Title
The Liliid and the Oddity: Macroevolution and development of underground storage organs in the order Liliales

Author
Tribble, Carrie M

Publication Date
2020

Peer reviewed|Thesis/dissertation

The Liliid and the Oddity: Macroevolution and development of underground storage organs in the order Liliales

by

Carrie M Tribble

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Carl J. Rothfels, Chair
Professor Chelsea D. Specht
Professor David D. Ackerly
Assistant Professor Benjamin K. Blackman

Summer 2020

The Liliid and the Oddity: Macroevolution and development of underground storage organs in the order Liliales

Abstract

The Liliid and the Oddity: Macroevolution and development of underground storage organs in the order Liliales

by

Carrie M Tribble

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Assistant Professor Carl J. Rothfels, Chair

Traditional botany focuses on the morphology, anatomy, and evolution of above-ground plant parts, but remarkable variation also exists underground. Underground storage organs (USOs), one example of understudied underground botany, include corms, bulbs, rhizomes, and stem- and root-derived tubers. These odd organs characterize geophytes, plants that produce perennating buds below ground and often store nutrients such as starch and water in USOs. Diverse underground morphology is particularly evident in the monocotyledenous order Liliales.

In this dissertation, I examine the evolution and development of USOs across the order Liliales, or the 'liliids'. In Chapter 1, I take a macroevolutionary perspective to ask if plants with different USOs are evolving towards different climatic niche adaptive peaks across the order. I find that the presence of root tubers, especially rotund root tubers, is associated with lower temperature seasonality. Furthermore, I develop and describe a new analysis pipeline in statistical comparative phylogenetics for testing adaptive hypotheses. In Chapter 2, I zoom in on a particular liliid geophyte, *Bomarea multiflora*, to identify genes underlying root tuber formation by comparing the transcriptomes of root tubers vs. fibrous roots. I compare the genes identified in this study with patterns from USOs produced by other taxa to characterize to what extent processes are shared across non-homologous USOs and across deep evolutionary divergences. I find that many processes are shared despite these differences, indicating that parallel molecular mechanisms may underlie USO development. In Chapter 3, I describe a new `R` package, `RevGadgets`, that can process and visualize the output of complex phylogenetic analyses from the `RevBayes` phylogenetic graphical modelling software. `RevGadgets` is designed to provide user-friendly modular workflows and thus increase accessibility to more complex phylogenetic models. I illustrate core `RevGadgets` functionality through six use cases and provide examples of code and resulting figures.

Together, these projects bring light to the outstanding diversity of below ground forms and begin the work of characterizing the evolution and development of this diversity. This work also illustrates the utility of establishing reproducible and user-friendly pipelines to increase the accessibility and versatility of complex statistical methods in comparative biology.

Contents

Contents **i**

List of Figures **iii**

List of Tables **iv**

1 Root morphology variation underlies differences in climatic niche optima in the geophyte-rich Liliales **1**
1.1 Introduction 1
1.2 Materials and Methods 5
1.3 Results 11
1.4 Discussion 15

2 Comparative transcriptomics of root tuber formation in a monocotyledonous geophyte **19**
2.1 Introduction 19
2.2 Materials and Methods 22
2.3 Results 25
2.4 Discussion 35

3 Data processing and plotting pipelines in R for Bayesian phylogenetic analyses with RevGadgets **40**
3.1 Introduction 40
3.2 Visualizing MCMC Output 42
3.3 Visualizing Phylogenies 46
3.4 Ancestral-State Reconstruction 49
3.5 Episodic Diversification Analysis 53
3.6 Posterior-Predictive Analysis 55
3.7 Summary 57

Bibliography **58**

A Phylogenetic systematics of *Bomarea* Mirb. **72**

B Chapter 1 Supplemental Code **74**

C Chapter 2 Supplemental Figures **75**

D Chapter 2 Supplemental Tables **80**

E Chapter 2 Supplemental Code **82**

F Chapter 3 Supplemental Code **83**

List of Figures

1.1 Illustration of underground morphologies in Liliales 2
1.2 Hierarchy of morphological states used in PARAMO analysis. 7
1.3 Methods diagram 10
1.4 Stochastic map results 12
1.5 Posterior distributions of state-specific θs 13
1.6 Test statistic (S) distributions 14

2.1 Sampling scheme of tissue types 23
2.2 Principal component analysis of all samples 26
2.3 Differential expression of candidate gene groups 30
2.4 Evolution of PEBP genes 35

3.1 Coda plots 43
3.2 Quantitative trace plot 45
3.3 QUalitative trace plot 45
3.4 Fossilized birth-death process plot 47
3.5 Branch rates plot 48
3.6 Standard ancestral state reconstruction plot of MAP states 51
3.7 Biogeographic range reconstruction plot 52
3.8 Episodic birth-death analysis plot 54
3.9 Posterior predictive analysis plots 56

C.1 Rhizome replicate analysis 76
C.2 Root replicate analysis 77
C.3 SAM replicate analysis 78
C.4 Tuber replicate analysis 79

List of Tables

1.1 List of gene regions and percent coverage used in phylogenetic reconstruction. 5
1.2 Constraints used for dating analysis. 6
1.3 `bayou` priors 8

2.1 Top ten GO annotations for all differentially expressed isoforms. 27
2.2 Top ten most differentially expressed isoforms (with $padj < 0.01$) and their corresponding annotations. 28
2.3 Gene groups from the literature 29
2.4 Differentially expressed isoforms (with $padj < 0.01$) in specific gene groups and their corresponding annotations. 31
2.5 Specific candidate genes and results from blasting to assembled transcriptome . 34

3.1 Summary of the primary `RevGadgets` functions 41

A.1 *Bomarea* species list 73

D.1 Candidate gene groups search terms 80
D.2 Genes used in blast analysis 81

Acknowledgments

This work would not have been possible without the support of many individuals and institutions.

I was funded by a National Science Foundation Graduate Research Fellowship Program grant. Transcriptome sequencing for Chapter 2 was funded by UC Berkeley College of Natural Resources and the University and Jepson Herbaria. Fieldwork and sequencing costs for the *Bomarea* phylogeny were provided by grants from the American Society of Plant Taxonomists, the Pacific Bulb Society, the Garden Club of America, the Society for Systematic Biology, The American Philosophical Society, the Torrey Botanical Society, the Tinker Foundation, and the Integrative Biology Department at UC Berkeley.

The following herbaria graciously provided access to collections for destructive sampling for the *Bomarea* phylogeny: The Missouri Botanical Garden, The New York Botanical Garden, The Field Museum, the United States National Herbarium (SI), the UC Herbarium, and the Herbario Nacional de México.

My work is a product of an incredible academic community, including formal and informal mentors, peers, and teachers. First, many thanks to my advisors and committee members: Carl Rothfels, Chelsea Specht, David Ackerly, Ben Blackman, Patrick O'Grady, and Bruce Baldwin. Carl and Chelsea in particular have fundamentally shaped who am I as a scholar and provided exceptional advising and guidance as I ricocheted between projects and directions. The members of the Specht and Rothfels labs made lab feel like home for the past five years, and I learned so much alongside them. Álvaro Idárraga and Mauricio Bonifacino taught me how to be a real botanist. Fernando Alzate showed me how truly magical *Bomarea* are. Will Freyman opened my eyes to the field of statistical phylogenetics and Mike May held me to high standards while helping me to refine my knowledge. I am incredibly grateful for their support and ongoing collaborations. Jesus Martinez-Gomez assisted with lab work and, along with Cody Howard, helped me to develop theories and ideas on geophytic evolution and development through years of GOOPhy conversations.

To the people who kept me grounded throughout it all, thank you. Book club (Ashley Smiley, Jeff Frederick, Ryan Murphy, Brian Egdorf, and Nicolas Alexandre) provided a space of unapologetic joy and gave me something to look forward to at the end of difficult weeks. Jenna Ekwealor and Joyce Chery are the life-long friends I know I can always count on. Writing group inspired and motivated me in the final stretch (Roxanne Marie Cruz de Hoyos, Aidee Guzman, Sarick Matze, and Adrian Lu). The Mango crew brought extra joy to my Saturdays (Giovanna Figueroa, Cat Taylor, Mahya Krouse, et al.). Keir Wefferling provided me with the soundtrack and 'ocotillo' to keep me going. Mary Wilson has been a brightness in my life during the darkness of 2020 - I would not have made it through finishing my PhD during a pandemic without you. Betsabe Castro Escobar, Mallory Ballinger, Ixchel Gonzalez Ramirez, and Mackenzie Kirchner-Smith baked the most amazing cakes, dropped off food during times of need, and encouraged parking lot puppy play-dates. My dog Kula reminded me to take breaks even when nothing seemed

to be going well. Elise Baker and Signe Chambers reminded me that I am a person beyond biology and made the Bay feel a little more like home. Finally, my parents Jane Schoonmaker and Gordon Tribble raised me to follow my curiosity about the natural world, and I would not have made it to where I am without their love and support.

Chapter 1

Root morphology variation underlies differences in climatic niche optima in the geophyte-rich Liliales

1.1 Introduction

The evolution of major innovations in life history strategies (how organisms gather and store energy and reproduce) is one of the primary themes of biodiversity research (Adler et al., 2014; Enquist et al., 1999). In one remarkable example of a life history innovation, certain plants can retreat underground by producing the buds of new growth on structures below the soil surface (Raunkiaer et al., 1934) while also storing nutrients to fuel this growth in highly modified, specialized underground storage organs (USOs).These plants are collectively referred to as geophytes. Geophytes have evolved independently many times across the plant tree of life, including in diverse and distantly related lineages within the ferns and flowering plants. Even within closely related lineages, geophytes show remarkable variation in the particular morphological modifications that allow them to survive underground. By differentially modifying leaves, stems, and roots, geophytes produce complex storage structures through distinct developmental and evolutionary means (reviewed in Tribble et al., in review; see Figure 1.1).

Previous work has suggested that the geophytic habit is correlated with more seasonal climatic conditions and higher-disturbance regimes (Cuéllar-Martínez and Sosa, 2016; Sosa et al., 2016; Sosa and Loera, 2017; Howard et al., 2019). Geophytes are particularly diverse in seasonally dry climates such as Mediterranean ecosystems, where they survive hot, dry summers underground and emerge during cool, dry winters to photosynthesize and reproduce, a pattern particularly prominent in the Cape region of South Africa, where almost 15% of native plant species are geophytic (Parsons and Hopper, 2003). Geophytes are also common in deciduous woodland habitats, where their USOs fuel quick spring regrowth to maximize photosynthetic opportunities before trees have

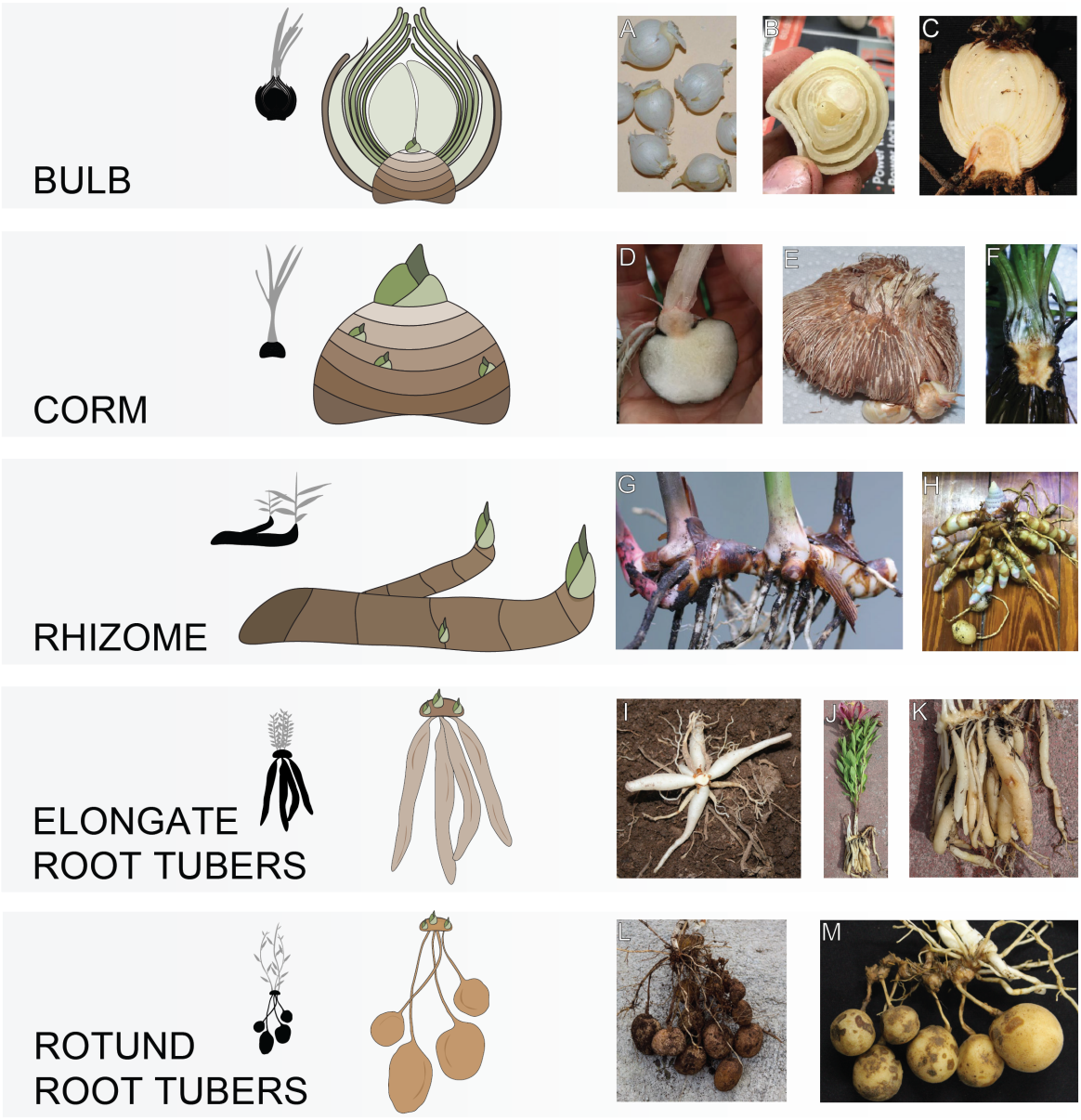

Figure 1.1: Diverse underground morphologies represented in Liliales: bulb, corm, rhizome, elongate root tubers, and rotund root tubers. Colored illustrations of USOs are presented within the whole plant context (grey and black drawings). Associated photos show exemplar variants of these traits: bulb: A) *Allium oreophyllym*, B) *Tulipa* cultivar, C) *Ledebouria* sp.; corm: D) *Amorphophallus* sp., E) *Chasmanthe* sp., F) *Isoetes storkii*; rhizome: G) *Hedychium* sp., H) *Cucurma* hybrid; tuberous roots: I) *Burchardia congesta*, J-K) *Alstroemeria* cultivar, L) *Bomarea edulis*, and M) *Bomarea obovata*. Modified from Tribble et al. (in review). Photo credits as follow: Robbin Moran (A, E, F, G), Cody C. Howard (B,C,D), Zach Siders (H), Barb Dobson and Ken Macintyre (I), and Carrie M. Tribble (J, K, L, and M).

regrown their canopies in spring (Whigham, 2004). These biodiversity and distribution patterns have led researchers to propose that geophytic USOs are adaptations to seasonal climates (Rees, 1989).

In a recent study of monocotyledonous geophytes, Howard et al. (2019) found that geophytism is correlated with areas of lower temperature and precipitation and higher temperature variation. Howard et al. (2019) also tested for different climate preferences among geophytes with different categories of USOs, and were unable to recover significant correlations with climate, with the exception of an association between rhizomatous geophytes and areas of increased temperature variation. The authors suggest that more detailed morphological data, as well as data that takes into account the developmental origin of USOs (leaf, stem, or root tissue), may be necessary to address within-geophytic variation in environmental preferences.

While this body of work provides compelling evidence for correlations between seasonal climates and the geophytic habit, previous phylogenetic-based research has relied on methods that fail to account for the complexities of morphological and adaptive evolutionary processes. While the goal of many phylogenetic comparative methods is to model the evolutionary relationships between (often multiple) traits and species, incorporating diverse data types into a cohesive analytical framework is often stymied by underlying differences in how different types of traits are expected to evolve across a tree. Specifically, including continuous and discrete traits in a single analysis is a longstanding statistical challenge in phylogenetics. Phylogenetic generalized linear models (Garland Jr et al., 1993), such as the phylogenetic ANOVA model used in Howard et al. (2019), are flexible for accommodating different data types (such as discrete morphology data and continuous climatic variables) by modifying the expected variance structure of continuous response variables using the phylogeny and an assumed model of continuous trait evolution. However, the continuous trait models commonly used to establish this expected variance structure are generally considered non-adaptive (e.g., Brownian motion; Felsenstein, 1985), whereas the studies that use such methods are often motivated by adaptive hypotheses. For example, if species have evolved to occupy different climatic niches, then the Brownian motion process, which represents a random-walk, often interpreted as neutral continuous trait evolution, will likely do a poor job of modeling how climatic niche axes have evolved across the phylogeny. Phylogenetic ANOVA tries to partition the variance among discrete categories. However, it does not explicitly model the evolution of these categories on the tree. If a discrete character does explain the adaptive optima of a continuous trait, then the evolutionary history of that discrete character is important to understanding the expected distribution of the continuous trait. In light of these considerations, it is unclear how powerful linear models are at distinguishing finer-scale differences in climatic niche between geophytic taxa.

To date, no study has explicitly tested if geophytes with different USOs are adapted to particular climatic niches within an adaptive evolutionary framework that models discrete trait evolution over time. Furthermore, more work is needed to test the extent to which the geophytic life history strategy is non-uniform and if plants with different USOs

converge on different climatic realized niches (the set of climatic conditions that describe where the organism successfully lives and reproduces; Holt, 2009).

If geophytes with independently-evolved and developmentally-distinct morphologies converge on the same climatic niche, then the diverse types of USOs may represent different evolutionary paths towards an effectively similar ecological strategy: retreating underground. This result would imply that the diversity of underground forms are due to developmental or genetic differences in the ancestors of geophytes that predisposed plants to modifying particular types of tissue in different ways when presented with the same types of environmental conditions. Conversely, if plants with different USOs occupy different climatic niches, variation in underground morphology may underlie differences in how these plants relate to their environment. In this case, geophytes may not be a uniform life-history strategy at all, but rather a set of different morphologies that allow plants to survive in different conditions.

Addressing these questions requires a nuanced and detailed approach to the evolution of both morphology and climatic niche. Some types of USOs may be more close in morphological and developmental space because they are modifications of the same type of tissue. For example, corms and rhizomes are both modifications of stem tissue, while in bulbs, the primary storage tissue is derived from modified leaves (see Figures 1.1 and 1.2). A model that represents transitions between corms and rhizomes in the same way that it represents transitions between rhizomes and bulbs effectively erases the complexity of these morphologies and ignores the role that shared developmental mechanisms may play in morphological disparification. Additionally, a non-adaptive model of climatic niche axes is inappropriate for testing adaptive hypotheses. The Ornstein Uhlenbeck (OU) process modifies Brownian motion such that trait values are expected to evolve towards an optimum, θ (Hansen, 1997; Butler and King, 2004), approximating the evolution of a continuous trait towards one or more adaptive peaks. Further modifications of these processes allow for the strength of selection towards those peaks, the number of peaks, and the baseline rates of evolution to vary (Beaulieu et al., 2012; Uyeda and Harmon, 2014). To test if plants with the same type of USO are evolving towards a shared optimal climatic niche, I develop a new analytical pipeline that capitalizes on recently-developed comparative methods that allow explicit testing of adaptive hypotheses (OU models; Uyeda and Harmon, 2014) and the incorporation of complex, nested relationships in morphological characters (Tarasov et al., 2019).

Representation of diverse geophytic morphologies is particularly high in the order Liliales, which contains roughly 1200 species distributed across the globe (Givnish et al., 2016), including geophytic taxa with a striking diversity of USOs (Figure 1.1). Liliales is an excellent group in which to test if geophytic organs are adaptation to different climatic niches, as there appear to have been many transitions between geophytes with diverse USOs, encompassing most of the main types of underground organs found across plants.

In this study, I build a species-level phylogeny of roughly 50% of the taxa in Liliales by capitalizing on the growing availability of published genetic data (Benson et al., 2018), advances in supermatrix construction (de Queiroz and Gatesy, 2007), and model-based tree-

Table 1.1: List of gene regions and percent coverage used in phylogenetic reconstruction.

Gene Region	Genome	Taxon Coverage	Length (bp)
matK	chloroplast	51.2%	1699
trnL-trnF spacer	chloroplast	50.6%	1169
ITS	nuclear	35.6%	979
atpB	chloroplast	33.7%	1500
psba	chloroplast	28.2%	1127
rpl16	chloroplast	26.9%	1371
rbcL	chloroplast	21.9%	733
nadhF	chloroplast	21.7%	701
atpB-rbcL spacer	chloroplast	20.3%	940
rps16	chloroplast	19.8%	868

building algorithms (Ronquist et al., 2012), which collectively have widened the scope of phylogenetic reconstruction and allowed for increasingly large trees and thus larger datasets necessary to obtain statistical power for these more complex phylogenetic comparative methods. I use this phylogeny and my newly-developed analysis pipeline to test the relationship between underground morphologies and modes of evolution in climate seasonality. Specifically, I hypothesize that different USOs are adaptations to specific climatic niches and thus species with those USOs should correspond to different climate values in analyses.

1.2 Materials and Methods

Data

I generated three data sets for downstream analysis: a distribution of species-level phylogenies including 50% of the species in Liliales, a modeled climatic niche for each species based on 19 climatic variables (Fick and Hijmans, 2017), and a detailed underground morphology database for all species.

Phylogeny

I used SUMAC 2.0 (Freyman, 2015) to download gene regions from NCBI GenBank (Benson et al., 2018) for all species in the order Liliales. I targeted genes that clustered with specific guide sequences to specifically obtain gene regions for 10 commonly sequenced genes in Liliales (Table 1.1).

I filtered the resulting sequences using custom python scripts (Appendix B) to remove gene regions with especially high percentages of missing data (fewer than 150 taxa out of 621), remove sites with more than 95% missing data, align and edit sequences, and

Table 1.2: Constraints used for dating analysis.

Constraint	Age (mya)	Calibration Type
Luzuriaga stem node	23.2	Fossil from Iles et al. (2015)
Ripogonaceae stem node	$51 - 52$	Fossil from Iles et al. (2015)
Liliales stem node	$115.6 - 131.1$	Secondary from Givnish et al. (2016)

rename taxa with taxonomic errors. All regions were aligned using MAFFT v7.271 (Katoh and Standley, 2013); some alignments (ITS, psbA, and rpl16, and trnL-trnF spacer) failed to align well under MAFFT and were subsequently aligned using PASTA (Mirarab et al., 2015) to improve alignment accuracy. I concatenated the filtered and edited alignments using Sequence Matrix (Vaidya et al., 2011). I reconstructed the phylogeny using `MrBayes v3.2.6` (Ronquist et al., 2012) on CIPRES (Miller et al., 2010) with two independent runs of four chains, partitioned by gene region, each under the $GTR + \gamma$ model with default priors. I constrained tree space to the known family-level relationships within the order according to the Angiosperm Phylogeny Website (Stevens et al., 2016) to reduce run times.

For downstream analyses, I randomly selected five trees from the posterior distribution; performing analyses over a distribution of trees accounts for phylogenetic uncertainty, which may be particularly important in large supermatrix approaches with high percentages of missing data and low support for some branches of the tree. I dated each of the selected trees in `R` (R Core Team, 2013) using the `chronos()` function from the `R` package `ape` (Paradis and Schliep, 2019)—an implementation of the penalized likelihood approach—using data from two fossils and a secondary calibration (See Table 1.2; Iles et al., 2015; Givnish et al., 2016).

Climate

I modeled the climatic niche of each species in Liliales using a newly-developed `R` pipeline, Climate and Niche Distribution Inference (`CaNDI`, manuscript in preparation, all code available on GitHub; Appendix B) that gathers and cleans species occurrences, downloads climate data (among other variables), and estimates niches for hundreds of taxa at a time. `CaNDI` takes as input a list of species, queries the Global Biodiversity Information Facility (GBIF; Flemons et al., 2007) and the Botanical Information and Ecology Network (BIEN Maitner et al., 2018) for occurrence records, and cleans those records using a series of filters designed to remove latitude and longitude records that fall outside of the species' native range, exactly at 0°, 90°, or 180°, or in the ocean. `CaNDI` then passes these climate data and occurrences to MaxEnt (Phillips and Dudík, 2008) to estimate the climatic niche using three replicates. For each species, `CaNDI` returns the probability of occurrence across the landscape.

I used climate data from the WorldClim database (Fick and Hijmans, 2017): 19 bioclimatic variables which describe various aspects of temperature and precipitation. Collinearity of predictor variables does not affect model performance (except in cases of

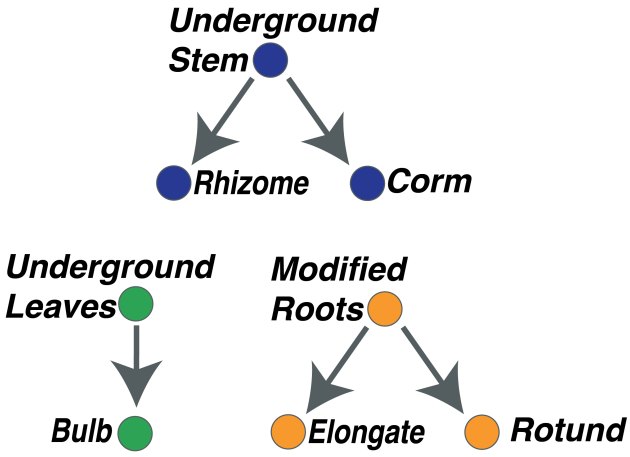

Figure 1.2: Hierarchy of morphological states used in PARAMO analysis.

model transfer Feng et al., 2019), so I included all 19 variables in niche estimation. To obtain a single estimate of the optimal value for each climate variable, I selected the value that corresponded to the part of the species' range with the highest probability of occurrence. Downstream analyses focused on the two axes of the multidimensional niche that describe seasonality: seasonality of precipitation and seasonality of temperature.

Morphology

I used morphological data from Kew's World Checklist of Selected Plant Families (WCSP WCSP, 2020) to code the USOs associated with species in Liliales. For taxa listed as tuberous, I referred to morphological literature (Kubitzki and Huber, 1998; Sanso and Xifreda, 2001; Pate and Dixon, 1982) for more detailed descriptions of their USOs, as the WCSP uses tuber as a catch-all category, including for corms, root tubers, and other organs. The final coding scheme consisted of the presence and absence of eight characters, grouped into three hierarchical clusters based on tissue type (leaf, stem, and root; see Figure 1.2).

Analysis

I combined the climate data, morphology data, and phylogeny in a novel analysis pipeline that integrates newly-developed methods for modeling continuous characters (such as climate) and discrete characters (such as morphological categories). This novel pipeline models the continuous variables (climate) and the discrete variable (morphology) independently, and then asks if variation in adaptive optima for continuous characters are explained by the evolution of the discrete trait, allowing for complex models of the discrete trait and for an imperfect correspondence between adaptive optima and the discrete trait. I applied this pipeline to test the hypothesis that underground mor-

Table 1.3: Priors used for `bayou` analysis of climate variables. In this context, climate refers to the particular climatic niche variable used in the analysis, so the priors on θ differed for the two analyses.

Model Parameter	Distribution	Distribution Parameter Value(s)
α	Half Cauchy	$scale = 0.1$
σ^2	Half Cauchy	$scale = 0.1$
K	Geometric	$p = 1/30$
θ	Normal	$mean = mean(climate)$
		$sd = 1.5 \cdot sd(climate)$

phology (discrete) explains the patterns of adaptive optima in two axes of climatic niche (continuous).

Climatic Data Analysis with`bayou`

The Bayesian reversible-jump Ornstein-Uhlenbeck (OU) process implemented in the `R` package `bayou` (Uyeda and Harmon, 2014) tests adaptive hypotheses by modeling the stochastic process by which a continuous trait evolves across a phylogeny towards various optimal values. This method, an extension of the Brownian motion process (Felsenstein, 1985), models the number and placement of adaptive regimes across the branches of the phylogeny, where each adaptive regime is characterized by a unique optimal continuous trait value, θ; the rate of evolution, σ^2, and the strength of selection, α. These regimes and their associated parameter values are sampled in proportion to their posterior probabilities. I use `bayou` to describe the modes evolution of climatic niche in Liliales.

For both temperature and precipitation seasonality I log-transformed the variable and ran `bayou` for 3.5 million generations using the priors specified in Table 1.3. I assessed convergence using the `R` package `coda` (Plummer et al., 2006) and discarded the first 1% of samples as burnin. I drew 1000 samples from the posterior distribution of adaptive regimes for each climatic niche variable to use in subsequent calculations. Each of these samples contains a history of continuous trait evolution, where the estimated adaptive optima are estimated to evolve and shift along branches of the phylogeny.

Morphological Ancestral State Reconstructions with PARAMO

For each of the five trees, I used the PARAMO pipeline (Tarasov et al., 2019) to reconstruct the evolution of underground morphologies using hidden, structured Markov models and stochastic mapping. The hidden states represent the 'predisposition' to evolve USOs of different tissue types, incorporating the hierarchy of states illustrated in Figure 1.2. Each cluster (leaf, stem, and root) corresponds to a distinct evolutionary model such that the clusters evolve independently. First, for each of the three clusters, I estimated evolutionary rates using the `R`package `corHMM` (Beaulieu et al., 2013). I then reconstructed their evolutionary histories by simulating the evolution of each cluster 1000 times under

the inferred model of evolution and conditioning on the observed data at the tips, a process termed stochastic mapping. Each simulation produces a single stochastic map that illustrates a possible scenario of character evolution under the inferred model. PARMO performs simulations using the textttmake.simmap() function in `phytools` (Revell, 2012). I produced a set of 1000 combined phenotype maps by overlaying the maps of each cluster. In isolation, each cluster only contains information about some of the USOs used the study. The combined phenotype map illustrates the history of underground morphology as the entire underground phenotype with many character states.

Calculating State-Specific Climatic Optima

For a given `bayou` sample and a given morphological stochastic map, I calculated the optimal climate value for each morphological trait weighted by the amount of time spent in the particular combination of optimum and morphological trait (Figure 1.3 B). As the observed morphology rather than the hidden states are expected to affect an organism's relationship to its environment, I collapsed the full set of character states (observed and hidden) into the observed states by combining the state-dependent optima of states with the same observed state but different hidden states. This procedure produced a set of state-specific optimal values per character state for each of 1000 stochastic map pairs (1000 stochastic map simulations from PARAMO and 1000 stochastic map samples from the `bayou` posterior) and for each of the 5 trees, for a total of 5000 state-specific optima per comparison. Some states were visited infrequently during stochastic mapping and thus have high percentages of missing data, which makes estimates of their state-specific optima more uncertain. To avoid these uncertain estimates, I dropped from analysis those states represented in less than 50% of the samples. I performed these comparisons for both precipitation and temperature seasonality and for the three morphological clusters (leaf, stem, and root) and the combined morphological phenotype, for a total of eight comparisons.

Hypothesis Testing

My null hypothesis is that plants with different USOs are not evolving towards different climatic niche optima; in other words, state-specific climatic niche optima (θ) are equal. Correspondingly, under the alternative hypothesis, state-specific θs are different. However, for any finite dataset, state-specific θs will be different, even under the null hypothesis, and this issue is exacerbated by the phylogenetic structure to the inferred θ values. To account for this structure, and the finite sample, I calculated a test statistic that summarizes the overall difference between any particular set of state-specific θs. I simulated the distribution of that test statistic under the null model and checked whether the differences generated under the null model are less than the differences of the empirical estimates.

I simulated 1000 null histories of the three characters using the `sim.history()` function in the `R`package `phytools` (Revell, 2012) and the estimated Q-matrices from the em-

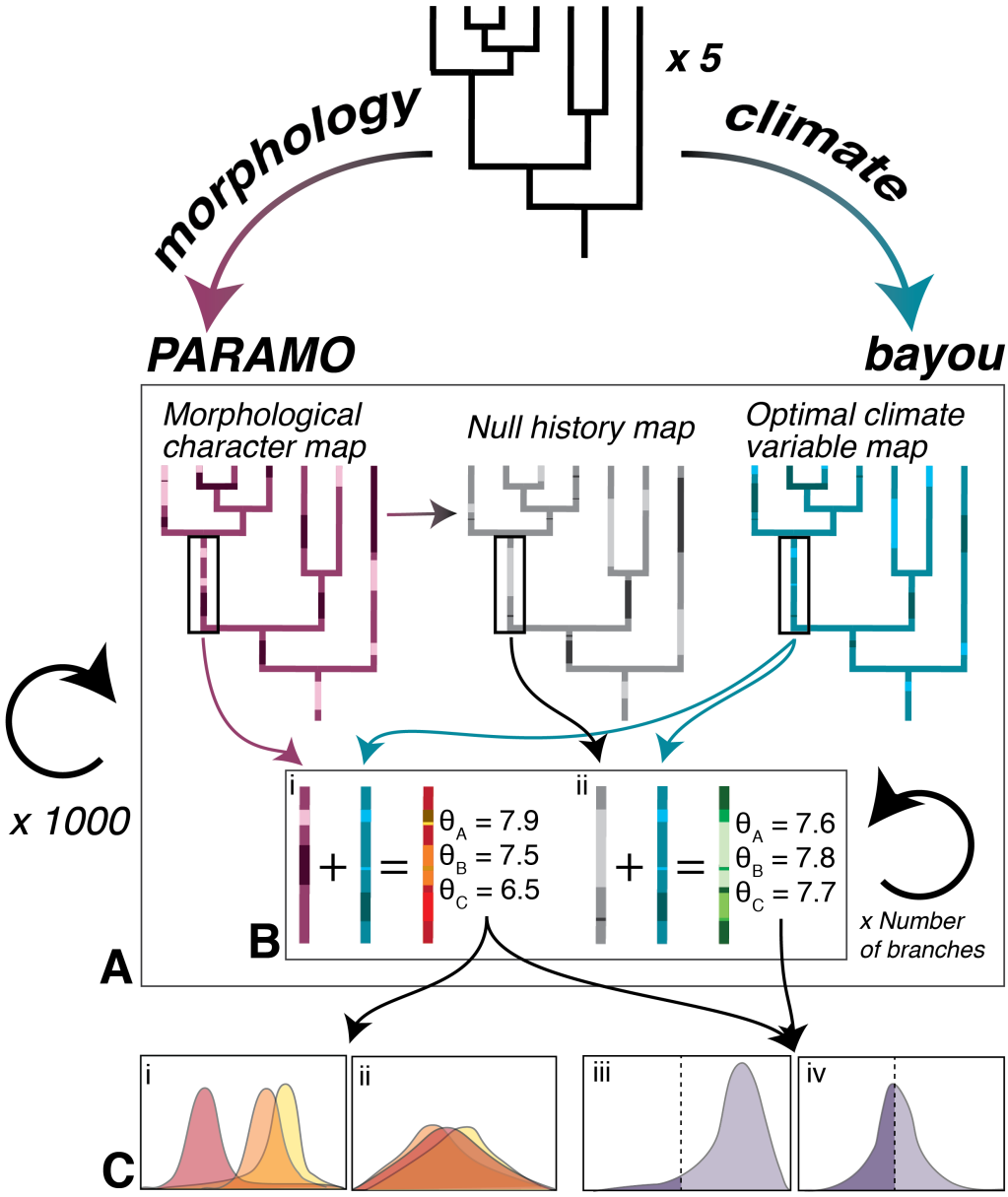

Figure 1.3: Schematic representation of methods used in this study. For each of five phylogenies sampled from the posterior distribution, I: A) used PARAMO and `bayou` to generate 1000 stochastic maps of the morphology and estimated climatic niche optima (θ), respectively, and simulate 1000 null histories (see Methods: Hypothesis testing) using the estimated evolutionary models from PARAMO. B) I combined each of the 1000 optimal climate variable maps with a morphological character map, and separately with a null history map, to create two sets of composite maps that show the state-specific θs per character state. C) I plot the densities of all estimated state-specific θs (i, ii) and the distribution of a test statistic (iii, iv; see Methods: Hypothesis testing). Plots i and iii represent a case where I would reject the null model, while in plots ii and iv I would fail to reject the null.

pirical analyses. This procedure differs from stochastic mapping in the empirical analysis in that simulations are not conditioned on the observed character data. As in the empirical analysis, I also combined the three characters to produce 1000 null histories of the combined phenotype.

To calculate the test statistic, for a given vector of state-specific θ values, I calculated the pairwise distance between all state-specific θ, which results in a distance matrix, D:

$$D = \begin{pmatrix} d(\theta_{1,1}) & d(\theta_{1,2}) \\ d(\theta_{2,1}) & d(\theta_{2,2}) \end{pmatrix},$$

where $d(\theta_{i,j})$ is the distance between θ_i and θ_j:

$$d(\theta_{i,j}) = \sqrt{(\theta_i - \theta_j)^2}.$$

I then measure overall amount of difference using the Frobenius norm, $F(D)$, which summarizes the magnitude of the distance matrix. For a given `bayou` sample i, I calculate, S_i, the difference between the $F(D_i)$ given the stochastic map and $F(D)_i$ given the null history. I then compute S_i for each `bayou` sample; if 0 is in the 95% credible interval of S, then the null cannot be rejected.

1.3 Results

Across all five trees, the ancestor of Liliales is estimated to have a rhizome with no leaf or root modifications (Figure 1.4 A–D shows the reconstructions for one tree), in agreement with previous work (Patterson and Givnish, 2002; Howard et al., 2019). For both temperature and precipitation seasonality, the estimated optima at basal branches in the phylogeny are highly seasonal, with subsequent shifts into less seasonal optima along more recent branches (Figure 1.4 E and F).

Figure 1.5 illustrates the distribution of estimated state-specific optima. Overall, the distributions of state-specific optima are more dispersed across the temperature seasonality axis than the precipitation seasonality access. For the leaf cluster, the estimated bulb-averaged optima are more seasonal than those for no bulb. For the stem cluster, both rhizomes and corms are estimated to be less seasonal than non-stem modification, though for precipitation just rhizomes are estimated to be less seasonal than the other states. In roots, root tubers (especially rotund root tubers) have lower temperature seasonality than no root tubers, but for precipitation, rotund and no root tubers overall overlap in their distributions while elongate root tubers are estimated to be more seasonal. In the combined phenotype for temperature, the most seasonal state for precipitation is bulb with rhizome and the least seasonal state is rhizome and rotund root tubers. For precipitation seasonality in the combined phenotype, most state-specific optima have highly overlapping distributions, though it appears that rhizomes are slightly less seasonal and rhizomes with

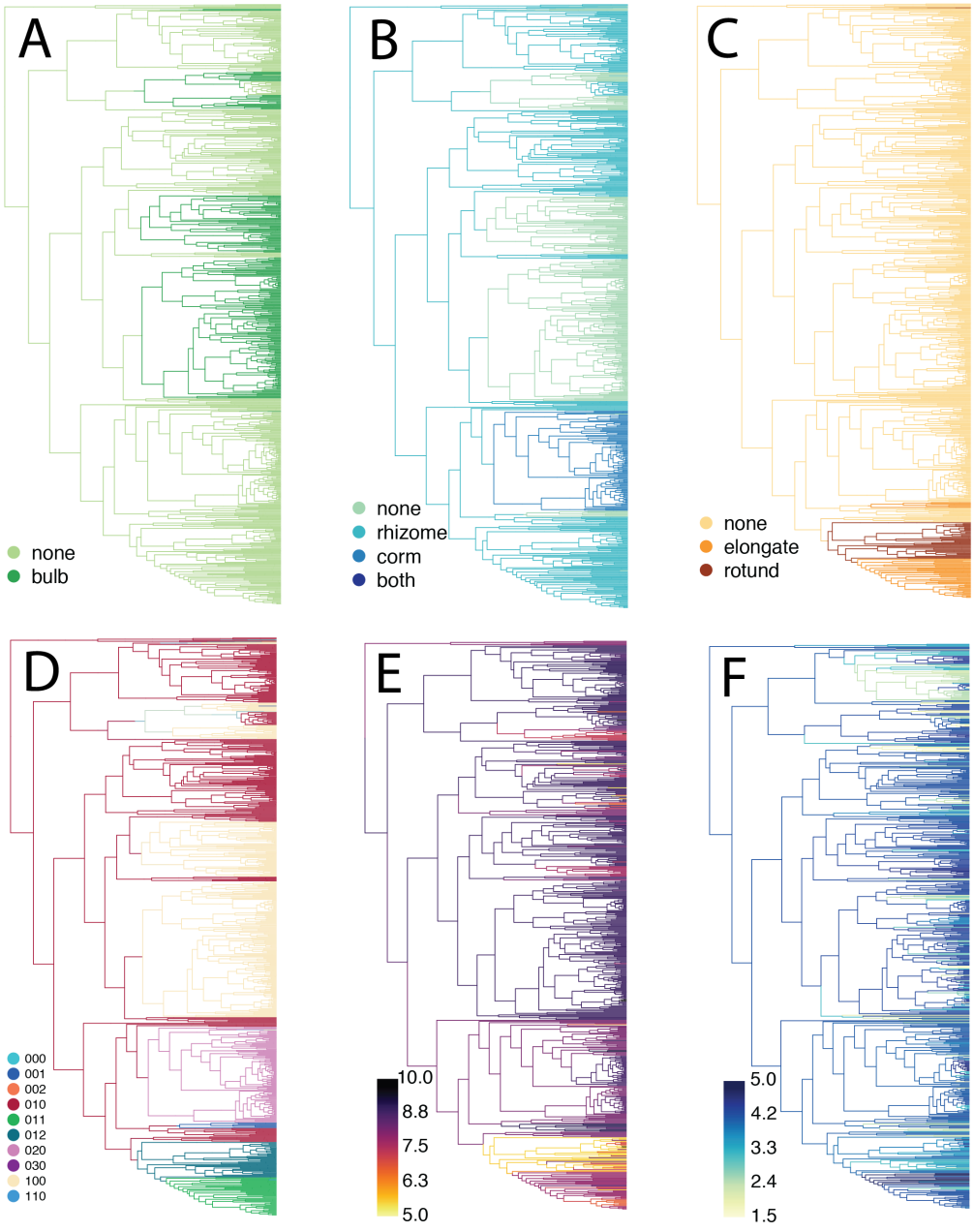

Figure 1.4: Stochastic maps of intermediary results for one of the trees used in analysis: A) leaf character from PARAMO, B) stem character from PARAMO, C) root character from PARAMO, D) combined phenotype character from PARAMO, E) posterior mean branch-specific θ for temperature seasonality F) posterior mean branch-specific θ for precipitation seasonality. In A, B, and C, the "none" category applies to absence of modifications of the relevant tissue type; those taxa may have modifications of the other tissues.

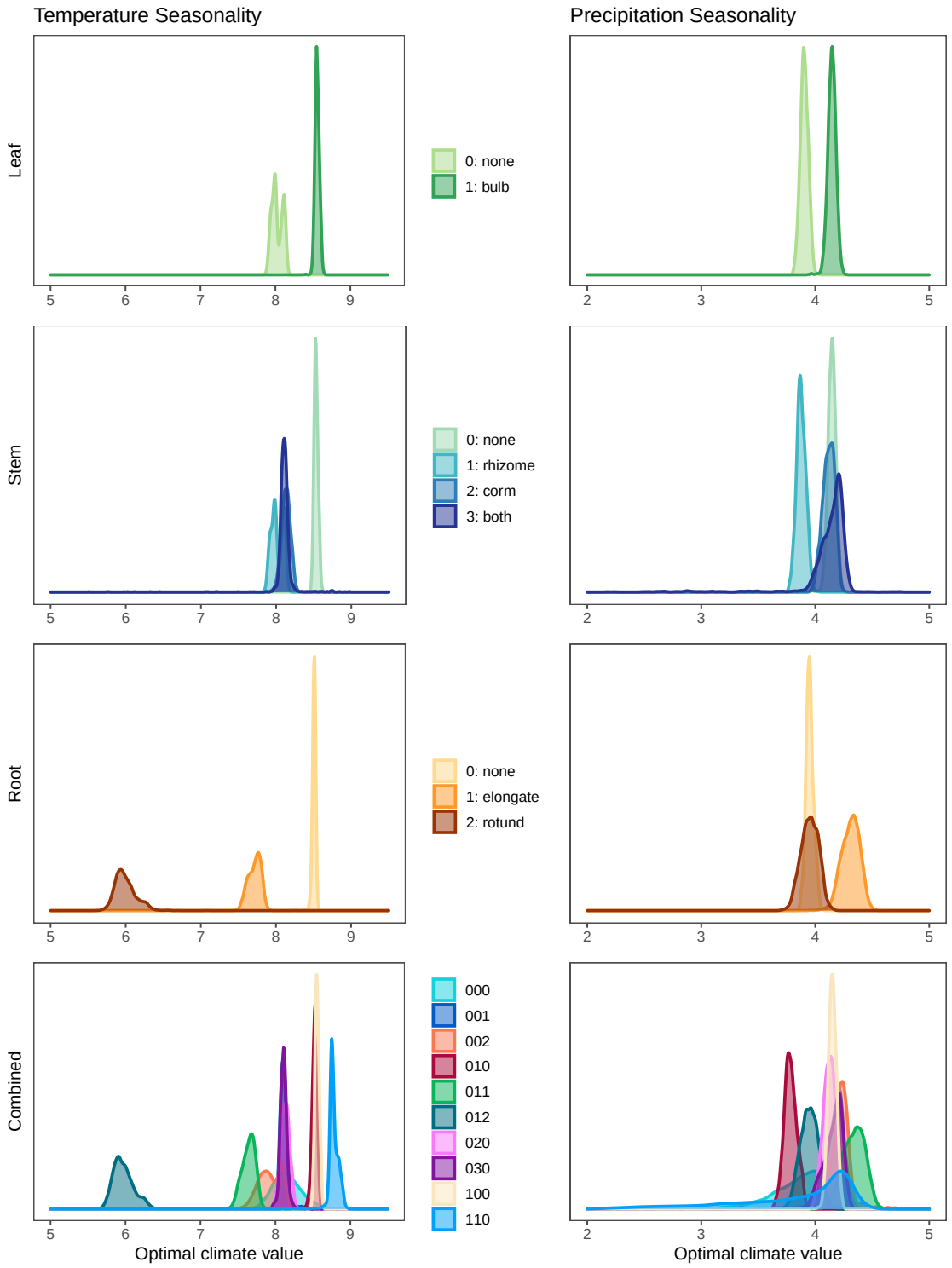

Figure 1.5: Posterior distributions of state-specific θs. For the combined morphology densities, digits in the state labels refer to states for each cluster. The first digit corresponds to the leaf state, the second digit corresponds to the stem state, and the third digit corresponds to the root state. For example, the 000 state refers to the absence of any USO (non-geophytes), and 012 refers to rhizomes and rotund root tubers.

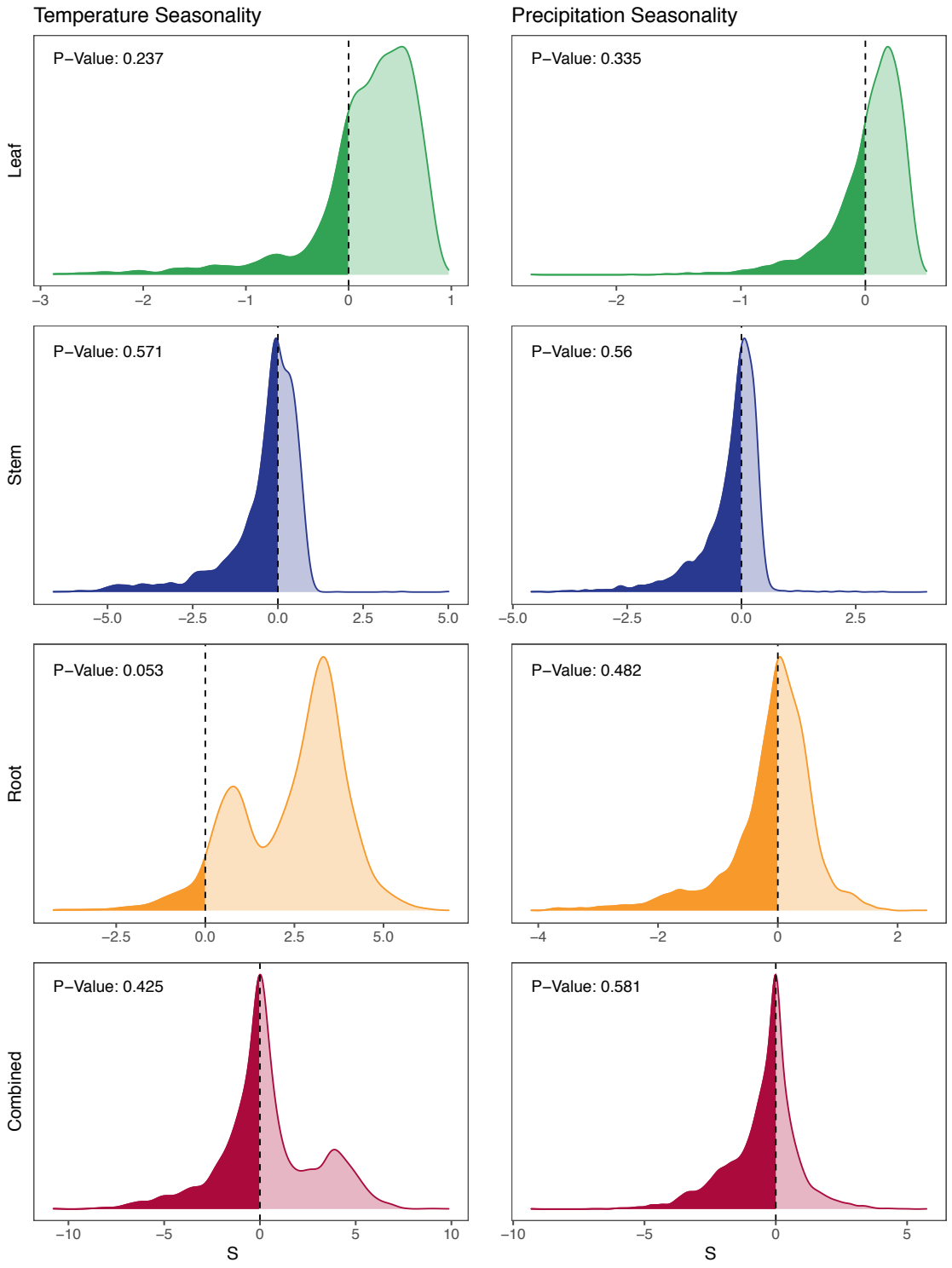

Figure 1.6: Test statistic (S) distributions. Distributions with 95% > 0 are considered evidence for a statistically significant difference in state-specific θ. P-Values correspond to the percent of values equal to or less than zero.

elongate root tubers are slightly more seasonal than other distributions. For both temperature and precipitation, the distribution corresponding to no modified underground organs (non-geophytes) falls out intermediate along both climatic niche axes and overlaps with many other states, signifying that non-geophytes are not more or less seasonal than geophytes.

While the state-specific optima curves in Figure 1.5 appear distinct for many of the morphology-climate comparisons, across both climate variables, no distributions are more different than expected by chance (P-Values 0.237 - 0.581), with the exception of the root cluster for temperature seasonality, which is marginally significant (P-Value 0.053; Figure 1.6).

1.4 Discussion

Getting at the root of the problem: are USOs adaptations to particular climatic niches?

Together, these analyses demonstrate that plants in Liliales with the same underground storage organ do not share similar climatic niches more than expected by chance, with the exception of root morphology, where the presence of modified roots, especially rotund root tubers, is associated with lower temperature seasonality. Furthermore, non-geophytes in Liliales are not associated with more or less seasonal climates than their geophytic counterparts. While many of the state-specific optima curves (Figure 1.5) appear distinct, the null model suggests that these differences could be observed even with no correspondence between the trait and climate (Figure 1.6), due to the phylogenetic structure of environmental niche preference in the underlying data. This null model represents that case where a simulated discrete trait evolves under the same evolution model as the estimated models for USOs, but without the observed pattern at the tips. Any correspondence between the simulated traits and climate is due to chance, the distribution of the climate data on the tree, and/or the rates of evolution and stationary frequencies of the model, rather than the distribution of morphology across the tree.

My findings suggest that root tubers may be an adaptation to distinct ecological conditions or that they experience physiological constraints that restrict root tubers to particular climatic niches, distinguishing them from the other USOs included in this study. Unlike corms, rhizomes, and bulbs, root tubers are rarely the source of perenniating underground buds. Most geophytes in Liliales (liliid geophytes) with root tubers also have rhizomes (e.g., *Bomarea*, *Alstroemeria*), which may serve as the source of underground buds while the root tubers store nutrients and water (Tribble et al., in review). Many geophytes regenerate their USOs annually (especially bulbs and corms; Pate and Dixon, 1982; Kamenetsky and Okubo, 2012), and thus the processes of nutrient flow between the USO and the above ground plant are necessarily linked to the same seasonal cycles. Partitioning growth and storage between organs (as in the case of the species with tu-

ber and rhizomes) may be particularly advantageous in climates with less temperature seasonality, as it allows for the continuous production of aerial shoots and the periodic replacement of stored nutrients as needed. Alternatively, places with less seasonal temperatures may be less likely to reach the freezing point, and root tubers may be particularly maladapted to frost compared to USOs derived from stem or leaf tissue. Future research should include more axes of the climatic niche to more specifically characterize the distinctions between niches of plants with different USOs.

These results differ from previous work in three primary ways. First, while I found no significant difference in environmental niche geophytes and non-geophytes, prior work suggested that being geophytic is associated with lower temperatures and precipitation and higher temperature variation compared to not being geophytic (Howard et al., 2019). The sparse representation of non-geophytes in Liliales indicates that these results may not reflect generalizable patterns of climatic niche occupancy between geophytes and non-geophytes, as these data likely lack the power to make this distinction. Liliales contains many taxa with diverse USOs, but non-geophytes are limited to a few exceptional plants, including *Alstroemeria graminea*, an annual native to the Atacama desert, and *Campynema lineare*, a curious herb distantly related to the rest of the order. Secondly, previous work found that rhizomes are correlated with increased temperature variation and found no evidence for difference niches between tuberous and non-tuberous taxa (Howard et al., 2019), while my analysis found no significant association between rhizomatous or non-rhizomatous taxa but instead suggests that taxa with root tubers are evolving towards lower optimal values of temperature seasonality. Thirdly, Patterson and Givnish (2002) found evidence that in the core Liliales, convergence on bulbs correlated with independent transitions into seasonal and high light intensity habitats, but while I also find evidence for several independent transitions to bulbs, the association between bulbs and increased seasonality is not statistically significant in my results (P-Values 0.237 and 0.335). The present study differs from prior work in terms of level of detail for morphological characters, number of included climate variables, methods of analysis, and taxonomic scope, so it is possible that differences are due to any combination of these factors. In particular, the Howard et al. (2019) study does not distinguish between root tubers and stem tubers, and thus may not have recovered the association I find here.

In Liliales, rotund and elongate root tubers are mostly found in a few clades (namely *Bomarea*, *Alstroemeria*, *Burchardia*, and a few additional taxa), so it is possible that the strong association between decreased temperature seasonality and the presence of root tubers is driven by an unmeasured trait that happens to co-occur in those clades. For this reason, extrapolating my results to non-liliid geophytes may not be appropriate, so follow-up studies should address variation in root morphology and climate in other clades, particularly in groups with many independent transitions between the absence and presence of root tubers and between different root tuber morphologies. *Asparagus* would be a particular appropriate system in which to further test these associations, as root morphology is highly variable in the genus (Leebens-Mack, pers. comm.). However, there are few other clades known for well-characterized variation in the presence and ab-

sence of root tubers or for variation in root tuber morphology. Root morphology is vastly undercharacterized in many plant clades (Tribble et al., in review; Janzen et al., 1975), so this work also motivates increased morphological characterization of underground morphology and root morphology in particular and demonstrates the importance continued emphasis on classic botanical techniques for understanding biodiversity. Studies that characterize the functional ecology and physiology of root tubers will likely yield important insights into how and why they differ from other USOs.

Applicability of analysis pipeline beyond the present study

This project demonstrates the utility of adapting existing, complex evolutionary models to address a long-standing challenge in statistical comparative biology: modeling discrete and continuous traits. The methodological pipeline presented in this study uses PARAMO to estimate the evolutionary history of a discrete trait while taking into account nested relationships between character states and `bayou` to estimate regimes of continuous trait evolution (including number and location of optimal continuous trait values). It then combines these estimates by calculating the average optimal values for each discrete trait and asks if these averaged optima are more different than expected under a null model.

This pipeline has several advantages over existing methods. First, the model parameters have clear biological interpretations. Densities from Figure 1.5 show the distributions of probable optimal phenotypes (θ) of the continuous trait by discrete category. These θs correspond to peaks in the adaptive landscape of the continuous trait, so the estimated parameter values have a clear and direct link to adaptive evolutionary theory, unlike the estimated effect sizes from linear models. Second, the methods directly model the evolutionary processes of both the continuous and discrete traits and thus are more appropriate for addressing hypotheses of adaptive evolution. Third, the pipeline can accommodate any number of models for discrete traits. While I use PARAMO (Tarasov et al., 2019) to model the nested relationships between USOs derived from the same tissue, any model of evolution which can be simulated under could generate stochastic map histories and thus be incorporated into the pipeline. In some empirical cases, not including the present study, traits are explicitly defined in ontologies (e.g. `phenoscape.org`), and PARAMO uses these definitions directly to establish the hierarchy between character states. As the use of ontologies in comparative biology becomes more common, incorporating hierarchies codified in ontologies will greatly expand the ability of researchers to test increasingly complex hypotheses, including how developmental processes impact trait evolution and the relationships between traits. This modularity means that the pipeline is flexible and can incorporate a wide range of plausible empirical datasets and hypotheses and thus could be adopted widely across macroevolutionary research. All code and scripts used in this pipeline will be made publicly available upon publication to facilitate adoption of the methods.

Conclusions

The results of this study support differences in the optimal climatic niches of plants with some underground storage organs, namely between plants with and without root tubers. Having root tubers, particularly rotund root tubers, is associated with lower temperature seasonality in Liliales, suggesting that root-derived underground morphologies may experience different ecological constraints or differentially mediate ecological interactions than do shoot-derived USOs. This implies that the developmental origin of the structure influences the way it mediates ecological relationships and draws into question the appropriateness of ascribing broad ecological patterns uniformly across geophytes. My work highlights an important clade of geophytes within the monocotyledons, and zooming in on this specific group allows me to be detailed in describing morphology and may illuminate clade-specific trends not detected in a larger analysis. Furthermore, this study serves a proof-of-concept for a new analysis pipeline that models the adaptive evolution of a continuous trait based on the hierarchical, nested evolution of a discrete trait. This pipeline is applicable across many areas of evolutionary biology and may serve as a model for future hypothesis-driven comparative research.

Chapter 2

Comparative transcriptomics of root tuber formation in a monocotyledonous geophyte

2.1 Introduction

The vast majority of scientific attention in botanical fields focuses exclusively on aboveground organs and 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 et al., 1934; Dafni et al., 1981b,a; 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., 2019, 2020). 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 analogous functions: 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 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 detailed 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 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), and more. Fewer studies have focuses on monocotyledonous taxa (including important studies in onion, such as in Lee et al., 2013), and these studies do not encompass the diversity of underground forms present 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 and Sosa, 2016; Sosa and Loera, 2017; Ott et al., 2019; Howard et al., 2020). 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., 2019), 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 (leaf tissue), corms (stem), rhizomes(stem), and tubers (various) 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 (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 and 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, members of the phosphatidylethanolamine-binding protein (PEBP) family, 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, but to date no study has focused on the evolutionary history of these genes in underground structures.

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 and 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 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 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 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 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.

2.2 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 (University Herbarium at UC Berkeley)] and germinated in greenhouse conditions at the University of California, Berkeley designed to replicate emphB. multiflora's native conditions (70°F – 85°F and 50% humidity). Six months after germination, three sibling individuals were harvested as biological replicates. The aerial shoot apical meristem (SAM) of a single branch, the rhizome apical meristem (RHI), root apical meristems (RAM) of several fibrous roots, and the growing tip of a tuberous root (TUB; Figure 2.1) were dissected from each of three individuals, for a total of 12 tissue samples.

Samples were immediately frozen in liquid nitrogen and maintained at -80°C until extraction. Total RNA was extracted 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. As this protocol did not work well for the vegetative SAM samples, RNA was extracted from two dissected SAMs using an alternative plant tissue RNA extraction protocol (Yockteng et al., 2013). Quality of total RNA was measured with Qubit (ThermoFisher, Waltham MA) and Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA); if needed, a Sera-Mag bead clean-up was used to further clean extracted RNA (Yockteng et al., 2013). Two SAM samples failed to extract at sufficient

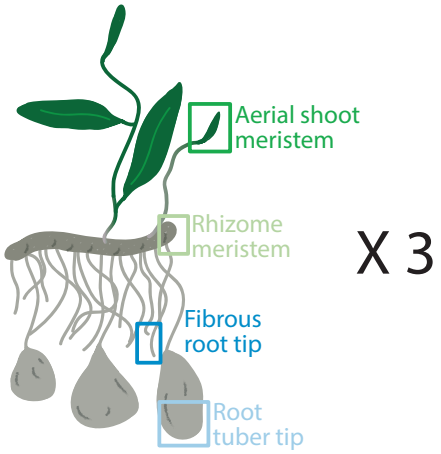

Figure 2.1: Sampling scheme of tissue types. *Bomarea multiflora* has modified underground stems (rhizomes) and modified roots (tuberous roots). Aerial shoots and roots develop from the underground rhizome. We extracted RNA from the aerial shoot meristem, rhizome meristem, fibrous root tip, and tuberous root tip of three *B. multiflora* individuals.

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

Bioinformatics Data Processing

Raw reads were cleaned, processed, and assembled using the Trinity RNA-Seq De novo Assembly pipeline (Grabherr et al., 2011) under the default settings unless otherwise stated in associated scripts. All analyses were run using the Savio supercomputing resource from the Berkeley Research Computing program at UC Berkeley. Reads were cleaned with Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), keeping un-

paired reads and using a minimum fragment length of 36 base pairs. Data from all tissue types and biological replicates were concatenated in order to generate a consensus transcriptome, assembled de novo from the concatenated data using Trinity. Each sample was compared back to the assembled consensus transcriptome, aligned using Bowtie 2 (Langmead and Salzberg, 2012), and quantified using RSEM (RNASeq by Expectation Maximization; Li and Dewey, 2011). The consensus transcriptome was annotated with a standard Trinotate pipeline (https://trinotate.github.io/), comparing assembled transcripts 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. The concordance of biological replicates was tested by looking for significant differences between the total number of fragments per replicate, by comparing the transcript quantities of all replicates to each other, and by checking the correlations between replicates. Transcripts with 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/ genes (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 D.1 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 D.2 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 pre-

viously 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-Gutiérrez 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).

2.3 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 C.1 - C.4) 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 2.2) 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.

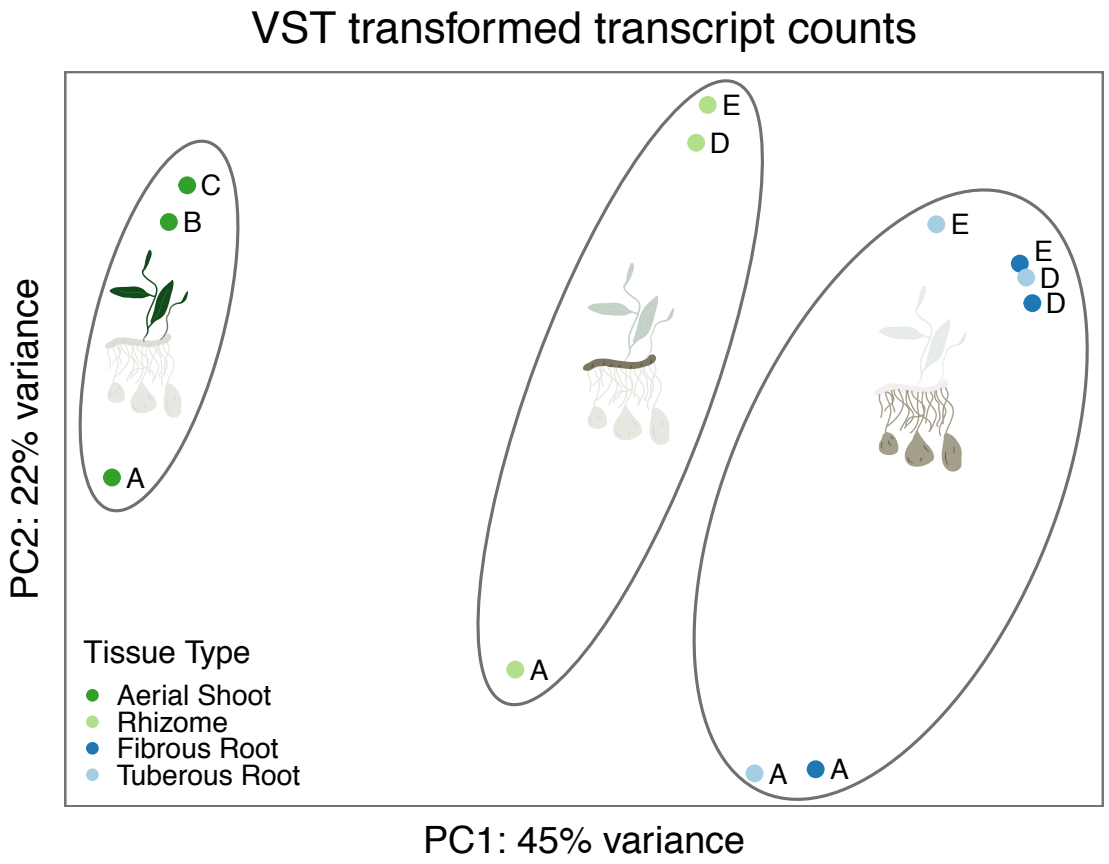

Figure 2.2: Principal component analysis of all samples. We performed a principal component analysis of variance stabling transformed (VST) transcript counts from all biological replicates of all tissue types. Point color indicates the tissue type of the sample and the letter corresponds to individual 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.

Differential Expression

A total of 271 differentially expressed isoforms (DEGs) were recovered between fibrous and tuberous roots (FR vs. TR). 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 2.1. 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

Table 2.1: Top ten GO annotations for all differentially expressed isoforms.

Cellular Components		Molecular Functions		Biological Processes	
Annotation	**Freq.**	**Annotation**	**Freq.**	**Annotation**	**Freq.**
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

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 2.2). 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 2.3 and Table 2.3). 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

Table 2.2: 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 2.3: Gene families, physysiological signalling pathways, and biosynthesis pathways identified in the literature as implicated in USO formation or root thickening. Up-regulation corresponds to an increase of the gene in modified organs vs. non-modified tissue or earlier developmental stages. Up vs. down regulation in *Bomarea edulis* corresponds either to the general pattern of identified members of that group, or those individual genes with significantly different expression levels, visualized in Figure 2.3.

Gene Group	Original Taxa	Up/ down regulation from literature	Up/ down regulation in *B. multiflora*
Abscisic Acid	lotus (Yang et al., 2015), sweet potato (Noh et al., 2010; Dong et al., 2019), and potato (Xu et al., 1998)	up, then down	group: down; individual genes: down
Auxin	many (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)	up	none
CDPK	cassava (Sojikul et al., 2010), potato (Raíces et al., 2003), and lotus (Cheng et al., 2013b)	up	group: down; individual gene: up
Cytokinin	sweet potato (Noh et al., 2010)	up	group: none; individual genes: up and down
Expansins	cassava (Sojikul et al., 2015), *Callerya speciosa* (Xu et al., 2016), lotus (Cheng et al., 2013b), Convulvulaceae (Eserman et al., 2018), and potato (Jung et al., 2010)	up	group: down
Gibberellin	various roots (Tanimoto, 2012), *Callerya speciosa* (Xu et al., 2016), wild sweet potato (Li et al., 2019), and sweet potato (Dong et al., 2019)	down	group: none; individual gene: down
Knox	sweet potato (Tanaka et al., 2008)	up	none
Lignin	cassava (Sojikul et al., 2015), wild sweet potato (Li et al., 2019), and *Callerya speciosa* (Xu et al., 2016)	down	group down; individual gene: down
MADS-Box	wild sweet potato (Li et al., 2019), sweet potato (Noh et al., 2010; Dong et al., 2019), lotus (Cheng et al., 2013b), and *Sagittaria trifolia* (Cheng et al., 2013a)	up	group: up; individual gene: up
Starch	cassava (Sojikul et al., 2010, 2015), lotus (Cheng et al., 2013b; Yang et al., 2015), wild and domestic sweet potato (Eserman et al., 2018; Li et al., 2019; Dong et al., 2019), and potato (Xu et al., 1998)	up	group: up; individual genes: up
14-3-3	*Arabidopsis thaliana* (Van Kleeff et al., 2014; Mayfield et al., 2012; He et al., 2015)	up (in stout roots)	none

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 significantly greater than expected (generally over-expressed in TR compared to FR) and in the remaining four cases (abscisic acid, CDPK, expansins, and lignin biosynthesis) the groups' log2-fold change values are under-expressed.

Fifteen individual DEGs were identified in these gene groups (Table 2.4); 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 log2-fold change values 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 2.4).

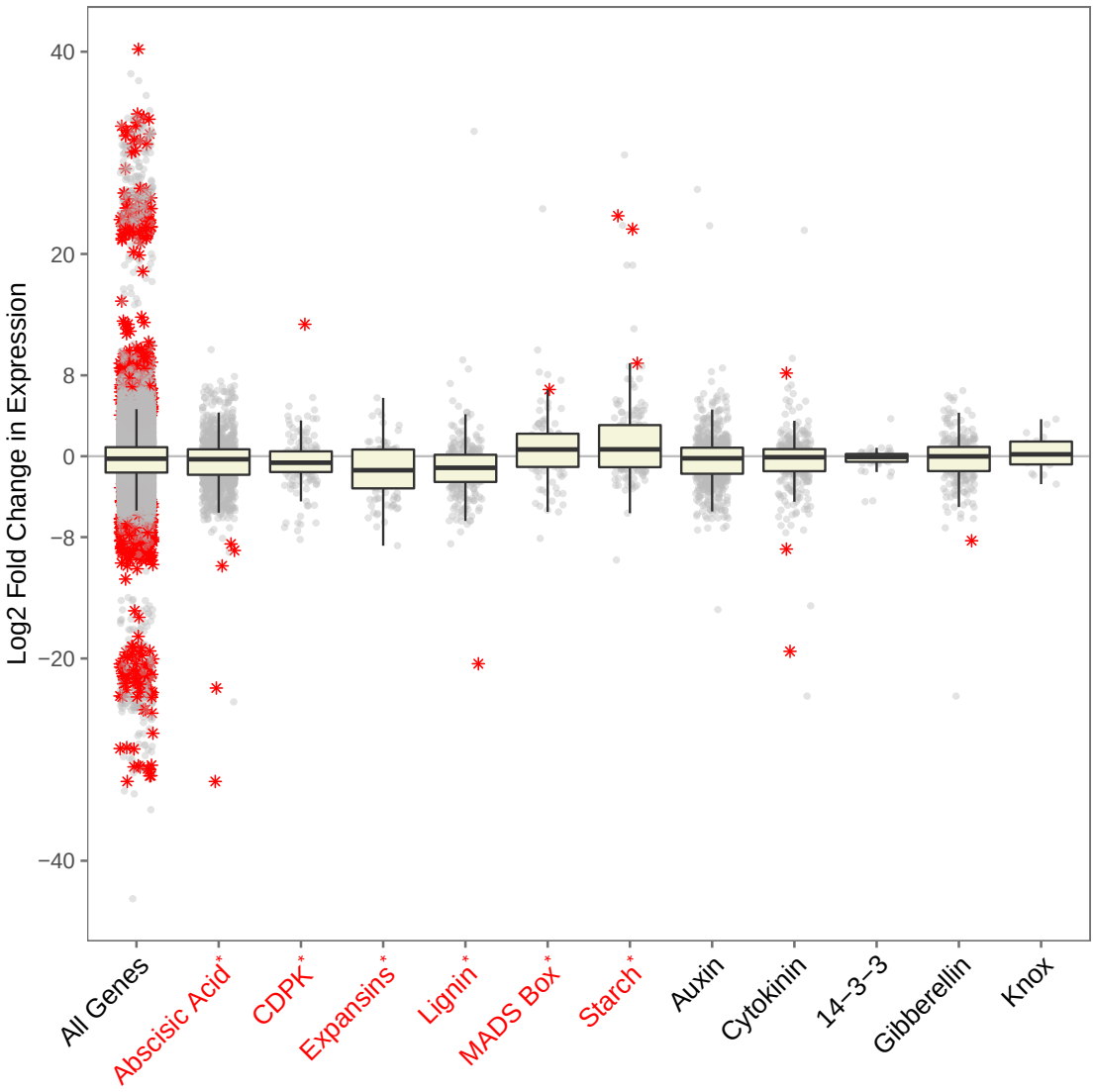

Figure 2.3: Differential expression of candidate gene groups. We identify isoforms corresponding to specific pathways and gene families (gene groups) and categorize the log2-fold change of expression between tuberous and fibrous roots of those groups. Positive values correspond to overexpression in tuberous vs. fibrous roots. 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 that are significantly differentially expressed (padj <0.01) are labeled as red asterisks within the scatter plots.

Table 2.4: 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

Table 2.5: 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

Five specific candidate genes were identified from the literature: 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 using a BLAST E-value cut-off of 0.01 (Table 2.5), but only one putative homolog is significantly differentially expressed ($padj < 0.01$): a putative FT homolog (TRINITY_DN129076_c1_g1_i1), further investigated in PEBP Gene Family Evolution (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 the result is spurious and we do not follow up with further analysis.

PEBP Gene Family Evolution

Thirty-seven *Bomarea* isoforms were identified as putative FLOWERING LOCUS T (FT) homologs. The longest isoform per gene was selected to include in an alignment of phosphatidylethanolamine-binding protein (PEBP) amino acid sequences. Coding sequences from the *Bomarea multiflora* transcriptome were translated to amino acid sequences using TransDecoder v5.5.0 (Haas et al., 2013), removing isoforms that failed to align properly; ultimately, we include six sequences, including the significantly differentially expressed copy. Using the JTT+G4 amino acid substitution models (Jones et al., 1992) we recover three major clusters in our unrooted gene tree, all with strong bootstrap support (Figure 2.4a); 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 with FT genes and three fall out with TFL1 genes. The *Bomarea* DEG homolog is highly supported in the TFL1 cluster with sequences from other monocot taxa (Figure 4b). FT homologs from

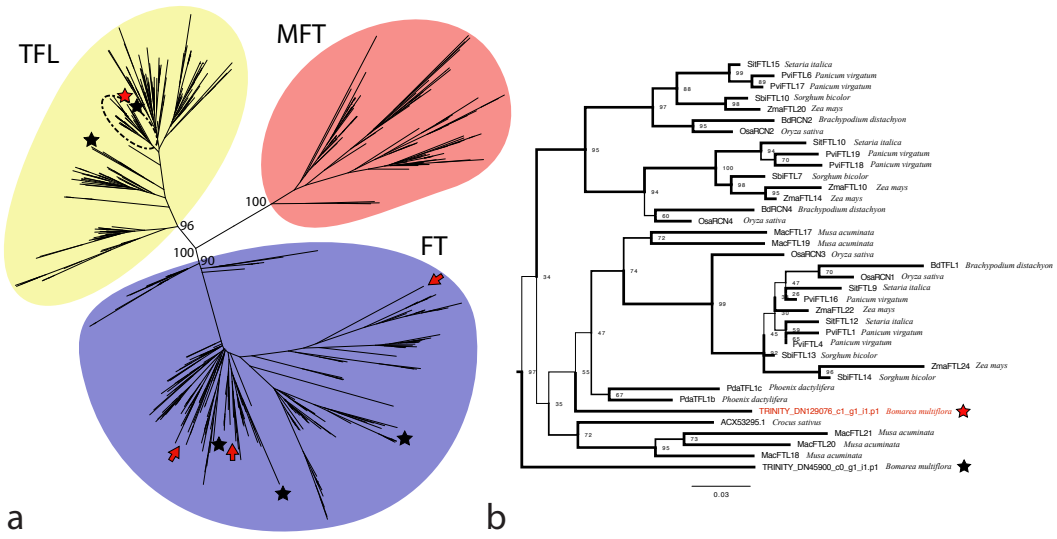

Figure 2.4: Evolution of PEBP genes: (a) Unrooted gene tree of 540 PEBP gene copies from across land plants. Stars indicate all included copies from *Bomarea multiflora*; the red star corresponds to the significantly differentially expressed isoform TRINITY_DN129076_c1_g1_i1). Red and arrows indicate 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.) are labeled as red asterisks within the scatter plots.

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.

2.4 Discussion

How to Make a Tuberous Root

The top 10 most highly represented biological processes of the 271 Differentially Expressed Isoforms (DEGs) include cell wall organization, responses to environmental signals, growth and development, and carbohydrate biosynthesis (Table 1). 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 and 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 and 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 2.2). 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 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 2.2, 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 2.2) may be differentially expressed in tuberous roots to protect starch reserves against potential predation by belowground herbivores. LRR receptors have been implicated in trigger various downstream plant immune responses (Liang and Zhou, 2018).

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-expressed in the tuberous roots of *Bomarea multiflora* (Figure 3). These processes show parallel function across deeply divergent evolutionary distances and in distinct plant structures.

Abscisic Acid (ABA) has been shown to increase initially and then decrease during USO formation in the rhizomes of *Nelumbo nucifera* (Yang et al., 2015), the tuberous roots of sweet potato *Ipomoea batatas* (Dong et al., 2019), and the stem tubers of potato *Solanum tuberosum* (Xu et al., 1998). We find that ABA signaling genes are under-expressed in tuberous roots (Figure 3), perhaps indicating that our samples were in a later developmental stage, characterized by lower ABA levels. Experimental manipulation of *B. multiflora* tuberous roots and developmental time series would be needed to confirm the role of ABA in tuberous root formation. An ABA signaling DEG (Table 2.4) is one of the most differentially expressed isoforms (TRINITY_DN122787_c0_g1_i1: Protein IQ-DOMAIN 32, Table 2.2). This isoform and other ABA genes may signal the cessation of continued elongation in monocot tuberous roots, similar to their role in deeply divergent taxa and non-homologous USOs.

Calcium-dependent Protein Kinases (CDPKs) play an integral role in tuber formation in cassava *Manihot esculenta* (Sojikul et al., 2010). Sojikul et al. (2010) propose that CDPKs may signal the initiation of tuberous root development, similarly to the process described in stem tuber formation in potato (Raíces et al., 2003) and rhizome development in *Nelumbo nucifera* (Cheng et al., 2013b). In these studies, CDPKs are over-expressed and signal the initiation of USO development. Our results show that on average CDPKs as a group are under-expressed in the tuberous roots of *B. multiflora*. However, the only significant CDPK DEG (TRINITY_DN124121_c3_g1_i11: Calcium-dependent protein kinase 2) is over-expressed, suggesting that CDPK 2 expression plays a role in initiating tuberous root development and that CDPK involvement in tuberous root formation is ubiquitous in all studied taxa.

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 expansins are under-expressed in tuberous compared to fibrous roots, but none are statistically significant, so it seems unlikely that expansins play an important role in tuberous root formation in *Bomarea multiflora*. It is possible that expansin involvement in USO formation is unique to eudicots.

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 biosyn-

thesis 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 2.4). 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 neofunctionalizations of MADS-Box genes from other aspects of plant development, or if they form a clade of USO-specific copies.

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 2.4).

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 that 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 function is not utilized in the tuberous root formation for monocot taxa.

PEBP and FT-Like Gene Evolution in Geophytic Taxa

Gene tree analysis of PEBPs indicates that FT and TFL1 genes have been independently co-opted several times in 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 flower as a reproductive structure. 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. Subsequent specialization of these genes in the timing of shifts to reproductive development as well as USO development occurred. Given our results, in seems likely that USO-specialized FT and TFL1 genes arose at least four times independently, indicating broadly parallel molecular evolution underlies the convergent morphological evolution of USOs. Additional USO-specific PEBP genes would shed more light on this pattern, but the dearth of studies on the molecular basis of USO development impedes such analysis. With increased sampling, follow-up studies could identify unique patterns of convergent molecular evolution on the USO-specific FT genes. Do these copies share independently derived subsequences or motifs that could reflect or cause shared function?

In *Bomarea multiflora*, TFL1 involvement in USO formation is surprising, as neotropical cloud forests are generally considered relatively aseasonal environments. However, *B. multiflora* shows two annual peaks in flowering corresponding to two peaks in annual rainfall (Ortiz and Idárraga Piedrahita, 2011), suggesting that important aspects of the plant's phenology are tied to seasonal fluctuations. While no previous studies have looked at the timing of tuber production in *Bomarea*, it is possible that tuber development is also tied to seasonal cues.

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 many molecular processes are shared across geophytic taxa, suggesting that deep parallel evolution at the molecular level underlies the convergent evolution of an adaptive trait. In particular, we demonstrate that PEBP genes implicated in underground storage organ formation have been recruited multiple times across the gene tree, demonstrating that repeated morphological convergence is matched by repeated molecular convergence. 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.

Chapter 3

Data processing and plotting pipelines in R for Bayesian phylogenetic analyses with RevGadgets

3.1 Introduction

The rise of increasingly complex methods for modeling macroevolutionary processes and the availability of "big data" from genome-scale sequencing has driven the development of novel software tools for building and running customized phylogenetic models. For example, researchers can now integrate fossils into phylogenetic reconstruction and divergence time estimation (Heath et al., 2014), explicitly model the tempo of speciation and extinction through time (Stadler, 2011; Höhna, 2014), reconstruct the ancestral states of species' characteristics through increasingly complex models (Lewis, 2001; Goldberg and Igić, 2008; Beaulieu et al., 2013), and reconstruct species' ancestral ranges while accounting for time and geographic constraints on areas (Ree and Smith, 2008; Landis et al., 2018). However, development of these analytical methods has outpaced the development of tools to explore and visualize their data-rich outputs, especially when implemented in a Bayesian framework. This is particularly true for modular approaches that allow users to customize and combine methods, creating new models for which visualization tools may not exist. This deficit in processing tools has hampered the ability of empiricists to interpret and present their results but also to apply these methods to their data in the first place, as meaningful results hidden within the outputs of complex models can be intractable without specialized processing pipelines, and extensive visualize steps can be necessary to trouble-shoot and optimize analyses.

The release of `RevBayes`, a Bayesian phylogenetic modeling program, provides users with unprecedented control over model specifications through the modularity of probabilistic graphical models (Höhna et al., 2014, 2016). In `RevBayes`, researchers use graphical models to specify the underlying dependencies and correspondences of modular compo-

Table 3.1: Summary of the primary `RevGadgets` functions

Function	Description
readTrace()	reads and processes MCMC trace files
summarizeTrace()	summarizes the posterior distributions of parameters in a trace file
plotTrace()	plots the distributions and 95% CI/ CS of quantitative and qualitative parameters
readTrees()	reads and processes phylogenies
plotTree()	plots a phylogeny with options to label posterior probabilities, branch rates, and ages
plotFBDTree()	plots a fossilized-birth-death time tree with optional sampled ancestors
processAncStates()	processes annotated tree files for plotting ancestral states
plotAncStatesMAP()	plots the maximum *a posteriori* ancestral state
plotAncStatesPie()	plots ancestral states as pies
processDivRates()	processes diversification rate traces for plotting
plotDivRates()	plots diversification rates at discrete time intervals
processPostPredStats()	processes test statistics' distributions for plotting
plotPostPredStats()	plots the distribution of test statistics from simulated data and the empirical value
removeBurnin()	removes specified burnin from trace files
rerootPhylo()	reroots a phylogeny and organizes associated data

nents (similar to individual Legos being used to build a complex city) and to analyze these models in a Bayesian framework. This modularity allows users to design custom models and to tailor analyses to their particular research question, and facilitates the implementation of new models. These advantages, however, also motivate the development of novel tools to explore, diagnose, and visualize the results of `RevBayes` analyses. As with most implementations of Bayesian inference, `RevBayes` applies Markov chain Monte Carlo (MCMC), which results in oftentimes very large log files in non-standardized text formats that require custom processing to diagnose MCMC performance and estimate parameter values. Furthermore, `RevBayes` has implemented several methods that incorporate cladogenetic events into character evolution models—for example, `ClaSSE` (Goldberg, Emma E and Igić, Boris, 2012), DEC (Ree and Smith, 2008), ChromoSSE (Freyman and Höhna, 2017), and more—where change in a character state may occur precisely at the moment of speciation. Cladogenetic models present a challenge for visualizing trait evolution on phylogenies, as the character states immediately prior to and subsequent of speciation must be plotted separately; this feature is not currently offered by the primary methods for visualizing ancestral character reconstructions on phylogenies (*e.g.*, FigTree; Rambaut, 2014). One of the main advantages of `RevBayes` is the ability to build increasingly complex models, which increases the importance of closely examining model outputs, as users may encounter unexpected behavior of methods that are new or rarely used, with fewer clear expectations for standard behavior.

To address these challenges, I introduce a novel `R` package (R Core Team, 2013), `RevGadgets`, to provide data processing pipelines and visualization tools for the outputs of analyses performed in `RevBayes`. Through user-friendly data pipelines, `RevGadgets`

guides users through importing RevBayes output into R, processing the output, and producing figures or other summaries of the results. RevGadgets builds off of many of the existing tools for phylogenetic data processing and plotting in R, especially the ggtree and treeio packages (Wang et al., 2020; Yu et al., 2017). As the Rev language is quite similar to R, the interaction between RevBayes and RevGadgets will feel natural for most users. RevGadgets provides paired processing and plotting functions built around commonly implemented analyses, such as tree building, divergence-time estimation, diversification-rate estimation, ancestral-state reconstruction, biogeographic range reconstruction, and posterior predictive simulation. A description of the primary functions in RevGadgets is found in Table 3.1. Using the general framework of ggplot2, the tidyverse, and associated packages (Wickham, 2011; Wickham et al., 2019), plotting functions return plot objects with default aesthetics that users may customize. The primary goals for generating these functions in a paired processing and plotting approach are to prioritize user-friendly interfaces and increase accessibility to complex analyses while encouraging users to be thorough in examining and checking their outputs, and to produce novel solutions to existing challenges in phylognetic visualization. Here, the core RevGadgets functionality is introduced through five use cases that illustrate the paired processing and plotting approach for some of the most commonly performed analyses in RevBayes.

3.2 Visualizing MCMC Output

The output of most RevBayes analyses is a tab-delimited file where rows correspond to samples of an MCMC algorithm and columns correspond to parameters in the model. Most information of interest to researchers, such as the most probable parameter values and the 95% credible interval or set, require processing this raw MCMC output. Visualizing MCMC outputs also encourages users to interact with their data, and, in some cases, MCMC visualization may be used directly in publications to compare values of particular parameters, such as transition rates between character states. Perhaps most importantly, visualizing MCMC output is critical for evaluating and troubleshooting analyses, especially for diagnosing MCMC convergence and unexpected model pathologies. As such, methods for processing and visualizing MCMC output for both quantitative and qualitative parameters are presented.

The following code demonstrates how to read in the MCMC trace file of a general time-reversible (GTR) substitution model analysis (Tavaré, 1986), in which a user has estimated substitution rate and stationary frequency parameters for a single gene in a sample of 23 primates. (Springer et al., 2012).

First, the user reads in the MCMC trace file with the readTrace() function.

```
> file <- system.file("extdata", "sub_models/primates_cytb_covariotide.
    p", package="RevGadgets")
> trace_quant <- readTrace(path = file)
```

The user may remove Burnin (the samples taken before the Markov chain reached stationarity) at this stage or after examining the trace file further. The output of `readTrace()` may be passed to `R` packages specializing in MCMC diagnoses such as `Coda` (Plummer et al., 2006). `Coda` functions may then be used to calculate convergence diagnostics and show trace plots (Figure 3.1.

```
> trace_quant_MCMC <- coda::as.mcmc(trace_quant[[1]])
> coda::effectiveSize(trace_quant_MCMC)[c("pi[1]", "pi[2]","pi[3]","pi
    [4]")]

    pi[1]     pi[2]     pi[3]     pi[4]
2139.1946  587.9894 1500.8834  580.5132
> coda::traceplot(trace_quant_MCMC)
```

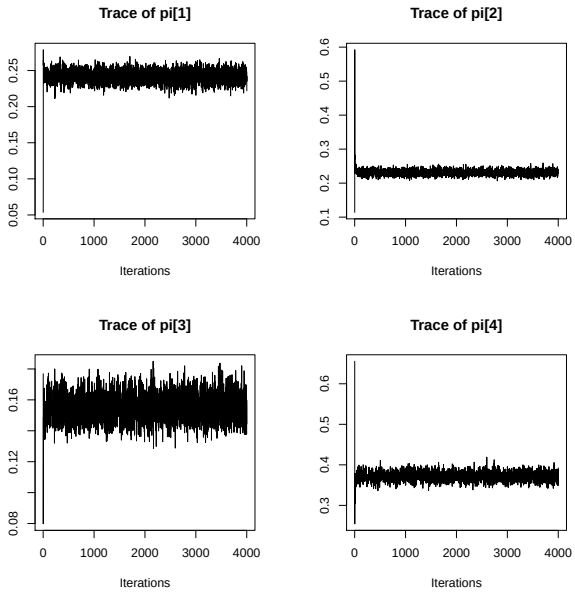

Figure 3.1: `Coda` trace plots of the four stationary frequency (π) parameters for the CYTB gene from 23 primates from (Springer et al., 2012).

In addition to its interactivity with `coda`, `RevGadgets` provides its own core functions for summarizing and visualizing traces of specific parameters. The `summarizeTrace()` function calculates the mean and 95% credible interval for quantitative variables and the 95% credible set for qualitative variables. For example, the user may plot stationary frequency (*pi*) parameter values and summarize their distributions from the trace file.

```
> summarizeTrace(trace = trace_quant, vars =  c("pi[1]","pi[2]","pi
    [3]","pi[4]"))

$`pi[1]`
```

```
$`pi[1]`$trace_1
          mean quantile_2.5 quantile_97.5
2.5% 0.3282636      0.30862      0.347765
...
$`pi[4]`
$`pi[4]`$trace_1
          mean quantile_2.5 quantile_97.5
2.5% 0.2146207     0.202042      0.227514
```

Base frequency parameters are quantitative variables, but similar functionality exists for qualitative variables, such as the binary character of if certain transition rates among character states exist, from a reversible jump MCMC (rjMCMC) ancestral-state reconstruction analysis.

```
> file <- system.file("extdata", "comp_method_disc/freeK_RJ.p", package
    = "RevGadgets")
> trace_qual <- readTrace(path = file)
> summarizeTrace(trace_qual, vars = c("prob_rate_12", "prob_rate_13", "
   prob_rate_21","prob_rate_23", "prob_rate_31", "prob_rate_32"))

$prob_rate_12
$prob_rate_12$trace_1
credible_set
        1         0
0.6440396 0.3559604
...
$prob_rate_32
$prob_rate_32$trace_1
        0
0.9724475
```

Users may visualize these results with `plotTrace()`, which generates density curves for quantitative parameters (Figure 3.2) and bar plots for qualitative parameters (Figure 3.3). In both cases, areas filled in with color indicate the 95% credible set or interval.

```
> plotTrace(trace = trace_quant, vars = c("pi[1]","pi[2]","pi[3]","pi
   [4]"))
> plotTrace(trace = trace_qual, vars = c("prob_rate_12", "prob_rate_13
   ", "prob_rate_31", "prob_rate_32"))
```

As with all `RevGadgets` plots, the output of `plotTrace()` is a ggplot-class object that can be modified by adding additional specifications. While the immediate output of all `RevGadgets` plotting functions are designed to be publication quality without further specification by the user, user may customize visual aesthetics such as color scheme using arguments in the function or by modifying the ggplot-class output (further examples provide below, in the other case studies).

Informative plots of the posterior distributions of parameter values are key to understanding the results of any Bayesian analysis; these tools encourage users to explore their results thoroughly rather than relying purely on single summary statistics. In addition, these products (summaries and plots) may be useful as tools for science communication

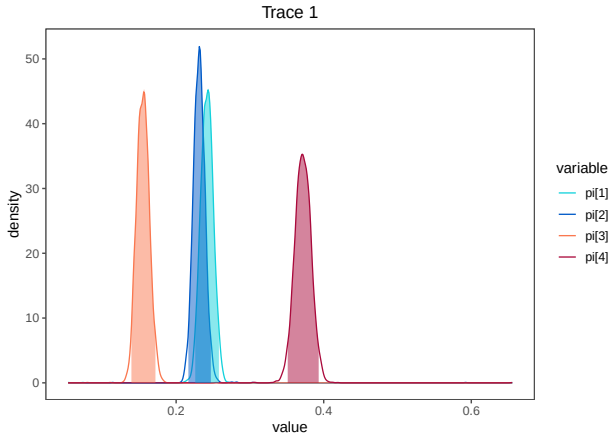

Figure 3.2: Output from plotTrace() for four stationary frequency (π) parameters from a GTR substitution model analysis. Colored areas under the curve indicate the 95% credible interval, showing that while π_3 and π_4 have statistically different distributions,π_1 and π_2 have highly overlapping distributions.

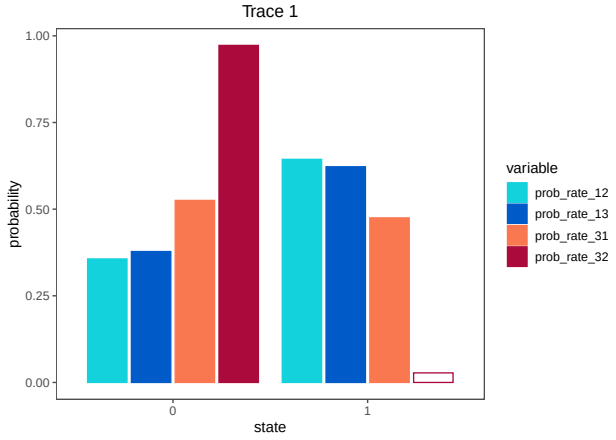

Figure 3.3: Output from plotTrace() for four binary qualitative variables from a reversible-jump analysis. Variables correspond to the probability that substitution rates should be included in the molecular evolution model. Colored areas within bars indicate credible set. For parameters other than the probability of rate 32, both state 0 and state 1 are in the credible set.

and education related to statistical phylogenetics, as they illustrate the output of Bayesian analyses and can easily be used to demonstrate differences in parameter distributions that result from changes to basic phylogenetic models.

3.3 Visualizing Phylogenies

Phylogenies are central to all analyses in `RevBayes`, and accurate and information-rich visualizations of evolutionary trees are thus critical. In this case study, I demonstrate the tree-plotting functionality of `RevGadgets` with methods developed to visualize phylogenies and their associated posterior probabilities, divergence time estimates, geological time scales, and branch rates. Additionally, text annotation may be added to specify any type of associated data, such as posterior probabilities of nodes or node ages. Users may modify aesthetics such as colors, sizes, branch thickness, and tip label formatting through specific function arguments or by adding layers to the resulting ggplot object. I demonstrate plotting discrete ancestral states, including biogeographic ranges, below in the ancestral-state reconstruction case study. Similar to the paired reading and plotting functions for MCMC traces described above, users plot trees through a reading and processing function `readTrees()` and a plotting function `plotTree()`. `RevBayes` saves trees as annotated nexus files (Maddison et al., 1997), and `RevGadgets` reads trees into `R` and processes them into treedata objects from `treeio` (Wang et al., 2020).

A special case of the `plotTree()` functionality allows for plotting fossilized birth-death (FBD) analyses, with `plotFBDTree()` (Figure 3.4); these analyses can infer fossilized ancestors (a fossil sample that is directly ancestral to another sample in the analysis and thus not on a branch of its own) so are not handled well by conventional plotting functions. In this case, a user has performed an FBD analysis of bears including fossil taxa and visualizes the resulting phylogeny to illustrate the phylogenetic relationships and divergence times among taxa. The user may include node-age bars colored by the node's posterior probability, a geological time scale and labeled epochs, and fossils estimated to be direct ancestors of other samples (sampled ancestors), as illustrated. The output of `plotFBDTree()`, like all `RevGadgets` plotting functions, is a ggplot-type object, so the users may further customize the plot using `ggplot2` or `ggtree` functions. This feature can be used to edit the legend position with `ggplot2::theme()`.

```
> file <- system.file("extdata", "fbd/bears.mcc.tre", package="
    RevGadgets")
> tree <- readTrees(paths = file)
> plot <- plotFBDTree(tree = tree, timeline = TRUE, tip_labels_italics
    = FALSE, node_age_bars = TRUE, node_age_bars_colored_by = "posterior
    ", node_age_bars_color = rev(RevGadgets:::.colFun(2))) + ggplot2::
    theme(legend.position=c(.25, .85))
```

The `plotTree()` function can also be used to visualize how quantitative variables vary across the phylogeny. Figure 3.5 demonstrates this functionality by visualizing the estimated optima of a relaxed Ornstein-Uhlenbeck process (Butler and King, 2004; Uyeda and Harmon, 2014), a model in which a quantitative character (such as body size), evolves towards shifting adaptive optima. In this example, a user visualizes the optima of body size evolution in surgeonfish.

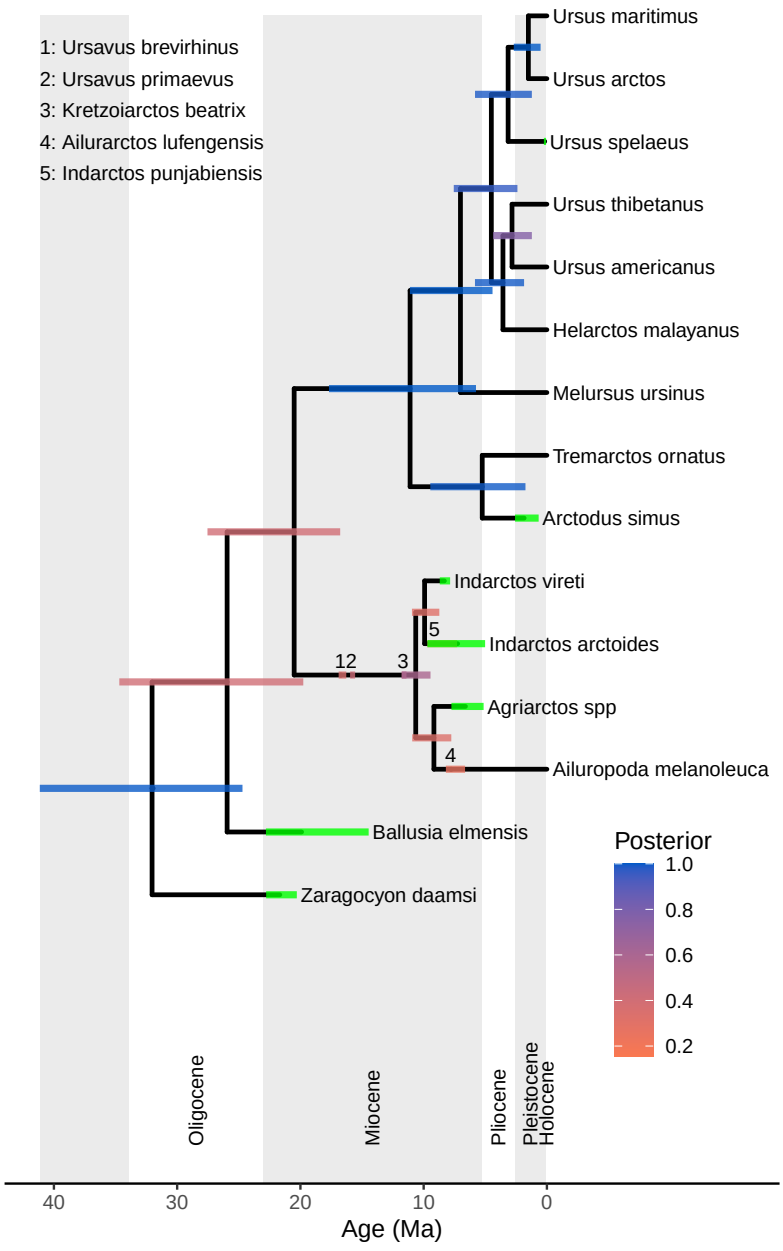

Figure 3.4: The output of a fossilized birth-death analysis may be plotted using `plotFBDTree()`. This phylogeny shows the evolutionary history of 20 bear species, including 12 fossil taxa, five of which are estimated to be direct ancestors of other taxa in the tree. Bars at the nodes show the 95% confidence interval for node age estimates and are colored according to the posterior probability of the node. Bars along branches correspond to sampled ancestors and are also colored based on the posterior probability of their placement on the tree. Green bars at the tips indicate age uncertainty for fossil tips. Data from Krause et al. (2008) and Abella et al. (2012).

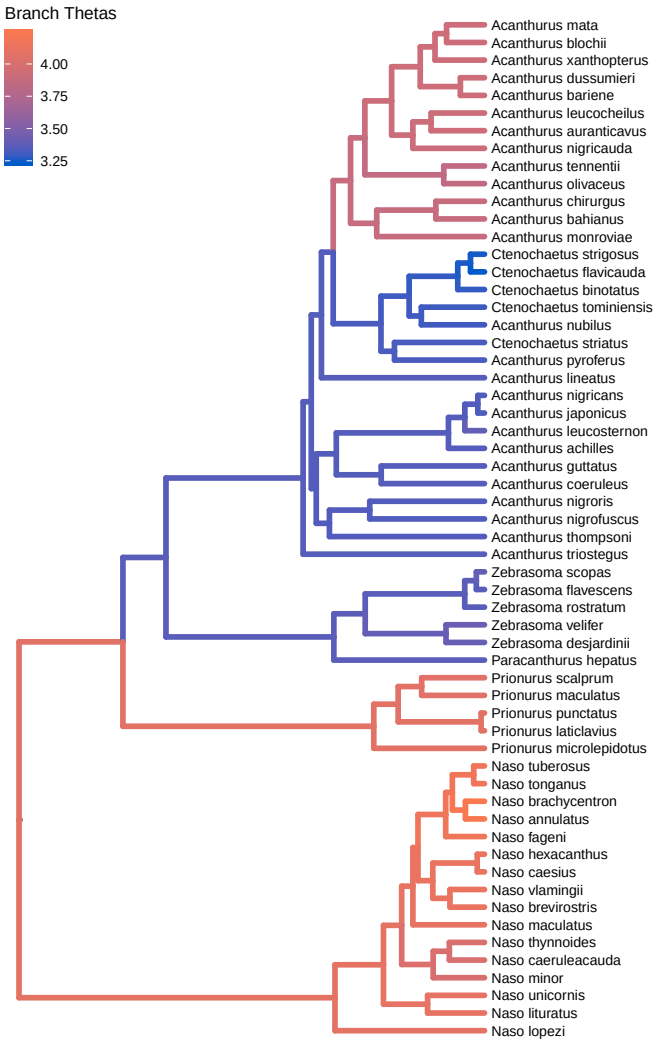

Figure 3.5: `plotTree()` can also plot the optima estimated from a relaxed Ornstein-Uhlenbeck model using plotTree(), in this case of body size evolution in surgeonfish (Acanthuridae). Branch color indicates the value of theta (optimum value) that corresponds to each branch, showing three clear optima across the phylogeny. Data from (Landis and Schraiber, 2017).

```
> file <- system.file("extdata", "relaxed_ou/relaxed_OU_MAP.tre",
    package="RevGadgets")
> tree <- readTrees(paths = file)
> plotTree(tree = tree, tip_labels_italics = FALSE, color_branch_by = "
    branch_thetas", line_width = 1.7) + ggplot2::theme(legend.position=c
    (.1, .9))
```

This code could also be used to visualize diversification rates or rates of molecular or morphological evolution across a phylogeny.

Together, these functions provide user-friendly and customizable tree-plotting functionality for a variety of core research questions in evolutionary biology. The `plotFBDTree()` function, in particular, represents newly developed methodology for processing and plotting taxa inferred to be direct ancestors of other samples in the phylogeny. As sampled ancestors can be expected under a wide range of evolutionary scenarios (Foote, 1996) and tree-building methods that include fossils become increasingly common (Loiseau et al., 2020; O'Reilly and Donoghue, 2020; Šmíd and Tolley, 2019; Hernandez-Gutierrez and Magallon, 2019; Gavryushkina et al., 2017; Zhang et al., 2016), these visualization tools will be instrumental for researchers to visualize all aspects of their results.

3.4 Ancestral-State Reconstruction

The estimation of phylogenetic trees is a core endeavor of statistical phylogenetics, but many researchers are more interested in the use of such trees to understand evolutionary dynamics through time, in particular the estimation of ancestral states and ancestral ranges, rather than in the inference of the trees themselves. These reconstruction methods allow users to model how traits, including species' distributions, have evolved across a phylogeny, producing probability distributions of states for each node of the phylogeny. The core tree plotting functions described above can be modified to visualize estimated ancestral states and ancestral ranges, in addition to annotations at the tips of the phylogeny. This aspect of `RevGadgets` functionality allows users to plot the maximum *a posteriori* (MAP) estimate of ancestral states via `plotAncStatesMAP()` or a pie chart showing the three most probable states via `plotAncStatesPie()`.

For visualization of the outputs of standard ancestral state reconstruction models, modelled ancestral states are plotted at the nodes. However, in cladogenetic models of ASR there are two ways that character states can be predicted to change on the phylogeny: shifts can occur along branches of the tree (anagenetic change) or can occur precisely at the moment of speciation (cladogenetic change; Ree and Smith, 2008). For cladogenetic models, classic visualizations of ancestral states are inadequate, as limiting ancestral-state visualizations to nodes prohibits the visualization of the actual timing of cladogenetic events. To remedy this problem `RevGadgets` plots the results of cladogenetic models on "shoulders" as well as the nodes. Cladogenetic models are becoming increasingly common, especially for biogeographic reconstructions, but currently no standard plotting tools for ancestral-state reconstructions allow for users to automatically or simultaneously visualize anagenetic and cladogenetic trait changes on a phylogeny. Given the increasing use of such methods, this functionality fills a gap in existing tools for statistical phylogenetic visualization. The ancestral-state plotting functions in `RevGadgets` allow users specify if a model is cladogenetic and to demarcate character states and their posterior probabilities by modifying the colors, shapes, and sizes of node and shoulder

symbols. Text annotations may be added to specify states, state posterior probabilities, and the posterior probabilities of nodes.

For example, a user may plot the result of an asymmetrical model for discrete ancestral-state reconstruction (Nylander et al., 2004) of placenta type in mammals. The `processAncStatesMAP()` function reads in, parses, and relabels an annotated summary tree, which was saved as a nexus-formatted file by `RevBayes`. Since discrete data are coded as integer values for analysis in `RevBayes`, users may provide state labels to produce an informative legend with the true character states (in this case, the types of placentas included in the ancestral-state reconstruction). The user employs the command `processAncStatesMAP()` to indicate the MAP state for each node with symbol color and the state's posterior probability by symbol size, as seen in Figure 3.6. Because of the size of the phylogeny, the user could choose to plot the reconstruction on a circular tree by changing the tree layout parameter.

```
> file <- system.file("extdata", "comp_method_disc/ase_freeK.tree",
    package="RevGadgets")
> example <- processAncStates(file, state_labels = c("1" = "
    Epitheliochorial", "2" = "Endotheliochorial", "3" = "Hemochorial"))
> plot <- plotAncStatesMAP(t = example, tree_layout = "circular") +
    ggplot2::theme(legend.position = c(0.57,0.41))
```

The `plotAncStatesMAP()` function also allows users to indicate the character state by the size of the symbol, useful for visualizing reconstructions with many ordered character states such as chromosome counts.

The `plotAncStatesPie()` function is a special case of `plotAncStatesMAP()` where the symbols at nodes are pie charts of the three most probable states for that node plus an "other" category of any remaining probability. For example, a user may employ this functionality to visualize the ancestral ranges of Hawaiian silverswords, estimated using a dispersal, extinction, and cladogenesis (DEC) biogeographic analysis and include shoulder states to indicate cladogenetic as well as anagenetic changes (see Figure 3.7). Because of the large number of states in this analysis (15 possible ranges and one "other" category), more pre-plotting processing is necessary. The user passes the appropriate ancestral area names to `processAncStates()` and specify custom colors in a named vector. To plot the ancestral states, the user provides the processed data, specifies that the data are "cladogenetic", adds text labels to the tips specifying the character state, and modifies sizes and horizontal positions for aesthetics. The user can also modify the legend position.

```
> file <- system.file("extdata", "dec/simple.ase.tre", package = "
    RevGadgets")
> labs <- c("1" = "K", "2" = "O", "3" = "M", "4" = "H", "5" = "KO", "6"
     = "KM", "7" = "OM", "8" = "KH", "9" = "OH", "10" = "MH", "11" = "
    KOM", "12" = "KOH", "13" = "KMH", "14" = "OMH", "15" = "KOMH")
> dec_example <- processAncStatesDiscrete(file, state_labels = labs)
> colors <- colorRampPalette(RevGadgets:::.colFun(12))(length(
    dec_example@state_labels))
```

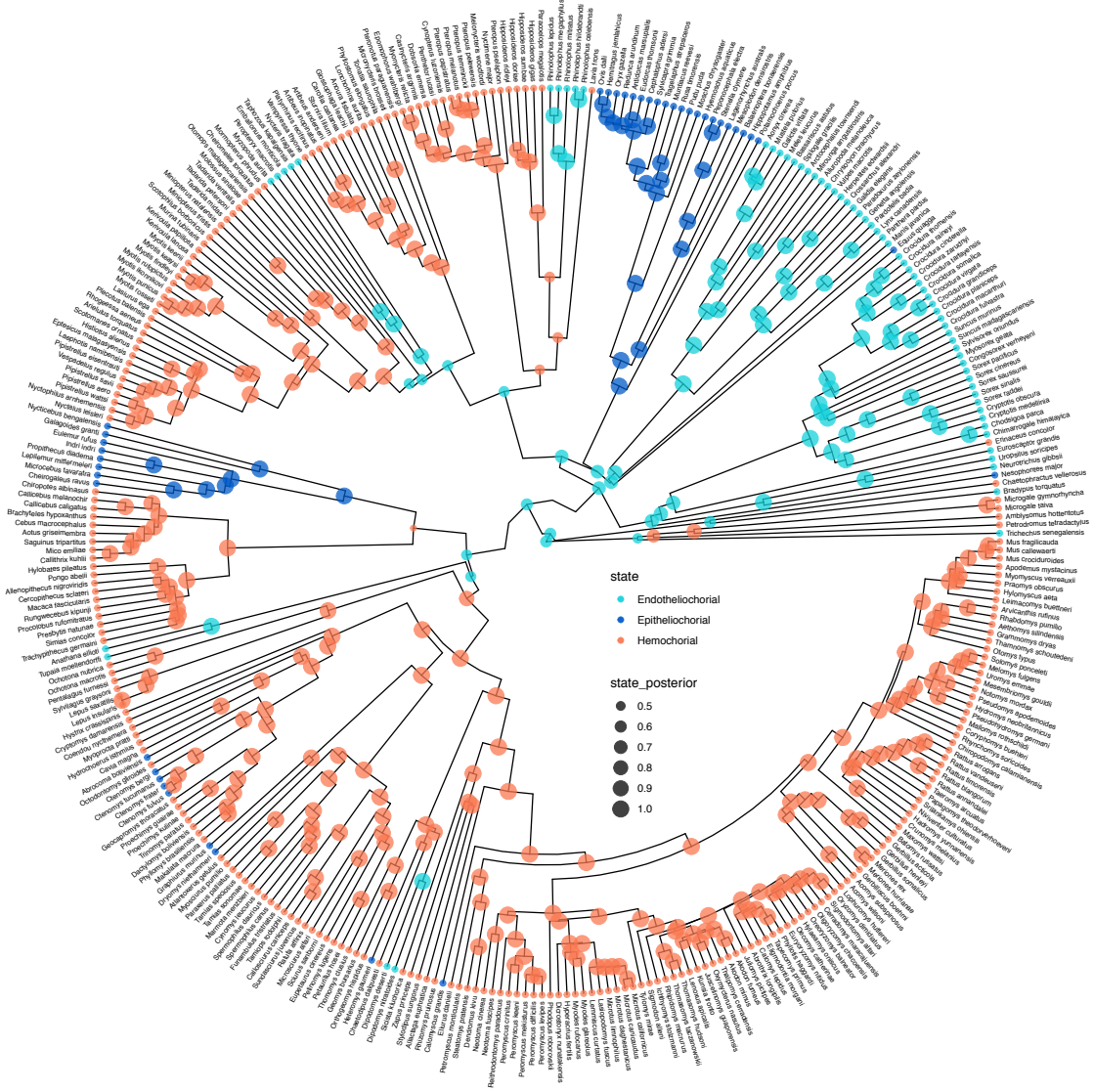

Figure 3.6: plotAncStatesMAP() plots the MAP states on the specified phylogeny, in this case used to visualize the results of an ancestral-state reconstruction of placenta type across mammals, under a model of anagenetic change (states plotted on nodes correspond to change that occured along the subtending branch). Color corresponds to the character state and size of the symbols corresponds to the posterior probability of that state. Data from (Elliot and Crespi, 2006; Faurby and Svenning, 2015).

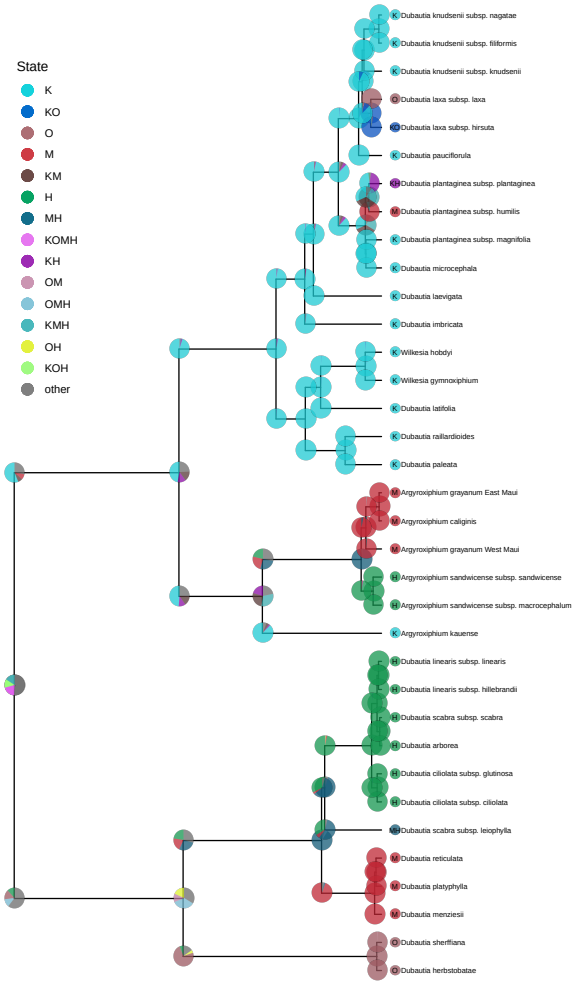

Figure 3.7: plotAncStatesPie() visualizes character states and their associated probabilities as pie charts at the nodes (and shoulders, for cladogenetic models) of the provided phylogeny. As an example, the estimated ancestral ranges of the Hawaiian silversword clade from a standard DEC (dispersal, extinction, and cladogenesis) model is shown. Pies at the shoulders represent the start state(s) of the branch, and pies at nodes represent the end state(s) such that change occurring along the branch corresponds to the difference between start and end states. Letters in state labels correspond to combinations of areas sampled from the model, where K = Kauai, O = Oahua, M = Maui Nui, and H = Hawaii Island. Data from (Landis et al., 2018).

```
> names(colors) <- dec_example@state_labels
> plot <- plotAncStatesPie(t = dec_example, cladogenetic = TRUE,
    tip_labels_states = TRUE, pie_colors = colors, tip_labels_offset =
    .2, tip_pie_nudge_x = -.15, node_pie_size = 1.2, tip_pie_size =
    0.12, tip_labels_states_offset = .05) + ggplot2::theme(legend.
    position = c(0.1, 0.75))
```

In addition to the above examples, these functions provide plotting tools for any discrete ancestral-state estimation including the results of chromosome count reconstructions and discrete state-dependent speciation and extinction (SSE) models.

3.5 Episodic Diversification Analysis

The rates at which diversification (speciation, extinction, and their difference) occurs are of great interest to evolutionary biologists (Morlon, 2014). Such rates may be modelled as constant (as in a constant-rate birth-death process (Kendall et al., 1948; Nee et al., 1994)), or allowed to vary over time (Stadler, 2011; May et al., 2016), across branches of a phylogeny (Rabosky, 2014; Höhna et al., 2019), or based on the character states of taxa (Maddison et al., 2007). For example, rates that vary across branches of the phylogeny may be visualized using `plotTree()` to color the branches by their inferred rate. State-dependent diversification models output a set of estimated ancestral-states and the estimated speciation and extinction rates corresponding to those states; `plotAncStatesMAP()` or `plotAncStatesPie()` may be used to visualize the ancestral-state estimates, and `plotTrace()` may be used to visualize the estimated rates. Here, I describe a method for plotting the result of a time-varying model, the episodic birth-death process (Stadler, 2011; Höhna, Sebastian, 2015), using `processDivRates()` and `plotDivRates()`. These functions process the MCMC trace files, estimate the mean and 95% confidence interval of rates for discrete periods of time, and visualize how these rates vary over time. The episodic birth-death process model in `RevBayes` produces separate trace files for times and rates for each type of rate (currently, speciation, extinction, and fossilization rates are supported). For example, a user may visualize a time-variable diversification analysis of primate evolution (see Figure 3.8). MCMC log files from an episodic birth-death process model are read into `R` using `processDivRates()`:

```
> speciation_time_file <- system.file("extdata", "epi_bd/
    primates_EBD_speciation_times.p", package = "RevGadgets")
> speciation_rate_file <- system.file("extdata", "epi_bd/
    primates_EBD_speciation_rates.p", package = "RevGadgets")
> extinction_time_file <- system.file("extdata", "epi_bd/
    primates_EBD_extinction_times.p",  package = "RevGadgets")
> extinction_rate_file <- system.file("extdata", "epi_bd/
    primates_EBD_extinction_rates.p", package = "RevGadgets")
> rates <- processDivRates(speciation_time_log = speciation_time_file,
    speciation_rate_log = speciation_rate_file, extinction_time_log =
```

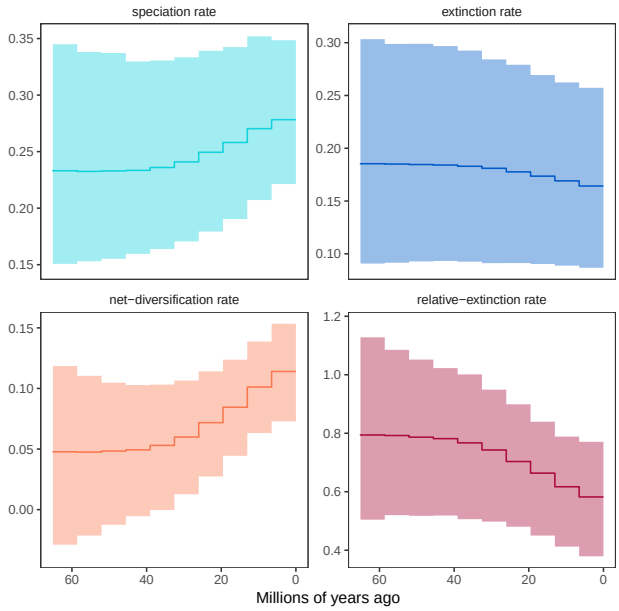

Figure 3.8: How speciation, extinction, net-diversification, and relative-extinction rates change through time, as estimated in an episodic birth-death model of primate diversification. The colored line represents the average estimate through time, and the shaded area around the line represents the 95% credible interval of those estimates. The speciation rate tends to increase over time, and extinction rate to decrease, so the derived metrics net-diversification and relative-extinction increase and decrease, respectively. Data from (Springer et al., 2012).

```
extinction_time_file, extinction_rate_log = extinction_rate_file,
burnin = 0.25)
```

The `processDivRates()` function automatically calculates net diversification (speciation - extinction) and relative extinction (extinction/speciation) rates from the speciation and extinction trace files. The user can plot these rates with `plotDivRates()` and modify the output using `ggplot2()`, shown in Figure 3.8:

```
> plotDivRates(rates = rates) + ggplot2::xlab("Millions of years ago")
```

Users may choose to plot only some of these rates by providing the plotting script with a subset of the processed data. Further customization of aesthetics such as colors may be achieved by modifying the output using `ggplot2()`.

Together with the above-described functions for plotting rate parameter distributions, `plotDivRates()` allows users to visualize the outputs of all diversification analyses available in `RevBayes`. As the analysis of diversification rates has become an increasingly important aspect of statistical phylogenetics, these visualization tools will allow researchers to explore the results of their models and perhaps open the door to the development and exploration of new analytical methods.

3.6 Posterior-Predictive Analysis

In addition to visualizing the results of phylogenetic models, `RevGadgets` provides tools for examining the adequacy of models used for analyzing data, an application that can be even more important than visualizing the results themselves (Uyeda et al., 2018; Pennell et al., 2015). Posterior-predictive analysis is a method for evaluating model adequacy that allows users to test if their model predicts a set of values that include the observed, empirical data (Bollback, 2002; Brown, 2014; Höhna et al., 2018). In other words, does simulating under the inferred parameter values from an analysis yield datasets that are consistent with the original input data? The analysis is essentially run backwards. First, users analyze their data with the model of interest and then use the inferred parameter values from that analysis to simulate (under the same model) a number of new data sets. The user then selects test statistics or metrics to describe the data and calculates those metrics from both the original, empirical dataset and the simulated datasets. If the metric from the empirical data is reasonably included within the distribution of metrics from simulated datasets, the model is considered adequate. `RevGadgets` provides functionality to visualize the comparison of the simulated data to the empirical data for any number of user-designed metrics through a processing function `processPostPredStats()` and a plotting function `plotPostPredStats()`. For example, a user may employ this workflow for a posterior-predictive simulation to test the adequacy of the Jukes-Cantor model for nucleotide sequence evolution (Jukes et al., 1969) in a single gene across a sample of 23 primates. The user can read in .csv files generated by `RevBayes` containing the calculated metrics from the simulated datasets and the empirical data. The `plotPostPredStats()` function produces a list of ggplot objects (one for each metric); Figure 3.9 shows a subset of these plots.

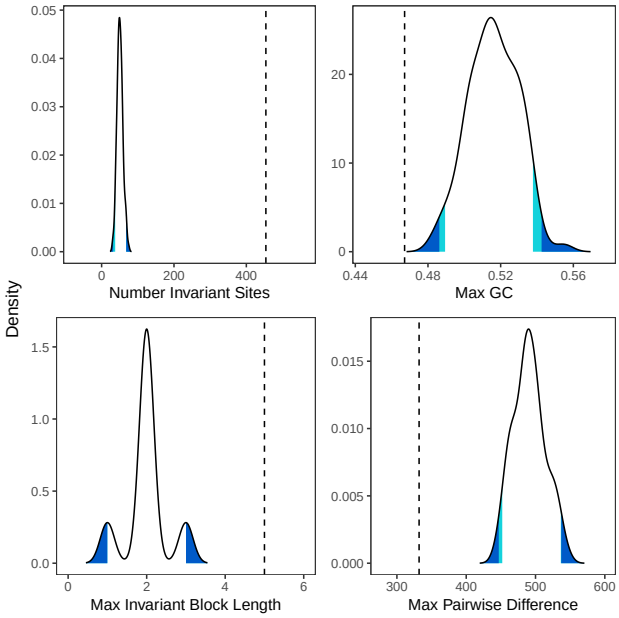

Figure 3.9: Four metrics from a posterior-predictive analysis of the Jukes-Cantor model for nucleotide sequence data. Light blue shaded areas under the curve correspond to the 5% quantile and dark blue areas correspond to the 2% quantile. The dashed line shows the value from empirical data. For all metrics shown, the model is inadequate. Data from Springer et al. (2012).

```
> file_sim <- system.file("extdata", "PPS/simulated_data_pps_example.
    csv", package="RevGadgets")
> file_emp <- system.file("extdata", "PPS/empirical_data_pps_example.
    csv", package="RevGadgets")
> t <- processPostPredStats(path_sim = file_sim, path_emp = file_emp)
> plots <- plotPostPredStats(data = t)
```

As increasingly complex statistical methods become more accessible to researchers, model adequacy analyses will be critical to testing how well various models describe generative processes. Additionally, the growing size of datasets available for phylogenetic reconstruction (genome-scale data) heightens the risk for biases in the data, such as base composition biases, to manifest as systematic errors in the results (Phillips et al., 2004). Posterior predictive analyses can help reveal how systematic errors in large datasets and resulting incongruity with pre-established models yields inaccurate parameter estimates. However, posterior predictive approaches remain relatively uncommon in empirical studies—this component of `RevGadgets` functionality and the associated clear workflows for performing and interpreting posterior predictive tests will hopefully increase the application of this important tool.

3.7 Summary

The case studies described above illustrate some of the core functionality available in `RevGadgets` and cover most methodologies used by empiricists. `RevGadgets` relies on established data structures from `ggtree` and `treeio`, so many of the plotting functions may be used for non-`RevBayes` analyses if the results are imported to R and converted to `treedata` objects. In addition to available tools, `RevGadgets` provides a framework and supporting infrastructure for future development. In particular, development is ongoing of methods for MCMC diagnosis and visualization of posterior distributions of trees and more nuanced and complicated visualizations of posterior predictive simulations. As new methods continue to be developed in `RevBayes`, `RevGadgets` will continue to provide pipelines for processing and visualizing the results. Updated releases of `RevGadgets` will be available on GitHub and advertised on the `RevBayes` website.

Bibliography

Abella, J., D. M. Alba, J. M. Robles, A. Valenciano, C. Rotgers, R. Carmona, P. Montoya, and J. Morales, 2012. *Kretzoiarctos* gen. nov., the oldest member of the giant panda clade. PLoS One 7.

Adler, P. B., R. Salguero-Gómez, A. Compagnoni, J. S. Hsu, J. Ray-Mukherjee, C. Mbeau-Ache, and M. Franco, 2014. Functional traits explain variation in plant life history strategies. Proceedings of the National Academy of Sciences 111:740–745.

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

Beaulieu, J. M., D.-C. Jhwueng, C. Boettiger, and B. C. O'Meara, 2012. Modeling stabilizing selection: expanding the ornstein–uhlenbeck model of adaptive evolution. Evolution: International Journal of Organic Evolution 66:2369–2383.

Beaulieu, J. M., B. C. O'Meara, and M. J. Donoghue, 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Systematic Biology 62:725–737.

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

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt, and E. W. Sayers, 2018. Genbank. Nucleic acids research 46:D41–D47.

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

Bollback, J. P., 2002. Bayesian model adequacy and choice in phylogenetics. Molecular Biology and Evolution 19:1171–1180.

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

Brown, J. M., 2014. Predictive approaches to assessing the fit of evolutionary models. Systematic Biology 63:289–292.

Butler, M. A. and A. A. King, 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. The American Naturalist 164:683–695.

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

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

Chabannes, M., A. Barakate, C. Lapierre, J. M. Marita, J. Ralph, M. Pean, S. Danoun, C. Halpin, J. Grima-Pettenati, and A. M. Boudet, 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 28:257–270.

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

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

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

Cuéllar-Martínez, M. and V. Sosa, 2016. Diversity patterns of monocotiledonous geophytes in Mexico. Botanical Sciences 94:699.

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

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

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

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

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

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

Elliot, M. G. and B. J. Crespi, 2006. Placental invasiveness mediates the evolution of hybrid inviability in mammals. The American Naturalist 168:114–120.

Enquist, B. J., G. B. West, E. L. Charnov, and J. H. Brown, 1999. Allometric scaling of production and life-history variation in vascular plants. Nature 401:907–911.

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

Faurby, S. and J.-C. Svenning, 2015. A species-level phylogeny of all extant and late Quaternary extinct mammals using a novel heuristic-hierarchical Bayesian approach. Molecular phylogenetics and evolution 84:14–26.

Felsenstein, J., 1985. Phylogenies and the comparative method. The American Naturalist 125:1–15.

Feng, X., D. S. Park, Y. Liang, R. Pandey, and M. Papeş, 2019. Collinearity in ecological niche modeling: Confusions and challenges. Ecology and evolution 9:10365–10376.

Fick, S. E. and R. J. Hijmans, 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology 37:4302–4315.

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

Flemons, P., R. Guralnick, J. Krieger, A. Ranipeta, and D. Neufeld, 2007. A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). Ecological informatics 2:49–60.

Foote, M., 1996. On the probability of ancestors in the fossil record. Paleobiology Pp. 141–151.

Freyman, W. A., 2015. SUMAC: Constructing phylogenetic supermatrices and assessing partially decisive taxon coverage. Evolutionary Bioinformatics 11:EBO–S35384.

Freyman, W. A. and S. Höhna, 2017. Cladogenetic and anagenetic models of chromosome number evolution: A Bayesian model averaging approach. Systematic Biology 67:195–215. URL `https://doi.org/10.1093/sysbio/syx065`.

Garland Jr, T., A. W. Dickerman, C. M. Janis, and J. A. Jones, 1993. Phylogenetic analysis of covariance by computer simulation. Systematic Biology 42:265–292.

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

Gavryushkina, A., T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond, 2017. Bayesian total-evidence dating reveals the recent crown radiation of penguins. Systematic Biology 66:57–73.

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

Givnish, T. J., A. Zuluaga, I. Marques, V. K. Lam, M. S. Gomez, W. J. Iles, M. Ames, D. Spalink, J. R. Moeller, B. G. Briggs, et al., 2016. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. Cladistics 32:581–605.

Goldberg, E. E. and B. Igić, 2008. On phylogenetic tests of irreversible evolution. Evolution: International Journal of Organic Evolution 62:2727–2741.

Goldberg, Emma E and Igić, Boris, 2012. Tempo and mode in plant breeding system evolution. Evolution: International Journal of Organic Evolution 66:3701–3709.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29:644–652.

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

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

Hansen, T. F., 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

He, Y., J. Wu, B. Lv, J. Li, Z. Gao, W. Xu, F. Baluška, W. Shi, P. C. Shaw, and J. Zhang, 2015. Involvement of 14-3-3 protein GRF9 in root growth and response under polyethylene glycol-induced water stress. Journal of experimental botany 66:2271–2281.

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

Heath, T. A., J. P. Huelsenbeck, and T. Stadler, 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences 111:E2957–E2966.

Hernandez-Gutierrez, R. and S. Magallon, 2019. The timing of Malvales evolution: Incorporating its extensive fossil record to inform about lineage diversification. Molecular phylogenetics and evolution 140:106606.

Hoffmann, A., F. Liberona, and A. Hoffmann, 1998. Distribution and ecology of geophytes in Chile. Conservation threats to geophytes in Mediterranean-type regions. Pp. 231–253, *in* Landscape disturbance and biodiversity in Mediterranean-type ecosystems. Springer.

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

Höhna, S., 2014. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. PLoS one 9.

Höhna, S., L. M. Coghill, G. G. Mount, R. C. Thomson, and J. M. Brown, 2018. P3: Phylogenetic posterior prediction in RevBayes. Molecular biology and evolution 35:1028–1034.

Höhna, S., W. A. Freyman, Z. Nolen, J. Huelsenbeck, M. R. May, and B. R. Moore, 2019. A Bayesian approach for estimating branch-specific speciation and extinction Rates. bioRxiv P. 555805.

Höhna, S., T. A. Heath, B. Boussau, M. J. Landis, F. Ronquist, and J. P. Huelsenbeck, 2014. Probabilistic graphical model representation in phylogenetics. Systematic biology 63:753–771.

Höhna, S., M. J. Landis, T. A. Heath, B. Boussau, N. Lartillot, B. R. Moore, J. P. Huelsenbeck, and F. Ronquist, 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Systematic Biology 65:726–736.

Höhna, Sebastian, 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. Journal of theoretical biology 380:321–331.

Holt, R. D., 2009. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. Proceedings of the National Academy of Sciences 106:19659–19665.

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

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

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

Iles, W. J., S. Y. Smith, M. A. Gandolfo, and S. W. Graham, 2015. Monocot fossils suitable for molecular dating analyses. Botanical Journal of the Linnean Society 178:346–374.

Janzen, D. H. et al., 1975. Ecology of plants in the tropics .

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

Jukes, T. H., C. R. Cantor, et al., 1969. Evolution of protein molecules. Mammalian protein metabolism 3:132.

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

Kamenetsky, R. and H. Okubo, 2012. Ornamental geophytes: from basic science to sustainable production. CRC press.

Katoh, K. and D. M. Standley, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30:772–780.

Kendall, D. G. et al., 1948. On the generalized "birth-and-death" process. The annals of mathematical statistics 19:1–15.

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

Krause, J., T. Unger, A. Noçon, A.-S. Malaspinas, S.-O. Kolokotronis, M. Stiller, L. Soibelzon, H. Spriggs, P. H. Dear, A. W. Briggs, et al., 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. BMC Evolutionary Biology 8:220.

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

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

Landis, M. J., W. A. Freyman, and B. G. Baldwin, 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. Evolution 72:2343–2359.

Landis, M. J. and J. G. Schraiber, 2017. Pulsed evolution shaped modern vertebrate body sizes. Proceedings of the National Academy of Sciences 114:13224–13229.

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

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

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

Leebens-Mack, J., pers. comm. personal communication.

Leeggangers, H. A., H. Nijveen, J. N. Bigas, H. W. Hilhorst, and R. G. Immink, 2017. Molecular Regulation of Temperature-Dependent Floral Induction in *Tulipa gesneriana*. Plant Physiology 173:1904–1919. URL `http://www.plantphysiol.org/content/173/3/1904`.

Lewis, P. O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic biology 50:913–925.

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

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

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

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

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

Liu, Y.-Y., K.-Z. Yang, X.-X. Wei, and X.-Q. Wang, 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 212:730–744.

Loiseau, O., A. Weigand, S. Noben, J. Rolland, D. Silvestro, M. Kessler, M. Lehnert, and N. Salamin, 2020. Slowly but surely: gradual diversification and phenotypic evolution in the hyper-diverse tree fern family Cyatheaceae. Annals of Botany 125:93–103.

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

Maddison, D. R., D. L. Swofford, and W. P. Maddison, 1997. NEXUS: an extensible file format for systematic information. Systematic biology 46:590–621.

Maddison, W. P., P. E. Midford, and S. P. Otto, 2007. Estimating a binary character's effect on speciation and extinction. Systematic biology 56:701–710.

Maitner, B. S., B. Boyle, N. Casler, R. Condit, J. Donoghue, S. M. Durán, D. Guaderrama, C. E. Hinchliff, P. M. Jørgensen, N. J. Kraft, et al., 2018. The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods in Ecology and Evolution 9:373–379.

May, M. R., S. Höhna, and B. R. Moore, 2016. A bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. Methods in Ecology and Evolution 7:947–959.

Mayfield, J. D., A.-L. Paul, and R. J. Ferl, 2012. The 14-3-3 proteins of arabidopsis regulate root growth and chloroplast development as components of the photosensory system. Journal of experimental botany 63:3061–3070.

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

Miller, M. A., W. Pfeiffer, and T. Schwartz, 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees Pp. 1–8.

Mirarab, S., N. Nguyen, S. Guo, L.-S. Wang, J. Kim, and T. Warnow, 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. Journal of Computational Biology 22:377–386.

Morlon, H., 2014. Phylogenetic approaches for studying diversification. Ecology letters 17:508–525.

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

Nee, S., R. M. May, and P. H. Harvey, 1994. The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 344:305–311.

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

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

Noh, S. A., H.-S. Lee, E. J. Huh, G. H. Huh, K.-H. Paek, J. S. Shin, and J. M. Bae, 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.

Nylander, J. A., F. Ronquist, J. P. Huelsenbeck, and J. Nieves-Aldrey, 2004. Bayesian phylogenetic analysis of combined data. Systematic biology 53:47–67.

Oppenheim, S. J., R. H. Baker, S. Simon, and R. DeSalle, 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, R. d. C. and Á. Idárraga Piedrahita, 2011. Catálogo de las Plantas Vasculares del Departamento de Antioquia. Tropicos, Botanical Information System at the MIssouri Botanical Garden. Accessed: 2019-10-01.

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

O'Reilly, J. E. and P. C. Donoghue, 2020. The effect of fossil sampling on the estimation of divergence times with the fossilized birth–death process. Systematic biology 69:124–138.

Paradis, E. and K. Schliep, 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Parsons, R. and S. D. Hopper, 2003. Monocotyledonous geophytes: comparison of south-western Australia with other areas of mediterranean climate. Australian Journal of Botany 51:129–133.

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

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

Patterson, T. B. and T. J. Givnish, 2002. Phylogeny, concerted convergence, and phylogenetic niche conservatism in the core Liliales: insights from rbcL and ndhF sequence data. Evolution 56:233–252.

Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon, 2015. Model adequacy and the macroevolution of angiosperm functional traits. The American Naturalist 186:E33–E50.

Phillips, M. J., F. Delsuc, and D. Penny, 2004. Genome-scale phylogeny and the detection of systematic biases. Molecular biology and evolution 21:1455–1458.

Phillips, S. J. and M. Dudík, 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175.

Plummer, M., N. Best, K. Cowles, and K. Vines, 2006. CODA: convergence diagnosis and output analysis for MCMC. R news 6:7–11.

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

Proches, S., R. M. Cowling, P. Goldblatt, J. C. Manning, and D. A. Snijman, 2006. An overview of the Cape geophytes. Biological Journal of the Linnean Society 87:27–43.

de Queiroz, A. and J. Gatesy, 2007. The supermatrix approach to systematics. Trends in ecology & evolution 22:34–41.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Rabosky, D. L., 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PloS one 9.

Raíces, M., P. R. Gargantini, D. Chinchilla, M. Crespi, M. T. Téllez-Iñón, and R. M. Ulloa, 2003. Regulation of CDPK isoforms during tuber development. Plant molecular biology 52:1011–1024.

Rambaut, A., 2014. FigTree 1.4. 2 software. Institute of Evolutionary Biology, Univ. Edinburgh .

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. .

Ree, R. H. and S. A. Smith, 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic biology 57:4–14.

Rees, A., 1989. Evolution of the geophytic habit and its physiological advantages. Herbertia 45:104–110.

Revell, L. J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in ecology and evolution 3:217–223.

Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck, 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology 61:539–542.

Rundel, P. W., 1996. Monocotylendous geophytes in the California flora. Madroño 43:355–368.

Sanso, A. M. and C. C. Xifreda, 2001. Generic delimitation between *Alstroemeria* and *Bomarea* (Alstroemeriaceae). Annals of Botany 88:1057–1069.

Šmíd, J. and K. A. Tolley, 2019. Calibrating the tree of vipers under the fossilized birth-death model. Scientific reports 9:1–10.

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

Sosa, V., K. M. Cameron, D. F. Angulo, and T. Hernández-Hernández, 2016. Life form evolution in epidendroid orchids: ecological consequences of the shift from epiphytism to terrestrial habit in *Hexalectris*. Taxon 65:235–248.

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

Springer, M. S., R. W. Meredith, J. Gatesy, C. A. Emerling, J. Park, D. L. Rabosky, T. Stadler, C. Steiner, O. A. Ryder, J. E. Janečka, et al., 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PloS one 7.

Stadler, T., 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proceedings of the National Academy of Sciences 108:6187–6192.

Stevens, P. F. et al., 2016. Angiosperm Phylogeny Website. Version 13. Angiosperm Phylogeny Website. Version 13. .

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

Tanaka, M., N. Kato, H. Nakayama, M. Nakatani, and Y. Takahata, 2008. Expression of class I knotted1-like homeobox genes in the storage roots of sweetpotato (*Ipomoea batatas*). Journal of plant physiology 165:1726–1735.

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

Tarasov, S., I. Mikó, M. J. Yoder, and J. C. Uyeda, 2019. PARAMO: A pipeline for reconstructing ancestral anatomies using ontologies and Stochastic mapping. Insect Systematics and Diversity 3:1.

Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on mathematics in the life sciences 17:57–86.

Tribble, C. M., J. Martínez-Gómez, C. C. Howard, J. Males, V. Sosa, E. Sessa, N. Cellinese, and C. D. Specht, in review. Get the shovel: evolutionary complexities of belowground organs in geophytes.

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

Uyeda, J. C. and L. J. Harmon, 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Systematic biology 63:902–918.

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell, 2018. Rethinking phylogenetic comparative methods. Systematic Biology 67:1091–1109.

Vaidya, G., D. J. Lohman, and R. Meier, 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27:171–180.

Van Kleeff, P., N. Jaspert, K. Li, S. Rauch, C. Oecking, and A. De Boer, 2014. Higher order *Arabidopsis* 14-3-3 mutants show 14-3-3 involvement in primary root growth both under control and abiotic stress conditions. Journal of experimental botany 65:5877–5888.

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

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

Wang, L.-G., T. T.-Y. Lam, S. Xu, Z. Dai, L. Zhou, T. Feng, P. Guo, C. W. Dunn, B. R. Jones, T. Bradley, et al., 2020. treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Molecular biology and evolution 37:599–603.

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

WCSP, 2020. World checklist of selected plant families, facilitated by the Royal Botanic Gardens, Kew. `http://wcsp.science.kew.org`. Accessed: 2017-09-04.

Whigham, D. F., 2004. Ecology of woodland herbs in temperate deciduous forests. Annu. Rev. Ecol. Evol. Syst. 35:583–621.

Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3:180–185.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, et al., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4:1686.

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

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

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

Yockteng, R., A. M. R. Almeida, S. Yee, T. Andre, C. Hill, and C. D. Specht, 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.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam, 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8:28–36.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist, 2016. Total-evidence dating under the fossilized birth-death process. Systematic Biology 65:228–249.

Appendix A

Phylogenetic systematics of *Bomarea* Mirb.

To bridge the macro and microevolutionary scales of Chapters 1 and 2, respectively, I originally planned to include a phylogeny of *Bomarea* Mirb., a fascinating clade of climbing Liliid geophytes with both rhizomes and root tubers, with an emphasis on the ethnobotanically important *Bomarea edulis*. Most species of *Bomarea* are climbing tropical vines with rhizomes and rotund root tubers, the latter being one of the types of underground storage organs (USOs) that emerged as of particular interest in Chapter 1. In Chapter 2, I investigate the genes involved in the development of these root tubers in one species: *Bomarea multiflora*. The planned *Bomarea* phylogeny would have provided a fundamental base of knowledge of relationships within the genus to contextualize Chapter 2 and lay the groundwork for a more detailed follow-up to the patterns observed in Chapter 1.

Samples for the phylogeny include silica dried specimens from fieldwork I conducted during my dissertation, samples contributed by collaborators, and herbarium samples that I collected from five major US herbaria (US, F, MO, NY, UC) and one international herbarium (MEX). Together, the samples include 19 outgroups from other genera in Alstroemeriaceae, 113 of the 120 species of *Bomarea*, and 47 individuals of *Bomarea edulis*, sampled from across the wide range of the species (see Table A.1. These samples will provide the most comprehensive picture of evolutionary relationships within the genus available, as well as an initial examination of the population-level relationships and possible (lack of) monophyly of *Bomarea edulis*.

However, due to the coronavirus pandemic, the samples are just now being processed by the sequencing facility, which has pivoted to provide viral sequencing and vaccine detection services since the outbreak began. While this project cannot be included in the formal dissertation, it certainly has been a fundamental part of my graduate training and will be a central element of my research program moving forward. My observations in the field have hinted at a wide variety of undescribed morphological variation in the root tubers of *Bomarea*. In particular, I observed smaller, string-of-pearls like tubers in the highest elevation species. This variation is especially interesting in light of the findings

of Chapter 1, which indicate that root tuber morphology is associated with differences in temperature seasonality in Liliales. Perhaps finer scale differences in root tuber morphology also affect certain aspects of climatic niche. The *Bomarea* phylogeny will provide the evolutionary framework to explore such relationships in future work.

Table A.1: List of species included in the ongoing *Bomarea* phylogeny project

Alstroemeria apertiflora	*Bomarea campanularia*	*Bomarea graminifolia*	*Bomarea pumila*
Alstroemeria aurea	*Bomarea campylophylla*	*Bomarea hartwegii*	*Bomarea puracensis*
Alstroemeria crispata	*Bomarea carderi*	*Bomarea herbertiana*	*Bomarea purpurea*
Alstroemeria haemantha	*Bomarea caucana*	*Bomarea hieronymi*	*Bomarea rosea*
Alstroemeria inodora	*Bomarea caudata*	*Bomarea hirsuta*	*Bomarea salsilla*
Alstroemeria isabellana	*Bomarea caudatisepala*	*Bomarea huanuco*	*Bomarea sclerophylla*
Alstroemeria ligtu	*Bomarea ceratophora*	*Bomarea involucrosa*	*Bomarea secundifolia*
Alstroemeria nervosa	*Bomarea cf. anceps*	*Bomarea killipii*	*Bomarea setacea*
Alstroemeria pallida	*Bomarea chaparensis*	*Bomarea lancifolia*	*Bomarea* sp. 'catanata soya'
Alstroemeria presliana	*Bomarea chimborazensis*	*Bomarea lehmannii*	*Bomarea* sp. 'enano rojo'
Alstroemeria pulchra	*Bomarea chiriquina*	*Bomarea libertadensis*	*Bomarea* sp. 'enano verde'
Alstroemeria pygmaea	*Bomarea coccinea*	*Bomarea linifolia*	*Bomarea* sp. 'oso'
Alstroemeria revoluta	*Bomarea colombiana*	*Bomarea longipes*	*Bomarea* sp. 'ponillal soya'
Alstroemeria stenosepala	*Bomarea cordifolia*	*Bomarea longistyla*	*Bomarea speciosa*
Bomarea acuminata	*Bomarea cornigera*	*Bomarea lopezii*	*Bomarea straminea*
Bomarea acutifolia	*Bomarea cornuta*	*Bomarea lutea*	*Bomarea suberecta*
Bomarea aff. cruenta	*Bomarea costaricensis*	*Bomarea macrocephala*	*Bomarea superba*
Bomarea albimontana	*Bomarea crassifolia*	*Bomarea macusani*	*Bomarea tarmensis*
Bomarea allenii	*Bomarea crocea*	*Bomarea moritziana*	*Bomarea torta*
Bomarea alstroemeriodes	*Bomarea densiflora*	*Bomarea multiflora*	*Bomarea tribachiata*
Bomarea amazonica	*Bomarea denticulata*	*Bomarea multipes*	*Bomarea trichophylla*
Bomarea amilcariana	*Bomarea diffracta*	*Bomarea nematocaulon*	*Bomarea trimorphophylla*
Bomarea ampayesana	*Bomarea dispar*	*Bomarea nervosa*	*Bomarea uncifolia*
Bomarea andimarcana	*Bomarea dissitifolia*	*Bomarea obovata*	*Bomarea uniflora*
Bomarea andreana	*Bomarea distichifolia*	*Bomarea ovallei*	*Bomarea velscoana*
Bomarea angulata	*Bomarea distichophylla*	*Bomarea ovata*	*Bomarea vestita*
Bomarea angustipetala	*Bomarea dolichocarpa*	*Bomarea pardina*	*Bomarea vitellina*
Bomarea angustissima	*Bomarea dulcis*	*Bomarea parvifolia*	*Bomarea weigendii*
Bomarea aurantiaca	*Bomarea edulis*	*Bomarea patacocensis*	*Drymophila cyanocarpa*
Bomarea boliviensis	*Bomarea euryphylla*	*Bomarea patinii*	*Drymophila moorei*
Bomarea brachysepala	*Bomarea fimbriata*	*Bomarea pauciflora*	*Luzuriaga marginata*
Bomarea bracteolata	*Bomarea foliosa*	*Bomarea perglabra*	*Luzuriaga polyphylla*
Bomarea bredemeyerana	*Bomarea formosissima*	*Bomarea peruviana*	*Luzuriaga radicans*
Bomarea bredemeyerana	*Bomarea glaucescens*	*Bomarea phyllostachya*	
Bomarea brevis	*Bomarea goniocaulon*	*Bomarea porrecta*	

Appendix B

Chapter 1 Supplemental Code

All code is freely available on GitHub: `https://github.com/cmt2/underground_evo`

Appendix C

Chapter 2 Supplemental Figures

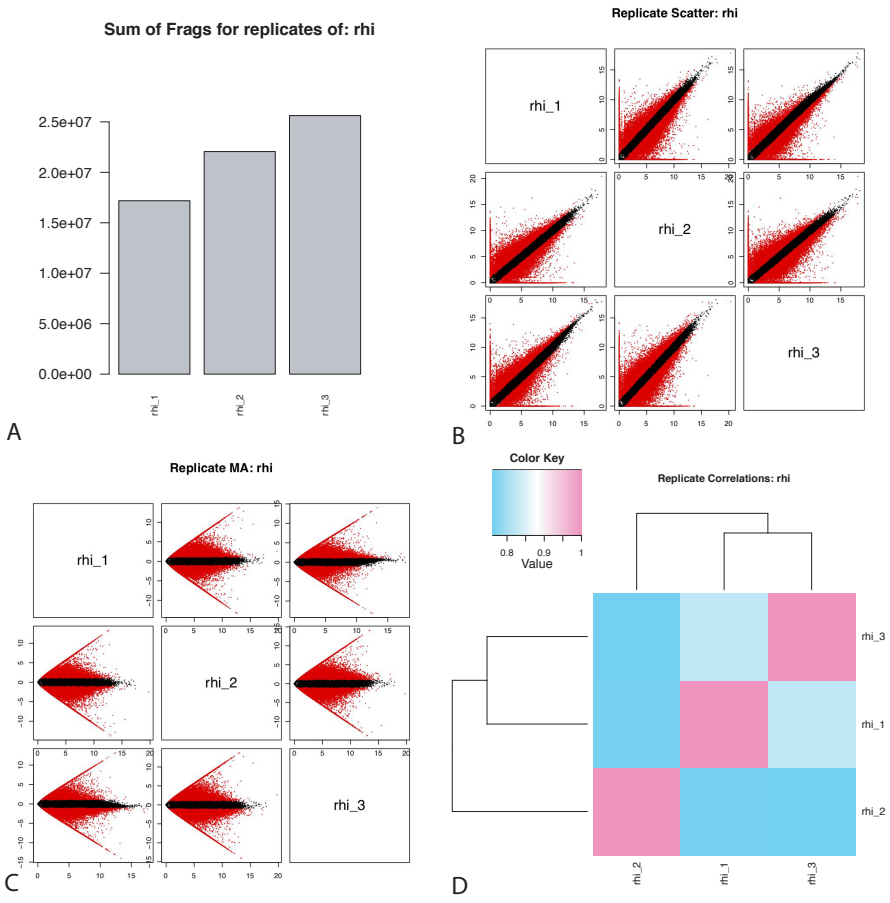

Figure C.1: Replication information for the three biological replicates of rhizome meristem (RHI) tissue, showing A) the sum of fragments for each replicate, B) 1:1 scatterplots of transcript abundance for all pairwise comparisons of the replicates, C) MA plots (Bland Altman plot) for all pairwise comparisons of the replicates, and D) the correlation coefficients between all replicates.

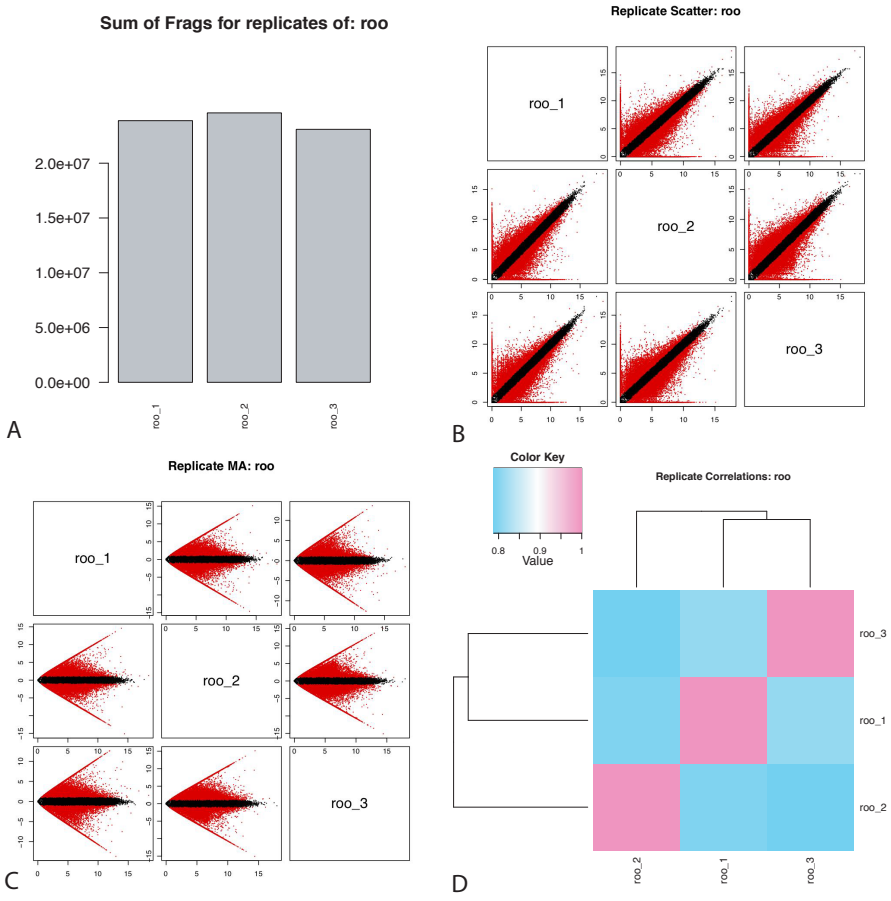

Figure C.2: Replication information for the three biological replicates of root (ROO) tissue, showing A) the sum of fragments for each replicate, B) 1:1 scatterplots of transcript abundance for all pairwise comparisons of the replicates, C) MA plots (Bland Altman plot) for all pairwise comparisons of the replicates, and D) the correlation coefficients between all replicates.

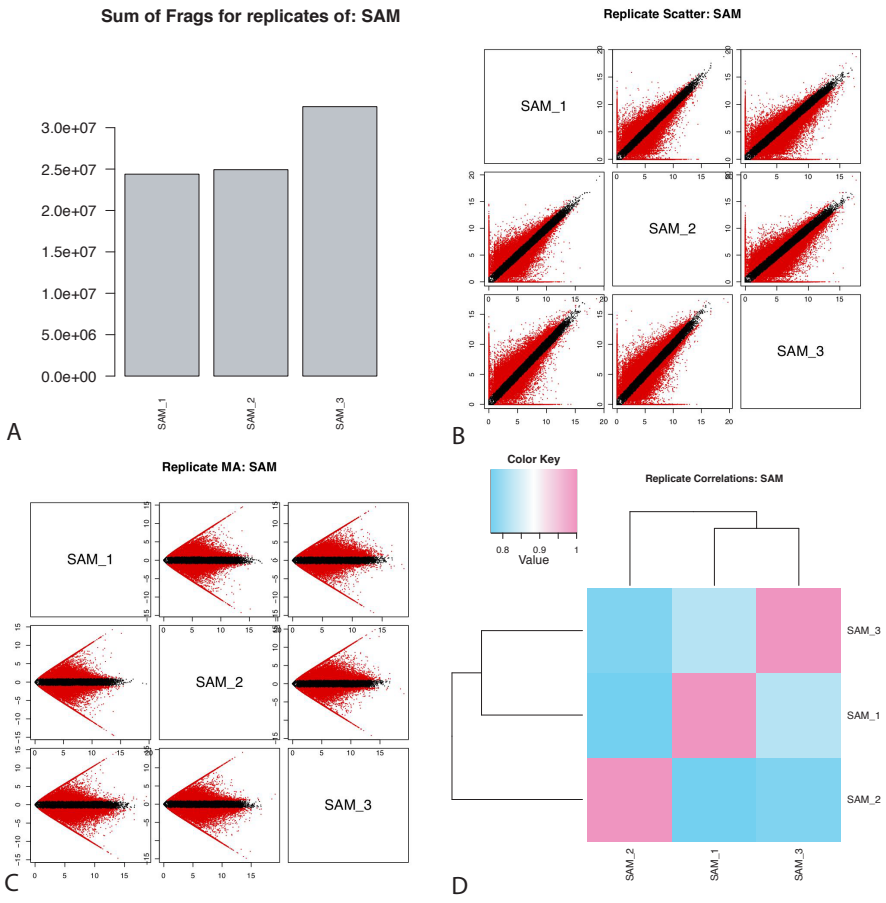

Figure C.3: Replication information for the three biological replicates of shoot apical meristem (SAM) tissue, showing A) the sum of fragments for each replicate, B) 1:1 scatterplots of transcript abundance for all pairwise comparisons of the replicates, C) MA plots (Bland Altman plot) for all pairwise comparisons of the replicates, and D) the correlation coefficients between all replicates.

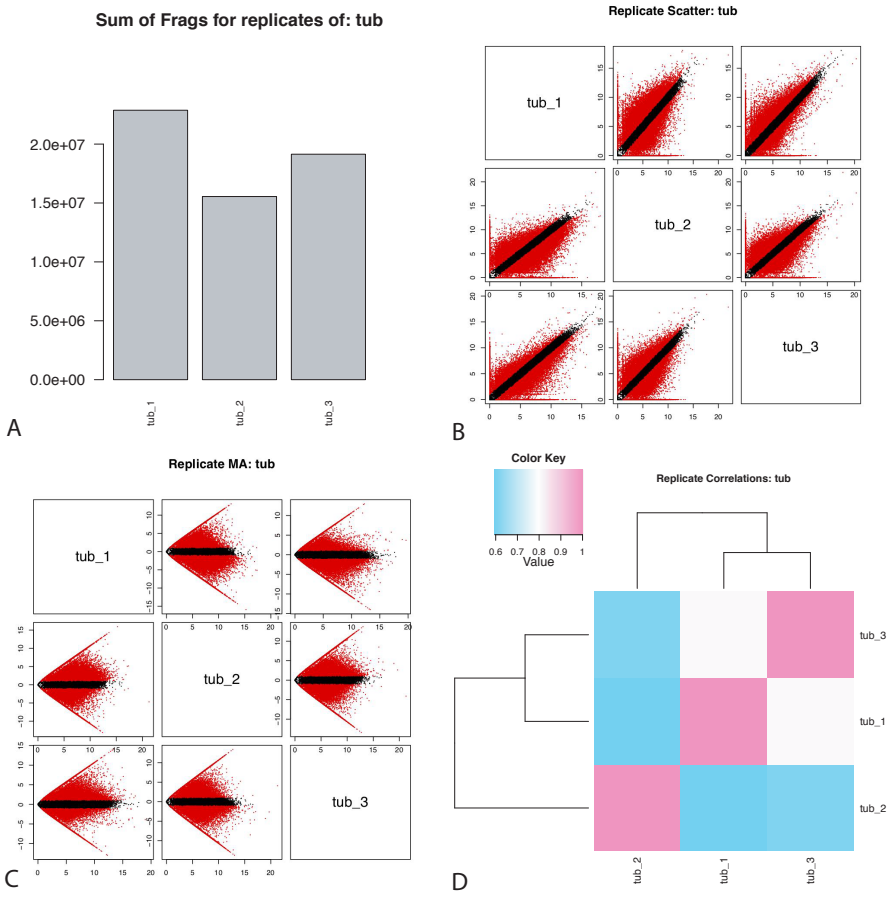

Figure C.4: Replication information for the three biological replicates of tuberous root (TUB) tissue, showing A) the sum of fragments for each replicate, B) 1:1 scatterplots of transcript abundance for all pairwise comparisons of the replicates, C) MA plots (Bland Altman plot) for all pairwise comparisons of the replicates, and D) the correlation coefficients between all replicates.

Appendix D

Chapter 2 Supplemental Tables

Table D.1: Search terms used to identify isoforms in candidate gene groups. We queried the annotated transcriptome using R's grep() function to identify the isoforms and then checked their expression levels in the RSEM data using deseq2.

Gene Group	Search Term	Search Column	Num Isoforms
Starch	GO:0019252	gene_ontology_blast	168
Cytokinin	GO:0009735	gene_ontology_blast	356
Abscisic Acid	GO:0009737	gene_ontology_blast	1357
Auxin	GO:0009733	gene_ontology_blast	716
MADS-Box	MADS-box	sprot_Top_BLASTX_hit	116
KNOX	KNOX	Pfam	18
Gibberellin	GO:0009739	gene_ontology_blast	295
Expansins	Expansin	sprot_Top_BLASTX_hit	117
Lignin	GO:0009809	gene_ontology_blast	325
14-3-3 genes	14-3-3-like protein	sprot_Top_BLASTX_hit	62
CDPK	Calcium-dependent protein kinase	sprot_Top_BLASTX_hit	180

Table D.2: To identify the expression levels of isoforms corresponding to specific candidate genes, we blasted amino acid sequences of the specific candidate gene to a blast-database of the assembled transcriptome.

Candidate Gene	Source	Accession Number	Taxon
FT-like	GenBank	AGZ20207.1	*Allium cepa*
FT-like	GenBank	AGZ20210.1	*Allium cepa*
FT-like	SPROT	tr|M1C558|M1C558_SOLTU	*Solanum tuberosum*
Sulfite reductase	GenBank	AAC24584.1	*Prunus armeniaca*
WOX4	SPROT	sp|Q6X7J9|WOX4_ARATH	*Arabidopsis thaliana*
OsbHLH120	SPROT	tr|Q67TR8|Q67TR8_ORYSJ	*Oryza sativa*
IDD5	SPROT	sp|Q9ZUL3|IDD5_ARATH	*Arabidopsis thaliana*

Appendix E

Chapter 2 Supplemental Code

All code is freely available on GitHub: `https://github.com/cmt2/bomTubers`

Appendix F

Chapter 3 Supplemental Code

The `RevGadgets` R package and the datasets referenced in example code chunks and used to generate figures are freely available on GitHub: `https://github.com/cmt2/RevGadgets`