

## REVIEW PAPER

# Genetics as a key to improving crop photosynthesis

Tom P. J. M. Theeuwen<sup>1,\*</sup>, Louise L. Logie<sup>1</sup>, Jeremy Harbinson<sup>2</sup> and Mark G. M. Aarts<sup>1</sup>

<sup>1</sup> Laboratory of Genetics, Wageningen University & Research, Wageningen, The Netherlands

<sup>2</sup> Biophysics, Wageningen University & Research, Wageningen, The Netherlands

\* Correspondence: [tom.theeuwen@wur.nl](mailto:tom.theeuwen@wur.nl)

Received 19 November 2021; Editorial decision 22 February 2022; Accepted 23 February 2022

Editor: Johannes Kromdijk, University of Cambridge, UK

## Abstract

Since the basic biochemical mechanisms of photosynthesis are remarkably conserved among plant species, genetic modification approaches have so far been the main route to improve the photosynthetic performance of crops. Yet, phenotypic variation observed in wild species and between varieties of crop species implies there is standing natural genetic variation for photosynthesis, offering a largely unexplored resource to use for breeding crops with improved photosynthesis and higher yields. The reason this has not yet been explored is that the variation probably involves thousands of genes, each contributing only a little to photosynthesis, making them hard to identify without proper phenotyping and genetic tools. This is changing, though, and increasingly studies report on quantitative trait loci for photosynthetic phenotypes. So far, hardly any of these quantitative trait loci have been used in marker assisted breeding or genomic selection approaches to improve crop photosynthesis and yield, and hardly ever have the underlying causal genes been identified. We propose to take the genetics of photosynthesis to a higher level, and identify the genes and alleles nature has used for millions of years to tune photosynthesis to be in line with local environmental conditions. We will need to determine the physiological function of the genes and alleles, and design novel strategies to use this knowledge to improve crop photosynthesis through conventional plant breeding, based on readily available crop plant germplasm. In this work, we present and discuss the genetic methods needed to reveal natural genetic variation, and elaborate on how to apply this to improve crop photosynthesis.

**Keywords:** Bi- and multiparental populations, diversity panels, gene validation, genome wide association studies (GWAS), improving photosynthesis, natural genetic variation, quantitative trait locus (QTL) mapping.

## Introduction

Plants are largely sessile organisms that depend on their genetic composition to survive in a given environment. The environment will normally change between geographic locations and over time. When species evolve, selection pressures from these environments (re)shape their genomes by incorporating genetic variations that improve their fitness. Approximately 1.5 billion

years ago a cyanobacterium was engulfed by a phagotrophic eukaryote (Löffelhardt, 2014). This unique symbiotic incident eventually led to the transformation of the cyanobacterium into a chloroplast, and formed the basis for the evolution of plants. It brought together three genomes, the nuclear, the mitochondrial, and the plastidial genomes. Coevolution

between these shaped photosynthesis while plants spread to all corners of the globe. The evolution of, for example, carbon concentrating mechanisms to overcome high photorespiration rates in  $C_3$  photosynthesis, and the occurrence of chlorophyll *d* in some cyanobacteria to make better use of the light spectrum available are a testament to the adaptability of photosynthesis (Keeley and Niyogi, 2003; Gloag *et al.*, 2007). Nowadays, with so many plant species occupying a wide diversity of niches and dynamic environments, differences in photosynthetic performance due to natural genetic variation occur for a range of processes, such as metabolism, growth, and responsiveness to environmental cues (Björkman and Holmgren, 1963; Flood *et al.*, 2011; Yamori *et al.*, 2014; Soleh *et al.*, 2016, 2017; Arrivault *et al.*, 2019; Rungrat *et al.*, 2019; Faralli and Lawson, 2020; Acevedo-Siaca *et al.*, 2020).

Despite this variation in photosynthetic performance, the basic biochemical mechanisms of photosynthesis have remained remarkably conserved within  $C_3$ , the different forms of  $C_4$ , and CAM plants (Bungard *et al.*, 1999; Flood *et al.*, 2011). These basic biochemical mechanisms take place in the core photosynthetic machinery, and are encoded by around a hundred genes (Tyagi and Gaur, 2003; Berry *et al.*, 2013). Here we define the core photosynthetic machinery as the enzymes and the multi-molecular complexes required for the light and dark reactions of photosynthesis. The tight interaction within the multi-molecular complexes is likely to limit evolution of the core photosynthetic mechanisms (Shi *et al.*, 2005). Consequently, the observed photosynthetic variation between species (e.g. Wullschleger, 1993) and within species (e.g. Driever *et al.*, 2014; Prinzenberg *et al.*, 2018; Taniyoshi *et al.*, 2020; Faralli and Lawson, 2020; Acevedo-Siaca *et al.*, 2020, 2021a; McAusland *et al.*, 2020) is most prominent in the remaining 3000 genes, whose coordinated action mediates photosynthesis (P. Wang *et al.*, 2017). It should be noted that in some cases photosynthetic variation is explained by variation in the abundance of core photosynthetic components—while maintaining the original function (Yin *et al.*, 2010; Kasajima *et al.*, 2011; Chao *et al.*, 2014; Simkin *et al.*, 2019; Rungrat *et al.*, 2019). Furthermore, in line with the omnigenic model, which argues that essentially any gene expressed in a tissue will be in some way involved in the complex phenotype of that tissue (Boyle *et al.*, 2017), many more genes are required for general functioning of a plant to ensure that photosynthesis can occur. The notion that natural genetic variation for photosynthetic functioning occurs primarily outside the basic biochemical mechanisms is further supported by studies in which phenotypic variation in photosynthetic traits is linked to the underlying genetic variation. The absence of core photosynthetic components amongst these genes confirms that variation in photosynthesis lies primarily outside the core photosynthetic processes (e.g. Q. Wang *et al.*, 2017; Van Rooijen *et al.*, 2017; Oakley *et al.*, 2018; Rungrat *et al.*, 2019; Adachi *et al.*, 2019; Prinzenberg *et al.*, 2020). With few genes identified so far in genetic studies, we do not state there is no functional

natural genetic variation in core photosynthetic components, but the evidence points to the thousands of genes outside the core photosynthetic components as holding most of the genetic variation underlying variation in photosynthesis.

A lot of what we know now about the genetics of photosynthesis is based on the studies of induced mutations in *Chlamydomonas reinhardtii* (Levine, 1968) and *Arabidopsis* (Scheller *et al.*, 2001; Alonso *et al.*, 2003). The identified photosynthetic mutants were essential to assign biochemical functions to the corresponding genes and proteins (Rochaix, 2004). While mutants selected upon induced mutagenesis often display a drastic and obvious phenotype, as they are frequently caused by loss-of-function, or ‘knock-out’, mutations that disrupt gene function completely, natural phenotypic variation is generally much more subtle than the phenotypic variation seen in selected mutants. The drastic phenotype of knock-out mutants hardly ever allows the plants to survive the dynamic conditions encountered in the field. Natural genetic variation rarely involves knock-out mutations, but rather mutations that modify the function of the gene, often only slightly, leading to fitness enhancing, rather than disrupting, phenotypic changes.

The main route to explore natural genetic variation is using genetic mapping approaches to reveal quantitative trait loci (QTLs) underlying phenotypic differences in photosynthetic processes (Box 1; Fig. 1). Besides the already mentioned studies on linking genetic variation to photosynthetic variation, there is a growing body of literature on mapping studies in many plant species (e.g. Jung and Niyogi, 2009; Lowry *et al.*, 2013; Chao *et al.*, 2014; Ortiz *et al.*, 2017; Feldman *et al.*, 2018; Joynson *et al.*, 2021; Ferguson *et al.*, 2021). These QTLs may be selected in marker assisted breeding for improved photosynthesis, where information on the candidate genes is not essential to improve the trait. However, identifying a QTL will not provide much information on the biological mechanism responsible for the phenotypic differences caused by the QTL, which would be essential in expanding our knowledge on photosynthesis. While many QTLs for photosynthetic phenotypes are known, each QTL typically covers a genomic region containing dozens of candidate genes (Box 1), and only rarely have the causal genes, and the allelic DNA sequence variation, been functionally validated (Bernardo, 2016). If we want to know more about the role of standing genetic variation for photosynthesis in plant growth and crop yield, the causal genes need to be identified, and the allelic variation for these genes needs to be studied to understand how it causes phenotypic differences. In this review we discuss why photosynthesis is rarely maximal in nature and difficult to improve through plant breeding without detailed understanding on genotype–phenotype relations. We also show how genetics can be used to reveal standing genetic variation for photosynthetic traits and to learn more about photosynthetic regulation, and how this may be applied to improve the photosynthetic performance of crops, and ultimately crop yield.

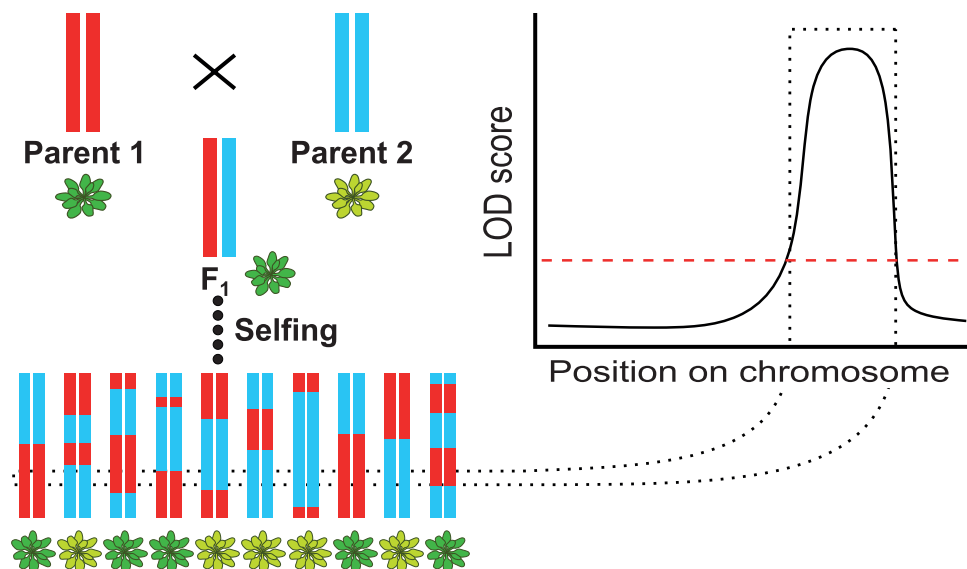
### Box 1. The concept of genetic mapping

To reveal genetic variation for a given phenotype, it is essential to correlate phenotypic variation to causal genotypic variation. For decades this has been done by genetic linkage mapping approaches (Nordborg and Weigel, 2008). A linkage map represents the order of and distance between genetic markers, based on the recombination frequency between the markers, for each chromosome.

Recombination frequencies are best determined in genetically segregating populations, often progeny of a cross between two genetically distinct parents. When such parents are diploid and inbred, and thus homozygous, there will at most be two different alleles per marker or gene. Alleles are sequence variants of the same stretch of DNA, which can comprise a gene or part of a gene, but can also correspond to non-coding DNA. Alleles can only be distinguished based on DNA sequence differences, such as a single nucleotide polymorphism (SNP), an insertion or deletion (InDel), or another kind of sequence variation. Different alleles may confer the same plant phenotype. In mapping approaches the variation for a particular trait, say chlorophyll content, will be determined and correlated with variation between the marker alleles at each genetic locus. If a genetic locus contributes to the trait, this will be due to an allelic difference in one of the genes residing at that locus, meaning one allele contributing to below average chlorophyll content, the other allele contributing to above average chlorophyll content (Fig. 1). Often the marker is not a genetic sequence variant of the gene involved, but genetically closely linked to it. This means, the marker identifies a genetic locus, as it resides in the vicinity of a gene for which each of the two parents of the segregating population carries a different allele, one contributing to low chlorophyll content, the other to high chlorophyll content. Since chlorophyll content is a quantitative trait, expressed in values rather than a classification, such a locus is generally referred to as a quantitative trait locus, or QTL. The likelihood of the association between a marker and the phenotype depends on the mean effect size difference between the allelic groups and the standard deviation around the mean of the phenotype (Xu, 1995). Consequently, QTLs for phenotypes with larger effects and smaller standard deviations will be mapped with more confidence than those for phenotypes with smaller effects or larger standard deviations. Likewise, if a trait is highly polygenic and affected by many genes with different alleles in the two parents, the effect size attributable to each gene is smaller than where the trait is only affected by one variant gene, and the QTLs are harder to map (Korte *et al.*, 2013).

Often biparental mapping approaches are employed in linkage mapping that can use different types of segregating populations. This could be single-use F<sub>2</sub> populations, or it could be ‘immortal’ populations of recombinant inbred lines (RILs), near-isogenic line (NILs), back-cross inbred lines (BILs), or doubled haploid (DH) lines (Bazakos *et al.*, 2017). RILs are constructed via repeated selfing of F<sub>2</sub> progeny of one F<sub>1</sub> hybrid, propagated through single-seed-descent. NILs and BILs are constructed via recurrent backcrossing of F<sub>1</sub> or F<sub>2</sub> progeny to one of the parents, and DHs are constructed through microspore culture or parental genome elimination in the F<sub>1</sub>, and subsequent (spontaneous or induced) doubling of the haploid genome (Bazakos *et al.*, 2017). Besides biparental populations, multi-parent advanced generation inter-cross (MAGIC) populations may be used, derived from crosses between different F<sub>1</sub> progeny with more than two parental genotypes (Beyer *et al.*, 2008). While linkage mapping provides strong statistical power to identify a QTL, it does not provide a high resolution as to which allelic difference causes a QTL. A higher resolution can be achieved with subsequent fine-mapping so as to identify sufficiently small regions of the genome holding only one or a few genes associated with the QTL (Jaganathan *et al.*, 2020).

The increasing ease of generating whole genome sequences has made higher resolution physical maps available, based on a large set of densely spaced sequence markers, often SNPs or InDels (Zargar *et al.*, 2015). Diversity panels consisting of different, not obviously related, genotypes, representing many different recombination events are increasingly used to provide marker-dense maps to identify QTLs (van Bezouw *et al.*, 2019). They often consist of different (natural) accessions of a (wild) species collected at different geographic locations and different niches, but can also be composed of different varieties or breeding lines of a crop species. Such populations are especially attractive for genome wide association studies (GWAS), in which phenotypes are tested for association with small genomic regions, due to the high frequency of recombination events in the population. QTLs identified this way are far narrower than those identified upon linkage mapping in a biparental population, and thus contain far fewer candidates for the causal gene, though the confidence to map such loci is lower than for QTLs identified in biparental populations (Nordborg and Weigel, 2008).



**Fig. 1.** Concept of QTL mapping. The example illustrates how a recombinant inbred line (RIL) population is used to correlate genetic variation (depicted in blue and red) and phenotypic variation (depicted in light green and dark green). The likelihood of the association is given as logarithm of odds (LOD) score, where higher values point to stronger associations. The regions on the genome, a locus, with a LOD score above the multiple-testing-corrected threshold is termed a QTL. The principle shown here for a RIL population can be used in all types of bi- and multiparental and GWA mapping approaches.

## Improving the photosynthetic performance of crops

### *Can photosynthesis be improved?*

While there is plenty of natural and induced genetic variation for photosynthetic functioning, the question arises whether photosynthetic performance can actually be improved by breeding, using either conventional or novel plant breeding techniques. The term ‘improving’ is very broad, and depending on the perspective it may have different meanings (Zelitch, 1975). In this context we refer to improving photosynthetic performance, as either capturing more light energy or using each unit of absorbed light energy to more efficiently fix  $\text{CO}_2$ , under some set of environmental conditions, relative to what is currently possible with a reference genotype or an elite cultivar. The ‘improvement’ may be realized at any of the levels of organization at which photosynthesis is normally measured (e.g. enzyme, protein complex, thylakoid, chloroplast, leaf, and canopy), but in an agricultural context it will be important that the improvement can be seen to act at the level of the canopy or field. With the global human population increasing and the climate changing rapidly, the need for agriculture to constantly push the limits of crop production is likely to increase.

Crop yield can be understood in terms of the efficiency indices: (i) light energy interception by the canopy, (ii) the conversion of that absorbed light energy into the chemical energy of biomass, and (iii) the harvest index (Monteith, 1994; Long *et al.*, 2006). Yield increases following the Green Revolution were remarkable and largely inspired by increases in the harvest index and light interception index of crops. By further

improving these two efficiency indices, plant breeding contributes to an annual increase in crop productivity. However, this increase is levelling off for the most important staple crops, making it very challenging to meet the increasing demand for higher crop yield (Zhu *et al.*, 2010; Ray *et al.*, 2012). The third index contributing to crop yield, describing the efficiency of the conversion of intercepted radiation into the chemical energy of biomass, is largely determined by photosynthesis and respiration (Monteith, 1994; Long *et al.*, 2006). To illustrate the scope of improvement, for soybean, a  $\text{C}_3$  species, the current efficiency of intercepted radiation is such that 1.5% of the full spectrum of solar energy per unit area is converted into biomass (Zhu *et al.*, 2010). The theoretical maximum calculated for  $\text{C}_3$  species is 4.6%, implying there is potential for a staggering 3-fold improvement in biomass production by improving photosynthesis (Zhu *et al.*, 2010).

### *Why has natural selection not led to improved photosynthesis?*

If energy conversion in photosynthesis is so important, then why has this not been maximized by natural selection and evolution? The answer to this is complex, but it is likely that photosynthesis has been optimized rather than maximized. Photosynthesis is clearly an important process for plants that brings with it the benefits of fixed carbon and energy, but it also brings with it costs; reconciling these costs with the benefits of the process in different environments results in different optimal combinations of photosynthetic properties. In a rather crude way this can be seen in the morphology of parasitic

plants in which photosynthesis is no longer used. The parasitic orchid *Rhizanthella gardneri*, for example, has no above ground parts and even flowers underground. It (and many other parasitic plants) dispenses with the costs of the above ground architecture necessary to support and position its leaves so they can function well as light absorbers for photosynthesis and at the same time be supplied with water and be able to export assimilates. This also implies that understanding the optimization of photosynthesis is not a matter that can be understood by analysing just leaf-level photosynthesis, but needs to be understood at the whole plant or, as is the case in agriculture, the canopy level.

Moving onto a simpler (albeit more restrictive) view of photosynthesis and focusing on leaves, there are numerous parameters that can be used to define photosynthesis. One of the most important of these is the maximum rate of photosynthesis as this defines a boundary to the effectiveness of the process as a source of carbon and a means for storing energy. The maximum rate can be that of a leaf, a genotype, a species, a type of plant, etc. Surprisingly, we know of no thorough theoretical analysis of the maximum leaf-level photosynthetic rate since that of Nobel (1991). He estimated for  $C_3$  photosynthesis a maximum theoretical  $CO_2$  assimilation rate of  $55 \mu\text{mol m}^{-2} \text{s}^{-1}$  at an irradiance  $2000 \mu\text{mol m}^{-2} \text{s}^{-1}$ . This compares favourably with the highest measured rate of  $C_3$  photosynthesis (see Box 2)

### Box 2. Maximizing photosynthesis in nature

There are examples in nature where optimal photosynthesis is close to maximal photosynthesis. For some species growing in a sunny, warm, nutritious environment with ample water, but little plant competition and low pathogen and herbivore exposure, it may be advantageous to maximize photosynthesis, especially if the growing season is (very) short. Such conditions occur, but are rare. Think about semi-deserts, with generally sparse vegetation, but occasionally sufficient rainfall to allow short-lived, abundant plant growth. Some of the species with the highest photosynthetic rates are found at such sites, such as the winter annual *Chylismia claviformis*, occurring in the dry semi-deserts of the southwestern part of North America. This species has one of the highest  $CO_2$  assimilation rates reported for  $C_3$  plants, exceeding  $60 \mu\text{mol m}^{-2} \text{s}^{-1}$  (Mooney *et al.*, 1976). Another remarkable species is *Amaranthus palmeri*, also from North American deserts, which is reported to show assimilation rates exceeding  $70 \mu\text{mol m}^{-2} \text{s}^{-1}$ , at a leaf optimum of  $42^\circ\text{C}$  (Ehleringer, 1983). These examples can serve as models to understand how maximal photosynthesis has been selected in nature, and as examples of what may be possible by breeding.

of about  $60 \mu\text{mol m}^{-2} \text{s}^{-1}$ . Not many plants, however, come even close to these very high rates of photosynthesis. Annual  $C_3$  crop plants, species that are normally considered to have high photosynthetic rates, have rates of  $20\text{--}30 \mu\text{mol m}^{-2} \text{s}^{-1}$  (Nobel, 1991), and there is considerable variation in the maximum photosynthetic capacity between types of plants (e.g. Larcher, 1995). In addition to maximum photosynthetic capacity, there are other photosynthetic parameters that contribute to the photosynthetic properties of a plant, for example water use efficiency (e.g. van den Boogaard *et al.*, 1997), phosphorous use efficiency (e.g. Denton *et al.*, 2007), nitrogen use efficiency (e.g. Evans, 1989), light-limited quantum yield (largely on a  $C_3\text{--}C_4$  axis) (e.g. Ehleringer and Pearcy, 1983), and responses to fluctuating light (e.g. Harbinson and Woodward, 1984). Photosynthesis therefore exhibits rich variation in its properties throughout the plant kingdom though this variation may not always be completely understood in terms of a process of optimization. There are some examples where the link between the costs and benefits of a photosynthetic property are more straightforward. Across species, leaf nitrogen is broadly related to maximum photosynthetic rate (e.g. Evans, 1989; Reich *et al.*, 1994) and within a species this relationship is much stronger (e.g. Evans, 1988; Makino *et al.*, 2003). More photosynthesis needs more nitrogen but many habitats are nitrogen limited, making higher photosynthetic rates more uneconomic, resulting in an optimum for photosynthesis that is less than the potential maximum rate of  $60 \mu\text{mol m}^{-2} \text{s}^{-1}$ . If a plant has an insufficiency of nitrogen, photosynthesis will be depressed relative to the maximum achieved under conditions of adequate nitrogen nutrition (Evans, 1988). So the maximum rate of photosynthesis actually varies a lot from leaf to leaf and in most cases needs to be seen as the result of an optimization process rather than as a biophysical limit.

To place this into perspective, in natural environments the fittest genotypes are selected, where fittest is defined as the ability to reproduce best (Popper, 1959). Even though improved photosynthesis might result in increased growth or higher seed yield, this might not be advantageous in a natural setting. Biotic and abiotic stresses, rather than suboptimal photosynthesis, are likely to impose more limitations on plant growth and reproduction in the field. Therefore photosynthesis is unlikely to be a permanent limitation on growth, and selection for improved photosynthesis may only infrequently occur, providing a poor driver of evolution. Furthermore, once a species has evolved its physiology to support photosynthesis in a given environment, adapting to a new environment can prove challenging (Leimu and Fischer, 2008). While in some cases this happens naturally, in other cases the required natural genetic variation is not available. If local populations of a species are geographically dispersed, gene flow will be limited making exchange of adaptive solutions difficult. Adaptation may also require too many steps to be taken before an optimum is reached, and if none of the intermediate steps improves fitness, the chances of reaching the optimum will be very low (as described by fitness

landscapes). In other words, the complex nature of photosynthesis, which brings with it interdependency between components of the system, limits the options for changing any part of the system, and as such a change might prove to be disruptive for the system as a whole. Consequently, photosynthetic capacity can be fixed at a level well below the maximal for a given environment.

### *Why has plant breeding not led to improved photosynthesis?*

While natural systems may rarely select for plants with maximum photosynthesis due to various environmental and genetic constraints, these constraints may be removed for crop species, as agricultural systems are fundamentally different from most natural environments. Environmental conditions impairing production are actively minimized by watering, fertilizing, weeding, and pest management of crops. Trade-offs that may exist in a natural condition, favouring optimal over maximal photosynthesis, may be irrelevant in agriculture. Since natural selection for optimal photosynthesis has resulted in genetic variation for photosynthetic traits in many species, such could be exploited to improve photosynthetic performance of crops in agriculture. Plant breeding can facilitate gene flow to interconnect the optima of fitness landscapes, and allow the best photosynthetic performance to be selected in a way that would be difficult to achieve in nature. Studies into historically released cultivars show that in some crops increases in photosynthetic rate have been made, but in other crops there is no sign of such increase (Table 1). So, plant breeding has occasionally contributed to improved photosynthesis, though in part unintentionally due to a correlation between yield and photosynthesis. Yet, the absence of widespread photosynthetic

improvements in crops, and the presence of considerable variation in photosynthetic performance in elite cultivars, demonstrates that there is no simple correlation between yield and photosynthetic performance (e.g. Driever *et al.*, 2014). This is likely due to the complex interactions between different mechanisms that act as bottlenecks for improving photosynthetic performance indirectly via selection for higher yield. It is here where direct phenotyping of components of photosynthetic functioning can contribute to improved photosynthesis and to identification of the bottlenecks that currently prevent yield increases.

### *Status quo in improving photosynthesis*

To assess the bottlenecks that can be targeted for photosynthetic improvements, the energy losses in photosynthesis have been modelled (Zhu *et al.*, 2010). The current understanding of the biochemical function of especially the core of the photosynthetic components allows the pinpointing of mechanisms that form bottlenecks. Alleviating these bottlenecks can range from relatively simple solutions, such as bypassing photorespiration and increasing the recovery from the photoprotective state, to more complex solutions such as converting crops from C<sub>3</sub> to C<sub>4</sub> photosynthesis (Zhu *et al.*, 2010). These elegant advances, achieved through genetic modifications, have been shown to be very effective (Kromdijk *et al.*, 2016; Driever *et al.*, 2017; South *et al.*, 2019; Simkin *et al.*, 2019; López-Calcagno *et al.*, 2020), but can also be very challenging to achieve (Ermakova *et al.*, 2020). It is noteworthy that the effective improvements in recovery from the photoprotective state that were shown to result in increased yields in tobacco (Kromdijk *et al.*, 2016) were not reproducible in Arabidopsis (García-Molina and Leister, 2020), meaning that it is not a one-size-fits-all solution.

**Table 1.** Overview of photosynthetic improvements in historically released cultivars in four major crops

Crop	Range of released date cultivars used in study	Main finding on photosynthetic improvements	Reference
Rice	1882–1976	Photosynthetic rate increased in the first half of 20th century, but afterwards improvement was less pronounced	Sasaki and Ishii (1992)
Rice	1893–1991	Photosynthetic rate improved only in some cultivars, but overall the photosynthetic rate correlated poorly with biomass	Zhang and Kokubun (2004)
Rice	1966–1997	Maximum photosynthetic rate decreased until 1980, but recovered slightly afterwards	Hubbart <i>et al.</i> (2007)
Wheat	1981–2008	Photosynthetic rate increased, but after early 2000 improvement was less pronounced	Zheng <i>et al.</i> (2011)
Wheat	1958–2007	No increase in conversion efficiency	Sadras <i>et al.</i> (2012)
Wheat	1967–2010	Photosynthetic rate increased	Ding <i>et al.</i> (2020)
Maize	1931–~1990	No increase in conversion efficiency	Ding <i>et al.</i> (2005)
Soybean	1934–1992	Photosynthetic rate increased	Morrison <i>et al.</i> (1999)
Soybean	1951–2006	Photosynthetic rate increased	Jin <i>et al.</i> (2010)
Soybean	1923–2007	Conversion efficiency increased	Koester <i>et al.</i> (2014)
Soybean	1923–2007	Maximum photosynthetic capacity has not increased, but daily carbon gain has increased	Koester <i>et al.</i> (2016)

Nevertheless, these approaches underline that improvements can be made, and undoubtedly more of these developments will follow.

So far most progress has been made in mechanisms that are part of the core photosynthetic apparatus, or processes directly linked to them. The absence of natural genetic variation in the core photosynthetic machinery makes genetic modification a suitable approach (Ort *et al.*, 2015). However, this dismisses the potentially thousands of other genes, for which there is natural genetic variation, that hold the capacity to improve photosynthesis. Moreover, genetic modification limits the range of crops that can be improved and the countries in which such crops can be grown. This is another reason to consider improving photosynthetic performance of crops via natural genetic variation as an attractive alternative. To improve photosynthetic performance through selection of standing genetic variation, two approaches can be followed, which are not mutually exclusive (Fig. 2). The most straightforward way is to exploit genetic variation for photosynthetic phenotypes in marker assisted breeding and genomic selection approaches of crops. This will of course need awareness of this option among plant breeders, as well as suitable germplasm, appropriate phenotyping facilities, and the right statistical framework, as support. A more sophisticated way would be to study the genes involved in photosynthesis that are not yet identified as such, and determine the genetic constraints underlying physiological bottlenecks in crop photosynthesis. With such knowledge, a systematic approach can be designed to improve crop photosynthesis and yield in a targeted way. This would open up knowledge on the thousands of genes involved in photosynthesis for which there is genetic variation, which so far have remained undiscovered and cannot be used to improve photosynthesis. To illustrate this, of the ~3000 Arabidopsis nuclear genes encoding a protein predicted to have a chloroplast target peptide, and to likely play a role in photosynthesis, only 15% have a known role in photosynthetic performance (Fristedt, 2017). Only for a very few of these genes has the phenotypic effect of allelic variation been explored.

## Quantitative and molecular genetics

So, what are the options on how quantitative and molecular genetics can help to reveal natural genetic variation for photosynthetic traits and establish the identity and function of the genes involved? To guide this discussion, a schematic overview is given in Fig. 2.

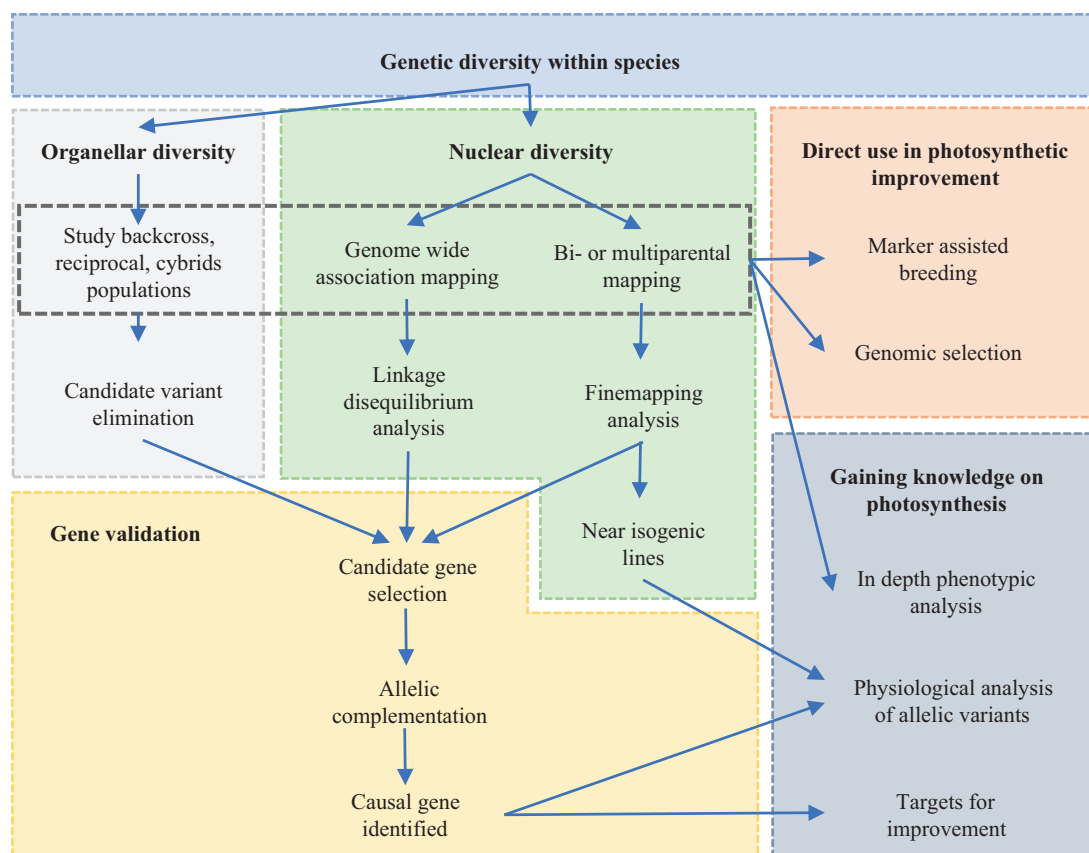
### *Genetic mapping approaches to reveal nuclear genetic variation*

Most plant traits are quantitative, meaning they are expressed in (non-discrete) values. Examples of quantitative traits are flowering time, plant biomass, and the efficiency of

photosynthesis. The identification of genes underlying quantitative traits can be resolved by genetic mapping (Box 1). In order to decide on which mapping approach to use, it is important to discuss the advantages and disadvantages of the most common approaches. In the past, most genetic mapping involved bi- and multiparent mapping populations, but in the last decade, the use of diversity panels for genome wide association studies (GWAS) has gained importance (Bazakos *et al.*, 2017). While we will focus on discussing approaches for diploid inbreeding species, the concepts also apply to outbreeding and polyploid species.

Bi- and multiparental populations (Box 1) are very useful in revealing most of the heritability for a specific phenotype. The low genetic diversity in these populations, derived from only a few parental lines, contributes to a high statistical power to detect QTLs, due to the roughly equal distribution of alleles (e.g. 50:50 at each locus in a biparental population or 25:25:25:25 in a tetra-parental population). Though they require several generations to construct, once available they allow the use of replicate plants per genotype in the study, thereby providing high confidence of plant phenotypes. Immortal populations also provide the opportunity to use the same population for replicate experiments in different environmental conditions, allowing the identification of robust QTLs, expressed in several environments or over several years. Bi- and multiparental populations can also be effective in resolving non-additive genetic variation, such as that caused by epistatic interactions. Epistasis is a very common biological phenomenon, meaning that the phenotypic effect of an allele at a specific locus depends on the effect of an allele at another locus (Mackay, 2014). Due to the interaction of these alleles, the statistical power to detect epistatic QTLs is lower than for additive QTLs, and sometimes insufficient, although always better for bi- and multiparental populations than for diversity panels used for GWAS (Nelson *et al.*, 2013; Mackay, 2014). Since photosynthesis is so complex, it is a prime example of a process in which epistatic interactions are likely to play a role. Thus, bi- and multiparental mapping holds great potential for revealing genetic variation for photosynthesis, especially when a large population can be screened, and even more so if the population is derived from phenotypically distinct parental lines, previously identified in a diversity screen for GWAS.

GWAS make use of diversity panels, rather than bi- or multiparent populations, which means a GWAS population, or diversity panel, typically represents much more genetic diversity and a higher density of recombination events between genomic regions than present in bi- or multiparental populations. The genomic region in which no recombination is found, when comparing all accessions in the diversity panel, is often referred to as being in linkage disequilibrium (LD) (Flint-Garcia *et al.*, 2003). In wild species, the decay of LD may be in the order of hundreds to a few thousands of base pairs, while in domesticated species, which have gone through a genetic bottleneck, the decay of LD can be in the range of



**Fig. 2.** Schematic overview of how natural genetic variation within a species can contribute to improving photosynthesis. Through analysis of nuclear and organellar genetic diversity, interesting marker phenotype associations can be revealed. These can be used directly in marker assisted breeding and genomic selection programmes, and the populations can be used to study correlations and responses of photosynthetic phenotypes. In order to gain more knowledge on how natural genetic variation contributes to photosynthesis, the casual genes have to be identified via gene validation methods. Near isogenic lines and transformed lines containing different alleles can be used to deepen physiological processes in which these genes play a role. Ultimately some of these will form new targets for photosynthetic improvements.

100-kbp (Flint-Garcia *et al.*, 2003). The high density of recombination events in wild species gives GWAS a very high mapping resolution (Bazakos *et al.*, 2017), and often restricts the QTL region to cover only a small number of candidate genes. Additionally, a diversity panel for GWAS is easily composed from available crop germplasm, or may be collected in a few months for wild species. If the species is autogamous, collected accessions may already be largely homozygous and nearly ‘immortal’, and thus may only need to be genotyped once. This has prompted GWAS as the upcoming approach to map genetic diversity in the past decade (Bazakos *et al.*, 2017). It was first established for *Arabidopsis* (Atwell *et al.*, 2010), but since then has been used for many more plant species (e.g. as reviewed in van Bezouw *et al.*, 2019; Gupta *et al.*, 2019; Hao *et al.*, 2020).

The high genetic variation of the diversity panels used in GWAS also means there may be several alleles for a locus, some of which may be represented at a high frequency in the panel, while others may be rare (Forsberg *et al.*, 2015). Alleles that are present at low frequencies are unsuitable for QTL analysis, as

the statistical power to detect the QTL will be too low (Korte and Farlow, 2013). Adaptations to specific environments, as is often the case for photosynthesis-related traits, may result in high allele frequencies in local, adapted, populations, but when considered over the full diversity panel, comprised of accessions from many, global populations, the favoured allele may be rare, resulting in poor statistical power to detect it (Barboza *et al.*, 2013; Lopez-Arboleda *et al.*, 2021). Another important aspect to consider in GWAS is the effect of kinship in the ability to detect QTLs. If the diversity panel contains many accessions that are much more closely related to each other than to other (groups of) accessions, the population structure that results from this kinship may enhance the false detection of a QTL, as several markers will be shared among those related accessions, and a marker–phenotype association may occur just by chance. This means population structure must be accounted for, and can also statistically be corrected for, in GWAS, to avoid false associations to be detected as QTLs (Korte and Farlow, 2013). This does, however, mean that genetic variation that is tightly associated with a local lineage is difficult to detect, as



the statistical correction ignores such associations. To get high resolution mapping, a very high genetic marker density will be needed, often in the range of hundreds of thousands or even millions of single nucleotide polymorphism (SNP) or insertion or deletion (InDel) markers. While this is now technically feasible (Jaganathan *et al.*, 2020), it poses problems with respect to the statistical analysis. With more markers to be tested for marker–phenotype associations, the chance of finding an association at random, without an actual causal relation between genetic variant and phenotype, will increase. This ‘multiple testing error’ can be corrected for by adjusting the significance threshold needed to classify a marker–trait association probability as significant. Most commonly used methods correct for multiple testing errors by taking a naïve Bonferroni or false-discovery rate threshold, or alternatively by performing permutation testing to take the underlying phenotypic distribution into account (Storey and Tibshirani, 2003; Freudenthal *et al.*, 2019, Preprint). QTLs identified with probability scores (often expressed as the minus logarithm of the *P*-value) above these thresholds are most likely to be valid QTLs that warrant further follow-up study. However, one should realize that QTLs that do not reach the significance threshold may still be valid, and QTLs that exceed the threshold may be false (Korte and Farlow, 2013). As a consequence of these issues, GWAS diversity panels often display large phenotypic diversity, and high heritability, but may only reveal a few QTLs. In most cases, QTLs with relatively low explained variance are detected, often barely reaching the significance threshold. This implies that while GWAS will reveal associations, the number of identified QTLs is often an under-representation of the total number of available QTLs present in the diversity panel. Especially when very few QTLs are found, one may want to resort to additional bi- and multiparental mapping approaches, using new or existing populations generated by crossing genotypes with interesting phenotypes as identified in the phenotypic screen that initiated the GWAS.

All of these genetic mapping methods, with their advantages and disadvantages, are commonly used in photosynthetic research. While somewhat variable between studies, biparental mapping (e.g. Jung and Niyogi, 2009; Yin *et al.*, 2010; Lowry *et al.*, 2013; Oakley *et al.*, 2018; Feldman *et al.*, 2018) seems equally successful in finding QTLs as GWAS (e.g. Chao *et al.*, 2014; P. Wang *et al.*, 2017; Ortiz *et al.*, 2017; Van Rooijen *et al.*, 2017; Rungrat *et al.*, 2019; Prinzenberg *et al.*, 2020; Joynson *et al.*, 2021; Ferguson *et al.*, 2021). Initially much of this work was performed in model species like *Arabidopsis*, for which suitable mapping populations are readily available, but these approaches are increasingly feasible in crop species. In order to develop immortal mapping populations as efficiently as possible, speed breeding allows for more generations per year, and fast construction of mapping populations in crops (Watson *et al.*, 2018). The costs of genotyping are still decreasing, and novel, cheaper genotyping approaches are still being developed (Gaio *et al.*, 2019, Preprint). Consequently, depending on the crop, and the

time, budget, phenotype, and aim of the project, one should be conscientious in selecting the appropriate mapping approach.

### *Marker assisted breeding and genomic selection*

Once a QTL is identified as holding interesting genetic variation for a trait, it can feed directly into a marker assisted breeding programme to introduce and establish the favourable allele in elite crop lines, thereby improving photosynthesis. The success and impact of this will rely on the number of QTLs affecting the trait, and the percentage of phenotypic variance that is explained by the QTL. Photosynthesis is a highly polygenic trait, affected by many QTLs, often with relatively small effect sizes. Consequently, introgression of individual QTLs will be logistically complex, requiring large recombinant populations or many subsequent cycles of crossing and selection, to be able to select the rare recombinants that will combine several alleles with positive effects on the trait, while retaining all other important crop traits. Instead, genomic selection approaches might be more suited. Genomic selection attributes a weight to each individual marker, which depends on its association with the trait (Meuwissen *et al.*, 2001). So rather than identifying individual QTLs, it predicts the breeding value of genotypes based on the overall prediction of all markers combined, also known as genomic estimated breeding value for a genotype (Crossa *et al.*, 2017). In this way, the underlying function remains unknown, but in contrast to classical breeding where individual alleles have to be incorporated, in genomic selection approaches many alleles with a combined large effect on the trait can quickly be incorporated into elite breeding material. This has been shown to be very effective in yield improvements and disease resistance breeding (Beyene *et al.*, 2015; Rutkoski *et al.*, 2015), and is used in commercial breeding programmes for soybean and maize (Bernardo, 2016). Genomic selection has not yet contributed to improvements in photosynthetic performance, but this is likely to change when using natural genetic variation for improving photosynthesis becomes more mainstream. This will be especially the case if more becomes known on the interaction of photosynthesis QTLs with the environment, the so-called genotype × environment interactions. Photosynthesis phenotypes are prone to be affected by the environment (Murchie *et al.*, 2018), which means that alleles beneficial in certain crop production conditions may not be beneficial in another environment or agricultural system. Due to the speed at which genomic selection can proceed, different elite cultivars may rapidly be generated for a range of environments.

### *Fine mapping and candidate gene validation*

The identification of QTLs for photosynthetic traits in model and crop species has become more common in recent years, and QTLs can feed into marker assisted breeding programmes to improve photosynthesis. However, it is also relevant to study

the function of the gene and alleles underlying the QTL, to better understand their role in photosynthesis (Fristedt, 2017). Revealing their function can increase the versatility of genetic modification studies and pinpoint relevant physiological mechanisms, to subsequently improve photosynthesis (Fig. 2). While, in principle, mutant screens are very effective for revealing the function of genes (Belcher *et al.*, 2015; Li *et al.*, 2019), a gene knock-out can be lethal, making it impossible to study the gene function via a knock-out mutant. Moreover, a knock-out mutant may show an altered photosynthetic phenotype, but such may be a pleiotropic effect, and only one aspect of a complex mutant phenotype, as the gene is only indirectly involved in photosynthesis. In such cases, it will remain difficult to determine the actual function of the gene with respect to photosynthesis. It is here that the study of allelic variation, in which functional differences convey more subtle differences, is useful. Knowledge on such subtle allelic variation is also much more likely to contribute to improving photosynthesis by breeding. As gene function analysis can be time consuming, such is especially feasible in model species like Arabidopsis, whereupon the knowledge can be translated to crops, or serve as an example for targeted follow-up studies in crops.

In revealing the responsible gene(s) underlying a QTL, a common mistake is to assume that the causal gene is the one closest to the marker with the highest genotype–phenotype association. Such is generally not the case, and proper gene validation is needed using a limited set of candidate genes. In the case of QTL mapping in bi- or multiparental mapping approaches, QTLs are often mapped to a large genomic region containing hundreds of genes. Resolving these will need one or more rounds of fine-mapping, using the generation of additional segregating populations, to reduce the QTL region by means of additional recombinations (e.g. Adachi *et al.*, 2019). Fine-mapping can benefit from the use of near isogenic lines (NILs), which vary for the QTL region, but are otherwise isogenic (Alonso-Blanco and Koornneef, 2000). Especially since photosynthesis is affected by many genes, by using NILs one will be able to examine the consequences of the variation in the target region only, as there will be no genetic variation in the rest of the genome. NILs can also be very useful to study the physiological impact of the QTL variation (e.g. Adachi *et al.*, 2014). To facilitate the speed in which NILs can be developed and used, especially for crops, it is worth considering the use of heterogeneous inbred families as a bi- or multiparental population, as these hold regions that are still heterozygous, allowing the quick identification of NILs (Tuinstra *et al.*, 1997). In GWAS, the mapping resolution is sometimes enough to directly pinpoint the causal gene, although this will require the availability of all sequence polymorphisms in all genotypes of the GWAS population, which is rare (Jaganathan *et al.*, 2020). Consequently, the marker is often simply a pointer to a region without any recombination in any of the genotypes in the studied population, also known as a haplotype block (Gabriel *et al.*, 2002). The decay of LD will determine the size of the

haplotype block that carries the genetic variant responsible for the genotype–phenotype association. In principle all genes within this haplotype block could represent the variant allele causal for the target phenotype. Once the haplotype block is identified, it is tempting to determine possible causality of the obvious candidates in the region based on what is known about the predicted functions of genes in the block. One needs to be cautious though, as the function of many plant genes is still unknown, even for Arabidopsis ([www.arabidopsis.org](http://www.arabidopsis.org)), because of which one may focus on a likely candidate and miss the actual causal gene, and thus the opportunity to shed further light on the biology underlying the target phenotype (Baxter, 2020). As further outlined below, the actual identification of the DNA sequence variant, be it a SNP, InDel, or other, is not trivial, is often laborious, and is sometimes unachievable, which may be frustrating.

When the number of candidate genes has been narrowed down to about 10, either upon fine-mapping in the biparental population, or upon LD analysis in GWAS, the gene validation can start. As a first step, and as a first step only, a mutant analysis of the remaining candidate genes is often the preferred next step. If a mutant with a loss-of-function allele produces a phenotype consistent with the function of the hypothetical wild-type gene giving rise to the phenotype used in the QTL analysis, then that mutant may point to the causal gene. Knock-out lines may be obtained via available stock centres, such as the T-DNA insertion lines for Arabidopsis (<https://arabidopsis.info> or <https://abrc.osu.edu/>), or by CRISPR–Cas-mediated gene editing. As genotypes within a species can vary substantially for photosynthetic traits (Wójtowicz and Gieczewska, 2021), it is important to realize that the phenotype of a loss-of-function mutant may depend on the genetic background. Since most Arabidopsis T-DNA mutants have been generated in a Columbia (Col) background, there will not be any mutants for genes that are absent from Col or genes for which Col has a natural loss-of-function allele, both of which are common (Gan *et al.*, 2011). However, it is insufficient or even inappropriate to conclude on causality of the QTL by establishing that the phenotype of a knock-out allele of one of the candidate gene involves the same biological process that is studied in the QTL analysis. Note that with over 3000 genes involved in photosynthesis in Arabidopsis (which is about 1 in 10), there is a good chance that one of the genes of a haplotype block covering 10 genes will give a knock-out mutant photosynthesis phenotype. While this may indeed be the gene underlying the QTL, additional validation is needed to confirm this. Without such confirmation, one may have identified a gene that is somehow involved in photosynthesis, but not necessarily the gene underlying the QTL! A logical next step is to establish that there is genetic sequence variation between the alleles with contrasting phenotypes, to explain such phenotypic differences. Confirming variation in alleles may not be easy to do, as allelic variation may not be in the coding region, but in sequences regulating transcription. It may even be of

an epigenetic nature, such as for the *FWA* gene, controlling flowering time in *Arabidopsis* (Soppe *et al.*, 2000).

Once there are only a few candidate genes remaining, meaning genes for which there is allelic variation between genotypes that may explain the phenotypic differences between the genotypes, and which have a function in line with the studied trait, the final step of the gene validation can start. In general two approaches are used to confirm the identity of the causal gene underlying a QTL. Of those two approaches transgenic complementation is most often used. This involves transforming the allelic variants into a loss-of-function mutant background, to recreate the phenotypic difference or variation initially used to identify the QTL (e.g. Alonso-Blanco *et al.*, 2005; Bentsink *et al.*, 2006; Loudet *et al.*, 2007). One will need to generate several independent transformation events, especially if the phenotype depends on the expression level of the introduced allele, which may be higher if more copies are introduced, and may depend on the site of T-DNA insertion. Transgenic complementation may be replaced by gene editing approaches, to substitute one allele for the other (Molla *et al.*, 2021), but this may not be feasible for all types of allelic variants. The alternative approach is to use quantitative allelic complementation (Weigel, 2012; Turner, 2014). This relies on crossing one or more accessions carrying one allele of the QTL with both a wild type and a knock-out mutant of the target gene, and do the same with one or more accessions carrying the alternative allele. If the target gene is not the one underlying the QTL, the phenotypic difference between F<sub>1</sub>s with wild-type or mutant plants will be similar for both allelic variants, while it will be different if the target gene is indeed the one for which allelic variation underlies the QTL. A case in which this has been successfully used was presented by Van Rooijen *et al.* (2017), showing the role of *YS1* in photosynthetic response to an increase in irradiance.

Once the causal gene is convincingly identified, the different alleles can be used to study the physiological role of the gene (Fig. 2) and learn more about its significance for photosynthesis and the potential to use it to improve photosynthesis, either through breeding or genetic modification. So while there are several approaches to gain new insights in physiological functioning of photosynthesis, in attempts to improve photosynthesis it will be best to be aware of the allelic variation and have a deeper understanding of the physiological functioning associated with this variation.

### *Revealing organellar genetic variation*

All of the methods described above focus on the exploitation of genetic variation in nuclear genomes, thus ignoring the genetic contribution from organellar genomes. Photosynthesis, however, is predominantly associated with processes occurring in the chloroplasts. The chloroplast holds roughly 70 protein-coding genes, most of which are essential for photosynthetic performance (Rochaix 1997). Also, mitochondria play a role

in supporting photosynthesis (Nunes-Nesi *et al.*, 2008; Fan *et al.*, 2021), and they contain roughly 30 protein-coding genes. As organelles inherit mostly uniparentally, and recombination does not take place, beneficial alleles of organellar genes spread much less easily through the population of a plant species than alleles of nuclear genes. The uniparental inheritance also means that the study of natural genetic variation in these organelles through conventional mapping populations is much more difficult (Joseph *et al.*, 2013; Tang *et al.*, 2014).

Consequently, when studying genetic variation for photosynthetic traits, not only the nuclear genome, but also the organellar genomes and the nuclear–organellar interaction should be considered. This variation can be exploited for breeding purposes using recurrent backcrossing or the construction of cybrids, which are genotypes with novel combinations of nuclear and organellar genomes (Evans, 2007; Miclaus *et al.*, 2016; Roux *et al.*, 2016; Flood *et al.*, 2020; Lv *et al.*, 2020). Using cybrids it was conclusively revealed that phenotypic differences for photosynthetic traits can be caused by natural genetic variation in organelles (Flood *et al.*, 2020). As for the nuclear counterpart, it is also relevant to identify the causal organellar gene for a cytoplasmic trait. As fine-mapping is not possible in organellar genomes, cybrids with the same nuclear genome but different organellar genomes that differ in the candidate genetic variants can be used for gene identification. When the number of candidates is sufficiently low, organellar transformation methods can be used to reveal the causal gene. While recently many advances have been made, editing of chloroplast genes remains difficult (Molla *et al.*, 2021).

## **From natural genetic variation to crop improvement of photosynthesis**

In this review we have described how to identify natural genetic variation for photosynthetic traits, and perform validation of the genes and alleles involved. Photosynthesis is not an easy trait to study: it is highly polygenic and phenotypically highly responsive to environmental conditions (Zargar *et al.*, 2017; Kaiser *et al.*, 2018; Vico *et al.*, 2019). In crop production, environments are rarely constant, but typically highly dynamic, and inevitably, it will be challenging to reliably phenotype photosynthesis in dynamic conditions to support analysis of genetic variation (Soleh *et al.*, 2016, 2017; Flood *et al.*, 2020; McAusland *et al.*, 2020; Acevedo-Siaca *et al.*, 2021b). It will be more the rule than the exception that apparently similar phenotypes are caused by different QTLs. In addition, there is likely to be a strong genotype by environment interaction, meaning that alleles that improve photosynthesis in one condition may be unfavourable in another. Adding to the challenge is that there is poor correlation between photosynthesis in constant and dynamic conditions (Acevedo-Siaca *et al.*, 2021b). Therefore, to allow the screening of functional variation for photosynthesis in dynamic conditions, it is essential

to reproducibly mimic such conditions in controlled environmental facilities (as described by Murchie *et al.*, 2018) and to be able to phenotype photosynthesis at high throughput in such facilities, to reliably identify genetic variation amenable to breeding (van Bezouw *et al.*, 2019). At the moment this is possible for a range of photosynthetic parameters, but not for all (Siebers *et al.*, 2021). Especially the development of high-throughput phenotyping techniques to determine photosynthesis in controlled, but dynamic environments is required, as this will provide the reproducibility needed for genetic studies (e.g. Cruz *et al.*, 2016), and the opportunity to establish how to best breed for improved photosynthesis under field conditions. To assess the overall impact of the variation for a photosynthetic trait in a given environment, it is essential to use crop models to take into account the dynamic properties and broad ranges of environments (Wu *et al.*, 2017, 2019; Coast *et al.*, 2021). Models will also be very useful to identify the potential contribution of photosynthesis improvements on yield, given the variation on other yield components, such as root nutrient uptake or sink capacity (Yin *et al.*, 2022). Using an approach where the impact of genetic variation on crop performance in a given agricultural context can be predicted will allow the identification of alleles that are deemed interesting to incorporate in breeding material. In the case of genomic selection procedures, the weight of an individual marker can be tested via crop models in the environment of interest, to steer the breeding for improved photosynthesis in a range of environments.

## Conclusion

Breeding for elite cultivars has largely ignored improvements in photosynthesis. While promising advances in improved photosynthesis may be achieved by genetic modification (Ort *et al.*, 2015; Hitchcock *et al.*, 2022), we argue that using natural genetic variation for photosynthesis holds an equally promising potential for improvement. While biotechnology applications largely focus on the core photosynthetic pathway, natural genetic variation will reveal the benefit of knowledge on the thousands of genes that ensure proper embedding of photosynthesis in plant metabolism and growth. Flood *et al.* (2011) already outlined natural genetic variation as a promising route, and indeed in the past decade an increasing number of studies have appeared revealing QTLs in different crops in a plethora of environments and photosynthetic traits. In the absence of a focused effort, largely caused by the complexity of the trait, it appears that few of these QTLs have been used to support breeding programmes, or even to gain an in-depth understanding of the physiological processes they could unveil. Natural plant species may not have maximized photosynthesis, as is desirable for high-yielding crops, but they evolved robust forms of photosynthesis that allow them to cope with many different, dynamic environments. In this sense, nature

has taken millions of years to try new or alternative methods of converting sunlight, in both dynamic and broad-ranging environments. By identifying the bottlenecks in the photosynthetic function in these environments, and through targeted genetic studies on these traits, nature's often elegant solutions to problems can be explored and learned from. As far as natural genetic variation for photosynthesis goes, we live in exciting times.

## Acknowledgements

We thank René Boesten for his help in designing the figures, Ben Auxier, Roel van Bezouw, Phuong Nguyen, and Maarten Koornneef for their insightful discussions and comments on the manuscript, and two anonymous reviewers for their helpful comments and insights, which have improved the manuscript.

## Author contributions

TPJMT, LLL, JH, and MGMA: conceptualization; TPJMT: writing - original draft; TPJMT, LLL, JH, and MGMA: writing - review & editing.

## Conflict of interest

The authors declare no competing interests.

## Funding

This work was, in part, supported by the Netherlands Organization for Scientific Research (NWO) through ALWGS.2016.012 (TPJMT) and GSGT.2019.023 (LLL) and the EU-Horizon2020 Research Innovation Program projects CropBooster-P, [www.cropbooster-p.eu](http://www.cropbooster-p.eu) (grant number 817690; JH and MGMA), CAPITALISE (grant number 862201; JH and MGMA).

## References

- Acevedo-Siaca LG, Coe R, Quick WP, Long SP. 2021a. Variation between rice accessions in photosynthetic induction in flag leaves and underlying mechanisms. *Journal of Experimental Botany* **72**, 1282–1294.
- Acevedo-Siaca LG, Coe R, Wang Y, Kromdijk J, Quick WP, Long SP. 2020. Variation in photosynthetic induction between rice accessions and its potential for improving productivity. *New Phytologist* **227**, 1097–1108.
- Acevedo-Siaca LG, Dionora J, Laza R, Paul Quick W, Long SP. 2021b. Dynamics of photosynthetic induction and relaxation within the canopy of rice and two wild relatives. *Food and Energy Security* **10**, e286.
- Adachi S, Baptista LZ, Sueyoshi T, Murata K, Yamamoto T, Ebitani T, Ookawa T, Hirasawa T. 2014. Introgression of two chromosome regions for leaf photosynthesis from an *indica* rice into the genetic background of a *japonica* rice. *Journal of Experimental Botany* **65**, 2049–2056.
- Adachi S, Yamamoto T, Nakae T, *et al.* 2019. Genetic architecture of leaf photosynthesis in rice revealed by different types of reciprocal mapping populations. *Journal of Experimental Botany* **70**, 5131–5144.
- Alonso-Blanco C, Gomez-Mena C, Llorente F, Koornneef M, Salinas J, Martínez-Zapater JM. 2005. Genetic and molecular analyses of natural variation indicate *CBF2* as a candidate gene for underlying a freezing

- tolerance quantitative trait locus in *Arabidopsis*. *Plant Physiology* **139**, 1304–1312.
- Alonso-Blanco C, Koornneef M.** 2000. Naturally occurring variation in *Arabidopsis*: An underexploited resource for plant genetics. *Trends in Plant Science* **5**, 22–29.
- Alonso JM, Stepanova AN, Leisse TJ, et al.** 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–7.
- Arrivault S, Alexandre Moraes T, Obata T, et al.** 2019. Metabolite profiles reveal interspecific variation in operation of the Calvin-Benson cycle in both C<sub>4</sub> and C<sub>3</sub> plants. *Journal of Experimental Botany* **70**, 1843–1858.
- Atwell S, Huang YS, Vilhjálmsson BJ, et al.** 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–31.
- Barboza L, Effgen S, Alonso-Blanco C, Kooke R, Keurentjes JJB, Koornneef M, Alcázar R.** 2013. *Arabidopsis* semidwarfs evolved from independent mutations in *GA20ox1*, ortholog to green revolution dwarf alleles in rice and barley. *Proceedings of the National Academy of Sciences, USA* **110**, 15818–15823.
- Baxter I.** 2020. We aren't good at picking candidate genes, and it's slowing us down. *Current Opinion in Plant Biology* **54**, 57–60.
- Bazakos C, Hanemian M, Trontin C, Jiménez-Gómez JM, Loudet O.** 2017. New strategies and tools in quantitative genetics: how to go from the phenotype to the genotype. *Annual Review of Plant Biology* **68**, 435–455.
- Belcher S, Williams-Carrier R, Stiffler N, Barkan A.** 2015. Large-scale genetic analysis of chloroplast biogenesis in maize. *Biochimica et Biophysica Acta – Bioenergetics* **1847**, 1004–1016.
- Bentsink L, Jowett J, Hanhart CJ, Koornneef M.** 2006. Cloning of *DOG1*, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **103**, 17042–17047.
- Bernardo R.** 2016. Bandwagons I, too, have known. *Theoretical and Applied Genetics* **129**, 2323–2332.
- Berry JO, Yerramsetty P, Zielinski AM, Mure CM.** 2013. Photosynthetic gene expression in higher plants. *Photosynthesis Research* **117**, 91–120.
- Beyene Y, Semagn K, Mugo S, et al.** 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Science* **55**, 154–163.
- Beyer P, Morell M, Mackay I, Powell W.** 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* **11**, 215–232.
- Björkman O, Holmgren P.** 1963. Adaptability of the photosynthetic apparatus to light intensity in ecotypes from exposed and shaded habitats. *Physiologia Plantarum* **16**, 889–914.
- Boyle EA, Li YI, Pritchard JK.** 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186.
- Bungard RA, Ruban AV, Hibberd JM, Press MC, Horton P, Scholes JD.** 1999. Unusual carotenoid composition and a new type of xanthophyll cycle in plants. *Proceedings of the National Academy of Sciences, USA* **96**, 1135–1139.
- Chao M, Yin Z, Hao D, Zhang J, Song H, Ning A, Xu X, Yu D.** 2014. Variation in Rubisco activase (RCAβ) gene promoters and expression in soybean [*Glycine max* (L.) Merr.]. *Journal of Experimental Botany* **65**, 47–59.
- Coast O, Posch BC, Bramley H, Gaju O, Richards RA, Lu M, Ruan Y-L, Trethowan R, Atkin OK.** 2021. Acclimation of leaf photosynthesis and respiration to warming in field-grown wheat. *Plant, Cell & Environment* **44**, 2331–2346.
- Crossa J, Pérez-Rodríguez P, Cuevas J, et al.** 2017. Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science* **22**, 961–975.
- Cruz JA, Savage LJ, Zegarac R, Hall CC, Satoh-Cruz M, Davis GA, Kovac WK, Chen J, Kramer DM.** 2016. Dynamic environmental photosynthetic imaging reveals emergent phenotypes. *Cell Systems* **2**, 365–77.
- Denton MD, Veneklaas EJ, Freimoser FM, Lambers H.** 2007. *Banksia* species (Proteaceae) from severely phosphorus-impoverished soils exhibit extreme efficiency in the use and re-mobilization of phosphorus. *Plant, Cell & Environment* **30**, 1557–1565.
- Ding J, Liang P, Wu P, Zhu M, Li C, Zhu X, Gao D, Chen Y, Guo W.** 2020. Effects of waterlogging on grain yield and associated traits of historic wheat cultivars in the middle and lower reaches of the Yangtze River, China. *Field Crops Research* **246**, 107695.
- Ding L, Wang KJ, Jiang GM, Liu MZ, Niu SL, Gao LM.** 2005. Post-anthesis changes in photosynthetic traits of maize hybrids released in different years. *Field Crops Research* **93**, 108–115.
- 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* **65**, 4959–4973.
- Driever SM, Simkin AJ, Alotaibi S, Fisk SJ, Madgwick PJ, Sparks CA, Jones HD, Lawson T, Parry MAJ, Raines CA.** 2017. Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20160384.
- Ehleringer J.** 1983. Ecophysiology of *Amaranthus palmeri*, a Sonoran desert summer annual. *Oecologia* **57**, 107–112.
- Ehleringer J, Pearcy RW.** 1983. Variation in quantum yield for CO<sub>2</sub> uptake among C<sub>3</sub> and C<sub>4</sub> plants. *Plant Physiology* **73**, 555–559.
- Ermakova M, Danila FR, Furbank RT, von Caemmerer S.** 2020. On the road to C<sub>4</sub> rice: advances and perspectives. *The Plant Journal* **101**, 940–950.
- Evans J.** 1988. Acclimation by the thylakoid membranes to growth irradiance and the partitioning of nitrogen between soluble and thylakoid proteins. *Functional Plant Biology* **15**, 93.
- Evans JR.** 1989. Photosynthesis and nitrogen relationships in leaves of C<sub>3</sub> plants. *Oecologia* **78**, 9–19.
- Evans MMS.** 2007. The *indeterminate gametophyte1* gene of maize encodes a LOB Domain protein required for embryo sac and leaf development. *The Plant Cell* **19**, 46–62.
- Fan Y, Asao S, Furbank RT, Caemmerer S von, Day DA, Tcherkez G, Sage TL, Sage RF, Atkin OK.** 2021. The crucial roles of mitochondria in supporting C<sub>4</sub> photosynthesis. *New Phytologist* **233**, 1083–1096.
- Faralli M, Lawson T.** 2020. Natural genetic variation in photosynthesis: an untapped resource to increase crop yield potential? *The Plant Journal* **101**, 518–528.
- Feldman MJ, Ellsworth PZ, Fahlgren N, Gehan MA, Cousins AB, Baxter I.** 2018. Components of water use efficiency have unique genetic signatures in the model C<sub>4</sub> grass *Setaria*. *Plant Physiology* **178**, 699–715.
- Ferguson JN, Fernandes SB, Monier B, et al.** 2021. Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. *Plant Physiology* **187**, 1481–1500.
- Flint-Garcia SA, Thornsberry JM, Edward SB IV.** 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* **54**, 357–374.
- Flood PJ, Harbinson J, Aarts MGM.** 2011. Natural genetic variation in plant photosynthesis. *Trends in Plant Science* **16**, 327–335.
- Flood PJ, Theeuwes TPJM, Schneeberger K, et al.** 2020. Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *Nature Plants* **6**, 13–21.
- Forsberg SKG, Andreatta ME, Huang XY, Danku J, Salt DE, Carlborg Ö.** 2015. The multi-allelic genetic architecture of a variance-heterogeneity locus for molybdenum concentration in leaves acts as a source of unexplained additive genetic variance. *PLoS Genetics* **11**, e1005648.
- Freudenthal JA, Ankenbrand MJ, Grimm DG, Korte A.** 2019. GWAS-Flow: A GPU accelerated framework for efficient permutation based genome-wide association studies. *bioRxiv*, **783100** [Preprint].
- Fristedt R.** 2017. Chloroplast function revealed through analysis of GreenCut2 genes. *Journal of Experimental Botany* **68**, 2111–2120.
- Gabriel SB, Schaffner SF, Nguyen H, et al.** 2002. The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229.
- Gaio D, To J, Liu M, Monahan L, Anantanawat K, Darling AE.** 2019. Hackflex: Low cost Illumina sequencing library construction for high sample counts. *bioRxiv*, **779215** [Preprint].
- Gan X, Stegle O, Behr J, et al.** 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* **477**, 419–423.

- Garcia-Molina A, Leister D.** 2020. Accelerated relaxation of photoprotection impairs biomass accumulation in *Arabidopsis*. *Nature Plants* **6**, 9–12.
- Gloag RS, Ritchie RJ, Chen M, Larkum AWD, Quinnell RG.** 2007. Chromatic photoacclimation, photosynthetic electron transport and oxygen evolution in the chlorophyll *d*-containing oxyphotobacterium *Acaryochloris marina*. *Biochimica et Biophysica Acta – Bioenergetics* **1767**, 127–135.
- Gupta PK, Kulwal PL, Jaiswal V.** 2019. Association mapping in plants in the post-GWAS genomics era. *Advances in Genetics* **104**, 75–154.
- Hao N, Han D, Huang K, Du Y, Yang J, Zhang J, Wen C, Wu T.** 2020. Genome-based breeding approaches in major vegetable crops. *Theoretical and Applied Genetics* **133**, 1739–1752.
- Harbinson J, Woodward FI.** 1984. Field measurements of the gas exchange of woody plant species in simulated sunflecks. *Annals of Botany* **53**, 841–851.
- Hitchcock A, Hunter CN, Sobotka R, Komenda J, Dann M, Leister D.** 2022. Redesigning the photosynthetic light reactions to enhance photosynthesis – the PhotoRedesign consortium. *The Plant Journal* **109**, 23–34.
- 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.
- Jaganathan D, Bohra A, Thudi M, Varshney RK.** 2020. Fine mapping and gene cloning in the post-NGS era: advances and prospects. *Theoretical and Applied Genetics* **133**, 1791–1810.
- Jin J, Liu X, Wang G, Mi L, Shen Z, Chen X, Herbert SJ.** 2010. Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in Northeast China. *Field Crops Research* **115**, 116–123.
- Joseph B, Corwin JA, Züst T, Li B, Irvani M, Schaeppman-Strub G, Turnbull LA, Kliebenstein DJ.** 2013. Hierarchical nuclear and cytoplasmic genetic architectures for plant growth and defense within *Arabidopsis*. *The Plant Cell* **25**, 1929–1945.
- Joynson R, Molero G, Coombes B, Gardiner LJ, Rivera-Amado C, Piñera-Chávez FJ, Evans JR, Furbank RT, Reynolds MP, Hall A.** 2021. Uncovering candidate genes involved in photosynthetic capacity using unexplored genetic variation in Spring Wheat. *Plant Biotechnology Journal* **19**, 1537–1552.
- Jung HS, Niyogi KK.** 2009. Quantitative genetic analysis of thermal dissipation in *Arabidopsis*. *Plant Physiology* **150**, 977–986.
- Kaiser E, Morales A, Harbinson J.** 2018. Fluctuating light takes crop photosynthesis on a rollercoaster ride. *Plant Physiology* **176**, 977–989.
- Kasajima I, Ebana K, Yamamoto T, Takahara K, Yano M, Kawai-Yamada M, Uchimiya H.** 2011. Molecular distinction in genetic regulation of nonphotochemical quenching in rice. *Proceedings of the National Academy of Sciences, USA* **108**, 13835–13840.
- Keeley JE, Rundel PW.** 2003. Evolution of CAM and C<sub>4</sub> carbon-concentrating mechanisms. *International Journal of Plant Sciences* **164**, 55–77.
- Koester RP, Nohl BM, Diers BW, Ainsworth EA.** 2016. Has photosynthetic capacity increased with 80 years of soybean breeding? An examination of historical soybean cultivars. *Plant, Cell & Environment* **39**, 1058–1067.
- Koester RP, Skoneczka JA, Cary TR, Diers BW, Ainsworth EA.** 2014. Historical gains in soybean (*Glycine max* Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *Journal of Experimental Botany* **65**, 3311–3321.
- Korte A, Farlow A.** 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 1–9.
- Kromdijk J, Głowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, Long SP.** 2016. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* **354**, 857–861.
- Larcher W.** 1995. *Physiological plant ecology: ecophysiology and stress physiology of functional groups*. Berlin, Heidelberg, New York: Springer-Verlag.
- Leimu R, Fischer M.** 2008. A meta-analysis of local adaptation in plants. *PLoS ONE* **3**, e4010.
- Levine RP.** 1968. Genetic dissection of photosynthesis. *Science* **162**, 768–771.
- Li J, Tietz S, Cruz JA, Strand DD, Xu Y, Chen J, Kramer DM, Hu J.** 2019. Photometric screens identified *Arabidopsis* peroxisome proteins that impact photosynthesis under dynamic light conditions. *The Plant Journal* **97**, 460–474.
- Löffelhardt W.** 2014. The single primary endosymbiotic event. In: Löffelhardt W (ed) *Endosymbiosis*. Vienna: Springer, 39–52.
- Long SP, Zhu XG, Naidu SL, Ort DR.** 2006. Can improvement in photosynthesis increase crop yields? *Plant, Cell & Environment* **29**, 315–330.
- Lopez-Arboleda WA, Reinert S, Nordborg M, Korte A.** 2021. Global genetic heterogeneity in adaptive traits. *Molecular Biology and Evolution* **38**, 4822–4831.
- López-Calcagno PE, Brown KL, Simkin AJ, Fisk SJ, Vialet-Chabrand S, Lawson T, Raines CA.** 2020. Stimulating photosynthetic processes increases productivity and water-use efficiency in the field. *Nature Plants* **6**, 1054–1063.
- Loudet O, Saliba-Colombani V, Camilleri C, Calenge F, Gaudon V, Koprivova A, North KA, Kopriva S, Daniel-Vedele F.** 2007. Natural variation for sulfate content in *Arabidopsis thaliana* is highly controlled by APR2. *Nature Genetics* **39**, 896–900.
- Lowry DB, Logan TL, Santuari L, Hardtke CS, Richards JH, DeRose-Wilson LJ, McKay JK, Sen S, Juenger TE.** 2013. Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in *Arabidopsis*. *The Plant Cell* **25**, 3266–3279.
- Lv J, Yu K, Wei J, et al.** 2020. Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. *Nature Biotechnology* **38**, 1397–1401.
- Mackay TFC.** 2014. Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nature Reviews. Genetics* **15**, 22–33.
- Makino A, Sakuma I, Sudo E, Mae T.** 2003. Differences between maize and rice in N-use efficiency for photosynthesis and protein allocation. *Plant and Cell Physiology* **44**, 952–956.
- McAusland L, Vialet-Chabrand S, Jauregui I, et al.** 2020. Variation in key leaf photosynthetic traits across wheat wild relatives is accession dependent not species dependent. *New Phytologist* **228**, 1767–1780.
- Meuwissen THE, Hayes BJ, Goddard ME.** 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.
- Miclaus M, Balacescu O, Has I, Balacescu L, Has V, Suteu D, Neuenschwander S, Keller I, Bruggmann R.** 2016. Maize cytolines unmask key nuclear genes that are under the control of retrograde signaling pathways in plants. *Genome Biology and Evolution* **8**, 3256–3270.
- Molla KA, Sretenovic S, Bansal KC, Qi Y.** 2021. Precise plant genome editing using base editors and prime editors. *Nature Plants* **7**, 1166–1187.
- Monteith JL.** 1994. Validity of the correlation between intercepted radiation and biomass. *Agricultural and Forest Meteorology* **68**, 213–220.
- Mooney HA, Ehleringer J, Berry JA.** 1976. High photosynthetic capacity of a winter annual in Death Valley. *Science* **194**, 322–324.
- Morrison MJ, Voldeng HD, Cober ER.** 1999. Physiological changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agronomy Journal* **91**, 685–689.
- Murchie EH, Kefauver S, Araus JL, Muller O, Rascher U, Flood PJ, Lawson T.** 2018. Measuring the dynamic photosynthetic. *Annals of Botany* **122**, 207–220.
- Nelson RM, Pettersson ME, Carlborg Ö.** 2013. A century after Fisher: time for a new paradigm in quantitative genetics. *Trends in Genetics* **29**, 669–676.
- Nobel PS.** 1991. Achievable productivities of certain CAM plants: basis for high values compared with C<sub>3</sub> and C<sub>4</sub> plants. *New Phytologist* **119**, 183–205.

- Nordborg M, Weigel D.** 2008. Next-generation genetics in plants. *Nature* **456**, 720–723.
- Nunes-Nesi A, Sulpice R, Gibon Y, Fernie AR.** 2008. The enigmatic contribution of mitochondrial function in photosynthesis. *Journal of Experimental Botany* **59**, 1675–1684.
- Oakley CG, Savage L, Lotz S, Larson GR, Thomashow MF, Kramer DM, Schemske DW.** 2018. Genetic basis of photosynthetic responses to cold in two locally adapted populations of *Arabidopsis thaliana*. *Journal of Experimental Botany* **69**, 699–709.
- Ort DR, Merchant SS, Alric J, et al.** 2015. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences, USA* **112**, 8529–8536.
- Ortiz D, Hu J, Salas Fernandez MG.** 2017. Genetic architecture of photosynthesis in *Sorghum bicolor* under non-stress and cold stress conditions. *Journal of Experimental Botany* **68**, 4545–4557.
- Popper KR.** 1959. The propensity interpretation of probability. *British Journal for the Philosophy of Science* **10**, 25–42.
- Prinzenberg AE, Campos-Dominguez L, Kruijer W, Harbinson J, Aarts MGM.** 2020. Natural variation of photosynthetic efficiency in *Arabidopsis thaliana* accessions under low temperature conditions. *Plant Cell and Environment* **43**, 2000–2013.
- Prinzenberg AE, Viquez-Zamora M, Harbinson J, Lindhout P, van Heusden S.** 2018. Chlorophyll fluorescence imaging reveals genetic variation and loci for a photosynthetic trait in diploid potato. *Physiologia Plantarum* **164**, 163–175.
- Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA.** 2012. Recent patterns of crop yield growth and stagnation. *Nature Communications* **3**, 1293.
- Reich PB, Walters MB, Ellsworth DS, Uhl C.** 1994. Photosynthesis-nitrogen relations in Amazonian tree species – I. Patterns among species and communities. *Oecologia* **97**, 62–72.
- Rochaix JD.** 1997. Chloroplast reverse genetics: New insights into the function of plastid genes. *Trends in Plant Science* **2**, 419–425.
- Rochaix J-D.** 2004. Genetics of the biogenesis and dynamics of the photosynthetic machinery in eukaryotes. *The Plant Cell* **16**, 1650.
- Roux F, Mary-Huard T, Barillot E, Wenes E, Botran L, Durand S, Villoutreix R, Martin-Magniette M-L, Camilleri C, Budar F.** 2016. Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proceedings of the National Academy of Sciences, USA* **113**, 3687–92.
- Rungrat T, Almonte AA, Cheng R, Gollan PJ, Stuart T, Aro EM, Borevitz JO, Pogson B, Wilson PB.** 2019. A genome-wide association study of non-photochemical quenching in response to local seasonal climates in *Arabidopsis thaliana*. *Plant Direct* **3**, e00138.
- Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink JL, Sorrells ME.** 2015. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. *The Plant Genome* **8**, plantgenome2014.10.0074.
- Sadras VO, Lawson C, Montoro A.** 2012. Photosynthetic traits in Australian wheat varieties released between 1958 and 2007. *Field Crops Research* **134**, 19–29.
- Sasaki H, Ishii R.** 1992. Cultivar differences in leaf photosynthesis of rice bred in Japan. *Photosynthesis Research* **32**, 139–146.
- Scheller HV, Jensen PE, Haldrup A, Lunde C, Knoetzel J.** 2001. Role of subunits in eukaryotic Photosystem I. *Biochimica et Biophysica Acta – Bioenergetics* **1507**, 41–60.
- 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.
- Siebers MH, Gomez-Casanovas N, Fu P, Meacham-Hensold K, Moore CE, Bernacchi CJ.** 2021. Emerging approaches to measure photosynthesis from the leaf to the ecosystem. *Emerging Topics in Life Sciences* **5**, 261.
- Simkin AJ, López-Calcano PE, Raines CA.** 2019. Feeding the world: Improving photosynthetic efficiency for sustainable crop production. *Journal of Experimental Botany* **70**, 1119–1140.
- Soleh MA, Tanaka Y, Kim SY, Huber SC, Sakoda K, Shiraiwa T.** 2017. Identification of large variation in the photosynthetic induction response among 37 soybean [*Glycine max* (L.) Merr.] genotypes that is not correlated with steady-state photosynthetic capacity. *Photosynthesis Research* **131**, 305–315.
- Soleh MA, Tanaka Y, Nomoto Y, Iwahashi Y, Nakashima K, Fukuda Y, Long SP, Shiraiwa T.** 2016. Factors underlying genotypic differences in the induction of photosynthesis in soybean [*Glycine max* (L.) Merr.]. *Plant, Cell & Environment* **39**, 685–693.
- Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJM.** 2000. The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Molecular Cell* **6**, 791–802.
- South PF, Cavanagh AP, Liu HW, Ort DR.** 2019. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* **363**, eaat9077.
- Storey JD, Tibshirani R.** 2003. Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences, USA* **100**, 9440–9445.
- Tang Z, Hu W, Huang J, Lu X, Yang Z, Lei S, Zhang Y, Xu C.** 2014. Potential involvement of maternal cytoplasm in the regulation of flowering time via interaction with nuclear genes in maize. *Crop Science* **54**, 544–553.
- Taniyoshi K, Tanaka Y, Shiraiwa T.** 2020. Genetic variation in the photosynthetic induction response in rice (*Oryza sativa* L.). *Plant Production Science* **23**, 513–521.
- Tuinstra MR, Ejeta G, Goldsbrough PB.** 1997. Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci. *Theoretical and Applied Genetics* **95**, 1005–1011.
- Turner TL.** 2014. Fine-mapping natural alleles: quantitative complementation to the rescue. *Molecular Ecology* **23**, 2377.
- Tyagi AK, Gaur T.** 2003. Light regulation of nuclear photosynthetic genes in higher plants. *Critical Reviews in Plant Sciences* **22**, 417–452.
- van Bezouw RFHM, Keurentjes JJB, Harbinson J, Aarts MGM.** 2019. Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *The Plant Journal* **97**, 112–133.
- van den Boogaard R, Alewijnse D, Veneklaas EJ, Lambers H.** 1997. Growth and water-use efficiency of 10 *Triticum aestivum* cultivars at different water availability in relation to allocation of biomass. *Plant, Cell & Environment* **20**, 200–210.
- Van Rooijen R, Kruijer W, Boesten R, Van Eeuwijk FA, Harbinson J, Aarts MGM.** 2017. Natural variation of YELLOW SEEDLING1 affects photosynthetic acclimation of *Arabidopsis thaliana*. *Nature Communications* **8**, 1421.
- Vico G, Way DA, Hurry V, Manzoni S.** 2019. Can leaf net photosynthesis acclimate to rising and more variable temperatures? *Plant, Cell & Environment* **42**, 1913–1928.
- Wang P, Hendron RW, Kelly S.** 2017. Transcriptional control of photosynthetic capacity: conservation and divergence from *Arabidopsis* to rice. *New Phytologist* **216**, 32–45.
- Wang Q, Zhao H, Jiang J, Xu J, Xie W, Fu X, Liu C, He Y, Wang G.** 2017. Genetic architecture of natural variation in rice nonphotochemical quenching capacity revealed by genome-wide association study. *Frontiers in Plant Science* **8**, 1773.
- Watson A, Ghosh S, Williams MJ, et al.** 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* **4**, 23–29.
- Weigel D.** 2012. Natural variation in *Arabidopsis*: From molecular genetics to ecological genomics. *Plant Physiology* **158**, 2–22.
- Wójtowicz J, Gieczewska KB.** 2021. The *Arabidopsis* accessions selection is crucial: Insight from photosynthetic studies. *International Journal of Molecular Sciences* **22**, 9866.
- Wu A, Doherty A, Farquhar GD, Hammer GL.** 2017. Simulating daily field crop canopy photosynthesis: an integrated software package. *Functional Plant Biology* **45**, 362–377.

- Wu A, Hammer GL, Doherty A, von Caemmerer S, Farquhar GD.** 2019. Quantifying impacts of enhancing photosynthesis on crop yield. *Nature Plants* **5**, 380–388.
- Wullschlegel SD.** 1993. Biochemical limitations to carbon assimilation in  $C_3$  plants—a retrospective analysis of the  $A/C_i$  curves from 109 species. *Journal of Experimental Botany* **44**, 907–920.
- Xu S.** 1995. A comment on the simple regression method for interval mapping. *Genetics* **141**, 1657–1659.
- Yamori W, Hikosaka K, Way DA.** 2014. Temperature response of photosynthesis in  $C_3$ ,  $C_4$ , and CAM plants: Temperature acclimation and temperature adaptation. *Photosynthesis Research* **119**, 101–117.
- Yin X, Gu J, Dingkuhn M, Struik PC.** 2022. A model-guided holistic review of exploiting natural variation of photosynthesis traits in crop improvement. *Journal of Experimental Botany* 73, in press.
- 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.
- Zargar SM, Gupta N, Nazir M, Mahajan R, Malik FA, Sofi NR, Shikari AB, Salgotra RK.** 2017. Impact of drought on photosynthesis: Molecular perspective. *Plant Gene* **11**, 154–159.
- Zargar SM, Raatz B, Sonah H, Nazir M, Bhat JA, Dar ZA, Agrawal GK, Rakwal R.** 2015. Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *Journal of Crop Science and Biotechnology* **18**, 293–308.
- Zelitch I.** 1975. Improving the efficiency of photosynthesis. *Science* **188**, 626–633.
- Zhang WH, Kokubun M.** 2004. Historical changes in grain yield and photosynthetic rate of rice cultivars released in the 20th century in Tohoku region. *Plant Production Science* **7**, 36–44.
- Zheng TC, Zhang XK, Yin GH, Wang LN, Han YL, Chen L, Huang F, Tang JW, Xia XC, He ZH.** 2011. Genetic gains in grain yield, net photosynthesis and stomatal conductance achieved in Henan Province of China between 1981 and 2008. *Field Crops Research* **122**, 225–233.
- Zhu XG, Long SP, Ort DR.** 2010. Improving photosynthetic efficiency for greater yield. *Annual Review of Plant Biology* **61**, 235–261.