the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) doi: 10.1111/tpj.17009

Bivariate GWA mapping reveals associations between aliphatic glucosinolates and plant responses to thrips and heat stress

Bader Arouisse[1], Manus P.M. Thoen[2,3], Willem Kruijer[1], Jonathan F. Kunst[1], Maarten A. Jongsma[4], Joost J.B. Keurentjes[5], Rik Kooke[1,5], Ric C.H. de Vos[4], Roland Mumm[4], Fred A. van Eeuwijk[1], Marcel Dicke[2] and Karen J. Kloth[2,*]

[1]*Biometris, Wageningen University and Research, Wageningen, the Netherlands,*
[2]*Laboratory of Entomology, Wageningen University & Research, Wageningen, the Netherlands,*
[3]*Enza Seeds, Enkhuizen, the Netherlands,*
[4]*Bioscience, Wageningen Plant Research, Wageningen University and Research, Wageningen, the Netherlands, and*
[5]*Laboratory of Genetics, Wageningen University and Research, Wageningen, the Netherlands*

Received 16 June 2022; accepted 20 August 2024.
*For correspondence (e-mail karen.kloth@wur.nl).

SUMMARY

Although plants harbor a huge phytochemical diversity, only a fraction of plant metabolites is functionally characterized. In this work, we aimed to identify the genetic basis of metabolite functions during harsh environmental conditions in *Arabidopsis thaliana*. With machine learning algorithms we predicted stress-specific metabolomes for 23 (a)biotic stress phenotypes of 300 natural Arabidopsis accessions. The prediction models identified several aliphatic glucosinolates (GLSs) and their breakdown products to be implicated in responses to heat stress in siliques and herbivory by Western flower thrips, *Frankliniella occidentalis*. Bivariate GWA mapping of the metabolome predictions and their respective (a)biotic stress phenotype revealed genetic associations with *MAM*, *AOP*, and *GS-OH*, all three involved in aliphatic GSL biosynthesis. We, therefore, investigated thrips herbivory on *AOP*, *MAM*, and *GS-OH* loss-of-function and/or overexpression lines. Arabidopsis accessions with a combination of MAM2 and AOP3, leading to 3-hydroxypropyl dominance, suffered less from thrips feeding damage. The requirement of MAM2 for this effect could, however, not be confirmed with an introgression line of ecotypes Cvi and L*er*, most likely due to other, unknown susceptibility factors in the L*er* background. However, AOP2 and GS-OH, adding alkenyl or hydroxy-butenyl groups, respectively, did not have major effects on thrips feeding. Overall, this study illustrates the complex implications of aliphatic GSL diversity in plant responses to heat stress and a cell-content-feeding herbivore.

Keywords: (a)biotic stresses, MAM, AOP, metabolomics, quantitative genetics.

INTRODUCTION

Plants are exposed to a wide range of stresses, such as drought, heat, salinity, pathogens, and herbivorous insects. To maximize their fitness, they evolved sophisticated strategies to cope with their unpredictable environment, including reallocation or recycling of resources for the biosynthesis of metabolites and protective structures. Plant resistance strategies involve both constitutive and induced strategies (Kant et al., 2015). Constitutive thick epicuticular wax layers, thorns, and protective compounds are eminent adaptations to severe (a)biotic stresses (Gómez & Zamora, 2002; War et al., 2012). On the other hand, induced responses act upon unpredictable environmental changes or infections and trigger *de novo* biosynthesis or modifications of existing metabolites (Kant et al., 2015). Depending on the cost–benefit analysis, plants invest in these different strategies to improve their fitness (Neilson et al., 2013; Steppuhn & Baldwin, 2008). The biosynthesis of plant secondary metabolites is considered to be an adaptive and dynamic strategy to cope with stressful conditions in a fluctuating environment (Isah, 2019). Apart from their central role in regulating plant development and growth, via their interactions through signaling processes and pathways, secondary metabolites play a major part in orchestrating plant reactions to biotic and abiotic stresses (Ashraf et al., 2018; Grotewold, 2005).

Secondary metabolites, also called specialized compounds, can be classified into three major groups based

on their biosynthetic pathways: terpenes, phenolics (flavonoids and phenylpropanoids), and nitrogen-containing compounds, the latter including cyanogenic glycosides, alkaloids, and glucosinolates (Degenhardt, 2008; Qualley & Dudareva, 2008). Glucosinolates (GSLs) are a distinct class of sulfur-containing secondary metabolites mainly found in the Brassicaceae family. Structurally, they are anions composed of three moieties: thiohydroximates containing an S-linked β-glucopyranosyl residue and an O-linked sulfate residue with a variable aglycone side chain derived from an α-amino acid which is then classified into (1) aliphatic GSLs (derived from alanine, leucine, isoleucine, valine, or methionine), (2) indole GSLs (derived from tryptophan), and (3) aromatic GSLs (derived from phenylalanine or tyrosine) (Halkier & Gershenzon, 2006). GSLs are usually stored in vacuoles and chemically stable, but when plant cells are damaged, they come into contact with myrosinase enzymes that hydrolyze GSLs into electrophilic compounds (Bones & Rossiter, 1996; Pang et al., 2009). Several GSL hydrolysis products, such as isothiocyanates, nitriles, epithionitriles, and thiocyanates, play a central role in plant defense against herbivorous insects and microbial pathogens. Of these hydrolysis products, isothiocyanates are the most studied and known to be toxic to a wide range of plant pathogens and herbivores (Gimsing & Kirkegaard, 2009; Halkier & Gershenzon, 2006) and considered to be involved in drought-induced stomatal closure (Salehin et al., 2019).

In the last decade, genome-wide association (GWA) studies have become a standard tool for studying the genetic architecture of quantitative traits (Tibbs Cortes et al., 2021). With the advances in high-throughput genotyping technologies and the progress in high-resolution liquid and gas chromatography–mass spectrometry (LC- and GC/MS), plant metabolomes, or metabolite traits, can be characterized and compared in a comprehensive manner (Keurentjes et al., 2008). Consequently, the challenge of including such omics data in genetic analyses is pertinent. Joint association analysis of multiple traits in a genome-wide association study offers several advantages and opportunities over individual association analysis of single traits or metabolites. Multi-trait GWA mapping makes use of the extra information provided by the correlation between traits to gain statistical power (Korte et al., 2012; Zhou & Stephens, 2014). Additionally, it has the ability to identify the multigenic basis of complex traits and to unravel genotypic variants with pleiotropic effects where single genetic variants are involved in the expression of multiple traits. Nevertheless, multi-trait GWA studies are often limited to small numbers of traits due to the upscaling of the computational cost, hence precluding the integration of large omics data with high-resolution SNP arrays (Porter & O'Reilly, 2017). Dimension reduction techniques (Arouisse et al., 2021) are a viable option to overcome these limitations.

In this study, we aimed to dissect the genetic basis of biochemical pathways of individual plant stress responses in *Arabidopsis thaliana*, using a bivariate GWA mapping of the metabolome, consisting of both primary and secondary metabolites, and the stress response of a natural GWA population. When exposing this population to a diverse set of 23 (a)biotic stresses (Data S1), including caterpillars, whiteflies, aphids, *Botrytis cinerea*, drought and salt, strong associations were revealed between certain biochemical pathways and feeding damage of the Western flower thrips, *Frankliniella occidentalis*, and heat stress. Thrips are cell-content feeders and particularly *F. occidentalis* is known to be affected by the jasmonic acid pathway (Abe et al., 2008; Escobar-Bravo et al., 2017), which is renowned for the induction of several secondary metabolite classes, including terpenes and GSLs (van der Fits & Memelink, 2000). Heat stress, in contrast, induces abscisic acid and is often connected to the accumulation of flavonoids that function as antioxidants and UV filters (Li et al., 2021). Here, we found that Arabidopsis responses to both stresses are affiliated with natural variation in aliphatic GSLs and their biosynthetic loci.

RESULTS

Bivariate GWA mapping with metabolomic prediction reveals new QTLs

To decipher the complex biochemical processes underlying Arabidopsis resistance to 23 (a)biotic stresses, penalized regression analysis on constitutive primary and secondary metabolomic profiles was performed, resulting in a predicted response to different stress treatments. This initial step reduced the multidimensionality of the secondary and primary metabolites to a single metabolomic estimate for each genotype. Bivariate GWA mapping on the stress response and its metabolic prediction identified several SNPs located across various candidate genes that are putatively involved in the regulation of stress responses in Arabidopsis (Data S1). Among these were associations between the number of eggs laid by *Aleyrodes proletella* whiteflies and SNPs in TREHALOSE PHOSPHATE SYNTHASE10 (TPS10), an interesting candidate as TPS11 was previously found to be involved in Arabidopsis defense against the phloem feeder *Myzus persicae* (Singh et al., 2011). Also, SNPs in genes involved in the biosynthesis of GSLs and their hydrolysis products were identified in relation to thrips and heat stress. These SNPs were neither captured by previous univariate GWA mapping of the stress responses (Thoen et al., 2017), nor by univariate GWA mapping with SNP imputations (Figure 1; Data S1).

MAM and *AOP* loci associated with thrips feeding damage

Out of the total of 449 considered metabolites, 13 were selected by LASSO to build the prediction model to infer

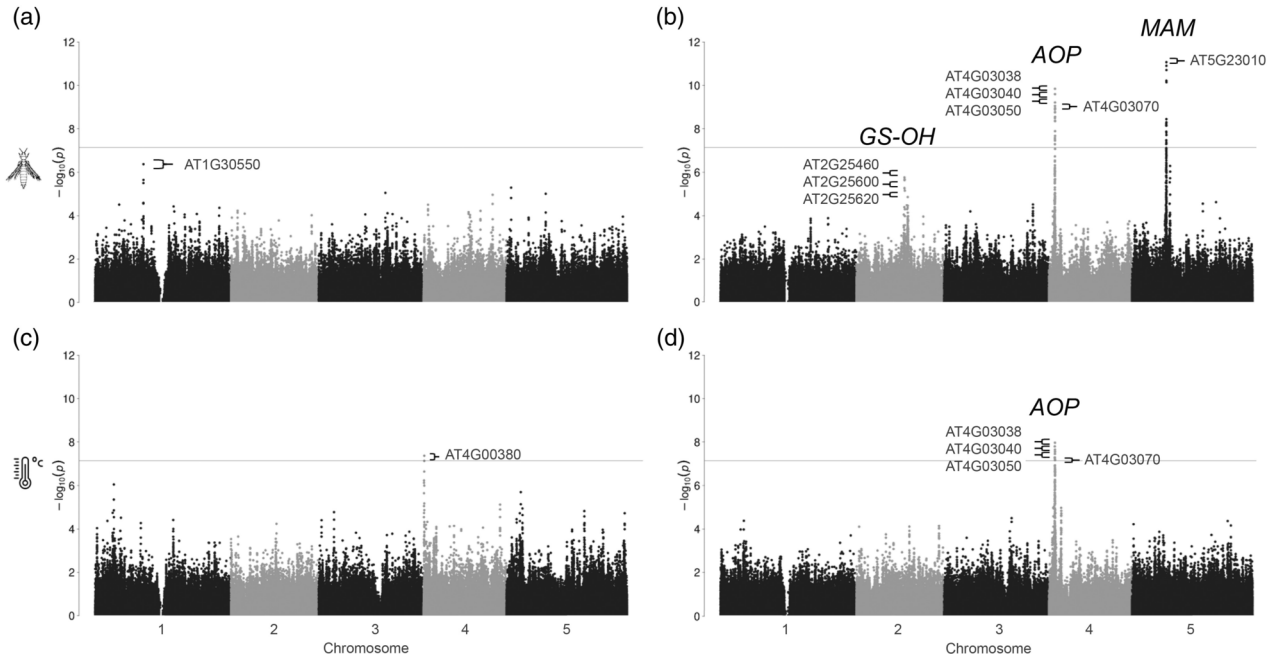

FIGURE 1. Univariate GWA mapping of stress responses and bivariate GWA mapping of stress responses and their metabolomic prediction in a population of 300 natural Arabidopsis accessions. (a) Univariate GWA mapping of thrips, *F. occidentalis*, feeding damage, (b) Bivariate GWAS of thrips feeding damage and its metabolomic prediction, (c) Univariate GWAS of the number of siliques after heat stress, (d) Bivariate GWAS of the number of siliques after heat stress and its metabolomic prediction. Left column represents the -log(p) of SNPs for effect on only the stress response. Right column represents the -log(p) of SNPs for a common effect on the stress response and its metabolomic prediction. Gray line indicates the Bonferroni threshold [$-$ log 10 (p) $\approx$ 7.13].

the feeding damage by thrips, *F. occidentalis*. This metabolite prediction correlated with thrips feeding damage with a Pearson correlation coefficient of 0.49. Among these metabolites, the aliphatic GSL glucoibarin and the GSL breakdown product 1-butene-4-isothiocyanate showed the highest correlation coefficients, that is, 0.08 and 0.12, respectively (Table S1). Three loci were significantly associated with thrips feeding damage: *MAM1* (*AT5G23010*), *AOP1* (*AT4G03070*), and *AOP3* (*AT4G03050*). In addition, polymorphisms near GS-OH (*AT2G25450*) were just above the threshold for significance (Figure 1b; Data S1). The *MAM* and *AOP* loci showed the highest $-$log10(p) values ranging from 9 to 11. Methylthioalkylmalate synthases (MAMs) determine the number of methionine side-chain elongations in aliphatic GSLs and play a central role in the diversification of aliphatic GSL structures in Arabidopsis and related plants (Kroymann et al., 2003). The 2-oxoglutarate-dependent dioxygenases AOP2 and AOP3 are known for the conversion of methylsulfinyl groups on the aliphatic GSL side chains, resulting in alkenyl or hydroxyalkyl GSLs, respectively (Halkier & Gershenzon, 2006; Kliebenstein, Gershenzon, & Mitchell-Olds, 2001; Kliebenstein, Kroymann, et al., 2001; Kliebenstein, Lambrix, et al., 2001). GS-OH can hydroxylate the alkenyl group that resulted from combined MAM1 and AOP2 activity (Kliebenstein, Gershenzon, & Mitchell-Olds, 2001; Kliebenstein, Kroymann, et al., 2001; Kliebenstein, Lambrix, et al., 2001).

Haplotype analysis based on the significant SNPs in the *MAM* region for thrips feeding damage showed the presence of two haplotype blocks. Block 1 contained six haplotypes and block 2 four haplotypes (Figure S1). Analysis of variance was performed on two contrasting haplotypes that occurred in more than 10% of the accessions. For both blocks, the Col-0 haplotype 2 had significantly more feeding damage than haplotype 1 ($P = 2.41 \times 10^{-4}$ and $P = 4.37 \times 10^{-3}$, respectively, Figure 2a,b; Table S2). The remaining haplotypes 3 to 6 were a combination of the two extreme haplotypes 1 and 2 (Figure S1b). Haplotype analysis of the significant SNPs in the *AOP* region revealed the presence of only one block, with two major haplotypes (Figure S2). Analysis of the variance of the inferred haplotypes, with correction for the population structure, revealed the presence of two contrasting haplotype groups in terms of feeding damage ($P = 7.83 \times 10^{-3}$, Table S2). The Col-0 *AOP* haplotype 2 exhibited significantly less feeding damage than haplotype 1 (Figure 2c). Overall, these results indicate that the two Col-0 haplotypes had contrasting effects on thrips, with a susceptible haplotype for *MAM* and a resistant haplotype for the *AOP* region.

AOP locus is associated with seed development after heat stress

For heat stress during flowering (Bac-Molenaar et al., 2015), we observed significant associations in both

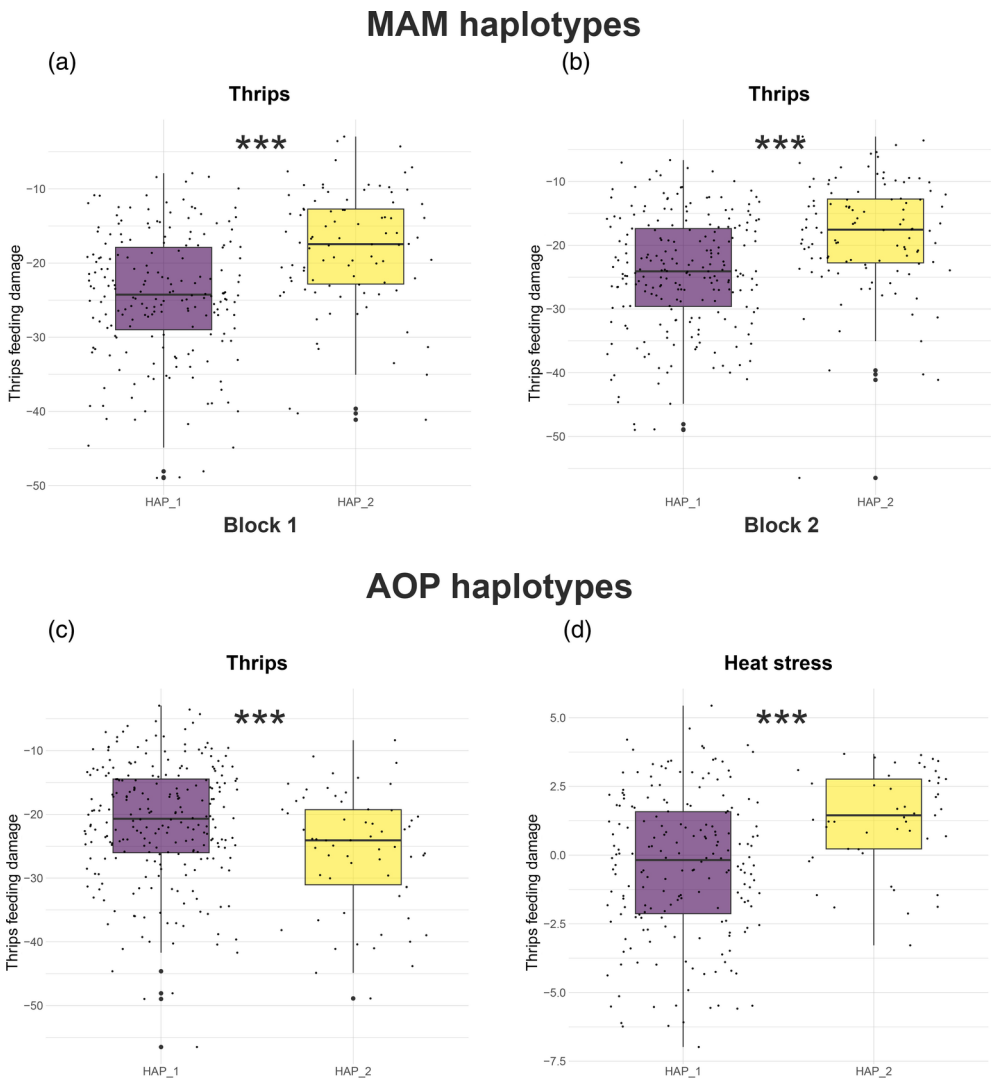

FIGURE 2. Feeding damage by *F. occidentalis* thrips and silique development during heat stress between inferred haplotypes constructed with significant SNPs. (a) Thrips feeding damage in relation to *MAM* haplotypes in block 1, and (b) in block 2. (c) Thrips feeding damage in relation to haplotypes of the *AOP* locus. (d) Number of siliques after heat stress in relation to *AOP* haplotypes. Phenotypes are displayed as best linear unbiased estimators (BLUEs). Inside boxplots, the bold line represents the median, box edges first and third quartile, and whiskers extend 1.5 × the interquartile distance (third quartile – first quartile) beyond the quartiles. Symbol *** is used to indicate $P < 0.001$ for test on the difference between haplotypes (Table S2).

univariate- and bivariate GWA mapping (Figure 1c,d). Twenty-four metabolites were selected by LASSO to build the model to predict silique development after heat stress with a Pearson correlation of 0.58. Among these metabolites, allyl-isothiocyanate, a breakdown product of aliphatic GSLs, had the highest coefficient (−0.17, Table S1). Heat stress revealed significant associations with the Arabidopsis *AOP* locus that is involved in GSL side-chain modifications. Several SNPs located in *AOP3* and several SNPs located in *AOP1* were significantly associated with silique formation after heat stress (−log10(p) value of 7.55 (Figure 1d; Data S1)). Haplotype analysis showed the presence of four haplotypes based on the significant SNPs located in *AOP1* and *AOP3* loci (Figure S2). Analysis of variance of the inferred haplotypes revealed the presence of two contrasting haplotype groups ($P = 5.30 \times 10^{-5}$), with the Col-0 haplotype 1 being associated with the formation of fewer siliques during heat stress (Figure 2d; Table S2). Other candidate genes presenting a significant common effect relating to heat stress included *AtNDX* (AT4G03090), which is involved in ABA signaling, a pathway involved in drought and heat stress responses (Wang et al., 2017), *BRI1 SUPPRESSOR 1* (AT4G03080), which interferes with the brassinosteroid (BR) signal-transduction pathway that regulates a wide range of developmental and physiological processes, including stress tolerance (Nolan et al., 2020)

TABLE 1 Contingency table with counts of individual Arabidopsis accessions per GSL chemotype

		MAM status		*AOP* status			*GS-OH* functionality		Aliphatic GSL chemotypes				
		C3	C4	Alk	MSO	OH	F	NF	3MSO	3OHP	4MSO	Allyl	But	OH-But
MAM haplotypes	HAP1	76	5	43	10	28	16	65	7	28	3	41	1	1
	HAP2	1	35	14	21	1	16	20	0	1	21	0	0	14
	Pearson's Chi-squared test													
	X-squared			31.663			8.0719		8.897					
	P-value			2.242e-06			0.01767		2.2e-16					
AOP haplotypes	HAP1	41	36	53	23	1	30	47	3	1	20	37	1	15
	HAP2	22	3	1	4	20	1	24	1	20	3	1	0	0
	Pearson's Chi-squared test													
	X-squared	16.717		71.254			15.614		78.147					
	P-value	0.0002343		1.234e-14			0.000407		1.157e-12					

Results of independence tests between the two most frequent haplotypes of *AOP* and *MAM* loci and GSL chemotypes are displayed in the two lower rows for each locus.
Alk, alkenyl; Allyl, allyl-glucosinolate; But, 3-butenyl; C3, carbon side chain length 3; C4, carbon side chain length 4; F, functional; MSO, methylsulfinyl; NF, nonfunctional; OH, hydroxyl; OH-But, 2S-OH-3-Butenyl; 3MSO, 3-methylsulfinylpropyl; 3OHP, 3-hydroxypropyl; 4MSO, 4-methylsulfinylbutyl.

and stomatal aperture (Kim et al., 2018), and *ABP1* (AT4G02980), which is involved in a broad range of growth and developmental processes and responses to heat stress (Gelová et al., 2021).

Haplotypes for thrips feeding link to GSL chemotypes

The natural diversity in aliphatic GSLs in Arabidopsis is largely determined by the presence or absence of functional *MAM*, *AOP*, and *GS-OH* loci. With the GSL classification system from Katz et al. (2021) we could match 102 accessions to a GSL chemotype, and test whether these chemotypes were related to a particular haplotype for heat stress and thrips feeding damage.

As expected, the independence test revealed strong and significant associations between *MAM* haplotypes and carbon side chain length and *AOP* haplotypes and alkenyl (Alk) or hydroxy-alkyl groups (OH, Table 1). We also observed reciprocal effects of *AOP* on carbon chain length (the *MAM* status) and effects of *MAM* on methylsulfinyl modifications (the *AOP* status), indicative of long-range LD between the *MAM* and *AOP* loci putatively caused by co-selection (Katz et al., 2021). The *MAM* haplotype 1 is associated with C3 chemotypes (MAM2 functionality), haplotype 2 with C4 chemotypes (MAM1 functionality), while the *AOP* haplotype 1 is associated with alkenyl (AOP2 functionality) and methylsulfinyl (no AOP functionality) side chains, while the *AOP* haplotype 2 associated with hydroxy-alkyl side chains (AOP3 functionality). Moreover, a strong association was revealed between these haplotypes and the functionality of the other aliphatic GSL modification gene, *GS-OH* (Table 1). From this, we conclude that haplotypes associated with reduced thrips herbivory (Figure 2) involved chemotypes with C3 methylsulfinyls (via MAM2 in combination with no AOP functionality) and 3-hydroxypropyl GSLs (3OHP, via functional MAM2 and AOP3).

Accessions with 3OHP dominance suffer less from thrips

The involvement of AOP in responses to herbivory was also found when comparing chemotypes at the accession level. Accessions with *AOP2* suffered from more feeding damage than accessions with *AOP3* or non-functional *AOP*s ($P = 0.012$, Figure 3b–e and Table S3). Functional *GS-OH*, which is indirectly associated with a functional *AOP2*, also increased herbivory ($P = 0.005$). In contrast to the haplotype-based approach, however, there was no effect of *MAM* status ($P = 0.153$), implying that methylsulfinyl chain length on itself was not relevant for thrips. However, accessions with the combination of *MAM2* and *AOP3* (3OHP dominant), experienced less feeding damage than accessions with *AOP2*-derived alkenyls, confirming the haplotype effect. For heat stress, chemotypes did not affect the level of stress resistance of accessions ($P = 0.075$, $P = 0.817$, $P = 0.55$ for *AOP*, *GS-OH*, and *MAM* status,

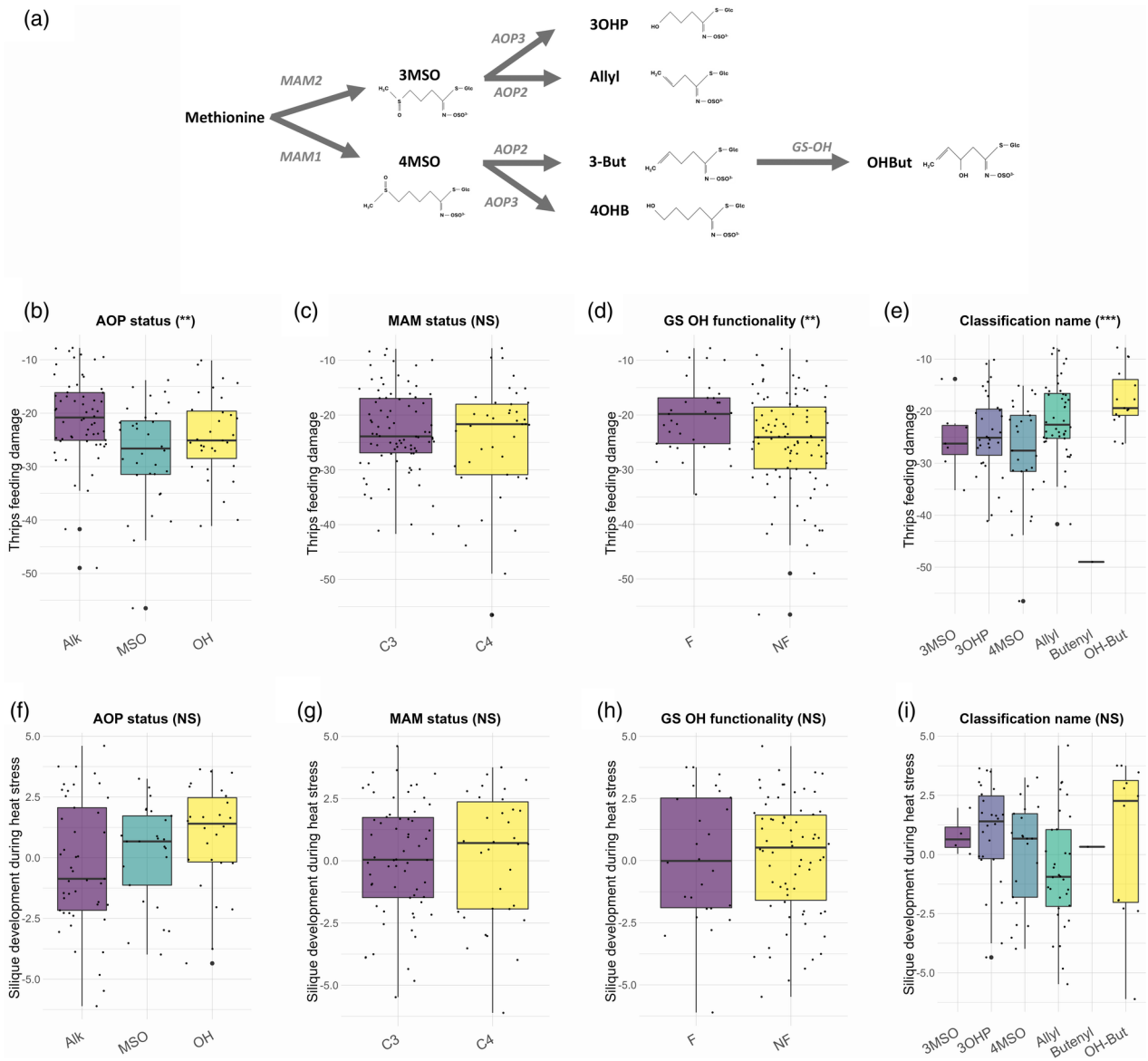

FIGURE 3. Aliphatic GSL chemotype effects on thrips and heat stress. (a) Involvement of MAM, AOP, and GS-OH enzymes in the biosynthesis of dominant GSL chemotypes (adapted from Katz et al. 2021). (b–e) Chemotype effects on thrips feeding damage, and (f–i) number of siliques after heat stress. Phenotypes are displayed as best linear unbiased estimators (BLUEs). Inside boxplots, the bold line represents the median, box edges represent the first and third quartile, and whiskers extend 1.5 × the interquartile distance (= third quartile − first quartile) beyond the quartiles. Indications of significance represent the probability of no difference between chemotype effects on the phenotype (***: $P < 0.001$, **: $P < 0.05$ and NS: $P > 0.05$, Alk, alkenyl (via AOP2); Allyl, allyl-glucosinolate; C3, carbon side chain length 3; C4, carbon side chain length 4; F, Functional; MSO, methylsulfinyl; NF, Non-Functional; OH, hydroxyl (via AOP3); OHBut, 2S-OH-3-Butenyl; 3-But, 3-butenyl; 3MSO, 3-methylsulfinylpropyl; 3OHP, 3-hydroxypropyl; 4MSO, 4-methylsulfinylbutyl; 4OHB, 4-Hydroxybutyl; Data S1; Table S3).

respectively, Figure 3f–i and Table S3). As the combination of *MAM1* and *AOP3* was rare (only 2 accessions), the effect of 4-hydroxybutyl GSL (4OHB) remained unclear.

Validation of GSL chemotype effects on thrips herbivory

To further elucidate the effects of aliphatic GSLs on thrips, several Arabidopsis transgenic mutants were screened for thrips damage. The *myc2myc3myc4* triple mutant, which is defective in JA signaling and GSL induction and highly susceptible to the generalist caterpillar *Spodoptera littoralis* (Schweizer et al., 2013) showed increased *F. occidentalis* feeding damage compared to the Col-0 wild type ($P = 0.002$, Figure 4a; Table S4), indicating that GSLs may be involved in Arabidopsis resistance to thrips. No effects of knocking out *MAM1* and *MAM3*, leading to more C3 and less long-chain GSLs, respectively (Textor et al., 2007), on thrips were observed in the Col-0 background ($P = 0.238$, Figure 4b; Table S4), confirming that variations in

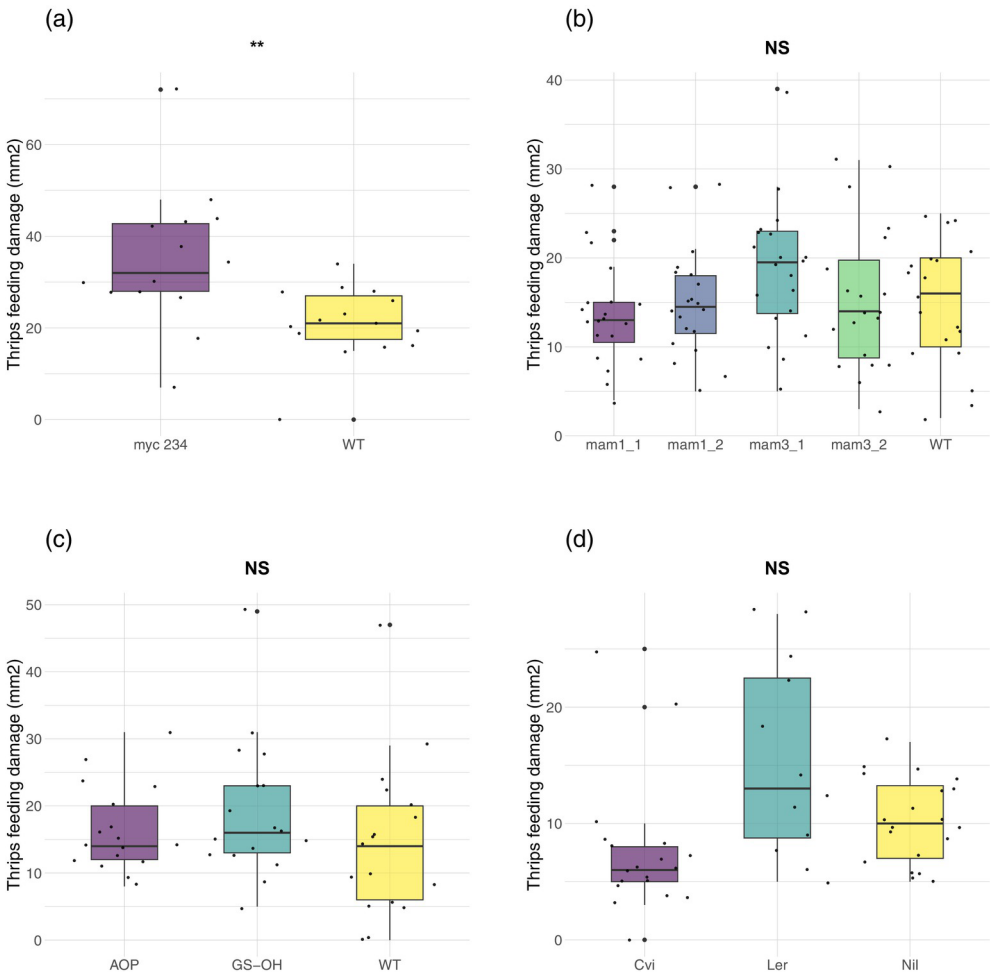

FIGURE 4. Validation of thrips feeding damage on plant lines with manipulated aliphatic GSL composition. *F. occidentalis* feeding damage on (a) *myc2myc3-myc4* triple mutants and the Col-0 wild type (*myc234*: $n = 14$, WT: $n = 15$), (b) Two *mam1* and two *mam3* knockout mutants in the Col-0 background (*mam1_1*: $n = 20$, *mam1_2*: $n = 20$, *mam3_1*: $n = 20$, *mam3_2*: $n = 20$, Col-0 WT: $n = 19$), (c) an *AOP2* overexpression line in Col-0, and an *AOP2* overexpression line in the background of a *gs-oh* knockout of Col-0 (*AOP*: $n = 17$, *GS-OH* knockout: $n = 17$, Col-0 WT: $n = 17$), and (d) Cvi, L*er* and the LCN5-2 near-isogenic line in the L*er* background (Cvi: $n = 19$, L*er*: $n = 12$, NIL: $n = 20$). Inside boxplots, the bold line represents the median, box edges represent the first and third quartile, and whiskers extend 1.5 × the interquartile distance (= third quartile − first quartile) beyond the quartiles. Indications of significance represent the probability of no difference between genotype effects on the phenotype using ANOVA (***: $P < 0.001$, **: $P < 0.05$ and NS: $P > 0.05$).

methylsulfinyl chain length did not affect *F. occidentalis*. In addition, no differences were observed in feeding damage when *AOP2* was overexpressed in the Col-0 background (expected to result in more 2-hydroxy-3-butenyl and 3-butenyl GSLs (Burow et al., 2015)), or when *AOP2* was overexpressed in a *gs-oh* knockout of Col-0 (expected to deliver only 3-Butenyl GSLs (Burow et al., 2015), $P = 0.3698$, Figure 4c; Table S4). To test if the AOP3 effect was dependent on MAM2 functionality, we used a near-isogenic introgression line (NIL) that allowed us to compare a 3OHP-dominant accession (L*er*, having *MAM2* and *AOP3*), with a 3-butenyl-dominant accession (Cvi, having *MAM1* and *AOP2*), and a NIL that was presumably 4OHB-dominant (*MAM1* and *AOP3* in L*er* background) (Keurentjes et al., 2007; Kliebenstein, Gershenzon, & Mitchell-Olds, 2001; Kliebenstein, Kroymann, et al., 2001; Kliebenstein, Lambrix, et al., 2001). According to trips feeding damage in the GWA population, L*er* was particularly susceptible in comparison to other 3OHP-dominant lines (Data S1). Assessment of feeding damage on the three lines confirmed the susceptibility of L*er* in comparison to Cvi, but no further increase in susceptibility was observed in the NIL due to the change from *MAM2* to *MAM1* ($P = 0.0025$, Figure 4d; Table S4). In summary, from the total population of 300 Arabidopsis lines, particularly accessions with 3OHP dominance suffered less from *F. occidentalis* infestations, suggesting that the combination of *MAM2* and *AOP3*, which results in the accumulation of C3-hydroxy-alkyl GLSs, reduces thrips infestations. Whether *MAM2* is required for this phenotype, or whether

4OHB-dominant accessions with the *MAM1* and *AOP3* combination show a comparable reduction in feeding damage, is still unclear due to the limited amount of accessions with this specific chemotype in our population ($n = 2$), and due to the potential influence of the susceptible L*er* background on our results.

DISCUSSION

Implications of aliphatic GSL recombination events

The associations of *AOP1*, *AOP3*, and/or *MAM1* with thrips feeding and heat stress represent both the wide biological importance of aliphatic GSLs in (a)biotic stress responses, as well as the major recombination events that have taken place in these loci. In Arabidopsis, the *MAM* locus contains up to three genes, *MAM1*, *MAM2*, and *MAM3*, with, depending on the specific accession, functionality of all or some of these genes. *MAM3* is responsible for the basic confirmation of the GSL side chains, *MAM2* is specifically involved in the synthesis of short (C3) side chains, while *MAM1* is required for C4 GSLs, with four methylene groups in the side chain (Textor et al., 2007). Similarly, the *AOP* locus contains different conformations of three *AOP* genes and is in LD with the *MAM* locus (Chan et al., 2010). *AOP2* and *AOP3* are known to be involved in the conversion of specific short-chain methylsulfinylalkyl GSL precursors to alkenyl GSLs and hydroxy-alkyl GSLs, respectively (Jensen et al., 2015; Kliebenstein, Gershenzon, & Mitchell-Olds, 2001; Kliebenstein, Kroymann, et al., 2001; Kliebenstein, Lambrix, et al., 2001). Noteworthy, the genotypic data used in our GWA mapping are based on the Col-0 genome, which lacks the *MAM2* gene and *AOP* functionality (Kroymann et al., 2003). Despite this, significant associations were found that can be considered as a hallmark for associations with the *MAM* and *AOP* genes in these two LD regions. Katz et al. (2021) reported that *MAM1* and *AOP2* plus *AOP3* manifest an epistatic interaction for GSL accumulation. The epistatic interaction is probably linked to the differences in *AOP2* and *AOP3* substrate availability that depend on the allelic state at the *MAM1* locus, which suggests feedback effects between these loci. This epistatic interaction should be taken into consideration when further validation analyses are conducted.

Aliphatic GSLs and their effects on thrips

The associations we observed between thrips feeding damage and constitutive levels of glucoibarin (a C7 GSL), 1-butene 4-isothiocyanate, and SNPs in the *MAM* and *AOP* loci align with previously observed negative correlations between GSLs and thrips feeding and detrimental effects of jasmonic acid-induced defense on the Western flower thrips (Abe et al., 2008; Thoen, 2016). As the *MAM* locus comprises three genes, *MAM1*, *MAM2*, and *MAM3*, that have been shown to be in strong LD in certain populations (Chan et al., 2010), the significant association of *MAM1* with thrips feeding damage could be considered a hallmark for natural variation presented in the *MAM* locus as a whole. Our results indicate that side chain length (*MAM1* or *MAM2*) on itself does not affect Western flower thrips. Also, *AOP2* functionality, leading to alkenyl GSLs, did not seem to suppress thrips herbivory, indicating that, for example, sinigrin (allyl) and gluconapin (3-butenyl) do not affect thrips. Nevertheless, oxidative cleavage of methylsulfinyl side chains by AOP3 was related to reduced feeding damage. Within the tested Arabidopsis population we observed that accessions holding a functional *MAM2* and *AOP3* gene, thereby producing 3OHP consisting of a C3 side chain with a hydroxy-alkyl group, suffered less from the Western flower thrips than other accessions. The fact that we could not confirm this in the L*er* background raises the question of whether the combined MAM2-AOP3 effect is merely an artifact of the specific population structure. Although our analyses did correct for kinship, we cannot completely exclude this option. Simultaneously, it is also well known that individual accessions differ in numerous other traits, including metabolites that were not considered by the applied GCMS and LCMS approaches, such as highly polar and apolar compounds, that could have influenced the eventual outcome for plant resistance. As the L*er* ecotype was more susceptible to thrips compared to other 3OHP-dominant accessions, unknown susceptibility factors may have affected our observations.

Aliphatic GSLs and their effects on heat stress

In our study, we found strong significant associations between heat stress in reproductive organs and allyl-isothiocyanate and SNPs in the AOP locus. The connection between aliphatic GSLs and reproductive organs is not new, as high hydroxy-alkyl GSL pools are found in Arabidopsis seeds and flowers (Brown et al., 2003; Jensen et al., 2015; Kliebenstein, Gershenzon, & Mitchell-Olds, 2001; Kliebenstein, Kroymann, et al., 2001; Kliebenstein, Lambrix, et al., 2001). In addition, aliphatic GSLs are known to be upregulated by heat stress (Guo et al., 2016; Ludwig-Müller et al., 2000) and previous studies reported that the administration of isothiocyanates mitigates growth inhibition after heat stress and induces transcriptional reprogramming of heat stress responses in *A. thaliana* (Hara et al., 2013; Kissen et al., 2016). Our metabolite prediction revealed a significant effect of the electrophylic allyl-isothiocyanates on seed embryo recovery after heat stress, which may be related to oxidative stress responses.

Integration of metabolome profiles in GWA studies

To maximize their fitness in a dynamic environment, plants show high plasticity and adapt their secondary metabolite profiles accordingly (Mertens et al., 2021). Although stress response and metabolome response are intertwined, there

are only few papers, such as Wang et al. (2015), that try to model jointly metabolite and downstream phenotypic trait variation in relation to DNA variation. Such integrated modeling may provide deeper insight into the genetic drivers of plant defenses. With a two-stage approach relying on work by Arouisse et al. (2021) and van Heerwaarden et al. (2015), we first reduced the metabolic dimensionality by penalized regression of the stress response on the full metabolome to select the most relevant metabolites. For convenient interpretation of our metabolomic predictions, we have chosen to use a form of penalized linear regression in this paper, although it may be advantageous to use nonlinear methods, such as random forest regression (Costa-Neto et al., 2021; Pérez-Rodríguez et al., 2012). Subsequently, the predicted values of this regression served as input to bivariate GWA mapping together with the target stress trait itself. The purpose of this bivariate approach is to magnify the genetic signal of the metabolomic stress response. Our results revealed new associations between plant responses to (a)biotic stresses and genes that would otherwise not have been detected. In particular, we focused on the associations between aliphatic GSL biosynthesis genes and resistance to thrips and heat stress.

Considerations on bivariate GWA mapping

Van Heerwaarden et al. (2015) showed that the power to detect a QTL for a target trait in a single trait GWA analysis can be increased by adding a correlated trait in a bivariate GWA analysis, where that correlated trait is a prediction for the target trait from a set of environmental covariates that represented selection forces. Here, we used a similar strategy, but now with the correlated trait being a prediction for the target trait from the plant metabolome. Therefore, unlike in Van Heerwaarden et al. (2015), our correlated trait did not represent a selection pressure, but rather was hypothesized to be part of the causal, multigenic architecture of the (a)biotic stress response, which was our target trait. We chose to use the full set of accessions to perform the bivariate GWA analysis. This is the usual approach in plant genetics because the number of genotypes is typically small, although it may lead to some overfitting. Splitting up the full set of accessions in a training and validation set can be an effective way of avoiding overfitting, but such a strategy is sensible only when the training set is large enough to attain sufficient power. Therefore, we chose to accept the risk of some overfitting to increase our QTL detection power, following the suggestion made by van Heerwaarden et al. (2015). Given the metabolomic prediction of the stress response (M) and the stress response of interest itself (Y), the bivariate GWA mapping focused on QTLs affecting both M and Y. One may question whether the detected QTLs affect both M and Y simultaneously or only one of the two traits, as it is known that a bivariate GWA mapping will also detect a QTL when a sufficiently strong signal is present for only one of the traits (Korte et al., 2012). First, even when in our bivariate GWA mapping a QTL is detected because of mainly M or only M, we still consider this QTL associated with the stress response, since the prediction M reflects an important part of the stress response. Second, we imposed a constraint on the QTL effects of M and Y that they should have the same allele substitution effect, which will lead to a somewhat conservative test. As an alternative, we could have tested for QTL effects that were unique to M and Y, which would have been more prone to overfitting. In addition, the haplotype analyses showed significant effects of the identified loci on Y. In case the metabolome would mediate the QTL effects to the stress response (Haplotype $\rightarrow$ M $\rightarrow$ Y), QTLs will be more significant in univariate GWA mappings on M alone (Zhou & Stephens, 2014). In case of a pleiotropic effect of the haplotype on both M and Y (M $\leftarrow$ Haplotype $\rightarrow$ Y), the bivariate GWA mapping is expected to be more powerful. As we did not know the full causal chain between genes, metabolites, and stress responses, but hypothesized that QTLs would have pleiotropic effects on both M and Y (the latter scenario), we decided to use a bivariate GWA mapping approach on M and Y to gain power in QTL detection. Simultaneously, we tried to reduce the risk of overfitting by using a constrained QTL model that required the allele substitution effect for M and Y to be the same.

Genetic associations beyond experimental and tissue-specific constraints

Although the metabolomic features used in this work were captured in different environmental conditions as well as in different experimental setups than the phenotypic stress responses, we were able to detect significant associations between several genes that orchestrate metabolomic compositions and biotic and abiotic stress responses. The bivariate GWA mapping identified substantially more significant QTLs than the univariate approach using only Arabidopsis responses to stress. Among the candidate genes were several loci involved in aliphatic GSLs and trehalose. Glucosinolates are typically found in all parts of Arabidopsis plants with variable concentrations depending on the developmental stages and types of plant tissues. Generally, the highest levels of GSLs are present in young leaves and reproductive tissues, such as siliques and seeds, whereas the GSL content was lower in mature leaves, which constituted the main part of the biological material used for our secondary metabolite dataset (Martínez-Ballesta et al., 2013; Porter et al., 1991). Even with this limitation, the bivariate GWA mapping revealed associations between seed-related effects of heat stress and GSL composition in unchallenged leaves. This is not surprising in view of the correlation between GSL content of leaves and seeds (Kliebenstein, Gershenzon, & Mitchell-Olds, 2001;

Kliebenstein, Kroymann, et al., 2001; Kliebenstein, Lambrix, et al., 2001), and illustrates the relevance of local constitutive metabolite profiles for predicting induced responses in systemic tissues. In conclusion, this research presents a study case how multivariate trait data can be integrated into a quantitative genetics and QTL mapping framework that can inspire future omics studies.

EXPERIMENTAL PROCEDURES

Plant material and phenotypic data

The 300 *Arabidopsis thaliana* accessions from the HapMap population were originally obtained from the ABRC Stock Center (Baxter et al., 2010), and previously phenotyped for responses to 23 single biotic or abiotic stresses and their combinations by Thoen et al., (2017) (see Data S1 for an overview of the stress descriptions). Abiotic stresses included salt, drought, osmotic, and heat stress. Biotic stresses involved the necrotrophic fungus *Botrytis cinerea*, the insect herbivores Western flower thrips, *Frankliniella occidentalis*, *Pieris rapae* caterpillars, *Myzus persicae* aphids, *Aleyrodes proletella* whiteflies, the root knot nematode *Meloidogyne incognita*, and the parasitic plant *Phelipanche ramosa*.

Arabidopsis *mam-1* (SALK_116223C and SALK_086935C) and *mam-3* (SALK_004536C and SALK_007222C) knock-out mutants were obtained from the Nottingham Arabidopsis Stock Centre (NASC), and tested for homozygosity of the T-DNA insert with PCR. Seeds from the *myc234* line (Schweizer et al., 2013) were kindly provided by Philip Reymond (Lausanne University), seeds from the *aop2* and *aop2/gs-oh* lines were kindly provided by Daniel Kliebenstein (University of California, Davis) (Jensen et al., 2015). The NIL LCN5-2 with the L*er* background and Cvi introgression around the *MAM* locus was established by Keurentjes et al. (2007). All plants were grown in climate-controlled chambers at 23°C, 70% RH, 100 μmol m^{-2} sec^{-1} light intensity, and a 8 h:16 h L:D photoperiod. Western flower thrips, *Frankliniella occidentalis* (Pergande), were reared in glass bottles on green common bean pods (*Phaseolus vulgaris*). To keep the offspring synchronized, 200 adult females were transferred to bottles with fresh bean pods twice a week. In all experiments, L1 or L2 thrips juveniles of approximately 5 days old were used. One mature leaf per five-week-old Arabidopsis plant was cut and inserted with its petiole in a film of 1% agar (Oxoid™ Agar Technical, Thermo Fisher) in a 5-cm diameter Petri dish. Each Petri dish was infested with three juvenile thrips, and the amount of feeding damage was determined after 6 days. Each accession was screened for 20 biological replicates.

Metabolite data

Secondary metabolites for 300 natural accessions of the Arabidopsis HapMap panel were measured by Kooke et al. (2021). In summary, metabolites were analyzed in unchallenged 4-week-old rosettes grown under short-day conditions (8 h:16 h L:D photoperiod). Two biological replicates of each accession, each consisting of a pool of three plants, were harvested in bulk and analyzed with (I) UPLC-Orbitrap-FTMS in order to profile semi-polar compounds including phenylpropanoids, flavonoids, GSLs etcetera, and (II) headspace SPME-GC-MS in order to profile volatile organic compounds. In total, 567 non-volatile and 603 volatile compounds were detected, respectively. To avoid false-positive associations, metabolites with missing data (non-detects) for both biological replicates of one or more accessions were discarded, resulting in a final set of 383 metabolites, of which the average of the two replicates per accession was subsequently computed. When an accession had only one replicate missing, this single replicate was taken as considered as representative of the particular accession, and its data were included in the further statistical analysis. For most GSLs both the intact (sulfur-containing) molecule and one or more of its specific volatile breakdown products were detected (e.g., sinigrin with its corresponding allyl isothiocyanates, and glucoraphanin with its butene isothiocyanates). In addition, a non-volatile primary metabolite collection that comprised 26 amino acids, 23 organic acids, 17 sugars, and 28 other unidentified metabolites, as detected by GCMS analysis of derivatized polar extracts, was obtained from Wu et al., (2016). Here, metabolites were analyzed by GCMS after derivatization of polar extracts from unchallenged rosette material of 5-week-old Arabidopsis plants, with one biological replicate per accession. All metabolite data were subjected to a log transformation and standardized by subtracting the mean and dividing by the standard deviation.

Dimension reduction and metabolomic prediction

Due to the high dimensionality of the metabolomics dataset, we carried out a dimension reduction prior to GWA mapping. Following Arouisse et al., (2021) and van Heerwaarden et al., (2015), we performed penalized regression of each stress response on all available metabolites. For a better interpretability, we used the least absolute shrinkage and selection operator (LASSO) for this regression analysis. The LASSO prediction was implemented using the "glmnet" package in R (Friedman et al., 2010) using 10-fold cross-validation and default settings. In the second step, we combined each individual Arabidopsis response during one of the (a)biotic stresses (Y) with its metabolomic prediction (M) in a bivariate genome-wide association study.

Genome-wide association studies

GWA analysis for diverse stresses and their metabolomic prediction was conducted on 1 million imputed SNPs with a minor allele frequency above 5% across the HapMap accessions (Arouisse et al., 2020). The bivariate GWA mapping was carried out using the statgenQTLxT package (van Rossum, 2024) in an R software environment following the multi-trait mixed model of Korte et al. (2012):

$$\begin{pmatrix} Y \\ M \end{pmatrix} = \begin{pmatrix} 1_n\mu_Y \\ 1_n\mu_M \end{pmatrix} + \begin{pmatrix} x\beta_Y \\ x\beta_M \end{pmatrix} + \begin{pmatrix} G_Y \\ G_M \end{pmatrix} + \begin{pmatrix} E_Y \\ E_M \end{pmatrix}$$

Where Y and M are $n \times 1$ vector of observed phenotypic values for the stress response (best linear unbiased estimators (BLUEs), calculated by Thoen et al. (2017)) and its metabolomic prediction for n genotypes respectively. 1_n is the $n \times 1$ column vector of ones. μ_Y and μ_M are trait-specific intercepts. x is the $n \times 1$ vector of scores of the marker under study, β_Y and β_M are the marker effects on the two traits. The vector of genetic background effects G is Gaussian with zero mean and covariance $V_g \otimes K$, where V_g is a 2×2 genetic covariance matrix, and K is an $n \times n$ kinship matrix (marker-based kinship). Similarly, the residual errors E have a covariance $V_e \otimes I_n$, where V_e is a 2×2 residual covariance matrix and I_n, is the $n \times n$ diagonal matrix. To identify quantitative trait loci (QTLs) with an effect on both the phenotypic trait and its metabolomic prediction, we tested for a common effect, by fitting the reduced model $\beta_Y = \beta_M = \beta$ and testing if $\beta = 0$ as in van Heerwaarden et al. (2015). The motivation for this test is that each metabolomic response is by construction positively correlated with the corresponding phenotypic trait. QTLs were defined as loci

with genome-wide significance based on the Bonferroni threshold of $7.13 \approx -\log10(0.05/\text{number of tested SNPs})$, and a 20 kb LD window. The reported effect estimates are obtained from the general model, where β_1 and β_2 can be different. For all plant stress responses, the obtained estimates are very similar to the ones obtained from univariate GWA mapping. The common QTL effect size for each bivariate GWA mapping is documented in standardized units for both traits in Data S1, tables 3 to 25, column "Pheno_data.effect".

Haplotype and chemotype analyses

Haplotype analysis was conducted based on significant SNPs that were present in the candidate genes. The construction of strong linkage disequilibrium (LD) blocks and haplotypes was performed with Haploview software (Barrett et al., 2004) using the algorithm of Gabriel et al. (2002). To study the effect of the two most frequent haplotypes on plant responses under different stresses, we fitted the following mixed model on the subset of accessions that contained those haplotypes with ASReml version 4.0 (Butler et al., 2018):

$$y = \mu + x\beta + G + E; G \sim N(0, \sigma_A^2 \mathrm{K}), E \sim N(0, \sigma_E^2 I_n)$$

where y is the $n \times 1$ vector of observed phenotypic values for n genotypes assigned to one of the two most frequent haplotypes (Hap_1 and Hap_2), μ is the general intercept, x is the $n \times 1$ binary vector assigning genotypes to the two most frequent haplotypes, β is the effect of the reference haplotype, G and E are $n \times 1$ vectors of random genetic and residual effects, with corresponding variance components σ_A^2 and σ_E^2; K is a known $n \times n$ relatedness matrix and In is the $n \times n$ identity matrix. To compare the effect of the haplotypes on plant response under stress we tested the null hypothesis $\beta = 0$ with a Wald test, conform the ASReml procedure (Butler et al., 2018). Additionally, we performed independence tests to verify whether haplotypes are associated with specific GSL chemotypes. The effects of chemotypes and Arabidopsis mutants and their wild type on stress responses were compared with a one-way ANOVA.

ACKNOWLEDGMENTS

This research is supported by the Perspective Programme "Learning from Nature" of the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research (NWO) and is partly funded by the Ministry of Economic Affairs.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Data and scripts processed in this study are available as follows:

- BLUEs of the stress responses: Data S1 and Thoen et al. (2017).
- Secondary metabolites: Supplemental data of Kooke et al. (2021).
- Primary metabolites: Supporting information of Wu et al. (2016).
- Metabolomic predictions: Data S1.
- Haplotypes: Data S1.
- Chemotypes: Data S1 and supplementary file 1 of Katz et al. (2021).
- Scripts and source files: https://zenodo.org/doi/10.5281/zenodo.12658371.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Haplotype structure of the *MAM* locus.

Figure S2. Haplotype structure of the *AOP* locus.

Table S1. Candidate genes containing SNPs with –log10(P) score above Bonferroni threshold.

Table S2. *MAM* and *AOP* haplotype effects on thrips feeding damage and silique development after heat stress.

Table S3. GSL chemotype effects on thrips feeding damage and silique development after heat stress.

Table S4. Thrips feeding damage on Arabidopsis mutants and the Cvi-L*er* NIL line.

Data S1. Phenotypic data (BLUEs), bivariate GWA mapping outcomes, and haplotype and chemotype assignments and metabolomic predictions.

REFERENCES

Abe, H., Ohnishi, J., Narusaka, M., Seo, S., Narusaka, Y., Tsuda, S. *et al.* (2008) Function of jasmonate in response and tolerance of Arabidopsis to thrip feeding. *Plant & Cell Physiology*, **49**, 68–80.

Arouisse, B., Korte, A., van Eeuwijk, F. & Kruijer, W. (2020) Imputation of 3 million SNPs in the Arabidopsis regional mapping population. *The Plant Journal*, **102**, 872–882.

Arouisse, B., Theeuwen, T.P.J.M., van Eeuwijk, F.A. & Kruijer, W. (2021) Improving genomic prediction using high-dimensional secondary phenotypes. *Frontiers in Genetics*, **12**, 715.

Ashraf, M.A., Iqbal, M., Rasheed, R., Hussain, I., Riaz, M. & Arif, M.S. (2018) Chapter 8 - Environmental stress and secondary metabolites in plants: an overview. In: Ahmad, P., Ahanger, M.A., Singh, V.P., Tripathi, D.K., Alam, P. & Alyemeni, M.N. (Eds.) *Plant metabolites and regulation under environmental stress*. Academic Press, pp. 153–167.

Bac-Molenaar, J.A., Fradin, E.F., Becker, F.F.M., Rienstra, J.A., van der Schoot, J., Vreugdenhil, D. *et al.* (2015) Genome-wide association mapping of fertility reduction upon heat stress reveals developmental stage-specific QTLs in *Arabidopsis thaliana*. *Plant Cell*, **27**, 1857–1874.

Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. (2004) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E. *et al.* (2010) A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genetics*, **6**, 1–8.

Bones, A.M. & Rossiter, J.T. (1996) The myrosinase-glucosinolate system, its organisation and biochemistry. *Physiologia Plantarum*, **97**, 194–208.

Brown, P.D., Tokuhisa, J.G., Reichelt, M. & Gershenzon, J. (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry*, **62**, 471–481.

Burow, M., Atwell, S., Francisco, M., Kerwin, R.E., Halkier, B.A. & Kliebenstein, D.J. (2015) The Glucosinolate biosynthetic gene AOP2 mediates feed-back regulation of Jasmonic acid signaling in Arabidopsis. *Molecular Plant*, **8**, 1201–1212.

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J. & Thompson, R. (2018) ASReml-R Reference Manual Version 4.

Chan, E.K.F., Rowe, H.C. & Kliebenstein, D.J. (2010) Understanding the evolution of defense metabolites in *Arabidopsis thaliana* using genome-wide association mapping. *Genetics*, **185**, 991–1007.

Costa-Neto, G., Fritsche-Neto, R. & Crossa, J. (2021) Nonlinear kernels, dominance, and envirotyping data increase the accuracy of

genome-based prediction in multi-environment trials. *Heredity*, **126**, 92–106.

Degenhardt, J. (2008) Ecological Roles of Vegetative Terpene Volatiles. In: Schaller, A. (Ed.) *Induced Plant Resistance to Herbivory*. Dordrecht: Springer Netherlands, pp. 433–442.

Escobar-Bravo, R., Klinkhamer, P.G.L. & Leiss, K.A. (2017) Induction of Jasmonic acid-associated defenses by Thrips alters host suitability for conspecifics and correlates with increased Trichome densities in tomato. *Plant & Cell Physiology*, **58**, 622–634. Available from: https://doi.org/10.1093/pcp/pcx014

Friedman, J., Hastie, T. & Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1–22.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.

Gelová, Z., Gallei, M., Pernisová, M., Brunoud, G., Zhang, X., Glanc, M. *et al.* (2021) Developmental roles of auxin binding protein 1 in *Arabidopsis thaliana*. *Plant Science*, **303**, 110750.

Gimsing, A.L. & Kirkegaard, J.A. (2009) Glucosinolates and biofumigation: fate of glucosinolates and their hydrolysis products in soil. *Phytochemistry Reviews*, **8**, 299–310.

Gómez, J.M. & Zamora, R. (2002) Thorns as induced mechanical defense in a long-lived shrub (*Hormathophylla spinosa*, Cruciferae). *Ecology*, **83**, 885–890.

Grotewold, E. (2005) Plant metabolic diversity: a regulatory perspective. *Trends in Plant Science*, **10**, 57–62.

Guo, L., Yang, R., Zhou, Y. & Gu, Z. (2016) Heat and hypoxia stresses enhance the accumulation of aliphatic glucosinolates and sulforaphane in broccoli sprouts. *European Food Research and Technology*, **242**, 107–116.

Halkier, B.A. & Gershenzon, J. (2006) Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology*, **57**, 303–333.

Hara, M., Harazaki, A. & Tabata, K. (2013) Administration of isothiocyanates enhances heat tolerance in *Arabidopsis thaliana*. *Plant Growth Regulation*, **69**, 71–77.

Isah, T. (2019) Stress and defense responses in plant secondary metabolites production. *Biological Research*, **52**, 39.

Jensen, L., Jepsen, H., Halkier, B., Kliebenstein, D. & Burow, M. (2015) Natural variation in cross-talk between glucosinolates and onset of flowering in Arabidopsis. *Frontiers in Plant Science*, **6**, 697.

Kant, M.R., Jonckheere, W., Knegt, B., Lemos, F., Liu, J., Schimmel, B.C.J. *et al.* (2015) Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Annals of Botany*, **115**, 1015–1051.

Katz, E., Li, J.-J., Jaegle, B., Abrahams, S.R., Bagaza, C., Holden, S. *et al.* (2021) Genetic variation, environment and demography intersect to shape Arabidopsis defense metabolite variation across Europe M. C. Schuman and A. Korte, eds. *eLife*, **10**, e67784.

Keurentjes, J.J.B., Bentsink, L., Alonso-Blanco, C., Hanhart, C.J., Blankestijn-De Vries, H., Effgen, S. *et al.* (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics*, **175**, 891–905.

Keurentjes, J.J.B., Koornneef, M. & Vreugdenhil, D. (2008) Quantitative genetics in the age of omics. *Current Opinion in Plant Biology*, **11**, 123–128.

Kim, E.J., Lee, S.H., Park, C.H. & Kim, T.W. (2018) Functional role of BSL1 subcellular localization in Brassinosteroid signaling. *Journal Of Plant Biology*, **61**, 40–49.

Kissen, R., Øverby, A., Winge, P. & Bones, A.M. (2016) Allyl-isothiocyanate treatment induces a complex transcriptional reprogramming including heat stress, oxidative stress and plant defence responses in *Arabidopsis thaliana*. *BMC Genomics*, **17**, 1–26.

Kliebenstein, D.J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J. *et al.* (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. *Plant Physiology*, **126**, 811–825.

Kliebenstein, D.J., Lambrix, V.M., Reichelt, M., Gershenzon, J. & Mitchell-Olds, T. (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in arabidopsis. *Plant Cell*, **13**, 681–693.

Kliebenstein, D.J., Gershenzon, J. & Mitchell-Olds, T. (2001) Comparative quantitative trait loci mapping of aliphatic, Indolic and benzylic Glucosinolate production in. *Genetics*, **159**, 359–370.

Kooke, R., Kruijer, W., van Eekelen, H.D.L.M., Becker, F.F.M., Wehrens, R., Hall, R.D. *et al.* (2021) Local adaptation shapes metabolic diversity in the global population of *Arabidopsis thaliana*. *bioRxiv*. https://doi.org/10.1101/2021.09.13.460026

Korte, A., Vilhjálmsson, B.J., Segura, V., Platt, A., Long, Q. & Nordborg, M. (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*, **44**, 1066–1071.

Kroymann, J., Donnerhacke, S., Schnabelrauch, D. & Mitchell-Olds, T. (2003) Evolutionary dynamics of an Arabidopsis insect resistance quantitative trait locus. *Proceedings of the National Academy of Sciences*, **100**, 14587–14592.

Li, N., Euring, D., Cha, J.Y., Lin, Z., Lu, M., Huang, L.J. *et al.* (2021) Plant hormone-mediated regulation of heat tolerance in response to global climate change. *Frontiers in Plant Science*, **11**, 627969.

Ludwig-Müller, J., Krishna, P. & Forreiter, C. (2000) A glucosinolate mutant of Arabidopsis is thermosensitive and defective in cytosolic Hsp90 expression after heat stress. *Plant Physiology*, **123**, 949–958.

Martínez-Ballesta, M.D.C., Moreno, D.A. & Carvajal, M. (2013) The physiological importance of glucosinolates on plant response to abiotic stress in brassica. *International Journal of Molecular Sciences*, **14**, 11607–11625.

Mertens, D., Boege, K., Kessler, A., Koricheva, J., Thaler, J.S., Whiteman, N.K. *et al.* (2021) Predictability of biotic stress structures plant defence evolution. *Trends in Ecology & Evolution*, **36**, 444–456.

Neilson, E.H., Goodger, J.Q.D., Woodrow, I.E. & Møller, B.L. (2013) Plant chemical defense: At what cost? *Trends in Plant Science*, **18**, 250–258. Available from: https://doi.org/10.1016/j.tplants.2013.01.001

Nolan, T.M., Vukasinović, N., Liu, D., Russinova, E. & Yin, Y. (2020) Brassinosteroids: multidimensional regulators of plant growth, development, and stress responses. *Plant Cell*, **32**, 298–318.

Pang, Q., Chen, S., Li, L. & Yan, X. (2009) Characterization of glucosinolate—myrosinase system in developing salt cress *Thellungiella halophila*. *Physiologia Plantarum*, **136**, 1–9.

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J.M., Crossa, J., Manès, Y. & Dreisigacker, S. (2012) Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3*, **2**, 1595–1605.

Porter, A.J.R., Morton, A.M., Kiddle, G., Doughty, K.J. & Wallsgrove, R.M. (1991) Variation in the glucosinolate content of oilseed rape (*Brassica napus* L.) leaves. *The Annals of Applied Biology*, **118**, 461–467.

Porter, H.F. & O'Reilly, P.F. (2017) Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Scientific Reports*, **7**, 1–12.

Qualley, A.V. & Dudareva, N. (2008) Aromatic Volatiles and Their Involvement in Plant Defense. In: Schaller, A. (Ed.) *Induced Plant Resistance to Herbivory*. Dordrecht: Springer Netherlands, pp. 409–432.

Salehin, M., Li, B., Tang, M., Katz, E., Song, L., Ecker, J.R. *et al.* (2019) Auxin-sensitive aux/IAA proteins mediate drought tolerance in Arabidopsis by regulating glucosinolate levels. *Nature Communications*, **10**, 4021. Available from: https://doi.org/10.1038/s41467-019-12002-1

Schweizer, F., Fernández-Calvo, P., Zander, M., Diez-Diaz, M., Fonseca, S., Glauser, G. *et al.* (2013) Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3, and MYC4 regulate Glucosinolate biosynthesis, insect performance, and feeding behavior. *Plant Cell*, **25**, 3117–3132.

Singh, V., Louis, J., Ayre, B.G., Reese, J.C. & Shah, J. (2011) TREHALOSE PHOSPHATE SYNTHASE11-dependent trehalose metabolism promotes *Arabidopsis thaliana* defense against the phloem-feeding insect Myzus persicae. *The Plant Journal*, **67**, 94–104.

Steppuhn, A. & Baldwin, I.T. (2008) Induced Defenses and the Cost-Benefit Paradigm. In: Schaller, A. (Ed.) *Induced Plant Resistance to Herbivory*. Dordrecht: Springer Netherlands, pp. 61–83.

Textor, S., de Kraker, J.-W., Hause, B., Gershenzon, J. & Tokuhisa, J.G. (2007) MAM3 catalyzes the formation of all aliphatic Glucosinolate chain lengths in Arabidopsis. *Plant Physiology*, **144**, 60–71.

Thoen, M.P.M. (2016) *Host-plant resistance to western flower thrips in Arabidopsis*. Wageningen: Wageningen University.

Thoen, M.P.M., Davila Olivas, N.H., Kloth, K.J., Coolen, S., Huang, P.P., Aarts, M.G.M. *et al.* (2017) Genetic architecture of plant stress resistance:

multi-trait genome-wide association mapping. *The New Phytologist*, **213**, 1346–1362.

Tibbs Cortes, L., **Zhang, Z.** & **Yu, J.** (2021) Status and prospects of genome-wide association studies in plants. *Plant Genome*, **14**, e20077. Available from: https://doi.org/10.1002/tpg2.20077

van der Fits, L. & **Memelink, J.** (2000) ORCA3, a Jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. *Science*, **289**, 295–297. Available from: https://doi.org/10.1126/science.289.5477.295

van Heerwaarden, J., **van Zanten, M.** & **Kruijer, W.** (2015) Genome-wide association analysis of adaptation using environmentally predicted traits. *PLoS Genetics*, **11**, 1–23.

van Rossum, B. (2024) statgenGxE: Genotype by Environment (GxE) Analysis. Available at: https://github.com/Biometris/statgenGxE/

Wang, H., **Paulo, J.**, **Kruijer, W.**, **Boer, M.**, **Jansen, H.**, **Tikunov, Y.** ***et al.*** (2015) Genotype–phenotype modeling considering intermediate level of biological variation: a case study involving sensory traits, metabolites and QTLs in ripe tomatoes. *Molecular BioSystems*, **11**, 3101–3110.

Wang, X., **Zhuang, L.**, **Shi, Y.** & **Huang, B.** (2017) Up-regulation of HSFA2c and HSPs by ABA contributing to improved heat tolerance in tall fescue and arabidopsis. *International Journal of Molecular Sciences*, **18**, 18.

War, A.R., **Paulraj, M.G.**, **Ahmad, T.**, **Buhroo, A.A.**, **Hussain, B.**, **Ignacimuthu, S.** ***et al.*** (2012) Mechanisms of plant defense against insect herbivores. *Plant Signaling & Behavior*, **7**, 1306–1320.

Wu, S., **Alseekh, S.**, **Cuadros-Inostroza, Á.**, **Fusari, C.M.**, **Mutwil, M.**, **Kooke, R.** ***et al.*** (2016) Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*. *PLoS Genetics*, **12**, 1–31.

Zhou, X. & **Stephens, M.** (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, **11**, 407–409.