

Oxygenic photosynthesis is the gateway of the sun's energy into the biosphere, it is where light becomes life. Genetic variation is the fuel of evolution, without it natural selection is powerless and adaptation impossible. In this thesis I have set out to study a relatively unexplored field which sits at the intersection of these two topics, namely natural genetic variation in plant photosynthesis. To begin I reviewed the available literature (**Chapter 2**), from this it became clear that the main bottleneck restricting progress was the lack of high-throughput phenotyping platforms for photosynthesis. To address this an automated high-throughput chlorophyll fluorescence phenotyping system was developed, which could measure 1440 plants in less than an hour for Φ_{PSII}, a measure of photosynthetic efficiency (**Chapter 3**). Using this phenotyping platform I screened five populations of *Arabidopsis thaliana*. Three of these populations resulted from bi-parental crosses and segregated for only two genomes, using these I conducted family mapping (**Chapter 4**). The final two populations were composed of natural, field collected, accessions and were analysed using a genome wide association approach (**Chapter 5**). The family mapping approach had greater statistical power due to within population replication and the genome wide association approach had higher mapping resolution due to historical recombination. Both approaches were used to identify genomic regions (loci) which were responsible for some of the variation in photosynthesis observed. The number and average effect of these loci was used to infer the genetic architecture of photosynthesis as a highly complex polygenic trait for which there are many loci of very small effect. In addition to screening these large populations a smaller subset of 18 lines was assayed for natural variation in phosphorylation of photosystem II (PSII) proteins in response to changing light (**Chapter 6**). This exploratory study indicated that this process shows considerable variation and may be important for adaptation of the photosynthetic apparatus to photosynthetic extremes. The genetic mapping studies just described, focus exclusively on genetic variation in the nuclear genome, whilst this contains the majority of the plants genetic information there is also a store of genetic information in the chloroplast and mitochondria. These genetic repositories contain genes which are essential for photosynthesis and energy metabolism. Any variation in these genes could have a large impact on photosynthesis. To study natural variation in these genomes I developed a new population of reciprocal nuclear-organellar hybrids (cybrids) which could be used to study the effect of genetic variation in organelles whilst controlling for nuclear genetic variation (**Chapter 7**). Preliminary results indicate that this resource will be of great use in disentangling natural genetic variation in nucleo-organelle interactions. Finally I looked at one chloroplast encoded photosynthetic mutation in more detail (**Chapter 8**). This mutation had evolved in response to herbicide application and had spread along British railways. When

studying this population of resistant plants I found empirical evidence for organelle mediated nuclear genetic hitchhiking. This is a previously undescribed evolutionary phenomenon and is likely to be quite common. In conclusion there is an abundance of genetic variation in photosynthesis which can be used to improve the trait for agriculture and provide insights into novel evolutionary phenomena in the field.

Acknowledgments

Acknowledgements

I started this PhD six years ago yet it feels like only yesterday, a feeling which worries me as I expect time will continue to fly until I suddenly find myself on death's door, although the acknowledgements it not really the place for such melancholic reflections. On a more upbeat note the main reason the time has passed so fast is due to all of the wonderful people I have had the pleasure of getting to know during these six years.

To begin with my two supervisors Mark and Jeremy, you make a wonderful team. From a student's perspective you are often polar opposites, a contrast which greatly enhanced my learning experience as from the outset there were always two ways of seeing the same thing. Mark I learnt much from you, not only do you have a detailed understanding of molecular biology and genetics but you also have a pragmatic perspective on science as a social and political enterprise, a perspective which is most useful. I also enjoyed discussing many topics with you ranging from science policy to the best way to cook asparagus, you are a formidable debating opponent which I truly enjoyed. Jeremy your knowledge on pretty much every topic is amazing, not only can you remember the trivia and facts but you always impart your own nuanced perspective on the subject, be it photosynthetic physiology or the political situation of 1980's Belfast. I thoroughly enjoyed our many discussions, especially those conducted over a Weihenstephaner in De Zaaier. Mark and Jeremy thank you for your patience, I know I was a proper pain in the ass on many occasions, but in the end I think we had a lot of fun and managed to do some great science together.

Maarten as my promoter your involvement in my project was to check in once in a while and see how things were going. Your impact on my development as a scientist has been profound. I truly admire your lifelong enthusiasm for Arabidopsis and I enjoyed sharing in the fun of looking at the many new accessions that were growing in the greenhouse. You have also gave me much needed and appreciated help by employing me in my predoc-postdoc position, allowing me great freedom and the chance to finish this thesis, for that I will quite simply be forever in your debt. Also thanks for being patient with me, I know I tested the limits of what could be reasonably expected, and yet you always gave me space to do the many other things I wanted to do. Bas you are a wonderful Leerstoelhouder exhibiting true leadership by making everybody feel that they matter. You also really opened my eyes (or should I say mind) to how simple quantitative genetic measures such as heritability could impart deep evolutionary insights, for my intellectual development this has been revolutionary. I found a new enthusiasm for my research which up until then I viewed as a gene finding exercise. Working at the Labs of Genetics, HPC and Plant Physiology (my

unofficial third group) has been a true pleasure largely due to all of the people I could discuss science with, particularly the many P.I.'s Duur, Joost K, Arjan, Bart P, Fons, Klaas, Wim, Ernst, Leo M, Rob S, Uulke, Henk, Fred, Dick, Leonie, Ep, Eric S, Freek, Harro, Rolf, Hans, and Sijmen. Thank you for always being willing to chat and discuss, without your openness to casual questions over coffee, or making time when I knock on your door I doubt I would have gained much of the understanding of genetics, evolution, and plant physiology pre and post harvest that I now have. I would also like to thank my supervisors and mentors during my bachelor's in Dublin, Jenny, Carl, Bruce and Jim, you really fanned the spark that set me on this path in the first place, thank you.

Of course it is not all about P.I.'s much of what I learned came from the many PhD's, PostDoc's, technicians, Masters and Bachelor's students that I spent so much time with, thanks to (I hope I didn't forget somebody): Aarón, Aina, Alessandra, Alex, Alexandre, Ana Carolina, Ana Paula, Anderson, André, Andreas, Arjen vdP, Arman, Bart N, Bas D, Benyamin H, Bert, Bertha, Brian, Catarina, Cecília, Cezary, Charles, Cimille, Claudio, Corrie, Craig, Cris, Daniela, Deborah, Desalegn, Diana, Dimitrios, Dora, Edouard, Elias, Eric, Erik, Felipe, Florien, Frank, Gabriella, Gerrit, Giovanni, Gonda, Graham, Greice, Gulia, Hans D, Hedayat, Hetty, Jelle, Jenny, Jianhua, Jimeng, Jimmy T, Joana, Johanna, Johannes, Joke, Jonathan, Jordi, Jose, Joost vdH, Júlia, Julio, Karen, Kris, Laurens, Lea, Lemeng, Leo W, Maarten W, Manickam, Manuela, Marcin, Marcio, Marijke, Mark Z, Menno, Menno P, Merijn, Mina, Mohamed, Natalia C, Natalia M, Neli, Nelson, Niall, Nihal, Nikos, Nilma, Okello, Paola, Pavlos, Peter, Phoung, Pingping, Ralph, Ramon, Rashid, Renake, Rene K, Rik, Rini, Robert, Ronny, Ross, Roxanne, Sabina, Sam, Sasan, Setareh, Taede, Thierry, Thomas, Tina, Tobias, Valeria, Victoria, Vid, Wei, Wilco (for making Friday means as it sounds), Willem, Wouter, Wytske, Xianwen, Ya-Fen, Yanli, and Zeshan. What a list, and that is only the people in the plant science group! Thanks to all of you, you made my experience at the Radix so varied and rich, you created a wonderful environment within which I could grow both as a scientist and a person (and grow my moustache once in a while ☺).

During my PhD I also had the chance to supervise three students Guang, Daniel, and Vangelis. You were very different people and convinced me that there is no silver bullet when it comes to supervising and mentoring, it also made me realise how difficult it can be. I hope you learned as much from me as I did from you, thank you for giving me the chance to test myself. Also thanks to Jenny, Martine, Vid, Leonie, Cindy, Wilma, Juliana, Pierre, Julio and

all the other members of the EPS PhD council with whom we arranged many seminars, symposia and of course parties!

A few other Key Radix members made quite a difference, in Biometris Fred, Jos, Paul, Sabine and particularly Willem really helped with the analysis of the huge (for me at least) amounts of data I had to handle. Joost vdH you also provided invaluable aid in finding insights in the numbers, also your often sardonic emails always lightened my day. Henk J, Rob vdS and Jan you were all key members of my project without whom the phenotyper would never have been possible, thanks for all the help, particularly Rob for your constant presence whenever the camera decided it wanted a break and would not measure anymore. Also as part of the Photosynthesis and/or GWAS team thanks to Aina, Roxanne, Ana Carolina, Charles, Nihal, Mohamed, Nelson, Rik, Karen and Johanna, it was an adventure. Ross without you the lab would have been forever *terra incognita*, thanks so much for all your help, humour and great taste in beer! Also I had the fortune of receiving three hotel projects to work on the Cyto Swaps, thanks to Wilco and Leo for the seed phenotyping, Roland and Jeroen for the metabolomics assays and Edouard, Elio and Gabino for the RNA seq.

I also made many friends outside the Radix, in particular my time in Wageningen was made all the more fun and enriching by having the chance to live on a Dutch student corridor. Living with 16 people can only be described as a unique experience. Many people came and went but all left fond, funny memories. Thanks to Chris, Tjalle, Maxime, Laura, Laurina, Laure, Tessa, Fedor, Merel, Jasper, Jesse, Roel, Lika, Nino, Elke, Victor, Hannah, Max, Leonor, Maria, Mathias, Merel, Quentin, Willemijn, Rian, and Mengxuan. Particular thanks to Chris, for always being there to listen, and to Jesse, Jasper, Maxime and Roel, you lovable bunch of couch stealing, chicken killing, beer drinking ragamuffins! Outside of the corridor I also met and enjoyed time with many other fantastic people, thanks to Agatha, Aina (Poland), Alena, Andrea, Angela, Arie, Ben, Benjamin J, Bram, Carla (XuXu), Carlos, Cata, Christina, Christos, Diego, Dominika, Erika, Fabrizio, Gabriel, Gareth, Harry, Horacio, Iliana, Ilias, Imanol, Jared, Jimi, Jimmy M, Johan, Kaitlyn, Kune, Lara, Livia, Manos, Mathis, Mauricio, Milan, Peter, Rafael, Ragel, Remko, Roberta, Ruth, Sidnei, Stefania, Tom R, Valentina, Veronika, Victor L, Yasmina. Thanks also to Adriana and to Rasa, you didn't live in Wageningen but you are both very special people.

Out of so many wonderful people a few deserve a special mention.

Erik, honestly what do I say, you have been a debating partner with whom science and fun are one. You have also been my muse, for you possess not only prodigious knowledge but also the wisdom of one who has reflected deeply on both himself and the world around him, and best of all you are more than happy to share the insights you have gained. I thank you for listening and discussing everything and for agreeing to be my paranymph, I was happy I could return the honour. Frank my second paranymph you have always brightened my day, your candid and enthusiastic approach to life have always energised me. Your passion for good food and drink is contagious. We have had many fun adventures together from Berlin to frying mushrooms, to a messy Saint Paddy's day or two. One day we will go fishing!

Charles, if it was not for my misfortune of sitting opposite you I am sure I would have finished my PhD in three years! Joking aside you are a great friend, always willing to lend a helping hand and improving the mood in the office, even if it was by repeatedly unplugging my mouse. Julio I also reckon you are also implicit in me taking so long to finish, but not so much for distractions in the office but rather all of the bizarre adventures that happened once we stepped outside of the Radix. You have a great way with people and spread smiles wherever you go. Thank you for your energy and enthusiasm, we got in trouble but we always had fun. Felipe, the third member of the three Brazilian musketeers who so marked my time in Wageningen. With you I shared many discussions and a great appreciation literature in all its forms, our adventures while cycling to Poland are most memorable, as always 'freedom is just another word' Ilias, not quite a Brazilian Musketeer, perhaps a Greek D'Artagnan, you were the final ingredient that we needed to 'role like the gang we know we are' and 'live like there was no day before tomorrow'! I look forward to many many more adventures, you are a unique character, other descriptions are not suitable for a PhD thesis.

Thomas you are another who has left an indelible mark on my intellectual development, our discussions were wide ranging and I enjoyed your different perspective, even if we differ in some fundamental axioms, this foundational differences make the discussion all the more enlightening. Thanks for the many walks and diverse and challenging conversation. Marcin, where do I begin, I think in you I met someone who likes to stir shit even more than I do. Talking and debating with you is always refreshing, you never cease to challenge and are not afraid to provoke. The parties at the firehouse were some of the best I have ever experienced, and your 13min eggs were divine! Rafael, you are another with whom I enjoyed many discussion, you brought a calming flavour of wisdom which I very much

needed, along with Bart N, Ben, Renake, Erik, Julio, Paola, and many others who pointed out that sometimes I was perhaps going a little overboard in how strong I held my opinions. John K, you are also one who deserves special thanks, you have really made me think, and sing for that matter, spending time with you is a pleasure and I always feel I have learnt something new.

Speaking of music, my time in Wageningen was much enhanced by all the music and musicians around me, thanks to Alexandre, André, Ben, Bernardo, Cata, Damian, Ilias, Jimi, Johan, Jasper, Julio, Kris (it is about time for another street dance), Nikos, Ragel, and Remko (I have never heard a stool played so well) for the endless fun and for allowing me to sing once in a while, even though it may have hurt your ears.

Natalia to you I owe so much, you helped me develop as a person and were so unbelievably patient with me. Thank you so much for all the fun and laughter, there are so many memories to cherish. Kaitlyn you were a breath of fresh air, while I love to talk about science it was great to shift gears once in a while and discuss poetry and how the absence of words can bring more meaning than their addition. Thanks for the many late night discussions.

I have lived in Cologne for over a year now, and many of the people I have met since I have arrived really helped me get through the final stretch of the PhD. At the Max Planck I owe thanks to all the members of Maarten's group Agatha, Adriaan, Birgit, Corinna, Dorothee, Neils, Wilma and Jonas, with special thanks to Ralph and Jia Ding for putting up with me in the office. Also a very special thanks goes to Sigi and Marianne without whom I would never have managed to get all my plants phenotyped and genotyped. In the Max Planck another group were essential to maintaining my sanity and energy levels, the coffee drinkers club! Eva, Jonas (the E. Pearson fan), Susanne, Felix (the Bee man), Corinna, and Ben, thank you so much for putting up with all my crappy jokes and even laughing the odd time. Outside of the Max Planck I was very lucky to land on my feet with no real effort on my part. Pearl, you have helped so much, first you found me a place to live in Cologne and second you showed me how to paint, without your tips and guidance my thesis cover would have been pretty crap. Also thanks for all of the dinners and chats, they really help keep me sane. In the place you found for me to live I have been very lucky to have a great flatmate Mr. Thibault! Thibault you are endlessly entertaining, energetic and great fun to live with, even if I get woken up at 5 in the morning by your drunk friends, I enjoy living with you. And Janna, my almost flatmate, you always greet me with a smile and cook an awesome chicken soup which

has been very welcome on many an occasion that I returned home after yet another unproductive day fighting with myself and my thesis.

So finally I come to Ireland, Thanks to Allan, Stefan and the entire Stewart household, for hosting me in your house whenever I returned to Dublin. To all the UCD crew Fionn, Seán D, Hazel, Becky, Nathan, Kev, Florentina, and special thanks to James, John K, John L, Letizia, and Di-sein for traipsing all the way to Wageningen just to see me, and Rosie for the interview trip! Adam T, without you it is doubtful I would have gotten the PhD as you were the one who woke me up and race drove me to the airport when I almost overslept and missed my flight for the interview! Also thanks for being an awesome friend. Thanks also to my dear friends Dave, Sally, Tadgh, Sen N, Tommy, Shane, particularly thanks to Derek and Eoin for the endless fun we have had together.

To my sisters, Sarah and Áine, thank you for always looking out for me, reminding me of birthdays and anniversaries I am prone to forgetting, and always challenging me, discussing over the dinner table has had its ups and downs over the years but I have learnt much. Most of all, thanks for being there, I know I can always rely on you, and I hope I can return the favour. My entire extended family Aunts, Uncles, cousins and especially Grandparents, thanks for just being there, you mean more to me than you know. Special thanks have to go to my Godmother Ann for always spoiling me and to Grandad T for the many phone conversations, fry-ups, pancake eating competitions, and trips to the botanic gardens, you are an inspiration!

Renake how do I thank you? Usually words are my forte but some emotions do not exist in the spoken realm. You have balanced me and made me grow up, even though I fought back and was a right difficult pain in the arse, and for some unfathomable reason you still love me. I thank you from deep in my heart, you make me a better person. I look forward to our years ahead, particularly now that I am finally finished this PhD!

Mam and Dad, I owe you, quite literally, everything. All that I have achieved I have done so from the foundation of your love. You made me who I am, you instilled within me a mental and emotional fortitude which has got me through everything, to say ‘I thank you’ is an understatement extraordinaire. As a gesture of my love I dedicate this thesis to you both.

Acknowledgements

Wow I didn't think this would be so hard to write, I get the feeling that I just keep repeating myself, I guess I am running out of original and personal ways to compliment all of the wonderful people I have had the honour of spending time with. I guess this is perhaps the best complaint one could have, knowing too many wonderful people to be ever able to thank them all properly.

So finally I will leave you all with an Irish blessing

Seo sláinte an tséitéara, an ghadaí, an trodaí, agus an óltóra!

Má dhéanann tu séitéireacht, go ndéana tú séitéireacht ar an mbás,

Má ghoideann tú, go ngoide tú croí mná,

Má throideann tú, go dtroide tú i leith do dheartháir,

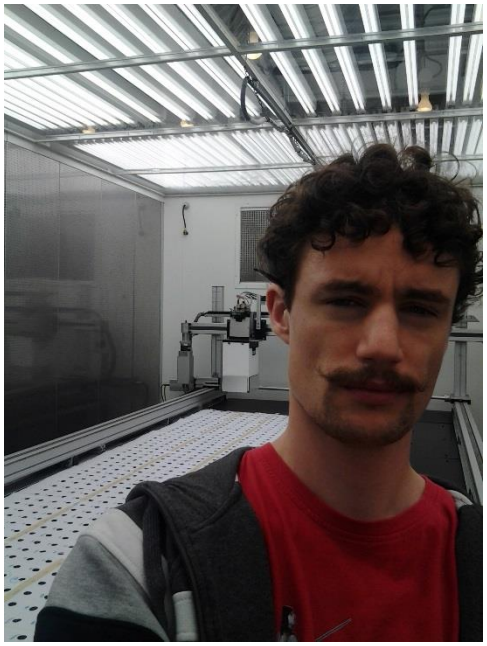
Agus má ólann tú, go n-óla tú liom féin.

About the Author

Biography

List of publications

Education statement



In situ with the phenotyper

Pádraic Joseph Flood was born in Holles Street Hospital, Dublin, Ireland, at 8:30am Monday the 5th of May 1986. Pádraic was a curious (perhaps better described as unusual) child who had a penchant for plants and insects. His father recounts stories of him navigating the neighbourhood using tree species as landmarks, and his mother recalls him disgusting her friends by showing them a selection of invertebrate fauna he found under rocks in their garden. His fascination with all things green was encouraged by his parents and his grandfather, who would save him from the trauma of shopping for clothes and bring him to the haven of the Dublin Botanic Gardens[46] where they would wonder at the cycads, the bamboo, and the leaves of the banana.

Pádraic graduated from Saint Benildus College, Dublin in 2004 and began a degree in general science in University College Dublin. In his third year he specialised in Botany and in the summer of 2007 received an undergraduate research experience and knowledge award (UREKA), which allowed him to conduct his first research project. The project was a collaboration between paleobotanist Jennifer McElwain and cell biologist Carl Ng. The aim was to study a broad evolutionary spectrum of vascular plants, ranging from lycopods to derived angiosperms in order to elucidate the possible role guard cell chloroplast had in stomatal functioning. The project was an utter failure as the techniques used for isolating the epidermis of the model plant *Commelina communis* could not be used on any of the other species. After ten weeks of trying everything from dental impression paste to chucking the plants in a blender, the final report was a list of unsuccessful techniques. Undaunted by this, Pádraic actually realised that he really quite liked research. For his Bachelor's thesis he once again teamed up with Jennifer and in a collaboration with ecophysiologist Bruce Osbourne set about studying the effects of low oxygen on the performance of a range of plant species. This project was marginally more successful but generated no conclusive results. Having graduated *cum laude* from his bachelors and winning the Botany Medal, yes this is a thing, Pádraic started a bursary in the Marine Institute and moved to the south west of Ireland to study the

[46] His love of plants was only matched by his hatred of clothes shopping, these two emotions acted synergistically to encourage his development as a botanist.

plants of the sea, phytoplankton. This was very different from his previous projects as it mainly involved monitoring species diversity and involved little in the way of hypothesis testing, nevertheless it was an interesting experience which left him with a lasting impression of the beauty of the microscopic and the importance of good science communication[47].

After finishing this bursary Pádraic got a job in an academic bookshop where he spent the next few months thinking about what topic to pursue for his PhD. Finally he settled on three topics, evolution, systematics and photosynthesis. Soon after he saw a PhD position advertised combining photosynthesis with evolution, he applied and the rest, as they say, is history[48]. In addition to conducting his doctoral research, Pádraic also took an interest in communicating the wonders of science to the public, this culminated in him winning the 2014 International FameLab competition.

[47] One seafood company failed to listen to the phytoplankton warning resulting in shellfish poisoning, the loss of millions of euros and ultimately bankruptcy.
[48] For an overview of the thesis work please consult the personal narrative section of the general introduction.

Flood PJ, Harbinson J, Aarts MGM (2011) Natural genetic variation in plant photosynthesis. Trends in Plant Science **16:** 327-335.

Flood PJ, Yin L, Herdean A, Harbinson J, Aarts MGM, Spetea C (2014) Natural variation in phosphorylation of photosystem II proteins in *Arabidopsis thaliana*: is it caused by genetic variation in the STN kinases? Philosophical Transactions of the Royal Society B: Biological Sciences **369**

Kruijer W, Boer M, Malosetti M, **Flood PJ**, Engel B, Kooke R, Keurentjes J, Eeuwijk Fv (2015) Marker-Based Estimation of Heritability in Immortal Populations. Genetics **199:** 1-20.

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to: **Pádraic Joseph Flood**
Date: **30 October 2015**
Group: **Laboratory of Genetics**
University: **Wageningen University & Research Centre**

1) Start-up phase

	date
► **First presentation of your project**	
Genetic analysis of plant plasticity of photosynthetic capacity and its tolerance to a fluctuating environment	Feb 15, 2010
► **Writing or rewriting a project proposal**	
► **Writing a review or book chapter**	
Natural Genetic Variation in Plant Photosynthesis, Trends in Plant Sciences, Vol 16, pp 327-335 (2011)	Feb 2011
► **MSc courses**	
Genetic Analysis, Tools and Concepts (GEN-30306)	Sep-Oct 2010
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*10.5 credits**

2) Scientific Exposure

	date
► **EPS PhD student days**	
EPS PhD student day 2011, Wageningen University	May 20, 2011
EPS PhD student day 2012, University of Amsterdam	Nov 30, 2012
EPS PhD student day 2013, Leiden University	Nov 29, 2013
► **EPS theme symposia**	
EPS Theme 4 symposium 'Genome Plasticity', Radboud University	Dec 11, 2009
EPS Theme 1 symposium 'Developmental Biology of Plants', Wageningen University	Jan 28, 2010
EPS Theme 3 symposium 'Metabolism and Adaptation', Leiden University	Feb 19, 2010
EPS Theme 4 symposium 'Genome Plasticity', Wageningen University	Dec 10, 2010
EPS Theme 3 symposium 'Metabolism and Adaptation', Wageningen University	Feb 10, 2011
EPS Theme 4 symposium 'Genome Plasticity', Wageningen University	Dec 09, 2011
EPS Theme 3 symposium 'Metabolism and Adaptation', Utrecht University	Apr 26, 2012
► **NWO Lunteren days and other National Platforms**	
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 19 & 20, 2010
TTI-GG Networking Event	Sep 22, 2010
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 04 & 05, 2011
TTI-GG Networking Event	Sep 21, 2011
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 02 & 03, 2012
5th National Ecogenomics Day	Jun 07, 2012
TTI-GG Networking Event	Sep 19, 2012
NERN Dutch National Ecology Meeting (NAEM)	Feb 05, 2013
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 22 & 23 2013
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 14 & 15 2014
► **Seminars (series), workshops and symposia**	

EPS Symposium Ecology and Experimental Plant Sciences 2, Wageningen	Sep 22, 2009
Biosolar Cells Workshop, Wageningen	Oct 18, 2010
Mini-symposim "Plant Breeding in the Genomics Era", Wageningen	Nov 25, 2011
Experimental Evolution Symposium	Mar 14, 2013
Invited seminar Otoline Leyser *'Auxin-Strigalactone cross-talk'*	Aug 28, 2009
Invited seminar Wim Soppe *'The molecular regulation of seed dormancy'*	Oct 20, 2009
Invited seminar Justin Borevitz *'Genetics of Adaptation: From model organisms to model ecosystems'*	Jan 12, 2010
Invited seminar Christiane Gebhardt *'The Molecular Basis of Quantitative Traits in Potato'*	Feb 05, 2010
Invited seminar Eric Schranz *'Genome archeology in the Brassicales'*	Mar 17, 2010
Invited seminar Bas Haring *'The value of Biodiversity'*	Sep 16, 2010
Invited seminar Jennifer A. Marshall Graves *'Weird animal genomes, sex and the future of men'*	Nov 16, 2010
Invited seminar Kirsten Bomblies *'Genetic incompatibility and the plant immune system'*	Nov 18, 2010
Invited seminar Jose Jimenz-Gomez *'Next generation quantitative genetics'*	Nov 29, 2010
Invited seminar Ales Pecinka *'Genome and epigenome stability under abiotic stress'*	Nov 29, 2010
Invited seminar Christian Hermans *'Molecular basis of plant nutrition: Insights into the responses to magnesium and nitrate availability'*	Dec 01, 2010
Invited seminar Eric Visser *'Exploring roots - selective root placement in nutrient-rich hotspots'*	Dec 08, 2010
Invited seminar William Ratcliff *'Experimental Evolution of Multicellularity'*	Aug 26, 2011
Invited seminar Robert Furbank *'Plant Phenomics, photosynthesis and the global food security challenge'*	Sep 02, 2011
Invited seminar Angus Buckling *'Bacteria-phage evolutionary ecology: lab, wild and applications'*	Oct 20, 2011
Invited seminar Neil Baker *'Evaluation of the role of the water-water cycle as a mechanism for protecting the photosynthetic apparatus from high light'*	Dec 06, 2011
Invited seminar Marc van Roosmalen *'The Brazilian Amazon: Hotspot of Biodiversity - New Species and Threats'*	Dec 07, 2011
Invited seminar Jennifer McElwain *'Impacts of global warming on plant biodiversity and ecosystem function: A 200 million year old case study form East Greenland'*	Jan 20, 2012
Invited seminar Graham Seymour *'Regulation of Ripening in Fleshy Fruits'*	Jan 24, 2012
Invited seminar Cornelia Spetea Wiklund *'Lessons from photosynthetic analysis in three widely used Arabidopsis ecotypes'*	Feb 21, 2012
Invited seminar José Crossa *'Genomic-enabled prediction in plant breeding when modeling G×E interaction'*	Jun 14, 2012
Invited seminar Jill M. Farrant *'Use of resurrection plants as models to understand how plants tolerate extreme water loss'*	Jun 26, 2012
Invited seminar Inez Hortenze Slamet-Loedin *'Genetic modification for iron biofortification and drought tolerance in rice'*	Jun 29, 2012

Invited seminar Patrick Forterre *'New concepts on the origin and nature of viruses: their major role in both ancient and recent biological evolution'*	Oct 18, 2012
Invited seminar Etienne Rampal *'A conceptual and statistical framework for adaptive radiations'*	Nov 05, 2012
Invited seminar Graham Farquhar	Mar 13, 2013
Invited seminar Marcus Koch *'Arabidopsis hybrid speciation - rare exception or simply overlooked?'*	Feb 14, 2013
Invited seminar Wolfram Mobius *'Range expansions of microrganisms in a structured environment'*	Oct 23, 2013
Invited seminar Dani Zamir *'Geno-Pheno in plant breeding'*	Feb 10, 2014

► **Seminar plus**

Justin Borevitz	Jan 12, 2010
Robert Furbank	Sep 02, 2011
Neil Baker	Dec 06, 2011
Jennifer McElwain	Jan 20, 2012
Cornelia Spetea Wiklund	Feb 21, 2012
Patrick Forterre	Oct 18, 2012
Richard Lenski	Mar 13, 2013
Graham Farquhar	Mar 14, 2013

► **International symposia and congresses**

Regulation of Plant Growth, Potsdam, Germany	Apr 12-14, 2010
International Congress on Photosynthesis Beijing, China	Aug 22-27, 2010
Gordon Conference; CO2 Assimilation in Plants: Genome to Biome, Geneva, Switzerland	May 29-Jun 03, 2011
3rd Joint retreat of PhD students in Plant Sciences	Jul 05-08, 2011
2nd International Plant Phenotyping Symposium, Jülich, Germany	Sep 05-07, 2011
Molecular Ecology	Feb 04-07, 2012
Molecular Mapping & Marker Assisted Selection, Vienna, Austria	Feb 08-11, 2012
4th European Plant Science Retreat for PhD students	Aug 15-17, 2012
PhenoDays 2012, Wageningen	Oct 10-12, 2012
International Congress on Photosynthesis, St Louis, USA	Aug 11-16, 2013
SEB Manchester 2014, UK	Jul 01-02, 2014

► **Presentations**

ALW meeting 'Experimental Plant Sciences', Lunteren (Poster)	Feb 19, 2010
International Congress on Photosynthesis Beijing (Poster)	Aug 23, 2010
TTI-GG Networking Event (Talk)	Sep 22, 2010
Gordon Conference; CO2 Assimilation in Plants: Genome to Biome (Poster)	May 29-Jun 03, 2011
Increasing photosynthesis in plants (Poster)	Aug 22, 2011
TTI-GG Networking Event (Talk)	Sep 21, 2011
Molecular Mapping & Marker Assisted Selection (Award for Best Talk)	Feb 08-11, 2012
ALW meeting 'Experimental Plant Sciences', Lunteren	Apr 05, 2012
4th European Plant Science Retreat for PhD students (Award for Best Talk)	Aug 16, 2012
Natural Variation of Plants (Talk)	Aug 22, 2012
TTI-GG Networking Event (Plenary Talk)	Sep 19, 2012
PhenoDays 2012 (Invited Speaker)	Oct 11, 2012
NERN Dutch National Ecology Meeting 2013 (Invited Speaker)	Feb 05, 2013
Wageningen University 95th Dies Natalis Symposium (Invited Speaker)	Mar 15, 2013
PABC Lecture NUIG Ireland (Invited Speaker)	Jun 13, 2013
Rijk Zwaan Seminar	Jul 02, 2013

International Congress on Photosynthesis St Louis (Poster)	Aug 13, 2013
Seminar IBMP Strasbourg (Invited Speaker)	Nov 15, 2013
SEB Manchester 2014 (Invited Speaker)	Jul 01, 2014
Measuring the Photosynthetic Phenome (Talk)	Jul 08, 2014
► **IAB interview**	
Meeting with a member of the International Advisory Board of EPS	Nov 14, 2012
► **Excursions**	
Monsanto site visit	Jan 27, 2011
Subtotal Scientific Exposure	*43.7 credits**

3) In-Depth Studies

	date
► **EPS courses or other PhD courses**	
ETNA School Plant Phenotyping, Jülich, Germany	Nov 01-10, 2009
VIBes in Bioscience symposium, Ghent, Belgium	Oct 13-15, 2010
Statistical analysis of ~omics data, Wageningen, NL	Dec 13-17, 2010
Increasing Photosynthesis in Plants, Wageningen, NL	Aug 21-26, 2011
Association mapping, Wageningen, NL	Feb 23, 2012
Natural Variation of Plants, Wageningen, NL	Aug 21-24, 2012
Introduction to R for statistical analysis, Wageningen, NL	Oct 22-23, 2012
► **Journal club**	
Literature discussion at Laboratory of Genetics	2009-2014
FLOP	2010-2013
Experimental Evolution Discussion Group	2011-2013
Evolution Lunch Meeting	2012-2014
► **Individual research training**	
Subtotal In-Depth Studies	*12.2 credits**

4) Personal development

	date
► **Skill training courses**	
PhD Competence Assessment	Jan 15, 2010
Basic Statistics	May 26-Jun 01 2010
ExPectationsS Career Day (EPS)	Nov 19, 2010
ExPexctationS Career Day (EPS)	Nov 18, 2011
ExPectationsS Career Day (EPS), Creativity in Science	Feb 01, 2013
How to write a world class paper	Oct 17, 2013
FameLab Master Class	Apr 24 & 25, 2014
► **Organisation of PhD students day, course or conference**	
ExPectationS Career Day Organiser	Nov 19, 2010
Photosynthesis Seminar Series, 22 seminars	Mar 01-Jul 19, 2011
EPS Flying Seminars, 6 Seminars	Sep 2011-June 2012
► **Membership of Board, Committee or PhD council**	
Member of EPS PhD Student Council	2010-2013
Subtotal Personal Development	*6.3 credits**

TOTAL NUMBER OF CREDIT POINTS* 72.7

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*

The research in this thesis was carried out at the Laboratories of Genetics and Horticulture & Product Physiology at Wagebningen University, Wageningen, The Netherlands, and was finacially supported by the Netherlands Scientific Organisation (NWO) and the Top Technological Insitute Green Genetics (TTI-GG).

Colophon

Cover by: Pádraic J Flood
Photosynthetic Variation (digital adaptation of water painting)

Thesis layout by: Pádraic J Flood

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress
Published by: Uitgeverij BOXPress, ‘s-Hertogenbosch