the plant journal

SEB Society for Experimental Biology

The Plant Journal (2018) **93,** 1102–1115 doi: 10.1111/tpj.13833

Unraveling lipid metabolism in maize with time-resolved multi-omics data

Francisco de Abreu e Lima[1], Kun Li[2], Weiwei Wen[1], Jianbing Yan[2], Zoran Nikoloski[1,3,*], Lothar Willmitzer[1] and Yariv Brotman[4]

[1]*Max-Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany,*

[2]*National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Shizishan Lu 1, 430070, Hongshan, Wuhan, China,*

[3]*Bioinformatics Group, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany, and*

[4]*Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel*

Received 16 August 2017; revised 3 January 2018; accepted 4 January 2018; published online 31 January 2018.

*For correspondence (e-mail nikoloski@mpimp-golm.mpg.de).

Francisco de Abreu e Lima and Kun Li are the authors who contributed equally to this study.

SUMMARY

Maize is the cereal crop with the highest production worldwide, and its oil is a key energy resource. Improving the quantity and quality of maize oil requires a better understanding of lipid metabolism. To predict the function of maize genes involved in lipid biosynthesis, we assembled transcriptomic and lipidomic data sets from leaves of B73 and the high-oil line By804 in two distinct time-series experiments. The integrative analysis based on high-dimensional regularized regression yielded lipid–transcript associations indirectly validated by Gene Ontology and promoter motif enrichment analyses. The co-localization of lipid-transcript associations using the genetic mapping of lipid traits in leaves and seedlings of a B73 × By804 recombinant inbred line population uncovered 323 genes involved in the metabolism of phospholipids, galactolipids, sulfolipids and glycerolipids. The resulting association network further supported the involvement of 50 gene candidates in modulating levels of representatives from multiple acyl-lipid classes. Therefore, the proposed approach provides high-confidence candidates for experimental testing in maize and model plant species.

Keywords: *Zea mays*, lipid metabolism, omics, GFLASSO, QTL.

INTRODUCTION

One of the many determinants of kernel quality in maize (*Zea mays* L.) is its oil content and composition (Watson, 1987). Maize oil is highly calorific, rich in polyunsaturated fatty acids (e.g. linoleic acid) and is largely composed of triacylglycerols (TAGs), high-energy lipids that nourish germinating seedlings (Baud and Lepiniec, 2010). Since maize oil is widely used as a food, feed and biofuel resource, obtaining a full characterization of the genetic underpinnings of lipid metabolism in maize is fundamental for improving oil quality, quantity and composition.

In recent years progress has been made towards the elucidation of the acyl-lipid metabolism in *Arabidopsis thaliana* (Graham and Eastmond, 2002; Beisson *et al.*, 2003; Li-Beisson *et al.*, 2013), accompanied by technological advances that facilitate the identification and quantification of hundreds of lipid species from a single sample (Wenk, 2005, 2010; Hummel *et al.*, 2011). *De novo* biosynthesis of fatty acids (FAs) takes place in the plastid with successive incorporations of acetyl-coenzyme A (CoA). The FAs can then be subjected to elongation, desaturation and eventual export from the plastid, and can ultimately give rise to the distinct acyl-lipid species, including phosphatidylinositols (PIs), phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), phosphatidylglycerols (PGs), monogalactosyldiacylglycerols (MGDGs), digalactosyldiacylglycerols (DGDGs), sulfoquinovosyldiacylglycerols (SQDGs), diacylglycerols (DAGs) and TAGs (Li-Beisson *et al.*, 2013). However, the large number of intermediates and the interaction between lipid pathways render the characterization of lipid metabolism quite challenging.

The advent of powerful statistical methods for genome-wide association studies (GWASs) and linkage-based mapping of quantitative trait loci (QTLs), combined with the available diversity of maize germplasm, allow the genetic architecture of complex traits in maize to be uncovered. With the prospect of enhancing kernel quality, many such

efforts have provided insights into the genetic blueprint underlying lipid metabolism in maize (Laurie *et al.*, 2004; Yang *et al.*, 2010; Cook *et al.*, 2012). However, kernel oil accumulation is a polygenic trait, as substantiated by linkage mapping and the repeated selection of Illinois high-oil (IHO) lines over 100 generations to achieve a 20% increase in oil concentration (Laurie *et al.*, 2004; Moose *et al.*, 2004). This renders the characterization of the genetic determinants of lipid metabolism a daunting task.

One way to identify genes involved in lipid metabolism is through co-expression analysis, since this method has been found to be effective in deciphering gene functions based on the 'guilt-by-association' paradigm (Aoki *et al.*, 2007; Saito *et al.*, 2008). Co-expression studies have already investigated oil biosynthesis in Arabidopsis (Mentzen *et al.*, 2008), *Physaria fendleri* (Troncoso-Ponce *et al.*, 2011), oil palm (Guerin *et al.*, 2016), the green alga *Chlamydomonas reinhardtii* (Gargouri *et al.*, 2015) and in a comparative analysis using four oilseed species (Horn *et al.*, 2016). However, such studies depend on existing knowledge of the function of genes, and it remains partial even for the most extensively studied model plant *A. thaliana* (Rhee and Mutwil, 2014). Therefore, the consideration of additional data profiles, such as those of metabolites (and lipids), may help decipher the role of candidate genes (Urbanczyk-Wochniak *et al.*, 2003; Rischer *et al.*, 2006; Bylesjö *et al.*, 2007; Mintz-Oron *et al.*, 2008; Szymanski *et al.*, 2014; Cavill *et al.*, 2016).

If conserved, knowledge about lipid pathways can be transferred between plant species. The oil content in the seeds of *A. thaliana* mutants lacking *WRINKLED1* (*WRI1*), a key master regulator of FA biosynthesis in higher plants, was restored with the respective orthologs from rapeseed (Liu *et al.*, 2010), *Camelina sativa* L. (An *et al.*, 2017), maize (Pouvreau *et al.*, 2011), oil palm (Ma *et al.*, 2013) and coconut (Sun *et al.*, 2017). In maize, *WRI1a* and *WRI1b* encode transcription factors of the APETALA2/ethylene-responsive element-binding (AP2/EREB) family that increase FA content in seeds and are co-expressed with glycolytic enzymes and genes involved in FA and TAG biosynthesis in the kernel (Pouvreau *et al.*, 2011). In oil palm, too, a *WRI1* gene was co-expressed with glycolytic enzymes and genes involved in FA and TAG biosynthesis in the fruit mesocarp (Guerin *et al.*, 2016). With respect to the downstream steps in oil biosynthesis, the over-expression of a type-one acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) from Arabidopsis greatly increased the TAG content in the leaves of tobacco plants (Bouvier-Navé *et al.*, 2000; Andrianov *et al.*, 2010). Similarly, a type-two DGAT from *C. reinhardtii* was shown to increase TAG concentrations in Arabidopsis (Sanjaya *et al.*, 2013). This is all the more spectacular given the divergence of Arabidopsis from the far-related monocots oil palm and coconut (120 million years ago) and from *C. reinhardtii* (1 billion years ago), and the conservation of oil biosynthetic pathways in seed and non-seed tissues. Altogether, this suggests the existence of highly conserved lipid biosynthetic pathways in higher plants.

However, determination of all the genes underlying lipid metabolism is complicated by targeted expression of multiple genes in specific tissues and organs (Ohlrogge *et al.*, 1991). Lipid composition further depends on the functional activities of the downstream protein products and their interrelation in the larger genome-wide metabolic network (Haslam *et al.*, 2016). Nevertheless, identification of a candidate gene underlying lipid metabolism based on data integrative studies in a single organ provides a stepping stone in revealing the role of the gene in other tissue types. In addition, given the tight connection between leaf physiology and kernel yield in maize (Cañas *et al.*, 2017), it is expected that an understanding of lipid metabolism in leaves and how it contributes to carbon partitioning will contribute to revealing their role in other tissues (e.g. kernel).

The combination of quantitative genetics and co-expression studies provides a powerful framework for functional characterization of genes (Mackay *et al.*, 2009). For instance, this strategy has been used to identify genes controlling glucosinolates and anti-herbivore defense in Arabidopsis (Chan *et al.*, 2011), micronutrients in chickpea (Upadhyaya *et al.*, 2016), leaf morphology in oilseed rape (Jian *et al.*, 2017) and phosphorus stress tolerance in soybean (Zhang *et al.*, 2017). In this class of approaches, one may utilize a single data set to determine loci associated with a trait of interest and then narrow down the search to those whose respective gene transcripts are correlated with the trait (Chen *et al.*, 2016); moreover, streamlining the list can be carried out by investigation of gene co-expression based on publically available data sets (Serin *et al.*, 2016). However, these approaches focus on analysis of a single trait at a time and neglect the inherent dependence between multiple traits in attempting to identify the associated loci.

With respect to lipid metabolism in maize, a recent effort based on the integration of the findings from (i) a GWAS of fatty acids using 368 maize inbred lines, (ii) QTL mapping of oil concentration and composition in the maize kernel of maize, using a B73 $\times$ By804 recombinant inbred ine (RIL) population, and (iii) a co-expression network based on the expression landscape of seeds obtained 15 days after pollination from the population indicated in (i), identified 43 high-confidence candidates involved in oil biosynthesis in maize (Li *et al.*, 2013). A similar methodology was adopted for testing the association with secondary metabolites in maize, providing functional prediction for 238 candidate genes, two of which were experimentally validated (Wen *et al.*, 2014).

Here we report a multi-omics, two-step integrative analysis of data sets comprising the circadian time-resolved

gene expression and lipid profiles of B73 and the high-oil line By804. Our unique data set incorporates molecular read-outs from two distinct field trials, providing greater confidence in our findings (based on the combination of two data sets), and relies on well-annotated acyl-lipid compounds. Unlike existing approaches, we do not consider individual lipid species but harness the presence of correlated traits to identify loci simultaneously associated with such traits. To this end, we use the graph-guided fused least absolute shrinkage and selection operator (GFLASSO; https://github.com/krisrs1128/gflasso) (Kim *et al.*, 2009). We then refine the findings with results from QTL mapping of lipid traits from a B73 × By804 RIL population. This step narrows down the list of genes putatively involved in the metabolism of the investigated lipid traits. We show that such a combination of modeling and QTL studies, which takes into consideration the correlation structure of multiple traits, is powerful in identifying high-confidence genes involved in lipid metabolism. This is supported by Gene Ontology (GO) and promoter motif enrichment analyses to indirectly test the involvement of our list of candidate genes in the metabolism of acyl-lipids. Finally, we discuss the predicted functions of the high-confidence candidates in the context of lipid metabolism.

RESULTS

Experimental design facilitating the application of two-step, multi-omics analysis

Our study relied on the integration of data from two time-series experiments. The time-series experiments consisted of monitoring gene expression and lipid levels in leaves of B73 and By804 (a high-oil line) subjected to a 12-h light followed by a 12-h dark photoperiod (12L12D). These were determined over 16 and 12 time points in two distinct trials, designated as 'winter' and 'summer' trials, respectively (Figure 1a). The QTL mapping experiment consisted of determining lipid levels in leaf and seedling samples in a B73 × By804 RIL population (Figure 1b). Having access to two data sets that capture seasonal variation, we expect to detect prevalent lipid–transcript associations, i.e. associations that are not present in a single season only.

To uncover lipid–transcript associations, we opted to analyze the data on all lipid species simultaneously, thus capturing the dependence inherent in metabolism. Therefore, rather than predicting the transcripts which best explain a single lipid profile, we aimed to identify transcripts which jointly explain latent variables representing distinct lipid classes. To this end (i) we employed the GFLASSO modeling approach and selected the best predictive genes then (ii) identified the genes in the QTL from acyl-lipids measured in the QTL mapping experiment (Figure 1c). We then (iii) determined the intersection of genes obtained from (i) and (ii), and finally (iv) investigated the network of lipid–gene interactions containing the genes obtained from step (iii) (Figure 1d). Therefore, our approach provides a novel methodological integration of the data to identify genes controlling metabolism of acyl-lipids.

Dynamics of lipids and gene expression

A total of 18 028 transcripts and 132 acyl-lipid species were analyzed in the time-series experiments (Data S1 in the online Supporting Information). The lipid data set included representatives from nine classes of acyl-lipids: PIs, PCs, PEs, PGs, MGDGs, DGDGs, SQDGs, DAGs and TAGs. Prior to modeling, we investigated the main sources of variation in the transcriptome and lipidome in the time-series experiments and the degree of similarity of the dynamics in the two trials (i.e. 'winter' and 'summer'). We first conducted separate principal component analyses (PCAs) of the lipid and expression data sets from the two trials. We found that for the lipidome PC1 separates the two genotypes (Figure 2a,b), while for the transcriptome PC1 separates early and late time points, reflecting circadian changes (Figure 2c,d).

Next, we computed the Pearson correlation coefficient (r) of the levels from each transcript or acyl-lipid between the two trials using the 12 overlapping time points (cross-trial correlation) and between B73 and By804 (cross-genotype correlation). Both analyses pointed to median Pearson correlation coefficients of $r = .62$ and $r = .58$ over individual transcripts across trials and across genotypes, respectively. Lipid levels, on the other hand, were comparatively more affected by the two factors, resulting in median Pearson correlation coefficients of $r = .34$ and $r = .19$ across trials and across genotypes, respectively (Figure S1). For the subsequent analyses, we concatenated the two trials, resulting in 56 data points, i.e. (16 × 2) + (12 × 2). The reason for concatenation is that co-expression analyses usually gather all available data, thus increasing the power of detecting truly functional associations (Usadel *et al.*, 2009).

Unraveling functional lipid-transcript associations

Considering the disproportion among lipid classes (Figure 3a) and the correlation among lipid species (Figure 3b), for each of the nine lipid classes we processed the corresponding lipid species with a PCA and retrieved the coordinates of the first principal component (latent variable), referred to as lipid score vectors (LSVs), and used these as responses to be predicted. In summary, each of the nine LSVs results in finding the linear combination of acyl lipids from a particular class that maximizes the spread of the data points. Interestingly, considering that trials were analyzed jointly, all LSVs except DAG separated the two genotypes (2-way ANOVA $P < .01$; Table S1). In addition, the variance explained with each LSV (R^2) ranged from 0.37 to 0.85 across all nine classes (Figure 3c).

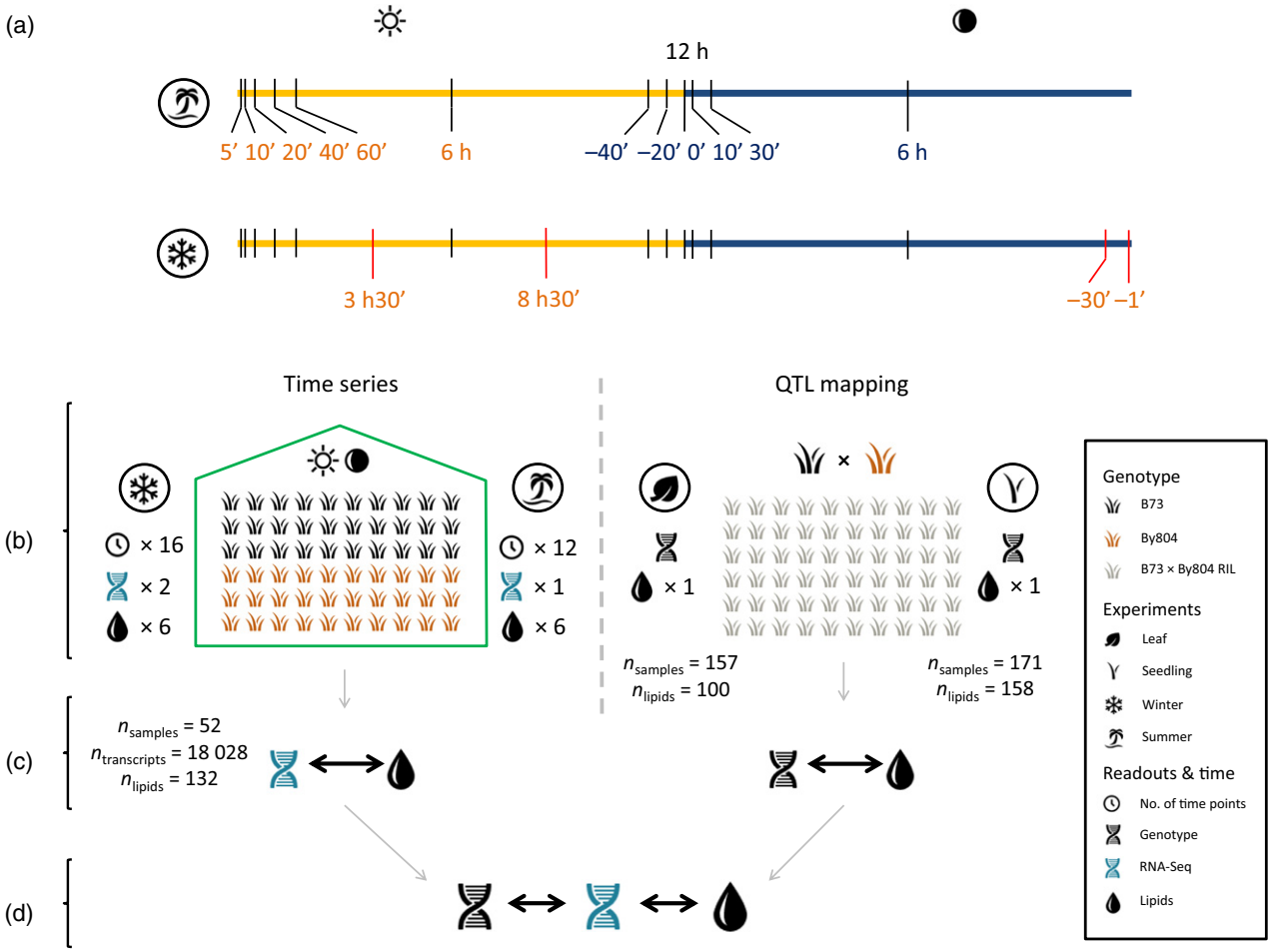

Figure 1. Schematic representation of the experimental design and data analysis.
(a) Schematic representation of the time points in the trials of 'summer' (top) and 'winter' (bottom), in the course of a 12-h light/12-h dark photoperiod. The time points shown are relative to either the beginning of the day (orange) or night (dark blue). The time series from the 'winter' trial differs from 'summer' by four time points marked in red.
(b) Experimental design of the time-series experiment ('Time series' left) and the quantitative trait locus (QTL) mapping experiment ('QTL mapping', right). The first comprised the measurement of gene expression and lipid levels in B73 and By804 plants in the 'winter' and 'summer' trials following the series given in (a). The second consisted of the measurement of lipid levels in leaf and seedling samples of a B73 × By804 RIL population.
(c) GFLASSO modeling (left) and mapping of lipid QTLs (right), using the data acquired from the time series and QTL mapping experiments, respectively.
(d) Lipid–gene association network analysis, using the intersection of the genes selected in (c), therefore integrating genomic, transcriptomic and lipidomic data (RIL, recombinant inbred lines). The multiplicative factors next to a symbol represent the number of replicates. The displayed icons are freely available at Icons8 (https://icons8.com/).

Overall, distinct acyl-lipid species were weighted evenly in the deployment of the LSVs used here (Figure S2). We also note that the corresponding loading weights can be either positive or negative, implying that LSVs might be negatively correlated to certain acyl-lipid species, thus warranting the subsequent selection of both positively and negatively associated transcripts.

To identify lipid–transcript associations, we used the genome-wide expression data set to determine the transcripts that are predictive of the acyl-lipid levels in the different classes represented by the LSVs, using a regularized high-dimensional regression framework. The regularized regression framework is the most suitable candidate, due to the substantially larger number of predictors in comparison to the number of samples (Figure 1). To this end, we used the profiles of all 18 028 transcripts as predictors in a unified GFLASSO framework, with the nine LSVs of the acyl-lipid classes as responses. GFLASSO allowed us to select genes that jointly predict multiple related traits, i.e. acyl-lipid latent variables, in a single framework.

After determining the model parameters which optimize the prediction accuracy of the model (Figure 4a), we investigated the model performance on a permutation of the samples. We observed that on permuted samples the error of the model was significantly larger, thus demonstrating that the predictability of LSVs from the whole transcriptome is not random (Figure 4b).

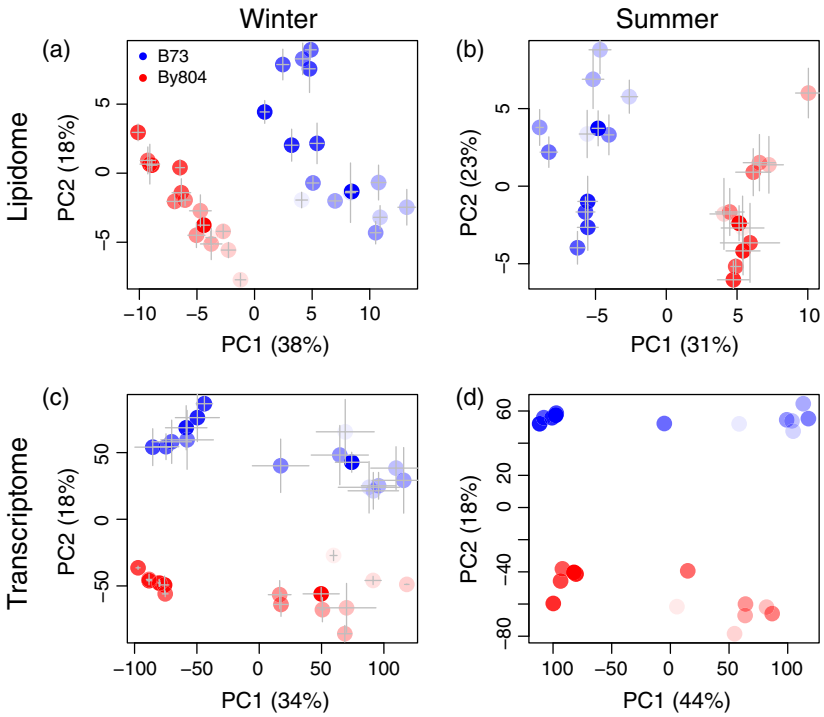

Figure 2. Principal component analysis of the data collected from the time-series experiment.
(a) Lipid data set from the 'winter' trial.
(b) Lipid data set from the 'summer' trial.
(c) Transcriptome data set from the 'winter' trial.
(d) Transcriptome data from the 'summer' trial.
The first two principal components (PCs) separate B73 (blue) from By804 (red) and clusters of time points. Percentages denote the amount of variance explained by each different PC. The time-point sequences are represented with increasing color transparency. Error bars represent the sample ± SE (standard error). No replicates were assessed for the expression data from the 'summer' trial.

To determine the list of predictive genes, we selected those whose respective regression coefficients were either below the 2.5th or above the 97.5th percentiles for each LSV. This rendered nine gene sets of size 902 each, containing approximately 5% of the entire expression data set (Figure S3, Table S2). Altogether, there were 3537 unique genes across all gene sets and 62.3% with unknown function.

Gene Ontology and promoter motif enrichment analysis

The gene sets were separately subjected to GO enrichment analysis. Given the high redundancy of many hierarchical terms (e.g. 'cellular response to starvation' and 'response to starvation'), we considered the top 50 most enriched categories (Fisher exact test $P < .05$). Across the gene sets, the most common functions included lipid metabolism, auxin biosynthesis and gravitropism, glycolysis, epigenetic and transcriptional regulation, cytoskeleton maintenance, ethylene biosynthesis and stress response (Table S3). With respect to lipid metabolism, we identified the terms 'galactolypid biosynthetic process' (SQDGs and DGDGs), 'unsaturated fatty acid biosynthetic process' (PGs and PCs), 'isoprenoid biosynthetic process' (a process that depletes acyl-coA, precursor of fatty acids: PGs, PCs, MGDGs and TAGs), 'oxylipin biosynthetic process' (fatty acid derivative and precursor of jasmonate; PIs), 'lipoate metabolic process' (fatty acid derivative; PEs), 'glycerol metabolic process' (DGDGs), 'glycolipid metabolic process' (DGDGs), 'liposaccharide metabolic process' (DGDGs) and 'polyol metabolic process' (a chemical group that includes glycerol; DAGs and TAGs). Therefore, our enrichment analysis demonstrated that the gene sets obtained by modeling the LSVs with the help of GFLASSO are involved in lipid metabolism and associated processes.

To complement these findings, we next conducted *in silico* promoter motif enrichment analysis for the gene sets identified in the first step of the analysis. For this purpose, we first performed untargeted motif identification and then compared the best hit to existing annotated Arabidopsis promoter motifs and retrieved the best match. We found that six out of the nine gene sets were significantly enriched for motifs associated with the AP2/EREB transcription factor family (SQDGs, DAGs, PEs, PIs, PGs and PCs). The DGDG and TAG gene sets were enriched for motifs that matched a LOB/AS2-family transcription factor and the MGDG gene set for motifs from a mitochondrial

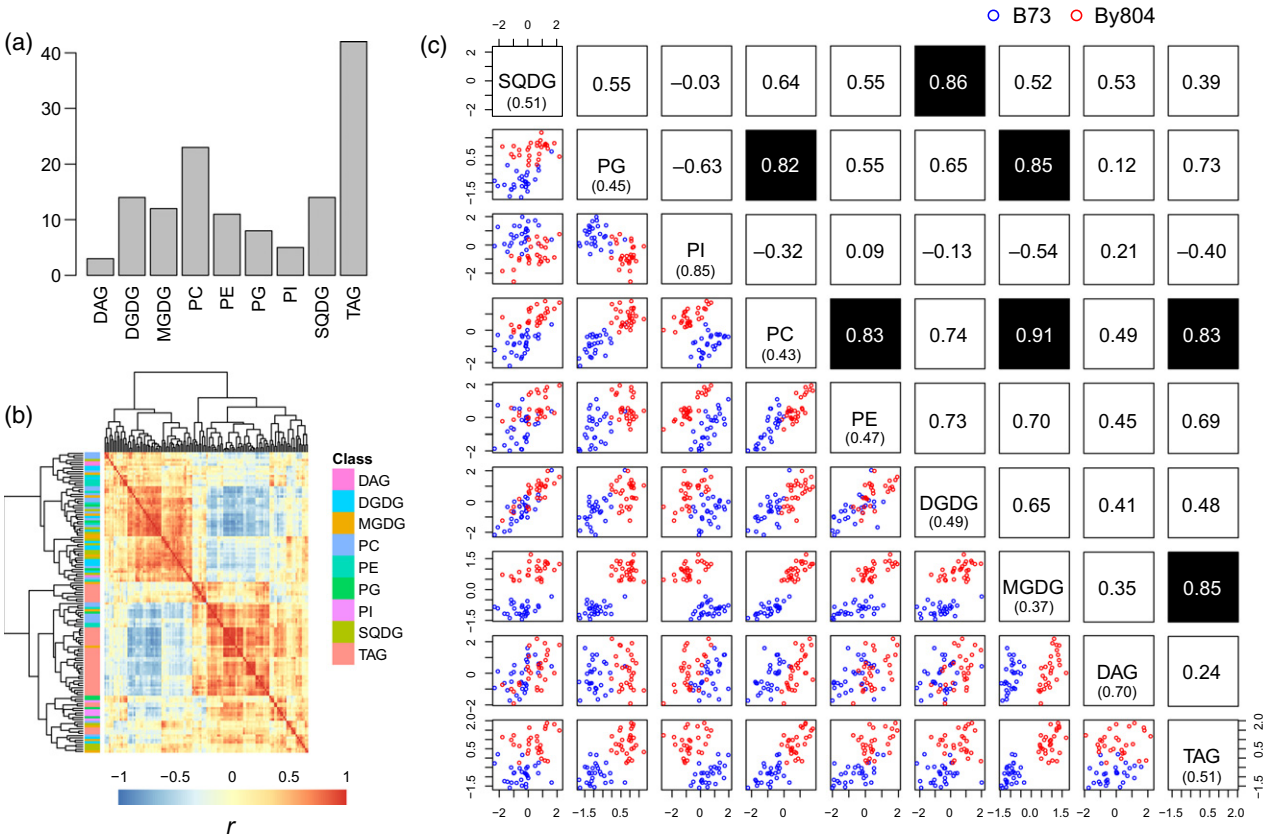

Figure 3. Reduction of acyl-lipid species into lipid score vectors.
(a) Absolute frequency of acyl-lipid classes in the lipid data set. The bar heights correspond to the total number of compounds representing each respective acyl-lipid class.
(b) Distance-based hierarchical cluster of the acyl-lipid pairwise Pearson correlation coefficients (*r*).
(c) Lower triangle: pairing scatter plot of the lipid score vectors (LSVs) from all nine lipid classes. Acyl-lipid classes are represented by the scores obtained from the first component (latent variable) in the corresponding principal component analysis (PCA). Values under parenthesis on the diagonal correspond to the variance explained (R^2) by each LSV in the respective PCA. Upper triangle: LSV pairwise *r*. Cells marked in black denote $r > 0.8$, and consequently the associations given to GFLASSO for the fusion constraint. Blue and red symbols represent the two genotypes, B73 and By804, respectively.
PI, phosphatidylinositol; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; SQDG, sulfoquinovosyldiacylglycerol; DAG, diacylglycerol; TAG, triacylglycerol.

transcription termination factor (mTERF) transcription factor (Table 1).

Co-localization of transcript–lipid associations with lipid QTLs

Finally, we identified the GFLASSO-selected candidate genes that co-localized with the lipid QTLs obtained in seedling and leaf samples from a B73 × By804 RIL population. In total, 158 and 100 acyl-lipid species were mapped from seedling and leaf samples in 171 and 157 RILs, respectively (Figure 1). By setting a logarithm of odds (LOD) threshold of 4.5 for the QTLs mapped to the different acyl-lipid classes, we were able to co-localize 12, 63, 81, 56, 63, 15, 4, 32 and 48 candidate genes with the DAG, DGDG, MGDG, PC, PE, PG, PI, SQDG (Figs S4–S11) and TAG (Figure 5, Table S4) QTLs, respectively. Altogether, there were 323 unique co-localized genes. To identify GFLASSO-selected genes co-localizing with QTLs of different lipid classes, and thus more likely to be functionally associated with lipid metabolism, we examined the corresponding gene–lipid association network. The resulting network connects all gene sets, except for that of PI. Notably, genes associated with MGDG and co-localized with the respective QTL were also associated with and co-localized with QTLs from PC, DAG, TAG, PC, PE and SQDG. In total we obtained 49 genes with a degree of two and a single one, a putative P-loop containing nucleoside triphosphate hydrolase (GRMZM2G313020), with degree of three, i.e. associated and co-localized with MGDG, PE and SQDG (Figure 6). In this list, we were also able to identify the vacuolar aquaporin TIP2-1 (GRMZM2G027098), a cytosolic glyceraldehyde-3-phosphate dehydrogenase (GRMZM 2G046804), histone H3C4 (GRMZM2G418258), 5′–3′ exonuclease (GRMZM2G096920), a carbohydrate transporter (GRMZM2G102683), an erythronate-4-phosphate dehydrogenase (GRMZM2G070780) and the VP1/ABI3 transcription

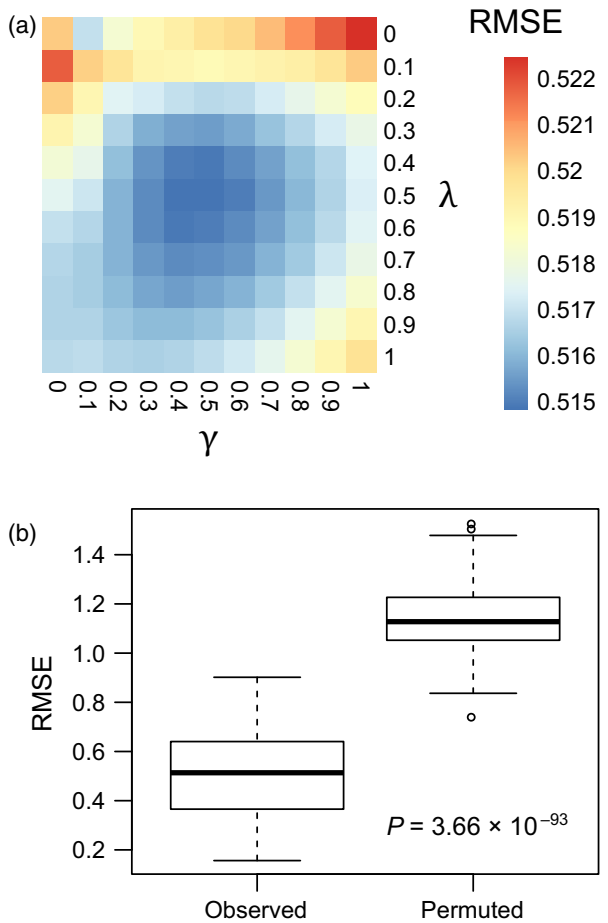

Figure 4. GFLASSO-based prediction of the lipid score vectors with the entire transcriptome.
(a) Cross-validation of GFLASSO. Each cell represents the average root mean square error (RMSE) across 135 (5 trials × 3 repeats × 9 lipid score vectors) estimates. The minimum average RMSE value (0.515) was obtained using $\lambda = 0.5$ and $\gamma = 0.5$.
(b) Permutation scheme. Comparison of the RMSE estimates in the GFLASSO model cross-validated with $\lambda = 0.5$ and $\gamma = 0.5$ ('Observed') compared with the same model upon permutation of the data points ('Permuted'). The *P*-value displayed derives from a two-tailed *t*-test.

factor *ZmAFL6* (GRMZM2G125596). We also identified the orthologs of Arabidopsis PVA12 (GRMZM2G006219), peroxin PEX14 (GRMZM2G169486), ubiquitin ATGP4 (GRMZM2G086267), auxin-responsive factor ARF1 (GRMZM2G137413) and MED21 (GRMZM2G074280) among others, including 29 of unknown function (Table S5).

DISCUSSION

Here we used time-resolved leaf transcriptome and lipidome data sets gathered from maize lines grown in two independent trials to determine lipid–gene associations. To this end, we used a two-step analysis that combines the findings from a QTL mapping with results from a novel multivariate integrative analysis, namely GFLASSO-based prediction of latent variables. The novelty of the approach is that the lipid–transcript associations are not built independently for each lipid species but rather the dependence between the species and their latent variable representatives is explicitly considered. Such an approach has the capacity to direct the search of explanatory variables (here transcript profiles) and narrow down the list of candidates.

Table 1 Gene set promoter motif enrichment analysis results

Lipid class	Enriched motifs[a]	MEME *E*[b]	Tomtom *E*[c]
SQDG	AP2/EREB (ESE1)	6.20×10^{-64}	1.81×10^{-1}
DGDG	LOB/AS2 (ASL18)	1.30×10^{-56}	2.28×10^{-2}
DAG	AP2/EREB (CRF4)	1.30×10^{-53}	1.58×10^{-1}
PE	AP2/EREB (ERF73)	2.00×10^{-37}	2.16×10^{-3}
PI	AP2/EREB (AT4G18450)	1.80×10^{-22}	1.05×10^{-2}
TAG	LOB/AS2 (ASL18)	1.40×10^{-56}	1.08×10^{-1}
PG	AP2/EREB (ERF105)	1.30×10^{-55}	1.17×10^{-3}
PC	AP2/EREB (ERF73)	2.20×10^{-80}	1.33×10^{-1}
MGDG	mTERF (AT5G23930)	3.90×10^{-59}	3.68×10^{-3}

[a]Enriched motifs: transcription factor families that putatively bind the motifs most enriched in each gene set. The Arabidopsis best hits are shown in parenthesis.
[b]MEME *E*: number of enriched motifs that, given the same width and site count, one would expect to find in a similarly sized set of randomly shuffled sequences.
[c]Tomtom *E*: number of times that the queried motif would be expected to match a known motif as well or better than the observed match in a randomized target database of the same size ($E = P \times n_{motifs}$).
PI, phosphatidylinositol; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; SQDG, sulfoquinovosyldiacylglycerol; DAG, diacylglycerol; TAG, triacylglycerol.

In conducting the analysis, we concatenated the two data sets. The idea behind this choice was to increase the power of detecting prevalent functional lipid–transcript associations, rather than associations specific to either trial; however, in this approach we cannot exclude that the detected associations reflect, to an extent, genotype–environment interactions. This approach was in line with our focus to identify a genetic basis for acyl-lipid metabolism that can be transferred to other tissues (e.g. maize kernel) as well as tested in species other than maize.

Our PCA revealed that lipid and transcript levels separate samples from genotypes and groups of early/late time points, respectively. Considering that By804 is the result of continuous selection for lipid traits, it is notable that acyl-lipid levels, in comparison with transcripts, better separate the genotypes. The reason for the lack of separation based on the level of transcripts was likely due to the consideration of the entire transcriptome, which also includes core functions conserved across diverse maize lines. This was consistent with the cross-trial and cross-genotype correlation analyses, which indicated that most of the transcriptome was less affected by either factor (i.e. genotype and time), whereas lipids were particularly variable with

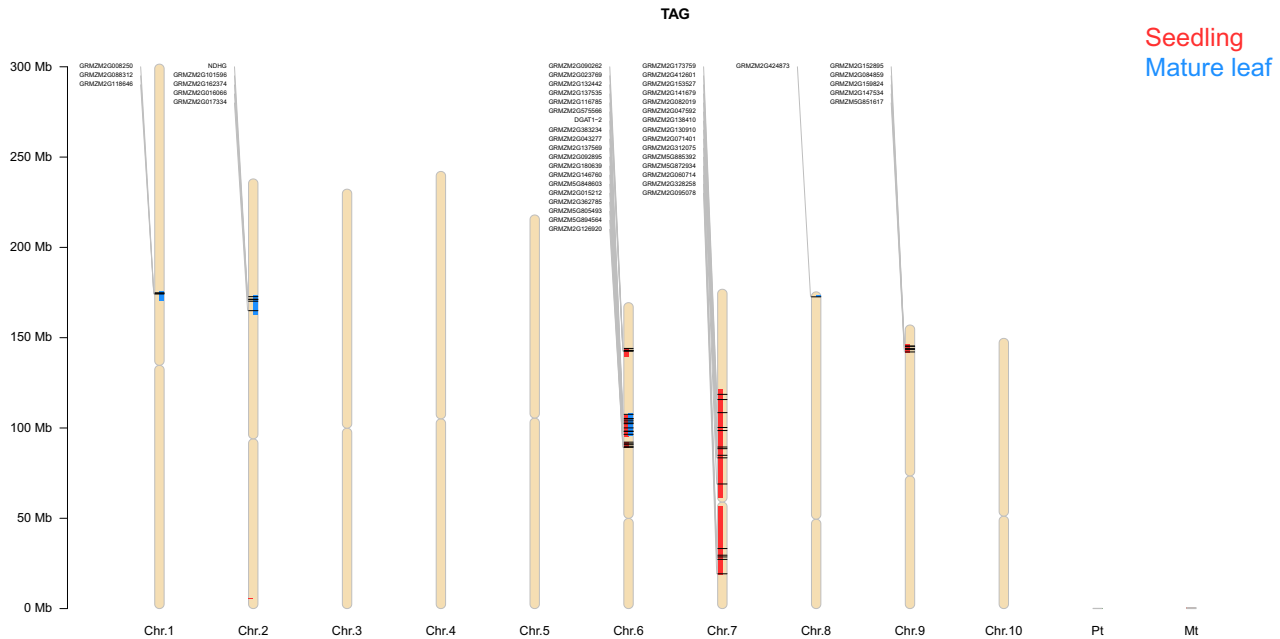

Figure 5. Co-localization of quantitative trait loci (QTLs) and GFLASSO candidates for triacylglycerols (TAGs).
All 10 maize chromosomes (Chr. 1–10), including the plastid (Pt) and mitochondrial (Mt) genomes, are depicted to scale (Mb, million base pairs). Genes whose transcripts were associated with lipid classes via GFLASSO (lipid–transcript associations) co-localizing to the QTL of TAG in seedlings (red) and mature leaves (blue) are shown as black lines. *DGAT1-2* is one of 48 co-localized candidates. A detailed gene list is available in Table S4.

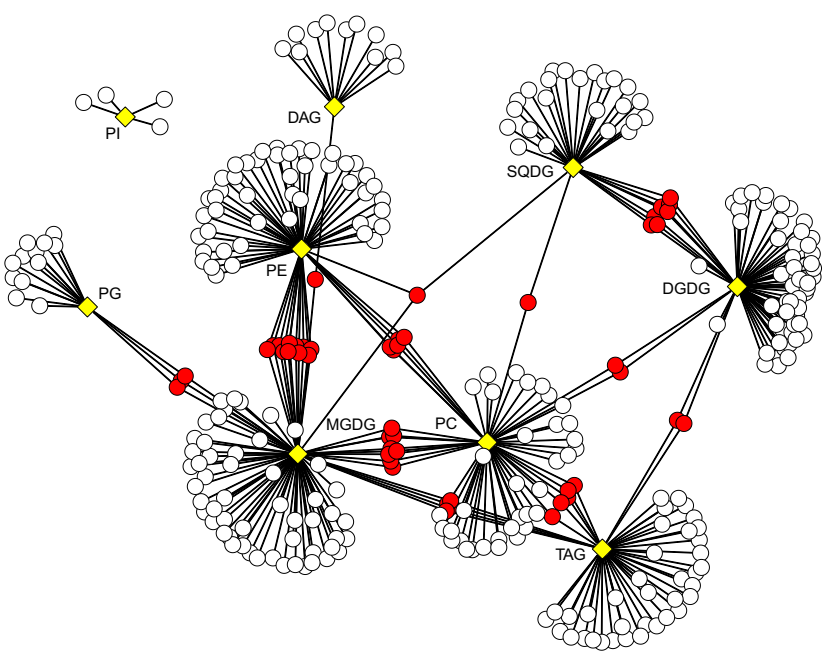

Figure 6. Lipid–gene association network.
Transcripts associated with lipid classes via GFLASSO (lipid–transcript associations) whose genes also co-localized to the QTL of those lipid classes (lipid–gene associations) are shown as white or red circles depending on having a single association or more, respectively. The nine distinct lipid classes are shown as yellow diamonds. PI, phosphatidylinositol; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; MGDG, monogalactosyldiacylglycerol; DGDG, digalactosyldiacylglycerol; SQDG, sulfoquinovosyldiacylglycerol; DAG, diacylglycerol; TAG, triacylglycerol.

respect to genotype. In addition, most LSVs separated B73 and By804 irrespective of the time course and trial, again underlining the large difference in lipid metabolism between the two genotypes.

The applied GFLASSO framework inherits the sparsity of LASSO (Tibshirani, 1996) and provides similar models for strongly correlated traits. Therefore, the derived optimal models simultaneously leveraged gene selection and

lessening of differences in the corresponding coefficient estimates for any given pair of correlated LSVs. Comparison of the models with those derived by permutation of transcriptomics data showed that the time-resolved gene expression levels were predictive of the lipid time profiles. Here, the use of PCA-derived components had the advantage of: (i) a reduced model bias, concerning the cross-validation error minimization, (ii) projections (i.e. LSVs) where the data points spread the most, thus refining generalizability, and finally (iii) feasible computations. However, we refrained from interpreting relationships among LSVs, as no biological association can be expected.

The enrichment analyses indicated the involvement of model-selected genes in lipid metabolism and related processes. Overall, lipid metabolism, auxin biosynthesis and gravitropism (governed by the polar auxin transport; Rashotte *et al.*, 2000), glycolysis, epigenetic and transcriptional regulation and cytoskeleton maintenance were GO terms enriched in at least one of the selected gene sets. A very recent study identified all these processes to be regulated by genes co-expressing with selected lipid biosynthesis genes in the oil palm mesocarp (Guerin *et al.*, 2016). Auxin biosynthetic and responsive genes were shown to be upregulated along with an increase in the accumulation of TAGs in the leaves of LEAFY COTYLEDON 2 (*LEC2*)-overexpressing Arabidopsis plants (Stone *et al.*, 2008). In addition, genes involved in fatty acid biosynthesis including *WRI1* were found to be co-expressed with auxin transporter genes in *Brassica napus* (Deng *et al.*, 2015). Glycolytic pathways make pyruvate available for FA biosynthesis (Rawsthorne, 2002), explaining why genes involved in FA biosynthesis together with those that hydrolyze cell wall components are predictive of the LSVs. In addition, cytoskeletal organization and sterols determine polar auxin transport and signaling (Yang *et al.*, 2013; Li *et al.*, 2014), suggestive of how these processes might be intertwined. Stress-related terms could indicate a role for membrane lipids upon stimulation by biotic and abiotic factors and rearrangement of cutin and cuticular waxes in the leaf surface (Shepherd and Wynne Griffiths, 2006; Kosma *et al.*, 2009; Le Provost *et al.*, 2013). In addition, oxylipins (here constituting an enriched GO term) such as jasmonate mediate stress responses and result from the oxidation of FAs (Andreou *et al.*, 2009). Ethylene plays a role in leaf senescence, a process that involves the degradation of organelle membranes (Lim *et al.*, 2007). Concretely, it was shown that ethylene causes a decrease in the levels of acyl-lipids such as MGDGs, DGDGs, PGs, PCs, PEs and PIs in Arabidopsis leaves (Jia and Li, 2015).

With respect to the motif enrichment analysis, targets of the AP2/EREB transcription factor family were consistent with knowledge about lipid metabolism. *WRI1a* and *WRI1b* are AP2/EREB transcription factors and two of the best characterized regulators of oil biosynthesis in maize (Pouvreau *et al.*, 2011). The implication of their involvement in the biosynthesis of FAs, the building blocks of all acyl-lipid species, was consistent with having AP2/EREB targets in multiple gene sets in this study. Also, Li and colleagues had previously identified, among 200 genes co-expressed with *WRI1a* in filling-stage kernels, 11 transcription factors structurally similar to either VP1/ABI3 or AP2/EREB transcription factor families (Li *et al.*, 2013). The mTERF transcription factors are known to control gene expression in mitochondria and plastids (Kleine, 2012). Interestingly, the motif of the mTERF protein we identified was very similar to AP2/EREB motifs (Machanick and Bailey, 2011). Nevertheless, the association of a mTERF transcription factor a MGDG, a plastidial galactolipid (and considering FAs are synthesized in the plastids; Post-Beittenmiller *et al.*, 1992), might help explain this finding. Finally, LOB/AS2 transcription factors, developmental regulators like AP2/EREB, are involved in auxin-regulated lateral root development (Matsumura *et al.*, 2009), again evoking co-regulation of lipid metabolism and auxin signaling. Auxin response factors (ARFs) have already been shown to be part of this interplay (Li *et al.*, 2015).

Our integrative approach enabled narrowing down the list of candidates drawn from the GFLASSO framework by co-localization with the QTLs of the corresponding acyl-lipid classes. For TAGs, in particular, we identified a well-characterized acyl-CoA:diacylglycerol acyltransferase, *DGAT1-2* (Zheng *et al.*, 2008) among 47 other candidates from the co-localization of lipid–transcript associations with lipid QTLs.

Finally, we inspected the network of co-localized genes to identify highly connected genes, i.e. genes contained in QTLs and associated with multiple lipid classes and thus more likely to be associated with lipid metabolism. The recruitment of aquaporins to the tonoplast is mediated by vacuolar trafficking, which was shown to be defective after inhibiting the biosynthesis of sterols and sphingolipids (Li *et al.*, 2015). On account of the importance of lipids for vacuolar sorting, synthesis of vacuolar proteins should then be coordinated with lipid metabolism. The identification of a glyceraldehyde-3-phosphate dehydrogenase, a putative erythronate-4-phosphate dehydrogenase and a carbohydrate transporter is supportive of the role of glycolysis in oil biosynthesis. The discovery of histone H3C4 and a 5′–3′ exonuclease is also supportive of the transcriptional and epigenetic regulatory functions uncovered by the GO enrichment analysis. In Arabidopsis, PVA12 coordinates the distribution of the sterol-binding protein ORP3a at the endoplasmic reticulum, facilitating sterol trafficking (Saravanan *et al.*, 2009). In Arabidopsis peroxisomes, PEX14 was shown to facilitate the unloading of thiolase from PTS2 cargos (Germain *et al.*, 2001; Lanyon-Hogg *et al.*, 2014). Arabidopsis ubiquitin ATGP4 was reported in a comprehensive proteomic study of the plasma membrane

(Marmagne *et al.*, 2007). Arabidopsis MED21 is a subunit of the Mediator complex that regulates RNA polymerase II and is important for resistance to fungi (Dhawan *et al.*, 2009). Finally, expression of the VP1/ABI3 member *ZmAFL6* was shown to peak in young kernels, and it was considered an ortholog of *AtLEC2*, a well-known master regulator in oil biosynthesis (Baud *et al.*, 2007; Grimault *et al.*, 2015). The regulators of oil biosynthesis *ZmWri1a* and *ZmWri1b* were both missing in our expression data set, probably because their expression is confined to the embryo (Pouvreau *et al.*, 2011).

We counterbalanced the large size of the gene sets from GFLASSO (each accounting for about 5% of the whole transcriptome) by raising the LOD threshold in the QTL mapping to 4.5. For LOD thresholds greater than 4.5 (e.g. 5), we could not identify QTLs for TAGs. Although a false discovery rate (FDR) < 0.05 was attained with LOD = 2.5, here we merge the QTLs of individual acyl-lipid species, allowing us to use a higher threshold. We note that from LOD = 2.5 to LOD = 4.5 the number of co-localized genes reduced from 254 to 81 (Figure S12). With respect to GFLASSO-based gene selection, the choice of the 2.5th and 97.5th percentiles is arbitrary and aimed at ultimately extracting a tractable number of candidates. Alternatively, when we selected candidates from GFLASSO with coefficient estimates either below or above the 5th or 95th percentiles, thus doubling the number of genes per gene set (from 902 to 1804), we noted that results were consistent with our previous enrichment and network analyses. The GO enrichment, in addition to the terms discussed above, included starch biosynthesis in multiple gene sets (Table S6). This is yet another cellular process that was found to be co-regulated with lipid metabolism in a co-expression study of oil palm (Guerin *et al.*, 2016). The motif enrichment consistently showed a general enrichment for AP2/EREB (DGDGs, DAGs, PIs, PCs) and LOB/AS2 (PEs, TAGs, PGs, MGDGs) transcription factors, and DBP (DNA-binding protein phosphatase) in the SQDG gene set. Compared with the previous result, the use of larger gene sets yielded smaller *E*-value estimates (Table S7). The co-localization network is approximately twice as large, displaying genes with a degree as high as four (Figure S13).

Our rich transcriptomics and lipidomics data sets provided a resource for conducting robust computational analysis to predict the function of maize genes involved in lipid biosynthesis. By conducting the two-step analysis that combined modern regression techniques with the findings from QTL analyses, we generated high-confidence gene candidates that could drive future experimental efforts aimed at characterizing this molecular process with important economic consequences. If lipid metabolism is indeed a highly conserved process, we hope our modeling results drawn from leaf material will partly qualify for understanding lipid metabolism in the kernel, essential for any undertaking aimed at improving the quality and quantity of maize oil.

EXPERIMENTAL PROCEDURES

Plant material

The time-series experiment involved B73 and By804 plants grown in a greenhouse in Huazhong Agricultural University, China. The plants were grown under a 12-h light/12-h dark (12L12D) photoperiod with a light intensity of about 7000 $\pm$ 200 lx and sampled 45 days after sowing. The 'summer' trial was conducted in July 2014, in which the eighth leaf from three biological replicates were sampled at 5 min, 10 min, 20 min, 40 min, 60 min, 6 h, 11 h 20 min, 11 h 40 min and 12 h after turning on the light and 10 min, 30 min and 6 h after onset of darkness (12 time points). The 'winter' trial was performed in January 2015, in which the eighth leaf from three biological replicates were sampled at 30 min before turning on the light, 1 min, 5 min, 10 min, 20 min, 40 min, 60 min, 3 h 30 min, 6 h, 8 h 30 min, 11 h 20 min, 11 h 40 min and 12 h after turning on the light, and 10 min, 30 min and 6 h after onset of darkness (16 time points, 12 of which are common to the 'summer' trial). With respect to time points coinciding with the moment the light was turned off (i.e. 12 h after turning on the light), a 10-sec tolerance was given for the light intensity to go down to 0 lx before sampling.

The QTL mapping experiment comprised 196 RILs from a B73 $\times$ By804 bi-parental population (Chander *et al.*, 2008). The seedling samples were harvested from the fifth leaf of seedlings, germinated in Huazhong Agricultural University experimental field (Wuhan, 109°51′ E, 18°25′ N) in 2013, and the leaf samples, at 15 days after pollination, from ear leaves in plants grown in the same field experimental station.

Genotyping and high-density bin map

All RILs were genotyped with an Illumina MaizeSNP50 BeadChip (https://www.illumina.com/). The bin map was constructed according to in-house Perl scripts (https://github.com/panqingchun/linkage_map). More information [e.g. single nucleotide polymorphism (SNP) missing rate, SNP heterozygosity, methods of bin map construction, marker number, bin length] can be found in our previous study (Wen *et al.*, 2015).

Lipid profiling

Samples from both the time series and QTL mapping experiments were lyophilized and stored for subsequent analysis. Samples were processed using ultra-performance liquid chromatography coupled with Fourier transform mass spectrometry (UPLC-FT-MS) on a C8 reverse phase column coupled with an Exactive mass spectrometer (Thermo-Fisher, http://www.thermofisher.com) in positive and negative ionization modes. Processing of chromatograms, peak detection and integration were performed using REFINER MSH 10 (GeneData, http://www.genedata.com). Processing of mass spectrometry data included the removal of the fragmentation information, isotopic peaks and chemical noise. Chromatographic peaks were annotated using an in-house lipid database.

Transcriptomic profiling

Transcript levels in the samples from the time-series experiment were determined by RNA sequencing (RNA-Seq). RNA extraction, library construction with an insert size of 300–500 bp, 150-bp

paired-end Illumina sequencing, read mapping and SNP calling followed published protocols. Briefly, total RNA was extracted using a Quick RNA Isolation Kit (Huayueyang, http://www.huayueyang.com/) according to its protocol. Bio-Rad Experion (http://www.bio-rad.com/) was used to evaluate the quality of total RNA. The RNA samples that passed quality control were used to construct Illumina stranded mRNA libraries with a Truseq Stranded mRNA sample preparation kit (Low Sample protocol; Illumina) and then subjected to pair-end sequencing on Illumina Hiseq 3000 platform. Trimmomatic (v.3.0) (Bolger *et al.*, 2014) was used to remove low-quality bases and reads. The quality of the raw reads and trimmed reads was checked using the FastQC (v.0.11.3) program (Andrews, 2010). Then we mapped those trimmed reads to the maize reference genome AGPv3 using Tophat2 (v.2.0.13) (Kim *et al.*, 2013). Cufflinks (v.2.0.2) was used to assemble transcripts and estimate their abundances (Trapnell *et al.*, 2013). The expression data were acquired from two technical replicates for the winter trial and a single sample for the summer trial.

Statistical treatment

The lipid data from both the time series and QTL experiments were normalized for (i) the initial amount of sample injected (mg) and (ii) differences in the ion intensities among UPLC-FT-MS batches (i.e. the batch effect). Next, features with missing values exceeding 20% of the total number of samples were discarded. The remaining missing values were imputed using a Random Forest regression model, with 500 trees per forest, using the 'missForest' R package (Stekhoven and Bühlmann, 2012). The standardized median genotypic values were computed after $\log_{10}$-transforming the data. The standardization (i.e. unit-variance scaling) was conducted separately in the two time-series experiments (i.e. 'winter' and 'summer' trials). Data from both trials were concatenated into a single data set.

The expression data from the time-series experiment were normalized for sequencing depth and gene size (fragments per kilobase of transcript per million mapped reads, FPKM). Next, reads with missing values exceeding 20% of the total number of samples were discarded. The remaining missing values were imputed using the first three principal components from a Bayesian PCA, using the 'pcaMethods' R package (Stacklies *et al.*, 2007). The standardized median genotypic values were computed after $\log_2$-transforming the data. The standardization was conducted separately in the two time-series experiments. Data from both trials were concatenated into a single data set.

Multitask-regression of lipid levels

The LSVs were obtained from the first principal component (latent variable) in the PCA of the corresponding lipid species.

All transcript profiles, concatenated with respect to genotype and trial, were used as predictors in the GFLASSO modeling framework (Kim *et al.*, 2009), aiming to jointly predict the LSVs. The LSV correlation-based network that determines solution parsimony can be either weighted (i.e. Pearson correlation coefficient) or unweighted (i.e. presence or absence of association by imposing a correlation threshold). We opted for a unweighted network structure, since the weighted counterpart has shown little improvement in accuracy with great expense of computational power (Kim *et al.*, 2009). A Pearson correlation threshold of $r = .80$ provided a good compromise between running time and network density. The regularization and fusion parameters (λ and γ, respectively) were determined from the smallest root mean squared error (RMSE) estimate in a three-fold cross-validation repeated five times. The tested parameters encompassed all combinations between λ and γ with values ranging from 0 to 1 (inclusive) in step increments of 0.1. GFLASSO was refitted with the optimal parameters and the coefficient (β) matrix was thresholded so that values below the 2.5th percentile or above the 97.5th percentile for each LSV were set to 1 and otherwise to 0.

Gene Ontology enrichment analysis

Gene Ontology enrichment was conducted using the R package 'topGO' (Alexa and Rahnenfuhrer, 2016) with the Fisher exact test. The GO term library used for the enrichment was downloaded from the Ensembl Gramene Biomart (http://ensembl.gramene.org/biomart/martview).

Promoter motif enrichment analysis

Promoter sequences of 5 kb in length were obtained from the GRASSIUS Promoter Sequence Database (http://grassius.org/grasspromdb.html) (Yilmaz *et al.*, 2009). Untargeted motif discovery (between 6 and 30 wide) was performed on the promoter sequences using MEME-ChIP (http://meme-suite.org/tools/meme-chip) (Machanick and Bailey, 2011), employing the normal enrichment mode with MEME (Bailey and Elkan, 1994). The obtained *E*-values are estimates of the expected number of motifs that, given the same width and site count, one would find in a similarly sized set of randomly shuffled sequences. Motif–motif similarity between the most significantly enriched motif and those 872 in the Arabidopsis DNA affinity purification (DAP) atlas generated from O'Malley *et al.* (2016) was evaluated with the Tomtom algorithm (Gupta *et al.*, 2007). Here, similarly, the reported *E*-values are *P*-values adjusted for multiple testing, due to querying a single motif against an entire library. Essentially, *E* is the expected number of times that the queried motif would be expected to match a known motif as well as or better than the observed match in a randomized target database of the same size.

Network analysis

Network analysis was conducted using the R package 'statnet' (Handcock *et al.*, 2003).

Association mapping

The mapping of lipid QTLs using seedling and leaf samples from the QTL experiment was performed using QTLCartographer (Wang *et al.*, 2012). A LOD threshold of 4.5 was used for delineating the QTLs. Lipid QTLs were merged based on lipid classes. More detailed information can be found in our previous study (Wen *et al.*, 2015).

AUTHORS' CONTRIBUTIONS

YB, WW, JY, ZN and LW conceived and designed the experiments. FAL and KL performed the experiments and analyzed data. All authors contributed to the written manuscript.

ACKNOWLEDGEMENTS

We thank Dr Álvaro Cuadros-Inostroza for assisting in the design of the co-localization maps, Kris Sankaran for making his GFLASSO implementation publicly available and the National Key Research and Development Program of China (2016YFD0101003) for funding.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

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

Figure S1. Cross-trial and cross-genotype correlations in the transcriptome and lipidome data sets.

Figure S2. Contribution of the different acyl-lipids in the deployment of the lipid score vectors.

Figure S3. Candidate gene selection from the fitted GFLASSO model.

Figure S4. Co-localization of quantitative trait loci and GFLASSO candidates for diacylglycerols.

Figure S5. Co-localization of quantitative trait loci and GFLASSO candidates for digalactosyldiacylglycerols.

Figure S6. Co-localization of quantitative trait loci and GFLASSO candidates for monogalactosyldiacylglycerols.

Figure S7. Co-localization of quantitative trait loci and GFLASSO candidates for phosphatidylcholines.

Figure S8. Co-localization of quantitative trait loci and GFLASSO candidates for phosphatidylethanolamines.

Figure S9. Co-localization of quantitative trait loci and GFLASSO candidates for phosphatidylglycerols.

Figure S10. Co-localization of quantitative trait loci and GFLASSO candidates for phosphatidylinositols.

Figure S11. Co-localization of quantitative trait loci and GFLASSO candidates for sulfoquinovosyldiacylglycerols.

Figure S12. Effect of quantitative trait locus logarithm of odds (LOD) thresholds on the number of co-localized genes per lipid class.

Figure S13. Lipid-gene association network using the 5th and 95th percentiles in selecting GFLASSO candidates.

Table S1. Two-way analysis of variance on the lipid score vectors.

Table S2. Gene sets selected from the GFLASSO after thresholding coefficients with the 2.5th and 97.5th percentiles.

Table S3. Top 50 Gene Ontology enriched categories for biological process.

Table S4. Gene sets shortlists after co-localization with the quantitative trait locus of the corresponding lipid class.

Table S5. Genes transcriptionally and genetically associated with multiple lipid classes.

Table S6. Top 50 Gene Ontology enriched categories for biological process using the 5th and 95th percentiles in selecting GFLASSO candidates.

Table S7. Gene set promoter motif enrichment analysis results using the 5th and 95th percentiles in selecting GFLASSO candidates.

Data S1. Lipidomics and RNA sequencing data from the time-series experiment. Lipid data from the time-series experiment. Expression data from the time-series experiment.

REFERENCES

Alexa, A. and Rahnenfuhrer, J. (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.30.0.

An, D., Kim, H., Ju, S., Go, Y.S., Kim, H.U. and Suh, M.C. (2017) Expression of Camelina WRINKLED1 Isoforms Rescue the Seed Phenotype of the Arabidopsis wri1 Mutant and Increase the Triacylglycerol Content in Tobacco Leaves. *Frontiers in Plant Science*, **8**, 34.

Andreou, A., Brodhun, F. and Feussner, I. (2009) Biosynthesis of oxylipins in non-mammals. *Prog. Lipid Res.*, **48**, 148–170. https://doi.org/10.1016/j.plipres.2009.02.002

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Andrianov, V., Borisjuk, N., Pogrebnyak, N. *et al.* (2010) Tobacco as a production platform for biofuel: Overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnol. J.*, **8**, 277–287.https://doi.org/10.1111/j.1467-7652.2009.00458.x

Aoki, K., Ogata, Y. and Shibata, D. (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.*, **48**, 381–390. https://doi.org/10.1093/pcp/pcm013

Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings/International Conference on Intelligent Systems for Molecular Biology; ISMB*, **2**, 28–36.

Baud, S. and Lepiniec, L. (2010) Physiological and developmental regulation of seed oil production. *Prog. Lipid Res.*, **49**, 235–249. https://doi.org/10.1016/j.plipres.2010.01.001

Baud, S., Mendoza, M.S., To, A., Harscoët, E., Lepiniec, L. and Dubreucq, B. (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. *Plant J.*, **50**, 825–838. https://doi.org/10.1111/j.1365-313X.2007.03092.x

Beisson, F., Koo, A.J.K., Ruuska, S. *et al.* (2003) Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol.*, **132**, 681–697.https://doi.org/10.1104/pp.103.022988

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouvier-Navé, P., Benveniste, P., Oelkers, P., Sturley, S.L. and Schaller, H. (2000) Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:Diacylglycerol acyltransferase. *Eur. J. Biochem.*, **267**, 85–96. https://doi.org/10.1046/j.1432-1327.2000.00961.x

Bylesjö, M., Eriksson, D., Kusano, M., Moritz, T. and Trygg, J. (2007) Data integration in plant biology: The O2PLS method for combined modeling of transcript and metabolite data. *Plant J.*, **52**, 1181–1191. https://doi.org/10.1111/j.1365-313X.2007.03293.x

Cañas, R.A., Yesbergenova-Cuny, Z., Simons, M. *et al.* (2017) Exploiting the Genetic diversity of maize using a combined metabolomic, enzyme activity profiling, and metabolic modeling approach to link leaf physiology to Kernel yield. *Plant Cell*, **5**, 919–943.

Cavill, R., Jennen, D., Kleinjans, J. and Briedé, J.J. (2016) Transcriptomic and metabolomic data integration. *Brief. Bioinform.*, **17**, 891–901. https://doi.org/10.1093/bib/bbv090

Chan, E.K.F., Rowe, H.C., Corwin, J.A., Joseph, B. and Kliebenstein, D.J. (2011) Combining genome-wide association mapping and transcriptional networks to identify novel genes controlling glucosinolates in *Arabidopsis thaliana*. *PLoS Biol.*, **9**(8), e1001125. https://doi.org/10.1371/journal.pbio.1001125

Chander, S., Guo, Y.Q., Yang, X.H., Zhang, J., Lu, X.Q., Yan, J.B., Song, T.M., Rocheford, T.R. and Li, J.S. (2008) Using molecular markers to identify two major loci controlling carotenoid contents in maize grain. *Theoretical and Applied Genetics*, **116**, 223–233. https://doi.org/10.1007/s00122-007-0661-7

Chen, W., Wang, W., Peng, M. *et al.* (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.*, **7**, 12767.https://doi.org/10.1038/ncomms12767

Cook, J.P., McMullen, M.D., Holland, J.B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E.S. and Flint-Garcia, S.A. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol.*, **158**, 824–834. https://doi.org/10.1104/pp.111.185033

Deng, W., Yan, F., Zhang, X., Tang, Y. and Yuan, Y. (2015) Transcriptional profiling of canola developing embryo and identification of the important roles of BnDof5.6 in embryo development and fatty acids synthesis. *Plant Cell Physiol.*, **56**, 1624–1640. https://doi.org/10.1093/pcp/pcv074

Dhawan, R., Luo, H., Foerster, A.M., AbuQamar, S., Du, H.-N., Briggs, S.D., Scheid, O.M. and Mengiste, T. (2009) HISTONE MONOUBIQUITINATION1 Interacts with a Subunit of the Mediator Complex and Regulates Defense against Necrotrophic Fungal Pathogens in Arabidopsis. *Plant Cell*, **21**, 1000–1019.

Gargouri, M., Park, J.-J., Holguin, F.O., Kim, M.-J., Wang, H., Deshpande, R.R., Shachar-Hill, Y., Hicks, L.M. and Gang, D.R. (2015) Identification of regulatory network hubs that control lipid metabolism in Chlamydomonas reinhardtii. *J. Exp. Bot.*, **66**, 4551–4566. https://doi.org/10.1093/jxb/erv217

Germain, V., Rylott, E.L., Larson, T.R., Sherson, S.M., Bechtold, N., Carde, J.-P., Bryce, J.H., Graham, I.A. and Smith, S.M. (2001) Requirement for 3-ketoacyl-CoA thiolase-2 in peroxisome development, fatty acid β-oxidation and breakdown of triacylglycerol in lipid bodies of Arabidopsis seedlings. *Plant J.*, **28**, 1–12. https://doi.org/10.1046/j.1365-313X.2001.01095.x

Graham, I.A. and Eastmond, P.J. (2002) Pathways of straight and branched chain fatty acid catabolism in higher plants. *Prog. Lipid Res.*, **41**, 156–181. https://doi.org/10.1016/S0163-7827(01)00022-4

Grimault, A., Gendrot, G., Chaignon, S., Gilard, F., Tcherkez, G., Thévenin, J., Dubreucq, B., Depège-Fargeix, N. and Rogowsky, P.M. (2015) Role of B3 domain transcription factors of the AFL family in maize kernel filling. *Plant Sci.*, **236**, 116–125. https://doi.org/10.1016/j.plantsci.2015.03.021

Guerin, C., Joët, T., Serret, J. *et al.* (2016) Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. *Plant J.*, **87**, 423–441. https://doi.org/10.1111/tpj.13208

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. and Noble, W.S. (2007) Quantifying similarity between motifs. *Genome Biol.*, **8**, R24. https://doi.org/10.1186/gb-2007-8-2-r24

Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M. and Morris, M. (2003) statnet: Software tools for the Statistical Modeling of Network Data.

Haslam, R.P., Sayanova, O., Kim, H.J., Cahoon, E.B. and Napier, J.A. (2016) Synthetic redesign of plant lipid metabolism. *Plant J.*, **87**, 76–86. https://doi.org/10.1111/tpj.13172

Horn, P.J., Liu, J., Cocuron, J.-C., McGlew, K., Thrower, N.A., Larson, M., Lu, C., Alonso, A.P. and Ohlrogge, J. (2016) Identification of multiple lipid genes with modifications in expression and sequence associated with the evolution of hydroxy fatty acid accumulation in Physaria fendleri. *Plant J.*, **86**, 322–348. https://doi.org/10.1111/tpj.13163

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

Jia, Y. and Li, W. (2015) Characterisation of lipid changes in ethylene-promoted senescence and its retardation by suppression of phospholipase Dδ in Arabidopsis Leaves. *Frontiers in Plant Science*, **6**, 1045.

Jian, H., Yang, B., Zhang, A., Zhang, L., Xu, X., Li, J. and Liu, L. (2017) Screening of Candidate Leaf Morphology Genes by Integration of QTL Mapping and RNA Sequencing Technologies in Oilseed Rape (Brassica napus L.). *PLoS ONE*, **12**, e0169641. https://doi.org/10.1371/journal.pone.0169641

Kim, S., Sohn, K.A. and Xing, E.P. (2009) A multivariate regression approach to association analysis of a quantitative trait network. *Bioinformatics*, **25**, I204–I212. https://doi.org/10.1093/bioinformatics/btp218

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kleine, T. (2012) Arabidopsis thaliana mTERF proteins: Evolution and functional classification. *Frontiers in Plant Science*, **3**, 233.

Kosma, D.K., Bourdenx, B., Bernard, A., Parsons, E.P., Lü, S., Joubès, J. and Jenks, M.A. (2009) The impact of water deficiency on leaf cuticle lipids of Arabidopsis. *Plant Physiol.*, **151**, 1918–1929. https://doi.org/10.1104/pp.109.141911

Lanyon-Hogg, T., Hooper, J., Gunn, S., Warriner, S.L. and Baker, A. (2014) PEX14 binding to Arabidopsis PEX5 has differential effects on PTS1 and PTS2 cargo occupancy of the receptor. *FEBS Lett.*, **588**, 2223–2229. https://doi.org/10.1016/j.febslet.2014.05.038

Laurie, C.C., Chasalow, S.D., LeDeaux, J.R., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T.R. and Dudley, J.W. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics*, **168**, 2141–2155. https://doi.org/10.1534/genetics.104.029686

Le Provost, G., Domergue, F., Lalanne, C., Ramos Campos, P., Grosbois, A., Bert, D., Meredieu, C., Danjon, F., Plomion, C. and Gion, J.-M. (2013) Soil water stress affects both cuticular wax content and cuticle-related gene expression in young saplings of maritime pine (Pinus pinasterAit). *BMC Plant Biol.*, **13**, 95. https://doi.org/10.1186/1471-2229-13-95

Li, H., Peng, Z., Yang, X. *et al.* (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.*, **45**, 43–50. https://doi.org/10.1038/ng.2484

Li, G., Liang, W., Zhang, X., Ren, H., Hu, J., Bennett, M.J. and Zhang, D. (2014) Rice actin-binding protein RMD is a key link in the auxin–actin regulatory loop that controls cell growth. *Proc. Natl Acad. Sci.*, **111**, 10377–10382.https://doi.org/10.1073/pnas.1401680111

Li, R., Sun, R., Hicks, G.R. and Raikhel, N.V. (2015) Arabidopsis ribosomal proteins control vacuole trafficking and developmental programs through the regulation of lipid metabolism. *Proc. Natl Acad. Sci.*, **112**, E89–E98. https://doi.org/10.1073/pnas.1422656112

Li-Beisson, Y., Shorrosh, B., Beisson, F, *et al.* (2013) Acyl-Lipid Metabolism. *Arabidopsis Book*, **11**, e0161. https://doi.org/10.1199/tab.0161

Lim, P.O., Kim, H.J. and Nam, H.G. (2007) Leaf senescence. *Annu. Rev. Plant Biol.*, **58**, 115–136. https://doi.org/10.1146/annurev.arplant.57.032905.105316

Liu, J., Hua, W., Zhan, G., Wei, F., Wang, X., Liu, G. and Wang, H. (2010) Increasing seed mass and oil content in transgenic Arabidopsis by the overexpression of wri1-like gene from Brassica napus. *Plant Physiol. Biochem.*, **48**, 9–15. https://doi.org/10.1016/j.plaphy.2009.09.007

Ma, W., Kong, Q., Arondel, V., Kilaru, A., Bates, P.D., Thrower, N.A., Benning, C. and Ohlrogge, J.B. (2013) WRINKLED1, A Ubiquitous Regulator in Oil Accumulating Tissues from Arabidopsis Embryos to Oil Palm Mesocarp. *PLoS ONE*, **8**, e68887. https://doi.org/10.1371/journal.pone.0068887

Machanick, P. and Bailey, T.L. (2011) MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696–1697. https://doi.org/10.1093/bioinformatics/btr189

Mackay, T.F.C., Stone, E.A. and Ayroles, J.F. (2009) The genetics of quantitative traits: Challenges and prospects. *Nat. Rev. Genet.*, **10**, 565EP. https://doi.org/10.1038/nrg2612

Marmagne, A., Ferro, M., Meinnel, T., Bruley, C., Kuhn, L., Garin, J., Barbier-Brygoo, H. and Ephritikhine, G. (2007) A High Content in Lipid-modified Peripheral Proteins and Integral Receptor Kinases Features in the Arabidopsis Plasma Membrane Proteome. *Mol. Cell Proteomics*, **6**, 1980–1996. https://doi.org/10.1074/mcp.M700099-MCP200

Matsumura, Y., Iwakawa, H., Machida, Y. and Machida, C. (2009) Characterization of genes in the ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES (AS2/LOB) family in Arabidopsis thaliana, and functional and molecular comparisons between AS2 and other family members. *Plant J.*, **58**, 525–537. https://doi.org/10.1111/j.1365-313X.2009.03797.x

Mentzen, W.I., Peng, J., Ransom, N., Nikolau, B.J. and Wurtele, E.S. (2008) Articulation of three core metabolic processes in Arabidopsis: Fatty acid biosynthesis, leucine catabolism and starch metabolism. *BMC Plant Biol.*, **8**, 76. https://doi.org/10.1186/1471-2229-8-76

Mintz-Oron, S., Mandel, T., Rogachev, I. *et al.* (2008) Gene expression and metabolism in Tomato fruit surface tissues. *Plant Physiol.*, **147**, 823–851.https://doi.org/10.1104/pp.108.116004

Moose, S.P., Dudley, J.W. and Rocheford, T.R. (2004) Maize selection passes the century mark: A unique resource for 21st century genomics. *Trends Plant Sci.*, **9**, 358–364. https://doi.org/10.1016/j.tplants.2004.05.005

Ohlrogge, J.B., Browse, J. and Somerville, C.R. (1991) The genetics of plant lipids. *Biochem. Biophys. Acta.*, **1**, 1–26. https://doi.org/10.1016/0005-2760(91)90294-R

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A. and Ecker, J.R. (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, **165**, 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Post-Beittenmiller, D., Roughan, G. and Ohlrogge, J.B. (1992) Regulation of plant fatty acid biosynthesis: Analysis of Acyl-Coenzyme A and Acyl-Acyl Carrier Protein Substrate Pools in Spinach and Pea Chloroplasts. *Plant Physiol.*, **100**, 923–930. https://doi.org/10.1104/pp.100.2.923

Pouvreau, B., Baud, S., Vernoud, V., Morin, V., Py, C., Gendrot, G., Pichon, J.P., Rouster, J., Paul, W. and Rogowsky, P.M. (2011) Duplicate maize

Wrinkled1 transcription factors activate target genes involved in seed oil biosynthesis. *Plant Physiol.*, **156**, 674–686. https://doi.org/10.1104/pp.111.173641

Rashotte, A.M., Brady, S.R., Reed, R.C., Ante, S.J. and Muday, G.K. (2000) Basipetal Auxin Transport Is Required for Gravitropism in Roots of Arabidopsis. *Plant Physiol.*, **122**, 481–490. https://doi.org/10.1104/pp.122.2.481

Rawsthorne, S. (2002) Carbon flux and fatty acid synthesis in plants. *Prog. Lipid Res.*, **41**, 182–196. https://doi.org/10.1016/S0163-7827(01)00023-6

Rhee, S.Y. and Mutwil, M. (2014) Towards revealing the functions of all genes in plants. *Trends Plant Sci.*, **19**, 212–221. https://doi.org/10.1016/j.tplants.2013.10.006

Rischer, H., Orešič, M., Seppänen-Laakso, T., Katajamaa, M., Lammertyn, F., Ardiles-Diaz, W., van Montagu, M.C.E., Inzé, D., Oksman-Caldentey, K.-M. and Goossens, A. (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *Proc. Natl Acad. Sci.*, **103**, 5614–5619. https://doi.org/10.1073/pnas.0601027103

Saito, K., Hirai, M.Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. *Trends Plant Sci.*, **13**, 36–43. https://doi.org/10.1016/j.tplants.2007.10.006

Sanjaya, , Miller, R., Durrett, T.P. *et al.* (2013) Altered Lipid Composition and Enhanced Nutritional Value of Arabidopsis Leaves following Introduction of an Algal Diacylglycerol Acyltransferase 2. *The Plant Cell Online*, **25**, 677–693.https://doi.org/10.1105/tpc.112.104752

Saravanan, R.S., Slabaugh, E., Singh, V.R., Lapidus, L.J., Haas, T. and Brandizzi, F. (2009) The targeting of the oxysterol-binding protein ORP3a to the endoplasmic reticulum relies on the plant VAP33 homolog PVA12. *Plant J.*, **58**, 817–830. https://doi.org/10.1111/j.1365-313X.2009.03815.x

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M. and Ligterink, W. (2016) Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science*, **7**, 444.

Shepherd, T. and Wynne Griffiths, D. (2006) The effects of stress on plant cuticular waxes. *New Phytol.*, **171**, 469–499. https://doi.org/10.1111/j.1469-8137.2006.01826.x

Stacklies, W., Redestig, H., Scholz, M., Walther, D. and Selbig, J. (2007) pcaMethods - A bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, **23**, 1164–1167. https://doi.org/10.1093/bioinformatics/btm069

Stekhoven, D.J. and Bühlmann, P. (2012) Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, **28**, 112–118. https://doi.org/10.1093/bioinformatics/btr597

Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.-F., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl Acad. Sci.*, **105**, 3151–3156. https://doi.org/10.1073/pnas.0712364105

Sun, R., Ye, R., Gao, L., Zhang, L., Wang, R., Mao, T., Zheng, Y., Li, D. and Lin, Y. (2017) Characterization and Ectopic Expression of CoWRI1, an AP2/EREBP Domain-Containing Transcription Factor from Coconut (*Cocos nucifera* L.) Endosperm, Changes the Seeds Oil Content in Transgenic Arabidopsis thaliana and Rice (*Oryza sativa* L.). *Frontiers in Plant Science*, **8**, 63.

Szymanski, J., Brotman, Y., Willmitzer, L. and Cuadros-Inostroza, Á. (2014) Linking gene expression and membrane lipid composition of Arabidopsis. *Plant Cell*, **26**, 915–928. https://doi.org/10.1105/tpc.113.118919

Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267–288.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L. and Pachter, L. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotech*, **31**, 46–53. https://doi.org/10.1038/nbt.2450

Troncoso-Ponce, M.A., Kilaru, A., Cao, X., Durrett, T.P., Fan, J., Jensen, J.K., Thrower, N.A., Pauly, M., Wilkerson, C. and Ohlrogge, J.B. (2011) Comparative deep transcriptional profiling of four developing oilseeds. *Plant J.*, **68**, 1014–1027. https://doi.org/10.1111/j.1365-313X.2011.04751.x

Upadhyaya, H.D., Bajaj, D., Das, S., Kumar, V., Gowda, C.L., Sharma, S., Tyagi, A.K. and Parida, S.K. (2016) Genetic dissection of seed-iron and zinc concentrations in chickpea. *Sci. Rep.*, **6**, 24050. https://doi.org/10.1038/srep24050

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L. and Fernie, A.R. (2003) Parallel analysis of transcript and metabolic profiles: A new approach in systems biology. *EMBO Rep.*, **4**, 989–993. https://doi.org/10.1038/sj.embor.embor944

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S. and Provart, N.J. (2009) Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. *Plant, Cell Environ.*, **32**, 1633–1651. https://doi.org/10.1111/j.1365-3040.2009.02040.x

Wang, S., Basten, C.J. and Zeng, Z.B. (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.

Watson, S.A. (1987) Structure and composition. In *Corn: Chemistry and Technology* (Watson, S.A. and Ramstad, P.E., eds.). St. Paul, MN: American Association of Cereal Chemists, Inc., pp. 53–82.

Wen, W., Li, D., Li, X., Gao, Y, *et al.* (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.*, **5**, 3438.

Wen, W., Li, K., Alseek, S. *et al.* (2015) Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell*, **27**, 1839–1856. https://doi.org/10.1105/tpc.15.00208

Wenk, M.R. (2005) The emerging field of lipidomics. *Nat. Rev. Drug Discovery*, **4**, 594–610. https://doi.org/10.1038/nrd1776

Wenk, M.R. (2010) Lipidomics: New tools and applications. *Cell*, **143**, 888–895. https://doi.org/10.1016/j.cell.2010.11.033

Yang, X., Guo, Y., Yan, J., Zhang, J., Song, T., Rocheford, T. and Li, J.S. (2010) Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. *Theoretical and Applied Genetics*, **120**, 665–678. https://doi.org/10.1007/s00122-009-1184-1

Yang, H., Richter, G.L., Wang, X., Młodzińska, E., Carraro, N., Ma, G., Jenness, M., Chao, D.Y., Peer, W.A. and Murphy, A.S. (2013) Sterols and sphingolipids differentially function in trafficking of the Arabidopsis ABCB19 auxin transporter. *Plant Journal*, **74**, 37–47. https://doi.org/10.1111/tpj.12103

Yilmaz, A., Nishiyama, M.Y., Fuentes, B.G., Souza, G.M., Janies, D., Gray, J. and Grotewold, E. (2009) GRASSIUS: A Platform for Comparative Regulatory Genomics across the Grasses. *Plant Physiol.*, **149**, 171–180. https://doi.org/10.1104/pp.108.128579

Zhang, D., Zhang, H., Chu, S., Li, H., Chi, Y., Triebwasser-Freese, D., Lv, H. and Yu, D. (2017) Integrating QTL mapping and transcriptomics identifies candidate genes underlying QTLs associated with soybean tolerance to low-phosphorus stress. *Plant Mol. Biol.*, **93**, 137–150. https://doi.org/10.1007/s11103-016-0552-x

Zheng, P., Allen, W.B., Roesler, K. *et al.* (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.*, **40**, 367–372. https://doi.org/10.1038/ng.85

