Co-option of a seed-like proteome by oil-rich tubers

Philipp William Niemeyer¹, Kerstin Schmitt², Oliver Valerius², Gerhard H. Braus², Jan deVries³, Anders Sven Carlsson⁴, Per Hofvander^{4*}, Till Ischebeck^{1*}

¹University of Göttingen, Albrecht-von-Haller-Institute for Plant Sciences and Göttingen Center for Molecular Biosciences (GZMB), Department of Plant Biochemistry, 37077 Göttingen, Germany

²University of Göttingen, Institute for Microbiology and Genetics and Göttingen Center for Molecular Biosciences (GZMB), Department for Molecular Microbiology and Genetics, Göttingen, 37077 Germany

³University of Göttingen, Institute for Microbiology and Genetics, Göttingen Center for Molecular Biosciences (GZMB) and Campus Institute Data Science (CIDAS), Department of Applied Bioinformatics, 37077 Göttingen, Germany

⁴Swedish University of Agricultural Sciences, Department of Plant Breeding, 230 53 Alnarp, Sweden

*Correspondence:

tischeb@uni-goettingen.de

per.hofvander@slu.se

Keywords:

Lipid droplets; Yellow nutsedge; *Cyperus esculentus*; Seeds; Proteome; Co-option

Abstract:

Co-option is an important aspect of evolution that can occur on several levels. Genes, whose function was molded by selection in the evolutionary past, are readily observed to serve a new function when acting in a different context in an extant system. Whole organs can be co-opted for new roles as well. For example, roots that evolved from shoot-like axes. Finally a framework of genes and its coded proteins can be co-opted to serve a similar molecular function but in a completely different organ, drastically changing its properties. Here, we describe such an example, where a set of proteins important for desiccation tolerance and oil accumulation in seeds of most angiosperms was co-opted in the tubers of yellow nutsedge (*Cyperus esculentus*). These tubers are not only desiccation tolerant but also store a large amount of lipids—especially TAG, similar to seeds. We generated nanoLC-MS/MS-based proteomes in five replicates of four stages of tuber development and compared them to the proteomes of roots and leaves, yielding 2257 distinct protein groups. Our data reveal a striking upregulation of hallmark proteins of seeds in the tubers. A deeper comparison to a previously published proteome of Arabidopsis seeds and seedlings indicate that indeed a seed-like proteome was co-opted. This was further supported by an analysis of the proteome of a lipid-droplet enriched fraction of yellow nutsedge, which also displayed seed-like characteristics.

Introduction:

Drought is the prime stressors that land plants have to overcome since their dawn (1). One strategy that land plants employ is to limit water loss and to transport water via a vascular system (2, 3). Another is to produce desiccation tolerant cells and tissues (4). Common to drought tolerance and desiccation is the accumulation of small osmolytes and proteins with protective functions (4). Also common to both strategies is the accumulation of neutral lipids, foremost triacylglycerol (TAG), in cytosolic lipid droplets (LDs) with especially high levels being reached in embryonic tissues (5). In most flowering plants, desiccation tolerance is limited to seeds and to some extend pollen. However, there are also plants that have desiccation tolerant vegetative organs some of the also referred to as “resurrection plants” such as *Sporobolus stapfianus* (6) and *Craterostigma plantagineum* (7) and it is likely that this tolerance evolved several times independently (8). It is however likely that these independent origins of desiccation tolerance are underpinned by the co-option of existing regulatory programs for resilience (9).

Here, we investigated yellow nutsedge (*Cyperus esculentus*), a monocot, perennial C4 plant (10). This species produces stolon-derived underground tubers that can fully desiccate and remain viable for years (11). Furthermore they can accumulate 25-30 % of their dry mass in lipids especially TAG (11, 12). In this regard they are unique as even the tubers of a close relative, purple nutsedge (*Cyperus rotundus*), are neither desiccation resistant nor oil accumulating (13, 14). We studied the proteomes of yellow nutsedge tubers and compared them to previously published proteomes of *Arabidopsis* seeds and seedlings. This comparison revealed striking similarities between tubers and seeds, and tuber sprouting and seedling establishment. We conclude that this species evolved the desiccation tolerance and oil-richness in their tubers by co-opting a protein framework normally present predominantly in seeds.

Results and Discussion:

Tubers of yellow nutsedge are enriched in hallmark proteins of seeds

To understand the molecular basis of tuber resilience in yellow nutsedge, we studied the proteome of four stages of tuber development (freshly harvested, dried, rehydrated for 48 h, and sprouted), and compared it to roots and leaves (Figure S1). From the first three tuber stages, we additionally isolated LD-enriched fractions. After a tryptic digest,

all peptide samples were analyzed by LC-MS/MS in five biological replicates and all protein groups were quantified using a label-free MS1-based quantification algorithm resulting in relative iBAQ (riBAQ) intensities. For functional annotation of these proteins, we used all nutsedge library entries as query for a BLASTp against the Arabidopsis TAIR10 primary transcript protein release library (15).

First, we studied the total proteomes. A principal component analysis (PCA) revealed that the tuber proteomes were mostly separated by developmental stage but much more distinct from roots and leaves (Figure 1A). In order to identify tuber-specific proteins, the data was hierarchically clustered (Figure 1A). Nine of the resulting thirty clusters (clusters 3-6, 8-11 and 13) contained 477 protein groups highly enriched in at least one of the tuber stages and far less abundant in leaves and roots. These protein groups corresponded to 433 Arabidopsis homologs that were subjected to a gene ontology search (16). This search indicated that 54 of these enriched proteins were associated with carbohydrate metabolism and 41 with cellular responses to stress but similar numbers were found in the other clusters combined. Overrepresented, however, were starch metabolism and several seed-related processes including seed maturation, dormancy, germination and seedling development. None of these seed-related terms were overrepresented in the other combined clusters.

Seven of the ten most abundant proteins that were at least tenfold enriched in one of the tuber stages in comparison to leaves and roots (Tables 1) have Arabidopsis homologs that are almost exclusively expressed in seeds based on transcript data (17); these homologs are further strongly enriched in proteomes of Arabidopsis seeds compared to 60 h old seedlings (18). Data gleaned from the total proteome thus suggest that yellow nutsedge might co-opt a seed-like proteome to sustain desiccation tolerant oil rich tubers.

Proteomes of tubers of yellow nutsedge and Arabidopsis seeds correlate

To further explore the similarity of tubers of yellow nutsedge and seeds of Arabidopsis, we compared the developmental pattern between the six yellow nutsedge stages/tissues investigated in this study with six stages of seedling establishment from Arabidopsis (18). 1200 protein homologs were present in the proteomes of both species and in all replicates of at least one stage. To allow for meaningful between-species comparisons, we defined, in each species and for each protein the highest

average abundance in a given stage as 1. This dataset was then analyzed in three ways.

First, a PCA revealed that the tuber and seed samples clustered closely together with the exception of fresh tubers that were more closely related to seedlings (Figure 2A).

Second, hierarchical clustering of the data (Figure 2B) yielded two clusters (4+5), comprising 90 proteins with a marked elevation in seeds rehydrated for 30 min and seeds stratified for 74 h at 4°C, and in dry and rehydrated tubers. These proteins included known LD-associated proteins such as oleosins and caleosins, late embryogenesis abundant (LEA) proteins and class I small heat shock proteins. All these protein families are known to be highly abundant in seeds (19-21) but likewise important in stress responses (22-25). Cluster 24 noteworthy contains 48 proteins high in sprouting tubers and seedlings, speaking to similarities between seedling establishment and tuber sprouting. Included in this cluster are proteins of the β -oxidation and glyoxylate cycle, needed for the conversion of lipids into carbohydrates. Further, tubers and seeds again clustered together.

Third, we calculated correlation coefficients between the averages of the different developmental stages to estimate similarities (Figure 2C). Here it was evident that, while tubers were most similar to each other, they also showed a positive correlation with the Arabidopsis seed and young seedling stages. Reversely, Arabidopsis seeds had higher correlation coefficients to nutsedge tubers than to 36 h old Arabidopsis seedlings.

Homologs to proteins with important functions in seeds and seedlings are enriched in tubers

GO term analyses based on the total abundance of proteins bolstered the similarities of tubers and seeds (Figure 2D). For instance, the terms embryo development, lipid storage, maintenance of seed dormancy and seed germination were especially high in seeds and tubers but reduced in older stages. The term maintenance of seed dormancy only comprises one protein, a homolog to the Arabidopsis protein 1-CYSTEINE PEROXIREDOXIN 1 (PER1) that enhances primary seed dormancy by suppressing ABA catabolism and GA biosynthesis (26). Another protein, homolog to GEM-RELATED 5 (GER5, AT5G13200), displays a similar pattern. GER5 is an ABA responsive gene that also regulates germination in Arabidopsis seeds (27). As of this,

it can be speculated that ABA and other hormones could play a role in oil accumulation, desiccation tolerance, and dormancy in tubers—mirroring seeds (28).

β -Oxidation and the glyoxylate cycle are key during *Arabidopsis* seedling establishment. This was further reflected in GO terms of sprouting tubers—although overall not reaching the same levels as *Arabidopsis* does, especially in case of the glyoxylate cycle. Yellow nutsedge thus seems to utilize its oil to synthesize carbohydrates, despite storing additionally large amounts of sugars and starch that are also degraded during sprouting (29).

The GO term analysis, however, also displayed differences between tubers and seeds. Tubers had higher levels of proteins associated with starch biosynthesis and glycolytic processes, likely reflecting that the tubers store starch and free sugars in higher amounts (30). Another difference is the amount of storage proteins present. While members of cupin storage protein families are enriched in tubers and decrease during sprouting, *Arabidopsis* seeds contain a far greater amount of storage proteins amounting to more than half of the total protein based on the proteomic data; this amount is far lower in tubers (< 0.5 %). Tubers on the other hand contain many proteins associated with protein folding, including various heat shock protein families that might protect proteins during desiccation. In contrast, these types of proteins are less abundant in *Arabidopsis* seeds and accumulate only during seedling establishment indicating that here other strategies are employed to preserve protein integrity.

Lipid droplet proteins known from seeds abound in tubers of yellow nutsedge

Most plant cells store TAG predominantly in cytosolic lipid droplets (LDs)(31). LDs are presumably present in all cell types but can strongly differ in their protein composition (18, 32-35). The described similarities between tubers and seeds and the notion that both accumulate oil raises the question if the tuber proteins that are associated with LDs are also similar to seeds. Therefore, we analyzed LD-enriched fractions of three tuber stages.

Homologs to known LD proteins made up ~60-70% of the proteins in LD-enriched fractions—with only ~5% being found in total extracts. With one exception, all these homologs were significantly enriched in the LD fraction (Figure 3A) indicating that these homologs are also LD associated. The most abundant among these proteins were the oleosins (Figure 3B)—especially the isoforms with higher sequence similarity to the *Arabidopsis* main seed oleosins OLE1, 2, 4 and 5 (17, 18)(Figure S2A). Oleosins

have been described to have a structural role as surfactants and to be seed and pollen-specific proteins (36). To our knowledge, they were so far not found in the proteome of vegetative tissues from flowering plants (33, 34). That said, *OLE* transcripts occur in yellow nutsedge tubers (11) and the leaves of the resurrection grass *Oropetium thomaeum*, after desiccation (9). Based on these findings and the presence of oleosin transcripts in *Physcomitrium* (*Physcomitrella*) *patens* spores and drought stressed algae, it is conceivable that oleosins are generally needed to shelter LDs under desiccation (5).

Steroleosins are likewise predominantly found in seeds and not in vegetative tissues (37). They are speculated to play a role in the metabolism of brassinosteroids (38, 39), hormones involved in development, and seed dormancy (40). There were two steroleosins isoforms in tubers (Figure S2B). This warrants future work on their role in development, dormancy or sprouting. Interestingly, seed lipid droplet protein (SLDP), another LD-associated protein (18) that is seed and seedling specific in *Arabidopsis* (17), was also detected in tubers but its molecular function is so far unknown.

Caleosins (CLO), in comparison, have been found in seeds, pollen and leaves (41-43). In *Arabidopsis*, CLO1 and CLO2 are the dominant caleosins of seeds (18); CLO3 is likely the most abundant leaf LD protein (34, 35) and is upregulated under stress (24). We explored whether the caleosins of yellow nutsedge showed a specific phylogenetic relationship to the seed or leaf isoforms of *Arabidopsis*. We however found that CLO1, 2 3 belongs to a larger clade within CLO gene family (here coined 'CLO1/2/3/6') within which lineage-specific duplications occurred; the most recent common ancestor of monocots and dicots likely had a single CLO1/2/3/6 homolog (Figure 3C).

Two further ubiquitously present LD proteins involved in the proper formation of LDs are the lipid droplet associated protein (LDAP) (44) and its interaction partner, LDAP interacting protein (LDIP) (45). LDAPs are low abundant in both desiccated tubers (Figure 3B) and seeds (18) while their interaction partner LDIP is also abundant in the desiccated structures of both species.

Apart from known LD proteins, also further proteins were found enriched in the LD fraction. Especially strongly enriched was a small G protein, RAB1C. Its mammalian homolog RAB18 is known to localize to LDs in mammals (46) but LD-binding G proteins are so far not described in plants. It appears possible, though that they are involved in LD formation or trafficking.

Another LD-enriched protein, the E3 ubiquitin ligase RGLG1 (RING domain ligase 1) (47), would be interesting to study in the future considering that several LD proteins get ubiquitinated prior to degradation (32, 48-50). RGLG1 from Arabidopsis is predominantly expressed in seeds (17) and was enriched in the LD-fractions of seedlings (18). It also plays a role in drought responses (47).

Conclusions and outlook

There are striking similarities between Arabidopsis seeds and tubers of yellow nutsedge. It is therefore conceivable that oil accumulation and desiccation resistance of both organs is based on a similar proteome and ultimately a similar gene expression pattern that could be regulated by a similar network of hormones and transcription factors. This network simply shifted its tissue-specificity, resulting in the co-option of an established molecular program by a different organ. Indeed, comparable protein patterns might not have occurred first in seeds, but across land plants to sustain a variety of oil-rich desiccation tolerant structures (5). For example, also the oil rich and drought resistant spores of *P. patens* abound in transcripts of typical LD and LEA proteins (51, 52).

Apparently, the reprogramming of yellow nutsedge tubers to seed-like characteristics evolved during a comparable short time frame. This underlines the potential for a complete reprogramming of vegetative tissues in crop plants by genetic engineering (53-55). Strategies observed in tubers of yellow nutsedge have the potential to guide such biotechnological approaches.

Material and methods

Plant Material:

Yellow nutsedge (*Cyperus esculentus* L. var. *sativus*) tubers were either taken dry, rehydrated for two days in water under continuous airflow or allowed to sprout in soil and harvested 1-2 days after sprouting. Yellow nutsedge plants were harvested from plants grown in a greenhouse under 14-16 h artificial light of mercury-vapor lamps (150 $\mu\text{mol m}^{-2} \text{sec}^{-1}$) complemented with sunlight and a temperature of 20-22°C at daytime or 16-18°C at nighttime. Roots and leaves were harvested from 33 d old plants and fresh tubers from 4 month old plants.

Isolation of total and LD-enriched protein fractions:

For each condition, 5 biological replicates were processed. Biological replicates used for lipid droplet isolation comprises 5 individual tubers, tissues of germinated tubers, leaves and roots derived from single individuals. The tubers were ground in a precooled mortar with 1-2 g sea-sand and 10 ml grinding buffer (50 mM Tris pH 7.4, 10 mM KCl, 200 mM Sucrose, 200 μ M PMSF). Subsequently, 2 ml of the suspension were transferred into a 2 ml reaction tube and spun for 10 s at 1000 g. A 50 μ l aliquot, referred to as total extract, was taken from the supernatant into 1 ml ethanol. The remaining material was centrifuged for 20 min at 20,000 g at 4°C. The floating fat pad was mechanically picked with a spatula and washed three times in 1.7 ml fresh grinding buffer by 20 min at 20,000 g at 4°C centrifugation. The resulting crude fat pad was resuspended in 1 ml ethanol and referred to as LD-enriched fraction. Total extract and LD-enriched fraction were stored at -20°C for at least one day to enhance protein precipitation.

Peptide sample preparation

The protein pellets were dissolved in 6 M urea and 5 % (w/v) SDS. Protein concentrations were determined with a Pierce BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA, USA). 10 μ g of protein was run on a SDS-PAGE gel until they entered the separation gel. A single gel piece per sample containing all proteins was excised, tryptic digested and derivatised (56). Peptides were desalted over an Empore™ Octadecyl C18 47 mm extraction disks 2215 (Supelco, St. Paul, MN, USA) as described (57).

LC-MS/MS analysis

Dried peptide samples were reconstituted in 20 μ l LC-MS sample buffer (2% acetonitrile, 0.1% formic acid). 3 μ l of each sample were subjected to reverse phase liquid chromatography for peptide separation using an RSLCnano Ultimate 3000 system (Thermo Fisher Scientific). Therefore, peptides were loaded on an Acclaim PepMap 100 pre-column (100 μ m x 2 cm, C18, 5 μ m, 100 Å; Thermo Fisher Scientific) with 0.07% trifluoroacetic acid at a flow rate of 20 μ L/min for 3 min. Analytical separation of peptides was done on an Acclaim PepMap RSLC column (75 μ m x 50 cm, C18, 2 μ m, 100 Å; Thermo Fisher Scientific) at a flow rate of 300 nL/min. The solvent composition was gradually changed within 94 min from 96 % solvent A (0.1 %

formic acid) and 4 % solvent B (80 % acetonitrile, 0.1 % formic acid) to 10 % solvent B within 2 minutes, to 30 % solvent B within the next 58 min, to 45% solvent B within the following 22 min, and to 90 % solvent B within the last 12 min of the gradient. All solvents and acids had Optima grade for LC-MS (Thermo Fisher Scientific). Eluting peptides were on-line ionized by nano-electrospray (nESI) using the Nanospray Flex Ion Source (Thermo Fisher Scientific) at 1.5 kV (liquid junction) and transferred into a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Full scans in a mass range of 300 to 1650 m/z were recorded at a resolution of 30,000 followed by data-dependent top 10 HCD fragmentation at a resolution of 15,000 (dynamic exclusion enabled). LC-MS method programming and data acquisition was performed with the XCalibur 4.0 software (Thermo Fisher Scientific).

Generation of protein library of *Cyperus esculentus* for peptide identification

Total RNA was extracted from frozen tissues of *Cyperus esculentus* mixed tissue sample and 8 different stages of tuber development by homogenizing material in Plant RNA Reagent according to instructions (Invitrogen, Carlsbad, USA). RNA integrity and concentration were determined using Experion RNA StdSens analysis kit (BioRad, Hercules, USA). Total RNA was DNase treated (TurboDNase, Ambion, Carlsbad, USA) before sequencing. Libraries for RNAseq were prepared at BGI (Shenzhen, China). Libraries were sequenced through the Illumina sequencing platform HiSeq 2000 as unpaired-end reads. Two different software were used to predict open-reading frames or proteins potentially encoded by the produced contigs (putative transcripts). TransDecoder (<https://github.com/TransDecoder>) Release v5.0.1 (58) was applied using default settings. GeneMark.hmm eukaryotic (<http://exon.gatech.edu/GeneMark/gmhmm.cgi>) was applied (59) using Arabidopsis as species model and with protein sequence as selected output for processing of the contigs. An additional protein reference was based on a previously published yellow nutsedge transcriptome (11). ORFs were extracted and translated using Geneious 8.1.8 (Biomatters Ltd., Auckland, New Zealand).

Data processing Calculation of relative intensity-based absolute quantification (riBAQ) values.

MS data was processed with Max Quant software version 1.6.2.10 (60, 61) with standard settings except for: „Match between runs“ was turned on; „iBAQ“ was

selected for label free quantification; FTMS recalibration was turned on. The three above mentioned nutsedge protein libraries were used. iBAQ values were determined with Max Quant 1.6.2.10. The data was further processed with Perseus software version 1.6.2.2 (62). Reverse hits, contaminants and proteins only identified by a modified peptide were removed from the data matrix. Then, all iBAQ values were divided by the total iBAQ value per sample and multiplied by 1000 yielding riBAQ in per mill. These values were the basis for all further data analysis.

BLAST

Homologs of nutsedge proteins in *Arabidopsis thaliana* were detected by using a BLASTp approach (63, 64). For this, protein data predicted from nutsedge transcriptomes were used as queries against the *Arabidopsis thaliana* TAIR10 primary transcript protein release using BLAST 2.5.0+.

PCA

The PCA of the nutsedge total extracts and the comparison of total and LD enriched fractions was performed on riBAQ values. Only proteins were considered that were identified by a least two peptides and that were found in at least three replicates of one of the stages. For the PCA comparing yellow nutsedge and Arabidopsis data, only proteins were considered that had a homolog in both species and were found in four out of five replicates of at least one stage. The values for each protein was normalized within the individual species setting the average of the highest stage to 1. PCA plots were created with Perseus 1.6.2.2 (62) using standard settings (Category enrichment in components was turned off).

Hierarchical clustering

The clustering was performed on average values for each stage/tissue. Furthermore, the highest average of each protein within each species was set to 1. Only proteins were considered that were found in all samples of at least one stage. In the comparison of yellow nutsedge and Arabidopsis only proteins were incorporated that had a homolog in both species. Plots were created with Perseus 1.6.2.2 (62) using Euclidian distance, average linkage and no constrains. The data was preprocessed with k-means. The number of clusters was set to 300, the maximum number of iterations to

100 and the number of restarts to 100. In the end, the proteins were grouped in 30 clusters.

Pearson correlation

Pearson correlation was performed on average values for each stage/tissue. Furthermore, the highest average of each protein within each species was set to 1. Only proteins were considered that were found in all samples of at least one stage and that had a homolog in both species. The correlation coefficients were calculated with Perseus 1.6.2.2 (62).

GO term enrichment

Go term enrichment analysis of proteins contained in certain clusters was done using an online tool based on the PANTHER classification system (<http://geneontology.org/>; <http://pantherdb.org/webservices/go/overrep.jsp>) (16).

Quantitative GO-term analysis

The quantitative GO-term analysis was based on riBAQ values. First, yellow nutsedge Proteins were assigned Arabidopsis identifiers if the e-value of the BLASTp result was lower than 10^{-5} . If several yellow nutsedge proteins were assigned the same Arabidopsis identifier, the values were added. Then, all proteins were assigned one or more GO terms and the values added for each term.

Volcano plots

Only proteins were considered that were found in four out of five replicates of at least one stage/subcellular fraction. Missing values were imputed with Perseus 1.6.2.2 (62) using a width of 0.3 and a downshift of 1.8 separately for each column. The volcano plot was then created using a t-test, a number of 250 randomizations, a false discovery rate of 0.01 and an S_0 of 2.

Analysis of LD proteins

All proteins that were homolog to a known LD protein from Arabidopsis were considered as LD proteins. The abundance of each LDs protein was divided by the total abundance of all LD proteins in the sample.

Phylogenetic analyses

In order to construct phylogenies, the datasets of the proteins of interest were supplemented with additional homologs from other seed plants and, in case of LDAPs, land plants and green algae. For this, well-described *Arabidopsis thaliana* proteins were used as query sequences in a BLASTp search against protein data from genome releases. In each case, the detected protein homologs in the yellow nutsedge transcriptomes were added.

For the phylogenetic analysis of CLO and HSD, homologs were mined from genome data of *Theobroma cacao* (65), *Picea abies* (66), *Oryza sativa* (67), *Nicotiana tabacum* (68), *Gnetum montanum* (69), *Gossypium hirsutum* (70), *Capsella grandiflora* (71), *Brassica rapa* (72), *Brachypodium distachyon* (73), *Amborella trichopoda* (74), *Arabidopsis thaliana* (15), and *Solanum lycopersium* (75).

For the phylogenetic analysis of OLE, homologs were mined from genome data of *Oryza sativa* (67), *Arabidopsis thaliana* (15), and *Solanum lycopersium* (75).

For the phylogenetic analysis of LDAPs, the same selection of homologs as in de Vries and Ischebeck (2020) were chosen, i.e. *Theobroma cacao* (65), *Triticum aestivum* (76), *Selaginella moellendorffii* (77), *Solanum lycopersium* (75), *Physcomitrium patens* (78), *Picea abies* (66), *Oryza sativa* (67), *Nicotiana tabacum* (68), *Marchantia polymorpha* (79), *Gnetum montanum* (69), *Gossypium hirsutum* (70), *Carica papaya* (80), *Capsella grandiflora* (71), *Brassica rapa* (72), *Brassica oleracea* (81), *Brachypodium distachyon* (73), *Amborella trichopoda* (74), *Arabidopsis thaliana* (15), *Arabidopsis lyrata* (82), the hornworts *Anthoceros agrestis* and *Anthoceros punctatus* (83), the ferns *Azolla filiculoides* and *Salvinia cucullata* (84), the Charophyceae *Chara braunii* (85), the genomes of the Zygnematophyceae *Mesotaenium endlicherianum* and *Spirogloea muscicola* (86) and the transcriptomes of the filamentous Zygnematophyceae *Zygnema circumcarinatum* SAG2419 (87) and SAG698-1a (88) as well as *Spirogyra pratensis* MZCH10213 and *Mougeotia* sp. MZCH240 (89), and genomes of the early-diverging streptophyte algae *Mesostigma viride* and *Chlorokybus atmophyticus* (90).

All protein sequences were aligned using MAFFT v7.453 L-INS-I (91). Sequences were cropped to retain the conserved region/protein features and re-aligned; sequences that did not cover at least 50% of the conserved region were removed. Maximum likelihood phylogenies were computed using IQ-TREE (92) multicore version 1.5.5 for Linux 64-bit. 500 bootstrap replicates were computed. Each

run of IQ-TREE entailed determining the best model for protein evolution according to the Bayesian information criterion (BIC) score via ModelFinder (93). The best models were: LG+G4 for CLO, JTT+G4 for OLE and HSD, and JTT+I+G4 for LDAP.

Acknowledgements:

PWN, OV, KS, GB and TI thanks the German research foundation (DFG, Grants IS 273/7-1, IRTG 2172 PRoTECT, BR1502-15-1 and INST 186/1230-1 FUGG to Stefanie Pöggeler) and JdV the European Research Council for funding (Grant Agreement No. 852725; ERC-StG 'TerreStriAL' under the European Union's Horizon 2020 research and innovation programme). OV and KS on behalf of the Service Unit LCMS Protein Analytics of the Göttingen Center for Molecular Biosciences (GZMB) thank the DFG for funding (INST 186/1230-1 FUGG to Stefanie Pöggeler). PH was financed by grants from the Swedish Foundation for Strategic Research (SSF) and Trees and Crops for the Future (TC4F), a Strategic Research Area at SLU, supported by the Swedish Government.

Author contributions:

P.W.N., G.H.B., P.H. and T.I. designed the work, P.W.N., K.S., O.V., J.d.V., A.S.C., P.H. and T.I. performed research, P.W.N., K.S., G.H.B., J.d.V., A.S.C., P.H. and T.I. analyzed data, and P.W.N., K.S., J.d.V., P.H., and T.I. wrote the manuscript. All authors critically read and revised the manuscript and approved the final version.

Declaration of Interests:

The authors declare no competing interests.

References

1. J. M. Fürst-Jansen, S. de Vries, J. d. Vries, Evo-physio: on stress responses and the earliest land plants. *Journal of Experimental Botany* (2020).
2. C. J. Harrison, J. L. Morris, The origin and early evolution of vascular plant shoots and leaves. *Philos Trans R Soc Lond B Biol Sci* **373** (2018).
3. K. J. Lu *et al.*, Evolution of vascular plants through redeployment of ancient developmental regulators. *Proc. Natl. Acad. Sci. USA* **117**, 733-740 (2020).
4. M. J. Oliver *et al.*, Desiccation tolerance: Avoiding cellular damage during drying and rehydration. *Annu Rev Plant Biol* **71**, 435-460 (2020).
5. J. de Vries, T. Ischebeck, Ties between stress and lipid droplets pre-date seeds. *Trends in Plant Science* <https://doi.org/10.1016/j.tplants.2020.07.017> (2020).
6. A. D. Neale *et al.*, The isolation of genes from the resurrection grass *Sporobolus stapfianus* which are induced during severe drought stress. *Plant, Cell & Environment* **23**, 265-277 (2000).
7. D. Bartels, Desiccation tolerance studied in the resurrection plant *Craterostigma plantagineum*. *Integr. Comp. Biol.* **45**, 696-701 (2005).
8. T. N. Le, S. J. McQueen-Mason, Desiccation-tolerant plants in dry environments. *Reviews in Environmental Science and Bio/Technology* **5**, 269 (2006).
9. R. VanBuren *et al.*, Seed desiccation mechanisms co-opted for vegetative desiccation in the resurrection grass *Oropetium thomaeum*. *Plant, Cell & Environment* **40**, 2292-2306 (2017).
10. M. S. Defelice, Yellow nutsedge *Cyperus esculentus* L.—snack food of the Gods. *Weed Technol.* **16**, 901-907 (2002).
11. Z. Yang, H. Ji, D. Liu, Oil biosynthesis in underground oil-rich storage vegetative tissue: comparison of *Cyperus esculentus* tuber with oil seeds and fruits. *Plant Cell Physiol* **57**, 2519-2540 (2016).
12. H. Turesson *et al.*, Characterization of oil and starch accumulation in tubers of *Cyperus esculentus* var. *sativus* (Cyperaceae): A novel model system to study oil reserves in nonseed tissues. *American Journal of Botany* **97**, 1884-1893 (2010).
13. J. Iqbal, S. Hussain, A. Ali, A. Javaid, Biology and management of purple nutsedge (*Cyperus rotundus* L.). *The Journal of Animal & Plant Sciences* **22**, 384-389 (2012).
14. G. D. Wills, Description of purple and yellow nutsedge (*Cyperus rotundus* and *C. esculentus*). *Weed Technol.*, 2-9 (1987).
15. P. Lamesch *et al.*, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202-D1210 (2012).
16. H. Mi *et al.*, Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols* **14**, 703-721 (2019).
17. A. V. Klepikova, A. S. Kasianov, E. S. Gerasimov, M. D. Logacheva, A. A. Penin, A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *The Plant Journal* **88**, 1058-1070 (2016).
18. F. K. Kretschmar *et al.*, Identification of low-abundance lipid droplet proteins in seeds and seedlings. *Plant Physiol* **182**, 1236-1245 (2020).
19. M. Battaglia, Y. Olvera-Carrillo, A. Garciarrubio, F. Campos, A. A. Covarrubias, The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* **148**, 6-24 (2008).
20. A. H. C. Huang, Plant lipid droplets and their associated proteins: Potential for rapid advances. *Plant Physiol* **176**, 1894-1918 (2018).
21. N. Wehmeyer, L. D. Hernandez, R. R. Finkelstein, E. Vierling, Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. *Plant Physiol* **112**, 747-757 (1996).
22. M. Hanin *et al.*, Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal Behav* **6**, 1503-1509 (2011).

23. T. L. Shimada, T. S. H. T. Y. F. I. Hara-Nishimura, A novel role for oleosins in freezing tolerance of oilseeds in *Arabidopsis thaliana*. *The Plant Journal* **55**, 798-809 (2008).
24. Y. Aubert *et al.*, RD20, a stress-inducible caleosin, participates in stomatal control, transpiration and drought tolerance in *Arabidopsis thaliana*. *Plant and Cell Physiology* **51**, 1975-1987 (2010).
25. L.-M. Guo, J. Li, J. He, H. Liu, H.-M. Zhang, A class I cytosolic HSP20 of rice enhances heat and salt tolerance in different organisms. *Scientific Reports* **10**, 1383 (2020).
26. H. Chen *et al.*, AtPER1 enhances primary seed dormancy and reduces seed germination by suppressing the ABA catabolism and GA biosynthesis in *Arabidopsis* seeds. *The Plant Journal* **101**, 310-323 (2020).
27. K. N. Baron, D. F. Schroeder, C. Stasolla, GEm-Related 5 (GER5), an ABA and stress-responsive GRAM domain protein regulating seed development and inflorescence architecture. *Plant Science* **223**, 153-166 (2014).
28. K. Chen *et al.*, Absciscic acid dynamics, signaling, and functions in plants. *J Integr Plant Biol* **62**, 25-54 (2020).
29. E. W. Stoller, D. P. Nema, V. M. Bhan, Yellow nutsedge tuber germination and seedling development. *Weed Science* **20**, 93-97 (1972).
30. J. P. Linssen, J. L. Cozijnsen, W. Pilnik, Chufa (*Cyperus esculentus*): a new source of dietary fibre. *Journal of the Science of Food and Agriculture* **49**, 291-296 (1989).
31. T. Ischebeck, H. E. Krawczyk, R. T. Mullen, J. M. Dyer, K. D. Chapman, Lipid droplets in plants and algae: Distribution, formation, turnover and function. *Semin Cell Dev Biol* 10.1016/j.semcdb.2020.02.014 (2020).
32. F. K. Kretschmar *et al.*, PUX10 is a lipid droplet-localized scaffold protein that interacts with CDC48 and is involved in the degradation of lipid droplet proteins. *The Plant Cell* **30**, 2137-2160 (2018).
33. P. J. Horn *et al.*, Identification of a new class of lipid droplet-associated proteins in plants. *Plant Physiology* **162**, 1926-1936 (2013).
34. L. Brocard *et al.*, Proteomic analysis of lipid droplets from arabidopsis aging leaves brings new insight into their biogenesis and functions. *Front Plant Sci* **8**, 894 (2017).
35. R. Fernández-Santos *et al.*, Protein profiles of lipid droplets during the hypersensitive defense response of *Arabidopsis* against *Pseudomonas* Infection. *Plant Cell Physiol* **61**, 1144-1157 (2020).
36. K. D. Chapman, J. M. Dyer, R. T. Mullen, Biogenesis and functions of lipid droplets in plants. *Journal of Lipid Research* **53**, 215-226 (2012).
37. T. L. Shimada, M. Hayashi, I. Hara-Nishimura, Membrane dynamics and multiple functions of oil bodies in seeds and leaves. *Plant Physiol* **176**, 199-207 (2018).
38. F. Li, T. Asami, X. Wu, E. W. Tsang, A. J. Cutler, A putative hydroxysteroid dehydrogenase involved in regulating plant growth and development. *Plant Physiol* **145**, 87-97 (2007).
39. S. Baud *et al.*, Regulation of HSD1 in Seeds of *Arabidopsis thaliana*. *Plant Cell Physiol*. **50**, 1463-1478 (2009).
40. A. L. G. Peres *et al.*, Brassinosteroids, the sixth class of phytohormones: a molecular view from the discovery to hormonal interactions in plant development and stress adaptation. *International Journal of Molecular Sciences* **20**, 331 (2019).
41. J. C. F. Chen, C. C. Y. Tsai, J. T. C. Tzen, Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds. *Plant and Cell Physiology* **40**, 1079-1086 (1999).
42. T. L. Shimada, I. Hara-Nishimura, Leaf oil bodies are subcellular factories producing antifungal oxylipins. *Curr. Opin. Plant Biol.* **25**, 145-150 (2015).
43. P.-L. Jiang, G.-Y. Jauh, C.-S. Wang, J. T. C. Tzen, A unique caleosin in oil bodies of lily pollen. *Plant Cell Physiol.* **49**, 1390-1395 (2008).
44. S. K. Gidda *et al.*, Lipid droplet-associated proteins (LDAPs) are required for the dynamic regulation of neutral lipid compartmentation in plant cells. *Plant Physiology* **170**, 2052-2071 (2016).

45. M. Pyc *et al.*, Arabidopsis lipid drop-associated protein (LDAP) - interacting protein (LDIP) influences lipid droplet size and neutral lipid homeostasis in both leaves and seeds. *The Plant Journal* **92**, 1182-1201 (2017).
46. C. B. K. Jayson *et al.*, Rab18 is not necessary for lipid droplet biogenesis or turnover in human mammary carcinoma cells. *Molecular biology of the cell* **29**, 2045-2054 (2018).
47. M. C. Cheng, E. J. Hsieh, J. H. Chen, H. Y. Chen, T. P. Lin, Arabidopsis RGLG2, functioning as a RING E3 ligase, interacts with AtERF53 and negatively regulates the plant drought stress response. *Plant Physiol* **158**, 363-375 (2012).
48. C. Deruyffelaere *et al.*, Ubiquitin-mediated proteasomal degradation of oleosins is involved in oil body mobilization during post-germinative seedling growth in Arabidopsis. *Plant and Cell Physiology* **56**, 1374-1387 (2015).
49. E. S. L. Hsiao, J. T. C. Tzen, Ubiquitination of oleosin-H and caleosin in sesame oil bodies after seed germination. *Plant Physiology and Biochemistry* **49**, 77-81 (2011).
50. C. Deruyffelaere *et al.*, PUX10 associates with CDC48A and regulates the dislocation of ubiquitinated oleosins from seed lipid droplets. *Plant Cell* **30**, 2116-2136 (2018).
51. N. Fernandez-Pozo *et al.*, PEATmoss (Physcomitrella Expression Atlas Tool): a unified gene expression atlas for the model plant *Physcomitrella patens*. *Plant J* 10.1111/tpj.14607 (2019).
52. C.-Y. Huang, C.-I. Chung, Y.-C. Lin, Y.-I. C. Hsing, A. H. Huang, Oil bodies and oleosins in *Physcomitrella* possess characteristics representative of early trends in evolution. *Plant Physiology* **150**, 1192-1203 (2009).
53. P. Hofvander *et al.*, Potato tuber expression of Arabidopsis WRINKLED1 increase triacylglycerol and membrane lipids while affecting central carbohydrate metabolism. *Plant Biotechnology Journal* **14**, 1883-1898 (2016).
54. T. Vanhercke *et al.*, Metabolic engineering for enhanced oil in biomass. *Prog. Lipid Res.* **74**, 103-129 (2019).
55. T. Vanhercke *et al.*, Step changes in leaf oil accumulation via iterative metabolic engineering. *Metab Eng* **39**, 237-246 (2017).
56. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856-2860 (2006).
57. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* **2**, 1896 (2007).
58. B. J. Haas *et al.*, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512 (2013).
59. A. Lomsadze, V. Ter-Hovhannisyanyan, Y. O. Chernoff, M. Borodovsky, Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* **33**, 6494-6506 (2005).
60. J. Cox *et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526 (2014).
61. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26**, 1367 (2008).
62. S. Tyanova *et al.*, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731-740 (2016).
63. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
64. W. Gish, D. J. States, Identification of protein coding regions by database similarity search. *Nat Genet* **3**, 266-272 (1993).
65. X. Argout *et al.*, The genome of *Theobroma cacao*. *Nat Genet* **43**, 101-108 (2011).
66. B. Nystedt *et al.*, The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579-584 (2013).
67. S. Ouyang *et al.*, The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res* **35**, D883-887 (2007).

68. N. Sierro *et al.*, The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun* **5**, 3833 (2014).
69. T. Wan *et al.*, A genome for gnetophytes and early evolution of seed plants. *Nat Plants* **4**, 82-89 (2018).
70. F. Li *et al.*, Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol* **33**, 524-530 (2015).
71. T. Slotte *et al.*, The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet* **45**, 831-835 (2013).
72. X. Wang *et al.*, The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* **43**, 1035-1039 (2011).
73. I. B. Initiative, Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763-768 (2010).
74. P. Amborella Genome, The Amborella genome and the evolution of flowering plants. *Science* **342**, 1241089 (2013).
75. T. G. Consortium, The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635-641 (2012).
76. I. W. G. S. Consortium, Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361** (2018).
77. J. A. Banks *et al.*, The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *science* **332**, 960-963 (2011).
78. D. Lang *et al.*, The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J* **93**, 515-533 (2018).
79. J. L. Bowman *et al.*, Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287-304 e215 (2017).
80. R. Ming *et al.*, The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452**, 991-996 (2008).
81. S. Liu *et al.*, The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun* **5**, 3930 (2014).
82. T. T. Hu *et al.*, The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet* **43**, 476-481 (2011).
83. F. W. Li *et al.*, Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat Plants* **6**, 259-272 (2020).
84. F. W. Li *et al.*, Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants* **4**, 460-472 (2018).
85. T. Nishiyama *et al.*, The Chara genome: Secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448-464 e424 (2018).
86. S. Cheng *et al.*, Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* **179**, 1057-1067 e1014 (2019).
87. M. Rippin, B. Becker, A. Holzinger, Enhanced desiccation tolerance in mature cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics. *Plant and Cell Physiology* **58**, 2067-2084 (2017).
88. J. de Vries, B. A. Curtis, S. B. Gould, J. M. Archibald, Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proc. Natl. Acad. Sci. USA* **115**, E3471-E3480 (2018).
89. J. de Vries *et al.*, Heat stress response in the closest algal relatives of land plants reveals conserved stress signalling circuits. *Plant Journal* (2020).
90. S. Wang *et al.*, Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat Plants* **6**, 95-106 (2020).
91. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780 (2013).
92. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **32**, 268-274 (2015).

93. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* **14**, 587-589 (2017).

Table

Table 1 The majority of the most abundant proteins strongly enriched in tubers have seed-specific homologs in *Arabidopsis*. The protein compositions of yellow nutsedge tubers (fresh, dry, rehydrated and sprouted) were compared to the ones of leaves and roots. Displayed are the ten most abundant proteins found in the tuber samples that were at least tenfold enriched in comparison to leaves and roots (by highest average). Enrichment in seeds was calculated between rehydrated seeds and 60 h old seedlings (18). *Arabidopsis* homologs were determined by BLASTp analysis.

Main protein ID	Enrichmenttuber		Enrichment		ArabidopsisArabidopsis	Arabidopsis
	tuber	%	Cluster	seeds	homolog	Gene/function
c37793_g1	37	107	10	0.96	AT2G36460	Fructose-BP aldolase
c42030_g1	44006	42	13	21.3	AT5G01670	Aldose reductase
contig_283 m.433	37	40	10	0.57	AT2G32520	Dienelactone hydrolase
contig_1251 m.1866	1162	37	11	n.f. in control	AT4G16160	Tim family protein
c43378_g1	7985	35	13	57.3	AT1G48130	ATPER1
c25411_g1	6922	30	13	83.5	AT1G54870	Oxidoreductase
c45866_g1	45	22	13	0.02	AT5G65020	ANNAT2
contig_4503 m.6288	132	22	3	50.2	AT4G25140	OLEOSIN 1
c39159_g1	1425	19	10	17.3	AT4G26740	CALEOSIN 1
contig_8289 m.10932	121	13	9	n.f. in control	AT3G18570	OLEOSIN 8

Supplemental information

Figure S1 Yellow nutsedge developmental stages investigated in this study.

Figure S2. LD enriched fractions could be reproducibly isolated.

Figure S3. Oleosin (OLE) and steroleosin (HSD) phylogeny.

Figure S4. Phylogeny of lipid droplet associated proteins (LDAPs).

Figure 1 Developmental tuber stages have a distinct proteome. Comparative proteomics of total protein extracts of tubers (fresh, FT; dry, DT, rehydrated for 48 h (RT) and sprouted (ST), leaves (L) and roots (R; Table S2). (A) Principal component analysis plot. (B) The values were normalized setting the average of the stage with the highest abundance to 1 and these values were hierarchically clustered. Proteins sorted by cluster can be found in (Table S3A). n=5 biological replicates per stage.

Figure 2 Comparison of protein expression pattern between nutsedge tissues and *Arabidopsis* seeds and seedlings. Total protein extracts of yellow nutsedge tubers (fresh, FT; dry, DT, rehydrated for 48 h, RT; and sprouted, ST), leaves (L) and roots (R) were compared to *Arabidopsis* seeds (rehydrated, RS, and stratified, StS) and seedlings (24 h – 60 h in the light after stratification). The principal component analysis (A) compares the individual samples. For the hierarchical clustering (B, see Table S5 for clusters) and the Pearson correlation analysis (C), samples were averaged for each stage. (D) Several GO terms are displayed that show similarities but also differences between tubers and seeds. Proteins were grouped by GO-terms and their relative abundance added up indicating per mill of the total proteome. A list of all GO terms can be found in Table S6 and S7. n=5 for all stages. *Arabidopsis* protein data was previously published in Kretzschmar et al. 2020.

Figure 3. LD proteomes of yellow nutsedge tubers resemble those of seeds and seedlings. A volcano plot was constructed to visualize proteins, which are significantly and consistently LD-enriched in yellow nutsedge (A). All 15 LD-enriched samples were compared to all total cellular fractions of fresh, dry, and rehydrated tubers (Table S2). The $\log_2$ -transformed values and p-values were calculated. Known LD proteins are indicated in green, while interesting candidates are in blue (see Table S8 for a list of all candidates). Black lines indicate a false discovery rate of 0.01. (B) The riBAQ intensities of LD-associated proteins in the LD-enriched fraction were calculated as a percentage of the riBAQ of all known LD-associated proteins (see Table S8 for data). (C) Unrooted maximum likelihood phylogeny of caleosin homologs detected in the predicted proteomes of spermatophyte genomes and the yellow nutsedge transcriptomes. Yellow nutsedge and Arabidopsis caleosins are depicted in blue and red, respectively. Abbreviations: CLO, caleosin; HSD, steroleosin; LDAP, LD-associated protein; LDIP, LDAP interacting protein; OLE, oleosin; PUX, plant UBX domain-containing protein; RAB, Rab GTPase; RGLG, RING domain ligase, SLDP, seed LD protein.

Figure S1 Yellow nutsedge developmental stages investigated in this study. Rehydrated tubers were soaked in water for 48 h and tubers sprouted in the soil were harvested 1-2 days after onset of sprouting. Roots and leaves were taken from plants grown in soil for 33 days omitting the tuber.

Figure S2. LD enriched fractions could be reproducibly isolated. Total protein extracts and LD-enriched fractions of yellow nutsedge tubers (fresh, FT; dry, DT, rehydrated for 48 h (RT)) were compared in respect to their proteome (Table S2). A principal component analysis was performed with all proteins that were found in at least three replicates of one stage (A). n=5 biological replicates per stage and fraction.

Figure S3. Oleosin (OLE) and steroleosin (HSD) phylogeny. Unrooted maximum likelihood phylogeny of oleosin (A) and steroleosin (B) homologs detected in the predicted proteomes of spermatophyte genomes and the yellow nutsedge transcriptomes. Yellow nutsedge and Arabidopsis proteins are depicted in blue and red, respectively.

Figure S4. Phylogeny of lipid droplet associated proteins (LDAPs). Unrooted maximum likelihood phylogeny of homologs detected in the predicted proteomes of spermatophyte genomes and the yellow nutsedge transcriptomes. Yellow nutsedge and Arabidopsis proteins are depicted in blue and red, respectively.