

Article

RNA-Seq Analysis of Developing Pecan (*Carya illinoensis*) Embryos Reveals Parallel Expression Patterns Among Allergen and Lipid Metabolism Genes

Christopher P Mattison, Ruhi Rai, Robert Settlege, Doug J. Hinchliffe, Crista Madison, John M Bland, Suzanne Brashear, Charles J Graham, Matthew R Tarver, Christopher Florane, and Peter J Bechtel

J. Agric. Food Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.jafc.6b04199 • Publication Date (Web): 25 Jan 2017

Downloaded from <http://pubs.acs.org> on February 5, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

RNA-Seq Analysis of Developing Pecan (*Carya illinoensis*) Embryos Reveals Parallel Expression Patterns Among Allergen and Lipid Metabolism Genes

Christopher P Mattison^{1*}, Ruhi Rai², Robert E Settlage², Doug J Hinchliffe³, Crista Madison³, John M Bland¹, Suzanne Brashear¹, Charles J Graham⁴, Matthew R Tarver⁵, Christopher Florane⁶, and Peter J Bechtel¹

¹ Southern Regional Research Center, FPSQ, ARS, U.S. Department of Agriculture, 1100 Robert E. Lee Boulevard, New Orleans, Louisiana, 70124, United States

² Virginia Bioinformatics Institute, 1015 Life Science Circle, Blacksburg, Virginia, 24061, United States

³ Southern Regional Research Center, CCU, ARS, U.S. Department of Agriculture, 1100 Robert E. Lee Boulevard, New Orleans, Louisiana, 70124, United States

⁴ Pecan Research & Extension Station, Agricultural Experiment Station, Louisiana State University-AgCenter, 10300 Harts Island Road, Shreveport, Louisiana 71115, United States

⁵ Biologics, Bayer CropScience, 890 Embarcadero Drive, West Sacramento, California 95605, United States

⁶ Southern Regional Research Center, CFB, ARS, U.S. Department of Agriculture, 1100 Robert E. Lee Boulevard, New Orleans, Louisiana, 70124, United States

*Corresponding author: (C.P.M.) E-mail: chris.mattison@ars.usda.gov

ORCID - Christopher P. Mattison: 0000-0001-7819-0914

Keywords: pecan, development, nut, transcriptome, allergy, fatty acid, metabolism

Running title: temporal pecan embryo transcriptome

Abstract

The pecan nut is a nutrient rich part of a healthy diet full of beneficial fatty acids and antioxidants, but can also cause allergic reactions in people suffering from food allergy to the nuts. The transcriptome of a developing pecan nut was characterized to identify the gene expression occurring during the process of nut development and to highlight those genes involved in fatty acid metabolism and those that commonly act as food allergens. Pecan samples were collected at several time points during the embryo development process including the water, gel, dough, and mature nut stages. Library preparation and sequencing was performed using Illumina based mRNA HiSeq with RNA from 4 time points during the growing season during August and September 2012. Sequence analysis with Trinotate software following the Trinity protocol identified 133,000 unigenes with 52,267 named transcripts and 45,882 annotated genes. A total of 27,312 genes were defined by GO annotation. Gene expression clustering analysis identified 12 different gene expression profiles, each containing a number of genes. Three pecan seed storage proteins that commonly act as allergens, Car i 1, 2, and 4 were significantly upregulated during the time course. Upregulated fatty acid metabolism genes that were identified included acyl-[ACP] desaturase and omega-6 desaturase genes involved in oleic and linoleic acid metabolism. Notably, a few of the upregulated acyl-[ACP] desaturase and omega-6 desaturase genes that were identified have expression patterns similar to the allergen genes based upon gene expression clustering and qPCR analysis. These findings suggest the possibility of coordinated accumulation of lipids and allergens during pecan nut embryogenesis.

Introduction

Pecan (*Carya illinoensis*) species are native to North America and are an important agricultural crop in the US. The US leads the world in pecan nut production accounting for approximately 80% of total production with an estimated value for the 2011 US crop at close to \$700 million¹. Pecan nuts are valued for their nutritional characteristics providing protein and unsaturated fats, as well as their enjoyable sensory qualities. Nuts contain high levels of antioxidants that are beneficial to health², and pecan's have been shown to contain very high levels of antioxidants³. Since 1980, Americans have consumed roughly half a pound of pecans per year⁴. Pecan nut and other tree nut consumption have been associated with several health benefits including improved serum lipid profile^{2,5}.

Unfortunately, while increased pecan nut consumption provides health benefits to most of the US population, a small but apparently increasing percentage suffers from tree nut allergies. The frequency of childhood tree nut allergy has steadily risen over the past decade⁶, and tree nut allergies are rarely outgrown⁷. There are 3 seed storage proteins that commonly act as tree nut and peanut allergens, 2S albumins, 7S vicilins, and 11S legumins⁸. In pecan nuts this includes the 2S albumin Car i 1⁹, the 11S legumin Car i 4¹⁰, and Car i 2 the 7S vicilin^{11,12}. How tree nuts regulate allergen gene expression and accumulation within developing nuts is not well understood.

Pecan and other nuts are good sources of healthy fats, and pecans are composed of 62-89% lipid^{13,14}. Pecan oil is high in oleic and linoleic unsaturated fats and is only about 5-10% saturated fat^{13,14}. However, there is evidence that dietary lipids such as those found in pecans and other tree

nuts may act as adjuvants for allergic sensitization to seed storage proteins found in nuts¹⁵. For example, the Brazil nut 2S albumin required the presence of Brazil nut lipids to induce antibody response in mice^{16, 17}. Further, the major mustard and peanut allergens, Sin a 1 and Ara h 1, have recently been shown to associate with phosphatidylglycerol, and this interaction has several effects including protecting the proteins from gastric, intestinal, and endolysosomal digestion, modifying their uptake by dendritic cells, and modulating cytokine production¹⁸. When and how fatty acids and nut allergens accumulate during kernel development likely depends upon the coordinated expression of genes that respond to nutritional and environmental factors that may influence the process.

Fatty acids perform multiple functions in plants including serving as an important source of energy reserves, essential membrane components, signaling molecules, and can play roles in plant defense^{19, 20}. Fatty acid synthesis is well characterized involving many enzymes including β -ketoacyl-ACP synthase III enzyme (GO:0033818)^{21, 22}, Beta-ketoacyl-CoA synthase I (GO:0004315) and II (GO:0033817), stearoyl-ACP desaturase (GO:0045300), and Acyl-ACP thioesterase enzymes (GO:0016297). There are likely multiple isoforms for each of the enzymes involved in fatty acid metabolism in a given organism, and the regulated expression of the genes encoding these enzymes continues to be an area of intense investigation²³⁻²⁶.

Underlying pecan nut development and the accumulation of allergens and fatty acids is a well-orchestrated set of genetic factors that integrate environmental cues, nutritional state, and plant stress. Real-time quantitative PCR has demonstrated that pecan allergen gene expression in developing ‘Sumner’ and ‘Desirable’ pecan cultivars peaks in September, and then declines

gradually as nut maturation progresses²⁷. To further characterize pecan nut development at the molecular level, gene expression in the internal cavity of the pecan kernel during development was analyzed by RNA transcriptome sequencing. The timing of allergen and fatty acid metabolism genes was highlighted to elucidate the relationship between allergen and fatty acid accumulation in developing pecan nuts.

Materials and Methods

Pecan Nut Samples

Nut samples (n = 10-20) were collected based on morphology from the same individual ‘Sumner’ or ‘Desirable’ trees at the Louisiana State University Pecan Research Station in Shreveport, LA at various dates from August through October of 2012. Each tree received standard agronomic practices²⁸. Nuts were collected at “late water” stage (August 11, 17), “gel stage” (August 23, 29), “dough stage” (September 4, 10), and “mature nut” (September 20, October 2). All samples were frozen immediately in liquid nitrogen and stored at –80°C for later chemical analysis. Tissue samples, either liquid, gel, or kernel depending upon the time point, were collected from the internal nut cavity for RNA isolation. The nut cavity tissues in this collection included the contents of the central vacuole, endosperm, embryo, placenta, seed coat, ovary packing tissue, developing fruit, and the cotyledon lobes as depicted by McKay²⁹. Structural components of the nut such as the hull, shell, and middle septum were not included in this analysis.

Proximate Analysis of Developing Pecan Nut Tissue

Proximate analysis of pecan nut samples for moisture, ash, and lipid content was performed using a single biological replicate for each individual time point from each cultivar,

and in some cases was not able to be performed due to the relatively minute amount and immature nature of the samples. Moisture content of developing pecan nut samples was determined by weight prior to and after drying for 5 days in a Virtis Freezemobile 12ES Freeze Dryer. Ash content was evaluated using between 0.5 to 2 grams of pecan nut sample. The samples were placed into a Lindberg/Blue muffle oven for 4 hours at 550 °C, removed from the oven and cooled for 20 minutes in a desiccator, and weighed again to determine mineral content. Lipid extractions were performed on freeze-dried sample using a Thermo Scientific Dionex ASE 350 Accelerated Solvent Extractor. Using optimized conditions, approximately 2 grams of developing pecan nut material was ground into fine particles with dry ice. The sample was mixed with 4.8 grams of diatomaceous earth (DE, Thermo Scientific) and placed into a 34 ml ASE 350 cell. When needed, a small amount of sand was placed on the top to fill the cell completely. Lipid extraction was performed with methylene chloride at 100 °C, 1500 psi, 2 min pre-heat, 5 min heat, 2 min static times, 60% flush volume, and 90 sec purge, and using the solvent saver/pressure mode. The extraction was repeated into separate 60 ml collection vials to obtain an extraction efficiency of 99%. A RapidVAP Vacuum Evaporation System (Labconco) operated at 35 °C, vortex speed at 30%, and vacuum at 300 mBar for 60 minutes was used to evaporate methylene chloride from the extracted lipids. Solvent evaporation was repeated until the mass of the oil was stable and no methylene chloride remained. Protein content determination was performed on a LECO FP628 using 0.15 grams of sample, in triplicate with standard deviations indicated in Table 1. Protein percent was calculated using the formula $N (\%) \times 5.32^{13}$.

Fatty Acid Analysis

Extracted oil samples (1 mg) were weighed into a 1.5 ml vial and mixed with 1 ml ethyl ether (dried over sodium metal), 20 μ l methyl acetate, and 40 μ l 0.5 M sodium methoxide (in methanol). The solution was mixed and after 5 min, the reaction was quenched with the addition of 30 μ l saturated oxalic acid/ether. The white precipitate was removed by centrifugation and the supernatant, with 100 μ l ether rinse of the pellet, was dried by nitrogen. The residue was dissolved in 500 μ l hexane with sonication and centrifuged to remove additional precipitate. The supernatant was analyzed by GC-MS using an Agilent 6890 GC, 5973 MS and Gerstel MACH (Agilent LTM) fast GC adaptation, using an HP-88 LTM column (30m x 0.18mm, 0.18 μ m; Agilent). The column heat gradient was 50 °C (1 min hold) to 140 °C at 25 °C/min, followed by a 4 °C/min gradient to 177 °C and a 2 °C/min gradient to 210 °C (1.65 min hold). To produce a 1 ml/min helium flow rate, a pressure ramp program was set up as 29.2 psi (1 min hold) to 39.3 psi at 2.81 psi/min, followed by a 0.44 psi/min gradient to 43.4 psi and a 0.22 psi/min gradient to 47.1 psi (1.65 min hold). The inlet was set to a 50:1 split at 220 °C, using an insert with glass wool (Agilent 5062-3587). The GC oven and MS transfer were set at 250 °C. A 5 μ l syringe was used in a 7683 autosampler. A mass range of 35-500 m/z was acquired. A certified Supelco 37-component FAME mix (C4-C24) (Sigma-Aldrich-CRM47885) was used to generate standard curves for all FAMEs. A 1 μ l injection was used for each sample analyzed by GC-MS.

Total RNA Isolation

For each RNA sample generated for sequencing or quantitative PCR, tissue from 4 to 5 nuts was combined for each biological replication, with a total of 3 biological replications per time-point and cultivar. Frozen pecan nut cavity tissues were ground to a fine powder using a mortar and pestle and liquid nitrogen. Total RNA was isolated from pecan nut cavity tissues using a protocol specifically developed for pine needles and other recalcitrant plant materials³⁰

with the following modifications. The amount of ground nut cavity tissue was 100 mg for each sample which was transferred to 1.7 ml microcentrifuge tubes. The volume of extraction buffer added to the tissue was 600 μ l followed by 2 extractions with 600 μ l chloroform:isoamyl alcohol (24:1). The aqueous phase was transferred to a new 1.7 ml microcentrifuge tube and 175 μ l of 8 M LiCl was added and the RNA precipitated overnight at 4°C. The remainder of the protocol was followed as described by Chang et al., (1993). The concentration and purity of each RNA sample was determined using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE) and a Qubit 2.0 fluorometer (Life Technologies, Grand Island, NY). The RNA quality for each sample was determined by RNA integrity number (RIN) using an Agilent Bioanalyzer 2100 and the RNA 6000 Nano Kit Chip (Agilent Technologies Inc., Santa Clara, CA) with 250 ng of total RNA per sample.

RNA-Seq Library Construction and Sequencing

Isolated RNA from the Sumner cultivar was used for library construction and sequencing. All sequencing library preparations were performed using an Apollo 324 Robot (Wafergen, CA, USA). Total RNA quality was validated on an Agilent BioAnalyzer 2100 (Agilent Technologies, Santa Clara CA, USA). A PrepX PolyA mRNA Isolation Kit (Wafergen, Fremont, CA, USA) was used to enrich 250 ng of total RNA for polyA RNA. The enriched PolyA RNA was then converted into a library of template molecules using a PrepX RNA-Seq for Illumina Library Kit (Wafergen, Fremont, CA, USA) for subsequent cluster generation and sequencing by Illumina HiSeq according to the manufacturer's protocol. Individually indexed cDNA libraries were clustered onto a flow cell using Illumina's TruSeq PE Cluster Kit v3-cBOT-HS, and sequenced with 101 Paired-End Sequencing on a HiSeq 2500 Ultra-High-Throughput Sequencing System using a TruSeq SBS Kit v3-HS.

Assembly, Annotation, and Gene Expression Analysis

Following sequencing, data was trimmed for both adaptor and quality using a combination of ea-utils and Btrim^{31,32}. Trimmed-paired reads were then *de novo* assembled using Trinity software³³ as strand specific reads. Annotation of transcripts (including identification of ORFs and attaching both description and function) was performed using the Trinotate package (<http://www.vcru.wisc.edu/simonlab/bioinformatics/programs/trinity/docs/annotation/Trinotate.html>) using default parameters as described in the Trinotate protocol. Reads were then aligned to the assembled transcriptome using Bowtie2³⁴. HTSeq and DESeq2 were used to count and determine significance of gene expression changes as described in the Trinity abundance estimation and identifying DE feature protocols^{35,36}. Cluster profile plots were created in R using kmeans to define 12 unique clusters of gene profiles.

Quantitative PCR (qPCR)

The experimental procedures and data analysis related to qPCR were performed according to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines³⁷. Isolated RNA from both Sumner and Desirable cultivars was used for qPCR experiments. The cDNA synthesis reactions were performed using the iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA) according to the manufacturer's protocol with 1 µg of total RNA per reaction used as template. Control cDNA synthesis reactions to check for genomic DNA contamination during qPCR consisted of the same template and components as the experimental reactions without the reverse transcriptase enzyme. The qPCR reactions were performed with iTaq™ SYBR® Green Supermix (Bio-Rad Laboratories) in a Bio-Rad CFX96 real time PCR detection system. Thermal cycler parameters for qPCR were

as follows: 95°C 3 minutes, 50 cycles of 95°C 15 seconds, 60°C 30 seconds. A dissociation curve was generated and used to validate that a single amplicon was present for each qPCR reaction. The calculations for amplification efficiencies of the target and reference genes, RNA inhibition assays, and the relative quantifications of the different target gene transcript abundances were performed using the comparative C_q method as described in the ABI Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR (Applied Biosystems, Foster City, CA) with the following modification: the average of two reference gene C_q values was determined by taking the geometric mean which was used to calculate the ΔC_q values for the individual target genes³⁸. The endogenous reference genes used in the qPCR reactions were the *C. illinoensis* 18S ribosomal RNA gene and the trnI-tmF chloroplast intergenic region as utilized in a previous study²⁷. Primer pairs used for the qPCR analysis are listed in Table 2. Developmental time points denoted with an asterisk have ≥ 2 -fold difference in transcript abundance levels between the two pecan varieties and are significantly different as determined by paired t-test.

Results

Sample Collection and Proximate Analysis

The moisture, mineral, protein, and fatty acid content of developing pecan nut samples from both Sumner and Desirable cultivars were characterized. Developing nut samples were collected at 7 time points correlating to the water, gel, dough, and mature nut stages during the 2012 growing season from August to October based upon nut morphology (Figure 1). Although there were slight variations between the cultivars, in most cases the proximate analysis characteristics were very similar. In both of the cultivars, moisture, mineral content, and

206 estimated carbohydrate level decreased during kernel development and was inversely correlated
207 to increases in protein and lipid content as nut development progressed (Table 1). Moisture
208 content was reduced from 46% to 17% in Desirable pecans, and from 43% to 20% in Sumner.
209 Mineral content in both cultivars dropped from around 3% to 1.8% in Desirable and 4% to 1.5%
210 in Sumner. The estimated carbohydrate percentage dropped from about 50% (48.76% for
211 Desirable and 50.85% for Sumner) to the mid teens in both cultivars (17.76% for Desirable and
212 15.23% for Sumner) during development. Lipid content rose from 2-3% in August to above
213 50% for both cultivars in October (Table 1). FAME (fatty acid methyl ester) analysis was used
214 to determine fatty acid content. Oleic acid was relatively stable for each time point for the
215 Sumner cultivar, but was observed to increase from August to September in the Desirable
216 cultivar (Figure 2). In comparison, linoleic acid content dropped from August to September in
217 the Desirable cultivar, but was relatively stable in the Sumner cultivar. The mean protein content
218 gradually increased during development within each sample, but appeared to initiate slightly
219 later in the Sumner cultivar (Table 1). In some samples, there was not enough material to allow
220 detectable protein from the nuts in early August from either cultivar, but soon after protein
221 content rose from about 3% in mid August to 10-12% in the October samples.

222 RNA Sequence Assembly and Gene Annotation

223 The genes expressed within the kernel were sequenced to try to correlate the
224 physiological and biochemical developments within the developing kernel to changes in gene
225 expression using 4 time points. RNA from four time points, August 11, August 23, September 4,
226 and September 20, 2012, were evaluated using the Sumner cultivar samples. Sequencing was
227 performed using Illumina based mRNA HiSeq and 200ng of total RNA isolated from each
228 developing pecan nut sample. A total of 86,526,858 bases (40.4% GC, a maximum contig length

of 12,160 bases, average length of 607 bases) were sequenced and assembled. ORF annotation was performed using the SwissProt/Uniprot/Uniref90 databases using Trinotate. Following the Trinity protocol, 142K transcripts were observed and collapsed into ca. 133K unigenes. Further attempts at clustering transcripts into genes (TGI-CL, CD-HIT-EST) did not result in an appreciable reduction in gene number presumably due to issues in assembly arising from either sample quality or population polymorphisms. There were 45,882 named Trinity genes with annotation and 52,267 named Trinity transcripts.

Annotated genes were given GO category assignments (biological process, cellular component, and molecular function). A total of 27,312 genes received GO annotations including 836 unique cellular component, 2,355 unique molecular function, and 4,309 unique biological process annotations. A large number of gene products functioning in biological processes such as carbohydrate and lipid metabolism, defense, stress response, and signaling were identified; including those localizing to or within membranes and those involved in molecular functions related to ATP binding and protein serine/threonine kinase activity. The top 25 GO groups having significant temporal gene expression differences throughout the time course are represented in Figure 3, and they are plotted with respect to the number of genes within the biological process (3A), cellular component (3B), and molecular function (3C) categories. There were a large number of genes involved in transcription, components of the nucleus or membranes, and ATP binding during embryo development that stood out from these plots.

The relative magnitude and directional change in gene expression within each category was also compared between time points, transition 1 (8/11-8/23), transition 2 (8/23-9/4), and transition 3 (9/4-9/20) (Figure 4 A-C). Amino acid transport, cellulose biosynthesis, mucilage extrusion, and lipid transport were among the biological processes that were generally

upregulated at each transition, while those associated with protein folding were generally downregulated at each transition. Genes involved in transmembrane transport were upregulated during transition 1 and 3. Conversely, translation associated genes were upregulated during the transition 2. As groups, anchored membrane components, anchored components of the plasma membrane, golgi membrane, and plasma membrane constituents were mostly upregulated at each of the transitions, while monolayer-surrounded lipid storage body components and nucleolus components were down regulated at transition1 and 3, but upregulated at transition 2. Lysosome and microtubule associated proteins were upregulated during transition 1 and 3, but downregulated during transition 2. Molecular functions that were generally upregulated at each transition included acid phosphatase activity, cellulose synthesis, and serine-type endopeptidase activity. While these observations provide a broad view of gene expression changes, continued analysis was focused on analyzing allergen and fatty acid metabolism genes.

Allergen Gene Expression Analysis

The mRNA sequencing data was inspected for temporal gene expression changes in the 3 conserved pecan allergens. The absolute expression values of each of the 3 allergen genes, Car i 1 (comp34075_c0), Car i 2 (comp34067_c0), and Car i 4 (comp34066_c0) demonstrate that expression of the 7S vicilin, Car i 2, increases the most during development (Figure 5A). A great deal of the increase in expression of Car i 2 occurred during the first and third transitions. When the expression changes were normalized, clear differences in expression profiles of the allergens were easier to discern. While expression of each of the allergen genes increased at the first transition, the expression of the Car i 1 2S albumin continued to increase during the second transition. During this time, expression of Car i 2 (comp34067_c0) and Car i 4 (comp34066_c0) leveled off or decreased slightly. Further, there was a sharp decrease in expression of the Car i 1

2S albumin gene (comp34075_c0) at the third transition as expression of the Car i 4 (comp34066_c0) and Car i 2 (comp34067_c0) genes was elevated at the same time (Figure 5B).

Gene Expression Profile Clustering

The three pecan allergen genes had different expression profiles, although the overall pattern of Car i 4 (comp34066_c0) and Car i 2 (comp34067_c0) genes were more similar. To identify groups of genes with comparable expression profiles, and identify those with expression patterns similar to the allergen genes, the gene expression data was evaluated using a cluster analysis of scaled gene expression values. The three allergen genes were categorized into cluster one for Car i 1 (comp34075_c0), cluster two for Car i 2 (comp34067_c0), and cluster three for Car i 4 (comp34066_c0) by this analysis (Figure 6). These three clusters (1-3) are similar in that there was an overall increase in expression with time. There were 233, 594, and 422 genes in clusters one, two, and three respectively, and 80, 245, and 169 genes were annotated in clusters 1-3 respectively. Included in these clusters were numerous genes involved in DNA binding and transcription, transmembrane transport functions, nutrient reservoir activity, and lipid metabolism (Supplementary material Table 1). For example, the expression of several fatty acid metabolism and storage genes were identified in cluster three including phospholipid:diacylglycerol acyltransferase 2 (PDAT, comp28848_c0), acyltransferase (comp28848_c0), desaturase (comp17048_c0 and comp34232_c1), beta-ketoacyl-CoA synthase I (comp34322_c0), and oleosins (comp34102_c0 and comp34087_c0). The expression pattern of a couple keto-ACP synthase genes (comp38847_c0 and comp126782_c0) matched that of the other genes included in cluster two. Genes in cluster one whose expression was similar to Car i 1, the 2S albumin, peaked at transition two (Figure 6). Genes within cluster one that were upregulated during transition two were involved in ethylene-activated signaling, response to

stress, as well as fatty acid metabolism (Supplementary material Table 1). The expression pattern of a long-chain fatty acid-CoA ligase enzyme (comp39519_c0) matched it to cluster one.

Nine other distinct expression profile clusters were differentiated from the gene expression data (Figure 6 and Sup Mat Table 1). For several of the clusters including four, five, seven, eight and nine, gene expression showed an overall decline (Figure 6), however the other cluster profiles had either maintained or increased overall gene expression. There were 455, 200, 741, 248, and 72 genes in clusters four, five, seven, eight and nine respectively. The types of gene functions in each of the clusters that were upregulated varied. For example, genes within cluster ten (containing 120 genes) included those involved in ATP biosynthesis, cellulose biosynthesis, seed coat development, and xylem development (Figure 6 and Supplementary material Table 1). Genes in cluster six (containing 323 genes) and cluster twelve (containing 391 genes) showed an increase in expression during the first transition (Figure 6 and Sup Mat Table 1), and included genes involved in pollen and cell wall-related synthesis or function. Genes within cluster six and eleven (containing 278 genes) included those involved in stress response or defense, protein synthesis, transport functions, and fatty acid metabolism.

Expression of Genes Involved in Fatty Acid Metabolism

Pecan oil is highly nutritious and studies have provided support for the health benefits of pecan consumption³⁹. Pecan oil contains very low levels of saturated fat, and is high in monounsaturated oleic and polyunsaturated linoleic acids^{40, 41}. The gene expression data was inspected to identify those annotated genes thought to be involved in fatty acid metabolism. There were 299 genes (52 whose expression was significantly altered) involved in various steps of fatty acid metabolism (Sup Mat Table 2), and these had a wide range of expression profiles, some of which fit into the clusters described above. These included a putative

phospholipid:diacylglycerol acyltransferase 1 (PDAT, comp34721_c1) in cluster 4, a 1-acyl-sn-glycerol-3-phosphate acyltransferase (LPAT, comp29904_c0) in cluster 7, and a diacylglycerol O-acyltransferase 1 (DGAT, comp32073_c0) in cluster 12. Several of these fatty acid metabolism genes were distinctly upregulated during the time course of nut filling that was studied. For example, there were ten acyl carrier protein (ACP) encoding genes, and among them all but two were upregulated during the time course. Seven of the ACP encoding genes had increased expression during the first transition, and in particular comp34310_co was elevated ~600-fold, the most based upon scaled expression changes. There were three malonyl-CoA:ACP transacylase genes, and one of these, comp40800_co, was upregulated ~100-fold, and much of this occurred during the third transition. There were thirty-six ketoacyl-ACP synthase genes, and many of them were downregulated during the time course. Exceptions to this included comp38847_c0, comp126782, and comp91447_co which were each highly upregulated (9,859, 4,564, and, 969-fold respectively).

Two hydroxyacyl-ACP dehydratase genes were identified, and comp40728_co and comp39981_co were both upregulated during the first transition. Comp39981_co remained elevated during transition one, two, and three, while comp40728_co expression was elevated sharply during transition one and then steadily declined during transition two and three. Two enoyl-ACP reductase genes were identified, comp25799_co and comp40412_co, but comp40412_co had the most noticeable change and was upregulated 205-fold during the first transition while comp25799_co was up 149-fold. Seven of the eleven acyl-CoA synthetase genes were downregulated during the time course, but comp33533_co, comp17321_co, comp39519_c0, and comp36497_co were exceptions. Comp39519_co, comp33533_co, and comp17321_co were up sharply during the first transition (627, 408, and 303-fold respectively),

344 while comp36497_co was relatively flat during transition one and two, and was upregulated 274-
345 fold during the third transition. Only one of the two acyl-ACP thioesterase genes,
346 comp33673_co, was upregulated during the time course with a marked increase of 513-fold
347 occurring during the first transition.

348 There were 103 acyltransferase genes identified in this analysis and forty-seven were
349 upregulated over the duration of the developmental period that was examined. Of these forty-
350 seven, there were seven genes including comp64068_co (19,154-fold), comp16498_co (3,348-
351 fold), comp60602_co (2,624-fold), comp20325_co (2,493-fold), comp122400_co (2,290-fold),
352 comp22057_c0 (1,920-fold), and comp84270_c0 (1,174-fold) that were markedly upregulated.
353 Three of the thirty-eight desaturase genes that were sequenced were upregulated over 500-fold
354 during the time course. Comp34232_c1 was up 2,299-fold, comp34153_co was up 969-fold, and
355 comp17048_co was up 722-fold. Another desaturase, comp17172_c0, had a sharp increase in
356 expression during the first transition (1,539-fold) similar to comp34232_c1 and comp34153_c0,
357 but its level dropped substantially during the second and third transitions.

358 Among the eight acetyl-CoA carboxylase genes that were identified, comp34460_c0
359 (from cluster 1), comp17318_c0, and comp43181_c0 were upregulated. Comp34460_c0,
360 previously highlighted in cluster one, and comp43181_c0 had a similar expression spike at the
361 first transition, while comp17318_c0 showed a more steady increase over the time course. There
362 were a pair of 3-oxoacyl-ACP synthase III genes and one of them, comp21054_c0, was
363 upregulated sharply during the first transition like the acetyl-CoA carboxylase comp34460_c0.
364 Four 3-oxoacyl-[acyl-carrier-protein] synthase II genes were identified and comp34322_c0 had
365 the largest (856-fold) change during the time course. There were two of five 3-oxoacyl-ACP
366 reductase genes, comp37759_c0 and comp25546_c0, whose expression was upregulated, 855

and 636-fold respectively, over the time course. Notably, comp37759_c0 was upregulated 1,321-fold during the third transition. A third 3-oxoacyl-ACP reductase gene, comp34514_c0, was upregulated over 400-fold during the first transition, but its level decreased substantially during transition 2 and 3. None of the five 3-hydroxyacyl-ACP dehydratase genes was substantially changed cumulatively during the time course, but comp40728_c0 was upregulated 344-fold during the first transition and declined to initial levels after that. Both of the enoyl-ACP reductase I genes that were identified had approximately 400-fold increased expression during transition one, but declined rapidly to initial levels during transition two and three.

Quantitative PCR of Selected Pecan Allergen and Fatty Acid Metabolism Gene Expression

Twelve of the fatty acid metabolism genes that were identified from the transcriptome sequencing were also evaluated for gene expression using qPCR. This analysis was performed with samples from both the Sumner cultivar used for transcriptome analysis and a second cultivar, Desirable, for comparison (Figure 7). Each of these genes was significantly upregulated at some point during the time course that was examined, and for some including a 3-oxoacyl-ACP synthase III (comp21054_c0), an acyl-[ACP] desaturase (comp34153_c0), an acyl-ACP thioesterase (comp33673_c0), and an acyl carrier protein (comp34310_c0) there was a good correlation to the expression pattern observed using the transcriptome analysis.

There were a few distinct expression patterns revealed by this analysis. In general terms, the expression pattern of a hydroxyl-ACP dehydratase (comp39981_c0), an acyl-activating enzyme (comp40417_c0), and a malonyl-CoA:ACP transacylase (comp40800_c0) were alike as they all peaked at the first sampling time point in August (Figure 7). The comp40417_c0 expression data from the transcriptome analysis agreed well with the qPCR data and showed a

greater than 90-fold decrease in the first transition. The expression pattern of a 3-oxoacyl-ACP synthase III (comp21054_c0), an acyl-ACP thioesterase (comp33673_c0), an acyl-[ACP] desaturase (comp34153_c0), and an acyl carrier protein (comp34310_c0) assessed by qPCR all peaked at the second time point in August. Other transcripts including two acyltransferases (comp16498_c0 and comp64068_c0), a desaturase (comp34232_c1), and a keto-ACP synthase (comp38847_c0) peaked at one or both of the latter time points in September (Figure 7).

The same general gene expression trends were observed between the two cultivars for some of the genes evaluated by qPCR. For example, relative transcript abundance of a hydroxyl-ACP dehydratase (comp39981_c0), an acyl-activating enzyme (comp40417_c0), and a malonyl-CoA:ACP transacylase (comp40800_c0) peaked early in the time course for both cultivars, and decreased at later time points (Figure 6). In several instances the correlation was not as obvious, and this could be due to primer mismatches due to differences in gene sequence between cultivars, or variations in the timing or selection of gene expressed during kernel development between the cultivars.

Discussion

Recent studies have characterized transcriptomes for tree nuts and other oil rich crops including hickory, tung, and physic nut^{24, 42, 43}. The study presented here characterized the gene expression within a developing pecan kernel, and highlighted the expression of fatty acid metabolism and allergen genes. However, expression was only analyzed during a single growing season, and changes in moisture, temperature, or pest attack could alter gene expression during growing seasons. For example, the expression of heat shock protein and acetyl-CoA carboxylase enzyme complex genes from developing hickory nuts (*Carya cathayensis* Sarg.) were fit to

different expression clusters in successive growth years during 2012 and 2013, possibly due to increased temperatures ⁴².

Predictably, an increase in pecan nut lipid content (to over 50% in the final sample) was observed as development progressed from August to October. The values are similar to those reported for Western Schley pecans at a comparable stage of development which averaged 19% moisture content (MC) and 60% lipid content ⁴⁴. Nuts collected with a MC near 4% had an average lipid content of 71% in the same study. The values collected here are lower than a previous study that reported a lipid content of 67 and 71% for the Sumner (MC 6.3%) and Desirable (MC 4.8%) cultivars respectively ¹³. Lipid content was dominated by oleic and linoleic acids, and the oleic acid values observed were close to those previously reported in the Desirable cultivar (56%), but were lower than those reported for the Sumner cultivar (50% versus 73%) ¹³. Both cultivars in the analysis had linoleic acid content (40 and 50% for Desirable and Sumner respectively) higher than previously reported values of 35 and 19% ¹³. In contrast, moisture values in the final samples were much higher than those published by Venkatachalam et al., (2007), suggesting the samples collected directly from the tree were not as dry as commercially available nuts. Rudolph et al., (1992) reported that increases in percent oil on a fresh weight basis over the last four weeks before harvest of Schley and Delmas nuts was primarily due to water loss rather than increased lipid synthesis. The variation in lipid and moisture content could reflect differences in growth conditions, and collection or storage time prior to analysis. The protein and ash values observed in the final samples were similar to the values described by Venkatachalam et al., (2007). Overall, the Sumner and Desirable cultivars showed similar changes over time and were similar in their nutrient content values.

Nearly 300 unique transcripts potentially involved in fatty acid metabolism were identified. For comparison, an analysis of nuts from the tung tree (*Vernicia fordii*) transcriptome identified 136 unigenes involved in fatty acid metabolism and another 126 lipid biosynthesis related genes⁴³. Acyl-[ACP] desaturase proteins are essential for oleic acid formation in plants, and are important for plant defense signaling⁴⁵. Eleven potential acyl-[ACP] desaturases in the pecan nut transcriptome were identified and five of these were highlighted in the expression pattern analysis. The Arabidopsis genome is predicted to harbor seven acyl-[ACP] desaturases genes, but one of these, SSI2/FAB2, plays a major role in oleic acid production⁴⁶. Two of the acyl-[ACP] desaturases expression patterns fit into cluster three (comp17048_c0 and comp34617_c0), two into cluster six (comp17172_c0 and comp34153_c0) and one into cluster ten (comp135859_c0). A sharp increase in expression between the August 11th and 23rd sampling dates for all five of these acyl-[ACP] desaturases was observed. The sharpest increases were observed for the comp34153_c0, comp17172_c0, and comp17048_c0 acyl-[ACP] desaturases which were up 10637, 2737, and 825-fold respectively between August 11th and 23rd. Similar increases in the expression of fatty acid metabolism genes including acetyl-CoA carboxylase, 3-ketoacyl-ACP reductase, and omega-3 fatty acid desaturase were observed in a study of the developing hickory nut (*Carya cathayensis* Sarg.)⁴². The expression pattern of the comp34153_c0 acyl-[ACP] desaturase was confirmed by qPCR analysis, and was shown to peak in late August and early September in the Sumner cultivar, but the same analysis was inconclusive in the Desirable cultivar possibly due to primer mis-match. The results suggest the comp34153_c0 transcript encodes an acyl-[ACP] desaturase that is likely important for the production of oleic acid.

Synthesis of linoleic acid requires omega-6 (delta 12) desaturase activity. Like oleic acid, linoleic acid plays a role in plant defense signaling and stress tolerance^{47,48}. For example, the Arabidopsis FAD2 omega-6 desaturase localizes to the endoplasmic reticulum and has been shown to be important for salt tolerance during germination⁴⁹. The transcriptome analysis identified three putative omega-6 desaturase enzymes, comp29832_c0, comp34232_c1, and comp35196_c0. The comp34232_c1 transcript was noticeably upregulated early during nut development and its expression profile fit into cluster three. The comp34232_c1 transcript was sharply upregulated between August 11th and August 23rd in a manner similar to the comp34153_c0 transcript encoding the putative acyl-[ACP] desaturase discussed above. Similar to the analysis presented here, FAD2 expression was also upregulated early during hickory nut development, but was downregulated thereafter⁴². The qPCR findings for the comp34232_c1 transcript were similar to the transcriptome analysis, showing peak levels during the September 10th time point in the Sumner cultivar. Similar to the comp34153_c0 transcript, the qPCR analysis of the comp34232_c1 transcript was inconclusive in the Desirable cultivar.

There is evidence that allergen sensitization can be influenced by the presence of lipids associated with the allergen or included within the sensitizing agent, including nut allergens¹⁶⁻¹⁸. In contrast, a link between fatty acid metabolites, T-cell function, and the presence of specific gut flora in infants has been suggested to be a positive influence on immune function⁵⁰. The acyl-[ACP] desaturase (comp34153_c0) and omega-6 desaturase (comp34232_c1) transcripts were highlighted due to their sharply increased expression during pecan nut development, and in the case of the omega-6 desaturase (comp34232_c1) because it exhibited an expression pattern similar to the Car i 4 pecan nut allergen in cluster 3. While expression clusters were not characterized in a tung tree nut study, desaturase enzymes were upregulated in the latter part of

480 nut development, and expression of an 11S legumin seed storage family protein was among the
481 top fifty most upregulated⁴³. Similarly, during physic nut development, enzymes for production
482 of oleic and linoleic acid were upregulated and transcription of 11S globulin family proteins was
483 among the most highly upregulated²⁴.

484 The findings presented here suggest coordinated regulation among select fatty acid
485 metabolism and pecan allergen genes during pecan kernel development. The applied extension
486 of these findings could provide information that may identify solutions helping those suffering
487 from tree nut allergies and facilitate additional food safety technologies. Continued research of
488 this type could also enable a better understanding of how the genes identified here are related to
489 nutritional value, sensory qualities, and provide hints for pecan cultivar improvements through
490 rational breeding design.

491

492 **Funding**

493 This research was supported by the Specialty Crop Block Grant Program at the U.S. Department
494 of Agriculture through grant agreement number 12-25-B-1464. Its contents are solely the
495 responsibility of the authors and do not necessarily represent the official views of the USDA.
496 Mention of trade names, commercial products, or companies in this article is solely for the
497 purpose of providing specific information and does not imply recommendation or endorsement
498 by the U.S. Department of Agriculture.

499

500 **Acknowledgments**

501 We thank Heping Cao and Jay Shockey for helpful discussion and critical evaluation of the
502 material presented in this paper.

References

- (1) Pollack, S. A. Noncitrus Fruits and Nuts 2011 Preliminary Summary. *Econ. Res. Ser. Rep.* **2012**, 7-9.
- (2) Blomhoff, R.; Carlsen, M. H.; Andersen, L. F.; Jacobs, D. R., Jr. Health benefits of nuts: potential role of antioxidants. *Brit. J. Nutr.* **2006**, 96, 52-60.
- (3) Kornsteiner-Krenn, M.; Wagner, K. H.; Elmadfa, I. Phytosterol content and fatty acid pattern of ten different nut types. *Int. J. Vitam. Nutr. Res.* **2013**, 83, 263-270.
- (4) Pollack, S. A. Fruit and Tree Nuts Situation and Outlook Yearbook 2008. *Econ. Res. Ser. Rep.* **2008**, 1-200.
- (5) Dominguez-Avila, J. A.; Alvarez-Parrilla, E.; Lopez-Diaz, J. A.; Maldonado-Mendoza, I. E.; Gomez-Garcia Mdel, C.; de la Rosa, L. A. The pecan nut (*Carya illinoensis*) and its oil and polyphenolic fractions differentially modulate lipid metabolism and the antioxidant enzyme activities in rats fed high-fat diets. *Food Chem.* **2015**, 168, 529-537.
- (6) Sicherer, S. H.; Munoz-Furlong, A.; Godbold, J. H.; Sampson, H. A. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *J. Allergy Clin. Immunol.* **2010**, 125, 1322-1326.
- (7) Fleischer, D. M. The natural history of peanut and tree nut allergy. *Curr. Allergy Asthma Rep.* **2007**, 7, 175-181.
- (8) Breiteneder, H.; Radauer, C. A classification of plant food allergens. *J. Allergy Clin. Immunol.* **2004**, 113, 821-830.
- (9) Sharma, G. M.; Irsigler, A.; Dhanarajan, P.; Ayuso, R.; Bardina, L.; Sampson, H. A.; Roux, K. H.; Sathe, S. K. Cloning and characterization of 2S albumin, Car i 1, a major allergen in pecan. *J. Agric. Food Chem.* **2011**, 59, 4130-4139.
- (10) Sharma, G. M.; Irsigler, A.; Dhanarajan, P.; Ayuso, R.; Bardina, L.; Sampson, H. A.; Roux, K. H.; Sathe, S. K. Cloning and characterization of an 11S legumin, Car i 4, a major allergen in pecan. *J. Agric. Food Chem.* **2011**, 59, 9542-9552.
- (11) Lee, B.; Zhang, R.; Du, W. X.; Grauke, L. J.; McHugh, T. H.; Zhang, Y. Z. Expression, purification and crystallization of pecan (*Carya illinoensis*) vicilin. *Acta Crystallogr Struct. Biol. Commun.* **2014**, 70, 1049-1052.
- (12) Zhang, Y.; Lee, B.; Du, W. X.; Lyu, S. C.; Nadeau, K. C.; Grauke, L. J.; Wang, S.; Fan, Y.; Yi, J.; McHugh, T. H. Identification and characterization of a new pecan [*Carya illinoensis* (Wangenh.) K. Koch] allergen, Car i 2. *J. Agric. Food Chem.* **2016**, 64, 4146-4151.
- (13) Venkatachalam, M.; Kshirsagar, H. H.; Seeram, N. P.; Heber, D.; Thompson, T. E.; Roux, K. H.; Sathe, S. K. Biochemical composition and immunological comparison of select pecan [*Carya illinoensis* (Wangenh.) K. Koch] cultivars. *J. Agric. Food Chem.* **2007**, 55, 9899-9907.
- (14) Robbins, K. S.; Shin, E. C.; Shewfelt, R. L.; Eitenmiller, R. R.; Pegg, R. B. Update on the healthful lipid constituents of commercially important tree nuts. *J. Agric. Food Chem.* **2011**, 59, 12083-12092.
- (15) Bublin, M.; Eiwegger, T.; Breiteneder, H. Do lipids influence the allergic sensitization process? *J. Allergy Clin. Immunol.* **2014**, 134, 521-529.
- (16) Dearman, R. J.; Alcocer, M. J.; Kimber, I. Influence of plant lipids on immune responses in mice to the major Brazil nut allergen Ber e 1. *Clin. Exp. Allergy* **2007**, 37, 582-591.
- (17) Mirotti, L.; Florsheim, E.; Rundqvist, L.; Larsson, G.; Spinozzi, F.; Leite-de-Moraes, M.; Russo, M.; Alcocer, M. Lipids are required for the development of Brazil nut allergy: the role of mouse and human iNKT cells. *Allergy* **2013**, 68, 74-83.
- (18) Angelina, A.; Sirvent, S.; Palladino, C.; Vereda, A.; Cuesta-Herranz, J.; Eiwegger, T.; Rodriguez, R.; Breiteneder, H.; Villalba, M.; Palomares, O. The lipid interaction capacity of Sin a 2 and Ara h 1, major mustard and peanut allergens of the cupin superfamily, endorses allergenicity. *Allergy* **2016**, 71, 1284-1294.
- (19) Okazaki, Y.; Saito, K. Roles of lipids as signaling molecules and mitigators during stress response in plants. *Plant J.* **2014**, 79, 584-596.

- (20) Kachroo, A.; Kachroo, P. Fatty acid-derived signals in plant defense. *Annu. Rev. Phytopathol* **2009**, *47*, 153-176.
- (21) Ohlrogge, J. B.; Jaworski, J. G. Regulation of fatty acid synthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1997**, *48*, 109-136.
- (22) Voelker, T.; Kinney, A. J. Variations in the biosynthesis of seed-storage lipids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **2001**, *52*, 335-361.
- (23) Wang, X.; Long, Y.; Yin, Y.; Zhang, C.; Gan, L.; Liu, L.; Yu, L.; Meng, J.; Li, M. New insights into the genetic networks affecting seed fatty acid concentrations in *Brassica napus*. *BMC Plant Biol.* **2015**, *15*, 1-18.
- (24) Costa, G. G.; Cardoso, K. C.; Del Bem, L. E.; Lima, A. C.; Cunha, M. A.; de Campos-Leite, L.; Vicentini, R.; Papes, F.; Moreira, R. C.; Yunes, J. A.; Campos, F. A.; Da Silva, M. J. Transcriptome analysis of the oil-rich seed of the bioenergy crop *Jatropha curcas* L. *BMC Genomics* **2010**, *11*, 1-9.
- (25) Cao, H.; Shockey, J. M.; Klasson, K. T.; Chapital, D. C.; Mason, C. B.; Scheffler, B. E. Developmental Regulation of diacylglycerol acyltransferase family gene expression in tung tree tissues. *PLoS ONE* **2013**, *8*, e76946.
- (26) Li-Beisson, Y.; Shorrosh, B.; Beisson, F.; Andersson, M. X.; Arondel, V.; Bates, P. D.; Baud, S.; Bird, D.; Debono, A.; Durrett, T. P.; Franke, R. B.; Graham, I. A.; Katayama, K.; Kelly, A. A.; Larson, T.; Markham, J. E.; Miquel, M.; Molina, I.; Nishida, I.; Rowland, O.; Samuels, L.; Schmid, K. M.; Wada, H.; Welti, R.; Xu, C.; Zallot, R.; Ohlrogge, J. Acyl-lipid metabolism. *Arabidopsis Book* **2013**, *11*, 1-70.
- (27) Mattison, C. P.; Tarver, M. R.; Florane, C.; Graham, C. J. Temporal expression of pecan allergens during nut development. *J. Hort. Sci. Biotech.* **2013**, *88*, 173-178.
- (28) Wells, L. Cultural Management of Commercial Pecan Orchards. *Southeastern Pecan Growers' Handbook*, Wells, L. Ed. University of Georgia Publishing: Tifton, GA, USA, 2007; pp 61-82.
- (29) McKay, J. W. Embryology of pecan. *J. Agr. Res.* **1947**, *74*, 263-283.
- (30) Chang, S.; Puryear, J.; Cairney, J., A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **1993**, *11*, 113-116.
- (31) Aronesty, E. ea-utils: Command-line tools for processing biological sequencing data. (<https://expressionanalysis.github.io/ea-utils/>) (2011).
- (32) Kong, Y. Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics.* **2011**, *98*, 152-153.
- (33) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644-652.
- (34) Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357-359.
- (35) Anders, S.; Pyl, P. T.; Huber, W. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics.* **2015**, *31*, 166-169.
- (36) Love, M. I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550.
- (37) Bustin, S. A.; Benes, V.; Garson, J. A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M. W.; Shipley, G. L.; Vandesompele, J.; Wittwer, C. T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **2009**, *55*, 611-622.
- (38) Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **2002**, *3*, 1-11.
- (39) Rajaram, S.; Burke, K.; Connell, B.; Myint, T.; Sabate, J. A monounsaturated fatty acid-rich pecan-enriched diet favorably alters the serum lipid profile of healthy men and women. *J. Nutr.* **2001**, *131*, 2275-2279.

- (40) Ryan, E.; Galvin, K.; O'Connor, T. P.; Maguire, A. R.; O'Brien, N. M. Fatty acid profile, tocopherol, squalene and phytosterol content of brazil, pecan, pine, pistachio and cashew nuts. *Int. J. Food Sci. Nutr.* **2006**, *57*, 219-228.
- (41) Rudolph, C. J.; Odell, G. V.; Hinrichs, H. A. H., D.A. ; Kays, S. J. Genetic, environmental, and maturity effects on pecan kernel lipid, fatty acid, tocopherol, and protein composition. *J. Food Qual.* **1992**, *15*, 263-278.
- (42) Huang, J.; Zhang, T.; Zhang, Q.; Chen, M.; Wang, Z.; Zheng, B.; Xia, G.; Yang, X.; Huang, C.; Huang, Y. The mechanism of high contents of oil and oleic acid revealed by transcriptomic and lipidomic analysis during embryogenesis in *Carya cathayensis* Sarg. *BMC Genomics* **2016**, *17*, 113.
- (43) Galli, V.; Guzman, F.; Messias, R. S.; Körbes, A. P.; Silva, S. D. A.; Margis-Pinheiro, M.; Margis, R. Transcriptome of tung tree mature seeds with an emphasis on lipid metabolism genes. *Tree Genetics & Genomes* **2014**, *10*, 1353–1367.
- (44) Singanusong, R.; Mason, R. L.; D'Arcy, B. R.; Nottingham, S. M. Compositional changes of Australia-grown Western Schley pecans [*Carya illinoensis* (Wangenh.) K. Koch] during maturation. *J. Agric. Food Chem.* **2003**, *51*, 406-412.
- (45) Kachroo, P.; Shanklin, J.; Shah, J.; Whittle, E. J.; Klessig, D. F. A fatty acid desaturase modulates the activation of defense signaling pathways in plants. *Proc. Natl. Acad. Sci. U S A* **2001**, *98*, 9448-9453.
- (46) Kachroo, A.; Shanklin, J.; Whittle, E.; Lapchyk, L.; Hildebrand, D.; Kachroo, P. The Arabidopsis stearoyl-acyl carrier protein-desaturase family and the contribution of leaf isoforms to oleic acid synthesis. *Plant Mol. Biol.* **2007**, *63*, 257-271.
- (47) Mata-Perez, C.; Sanchez-Calvo, B.; Begara-Morales, J. C.; Luque, F.; Jimenez-Ruiz, J.; Padilla, M. N.; Fierro-Risco, J.; Valderrama, R.; Fernandez-Ocana, A.; Corpas, F. J.; Barroso, J. B. Transcriptomic profiling of linolenic acid-responsive genes in ROS signaling from RNA-seq data in Arabidopsis. *Front. Plant Sci.* **2015**, *6*, 122.
- (48) War, A. R.; Paulraj, M. G.; Ahmad, T.; Buhroo, A. A.; Hussain, B.; Ignacimuthu, S.; Sharma, H. C. Mechanisms of plant defense against insect herbivores. *Plant Signal Behav.* **2012**, *7*, 1306-1320.
- (49) Zhang, J.; Liu, H.; Sun, J.; Li, B.; Zhu, Q.; Chen, S.; Zhang, H. Arabidopsis fatty acid desaturase FAD2 is required for salt tolerance during seed germination and early seedling growth. *PLoS ONE* **2012**, *7*, e30355.
- (50) Fujimura, K. E.; Sitarik, A. R.; Havstad, S.; Lin, D. L.; Levan, S.; Fadrosh, D.; Panzer, A. R.; LaMere, B.; Rackaityte, E.; Lukacs, N. W.; Wegienka, G.; Boushey, H. A.; Ownby, D. R.; Zoratti, E. M.; Levin, A. M.; Johnson, C. C.; Lynch, S. V. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat. Med.* **2016**, *22*, 1187-1191.

635 **Figure Captions**

636 **Figure 1.** Cross-section of developing Desirable and Sumner pecan embryos. Samples were
637 collected based on morphology from an individual ‘Sumner’ or ‘Desirable’ cultivar pecan tree at
638 the Louisiana State University Pecan Research Station pecan orchard (Shreveport, LA) during
639 August and September of the 2012 growing season. The trees received standard agronomic
640 practices and the nuts are representative of “water” stage (August 11, 17), “gel” stage (August
641 23), “dough” stage (September 4, 20), and “mature” nuts (September 20, October 2). Samples
642 boxed in yellow from the Sumner cultivar (8-11, 8-23, 9-4, and 9-20) were used for gene
643 expression analysis.

644 **Figure 2.** FAME analysis of developing Desirable and Sumner pecan embryos.
645 Lipids extracted from pecan samples at the indicated time points were analyzed for FAMES, with
646 those over 2% including palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic
647 acid (C18:2), and alpha-linolenic acid (C18:3) content using an Agilent GC-MS with standard
648 curves developed in Chemstation software. Quantification was based on the prominent mass ion
649 of each FAME.

650 **Figure 3.** GO annotation of developing Sumner pecan embryo transcriptome.
651 Sequenced and assembled RNA samples from developing Sumner pecan embryos were
652 annotated using the SwissProt, Uniprot, and Uniref90 databases with Trinotate software.
653 Annotated genes were categorized based upon A) biological process (BP), B) cellular component
654 (CC), and C) molecular function (MF) categories.

655 **Figure 4.** Characterization of significant gene expression changes during Sumner pecan embryo
656 development. HTSeq and DESeq2 were used to count and determine significance of gene
657 expression changes as described in the Trinity abundance estimation and identifying DE feature

protocols. The relative magnitude of gene expression changes during Sumner pecan embryo development is shown for transitions between the following dates A) 8/11-8/23, B) 8/23-9/4, and C) 9/4-9/20.

Figure 5. Pecan allergen gene expression analysis in developing embryos. The gene expression changes of 3 pecan nut allergens, Car i 1 (comp34075_c0, green line), Car i 4 (comp34066_c0, blue line), and Car i 2 (comp34067_c0, red line) during development were plotted based upon A) absolute and B) normalized gene expression values as noted on the y-axis and dates of sample collection on the x-axis.

Figure 6. Clustered gene expression profiles from developing Sumner pecan embryos. Genes were clustered based upon their temporal expression profile. Temporal gene expression profiles were created and evaluated in R using kmeans to define 12 unique clusters. Scaled gene expression units are indicated on the Y-axis and date of collection on the X-axis.

Figure 7. Quantitative PCR of select allergen and fatty acid metabolism genes from developing Sumner and Desirable pecan embryos. Gene expression profiles from Sumner and Desirable cultivars were evaluated by qPCR. Four developmental time points are represented. A) Expression of known pecan genes whose products elicit an allergic response in humans. B) Expression of genes involved in various stages of fatty acid biosynthesis. Developmental time points denoted with an asterisk have ≥ 2 -fold difference in transcript abundance levels between the two pecan varieties and are significantly different as determined by paired t-test.

Table 1. Proximate analysis values on a wet weight basis of developing Desirable and Sumner pecan cultivars

Cultivar Desirable	Moisture %	Ash %	Lipid %	Protein %*	CHO%***
8/11/2012	45.65	3.06	2.53	**	NC
8/17/2012	50.94	1.90	2.67	3.31+/-0.37	41.18
8/23/2012	45.68	2.98	10.27	5.90+/-0.57	35.17
8/29/2012	40.41	2.47	26.42	9.05+/-0.89	21.65
9/10/2012	29.55	1.52	43.74	9.86+/-0.93	15.33
9/20/2012	22.50	1.60	49.12	10.35+/-0.66	16.44
10/2/2012	16.99	1.83	50.83	12.59+/-1.30	17.76
Cultivar Sumner	Moisture %	Ash %	Lipid %	Protein %*	CHO%***
8/11/2012	43.09	3.57	2.50	**	NC
8/17/2012	43.89	4.31	4.74	**	NC
8/23/2012	42.26	3.55	11.17	6.80+/-2.3	36.22
8/29/2012	37.40	2.82	26.94	7.63+/-1.27	25.21
9/4/2012	31.78	2.35	35.64	8.53+/-1.27	21.71
9/20/2012	18.90	**	51.34	9.82+/-1.10	NC
10/2/2012	19.88	1.48	54.38	10.50+/-0.94	13.76

* Mean values +/- standard deviation

** Insufficient sample for determination

*** Calculated as 100 -% moisture -% lipid - % protein

NC – not calculated

Table 2. Primers for quantitative Polymerase Chain Reaction (qPCR)

Gene Target	Predicted product	Gene Cluster	Forward primer sequence 5'-3'	Reverse primer sequence 5'-3'
comp31874_c1	13S legumin Car i 4 allergen, nutrient reservoir activity	1	CCACAACAATGCCAATCAG	GTATTCCTGCTGACCTTGT
comp17358_c0	flavonoid biosynthetic process, integral to membrane	2	CCAGAAGCAGCAGAGAGTTTA	GTGTTGCGTCTTCTGGGTAT
comp135859_c0	fatty acid biosynthetic process, acyl-[acyl-carrier-protein] desaturase activity	3	ACAATTGATAGAGACCCGACAT T	GTGTACACGAAGCCCAAGTA
comp15104_c1	transmembrane receptor protein serine/threonine kinase activity	4	TCAGAACCTCCACTCGTAGAT	CTAATGGCAACCCGGATATAGG
comp34067_c0	7S vicilin Car i 7 allergen, nutrient reservoir, vicilin-like antimicrobial peptides	5	GGCAGAACTATCATGGTGAC	AGCATACTAAGGCGAGGAA
comp34075_c0	2S albumin Car i 1 allergen, nutrient reservoir activity	6	AGTAGCAGCTCTCCTTGTA	GGTTGTCAATGTCCTCGTC
comp39405_c0	protein serine/threonine phosphatase activity	7	TCGTAGGACGATGTCAGAGATA A	GAGCGACAGAAAGGGAACA
comp42176_c0	plasma membrane, protein transport	8	CTGGCTGGAGTTCTCTCATTATC	TGTGTCTGTGCTTCTCAGTTT
comp49987_c0	cyclin-dependent protein kinase inhibitor activity	9	TCCAGGTTTCATCAGCATCAC	GGGAACCTCAAGCAGGAGAAA
comp30247_c0	ethylene-responsive transcription factor	10	CTTCTTCACGCCCTCTCTTC	CACCTCTGTCTGCTTACTCTTC
comp40754_c0	regulation of ARF GTPase activity, metal ion binding	11	CTATCTCGGCATCACCCATT	GACGCACTGACTCTTCTGTGA
comp17793_c0	chloroplast thylakoid membrane, photosynthetic electron transport in photosystem II	12	CGATCCTCAGCAAATAGGAGTT	GCCGCTGGGTTAGGTTTAT
comp18672_c0	fatty acid oxidation	10	CCTGGTAACGAAGCCAAGAA	GGAAGCCCTCTTCAGCATATT
comp39590_c0	unknown	10	TCTGCTGCATGTGAGAGTATTG	GCTAGAGGATGATGGTGAAGA AG
comp36103_c0	protein-disulfide reductase activity	10	CCTCGTAGATTCCAACCAGAAC	GGATCTTGAGAGCAAGGTCATA G
comp53691_c0	serine carboxypeptidase-like, proteolysis	10	TAGTGACGTATCCTGCGATTG	GGGCACTGAAGAATGGATAGA A
comp66643_c0	unknown	9	TCACCGTAACCAGAACCTTTAC	GATCAGGATCAGGAGGCAATC
comp17308_c0	alpha tubulin 2	control	CAGACCAACCTCCTCATAATC	GGCATTCTGTTCACTGGTATGT
AF174619.1	18S ribosomal RNA	control	ACATCTTACCACGATACATAAC	AACTTGC GTTCAAAGACTC
AF076774	trnI-tmF chloroplast intergenic regions	control	CGGACGAGAATAAAGATAGAG T	TTTTGGGGATAGAGGGACTTG

Figure 1.

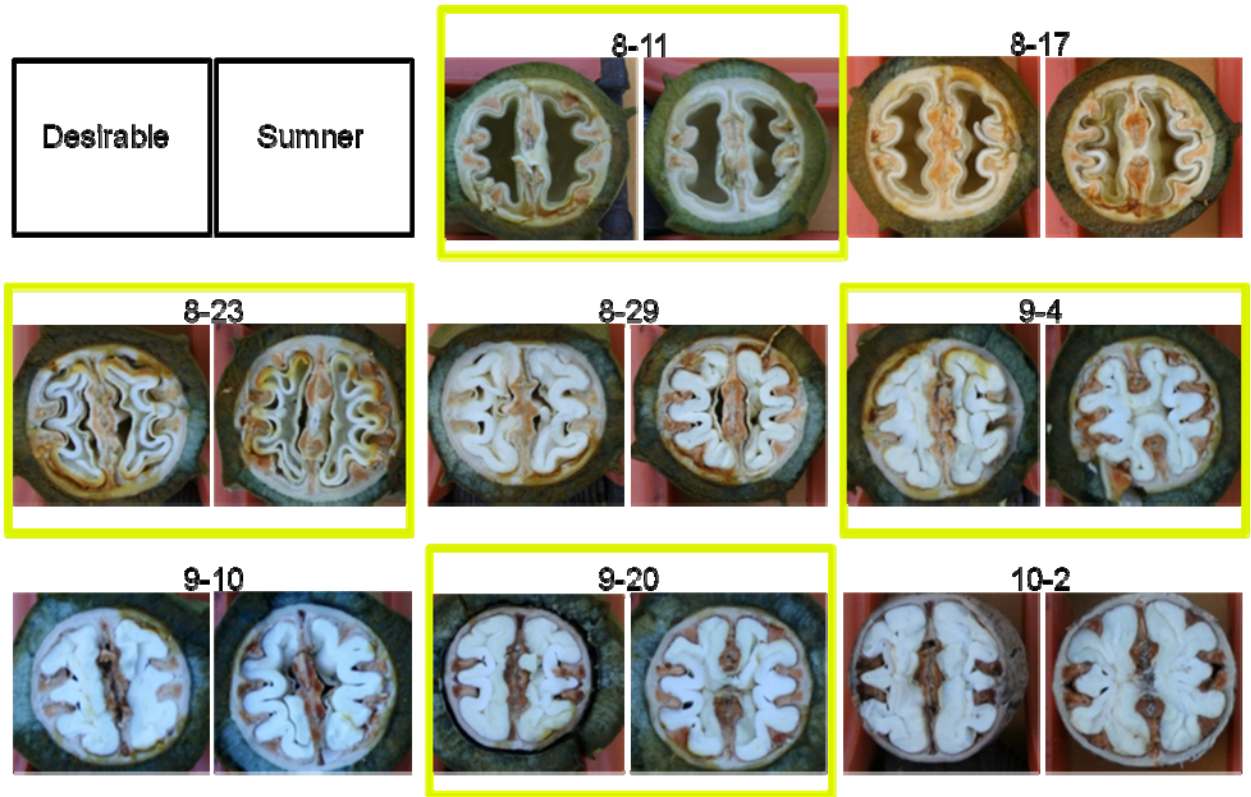


Figure 2.

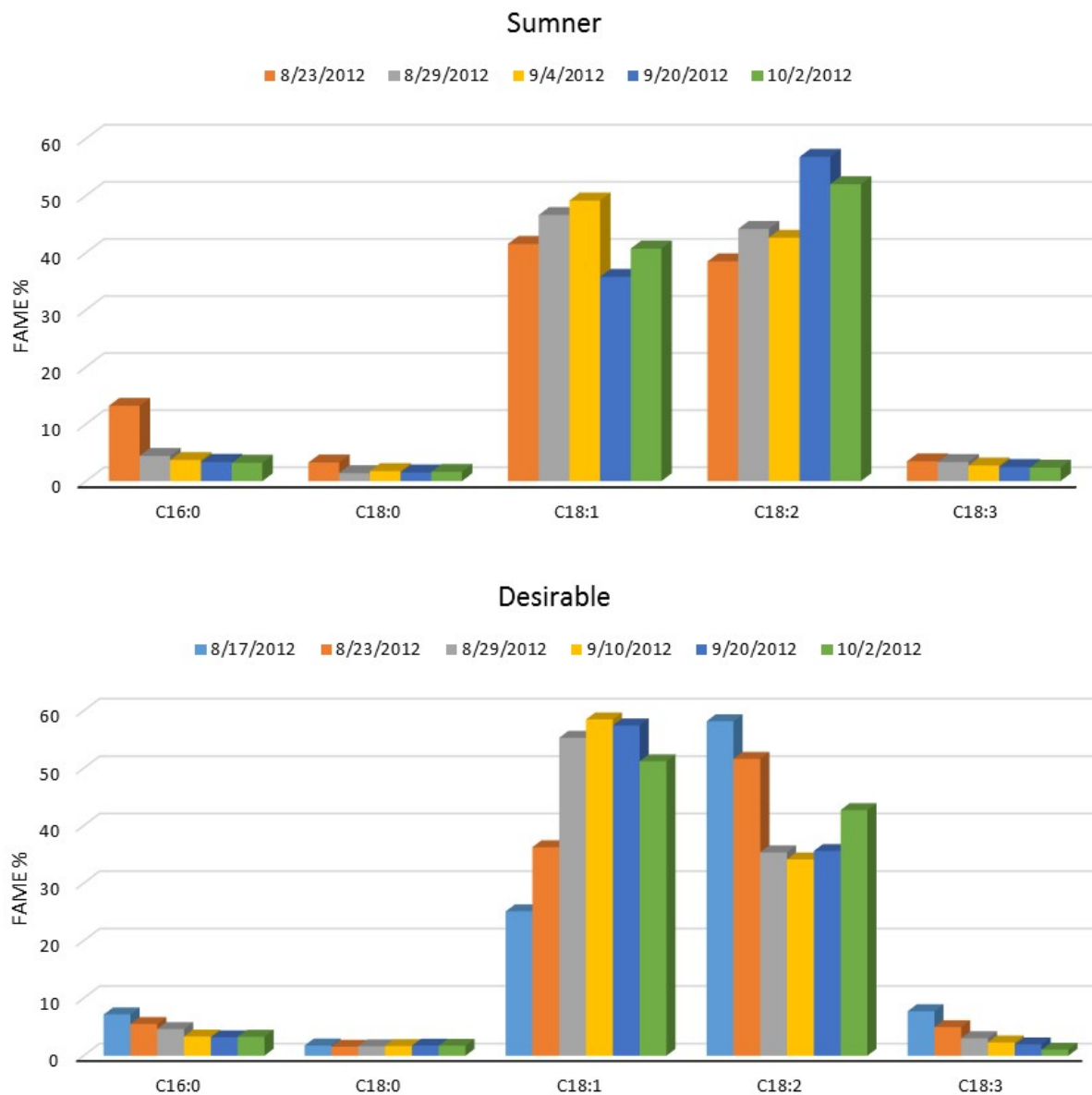
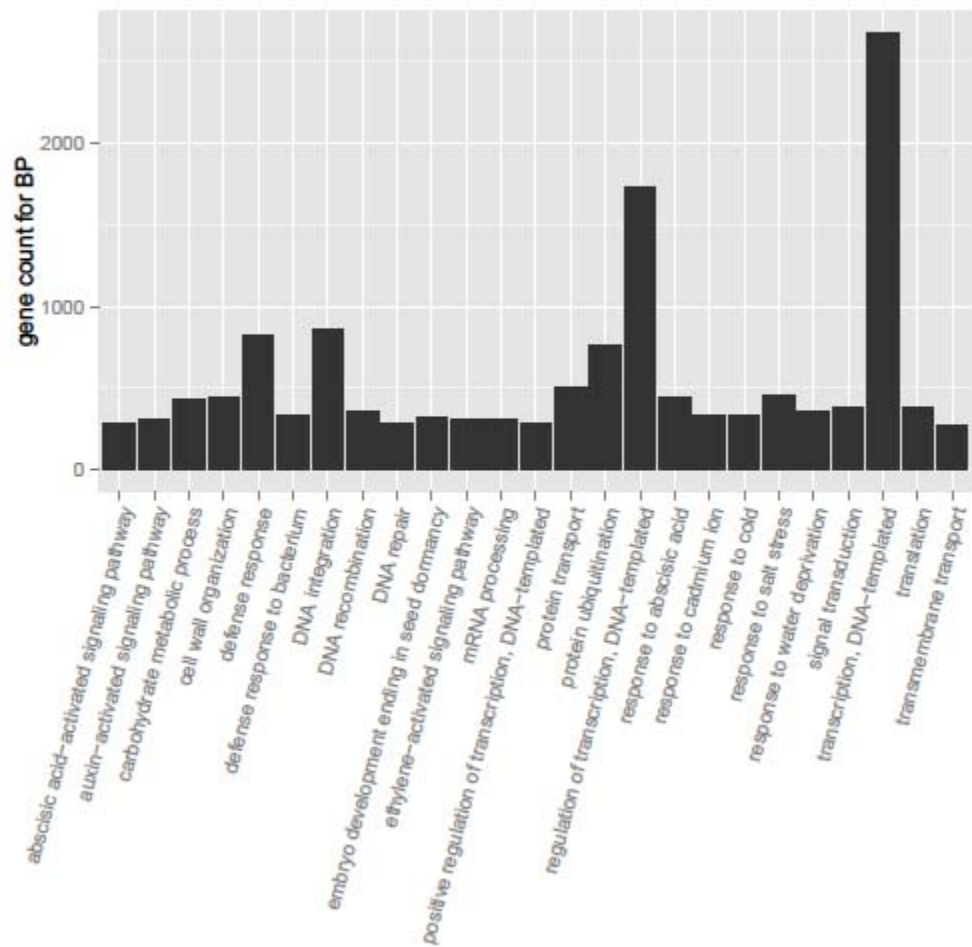
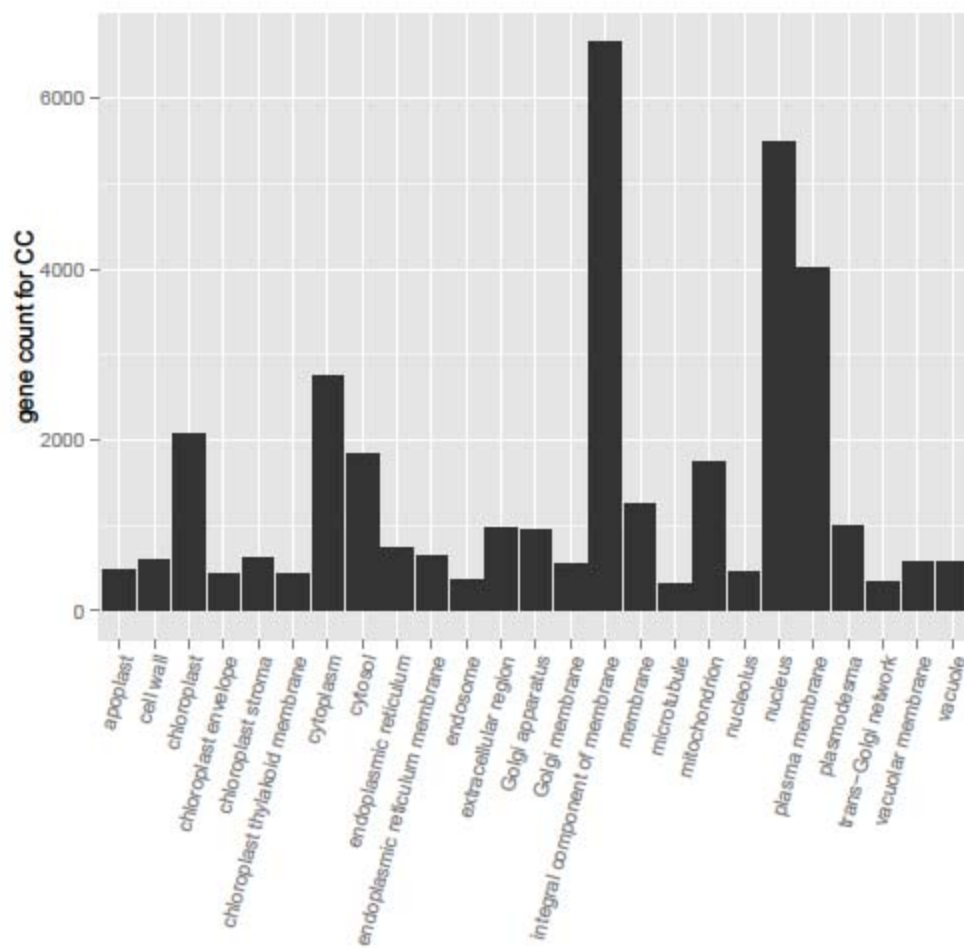


Figure 3.

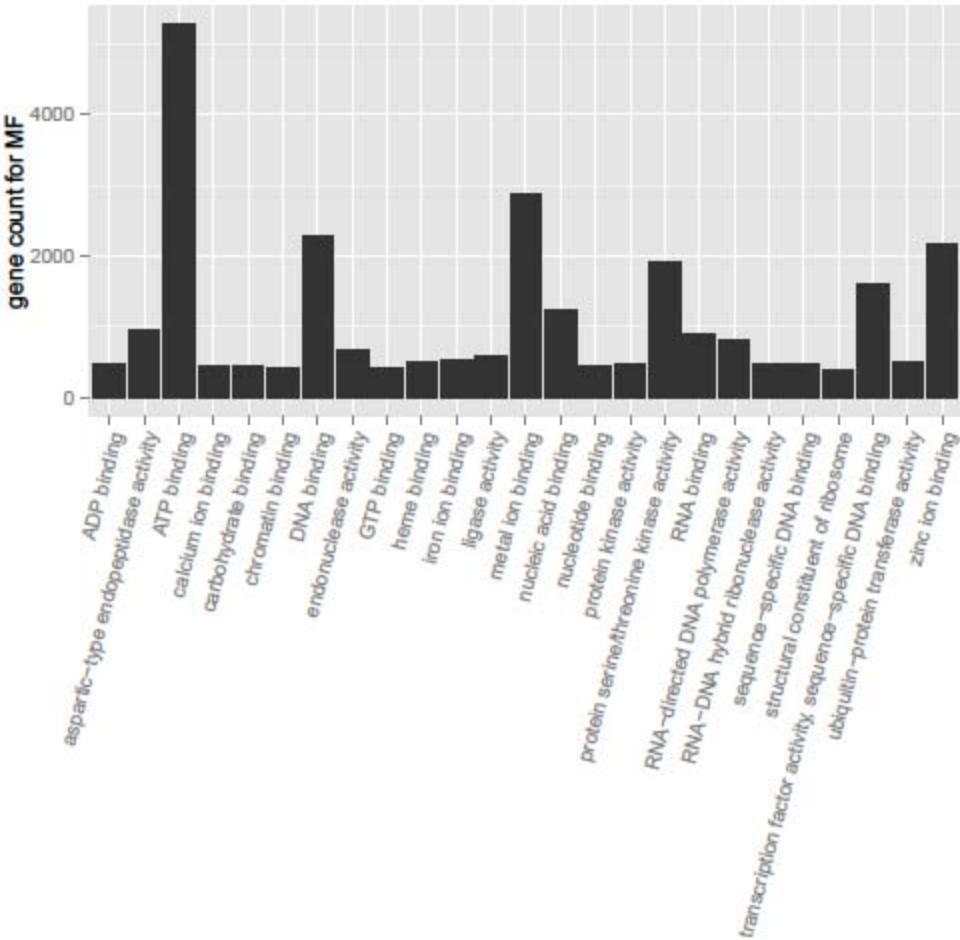
A



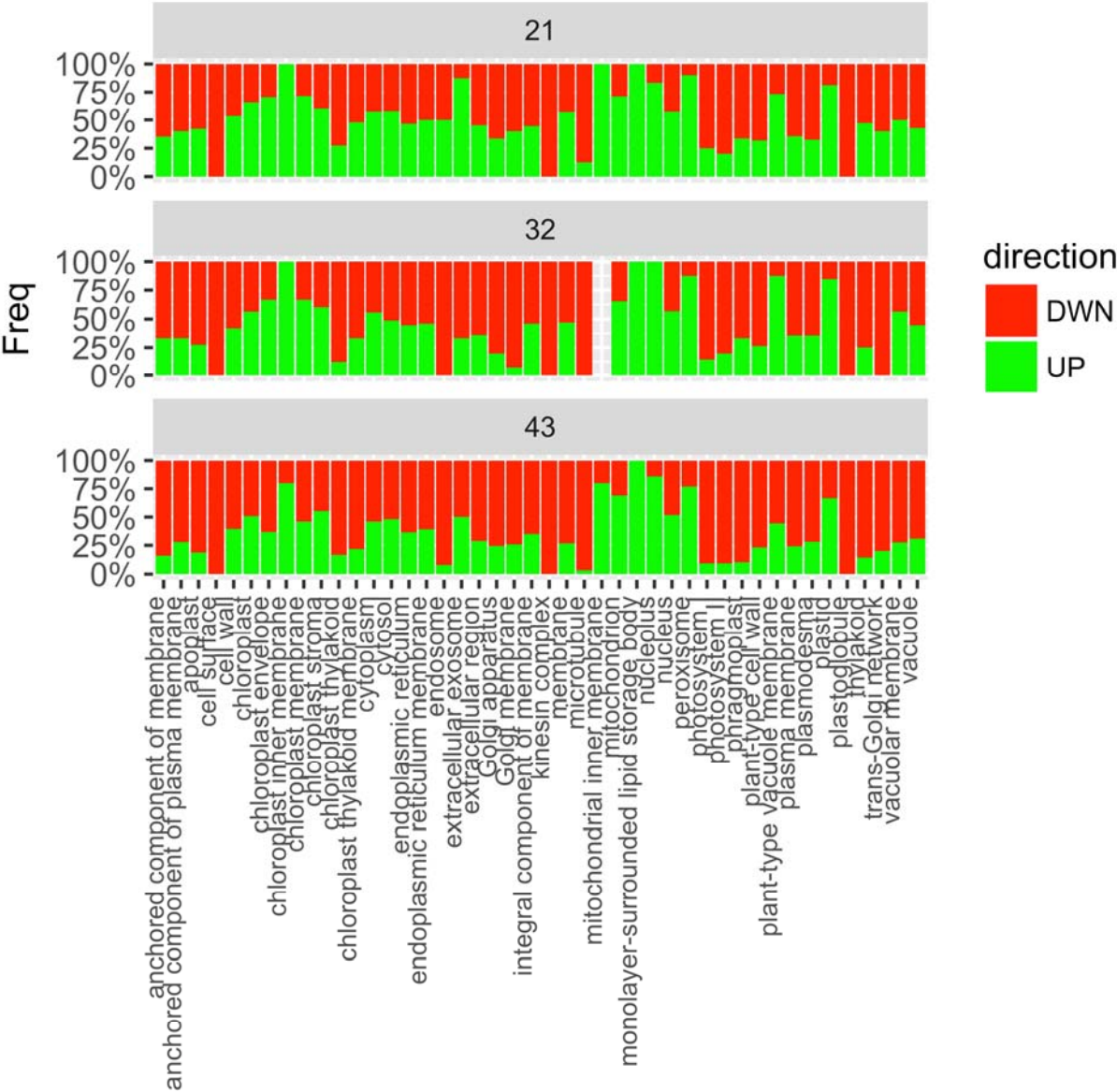
B



C



B.



C.

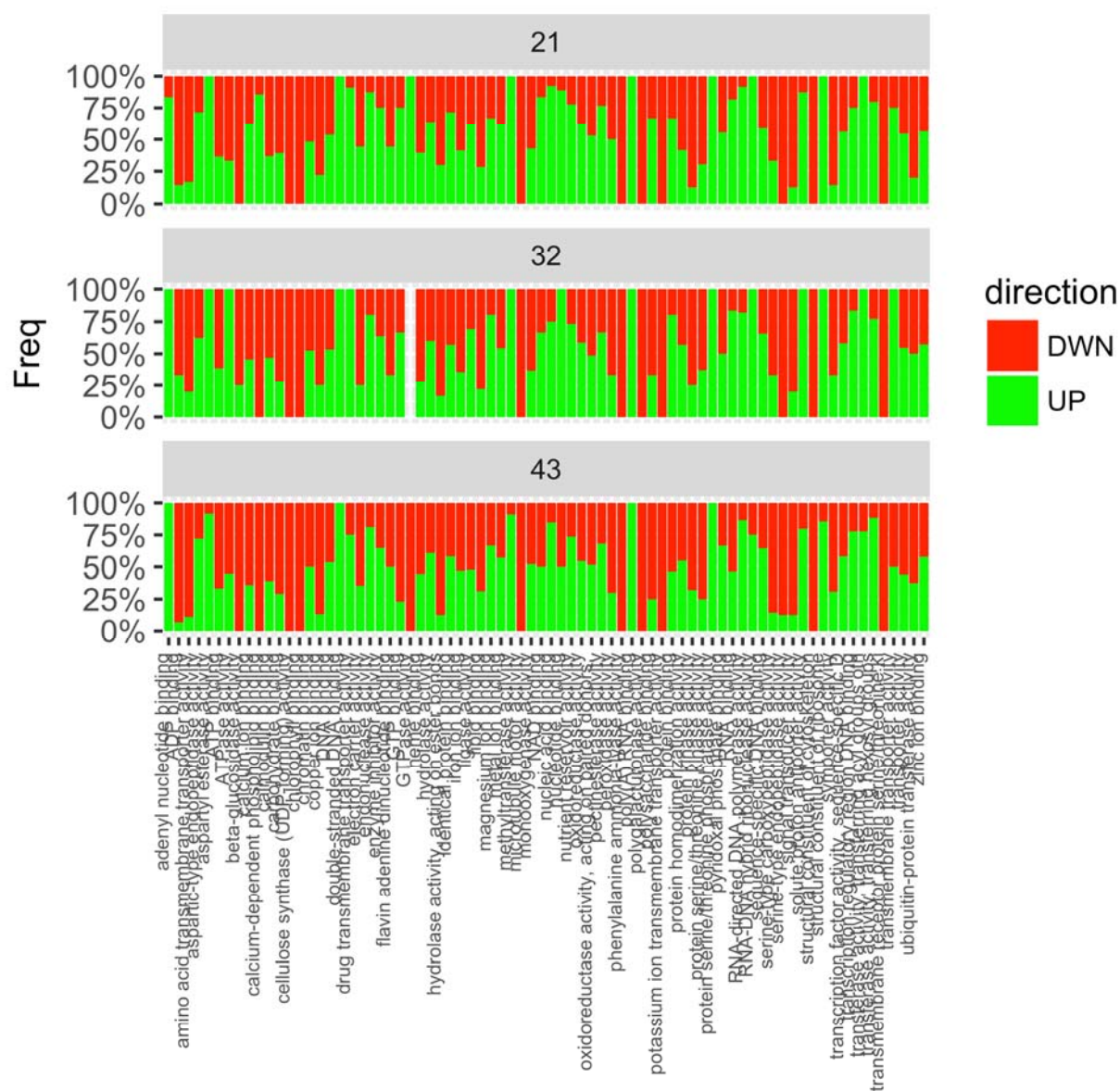


Figure 5.

A.

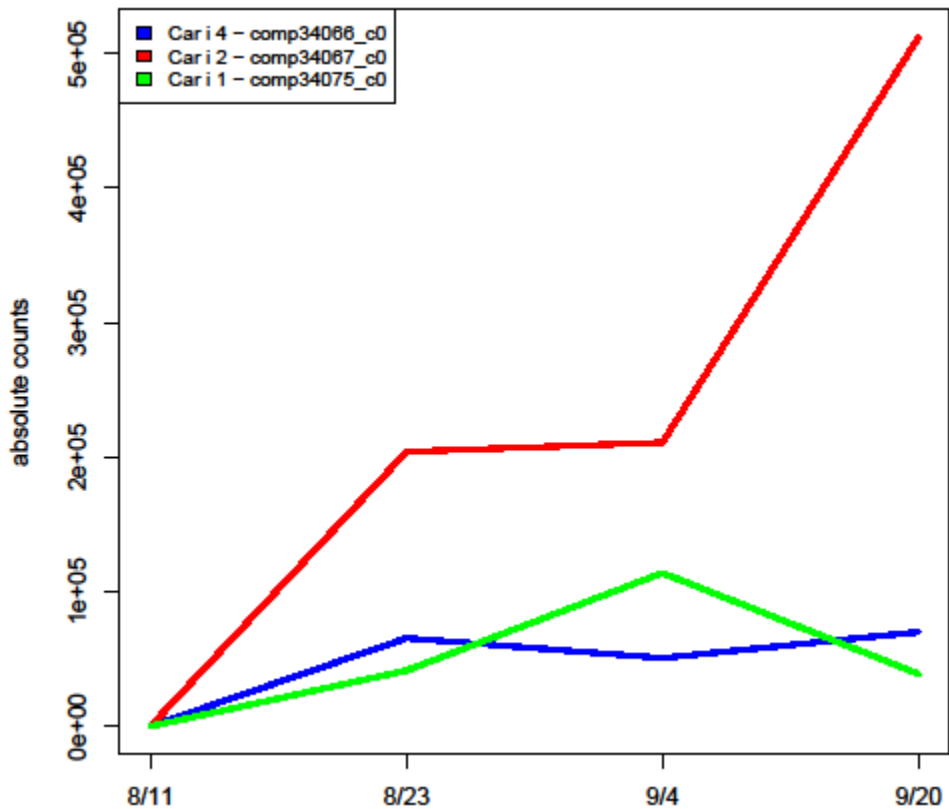


Figure 5.

B.

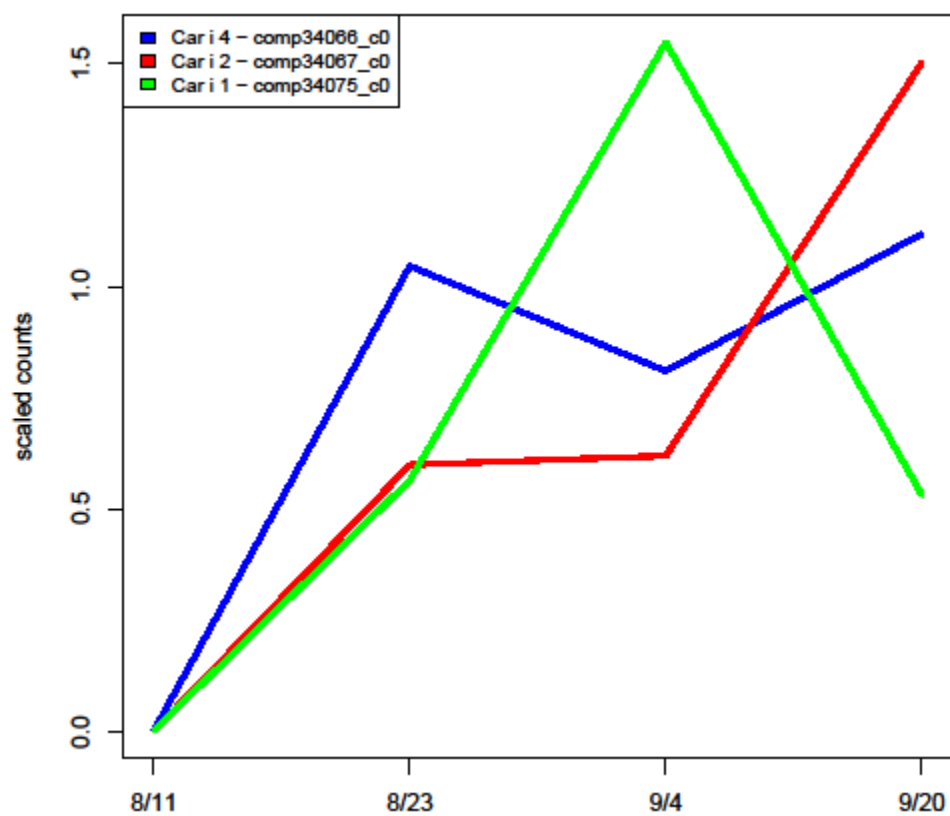


Figure 6.

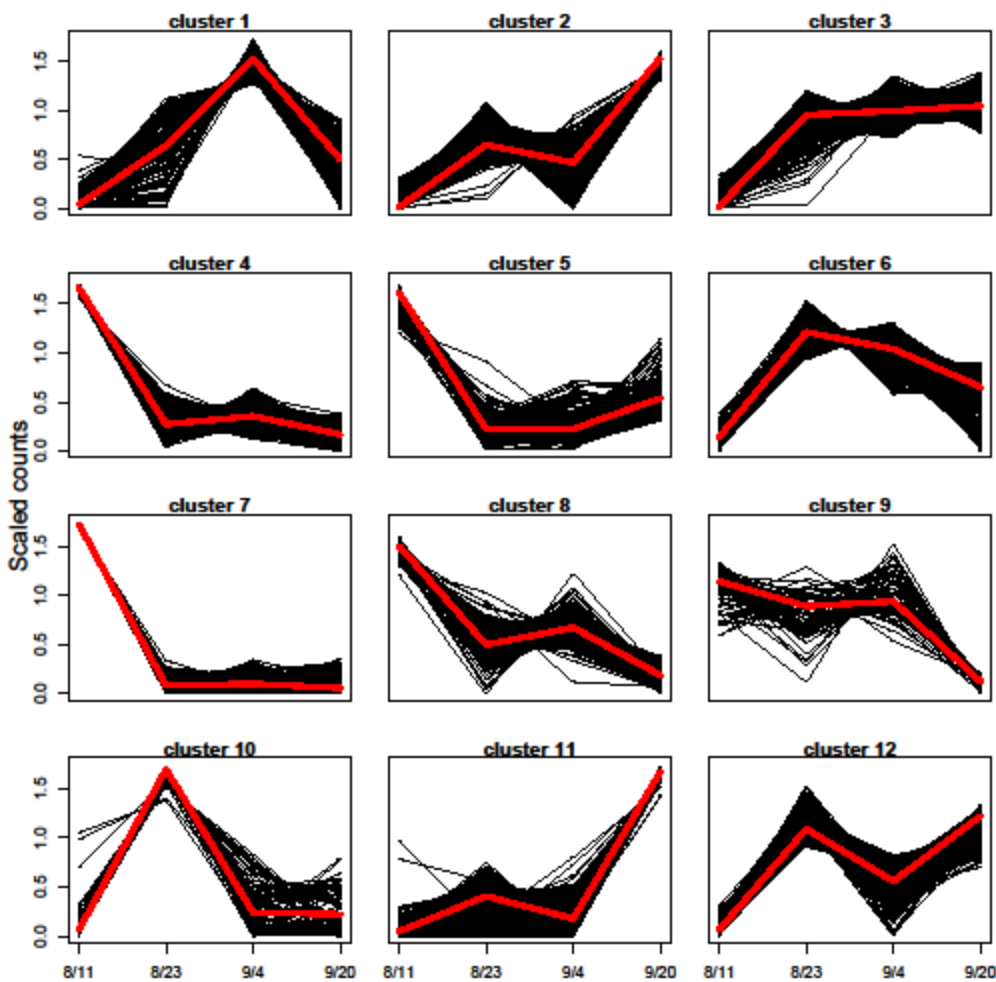


Figure 7.

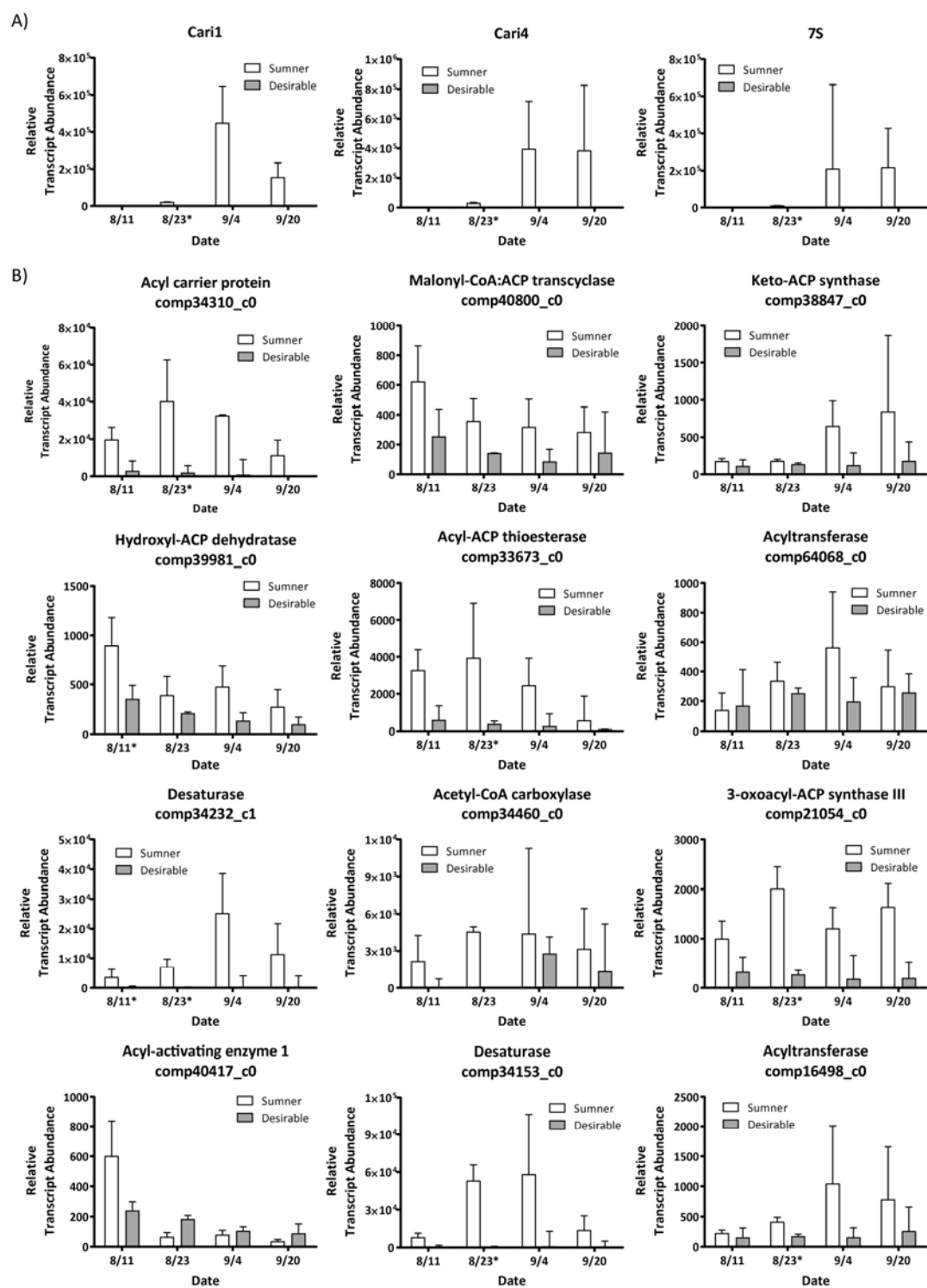


Table of Contents Graphic

