






















# Hello darkness, my old friend: 3-KETOACYL-COENZYME A SYNTHASE4 is a branch point in the regulation of triacylglycerol synthesis in *Arabidopsis thaliana*

Urszula Luzarowska <sup>1</sup>, Anne-Kathrin Ruß <sup>2</sup>, Jérôme Joubès <sup>3</sup>, Marguerite Batsale <sup>3</sup>,  
Jędrzej Szymański <sup>4,5</sup>, Venkatesh P. Thirumalaikumar <sup>2</sup>, Marcin Luzarowski <sup>2</sup>, Si Wu <sup>6</sup>,  
Feng Zhu <sup>2,7</sup>, Niklas Endres <sup>2</sup>, Sarah Khedhayir <sup>2</sup>, Julia Schumacher <sup>8</sup>, Weronika Jasinska <sup>1</sup>,  
Ke Xu <sup>2</sup>, Sandra Marcela Correa Cordoba <sup>1</sup>, Simy Weil <sup>1</sup>, Aleksandra Skirycz <sup>2</sup>, Alisdair  
Robert Fernie <sup>2</sup>, Yonghua Li-Beisson <sup>9</sup>, Corina M. Fusari <sup>10,\*†</sup> and Yariv Brotman <sup>1,\*†</sup>

- 1 Department of Life Sciences, Ben Gurion University of the Negev, 8410501 Beer-Sheva, Israel
- 2 Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany
- 3 Laboratoire de Biogenèse Membranaire, UMR 5200, CNRS, University Bordeaux, F-33140 Villenave d'Ornon, France
- 4 Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research, OT Gatersleben, 06466 Seeland, Germany
- 5 IBG-4 Bioinformatics, Forschungszentrum Jülich, 52428 Jülich, Germany
- 6 Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA
- 7 National R&D Center for Citrus Preservation, Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, Wuhan 430070, P.R. China
- 8 Department for Plant Cell and Molecular Biology, Institute for Biology, Humboldt-Universität zu Berlin, Philipstraße 13, 10115 Berlin, Germany
- 9 CEA, CNRS, BIAM, Institute de Biosciences et Biotechnologies Aix-Marseille, Aix Marseille Univ., F-13108 Saint Paul-Lez-Durance, France
- 10 Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET-UNR), Suipacha 570, S2000LRJ Rosario, Argentina

\*Author for correspondence: brotmany@post.bgu.ac.il (Y.B.), fusari@cefobi-conicet.gov.ar (C.M.F.)

†Senior authors.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plcell>) is: Corina M. Fusari and Yariv Brotman.

## Abstract

Plant lipids are important as alternative sources of carbon and energy when sugars or starch are limited. Here, we applied combined heat and darkness or extended darkness to a panel of ~300 *Arabidopsis* (*Arabidopsis thaliana*) accessions to study lipid remodeling under carbon starvation. Natural allelic variation at 3-KETOACYL-COENZYME A SYNTHASE4 (*KCS4*), a gene encoding an enzyme involved in very long chain fatty acid (VLCFA) synthesis, underlies the differential accumulation of polyunsaturated triacylglycerols (puTAGs) under stress. Ectopic expression of *KCS4* in yeast and plants proved that *KCS4* is a functional enzyme localized in the endoplasmic reticulum with specificity for C22 and C24 saturated acyl-CoA. Allelic mutants and transient over-expression in planta revealed the differential role of *KCS4* alleles in VLCFA synthesis and leaf wax coverage, puTAG accumulation, and biomass. Moreover, the region harboring *KCS4* is under high selective pressure and allelic variation at *KCS4* correlates with environmental parameters from the locales of *Arabidopsis* accessions. Our results provide evidence that *KCS4* plays a decisive role in the subsequent fate of fatty acids released from chloroplast membrane lipids under carbon starvation. This work sheds light on both plant response mechanisms and the evolutionary events shaping the lipidome under carbon starvation.

## IN A NUTSHELL

**Background:** In the light, plants use photosynthesis to produce sugars and grow, while in the dark, they mainly use starch as a source of energy for maintenance throughout the night. When dark conditions last for a long time, plants consume all the available starch and enter a starvation mode, where they start feeding on lipids to support their metabolism. Although metabolism affects many responses, its genetic regulation is still underexplored. The study of complex traits such as metabolism can be facilitated by combining the use of natural populations and metabolomics. This strategy, called metabolomic-GWAS (genome-wide association studies), enables the identification of genes involved in the variation of multiple metabolic features under diverse environmental set-ups.

**Question:** How is metabolism controlled upon stress? What are the genes acting behind a metabolic response to environmental fluctuations? How did plants evolve to cope with adverse environmental conditions?

**Findings:** We were interested in determining the genetic regulation behind the adjustment in lipid metabolism in Arabidopsis after the sudden onset of abiotic stresses (focusing on heat and darkness). The interesting finding of our work is that KCS4, an enzyme involved in the elongation of fatty acids and wax synthesis, determines the differential accumulation of polyunsaturated triacylglycerols (puTAGs). By sequestering the saturated triacylglycerols into wax, KCS4 leaves free puTAGs to accumulate in lipid bodies and be used as an alternative source of energy.

**Next steps:** This approach could further be used to explore the fine-tuning of lipid metabolism in diverse environments and in other plant species.

## Introduction

Plant metabolism profoundly influences or even coordinates signaling, physiological, and defense responses. When the prevailing environment is adverse, biosynthesis, concentration, transport, and storage of primary and secondary metabolites are affected in a stress-dependent manner. These metabolic responses to abiotic stress involve fine adjustments in carbohydrate, amino acid, amine, and lipid pathways. Proper activation of early metabolic responses helps cells restore chemical and energetic imbalances imposed by stress and is crucial to acclimation and survival (Fraire-Velázquez and Balderas-Hernández 2013).

Like many other complex traits, plant metabolism shows extended variation within natural populations and therefore has been widely investigated using genome-wide association studies (GWAS). Genetic factors involved in variation for key metabolic traits have been successfully identified and characterized (i.e. primary and secondary metabolites and enzyme activities) (Wu et al. 2016, 2018; Fusari et al. 2017). A number of GWAS regarding lipid metabolism have been reported and resulted in the identification of new enzymes and regulators in species like Arabidopsis (*Arabidopsis thaliana*), maize (*Zea mays*), coffee (*Coffea arabica*), rice (*Oryza sativa*), and canola (*Brassica napus*) (Riedelsheimer et al. 2013; Gacek et al. 2017; Sant’Ana et al. 2018; Li et al. 2019; Hong et al. 2022). These studies focused on lipid levels under optimal growth conditions. However, the accumulation of lipids is strongly affected by fluctuations in environmental conditions, such as light, temperature, and nutrient availability (Burgos et al. 2011; Caldana et al. 2011; Szymanski et al. 2014). A classic example involves compensation for decreased membrane fluidity under cold stress (Welti et al. 2002). Moreover, adaptation to a high temperature has been associated with an increase in the

relative content of digalactosyldiacylglycerol (DGDG) and the ratio of DGDG to monogalactosyldiacylglycerol (MGDG), as well as a moderate increase in fatty acyl saturation of plant lipids (Han 2016).

In plants, the major source of energy in the dark is starch, which is degraded linearly according to the duration of the previous night (Graf and Smith 2011). Therefore, plants subjected to a sudden extended night (dark stress) experience carbon starvation due to limited starch availability. Under these conditions, polyunsaturated fatty acids (PUFA) released from membrane lipids are used as an alternative substrate for respiration via  $\beta$ -oxidation. These released fatty acids are often sequestered into triacylglycerols (TAGs) to avoid the adverse effects of the detergent-like properties of free fatty acids (FFAs). Therefore, it follows that TAG synthesis and lipid droplet formation sometimes constitute a cellular response to lipotoxicity. This sequestration is especially true for PUFA species because they contain multiple double bonds and are liable to produce reactive oxygen species (ROS) (Kao et al. 2018). It has been shown that the process of FFA sequestration into TAGs protects the cell from ROS generated during PUFA peroxidation (Fan et al. 2017).

Major changes in lipidomic composition under stress are accompanied by a remarkable remodeling of gene expression, both at the level of whole metabolic pathways and also involving activation of specific stress-induced isozymes (Szymanski et al. 2014). This work suggests that only by causing a perturbation of the metabolic homeostasis is GWAS likely to capture genes that evolved to stand at critical branch points in the regulation of the lipidome under naturally fluctuating environments.

Here, we performed GWAS using lipidomic data obtained in 4 different environments: control, combined heat and

darkness, and 2 different extended darkness conditions. The locus harboring *KCS4*, encoding 3-KETOACYL-COENZYME A SYNTHASE4 (Joubès et al. 2008; Kim et al. 2021), was strongly associated with a high number of polyunsaturated TAGs (puTAGs) in stress conditions but was not detected in control condition. Thus, *KCS4* association is driven by carbon starvation. Using mutants for alternative *KCS4* alleles and ectopic expression in yeast and plants, we explored the biochemical properties of *KCS4* and validated its role in lipid metabolism. We place a model for *KCS4* enzymatic and regulatory function, where *KCS4* acts as a branch point in the fate of FAs, directing saturated FA to the very long chain fatty acid pathway (VLCFA) and facilitating the pool of polyunsaturated FA in the accumulation of puTAGs. Two *KCS4* alleles have the differential capacities to process saturated FA. Accessions carrying the strong allele show higher levels of VLCFA-derived cuticular waxes and higher levels of puTAGs. In fully exploiting the genetic data available for Arabidopsis, this work sheds light on the evolutionary events shaping the lipidome and further highlights the importance of lipid remodeling in plant response to changing environments.

## Results

### Plasticity in Arabidopsis lipid metabolism under different environmental conditions

To investigate the genetic regulation of lipid biosynthesis in Arabidopsis, we conducted LC-MS-based lipidomic analysis of 301 and 309 Arabidopsis accessions grown in 2 environmental conditions: control (21/16 °C in 16/8-h light/dark cycle, termed CHD) and stress (heat 32 °C + darkness for 24 h, termed HD), respectively. HD was selected as it has previously been shown to have a strong effect on the lipidome (Caldana et al. 2011). The mapping population was phenotyped for 98 and 109 glycerolipids in CHD and HD, respectively (Supplemental Data Set S1).

Principal component analysis (PCA) of lipid data clearly separated the natural ecotypes into 2 different groups according to the 2 environmental conditions tested (Supplemental Fig. S1A). Furthermore, differences in the lipid status between CHD and HD samples showed that TAGs accumulated to higher levels while the levels of monogalactosyldiacylglycerols (MGDGs) and digalactosyldiacylglycerols (DGDGs) mostly decreased under stress (Supplemental Fig. S1B). In addition, there is substantial positive correlation between lipid levels from a given lipid class, suggesting coordinated regulation at the level of lipid classes (Supplemental Fig. S1C). Notably, TAGs showed the highest variation across accessions for both stress and control conditions, with many species falling in the range of >10-fold difference (Supplemental Fig. S1D).

### Genetic basis underlying Arabidopsis lipid metabolism

Lipid intensities from CHD and HD and the genotypic information from the 250 K SNPchip (Horton et al. 2012) were

used to perform 2 independent GWAS for the identification of lipid quantitative trait loci (QTL). GWAS revealed 93 and 544 SNP-trait associations in control and stress conditions, respectively, at a genome-wide significance level of  $\text{LOD} \geq 5.3$  (Supplemental Data Set S2). These SNP-trait associations were further grouped into 51 and 66 QTL for CHD and HD conditions, respectively. In total, only 5 QTL were shared between CHD and HD, revealing the great plasticity of the lipidome in response to different environmental conditions (Supplemental Data Set S3).

Regarding functional annotation of candidate genes, we mapped 12 and 22 genes putatively involved in lipid metabolism for control and stress conditions, respectively (Supplemental Data Set S3, Acyl-Lipid Metabolism database [Li-Beisson et al. 2010] <http://aralip.plantbiology.msu.edu/>) (Fig. 1). After filtering for robust associations colocalizing for 2 or more traits, 4 QTL harbored genes with annotated functions related to lipid metabolism (Fig. 1A). Two QTL in Chromosome 1, 1 harboring the *KCS4*(AT1G19440) gene and other including a *LIPASE CLASS 3* (AT1G06250), 2 *ACYL-COA OXIDASES* (*ACX3*: AT1G06290 and *ACX6*: AT1G06310) and 2 *FATTY ACID DESATURASE* (*CER17*: AT1G06350, *ADS4.2*: AT1G063609); 1 in Chromosome 3 harboring the *FATTY ACID DESATURASE* (*FAD2*: AT3G12120) and the last QTL in Chromosome 5, harboring a *LYSOPHOSPHOLIPASE* (AT5G20060) (Supplemental Data Set S3, Fig. 1A).

### *KCS4* is involved in fatty acid elongation (FAE) and has major impact on polyunsaturated TAG levels under heat and dark stress

Remarkably, 22 puTAGs colocalized to the QTL on Chromosome 1, harboring *KCS4* (Fig. 1A). This QTL was not found in control conditions, and the lead SNP in the locus (m11481) colocalized with  $\text{LOD} \geq 8$  (Bonferroni corrected *P*-value < 0.05) for 16 of the 22 associated TAGs (e.g. TAG 54:9, Fig. 1B; Supplemental Data Set S2). Besides *KCS4*, 33 other genes were mapped on the QTL, spanning 107,318 bp (Supplemental Data Set S3). However, with the exception of *KCS4*, none of the others showed any sequence similarity to proteins known to be involved in lipid metabolism.

*KCS4* belongs to a gene family involved in the elongation of VLCFA. Twenty-one KCS genes have been previously identified in the Arabidopsis genome. The KCS family encodes enzymes involved in the first and often described as rate-limiting step of FAE (Blacklock and Jaworski 2006; Joubès et al. 2008). *KCS4*'s closest homologues are *KCS9* (AT2G16280; 78% aa identity), involved in the elongation of C22 to C24 acyl-CoAs (Kim et al. 2013); *KCS17* (AT4G34510; 72% aa identity); and *KCS18* (AT4G34520; 64% aa identity), involved in elongating C18 acyl-CoAs to C20 and C22 in seeds and shown to drive the natural variation in the abundance of VLCFAs in this organ (Rossak et al. 2001). However, *KCS4* specificity and function have not been clearly characterized so far. TAGs associated with

the *KCS4* QTL were exclusively polyunsaturated (e.g. puTAG 52:6, 52:7, 54:6, 54:8, 54:9, 56:6, Supplemental Data Sets S2 and S3) and correlated positively with each other, grouping together in a single well-defined cluster (Fig. 1C).

To identify the alleles involved in puTAG variation under HD, we constructed haplotypes considering the 4 SNPs from *KCS4* scored in the 250 K SNPchip (m11502, m11503, m11504, m11505, Supplemental Data Set S2). The population was divided into 9 nonunique haplotypes, which were, therefore, further considered as being putatively involved in the QTL variation. Mean levels of puTAG 54:9 varied significantly among haplotypes (Fig. 1D). Furthermore, m11503 was identified as a nonsynonymous SNP (nsSNP). In the majority of accessions, nsSNP m11503 encodes for a methionine (hereafter *KCS4*[Met]) Accessions carrying this allele have higher levels of puTAG in the stress condition than accessions carrying the minor allele, which encodes for a threonine (hereafter *KCS4*[Thr]) (Fig. 1E). These results suggest that the nsSNP (m11503) in *KCS4* is the causal mutation underlying puTAG variation under HD conditions in *Arabidopsis*.

### *KCS4* associates with puTAGs in extended darkness

Thus far, we found that *KCS4* associated with puTAGs in a combined heat and darkness treatment. In prolonged darkness, starch is prematurely exhausted and plants are carbon-starved (Stitt and Zeeman 2012). Therefore, we determined whether *KCS4* also associates with puTAGs, when darkness alone is applied.

For this purpose, we utilized our own data sets: control (CD), 3 days (3D) extended darkness, and 6 days (6D) extended darkness (Supplemental Data Set S1) (Zhu et al. 2022). The GWAS at a genome-wide significance level of  $\text{LOD} \geq 5.3$  resulted in 230, 255, and 342 SNP-trait associations grouped into 73, 81, and 70 QTL for CD, 3D, and 6D, respectively (Supplemental Data Sets S2 and S3). The locus harboring *KCS4* was associated strongly in 3D and 6D with 8 and 13 puTAGs, respectively. Lead SNP and other features of the association matched with GWAS results from the HD experiment (Supplemental Fig. S2). Thus, darkness alone is enough to trigger the association of *KCS4* with puTAGs and could be related to the carbon status of the plant. Expression of *DIN1* (AT4G35770) and *BCAT2* (AT1G10070), 2 genes usually used as starvation markers (Morales et al. 2019) is upregulated under HD conditions (Supplemental Fig. S3), showing that HD is also triggering carbon starvation. In summary, an additive effect is observed when both stresses are applied together, determining a stronger association of *KCS4* to puTAGs under HD.

### Dissection of *KCS4* alleles involved in TAG variation, pattern of linkage disequilibrium (LD), and signatures of selection

We next mined the full sequence information available for 149 accessions from the GWAS population (<http://signal.salk.edu/atg1001/3.0/gebrowser.php>), for 80 kb around SNP m11481 on Chromosome 1. The sequence alignment (Supplemental File S1) was used to build an SNP matrix (Supplemental Data Set S4) to perform GWAS with more SNPs than those included in the previous experiments using the 250 K SNPchip (Horton et al. 2012). We analyzed a total of 1,082 SNPs and short insertion–deletions, in addition to the 131 SNPs previously tested in this region. These analyses aim to discard genetic heterogeneity among other genes in the locus, analyze the pattern of LD in the region, determine whether more SNPs in *KCS4* are additionally involved in the variation, and evaluate signatures of selection. The results are shown in Fig. 2, exemplified for TAG 54:9. According to the *KCS4* gene model (Fig. 2A), we identified 35 additional SNPs in the *KCS4* genomic region apart from the 4 SNPs scored in the 250 K SNPchip (19 in the promoter, 2 in the 5'UTR, 13 in the coding region, and 1 in the 3'UTR; Supplemental Data Set S5). Fewer TAGs with lower LOD values showed an association with *KCS4*, as expected when analyzing fewer accessions (<100) (Supplemental Data Set S5, Fig. 2B). The lead SNP using resequencing information was a polymorphism not tested before (rs21933), but in high LD with m11481 ( $r^2 = 0.73$ ,  $P < 0.001$ ). In addition, a number of polymorphisms in the upstream region of *KCS4*, not tested in the SNPchip, were identified as being significant, pointing to a putative combined regulation at the promoter region mediated by rs39451, rs39499, and m11502 and at the coding region mediated by the nsSNP m11503 (Fig. 2A). Surprisingly, several SNPs in the *KCS4* coding region did not associate (Fig. 2B).

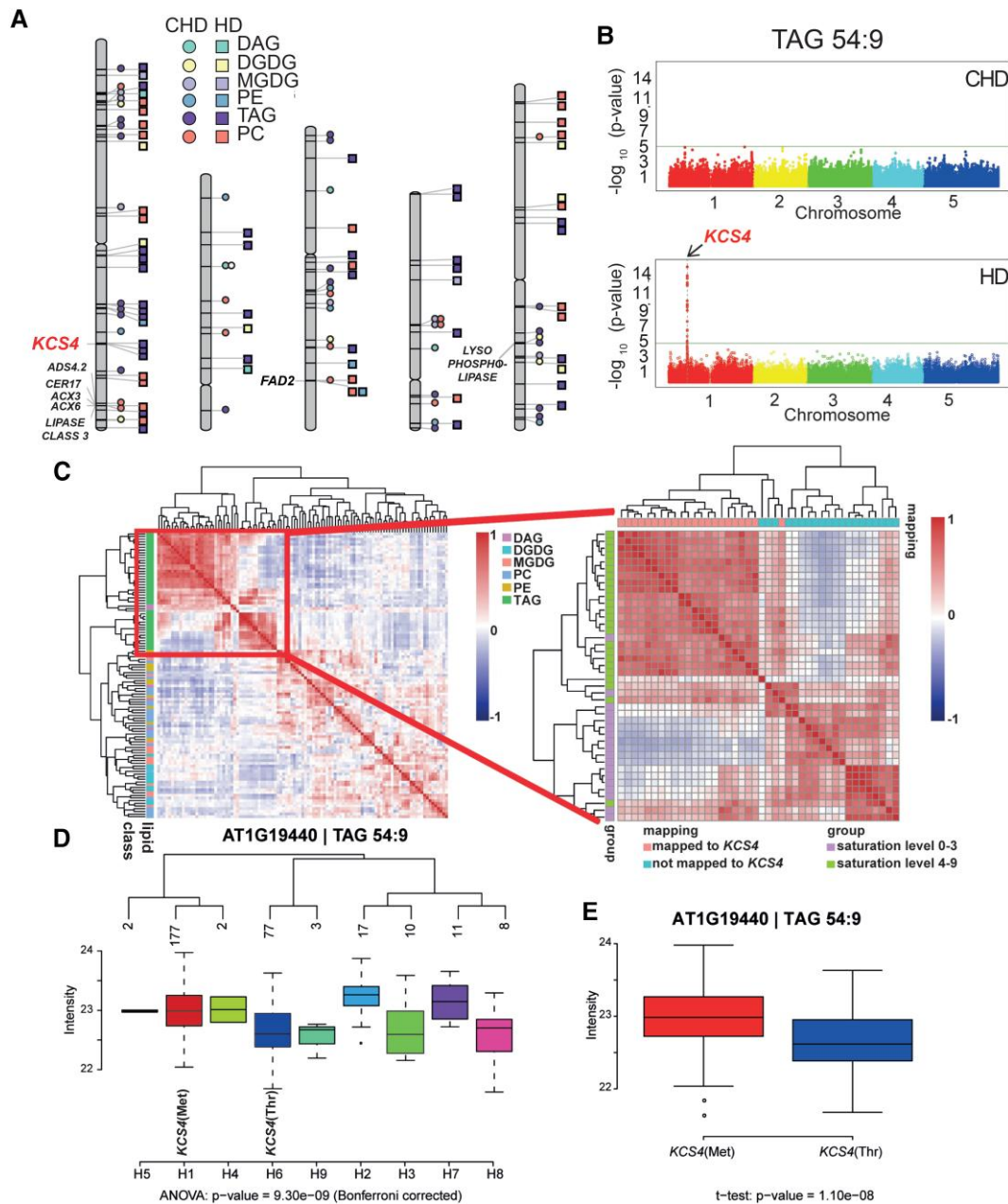
All SNPs in the locus harboring *KCS4* with high LOD values are in high LD with each other and in high LD with the lead SNPs rs21933 and m11481 ( $r^2 > 0.2$ ,  $P < 0.001$ , Fig. 2C). The remaining SNPs in the *KCS4* coding region are not in LD with the lead SNPs nor with the nsSNP m11503 in the *KCS4* coding region. This result suggests that these mutations may have appeared at a different time during evolution or they may have accumulated later, as a consequence of specific recombination, to increase allele frequency at other SNP positions. To evaluate signatures of selection, we tested Tajima's *D* for the 80 kb around m11481 (Fig. 2D). Tajima's *D* reached significantly negative values along the 80 kb, particularly around SNPs significantly associated with the QTL: rs21933, m11481, rs39451, rs39499, m11502, m11503. This result indicates that the region is under high selective pressure and supports the selection of *KCS4* as the candidate gene involved in puTAG variation.

Transcriptional regulation of *KCS4* in *Arabidopsis*

Several SNPs in the promoter region showed significant LD with the nsSNP m11503. Therefore, we focused on the transcriptional regulation of *KCS4* in *Arabidopsis* (Fig. 3). We analyzed 1 kb upstream of the transcription start site for known transcription factor (TF) binding sites using the DAP-Seq database (O'Malley et al. 2016). Several TF families belonging to land plants could likely bind to the *KCS4* promoter region,

### Transcriptional regulation of *KCS4* in *Arabidopsis*

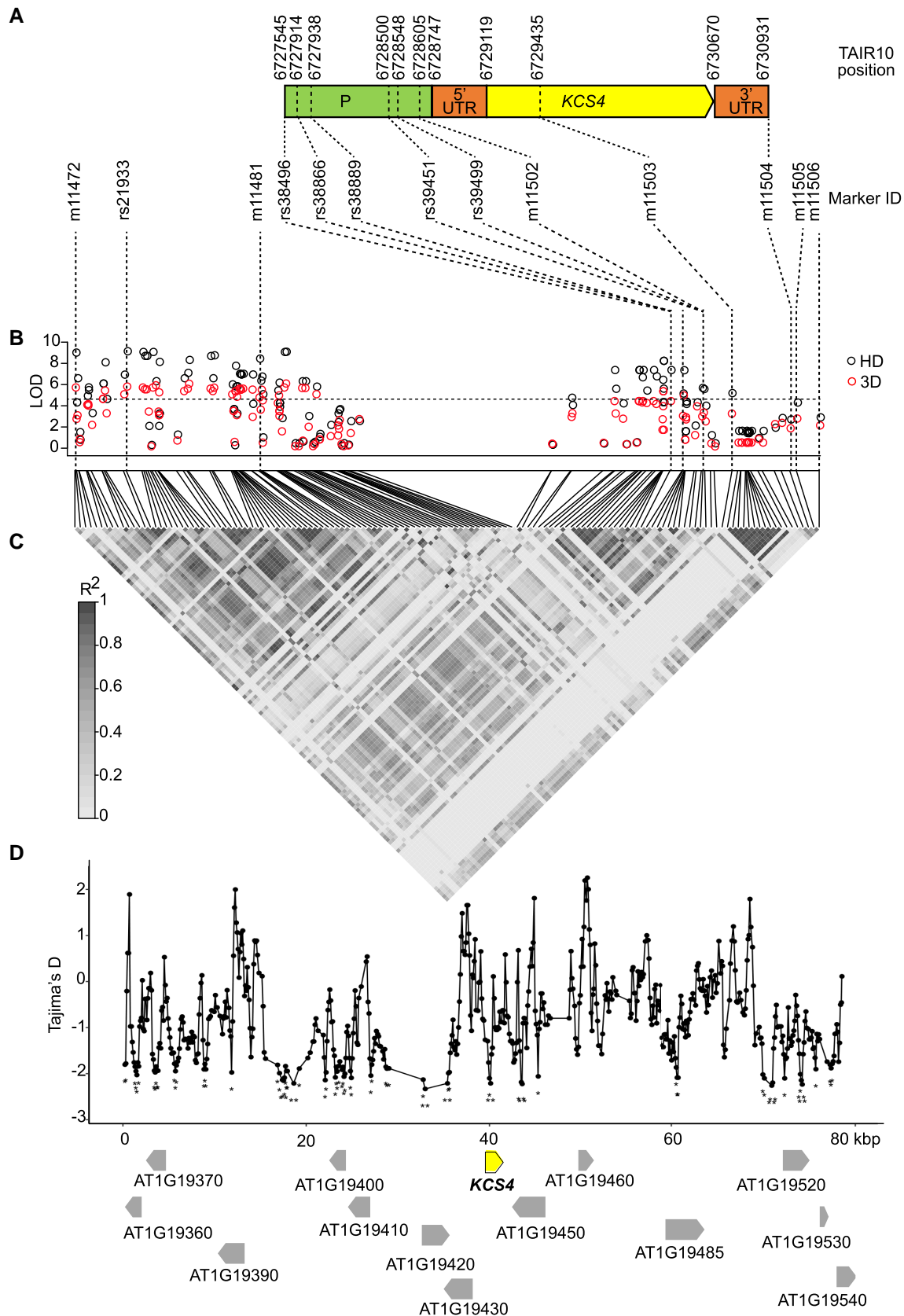
Several SNPs in the promoter region showed significant LD with the nsSNP m11503. Therefore, we focused on the transcriptional regulation of *KCS4* in *Arabidopsis* (Fig. 3). We analyzed 1 kb upstream of the transcription start site for known transcription factor (TF) binding sites using the DAP-Seq database (O'Malley et al. 2016). Several TF families belonging to land plants could likely bind to the *KCS4* promoter region,



**Figure 1.** Results from GWAS analysis on lipidomic data from plants grown under control (CHD) and stress (HD) conditions. **A**) Chromosome scheme and QTL identified in the 2 experimental conditions (CHD = circles, HD = squares), for different lipid classes: phosphatidylethanolamine (PE), phosphatidylcholine (lecithin) (PC), MGDG, DGDG, diacylglycerol (DAG), and triacylglycerol (TAG). Color/shape references are included in the figure. Gene names are included only for colocalizing QTL for 2 or more lipid species with LOD > 5.3. **B**) Manhattan plots obtained for TAG 54:9 in CHD (top) and in HD (bottom). The *KCS4* (AT1G19440) QTL was found at Chromosome 1 only under HD condition. **C**) Pearson correlations coefficients calculated between all lipid species across all accessions, for HD condition. Lipids are color-coded according to class (see reference at right) and clustered using Ward's distance on pairwise dissimilarity. A zoom-in of the TAG cluster shows the grouping pattern of TAGs according to the number of unsaturated bonds and to the GWAS result (left). **D**) Average trait value (intensity of TAG 54:9,  $\log_2$  scale) for the different *KCS4* haplotypes using SNPs m11502, m11503, m11504, and m11505. The 2 major haplotypes have the alternative alleles for m11503 (Met or Thr). **E**) TAG 54:9 is the average value for accessions carrying the *KCS4*(Met) (left) and *KCS4*(Thr) (right) alleles.

indicating that *KCS4* expression is controlled by several regulatory processes (Fig. 3A). We furthered our analysis by looking at the core binding sites of the highly enriched TF families that contained binding motifs across SNPs in the promoter region of *KCS4* (rs38889, rs39451, rs39499, and m11502).

The top-enriched DOF family (DNA-binding One zinc Finger) does not have a binding site colocalizing with polymorphisms but does in other regions of the promoter (Fig. 3B). Similarly, the NAC and WRKY TF families bind to nonpolymorphic regions in the *KCS4* promoter, and MYB



**Figure 2.** Dissection of *KCS4* alleles involved in TAG variation, pattern of LD, and signatures of selection. **A**) *KCS4* gene model depicting promoter region (P), 5'UTR, coding region (*KCS4*), and 3'UTR of *KCS4* (AT1G19440). Beginning, end and relevant SNP positions are (continued)

TF binds near an SNP but in the coding region. Interestingly, heat shock factor (HSF), the second highly enriched TF family, has a core binding motif (heat shock element: 5' nnGnAnnTnCtn 3') within SNP m11502 in the promoter region.

We performed an electrophoretic mobility shift assay (EMSA) to assess the physical interaction of HSFA2 (AT2G26150) with the *KCS4* promoter sequence. Additionally, we tested if the alleles (C and T) at SNP m11502 influence the binding. HSFA2, a well-studied HSF family member, has been shown to bind to the heat shock element in *Arabidopsis* (Lämke et al. 2016). As shown in Fig. 3C, HSFA2-GST fusion protein bound to the 5'-DY682-labeled *KCS4* promoter fragment in vitro with the C allele at m11502, resulting in a retardation band on the gel (Fig. 3C, Lane 2) compared to the free probe (Lane 1). Interaction between HSFA2 and the *KCS4* promoter region containing the binding site appears to be specific, as the addition of an excessive unlabeled competitor (same sequence, 200× more) significantly reduced the binding to the labeled probe (Fig. 3C, Lane 3). Finally, we tested a competitor sequence having the alternative allele at m11502 (T). The latter reduced the binding compared to the labeled probe (Fig. 3C, Lane 6), suggesting that the T allele at m11502 promotes stronger binding of HSF2A to the *KCS4* promoter region. Overall, this indicates that polymorphisms at the promoter region could exert a dosage effect on TFs, influencing the expression of *KCS4*.

Furthermore, we analyzed natural variation in transcript levels of *KCS4* for the CHD and HD conditions in the complete GWAS panel (Fig. 3D). *KCS4* transcripts were significantly lower in HD compared to CHD. This correlates significantly negatively with highly unsaturated TAG levels ( $r^2 = -0.66$ ,  $P < 0.001$ ), which increased in HD (e.g. Fig. 3E, TAG 54:9). In HD, *KCS4* transcript levels were not significantly different for accessions carrying different alleles at m11502. This confirms that neither differences in TF binding affinity nor expression levels of *KCS4* determine the variation of puTAG levels under HD (Fig. 3E). In other words, polymorphisms at the *KCS4* promoter region are not causal for puTAG variation in *Arabidopsis* natural populations.

### Expression patterns of KCS family members under diverse environmental conditions

We analyzed the time-resolved response of *Arabidopsis* Col-0 accession towards light and temperature for 16 of the 21 gene members using data previously obtained (Caldana et al. 2011). Focusing on 24 h, most KCS members, including

*KCS4*, are upregulated under light conditions and downregulated in darkness, irrespective of the temperature. However, *KCS8* (AT2G15090), *KCS12* (AT2G28630), *KCS16* (AT4G34250), *KCS18*, and *KCS21* (AT5G49070) showed the opposite pattern. *KCS4* transcript levels were downregulated both under darkness and under HD, but transcript levels were much lower in HD, strengthening the idea of a synergistic effect between light and temperature stress in the *KCS4* response. In addition, *KCS5* (AT1G25450), *KCS9*, and *KCS20* (AT5G43760) showed a similar transcriptional response to *KCS4* across several environmental stimuli (Fig. 3F).

### Functional validation of the role of *KCS4* in puTAG variation

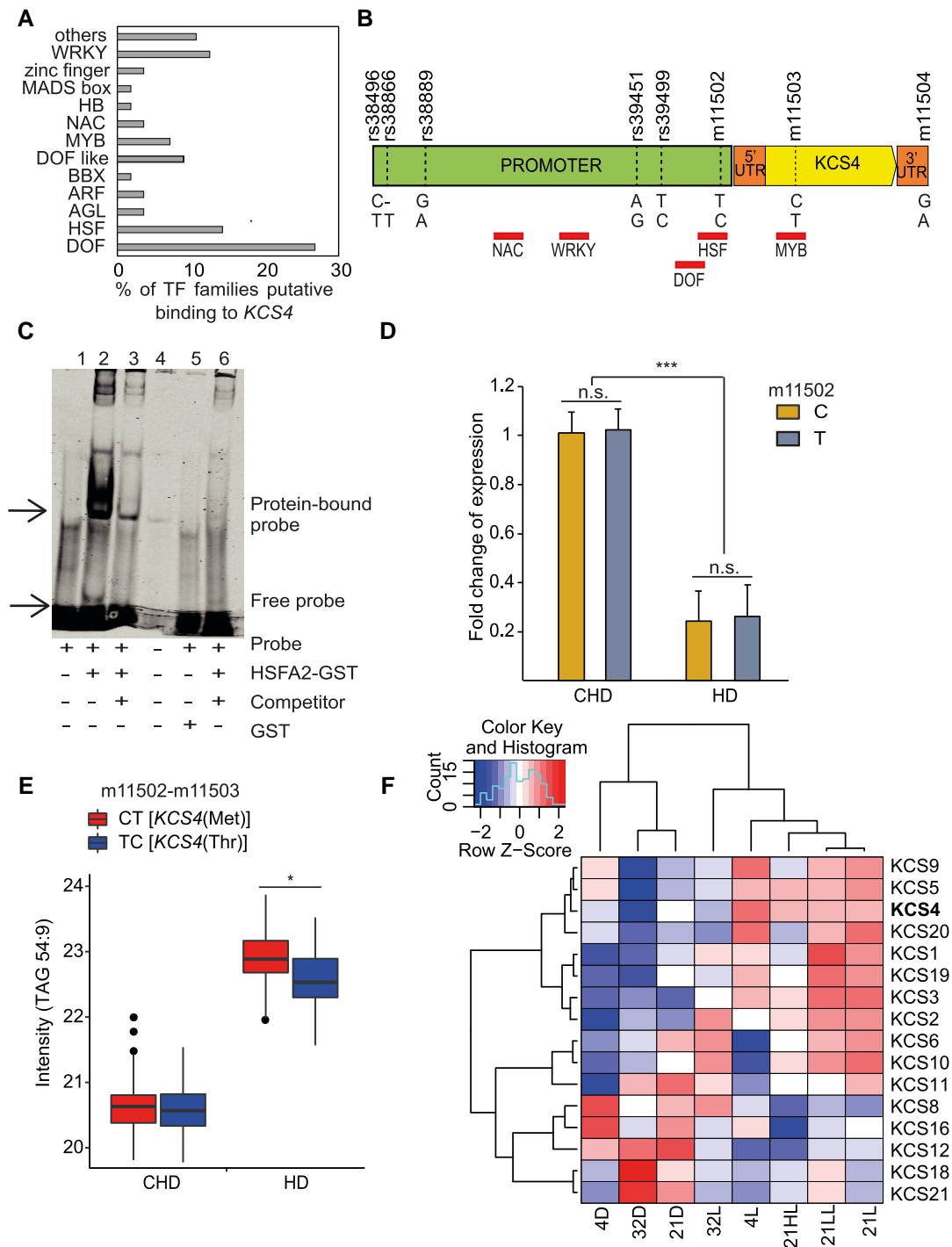
To validate the role of *KCS4* alleles in puTAG variation under HD stress, we selected 5 accessions: 3 carrying *KCS4*(Met) and 2 carrying *KCS4*(Thr). Selection of these accessions was based on puTAG levels measured in the HD experiment (Supplemental Data Set S1). Mutant lines (*kcs4*) were obtained through CRISPR-Cas9 by introducing a 57/58-bp deletion at position +18 from the ATG (Supplemental File S2). This deletion introduced a premature stop codon, resulting in a truncated protein of 16 aa compared to the wild-type version of 516 aa. Lines were grown under CHD or HD as described in materials and methods and harvested to perform lipidomic analysis. Since natural variation is detected at a population level, we compared the average values for *kcs4*(Met) and *kcs4*(Thr) allelic mutants against wild-type *KCS4*(Met) and *KCS4*(Thr) alleles (Fig. 4).

Under control conditions (CHD), puTAG levels (X:4 to X:9) were significantly different between *KCS4*(Met) and *kcs4*(Met) lines. Mutants for *kcs4*(Met) presented, on average, lower levels of puTAGs, similar to the levels found for *KCS4*(Thr) accessions (Fig. 4, A–C and G). This phenomenon was seen for the accessions pooled together as well as for individual accessions (Supplemental Data Set S6). On the other hand, puTAG levels did not differ between *KCS4*(Thr) and *kcs4*(Thr). These results support *KCS4*(Met) as the stronger allele and the causal polymorphism for polyunsaturated TAG variation.

Under HD condition, a higher accumulation of puTAGs was seen for accessions carrying *KCS4*(Met) allele (Fig. 4, D–G). However, overall levels of TAGs (saturated and polyunsaturated) increased independently of *KCS4* alleles. This is in agreement with the phenotype observed for the complete mapping panel. Here, mutants didn't differ in puTAG

#### Figure 2. (Continued)

indicated above (TAIR10) and below (SNP ID) gene structure. **B**) Manhattan plot of GWAS results (HD: black circles; 3D: red circles) obtained for TAG 54:9 using re-sequenced information for 149 accessions. The re-sequenced region on Chromosome 1 (positions 6689119 to 6769118, TAIR10) included 131 SNPs genotyped in the 250 K SNPchip plus 1,137 additional polymorphisms. Bonferroni threshold (LOD = 4.4) is indicated with a horizontal dashed line. **C**) LD is depicted as a heat map of the coefficient of correlation ( $r^2$ ). The LD block ( $0.8 > r^2 > 0.2$ ,  $P < 0.001$ ) includes the lead SNPs rs21933 and m11481, SNPs in *KCS4* promoter (rs39451, rs39499, m11502), in *KCS4* genomic region (m11503) and in *KCS4* 3' UTR (m11504). **D**) Tajima's *D* value in sliding windows for the genomic region analyzed in (B). Gene models and gene orientation are depicted with arrows, *KCS4* (in bold) is highlighted. Significantly negative Tajima's *D* values are marked with asterisks (\* $P < 0.05$ , \*\* $P < 0.01$ ).



**Figure 3.** KCS4 transcriptional regulation. **A**) Percentage of TFs classes predicted to bind KCS4 genomic region according to in silico analysis of binding motives using DAP-seq database (O'Malley et al. 2016). **B**) Position of the binding motifs (red rectangles, names below) for the main TF classes. SNP positions in the promoter and in coding region are marked with vertical dash-lines and IDs are stated above gene model. **C**) EMSA showing the interaction between HEAT SHOCK FACTOR A2 (HSFA2/AT2G26150) and the HSE binding motif (heat shock element: 5'-nnGnAnnTnTcn-3') found within the SNP m11502 in KCS4 promoter region; 1: labeled probe (5'-DY682-labeled double-stranded oligonucleotide) only; 2: labeled probe plus HSFA2-GST protein; 3: labeled probe, HSFA2-GST protein and 200× competitor DNA (unlabeled oligonucleotide containing HSE binding site with the "C" allele at m11502); 4: well spill; 5: labeled probe plus free GST; 6: labeled probe, HSFA2-GST protein and 200× competitor DNA (unlabeled oligonucleotide containing HSE binding site with the "T" allele at m11502). **D**) KCS4 expression for the complete GWAS population in CHD and HD conditions, grouped based on C or T allele at SNP m11502 in KCS4 promoter region. Expression (fold change) was measured by RT-qPCR. Significant differences between CHD and HD transcript levels and between C and T alleles are shown (*t*-test: \**P* < 0.05,

(continued)



levels from their wild-type counterparts, as if the massive production of puTAGs under stress overrides the effect of KCS4 mutation.

Transient expression of AtKCS4(Met) in *Nicotiana benthamiana* leaves showed significantly higher levels of puTAGs for classes 52, 54, and 56 compared to control plants (Fig. 4, H–J). Conversely, TAGs with none or a few unsaturated bonds tended to be higher in control plants. These results show that the transient expression of AtKCS4(Met) changes the TAG profile in *N. benthamiana*.

In summary, the reduction of puTAGs in the *kcs4*(Met) mutant (Fig. 4, A–C and G), together with the increase of puTAGs when KCS4(Met) was expressed in *N. benthamiana* (Fig. 4, H–J), confirms that KCS4(Met) is the causal polymorphism for polyunsaturated TAG variation and validates GWAS results.

### Functional validation of the role of KCS4 in the accumulation of leaf cuticular waxes

VLCFAs are produced by the elongation complex (i.e. KCS, KCR:  $\beta$ -KETOACYL REDUCTASE, HCD:  $\beta$ -HYDROXYACYL-COA DEHYDRATASE and ECR: ENOYL-COA REDUCTASE) in the endoplasmic reticulum (ER) and used as precursors for the synthesis of lipid barriers such as cuticular waxes. Therefore, we determined the leaf wax coverage for KCS4 wild-type alleles and mutants as a proxy of the enzyme activity in planta (Fig. 4, K–M). In general, lines carrying KCS4(Met) showed significantly higher content of major components of leaf waxes than lines carrying KCS4(Thr) in both CHD and HD conditions (Fig. 4M). Particularly for CHD condition, the mutant *kcs4*(Met) showed a significant reduction of the compounds ALK30 and ALC24 compared to the wild-type allele KCS4(Met). Other wax compounds such as ALK27, ALD30 + FA28, and ALD32 + FA30 showed the same tendency, although values were not significantly different from the wild-type KCS4(Met) allele (Fig. 4K). On the other hand, *kcs4*(Thr) did not show any differences from its wild-type counterpart. These results confirm that the wax coverage and composition are different for both alleles and that other members of the KCS family cannot fully complement the synthesis of waxes coming from FA elongated by KCS4.

### KCS4 enzymatic activity and cellular localization

KCS4 shows sequence similarity to the remaining 20 members of the Arabidopsis KCS multigene family (Joubès et al. 2008). As stated before, KCS enzymes catalyze the first rate-limiting step in the VLCFA elongation, and they determine the substrate specificity.

To evaluate the very-long-chain 3-ketoacyl-CoA synthase activity and substrate specificity of KCS4, we expressed AtKCS4(Met) in wild-type yeast and profiled VLCFAs (Fig. 5). The strain expressing AtKCS4(Met) synthesized significantly higher levels of VLCFA from C20 to C26 compared to the strain transformed with an empty vector (Fig. 5A).

We further expressed AtKCS4(Met) in the  $\Delta$ elo3 yeast strain, lacking the endogenous ELO3 elongase (Fig. 5B). The  $\Delta$ elo3 strain accumulates C20 and C22 VLCFA (Han et al. 2002). Expression of KCS4 led to the elongation of VLCFAs from C22 to C24 and from C24 to C26 (Fig. 5B). We also analyzed the heterologous expression of AtKCS9 as it is the closest KCS4 homolog (Fig. 5B). KCS9 elongates VLCFAs from C22 to C24 (Kim et al. 2013 and Fig. 5B). These results suggest that KCS4 has differential activity compared to KCS9, pushing further elongation of VLCFAs up to C26.

Transient expression of AtKCS4(Met) in *N. benthamiana* leaves produced a significant decrease of C20 and C22 VLCFAs and a significant increase of VLCFAs from C26 up to C30 compared to the control (Fig. 5C). The elongation of VLCFAs up to C30 might be the result of the action of the native elongase complex in *N. benthamiana* due to the increased availability of C26 when KCS4 is transiently expressed. AtKCS4(Met)-YFP was found to label a polygonal and tubular network characteristic of the ER (Fig. 5D), in agreement with previous results shown for other KCS family members (Joubès et al. 2008).

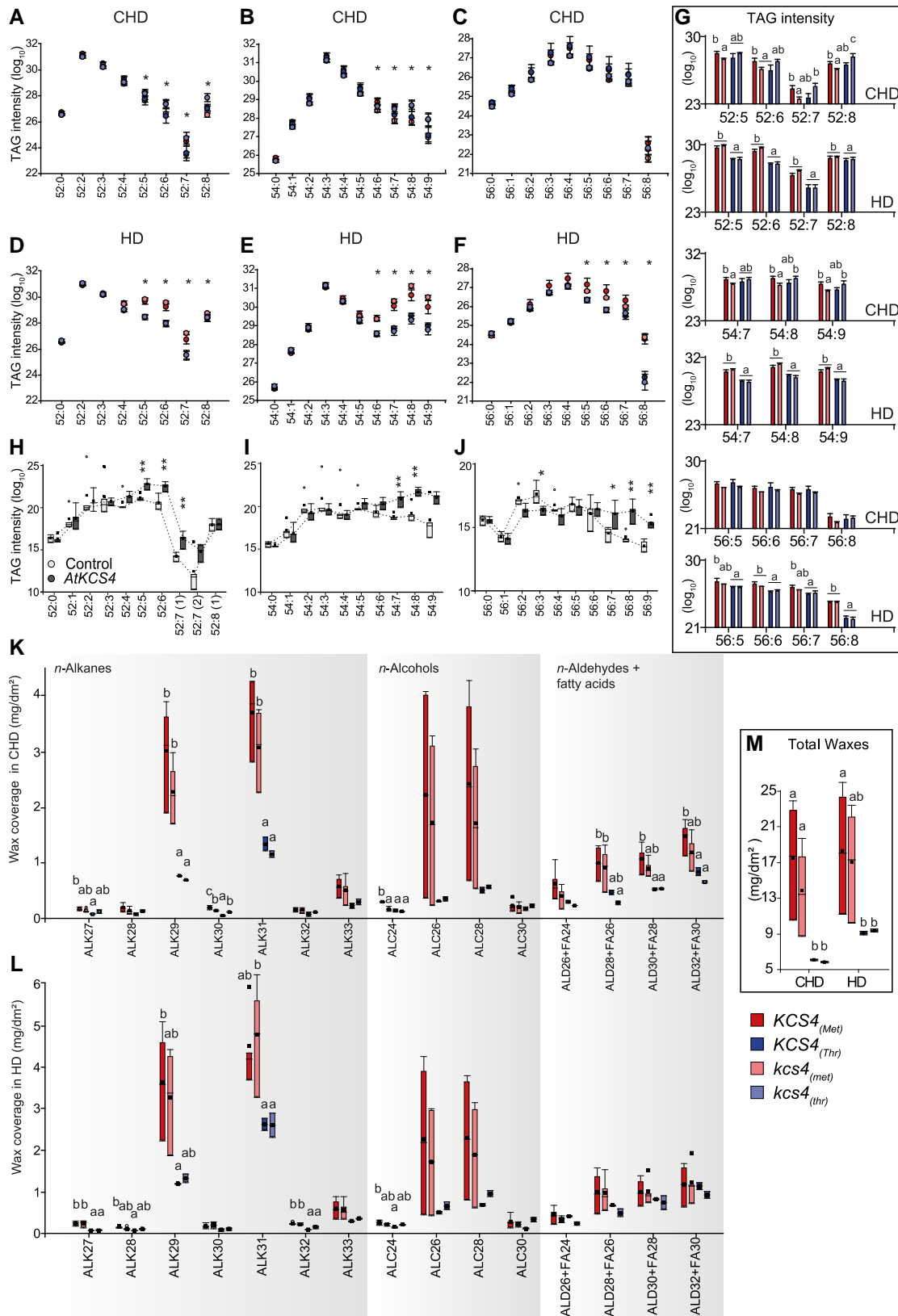
We further transformed  $\Delta$ elo3 strain with KCS4 wild-type and mutant alleles (Supplemental Fig. S4). Expression of both KCS4(Met) and KCS4(Thr) alleles led to the elongation of C22 into C24 up to C26 VLCFA in the  $\Delta$ elo3 strain with slight differences in the amount of C22 and 2OH:C24 (Supplemental Fig. S4A). On the other hand, yeast expressing allelic mutants did not show exactly the same composition as the control line, C26 did not accumulate, but C22 and C24 VLCFA were still produced at a low level compared to yeast expressing wild-type alleles as if some activity remained. This residual accumulation of VLCFA in the mutants indicates that *kcs4*(Met) and *kcs4*(Thr) lines still have some enzymatic activity, although very weak.

In addition, we performed a time-course experiment at 48, 72, and 96 h for  $\Delta$ elo3 strain expressing KCS4(Met) and KCS4(Thr) wild-type alleles (Supplemental Fig. S4B). Yeasts expressing KCS4(Met) ended up with more C26 and 2OH:C26 than yeasts expressing KCS4(Thr), supporting the hypothesis of a differential enzymatic activity between KCS4 alleles.

KCS1 (AT1G01120) expression in yeast leads to the production of C18, C20 and C22 VLCFA (Trenkamp et al. 2004; Tresch et al. 2012), KCS2 (AT1G04220) can elongate

#### Figure 3. (Continued)

\*\* $P < 0.001$ , \*\*\* $P < 0.0001$ , n.s. not significant). Data represent means  $\pm$  sd. For CHD  $n = 193$  for the C allele and  $n = 94$  for the T allele. For HD  $n = 163$  for the C allele  $n = 75$  for the T allele. **E**) Average trait value (TAG 54:9 intensity, in  $\log_2$ ) for the 2 major haplotypes determined by SNPs m11502 (promoter) and m11503 (coding region). CT and TC haplotypes encodes for KCS4(Met) and KCS4(Thr), respectively. Significant differences between haplotypes ( $t$ -test: \* $P < 0.05$ ). **F**) Heatmap of expression pattern for 16 members of the KCS multigene family clustered using complete-average method. Transcript abundance was measured 24 h after treatment. Data used in F was obtained previously (Caldana et al. 2011).



**Figure 4.** KCS4 functional validation using allelic mutants in Arabidopsis and heterologous expression in *N. benthamiana*. **A–G)** Average TAG profiles for wild-type (WT) accessions and allelic mutants carrying the KCS4(Met) allele (WT: red; mutant: pink) and the KCS4(Thr) allele (WT: blue; mutant: light-blue) in CHD and HD conditions. Dot plots (**A–F**) and bars (**G**) represent means  $\pm$  SE,  $n = 14$  to 25. One-way ANOVA followed by

(continued)

from C20 to C22 (Millar et al. 1999; Franke et al. 2009), KCS20 mainly produces C22 and C24 VLCFA, and KCS5 and KCS6 (AT1G68530) mainly produce C24 to C28. All these proteins have different expression profiles under heat and dark stress, with some being upregulated and some being downregulated (Fig. 3F). Expression profiles of KCS4, KCS5, KCS9, and KCS12 are also altered in *kcs4* CRISPR-Cas lines, evidencing a delicate expression balance under stress, between KCS family proteins (Supplemental Fig. S5).

### Biomass is affected by KCS4 function

A mild reduction in VLCFA content improves cell proliferation and shoot growth (Nobusawa et al. 2013). To investigate if KCS4 has an effect on plant growth, we compared biomass of KCS4 wild-type accessions and mutant lines. We collected full rosettes from plants grown in CHD and subjected to HD, immediately after treatment. The *kcs4* mutants grew significantly bigger than their WT counterparts (Fig. 6). This indicates that biomass decreases with increasing KCS4 activity and further supports both the idea that mutations at KCS4 reduce the overall content of VLCFA and that KCS4 activity cannot be overcome by other family members.

### Phylogenetic signal of the lipidome and correlation with climatic and geographical conditions

To test if there is a correlation between lipid profiles and genetic distance (a proxy for phylogeny), *K* statistics and empirical *P*-values were calculated for all the measured lipid features in control (CHD) and stress (HD) conditions (Blomberg et al. 2003). The results indicate a correlation between genetic distance and the similarity of lipidomic profiles. Lipids exhibiting a significant phylogenetic signal in CHD are enriched in MGDGs ( $P \leq 0.001$ ), while in HD, we observed enrichment in TAGs ( $P \leq 0.05$ ). Five of the 6 TAGs with a significant phylogenetic signal are polyunsaturated species associated with the KCS4 QTL (Supplemental Fig. S6).

In addition, climate conditions from the geographical origins of accessions (Fick and Hijmans 2017) were also analyzed to identify putative correlations with lipidomic profiles. We used a lasso-regularized linear model to predict each climate parameter from the lipidomic profiles. In result, we estimated the prediction power for each best-fitting model, and the lipids, being the best predictors for the climate parameters. While the CHD condition did not provide significant results, the HD condition highlighted in total 15 significant model fits,

5 of which also gave significant results in a cross-validation test (Supplemental Fig. S7). All of these 5 significant variables are related to temperature. In total, 6 “climate-predictive lipids” were also significantly associated with KCS4, with TAG 58:6, TAG 60:6, and TAG 56:8 being linked to the temperature in the driest and the coldest quarters of the year.

Finally, cluster analysis was performed for climate parameters and TAG levels by grouping accessions according to KCS4 haplotypes (m11502-m11503: CC, CT, TT, TC), and significant differences between mean values for haplotypes were tested using nonparametric ANOVA (Fig. 7). Eight climate parameters plus latitude and longitude differed significantly between haplotypes ( $P < 0.05$ , Fig. 7A). Moreover, 5 of them coclustered with TAGs associated with KCS4, 4 of them related to precipitation, and 1 to temperature (isothermality). The “TT” haplotype in most cases, presented the lowest value for these parameters, suggesting that the allelic effect of these 2 SNPs is opposite for climate variables (Fig. 7B).

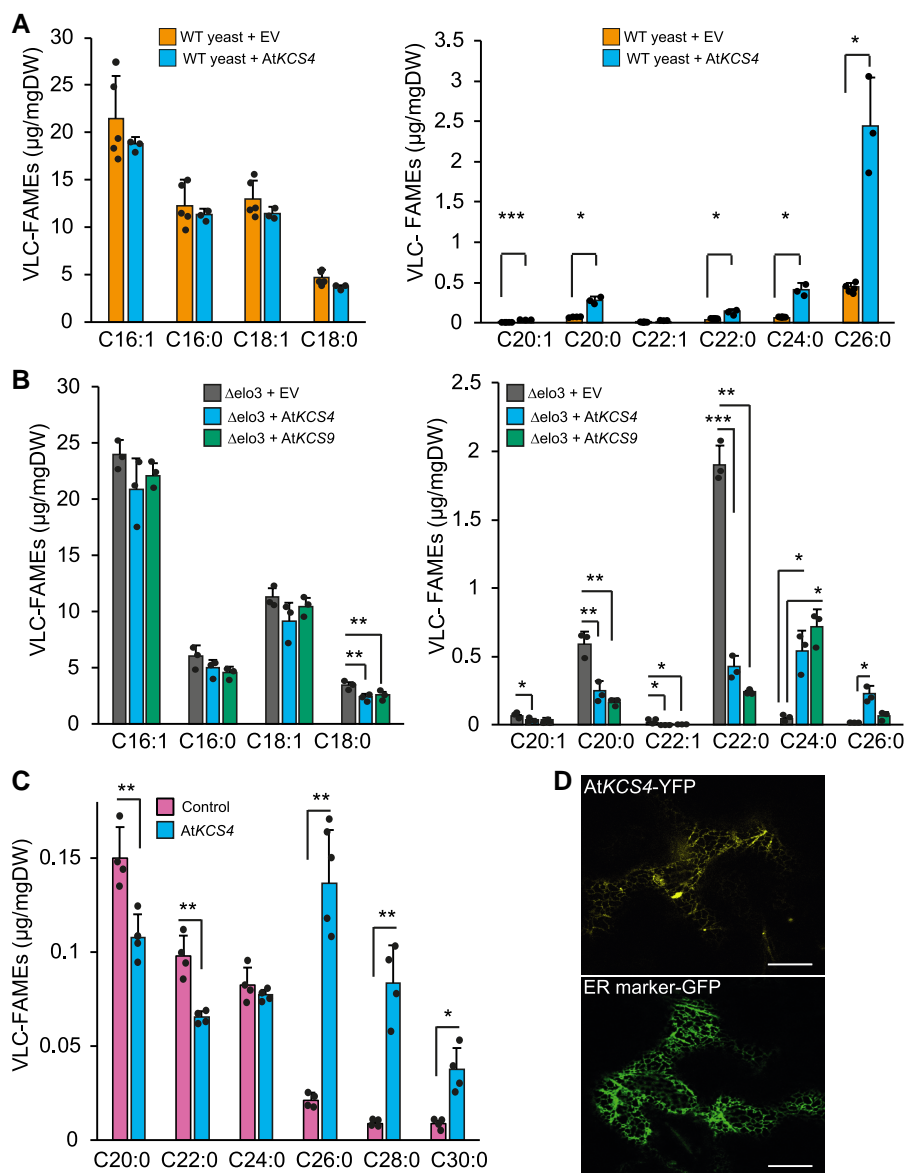
## Discussion

Recent years have witnessed the successful application of metabolic GWAS (Luo 2015). In lipids, the effort has focused more on dissecting the genetic architecture of oil biosynthesis (Correa et al. 2020). Here, we focus on glycerolipid profiles of ~300 Arabidopsis natural accessions grown under control conditions and subjected to a sudden environmental change given by combined stress of heat and darkness. Our GWAS exposed the high plasticity of the Arabidopsis genome, allowing the identification of genes involved in the regulation of lipid metabolism, in an environment-dependent manner (Fig. 1A).

Lipid-based GWAS in control conditions showed several mild-strength associations, similar to what was found for glycerolipids in rice (Hong et al. 2022) and for primary metabolites in maize and Arabidopsis (Wen et al. 2015; Wu et al. 2016), which are controlled by a large number of small-effect loci. This could be explained by the fact that both lipids and primary metabolites are essential for the survival of the cell and thus, do not show on/off phenotypes. This is different from natural variation in secondary metabolites, which is controlled, to a large extent, by major large-effect loci (Chan et al. 2011; Matsuda et al. 2012; Wen et al. 2014; Wu et al. 2018).

### Figure 4. (Continued)

Fisher-LSD post hoc test was performed to detect differences between lines for each TAG species. Significant changes ( $P < 0.05$ ) are marked with asterisks (A–F) or different letters (G). H–J) Transient expression of AtKCS4(Met) in *N. benthamiana* leaves. Plants were grown in CHD conditions for the whole experiment. TAG levels for classes 52 (52:0 to 52:8), 54 (54:0 to 54:9) and 56 (56:0 to 56:0) for the control line (white) and for the AtKCS4(Met) line (dark grey). Box-plots ( $n = 3$ ). Significant values (*t*-test) are indicated (\* $P < 0.05$ , \*\* $P < 0.01$ ). K–M) Leaf cuticular wax coverage for WT accessions and allelic mutants carrying the KCS4(Met) allele (WT: red; mutant: pink) and the KCS4(Thr) allele (WT: blue; mutant: light-blue) in CHD and HD conditions. ALK: alkane, ALC: alcohol, ALD: aldehyde, FA: fatty acid. Box-plots ( $n = 4$ ). One-way ANOVA followed by Fisher-LSD post hoc test was performed to detect differences between lines for each wax species. Means with a common letter are not significantly different ( $P < 0.05$ ).

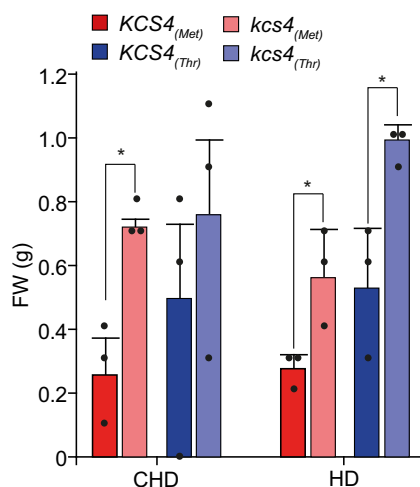


**Figure 5.** KCS4 activity and specificity in FAE and subcellular localization. **A**) Very long chain fatty acid methyl ester (VLC-FAME) profiles (16 to 18 and 20 to 26, separate panels) of WT yeast transformed with an empty vector and *AtKCS4*(Met) (reference included in the figure). **B**) VLC-FAME profiles (16 to 18 and 20 to 26, separate panels) of  $\Delta$ elo3 yeast mutant transformed with an empty vector (EV, grey), *AtKCS4*(Met) (blue), and *AtKCS9* (green). **C**) VLC-FAME profiles (20 to 30) from leaves of *N. benthamiana* control plants or transformed with *AtKCS4*(Met). Bars represent means  $\pm$ SD ( $n = 3$  to 5), individual values are plotted as filled circles. Significant differences for *t*-test are shown (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). **D**) Confocal images of *N. benthamiana* leaves transiently transformed with *KCS4*(Met)-YFP construct. The plasmid ER-GK CD3-955 was used as an ER marker. Scale = 20  $\mu$ m. DW: dry weight.

In our study, applying combined heat and darkness resulted in a strong perturbation of lipid homeostasis. This led to the identification of a large-effect QTL at Chromosome 1, harboring *KCS4*. *KCS4* is associated with several puTAGs specifically under carbon starvation conditions (HD, 3D, and 6D) in a highly robust manner (Fig. 1B, Supplemental Fig. S2B and Data Sets S2/S3). Under HD, all puTAGs mapping to *KCS4* are higher than under CHD, but accessions carrying *KCS4*(Met) always have higher levels of puTAGs than accessions carrying *KCS4*(Thr). PuTAGs that

mapped to *KCS4* clustered together and correlated negatively to saturated TAGs, further illustrating the role of *KCS4* in the specific regulation of puTAGs (Fig. 1C).

We have provided several lines of evidence that *KCS4* is the causal gene for TAG variation and that the *KCS4*(Met) allele determines higher levels of puTAGs under stresses (HD, 3D, 6D): (i) LD, correlation among lipid species and haplotypic analysis point *KCS4* as the candidate gene for puTAG variation under stress (Figs. 1 and 2), (ii) transient overexpression of *KCS4*(Met) in *N. benthamiana* mimics the phenotype we



**Figure 6.** Biomass changes in *KCS4* WT and allelic mutants. Rosettes fresh-weight measured in accessions carrying *KCS4*<sub>(Met)</sub> allele (WT: red, mutant: pink) or *KCS4*<sub>(Thr)</sub> allele (WT: blue, mutant: light-blue). Plants were grown in control conditions (CHD) or subjected to heat and darkness for 24 h (HD). Data represent means  $\pm$  SE,  $n = 3$ . Significant differences for  $t$ -test are shown (\* $P < 0.05$ ).

observed in the GWAS population grown under stress conditions (Fig. 4, H–J), (iii) lipidomic analysis of allelic mutants (*kcs4*<sub>[Met]</sub>/*kcs4*<sub>[Thr]</sub>) and their respective wild-type accessions (*KCS4*<sub>[Met]</sub>/*KCS4*<sub>[Thr]</sub>) shows a clear phenotype on puTAG, VLCFA, and leaf wax components (Figs. 4 and 5, Supplemental Fig. S5). Moreover, our detailed expression analysis of the entire population indicates that TAG variation in HD conditions is not due to differences in the expression of *KCS4* alleles (Fig. 3D). This finding further supports the hypothesis that the causal polymorphism is the nsSNP m11503, leading to a Met to Thr amino acid change.

*KCS4* is a functional enzyme, localized in the ER, with substrate specificity for VLCFAs of C22 and C24 length, and has the ability to elongate VLCFAs to C24 and up to C26 (Fig. 5). This result is in agreement with Kim et al. (2021) findings for *kcs4* T-DNA mutant, which accumulates C22 and C24. Moreover, the results from ectopic expression in yeast and plants confirmed that *KCS4* has specificity for the elongation of saturated FAs (Fig. 5), in agreement with previous findings that showed that elongase-condensing enzymes are mostly active with saturated FA (Blacklock and Jaworski 2006).

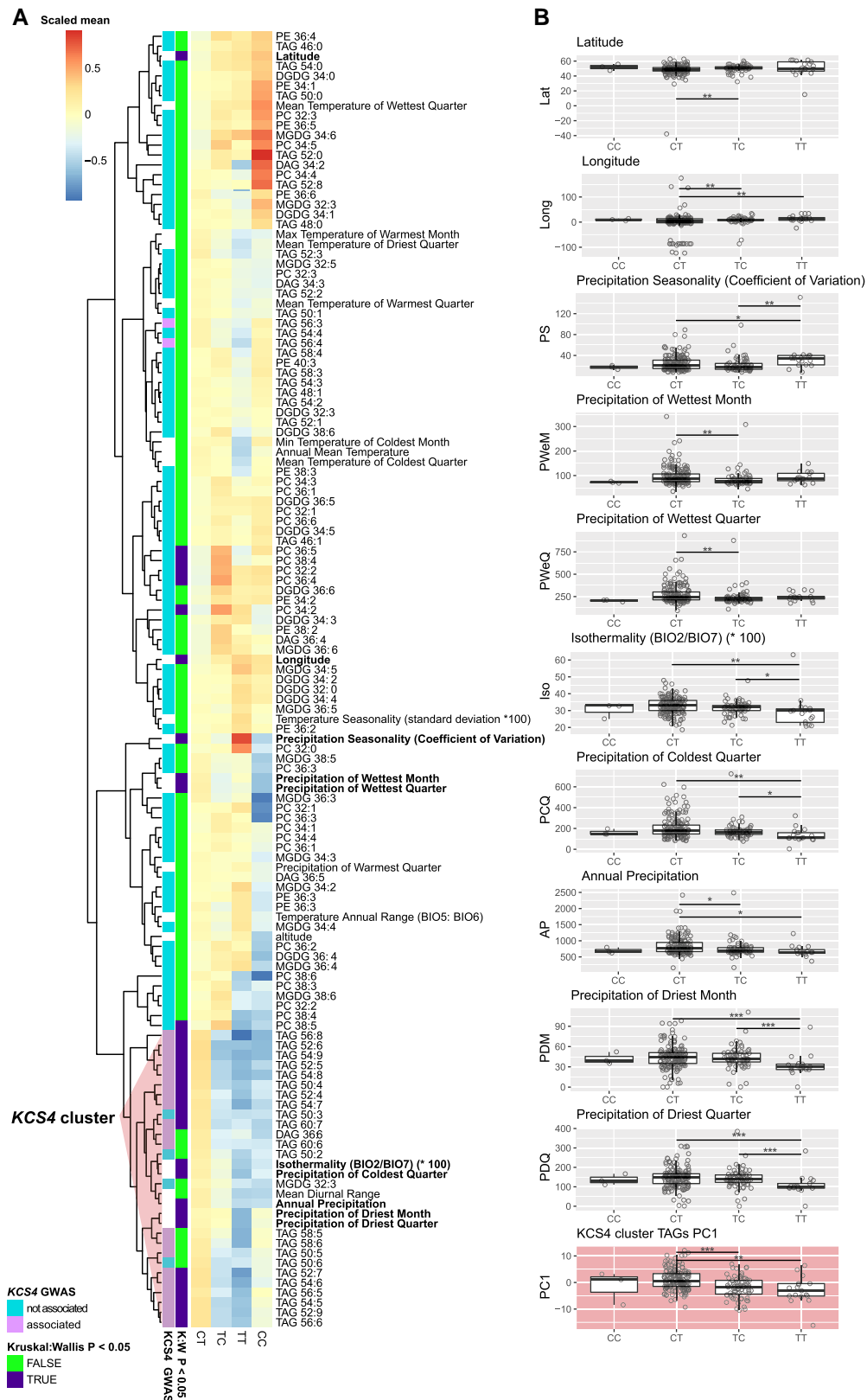
Under HD, differences in TAG levels for accessions carrying *KCS4* opposite alleles (*KCS4*<sub>[Met]</sub> vs. *KCS4*<sub>[Thr]</sub>) become more significant as the number of double bonds in TAGs increases (Fig. 4, D–G). In addition, TAG levels are significantly different when comparing *KCS4*<sub>(Met)</sub> with its allelic mutant, *kcs4*<sub>(Met)</sub>, in control conditions (CHD) (Fig. 4, A–C and G). In the same line of evidence, the mutant *kcs4*<sub>(Met)</sub> showed a significant reduction of the major components of leaf waxes compared to its wild-type allele in CHD (Fig. 4, K–M). Not unexpectedly, *kcs4*<sub>(Thr)</sub> did not show any differences in TAG levels or wax compounds compared to *KCS4*<sub>(Thr)</sub>. It has been previously shown in GWAS that mutating the

weaker allele did not always result in clear phenotypic changes for the traits under study (Todesco et al. 2010; Fusari et al. 2017; Wang et al. 2022; Yang et al. 2022).

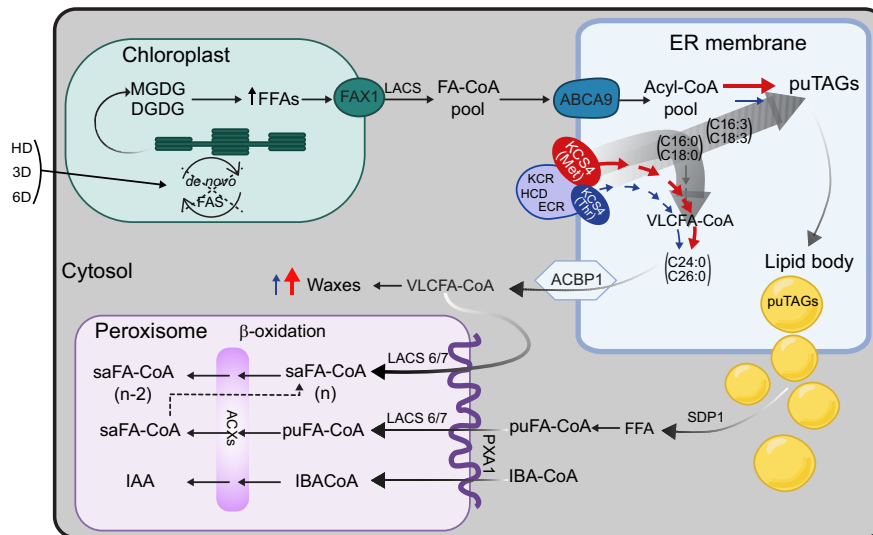
The phenotype can also depend on the context, and a change in either the genetic or the environmental background can amplify or remove the effect (Tonsor et al. 2005). Under HD, there is a strong metabolomic and expression reprogramming as shown before for lipids and primary and secondary metabolites (Burgos et al. 2011; Szymanski et al. 2014; Wu et al. 2016, 2018). This could explain why we did not observe a phenotype by mutating only 1 gene. In addition, sometimes effects are not evident when analyzing single mutants, particularly if there are gene families encoding the same or very similar enzymatic activities. For example, in studies of auxin conjugates, indole-3-acetic acid showed a reduction only when triple amidohydrolase mutants were analyzed (Ludwig-Müller 2011). *KCS4* is 1 of 21 genes encoding KCS activity (Joubès et al. 2008; Guo et al. 2016) and we have shown that expression of other KCS members is altered in *kcs4* allelic mutants, indicating a complex regulation between KCS genes (Supplemental Fig. S5). Of note, the nature of this complex regulation was recently demonstrated to involve protein–protein interactions within the elongase complex (Kim et al. 2022).

During plant evolution, the KCS family began with a single gene in algae and evolved via numerous gene duplications into a 21-member gene family in Arabidopsis (Guo et al. 2016). Analysis of LD and Tajima's  $D$  parameters in the region harboring *KCS4* showed that diversity is limited to 2 major haplotypes and that the region is under strong selective pressure (Fig. 2). As *KCS4* is involved in TAG regulation under conditions where carbon reserves are exhausted, it might be possible that *KCS4* has been targeted for natural selection to enhance adaptation in stressful environments. In nature, 24 h of darkness do not occur, but carbon starvation is also observed when plants are exposed to very limited light periods in regions with high amplitudes of day length. In agreement with the adaptive role of *KCS4* to environmental stresses, the geographical origin of the accessions (i.e. latitude) correlated with *KCS4* haplotypes (Fig. 7). Together, these analyses indicate that polymorphism variation at *KCS4* is important for adaptation in different climatic zones by adjustment of lipid metabolism.

In an event of sudden extended darkness (3D, 6D) or combined heat and darkness (HD), plants become carbon-starved (Supplemental Fig. S3) (Stitt and Zeeman 2012). Increasing night temperatures accelerates enzyme activities, exhausting carbon resources prematurely (Pilkington et al. 2015). In this case, FAs are used as alternative substrates for respiration via  $\beta$ -oxidation (Fan et al. 2014). There are many lines of evidence that support the idea that plant lipids are alternative sources of carbon and energy under stress, particularly under carbon starvation. For instance, Kunz et al. (2009) showed that by knocking out the PEROXISOMAL ABC-TRANSPORTER1 (*PXA1*, AT4G39850), the import of FFA into the peroxisome is blocked, resulting in plants dying faster in extended



**Figure 7.** KCS4 haplotype correlation with climate and geographic parameters. **A**) Heatmap of the mean values of lipid, climate, and geographical variables for each of the 4 KCS4 haplotypes (SNPs m11502-m11503). Results of the KCS4 locus association as well as nonparametric ANOVA test (Kruskal–Wallis) are highlighted in the row annotation bar. Cluster dominated by KCS4-associated TAGs is marked. **B**) Average haplotype value for the principal component 1 (PC1) obtained using the data in KCS4 cluster (in **A**), and the climate and geographical parameters showing significant Kruskal–Wallis in **A**. Significant differences are shown (pairwise Wilcoxon rank sum test: \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001).



**Figure 8.** Model summarizing the role of KCS4 on FA fate in Arabidopsis under carbon starvation. The lack of de novo FA synthesis under heat and darkness (HD) or extended darkness (3D, 6D) induces the degradation of galactolipids (MGDGs and DGDGs) from the chloroplasts' thylakoids. FFAs are then exported to the ER (FA-CoA pool), where KCS4 is localized and acts as a branch point in the fate of FAs, channeling saturated FAs to VLCFA elongation and tipping the balance to a higher polyunsaturated-to-saturated-FA ratio for puTAG accumulation (degrade grey arrows). Two KCS4 alleles have differential capacity to allocate saturated FAs into the VLCFA pathway, with a major -more efficient- allele, KCS4(Met), and a minor -weaker allele, KCS4(Thr), exemplified here in different sizes. KCS4(Met) accessions efficiently channel the saturated FA from the Acyl-CoA pool to cuticular waxes, as they present higher accumulation of waxes than KCS4(Thr) accessions under stress. In addition, polyunsaturated TAG (puTAGs) levels are, therefore, higher in KCS4(Met) accessions compared to KCS4(Thr) accessions, as the result of a rise in the ratio of polyunsaturated-to-saturated FAs going to TAG formation. VLCFA-CoA produced by the elongase complex (here exemplified by KCS4, KCR, HCD, ECR) go to increase the cuticular waxes under HD and/or directly serve as a source of carbon undergoing degradation in the peroxisome; whereas puTAGs accumulate into lipid bodies first. FFAs are then released from lipid bodies by the action of lipases (SDP1) and further imported to the peroxisome to undergo  $\beta$ -oxidation. Acronyms: ABCA9 (ATP-BINDING CASSETTE A9; AT5G61730), ACBP-1 (Acyl-CoA-binding protein 1; AT5G53470); ACX (ACYL-COA OXIDASE); CoA (coenzyme A); DGDG (digalactosyldiacylglycerol); ECR (ENOYL-COA REDUCTASE; AT3G55360); FAs (fatty acids); FAS (fatty acid synthesis); FAX1 (FATTY ACID EXPORT1; AT3G57280); FFAs (free fatty acids); HCD ( $\beta$ -HYDROXYACYL-COA DEHYDRATASE); IAA (indol-3-acetic-acid), IBA (indol-3-butyric-acid), KCR ( $\beta$ -KETOACYL REDUCTASE); KCS4(Met/Thr) (3-KETOACYL-COA SYNTHASE4); LACS (LONG-CHAIN ACYL-COA SYNTHETASE); SDP1 (SUGAR DEPENDENT1; AT5G04040); MGDG (monogalactosyldiacylglycerol); PXA1 (PEROXISOMAL ABC-TRANSPORTER 1; AT4G39850); pu (polyunsaturated), sa (saturated), TAGs (triacylglycerols); VLCFA (very long chain fatty acid).

darkness. Fan et al. (2017), through the study of *sdp1* mutants (SUGAR DEPENDENT1, SDP1, AT5G04040), could define the crucial role that TAG metabolism has in membrane lipid breakdown, FA turnover, and plant survival under extended darkness. Another study showed that there are different metabolic strategies employed by leaves in response to different darkening treatments (Law et al. 2018). These strategies involve either importing sugars from sink sources, metabolizing amino acids, or activating the turnover of lipids from the plastid membranes to the  $\beta$ -oxidation in the peroxisome. The last strategy is particularly important after 6 d of dark treatment. Finally, Yu et al. (2018) demonstrated that the disruption of starch synthesis resulted in a significant increase in FA synthesis via post-translational regulation of the plastidial acetyl-coenzyme A carboxylase and a concurrent increase in the synthesis and turnover of membrane lipids and triacylglycerol.

To summarize our findings, we present a model of lipid remodeling under heat and dark stress in Arabidopsis (Fig. 8).

In the proposed model, darkness induces the predominant response within the combined stress, blocking de novo FA synthesis (de novo FAS) and producing considerable membrane-lipid remodeling. We observed decreased levels of MGDGs and DGDGs across accessions, as previously shown (Burgos et al. 2011). Under carbon starvation, FAs, mostly 16:3 and 18:3, are released from galactolipids by lipases (Higashi et al. 2015; Fan et al. 2017). FAs are then exported to the ER, where KCS4 is localized and acts as a branch point in the fate of FAs, directing the saturated ones to the VLCFA. Two KCS4 alleles have differential capacities to allocate saturated FAs into the VLCFA pathway, with a major, more efficient allele, KCS4(Met), and a minor, weaker-allele, KCS4(Thr). KCS4(Met) accessions efficiently channel the saturated FA from the Acyl-CoA pool to cuticular waxes, as they present a higher accumulation of waxes than KCS4(Thr) accessions under stress (Fig. 4). Polyunsaturated TAG levels are, therefore, higher in KCS4(Met) accessions compared to KCS4(Thr) accessions, as the result of a rise in

the ratio of polyunsaturated-to-saturated FAs that proceeds to TAG formation (Fig. 4). This effect of KCS4(Met) on TAG levels is only observed under stress conditions, when there is an increased amount of free-PUFA coming from membrane lipids, and when it is necessary to sequester them into TAGs, to protect the cell from the FFA-toxicity induced damage (Kao et al. 2018).

Our work demonstrates that KCS4 is a key enzyme involved in the regulation of puTAG levels under heat and darkness and under extended darkness alone, adding complexity to lipid metabolism regulation under carbon starvation. The enzymatic function of KCS4 in the synthesis of VLCFA affects the accumulation of the unsaturated Acyl-CoA pool (e.g. 16:3 and 18:3) and the sequestration of unsaturated Acyl-CoA into puTAGs. Despite being a member of a 21-gene family with overlapping function in VLCFA synthesis and some redundancy in spatial and temporal regulation of expression, these results single out KCS4 as functioning in fine-tuning lipid regulation, and show that polymorphisms in KCS4 might have helped plants adapt to different geographic regions.

In conclusion, our study demonstrates that lipidomic GWAS performed under stress conditions is a powerful and unbiased tool for shedding light on the genetic control of lipid metabolism. We identified a QTL to which lipids are mapped in an environment-specific fashion, thus illustrating the plasticity of the genome in the regulation of lipid metabolism in changing environments. This approach fleshed out the fundamental role that KCS4 has in FA channeling under stress and proved to be powerful in unraveling the genetic regulation of complex phenotypes.

## Materials and methods

### Plant material

The GWA mapping population comprised a total of 330 *A. thaliana* natural accessions belonging to the HapMap panel (Horton et al. 2012). This panel was selected because the accessions maximize diversity and minimize redundancy and close family relatedness (Baxter et al. 2010).

### GWAS experimental design

Accessions or mutant lines were grown in soil (potting compost) in short days (8 h light) for 4 wk, then transferred to long days (16 h light) for an additional period of 2 wk. Control plants were harvested immediately and stressed plants were subjected to the treatment and harvested right after it.

In control conditions (for heat/dark experiment: CHD, for dark experiment: CD), plants were grown on long days (16 h light, 8 h dark), temperature was kept at 21/16 °C (day and night, respectively), light intensity at  $150 \mu\text{E m}^{-2} \text{s}^{-1}$ , and humidity at 75%. In the heat and dark experiment (HD), plants were grown like the plants in CHD for 6 wk, then subjected to dark and 32 °C continuously for 24 h, and then harvested. In

the dark treatment, plants were grown like the plants in CD for 6 wk and then transferred to continuous darkness either for 72 h (3D) or 144 h (6D). Plants from HD, 3D, and 6D were harvested 24, 72, and 144 h after the beginning of the treatment, respectively. Between 3 and 6 plants from each accession/line were collected for each condition tested and extracted for lipid extraction (see below). We followed a similar experimental design to that described in our recently published study (Zhu et al. 2022)

### Targeted lipid profiling by LC-MS and data acquisition

Lipid extraction was done, and measurements from *Arabidopsis* leaves (100 mg for each sample) were performed as described previously (Hummel et al. 2011). In brief, dried organic phase was measured using Waters Acquity ultra-performance liquid chromatography system (Waters, <http://www.waters.com>) coupled with Fourier transform mass spectrometry (UPLC-FT-MS) in positive ionization mode. Analysis and processing of raw MS data were done with REFINER MS 10.0 (GeneData, <http://www.genedata.com>). Workflow included peak detection, retention time alignment, and removal of chemical noise and isotopic peaks from the MS data. Obtained mass features characterized by specific peak ID, retention time, *m/z* values, and intensity were further processed using custom scripts in R (R Core Team 2019). Clusters with means of signal intensities lower than 50,000 were removed, and only peaks present in at least 70% of the samples were kept for further analysis. Peak intensities were normalized by the day of measurement and median and subsequently  $\log_2$  transformed. After that, obtained molecular features were queried against a lipid database for annotation. The in-house database used includes 219 lipid species of the following classes: diacylglycerols (DAGs), DGDGs, MGDGs, phosphatidylcholines, phosphatidylethanolamines, and TAGs. This was carried out by comparing their retention times and exact masses against those of the reference compounds, allowing maximal deviations of 0.1 min and 10 ppm. Identified lipids were confirmed by manual verification of the chromatograms using XCalibur (Version 3.0, Thermo-Fisher, Bremen, Germany) (Alseekh et al. 2021).

### General statistical and multivariate analysis of lipidomic data

To calculate the correlation between lipids, the Pearson correlation was used. The correlation plots and heatmaps were created using the package pheatmap (Raivo Kolde 2019). The lipids were clustered using complete-linkage clustering. PCA was obtained from the correlation matrix of lipidomic data (CHD and HD together) using the pcaMethod (Stacklies et al. 2007) package.

To detect the fold change in lipid feature levels across all accessions for each lipid feature, fold change was calculated by dividing the average of the 5 highest intensities measured



in a given accession by the average of the 5 lowest intensities, (i) for all annotated lipids ( $N_{\text{CHD}} = 98$ ,  $N_{\text{HD}} = 109$ ), (ii) only for TAGs ( $N_{\text{CHD}} = 30$ ,  $N_{\text{HD}} = 41$ ), and (iii) for lipid classes other than TAGs ( $N_{\text{CHD}} = 68$ ,  $N_{\text{HD}} = 68$ ).

### GWAS analysis

Data acquisition for GWAS, mapping, and locus identification were performed as described previously (Wu et al. 2016; Fusari et al. 2017). GWAS was performed using a compressed mixed linear model (MLM) implemented in the GAPIT (Lipka et al. 2012) package. The 109 (HD), 98 (CHD), 80 (3D), 109 (6D), and 122 (CD) lipid features were used as phenotypes (Supplemental Data Set S1), whereas a set of 214,051 SNPs was used as genotypic data (250 K SNPchip). A total of 3 principal components were included in the MLM to account for population structure, and the SNP fraction considered to perform PCA was set to 0.1 to avoid excessive computations. Kinship matrix and other parameters were set to default values.

Additionally, we performed GWAS using a full genome sequence from a section of Chromosome 1. Shortly, sequences for 80 kb around *KCS4* (AT1G19440, TAIR10 position: 6689119–6769118) were obtained from the Arabidopsis 1001 Genomes project (<http://signal.salk.edu/atg1001/3.0/gbrowser.php>, 162 accessions). The alignment was performed using the default routine implemented in MAFFT (Kato et al. 2019) with a final manual inspection to discard alignment problems due to big gaps or missing data (Supplemental File S1). For 9 accessions, 2 sequences were available. The sequences included further in the analysis were selected based on the agreement of alleles with the SNPchip data and/or on the percentage of nonmissing data ( $\geq 90\%$ ). Final SNP matrix, obtained by filtering for minor allele frequency  $\geq 5\%$ , contained 1,213 SNPs and indels (Supplemental Data Set S4). This matrix was used as the genotypic data. Population structure and cryptic relationships among accessions were taken into account by including principal components as cofactors and a kinship matrix obtained with SNPs in the 250 K SNPchip for the subset of 162 accessions (Supplemental Data Set S4). Lipid features collected previously (HD and 3D) Supplemental Data Set S1) were used as phenotypes. The 3 data sets (genotype, population structure, and phenotypes) matched for 149 accessions in the HD experiment and for 134 accessions in the 3D experiment. MLM was implemented using TASSEL software with the default settings for compressed MLM (Bradbury et al. 2007).

### SNP filtering, QTL identification, and candidate gene selection

For results obtained using GAPIT package, only SNPs having a  $P$ -value  $< 1/N$  ( $N = 214,051$ ) were retained to correct for multiple comparisons. These SNPs were assigned to the same QTL if the genomic distance between them was  $\leq 10$  kb. All genes in a given QTL were taken as putative candidates. Further

candidate gene selection for validation was done based on functional annotation or previous knowledge.

For results obtained using Tassel software, SNPs were regarded as significantly associated with lipid traits after setting Bonferroni threshold ( $\alpha = 0.05/N$ ,  $N = 1,213$  polymorphism tested).

### LD, nucleotide diversity, and haplotype analysis

LD was computed as the coefficient of correlation between pair of sites ( $r^2$ ) using the LDheatmap package (Shin et al. 2006), either with the SNPchip (for 1,300 and 314 accessions) or with the full sequence data (for 149 accessions). For deviations from neutrality, Tajima's  $D$  statistic (Tajima 1989) was computed as described previously (Fusari et al. 2017).

Haplotypes for *KCS4* were obtained by using SNPchip data considering 1,000 bp upstream of the ATG, the coding region and 1,000 bp downstream of the stop codon (SNPs = m11502, m11503, m11504, m11505). Accessions belonging to the same haplotype were grouped and the haplotype median was obtained for each lipid feature. To cluster haplotypes, a combination of allele sharing distance and Ward's minimum variance was used, as described before (Gao and Starmer 2007; Qiu et al. 2015). One-way ANOVA followed by Bonferroni correction for multiple comparisons was performed to test differences between haplotypes ( $P < 0.05$ ).

### Mutant line selection and genotyping

For Arabidopsis CRISPR-Cas9 lines, the binary vector containing guide RNAs (gRNAs) and the Cas9 nuclease were generated based on a classic cloning procedure described previously (Schumacher et al. 2017). A pair of gRNAs was used to target the N-terminal coding sequence of *KCS4* using the spacer sequences indicated in Supplemental Data Set S7. As background, wild-type accessions carrying *KCS4*(Met): Pro-0, Petergof, Per-1, Col-0; and *KCS4*(Thr): Si-0, Bu-8, Sh-0 were chosen based on TAG levels phenotyped in HD. Mutations were confirmed by PCR and sequencing of *KCS4* using primers listed in Supplemental Data Set S7. Sequence contigs and alignment were done using Bioedit (Hall 1999).

For heterologous expression in yeast and transient expression in *N. benthamiana*, *KCS4* (AT1G19440) and *KCS9* (AT2G16280) open reading frame was amplified from Arabidopsis flower cDNA using primers listed in Supplemental Data Set S7. The corresponding PCR fragments were inserted into the pDONR 221 ENTRY vector using the GATEWAY recombinational cloning technology (Invitrogen, Darmstadt, Germany) and subsequently transferred into pK7YWG2, pK7WG2D, and pvtLEU DESTINATION vectors (Karimi et al. 2002; Ditttrich-Domergue et al. 2014) by using LR clonase II. All transient heterologous expressions in *N. benthamiana* were made using the *Agrobacterium tumefaciens* strain GV3101 and 5-week-old plants cultivated in controlled conditions (16-h light photoperiod, 25 °C) (Perraki et al. 2012). *N. benthamiana* plants were agroinfiltrated on the abaxial side of the leaves and analyzed 2 and 5 d after

agroinfiltration for subcellular localization experiments and for overexpression, respectively.

For yeast experiments, *Saccharomyces cerevisiae* strains INVSc1 (MATa *his3Δ1 leu2 trp1-289 ura3-52*) and  $\Delta$ *elo3* (MATa *his3Δ1 leu2 trp1-289 ura3-52 elo3Δ*) were used to express *KCS4* wild-type and allelic mutants and *KCS9*.

### RT-qPCR

*RGS1-HXK1 INTERACTING PROTEIN1/RHIP1* (AT4G26410) was used as a reference gene for *KCS4* transcript analysis in CHD and HD for the complete GWAS population (Czechowski et al. 2005). *GAPDH3* (AT1G13440) was used as a reference gene for starvation markers and *KCS* transcript analysis. All primers used are listed in Supplemental Data Set S7.

RT-qPCR was performed on 3 individual plants per ecotype. Each sample was tested in 3 technical replicates. Averaged values for each ecotype were pooled together according to the allele carried. For CHD  $n = 193$  for the C allele and  $n = 94$  for the T allele. For HD  $n = 163$  for the C allele and  $n = 75$  for the T allele. RT-qPCR was performed on an ABI Prism 7900 HT real-time PCR system (Applied Biosystems/Life Technologies, Darmstadt, Germany) in 384-well PCR plates with a total reaction volume of 5  $\mu$ L (2  $\mu$ L forward and reverse primer mixture, 0.5  $\mu$ M each), 0.5  $\mu$ L cDNA and 2.5  $\mu$ L Power SYBR Green-PCR Master Mix (Applied Biosystems/Life Technologies) using the following cycling program: 50 °C for 2 min, 95 °C for 5 min, 40 cycles of 95 °C for 15 s, and 60 °C for 60 s, followed by a denaturation step of 95 °C for 15 s and 60 °C for 15 s, followed by a continuous temperature increase (0.3 °C s<sup>-1</sup>) to 95 °C (for 15 s). Steps 5, 6, and 7 were introduced to record a dissociation or melting curve for each product in order to detect nonspecific amplification. Expression values were calculated by  $2^{(-\Delta\Delta C_t)}$  using either *ACTIN 2* (AT3G18780) or *GAPDH3* (AT1G13440) as reference genes. Errors are given as the lower ( $2^{(-\Delta\Delta C_t - sd)}$ ) and upper ( $2^{(-\Delta\Delta C_t + sd)}$ ) limit of the standard deviation around the mean.

### EMSA

EMSA reactions were performed as previously described (Thirumalaikumar et al. 2018). Briefly, HSFA2-GST proteins were prepared as stated before (Puranik et al. 2011). *HSFA2* coding sequence was PCR-amplified from Arabidopsis heat-stressed cDNA using primers listed in Supplemental Data Set S7. PCR products were recombined in pDEST24 using the GATEWAY cloning system (Invitrogen). Recombinant vectors were transformed into *Escherichia coli* Star (DE3) pRARE generated by transforming the pRARE plasmid isolated from Rosetta (DE3). 5' labeled with fluorescent dye DY682 and unlabeled *KCS4* promoter regions spanning the SNP position m11502 (C or T allele) were purchased from Eurofins MWG operon. DNA EMSA reactions were performed using Odyssey Infrared EMSA kit (Li-COR, Bad Homburg, Germany). Competitor was added at 200 $\times$  the amount of probe.

### Fatty acid composition and content analysis for *N. benthamiana* and yeast

For fatty acid methyl esters (FAMES) analysis in *N. benthamiana*, 3 agroinfiltrated leaf disks were transferred to tubes containing 1 mL of 0.5 M sulfuric acid in methanol containing 2% (v/v) dimethoxypropane and 20  $\mu$ g of heptadecanoic acid (C17:0) as an internal standard. Transmethylation was performed at 85 °C for 3 h. For FAMES analysis in yeast, yeasts were grown in 5 mL of appropriate liquid minimal medium for 1 wk at 30 °C, then pelleted and washed in 2 mL of 2.5% (w/v) NaCl. Yeast cell pellets were resuspended in 1 mL of 0.5 M sulfuric acid in methanol containing 2% (v/v) dimethoxypropane and 20  $\mu$ g of heptadecanoic acid (C17:0) as internal standard and transmethylation was performed at 85 °C for 1 h. After cooling, 2 mL of 2.5% (w/v) NaCl was added, and fatty acyl chains were extracted in 300  $\mu$ L of hexane. Samples were analyzed by GC-FID and GC-MS as previously described (Domergue et al. 2010).

### Cuticular wax analysis

Epicuticular waxes were extracted from leaves from 2 to 4 different plants for each allele (*KCS4*[Met]/*KCS4*[Thr]) by immersing tissues for 30 s in chloroform containing docosane (C22 alkane) as an internal standard. Extracts were derivatized and analyzed as previously described (Bourdenx et al. 2011).

### Confocal microscopy

Live imaging was performed using a Leica SP5 confocal laser scanning microscopy system (Leica, Wetzlar, Germany) equipped with Argon, DPSS, He-Ne lasers, hybrid detectors, and 63 $\times$  oil-immersion objective. Two days post agroinfiltration, *N. benthamiana* leaf samples were gently transferred between a glass slide and a cover slip in a drop of water. The plasmid ER-gk CD3-955 was used as a fluorescent ER marker (Nelson et al. 2007). YFP and GFP fluorescence were observed using excitation wavelength of 488 nm and their fluorescence emission was collected at 490 to 540 nm.

### Biomass determination

Plants were grown and treated as described for lipid measurements. Plants from wild-type accessions carrying *KCS4*(Met) and *KCS4*(Thr) as well as allelic mutants *kcs4*(Met) and *kcs4*(Thr) were grown in CHD or treated for 24 h with HD and harvested right after (CHD and HD at the same time). Biomass measurement was performed on complete rosettes for each line. Results from accessions or mutants were then pooled according to the allele carried ( $n = 3 \pm$  SD).

### Geographic and phylogenetic signal of the lipidome

The neighbor-joining tree was constructed by BioNJ algorithm (Gascuel 1997), using J-C distance in SeaView 4.4.3 software (Gouy et al. 2010). Phylogenetic signal was computed using *K* statistics and as a background a Brownian motion model of

trait evolution (Blomberg et al. 2003) in the “picante” R package (Kembel et al. 2010). Significance of the phylogenetic signal was estimated by comparison of observed variance patterns to a null model obtained by 9,999 random permutations of the accession labels across the phylogeny. The altitude at the sites was estimated using the geomapdata database (Lees 2000) based on the latitude and longitude coordinates. The climate conditions have been downloaded from the WorldClim database (www.worldclim.org, Fick and Hijmans 2017). The lipidomic data were regressed against the geodesic distance matrix using nonparametric multivariate variance analysis (Anderson 2001). The relationship between lipid levels and the climate parameters was analyzed using regularized regression model (Friedman et al. 2010). The regularization parameter  $\alpha$  and penalty parameters  $\lambda$  were estimated using 10-fold cross-validation within the training set. In result, the Elastic Net regularization with parameter  $\alpha = 0.5$  was selected as optimal and applied for all regression models and the “ $\lambda_{\min} + 1se$ ” rule was applied to choose the penalty parameter for each model separately. Prediction power was estimated in the Monte Carlo cross-validation test with 100 samples with 57% and 25% of samples taken as a training and test set, respectively. Kruskal–Wallis and Wilcoxon rank sum tests have been performed using the R stats package.

### Statistics

All statistical tests (ANOVA and *t*-test results showing variables, parameters, degrees of freedom, and tests) are included in [Supplemental Data Set S8](#).

### Accession numbers

AT1G01120: 3-KETOACYL-COENZYME A SYNTHASE1 (KCS1); AT1G04220: 3-KETOACYL-COENZYME A SYNTHASE2 (KCS2); AT1G06250: LIPASE CLASS 3; AT1G06290: ACYL-COA OXIDASE3 (ACX3); AT1G06310: ACYL-COA OXIDASE6 (ACX6); AT1G06350: FATTY ACID DESATURASE (CER17); AT1G063609: FATTY ACID DESATURASE (ADS4.2); AT1G13440: GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (GAPDH3); AT1G19440: 3-KETOACYL-COENZYME A SYNTHASE4 (KCS4); AT1G25450: 3-KETOACYL-COENZYME A SYNTHASE5 (KCS5); AT1G68530: 3-KETOACYL-COENZYME A SYNTHASE6 (KCS6); AT2G15090: 3-KETOACYL-COENZYME A SYNTHASE8 (KCS8); AT2G16280: 3-KETOACYL-COENZYME A SYNTHASE4 (KCS9); AT2G26150: HEAT SHOCK TRANSCRIPTION FACTOR A2 (HSFA2), AT2G28630: 3-KETOACYL-COENZYME A SYNTHASE12 (KCS12); AT3G12120: FATTY ACID DESATURASE2 (FAD2); AT3G18780: ACTIN 2; AT4G26410: RGS1-HXK1 INTERACTING PROTEIN1/RHIP1; AT4G34250: 3-KETOACYL-COENZYME A SYNTHASE16 (KCS16); AT4G34520: 3-KETOACYL-COENZYME A SYNTHASE18 (KCS18); AT4G39850: PEROXISOMAL ABC-TRANSPORTER1 (PXA1); AT5G04040: SUGAR DEPENDENT1 (SDP1); AT5G20060: LYSOPHOSPHOLIPASE; AT5G43760: 3-KETOACYL-COENZYME A SYNTHASE20 (KCS20); AT5G49070: 3-KETOACYL-COENZYME A SYNTHASE21 (KCS21).

### Acknowledgments

We would like to thank Prof. Dr Lothar Willmitzer for discussion, support, and the opportunity to develop the project. In addition, we acknowledge Anne Michaelis and Gudrun Wolter for their outstanding technical assistance in LC-MS measurements, and Dr Karin Kohl for her help in plant cultivation. We thank Hezi Tenenboim for his editorial aid. CMF is a career member of the National Research Council (CONICET) and leader of a Max Planck Partner Group.

### Author contributions

Y.B. and C.M.F. conceived the project and supervised the study. U.L. conducted GWAS experiments, obtained and selected KO mutants, performed experimental validation with KO, statistical analysis, and interpretation of results. A.-K.R. helped U.L. in experimental work and performed analysis of the phenotypic data. J.J. and M.B. conducted experiments on tobacco and yeast lines. J.Sz. conducted the phylogenetic and geographical-signal analysis. V.P.T. provided expertise and assistance on TF analysis and advised on EMSA experiment and RT-qPCR assay. M.L. and S.W. assisted in GWAS and multivariate analysis. F.Z. provided the plant material and carried out the lipid profiling for 3D and 6D GWAS. N.E., S.K., and S.M. helped U.L. in RT-qPCR assay for GWAS population and KO validation. H.X. and W.J. performed lipid profiling for *kcs4* mutants. J.Sc. obtained KO constructs for transformation of Arabidopsis. A.S. provided access to LC-MS. A.R.F. and A.S. helped with discussion. C.M.F. performed TASSEL, LD, Tajima’s *D* and haplotype analyses, analyzed and interpreted the results, and helped with data visualization. C.M.F., U.L., and Y.B. wrote the paper with input from, J.J., J.Sz. and Y.L.-B.. All authors read and drafted the manuscript.

### Supplemental data

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Phenotypic analyses for the Arabidopsis GWAS panel under control (CHD) and stress (HD) conditions.

**Supplemental Figure S2.** Results from GWAS analysis on lipidomic data from plants grown under extended darkness (3D, 6D) conditions.

**Supplemental Figure S3.** Quantification of starvation markers in *KCS4* wild-type accessions and allelic mutants.

**Supplemental Figure S4.** *KCS4* enzymatic activity.

**Supplemental Figure S5.** Quantification of *KCS4*, *KCS5*, *KCS9*, and *KCS12* transcripts in wild-type *KCS4*(Met)/*KCS4*(Thr) accessions and *kcs4*(Met)/*kcs4*(Thr) allelic mutants.

**Supplemental Figure S6.** Phylogenetic signal of the lipidome.

**Supplemental Figure S7.** Prediction of climate parameters from lipidomic profiles.

**Supplemental Data Set S1.** Phenotypic data used for GWAS.

**Supplemental Data Set S2.** GWAS results.

**Supplemental Data Set S3.** Candidate genes and QTL colocalization.

**Supplemental Data Set S4.** Data used for GWAS using complete sequenced information from 162 *Arabidopsis* ecotypes.

**Supplemental Data Set S5.** GWAS result using complete sequenced information from 162 *Arabidopsis* ecotypes.

**Supplemental Data Set S6.** Metabolomic analysis of KCS4 wild-type alleles and *kcs4* allelic mutants.

**Supplemental Data Set S7.** Primers and probes used in this study.

**Supplemental Data Set S8.** Statistics **Supplemental File S1.** Alignment of KCS4 QTL using complete sequenced information from 162 *Arabidopsis* ecotypes.

**Supplemental File S2.** Alignment of KCS4 genomic region from wild-type *Arabidopsis* ecotypes and CRISPR-Cas allelic mutants used in this study.

## Funding

This research was supported by the Israel Science Foundation (grant no. 859/19).

*Conflict of interest statement.* None declared.

## Data availability

The data used to support the results of this study are provided within the article.

## References

- Aleekh S, et al.** Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods*. 2021;**18**(7):747–756. <https://doi.org/10.1038/s41592-021-01197-1>
- Anderson MJ.** A new method for non-parametric multivariate analysis of variance. *Austral Ecol*. 2001;**26**(1):32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, et al.** A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1; 1. *PLoS Genet*. 2010;**6**(11): e1001193. <https://doi.org/10.1371/journal.pgen.1001193>
- Blacklock BJ, Jaworski JG.** Substrate specificity of *Arabidopsis* 3-ketoacyl-CoA synthases. *Biochem Biophys Res Commun*. 2006;**346**(2):583–590. <https://doi.org/10.1016/j.bbrc.2006.05.162>
- Blomberg SP, Garland Jr T, Ives AR.** Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*. 2003;**57**(4):717–745. <http://doi.org/10.1111/j.0014-3820.2003.tb00285.x>
- Bourdenx B, Bernard A, Domergue F, Pascal S, Léger A, Roby D, Pervent M, Vile D, Haslam RP, Napier JA, et al.** Overexpression of *Arabidopsis* ECERIFERUM1 promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant Physiol*. 2011;**156**(1):29–45. <https://doi.org/10.1104/pp.111.172320>
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES.** TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;**23**(19):2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Burgos A, Szymanski J, Seiwert B, Degenkolbe T, Hannah MA, Giavalisco P, Willmitzer L.** Analysis of short-term changes in the *Arabidopsis thaliana* glycerolipidome in response to temperature and light. *Plant J*. 2011;**66**(4):656–668. <https://doi.org/10.1111/j.1365-313X.2011.04531.x>
- Caldana C, Degenkolbe T, Cuadros-Inostroza A, Klie S, Sulpice R, Leisse A, Steinhauser D, Fernie AR, Willmitzer L, Hannah MA.** High-density kinetic analysis of the metabolomic and transcriptomic response of *Arabidopsis* to eight environmental conditions. *Plant J*. 2011;**67**(5): 869–884. <https://doi.org/10.1111/j.1365-313X.2011.04640.x>
- Chan EKF, Rowe HC, Corwin JA, Joseph B, Kliebenstein DJ.** Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol*. 2011;**9**(8):e1001125. <https://doi.org/10.1371/journal.pbio.1001125>
- Correa SM, Fernie AR, Nikoloski Z, Brotman Y.** Towards model-driven characterization and manipulation of plant lipid metabolism. *Prog Lipid Res*. 2020;**80**:101051. <https://doi.org/10.1016/j.plipres.2020.101051>
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR.** Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol*. 2005;**139**(1):5–17. <https://doi.org/10.1104/pp.105.063743>
- Dittrich-Domergue F, Joubès J, Moreau P, Lessire R, Stymne S, Domergue F.** The bifunctional protein TtFARAT from *Tetrahymena thermophila* catalyzes the formation of both precursors required to initiate ether lipid biosynthesis. *J Biol Chem*. 2014;**289**(32):21984–21994. <https://doi.org/10.1074/jbc.M114.579318>
- Domergue F, Vishwanath SJ, Joubès J, Ono J, Lee JA, Bourdon M, Alhattab R, Lowe C, Pascal S, Lessire R.** Three *Arabidopsis* fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. *Plant Physiol*. 2010;**153**(4):1539–1554. <https://doi.org/10.1104/pp.110.158238>
- Fan J, Yan C, Roston R, Shanklin J, Xu C.** *Arabidopsis* lipins, PDAT1 acyltransferase, and SDP1 triacylglycerol lipase synergistically direct fatty acids toward  $\beta$ -oxidation, thereby maintaining membrane lipid homeostasis. *Plant Cell*. 2014;**26**(10):4119–4134. <https://doi.org/10.1105/tpc.114.130377>
- Fan J, Yu L, Xu C.** A central role for triacylglycerol in membrane lipid breakdown, fatty acid  $\beta$ -oxidation, and plant survival under extended darkness. *Plant Physiol*. 2017;**174**(3):1517. <https://doi.org/10.1104/pp.17.00653>
- Fick SE, Hijmans RJ.** WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. 2017;**37**(12):4302–4315. <https://doi.org/10.1002/joc.5086>
- Fraire-Velázquez S, Balderas-Hernández VE.** Abiotic stress in plants and metabolic responses. In: **Vahdati K**, editor. *Abiotic stress—plant responses and applications in agriculture*. London: Intech Open; 2013. p.25
- Franke R, Höfer R, Briesen I, Emsermann M, Efremova N, Yephremov A, Schreiber L.** The DAISY gene from *Arabidopsis* encodes a fatty acid elongase condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza-micropyle region of seeds. *Plant J*. 2009;**57**(1):80–95. <https://doi.org/10.1111/j.1365-313X.2008.03674.x>
- Friedman J, Hastie T, Tibshirani R.** Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;**33**(1): 1. <https://doi.org/10.18637/jss.v033.i01>
- Fusari CM, Kooke R, Lauxmann MA, Annunziata MG, Enke B, Hoehne M, Krohn N, Becker FFMM, Schlereth A, Sulpice R, et al.** Genome-wide association mapping reveals that specific and

- pleiotropic regulatory mechanisms fine-tune central metabolism and growth in *Arabidopsis*. *Plant Cell*. 2017;**29**(10):2349–2373. <https://doi.org/10.1105/tpc.17.00232>
- Gacek K, Bayer PE, Bartkowiak-Broda I, Szala L, Bocianowski J, Edwards D, Batley J.** Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*. *Front Plant Sci*. 2017;**7**:2062. <https://doi.org/10.3389/fpls.2016.02062>
- Gao X, Starmer J.** Human population structure detection via multilocus genotype clustering. *BMC Genet*. 2007;**8**(1):34. <https://doi.org/10.1186/1471-2156-8-34>
- Gascuel O.** BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*. 1997;**14**(7):685–695. <https://doi.org/10.1093/oxfordjournals.molbev.a025808>
- Gouy M, Guindon S, Gascuel O.** Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 2010;**27**(2):221–224. <https://doi.org/10.1093/molbev/msp259>
- Graf A, Smith AM.** Starch and the clock: the dark side of plant productivity. *Trends Plant Sci*. 2011;**16**(3):169–175. <https://doi.org/10.1016/j.tplants.2010.12.003>
- Guo HS, Zhang YM, Sun XQ, Li MM, Hang YY, Xue JY.** Evolution of the KCS gene family in plants: the history of gene duplication, sub/neofunctionalization and redundancy. *Mol Genet Genomics*. 2016;**291**(2):739–752. <https://doi.org/10.1007/s00438-015-1142-3>
- Hall TA.** Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*. London: Information Retrieval Ltd., c1979-c2000; 1999. p. 95–98
- Han X.** In **Desiderio DM, Loo JA**, eds. *Lipidomics: Comprehensive Mass Spectrometry of Lipids*, Ed editors. London: Wiley; 2016
- Han G, Gable K, Kohlwein SD, Beaudoin F, Napier JA, Dunn TM.** The *Saccharomyces cerevisiae* YBR159w gene encodes the 3-ketoreductase of the microsomal fatty acid elongase. *J Biol Chem*. 2002;**277**(38):35440–9. <https://doi.org/10.1074/jbc.M205620200>
- Higashi Y, Okazaki Y, Myouga F, Shinozaki K, Saito K.** Landscape of the lipidome and transcriptome under heat stress in *Arabidopsis thaliana*. *Sci Rep*. 2015;**5**(1):10533. <https://doi.org/10.1038/srep10533>
- Hong J, et al.** Genetic architecture of seed glycerolipids in Asian cultivated rice. *Plant Cell Environ*. 2022. **46**(4):1278–1294. <https://doi.org/10.1111/pce.14378>
- Horton MW, et al.** Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet*. 2012;**44**(2):212–216. <https://doi.org/10.1038/ng.1042>
- Hummel J, Segu S, Li Y, Irgang S, Jueppner J, Giavalisco P.** Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. *Front Plant Sci*. 2011;**2**:54. <https://doi.org/10.3389/fpls.2011.00054>
- Joubès J, Raffaele S, Bourdenx B, Garcia C, Laroche-Traineau J, Moreau P, Domergue F, Lessire R.** The VLCFA elongase gene family in *Arabidopsis thaliana*: phylogenetic analysis, 3D modelling and expression profiling. *Plant Mol Biol*. 2008;**67**(5):547–566. <https://doi.org/10.1007/s11103-008-9339-z>
- Kao YT, Gonzalez KL, Bartel B.** Peroxisome function, biogenesis, and dynamics in plants. *Plant Physiol*. 2018;**176**(1):162–177. <https://doi.org/10.1104/pp.17.01050>
- Karimi M, Inzé D, Depicker A.** GATEWAY<sup>TM</sup> vectors for agrobacterium-mediated plant transformation. *Trends Plant Sci*. 2002;**7**(5):193–195. [https://doi.org/10.1016/S1360-1385\(02\)02251-3](https://doi.org/10.1016/S1360-1385(02)02251-3)
- Katoh K, Rozewicki J, Yamada KD.** MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;**20**(4):1160–1166. <https://doi.org/10.1093/bib/bbx108>
- Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO.** Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*. 2010;**26**(11):1463–1464. <https://doi.org/10.1093/bioinformatics/btq166>
- Kim J, Jung JH, Lee SB, Go YS, Kim HJ, Cahoon R, Markham JE, Cahoon EB, Suh MC.** Arabidopsis 3-ketoacyl-coenzyme a synthase9 is involved in the synthesis of tetracosanoic acids as precursors of cuticular waxes, suberins, sphingolipids, and phospholipids. *Plant Physiol*. 2013;**162**(2):567–580. <https://doi.org/10.1104/pp.112.210450>
- Kim J, Kim RJ, Lee SB, Suh MC.** Protein–protein interactions in fatty acid elongase complexes are important for very-long-chain fatty acid synthesis. *J Exp Bot*. 2022;**73**(9):3004–3017. <https://doi.org/10.1093/jxb/erab543>
- Kim J, Lee SB, Suh MC.** Arabidopsis 3-ketoacyl-CoA synthase 4 is essential for root and pollen tube growth. *J Plant Biol*. 2021;**64**(2):155–165. <https://doi.org/10.1007/s12374-020-09288-w>
- Kunz HH, Scharnewski M, Feussner K, Feussner I, Flügge UI, Fulda M, Gierth M.** The ABC transporter PXA1 and peroxisomal  $\beta$ -oxidation are vital for metabolism in mature leaves of *Arabidopsis* during extended darkness. *Plant Cell*. 2009;**21**(9):2733–2749. <https://doi.org/10.1105/tpc.108.064857>
- Lämke J, Brzezinka K, Altmann S, Bäurle I.** A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO J*. 2016;**35**(2):162–175. <https://doi.org/10.15252/embj.201592593>
- Law SR, Chrobok D, Juvany M, Delhomme N, Lindén P, Brouwer B, Ahad A, Moritz T, Jansson BS, Gardeström P, et al.** Darkened leaves use different metabolic strategies for senescence and survival. *Plant Physiol*. 2018;**177**(1):132–150. <https://doi.org/10.1104/pp.18.00062>
- Lees JM.** Geotouch: software for three and four dimensional GIS in the earth sciences. *Comput Geosci*. 2000;**26**(7):751–761. [https://doi.org/10.1016/S0098-3004\(99\)00133-8](https://doi.org/10.1016/S0098-3004(99)00133-8)
- Li-Beisson Y, et al.** Acyl-lipid metabolism. *Arabidopsis Book*. 2010;**8**. <https://doi.org/10.1199/tab.0133>
- Li H, Thrash A, Tang JD, He L, Yan J, Warburton ML.** Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *Plant J*. 2019;**98**(5):853–863. <https://doi.org/10.1111/tjp.14282>
- Lipka AE, et al.** GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012;**28**(18):2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Ludwig-Müller J.** Auxin conjugates: their role for plant development and in the evolution of land plants. *J Exp Bot*. 2011;**62**(6):1757–1773. <https://doi.org/10.1093/jxb/erq412>
- Luo J.** Metabolite-based genome-wide association studies in plants. *Curr Opin Plant Biol*. 2015;**24**(2):31–38. <https://doi.org/10.1016/j.pbi.2015.01.006>
- Matsuda F, Okazaki Y, Oikawa A, Kusano M, Nakabayashi R, Kikuchi J, Yonemaru JI, Ebana K, Yano M, Saito K.** Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J*. 2012;**70**(4):624–636. <https://doi.org/10.1111/j.1365-313X.2012.04903.x>
- Millar AA, Clemens S, Zachgo S, Michael Giblin E, Taylor DC, Kunst L.** CUT1, an arabidopsis gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell*. 1999;**11**(5):825–838. <https://doi.org/10.1105/tpc.11.5.825>
- Moraes TA, Mengin V, Annunziata MG, Encke B, Krohn N, Höhne M, Stitt M.** Response of the circadian clock and diel starch turnover to one day of low light or low CO<sub>2</sub>. *Plant Physiol*. 2019;**179**(4):1457–1478. <https://doi.org/10.1104/pp.18.01418>
- Nelson BK, Cai X, Nebenführ A.** A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J*. 2007;**51**(6):1126–1136. <https://doi.org/10.1111/j.1365-313X.2007.03212.x>
- Nobusawa T, Okushima Y, Nagata N, Kojima M, Sakakibara H, Umeda M.** Synthesis of very-long-chain fatty acids in the epidermis controls plant organ growth by restricting cell proliferation. *PLoS Biol*. 2013;**11**(4):e1001531. <https://doi.org/10.1371/journal.pbio.1001531>

- O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. Cistrome and episcistrome features shape the regulatory DNA landscape. *Cell*. 2016;**165**(5):1280–1292. <https://doi.org/10.1016/j.cell.2016.04.038>
- Perraki A, Cacas J-L, Crowet J-M, Lins L, Castroviejo M, German-Retana S, Mongrand S, Raffaele S. Plasma membrane localization of *Solanum tuberosum* remorin from group 1, homolog 3 is mediated by conformational changes in a novel C-terminal anchor and required for the restriction of potato virus X movement. *Plant Physiol*. 2012;**160**(2):624–637. <https://doi.org/10.1104/pp.112.200519>
- Pilkington SM, Encke B, Krohn N, Höhne M, Stitt M, Pyl ET. Relationship between starch degradation and carbon demand for maintenance and growth in *Arabidopsis thaliana* in different irradiance and temperature regimes. *Plant Cell Environ*. 2015;**38**(1):157–171. <https://doi.org/10.1111/pce.12381>
- Puranik S, Kumar K, Srivastava PS, Prasad M. Electrophoretic mobility shift assay reveals a novel recognition sequence for *Setaria italica* NAC protein. *Plant Signal Behav*. 2011;**6**(10):1588–1590. <https://doi.org/10.4161/psb.6.10.17130>
- Qiu X, Pang Y, Yuan Z, Xing D, Xu J, Dingkuhn M, Li Z, Ye G. Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. *PLoS One*. 2015;**10**(12):e0145577. <https://doi.org/10.1371/journal.pone.0145577>
- Raivo Kolde. pheatmap: Pretty Heatmaps. R package version 1.0.12. 2019.
- R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020. <https://www.R-project.org/>
- Riedelsheimer C, Brotman Y, Méret M, Melchinger AE, Willmitzer L. The maize leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits. *Sci Rep*. 2013;**3**(1):2479. <https://doi.org/10.1038/srep02479>
- Rossak M, Smith M, Kunst L. Expression of the FAE1 gene and FAE1 promoter activity in developing seeds of *Arabidopsis thaliana*. *Plant Mol Biol*. 2001;**46**(6):717–725. <https://doi.org/10.1023/A:1011603923889>
- Sant'Ana GC, et al. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. *Sci Rep*. 2018;**8**:465. <https://doi.org/10.1038/s41598-017-18800-1>
- Schumacher J, Kaufmann K, Yan W. Multiplexed GuideRNA-expression to efficiently mutagenize multiple loci in arabidopsis by CRISPR-Cas9. *Bio Protoc*. 2017;**7**(5):e2166. <https://doi.org/10.21769/BioProtoc.2166>
- Shin J-H, Blay S, McNeney B, Graham J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw*. 2006;**16**(3):1–10. <https://doi.org/10.18637/jss.v016.c03>
- Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. pcaMethods-a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*. 2007;**23**(9):1164–1167. <https://doi.org/10.1093/bioinformatics/btm069>
- Stitt M, Zeeman SC. Starch turnover: pathways, regulation and role in growth. *Curr Opin Plant Biol*. 2012;**15**(3):282–292. <https://doi.org/10.1016/j.pbi.2012.03.016>
- Szymanski J, Brotman Y, Willmitzer L, Cuadros-Inostroza Á. Linking gene expression and membrane lipid composition of *Arabidopsis*. *Plant Cell*. 2014;**26**(3):915–928. <https://doi.org/10.1105/tpc.113.118919>
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;**123**(3):585–595. <https://doi.org/10.1093/genetics/123.3.585>
- Thirumalaikumar VP, Devkar V, Mehterov N, Ali S, Ozgur R, Turkan I, Mueller-Roeber B, Balazadeh S. NAC Transcription factor JUNGBRUNNEN 1 enhances drought tolerance in tomato. *Plant Biotechnol J*. 2018;**16**(2):354–366. <https://doi.org/10.1111/pbi.12776>
- Todesco M, et al. Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature*. 2010;**465**(7298):632–636. <https://doi.org/10.1038/nature09083>
- Tonsor SJ, Alonso-Blanco C, Koornneef M. Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell Environ*. 2005. **28**(1):2–20. <https://doi.org/10.1111/j.1365-3040.2004.01264.x>
- Trenkamp S, Martin W, Tietjen K. Specific and differential inhibition of very-long-chain fatty acid elongases from *Arabidopsis thaliana* by different herbicides. *Proc Natl Acad Sci U S A*. 2004;**101**(32):11903–11908. <https://doi.org/10.1073/pnas.0404600101>
- Tresch S, Heilmann M, Christiansen N, Looser R, Grossmann K. Inhibition of saturated very-long-chain fatty acid biosynthesis by mefluidide and perfluidone, selective inhibitors of 3-ketoacyl-CoA synthases. *Phytochemistry*. 2012;**76**(4):162–171. <https://doi.org/10.1016/j.phytochem.2011.12.023>
- Wang P, Guo L, Morgan J, Dudareva N, Chapple C. Transcript and metabolite network perturbations in lignin biosynthetic mutants of *Arabidopsis*. *Plant Physiol*. 2022;**190**(4):2828–2846. <https://doi.org/10.1093/plphys/kiac344>
- Welti R, Li W, Li M, Sang Y, Biesiada H, Zhou HE, Rajashekar CB, Williams TD, Wang X. Profiling membrane lipids in plant stress responses: role of phospholipase D $\alpha$  in freezing-induced lipid changes in *Arabidopsis*. *J Biol Chem*. 2002;**277**(35):31994–32002. <https://doi.org/10.1074/jbc.M205375200>
- Wen W, Li K, Alseekh S, Omranian N, Zhao L, Zhou Y, Xiao Y, Jin M, Yang N, Liu H, et al. Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell*. 2015;**27**(7):1839–1856. <https://doi.org/10.1105/tpc.15.00208>
- Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J, et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun*. 2014;**5**(1):3438. <https://doi.org/10.1038/ncomms4438>
- Wu S, Alseekh S, Cuadros-Inostroza Á, Fusari CM, Mutwil M, Kooke R, Keurentjes JB, Fernie AR, Willmitzer L and Brotman Y. Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*. *PLoS Genet*. 2016;**12**(10):1–31. <https://doi.org/10.1371/journal.pgen.1006363>
- Wu S, Tohge T, Cuadros-Inostroza Á, Tong H, Tenenboim H, Kooke R, Méret M, Keurentjes JB, Nikoloski Z, Fernie AR, et al. Mapping the *Arabidopsis* metabolic landscape by untargeted metabolomics at different environmental conditions. *Mol Plant*. 2018;**11**(1):118–134. <https://doi.org/10.1016/j.molp.2017.08.012>
- Yang L, Wang Z, Zhang A, Bhawal R, Li C, Zhang S, Cheng L, Hua J. Reduction of the canonical function of a glycolytic enzyme enolase triggers immune responses that further affect metabolism and growth in *Arabidopsis*. *Plant Cell*. 2022;**34**(5):1745–1767. <https://doi.org/10.1093/plcell/koab283>
- Yu L, Fan J, Yan C, Xu C. Starch deficiency enhances lipid biosynthesis and turnover in leaves. *Plant Physiol*. 2018;**178**(1):118–129. <https://doi.org/10.1104/pp.18.00539>
- Zhu Feng, Alseekh S, Koper K, Tong H, Nikoloski Z, Naake T, Liu H, Yan J, Brotman Y, Wen W, et al. Genome-wide association of the metabolic shifts underpinning dark-induced senescence in *Arabidopsis*. *Plant Cell*. 2022;**34**(1):557–578. <https://doi.org/10.1093/plcell/koab251>