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

Carrie M. Tribble[1,*], Jesús Martínez-Gómez[1,2], Fernando Alzate-Guarin[3], Carl J. Rothfels[4], and Carl J. Rothfels[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 N° 53-108, Medellín, Colombia*
[4]*University Herbarium and Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720*
[*]*E-mail: ctribble09@gmail.com*

September 20, 2020

Abstract 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), unique innovations which allow these plants to retreat underground. Our ability to understand the developmental and evolutionary forces that shape these morphologies is limited by a lack of studies on certain USOs and plant clades. *Bomarea multiflora* (Alstroemeriaceae) is a monocot with tuberous roots, filling a key gap in our understanding of USO development. 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 between tuberous and non-tuberous roots and test the expression of a set of *a priori* candidate genes that have been implicated in underground storage in other taxa. We identify 271 genes that are differentially expressed in root tubers versus non-tuberous roots, including genes implicated in cell wall modification, defense response, and starch biosynthesis. We also identify a phosphatidylethanolamine-binding protein (PEBP), which has been implicated in tuberization signalling in other taxa and, through gene-tree analysis, place this copy in a phylogenytic context. These findings suggest that some similar molecular processes underlie the formation of underground storage structures across flowering plants despite the long evolutionary distances among taxa and non-homologous morphologies (*e.g.*, bulbs versus tubers). [Plant development, tuberous roots, comparative transcriptomics, geophytes]

1 Introduction

Scientific attention in botanical fields focuses almost exclusively on aboveground organs and biomass. However, a holistic understanding of land plant evolution, morphology, and ecology requires a comprehensive understanding of belowground structures: on average 50% of an individual plant's biomass lies beneath the ground (Niklas, 2005), and these portions of a plant are critical for resource acquisition, resource storage, and mediating the plant's interactions with its environment. 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 fall toward the extreme end of this belowground/aboveground allocation spectrum. In a remarkable example of convergent evolution of an innovative life history strategy, geophytes retreat underground by producing the buds of new growth on structures below the soil surface, while also storing nutrients to fuel this growth in highly modified, specialized underground storage organs (USOs) (Raunkiaer et al., 1934; Dafni et al., 1981b,a; Al-Tardeh et al., 2008; Veselý et al., 2011). Many geophytes also have the capacity to reproduce asexually through underground offshoots in addition to sexual reproduction. Geophytes are ecologically and economically important, morphologically diverse, and have evolved independently in all major groups of vascular plants ex-

cept gymnosperms (Howard et al., 2019, 2020). These plants and their associated underground structures are a compelling example of evolutionary convergence; diverse taxa produce a variety of structures, often from different tissues, that serve the analogous function of underground nutrient storage. However, our understanding of the molecular processes that drive this convergence, and the extent to which these processes are themselves parallel, remains limited, due in part to the lack of molecular studies in diverse geophyte lineages. This lack of study is particularly true for monocotyledonous geophytic taxa, which comprise the majority of ecologically and economically important geophyte diversity, but have not be subject to wide scientific attention beyond a select few crops.

Some of the world's most important crop plants have underground storage organs, 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*), various *Brassica* crops including kohlrabi and turnip (Hearn et al., 2018), and more. While several of these crops 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 detailed research has focused on a select few, which that do not represent the diversity of geophyte morphology, phylogeny, or ecology. Hearn (2006, among others) has proposed that "switches" in existing developmental programs can explain transitions between major growth forms; such a hypothesis requires broad sampling across the evolutionary breadth of taxa demonstrating the growth form. In particular, most genetic research on geophytes and their associated underground storage organs has been conducted in eudicots such as potato (Hannapel et al., 2017), sweet potato (Eserman et al., 2018; Li et al., 2019), cassava (Sojikul et al., 2010, 2015; Chaweewan and Taylor, 2015), *Brassica* (Hearn et al., 2018), and *Adenia* (Hearn, 2009). Fewer studies have focused on monocots (but see important studies in onion, such as in Lee et al., 2013), and these studies focus solely on bulbs; to date no study has characterized the molecular underpinnings of tuber formation in a monocotyledenous taxon.

This limited phylogentic breadth is particularly important in light of the findings of Hearn et al. (2018). This study provide compelling evidence that within closely related *Brassica* taxa, molecular mechanisms are shared between stem and hypocotyl/ root modifications. They also demonstrate that these mechanisms have been implicated in the development of other USOs, namely the eudicot lineages potato and sweet potato, and increased phylogenetic sampling could confirm and expand these findings or suggest that such results are clade-specific.

Underground storage organs originate from all major types of plant vegetative tissue: roots, stems, leaves, and hypocotyls. Bulbs (leaf tissue), corms (stem), rhizomes (stem), and tubers (stem or root) are some of the most common underground storage organ morphologies (Pate and 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, storage proteins, and water. The functional and physiological similarities of underground storage organs may drive or 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, as is suggested in Hearn et al. (2018).

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 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 and van Staden, 1985; Sojikul et al., 2015), in rhizome formation in *Panax japonicus* (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. FT-like genes, members of the phosphatidylethanolamine-binding protein (PEBP) family, have been implicated in USO formation in potato (Navarro et al., 2011; Hannapel et al., 2017), *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 involvements of FT orthologues in geophytic taxa.

The lateral expansion of roots into tuberous roots may be driven by cellular proliferation, by cellular expansion, or by a combination of these processes. Expansion in plant cells requires modification of the rigid plant cell wall to accommodate increases in cellular volume (Dolan and Davies, 2004; Humphrey et al., 2007), and genes such as expansins have been implicated in cellular expansion during tuberous root development in cassava and *Callerya speciosa* (Sojikul et al., 2015; Xu et al., 2016). Recent studies of the tuberous roots of sweet potato (*Ipomoea batatas*) and related species 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 and Taylor, 2015). However, geophytes are especially common in monocotyledonous plants (Howard et al., 2019, 2020), which lack a vascular cambium entirely. No previous study has addressed the molecular mechanisms of root tuber 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?

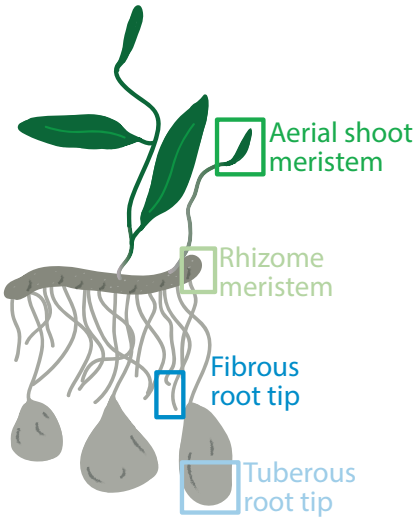

Figure 1: Sampling scheme of tissue types. *Bomarea multiflora* has modified underground stems (rhizomes) and modified roots (tuberous roots). We extracted RNA from the aerial shoot meristem, rhizome meristem, fibrous root tip, and tuberous root tip.

Bomarea multiflora (L. f.) Mirb. is a climbing monocotyledonous geophyte native to Venezuela, Colombia, and Ecuador, where 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 by comparing the genes expressed in different organs, developmental stages, or ecological conditions (Ekblom and 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.

In this study, we investigate the molecular mechanisms underlying the formation of tuberous roots in *Bomarea multiflora*, the first in a monocotyledonous taxon, using a comparative transcriptomics approach and quantify the extent to which these mechanisms are shared across the taxonomic and morphological breadth of geophytic taxa. Specifically, we ask by which developmental mechanisms does the plant modify fibrous roots into tuberous roots: (1) *how* expansion occurs, (2) *when* tuberization is triggered, and (3) *what* the tuberous roots store.

2 Materials and Methods

2.1 Greenhouse and Laboratory Procedures

We collected seeds from a single inflorescence of *Bomarea multiflora* in Antioquia, Colombia [vouchered as Tribble 194, deposited at UC (University Herbarium at UC Berkeley)] and germinated them in greenhouse conditions at the University of California, Berkeley designed to replicate native conditions for emphB. multiflora (70°F − 85°F and 50% humidity). Six months after germination, we harvested three sibling individuals as biological replicates. We dissected a single aerial shoot apical meristem (SAM), the rhizome apical meristem (RHI), root apical meristems (RAM) of several fibrous roots, and the growing tip of a tuberous root (TUB; Figure 1) from each of the three individuals for a total of 12 tissue samples.

We immediately froze samples in liquid nitrogen and maintained them at −80°C until extraction. We extracted total RNA from all samples using the Agilent Plant RNA Isolation Mini Kit (Agilent, Santa Clara, Ca), optimized for non-standard plant tissues, especially those that may be high in starch. Quality of total RNA was measured with Qubit (ThermoFisher, Waltham MA) and Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA); if needed, we used a Sera-Mag bead clean-up to further clean extracted RNA (Yockteng et al., 2013). Two SAM samples failed to extract at sufficient concentrations, so

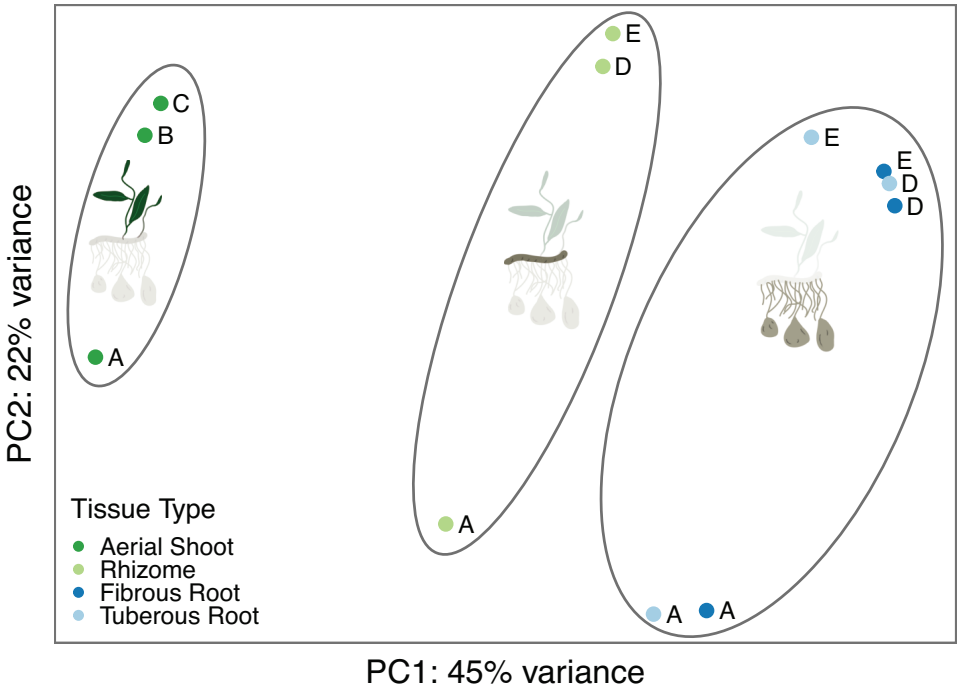

Figure 2: Principal component analysis of VST-transformed transcript counts from all samples. We performed a principal component analysis of variance-stabilizing-transformed (VST) transcript counts from all biological replicates of all tissue types. Points are colored by tissue type, and letters correspond to the individual plants sampled.

we harvested the SAMs of two additional individuals and extracted using the Yockteng et al. (2013) protocol. Samples with an RNA integrity (RIN) score >7 proceeded directly to library prep. We used the KAPA Stranded mRNA-Seq Kit (Kapa Biosystems, Waltham MA) protocol for library prep, applying half reactions with an input of at least 500 ng of RNA; however, all but two samples had 1 ug of RNA. RNA fragmentation time depended on RIN score (7 <RIN <8: 4 min; 8 <RIN <9: 5 min; 9 <RIN: 6 min). We split samples in half after the second post-ligation clean up (Step 10 in the Kapa protocol) in order to fine-tune the enrichment step. We amplified the first half of the samples with 12 PCR cycles; this proved too low and we increased to 15 cycles for the second half of the samples. We combined samples and assessed library quality with a Bioanalyzer 2100 using the DNA 1000 kit. We performed a bead clean-up on libraries showing significant adaptor peaks (Yockteng et al., 2013). We cleaned, multiplexed, and sequenced samples on a single lane of HiSeq4000 at the California Institute for Quantitative Biosciences (QB3) Vincent J. Coates Genomics Sequencing Lab.

2.2 Transcriptome Assembly, Annotation, and Quantification

We cleaned, processed, and assembled the raw reads using the Trinity RNA-Seq De novo Assembly pipeline (Grabherr et al., 2011) under the default settings unless otherwise stated in associated scripts. We ran all analyses using the Savio supercomputing resource from the Berkeley Research Computing program at UC Berkeley. We cleaned reads with Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), keeping unpaired reads and using a minimum fragment length of 36 base pairs. We used all reads from all samples to generate a consensus transcriptome, assembled de novo from the concatenated data using Trinity. We aligned each sample back to the assembled consensus transcriptome using Bowtie2 (Langmead and Salzberg, 2012), and quantified isoform abundances using RSEM (RNASeq by Expectation Maximization; Li and Dewey, 2011). We assessed transcriptome assembly quality by comparing the percentages of reads that mapped back to the assembled transcriptome using Bowtie2 (Langmead and Salzberg, 2012) to align reads, among other standard metrics. We annotated the consensus transcriptome with a standard Trinotate pipeline (https://trinotate.github.io/), comparing assembled isoforms to SWISS-PROT (Boeckmann et al., 2003), RNAmmer (Lagesen et al., 2007), Pfam (Finn et al., 2014), eggNOG (Powell et al., 2014), KEGG (Tanabe and Kanehisa, 2012), and Gene Ontology (Gene Ontology Consortium, 2004) databases. We tested the concordance of biological replicates by looking for significant differences between the total number of fragments per repli-

cate, by comparing the transcript quantities of all replicates to each other, and by checking the correlations between replicates. Isoforms with fewer than 10 total counts were discarded prior to subsequent analyses. We transformed transcript counts using the variance-stabilized transformation (VST) and compared all 12 samples using a principal components analysis.

2.3 Differential Expression

We identified differentially expressed isoforms (hereafter referred to as DEGs) between fibrous (FR) and tuberous (TR) roots with the R (R Core Team, 2013) DESeq2 package (Love et al., 2014). We used a p-adjusted cut-off (padj, using a Benjamini-Hochberg correction for false discovery rate) of 0.01 and a log2-fold change cut-off of 2 to determine statistically significant and sufficiently differentially expressed isoforms for downstream analyses. To test if the distribution of functional annotations for the DEGs is statistically different from the overall pool of annotated isoforms, we performed Fisher's exact tests in R (R Core Team, 2013) to compare the distributions of number of isoforms associated with (1) biological process, (2) molecular function, and (3) cellular component GO annotations. If the overall distribution of annotations differed, we identified the specific GO terms that are enriched in the DEG dataset relative to the pool of all annotated isoforms. To identify enriched GO terms, we consider the number of isoforms associated with a particular GO category to be drawn from a binomial distribution:

$$g_i \sim \text{Binomial}(n, p_i),$$

where g_i is the number of non-differentially expressed isoforms in GO category i, n is the total number of non-differentially expressed isoforms, and p_i is the probability that a given isoform is in GO category i. Thus, the maximum likelihood estimator of p_i is given by:

$$p_i = g_i/n$$

Our null hypothesis is that the probability that a given transcript is in GO category i is not greater in the pool of differentially expressed isoforms than the pool of non-differentially expressed isoforms. Under our null, the expected number of differentially expressed isoforms associated with GO category i (k_i) is also defined by a binomial distribution:

$$k_i \sim \text{Binomial}(m, p_i),$$

where m is the total number of differentially expressed isoforms and p_i is defined above. To test our null hypothesis, we compute the probability that the observed value is greater than k_i, conditioning on the probability of at least one transcript count per GO category (as GO categories with zero differentially expressed transcript counts are not represented in the differentially expressed dataset) and using a Bonferroni correction (Bonferroni, 1935) for multiple comparisons.

2.4 Parallel Processes Across Taxa

To complement the broad survey of expression patterns, we additionally identified specific candidate genes, gene families, and molecular processes that might be involved in the development of underground storage organs via a survey of the recent literature 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), we first queried the annotated transcriptome for isoforms with annotations matching the associated process or family (see Supplemental Table 2 for the specific search terms used), and then tested if that group was more or less differentially expressed than expected by chance. Specifically, for each focal group of n isoforms, we randomly sampled n isoforms 10,000 times from the pool of all isoforms. For each of the 10,000 samples, we (1) compared the distribution of absolute log2-fold change values of the sampled isoforms to the distribution of log2-fold change values of the full dataset and calculated the effect size of a non-parametric Mann–Whitney–Wilcoxon test, generating an expected distribution of effect sizes for a random group of genes of size n, and (2) counted the number of significant DEGs. Under our null model, we expect the effect sizes of the focal group and the randomly sampled groups to be the same. To determine significance of the overall distributions of log2-fold change values, we compared the effect size of the focal group to the distribution of simulated effect sizes and determined if the focal group effect size fell within the 95% credible interval of the simulated distribution. Focal effect sizes that were larger than 97.5% of the simulated effect sizes indicated that the focal group is generally more differentially expressed than the null expectation; similarly, focal effect sizes smaller than 97.5% of the simulated effect sizes indicate that the focal group is less differentially expressed than the null expectation. Note that as we performed this analysis on the absolute values of log2-fold changes, less differentially expressed refers to expression levels that are more similar between the focal groups and the null than expected by chance, rather than negative log2-fold change values. Our null model also predicts that the number of significant DEGs in each focal group falls within the 95% credible set of the distributions of number of significant DEGs. We compared the number of DEGs in the focal group to distribution of number of DEGs from our 10,000 random samples. We determined significance by identifying groups that contain more or fewer DEGs than the 95% credible set of the simulated distribution.

Table 1: 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_i: 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_i: 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_i: 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_: 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_i: 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

For all targeted candidate genes, we blasted the amino acid sequence of the candidate gene to the assembled consensus transcriptome (see Supplemental Table 3 for the blasted sequence specifications) using an e-value cut-off of 0.01 to assess if the identified homologs were differentially expressed.

2.5 PEBP Gene Family Evolution

We reconstructed the evolutionary history of the phosphatidylethanolamine-binding protein (PEBP) gene family 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. For copies from *B. multiflora*, we selected the longest isoform per gene to include in the alignment. We translated coding sequences from the *Bomarea multiflora* transcriptome to amino acid sequences using TransDecoder v5.5.0 (Haas et al., 2013), removing isoforms that failed to align properly. We aligned amino acids with MAFFT as implemented in AliView v1.18 (Larsson, 2014), using trimAl v1.4.rev15 (Capella-Gutiérrez et al., 2009) with the -gappyout option. We selected the best evolutionary model with ModelTest-NG v0.1.5 (Darriba et al., 2016) (JTT+G4 amino acid substitution model (Jones et al., 1992)) and reconstructed unrooted gene trees under maximum likelihood as implemented in IQtree (Nguyen et al., 2014), run on XSEDE using the CIPRES portal (Miller et al., 2010). Using the amino acid alignment, we compared amino acid residues from *Bomarea multiflora* orthologs to those that have been functionally characterized in the *Arabidopsis* FT orthologs (reviewed

in Ho and Weigel, 2014).

All scripts used in Sections 2.2, 2.3, 2.4, and 2.5 are available on GitHub at (`github.com/cmt2/bomTubers`)

3 Results

3.1 Transcriptome Assembly, Annotation, and Quantification

We recovered a total of 359 M paired-end 100 bp reads from the single HiSeq 4000 lane for the multiplexed 12 samples (NCBI BioProject #####). The assembled consensus transcriptome consists of 370,672 loci, corresponding to 224,661 Trinity "genes" (Drayd #####). The consensus transcriptome has a GC content of 45.14%, N50 of 1191 bp, median transcript length of 317 bp, and mean transcript length of 556.95 bp (see also Supplemental Materials Section 1.1). Of all reads, 85.54% aligned concordantly (in a way which matches Bowtie2's (Langmead and Salzberg, 2012) expectation for paired-end reads) to the assembled transcriptome, indicating sufficient assembly quality to proceed with downstream analyses (see Supplemental Materials section 1.2). 8.15% of genes had at least 10% sequence identity with the UniProt database (See Supplemental Material section 1.3, Table 1 and Figure 1; Consortium, 2019). Of those, 53.49% blasted with at least 80% sequence identity. The Trinotate pipeline annotated 1.70% of all isoforms. All four tissue types showed concordance between the three biological replicates with generally 1:1 ratios of transcript quantities to each other (see Supplemental Materials section 2: Figures 2 - 5) so we proceeded with analyses using 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 in all root samples is similar, especially in contrast to the very distinct expression profiles of the shoot samples. The second PC axis (22% of variance) generally explains variance among biological replicates, with Individual A particularly distinct from other individuals, perhaps due to microhabitat variation in the greenhouse, genotype differences, increased or decreased herbivory compared to other individuals.

3.2 Differential Expression

We recovered a total of 271 differentially expressed isoforms (DEGs) between fibrous and tuberous roots (FR vs. TR). Of these, 226 correspond to regions of the assembled consensus transcriptome with functional annotations.

Of the three types of Gene Ontology (GO) annotations, we recovered significant differences in the distributions of the number of isoforms associated with GO categories between the differentially expressed dataset and the pool of all isoforms for molecular functions ($p = 1x10^{-4}$) and cellular components ($p = 0.031$) but not for biological processes ($p = 0.921$). We recovered no significantly enriched individual GO annotations for biological processes. For cellular components, we found that *cytoplasm*, *integral component of membrane*, *nucleus*, and *plasma membrane* were enriched in the differentially expressed dataset relative to non-differentially expressed isoforms. For molecular functions, we found that *ATP binding* and *metal ion binding* were enriched.

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) are implicated in cellular and biological processes (Table 1). All but one 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 most differentially expressed isoform (with a 40.25 log2-fold change), 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.

3.3 Parallel Processes Across Taxa

Based on our literature survey, we identified 12 groups of genes that have been implicated in USO formation in other geophytes: 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, KNOX genes, and lipoxygenases (See Figure 3 and Supplemental Table 4). Of these 12 gene groups, six have significantly different

Table 2: Differentially expressed isoforms (with $padj < 0.01$) in specific gene groups and their corresponding annotations.

Process Group	Transcript ID: Annotated Name (SPROT)	Log2 Fold Change	Gene Ontology		
			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	-22.91		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	-19.35	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	23.73	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	22.47	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

expression values than the overall distribution of expression for all isoforms. In four cases (expansins, lignin, MADS-Box, and starch biosynthesis) the overall expression levels are significantly more differentially expressed than expected by chance. Expansins and lignin are generally under-expressed in TR compared to FR while MADS-Box and starch are generally over-expressed in TR compared to FR. In two cases (14-3-3 and lipoxygenases), the overall expression levels are less differentially expressed than expected by chance; their expression levels are generally conserved between TR and FR. In the remaining six cases (abscisic acid, auxin, CDPK, cytokinin, gibberellin, and knox) we recover no significant pattern in overall expression levels.

Table 3: Specific candidate genes and results from blasting to assembled transcriptome. Asterisk indicates marginal significance. Isoforms found corresponds to the number of copies identified in the assembled *Bomarea multiflora* transcriptome, and Number DE corresponds to the number of those copies which are significantly differentially expressed.

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

We identified fifteen individual DEGs in these gene groups (Table 2); interestingly, there seems to be no generalizable relationships between the significance and directionality of a particular group's distribution with the presence and directionality of expression for individual DEGs. For example, the expression distribution of gibberellin genes does not deviate significantly from the global pool of isoforms, but there is one significantly under-expressed DEG in the gibberellin group; similarly, the CDPK genes are under-expressed as a group, but the only CDPK-related significant DEG is over-expressed (Table 2).

In addition to the gene groups, we identified five specific candidate genes from the literature: OsbHLH120 (qRT9) has been implicated in root thickening in rice (Li et al., 2015); IDD5 and WOX4 are implicated in starch biosynthesis and TR formation, respectively, in Convolvulaceae (Eserman et al., 2018); sulfite reductase is associated with TR formation in *Manihot esculenta* (Sojikul et al., 2010); and FLOWERING LOCUS T (FT) has been implicated in signaling the timing of USO formation in a variety of taxa, notably *Allium cepa* and *Solanum tuberosum* (Navarro et al., 2011; Hannapel et al., 2017). We recover between nine and 63 putative homologs of these candidates (Table 3), but only one is significantly differentially expressed ($padj < 0.01$): a putative FT homolog (TRINITY_DN129076_c1_g1_i1), further investigated in the PEBP Gene Family Evolution section (below). One putative qRT9 homolog is marginally significant ($padj = 0.050$), and the E-value from the BLAST result to this isoform was 0.09. Given these marginal significance values, it is likely that the result is spurious and we do not follow up with further analysis.

3.4 PEBP Gene Family Evolution

We identified thirty-seven *Bomarea* isoforms as putative FLOWERING LOCUS T (FT) homologs. After filtering for the longest isoform per gene and filtering out sequences which failed to align properly, ultimately, we include five sequences in addition to the significantly differentially expressed copy. We recover three major clusters in our unrooted gene tree, all with strong bootstrap support (Figure 4a); these correspond to the FT cluster, TERMINAL FLOWER 1 (TFL1) cluster, and MOTHER OF FT AND TFL1 (MFT) cluster recovered in previous analyses (Liu et al., 2016). Three of the six *Bomarea multiflora* isoforms fall out in the FT cluster and three in the TFL1 cluster. The *Bomarea* DEG homolog is highly supported in the TFL1 cluster with sequences from other monocot taxa (Figure 4b). Members of this TFL1 clade have been functionally characterized in *Oryza sativa*, where four orthologs (OsRNC1, OsRNC2, OsRNC3 and OsRNC14) antagonize the rice ortholog of FT to regulate inflorescence development (Kaneko-Suzuki et al., 2018). mRNA isoforms of all OsRNC are expressed in the root and transported to the SAM. Interestingly, the Crocus sativus ortholog (CsatCEN/TFL1) belongs to the same clade and is also expressed underground (in corms; Tsaftaris et al., 2012). Amino acid analysis of TRINITY_DN129076_c1_g1_i1.p1 shows that it convergently shares a glycine residue at AthFT position G137, typically characteristic of FT rather than TFL genes (Supplemental Figure 8 Pin et al., 2010). FT homologs from *Allium cepa* and *Solanum tuberosum* that have been functionally implicated in stem tuber and bulb formation, respectively, are in the FT cluster but do not cluster together; rather all USO-implicated PEBP genes are more closely related to non-USO copies than to each other.

4 Discussion

4.1 How to Make a Tuberous Root

Our results suggest potential developmental mechanisms 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 primary thickening growth via cellular expansion *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 and van Staden, 1985).

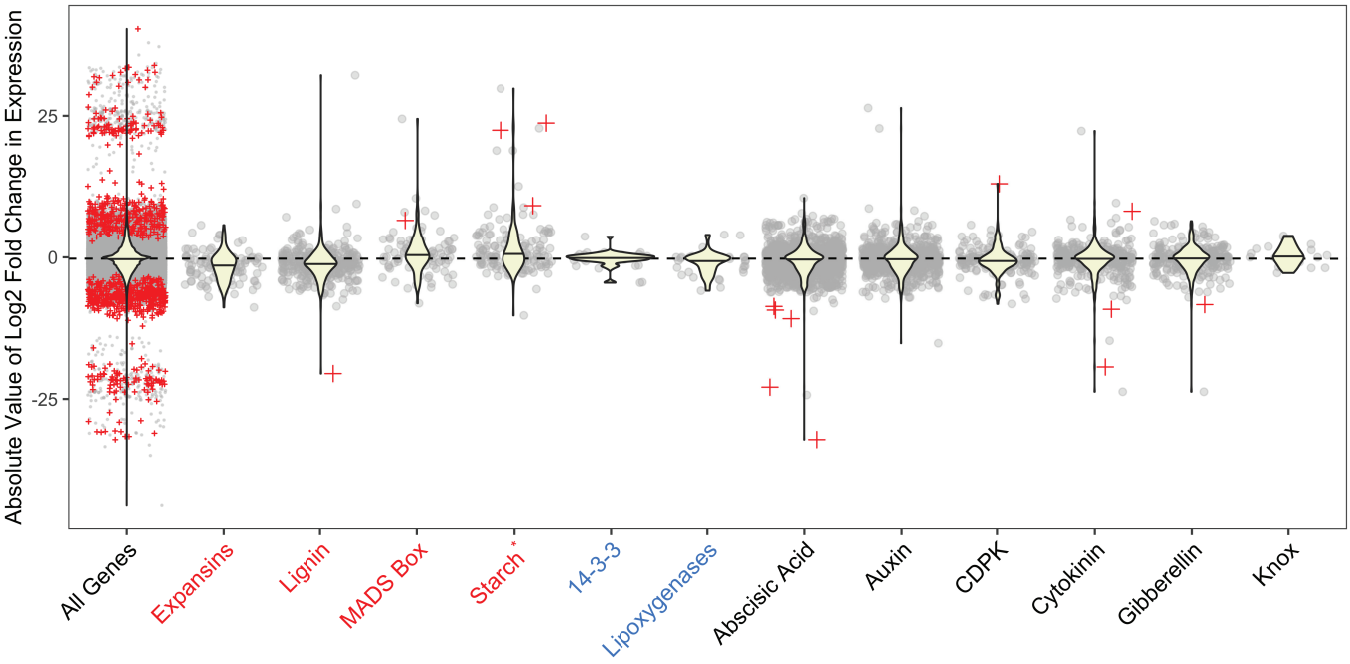

Figure 3: Differential expression of candidate gene groups. We identified isoforms corresponding to specific pathways and gene families (groups of genes) and categorize the log2-fold change of expression between tuberous and fibrous roots of those groups. Positive values correspond to overexpression in tuberous vs. fibrous roots. Box plots correspond to the log2-fold change value for the gene groups and grey points correspond to individual isoforms. Isoforms that are significantly differentially expressed (padj <0.01) are labeled as red asterisks within the scatter plots. Groups with absolute value log2-fold change distributions that are significantly larger from the entire dataset (shown in All Genes) are labeled in red on the X-axis, indicating that these groups are more differentially expressed (either generally up or down in tuberous roots) than expected:expansins, lignin, MADS box, and starch. Groups with less differential expression than expected by chance are labeled in blue: 14-3-3 and lipoxygenases. Groups with significantly more differentially expressed isoforms that expected by chance are labeled with an asterisk on the X-axis: starch.

Cell wall-related genes include pectinesterase TRINITY_DN122210_c6_g1_i1 (log2-fold change = 21.91; padj = 1.22E-8), which modifies pectin in cell walls leading to cell wall softening, as demonstrated for example in *Arabidopsis* (Braybrook and Peaucelle, 2013). Interestingly, expansins as a group were under- rather than over overexpressed in *B. multiflora*, though this is the mechanism by which cell expansion occurs in other taxa (see Expansins discussion below). Two of the enriched cellular component GO categories involve modifications to the cell membrane (*integral component of membrane* and *plasma membrane*), which suggests that modifications to the membrane may also be necessary in cellular expansion.

Flowering development genes such as TFL genes may be involved in mediating environmental signals and inducing tuber formation. 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 1). Callose is a much less common component of cell walls than 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 may play an integral role in triggering tuberous roots development in *B. multiflora* through symplastic signaling pathways and/or in modifying cell walls to accomodate expansion. Callose involvment in USO formation has not previously been reported and may be unique to *B. multiflora* or to monocots.

Finally, starch is thought to be the primary nutrient reserve in *Bomarea* tubers (Kubitzki and Huber, 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 1, TRINITY_DN121430_c10_g2_i1) and participates in carbohydrate biosynthesis, demonstrating that *B. multiflora* tubers

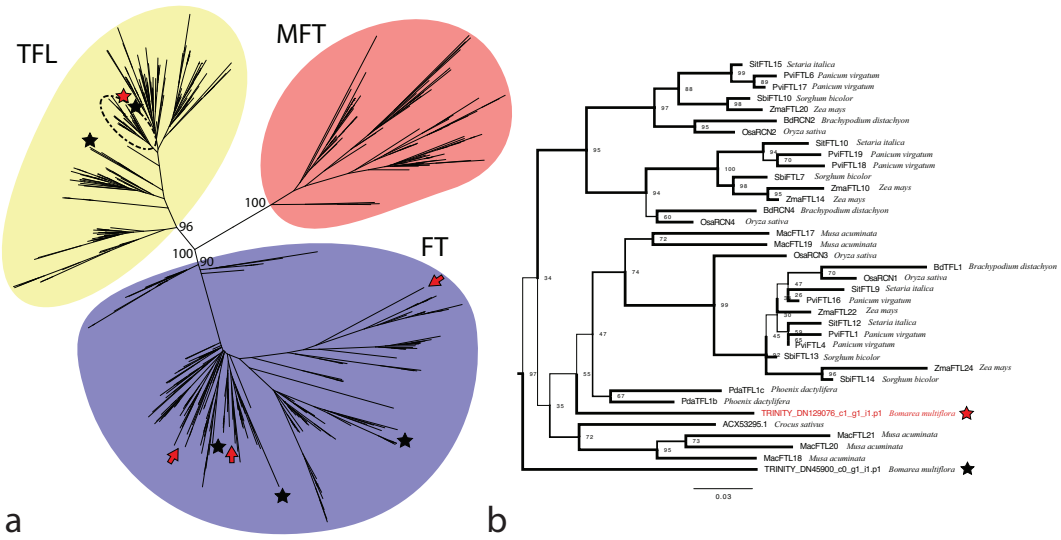

Figure 4: Evolution of PEBP genes: (a) Unrooted gene tree of 540 PEBP gene copies from across land plants. Stars indicate all included copies from *Bomarea multiflora*; the red star corresponds to the significantly differentially expressed isoform TRINITY_DN129076_c1_g1_i1). Red and arrows indicate PEBP 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.)

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 1) may be differentially expressed in tuberous roots to protect starch reserves against potential predation by belowground herbivores. LRR receptors have been implicated in triggering various downstream plant immune responses (Liang and Zhou, 2018).

We emphasize that more detailed work to follow up on these aspects of root tuber development should include morphological, anatomical, and developmental characterization of the tuberization process. The integration of these methods with genetic and molecular characterization will likely provide important functional links between the observed expression patterns we report here and the specific effects on plant form and function.

4.2 Similarities in Molecular Mechanisms of USO Formation

We identify four molecular processes, previously implicated in USO formation in other taxa, which are either over- or under-expressed in the tuberous roots of *Bomarea multiflora* (Figure 3). These processes suggest potential parallel function across deeply divergent evolutionary distances and in distinct plant structures.

Expansins are cell wall modifying genes known to loosen cell walls in organ formation (Dolan and Davies, 2004; Humphrey et al., 2007; Braybrook and Peaucelle, 2013). Their involvement in USO formation has been documented in the tuberous roots of cassava (Sojikul et al., 2015) and *Callerya speciosa* (Xu et al., 2016), the rhizomes of *Nelumbo nucifera* (Cheng et al., 2013b), the tuberous roots of various Convolvulaceae (Eserman et al., 2018), and the stem tubers of potato (Jung et al., 2010). As a group, in our data expansins are under-expressed in tuberous compared to fibrous roots, but no individual DEGs are statistically significant. These results suggest, surprisingly, that expansins likely do not play an important role in tuberous root formation in *Bomarea multiflora*. We do identify callose synthase 3 as one of the most differentially expressed genes (Table 1, so it is possible that cell wall modification during tuber development is primarily driven by callose rather expansin action in *B. multiflora*.

Lignin biosynthesis genes are under-expressed in several geophytic taxa with tuberous roots, including cassava (Sojikul et al., 2015), wild sweet potato (*Ipomoea trifida*; (Li et al., 2019), and *Callerya speciosa* (Xu et al., 2016). Similarly, we find that lignin biosynthesis overall is under-expressed in tuberous compared to fibrous roots, and one isoform in particular is significantly under-expressed: TRINITY_DN125451_c5_g1_i3: Cinnamoyl-CoA reductase-like SNL6 (Table 2). This gene has been found to significantly decrease lignin content without otherwise affecting development in tobacco (Chabannes et al., 2001). Decreased lignin in tuberous roots may further allow for cell expansion and permit lateral swelling of tuberous roots during development.

MADS-Box genes are implicated in USO formation in the tuberous roots of wild sweet potato (*Ipomoea trifida* (Li et al., 2019) and sweet potato (*Ipomoea batatas* (Noh et al., 2010; Dong et al., 2019), the rhizomes of *Nelumbo nucifera* (Cheng et al., 2013b), and the corms of *Sagittaria trifolia* (Cheng et al., 2013a), indicating widespread parallel use of MADS-Box genes in the formation of USOs. Similarly, we find that MADS-Box genes overall, and one DEG in particular, are over-expressed in *Bomarea* tuberous roots. MADS-Box genes are implicated widely as important transcription factors regulating plant development (Buylla et al., 2000). It is thus unsurprising that MADS-Box genes are regularly implicated in USO formation. It remains unclear if the MADS-Box genes identified in the aforementioned studies represent independent involvement of MADS-Box genes in USOs from other aspects of plant development, or if they form a clade of USO-specific copies. Follow-up phylogenetic analyses of these genes could prove informative in determining if MADS-Box genes involved in USO development form a clade (perhaps indicating a common origin for MADS-Box involvement in USOs or molecular convergence) or if they fall out independently (perhaps indicating multiple events of MADS-box involvement through distinct molecular changes).

Starch biosynthesis genes are very commonly identified in the formation of USOs, including in cassava (Sojikul et al., 2010, 2015), *Nelumbo nucifera* (Cheng et al., 2013b; Yang et al., 2015), wild and domesticated sweet potatoes (Eserman et al., 2018; Li et al., 2019; Dong et al., 2019), and potato (Xu et al., 1998). Since starch is so ubiquitous in USOs, this is unsurprising. We also find starch isoforms overall to be over-expressed in *Bomarea* tuberous roots, and three genes in particular are significantly over-expressed (see Table 2).

Unexpectedly, two of our candidate gene groups, 14-3-3 genes and lipoxygenases, show significantly less differential expression that expected by chance. This result suggests that some groups of genes previously identified as involved in USO formation may be conserved in their expression patterns between tuberous and fibrous roots in *Bomarea multiflora* unlike their differential expression patterns in other taxa. The evolutionary origin of 14-3-3 and lipoxygenase gene families likely predates the divergence of plant groups and the evolution of any kind of USO (both gene families are found across eukaryotes), implying that their involvement in USO development is perhaps due to neofunctionalization in specific plant clade(s). To our knowledge, the gene families' involvement in USO development has so far been only documented in eudicots, suggesting that this neofunctionalization may have occurred in the eudicot clade and would thus not be documented in monocots, a counter-example to parallelism in USO development across land plants. Gene tree analysis of these families and the homologs implicated in USO development would yield further insight into these potential patterns.

The other molecular processes we tested failed to show group-level differences from the global distribution of expression levels. However, the presence of DEGs in some of these groups indicates that the phytohormones in particular may play a role in tuberous root formation. One gibberellin response isoform is significantly under-expressed in tuberous roots, which aligns with previous research suggesting that decreased gibberellin concentrations in roots can lead to root enlargement (Tanimoto, 2012) and tuber formation (Xu et al., 2016; Li et al., 2019; Dong et al., 2019). The lack of significant auxin-related isoforms as differentially expressed is surprising, as auxin has been implicated in USO formation in several previous studies (Noh et al., 2010; Cheng et al., 2013b; Sojikul et al., 2015; Yang et al., 2015; Xu et al., 2016; Hannapel et al., 2017; Li et al., 2019; Dong et al., 2019; Kolachevskaya et al., 2019), but it is possible this paritcular role of auxin is not part of tuberous root development in monocot taxa, or that it is simply not identified in this study.

4.3 PEBP and FT-Like Gene Evolution in Geophytic Taxa

Gene tree analysis of PEBPs indicates that copies of these genes have independently evolved several times to signal USO formation in diverse angiosperms (including monocots and eudicots) and in diverse USO morphologies (including tuberous roots, bulbs, and stem tubers). Furthermore, the presence of TFL1 and FT homologs in gymnosperms and other non-flowering plants (Liu et al., 2016) suggests that the origin of these genes predates the evolution of the flower, their name notwithstanding. Instead, it seems likely that these genes originally evolved as environmental signaling genes with wider involvement in triggering the seasonality of various aspects of plant development. Subsequently, it is possible that gene duplication followed by neofunctionalization caused many copies in flowering plants to signal flower develepment and other copies to signal USO development. Given our results, it is possible that USO-specialized FT and TFL1 genes arose at least four times independently, suggesting that broadly parallel molecular evolution may underlie the convergent morphological evolution of USOs. However, without further work to quantify and characterize the differentially expressed PEBP gene in *Bomarea multiflora*, it is unclear if this gene is involved in tuberization signalling. Previously verified USO-specialized PEBP genes are clearly FT genes, but the *Bomarea multiflora* copy is a TFL1 ortholog. In the *Beta vulgaris* FT ortholog BvFT1, a mutation from glycine to glutamine along with two other mutations are sufficient to

turn FT into an antagonist of the functional FT orthology BvFT2 (Pin et al., 2010). None of the other TFL1 orthologs in the monocot clade share this mutation. While the *B. multiflora* ortholog does show signs of molecular convergence with FT genes at one residue known to induce TFL-like function in FT genes 8, it is unknown if function recovery can occur from TFL to FT and if only one mutation out of three is sufficient to recover function. Follow up studies to characterize the functions of various TFL1 and FT orthologs in *Bomarea multiflora* are necessary understand the role this DEG plays in tuber development.

Furthermore, the identification of additional USO-specific PEBP genes would shed more light on patterns of PEBP family involvement in USOs, but the dearth of studies on the molecular basis of USO development impedes such analyes. With increased sampling, follow-up studies could identify unique patterns of convergent molecular evolution on the USO-specific FT genes, such as tests for selection or further characterization of independently derived subsequences or motifs that could reflect or cause shared function.

4.4 Conclusions

We provide the first evidence of the molecular mechanisms of tuberous root formation in a monocotyledonous taxon, filling a key gap in understanding the commonalities of storage organ formation across taxa. We demonstrate that several groups of genes shared across geophytic taxa are implicated in tuberous root development in *Bomarea multiflora*, patterns which suggest that deep parallel evolution at the molecular level may underlie the convergent evolution of an adaptive trait. In particular, we demonstrate that PEBP genes previously implicated in underground storage organ formation appear multiple times across the gene tree, and we suggest that a TFL1 gene in *Bomarea multiflora* may also be involved in USO development. These patterns imply that repeated morphological convergence may be matched by independent evolutions of similar molecular mechanisms. However, we also demonstrate two counter examples to this pattern: groups of genes previously implicated in USO development whose patterns of expression are more similar in tuberous and fibrous roots than expected by chance (14-3-3 genes and lipoxygenases). These findings suggest further avenues for research on the molecular mechanisms of how plants retreat underground and evolve strategies enabling adaptation to environmental stresses. More molecular studies on diverse, non-model taxa and more thorough sampling of underground morphological diversity will enhance our understanding of the full extent of these convergences and add to our general understanding of the molecular basis for adaptive, convergent traits.

4.5 Acknowledgements

This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). We additionally thank Lydia Smith (Evolutionary Genetics Lab, UC Berkeley) for training and sharing her expertise on RNA-Seq, NSF GRFP, SSB, ASPT, Pacific Bulb Society, UC Berkeley's Integrative Biology Department, and the Tinker Foundation for support to CMT, and UC Berkeley CNR and the University and Jepson Herbarium for supporting sequencing costs. Michael R. May, David D. Ackerly, and Benjamin K. Blackman, and two anonymous reviewers provided feedback on the manuscript.

References

Al-Tardeh, S., Sawidis, T., Diannelidis, B.-E., and Delivopoulos, S. (2008). Water content and reserve allocation patterns within the bulb of the perennial geophyte red squill (Liliaceae) in relation to the Mediterranean climate. *Botany*, 86(3):291–299.

Benitez-Alfonso, Y., Faulkner, C., Pendle, A., Miyashima, S., Helariutta, Y., and Maule, A. (2013). Symplastic intercellular connectivity regulates lateral root patterning. *Developmental cell*, 26(2):136–147.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research*, 31(1):365–370.

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. *Studi in onore del professore salvatore ortu carboni*, pages 13–60.

Braybrook, S. A. and Peaucelle, A. (2013). Mechano-chemical aspects of organ formation in *Arabidopsis thaliana*: the relationship between auxin and pectin. *PloS one*, 8(3):e57813.

Buylla, E. R. A., Liljegren, S. J., Pelaz, S., Gold, S. E., Burgeff, C., Ditta, G. S., Silva, F. V., and Yanofsky, M. F. (2000). MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *The Plant Journal*, 24(4):457–466.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–1973.

Chabannes, M., Barakate, A., Lapierre, C., Marita, J. M., Ralph, J., Pean, M., Danoun, S., Halpin, C., Grima-Pettenati, J., and Boudet, A. M. (2001). Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. *The Plant Journal*, 28(3):257–270.

Chaweewan, Y. and Taylor, N. (2015). Anatomical assessment of root formation and tuberization in cassava (*Manihot esculenta* Crantz). *Tropical Plant Biology*, 8(1-2):1–8.

Cheng, L., Li, S., Xu, X., Hussain, J., Yin, J., Zhang, Y., Li, L., and Chen, X. (2013a). Identification of differentially expressed genes relevant to corm formation in *Sagittaria trifolia*. *PLoS One*, 8(1):e54573.

Cheng, L., Li, S., Yin, J., Li, L., and Chen, X. (2013b). Genome-wide analysis of differentially expressed genes relevant to rhizome formation in lotus root (*Nelumbo nucifera* Gaertn). *PloS one*, 8(6):e67116.

Consortium, U. (2019). Uniprot: a worldwide hub of protein knowledge. *Nucleic acids research*, 47(D1):D506–D515.

Dafni, A., Cohen, D., and Noy-Mier, I. (1981a). Life-cycle variation in geophytes. *Annals of the Missouri Botanical Garden*, pages 652–660.

Dafni, A., Shmida, A., and Avishai, M. (1981b). Leafless autumnal-flowering geophytes in the Mediterranean region—phytogeographical, ecological and evolutionary aspects. *Plant Systematics and Evolution*, 137(3):181–193.

Darriba, D., Weiß, M., and Stamatakis, A. (2016). Prediction of missing sequences and branch lengths in phylogenomic data. *Bioinformatics*, 32(9):1331–1337.

Dolan, L. and Davies, J. (2004). Cell expansion in roots. *Current opinion in plant biology*, 7(1):33–39.

Dong, T., Zhu, M., Yu, J., Han, R., Tang, C., Xu, T., Liu, J., and Li, Z. (2019). RNA-Seq and iTRAQ reveal multiple pathways involved in storage root formation and development in sweet potato (*Ipomoea batatas* L.). *BMC plant biology*, 19(1):1–16.

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

Eserman, L. A., Jarret, R. L., and Leebens-Mack, J. H. (2018). Parallel evolution of storage roots in morning glories (Convolvulaceae). *BMC plant biology*, 18(1):95.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. *Nucleic acids research*, 42(D1):D222–D230.

Gaudioso-Pedraza, R., Beck, M., Frances, L., Kirk, P., Ripodas, C., Niebel, A., Oldroyd, G. E., Benitez-Alfonso, Y., and de Carvalho-Niebel, F. (2018). Callose-regulated symplastic communication coordinates symbiotic root nodule development. *Current Biology*, 28(22):3562–3577.

Gene Ontology Consortium (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*, 32(suppl_1):D258–D261.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., 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(7):644–652.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8(8):1494–1512.

Hannapel, D. J., Sharma, P., Lin, T., and Banerjee, A. K. (2017). The multiple signals that control tuber formation. *Plant Physiology*, 174(2):845–856.

Hearn, D. J. (2006). Adenia (passifloraceae) and its adaptive radiation: phylogeny and growth form diversification. *Systematic Botany*, 31(4):805–821.

Hearn, D. J. (2009). Descriptive anatomy and evolutionary patterns of anatomical diversification in *Adenia* Passifloraceae). *Aliso: A journal of systematic and evolutionary botany*, 27(1):13–38.

Hearn, D. J., O'Brien, P., and Poulsen, T. M. (2018). Comparative transcriptomics reveals shared gene expression changes during independent evolutionary origins of stem and hypocotyl/root tubers in Brassica (Brassicaceae). *PLOS ONE*, 13(6):e0197166–25.

Ho, W. W. H. and Weigel, D. (2014). Structural features determining flower-promoting activity of arabidopsis flowering locus t. *The Plant Cell*, 26(2):552–564.

Hofreiter, A. (2008). A revision of *Bomarea* subgenus *Bomarea* s. str. section *Multiflorae* (alstroemeriaceae). *Systematic Botany*, 33(4):661–684.

Howard, C. C., Folk, R. A., Beaulieu, J. M., and Cellinese, N. (2019). The monocotyledonous underground: global climatic and phylogenetic patterns of geophyte diversity. *American journal of botany*, 106(6):850–863.

Howard, C. C., Landis, J. B., Beaulieu, J. M., and Cellinese, N. (2020). Geophytism in monocots leads to higher rates of diversification. *New Phytologist*, 225(2):1023–1032.

Humphrey, T. V., Bonetta, D. T., and Goring, D. R. (2007). Sentinels at the wall: cell wall receptors and sensors. *New Phytologist*, 176(1):7–21.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, 8(3):275–282.

Jung, J., O'Donoghue, E. M., Dijkwel, P. P., and Brummell, D. A. (2010). Expression of multiple expansin genes is associated with cell expansion in potato organs. *Plant Science*, 179(1-2):77–85.

Kaneko-Suzuki, M., Kurihara-Ishikawa, R., Okushita-Terakawa, C., Kojima, C., Nagano-Fujiwara, M., Ohki, I., Tsuji, H., Shimamoto, K., and Taoka, K.-I. (2018). Tfl1-like proteins in rice antagonize rice ft-like protein in inflorescence development by competition for complex formation with 14-3-3 and fd. *Plant and Cell Physiology*, 59(3):458–468.

Kolachevskaya, O. O., Lomin, S. N., Arkhipov, D. V., and Romanov, G. A. (2019). Auxins in potato: molecular aspects and emerging roles in tuber formation and stress resistance. *Plant cell reports*, 38(6):681–698.

Kubitzki, K. and Huber, H. (1998). *Flowering plants, monocotyledons: Lilianae (except Orchidaceae)*. Springer.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research*, 35(9):3100–3108.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22):3276–3278.

Lee, R., Baldwin, S., Kenel, F., McCallum, J., and Macknight, R. (2013). FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nature communications*, 4(1):1–9.

Leeggangers, H. A., Nijveen, H., Bigas, J. N., Hilhorst, H. W., and Immink, R. G. (2017). Molecular Regulation of Temperature-Dependent Floral Induction in *Tulipa gesneriana*. *Plant Physiology*, 173(3):1904–1919.

Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12(1):323.

Li, J., Han, Y., Liu, L., Chen, Y., Du, Y., Zhang, J., Sun, H., and Zhao, Q. (2015). qRT9, a quantitative trait locus controlling root thickness and root length in upland rice. *Journal of experimental botany*, 66(9):2723–2732.

Li, M., Yang, S., Xu, W., Pu, Z., Feng, J., Wang, Z., Zhang, C., Peng, M., Du, C., Lin, F., Wei, C., Qiao, S., Zou, H., Zhang, L., Li, Y., Yang, H., Liao, A., Song, W., Zhang, Z., Li, J., Wang, K., Zhang, Y., Lin, H., Zhang, J., and Tan, W. (2019). The wild sweetpotato (*Ipomoea trifida*) genome provides insights into storage root development. *BMC Plant Biology*, 19(1):119–17.

Li, X.-F., Jia, L.-Y., Xu, J., Deng, X.-J., Wang, Y., Zhang, W., Zhang, X.-P., Fang, Q., Zhang, D.-M., Sun, Y., and Xu, L. (2013). FT-Like NFT1 Gene May Play a Role in Flower Transition Induced by Heat Accumulation in Narcissus tazetta var. chinensis. *Plant & cell physiology*, 54(2):270–281.

Liang, X. and Zhou, J.-M. (2018). Receptor-like cytoplasmic kinases: central players in plant receptor kinase–mediated signaling. *Annual review of plant biology*, 69:267–299.

Liu, Y.-Y., Yang, K.-Z., Wei, X.-X., and Wang, X.-Q. (2016). Revisiting the phosphatidylethanolamine-binding protein (PEBP) gene family reveals cryptic FLOWERING LOCUS Tgene homologs in gymnosperms and sheds new light on functional evolution. *New Phytologist*, 212(3):730–744.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):31–21.

Melis, R. J. and van Staden, J. (1985). Tuberization in Cassava (*Manihot esculenta*): Cytokinin and Abscisic Acid Activity in Tuberous Roots. *Journal of plant physiology*, 118(4):357–366.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the cipres science gateway for inference of large phylogenetic trees. In *2010 gateway computing environments workshop (GCE)*, pages 1–8. Ieee.

Navarro, C., Abelenda, J. A., Cruz-Oró, E., Cuéllar, C. A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature*, 478(7367):119–122.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1):268–274.

Niklas, K. J. (2005). Modelling below- and above-ground biomass for non-woody and woody plants. *Annals of Botany*, 95(2):315–321.

Noh, S. A., Lee, H.-S., Huh, E. J., Huh, G. H., Paek, K.-H., Shin, J. S., and Bae, J. M. (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(5):1337–1349.

Oppenheim, S. J., Baker, R. H., Simon, S., and 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(2):139–154.

Pate, J. S. and Dixon, K. W. (1982). *Tuberous, cormous and bulbous plants*. International Scholarly Book Services Inc.[distributor].

Pin, P. A., Benlloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J. J., and Nilsson, O. (2010). An antagonistic pair of ft homologs mediates the control of flowering time in sugar beet. *Science*, 330(6009):1397–1400.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldón, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L. J., von Mering, C., and Bork, P. (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic acids research*, 42(Database issue):D231–9.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raunkiaer, C. et al. (1934). The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer. *The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer.*

Schneider, R., Hanak, T., Persson, S., and Voigt, C. A. (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., and Svasti, M. R. J. (2010). AFLP-based transcript profiling for cassava genome-wide expression analysis in the onset of storage root formation. *Physiologia Plantarum*, 140(2):189–198.

Sojikul, P., Saithong, T., Kalapanulak, S., Pisuttinusart, N., Limsirichaikul, S., Tanaka, M., Utsumi, Y., Sakurai, T., Seki, M., and Narangajavana, J. (2015). Genome-wide analysis reveals phytohormone action during cassava storage root initiation. *Plant molecular biology*, 88(6):531–543.

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

Tang, J. R., Lu, Y. C., Gao, Z. J., Song, W. L., Wei, K. H., Zhao, Y., Tang, Q. Y., Li, X. J., Chen, J. W., Zhang, G. H., Long, G. Q., Fan, W., and Yang, S. C. (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(2):373–381.

Tsaftaris, A., Pasentsis, K., Kalivas, A., Michailidou, S., Madesis, P., and Argiriou, A. (2012). Isolation of a CENTRORADIALIS/TERMINAL FLOWER1 homolog in saffron (*Crocus sativus* L.): characterization and expression analysis. *Molecular Biology Reports*, 39(8):7899–7910.

Vatén, A., Dettmer, J., Wu, S., Stierhof, Y.-D., Miyashima, S., Yadav, S. R., Roberts, C. J., Campilho, A., Bulone, V., Lichtenberger, R., Lehesranta, S., Mähönen, A. P., Kim, J.-Y., Jokitalo, E., Sauer, N., Scheres, B., Nakajima, K., Carlsbecker, A., Gallagher, K. L., and Helariutta, Y. (2011). Callose biosynthesis regulates symplastic trafficking during root development. *Developmental Cell*, 21(6):1144–1155.

Veselý, P., Bureš, P., Šmarda, P., and Pavlíček, T. (2011). Genome size and DNA base composition of geophytes: the mirror of phenology and ecology? *Annals of Botany*, 109(1):65–75.

Wang, Y., Liu, L., Song, S., Li, Y., Shen, L., and Yu, H. (2017). DOFT and DOFTIP1 affect reproductive development in the orchid *Dendrobium* Chao Praya Smile. *Journal of experimental botany*, 68(21-22):5759–5772.

Wilson, C. A. (2006). Patterns in evolution in characters that define *Iris* subgenera and sections. *Aliso*, 22(1):425–433.

Xu, L., Wang, J., Lei, M., Li, L., Fu, Y., Wang, Z., Ao, M., and Li, Z. (2016). Transcriptome analysis of storage roots and fibrous roots of the traditional medicinal herb *Callerya speciosa* (Champ.) ScHot. *PLOS ONE*, 11(8):e0160338–20.

Xu, X., van Lammeren AA, Vermeer, E., and Vreugdenhil, D. (1998). The role of gibberellin, abscisic acid, and sucrose in the regulation of potato tuber formation in vitro. *Plant Physiology*, 117(2):575–584.

Yang, M., Zhu, L., Pan, C., Xu, L., Liu, Y., Ke, W., and Yang, P. (2015). Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical lotus (*Nelumbo nucifera*). *Nature Publishing Group*, 5(1):13059.

Yockteng, R., Almeida, A. M. R., Yee, S., Andre, T., Hill, C., and Specht, C. D. (2013). A method for extracting high-quality rna from diverse plants for next-generation sequencing and gene expression analyses. *Applications in Plant Sciences*, 1(12):1300070–6.