SUPPLEMENTARY FILES

Supplementary File Text 1: Hallmark proteins with an association to dormancy

a) LEA proteins: We searched for the homology of lea-isoform k and the late embryogenesis abundant protein family protein (Protein ID1 corresponding with Contig9631 and Protein ID2, corresponding with Contig227 and Contig3165), with two previously identified LEA (Lea-1 and Lea-2) proteins (54). Sequence comparison showed that Lea-1 was 100% identical to Contig9631 protein lea- isoform k, while Lea-2 was 100% identical to Contig227 and Contig3165, late embryogenesis abundant protein family protein, although in both cases the contigs were longer than the Lea-1 and Lea-2 proteins from GenBank. Three additional translated transcripts were found highly similar (BLAST e-value < 10e-10) to both Lea-1 and Lea-2 proteins from GenBank: Contig4178 late embryogenesis abundant protein 76-like, Contig469 and Singlet16354 hypothetical protein. All three transcripts were expressed at the mRNA level, with very high expression in RE and low expression in AM. Only Contig469 was expressed at the protein level, with very high expression in RE and E0 and very low expression in AM. Pairwise comparison of translated Contig469 and Lea-1 shows many small regions with similarity (not identity), but with a different order between them. Possibly, Contig469 was assembled in a wrong manner, due to the repeats in the sequence, as LEA proteins are characterized by repeated sequences (Tunnacliffe and Wise, 2007).

b) Trehalose: Trehalose synthesis in all egg types was suggested by four trehalose–phosphate synthase proteins of ID135 (contig 14039), ID1088 (Contig 14054), ID2848 (contig 8645), ID2919 (singlet 16543), with the highest expression level of ID135.

c) Proteins with an association to the response of ROS and Oxidoreductases:glutathione-S-transferase, peroxiredoxins, superoxide dismutase and catalase showed differential abundance in the different the egg types. A few ferritin proteins also showed differential abundance in the comparison of AM and RE.

Cited reference: Tunnacliffe, A., and Wise, M. J. (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften 94, 791-812.

Supplementary File Text 2: Metabolome profiling

1. LC-MS Measurements (Hydrophilic and Lipophilic Analytes): The samples were measured with the use of a Reversed Phase Ultra Performance Liquid Chromatography (RP-UPLC, Waters ACQUITY) coupled to a mass spectrometer which consists of an electrospray ionization source (ESI) and an Orbitrap-type mass analyzer (Thermo-Fisher, Bremen, Germany). Chroamtographic separations was carried on with use of Reversed Phase Bridge Ethyl Hybrid (BEH) C_8 and Reversed Phase High Strength Silica (HSS) C_{18} columns (Waters) for lipophilic and hydrophilic compounds respectively. The mobile phase composition used for separation was previously described in Giavalisco et al., 2009. The mass spectra were acquired in survey full scan MS in positive and negative mode (Mass Range [100-1500]) for polar and lipophilic analytes. Additionally, the lipophilic compounds were measured in dd-MS/MS (data dependent tandem acquisition mode) collision energy 40 eV) in positive and negative mode in mass range (100-1500).

2. LC-MS Data Processing (Hydrophilic Analytes): The data analysis was performed with the software REFINER MS® 10.0 (GeneData, http://www.genedata.com). Alignment and filtration of the LC-MS data were completed using in-house R-based software. After extraction of the peak list from the chromatograms, the data is processed, aligned and filtered. At this stage an average RT and an average *m/z* values were given to the features. Thereafter, isotopic peaks, in-source fragments and lower intensity adducts were removed from the data. The annotation of the content of the sample was accomplished by matching the extracted data from the chromatograms with our library (metaSysX GmbH, Potsdam, Germany) of reference compounds.. For quantitation of metabolites the most intense adduct was used.

3. GC-MS Measurements: The samples derivatization and analysis were carried as described by Lisec et al., 2006. The GC-MS data were obtained using an *Agilent* Technologies GC coupled to a *Leco Pegasus HT* mass spectrometer which consists of an EI ionization source and a time-of-flight (TOF) mass analyzer.

Sample volumes of 1µl were injected into the GC column: 30 meters DB35; Starting temp: 85 °C for 2min; Gradient: 15 °C per min up to 360 °C.

4. GC-MS Data Processing and Annotation: NetCDF files that were exported from the Leco Pegasus software were imported to R program. The Bioconductor package TargetSearch was used to transform retention time to retention index (RI), to align the chromatograms, to extract the peaks, and to annotate them by comparing the spectra and the RI to the Fiehn Library and to a user created library. Annotation of peaks was manually confirmed in Leco Pegasus. Analytes were relatively quantified using a unique mass. Metabolites with a RT and a mass spectrum that did not result in a match in database were labeled as unknown metabolites.

5. Data normalization and statistical analysis: The data were sample median normalized.

To identify metabolites that were significantly altered between the groups, *t* test was run. The statistical significance was computed with two-tailed *t* test.

Cited References:

Giavalisco, P., Köhl, K,, Hummel, J., Seiwert, B., Willmitzer, L. (2009) 13C isotope-labeled metabolomes allowing for improved compound annotation and relative quantification in liquid chromatography-mass spectrometry-based metabolomic research. Anal. Chem. 81, 6546-6551

Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., Fernie, A. R. (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387-96

LEGENDS TO SUPPLEMENTARY TABLES

Table S1: Peptide report output from MaxQuant, which provides detailed information on all identified peptides.

Table columns: Sequence- The identified AA sequence of the peptide. Charges - All charge states that have been observed. PEP - Posterior Error Probability of the identification. This value essentially operates as a p-value, where smaller is more significant. Score- Andromeda score for the best identified among the associated MS/MS spectra. Delta Score- Score difference to the second best identified peptide. Intensity - Summed up extracted ion current (XIC) of all isotopic clusters associated with the peptide sequence.

Table S2: A. Detailed blast results for the best hit of each translated transcript. The column "Sequence description" presents the final description provided by Blast2GO based on all hits for that transcript. B. List of protein (internal) IDs with the corresponding transcript(s), description and best blast hit. Table columns: Protein ID – internal ID for the protein group; Majority Protein IDs – IDs of translated transcripts from the reference transcriptome to which peptides were mapped by MaxQuant, followed by their Blast2GO annotation; blast_hit – RefSeq ID of the best blast hit.

Table S3: The list of proteins that were identified in the three egg types and subsets thereof, with protein annotation, peptide counts, intensity values and 1-Way ANOVA results.

Table columns: Protein ID – internal ID; Majority Protein IDs – IDs of translated transcripts from the reference transcriptome to which peptides were mapped by MaxQuant; Majority protein titles – Blast2GO annotation for the transcripts; Sequence length – number of amino acids in a translated transcript; KO – KEGG orthology ID assigned to the translated transcript by KEGG KAAS server; pep – number of peptides (razor + unique) mapped to the translated transcriptIntens – intensity values (a.k.a. expression levels) in Log2 scale, summed up extracted ion current (XIC) of all isotopic clusters associated with all peptides assigned to the protein group, empty cells denote intensity=0; FDR p-value and Fold-Change are from the

1-Way ANOVA test. FDR p-values smaller than 0.05, are marked with a yellow background. Intensity and fold change values are colored using a color-scale.**Table S3A:** A list of all 2,928 proteins that were identified in any of the three egg types (AM, RE and E0). Lines marked with gray font indicate proteins not "present" in any of the egg types, where presence is defined as having at least two peptides in at least two replicates of an egg type. These proteins were omitted from subsequent analyses. **Table S3B:** A list of 2,631 proteins having at least two peptides in at least two replicates of an egg type ("present proteins"). This list was used in all further analyses. Additional columns show: Proteins that were present in the AM, RE and E0 groups, respectively; proteins showing differentially abundance in the comparison AM vs. RE and AM vs. E0, respectively; Proteins present in certain egg type(s) and absent in other(s). Columns marked by yellow headers display proteins showing differential abundance in the comparisons of AM vs. RE and AM vs. E0 belonging to pathways of specific interest. The last two columns in blue show the comparison of AM vs. RE and AM vs. E0, after LFQ-normalization, using less stringent FC cutoff for differential abundance (see details under "Experimental Procedures").

Table S4: Enriched GO terms for proteins present in AM (S4A), E0 (S4B) and RE (S4C). GO terms for GO biological processes (P), GO molecular functions (F) and GO cellular functions (C), are also given in this table for the three egg types. Clicking on the (+) sign below/next to the GO term opens a window with the full list of transcripts corresponding with the specific GO term.

Table S5: Enriched KEGG pathways in AM (S5A), RE (S5B) and E0 (S5C) and of the comparison of AM vs. RE (S5D) and AM vs. E0 (S5E). Two additional sub-tables show the enriched KEGG pathways in the sets of proteins which showed concordance (see Fig 13) with RNA levels, of AM (S5F) and of RE (S5G).

Table S6: Enriched GO terms for differentially abundant proteins in the comparison of AM vs. RE (S6A) and AM vs. E0 (S6B). Subsequent sheets show the same results but with GO terms separated to biological processes (P), molecular functions (F) and cellular components (C). Clicking on the (+) sign below/next to the GO term opens a window with the full list of transcripts corresponding with the specific GO term.

Table S7: Pathways showing differentially abundant kinases in the comparisons of AM vs. RE and AM vs. E0. The right column shows the number of differentially abundant kinase proteins in each pathway for both comparisons.

Table S8: The correlation in expression of AM proteins (A) or RE proteins (B) and their corresponding transcript (RNA) levels from RNA-Seq (51). The table displays the mean values of the intensity of proteins (log2 scale) and the matching signal intensity (log2 scale) of the corresponding transcripts. Based on the studentized test result, the last column specifies the color code for concordance used in Fig 13.

Table S9: The enriched GO terms for the proteins displayed in Fig 13 showing the correlation in expression of AM proteins (S9A, S9B) or RE proteins (S9C, S9D) and their corresponding transcript (RNA) levels. Clicking on the (+) sign below/next to the GO term opens a window with the full list of contigs corresponding with the specific GO term. Subsequent sheets show the same results but with GO terms separated to biological processes (P), molecular functions (F) and cellular components (C).