Hello darkness, my old friend: 3-Ketoacyl-Coenzyme A Synthase4 is a branch point in the regulation of triacylglycerol synthesis in *Arabidopsis thaliana*

Urszula Luzarowska[1], Anne-Kathrin Ruß[2], Jérôme Joubès[3], Marguerite Batsale[3], Jędrzej Szymański[4], Venkatesh Periyakavanam Thirumalaikumar[2], Marcin Luzarowski[2], Si Wu[5], Feng Zhu[2,6], Niklas Endres[2], Sarah Khedhayir[2], Julia Schumacher[7], Weronika Jasinska[1], Ke Xu[2], Sandra Marcela Correa Cordoba[1], Simy Weil[1], Aleksandra Skirycz[2], Alisdair Robert Fernie[2], Yonghua Li-Beisson[8], 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] University Bordeaux, CNRS, Laboratoire de Biogenèse Membranaire, UMR 5200, F-33140 Villenave d'Ornon, France.

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

[5] 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

† Senior Authors

***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 under carbon starvation, by acting as a regulatory branch point in the fate of fatty acids.

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 in very long chain fatty-acid (VLCFA) synthesis, underlies a differential accumulation of polyunsaturated triacylglycerols (puTAGs) 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. Allelic mutants and transient overexpression *in planta* revealed the differential role of *KCS4* alleles in VLCFA synthesis and wax coverage, puTAG accumulation and biomass. Moreover, the region harboring *KCS4* is under high selective pressure and 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 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. These metabolic responses to abiotic stress involve fine adjustments in carbohydrate, amino acid, amine and lipid pathways. Proper activation of early metabolic responses helps cells restore chemical and energetic imbalances imposed by the stress and is crucial to acclimation and survival (Fraire-Velázquez and Balderas-Hernández, 2013).

Like many other complex traits, plant metabolism has an extended variation within natural populations, and therefore has been widely investigated by using genome-wide association studies (GWAS). Genetic factors involved in variation for key metabolic traits have been successfully identified and characterized (i.e., primary and secondary metabolites and enzyme activities) (Wu et al., 2016, 2018; Fusari et al., 2017). A number of GWAS regarding lipid metabolism have been reported and resulted in the identification of new enzymes and regulators in species like Arabidopsis, maize, coffee, rice and canola (Riedelsheimer et al., 2013; Gacek et al., 2017; Sant'Ana et al., 2018; Li et al., 2019; Hong et al., 2022). These studies focused on lipid levels under optimal growth conditions. However, accumulation of lipids is highly affected by fluctuations in the environmental conditions, such as light, temperature and nutrient availability (Burgos et al., 2011; Caldana et al., 2011; Szymanski et al., 2014). A classic example involves compensation for 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, the major source of energy in the dark is starch, which is degraded linearly according to the duration of the previous night (Graf and Smith, 2011). Therefore, plants subjected to a sudden extended night (dark stress) experience carbon starvation due to limited starch availability. Under these conditions, polyunsaturated fatty acids (PUFA) released from membrane lipids are used as an alternative substrate for respiration via β-oxidation. These released fatty acids are often sequestered into triacylglycerols (TAGs) to avoid the adverse effects of the detergent-like properties of free fatty acids (FFAs). Therefore, it follows that TAG synthesis and lipid-droplet formation sometimes

84 constitute a cellular response to lipotoxicity. This sequestration is especially true for PUFA species
85 because they contain multiple double bonds and are liable to production of reactive oxygen species
86 (ROS) (Kao et al., 2018). It has been shown that the process of FFA sequestration into TAGs protects
87 the cell from ROS generated during PUFA peroxidation (Fan et al., 2017).

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

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

107 **Results**

108 **Plasticity in Arabidopsis lipid metabolism under different environmental conditions**

109 In order to investigate the genetic regulation of lipid biosynthesis in Arabidopsis, we conducted LC–
110 MS-based lipidomic analysis of 301 and 309 Arabidopsis accessions grown in two environmental
111 conditions: control (21/16°C in 16/8h light/dark cycle, termed CHD) and stress (heat 32°C + darkness
112 for 24 h, termed HD), respectively. HD was selected as it has previously been shown to have a strong
113 effect on the lipidome (Caldana et al., 2011). The mapping population was phenotyped for a total of
114 98 and 109 glycerolipids in CHD and HD, respectively (Supplementary Dataset 1).

115 Principal component analysis (PCA) of lipid data clearly separates the natural ecotypes into two
116 different groups according to the two environmental conditions tested (Supplementary Figure 1a).
117 Furthermore, differences in the lipid status between CHD and HD samples showed that TAGs were
118 strongly accumulated while MGDGs and DGDGs were mostly decreased under stress (Supplementary
119 Figure 1b). In addition, there is substantial positive correlation between lipid levels from a given lipid
120 class, suggesting coordinated regulation at the level of lipid classes (Supplementary Figure 1c).
121 Notably, TAGs showed the highest variation across accessions for both stress and control conditions,
122 with many species falling in the range of >10-fold difference (Supplementary Figure 1d).

123 **Genetic basis underlying Arabidopsis lipid metabolism**

124 Lipid intensities from CHD and HD and the genotypic information from the 250K SNPchip (Horton et
125 al., 2012) were used to perform two independent genome-wide association studies (GWAS) for the
126 identification of lipid quantitative trait loci (QTL). GWAS revealed 93 and 544 SNP–trait associations
127 in control and stress conditions, respectively, at a genome-wide significance level of LOD ≥ 5.3
128 (Supplementary Dataset 2). These SNP–trait associations further grouped into 51 and 66 QTL for

129 CHD and HD conditions, respectively. In total, only five QTL were shared between CHD and HD,
130 revealing the great plasticity of the lipidome in response to different environmental conditions
131 (Supplementary Dataset 3).

132 Regarding functional annotation of candidate genes, we mapped 12 and 22 genes putatively
133 involved in lipid metabolism for control and stress condition, respectively (Supplementary Dataset 3,
134 Acyl-Lipid Metabolism database (Li-Beisson et al., 2010) http://aralip.plantbiology.msu.edu/) (Figure
135 1). After filtering for robust associations co-localizing for two or more traits, four QTL harbored lipid-
136 genes (Figure 1a). Two QTL in Chromosome 1, one harboring the *3-KETOACYL-COENZYME A*
137 *SYNTHASE4* (*KCS4*, AT1G19440) and other including a *LIPASE CLASS 3* (AT1G06250), two *ACYL-COA*
138 *OXIDASES* (*ACX3*: AT1G06290 and *ACX6*: AT1G06310) and two *FATTY ACID DESATURASE* (*CER17*:
139 AT1G06350, ADS4.2: AT1G063609); one in Chromosome 3 harboring the *FATTY ACID DESATURASE*
140 (*FAD2:* AT3G12120) and the last QTL in Chromosome 5, harboring a *LYSOPHOSPHOLIPASE*
141 (AT5G20060) (Supplementary Dataset 3, Figure 1a).

142 ***3-KETOACYL-COENZYME A SYNTHASE4* is involved in fatty acid elongation and has major impact on**
143 **polyunsaturated TAG levels under heat and dark stress**

144 Remarkably, 22 polyunsaturated TAGs (puTAGs) co-localized to the QTL on chromosome 1,
145 harboring *KCS4* (Figure 1a). This QTL was not found in control conditions, and the lead SNP in the
146 locus (m11481) co-localized with LOD ≥ 8 (Bonferroni corrected p-value < 0.05) for 16 of the 22
147 associated TAGs (e.g., TAG 54:9, Figure 1b; Supplementary Dataset 2). Besides *KCS4*, 33 other genes
148 mapped on the QTL, spanning 107,318 bp (Supplementary Dataset 3). However, with the exception
149 of *KCS4,* none of the others showed any homology to proteins known to be involved in lipid
150 metabolism.

151 *KCS4* belongs to a gene family involved in the elongation of VLCFA. Twenty-one *KCS* genes have been
152 previously identified in the Arabidopsis genome. The *KCS* family encodes for enzymes involved in the
153 first and often described as rate-limiting step of fatty–acid elongation (FAE) (Blacklock and Jaworski,
154 2006; Joubès et al., 2008). KCS4's closest homologues are: KCS9 (78% aa identity), involved in the
155 elongation of C22 to C24 acyl-CoAs (Kim et al., 2013); KCS17 (72% aa identity); and KCS18 (64% aa
156 identity), involved in elongating C18 acyl-CoAs to C20 and C22 in seeds and shown to drive the
157 natural variation in the abundance of VLCFAs in this organ (Rossak et al., 2001). However, KCS4
158 specificity and function have not been clearly characterized so far. TAGs associated with the *KCS4*
159 QTL were exclusively polyunsaturated (e.g., puTAG 52:6, 52:7, 54:6, 54:8, 54:9, 56:6, Supplementary.
160 Table 2-3) and correlated positively with each other, grouping together in a single well-defined
161 cluster (Figure 1c).

162 To identify the alleles involved in puTAG variation under HD, we constructed haplotypes considering
163 the four SNPs from *KCS4* scored in the 250K SNPchip (m11502, m11503, m11504, m11505,
164 Supplementary Dataset 2). The population divided into nine non-unique haplotypes, which were
165 therefore, further considered as being putatively involved in the QTL variation. Mean levels of
166 puTAG 54:9 varied significantly among haplotypes (Figure 1d). Furthermore, m11503 was identified
167 as a non-synonymous SNP (nsSNP). In the majority of accessions, SNP m11503 encodes for a
168 methionine [hereafter *KCS4*(Met)]. Accessions carrying this allele have higher levels of puTAG in the
169 stress condition than accessions carrying the minor allele, which encodes for a threonine [hereafter
170 *KCS4*(Thr)] (Figure 1e). These results suggest that the nsSNP (m11503) in *KCS4* is the causal mutation
171 underlying puTAG variation under HD conditions in Arabidopsis.

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

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

177 For this purpose, we utilized own data sets: control (CD), 3 days (3D) extended darkness, and 6 days
178 (6D) extended darkness (Supplementary Dataset 1) (Zhu et al., 2022). The GWAS at a genome-wide
179 significance level of LOD ≥ 5.3 resulted in 230, 255 and 342 SNP–trait associations grouped in 73, 81,
180 and 70 QTL for CD, 3D, and 6D, respectively (Supplementary Dataset 2, Supplementary Dataset 3).
181 The locus harboring *KCS4* associated strongly in 3D and 6D with 8 and 13 puTAGs, respectively. Lead
182 SNP and other features of the association matched with GWAS results from the HD experiment
183 (Supplementary Figure 2). Thus, darkness alone is enough to trigger the association of *KCS4* with
184 puTAGs and could be related to the carbon status of the plant. Expression of *DIN1* (AT4G35770) and
185 *BCAT2* (AT1G10070), two genes usually used as starvation markers (Moraes et al., 2019) are
186 upregulated under HD conditions (Supplementary Figure 3), showing that HD is also triggering
187 carbon starvation. In summary, an additive effect is observed when both stresses are applied
188 together, determining a stronger association of *KCS4* to puTAGs under HD.

189 **Dissection of *KCS4* alleles involved in TAG variation, pattern of linkage disequilibrium and**
190 **signatures of selection**

191 We next mined the full sequence information available for 149 accessions from the GWAS
192 population (http://signal.salk.edu/atg1001/3.0/gebrowser.php), for 80 kb around SNP m11481 on
193 chromosome 1. The sequence alignment (Supplementary File 1) was used to build a SNP matrix
194 (Supplementary Dataset 4) to perform GWAS with more SNPs than those included in the previous
195 experiments using the 250K SNPchip (Horton et al., 2012). We analyzed a total of 1,082 SNPs and
196 short insertion-deletions, in addition to the 131 SNPs previously tested in this region. These analyses
197 aim to discard genetic heterogeneity among other genes in the locus, analyze the pattern of linkage
198 disequilibrium (LD) in the region, determine whether more SNPs in *KCS4* are additionally involved in
199 the variation, and evaluate signatures of selection. The results are shown in Figure 2, exemplified for
200 TAG 54:9. According to the *KCS4* gene model (Figure 2a), we identified 35 additional SNPs in *KCS4*
201 genomic region apart from the 4 SNPs scored in the 250K SNPchip (19 in the promoter, two in the
202 5'UTR, 13 in the coding region, and one in the 3'UTR; Supplementary Dataset 5). Fewer TAGs with
203 lower LOD values showed association with *KCS4* gene region, as expected when analyzing a reduced
204 number of accessions (< 100) (Supplementary Dataset 5, Figure 2b). The lead SNP using
205 resequencing information was a polymorphism not tested before (rs21933), but in high LD with
206 m11481 (r^2 =0.73, p << 0.001). In addition, a number of polymorphisms in the upstream region of
207 *KCS4,* not tested in the SNPchip, were identified as being significant, pointing to a putative combined
208 regulation at the promoter region mediated by rs39451, rs39499 and m11502 and at the coding
209 region mediated by the nsSNP m11503 (Figure 2a). Surprisingly, several SNPs in the *KCS4* coding
210 region did not associate (Figure 2b).

211 All SNPs in the locus harboring KCS4 with high LOD values are in high LD with each other and in high
212 LD with the lead SNPs rs21933 and m11481 (r^2 > 0.2, p < 0.001, Figure 2c). The remaining SNPs in the
213 *KCS4* coding region are not in LD with the lead SNPs nor with the nsSNP m11503 in the *KCS4* coding
214 region. This result suggests that these mutations may have appeared in a different time during
215 evolution or they may have accumulated later, as a consequence of specific recombination, to
216 increase allele frequency at other SNP positions. To evaluate signatures of selection, we tested
217 Tajima's D for the 80 kb around m11481 (Figure 2d). Tajima's D reached significantly negative values
218 along the 80 kb, particularly around SNPs significantly associated in the QTL: rs21933, m11481,

219 rs39451, rs39499, m11502, m11503. This result indicates that the region is under a high selective
220 pressure and supports the selection of *KCS4* as the candidate gene involved in puTAG variation.

221 **Transcriptional regulation of *KCS4* in Arabidopsis**

222 Several SNPs in the promoter region showed significant LD with the non-synonymous SNP m11503.
223 Therefore, we deepened on the transcriptional regulation of *KCS4* in Arabidopsis (Figure 3). We
224 analyzed 1 kb upstream of the transcription start site for the known transcription factor binding sites
225 using the DAP-Seq database (O'Malley et al., 2016). Several transcription factor (TF) families
226 belonging to land plants could likely bind to the *KCS4* promoter region, indicating that *KCS4*
227 expression is controlled under several regulatory processes (Figure 3a). We furthered our analysis by
228 looking at the core binding sites of the highly enriched TF families that contained binding motifs
229 across SNPs in the promoter region of *KCS4* (rs38889, rs39451, rs39499 and m11502). The top-
230 enriched DOF family (DNA-binding One zinc Finger) does not have a binding site colocalizing with
231 polymorphisms, but in other regions of the promoter (Figure 3b). Similarly, the NAC and WRKY TF
232 families bind to non-polymorphic regions in the *KCS4* promoter, and MYB TF binds near a SNP but in
233 the coding region. Interestingly, HSF (heat shock factor), the second highly enriched TF family has a
234 core binding motif (heat shock element: 5'nnGnAnnTnCtn 3') within SNP m11502 in the promoter
235 region.

236 We performed an electrophoretic mobility shift assay (EMSA) to assess the physical interaction of
237 Heat Shock Factor A2 (HSFA2) with the *KCS4* promoter sequence. Additionally, we tested if alleles
238 (C/T) at SNP m11502 influence the binding affinity. HSFA2, a well-studied HSF family member, has
239 been shown to bind to the heat shock element in Arabidopsis (Lämke et al., 2016). As shown in
240 Figure 3c, HSFA2-GST fusion protein bound to the 5-DY682-labeled *KCS4* promoter fragment *in vitro*
241 with the C allele at m11502, resulting in a retardation band on the gel (Figure 3c, lane 2) compared
242 to the free probe (lane 1). Interaction between HSFA2 and the *KCS4* promoter region containing the
243 binding site appears to be specific, as the addition of an excessive unlabeled competitor (same
244 sequence, 200x more) significantly reduced the binding to the labeled probe (Figure 3c, lane 3).
245 Finally, we tested a competitor sequence having the alternative allele at m11502 (T). The latter
246 reduced the binding compared to the labeled probe (Figure 3c, lane 6), suggesting that the T allele at
247 m11502 determines a stronger binding of HSF2A to the *KCS4* promoter region. Overall, this indicates
248 that polymorphisms at the promoter region could exert a dosage effect of transcription factors,
249 influencing the expression of *KCS4*.

250 Furthermore, we analyzed natural variation in transcript levels of *KCS4* for the CHD and HD
251 conditions in the complete GWAS panel (Figure 3d). *KCS4* transcripts are significantly lower in HD
252 compared to CHD. This correlates significantly negatively with highly unsaturated TAG levels (r^2 = -
253 0.66, $p < 0.001$), which increased in HD (e.g., Figure 3e, TAG 54:9). In HD, *KCS4* transcript levels are
254 not significantly different for accessions carrying different alleles at m11502. This confirms that
255 neither differences in TF binding-affinity nor expression levels of *KCS4* determine variation of puTAG
256 levels under HD (Figure 3e). In other words, polymorphisms at *KCS4* promoter region are not the
257 causal for puTAG variation in Arabidopsis natural population.

258 **Expression patterns of KCS family members under diverse environmental conditions**

259 We analyzed the time-resolved response of Arabidopsis Col-0 accession towards light and
260 temperature for 16 of the 21 gene members using data previously obtained (Caldana et al., 2011).
261 Focusing on 24 h, most *KCS* members, including *KCS4* are upregulated under light conditions and
262 downregulated in darkness, irrespective of the temperature. However, *KCS8*, *KCS12*, *KCS16*, *KCS18*,
263 and *KCS21* showed an opposite behavior. *KCS4* transcript levels are downregulated both under

264 darkness and under HD, but transcript levels are much lower in HD, strengthening the idea of a
265 synergistic effect between light and temperature stress in the *KCS4* response. In addition, *KCS5*, *KCS9*
266 and *KCS20* showed a similar transcriptional response to *KCS4* across several environmental stimuli
267 (Figure 3f).

268 ***KCS4* functional validation in puTAG variation**

269 To validate the role of *KCS4* alleles in puTAG variation under HD stress, we selected five accessions:
270 three carrying *KCS4*(Met) and two carrying *KCS4*(Thr). Selection of these accessions was based on
271 puTAG levels measured in the HD experiment (Supplementary Dataset 1). Mutant lines (*kcs4*) were
272 obtained through CRISPR-Cas9 by introducing a 57/58–bp deletion at position +18 from the ATG
273 (Supplementary File 2). This deletion introduced a premature stop codon resulting in a truncated
274 protein of 16 aa compared to the wild-type version of 516 aa. Lines were grown under CHD or HD as
275 described in materials and methods and harvested to perform lipidomic analysis. Since natural
276 variation is detected at a population level, we compared average values for *kcs4*(Met) and *kcs4*(Thr)
277 allelic mutants against wild-type *KCS4*(Met) and *KCS4*(Thr) alleles (Figure 4).

278 Under control conditions (CHD), puTAG levels (X:4 to X:9) are significantly different between
279 *KCS4*(Met) and *kcs4*(Met) lines. Mutants for *kcs4*(Met) presented, in average, lower levels of
280 puTAGs, similar to the levels found for *KCS4*(Thr) accessions (Figure 4 a-c, j-k). This phenomenon is
281 seen for the accessions pooled together as well as for individual accessions (data not shown). On the
282 other hand, puTAG levels did not differ between *KCS4*(Thr) and *kcs4*(Thr). These results support
283 *KCS4*(Met) as the stronger allele and the causal polymorphism for polyunsaturated TAG variation.

284 Under HD condition, a higher accumulation of puTAGs is seen for accessions carrying *KCS4*(Met)
285 allele (Figure 4 d-f, j-k). However, overall levels of TAGs (saturated and polyunsaturated) increase
286 independently of *KCS4* alleles. This is in agreement with the phenotype observed for the complete
287 mapping panel. Here, mutants don't differ in puTAG levels from their wild-type counterparts, as if
288 the massive production of puTAGs under stress overrides the solely effect of *KCS4* mutation.

289 Transient expression of At*KCS4*(Met) in *N. benthamiana* leaves showed significantly higher levels of
290 puTAGs for classes 52, 54, and 56 compared to control plants (Figure 4 g-i). Conversely, TAGs with
291 none or a few unsaturated bonds tend to be higher in control plants. These results show that the
292 transient expression of At*KCS4*(Met) changes the TAG profile in *N. benthamiana*.

293 In summary, the reduction of puTAGs in the *kcs4*(Met) mutant (Figure 4 a-c, j-k), together with the
294 increasement of puTAGs when *KCS4*(Met) in ectopically expressed in *N. benthamiana* (Figure 4 g-i),
295 confirms that *KCS4*(Met) is the causal polymorphism for polyunsaturated TAG variation and validate
296 GWAS results.

297 ***KCS4* functional validation in the accumulation of cuticular waxes**

298 VLCFAs are produced by the elongation complex (i.e., KCS, KCR: β-KETOACYL REDUCTASE, HCD: β-
299 HYDROXYACYL-COA DEHYDRATASE and ECR: ENOYL-COA REDUCTASE) in the Endoplasmic Reticulum
300 (ER) and used as precursors for the synthesis of lipid barriers such as cuticular waxes. Therefore, we
301 determined the wax coverage for *KCS4* wild-type alleles and mutants as a proxy of the enzyme
302 activity *in planta* (Figure 4 l-m). Under CHD, lines carrying *KCS4*(Met) showed higher content of
303 waxes than lines carrying *KCS4*(Thr). The mutant *kcs4*(Met) showed a significant reduction of the
304 major components of leaf waxes: C29 and C31 *n*-alkanes, C26 and C28 *n*-alcohols, C30 and C32
305 aldehydes, C28 and C30 fatty acids. On the contrary, *kcs4*(Thr) does not show differences to its wild-
306 type counterpart. Under HD, the total wax content is increased in the different lines and significant
307 differences are seen between Met and Thr alleles but not between wild type and mutants. These

308 results confirm that the wax coverage and composition is different for both alleles, and that other
309 members of the KCS family cannot fully complement the synthesis of waxes coming from FA
310 elongated by KCS4.

311 **KCS4 enzymatic activity and cellular localization**

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

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

319 We further expressed *AtKCS4*(Met) in the Δ*elo3* yeast strain, lacking the endogenous ELO3 elongase
320 (Figure 5b). The Δ*elo3* strain accumulates C20 and C22 VLCFA (Han et al., 2002). Expression of KCS4
321 led to the elongation of VLCFAs from C22 to C24 and from C24 to C26 (Figure 5b). We also analyzed
322 heterologous expression of At*KCS9* as it is the closest *KCS4* homolog (Figure 5b). KCS9 elongates
323 VLCFAs from C22 to C24 (Kim et al., 2013 and Figure 5b). These results suggest that KCS4 has a
324 differential activity compared to KCS9, pushing further the elongation of VLCFAs up to C26.

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

332 We further transformed Δ*elo3* strain with *KCS4* wild-type and mutant alleles (Supplementary Figure
333 4). Expression of both *KCS4*(Met) and *KCS4*(Thr) alleles led to the elongation of C22 into C24 up to
334 C26 VLCFA in the Δ*elo3* strain with slight differences in the amount of C22 and 2OH-C24
335 (Supplementary Figure 4a). On the other hand, yeast expressing allelic mutants do not show exactly
336 the composition of the control line, C26 is not accumulated but C22 and C24 VLCFA are still
337 produced at low level compared to yeast expressing wild-type alleles, as if some activity remains.
338 This residual accumulation of VLCFA in the mutants indicates that *kcs4*(Met) and *kcs4*(Thr) lines still
339 have some enzymatic activity, although very weak.

340 In addition, we performed a time-point experiment at 48, 72 and 96 h for Δ*elo3* strain expressing
341 *KCS4*(Met) *and KCS4*(Thr) wild-type alleles (Supplementary Figure 4b). We could detect that, yeasts
342 expressing *KCS4*(Met) ended up with more C26 and 2OH:C26 than yeasts expressing *KCS4*(Thr),
343 supporting the hypothesis of a differential enzymatic activity between *KCS4* alleles.

344 KCS1 expression in yeast leads to the production of C18, C20 and C22 VLCFA (Trenkamp et al., 2004;
345 Tresch et al., 2012), KCS2 can elongate from C20 to C22 (Millar et al., 1999; Franke et al., 2009),
346 KCS20 mainly produces C22 and C24 VLCFA, KCS5 and KCS6 mainly produce C24 to C28. All these
347 proteins have different expression profiles under heat and dark stress, some of them upregulated
348 and some of them downregulated (Figure 3f). Expression profiles of KCS4, KCS5, KCS9 and KCS12 are
349 also altered in *kcs4* CRISPR-Cas lines, evidencing a delicate expression balance under stress, between
350 KCS family proteins (Supplementary Figure 5).

351 **Biomass is affected by KCS4 function**

352 It has been demonstrated that a mild reduction in VLCFA content improve cell proliferation and
353 shoot growth (Nobusawa et al., 2013). To investigate if *KCS4* exert an effect in plant growth, we
354 compared biomass for *KCS4* wild-type accessions and mutant lines. We collected full rosettes from
355 plants grown in CHD and subjected to HD, immediately after treatment. The *kcs4* mutants grew
356 significantly bigger than their WT counterparts (Figure 6). This indicates that biomass is negatively
357 correlated with KCS4 activity and further supports both the idea that mutations at *KCS4* reduce the
358 overall content of VLCFA and that KCS4 activity cannot be overcome by other family members.

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

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

367 In addition, climate conditions from the geographical origins of accessions (Fick and Hijmans, 2017)
368 were also analyzed to identify putative correlations with lipidomic profiles. We used a lasso
369 regularized linear model to predict each climate parameter from the lipidomic profiles. In result, we
370 estimated the prediction power for each best-fitting model, and the lipids, being the best predictors
371 for the climate parameters. While the CHD condition did not provide significant results, the HD
372 condition highlighted in total 15 significant model fits, five of which gave also significant results in a
373 cross-validation test (Supplementary Figure 7). All of these five significant variables are related to
374 temperature. In total, six "climate-predictive lipids" were also significantly associated with *KCS4*,
375 with TAG 58:6, TAG 60:6, and TAG 56:8 being linked to the temperature in the driest and the coldest
376 quarter of the year.

377 Finally, cluster analysis was performed for climate parameters and TAG levels, by grouping
378 accessions according to *KCS4* haplotypes (m11502-m11503: CC, CT, TT, TC), and significant
379 differences between mean values for haplotypes were tested using non-parametric ANOVA (Figure
380 7). Eight climate parameters plus latitude and longitude differed significantly between haplotypes (p
381 < 0.05, Figure 7a). Moreover, five of them co-clustered with TAGs associated with *KCS4*, four of them
382 related to precipitation and one to temperature (isothermality). The "TT" haplotype in most cases,
383 presented the lowest value for these parameters, suggesting that the allelic effect of these two SNPs
384 is opposite for climate variables (Figure 7b).

385 **Discussion**

386 Recent years have witnessed the successful application of metabolic GWAS (Luo, 2015). In lipids, the
387 effort has focused more on dissecting the genetic architecture of oil biosynthesis (Correa et al.,
388 2020). Here, we focus on glycerolipid profiles of ~ 300 Arabidopsis natural accessions grown under
389 control conditions and subjected to a sudden environmental change, given by a combined stress of
390 heat and darkness. Our GWAS exposed the high plasticity of the Arabidopsis genome allowing the
391 identification of genes involved in the regulation of lipid metabolism, in an environment-dependent
392 manner (Figure 1a).

393 Lipid-based GWAS in control conditions showed several mild-strength associations, similar to what
394 was found for glycerolipids in rice (Hong et al., 2022) and for primary metabolites in maize and
395 Arabidopsis (Wen et al., 2015; Wu et al., 2016), which are controlled by a large number of small-

396 effect loci. This could be explained by the fact that both lipids and primary metabolites are essential
397 for the survival of the cell and thus, do not show on/off phenotypes. This is different from natural
398 variation in secondary metabolites, which are controlled, to a large extent, by major large-effect loci
399 (Chan et al., 2011; Matsuda et al., 2012; Wen et al., 2014; Wu et al., 2018).

400 In our study, applying combined heat and darkness resulted in a strong perturbation of lipid
401 homeostasis. This led to the identification of a large-effect QTL at Chromosome 1, harboring *KCS4*.
402 *KCS4* associated with several polyunsaturated TAGs (puTAGs) specifically under carbon starvation
403 conditions (HD, 3D and 6D) in a highly robust manner (Figure 1b, Supplementary Figure 2b,
404 Supplementary Dataset 2/3). Under HD, all puTAGs mapping to *KCS4* are higher than under CHD, but
405 accessions carrying *KCS4*(Met) always have higher levels of puTAGs than accessions carrying
406 *KCS4*(Thr). PuTAGs that mapped to *KCS4* clustered together and correlated negatively to saturated
407 TAGs, further illustrating the role of *KCS4* in the specific regulation of puTAGs (Figure 1c).

408 We have provided several lines of proof that both *KCS4* is the causal gene for TAG variation and
409 *KCS4*(Met) allele determines higher levels of puTAGs under stresses (HD, 3D, 6D): (i) linkage
410 disequilibrium, correlation among lipid species and haplotypic analysis point *KCS4* as the candidate
411 gene for puTAG variation under stress (Figure 1, Figure 2), (ii) transient overexpression of *KCS4*(Met)
412 in *N. benthamiana* mimics the phenotype we observed in the GWAS population grown under stress
413 conditions (Figure 4 g-i), (iii) lipidomic analysis of allelic mutants (*kcs4*(Met)/*kcs4*(Thr)) and their
414 respective wild-type accessions (*KCS4*(Met)/*KCS4*(Thr)) shows a clear phenotype on puTAG, VLCFA
415 and wax components (Figure 4, Figure 5, Supplementary Figure 5).

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

423 Under HD, differences in TAG levels for accessions carrying *KCS4* opposite alleles [*KCS4*(Met) vs.
424 *KCS4*(Thr)] become more significant as the number of double bonds in the TAGs increases (Figure 4
425 d-f, g-i). In addition, TAG levels are significantly different when comparing *KCS4*(Met) with its allelic
426 mutant *kcs4*(Met) in control conditions (CHD) (Figure 4 a-k). In the same line of evidence, the mutant
427 *kcs4*(Met) showed a significant reduction of the major components of leaf waxes compare to its
428 wild-type allele in CHD (Figure 4 l-m). Not unexpectedly, *kcs4*(Thr) did not show any differences in
429 TAG levels of wax compounds compared to *KCS4*(Thr). It has been previously shown in GWAS, that
430 mutating the weaker allele did not always result in clear phenotypic changes for the traits under
431 study (Todesco et al., 2010; Fusari et al., 2017; Wang et al., 2022; Yang et al., 2022).

432 The phenotype can also depend on the context and a change in either the genetic or the
433 environmental background can amplify or remove the effect (Tonsor et al., 2005). Under HD, there is
434 a strong metabolomic and expression reprogramming as shown before for lipids and primary and
435 secondary metabolites (Burgos et al., 2011; Szymanski et al., 2014; Wu et al., 2016, 2018). This could
436 explain why we cannot observe a phenotype by mutating only one gene. In addition, sometimes
437 effects are not evident when analyzing single mutants, particularly if there are gene families
438 encoding for the same or very similar enzymatic activities. For example, in studies of auxin
439 conjugates, IAA showed a reduction only when triple amidohydrolase mutants were analyzed
440 (Ludwig-Müller, 2011). *KCS4* is one of 21 genes encoding for KCS activity (Joubès et al., 2008; Guo et

441 al., 2016) and we have shown that expression of other KCS members is altered in *kcs4* allelic
442 mutants, indicating a complex regulation between *KCS* genes (Supplementary Figure 5). Of note, the
443 nature of this complex regulation was recently demonstrated to involve protein-protein interactions
444 within the elongase complex (Kim et al., 2022).

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

457 In an event of sudden extended darkness (3D, 6D) or combined heat and darkness (HD), plants
458 become carbon starved (Supplementary Figure 3) (Stitt and Zeeman, 2012). It has been shown that
459 increasing night temperature accelerates enzyme activities exhausting carbon resources prematurely
460 (Pilkington et al., 2015). In this case, FAs are used as alternative substrates for respiration via β-
461 oxidation (Fan et al., 2014). There are many lines of evidence that support the idea that plant lipids
462 are alternative sources of carbon and energy under stress, particularly under carbon starvation. For
463 instance, Kunz et al. (2009) showed that by knocking out the PXA1 transporter, the import of FFA
464 into the perixosome is blocked, resulting in plants dying faster in extended darkness. Fan et al.
465 (2017), through the study of SDP1/LIP1 mutants, could define the crucial role that TAG metabolism
466 has in membrane lipid breakdown, fatty acid turnover, and plant survival under extended darkness.
467 Other study showed that there are different metabolic strategies employed by leaves in response to
468 different darkening treatments (Law et al., 2018). These strategies involve either importing sugars
469 from sink sources, metabolizing amino acids or activating the turnover of lipids from the plastid
470 membranes to the b-oxidation in the peroxisome. The last strategy is particularly important after 6
471 days of dark treatment. Finally, Yu et al. (2018) demonstrated that the disruption of starch synthesis
472 resulted in a significant increase in FA synthesis via post-translational regulation of the plastidial
473 acetyl-coenzyme A carboxylase and a concurrent increase in the synthesis and turnover of
474 membrane lipids and triacylglycerol.

475 To summarize our findings, we present a model of lipid remodeling under heat and dark stress in
476 Arabidopsis (Figure 8). In the proposed model, darkness induces the predominant response within
477 the combined stress, blocking *de novo* FA synthesis (*de novo* FAS) and producing considerable
478 membrane-lipid remodeling. We observed decreased levels of MGDGs and DGDGs across accessions,
479 as previously shown (Burgos et al., 2011). Under carbon starvation, FAs, mostly 16:3 and 18:3, are
480 released from galactolipids by lipases (Higashi et al., 2015; Fan et al., 2017). FAs are then exported to
481 the ER, where KCS4 is localized and acts as a branch point in the fate of FAs, directing the saturated
482 ones to the very-long-chain fatty-acid pathway (VLCFA). Two *KCS4* alleles have differential capacity
483 to allocate saturated FAs into the VLCFA pathway, with a major -more efficient- allele, *KCS4*(Met),
484 and a minor -weaker- allele, *KCS4*(Thr). *KCS4*(Met) accessions are, as well, efficiently channeling the
485 saturated FA from the Acyl-CoA pool to cuticular waxes, as they present higher accumulation of
486 waxes than *KCS4*(Thr) accessions under stress (Figure 4). Polyunsaturated TAG levels are, therefore,

487 higher in *KCS4*(Met) accessions comparing to *KCS4*(Thr) accessions, as the result of a rise in the ratio
488 of polyunsaturated-to-saturated FAs that proceed to TAG formation (Figure 4). This effect of
489 *KCS4*(Met) on TAG levels is only observed under stress conditions, when there is an increased
490 amount of free-polyunsaturated fatty acids coming from membrane lipids, and when it necessary to
491 sequestered them into TAGs, to protect the cell from the free-fatty-acid-toxicity induced damage
492 (Kao et al., 2018).

493 Our work demonstrates that KCS4 is a key enzyme involved in the regulation of puTAG levels under
494 heat and darkness and under extended darkness alone, adding complexity to lipid-metabolism
495 regulation under carbon starvation. The enzymatic function of KCS4 in the synthesis of VLCFA is
496 affecting the accumulation of the unsaturated Acyl-CoA pool (16:3 and 18:3, for example) and the
497 sequestration of them into puTAGs.

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

504

505 **Methods**

506 **Plant material**

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

510 **Experimental design**

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

515 In control conditions (for heat/dark experiment: CHD, for dark experiment: CD), plants were grown
516 in long days (16h light, 8 h dark), temperature was kept at 21/16°C (day and night, respectively), light
517 intensity at 150 µE m^{-2} sec^{-1}, humidity 75%. In the heat and dark experiment (HD), plants were grown
518 like the plants in CHD for six weeks, then subjected to dark and 32°C continuously for 24 h, and then
519 harvested. In the dark treatment, plants were grown like the plants in CD for six weeks and then
520 transferred to continuous darkness either for 72 h (3D) or 144 h (6D). Plants from HD, 3D, 6D were
521 harvested 24 h, 72 h, and 144 h after the beginning of the treatment, respectively. Between three
522 and six plants from each accession/line were collected for each condition tested.

523 **Targeted lipid profiling by LC–MS and data acquisition**

524 Lipid extraction and measurement from Arabidopsis leaves (100 mg for each sample) were
525 performed as described previously (Hummel et al., 2011). In brief, dried organic phase was
526 measured using Waters Acquity ultra-performance liquid chromatography system (Waters,
527 http://www.waters.com) coupled to Fourier transform mass spectrometry (UPLC–FT–MS) in positive
528 ionization mode. Analysis and processing of raw MS data were done with REFINER MS® 10.0

529 (GeneData, http://www.genedata.com). Workflow included peak detection, retention time
530 alignment and removal of chemical noise and isotopic peaks from the MS data. Obtained mass
531 features characterized by specific peak ID, retention time, m/z values and intensity were further
532 processed using custom scripts in R (R Core Team, 2019). Clusters with mean of signal intensities
533 lower than 50000 were removed and only peaks present in at least 70% of the samples were kept for
534 further analysis. Peak intensities were normalized by the day of measurement and median and
535 subsequently $\log_2$ transformed. After that, obtained molecular features were queried against a lipid
536 database for annotation. The in-house database used includes 219 lipid species of the following
537 classes: diacylglycerols (DAGs), digalactosyldiacylglycerols (DGDGs), monogalactosyldiacylglycerols
538 (MGDGs), phosphatidylcholines (PCs), phosphatidylethanolamines (PEs) and triacylglycerols (TAGs).
539 This was carried out by comparing their retention times and exact masses against those of the
540 reference compounds, allowing maximal deviations of 0.1 minutes and 10 ppm. Identified lipids
541 were confirmed by manual verification of the chromatograms using XCalibur (Version 3.0, Thermo-
542 Fisher, Bremen, Germany) (Alseekh et al., 2021).

543 **General statistical and multivariate analysis of lipidomic data**

544 To calculate the correlation between lipids, Pearson correlation was used. The correlation plots and
545 heatmaps were created using the package pheatmap (Raivo Kolde, 2019). The lipids were clustered
546 using complete-linkage clustering. Principal component analysis (PCA) was obtained from the
547 correlation matrix of lipidomic data (CHD and HD together) using the pcaMethod (Stacklies et al.,
548 2007) package.

549 To detect the fold change in lipid-feature levels across all accessions for each lipid feature, fold
550 change was calculated by dividing the average of the five highest intensities measured in a given
551 accession by the average of the five lowest intensities, A) for all annotated lipids ($N_{CHD} = 98$,
552 $N_{HD} = 109$), B) only for TAGs ($N_{CHD} = 30$, $N_{HD} = 41$), C) for lipid classes other than TAGs ($N_{CHD} = 68$,
553 $N_{HD} = 68$).

554 **GWAS analysis**

555 Data acquisition for GWAS, mapping and locus identification were performed as described
556 previously (Wu et al., 2016; Fusari et al., 2017). Genome-wide association analysis (GWAS) was
557 performed using a compressed mixed linear model (MLM) implemented in the GAPIT (Lipka et al.,
558 2012) package. The 109 (HD), 98 (CHD), 80 (3D), 109 (6D), and 122 (CD) lipid features were used as
559 phenotypes (Supplementary Dataset 1), whereas a set of 214,051 SNPs was used as genotypic data
560 (250K SNPchip). A total of three principal components were included in the MLM to account for
561 population structure and the SNP fraction considered to perform PCA was set to 0.1 to avoid
562 excessive computations. Kinship matrix and other parameters were set to default values.

563 Additionally, we performed GWAS using full genome sequence from a section of chromosome 1.
564 Shortly, sequences for 80 kb around *KCS4* (AT1G19440, TAIR10 position: 6689119 - 6769118) were
565 obtained from the Arabidopsis 1001 Genomes project
566 (http://signal.salk.edu/atg1001/3.0/gebrowser.php, 162 accessions). The alignment was performed
567 using the default routine implemented in MAFFT (Katoh et al., 2019) with a final manual inspection
568 to discard alignment problems due to big gaps or missing data (Supplementary file 1). For nine
569 accessions, two sequences were available. The sequences included further in the analysis were
570 selected based on agreement of alleles with the SNPchip data and/or on % of non-missing data (≥
571 90%). Final SNP matrix, obtained by filtering for minor allele frequency ≥ 5%, contained 1,213 SNPs
572 and indels (Supplementary Dataset 4). This matrix was used as the genotypic data. Population
573 structure and cryptic relationships among accessions were taken into account by including principal

574 components as cofactors and a kinship matrix obtained with SNPs in the 250K SNPchip for the subset
575 of 162 accessions (Supplementary Dataset 4). Lipid features collected formerly (HD and 3D)
576 Supplementary Dataset 1) were used as phenotypes. The three data sets (genotype, population
577 structure and phenotypes) matched for 149 accessions in the HD experiment and for 134 accessions
578 in the 3D experiment. MLM was implemented using TASSEL software with the default settings for
579 compressed MLM (Bradbury et al., 2007).

580 **SNP filtering, QTL identification and candidate gene selection**

581 For results obtained using GAPIT package, SNPs associated with lipid species having a p value < 1/N
582 (N=214,051) were extracted. SNPs were assigned to the same QTL if the genomic distance between
583 them was ≤ 10 kb. All genes in a given QTL were taken as putative candidates. Further candidate
584 gene selection for validation was done based on functional annotation or based on previous
585 knowledge.

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

588 **Linkage disequilibrium, nucleotide diversity, and haplotype analysis**

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

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

600 **Mutant line selection and genotyping**

601 For Arabidopsis CRISPR-Cas9 lines, the binary vector containing multiple gRNAs and the Cas9
602 nuclease were generated based on a classic cloning procedure described previously (Schumacher et
603 al., 2017). A pair of guideRNAs was used to target the N-terminal coding sequence of *KCS4* using the
604 spacer sequences indicated in Supplementary Dataset 8. As background, wild-type accessions
605 carrying *KCS4*(Met): Pro-0, Petergof, Per-1, Col-0; and *KCS4*(Thr): Si-0, Bu-8, Sh-0 were chosen based
606 on TAG levels phenotyped in HD. Mutations were confirmed by PCR and sequencing of *KCS4* using
607 primers listed in Supplementary Dataset 7. Sequence contigs and alignment were done using Bioedit
608 (Hall, 1999).

609 For heterologous expression in yeast and transient expression in *Nicotiana benthamiana, KCS4* and
610 *KCS9* open reading frame was amplified from Arabidopsis flower cDNA using primers listed in
611 Supplementary Dataset 7. The corresponding PCR fragments were inserted into the pDONR™221
612 ENTRY vector using the GATEWAY® recombinational cloning technology (Invitrogen, Darmstadt,
613 Germany) and subsequently transferred into pK7YWG2, pK7WG2D and pvtLEU DESTINATION vectors
614 (Karimi et al., 2002; Dittrich-Domergue et al., 2014) by using LR clonase II. All transient heterologous
615 expressions in *N. benthamiana* were made using the *Agrobacterium tumefaciens* strain GV3101 and
616 five-week-old plants cultivated in controlled conditions (16 h light photoperiod, 25 °C) (Perraki et al.,
617 2012). *N. benthamiana* plants were agroinfiltrated on the abaxial side of the leaves and analyzed 2

618 and 5 days after agroinfiltration for subcellular localization experiments and for overexpression,
619 respectively.

620 For yeast experiments, *Saccharomyces cerevisiae* strains INVSc1 [MATa *his3Δ1 leu2 trp1-289 ura3-*
621 *52*] and Δ*elo3* [MATa *his3Δ1 leu2 trp1-289 ura3-52 elo3Δ*] were used to express *KCS4* wild-type and
622 allelic mutants and *KCS9*.

623 **qRT-PCR and eQTL analysis**

624 *RGS1-HXK1 INTERACTING PROTEIN1/RHIP1* (AT4G26410) was used as reference gene for *KCS4*
625 transcript analysis in CHD and HD for the complete GWAS population (Czechowski et al., 2005).
626 *GAPDH3* (AT1G13440) was used as reference gene for starvation markers and *KCS* transcripts
627 analysis. All primers used are listed in Supplementary Dataset 7.

628 qRT-PCR was performed on three biological replicates per ecotype. Each sample was tested in three
629 technical replicates. qRT-PCR was performed on an ABI Prism® 7900 HT real-time PCR system
630 (Applied Biosystems/Life Technologies, Darmstadt, Germany) in 384-well PCR plates with a total
631 reaction volume of 5µL (2 µL forward and reverse primer mixture, 0.5 µM each), 0.5 µL cDNA and 2.5
632 µL Power SYBR® Green-PCR Master Mix (Applied Biosystems/Life Technologies) using the following
633 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
634 by a denaturation step of 95°C for 15 s, 60°C for 15 s, followed by a continuous temperature increase
635 (0.3°C per s) to 95°C (for 15 s). Steps 5, 6 and 7 were introduced to record a dissociation or melting
636 curve for each product in order to detect non-specific amplification. Expression values were
637 calculated by $2^{(-\Delta\Delta Ct)}$ using either *ACTIN2* or *GAPDH3* as reference genes. Errors are given as the lower
638 ($2^{-(\Delta\Delta Ct-SD)}$) and upper ($2^{-(\Delta\Delta Ct+SD)}$) limit of the standard deviation around the mean.

639 **EMSA**

640 EMSA reactions were performed as previously described (Thirumalaikumar et al., 2018). Briefly,
641 HSFA2-GST proteins were prepared as stated before (Puranik et al., 2011). *HSFA2* coding sequence
642 was PCR-amplified from Arabidopsis heat-stressed cDNA using primers listed in Supplementary
643 Dataset 7. PCR products were recombined in pDEST24 using the GATEWAY cloning system
644 (Invitrogen). Recombinant vectors were transformed into *Escherichia coli* Star (DE3) pRARE
645 generated by transforming the pRARE plasmid isolated from Rosetta (DE3). 5'- labeled with
646 fluorescent dye DY682 and unlabeled *KCS4* promoter regions spanning the SNP position m11502 (C
647 or T allele) were purchased from Eurofins MWG operon. DNA EMSA reactions were performed using
648 Odyssey Infrared EMSA kit (Li-COR, Bad Homburg, Germany). Competitor was added at 200x the
649 amount of probe.

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

651 For fatty acid methyl esters (FAMEs) analysis in *N. benthamiana*, three agroinfiltrated leaf discs were
652 transferred to tubes containing 1 mL 0.5 M sulfuric acid in methanol containing 2% (v/v)
653 dimethoxypropane and 20 µg of heptadecanoic acid (C17:0) as internal standard. Transmethylation
654 was performed at 85°C for 3 hours. For FAMEs analysis in yeast, yeasts were grown in 5 mL of
655 appropriate liquid minimal medium for 1 week at 30°C, then pelleted and washed in 2 mL of 2.5%
656 (w/v) NaCl. Yeast cell pellets were resuspended in 1 mL 0.5 M sulfuric acid in methanol containing
657 2% (v/v) dimethoxypropane and 20 µg of heptadecanoic acid (C17:0) as internal standard and
658 transmethylation was performed at 85°C for 1 hour. After cooling, 2 mL of 2.5% (w/v) NaCl was
659 added, and fatty acyl chains were extracted in 300 µL of hexane. Samples were analyzed by GC-FID
660 and GC-MS as previously described (Domergue et al., 2010).

661 **Cuticular wax analysis**

662 Epicuticular waxes were extracted from leaves by immersing tissues for 30 s in chloroform
663 containing docosane (C22 alkane) as internal standard. Extracts were derivatized and analyzed as
664 previously described (Bourdenx et al., 2011).

665 **Confocal microscopy**

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

672 **Biomass determination**

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

678 **Geographic and phylogenetic signal of the lipidome**

679 The neighbor-joining tree was constructed by BioNJ algorithm (Gascuel, 1997), using J-C distance in
680 SeaView 4.4.3 software (Gouy et al., 2010). Phylogenetic signal was computed using K statistics and
681 as a background a Brownian motion model of trait evolution (Blomberg et al., 2003) in the 'picante'
682 R package (Kembel et al., 2010). Significance of the phylogenetic signal was estimated by comparison
683 of observed variance patterns to a null model obtained by 9,999 random permutations of the
684 accession labels across the phylogeny. The altitude at the sites was estimated using the geomapdata
685 database (Lees, 2000) based on the latitude and longitude coordinates. The climate conditions have
686 been downloaded from the WorldClim database (www.worldclim.org (Fick and Hijmans, 2017)). The
687 lipidomic data were regressed against the geodesic distance matrix using non-parametric
688 multivariate variance analysis (Anderson, 2001). The relationship between lipid levels and the
689 climate parameters was analyzed using regularized regression model (Friedman et al., 2010). The
690 regularization parameter alpha and penalty parameters lambda were estimated using 10-fold cross-
691 validation within the training set. In result the Elastic Net regularization with parameter alpha = 0.5
692 was selected as optimal and applied for all regression models and the 'λmin + 1SE'-rule was applied
693 to choose the penalty parameter for each model separately. Prediction power was estimated in
694 Monte Carlo cross-validation test with 100 samples with 57% and 25% of samples taken as a training
695 and test set respectively. Kruskal-Wallis and Wilcoxon rank sum tests have been performed using the
696 R stats package.

697 **References**

698 Alseekh, S. et al. (2021). Mass spectrometry-based metabolomics: a guide for annotation,
699 quantification and best reporting practices. Nature Methods 2021 18:7 18: 747–756.

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

702 Baxter, I., Brazelton, J.N., Yu, D., Huang, Y.S., Lahner, B., Yakubova, E., Li, Y., Bergelson, J.,
703 Borevitz, J.O., Nordborg, M., Vitek, O., and Salt, D.E. (2010). A coastal cline in sodium

704 accumulation in Arabidopsis thaliana is driven by natural variation of the sodium
705 transporter AtHKT1;1. PLoS Genet 6: e1001193.

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

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

710 Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Léger, A., Roby, D., Pervent, M., Vile, D.,
711 Haslam, R.P., Napier, J.A., Lessire, R., and Joubès, J. (2011). Overexpression of Arabidopsis
712 ECERIFERUM1 promotes wax very-long-chain alkane biosynthesis and influences plant
713 response to biotic and abiotic stresses. Plant Physiol 156: 29–45.

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

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

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

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

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

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

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

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

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

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

746 Fick, S.E. and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces
747 for global land areas. International journal of climatology 37: 4302–4315.

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

751 Franke, R., Höfer, R., Briesen, I., Emsermann, M., Efremova, N., Yephremov, A., and Schreiber,
752 L. (2009). The DAISY gene from Arabidopsis encodes a fatty acid elongase condensing
753 enzyme involved in the biosynthesis of aliphatic suberin in roots and the
754 chalaza-micropyle region of seeds. The Plant Journal 57: 80–95.

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

757 Fusari, C.M., Kooke, R., Lauxmann, M.A., Annunziata, M.G., Enke, B., Hoehne, M., Krohn, N.,
758 Becker, F.F.M.M., Schlereth, A., Sulpice, R., Stitt, M., and Keurentjes, J.J.B. (2017).
759 Genome-wide association mapping reveals that specific and pleiotropic regulatory
760 mechanisms fine-tune central metabolism and growth in Arabidopsis. Plant Cell 29: 2349–
761 2373.

762 Gacek, K., Bayer, P.E., Bartkowiak-Broda, I., Szala, L., Bocianowski, J., Edwards, D., and Batley, J.
763 (2017). Genome-wide association study of genetic control of seed fatty acid biosynthesis
764 in Brassica napus. Front Plant Sci 7: 1–13.

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

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

769 Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: a multiplatform graphical
770 user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:
771 221–224.

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

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

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

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

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

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

787 Hong, J. et al. (2022). Genetic architecture of seed glycerolipids in Asian cultivated rice. Plant
788 Cell Environ.

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

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

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

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

799 Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY^{TM} vectors for Agrobacterium-mediated
800 plant transformation. Trends Plant Sci 7: 193–195.

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

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

806 Kim, J., Jung, J.H., Lee, S.B., Go, Y.S., Kim, H.J., Cahoon, R., Markham, J.E., Cahoon, E.B., and
807 Suh, M.C. (2013). Arabidopsis 3-ketoacyl-coenzyme a synthase9 is involved in the
808 synthesis of tetracosanoic acids as precursors of cuticular waxes, suberins, sphingolipids,
809 and phospholipids. Plant Physiol 162: 567–580.

810 Kim, J., Kim, R.J., Lee, S.B., and Suh, M.C. (2022). Protein–protein interactions in fatty acid
811 elongase complexes are important for very-long-chain fatty acid synthesis. J Exp Bot 73:
812 3004–3017.

813 Kim, J., Lee, S.B., and Suh, M.C. (2021). Arabidopsis 3-Ketoacyl-CoA Synthase 4 is Essential for
814 Root and Pollen Tube Growth. Journal of Plant Biology 64: 155–165.

815 Kunz, H.H., Scharnewski, M., Feussner, K., Feussner, I., Flügge, U.I., Fulda, M., and Gierth, M.
816 (2009). The ABC Transporter PXA1 and Peroxisomal β-Oxidation Are Vital for Metabolism
817 in Mature Leaves of Arabidopsis during Extended Darkness. Plant Cell 21: 2733–2749.

818 Lämke, J., Brzezinka, K., Altmann, S., and Bäurle, I. (2016). A hit-and-run heat shock factor
819 governs sustained histone methylation and transcriptional stress memory. EMBO J 35:
820 162–175.

821 Law, S.R., Chrobok, D., Juvany, M., Delhomme, N., Lindén, P., Brouwer, B., Ahad, A., Moritz, T.,
822 Jansson, B.S., Gardeström, P., and Keech, O. (2018). Darkened Leaves Use Different
823 Metabolic Strategies for Senescence and Survival. Plant Physiol 177: 132–150.

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

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

829 Li-Beisson, Y. et al. (2010). Acyl-Lipid Metabolism. Arabidopsis Book 8: e0133.

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

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

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

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

839 Millar, A.A., Clemens, S., Zachgo, S., Michael Giblin, E., Taylor, D.C., and Kunst, L. (1999). CUT1,
840 an Arabidopsis Gene Required for Cuticular Wax Biosynthesis and Pollen Fertility, Encodes
841 a Very-Long-Chain Fatty Acid Condensing Enzyme. Plant Cell 11: 825–838.

842 Moraes, T.A., Mengin, V., Annunziata, M.G., Encke, B., Krohn, N., Höhne, M., and Stitt, M.
843 (2019). Response of the Circadian Clock and Diel Starch Turnover to One Day of Low Light
844 or Low CO2. Plant Physiol 179: 1457–1478.

845 Nelson, B.K., Cai, X., and Nebenführ, A. (2007). A multicolored set of in vivo organelle markers
846 for co-localization studies in Arabidopsis and other plants. The Plant Journal 51: 1126–
847 1136.

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

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

854 Perraki, A., Cacas, J.-L., Crowet, J.-M., Lins, L., Castroviejo, M., German-Retana, S., Mongrand,
855 S., and Raffaele, S. (2012). Plasma membrane localization of Solanum tuberosum remorin
856 from group 1, homolog 3 is mediated by conformational changes in a novel C-terminal
857 anchor and required for the restriction of potato virus X movement. Plant Physiol 160:
858 624–637.

859 Pilkington, S.M., Encke, B., Krohn, N., Höhne, M., Stitt, M., and Pyl, E.T. (2015). Relationship
860 between starch degradation and carbon demand for maintenance and growth in

861 Arabidopsis thaliana in different irradiance and temperature regimes. Plant Cell Environ
862 38: 157–171.

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

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

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

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

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

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

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

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

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

883 Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). pcaMethods-a
884 bioconductor package providing PCA methods for incomplete data. Bioinformatics 23:
885 1164–1167.

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

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

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

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

895 Todesco, M. et al. (2010). Natural allelic variation underlying a major fitness trade-off in
896 Arabidopsis thaliana. Nature 465: 632–636.

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

900 Trenkamp, S., Martin, W., and Tietjen, K. (2004). Specific and differential inhibition of very-
901 long-chain fatty acid elongases from Arabidopsis thaliana by different herbicides.
902 Proceedings of the National Academy of Sciences 101: 11903–11908.

903 Tresch, S., Heilmann, M., Christiansen, N., Looser, R., and Grossmann, K. (2012). Inhibition of
904 saturated very-long-chain fatty acid biosynthesis by mefluidide and perfluidone, selective
905 inhibitors of 3-ketoacyl-CoA synthases. Phytochemistry 76: 162–171.

906 Wang, P., Guo, L., Morgan, J., Dudareva, N., and Chapple, C. (2022). Transcript and metabolite
907 network perturbations in lignin biosynthetic mutants of Arabidopsis. Plant Physiol.

908 Welti, R., Li, W., Li, M., Sang, Y., Biesiada, H., Zhou, H.E., Rajashekar, C.B., Williams, T.D., and
909 Wang, X. (2002). Profiling membrane lipids in plant stress responses: Role of
910 phospholipase Dα in freezing-induced lipid changes in arabidopsis. Journal of Biological
911 Chemistry 277: 31994–32002.

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

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

918 Wu, S. et al. (2016). Combined use of genome-wide association data and correlation networks
919 unravels key regulators of primary metabolism in Arabidopsis thaliana. PLoS Genet 12: 1–
920 31.

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

925 Yang, L., Wang, Z., Zhang, A., Bhawal, R., Li, C., Zhang, S., Cheng, L., and Hua, J. (2022).
926 Reduction of the canonical function of a glycolytic enzyme enolase triggers immune
927 responses that further affect metabolism and growth in Arabidopsis. Plant Cell 34: 1745–
928 1767.

929 Yu, L., Fan, J., Yan, C., and Xu, C. (2018). Starch deficiency enhances lipid biosynthesis and
930 turnover in leaves. Plant Physiol 178: 118–129.

931 Zhu, F. et al. (2022). Genome-wide association of the metabolic shifts underpinning dark-
932 induced senescence in Arabidopsis. Plant Cell 34: 557–578.

933

934 **Author Contribution**

935 YB and CMF conceived the project and supervised the study. UL conducted GWAS experiments,
936 obtained and selected KO mutants, performed experimental validation with KO, statistical analysis
937 and interpretation of results. A-KR help UL in experimental work and performed analysis of the
938 phenotypic data. JJ and MB conducted experiments on tobacco and yeast lines. JSz conducted the
939 phylogenetic and geographical-signal analysis. VPTK provided expertise and assistance on TF analysis
940 and advised on EMSA experiment and qRT-PCR assay. ML and SW assist in GWAS and multivariate

941 analysis. FZ provided the plant material and carried out the lipid profiling for 3D and 6D GWAS. NE,
942 SK and SM helped UL in qRT-PCR assay for GWAS population and KO validation. HX and WJ
943 performed lipid profiling for *kcs4* mutants. JSc obtained KO constructs for transformation of
944 Arabidopsis. AS provided access to LC-MS. ARF and AS help with discussion. CMF performed TASSEL,
945 LD, Tajima´s D and haplotype analyses, analysed and interpreted the results and helped with data
946 visualization. CMF, UL and YB wrote the paper with input from, JJ, JSz and YL-B. All authors read and
947 draft the manuscript.

948 **Acknowledgements**

949 We would like to thank Prof. Dr. Lothar Willmitzer for discussion, support and the opportunity to
950 develop the project. In addition, we acknowledge Änne Michaelis and Gudrun Wolter for their
951 outstanding technical assistance in LC-MS measurements, and Dr. Karin Kohl for her help in plant
952 cultivation. We thank Hezi Tenenboim for his editorial aid. CMF is a career member of the National
953 Research Council (CONICET) and leader of a Max Planck Partner Group. This research was supported
954 by the Israel Science Foundation (grant no. 859/19).

955 **Competing interests**

956 The authors declare no competing interests.

957 **Figure legends**

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

960 **a.** Chromosome scheme and Quantitative Trait Loci (QTL) identified in the two experimental
961 conditions (CHD = circles, HD = squares), for different lipid classes: phosphatidylethanolamine (PE),
962 phosphatidylcholine (lecithin) (PC), monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol
963 (DGDG), diacylglycerol (DAG) and triacylglycerol (TAG). Color/shape references are included in the
964 figure. Gene names are included only for co-localizing QTL for two or more lipid species with LOD >
965 5.3. **b.** Manhattan plots obtained for TAG 54:9 in CHD (top) and in HD (bottom). The *KCS4* QTL was
966 found at Chromosome 1 only under HD condition. **c.** Pearson correlations coefficients calculated
967 between all lipid species across all accessions, for HD condition. Lipids are color-coded according to
968 class (see reference at right) and clustered using Ward´s distance on pairwise dissimilarity. A zoom-in
969 of the TAG cluster shows the grouping pattern of TAGs according to number of unsaturated bonds
970 and to the GWAS result (left). **d.** Average trait value (intensity of TAG 54:9, log2 scale) for the
971 different *KCS4* haplotypes using SNPs m11502, m11503, m11504 and m11505. The two major
972 haplotypes have the alternative alleles for m11503 (Met or Thr) **e.** TAG 54:9 average value for
973 accessions carrying the *KCS4*(Met) and *KCS4*(Thr) alleles.

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

976 **a.** *KCS4* gene model depicting promoter region (P, green), 5'UTR (orange), coding region (yellow) and
977 3'UTR (orange) of AT1G19440/*KCS4*. Beginning, end and relevant SNP positions are indicated above
978 (TAIR10) and bellow (SNP ID) gene structure. **b.** Manhattan plot of GWAS results (HD: black circles;
979 3D: red circles) obtained for TAG 54:9 using re-sequenced information for 149 accessions. The re-
980 sequenced region on Chromosome 1 (positions 6689119 to 6769118, TAIR10) included 131 SNPs
981 genotyped in the 250K SNPchip plus 1,137 additional polymorphisms. Bonferroni threshold (LOD =
982 4.4) is indicated with a horizontal dashed line. **c**. Linkage disequilibrium (LD) is depicted as a heat
983 map of the coefficient of correlation (r^2). The LD block ($0.8 > r^2 > 0.2$, $P < 0.001$) includes the lead

984 SNPs rs21933 and m11481, SNPs in *KCS4* promoter (rs39451, rs39499, m11502), in *KCS4* genomic
985 region (m11503) and in *KCS4* 3'UTR (m11504). **d.** Tajima's D value in sliding windows for the genomic
986 region analyzed in **(b)**. Gene models and gene orientation are depicted with grey arrows, *KCS4* is
987 highlighted in yellow. Significantly negative Tajima's D values are marked with asterisks (*P < 0.05,
988 **P < 0.01).

989 **Figure 3. KCS4 transcriptional regulation.**

990 **a.** Percentage of transcription factors (TF) classes predicted to bind *KCS4* genomic region according
991 to *in-silico* analysis of binding motives using DAP-seq database (O'Malley et al., 2016). **b.** Position of
992 the binding motifs (red rectangles) for the main TF classes. SNP positions in the promoter and in
993 coding region are marked with horizontal dash-lines and names are stated above gene model. **c.**
994 Electrophoretic Mobility Shift Assay (EMSA) showing the interaction between *HEAT SCHOCK FACTOR*
995 *A2* (*HSFA2*/AT2G26150) and the HSE binding motif (heat shock element: 5'- nnGnAnnTnCtn -3')
996 found within the SNP m11502 in *KCS4* promoter region; 1: labelled probe (5'-DY682-labelled
997 double-stranded oligonucleotide) only; 2: labelled probe plus HSFA2-GST protein; 3: labelled probe,
998 HSFA2-GST protein and 200× competitor DNA (unlabeled oligonucleotide containing HSE binding site
999 with the "C" allele at m11502), 4: well spill, 5: labelled probe plus free GST, 6: labelled probe,
1000 HSFA2-GST protein and 200× competitor DNA (unlabeled oligonucleotide containing HSE binding site
1001 with the "T" allele at m11502). **d.** *KCS4* expression for the complete GWAS population in CHD and HD
1002 conditions, grouped based on C or T allele at SNP m11502 in *KCS4* promoter region. Expression (fold
1003 change) was measured by qRT-PCR. Significant differences between CHD and HD transcript levels
1004 and between C and T alleles are shown (*t*-Test: * P < 0.05, ** P < 0.001, *** P < 0.0001, n.s. not
1005 significant). Data represent means ± SD (three technical replicates per sample). **e.** Average trait value
1006 (TAG 54:9 intensity, in $\log_2$) for the two major haplotypes determined by SNPs m11502 (promoter)
1007 and m11503 (coding region). CT and TC haplotypes encodes for *KCS4*(Met) and *KCS4*(Thr),
1008 respectively. Significant differences between haplotypes (*t*-Test: * P < 0.05). **f.** Heatmap of
1009 expression pattern for 16 members of the *KCS* multigene family clustered using complete-average
1010 method. Transcript abundance was measured 24 hours after treatment. Data was obtained
1011 previously (Caldana et al., 2011).

1012 **Figure 4. *KCS4* functional validation using allelic mutants in Arabidopsis and heterologous**
1013 **expression in *Nicotiana benthamiana*.**

1014 **a-f and j-k.** Average TAG profiles for wild-type (WT) accessions and allelic mutants carrying the
1015 *KCS4*(Met) allele (WT: red; mutant: pink) and the *KCS4*(Thr) allele (WT: blue; mutant: light-blue) in
1016 CHD and HD conditions. One-way ANOVA followed by Tukey post-hoc test was performed to detect
1017 differences between lines. Significant changes (* P < 0.05) are marked with asterisks (**a-f**) and with
1018 letters for 52 TAG series and 54 TAG series (**j-k**). **g-h-i.** Transient expression of At*KCS4* in *Nicotiana*
1019 *benthamiana* leaves. TAG levels for classes 52 (52:0 - 52:8), 54 (54:0 - 54:9) and 56 (56:0 - 56:0) for
1020 the control line (grey circles) and for the At*KCS4* line (black circles). Significant values (*t*-Test) are
1021 indicated (° P < 0.1, * P < 0.01, ** P < 0.001). **l-m**. Cuticular wax coverage for the same accessions
1022 grown in CHD or HD conditions. One-way ANOVA followed by Tukey post-hoc test was performed to
1023 detect differences between lines (* P < 0.05). n.s.: Not significantly different.

1024 **Figure 5. KCS4 activity and specificity in fatty-acid elongation and subcellular localization.**

1025 **a.** Very long chain fatty acid methyl ester (VLC-FAME) profiles (16 to 18 and 20 to 26, separate
1026 panels) of wild-type yeast transformed with an empty vector (orange) and *AtKCS4* (blue). **b.** VLC-
1027 FAME profiles (16 to 18 and 20 to 26, separate panels) of Δelo3 yeast mutant transformed with an
1028 empty vector (grey), *AtKCS4* (blue bars) and *AtKCS9* (green). **c.** VLC-FAME profiles (20 to 30) from

1029 leaves of *Nicotiana benthamiana* control plants (pink) or transformed with *AtKCS4* (blue). Significant
1030 differences are shown (*t*-Test: * P < 0.05, ** P < 0.001, *** P < 0.0001). **d.** Confocal images of *N.*
1031 *benthamiana* leaves transiently transformed with KCS4-YFP construct. The plasmid ER-GK CD3-955
1032 was used as an ER marker. Bars = 20 µm.

1033 **Figure 6. Biomass changes in *KCS4* wild-type and allelic mutants.**

1034 Rosettes fresh-weight measured in accessions carrying KCS4(Met) or KCS4(Thr), and their respctive
1035 allelic mutants. Plants were grown in control conditions (CHD) or subjected to heat and darkness for
1036 24 h (HD) and harvested after the treatment. Significant differences were obtained using *t*-Test (* P
1037 < 0.05).

1038 **Figure 7. *KCS4* haplotype correlation with climate and geographic parameters.**

1039 **a.** Heatmap of the mean values of lipid, climate and geographical variables for each of the four KCS4
1040 haplotypes (SNPs m11502-m11503). Results of the *KCS4* locus association as well as nonparametric
1041 ANOVA test (Kruskal-Wallis) are highlighted in the row-annotation bar. Cluster dominated by KCS4-
1042 associated TAGs is highlighted in red. **b.** Average haplotype value for: the principal component 1
1043 (PC1) obtained using the data in KCS4 cluster (red triangle in **a**), and the climate and geographical
1044 parameter showing significant Kruskal-Wallis in **a**. Significant differences are shown (pairwise
1045 Wilcoxon rank sum test: * P < 0.05, ** P < 0.01, *** P < 0.001).

1046 **Figure 8. Model summarizing the role of KCS4 on FA fate in Arabidopsis under carbon starvation.**

1047 Figure 8. Model summarizing the role of KCS4 on FA fate in Arabidopsis under carbon starvation.

1048 The lack of de novo FA synthesis under HD induces the degradation of galactolipids (MGDGs and
1049 DGDGs) from the chloroplasts' thylakoids. FFAs are then exported to the ER (FA-CoA pool), where
1050 KCS4 is localized and acts as a branch point in the fate of FAs, channeling saturated FAs to VLCFA
1051 elongation and tipping the balance to a higher polyunsaturated-to-saturated-FA ratio for puTAG
1052 accumulation (degrade grey arrows). Two *KCS4* alleles have differential capacity to allocate
1053 saturated FAs into the VLCFA pathway, with a major -more efficient- allele, *KCS4*(Met), and a minor -
1054 weaker allele, *KCS4*(Thr), exemplified here as red and blue and in different sizes, respectively.
1055 *KCS4*(Met) accessions efficiently channel the saturated FA from the Acyl-CoA pool to cuticular waxes,
1056 as they present higher accumulation of waxes than *KCS4*(Thr) accessions under stress. In addition,
1057 polyunsaturated TAG (puTAGs) levels are, therefore, higher in *KCS4*(Met) accessions comparing to
1058 *KCS4*(Thr) accessions, as the result of a rise in the ratio of polyunsaturated-to-saturated FAs going to
1059 TAG formation. VLCFA-CoA produced by the elongase complex (here exemplified by KCS4, KCR, HCD,
1060 ECR) go to increase the cuticular waxes under HD, and/or directly serve as a source of carbon
1061 undergoing degradation in the peroxisome; whereas puTAGs accumulated into lipid bodies first.
1062 FFAs are then released from lipid bodies by the action of lipases (SDP1/LIP1) and further imported to
1063 the peroxisome to undergo β-oxidation. Acronyms: ABCA9 (ATP-BINDING CASSETTE A9), ACBP-1
1064 (Acyl-CoA-binding protein 1); ACX (ACYL-COA OXIDASE); CoA (Coenzyme A); DGDG
1065 (Digalactosyldiacylglycerol); ECR (ENOYL-COA REDUCTASE); FAs (fatty acids); FAS (fatty-acid
1066 synthesis); FAX1 (FATTY ACID EXPORT1); FFAs (free fatty acids); HCD (β-HYDROXYACYL-COA
1067 DEHYDRATASE); IAA (indol-3-acetic-acid), IBA (indol-3-butyric-acid), KCR (β-KETOACYL REDUCTASE);
1068 KCS4(Met/Thr) (3-KETOACYL-COA SYNTHASE4); LACS (LONG-CHAIN ACYL-COA SYNTHETASE);
1069 SDP1/LIP1 (LIPASE 1); MGDG (Monogalactosyldiacylglycerol); PXA1 (PEROXISOMAL ABC-
1070 TRANSPORTER 1); pu (polyunsaturated), sa (saturated), TAGs (Triacylglycerols); VLCFA (Very-Long-
1071 Chain-Fatty-Acid).

1072 **Supplementary Figure 1. Phenotypic analyses for the Arabidopsis GWAS panel under control and**
1073 **stress conditions**

1074 **a.** Principal component analysis (PCA) of lipid data for control condition (CHD, blue circles) and heat
1075 and dark condition (HD, red diamonds). Each dot represents a single accession. PC1 and PC2 explain
1076 together 67.5% of the variance observed in the phenotypic data. **b.** Heatmap of differential
1077 accumulation of lipid levels in stress (HD) and control (CHD) conditions. Lipids are clustered by lipid
1078 class. For each accession, the lipid level in HD was subtracted by the lipid level in CHD. Positive
1079 values (red) denote higher levels of lipid in HD condition, negative values (blue) denote higher level
1080 of lipid in CHD condition (scale= log2 (Δ of lipid intensity). **c**. Correlation pattern among lipid levels in
1081 both conditions: CHD (blue rectangle) and HD (red rectangle). Pairwise Pearson correlations (r^2) were
1082 calculated between each lipid across all accessions. Lipids are highlighted with colors according to
1083 lipid classes (references included in the figure). **d.** Distribution of fold change of lipid feature levels
1084 across all accessions. For each lipid feature, fold change was calculated by dividing the average of
1085 the five highest intensities measured in a given 5 accessions by the average of the five lowest
1086 intensities. Y-axes is expressed in percentage of (A) all metabolite features (N=109 in stress, N=98 in
1087 CHD), (B) only TAG features (N=41 in stress, N= 30 in CHD) and (C) other lipid features (N=68 in stress
1088 and N=68 in CHD).

1089 **Supplementary Figure 2. Results from GWAS analysis on lipidomic data from plants grown under**
1090 **extended darkness (3D, 6D) conditions.**

1091 **a.** Chromosome scheme and Quantitative Trait Loci (QTL) identified in the two experimental
1092 conditions (3D = circles, 6D = squares), for different lipid classes: phosphatidylethanolamine (PE),
1093 phosphatidylcholine (lecithin) (PC), monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol
1094 (DGDG), diacylglycerol (DAG) and triacylglycerol (TAG). Color/shape references are included in the
1095 figure. Gene IDs are included only for QTL co-localizing for two or more lipid species with LOD score
1096 > 5.3. **b.** Manhattan plots obtained for TAG 54:9 in 3D (left) and in 6D (right). **c.** Average trait value
1097 (intensity of TAG 54:9, $\log_2$ scale) for the different *KCS4* haplotypes using SNPs m11502, m11503,
1098 m11504 and m11505. **e.** TAG 54:9 average value for accessions carrying the KCS4(Met) and
1099 KCS4(Thr) alleles.

1100 **Supplementary Figure 3. Quantification of starvation markers in KCS4 wild-type accessions and**
1101 **allelic mutants.**

1102 Transcripts were measured in 5 accessions, 3 carrying *KCS4*(Met) and two carrying *KCS4*(Thr), and
1103 five allelic mutants from those 5 accessions. Plants were grown in control conditions (CHD) or
1104 subjected to heat and darkness for 24 h (HD) and harvested both group at the same time point.
1105 Three biological replicates were used for each line, results from separate accessions were pooled
1106 according to the allele carried. Fold change of expression was obtained using the $2^{-(\Delta\Delta CT)}$ method
1107 using GAPDH3 (At1g13440) as housekeeping gene. Statistical significance was performed with
1108 ANOVA followed by Fisher LSD. Means with a common letter are not significantly different (P > 0.05).

1109 **Supplementary Figure 4. KCS4 enzymatic activity**

1110 **a.** Very long chain fatty acid methyl ester (VLC-FAME) profiles (22 to 26) of Δelo3 yeast mutant
1111 transformed with wild-type At*KCS4*(Met)/At*KCS4*(Thr) and allelic mutans *kcs4*(Met)/*kcs4*(Thr).
1112 References for colors are in the figure. Significant differences between alleles obtained with *t*-Test:
1113 ° P < 0.1, ** P < 0.01, *** P < 0.001. **b.** VLC-FAME profiles (22 to 26) at different times of harves (48,
1114 72 and 96 h) for At*KCS4*(Met) and At*KCS4*(Thr) wild-type alleles. Significant differences were
1115 obtained with ANOVA followed by Tukey test (P < 0.05).

1116 **Supplementary Figure 5. Quantification of *KCS4*, *KCS5*, *KCS9* and *KCS12* transcripts in wild-type**
1117 ***KCS4*(Met)/*KCS4*(Thr) accessions and kcs4(Met)/kcs4(Thr) allelic mutants.**

1118 Transcripts were measured in 5 accessions, 3 carrying *KCS4*(Met) and two carrying *KCS4*(Thr), and
1119 five allelic mutants from those 5 accessions. Plants were grown in control conditions (CHD) or
1120 subjected to heat and darkness for 24 h (HD) and harvested both group at the same time point.
1121 Three biological replicates were used for each line, results from separate accessions were pooled
1122 according to the allele carried. Fold change of expression was obtained using the $2^{-(\Delta\Delta CT)}$ method
1123 using GAPDH3 (At1g13440) as housekeeping gene. Statistical significance was performed with
1124 ANOVA followed by Fisher LSD. Means with a common letter are not significantly different ($P > 0.05$).

1125 **Supplementary Figure 6. Phylogenetic signal of the lipidome**

1126 **a.** Distribution of the phylogenetic signal calculated for all measured lipids in HD and CHD conditions.
1127 Only lipids exhibiting significant phylogenetic signal are labelled. Lipids significant in CHD and HD are
1128 marked with blue and red dots, respectively. Compounds significant in both conditions ($p<0.05$) have
1129 overlapping red and blue markers. **b.** K parameter distribution obtained for the TAG profiles in CHD
1130 (control) and HD (treatment) condition according to Blomberg et al. (2003).

1131 **Supplementary Figure 7. Prediction of climate parameters from lipidomic profiles**

1132 Regression coefficients of selected predictive lipidomic features towards 19 climate variables and 3
1133 geographical coordinates (latitude, longitude and altitude) of the geographical origin of analyzed
1134 accessions. Statistical evaluation of the fitted models is provided on the right-hand panel. The model
1135 fit significance (red asterisks) is estimated by the correlation coefficient between the original and
1136 model-fitted variable values. The significance of Cross-Validation test (black asterisks) represents the
1137 significance of the correlation coefficient between the predicted and real values of the test sample.
1138 The cross-validation has been performed 100 times with the 3:1 proportion of the training to test
1139 sample set size.

1140 **Supplementary File 1.**

1141 Alignment used for the generation of the SNP matrix used in TASSEL to analyze more polymorphisms
1142 in the QTL associated with *KCS4*. Sequences were obtained from the Arabidopsis 1001 genome
1143 project (http://signal.salk.edu/atg1001/3.0/gebrowser.php, 162 accessions). TAIR10 positions of the
1144 extract are chromosome 1: 6689119 - 6769118. Alignment was performed using MAFFT (Katoh et al.,
1145 2019) and manually curated when needed.

1146 **Supplementary File 2.**

1147 Alignment of KCS4 sequences from wild type and knockout accessions used in this study.

1148 **Supplementary Dataset 1.**

1149 Phenotypic data used for GWAS and collected in: (1) control conditions for the heat and dark stress
1150 treatment (CHD), (2) heat and dark stress (HD), (3) control conditions for the dark treatment (CD), (4)
1151 three days extended darkness (3D) and (5) six days extended darkness. Each phenotype is in
1152 different excel sheets. Plants were grown and harvested as described in Materials and Methods.
1153 Lipid classes detected are DAG: dyacylglicerols, DGDG: digalactosyldiacylglycerol, MGDG:
1154 monogalactosyldiacylglycerol, PC: phosphatidylcholine (lecithin), PE: phosphatidylethanolamine and
1155 TAG: triacylglycerol. For each species the number of carbons and the degree of bond-saturation is
1156 stated. Phenotypic value is the logarithm in the second base of intensity obtained in LC-MS.

1157 **Supplementary Dataset 2.**

1158 Table with GWAS results obtained from GAPIT for each phenotype presented in Supp. Table 1. All
1159 SNPs with a LOD score > 5.3 for all traits obtained in the same condition are summarized in the same
1160 table. Each sheet presents the results for each condition (see Supp. Table 1). MAF: minor allele
1161 frequency.

1162 **Supplementary Dataset 3.**

1163 Full list of GWAS results summarized in number of QTL (quatitative trait loci) and genes detected per
1164 condition, for all lipid species phenotyped (LOD score >5.3). Co-localization between traits in the
1165 same conditions are stated in "Peak.ID" columns. Co-localizing genes between CHD and HD
1166 condition are also mentioned. Putative lipid functions for genes related to lipid biosynthesis
1167 pathways are included in the last column. KCS4 association is highlighted in red. Each sheet compiles
1168 results for all traits phenotyped in one condition (see Supp. Table 1)

1169 **Supplementary Dataset 4.**

1170 Data used for TASSEL analysis: Genotypic matrix, generated from the alignment in Supp. File 1;
1171 principal components and kinship matrix used to control for population structure and confounding
1172 effects. Phenotypic matrixes for HD and 3D are included in Supplementary Dataset 1.

1173 **Supplementary Dataset 5.**

1174 GWAS results for HD and 3D data (see Supp. Table 1 for phenotype matrix and Supp. Table 4 for
1175 genotype matrix) using complete sequence information for 80 kb around KCS4. SNPs super-passing
1176 Bonferroni-corrected p-values (LOD > 4.38) are marked in bold. Markers ID in Tassel are referenced
1177 to positions in Supp. File 1, marker ID in GAPIT and TAIR10 position.

1178 **Supplementary Dataset 6.**

1179 All lipids measured in Arabidopsis wild-type accessions and *kcs4* allelic mutants. Accession ID
1180 (common name), Type (wild type or mutant), KCS4_allele (Methionine or Threonine), Condition
1181 (control or HD) are included in different columns. The table presents trait values for each biological
1182 replicate for accessions separately. Values in Figure 4 were obtained by averaging accessions
1183 encoding for a Methionine or Threonine, respectively for each type (WT and KO).

1184 **Supplementary Dataset 7.**

1185 Primers and oligonucleotide sequences used in this study.

1186

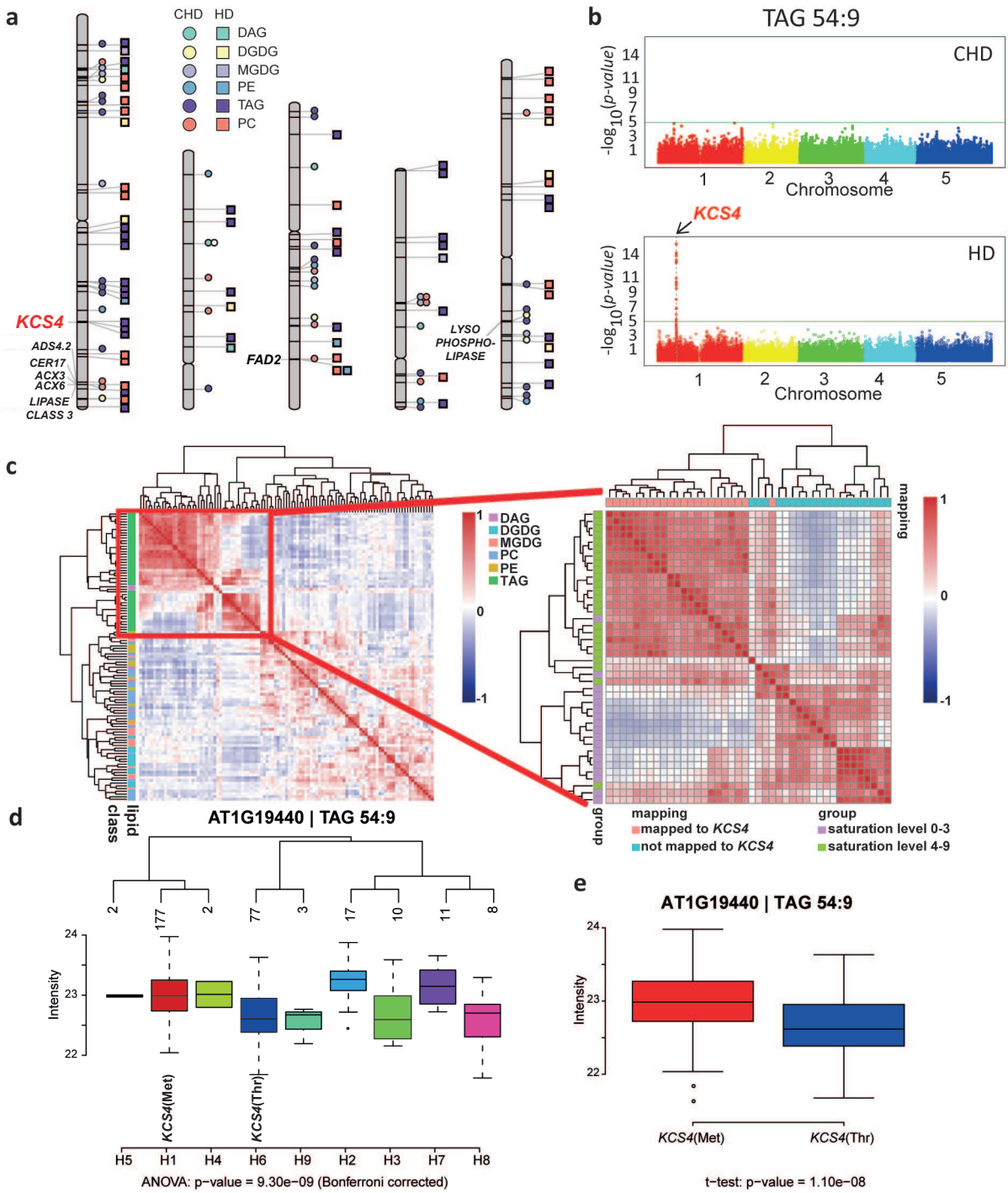

Figure 1. Results from GWAS on lipidomic data from plants grown under control (CHD) and heat and darkness (HD) conditions.

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 names 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 (see reference at right) and clustered using Ward´s distance on pairwise dissimilarity. A zoom-in of the TAG cluster shows the grouping pattern of TAGs according to number of unsaturated bonds and to the GWAS result (left) **d.** Average trait value (intensity of TAG 54:9, $\log_2$ scale) for the different *KCS4* haplotypes using SNPs m11502, m11503, m11504 and m11505. The two major haplotypes have the alternative alleles for m11503 (Met or Thr) **e.** TAG 54:9 average value for accessions carrying the *KCS4*(Met) and *KCS4*(Thr) alleles.

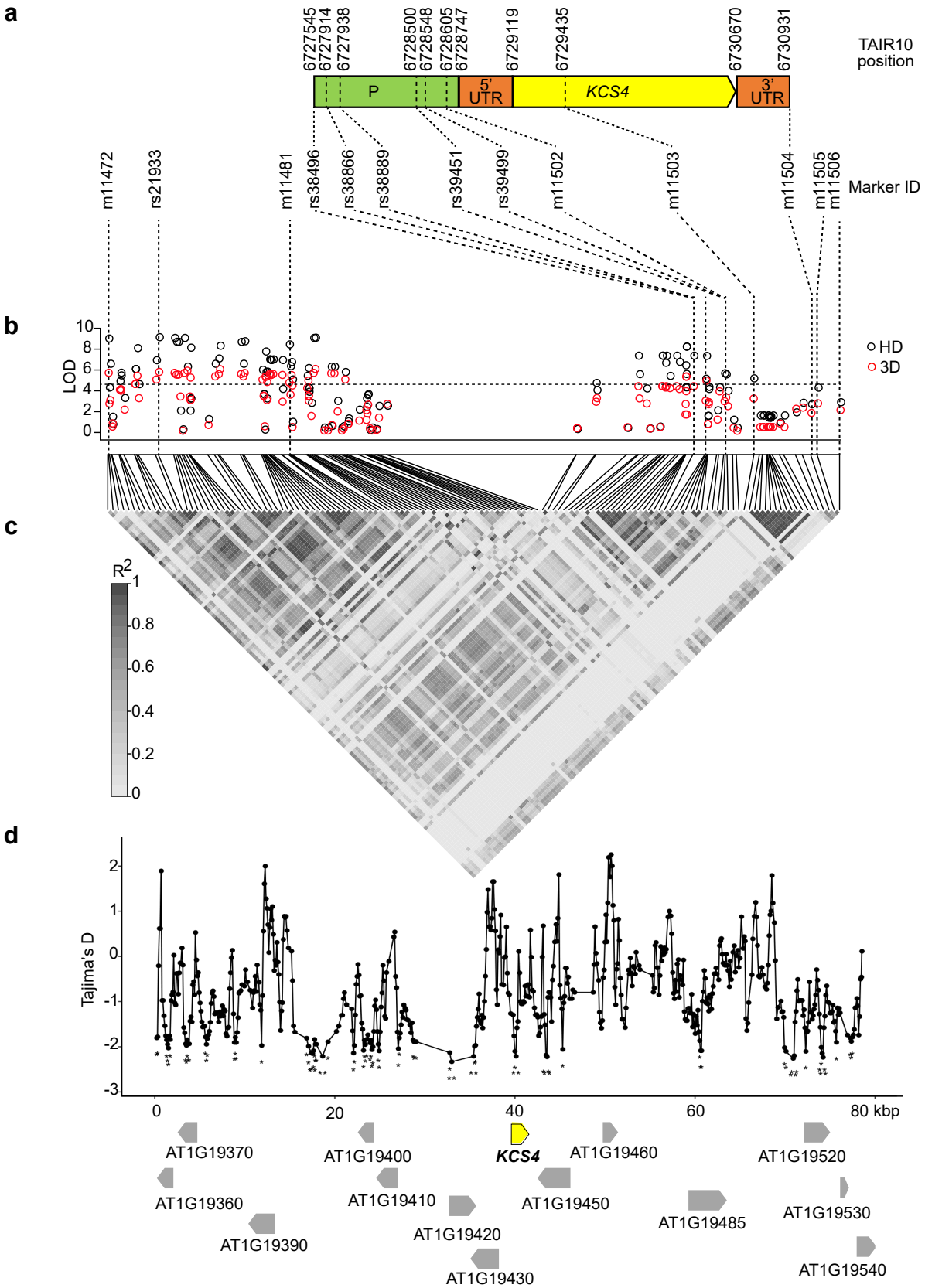

Figure 2. Dissection of *KCS4* alleles involved in TAG variation, pattern of linkage disequilibrium and signatures of selection.

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 bellow (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).

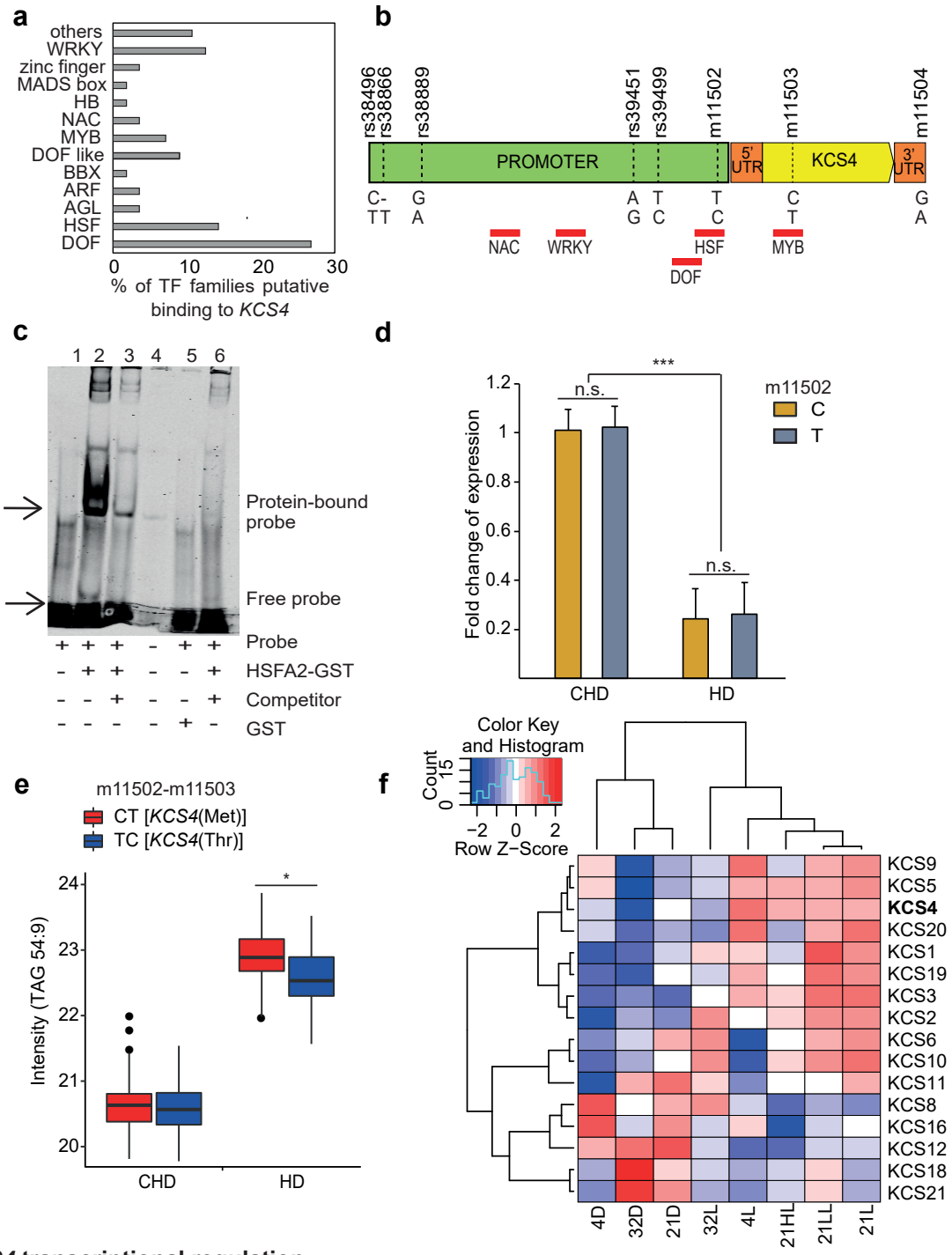

Figure 3. *KCS4* transcriptional regulation.

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 SCHOCK 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 (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 ± 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 4

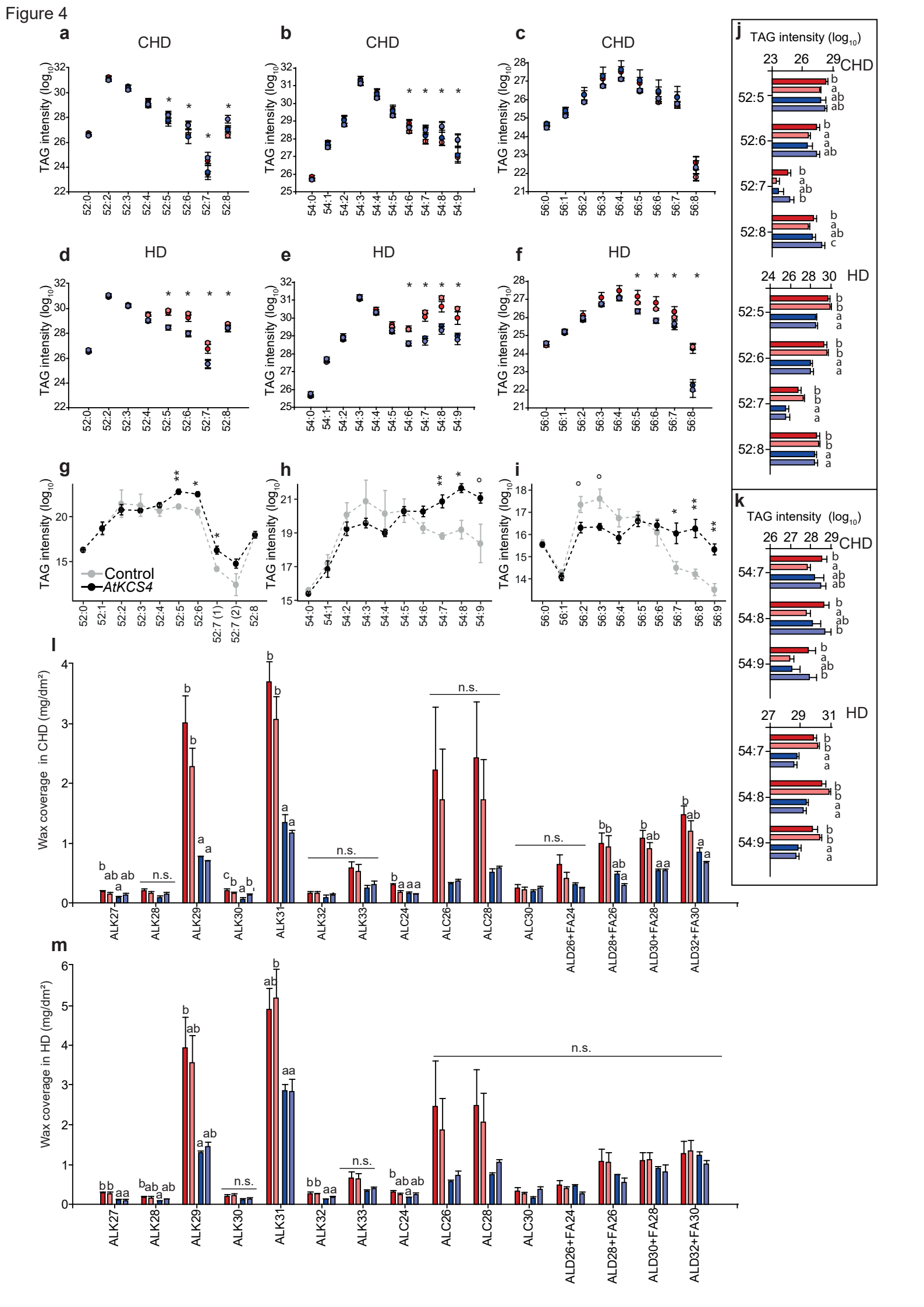
a
CHD
b
CHD
c
CHD
d
HD
e
HD
f
HD
TAG intensity (log10)
52:0
52:2
52:3
52:4
52:5
52:6
52:7
52:8
54:0
54:1
54:2
54:3
54:4
54:5
54:6
54:7
54:8
54:9
56:0
56:1
56:2
56:3
56:4
56:5
56:6
56:7
56:8
g
h
i
Control
AtKCS4
52:7 (1)
52:7 (2)
56:9
j
TAG intensity (log10)
CHD
HD
k
TAG intensity (log10)
CHD
HD
l
Wax coverage in CHD (mg/dm²)
m
Wax coverage in HD (mg/dm²)
ALK27
ALK28
ALK29
ALK30
ALK31
ALK32
ALK33
ALC24
ALC26
ALC28
ALC30
ALD26+FA24
ALD28+FA26
ALD30+FA28
ALD32+FA30
n.s.

Figure 4. *KCS4* functional validation using allelic mutants in Arabidopsis and heterologous expression in *Nicotiana benthamiana*.

a-f. Average TAG profiles for wild-type (WT) accessions and allelic mutants carrying the *KCS4*(Met) allele (WT: red; mutant: pink) and the *KCS4*(Thr) allele (WT: blue; mutant: light-blue) in CHD and HD conditions. One-way ANOVA followed by Tukey post-hoc test was performed to detect differences between lines. Significant changes (* $P < 0.05$) are marked with asterisks **(a-f)** and in detail for 52 TAG series and 54 TAG series **(j-k)**. **g-h-i.** 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 - 56:0) for the control line (grey circles) and for the At*KCS4*(Met) line (black circles). Significant values (*t*-Test) are indicated (° $P < 0.1$, * $P < 0.01$, ** $P < 0.001$). **l-m.** Cuticular wax coverage for the same accessions grown in CHD or HD conditions. One-way ANOVA followed by Tukey post-hoc test was performed to detect differences between lines (* $P < 0.05$). n.s.: Not significantly different.

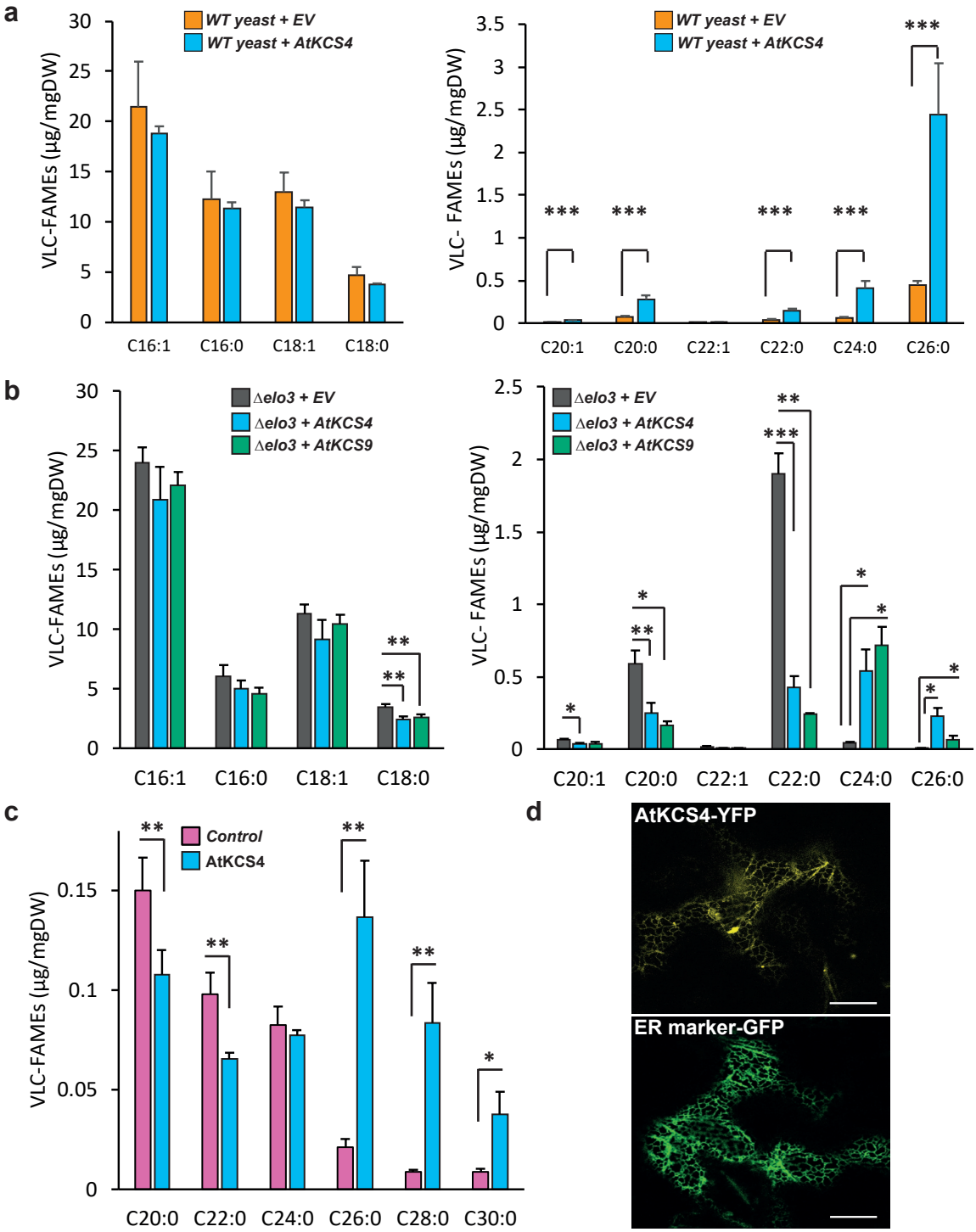

Figure 5. *KCS4* activity and specificity in fatty-acid elongation and subcellular localization. **a.** Very long chain fatty acid methyl ester (VLC-FAME) profiles (16 to 18 and 20 to 26, separate panels) of wild-type yeast transformed with an empty vector (orange) and *AtKCS4* (blue). **b.** VLC-FAME profiles (16 to 18 and 20 to 26, separate panels) of Δelo3 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.

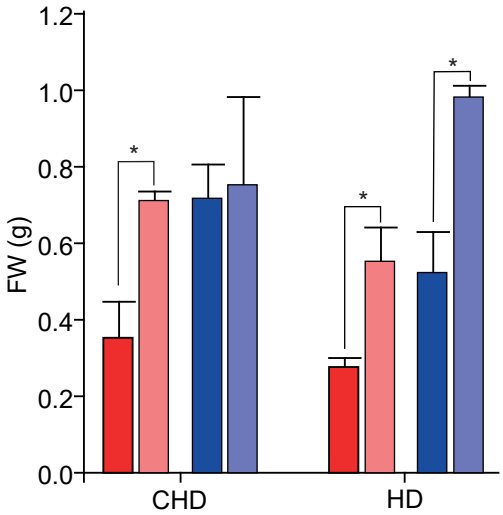

Figure 6. Biomass changes in KCS4 wild-type accessions and allelic mutants.
Rosettes fresh-weight measured in accessions carrying *KCS4*(Met) or *KCS4*(Thr), and their respctive allelic mutants. Plants were grown in control conditions (CHD) or subjected to heat and darkness for 24 h (HD) and harvested after the treatment. Significant differences were obtained using *t*-Test (* $P < 0.05$).

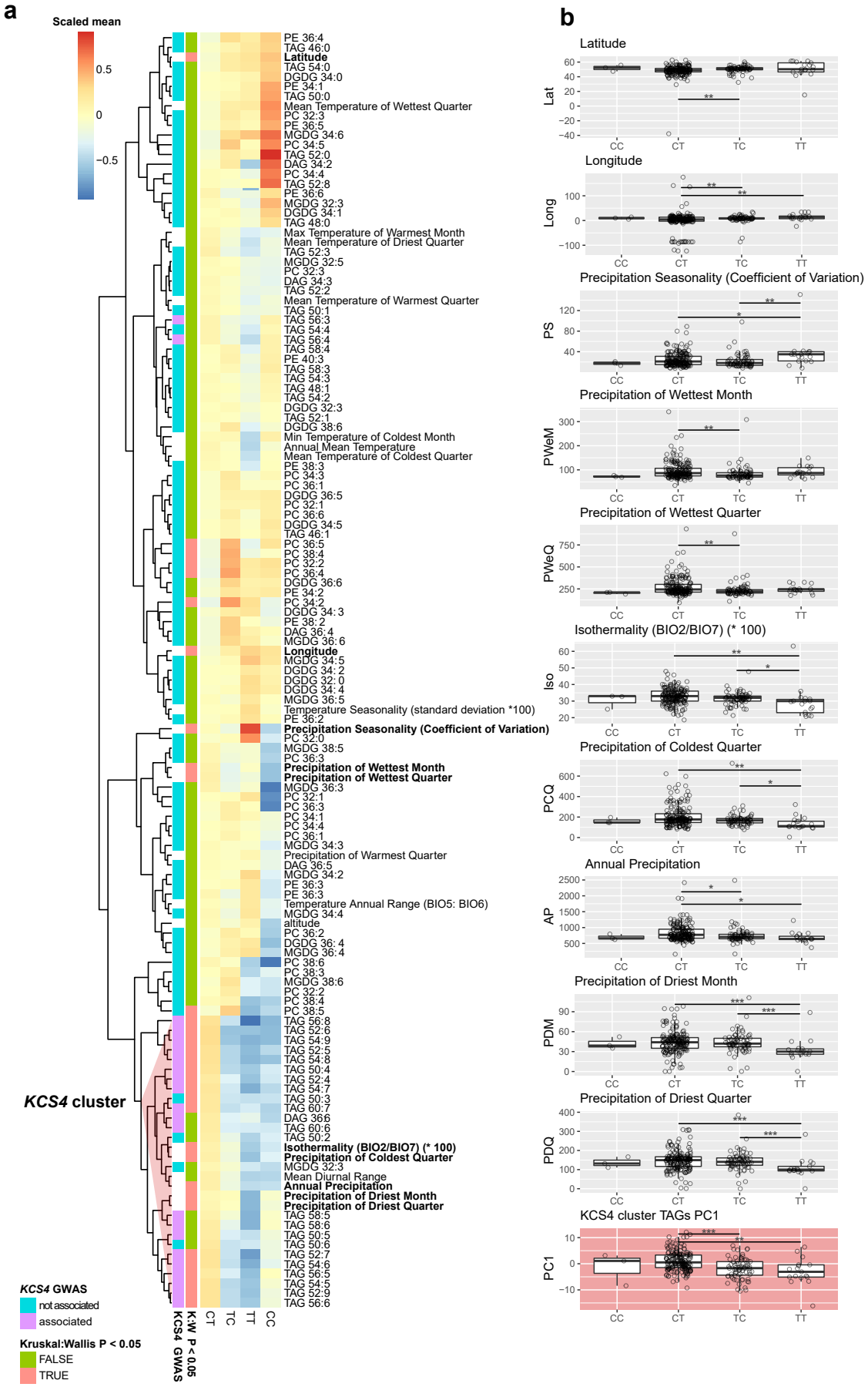

Figure 7. *KCS4* haplotype correlation with climate and geographic parameters.

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).

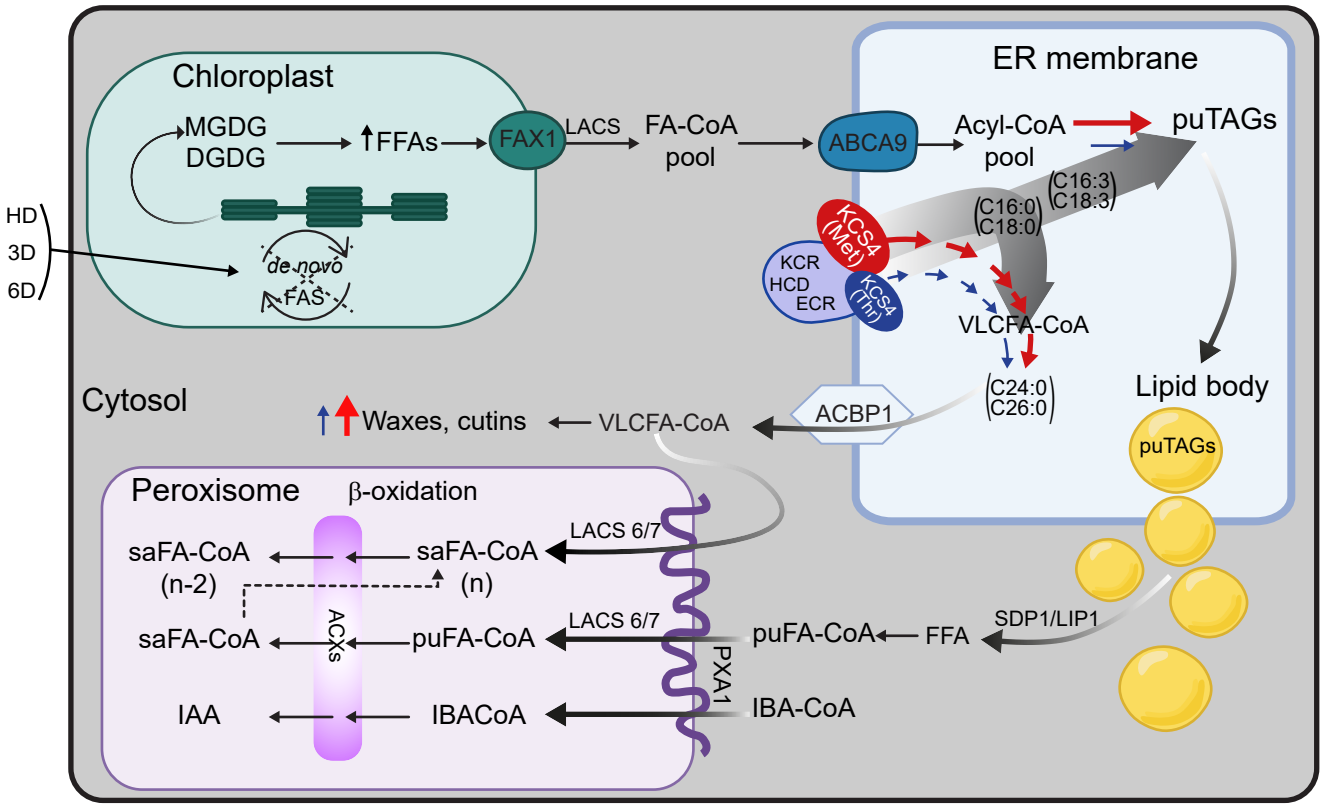

Figure 8. Model summarizing the role of KCS4 on FA fate in Arabidopsis under carbon starvation. The lack of *de novo* FA synthesis under HD induces the degradation of galactolipids (MGDGs and DGDGs) from the chloroplasts' thylakoids. FFAs are then exported to the ER (FA-CoA pool), where KCS4 is localized and acts as a branch point in the fate of FAs, channeling saturated FAs to VLCFA elongation and tipping the balance to a higher polyunsaturated-to-saturated-FA ratio for puTAG accumulation (degrade grey arrows). Two KCS4 alleles have differential capacity to allocate saturated FAs into the VLCFA pathway, with a major -more efficient- allele, *KCS4*(Met), and a minor -weaker- allele, *KCS4*(Thr), exemplified here as red and blue and in different sizes, respectively. *KCS4*(Met) accessions efficiently channel the saturated FA from the Acyl-CoA pool to cuticular waxes, as they present higher accumulation of waxes than KCS4(Thr) accessions under stress. In addition, polyunsaturated TAG (puTAGs) levels are, therefore, higher in *KCS4*(Met) accessions comparing to *KCS4*(Thr) accessions, as the result of a rise in the ratio of polyunsaturated-to-saturated FAs going to TAG formation. VLCFA-CoA produced by the elongase complex (here exemplified by KCS4, KCR, HCD, ECR) go to increase the cuticular waxes under HD, and/or directly serve as a source of carbon undergoing degradation in the peroxisome; whereas puTAGs accumulated into lipid bodies first. FFAs are then released from lipid bodies by the action of lipases (SDP1/LIP1) and further imported to the peroxisome to undergo β-oxidation. Acronyms: ABCA9 (ATP-BINDING CASSETTE A9), ACBP-1 (Acyl-CoA-binding protein 1); ACX (ACYL-COA OXIDASE); CoA (Coenzyme A); DGDG (Digalactosyldiacylglycerol); ECR (ENOYL-COA REDUCTASE); FAs (fatty acids); FAS (fatty-acid synthesis); FAX1 (FATTY ACID EXPORT1); FFAs (free fatty acids); HCD (β-HYDROXYACYL-COA DEHYDRATASE); 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).