

EVOLUTIONARY PHYSIOLOGY AND TRANSCRIPTOMICS OF CRASSULACEAN ACID
METABOLISM (CAM) IN THE AGAVOIDEAE (ASPARAGACEAE)

by

CAROLINE HEYDUK

(Under the Direction of Jim Leebens-Mack)

ABSTRACT

Crassulacean acid metabolism (CAM) is a mode of photosynthesis found in ~6% of flowering plants and serves as an adaptation to water-limited habitats. CAM plants open their stomata for gas exchange at night, when transpiration rates are lower, and fix CO₂ via an alternative pathway. Carbon is stored as organic acids during the night, then decarboxylated during the day behind closed stomata. CAM results in high levels of CO₂ around RuBisCO, the primary carbon-fixing enzyme in all green plants, with minimal water loss. Although CAM occurs in at least 35 separate lineages, its evolutionary trajectory from C₃ is unknown. Here we explore the evolutionary patterns of CAM across the Agavoideae, a subfamily of species that includes *Agave* and *Yucca*. Anatomical observations paired with character evolution show that species of the Agavoideae may have been preadapted to the CAM syndrome, with many C₃ species showing CAM-like morphology. Comparative physiology was explored in more detail in a *Yucca* hybrid system, where a CAM and C₃ species hybridized to form a C₃-CAM intermediate. The parents and hybrid offspring were characterized for anatomical and physiological traits and show the hybrid is able to convert from C₃ carbon fixation to 100% CAM uptake under periods of drought stress. Finally, the hybrid system in *Yucca* was used to

understand the transcriptional regulation of the CAM pathway; despite lacking any CAM anatomy or physiology, the C₃ parental species shows similar gene expression patterns as the CAM species, indicating perhaps an ancestral gene expression pattern that enabled the evolution of CAM in a subset of *Yucca* species.

INDEX WORDS: Crassulacean acid metabolism, physiology, phylogenomics, transcriptomics, *Yucca*, hybrid

EVOLUTIONARY PHYSIOLOGY AND TRANSCRIPTOMICS OF CRASSULACEAN ACID
METABOLISM (CAM) IN THE AGAVOIDEAE (ASPARAGACEAE)

by

CAROLINE HEYDUK

BS, University of Wisconsin, 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

© 2015

Caroline Heyduk

All Rights Reserved

EVOLUTIONARY PHYSIOLOGY AND TRANSCRIPTOMICS OF CRASSULACEAN ACID
METABOLISM (CAM) IN THE AGAVOIDEAE (ASPARAGACEAE)

by

CAROLINE HEYDUK

Major Professor:	Jim Leebens-Mack
Committee:	John Burke
	Katrien Devos
	Lisa Donovan
	Chung-Jui Tsai

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

ACKNOWLEDGEMENTS

Live tissue sampling of *Yuccas* was assisted by the park rangers that work for the Florida State Parks, South Carolina State Parks, North Carolina State Parks, and the National Park Service. Darren Loomis from the Virginia Park Service went above and beyond to help me find *Y. gloriosa* samples. Joni Ward at the Desert Botanical Garden assisted with sampling (and re-sampling) their living collection. Additionally, Olle Pellmyr and Jim Leebens-Mack's *Yucca* samples were enormously helpful. For isotopic collections, I thank the Missouri Botanical Garden Herbarium (and Michael McKain), the Desert Botanical Garden Herbarium (Wendy Hodgson), and the Georgia Herbarium (Wendy Zomelefer). Saravanaraj Ayyampalayam assisted with bioinformatics. I had a number of generous funding sources, including the Society for the Study of Evolution, the Society for Comparative and Integrative Biology, the American Society of Plant Taxonomists, the Botanical Society of America, the University of Georgia Graduate School, and NSF. Finally, without the endless support Jim Leebens-Mack, my committee members, and my husband Edward McAssey, none of this work would have been possible.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER	
I INTRODUCTION AND LITERATURE REVIEW	1
The problem with RuBisCO	1
Crassulacean acid metabolism	2
Physiology of CAM plants.....	5
Genetics of CAM plants.....	6
Study system	7
References.....	9
II EVOLUTION OF CRASSULACEAN ACID METABOLISM IN THE AGAVOIDEAE (ASPARAGACEAE)	15
Abstract	16
Introduction.....	17
Methods.....	20
Results.....	26
Discussion.....	30
References.....	36

III GAS EXCHANGE AND LEAF ANATOMY OF A C ₃ -CAM HYBRID, <i>YUCCA</i> <i>GLORIOSA</i> (ASPARAGACEAE).....	52
Abstract.....	53
Introduction.....	54
Methods.....	58
Results.....	64
Discussion.....	67
References.....	74
IV TIME-COURSE COMPARATIVE TRANSCRIPTOMICS BETWEEN C ₃ , CAM, AND INTERMEDIATE <i>YUCCA</i> SPECIES.....	86
Abstract.....	87
Introduction.....	88
Methods.....	91
Results.....	96
Discussion.....	101
References.....	106
V CONCLUSION AND DISCUSSION	124
CAM in the Agavoideae	124
The CAM continuum	125
Remaining questions.....	126
References.....	127
APPENDICES	
A SUPPLEMENTARY FIGURES AND TABLES CH. II.....	129

B SUPPLEMENTARY FIGURES AND TABLES CH. III	133
C SUPPLEMENTARY FIGURES AND TABLES CH. IV	144

LIST OF TABLES

	Page
Table 2.1: Source information for both DNA tissue and for tissue analyzed for $\delta^{13}\text{C}$ ratios	42
Table 2.2: Average values (and standard errors) for cross sectional anatomical traits in the Agavoideae	45
Table 3.1: Nighttime acid accumulation (ΔH^+ , $\mu\text{equivalents g}^{-1}$) derived from difference in dusk and dawn titratable acidity measurements.....	78
Table 3.2: Mean and standard error for traits in the three species and significance of pairwise comparisons	79
Table 4.1: Genotypes of each species used for RNA sequencing.....	110
Table 4.2: Module name, size in terms of number of transcripts, and correlation and p-value to nighttime CO_2 uptake rates	111
Table S2.1: Assembly statistics for each library.....	129
Table S2.2: Delta carbon isotope values for each species	131
Table S3.1: Locality information for each genotype used for physiological analysis.....	133
Table S3.2: Microsatellite primer sequences and repeat motif/length.....	134
Table S3.3: Soil moisture probe measures for % soil water content	135
Table S4.1: Coexpressed gene network module information for all three species of <i>Yucca</i>	144

LIST OF FIGURES

	Page
Figure 2.1: Phylogenetic reconstruction of 66 taxa from the Agavoideae	47
Figure 2.2: Ancestral state reconstruction of photosynthetic pathway in the Agavoideae	48
Figure 2.3: Principal coordinates analysis of the 19 BioClim climate variables for species in the Agavoideae	49
Figure 2.4: Cross section images from select Agavoideae species.....	50
Figure 3.1: Microsatellite variation among genotypes	80
Figure 3.2: Gas exchange patterns for <i>Y. aloifolia</i> , <i>Y. filamentosa</i> , and <i>Y. gloriosa</i>	81
Figure 3.3: Cross sections of <i>Y. aloifolia</i> , <i>Y. gloriosa</i> , and <i>Y. filamentosa</i> at 5x and 10x magnification	82
Figure 3.4: Intercellular airspace (IAS) vs. cell size.....	83
Figure 3.5: Principal coordinates analysis of phenotypic data	84
Figure 3.6: Correlation matrix of phenotypic data.....	85
Figure 4.1: Representation of the CAM pathway	112
Figure 4.2: Description of gene family content and annotation overlap.....	113
Figure 4.3: – For every time point, differentially expressed transcripts under drought conditions relative to well watered.....	114
Figure 4.4: Module eigengenes for the three modules most correlated to nighttime CO ₂ uptake rates	115
Figure 4.5: Gene expression of canonical CAM genes in <i>Y. aloifolia</i>	117

Figure 4.6: Gene expression of canonical CAM genes in <i>Y. gloriosa</i>	118
Figure 4.7: Gene expression of canonical CAM genes in <i>Y. filemantosa</i>	119
Figure 4.8: Gene trees of PEPC (A), PPCK (B), and PEPCK (C), with tips color coded by their module membership.....	120
Figure 4.9: Gene trees of NAD/P-me (A), PPDK (B), and CA (C), with tips color coded by their module membership.....	122
Figure S3.1: Individual genotype gas exchange curves for the 3 samples of <i>Y. aloifolia</i>	136
Figure S3.2: Individual genotype gas exchange curves for the 4 samples of <i>Y. filamentosa</i>	137
Figure S3.3: Individual genotype gas exchange curves for the 9 samples of <i>Y. gloriosa</i>	138
Figure S3.4: Genotypic deltaH+ (H+ μ equivalents AM – H+ μ equivalents PM, per gram of tissue) values across the four days measured.....	141
Figure S3.5: Raw Spearman correlation coefficients of phenotypic traits	142
Figure S3.6: Uncorrected and Holm-Bonferroni corrected p-values for correlations described in Fig. S3.5	143
Figure S4.1: Heatmap (hierarchical clustering) of read counts (normalized as counts per million) for each of the 6004 gene families shared between the three species.....	146

CHAPTER I:

INTRODUCTION AND LITERATURE REVIEW

The problem with RuBisCO

Photosynthesis is the primary producer of the earth's atmosphere, and is the only way in which inorganic, atmospheric carbon is converted to organic carbon. The majority of photosynthetic organisms fix carbon via the Calvin-Benson cycle, otherwise known as "C₃" photosynthesis after the 3 carbon molecule that acts as an intermediate holder of atmospheric CO₂. The primary carboxylating enzyme in C₃ plants is Ribulose-1,5-Bisphosphate carboxylase/oxygenase (RuBisCO); as its name implies, it has the ability to fix both carbon and oxygen from the atmosphere. While carbon that is fixed via RuBisCO is moved into the Calvin cycle to produce carbohydrates, oxygen that is fixed forces plants to undergo photorespiration, an energetically costly process that is required to rid oxygen from Calvin-Benson cycle substrate at the expense of ATP. The oxygenase activity of RuBisCO is particularly pronounced under high temperatures (von Caemmerer and Quick 2000) and high concentrations of oxygen relative to carbon (Forrester et al. 1966). When RuBisCO evolved ~3 billion years ago (Bya) (Nisbet et al. 2007), the earth's atmosphere was high in CO₂ but lacking in oxygen (Christin and Osborne 2013). Today, however, the proportion of the atmosphere that is oxygen is just over 20%, and conditions that promote oxygenation by RuBisCO are not uncommon: many plant species inhabit the tropics, subtropics, and arid regions of the world that experience high daily temperatures; plants that are subject to regular water stress will close their stomata to conserve water, thereby

depleting internal leaf carbon concentrations (C_i) and simultaneously increasing the ratio of oxygen to CO_2 .

To circumvent the oxygenating drawbacks of RuBisCO, many photosynthetic organisms have developed what are known as carbon concentrating mechanisms (CCMs) to enrich the air around RuBisCO with carbon; it's estimated that half of all carbon fixed is acquired through a CCM (Raven et al. 2008). The prokaryotic lineage of cyanobacteria is likely the oldest of the known CCMs, evolving in the low CO_2 atmosphere of the Carboniferous (Raven et al. 2008). A large burst in the evolution of CCMs occurred 20-30 million years ago (Mya) during the Oligocene, when CO_2 levels in the atmosphere dropped significantly (and remain at similar levels today) (Christin and Osborne 2013). During this period, C_4 photosynthesis evolved, which is a more complex CCM that uses spatial separation of the initial carbon fixation and RuBisCO activity to create a carbon "pump." C_4 plants use phosphoenolpyruvate carboxylase (PEPC) as the initial carbon fixing enzyme and concentrate CO_2 as an organic acid (Edwards et al. 2010). The PEPC activity happens in mesophyll cells, but the organic acid is actively shuttled to bundle sheath cells, which is where RuBisCO and the remaining Calvin cycle enzymes are found. When the organic acid is decarboxylated in the bundle sheath cells, the result is a very high concentration of carbon around RuBisCO and highly efficient carbon fixation with substantially less photorespiration.

Crassulacean acid metabolism

A second major CCM, known as Crassulacean acid metabolism (CAM), allows plants to cope with prolonged or frequent water stress by separating temporally the acquisition of atmospheric carbon and Calvin-Benson cycle carbon fixation. At night, when transpiration rates

are lowest, CAM plants open their stomata and fix carbon via PEPC. The carbon is stored as C₄ malic acid in the vacuole overnight, until stomata close during the day and the malic acid is decarboxylated. Like other CCMs, this carbon “pump” results in high carbon concentrations around RuBisCO and more efficient carbon fixation via the Calvin cycle. In addition, because stomata are open mostly during the night (but with morning and afternoon daytime carbon fixation common in CAM plants), plants using the CAM pathway increase their water use efficiency (WUE) by acquiring more carbon per molecule of water lost (Nobel 1988; Woerner and Martin 1999). Unlike C₄, which is thought to have evolved in response to a steep drop in atmospheric CO₂ (but see Arakaki et al. 2011 and Edwards et al. 2010, which highlight the more recent evolution of the C₄ grassland), CAM appears in the flowering plant phylogenies between 5-10 Mya (Good-Avila et al. 2006; Ocampo and Columbus 2010; Bone et al. 2015). Although a number of non-angiosperm species, such as *Welwitschia*, *Isoetes*, and a number of cycads, are all reportedly CAM, these species may have acquired the ability to use CAM more recently (Edwards and Ogburn 2012).

While plants are typically classified as CAM or C₃, it has been largely agreed that a such a binomial classification ignores a large spectrum of variation in CAM photosynthetic ability (Cushman 2001; Winter et al. 2015). Examples of strong CAM plants, which constitutively fix carbon at night, include iconic desert species such as the cacti and agaves. Traditionally, constitutive CAM plants are defined as having little to none of their carbon come from daytime fixation, unless plants are exceptionally well-watered or at the seedling stage (Ting et al. 1996; Winter et al. 2008). Even iconic CAM *Agave* species increase the proportion of carbon fixed via the C₃ pathway when well-watered (Hartsock and Nobel 1976). C₃ plants are likewise easy to distinguish by their inability to nocturnally fix carbon and a lack of acid accumulation in the

leaves, no matter how dry the conditions. Between these two extremes lies a variety of intermediate forms, including: CAM idling, in which plants keep stomata closed day and night but re-fix respired CO₂ (Sipes and Ting 1985); CAM cycling, where in addition to recycling nocturnally respired CO₂, plants also open stomata during the day for carbon fixation via C₃ photosynthesis (Sipes and Ting 1985); facultative CAM, where CAM can be induced or up-regulated in response to abiotic stressors, such as drought or salt stress (Lüttge 2006; Winter et al. 2008, 2011; Herrera 2009; Winter and Holtum 2014); and weak CAM, which describes plants that have high levels of daytime carbon fixation supplemented by low level CAM (Silvera et al. 2005, 2010b). The classification of these plants as “weak” is somewhat misleading, and implies that their level of CAM use is a lesser form than constitutive CAM. This implication has led perhaps to the treatment of weak CAM species as intermediate steps toward full CAM, which ignores the possibility that weak CAM plants are using as much CAM as necessary for their optimal growth, and that intermediate CAM phenotypes can represent an evolutionary endpoint in its own right.

CAM is found in ~6% of flowering plants, but this is likely an underestimate due to the difficulties in assessing CAM, especially in plants expressing low levels of CAM activity. Unlike C₄, which has a high degree of phylogenetic clustering (Sage et al. 1999), CAM is spread across more than 35 families (Silvera et al. 2010a). CAM lineages have been fairly well described to date, but how lineages transition from C₃ to CAM is less well understood. In the C₄ literature, a stepping-stone model of evolution has been proposed in which plants gradually progress from C₃ to C₄ through an intermediate stage known as C₂ (Monson and Rawsthorne 2000; Sage 2004). No such transitory model exists for the evolution from C₃ to CAM, though C₃ plants have the major CAM pathway components in stomatal guard cells, and it has been proposed that CAM evolution

required shifting these reactions to mesophyll cells (Cockburn 1981). In addition, intermediate or weak CAM species have been implicated in the evolution between C_3 and full CAM. For example, an anciently derived gymnosperm, *Welwitschia mirabilis*, employs low levels of CAM or CAM cycling (Ting and Burk 1983), and there is some evidence from the *Crassulaceae* that implicate CAM cycling or a similar weak form was an evolutionary stepping stone to full constitutive CAM (Teeri 1982).

Physiology of CAM plants

As described above, the variations in CAM create a spectrum of phenotypes that can be attributed to CAM plants. For example, although nighttime gas exchange is characteristic of CAM, most CAM plants show peaks of carbon uptake both in the early morning and in the late afternoon that are mediated via the C_3 -pathway (Neales 1975). Daytime carbon uptake in even constitutive CAM plants can be especially pronounced when well watered (Hartsock and Nobel 1976; Keeley and Rundel 2003). Although all CAM plants should show detectable levels of leaf acid accumulation over the night period, the magnitude of titratable acid is variable and can be affected by both light intensity (Kluge and Ting 1978) and day length (Queiroz 1974), and is impeded by low levels of CO_2 (Kluge and Ting 1978). The use of CAM can be affected by environmental conditions including temperature (Sipes and Ting 1985), humidity (Griffiths et al. 1986), drought (Winter et al. 2011), and seasonality (Guralnick et al. 1984).

CAM plants often, but not always, share a common suite of morphological traits, although many of the same traits can be found in C_3 plants. An increase in leaf succulence is the most obvious trait of CAM plants, as demonstrated by iconic CAM taxa like cacti and *Agave*. Succulent tissue in CAM plants forms from the presence of large mesophyll cells for the storage

of malic acid (though succulence can also evolve as a water storage option in non-CAM lineages, as in members of the Agavaceae (Martin et al. 1982; Huxman et al. 1998)) (Keeley and Rundel 2003). Anatomically, CAM leaves show a decrease in the intercellular airspace (IAS) relative to C_3 plants (Nelson and Sage 2008; Zambrano et al. 2014), which is thought to decrease the potential of the concentrated CO_2 to leave cells. Studies have proposed that an ability for 3D venation is critical for increasing leaf succulence without decreasing hydraulic conductivity in leaves (Griffiths 2013; Ogburn and Edwards 2013). In addition, observations on stomata have shown that plants with high levels of nighttime carbon fixation typically have fewer stomata (Zambrano et al. 2014).

Genetics of CAM plants

Although CAM is a derived phenotype, the genetic components of CAM are found in all plant lineages and have been co-opted for use in carbon metabolism (West-Eberhard et al. 2011). Co-option of plant genes for the CAM pathway may have happened through duplication and subsequent neo-functionalization: PEPC has been shown to have a CAM-specific isoform in the *Oncidiinae* (Orchidaceae) (Silvera et al. 2014), and multiple CAM gene families expanded in *Phalenopsis* relative to C_3 ancestors (Cai et al. 2015). Another way plants may have gained CAM is through changes in regulatory mechanisms of CAM genes. Both circadian regulation (Hartwell 2005) and epigenetic changes (Huang et al. 2010) appear to have important roles to play in CAM expression. Although progress has been made in understanding the roles of various CAM pathway genes, particularly with the aid of transgenic approaches (for example, see Dever et al. 2015), still little is known about the regulatory mechanisms of CAM genes or their interaction with clock genes.

For over a decade, *Mesembryanthemum crystallinum* and *Kalanchoe spp.* have been the focus of investigations on the molecular basis of CAM. *Mesembryanthemum crystallinum* is a halophyte that transitions from C₃ to CAM under salt stress, and thus provides an ideal comparative system under which to explore the genetics of CAM induction (Cushman 2001). Species of *Kalanchoe* are ontogenetically programmed for CAM transitions, and comparisons between C₃ and CAM life stages has the added advantage of not requiring the imposition of abiotic stress (Garcia et al. 2014, Hartwell, *unpublished*). With the advent of high throughput sequencing, however, a greater diversity of CAM plants are now being developed as genomic models to study photosynthetic pathway evolution. The genome of the CAM orchid *Phalaneopsis* was published in 2014 (Cai et al. 2015), and a number of other genomes are in the works, including *Ananas comosus* (pineapple, see Zhang et al. 2014b), *Agave*, and *Kalanchoe*. Genomics information, paired with transcriptomic data sampled temporally to observe changes in expression across the diel cycle, are poised to contribute significant advances in our understanding of genome evolution and its impact on CAM evolution. Importantly, however, CAM genomic resources by themselves will be limited in what they provide toward understanding the evolution of CAM from C₃; to thoroughly address that question, -omics level data from closely related C₃ and CAM species are required for comparative genomics analyses. Such methods have been successful in discovering key components of C₄ evolution (Bräutigam et al. 2011) and avian diversity and adaptation (Zhang et al. 2014a), to name a few, and should be similarly applied to the question of CAM evolution.

Study system

Attempts to understand the evolution of CAM have used plant groups with recurrent and independent origins of CAM, such as the Orchidaceae (Silvera et al. 2005, 2010b; Bone et al.

2015; Givnish et al. 2015), Bromeliaceae (Givnish et al. 2014; Quezada et al. 2014; Crayn et al. 2015), and Crassulaceae. In addition, detailed physiological and genetic research into transitions from C_3 to CAM has focused on intermediate C_3 -CAM species. Intermediate species, particularly those that are facultatively CAM, may be especially informative for understanding the requirements and tradeoffs between C_3 and CAM (Winter and Holtum 2014). The Agavoideae, a subfamily of Asparagaceae (Asparagales), is exceptionally suited to both levels of research. At 600 species, it has examples of both mesic and xeric adapted species, including the genera *Agave* and *Yucca*. Further, there are a number of closely related C_3 -CAM species pairs that allow for comparative studies at both the physiological and genetic level. Classic work by Park Nobel used species of *Agave* to understand nearly all aspect of the ecophysiology of CAM, from root growth to water use (Nobel 1976, 1988).

The second chapter of my dissertation describes how characterization of macroevolutionary transitions between C_3 and CAM sets the stage for physiological and genetic investigations of CAM within the Agavoideae. *Yucca* emerges as a particularly attractive model for subsequent investigation, as half of the genus is C_3 while the other is CAM and hybridization between the two subgenera is not uncommon (Webber 1960; Rentsch and Leebens-Mack 2012). The focus shifts in the last two chapters to developing *Yucca gloriosa*, a natural hybrid species derived from a C_3 parent, *Y. filamentosa*, and the CAM species *Y. aloifolia*, as a model for understanding the physiological requirements, genetic basis, and evolution of CAM. The basic physiology, including gas exchange patterning, of the three species has never been investigated in any great detail. Therefore, chapter III set out to measure a number of physiological traits in accessions of all three species, confirming the respective photosynthetic pathways of each parental species and describing in more detail the intermediate nature of *Y. gloriosa*.

Transcriptomic analysis of all three species was used in chapter IV to understand the genetic changes underlying the phenotypic changes described in chapter III. Together, this work advances the Agavoideae as a model system for understanding the evolution of CAM photosynthesis.

References

Arakaki M., Christin P.-A., Nyffeler R., Lendel A., Eggli U., Ogburn R.M., Spriggs E., Moore M.J., Edwards E.J. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. U. S. A.* 108:8379–84.

Bone R.E., Smith J.A.C., Arrigo N., Buerki S. 2015. A macro-ecological perspective on crassulacean acid metabolism (CAM) photosynthesis evolution in Afro-Madagascan drylands: Eulophiinae orchids as a case study. *New Phytol.* .

Bräutigam A., Kajala K., Wullenweber J., Sommer M., Gagneul D., Weber K.L., Carr K.M., Gowik U., Mass J., Lercher M.J., Westhoff P., Hibberd J.M., Weber A.P.M. 2011. An mRNA blueprint for C₄ photosynthesis derived from comparative transcriptomics of closely related C₃ and C₄ species. *Plant Physiol.* 155:142–56.

Von Caemmerer S., Quick W.P. 2000. Rubisco: Physiology in Vivo. In: Leegood R.C., Sharkey T.D., von Caemmerer S. *Photosynthesis: Physiology and Metabolism*. Kluwer Academic Publishers.

Cai J., Liu X., Vanneste K., Proost S., Tsai W.-C., Liu K.-W., Chen L.-J., He Y., Xu Q., Bian C., Zheng Z., Sun F., Liu W., Hsiao Y.-Y., Pan Z.-J., Hsu C.-C., Yang Y.-P., Hsu Y.-C., Chuang Y.-C., Dievart A., Dufayard J.-F., Xu X., Wang J.-Y., Wang J., Xiao X.-J., Zhao X.-M., Du R., Zhang G.-Q., Wang M., Su Y.-Y., Xie G.-C., Liu G.-H., Li L.-Q., Huang L.-Q., Luo Y.-B., Chen H.-H., Van de Peer Y., Liu Z.-J. 2015. The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* 47:65–72.

Christin P.-A., Osborne C.P. 2013. The recurrent assembly of C₄ photosynthesis, an evolutionary tale. *Photosynth. Res.* 117:163–75.

Cockburn W. 1981. The evolutionary relationship between stomatal mechanism, crassulacean acid metabolism and C₄ photosynthesis. *Plant, Cell Environ.* 4:417–418.

Crayn D.M., Winter K., Schulte K., Smith J.A.C. 2015. Photosynthetic pathways in Bromeliaceae: phylogenetic and ecological significance of CAM and C₃ based on carbon isotope ratios for 1893 species. *Bot. J. Linn. Soc.* 178:169–221.

Cushman J.C. 2001. Crassulacean Acid Metabolism. A Plastic Photosynthetic Adaptation to Arid Environments. *Plant Physiol.* 127:1439–1448.

Dever L. V, Boxall S.F., Kneřová J., Hartwell J. 2015. Transgenic perturbation of the decarboxylation phase of Crassulacean acid metabolism alters physiology and metabolism but has only a small effect on growth. *Plant Physiol.* 167:44–59.

Edwards E.J., Ogburn R.M. 2012. Angiosperm Responses to a Low-CO₂ World: CAM and C₄ Photosynthesis as Parallel Evolutionary Trajectories. *Int. J. Plant Sci.* 173:724–733.

Edwards E.J., Osborne C.P., Strömberg C.A.E., Smith S.A., Bond W.J., Christin P.-A., Cousins A.B., Duvall M.R., Fox D.L., Freckleton R.P., Ghannoum O., Hartwell J., Huang Y., Janis C.M., Keeley J.E., Kellogg E.A., Knapp A.K., Leakey A.D.B., Nelson D.M., Saarela J.M., Sage R.F., Sala O.E., Salamin N., Still C.J., Tiplle B. 2010. The origins of C₄ grasslands: integrating evolutionary and ecosystem science. *Science* (80-.). 328:587–91.

Forrester M.L., Krotkov G., Nelson C.D. 1966. Effect of Oxygen on Photosynthesis, Photorespiration and Respiration in Detached Leaves. I. Soybean. *Plant Physiol.* 41:422–427.

Garcia T.M., Heyduk K., Kuzmick E., Mayer J.A. 2014. Crassulacean acid metabolism biology. *New Phytol.* 204:738–40.

Givnish T.J., Barfuss M.H.J., Van Ee B., Riina R., Schulte K., Horres R., Gonsiska P.A., Jabaily R.S., Crayn D.M., Smith J.A.C., Winter K., Brown G.K., Evans T.M., Holst B.K., Luther H., Till W., Zizka G., Berry P.E., Sytsma K.J. 2014. Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Mol. Phylogenet. Evol.* 71:55–78.

Givnish T.J., Spalink D., Ames M., Lyon S.P., Hunter S.J., Zuluaga A., Iles W.J.D., Clements M.A., Arroyo M.T.K., Leebens-Mack J., Endara L., Kriebel R., Neubig K.M., Whitten W.M., Williams N.H., Cameron K.M. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. Biol. Sci.* 282:20151553–.

Good-Avila S. V, Souza V., Gaut B.S., Eguiarte L.E. 2006. Timing and rate of speciation in Agave (Agavaceae). *Proc. Natl. Acad. Sci. U. S. A.* 103:9124–9.

Griffiths H. 2013. Plant venation: from succulence to succulents. *Curr. Biol.* 23:R340–1.

Griffiths H., Luttge U., Stimmel K.H., Crook C.E., Griffiths N.M., Smith J.A.C. 1986. Comparative ecophysiology of CAM and C₃ bromeliads. III. Environmental influences on CO₂ assimilation and transpiration. *Plant, Cell Environ.* 9:385–393.

Guralnick L.J., Rorabaugh P.A., Hanscom Z. 1984. Seasonal Shifts of Photosynthesis in *Portulacaria afra* (L.) Jacq. *Plant Physiol.* 76:643–646.

Hartsock T.L., Nobel P.S. 1976. Watering converts a CAM plant to daytime CO₂ uptake. *Nature.* 262:574–576.

- Hartwell J. 2005. The co-ordination of central plant metabolism by the circadian clock. *Biochem. Soc. Trans.* 33:945–8.
- Herrera A. 2009. Crassulacean acid metabolism and fitness under water deficit stress: if not for carbon gain, what is facultative CAM good for? *Ann. Bot.* 103:645–53.
- Huang N.-C., Li C.-H., Lee J.-Y., Yen H.E. 2010. Cytosine methylation changes in the ice plant *Ppc1* promoter during transition from C3 to Crassulacean acid metabolism. *Plant Sci.* 178:41–46.
- Huxman T.E., Hamerlynck E.P., Loik M.E., Smith S.D. 1998. Gas exchange and chlorophyll fluorescence responses of three south-western *Yucca* species to elevated CO₂ and high temperature. *Plant, Cell Environ.* 21:1275–1283.
- Keeley J.E., Rundel P.W. 2003. Evolution of CAM and C 4 Carbon-Concentrating Mechanisms. *Int. J. Plant Sci.* 164:S55–S77.
- Kluge M., Ting I.P. 1978. Morphology, anatomy, and ultrastructure of CAM plants. In: Billings W.D., Golley F., Lange O.L., Olson J.S. *Crassulacean Acid Metabolism*. Berlin: Springer-Verlag. p. 29–38.
- Lüttge U. 2006. Photosynthetic flexibility and ecophysiological plasticity: questions and lessons from *Clusia*, the only CAM tree, in the neotropics. *New Phytol.* 171:7–25.
- Martin C.E., Lubbers A.E., Teeri J.A. 1982. Variability in Crassulacean acid metabolism: A survey of North Carolina succulent species. *Bot. Gaz.* 143:491–497.
- Monson R.K., Rawsthorne S. 2000. Carbon dioxide assimilation in C3-C4 intermediate plants. In: Leegood R., Sharkey T., von Caemmerer S. *Photosynthesis: Physiology and Metabolism*. Kluwer Academic Publishers. p. 533–550.
- Neales T.F. 1975. The gas exchange patterns of CAM plants. In: Marcelle R. *Environmental and Biological Control of Photosynthesis*. p. 299–310.
- Nelson E.A., Sage R.F. 2008. Functional constraints of CAM leaf anatomy: tight cell packing is associated with increased CAM function across a gradient of CAM expression. *J. Exp. Bot.* 59:1841–50.
- Nisbet E.G., Grassineau N. V., Howe C.J., Abell P.I., Regelous M., Nisbet R.E.R. 2007. The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology.* 5:311–335.
- Nobel P.S. 1976. Water Relations and Photosynthesis of a Desert CAM Plant, *Agave deserti*. *Plant Physiol.* 58:576–582.
- Nobel P.S. 1988. *Environmental Biology of Agaves and Cacti*. Cambridge University Press

- Ocampo G., Columbus J.T. 2010. Molecular phylogenetics of suborder Cactineae (Caryophyllales), including insights into photosynthetic diversification and historical biogeography. *Am. J. Bot.* 97:1827–47.
- Ogburn R.M., Edwards E.J. 2013. Repeated origin of three-dimensional leaf venation releases constraints on the evolution of succulence in plants. *Curr. Biol.* 23:722–6.
- Quezada I.M., Zotz G., Gianoli E. 2014. Latitudinal variation in the degree of crassulacean acid metabolism in *Puya chilensis*. *Plant Biol. (Stuttg)*. 16:848–52.
- Raven J.A., Cockell C.S., De La Rocha C.L. 2008. The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363:2641–50.
- Rentsch J.D., Leebens-Mack J. 2012. Homoploid hybrid origin of *Yucca gloriosa*: intersectional hybrid speciation in *Yucca* (Agavoideae, Asparagaceae). *Ecol. Evol.* 2:2213–22.
- Sage R.F. 2004. The evolution of C₄ photosynthesis. *New Phytol.* 161:341–370.
- Sage R.F., Li M., Monson R.K. 1999. The taxonomic distribution of C₄ photosynthesis. In: Sage R.F., Monson R.K. *C₄ Plant Biology*. San Diego: Academic Press. p. 551–584.
- Silvera K., Neubig K.M., Whitten W.M., Williams N.H., Winter K., Cushman J.C. 2010a. Evolution along the crassulacean acid metabolism continuum. *Funct. Plant Biol.* 37:995.
- Silvera K., Santiago L.S., Cushman J.C., Winter K. 2010b. The incidence of crassulacean acid metabolism in Orchidaceae derived from carbon isotope ratios: a checklist of the flora of Panama and Costa Rica. *Bot. J. Linn. Soc.* 163:194–222.
- Silvera K., Santiago L.S., Winter K. 2005. Distribution of crassulacean acid metabolism in orchids of Panama: evidence of selection for weak and strong modes. *Funct. Plant Biol.* 32:397.
- Silvera K., Winter K., Rodriguez B.L., Albion R.L., Cushman J.C. 2014. Multiple isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid metabolism. *J. Exp. Bot.* 65:3623–36.
- Sipes D.L., Ting I.P. 1985. Crassulacean Acid Metabolism and Crassulacean Acid Metabolism Modifications in *Peperomia camptotricha*. *Plant Physiol.* 77:59–63.
- Teeri J. 1982. Photosynthetic Variation in the Crassulaceae. In: Ting I.P., Gibbs M. *Crassulacean Acid Metabolism*. Rockville, MD: American Society of Plant Physiologists. p. 244–259.
- Ting I.P., Burk J.H. 1983. Aspects of carbon metabolism in *Welwitschia*. *Plant Sci. Lett.* 32:279–285.

Ting I.P., Patel A., Kaur S., Hann J., Walling L. 1996. Ontogenetic Development of Crassulacean Acid Metabolism as Modified by Water Stress in Peperomia. In: Winter K., Smith J.A.C. Crassulacean Acid Metabolism. Berlin, Heidelberg: Springer Berlin Heidelberg.

Webber J.M. 1960. Hybridization and instability of Yucca. Madrono. 15:187–192.

West-Eberhard M.J., Smith J.A.C., Winter K. 2011. Plant science. Photosynthesis, reorganized. Science (80-.). 332:311–2.

Winter K., Garcia M., Holtum J.A.M. 2008. On the nature of facultative and constitutive CAM: environmental and developmental control of CAM expression during early growth of Clusia, Kalanchoe, and Opuntia. J. Exp. Bot. 59:1829–40.

Winter K., Garcia M., Holtum J.A.M. 2011. Drought-stress-induced up-regulation of CAM in seedlings of a tropical cactus, Opuntia elatior, operating predominantly in the C3 mode. J. Exp. Bot. 62:4037–42.

Winter K., Holtum J.A.M. 2014. Facultative crassulacean acid metabolism (CAM) plants: powerful tools for unravelling the functional elements of CAM photosynthesis. J. Exp. Bot. 65:3425–41.

Winter K., Holtum J.A.M., Smith J.A.C. 2015. Crassulacean acid metabolism: a continuous or discrete trait? New Phytol. .

Woerner A.C., Martin C.E. 1999. Mechanistic basis of differences in water-use efficiency between a CAM and a C3 species of Peperomia (Piperaceae). New Phytol. 144:307–312.

Zambrano V.A.B., Lawson T., Olmos E., Fernández-García N., Borland A.M. 2014. Leaf anatomical traits which accommodate the facultative engagement of crassulacean acid metabolism in tropical trees of the genus Clusia. J. Exp. Bot. 65:3513–23.

Zhang G., Li C., Li Q., Li B., Larkin D.M., Lee C., Storz J.F., Antunes A., Greenwold M.J., Meredith R.W., Odeen A., Cui J., Zhou Q., Xu L., Pan H., Wang Z., Jin L., Zhang P., Hu H., Yang W., Hu J., Xiao J., Yang Z., Liu Y., Xie Q., Yu H., Lian J., Wen P., Zhang F., Li H., Zeng Y., Xiong Z., Liu S., Zhou L., Huang Z., An N., Wang J., Zheng Q., Xiong Y., Wang G., Wang B., Fan Y., da Fonseca R.R., Alfaro-Nunez A., Schubert M., Orlando L., Mourier T., Howard J.T., Ganapathy G., Pfenning A., Whitney O., Rivas M. V., Hara E., Smith J., Farre M., Narayan J., Slavov G., Romanov M.N., Borges R., Machado J.P., Khan I., Springer M.S., Gatesy J., Hoffmann F.G., Opazo J.C., Hastad O., Sawyer R.H., Kim H., Kim K.-W., Kim H.J., Cho S., Li N., Huang Y., Bruford M.W., Zhan X., Dixon A., Bertelsen M.F., Derryberry E., Warren W., Wilson R.K., Li S., Ray D.A., Green R.E., O'Brien S.J., Griffin D., Johnson W.E., Haussler D., Ryder O.A., Willerslev E., Graves G.R., Alstrom P., Fjeldsa J., Mindell D.P., Edwards S. V., Braun E.L., Rahbek C., Burt D.W., Houde P., Zhang Y., Yang H., Jarvis E.D., Gilbert M.T.P., Ye C., Liang S., Yan Z., Zepeda M.L., Campos P.F., Velazquez A.M. V., Samaniego J.A., Avila-Arcos M., Martin M.D., Barnett R., Ribeiro A.M., Mello C. V., Lovell P. V., Almeida D., Maldonado E., Pereira J., Sunagar K., Philip S., Dominguez-Bello M.G., Bunce M., Lambert D.,

Brumfield R.T., Sheldon F.H., Holmes E.C., Gardner P.P., Steeves T.E., Stadler P.F., Burge S.W., Lyons E., McCarthy F., Pitel F., Rhoads D., Froman D.P. 2014a. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* (80-). 346:1311–1320.

Zhang J., Liu J., Ming R. 2014b. Genomic analyses of the CAM plant pineapple. *J. Exp. Bot.* 65:3395–404.

CHAPTER II:

EVOLUTION OF CRASSULACEAN ACID METABOLISM IN THE AGAVOIDEAE (ASPARAGACEAE)¹

¹ Heyduk K, Lalani F, McKain MR, Leebens-Mack J. To be submitted to *New Phytologist*.

Abstract

Crassulacean acid metabolism (CAM) is a modified form of photosynthesis that has arisen independently at least 35 times in flowering plants. This novel form of photosynthesis is thought to allow plants to inhabit water stressed environments through the opening stomata for gas exchange primarily at night, when transpiration rates are lowest. The carbon acquired at night is stored as malate in vacuoles until the day, when stomata close and malate is decarboxylated. The result is a very high concentration of CO₂ around RuBisCO with reduced photorespiration and a decrease in water loss per molecule carbon fixed by RuBisCO. The occurrence of CAM in plant lineages is often correlated with shifts to arid, semiarid, or epiphytic habits, as well as transitions in leaf morphology (e.g. leaf thickness) and anatomy (e.g. cell size and packing) from C₃-like to more CAM-like. Here we assess the evolution of CAM in the Agavoideae, a subfamily of Asparagaceae which consists of ~600 species and includes iconic desert genera such as *Agave* and *Yucca* as well as genera such as *Hosta* that inhabit mesic environments. In order to understand the independent origins of CAM within the Agavoideae, we developed target enrichment baits to selectively sequence a small portion of the genome of over 60 species. Both nuclear and chloroplast data were used to reassess the phylogeny of the group, and carbon isotope ratios were used as a proxy for CAM presence in a species. Ancestral character state mapping suggests three independent origins of CAM in the Agavoideae. CAM species and C₃ species are separated in a PCA of climate space, suggesting that the evolution of CAM in species in the Agavoideae allowed for movement into novel environments. Furthermore, these shifts from C₃ to CAM are associated with changes in leaf morphology, but even C₃ ancestors of the CAM Agavoideae species show a predisposition toward CAM-like morphology, including an increase in leaf thickness and cell size with the aid of 3D venation.

Introduction

Most plants use C₃ photosynthesis whereby the Calvin-Benson cycle is employed to fix atmospheric carbon into usable sugars. RuBisCO is the primary carboxylating enzyme in C₃ plants, but under certain conditions RuBisCO acts as an oxygenase in addition to its central carboxylating activity. High temperatures and a high O₂:CO₂ ratio promote oxygenation, which initiates the photorespiration pathway to remove oxygen from the RuBisCO and its substrate, RuBP. Photorespiration is energetically costly for plants and no net carbon is fixed. Current day levels of oxygen in the atmosphere (20%) far outweigh the amount of CO₂ (0.04%); furthermore, conditions that lead plants to close their stomata, such as drought or high evapotranspiration rates, cause a drawdown of leaf CO₂ levels and an increase in the leaf O₂/CO₂ ratio. The latter situation in which plants are regularly water limited has spurred the repeated evolution of a carbon concentrating mechanism known as Crassulacean acid metabolism (CAM), whereby plants increase their water use efficiency (WUE, molecules of water lost per molecule of carbon gained) by opening stomata only at night. In CAM plants, CO₂ is fixed by phosphoenolpyruvate carboxylase (PEPC) and stored as the C₄ compound malic acid in the vacuole until the daytime, when stomata close and malic acid is decarboxylated, resulting in increased concentrations of CO₂ around RuBisCO. The concentration of carbon within CAM plant cells results in an increased WUE, and allows CAM plants to grow in some of the most water stressed environments on earth.

The evolution of CAM from C₃ ancestors has occurred at least 35 times independently across angiosperms (Silvera et al. 2010). Though CAM is not as clustered phylogenetically as C₄ photosynthesis, with as many as 60% of all C₄ species being members of the Poales (Edwards et al. 2010; Sage et al. 2011), a few large lineages seem to have multiple repeated origins of CAM.

Major lineages which contain many CAM species include the Orchidaceae, Cactaceae, Bromeliaceae, Crassulaceae, Euphorbiaceae, and the Agavoideae within the Asparagaceae. Surveys of CAM plants across broad phylogenetic scales typically include studies of carbon isotope ratios ($\delta^{13}\text{C}$): CAM and C_4 plants, via the enzyme PEPC, discriminate less against the heavier isotope ^{13}C isotope of carbon than RuBisCO does, resulting in a ratio between ^{13}C and ^{12}C that is more similar to atmospheric ratios in CAM plants relative to C_3 . CAM plants have $\delta^{13}\text{C}$ ratios between -10‰ and -20‰, while C_3 plants have ratios more negative than -22‰. While a quick diagnostic, especially for a large number of taxa, $\delta^{13}\text{C}$ can be misleading when plants are neither strong CAM nor only C_3 ; in other words, $\delta^{13}\text{C}$ ratios cannot differentiate plants that use CAM part-time from full C_3 plants (Winter and Holtum 2002). Additionally, carbon isotope ratios for CAM and C_4 plants are indiscernible, although climate and anatomical observations can help separate species using the two pathways. Despite these limitations, $\delta^{13}\text{C}$ still remains the best option for quick surveys of entire plant lineages, and can provide a starting point for more detailed physiological analysis.

Anatomically, CAM plants share a common suite of traits that are directly related to the ability of the CAM pathway to function efficiently. Large cells, and particularly large vacuoles, are required for storing malic acid; the amount of malic acid stored is highly predictive of the amount of CO_2 fixed from the atmosphere via PEPC. To accommodate large cells, CAM plants are thought to decrease the intercellular airspace (IAS) in leaves, which has the added benefit of limiting the diffusion of decarboxylated CO_2 in CAM plant cells during the day (Nelson et al. 2005; Nelson and Sage 2008; Zambrano et al. 2014). A number of CAM traits represent tradeoffs with the C_3 photosynthetic pathway: for example, a lack of IAS in C_3 plants, which are already susceptible to low carbon levels and photorespiration, will limit the ability of CO_2 to

diffuse throughout the layers of mesophyll cells. CAM plants therefore underwent a transition in their evolution from C_3 and acquired anatomical modifications that promote optimal CAM pathway function. Evolutionary studies of C_3 to C_4 transitions have found that anatomical differences arise before a fully optimized C_4 pathway (Christin et al. 2011, 2013a), but few studies have examined anatomical preconditioning for the evolution of CAM (but see Griffiths 2013).

The question of preconditioning for the evolution of CAM was explored in the subfamily Agavoideae (Asparagaceae) (APG III, Chase et al. 2009), a group of ~600 species that includes the iconic desert genera *Yucca* and *Agave*, in addition to a number of mesic genera, like *Camassia* and *Hosta*. The Agavoideae are widespread but largely restricted to the new world (although *Hosta* is an exception, with a number of species found in eastern Asia (Rocha et al. 2006)). The center of diversity for many, but not all, of the species is the deserts of North and Central America (Bogler and Simpson 1996). Morphologically, the Agavoideae display a range of floral types and pollination syndromes, including specialized pollinators, like bats in *Agave* (Howell and Roth 1981; Gentry 1982) and moths in *Yucca* (Pellmyr 1999, 2003), as well as more generalist bee pollinators (Suzuki et al. 2002; Rentsch and Leebens-Mack 2014). To date, analyses of relationships of Agavoideae members have been plagued by low and mixed phylogenetic signals (Good-Avila et al. 2006; Smith et al. 2008), particularly within genera. The phylogenetic difficulties largely arise from the young age of the group: the core Agavoideae are estimated to have diverged 20-26 Mya, while *Agave* and *Yucca* represent much more recent diversification events at 8-10 and 9-17 Mya, respectively (Good-Avila et al. 2006; Smith et al. 2008). Previous work on photosynthesis within the Agavoideae has focused almost exclusively on *Agave* (but see Smith et al. 1983; Huxman et al. 1998 for examples in *Yucca*), and the

remaining desert species were assumed to be CAM. Here, I assessed photosynthetic pathway in over 60 species from 12 genera and estimate a phylogenetic tree of these taxa using target enrichment techniques (Heyduk et al. 2015) with 68 nuclear loci and ~72kb of the plastid genome. Using these relationships, ancestral states of photosynthetic pathway were used to estimate the number of origins of CAM. In addition, leaf cross sections were used to explore whether certain CAM promoting traits – such as large cells or decreased IAS – evolved prior to or concurrently with CAM photosynthesis.

Methods

Tissue sampling

For phylogenetic analyses, many of the *Yucca* tissues were collected from natural populations by Olle Pellmyr and Jim Leebens-Mack and stored at -80°C for 10-20 years, while tissue for all *Agave* accessions came from the Desert Botanical Garden's live collection. Additional sources of tissue are described in Table 2.1. For each species, the source of tissue for carbon isotope measurements is also indicated in Table 2.1. Most isotope tissue was sourced from the Missouri Botanical Garden Herbarium, the University of Georgia State Herbarium, and the Desert Botanical Garden Herbarium. Tissue for anatomical cross sections was harvested from germplasm growing at the University of Georgia plant biology greenhouses.

RNA target enrichment bait design

Transcriptomes from *Yucca filamentosa*, *Chlorogalum pomeridianum*, *Hesperaloe parviflora*, *Hosta venusta*, and *Yucca brevifolia*, as well as available EST sequences from *Agave tequilana* (from Simpson et al. 2011) were assembled with Trinity (release: 2012-06-08).

Transcripts of each species were translated by tBLASTx against 10 sequenced angiosperm

genomes. The best hit of each transcript was retained and used as the guide for translation using in the TransPipe pipeline (Barker et al. 2010). The resulting amino acid and nucleotide coding sequences were clustered into gene families using OrthoMCL (Li et al. 2003) as in McKain et al. 2012. Amino acids of all sequences in each gene family were aligned in MUSCLE (Edgar 2004) and the coding sequences were mapped to the protein alignments using PAL2NAL (Suyama et al. 2006). Using rice intron boundaries, each gene family alignment had intron-exon boundaries estimated using intronFinder from the Solanaceae Genomics Network (Fernandez-Pozo et al. 2015). Gene alignments were then broken into putative exon sequences and introns were removed. All sequences within a single exon alignment had pairwise distance calculated to estimate the percentage of divergence between taxa for a given exon. Sequences were clustered within an exon so that no member of a cluster was more than 10% divergent from any other member. From these clustered exons, baits were designed for putative single copy genes (Duarte et al. 2010), and resulted in 1200 exons from 776 genes.

Library construction and enrichment

DNA was isolated from all tissue using a modified CTAB protocol (Doyle 1987), using an additional sorbitol cleaning step to remove plant secondary compounds (Štorchová et al. 2000). DNA was sheared with a Covaris sonicator (Covaris Inc.) to an average size of 350bp. Libraries were constructed using either single-ended or dual-indexed adapters with an in-house protocol modified from Fisher et al. 2011. Libraries were then used as the template for hybridization with the RNA baits as described in the Myselect protocol and in Heyduk *et al* 2015. Briefly, libraries are hybridized to biotinylated RNA baits for 36 hours, with 4-5 libraries per pool. Nuclear and chloroplast baits were added to the same hybridization reaction, at a 4:1 nuclear:chloroplast volume ratio. Nuclear baits were added at original concentration, but

chloroplast baits were diluted to 1/500x prior to hybridization. Libraries were pooled based on phylogenetic distance to minimize preferential hybridization and capture. Post-hybridization, libraries were assessed for target enrichment using quantitative real time PCR (qRT-PCR) of two nuclear targets and a chloroplast target. Levels of targeted loci were compared to non-targeted ribosomal ITS copy numbers in the enriched pools, using unenriched libraries as a control. Enriched pools were then quantified using qRT-PCR and pooled for sequencing. Three separate sequencing runs were carried out, including one preliminary run on a MisSeq v 2500 using 150bp PE reads, one run on a HiSeq v. 2500 with 100bp PE reads, and a final sequencing run on a HiSeq. v. 2500 rapid run with 150bp PE reads, with a final total of 69 species sequenced.

Assembly

Reads were quality trimmed and removed of adapters using FASTX tools and FAR, respectively. Libraries for two samples of *Camassia quamash* and the sample of *Chlorophytum rhizopendulum* were removed due to low read counts. Trinity (Grabherr et al. 2011) was used for assembly because sequence capture data, much like RNA-seq for which Trinity was designed, consists of short fragments with variable coverage. Trinity isoforms were collapsed using CAP3 (Huang and Madan 1999) with at least 95% identity and a 20bp overlap. For chloroplast alignments, cleaned reads were first mapped to a reference sequence from *Yucca filamentosa* (McKain et. al, *in prep*) using Bowtie2 (Langmead et al. 2009). Only reads that mapped to the reference were used in subsequent YASRA (Ratan 2009) assemblies, each of which required a minimum contig coverage of 10x for assembly. Contigs from YASRA were merged end-to-end for each species. Coverage was calculated using Bowtie v.2 (Langmead et al. 2009) and bedtools (Quinlan and Hall 2010).

Phylogenetic reconstruction

To reconstruct phylogenetic relationships, only target sequences were kept using methods described in (Heyduk et al. 2015), but were modified as follows. To minimize noise but maximize signal, all intron sequences were trimmed to include only 100bp of intron on either side of an exon via BLAST alignments to the target exons. Exons were then merged together into genes based on their order in the original target sequences. Genes were then aligned using PRANK (Löytynoja and Goldman 2005) and cleaned with Gblocks (Castresana 2000), requiring half of the taxa for a given gene to have sequence in a block for it to be kept. To reduce spurious alignments, sequences within each gene were removed if they did not comprise at least 50% of the total alignment length. Genes were only considered for further analyses if they were at least 200bp long, contained at least one parsimony informative single nucleotide polymorphism, and had at least 34 of the 66 taxa in the alignment. Gene trees were estimated independently in RAxML (Stamatakis 2006) with 500 bootstrap replicates and the GTRGAMMA model of evolution.

The RAxML best trees and bootstrap trees were used in a subsequent ASTRAL (Mirarab et al. 2014) analysis with 400 bootstrap replicates; ASTRAL bootstrap replicate trees were imported into Geneious 8.1.6 (Kearse et al. 2012) to estimate a consensus tree. Multiple samples for some species were treated independently (versus being forced to represent a single lineage in the species tree). In addition, individual gene alignments were concatenated into a supermatrix, and a species tree was estimated using RAxML and the GTRGAMMA model as above. Chloroplast sequences were aligned with MAFFT (Kato and Standley 2013) and were cleaned with Gblocks, requiring 34 of the 66 taxa to be present in a block for it to be retained. Species relationships were estimated in RAxML with GTRGAMMA and 500 bootstrap replicates.

ASTRAL implements a coalescence-based approach to species tree estimation, whereas the supermatrix analysis assumes that all genes share the same history and does not take into account deep coalescence that arises from incomplete sorting of ancestral alleles during speciation events. The ASTRAL and supermatrix trees were compared to assess the influence of ancestral gene coalescence in supermatrix analysis.

Isotopic analysis

Tissue for isotopic analysis was largely collected from either the Desert Botanical Garden herbarium, The Missouri Botanical Gardens, or from the Georgia State Herbarium (for all accession information, see Table 2.1). Tissue was selected from non-horticultural accessions whenever possible to avoid effects of watering on the plasticity of mode of carbon uptake. Samples were dried at 50°C for two days then hand ground with mortar and pestle. Isotopic composition was assessed at the University of Georgia Soil Analysis Laboratory against a background of Pee Dee belemnite.

Ancestral state reconstruction

Estimation of ancestral states at the nodes of the Agavoideae tree was conducted in diversitree (FitzJohn 2012) in R 3.2.2 (R Core Team 2013) using a time calibrated version of the supermatrix tree. To generate a time-calibrated tree, we used the ape package with an age estimate for the root (Agavoideae+*Hosta*) from Smith et al. 2008, with a maximum age of 15.14 My and a minimum age of 13.26 My. Although diversification rates might vary by whether a lineage is C₃ or CAM, simulation work has shown that models that incorporate diversification along with transitions from one character state to another (BiSSE, Davis et al. 2013) can be inaccurate when the number of taxa is low. Instead, a Markov model of discrete character

evolution was used to first assess whether transitions from C₃ to CAM were equally likely as reversions from CAM to C₃ in the diversitree package in R. As there was no improvement in model fit when allowing for different transition rates, a single rate was used to estimate ancestral states at the nodes of the calibrated phylogeny using marginal probabilities.

Climate space analysis

GPS coordinates were taken from herbarium records of species used in the phylogeny, excluding records from gardens or those from non-native ranges. For those coordinates, 19 bioclimatic variables were downloaded from the WorldClim database (<http://www.worldclim.org/bioclim>) using the raster package in R. A principal coordinates analysis was used to assess climate similarities between estimated C₃/CAM species. Samples of species were designated according to the exemplar species measured for carbon isotope ratios in this study; although this may be a generalization for more intermediate species, the majority of species are expected to be fixed for CAM or C₃ (Silvera et al. 2005) and can be represented by a single isotope value.

Anatomical evolution

Cross sections from exemplars from the following species were collected from fresh tissue: *Agave schottii*, *Agave palmeri*, *Polianthes tuberosa*, *Manfreda scabra*, *Manfreda virginica*, *Yucca brevifolia*, *Yucca angustissima*, *Beschorneria yuccoides*, *Hosta ventricosa*, *Chlorophytum rhizopendulum*, *Hesperaloe funifera*, and *Hesperaloe parviflora*. Leaves were cut, fixed in formalin and embedded in paraffin, then sectioned with a microtome and stained with Toluidine blue at the UGA Veterinary Histology Laboratory. When available, two separate plants per species were used for cross sectioning; for every cross section, 2-3 images were

captured across the leaf, measured, and then averaged. Cross sections were measured for traits as in Heyduk et al. 2015b (in review), and are described here briefly. Leaf thickness, average cell size (adaxial and abaxial), intercellular air space (IAS), and the number of planes of fully developed vasculature were counted and measured in ImageJ (Rashband). IAS as a percent of mesophyll area. Student's T-test were used to compare CAM and C₃ species for anatomical traits.

Results

Sequence capture

After read-cleaning, libraries had an average of 7 million reads, of which 3.28% mapped to the chloroplast reference, and an additional 2.7% mapped to the single copy genes of interest (for library specific data, see Table S2.1). On average 450 exons were successfully captured in each library, originating from an average of 355 genes. After filtering for paralogous sequences, trimming to retain only 100bp of intron, and removing short alignments and genes with less than 35 taxa, 68 genes remained. Exons from those remaining genes had an average coverage of 49x, while the adjacent intron sequence had an identical average coverage (for library-specific coverage values, see Table S2.1). Average coverage over the ~78kb of chloroplast genome assembled was 121x. The baits designed were successful across the entire subfamily, though an attempt to capture an outgroup species from the Anthericaceae failed. This may have been a result of poor DNA quality or divergence from bait sequences.

Phylogenetic inference

Although relationships among genera of the Agavoideae have been fairly well described to date (Smith et al. 2008; McKain et al. 2012; Archibald et al. 2015), species-level relationships

have proven to be more difficult to resolve with any confidence, despite interest in both *Yucca* and *Agave* for pollination syndrome evolution (Pellmyr 2003; Rocha et al. 2006). Using 68 genes spanning a total of 35,660bp of nuclear information across, including both exonic and intronic sequences, we were able to resolve finer scale resolution among genera (Fig. 2.1). Furthermore, results from a coalescent approach indicate a potential hard polytomy at the base of the Agavoideae, which is in line with the rapid radiation of the group 20-26 Mya (Rocha et al. 2006; Smith et al. 2008). Aside from the well-documented paraphyly of *Agave* (e.g. Bogler and Simpson 1996; Bogler et al. 2005; Rocha et al. 2006; Smith et al. 2008), all genera were supported as monophyletic, with the exception of *Manfreda* in the ASTRAL and chloroplast analyses. The two sampled *Manfreda* species formed a clade in the supermatrix tree, but the placement of *Manfreda*, *Polianthes*, and *Prochyanthes* (not sampled here) within *Agave* sensu lato is well established (Bogler et al. 2005; Rocha et al. 2006).

Resolution within the genera *Agave* and *Yucca* is not much improved from previous analyses, and is likely a result of the rapid radiation in each genus. For *Agave* sensu lato in particular, the polytomy at the backbone of the genus was never able to be resolved, suggesting current gene flow among species (Gentry 1982) as well as retention of ancestral polymorphism due to short branches on the tree (Fig. 2.1). Even with 68 genes and 72kb of chloroplast sequence, resolution of this genus is poor, suggesting that *Agave* represents a hard polytomy, one that will not be able to be resolved with additional data. In agreement with previous analyses of plastid and ribosomal markers (Bogler and Simpson 1996; Bogler et al. 2005; Rocha et al. 2006; Smith et al. 2008), the placement of *Manfreda* and *Polianthes* remains within *Agave*, indicating a recent divergence of these genera from *Agave* ancestors. *Yucca* is better resolved, perhaps because its specialized moth pollination system may have kept some (but not all) hybridization

between species to a minimum. The three subgenera typically identified by their fruit types – *Chaenocarpa* (dehiscent fruit), *Sarcocarpa* (fleshy fruit), and *Clistocarpa* (spongy, monospecific containing only *Y. brevifolia*) – all fall out separately on all three tree methods and are well supported in their division (Fig. 2.1). The relationships between these subgenera, however, are less clear, as the backbone of *Yucca* is also a polytomy.

Lineages sister to *Yucca* and *Agave* largely confirm previous work, although whether *Camassia*+*Chlorogalum*+*Hastingsia* is sister to *Hesperaloe*+*Schoenolirion*+*Hesperoyucca* is dependent on the tree method used; ASTRAL was unable to determine relationships between any of the major clades in the Agavoideae, while the supermatrix and chloroplast estimations place *Camassia*+*Chlorogalum*+*Hastingsia* sister to *Hesperaloe*+*Schoenolirion*+*Hesperoyucca*. In addition, the concatenated tree has *Hosta* sister to both of these clades; this is consistent with uncertain placement of the genus relative to the rest of the Agavoideae in a number of other studies (Archibald et al. 2015).

Ancestral states and climate space

Carbon isotope values for all species are listed in Table S2.2 Ancestral state reconstruction on a time calibrated supermatrix phylogeny of the Agavoideae showed high support for three independent origins of CAM and one reversion: one gain at the base of *Hesperaloe*, one at the base of subgenus *Sarcocarpa* in *Yucca*, and a third at the most recent common ancestor of *Agave*, *Manfreda*, and *Polianthes* (Fig. 2.2). A single reversion is seen at *Polianthes tuberosa*. Although the phylogeny is calibrated with a single node, the evolution of CAM in all three instances occurs at roughly the same time, about 2.5-5 Mya.

A PCA of the 19 climate variables from the WorldClim database show a large degree of separation between C₃ and CAM taxa, with 66% of the total variation explained by the first two axes (Fig. 2.3). The first component, which separates C₃ from CAM plants, is largely driven by the minimum temperature of the coldest month (loading score of 0.33), the mean temperature of the coldest quarter (0.33), the annual temperature range (-0.32), and temperature seasonality (standard deviation of temperatures) (-0.31). The second axis largely separates taxa by the maximum temperature in the warmest month (-0.38), the mean temperature of the warmest quarter (-0.35), precipitation of the driest quarter (0.33), and precipitation of the driest month (0.32). Temperature, particularly that of the coldest and warmest times of the year, impacts where CAM and C₃ species are distributed, but the large degree in overlap between C₃ and CAM indicates a range of seasonal variation in temperature that permits the co-occurrence of both C₃ and CAM.

Anatomical evolution

Average leaf thickness and average cell area (including both abaxial and adaxial mesophyll cells) were significantly different between C₃ and CAM plants (Table 2.2). In addition, CAM plants had a larger number of planes of vascular bundles than C₃ plants, although by specifically avoiding the large central mid-rib to minimize heterogeneity in leaf measurements, these results may be biased (unmeasured cross sections from thicker midrib regions are shown in Fig. 2.4). A number of C₃ species, including *B. yuccoides* and *Y. angustissima*, had a propensity for succulence, albeit restricted to the center of the leaf (Fig. 2.4). Further, although CAM plants have a greater degree of 3D venation, both C₃ and CAM plants in the Agavoideae have significantly more planes of vascular bundles than 1 (one sample T-test, df=37, p=0). Although the degree of intercellular airspace was significantly different between C₃

and CAM species (Table 2.2), a number of C₃ genera, including *Polianthes*, *Manfreda*, and *Beschorneria*, had values more similar to their CAM relatives in *Agave* (Table 2.2).

Discussion

Evolution of CAM

Historically, phylogenetic analyses of the Agavoideae have been able to identify major clades, but relationships among these clades has been difficult to resolve (Bogler and Simpson 1996; Bogler et al. 2005; Rocha et al. 2006; Smith et al. 2008). This work, which comprises 68 single copy nuclear genes and an addition 72kb of chloroplast genome sequence, found strong support for five major clades: *Agave sensu lato*+*Beschorneria*, *Yucca*, *Camassia*+*Chlorogalum*+*Hastingisa*, *Hesperaloe*+*Schoenolirion*+*Hesperoyucca*, and *Hosta*. This study included comprehensive sampling of the Agavoideae and many more loci than had been analyzed in previous studies, yet we failed to resolve a bifurcating set of relationships between the five major clades on the backbone of the Agavoideae phylogeny. This strongly implicates rapid radiation among these five lineages, perhaps spurred by parallel divergence from a single ancestral polyploidy population (McKain et al. 2012). At the very least, we must acknowledge that the true history fo the Agavoideae is not likely to the fit the model of simple bifurcating cladogenesis assumed by most phylogenetic inference approaches.

Within genera, particularly *Agave*, resolution is even poorer. Again this likely reflects the biology of the genus, which is known to hybridize and is estimated to have speciated in the last 5-10 My (Rocha et al. 2006; Smith et al. 2008, this paper). For long-lived perennials that have a lifespan of a century and, in some cases, flower once at the end of their life, this quick radiation is further compounded by long generation time. Combined with current day hybridization, gene

flow both past and present will prevent the resolution of the genus, even using a larger set of data. In the case of *Yucca*, the specialization of moth species to one or a few *Yucca* species has promoted diversification, despite the age of *Yucca* being nearly the same as *Agave* (Smith et al. 2008). Although traditionally defined subgenera in *Yucca* – *Sarcocarpa*, *Chaenocarpa*, and *Clistocarpa* – are largely monophyletic, their relationships to each other are not resolved in either nuclear analysis, again possibly implicating a species radiation rather than bifurcating cladogenesis.

The issues of phylogenetic resolution are further amplified by the inability of ancestral state reconstruction programs to integrate tree uncertainty into their algorithms. Diversitree, for example, required a fully bifurcating tree; the one provided was from the supermatrix analysis, but the majority of nodes shown in Figure 2.2 have bootstrap support much lower than 50. Importantly for this subfamily, however, is that the main transitions to CAM from C_3 are placed within well-supported clades (*Agave*, *Yucca*, and *Hesperaloe*). The *Yucca* section that evolved CAM is monophyletic and strongly supported as nested within the C_3 *Yucca* species, and is therefore estimated to be an independent origin of CAM. In addition, placement of the *Hesperaloe* clade is ambiguous (Fig. 2.1), but the presence of C_3 sister species (*Schoenolirion* and *Hesperocallis*) that are strongly supported in their placement again indicates an independent origin of CAM in this lineage. And although *Polinathes* is unresolved within the genus *Agave*, it has support for its placement within the genus, rather than sister to it, strongly indicating a reversion from CAM back to the C_3 state. It is therefore unlikely that using a partially unresolved tree for ancestral state estimation biased the results in any way. Whereas further taxon sampling is not likely to improve resolution of a bifurcating tree, more complete species sampling could allow for better estimation of diversification rates in CAM and C_3 lineages.

The evolution of CAM within the Agavoideae is of particular interest due to the high diversity of habitats that species in this subfamily inhabit. Species like *Camassia* and *Hosta* inhabit mesic regions, including wet meadows for *Camassia* and shady forest understories for species of *Hosta*. In addition, although the large genera of *Agave* and *Yucca* have a center of diversity in the southwestern deserts of North America, a number of species – particularly those nearer the base of the tree – have distributions outside of the iconic desert range. These more anciently derived species' ranges may suggest that ancestrally the Agavoideae was composed of non-desert dwellers and established lineages migrated into more arid regions after an early radiation within the group. A movement into more arid regions would require that those desert regions were in place already, and that species that moved there had an ability to grow in arguably some of the harshest conditions on the planet. It is equally possible that species adapted in place as aridification happened, as well. North American deserts are thought to have arisen as recently as 2-5 Mya on a widespread scale (Moore and Jansen 2006); this date corresponds nearly identically to both the estimated age of CAM lineages in Figure 2.2, as well as previous molecular dating analyses (Smith et al. 2008). In addition, the PCA of climate variables (Fig. 2.3) indicates that although there is a level of overlap in climate space between C₃ and CAM species, there is a significant degree of separation along PC1. PC1 reflects largely temperature variation, with warmest temperature positively correlated with the first component, and coldest temperature negatively correlated it, indicating that the most arid adapted species are CAM.

The degree of overlap in the climate PCA is driven largely by the distribution of *Yucca* species, which are known to overlap, particularly when comparing species from the separate subgenera (Althoff et al. 2012). The overlap in range between C₃ and CAM *Yucca* species suggests that whereas CAM may be an adaptation to arid environmental conditions, some C₃

Yucca growing under arid conditions may not have the molecular building blocks for the evolution of CAM, and/or may have evolved alternative physiological adaptations to extreme water stress. In either case, there is currently no data that speaks to why members of *Chaenocarpa* did not subsequently evolve CAM photosynthesis. C_3 *Yuccas* are by no means marginal species in the deserts of North America; for example, the C_3 species *Yucca elata* and can be a dominant component of Sonoran desert plant communities.

The reversion to C_3 physiology from a CAM ancestor in *Polianthes* is rare case of transition away from the derived CAM state (but see Crayn et al. 2004 for an example in the Bromeliaceae). Given the high level resolution of the WorldClim database, *Polianthes* is more similar to CAM than other C_3 species in climate space, but generally grows in areas that are less arid than most *Agaves*. *Polianthes* species are native to Mexico and grow largely in pine-oak forests or grassland habitats (Solano and Feria 2007) – certainly dry and hot habitats, but not quite as extreme as species of *Agave*. Whether the shift in habitat preceded the loss of CAM or vice versa remains unknown, however *Polianthes tuberosa* retains similarities to its CAM relatives anatomically, with %IAS levels nearly as low as those in *Agave* species (Fig. 2.4, Table 2.2).

Evolution of CAM leaf anatomy

Although differences between CAM and C_3 leaves exist in the Agavoideae, particularly thickness and cell size, the number of similarities between CAM and C_3 speaks to the idea of preadaptation for the evolution of complex traits. Preadaptation has been shown to be important for the evolution of sexual mimicry in orchid pollination (Schiestl and Cozzolino 2008), in the repeated origin of a snapping claw in shrimp (Anker et al. 2006), and for the repeated evolution

of C₄ photosynthesis in the grasses (Christin et al. 2013b). In the Agavoideae, all lineages sampled for cross sectional anatomy, with the exception of *Hosta* (Fig. 2.4) and *Chlorophytum*, had a degree of 3D venation in that vascular bundles were aligned on more than a single plane. The propensity toward 3D venation was irrespective of photosynthetic pathway. The trend is particularly striking in C₃ *Yuccas*, where central portions of leaves in *Y. angustissima* and *Y. brevifolia* (Fig. 2.4) are as thick as *Agave* leaves, and in C₃ *Beschorneria*, which is the sister genus to the CAM *Agave* and shows a similar thickening of the leaf and 3D venation restricted to the midrib area. In both of these cases, the predisposition toward 3D venation potentially allowed for an increase in leaf thickness and cell succulence in these lineages. Large cells and 3D venation were likely selected for as mechanisms to store water initially, as *Beschorneria*, *Y. angustissima*, and *Y. brevifolia* inhabit the deserts of the southwestern U.S. and Mexico. CAM could then develop in both *Yucca* and *Agave* lineages without much in the way of structural change in the leaf – large cells were already in place to store malic acid nocturnally. The shared anatomical tendencies again raise the question why all species of the Agavoideae are not employing CAM, especially C₃ *Yuccas* that have both anatomical characteristics and the environmental pressure thought to be required for CAM evolution.

Additional sampling for cross sectional anatomy of leaves may elucidate patterns that are currently unclear. *Polianthes*, which has reverted back to C₃ from CAM, has larger cells than other C₃ taxa measured but smaller than its CAM *Agave* relatives; while most of the CAM anatomical traits were lost in *Polianthes*, large cells were either ancestrally retained and has not yet completely been lost, or is selected for under the semi-arid habitats of *Polianthes* in Mexico. Additional species of this genus would help resolve both if all species in *Polianthes* have reverted to C₃, and whether they all share the same loss of ancestral CAM anatomy. Near the

base of the tree, current sampling is sparse; there is a large jump to thick, large-celled leaves from *Hosta* to *Hesperaloe* (Fig. 2.4), but that comparison is missing a number of species that may help describe that transition more carefully. Sampling *Hesperocallis*, *Shoenolirion*, and *Camassia* leaves will shape our understanding of C₃ to CAM anatomical transitions at the base of the Agavoideae phylogeny. At the same time, sampling will not likely lead to inference of a bifurcating tree with straightforward mapping of transitions in anatomical traits. Nonetheless, the possibility that the last common ancestor of all extant CAM Agavoideae lineages was predisposed to the evolution of CAM may implicate ancient polyploidy (McKain et al. 2012) as a contributing factor in the evolution of a novel trait. Polyploidy has been associated with the origin of innovations in other groups including the origin of nodulation in legumes (Cannon et al. 2015) and the origin of the flower (Jiao et al. 2011; Amborella Genome Project 2013), but this speculative hypothesis remains to be tested for CAM evolution.

Conclusions

Despite difficulties in resolving the phylogeny of the subfamily Agavoideae, ancestral state reconstructions shows three independent origins of CAM in the group. These origins are associated with a shift in climate space toward warmer, drier habitats; in *Yucca*, the shift is less clear, as ranges between the C₃ and CAM clades overlap, and the driver for CAM evolution in one clade but not the other continues to be an area of interest. Assessment of leaf characters shows a propensity for the subfamily to have thick leaves enabled by 3D venation. It's possible that veins arranged in multiple planes allowed for larger cells for water storage in dry habitats. Larger cells and thicker leaves acted as a preadaptation for CAM evolution, as large cells are required for efficient storage of malic acid. Preadaptation in the Agavoideae likely assisted in the repeated evolution of CAM in three separate lineages, occurring at nearly the same time, and

allowed for shifts in habitat occupation. Further work on whether diversification rates vary between C₃ and CAM lineages will give insight into how the evolution of CAM might be promoting biodiversity in the Agavoideae.

References

- Althoff D.M., Segraves K.A., Smith C.I., Leebens-Mack J., Pellmyr O. 2012. Geographic isolation trumps coevolution as a driver of yucca and yucca moth diversification. *Mol. Phylogenet. Evol.* 62:898–906.
- Amborella Genome Project. 2013. The Amborella genome and the evolution of flowering plants. *Science.* 342:1241089.
- Anker A., Ahyong S.T., Noel P.Y., Palmer A.R. 2006. Morphological phylogeny of Alpheid shrimps: parallel preadaptation and the origin of a key morphological innovation, the snapping claw. *Evolution (N. Y).* 60:2507–2528.
- Archibald J.K., Kephart S.R., Theiss K.E., Petrosky A.L., Culley T.M. 2015. Multilocus phylogenetic inference in subfamily Chlorogaloideae and related genera of Agavaceae - informing questions in taxonomy at multiple ranks. *Mol. Phylogenet. Evol.* 84:266–83.
- Barker M.S., Dlugosch K.M., Dinh L., Challa R.S., Kane N.C., King M.G., Rieseberg L.H. 2010. EvoPipes.net: Bioinformatic Tools for Ecological and Evolutionary Genomics. *Evol. Bioinform. Online.* 6:143–9.
- Bogler D.J., Pires C.J., Francisco-Ortega J. 2005. Phylogeny of the Agavaceae based on ndhF, rbcL, and its sequences: implications of molecular data for classification. *Aliso.* 22:311–326.
- Bogler D.J., Simpson B.B. 1996. Phylogeny of the Agavaceae based on ITS rDNA sequence variation. *Am. J. Bot.* 83:1225–1235.
- Cannon S.B., McKain M.R., Harkess A., Nelson M.N., Dash S., Deyholos M.K., Peng Y., Joyce B., Stewart C.N., Rolf M., Kutchan T., Tan X., Chen C., Zhang Y., Carpenter E., Wong G.K.-S., Doyle J.J., Leebens-Mack J. 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol. Biol. Evol.* 32:193–210.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17:540–52.
- Chase M.W., Reveal J.L., Fay M.F. 2009. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. *Bot. J. Linn. Soc.* 161:132–136.

Christin P.-A., Osborne C.P., Chatelet D.S., Columbus J.T., Besnard G., Hodkinson T.R., Garrison L.M., Vorontsova M.S., Edwards E.J. 2013a. Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U. S. A.* 110:1381–6.

Christin P.-A., Osborne C.P., Chatelet D.S., Columbus J.T., Besnard G., Hodkinson T.R., Garrison L.M., Vorontsova M.S., Edwards E.J. 2013b. Anatomical enablers and the evolution of C4 photosynthesis in grasses. *Proc. Natl. Acad. Sci. U. S. A.* 110:1381–6.

Christin P.-A., Sage T.L., Edwards E.J., Ogburn R.M., Khoshravesh R., Sage R.F. 2011. Complex evolutionary transitions and the significance of c(3)-c(4) intermediate forms of photosynthesis in Molluginaceae. *Evolution.* 65:643–60.

Crayn D.M., Winter K., Smith J.A.C. 2004. Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. *Proc. Natl. Acad. Sci. U. S. A.* 101:3703–8.

Davis M.P., Midford P.E., Maddison W. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol.* 13:38.

Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.

Duarte J.M., Wall P.K., Edger P.P., Landherr L.L., Ma H., Pires J.C., Leebens-Mack J., dePamphilis C.W. 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.* 10:61.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–7.

Edwards E.J., Osborne C.P., Strömberg C.A.E., Smith S.A., Bond W.J., Christin P.-A., Cousins A.B., Duvall M.R., Fox D.L., Freckleton R.P., Ghannoum O., Hartwell J., Huang Y., Janis C.M., Keeley J.E., Kellogg E.A., Knapp A.K., Leakey A.D.B., Nelson D.M., Saarela J.M., Sage R.F., Sala O.E., Salamin N., Still C.J., Tipple B. 2010. The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science* (80-). 328:587–91.

Fernandez-Pozo N., Menda N., Edwards J.D., Saha S., Teclé I.Y., Strickler S.R., Bombarely A., Fisher-York T., Pujar A., Foerster H., Yan A., Mueller L.A. 2015. The Sol Genomics Network (SGN)--from genotype to phenotype to breeding. *Nucleic Acids Res.* 43:D1036–41.

Fisher S., Barry A., Abreu J., Minie B., Nolan J., Delorey T.M., Young G., Fennell T.J., Allen A., Ambrogio L., Berlin A.M., Blumenstiel B., Cibulskis K., Friedrich D., Johnson R., Juhn F., Reilly B., Shammass R., Stalker J., Sykes S.M., Thompson J., Walsh J., Zimmer A., Zwirko Z., Gabriel S., Nicol R., Nusbaum C. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 12:R1.

FitzJohn R.G. 2012. Diversitree : comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* 3:1084–1092.

Gentry H.S. 1982. *Agaves of continental North America*. Tuscon, AZ: The University of Arizona Press.

Good-Avila S. V, Souza V., Gaut B.S., Eguiarte L.E. 2006. Timing and rate of speciation in *Agave* (Agavaceae). *Proc. Natl. Acad. Sci. U. S. A.* 103:9124–9.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D. a, Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–52.

Griffiths H. 2013. Plant venation: from succulence to succulents. *Curr. Biol.* 23:R340–1.

Heyduk K., Trapnell D.W., Barrett C.F., Leebens-Mack J. 2015. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biol. J. Linn. Soc.:*n/a–n/a.

Howell D.J., Roth B.S. 1981. Sexual reproduction in *Agaves*: the benefits of bats; the cost of semelparous advertising. *Ecology* 62:1–7.

Huang X., Madan a. 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9:868–77.

Huxman T.E., Hamerlynck E.P., Loik M.E., Smith S.D. 1998. Gas exchange and chlorophyll fluorescence responses of three south-western *Yucca* species to elevated CO₂ and high temperature. *Plant, Cell Environ.* 21:1275–1283.

Jiao Y., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho L.P., Hu Y., Liang H., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H., Leebens-Mack J., dePamphilis C.W. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature.* 473:97–100.

Katoh K., Standley D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30:772–780.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Mentjies P., Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28:1647–1649.

Langmead B., Trapnell C., Pop M., Salzberg S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

- Li L., Stoeckert C.J., Roos D.S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178–89.
- Löytynoja A., Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U. S. A.* 102:10557–62.
- McKain M.R., Wickett N., Zhang Y., Ayyampalayam S., McCombie W.R., Chase M.W., Pires J.C., dePamphilis C.W., Leebens-Mack J. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *Am. J. Bot.* 99:397–406.
- Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics.* 30:i541–8.
- Moore M.J., Jansen R.K. 2006. Molecular evidence for the age, origin, and evolutionary history of the American desert plant genus *Tiquilia* (Boraginaceae). *Mol. Phylogenet. Evol.* 39:668–687.
- Nelson E.A., Sage R.F. 2008. Functional constraints of CAM leaf anatomy: tight cell packing is associated with increased CAM function across a gradient of CAM expression. *J. Exp. Bot.* 59:1841–50.
- Nelson E.A., Sage T.L., Sage R.F. 2005. Functional leaf anatomy of plants with crassulacean acid metabolism. *Funct. Plant Biol.* 32:409.
- Pellmyr O. 1999. Forty million years of mutualism: Evidence for Eocene origin of the yucca-yucca moth association. *Proc. Natl. Acad. Sci.* 96:9178–9183.
- Pellmyr O. 2003. *Yucca*, *Yucca* moths, and coevolution: a review. *Ann. Missouri Bot. Gard.* 90:35–55.
- Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 26:841–2.
- R Core Team. 2013. R: A Language and Environment for Statistical Computing. .
- Rashband W.S. ImageJ. .
- Ratan A. 2009. Assembly algorithms for next-generation sequence data. .
- Rentsch J.D., Leebens-Mack J. 2014. *Yucca aloifolia* (Asparagaceae) opts out of an obligate pollination mutualism. *Am. J. Bot.* 101:2062–7.
- Rocha M., Good-Avila S. V, Molina-Freaner F., Arita H.T., Castillo A., Garcia-Mendoza A., Silva-Montellano A., Gaut B.S., Souza V., Eguiarte L.E. 2006. Pollination biology and adaptive radiation of Agavaceae, with special emphasis on the genus *Agave*. *Aliso.* 22:329–344.

- Sage R.F., Christin P.-A., Edwards E.J. 2011. The C(4) plant lineages of planet Earth. *J. Exp. Bot.* 62:3155–69.
- Schiestl F.P., Cozzolino S. 2008. Evolution of sexual mimicry in the orchid subtribe orchidinae: the role of preadaptations in the attraction of male bees as pollinators. *BMC Evol. Biol.* 8:27.
- Silvera K., Neubig K.M., Whitten W.M., Williams N.H., Winter K., Cushman J.C. 2010. Evolution along the crassulacean acid metabolism continuum. *Funct. Plant Biol.* 37:995.
- Silvera K., Santiago L.S., Winter K. 2005. Distribution of crassulacean acid metabolism in orchids of Panama: evidence of selection for weak and strong modes. *Funct. Plant Biol.* 32:397.
- Simpson J., Martinez Hernandez A., Jazmin Abraham Juarez M., Delgado Sandoval S., Sanchez Villarreal A., Cortes Romero C. 2011. Genomic resources and transcriptome mining in *Agave tequilana*. *GCB Bioenergy.* 3:25–36.
- Smith C.I., Pellmyr O., Althoff D.M., Balcázar-Lara M., Leebens-Mack J., Segraves K.A. 2008. Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. *Proc. Biol. Sci.* 275:249–58.
- Smith S.D., Hartsock T.L., Nobel P.S. 1983. Ecophysiology of *Yucca brevifolia*, an arborescent monocot of the Mojave desert. *Oecologia.* 60:10–17.
- Solano E., Feria P.T. 2007. Ecological niche modeling and geographic distribution of the genus *Polianthes* L. (Agavaceae) in Mexico: using niche modeling to improve assessments of risk status. In: Hawksworth D.L., Bull A.T., editors. *Plant Conservation and Biodiversity*. Springer. p. 311–326.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 22:2688–90.
- Štorchová H., Hrdličková R., Chrtěk J., Tetera M., Fitze D., Fehrer J. 2000. An Improved Method of DNA Isolation from Plants Collected in the Field and Conserved in Saturated NaCl/CTAB Solution on JSTOR. *Taxon.* 49:79–84.
- Suyama M., Torrents D., Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–12.
- Suzuki K., Dohzono I., Hiei K., Fukuda Y. 2002. Pollination effectiveness of three bumblebee species on flowers of *Hosta sieboldiana* (Liliaceae) and its relation to floral structure and pollinator sizes. *Plant Species Biol.* 17:139–146.
- Winter K., Holtum J.A.M. 2002. How Closely Do the delta 13C Values of Crassulacean Acid Metabolism Plants Reflect the Proportion of CO₂ Fixed during Day and Night? *Plant Physiol.* 129:1843–1851.

Zambrano V.A.B., Lawson T., Olmos E., Fernández-García N., Borland A.M. 2014. Leaf anatomical traits which accommodate the facultative engagement of crassulacean acid metabolism in tropical trees of the genus *Clusia*. *J. Exp. Bot.* 65:3513–23.

Table 2.1 – Source information for both DNA tissue and for tissue analyzed for $\delta^{13}\text{C}$ ratios. Blank spaces indicates no sampling. Live accession IDs and all herbarium voucher IDs are included. DBG – Desert Botanical Garden (Phoenix, AZ), UGAG – University of Georgia Plant Biology greenhouse (Athens, GA), UGAH – University of Georgia Herbarium (Athens, GA), MOBOT – Missouri Botanical Garden Herbarium, Pellmyr – Olle Pellmyr (donation of tissue to Leebens-Mack lab), Kephart – Susan Kephart (collaborator, Willamette University), Rentsch – Jeremy Rentsch (Francis Marion University), Heyduk – self collected.

Species	DNA source	DNA accession ID	Isotope source	Isotope voucher ID
<i>Agave</i>				
<i>A. bovicornuta</i>	DBG	2004-0254-01	DBG	DES00042498
<i>A. parryi</i>	DBG	1993-0940-01-1	DBG	DES00040639
<i>A. glomeruliflora</i>	DBG	1967-9017-01-2	DBG	DES00034189
<i>A. scabra</i>	DBG	1973-0062-01-8	DBG	DES00027354
<i>A. potrerana</i>	DBG	2010-0002-01	DBG	DES00007640
<i>A. schidigera</i>	DBG	2012-0516-01-1	DBG	DES00051834
<i>A. parviflora</i>	DBG	2013-0055-01-1	DBG	DES00001554
<i>A. arizonica</i>	DBG	1989-0106-02-5	DBG	DES00030416, DES00026323
<i>A. xajoensis</i>	DBG	1987-0212-01-1		
<i>A. ocahui</i>	DBG	1988-0360-11-09	DBG	DES00034299
<i>A. sobria</i>	DBG	1962-7170-02-4	DBG	DES00037931
<i>A. tequilana</i>	DBG	1978-0505-02-4	DBG	DES00064464
<i>A. deserti</i>	DBG	2010-0007-01-4	DBG	DES00034176
<i>A. utahensis</i> var. <i>utahensis</i>	DBG	1986-0055-21-19	DBG	DES00061701
<i>A. murpheyi</i>	DBG	2003-0373-01-3	DBG	DES00032687
<i>A. delamateri</i>	DBG	1990-0243-01-1		
<i>A. angustifolia</i>	DBG	1991-0480-02-03	DBG	DES00027936
<i>A. aktites</i>	DBG	1966-8590-02-28	DBG	DES00043363
<i>A. aurea</i>	DBG	1998-0143-01-02	DBG	DES00008277
<i>A. colimana</i>	DBG	1994-0649-1005	DBG	DES00026897
<i>A. seemanniana</i>	DBG	1990-0528-01-01	DBG	DES00064477
<i>A. schotti</i>	DBG	1990-0389-02-28	DBG	DES00033044, DES00069471
<i>A. americana</i>	UGAG	n.a.	DBG	DES00026847
<i>A. cerulata</i>	UGAG	n.a.	DBG	DES00037928
<i>A. attenuata</i>	UGAG	n.a.	DBG	1983-0526-02-02 (living)
<i>A. palmeri</i>	UGAG	n.a.		
<i>Anemarrhena</i>				
<i>A. asphodeloides</i>	UGAG	n.a.		
<i>Beschorneria</i>				

<i>B. septentrionalis</i>	UGAG	n.a.		
<i>B. yuccoides</i>			MOBOT	Abiai, Mendoza, Meruda 1402
<i>Camassia</i>				
<i>C. howellii</i>	Kephart	CAHO_BBR_16M	Kephart	CAHO_BBR_14M
<i>C. quamash linearifolia</i>	Kephart	QL12		
<i>C. quamash spp. utahensis</i>	Kephart	CAQUuta_SR_15	Kephart	CAQUuta_SR_16
<i>C. quamash spp. utahensis</i>	Kephart	CAQUuta_SG_15	Kephart	CAQUuta_SG_16
<i>Chlorogalum</i>				
<i>C. pomeridianum</i>	UGAG	n.a.	MOBOT	Beard (6/2/1967)
<i>Echeandia</i>				
<i>E. spp.</i>	UGAG	n.a.		
<i>E. leucantha</i>			MOBOT	Davidse, Sousa, Linares 35381
<i>E. luteola</i>			MOBOT	Duran, Dorantes, Sima 3507
<i>E. mexicana</i>			MOBOT	McVaugh 17497
<i>E. skinneri</i>			MOBOT	Martinez 31235
<i>Hastingisa</i>				
<i>H. alba</i>	Kephart	HAAL_H70_3M	Kephart	HAAL_H70_25E
<i>Hesperaloe</i>				
<i>H. nocturna</i>	DBG	1991-0478-01-1		
<i>H. funifera</i>	DBG	1976-0106-01-11		
<i>H. campanulata</i>	UGAG	n.a.	UGAG	n.a.
<i>H. parviflora</i>	UGAG	n.a.	DBG	1972-0160-01-2
<i>Hesperoyucca</i>				
<i>H. whippeli</i>	UGAG	n.a.	DBG	2002-0325-10-4
<i>Hosta</i>				
<i>H. ventricosa</i>	UGAG	n.a.		
<i>H. venusta</i>	UGAG	n.a.		
<i>H. lancifolia</i>			UGAH	235540
<i>Manfreda</i>				
<i>M. virginica</i>	UGAG	n.a.	UGAH	251383
<i>M. undulata</i>	UGAG	n.a.		
<i>Polianthes</i>				
<i>P. tuberosa</i>	UGAG	n.a.	UGAH	242795
<i>Schoenolirion</i>				
<i>S. croceum</i>	UGAG	n.a.	UGAH	232044, 235083
<i>Yucca</i>				
<i>Y. schidigera</i>	Pellmyr	321	Pellmyr	321
<i>Y. elephantipes</i>	Pellmyr	275	Pellmyr	179
<i>Y. queretaroensis</i>	Pellmyr	311	Pellmyr	146

<i>Y. aloifolia</i>	Rentsch	45	Heyduk	24
<i>Y. filamentosa</i>	Heyduk	9.2	Rentsch	AH110409-B#90
<i>Y. schotti</i>	Pellmyr	116	Pellmyr	115
<i>Y. carnerosana</i>	Pellmyr	263	Pellmyr	260
<i>Y. jaliscensis</i>	Pellmyr	298	Pellmyr	296
<i>Y. linearifolia</i>	Pellmyr	143	Pellmyr	143
<i>Y. louisianensis</i>	Pellmyr	193		
<i>Y. schotti</i>	Pellmyr	115		
<i>Y. filifera</i>	Pellmyr	130	Pellmyr	286
<i>Y. baileyi</i>	Pellmyr	181	Pellmyr	181
<i>Y. angustissima</i>	Pellmyr	123	Pellmyr	123
<i>Y. brevifolia</i>	Pellmyr	127	Pellmyr	126
<i>Y. glauca</i>	Pellmyr	226	Pellmyr	238
<i>Y. pallida</i>	Pellmyr	156	Pellmyr	156
<i>Y. constricta</i>	Pellmyr	147	Pellmyr	147
<i>Y. capensis</i>	Pellmyr	172	Pellmyr	169
<i>Y. brevifolia</i>	Pellmyr	111	Pellmyr	125
<i>Y. elata</i>	Pellmyr	110	Pellmyr	122,236

Table 2.2 - Average values (and standard errors) for cross sectional anatomical traits in the Agavoideae. The number of images measured for each plant is indicated (“n”), although this varied by trait measured: if SE= “n.a.,” only one image was measured for that trait. Average leaf thickness, average cell area, and the area of veins are all measured in μm . %IAS is the percent of mesophyll area that is composed of airspace. Missing values were not able to be calculated due to cross section error (i.e., the full width of the leaf was not included).

Species	CAM/C	n	Avg. leaf thickness	Avg. cell area	%IAS	Planes
	3					
Agave palmeri	CAM	3	n.a.	2888.46±336.74	11.92±3.22	n.a.
Agave schottii	CAM	3	1315.43±370.63	2365.22±188.95	7.53±1.88	1.33
Beschorneria yuccoides	C ₃	3	996.74±262.52	819.35±152.28	11.97±3.99	2.33
Beschorneria yuccoides	C ₃	3	488.53±82.30	853.03±59.86	12.13±2.74	1.67
Chlorophytum rhizopendulum	C ₃	3	228.62±26.32	308.46±10.05	33.07±2.53	1.00
Chlorophytum rhizopendulum	C ₃	2	235.81±2.36	293.55±1.29	33.91±2.01	1.00
Hesperaloe funifera	CAM	3	1282.35±14152	846.19±72.56	4.26±0.23	3.33
Hesperaloe parviflora	CAM	2	1741.93±n.a.	913.64±145.65	7.56±n.a.	3.00
Hosta ventricosa	C ₃	3	306.32±25.30	649.69±59.08	26.23±3.09	1.00
Hosta ventricosa	C ₃	3	264.40±20.49	662.35±28.98	27.14±2.25	1.00
Manfreda scabra	CAM	1	634.59±n.a.	1570.33±n.a.	5.79±n.a.	1.00
Manfreda virginica	CAM	3	1556.15±145.56	673.90±51.83	10.12±1.89	2.33
Polianthes tuberosa	C ₃	3	416.00±71.40	1133.18±46.86	8.22±1.23	1.00
Polianthes tuberosa	C ₃	2	419.40±24.59	1047.76±13.42	10.73±2.08	1.00
Yucca aloifolia	CAM	3	1326.41±131.62	1652.04±179.18	13.25±1.47	2.00
Yucca angustissima	C ₃	2	1394.68±821.49	595.05±20.42	5.11±1.36	3.50
Yucca brevifolia	C ₃	2	1121.51±368.25	591.13±46.98	2.67±0.49	1.50
CAM mean±SE		19	1294.14±76.87	1484.87±195.44	8.42±1.01	2.18±0.21
C ₃ mean±SE		24	509.26±115.31	897.21±47.38	18.99±2.27	1.46±0.18
T.test C₃ vs. CAM			t=-5.95, df=35.41, p=0	t=-3.77, df=22.5, p=0.001	t=4.25, df=31.42, p=0.0002	t=-2.45, df= 32.97 p=0.02

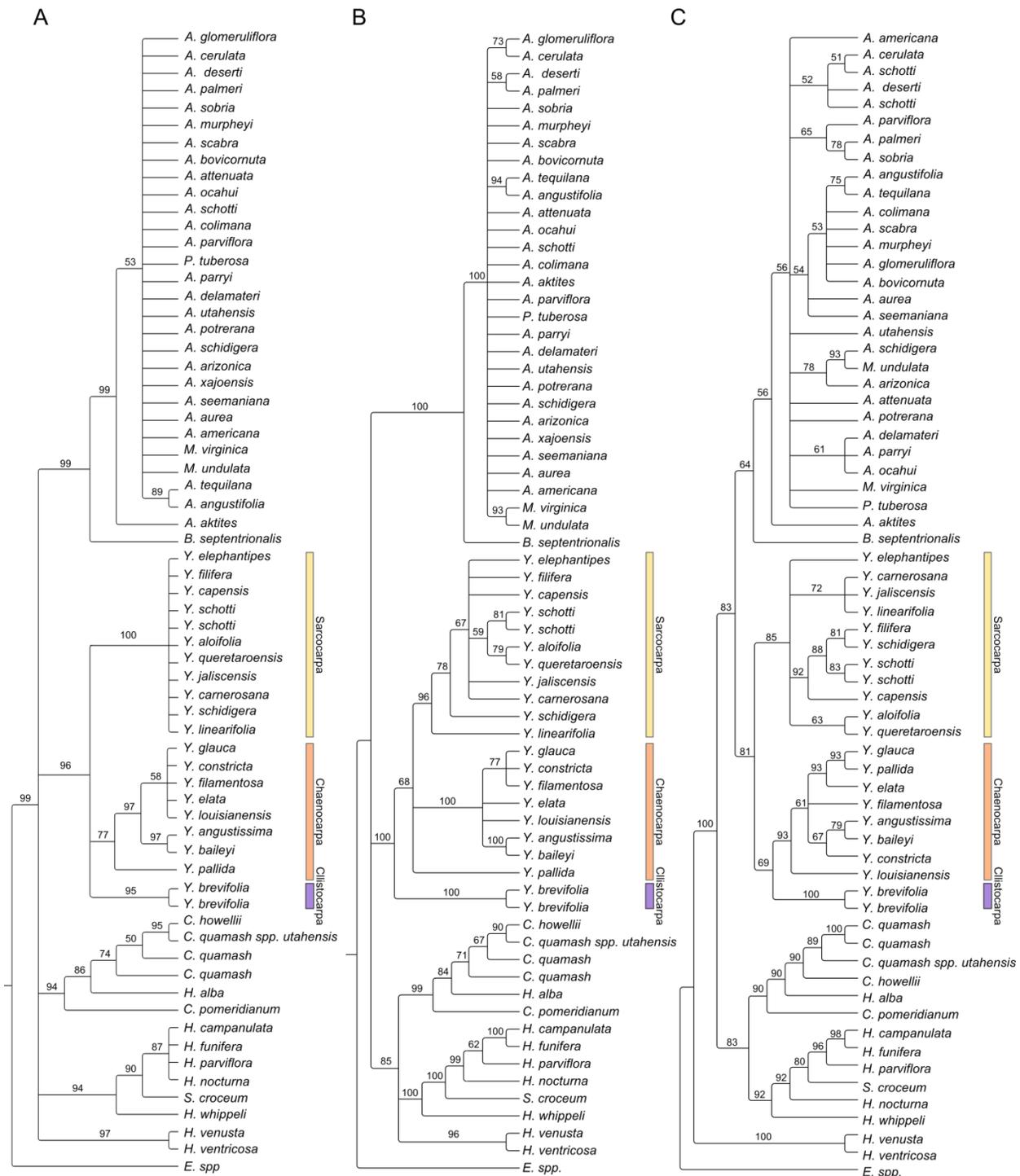


Figure 2.1 – Phylogenetic reconstruction of 66 taxa from the Agavoideae, using a) ASTRAL, b) a supermatrix analysis through RAXML, and c) the chloroplast genome of 72,220bp. Branches with bootstraps less than 50 were collapsed.

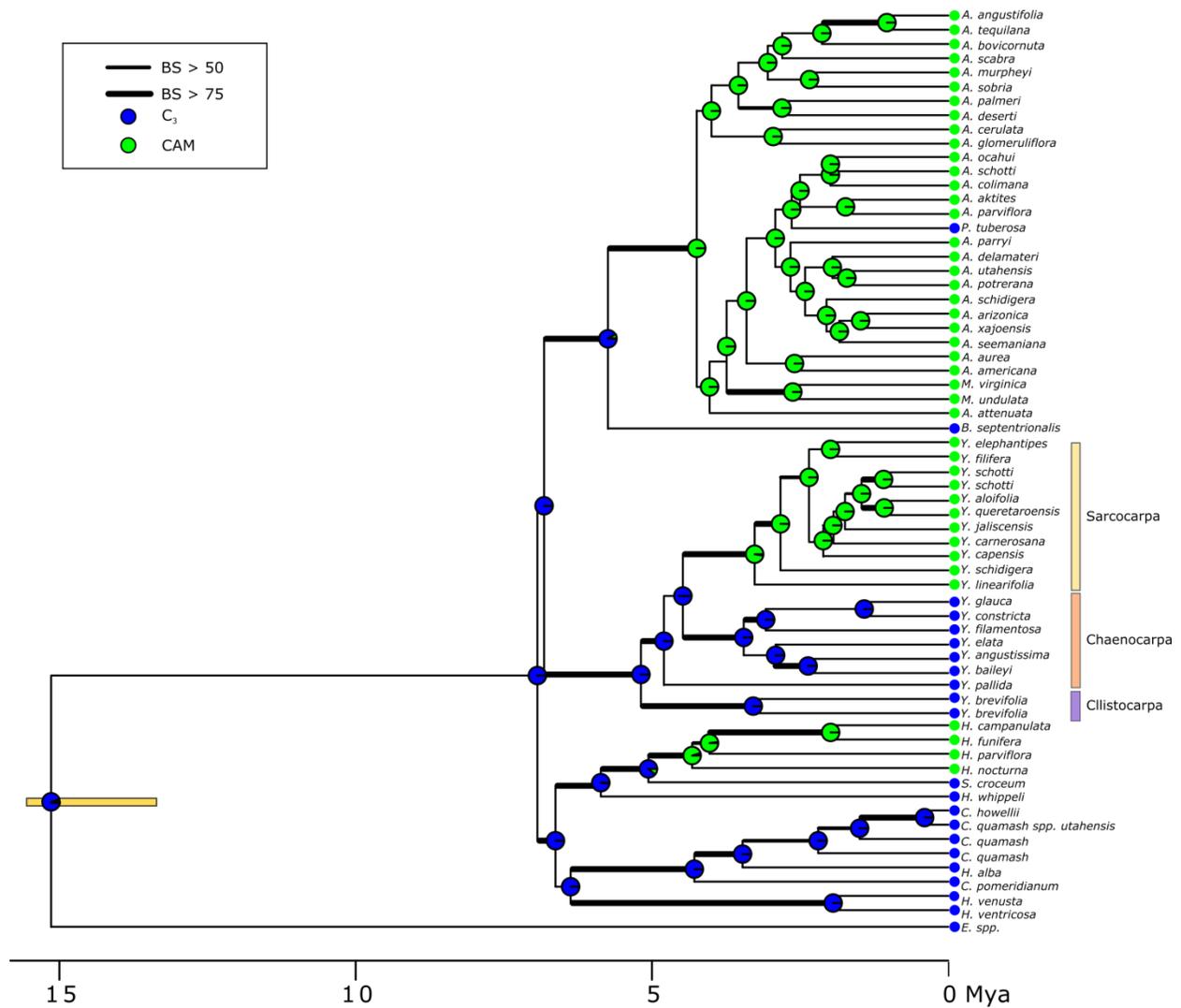


Figure 2.2 – Ancestral state reconstruction of photosynthetic pathway in the Agavoideae. Taxa presented are the same as those in Figure 2.1 with the exception of *Y. louisianensis*, which did not was used in isotopic analysis. Tree is calibrated phylogeny derived from supermatrix analysis. Blue circles are C₃, green are CAM, with the probability of each state at each node indicated in pie graphs. Nodes with greater than 75BS support in supermatrix analysis are strongly bolded, while nodes with BS support of less than 75 but greater than 50 are a medium thickness. The scale is based on a single calibration point, marked in by a bar indicating the maximum and minimum value used.

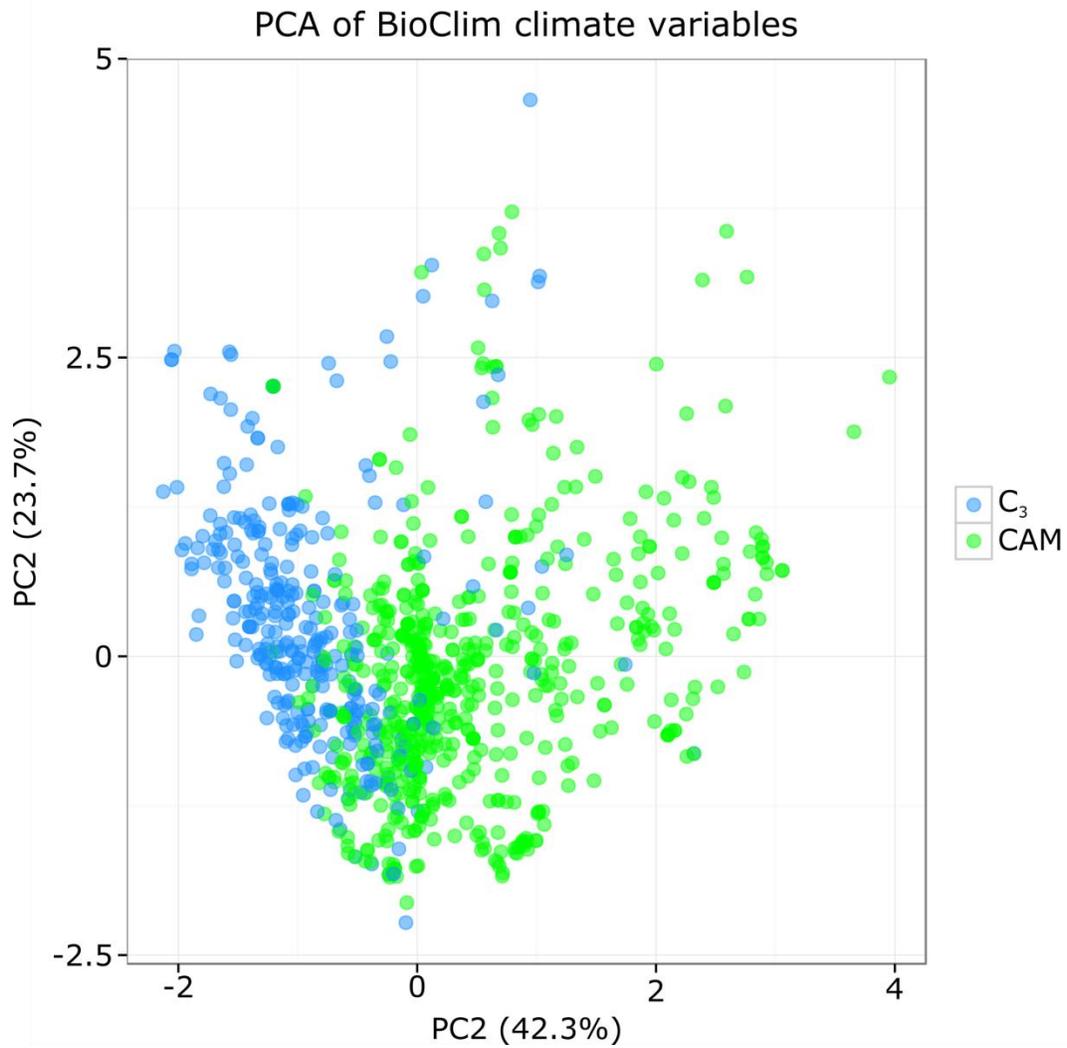


Figure 2.3 – Principal coordinates analysis of the 19 BioClim climate variables for species in the Agavoideae, coded by whether they are C₃ (blue) or CAM (green) based on isotopes assessed for an exemplar species.

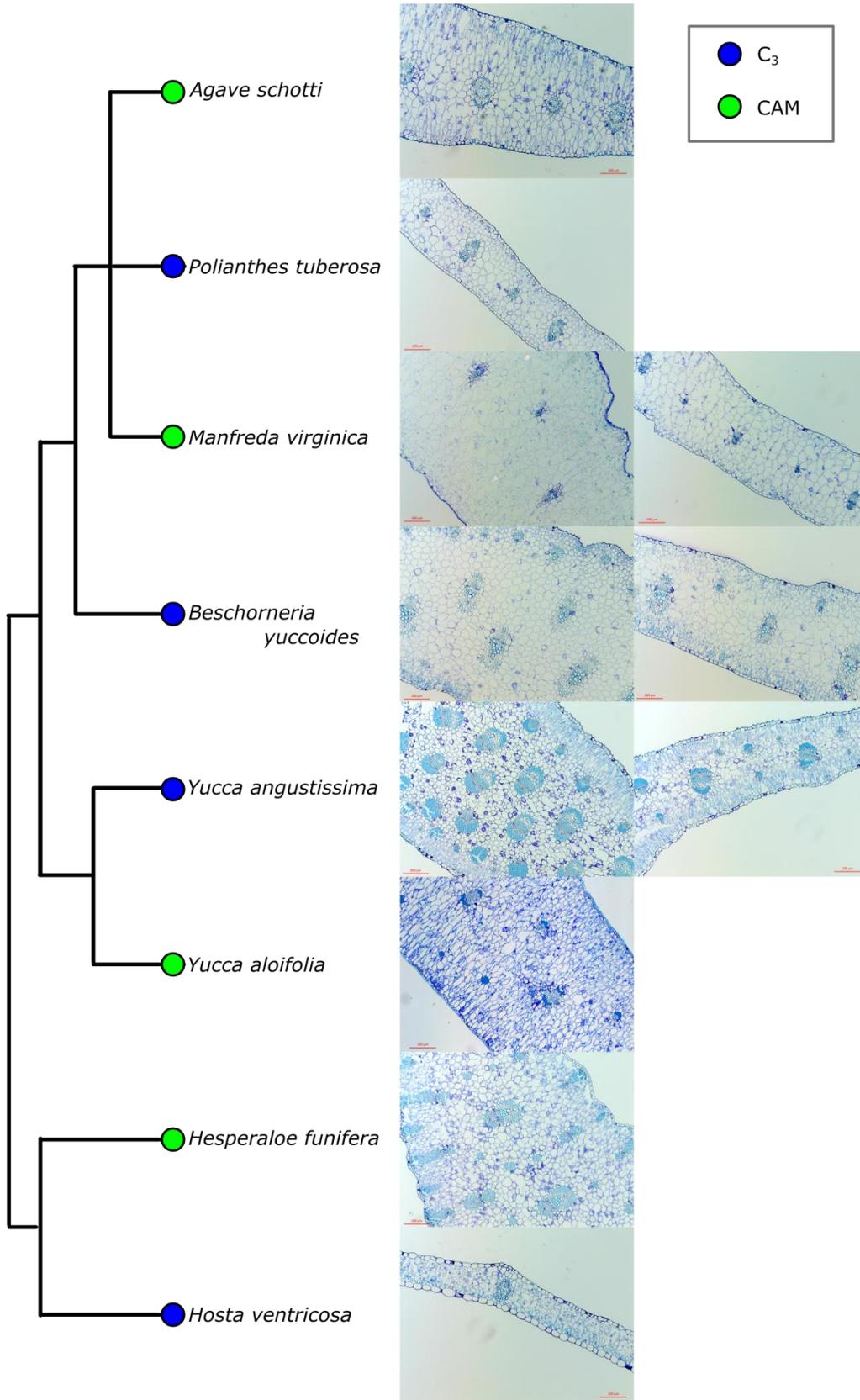


Figure 2.4 – Cross section images from select Agavoideae species. Phylogenetic tree to the left indicates relationships, and circles on tips indicate C₃ (blue) or CAM (green). For species that have high leaf heterogeneity in terms of thickness, both a central, thicker portion of the leaf is shown as well as the thinner outer leaf. Scale bar = 200μm.

CHAPTER III:

GAS EXCHANGE AND LEAF ANATOMY OF A C₃-CAM HYBRID, *YUCCA GLORIOSA* (ASPARAGACEAE)²

² Heyduk K, Burrell N, Lalani F, Leebens-Mack. J. Journal of Experimental Botany. Accepted 11/19/2015. Reprinted with publisher's permission.

Abstract

While the majority of plants use the typical C₃ carbon metabolic pathway, ~6% of angiosperms have adapted to carbon limitation as a result of water stress by employing a modified form of photosynthesis known as Crassulacean acid metabolism (CAM). CAM plants concentrate carbon in the cells by temporally separating atmospheric carbon acquisition and Calvin-Benson cycle CO₂ sugar production. CAM has been studied for decades, but the evolutionary progression from C₃ to CAM remains obscure. In order to better understand the morphological and physiological characteristics associated with CAM photosynthesis, phenotypic variation was assessed in *Yucca aloifolia*, a CAM species, *Yucca filamentosa*, a C₃ species, and *Yucca gloriosa*, a hybrid species derived from these two yuccas exhibiting intermediate C₃-CAM characteristics. Gas exchange, titratable leaf acidity, and leaf anatomical traits of all three species were assayed in a common garden under well watered and drought stressed conditions. *Yucca gloriosa* showed intermediate phenotypes for nearly all traits measured, including the ability to acquire carbon at night. Using the variation found among individuals of all three species, correlations between traits were assessed to better understand how leaf anatomy and CAM physiology are related. *Yucca gloriosa* may be constrained by a number of traits which prevent it from using CAM to as high a degree as *Y. aloifolia*. The intermediate nature of *Y. gloriosa* makes it a promising system in which to study the evolution of CAM.

Introduction

The daily fixation of atmospheric carbon dioxide is a defining trait of green plants, and is arguably the basis for the majority of terrestrial biodiversity. Plants experience a number of stresses that make the photosynthetic machinery less than optimal, including drought, shade, and high temperatures. As they cannot relocate to avoid the stress, plants have instead modified their photosynthetic pathways in ways to circumvent limitations caused by abiotic stress. One such modification, Crassulacean acid metabolism (CAM), has evolved a number of times independently across angiosperms in response to carbon limitation due to water stress. Most plants that use the C₃ carbon metabolism pathway fix CO₂ during the day via the carboxylation of Ribulose-1,5-bisphosphate by the enzyme Rubisco. However, Rubisco has both carboxylase and oxygenase activity, the latter of which is favored in high temperatures or under low CO₂ levels, resulting in costly photorespiration. For C₃ plants during times of water stress, stomata close to prevent water loss, which leads to a depletion of internal CO₂ levels and initiation of photorespiration. To simultaneously increase carbon uptake while minimizing water loss, CAM plants instead open their stomata at night, when transpirational rates are lowest. Atmospheric CO₂ is converted to a four carbon intermediate in the form of malic acid and stored in the vacuole until day time, when stomata close and the malic acid is decarboxylated (Osmond 1978; 1996). The resulting internal CO₂ concentrations, which are fixed by Rubisco behind closed stomata, are as high as 2% (Cockburn et al. 1979), compared to 0.38-0.4% outside the leaf. CAM species have estimated water use efficiency (WUE, mols of CO₂ fixed to mol H₂O transpired) from $6-30 \times 10^{-3}$, compared to $0.6-1.3 \times 10^{-3}$ for C₃ plants (Lüttge 2004).

CAM is found in about 6% of flowering plants and is distributed across 35 plant families (Cushman 2001; Silvera et al. 2010). CAM species share a suite of physiological characteristics,

including the ability to fix carbon at night, the accumulation of malic acid in the vacuoles over the dark period, increased leaf succulence due to enlarged vacuoles for the storage of malic acid, and daily carbohydrate turnover for the regeneration of phosphoenolpyruvate (*PEP*), the molecule carboxylated for the initial fixation of carbon by the enzyme PEP carboxylase (*PEPC*). Leaf anatomy has also been implicated in the evolution of CAM, with previous research indicating tight cell packing (or low intercellular airspace (IAS)) and large mesophyll cells as requirements for the optimal CAM function (Nelson and Sage 2008; Zambrano et al. 2014). Changes in these two traits may be linked responses to selection for increased vacuolar storage of malic acid in CAM species, but this hypothesis needs to be tested. Vein density has been shown to be a critical trait for the evolution of C₄ photosynthesis and is central to Kranz anatomy (Hattersley 1984; Ueno et al. 2006; McKown and Dengler 2007; Christin et al. 2013), but venation has been largely overlooked in anatomical studies of CAM plants. Interveinal distance has been found to correlate positively with succulence (Ogburn and Edwards 2013). In addition, stomatal densities in *Clusia* were found to be lower in plants with higher nighttime carbon uptake (Zambrano et al. 2014). In general, it is thought that there are tradeoffs between CAM-promoting traits and efficient C₃ photosynthesis. The optimal anatomy for CAM plants includes large cells and a decrease in the amount of IAS; these same traits would limit efficient conductance of gas throughout the leaf (Nelson et al. 2005; Nelson and Sage 2008; Zambrano et al. 2014).

Despite the extremes in phenotypes, CAM is often described as a continuum or spectrum, with full CAM at one end, C₃ at the other, and various intermediate forms between (Winter et al. 2015). Although CAM plants are defined by nighttime carbon uptake, this dogma ignores the high degree of plasticity found in CAM plants (Dodd et al. 2002). Some CAM lineages have

“weak” CAM plants, predominantly C₃ plants with very low levels of CAM expression (as measured by leaf titratable acidity) (Silvera et al. 2005). In addition, a handful of species, including those in *Clusia* and *Mesembryanthemum*, are known to be facultatively CAM, whereby they use the C₃ pathway under non-stressed conditions but can upregulate the CAM cycle in response to a variety of abiotic stresses. Although these intermediate CAM species are less common than constitutive CAM or C₃ plants (Borland et al. 2011), detailed comparative studies between C₃, CAM, and intermediate forms may help bridge the evolutionary gap between the ends of the spectrum (Silvera et al. 2010; Garcia et al. 2014; Winter and Holtum 2014). Correlations between the ability of intermediate species to use the CAM pathway and their physiology and leaf anatomy could pinpoint traits that are vital – or not – to nighttime carbon fixation.

To assess how physiology and anatomy correlate to a species’ ability to use CAM, we explored a natural hybrid system in the genus *Yucca* L. (Asparagaceae); *Yucca aloifolia* (CAM) and *Yucca filamentosa* (C₃), which are sympatric in the southeastern US and have hybridized to form *Yucca gloriosa* (Rentsch and Leebens-Mack 2012). *Yucca gloriosa* is unlikely to be a recent F1 hybrid (Rentsch and Leebens-Mack 2012), and this species may be segregating for parental phenotypes as well. The parental species and the hybrid overlap in habitats along the dunes of the southeastern coastline, but grow in different parts of the dune system and are morphologically distinct. *Yucca filamentosa* inhabits the scrub-pine forests behind the dunes, except in the northern parts of its range, where it lives on the dunes under the protection of nurse plants. *Yucca aloifolia* and *Y. gloriosa* are both foredune species, with *Y. gloriosa* typically found on the oceanside of the foredune. *Yucca gloriosa* was previously shown *in situ* to accumulate significant amounts of malic acid during the night, but showed no concurrent nighttime ¹⁴CO₂

uptake (Martin et al. 1982). The lack of nighttime carbon uptake could be an artifact of sampling methodology used, especially if gas exchange rates are low. Genotypes from each of the three species were collected from across the species' ranges and assessed for photosynthetic pathway. Plants were assayed for carbon uptake in growth chambers under well watered conditions, as well as while drought stressed, as CAM has been shown to be up-regulated under drought in various species (Ting 1985; Lee and Griffiths 1987; Winter and Ziegler 1992; Winter et al. 2011). These same *Yucca* individuals were phenotyped for titratable acidity and leaf anatomical characteristics. To better understand what suites of traits are important for CAM function, and how they impact the ability of *Y. gloriosa* to use either photosynthetic pathway, physiological traits were assessed for correlations.

Using the *Yucca* system, we investigate CAM-related trait variation between a set of sympatric C₃, CAM, and hybrid *Yucca* species. We also evaluate the extent the hybrid species *Y. gloriosa* exhibits characteristics of its C₃ and CAM parents. Phenotypic assessment shows parental species are true to type, displaying gas exchange patterns and leaf anatomical traits that are predicted by their respective photosynthetic pathways. The hybrid species *Y. gloriosa* uses both C₃ and CAM pathways to assimilate carbon and converts to fully CAM when drought stressed. Moreover, we argue, that *Y. gloriosa* can serve as a new study system for investigating the genetic architecture and evolution of CAM within the Agavoideae, a group that includes some of the most iconic CAM species.

Methods

Plant acquisition and maintenance

All three *Yucca* species are clonal and readily generate ramets from the base of larger maternal plants. In summer 2013, ramets between 10-15cm tall were collected from all three species across the southeastern US seaboard, from the Outer Banks of North Carolina to the barrier islands of Georgia (Table S3.1). Two clonal ramets were collected from each maternal plant and transplanted to the University of Georgia greenhouses in a 50:50 by volume mix of sand:pine bark (with vermiculite and limestone added). Plants were watered and fertilized as needed for at least 9 months prior to experimentation (described below). Diseased plants or those where size dimorphism between clones was very large were excluded.

Genotyping of individuals

To assess a degree of genetic variation among individuals of all three species, microsatellite markers were used from Flatz *et al.*, 2011, Rentsch and Leebens-Mack 2012, or were developed *de novo* for this study (see Table S3.2). *De novo* development of primers took advantage of unpublished sequence capture data in *Y. aloifolia* (see Heyduk *et al.* 2015 for an overview of methods). An assembled set of contigs from the sequence capture data was first analyzed for a single *Y. aloifolia* individual (Y45) through BatchPrimer3 for microsatellite discovery. These putative loci were then screened in the sequence capture assemblies from an additional five genotypes of *Y. aloifolia*. Polymorphism across all six genotypes was assessed *in silico*, and loci that varied in repeat length across the individuals of *Y. aloifolia* were kept for screening via PCR. A final screening left 7 loci from each of the three resources (Flatz *et al.*

2011, Rentsch and Leebens-Mack, 2012, this study) that amplified successfully in all three species.

DNA from individuals used in phenotypic analysis (Table S3.1) was isolated using a modified CTAB method (Doyle 1987; Štorchová et al. 2000). DNA was amplified for the microsatellite loci using a 3-primer system, where one primer (M13) is fluorescently tagged with either a FAM or HEX fluorophore, and the forward primer has an additional sequence that is complementary to the M13 primer sequence (Schuelke 2000). Loci were amplified with the following PCR mix: 3.6uL of PCR buffer (100mM Tris-HCl pH 8.0, 500mM KCl), 0.9uL of 25mM MgCl₂, 0.6uL of 10mM dNTPs, 0.4uL of reverse primer (10uM), 0.4uL of M13 (10uM), 0.375uL forward primer (2uM), 6.725uL of H₂O, 1uL of taq polymerase, and 1uL of template DNA diluted to ~10ng/uL for a total of 15uL per PCR reaction. Amplification used a touchdown program as follows: initial denaturation at 95°C for 2 min; 10 cycles of 95°C for 15 sec, 64°C for 15 sec with a 1 degree drop per cycle, and 72°C for 30 sec; 25 cycles of 95°C for 15 sec, 54°C for 15 sec, and 72°C for 30 sec; a final extension at 72°C for 1 min. PCR products were diluted 1:15, pooled when appropriate, then 3 uL was mixed with 10uL of a formamide:ROX dye-labeled size standard (1mL formamide, 100uL ROX ladder). Fragment analysis was conducted on an Applied Biosystems 3730xl DNA Analyzer. Alleles were called using Geneious 8.1.6 (Kearse et al. 2012), exported, and analyzed for hybrid index scores using the introgress package (Gompert and Buerkle 2010) in R 3.2.2 (R Core Team 2013).

Gas exchange

A total of 16 genotypes were phenotyped: three from *Y. aloifolia*, four from *Y. filamentosa*, and 9 from *Y. gloriosa*. In order to assess gas exchange patterns on a large number

of plants, genotypes from all three species were split between three independent experiments conducted in July and October 2014 and February 2015 (Table S3.1). For a given experimental period, 5-6 genotypes each with 2 clones were measured. Genotypes were randomly assigned to one of three time blocks, and individual clones of each genotype were randomly assigned to the water or drought treatment. Plants were moved into the growth chamber one week prior to the onset of any experimental treatment and all plants regardless of assigned treatment were watered daily. Growth chamber conditions were set to day/night temps of 32/17°C, with a relative humidity of 30% and day length of 12 hours. Photosynthetically active radiation (PAR) at plant level was between 400 and 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

Leaf gas exchange measurements were conducted with a LiCOR-6400XT portable photosynthesis system (LiCOR, Lincoln, NE). The first set of measurements was taken while soil for all plants was still at field capacity (day 1). Gas exchange measurements were taken every four hours for 24 hours, beginning one hour after the onset of light in the growth chamber. After the initial 24-hour interval of well-watered gas exchange measurements was complete, water was withheld from the drought treatment plants to initiate a dry down, while well watered treatment plants continued to be watered daily. Soil moisture probe measurements taken every other day were used to ensure a relatively even dry down (Table S3.3). Well watered and drought treated plants were then measured for gas exchange rates every other day for three 24-hour intervals (days 3, 5, and 7). For each measurement, plants were measured in the same order and the same leaf was used for the gas exchange measurements, unless the leaf showed signs of damage from the LiCor chamber. After the fourth 24-hour interval (day 7) of gas exchange measurements were complete, all plants were watered, and a final day of gas exchange measurements was conducted one day later (day 9).

Leaf titratable acidity

Discs of leaf tissue were taken from plants two hours before lights turned off in the evening (PM) and two hours before lights turned back on in the morning (AM). Samples were taken on the initial day while all plants were well watered (day 1), on two of the three drought-treatment days (days 5 and 7), and on the final re-watered day (day 9). Because of limited tissue, 2-3 disc samples per plant were taken. Tissue was immediately frozen in N₂ and stored at -80°C until leaf titrations commenced. Titrations were done on each sample independently by first measuring frozen weight, then boiling for 20 minutes in 20mL of H₂O. Boiled samples were allowed to cool to room temperature, then titrated with 100mM NaOH to the initial pH of the water used, which varied slightly in pH from 7.0-8.0. DeltaH⁺ was calculated as the difference in μ equivalents of H⁺ measured in the morning and the evening, with values above 0 indicating CAM activity. Student's T-tests were performed to check for deltaH⁺ values that were significantly different from zero, as well as to check for significant differences between watered and drought stressed samples on any given day.

Leaf anatomy

Leaf thickness was averaged from five replicate leaves per plant, on 1-2 plants per genotype; the midrib was avoided, as were young or very old leaves. To measure succulence, the most recent, fully mature leaf per plant was collected, cutting just above the lighter-colored petiole. Fresh leaves were immediately weighed, scanned, and dried in an oven at 60°C. The final dry weight was recorded when the mass changed less than 0.01g per day. Leaf area was calculated in ImageJ (Rashband). Succulence was calculated as grams of water per cm of leaf area: $(F.W.g - D.W.g) / (cm^2)$. Genotype succulence values were calculated by averaging across

replicate plants, and as residuals were not normally distributed, a Wilcoxon rank sum test was used to test for significant differences.

Leaf cross sections from all species (all individuals phenotyped with the LiCOR, plus additional genotypes, Table S3.1) were prepared as follows: tissue was harvested from leaves, fixed in formalin, and embedded in paraffin. Samples were then sliced with a microtome, mounted, and stained with Toluidine blue by the University of Georgia Veterinary Histology Laboratory. Slides were imaged by a Zeiss microscope using ZEN software with 1-3 images taken per slide. Leaf anatomical characteristics were measured in ImageJ. All traits were measured independently on each image, then averaged across all images for a given genotype. Cell size was measured separately on adaxial and abaxial portions of the leaf and also averaged together for a cumulative average cell size. In many CAM plants, mesophyll cells fail to differentiate into palisade or spongy forms (Gibson 1982), so these subdivisions were not taken into account when measuring cell sizes for either *Y. aloifolia* or *Y. gloriosa*. *Yucca filamentosa* does have clear mesophyll differentiation, and palisade and spongy mesophyll cells were used for adaxial and abaxial measurements, respectively. Leaf IAS was calculated by measuring the airspace in a given demarcated area of each cross section and dividing by the total area of mesophyll tissue in that area.

To address whether vein density is related to a plant's use of CAM, we measured vein spacing and the degree of 3D venation in the three species. Average distances between major and minor veins were measured horizontally only when veins were clearly in the same plane and when xylem and phloem were visibly developed. The degree of 3D venation was assessed by counting the number of independent planes of veins (major or minor) in each cross section.

Residuals of traits measured from cross sections were not normally distributed and Wilcoxon rank sum tests were used to assess species trait differences.

Stomatal density was measured using fresh tissue for the same genotypes used in gas exchange experiments many months after any treatment was imposed to avoid any effects of drought stress on stomata density. In addition, leaves that were fully developed before treatment in the growth chamber were used. Leaves were painted with varnish on both adaxial and abaxial sides. The varnish was allowed to dry, then removed by adhering tape and gently removing the imprinted varnish. Images of stomatal peels were captured using the same Zeiss microscope and ZEN software as was used for cross sections. The stomatal density was calculated as number of stomata per area of epidermal tissue. An ANOVA was used to test for effects of species, side of leaf (upper or lower), and an interaction of the two. Neither the side of leaf nor the interaction of side with species was significant ($F=1.174$, $p=0.286$ for side of leaf, $F=0.284$, $p=0.754$ for the interaction of side*species), so these were not considered in further tests of species differences, which were evaluated with Wilcoxon rank sum tests.

To assess the phenotypic space described by the traits measured, anatomical traits (with the exception of the number of planes of veins in the leaf, which was invariable) were combined with titratable acidity ΔH^+ values and gas exchange data to perform a Principle Coordinates Analysis (PCA). ΔH^+ values were averaged for both clones of a genotype measured from day 1, as all plants were well-watered. ΔH^+ values were included from drought-stressed clones from day 7 as a measure of the ability of a plant to accumulate malic acid under drought stress. As a proxy for the ability to use the CAM pathway, we summed the carbon uptake values that occurred at night during a given day-night cycle, and divided by the total carbon uptake value (the sum of all CO_2 uptake values measured for that plant across a 24 hour period). While

it excludes the carbon uptake that occurred in the 4 hours between measurements, those values are not likely to be largely different than the ones recorded, and the overall proportion calculated is an accurate estimate of relative contribution of nighttime carbon uptake. On the same matrix of phenotype data, we calculated a Spearman rank correlation matrix with Holm-Bonferroni adjusted p-values. Each trait was also tested for bimodality using Hartigan's Dip Test, as correlation results may be influenced by strongly bimodal distributions.

Results

Genetic diversity

Hybrid index scores for the hybrid ranged from ~0.35 to ~0.65, with two individuals having an index of 0.5 (Fig. 3.1A). These two individuals had loci that were monomorphic for parental alleles, indicating they are not F1 hybrid genotypes. Hybrid indices indicate *Y. gloriosa* samples used in this study were largely later generational hybrids segregating for alleles from each parent. The first two principal components of the genetic distance PCA explain 81.1% of the variation among the three species (Fig. 3.1B). The species cluster into three distinct groups; there is no evidence of ongoing back-crossing of *Y. gloriosa* to either parental species in the samples used in this study. Rather, the marker data suggest that *Y. gloriosa* is distinct from both *Y. aloifolia* and *Y. filamentosa*, and is likely on an independent evolutionary trajectory from either parental species.

Gas exchange pattern

Both parental species behaved as their photosynthetic types would predict under well-watered and drought conditions. *Yucca aloifolia* showed predominantly nighttime CO₂ uptake along with late afternoon uptake (Fig. 3.2). This pattern remained under well-watered for all five days.

Under drought-stressed conditions, carbon still entered the leaves of *Y. aloifolia* at night but photosynthetic rates were reduced, and day time uptake became negligible. For *Y. filamentosa*, well-watered plants showed no ability to take up carbon at night (Fig. 3.2). Drought-stressed *Y. filamentosa* plants likewise showed no transition to nighttime uptake, and total photosynthetic rates were reduced to nearly zero by day 5 across all time points measured. *Yucca gloriosa* showed high levels of daytime carbon uptake, but nighttime carbon gain happened under well-watered conditions as well (Fig. 3.2). As drought stress was induced in the hybrid, daytime uptake of carbon dropped to zero and net carbon gain occurred entirely in the dark, although absolute values were never as high in the hybrid as in *Y. aloifolia*. Upon re-watering, the drought stressed clones of each species began returning to their original, well-watered phenotype.

All genotypes of the parental species were consistent in the general pattern of gas exchange expected for each photosynthetic pathway. Although overall rates of photosynthetic activity may have varied between genotypes, the CAM and C₃ patterns were maintained in all genotypes of *Y. aloifolia* and *Y. filamentosa*, respectively (Fig. S3.1 and S3.2). *Yucca gloriosa* had somewhat more variable responses to drought, though nearly all genotypes used nighttime carbon uptake at low levels even under well watered conditions (Fig. S3.3). Some genotypes of the hybrid had a more pronounced level of nighttime carbon uptake under drought, though most maintained well-watered levels even while stressed.

Titrateable acidity

Yucca aloifolia had the highest levels of acid accumulation (Table 3.1) as indicated by ΔH^+ and was the only species to show a significant effect of drought on the degree of leaf acidity (day 7, Student's T-test, $df=8$, $P<0.001$). No sample of *Y. filamentosa* ever required

titration, demonstrating a complete lack of acid accumulation in this species. *Yucca gloriosa* had variable acid accumulation across time, with a species average value significantly greater than zero on days 1,7, and 9 but not 5 (Table 3.1). Some genotypes of *Y. gloriosa* had no acid accumulations (Fig. S3.4), while *Y. aloifolia* had variable but positive accumulation across all genotypes and all treatments (Fig. S3.4).

Leaf anatomy

Cross sections of each of the three species were easily distinguishable (Fig. 3.3) and almost all anatomical traits measured were significantly different between all three species (Table 3.2). *Yucca aloifolia* had the most succulent leaves, followed by *Y. gloriosa*, with *Y. filamentosa*'s leaves being the least succulent. Leaf IAS was significantly smaller in *Y. aloifolia* and was the largest in *Y. filamentosa*. Similarly, *Y. aloifolia* had the largest cells. Across species, average cell size was negatively correlated to IAS (Fig. 3.4, $R^2=0.4122$, $p<0.01$), but this relationship did not hold within a species; in *Y. gloriosa*, for example, average cell size was not correlated significantly to IAS ($R^2= 0.0737$, $P=0.42$). However, there was no significant difference in the number of stomata on either adaxial or abaxial sides of the leaf between species, although *Y. gloriosa* has a weakly significant difference ($p=0.04955$) on the adaxial leaf surface compared to *Y. filamentosa*. The average distance between major veins was not significantly different between *Y. gloriosa* and *Y. filamentosa*, and *Y. gloriosa* did not differ in distance between minor veins from either parent, although the parental phenotypes were significantly different from each other. All three species showed similar propensities for 3D venation and there were no significant difference between them (Table 3.1).

The PCA (Fig. 3.5) of all phenotypic data shows two ends of trait space defined by *Y. aloifolia* and *Y. filamentosa*. Further, *Y. gloriosa* appears slightly closer to *Y. filamentosa* but falls between the two clusters of parental species. The five traits with highest loading scores for the first principal component include adaxial and abaxial cell size (loading scores of -0.30 and -0.31, respectively), average proportion of dark CO₂ uptake under watered conditions (-0.32), maximum dark CO₂ uptake (-0.32), and leaf thickness (-0.31). The traits with highest loading scores for component two include adaxial and abaxial stomatal density (-0.53 and -0.55, respectively), the average distance between major veins (0.44), succulence (0.22), and leaf IAS (0.21). The large loading scores for stomatal densities on PC2 are driven largely by a single genotype of *Y. filamentosa*. A number of traits were correlated, including positive relationships between leaf density (succulence, thickness, and cell sizes) and maximum dark CO₂ uptake rates and proportion of CO₂ taken up at night (Fig. 3.6). DeltaH⁺ under watered conditions was positively correlated to proportion of dark CO₂ uptake both under well watered and drought conditions, but negatively correlated to maximum CO₂ uptake in the light. Negatively correlations existed between IAS and the maximum dark CO₂ uptake rate, as well as IAS and leaf thickness. Only two traits – proportion of nighttime CO₂ uptake in watered and drought conditions – were not unimodal across species; correlations between proportions of CO₂ uptake and any other trait should therefore be treated with caution.

Discussion

Photosynthesis in Yucca

Using a combination of gas exchange measurements, titratable acidity, and leaf anatomy, we verified the use of the CAM and C₃ photosynthetic pathways by *Y. aloifolia* and *Y.*

filamentosa, respectively. These species of *Yucca* crossed to form a natural hybrid species *Y. gloriosa* (Rentsch and Leebens-Mack 2012), in which previous natural surveys showed nightly acid accumulation but no detectable nocturnal gas exchange (Martin et al. 1982). Our results show that *Y. gloriosa* is an intermediate C₃-CAM species, with the ability to uptake CO₂ nocturnally but with relatively low levels of nightly acid accumulation (Fig. 3.2 and Table 3.2). Anatomically, the leaves of the three species are distinct, and *Y. gloriosa* has intermediate phenotypes for a variety of traits measured (Table 3.2). Trait values reported here for different *Y. gloriosa* genotypes are not only intermediate relative to the parental species, but also “fill” the phenotypic space between the extremes of C₃ and CAM. Individuals of *Y. gloriosa* sampled from natural populations are segregating for parental markers and none of the nine *Y. gloriosa* genotypes used in our analyses are F1’s (Fig. 3.1). These results are in agreement with an earlier study suggesting that *Y. gloriosa* is genetically distinct from either parental species and is largely evolving independent of *Y. aloifolia* and *Y. filamentosa* (Rentsch and Leebens-Mack 2012). The range of hybrid index scores for genotypes sampled in this and the earlier study raises the possibility that populations are segregating for CAM-related traits, but additional replication of individual *Y. gloriosa* genotypes is needed to assess intraspecific variation for physiological traits.

Gas exchange patterns for *Y. aloifolia* and *Y. filamentosa* are representative of their photosynthetic pathways, as are ΔH^+ values. For *Y. gloriosa*, the photosynthetic machinery is intermediate; the hybrid has daytime CO₂ uptake levels that are comparable to the C₃ parent, and while it has the ability to use the CAM pathway at low levels, its nighttime rate of CO₂ uptake and acid accumulation never reached the levels in *Y. aloifolia*. Non-zero but low CO₂ uptake rates at night are common in facultative CAM plants, as the facultative upregulation of the CAM

cycle is often coincident with abiotic stress (Winter and Holtum 2014). Since water stress decreased the magnitude of nighttime CAM in *Y. aloifolia* (Fig. 3.2), it is unsurprising that drought stress likewise limited nocturnal CO₂ uptake in *Y. gloriosa*. More important than the magnitude of nighttime CO₂ uptake is the proportion of daily carbon acquired at night relative to total carbon gain; for drought-stressed *Y. gloriosa* plants, 100% of carbon acquisition occurred during the night.

Acid accumulation in the hybrid was highly variable, though more acid accumulated in *Y. gloriosa* leaves than the C₃ parent, paralleling its ability to use the CAM cycle at night. The lack of an increase in acidification under drought stress in *Y. gloriosa* is correlated to little increase in net CO₂ uptake rates in drought stressed plants relative to well watered. The variability among genotypes and between time points in acid accumulations (Fig. S3.4) may be partly due to leaf heterogeneity and the inability to sample exact replicates in terms of leaf age and position. DeltaH⁺ levels may also be so low in the hybrid that they were not detected consistently with titration methods used. Alternatively, the ability to accumulate acid in the leaves – a proxy for CO₂ fixation by *PEPC* – may be segregating in genotypes of the hybrid, despite a relatively consistent ability across genotypes to acquire carbon via stomatal opening at night.

In species that are facultatively CAM, the use of the CAM pathway has been shown to relate to environment or seasonality (Guralnick et al. 1984; Ball et al. 1991; Borland et al. 1992; Lüttge 2006; Winter and Holtum 2007). For example, in the annual *Mesembryanthemum crystallinum*, induction of CAM occurs at the start of the dry season, with C₃ photosynthesis being the primary mode of carbon acquisition during the wet period (Bloom and Troughton 1979; Winter and Holtum 2007). Similarly, *Clusia uvitana*, a weak CAM species, was shown to increase the proportion of carbon acquired via CAM from 27% in the wet season to 42% during

the dry season (Zotz and Winter 1994). In theory, seasonal or environmentally-induced photosynthetic switching should allow plants to grow rapidly by using C_3 photosynthesis during favorable conditions, and allow these plants to continue to grow by utilizing CAM during dry conditions or seasons. Even under extreme drought conditions, plants with CAM phenotypes can shut stomata completely and keep carbon metabolism primed by recycling respired CO_2 . This process, known as CAM “cycling,” produces no net growth as net CO_2 intake is zero, but allows the photosynthetic machinery to stay active until conditions become more favorable. While drought stress forces *Y. gloriosa* from mostly C_3 carbon gain to 100% CAM carbon gain, how this plant responds to drought in its natural setting is not known. *In situ* studies are required to better describe seasonal and environmental impacts on the frequency of CAM use in *Y. gloriosa*.

Leaf anatomy

Yucca aloifolia possessed traits expected for CAM plants, including increased succulence and decreased IAS. Previous work in *Clusia*, *Annanas*, and *Kalanchoe* (Nelson and Sage 2008) showed similar relationships between IAS and strong CAM. Zambrano *et al.* (2014) found CAM *Clusia* species have thicker leaves, are more succulent, and have lower IAS than their C_3 counterparts. The correlation between leaf thickness or succulence, internal air space, and CAM across different plant lineages indicates that these traits represent fundamental requirements for the CAM cycle. An increase in succulence, and the corresponding decrease in IAS, has been hypothesized to limit conductance of gases through the mesophyll. For CAM plants that generate very high concentrations of CO_2 in the cells during daytime decarboxylation of malic acid, this limitation in conductance serves to keep captured carbon in the leaf. The intermediate levels of succulence and IAS found in *Y. gloriosa*, however, present a unique challenge to this species, as the majority of its carbon is assimilated during the day under well-watered conditions (Fig. 3.2).

Movement of CO₂ through the leaf is imperative for efficient C₃ photosynthesis, which is typically inhibited by leaf CO₂ levels. The proportion of IAS in *Y. gloriosa* measured here approaches levels found in the fully CAM *Y. aloifolia*, which would limit C₃ photosynthesis during the day in *Y. gloriosa*. In addition, stomatal density was nearly identical in *Y. aloifolia* and *Y. gloriosa*, precluding the hybrid from using a greater number of stomata to compensate for lower leaf CO₂ movement. How *Y. gloriosa* is able to conduct CO₂ at the same level as *Y. filamentosa* could be addressed by stomatal aperture size, but the sunken nature and large subsidiary cells in stomata of *Yucca* make measuring aperture size difficult. In addition, IAS in *Y. gloriosa* is higher than in *Y. aloifolia*, and may conversely be limiting the hybrid's ability to use the CAM pathway to as great of a degree as is found in *Y. aloifolia*.

While research into the vascular architecture of CAM plants is lacking, a wealth of information about venation and photosynthetic efficiency comes from the C₄ literature, an independent modification to the photosynthetic pathway. An increase in leaf vein density is the basis of Kranz anatomy for C₄ plants, enabling a high mesophyll to bundle sheath cell ratio required for efficient spatial concentration of CO₂. For CAM, it is unlikely that vein density plays a direct role in photosynthetic efficiency as it does in C₄ species; rather, the increase in succulence in CAM plants likely leads to modified venation patterns to maintain hydraulic connectivity. While increasing succulence is predicted to be correlated to a decrease in vein density and a decrease in the distance between veins (Noblin et al. 2008), succulent species have been shown to circumvent limitations to hydraulic connectivity by evolving 3D venation (Ogburn and Edwards 2013). All three species of *Yucca*, including the C₃ *Y. filamentosa*, had more than one plane of veins in their cross sections, indicating that the tendency toward 3D venation may be ancestral in *Yucca*, perhaps as a response to arid environments, and may have

allowed for the further evolution of succulent CAM species in this genus (Griffiths 2013). If propensities for thick leaves and 3D venation allow for repeated independent of origins of CAM in a lineage, understanding why *Y. filamentosa* is not CAM will be important for describing the evolutionary trajectory from C₃ to CAM.

Evolutionary implications

Recent work on transitions from different photosynthetic states has focused on using intermediate plants that have varying propensities or similarities towards one state or another. *Flaveria*, which has C₃, C₄, and C₃-C₄ intermediate species, has allowed for detailed studies on the trajectory of anatomical, physiological, and genetic changes required to evolve C₄ from C₃ (Huxman and Monson 2003; McKown and Dengler 2007). Similarly, evolutionary studies within the grasses, have described an anatomical progression from C₃ to C₄, and have shown that certain grass lineages possess pre-requisites for C₄ photosynthesis (Christin et al. 2013). While there are many evolutionary model systems in CAM plants, including facultative species in *Clusia*, *Mesembryanthemum*, *Kalanchoe*, and the *Orchidaceae*, *Yucca* holds particular promise due to closely related species that are C₃, CAM, and C₃-CAM intermediates. In addition, molecular analysis has shown that individuals of *Y. gloriosa* are later generation hybrids that segregate for molecular markers (Rentsch and Leebens-Mack, 2012, this study); photosynthetic phenotypes may segregate as well, but this hypothesis needs to be tested. While this study does not have the necessary replication of genotype gas exchange patterns to conclusively shown variation in the ability to use CAM, leaf anatomical traits do differ between genotypes, and indicate that parental species physiology is likewise segregating among individuals of the hybrid. Future replicated analyses of gas exchange patterns across time points for more genotypes of *Y.*

gloriosa will elucidate the degree of variation and genetic architecture of CAM-related traits within this species.

Finally, the intermediate nature of *Y. gloriosa* indicates it should be classified as a ‘C₃-CAM species’ according to Winter *et al.* (2015) and raises the question of whether such intermediacy represents a stable state or a point on an evolutionary trajectory toward C₃ or CAM. The traditional classifications of intermediate CAM include facultative species and CAM cyclers, although recent work in the *Orchidaceae* has described the prevalence of “weak” CAM, a largely C₃ plant with low levels of nighttime acid accumulation (Silvera *et al.* 2005). *Yucca gloriosa* does not fit into any of these traditional categories, and therefore prompts questions about the defining features of CAM plants (Winter *et al.* 2015). Certainly, the ability of *Y. gloriosa* genotypes to obtain carbon entirely via the CAM cycle under drought stress should place it onto the CAM spectrum, but where along the continuum is less clear. Using *Y. gloriosa* and other intermediate species as models for evolution from C₃ to CAM will help clarify the distribution of anatomical, physiological, and molecular traits along a CAM continuum. Better characterization of this continuum will ultimately inform understanding of CAM evolution and the potential for engineering CAM in a C₃ species (Borland *et al.* 2011, 2014; Borland and Yang 2013).

Conclusions

Assessment of photosynthetic pathway and leaf anatomy reveals *Y. gloriosa* as an intermediate C₃-CAM species, possessing the ability to rely entirely on nocturnal carbon uptake during drought stress. This species’ ability to use CAM may be limited by a number of anatomical features, including smaller cells which prohibit accumulation of malic acid and

greater intercellular airspace which might promote rapid loss of nocturnally fixed CO₂ during the day. *Y. gloriosa*'s intermediate nature poises it as an ideal system to study the evolution of CAM photosynthesis from a C₃ ancestor, and future work will focus on understanding variation among hybrid genotypes.

References

- Ball E., Hann J., Kluge M., Lee H.S.J., Luttge U., Orthen B., Popp M., Schmitt A., Ting I.P. 1991. Ecophysiological compartment of the tropical CAM-Tree *Clusia* in the Field II: Modes of photosynthesis in trees and seedlings. *New Phytol.* 117:483–491.
- Bloom A.J., Troughton J.H. 1979. High productivity and photosynthetic flexibility in a CAM plant. *Oecologia.* 38:35–43.
- Borland A.M., Barrera Zambrano V.A., Ceusters J., Shorrocks K. 2011. The photosynthetic plasticity of crassulacean acid metabolism: an evolutionary innovation for sustainable productivity in a changing world. *New Phytol.* 191:619–33.
- Borland A.M., Griffiths H., Maxwell C., Broadmeadow M.S.J., Griffiths N.M., Barnes J.D. 1992. On the ecophysiology of the Clusiaceae in Trinidad: expression of CAM in *Clusia minor* L. during the transition from wet to dry season and characterization of three endemic species. *New Phytol.* 122:349–357.
- Borland A.M., Hartwell J., Weston D.J., Schlauch K.A., Tschaplinski T.J., Tuskan G.A., Yang X., Cushman J.C. 2014. Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends Plant Sci.* 19:327–38.
- Borland A.M., Yang X. 2013. Informing the improvement and biodesign of crassulacean acid metabolism via system dynamics modelling. *New Phytol.* 200:946–9.
- Christin P.-A., Osborne C.P., Chatelet D.S., Columbus J.T., Besnard G., Hodkinson T.R., Garrison L.M., Vorontsova M.S., Edwards E.J. 2013. Anatomical enablers and the evolution of C₄ photosynthesis in grasses. *Proc. Natl. Acad. Sci. U. S. A.* 110:1381–6.
- Cockburn W., Ting I.P., Sternberg L.O. 1979. Relationships between Stomatal Behavior and Internal Carbon Dioxide Concentration in Crassulacean Acid Metabolism Plants. *Plant Physiol.* 63:1029–1032.
- Cushman J.C. 2001. Crassulacean Acid Metabolism. A Plastic Photosynthetic Adaptation to Arid Environments. *Plant Physiol.* 127:1439–1448.

- Dodd A.N., Borland A.M., Haslam R.P., Griffiths H., Maxwell K. 2002. Crassulacean acid metabolism: plastic, fantastic. *J. Exp. Bot.* 53:569–580.
- Doyle J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Flatz R., Yoder J.B., Lee-Barnes E., Smith C.I. 2011. Characterization of microsatellite loci in *Yucca brevifolia* (Agavaceae) and cross-amplification in related species. *Am. J. Bot.* 98:e67–9.
- Garcia T.M., Heyduk K., Kuzmick E., Mayer J.A. 2014. Crassulacean acid metabolism biology. *New Phytol.* 204:738–40.
- Gibson A.C. 1982. The Anatomy of Succulence. In: Ting I.P., Gibbs M., editors. *Crassulacean Acid Metabolism: Proceedings of the fifth annual symposium in botany*. Rockville, MD: American Society of Plant Physiologists. p. 1–17.
- Gompert Z., Buerkle A.C. 2010. introgress: a software package for mapping components of isolation in hybrids. *Mol. Ecol. Resour.* 10:378–384.
- Griffiths H. 2013. Plant venation: from succulence to succulents. *Curr. Biol.* 23:R340–1.
- Guralnick L.J., Rorabaugh P.A., Hanscom Z. 1984. Seasonal Shifts of Photosynthesis in *Portulacaria afra* (L.) Jacq. *Plant Physiol.* 76:643–646.
- Hattersley P.W. 1984. Characterization of C4 Type Leaf Anatomy in Grasses (Poaceae). Mesophyll: Bundle Sheath Area Ratios. *Ann. Bot.* 53:163–180.
- Heyduk K, Trapnell DW, Barrett CF, Leebens-Mack J. 2015. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society*, doi: 10.1111/bij.12551.
- Huxman T.E., Monson R.K. 2003. Stomatal responses of C3, C3-C4 and C4 *Flaveria* species to light and intercellular CO₂ concentration: implications for the evolution of stomatal behaviour. *Plant, Cell Environ.* 26:313–322.
- Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Mentjies P., Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28:1647–1649.
- Lee H.S.J., Griffiths H. 1987. Induction and Repression of CAM in *Sedum telephium* L. in Response to Photoperiod and Water Stress. *J. Exp. Bot.* 38:834–841.
- Lüttge U. 2004. Ecophysiology of Crassulacean Acid Metabolism (CAM). *Ann. Bot.* 93:629–52.

- Lüttge U. 2006. Photosynthetic flexibility and ecophysiological plasticity: questions and lessons from *Clusia*, the only CAM tree, in the neotropics. *New Phytol.* 171:7–25.
- Martin C.E., Lubbers A.E., Teeri J.A. 1982. Variability in Crassulacean acid metabolism: A survey of North Carolina succulent species. *Bot. Gaz.* 143:491–497.
- McKown A.D., Dengler N.G. 2007. Key innovations in the evolution of Kranz anatomy and C4 vein pattern in *Flaveria* (Asteraceae). *Am. J. Bot.* 94:382–399.
- Nelson E.A., Sage R.F. 2008. Functional constraints of CAM leaf anatomy: tight cell packing is associated with increased CAM function across a gradient of CAM expression. *J. Exp. Bot.* 59:1841–50.
- Nelson E.A., Sage T.L., Sage R.F. 2005. Functional leaf anatomy of plants with crassulacean acid metabolism. *Funct. Plant Biol.* 32:409.
- Noblin X., Mahadevan L., Coomaraswamy I.A., Weitz D.A., Holbrook N.M., Zwieniecki M.A. 2008. Optimal vein density in artificial and real leaves. *Proc. Natl. Acad. Sci.* 105:9140–9144.
- Ogburn R.M., Edwards E.J. 2013. Repeated origin of three-dimensional leaf venation releases constraints on the evolution of succulence in plants. *Curr. Biol.* 23:722–6.
- Osmond C.B. 1978. Crassulacean Acid Metabolism: A Curiosity in Context. *Annu. Rev. Plant Physiol.* 29:379–414.
- R Core Team. 2013. R: A Language and Environment for Statistical Computing. .
- Rentsch J.D., Leebens-Mack J. 2012. Homoploid hybrid origin of *Yucca gloriosa*: intersectional hybrid speciation in *Yucca* (Agavoideae, Asparagaceae). *Ecol. Evol.* 2:2213–22.
- Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18:233–4.
- Silvera K., Neubig K.M., Whitten W.M., Williams N.H., Winter K., Cushman J.C. 2010. Evolution along the crassulacean acid metabolism continuum. *Funct. Plant Biol.* 37:995.
- Silvera K., Santiago L.S., Winter K. 2005. Distribution of crassulacean acid metabolism in orchids of Panama: evidence of selection for weak and strong modes. *Funct. Plant Biol.* 32:397.
- Štorchová H., Hrdličková R., Chrtěk J., Tetera M., Fitze D., Fehrer J. 2000. An Improved Method of DNA Isolation from Plants Collected in the Field and Conserved in Saturated NaCl/CTAB Solution on JSTOR. *Taxon.* 49:79–84.
- Ting I.P. 1985. Crassulacean Acid Metabolism. *Annu. Rev. Plant Physiol.* 36:595–622.

Ueno O., Kawano Y., Wakayama M., Takeda T. 2006. Leaf vascular systems in C(3) and C(4) grasses: a two-dimensional analysis. *Ann. Bot.* 97:611–21.

Winter K., Garcia M., Holtum J.A.M. 2011. Drought-stress-induced up-regulation of CAM in seedlings of a tropical cactus, *Opuntia elatior*, operating predominantly in the C3 mode. *J. Exp. Bot.* 62:4037–42.

Winter K., Holtum J.A.M., Smith J.A.C. 2015. Crassulacean acid metabolism: a continuous or discrete trait? *New Phytol.*

Winter K., Holtum J.A.M. 2007. Environment or development? Lifetime net CO₂ exchange and control of the expression of Crassulacean acid metabolism in *Mesembryanthemum crystallinum*. *Plant Physiol.* 143:98–107.

Winter K., Holtum J.A.M. 2014. Facultative crassulacean acid metabolism (CAM) plants: powerful tools for unravelling the functional elements of CAM photosynthesis. *J. Exp. Bot.* 65:3425–41.

Winter K., Smith J.A.C., editors. 1996. Crassulacean acid metabolism - biochemistry, ecophysiology, and evolution. Berlin, Germany: Springer.

Winter K., Ziegler H. 1992. Induction of crassulacean acid metabolism in *Mesembryanthemum crystallinum* increases reproductive success under conditions of drought and salinity stress. *Oecologia.* 92:475–479.

Zambrano V.A.B., Lawson T., Olmos E., Fernández-García N., Borland A.M. 2014. Leaf anatomical traits which accommodate the facultative engagement of crassulacean acid metabolism in tropical trees of the genus *Clusia*. *J. Exp. Bot.* 65:3513–23.

Zotz G., Winter K. 1994. A one-year study on carbon, water and nutrient relationships in a tropical C3-CAM hemi-epiphyte, *Clusia uvitana* Pittier. *New Phytol.* 127:45–60.

Table 3.1 – Nighttime acid accumulation (DeltaH+, $\mu\text{equivalents g}^{-1}$) derived from difference in dusk and dawn titratable acidity measurements, averaged across samples for all three species with standard error. Unshaded lines are averages for plants kept well watered the duration of the experiment, grey lines are averages for plants under drought stress (which began on day 3, not sampled, and ended after day 7). Significance values indicate species by treatment measurements that are significantly different from zero by Student’s T-test: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s., not significant.

Species	Day 1		Day 5		Day 7		Day 9	
	DeltaH+	Sig	DeltaH+	Sig	DeltaH+	Sig	DeltaH+	Sig
<i>Y. aloifolia</i>	144.2±16.0	***	106.7±16.2	***	133.5±14.6	***	125.5±7.3	***
<i>Y. aloifolia</i>	135.1±6.9	***	119.8±6.7	***	48.0±14.2	**	99.5±9.1	***
<i>Y. gloriosa</i>	2.4±0.97	*	0.44±0.26	n.s.	2.0±0.71	*	1.6±0.67	*
<i>Y. gloriosa</i>	3.7±1.4	*	1.79±0.87	n.s.	1.7±1.0	n.s.	2.4±0.77	**
<i>Y. filamentosa</i>	0±0	n.s.	0±0	n.s.	0±0	n.s.	0±0	n.s.
<i>Y. filamentosa</i>	0±0	n.s.	0±0	n.s.	0±0	n.s.	0±0	n.s.

Table 3.2 – Mean and standard error for traits in the three species and significance of pairwise comparisons: YA = *Y. aloifolia*, YF= *Y. filamentosa*, YG=*Y. gloriosa*. Sample sizes are 4, 10, and 7 genotypes of *Y. aloifolia*, *Y. gloriosa*, and *Y. filamentosa*, respectively, except for succulence, which had $n=6$, 12, and 8. All units are in μm except succulence (g/cm^2) and IAS (percent of mesophyll area). Stomatal and cell size values calculated separately for adaxial (ad) and abaxial (ab) surface areas. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, N.S. not significant.

Trait	<i>Y. aloifolia</i>	<i>Y. gloriosa</i>	<i>Y. filamentosa</i>	YA vs. YF	YA vs. YG	YG vs. YF
Succulence	0.1172±0.0027	0.0699±0.0022	0.0396±0.0022	***	***	***
Stomata (mm^2)⁻¹ (ad)	75.02±5.78	125.11±17.14	79.06±5.60	N.S.	N.S.	*
Stomata (mm^2)⁻¹ (ab)	93.06±13.07	125.12±17.15	92.88±8.63	N.S.	N.S.	N.S.
Thickness	1331.29±90.97	748.23±24.20	520.00±27.67	**	**	***
IAS(%)	14.5±0.54	21.5±1.28	28.7±2.38	**	**	*
Avg. cell size ad	2124.28±169.19	1153.99±95.21	696.31±47.85	**	**	***
Avg. cell size ab	1960.99±242.04	985.08±88.52	634.18±27.50	**	**	**
Avg. dist major vein	701.56±34.37	535.59±37.08	450.86±61.73	*	*	N.S.
Avg. dist minor vein	673.01±8.04	551.02±46.42	460.37±38.87	**	N.S.	N.S.
Planes of veins	2.46±0.30	2.55±0.19	2.55±0.14	N.S.	N.S.	N.S.

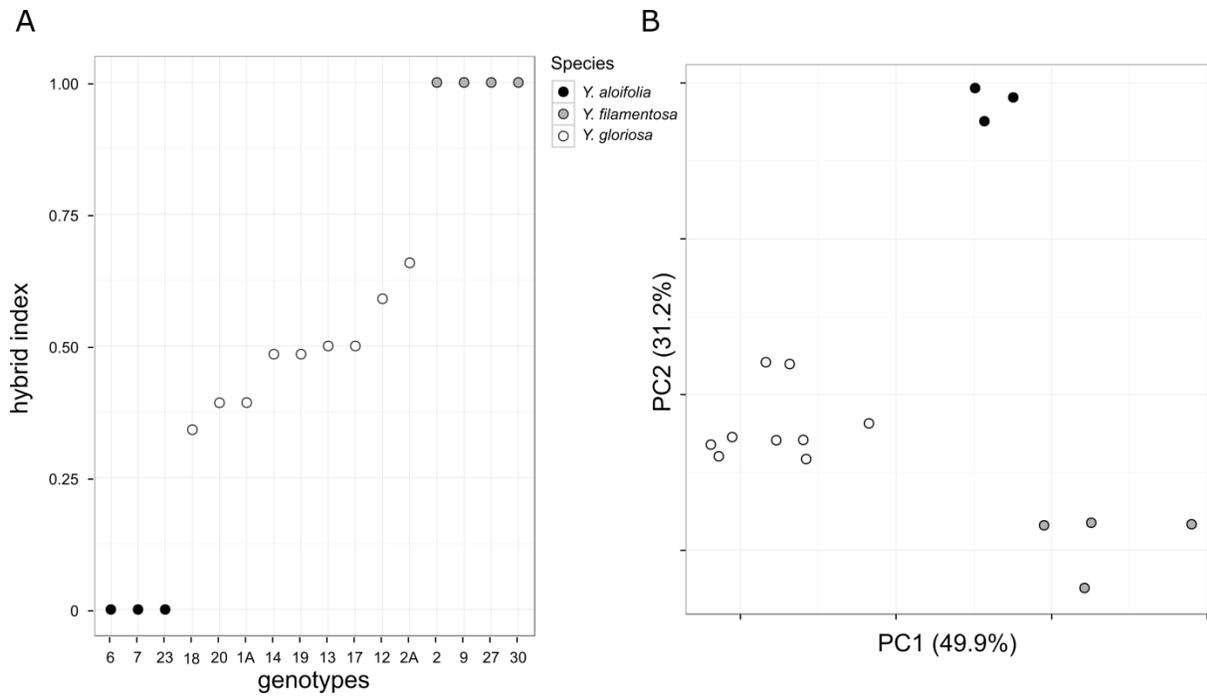


Figure 3.1 – Microsatellite variation among genotypes represented by A) hybrid index scores, with parental scores set to 0 or 1 and the index value for the hybrid species representing proportion of ancestry from each parent, and B) PCA of distance matrix of multi-locus genotypes.

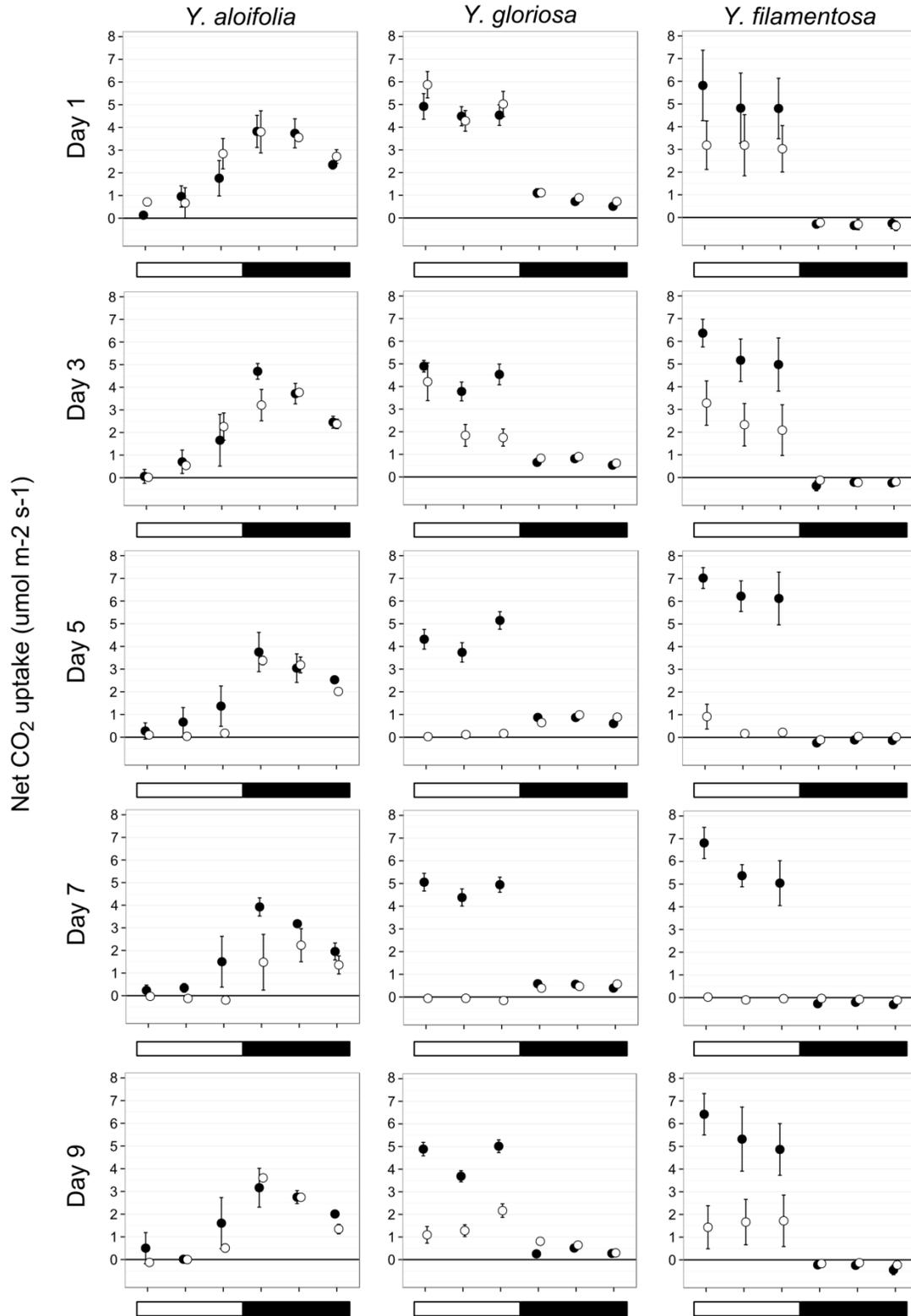


Figure 3.2 – Gas exchange patterns for *Y. aloifolia*, *Y. filamentosa*, and *Y. gloriosa* across all days of the growth chamber dry down. Filled circles indicate the clone kept under well watered conditions, open circles indicate clones which were subjected to dry down starting after Day 1.

The open bar indicates hours under light, the filled bar indicates time when lights were off. Points represent averages across all genotypes measured for a given species, with standard errors (too small to be visible for some points).

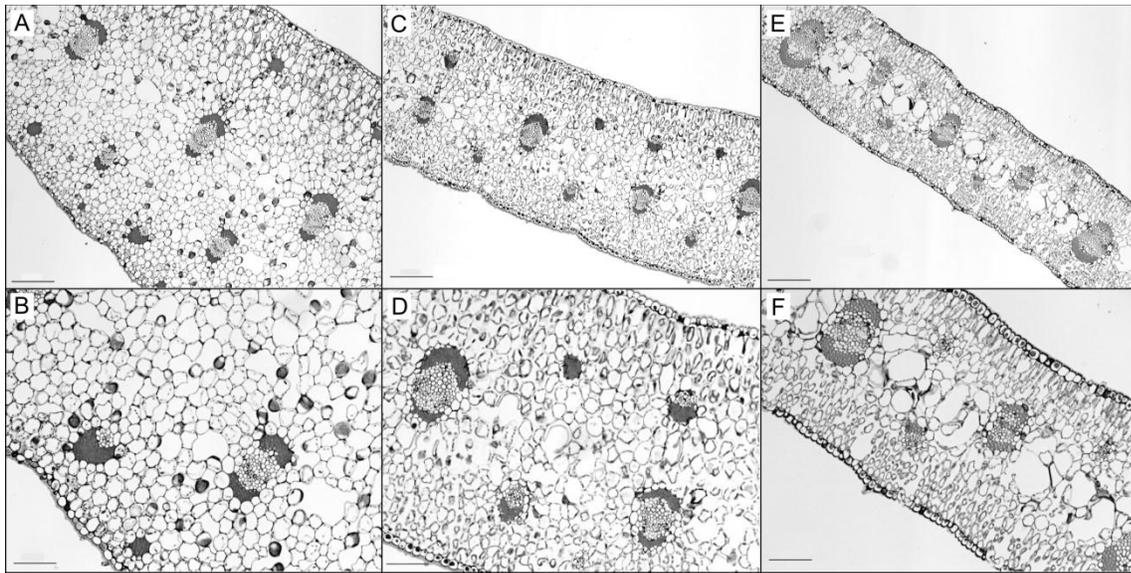


Figure 3.3 – Cross sections of *Y. aloifolia* (A and B), *Y. gloriosa* (C and D), and *Y. filamentosa* (E and F) at 5x (A, C, E) and 10x (B, D, F) magnification. Scale bars for 5x=200um and at 10x=100um.

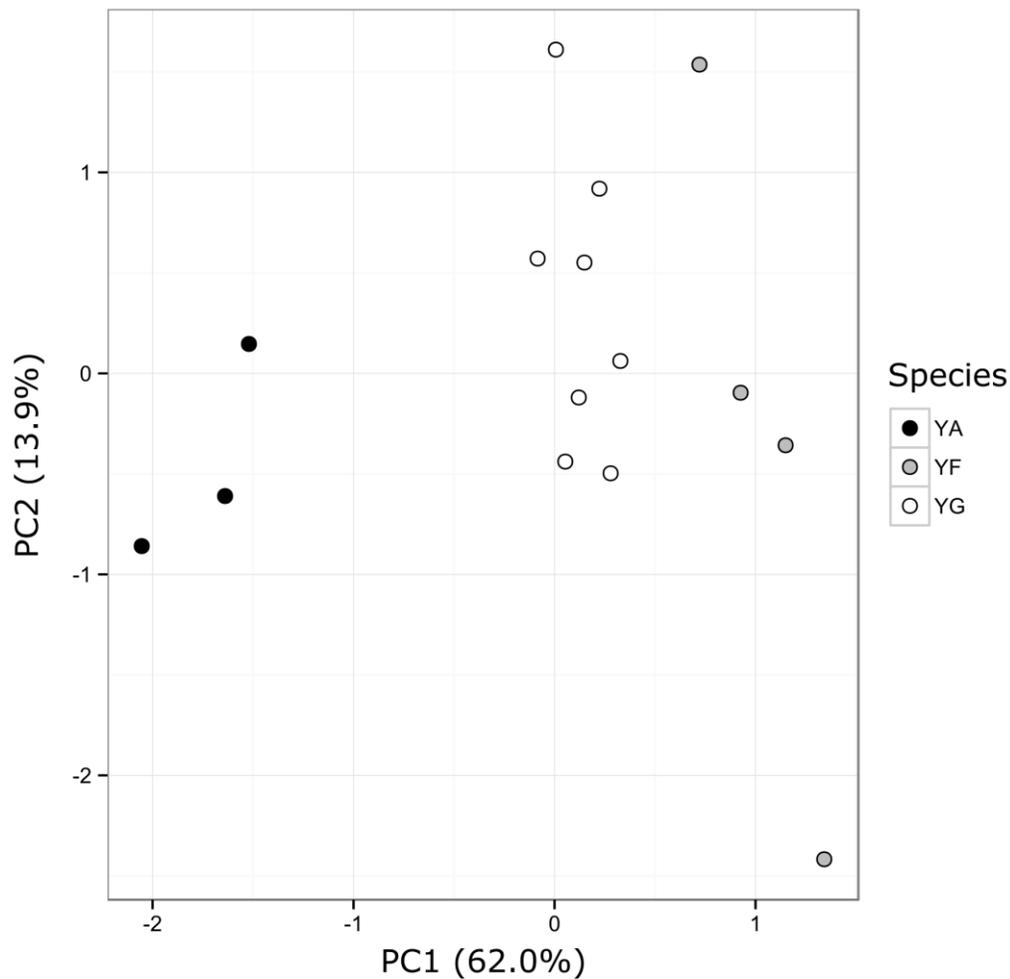


Figure 3.5 – Principal coordinates analysis of phenotypic data. *Y. aloifolia* is represented by black filled circles, *Y. gloriosa* by open unfilled circles, *Y. filamentosa* by grey circles. Phenotypes include those listed in Table 3.2: succulence, stomatal densities for adaxial and abaxial sides, leaf thickness, %IAS, average cell size on adaxial and abaxial sides, average distance between major veins, average distance between minor veins, and additionally maximum dark CO₂ uptake rate, maximum light CO₂ uptake rate, deltaH+ well watered, deltaH+ drought conditions, proportion of CO₂ taken up at night well watered and under drought.

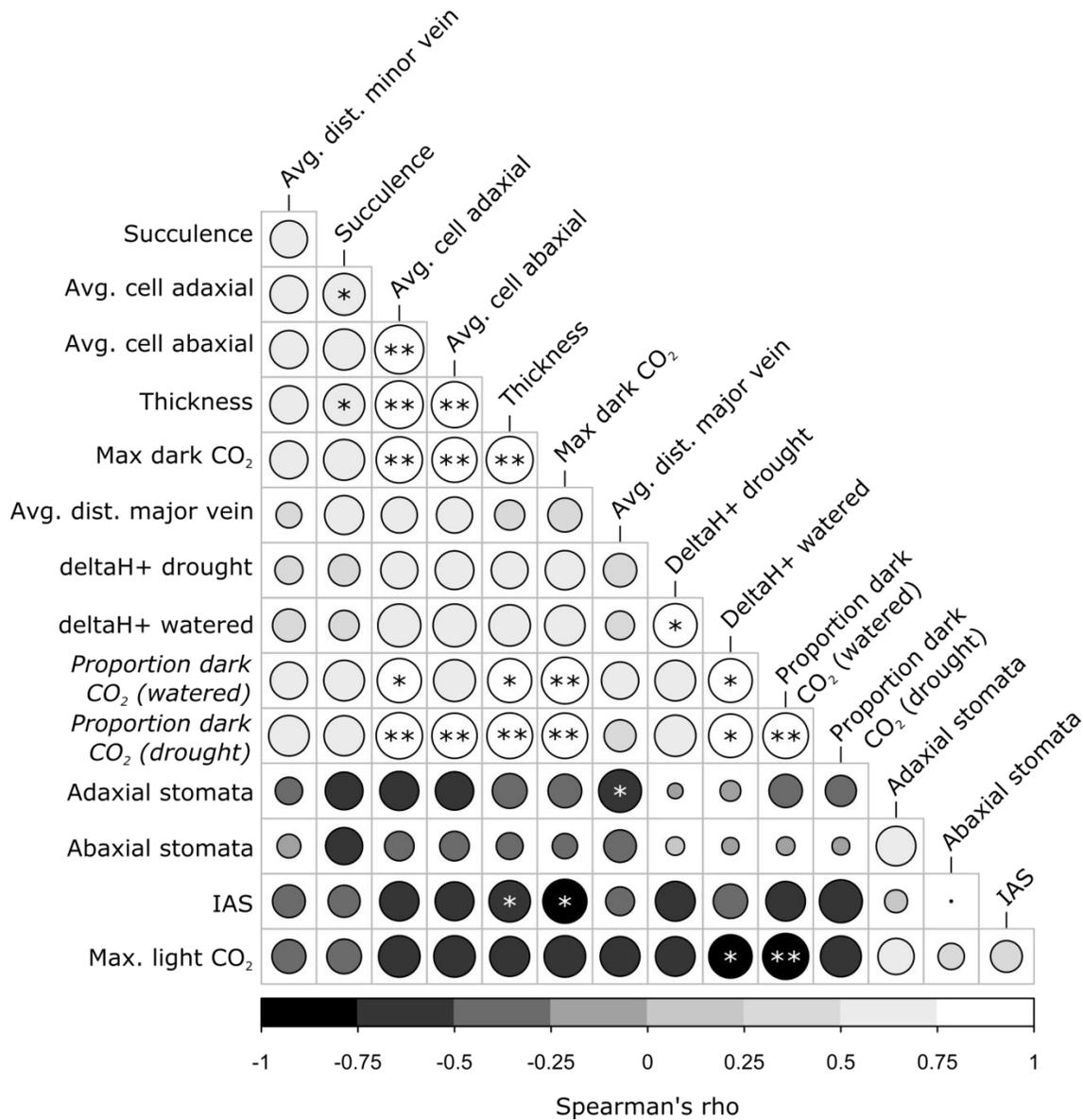


Figure 3.6 – Correlation matrix of phenotypic data, based on Spearman’s rho. Rho of -1 is black, rho of 1 is white. The size of the circles indicates the absolute value of the correlation, with larger circles referencing larger absolute correlations. Holm-Bonferoni corrected p-values: * $p \leq 0.05$, ** $p \leq 0.01$. Trait names that are italicized on the vertical axis are traits that were not unimodal according to Hartigan’s Dip Test.

CHAPTER IV:

TIME-COURSE COMPARATIVE TRANSCRIPTOMICS BETWEEN C₃, CAM, AND INTERMEDIATE *YUCCA*
SPECIES³

³ Heyduk K, Ayyampalayam S, Ray J, Leebens-Mack J. To be submitted to *New Phytologist*.

Abstract

The multiple independent origins of carbon concentrating mechanisms in plants constitute one of the most remarkable examples of parallel evolution. Though most plants use the C₃ pathway, two alternative carbon concentrating photosynthetic pathways have evolved: C₄ and Crassulacean acid metabolism, or CAM. The 6% of angiosperm species that employ CAM increase the efficiency of carbon fixation in water-limited environments by opening their stomata for gas exchange only at night when transpiration rates are lower. CO₂ is stored as a C₄ organic acid until the daytime, when decarboxylation of the acid results in high level of CO₂ around RuBisCO behind closed stomata. The ecophysiology and biochemistry of CAM plants has long fascinated researchers, and the advent of high throughput genomics has allowed for the exploration of the genetic changes required to transition from C₃ to CAM. Here we compare the transcriptomes of three *Yucca* species: *Y. aloifolia* (CAM), *Y. filamentosa* (C₃), and a hybrid that resulted from their natural crossing, *Y. gloriosa* (C₃-CAM intermediate). Using RNA-seq, we sequenced each species under well-watered and drought conditions across 6 timepoints, taken every four hours over a 24 hour diel-cycle. Using correlated network analyses, the three species transcriptomes were compared to each other in terms of gene network correlation to carbon uptake during the night, gene content of coexpressed networks, and overall gene expression of canonical CAM genes. The findings of my work indicate that although three species are divergent in their photosynthetic pathway, all three have expression of some CAM genes in a Circadian-controlled manner suggestive of CAM activity. Other CAM genes, like carbonic anhydrase and enzymes implicated in daytime decarboxylation, had unclear expression patterns that throw into question their role in CAM functioning. Together these data indicate the ancestral

evolution of CAM-like expression in C₃ *Yucca* species and implicate at least another layer of regulation over nighttime CO₂ fixation.

Introduction

Crassulacean acid metabolism (CAM) is a mode of photosynthesis whereby plants open their stomata largely at night and initially assimilate carbon using phosphoenolpyruvate carboxylase (PEPC) rather than RuBisCO (Fig. 4.1). Carbon is temporarily stored as malic acid within the vacuole, and only when stomata close during the day is the malic acid decarboxylated, resulting in high concentrations of CO₂ around RuBisCO. The extra steps of CAM – carboxylation of PEP, decarboxylation of malic acid, transport into and out of the vacuole - are energetically more costly than C₃ photosynthesis, but CAM plants have the advantage of acquiring carbon at a lower water cost (increased water use efficiency, WUE). In addition, RuBisCO is able to act more efficiently with a high concentration of CO₂ and the risk of photorespiration is significantly minimized. CAM plants are thought to be adapted to arid regions of the world where water stress is unavoidable and where the energetic cost of CAM is offset by the carbon gains. CAM has evolved at least 35 independent times in flowering plants (Silvera et al. 2010), thus making it a remarkable example of parallel evolution of a complex trait.

CAM function in plants requires a suite of characteristics to be in place, from physiological (see Ch. III) and biochemical to genomic. CAM has been studied from a biochemical standpoint for decades, and much is known about the metabolic turnover, starch cycling, and enzymatic machinery of CAM plants (Cushman and Bohnert 1997; Chen et al. 2002; Dodd et al. 2002). Similarly, physiological studies of CAM plants have revealed the

importance of succulence and large cells (Kluge and Ting 1978; Nelson et al. 2005; Nelson and Sage 2008; Zambrano et al. 2014). The genotypic changes required to go from C₃ to CAM are less well-studied, although an understanding of the basic machinery that must be in place has been developed based on biochemical and physiological work. At the beginning of the night period, carbonic anhydrase (CA) converts CO₂ to HCO₃⁻ (Fig. 4.1). PEPC fixes the carbon from CA into oxaloacetate (OAA), but its activity is regulated by a dedicated kinase, PEPC kinase (PPCK). Phosphorylated PEPC is able to fix carbon in the presence of its downstream product, malate, whereas the un-phosphorylated form is sensitive to malate (Nimmo 2000; Taybi et al. 2000). As day approaches, PPCK is down-regulated either because its Circadian regulator is turned off (Carter et al. 1991; Hartwell et al. 1996) or because high levels of cytosolic malate repress its expression (Borland et al. 1999), leaving only un-phosphorylated, malate-sensitive PEPC in the cytosol. During the day, the stored malic acid is removed from the vacuole and decarboxylated by either phosphoenolpyruvate carboxykinase (PEPCK) or NADP/NAD-malic enzyme, and it is thought that CAM plants use one or the other, though research indicates both types may be operating in a single plant (Holtum and Osmond 1981). NADP/NAD-me CAM plants additionally have high levels of PPK, which converts pyruvate to PEP. This final step is important for CAM plants, as PEP is then used in the gluconeogenesis pathway to synthesize soluble sugars. At night, those sugars must be re-mobilized to form PEP as the substrate for PEPC.

Recent studies have profiled expression before and after CAM induction in *Mesembryanthemum crystallinum* (Cushman et al. 2008) or in green (CAM) and white (non-photosynthetic) parts of the leaf blade in *Annanas comusus* (Zhang et al. 2014, Ming et al in press). Species of *Kalanchoe* have been developed for transgenic experiments, where gene

knockouts of PEPCK and NADP/NAD-me show reduced growth and photosynthetic ability in constitutive CAM plants (Dever et al. 2015). All genetic/genomic studies published to date involve comparisons between C₃ and CAM tissue types sampled from different species, or the same species under different age or environmental conditions (Taybi et al. 2004; Cushman et al. 2008; Gross et al. 2013), but even these are limited in number. While CAM is often described as a continuum of phenotypes (Silvera et al. 2010), ranging from strong CAM species on one end to C₃ on the other, the comparative analysis of strong CAM and C₃ species can provide the most pronounced contrast of traits and thus the molecular mechanisms contributing to CAM anatomy and physiology. At the same time, plants that are able to facultatively use C₃ or CAM as the environment dictates are especially useful for understanding how and when C₃ plants switch to CAM from a transcription perspective (Cushman et al. 2008; Winter and Holtum 2014).

Here I focus on interspecific comparisons among a set of C₃, CAM, and C₃-CAM intermediate *Yucca* species, using transcriptomics to characterize gene expression levels and gene interaction networks including canonical CAM genes. Gene expression and physiology were assayed in a C₃-CAM hybrid species, *Yucca gloriosa*, alongside its parental species *Y. aloifolia* (CAM) and *Y. filamentosa* (C₃). All three species have been previously characterized with respect to their photosynthetic pathway (Ch. III): *Y. aloifolia* consistently had nighttime uptake of CO₂ with concomitant malic acid accumulation in leaf tissue, as well as anatomical characteristics indicative of CAM function; *Y. filamentosa* had typical C₃ leaf anatomy and showed no positive CO₂ uptake at night or malic acid accumulation; the hybrid species *Y. gloriosa* was largely C₃ under well watered conditions, but when drought stressed transitions to 100% nighttime carbon uptake. We grew clones of all three species in a common garden setting

under both well watered and drought stressed conditions to assay transcriptional differences between species under different environmental conditions over a 24-hour diel cycle.

Methods

Plant material and RNA sequencing

RNA was collected during experiments described in Chapter III. Briefly, clones of the three species of *Yucca* were acclimated to growth chambers with a day/night temperature of 30/17°C and 30% humidity. One clone was kept well-watered for 10 days while the second clone was subjected to drought stress via dry down beginning after the end of day 1. On the 7th day of the experiment, after plants had water withheld for 6 days and soil water content dropped from ~20% under well watered to ~5% under drought (see Table S3.3), RNA was sampled every four hours beginning one hour after lights turned on, for a total of 6 time points. RNA sampling was conducted on 3 genotypes of *Y. aloifolia* and 4 genotypes of both *Y. filamentosa* and *Y. gloriosa* (Table 4.1). Due to size limitations, genotypes from the three species were randomly assigned to three different growth chamber dates: July 2014, October 2014, and February 2015. RNA was isolated from a total of 130 samples (n=36, 47, and 47 for *Y. aloifolia*, *Y. filamentosa*, and *Y. gloriosa*, respectively) using Qiagen's RNeasy mini kit (www.qiagen.com). DNA was removed from RNA samples with Ambion's Turbo DNase, then assessed for quality on an Agilent Bioanalyzer 2100. RNA libraries were constructed with 1ug of input RNA using Kapa Biosystem's stranded mRNA kit and a dual-index barcoding scheme. Libraries were quantified via qPCR then randomly combined into 4 pools of 30-36 libraries for PE75 sequencing on the NextSeq 500 at the Georgia Genomics Facility.

Assembly and read mapping

Reads were cleaned using Trimmomatic (Bolger et al. 2014) to remove adapter sequences, as well as low-quality bases and reads less than 40bp. Reads that were unpaired were also removed from analysis. After cleaning and retaining only paired reads, *Y. aloifolia* had 439,504,093 pairs of reads, *Y. filamentosa* had 675,702,853 pairs of reads, and *Y. gloriosa* had 668,870,164 pairs. Due to the sheer number of reads for each species, a subset of reads were used to construct reference assemblies for each species. This prevented erroneous reads from piling up to create false support for a mis-sequenced basepair, and allowed for more efficient assembly (Haas et al. 2013). To take a subset of reads, 14% of the total reads for each species was used, resulting in an average of 83 million pairs of reads per species for assembly. Trinity v. 2.0.6 (Grabherr et al. 2011) was used for digital normalization as well as assembly. The full set of reads from each species library were mapped back to that species' transcriptome assembly with Bowtie (Langmead et al. 2009); read mapping information was then used to calculate transcript abundance metrics in RSEM v.1.2.7 (Li and Dewey 2011; Haas et al. 2013). Trinity transcripts that had a calculated FPKM < 1 were removed, and an isoform from a component was discarded if less than 25% of the total component reads mapped to it.

To further simplify the assemblies and remove assembly artifacts and incompletely processed RNA reads, the filtered set of transcripts for each species was independently sorted into orthogroups that circumscribe gene families from 22 sequenced land plant genomes. The inferred coding sequences filtered through Transdecoder (<http://transdecoder.github.io/>), which searches for open reading frames in assembled RNA-sequencing data, were sorted in gene families as circumscribed by the Amborella Genome Project (2013). Coding sequences for each species were individually matched to a protein database derived from gene models from the 22

genome dataset using BLASTx (Altschul et al. 1990). Best hits for each query sequence were retained and were used to sort the *Yucca* transcript into the same orthogroup as the query sequence. Assembled *Yucca* sequences were further filtered to retain only putatively full length sequences; *Yucca* transcripts that were shorter or longer than the minimum or maximum length, respectively, of an orthogroup were removed. Transdecoder produces multiple reading frames per transcript, so only the longest was retained. Read counts for the final set of orthogrouped transcripts were re-calculated and analyzed in EdgeR (Robinson et al. 2010).

Species' expression patterns were assessed for outlier libraries, which were subsequently removed. A single outlier for *Y. aloifolia* was removed, as it was divergent from its biological replicates as determined by a multidimensional scaling plot of all data (not shown). For a qualitative description of gene family gene expression between species, read counts for the longest transcript per species was used as the gene family expression value for that species, then counts per million (cpm) were calculated for each library. In a given *Yucca* species, all libraries were separated by time point, then Student's T-test between treatments (watered and drought) was conducted in EdgeR to find the number of up and down regulated genes at each time point, using a p-value cutoff of 0.05 and adjusting for multiple testing with a Holm-Bonferroni correction.

Weighted gene correlated network analysis

The R package WGCNA (Langfelder and Horvath 2008) was used to construct species-specific correlated networks using counts per million normalized read counts, as computed in EdgeR. Modules represent gene interaction networks comprising genes with highly correlated expression patterns, as determined by an adjacency matrix of gene expression similarity across

all libraries of interest. First, the similarity between genes is calculated as the absolute value of the correlation coefficient between two genes i and j :

$$s_{ij} = |cor(x_i, x_j)|.$$

The weighted adjacency between genes i and j is calculated as:

$$a_{ij} = s_{ij}^\beta$$

where β is the soft threshold parameter. As opposed to a ‘hard’ threshold τ , which requires $s_{ij} > \tau$ for $a_{ij} = 1$, a soft threshold instead raises similarity to the power β , and therefore more heavily weights highly correlated genes (those with a high s_{ij}) to have values closer to 1 (Langfelder and Horvath 2008). The value of β was determined by calculating a range of threshold values and plotting them against the R^2 of the scale-free model fit for that power. For all three species eight was chosen as the soft threshold, as increasing past a threshold of eight did not yield a higher R^2 of the model. As WGCNA does not allow for more than 5000 genes to be analyzed in a single module estimation step, modules were calculated using the blockwise function in WGCNA, which allows for subsets of genes <5000 to be clustered before the entire dataset is finally merged. Blockwise estimation has been shown to result in no major loss of information compared to a single step estimation (Langfelder and Horvath 2008). A module required at least 30 transcripts in order to be recognized as a distinct module.

For every library in this dataset there exists a value of CO₂ uptake, measured with a LiCOR 6400XT at the same time that RNA was harvested (see Ch. III). These values, which are expressed in μmol of CO₂ per m² per second, represent the amount of carbon fixation occurring in the leaves; because these gas exchange measurements were conducted every four hours day

and night, they can be used to infer the degree of CAM a plant uses by examining nighttime CO₂ uptake rates (C₃ plants will have none, or negative values due to respiratory loss of CO₂). Two phenotype columns were created – one for nighttime uptake, a separate one for daytime carbon fixation – as different genes are expected to be correlated with each. Gas exchange values were correlated to modules as well as genes to narrow which modules and specific genes are related to the ability to use the CAM pathway. Each module's eigengene – a summary expression profile for all genes in a module – was calculated for each library and averaged for all libraries in a time point, irrespective of drought or watered treatment. Gene family representation was compared between the top three modules that correlate to nighttime CO₂ uptake across all three species. The average connectivity (k) within a module (defined as the sum of all adjacency values between a gene and all other genes in a module), the correlation to CO₂ uptake rate, and the p-value of that correlation were calculated for every transcript.

CAM genes

All transcripts used in the cluster analysis were screened for CAM-related annotations, particularly for the major genes involved in the CAM pathway (Fig. 4.1). These included PEPC, the initial carbon-fixing enzyme in CAM plants, as well as its regulator PPCK. Carbonic anhydrase (CA), which initially converts CO₂ to HCO₃⁻, NADP-me and PEPC (both of which convert malic acid back to pyruvate and CO₂), and PDK (converts pyruvate to PEP). Annotations used were based on the orthogroup annotation, all of which were annotated against Pfam, gene ontology (GO), and *Arabidopsis* (TAIR10) databases. Modules that contained any of these genes were compared across the three species for the gene family content. Transcripts with CAM annotations were plotted across time, with average FPKM and standard error across biological replicates reported. Well-watered and drought stressed plants had average and

standard errors calculated separately. CAM-annotated transcripts were aligned for each of the six genes studied in MUSCLE (Edgar 2004), and gene trees were estimated in RAxML (Stamatakis et al. 2008) with 500 bootstraps and a GTRGAMMA model, using sequences from *Elaeis guineensis* as an outgroup (Singh et al. 2013). Relationships among gene copies from all three species were compared to the species-specific modules which housed those genes.

Results

Assembly and gene family circumscription

After an initial pass of filtering to remove lowly expressed transcripts or those that were minor isoforms (<25% total component expression), the average of 55k transcripts remained per species. Transcripts that were retained after matching to the 22 plant genomes used for gene family circumscription (Amborella Genome Project 2013) were much reduced: 19,320 for *Y. aloifolia*, 23,362 for *Y. filamentosa*, and 21,716 for *Y. gloriosa*. Removing transcripts that were out of bounds of the minimum and maximum length for their respective gene family, as well as removing multiple coding sequence predictions for a given transcript, reduced final transcript numbers: 14,185 transcripts remained for *Y. aloifolia*, 14,376 for *Y. filamentosa*, and 13,313 for *Y. gloriosa*. A total of 6,004 gene families were represented in all three species (Fig. 4.2A); this subset of shared orthogroups was no different in terms of gene content than the entire set of gene families found across all three species (Fig. 4.2B). Expression across shared gene families shows strong contrasts between species (Fig. S4.1). Patterns of differential expression between watered and drought-stressed plants were drastically different between the three species (Fig. 4.3), with the largest change occurring one hour after lights turned on for *Y. aloifolia* and just before the lights turned on in *Y. filamentosa*. *Yucca gloriosa* had a larger number of differentially expressed

genes at nearly every time point, indicating a larger transcriptional response to drought stress in these genotypes. The variation in differentially expressed genes across species could be attributed to changes in variability between replicates at a given time point. However, the biological coefficient of variation (BCV), which measures the variation in gene expression between replicate samples that is due to biological – versus technical – causes, was similar in all three species (*Y. aloifolia* = 37%, *Y. filamentosa* = 51%, *Y. gloriosa* = 44%). As expected, BCV of replicates at a time point was not significantly correlated with the total number of genes that were differentially expressed in that time point ($R^2=0.023$, $p=0.2528$), implying that the variation in the number of differentially expressed genes across time points is more likely related to biology of each species rather than variation among biological replicates.

Correlated network analysis and CAM genes

Modules, representing inferred gene interaction networks comprised of coexpressed genes, varied in size and number between the three species. WGCNA modules are arbitrarily named by color, so modules of the same name in different species are not homologous, as each species was clustered independently. The ten most correlated modules to nighttime carbon uptake for each species are shown in Table 4.2, with the number of transcripts in each module, their correlation to carbon uptake rates and the p-value of that correlation listed separately for each species. There was some but not complete overlap in modules that were highly correlated to nighttime CO₂ fixation and modules that contained expected canonical CAM genes (Table 4.2, bold modules). *Yucca aloifolia* had a total of 31 modules, of which 9 had a significant correlation to gas exchange rate (with $p \leq 0.01$). *Yucca filamentosa* had a total of 39 modules with 6 significantly correlated to gas exchange, and *Y. gloriosa* had 27 modules and 6 of those had a significant correlation to gas exchange. *Yucca aloifolia* exhibited the strongest connections

between genes in modules with an average connectivity (the average of weighted adjacencies, $\sum a_{ij}$, for each gene) of 57.24. In contrast, average gene connections in *Y. filamentosa* and *Y. gloriosa* were at 27.74 and 37.10, respectively.

Module eigengenes expression for all three species showed a high day-night variation in modules that were highly correlated to nighttime CO₂ uptake (Fig. 4.4). Curiously, the module most significantly correlated to night CO₂ assimilation in *Y. aloifolia*, ‘dark grey’ (Fig. 4.4A) didn’t contain a single CAM annotated gene and only one putative clock gene, ELF3, which has been shown to be responsible for mediating light signals on the Circadian clock in *Arabidopsis* (Covington et al. 2001). Compared to the total set of gene families in the data, the dark grey module in *Y. aloifolia* was enriched for GO terms associated with metabolic processes, and regulation of cellular and biological processes. The ‘pink’ module in *Y. aloifolia* contained a copy of PEPC, though not the version which had strong temporal variation in expression (Fig. 4.5A). The putative CAM copy of PPCK in *Y. aloifolia*, as well as PEPCK, were both found in the pink module correlated to nighttime CO₂ uptake. *Y. aloifolia* transcripts with CAM-related annotations were generally dispersed and were found in 14 different modules. Transcripts with CAM-related annotations were equally dispersed in the other two species, with 12 and 14 modules containing CAM genes in each of *Y. filamentosa* and *Y. gloriosa*, respectively.

Expression varied in temporal patterning and degree of expression (in FPKM) across CAM genes in the three *Yucca* species (Figs. 4.5, 4.6, and 4.7). There are two copies of PEPC in *Y. aloifolia* with appreciable expression levels (Fig. 4.5A), and only one with a strong increase in expression during the pre-night/night phase. Daytime levels of expression in the PEPC homologs with putative CAM function were surprisingly high, a pattern which was amplified under drought stress. PPCK’s expression largely coincided with the expression of PEPC (Fig. 4.5B),

but was expressed only at night, whereas PEPC peaked in expression in the hours before the onset of night. The two potential decarboxylating enzymes, PEPCK (Fig. 4.5C) and NADP-me (Fig. 4.5D), had variable expression patterns, but the variation across time points did not fit the expected increase in transcription in the day period for CAM function. PEPCK was expressed at much lower levels than NADP-me, indicating its lesser role in decarboxylation in *Yucca* species. PPDK was expressed at the onset of the light period and its expression gradually decreased over the course of the day (Fig. 4.5E). The carbonic anhydrase transcript with highest expression did not have a day-night pattern of expression; no other copy of CA had an increase in expression at night relative to the day.

Yucca gloriosa had nearly all CAM genes expressed in the same temporal pattern as *Y. aloifolia*, even under well-watered conditions when *Y. gloriosa* largely conducts C₃ photosynthesis (see Ch. III) (Fig. 4.6). PEPC expression appears to be turned on even earlier in *Y. gloriosa*, with high expression at the second and third time points from the daytime, although FPKM levels for PEPC are lower in *Y. gloriosa* than in *Y. aloifolia*. PEPCK expression was nearly identical between the hybrid and *Y. aloifolia*, and PEPCK and NADP-me were as non-circadian as they were in *Y. aloifolia*. In comparison, the C₃ *Y. filamentosa* had surprisingly CAM-like expression patterns for PEPC and PEPCK. PEPCK in particular increases in the magnitude of expression at night under drought stress, a response similar to that of *Y. gloriosa*, whose carbon fixation transitions to 100% nighttime fixation under drought stress. Modules that were highly associated with nighttime CO₂ expression had relatively little overlap in terms of gene families represented in the three species (Fig. 3D); in contrast, modules that contained CAM genes had higher overlap in terms of gene family representation between all three species. In addition, *Y. aloifolia* and *Y. gloriosa* were more similar in their gene family overlap, indicating modules

related to CAM genes were similar in these two species. Even with these similarities, each species had a share of gene families that were unique to its CAM-annotated modules.

Gene tree analysis of the six CAM genes of interest showed that the homologs of PEPC and PPCK which showed CAM-like expression in each species were also the most closely related (Fig. 4.8). PEPC-associated modules did not show a strong circadian pattern in any species (though see the yellow module for *Y. aloifolia*, Fig. 4.8), again implying the limited role of PEPC as a decarboxylase in *Yucca*. NAD/P malic enzymes had slight circadian rhythm in their module eigengenes, though surprisingly *Y. filamentosa* had the strongest upregulation of modules containing NAD/P-me during the day (Fig. 4.9). However, the individual gene expression profiles of NAD/P-me transcripts (Figs. 4.5-7D) for all three species show that the upregulation during the day is relative to fairly high FPKM during the night as well, indicating that NAD/P-me enzymes are never fully down-regulated in any of the three *Yuccas*. The relationships among putative CAM copies of PPDK were not monophyletic, with *Y. gloriosa* and *Y. aloifolia* transcripts being more closely related than either are to the CAM-like copy of PPDK in *Y. filamentosa* from the ‘turquoise’ module. Lastly, carbonic anhydrases represent a large gene family with α , β , and γ versions, all of which were separate clades in the gene tree (Fig. 4.9). Which homologs of CA are performing CAM function is difficult to delineate, as gene expression plots again do not show an upregulation of expression at night in any of the three species (Figs. 4.5-7D).

Discussion

Comparative coexpression networks

While WGCNA is a powerful tool for understanding coexpression of genes across time, methods for comparing networks between species are less well-developed. Comparative studies without known orthology between transcripts are particularly challenging. Here we clustered transcripts into gene families, which had the advantage of immediately allowing comparisons between the three *Yucca* species. While it still did not give one-to-one orthologous transcripts, it did allow for assessment of cross-species overlap in the gene family representation within WGCNA defined gene interaction networks. Modules that contained CAM annotated genes in all three species had a large degree of gene family overlap (Fig. 4.4E), indicating shared structure between modules which are correlated to key CAM gene's expression profiles. Still, each species contained gene families unique to its own CAM-related modules; for the two parental species, *Y. aloifolia* and *Y. filamentosa*, this likely reflects divergence over the last 5-10 million years (Smith et al. 2008), where *Y. aloifolia* evolved CAM which perhaps required a rearrangement of coexpressed genes. The hybrid nature of *Y. gloriosa* means its network of coexpressed genes is some intermediate version of its parental networks, and warrants further investigation as a case study for hybrid species gene expression evolution.

Expression of PEPC and PPCK

Gene expression studies in other systems have found contrasting results for the pattern of expression and role of various CAM genes, including the initial carbon fixing enzyme PEPC and its dedicated kinase, PPCK, whose activity phosphorylates PEPC and decreases its sensitivity to cytosolic malate, a product that is rapidly accumulated at night in CAM plants. Work in *Clusia*

showed that C₃ species had nearly constant expression of PPCK across the day-night cycle, but PEPC was largely absent, indicating that any PEPC protein present was constantly phosphorylated and that the amount of PEPC was a bigger limiting factor for CAM expression in these species (Taybi et al. 2004). In contrast, work in *Mesembryanthemum crystallinum* and *Kalanchoe fedtschenkoi* implicated the phosphorylating activity of PPCK as the main determinant of CAM expression (Hartwell et al. 1999; Taybi et al. 2000). In *Yucca*, a third scenario emerges where all three species have nocturnal expression of both PEPC and PPCK at appreciable levels, regardless of photosynthetic pathway (Figs. 4.5-7A,B). PEPC expression peaks in the late afternoon in the CAM *Y. aloifolia* (Fig. 4.5A), whereas the C₃-CAM intermediate *Y. gloriosa* has an up-regulation of PEPC in the middle of the afternoon. In both species, PEPC transcription occurs significantly earlier than the protein is needed. In theory, PEPC should not be present and capable of fixing CO₂ in the cell until RuBisCO is down-regulated, as the two enzymes compete for CO₂. In addition, PPCK has an offset expression pattern relative to PEPC in *Yucca*, with highest expression occurring in our dataset after lights have turned off. If PEPC protein is present in the cell before PPCK begins to be transcribed and translated, PEPC should be inhibited by the product of its own activity, malic acid. Indeed, PPCK's expression is the most starkly Circadian of all genes examined in detail, with barely detectable levels of expression during the day and a sharp increase of transcript abundance at night.

In *Y. aloifolia*, the levels of both PEPC and PPCK are decreased slightly under drought, which coincides with the eventual decline of carbon fixation rates at night in this species as drought progressed (Ch. III). In *Y. gloriosa*, low levels of CAM activity even under well-watered conditions (Ch. III) are associated with relatively high levels of expression of CAM genes,

including PEPC and PPCK. PPCK is further up-regulated under drought stress in the hybrid, which parallels physiological data that showed 100% carbon uptake at night when *Y. gloriosa* was drought stressed (Ch. III). In addition, the hybrid is notable for the high levels of PEPC expression in the middle of the day period; unlike *Y. aloifolia*, who had relatively low levels of C₃ gas exchange under well-watered conditions (and nearly none under drought stress), *Y. gloriosa* fixed a significant proportion of its CO₂ during the day when well-watered. The presence of both PEPC and RuBisCO in the leaves of this species in the middle of the day is therefore even more puzzling, and requires subsequent work to determine whether CO₂ is fixed in both pathways simultaneously, or if the lack of phosphorylation of PEPC during the day in *Y. gloriosa* causes it to be largely inactive. Lack of phosphorylated PEPC during the day has been shown in other CAM studies (Dever et al. 2015), and if true in *Yucca* the question remains why PEPC is expressed at such high levels when it is photosynthetically obsolete during the day.

The presence of both PEPC and PPCK in the C₃ *Y. filamentosa* with expression patterns similar to those found in CAM *Y. aloifolia* is surprising, given a total lack of nighttime CO₂ fixation in *Y. filamentosa* (Ch. III). Although low level CAM activity can be hard to detect with gas exchange methods, *Y. filamentosa* had C₃ characteristics for all phenotypes measured, including titratable acidity, leaf anatomy, and carbon isotopes. Combined, the phenotypes for *Y. filamentosa* accessions leave little room for doubt regarding its photosynthetic pathway. The combined Circadian expression of PEPC and PPCK and their subsequent role in a C₃ plant remain unknown, though an ancestral evolution of the genetic machinery for CAM cannot be ruled out. *Yucca* species that are C₃ based on carbon isotope ratios are succulent and largely overlap in terms of habitat with closely related CAM *yucca* species (Ch. II). It's possible the section of *Yucca* species that are entirely C₃ diverged from their ancestors before the origin of

fully developed CAM pathway; CAM subsequently evolved in extant CAM species of *Yucca* but never in modern day C_3 species, yet vestiges of their shared ancestry can be found in the expression of canonical CAM genes. Whether the ancestral expression pattern for PEPC and PPCK facilitated the origin of CAM is unknown, but gene tree analysis indicates that PEPC and PPCK orthologs from *Y. aloifolia* and *Y. filamentosa* exhibit CAM-like expression, implying that the last common ancestor of C_3 and CAM *Yucca* species had PEPC and PPCK genes with CAM-like expression (Fig. 4.8). Why C_3 *Yucca* species failed to evolve CAM remains a mystery, though variation at the nucleotide level should not be overlooked. For example, the transcript of PEPC that shows CAM-like expression in *Y. filamentosa* is much shorter than (617 amino acids) than putative CAM PEPC transcripts from *Y. aloifolia* and *Y. gloriosa*, which are both 1016 amino acids long, although all three copies appear to be full length with start and stop codons. Further analysis of assemblies and alignments of CAM genes across all three species might elucidate nucleotide or protein level changes that subsequently change the function of individual species' transcripts, despite a high overall similarity in expression patterns.

Decarboxylation in Yucca

CAM plants are classified as either NAD-me, NADP-me, or PEPCK based on which enzyme is used to decarboxylate malic acid in the light period. Closely related *Agave* species have been described as malic-enzyme species, using either NAD- or NADP-me as their primary decarboxylating enzyme (Raveh et al. 1998). Expression of both the malic enzymes (Figs. 4.5-7D) and PPDK, which is necessary for the conversion of malic enzyme's product pyruvate to PEP, together support *Yucca* as a malic enzyme CAM genus as well. While the expression of PPDK in all three species showed a consistent up-regulation with the beginning of the photo period, the temporal pattern of malic enzymes was less clear. In *Y. aloifolia* and *Y. filamentosa*,

the most highly expressed transcript was NADP-me in both cases, though it lacked Circadian expression patterns. In *Y. gloriosa*, the most dominant transcript in terms of FPKM was NAD-me. In no species did any transcript have a strong diurnal expression pattern, and additional work will be needed to describe how decarboxylation is occurring. As is, malic enzyme expression across all time points would be problematic if located in the cytosol, since the presence of NAD- or NADP-me at night in the cytosol would conflict immediately decarboxylate any malic acid produced as a result of PEPC activity.

Future work and conclusions

RNA sequencing has the advantage over lower scale methods, like qRT-PCR, in that a whole transcriptional phenotype can be found across all genes, rather than just a few. Here we begin to scratch the surface of interesting patterns of gene expression by focusing on the CAM pathway, an obvious place to start. But numerous other pathways and genes are implicated in CAM function. The expression of the Circadian clock is an excellent next candidate, though its full pathway description in plants is relatively weak. Using our detailed time-course study could help elucidate additional parts of the pathway, and in particular the effect of various clock genes on the CAM pathway, which is known to be Circadianally regulated. Additionally, starch turnover is an important aspect to CAM function. CAM plants must regenerate PEP every night from sugar or starch reserves, then return PEP back to those reserves during the day. Numerous physiological studies have shown daily starch turnover as an indicator and regulator of CAM expression (Black et al. 1996; Borland and Dodd 2002; Taybi et al. 2004), but little subsequent work has been done on the molecular mechanisms of starch re-allocation. Finally, the large time shift in expression of CAM genes compared to when their proteins are hypothesized to be active is worthy of additional work; proteomics analysis in the same tissues used for RNA-sequencing

here would determine whether early expression of CAM genes results in functional proteins, or if the transcripts are rapidly degraded until the night period.

In terms of understanding evolutionary transitions between C₃ and CAM, we find that ancestrally *Yucca* may have been set up with key enzymes needed for CAM function, as evident from the presence and diurnal expression of some CAM genes in the C₃ *Y. filamentosa* and shared ancestry of CAM-copies based on phylogenetic history of homologs. The idea that there was, at some point, a population of *Yucca* progenitors that could evolve CAM more readily is similarly supported by the presence of anatomical similarities in some CAM and C₃ species of *Yucca*, including thick leaves, 3D venation, and large mesophyll cells (Ch. II). Additionally, the hybrid C₃-CAM intermediate species *Y. gloriosa* has strong CAM expression on both a physiological and transcriptomic level, and is likely the key to discovering what *Y. filamentosa* might be lacking in terms of expression or regulation of CAM genes that prevent it from becoming fully CAM.

References

- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–10.
- Amborella Genome Project. 2013. The Amborella genome and the evolution of flowering plants. *Science.* 342:1241089.
- Black C.C., Chen J.-Q., Doong R.L., Angelov M.N., Sung S.J.S. 1996. Alternative carbohydrate reserves used in the daily cycle of Crassulacean acid metabolism. In: Winter K., Smith J.A.C., editors. *Crassulacean Acid Metabolism*. Springer. p. 31–45.
- Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114–20.
- Borland A.M., Dodd A.N. 2002. Carbohydrate partitioning in crassulacean acid metabolism plants: reconciling potential conflicts of interest. *Funct. Plant Biol.* 29:707–716.

- Borland A.M., Hartwell J., Jenkins G.I., Wilkins M.B., Nimmo H.G. 1999. Metabolite Control Overrides Circadian Regulation of Phosphoenolpyruvate Carboxylase Kinase and CO₂ Fixation in Crassulacean Acid Metabolism. *Plant Physiol.* 121:889–896.
- Carter P.J., Nimmo H.G., Fewson C.A., Wilkins M.B. 1991. Circadian rhythms in the activity of a plant protein kinase. *EMBO J.* 10:2063–8.
- Chen L.-S., Lin Q., Nose A. 2002. A comparative study on diurnal changes in metabolite levels in the leaves of three crassulacean acid metabolism (CAM) species, *Ananas comosus*, *Kalanchoe daigremontiana* and *K. pinnata*. *J. Exp. Bot.* 53:341–350.
- Covington M.F., Panda S., Liu X.L., Strayer C.A., Wagner D.R., Kay S.A. 2001. ELF3 modulates resetting of the circadian clock in *Arabidopsis*. *Plant Cell.* 13:1305–15.
- Cushman J.C., Bohnert H.J. 1997. Molecular Genetics of Crassulacean Acid Metabolism. *Plant Physiol.* 113:667–676.
- Cushman J.C., Tillett R.L., Wood J.A., Branco J.M., Schlauch K.A. 2008. Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C₃ photosynthesis and Crassulacean acid metabolism (CAM). *J. Exp. Bot.* 59:1875–94.
- Dever L. V, Boxall S.F., Kneřová J., Hartwell J. 2015. Transgenic perturbation of the decarboxylation phase of Crassulacean acid metabolism alters physiology and metabolism but has only a small effect on growth. *Plant Physiol.* 167:44–59.
- Dodd A.N., Borland A.M., Haslam R.P., Griffiths H., Maxwell K. 2002. Crassulacean acid metabolism: plastic, fantastic. *J. Exp. Bot.* 53:569–580.
- Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–7.
- Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D. a, Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644–52.
- Gross S.M., Martin J.A., Simpson J., Abraham-Juarez M.J., Wang Z., Visel A. 2013. De novo transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics.* 14:563.
- Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., Macmanes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., Leduc R.D., Friedman N., Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8:1494–512.

- Hartwell J., Gill A., Nimmo G.A., Wilkins M.B., Jenkins G.I., Nimmo H.G. 1999. Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. *Plant J.* 20:333–42.
- Hartwell J., Smith L.H., Wilkins M.B., Jenkins G.I., Nimmo H.G. 1996. Higher plant phosphoenolpyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. *Plant J.* 10:1071–1078.
- Holtum J., Osmond C. 1981. The Gluconeogenic Metabolism of Pyruvate During Deacidification in Plants With Crassulacean Acid Metabolism. *Aust. J. Plant Physiol.* 8:31.
- Kluge M., Ting I.P. 1978. Morphology, anatomy, and ultrastructure of CAM plants. In: Billings W.D., Golley F., Lange O.L., Olson J.S., editors. *Crassulacean Acid Metabolism*. Berlin: Springer-Verlag. p. 29–38.
- Langfelder P., Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 9:559.
- Langmead B., Trapnell C., Pop M., Salzberg S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.
- Li B., Dewey C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 12:323.
- Nelson E.A., Sage R.F. 2008. Functional constraints of CAM leaf anatomy: tight cell packing is associated with increased CAM function across a gradient of CAM expression. *J. Exp. Bot.* 59:1841–50.
- Nelson E.A., Sage T.L., Sage R.F. 2005. Functional leaf anatomy of plants with crassulacean acid metabolism. *Funct. Plant Biol.* 32:409.
- Nimmo H.G. 2000. The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends Plant Sci.* 5:75–80.
- Raveh E., Wang N., Nobel P.S. 1998. Gas exchange and metabolite fluctuations in green and yellow bands of variegated leaves of the monocotyledonous CAM species *Agave americana*. *Physiol. Plant.* 103:99–106.
- Robinson M.D., McCarthy D.J., Smyth G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 26:139–40.
- Silvera K., Neubig K.M., Whitten W.M., Williams N.H., Winter K., Cushman J.C. 2010. Evolution along the crassulacean acid metabolism continuum. *Funct. Plant Biol.* 37:995.
- Singh R., Ong-Abdullah M., Low E.-T.L., Manaf M.A.A., Rosli R., Nookiah R., Ooi L.C.-L., Ooi S.-E., Chan K.-L., Halim M.A., Azizi N., Nagappan J., Bacher B., Lakey N., Smith S.W., He

- D., Hogan M., Budiman M.A., Lee E.K., DeSalle R., Kudrna D., Goicoechea J.L., Wing R.A., Wilson R.K., Fulton R.S., Ordway J.M., Martienssen R.A., Sambanthamurthi R. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*. 500:335–9.
- Smith C.I., Pellmyr O., Althoff D.M., Balcázar-Lara M., Leebens-Mack J., Segraves K.A. 2008. Pattern and timing of diversification in *Yucca* (Agavaceae): specialized pollination does not escalate rates of diversification. *Proc. Biol. Sci.* 275:249–58.
- Stamatakis A., Hoover P., Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57:758–71.
- Taybi T., Nimmo H.G., Borland A.M. 2004. Expression of phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxylase kinase genes. Implications for genotypic capacity and phenotypic plasticity in the expression of crassulacean acid metabolism. *Plant Physiol.* 135:587–98.
- Taybi T., Patil S., Chollet R., Cushman J.C. 2000. A minimal serine/threonine protein kinase circadianly regulates phosphoenolpyruvate carboxylase activity in crassulacean acid metabolism-induced leaves of the common ice plant. *Plant Physiol.* 123:1471–82.
- Winter K., Holtum J.A.M. 2014. Facultative crassulacean acid metabolism (CAM) plants: powerful tools for unravelling the functional elements of CAM photosynthesis. *J. Exp. Bot.* 65:3425–41.
- Zambrano V.A.B., Lawson T., Olmos E., Fernández-García N., Borland A.M. 2014. Leaf anatomical traits which accommodate the facultative engagement of crassulacean acid metabolism in tropical trees of the genus *Clusia*. *J. Exp. Bot.* 65:3513–23.
- Zhang J., Liu J., Ming R. 2014. Genomic analyses of the CAM plant pineapple. *J. Exp. Bot.* 65:3395–404.

Table 4.1 – Genotypes of each species used for RNA sequencing, with the date of when each was in a growth chamber, and the state and GPS coordinates of where the genotype was initially collected from.

Species	Genotype	Date of GC	State	Lat	Lon
<i>Y. aloifolia</i>	6	July 2014	SC	33.64912	-78.92852
<i>Y. aloifolia</i>	7	July 2014	SC	33.6501	-78.9256
<i>Y. aloifolia</i>	23	February 2015	GA	33.92746	-83.37906
<i>Y. filamentosa</i>	2	October 2014	NC	35.95818	-75.63808
<i>Y. filamentosa</i>	9	October 2014	SC	33.5044	-79.06227
<i>Y. filamentosa</i>	27	February 2015	GA	33.92129	-83.39032
<i>Y. filamentosa</i>	30	July 2014	GA	33.93197	-83.33298
<i>Y. gloriosa</i>	12	July 2014	SC	33.50278	-79.06448
<i>Y. gloriosa</i>	17	July 2014	SC	32.39087	-80.43033
<i>Y. gloriosa</i>	19	October 2014	GA	31.02172	-81.4348
<i>Y. gloriosa</i>	20	February 2015	GA	31.14895	-81.36608

Table 4.2 – Module name, size in terms of number of transcripts, and correlation and p-value to nighttime CO₂ uptake rates for the top ten most highly correlated modules in each species. Modules that contain putative CAM genes in each species are in bold.

<i>Y. aloifolia</i>				<i>Y. gloriosa</i>				<i>Y. filamentosa</i>			
Module	Size	R ²	P	Module	Size	R ²	P	Module	Size	R ²	P
dark grey	143	0.81	0.000	midnight blue	302	0.818	0.000	purple	565	-0.46	0.001
pink	493	0.71	0.000	green	826	0.67	0.000	green yellow	542	-0.45	0.002
green	705	0.70	0.000	green yellow	515	-0.62	0.000	steel blue	89	-0.39	0.006
dark green	173	-0.60	0.000	red	766	-0.62	0.000	pale turquoise	81	0.35	0.017
magenta	378	-0.59	0.000	light yellow	195	-0.51	0.000	grey	272	-0.34	0.018
saddle brown	46	-0.55	0.001	pink	718	-0.43	0.003	light cyan	337	0.34	0.021
dark red	173	0.55	0.001	magenta	665	-0.40	0.005	tan	520	0.33	0.023
red	588	0.54	0.001	orange	46	-0.36	0.012	light yellow	244	0.32	0.028
purple	360	-0.46	0.006	dark grey	48	0.36	0.012	saddle brown	103	-0.30	0.044
orange	135	0.40	0.018	royal blue	172	-0.35	0.017	violet	56	-0.29	0.047

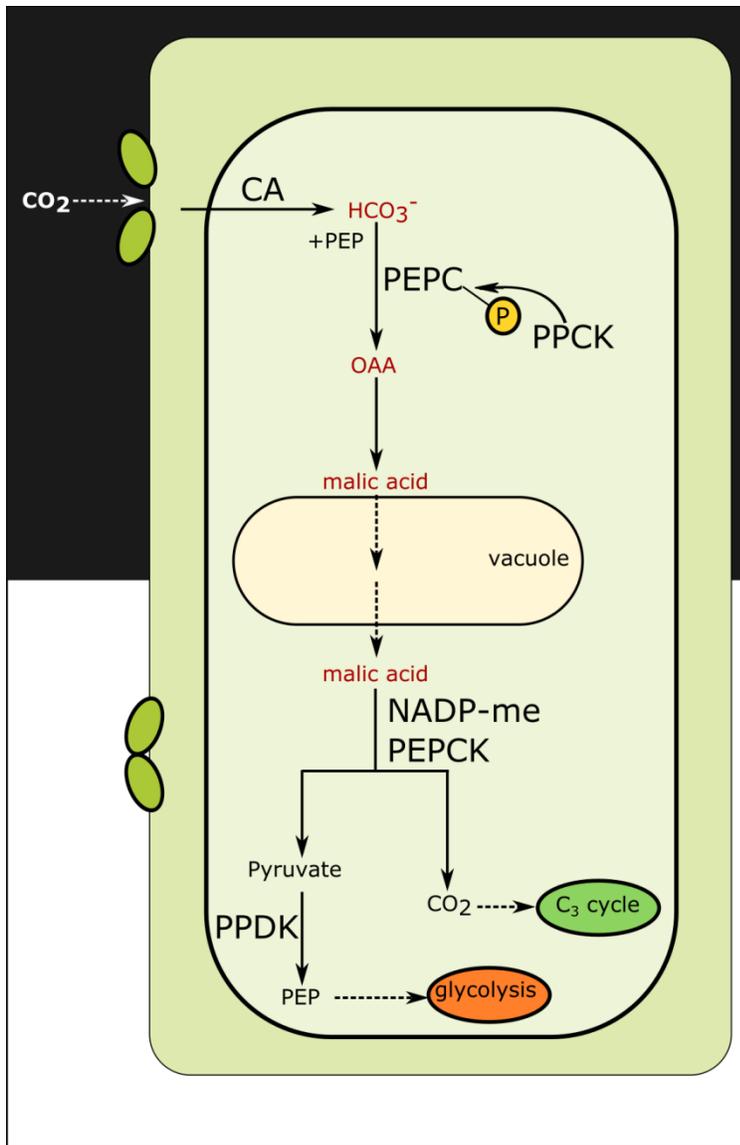
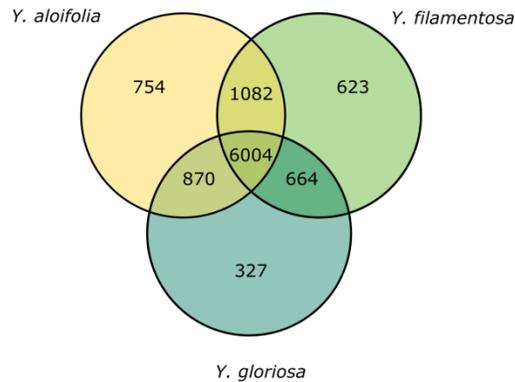


Figure 4.1 – Representation of the CAM pathway. Solid arrows represent enzymatic reactions, while dashed arrows indicate physical movement of molecules. CA – carbonic anhydrase, PEPC – phosphoenolpyruvate carboxylase, PPCK – phosphoenolpyruvate carboxylase kinase, NADP-me – NADP malic enzyme, PEPCK – phosphoenolpyruvate carboxykinase, PPDK – pyruvate, phosphodikinase.

A



B

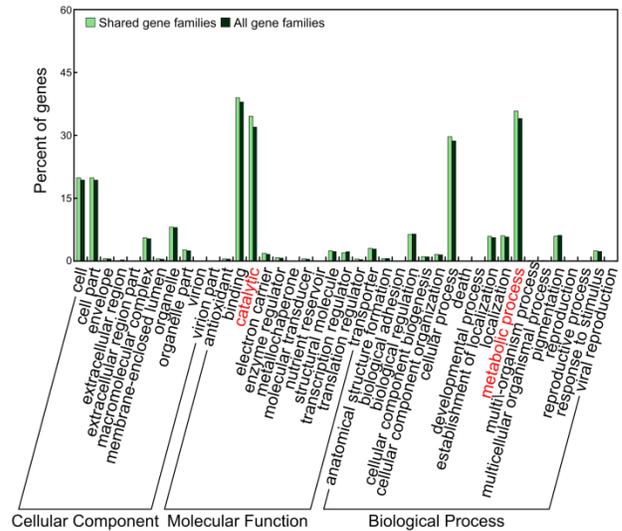


Figure 4.2 – Description of gene family content and annotation overlap. A) Venn diagram of shared gene families between the three species, the result of BLASTn of Transdecoder coding sequence files against the database of 22 sequence land plant genomes (Amborella Genome Project 2013). B) Gene Ontology enrichment of all gene families present in three species, versus the 6,004 gene families shared among all three species. Categories in red indicate significant differences via χ^2 test.

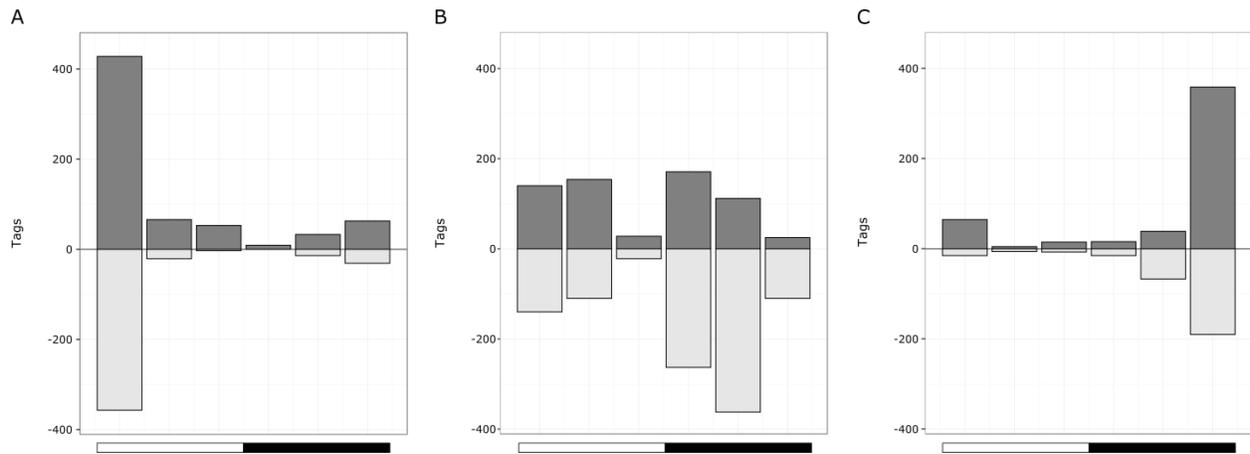


Figure 4.3 – For every time point, differentially expressed transcripts (up regulated=dark grey, down regulated=light grey) under drought conditions relative to well watered in A) *Y. aloifolia*, B) *Y. gloriosa*, and C) *Y. filamentosa*. The first time point corresponds to 1 hour after lights turn on in the morning, then every four hours afterwards. White and black bar beneath each graph represents day and night, respectively.

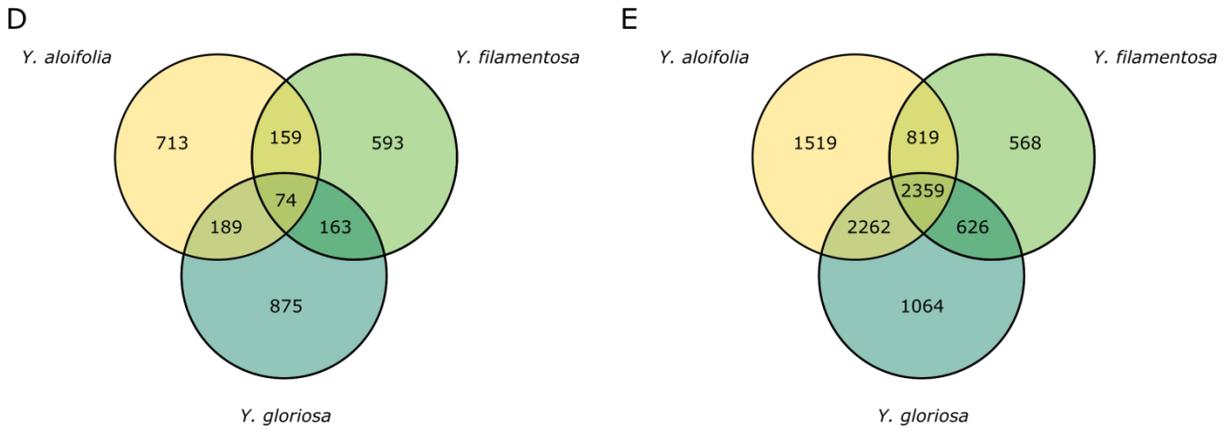
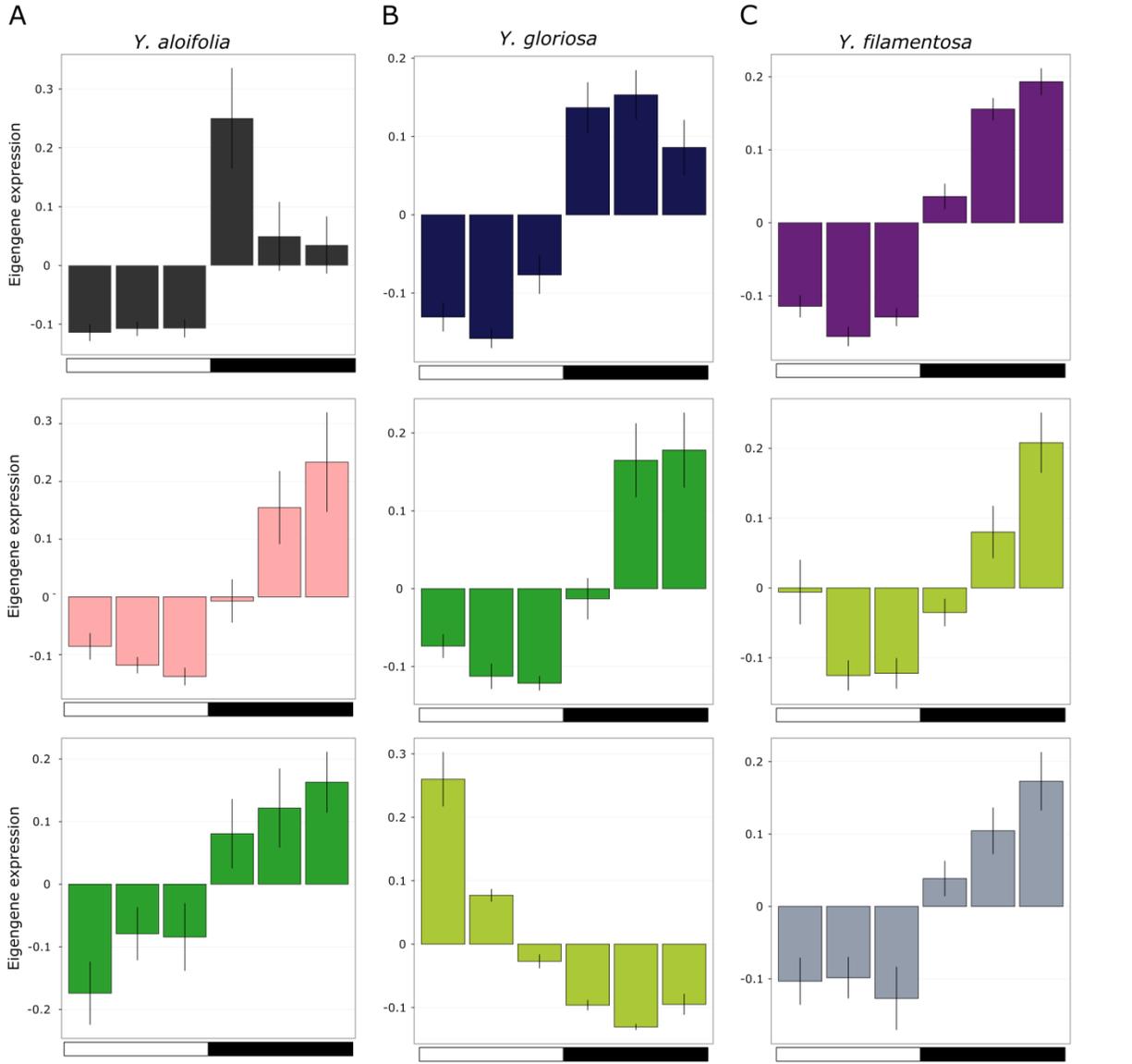


Figure 4.4 – Module eigengenes for the three modules most correlated to nighttime CO₂ uptake rates: Bars are colored according to their module name as output by WGCNA (which is arbitrary, and not comparable across species) for each of A) *Y. aloifolia*, B) *Y. gloriosa*, and C) *Y. filamentosa*. Average eigenvalue for each time point was calculated \pm SE. Open horizontal bar indicates time points when lights were on, black filled bar is night time points. Gene family content was compared for the top three modules across all species (D) and across all modules that contain CAM genes of interest (E).

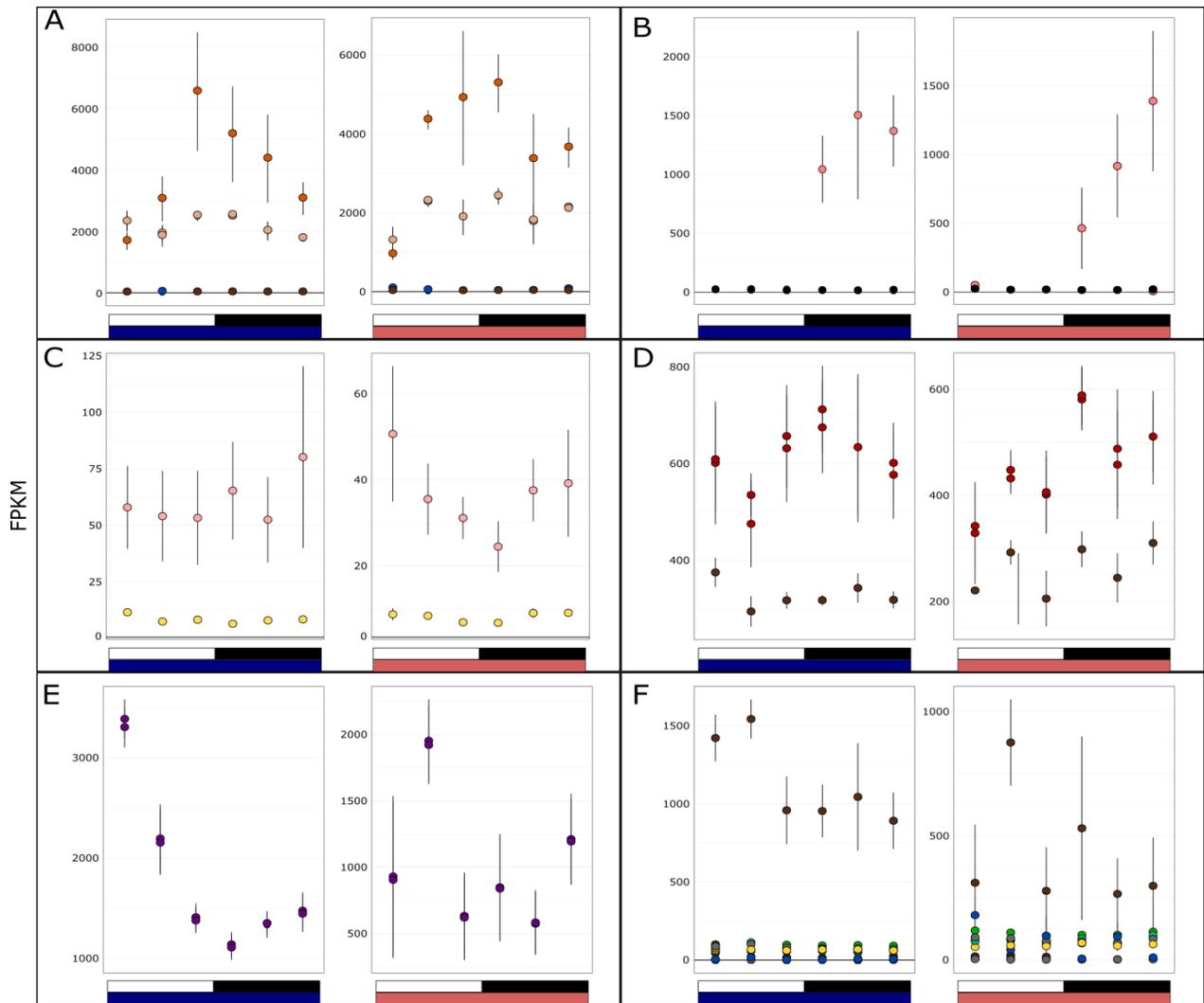


Figure 4.5 – Gene expression of canonical CAM genes in *Y. aloifolia*. A) PEPC, B) PPCK, C) PEPCK, D) malic enzymes, E) PPDK, F) CA. White bar below each plot indicates day light, filled black bar are samples taken at night. Watered vs. drought is indicated by a blue or red bar below each graph, respectively. Genes are color coded by the module to which they belong. Average fpkm across replicates at a given time point were calculated \pm standard error.

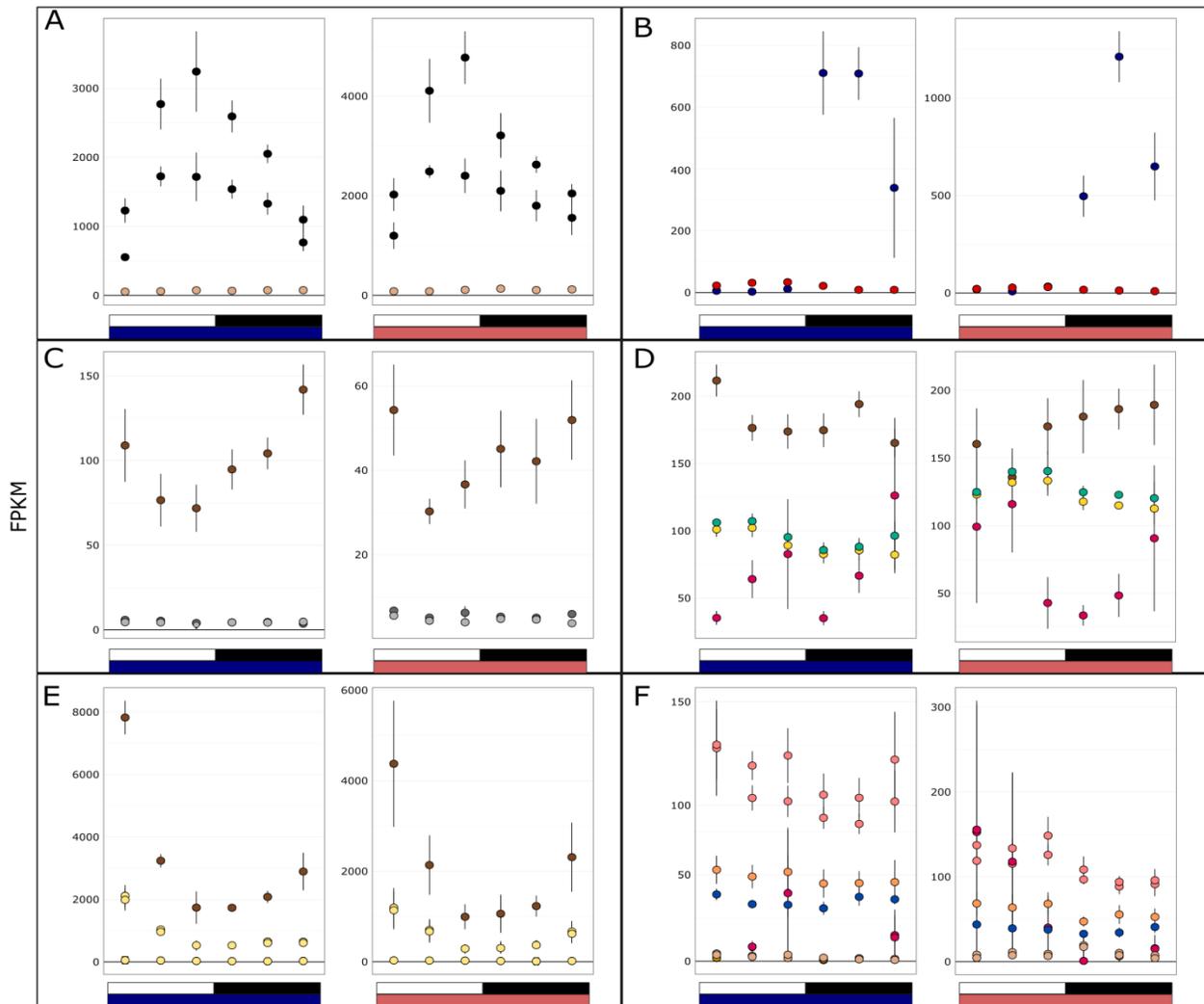


Figure 4.6 – Gene expression of canonical CAM genes in *Y. gloriosa*. A) PEPC, B) PPCK, C) PEPCK, D) malic enzymes, E) PPDK, F) CA. White bar below each plot indicates day light, filled black bar are samples taken at night. Watered vs. drought is indicated by a blue or red bar below each graph, respectively. Genes are color coded by the module to which they belong. Average fpkm across replicates at a given time point were calculated \pm standard error.

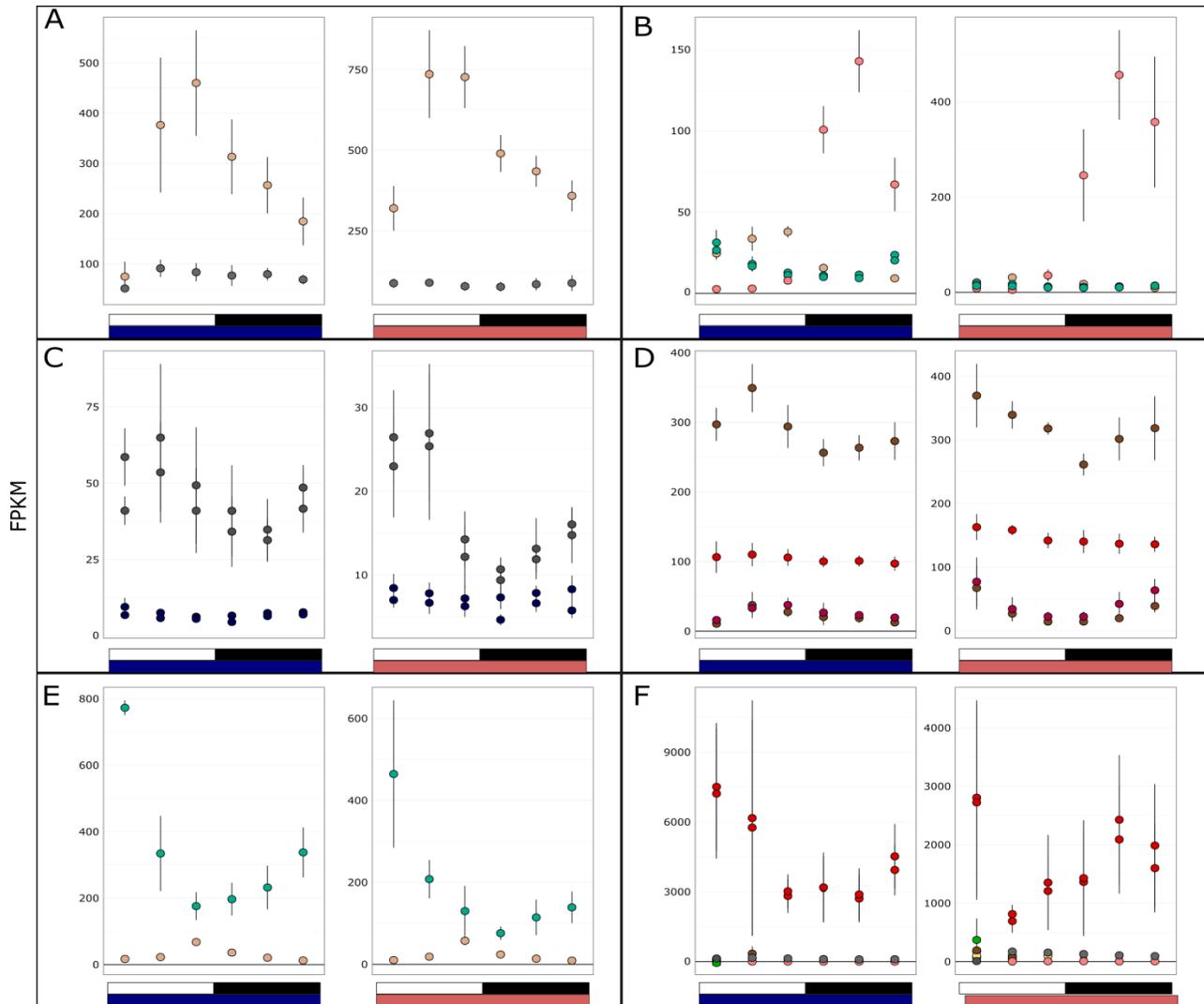


Figure 4.7 – Gene expression of canonical CAM genes in *Y. filamentosa*. A) PEPC, B) PPCK, C) PEPCK, D) malic enzymes, E) PPDK, F) CA. White bar below each plot indicates day light, filled black bar are samples taken at night. Watered vs. drought is indicated by a blue or red bar below each graph, respectively. Genes are color coded by the module to which they belong. Average fpkm across replicates at a given time point were calculated \pm standard error.

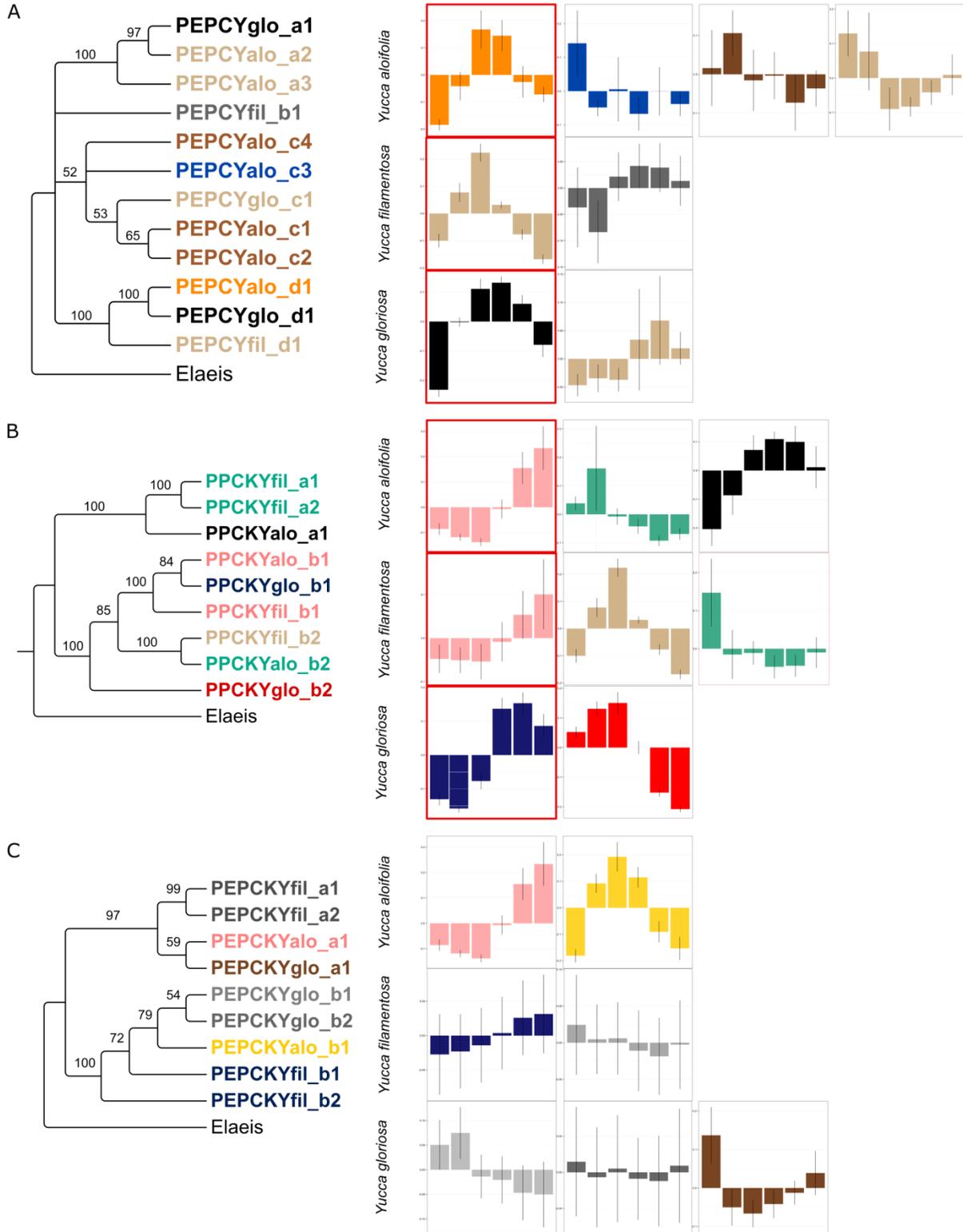
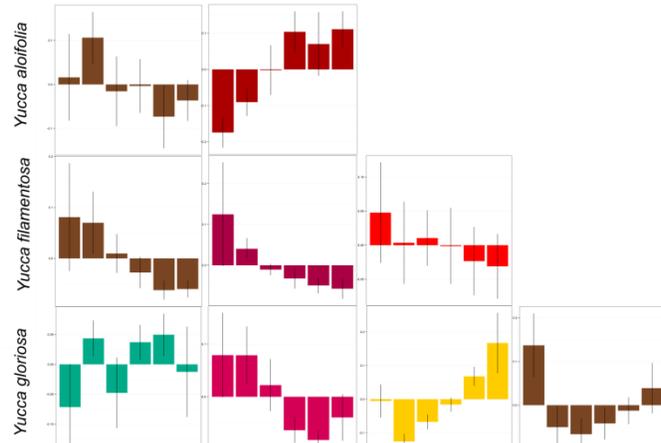
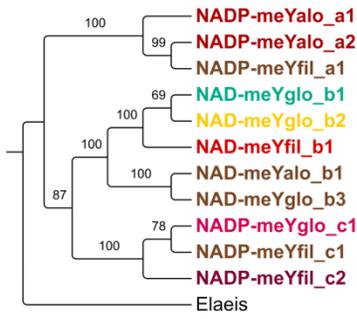


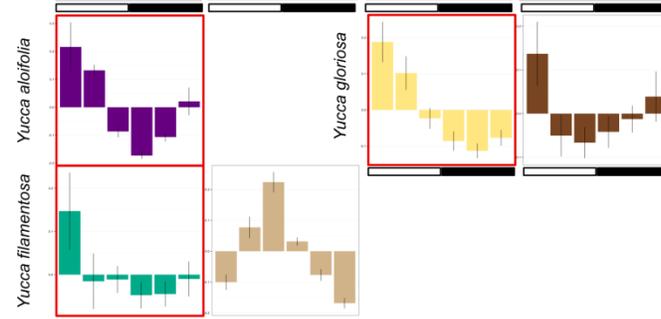
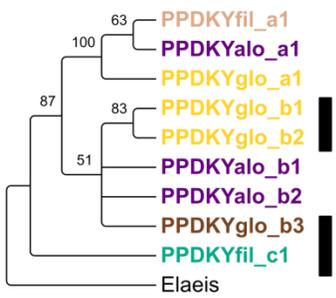
Figure 4.8 – Gene trees of PEPC (A), PPCK (B), and PEPCK (C), with tips color coded by their module membership. Module eigengenes for each species for each gene are to the right. Gene

tree tips are arbitrarily labeled by clade; black bars to the right of a clade indicate putative CAM copies of a gene based on eigengene expression as well as individual expression of transcripts (Figs. 4.5-7). Thicker red lines around an eigengene expression plot indicate which module putative CAM genes are clustered into for each species.

A



B



C

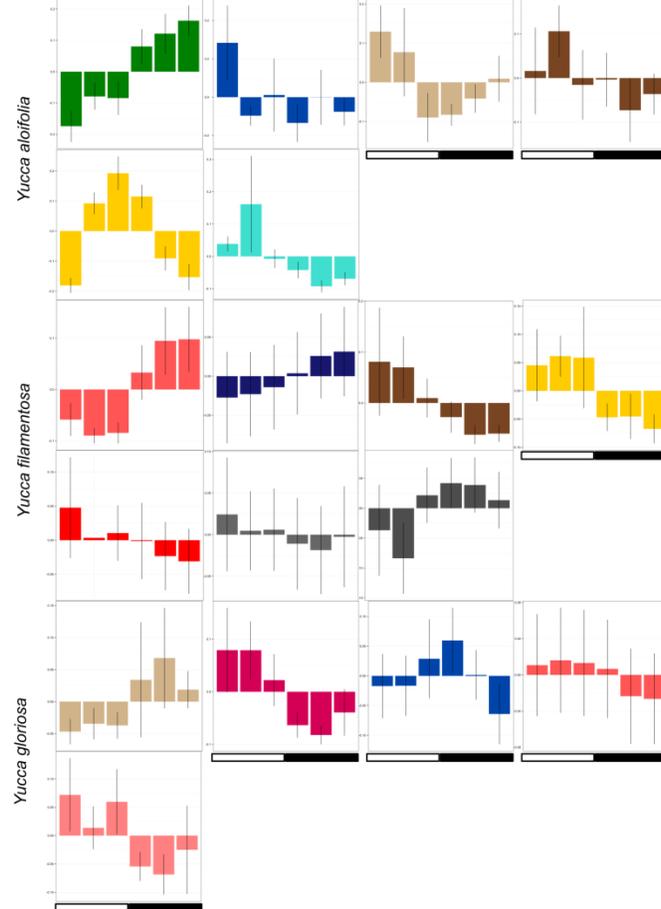
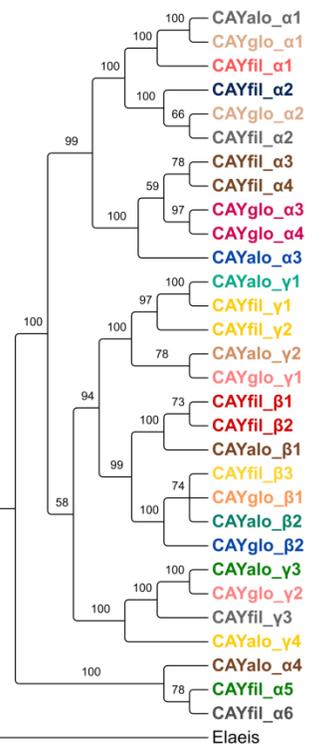


Figure 4.9 – Gene trees of NAD/P-me (A), PPDK (B), and CA (C), with tips color coded by their module membership. Module eigengenes for each species for each gene are to the right. Gene tree tips are arbitrarily labeled by clade; black bars to the right of a clade indicate putative CAM copies of a gene based on eigengene expression as well as individual expression of transcripts (Figs. 4.5-7). Thicker lines around an eigengene expression plot indicate which module putative CAM genes are clustered into for each species.

CHAPTER V:

CONCLUSION AND DISCUSSION

CAM in the Agavoideae

The Agavoideae has a long history of research related to CAM photosynthesis, most notably that of Park Nobel, who did both field and greenhouse studies with a number of *Agave* species. Partly because of his work, and partly because of the iconic nature of *Agaves* as charismatic representatives of New World desert flora, the Agavoideae have been largely considered an entirely CAM lineage and not relevant for comparative research. The results presented here indicate that this is not the case, and that the Agavoideae in fact constitutes an ideal system for understanding the evolution of CAM at all levels – ecological, physiological, and genomic. In particular, the three independent origins of CAM in the Agavoideae described in Chapter II allow for parallel comparisons of how CAM evolves in closely related species. Chapters III and IV took advantage of these closely related C₃/CAM species and investigated C₃, CAM, and C₃-CAM intermediate *Yuccas*. Together, the work presented here represents a coherent synthesis of CAM evolution within the Agavoideae and it sets a foundation for extending this important group for understanding the genetic basis for the evolution of CAM-related traits.

Investigation of leaf anatomy in C₃ and CAM species from across the Agavoideae answered a long standing question: do lineages that evolve CAM have a predisposition toward

CAM evolution? Many species that live in arid regions of the world are *not* CAM, and logically raises questions about why some lineages evolve CAM while others do not. Those that do evolve CAM, on the other hand, tend to have multiple origins; examples include the Bromeliaceae (Givnish et al. 2014), Orchidaceae (Givnish et al. 2015), Euphorbiaceae (Horn et al. 2014). The same pattern was seen in the Agavoideae, where three independent origins all occurred at about 5Mya in *Hesperaloe*, *Yucca*, and *Agave*. Further, it was determined that many lineages in the Agavoideae, including those that are C₃, have a propensity for thick leaves, particularly in the mid vein area, and 3D venation. The latter trait is thought to allow plants to increase their leaf succulence without a major loss of hydraulic conductivity; ancestrally, the Agavoideae may have evolved 3D venation and succulence in response to water limitation, as many extant lineages are found in arid habitats. Subsequently, that increase in succulence may have allowed for CAM to evolve without much structural change in leaf anatomy, as the CAM pathway requires large, succulent cells for malic acid storage at night.

The CAM continuum

The presence of CAM photosynthesis is often discussed as a binary trait, where a plant is – or isn't – a CAM species. It's clear, however, that the CAM phenotype is a spectrum ranging from strong, constitutive CAM to lower CAM activity in otherwise predominantly C₃ plants (Winter and Holtum 2014; Winter et al. 2015). The research presented here adds to the growing body of evidence that CAM is not simply present or absent. Across the Agavoideae, both C₃ and CAM lineages appear to have anatomical characteristics that are CAM-like, including reduced air space between cells and large, thick leaves. Within closely related *Yucca* species, gene expression patterns of canonical CAM genes are similar in both CAM and C₃. PEPC, the main carbon fixing enzyme, was expressed at night in *Y. filamentosa* despite no measurable carbon

uptake or malic acid accumulation. Similarly, PEPC's dedicated kinase, PPCK, showed strong diurnal expression in the CAM, C₃, and C₃-CAM intermediate species of *Yucca*. The gene expression data indicates that, much like the propensity for certain anatomical characteristics that allowed for the repeated evolution of CAM, the Agavoideae may similarly have had a genetic predisposition for CAM-like gene expression. Why ancestral gene expression would have evolved remains to be seen; although these genes are known to function in CAM, they may have other roles yet unknown. It's clear, however, that if we define CAM plants as having particular physiological and genetic traits which C₃ species do not have, we miss the potential of exploring the nuances of CAM evolution in species that exhibit rare intermediate traits (e.g., CAM-like expression of genes but no CAM physiology). Non-CAM species that share similarities with CAM in terms of physiology or gene expression are ideal candidates for elucidating how plants gradually evolve CAM over time.

Remaining questions

While this dissertation work has expanded what is known about CAM evolution in the Agavoideae and moved the subfamily from simply a charismatic desert lineage to one that can serve as a model system for CAM research, there is much that remains to be addressed. Additional anatomical and physiological analysis of members of the Agavoideae will describe more completely the variation in this group of plants. Particularly, more anciently derived species, including the genus *Hesperaloe* and its closely related C₃ lineages warrant further examination, as their representation in this work was relatively limited. As they represent some of the earliest diverging genera of the subfamily, their place in the story of CAM evolution in the Agavoideae is critical to understanding questions of anatomical and physiological preadaptations. Additionally, the genus *Polianthes* represents a reversion from C₃ to CAM;

assays for gas exchange and titratable acidity are ongoing and, when paired with gene expression information, may give insight into how plants lose CAM function. This undoubtedly will be an important piece of the puzzle of how CAM is gained from the ancestral C₃ state.

The hybrid *Yucca* species, *Y. gloriosa*, explored in detail in chapters III and IV, is certain to become a model for investigating the genetic architecture of CAM physiology and anatomy. Genetic markers indicate the species is segregating for parental alleles; one might expect a similar segregation of phenotypes, including the propensity to use the CAM pathway. While the experimental design in Ch. 2 did not permit conclusions to be drawn about differences among *Y. gloriosa* genotypes, future work will focus extensively on this species. Increased replication and sampling a greater breadth of genotypes from the entire range will likely reveal that this incipient species (estimated to be less than 500 years old) may still be fine tuning its photosynthetic machinery. If *Y. gloriosa* is segregating for CAM ability, RNA seq work paired with proteomic information will describe genetic differences that co-vary with CAM, and will represent one of the most detailed and extensive molecular studies of CAM to date.

Overall the future for CAM research in the Agavoideae is exciting. Although this dissertation has raised many new questions regarding CAM evolution, it has more importantly uncovered a number of previously unknown details of CAM evolution, physiology, and gene expression.

References

Givnish T.J., Barfuss M.H.J., Van Ee B., Riina R., Schulte K., Horres R., Gonsiska P.A., Jabaily R.S., Crayn D.M., Smith J.A.C., Winter K., Brown G.K., Evans T.M., Holst B.K., Luther H., Till W., Zizka G., Berry P.E., Sytsma K.J. 2014. Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. *Mol. Phylogenet. Evol.* 71:55–78.

- Givnish T.J., Spalink D., Ames M., Lyon S.P., Hunter S.J., Zuluaga A., Iles W.J.D., Clements M.A., Arroyo M.T.K., Leebens-Mack J., Endara L., Kriebel R., Neubig K.M., Whitten W.M., Williams N.H., Cameron K.M. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. Biol. Sci.* 282:20151553–.
- Horn J.W., Xi Z., Riina R., Peirson J.A., Yang Y., Dorsey B.L., Berry P.E., Davis C.C., Wurdack K.J. 2014. Evolutionary bursts in Euphorbia (Euphorbiaceae) are linked with photosynthetic pathway. *Evolution* (N. Y). 68:3485–3504.
- Winter K., Holtum J.A.M., Smith J.A.C. 2015. Crassulacean acid metabolism: a continuous or discrete trait? *New Phytol.*
- Winter K., Holtum J.A.M. 2014. Facultative crassulacean acid metabolism (CAM) plants: powerful tools for unravelling the functional elements of CAM photosynthesis. *J. Exp. Bot.* 65:3425–41.

APPENDIX A

SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER II

Table S2.1 – Assembly statistics for each library, including total number of raw reads, number of cleaned reads, number of exons after filtering, number of genes after filtering, average coverage of exons (E) and introns (I), average coverage of the chloroplast.

Species	Raw reads	Cleaned reads	# Exons	# Genes	Cov E	Cov I	Cov cp
Agave bovicornuta	6459324	6268758	471	345	35.18	20.20	102.29
Agave parryi	4244268	4000574	474	345	9.45	5.71	18.27
Agave glomeruliflora	21943776	21440816	466	326	108.98	59.24	868.05
Agave scabra	7950274	7703206	468	349	48.99	32.56	61.10
Agave potrerana	14747558	14346306	474	365	84.91	49.89	370.26
Agave schidigera	19900360	19337346	473	353	130.35	78.55	402.27
Agave parviflora	4700580	4455050	476	369	19.62	12.04	55.03
Agave arizonica	12072806	11763848	474	354	74.84	48.09	233.89
Agave xajoensis	7565546	7352304	471	350	47.31	24.74	55.80
Agave ocahui	1039006	997676	358	340	9.63	11.70	26.87
Agave sobria	6872004	6660266	473	332	37.40	22.84	135.64
Agave tequilana	7535836	7328140	472	372	59.45	38.00	79.31
Agave deserti	2070394	1959426	474	380	8.72	7.71	26.21
Agave utahensis var utahensis	4764842	4629534	481	355	35.18	26.85	54.62
Agave murpheyi	11372240	11047656	472	347	74.34	44.92	322.79
Agave delamateri	3452162	3352910	476	362	33.53	29.92	49.93
Agave angustifolia	1414858	1341168	473	355	8.75	7.48	11.66
Agave aktites	20517480	20130324	351	314	59.13	27.77	53.94
Agave aurea	4499458	4365392	468	367	50.73	27.97	28.13
Agave colimana	6190350	6001826	477	356	38.31	25.05	62.99
Agave seemanniana	1363268	1290158	476	372	8.74	7.51	11.65
Agave schottii	4624438	4470552	473	378	39.73	24.22	66.60
agave americana	3200672	6362814	465	356	26.40	18.00	73.04
Agave cerulata	6401344	7181144	465	334	31.58	18.11	72.59
Agave attenuata	7817084	7769880	466	339	34.57	14.87	55.32
Agave palmeri	5967490	5933208	466	365	29.96	17.29	49.15
Beschorneria yuccoides	6011456	5976670	467	335	42.43	20.34	102.32
Camassia howellii	1705916	1657316	239	297	54.87	55.23	101.25
Camassia quamash	850810	825936	239	283	22.66	22.16	46.82
Camassia quamash	862742	857482	578	426	9.22	4.65	63.36

Camassia linearifolia	10709550	10644214	460	365	90.31	42.13	625.11
Chlorogalum pomeridianum	4799854	4770692	571	389	38.23	17.48	104.70
Echeandia spp	3871790	3845874	575	467	15.81	8.74	325.42
Hastingsia alba	37953820	36655052	476	377	175.26	147.51	243.17
Hesperaloe nocturna	4960036	4814104	357	342	52.45	34.44	121.02
Hesperaloe funifera	7448118	7225018	359	350	56.99	37.04	137.56
Hesperaloe campanulata	7221972	14578696	346	291	55.36	28.43	76.42
Hesperaloe parviflora	1453666	1413297	344	307	53.05	1058.11	110.45
Hesperoyucca whipplei	10125136	10074580	347	319	41.44	37.16	242.86
Hosta ventricosa	12240902	11914404	364	327	63.76	51.12	178.08
Hosta venusta	11362304	11047496	360	334	53.52	46.03	119.22
Manfreda virginica	4724872	4696410	470	352	27.71	14.80	43.44
Manfreda undulata	3051966	3033776	470	358	26.61	11.84	64.63
Polianthes tuberosa	2772698	2756320	463	374	16.26	9.45	40.50
Schoenolirion croceum	4398068	4371664	585	412	18.58	11.49	14.64
Yucca schidigera	8758314	8711430	461	349	41.85	14.65	14.12
Yucca elephantipes	220120	218824	353	361	5.05	3.98	3.60
Yucca queretaroensis	6772776	6733952	459	355	36.95	19.54	67.99
Yucca aloifolia	2501464	2487522	464	367	20.15	9.48	26.21
Yucca filamentosa	5940034	5775626	466	359	41.15	27.81	358.29
Yucca schottii	4147958	4021610	471	354	38.55	53.37	49.98
Yucca carnerosana	4022738	3904914	466	374	39.66	20.67	50.13
Yucca jaliscensis	2918294	2830646	474	368	32.64	72.12	24.95
Yucca linearifolia	3029668	2939208	472	374	36.40	29.21	32.95
Yucca louisianensis	11242774	10911738	464	360	88.66	76.59	114.04
Yucca schottii	4225156	4017392	462	368	13.70	6.74	25.68
Yucca filifera	4220594	3996202	468	349	12.79	6.38	20.70
Yucca baileyi	6959930	6588480	465	368	23.20	10.97	29.78
Yucca angustissima	17425926	17072766	349	316	110.12	93.88	98.33
Yucca brevifolia	1983154	1879500	470	376	13.09	7.54	46.94
Yucca glauca	7208520	7000098	469	350	65.62	39.16	99.65
Yucca pallida	8997172	8724316	471	359	91.61	44.18	87.99
Yucca constricta	17144328	16633760	468	374	146.90	65.85	169.57
Yucca capensis	10093724	9802480	472	379	99.61	99.81	119.49
Yucca brevifolia	9690578	9424596	473	363	94.63	69.52	193.70
Yucca elata	10571708	10267342	473	356	105.05	63.91	247.72

Table S2.2 – $\delta^{13}\text{C}$ values for species.

Species	$\delta^{13}\text{C}$
<i>A. bovicornuta</i>	-14.56
<i>A. parryi</i>	-14.66
<i>A. glomeruliflora</i>	-14.89
<i>A. scabra</i>	-13.78
<i>A. schidigera</i>	-13.38
<i>A. parviflora</i>	-11.32
<i>A. arizonica</i> (26323)	-13.03
<i>A. arizonica</i> (30416)	-12.33
<i>A. ocahui</i>	-12.37
<i>A. sobria</i>	-12.39
<i>A. tequilana</i>	-14.14
<i>A. deserti</i>	-14.99
<i>A. utahensis</i> var. <i>utahensis</i>	-16.88
<i>A. murpheyi</i>	-14.80
<i>A. angustifolia</i>	-13.14
<i>A. aktites</i>	-15.68
<i>A. aurea</i>	-13.62
<i>A. colimana</i>	-16.10
<i>A. seemanniana</i>	-13.59
<i>A. schotti</i> (69071)	-12.49
<i>A. schotti</i> (33044)	-13.02
<i>A. americana</i>	-11.61
<i>A. cerulata</i>	-11.41
<i>B. yuccoides</i>	-25.47
<i>C. howellii</i>	-29.09
<i>C. quamash</i> spp. <i>utahensis</i>	-27.73
<i>C. quamash</i> spp. <i>utahensis</i>	n.a.
<i>C. pomeridianum</i>	-29.11
<i>E. leucantha</i>	-27.85
<i>E. luteola</i>	-28.04
<i>E. mexicana</i>	-25.04
<i>E. skinneri</i>	-27.83
<i>H. alba</i>	-30.28
<i>H. parviflora</i>	-17.88
<i>H. whippeli</i>	-23.15
<i>H. lancifolia</i>	-28.04
<i>M. virginica</i>	-19.05
<i>M. undulata</i>	n.a.
<i>P. tuberosa</i>	-29.86
<i>S. croceum</i> (232044)	-28.39
<i>S. croceum</i>	-28.81

(235083)	
<i>Y. schidigera</i>	-12.98
<i>Y. elephantipes</i>	-12.88
<i>Y. queretaroensis</i>	-15.24
(146)	
<i>Y. queretaroensis</i>	-15.10
(123)	
<i>Y. aloifolia</i>	-15.33
<i>Y. filamentosa</i>	-28.19
<i>Y. schotti</i>	-14.30
<i>Y. carnerosana</i>	-13.75
<i>Y. jaliscensis</i>	-17.00
<i>Y. linearifolia</i>	-15.63
<i>Y. schotti</i>	-14.30
<i>Y. filifera</i>	-13.64
<i>Y. baileyi</i>	-22.29
<i>Y. angustissima</i>	-21.59
<i>Y. brevifolia</i> (126)	-22.92
<i>Y. brevifolia</i> (125)	-20.12
<i>Y. glauca</i>	-24.25
<i>Y. pallida</i>	-24.29
<i>Y. constricta</i>	-25.84
<i>Y. capensis</i>	-14.74
<i>Y. elata</i>	-23.25

APPENDIX B

SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER III

Table S3.1 – Locality information for each genotype used for physiological analysis. Latitude and longitude reported in decimal degrees. States: NC = North Carolina, SC = South Carolina, GA = Georgia. Those designated with * were used for gas exchange measurement, and “GC block” indicates what date the clones were in the growth chamber.

Species	Genotype ID	GC block	State	Lat	Lon
<i>Y. aloifolia</i> *	6	July 2014	SC	33.64912	-78.92852
<i>Y. aloifolia</i> *	7	July 2014	SC	33.6501	-78.9256
<i>Y. aloifolia</i> *	23	Feb 2015	GA	33.92746	-83.37906
<i>Y. aloifolia</i>	24	NA	GA	33.94787	-83.4092
<i>Y. filamentosa</i>	1	NA	NC	35.95847	-75.63843
<i>Y. filamentosa</i> *	2	Oct 2014	NC	35.95818	-75.63808
<i>Y. filamentosa</i>	4	NA	NC	35.96482	-75.63393
<i>Y. filamentosa</i> *	9	Oct 2014	SC	33.5044	-79.06227
<i>Y. filamentosa</i>	22	NA	GA	33.90708	-83.34247
<i>Y. filamentosa</i> *	27	Feb 2015	GA	33.92129	-83.39032
<i>Y. filamentosa</i> *	30	July 2014	GA	33.93197	-83.33298
<i>Y. gloriosa</i> *	1	Feb 2015	GA	31.14714	-81.36619
<i>Y. gloriosa</i> *	2	Oct 2014	GA	31.14633	-91.3664
<i>Y. gloriosa</i> *	12	July 2014	SC	33.50278	-79.06448
<i>Y. gloriosa</i> *	13	Feb 2015	SC	32.50592	-80.2931
<i>Y. gloriosa</i> *	14	Oct 2014	SC	32.50398	-80.29545
<i>Y. gloriosa</i>	15	NA	SC	32.50393	-80.2957
<i>Y. gloriosa</i> *	17	July 2014	SC	32.39087	-80.43033
<i>Y. gloriosa</i> *	18	Oct 2014	GA	31.11907	-81.40818
<i>Y. gloriosa</i> *	19	Oct 2014	GA	31.02172	-81.4348
<i>Y. gloriosa</i> *	20	Feb 2015	GA	31.14895	-81.36608

Table S3.2 – Microsatellite primer sequences and repeat motif/length. Source indicates where the primers were derived from: 2011 indicates a source of Flatz et. al 2011, *de novo* is a new SSR developed for this study, and 2012 is from Rentsch and Leebens-Mack 2012.

Name	Source	Sequence	Motif	Length
Yb12-forward Yb12-reverse	2011	AACTCCCGTGTTTTGGTGTG AACTCTACTGCCATGTATGTACGC	TACA	124-132
Yb04-forward Yb04-reverse	2011	GCGCATTTTTGTATTCTATGC TCAGCAGCAACCGACAATAG	CT/GT	162-184
36-forward 36-reverse	2011	TACCCGTTCTTGCGGATAGT GCTGAGTTCATCGTCGTCCT	CT	165-179
34690-forward 34690-reverse	<i>de novo</i>	CTGAGTGTGCTACACTTTTCC CGTAAACATGATTTCTCATCC	CTTACAC	167-173
24908-forward 24908-reverse	<i>de novo</i>	GACGTAGCTGTCACAGATGTT TTTCTTCGTTTCGTTAACTTG	CT	159-165
13-forward 13-reverse	2012	TTACCGAAGCCAGCTCTGC GGAGTGAGAGAGGGAGTGG	AG	234-243
1-forward 1-reverse	2012	CCGACTTCCACCGAACTTG AGACCCAGCGATGATGGAG	CAG	181-201

Table S3.3 – Soil moisture probe measures for % soil water content, taken every other day prior to first LiCor measurement with a soil moisture probe. Plants are indicated by species identifiers (“YA” = *Y. aloifolia*, “YF” = *Y. filamentosa*, and “YG” = *Y. gloriosa*) as well as their genotype designation (number/letter). Grey shaded rows are plants that were assigned to the drought stress treatment.

Plant	Day1	Day3	Day5	Day7	Day9
YA6.1	20.2	15.6	6.3	6.4	11.7
YA6.2	17.5	28.3	20.4	17.2	15.8
YA7.1	6.1	13.3	13.8	15.3	15.5
YA7.2	14.2	12.9	6.0	3.5	8.6
YA23.1	9.9	8.4	11.0	10.8	12.6
YA23.2	14.7	14.8	10.5	5.6	5.6
YF2.1	5.9	1.9	0.2	0	12.1
YF2.2	7.9	14.5	14.2	11	13.7
YF9.1	14.1	10.4	5.1	0.7	15.4
YF9.2	9.0	15.2	16.4	14.8	17.6
YF27.1	8.7	8.6	14.3	14.7	14.3
YF27.2	6.9	6.4	2.6	0	14.9
YF30.1	17.8	5.4	4.5	0.7	19.5
YF30.2	24.7	20.1	20.4	21.5	19.4
YG1A	12.5	9.5	25.1	16.9	19.7
YG1B	13.5	7.0	3.3	5.0	13.6
YG2A	12.3	7.4	2.1	0	17.2
YG2B	8.4	17.6	24.3	20.1	17.2
YG12.1	15.5	10.9	10.3	15.2	18.6
YG12.2	14.4	12.5	2.8	0	12.9
YG13.1	10.0	14.2	23.9	15.1	19.5
YG13.2	8.7	7.5	3.1	1.8	11.0
YG14.1	7.7	7.4	3.2	0	9.9
YG14.2	11.6	7.3	11.9	14.7	13.8
YG17.1	5.4	1.4	2.3	0.7	8.4
YG17.2	14.2	17	16	14.7	13.2
YG18.1	11.5	6.7	0.4	0	12.7
YG18.2	10.4	13.4	14.8	12.5	15.7
YG19.1	7.6	3.5	2.6	0.6	10.4
YG19.2	5.1	10.2	10.5	14.2	15.7
YG20.1	7.0	4.3	0.9	3.8	14.1
YG20.2	7.1	10.5	16.0	9.1	12.6

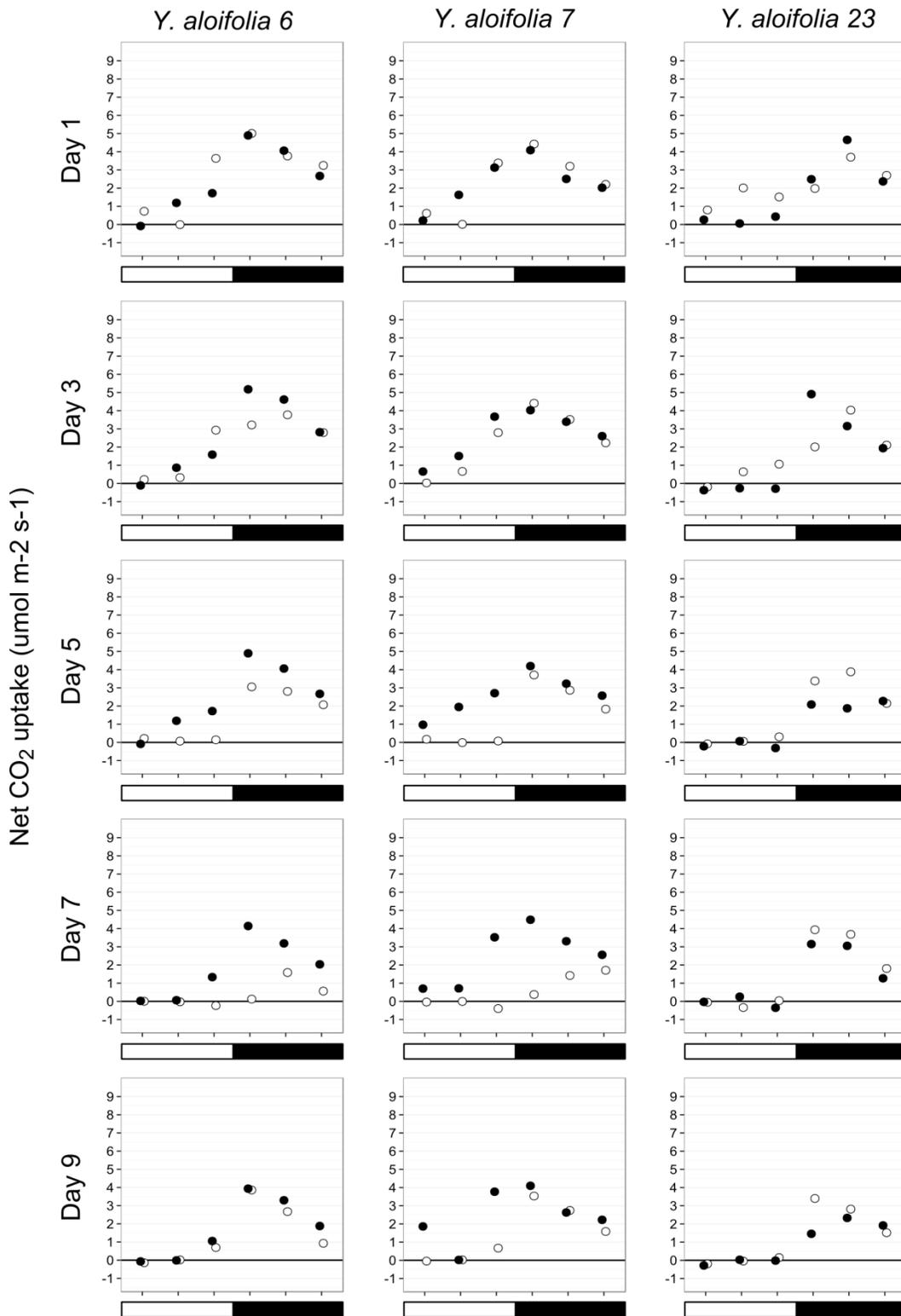


Figure S3.1 – Individual genotype gas exchange curves for the 3 samples of *Y. aloifolia*. Filled circles indicate the clone kept under well watered conditions, open circles indicate clones which were subjected to dry down starting after Day 1. The open bar indicates hours under light, the filled bar indicates time when lights were off.

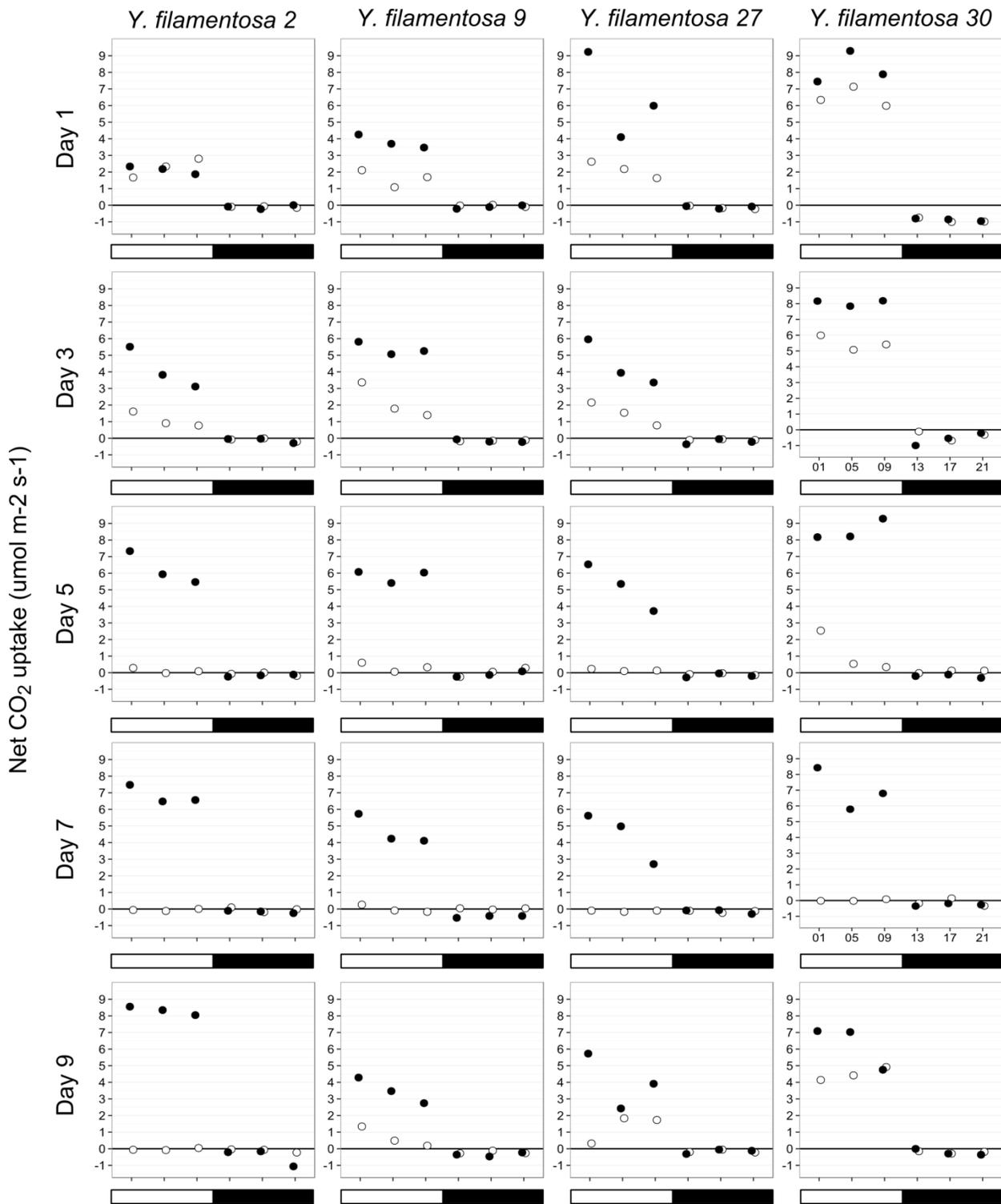


Figure S3.2 – Individual genotype gas exchange curves for the 4 samples of *Y. filamentosa*. Filled circles indicate the clone kept under well watered conditions, open circles indicate clones which were subjected to dry down starting after Day 1. The open bar indicates hours under light, the filled bar indicates time when lights were off.

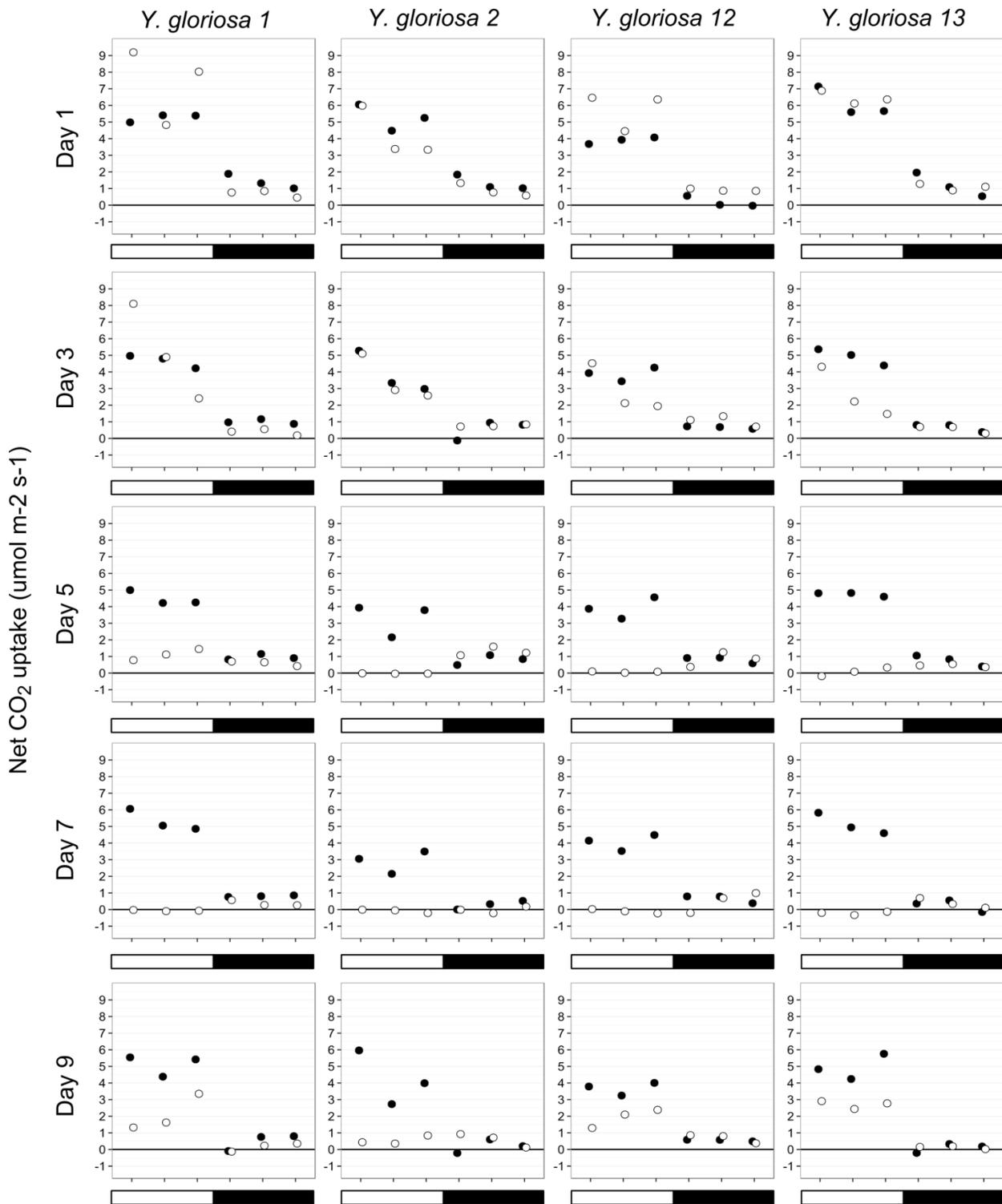


Figure S3.3 – Individual genotype gas exchange curves for the 9 samples of *Y. gloriosa*, divided into three panels. Filled circles indicate the clone kept under well watered conditions, open circles indicate clones which were subjected to dry down starting after Day 1. The open bar indicates hours under light, the filled bar indicates time when lights were off.

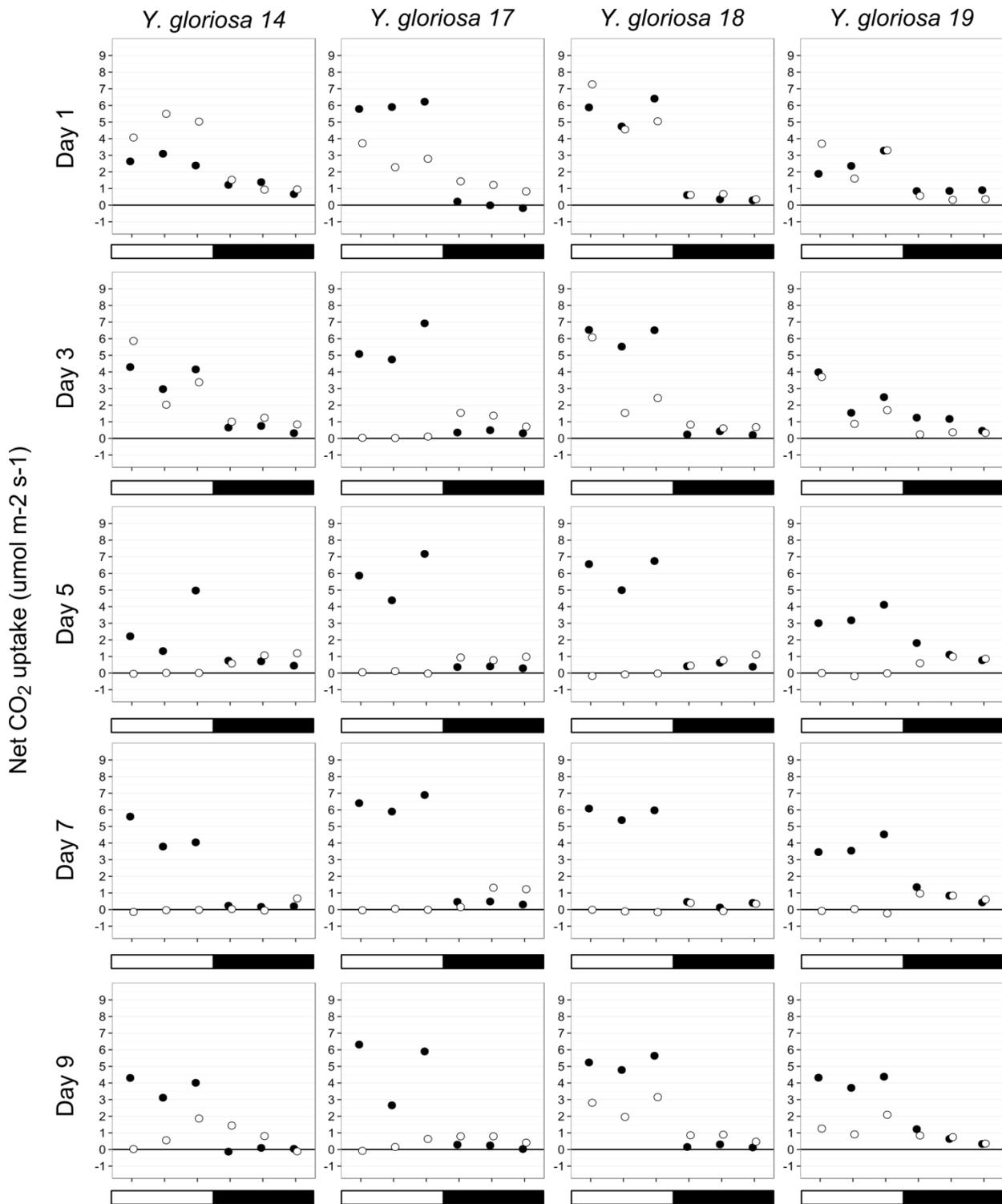


Figure S3.3 – Individual genotype gas exchange curves for the 9 samples of *Y. gloriosa*, divided into three panels. Filled circles indicate the clone kept under well watered conditions, open circles indicate clones which were subjected to dry down starting after Day 1. The open bar indicates hours under light, the filled bar indicates time when lights were off.

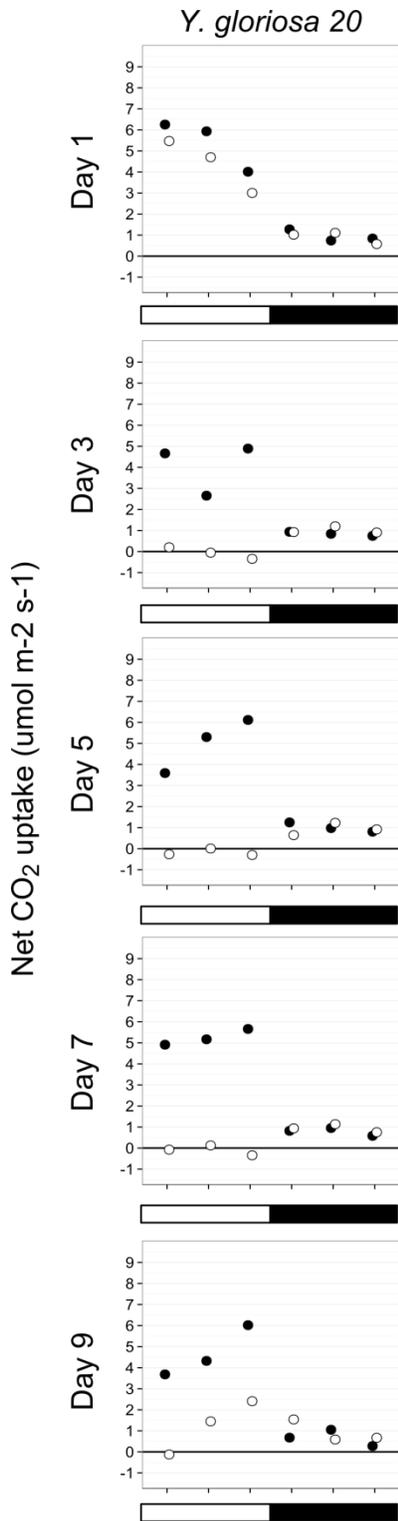


Figure S3.3 – Individual genotype gas exchange curves for the 9 samples of *Y. gloriosa*, divided into three panels. Filled circles indicate the clone kept under well watered conditions, open circles indicate clones which were subjected to dry down starting after Day 1. The open bar indicates hours under light, the filled bar indicates time when lights were off.

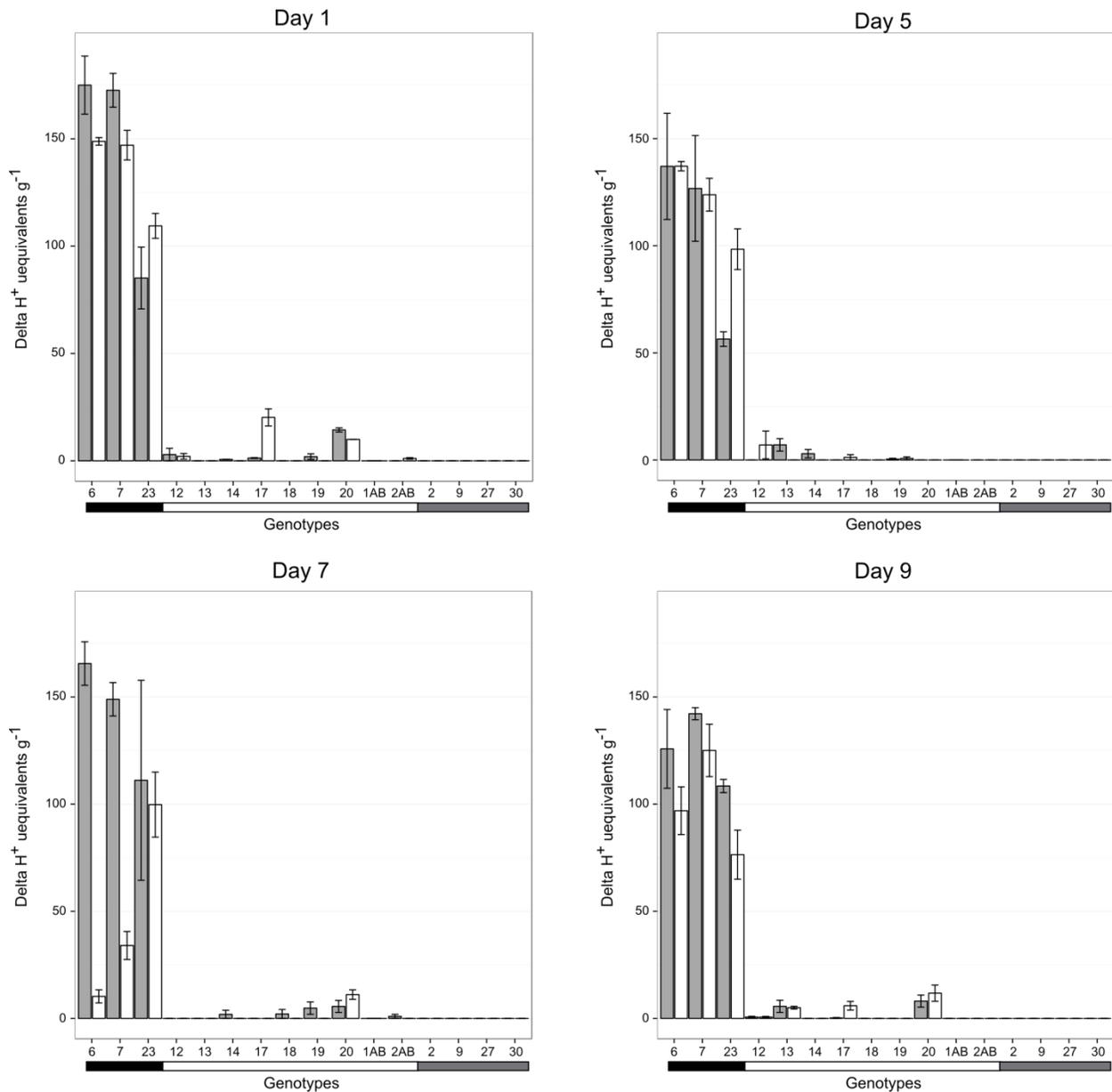


Figure S3.4 – Genotypic deltaH⁺ (H⁺ uequivalents AM – H⁺ uequivalents PM, per gram of tissue) values across the four days measured. Black line under genotypes are *Y. aloifolia*, open bar is *Y. gloriosa*, and grey bar indicates genotypes of *Y. filamentosa*. Grey vertical bars indicate measurements from watered plants (except day 1, where all plants were well-watered), open bars indicate plants in the drought treatment. Standard error reported with error bars.

	Succulence	Adaxial stomata	Abaxial stomata	Leaf thickness	IAS	Avg. dist. Major vein	Avg. dist minor vein	Max. dark CO ₂	Max. light CO ₂	Proportion dark CO ₂ (watered)	Proportion dark CO ₂ (watered)	deltaH+ (w)	deltaH+ (d)	Avg. cell size adaxial	Avg. cell size abaxial
Succulence															
Adaxial stomata	-0.57														
Abaxial stomata	-0.55	0.63													
Leaf thickness	0.68	-0.49	-0.29												
IAS	-0.41	0.21	0.00	-0.69											
Avg. dist. major vein	0.61	-0.74	-0.43	0.37	-0.34										
Avg. dist minor vein	0.55	-0.30	-0.23	0.58	-0.43	0.27									
Max. dark CO ₂	0.67	-0.46	-0.26	0.89	-0.78	0.47	0.58								
Max light CO ₂	-0.49	0.53	0.28	-0.64	0.41	-0.64	-0.46	-0.70							
Proportion dark CO ₂ (drought)	0.67	-0.45	-0.14	0.78	-0.63	0.56	0.56	0.88	-0.85						
Proportion dark CO ₂ (watered)	0.65	-0.39	-0.13	0.84	-0.74	0.41	0.67	0.83	-0.67	0.84					
DeltaH+ (w)	0.37	-0.17	-0.12	0.69	-0.48	0.33	0.43	0.66	-0.76	0.77	0.77				
DeltaH+ (d)	0.40	-0.10	0.14	0.55	-0.64	0.45	0.32	0.62	-0.63	0.67	0.69	0.77			
Avg. cell size adaxial	0.70	-0.61	-0.35	0.92	-0.62	0.52	0.57	0.86	-0.70	0.79	0.87	0.75	0.55		
Avg. cell size abaxial	0.68	-0.59	-0.35	0.86	-0.61	0.53	0.58	0.81	-0.69	0.71	0.83	0.74	0.60	0.96	

Figure S3.5 – Raw Spearman correlation coefficients of phenotypic traits.

	Succulence	Adaxial stomata	Abaxial stomata	Leaf thickness	IAS	Avg. dist. Major vein	Avg. dist minor vein	Max. dark CO ₂	Max. light CO ₂	Proportion dark CO ₂ (watered)	Proportion dark CO ₂ (watered)	deltaH+ (w)	deltaH+ (d)	Avg. cell size adaxial	Avg. cell size abaxial
Succulence		0.49	0.63	0.05	1	0.22	0.46	0.28	1	0.31	0.37	1	1	0.03	0.06
Adaxial stomata	0.01		0.25	1	1	0.03	1	1	1	1	1	1	1	0.32	0.41
Abaxial stomata	0.01	0		1	1	1	1	1	1	1	1	1	1	1	1
Leaf thickness	0	0.03	0.23		0.04	1	0.34	0	0.4	0.03	0	0.22	1	0	0
IAS	0.06	0.39	0.99	0		1	1	0.03	1	0.44	0.07	1	0.41	0.2	0.24
Avg. dist. major vein	0	0	0.07	0.1	0.14		1	1	0.41	0.99	1	1	1	0.67	0.65
Avg. dist minor vein	0.01	0.21	0.35	0.01	0.05	0.24		0.8	1	1	0.31	1	1	0.37	0.33
Max. dark CO ₂	0	0.09	0.35	0	0	0.07	0.02		0.2	0	0.01	0.34	0.49	0	0.01
Max light CO ₂	0.06	0.04	0.31	0.01	0.12	0.01	0.07	0		0	0.28	0.05	0.44	0.2	0.21
Proportion dark CO ₂ (drought)	0	0.09	0.63	0	0.01	0.02	0.02	0	0		0.01	0.04	0.31	0.03	0.16
Proportion dark CO ₂ (watered)	0.01	0.15	0.66	0	0	0.11	0	0	0	0		0.04	0.22	0	0.01
DeltaH+ (w)	0.16	0.54	0.68	0	0.06	0.21	0.09	0.01	0	0	0		0.04	0.07	0.08
DeltaH+ (d)	0.12	0.73	0.63	0.03	0.01	0.08	0.23	0.01	0.01	0	0	0		1	0.61
Avg. cell size adaxial	0	0.01	0.14	0	0	0.02	0.01	0	0	0	0	0	0.03		0
Avg. cell size abaxial	0	0.01	0.14	0	0	0.01	0.01	0	0	0	0	0	0.01	0	

Figure S3.6 – Uncorrected (below diagonal) and Holm-Bonferroni corrected (above diagonal) p-values for correlations described in Fig. S5.

APPENDIX C

SUPPLEMENTARY FIGURES AND TABLES FROM CHAPTER IV

Table S4.1 – Coexpressed gene network module information for all three species of *Yucca*. Module name is described as the arbitrary color given each module by WGCNA, the size is in reference to the number of transcripts present in each, the correlation is against nighttime CO₂ uptake rates, and the significance of that correlation (p).

<i>Y. aloifolia</i>				<i>Y. gloriosa</i>				<i>Y. filamentosa</i>			
Module	Size	R ²	p	Module	Size	R ²	p	Module	Size	R ²	p
Dark grey	143	0.835	0.000	Midnight blue	302	0.818	0.000	Purple	565	-0.457	0.001
Pink	493	0.708	0.000	Green	826	0.673	0.000	Green yellow	542	-0.449	0.002
Green	705	0.696	0.000	Green yellow	515	-0.617	0.000	Steel blue	89	-0.393	0.006
Dark green	173	-0.604	0.000	Red	766	-0.616	0.000	Pale turquoise	81	0.347	0.017
Magenta	378	-0.586	0.000	Light yellow	195	-0.505	0.000	Grey	272	-0.343	0.018
Saddle brown	46	-0.553	0.001	Pink	718	-0.429	0.003	Light cyan	337	0.335	0.021
Dark red	173	0.547	0.001	Magenta	665	-0.404	0.005	Tan	520	0.331	0.023
Red	588	0.543	0.001	Orange	46	-0.364	0.012	Light yellow	244	0.321	0.028
Purple	360	-0.458	0.006	Dark grey	48	0.363	0.012	Saddle brown	103	-0.296	0.044
Orange	135	0.396	0.018	Royal blue	172	-0.346	0.017	Violet	56	-0.291	0.047
Dark turquoise	167	-0.384	0.023	Black	719	0.316	0.030	Dark turquoise	189	0.254	0.084
Grey	134	-0.379	0.025	Light green	261	-0.312	0.033	Yellow	934	0.246	0.095
Black	587	0.350	0.039	Tan	431	0.298	0.042	Dark magenta	48	0.219	0.139
Blue	1993	-0.343	0.044	Yellow	855	0.292	0.046	Dark green	192	-0.208	0.160
Tan	344	-0.315	0.065	Dark orange	45	0.231	0.117	Plum1	35	0.201	0.175
Cyan	301	-0.303	0.077	turquoise	1483	0.207	0.162	Green	884	-0.194	0.192
Grey60	213	-0.279	0.105	Dark turquoise	66	-0.160	0.284	Sienna3	44	0.190	0.200
Turquoise	2189	-0.259	0.134	Dark red	144	-0.146	0.328	Dark orange	171	0.166	0.265
Steel blue	38	0.242	0.161	Cyan	355	-0.145	0.331	Dark grey	186	-0.164	0.271

Light green	212	0.240	0.166	Blue	1426	0.125	0.401	Orange	181	-0.160	0.281
Midnight blue	280	0.238	0.169	Grey	404	-0.119	0.426	Royal blue	239	0.154	0.300
Sky blue	56	-0.158	0.364	Light cyan	297	-0.107	0.475	Salmon	516	-0.153	0.305
White	84	0.116	0.506	Grey60	291	-0.060	0.688	Dark olive green	55	0.151	0.311
Green yellow	354	0.099	0.572	Salmon	429	0.046	0.759	Light green	254	0.142	0.343
Brown	1676	0.084	0.632	Purple	633	0.043	0.774	Brown	1022	0.125	0.403
Light yellow	177	-0.066	0.704	Brown	1112	-0.023	0.879	Sky blue3	35	0.123	0.410
Royal blue	176	-0.063	0.719	Dark green	109	0.008	0.957	Yellow green	40	-0.119	0.424
Dark orange	84	0.047	0.789					Cyan	383	0.119	0.426
Salmon	310	0.046	0.791					Dark red	210	-0.115	0.443
Light cyan	280	-0.033	0.853					Red	744	0.110	0.462
Yellow	1336	0.031	0.860					Grey60	256	-0.106	0.479
								Pink	720	-0.105	0.484
								Midnight blue	345	-0.088	0.558
								Magenta	612	0.061	0.685
								Black	722	0.057	0.704
								White	167	-0.055	0.712
								Turquoise	1117	0.045	0.764
								Sky blue	164	0.045	0.765
								Blue	1102	-0.033	0.824

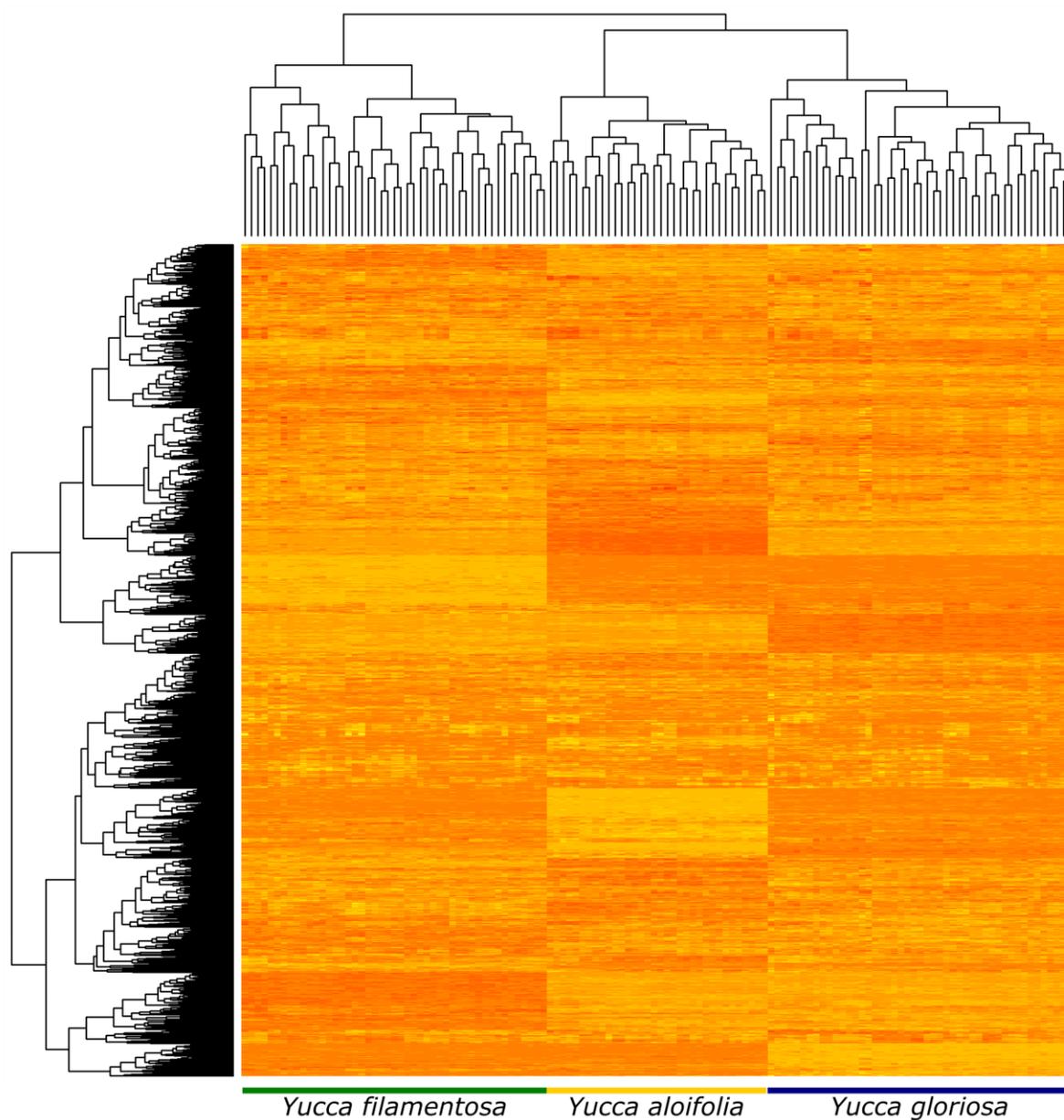


Figure S4.1 - Heatmap (hierarchical clustering) of read counts (normalized as counts per million) for each of the 6004 gene families shared between the three species. The longest member of the orthogroup for each species was used as the representative for read counts.