Hello darkness, my old friend: 3-Ketoacyl-Coenzyme A Synthase4 is a branch point in the regulation of triacylglycerol synthesis in Arabidopsis by re-channeling fatty acids under carbon starvation

Urszula Luzarowska^{1§}, Anne-Kathrin Ruß², Jérôme Joubès³, Marguerite Batsale³, Jędrzej Szymański⁴, Venkatesh Periyakavanam Thirumalaikumar², Marcin Luzarowski², Si Wu⁵, Feng Zhu^{2,6}, Niklas Endres², Sarah Khedhayir², Julia Schumacher⁷, Sandra Marcela Correa Cordoba¹, Aleksandra Skirycz², Alisdair Robert Fernie², Yonghua Li-Beisson⁸, Corina Mariana Fusari^{9*}, Yariv Brotman^{1*}

¹ Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel

² Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

³ Université de Bordeaux, Laboratoire de Biogenèse Membranaire, CNRS UMR 5200, Villenave d'Ornon, France

⁴ Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research, OT Gatersleben, 06466, Seeland, Germany

⁵ Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

⁶ National R&D Center for Citrus Preservation; Key Laboratory of Horticultural Plant Biology, Ministry of Education; Huazhong Agricultural University, Wuhan 430070, P.R. China

⁷ Department for Plant Cell and Molecular Biology, Institute for Biology, Humboldt-Universität zu Berlin, Philippstraße 13, 10115 Berlin, Germany

⁸ Aix Marseille Univ., CEA, CNRS, BIAM, Institute de Biosciences et Biotechnologies Aix-Marseille, F-13108 Saint Paul-Lez-Durance, France

⁹ Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET-UNR), Suipacha 570, S2000LRJ Rosario, Argentina

[§] Current address: Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

***Corresponding authors:**

Yariv Brotman: brotmany@post.bgu.ac.il; Corina M. Fusari: fusari@cefobi-conicet.gov.ar

Running title: KCS4 impacts lipid remodeling under C-starvation

Keywords: Arabidopsis, GWAS, natural variation, triacylglycerols, fatty acids, KCS4, carbon starvation

One sentence summary:

Natural variation at *KCS4* underlies a differential accumulation of polyunsaturated triacylglycerols, by acting as a regulatory branch point in the fate of fatty acids under carbon starvation.

Abstract

Due to their sessile lifestyle, plants have evolved unique mechanisms to deal with environmental challenges. Under stress, plant lipids are important as alternative sources of carbon and energy when sugars or starch are limited. Here, we applied combined heat and darkness and extended darkness to a panel of ~ 300 Arabidopsis accessions to study lipid remodeling under carbon starvation. Natural allelic variation at 3-KETOACYL-COENZYME A SYNTHASE4 (KCS4), a gene encoding for an enzyme involved fatty-acid elongation, underlies a differential accumulation of polyunsaturated triacylglycerols (TAGs) under stress. Ectopic expression in yeast and plants proved that KCS4 is a functional enzyme localized in the ER with specificity for C22 and C24 saturated acyl-CoA. Loss-of-function mutants and transient overexpression *in planta* revealed the role of KCS4 alleles in TAG synthesis and biomass accumulation. The region harboring KCS4 is under high selective pressure. Furthermore, allelic variation at KCS4 correlated 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 to abiotic stress and the evolutionary events shaping the lipidome under carbon starvation.

Introduction

Plant metabolism profoundly influences or even co-ordinates 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. 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 the stress and is crucial to acclimation and survival (Fraire-Velázquez and Balderas-Hernández, 2013).

Plant metabolism has been investigated by genome-wide association studies (GWAS), and the genetic determinants of key metabolic traits have been successfully characterized (i.e. primary and secondary metabolites and enzyme activities) (Wu et al., 2018, 2016; Fusari et al., 2017). Regarding lipid metabolism, a number of GWAS have been reported, for instance in Arabidopsis, maize, coffee, and canola, resulting in the identification of new enzymes and regulators (Riedelsheimer et al., 2013; Li et al., 2019; Sant'Ana et al., 2018; Gacek et al., 2017). These studies focused on lipid levels under optimal growth conditions. However, accumulation of lipids is highly affected by fluctuations in the environmental conditions, such as light, temperature and nutrient availability (Burgos et al., 2011; Szymanski et al., 2014; Caldana et al., 2011). A classic example involves compensation for the 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, starch, the major source of energy in the dark, 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. Under these conditions, polyunsaturated fatty acids (FA) released from membrane lipids are used as an alternative substrate for respiration via β -oxidation. Free fatty acids (FFAs) have detergent-like properties and are often sequestered into triacylglycerols (TAGs). TAG synthesis and lipid-droplet formation sometimes constitute a cellular response to lipotoxicity. This is especially true for polyunsaturated fatty acid (PUFA) species because they contain multiple double bonds and are liable to production of reactive oxygen species (ROS). 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, but also involving activation of specific stress-induced isozymes (Szymanski et al., 2014). All these observations indicate that only by causing a perturbation of the metabolic homeostasis are GWAS likely to capture genes that evolved to stand at critical branch points in the regulation of the lipidome under natural fluctuating environments.

Here, we performed GWAS using lipidomic data obtained in three different environments: control, combined heat and darkness, and extended darkness. We observed that many associations were detected in specific conditions. The locus harboring *KCS4*, an as yet uncharacterized gene encoding for a 3-KETOACYL-COENZYME A SYNTHASE (Joubès et al., 2008), was strongly associated with a high number of polyunsaturated TAGs in stress conditions, but was not detected in control condition. Thus, *KCS4* association is driven by carbon starvation. Using loss-of-function mutants and ectopic expression in yeast and plants we explored the biochemical properties of *KCS4* and validated its role in polyunsaturated TAG metabolism. Taken together, we place a model for *KCS4* function, where *KCS4* acts as a regulatory branch point in the fate of saturated FAs, directing them either to the very-long chain fatty acid pathway (VLCFA) or to TAG. Two *KCS4* alleles have differential capacity to process saturated FA, resulting in higher levels of polyunsaturated TAGs in accessions carrying the strong allele. 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 growth.

Results

Plasticity in Arabidopsis lipid metabolism under different environmental conditions

In order 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 two environmental conditions: control (21/16°C in 16/8h 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 a total of 98 and 109 glycerolipids in CHD and HD, respectively (Supplementary Table 1).

Principal component analysis (PCA) of lipid data clearly separates the natural ecotypes into two different groups according to the two environmental conditions tested (Supplementary Figure 1a). Furthermore, differences in the lipid status between CHD and HD samples showed that TAGs were strongly accumulated while MGDGs and DGDGs were mostly decreased under stress (Supplementary Figure 1b). In addition, there is substantial positive correlation between lipid levels from a given lipid class, suggesting coordinated regulation at the level of lipid classes (Supplementary Figure 1c). 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 (Supplementary Figure 1d).

Genetic basis underlying Arabidopsis lipid metabolism

Lipid intensities from CHD and HD and the 250K SNPchip (Horton et al., 2012) were used to perform two independent genome-wide association studies (GWAS) for the identification of lipid quantitative trait loci (QTL). The GWAS revealed 93 and 544 SNP-trait associations in control and stress conditions, respectively, at a genome-wide significance level of $\text{LOD} \geq 5.3$ (Supplementary Table 2). These SNP-trait associations further grouped into 51 and 66 QTL for CHD and HD conditions, respectively. In total, only five QTL were shared between CHD and HD, revealing the great plasticity of the lipidome in response to different environmental conditions (Supplementary Table 3).

Regarding functional annotation of candidate genes, we mapped 12 and 22 genes putatively involved in lipid metabolism for control and stress condition, respectively (Supplementary Table 3, Acyl-Lipid Metabolism database (Li-Beisson et al., 2010) <http://aralip.plantbiology.msu.edu/>). After filtering for robust associations co-localizing for two or more traits, four QTL harbored eight lipid-genes (Figure 1a), including a *LIPASE CLASS 3* (AT1G06250), two *ACYL-COA OXIDASE* (ACX3: AT1G06290 and ACX6: AT1G06310), two *FATTY ACID DESATURASE* (CER17: AT1G06350 and FAD2:AT3G12120) and a *LYSOPHOSPHOLIPASE* (AT5G20060) (Supplementary Table 3, Figure 1a).

A QTL harboring an enzyme involved in fatty acid elongation has major impact on TAG levels under heat and dark stress

Remarkably, 22 TAGs co-localized to a single QTL on chromosome 1, harboring *3-KETOACYL-COENZYME A SYNTHASE4* (*KCS4*, AT1G19440, Figure 1a). This QTL was not found in control conditions, and the lead SNP in the locus (m11481) co-localized with LOD ≥ 8 (Bonferroni corrected $p < 0.05$) for 16 of the 22 associated TAGs (e.g. Figure 1b). An additional 33 genes mapped on the QTL, spanning 107,318 bp (Supplementary Table 3). However, with the exception of *KCS4*, none of the others showed any homology 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 for enzymes involved in the first and often described as rate-limiting step of fatty-acid elongation (FAE) (Blacklock and Jaworski, 2006; Joubès et al., 2008). *KCS4*'s closest homologues are: *KCS9* (78% aa identity), involved in the elongation of C22 to C24 acyl-CoAs (Kim et al., 2013); *KCS17* (72% aa identity); and *KCS18* (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 (Patrick and Colyvas, 2014; Rossak et al., 2001). However, *KCS4* specificity and function have not been characterized so far.

TAGs associated with the *KCS4* QTL were mostly polyunsaturated (e.g. 52:6, 52:7, 54:6, 54:8, 54:9, 56:6) and correlated positively with each other, grouping together in a single very-well defined cluster (Figure 1c).

To identify the alleles involved in TAG variation under HD, we constructed haplotypes considering the four SNPs from *KCS4* (m11502, m11503, m11504, m11505). The population divided into nine non-unique haplotypes, which were therefore, further considered as being putatively involved in the QTL variation. Mean levels of TAG 54:9 varied significantly among haplotypes (Figure 1d). Furthermore, m11503 was identified as a non-synonymous SNP (nsSNP). In the majority of accessions, SNP m11503 encodes for a methionine [hereafter *KCS4*(Met)]. Accessions carrying this allele have higher levels of TAG in the stress condition than accessions carrying the minor allele, which encodes for a threonine [hereafter *KCS4*(Thr)] (Figure 1e). These results suggest that the nsSNP in *KCS4* is a causal mutation underlying polyunsaturated TAG variation in Arabidopsis, under stress conditions.

***KCS4* associates with polyunsaturated TAGs in extended darkness**

Thus far, we found that *KCS4* associates with polyunsaturated TAGs in a combined heat and dark treatment. In prolonged darkness, starch is prematurely exhausted and plants are carbon starved (Stitt and Zeeman, 2012). Therefore, we determined whether *KCS4* associates with TAGs, when only extended darkness is applied.

For this purpose, we conducted three additional independent lipid GWAS experiments: control (CD), 3 days (3D) extended darkness, and 6 days (6D) extended darkness (Supplementary Table 1). The GWAS at a genome-wide significance level of LOD ≥ 5.3 resulted in 230, 255 and 342 SNP-trait associations grouped in 73, 81, and 70 QTL for CD, 3D, and 6D, respectively (Supplementary Table 2,

Supplementary Table 3). The locus harboring *KCS4* associated strongly in 3D and 6D with 8 and 13 TAGs, respectively. Lead SNP and other features of the association matched with GWAS results from the HD experiment (Supplementary Figure 2). Therefore, darkness alone is enough to trigger the association of *KCS4* with polyunsaturated TAGs.

Dissection of *KCS4* alleles involved in TAG variation, pattern of linkage disequilibrium 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 (Supplementary File 1) was used to build a SNP matrix (Supplementary Table 4) to perform GWAS with more SNPs than those included in the previous experiments using the 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 linkage disequilibrium (LD) in the region, determine whether other SNPs in *KCS4* are additionally involved in the variation, and evaluate signatures of selection. The results are shown in Figure 2, exemplified for TAG 54:9. According to the *KCS4* gene model (Figure 2a), we found 35 additional SNPs in *KCS4* apart from the 4 SNPs in the SNPchip (19 in the promoter, two in the 5'UTR, 13 in the coding region, and one in the 3'UTR; Supplementary Table 5). Overall, fewer TAGs with lower association values co-localized for 3D than for HD, mimicking GWAS results obtained using the 250K SNPchip (Supplementary Table 5, Figure 2b). This suggests that there is an additive or synergistic effect of both heat and dark stresses in the regulation of TAGs mediated by *KCS4*. 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 (Figure 2a). Furthermore, several SNPs in the *KCS4* coding region surprisingly did not associate with the QTL on chromosome 1 (Figure 2b).

The pattern of LD showed that all SNPs associated to the QTL in chromosome 1, are in high LD with each other and in high LD with the lead SNPs rs21933 and m11481 ($r^2 > 0.2$, $p < 0.001$, Figure 2c). The remaining SNPs in the *KCS4* coding region are not in LD either with the lead SNPs or with the nsSNP m11503 in the *KCS4* coding region. This result suggests that these mutations may have appeared in 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 (Figure 2d). Tajima's D reached significantly negative values along the 80 kb, particularly around SNPs significantly associated in the QTL: rs21933, m11481, rs39451, rs39499, m11502, m11503. This result indicates that the region is under a high selective pressure.

KCS4* functional validation using knockout mutants in *Arabidopsis* and transient overexpression in *Nicotiana benthamiana

To validate the role of *KCS4* alleles in TAG variation under HD stress, we selected six accessions: three carrying *KCS4*(Met) and three carrying *KCS4*(Thr). Selection of these accessions was based on TAG levels measured in the HD experiment. Knockout (KO) lines were obtained through CRISPR-Cas9 by introducing a 57/58-bp deletion at position +18 from the ATG (Supplementary File 2). 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 conditions, and harvested in two independent experiments (Exp1, Exp2; Supplementary Table 6). Since natural variation is detected at

a population level, TAGs were analyzed by comparing the average behavior of *kcs4* and wild-type *KCS4* accessions, according to the allele carried (Figure 3). Accessions carrying *KCS4*(Met) presented—on average—significantly higher polyunsaturated TAG levels under HD conditions than accessions carrying *KCS4*(Thr). *Kcs4*(Met) and *kcs4*(Thr) did not show any significant difference compared to *KCS4*(Met) and *KCS4*(Thr), respectively (Figure 3a/3b). The fact that knocking out *KCS4* does not seem to produce an effect on TAG levels could be due to a feedback loop activated on other members of the *KCS* family to compensate for *KCS4* activity, when there is complete loss of the *KCS4* transcript.

On the other hand, transient expression of At*KCS4*(Met) in *N. benthamiana* leaves shows significantly higher levels of polyunsaturated TAGs for classes 52, 54, and 56 compared to control plants (Figure 3c). Conversely, TAGs with one or a few unsaturated bonds tend to be higher in control plants. These results show that the transient expression of At*KCS4*(Met) changes the TAG profile in *N. benthamiana*. Saturated FAs are channeled to VLCFA elongation, tipping the balance to a higher polyunsaturated-to-saturated-FA ratio and driving the production of polyunsaturated TAGs at the expense of saturated TAGs.

Transcriptional regulation of *KCS4* in Arabidopsis

Several SNPs in the promoter region induced us to attempt to identify the transcriptional regulators of *KCS4* in Arabidopsis. We analyzed 1 kb upstream of the transcription start site for the known transcription factor binding sites using the DAP-Seq database (O'Malley et al., 2016). We found several transcription factor (TF) families belonging to land plants that could likely bind to the *KCS4* promoter region, indicating that *KCS4* expression might be controlled under several regulatory processes (Figure 4a).

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 within the SNP region, but in other regions of the promoter (Figure 4b). Similarly, the NAC and WRKY TF families bind to non-polymorphic regions in the *KCS4* promoter, and MYB TF binds near a SNP but in the coding region. Interestingly, HSF (heat shock factor), the second highly enriched TF family has a core binding motif (heat shock element: 5'nnGnAnnTnCtn 3') within SNP m11502 in the promoter region.

Next, we performed an electrophoretic mobility shift assay (EMSA) to assess the physical interaction of Heat Shock Factor A2 (HSFA2) with the *KCS4* promoter sequence. Additionally, we tested if alleles (C/T) at SNP m11502 influence the binding affinity. *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 Figure 4c, 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 (Figure 4c, 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, 200x more) significantly reduced the binding to the labeled probe (Figure 4c, 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 (Figure 4c, lane 6), suggesting that the T allele at m11502 determines a stronger binding of HSF2A to the *KCS4* promoter region. Overall, this indicates that natural variation at the promoter region could contribute to the final effect and function of the different *KCS4* alleles.

Furthermore, we analyzed natural variation in transcript levels of *KCS4* for thevCHD and HD conditions in the complete GWAS panel (Figure 4d). *KCS4* transcripts are 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. Figure 4e, TAG 54:9). In HD, *KCS4* transcript levels are not significantly different for accessions carrying different alleles at m11502 (Figure 4e).

To gain further insight on the regulation of *KCS4* transcript, we performed expression QTL (eQTL) analysis using *KCS4* expression data for the complete GWAS population in CHD and HD conditions (Supplementary Table 7). We identified 29 eQTL for a LOD ranging from 5 to 6. We detected a QTL harboring *MYB96* (AT5G62470) in the CHD condition. It was shown that *MYB96* acts as a transcriptional activator of the *KCS* genes(Seo et al., 2011). Interestingly, the *ACX3* (AT1G06290) and *ACX6* (AT1G06310) genes were also found associated with *KCS4* transcript levels in HD (Supplementary Table 7). The first step of peroxisomal FA β -oxidation is catalyzed by *ACX* enzymes(Khan et al., 2012; Eastmond et al., 2000). *ACX3* has specificity for medium-chain FA, and its function has also been linked to the indole-3-butyric acid (IBA) to indole-3-acetic acid (IAA) β -oxidation in the peroxisome (Adham et al., 2005). This result might indicate a regulatory feedback among IAA response to stress, FA degradation and VLCFA synthesis mediated by *ACX* genes and *KCS4* transcript.

Expression patterns of the KCS gene family in eight environmental conditions

To evaluate the effect of heat and darkness on the expression of *KCS4* and to compare expression patterns in different environmental conditions across the *KCS* family, we analyzed the time-resolved response of Arabidopsis Col-0 accession towards light and temperature for 16 of the 21 gene members using previously obtained data (Caldana et al., 2011). Because of the complexity of the response, we focus on the last time-point, i.e. 24 h after treatment (Figure 4f). Most *KCS* members, including *KCS4* are upregulated under light conditions and downregulated in darkness, irrespective of the temperature. However, *KCS8*, *KCS12*, *KCS16*, *KCS18*, and *KCS21* showed an opposite behavior. *KCS4* transcript levels are downregulated both under darkness and under HD, but they are much lower in HD, reinforcing the idea of a synergistic effect between light and temperature stress in the *KCS4* response. In addition, *KCS5*, *KCS9* and *KCS20* showed a similar transcriptional response as *KCS4* across several environmental stimuli (Figure 4f).

KCS4 shows specificity for elongation of VLCFA up to 26-carbon chains

KCS4 shows sequences homology 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 determine the substrate specificity.

To evaluate if *KCS4* has a very-long-chain 3-ketoacyl-CoA synthase activity and to investigate its substrate specificity, we heterologously expressed *AtKCS4*(Met) in yeast and evaluated the VLCFA profile. The strain expressing *AtKCS4* synthesized significantly higher levels of VLCFA from C20 to C26 compared to the strain transformed with an empty vector (Figure 5a). Because yeast has its own well-described machinery for VLCFA synthesis, we further expressed *AtKCS4* in a yeast strain that lacks the *ELO3* gene encoding elongase ($\Delta elo3$, Figure 5b). Because the yeast *ELO3* elongase is essential for the elongation of VLCFA from C22 to C24 and C26 (Han et al., 2002), the $\Delta elo3$ mutant showed high accumulation of C20 and C22 VLCFA. The $\Delta elo3$ mutant expressing *AtKCS4*(Met) accumulated substantial amounts of C24 and C26 VLCFA, confirming that *KCS4* is involved in the elongation of VLCFAs from C22 to C24 and from C24 to C26 (Figure 5b). We also analyzed *KCS9* heterologous expression in yeast (Figure 5b). *KCS9* is the closest *KCS4* homolog and its specificity is to elongate VLCFAs from C22 to C24 (Kim et al., 2013). The $\Delta elo3$ mutant expressing *KCS9*, as

expected, showed a peak in C24 accumulation. These results suggest that KCS4 has a differential activity compared to KCS9, pushing further the elongation of VLCFAs up to C26.

Transient expression of *AtKCS4* in *N. benthamiana* leaves showed a significant decrease of C20 and C22 VLCFAs and a significant increase of VLCFAs from C26 up to C30 in comparison to the control (Figure 5c). The elongation of VLCFAs from C26 to C30 in *N. benthamiana* might be the result of the action of the native elongase complex due to increased availability of C26 when *KCS4* is transiently expressed.

KCS4 is localized in the endoplasmic reticulum (ER)

In plants, the elongase complex is localized in the ER (Joubès et al., 2008). To monitor KCS4 localization we fused YFP to the C-terminal end of *AtKCS4* and visualized it with confocal microscopy (Figure 5d).

AtKCS4-YFP was found to label a polygonal and tubular network characteristic of the endoplasmic reticulum (ER), in agreement with previous results shown for other KCS family members (Joubès et al., 2008).

Biomass is affected by lipid rearrangement mediated by KCS4 in stress conditions

It has been demonstrated that a mild reduction in VLCFA content improve cell proliferation and shoot growth (Nobusawa et al., 2013). To investigate if knocking out *KCS4* has such an effect, we compared biomass of *kcs4* mutants and their respective WT for the accessions Pro-0 [*KCS4*(Met)] and Sh-0 [*KCS4*(Thr)].

The *kcs4* lines, grew significantly bigger than their WT counterparts, indicating that biomass is negatively correlated with KCS4 activity (Figure 6).

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

To test if there is 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 six TAGs with a significant phylogenetic signal are polyunsaturated species associated with the *KCS4* QTL (Supplementary Figure 3).

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, five of which gave also significant results in a cross-validation test (Supplementary Figure 4). All of these five significant variables are related to temperature. In total, six “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 quarter 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 non-parametric ANOVA. Eight climate parameters plus latitude and longitude differed significantly between haplotypes ($p < 0.05$,

Figure 7a). Moreover, five of them co-clustered with TAGs associated with *KCS4*, four of them related to precipitation and one to temperature (isothermality). The “TT” haplotype in most cases, presented the lowest value for these parameters, suggesting that the allelic effect of these two SNPs is opposite for climate variables (Figure 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 different environmental conditions. 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 (Figure 1a).

Lipid-based GWAS in control conditions showed several mild-strength associations, similar to the pattern obtained for primary metabolites (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 are controlled, to a large extent, by major large-effect loci (Wu et al., 2018; Chan et al., 2011; Wen et al., 2014; Matsuda et al., 2012).

In our study, applying stress resulted in a strong perturbation of lipid homeostasis. This resulted in the identification of a large-effect QTL at the *KCS4* gene. *KCS4* associated with several polyunsaturated TAGs specifically under stress conditions (HD, 3D and 6D) in a highly robust manner (Figure 1b, Supplementary Figure 2b). Under HD, all TAGs mapping to *KCS4* are higher than under CHD. In contrast, under 3D there is a strong reduction in polyunsaturated TAGs in comparison to CD. Under 6D, levels are similar to the TAG values found under CD. Irrespective of this, accessions carrying *KCS4*(Met) always presented higher levels of polyunsaturated TAGs than accessions carrying *KCS4*(Thr). Polyunsaturated TAGs that mapped to *KCS4* clustered together and negatively correlated to saturated TAGs, further illustrating the role of *KCS4* in the specific regulation of polyunsaturated TAGs (Figure 1c).

Repression of several KCS family members has been shown to occur during darkness due to the activation of a TF that downregulates the wax biosynthetic pathway (Go et al., 2014; Kim et al., 2018). We, too, observed downregulation of a number of KCSs under darkness alone, and under combined HD (Figure 4f), including *KCS4* (Figure 4d). These findings are in line with the notion that VLCFA elongation slows down upon stress in order to re-allocate carbon resources under carbon-limiting conditions.

So far, *KCS4* enzymatic characterization has not yet been achieved. Our experiments indicate that *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 (Figure 5). Moreover, the results from ectopic expression in yeast and plants confirmed that *KCS4* has specificity for elongation of saturated FAs (Figure 5), in agreement with previous findings that shows 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*(Met) vs. *KCS4*(Thr)] become more significant as the number of double bonds in the TAGs increases (Figure 3a). However, TAG levels are not significantly different when comparing knock-outs to their wild-type counterparts. Limitations in the study of knock-out mutants have been previously pointed out. For instance, the phenotype can 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). In addition, sometimes effects are not evident when analyzing single mutants, particularly if there are gene families encoding for the same or very similar enzymatic activities. For example, in studies of auxin conjugates, IAA showed a reduction only when triple amidohydrolase mutants were analyzed (Ludwig-Müller, 2011). *KCS4* is one of 21 genes encoding for KCS activity (Guo et al., 2016; Joubès et al., 2008). Despite each KCS having specificity for VLCFAs with different acyl-chain lengths, there is some redundancy and overlap (Blacklock and Jaworski, 2006). This redundancy is also observed in their expression patterns (Figure 4f).

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 two major haplotypes and that the region is under strong selective pressure (Figure 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 does 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 an adaptative role of *KCS4* to environmental stresses, the geographical origin of the accessions (i.e. latitude) correlated with *KCS4* haplotypes (Figure 7). Together, these analyses indicate that polymorphism variation at *KCS4* is important for adaptation in different climatic zones by adjustment of lipid metabolism.

To summarize our findings, we present a model of lipid remodeling under heat and dark stress in *Arabidopsis* (Figure 8). In an event of sudden extended darkness, plants become carbon starved (Stitt and Zeeman, 2012). It has been shown that increasing night temperature accelerates enzyme activities exhausting carbon resources prematurely (Pilkington et al., 2015). In this case, FAs are used as alternative substrates for respiration via β -oxidation (Fan et al., 2014). In the proposed model, darkness induces the predominant response within the combined stress, blocking *de novo* FA synthesis (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 (Fan et al., 2017; Higashi et al., 2015). 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 very-long-chain fatty-acid pathway (VLCFA). 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). Polyunsaturated TAG levels are, therefore, higher in *KCS4*(Met) accessions comparing to *KCS4*(Thr) accessions, as the result of a rise in the ratio of polyunsaturated-to-saturated FAs that proceed to TAG formation. This effect of *KCS4*(Met) on TAG levels is only observed under stress conditions, when there is an increased amount of free-polyunsaturated fatty acids coming from membrane lipids, and when it necessary to sequestered them into TAGs, to protect the cell from the free-fatty-acid-toxicity induced damage (Kao et al., 2018). A number of additional regulatory points have been identified to influence VLCFA production, and more general lipid remodeling under HD. For instance, as the wax pathways is downregulated due to the activation of DEWAX TF under darkness (Go et al., 2014; Kim et al., 2018), VLCFA-CoA could directly serve as a source of carbon, and undergo degradation in the peroxisome. In addition, *KCS4* transcripts mapped to *ACX3* and *ACX6*, indicating a genetic interaction between these two loci. As *ACXs* are key enzymes for FA degradation, and they have been also implicated in the IBA to IAA conversion in the peroxisome, this interaction might help both to adjust the energy demand of the cell under carbon starvation and to activate auxin responses against stress.

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 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 re-channeling under stress and proved to be powerful in unraveling the genetic regulation of complex phenotypes.

Methods

Plant material

The GWA mapping population comprised a total of 330 *Arabidopsis 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).

Experimental design

Accessions or mutant lines were grown in soil (potting compost) in short days (8 h light) for four weeks, then transferred to long days (16 h light) for an additional period of two weeks. 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 in long days (16h light, 8 h dark), temperature was kept at 21/16°C (day and night, respectively), light intensity at 150 $\mu\text{E m}^{-2} \text{sec}^{-1}$, humidity 75%. In the heat and dark experiment (HD), plants were grown like the plants in CHD for six weeks, 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 six weeks and then transferred to continuous darkness either for 72 h (3D) or 144 h (6D). Plants from HD, 3D, 6D were harvested 24 h, 72 h, and 144 h after the beginning of the treatment, respectively. Between three and six plants from each accession/line were collected for each condition tested.

Targeted lipid profiling by LC–MS and data acquisition

Lipid extraction and measurement from *Arabidopsis* leaves 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 to 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 mean of signal intensities lower than 50000 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), digalactosyldiacylglycerols (DGDGs), monogalactosyldiacylglycerols (MGDGs), phosphatidylcholines (PCs), phosphatidylethanolamines (PEs) and triacylglycerols (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 minutes and 10 ppm. Identified lipids were confirmed by manual verification of the chromatograms using XCalibur (Version 3.0, Thermo-Fisher, Bremen, Germany).

General statistical and multivariate analysis of lipidomic data

To calculate the correlation between lipids, 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. Principal component analysis (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 five highest intensities measured in a given accession by the average of the five lowest intensities, A) for all annotated lipids ($N_{CHD} = 98$, $N_{HD} = 109$), B) only for TAGs ($N_{CHD} = 30$, $N_{HD} = 41$), C) for lipid classes other than TAGs ($N_{CHD} = 68$, $N_{HD} = 68$).

GWAS analysis

Data acquisition for GWAS, mapping and locus identification were performed as described previously (Wu et al., 2016; Fusari et al., 2017). Genome-wide association analysis (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 (Supplementary Table 1), whereas a set of 214,051 SNPs was used as genotypic data (250K SNPchip). A total of three 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 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/gebrowser.php>, 162 accessions). The alignment was performed using the default routine implemented in MAFFT (Katoh et al., 2019) with a final manual inspection to discard alignment problems due to big gaps or missing data (Supplementary file 1). For nine accessions, two sequences were available. The sequences included further in the analysis were selected based on agreement of alleles with the SNPchip data and/or on % of non-missing data ($\geq 90\%$). Final SNP matrix, obtained by filtering for minor allele frequency $\geq 5\%$, contained 1,213 SNPs and indels (Supplementary Table 4). 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 250K SNPchip for the subset of 162 accessions (Supplementary Table 4). Lipid features collected formerly (HD and 3D) Supplementary Table 1) were used as phenotypes. The three 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, SNPs associated with lipid species having a p value $< 1/N$ ($N=214,051$) were extracted. SNPs were assigned to the same QTL if the genomic distance between them was ≤ 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 based on 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).

Linkage disequilibrium, nucleotide diversity, and haplotype analysis

Linkage disequilibrium (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 the ATG, the coding region and 1,000 bp downstream 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 KO lines, the binary vector containing multiple gRNAs and the Cas9 nuclease were generated based on a classic cloning procedure described previously (Schumacher et al., 2017). A pair of guideRNAs was used to target the N-terminal coding sequence of *KCS4* using the spacer sequences indicated in Supplementary Table 8. 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 Supplementary Table 8. Sequence contigs and alignment were done using Bioedit (Hall, 1999).

For heterologous expression in yeast and transient expression in *Nicotiana benthamiana*, *KCS4* and *KCS9* open reading frame was amplified from Arabidopsis flower cDNA using primers listed in Supplementary Table 8. 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; Dittrich-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 five-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 days 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 *Δelo3* [MATa *his3Δ1 leu2 trp1-289 ura3-52 elo3Δ*] were used to express *KCS4* and *KCS9*.

qRT-PCR and eQTL analysis

RNA extraction followed by cDNA synthesis and qRT-PCR were performed as described before (Thirumalaikumar et al., 2018). *ACTIN2* (AT3G18780) was used as a reference gene for *HSFA2/KCS4* transcript analysis in Col0 and *hsfa2* mutant. *RGS1-HXK1 INTERACTING PROTEIN1/RHIP1* (AT4G26410) was used as reference gene for *KCS4* transcript analysis in CHD and HD for the complete GWAS population (Czechowski et al., 2005). Primers used to test *KCS4*, *HSFA2* and reference genes are listed in Supplementary Table 8.

qRT-PCR was performed on samples made out of three biological replicates pooled together. Each sample was tested in three technical replicates. qRT-PCR 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 µL (2 µL forward and reverse primer mixture, 0.5 µM each), 0.5 µL cDNA and 2.5 µL 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, 60°C for 15 s, followed by a continuous temperature increase (0.3°C per s) 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 non-specific amplification. Expression values were calculated by $2^{(-\Delta\Delta Ct)}$ using either *ACTIN2* or *RHIP1* as reference genes. Errors are given as the lower ($2^{-(\Delta\Delta Ct-SD)}$) and upper ($2^{-(\Delta\Delta Ct+SD)}$) limit of the standard deviation around the mean.

eQTL analysis was performed using *KCS4* expression data measured in CHD and 24 h after HD for the complete GWAS population (Supplementary Table 7). ΔCt was used as trait for the mapping and it was calculated using *RHIP1* as reference gene. The GWAS was performed as described in “GWAS analysis” section.

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 Supplementary Table 8. 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 200x the amount of probe.

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

For fatty acid methyl esters (FAME) in *N. benthamiana*, three agroinfiltrated leaf discs were transferred to tubes containing 1 mL 0.5 M sulfuric acid in methanol containing 2% (v/v) dimethoxypropane and 20 µg of heptadecanoic acid (C17:0) as internal standard. Transmethylation was performed at 85°C for 3 hours. After cooling, 2 mL of 2.5% (w/v) NaCl was added, and fatty acyl chains were extracted in 300 µL of hexane. Samples were analyzed by GC-FID and GC-MS as previously described (Domergue et al., 2010).

For FAME analysis in yeast, wild-type *Saccharomyces cerevisiae* or INVSc1 strains transformed with *KCS4*, *KCS9* or the empty vector (EV), were grown in 5 mL of appropriate liquid minimal medium for 1 week at 30°C, then pelleted and washed in 2 mL of 2.5% (w/v) NaCl. Yeast cell pellets were resuspended in 1 mL 0.5 M sulfuric acid in methanol containing 2% (v/v) dimethoxypropane and 20 µg of heptadecanoic acid (C17:0) as internal standard and transmethylation was performed at 85°C for 1 hour.

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 63x 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 490–540 nm.

Biomass determination

Plants (WT and KO lines) from Pr0 and Sh0 accessions were grown as described in experimental design for CHD and immediately weighed after harvested.

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 non-parametric 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 ‘λmin + 1SE’-rule was applied to choose the penalty parameter for each model separately. Prediction power was estimated in 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.

References

- Adham, A.R., Zolman, B.K., Millius, A., and Bartel, B.** (2005). Mutations in Arabidopsis acyl-CoA oxidase genes reveal distinct and overlapping roles in β-oxidation. *Plant J.* **41**: 859–874.
- Anderson, M.J.** (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**: 32–46.
- Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J., Borevitz, J.O., Nordborg, M., Vitek, O., and Salt, D.E.** (2010). A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genet.* **6**: e1001193.
- Blacklock, B.J. and Jaworski, J.G.** (2006). Substrate specificity of Arabidopsis 3-ketoacyl-CoA synthases. *Biochem. Biophys. Res. Commun.* **346**: 583–590.
- Blomberg, S.P., Garland Jr, T., and Ives, A.R.** (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution (N. Y.)*. **57**: 717–745.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S.** (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–5.
- Burgos, A., Szymanski, J., Seiwert, B., Degenkolbe, T., Hannah, M.A., Giavalisco, P., and Willmitzer, L.** (2011). Analysis of short-term changes in the Arabidopsis thaliana glycerolipidome in

response to temperature and light. *Plant J.* **66**: 656–668.

Caldana, C., Degenkolbe, T., Cuadros-Inostroza, A., Klie, S., Sulpice, R., Leisse, A., Steinhauser, D., Fernie, A.R., Willmitzer, L., and Hannah, M.A. (2011). High-density kinetic analysis of the metabolomic and transcriptomic response of *Arabidopsis* to eight environmental conditions. *Plant J.* **67**: 869–884.

Chan, E.K.F., Rowe, H.C., Corwin, J.A., Joseph, B., and Kliebenstein, D.J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol.* **9**.

Correa, S.M., Fernie, A.R., Nikoloski, Z., and Brotman, Y. (2020). Towards model-driven characterization and manipulation of plant lipid metabolism. *Prog. Lipid Res.*: 101051.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.*

Dittrich-Domergue, F., Joubès, J., Moreau, P., Lessire, R., Stymne, S., and Domergue, F. (2014). The bifunctional protein TtFARAT from *Tetrahymena thermophila* catalyzes the formation of both precursors required to initiate ether lipid biosynthesis. *J. Biol. Chem.* **289**: 21984–21994.

Domergue, F., Vishwanath, S.J., Joubès, J., Ono, J., Lee, J.A., Bourdon, M., Alhattab, R., Lowe, C., Pascal, S., and Lessire, R. (2010). Three *Arabidopsis* fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. *Plant Physiol.* **153**: 1539–1554.

Eastmond, P.J., Hooks, M.A., Williams, D., Lange, P., Bechtold, N., Sarrobert, C., Nussaume, L., and Graham, I.A. (2000). Promoter trapping of a novel medium-chain acyl-CoA oxidase, which is induced transcriptionally during *Arabidopsis* seed germination. *J. Biol. Chem.*

Fan, J., Yan, C., Roston, R., Shanklin, J., and Xu, C. (2014). *Arabidopsis* lipins, PDAT1 acyltransferase, and SDP1 triacylglycerol lipase synergistically direct fatty acids toward β -oxidation, thereby maintaining membrane lipid homeostasis. *Plant Cell* **26**: 4119–4134.

Fan, J., Yu, L., and Xu, C. (2017). A Central Role for Triacylglycerol in Membrane Lipid Breakdown, Fatty Acid β -Oxidation, and Plant Survival under Extended Darkness. *Plant Physiol.* **174**: 1517–1530.

Fick, S.E. and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**: 4302–4315.

Fraire-Velázquez, S. and Balderas-Hernández, V.E. (2013). Abiotic Stress in Plants and Metabolic Responses. In *Abiotic Stress - Plant Responses and Applications in Agriculture*, K. Vahdati, ed (Intech open), p. 25.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**: 1.

Fusari, C.M., Kooke, R., Lauxmann, M.A., Annunziata, M.G., Enke, B., Hoehne, M., Krohn, N., Becker, F.F.M.M., Schlereth, A., Sulpice, R., Stitt, M., and Keurentjes, J.J.B. (2017). Genome-wide association mapping reveals that specific and pleiotropic regulatory mechanisms fine-tune central metabolism and growth in *Arabidopsis*. *Plant Cell* **29**: 2349–2373.

Gacek, K., Bayer, P.E., Bartkowiak-Broda, I., Szala, L., Bocianowski, J., Edwards, D., and Batley, J. (2017). Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*. *Front. Plant Sci.* **7**: 1–13.

- Gao, X. and Starmer, J.** (2007). Human population structure detection via multilocus genotype clustering. *BMC Genet.* **8**: 34.
- Gascuel, O.** (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**: 685–695.
- Go, Y.S., Kim, H., Kim, H.J., and Suh, M.C.** (2014). Arabidopsis cuticular wax biosynthesis is negatively regulated by the DEWAX gene encoding an AP2/ERF-type transcription factor. *Plant Cell* **26**: 1666–1680.
- Gouy, M., Guindon, S., and Gascuel, O.** (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**: 221–224.
- Graf, A. and Smith, A.M.** (2011). Starch and the clock: The dark side of plant productivity. *Trends Plant Sci.* **16**: 169–175.
- Guo, H.S., Zhang, Y.M., Sun, X.Q., Li, M.M., Hang, Y.Y., and Xue, J.Y.** (2016). Evolution of the KCS gene family in plants: The history of gene duplication, sub/neofunctionalization and redundancy. *Mol. Genet. Genomics* **291**: 739–752.
- Hall, T.A.** (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series* (Information Retrieval Ltd., c1979-c2000.: London), pp. 95–98.
- Han, G., Gable, K., Kohlwein, S.D., Beaudoin, F., Napier, J.A., and Dunn, T.M.** (2002). The *Saccharomyces cerevisiae* YBR159w gene encodes the 3-ketoreductase of the microsomal fatty acid elongase. *J. Biol. Chem.*
- Han, X.** (2016). *Lipidomics: Comprehensive Mass Spectrometry of Lipids* D.M. Desiderio and J.A. Loo, eds (Wiley).
- Higashi, Y., Okazaki, Y., Myouga, F., Shinozaki, K., and Saito, K.** (2015). Landscape of the lipidome and transcriptome under heat stress in *Arabidopsis thaliana*. *Sci. Rep.* **5**: 10533.
- Horton, M.W. et al.** (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat. Genet.* **44**: 212–6.
- Hummel, J., Segu, S., Li, Y., Irgang, S., Jueppner, J., and Giavalisco, P.** (2011). Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. *Front Plant Sci* **2**: 54.
- Joubès, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., and Lessire, R.** (2008). The VLCFA elongase gene family in *Arabidopsis thaliana*: Phylogenetic analysis, 3D modelling and expression profiling. *Plant Mol. Biol.* **67**: 547–566.
- Kao, Y.T., Gonzalez, K.L., and Bartel, B.** (2018). Peroxisome function, biogenesis, and dynamics in plants. *Plant Physiol.* **176**: 162–177.
- Karimi, M., Inzé, D., and Depicker, A.** (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**: 193–195.
- Katoh, K., Rozewicki, J., and Yamada, K.D.** (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**: 1160–1166.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O.** (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.

- Khan, B.R., Adham, A.R., and Zolman, B.K.** (2012). Peroxisomal Acyl-CoA oxidase 4 activity differs between Arabidopsis accessions. *Plant Mol. Biol.* **78**: 45–58.
- Kim, H., Go, Y.S., and Suh, M.C.** (2018). DEWAX2 Transcription Factor Negatively Regulates Cuticular Wax Biosynthesis in Arabidopsis Leaves. *Plant Cell Physiol.* **59**: 966–977.
- Kim, J., Jung, J.H., Lee, S.B., Go, Y.S., Kim, H.J., Cahoon, R., Markham, J.E., Cahoon, E.B., and Suh, M.C.** (2013). 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.* **162**: 567–580.
- Lämke, J., Brzezinka, K., Altmann, S., and Bäurle, I.** (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *EMBO J.* **35**: 162–175.
- Lees, J.M.** (2000). Geotouch: Software for Three and Four Dimensional GIS in the Earth Sciences. *Comput. Geosci.* **26**: 751–761.
- Li-Beisson, Y. et al.** (2010). Acyl-Lipid Metabolism. *Arab. B.* **8**: e0133.
- Li, H., Thrash, A., Tang, J.D., He, L., Yan, J., and Warburton, M.L.** (2019). Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *Plant J.* **98**: 853–863.
- Lipka, A.E. et al.** (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**: 2397–2399.
- Ludwig-Müller, J.** (2011). Auxin conjugates: Their role for plant development and in the evolution of land plants. *J. Exp. Bot.* **62**: 1757–1773.
- Luo, J.** (2015). Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* **24**: 31–38.
- Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., Yonemaru, J.I., Ebana, K., Yano, M., and Saito, K.** (2012). Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* **70**: 624–636.
- Nelson, B.K., Cai, X., and Nebenführ, A.** (2007). A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* **51**: 1126–1136.
- Nobusawa, T., Okushima, Y., Nagata, N., Kojima, M., Sakakibara, H., and Umeda, M.** (2013). Synthesis of very-long-chain fatty acids in the epidermis controls plant organ growth by restricting cell proliferation. *PLoS Biol.* **11**.
- O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R.** (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**: 1280–1292.
- Patrick, J.W. and Colyvas, K.** (2014). Crop yield components-photoassimilate supply-or utilisation limited-organ development? *Funct. Plant Biol.*
- Perraki, A., Cacas, J.-L., Crowet, J.-M., Lins, L., Castroviejo, M., German-Retana, S., Mongrand, S., and Raffaele, S.** (2012). 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.* **160**: 624–637.
- Pilkington, S.M., Encke, B., Krohn, N., Höhne, M., Stitt, M., and Pyl, E.T.** (2015). Relationship between starch degradation and carbon demand for maintenance and growth in Arabidopsis thaliana in different irradiance and temperature regimes. *Plant, Cell Environ.* **38**: 157–171.

786 **Puranik, S., Kumar, K., Srivastava, P.S., and Prasad, M.** (2011). Electrophoretic mobility shift assay
787 reveals a novel recognition sequence for *Setaria italica* NAC protein. *Plant Signal. Behav.* **6**:
788 1588–1590.

789 **Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., Li, Z., and Ye, G.** (2015). Genome-wide
790 association study of grain appearance and milling quality in a worldwide collection of indica
791 rice germplasm. *PLoS One* **10**: e0145577.

792 **R Core Team** (2019). R: A language and environment for statistical computing.

793 **Raivo Kolde** (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12.

794 **Riedelsheimer, C., Brotman, Y., Méret, M., Melchinger, A.E., and Willmitzer, L.** (2013). The maize
795 leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits.
796 *Sci. Rep.* **3**.

797 **Rossak, M., Smith, M., and Kunst, L.** (2001). Expression of the FAE1 gene and FAE1 promoter activity
798 in developing seeds of *Arabidopsis thaliana*. *Plant Mol. Biol.* **46**: 717–725.

799 **Sant’Ana, G.C. et al.** (2018). Genome-wide association study reveals candidate genes influencing
800 lipids and diterpenes contents in *Coffea arabica* L. *Sci. Rep.* **8**: 1–12.

801 **Schumacher, J., Kaufmann, K., and Yan, W.** (2017). Multiplexed GuideRNA-expression to Efficiently
802 Mutagenize Multiple Loci in *Arabidopsis* by CRISPR-Cas9. *Bio-Protocol* **7**.

803 **Seo, P.J., Lee, S.B., Suh, M.C., Park, M.J., and Park, C.M.** (2011). The MYB96 transcription factor
804 regulates cuticular wax biosynthesis under drought conditions in *arabidopsis*. *Plant Cell* **23**:
805 1138–1152.

806 **Shin, J.-H., Blay, S., McNeney, B., and Graham, J.** (2006). LDheatmap: an R function for graphical
807 display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat.*
808 *Softw.* **16**: 1–10.

809 **Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J.** (2007). pcaMethods-a
810 bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**: 1164–
811 1167.

812 **Stitt, M. and Zeeman, S.C.** (2012). Starch turnover: pathways, regulation and role in growth. *Curr.*
813 *Opin. Plant Biol.* **15**: 282–292.

814 **Szymanski, J., Brotman, Y., Willmitzer, L., and Cuadros-Inostroza, Á.** (2014). Linking gene expression
815 and membrane lipid composition of *Arabidopsis*. *Plant Cell* **26**: 915–928.

816 **Tajima, F.** (1989). Statistical method for testing the neutral mutation hypothesis by DNA
817 polymorphism. *Genetics* **123**: 585–595.

818 **Thirumalaikumar, V.P., Devkar, V., Mehterov, N., Ali, S., Ozgur, R., Turkan, I., Mueller-Roeber, B.,
819 and Balazadeh, S.** (2018). NAC transcription factor JUNGBRUNNEN 1 enhances drought
820 tolerance in tomato. *Plant Biotechnol. J.* **16**: 354–366.

821 **Tonsor, S.J., Alonso-Blanco, C., and Koornneef, M.** (2005). Gene function beyond the single trait:
822 Natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant, Cell*
823 *Environ.*

824 **Welti, R., Li, W., Li, M., Sang, Y., Biesiada, H., Zhou, H.E., Rajashekar, C.B., Williams, T.D., and
825 Wang, X.** (2002). Profiling membrane lipids in plant stress responses: Role of phospholipase D α
826 in freezing-induced lipid changes in *arabidopsis*. *J. Biol. Chem.* **277**: 31994–32002.

Wen, W. et al. (2015). Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell* **27**: 1839–1856.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., and Yan, J. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* **5**.

Wu, S. et al. (2016). Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*. *PLoS Genet.* **12**: 1–31.

Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooke, R., Méret, M., Keurentjes, J.B., Nikoloski, Z., Fernie, A.R., Willmitzer, L., and Brotman, Y. (2018). Mapping the *Arabidopsis* Metabolic Landscape by Untargeted Metabolomics at Different Environmental Conditions. *Mol. Plant* **11**: 118–134.

Author contribution

YB and CMF conceived the project and supervised the study. UL conducted GWAS experiments, obtained and selected KO mutants, performed experimental validation with KO, statistical analysis and interpretation of results. A-KR help UL in experimental work and performed analysis of the phenotypic data. JJ and MB conducted experiments on tobacco and yeast lines. JSz conducted the phylogenetic and geographical-signal analysis. VPTK provided expertise and assistance on TF analysis and advised on EMSA experiment and qRT-PCR assay. ML and SW assist in GWAS and multivariate analysis. FZ provided the plant material and carried out the lipid profiling for 3D and 6D GWAS. NE and SK helped UL in qRT-PCR assay for GWAS population and KO validation. JSc obtained KO constructs for transformation of *Arabidopsis*. AS provided access to LC-MS. ARF and AS help with discussion. CMF performed TASSEL, LD, Tajima's D and haplotype analyses, analysed and interpreted the results and helped with data visualization. CMF and UL wrote the paper with input from YB, JJ, JSz and YL-B. All authors read and draft the manuscript.

Acknowledgements

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, Gudrun Wolter and Simy Weil 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). This research was supported by the Israel Science Foundation (grant no. 859/19).

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. Results from GWAS analysis on lipidomic data from plants grown under control (CHD) and stress (HD) conditions.

a. Chromosome scheme and Quantitative Trait Loci (QTL) identified in the two experimental conditions (CHD = circles, HD = squares), for different lipid classes: phosphatidylethanolamine (PE), phosphatidylcholine (lecithin) (PC), monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), diacylglycerol (DAG) and triacylglycerol (TAG). Color/shape references are included in the figure. Gene IDs are included only for co-localizing QTL for two 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* 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 and clustered using Ward's distance on pairwise dissimilarity. **d.** Average trait value (intensity of TAG 54:9, log₂ scale) for the different *KCS4* haplotypes using SNPs m11502, m11503, m11504 and m11505. **e.** TAG 54:9 average value for accessions carrying the *KCS4*(Met) and *KCS4*(Thr) alleles.

Figure 2. Dissection of *KCS4* alleles involved in TAG variation, pattern of linkage disequilibrium and signatures of selection.

a. *KCS4* gene model depicting promoter region (P, green), 5'UTR (orange), coding region (yellow) and 3'UTR (orange) of AT1G19440/*KCS4*. Beginning, end and relevant SNP positions are 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 250K SNPchip plus 1,137 additional polymorphisms. Bonferroni threshold (LOD = 4.4) is indicated with a horizontal dashed line. **c.** Linkage disequilibrium (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 grey arrows, *KCS4* is highlighted in yellow. Significantly negative Tajima's D values are marked with asterisks (* $P < 0.05$, ** $P < 0.01$).

Figure 3. *KCS4* functional validation using knock-out mutants in *Arabidopsis* and heterologous expression in *Nicotiana benthamiana*.

a and b. Average TAG profiles for wild-type (WT) accessions and knock-out (KO) mutants carrying the *KCS4*(Met) allele (WT: red; KO: pink) and the *KCS4*(Thr) allele (WT: blue; KO: light-blue). TAG class 54 (from 54:1 to 54:9) is plotted for CHD and HD for two independent experiments. TAG class 56 was only detected in the second experiment for 56:4, 56:5 and 56:6. One-way ANOVA followed by Tuckey post-hoc test was performed to detect differences between lines. Significant changes (* $P < 0.05$) were detected between accessions with different genetic background (red and pink vs. blue and light-blue). No significant differences were found for the KO lines compared with their WT counterparts. **c.** Transient expression of At*KCS4* in *Nicotiana benthamiana* leaves. TAG levels for classes 52 (52:0 - 52:8), 54 (54:0 - 54:9) and 56 (56:0 - 56:0) for the control line and for the At*KCS4* line (At*KCS4*, black circles). Significant values (*t*-Test) are indicated (° $P < 0.1$, * $P < 0.01$, ** $P < 0.001$).

Figure 4. *KCS4* transcriptional regulation.

a. Percentage of transcription factors (TF) 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) for the main TF classes. SNP positions in the promoter and in coding region are marked with horizontal dash-lines and names are stated above gene model. **c.** Electrophoretic Mobility Shift Assay (EMSA) showing the interaction between *HEAT SHOCK FACTOR A2* (HSFA2/AT2G26150) and the HSE binding motif (heat shock element: 5'- nnGnAnnTnCtn -3') found within the SNP m11502 in *KCS4* promoter region; 1: labelled probe (5'-DY682-labelled double-stranded oligonucleotide) only; 2: labelled probe plus HSFA2-GST protein; 3: labelled probe, HSFA2-GST protein and 200× competitor DNA (unlabeled oligonucleotide containing HSE binding site with the "C" allele at m11502), 4: well spill, 5: labelled probe plus free GST, 6: labelled 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 qRT-PCR. Significant differences between CHD and HD transcript levels and between C and T alleles are shown (*t*-Test: * $P < 0.05$, ** $P < 0.001$, *** $P < 0.0001$, n.s. not significant). Data represent means $\pm$ SD (three technical replicates per sample). **e.** Average trait value (TAG 54:9 intensity, in $\log_2$) for the two 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 hours after treatment. Data was obtained previously (Caldana et al., 2011).

Figure 5. *KCS4* activity and specificity in fatty-acid elongation and subcellular localization.

a. Very long chain fatty acid methyl ester (VLC-FAME) profiles (16 to 18 and 20 to 26, separate panels) of wild-type yeast transformed with an empty vector (orange) and *AtKCS4* (blue). **b.** VLC-FAME profiles (16 to 18 and 20 to 26, separate panels) of Δ elo3 yeast mutant transformed with an empty vector (grey), *AtKCS4* (blue bars) and *AtKCS9* (green). **c.** VLC-FAME profiles (20 to 30) from leaves of *Nicotiana benthamiana* control plants (pink) or transformed with *AtKCS4* (blue). Significant differences are shown (*t*-Test: * $P < 0.05$, ** $P < 0.001$, *** $P < 0.0001$). **d.** Confocal images of *N. benthamiana* leaves transiently transformed with *KCS4*-YFP construct. The plasmid ER-GK CD3-955 was used as an ER marker. Bars = 20 μ m.

Figure 6. Biomass changes in *kcs4* mutants.

Wild-type Arabidopsis accessions and knock-out mutants for Pro-0 and Sh-0, carrying *KCS4*(Met) or *KCS4*(Thr), respectively. *t*-Test for rosette fresh weight (g) (* $P < 0.05$, *** $P < 0.001$).

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 four *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 highlighted in red. **b.** Average haplotype value for: the principal component 1 (PC1) obtained using the data in *KCS4* cluster (red triangle in **a**), and the climate and geographical parameter 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 HD induces the degradation of galactolipids (MGDGs and DGDGs) from the chloroplasts' thylakoids by lipases (HIL1). 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, directing the saturated ones to the very-long-chain fatty-acid pathway (VLCFA). 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 as red and blue, respectively. Polyunsaturated TAG (pu TAGs) levels are, therefore, higher in *KCS4*(Met) accessions comparing to *KCS4*(Thr) accessions, as the result of a rise in the ratio of polyunsaturated-to-saturated FAs (pu/sa FA-CoA). The activation of DEWAX transcription factors under darkness inhibits wax biosynthesis. Therefore, VLCFA-CoA produced by the elongase complex (here exemplified by *KCS4*, *KCR*, *HCD*, *ECR*) could directly serve as a source of carbon, and undergo degradation in the peroxisome. pu TAGs are

accumulated in lipid bodies. FFAs are released from lipid bodies by the action of lipases (SDP1/LIP1) and further imported to the peroxisome to undergo β -oxidation. ACXs are key enzymes for FA degradation and they have been also implicated in the IBA to IAA conversion in the peroxisome. ACX3 and ACX6 have been detected as putative regulators of KCS4. Acronyms: ABCA9 (ATP-BINDING CASSETTE A9), ACBP-1 (Acyl-CoA-binding protein 1); ACX (ACYL-COA OXIDASE); CoA (Coenzyme A); DGDG (Digalactosyldiacylglycerol); DEWAX (DECREASE WAX BIOSYNTHESIS), ECR (ENOYL-COA REDUCTASE); FAs (fatty acids); FAS (fatty-acid synthesis); FAX1 (FATTY ACID EXPORT1); FFAs (free fatty acids); HCD (β -HYDROXYACYL-COA DEHYDRATASE); HIL1 (HEAT INDUCIBLE LIPASE1); IAA (indol-3-acetic-acid), IBA (indol-3-butyric-acid), KCR (β -KETOACYL REDUCTASE); KCS4(Met/Thr) (3-KETOACYL-COA SYNTHASE 4); LACS (LONG-CHAIN ACYL-COA SYNTHETASE); SDP1/LIP1 (LIPASE 1); MGDG (Monogalactosyldiacylglycerol); PXA1 (PEROXISOMAL ABC-TRANSPORTER 1); pu (polyunsaturated), sa (saturated), TAGs (Triacylglycerols); VLCFA (Very-long-chain-fatty-acid).

Supplementary Figure 1. Phenotypic analyses for the Arabidopsis GWAS panel under control and stress conditions

a. Principal component analysis (PCA) of lipid data for control condition (CHD, blue circles) and heat and dark condition (HD, red diamonds). Each dot represents a single accession. PC1 and PC2 explain together 67.5% of the variance observed in the phenotypic data. **b.** Heatmap of differential accumulation of lipid levels in stress (HD) and control (CHD) conditions. Lipids are clustered by lipid class. For each accession, the lipid level in HD was subtracted by the lipid level in CHD. Positive values (red) denote higher levels of lipid in HD condition, negative values (blue) denote higher level of lipid in CHD condition (scale= $\log_2$ (Δ of lipid intensity)). **c.** Correlation pattern among lipid levels in both conditions: CHD (blue rectangle) and HD (red rectangle). Pairwise Pearson correlations (r^2) were calculated between each lipid across all accessions. Lipids are highlighted with colors according to lipid classes (references included in the figure). **d.** Distribution of fold change of lipid-feature levels across all accessions. For each lipid feature, fold change was calculated by dividing the average of the five highest intensities measured in a given 5 accession by the average of the five lowest intensities. Y-axes is expressed in percentage of (A) all metabolite features (N=109 in stress, N=98 in CHD), (B) only TAG features (N=41 in stress, N= 30 in CHD) and (C) other lipid features (N=68 in stress and N=68 in CHD).

Supplementary Figure 2. Results from GWAS analysis on lipidomic data from plants grown under extended darkness (3D, 6D) conditions.

a. Chromosome scheme and Quantitative Trait Loci (QTL) identified in the two experimental conditions (3D = circles, 6D = squares), for different lipid classes: phosphatidylethanolamine (PE), phosphatidylcholine (lecithin) (PC), monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), diacylglycerol (DAG) and triacylglycerol (TAG). Color/shape references are included in the figure. Gene IDs are included only for QTL co-localizing for two or more lipid species with LOD score > 5.3. **b.** Manhattan plots obtained for TAG 54:9 in 3D (left) and in 6D (right). **c.** Average trait value (intensity of TAG 54:9, $\log_2$ scale) for the different KCS4 haplotypes using SNPs m11502, m11503, m11504 and m11505. **e.** TAG 54:9 average value for accessions carrying the KCS4(Met) and KCS4(Thr) alleles.

Supplementary Figure 3. Phylogenetic signal of the lipidome

a. Distribution of the phylogenetic signal calculated for all measured lipids in HD and CHD conditions. Only lipids exhibiting significant phylogenetic signal are labelled. Lipids significant in CHD and HD are marked with blue and red dots, respectively. Compounds significant in both conditions ($p < 0.05$) have

overlapping red and blue markers. **b.** K parameter distribution obtained for the TAG profiles in CHD (control) and HD (treatment) condition according to (Blomberg et al., 2003).

Supplementary Figure 4. Prediction of climate parameters from lipidomic profiles

Regression coefficients of selected predictive lipidomic features towards 19 climate variables and 3 geographical coordinates (latitude, longitude and altitude) of the geographical origin of analyzed accessions. Statistical evaluation of the fitted models is provided on the right-hand panel. The model fit significance (red asterisks) is estimated by the correlation coefficient between the original and model-fitted variable values. The significance of Cross-Validation test (black asterisks) represents the significance of the correlation coefficient between the predicted and real values of the test sample. The cross-validation has been performed 100 times with the 3:1 proportion of the training to test sample set size.

Supplementary File 1.

Alignment used for the generation of the SNP matrix used in TASSEL to analyze more polymorphisms in the QTL associated with *KCS4*. Sequences were obtained from the Arabidopsis 1001 genome project (<http://signal.salk.edu/atg1001/3.0/gebrowser.php>, 162 accessions). TAIR10 positions of the extract are chromosome 1: 6689119 - 6769118. Alignment was performed using MAFFT (Katoh et al., 2019) and manually curated when needed.

Supplementary File 2.

Alignment of *KCS4* sequences from wild type and knockout accessions used in this study.

Supplementary Table 1.

Phenotypic data used for GWAS and collected in: (1) control conditions for the heat and dark stress treatment (CHD), (2) heat and dark stress (HD), (3) control conditions for the dark treatment (CD), (4) three days extended darkness (3D) and (5) six days extended darkness. Each phenotype is in different excel sheets. Plants were grown and harvested as described in Materials and Methods. Lipid classes detected are DAG: diacylglycerols, DGDG: digalactosyldiacylglycerol, MGDG: monogalactosyldiacylglycerol, PC: phosphatidylcholine (lecithin), PE: phosphatidylethanolamine and TAG: triacylglycerol. For each species the number of carbons and the degree of bond-saturation is stated. Phenotypic value is the logarithm in the second base of intensity obtained in LC-MS.

Supplementary Table 2.

Table with GWAS results obtained from GAPIT for each phenotype presented in Supp. Table 1. All SNPs with a LOD score > 5.3 for all traits obtained in the same condition are summarized in the same table. Each sheet presents the results for each condition (see Supp. Table 1). MAF: minor allele frequency.

Supplementary Table 3.

Full list of GWAS results summarized in number of QTL (quantitative trait loci) and genes detected per condition, for all lipid species phenotyped (LOD score > 5.3). Co-localization between traits in the same conditions are stated in "Peak.ID" columns. Co-localizing genes between CHD and HD condition are also mentioned. Putative lipid functions for genes related to lipid biosynthesis pathways are included in the last column. *KCS4* association is highlighted in red. Each sheet compiles results for all traits phenotyped in one condition (see Supp. Table 1)

Supplementary Table 4.

Data used for TASSEL analysis: Genotypic matrix, generated from the alignment in Supp. File 1; principal components and kinship matrix used to control for population structure and confounding effects. Phenotypic matrixes for HD and 3D are included in Supplementary Table 1.

Supplementary Table 5.

GWAS results for HD and 3D data (see Supp. Table 1 for phenotype matrix and Supp. Table 4 for genotype matrix) using complete sequence information for 80 kb around KCS4. SNPs super-passing Bonferroni-corrected p-values (LOD > 4.38) are marked in bold. Markers ID in Tassel are referenced to positions in Supp. File 1, marker ID in GAPIT and TAIR10 position.

Supplementary Table 6.

TAGs associated with *KCS4*, measured in Arabidopsis *kcs4* lines. Accession ID (common name), Type (wild type or mutant), *KCS4_allele* (Methionine or Threonine), Condition (control or HD) and number of experiment (Exp) are included in different columns. The table presents trait values for each biological replicate for accessions separately. Values in Figure 3 were obtained by averaging accessions encoding for a Methionine or Threonine, respectively for each type (WT and KO).

Supplementary Table 7.

GWAS result summarized in number of eQTL (expression QTL) and genes detected using *KCS4* expression data in CHD and HD (LOD score of association > 5.0). Putative lipid functions for genes related to lipid biosynthesis pathways are included in the last column. Each sheet compiles results for all traits phenotyped in one condition (see Supp. Table 7).

Supplementary Table 8.

Primers and oligonucleotide sequences used in this study.

Figure 1. Results from GWAS analysis on lipidomic data from plants grown under control (CHD) and stress (HD) conditions.

a. Chromosome scheme and Quantitative Trait Loci (QTL) identified in the two experimental conditions (CHD = circles, HD = squares), for different lipid classes: phosphatidylethanolamine (PE), phosphatidylcholine (lecithin) (PC), monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), diacylglycerol (DAG) and triacylglycerol (TAG). Color/shape references are included in the figure. Gene IDs are included only for co-localizing QTL for two 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* 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 and clustered using Ward's distance on pairwise dissimilarity. **d.** Average trait value (intensity of TAG 54:9, log₂ scale) for the different *KCS4* haplotypes using SNPs m11502, m11503, m11504 and m11505. **e.** TAG 54:9 average value for accessions carrying the *KCS4*(Met) and *KCS4*(Thr) alleles.

Figure 2. Dissection of *KCS4* alleles involved in TAG variation, pattern of linkage disequilibrium and signatures of selection.

a. *KCS4* gene model depicting promoter region (P, green), 5'UTR (orange), coding region (yellow) and 3'UTR (orange) of AT1G19440/*KCS4*. Beginning, end and relevant SNP positions are 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 250K SNPchip plus 1,137 additional polymorphisms. Bonferroni threshold ($LOD = 4.4$) is indicated with a horizontal dashed line. **c.** Linkage disequilibrium (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 grey arrows, *KCS4* is highlighted in yellow. Significantly negative Tajima's D values are marked with asterisks (* $P < 0.05$, ** $P < 0.01$).

Figure 3. KCS4 functional validation using knock-out mutants in Arabidopsis and heterologous expression in *Nicotiana benthamiana*.

a and **b**. Average TAG profiles for wild-type (WT) accessions and knock-out (KO) mutants carrying the *KCS4*(Met) allele (WT: red; KO: pink) and the *KCS4*(Thr) allele (WT: blue; KO: light-blue). TAG class 54 (from 54:1 to 54:9) is plotted for CHD and HD for two independent experiments. TAG class 56 was only detected in the second experiment for 56:4, 56:5 and 56:6. One-way ANOVA followed by Tuckey post-hoc test was performed to detect differences between lines. Significant changes (* $P < 0.05$) were detected between accessions with different genetic background (red and pink vs. blue and light-blue). No significant differences were found for the KO lines compared with their WT counterparts. **c**. Transient expression of *AtKCS4* in *Nicotiana benthamiana* leaves. TAG levels for classes 52 (52:0 - 52:8), 54 (54:0 - 54:9) and 56 (56:0 - 56:9) for the control line and for the *AtKCS4* line (*AtKCS4*, black circles). Significant values (*t*-Test) are indicated (° $P < 0.1$, * $P < 0.01$, ** $P < 0.001$).

Figure 4. KCS4 transcriptional regulation.

a. Percentage of transcription factors (TF) classes predicted to bind *KCS4* genomic region according to in-silico analysis of binding motifs using DAP-seq database (O'Malley et al., 2016). **b.** Position of the binding motifs (red rectangles) for the main TF classes. SNP positions in the promoter and in coding region are marked with horizontal dash-lines and names are stated above gene model. **c.** Electrophoretic Mobility Shift Assay (EMSA) showing the interaction between HEAT SHOCK FACTOR A2 (AT2G26150/*HSFA2*) and the HSE binding motif (heat shock element: 5' nnGnAnnTnCtn -3') found within the SNP m11502 in *KCS4* promoter region; 1: labelled probe (5'-DY682-labelled double-stranded oligonucleotide) only; 2: labelled probe plus *HSFA2*-GST protein; 3: labelled probe, *HSFA2*-GST protein and 200× competitor DNA (unlabelled oligonucleotide containing HSE binding site with the "C" allele at m11502), 4: well spill, 5: labelled probe plus free GST, 6: labelled probe, *HSFA2*-GST protein and 200× competitor DNA (unlabelled 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 qRT-PCR. Significant differences between CHD and HD transcript levels and between C and T alleles are shown (*t*-Test: * $P < 0.05$, ** $P < 0.001$, *** $P < 0.0001$, n.s. not significant). Data represent means $\pm$ SD (three technical replicates per sample). **e.** Average trait value (TAG 54:9 intensity, in log2) for the two 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 hours after treatment. Data was obtained previously (Caldana et al., 2011).

Figure 5. KCS4 activity and specificity in fatty-acid elongation and subcellular localization.

a. Very long chain fatty acid methyl ester (VLC-FAME) profiles (16 to 18 and 20 to 26, separate panels) of wild-type yeast transformed with an empty vector (orange) and *AtKCS4* (blue). **b.** VLC-FAME profiles (16 to 18 and 20 to 26, separate panels) of *elo3* yeast mutant transformed with an empty vector (grey), *AtKCS4* (blue bars) and *AtKCS9* (green). **c.** VLC-FAME profiles (20 to 30) from leaves of *Nicotiana benthamiana* control plants (pink) or transformed with *AtKCS4* (blue). Significant differences are shown (*t*-Test: * $P < 0.05$, ** $P < 0.001$, *** $P < 0.0001$). **d.** Confocal images of *N. benthamiana* leaves transiently transformed with KCS4-YFP construct. The plasmid ER-GK CD3-955 was used as an ER marker. Bars = 20 μm.

Figure 6. Biomass changes in *kcs4* mutants.

Wild-type Arabidopsis accessions and knock-out mutants for Pro-0 and Sh-0, carrying *KCS4*(Met) or *KCS4*(Thr), respectively. *t*-Test for rosette fresh weight (g) (* $P < 0.05$, *** $P < 0.001$).

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 four *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 highlighted in red. **b.** Average haplotype value for: the principal component 1 (PC1) obtained using the data in *KCS4* cluster (red triangle in **a**), and the climate and geographical parameter 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 HD induces the degradation of galactolipids (MGDGs and DGDGs) from the chloroplasts' thylakoids by lipases (HIL1). 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, directing the saturated ones to the very-long-chain fatty-acid pathway (VLCFA). 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 as red and blue, respectively. Polyunsaturated TAG (pu TAGs) levels are, therefore, higher in KCS4(Met) accessions comparing to KCS4(Thr) accessions, as the result of a rise in the ratio of polyunsaturated-to-saturated FAs (pu/sa FA-CoA). The activation of *DEWAX* transcription factors under darkness inhibits wax biosynthesis. Therefore, VLCFA-CoA produced by the elongase complex (here exemplified by KCS4, KCR, HCD, ECR) could directly serve as a source of carbon, and undergo degradation in the peroxisome. pu TAGs are accumulated in lipid bodies. FFAs are released from lipid bodies by the action of lipases (SDP1/LIP1) and further imported to the peroxisome to undergo β -oxidation. ACXs are key enzymes for FA degradation and they have been also implicated in the IBA to IAA conversion in the peroxisome. ACX3 and ACX6 have been detected as putative regulators of KCS4. Acronyms: ABCA9 (ATP-BINDING CASSETTE A9), ACBP-1 (Acyl-CoA-binding protein 1); ACX (ACYL-COA OXIDASE); CoA (Coenzyme A); DGDG (Digalactosyldiacylglycerol); *DEWAX* (*DECREASE WAX BIOSYNTHESIS*), ECR (ENOYL-COA REDUCTASE); FAs (fatty acids); FAS (fatty-acid synthesis); FAX1 (FATTY ACID EXPORT1); FFAs (free fatty acids); HCD (β -HYDROXYACYL-COA DEHYDRATASE); HIL1 (HEAT INDUCIBLE LIPASE1); IAA (indol-3-acetic-acid), IBA (indol-3-butyric-acid), KCR (β -KETOACYL REDUCTASE); KCS4(Met/Thr) (3-KETOACYL-COA SYNTHASE4); LACS (LONG-CHAIN ACYL-COA SYNTHETASE); SDP1/LIP1 (LIPASE 1); MGDG (Monogalactosyldiacylglycerol); PXA1 (PEROXISOMAL ABC-TRANSPORTER 1); pu (polyunsaturated), sa (saturated), TAGs (Triacylglycerols); VLCFA (Very-long-chain-fatty-acid).

Parsed Citations

Adham, A.R., Zolman, B.K., Millius, A., and Bartel, B. (2005). Mutations in *Arabidopsis* acyl-CoA oxidase genes reveal distinct and overlapping roles in β -oxidation. *Plant J.* 41: 859–874.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26: 32–46.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J., Borevitz, J.O., Nordborg, M., Vitek, O., and Salt, D.E. (2010). A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter *AtHKT1;1*. *PLoS Genet.* 6: e1001193.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Blacklock, B.J. and Jaworski, J.G. (2006). Substrate specificity of *Arabidopsis* 3-ketoacyl-CoA synthases. *Biochem. Biophys. Res. Commun.* 346: 583–590.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Blomberg, S.P., Garland Jr, T., and Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution (N. Y.)* 57: 717–745.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–5.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Burgos, A., Szymanski, J., Seiwert, B., Degenkolbe, T., Hannah, M.A., Giavalisco, P., and Willmitzer, L. (2011). Analysis of short-term changes in the *Arabidopsis thaliana* glycerolipidome in response to temperature and light. *Plant J.* 66: 656–668.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Caldana, C., Degenkolbe, T., Cuadros-Inostroza, A., Klie, S., Sulpice, R., Leisse, A., Steinhauser, D., Fernie, A.R., Willmitzer, L., and Hannah, M.A. (2011). High-density kinetic analysis of the metabolomic and transcriptomic response of *Arabidopsis* to eight environmental conditions. *Plant J.* 67: 869–884.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Chan, E.K.F., Rowe, H.C., Corwin, J.A., Joseph, B., and Kliebenstein, D.J. (2011). Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol.* 9.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Correa, S.M., Fernie, A.R., Nikoloski, Z., and Brotman, Y. (2020). Towards model-driven characterization and manipulation of plant lipid metabolism. *Prog. Lipid Res.*: 101051.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.*

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Dittrich-Domergue, F., Joubès, J., Moreau, P., Lessire, R., Stymne, S., and Domergue, F. (2014). The bifunctional protein TtFARAT from *Tetrahymena thermophila* catalyzes the formation of both precursors required to initiate ether lipid biosynthesis. *J. Biol. Chem.* 289: 21984–21994.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Domergue, F., Vishwanath, S.J., Joubès, J., Ono, J., Lee, J.A., Bourdon, M., Alhattab, R., Lowe, C., Pascal, S., and Lessire, R. (2010). Three *Arabidopsis* fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. *Plant Physiol.* 153: 1539–1554.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Eastmond, P.J., Hooks, M.A., Williams, D., Lange, P., Bechtold, N., Sarrobert, C., Nussaume, L., and Graham, I.A. (2000). Promoter trapping of a novel medium-chain acyl-CoA oxidase, which is induced transcriptionally during *Arabidopsis* seed germination. *J. Biol. Chem.*

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Fan, J., Yan, C., Roston, R., Shanklin, J., and Xu, C. (2014). Arabidopsis lipins, PDAT1 acyltransferase, and SDP1 triacylglycerol lipase synergistically direct fatty acids toward β -oxidation, thereby maintaining membrane lipid homeostasis. *Plant Cell* 26: 4119–4134.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Fan, J., Yu, L., and Xu, C. (2017). A Central Role for Triacylglycerol in Membrane Lipid Breakdown, Fatty Acid β -Oxidation, and Plant Survival under Extended Darkness. *Plant Physiol.* 174: 1517–1530.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Fick, S.E. and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37: 4302–4315.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Fraire-Velázquez, S. and Balderas-Hernández, V.E. (2013). Abiotic Stress in Plants and Metabolic Responses. In *Abiotic Stress - Plant Responses and Applications in Agriculture*, K. Vahdati, ed (Intech open), p. 25.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33: 1.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Fusari, C.M., Kooke, R., Lauxmann, M.A., Annunziata, M.G., Enke, B., Hoehne, M., Krohn, N., Becker, F.F.M.M., Schlereth, A., Sulpice, R., Stitt, M., and Keurentjes, J.J.B. (2017). Genome-wide association mapping reveals that specific and pleiotropic regulatory mechanisms fine-tune central metabolism and growth in Arabidopsis. *Plant Cell* 29: 2349–2373.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Gacek, K., Bayer, P.E., Bartkowiak-Broda, I., Szala, L., Bocianowski, J., Edwards, D., and Batley, J. (2017). Genome-wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*. *Front. Plant Sci.* 7: 1–13.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Gao, X. and Starmer, J. (2007). Human population structure detection via multilocus genotype clustering. *BMC Genet.* 8: 34.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14: 685–695.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Go, Y.S., Kim, H., Kim, H.J., and Suh, M.C. (2014). Arabidopsis cuticular wax biosynthesis is negatively regulated by the DEWAX gene encoding an AP2/ERF-type transcription factor. *Plant Cell* 26: 1666–1680.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27: 221–224.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Graf, A. and Smith, A.M. (2011). Starch and the clock: The dark side of plant productivity. *Trends Plant Sci.* 16: 169–175.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Guo, H.S., Zhang, Y.M., Sun, X.Q., Li, M.M., Hang, Y.Y., and Xue, J.Y. (2016). Evolution of the KCS gene family in plants: The history of gene duplication, sub/neofunctionalization and redundancy. *Mol. Genet. Genomics* 291: 739–752.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series* (Information Retrieval Ltd., c1979-c2000.: London), pp. 95–98.

Pubmed: [Author and Title](#)
Google Scholar: [Author Only Title Only Author and Title](#)

Han, G., Gable, K., Kohlwein, S.D., Beaudoin, F., Napier, J.A., and Dunn, T.M. (2002). The *Saccharomyces cerevisiae* YBR159w gene encodes the 3-ketoreductase of the microsomal fatty acid elongase. *J. Biol. Chem.*

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Han, X. (2016). Lipidomics: Comprehensive Mass Spectrometry of Lipids D.M. Desiderio and J.A. Loo, eds (Wiley).

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Higashi, Y., Okazaki, Y., Myouga, F., Shinozaki, K., and Saito, K. (2015). Landscape of the lipidome and transcriptome under heat stress in *Arabidopsis thaliana*. Sci. Rep. 5: 10533.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Horton, M.W. et al. (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. Nat. Genet. 44: 212–6.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hummel, J., Segu, S., Li, Y., Irgang, S., Jueppner, J., and Giavalisco, P. (2011). Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. Front Plant Sci 2: 54.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Joubès, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., and Lessire, R. (2008). The VLCFA elongase gene family in *Arabidopsis thaliana*: Phylogenetic analysis, 3D modelling and expression profiling. Plant Mol. Biol. 67: 547–566.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kao, Y.T., Gonzalez, K.L., and Bartel, B. (2018). Peroxisome function, biogenesis, and dynamics in plants. Plant Physiol. 176: 162–177.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAYTM vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci. 7: 193–195.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20: 1160–1166.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26: 1463–1464.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Khan, B.R., Adham, A.R., and Zolman, B.K. (2012). Peroxisomal Acyl-CoA oxidase 4 activity differs between *Arabidopsis* accessions. Plant Mol. Biol. 78: 45–58.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kim, H., Go, Y.S., and Suh, M.C. (2018). DEWAX2 Transcription Factor Negatively Regulates Cuticular Wax Biosynthesis in *Arabidopsis* Leaves. Plant Cell Physiol. 59: 966–977.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kim, J., Jung, J.H., Lee, S.B., Go, Y.S., Kim, H.J., Cahoon, R., Markham, J.E., Cahoon, E.B., and Suh, M.C. (2013). *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. 162: 567–580.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Länke, J., Brzezinka, K., Altmann, S., and Bäurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. EMBO J. 35: 162–175.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lees, J.M. (2000). Geotouch: Software for Three and Four Dimensional GIS in the Earth Sciences. Comput. Geosci. 26: 751–761.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Li-Beisson, Y. et al. (2010). Acyl-Lipid Metabolism. Arab. B. 8: e0133.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Li, H., Thrash, A., Tang, J.D., He, L., Yan, J., and Warburton, M.L. (2019). Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *Plant J.* 98: 853–863.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lipka, A.E. et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28: 2397–2399.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ludwig-Müller, J. (2011). Auxin conjugates: Their role for plant development and in the evolution of land plants. *J. Exp. Bot.* 62: 1757–1773.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* 24: 31–38.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J., Yonemaru, J.I., Ebana, K., Yano, M., and Saito, K. (2012). Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *Plant J.* 70: 624–636.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Nelson, B.K., Cai, X., and Nebenführ, A. (2007). A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J.* 51: 1126–1136.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Nobusawa, T., Okushima, Y., Nagata, N., Kojima, M., Sakakibara, H., and Umeda, M. (2013). Synthesis of very-long-chain fatty acids in the epidermis controls plant organ growth by restricting cell proliferation. *PLoS Biol.* 11.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 165: 1280–1292.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Patrick, J.W. and Colyvas, K. (2014). Crop yield components-photoassimilate supply-or utilisation limited-organ development? *Funct. Plant Biol.*

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Perraki, A., Cacas, J.-L., Crowet, J.-M., Lins, L., Castroviejo, M., German-Retana, S., Mongrand, S., and Raffaele, S. (2012). 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.* 160: 624–637.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Pilkington, S.M., Encke, B., Krohn, N., Höhne, M., Stitt, M., and Pyl, E.T. (2015). Relationship between starch degradation and carbon demand for maintenance and growth in *Arabidopsis thaliana* in different irradiance and temperature regimes. *Plant, Cell Environ.* 38: 157–171.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Puranik, S., Kumar, K., Srivastava, P.S., and Prasad, M. (2011). Electrophoretic mobility shift assay reveals a novel recognition sequence for *Setaria italica* NAC protein. *Plant Signal. Behav.* 6: 1588–1590.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., Li, Z., and Ye, G. (2015). Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. *PLoS One* 10: e0145577.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

R Core Team (2019). R: A language and environment for statistical computing.

Raivo Kolde (2019). pheatmap: Pretty Heatmaps. R package version 1.0.12.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Riedelsheimer, C., Brotman, Y., Méret, M., Melchinger, A.E., and Willmitzer, L. (2013). The maize leaf lipidome shows multilevel genetic control and high predictive value for agronomic traits. *Sci. Rep.* 3.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Rossak, M., Smith, M., and Kunst, L. (2001). Expression of the FAE1 gene and FAE1 promoter activity in developing seeds of *Arabidopsis thaliana*. Plant Mol. Biol. 46: 717–725.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sant'Ana, G.C. et al. (2018). Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. Sci. Rep. 8: 1–12.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Schumacher, J., Kaufmann, K., and Yan, W. (2017). Multiplexed GuideRNA-expression to Efficiently Mutagenize Multiple Loci in *Arabidopsis* by CRISPR-Cas9. Bio-Protocol 7.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Seo, P.J., Lee, S.B., Suh, M.C., Park, M.J., and Park, C.M. (2011). The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in *arabidopsis*. Plant Cell 23: 1138–1152.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. J. Stat. Softw. 16: 1–10.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). pcaMethods-a bioconductor package providing PCA methods for incomplete data. Bioinformatics 23: 1164–1167.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Stitt, M. and Zeeman, S.C. (2012). Starch turnover: pathways, regulation and role in growth. Curr. Opin. Plant Biol. 15: 282–292.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Szymanski, J., Brotman, Y., Willmitzer, L., and Cuadros-Inostroza, Á. (2014). Linking gene expression and membrane lipid composition of *Arabidopsis*. Plant Cell 26: 915–928.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Thirumalaikumar, V.P., Devkar, V., Mehterov, N., Ali, S., Ozgur, R., Turkan, I., Mueller-Roeber, B., and Balazadeh, S. (2018). NAC transcription factor JUNGBRUNNEN 1 enhances drought tolerance in tomato. Plant Biotechnol. J. 16: 354–366.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tonsor, S.J., Alonso-Blanco, C., and Koornneef, M. (2005). Gene function beyond the single trait: Natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. Plant, Cell Environ.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Welti, R., Li, W., Li, M., Sang, Y., Biesiada, H., Zhou, H.E., Rajashekar, C.B., Williams, T.D., and Wang, X. (2002). Profiling membrane lipids in plant stress responses: Role of phospholipase D α in freezing-induced lipid changes in *arabidopsis*. J. Biol. Chem. 277: 31994–32002.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wen, W. et al. (2015). Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. Plant Cell 27: 1839–1856.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., Liu, J., Liu, H., Chen, W., Luo, J., and Yan, J. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat. Commun. 5.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, S. et al. (2016). Combined use of genome-wide association data and correlation networks unravels key regulators of primary metabolism in *Arabidopsis thaliana*. PLoS Genet. 12: 1–31.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooke, R., Méret, M., Keurentjes, J.B., Nikoloski, Z., Fernie, A.R.,

Willmitzer, L., and Brotman, Y. (2018). Mapping the Arabidopsis Metabolic Landscape by Untargeted Metabolomics at Different Environmental Conditions. Mol. Plant 11: 118–134.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Author contribution

YB and CMF conceived the project and supervised the study. UL conducted GWAS experiments, obtained and selected KO mutants, performed experimental validation with KO, statistical analysis and interpretation of results. A-KR help UL in experimental work and performed analysis of the phenotypic data. JJ and MB conducted experiments on tobacco and yeast lines. JSz conducted the phylogenetic and geographical-signal analysis. VPTK provided expertise and assistance on TF analysis and advised on EMSA experiment and qRT-PCR assay. ML and SW assist in GWAS and multivariate analysis. FZ provided the plant material and carried out the lipid profiling for 3D and 6D GWAS. NE and SK helped UL in qRT-PCR assay for GWAS population and KO validation. JSc obtained KO constructs for transformation of Arabidopsis. AS provided access to LC-MS. ARF and AS help with discussion. CMF performed TASSEL, LD, Tajima's D and haplotype analyses, analysed and interpreted the results and helped with data visualization. CMF and UL wrote the paper with input from YB, JJ, JSz and YL-B. All authors read and draft the manuscript.

Acknowledgements

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, Gudrun Wolter and Simy Weil 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). This research was supported by the Israel Science Foundation (grant no. 859/19).

Competing interests

The authors declare no competing interests.