Running head: System interactions of Lys metabolism

*Corresponding author:
Gad Galili, 1Department of Plant Sciences
The Weizmann Institute of Science
76100 Rehovot
Israel

Phone: +972-8-9343511,
Fax: +972-8-9344181
E-mail: gad.galili@weizmann.ac.il

Journal research area: Biochemical Processes and Macromolecular Structures
Title:
Deciphering transcriptional and metabolic networks associated with lysine metabolism
during Arabidopsis seed development

Ruthie Angelovici¹#, Aaron Fait²#, Xiaohong Zhu¹, Jedrzej Szymanski³, Ester Feldmesser⁴,
Alisdair R. Fernie³ and Gad Galili¹*

¹Department of Plant Sciences, the Weizmann Institute of Science, 76100 Rehovot, Israel; ²French
Associates Institute for Agriculture & Biotechnology of Drylands, Jacob Blaustein Institute for
Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, 84990, Israel;
³Abteiitung Willmitzer, Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg
1, 1447 Potsdam-Golm, Germany; ⁴Department of Biological Services, the Weizmann Institute of
Science, 76100 Rehovot, Israel.
#The authors contributed equally to this paper

*Corresponding author
This collaborative research was supported by The Israel Academy of Sciences and Humanities, National Council for Research and Development, The German Federal Ministry of Education and Research (BMBF) within the framework of German-Israeli Project Cooperation (DIP), as well as by BARD, the United States - Israel Binational Agricultural Research and Development foundations and the German Israeli Foundation, GIF grant no. 933-239.12/2006.

Corresponding author:
Gad Galili
E-mail: gad.galili@weizmann.ac.il
ABSTRACT

In order to elucidate transcriptional and metabolic networks associated with Lys metabolism, we utilized developing seeds as a system in which Lys synthesis could be stimulated developmentally without application of chemicals and coupled this to a T-DNA insertion knockout mutation impaired in Lys catabolism. This seed-specific metabolic perturbation stimulated Lys accumulation starting from the initiation of storage reserve accumulation. Our results revealed that the response of seed metabolism to the inducible alteration of Lys metabolism was relatively minor, however, that which was observable operated in a modular manner. They also demonstrated that Lys metabolism is strongly associated with the operation of the TCA cycle, whilst largely disconnected from other metabolic networks. In contrast, the inducible alteration of Lys metabolism was strongly associated with gene networks, stimulating the expression of hundreds of genes controlling anabolic processes that are associated with plant performance and vigor, whilst suppressing a small number of genes associated with plant stress interactions. The most pronounced effect of the developmentally-inducible alteration of Lys metabolism was an induction of expression of a large set of genes encoding ribosomal proteins as well as genes encoding translation initiation and elongation factors, all of which are associated with protein synthesis. With respect to metabolic regulation, the inducible alteration of Lys metabolism was primarily associated with altered expression of genes belonging to networks of amino acids and sugar metabolism. The combined data are discussed within the context of network interactions both between and within metabolic and transcriptional control systems.
INTRODUCTION

Metabolism is one of the most important and best recognized networks within biological systems. However, advances in the understanding of metabolic regulation still suffer from insufficient research concerning the modular operation of such networks. Furthermore, elucidation of metabolic regulation within the context of the entire system including transcriptional, translational and post-translational mechanisms is rarely attempted (Sweetlove et al., 2008). Instead, to date, studies on metabolic regulation have mostly been limited to regulatory interactions within the metabolic pathways themselves. These studies have, amongst other things, revealed that amino acid biosynthesis generally occurs through branched pathways, which are considerably more complex than the non-branched pathway of glycolysis (Sweetlove and Fernie, 2005; Sweetlove et al., 2008). Of the pathways associated with plant amino acid metabolism, studies of the branched Asp-family pathway involved in the synthesis of Lys, Met, Thr and Ile as well as the conversion of Thr into Gly (see schematic diagram in Fig. 1) led to a considerable understanding of the intra-pathway regulation of the synthesis of these amino acids as well as the inter-regulation of Asp and Glu metabolism (Karchi et al., 1994; Galili, 2002; Zhu and Galili, 2003; 2004; Joshi et al., 2006; Stepansky et al., 2006; Less and Galili, 2009). Lys metabolism was also shown to be regulated both by the rate of its synthesis and catabolism via the α-amino adipic acid pathway (Zhu and Galili, 2004; Less and Galili, 2009). In addition, two genes associated with response of plants to biotic stresses have been shown to encode enzymes of Lys biosynthesis (AGD2) and likely also catabolism (ALD1) into a yet unknown compound, important for pathogen resistance (Song et al., 2004; Song et al., 2004). Notwithstanding the significant complexity of the pathway of Lys metabolism in plants, remarkably little is known concerning the influence of this pathway on genome-wide patterns of gene expression or of its impact on primary metabolism as a whole.

Developing seeds can serve as an excellent system for studying developmentally controlled metabolic regulation within a broad context, including both metabolic and transcriptional parameters alongside their associated networks (Sweetlove and Fernie, 2005). Seed development initiates by embryogenesis and is followed by a period characterised by highly active sugar, amino acids and lipids metabolism (Weber et al., 2005). During this metabolically active period, developing seeds also induce a massive synthesis of reserve compounds, including storage proteins, starch and oil, whose relative proportions vary between seeds of different plant species (Weber et al., 2005; Gallardo et al., 2008). Furthermore, the synthesis of seed storage proteins is
entirely seed-specific, occurring via a strict seed-specific stimulation of expression of a family of
genes encoding the different groups of storage proteins (Verdier and Thompson, 2008). In fact, the
induction of expression of the storage protein genes is very fast, initiating from non-detectable
levels at the initiation of the period of storage reserve accumulation and reaching massive amounts
of both storage protein mRNAs and storage proteins themselves in a relatively short time periods
of up to a few days (Verdier and Thompson, 2008). This renders developing seeds as an excellent
developmentally inducible system in which to study metabolic regulation.

We have previously generated transgenic Arabidopsis plants expressing a bacterial
feedback insensitive dihydrodipicolinate synthase (DHDPS) of Lys biosynthesis under a seed-
specific storage protein promoter and crossed these transgenic plants with a T-DNA knockout
mutant in the \textit{LKR/SDH} gene of Lys catabolism (Zhu and Galili, 2003). This crossed transgenic
genotype (termed the “KD” genotype) possesses a highly efficient, developmentally induced
accumulation of Lys, initiated at the start of the period of reserve accumulation (Zhu and Galili,
2003). In the present study, we have subjected developing seeds of this KD genotype to detailed
transcriptomics, metabolomics and bioinformatics analyses to decipher how Lys metabolism cross-
interacts with primary metabolism as well as genome-wide gene expression programs. Our results
revealed a strong connection between Lys metabolism and the TCA cycle and also shed new light
on the modular operation of plant metabolism. They also suggest that Lys metabolism has a
relatively small impact on the network of primary metabolism, but in contrast, has surprisingly
wide impact on specific genome-wide gene expression networks.

\textbf{RESULTS}

\textbf{A Seed-Specific Developmentally Inducible System To Decipher Gene Expression And
Metabolic Networks Associated With Lys Metabolism}

In order to elucidate novel network interactions of Lys metabolism with genome-wide gene
expression programs and networks of primary metabolism, we utilized an Arabidopsis mutant
(termed KD) (Zhu and Galili, 2003) displaying a seed-specific, developmentally-induced bacterial
feedback-insensitive DHDPS enzyme of Lys biosynthesis in addition to a T-DNA knockout
mutation in the \textit{LKR/SDH} gene affected in Lys catabolism (Zhu and Galili, 2003). Maturing seeds
of the wild type (WT) and KD genotype showed comparable morphologies across seed
development (Fig. 2A), as well as efficient developmental induction of the bacterial DHDPS from
negligible levels at ~10 days after flowering (DAF) to high expression levels during the later stages of seed maturation (Fig. 2B). Metabolic analysis also showed a developmentally inducible accumulation of Lys from ~10 DAF to mature dry seeds (Fig. 2C), proving the functionality of our genetic system in altering the flux of Lys metabolism.

**Effect Of The Developmentally-Inducible Alteration Of Lys Metabolism On Primary Metabolism During Seed Maturation**

To study the effect of altered Lys metabolism on networks of primary metabolism, maturing seeds of the WT and KD genotype were profiled using an established GC-MS protocol (Roessner et al., 2001), which allowed us to quantify the relative content of over 70 metabolites (see details in the Experimental procedures section). Since environmental differences can have a great impact on metabolite content, GC-MS analysis was performed on two independently grown and sampled seed batches. However, although these data confirm the consistency across experiments and the full data set is supplied in Supplementary Tables S1-2), here for the sake of simplicity we present the analysis of a single experiment. In the case of specific metabolites, we focused on those that were significantly different between the WT and KD genotype in both experiments when assessed by a two-way analysis of variance (ANOVA) (for details see Experimental Procedures and Supplementary Table S3) and which displayed similar patterns of change in both experiments. When analyzing the global trends of changes of metabolites in developing seeds by principal component analysis (PCA), the WT and KD genotypes exhibited comparable changes in the metabolic profile, implying a relatively small effect of the genetic manipulations exhibited by KD on primary metabolism (Supplementary Fig. 1). Interestingly, we observed a major switch in metabolism during desiccation in both genotypes (transition from 16 DAF to dry seeds; Supplementary Fig. 1), which is in accordance with our previous report (Fait et al., 2006). These lines of evidence further demonstrate that the KD seeds exhibit mainly a WT-like metabolism throughout seed development.

In order to gain a higher resolution of the metabolic changes than that available following global PCA analysis, we next tested the effect of the KD mutation on the levels of the independent metabolites detected by the GC-MS analysis. Surprisingly, as shown in Fig. 3 A-D, the two way ANOVA analysis indicated that among the ~70 metabolites identified in the GC-MS analysis, the levels of only four metabolites were consistently different between genotypes across seed
maturation, implying a relatively minor interaction of Lys metabolism with primary metabolism. Interestingly, three of these metabolites (fumarate, citrate and 2-oxoglutarate) (Fig. 3 A-C) are TCA cycle intermediates, suggesting a strong interaction of Lys metabolism with mitochondrial energy metabolism.

**Effect Of The Developmental-Inducible Alteration Of Lys Metabolism On Metabolic Networks In Seeds**

Despite the importance of metabolism in the regulation of plant growth, the structure and dynamics of the metabolic networks in plants have received relatively little attention to date with only a handful of studies focusing on this aspect (Weckwerth et al., 2004; Schauer et al., 2006; Schauer et al., 2008; Sulpice et al., 2009). The mode of interaction between metabolites can provide important insight into the modular behavior of biochemical processes and their regulation (Camacho et al., 2005; Steuer et al., 2006). We were therefore interested to study whether the perturbation of Lys metabolism influences the association of individual metabolites and/or of groups of metabolites within developing seeds. Examination of the metabolite correlation networks was performed exactly as described in Experimental Procedures following the method of Nikiforova and co-workers (Nikiforova et al., 2005). A matrix of Spearman correlation coefficients was independently constructed for each genotype and was subsequently filtered using common significance threshold of $10^{-9}$. A strict threshold was chosen because of highly synergistic behavior of all metabolites in the analyzed datasets. The obtained adjacency matrix was used to create a so-called community-based network, in which metabolites are represented as nodes and significant correlations as lines (edges) that connect the nodes (detailed in Experimental procedures and Newman, 2006). We first built the metabolic network of the WT genotype in which we identified the number and composition of groups (communities), of associated metabolites (Fig. 4A, five circles of metabolites having different colors and Table 1, columns 1-5). Network analysis emphasized links both between and within the various compound classes of metabolites detected. For example, the amino acids measured in the WT genotype split into three separate communities, of which only one consists exclusively of amino acids although it is important to note that the amino acids in this community are not necessary synthesized from the same metabolic pathway (Table 1). Several sugars and organic acids grouped to the same community (community no.5, Table 1), whilst fatty acids grouped to a separate community (community no.3, Table 1). TCA
cycle intermediates were, however, scattered across the different communities, although this may well reflect the complexity of their subcellular compartmentation (Winter et al., 1993; Martinoia and Rentsch, 1994). We next constructed the metabolic network of the KD genotype in the exact same way (Fig. 4B). Interestingly, the WT and KD genotypes differed in the metabolites comprising each community (Table 1 & 2, Fig. 4). For example, the amino acids that composed community-2 in the WT, grouped in the KD genotype into two different communities, community-2 and community-4, which consisted not only amino acids but also sugars, sugar phosphates and organic acids.

In order to highlight both commonalities and diversities in network topologies of the genotypes (Fig. 4 panels A, B), we also merged the networks of the WT and KD genotype (Fig. 4 panel C). In the resultant union network, different node colors represent the different WT communities (as displayed in Fig. 4 panel A), whilst the graphic grouping of the nodes into different circles represent the different KD communities (Fig. 4 panel B). This graphic reveals considerable scattering of nodes of identical colors across the combined network of panel C indicating the degree of network perturbation caused by Lys accumulation in the KD mutation during seed development. Although the major structure of metabolite-metabolite associations was conserved in the two genotypes (indeed a highly significant overlap between mutant and wild type network was confirmed by application of the Fisher exact test at a $p$-value 0.00001), the KD genotype revealed markedly higher connectivity than the WT (see correlation matrices visualized in Fig. 4). Comparative network analysis also suggested that the effect of the KD genotype on the associations between metabolites within the different communities was largely modular. Namely, specific metabolites formed more robust associations by preserving their links, but grouping linked metabolites into different communities in the two genotypes (comparison between Fig. 4 panels A and B in Table 2). These groups can therefore be regarded as basic metabolic modules (Table 2 Fig. 4B). As would be expected, some modules, contained metabolites belonging to the same biochemical pathway; for example module 6 contains sugars and sugar alcohols and module 1 contains intermediates in fatty acid biosynthesis. Surprisingly, among the TCA cycle metabolites, only succinate and 2-oxoglutarate maintained their associations between genotypes, whilst other TCA cycle intermediates such as GABA, citrate and fumarate did not display modular behaviour. This is likely to, at least in part, be due to the complex subcellular distribution of the GABA, citrate and fumarate. Within the amino acids, two different modules could be observed: the first
included Asn, Ile and Ser (Table 2, module 5), whilst the second contained Phe, Gly and Glu (Table 2, module 8). Thus, while the KD genotype neither displays a prominent effect on the levels of the different metabolites detected in the GC-MS analysis, nor on the global structure of the metabolic network of developing seeds, it does change specific associations between metabolites or modules thereof. Finally, the alteration of Lys content in the KD genotype resulted in the formation of Lys containing associations. For example Lys strongly associated with Ala, fatty acids and Gln in the KD genotype.

Effect Of The Developmentally-Inducible Alteration Of Lys Metabolism On Gene Expression Networks In Developing Seeds

We were also interested in studying the regulatory role of Lys metabolism in the context of the entire system, namely, elucidating the inducible effect of altered Lys metabolism in the KD genotype on networks of gene expression. For this purpose, we performed microarray analysis of developing seeds (stages 14 DAF, 16 DAF and dry seeds) of the WT and KD genotype, using the Affymetrix AtH1 chips. ANOVA analysis of the microarray results indicated that the expression of ~1400 genes showed a significant and at least 2-fold change between the WT and KD genotype in at least one time-point during seed development, with a false discovery rate (FDR) < 0.05 (Supplementary Table S4). We further identified the genes that were consistently up- (650) and down regulated (507) across all time points and subjected them to over-representation analysis using the tools embedded in the PageMan and MapMan software tools (http://mapman.mpimp-golm.mpg.de/general/ora/ora.shtml (Usadel et al., 2006) For a detailed prob-set list, and calculated ratios see Supplemental Tables S7 The major effects of the KD mutation included a stimulation of the expression of ~300 genes belonging to functional categories associated with anabolic processes and plant vigor (Table 3; Fig. 5 and Fig. 6). Amongst these are: (i) 221 genes associated with protein synthesis (including 149 genes encoding ribosomal proteins and 20 genes encoding translation initiation elongation factors); (ii) 25 genes encoding enzymes associated with amino acid metabolism, from which six genes interestingly belong to the Asp-family pathway that synthesizes Lys (AT3G25900 HMT-1; AT4G01850 SAM-2; AT2G45440 DHDPS2; AT2G59890 Dihydrodipicolinate reductase; AT5G13280 AK-LYS1; AT1G14810 Asp Semialdehyde dehydrogenase); (iii) 12 genes encoding enzymes associated with nucleotide metabolism; (iv) six methyl transferase genes; and (v) 20 genes associated with RNA processing.
In addition to stimulating the expression of genes associated with the above processes, the KD genotype also seems to have mixed effect (induction/reduction) on small sets of genes involved in several metabolic processes, as well as those associated with RNA transcription and processing, posttranslational modifications and protein turnover (Table 3; Fig. 5 and 6).

In contrast to the principal stimulatory effect of the KD genotype on genes associated with anabolic processes thought to be related to plant performance, this mutant had an inhibitory effect on the expression of a very small number of genes, particularly those associated with sugar synthesis (10 genes), amino acid degradation (3 genes), heat shock stress (14 genes, including a gene encoding a novel heat shock transcription factor; AT3G24520), DNA synthesis and chromatin structure (8 genes) and transcriptional regulation (3 genes). Given the enriched down-regulation of heat shock genes in the KD genotype (Table 3), we decided to focus further on the specific response to biotic and abiotic stress related transcripts and visualized the KD response by the MAPMAN tool (Fig. 7). In doing so we observed a down-regulation of the expression of genes associated with stress; however it is important to note that only heat stress response genes were significantly over-represented within the dataset (Table 3).

To test if the major reduction in certain TCA cycle intermediates observed in the KD genotype was limited to the metabolic level, we also searched for specific alteration in the expression of genes associated with the TCA cycle. The expression of a number of TCA cycle associated genes were affected in the KD genotype, of which two particularly interesting cases were the upregulation of succinate dehydrogenase 2 (AT3G2738) and alternative oxidase1A (AT3G22370). The former gene product converts succinate to fumarate and in doing so supplies electrons to the mitochondrial electron transport chain. In contrast, the later is often associated with inhibition of mitochondrial electron transport chain. However, it is important to note that the in vivo activity of alternative oxidise is most likely not controlled at the transcriptional but rather at the post-transcriptional influences (Day et al., 1994; Oliver et al., 2008; Rasmusson et al., 2009). Thus, taken together, it is apparent from our microarray analysis that the induced stimulation of Lys biosynthesis, coupled with a down-regulation of its degradation principally stimulates a large number of genes associated with plant vigor, whilst simultaneously suppressing a limited number of genes associated with plant response to stress, particularly heat stress. That said, our examination of whether the KD genotype exhibited a major influence on the proteome of mature dry seeds suggests that the proteomes of the two genotypes were highly comparable.
We also wished to test whether the KD genotype also alters the transcription patterns during seed maturation, compared to WT genotype. To address this issue, we clustered the genes that were consistently up- (650) or down regulated (507) in the KD genotype during seed maturation of both WT and KD genotype (three time points for each genotype; total six time points). Clustering analysis was performed, using the Expander tool (see Methods) as follows: (i) the six time points were clustered together with the three WT time points being on the left and the three KD time points being on the right, thus generating a single cluster of six points. This connected trend line has no biological relevance, but was used in order to enable us to compare both the pattern and the relative intensity of the same genes in WT and KD genotype; and (ii) the artificial line connecting the dry WT seeds to the 14DAF KD seed was eliminated and then we showed the cluster trend lines of the WT and the KD separately. As shown in Fig. 8 the genes data set grouped into four different clusters (panels A-D). In each pane, the trend lines of the WT and the KD genotype appear in the left and right, respectively. Interestingly, the mRNA expression patterns during seed maturation were quite comparable between the two genotypes. Next, we subjected the genes lists from the four different clusters to an over-representation analysis using the tools embedded in the PageMan and MapMan softwares (http://mapman.mpimp-mpg.de/general/ora/ora.shtml) (Usadel et al., 2006) in order to elucidate the expression pattern of the functional categories identified in Table 2.

Genes that were associated with amino acid metabolism, signalling processes and nucleotide synthesis were over represented in cluster-1. Genes associated with protein synthesis, including the ribosomal proteins, were over represented in clusters -1 and 4. Genes associated with heat shock were over represented in cluster -3. Genes associated with starch degradation were over represented in cluster -2 (Fig 8 A-D).

Does the high Lys accumulation in the KD genotype influences the expression of EF1α whose expression correlates with Lys level in maize high-Lys opaque mutants?

Lys is the most limiting essential amino acid in cereal crops and hence improving its content in seeds represent a major biotechnological challenge (Ufaz and Galili, 2008). Previous studies in maize, using a variety of high-Lys, Opaque and quality protein maize (QPM) genotypes, showed a strong correlation between seed Lys content and the expression level of the translation elongation
factor EF1α (Habben et al., 1995). Hence, even though distinct methodologies were used to improve Lys content in the opaque/QPM maize genotypes and in the Arabidopsis KD genotype, it was interesting to test whether the expression of the Arabidopsis EF1α homologues are upregulated in the high Lys KD genotype, compared to the WT genotypes. Notably, expression of the two Arabidopsis EF1α genes (AT1G07940 and AT2G38900) was comparable between the WT and KD genotypes, but interestingly, the expression level of two genes encoding other putative translational elongation factors (AT1G09640 and AT2G31060) were higher in the KD compared to the WT genotypes (Supplementary Table IV). We also tested the expression levels of nearly 20 Arabidopsis genes, known from the literature to encode Lys-rich proteins, in the WT and KD genotypes. We found that the expression level of all of these genes was not different between the WT and KD genotype (data not shown).

DISCUSSION

Suitability Of Maturing Seeds As A Developmentally-Inducible System To Study Gene Expression And Metabolic Networks That Interact With Lys Metabolism

In the present report we used maturing Arabidopsis seeds as a developmentally-inducible system to alter the metabolic flux of Lys metabolism, as a tool to investigate systems-wide interactions of Lys metabolism with the primary metabolome as well as with genome-wide transcriptional programs. As opposed to chemical-based inducible systems –such as the ethanol- and dexamethasone-inducible systems (Aoyama and Chua, 1997; Junker et al., 2003) - maturing seeds offer a chemical-free, rapid and efficient developmentally-inducible system that is based on a promoter of a storage protein gene. Our specific inducible alteration of Lys metabolism was based on a developmental-inducible stimulation of Lys biosynthesis by expressing a bacterial gene encoding feedback-insensitive DHDPS of Lys biosynthesis under the control of the phaseolin storage protein promoter in the background of a T-DNA knockout mutant in the LKR/SDH gene of Lys catabolism. The rapid induction of the bacterial DHDPS gene and the inducible accumulation of Lys metabolism across seed maturation are shown in Fig. 1 panels B and C, respectively.
Modularity Of The Metabolome Network Of Developing Seeds In Response To The Inducible Alteration Of Lys Metabolism

Interestingly, the inducible alteration of Lys metabolism in the KD genotype had no major effect on the levels of other primary metabolites, causing only a pronounced reduction in the levels of the TCA cycle intermediates fumarate and citrate. The reason for the major suppressive effect of the KD genotype on the TCA-cycle metabolites is yet unknown, but may be due to a competition of the Lys biosynthesis pathway with the TCA cycle on their common precursor metabolite pyruvate because pyruvate is also used as a co-substrate for the DHDPS enzyme of Lys metabolism (Fig. 1). An even more interesting observation was seen in the adjustment of the metabolic network of maturing seeds to the inducible alteration of Lys metabolism, which offers considerable insight into the operation of this network in maturing seeds. Our metabolite network analysis revealed that the altered Lys metabolism generally increases the number of significant correlations between metabolites in respect to the dynamic changes in their levels during seed maturation. In addition, as shown in Table 1 and Table 2, the alteration in Lys metabolism also modifies the degree of association between metabolites, resulting in their grouping within slightly different communities as compared to the WT. However, this reorganization of metabolic associations involved not only single metabolites, but also basic metabolic modules that were conserved between the WT and KD genotypes (Table 2). Examples of the modular organization of the metabolic network in response to the KD mutation included the TCA cycle intermediates 2-OG and succinate, but not fumarate, malate or citrate in both the WT and KD genotypes. This finding implies a complex modular regulation of different parts of this metabolic cycle and as such supports the contrasting phenotypes observed on perturbing various reactions involved in the TCA cycle in plants (Nunes-Nesi et al., 2005; Nunes-Nesi et al., 2007). The suggestion of such a modular function of the TCA cycle is also not without precedence, with such functional activities being commonly described in microbes (see for example (McCammon et al., 2003). Combination of results from these studies should provide the evidence required to initiate further studies on the modularity of metabolic networks in plants.

Interestingly, the KD genotype also displayed a dispersed association of amino acids, which were synthesized by different branches of primary metabolic networks (Table 2). These results imply that network associations of amino acids are not only due to their function in protein synthesis, but also due to their regulatory function in plant metabolism. An interesting example implying such
novel associations of amino acids is the observation that in both the WT and KD genotype, trehalose is more strongly associated with Met than with trehalose 6p (Fig. 4). Other interesting observations are the new association groups that are formed in response to the altered Lys metabolism. Three examples are: (i) the association of Met with Ile in the KD genotype (Table 2) may signify an adjusted operation of the newly discovered channeling of Met via Met γ-lyase into Ile (Re'beille' et al., 2006; Joshi and Jander, 2009); (ii) the association of Glu with GABA in the KD genotype may signify an adjustment of the established conversion of Glu into GABA (ref review Aaron); and (iii) the association of Gly and Ser in the WT, but not in the KD genotype, may imply a modified operation of photorespiration. Thus, our results imply that the metabolic network-association approach, coupled with the analysis metabolic mutant, may provide a novel insight into re-routing of metabolic pathways that could be further validated by metabolic flux analyses.

Response Of The Transcriptome Network Of Developing Seeds To The Inducible Alteration Of Lys Metabolism

Our results showed that the inducible alteration of Lys metabolism in the KD genotype had a significant effect on gene expression programs in the developing seeds. The most significant effect of the KD genotype included a stimulatory effect on a large array of genes principally controlling anabolic processes that are associated with plant vigor (Table 3), namely processes associated with active plant growth when growth conditions are usually optimal. In addition, the KD genotype also suppressed the expression of a smaller group of genes principally associated with amino acid degradation and response to stress conditions (particularly heat shock). Thus, our results imply that the Lys metabolism pathway participates in the regulation of plant growth and response to stress through modulating networks of gene expression. If true, this observation begs the question; How does such a system interaction operate? Active plant growth is principally associated with up-regulated expression of genes encoding enzymes of amino acid biosynthesis, which are particularly necessary for protein synthesis. In contrast, our recent bioinformatics analysis (Less and Galili, 2009), revealed that exposure to various stresses slightly suppresses the expression of some genes encoding biosynthetic enzymes of Lys biosynthesis, while significantly stimulating the LKR/SDH gene of Lys catabolism. Thus, we hypothesize that changes in either the flux of Lys metabolism and/or the level of Lys itself are recognized by the plant as signal[s] regulating gene expression.
This hypothesis is supported by two additional independent lines of evidence: (i) a mutation increasing the resistance of Arabidopsis plants to pathogen infection and also simulating the expression of a number of genes associated with pathogen resistance (termed the “agd2” mutant) has been shown to occur in a gene encoding an transaminase enzyme of Lys biosynthesis (Hudson et al., 2006); and (ii) a T-DNA knockout mutation in another Arabidopsis gene (termed the ALD1 gene), which is highly homologous to the AGD2 gene of Lys biosynthesis and was also shown to encode an enzyme that in vitro transaminates Lys into a yet unknown metabolite, was further shown to render Arabidopsis plants more sensitive to pathogens and also to affect the expression of various stress-associated genes in a different manner to that of the AGD2 gene (Song et al., 2004; Song et al., 2004).

Interestingly, as opposed to the major effect of the KD genotype on global gene expression program, metabolic engineering approaches generating rice Trp overproducing plants as well as Arabidopsis Phe overproducing plants resulted in only small effects on global gene expression programs (Dubouzet et al., 2007; Tzin et al., 2009). This illustrates that different amino acid metabolic networks may possess variable interactions with gene expression programs.

CONCLUSIONS

The present report describes the first example of an approach to use seeds as a developmentally-inducible system to study metabolic regulation in the contest of a comprehensive system that includes interactions with transcriptional networks. Our results expose novel metabolic and transcriptional network interactions associated with Lys metabolism and illustrate the modular behaviour of these interactions. They additionally, demonstrate that Lys metabolism is relatively inert with respect to its general impact on metabolic regulation, which is primarily associated with TCA cycle metabolites. This comparison provides an important advance in the understanding of the regulation of plant metabolism. Whilst much recent experimental and theoretical research effort has been focussed on understanding the intra-regulation of plant amino acid metabolism (Hare et al., 1999; Galili, 2002); (Curien et al., 2009) attempts to evaluate the role of amino acid metabolism with respect to primary metabolism has rarely been directly assessed. The results presented here are in accordance with general theories of metabolic robustness (Stephanopoulos and Vallino, 1991; Rontein et al., 2002), in that they suggest that, at least under the optimal conditions tested here, the alterations in Lys metabolism exhibited in the KD genotype did not
greatly influence metabolite partitioning between the genotypes. Yet, Lys metabolism did display a more pronounced influence in the context of the entire system, affecting the expression of large set of genes. The increased Lys accumulation in the KD genotype strongly stimulates gene expression programs associated with plant vigor, while suppressing a smaller number of genes principally associated with stress (heat shock) response. These results, thus support our previous bioinformatics analysis (Less and Galili, 2009), in revealing that the expression of genes encoding biosynthetic enzymes of the Asp-family network is principally co-regulated with genome-wide genes associated with plant growth, while the expression of genes encoding catabolic enzymes of this network are principally co-regulated with stress-associated genes.

MATERIAL AND METHODS

Chemicals.

All chemicals were purchased from Sigma-Aldrich Israel Ltd. (Jerusalem, Israel) with the exception of N-methyl-N-[trimethylsilyl] trifluoroacetamide (Macherey-Nagel GmbH & Co. KG, Düren, Germany).

Plant Growth And Seed Collection

*Arabidopsis thaliana* (WS) seeds of the WT and KD genotype (Zhu and Galili, 2003) were germinated on soil and grown for two consecutive rounds in the greenhouse (21°C). Flowers were marked and at given time intervals following flowering (10±1 to 18±1 DAF). Maturing siliques were collected, lyophilized to dryness and seeds were then dissected from the dried siliques and immediately frozen in liquid nitrogen. Mature seeds were collected at the end of the desiccation period and stored at 4°C. Three to five thousands seeds were harvested for each extraction.

Extraction, Derivatization And Analysis Of Seeds Metabolites Using GC-MS

Collected material was extracted and analyzed by GC-MS was carried out using a GC-TOF protocol optimized for Arabidopsis (Lisec et al., 2006) by adjusting the extraction protocol to seed material as described in Fait et al., (2006). Relative metabolite content was calculated as described in Roessner et al. (2001) following peak identification using TagFinder (Luedemann et al., 2008). Substances were identified by comparison to mass spectral tags represented in our in-house
database (Schauer et al., 2005; Erban et al., 2007; Hummel et al., 2007).

Statistical And Network Analysis

Statistical analysis, PCA and Two Way ANOVA, was performed on the data sets obtained from metabolite profiling with the software package TMEV (Saeed et al., 2003) using the default weighted covariance-estimation function. Prior to the analysis data were Log transformed and normalized to the median of the entire sample set for each metabolite. This transformation reduces the impact of outliers without affecting the discrimination within the data set. Furthermore to test statistical significance between specific time points t-tests were performed using the algorithm incorporated into Microsoft excel (Microsoft Corporation, Seattle, WA, USA) with Bonferroni correction of the critical p-value for multiple comparisons. Networks of metabolite correlations were reconstructed based on the method described by Nikiforova et al. (2005). From the entire set of metabolites response we first produced a correlation matrix based on the calculation of the Pearson product moment correlation coefficient between each pair of metabolites across the whole developmental period. To this aim we unified both batches in order to increase the robustness of the analysis and identify those correlations shared by both batches. We then calculate the significance of each correlation coefficient. We set the correlation significance threshold fixed at the level of Spearman correlation coefficient \( p-value < 10^{-10} \) Next we visualize the correlation matrix as a graph, wherein each metabolite is a vertex (or a node) and each significant correlation between two metabolites is a line between two vertices. Resulting graph was analyzed in context of its topology. A cumulative degree distribution and network modularity were investigated. The former defines a hierarchy in nodes connectivity, whereas the latter informs about presence of highly connected clusters. Newman’s algorithm based on matrix eigenvectors was used to detect a community structure (Newman, 2006), leading to partition of the network into communities of vertices characterized by a higher number of ties and higher structural equivalence within the community as compared to the their ties outside the community. These clusters, called communities contain metabolites, which are structurally equivalent and thus putatively functionally associated. Structural equivalence is a property, which can be scaled based on the number and nature of ties between metabolites. A subset of vertices is defined as approaching structurally equivalence when they share ties with identical vertices. To emphasize the changes in the structural feature of the merged metabolic networks we visualized the communities built from
the WT correlation matrix in different colours, whilst communities built in the mutant are shown as circles. Networks were visualized and merged by using the open source bioinformatics software platform for visualizing molecular interaction networks Cytoscape software 2.6.1, all the calculations and network analysis were performed using R 2.8.0 software (Gentleman et al., 2004)

**RNA Extraction and Microarray Analysis**

All experiments analyzing RNA expression levels were carried out using two replicates of seed materials, obtained from two independent seed lots, from plants grown under controlled-environment conditions. Fresh developing seeds at 10, 14 and 16 DAF were harvested and total RNA was extracted as previously described (Ruuska and Ohlrogge, 2001) as well as for dry seeds, imbibed seeds and germinating seeds as described above (seed germination). Total RNA was treated with DNAase RQ-1 (Promega) then RNA was amplified using 2-cycle Affimexrix labelling using the standard Affimexrix protocol. Hybridization, labelling, scanning and data extraction were performed according to the standard Affimexrix protocols. Transcriptome analysis was carried out using Partek Genome Suite software (Partek) (www.partek.com). Pre-processing was carried out using the Robust Microarray Averaging (RMA) algorithm (Irizarry et al., 2003) Two way ANOVA analysis was performed. Batch effects between the two duplicate experiments were removed. False discovery rate (FDR) was applied to correct form multiple comparisons (Hochberg and Benjamini, 1990) Differentially expressed genes were chosen according to FDR < 0.05 and a twofold change between genotypes at least in one time-point. Over/under representation analysis was performed by PageMan tool (http://mapman.mpimp-golm.mpg.de/general/ora/ora.shtml, Usadel B et al. (2006). Visualization of metabolic pathways and other functional categories was performed with the MapMan (Usadel et al., 2005) software tool. Clustering analysis was performed by Expander software (Shamir et al., 2005), http://acgt.cs.tau.ac.il/expander/) using the click algorithm.

**Protein Analysis:**

For two-dimensional IEF/SDS PAGE fractionation, 50 mg seeds were extracted as previously described (Ruuska and Ohlrogge, 2001), and proteins were then extracted from the phenol phase. IEF-SDS PAGE was then performed as previously described (Liska et al., 2004) and stained with coomassie blue.
Accession Numbers

Microarray data have been deposited to the Gene Expression Omnibus database under series number GSE18112.

Supplemental Data:

Supplemental Table S1: The entire dataset of the relative contents of metabolites in growing time 1.
Supplemental Table S2: The entire dataset of the relative content of metabolites in growing time 2.
Supplemental Table S3: The entire dataset of the p-values of the Two Way ANOVA of the data sets from the different seed growing times.
Supplemental Table S4: Expression levels of genes whose mRNA levels were significantly different between WT and KD during seeds maturation (data is presented in log₂ values).
Supplemental Table S5: An elaboration of the overrepresentation analysis of genes whose mRNA levels were significantly different between WT and KD during seeds maturation.
Supplemental Table S6: Complete microarray results of expression levels in the WT and KD seeds during seed maturation (data is presented in log₂ values).
Supplemental Table S7: Calculated ratios of transcript levels of KD vs WT genes whose mRNA levels were significantly different between WT and KD seeds
Supplemental Table S8: Abbreviation of the network metabolites of Fig. 4.
Supplemental Figure S1: Principal-component analysis (PCA) of metabolite profiles of distinct maturation stages of seeds of the WT and KD genotype.
Supplemental Figure S2: Effect of altered Lys metabolism in the KD genotype on the levels and deposition of the 12S cruciferin storage proteins.

ACKNOWLEDGEMENTS

We thank Prof. Dominique Job for helpful suggestions with two-dimensional PAGE. We also thank Dr. Alexander Erban for consultation on GC-MS analyses as well as Sarit Brog and Limor
Frid for excellent technical assistance. GG is an incumbent of the Bronfman Chair of Plant Science at the Weizmann Institute of Science. AF acknowledges the support of the Alexander von Humboldt foundation.
References:

Aoyama T, Chua N (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. The Plant Journal 11: 605-612

Camacho A, Rochera C, Silvestre JJ, Vicente E, Hahn MW (2005) Spatial dominance and inorganic carbon assimilation by conspicuous autotrophic biofilms in a physical and chemical gradient of a cold sulfurous spring: the role of differential ecological strategies. Microb Ecol 50: 172-184

Curien G, Bastien O, Robert-Genthon M, Cornish-Bowden A, Cardenas ML, Dumas R (2009) Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. Mol Syst Biol 5: 271

Day DA, Millar AH, Wiskich JT, Whelan J (1994) Regulation of Alternative Oxidase Activity by Pyruvate in Soybean Mitochondria. Plant Physiol. 106: 1421-1427

Dubouzet JG, Ishihara A, Matsuda F, Miyagawa H, Iwata H, Wakasa K (2007) Integrated metabolomic and transcriptomic analyses of high-tryptophan rice expressing a mutant anthranilate synthase alpha subunit. J Exp Bot 58: 3309-3321

Erban A, Schauer N, Fernie AR, Kopka J (2007) Nonsupervised construction and application of mass spectral and retention time index libraries from time-of-flight gas chromatography-mass spectrometry metabolite profiles. Methods Mol Biol 358: 19-38

Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol. 142: 839-854

Galili G (2002) New insights into the regulation and functional significance of lysine metabolism in plants. Annu Rev Plant Biol 53: 27-43

Gallardo K, Thompson R, Burstin J (2008) Reserve accumulation in legume seeds. C R Biol 331: 755-762

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80

Habben JE, Moro GL, Hunter BG, Hamaker BR, Larkins BA (1995) Elongation factor 1 alpha concentration is highly correlated with the lysine content of maize endosperm. Proc Natl Acad Sci U S A 92: 8640-8644

Hare P, Cress W, van Staden J (1999) Review article. Proline synthesis and degradation: a model system for elucidating stress-related signal transduction. J. Exp. Bot. 50: 413-434

Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Stat Med 9: 811-818

Hudson AO, Singh BK, Leustek T, Gilvarg C (2006) An LL-diaminopimelate aminotransferase defines a novel variant of the lysine biosynthesis pathway in plants. Plant Physiol 140: 292-301

Hummel J, Selbig J, Walthier D, Kopka J (2007) The Golm Metabolome Database: a database for GC-MS based metabolite profiling. In metabolomics, Vol Volume 18. Springer-Verlag Berline, Berline, Germany, pp 75-95

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264

Joshi V, Jander G (2009) Arabidopsis thaliana methionine gamma-lyase is regulated according to
isoleucine biosynthesis needs, but plays a subordinate role to threonine deaminase. Plant Physiol: (ahead of print)

Joshi V, Laubengayer KM, Schauer N, Fernie AR, Jander G (2006) Two Arabidopsis threonine aldolases are nonredundant and compete with threonine deaminase for a common substrate pool. Plant Cell 18: 3564-3575

Junker BrH, Chu C, Sonnewald U, Willmitzer L, Fernie AR (2003) In plants the alc gene expression system responds more rapidly following induction with acetaldehyde than with ethanol. FEBS letters 535: 136-140

Karchi H, Shaul O, Galli G (1994) Lysine synthesis and catabolism are coordinately regulated during tobacco seed development. Proc Natl Acad Sci U S A 91: 2577-2581

Less H, Galli G (2009) Coordinations between gene modules control the operation of plant amino acid metabolic networks. BMC Syst Biol 3: 14

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat Protoc 1: 387-396

Liska AJ, Shevchenko A, Pick U, Katz A (2004) Enhanced photosynthesis and redox energy production contribute to salinity tolerance in Dunaliella as revealed by homology-based proteomics. Plant Physiol. 136: 2806-2817

Luedemann A, Strassburg K, Erban A, Kopka J (2008) TagFinder for the quantitative analysis of gas chromatography–mass spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics 24: 732-737

Martinoia E, Rentsch D (1994) Malate Compartmentation - Responses to a Complex Metabolism. Annual Review of Plant Physiology and Plant Molecular Biology 45: 447-467

McCannmon MT, Epstein CB, Przybyla-Zawisla B, McAlister-Henn L, Butow RA (2003) Global Transcription Analysis of Krebs Tricarboxylic Acid Cycle Mutants Reveals an Alternating Pattern of Gene Expression and Effects on Hypoxic and Oxidative Genes. Mol. Biol. Cell 14: 958-972

Newman ME (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E Stat Nonlin Soft Matter Phys 74: 036104

Nikforova VJ, Daub CO, Hesse H, Willmitzer L, Hoefgen R (2005) Integrative gene-metabolite network with implemented causality deciphers informational fluxes of sulphur stress response. J Exp Bot 56: 1887-1896

Nunes-Nesi A, Carrari F, Lytovchenko A, Smith AMO, Ehlers Loureiro M, Ratcliffe RG, Sweetlove LJ, Fernie AR (2005) Enhanced Photosynthetic Performance and Growth as a Consequence of Decreasing Mitochondrial Malate Dehydrogenase Activity in Transgenic Tomato Plants. Plant Physiol. 137: 611-622

Nunes-Nesi A, Fernando Carrari, Yves Gibon, Ronan Sulpice, Anna Lytovchenko, Joachim Fisahn, James Graham, R. George Ratcliffe, Lee J, Sweetlove, Alisdair R. Fernie (2007) Deficiency of mitochondrial fumarase activity in tomato plants impairs photosynthesis via an effect on stomatal function. The Plant Journal 50: 1093-1106

Oliver SN, Lunn JE, Urbanczyk-Wochniak E, Lytovchenko A, van Dongen JT, Faix B, Schmalzin E, Fernie AR, Geigenberger P (2008) Decreased Expression of Cytosolic Pyruvate Kinase in Potato Tubers Leads to a Decline in Pyruvate Resulting in an in vivo Repression of the Alternative Oxidase. Plant Physiol. 148: 1640-1654

Rasmusson GA, Fernie AR, van Dongen JT (2009) Alternative oxidase: a defence against metabolic fluctuations? Physiologia Plantarum 9999

Re‘beille’ F, Jabrin S, Bligny R, Loizeau K, Gambonnet B, Van Wilder V, Douce R, Ravanel Sp (2006) Methionine catabolism in Arabidopsis cells is initiated by a ‘^-cleavage process and leads to S-methylcysteine and isoleucine synthases. Proceedings of the National Academy of Sciences 103: 15687-15692
Ufaz S, Galili G (2008) Improving the content of essential amino acids in crop plants: goals and opportunities. Plant Physiol 147: 954-961

Usadel B, Nagel A, Steinhauser D, Gibon Y, Blasing OE, Redestig H, Sreenivasulu N, Krall L, Hannah MA, Poree F, Fernie AR, Stitt M (2006) PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics 7: 535

Usadel B, Nagel A, Thimm O, Redestig H, Blasing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D, Scheible WR, Gibon Y, Morcuende R, Weigt D, Meyer S, Stitt M (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. Plant Physiol 138: 1195-1204

Verdier J, Thompson RD (2008) Transcriptional regulation of storage protein synthesis during dicotyledon seed filling. Plant Cell Physiol 49: 1263-1271

Weber H, Borisjuk L, Wobus U (2005) Molecular physiology of legume seed development. Annu. Rev. Plant Biol. 56: 253-279

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. Proceedings of the National Academy of Sciences of the United States of America 101: 7809-7814

Winter H, Robinson DG, Heldt HW (1993) Subcellular volumes and metabolite concentrations in barley leaves. Planta 191: 180-190

Zhu X, Galili G (2003) Increased lysine synthesis coupled with a knockout of its catabolism synergistically boosts lysine content and also trans regulates the metabolism of other amino acids in arabidopsis seeds. Plant Cell 15: 845-853

Zhu X, Galili G (2004) Lysine metabolism is concurrently regulated by synthesis and catabolism in both reproductive and vegetative tissues. Plant Physiol 135: 129-136
FIGURE LEGENDS

**Figure 1.** A schematic diagram of the Asp-family pathway. Only several enzymes are illustrated. Broken arrows represent several enzymatic steps. Abbreviations: AAT, Asp aminotransferase; AK, Asp kinase; CGS, Cystathionine γ-synthase; DHDPS, Dihydrodipicolinate synthase; LKR/SDH, Lysine-ketoglutarate reductase/saccharopine dehydrogenase; TS, Threonine synthase.

**Figure 2.** Effects of altered Lys metabolism in the KD genotype on seed maturation and free Lys level. The developmental stages of seed maturation are marked by days after flowering (DAF), while mature dry seeds are marked by “Dry”. (A) Morphology of maturing seeds of the WT (top) and KD (bottom) mutant. (B) Immunoblot analysis with anti bacterial DHDPS antibodies showing the level of expression of the bacterial DHPS during seed maturation in the KD genotype. (C) Relative Lys level in the WT and KD genotype at different stages of seed maturation. Values represent the mean of the response of the metabolite, expressed as peak area normalized to the internal standard ribitol as well as to dry weight. The relative metabolites levels of the WT and KD genotype along the different stages of seed maturation in days after flowering (DAF) are illustrated by white and black histograms respectively. Bars on top of the histograms represent the standard errors of four biological replicates grown together, each derived from 5mg of isolated lyophilized seeds, bulked from at least 10 plants for each time point.

**Figure 3.** Relative contents of the four metabolites whose levels were altered significantly in maturing seeds of the KD genotype, compared to the WT. Comparable results were obtained in both growing times. The names of the metabolites are given in panels A-D. The relative metabolites levels in the WT and KD genotype along the different stages of seed maturation in days after flowering (DAF) are illustrated by white and black histograms respectively. Bars on top of the histograms represent the standard errors. Relative values were analyzed as described in the legend of Fig. 2.

**Figure 4.** Network interactions between different metabolites during seed maturation in the WT and KD genotype. Metabolites were grouped into 5 distinct communities in the WT (A) and 4 distinct communities in the KD genotype (B). The merge of the networks visualization based on their coordinated levels during seed maturation of the KD genotype is represented by different
colours to the different metabolites, each colour representing a defined community in the WT genotype metabolic network (C). The edges (lines) connecting two nodes represent a significant correlation between metabolites, black when it occurs in both genotypes, blue when it occurs in the WT and fails in the KD genotype, and red if it occurs in the KD genotype but is absent in the WT genotype. Metabolite abbreviations are provided in Supplementary Table S8.

**Figure 5.** MapMan RNA-protein synthesis overview maps showing differences in transcript levels between the WT, and the KD genotype in 14 DAF (A); 16 DAF (B) and dry seeds (C). Average transcripts levels were calculated from two independent replicates of Affymetrix AtH1 GeneChips and Fold changes were calculated from the normalized genes expression data of the KD vs WT during 14DAF, 16DAF and dry seeds. The resulting file was loaded into the MapMan Image Annotator module to generate the RNA-protein synthesis overview map (Supplemental Table S7). The normalized expression levels (in Log2) of the genes that exhibit significant change between the WT and the KD genotypes are available in Supplemental Table S4.

**Figure 6.** MapMan metabolism overview maps showing differences in transcript levels between the WT and the KD genotype, in 14 DAF (A); 16 DAF (B) and dry seeds (C). Average transcripts levels were calculated from two independent replicates of Affymetrix AtH1 GeneChips and Fold changes were calculated from the normalized genes expression data of the KD vs WT during 14DAF, 16DAF and dry seeds. The resulting file was loaded into the MapMan Image Annotator module to generate the metabolism overview map (Supplemental Table S7). The normalized expression levels (in Log2) of the genes that exhibit significant change between the WT and the KD genotypes are available in Supplemental Table S4.

**Figure 7.** Abiotic and biotic stress response overview maps showing differences in transcript levels between the WT and the KD genotype, in 14 DAF (A); 16 DAF (B) and dry seeds (C). Average transcripts levels were calculated from two independent replicates of Affymetrix AtH1 GeneChips and Fold changes were calculated from the normalized genes expression data of the KD vs WT during 14DAF, 16DAF and dry seeds. The resulting file was loaded into the MapMan Image Annotator module to generate the cell response overview map (Supplemental Table S7). The normalized expression levels (in Log2) of the genes that exhibit significant change between the WT and the KD genotypes are available in Supplemental Table S4.
Figure 8. Clustering analysis of genes, whose mRNA levels were significantly different between the WT and KD genotype during seed germination. Clustering analysis was performed by comparing the average patterns of the normalized expression levels of identical groups of genes in the WT and KD genotype in maturing seeds, as described in M&M. Bars represent the standard error. Seed maturation stages in days after flowering (DAF) are provided in the bottom.
Table I. Metabolic composition of five communities occurring in the WT genotype. Abbreviation: DiEglicerol-Diethyleneglycol, DHA-Dehydroascorbic acid dimer, FA 16:00-Hexadecanoic acid, FA 18:00-Octadecanoic acid, FA 14:00-Tetradecanoic acid, FA 06:00-2-ethyl-Hexanoic acid; MyoInt-myo-Inositol; pGlu-Pyroglutamic acid.

| community-1       | community-2       | Community-3       | community-4       | community-5       |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| 2 Oxoglutarate    | Glutamate         | DiEglicerol       | Maltose           | Galactose         |
| Succinate         | Asparagine        | FA 06:00          | Laminaribiose     | Fructose          |
| Trehalose 6P      | Isoleucine        | FA 14:00          | Trehalose         | Glucose           |
| DHA               | Glycine           | FA 18:00          | Galactinol        | Ribose            |
| Threonic acid     | Serine            | FA 16:00          | Galactosylglycerol| Xylose            |
| Phenylalanine     | Melibiose         | MyoIno1P          | Tyrrosine         |                   |
| Keflose           | pGlu              | Threonine         |                   |                   |
| Saicin            | Tryptophan        | MyoInt            |                   |                   |
| Celliboise        | Methionine        | Glicerol          |                   |                   |
|                   | Guanosine         | Citrate           |                   |                   |
|                   |                   | Shikimate         |                   |                   |
|                   |                   | Malate            |                   |                   |
|                   |                   | Phosphoric acid   |                   |                   |
|                   |                   | Epicatechin       |                   |                   |
Table II. Metabolite composition of the four communities occurring in the KD genotype. Metabolites that are grouped together in both the WT and the KD are defined as modules and are shown in modules number 1-8. Abbreviation: DiEglicerol-Diethyleneglycol, DHA-Dehydroascorbic acid dimer, FA 16:00- Hexadecanoic acid, FA 18:00- Octadecanoic acid, FA 14:00- Tetradecanoic acid, FA 06:00- 2-ethyl- Hexanoic acid; GABA- 4-amino- Butyric acid MyoInt- myo-Inositol; pGlu- Pyroglutamic acid.

| community-1 module | community-2 module | community-3 module | community-4 module |
|-------------------|-------------------|-------------------|-------------------|
| FA 06:00          | Mallose            | Glactose          | Trehalose 6P      |
| FA 14:00          | Trehalose          | Fructose          | DHA               |
| FA 18:00          | Galactonic acid    | Glucose           | Threonate         |
| FA 16:00          | pGlu               | Ribose            | Phenylalanine     |
| Melibiose         | Methionine         | Xylose            | Glycine           |
| Kestose           | Guanosine          | Threonine         | Glutamate         |
| Salicin           | Fructose 2 Oxoglutarate | Glycemic acid     | Citrate           |
| Tryptophan        | Fructose GABA      | Shikimate         | GABA              |
| Laminaribiose     | Asparagine         | Malate            | Cellbiose         |
| DiEglycol         | Isoleucine         | Phosphoric acid   | Epicatechin       |
| Benzoic acid      | Serine             |                   |                   |
| Sinapate          | GlnA 1,4           |                   |                   |
| Alanine           | Tyrosine           | Galactosylglycerol|                   |
| Glutamine         | Glucose 6P         | Fumarate          |                   |
| GlnA 1,5          | Mallose            | Glycerol          |                   |
| Lysine            | MyoInt1P           |                   |                   |
|                   | Oxalate            |                   |                   |
Table III. Functional categories that are overrepresented in the list of genes that exhibit significant change in expression levels between the WT and the KD, and were consistently up regulated (650) genes or down regulated (507) genes at all time points. Overrepresentation analysis was performed by PageMan tool (see M&M). Only functional categories with more than three genes are shown, the elaborated analysis is available in Supplementary Table S5.

### Overrepresented categories of down regulated genes

| Main category                  | Sub categories                  | No. of elements | P-value       |
|--------------------------------|---------------------------------|-----------------|---------------|
| Major CHO metabolism           | Synthesis                       | 10              | 8.86E-05      |
|                                | Starch                          | 5               | 8.80E-04      |
|                                |                                 | 5               | 1.98E-04      |
| Amino acid metabolism          | Degradation:glycine             | 3               | 8.36E-04      |
| Abiotic Stress-Heat            | -                               | 14              | 6.55E-05      |
| DNA. Synthesis chromatin structure |                               | 8               | 4.83E-04      |
| RNA Regulation of transcription | NIH-like bZIP-related family    | 3               | 1.17E-03      |

### Overrepresented categories of up regulated genes

| Main category                  | Sub categories                  | No. of elements | P-value       |
|--------------------------------|---------------------------------|-----------------|---------------|
| Proteins:                      | Synthesis:                      | 221             | 2.16E-42      |
|                                | Misc. Ribosomal proteins:       | 176             | 6.46E-149     |
|                                | BRIX                            | 149             | 1.70E-156     |
|                                | Mito/Plastid                    | 6               | 5.32E-10      |
|                                | Initiation                      | 3               | 7.46E-04      |
|                                | Elongation                      | 14              | 1.106E-07     |
|                                | Protein targeting               | 6               | 1.776E-04     |
|                                | mitochondria                    | 5               | 1.26E-03      |
|                                | Degradation:                    | 18              | 1.79E-05      |
|                                | Ubiquitin                       | 14              | 2.25E-03      |
|                                | Proteasom E3 ligase             | 7               | 7.42E-04      |
|                                |                                 | 5               | 1.07E-05      |
| Amino acid metabolism          | Synthesis:                      | 25              | 2.81E-07      |
|                                | Aspartate family                | 18              | 2.11E-06      |
|                                | Glutamate family                | 6               | 2.56E-04      |
|                                | Arginine family                 | 4               | 2.72E-04      |
|                                |                                 | 4               | 1.20E-04      |
| Nucleotide metabolism          | Synthesis:                      | 12              | 8.46E-04      |
|                                | Purine                          | 8               | 2.44E-06      |
|                                | Pyrimidine                      | 5               | 4.44E-05      |
|                                |                                 | 3               | 2.39E-03      |
| O-methyl transferases:         | -                               | 6               | 5.03E-06      |
| RNA processing                 | -                               | 17              | 5.23E-04      |
| RNA regulation of transcription| -                               | 13              | 1.11E-03      |
| Signaling                      | -                               | 13              | 1.11E-03      |
RNA-protein synthesis KD vs WT seeds

A 14DAF

RNA transcription
RNA processing
Protein aa. activation
Ribosomal proteins

Proteins:
Initiation
Elongation
Release

B 16DAF

RNA transcription
RNA processing
Protein aa. activation
Ribosomal proteins

Proteins:
Initiation
Elongation
Release

C Dry

RNA transcription
RNA processing
Protein aa. activation
Ribosomal proteins

Proteins:
Initiation
Elongation
Release

Downloaded from on March 18, 2020 - Published by www.plantphysiol.org
Copyright © 2009 American Society of Plant Biologists. All rights reserved.
Amino acids and starch synthesis KD vs WT seeds

A 14DAF
B 16DAF
C Dry

Amino acids synthesis

Starch synthesis
Biotic and abiotic stress responses KD vs WT seeds

A 14DAF

B 16DAF

C Dry

Biotic Stress

Abiotic Stress

Heat

Cold

Drought/salt

Touch/wounding

Light

Misc

Heat

Cold

Drought/salt

Touch/wounding

Light

Misc

Downloaded from on March 18, 2020 - Published by www.plantphysiol.org
Copyright © 2009 American Society of Plant Biologists. All rights reserved.
