Gene expression evolution during sunflower (*Helianthus annuus)* domestication

Ray Hess Bair Watson
Greencastle, PA

Bachelor of Arts, Swarthmore College, 2013

A Thesis presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Master of Science

Department of Biology
University of Virginia
July, 2019

Abstract

Domestication is characterized by dramatic phenotypic evolution from a wild progenitor to a species adapted for human use. In crops, much research has been dedicated to identifying traits associated with domestication and identifying loci underlying these traits on an individual basis as part of breeding programs. However, identifying genes or loci contributing to domestication phenotypes which are also under selection at the genome/transcriptome-wide level has only been recently explored in a few crops. Further, little is understood about the genome-wide effects of domestication. In this study I used developmental and tissue-specific RNA sequencing in combination with existing selective-sweep data to characterize differentially expressed (DE) genes between wild and domesticated sunflowers, associate DE genes with domestication phenotypes, identify DE genes with a signature of selection, and characterize how domestication altered the transcriptional landscape. Domestication significantly altered the expression of between 16 and 48 percent of genes in a tissue-specific manner, with a majority (54-70%) of those DE genes being expressed at higher levels in domesticated plants. Of the 183 previously identified domestication candidates showing signatures of selection in their coding regions, 120 were differentially expressed in at least one tissue, with many (85/120) differentially expressed in multiple tissues. The sunflower transcriptome was also substantially altered by domestication, with an overall increase in gene co-expression network connectivity. However, some individual putative domestication genes also experienced a reduction in co-expression network connectivity. These results indicate that domestication had far reaching and dramatic impacts on the transcriptional landscape of sunflowers.

Keywords: domestication, comparative transcriptomics, gene co-expression, RNA sequencing

Acknowledgements

This work was made possible with professional and personal support from many people. First and foremost, I must thank my partner Brittany Watson. The ferocity of her support and love has enabled me to be a better person and she has served as a never ending example of what a true educator and mentor can be to her students. The beauty of her spirit and depth of her patience and kindness made this work possible. She was and is a constant source of light.

Many mentors, past and present, contributed greatly to both this research itself, as well as the skills it required. My advisor at UVA, Laura Galloway never failed to encourage me to grow as scientist and person. Laura's lab was a safe place to look for thoughtful feedback and support – even if it was on a very rough draft. She prevented me from getting lost in the weeds on many occasions. Laura's mentorship helped me believe in the beauty of my research and myself. My committee members, Butch Brodie and Bob Cox, also provided valuable feedback on this manuscript and support in many other endeavors. Vince Formica's unguarded enthusiasm for life and science motivated me to go into research in the first place, and his sense of wonder towards all things scientific and nature-related continue to inspire me. Finally, the lessons I learned in Kerrie Hefner's classroom still serve me daily. She taught me how to write and live with heart.

I would also like to thank the UVA Biology graduate student community. I received enormous scientific help with my work, but of even greater value was their comradery and friendship. Specifically, this work would not have been possible without the help of my co-conspirators, sometimes lab-mates, and friends: Brian Sanderson, Corlett Wood, Malcolm Augat, and Catherine Debban. Catherine and I entered graduate school together and she became one of my closest friends, her encouragement through many challenges was invaluable. Brian, Corlett and Malcom never complained when they found me working at their desks, and were more

generous with their time, space, and textbooks than a young graduate student had right to ask. Malcolm's indominable laughter never failed to brighten an awkward silence. Corlett's constant curiosity drove me to ask questions that did not have answers. Brian was always willing to provide much needed assistance and a deep well of compassion. He was particularly instrumental in the completion of this thesis. The Jefferson Scholars Foundation and its staff, particularly Ben Skipper, have been of incredible support in funding this work.

There were many helpers outside the academy that made this work possible and deserve thanks. Kelsie Morioka was an unfaltering pillar of love and support, without whom, this work would have never materialized. Kelsie was always generous with her time, energy, and healing mugs of tea. Sarah Floyd and Jim Mead provided some of the greatest generosity of spirit I have had the good fortune to encounter and reminded me of my purpose. Karl Adler put graduate school in a wider perspective and reminded me of the importance of self-care. Katie Goldman always kept me grounded no matter where she was in the world. I owe inestimable thanks to Danielle Ryder for her unquestioning love. She has been my sister, friend, and guardian since birth. She is the strongest and kindest person I have ever met. Danielle was always quick, but never hasty, to provide sage advice. Finally, while research can sometimes feel esoteric, Wallace Watson provided constant reminders of why domestication matters to us all, and for that unconditional love, I will always be thankful.

Introduction

Domestication has resulted in rapid and dramatic phenotypic changes from wild progenitors to modern crops and livestock. Across many crops, this transition has involved a suite of similar traits, termed the "domestication syndrome," which includes increased fruit and seed yields (Frary *et al.*, 2000), changes to flowering time (Blackman *et al.*, 2011b), as well as loss of branching, seed dormancy and dispersal (reviewed in Meyer & Purugganan (2013). A thorough understanding of the complex history of domestication traits and their underlying molecular mechanisms can provide valuable information to scientists for crop breeding and conservation. As human populations grow, climates shift, and global food security continues to be of concern, there is inarguable societal and financial value to understanding how past selection under domestication can be used to guide future improvement studies.

In addition to the applied benefits, the study of domestication has fascinated evolutionary biologists since the conception of the field as it provides repeated parallel case studies of selection and phenotypic evolution (Darwin, 1859; Meyer & Purugganan, 2013). The human-mediated, artificial selection during domestication often involved strong directional selection on visible phenotypes over a relatively short time period. This strong artificial selection on visible phenotypes can potentially increase the power to detect signatures of selection. Further, human involvement and interests in agriculture have led to detailed archeological and historical records for some crops, which enhance our understanding of phenotypic changes during domestication and can provide information about demographic shifts during the process (Smith, 2006; Purugganan & Fuller, 2011). The contextual information about selection during domestication makes it a valuable model for investigating broader evolutionary concepts. For example, the

impact of development and trait correlations on selection during domestication was recognized as early as *On the Origin of Species* (Darwin, 1859).

Since the modern synthesis, much work has been devoted to characterizing the traits under selection during domestication and their underlying loci, particularly in regard to plant breeding and improvement. Top-down approaches such as quantitative trait locus (QTL) mapping studies and genome wide association studies (GWAS) leverage visible or known phenotypes and controlled genetic crosses or mapping populations to determine the genomic regions or loci underlying domestication phenotypes (Ross-Ibarra *et al.*, 2007; Kantar *et al.*, 2017). However, in isolation, these studies cannot prove that a locus (or multiple loci) underlying a phenotype of interest was actually under selection during domestication, and these approaches are biased towards detecting loci of large effect (Doebley *et al.*, 2006; Ross-Ibarra *et al.*, 2007; Kantar *et al.*, 2017). Bottom-up studies in crops using extant genetic variation and population genetic/genomic analyses identify loci showing signatures of selection without regard to their phenotypic consequences, but are limited by technical considerations such as marker density and availability of extant populations in identifying causal genes (Ross-Ibarra *et al.*, 2007; Gepts, 2014). Combining both approaches can directly link selection to genotype and phenotype, but such multi-pronged approaches can be severely impacted by cost and feasibility concerns. To date, this has limited many studies of evolution during domestication to loci of major effect in the most highly characterized and valuable crops, such as maize (Doebley *et al.*, 2006; Schaefer *et al.*, 2017).

With the continued improvement of next-generation sequencing technologies, comparative transcriptomics (RNA-sequencing) between domesticated species and their wild ancestors is a viable genome-wide method for studying the impacts of evolution during

domestication and provides a bridge between top-down and bottom-up approaches. RNA-seq does not require *a priori* knowledge of phenotypes under selection and is not limited to discovering genes only related to visible or measurable phenotypic changes (Schaefer *et al.*, 2017). However, this independence can limit the ability of comparative transcriptomic studies to infer gene function. Examining differential gene expression in a tissue- and/or development-specific context can aid in identifying probable function in relation to domestication phenotypes. Gene co-expression networks can also provide putative functional information for unknown genes through their association with genes already characterized (Schaefer *et al.*, 2017), as well as insight into how domestication altered the transcriptional landscape as a whole. However, differential gene expression can be due to either selection or drift and could be caused by a coding mutation in the gene of interest, or due to regulatory changes. By combining developmentally and tissue-specific RNAseq with signature-of-selection studies, putative domestication candidates can be revealed and linked to domestication phenotypes of interest.

The common sunflower, *Helianthus annuus*, was domesticated roughly 4000 years ago in Eastern North America (Smith, 2006; Blackman *et al.*, 2011c; Smith, 2013), and exhibits some of the classic traits of the domestication syndrome – increased seed and inflorescence size, loss of shattering, loss of branching, and flowering time transitions (Heiser, 1976; Smith, 2013) – making it well suited to examining the evolution of these convergent traits (Lenser & Theißen, 2013). The extensive and well preserved archaeobotanical collections of sunflower achenes (seed + pericarp/hull) and discs enable the characterization of historic phenotypes and uses of sunflowers (Smith, 2013). As an oilseed crop of increasing value and market dominance (USDA, 2019), sunflower has been subject to intensive breeding and phenotypic characterization (Heiser, 1976), as well as genetic and genomic study (Burke *et al.*, 2002; Kane *et al.*, 2012; Badouin *et*

al., 2017), providing suitable resources for genomic studies of domestication. Finally, recent transcriptome-wide selective sweep studies have identified putative domestication and improvement genes under selection (Baute *et al.*, 2015). However, sunflower is unique in that most domestication traits appear to involve many loci of moderate to small effect (Burke *et al.*, 2002; Wills & Burke, 2007), as opposed few loci of large effect in many other domesticated crops (Koinange *et al.*, 1996; Cai & Morishima, 2002). Combined with the large genome size (~3.5 Gbb, Badouin *et al.*, 2017) and high repetitive DNA content (Staton *et al.*, 2012; Natali *et al.*, 2013), the unique genetic architecture of domestication traits in sunflower has limited the identification and functional/molecular characterization of the genes directly contributing to phenotypic evolution during domestication.

I quantified differential gene expression between wild and domesticated sunflowers using RNA-sequencing to identify and characterize the changes in gene expression contributing to sunflower domestication. I sequenced samples from tissues relevant to domestication phenotypes that spanned sunflower developmental stages in order to more accurately associate expression changes with phenotypic evolution. I also built gene co-expression networks for wild and domesticated sunflowers to determine whether domestication rewired transcriptional networks at the systemic level and identified modules of genes that showed significant changes in member connectivity during domestication. Finally, I examined the overlap between differentially expressed genes with loci previously identified by selective sweep analyses to associate differential expression with phenotypic evolution during domestication. I combined these analyses to characterize gene expression evolution during sunflower domestication.

Methods

Study system

The common sunflower, *Helianthus annuus*, was domesticated ~4000 years ago by Native Americans in eastern North America, a center of domestication for multiple species (Yarnell, 1978; Harter *et al.*, 2004; Smith, 2006; Blackman *et al.*, 2011c). Both wild and domesticated sunflowers served as important food sources for native peoples as well as sources of dye, oil, ceremonial flowers, and building materials (Heiser, 1951; 1976). During sunflower domestication, a syndrome of traits evolved, including losses of seed shattering, seed dormancy and self-incompatibility; increase in size and weight of fruits and inflorescences; and reduced branching (Heiser, 1976; Blackman *et al.*, 2011c; Smith, 2013).

For sunflowers, like most species, the transition from wild ancestor to modern crop is usually loosely divided into two distinct process, domestication and improvement. Domestication is the initiation of divergence from the wild ancestor, while improvement is the subsequent period of rapid diversification of the domesticated crop often associated with intense breeding beginning in the 19th century (Meyer & Purugganan, 2013; Baute *et al.*, 2015). Modern breeding programs in sunflower have also included deliberate introgression of genetic material from multiple wild *Helianthus* species to recover traits such as branching and improve disease resistance, resulting in distinct population structures between domesticated landraces and improved varieties (Baute *et al.*, 2015). In the first half of the 20th century, extant Native American varieties were collected in an effort to preserve valuable pre-improvement germplasm (Heiser, 1951; Wales *et al.*, 2018). I exclusively sampled pre-improvement sunflower varieties to highlight the changes in gene expression during the domestication process, and refer to them throughout as “domesticated accessions.”

Accessions, tissues, and developmental stages sampled

Seeds from five domesticated Native American *H. annuus* accessions (Arikara, Hidatsa, Hopi, Mandan, and Seneca) and five wild *H. annuus* accessions (Arkansas, Kentucky, Missouri, Nebraska/Ann1238, Tennessee) were obtained from the USDA-GRIN repository (Table S1). Domesticated accessions were chosen represent the bulk of extant North American biodiversity pre-improvement and span the geographic distribution of domesticated sunflowers (Wales *et al.*, 2018). Due to the extirpation of native peoples and cultures, no extant domesticated accessions remain from the southeastern region of North America, the likely center of domestication. Wild accessions were chosen to both include the area of domestication and span the same geographic region as extant landraces (Harter *et al.*, 2004).

I sampled sunflowers at 5 developmental stages targeting tissues related to domestication phenotypes to maximize my ability to infer the function of differentially expressed (DE) genes (Fig. 1). I sampled root and shoot tissues of 7-day old seedlings. I sequenced shoot tissues at the 7-day old stage for all accessions to enable a multi-genotype comparison of DE genes. At this early stage, phenotypic divergence between wild and domesticated plants, as well as between accessions within each, is modest compared to later stages, helping to minimize DE genes due to developmental disparity (Swanson-Wagner & Briskine, 2012; Bellucci *et al.*, 2014). For all subsequent tissues/stages, one wild and one domesticated accession were chosen as "focal" accessions for sequencing to defray costs. Ann1238 from Nebraska (NE) was chosen as the wild focal, as it is the predominant wild accession used in sunflower mapping, QTL, and breeding studies (Radanović *et al.*, 2018). Arikara was chosen as the domesticated focal genotype as it was collected closest to the domestication center (Harter *et al.*, 2004).

At the 20-day stage, I sampled leaf, apical meristem, and vegetative node tissues. This stage marks the initiation of branching in wild sunflowers under long-day conditions as well as the beginning of noticeable developmental divergence between wild and domesticated sunflower plants. I then sampled apical meristems at early budding (R1 stage) and late budding (five days post R1) to assess genes involved in the transition to flowering and domestication traits such as inflorescence size. Finally, I sampled seeds nine days post fertilization, again targeting domestication phenotypes such as seed size, seed number, seed oil content, and seed starch. In most domesticated varieties, a majority of fatty acid deposition occurs within the first 30 days of flowering.

Plant growth conditions

Plants from wild and domesticated accessions were grown under identical conditions at each developmental stage. For all stages, seeds were sterilized with a 2% bleach and 1% triton-X (Sigma Aldrich) solution and scarified. Seeds were germinated in the dark on filter paper for 6 days. Seeds were then moved to 16h light: 8h dark environment for 24h prior to sampling or sowing. Samples were collected 4h after initial light exposure. Three replicate RNA seq libraries we created for all tissues. Seven-day old seedlings were sampled directly from filter paper, and for each replicate, roots or stems were pooled from five individuals. No pooling was performed on leaf, node, meristem, early budding, and late budding tissues. For all other stages, seedlings were sown in a mixture of Fafard 3B soil (Sun Gro Horticulture, Agawam, Massachusetts, USA) and calcined clay (3:2 by mass) after seven days on filter paper and grown under long-day conditions (16L:8D). Accessions were evenly distributed in flats and randomized regularly. At 20 days after sowing into 10 cm pots in Conviron growth chambers (constant 25.5ºC), the shoot

apex, third set of true leaves, and axillary node at the base of the second true set of leaves, were sampled from a subset of plants. For plants allotted for sampling at later developmental stages, seedlings were planted into 7.5 liter pots and raised in standard glasshouse conditions (16L:8D with supplemental sodium-halide light, 21 °C days, 16.7 °C nights, daily watering, weekly feeding with MiracleGrow 20-20-20 fertilizer). Apical floral meristems were sampled at early and late budding stages. Plants were crossed within genotype when the second and third whorls of flowers of the disc were receptive to obtain seed and pericarp tissues. Inflorescences were then bagged to prevent inadvertent crossing, and samples were collected nine days post-fertilization from the second and third whirl of flowers. Tissue from 3-4 individual seeds was pooled per replicate from each maternal individual.

RNA-extraction, library preparation and sequencing

Libraries were made from each tissue. All samples were flash frozen and ground in liquid nitrogen. I extracted RNA from ~100mg of tissue following protocol B of the Spectrum Plant Total RNA Kit including on column DNase I treatment (Sigma-Aldrich). RNA samples were then purified and concentrated with the RNeasy MiniElute Cleanup kit (Qiagen). RNA quantity and quality were assayed using an Agilent Bioanalyzer RNA High sensitivity chip and by Qubit prior to library construction. For each sample, 750ng of total RNA with RIN > 8.0 was used to make 100bp paired end libraries with the NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra Directional Library prep kit for Illumina following the manufacturer's instructions (New England BioLabs). A total of 27 Arikara and 27 NE libraries (9 stages, 3 replicates each), and an additional 24 shoot libraries (8 additional accessions, 3 replicates each) were made. Library quality and concentration were determined by Bioanalyzer and Kapa qPCR

before pooling for sequencing. Libraries were sequenced on either an Illumina Hiseq 2500 instrument with v3 chemistry by BGI in Hong Kong or a HiSeq 4000 with v4 chemistry at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (Table S2).

Sequence processing, transcriptome assembly, custom genome generation.

Sequenced libraries were cleaned and used to make a reference transcriptome and custom reference genomes for read mapping and quantification. Adaptor sequences and poor quality bases were trimmed with the EA-Utils tool Fastq-mcf (v1.04.803, quality threshold > 25; % occurrence threshold before adapter clipping = 0.01). A genome-guided transcriptome was generated with Trinity (Grabherr *et al.*, 2011) by subsampling 10 million reads from all Arikara and NE libraries across tissues (9 tissues * 3 replicates = 270 million reads each from Arikara and NE) by Melis Akman. Assembled transcripts were filtered with RNA-seq by the Expectation Maximization (RSEM) method with TPM =0.03 (1st quantile) and isopct = 10. This resulted in a filtered set of 289,628 "genes" and 339,933 transcripts with a contig N50 of 947 and an average contig length of 662bp.

Custom genomes were generated for both of the focal accessions, NE and Arikara, that incorporated sequence divergence from the elite domesticated reference genome (HanXRQ v1). To do this, variants were identified with GATK following best practices for RNA-seq data by the gVCF method (McKenna *et al.*, 2010; DePristo *et al.*, 2011). Briefly, RNA-seq reads from each of the focal libraries were mapped to the XRQ reference genome with STAR-2 pass (Dobin & Gingeras, 2015), duplicates were marked and sorted, and readgroups added using Picard. The GATK tools SplitnTrim and BSQR were used to address reads spanning introns and to recalibrate base quality scores, respectively. Finally, variants were called in gVCF mode for each

sample, and those individual gVCF files were combined for each accession to generate NE and Arikara variants. A custom reference genome and associated GTF were generated for each accession by inserting SNPs and INDELS into the HanXRQv1 reference genome and GTF with g2gtools (henceforth referred to as the Arikara and NE genomes). If the major allele frequency for a given SNP was less than 0.9, IUPAC ambiguity codes were inserted. Only INDELs which occurred at a frequency greater than 0.75 or showed a frequency divergence between accessions greater than 0.6 were incorporated into the custom reference genomes.

Identification of differentially expressed genes (DE genes)

RNA-seq reads were aligned to the custom reference genomes and transcriptome and used to quantify gene expression. RNA-seq reads from all domesticated accessions were aligned to the Arikara reference genome and reads from all wild accessions were aligned to the NE reference genome with HISAT2 (v2.1.0 --dta –no-mixed –no-discordant –fr –max-intronlen 100000) (Kim *et al.*, 2015; Pertea *et al.*, 2016). Reads were then assembled and quantified with StringTie v1.3.2 (Pertea *et al.*, 2015; 2016). The HanXRQv1 genome has sequences from the chloroplast (CP) and mitochondrial (MT) genomes assembled into the nuclear genome. As a result, reads from the plastid genomes were multi-mapping to the nuclear genome, limiting the accuracy of plastid gene quantification. Plastid genes were identified in the chromosomes of the nuclear genome with BLAST and were removed from the StringTie assembled transcripts of each custom genome prior to re-running StringTie.

Differential gene expression was analyzed in R with three different packages because the significance of DE can vary substantially with analysis package (Zyprych-Walczak *et al.*, 2015; Conesa *et al.*, 2016). Genes with read counts below 5 CPM in more than 50% of libraries in a comparison were removed to minimize false positive DE calls from genes with low expression

rates. Raw read counts were normalized and transformed according to best practices for each package, EdgeR (Robinson *et al.*, 2009), DEseq2 (Love *et al.*, 2014), and Limmavoom (Ritchie *et al.*, 2015), prior to analysis. Genes were considered significantly DE when the Benjamini–Hochberg corrected FDR < 0.05 for at least two of the three packages (cf. Colicchio *et al.* 2015; Sanderson *et al.* 2018); referred to as the StringTie pipeline. To confirm DE at the gene level, RNA-seq reads were pseudo-aligned to the *de novo* transcriptome with Kallisto (Bray *et al.*, 2016) (v0.43.1 –bootstrap 100 –fr-stranded) before being analyzed in R with Sleuth (Patro *et al.*, 2017); referred to as the Kallisto pipeline.

Gene ontology (GO) analysis of DE genes

I performed a GO enrichment analysis of DE genes in a tissue specific manner to determine whether DE genes were enrichened for a particular biological process. DE genes were compared to the full set of *Arabidopsis thaliana* genes to identify their orthologs using reciprocal best-hit BLAST. DE genes with no *A. thaliana* ortholog were omitted from GO enrichment analysis. DE genes were classified as more highly expressed in either domesticated plants or wild plants for each individual tissue to determine whether the directionality of DE affected biological process. The list of *A. thaliana* orthologs for each tissue was then used in the GO analysis with PANTHER 13.1, using the *A. thaliana* GO Ontology database released 2018-04-04 and Fisher's Exact with FDR multiple test correction at 0.05.

Identification of DE genes with a signature of selection and involved in the oil biosynthesis pathway

Genes previously identified as showing a signature of selection during domestication (Chapman et al., 2008; Blackman et al., 2011c; Chapman & Burke, 2012; Mandel et al., 2013; Baute et al., 2015) were compared to the HanXRQv1 genome using BLAST to identify DE genes putatively under selection. Putative domestication candidates were then assessed for DE following previous methodology across all tissues and comparisons. Identical methods were used for identifying putative fatty acid biosynthetic pathway genes, except the reference sequence was derived from the *Arabidopsis* fatty acid biosynthetic pathway from the Arabidopsis Lipid Gene Database (http://aralip.plantbiology.msu.edu/) as per Ibarra-Laclette *et al.* (2015)

Identification of phenotype-specific differentially expressed genes

Hopi, a landrace endemic to the southwestern US, exhibits a suite of unique achene and flowering time traits and was utilized by Native American peoples primarily as a source of dye (Heiser, 1976; Wills *et al.*, 2010). I identified genes with a Hopi-specific expression pattern (Hopi vs Man, Sen, Hid – 4,426 DE genes) to determine if this novel use was accompanied by divergent gene expression patterns. Hopi-specific DE domestication genes in the shoot tissues were identified by taking the overlap of two comparisons – the Hopi-specific data set and the domestication-specific genes which were identified by comparing domestic to wild shoot tissues (Hopi, Mandan, Seneca, Hidatsa vs TN, MO, AR, KY – 11,585 DE genes).

I further sequenced apical meristem and vegetal node tissues for the Mandan accession at the 20-day stage, which exhibits stronger apical dominance (pers. obs.) As Arikara is not fully apically dominant, I sequenced a second wild accession, TN, at the same time as Mandan.

Branching-specific genes in both meristematic and node tissues were identified by comparing domesticated to wild libraries (Man & Ari vs NE & TN). Further, because Arikara does not exhibit complete apical dominance, Man vs Ari comparisons were performed, and branching-specific genes identified as the overlap of these two groups (i.e. Mandan-specific expression that is also domestication specific). As the four libraries used for branching comparisons were derived from two separate batches, I confirmed with MDS plots that batch was not driving the grouping of samples (Fig. S1) prior to these comparisons.

Gene co-expression network construction and analysis

Gene co-expression networks were constructed to determine if domestication altered the transcriptional landscape of sunflowers. Filtered and voom transformed reads counts from the StringTie analysis of all Arikara and NE samples were used in the construction of a Domesticated/Wild consensus gene co-expression network with the R package WGCNA (Langfelder & Horvath, 2008). After batch correction, libraries cluster first by wild vs domesticated (NE vs Ari), and then by tissue (Fig. S2). Analysis of network topology was performed to pick a soft threshold of six for a scale-free topology (Fig. S3). For each sample, the eigengene expression value, or first principal component of gene expression, was calculated. In addition, connectivity (kME), or the correlation between module Eigen gene and individual gene's expression was calculated for each sample. I used the ANOVA function in R, with ~tissue*domestication as a fixed effect and sequencing batch as a random effect, to determine whether domestication or tissue type have an effect on module membership. Post-hoc analyses were performed in R with lmer to determine if individual tissue type contributed to module membership, and p=values corrected with Satterthwaite according to Luke (2017).

I used a paired t-test to compare the absolute value of kME of individual genes between wild and domesticated modules to determine whether domestication has an effect on the strength of connectivity within modules. Bonferroni corrections were used on p-values for all module level tests. The number of genes with sign changes (positive-to-negative and vice versa) was calculated network-wide and within each module to determine whether domestication changed the direction of module-wide connections separate from changes to the strength of module connectivity.

Gene network focal gene selection methods.

Three modules (11, 12, 13) were chosen for in-depth analysis of member gene connectivity to determine whether domestication had an effect on the local sub-networks of putative domestication DE genes. These modules were chosen for their significant module-level changes in connectivity between wild and domesticated networks as well as the presence of putative domestication DE genes which served as focal genes. Local sub-networks for each focal DE gene were determined as the genes with a significant Pearson correlation of expression with the focal gene. Individual selection criteria for the local sub-network graphed varied and are detailed below. Briefly, significant correlations between members of the subnetwork were graphed and holm multiple testing corrections performed on the pairwise comparisons between genes in the sub-network (n=total pairwise comparisons in sub-network). For module 11, the set of local genes with expression that was significantly correlated to the focal gene differed completely between Arikara and NE co-expression networks. The local sub-network graphed represents top candidates in the NE sub-network. For module 12, the focal gene is a hub, meaning it is strongly and significantly correlated to a majority of genes in the module for both

the Arikara and NE networks. Hundreds of genes in both Arikara and NE networks had significant expression correlations (positive and negative) to the focal gene in module 12. I chose to depict previous domestication candidates that were significantly correlated in both Ari and NE networks as well as the top highly correlated (absolute value) genes from each. For module 13, the set of local genes with expression significantly correlated to the focal gene differed between Arikara and NE co-expression networks. I selected a subset (n=15) of the total genes correlated to the focal gene in the Arikara network for the figure based on putative function and signature of selection.

Results

RNA-seq mapping and read count

After adaptor trimming, each library contained between 14.8 and 51.3 million paired-end reads (median = 25.7 millions reads, Table S2). Neither tissue, domestication status, sequencing platform, nor batch had an impact on the post-trimming read library count ($P>0.5$, Table S3). The genome-guided transcriptome generated from subsampling 10 million reads from all Arikara and NE libraries resulted in a filtered set of 289,628 “genes” and 339,933 “transcripts” with a contig N50 of 947 and an average contig length of 662bp.

On average, about 65% of total cleaned read pairs per library aligned to the respective reference genome, but there was substantial variation between tissues (31.9-83.7%, Table S2). Tissue type, domestication status, and their interaction all have a significant impact on the percentage of reads aligned to the genome (all $P<0.01$, Tables S2 and S4). Clean read pairs pseudo-aligned to the transcriptome at a higher rate on average, 78%, with substantially less variation (68.3-92.9%, Table S2) than the genome-based method. Domestication had no effect on

the percent reads pseudo-aligned to the transcriptome ($p=0.5$, Table S5), while tissue type and the interaction of tissue and domestication both had a significant effect on percentage of reads pseudo-aligned ($p<0.005$ for both, Table S5).

Identification and characterization of differentially expressed genes

Both the StringTie and Kallisto pipelines tended to identify similar genes as being DE, with expression changes in the same direction. When gene expression was regressed between the two pipelines for each tissue, r-squared values averaged around 0.8 (Fig. S4). All remaining results reported are for the StringTie pipeline. Of the ~56,000 gene level annotations in the StringTie pipeline, between 16,844 and 22,370 were expressed in any given tissue (Fig. 1). Between 16-48% of expressed genes were differentially expressed (DE) between wild and domesticated plants within a tissue. Of the DE genes, 10-20% were DE in a single tissue only, with a majority being shared between at least two tissues. In all comparisons, DE genes were significantly more likely to be expressed at higher levels in domesticated plants than wild (Fig. 2).

Genes that were DE across multiple accessions were also more likely to be expressed at higher levels in domesticated accessions (Shoot-8 = 55%, and 4-accession comparisons of Meristem=78% and Node= 79%, $p < 0.00001$ for all). The single accession shoot comparison (Arikara vs NE) yielded a similarly high number of DE genes as the 8-accession comparison of shoot expression (10,166 and 11,584 DE genes, respectively). Only 33.7% of genes that were significantly DE in either comparison were shared between comparisons (Fig. 3). However, most of the genes that were not significantly DE still showed similar expression profiles with same direction of log2fold change across comparisons (Fig. S5). In the multiple accession comparison,

variation within wild accessions was visible in the 500 most variable genes between wild and domesticated accessions (Fig. 4). Two wild accessions, MO (Fig. 4, A) and TN (Fig. 4, B), exhibit expression patterns in some genes more closely resembling domesticated phenotypes, although this pattern was variable in TN, with one library (TN3), having only ~half the number of genes with a domesticated-like expression pattern. Hopi, a domesticated accession that underwent divergent selection after domestication for dye purposes instead of as an oilseed crop, had a substantial number of genes with a Wild-like expression pattern (Fig. 4, C). Within a tissue, DE genes were enriched for particular GO biological process related to domestication phenotypes of interest. For instance, genes expressed at higher levels in domesticated plants at the late budding stage are enriched for Fatty Acid, Lipid and Wax Biosynthesis/Metabolism (Fig. 1, Table 1).

Differential expression of putative domestication candidates

Of the 183 putative domestication candidate genes identified in previous studies as showing a signature of positive selection, 120 are significantly differentially expressed in at least one tissue (Fig. 5). Of those, 35 are DE in only a single tissue and the remainder are DE in multiple tissue. No domestication candidates are DE in all tissues, but two, HanXRQChr14g0442381 (SERAT3) and HanXRQChr01g0010191 (SNX1) are DE in eight tissues, and only differ in expression between seed and pericarp tissues. Across all tissues, domestication candidates are DE a total of 281 times, with 135 of those showing increased expression and 146 decreased expression in domesticated plants. Of genes expressed at higher levels in domesticated plants, the average $\log_2$ fold change ($l_2$fc) did not vary among tissues ($p = 0.3$). DE was lowest in Node tissues (mean= 1.0 $l_2$fc) and highest in pericarps (mean = 1.9 $l_2$fc),

with most tissues averaging a log2fold change ~1.2 (excluding seeds, which only had a single gene DE, with $l_2$fc of 2.28). Of genes differentially expressed at lower levels in domesticated plants, there was a significant effect of tissue on the average log_2fold change ($p < 0.0001$). DE was weakest in meristem tissues (mean= -0.94 $l_2$fc), while it was strongest in pericarps (mean = -2.07 $l_2$fc), with most tissues averaging a log_2fold change ~ -1.0 (excluding seeds, which only had two genes DE, with mean $l_2$fc of -1.8). Candidates negatively DE in the pericarp were expressed at significantly lower levels than all other tissues ($p<0.0001$ for all comparisons, excluding seeds).

Phenotype specific differential expression results

Fatty acid biosynthetic pathway DE – Of the 61 genes identified in the HanXRQ genome as putatively homologous to *A. thaliana* fatty acid biosynthetic genes, 47 were DE in at least one tissue (Fig. 6). Of those, 22 are DE in only a single tissue. A majority of single tissue genes (14/22) are found in either shoot or root tissues, most of which (13/14) show increased expression in domesticated plants. Across all tissues, fatty acid candidates are DE a total of 96 times, with 53 of those showing increased expression in domesticated plants (driven by shoot and root tissues – 42/53), and 43 with decreased expression. Of particular interest to domesticated phenotypes such as seed oil content, are the seed and pericarp DE genes. In the seed tissue, no fatty acid biosynthesis genes are expressed. In the pericarp, 16 candidate genes are DE, more than any other tissue except roots and shoots, and all of these genes show reduced expression with domestication (mean log2fold change = -1.99)

Hopi Specific DE. A total of 4426 genes were significantly differentially expressed between Hopi shoot tissues and other domesticated accessions. Of these, almost exactly half,

2209 genes, were also differentially expressed between wild and domesticated shoot tissues, indicating that for shoot tissues, Hopi DE expression patterns may be somewhere between wild and domesticated (Fig. 4, C). Of these, 1752 genes show a complete sign change in their expression pattern, indicating that their expression is much more similar to domesticated. Eleven of the Hopi-specific DE genes were previously identified as showing a signature of selection, including the gene *HaFAD2-1* (HanXRQChr01g0009721), which is a key component in sunflower oil biosynthesis and was under selection during improvement. Another gene in the oil biosynthesis pathway, ACPX (HanXRQChr14g0445471), although not a putative domestication candidate, also shows a Hopi-specific DE pattern. HanXRQChr09g0248051, a probable homolog of WNK1, is highly DE across many tissues, and is one of the top variably expressed genes in Shoot-8 tissues, where the expression pattern in Hopi more closely resembles that of wild accessions.

Gene co-expression network results

The consensus co-expression network shared between wild and domesticated sunflower resulted in 36 modules with anywhere from 46 to 5320 member genes (Table 2). There was no effect of domestication on module eigengene value for any gene module ($p > 0.05$). This is not surprising, as a consensus co-expression restricts network construction to shared co-expression modules to facilitate equal comparisons between treatments. Tissue type had a significant effect on module membership for all modules ($p < 0.001$). This is again not surprising given that tissue is a major driver of gene expression, and when clustered by gene expression, samples fall out largely in a tissue-specific manner (Fig. S2). Of the 36 modules, half showed a significant change in the strength of member gene connectivity between wild and domesticated plants

(Table 2). Of modules with change in the strength of connectivity, most (17/18) showed a stronger degree of module level connectivity in domesticated plants compared to wild plants. Only module 20 had a stronger degree of connectivity in wild plants. Of the 26,899 genes used to construct co-expression networks, 2758 (10.2%), showed a change in the sign of connectivity (positive-to-negative and vice versa) with domestication. Sign-change genes were unevenly distributed among modules. In three modules (12, 23, and 26) over 90% of the member genes showed a sign change with domestication.

Module 12 subnetwork: The focal gene, a putative *ACD1* (HanXRQChr14g0437821), shows a signature of selection during domestication and is involved in plant greening in *Arabidopsis*. *ACD1* changes direction of network connectivity in the entire module (NE kME = 0.91, Ari kME = -0.86), however, the strength of connectivity (absolute value kME) is relatively unchanged. *ACD1* expression is significantly correlated with a very large number of module member genes in both Ari (n=276) and NE (n=273), of which, most are shared (n=218), further supporting the genes role as a network hub (Fig 7). The number of genes with expression negatively correlated to *ACD1* increases with domestication (NE=82, Ari=99), but average strength remains unchanged (NE r=-0.73, Ari r=-0.7, p=0.2). The number of genes with expression positively correlated to *ACD1* decreases slightly with domestication (NE=191, Ari=177), and the average strength of the correlations significantly decreased with domestication (NE r=0.88, Ari r=0.82, $p<0.0001$). *ACD1* was significantly DE in many tissues (Fig. 8, A), with a reduction in expression in domesticated plants relative to wild plants. An exception was shoots, where expression was higher in domesticated plants.

Module 11 subnetwork results: The focal gene, a putative *HaFAD2-1* (HanXRQChr01g0009721), shows a signature of selection during improvement and is integral to

the fatty acid biosynthesis pathway. *HaFAD2-1* is in the 84th percentile for network connectivity change within module (Fig. 9). The number of genes significantly correlated to *HaFAD2-1* is halved during domestication (NE = 31, Ari = 15) and there is no overlap between the genes correlated to *HaFAD2-1* in Arikara and NE. Of the genes significantly correlated with *HaFAD2-1* in Arikara, most (13/15) show a strong negative correlation (mean r=-0.75), while in NE, most of the significantly correlated genes (28/31) show a strong positive correlation (mean r=0.75). *HaFAD2-1* is highly DE in pericarp (log2fold change = -4) as well as multiple other tissues (Fig. 10, A).

Module 13 subnetwork results: The focal gene, a putative *CDF* (HanXRQChr07g0199341), shows a signature of selection during domestication and is an integral part of the flowering time regulatory pathway. *CDF* has a very weak connection in the wild module with only a small nonsignificant kME (p=0.12). Only two genes were connected to *CDF* in the wild network (Fig. 11). However, in the domesticated network, *CDF* is moderately connected (kME= -0.56) and a total of 33 genes show a significant expression correlation with the *CDF,* most of which are positively correlated to *CDF* expression. *CDF* is DE in shoot and root tissues (Fig. 8B).

Discussion

Widespread differential expression with domestication

Differential gene expression studies between closely related species or populations have identified widely varying numbers of differentially expressed genes, from a low of 0.7% of expressed genes in the common bean (*Phaseolus vulgaris)* (Bellucci *et al.*, 2014) to as high as 78% in closely related *Drosophila* species (McManus *et al.*, 2010). Most studies average 2-25%,

with a variety of biological, ecological, and methodological factors impacting these numbers (reviewed in Pavey *et al.*, 2010; Alvarez *et al.*, 2015). Studies specifically looking at DE associated with domestication have identified relatively modest changes in the percent of genes differentially expressed, e.g., 3.3% of genes DE between maize and teosinte (Swanson-Wagner & Briskine, 2012), 0.7% between wild and domesticated bean (Bellucci *et al.*, 2014), and ~9.5-14% between wild and domesticated tomato (Koenig & Jiménez-Gómez, 2013; Sauvage *et al.*, 2017a). Across tissues and developmental stages, I observed substantially higher percentages of DE, with between 16-48% of expressed genes DE, indicating differential gene expression may contribute more substantially to the domestication syndrome in sunflowers than in other domesticated plants.

The higher percentages of DE in my data could also be associated with the experimental design. Within a tissue, comparisons included only a single wild (NE) and domesticated (Arikara) accession and thus these comparisons reflect accession-specific DE as well as domestication DE. However, in the 8-accession comparison of shoot tissues, the percentages of DE genes (47.1%) is even higher than the single accession comparison (43.6%). Similar results were found when an additional wild (TN) and domesticated (Mandan) accessions were sequenced for apical meristem and vegetative node tissues. These results suggest that accession-specific DE is not the likely driver of higher proportions of DE in sunflowers compared to other crops, particularly combined with consistently high DE across all tissues in sunflowers.

In all comparisons, DE genes tended to be expressed at higher levels in domesticated plants than in wild (Fig. 2). A bias toward increased expression in domesticated plants has been demonstrated in maize (Swanson-Wagner & Briskine, 2012; Lemmon *et al.*, 2014; Wang *et al.*, 2018), cassava (Xia *et al.*, 2016) and cotton (Rapp *et al.*, 2010). My data supports the trend of an

increase in gene expression with domestication, and the pattern is robust to development and tissue type. In addition, genes that were DE across multiple accessions were also more likely to be expressed at higher levels in domesticated accessions, indicating this finding is not due to accession-specific DE patterns. Exceptions to this trend do exist in a few domesticated species, with overall downregulation during domestication in the common bean (Bellucci *et al.*, 2014) and tomatoes (Sauvage *et al.*, 2017b). This pattern is attributed to the proposed importance of loss-of-function mutations over gain-of-function mutations (Bellucci *et al.*, 2014). However, we lack a robust explanation for the evolutionary importance of the general trend of increased expression and its role in domestication across species (Wang *et al.*, 2018). One possibility is that it may be driven by the increases in growth rate and overall plant size that generally accompanying domestication.

Signatures of selection

Differentially expressed genes identified through comparative transcriptomics could be directly responsible for phenotypic differences associated with domestication. However, expression differences could also be downstream of causative mutations or result from epistatic interactions. While a tissue- and developmentally-specific DE pattern can suggest which differentially expressed genes contribute to phenotypes under selection, without functional confirmation studies, these data do not provide definitive results. Further, extensive sampling across multiple tissues and accessions can identify tens of thousands of differentially expressed genes (Fig. 1). To identify differentially expressed genes that are likely responsible for domestication phenotypes, I compared DE genes with those previously identified as putatively under selection during domestication by showing a signature of selection in coding sequence.

A majority (66%) of putative domestication candidates were differentially expressed in at least one tissue (Fig. 5), consistent with patterns in maize (Swanson-Wagner & Briskine, 2012) where genomic regions with a signature of selection were enriched for DE genes. Most domestication candidate genes were differentially expressed in more than two tissues and spanned multiple developmental stages (Fig. 5). Unlike the entire population of DE genes, where domestication led to higher expression across all tissues, putative domestication candidates were no more likely to have greater expression in domesticated plants than in wild plants regardless of tissue or accession. This indicates that genome-wide increases in expression may be a result of pleiotropic effects of genes under selection, contrary to maize, where DE genes identified as having a signature of selection are more likely to be expressed higher levels in domesticated plants (Swanson-Wagner & Briskine, 2012).

Here I examine several candidates with putative functions associated with domestication to better understand the patterns of DE found in genes with a signature of selection. HanXRQChr13g0390401, a putative homolog of *HaFAD2-3*, a gene in the fatty acid biosynthetic pathway, has been identified as being under selection during domestication (Chapman & Burke, 2012). *HaFAD2-3* had a significant reduction in expression with domestication in late budding and pericarp tissues (Fig. 10, C). The *FAD2* gene family, with three members in *H. annuus*, has been important in the breeding of high oleic oilseed crops because the enzyme encoded by this gene converts oleic acid into linoleic acid. High oleic acid oil is preferable for both its improved storage, industrial use, and cardiovascular health benefits (Yamaki *et al.*2002). Across species, reduced expression of *FAD2* results in high oleic acid oil varieties and has been targeted for breeding, e.g., in sunflower (Schuppert *et al.*, 2006), peanuts (Patel *et al.*, 2004), and flax (Chen *et al.*, 2015), among others (Dar *et al.*, 2017).

In sunflowers, *HaFAD2-1*, has been the primary focus of breeding and molecular characterization, as it is the only *HaFAD2* gene expressed in high levels the developing seeds in improved varieties, and its silencing has been functionally linked to high oleic oil varieties (Martínez-Rivas *et al.*, 2001; Schuppert *et al.*, 2006). *HaFAD2-1* also shows a signature of selection with improvement, but not with domestication (Chapman & Burke, 2012). This pattern is congruent with the development of high oleic acid sunflower varieties through mutagenesis in the latter half of the 20th century (Miller *et al.*, 1987; Dimitrijević *et al* 2017.). Consistent with this previous work, I found no change in *HaFAD2-1* expression during domestication; it is highly expressed in both wild and domesticated seeds (Fig. 10A). However, *FAD2-3,* does show a significant reduction in expression in domesticated pericarps compared to wild, as does *FAD2-2*, and both genes show minimal expression in both wild and domesticated seeds (Fig. 10). Combined with the domestication signature of selection in *FAD2-3*, these data suggest that modification to fatty acid biosynthesis and composition in sunflower seeds may have preceded modern improvement through the reduction of expression of *HaFAD2*-2 and *HaFAD2*-3 in pericarp tissues. Selection for reduced *HaFAD2-2/3* gene expression during domestication may potentially explain their weak expression in modern improved varieties as well.

A finer scale examination of the timing and location of expression of all three *HaFAD2* genes in the seeds of multiple accessions of wild and domesticated landrace sunflowers is necessary to further elucidate the impact of this gene on fatty acid biosynthesis during and after domestication. However, results to date indicate the *HaFAD2* gene family is likely to have played a key role during domestication and improvement. In fact, all DE genes putatively identified in the fatty acid biosynthesis pathway (including those not showing a signature of selection), have reduced expression in domesticated pericarp tissues compared to wild (Fig. 6).

Further developmental dissection of gene expression and functional characterization of these genes have great potential to reveal valuable insight into both the evolution of domesticated sunflowers, and the details of an important agronomic trait.

Tissue and developmental specificity of DE genes.

I found that the developmental stage and tissue sampled had substantial impacts on differential gene expression, including the number of DE genes (Fig. 1), direction of DE (Fig. 2), type of genes (Table 1), and specific candidate gene expression (Figs. 5 & 6). The effects of tissue and developmental specificity are not surprising given the overwhelming support for these phenomena in evolutionary development (Gilbert & Epel, 2009). However, by leveraging the valuable biological data and unique evolutionary history of domesticated species, I was able to further elucidate the complex role of differential gene expression in the evolution of traits of interest in sunflowers.

I found dramatic variation in the direction and magnitude of DE by specifically targeting tissues at key developmental stages related to phenotypes under selection during domestication. For example, HanXRQChr05g0155691 is a putative *HaSWEET13* gene encoding a bidirectional sugar transporter that shows a signature of selection (Baute *et al.*, 2015). I found this putative *HaSWEET13* gene broadly shows increased expression in domesticated sunflower, with positive log2fold changes in root, late budding, and pericarp tissues (Fig. 12). Sugar transport can impact plant growth rate/size and seed yield through alterations to resource allocation between source (leaf) and sink (root/seed) tissues (Lemoine, 2013; White et al., 2015). Growth rate/size and seed yield broadly increase during domestication across species (Meyer & Purugganan, 2013; Milla et ˆal., 2014; Milla & Matesanz, 2017). Increased expression of *HaSWEET13* during the

domestication of sunflowers could play a role in altering these traits. In *A. thaliana*, overexpression of *AtSWEET4,* a hexose transporter gene, leads to increases in plant size, with the opposite for knock-down mutants (Liu *et al.*, 2016). In maize, *ZmSWEET4c* is expressed at greater levels in developing maize grains than in teosinte, shows a signature of selection, and in knockdown mutants seed filling is defective. The same is true of the rice ortholog, *OsSWEET4* (Sosso *et al.*, 2015). It is possible that convergent selection for seed traits may have acted on the same gene family in both monocots and dicots during domestication.

While the putative *HaSWEET13* gene shows increased expression with domestication in most sunflower tissues, there was decreased expression in node tissues at the 20-day stage ($l_2$fc=-2.9, Fig. 12). This developmental time point is when branching is initiated in wild sunflowers, while domesticated sunflowers retain apical dominance. Sugar transport genes may be associated with apical dominance and the initiation of branching. For example, an increased sugar supply has been shown to be both necessary and sufficient for suppressed buds to be released from apical dominance (Mason *et al.*, 2014; Van den Ende, 2014). A greater concentration of sugar transporter proteins produced by *HaSWEET13* in vegetative nodes in wild sunflowers could be one potential supplier of sugar for initiating branching. Without a sampling scheme spanning development, the nuanced manner in which domestication has shaped the timing and location changes to *HaSWEET13* expression demonstrated here would not be evident.

Gene co-expression network evolution

In the wild and domesticated consensus gene co-expression network, half of the modules identified showed a significant change in the strength of member gene connectivity, with almost all of those exhibiting a stronger degree of connectivity in domesticated plants. The few studies

that have directly compared the effects of domestication on the connectivity of gene co-expression patterns have found inconsistent results. In maize, some sub-networks surrounding DE genes showed a reduction in connectivity with domestication, while others were unchanged (Swanson-Wagner & Briskine, 2012). In tomato, of the three co-expression modules identified, two showed a decrease in connectivity with domestication, while one showed an increase (Koenig & Jiménez-Gómez, 2013). Finally, in the common bean, data suggested that genes under selection might have higher connectivity in domesticated plants compared to wild ones (Bellucci *et al.*, 2014). The prevalence of genome-wide increases in the strength of co-expression network connectivity in my data suggest that in sunflower domestication acted to increase gene-expression integration.

Evidence for integration in local sub-networks surrounding genes extends to those with a signature of selection. For instance in module 13, the focal gene, a putative cycling DOF factor, *CDF* (HanXRQChr07g0199341), is a repressor of flowering time, a phenotype under selection during sunflower domestication (Blackman *et al.*, 2010; 2011a; Blackman, 2013). Indeed, *CDF* has been identified as a putative domestication candidate with a signature of selection during domestication (Chapman *et al.*, 2008), linked to QTL for flowering time traits, and demonstrated a signature of selection in a transcriptomic analysis (Baute *et al.*, 2015). In this study, *CDF* showed reduced expression with domestication in shoot (and Shoot-8) and root tissue comparisons, and a Hopi specific DE pattern (Fig. 8 B, D). *CDF* is not integrated into the wild sunflower co-expression network, and is only co-expressed with two of the other 463 genes in the module (Fig. 9). However, in the domesticated network, *CDF* has become strongly connected to 33 genes and has a significant, negative KmE to the module eigengene, suggesting its expression pattern runs counter to the modules core co-expression (Fig. 9). Nearly all of the

genes co-expressed with *CDF* within the module show a strong positive correlation with *CDF* expression, indicating that while *CDF* expression is generally opposite that of most genes in the network, its subnetwork is integrated in the same direction of co-expression.

Two of the genes that are in the *CDF* sub-network are also flowering time regulatory genes, a putative *DELLA* (HanXRQChr08g0212091) and *LHY* (HanXRQChr09g0264411). *CDF, DELLA*, and *LHY* are all positively co-expressed with one another in the domesticated network, and with many of the other member genes of the subnetwork. Further supporting the co-expression integration of these three genes in domesticated plants, *DELLA* and *LHY* are both negatively co-expressed with module hub genes, similar to *CDF* (*DELLA* kME = -0.79, *LHY* kME= -0.75). The only member of the subnetwork that is negatively correlated to any other genes is HanXRQChr11g0332111, a probable beta glucosidase 41 (*BGLU41)* gene. *BGLU41* shows negative co-expression with only two of the 33 sub-network member genes, *CDF* and *LHY*, again suggesting that these flowering time regulators may be integrated. Finally, all three of these genes functionally act to repress flowering: *CDF1* through repressing floral activators *CONSTANS (CO)* and *FLOWERING LOCUS T* (FT) (Fornara *et al.*, 2009; Goralogia *et al.*, 2017); *DELLA* through interactions with the gibberellic acid pathway and repression of *CO, SOC1*, and *LFY* (Blackman *et al.*, 2011a; Li *et al.*, 2016; Wang *et al.*, 2016); *LHY* through its control of the circadian clock and interaction with photoperiodic flowering induction (Fujiwara *et al.*, 2008; Yoo & Wendel, 2014). Combined, the signature of selection in *CDF* and the functional and co-expression integration of *CDF1, DELLA,* and *LHY* flowering time genes suggests the potential for selection to have acted on them in concert during domestication. If domestication traits such as flowering time are controlled by many genes of small effect, as is the

proposed case in sunflowers (Burke *et al.*, 2002; Wills & Burke, 2007), more tightly integrated gene co-expression could allow for a more rapid response to selection at the phenotypic level.

However, there was also evidence within subnetworks of reduced connectivity around focal genes. In module 11, while domestication resulted in a more strongly connected co-expression network, the local network surrounding a putative improvement candidate, *HaFAD2-1*, showed a dramatic reduction in connectivity (Fig. 9). Further, of the 12 domestication candidates with the greatest (top 10%) change in connectivity module wide, around half (7/12) have reduced connectivity in the domesticated network. This included candidates (4/7) in modules which are on the whole, more strongly connected in domesticated plants. One possible explanation for the local reduction in gene connectivity with domestication is that strongly connected genes are more likely to be pleiotropic than weakly connected genes (He & Zhang, 2006). Genes with reduced pleiotropy could represent easier targets of selection, as they may be subject to reduced evolutionary constraint (Stern & Orgogozo, 2008). As such, the degree to which a trait is able to respond to selection during domestication could be impacted by the level of co-expression connectivity. My data suggests that in some cases, pleiotropy might be driving a reduced connectivity among co-expressed genes while in others, it could lead to functional integration. The potentially opposing global and local co-expression patterns highlights the importance of context-specific analyses of domestication on the transcriptional landscape of sunflowers.

Conclusion

The effects of human-mediated selection on gene expression evolution can be far-reaching and complex. Domestication of the common sunflower, *Helianthus annuus*, resulted in both substantial differential expression of individual genes across tissues and development, as well as broad changes to the entire transcriptional landscape through changes to gene co-expression patterns. Domestication has resulted in the widespread increase of gene expression across the genome and globally strengthened gene co-expression networks. While comparative transcriptomic studies alone are often unable to differentiate between the effects of selection and nonadaptive processes, combining these results with signature of selection studies identified differentially expressed putative domestication genes. This combined approach identified valuable candidates for molecular confirmation studies to enhance further understanding of gene function in relation to the evolution of domesticated species.

References

Alvarez M, Schrey AW, Richards CL. 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Molecular Ecology* 24: 710–725.

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Brière C, Owens GL, Carrère S, Mayjonade B, *et al.* 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546: 148–152.

Baute GJ, Kane NC, Grassa CJ, Lai Z, Rieseberg LH. 2015. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytologist* 206: 830–838.

Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, Minio A, Rau D, Rodriguez M, Panziera A, *et al.* 2014. Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *The Plant Cell* 26: 1901–1912.

Blackman BK. 2013. Interacting duplications, fluctuating selection, and convergence: the complex dynamics of flowering time evolution during sunflower domestication. *Journal of Experimental Botany* 64: 421–431.

Blackman BK, Michaels SD, Rieseberg LH. 2011a. Connecting the sun to flowering in sunflower adaptation. *Molecular Ecology* 20: 3503–3512.

Blackman BK, Rasmussen DA, Strasburg JL, Raduski AR, Burke JM, Knapp SJ, Michaels SD, Rieseberg LH. 2011b. Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* 187: 271–287.

Blackman BK, Scascitelli M, Kane NC, Luton HH, Rasmussen DA, Bye RA, Lentz DL, Rieseberg LH. 2011c. Sunflower domestication alleles support single domestication center in eastern North America. *Proceedings of the National Academy of Sciences of the United States of America* 108: 14360–14365.

Blackman BK, Strasburg JL, Raduski AR, Michaels SD, Rieseberg LH. 2010. The role of recently derived ft paralogs in sunflower domestication. *Current Biology* 20: 629–635.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34: 525–527.

Burke JM, Tang S, Knapp SJ, Rieseberg LH. 2002. Genetic analysis of sunflower domestication. *Genetics* 161: 1257–1267.

Cai H, Morishima H. 2002. QTL clusters reflect character associations in wild and cultivated rice. *Theoretical and Applied Genetics* 104: 1217–1228.

Chapman MA, Burke JM. 2012. Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower. *Theoretical and Applied Genetics* 125: 897–907.

Chapman MA, Pashley CH, Wenzler J, Hvala J, Tang S, Knapp SJ, Burke JM. 2008. A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *The Plant Cell* 20: 2931–2945.

Chen Y, Zhou X-R, Zhang Z-J, Dribnenki P, Singh S, Green A. 2015. Development of high oleic oil crop platform in flax through RNAi-mediated multiple FAD2 gene silencing. *Plant Cell Reports* 34: 643–653.

Colicchio JM, Miura F, Kelly JK, Ito T, Hileman LC. 2015. DNA methylation and gene expression in *Mimulus guttatus*. *BMC Genomics*: 1–15.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, *et al.* 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology*: 1–19.

Dar AA, Choudhury AR, Kancharla PK, Arumugam N. 2017. The FAD2 gene in plants: occurrence, regulation, and role. *Frontiers in Plant Science* 8: 1336–16.

Darwin C. 1859. *On the origin of species by means of natural selection*.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43: 491–498.

Dimitrijević A, Imerovski I, and DMCB, 2017. Oleic acid variation and marker-assisted detection of Pervenets mutation in high-and low-oleic sunflower cross. *SciELO Brasil*

Dobin A, Gingeras TR. 2015. Mapping RNA-seq reads with STAR. *Current Protocols in Bioinformatics.* 2015 51:11.14.1-19.

Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127: 1309–1321.

Fornara F, Panigrahi KCS, Gissot L, Sauerbrunn N, RUhl M, Jarillo JA, Coupland G. 2009. Arabidopsis DOF transcription factors act redundantly to reduce constans expression and are essential for a photoperiodic flowering response. *Developmental Cell* 17: 75–86.

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD. 2000. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289: 85–88.

Fujiwara S, Oda A, Yoshida R, Niinuma K, Miyata K, Tomozoe Y, Tajima T, Nakagawa M, Hayashi K, Coupland G, *et al.* 2008. Circadian clock proteins LHY and CCA1 regulate SVP protein accumulation to control flowering in *Arabidopsis*. *The Plant Cell* 20: 2960–2971.

Gepts P. 2014. The contribution of genetic and genomic approaches to plant domestication studies. *Current Opinion in Plant Biology* 18: 51–59.

Gilbert SF, Epel D. 2009. *Ecological Developmental Biology*. Sinauer Associates Incorporated.

Goralogia GS, Liu T-K, Zhao L, Panipinto PM, Groover ED, Bains YS, Imaizumi T. 2017. CYCLING DOF FACTOR 1 represses transcription through the TOPLESS co-repressor to control photoperiodic flowering in *Arabidopsis*. *The Plant Journal* 92: 244–262.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Harter AV, Gardner KA, Falush D, Lentz DL, Bye RA, Rieseberg LH. 2004. Origin of extant domesticated sunflowers in eastern North America. *Nature* 430: 201–205.

He X, Zhang J. 2006. Toward a molecular understanding of pleiotropy. *Genetics* 173: 1885–1891.

Heiser CB Jr. 1951. The Sunflower among the North American Indians. *Proceedings of the American Philosophical Society* 95: 432–448.

Heiser CB Jr. 1976. *The Sunflower*. University of Oklahoma Press.

Ibarra-Laclette E, Méndez-Bravo A, Pérez-Torres CA, Albert VA, Mockaitis K, Kilaru A, López-Gómez R, Cervantes-Luevano JI, Herrera-Estrella L. 2015. Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. *BMC Genomics*: 1–18.

Kane NC, Burke JM, Marek L, Seiler G, Vear F, Baute G, Knapp SJ, Vincourt P, Rieseberg LH. 2012. Sunflower genetic, genomic and ecological resources. *Molecular Ecology Resources* 13: 10–20.

Kantar MB, Nashoba AR, Anderson JE, Blackman BK, Rieseberg LH. 2017. The genetics and genomics of plant domestication. *Bioscience* 67: 971–982.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12: 357–360.

Koenig D, Jiménez-Gómez JM. 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proceedings of the National Academy of Sciences of the United States of America* 110: 2655–2662.

Koinange EMK, Singh SP, Gepts P. 1996. Genetic control of the domestication syndrome in common bean. *Crop Science* 36: 1037–1045.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 1–13.

Lemmon ZH, Bukowski R, Sun Q, Doebley JF. 2014. The role of cis regulatory evolution in maize domestication (H Fraser, Ed.). *PLoS genetics* 10: e1004745–15.

Lenser T, Theißen G. 2013. Molecular mechanisms involved in convergent crop domestication. *Trends in Plant Science* 18: 704–714.

Li M, An F, Li W, Ma M, Feng Y, Zhang X, Guo H. 2016. DELLA proteins interact with FLC to repress flowering transition. *Journal of Integrative Plant Biology* 58: 642–655.

Liu X, Zhang Y, Yang C, Tian Z, Li J. 2016. AtSWEET4, a hexose facilitator, mediates sugar transport to axial sinks and affects plant development. *Scientific Reports* 6:24563 1–12.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 31–21.

Luke SG. 2017. Evaluating significance in linear mixed-effects models in R. *Behavioral Research Methods* 49:1494-1502.

Martínez-Rivas JM, Sperling P, Lühs W, Heinz E. 2001. Spatial and temporal regulation of three different microsomal oleate desaturase genes (FAD2) from normal-type and high-oleic varieties of sunflower (*Helianthus annuus L.*). *Molecular Breeding* 8: 159–168.

Mason MG, Ross JJ, Babst BA, Wienclaw BN, Beveridge CA. 2014. Sugar demand, not auxin, is the initial regulator of apical dominance. *Proceedings of the National Academy of Sciences of the United States of America* 111: 6092–6097.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.

McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Research* 20: 816–825.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* 14: 840–852.

Miller JF, Zimmerman DC, Vick BA. 1987. Genetic control of high oleic acid content in sunflower oil. *Crop Science* 27: 923–926.

Natali L, Cossu RM, Barghini E, Giordani T, Buti M, Mascagni F, Morgante M, Gill N, Kane NC, Rieseberg L, *et al.* 2013. The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads. *BMC Genomics* 14: 686.

Patel M, Jung S, Moore K, Powell G, Ainsworth C, Abbott A. 2004. High-oleate peanut mutants result from a MITE insertion into the FAD2 gene. *Theoretical and Applied Genetics* 108: 1492–1502.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14: 417-419.

Pavey SA, Collin H, Nosil P, Rogers SM. 2010. The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences* 1206: 110–129.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11: 1650–1667.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33: 290–295.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. *Nature* 457: 843–848.

Purugganan MD, Fuller DQ. 2011. Archaeological data reveal slow rates of evolution during plant domestication. *Evolution* 65: 171–183.

Radanović A, Miladinović D, Cvejić S, Jocković M, Jocić S. 2018. Sunflower genetics from ancestors to modern hybrids - a review. *Genes* 9: 528–19.

Rapp RA, Haigler CH, Flagel L, Hovav RH, Udall JA, Wendel JF. 2010. Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum L.*) was massively altered by domestication. *BMC Biology* 8: 139–15.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43: e47–e47.

Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.

Ross-Ibarra J, Morrell PL, Gaut BS. 2007. Plant Domestication, a Unique Opportunity to Identify the Genetic Basis of Adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 8641–8648.

Sanderson BJ, Wang L, Tiffin P, Wu Z, Olson MS. 2018. Sex-biased gene expression in flowers, but not leaves, reveals secondary sexual dimorphism in *Populus balsamifera*. *New Phytologist* 221: 527–539.

Sauvage C, Rau A, Aichholz C, Chadoeuf J, Sarah G, Ruiz M, Santoni S, Causse M, David J, Glémin S. 2017a. Domestication rewired gene expression and nucleotide diversity patterns in tomato. *The Plant Journal* 91: 631–645.

Schaefer RJ, Michno J-M, Myers CL. 2017. Unraveling gene function in agricultural species using gene co-expression networks. *BBA - Gene Regulatory Mechanisms* 1860: 53–63.

Schuppert GF, Tang S, Slabaugh MB, Knapp SJ. 2006. The sunflower high-oleic mutant ol carries variable tandem repeats of FAD2-1, a seed-specific oleoyl-phosphatidyl choline desaturase. *Molecular Breeding* 17: 241–256.

Smith BD. 2006. Eastern North America as an independent center of plant domestication. *Proceedings of the National Academy of Sciences of the United States of America* 103: 12223–12228.

Smith BD. 2013. The domestication of *Helianthus annuus L*. (sunflower). *Vegetation History and Archaeobotany* 23: 57–74.

Sosso D, Luo D, Li Q-B, Sasse J, Yang J, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS, *et al.* 2015. Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. *Nature Genetics* 47: 1489–1493.

Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, Ungerer MC, Knapp SJ, Rieseberg LH, Burke JM. 2012. The sunflower (*Helianthus annuusL.)* genome reflects a recent history of biased accumulation of transposable elements. *The Plant Journal* 72: 142–153.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62: 2155–2177.

Swanson-Wagner R, Briskine R. 2012. Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences of the United States of America* 109: 11878–11883.

USDA. 2019. *Global Oilseed Consumption Continues to Grow Despite Slowing Trade and Production*.

Van den Ende W. 2014. Sugars take a central position in plant growth, development and, stress responses. A focus on apical dominance. *Frontiers in Plant Science 5*: 1–3.

Wales N, Akman M, Watson RHB, Sánchez Barreiro F, Smith BD, Gremillion KJ, Gilbert MTP, Blackman BK. 2018. Ancient DNA reveals the timing and persistence of organellar genetic bottlenecks over 3,000 years of sunflower domestication and improvement. *Evolutionary Applications* 12: 38–53.

Wang H, Pan J, Li Y, Lou D, Hu Y, Yu D. 2016. The DELLA-CONSTANS transcription factor cascade integrates gibberellic acid and photoperiod signaling to regulate flowering. *Plant Physiology* 172: 479–488.

Wang X, Chen Q, Wu Y, Lemmon ZH, Xu G, Huang C, Liang Y, Xu D, Li D, Doebley JF, *et al.* 2018. Genome-wide analysis of transcriptional variability in a large maize-teosinte population. *Molecular Plant* 11: 443–459.

Wills DM, Burke JM. 2007. Quantitative trait locus analysis of the early domestication of sunflower. *Genetics* 176: 2589–2599.

Wills DM, Abdel-Haleem H, Knapp SJ, Burke JM. 2010. Genetic architecture of novel traits in the hopi sunflower. *Journal of Heredity* 101: 727–736.

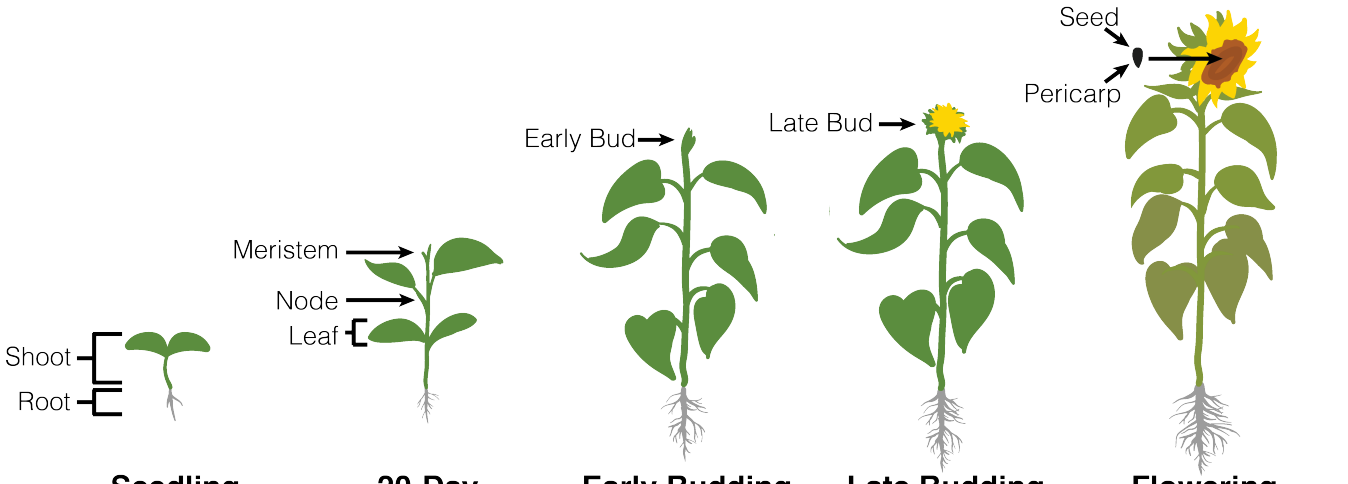
Xia Z, Chen X, Lu C, Zou M, Wang S, Zhang Y, Pan K, Zhou X, Wang H, Wang W. 2016. Comparative transcriptomics revealed enhanced light responses, energy transport and storage in domestication of cassava (*Manihot esculenta*). *Frontiers of Agricultural Science and Engineering* 3: 295–13.

Yamaki T, Yano T, Satoh H, Endo T, Lipids CM, 2002. High oleic acid oil suppresses lung tumorigenesis in mice through the modulation of extracellular signal-regulated kinase cascade. *Food Science Technology Research* 11:231-235

Yarnell R. 1978. Domestication of sunflower and sumpweed in Eastern North America. Ford RI ed. The nature and status of ethobotany. University of Michigan.

Yoo M-J, Wendel JF. 2014. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genetics* 10: e1004073.

Zyprych-Walczak J, Szabelska A, Handschuh L, Górczak K, Klamecka K, Figlerowicz M, Siatkowski I. 2015. The impact of normalization methods on RNA-Seq data analysis. *BioMed Research International* 2015: 1–10.



Seedling

Expr	DE	GO Enrichment
Root		
21,946	10,438	Vegetative Phase Change & Absisic Acid Biosynth
Shoot		
21,301	10,166	Photosynthesis

20-Day

Expr	DE	GO Enrichment
Apical Meristem		
21,027	5,581	Specification of Floral Organ
Vegetative Node		
21,128	7,052	Cell Wall Components & Transport
Leaf		
16,844	3,822	Jasmonic Acid Pathway

Early Budding

Expr	DE	GO Enrichment
Early Bud		
22,202	3,524	Photosynthesis & Carbon Fixation

Late Budding

Expr	DE	GO Enrichment
Late Bud		
22,370	4,687	Photosynthesis & Carbon Fixation

Flowering

Expr	DE	GO Enrichment
Seed		
18,152	1,859	Disaccharide Biosynthesis
Pericarp		
19,575	3,794	Plant Ovule Development, Sucrose metabolism, Fatty Acid Biosynth

Figure 1 - Developmental stages and tissues sampled for RNA sequencing. Number of genes expressed, differentially expressed, and GO enrichments of DE genes of interest to domestication phenotypes in Arikara vs Nebraska comparisons. See more comprehensive GO enrichment results in Table 1.

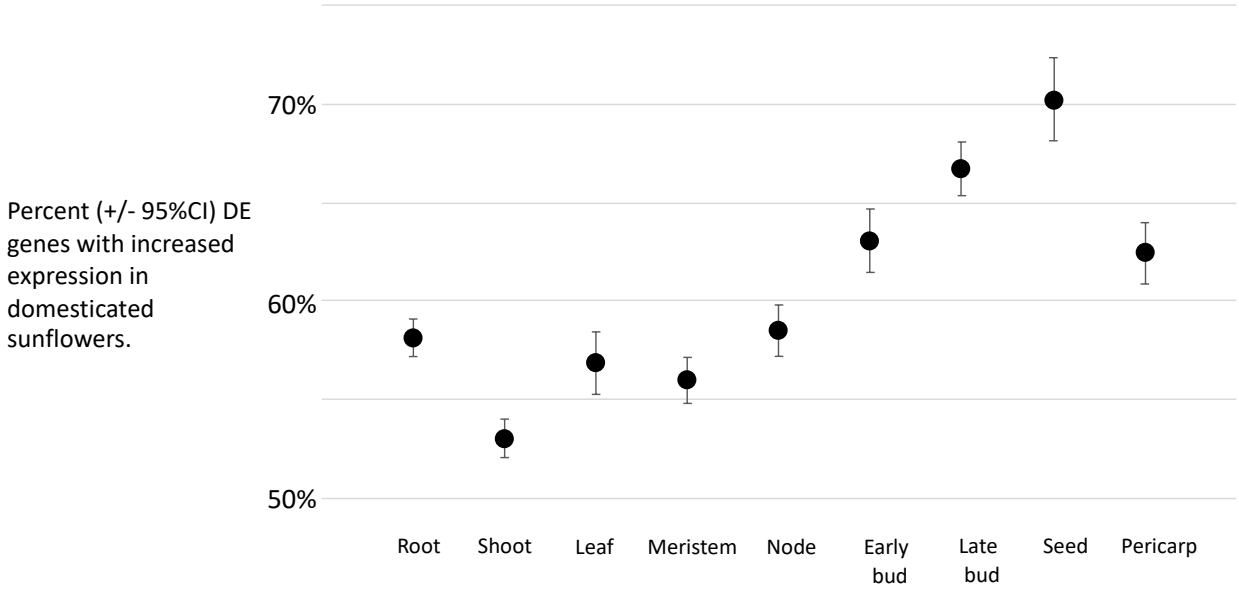


Figure 2 - Percent of differentially expressed genes (+/- 95%CI) with increased expression in domesticated sunflowers (Arikara) relative to wild sunflowers (NE) by individual tissue. All values are significantly different ($p < 0.00001$) from the null expectation of an equal distribution (50%) of genes.

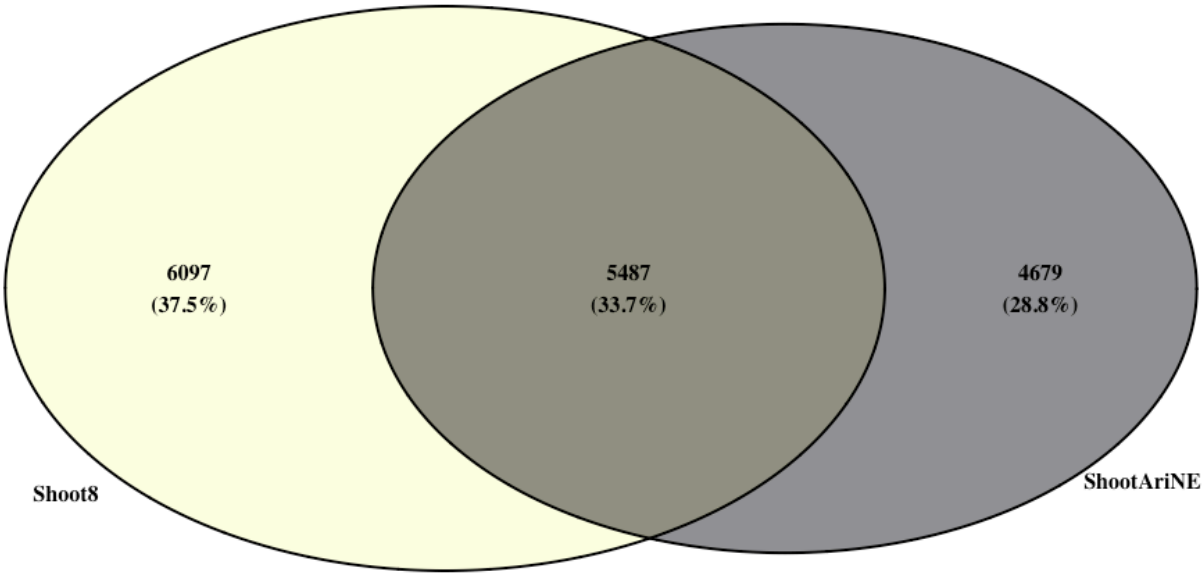


Figure 3 - Overlap of significantly differentially expressed genes between the two shoot tests, a comparison of a single domesticated and wild accession (Arikara vs NE) and a comparison of four domesticated to four wild accessions (Shoot8 – AR, TN, MO, KY vs Seneca, Mandan, Hopi, Hidatsa)

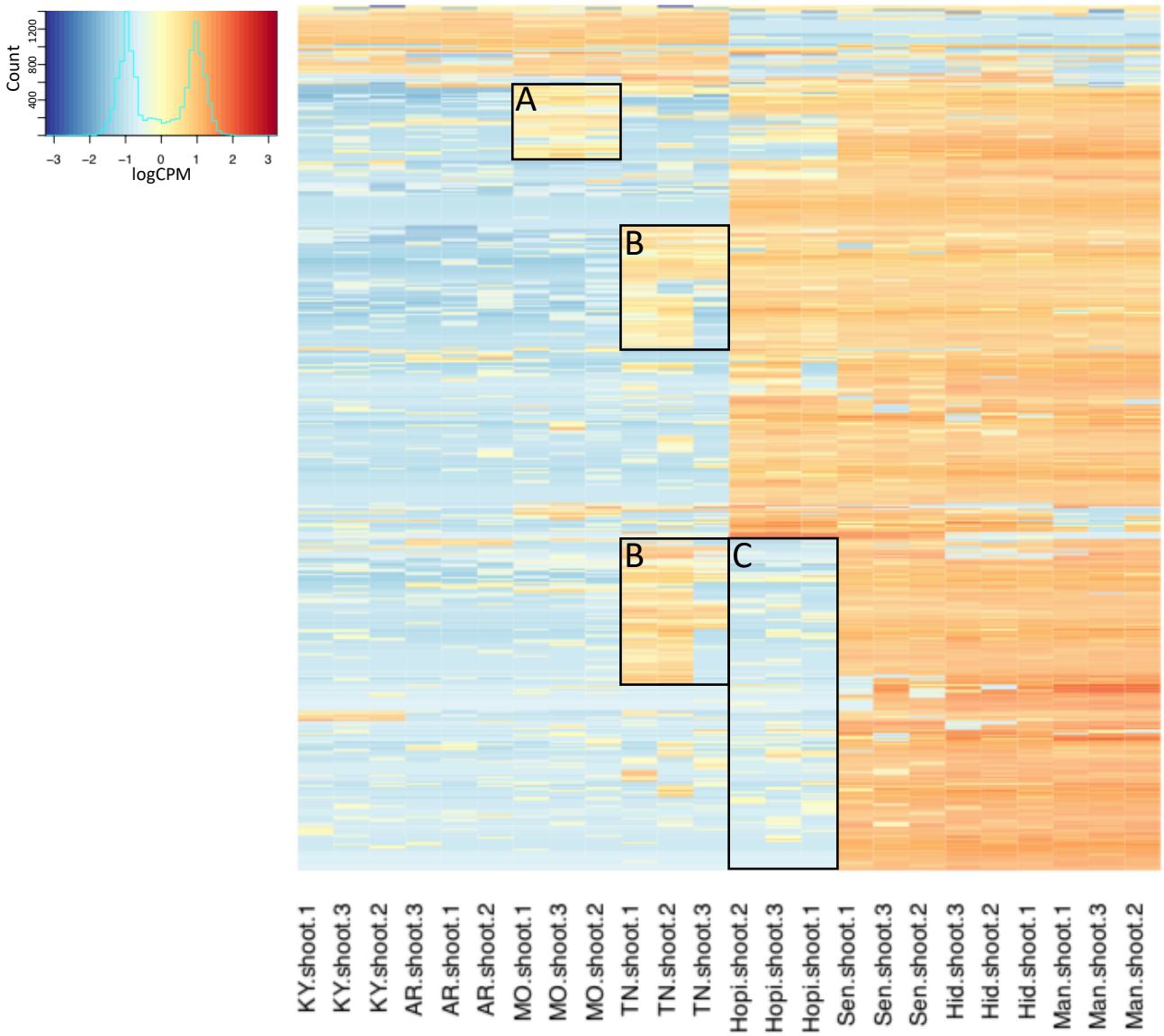


Figure 4 - Heatmap of expression (logCMP) of the top 500 most variably expressed genes in the Shoot8 comparison. Two wild accessions, MO (A) and TN (B), exhibit expression patterns in some genes more closely resembling domesticated accessions, although this pattern is variable in TN, with one library (TN.3), having only ~half the number of genes with a domesticated-like expression pattern. Hopi, a domesticated accession that underwent divergent selection after domestication for dye purposes instead of as an oilseed crop, has a substantial number of genes with a Wild-like expression pattern (C) including a putative Diacylglycerol acyltransferase, HanXRQChr15g0475811, implicated as a primary gene in seed oil deposition and WNK3, HanXRQChr09g0248051, a flowering time/photoperiod gene that shows high DE across many other tissues as well.

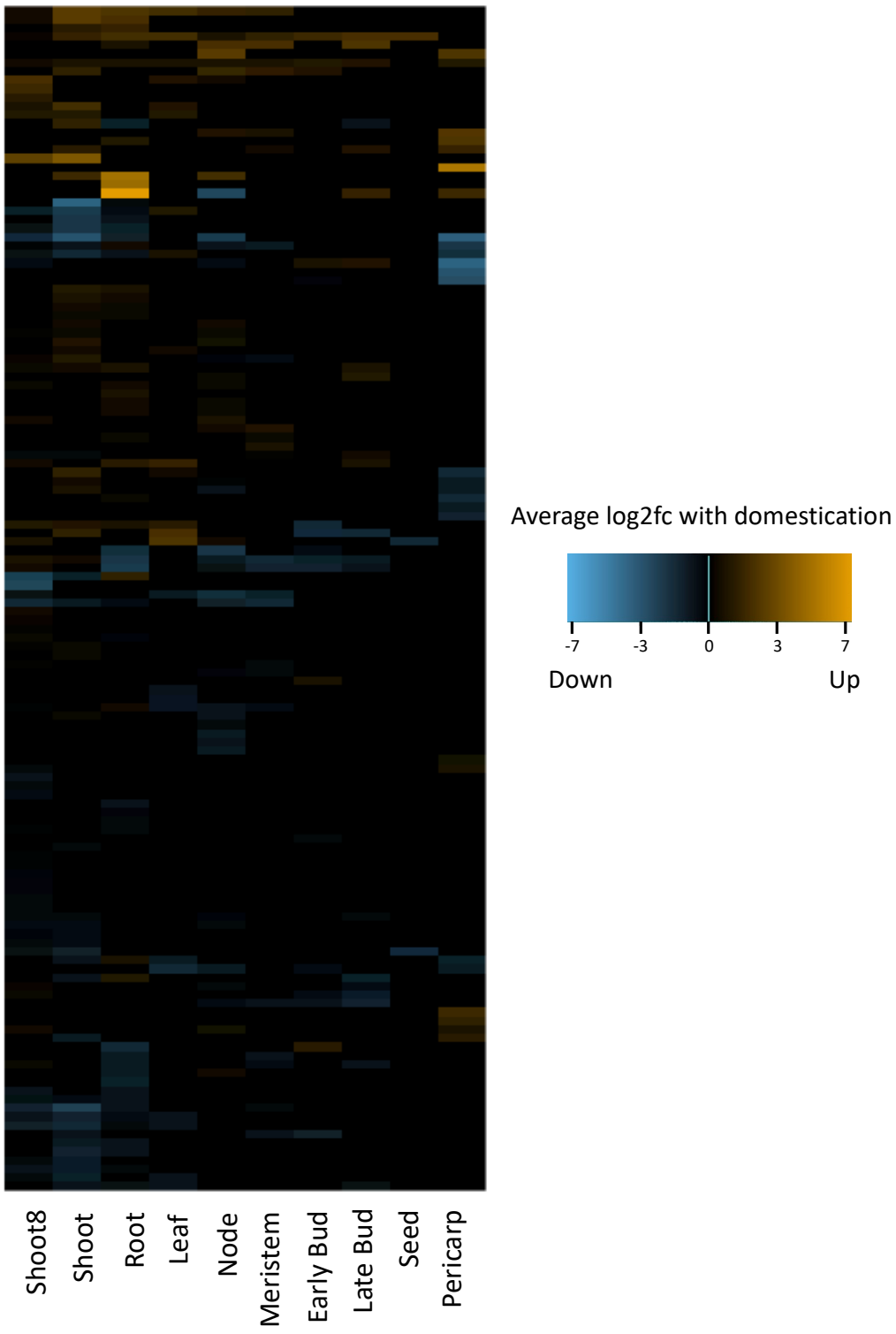


Figure 5 - Average log2-fold change in expression of significantly differentially expressed genes between wild (NE) and domesticated (Arikara) libraries as well as Shoot8, a combination of four domesticated and four wild accessions, for putative domestication candidates previously identified as showing a signature of selection. Genes in orange are expressed at higher levels in domesticated plants compared to wild, while those in blue are expressed at lower levels.

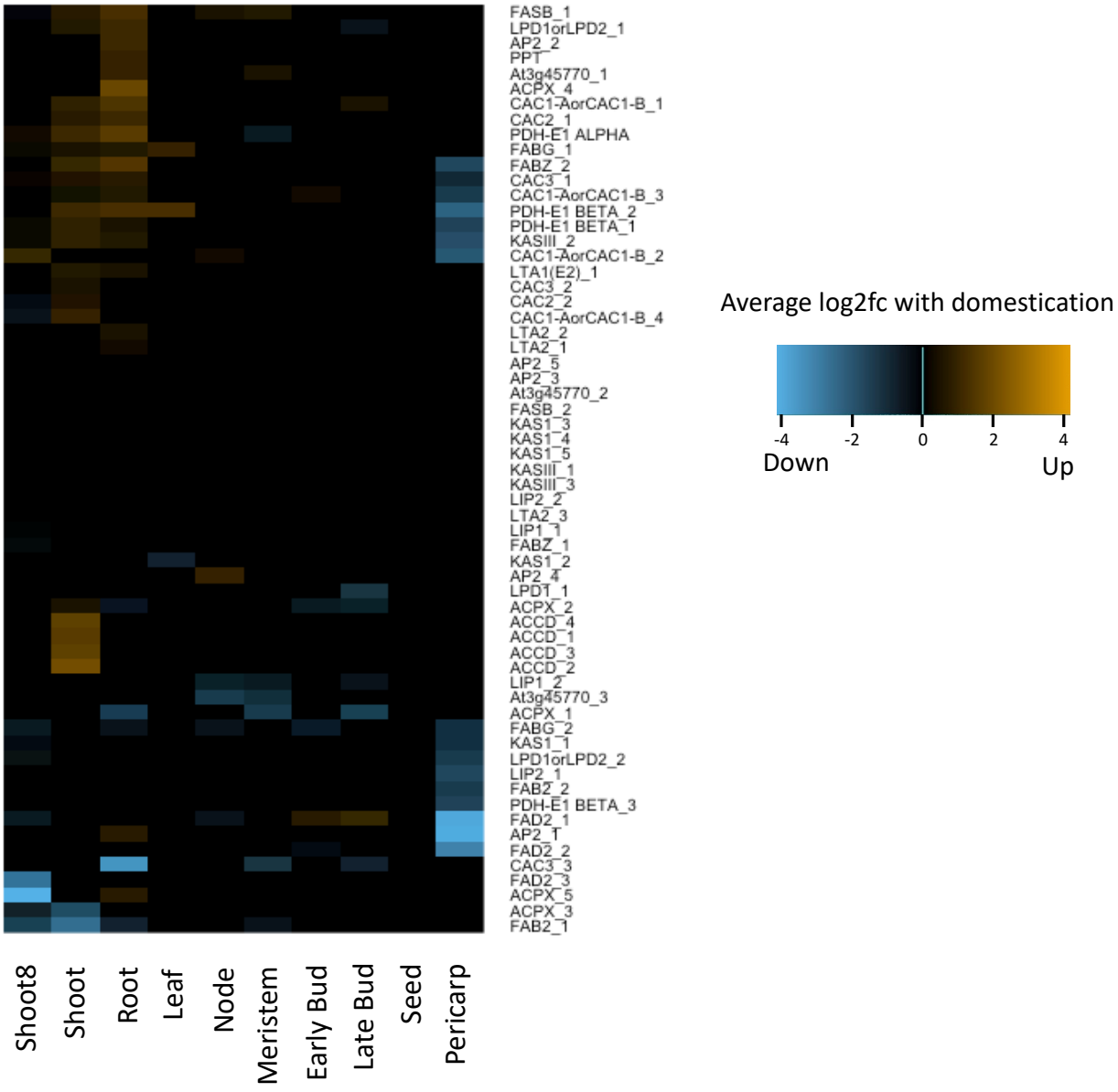


Figure 6 - Average log2-fold change in expression of significantly differentially expressed genes between wild (NE) and domesticated (Arikara) libraries as well as Shoot8, a combination of four domesticated and four wild accessions, for homologs of the fatty acid biosynthetic pathway in *Arabidopsis*. In early developmental stages, such as shoot (23/25) and root (21/26), many genes increase in expression with domestication (blue), while at later developmental stages, notably the pericarp (16/16), many genes show a significant decrease in expression with domestication. Genes shown in orange are expressed at higher levels in domesticated plants compared to wild, while those in blue are expressed at lower levels.

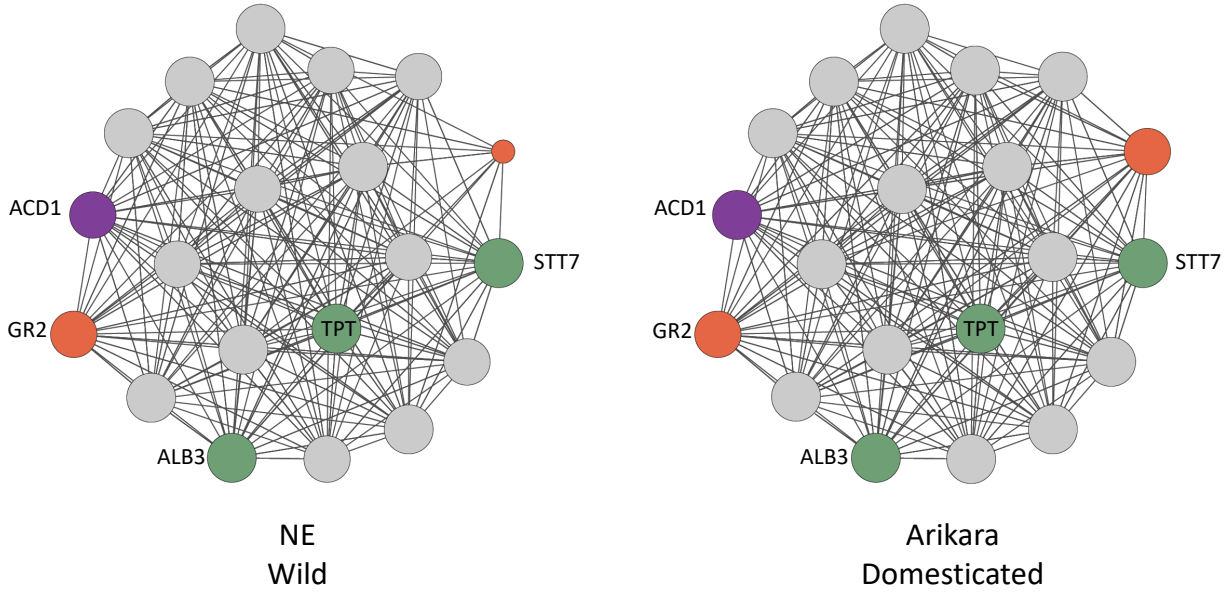


Figure 7 - Module 12 gene co-expression subnetwork. Focal gene, ACD1-like (Purple), a putative domestication gene and hub gene implicated in plant greening/chlorophyll functions, has significant correlated expression with many genes in the network (Ari n=276; NE n=273). Top highly correlated genes include two other putative domestication genes, GR2 and HanXRQChr09g0267541 (orange), as well as genes that are orthologs of AT genes (green) implicated in photosynthesis/circadian rhythm (STT7), environmental acclimation/sugar transport (TPT), and plant senescence/greening (ALB3).

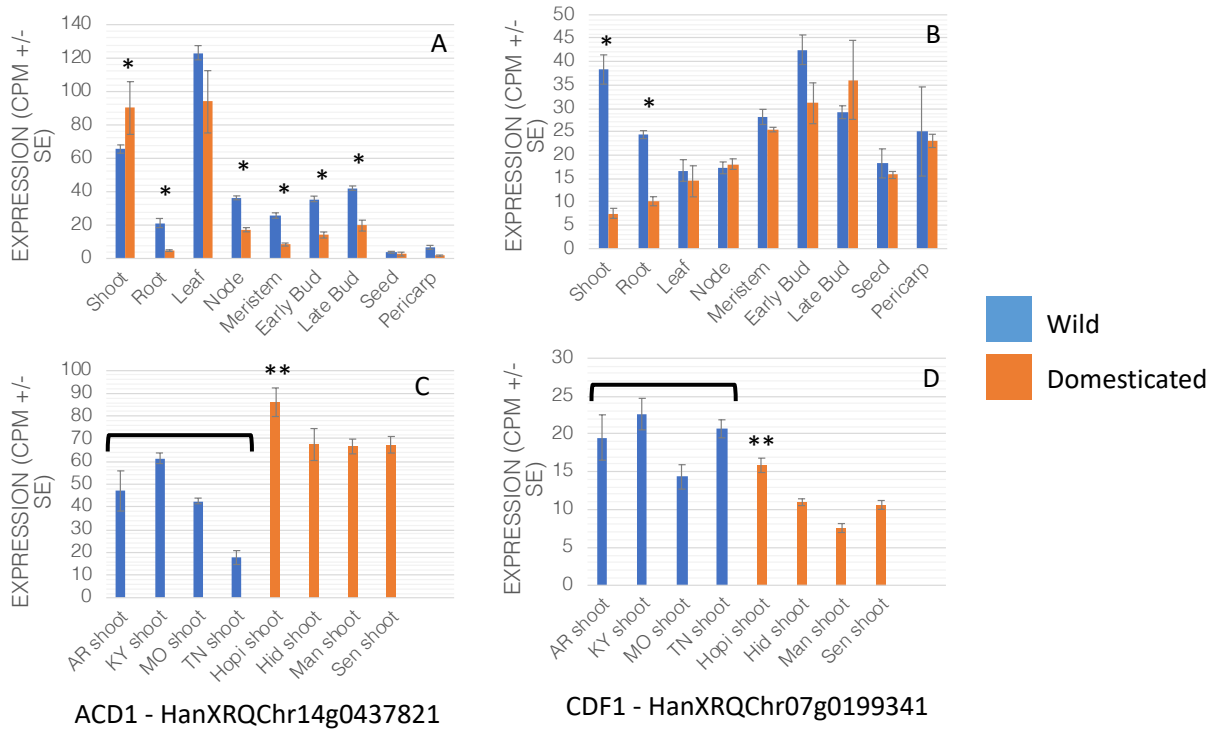


Figure 8 - Expression (CPM +/- SE) of putative domestication candidates chosen as focal genes of interest in their respective gene co-expression modules across tissue (Arikara vs NE) comparisons (A, B) and within shoot tissue across genotypes (C,D). Single asterisk indicates significant DE within tissue. Brackets indicate significant DE between all wild and domesticated genotypes within shoot tissues, and double asterisk indicates significant DE between Hopi and other domesticated genotypes.

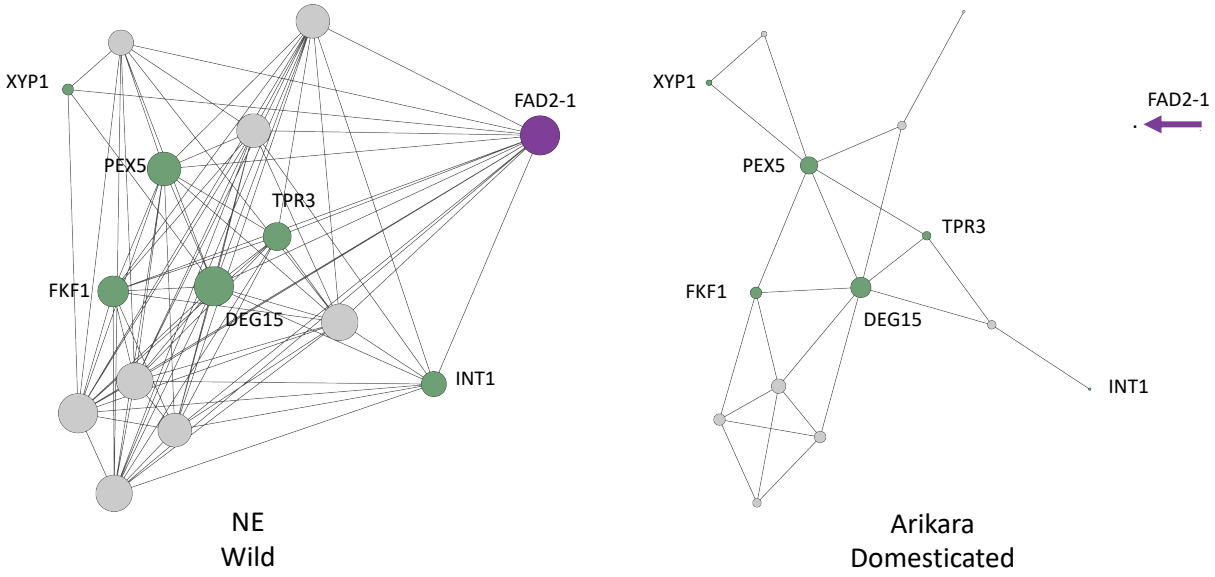


Figure 9 - Module 11 gene co-expression subnetwork. Focal gene, FAD2-1 (Purple), a putative improvement gene and an integral member of the fatty acid biosynthetic pathway. In wild (NE) networks, expression is significantly correlated to a number of genes of interest (green), e.g. FKF1 (a flowering time); XYP1, PEX5, and DEG15 (lipid transfer or metabolism); and TPR3 and INT1 (branching related through apical dominance or sugar transport and cell elongation). In the Ari subnetwork, all correlated expression to the focal gene of interest is lost, and many of the members of the subnetwork show a reduction in the number and strength of correlated expression with one another.

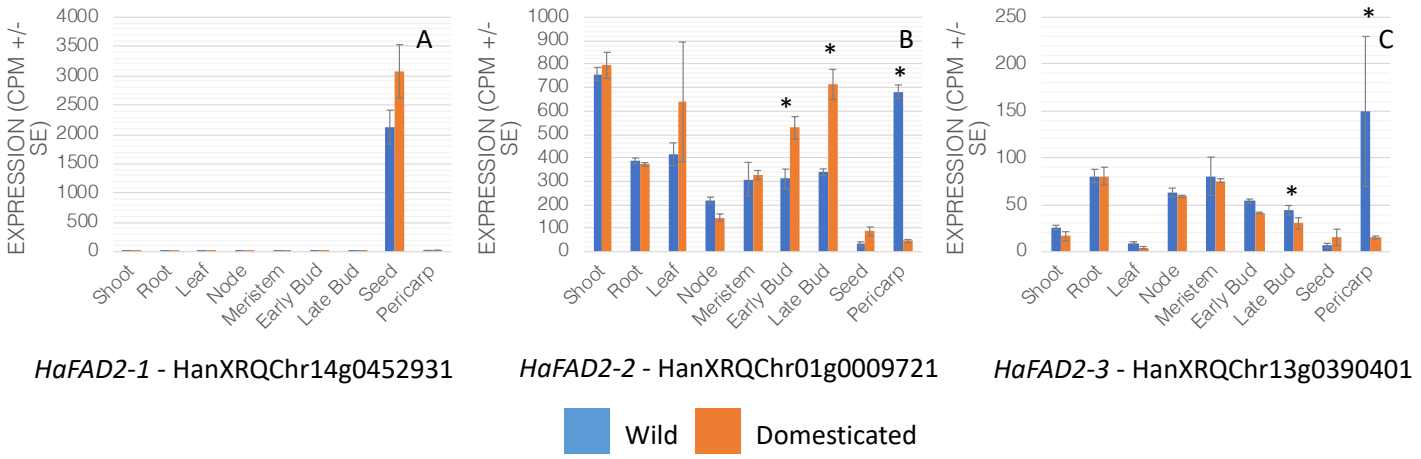


Figure 10 - Expression (CPM +/- SE) of the three FAD2 copies in the sunflower genome, including *HaFAD2-1,* a putative improvement gene, and *HaFAD2-3,* a putative domestication candidate. Asterisk indicates significant DE within tissue.

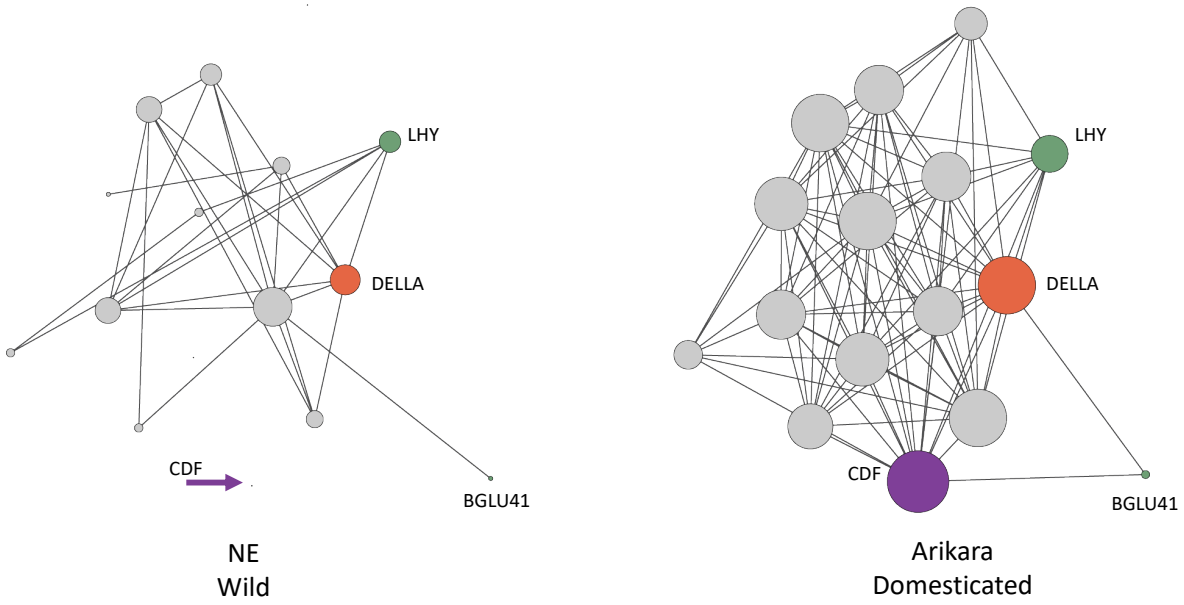


Figure 11 - Module 13 gene co-expression subnetwork. Focal gene, CDF (Purple), is a putative domestication gene and functionally involved in the sunflower flowering time pathway. In the wild (NE) network, expression is not significantly correlated to any other genes in the module. In the domesticated (Arikara) network, expression is correlated to a number of genes of interest, including two other flowering time genes, DELLA (orange, putative domestication candidate) and LHY (green). BGLU41 is the only gene in the subnetwork with a negative correlation in Arikara, and it is only correlated to the two genes in the subnetwork with a signature of selection during domestication.

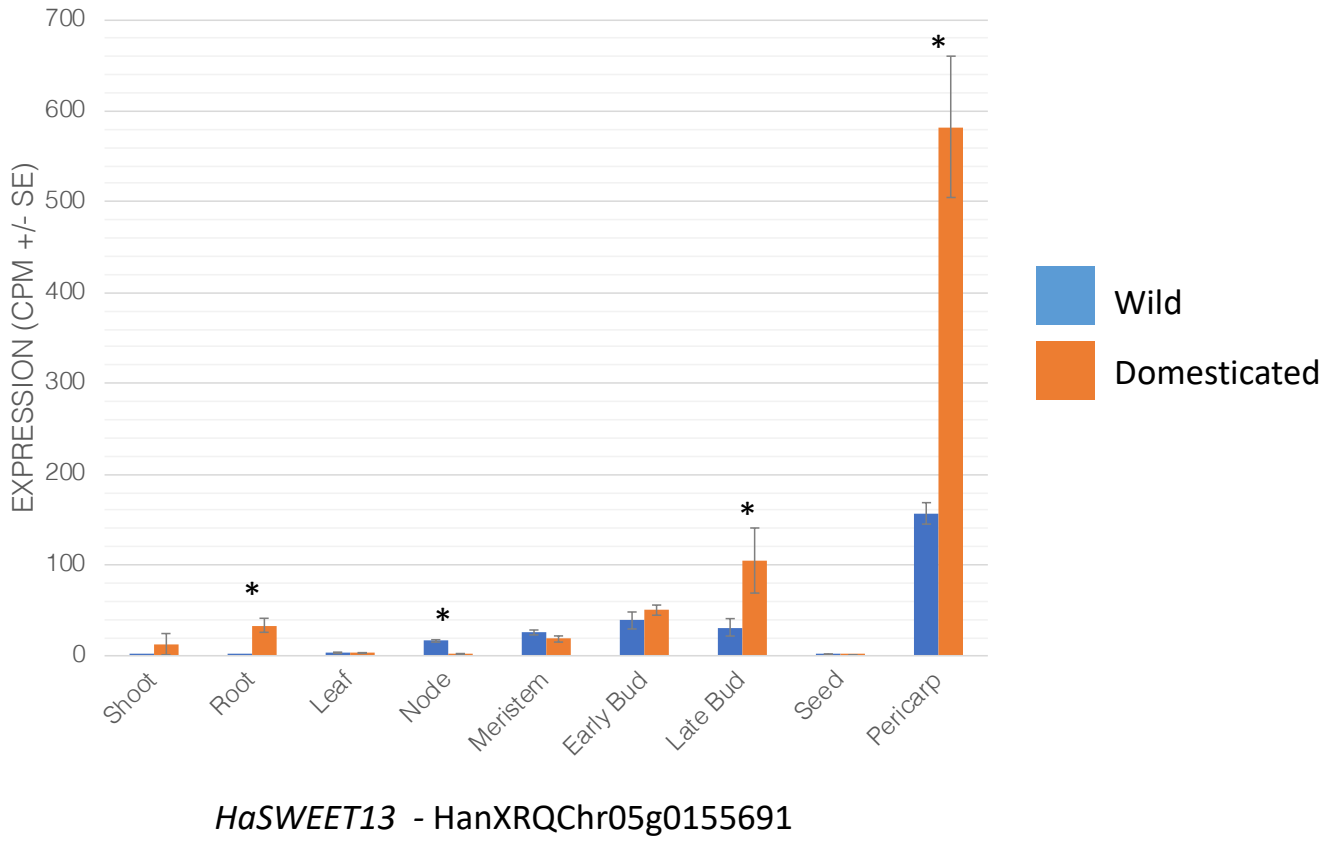


Figure 12 – Differential expression of putative domestication candidate *HaSWEET13*, a sugar transport gene, across tissues. Asterisk indicates significant DE within tissue

Table 1 – Key biological process of interest related to domestication phenotypes enriched in GO analysis of DE genes separated by tissue and direction of expression change (+ = positive log2fc with domestication). A full summary of GO analysis, including specific GO terms, P-values, and Fold-enrichments is available in Supplemental Table 6.

Tissue	Direction	GO Biological Process
Root	-	Peptide Transport
	+	Vegetative Phase Change & Absisic Acid Biosynth
Shoot	-	rRNA Processing & Ribosomal Subunits
	+	Photosynthesis
Leaf	-	DNA Replication
	+	Jasmonic Acid Pathway
Node	-	Metabolic & Biosynthetic Processes
	+	Cell Wall Components & Transport
Meristem	-	Specification of Floral Organ
	+	Regulation of Transcription and Translation
Early Bud	-	Photosynthesis & Carbon Fixation
	+	DNA biosynthesis & Transport
Late Bud	-	Photosynthesis
	+	Fatty Acid, Lipid and Wax Biosynthesis/Metabolism
Seed	-	Disaccharide Biosynthesis, Systemic Acquired Resistance
	+	Kinetichore Assembly
Pericarp	-	Plant Ovule Development, Sucrose metabolism, Fatty Acid Biosynth
	+	Nucleotide-sugar Biosynth & Cell Wall Formation

Table 2 – Summary of gene co-expression modules, gene membership, significant changes in member connectivity, percentage of genes within a module with a sign change, and the number of genes with a signature of selection during domestication.

Module number	Number member genes	Connectivity strength change	% genes sign change	Number domestication candidates
1	5319		5.0%	26
2	3638	+	5.3%	21
3	3479		5.9%	10
4	2514	+	6.6%	10
5	1183		6.1%	8
6	1136	+	11.7%	8
7	1092	+	4.9%	9
8	987	+	5.5%	5
9	937		1.5%	5
10	908	+	4.6%	2
11	901	+	12.7%	7
12	655	+	94.8%	3
13	463	+	11.9%	4
14	460	+	9.6%	2
15	446	+	11.7%	1
16	321	+	1.9%	3
17	265		3.8%	1
18	250	+	88.8%	0
19	223		4.0%	1
20	205	-	41.5%	1
21	156		1.3%	0
22	147		0.0%	0
23	123	+	97.6%	0
24	122		13.1%	0
25	113		3.5%	2
26	111		91.9%	0
27	108		0.9%	3
28	107	+	0.0%	0
29	101		4.0%	0
30	96		4.2%	0
31	75		16.0%	0
32	66		77.3%	0

33	49	+	59.2%	0
34	49	+	0.0%	0
35	48		0.0%	0
36	46		0.0%	1

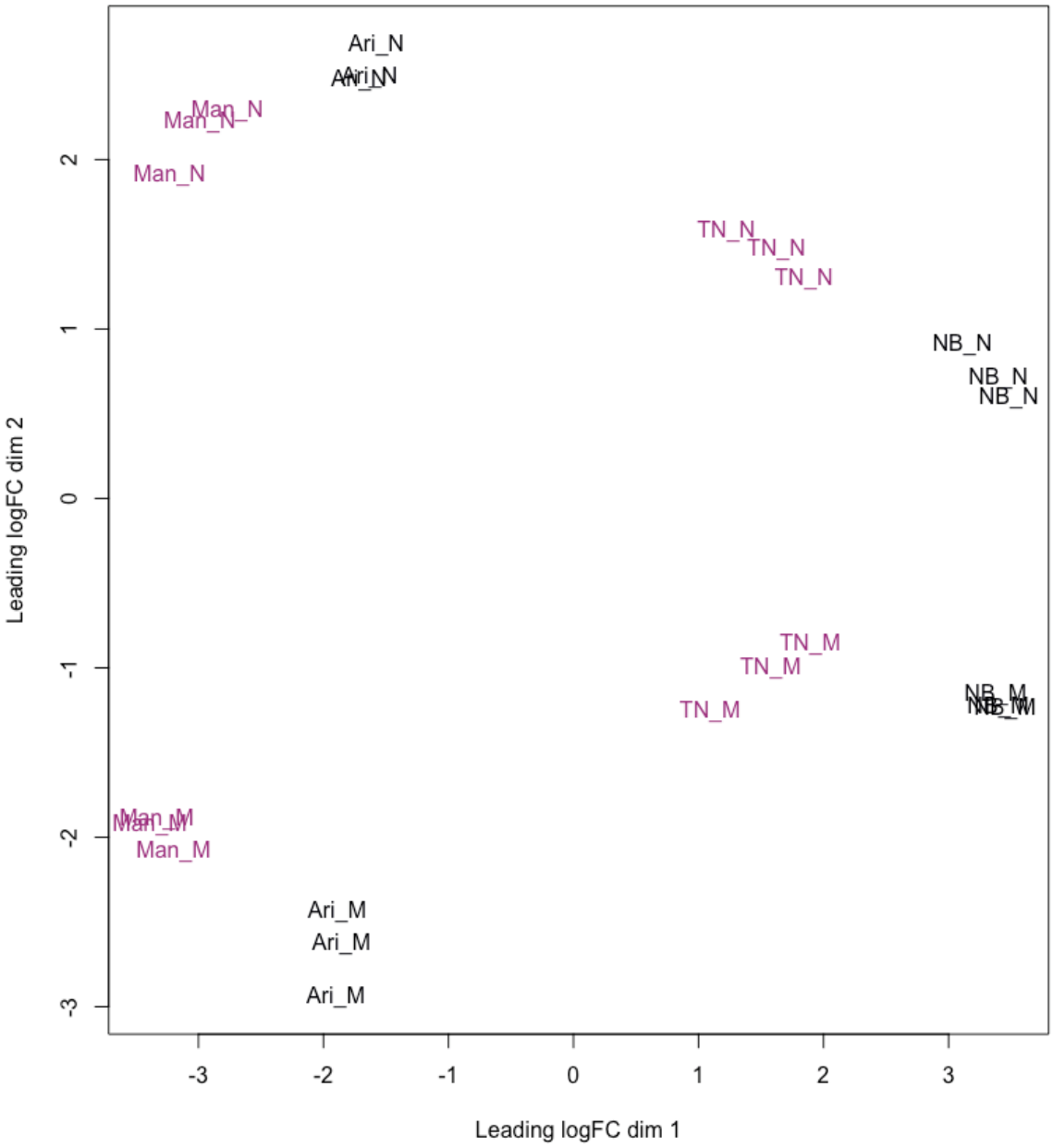


Figure S1 - Multidimensional Scaling Plot representing the log2fold change for all expressed genes in Meristem ("M") and Node ("N") tissue libraries. Libraries are colored by batch (purples/black). Dimension 2 separates libraries by tissue (Node+, Meristem-), while Dimension 1 separates libraries by domestication (Wild+, Domesticated-).

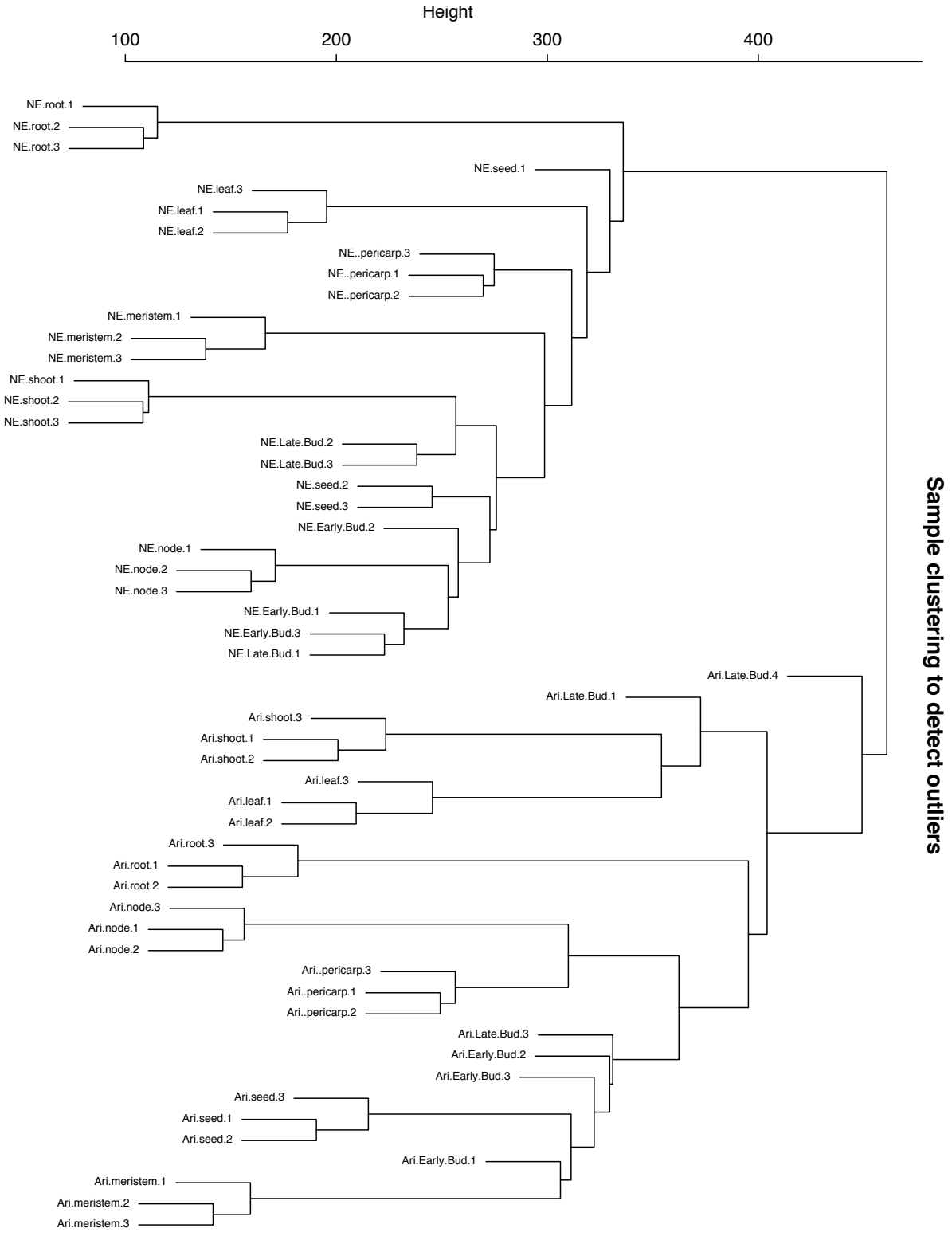


Figure S2 - Libraries clustered by euclidean distance following batch correction prior to gene co-expression network construction to detect outliers. Libraries cluster first by domestication, and then are separated primarily by tissue type.

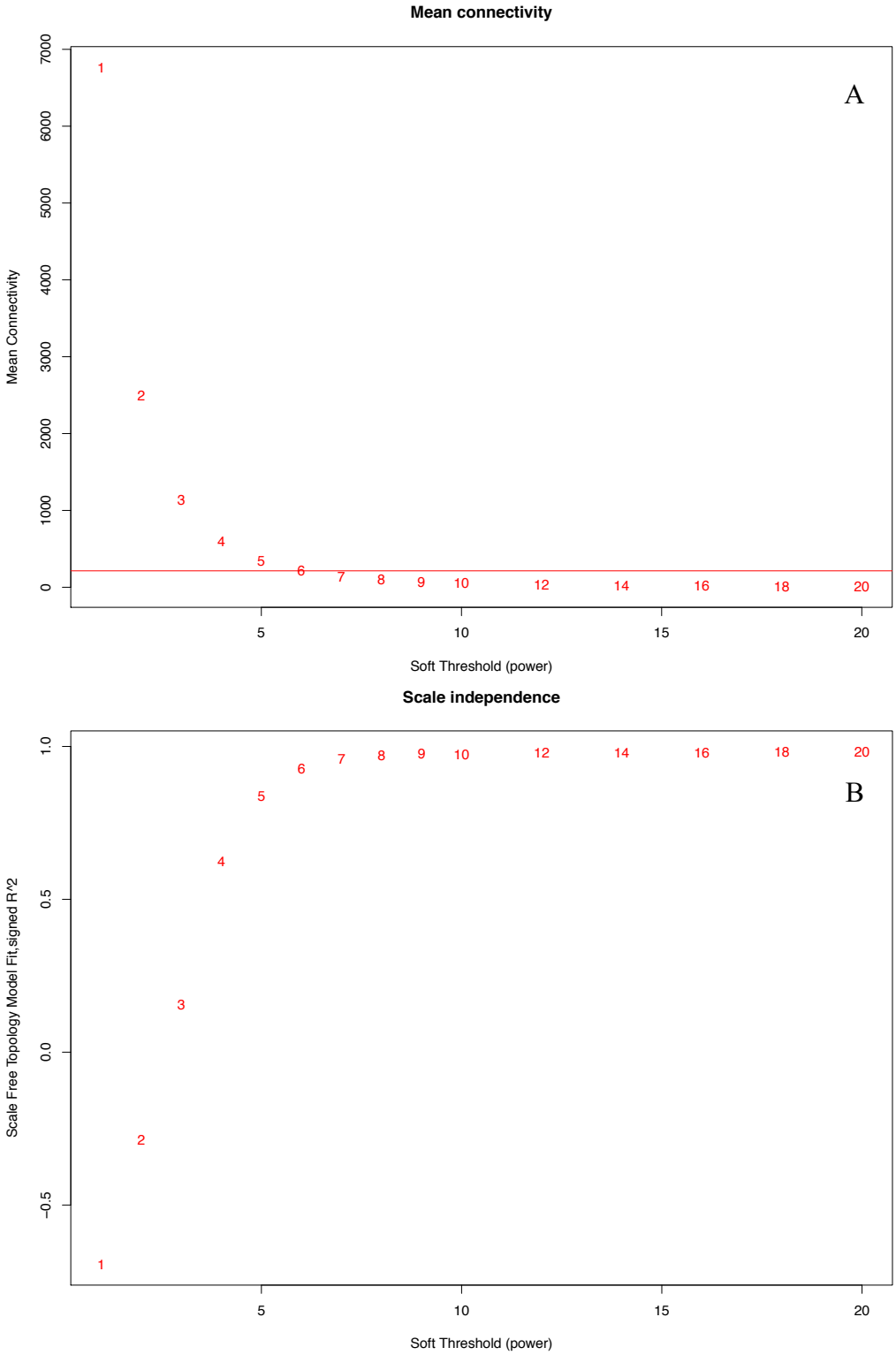


Figure S3 - Analysis of network topology for a range of soft-thresholding powers. Scale-free model fit as a function of the soft thresholding power (A). Mean degree of connectivity as a function of soft thresholding power. Soft-thresholding power of six was chosen, as it provides a scale free topology fit above 0.8 (R^2=0.93) with a mean connectivity substantial enough to calculate networks (mean.k=214).

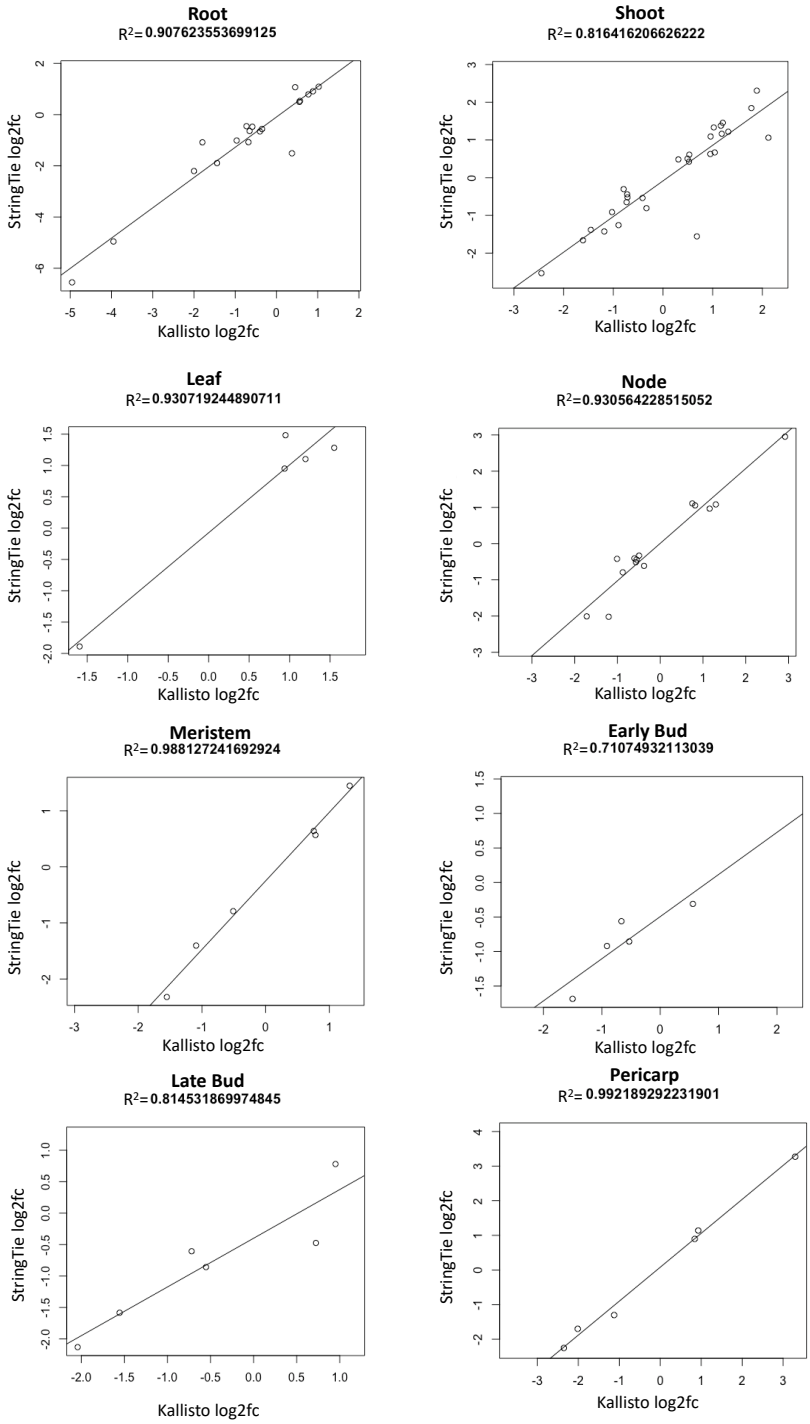


Figure S4 – Regression of expression for DE putative domestication candidate genes identified by both pipelines by tissue. For all tissues, genes exhibit similar expression patterns with substantial R^2 values. Seed is excluded as not enough domestication candidates were DE to perform regression

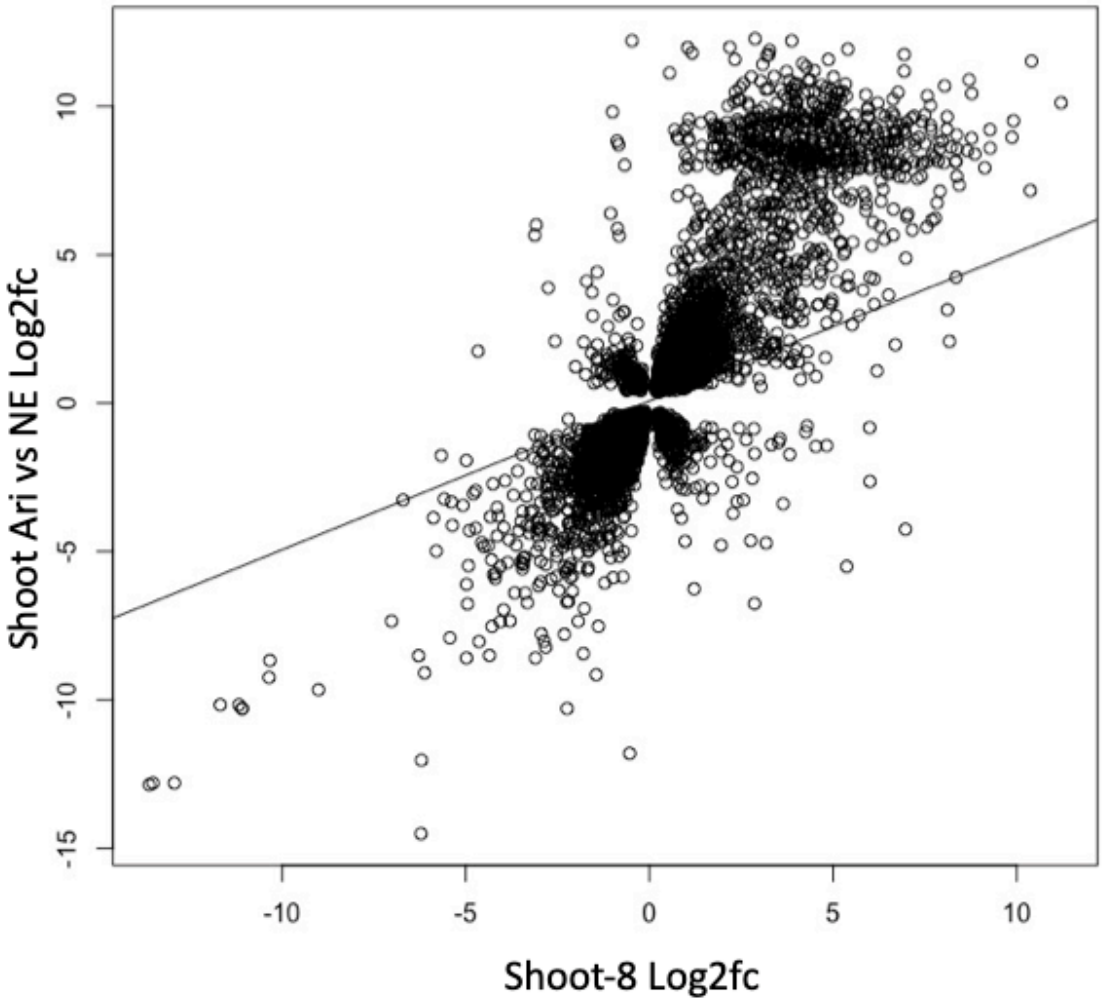


Figure S5 - Regression of average log2fc for genes expressed in both Shoot8 and Shoot Ari vs NE comparisons. R^2=0.71. Pattern is consistent when all genes, regardless of expression (zero enriched data), are analyzed as well, but R^2 drops to 0.56

Table S1 –Accession numbers, genotype names and abbreviations, and domestication status for all sunflower lines used in this study

Accession	Name	Abbreviation	Status
PI 369357	Arikara	Ari	Domesticated
PI 369359	Hopi	Hopi	Domesticated
PI 369360	Seneca	Sen	Domesticated
PI 600717	Mandan	Man	Domesticated
PI 600720	Hidatsa	Hid	Domesticated
PI 435552	Tennessee	TN	Wild
PI 435613	Kentucky	KY	Wild
PI 435616	Missouri	MO	Wild
PI 613727	Arkansas	AR	Wild
PI 659440	Nebraska	NE/Ann1238	Wild

Table S2 – Summary of RNA-sequencing library read counts for both pipelines. (%total aligned genome = StringTie total alignment rate, %pseudoaligned transcriptome = Kallisto total alignment rate)

Genotype	Tissue	Biological replicate	Trimmed read count	%aligned to genome singly	%aligned genome multi	%total aligned genome	%psuedo-aligned transcriptome	Status
AR	Shoot	1	23615723	67.6	4.4	72.0	73.3	Wild
AR	Shoot	2	18008353	66.9	3.8	70.7	73.2	Wild
AR	Shoot	3	14806958	64.8	4.2	69.0	74.0	Wild
Ari	Early Bud	1	21148955	50.1	4.9	55.0	83.6	Domesticated
Ari	Early Bud	2	20264255	55.6	4.9	60.6	82.4	Domesticated
Ari	Early Bud	3	23750287	54.7	4.8	59.4	82.4	Domesticated
Ari	Late Bud	1	20125052	75.8	3.6	79.4	85.3	Domesticated
Ari	Late Bud	2	20967891	48.6	4.9	53.5	84.5	Domesticated
Ari	Late Bud	3	16747022	42.4	5.0	47.4	76.8	Domesticated
Ari	Seed	1	23565440	76.2	7.5	83.7	74.3	Domesticated
Ari	Seed	2	18390483	73.3	9.2	82.5	70.4	Domesticated
Ari	Seed	3	26728580	76.1	4.2	80.3	73.3	Domesticated
Ari	Leaf	1	28729358	35.8	16.0	51.8	88.2	Domesticated
Ari	Leaf	2	30075298	29.7	16.1	45.8	89.1	Domesticated
Ari	Leaf	3	26418889	14.2	17.7	31.9	92.9	Domesticated
Ari	Meristem	1	25540524	56.6	4.5	61.1	82.2	Domesticated
Ari	Meristem	2	28542512	52.8	4.5	57.3	83.0	Domesticated
Ari	Meristem	3	27071991	54.3	4.6	58.9	82.7	Domesticated
Ari	Node	1	29166491	56.2	5.1	61.3	82.9	Domesticated
Ari	Node	2	29245886	62.9	4.6	67.5	81.6	Domesticated
Ari	Node	3	29281738	63.8	4.8	68.6	81.4	Domesticated
Ari	Root	1	27288209	45.0	4.7	49.8	81.9	Domesticated
Ari	Root	2	27864688	43.1	5.1	48.2	84.4	Domesticated
Ari	Root	3	24374405	48.7	4.6	53.4	82.9	Domesticated
Ari	Pericarp	1	35576227	78.3	3.1	81.4	76.2	Domesticated
Ari	Pericarp	2	19381368	76.9	3.1	80.0	76.2	Domesticated
Ari	Pericarp	3	31460033	76.6	3.0	79.5	75.4	Domesticated
Ari	Shoot	1	25010648	25.7	13.9	39.6	89.9	Domesticated
Ari	Shoot	2	31321214	26.2	13.5	39.7	89.8	Domesticated
Ari	Shoot	3	26441073	31.9	11.2	43.2	87.8	Domesticated
Hid	Shoot	1	25551584	71.9	4.9	76.8	75.7	Domesticated
Hid	Shoot	2	48315652	73.2	4.4	77.5	76.0	Domesticated
Hid	Shoot	3	18600637	73.3	4.5	77.8	75.1	Domesticated
Hopi	Shoot	1	19186477	69.0	4.1	73.1	73.9	Domesticated
Hopi	Shoot	2	22821933	70.2	4.4	74.6	73.9	Domesticated
Hopi	Shoot	3	23741833	68.1	4.1	72.2	74.0	Domesticated
KY	Shoot	1	18930597	64.6	3.6	68.1	73.7	Wild

KY	Shoot	2	18077080	67.6	3.4	71.0	72.9	Wild
KY	Shoot	3	23651427	66.6	3.7	70.3	73.1	Wild
MO	Shoot	1	32168645	67.9	3.8	71.6	73.8	Wild
MO	Shoot	2	51331088	66.5	4.4	71.0	73.3	Wild
MO	Shoot	3	25917534	65.9	3.7	69.5	74.8	Wild
Man	Meristem	1	31197943	74.7	2.9	77.6	74.5	Domesticated
Man	Meristem	2	23629382	73.1	2.9	76.0	74.0	Domesticated
Man	Meristem	3	24242588	73.7	2.5	76.2	74.1	Domesticated
Man	Node	1	24121266	75.2	2.9	78.1	74.3	Domesticated
Man	Node	2	30002870	73.2	3.0	76.2	74.0	Domesticated
Man	Node	3	19668391	72.4	3.0	75.4	73.4	Domesticated
Man	Shoot	1	46692603	74.6	4.8	79.3	75.2	Domesticated
Man	Shoot	2	25610265	73.9	5.2	79.0	75.2	Domesticated
Man	Shoot	3	21368014	44.0	6.1	50.2	83.9	Domesticated
NE	Early Bud	1	25159983	54.2	3.8	57.9	80.0	Wild
NE	Early Bud	2	22785514	49.8	3.6	53.5	81.3	Wild
NE	Early Bud	3	19889344	56.6	3.7	60.2	79.1	Wild
NE	Late Bud	1	23244458	48.4	5.3	53.7	82.3	Wild
NE	Late Bud	2	21115084	48.8	5.0	53.8	82.3	Wild
NE	Late Bud	3	18665301	48.5	4.9	53.4	82.8	Wild
NE	Seed	1	25436692	66.8	9.2	76.0	71.4	Wild
NE	Seed	2	31230682	63.2	8.3	71.5	68.3	Wild
NE	Seed	3	24439277	68.4	4.6	72.9	69.6	Wild
NE	Leaf	1	25771773	36.1	14.6	50.8	85.7	Wild
NE	Leaf	3	27713438	42.1	13.4	55.5	83.7	Wild
NE	Leaf	4	30983731	30.5	16.8	47.3	88.3	Wild
NE	Meristem	1	27194818	55.6	3.5	59.1	80.2	Wild
NE	Meristem	2	27569177	55.9	4.0	59.9	80.3	Wild
NE	Meristem	3	28576855	54.4	4.1	58.4	79.5	Wild
NE	Node	1	28204677	59.4	4.6	64.0	79.9	Wild
NE	Node	2	23227422	60.1	4.2	64.3	78.3	Wild
NE	Node	3	28315336	57.5	4.2	61.7	79.3	Wild
NE	Root	1	27505097	55.1	3.6	58.7	76.0	Wild
NE	Root	2	31066271	52.8	3.6	56.5	78.1	Wild
NE	Root	3	25174671	55.9	3.5	59.3	75.8	Wild
NE	Pericarp	1	33225305	69.7	2.6	72.3	72.1	Wild
NE	Pericarp	2	23415634	73.2	2.5	75.8	74.8	Wild
NE	Pericarp	3	27835057	63.1	2.8	65.9	75.5	Wild
NE	Shoot	1	25808005	44.3	7.2	51.6	81.8	Wild
NE	Shoot	2	24702368	41.9	7.6	49.5	83.2	Wild
NE	Shoot	3	25843353	40.9	7.8	48.7	82.7	Wild
Sen	Shoot	1	33954818	71.6	4.7	76.3	75.6	Domesticated
Sen	Shoot	2	38778950	67.1	4.5	71.6	71.0	Domesticated
Sen	Shoot	3	28729565	68.8	5.2	74.0	76.9	Domesticated
TN	Meristem	1	28414129	66.6	2.4	69.0	73.1	Wild

TN	Meristem	2	31818629	64.0	2.4	66.3	72.6	Wild
TN	Meristem	3	23201267	65.5	2.3	67.8	73.3	Wild
TN	Node	1	26479551	65.3	2.7	67.9	73.2	Wild
TN	Node	2	28413609	65.5	2.5	68.0	73.3	Wild
TN	Node	3	29733219	66.6	2.5	69.2	73.3	Wild
TN	Shoot	1	23156594	67.6	4.0	71.6	73.3	Wild
TN	Shoot	2	17093389	67.6	3.8	71.4	74.1	Wild
TN	Shoot	3	24783985	66.6	4.1	70.7	73.5	Wild

Table S3 – Statistical results from ANOVA for effects of tissue, domestication status, batch and sequencing platform on number of cleaned reads to determine if they affect library quality.

	Sum Sq	**Df**	**F value**	**Pr(>F)**
Tissue	2.126E+14	8	0.707	0.684
Status	1.689E+13	1	0.450	0.504
Batch	1.481E+13	1	0.394	0.532
Sequencing platform	1.308E+13	1	0.348	0.557
Residuals	2.930E+15	78	NA	NA

Table S4 – Results from ANOVA of percentage of total reads aligned to the reference genome with HISAT2. Tissue, domestication status, and their interaction are fixed effects with genotype and batch nested within sequencing platform as random effects (results not shown).

	Chisq	**Df**	**Pr(>Chisq)**
Tissue	141.702	8	1.05E-26
Status	11.389	1	0.0007
Tissue:Status	39.135	8	4.64E-06

Table S5 – Results from ANOVA of percentage of total reads aligned to the transcriptome with Kallisto. Tissue, domestication status, and their interaction are fixed effects with genotype and batch nested within sequencing platform as random effects (results not shown).

	Chisq	**Df**	**Pr(>Chisq)**
Tissue	131.521	8	1.37E-24
Status	0.003	1	0.954
Tissue:Status	24.120	8	0.002