Comparative transcriptomics of a monocotyledonous geophyte reveals shared molecular mechanisms of underground storage organ formation.

Carrie M. Tribble[1*], Jesús Martínez-Gómez[2], Fernando Alzate-Guarin[3], Carl J. Rothfels[1], Chelsea D. Specht[2]

1. Department of Integrative Biology and the UC and Jepson Herbaria, University of California, Berkeley, Berkeley, CA 94720

2. School of Integrative Plant Sciences and L.H. Bailey Hortorium, Cornell University, Ithaca, NY 14853 USA

3. Grupo de Estudios Botánicos (GEOBOTA) and Herbario Universidad de Antioquia (HUA), Instituto de Biología, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Calle 67 Nº 53-108, Medellín, Colombia

Corresponding author: Carrie M Tribble, ctribble09@gmail.com, (510) 643-0633

Summary

- Many species from across the vascular plant tree-of-life have modified standard plant tissues into tubers, bulbs, corms, and other underground storage organs (USOs). *Bomarea multiflora* (Alstroemeriaceae) is a tropical climbing monocot with unique underground morphology, including tuberous roots. We take a comparative transcriptomics approach to characterizing the molecular mechanisms of tuberous root formation in *B. multiflora* and compare these mechanisms to those identified in other underground storage structures across diverse plant lineages.
- We sequenced transcriptomes from the growing tip of four tissue types (aerial shoot, rhizome, fibrous root, and root tuber) of three individuals of *B. multiflora*. We identify differentially expressed isoforms and test the expression of candidate genes that have been implicated in underground storage in other taxa.
- We identify 271 genes that are differentially expressed in root tubers versus fibrous roots, including genes implicated in cell wall modification, environmental signaling and flowering time, and starch biosynthesis. We also identify a phosphatidylethanolamine-binding protein (PEBP) that is over-expressed in tuberous roots.
- These findings demonstrate that deeply parallel processes underlie the formation of underground storage structures despite long evolutionary distances between taxa and non-homologous morphologies, illustrating that repeated co-option of similar genetic pathways can lead to convergent morphologies.

Introduction

The vast majority of scientific attention in botanical fields focuses exclusively on aboveground biomass. However, a holistic understanding of plant morphology, ecology, and evolution requires that considerable research effort go towards generating a comprehensive understanding of belowground biomass. While most studies of plant form and function focus on aboveground organs, on average 50% of an individual plant's biomass lies beneath the ground (Niklas, 2005). Often, belowground biomass is thought to consist solely of standard root tissue, but in some cases, plants modify 'ordinary' structures for specialized underground functions. Plants called "geophytes" are toward the extreme end of this belowground/aboveground allocation spectrum. These species rely on nutrients stored in belowground organs (underground storage organs or "USOs"), and their ephemeral aboveground parts resprout from buds located on belowground organs (Raunkiaer, 1934; Dafni *et al.*, 1981a,b; Al-Tardeh *et al.*, 2008; Veselý *et al.*, 2011). Geophytes are ecologically and economically important, morphologically diverse, and have evolved independently in all major groups of vascular plants except gymnosperms (Howard *et al.*, 2019a,b). These plants and their associated underground structures are a compelling example of evolutionary convergence; diverse taxa form a variety of structures, often from different tissues, that serve parallel physiological and ecological functions. However, our understanding of the molecular processes that drive this convergence, and the extent to which these processes are themselves convergent, remains remarkably limited, due in part to the lack of molecular studies in geophyte lineages. This is particularly true for monocotyledonous geophytic taxa, which comprise the majority of ecologically and economically important geophyte diversity.

Some of the world's most important crop plants are geophytes, including potato (stem tuber, *Solanum tuberosum*), sweet potato (tuberous root, *Ipomoea batatas*), yam (epicotyl- and hypocotyl-derived tubers, *Dioscorea spp.*), cassava (tuberous root, *Manihot esculenta*), radish (swollen hypocotyl and taproot, *Raphanus raphanistrum*), onion (bulb, *Allium cepa*), lotus

(rhizome, *Nelumbo nucifera*), and more. While several of these crop plants are well studied and have sequenced genomes or other genetic or genomic data that may inform the molecular mechanisms underlying underground storage organ development, most research has focused on a select few taxa that do not represent the diversity of geophyte morphology, taxonomy, or ecology. In particular, most genetic research on geophytes and their associated underground storage organs has been conducted in eudicotyledons (with the exception of important studies in onion, such as in Navarro *et al*., 2011), which neglects a huge proportion of geophytic diversity in the monocots. Furthermore, geophytes form important components of many ecosystems, particularly Mediterranean biomes and other seasonal habitats, as their underground nutrient reserves fuel regrowth following periods of seasonal dormancy, prolonged dormancy, or short term resource limitation (Rundel, 1996; Hoffmann *et al.*, 1998; Parsons, 2000; Proches *et al.*, 2006; Cuéllar-Martínez & Sosa, 2016; Sosa & Loera, 2017; Ott *et al.*, 2019; Howard *et al.*, 2019b). The geophytic habit has evolved multiple times across the vascular plant tree of life, and geophyte lineages have been shown to diversify at faster rates than do related non-geophytic taxa, particularly in taxa with bulbs, corms, and tubers (Howard *et al.*, 2019b), indicating that the geophytic habit may be correlated with increased diversification and/or reduced extinction rates.

Underground storage organs originate from all major types of plant vegetative tissue: roots, stems, leaves, and hypocotyls. Bulbs, corms, rhizomes, and tubers are some of the most common underground storage organ morphologies (Pate & Dixon, 1982), but the full breadth of morphological variation in USOs includes various root modifications (tuberous roots, taproots, etc.), swollen hypocotyls that merge with swollen root tissue (e.g. *Adenia*: Hearn, 2009), and intermediate structures such as rhizomes where the terminal end of the rhizome forms a bulb from which aerial shoots emerge (e.g. *Iris*: Wilson, 2006). Despite this morphological complexity, USOs all develop through the expansion of standard plant tissue, either derived from the root or shoot, into swollen, discrete storage organs. These storage organs also serve similar functions as belowground nutrient reserves (Veselý *et al.*, 2011), often containing starch or other non-structural carbohydrates (NSCs), storage proteins, and water. The functional and physiological similarities of underground storage organs may be driven by deep molecular homology with parallel evolution in the underlying genetic architecture of storage organ development, despite differences in organismal level morphology and anatomy.

The economic importance of some geophytes and the relevance of understanding the formation of storage organs for crop improvement have motivated studies on the genetic basis for storage organ development in select taxa. Potato has become a model system for understanding the molecular basis of USO development, and numerous studies have demonstrated the complex and interacting roles of plant hormones such as auxin, abscisic acid, cytokinin, and gibberellin on the tuber induction process (reviewed in Hannapel *et al.*, 2017). These hormones have been additionally identified in USO formation in other tuberous root crops including sweet potato (Noh *et al.*, 2010; Dong *et al.*, 2019) and cassava (Melis & van Staden, 1985; Sojikul *et al.*, 2015), in rhizome formation in *Panax japonicus* var. *major* (Tang *et al.*, 2019) and *Nelumbo nucifera* (Cheng *et al.*, 2013b; Yang *et al.*, 2015), and in corm formation in *Sagittaria trifolia* (Cheng *et al.*, 2013a), suggesting that parallel processes trigger tuberization in both root- and stem-originating USOs. Three primary mobile signaling genes have been implicated in triggering the onset of tuberization in potato: SELF-PRUNING6A (StSP6A), a FLOWERING LOCUS T-like gene (FT gene); StBEL5 in the BEL1-like transcription factor family; and POTH1, a KNOX type transcription factor. FT-like genes have been additionally implicated in USO formation in *Dendrobium* (Wang *et al.*, 2017), *Callerya speciosa* (Xu *et al.*, 2016), tropical lotus (*Nelumbo*

nucifera; Yang *et al.*, 2015), and onion (*Allium cepa*; Lee *et al.*, 2013), indicating either deep homology of FT involvement in USO formation across angiosperms or multiple independent co-option of FT orthologues in geophytic taxa.

The lateral expansion of roots into tuberous roots may be driven by either cellular proliferation concurrent with primary growth (primary thickening growth; Kaplan, unpublished), by cellular proliferation subsequent to primary growth (secondary thickening growth; Kaplan, unpublished), by cellular expansion, where individual cells expand in size, or by a combination of these processes. Expansion in plant cells requires the modification of the rigid cell wall to accommodate increases in cellular volume (Dolan & Davies, 2004; Humphrey *et al.*, 2007), and genes such as expansins have been implicated in cellular expansion during tuberous root development in some taxa such as cassava and *Callerya speciosa* (Sojikul *et al*., 2015; Xu *et al.*, 2016). Recent studies of the tuberous roots of sweet potato (*Ipomoea batatas*) and other members of Convolvulaceae indicate that USO formation in these taxa involves a MADS-box gene implicated in the vascular cambium (SRD1; Noh *et al.*, 2010) and a WUSCHEL-related homeobox gene (WOX4; Eserman *et al.*, 2018), also involved in vascular cambium development. Additional work on cassava (*Manihot esculenta*) also suggests that tuberous root enlargement is due to secondary thickening growth originating in the vascular cambium (Chaweewan & Taylor, 2015). However, geophytes are especially common in monocotyledonous plants (Howard *et al.*, 2019a,b), which lack a vascular cambium entirely. No previous study has addressed the molecular mechanisms of USO development in this major clade, so the causes of root thickening are particularly enigmatic. Do monocots form tuberous roots through genetic machinery that shares deep homology with the eudicot vascular-cambium-related pathways, or have they evolved an entirely independent mechanism?

Bomarea multiflora (L. f.) Mirb. is a scandent monocotyledonous geophyte that is native to Venezuela, Colombia, and Ecuador (Hofreiter, 2008) and is invasive in New Zealand (National Pest Plant Accord: https://www.mpi.govt.nz/protection-and-response/long-term-pest-management/national-pest-plant-accord). It typically grows in moist cloud forests between 1800 3800 meters elevation (Hofreiter, 2008). *Bomarea multiflora* is an excellent model in which to study the molecular mechanisms underlying underground storage organ formation in the monocots because it has two types of underground modifications: tuberous roots and rhizomes. However, prior to this study, no genomic or transcriptomic data was available for any species of *Bomarea*. Comparative transcriptomics permits comprehensive examination of the molecular basis of development, tissue differentiation, and physiology in ecologically relevant taxa by comparing the genes expressed in different organs, developmental stages, or ecological conditions (Ekblom & Galindo, 2011; Oppenheim *et al.*, 2015). Because no prior genomic or transcriptomic data is needed for comparative transcriptomic studies, this method is especially appropriate for studies of non-model organisms and can yield novel insights into the expression profiles of specific tissues. In this study, we investigate the molecular mechanisms underlying the formation of tuberous roots in *Bomarea multiflora* using a comparative transcriptomics approach and quantify the extent to which these mechanisms are shared across the taxonomic and morphological breadth of geophytic taxa.

Materials and Methods

Greenhouse and laboratory procedures Seeds were collected from a single inflorescence of *Bomarea multiflora* in Antioquia, Colombia [vouchered as Tribble 194, deposited at UC; Index

193 Herbariorium http://sweetgum.nybg.org/science/ih/] and germinated in greenhouse conditions at
194 the University of California, Berkeley designed to replicate *B. multiflora*'s native conditions
195 (70°F – 85°F and ~50% humidity). Six months after germination, three sibling individuals were
196 harvested as biological replicates. The aerial shoot apical meristem (SAM) of a single branch, the
197 rhizome apical meristem (RHI), root apical meristems (RAM) of several fibrous roots, and the
198 growing tip of a tuberous root (TUB; Figure 1) were dissected from each of three individuals, for
199 a total of 12 tissue samples. Samples were immediately frozen in liquid nitrogen and maintained
200 at -80°C until extraction. Total RNA was extracted from all samples using the Agilent Plant
201 RNA Isolation Mini Kit (Agilent, Santa Clara, Ca), optimized for non-standard plant tissues,
202 especially those that may be high in starch. As this protocol did not work well for the vegetative
203 SAM samples, RNA was extracted from two dissected SAMs using an alternative plant tissue
204 RNA extraction protocol (Yockteng *et al.*, 2013). Quality of total RNA was measured with Qubit
205 (ThermoFisher, Waltham MA) and Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA);
206 if needed, a Sera-Mag bead clean-up was used to further clean extracted RNA (Yockteng *et al.*,
207 2013). Two SAM samples failed to extract at sufficient concentrations, so we harvested the
208 SAMs of two additional individuals, froze, and extracted using the Yockteng et al. 2013
209 protocol. Samples with an RNA integrity (RIN) score > 7 proceeded directly to library prep. The
210 KAPA Stranded mRNA-Seq Kit (Kapa Biosystems, Waltham MA) protocol was used for library
211 prep. Half reactions were used with an input of at least 500 ng of RNA; however, most samples
212 (all but two) had ~1 ug of RNA. RNA fragmentation time depended on RIN score (7< RIN < 8: 4
213 min; 8 < RIN < 9: 5 min; 9 < RIN: 6 min). Samples were split in half after the second post-
214 ligation clean up (Step 10 in the Kapa protocol) in order to fine-tune the enrichment step. The
215 first half of the samples were amplified with 12 PCR cycles; this proved too low and was
216 increased to 15 cycles for the second half of the samples. Samples were combined and library
217 quality was assessed with a Bioanalyzer 2100 using the DNA 1000 kit. A bead clean-up was
218 performed on libraries showing significant adaptor peaks (Yockteng *et al.*, 2013). Samples were
219 cleaned, multiplexed, and sequenced on a single lane of HiSeq4000 at the California Institute for
220 Quantitative Biosciences (QB3) Vincent J. Coates Genomics Sequencing Lab.
221
222 **Bioinformatics Data Processing** Raw reads were cleaned, processed, and assembled using the
223 Trinity RNA-Seq De novo Assembly pipeline (Grabherr *et al.*, 2011) under the default settings
224 unless otherwise stated in associated scripts. All analyses were run using the Savio
225 supercomputing resource from the Berkeley Research Computing program at UC Berkeley.
226 Reads were cleaned with Trim Galore!
227 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), keeping unpaired reads and
228 using a minimum fragment length of 36 base pairs. Data from all tissue types and biological
229 replicates were concatenated in order to generate a consensus transcriptome, assembled de novo
230 from the concatenated data using Trinity. Each sample was compared back to the assembled
231 consensus transcriptome, aligned using Bowtie 2 (Langmead & Salzberg, 2012), and quantified
232 using RSEM (RNASeq by Expectation Maximization; Li & Dewey, 2011). The consensus
233 transcriptome was annotated with a standard Trinotate pipeline (https://trinotate.github.io/),
234 comparing assembled transcripts to SWISS-PROT (Boeckmann *et al.*, 2003), RNAmmer
235 (Lagesen *et al.*, 2007), Pfam (Finn *et al.*, 2014), eggNOG (Powell *et al.*, 2014), KEGG (Tanabe
236 & Kanehisa, 2012), and Gene Ontology (Gene Ontology Consortium, 2004) databases. The
237 concordance of biological replicates was tested by looking for significant differences between
238 the total number of fragments per replicate, by comparing the transcript quantities of all
239 replicates to each other, and by checking the correlations between replicates. Transcripts with
240 less than 10 total counts were discarded for all downstream analyses. Transcript counts were

transformed using the variance-stabilized transformation (VST) and all 12 samples were compared using a principal components analysis. All scripts for these analyses are available on GitHub (github.com/cmt2/bomTubers).

Data Analysis Differentially expressed isoforms (hereafter referred to as DEGs) between fibrous (FR) and tuberous (TR) roots were identified with DESeq2 package in R (Love *et al.*, 2014), extracting the results of comparing tuberous roots to fibrous roots and using a p-adjusted cut-off (padj, uses a Benjamini-Hochberg correction for false discovery rate) of 0.01 and a log2-fold change cut-off of 2 to specifically only statistically significant and sufficiently differentially expressed isoforms for downstream analyses. To identify specific candidate genes, gene families, or molecular processes that might be involved in the development of underground storage organs, we surveyed the literature for recent publications on the molecular basis of USO formation. For each group of genes hypothesized to be involved in USO formation (either gene families or molecular/ physiological processes), the annotated transcriptome was queried for transcript annotations matching the associated process or family (See Table S1 for the specific search terms used). The resulting transcripts were 1) compared using the distribution of the group's log2-fold change values to the distribution of values from all isoform using a non-parametric Mann-Whitney test, and 2) tested for the presence of any statistically significant differentially expressed isoforms. For all targeted candidate genes, the amino acid sequence of the candidate gene was blasted to the assembled consensus transcriptome (see Table S2 for the blasted sequences specifications) using an e-value cut-off of 0.01 to assess if the identified homologs were differentially expressed. All associated scripts for data analysis are available on GitHub (github.com/cmt2/bomTubers).

The evolutionary history of the phosphatidylethanolamine-binding protein (PEBP) gene family was reconstructed by combining amino acid sequences from an extensive previously published alignment (Liu *et al.*, 2016) with the addition of sequences specifically implicated in USO formation in onion and potato or from geophytic taxa such as *Narcissus tazetta* (accession AFS50164.1), *Tulipa gesneriana* (accessions MG121853, MG121854, and MG121855), *Crocus sativa* (saffron, accession ACX53295.1), and *Lilium longiflorum* (accessions MG121858, MG121857, MG121859) (Navarro *et al.*, 2011; Tsaftaris *et al.*, 2012; Lee *et al.*, 2013; Li *et al.*, 2013; Leeggangers *et al.*, 2017) and with copies identified in our transcriptome. Amino acids were aligned with MAFFT as implemented in AliView v1.18 (Larsson, 2014), using trimAl v1.4.rev15 (Capella-Gutierrez *et al.*, 2009) with the -gappyout option. The best evolutionary model was selected with ModelTest-NG v0.1.5 (Darriba *et al.*, 2016) and unrooted gene trees were reconstructed under a maximum likelihood framework as implemented in IQtree (Nguyen *et al.*, 2014), run on XSEDE using the CIPRES portal (Miller *et al.*, 2010).

Results

Transcriptome Data A total of 359 M paired-end 100 bp reads were recovered from the single HiSeq 4000 lane for the multiplexed 12 samples, to be made available as SRAs in an NCBI BioProject prior to publication. The assembled consensus transcriptome consists of 370,672 transcripts, corresponding to 224,661 unigenes, also to be made available on Dryad prior to publication. The combined data have a GC content of 45.14%, N50 of 1191 bp, median transcript length of 317 bp, and mean transcript length of 556.95 bp. All four tissue types showed concordance between the three biological replicates with generally 1:1 ratios of transcript quantities to each other (see Figures S1 – S4) so no further replicates were discarded and all

differentially expression analyses incorporated data from all three biological replicates. A principal component analysis (PCA) of the VST transcript counts (Figure 2) shows that the first PC axis (45% of the variance in samples) generally explains the variation between tissue types. The shoot tissues (SAM and RHI) cluster separately, while the root tissues (ROO and TUB) cluster together. The underground rhizome samples (RHI) fall out intermediate between the aerial shoot samples (SAM) and the underground root and tuber samples (ROO and TUB) along this axis. The co-clustering of fibrous and tuberous root samples in the PCA indicates that the overwhelming, general pattern of expression between all root samples is similar, especially in contrast to the very distinct shoot samples. This broad pattern of similarity provides an excellent opportunity to identify the particular genes that cause such obvious morphological differences between the tuberous and fibrous roots, given their similar background patterns of expression. The second PC axis (22% of variance) generally explains variance between biological replicates, with Individual A particularly distinct from other individuals.

Differential Expression A total of 271 differentially expressed isoforms (DEGs) were recovered between fibrous and tuberous roots (FR vs. TR) (Table S3). Of these, 226 correspond to regions of the annotated consensus transcriptome with functional annotations. The most common gene ontology (GO) terms associated with these DEGs are presented in Table 1. The top cellular components include the nucleus, the cell membrane and cell wall, and plastids; the top molecular functions include kinase activity and binding to ATP, nucleic acids, proteins, and various metals; and the top biological functions include cell wall organization, morphological development, environmental sensing, and protein transport. Of the 271 DEGs, 126 (46.5%) were over-expressed in tuberous roots while the remaining 145 (53.5%) were under-expressed. All top ten most differentially expressed isoforms (the ten DEGs with the highest absolute value log2-fold change values between fibrous and tuberous roots) were functionally annotated. These ten DEGs are implicated in various cellular and biological processes (Table 2). All but nine of these top ten DEGs are overexpressed in tuberous roots and are generally implicated in nucleotide and ATP binding, cell wall modification, root morphogenesis, and carbohydrate and fatty acid biosynthesis. The top most differentially expressed isoform with a 40.25 log2-fold change value, TRINITY_DN116220_c0_g1_i4, is a Zinc finger CCCH domain-containing protein 55, a possible transcription factor of unknown function. Other notable top DEGs include TRINITY_DN128685_c1_g3_i4, callose synthase 3, which regulates cell shape, TRINITY_DN121298_c2_g2_i5, a heat shock protein, TRINITY_DN127064_c0_g3_i1, an LRR receptor-like serine implicated in lateral root morphogenesis, and TRINITY_DN121430_c10_g2_i, a carbohydrate metabolism protein.
The tenth most differentially expressed DEG, under-expressed in tuberous roots, is implicated in abscisic acid signaling.

Parallel Processes Across Taxa Eleven gene groups — either gene families, physiological signaling pathways, or biosynthesis pathways — were identified that have been implicated in USO formation by expression of functional analyses across various geophytic organisms: abscisic acid response genes, calcium-dependent protein kinases (CDPK), expansins, lignin biosynthesis, MADS-Box genes, starch biosynthesis, auxin response genes, cytokinin response genes, 14-3-3 genes, gibberellin response genes, and KNOX genes (See Figure 3). For each group, putative homologs were located in *Bomarea multiflora* and their expression patterns were analyzed. Of these 11 gene groups, the log2-fold change values of six are distributed significantly differently than the overall distribution of log2-fold change values for all isoforms; in two cases (starch biosynthesis and MADS-Box genes) the expression levels overall are

337 significantly greater than expected (generally over-expressed in TR compared to FR) and in the
338 remaining four cases (abscisic acid, CDPK, expansins, and lignin biosynthesis) the groups' log2-
339 fold change values are under-expressed. Fifteen individual DEGs were identified in these gene
340 groups (Table 3); interestingly, there seems to be no generalizable relationships between the
341 significance and directionality of a particular group's distribution with the presence and
342 directionality of log2-fold change values for individual DEGs. For example, the expression
343 distribution of gibberellin genes does not deviate significantly from the global pool of isoforms,
344 but there is one significantly under-expressed DEG in the gibberellin group; similarly, the CDPK
345 genes are under-expressed as a group, but the only CDPK-related significant DEG is over-
346 expressed (Table 3).
347
348 Five specific candidate genes were identified from the literature: qRT9 has been implicated in
349 root thickening in rice (Li *et al.*, 2015); IDD5 and WOX4 are implicated in starch biosynthesis
350 and TR formation, respectively, in Convolvulaceae (Eserman *et al.*, 2018); sulfite reductase is
351 associated with TR formation in *Manihot esculenta* (Sojikul *et al.*, 2010); and FLOWERING
352 LOCUS T (FT) has been implicated in signaling the timing of USO formation in a variety of
353 taxa, notably *Allium cepa* and *Solanum tuberosum* (Navarro *et al.*, 2011; Hannapel *et al.*, 2017).
354 We recover between nine and 63 putative homologs of these candidates using a BLAST E-value
355 cut-off of 0.01 (Table 4), but only one putative homolog is significantly differentially expressed
356 (padj < 0.01): a putative FT homolog (TRINITY_DN129076_c1_g1_i1), further investigated in
357 PEBP Gene Family Evolution (below). One putative qRT9 homolog is marginally significant
358 (padj = 0.050), and the E-value from the BLAST result to this isoform was 0.09. Given these
359 marginal significance values, it is likely the result is spurious and we do not follow up with
360 further analysis.
361
362 **PEBP Gene Family Evolution** Thirty-seven *Bomarea* isoforms were identified as putative
363 FLOWERING LOCUS T (FT) homologs. The longest isoform per gene was selected to include
364 in an alignment of phosphatidylethanolamine-binding protein (PEBP) amino acid sequences.
365 Coding sequences from the *Bomarea multiflora* transcriptome were translated to amino acid
366 sequences using TransDecoder v5.5.0 (Haas *et al.*, 2013), removing isoforms that failed to align
367 properly; ultimately, we include six sequences, including the significantly differentially
368 expressed copy. Using the JTT+G4 amino acid substitution models (Jones *et al.*, 1992) we
369 recover three major clusters in our unrooted gene tree, all with strong bootstrap support (Figure
370 4a); these correspond to the FT cluster, TERMINAL FLOWER 1 (TFL1) cluster, and MOTHER
371 OF FT AND TFL1 (MFT) cluster recovered in previous analyses (Liu *et al.*, 2016). Three of the
372 six *Bomarea multiflora* isoforms fall out with FT genes and three fall out with TFL1 genes. The
373 *Bomarea* DEG homolog is highly supported in the TFL1 cluster with sequences from other
374 monocot taxa (Figure 4b). FT homologs from *Allium cepa* and *Solanum tuberosum* that have
375 been functionally implicated in stem tuber and bulb formation, respectively, are in the FT cluster
376 but do not cluster together; rather all USO-implicated PEBP genes are more closely related to
377 non-USO copies than to each other.
378
379 **Discussion**
380
381 **How to Make a Tuberous Root** The top 10 most highly represented biological processes of the
382 271 Differentially expressed isoforms (DEGs) include cell wall organization, responses to
383 environmental signals, growth and development, and carbohydrate biosynthesis (Table 1).
384 Together, these processes describe the various components of development by which the plant

modifies fibrous roots into tuberous roots: 1) *how* expansion occurs, 2) *when* tuberization is triggered, and 3) *what* the tuberous roots store.

Root expansion likely occurs due to cellular expansion and primary thickening growth in *B. multiflora*. Due to the absence of a vascular cambium, secondary growth is not likely to be involved, despite the prevalence of this mechanism in other taxa such as sweet potato (Noh *et al.*, 2010; Eserman *et al.*, 2018) and cassava (Melis & van Staden, 1985). DEGs with cell wall functional annotations, the most common GO biological process, likely contribute to root enlargement through permitting cellular expansion. For example, pectinesterase TRINITY_DN122210_c6_g1_i1 (log2-fold change = 21.91; padj = $1.22E^{-8}$) modifies pectin in cell walls leading to cell wall softening, as demonstrated, for example, in Arabidopsis (Braybrook & Peaucelle, 2013). Interestingly, expansins were not over-expressed in *B. multiflora*, though this is the mechanism by which cell expansion occurs in other taxa (see Expansins discussion below). The DEGs functionally annotated as contributing to cell division (one of the top GO biological processes in the DEG dataset) may contribute to root enlargement through increased cellular proliferation at the growing tip of the tuberous root (primary thickening growth).

Several of the top GO biological processes are responses to environmental stimuli, including cold, water deprivation, flowering, and defense, which may trigger tuberization. Flowering development genes in particular may be co-opted for tuber formation, a hypothesis discussed in more detail below (PEBP gene family evolution). Tuberization signaling may also be mediated by callose production, influencing symplastic signaling pathways through plasmodesmata modification. Callose synthase 3 is one of the most highly differentially expressed DEGs (TRINITY_DN128685_c1_g3_i4, Table 2). Callose is a much less common component of cell walls than is cellulose (Schneider *et al.*, 2016), but it is often implicated in specialized cell walls and in root-specific expression (Vatén *et al.*, 2011; Benitez-Alfonso *et al.*, 2013). Callose synthase has been implicated in the development of other unique root-based structures such as root nodules (Gaudioso-Pedraza *et al.*, 2018) and mutations in callose synthase 3 affect root morphology (Vatén *et al.*, 2011), suggesting that callose synthase 3 plays an integral role in triggering tuberous roots development in *B. multiflora* through symplastic signaling pathways. Callose signaling-induced USO formation has not previously been reported and be unique to *B. multiflora* or to monocotylendous taxa.

Finally, starch is thought to be the primary nutrient reserve in *Bomarea* tubers (Kubitzki, 1998). Many previous studies have found evidence of overexpression of carbohydrate and starch synthesis molecules in USOs (for example in sweet potato; Eserman *et al.*, 2018)). Differentially expressed isoforms implicated in the carbohydrate metabolic process support the presence of active starch synthesis in our data. One of the most differentially expressed isoforms is a homolog of sucrose non-fermenting 4-like protein (Table 2, TRINITY_DN121430_c10_g2_i1) and participates in carbohydrate biosynthesis, demonstrating that *B. multiflora* tubers were actively synthesizing starch when harvested. Additionally, genes implicated in defense response, such as TRINITY_DN127064_c0_g3_i1 (LRR receptor-like serine/ threonine-protein kinase HSL2, Table 2) may be differentially expressed in tuberous roots to protect starch reserves against potential predation by belowground herbivores.

Similarities in Molecular Mechanisms of USO Formation We identify six molecular processes, previously implicated in USO formation in other taxa, which are either over- or under-

433 expressed in the tuberous roots of *Bomarea multiflora* (Figure 3). These processes show parallel
434 function across deeply divergent evolutionary distances and in distinct plant structures.
435
436 ***Abscisic Acid (ABA)*** has been shown to increase initially and then decrease during USO
437 formation in the rhizomes of *Nelumbo nucifera* (Yang *et al.*, 2015), the tuberous roots of sweet
438 potato *Ipomoea batatas* (Dong *et al.*, 2019), and the stem tubers of potato *Solanum tuberosum*
439 (Xu *et al.*, 1998). We find that ABA signaling genes are under-expressed in tuberous roots
440 (Figure 3), perhaps indicating that our samples were in a later developmental stage, characterized
441 by lower ABA levels. Experimental manipulation of *B. multiflora* tuberous roots and
442 developmental time series would be needed to confirm the role of ABA in tuberous root
443 formation. An ABA signaling DEG (Table 3) is one of the most differentially expressed isoforms
444 (TRINITY_DN122787_c0_g1_i1: Protein IQ-DOMAIN 32, Table 2). This isoform and other
445 ABA genes may signal the cessation of continued elongation in monocot tuberous roots, similar
446 to their role in deeply divergent taxa and non-homologous USOs.
447
448 ***Calcium-dependent Protein Kinases (CDPKs)*** play an integral role in tuber formation in cassava
449 *Manihot esculenta* (Sojikul *et al.*, 2010). Sujikul et al. (2010) propose that CDPKs may signal
450 the initiation of tuberous root development, similarly to the process described in stem tuber
451 formation in potato (Raíces *et al.*, 2003) and rhizome development in *Nelumbo nucifera* (Cheng
452 *et al.*, 2013b). In these studies, CDPKs are over-expressed and signal the initiation of USO
453 development. Our results show that on average CDPKs as a group are under-expressed in the
454 tuberous roots of *B. multiflora.* However, the only significant CDPK DEG
455 (TRINITY_DN124121_c3_g1_i11: Calcium-dependent protein kinase 2) is over-expressed,
456 suggesting that CDPK 2 expression plays a role in initiating tuberous root development and that
457 CDPK involvement in tuberous root formation is ubiquitous in all studied taxa.
458
459 ***Expansins*** are cell wall modifying genes known to loosen cell walls in organ formation (Dolan
460 & Davies, 2004; Humphrey *et al.*, 2007; Braybrook & Peaucelle, 2013). Their involvement in
461 USO formation has been documented in the tuberous roots of cassava (Sojikul *et al.*, 2015), and
462 *Callerya speciosa* (Xu *et al.*, 2016) the rhizomes of *Nelumbo nucifera* (Cheng *et al.*, 2013b), the
463 tuberous roots of various Convolvulaceae (Eserman *et al.*, 2018), and the stem tubers of potato
464 (Jung *et al.*). As a group expansins are under-expressed in tuberous compared to fibrous roots,
465 but none are statistically significant, so it seems unlikely that expansins play an important role in
466 tuberous root formation in *Bomarea multiflora*. It is possible that expansin involvement in USO
467 formation is unique to eudicots.
468
469 ***Lignin*** biosynthesis genes are under-expressed in several geophytic taxa with tuberous roots,
470 including cassava (Sojikul *et al.*, 2015), wild sweet potato (*Ipomoea trifida*; (Li *et al.*, 2019), and
471 *Callerya speciosa* (Xu *et al.*, 2016). Similarly, we find that lignin biosynthesis overall is under-
472 expressed in tuberous compared to fibrous roots, and one isoform in particular is significantly
473 under-expressed: TRINITY_DN125451_c5_g1_i3: Cinnamoyl-CoA reductase-like SNL6 (Table
474 3). This gene has been found to significantly decrease lignin content without otherwise affecting
475 development in tobacco (Chabannes *et al.*, 2001). Decreased lignin in tuberous roots may further
476 allow for cell expansion and permit lateral swelling of tuberous roots during development.
477
478 ***MADS-Box*** genes are implicated in USO formation in the tuberous roots of wild sweet potato
479 (*Ipomoea trifida*; (Li *et al.*, 2019), and sweet potato (*Ipomoea batatas*; (Noh *et al.*, 2010; Dong *et*
480 *al.*, 2019), the rhizomes of *Nelumbo nucifera* (Cheng *et al.*, 2013b), and the corms of *Sagittaria*

481 *trifolia* (Cheng *et al.*, 2013a), indicating widespread parallel use of MADS-Box genes in the
482 formation of USOs. Similarly, we find that MADS-Box genes overall, and one DEG in
483 particular, are over-expressed in *Bomarea* tuberous roots. MADS-Box genes are implicated
484 widely as important transcription factors regulating plant development (Buylla *et al.*, 2000). It is
485 thus unsurprising that MADS-Box genes are regularly implicated in USO formation. It remains
486 unclear if the MADS-Box genes identified in the aforementioned studies represent independent
487 neofunctionalizations of MADS-Box genes from other aspects of plant development, or if they
488 form a clade of USO-specific copies.
489
490 ***Starch biosynthesis genes*** are very commonly identified in the formation of USOs, including in
491 cassava (Sojikul *et al.*, 2010; 2015), *Nelumbo nucifera* (Cheng *et al.*, 2013b; Yang *et al.*, 2015),
492 wild and domesticated sweet potatoes (Eserman *et al.*, 2018; Li *et al.*, 2019; Dong *et al.*, 2019),
493 and potato (Xu *et al.*, 1998). Since starch is so ubiquitous in USOs, this is unsurprising. We also
494 find starch isoforms overall to be over-expressed in *Bomarea* tuberous roots, and three genes in
495 particular are significantly over-expressed (see Table 3).
496
497 The other molecular processes we tested failed to show group-level differences from the global
498 distribution of expression levels. However, the presence of DEGs in some of these groups
499 indicates that the phytohormones in particular may play a role in tuberous root formation. One
500 gibberellin response isoform that is significantly under-expressed in tuberous roots, which aligns
501 with previous research suggesting that decreased gibberellin concentrations in roots can lead to
502 root enlargement (Tanimoto, 2012) and tuber formation (Xu *et al.*, 2016; Li *et al.*, 2019; Dong *et*
503 *al.*, 2019). The lack of significant auxin-related isoforms as differentially expressed is surprising,
504 as auxin has been implicated in USO formation in several previous studies (Noh *et al.*, 2010;
505 Cheng *et al.*, 2013b; Sojikul *et al.*, 2015; Yang *et al.*, 2015; Xu *et al.*, 2016; Hannapel *et al.*,
506 2017; Li *et al.*, 2019; Dong *et al.*, 2019; Kolachevskaya *et al.*, 2019), but it is possible this
507 function is not utilized in the tuberous root formation for monocot taxa.
508
509 **PEBP and FT-Like Gene Evolution in Geophytic Taxa** Gene tree analysis of PEBPs indicates
510 that FT and TFL1 genes have been independently co-opted several times in USO formation in
511 diverse angiosperms (including monocots and eudicots) and in diverse USO morphologies
512 (including tuberous roots, bulbs, and stem tubers). Furthermore, the presence of TFL1 and FT
513 homologs in gymnosperms and other non-flowering plants (Liu *et al.*, 2016) suggests that the
514 origin of these genes predates the evolution of flower as a reproductive structure. Instead, it
515 seems likely that these genes originally evolved as environmental signaling genes with wider
516 involvement in triggering the seasonality of various aspects of plant development. Subsequent
517 specialization of these genes in the timing of shifts to reproductive development as well as USO
518 development occurred. Given our results, in seems likely that USO-specialized FT and TFL1
519 genes arose at least four times independently, indicating broadly parallel molecular evolution
520 underlies the convergent morphological evolution of USOs. Additional USO-specific PEBP
521 genes would shed more light on this pattern, but the dearth of studies on the molecular basis of
522 USO development impedes such analysis. With increased sampling, follow-up studies could
523 identify unique patterns of convergent molecular evolution on the USO-specific FT genes. Do
524 these copies share independently derived subsequences or motifs that could reflect or cause
525 shared function?
526
527 In *Bomarea multiflora*, TFL1 involvement in USO formation is surprising, as neotropical cloud
528 forests are generally considered relatively aseasonal environments. However, *B. multiflora*

529 shows two annual peaks in flowering corresponding to two peaks in annual rainfall (Ortiz &
530 Idárraga Piedrahita, 2009), suggesting that important aspects of the plant's phenology are tied to
531 seasonal fluctuations. While no previous studies have looked at the timing of tuber production in
532 *Bomarea*, it is possible that tuber development is also tied to seasonal cues.
533
534 **Conclusion** We provide the first evidence of the molecular mechanisms of tuberous root
535 formation in a monocotyledonous taxon, filling a key gap in understanding the commonalities of
536 storage organ formation across taxa. We demonstrate that many molecular processes are shared
537 across geophytic taxa, suggesting that deep parallel evolution at the molecular level underlies the
538 convergent evolution of an adaptive trait. In particular, we demonstrate that PEBP genes
539 implicated in underground storage organ formation have been recruited multiple times across the
540 gene tree, demonstrating that repeated morphological convergence is matched by repeated
541 molecular convergence. These findings suggest further avenues for research on the molecular
542 mechanisms of how plants retreat underground and evolve strategies enabling adaptation to
543 environmental stresses. More molecular studies on diverse, non-model taxa and more thorough
544 sampling of underground morphological diversity will enhance our understanding of the full
545 extent of these convergences and add to our general understanding of the molecular basis for
546 adaptive, convergent traits.
547
548 **Acknowledgements**
549 We thank Lydia Smith (Evolutionary Genetics Lab, UC Berkeley) for training and sharing her
550 expertise on RNA-Seq, NSF GRFP, SSB, ASPT, Pacific Bulb Society, UC Berkeley's
551 Integrative Biology Department, and the Tinker Foundation for support to CMT, and UC
552 Berkeley CNR and the University and Jepson Herbarium for supporting sequencing costs.
553
554 **Author Contribution**
555 FA provided materials for the experiment, CMT and CDS designed the experiment, CMT and
556 JMG executed the experiment, CMT, JMG, and CJR analyzed the data, and CMT, JMG, CJR
557 and CDS contributed to writing the manuscript.
558
559 **References**
560
561 **Al-Tardeh S, Sawidis T, Diannelidis B-E, Stylianos Delivopoulos**. **2008**. Water content and
562 reserve allocation patterns within the bulb of the perennial geophyte red squill (Liliaceae) in
563 relation to the Mediterranean climate. *Botany* **86**: 291–299.

564 **Benitez-Alfonso Y, Faulkner C, Pendle A, Miyashima S, Helariutta Y, Maule A**. **2013**.
565 Symplastic Intercellular Connectivity Regulates Lateral Root Patterning. *Developmental Cell* **26**:
566 136–147.

567 **Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ,**
568 **Michoud K, O'Donovan C, Phan I, *et al.* 2003**. The SWISS-PROT protein knowledgebase and
569 its supplement TrEMBL in 2003. *Nucleic acids research* **31**: 365–370.

570 **Braybrook SA, Peaucelle A**. **2013**. Mechano-chemical aspects of organ formation in
571 Arabidopsis thaliana: the relationship between auxin and pectin. *PLOS ONE* **8**: e57813.

572 **Buylla ERA, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Silva FV, Yanofsky MF**.
573 **2000**. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells,
574 roots and trichomes. *The Plant Journal* **24**: 457–466.

575 **Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T**. **2009**. trimAl: a tool for automated
576 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

577 **Chabannes M, Barakate A, Lapierre C, Marita JM, Ralph J, Pean M, Danoun S, Halpin C,**
578 **Pettenati JG, Boudet AM**. **2001**. Strong decrease in lignin content without significant alteration
579 of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase
580 (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. *The Plant Journal* **28**:
581 257–270.

582 **Chaweewan Y, Taylor N**. **2015**. Anatomical Assessment of Root Formation and Tuberization in
583 Cassava (Manihot esculenta Crantz). *Tropical Plant Biology* **8**: 1–8.

584 **Cheng L, Li S, Xu X, Hussain J, Yin J, Zhang Y, Li L, Chen X**. **2013a**. Identification of
585 differentially expressed genes relevant to corm formation in *Sagittaria trifolia*. *PLOS ONE* **8**:
586 e54573.

587 **Cheng L, Li S, Yin J, Li L, Chen X**. **2013b**. Genome-wide analysis of differentially expressed
588 genes relevant to rhizome formation in lotus root (*Nelumbo nucifera* Gaertn). *PLOS ONE* **8**:
589 e67116.

590 **Cuéllar-Martínez M, Sosa V**. **2016**. Diversity patterns of monocotiledonous geophytes in
591 Mexico. *Botanical Sciences* **94**: 699–13.

592 **Dafni A, Cohen D, Noy-Mier I**. **1981a**. Life-cycle variation in geophytes. *Annals of the*
593 *Missouri Botanical Garden* **68**: 652–660.

594 **Dafni A, Shmida A, Avishai M**. **1981b**. Leafless autumnal-flowering geophytes in the
595 Mediterranean region - phytogeographical, ecological and evolutionary aspects. *Plant*
596 *Systematics and Evolution* **137**: 181–193.

597 **Darriba D, Weiß M, Stamatakis A**. **2016**. Prediction of missing sequences and branch lengths
598 in phylogenomic data. *Bioinformatics* **32**: 1331–1337.

599 **Dolan L, Davies J**. **2004**. Cell expansion in roots. *Current Opinion in Plant Biology* **7**: 33–39.

600 **Dong T, Zhu M, Yu J, Han R, Tang C, Xu T, Liu J, Li Z**. **2019**. RNA-Seq and iTRAQ reveal
601 multiple pathways involved in storage root formation and development in sweet potato (*Ipomoea*
602 *batatas* L.). *BMC Plant Biology* **19**: 136–16.

603 **Ekblom R, Galindo J**. **2011**. Applications of next generation sequencing in molecular ecology
604 of non-model organisms. *Heredity* **107**: 1–15.

605 **Eserman LA, Jarret RL, Leebens-Mack JH**. **2018**. Parallel evolution of storage roots in
606 morning glories (Convolvulaceae). *BMC Plant Biology* **18**: 1–11.

607 **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A,**
608 **Hetherington K, Holm L, Mistry J, *et al.* 2014**. Pfam: the protein families database. *Nucleic*
609 *acids research* **42**: D222–30.

610 **Gaudioso-Pedraza R, Beck M, Frances L, Kirk P, Ripodas C, Niebel A, Oldroyd GED,**
611 **Benitez-Alfonso Y, de Carvalho-Niebel F**. **2018**. Callose-Regulated Symplastic
612 Communication Coordinates Symbiotic Root Nodule Development. *Current Biology* **28**: 3562–
613 3577.e6.

614 **Gene Ontology Consortium**. **2004**. The Gene Ontology (GO) database and informatics
615 resource. *Nucleic acids research* **32**: 258D–261.

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

619 **Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB,**
620 **Eccles D, Li B, Lieber M, *et al.* 2013**. De novo transcript sequence reconstruction from RNA-
621 seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**: 1494–
622 1512.

623 **Hannapel DJ, Sharma P, Lin T, Banerjee AK**. **2017**. The Multiple Signals That Control Tuber
624 Formation. *Plant Physiology* **174**: 845–856.

625 **Hearn DJ**. **2009**. Descriptive anatomy and evolutionary patterns of anatomical diversification in
626 *Adenia* (Passifloraceae). *Aliso* **27**: 13–38.

627 **Hoffmann AJ, Liberona F, Hoffmann AE**. **1998**. *Distribution and ecology of geophytes in*
628 *Chile. Conservation threats to geophytes in mediterranean-type regions*. Berlin, Germany:
629 Springer-Verlag.

630 **Hofreiter A**. **2008**. A Revision of *Bomarea* Subgenus *Bomarea* s.str. Section *Multiflorae*
631 (Alstroemeriaceae). *Systematic Botany* **33**: 661–684.

632 **Howard CC, Folk RA, Beaulieu JM, Cellinese N**. **2019a**. The monocotyledonous
633 underground: global climatic and phylogenetic patterns of geophyte diversity. *American Journal*
634 *of Botany* **106**: 850–863.

635 **Howard CC, Landis JB, Beaulieu JM, Cellinese N**. **2019b**. Geophytism in monocots leads to
636 higher rates of diversification. *New Phytologist*: 1-10.

637 **Humphrey TV, Bonetta DT, Goring DR**. **2007**. Sentinels at the wall: cell wall receptors and
638 sensors. *New Phytologist* **176**: 7–21.

639 **Jones DT, Taylor WR, Thornton JM**. **1992**. The rapid generation of mutation data matrices
640 from protein sequences. *Computer applications in the biosciences: CABIOS* **8**: 275–282.

641 **Jung J, O'Donoghue EM, Dijkwel PP, Brummell DA**. Expression of multiple expansin genes
642 is associated with cell expansion in potato organs. *Plant Science* **179**: 77–85.

643 **Kolachevskaya OO, Lomin SN, Arkhipov DV, Romanov GA**. **2019**. Auxins in potato:
644 molecular aspects and emerging roles in tuber formation and stress resistance. *Plant cell reports*
645 **38**: 681–698.

646 **Kubitzki K**. **1998.** Vol III Monocotyledons: Lilianae (except Orchidaceae). *The families and*
647 *genera of vascular plants*. Berlin, Germany: Springer-Verlag.
648
649 **Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW**. **2007**. RNAmmer:
650 consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research* **35**: 3100–3108.
651
652 **Langmead B, Salzberg SL**. **2012**. Fast gapped-read alignment with Bowtie 2. *Nature methods*
653 **9**: 357–359.

654 **Larsson A**. **2014**. AliView: a fast and lightweight alignment viewer and editor for large datasets.
655 *Bioinformatics* **30**: 3276–3278.

656 **Lee R, Baldwin S, Kenel F, McCallum J, Macknight R**. **2013**. FLOWERING LOCUS T genes
657 control onion bulb formation and flowering. *Nature Communications* **4**: 2884.

658 **Leeggangers HACF, Nijveen H, Bigas JN, Hilhorst HWM, Immink RGH**. **2017**. Molecular
659 Regulation of Temperature-Dependent Floral Induction in Tulipa gesneriana. *Plant Physiology*
660 **173**: 1904–1919.

661 **Li B, Dewey CN**. **2011**. RSEM: accurate transcript quantification from RNA-Seq data with or
662 without a reference genome. *BMC bioinformatics* **12**: 323.

663 **Li J, Han Y, Liu L, Chen Y, Du Y, Zhang J, Sun H, Zhao Q**. **2015**. qRT9, a quantitative trait
664 locus controlling root thickness and root length in upland rice. *Journal of experimental botany*
665 **66**: 2723–2732.

666 **Li M, Yang S, Xu W, Pu Z, Feng J, Wang Z, Zhang C, Peng M, Du C, Lin F, *et al.* 2019**.
667 The wild sweetpotato (*Ipomoea trifida*) genome provides insights into storage root development.
668 *BMC Plant Biology* **19**: 119–17.

669 **Li X-F, Jia L-Y, Xu J, Deng X-J, Wang Y, Zhang W, Zhang X-P, Fang Q, Zhang D-M, Sun**
670 **Y, *et al.* 2013**. FT-Like NFT1 Gene May Play a Role in Flower Transition Induced by Heat
671 Accumulation in Narcissus tazetta var. chinensis. *Plant & cell physiology* **54**: 270–281.

672 **Liu Y-Y, Yang K-Z, Wei X-X, Wang X-Q**. **2016**. Revisiting the phosphatidylethanolamine-
673 binding protein (PEBP) gene family reveals cryptic FLOWERING LOCUS Tgene homologs in
674 gymnosperms and sheds new light on functional evolution. *New Phytologist* **212**: 730–744.

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

677 **Melis RJ, van Staden J**. **1985**. Tuberization in Cassava (*Manihot esculenta*): Cytokinin and
678 Abscisic Acid Activity in Tuberous Roots. *Journal of plant physiology* **118**: 357–366.

Miller MA, Pfeiffer W, Schwartz T. **2010**. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 gateway computing environments workshop (GCE). Ieee: 1-10.

Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, Shimamoto K, Prat S. **2011**. Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* **478**: 119–122.

Nguyen L-T, Schmidt HA, Haeseler von A, Minh BQ. **2014**. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Niklas KJ. **2005**. Modelling below-and above-ground biomass for non-woody and woody plants. *Annals of Botany* **95**: 315–321.

Noh SA, Lee H-S, Huh EJ, Huh GH, Paek K-H, Shin JS, Bae JM. **2010**. SRD1 is involved in the auxin-mediated initial thickening growth of storage root by enhancing proliferation of metaxylem and cambium cells in sweetpotato (Ipomoea batatas). *Journal of experimental botany* **61**: 1337–1349.

Oppenheim SJ, Baker RH, Simon S, DeSalle R. **2015**. We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. *Insect molecular biology* **24**: 139–154.

Ortiz RDC, Idárraga Piedrahita Á. 2009. Catálogo de las Plantas Vasculares del Departamento de Antioquia. http://www.tropicos.org/Project/Catalogo-de-Antioquia.

Ott JP, Klimešová J, Hartnett DC. **2019**. The ecology and significance of below-ground bud banks in plants. *Annals of Botany* **123**: 1099–1118.

Parsons RF. **2000**. Monocotyledonous geophytes: comparison of California with Victoria, Australia. *Australian Journal of Botany* **48**: 39–43.

Pate JS, Dixon KW. **1982**. *Tuberous, cormous and bulbous plants: biology of an adaptive strategy in Western Australia*. Nedlands, Australia: University of Western Australia Press.

Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M, *et al.* 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic acids research* **42**: D231–9.

Proches S, Cowling RM, Goldblatt P, Manning JC, Snijman DA. **2006**. An overview of the Cape geophytes. *Biological Journal of the Linnean Society* **87**: 27–43.

Raíces M, Gargantini PR, Chinchilla D, Crespi M, Téllez-Iñón MT, Ulloa RM. **2003**. Regulation of CDPK isoforms during tuber development. *Plant molecular biology* **52**: 1011–1024.

Raunkiaer C. **1934**. *The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer.*

Rundel PW. **1996**. Monocotylendous geophytes in the California flora. *Madroño* **43**: 355–368.

Schneider R, Hanak T, Persson S, Voigt CA. **2016**. Cellulose and callose synthesis and organization in focus, what's new? *Current Opinion in Plant Biology* **34**: 9–16.

Sojikul P, Kongsawadworakul P, Viboonjun U, Thaiprasit J, Intawong B, Narangajavana J, Svasti MRJ. **2010**. AFLP-based transcript profiling for cassava genome-wide expression analysis in the onset of storage root formation. *Physiologia Plantarum* **140**: 189–198.

Sojikul P, Saithong T, Kalapanulak S, Pisuttinusart N, Limsirichaikul S, Tanaka M, Utsumi Y, Sakurai T, Seki M, Narangajavana J. **2015**. Genome-wide analysis reveals phytohormone action during cassava storage root initiation. *Plant molecular biology* **88**: 531–543.

Sosa V, Loera I. **2017**. Influence of current climate, historical climate stability and topography on species richness and endemism in Mesoamerican geophyte plants. *PeerJ* **5**: e3932–20.

Tanabe M, Kanehisa M. **2012**. Using the KEGG database resource. *Current protocols in bioinformatics* **Chapter 1**: Unit1.12–1.12.43.

Tang JR, Lu YC, Gao ZJ, Song WL, Wei KH, Zhao Y, Tang QY, Li XJ, Chen JW, Zhang GH, *et al.* 2019. Comparative transcriptome analysis reveals a gene expression profile that contributes to rhizome swelling in *Panax japonicus* var. *major. Plant Biosystems*.

Tanimoto E. **2012**. Tall or short? Slender or thick? A plant strategy for regulating elongation growth of roots by low concentrations of gibberellin. *Annals of Botany* **110**: 373–381.

Thiers B. Index Herbariorum: a global directory of public herbaria and associated staff. http://sweetgum.nybg.org/science/ih/

Tsaftaris A, Pasentsis K, Kalivas A, Michailidou S, Madesis P, Argiriou A. **2012**. Isolation of a CENTRORADIALIS/TERMINAL FLOWER1 homolog in saffron (Crocus sativus L.): characterization and expression analysis. *Molecular Biology Reports* **39**: 7899–7910.

Vatén A, Dettmer J, Wu S, Stierhof Y-D, Miyashima S, Yadav SR, Roberts CJ, Campilho A, Bulone V, Lichtenberger R, *et al.* 2011. Callose Biosynthesis Regulates Symplastic Trafficking during Root Development. *Developmental Cell* **21**: 1144–1155.

Veselý P, Bureš P, Šmarda P, Pavlíček T. **2011**. Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? *Annals of Botany* **109**: 65–75.

Wang Y, Liu L, Song S, Li Y, Shen L, Yu H. **2017**. DOFT and DOFTIP1 affect reproductive development in the orchid *Dendrobium* Chao Praya Smile. *Journal of experimental botany* **68**: 5759–5772.

Wilson CA. **2006**. Patterns in evolution in characters that define Iris subgenera and sections. *Aliso* **22**: 425–433.

Xu L, Wang J, Lei M, Li L, Fu Y, Wang Z, Ao M, Li Z. **2016**. Transcriptome analysis of storage roots and fibrous roots of the traditional medicinal herb *Callerya speciosa* (Champ.) ScHot (X-Q Li, Ed.). *PLOS ONE* **11**: e0160338–20.

Xu X, van Lammeren AA, Vermeer E, Vreugdenhil D. **1998**. The role of gibberellin, abscisic acid, and sucrose in the regulation of potato tuber formation in vitro. *Plant Physiology* **117**: 575–584.

Yang M, Zhu L, Pan C, Xu L, Liu Y, Ke W, Yang P. **2015**. Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical lotus (*Nelumbo nucifera*). *Nature Publishing Group* **5**: 13059.

Yockteng R, Almeida AMR, Yee S, Andre T, Hill C, Specht CD. **2013**. A Method for Extracting High-Quality RNA from Diverse Plants for Next-Generation Sequencing and Gene Expression Analyses. *Applications in Plant Sciences* **1**: 1300070–6.

New Zealand Ministry for Public Industries. National Pests Plant Accord. https://www.mpi.govt.nz/protection-and-response/long-term-pest-management/national-pest-plant-accord/. Accessed 1 October 2019.

Figures and Tables

Figure 1: Sampling scheme of tissue types. *Bomarea multiflora* has modified underground stems (rhizomes) and modified roots (tuberous roots). Aerial shoots and roots develop from the underground rhizome. We extracted RNA from the aerial shoot meristem, rhizome meristem, fibrous root tip, and tuberous root tip of three *B. multiflora* individuals.

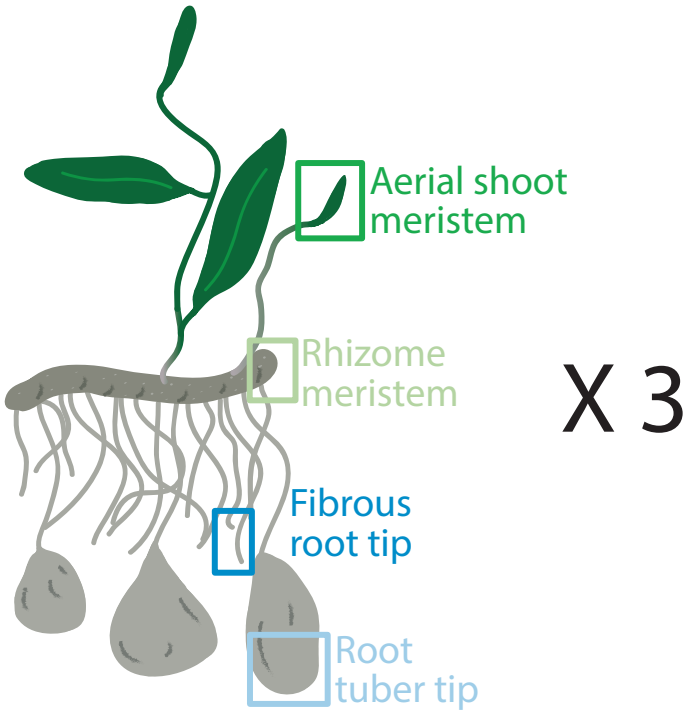

Figure 2: Principal component analysis of all samples. We performed a principal component analysis of variance stabling transformed (VST) transcript counts from all biological replicates of all tissue types. Point color indicates the tissue type of the sample and the letter corresponds to individual plants sampled. The first PC axis (PC1) explains 45% of the variance in the data, corresponding to variance due to tissue type. The second PC axis (PC2) explains 22% of the variance, corresponding to variance due to individual plant differences and other factors.

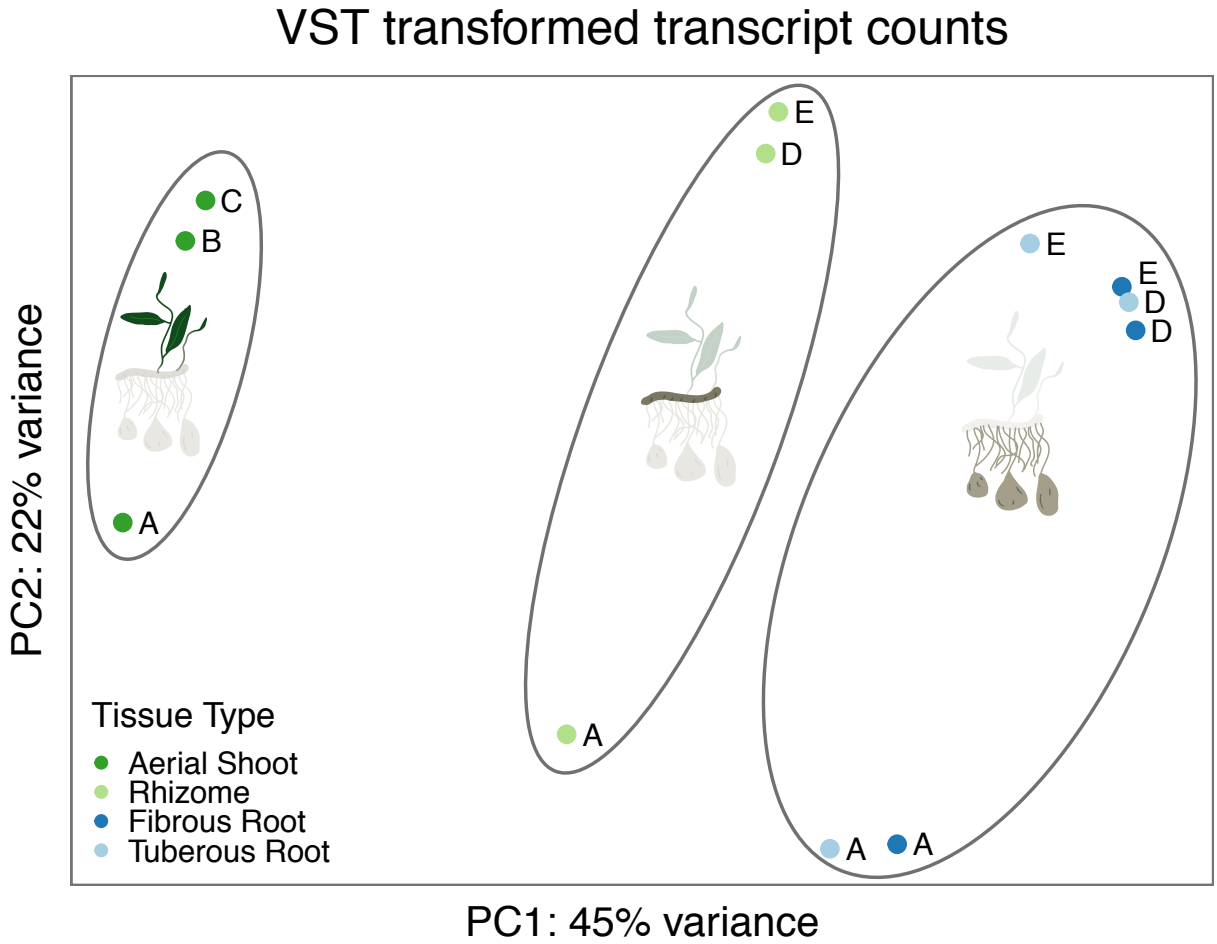

Figure 3: Differential expression of candidate gene groups. We identify isoforms corresponding to specific pathways and gene families (gene groups) and categorize the log2-fold change of expression between tuberous and fibrous roots of those groups. Boxplots correspond to the log2-fold change value for the gene groups and grey points correspond to individual isoforms. Groups with distributions that are significantly different from the entire dataset (All Genes) are labeled in red with asterisks on the X-axis. Isoforms are significantly differentially expressed (padj > 0.01) are labeled as red asterisks within the scatter plots.

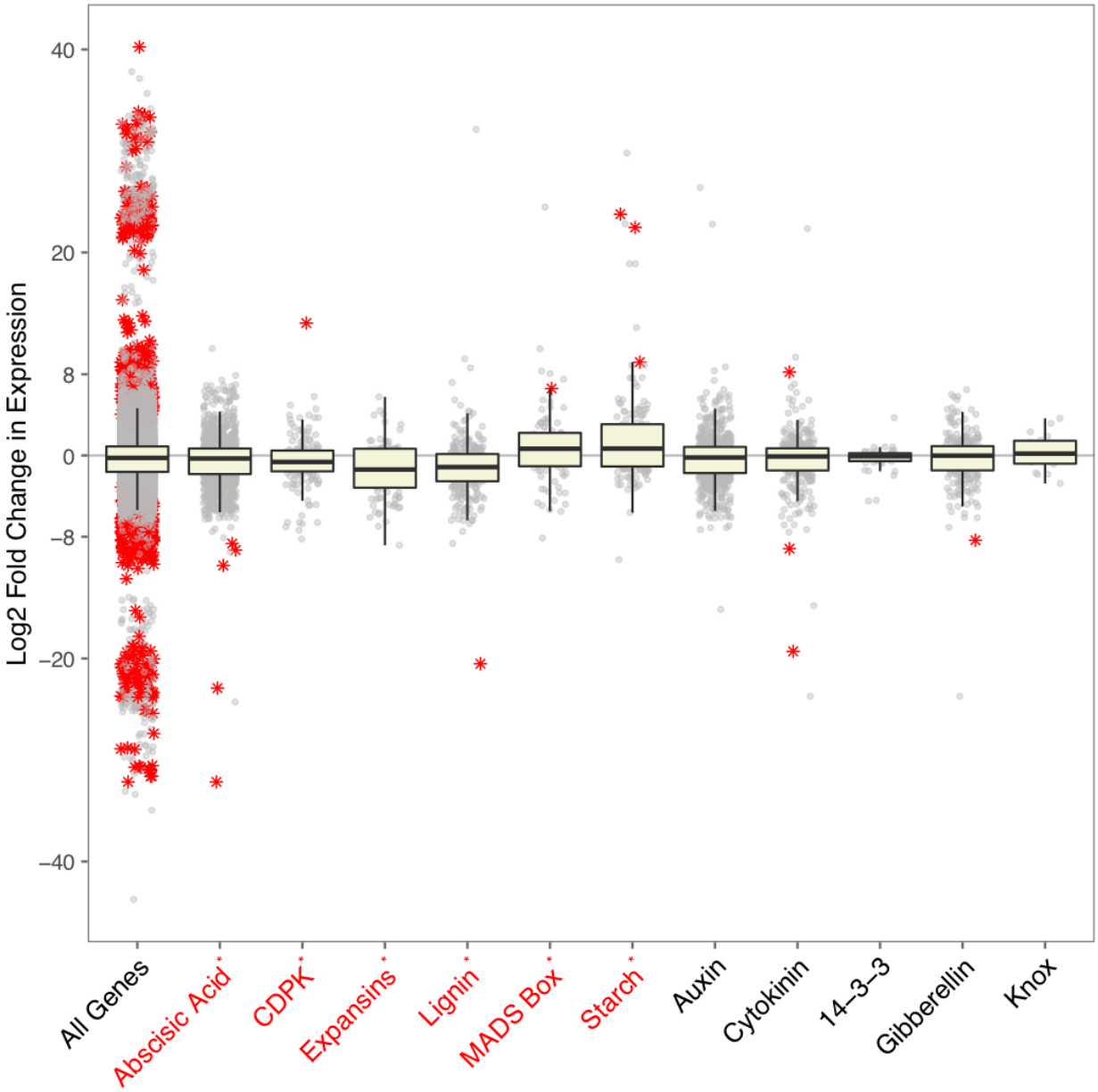

Figure 4: Evolution of PEBP genes: (a) Unrooted gene tree of 540 PEBP gene copies from across land plants. Stars indicate all included copies from *Bomarea multiflora*; the red star corresponds to the significantly differentially expressed isoform TRINITY_DN129076_c1_g1_i1). Red and arrows indicate PEBP copies that have been implicated in USO formation in other taxa (*Allium cepa* and *Solanum tuberosum*). The major clusters correspond to the FLOWERING LOCUS T (FT, in purple), TERMINAL FLOWER 1 (TFL1, in yellow), and the MOTHER OF FLOWERING LOCUS T AND TERMINAL FLOWER 1 (MFT, in red) gene groups and are labeled with high bootstrap support. (b) Detailed view of the cluster indicated by a dashed circle in (a), including monocot-specific copies of TERMINAL FLOWER 1 genes. Line thickness corresponds to bootstrap support.

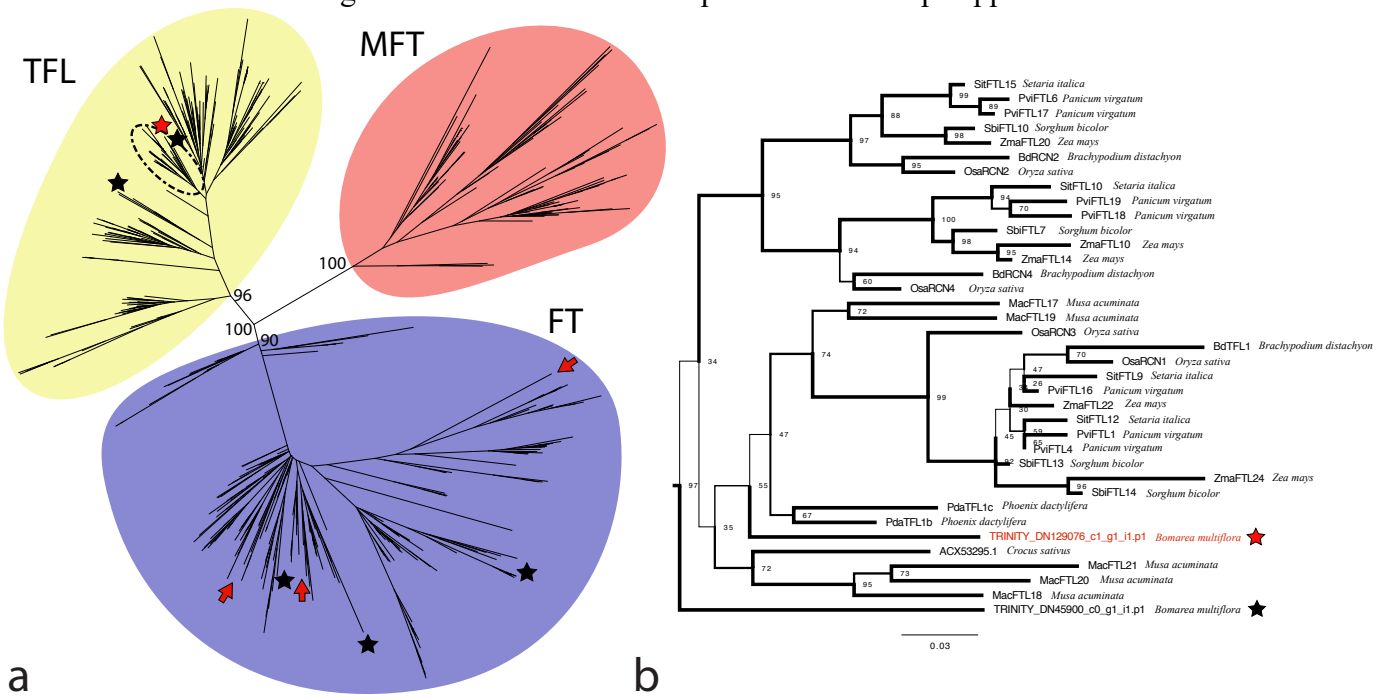

Table 1: Top ten GO annotations for all differentially expressed isoforms.

Cellular Components		Molecular Functions		Biological Processes	
Annotation	**Frequency**	**Annotation**	**Frequency**	**Annotation**	**Frequency**
nucleus	0.129	ATP binding	0.089	cell wall organization	0.020
integral component of membrane	0.098	metal ion binding	0.069	flower development	0.020
plasma membrane	0.078	DNA binding	0.044	multicellular organism development	0.016
cytoplasm	0.067	protein serine/threonine kinase activity	0.027	protein transport	0.016
cytosol	0.053	DNA-binding transcription factor activity	0.024	response to water deprivation	0.016
chloroplast	0.038	RNA binding	0.020	regulation of transcription, DNA-templated	0.013
mitochondrion	0.024	zinc ion binding	0.013	response to cold	0.013
plasmodesma	0.024	heme binding	0.011	carbohydrate metabolic process	0.011
apoplast	0.022	identical protein binding	0.011	cell division	0.011
cell wall	0.022	iron ion binding	0.011	defense response	0.011

Table 2: Top ten most differentially expressed isoforms (with padj < 0.01) and their corresponding annotations.

		Gene Ontology		
Transcript ID: Annotated Name (SPROT)	**Log2 Fold Change**	**Cellular Components**	**Molecular Functions**	**Biological Processes**
TRINITY_DN116220_c0_g1_i4: Zinc finger CCCH domain-containing protein 55	40.25		DNA binding, metal ion binding	
TRINITY_DN128685_c1_g3_i4: Callose synthase 3	33.85	1,3-beta-D-glucan synthase complex, integral component of membrane, plasma membrane	1,3-beta-D-glucan synthase activity	(1-¿3)-beta-D-glucan biosynthetic process, cell wall organization, regulation of cell shape
TRINITY_DN121298_c2_g2_i5: Heat shock 70 kDa protein 15	33.52	cell wall, cytosol, nucleus, plasma membrane, plasmodesma	ATP binding	
TRINITY_DN115892_c0_g1_i3: Pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH1	33.52	membrane, spliceosomal complex	ATP binding, ATP-dependent 3'-5' RNA helicase activity, RNA binding	mRNA processing, posttranscriptional gene silencing by RNA, RNA splicing
TRINITY_DN127064_c0_g3_i1: LRR receptor-like serine/threonine-protein kinase HSL2	33.32	integral component of membrane, plasma membrane	ATP binding, protein serine/threonine kinase activity	defense response to Gram-negative bacterium, lateral root morphogenesis, leaf abscission, regulation of gene expression
TRINITY_DN128839_c3_g1_i6: Palmitoyl-acyl carrier protein thioesterase, chloroplastic	33.01	chloroplast	thiolester hydrolase activity	fatty acid biosynthetic process
TRINITY_DN121430_c10_g2_i1 Sucrose nonfermenting 4-like protein	32.64	chloroplast, chloroplast starch grain, cytoplasm, nucleus	kinase binding, maltose binding, protein kinase activator activity, protein kinase regulator activity, protein serine/threonine kinase activity	carbohydrate metabolic process, cellular response to glucose starvation, mitochondrial fission, peroxisome fission, pollen hydration, protein autophosphorylation, regulation of protein kinase activity, regulation of reactive oxygen species metabolic process
TRINITY_DN124527_c1_g1_i5: Bifunctional TH2 protein, mitochondrial	32.64	cytosol, mitochondrion	thiaminase activity, thiamine phosphate phosphatase activity	thiamine biosynthetic process
TRINITY_DN120224_c4_g1_i1: Enhancer of polycomb homolog 2	32.25	Piccolo NuA4 histone acetyltransferase complex		DNA repair, histone acetylation, regulation of transcription by RNA polymerase II
TRINITY_DN122787_c0_g1_i1: Protein IQ-DOMAIN 32	-32.16	chloroplast envelope, cytosol, microtubule associated complex, nucleus, plasma membrane		response to abscisic acid

Table 3: Differentially expressed isoforms (with padj < 0.01) in specific gene groups and their corresponding annotations.

			Gene Ontology		
Process Group	**Transcript ID: Annotated Name (SPROT)**	**Log2 Fold Change**	**Cellular Components**	**Molecular Functions**	**Biological Processes**
Abscisic Acid Signaling	TRINITY_DN122359_c1_g1_i3: AMP deaminase	-8.67	cytosol, endoplasmic reticulum, integral component of mitochondrial outer membrane, intracellular membrane-bounded organelle, nucleus	AMP deaminase activity, ATP binding, metal ion binding, protein histidine kinase binding	embryo development ending in seed dormancy, IMP salvage, response to abscisic acid
Abscisic Acid Signaling	TRINITY_DN117636_c1_g2_i8: Dual specificity protein phosphatase PHS1	-9.32	cytoplasm	kinase activity, phosphoprotein phosphatase activity, protein tyrosine phosphatase activity, protein tyrosine/serine/threonine phosphatase activity	abscisic acid-activated signaling pathway, cortical microtubule organization, regulation of gene expression, regulation of stomatal movement, response to abscisic acid
Abscisic Acid Signaling	TRINITY_DN121543_c6_g2_i1: Probable RNA-binding protein ARP1	-10.84	nucleus, ribonucleoprotein complex	mRNA binding, RNA binding	mRNA processing, regulation of seed germination, response to abscisic acid, response to salt stress, response to water deprivation
Abscisic Acid Signaling	TRINITY_DN122705_c2_g1_i2: ACT domain-containing protein ACR8	-22.92		amino acid binding	response to abscisic acid
Abscisic Acid Signaling	TRINITY_DN122787_c0_g1_i1: Protein IQ-DOMAIN 32	-32.16	chloroplast envelope, cytosol, microtubule associated complex, nucleus, plasma membrane		response to abscisic acid
Calcium-Dependent Protein Kinases	TRINITY_DN124121_c3_g1_i11: Calcium-dependent protein kinase 2	13.05	cytoplasm, nucleus	ATP binding, calcium ion binding, calcium-dependent protein serine/threonine kinase activity, calmodulin binding, calmodulin-dependent protein kinase activity	abscisic acid-activated signaling pathway, intracellular signal transduction, peptidyl-serine phosphorylation, protein autophosphorylation
Cytokinin Signaling	TRINITY_DN128252_c1_g4_i1: Ferritin-3, chloroplastic	8.22	chloroplast, chloroplast envelope, chloroplast stroma, chloroplast thylakoid membrane, cytoplasm, membrane, mitochondrion, thylakoid	ferric iron binding, ferrous iron binding, ferroxidase activity, identical protein binding, iron ion binding	flower development, intracellular sequestering of iron ion, iron ion transport, leaf development, photosynthesis, response to bacterium, response to cold, response to cytokinin, response to hydrogen peroxide, response to iron ion, response to reactive oxygen species
Cytokinin Signaling	TRINITY_DN124688_c1_g1_i3: Temperature-induced lipocalin-1	-9.18	chloroplast envelope, chloroplast membrane, cytoplasm, cytoplasmic side of plasma membrane, endoplasmic reticulum, Golgi apparatus, mitochondrion, plasma membrane, plasmodesma, vacuolar membrane, vacuole	nutrient reservoir activity, transporter activity	cellular chloride ion homeostasis, cellular sodium ion homeostasis, heat acclimation, hyperosmotic salinity response, lipid metabolic process, positive regulation of response to oxidative stress, positive regulation of response to salt stress, response to cold, response to cytokinin, response to freezing, response to heat, response to high light intensity, response to light stimulus, response to paraquat, response to reactive oxygen species, response to water deprivation, seed maturation
Cytokinin Signaling	TRINITY_DN126720_c3_g1_i6: Two-component response regulator ARR2	-19.29	nucleus	DNA binding, DNA-binding transcription factor activity, phosphorelay response regulator activity	cellular response to cytokinin stimulus, cytokinin-activated signaling pathway, ethylene-activated signaling pathway, leaf senescence, regulation of root meristem growth, regulation of seed growth, regulation of stomatal movement, response to cytokinin, response to ethylene, root development
Lignin Biosynthesis	TRINITY_DN125451_c5_g1_i3: Cinnamoyl-CoA reductase-like SNL6	-20.52		3-beta-hydroxy-delta5-steroid dehydrogenase activity, oxidoreductase activity, oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	defense response to bacterium, lignin biosynthetic process, steroid biosynthetic process
MADS-Box Genes	TRINITY_DN127253_c5_g3_i1: MADS-box transcription factor 33	6.60	nucleus	DNA-binding transcription factor activity, protein dimerization activity, RNA polymerase II regulatory region sequence-specific DNA binding	positive regulation of transcription by RNA polymerase II
Response to Gibberellin	TRINITY_DN110988_c0_g1_i2: Gibberellin 3-beta-dioxygenase 1	-8.35	cytoplasm	dioxygenase activity, gibberellin 3-beta-dioxygenase activity, metal ion binding	gibberellic acid mediated signaling pathway, gibberellin biosynthetic process, response to gibberellin, response to red light, response to red or far red light
Starch Biosynthesis	TRINITY_DN119512_c6_g1_i4: Granule-bound starch synthase 1, chloroplastic/amyloplastic	23.77	amyloplast, chloroplast	alpha-1,4-glucan synthase activity, glycogen (starch) synthase activity, starch synthase activity	starch biosynthetic process
Starch Biosynthesis	TRINITY_DN118583_c0_g1_i7: 1,4-alpha-glucan-branching enzyme, chloroplastic/amyloplastic	22.46	amyloplast, chloroplast	1,4-alpha-glucan branching enzyme activity, 1,4-alpha-glucan branching enzyme activity (using a glucosylated glycogenin as primer for glycogen synthesis), cation binding, hydrolase activity, hydrolyzing O-glycosyl compounds	carbohydrate metabolic process, glycogen biosynthetic process, starch biosynthetic process, starch metabolic process
Starch Biosynthesis	TRINITY_DN114909_c0_g1_i2: Sedoheptulose-1,7-bisphosphatase, chloroplastic	9.19	apoplast, chloroplast envelope, chloroplast stroma, thylakoid	metal ion binding, sedoheptulose-bisphosphatase activity	defense response to bacterium, reductive pentose-phosphate cycle, starch biosynthetic process, sucrose biosynthetic process

Table 4: Specific candidate genes and results from blasting to assembled transcriptome. Asterisk indicates marginal significance.

Gene Names	Original Taxon	Isoforms Found	Number DE
OsbHLH120	*Oryza sativa*	22	1*
IDD5	*Ipomea batatas*	63	0
WOX4	*Ipomea batatas*	31	0
Sulfite reductase	*Manihot esculenta*	9	0
FT-like	*Solanum tuberosum, Allium cepa*	37	1