

DARWIN REVIEW

Plant biochemical genetics in the multiomics era

Saleh Alseekh^{1,2,*}, Esra Karakas¹, Feng Zhu³, Micha Wijesingha Ahchige¹ and Alisdair R. Fernie^{1,2,*}

¹ Max Planck Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany

² Center of Plant Systems Biology and Biotechnology, Plovdiv, Bulgaria

³ National R&D Center for Citrus Preservation, Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, 430070 Wuhan, China

* Correspondence: fernie@mpimp-golm.mpg.de or alseekh@mpimp-golm.mpg.de

Received 13 December 2022; Editorial decision 3 May 2023; Accepted 9 May 2023

Editor: Aleksandra Skirycz, Boyce Thompson Institute, USA

Abstract

Our understanding of plant biology has been revolutionized by modern genetics and biochemistry. However, biochemical genetics can be traced back to the foundation of Mendelian genetics; indeed, one of Mendel's milestone discoveries of seven characteristics of pea plants later came to be ascribed to a mutation in a starch branching enzyme. Here, we review both current and historical strategies for the elucidation of plant metabolic pathways and the genes that encode their component enzymes and regulators. We use this historical review to discuss a range of classical genetic phenomena including epistasis, canalization, and heterosis as viewed through the lens of contemporary high-throughput data obtained via the array of approaches currently adopted in multiomics studies.

Keywords: Biochemical genetics, canalization, epistasis, metabolomics, GWAS, QTL.

Introduction

The foundation of biochemical genetics

There has been much written about Gregor Mendel in the last 12 months due to the 200th anniversary of his birth (Eckardt *et al.*, 2022). He is often described as the founder of genetics but what is less commonly acknowledged is that he is without a doubt, somewhat serendipitously, the founder of biochemical genetics. This is because of the seven characteristics of pea plants he studied, one of which was a wrinkled phenotype and was a morphological change that was displayed due to a mutation in a starch branching enzyme—a fact that was first documented more than 30 years ago (Bhattacharyya *et al.*, 1990). Until recently, biochemical genetics was centered around those metabolites that conferred highly visual phenotypes, either with respect to pigmentation, shape, or size. That said, unlike the case

for Mendel, the majority of these studies followed the isolation of the first enzyme to be identified, diastase (currently known as amylase), in microbes the early 20th century (Needham, 1970; Kohler, 1972) and were also subsequent to what many regard as the foundation of biochemical genetics, namely the linking of enzymes to genes via the 'one gene, one enzyme' theory of (Beadle and Tatum (1941). Whilst this theory holds true (at least in part) in simple organisms such as *Neurospora crassa* for which it was developed, the extensive gene duplication that has characterized the plant kingdom (Zhang, 2003; Fernie and Tohge, 2017), means that this is not the case for plants. Although on the one hand this presents a complication in the study of plant metabolism (and especially that of specialized metabolism), on the other hand, as we will detail below, recent studies aimed at examining metabolic gene clusters in plants have benefitted

from this complexity (Zhan *et al.*, 2022). The term ‘biochemical genetics’, whilst less commonly used in plants than it was 50 years ago, refers to the combination of biochemistry and genetics. Prior to the greater availability of genome sequences over recent decades, this was largely performed by the use of molecular genetic markers and simple characteristics of gene products, for example based on their electrophoretic properties (Banuett-Bourrillon and Hague, 1979; Kawase and Sakamoto, 1984). Whilst these approaches were simple, they produced an incredible knowledge base. The marriage of genetics and biochemistry has a long history in the unravelling of plant metabolic pathways, and it has contributed to pivotal developments in phytochemistry (Somerville, 2000; Fernie, 2007), including the vital elucidation of various aspects of photosynthesis (Calvin, 1962; Kortschak *et al.*, 1965; Hatch and Slack, 1968), the discovery of transposable elements (McClintock, 1951), the establishment of the pathway of chlorophyll biosynthesis (Wolff and Price, 1957), the endosymbiont hypothesis of organelle acquisition (Gray *et al.*, 1999), and the high-level structural resolution of photosystems I and II (Zouni *et al.*, 2005; El-Mohsnawy *et al.*, 2010). In contemporary terms, biochemical genetics can be explained as the genetic diversity underpinning protein function and abundance.

Prior to the development of gene transformation and more latterly gene-editing techniques, many early studies exploited mutants in combination with either steady-state or isotope-labelled metabolite analyses (Calvin, 1962; Wheeler *et al.*, 1998), and this allowed the definition of the structures of the majority of the cardinal pathways of plant biochemistry. This approach has continued to be followed successfully in recent decades (Wheeler *et al.*, 1998; Schwender *et al.*, 2004; Tieman *et al.*, 2006; Dal Cin *et al.*, 2011). However, the dependence of metabolic studies on genetic polymorphism has also considerably diversified to include both natural variation (Kliebenstein *et al.*, 2001; Bentsink *et al.*, 2003; Schauer *et al.*, 2005; El-Lithy *et al.*, 2006) and forward- and reverse-genetic approaches (Fridman *et al.*, 2004; Sergeeva *et al.*, 2004; Schauer *et al.*, 2006; Tieman *et al.*, 2006; Zhang *et al.*, 2006; Liseč *et al.*, 2008; Rowe *et al.*, 2008; Joseph *et al.*, 2015a, 2015b). In parallel, integrated genomics approaches are beginning to yield an enhanced mechanistic understanding of biological systems, particularly when studied kinetically following an environmental perturbation or when spanning several developmental periods (Carrari *et al.*, 2006; Wang *et al.*, 2014; Yang *et al.*, 2022). In this paper, we review current and historical strategies for the elucidation of plant metabolic pathways and the genes that encode them. We additionally discuss a range of classical genetic phenomena including mode of inheritance, epistasis, canalization, and heterosis as viewed through the lens of the high-throughput data obtained via the array of approaches that are currently adopted in multiomics studies (Fig. 1). First, however, we describe the gradual shift away from using single-mutants to scanning natural variance in the wild, and discuss the difficulties and benefits of this transition.

Single-mutants to segregating populations to genome-wide association studies: an overview of population types underpinning the study of plant biochemical genetics

As was the case noted above for *N. crassa*, the earliest reports of biochemical genetics tended to focus on mutant screens targeting single genes. Indeed, much of our understanding of the biosynthesis of starch, lipids, cell walls, and flavonols has come from screens of populations mutagenized using ethyl methanesulfonate (Herlihy *et al.*, 2019). In the case of starch metabolism, these screens were based on simple iodine staining that highlights differences in amylose binding (Caspar *et al.*, 1985; Lin *et al.*, 1988). Similarly, flavonol biosynthesis was dissected on the basis of seed colour, since the lack of flavonols in the seed coat gives rise to a transparent testa phenotype (Appelhaagen *et al.*, 2014). In contrast, mutants in lipids and cell walls were largely identified by more laborious means or in developmental screens. For example, in the case of the identification of the Arabidopsis *dgd1* mutant radiolabel tracing was employed (Dörmann *et al.*, 1999), whilst mutants of cell wall biosynthesis have often been found on the basis of cell swelling phenotypes in root tissues (Doblin *et al.*, 2002). The use of stress to exacerbate phenotypes has additionally been widely employed, such as in screens for sensitivity to UV-B (Li *et al.*, 1993; Landry *et al.*, 1995; Conklin *et al.*, 1996). Since the turn of the century the availability (at least in Arabidopsis) of a T-DNA knockout collection and database (e.g. TAIR and ABRC) have greatly simplified gene functional analysis. Beyond Arabidopsis, the development of the TILLING method as an approach to accelerate the identification of mutant loci (Till *et al.*, 2003) and the development of genotyping by sequencing (Galvão *et al.*, 2012) have greatly accelerated the identification of the genetic basis of biochemical phenotypes (Fig. 1). In parallel, the advent and adoption of techniques that can simultaneously determine the levels of hundreds of metabolites (Fiehn *et al.*, 2000; Alseekh and Fernie, 2018) alongside the opportunities afforded by next-generation sequencing (Purugganan and Jackson, 2021; Gui *et al.*, 2023) have opened up the possibility of evaluating much broader genetic variance, such as that available in advanced breeding populations (Alseekh *et al.*, 2021; Sharwood *et al.*, 2022; Bulut *et al.*, 2023) and panels of natural variants (Fang and Luo, 2019). Although such populations have been reviewed extensively elsewhere (Fernie and Klee, 2011; Wijnen and Keurentjes, 2014), we will briefly describe their development here since they are highly relevant to the following sections.

Whilst the study of natural variance (at a narrow level) proceeded that of advanced breeding populations, given that it has only been possible to provide mechanistic rather than descriptive results since the recent advent of genome-wide association analysis, we will start this discussion with the use of advanced breeding populations. There are five types of population that are worthy of discussion in the context

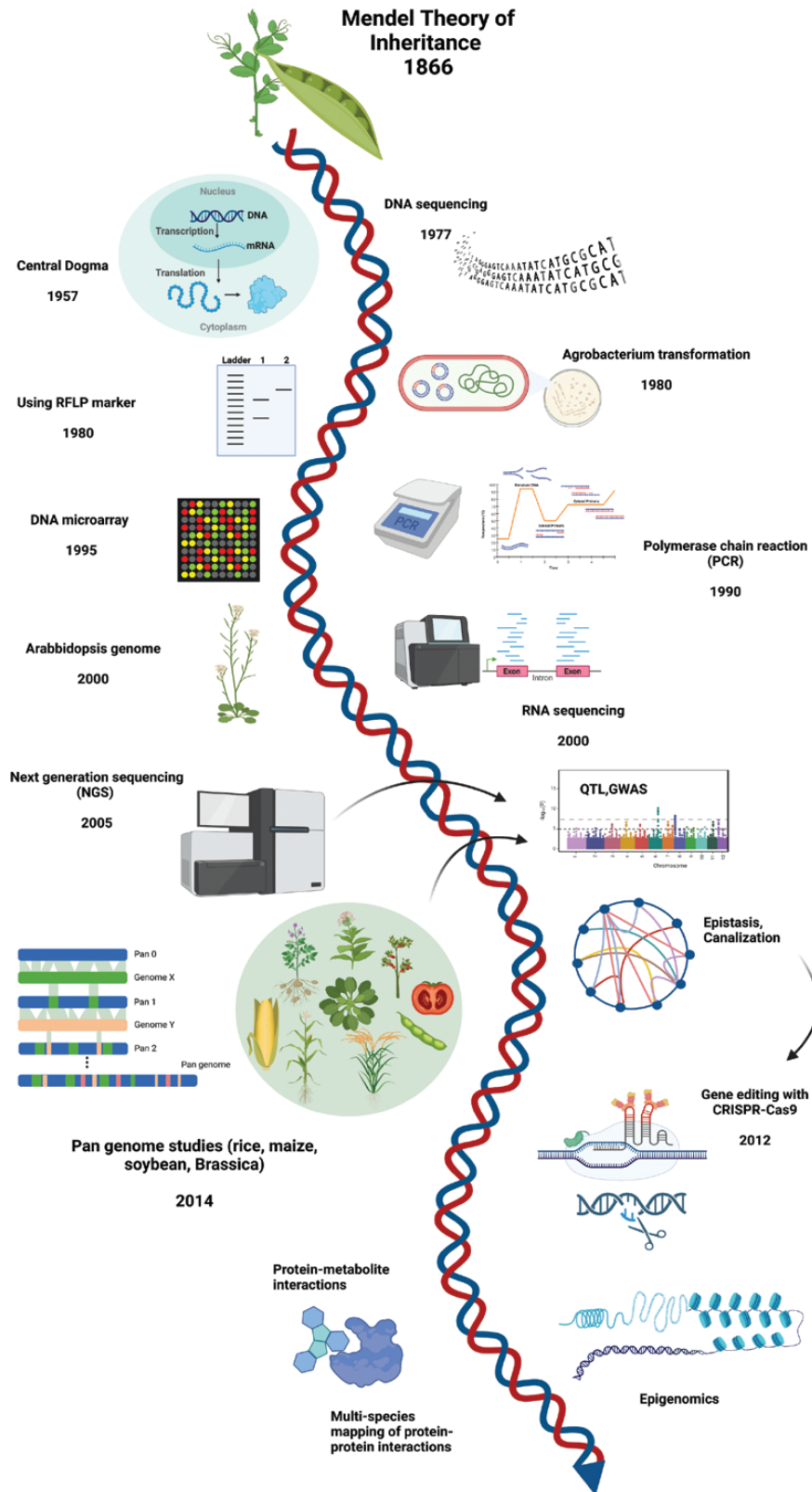


Fig. 1. Historical developments of genetics in the multiomics era (part of the figure generated by BioRender).

of biochemical genetics, namely recombinant inbred lines, introgression lines, backcrossed inbred lines, double-haploids, and heterogeneous inbred families (Bulut *et al.*, 2023). Populations can be generated by multiple rounds of selfing, as in the case for recombinant inbred lines (RILS; Fernie and Klee, 2011), or by repeated backcrossing and extensive genotyping, as in the case of introgression lines (ILs) and backcrossed inbred lines (BILs). Despite their similar routes of generation, BILs are more similar to RILs in that they have a mosaic of donor and recurrent genomes rather than a single (or at least a small number) of chromosomal segment substitutions (Ofner *et al.*, 2016; Brog *et al.*, 2019). Homozygous populations can also be induced by chromosomal doubling of haploids, such as for double-haploids (von Korff *et al.*, 2004). Heterogeneous inbred families are most frequently used in order to confirm quantitative trait loci (QTL) detected in a RIL population by taking a predecessor of a RIL that remains heterozygous for the region of interest but is otherwise homozygous and selfing it, following which the heterozygous region segregates in a Mendelian manner. Such studies thus enable comparison of the trait of interest for that specific region for both parental genotypes in an isogenic background. Following similar principles, multi-parental populations have been generated to increase the allelic variance in the resultant offspring, such as NAM and CUBIC populations (Gage *et al.*, 2020; Liu *et al.*, 2020). Again, we will detail the use of these populations in biochemical genetics in the following sections.

A complementary approach to that offered by breeding populations is that of directly assessing the broad-range natural variation of traits and associating this with differences in gene sequences via genome-wide association studies (GWAS; Alseekh *et al.*, 2021). The aim of GWAS is strikingly simple, namely to detect the association between allele or genotype frequency and trait status. The first step is to select an appropriate study population, considering both the size of the population and the amount of genetic and phenotypic variance that it possesses. Given that we have recently reviewed this extensively (Alseekh *et al.*, 2021), we will not detail it here; suffice to say that suitable collections now exist for a range of model species, common and even rare crop species, and also a range of non-cultivated species. These collections have made GWAS considerably easier to conduct, even compared with just a few years ago. Given that approaches based on breeding populations and GWAS are highly complementary, we will discuss them together in the following sections.

Epistasis

In population genetics, the total genetic variance is divided into orthogonal components attributable to additive, dominance (intralocus interactions), and epistatic variance (interlocus interactions) that depend on allele frequencies

(Vitezica *et al.*, 2017). Epistasis can be defined as deviation from the additivity of the combined effect of multiple variants (Falconer, 1989). At the beginning of the 20th century, the term epistasis was first brought into the scientific community by William Bateson as meaning a non-linear interaction between two or more segregating loci with different alleles across genetic backgrounds (Causse *et al.*, 2007; Liu and Yan, 2019). Most observed genetic variance for quantitative traits is additive; however, the great influence of epistatic gene action at many loci for quantitative traits should not be ignored (Liu and Yan, 2019). In the absence of epistasis, the estimates of additive and dominance effects at a given locus are the same regardless of the genotype of another locus; however, in the presence of epistasis, a substantial contribution is provided to each of these variance components.

Multiple genetic factors and their interactions (epistasis) are the key regulators for elucidating the genetic basis of phenotypic variances, and are also implicated in gene-environment interactions. The great influence of epistasis and complicated forms of environmental effects have been neglected in studies detecting phenotypic variance. The elucidation of statistical associations between millions of genetic variants and phenotypes has become possible with the use of single-nucleotide polymorphisms (SNPs), and these have also contributed to the investigation of genotype \times environment (G \times E) interactions (Liu and Yan, 2019). It has been suggested that G \times E interactions might have impacts in complex ways within populations (Scheiner, 1993; Juenger *et al.*, 2005).

In all studies of quantitative variation, statistical power is a central issue. Theoretical models and experimental results indicate that population bottlenecks and subdivisions expose hidden additive genetic variance selection due to epistasis. Many distinct methods are available to analyse and visualize one-, two-, or three-way epistatic interactions (Martínez *et al.*, 2018). Due to the statistical and computational complexities, most analyses are constrained to examine only pairwise interactions in order to detect epistasis (Weinreich *et al.*, 2013). The approaches mainly focus on the selection of SNPs for interactions, based on existing biological knowledge or statistical features (L. Li *et al.*, 2022). When considering biochemical pathways and gene networks, the presence of epistasis may be inevitable; however, evidence of epistasis in terms of quantitative character variation is surprisingly weak (Mackay, 2014; Labroo *et al.*, 2021). In quantitative genetics, the standard non-epistatic model is often considered to be a linear approximation of the complicated mapping from genotype to phenotype. Therefore, it is now considered normal to characterize gene interactions with non-linear functions when predicting phenotype from genetic metrics.

Over recent decades, the contribution of epistasis to quantitative trait phenotypes has been apparent in many studies, especially in GWAS and QTL analysis (Yi and Xu, 2002; Manicacci *et al.*, 2009; Würschum *et al.*, 2011; Wen *et al.*, 2015; Zhang *et al.*, 2015). GWAS have been mainly conducted based on the

assessment of statistical significance in the dissection of G×E interactions (Liu and Yan, 2019). Many studies have successfully mapped QTL; however, the variations that are caused by the tens to hundreds of genes underlying genomic regions remain unidentified (Allen Orr, 2001; Maloof, 2003; Koornneef *et al.*, 2004). The increasing use of QTL mapping has been cited as the key contributor for elucidating epistasis in quantitative trait variation (Erickson *et al.*, 2004).

Dominance and epistasis play key roles in the determination of complex traits of interest. The difficulties in detecting epistasis cannot be ignored, and therefore experimental set-ups capable of full characterization are necessary to test these higher-order interactions. An ideal experimental set-up would be the generation of isogenic lines harbouring distinct allelic combinations underlying the set of genes of interest. F₂ populations or recombinant inbred populations have been used to handle epistatic mapping (Kerwin *et al.*, 2017). A strong epistatic interaction can be detected for some metabolites in kernels in maize RIL populations (Wen *et al.*, 2015). Shang *et al.* (2016) have shown that epistasis contributes to yield heterosis in RIL and backcross populations of cotton. GWAS and epistasis studies performed using 214 soybean germplasm accessions have shown that additive and epistatic variance can explain almost half of the phenotypic variance for sudden death syndrome resistance (Zhang *et al.*, 2015).

The great influence of epistasis has been demonstrated in a wide range of crop species (Schnable and Springer, 2013; Labroo *et al.*, 2021; Monforte, 2021). For instance, in common bean significant epistasis is observed for seed yield, the number of seeds per plant, and the number of pods per plant (Moreto *et al.*, 2012). In wheat, epistatic analysis using RIL mapping populations has detected one pair of epistatic QTL for the first internode component index and three pairs for the third internode component index, thereby providing information about plant height components and associated increases in yield (Qin *et al.*, 2022). Another study of common bean has indicated the role of epistasis in the genetic control of traits associated with yield in inter-gene-pool crosses (Johnson and Gepts, 2002), and in rice it has been shown that epistasis regulates plant height on the genetic basis of midparent heterosis (Shen *et al.*, 2014). Epistatic QTL interactions including synergistic interactions have been identified in wheat populations that shed light on the inheritance of shattering resistance (Bokore *et al.*, 2022). Metabolite profiling using a maize backcross population detected ~15% epistatic interactions for primary metabolites that have impacts on maize quality, and hence could be utilized for improvement (K. Li *et al.*, 2019). Away from crop species, Kerwin *et al.*, (2017) conducted a detailed analysis that identified both additive and epistatic interactions for the polymorphic genes that control aliphatic glucosinolate in *Arabidopsis* that had not previously been assessed for potential interactions with the environment. In addition, crossing of C24 with Col-0 has indicated that dominance

and epistatic interactions have an important role in biomass-related traits (Kusterer *et al.*, 2007).

Canalization

Much less well understood than the genetics that contribute to the level of a given quantitative trait are those that are responsible for the robustness of the trait between or within environments (Fig. 2). An early principle for the concept of robustness in a developmental context was described by Waddington (1942). Waddington noticed that developmental processes were generally canalized so as to bring about a single clearly defined final stage, regardless of minor variations in the conditions under which the processes occurred. Based on this, Waddington suggested that there must be some capacity of the genotype to buffer the phenotype against these minor variations in genotype and environment. In that sense, a distinction can be made between genetic and environmental canalization. Perhaps the best-studied developmental process that shows considerable canalization against different sources of variation is the patterning of the vulva plate of *Caenorhabditis elegans* (Félix and Barkoulas, 2012). In both *Drosophila* and in *Arabidopsis*, the chaperone HSP90 has been shown to be a capacitor of phenotypic variation (Rutherford and Lindquist, 1998; Queitsch *et al.*, 2002), providing another example of developmental canalization. Genes such as HSP90 are often considered to be so-called gene network hubs, where the exponential distribution of connectivity is associated with robustness (Lachowiec *et al.*, 2016).

Quantitative traits seem to have received less attention than developmental ones in terms of canalization. When considering quantitative traits of different genotypes in different environments it becomes clear that there are different forms of canalization that can be considered, and it seems reasonable and necessary to also use a terminology that works in a quantitative sense. Using the reaction norm, that is how the trait of a genotype changes

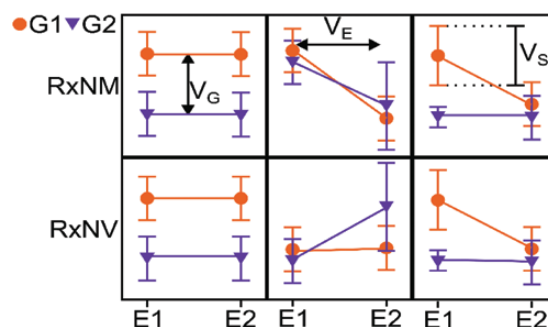


Fig. 2. Concepts of canalization and variance. Schematic plots are shown for the mean (\pm SD) of theoretical traits of two different genotypes, G1 and G2, in two different environments, E1 and E2. The top row shows plots for the reaction norm of the mean (RxNM) whilst the bottom row shows plots for the corresponding variance (RxNV). The sources of variation are either due to genetic (V_G), environmental (V_E), or stochastic (V_S) factors.

between environments, it is possible to distinguish the reaction norm of the mean ($R \times NM$) and the reaction norm of the variance ($R \times NV$) of a trait (Dworkin, 2005). While the former considers the change of mean values of genotypes between environments, the latter considers the change of variance between individuals between environments. Accordingly, the sources of phenotypic variation can be either genetic, environmental, or stochastic factors (Fig. 2; Laitinen and Nikoloski, 2019).

While there are examples with respect to canalization of quantitatively characterized traits ranging from single-celled organisms to higher plants (Alon *et al.*, 1999; Lehner, 2010; Fisher *et al.*, 2017; Lachowicz *et al.*, 2018), the application to metabolism is relatively recent and more limited. A flux balance analysis in *E. coli* identified six capacitor reactions that have a high influence on genetic canalization of metabolism (Ho and Zhang, 2016). In Arabidopsis, the *ELF3* gene shows an effect on the canalization of both the circadian clock as well as glucosinolate levels (Jimenez-Gomez *et al.*, 2011). By using transformed data to assess the reaction norm of the variance, several canalized metabolic QTL (cmQTL) responsible for genotype \times environment effects on the variance of various tomato fruit metabolites have been detected (Alseekh *et al.*, 2017). Interestingly the cmQTL only sparsely overlapped with the QTL for the levels of the metabolites, suggesting different loci were related to level and variance. Further validation of candidate genes supports the idea that genes that affect (e.g.) the cross-environment canalization do not necessarily also affect the inter-individual variance. The observation that loci responsible for variation in a trait are (at least partly) distinct and also fewer in number than loci responsible for the level of the trait is supported by other research (Hall *et al.*, 2007; Joseph *et al.*, 2015a; Li *et al.*, 2016; Kusmec *et al.*, 2017). Both the distinctiveness and the reduced number of loci could point to there being a few regulatory genes that simultaneously control several traits (Alseekh *et al.*, 2017). Depending on the trait, we can consider many different combinations of effects of loci on either the level or variance of that trait, or both (Fig. 2, columns 1, 2, and 3, respectively).

In this regard, datasets from large mapping studies that have previously been used simply to study the level of a given trait could be reused to study the inter-individual or cross-environment variation. While this has been done in some cases (Schauer *et al.*, 2006; Wentzell *et al.*, 2007; Schauer *et al.*, 2008; Jimenez-Gomez *et al.*, 2011; Alseekh *et al.*, 2017), this resource of legacy data is still underutilized. Together with new data generated through the high-throughput platforms of the multiomics era, it can be used to study canalization and variation more comprehensively (Wijesingha Ahchige *et al.*, 2023).

Organellar inheritance of biochemical traits

The organellar inheritance of biochemical genetic traits (also known as cytoplasmic genetic variation) is a further type of

epistasis that we consider separately here given that it has long been studied in its own right. Organellar genomic variation can be linked to dramatic phenotypic alterations in both mammals (Johns *et al.*, 1992; Taylor and Turnbull, 2005; Schon *et al.*, 2012) and plants (Schnable and Wise, 1998; Roubertoux *et al.*, 2003). The use of structured populations in yeast and animals have variously shown that cytoplasmic variance can influence fitness, cognition, and biomass as well as altering fitness and creating hybrid isolation (Dimitrov *et al.*, 2009; Wolf, 2009; Willett, 2012). Similarly, in plants mitochondrial genetic variation is linked to important quantitative phenotypes including cytoplasmic male sterility (Hanson, 1991; Schnable and Wise, 1998), whilst plant breeding efforts commonly employ diallele crosses to assess the presence of maternal effects on phenotypes such as height in maize (Tang *et al.*, 2013). Moreover, the presence of cytoplasmic and nuclear genome interactions has been found to influence a range of agronomic traits (Tao *et al.*, 2004); however, the genes involved are yet to be identified. Such variation is uncovered in experiments in which variance in organellar genomes is assessed alongside that of the nuclear genome. Due to their differing modes of inheritance, in practice this is largely achieved via varying the organellar genome within a population in which the nuclear genome has undergone considerable recombination. This is most simply achieved by the analysis of reciprocal F_2 populations; however, historically in plants such analyses have been taken to imply that cytoplasmic effects on phenotypic variation are quite small (Singh, 1966; Crane and Nyquist, 1967; Eenink and Garretsen, 1980; Bhatt *et al.*, 1983; Rehal *et al.*, 2022). However, in contrast to previous estimates of small effects, genomic sequencing within Arabidopsis has shown the presence of considerable genetic polymorphism in both the plastidic and mitochondrial genomes, suggesting the potential for broad phenotypic consequences (Moison *et al.*, 2010). In an attempt to quantify the importance of cytoplasmic variance, Joseph *et al.* (2013) utilized metabolomics to investigate how genetic variation in the cytoplasmic and nuclear genomes interacts to control metabolome variation in a reciprocal Arabidopsis Kas \times Tsu RIL population (Juenger *et al.*, 2006; McKay *et al.*, 2008). They found that variation in the organellar genome contributed to variation in the levels of more than 80% of the metabolites studied. Organellar genes also helped to regulate the effect of nuclear genes. This combination of direct and indirect influences helps to explain how a small number of organellar genes can have a disproportionately large effect on phenotype. However, despite the considerable insights provided by this study, it is our contention that a considerable deficit remains in our overall knowledge of the importance of these interactions, and it will be highly important to boost our understanding of their importance in crop species as well as to employ recently developed approaches (Alseekh *et al.*, 2021) to ascertain whether differences in their influence have arisen during the processes of domestication and crop improvement.

Use of populations in metabolite and gene annotation

Plant metabolic studies have traditionally focused on the role and regulation of the enzymes that catalyse key reactions within specific pathways (Ferne and Tohge, 2017). Although metabolic regulation is still being addressed via reverse-genetics approaches such as transgenesis, and via novel molecular techniques such as genome editing, within recent years broad natural variance in the form of linkage mapping populations and GWAS have been shown to be effective tools by which to deepen our understanding of plant metabolism (Alseikh *et al.*, 2021). As mentioned above, there has been a long history of utilizing genetics to understand plant biochemistry, as demonstrated by Mendel's early studies on peas (Mendel, 1865; White, 1917). In plant metabolism, a further seminal study was the discovery of transposable elements in maize kernels that regulate the flavonoid biosynthesis genes (McClintock, 1950). The genetic regulation of plant flower color, which is mainly an attribute of the abundance of secondary metabolites such as anthocyanins and flavonoids, has been investigated over the last two centuries (Hooker, 1837; Sink, 1973).

Sax (1923) was the first to describe QTL mapping in studying the seed size and color of common bean. Modern QTL mapping, which is defined in segregating populations, is based on the genotyping of progeny derived from a cross between distinct genotypes and the association of molecular markers with the phenotype of interest in the resultant offspring (Broman, 2001). It was facilitated by development of comprehensive DNA markers (Kumar, 1999; Phillips and Vasil, 2001) and more recently by the release of many plant genomes (Kress *et al.*, 2022). In terms of plant metabolism, early use of mapping populations focused mainly on important and easy-to-score metabolic traits such as carotenoid content in tomato, protein and oil content in maize, and starch content in potato and rice (Moose *et al.*, 2004; Ferne *et al.*, 2006). The initial QTL mapping studies using high-throughput mass spectrometry (MS) techniques and assessing the levels of multiple metabolic traits was first made in 2006. These studies looked at the primary metabolism of tomato (Schauer *et al.*, 2006) and the contents of secondary metabolites in Arabidopsis (Keurentjes *et al.*, 2006) to evaluate the natural variations in metabolism present in biparental segregating populations. They were followed up by further mapping studies in the same laboratories that addressed many aspects of the genetics of metabolism, including comparative analyses of population types and the evaluation of heterosis, heritability, and the environmental plasticity of the plant metabolome (Lisek *et al.*, 2008; Rowe *et al.*, 2008; Joseph *et al.*, 2015a, 2015b). This approach has been successfully applied to various types of segregating populations across a wide range of other important crop species, including maize (Wen *et al.*, 2015; K. Li *et al.*, 2019), rice (Matsuda *et al.*, 2012; Chen *et al.*, 2018), wheat (Hill *et al.*, 2013), barley (Zeng

et al., 2020), pepper (Wahyuni *et al.*, 2014), eggplant (Sulli *et al.*, 2021), and potato (Carreno-Quintero *et al.*, 2012). Examples such as these highlight the power of interspecific breeding populations for understanding the genetic bases of plant metabolism where large numbers of genes have been cloned and characterized. Although considerable biological insights have been obtained from using linkage mapping, the adoption of GWAS has allowed for the greatest breakthroughs given the advent of next-generation sequencing (Nguyen *et al.*, 2019). This approach relies on testing genetic variants across the genomes of many individuals of a population to identify genotype–phenotype associations (Uffelmann *et al.*, 2021). Similar to the linkage-mapping approach, GWAS has been successfully combined with metabolomics to assess the genetic bases of natural variance in plant metabolomes. This has been extensively reviewed elsewhere (Luo, 2015; Fang and Luo, 2019; Alseikh *et al.*, 2021), and here we will just highlight some examples using this approach that have identified effects of genetic variants on metabolic diversity across natural populations. GWAS in plants was initially applied to Arabidopsis (Chan *et al.*, 2010, 2011; Angelovici *et al.*, 2013; Verslues *et al.*, 2014; Strauch *et al.*, 2015) and then successfully extended to several crop species, covering primary and quality-related metabolites such as in tomato (Sauvage *et al.*, 2014; Tieman *et al.*, 2017; Ye *et al.*, 2017), wheat (Rathan *et al.*, 2022), jujube (Hou *et al.*, 2020), cassava (Rabbi *et al.*, 2022), peach (Y. Li *et al.*, 2019), and apple (Liao *et al.*, 2021). Similar approaches have been applied for specialized secondary metabolites in rice (Dong *et al.*, 2015; Matsuda *et al.*, 2015; Yamamura *et al.*, 2015; Chen *et al.*, 2016; Zhu *et al.*, 2018), maize (Wen *et al.*, 2014; Zhou *et al.*, 2019), wheat (Chen *et al.*, 2020), cucumber (Zhou *et al.*, 2016), and lettuce (Zhang *et al.*, 2020). Whilst many genes involved in plant metabolism pathways have been identified and cloned through large-scale GWAS, there are some examples of smaller-scale studies achieving good results. For example, the pathways of saiginols (phenylacylated flavonols) (Tohge *et al.*, 2016) and dolichols in Arabidopsis (Gawarecka *et al.*, 2022), lignin precursors in Arabidopsis, maize, and poplar (Vanholme *et al.*, 2019), and steroidal glycoalkaloids in solanaceous species (Schwahn *et al.*, 2014) have all been refined by comparisons of only a handful up to around 100 genotypes. In principal, mapping of an unknown compound to a gene gives hints as to the structure of that compound, especially in cases where the gene encodes an enzyme involved in the biosynthesis pathway. Indeed, application of cross-over theory can be employed to check for corresponding changes in the substrate(s) or product(s) of the reaction catalysed as an additional means of validation. Even if the associated gene is a regulatory rather than a structural one, the vast array of available co-expression datasets (Mutwil *et al.*, 2011) alongside the massive recent increases in our understanding of the functions of transcription factors (see for example Tang *et al.*, 2021) means that mapping these differences to the respective metabolic pathways is becoming easier.

Uncovering linkages from metabolites to morphology

As noted above, the linkage between morphological and metabolic phenotypes dates back to the foundation of genetics, even if Mendel himself was not aware of this. Much of the early study of biochemical genetics directly addressed these links since the majority of early genetic screens were based on morphological or other visual phenotypes. Following the development of genetic transformation protocols, however, much information was gleaned by the overexpression, down-regulation, and systematic knock out of genes (Stitt and Sonnewald, 1995; Alonso and Ecker, 2006). Whilst knockout mutants are random in nature, modern techniques now allow precisely targeted studies with genes essentially being knocked out to order. Such studies have now provided (almost) saturation coverage of the major of plant primary metabolic pathways, including the Calvin–Benson, tricarboxylic acid (TCA), and photorespiratory C₂ cycles, as well as gluconeogenesis, the sucrose-to-starch transition, and amino acid metabolism (Galili *et al.*, 2016; Fernie and Bauwe, 2020; Fernie *et al.*, 2020; Walker *et al.*, 2021; Rosado-Souza *et al.*, 2022). Since this body of work has been comprehensively reviewed elsewhere, we will not cover it in detail here; suffice to say that it has provided a vast number of targets for breeding and that many of the effects of the genetic interventions have now been verified as resulting in improved performance in the field as well as in the laboratory. In keeping with the theme of this historical review, we will instead focus the rest of this section on a classical theme of genetics, namely heterosis. In their review on this phenomenon 16 years ago, Lippman and Zamir (2007) described it a plant biological mystery that has endured since Charles Darwin famously described how hybrids display superior growth and fertility to their parents (Darwin, 1876). Such hybrid vigor was rediscovered in maize breeding some 30 years later (East, 1908; Shull, 1908), and it was subsequently demonstrated to occur in many other crop species. Indeed, it is now widely used in agriculture with in excess of 65% of worldwide maize, sorghum, and sunflower production being hybrid-based. Yield gains of 15–50% cited by Lippman and Zamir (2007) have provided a great incentive to understand the underlying mechanisms, and *yet although* we have made great inroads into understanding the genetic basis of yield heterosis, most notably in tomato (Lippman and Tanksley, 2001; Krieger *et al.*, 2010; MacAlister *et al.*, 2012; Park *et al.*, 2012; Jiang *et al.*, 2013; Soyk *et al.*, 2017; Torgeman and Zamir, 2023) and maize (Lai *et al.*, 2010; Yang *et al.*, 2017; Birdseye *et al.*, 2021; C. Li *et al.*, 2022), a detailed understanding of the molecular mechanisms and how they are integrated with metabolism remains lacking. There are excellent reviews available concerning the genetics underpinning heterosis (e.g. Lippman and Zamir, 2007), and so we will focus here on studies evaluating metabolic events underpinning growth and yield that have been elucidated via examination of either restricted or broad (natural) genetic variance. Our first example is provided by the

assessment of what are now very well characterized *Solanum penellii* introgression lines that contain individual chromosomal segmental substitutions of the donor species in the background of the cultivated tomato *S. lycopersicum* (Alseekh *et al.*, 2013). In their original approach, Schauer and co-workers utilized a then-novel metabolic profiling protocol to evaluate the contents of primary metabolites and uncovered a large number of QTL, and they also demonstrated clear links (mainly negative ones) between metabolite levels and traits associated with yield (Roessner *et al.*, 2001). In a follow-up study, Schauer *et al.* (2008) found that these links were not present when the introgression lines were in the heterozygous state, suggesting that metabolite contents could be enhanced following this approach in the absence of a yield penalty. At around the same time, the group of Thomas Altmann used the same approach to characterize metabolite levels and growth in Arabidopsis Col-0 × C24 introgression lines and using statistical methods they were able to detect a metabolic signature for growth (Meyer *et al.*, 2007). In follow-up studies in collaboration with the group of Albrecht Melchinger, they were able to apply the same approach to the growth of maize, including the use of metabolomic prediction of growth phenotypes (Riedelsheimer *et al.*, 2012). Importantly, further studies demonstrated the utility of this approach in showing that metabolite profiling of young leaves could predict biomass at harvest (Vergara-Diaz *et al.*, 2020), and similar findings have also been published for wheat and canola (Knoch *et al.*, 2021), and also with regard to predicting responses to stress (Villate *et al.*, 2021). Furthermore, this approach is not only useful for predicting yield, as recent work has illustrated its potential use in predicting properties related to fruit flavour (Colantonio *et al.*, 2022). The growing trend of characterizing large populations at the multiomics level offers yet further opportunities for us to obtain more complete insights into both the mysteries of heterosis *per se* and the linkages between metabolism and growth and development in general.

The metabolic basis of plant immunity

During their lifespan, plants are threatened by various different biotic challenges, including those posed by bacteria, fungi, and insects. To overcome these challenges, plants have evolved two layers of immunity response to pathogens, namely pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI), which require the function of different metabolites (Zeier, 2013; Piasecka *et al.*, 2015; Yuan *et al.*, 2021). The waxes and lignin of plant surfaces represent the first structural barriers for pathogens, and after passing these physical barriers, PAMPs of the microbes, which include bacterial proteins and endotoxins and the fungal cell-wall component chitin, are recognized by specific pattern-recognition receptors (PRRs), thereby activating various PTI defense responses (DeFalco and Zipfel, 2021).

Among the PAMPs, chitin (an insoluble polymer of β -1,4-linked *N*-acetylglucosamine) has long been well known and hence been the subject of a large amount of study. Chitinases secreted by host plants can hydrolyse the insoluble polymer to chitin oligosaccharides, which can then be recognised by plasma membrane-localized lysin-motif (LysM) proteins, and this in turn induces the accumulation of antimicrobial metabolites such as reactive oxygen species (ROS) and salicylic acid (SA) (Gong *et al.*, 2020). In an evolutionary arms race, pathogens developed specialized effector proteins to suppress PTI, in response to which plants evolved R-proteins to recognize the active effectors, thereby inducing ETI responses and leading to the activation of mitogen-activated protein kinase signalling pathways and the promotion of more SA accumulation (Tsuda *et al.*, 2013; Shinya *et al.*, 2015). In Arabidopsis, SA can induce the folding and docking of the SA-binding domain to the ankyrin repeats of NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1 (NPR1), which affects its activity as a transcriptional cofactor and activates the expression of NPR1-dependent systemic acquired resistance genes (Kumar *et al.*, 2022). In addition, another immunity-related hormone, jasmonic acid (JA) induces a different type of systemic resistance called induced systemic resistance (ISR), and studies have indicated that SA and JA immune signaling can interact with each other to balance plant responses against pathogens (Ton *et al.*, 2002; Hou and Tsuda, 2022).

In addition to functioning in pathogen recognition, immunity signalling, and response regulation, some primary metabolites can participate in the immunity process by acting as precursors of key immunity metabolites, and some secondary metabolites can act directly as antimicrobial metabolites (Hartmann and Zeier, 2018; Huang *et al.*, 2021). For example, the important amino acid lysine can be catalysed by AGD2-like defense response protein 1 and SAR-deficient 4 to synthesize pipercolic acid, which is then transferred to a cyclic, non-protein amino acid, *N*-hydroxypipercolic acid (NHP), catalysed by flavin-dependent-monooxygenase 1 (Návarová *et al.*, 2012; Hartmann *et al.*, 2018). Recently, NHP has been demonstrated to induce the SA-independent signal pathway of plant innate immunity and systemic acquired resistance (Holmes *et al.*, 2021; Mohnike *et al.*, 2021). Putrescine, the most abundant polyamine, is a product of ornithine and arginine catabolism and it can induce callose deposition and the expression of several PTI marker genes through the hydrogen peroxide and NADPH oxidase (RbohD and RbohF) signaling pathway (Liu *et al.*, 2019; Gerlin *et al.*, 2021). Another polyamine, spermine, can inhibit the ROS burst and the associated sharp increase in the cytosolic Ca^{2+} concentration, and it acts as the repressor of the earliest signaling events of PTI and downstream transcriptional and metabolic immunity responses (Zhang *et al.*, 2023). Plant secondary metabolites have been demonstrated to possess high bioactivity that efficiently controls both herbivores and pathogens (Zaynab *et al.*, 2018). Based on an integrative analysis of the genetic screening of a 26-parent recombinant

inbred line population with unbiased transcriptomic and MS-based metabolomic analyses, the novel compound scaffoylputrescine-5-(*Z*)-3-hexenal has been identified as playing a powerful role in non-host resistance against *Empoasca* leafhoppers (Bai *et al.*, 2022). In recent decades, many studies have reported that phenylpropanoid metabolism plays important roles in resistance to both abiotic stress and pathogen attack (Tohge *et al.*, 2016; Huang *et al.*, 2021). A combination of transcriptomics, metabolomics, and sensitivity assays of 580 tomato lines to *Botrytis cinerea* has shown that the predominant tomato flavonoid naringenin chalcone together with two flavonoid glycosides (3',5'-di-*C*-glucosylphloretin and phloretin-tri-hexose) are representative fungal resistance-related metabolites (Szymański *et al.*, 2020).

In recent decades, climate change has not only caused abnormal rainfall patterns and temperatures that induce drought and heat stress on plants, it has also changed the habits of temperature-dependent pathogens and increased the risk of infection of plants (Chaloner *et al.*, 2021). It is therefore necessary to enhance efforts to comprehensively elucidate the metabolomic basis of plant immunity. We envisage that such efforts will be accelerated by the development of efficient treatments to overcome plant microbes and to supply novel metabolic engineering genes and pathways to improve plant biotic stress tolerance. Before this is possible, however, a considerably higher number of multiomics studies of the effects of a given pathogen on a given crop need to be carried out. The study by Liu *et al.* (2018) of the role of OsWRKY67 in regulating responses to rice leaf panicle blast and bacterial blight diseases represents an interesting example of such a study, but we will need to examine far more pathogen-host responses in order to gather information regarding the commonalities and specificities of defense responses. As for understanding metabolite accumulation *per se*, utilizing broad natural variance as well as advanced breeding lines will probably represent invaluable tools to address this question in the coming decades.

The next 100 years

In the sections above, we have documented how the answers to many classical questions in genetic have either influenced, or been influenced by, biochemical factors. One notable omission from our review is the subject of the impact of epigenetics and epigenomics on metabolism and vice versa. There were two reasons for this: firstly, two recent reviews by Samo *et al.* (2021) and Hayashi *et al.* (2023) have covered this subject, the latter in a highly comprehensive manner, and secondly, we agree with the contention of (Weigel, 2016) that the majority of the influence is retained in the genome *per se*. This fact notwithstanding, the study of epigenetic and epigenomic control of biochemical traits is clearly an area in which we can anticipate considerable breakthroughs in the future. Another area of great promise is the prospect of gaining a far more comprehensive

understanding of gene–environment interactions that encompasses both canalization and plasticity, with the genetic populations and analytical tools required to address these topics now becoming increasingly available and well tested (Alseekh *et al.*, 2017; N. Liu *et al.*, 2021), as is the mathematical framework by which to assess them (Alseekh *et al.*, 2017; Laitinen and Nikoloski, 2019). The adaptation of such approaches should ultimately allow the development of truly sustainable agriculture (Fernie and Sonnewald, 2021; J. Liu *et al.*, 2021), which is arguably the greatest challenge for humanity of the 21st century.

Author contributions

All the authors jointly wrote the paper; EK and MWA prepared the figures.

Conflict of interest

The authors have no conflicts to declare in relation to this work.

Funding

SA and ARF acknowledge funding by the PlantaSYST project by the European Union's Horizon 2020 Research and Innovation Programme (SGA-CSA nos 664621 and 739582 under FPA no. 664620). SA acknowledges funding by the NatGenCrop project: HORIZON-WIDERA-2022-TALENTS-01, No. 101087091.

References

- Allen Orr H. 2001. The genetics of species differences. *Trends in Ecology and Evolution* **16**, 343–350.
- Alon U, Surette MG, Barkai N, Leibler S. 1999. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171.
- Alonso JM, Ecker JR. 2006. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nature Reviews Genetics* **7**, 524–536.
- Alseekh S, Fernie AR. 2018. Metabolomics 20 years on: what have we learned and what hurdles remain? *The Plant Journal* **94**, 933–942.
- Alseekh S, Kostova D, Bulut M, Fernie AR. 2021. Genome-wide association studies: assessing trait characteristics in model and crop plants. *Cellular and Molecular Life Sciences* **78**, 5743–5754.
- Alseekh S, Ofner I, Pleban T, Tripodi P, Di Dato F, Cammareri M, Mohammad A, Grandillo S, Fernie AR, Zamir D. 2013. Resolution by recombination: breaking up *Solanum pennellii* introgressions. *Trends in Plant Science* **18**, 536–538.
- Alseekh S, Scossa F, Wen W, Luo J, Yan J, Beleggia R, Klee HJ, Huang S, Papa R, Fernie AR. 2021. Domestication of crop metabolomes: desired and unintended consequences. *Trends in Plant Science* **26**, 650–661.
- Alseekh S, Tong H, Scossa F, Brotman Y, Vigroux F, Tohge T, Ofner I, Zamir D, Nikoloski Z, Fernie AR. 2017. Canalization of tomato fruit metabolism. *The Plant Cell* **29**, 2753–2765.
- Angelovici R, Lipka AE, Deason N, Gonzalez-Jorge S, Lin H, Cepela J, Buell R, Gore MA, Dellapenna D. 2013. Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. *The Plant Cell* **25**, 4827–4843.
- Appelhagen I, Thiedig K, Nordholt N, Schmidt N, Huep G, Sagasser M, Weisshaar B. 2014. Update on transparent *testa* mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. *Planta* **240**, 955–970.
- Bai Y, Yang C, Halitschke R, Paetz C, Kessler D, Burkard K, Gaquerel E, Baldwin IT, Li D. 2022. Natural history-guided omics reveals plant defensive chemistry against leafhopper pests. *Science* **375**, eabm2948.
- Banuett-Bourrillon F, Hague DR. 1979. Genetic analysis of alcohol dehydrogenase isozymes in pearl millet (*Pennisetum typhoides*). *Biochemical Genetics* **17**, 537–552.
- Beadle GW, Tatum EL. 1941. Genetic control of biochemical reactions in neurospora. *Proceedings of the National Academy of Sciences, USA* **27**, 499–506.
- Bentsink L, Yuan K, Koornneef M, Vreugdenhil D. 2003. The genetics of phytate and phosphate accumulation in seeds and leaves of *Arabidopsis thaliana*, using natural variation. *Theoretical and Applied Genetics* **106**, 1234–1243.
- Bhatt GM, Ellison FW, Mares DJ. 1983. Inheritance studies in dormancy in three wheat crosses. In: Kruger J, LaBerge D. eds. *Third International symposium on preharvest sprouting in cereals*. Oxford: Taylor & Francis, 274–278.
- Bhattacharyya MK, Smith AM, Ellis TH, Hedley C, Martin C. 1990. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* **60**, 115–122.
- Birdseye D, de Boer LA, Bai H, Zhou P, Shen Z, Schmelz EA, Springer NM, Briggs SP. 2021. Plant height heterosis is quantitatively associated with expression levels of plastid ribosomal proteins. *Proceedings of the National Academy of Sciences, USA* **118**, e2109332118.
- Bokore FE, Cuthbert RD, Knox RE, Campbell HL, Meyer B, N'Diaye A, Pozniak CJ, DePauw R. 2022. Main effect and epistatic QTL affecting spike shattering and association with plant height revealed in two spring wheat (*Triticum aestivum* L.) populations. *Theoretical and Applied Genetics* **135**, 1143–1162.
- Brog YM, Osorio S, Yichie Y, Alseekh S, Bensal E, Kochevenko A, Zamir D, Fernie AR. 2019. A *Solanum neorickii* introgression population providing a powerful complement to the extensively characterized *Solanum pennellii* population. *The Plant Journal* **97**, 391–403.
- Broman KW. 2001. Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal* **30**, 44–52.
- Bulut M, Alseekh S, Fernie AR. 2023. Natural variation of respiration-related traits in plants. *Plant Physiology* **191**, 2120–2132.
- Calvin M. 1962. The path of carbon in photosynthesis. *Science* **135**, 879–889.
- Carrari F, Baxter C, Usadel B, *et al.* 2006. Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiology* **142**, 1380–1396.
- Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CW, Mumm R, Bouwmeester H, Visser RG, Keurentjes JJ. 2012. Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *Plant Physiology* **158**, 1306–1318.
- Caspar T, Huber SC, Somerville C. 1985. Alterations in growth, photosynthesis, and respiration in a starchless mutant of *Arabidopsis thaliana* (L.) deficient in chloroplast phosphoglucomutase activity. *Plant Physiology* **79**, 11–17.
- Causse M, Chaïb J, Lecomte L, Buret M, Hospital F. 2007. Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. *Theoretical and Applied Genetics* **115**, 429–442.
- Chaloner TM, Gurr SJ, Bebbler DP. 2021. Plant pathogen infection risk tracks global crop yields under climate change. *Nature Climate Change* **11**, 710–715.
- Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ. 2011. Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biology* **9**, e1001125.

- Chan EK, Rowe HC, Kliebenstein DJ.** 2010. Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics* **185**, 991–1007.
- Chen J, Hu X, Shi T, et al.** 2020. Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnology Journal* **18**, 1722–1735.
- Chen J, Wang J, Chen W, et al.** 2018. Metabolome analysis of multi-connected biparental chromosome segment substitution line populations. *Plant Physiology* **178**, 612–625.
- Chen W, Wang W, Peng M, et al.** 2016. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nature Communications* **7**, 12767.
- Colantonio V, Ferrão LFV, Tieman DM, Bliznyuk N, Sims C, Klee HJ, Munoz P, Resende MFR.** 2022. Metabolomic selection for enhanced fruit flavor. *Proceedings of the National Academy of Sciences, USA* **119**, e2115865119.
- Conklin PL, Williams EH, Last RL.** 1996. Environmental stress sensitivity of an ascorbic acid-deficient *Arabidopsis* mutant. *Proceedings of the National Academy of Sciences, USA* **93**, 9970–9974.
- Crane PL, Nyquist WE.** 1967. Effects of different gene-cytoplasm systems on quantitative characters in reciprocal F₂ crosses of maize (*Zea mays* L.). *Crop Science* **7**, 376–378.
- Dal Cin V, Tieman DM, Tohge T, et al.** 2011. Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *The Plant Cell* **23**, 2738–2753.
- Darwin C.** 1876. *Cross and self-fertilisation of plants*. London: John Murray.
- DeFalco TA, Zipfel C.** 2021. Molecular mechanisms of early plant pattern-triggered immune signaling. *Molecular Cell* **81**, 3449–3467.
- Dimitrov LN, Brem RB, Kruglyak L, Gottschling DE.** 2009. Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains. *Genetics* **183**, 365–383.
- Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP.** 2002. Cellulose biosynthesis in plants: from genes to rosettes. *Plant & Cell Physiology* **43**, 1407–1420.
- Dong X, Gao Y, Chen W, Wang W, Gong L, Liu X, Luo J.** 2015. Spatiotemporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Molecular Plant* **8**, 111–121.
- Dörmann P, Balbo I, Benning C.** 1999. Arabidopsis galactolipid biosynthesis and lipid trafficking mediated by DGD1. *Science* **284**, 2181–2184.
- Dworkin I.** 2005. Canalization, cryptic variation, and developmental buffering: a critical examination and analytical perspective. In: Hallgrímsson B, Hall BK, eds. *Variation*. Burlington, VT: Academic Press, 131–158.
- East EM.** 1908. Inbreeding in corn. In: Connecticut Agricultural Experimental Station Report for 1907, 419–428.
- Eckardt NA, Birchler JA, Meyers BC.** 2022. Focus on plant genetics: celebrating Gregor Mendel's 200th birth anniversary. *The Plant Cell* **34**, 2453–2454.
- Enink AH, Garretsen F.** 1980. Research on the inheritance of fasciation in lettuce (*Lactuca sativa* L.). *Euphytica* **29**, 653–660.
- Ei-Lithy ME, Bentsink L, Hanhart CJ, Ruys GJ, Rovito D, Broekhof JL, van der Poel HJ, van Eijk MJ, Vreugdenhil D, Koornneef M.** 2006. New *Arabidopsis* recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. *Genetics* **172**, 1867–1876.
- Ei-Mohsnawy E, Kopczak MJ, Schlodder E, Nowaczyk M, Meyer HE, Warscheid B, Karapetyan NV, Rögner M.** 2010. Structure and function of intact photosystem 1 monomers from the cyanobacterium *Thermosynechococcus elongatus*. *Biochemistry* **49**, 4740–4751.
- Erickson DL, Fenster CB, Stenøien HK, Price D.** 2004. Quantitative trait locus analyses and the study of evolutionary process. *Molecular Ecology* **13**, 2505–2522.
- Falconer DS.** 1989. *Introduction to quantitative genetics*, 3rd edn. New York, NY: John Wiley and Sons.
- Fang C, Luo J.** 2019. Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. *The Plant Journal* **97**, 91–100.
- Félix M-A, Barkoulas M.** 2012. Robustness and flexibility in nematode vulva development. *Trends in Genetics* **28**, 185–195.
- Fernie AR.** 2007. The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. *Phytochemistry* **68**, 2861–2880.
- Fernie AR, Bachem CWB, Helariutta Y, et al.** 2020. Synchronization of developmental, molecular and metabolic aspects of source–sink interactions. *Nature Plants* **6**, 55–66.
- Fernie AR, Bauwe H.** 2020. Wasteful, essential, evolutionary stepping stone? The multiple personalities of the photorespiratory pathway. *The Plant Journal* **102**, 666–677.
- Fernie AR, Klee HJ.** 2011. The use of natural genetic diversity in the understanding of metabolic organization and regulation. *Frontiers in Plant Science* **2**, 59.
- Fernie AR, Sonnewald U.** 2021. Plant biotechnology for sustainable agriculture and food safety. *Journal of Plant Physiology* **261**, 153416.
- Fernie AR, Tadmor Y, Zamir D.** 2006. Natural genetic variation for improving crop quality. *Current Opinion in Plant Biology* **9**, 196–202.
- Fernie AR, Tohge T.** 2017. The genetics of plant metabolism. *Annual Review of Genetics* **51**, 287–310.
- Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L.** 2000. Metabolite profiling for plant functional genomics. *Nature Biotechnology* **18**, 1157–1161.
- Fisher J, Bensal E, Zamir D.** 2017. Bimodality of stable and plastic traits in plants. *Theoretical and Applied Genetics* **130**, 1915–1926.
- Fridman E, Carrari F, Liu YS, Fernie AR, Zamir D.** 2004. Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* **305**, 1786–1789.
- Gage JL, Monier B, Giri A, Buckler ES.** 2020. Ten years of the maize nested association mapping population: impact, limitations, and future directions. *The Plant Cell* **32**, 2083–2093.
- Galili G, Amir R, Fernie AR.** 2016. The regulation of essential amino acid synthesis and accumulation in plants. *Annual Review of Plant Biology* **67**, 153–178.
- Galvão VC, Nordström KJ, Lanz C, Sulz P, Mathieu J, Posé D, Schmid M, Weigel D, Schneeberger K.** 2012. Synteny-based mapping-by-sequencing enabled by targeted enrichment. *The Plant Journal* **71**, 517–526.
- Gawarecka K, Siwinska J, Poznanski J, et al.** 2022. cis-Prenyltransferase 3 and α/β -hydrolase are new determinants of dolichol accumulation in *Arabidopsis*. *Plant, Cell & Environment* **45**, 479–495.
- Gerlin L, Barouk C, Genin S.** 2021. Polyamines: double agents in disease and plant immunity. *Trends in Plant Science* **26**, 1061–1071.
- Gong B-Q, Wang F-Z, Li J-F.** 2020. Hide-and-seek: chitin-triggered plant immunity and fungal counterstrategies. *Trends in Plant Science* **25**, 805–816.
- Gray MW, Burger G, Lang BF.** 1999. Mitochondrial evolution. *Science* **283**, 1476–1481.
- Gui S, Martinez-Rivas FJ, Wei W, Meng M, Yan J, Usadel B, Fernie AR.** 2023. Going broad and deep: sequencing driven insights into plant physiology, evolution and crop domestication. *The Plant Journal* **113**, 446–459.
- Hall MC, Dworkin I, Ungerer MC, Purugganan M.** 2007. Genetics of microenvironmental canalization in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **104**, 13717–13722.
- Hanson MR.** 1991. Plant mitochondrial mutations and male sterility. *Annual Review of Genetics* **25**, 461–486.
- Hartmann M, Zeier J.** 2018. l-lysine metabolism to N-hydroxyphenylpyruvic acid: an integral immune-activating pathway in plants. *The Plant Journal* **96**, 5–21.
- Hartmann M, Zeier T, Bernsdorff F, et al.** 2018. Flavin monooxygenase-generated N-hydroxyphenylpyruvic acid is a critical element of plant systemic immunity. *Cell* **173**, 456–469.e16.

- Hatch MD, Slack CR.** 1968. A new enzyme for the interconversion of pyruvate and phosphopyruvate and its role in the C₄ dicarboxylic acid pathway of photosynthesis. *Biochemical Journal* **106**, 141–146.
- Hayashi K, Alseekh S, Fernie AR.** 2023. Genetic and epigenetic control of the metabolome. *Proteomics*. In press. doi:10.1002/pmic.202200104.
- Herlihy J, Ludwig NR, van den Ackerveken G, McDowell JM.** 2019. Oomycetes used in Arabidopsis research. *The Arabidopsis Book* **17**, 1–26.
- Hill CB, Taylor JD, Edwards J, Mather D, Bacic A, Langridge P, Roessner U.** 2013. Whole-genome mapping of agronomic and metabolic traits to identify novel quantitative trait loci in bread wheat grown in a water-limited environment. *Plant Physiology* **162**, 1266–1281.
- Ho W-C, Zhang J.** 2016. Adaptive genetic robustness of *Escherichia coli* metabolic fluxes. *Molecular Biology and Evolution* **33**, 1164–1176.
- Holmes EC, Chen Y-C, Mudgett MB, Sattely ES.** 2021. Arabidopsis UGT76B1 glycosylates N-hydroxy-pipecolic acid and inactivates systemic acquired resistance in tomato. *The Plant Cell* **33**, 750–765.
- Hooker WJ.** 1837. *Petunia violacea*; hybrida. Purple *Petunia*; hybrid var. Curtis's Botanical Magazine **11**, 3556.
- Hou L, Chen W, Zhang Z, Pang X, Li Y.** 2020. Genome-wide association studies of fruit quality traits in jujube germplasm collections using genotyping-by-sequencing. *The Plant Genome* **13**, e20036.
- Hou S, Tsuda K.** 2022. Salicylic acid and jasmonic acid crosstalk in plant immunity. *Essays in Biochemistry* **66**, 647–656.
- Huang X-X, Wang Y, Lin J-S, Chen L, Li Y-J, Liu Q, Wang G-F, Xu F, Liu L, Hou B-K.** 2021. The novel pathogen-responsive glycosyltransferase UGT73C7 mediates the redirection of phenylpropanoid metabolism and promotes SNC1-dependent Arabidopsis immunity. *The Plant Journal* **107**, 149–165.
- Jiang K, Liberatore KL, Park SJ, Alvarez JP, Lippman ZB.** 2013. Tomato yield heterosis is triggered by a dosage sensitivity of the florigen pathway that fine-tunes shoot architecture. *PLoS Genetics* **9**, e1004043.
- Jimenez-Gomez JM, Corwin JA, Joseph B, Maloof JN, Kliebenstein DJ.** 2011. Genomic analysis of QTLs and genes altering natural variation in stochastic noise. *PLoS Genetics* **7**, e1002295.
- Johns DR, Neufeld MJ, Park RD.** 1992. An ND-6 mitochondrial DNA mutation associated with leber hereditary optic neuropathy. *Biochemical and Biophysical Research Communications* **187**, 1551–1557.
- Johnson WC, Gepts P.** 2002. The role of epistasis in controlling seed yield and other agronomic traits in an Andean × Mesoamerican cross of common bean (*Phaseolus vulgaris* L.). *Euphytica* **125**, 69–79.
- Joseph B, Corwin JA, Kliebenstein DJ.** 2015a. Genetic variation in the nuclear and organellar genomes modulates stochastic variation in the metabolome, growth, and defense. *PLoS Genetics* **11**, e1004779.
- Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ.** 2013. Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. *eLife* **2**, e00776.
- Joseph B, Lau L, Kliebenstein DJ.** 2015b. Quantitative variation in responses to root spatial constraint within *Arabidopsis thaliana*. *The Plant Cell* **27**, 2227–2243.
- Juenger TE, Sen S, Stowe KA, Simms EL.** 2005. Epistasis and genotype–environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica* **123**, 87–105.
- Juenger TE, Wayne T, Boles S, Symonds VV, McKay J, Coughlan SJ.** 2006. Natural genetic variation in whole-genome expression in *Arabidopsis thaliana*: the impact of physiological QTL introgression. *Molecular Ecology* **15**, 1351–1365.
- Kawase M, Sakamoto S.** 1984. Variation, geographical distribution and genetical analysis of esterase isozymes in foxtail millet, *Setaria italica* (L.) P. Beauv. *Theoretical and Applied Genetics* **67**, 529–533.
- Kerwin RE, Feusier J, Muok A, et al.** 2017. Epistasis × environment interactions among *Arabidopsis thaliana* glucosinolate genes impact complex traits and fitness in the field. *New Phytologist* **215**, 1249–1263.
- Keurentjes JJ, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LH, Jansen RC, Vreugdenhil D, Koornneef M.** 2006. The genetics of plant metabolism. *Nature Genetics* **38**, 842–849.
- Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T.** 2001. Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiology* **126**, 811–825.
- Knoch D, Werner CR, Meyer RC, Riewe D, Abbadi A, Lücke S, Snowdon RJ, Altmann T.** 2021. Multi-omics-based prediction of hybrid performance in canola. *Theoretical and Applied Genetics* **134**, 1147–1165.
- Kohler RE.** 1972. The reception of Eduard Buchner's discovery of cell-free fermentation. *Journal of the History of Biology* **5**, 327–353.
- Koornneef M, Alonso-Blanco C, Vreugdenhil D.** 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* **55**, 141–172.
- Kortschak HP, Hartt CE, Burr GO.** 1965. Carbon dioxide fixation in sugarcane leaves. *Plant Physiology* **40**, 209–213.
- Kress WJ, Soltis DE, Kersey PJ, Wegrzyn JL, Leebens-Mack JH, Gostel MR, Liu X, Soltis PS.** 2022. Green plant genomes: what we know in an era of rapidly expanding opportunities. *Proceedings of the National Academy of Sciences, USA* **119**, e2115640118.
- Krieger U, Lippman ZB, Zamir D.** 2010. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. *Nature Genetics* **42**, 459–463.
- Kumar LS.** 1999. DNA markers in plant improvement: an overview. *Biotechnology Advances* **17**, 143–182.
- Kumar S, Zavaliev R, Wu Q, et al.** 2022. Structural basis of NPR1 in activating plant immunity. *Nature* **605**, 561–566.
- Kusmec A, Srinivasan S, Nettleton D, Schnable PS.** 2017. Distinct genetic architectures for phenotype means and plasticities in *Zea mays*. *Nature Plants* **3**, 715–723.
- Kusterer B, Muminovic J, Utz HF, Piepho H-P, Barth S, Heckenberger M, Meyer RC, Altmann T, Melchinger AE.** 2007. Analysis of a triple test-cross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in Arabidopsis. *Genetics* **175**, 2009–2017.
- Labroo MR, Studer AJ, Rutkoski JE.** 2021. Heterosis and hybrid crop breeding: a multidisciplinary review. *Frontiers in Genetics* **12**, 643761.
- Lachowiec J, Mason GA, Schultz K, Queitsch C.** 2018. Redundancy, feedback, and robustness in the *Arabidopsis thaliana* BZR/BEH gene family. *Frontiers in Genetics* **9**, 523.
- Lachowiec J, Queitsch C, Kliebenstein DJ.** 2016. Molecular mechanisms governing differential robustness of development and environmental responses in plants. *Annals of Botany* **117**, 795–809.
- Lai J, Li R, Xu X, et al.** 2010. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics* **42**, 1027–1030.
- Laitinen RAE, Nikoloski Z.** 2019. Genetic basis of plasticity in plants. *Journal of Experimental Botany* **70**, 739–745.
- Landry LG, Chapple CC, Last RL.** 1995. Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiology* **109**, 1159–1166.
- Lehner B.** 2010. Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS ONE* **5**, e9035.
- Li B, Zhang Y, Mohammadi SA, Huai D, Zhou Y, Kliebenstein DJ.** 2016. An integrative genetic study of rice metabolism, growth and stochastic variation reveals potential C/N partitioning loci. *Scientific Reports* **6**, 30143.
- Li C, Guan H, Jing X, et al.** 2022. Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. *Nature Plants* **8**, 750–763.
- Li J, Ou-Lee TM, Raba R, Amundson RG, Last RL.** 1993. Arabidopsis flavonoid mutants are hypersensitive to UV-B irradiation. *The Plant Cell* **5**, 171–179.
- Li K, Wen W, Alseekh S, et al.** 2019. Large-scale metabolite quantitative trait locus analysis provides new insights for high-quality maize improvement. *The Plant Journal* **99**, 216–230.
- Li L, Wu X, Chen J, Wang S, Wan Y, Ji H, Wen Y, Zhang J.** 2022. Genetic dissection of epistatic interactions contributing yield-related agronomic traits in rice using the compressed mixed model. *Plants* **11**, 2504.

- Li Y, Cao K, Zhu G, *et al.* 2019. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biology* **20**, 36.
- Liao L, Zhang W, Zhang B, *et al.* 2021. Unraveling a genetic roadmap for improved taste in the domesticated apple. *Molecular Plant* **14**, 1454–1471.
- Lin TP, Caspar T, Somerville C, Preiss J. 1988. Isolation and characterization of a starchless mutant of *Arabidopsis thaliana* (L.) Heynh lacking ADPglucose pyrophosphorylase activity. *Plant Physiology* **86**, 1131–1135.
- Lippman Z, Tanksley SD. 2001. Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom. *Genetics* **158**, 413–422.
- Lippman ZB, Zamir D. 2007. Heterosis: revisiting the magic. *Trends in Genetics* **23**, 60–66.
- Lisec J, Meyer RC, Steinfath M, *et al.* 2008. Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *The Plant Journal* **53**, 960–972.
- Liu C, Atanasov KE, Tiburcio AF, Alcázar R. 2019. The polyamine putrescine contributes to H₂O₂ and *RbohD/F*-dependent positive feedback loop in *Arabidopsis* PAMP-triggered immunity. *Frontiers in Plant Science* **10**, 894.
- Liu H-J, Wang X, Xiao Y, *et al.* 2020. CUBIC: an atlas of genetic architecture promises directed maize improvement. *Genome Biology* **21**, 20.
- Liu H-J, Yan J. 2019. Crop genome-wide association study: a harvest of biological relevance. *The Plant Journal* **97**, 8–18.
- Liu J, Fernie AR, Yan J. 2021. Crop breeding – from experience-based selection to precision design. *Journal of Plant Physiology* **256**, 153313.
- Liu N, Du Y, Warburton ML, Xiao Y, Yan J. 2021. Phenotypic plasticity contributes to maize adaptation and heterosis. *Molecular Biology and Evolution* **38**, 1262–1275.
- Liu Q, Li X, Yan S, *et al.* 2018. OsWRKY67 positively regulates blast and bacteria blight resistance by direct activation of PR genes in rice. *BMC Plant Biology* **18**, 257.
- Luo J. 2015. Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology* **24**, 31–38.
- MacAlister CA, Park SJ, Jiang K, Marcel F, Bendahmane A, Izkovich Y, Eshed Y, Lippman ZB. 2012. Synchronization of the flowering transition by the tomato *TERMINATING FLOWER* gene. *Nature Genetics* **44**, 1393–1398.
- Mackay TFC. 2014. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nature Reviews Genetics* **15**, 22–33.
- Maloo JN. 2003. Genomic approaches to analyzing natural variation in *Arabidopsis thaliana*. *Current Opinion in Genetics and Development* **13**, 576–582.
- Manicacci D, Camus-Kulandaivelu L, Fourmann M, *et al.* 2009. Epistatic interactions between *Opaque2* transcriptional activator and its target gene *CyPPDK1* control kernel trait variation in maize. *Plant Physiology* **150**, 506–520.
- Martínez H, Barrachina S, Castillo M, Quintana-Orti ES, Rambla de Argila J, Farre X, Navarro A. 2018. FaST-LMM for two-way epistasis tests on high-performance clusters. *Journal of Computational Biology* **25**, 862–870.
- Matsuda F, Nakabayashi R, Yang Z, Okazaki Y, Yonemaru J, Ebana K, Yano M, Saito K. 2015. Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *The Plant Journal* **81**, 13–23.
- Matsuda F, Okazaki Y, Oikawa A, Kusano M, Nakabayashi R, Kikuchi J, Yonemaru J, Ebana K, Yano M, Saito K. 2012. Dissection of genotype–phenotype associations in rice grains using metabolome quantitative trait loci analysis. *The Plant Journal* **70**, 624–636.
- McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences, USA* **36**, 344–355.
- McClintock B. 1951. Chromosome organization and genic expression. *Cold Spring Harbor Symposia on Quantitative Biology* **16**, 13–47.
- McKay JK, Richards JH, Nemali KS, Sen S, Mitchell-Olds T, Boles S, Stahl EA, Wayne T, Juenger TE. 2008. Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, KAS-1 × TSU-1. *Evolution* **62**, 3014–3026.
- Mendel G. 1865. Versuche über Pflanzen-Hybriden. *Verhandlungen des Naturforschenden Vereines in Brünn* **4**, 3–47.
- Meyer RC, Steinfath M, Lisec J, *et al.* 2007. The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **104**, 4759–4764.
- Mohnike L, Rekhter D, Huang W, Feussner K, Tian H, Herrfurth C, Zhang Y, Feussner I. 2021. The glycosyltransferase UGT76B1 modulates *N*-hydroxy-pipecolic acid homeostasis and plant immunity. *The Plant Cell* **33**, 735–749.
- Moison M, Roux F, Quadrado M, Duval R, Ekovich M, Lê DH, Verzaux M, Budar F. 2010. Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in *Arabidopsis thaliana*. *The Plant Journal* **63**, 728–738.
- Monforte AJ. 2021. Heterosis goes underground. *Journal of Experimental Botany* **72**, 6081–6083.
- Moose SP, Dudley JW, Rocheford TR. 2004. Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends in Plant Science* **9**, 358–364.
- Moreto AL, Ramalho MAP, Bruzi AT. 2012. Epistasis in an Andean × Mesoamerican cross of common bean. *Euphytica* **186**, 755–760.
- Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, Fernie AR, Usadel B, Nikoloski Z, Persson S. 2011. PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *The Plant Cell* **23**, 895–910.
- Návarová H, Bernsdorff F, Döring A-C, Zeier J. 2012. Pipecolic acid, an endogenous mediator of defense amplification and priming, is a critical regulator of inducible plant immunity. *The Plant Cell* **24**, 5123–5141.
- Needham J. 1970. *The chemistry of life: eight lectures on the history of biochemistry*. Cambridge, UK: Cambridge University Press.
- Nguyen KL, Grondin A, Courtois B, Gantet P. 2019. Next-generation sequencing accelerates crop gene discovery. *Trends in Plant Science* **24**, 263–274.
- Ofner I, Lashbrooke J, Pleban T, Aharoni A, Zamir D. 2016. *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. *The Plant Journal* **87**, 151–160.
- Park SJ, Jiang K, Schatz MC, Lippman ZB. 2012. Rate of meristem maturation determines inflorescence architecture in tomato. *Proceedings of the National Academy of Sciences, USA* **109**, 639–644.
- Phillips RL, Vasil IK. 2001. *DNA-based markers in plants*, 2nd edn. Dordrecht: Kluwer Academic Publishers.
- Piasecka A, Jedrzejczak-Rey N, Bednarek P. 2015. Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytologist* **206**, 948–964.
- Purugganan MD, Jackson SA. 2021. Advancing crop genomics from lab to field. *Nature Genetics* **53**, 595–601.
- Qin R, Ma T, Cao M, *et al.* 2022. QTL detection for internode component index in wheat using a RIL mapping population. *Plant Molecular Biology Reporter* **41**, 195–208.
- Queitsch C, Sangster TA, Lindquist S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624.
- Rabbi IY, Kayondo SI, Bauchet G, *et al.* 2022. Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. *Plant Molecular Biology* **109**, 195–213.
- Rathan ND, Krishna H, Ellur RK, *et al.* 2022. Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). *Scientific Reports* **12**, 7037.
- Rehal PK, Tuan PA, Nguyen T-N, Cattani DJ, Humphreys DG, Ayele BT. 2022. Genetic variation of seed dormancy in wheat (*Triticum aestivum* L.) is mediated by transcriptional regulation of abscisic acid metabolism and signaling. *Plant Science* **324**, 111432.

- Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE. 2012. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proceedings of the National Academy of Sciences, USA* **109**, 8872–8877.
- Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie A. 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *The Plant Cell* **13**, 11–29.
- Rosado-Souza L, Yokoyama R, Sonnewald U, Fernie AR. 2022. Understanding source–sink interactions: progress in model plants and translational research to crops in controlled and field conditions. *Molecular Plant* **16**, 96–121.
- Roubertoux PL, Sluiter F, Carlier M, *et al.* 2003. Mitochondrial DNA modifies cognition in interaction with the nuclear genome and age in mice. *Nature Genetics* **35**, 65–69.
- Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ. 2008. Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *The Plant Cell* **20**, 1199–1216.
- Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342.
- Samo N, Ebert A, Kopka J, Mozgová I. 2021. Plant chromatin, metabolism and development – an intricate crosstalk. *Current Opinion in Plant Biology* **61**, 102002.
- Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikolski Z, Fernie AR, Causse M. 2014. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiology* **165**, 1120–1132.
- Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**, 552–560.
- Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR. 2008. Mode of inheritance of primary metabolic traits in tomato. *The Plant Cell* **20**, 509–523.
- Schauer N, Semel Y, Roessner U, *et al.* 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology* **24**, 447–454.
- Schauer N, Zamir D, Fernie AR. 2005. Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *Journal of Experimental Botany* **56**, 297–307.
- Scheiner SM. 1993. Genetics and evolution of phenotypic plasticity. *Annual Review of Ecology and Systematics* **24**, 35–68.
- Schnable PS, Springer NM. 2013. Progress toward understanding heterosis in crop plants. *Annual Review of Plant Biology* **64**, 71–88.
- Schnable PS, Wise RP. 1998. The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends in Plant Science* **3**, 175–180.
- Schon EA, DiMauro S, Hirano M. 2012. Human mitochondrial DNA: roles of inherited and somatic mutations. *Nature Reviews Genetics* **13**, 878–890.
- Schwahn K, de Souza LP, Fernie AR, Tohge T. 2014. Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. *Journal of Integrative Plant Biology* **56**, 864–875.
- Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* **432**, 779–782.
- Sergeeva LI, Vonk J, Keurentjes JJ, van der Plas LH, Koornneef M, Vreugdenhil D. 2004. Histochemical analysis reveals organ-specific quantitative trait loci for enzyme activities in *Arabidopsis*. *Plant Physiology* **134**, 237–245.
- Shang L, Liang Q, Wang Y, Zhao Y, Wang K, Hua J. 2016. Epistasis together with partial dominance, over-dominance and QTL by environment interactions contribute to yield heterosis in upland cotton. *Theoretical and Applied Genetics* **129**, 1429–1446.
- Sharwood RE, Quick WP, Sargent D, Estavillo GM, Silva-Perez V, Furbank RT. 2022. Mining for allelic gold: finding genetic variation in photosynthetic traits in crops and wild relatives. *Journal of Experimental Botany* **73**, 3085–3108.
- Shen G, Zhan W, Chen H, Xing Y. 2014. Dominance and epistasis are the main contributors to heterosis for plant height in rice. *Plant Science* **215**, 11–18.
- Shinya T, Nakagawa T, Kaku H, Shibuya N. 2015. Chitin-mediated plant–fungal interactions: catching, hiding and handshaking. *Current Opinion in Plant Biology* **26**, 64–71.
- Shull GH. 1908. The composition of a field of maize. *Journal of Heredity* **4**, 296–301.
- Singh M. 1966. Cytoplasmic effects on agronomic characters in maize. *Indian Journal of Genetics and Plant Breeding* **26**, 386–390.
- Sink KC. 1973. The inheritance of apetalous flower type in *Petunia hybrida* Vilm. and linkage tests with the genes for flower doubleness and grandiflora characters and its use in hybrid seed production. *Euphytica* **22**, 520–526.
- Somerville C. 2000. The twentieth century trajectory of plant biology. *Cell* **100**, 13–25.
- Soyk S, Lemmon ZH, Oved M, *et al.* 2017. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell* **169**, 1142–1155. e12.
- Stitt M, Sonnewald U. 1995. Regulation of metabolism in transgenic plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **46**, 341–368.
- Strauch RC, Svedin E, Dilkes B, Chapple C, Li X. 2015. Discovery of a novel amino acid racemase through exploration of natural variation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **112**, 11726–11731.
- Sulli M, Barchi L, Toppino L, Diretto G, Sala T, Lanteri S, Rotino GL, Giuliano G. 2021. An eggplant recombinant inbred population allows the discovery of metabolic QTLs controlling fruit nutritional quality. *Frontiers in Plant Science* **12**, 638195.
- Szymański J, Bocobza S, Panda S, *et al.* 2020. Analysis of wild tomato introgression lines elucidates the genetic basis of transcriptome and metabolome variation underlying fruit traits and pathogen response. *Nature Genetics* **52**, 1111–1121.
- Tang M, Li B, Zhou X, *et al.* 2021. A genome-scale TF–DNA interaction network of transcriptional regulation of *Arabidopsis* primary and specialized metabolism. *Molecular Systems Biology* **17**, e10625.
- Tang Z, Yang Z, Hu Z, Zhang D, Lu X, Jia B, Deng D, Xu C. 2013. Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize. *Molecular Breeding* **31**, 1–14.
- Tao D, Hu F, Yang J, Yang G, Yang Y, Xu P, Li J, Ye C, Dai L. 2004. Cytoplasm and cytoplasm–nucleus interactions affect agronomic traits in japonica rice. *Euphytica* **135**, 129–134.
- Taylor RW, Turnbull DM. 2005. Mitochondrial DNA mutations in human disease. *Nature Reviews Genetics* **6**, 389–402.
- Tieman D, Taylor M, Schauer N, Fernie AR, Hanson AD, Klee HJ. 2006. Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proceedings of the National Academy of Sciences, USA* **103**, 8287–8292.
- Tieman D, Zhu G, Resende MF Jr, *et al.* 2017. A chemical genetic roadmap to improved tomato flavor. *Science* **355**, 391–394.
- Till BJ, Reynolds SH, Greene EA, *et al.* 2003. Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Research* **13**, 524–530.
- Tohge T, Wendenburg R, Ishihara H, *et al.* 2016. Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nature Communications* **7**, 12399.
- Ton J, Van Pelt JA, Van Loon LC, Pieterse CM. 2002. Differential effectiveness of salicylate-dependent and jasmonate/ethylene-dependent induced resistance in *Arabidopsis*. *Molecular Plant-Microbe Interactions* **15**, 27–34.
- Torgeman S, Zamir D. 2023. Epistatic QTLs for yield heterosis in tomato. *Proceedings of the National Academy of Sciences, USA* **120**, e2205787119.
- Tsuda K, Mine A, Bethke G, Igarashi D, Botanga CJ, Tsuda Y, Glazebrook J, Sato M, Katagiri F. 2013. Dual regulation of gene

- expression mediated by extended MAPK activation and salicylic acid contributes to robust innate immunity in *Arabidopsis thaliana*. *PLoS Genetics* **9**, e1004015.
- Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D.** 2021. Genome-wide association studies. *Nature Reviews Methods Primers* **1**, 59.
- Vanholme R, De Meester B, Ralph J, Boerjan W.** 2019. Lignin biosynthesis and its integration into metabolism. *Current Opinion in Biotechnology* **56**, 230–239.
- Vergara-Diaz O, Vatter T, Vicente R, Obata T, Nieto-Taladriz MT, Aparicio N, Carlisle Kefauver S, Fernie A, Araus JL.** 2020. Metabolome profiling supports the key role of the spike in wheat yield performance. *Cells* **9**, 1025.
- Verslues PE, Lasky JR, Juenger TE, Liu TW, Kumar MN.** 2014. Genome-wide association mapping combined with reverse genetics identifies new effectors of low water potential-induced proline accumulation in *Arabidopsis*. *Plant Physiology* **164**, 144–159.
- Villate A, San Nicolas M, Gallastegi M, Aulas PA, Olivares M, Usobiaga A, Etxebarria N, Aizpurua-Olaizola O.** 2021. Review: Metabolomics as a prediction tool for plants performance under environmental stress. *Plant Science* **303**, 110789.
- Vitezica ZG, Legarra A, Toro MA, Varona L.** 2017. Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Genetics* **206**, 1297–1307.
- von Korff M, Wang H, Léon J, Pillen K.** 2004. Development of candidate introgression lines using an exotic barley accession (*Hordeum vulgare* ssp. *spontaneum*) as donor. *Theoretical and Applied Genetics* **109**, 1736–1745.
- Waddington CH.** 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565.
- Wahyuni Y, Stahl-Hermes V, Ballester AR, et al.** 2014. Genetic mapping of semi-polar metabolites in pepper fruits (*Capsicum* sp.): towards unravelling the molecular regulation of flavonoid quantitative trait loci. *Molecular Breeding* **33**, 503–518.
- Walker RP, Chen ZH, Famiani F.** 2021. Gluconeogenesis in plants: a key interface between organic acid/amino acid/lipid and sugar metabolism. *Molecules* **26**, 5129.
- Wang L, Czedik-Eysenberg A, Mertz RA, et al.** 2014. Comparative analyses of C₄ and C₃ photosynthesis in developing leaves of maize and rice. *Nature Biotechnology* **32**, 1158–1165.
- Weigel D.** 2016. Why I love genetics: essay on occasion of being awarded the GSA medal 2016. *Genetics* **204**, 841–843.
- Weinreich DM, Lan Y, Wylie CS, Heckendorn RB.** 2013. Should evolutionary geneticists worry about higher-order epistasis? *Current Opinion in Genetics & Development* **23**, 700–707.
- Wen W, Li D, Li X, et al.** 2014. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications* **5**, 3438.
- Wen W, Li K, Alseekh S, et al.** 2015. Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *The Plant Cell* **27**, 1839–1856.
- Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ.** 2007. Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genetics* **3**, e162.
- Wheeler GL, Jones MA, Smirnoff N.** 1998. The biosynthetic pathway of vitamin C in higher plants. *Nature* **393**, 365–369.
- White O.** 1917. Studies of inheritance in *Pisum*. II. The present state of knowledge of heredity and variation in peas. *Proceedings of the American Philosophical Society* **56**, 487–588.
- Wijesingha Ahchige M, Fernie AR, Alseekh S.** 2023. PANTOTHENATE KINASE4, LOSS OF GDU2, and TRANSPOSON PROTEIN1 affect the canalization of tomato fruit metabolism. *Plant Physiology* **192**, 442–468.
- Wijnen CL, Keurentjes JJ.** 2014. Genetic resources for quantitative trait analysis: novelty and efficiency in design from an *Arabidopsis* perspective. *Current Opinion in Plant Biology* **18**, 103–109.
- Willett CS.** 2012. Quantifying the elevation of mitochondrial DNA evolutionary substitution rates over nuclear rates in the intertidal copepod *Tigriopus californicus*. *Journal of Molecular Evolution* **74**, 310–318.
- Wolf JB.** 2009. Cytonuclear interactions can favor the evolution of genomic imprinting. *Evolution* **63**, 1364–1371.
- Wolff JB, Price L.** 1957. Terminal steps of chlorophyll a biosynthesis in higher plants. *Archives of Biochemistry and Biophysics* **72**, 293–301.
- Würschum T, Maurer HP, Kraft T, Janssen G, Nilsson C, Reif JC.** 2011. Genome-wide association mapping of agronomic traits in sugar beet. *Theoretical and Applied Genetics* **123**, 1121–1131.
- Yamamura C, Mizutani E, Okada K, et al.** 2015. Diterpenoid phytoalexin factor, a bHLH transcription factor, plays a central role in the biosynthesis of diterpenoid phytoalexins in rice. *The Plant Journal* **84**, 1100–1113.
- Yang C, Shen S, Zhou S, et al.** 2022. Rice metabolic regulatory network spanning the entire life cycle. *Molecular Plant* **15**, 258–275.
- Yang J, Mezrouk S, Baumgarten A, Buckler ES, Guill KE, McMullen MD, Mumm RH, Ross-Ibarra J.** 2017. Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. *PLoS Genetics* **13**, e1007019.
- Ye J, Wang X, Hu T, et al.** 2017. An indel in the promoter of *AI-ACTIVATED MALATE TRANSPORTER9* selected during tomato domestication determines fruit malate contents and aluminum tolerance. *The Plant Cell* **29**, 2249–2268.
- Yi N, Xu S.** 2002. Mapping quantitative trait loci with epistatic effects. *Genetical Research* **79**, 185–198.
- Yuan M, Ngou BPM, Ding P, Xin XF.** 2021. PTI-ETI crosstalk: an integrative view of plant immunity. *Current Opinion in Plant Biology* **62**, 102030.
- Zaynab M, Fatima M, Abbas S, Sharif Y, Umair M, Zafar MH, Bahadar K.** 2018. Role of secondary metabolites in plant defense against pathogens. *Microbial Pathogenesis* **124**, 198–202.
- Zeier J.** 2013. New insights into the regulation of plant immunity by amino acid metabolic pathways. *Plant, Cell & Environment* **36**, 2085–2103.
- Zeng X, Yuan H, Dong X, et al.** 2020. Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of qingke. *Molecular Plant* **13**, 112–127.
- Zhan C, Shen S, Yang C, Liu Z, Fernie AR, Graham IA, Luo J.** 2022. Plant metabolic gene clusters in the multi-omics era. *Trends in Plant Science* **27**, 981–1001.
- Zhang C, Atanasov KE, Alcázar R.** 2023. Spermine inhibits PAMP-induced ROS and Ca²⁺ burst and reshapes the transcriptional landscape of PAMP-triggered immunity in *Arabidopsis*. *Journal of Experimental Botany* **74**, 427–442.
- Zhang J.** 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* **18**, 292–298.
- Zhang J, Singh A, Mueller DS, Singh AK.** 2015. Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. *The Plant Journal* **84**, 1124–1136.
- Zhang W, Alseekh S, Zhu X, Zhang Q, Fernie AR, Kuang H, Wen W.** 2020. Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *The Plant Journal* **104**, 613–630.
- Zhang Z, Ober JA, Kliebenstein DJ.** 2006. The gene controlling the quantitative trait locus *EPITHIOSPECIFIER MODIFIER1* alters glucosinolate hydrolysis and insect resistance in *Arabidopsis*. *The Plant Cell* **18**, 1524–1536.
- Zhou S, Kremling KA, Bandillo N, et al.** 2019. Metabolome-scale genome-wide association studies reveal chemical diversity and genetic control of maize specialized metabolites. *The Plant Cell* **31**, 937–955.
- Zhou Y, Ma Y, Zeng J, et al.** 2016. Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nature Plants* **2**, 16183.
- Zhu G, Wang S, Huang Z, et al.** 2018. Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**, 249–261.e12.
- Zouni A, Kern J, Frank J, Hellweg T, Behlke J, Saenger W, Irrgang KD.** 2005. Size determination of cyanobacterial and higher plant photosystem II by gel permeation chromatography, light scattering, and ultracentrifugation. *Biochemistry* **44**, 4572–4581.