GENOME EVOLUTION IN MONOCOTS

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

By

Kate L. Hertweck

Dr. J. Chris Pires, Dissertation Advisor

JULY 2011

The undersigned, appointed by the dean of the Graduate School,

have examined the dissertation entitled

GENOME EVOLUTION IN MONOCOTS

Presented by Kate L. Hertweck

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

Dr. J. Chris Pires

Dr. Lori Eggert

Dr. Candace Galen

Dr. Rose-Marie Muzika

ACKNOWLEDGEMENTS

I am indebted to many people for their assistance during the course of my graduate education. I would not have derived such a keen understanding of the learning process without the tutelage of Dr. Sandi Abell. Members of the Pires lab provided prolific support in improving lab techniques, computational analysis, greenhouse maintenance, and writing support. Team Monocot, including Dr. Mike Kinney, Dr. Roxi Steele, and Erica Wheeler were particularly helpful, but other lab members working on Brassicaceae (Dr. Zhiyong Xiong, Dr. Maqsood Rehman, Pat Edger, Tatiana Arias, Dustin Mayfield) all provided vital support as well. I am also grateful for the support of a high school student, Cady Anderson, and an undergraduate, Tori Docktor, for their assistance in laboratory procedures. Many people, scientist and otherwise, helped with field collections: Dr. Travis Columbus, Hester Bell, Doug and Judy McGoon, Julie Ketner, Katy Klymus, and William Alexander. Many thanks to Barb Sonderman for taking care of my greenhouse collection of many odd plants brought back from the field. I obtained irreplacable intellectual support from my peers at MU: Katy Frederick-Hudson, Corey Hudson, Ashley Siegel, Jen Holland, Dr. Elene Valdivia, and other members of our Think Tank. My perpetually patient and helpful committee included Dr. Candi Galen, Dr. Lori Eggert, and Dr. Rose-Marie Muzika. Finally, I owe deep thanks and appreciation to my advisor, Dr. J. Chris Pires. I am very proud to be the Pires lab “burnt pancake.”

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

GENOME EVOLUTION IN MONOCOTS

Kate L. Hertweck

Dr. J. Chris Pires, Dissertation Advisor

ABSTRACT

Monocotyledonous plants are a well-circumscribed lineage comprising 25% of all angiosperm species, including many agriculturally and ecologically important species (e.g., grasses, gingers, palms, orchids, lilies, yams, pondweeds, seagrasses, aroids). These taxa possess nearly the full breadth of vegetative and floral morphology seen across angiosperms, dominate a variety of ecosystems, and exhibit considerable genomic complexity, including the largest genome sizes of all plants. The opportunities afforded by this wealth of variation include evaluating patterns of morphological evolution, genomic change, and geographic radiation. This same variation, however, presents unique challenges to establishing an accurate phylogenetic framework as the foundation for evolutionary analysis.

This dissertation documents three vignettes in monocot evolution, each highlighting different taxonomic scales and relevant questions to the diversification and significance of both organismal (life history, biogeography, morphology) and genomic (genome size, molecular evolution) characteristics. Chapter 2 uses molecular sequence data from all three genomic partitions (nuclear and both organellar genomes) to infer evolutionary

relationships in monocots. Subsequent divergence time and diversification analysis suggests that radiation of major monocot lineages was highly dependent on the origin of other plant and animal lineages. Chapter 3 evaluates a taxonomic classification system in the Tradescantia alliance (Commelinaceae, Commelinales), a group of closely related genera exhibiting kaleidoscopic variation in life history and genomic traits. The phylogeny developed for the alliance is used to re-interpret evolution of taxonomically relevant morphological characters and to test for correlations between genome size and life history/biogeography. Finally, Chapter 4 evaluates a methodological approach to genome sequencing in two lineages of monocots. Grasses (Poaceae, Poales) as a model system are used to test the efficacy of such methods. Non-model Asparagales (agave, onion, asparagus), with large genomes and a paucity of published sequence data, are used to support the ability of these genome sequencing methods to provide ample data for ecological and evolutionary studies. Each of these examples highlights the ability of monocots to serve as test cases for different types of evolutionary questions.

CHAPTER 1

INTRODUCTION

Monocotyledenous plants are a well-defined and monophyletic group comprising over 60,000 species (25% of all angiosperm species). Monocots are characterized by presence of a single cotyledon, mainly herbaceous habit, parallel leaf venation, flowers with three parts, and a variety of other anatomical and morphological similarities [1]. They are the ecological cornerstone of many habitats (e.g., prairies and wetlands) and possess economic importance exceeding any other angiosperm clade. Cereal grasses and other dietary staples like taro and yams provide the primary source of carbohydrates in many cultures, and livestock from which meat protein is derived depend on pasture grasses. Additional edibles include agave, onion, asparagus, bananas, coconuts, palms (oil), and a variety of other fruits and vegetables. Turf grasses, orchids, and bulbs (e.g., Agapanthus, Amaryllis) are bred and propagated widely for horticultural purposes, while additional bulbous and epiphytic species are narrowly restricted, endangered, and/or protected by international law. Finally, many agriculturally and ecologically devastating invasive and noxious weeds are monocots (grasses, *Hydrilla*, *Eichornia*).

Despite widespread ecological and economic significance, classification within monocots has been contentious because of confounding morphological characters between lineages [e.g., 2]. The current classification system for monocots [3] describes eleven orders

(Acorales, Alismatales, Petrosaviales, Dioscoreales, Pandanales, Liliales, Asparagales, Arecales, Commelinales, Zinigiberales, Poales) and one unplaced family (Dasypogonaceae). The first molecular phylogeny of monocots utilized a single gene (rbcL) and revolutionized our understanding of organization of and relationships between these orders [4]. Current phylogenetic inference strongly supports monocots as a monophyletic lineage diverging from the rest of the angiosperms in the early Cretaceous, between 191-139 Ma [see 1 for a thorough review of divergence time studies]. Datasets with wide taxon sampling/few genes [5] and sparse taxon sampling/many genes [6] both have resolved many nodes within monocots, but several crucial nodes remain unresolved. A robust higher-level phylogenetic framework supported by multiple genes from each genomic partition, particularly the nuclear genome, is essential for inferring patterns of diversification in monocots.

Despite morphology uniting monocots, the lineage contains huge variation in life history and morphological traits. Dominance in both terrestrial and aquatic ecosystems highlights the importance of monocots in most habitats. Monocots represent the full range of growth forms, including but not limited to annuals, perennials, bulbs/rhizomes, succulents, erect, trailing, and epiphytes. Of the 400 species of mycoheterotrphic plants, 88% are monocots. They possess a suite of characteristics making them especially suited to the demands of mycoheterotrohy, including a primarily herbaceous habit and anatomically appropriate roots [7]. Both incredibly speciose (grasses, orchids) and taxonomically sparse (Acorales, Petrosaviales, Dasypogonaceae) lineages occur in monocots. Monocots also include a wide variety of inflorescence structures, include the largest unbranched inflorescence (*Amorphophallus*, Araceae, Alismatales), largest branched inflorescence

(*Corypha*, Arecaceae, Arecales) and smallest flower (*Wolffia*, Lemnaceae, Alismatales). These kinds of character variation provided the opportunity to test relationships between morphological traits, like the co-occurence of net venation and fleshy fruits with shaded habitats [8].

Like all plants, monocots contain three genomic partitions: two maternally inherited organellar genomes, the plastome (from the chloroplast) and mitogenome (from the mitochondria), and a biparentally inherited nuclear genome. The variation of monocot life history traits is reflected in nuclear genomic variation. The nuclear genome of monocots represents levels of genomic diversity similar to other angiosperms regarding range of chromosome numbers, polyploidy and GC content. However, monocots exhibit remarkable variation in chromosome packaging/organization and genome size [9], making them ideal models to study evolution of such characteristics. The organization of chromosomes into bimodal karyotypes, in which a genome contains two distinct sizes of chromosomes, is more common in monocots, including Asparagales [10]. Even more variable is the range in nuclear genome sizes (DNA content) in monocots, as they have some of the largest genome sizes recorded to date and exhibit various modes of genome expansion and contraction throughout lineages [9]. Large genomes consist of large chromosomes easily visualized with microscopy, making them early model systems for the study of cytogenetics [e.g., 11]. Several monocot lineages also include dioecious species with nascent sex chromosomes e.g., Asparagus [12], which provides opportunities to link cytogenetic traits with life history traits.

Substantial variation in life history traits and genome size have resulted in unique patterns of molecular evolution. Early studies identified several monocot lineages as possessing quite variable rates of molecular evolution [13]. Tests across angiosperms, including monocot Commelinids, identified varying evolutionary rates correlated to life history traits [14]. Monocots in particular exhibit heterogeneous rates of molecular evolution in mitochondrial genes [15]. Additionally, molecular evolutionary studies are complicated in monocots by the predominance of unique life history traits. Mycoheterotrophic taxa, for example, lack many chloroplast genes commonly used for such studies [16].

These patterns in life history traits, genomic characteristics, and molecular evolution likely contribute to the difficulty of phylogenetic reconstruction in monocots [5, 6]. However, associations between these factors also provide the opportunity to explore a variety of questions in systematics and evolutionary biology. A plastome phylogeny sparsely sampling across monocots, but more deeply within Poales, revealed multiple shifts to wind-pollination, a conclusion previously unattainable with a poorly resolved phylogeny. An understanding of how molecular rates vary across monocots [14] can help interpret evolutionary analyses of diversification across this problematic group. Finally, additional genomic information from some of the monocots with large genomes can help elucidate patterns of genome size expansion and contraction, as many monocot lineages remain poorly sampled [9]. We are moving towards a better understanding of how these factors affect monocot evolution, which will allow for more specific tests of the role each plays in diversification.

Apart from the biological questions highlighted above, a suite of methodological and epistemiological issues are addressed in the following chapters. Of particular interest is how scale informs analysis. Scale, in this case, refers to two different aspects of experimental design. First, the following chapters utilize different types of data in addressing evolutionary questions. Molecular sequence data represents the smallest scale, at which the genome can be analyzed at the nucleotide level. Whole genome data includes sampling from multiple genomes (nuclear, mitochondrial, and chloroplast), alterations to gene order and chromosome structure, and broad scale changes to genome size. At the largest scale, data representing the organism (rather than molecules) includes morphological and life history variation. These types of data vary according to inherent complexity and levels of diversity. Obtaining each type of data, as well as analyzing and interpreting requires particular technology and skills. Second, the taxonomic level being evaluated should be selected using the question as a guide. Higher taxonomic scales, at the level of orders or families, involve much older nodes and deeper divergences than do comparisons at the generic or specific level. Each level of scale contains associated levels of uncertainty. In designing my study, I repeatedly considered what level is appropriate taxonomically and for obtaining data when addressing particular evolutionary questions?

The preceding questions mainly involve practical issues related to methodological implementation. From a theoretical standpoint, however, we are experiencing a transition in evolutionary biology. Classic systematic treatments focused entirely on morphological characters to determine relationships. Molecular systematics emerged as a way to sample a genome for characters, and relationships were discerned from modeling evolution using

DNA sequences. Modern systematics is moving towards sampling whole genomes, which brings a wealth of information from which evolutionary patterns can be gleaned, as well as concomitant problems for analysis. Regardless, we are rapidly gaining ground in resolving the tree of life. As remaining questions in organismal phylogenetics are being answered, an increasing emphasis is being placed on using phylogenetics to test hypotheses and experimentally infer answers related to organismal diversification, population genetics, molecular/cellular/developmental biology, and a multitude of other areas of biological research. Rather than a phylogeny being the end result of a systematic study, a phylogenetic tree now serves as a tool with which to answer even more valuable questions about the manner in which life evolved.

The fusion between methodological considerations and the changing face of systematics provides the opportunity to explore two broad questions in evolution and ecology. First, what is the historical context for evolution of particular plant lineages? Extant diversity in plants includes amazing variation in morphology, life history, and biogeography. A phylogenetic context provides the best opportunity to explore the driving forces behind evolution of this diversity. Improved understanding of evolutionary relationships in plants will now allow determination of this historical context. Second, how do genomic characteristics affect plant evolution and adaptation? Whole-genome characteristics, like karyotype and genome size, represent an interesting juncture between molecular and morphological characters. These characteristics are especially labile in plant groups because of prolific and influential phenomena like hybridization and polyploidy. Little is known, however, about the role these genomic changes play across the plant kingdom in shaping

diversity of lineages. These two broad questions seek to explain the mechanisms and pressures associated with plant diversification.

The following chapters differ in their approach to addressing each of the preceding questions. Chapter 2 uses molecular data sampled from across the mitogenomic, plastome, and nuclear genomes to infer a robust phylogeny across monocots. A newly evaluated fossil dataset is used to calculate divergence times for each monocot order; when combined with extant species counts for each group, these dating estimates provide insight into the context of monocot diversification since the Cretaceous. Chapter 3 provides an example of monocot evolution on the lowest taxonomic level by evaluating taxonomic classification in the *Tradescantia* alliance, a group of closely related genera with wide variation in life history traits, biogeography, and genome size. Finally, Chapter 4 approaches monocot evolution on a narrower taxonomic scale, and investigates the effects of genome size and other characteristics on application of low-redundancy genome sequencing in the Asparagales, a non-model lineage. The methods described in this chapter provide an accessible method with which to obtain ample data for phylogenetic and ecological genetic purposes. Cumulatively, these chapters illustrate the manner in which different types of data and various taxonomic levels can provide the context for both asking and answering evolutionary questions.

Literature Cited

1. Stevens PF (2001 onwards) Angiosperm Phylogeny Website.

2. Stevenson DW, Davis JI, Freudenstein JV, Hardy CR, Simmons MP, et al. (2000) A phylogenetic analysis of monocotyledons based on morphological and molecular character sets, with comments on the placement of Acorus and Hydatellaceae. In: K. L. Wilson DAM, editor. Monocots: Systematics and Evolution. Collingwood, Victoria, Australia: CSIRO Publishing.

3. APGIII (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal Of The Linnean Society 161: 105-121.

4. Chase MW, Stevenson DW, Wilkin P, Rudall PJ (1995) Monocot systematics: a combined analysis. In: Rudall PJ, Cribb PJ, Cutler DF, Humphries CJ, editors. Monocotyledons: Systematics and Evolution. Richmond, Surrey, UK: Royal Botanic Gardens, Kew. pp. 685-730.

5. Chase MW, Fay MF, Devey DS, Maurin O, Ronsted N, et al. (2006) Multigene analyses of monocot relationships: A summary. Aliso 22: 63-75.

6. Graham SW, Zgurski JM, McPherson MA, Cherniawsky DM, Saarela JM, et al. (2006) Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. Aliso 22: 3-21.

7. Imhof S (2010) Are Monocots Particularly Suited to Develop Mycoheterotrophy? In: Seberg P, Barfod, Davis, editor. Diversity, Phylogeny, and Evolution in the Monocotyledons. Denmark: Aarhus University Press. pp. 11-23.

8. Givnish TJ, Pires JC, Graham SW, McPherson MA, Prince LM, et al. (2005) Repeated evolution of net venation and fleshy fruits among monocots in shaded habitats confirms a priori predictions: evidence from an ndhF phylogeny. Proceedings Of The Royal Society B-Biological Sciences 272: 1481.

9. Leitch IJ, Beaulieu JM, Chase MW, Leitch AR, Fay MF (2010) Genome size dynamics and evolution in monocots. Journal of Botany 2010: 18.

10. Pires JC, Maureira IJ, Givnish TJ, Sytsma KJ, Seberg O, et al. (2006) Phylogeny, genome size, and chromosome evolution of Asparagales. Aliso 22: 285-302.

11. Darlington CD (1929) Chromosome behavior and structural hybridity in the Tradescantiae I. Journal of Genetics 21: 207-286.

12. Telgmann-Rauber A, Jamsari A, Kinney MS, Pires JC, Jung C (2007) Genetic and physical maps around the sex-determining M-locus of the dioecious plant asparagus. Molecular Genetics and Genomics 278: 221-234.

13. Gaut BS, Muse SV, Clark WD, Clegg MT (1992) Relative rates of nucleotide substitution at the rbcl locus of monocotyledonous plants. Journal of Molecular Evolution 35: 292-303.

14. Smith SA, Donoghue MJ (2008) Rates of Molecular Evolution Are Linked to Life History in Flowering Plants. Science 322: 86-89.

15. Petersen G, Seberg O, Davis JI, Goldman DH, Stevenson DW, et al. (2006) Mitochondrial data in monocot phylogenetics. Aliso 22: 52-62.

16. Merckx V, Freudenstein JV (2010) Evolution of mycoheterotrophy in plants: a phylogenetic perspective. New Phytologist 185: 605-609.

CHAPTER 2

PHYLOGENETICS, DIVERGENCE TIMES, AND DIVERSIFICATION FROM THREE GENOMIC PARTITIONS IN MONOCOTS

ABSTRACT

Resolution of evolutionary relationships among monocot orders remains problematic despite the application of various taxon and molecular locus sampling strategies. In this study we sequenced and analyzed a small fragment of the low-copy, nuclear-encoded phytochrome C (*PHYC*) gene and combined these data with the multigene data set (four plastid, one mitochondrial, two nuclear ribosomal loci) of Chase et al. [1] to determine if adding this marker improved resolution and support of relationships among major lineages of monocots. The addition of *PHYC* to the multigene dataset increases support along the backbone of the monocot phylogeny, although relationships between orders of commelinids remain elusive. We also estimated divergence times in monocots by applying newly-evaluated fossil calibrations to the resolved phylogenetic tree. Our relaxed constraint for the age of angiosperms allowed estimation of the age of monocots (132-163 Ma for extant lineages), and improved estimates for each order of monocots that in some cases vary substantially from previous estimates. We used three tests of whole-tree diversification to determine that monocots exhibit a characteristic pattern of rapid early diversification from high speciation rates that decrease through time. Furthermore, three orders (Asparagales, Poales, and Commelinales) exhibit significant shifts in diversification

rate in recent evolutionary history. We finally describe resulting patterns in the context of radiation of other relevant plant and animal lineages on a similar timeframe. While much work is still required to fully understand the historical context of monocot evolution, we improve knowledge of monocot evolution with a more robust phylogeny and improved divergence time estimates.

INTRODUCTION

Molecular phylogenetics has greatly improved our understanding of the evolutionary origin of monocots as well as relationships within this diverse lineage. The results of a combined analysis of 17 plastid loci and nuclear phytochrome C *(PHYC)* across angiosperms inferred monocots as a monophyletic group sister to *Ceratophyllum* and eudicots with strong statistical support [2]. Angiosperm Phylogeny Group [3] segregated monocots into 81 families and 10 orders; two families (Dasypogonaceae, Petrosaviaceae) remain unplaced to order. The two most recent and comprehensive molecular phylogenetic studies improved resolution and support for major lineages by pursuing different sampling strategies. Graham et. al [4] used fewer taxa with more loci from only the plastid genome. Chase et. al [1] used more comprehensive taxon sampling with fewer loci from plastid, mitochondrial, and nuclear genomes. Both analyses provide strong support for the monophyly of all orders as defined by APG II and for the families Dasypogonaceae and Petrosaviaceae. There is some support for relationships among monocot orders; however, several higher relationships resolved with only low to moderate support (Figure 1). In

particular, while strongly supported as monophyletic, relationships among orders of commelinids are difficult to elucidate [1,4,5,6].

The limitations of phylogenetic reconstruction methods combined with a notable deficiency of fossil calibration points has limited previous studies, resulting in a wide range of uncertainty in divergence times in monocots. The first evaluation of monocot divergence times utilized extensive taxonomic sampling (878 taxa, or "800+") of a single plastid locus (rbcL), eight fossil calibrations, and non-parametric rate smoothing (NPRS) to date the divergence of all major monocot lineages to the early (lower) Cretaceous [7]. Anderson and Janssen [8] reanalyzed this dataset with five additional fossil calibrations and the application of two new dating methods, penalized likelihood (PL) and a sister-lineage smoothing method implemented in the program PATHd8. The additional fossils had little effect on divergence times for both NPRS and PL, but PATHd8 returned much younger divergence times for a number of monocot lineages, similar to other studies comparing divergence times resulting from these programs [9]. Magallon and Castillo [10] evaluated divergence times and diversification across angiosperms using a stricter set of criteria for fossil calibrations and Bayesian inference; dates from this analysis were intermediate to the NPRS/PL and PATHd8 analyses. Variation in parameters used to date lineages and/or differences in the datasets (taxa and data) leads to wide confidence intervals for each age [11]; in the case of monocots, major sources of variation include numbers of taxa and molecular loci.

There has been great progress in circumscribing relationships among monocot orders and in dating divergence times of major lineages using uniparentally inherited

organellar DNA of the chloroplast and the mitochondrion and high copy nuclear ribosomal (nrDNA) loci [7,8,10]. Low copy nuclear genes provide unlinked loci with which to independently test phylogenetic hypotheses derived primarily from uniparentally inherited and linked chloroplast markers. Moreover, the combination of low copy nuclear loci with other plastid, mitochondrial, and high-copy nuclear loci provide a robust dataset with which to evaluate both phylogenetic relationships and estimate divergence times.

In this study, we improved the resolution of estimates of monocot phylogeny and divergence times by adding low copy nuclear gene data and applying new fossil calibrations. DNA sequence variation in low-copy nuclear phytochrome genes was effective in resolving phylogenetic relationships across angiosperms [e.g., 12,13,14,15]. This family of red and far/red light sensing proteins is well characterized in several angiosperm species and comprises a small number of genes evolving independently in angiosperms; establishment of *PHYC* as single copy validates its use in phylogenetic analysis [16]. We sequenced and analyzed a small fragment from exon I of the nuclear encoded *PHYC* gene for most monocot and several outgroup families. *PHYC* data were combined with the multigene data set of Chase et al. [1] to determine if adding this marker improved resolution and support of relationships among the major lineages of monocots, particularly at unresolved or weakly supported nodes.

We also estimated divergence times by applying new, robust fossil calibrations to a resolved phylogenetic tree calculated from the multi-locus dataset representing all three plant genomes, including the low copy nuclear gene *PHYC*. We present an estimate for stem lineage (SL, includes first divergence of lineage) and crown group (CG, only extant taxa)

monocots that is slightly older than previous estimates. Our divergence estimates for monocot orders also vary substantially from previous dates for several lineages. We use three methods to evaluate diversification in monocots, and interpret resulting patterns in the context of other relevant plant and animal lineages radiating at the same time.

MATERIALS AND METHODS

Taxon Sampling

Taxon sampling was identical to the multilocus data sets of Chase and colleagues [1,17,18]. These data sets included 124 species representing all 11 orders of the monocots and Dasypogonaceae [19] and 17 taxa representing early-diverging angiosperm lineages [3,13,20]. Ten eudicot taxa were added to provide a more complete picture of the sister group to monocots, as well as to improve divergence time estimates. Taxon names (and substitutions), voucher information, and accession numbers are provided in Table 1. Tip labels in all trees correspond to the taxon name from Chase et. al [1].

DNA extraction, PCR, cloning, and sequencing

In most cases the DNA used for amplification was the same as used in previous molecular phylogenetic studies of the monocots (Table 1) [1,17,18]. Other samples represented the same genus or family when DNA accessions were unavailable and/or did not amplify; estimations of familial relationships using similar procedures have shown that such substitutions have not had adverse effects on phylogenetic studies at higher

taxonomic levels since these families are monophyletic [20,21]. Genomic DNA was extracted from fresh or silica-dried leaf material of replacement samples following a modified CTAB procedure [22] using 3X-6X CTAB and 2 M NaCl [23]. For most specimens approximately a 1.2 kb region within exon 1 of the nuclear encoded *PHYC* gene was amplified using primers c230f and c623r [13,14,16].

For taxa that did not amplify using this protocol, additional primers were designed manually based on the original primers but made less degenerate for specific orders (Table 2). Amplification with the newly designed primers used the Qiagen® *Taq* DNA polymerase system (Qiagen Inc. USA, Valencia, CA) in the following 50 µl reaction mixture: template DNA ~100 ng, 2 µl of each primer at 10 µM, 5 µl of 10X Qiagen® PCR Buffer (with 15 mM $MgCl_2$), an additional 2 µl of 25 mM $MgCl_2$, 4 µl of 2.5 mM each dNTPs, and 0.4 µl of Qiagen® *Taq* (5U/µl). PCR reactions utilized the following conditions: an initial denaturing step of 94° C for 5 minutes, 40 cycles at 94° C for 1 min., 55° C for 1 min., 72° C for 1 min. 30 sec., and a final extension step of 72° C for 20 min. All PCR products were visualized on a 1.5% agarose gel, and 1.2 kb bands were excised and purified, ligated into plasmid and cloned using the TOPO TA Cloning® Kit (Invitrogen Corp., Carlsbad, CA). We screened at least 10 positive (white) colonies using PCR and M13F and M13R primers using Sanger sequencing. The resulting products were purified prior to sequencing, and yielded at least 6 complete clone sequences per taxon.

PHYC phylogenetic analysis

Forward and reverse trace files for each sequenced clone were assembled into complete sequences using SeqMan Pro version 7.1.0 (DNASTAR, Madison, WI). Vector ends were identified and trimmed manually. The identity of edited *PHYC* sequences was verified by the presence of easily recognized amino acid sequence hallmarks. All *PHYC* clones were initially aligned for each monocot order using MegAlign version 7.1.0 (DNASTAR) followed by manual alignment as translated amino acids using MacClade 4.0 [24]. Nucleotide sequence alignments within order were unambiguous and did not contain large insertion/deletion polymorphisms. Preliminary phylogenetic analyses of all *PHYC* clones within each order indicated clones from the same taxon were monophyletic (data not shown). One clone from each taxon was randomly chosen to represent the species in final phylogenetic analysis.

One *PHYC* clone per taxon was added to the final dataset and aligned as amino acid sequences by MUSCLE [25,26] before back-translating to nucleotide sequences for maximum likelihood (ML) phylogenetic analysis. ML analyses were run with *Amborella trichopoda* as the outgroup using RAxML v. 7.0.4 [27] and a GTRCAT [28] approximation of molecular evolution, which is suitable for large datasets. Bootstrap analyses for phylogenies were calculated from 100 replicates.

Concatenated phylogenetic analysis

For combined analyses, the *PHYC* data set described above was added to the previous seven-gene data set of Chase et al. [1], which includes data from four chloroplast loci (*atp*B,

*mat*K, *ndh*F, *rbc*L), one mitochondrial locus (*atp*A), and two nuclear ribosomal loci (18S and 26S). As the original seven-gene matrix was not complete (all loci for all taxa), sequences made available on GenBank since initial construction of this matrix were added (Table 1). We excluded all characters previously removed in the Chase et al. [1] study. Alignment and ML tree building parameters were similar to those used in the PHYC alone dataset but were conducted as partitioned analyses. We constrained outgroup topology to the current best estimate of relationships [29] for more accurate placement of fossil taxa.

Divergence times and diversification

Fossils were selected from within monocots and from the basally derived angiosperm and eudicot outgroups to constrain divergence time estimates (Table 3) and generally followed the recommendations of Gandolfo et. al [30]. CG (crown group) refers to the node from which extant lineages of a group diverge, whereas SL (stem lineage) refers to the node directly below the CG; SL represents the divergence of both extant and extinct members of the lineage in question. Fossils 1-6 constrain basally derived angiosperm lineages and fossil 7 fixes the age of eudicots; these constraints were selected from applicable fossils in Magallon et. al [10]. We re-evaluated available monocot fossils for applicability and validity, and these calibrations represent substantial alterations to previous fossil selection for divergence times in monocots. Although *Mayoa portugalica* (fossil 8) is placed in tribe Spathiphyllae, there is not enough taxon sampling to allow the constraint of this fossil at this position; instead the fossil constrains the CG Alismatales. There is some debate regarding the placement of *Nuhliantha* and *Mabelia* (fossil 9) in the Triuridaceae, but

phylogenetic analysis of fossil flowers establish them as the oldest unequivocal monot flowers [31]; they serve as a constraint for the CG Pandanales based on our sampling. Pollen and leaves from *Sabalites carolinensis* [fossil 10, 32] allow constraint for SL Arecales. Fruits for *Spirematospermum chandlerae* [fossil 11, 33] as well as two other fossil genera [34] support constraint for SL Zingiberales (divergence from Commelinales). Finally, various phytoliths (fossil 12) constrain SL Poaceae to be nearly as old as continental drift evidence from the breakup of Gondwana [35]. The previous five fossils are the best estimates for age constraints across monocots (Gandolfo, pers. comm.); several other fossils were considered for inclusion as constraints but were excluded because their ages were too young to contribute meaningfully to the analysis [36, 37]. Stratigraphic positions of fossils for constraints were transformed to minimum ages using the upper (younger) bound of the interval based on the stratigraphic timescale of Gradstein and Ogg [38]. We allowed for maximum flexibility in estimation of basal nodes by setting the maximum age of angiosperms at 160 Ma, the median value for current angiosperm age estimates [39].

Previous work on sources of error in divergence time analysis suggests that alternative tree topologies do not affect dating estimates [11], presumably because branch lengths important to stem lineages and crown groups remain relatively constant. Divergence time analyses were calculated using the eight-gene combined ML tree and associated branch lengths (Figure 3). Divergence times were estimated using a semiparametric method implemented in r8s v1.70 [40] using penalized likelihood [41], TN algorithm with bound constraints, three initial starts and fossil-based cross validation [42]. A test for the application of a molecular clock failed, validating the use of relaxed molecular clock

approaches. An optimal smoothing parameter was estimated by testing values from log λ_{10}=0 to 1.4 at intervals of 0.2. We obtained confidence intervals for the PL analysis by testing the same calculations with the upper (140 Ma) and lower (200 Ma) bounds of the current angiosperm age estimates. See Bell [39] for a complete discussion of current dating of CG angiosperms.

We used two methods to evaluate diversification in monocots. First, a lineage through time [LTT, 43] plot was constructed in the R using the APE package [44] to visualize the rate of diversification across the tree. Second, we used SymmeTREE [45] to implement tests of diversification throughout the tree. This program uses tree topology and tree-wide species diversity to determine if branches of a tree have diversified under significantly different rates, and to identify branches along which shifts in diversification have occurred. We trimmed the tree to include only ingroup (monocot) taxa, cut out a few extraneous taxa for diversity estimate purposes, and obtained species counts for taxonomic groups from the Angiosperm Phylogeny Website [46]; each tip generally corresponded to a family or subfamily.

RESULTS

PHYC analysis

The final version of the *PHYC* alone data set used in this study included 132 taxa comprising 1113 bp of exon 1 of the *PHYC* gene corresponding to 371 aligned amino acids (Table 1); 81.4% of the positions in this matrix were variable positions and 12% missing

data/gaps (excluding taxa for which no *PHYC* data were available). ML analysis of *PHYC* resulted in a tree with final ML optimization likelihood of -283376.242765 and was fairly congruent to plastid phylogenies of monocots. While most orders are supported as monophyletic, there is little support for relationships among major lineages (Figure 2). The earliest diverging lineages in both Dioscoreales (Nartheciaceae) and Asparagales (Orchidaceae) are not included with their assigned orders, although paraphyly is not strongly supported.

Combined eight gene data set and analyses

The data set that includes the seven loci from Chase et al. [1] combined with the *PHYC* data presented in this study included 151 taxa, an aligned length of 11,459 bp, 61.1% of which were variable, 2.9% missing data/gaps, and a tree with final ML optimization of -56310.480359 (Figure 3). Because the sampling for this paper follows that of Chase et al. [1] we will only highlight areas of conflict or where there were differences in resolution/support (indicated by bootstrap support, or BS). Also, following Chase et al. [1] terminals will be described using family names and not the names of representative genera; we will focus on placement and support for major lineages (11 orders and Dasypogonaceae).

Acorales—The combined data set resulted in monophyly of the monocots including Acorales (BS=100). Acorales is strongly supported as sister to the rest of the monocots (BS=100); monophyly of this monogeneric order is also strongly supported (BS=100).

Alismatales—Placement of this order as the next branching lineage above Acorales is strongly supported as well as the monophyly of this order (BS=100). Sampling in this large lineage is somewhat sparse with fewer than half of extant families represented.

Petrosaviales—Both the monophyly of this order and its position as the next branching lineage above Alismatales are strongly supported (BS=100). Sampling of this order includes representatives of both genera.

Dioscoreales/Pandanales—There is support for the sister relationship of these two orders (BS=81) as well as their placement as the next branching lineage above Petrosaviales and sister to the rest of the monocots (BS=99). Monophyly of Dioscoreales is strongly supported (BS=94) and includes Nartheciaceae (unlike the *PHYC* alone analyses); all families of this order are represented. Monophyly of Pandanales is strongly supported (BS=100); sampling of this order includes representatives for all 5 families.

Liliales—The position of Liliales as the next branching lineage above Dioscoreales + Pandanales is moderately supported (BS=90). Monophyly of Liliales is also strongly supported (BS=95). All ten families were represented.

Asparagales—Support for the placement of Asparagales as the next branching lineage above Liliales and sister to the commelinids is weak (BS=62). The order (including Orchidaceae) is monophyletic (BS=93). Most families are represented.

Commelinids—The commelinid lineage is strongly monophyletic (BS=100), but resolution is still lacking among most of the orders and Dasypogonaceae. The placement of the four major clades in the commelinids (Arecales, Dasypogonaceae, Commelinales/Zingiberales, and Poales) remains uncertain.

Arecales—This monofamilial order is strongly monophyletic (BS=100). Association of this order with Dasypogonaceae is not supported (BS=25).

Dasypogonaceae—This small but distinct lineage is well represented in this study (3 of 4 genera) and is strongly monophyletic (PB=100, LB=100, PP=1.0).

Commelinales/Zingiberales—The sister relationship of these two orders is strongly supported as is the monophyly of each of these two orders (all with BS=100). Both of these orders are well sampled in this study with representatives from all 5 families of Commelinales and from all 8 families of Zingiberales.

Poales—The monophyly of the Poales is strongly supported (BS=100). We recovered weak support for the relationship of Poales as sister to Commelinales + Zingiberales (BS=53). Most diversity in this lineage is represented.

Divergence times and diversification

Cross validation for PL in r8s returned an optimal smoothing parameter of 4. Divergence times for stem lineages (SL) and crown groups (CG) for all major monocot lineages are shown in Table 4. We note differences between analysis types of 10 Ma years or more for a SL or CG as this generally corresponds to a clear shift from one geological stage to another.

Our relaxed constraint for CG angiosperms allowed estimation of the divergence time of monocots, which is substantially older than previous estimates (SL=152 Ma and CG=157 Ma, Figure 4). Our analyses suggest younger divergence times for several crown groups, including Zingiberales, Dasypogonaceae, Arecales, and Petrosaviales (Table 4). Additionally,

several lineages diverge earlier that previous estimates (SL/CG Poales, SL/CG Commelinids, SL Asparagales, SL/CG Liliales, SL Petrosaviales, and SL Alismatales). We also present the first divergence time for monogeneric Acorales of 11 Ma. Our confidence intervals substantially narrow the range for divergence times of monocot lineages.

The LTT plot visually represents diversification of monocots based on tree topology (branching patterns) in the combined eight-gene ML tree (Figure 5). These graphs plot the estimated time before present (x axis) against the number of lineages (log scale, y axis). The resulting line is a species accumulation curve, which indicates tree-wide net diversification rates (rate of speciation minus rate of extinction). Overall, the curve (rate of lineage accumulation) increases rapidly before slowing down and then leveling off, a signature indicative of explosive evolutionary radiations. Evolutionary modeling suggests that such patterns can only emerge from declining speciation rates [47], supporting higher rates of diversification from a rapid radiation near the root of the tree. After the initial rapid increase (late Jurassic), there are two additional periods of increased diversification: one from 130-138 Ma (early Cretaceous) and another from 45-60 Ma (early Cenozoic, directly after the K-T boundary). Although this graph represents all taxa in the combined eight-gene tree, the same pattern emerges if only monocots are included (data not shown).

Whereas the LTT analysis incorporates tree topology and divergence times, SymmeTREE [45] analysis involves tree topology and extant species diversity for each taxonomic group. It calculates several tests of whole-tree diversification, all of which were significant [highest p-value-0.02, see 45 for explanation of tests], indicating rates vary significantly on at least one branch in the tree. A significant result for shifts in diversification

rates on a tree-wide level allowed for implementation of tests to locate where such shifts occurred. We identified five branches on the tree where shifts in diversification occurred (Table 5); all nodes are relatively speciose, indicating an increase in diversification rate. Two of these branches were statistically significant: SL Hanguanaceae/Commelinaceae and the terminal *Agave* branch (family Agavaceae, Asparagales). The remaining three returned only marginally significant results, which still indicate potentially interesting areas of the tree: the terminal branches for Commelinaceae (Commelinales), *Herreria* (family Agavaceae, Asparagales), Eriocaulaceae (Poales), and the SL of Joinvilleaceae/Ecdeiocoleaceae/Poaceae (Poales).

DISCUSSION

In this study, we improved the resolution of estimates of monocot phylogeny and divergence times by adding low copy nuclear gene data (*PHYC*) and applying new fossil calibrations. We also evaluated tree-wide diversification patterns. We confirm the monophyly of monocot orders and resolve several key relationships along the backbone of the phylogeny. Our results support the divergence of most monocot orders in the lower Cretaceous, but identify secondary points of diversification later in the geologic timescale.

Our combination of *PHYC* with the previously analyzed chloroplast, mitochondrial, and nuclear ribosomal dataset increased support for some previously uncertain relationships. Our analysis again supports the recognition of Petrosaviaceae and Dasypogonaceae as separate orders. Dioscoreales (including Nartheciaceae) is strongly supported as sister to Pandanales, and we show increased support for the placement of

Liliales and Asparagales along the backbone of the tree. However, relationships between orders of Commelinids remain ambiguous.

We present improved estimates for divergence times between monocot orders, which in some cases vary substantially from previous estimates. There are several reasons why divergence time estimates for monocots differ between analyses, including variation in fossil calibrations, tree building methods, and dating methods. A better understanding of the fossil record allows for more stringent guidelines for accepting fossils as calibration points. Identification and/or phylogenetic placement for several commonly utilized fossils for monocot divergence time calibrations have recently been called into question [48,49], and an updated geologic timescale has similarly revised dating estimates for other fossils [38]. The fossil calibrations utilized in our study have been carefully selected to minimize redundancy, represent taxonomic diversity in the fossil record, and conservatively place constraints throughout the tree. Although most of our fossil constraints only differ slightly from previously utilized fossils, precise dating and placement of these fossils can alter divergence times for several monocot orders. Additionally, our relaxed maxage constraint for CG angiosperms allows for more flexibility in estimating ages for some of the basalmost nodes in our tree. A younger maxage constraint results in all nodes constrained by fossils returning the age of constraint as a divergence time (results not shown); given the paucity of the fossil record in monocots, it is highly unlikely all sampled fossils represent the optimal age of divergence for each node.

The placement of fossils, however, relies on an ability to reconstruct a phylogeny accurately and precisely. Previous divergence time analysis with thorough sampling in the

monocots relied on MP analyses, although branch lengths were sometimes transformed using a model of molecular evolution [7]. Furthermore, phylogenies on which divergence times were based were limited almost entirely to chloroplast and nrDNA. Tree topology and resulting branch lengths of previous analyses appear to have a much greater influence on divergence times than alternative fossil calibration points. Our results are quite similar to limited results for monocots of Magallon and Castillo [10], which used similarly conservative fossil calibration points and multiple sequence loci to infer the tree from which divergence analyses were obtained. Bell et. al [50] compared divergence time estimates across angiosperms obtained from various sources (i.e., genes or data partitions) and found that divergence estimates vary widely based on the type of molecular data used. Our results corroborate findings that divergence estimates obtained with the combination of data partitions from multiple genomes effectively smooth variation from each data partition and result in more robust and reliable estimates.

Our refined estimates of divergence times for monocot orders (Figure 4) indicate most monocot lineages diverged in the lower Cretaceous. Dioscoreales, Pandanales, Liliales, and Arecales all diverged more than 10 Ma earlier than previously thought [8]. However, Zingiberales and Commelinales appear to have split from other commelinids in the upper Cretaceous, and the CG of these and several other orders (Acorales, Arecales, Dasypogonales) have experienced more rapid, recent radiations. While the number of extant species in Acorales and Dasypogonales explains the very young ages of these orders, Arecales and Zingiberales are more anomalous. Our fossil calibration for Arecales was placed at the node of palm divergence from Dasypogonaceae because of low sampling in

this order, although we do include a species from the most basally derived palm lineage [51]. When low sampling is combined with low substitution rates due to a woody habit [52], both phylogenetics and divergence time estimates for this lineage remain uniquely challenging. However, these complications do not apply to Zingiberales, as sampling of families throughout the CG is comprehensive and life history varies among lineages. Our data support an even more rapid radiation for this diverse group than previously hypothesized [53] that occurs after the diversification of almost all other major angiosperm lineages.

The Lower Cretaceous (140-110 Mya) was the setting for divergence of most monocot stem lineages, as well as the emergence of some extant crown groups. Later in the Upper Cretaceous, angiosperm dominated forests composed primarily of rosids [54] arose and created an understory suitable for the diversification of ferns [55]. Animal lineages experiencing rapid diversification at this time include placental mammals [56], amphibians [57], weevils [58], and ants [59]. Extant monocots experienced an additional rapid period of diversification 45-60 Mya, nearly 50 My after the initial divergence of orders. Delayed diversification following early origins is consistent with a “long evolutionary fuse” [60], a pattern reflected in ants [59], mammals [56] and other animals but not yet applied to plants. Alternatively, monocots may have been historically diverse, experienced high extinction rates, and left only a few remnant lineages that persisted to present. However, the sparse monocot fossil record from the early to mid Cretaceous indicates low diversity of ancestral lineages, and the appearance of relatively high levels of fossil diversity around 65 Mya [e. g., 61] supports our hypothesis of rapid radiation at that time. Interestingly, the

only significant shifts in diversification detected in our phylogeny occur quite contemporaneously, and in a few notable lineages of speciose monocots (Poales, Commelinales, Asparagales).

What factors contribute to the diversification pattern in monocots? Fern diversification has been attributed to the radiation of angiosperm dominated forests and subsequent creation of “new ecospaces into which certain lineaeges could diversify” [55]. Ancestral monocots were likely understory herbs as well, but the period of most rapid monocot diversification post-dates the fern radiation. Monocot diversification and radiation into extant lineages accelerated after the diversification of other major lineages of plants and animals. Niches were appearing as the composition of forests changed, but more importantly, newly emerged diversity in animal lineages important to plant pollination and dispersal were now available. In fact, specialized pollination modes (including Hymenoptera) are found in 75% of basal monocot families without wind pollination, and specialized pollination increased during the late Cretaceous-early Paleogene [62]. Even more important than the presence of specialized pollinators in the late Cretaceous was the availability of new seed dispersal mechanisms providing for local adaptation and selection [61]. A comparison between 77 angiosperm ant dispersed/non ant dispersed sister pairs, including 12 monocot pairs, found that ant dispersed lineages have diversified more than their sister pairs [63]. The importance of dispersal modes also explains the relatively young age of the large and diverse order Zingiberales; the presence of fleshy fruits in this order [6].

The work presented here solidifies both the relationships among and divergence times for major monocot lineages. Reconciliaton between the fossil record, phylogenetic

inference, extant species diversity, and divergence times inferred from evolutionary rates provides the context for extrapolating historical patterns and evaluating contemporary patterns of diversity in monocots. We propose a hypothetical model of monocot evolution in which speciation rates, not extinction rates, initially resulted in high levels of diversification in monocot evolution. As speciation rates slowed during the Cretaceous, levels of diversification attenuated. The radiation of ants and other animal lineages relevant to plant pollination and dispersal allowed for rapid diversification in a few key orders, setting the stage for modern evolutionary patterns in monocots.

Acknowledgements

I thank all collaborators on this work, all of whom will be co-authors for publication: Michael S. Kinney, Jill LeRoy, Olivier Maurin, Stephanie A. Stuart, Sarah Mathews, Mark W. Chase, J. Chris Pires. I am grateful to Susana Magallon and Ruth Stockey for advise on fossil calibrations, and Mark Beilstein and Nathalie Nagalingum for assistance with divergence time estimation. This work was supported by the National Science Foundation (DEB 0829849).

Literature Cited

1. Chase MW, Fay MF, Devey DS, Maurin O, Ronsted N, et al. (2006) Multigene analyses of monocot relationships: A summary. Aliso 22: 63-75.

2. Saarela JM, Rai HS, Doyle JA, Endress PK, Mathews S, et al. (2007) Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. Nature 446: 312-315.

3. APGII (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Botanical Journal of the Linnean Society 141: 399.

4. Graham SW, Zgurski JM, McPherson MA, Cherniawsky DM, Saarela JM, et al. (2006) Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. Aliso 22: 3-21.

5. Davis JI, Stevenson DW, Petersen G, Seberg O, Campbell LM, et al. (2004) A phylogeny of the monocots, as inferred from rbcL and atpA sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. Systematic Botany 29: 467-510.

6. Givnish TJ, Evans TM, Pires JC, Sytsma KJ (1999) Polyphyly and convergent morphological evolution in Commelinales and Commelinidae: Evidence from rbcL sequence data. Molecular Phylogenetics And Evolution 12: 360.

7. Janssen T, Bremer K (2004) The age of major monocot groups inferred from 800+ rbcL sequences. Botanical Journal of the Linnean Society 146: 385-398.

8. Anderson CL, Janssen T (2009) Monocots. In: Kumar SBHaS, editor. Timetree of Life: Oxford University Press.

9. Brown J, Rest J, Garcia-Moreno J, Sorenson M, Mindell D (2008) Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biology 6: 6.

10. Magallon S, Castillo A (2009) Angiosperm diversification through time. American Journal Of Botany 96: 349-365.

11. Sanderson MJ, Doyle JA (2001) Sources of Error and Confidence Intervals in Estimating the Age of Angiosperms from rbcL and 18S rDNA Data. American Journal of Botany 88: 1499-1516.

12. Mathews S, Sharrock RA (1996) The phytochrome gene family in grasses (Poaceae): A phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. Molecular Biology and Evolution 13: 1141-1150.

13. Mathews S, Donoghue MJ (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. Science 286: 947-950.

14. Mathews S, Donoghue MJ (2000) Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. International Journal of Plant Sciences 161: S41-S55.

15. Bennett JR, Mathews S (2006) Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. American Journal of Botany 93: 1039-1051.

16. Mathews S, Lavin M, Sharrock RA (1995) Evolution of the Phytochrome Gene Family and Its Utility for Phylogenetic Analyses of Angiosperms. Annals of the Missouri Botanical Garden 82: 296-321.

17. Chase MW, Stevenson DW, Wilkin P, Rudall PJ (1995) Monocot systematics: a combined analysis. In: Rudall PJ, Cribb PJ, Cutler DF, Humphries CJ, editors. Monocotyledons: Systematics and Evolution. Richmond, Surrey, UK: Royal Botanic Gardens, Kew. pp. 685-730.

18. Chase MW, Soltis DE, Soltis PS, Rudall PJ, Fay MF, et al. (2000) Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. In: K. L. Wilson DAM, editor. Monocots: Systematics and Evolution. Collingwood, Victoria, Australia: CSIRO Publishing.

19. Givnish TJ, Pires JC, Graham SW, McPherson MA, Prince LM, et al. (2006) Phylogenetic relationships of monocots based on the highly informative plastid gene ndhF : Evidence for widespread concerted convergence. Monocots: Comparative biology and evolution (excluding Poales). Claremont, CA, USA: Rancho Santa Ana Botanic Garden.

20. Qiu Y-L, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis MJ, et al. (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402: 404-407.

21. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, et al. (2000) Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. Botanical Journal of the Linnean Society 133: 381-461.

22. Doyle JJaJLD (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15.

23. Smith JF, Sytsma KJ, Shoemaker JS, Smith RL (1991) A qualitative comparison of total cellular DNA extraction protocols. Phytochemical Bulletin 23: 2-9.

24. Maddison DR, Maddison WP (2001) MacClade. 4 ed. Sunderland, MA: Sinauer Associates, Inc.

25. Edgar R (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

26. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

27. Stamatakis A, Hoover P, Rougemont J (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology 57: 758 – 771.

28. Stamatakis A. Phylogenetic Models of Rate Heterogeneity: A High Performance Computing Perspective; 2006.

29. Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences of the United States of America 104: 19363-19368.

30. Gandolfo MA, Nixon KC, Crepet WL (2008) Selection of fossils for calibration of molecular dating models. Annals of the Missouri Botanical Garden 95: 34-42.

31. Friis EM, Pedersen KR, Crane PR (2006) Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. Palaeogeography, Palaeoclimatology, Palaeoecology 232: 251-293.

32. Berry EW (1914) The Upper Cretaceous and Eocene floras of South Carolina, Georgia. US Geological Survey, Professional Paper 84: 1-200.

33. Friis EM (1988) Spirematospermum chandlerae sp. nov., an extinct species of Zingiberaceae from the North American Cretaceous. Tertiary Research 9: 7-12.

34. Rodriguez-de la Rosa RA, Cevallos-Ferriz SRS (1994) Upper Cretaceous Zingiberalean Fruits with in Situ Seeds from Southeastern Coahuila, Mexico. International Journal of Plant Sciences 155: 786-805.

35. Prasad V, Stromberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur Coprolites and the Early Evolution of Grasses and Grazers. Science 310: 1177-1180.

36. Ramirez SR, Gravendeel B, Singer RB, Marshall CR, Pierce NE (2007) Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. Nature 448: 1042-1045.

37. Stockey RA, Rothwell GW, Johnson KR (2007) Cobbania corrugata gen. et comb. nov. (Araceae): A floating aquatic monocot from the upper cretaceous of western North America. American Journal of Botany 94: 609-624.

38. Gradstein FM, Ogg JG (2004) Geologic time scale 2004-Why, how and where next. Lethaia 37: 175-181.

39. Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-revisited. American Journal Of Botany 97: 1296-1303.

40. Sanderson MJ (2003) r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19: 301-302.

41. Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Molecular Biology and Evolution 19: 101-109.

42. Near TJ, Sanderson MJ (2004) Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. Philosophical Transactions of the Royal Society B: Biological Sciences 359: 1477-1483.

43. Nee S, Mooers AO, Harvey PH (1992) Tempo and mode of evolution revealed from molecular phylogenies. Proceedings of the National Academy of Sciences of the United States of America 89: 8322-8326.

44. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language 20: 289-290.

45. Chan KMA, Moore BR (2005) SYMMETREE: whole-tree analysis of differential diversification rates. Bioinformatics 21: 1709-1710.

46. Stevens PF (2001 onwards) Angiosperm Phylogeny Website. http://www.mobot.org/MOBOT/research/APweb/

47. Rabosky DL, Lovette IJ (2008) Explosive evolutionary radiations: Decreasing speciation or increasing extinction through time? Evolution 62: 1866-1875.

48. Crepet WL, Nixon KC, Gandolfo MA (2004) Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. Am J Bot 91: 1666-1682.

49. Crepet WL, Gandolfo MA (2008) Paleobotany in the post-genomics era: Introduction. Annals of the Missouri Botanical Garden 95: 1-2.

50. Bell CD, Soltis DE, Soltis PS (2005) The age of the angiosperms: A molecular timescale without a clock. Evolution 59: 1245-1258.

51. Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud JC, et al. (2006) A new subfamily classification of the palm family (Arecaceae): Evidence from plastid DNA phylogeny. Botanical Journal of the Linnean Society 151: 15-38.

52. Smith SA, Donoghue MJ (2008) Rates of Molecular Evolution Are Linked to Life History in Flowering Plants. Science 322: 86-89.

53. Kress WJ, Prince LM, Hahn WJ, Zimmer EA (2001) Unraveling the evolutionary radiation of the families of the Zingiberales using morphological and molecular evidence. Systematic Biology 50: 926.

54. Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, et al. (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. Proceedings of the National Academy of Sciences of the United States of America 106: 3853-3858.

55. Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallon S, et al. (2004) Ferns diversified in the shadow of angiosperms. Nature 428: 553-557.

56. Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, et al. (2007) The delayed rise of present-day mammals. Nature 446: 507-512.

57. Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, et al. (2007) Global patterns of diversification in the history of modern amphibians. Proceedings of the National Academy of Sciences of the United States of America 104: 887-892.

58. McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD (2009) Temporal lags and overlap in the diversification of weevils and flowering plants. Proceedings of the National Academy of Sciences of the United States of America 106: 7083-7088.

59. Moreau CS, Bell CD, Vila R, Archibald SB, Pierce NE (2006) Phylogeny of the ants: Diversification in the age of angiosperms. Science 312: 101-104.

60. Cooper A, Fortey R (1998) Evolutionary explosions and the phylogenetic fuse. Trends in Ecology & Evolution 13: 151-156.

61. Crane PR, Friis EM, Pedersen KR (1995) The origin and early diversification of angiosperms. Nature 374: 27-33.

62. Hu S, Dilcher DL, Jarzen DM, Taylor DW (2008) Early steps of angiosperm-pollinator coevolution. Proceedings of the National Academy of Sciences of the United States of America 105: 240-245.

63. Lengyel S, Gove AD, Latimer AM, Majer JD, Dunn RR (2009) Ants sow the seeds of global diversification in flowering plants. PLoS ONE 4.

64. Friis EM, Pedersen KR, Crane PR (2001) Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. Nature 410: 357-360.

65. Mohr B, Bernardes-de-Oliveira M (2004) Endressinia brasiliana, a Magnolialean Angiosperm from the Lower Cretaceous Crato Formation (Brazil). International Journal of Plant Sciences 165: 1121-1133.

66. Doyle JA, Hotton CL, Ward JV (1990) Early Cretaceous Tetrads, Zonasulculate Pollen, and Winteraceae. II. Cladistic Analysis and Implications. American Journal of Botany 77: 1558-1568.

67. Doyle JA (2000) Paleobotany, Relationships, and Geographic History of Winteraceae. Annals of the Missouri Botanical Garden 87: 303-316.

68. Mai DH (1995) Entwicklung der Wasser-und Sumpfpflanzen-Gesellschaften Europas von der Kreide bis ins Quartar. Flora 176: 449-511.

69. Hughes NF, McDougall AB (1987) Records of angiospermid pollen entry into the English Early Cretaceous succession. Review of Palaeobotany & Palynology 50: 255-272.

70. Doyle JA (1992) Revised palynological correlations of the lower Potomac Group (USA) and the Cocobeach sequence of Gabon (Barremian-Aptian). Cretaceous Research 13: 337-349.

71. Friis EM, Pedersen KR, Crane PR (2004) Araceae from the Early Cretaceous of Portugal: Evidence on the emergence of monocotyledons. Proceedings of the National Academy of Sciences of the United States of America 101: 16565-16570.

72. Gandolfo MA, Nixon KC, Crepet WL (2002) Triuridaceae fossil flowers from the Upper Cretaceous of New Jersey. American Journal of Botany 89: 1940-1957.

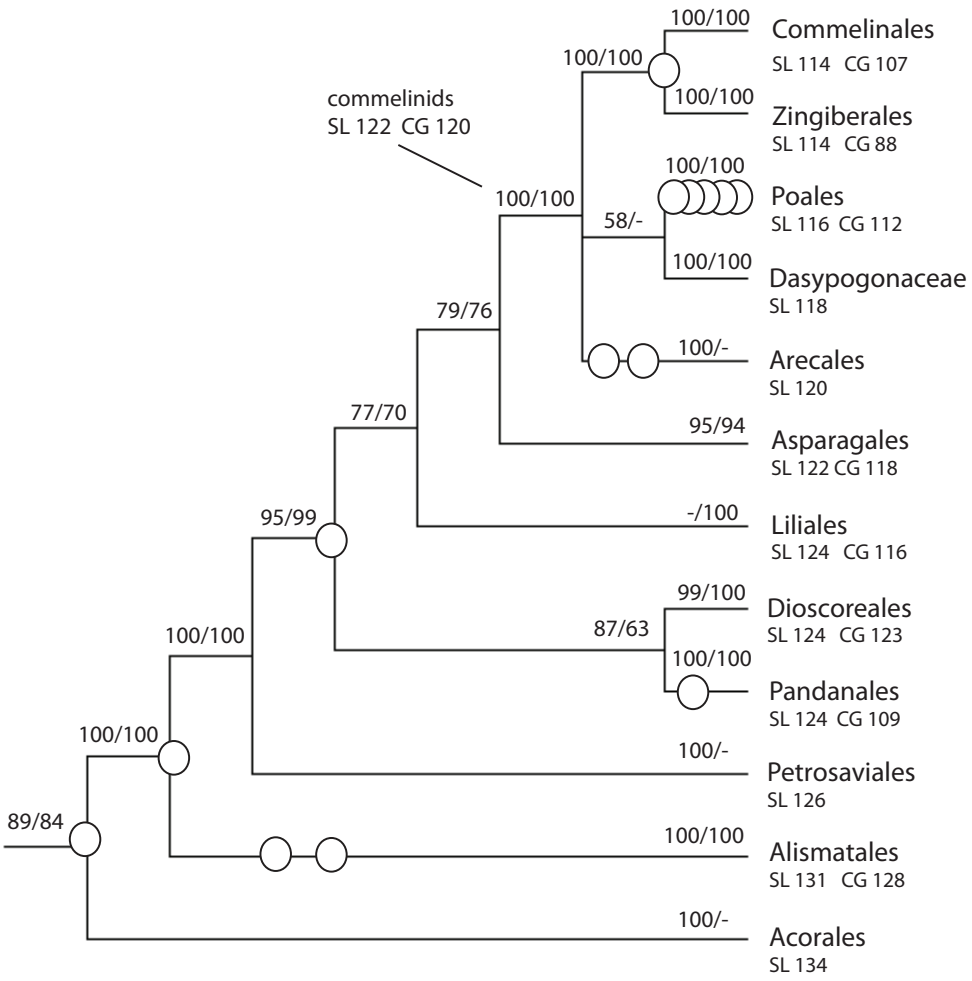

Figure 1. Summary of previously hypothesized relationships between monocots [1,4] and divergence time estimates. Numbers by nodes correspond to bootstrap values from Chase et. al [1] and Graham et. al [4], respectively. Open circles indicate fossil calibrations utilized by Anderson and Janssen [8], and values below order names indicate divergence time estimates for stem lineages (SL) and crown groups (CG) from the same study.

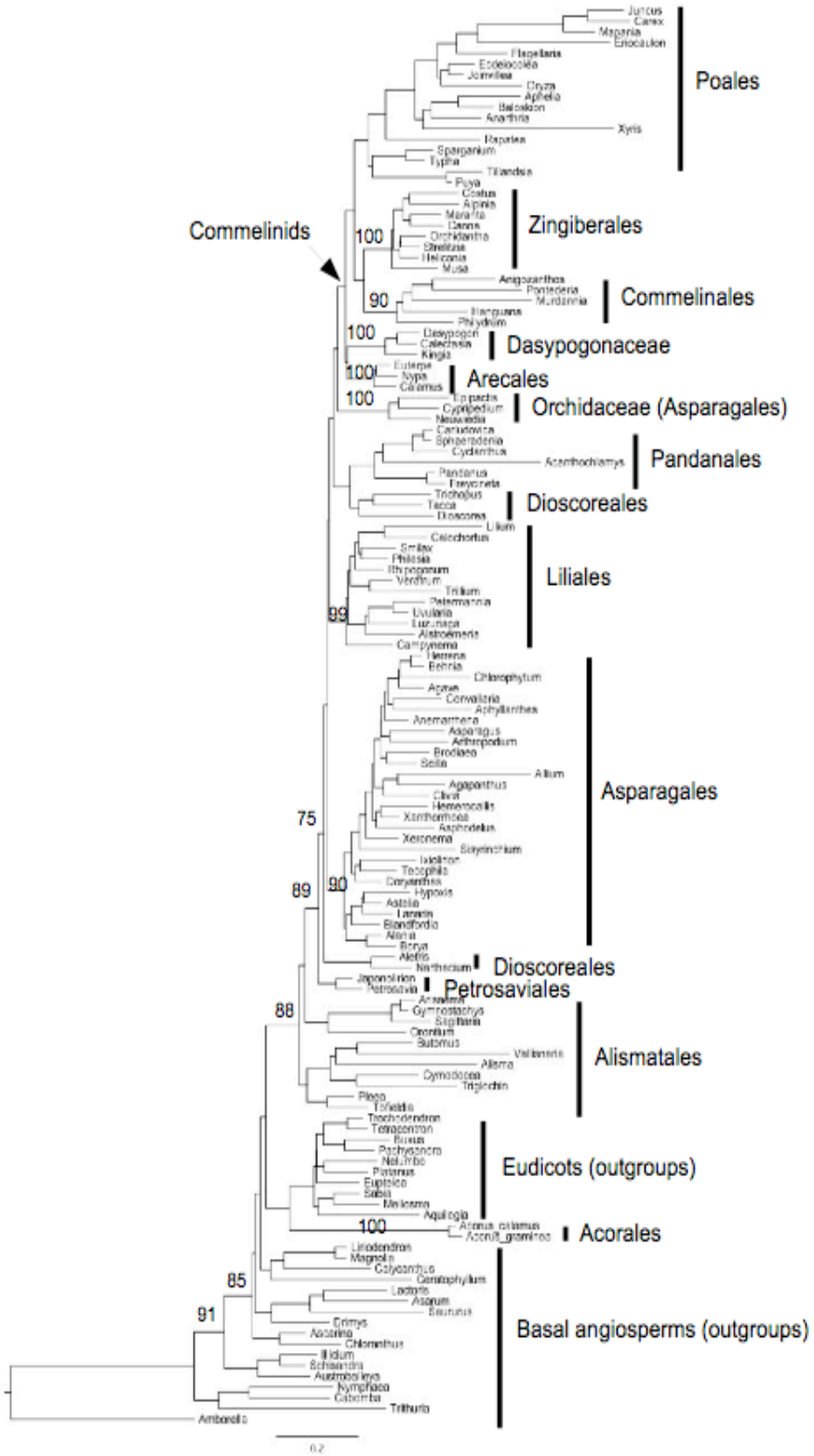
Commelinids
Poales
Zingiberales
Commelinales
Dasypogonaceae
Arecales
Orchidaceae (Asparagales)
Pandanales
Dioscoreales
Liliales
Asparagales
Dioscoreales
Petrosaviales
Alismatales
Eudicots (outgroups)
Acorales
Basal angiosperms (outgroups)
100
90
100
100
100
99
75
89
90
88
85
91
100
0.2

Figure 2. ML phylogram of monocots inferred from low copy nuclear gene PHYC. Bootstrap support (100 replicates) is shown along tree backbone and for crown groups when >70.

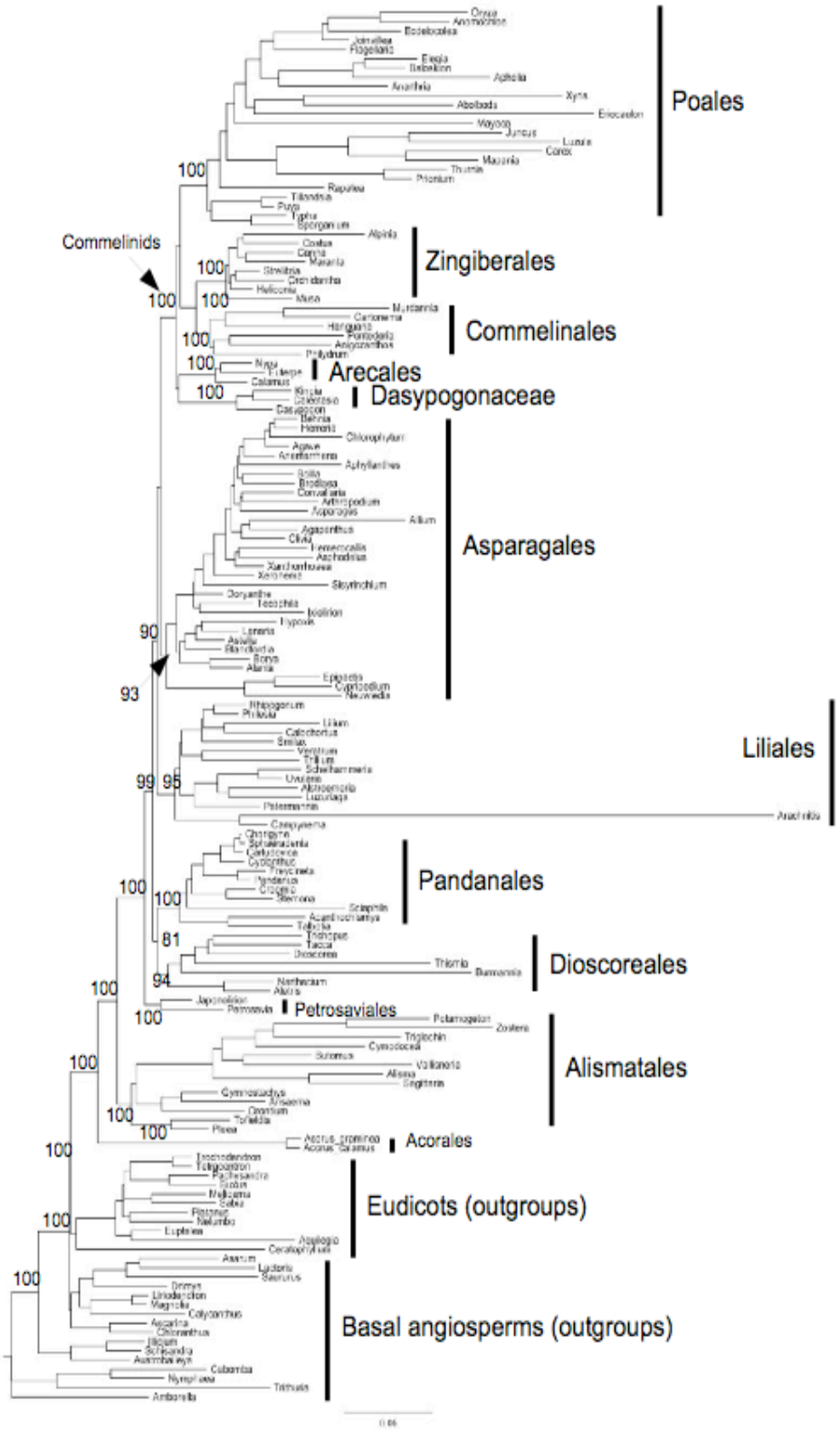
Poales
Commelinids
Zingiberales
Commelinales
Arecales
Dasypogonaceae
Asparagales
Liliales
Pandanales
Dioscoreales
Petrosaviales
Alismatales
Acorales
Eudicots (outgroups)
Basal angiosperms (outgroups)

Figure 3. ML phylogram of monocots inferred from eight gene matrix.

Bootstrap support (100 replicates) is shown along tree backbone and for crown groups when >70.

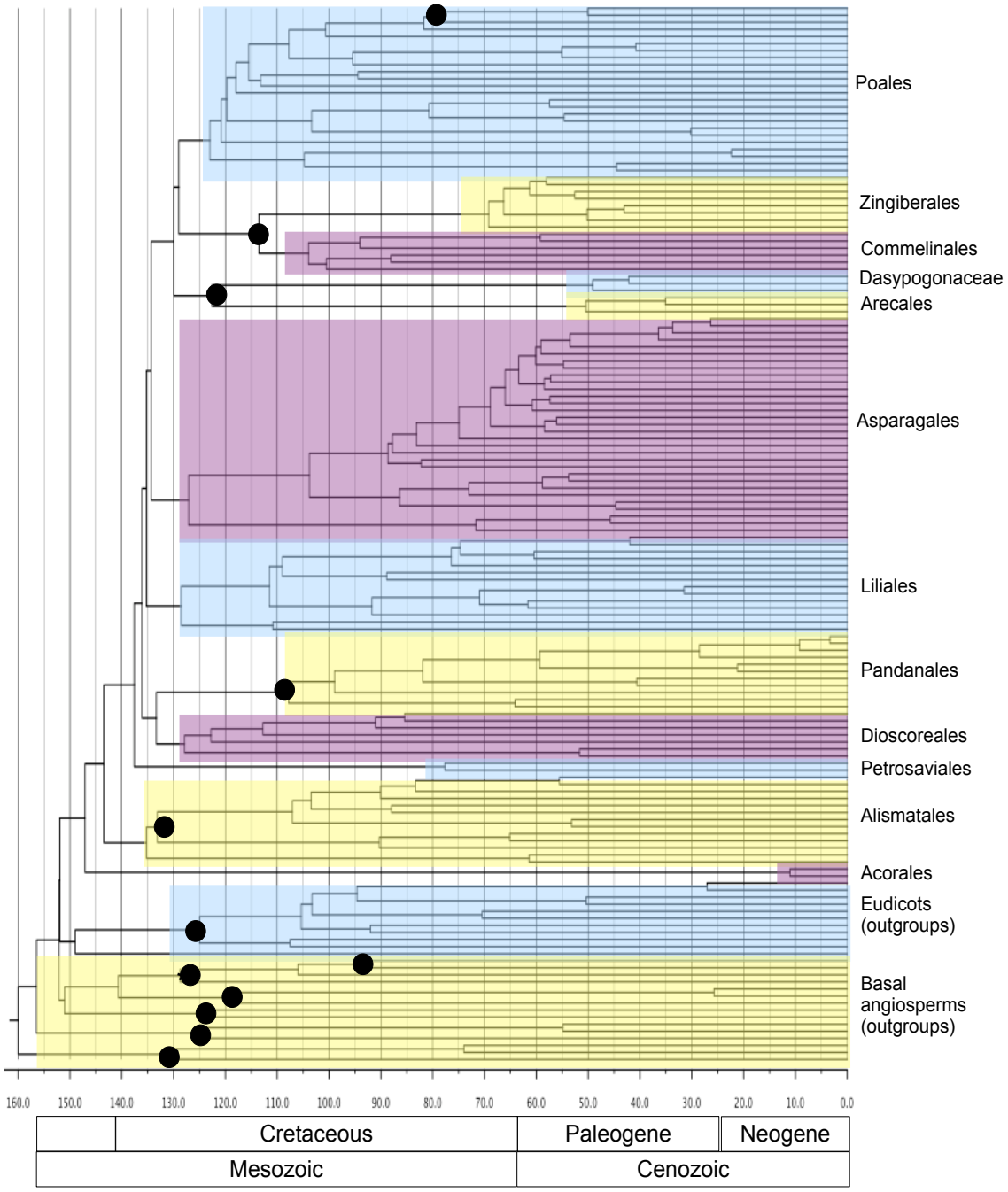
Poales
Zingiberales
Commelinales
Dasypogonaceae
Arecales
Asparagales
Liliales
Pandanales
Dioscoreales
Petrosaviales
Alismatales
Acorales
Eudicots (outgroups)
Basal angiosperms (outgroups)
160.0 150.0 140.0 130.0 120.0 110.0 100.0 90.0 80.0 70.0 60.0 50.0 40.0 30.0 20.0 10.0 0.0
Cretaceous
Paleogene
Neogene
Mesozoic
Cenozoic

Figure 4. Chronogram depicting divergence time estimates for monocot orders derived from the combined eight gene ML tree and PL. ML tree topology from Figure 4 displayed as a chronogram. Numbers by nodes report bootstrap support (BS, 100 replicates). Circles indicate placement of fossil calibrations listed in Table 3. Colored blocks represent the inclusion of taxa in crown groups. Fossils start with number 1 at the bottom and continue sequentially up the tree.

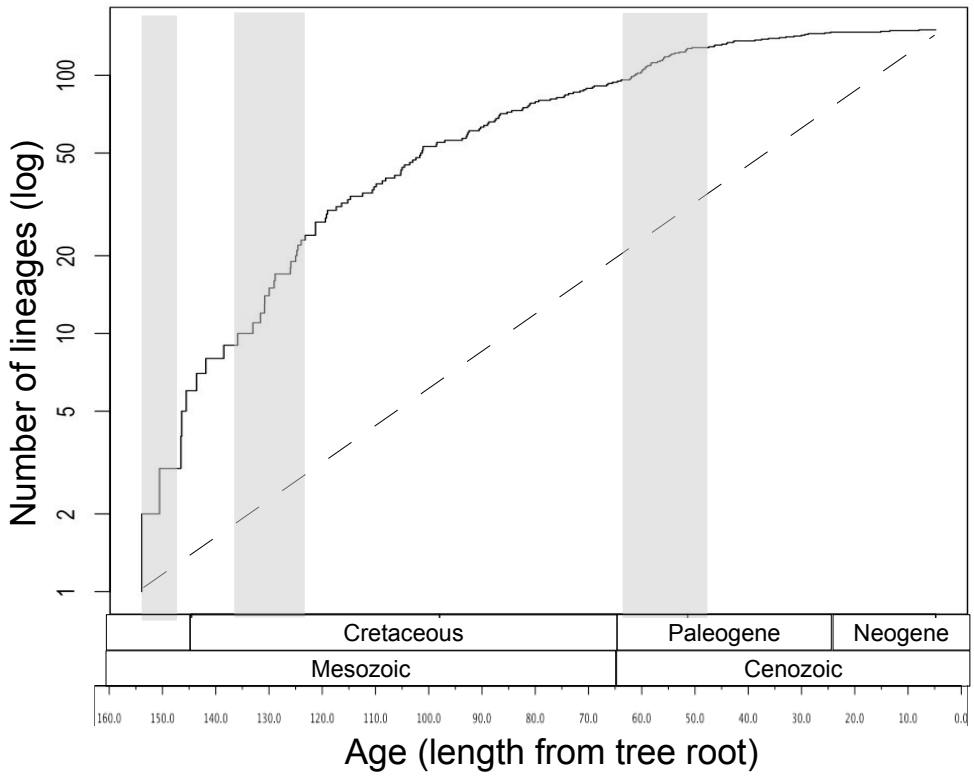

Figure 5. Lineage through time (LTT) plot of monocots from combine eight-gene chronogram. The dashed line indicates a constant diversification rate in the absence of extinction. Intervals with increased rates of diversification (steeper slope) are labeled in grey.

Table 1. Taxa and voucher information for monocot and outgroup taxa used in this study. Family assignations follow APG II [3]. A. PHYC data, B. Revised 7-gene data.

A.

Order	Family	Chase taxon	PHYC GenBank	PHYC Taxon	PHYC Collector - ID	PHYC Voucher
Amborellales	Amborellaceae	Amborella	AF190063	Amborella trichopoda	N/A	N/A
Austrobaileyales	Austrobaileyaceae	Austrobaileya	AF190069	Austrobaileya scandens	N/A	N/A
Austrobaileyales	Schisandraceae	Illicium	AF276729	Illicium oligandrum	N/A	N/A
Austrobaileyales	Schisandraceae	Schisandra	DQ981793	Schisandra chinensis_1949_6_1	N/A	N/A
Cannellales	Winteraceae	Drimys	AF190081	Drimys winteri	N/A	N/A
Ceratophyllales	Ceratophyllaceae	Ceratophyllum	AF276717	Ceratophyllum demersum	N/A	N/A
Chloranthales	Chloranthaceae	Ascarina	TBA	Ascarina_sp_1846_4_1	MWC 9601	TBA
Chloranthales	Chloranthaceae	Chloranthus	AF190077	Chloranthus spicatus	N/A	N/A
Laurales	Calycanthaceae	Calycanthus	AF190073	Calycanthus floridus	N/A	N/A
Magnoliales	Magnoliaceae	Liriodendron	AY396711	Liriodendron tulipifera	N/A	N/A
Magnoliales	Magnoliaceae	Magnolia	AF190095	Magnolia grandiflora_1856_3_1	N/A	N/A
Nymphaeales	Cambombaceae	Cabomba	AF190071	Cabomba sp.	N/A	N/A
Nymphaeales	Nymphaea	Nymphaea	AF190099	Nymphaea alba	N/A	N/A
Piperales	Aristolochiaceae	Asarum	AY396705	Asarum canadense	N/A	N/A
Piperales	Lactoridaceae	Lactoris	AF190092	Lactoris fernandeziana	N/A	N/A
Piperales	Saururaceae	Saururus	AF190107	Saururus cernuus	N/A	N/A
Acorales	Acoraceae	Acorus_cal	TBA	Acorus calamus_1845_2_1	MWC 2758	MWC 2758 K
Acorales	Acoraceae	Acorus_gram	AF190061	Acorus gramineus	N/A	N/A
Alismatales	Alismataceae	Alisma	TBA	Alisma_triviale_2075_11_4	MWC 10624	Buzgo 1013
Alismatales	Alismataceae	Sagitarria	AF190103	Sagittaria_sp	N/A	N/A
Alismatales	Araceae	Arisaema	TBA	Arisaema_sp_1846_3_1	MWC 8749	TCMK 27
Alismatales	Araceae	Gymnostachys	TBA	Gymnostachys anceps_1290_3_1	SM	SM
Alismatales	Araceae	Orontium	TBA	Orontium aquaticum_19212_4	SM	SM
Alismatales	Butomaceae	Butomus	TBA	Butomus_umbellatus_1846_5_1	MWC 11051	Mary Clare Sheahan, MCS 090 K
Alismatales	Cymodoceaceae	Cymodocea	TBA	Posidonia	TBA	TBA
Alismatales	Hydrocharitaceae	Vallisneria	TBA	C_Valisneria_asiatica_1840_6_6	MWC 6018	MWC 6018 K

Order	Family	Chase taxon	PHYC GenBank	PHYC Taxon	PHYC Collector - ID	PHYC Voucher
Petrosaviales	Petrosaviaceae	Japonolirion	TBA	Japonolirion_osense_1844_5_3	MWC 3000	Chase 2000 K
Petrosaviales	Petrosaviaceae	Petrosavia	TBA	C_Petrosavia_sp_1895_3_1	MWC 1933	K Cameron K
Alismatales	Potamogetonaceae	Potamogeton	N/A	N/A	N/A	N/A
Alismatales	Tofieldiaceae	Pleea	AF276736	Pleea_tenuifolia	N/A	N/A
Alismatales	Tofieldiaceae	Tofieldia	AY396715	Tofieldia_calyculata	N/A	N/A
Alismatales	Zosteraceae	Zostera	N/A	N/A	N/A	N/A
Asparagales	Agapanthaceae	Agapanthus	TBA	Agapanthus_campanulatus_1008	MWC 1008	TBA
Asparagales	Agavaceae	Agave	TBA	Agave_MWC.5	MWC	TBA
Asparagales	Alliaceae	Allium	TBA	Allium_haematochiton_JCP_1	JCP	WIS
Asparagales	Amaryllidaceae	Clivia	TBA	Amaryllis	M-379	TBA
Asparagales	Agavaceae	Anemarrhena	TBA	Anemarrhena asphdeloides	MWC 1022	N/A A
Asparagales	Agavaceae	Chlorophytum	TBA	Chlorophytum_K.2	TBA	TBA
Asparagales	Aphyllanthaceae	Aphyllanthes	TBA	Aphyllanthes monspeliensis	MWC 614	TBA
Asparagales	Asparagaceae	Asparagus	AF276715	Asparagus_falcatus	N/A	N/A
Asparagales	Asphodelaceae	Asphodelus	TBA	Eremurus_490K.5	MWC 490	TBA
Asparagales	Asteliaceae	Astelia	TBA	Astelia_banksii_1071	MWC 1071	TBA
Asparagales	Agavaceae	Behnia	TBA	Behnia_reticulata 419K.1	MWC 419	TBA
Asparagales	Blandfordiaceae	Blandfordia	TBA	Blandfordia_punicea_519	MWC 519	TBA
Asparagales	Boryaceae	Alania	TBA	Alania_endlicheri_JVF2944.5	JVF 2944	TBA
Asparagales	Boryaceae	Borya	TBA	Borya_sep_MWC.4	MWC	TBA
Asparagales	Ruscaceae	Convallaria	TBA	Convallaria_496.D2	MWC 496	TBA
Asparagales	Doryanthaceae	Doryanthes	TBA	Doryanthes_palmeri_19153	MWC 19153	TBA
Asparagales	Hemerocallidaceae	Hemerocallis	TBA	Hemerocallis_12067.2	MWC 12067	TBA
Asparagales	Agavaceae	Herreria	TBA	Herreria_2154.1	MWC 2154	TBA
Asparagales	Hyacinthaceae	Scilla	TBA	Scilla_JCP_PHYC_Clone2	JCP	TBA
Asparagales	Hypoxidaceae	Hypoxis	TBA	Hypoxis_hemerocallidea_1045	MWC 1045	TBA
Asparagales	Iridaceae	Sisyrinchium	TBA	Sisyrinchium_I208.11	MWC 1208	TBA
Asparagales	Ixiolirionaceae	Ixiolirion	TBA	Ixiolirion_tataricum_489K	MWC 489	TBA
Asparagales	Lanariaceae	Lanaria	TBA	Lanaria_lanata_458.7	MWC 458	TBA
Asparagales	Laxmanniaceae	Arthropodium	TBA	Arthropodium_cirratum_651	MWC 651	TBA
Asparagales	Orchidaceae	Cypripedium	TBA	Cypripedium_calceolus_O1116	MWC O1116	TBA

Order	Family	Chase taxon	PHYC GenBank	PHYC Taxon	PHYC Collector - ID	PHYC Voucher
Asparagales	Orchidaceae	Neuwiedia	TBA	Neuwiedia_veratrifolia_0883	MWC O883	TBA
Asparagales	Tecophilaeaceae	Tecophilaea	TBA	Tecophilaea_1498K.1	MWC 1498	TBA
Asparagales	Themidaceae	Brodiaea	TBA	BrodiaeaCoronariaJCP.4	JCP	WIS
Asparagales	Xanthorrhoeaceae	Xanthorrhoea	TBA	Xanthorrhoea_MWC_PHYC_Clone1	MWC	K
Asparagales	Xeronemataceae	Xeronema	TBA	Xeronema_callistmeon_653	MWC 653	K
Dioscoreales	Burmanniaceae	Burmannia	N/A	N/A	N/A	N/A
Dioscoreales	Thismiaceae	Thismia	N/A	N/A	N/A	N/A
Dioscoreales	Dioscoreaceae	Trichopsus	TBA	C_Trichopus_sempervirens_1846_9	MWC 15068	Wilkin et al 948 K
Dioscoreales	Dioscoreaceae	Dioscorea	AF276721	Dioscorea elephantipes	N/A	N/A
Dioscoreales	Dioscoreaceae	Tacca	TBA	Tacca_MPP01.4.seq		MU
Dioscoreales	Nartheciaceae	Aletris	TBA	C_Aletris_alba_1982_2_1	MWC 517	MWC 517 K
Dioscoreales	Nartheciaceae	Narthecium	TBA	Narthecium_610.2	MWC 610	K
Liliales	Alstroemeriaceae	Alstroemeria	TBA	Alstroemeria_19990_2	TBA	TBA
Liliales	Campynemataceae	Campynema	TBA	Campynema_19572_11	MWC 477	Walsh 3488 MEL
Liliales	Colchicaceae	Petermannia	TBA	Colchicum_speciosum_109	TBA	TBA
Liliales	Colchicaceae	Schelhammera	N/A	N/A	N/A	N/A
Liliales	Colchicaceae	Uvularia	TBA	C_Uvularia_perfoliata_1843_11_1	MWC 494	MWC 494 K
Liliales	Corsiaceae	Arachnitis	N/A	N/A	N/A	N/A
Liliales	Liliaceae	Calochortus	TBA	C_Calochortus_minimus_1868_1_1	MWC 239	Ness 606 PUA
Liliales	Liliaceae	Lilium	AF276733	Lilium_superbum	N/A	N/A
Liliales	Luzuriagaceae	Luzuriaga	TBA	C_Luzuriaga_radicans_1868_3_2	MWC 499	Chase 499 K
Liliales	Melanthiaceae	Trillium	TBA	C_Trillium_erectum_1982_6_2	MWC 444	MWC 444 K
Liliales	Melanthiaceae	Veratrum	TBA	C_Xerophyllum_tenax_1868_9_3	MWC 527	MWC 527 K
Liliales	Philesiaceae	Philesia	TBA	C_Philesia_buxifolia_1843_7_1	MWC 545	MWC 545 K
Liliales	Rhipogonaceae	Rhipogonum	TBA	Rhipogonum_187_8	MWC 187	MWC 187 NCU
Liliales	Smilacaceae	Smilax	AF276744	Smilax_rotundifolia_AF276744	N/A	N/A
Pandanales	Cyclanthaceae	Carludovica	AY396707	Carludovica_palmata_AY396707	N/A	N/A
Pandanales	Cyclanthaceae	Chorigyne	N/A	N/A	N/A	N/A
Pandanales	Cyclanthaceae	Cyclanthus	TBA	C_Cyclanthus_bipartitus_1845_3	MWC 1237	Chase 1237 K
Pandanales	Cyclanthaceae	Sphaeradenia	TBA	Sphaeradenia_222.7	SM	TBA
Pandanales	Pandanaceae	Freycinetia	TBA	C_Freycinetia_scandens_1868_2_5	MWC 191	Chase 191 NCU

Order	Family	Chase taxon	PHYC GenBank	PHYC Taxon	PHYC Collector - ID	PHYC Voucher
Pandanales	Stemonaceae	Croomia	N/A	N/A	N/A	N/A
Pandanales	Stemonaceae	Stemona	TBA	C_Stemona_javanica_1953_12_4[partial]	MWC 2156	MWC 2156 K
Pandanales	Triuridaceae	Sciaphila	N/A	N/A	N/A	N/A
Pandanales	Velloziaceae	Acanthochlamys	TBA	Vellozia_3477.9	TBA	TBA
Pandanales	Velloziaceae	Talbotia	N/A	N/A	N/A	N/A
Arecales	Arecaceae	Calamus	TBA	Calamus_12835.15	TBA	TBA
Arecales	Arecaceae	Euterpe	TBA	Euterpe_22038.3	TBA	TBA
Arecales	Arecaceae	Nypa	TBA	Nypa_12603.10	TBA	TBA
Dasypogonales	Dasypogonaceae	Calectasia	TBA	Calectasia_narragara_20213	TBA	TBA
Dasypogonales	Dasypogonaceae	Dasypogon	TBA	Dasypogon_20866_2	TBA	TBA
Dasypogonales	Dasypogonaceae	Kingia	TBA	Kingia_australis_2230	TBA	TBA
Commelinales	Commelinaceae	Cartonema	N/A	N/A	N/A	N/A
Commelinales	Commelinaceae	Murdannia	TBA	Murdannia_bracteata_KLH_11	TBA	MOBOT
Commelinales	Haemodoraceae	Anigozanthos	TBA	Anigozanthos_20849_2	TBA	TBA
Commelinales	Hanguanaceae	Hanguana	TBA	Hanguana_20016_5	TBA	TBA
Commelinales	Philydraceae	Philydrum	TBA	Helmholtzia_452_1	TBA	TBA
Commelinales	Pontederiaceae	Pontedaria	TBA	Pontederia_2996_3	TBA	TBA
Zingiberales	Cannaceae	Canna	TBA	Canna_paniculata_5572	TBA	TBA
Zingiberales	Costaceae	Costus	TBA	Costus_woodsonii_3911	TBA	TBA
Zingiberales	Heliconiaceae	Heliconia	TBA	Heliconia_rostrata_3907	TBA	TBA
Zingiberales	Lowiaceae	Orchidantha	TBA	Orchidantha_maxillarioides_3912	TBA	TBA
Zingiberales	Marantaceae	Maranta	TBA	Maranta_depressa_3858	TBA	TBA
Zingiberales	Musaceae	Musa	TBA	Musa_basjoo_3952	TBA	TBA
Zingiberales	Strelitziaceae	Strelitzia	TBA	Strelitzia_reginae?MPP086.1	TBA	TBA
Zingiberales	Zingiberaceae	Alpinia	TBA	Alpinia_calcarata_6171	TBA	TBA
Poales	Anarthriaceae	Anarthria	TBA	Anarthria_prolifera_437	TBA	TBA
Poales	Bromeliaceae	Puya	TBA	Puya_raimondii_2847	TBA	TBA
Poales	Bromeliaceae	Tillandsia	TBA	Tillandsia_albida_18963	TBA	TBA
Poales	Centrolepidaceae	Aphelia	TBA	Aphelia_14158_6	TBA	TBA
Poales	Cyperaceae	Carex	TBA	Carex_pleurocaula_16373	TBA	TBA
Poales	Cyperaceae	Mapania	TBA	Mapania_2713_B5	TBA	TBA
Poales	Ecdeiocoleaceae	Ecdeiocolea	TBA	Ecdeiocolea_12283_5	TBA	TBA

Order	Family	Chase taxon	PHYC GenBank	PHYC Taxon	PHYC Collector - ID	PHYC Voucher
Poales	Flagellariaceae	Flagellaria	U61204	Flagellaria_indica_206	N/A	
Nymphaeales	Hydatellaceae	Trithuria	DQ981794	Trithuria_submersa	N/A	
Poales	Joinvilleaceae	Joinvillea	AY396709	Joinvillea_ascendens_AY396709	N/A	
Poales	Juncaceae	Juncus	TBA	Juncus_effusus_MPP.4	TBA	TBA
Poales	Juncaceae	Luzula	N/A	N/A	N/A	N/A
Poales	Mayaceae	Mayaca	N/A	N/A	N/A	N/A
Poales	Poaceae	Anomochloa	N/A	N/A	N/A	N/A
Poales	Poaceae	Oryza	AB018442	Oryza_AB018442	N/A	N/A
Poales	Thurniaceae	Prionium	N/A	N/A	N/A	N/A
Poales	Rapateaceae	Rapatea	TBA	Stegolepis_sp_3486	TBA	TBA
Poales	Restionaceae	Baloskion	TBA	Baloskion_560_4	TBA	TBA
Poales	Restionaceae	Elegia	U61219 (Thamnochortus)	Thamnochortus	N/A	N/A
Poales	Sparganiaceae	Sparganium	TBA	Sparganium_latifolium_3786	TBA	TBA
Poales	Thurniaceae	Thurnia	N/A	N/A	N/A	N/A
Poales	Typhaceae	Typha	TBA	Typha_minima_6415	TBA	TBA
Poales	Xyridaceae	Abolboda	N/A	N/A	N/A	N/A
Poales	Xyridaceae	Xyris	TBA	Xyris_154	TBA	TBA
Proteales	Nelumbonaceae	Nelumbo	AF190097	Nelumbo	N/A	N/A
Proteales	Platanaceae	Platanus	AY396713	Platanus	N/A	N/A
Ranunculales	Ranunculaceae	Aquilegia	AF190067	Aquilegia	N/A	N/A
Ranunculales	Eupteleaceae	Euptelea	AY396708	Euptelea	N/A	N/A
Sabiales	Sabiaceae	Meliosma	AY396712	Meliosma	N/A	N/A
Sabiales	Sabiaceae	Sabia	AY396714	Sabia	N/A	N/A
Trochodendrales	Trochodendraceae	Tetracentron	AF276749	Tetracentron	N/A	N/A
Trochodendrales	Trochodendraceae	Trochodendron	AF190109	Trochodendron	N/A	N/A
Buxales	Buxaceae	Buxus	AY396706	Buxus	N/A	N/A
Buxales	Buxaceae	Pachysandra	AF276735	Pachysandra	N/A	N/A

B.

Order	Family	Chase taxon	atpA/1	rbcL	matK	ndhF	atpB	18S	26S
Amborellales	Amborellaceae	Amborella	AY009407	L12628	AF543721	AF235046	AF235041	U42497	AY095449
Austrobaileyales	Austrobaileyaceae	Austrobaileya	AY299723	L12632	DQ401347	AF238052	AJ235403	AF206858	AY292886
Austrobaileyales	Schisandraceae	Illicium	AY299786	L12652	AF543738	AF123808	U86385.2	L75832	EU161362
Austrobaileyales	Schisandraceae	Schisandra	AF197662	L12665	AY326509	AF238062	AJ235599	L75842	TBA
Cannellales	Winteraceae	Drimys	AY299761	AF093734	AJ581398 (Belliolum)	AF123806	AF093425	U42823	AF036491
Ceratophyllales	Ceratophyllaceae	Ceratophyllum	AY299743	D89473	AJ581400	AF130232	AJ235430.2	U42517	AY095456
Chloranthales	Chloranthaceae	Ascarina	AF197667	AF238050	AJ966795	AF238051	AJ235593 (Sarcandra)	AF207012 (Sarcandra)	TBA
Chloranthales	Chloranthaceae	Chloranthus	AY299746	L12640	AJ966796 (Sarcandra)	AF238053	AJ235431.2	AF206885	AF479245
Laurales	Calycanthaceae	Calycanthus	AY299739	AF022951 .2	AY525337	AF123802	AJ235422	U38318	AY095454
Magnoliales	Magnoliaceae	Liriodendron	AF197690	L12654	AF465298	AF123810	AJ235522	AF206954	AY292879
Magnoliales	Magnoliaceae	Magnolia	AY299800	AY298837	AB040152	AF238056	AJ235526	AF206956	AF479244
Nymphaeales	Cambombaceae	Cabomba	AF197641	M77027	AF092991 (Victoria)	AF123801	AF187058	AF096691	AF479239
Nymphaeales	Nymphaea	Nymphaea	AY299814	M77034	AY779190	AF188853	AJ235544	AF206973	AY292900
Piperales	Aristolochiaceae	Asarum	AF197671	L14290	AF465285 (Aristolochia)	AF123800	U86383	DQ472350	AY095450 (Aristolochia)
Piperales	Lactoridaceae	Lactoris	AF197710	L08763	N/A	AF123809	AJ235515	U42783	AY292898
Piperales	Saururaceae	Saururus	AY299833	L14294	AF465302	AF123811	AJ235596	U42805	AY095468
Acorales	Acoraceae	Acorus_cal	AF039256	M91625	AB040154	AY007647.2	AJ235381.2	TBA	TBA
Acorales	Acoraceae	Acorus_gram	AY299699	D28866	AB040155	AF546992	AF197616	AF197584	AF036490
Alismatales	Alismataceae	Alisma	AF197717	L08759	AB040179	AF546993	N/A	AF197585	TBA
Alismatales	Alismataceae	Sagitarria	AY299832	L08767	AB002580 (Hydrocleys)	AY007657.2	AF239788	TBA	TBA
Alismatales	Araceae	Gymnostachys	AF039244	M91629	AB040177	AY191196	AF168915	AF069200	TBA

Order	Family	Chase taxon	atpA/1	rbcL	matK	ndhF	atpB	18S	26S
Alismatales	Butomaceae	Butomus	AY299733	U80685	AY952416	AF546997	AY147593	TBA	TBA
Alismatales	Cymodoceaceae	Cymodocea	DQ859095	U80687	TBA (Amphibolis)	AY191197 (Halodule)	AF168887 (Aponogeton)	AF168826 (Aponogeton)	N/A
Alismatales	Hydrocharitaceae	Vallisneria	DQ859119	AF206832	AB002568 (Elodea)	N/A	AF209694	AF207050	TBA
Alismatales	Juncaginaceae	Triglochin	AY299852	U80714	AM920647	AF546998	AF197601	AF197586	TBA
Petrosaviales	Petrosaviaceae	Japonolirion	AY299790	AF206784	AB040161	AY191199	AF209608	AF206942	TBA
Petrosaviales	Petrosaviaceae	Petrosavia	AY299821	AF206806	AB040156	N/A	AF209649	AF206987	TBA
Alismatales	Potamogetonaceae	Potamogeton	AY299829	U03730	AB002581	N/A	AF197600	EF526336	N/A
Alismatales	Tofieldiaceae	Pleea	AY299827	AJ131774	AF465301	DQ008886	AJ235564	AF206995	AY095472
Alismatales	Tofieldiaceae	Tofieldia	AY299851	AJ286562	AM920648	AF547023	AJ235627.2	AF207043	TBA
Alismatales	Zosteraceae	Zostera	DQ859121	U03724	AB125356	AF547022	AF209700	AF207058	TBA
Asparagales	Agapanthaceae	Agapanthus	AY299701	Z69221	AB017306	TBA	AJ417568	AF168851 (Hippeastrum)	(Hippeastrum)
Asparagales	Agavaceae	Agave	AY299703	Z69227 (Polianthes)	TBA	AF508398	AF209521	AF206841	TBA
Asparagales	Alliaceae	Allium	AY299707	AF206731	AB017307	AF547000	AF209525	AF168825	TBA
Asparagales	Amaryllidaceae	Clivia	AY299749	AF116950	AB017278	AY225031	AF209566	AF206889	TBA
Asparagales	Agavaceae	Anemarrhena	AY299711	Z77251	TBA	AY191162	AJ417570	TBA	TBA
Asparagales	Agavaceae	Chlorophytum	DQ859074	L05031	AB020806	AY191163	AF168894	U42066	TBA
Asparagales	Aphyllanthaceae	Aphyllanthes	AY299714	Z77259	TBA	AY191167	TBA	TBA	TBA
Asparagales	Asparagaceae	Asparagus	AY299720	L05028	AB029804	AF508403	TBA	AF069205	TBA
Asparagales	Asteliaceae	Astelia	AY299722	AF307906	AY368372	AY191164	TBA	AF206963 (Milligania)	(Milligania)
Asparagales	Blandfordiaceae	Blandfordia	AY299727	Z73694	AB017315	AY191169	AJ235412	AF206869	TBA
Asparagales	Boryaceae	Alania	AY299705	Y14982	N/A	AY191170	N/A	N/A	N/A
Asparagales	Boryaceae	Borya	AY299728	Y14985	AY368373	AY225059	AF209543	AF206872	TBA
Asparagales	Ruscaceae	Convallaria	AY299752	AB089627	AB029771	AF508404	AF168897	AF168834	TBA
Asparagales	Doryanthaceae	Doryanthes	AY299760	Z73697	AJ580616	AY225060	AY465543	TBA	TBA

Order	Family	Chase taxon	atpA/1	rbcL	matK	ndhF	atpB	18S	26S
Asparagales	Hemerocallidaceae	Hemerocallis	AY299780	FJ707502	TBA	AY147780	AF168923	TBA	N/A
Asparagales	Hyacinthaceae	Scilla	AY299836	L05038 (Ledebouria)	TBA	AF508397	AF168925 (Hyacinthus)	AF069206 (Ledebouria)	TBA
Asparagales	Hypoxidaceae	Hypoxis	AY299784	Y14989	AY368375	AY191179	AJ235582.2 (Rhodohypoxis)	AF207008 (Rhodohypoxis)	(Rhodohypoxis)
Asparagales	Iridaceae	Sisyrinchium	AY299837	Z77290	AJ579982	AF547008	AF209592 (Gladiolus)	L54062 (Gladiolus)	(Gladiolus)
Asparagales	Ixiolirionaceae	Ixiolirion	AY299789	Z73704	AJ579965	AY147781	TBA	AF206940	TBA
Asparagales	Lanariaceae	Lanaria	AY299796	Z77313	TBA	AY191183	AJ417592	TBA	TBA
Asparagales	Laxmanniaceae	Arthropodium	AY299719	Z69233	TBA	AY191184	TBA (Sowerbaea)	TBA	TBA
Asparagales	Orchidaceae	Cypripedium	AY299755	AF074142	TBA	AY225063	AJ235448.2	TBA	TBA
Asparagales	Orchidaceae	Epipactis	AY299766	Z73707	AF263659	AY225064	AJ235548.2 (Oncidium)	U42791 (Oncidium)	TBA
Asparagales	Orchidaceae	Neuwiedia	AY299813	AF074200	TBA (Apostasia)	U20633	TBA (Apostasia)	TBA (Apostasia)	(Apostasia?)
Asparagales	Tecophilaeaceae	Tecophilaea	AY299848	Y17337	TBA	AY191193	AJ235620.2	AF168836 (Cyanella)	(Cyanaella)
Asparagales	Xanthorrhoeaceae	Xanthorrhoea	AF039250	Z73710	TBA	AY147785	AF168952	U42064	TBA
Asparagales	Xeronemataceae	Xeronema	AY299857	Z69235	TBA	AY191194	AJ235647.2	AF207056	TBA
Dioscoreales	Burmanniaceae	Burmannia	AY299732	AF206742	AY956483	N/A	AF209548	TBA	TBA
Dioscoreales	Dioscoreaceae	Trichopsus	AY299724	AY298818	TBA	AF546996	AF308019 (Avetra)	AF309395 (Avetra)	N/A
Dioscoreales	Dioscoreaceae	Dioscorea	AY299759	AJ235803	AB040208	AY007652.2	TBA	AF206903	TBA
Dioscoreales	Dioscoreaceae	Tacca	AY299845	AJ235810	TBA	AY191200	AF308025	TBA	TBA
Dioscoreales	Nartheciaceae	Aletris	AY299706	TBA	TBA	AY191201	AF308040	TBA	TBA
Dioscoreales	Nartheciaceae	Narthecium	AY299809	AJ286560	AB040162	AY191202	AF308042	TBA	TBA
Liliales	Alstroemeriaceae	Alstroemeria	AF039254	Z77254	AY624481	AF276011	(Bomarea)	AF206871 (Bomarea)	(Bomarea)
Liliales	Campynemataceae	Campynema	AY299740	Z77264	TBA	AY224997	AJ417573	N/A	TBA
Liliales	Colchicaceae	Petermannia	AY299820	AY298844	TBA	AY225001	N/A	N/A	N/A

Order	Family	Chase taxon	atpA/1	rbcL	matK	ndhF	atpB	18S	26S
Liliales	Colchicaceae	Uvularia	TBA	Z77315	AY624482	AF276023	AJ417574 (Iphigenia)	N/A	TBA
Liliales	Corsiaceae	Arachnitis	AY299715	N/A	N/A	N/A	N/A	TBA	AF364030
Liliales	Liliaceae	Calochortus	AY299737	Z77263	TBA	AF275994	TBA	TBA	TBA
Liliales	Liliaceae	Lilium	AY299797	L12682	TBA	AY007655	AF209618	AF206952	TBA
Liliales	Luzuriagaceae	Luzuriaga	AY299798	Z77300	TBA	AY225005	AY465548	AF233091	N/A
Liliales	Melanthiaceae	Trillium	AF039253	D28164	AB07392	AY191205	AF209692	AF207048	TBA
Liliales	Melanthiaceae	Veratrum	AF039255	D28168	AB017417	AF276024	TBA	AF207057 (Xerophyllum)	(Xerophyllum)
Liliales	Philesiaceae	Philesia	AY299822	Z77302	AY624479	AF276014	AY465551	TBA	TBA
Liliales	Rhipogonaceae	Rhipogonum	AY299831	Z77309	TBA	AF276016	AY465553	TBA	TBA
Liliales	Smilacaceae	Smilax	AF039251	Z77310	AB040204	AF276018	AF209677	AF207022	TBA
Pandanales	Cyclanthaceae	Chorigyne	AY299747	AY298823	N/A	N/A	N/A	N/A	N/A
Pandanales	Cyclanthaceae	Cyclanthus	AY299754	AY007660	TBA	AY224992	AF168904	AF168837	TBA
Pandanales	Cyclanthaceae	Sphaeradenia	AY299840	AJ235808	N/A	N/A	AJ235607.2	AF207024	TBA
Pandanales	Pandanaceae	Freycinetia	AY299770	AF206770	AB040209	N/A	AF209590	AF206915	TBA
Pandanales	Pandanaceae	Pandanus	AY299818	M91632	TBA	AY191203	AF308043	AY952391	N/A
Pandanales	Stemonaceae	Stemona	AY299842	AJ131948	TBA	AF547009	AF308037	AF207028	TBA
Pandanales	Triuridaceae	Sciaphila	AY299835	N/A	N/A	N/A	N/A	TBA	N/A
Pandanales	Velloziaceae	Acanthochlamys	AY299698	TBA	TBA	AY224993	TBA	AY952411	N/A
Pandanales	Velloziaceae	Talbotia	AF039247 (Vellozia)	AJ131946 (Barbacenia)	TBA (Vellozia)	AF546999 (Vellozia)	TBA (Barbacenia)	AF206861 (Barbacenia)	(Barbacenia)
Arecales	Arecaceae	Calamus	AY299734	AJ404775	TBA	AY044523	AF233081	AF168828	TBA
Arecales	Arecaceae	Euterpe	AY299769	AY298832	TBA	AY044535 (Areca)	AY044460 (Geonoma)	AF168831 (Caryota)	(Caryota)
Arecales	Arecaceae	Nypa	U58833	M81813	AM114552	AY044525	AY012414	AF168854 (Iriartea)	(Iriartea)
Dasypogonales	Dasypogonaceae	Calectasia	AY124505	AY123231	TBA	AY191208	AF168891	AF069209	AY079521
Dasypogonales	Dasypogonaceae	Dasypogon	AY124503	AY123229	TBA	AY191209	AF168907	AJ417898	TBA
Dasypogonales	Dasypogonaceae	Kingia	AY124506	AY123232	AM114718	AY465644	N/A	TBA	AF466385

Order	Family	Chase taxon	atpA/1	rbcL	matK	ndhF	atpB	18S	26S
Commelinales	Commelinaceae	Murdannia	AY299805	AY298838	TBA	AY624112 (Spatholirion)	AF168950 (Tradescantia)	AF168840 (Elasis)	N/A
Commelinales	Haemodoraceae	Anigozanthos	AF039246	AJ404843	AM114721	AF546994	TBA	TBA	TBA
Commelinales	Hanguanaceae	Hanguana	AY299775	AJ417896	AB088800	AY007654	AJ417579	AF387604	TBA
Commelinales	Philydraceae	Philydrum	AY299824	U41596.2	AF434870 (Philydrella)	U41622	AF209651 (Philydrella)	U42074 (Helmholtzia)	(Helmholtzia)
Zingiberales	Cannaceae	Canna	AY299741	AF378763	TBA	AY191214	AF168892	D29785	TBA
Zingiberales	Costaceae	Costus	AY299753	AY298826	TBA	AY191215	AF168899	U42080	TBA
Zingiberales	Heliconiaceae	Heliconia	AY299778	AF378765	TBA	AY656108	AF168917	U42082	TBA
Zingiberales	Lowiaceae	Orchidantha	AY299815	AF243841	TBA	AY191217	AF168933	AF168865	TBA
Zingiberales	Marantaceae	Maranta	AY299801	AF378768	TBA	AY191218	AF168927	U42079	TBA
Zingiberales	Musaceae	Musa	AY299806	AF378770	AJ581437	AY191219	AF168930	U42083	TBA
Zingiberales	Strelitziaceae	Strelitzia	AY299843	AF243846	TBA	AY191220	AF168948	AF069229	TBA
Poales	Anarthriaceae	Anarthria	AY124513	AF148760	DQ257499	N/A	AJ419129	TBA	TBA
Poales	Bromeliaceae	Puya	AY124508	L19973	EU780853	L75903	AF209661	AF069212 (Aechmea)	(Aechmea)
Poales	Bromeliaceae	Tillandsia	AY124507	L19971	AY614080	L75899	TBA	AF168847 (Glomeropitcairnia)	N/A
Poales	Centrolepidaceae	Aphelia	N/A	AY123233	DQ257500	EF153942	AJ419131	N/A	TBA
Poales	Cyperaceae	Carex	AY124514	Y12998	TBA	AF163455	AF168906 (Cyperus)	AF168838 (Cyperus)	TBA
Poales	Cyperaceae	Mapania	N/A	Y12955	TBA	AY129256	AF209667 (Rhyncospora)	AF207009 (Rhyncospora)	(Rhyncosperma)
Poales	Ecdeiocoleaceae	Ecdeiocolea	AY124516	AY123235	DQ257530	AY622313	AJ419136	TBA	TBA
Poales	Eriocaulaceae	Eriocaulon	AY124517	AY123236	TBA	AF547017	TBA	TBA	TBA
Poales	Flagellariaceae	Flagellaria	AF039248	L12678	AB040214	U22008	AF209589	AF206913	TBA
Nymphaeales	Hydatellaceae	Trithuria	N/A	DQ915188	N/A	AF547020	N/A	N/A	TBA
Poales	Joinvilleaceae	Joinvillea	AY124519	L01471	AF164380	U21973	AJ419143	AF168855	TBA
Poales	Juncaceae	Juncus	AY124520	L12681	TBA	AF547015	AJ235509.2	AF206944	TBA
Poales	Juncaceae	Luzula	AY124521	AJ419945	TBA	N/A	AJ419145	N/A	TBA

Order	Family	Chase taxon	atpA/1	rbcL	matK	ndhF	atpB	18S	26S
Poales	Poaceae	Oryza	X51422	D00207	AF148650	X15901	D00432	X00755	M11585
Poales	Thurniaceae	Prionium	AY124527	U49223	TBA	AF547019	AJ419153	N/A	TBA
Poales	Rapateaceae	Rapatea	AY124511	L19972 (Stegolepis)	TBA (Stegolepis)	AF207623	AJ419150	N/A	N/A
Poales	Restionaceae	Baloskion	AY124529	AF148761	DQ257501	AF251444	AF209666 (Restio)	AF207006 (Restio)	(Restio)
Poales	Restionaceae	Elegia	AY124530	AY123238	TBA	AF547016	AJ419151	AF069219	TBA
Poales	Sparganiaceae	Sparganium	AY124509	M91633	TBA	AY191213	AF209678	AF069220	TBA
Poales	Thurniaceae	Thurnia	AY124532	AY123239	TBA	AY208986	AJ419154	N/A	N/A
Poales	Xyridaceae	Abolboda	AY124533	AY123240	TBA	AY438616 (Orectanthe)	N/A	AF168824	TBA
Poales	Xyridaceae	Xyris	AY299859	AF206834	TBA	AF547021	AY465541	AF168881	TBA
Proteales	Nelumbonaceae	Nelumbo	AF197654	FJ626615	AM396514	EU642680	EU642740	L75835	FJ626483
Proteales	Platanaceae	Platanus	AF197655	L01943	AM396503	NC_008335	NC_008335	U42794	AF274662
Ranunculales	Ranunculaceae	Aquilegia	AY394727	FJ449851	EF437128	AF130233	EU053875	X63300	FJ626439
Ranunculales	Eupteleaceae	Euptelea	AF197650	AY048174	AM396510	AY394737	AF528850	L75831	AF389249
Sabiales	Sabiaceae	Meliosma	AF197656	AF197587	AM396513	AY394741	AF209626	AF206961	AF389271
Sabiales	Sabiaceae	Sabia	AF197657	FJ626616	AM396512	AJ236276	AF093395	L75840	AF389272
Trochodendrales	Trochodendraceae	Tetracentron	AF197647	L12668	AM396504	N/A	AF093422	AF094564	AF274670
Trochodendrales	Trochodendraceae	Trochodendron	AF197648	L01958	AF543751	EU002269	EU002169	AF094565	AF479205
Buxales	Buxaceae	Buxus	AF197636	NC_009599	NC_009599	NC_009599	NC_009599	L54065	AF389243
Buxales	Buxaceae	Pachysandra	AF197634	AF093718	AF542581	AF241601	AF528854	AF094533	AF389244

Table 2. *PHYC* primers used in this study

General *PHYC* primers [16]

c230f	5’	GAY TTR GAR CCW GTD AAY C
c623r	5’	GRA TKG CAT CCA TYT CMA YRT C

Asparagales

Asp_PhyC_1F	5’	GAG CCW GTT AAC CCW GCY GAT GTA CC
Asp_PhyC_1R	5’	GMA TCC ATY TCS AYR TCT TCC CA

Commelinales

Comm_phyC_P1F	5’	GAT GTY YTG GTT CGS GAR GTK AGY GAG C
Comm_phyC_P2F	5’	GAG CCT GTK AAC CCY RCC GAT G
Comm_PhyC_P1R	5’	ATC CAT TTC RAY RTC TTC CCA RGG

Dioscoreales

Diosc_PhyC_P1F	5’	CCW GCY GAT GTG CCA GTR ACW GCT GC
Diosc_PhyC_P1R	5’	TCC CAS GGA AWA CTY CTK YGC TTW ACC AC

Pandanales

Pand_phyC_P2F:	5’	GCC GAY GTV CCM GTS ASM GCY GCY GG
Pand_phyC_P2R:	5’	GGA AGR CTY CTT CGC TTC ACC AC

Poales

Poal_PhyC_P1F	5’	GAY TTR GAG CCW GTK AAY CC
Poal_PhyC_P1R	5’	GRA TGG MAT CCA TYT CVA YRT CYT CCC A

Pandanales

Pand_phyC_P2F:	5’	GCC GAY GTV CCM GTS ASM GCY GCY GG
Pand_phyC_P2R:	5’	GGA AGR CTY CTT CGC TTC ACC AC

Table 3. Fossils utilized for calibration of divergence times.

Node label refers to assignation on Figure 5. Constrained nodes relate to the lineage for which each minimum date is assigned (SL=stem lineage, CG=crown group). MRCA indicates the node placement for this study. Stratigraphic positions (stage) of fossils for calibrations were transformed to absolute ages using the upper (younger) bound of the interval based on the current stratigraphic timescale [38].

Node label	Constrained node	MRCA	Fossil taxon, basis for identification (reference)	Stage	Age (Ma)
1	CG Nymphaeales	*Nymphaea, Trithuria*	small peryginous flower [64]	Late Aptian-Early Albian	112
2	SL Schisandraceae	*Illicium, Austrobailey a*	seeds with epidermal cells with anticlinal undúlate walls [31]	Late Barremian-Early Aptian	125
3	CG Chloranthaceae	*Chloranthus, Ascarina*	*Clavatipollenites* and *Asteropollis* pollen and flowers [31]	Late Barremian-Early Aptian	125
4	SL Magnoliales	*Magnolia, Calycanthus*	*Endressinia brasiliana*; branching axis, leaves, flowers [65]	Late Aptian-Early Albian	112
5	SL Winteraceae	*Drimys, Asarum*	*Walkeripollis gabonensis*; pollen [66,67]	Late Barremian-Early Aptian	125
6	SL Lactoridaceae	*Asarum, Lactoris*	*Lactoripollenites africanus*; pollen [68]	Turonian-Campanian	89.3
7	CG eudicots	*Buxus, Euptelea*	tricolpate pollen grains [69,70];	Late Barremian-Early Aptian	125 (fixed)
8	SL Araceae	*Orontium, Alisma*	*Mayoa portugallica*; pollen [71]	Late Barremian-Early Aptian	125
9	CG Pandanales	*Stemona, Acanthochla mys*	Triuridaceae, *Mabelia, Nuhliantha*; flowers, pollen [72]	Turonian	89.3
10	SL Arecales	*Nypa, Kingia*	*Sabalites carolinensis*; pollen, leaves [32]	Coniacian-Santonian	85.8
11	SL Zingiberales	*Heliconia, Murdannia*	*Spirematospermum chandlerae*; fruits [33]	Santonian-Campanian	83.5
12	SL Poaceae	*Ecdeiocolea , Oryza*	phytoliths [35]	Maastrichtian-Campanian	70.3

Table 4. Results of divergence time estimates from different analyses.

Janssen is NPRS estimates from Janssen [7]; And PL/Pd are penalized likelihood/PATHd8, respectively from Anderson [8], with only one value shown when both are identical; Mag=constrained ages from Magallon [10]. PL160 are results from this study, with the range indicating alternative values from setting the maxage of the CG angiosperms at 180 Ma (lower bound) and 140 Ma (upper bound). SL=stem lineage, CG=crown group. An asterisk (*) indicates the tree root with fixed age. N/A indicates the date for that node was not reported or was not estimated because of taxonomic sampling. All units are Ma.

Lineage	Janssen	And PL/Pd	Mag	PL160	
				age	range
SL monocots	N/A	N/A	N/A	152	136-169
CG monocots/ SL Acorales	134*	134*	127	147	132-163
CG Acorales	N/A	N/A	N/A	11	10-12
SL Alismatales	131	131/124	126	143	130-159
CG Alismatales	128	128/123	125	135	126-149
SL Petrosaviales	126	126/107	123	138	125-152
CG Petrosaviales	123	N/A	N/A	78	70-80
SL Dioscoreales/ SL Pandanales	124	124/104	119	133	122-147
CG Dioscoreales	123	123/101	115	128	117-141
CG Pandanales	114	109/90	102	108	98-120
SL Liliales	124	124/104	120	135	123-149
CG Liliales	117	116/98	114	129	117-142
SL Asparagales	122	122/102	118	134	130-148
CG Asparagales	119	118/70	112	127	116-141
SL Commelinids	122	122/102	N/A	134	122-148
CG Commelinids	120	120/100	N/A	130	119-143
SL Arecales	120	120/100	114	123	112-136
CG Arecales	110	N/A	N/A	49	44-56
SL Dasypogonaceae	119	118/98	114	123	112-136
CG Dasypogonaceae	100	N/A	N/A	50	45-56
SL Poales	117	116/98	111	129	118-142
CG Poales	113	112/97	99	123	113-134
SL Zingiberales/ SL Commelinales	114	114/101	99	113	104-125
CG Zingiberales	88	88/36	79	69	62-77
CG Commelinales	110	107/97	N/A	104	95-115

Table 5. Whole-tree tests for shifts in diversification rate from SymmeTREE

[45]. Δ_1 and Δ_2 are two different calculations for likelihood ratio-based shift statistics. An asterisk (*) indicates a p-value of statistical significance; † indicates a p-value of marginal significance. All taxonomic clades listed are for terminal branches except for Joinvilleaceae/Ecdeiocoleaceae/Poaceae (internal Poales branch) and Hanguanaceae/Commelinaceae (internal Commelinales branch).

Clade	**Δ_1 p-value**	**Δ_2 p-value**
Commelinaceae (Commelinales)	0.06*	0.08*
Hanguanaceae/Commelinaceae (Commelinales)	0.01†	0.02†
Herreria (Agavaceae)	0.09*	0.1*
Agave (Agavaceae)	0.04†	0.06*
Eriocaulaceae (Poales)	0.06*	0.07*
Joinvilleaceae/Ecdeiocoleaceae/ Poaceae (Poales)	0.06*	0.08*

CHAPTER 3

SYSTEMATICS AND EVOLUTION OF LIFE HISTORY TRAITS AND GENOME SIZE IN THE TRADESCANTIA ALLIANCE (COMMELINACEAE)

Abstract

The *Tradescantia* alliance (subtribes Tradescantiinae and Thyrsantheminae of tribe Tradescantieae) comprises a group of closely related New World genera exhibiting considerable variation in life history and genomic traits. While historically difficult to circumscribe taxonomically, the degree of variation represents an opportunity to explore character evolution and correlations. We constructed a molecular phylogeny for the eighty five taxa in Commelinaceae, with sampling focused in the *Tradescantia* alliance, and found all but one currently defined genus (*Tinantia*) to be polyphyletic. *Tradescantia* and *Gibasis* are strongly supported as a single clade, as are *Callisia* and *Tripogandra*. Inflorescence morphology, an important character for generic identification, is revealed as labile and complex across the phylogeny. We used this phylogenetic framework to parsimoniously evaluate trait evolution of five life history traits (life history schedule, breeding system, Raunkiaer growth form, growth habit, and biogeography) and genome size evolution across the alliance. We tested for correlations between genome size and each life history trait using independent contrasts but found no significant relationships. We discuss limitations of this dataset for implementation of comparative biology methods.

Introduction

The Tradescantia alliance is a group of eleven genera comprising New World subtribes Tradescantiniae and Thyrsantheminiae of tribe Tradescantieae in the monocot family Commelinaceae (dayflower or spiderwort family). These genera (Tradescantia, Gibasis, Callisia, Tripogandra, Elasis, Tinantia, Thyrsanthemum, Weldenia, Gibasoides, Matudanthus, Sauvallea) maintain variable levels of genome change, including polyploidy, aneuploidy, hybridization, and genomic rearrangements. Commelinaceae is second only to grasses in respect to the number of weedy and polyploid species [1]. Despite the widespread significance of such species ecologically and cytogenetically, many outstanding questions remain in relation to the evolutionary framework of the Tradescantia alliance.

Systematics in Commelinaceae were historically problematic for several reasons. First, flowers in this group are short-lived and deliquescent; herbarium specimens rarely preserve important floral characteristics. Second, morphological characters are confusing and seem to have arisen via convergent evolution [2]. Floral characteristics are similar for several of the genera (Figure 1), and interpretation of inflorescence characterisistics varies greatly between researchers. A thorough discussion of difficulties in assigning morphological states to taxa in Commelinaceae can be found in Evans et. al [3]. Third, interspecific hybridization may have played a role in the evolution of the group, confounding efforts to resolve interspecific relationships [4]. As a result, many current genera are the result of dissolving, resurrecting, or recombining historic genera in the group. Species have been shuffled between many genera, and the discovery of new species and genera is ongoing.

Clarke [5] initially proposed a classification of sections for genus Tradescantia which, although it did not include the full complement of species now included in the genus, was gradually dismembered and reorganized by subsequent researchers. Tripogandra [6], Gibasoides, Matudanthus and Elasis [7] were each removed from Clarke's Tradescantia and given generic status. Clarke [5] also first described the genus Tinantia, at least one species of which had been previously described as Tradescantia [8]. One species, Tinantia anomala, was later transferred to a new genus, Commelinantia, because of morphological characters reminiscent of Commelina [9,10]. Subsequent researchers, however, rejected this analysis and instead grouped it with Tinantia [e. g.,11].

In his description of Mexican Commelinaceae, Hunt [12] favored the inclusion of several minor genera into larger, broader genera: Gibasis (including Aneilema sensu Matuda, in part), Tradescantia (including Campelia, Cymbispatha, Rhoeo, Separotheca, Setcreasea, Zebrina), Callisia (including Aploleia, Cuthbertia, Hadrodemas, Leptorrhoeo, Phyodina, Spironema) and Tripogandra (including Neodonellia). The current and most acceptable Commelinaceae classification divides tribe Tradescantieae into seven subtribes: three from the Old World and four from the New World. This system places Thyrsanthemum, Gibasoides, Tinantia, Elasis, Matudanthus, and Weldenia into subtribe Thyrsanthiminae; Gibasis, Tradescantia, Callisia and Tripogandra are placed in Tradescantiinae. One genus, Sauvallea, is an enigmatic genus from Cuba thought to belong in either of the two previously mentioned subfamilies [13]. Previous studies had also placed Tinantia and/or Thyrsanthemum in different groups [e.g., 11].

Phylogenetic analysis of morphological characters across Commelinaceae suggest a great deal of homoplasy in most characters previously used to classify groups [3]. The first molecular phylogeny of the family suggested that tribe Tradescantieae is monophyletic with the exception of Palisota. As sampling was limited to one species per genus, however, further exploration of the relationships among genera is needed [14]. A more recent phylogeny including comprehensive sampling of genera in tribe Tradescantieae exploited morphological and molecular data, and is the basis for sampling in the present study. It revealed a more derived New World clade composed of Tradescantia, Gibasis, Callisia, Tripogandra, Elasis, Tinantia, Thyrsanthemum, and Weldenia [Figure 2, 15]. A study examining invasiveness in a phylogenetic context focused sampling on taxa relevant to invasion biology. A combined analysis of a cpDNA locus (trnL-F) and a multiple copy nuclear locus (5S NTS) presents Tradescantia and Gibasis as monophyletic with Callisia is paraphyletic [16]. One final phylogenetic study focused sampling on Callisia; two cpDNA loci resolved a polyphyletic Callisia from the inclusion of Tripogandra as well as a monophyletic Tradescantia sister to the clade containing Callisia and Tripogandra clade [17].

While the systematic history of the Tradescantia alliance is complex, it provides ample opportunity to explore mode of character evolution over time. Additionally, plant groups with diverse life history and genomic traits are optimal systems in which to test hypotheses about relationships between genomic and organismal characteristics. Research in Veronica, for example, explored the relationship between genome size and life history, and found genome sizes of annuals had a lower upper limit than genome sizes of perennials [18]. A study of Mexican Commelinaceae species also suggested that specialized plants

(geophytes and hemicryptophytes) have larger genome sizes than plants living in unspecialized habitats. Furthermore, genome size in these species increased with latitude of native regions [19]. While the former study utilized a phylogenetic framework, the latter did not; a robust phylogeny of the Tradescantia alliance provides the context necessary to test each of these hypotheses while taking possible phylogenetic bias into account [20].

Given the and complicated nature of evolution and hypothesized hybridization in the Tradescantia alliance, a phylogeny utilizing only chloroplast loci can provide a simplified version of just the matrilineal relationships in the group. While it is clear which genera belong in the Tradescantia alliance, relationships among these genera remain confusing and unclear. Disagreement about generic and subtribal boundaries necessitates a more thorough examination of Commelinaceae phylogenetics with more data and thorough sampling. The questions addressed by this research are twofold. First, are subtribes and genera monophyletic? The current classifications of family Commelinaceae [13] and each of the genera [6,7,21,22,23] serve as hypotheses of phylogeny in this group. Second, are there correlations between genome size and life history traits in the Tradescantia alliance? These taxa provide a prime opportunity to test previously hypothesized relationships between genome size and life history schedule, breeding system, Raunkiaer growth form, growth habit, and biogeography.

Materials and Methods

Taxon selection

Sampling in our study includes eighty five taxa obtained from field collections,

botanical gardens, and commercial sources, as well as sequences previously published in GenBank (Table 1). When possible, living specimens were maintained in greenhouses at the University of Missouri for DNA extraction and trait analysis. Herbarium specimens have been deposited in the University of Missouri Dunn-Palmer Herbarium (UMO). The ingroup includes 58 taxa from eight genera, including 29 *Tradescantia* (ca. 70 species total in genus), nine *Gibasis* (11 spp.), 16 *Callisia* (ca. 20 spp.), five *Tripogandra* (ca. 22 spp.), one Thyrsanthemum (3 spp.), six *Tinantia* (14 spp.) and monotypic *Elasis* and *Weldenia*. Obtaining monotypic genera *Sauvallea, Gibasoides,* and *Matudanthus* was not possible for this study. Outgroup taxa were selected from other subtribes in tribe Tradescantieae [11 taxa, 15] and superoutgroups are represented by five taxa from tribe Commelinae [13]. Taxonomic assignments for this study follow the most current systematic treatments for particular groups available [6,7,21,22,23].

Molecular methods

DNA extraction necessitated a 3X-6X CTAB method [24] from fresh or frozen leaf tissue. We amplified two plastid loci generally following PCR parameters in Shaw et. al [25] with minor alterations in $MgCl_2$ concentrations for recalcitrant taxa. Conserved primers [F71, R1516,25] amplified the rpL16 intron and two additional internal primers assisted in sequencing (rpL16F692 ATGGAGAAGCTGTGGGAACGA, rpL16R690 CGTTCCCACAGCTTCTCCATTA). Conserved primers TabC and TabF amplified the trnL intron/trnL-trn-F intergenic spacer with additional sequencing via internal primers TabD and TabE [26]. The University of Missouri's DNA Core directly sequenced purified products.

Sequence alignment and phylogenetic analysis

We edited resulting sequences using DNAStar's Lasergene program suite [27] with manual curation and aligned each locus using MUSCLE [28,29]. We constructed all phylogenetic inferences using RAxML v7.2.8 [30] implemented on-line in RAxML BlackBox [31]. We partitioned the analysis into two loci (rpL16 and trnL-trnF and implemented a GTR+GAMMA model of molecular evolution for each partition. We used several methods to evaluate confidence intervals and explore alternative hypotheses in our resulting phylogeny. First, we obtained 100 bootstrap replicates in RAxML. Second, we conducted constraint tests to evaluate support for monophyly of subtribes (Tradescantiinae: *Tradescantia*, *Gibasis*, *Callisia*, *Tripogandra;* Thyrsantheminae: *Elasis*, *Thyrsanthemum, Tinantia*) and individual genera (*Tradescantia*, *Gibasis*, *Callisia*). Constraint trees were inferred using the same parameters as the unconstrained trees. We compared constraint trees using several topology-based tests implemented in CONSEL [32].

Genome size data

The Benaroya Research Institute at Virginia Mason in Seattle, Washington obtained genome size estimates using a flow cytometry protocol modified from Arumuganathan and Earle [33,34]. Additional accessions from similar collections are substituted for some taxa. If we were unable to obtain fresh leaf tissue for flow cytometry, we used values reported in the Plant DNA C-values Database [35]. When a range of values were available for a single taxon, we selected a median value for representation. Genome size is reported as pg/1C, or mass of DNA per haploid cell (Table 1).

Life history traits

We collected information regarding life history traits for taxa using both the literature and notes from our greenhouse collections. Our dataset included five discrete character traits: life history schedule, breeding system, Raunkiaer growth forms, growth habit, and biogeography. Reconciliation of multi-state taxa were guided by ancestral reconstructions (see Character Evolution below and Results).

Life history schedule. Plants were scored as perennial or annual based on growth in the native range in the wild from published species descriptons; “annuals or short lived perennials” were classified as annuals.

Breeding system. While there is a close connection between annuality and self compatibility, these characters varied independently in our dataset and are tested separately. Self compatibility (SC) and incompatibility (SI) largely followed Owens [36] and were scored as SC when accessions exhibiting both syndromes were reported in the literature or observed in the greenhouse (seed set from plants in the absence of pollinators or unrelated accessions).

Raunkiaer growth forms. We categorized plant growth life forms using an updated Raunkiaer system [37] by building upon Martinez's [19] dataset. According to this system, annual plants are therophytes. Assignments to perennials depended on the amount of growth during unfavorable (dry, cold) seasons. Geophytes include plants that persist as underground bulbs or rhizomes, hemicryptophytes persist just at ground level, and chamaephytes are herbaceous growth persisting above ground in unfavorable seasons.

Growth habit. Growth forms and growth systems are not completely independent characters, but represent two different strategies to describe the diversity in life form of the *Tradescantia* alliance. As Raunkiaer's system does not fully encompass the variation of life history traits in the *Tradescantia* alliance, we also assigned taxa to categories based on growth habit. Species growing with overlapping leaves reminiscent of bromeliads are labeled as rosettes. Plants that spread via trailing stems that root at the nodes are called creeping. Trailing or low-growing plants that do not (or rarely) root at the nodes are decumbent; erect plants are those which do not root at the nodes but stand upright and higher from the ground on longer stems.

Biogeography

Finally, taxa were assigned to a biogeographic categories, with priority given to Old World or more southern ranges when applicable: Old World (Africa, Asia), South America, Mesoamerica/Central America (including southern Mexico), Mexico (central, northern, eastern, western), and/or North America (United States).

Character evolution

We evaluated each life history trait by tracing character history on the ML tree using a parsimony criterion in Mesquite v2.74 [38]. The resulting tree graphically represents the evolution of each character across the tree and estimates the ancestral state of the the character at each node. Polarization of traits estimated using ancestral character states provided the context for correlational analyses. We explored correlations between genome size (a continuous trait) and life history traits (discrete traits) using PDAP v1.07 [39] implemented in Mesquite. This package is appropriate for the analysis in question because

it accepts missing values in the character matrix and calculates correlations among continuous characters using Felsenstein's Independent Contrasts [FIC, 20]. Branch lengths of the ML tree transformed using the “branch length method of Nee” [38] allowed the dataset to pass the standard assumptions check for independent contrasts.

Results

Phylogenetic inference

A description of each data partition and the combined two locus dataset is available in Table 2. The best-scoring ML tree is well supported along the backbone (Figure 3); specific taxonomic groups are discussed below. Results from constraint tests are found in Table 3.

Tradescantia. Topology tests do not support *Tradescantia* as monophyletic (Table 3). *Tradescantia* species comprise a strongly supported clade with the inclusion of *Gibasis geniculata* and *G. linearis* (BS=100), as well as the sister taxon *G. oaxacana* (BS=100). There is little reinforcement for taxonomic classification within *Tradescantia*, as only weak bootstrap support exists for most internal nodes in the clade. No currently named sections emerge as monophyletic; sect. Tradescantia series Tradescantia (the “erect” *Tradescantia*) appears as monophyletic albeit with very weak bootstrap support (Figure 3).

Gibasis. As two species of *Gibasis* are nested within *Tradescantia*, and a third species is sister to *Tradescantia*, there is no support for this genus as monophyletic (Figure 3). Topology tests reinforce this interpretation, as the constrained tree is significantly different from the unconstrained test for most of the topology tests. The exception is the SH test

(p=0.179), but this test is known to have a relatively high error rate in some cases [40]. With the exception of the three taxa mentioned in association with *Tradescantia*, *Gibasis* forms a strongly supported monophyletic clade (BS=97), and also with its sister taxon, the monotypic genus *Elasis* (BS=92). The latter clade is sister to the *Tradescantia* clade. The *Gibasis* taxa grouping together are all from sect. *Gibasis*; the only member of this section not in the clade is *G. linearis*. The other two Gibasis species, *G. geniculata* and *G. oaxacana*, comprise sect. Heterobasis.

Callisia and *Tripogandra*. All *Callisia* taxa are in a strongly supported clade (BS=97) sister to *Gibasis* + *Tradescantia* (Figure 3). All *Tripogandra* species are nested within this clade (BS=99 with inclusion of *Callisia gracilis*); as with *Gibasis*, most topological constraint tests support a significantly different tree than the unconstrained tree (although SH=0.19, Table 3). There is substantial substructure within the *Callisia* clade, including support for several taxonomic sections. Section Cuthbertia (BS=100) and sect. Brachyphylla (BS=100, including previously unplaced *C. hintoniorum*) are sister to each other (BS=100) as the first *Callisia* lineage to diverge. Three taxa of sect. Leptocallisia are monophyletic (BS=100) and next to diverge (BS=97). The two remaining clades are also strongly supported as sister (BS=95). One clade is the afore mentioned *Tripogandra* + *C. gracilis*, the other is *C. warscewicziana* (sect. Hadrodemas) sister to sect. Callisia (BS=100). Section Callisia is strongly supported as monophyletic (BS=100), and comprised of three “groups” that, despite little morphological separation, are supported in the phylogeny (Figure 3).

Subtribes Tradescantiinae and Thyrsantheminae. Neither of the subtribes comprising the *Tradescantia* alliance were supported by topology tests (Table 3). Subtribe

Tradescantiinae is well supported with the inclusion of *Elasis* (BS=97). Subtribe Thyrsantheminae is a parapyletic grade, with moderate support along the backbone of the tree (Figure 3). The largest genus in this subtribe, *Tinantia*, is the only genus in the *Tradescantia* alliance supported by our phylogeny (BS=89).

Character evolution and biogeography

We obtained several genome size estimates for several previously unreported taxa. Ancestral state reconstructions from parsimony suggest that for all taxa sampled (including outgroups), the ancestral states for Commelinaceae were perennial, SC, chamaephyte/rosette habit and origin in the Old World or South America (Table 4). The most likely ancestral state for the *Tradescantia* alliance was similar except for an erect growth habit. The ancestral genome size range for both nodes was 4.5-8.6 pg/1C. There were several notable patterns in switches between character states across the whole tree (Figure 4). First, there were three origins of annuality from perennial plants; once for *Tinantia* and twice in *Callisia* + *Tripogandra* (data not shown). Second, there was one major switch from SC to SI near the divergence of the *Tradescantia* alliance, followed by several reversals to SC (data not shown). Third, all Raunkiaer growth forms arise from the ancestral chamaephyte state, and there are few reversals (data not shown). Fourth, biogeographic patterns suggest three introductions to North America, once each in *Tinantia*, *Callisia*, and *Tradescantia* (Figure 4). Movement between divisions in other New World delimitations occurs throughout the tree. Finally, there are at least four major expansions in genome sizes, twice in *Callisia*, once in *Gibasis*, and at least twice in *Tradescantia*; the transitions in Tradescantia are towards very large genome sizes. There are no clear patterns discernable

from the complex switches in growth habit (data not shown).

We detected no significant correlations between life history traits and genome size (Table 4).

Discussion

A molecular phylogeny of the *Tradescantia* alliance from two chloroplast loci resolves relationships between notoriously difficult genera. Resulting implications for circumscription of genera provide insight into interpretation of morphological characters and their lability over evolutionary time. Reconstructions of ancestral states for a variety of life history traits related to habit, breeding system, biogeography, and genome size indicate multiple transitions for any character throughout the phylogeny. While we did not detect any significant correlations between each life history trait and genome size, the composition of our dataset may have limited ability to analyze these trends.

Phylogenetic classification

The phylogenetic reconstruction from two chloroplast loci recapitulates the evolutionary relationships between genera posited by previous studies that were limited to one taxon per genus (Figure 2). Topological constraint tests provide information about the monophyly of genera and subtribes, which as a result inform understanding of morphological characters used to define taxonomic groups. The ingroup of the *Tradescantia* alliance is comprised of two closely related subtribes, Tradescantiinae and Thyrsanthemineae, which while strongly supported as single clade are both paraphyletic according to current classification. The polyphyly of subtribe Thyrsantheminae confirms

previous findings from phylogenies constructed from both morphological and molecular loci [3,14,15]. The main distinction between these subtribes is the structure of the inflorescence. Tradescantiinae, and nearly all genera within it, are characterized by bifacially fused cincinni, although exceptions in *Gibasis* are noted [13]. Our results indicate this morphological feature to be labile throughout the phylogeny. The inclusion of *Elasis* into subtribe Tradescantiinae is strongly supported in this analysis by at least two robust nodes in the backbone of the phylogeny. As a result, the single cincinni of *Elasis* represents a reduced form of the two bifacially fused cincinni characteristic of subtribe Tradescantiinae, confirming the hypothesis of Evans et. al [14].

Increased sampling indicates additional problems to generic delimitations from previous studies [16,17]. None of the currently circumscribed genera in subtribe Tradescantiinae are monophyletic. Burns Moriuchi [16] found *Gibasis* to be strongly monophyletic; however, all three species included in that analysis were from section Gibasis. Our results suggest *Tradscantia* and *Gibasis* intergrade substantially with each other. In contrast to previous molecular systematic studies [16,17], we confirmed monophyly of most sections in *Callisia* and resolved relationships between them. Morphological features also support the association of *Tripogandra* with sect. Callisia. *Tripogandra* is a relatively clearly marked genus characterized by dimorphic stamens with protrusions on three filaments [6]. While sect. Callisia does not display these protrusions, taxa in this group differ from many others in the *Tradescantia* alliance in that they possess dimorphic stamens [23].

This is the first study to include substantial sampling from *Tinantia*, which we reveal to be the only genus in the alliance supported as monophyletic. Floral zygomorphy and corresponding staminal characteristics make this a robustly delineated genus morphologically. The two most problematic taxa in *Tinantia*, *T. pringlei* and *T. anomala* [10], are sister to the other species. Remaining genera in subtribe Thrysantheminae are monotypic or only represented by one species. Of particular interest to systematics of the alliance are still unsampled monotypic genera *Gibasoides*, *Matudanthus*, and *Sauvallea*; their inclusion could potentially solidify placement of the other genera and circumscription of subtribes.

Character evolution and biogeography

We detected no discernable correlations between genome size and life history traits. For biogeography and genome size, however, a visual inspection of trait evolution suggests a relationship (Figure 4). Each of the introductions to North America coincides with an expansion in genome size (with the exception of *Tinantia pringlei*), which reflects the pattern of increasing genome size and latitude in Mexican Commelinaceae [19]. Why is this pattern not reflected in a tree-wide correlation? First, the latter study analyzed data without the benefit of a phylogeny, so sampling of closely related lineages that share the same traits may have biased the test. Second, comparative biology studies are especially sensitive to the method with which data are handled. The correlational test implemented in PDAP, for example, requires forcing discrete characters (life history traits) into a continuous framework. On the other hand, ancestral state reconstructions bin continuous data, like genome size, into somewhat arbitrary categories. The decision-making strategy for data

management is partly limited by available data. Character state data was unavailable for some of the more enigmatic taxa in this study; such gaps in the dataset may dramatically alter the outcome of these analyses. In the case of ancestral state reconstructions, taxon (especially outgroup) sampling is vital to properly polarize characters. Additional taxon sampling assisted in resolving taxonomic relationships for the *Tradescantia* alliance, but even more sampling will likely be required to fully understand trait evolution in this group.

Limitations of data

Both loci sampled for this study are from the plant plastomes; their relatively high rates of evolution often result in complex insertion/deletion polymorphisms (indels) that cause alignment difficulties [41]. Additional methods for evaluating or modeling indel evolution simultaneously with tree estimation may assist in sorting phylogenetic signal from homoplasy in such datasets [42,43]. Despite the rapidly evolving nature of the two chloroplast loci utilized in this study, virtually no variation was found to differentiate the erect Tradescantia. Whole plastome sequencing promises to discern molecular variation between even closely related species [44]. Finally, greater taxon sampling and data sampling from the nuclear genome may resolve some of the more difficult questions in the group, including the placement of Elasis and additional taxa. As several members of the Tradescantia alliance are hypothesized to have arisen via hybridization [4], additional data will likely resolve some of these issues.

Acknowledgements

KLH is funded by an MU Life Sciences Fellowship and graduate research grants from

the Botanical Society of America, the Society for Systematic Biologists, and the MU Graduate School. The authors acknowledge the National Science Foundation (DEB 0829849) for funding and Tori Docktor for lab assistance.

Literature Cited

1. Jones K, Kenton A (1984) Mechanisms of chromosome change in the evolution of the tribe Tradscantieae (Commelinaceae). In: Sharma AK, Sharma A, editors. Chromosomes in Evolution of Eukaryotic Groups. Boca Raton, FL: CRC Press. pp. 143-168.

2. Tomlinson PB (1966) Anatomical data in the classification of the Commelinaceae. Journal of the Linnaean Society of London: Botany 59: 371-395.

3. Evans TM, Faden RB, Simpson MG, Sytsma KJ (2000) Phylogenetic Relationships in the Commelinaceae: I. A. Cladistic Analysis of Morphological Data. Systematic Botany 25: 668-691.

4. Anderson E (1936) Hybridization in American Tradescantias. Annals of the Missouri Botanical Garden 23: 511-525.

5. Clarke CB (1881) Commelinaceae. In: Candolle ADCaCD, editor. Monographiae Phanerogamarum. Paris: G. Masson. pp. 113-324.

6. Handlos WL (1975) The taxonomy of Tripogandra (Commelinaceae). Rhodora 77: 213-319.

7. Hunt DR (1978) Three new genera in Commelinaceae: American Commelinaceae VI. Kew Bulletin 33: 331-334.

8. Torrey J (1859) Botany of the Mexican Boundary.

9. Tharp BC (1922) Commelinantia, a New Genus of the Commelinaceae. Bulletin of the Torrey Botanical Club 49: 269-275.

10. Tharp BC (1956) Commelinantia (Commelineae): An Evaluation of Its Generic Status. Bulletin of the Torrey Botanical Club 83: 107-112.

11. Brenan JPM (1966) The classification of Commelinaceae. Journal of the Linnaean Society of London: Botany 59: 349-370.

12. Hunt DR (1993) The Commelinaceae of Mexico. In: Ramamoorthy TP, Bye R, Lot A, Fa J, editors. Biological Diversity of Mexico: Origins and Distribution. New York: Oxford University Press. pp. 421-437.

13. Faden RB (1991) The classification of the Commelinaceae. Taxon 40: 19-31.

14. Woodson RE, Jr. (1942) Commentary on the North American Genera of Commelinaceae. Annals of the Missouri Botanical Garden 29: 141-154.

15. Evans TM, Sytsma KJ, Faden RB, Givnish TJ (2003) Phylogenetic relationships in the Commelinaceae: II. A cladistic analysis of rbcL sequences and morphology. Systematic Botany 28: 270.

16. Wade DJ, Evans TM, Faden RB (2006) Subtribal relationships in the tribe Tradescantieae (Commelinaceae) based on molecular and morphological data. Proceedings for the Third International Symposium on Monocots Ontario, California

17. Burns Moriuchi JH (2006) A comparison of invasive and noninvasive Commelinaceae in a phylogenetic context: The Florida State University. 190 p.

18. Bergamo S (2003) A phylogenetic evaluation of Callisia Loefl. (Commelinaceae) based on molecular data. Athens, GA: University of Georgia, Athens. 160 p.

19. Albach DC, Greilhuber J (2004) Genome size variation and evolution in Veronica. Annals of Botany 94: 897-911.

20. Martinez A, Ginzo HD (1985) DNA Content In Tradescantia. Canadian Journal of Genetics & Cytology 27: 766-775.

21. Felsenstein J (1985) Phylogenies and the comparative method. American Naturalist 125: 1-15.

22. Hunt DR (1980) Sections and series in Tradescantia: American Commelinaceae IX. Kew Bulletin 35: 437-442.

23. Hunt DR (1985) A revision of Gibasis Rafin. Kew Bulletin 4: 107-129.

24. Hunt DR (1986) Amplification of Callisia Loefl.: American Commelinaceae XV. Kew Bulletin 41: 407-412.

25. Smith JF, Sytsma KJ, Shoemaker JS, Smith RL (1991) A qualitative comparison of total cellular DNA extraction protocols. Phytochemical Bulletin 23: 2-9.

26. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, et al. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92: 142-166.

27. Taberlet P, L. Geilly, G. Pautou, and J. Bouvet (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17: 1105-1109.

28. Blattner FR, Schwei TE (2007) Lasergene. DNAStar.

29. Edgar R (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

30. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

31. Stamatakis A (2006) RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics 22: 2688–2690.

32. Stamatakis A, Hoover P, Rougemont J (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology 57: 758 - 771.

33. Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17: 1246-1247.

34. Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9: 208-218.

35. Hertweck KL, Steele PR, Pires JC (in preparation) Obtaining DNA sequences from three genomic partitions using Illumina genomic survey sequences of monocots and reference based assembly methods.

36. Bennett MD, Leitch IJ (2010) Angiosperm DNA C-values database. http://www.kew.org/cvalues.

37. Owens SJ (1981) Self-incompatibility in the Commelinaceae. Annals Of Botany 47: 567-581.

38. Shimwell DW (1972) The description and classification of vegetation. Seatlle: University of Washington Press. 322 p.

39. Maddison W, Maddison DR (2010) Mesquite. 2.74 ed.

40. Midford PE, Garland TJ, Maddison WP (2005) PDAP Package of Mesquite. 1.07 ed.

41. Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Systematic Biology 49: 652-670.

42. Golubchik T, Wise MJ, Easteal S, Jermiin LS (2007) Mind the Gaps: Evidence of Bias in Estimates of Multiple Sequence Alignments. Mol Biol Evol 24: 2433-2442.

43. Suchard MA, Redelings BD (2006) BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. Bioinformatics 22: 2047-2048.

44. Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T (2009) Rapid and Accurate Large-Scale Coestimation of Sequence Alignments and Phylogenetic Trees. Science 324: 1561-1564.

45. Steele PR, Hertweck KL, Mayfield D, Pflug J, Pires JC (in prep) Species identification using evidence from total genomic data.

46. Shimodaira H (2002) An Approximately Unbiased Test of Phylogenetic Tree Selection. Syst Biol 51: 492-508.

47. Kishino H, Hasegawa M (1989) Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution 29: 170-179.

48. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16: 1114-1116.

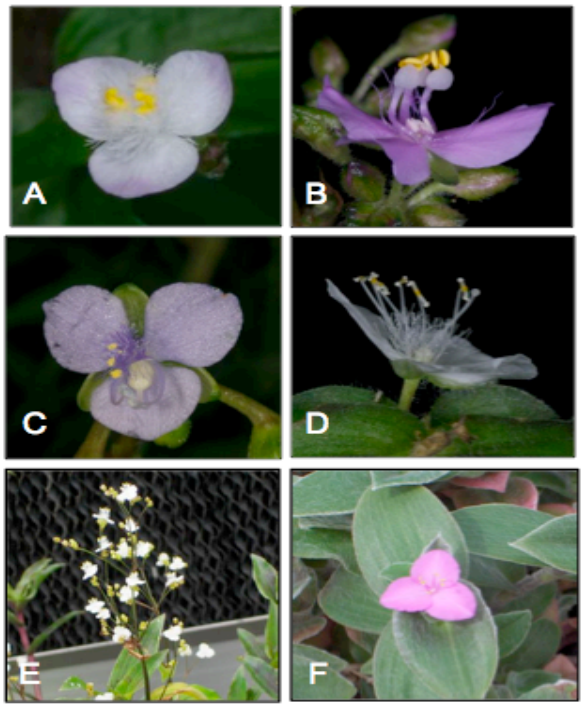

Figure 1. Floral morpological diversity in the *Tradscantia* alliance. Selected exemplars represent characteristic features of each genus. Floral morphology: A. *Gibasis*, B. *Tripogandra*, C. *Tinantia*, D. *Tradescantia*. Inflorescence morphology: E. *Gibasis*, F. *Tradescantia*.

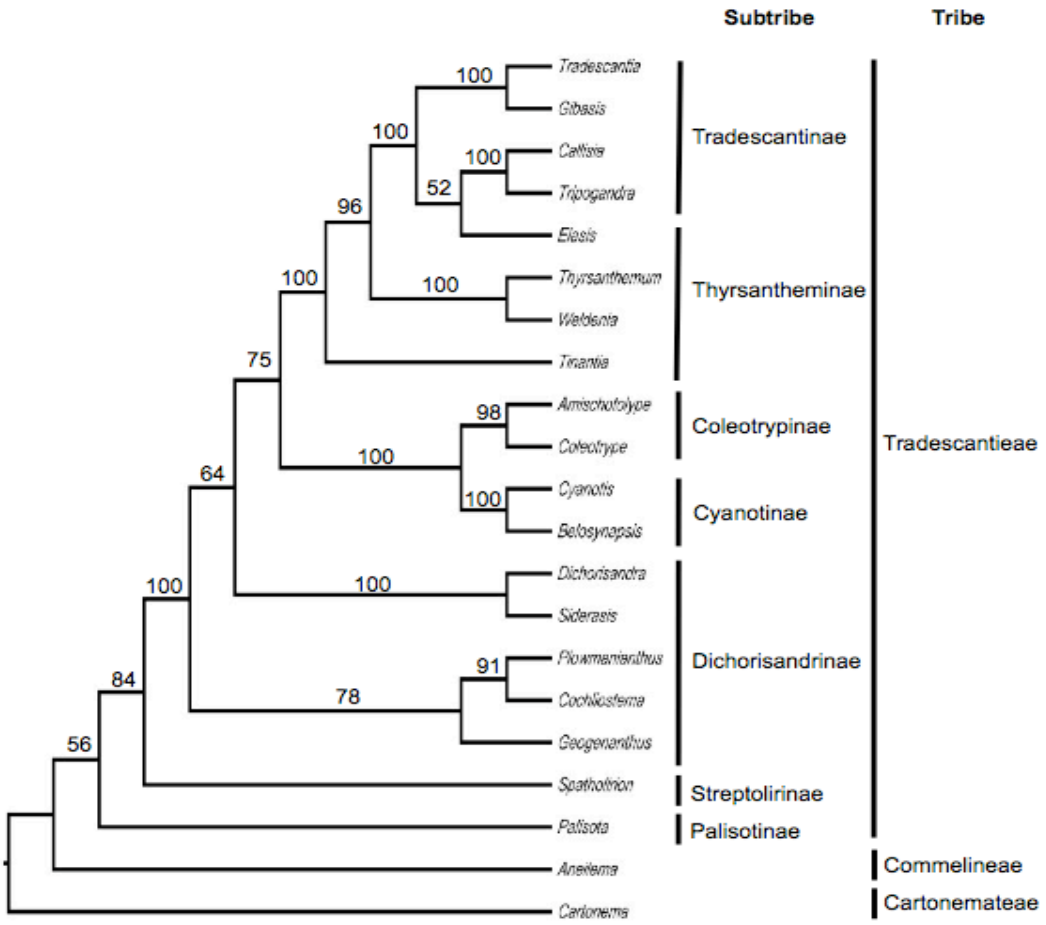

Figure 2. Previous hypothesis for phylogenetic relationships in tribe Tradescantieae. Modified from [15], inferred from one taxon per genus from morphological and molecular data. Numbers by nodes represent bootstrap support.

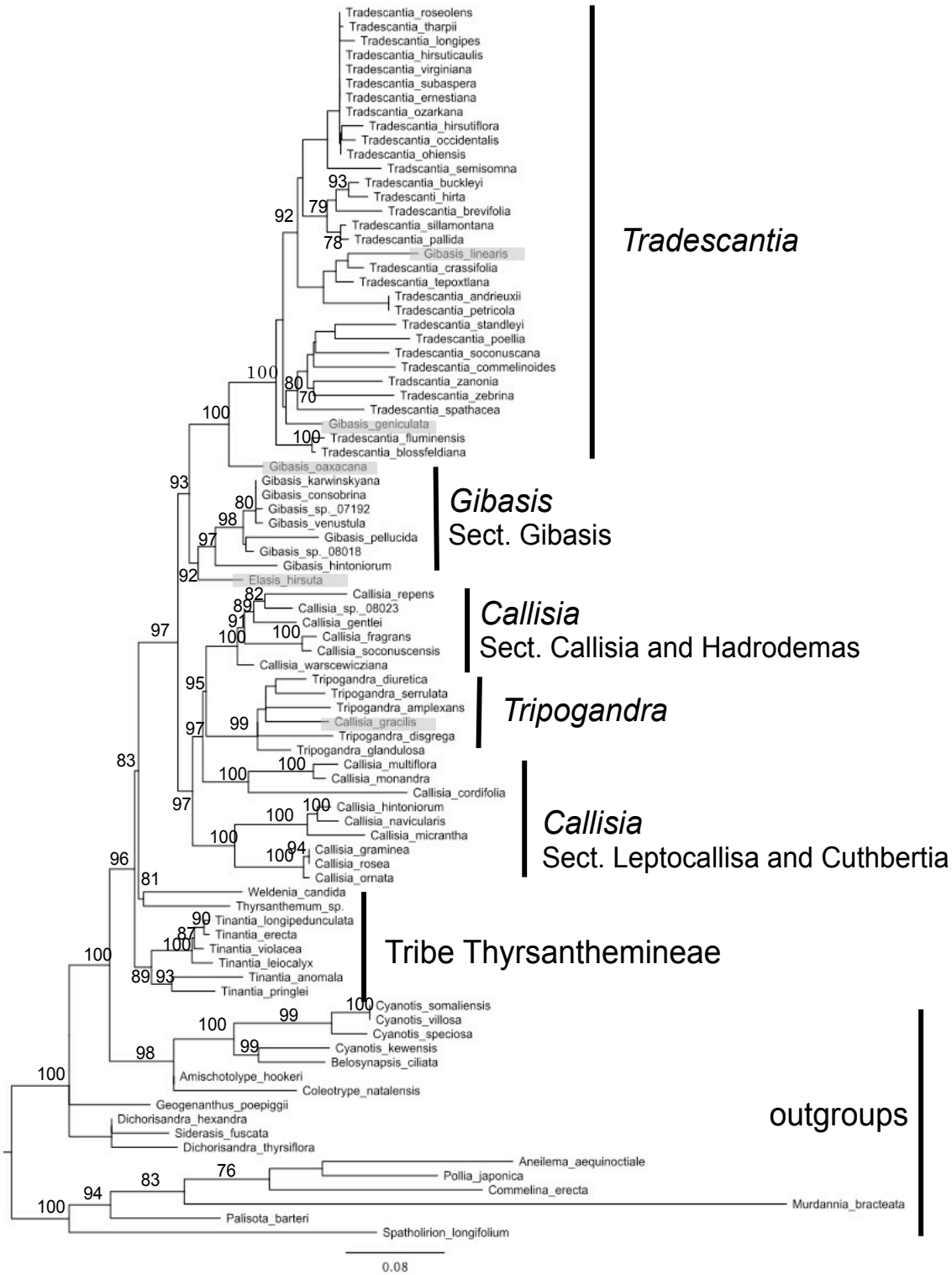
Tradescantia_roseolens
Tradescantia_tharpii
Tradescantia_longipes
Tradescantia_hirsuticaulis
Tradescantia_virginiana
Tradescantia_subaspera
Tradescantia_ernestiana
Tradscantia_ozarkana
Tradescantia_hirsutiflora
Tradescantia_occidentalis
Tradescantia_ohiensis
Tradscantia_semisomna
Tradescantia_buckleyi
Tradescanti_hirta
Tradescantia_brevifolia
Tradescantia_sillamontana
Tradescantia_pallida
Gibasis_linearis
Tradescantia_crassifolia
Tradescantia_tepoxtlana
Tradescantia_andrieuxii
Tradescantia_petricola
Tradescantia_standleyi
Tradescantia_poellia
Tradescantia_soconuscana
Tradescantia_commelinoides
Tradscantia_zanonia
Tradescantia_zebrina
Tradescantia_spathacea
Gibasis_geniculata
Tradescantia_fluminensis
Tradescantia_blossfeldiana
Gibasis_oaxacana
Gibasis_karwinskyana
Gibasis_consobrina
Gibasis_sp._07192
Gibasis_venustula
Gibasis_pellucida
Gibasis_sp._08018
Gibasis_hintoniorum
Elasis_hirsuta
Callisia_repens
Callisia_sp._08023
Callisia_gentlei
Callisia_fragrans
Callisia_soconuscensis
Callisia_warscewicziana
Tripogandra_diuretica
Tripogandra_serrulata
Tripogandra_amplexans
Callisia_gracilis
Tripogandra_disgrega
Tripogandra_glandulosa
Callisia_multiflora
Callisia_monandra
Callisia_cordifolia
Callisia_hintoniorum
Callisia_navicularis
Callisia_micrantha
Callisia_graminea
Callisia_rosea
Callisia_ornata
Weldenia_candida
Thyrsanthemum_sp.
Tinantia_longipeduncluta
Tinantia_erecta
Tinantia_violacea
Tinantia_leiocalyx
Tinantia_anomala
Tinantia_pringlei
Cyanotis_somaliensis
Cyanotis_villosa
Cyanotis_speciosa
Cyanotis_kewensis
Belosynapsis_ciliata
Amischotolype_hookeri
Coleotrype_natalensis
Geogenanthus_poepiggii
Dichorisandra_hexandra
Siderasis_fuscata
Dichorisandra_thyrsiflora
Aneilema_aequinoctiale
Pollia_japonica
Commelina_erecta
Murdannia_bracteata
Palisota_barteri
Spatholirion_longifolium
0.08
Tradescantia
Gibasis
Sect. Gibasis
Callisia
Sect. Callisia and Hadrodemas
Tripogandra
Callisia
Sect. Leptocallisa and Cuthbertia
Tribe Thyrsanthemineae
outgroups

Figure 3. cpDNA phylogram of the *Tradescantia* alliance from trnL-trn-F and rpL16. Numbers by nodes represent bootstrap support (BS, 100 replicates). Main taxonomic groups are highlighted; section. Taxa shaded in gray are displaced from their current taxonomically assigned clade. *Tinantia* alone is confirmed as monophyletic; *Callisia*, *Gibasis*, *Tradescantia* and *Tripogandra* are polyphyletic.

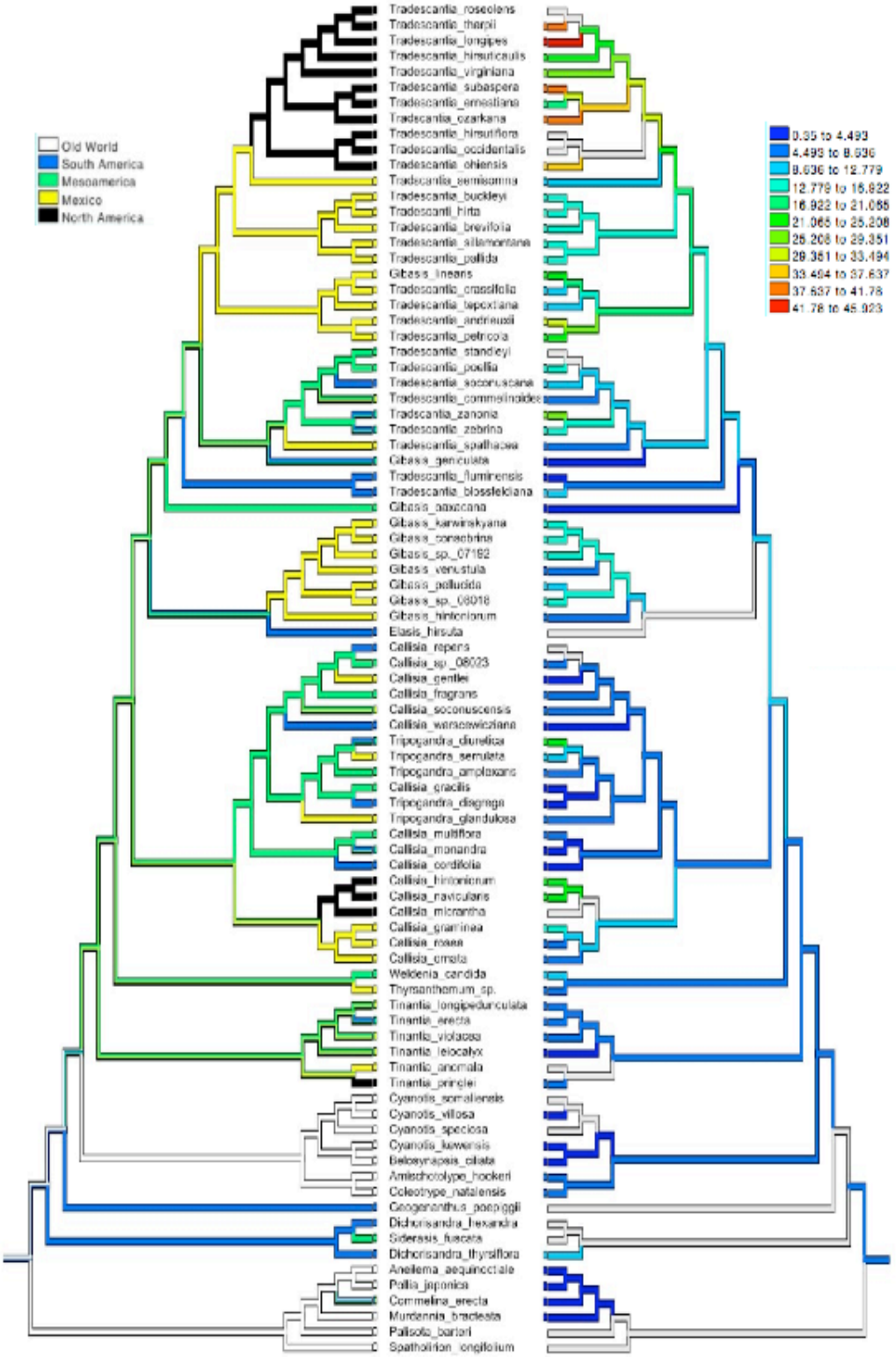
Old World
South America
Mesoamerica
Mexico
North America
0.35 to 4.493
4.493 to 8.636
8.636 to 12.779
12.779 to 16.922
16.922 to 21.065
21.065 to 25.208
25.208 to 29.351
29.351 to 33.494
33.494 to 37.637
37.637 to 41.78
41.78 to 45.923
Tradescantia_roseolens
Tradescantia_tharpii
Tradescantia_longipes
Tradescantia_hirsuticaulis
Tradescantia_virginiana
Tradescantia_subaspera
Tradescantia_ernestiana
Tradescantia_ozarkana
Tradescantia_hirsutiflora
Tradescantia_occidentalis
Tradescantia_ohiensis
Tradescantia_semiglabra
Tradescantia_buckleyi
Tradescantia_hirta
Tradescantia_brevifolia
Tradescantia_sillamontana
Tradescantia_pallida
Gibasis_linearis
Tradescantia_crassifolia
Tradescantia_tepoxtlana
Tradescantia_andrieuxii
Tradescantia_petricola
Tradescantia_standleyi
Tradescantia_pusilla
Tradescantia_soconuscana
Tradescantia_commelinoides
Tradscantia_zanonia
Tradescantia_zebrina
Tradescantia_spathacea
Gibasis_geniculata
Tradescantia_fluminensis
Tradescantia_blossfeldiana
Gibasis_oaxacana
Gibasis_karwinskyana
Gibasis_cerebrina
Gibasis_sp._07192
Gibasis_venustula
Gibasis_pellucida
Gibasis_sp._06018
Gibasis_hintoniorum
Elasis_hirsuta
Callisia_repens
Callisia_sp._08023
Callisia_gentlei
Callisia_fragrans
Callisia_soconuscensis
Callisia_warscewicziana
Tripogandra_diuretica
Tripogandra_serrulata
Tripogandra_amplexans
Callisia_gracilis
Tripogandra_disgrega
Tripogandra_glandulosa
Callisia_multiflora
Callisia_monandra
Callisia_cordifolia
Callisia_hintoniorum
Callisia_navicularis
Callisia_micrantha
Callisia_graminea
Callisia_rosea
Callisia_ornata
Weldenia_candida
Thyrsanthemum_sp.
Tinantia_longipeduniculata
Tinantia_erecta
Tinantia_violacea
Tinantia_leiocalyx
Tinantia_anomala
Tinantia_pringlei
Cyanotis_somaliensis
Cyanotis_villosa
Cyanotis_speciosa
Cyanotis_kuwaensis
Belosynapsis_ciliata
Amischotolype_hookeri
Coleotrype_natalensis
Geogenanthus_poeppigii
Dichorisandra_hexandra
Siderasis_fuscata
Dichorisandra_thyrsiflora
Anelema_aequinoctiale
Pollia_japonica
Commelina_erecta
Murdannia_bracteata
Pollia_condensata
Spatholirion_longifolium

Figure 4. Relationship between biogeography and genome size in the *Tradescantia* alliance. Cladogram on left shows biogeographic regions; cladogram on right shows genome size categories. Ancestral reconstructions were inferred using parsimony. There is no significant relationship between biogeography (discrete trait) and genome size (continuous trait), but movements to North America correspond with two of the expansions in genome size.

Table 1: Taxa and life history traits included in the *Tradescantia* alliance phylogeny. Taxa without previous affiliation with generic sections are placed according to the ML phylogeny. Accession information includes collector, collection number, location where taxon was collected, and voucher location; commercial indicates it was obtained from a horticultural source. Abbreviations: A/P=annual/perennial. SI/SC=self incompatible/self compatible. Raunkiaer growth form: C=chamaephyte, G=geophyte, H=hemicryptophyte, T=therophyte. Growth habit: C=creeping, E=erect, D=decumbent, R=rosette. Biogeography: O=Old World, S=South America, C=Mesoamerica/Central America, M=Mexico, N=North America (United States). A dash (-) indicates missing data. For genome sizes, a single asterisk (*) indicates values were obtained from the Plant DNA C-Value Database [35]. A double asterisk (**) indicates an alternate accession of that species from our living collections was used for genome sizing.

Taxon	Accession	Life history schedule	Breeding system	Raunkiaer growth form	Growth habit	Biogeo-graphy	Genome size (pg/1C)
TRIBE TRADESCANTIEAE MEISNER							
SUBTRIBE TRADESCANTIINAE ROHW.							
***Tradescantia* L.**							
Section Austrotradescantia D.R.Hunt							
Tradescantia fluminensis Vellozo	KH0676, commercial (UMO)	P	SC	C	C	S	4.49
Section Campelia (L.C.Rich)D.R.Hunt							
Tradescantia zanonia (L.)Sw.	KH0686, commercial (UMO)	P	SI	C	E	S	13.75
Section Corrina (D.R.Hunt)							
Tradescantia soconuscana Matuda	Faden 76/98, Smithsonian 80-365	P	SI	C	D	C	12.02
Section Cymbispatha (Pichon)D.R.Hunt							
Tradescantia commelinoides Schultes et Schultes f.	KH07161, Mexico (UMO)	P	SC	G	D	C	8.03
Tradescantia poelliae D.R.Hunt	Grant 92-1863, Costa Rica; SI 1992-049	P	SI	C	C	C	13.75*
Tradescantia standleyi Steyerm.	Kew 18847	P	SI	-	C	C	-
Section Mandonia D.R.Hunt							
Tradescantia petricola J.R.Grant	Grant 95-2347, Costa Rica, SI 1995-317	P	SC	G	E	M	31.3
Tradescantia crassifolia Cavanilles	Peterson et al. 16911, Mexico, SI 2003-010	P	SC	G	E	M	24.9*
Tradescantia tepoxtlana Matuda	KH07175, Mexico (UMO)	P	SC	G	E	M	9.48
Section Parasetcreasea D.R.Hunt							
Tradescantia andrieuxii C.B.Clark	KH08079, Mexico (UMO)	P	SC	G	E	M	21.53
Section Rhoeo (Hance) D.R.Hunt							
Tradescantia spathacea Sw.	KH0678, commercial (UMO)	P	SC	H	R	M	7.15**

Section Setcreasea (K.Schum.&Sydow)D.R.Hunt							
Tradescantia brevifolia (Torrey) Rose	Faden, Burns 283 (FSU)	P	SI	C	D	M	14.9
Tradescantia buckleyi (I.M.Johnston) D.R. Hunt	SI 1992-047	P	SI	C	D	M	16.26
Tradescantia hirta D.R.Hunt	KH07196, Mexico (UMO)	P	SI	G	E	M	14.74**
Tradescantia pallida (Rose) D.R.Hunt	KH0502, commercial (UMO)	P	SI	C	C	M	14.99
Section Tradescantia							
Tradescantia semisomna Standl.	KH07133, Mexico (UMO)	P	SI	G	E	M	12.65
Series Sillamontanae D.R.Hunt							
Tradescantia sillamontana Matuda	KH0682, commercial (UMO)	P	SC	C	C	M	14.13
Series Virginianae D.R.Hunt (erect *Tradescantia*)							
Tradescantia ernestiana Anderson&Woodson	KH0617, Arkansas (UMO)	P	SI	H	E	N	20.35*
Tradescantia hirsuticaulis Small	KH0735, Arkansas (UMO)	P	SI	H	E	N	21.6*
Tradescantia hirsutiflora Bush	Burns 279, Florida (FSU)	P	SI	H	E	N	-
Tradescantia longipes Anderson&Woodson	KH07123, Missouri (UMO)	P	SI	H	E	N	41.78*
Tradescantia occidentalis (Britton)Smyth	Burns 286, commercial (FSU)	P	SI	H	E	N	-
Tradescantia ohiensis Rafinesque	KH0637, Missouri (UMO)	P	SI	H	E	N	36.75**
Tradescantia ozarkana Anderson&Woodson	KH0610, Missouri (UMO)	P	SI	H	E	N	41.32**
Tradescantia roseolens Small	Bergamo 99-186, Florida (GA)	P	SI	H	E	N	
Tradescantia subaspera Ker Gawler	KH0646, Missouri (UMO)	P	SI	H	E	N	38.5**
Tradescantia tharpii Anderson&Woodson	KH07203, Missouri (UMO)	P	SI	H	E	N	39
Tradescantia virginiana L.	KH0631, Indiana (UMO)	P	SI	H	E	N	27.39

Section Zebrina (Schnizlein)D.R.Hunt							
Tradescantia blossfeldiana Mildbr.	Smithsonian 80-362	P	SC	C	D	S	8.75
Tradescantia zebrina hort ex. Bosse	KH0501, commercial (UMO)						
***Gibasis* Raf.**							
Section Gibasis							
Gibasis consobrina D.R.Hunt	Kew 18843, Mexico	P	SI	G	D	M	15.66**
Gibasis karwinskyana (Roem.&Schult.)Rohw.	Kew 18844, unknown	P	SI	G	D	M	12.94**
Gibasis hintoniorum Turner	KH07191, Mexico (UMO)	P	X	G	E	M	6.53
Gibasis linearis (Benth)Rohw.	KH07126, Mexico (UMO)	P	SI	G	E	M	12.5
Gibasis pellucida (M.Martens&Galeotti)D.R.Hunt	Burns 248, Florida (FSU)	P	SC	C	C	M	11.23
Gibasis pulchella Raf.	KH07192, Mexico (UMO)	P	-	G	E	M	15.41
Gibasis venustula (Kunth)D.R.Hunt	J. Bogner s.n. Mexico SI 2003-081	P	SI	G	E	M	7.88
Gibasis sp.	KH08018, Mexico (UMO)	P	SI	G	D	M	16.72
Section Heterobasis D.RHunt							
Gibasis geniculata (Jacq)Rohw.	KH0681, commercial (UMO)	P	SC	C	C	S	3.16
Gibasis oaxacana D.R.Hunt	Faden, SI	P	SI	C	C	C	2.94
***Callisia* Loefl.**							
Section Brachyphylla D.R.Hunt							
Callisia hintoniorum Turner	KH07197, Mexico (UMO)	P	-	G	E	M	8.36
Callisia micrantha (Torrey) D.R.Hunt	Bergamo 00-268 (GA)	P	SI	C	C	M	5.02
Callisia navicularis	KH0697, commercial (UMO)	P	SC	C	C	M	13.95
Section Callisia							
Group “gentlei”							
Callisia gentlei Matuda	KH0689, commercial (UMO)	P	SI	C	C	C	7.07

Group "fragrans"							
Callisia fragrans (Lindley) Woodson	KH0674, commercial (UMO)	P	SI	C	R	M	3.85
Callisia soconuscensis Matuda	Bergamo 86-203 (GA)	P	SI	C	C	C	1.13
Group "repens"							
Callisia repens (Jacquin) Linnaeus	KH07201, Mexico (UMO)	P	SC	C	C	C	24.5*
Callisia sp.	KH08023, Mexico (UMO)	P	-	C	C	M	9.0
Section Cuthbertia (Small)D.R.Hunt							
Callisia graminea (Small)G.Tucker	Bergamo 99-189, Giles 93L-1 (GA)	P	SI	G	E	N	47.22*
Callisia ornata (Small)G.C.Tucker	KH, Florida (UMO)	P	-	G	E	N	-
Callisia rosea (Ventenat)D.R.Hunt	Bergamo 99-198 (GA)	P		G	E	N	21.76**
Section Hadrodemas (H.E.Moore)D.R.Hunt							
Callisia warscewicziana (Kunth st Bouche) D.R.Hunt	Bergamo 97-068 (GA)	P	SI	C	R	M	5.02
Section Leptocallisia							
Callisia cordifolia (Swartz)E.S.Anderson&Woodson	Faden 83/37, Smithsonian 83-197	A	SC	T	C	S	4.05
Callisia gracilis (Kunth)D.R.Hunt	Faden 01-075, Grant 3984 (Smithsonian)	A	SC	T	C	C	4.96
Callisia monandra (Sw.)Schultes et Schultes f.	*J. Bogner* s.n., Munich Bot. Gart.; SI 1993-092	A	SC	T	C	S	2.7
Callisia multiflora (Mart&Gal)Standl.	Bergamo 80-395 (GA)	P	SC	C	C	C	6.65
***Tripogandra* Raf.**							
Tripogandra amplexans Handlos	KH07172, Mexico (UMO)	P	SC	C	D	S	8.75
Tripogandra disgrega (Kunth)Woodson	KH07159, Mexico (UMO)	A	SC	T	E	C	6.56**
Tripogandra diuretica (Mart.)Handlos	Plowman 10171, Brazil SI 1980-368	P	SC	C	C	S	-
Tripogandra glandulosa	Faden, SI	P	SC	3.86	C	C	S

(Seub.)Rohw.							
Tripogandra serrulata (Vahl) Handlos	KH0679, commercial (UMO)	P	SC	C	C	C	6.71
SUBTRIBE THYRSANTHEMINAE FADEN&D.R.HUNT							
Elasis hirsuta (Kunth)D.R.Hunt	MacDougal and Lalumondier 4953 (Kew)	P	-	C	D	S	-
Thyrsanthemum sp.	M. Chase 606 (Kew)	P	SI	G	E	M	7.23**
Weldenia candida Schultes f.	M. Chase 592 (Kew)	P	SI	G	R	C	10**
***Tinantia*Scheidw.**							
Tinantia anomala (Torrey) C.B.Clarke	KH07094, Texas (UMO)	A	SC	T	E	N	6.29
Tinantia erecta (Jacq.)Schlecht	KH07186, Mexico (UMO)	A	SC	T	E	S	8.5
Tinantia leiocalyx C.B.Clarke ex J.D.Smith	KH08077, Mexico (UMO)	A	SC	T	E	C	3.76
Tinantia longipedunculata Standl.&Steyerm,	KH08075, Mexico (UMO)	A	SC	T	E	C	6.78
Tinantia pringlei (S.Wats.)Rohw.	Faden, Burns 267 (FSU)	P	SC	T	E	M	-
Tinantia violacea Rohw.	KH07162, Mexico (UMO)	A	-	T	E	C	5.61
SUBTRIBE COLEOTRYPINAE FADEN&D.R.HUNT							
Amischotolype hookeri (Hassk.)Hara	Hahn 6041, Thailand, SI1990-023	P	-	C	E	O	8.33
Coleotrype natalensis C.B.Clarke	Faden 74/206, South Africa, SI 1983-399	P	SI	C	E	O	6.2
SUBTRIBE CYANOTINAE (PICHON)FADEN&D.R.HUNT							
Belosynapsis ciliata (Blume)R.S.Rao	Winters, Higgins & Higgins 186, New Guinea, SI 1982-232	P	-	C	C	O	0.35
Cyanotis kewensis C.B.Clarke	KH06105, commercial (UMO)	P	-	C	C	O	1.9
Cyanotis somaliensis C.B.Clarke	MOBOT 1972-1486	P	SC	C	C	O	2.63
Cyanotis speciosa (L.f.)Hassk.	Burns ? (FSU)	P	SI	-	D	O	-
Cyanotis villosa (Spreng.)Schult.f.	Faden 76/555 (GA)	-	SC	-	D	O	-
SUBTRIBE DICHORISANDRINAE (PICHON)FADEN&D.R.HUNT							

Dichorisandra hexandra (Aubl.)Standl.	DeGranville et. al s.n., French Guiana, Smithsonian 89-070	P	SI	C	D	S	-
Dichorisandra thyrsiflora Mikan.	MOBOT 1980-1258	P	SI	C	E	S	11.69
Geogenanthus poeppigii (Miq.)Faden	MOBOT 1998-1414	P	-	C	D	S	-
Siderasis fuscata (Lodd.)H.E.Moore	KH0699, commercial (UMO)	P	SC	C	D	S	-
SUBTRIBE PALISOTINAE FADEN&D.R.Hunt							
Palisota barteri Hook	Faden, SI	P	SC	C	R	O	-
TRIBE COMMELINEAE BRUCKNER							
Aneilema aequinoctiale (P.Beauv.) G.Don	Bolnick s.n., Mozambique, SI 2002-202	P	SC	C	C	O	1.87
Commelina erecta L.	Burns 250, Florida (FSU)	P	SC	G	D	O	2.58
Murdannia bracteata	MOBOT 1995-1919	P	SC	C	C	O	1.29
Pollia japonica Thunberg	MOBOT 1978-0933	P	SC	C	E	O	1.11
Spatholirion longifolium (Gagnep.)Dunn	Unknown, GenBank	P	-	-	-	O	-

Table 2. Characteristics of the two locus chloroplast gene dataset.

	rpL16	**trnL-trnF**	**Combined**
# included taxa	70	84	87
Total length (bp)	1989	1634	3623
Shortest sequence	645 (*Tradescantia* 07123)	270 (*Tripogandra glandulosa*)	N/A
Longest sequence	1243 (*Tradescantia petricola*)	1192 (*Dichorisandra hexandra*)	N/A
% variable	58.2	59.7	58.9
% missing/gaps	48.2	48.7	54.76

Table 3. Constraint tests for monophyly of taxonomic groups. Asterisks indicate constrainted trees that were not significantly different from the unconstrained tree. P-values are indicated for each of the following topological hypothesis tests: AU=Approximately Unbiased [45], KH=Kishino-Hasegawa [46], SH=Shimodaira-Hasegawa [47], WKH=weighted KH, WSH=weighted SH.

Taxonomic group	Likelihood of best tree	AU	KH/WKH/WSH	SH
unconstrained	-21647.129702	1.000	1.000	1.000
Tradescantia	-22573.861668	3e-05*	0*	4e-05*
Gibasis	-21831.247025	2e-07*	0*	0.179
Callisia	-22347.737631	7e-07*	0*	0.19
Subtribe Tradescantiinae	-24968.999745	2e-50*	0*	0*
Subtribe Thyrsantheminae	-21842.12	2e-49*	0*	0*

Table 4. Character evolution in the *Tradescantia* alliance. Ancestral state reconstructions are inferred from parsimony. Correlations with genome size results are p-values (two-tailed) from Felsenstein's Independent Constrasts.

Life history trait	**Ancestral state (whole tree)**	**Ancestral state (*Tradescantia* alliance)**	**Correlation with genome size**
Life history schedule	perennial	perennial	0.32
Breeding system	SC	SC	0.23
Raunkier growth form	chamaephyte	chamaephyte	0.64
Growth habit	rosette	erect	0.23
Biogeography	Old World/South America	Equivocal (New World)	0.15

CHAPTER 4

ASSEMBLY OF THREE GENOMIC PARTITIONS FROM ILLUMINA GENOME SURVEY SEQUENCES

Abstract

Low redundancy and shallow coverage genome survey sequences (GSS) from massively parallel sequencing have the potential to rapidly provide large, cost-effective datasets for phylogenetic inference, replace single gene or spacer regions as DNA barcodes, and provide a plethora of data for other comparative molecular evolution studies. The application of GSS to non-model systems, however, is hindered by a lack of understanding regarding how robustness of assembled plastomes, mitogenomes, and nuclear ribosomal (nrDNA) loci differ based on phylogenetic relatedness of reference sequences used to build contigs. Our goal was to determine the type (plastome, mitogenomic, and nrDNA sequences) and quality of assembled genomic data attainable from Illumina 80-100 bp single-end GSS. We tested our methods by sequencing total genomic DNA from taxa belonging to two lineages of monocotyledonous plants: the grass family (Poaceae), a model system, and the order Asparagales (including asparagus, onion and agave), a non-model system. We compared our reference-based assemblies to *de novo* contigs in three Poaceae taxa, for which complete genome sequences are available for confirmation of accuracy, to serve as a control. We also evaluated consistency of assemblies resulting from the use of different reference sequences, both closely and distantly related to the sequenced taxon, in YASRA. Our Asparagales sampling included 48 taxa representing broad variation in genome

size and life history traits; we evaluated the success of our methods to obtain assemblies from non-model taxa. We found that our easily implemented, low-cost approach to sequencing total genomic DNA can return reliable, robust organellar and nrDNA sequences in a variety of plant lineages. Additionally, high quality assemblies are not dependent on genome size, amount of plastid present in the total genomic DNA template, or relatedness of available reference sequences for assembly, allowing our methods to be implemented widely in plant groups.

Introduction

Massively parallel sequencing (MPS) has revolutionized molecular evolution by making genomic sequencing possible for many more organisms than previously attainable. While this technology is allowing unprecedented access to raw sequence data, storing, managing, and processing such data remains daunting. Genome survey sequences (GSS) present an enticing alternative to complete genome sequencing and assembly; this method utilizes non-targeted MP sequencing of total genomic DNA to shallowly sequence the entire genomic compliment with low coverage and redundancy. While GSS projects generally prohibit assembly of the complete genome, sequences present in high copy number, including organellar (plastid and mitochondrial) and nuclear ribosomal genes (nrDNA), are more easily assembled. The terms plastome and mitogenome have been described in various contexts; these terms may refer to just the genic (coding) portions of the genome, or the entire genomic compliment. For the purposes of our study, we will use plastome and mitogenome to refer to the complete genome in each respective organelle, including

intergenic and spacer regions. Reference taxa are the organisms to which GSS is being applied. A target taxon, conversely, is the organism with a previously sequenced genome that is used as a reference for assembly purposes.

Standards for complete genome sequencing require high coverage to ensure assembly and prevent sequencing errors. Releasing preliminary results from in-progress sequencing projects, like assemblies from 2X coverage of a genome, is often seen as a way to “whet users' appetites” for high coverage, fully sequenced versions of the same genome. Indeed, many questions in comparative genomics are impossible to answer with sparse coverage [1]. However, low coverage GSS has yielded impressive results when comparisons with closely related reference species are sought. For example, overlaying 0.66X coverage of the pig genome to a human-mouse alignment revealed comparisons between 38% of the coding fraction of the genome [2]. Similar coverage (0.1X) in scuttle fly allowed almost complete reconstruction of the mitogenome as well as information about repetitive elements and some functional genes [3]. When syntenically aligned to a well assembled and annotated reference genome, sparse sequencing of related taxa can even provide robust enough information to infer levels of recombination, introgression, and chromosomal restructuring [4].

The studies cited above used either conventional Sanger sequencing or 454 MPS data to obtain sequence information about genomes. While these methods provide relatively long sequence reads (~1000 and ~400 bp, respectively), they are more costly and/or labor intensive. Illumina (Solexa) sequencing is an alternative MPS technology that provides shorter sequence reads (for this study, ~80 bp) at a more reasonable cost per

taxon. Nock et al. [5] sequenced total genomic DNA on one Illumina lane (36 bp reads) per taxon for five grass species. When compared to a previously sequenced rice plastome reference, they were able to assemble complete plastomes for the target species with 100-750x median coverage. Their success contrasts with prior expectations that plastomes could only be assembled from GSS of DNA enriched for plastids [i.e., chloroplast isolations,6].

Plastomes are targeted for next-generation sequencing projects because of their phylogenetic utility [7,8] and high frequency relative to the nuclear genome in total genomic DNA extractions. Other genomic loci present in high copy number may be easily assembled from even relatively sparse GSS. Compared to the plastome, little is known about evolution of plant mitogenomes, partly due to larger size of this organellar genome [9], high rates of evolution [10], and fewer targeted sequencing efforts. Additional information about plant mitogenomes could prove useful for comparative studies. High-copy nrDNA loci should also be easy to assemble from the nuclear partition, and can provide independent confirmation of species identification or phylogenetic signal. Obtaining sequences from nuclear and organellar genomes from Illumina GSS has been proposed for a broad range of systematic applications [11].

Despite the apparent advantages to assembling plastomes, mitogenomes, and nrDNA from GSS, several outstanding questions hinder implementation of these methods in a wider breadth of taxa. First, most genome sequencing projects to date, including GSS, have targeted taxa with relatively small genome sizes. Larger genomes have higher repetitive element compliments that not only obscure genic content in genomes, but also confound efforts to reliably assemble large genomic contigs, or contiguous sections of

assembled short reads [12]. It is unclear how genome size, which can vary dramatically among plant lineages [13], can affect assembly quality for both nuclear genes and organellar genomes [5]. Second, current genome sequencing is focused on relatively few taxa distributed unevenly throughout the tree of life, so it is likely that a closely related reference taxon is unavailable for scientists unless they are working in a model system. Little work has investigated how phylogenetic distance of reference taxa affects assembly quality of the target genome [5]. To our knowledge, no research has examined how GSS assemblies in lesser studied taxa are affected by phylogenetic distance from reference sequences.

Our goal was to determine the type and quality of assembled genomic data (plastome, mitogenomic, and nuclear ribosomal sequence) attainable from Illumina GSS. We tested our methods in two lineages of monocotyledonous plants: family Poaceae (grasses, order Poales), and order Asparagales (which includes asparagus, orchids, irises, agave and onion). We sequenced total genomic DNA from leaf tissue with six taxa per Illumina lane and utilized a reference based assembly program to construct sequences and estimate the level of coverage for each partition. Using Poaceae taxa with published genomes available, we explored the effect phylogenetic relatedness of reference sequence to target assembly. We also compared these reference-based assemblies to *de novo* methods to discern the level of error associated with reconstruction. We tested some of the assumed limitations of these methods using non-model Asparagales taxa. We found that our easily implemented, low-cost approach to sequencing total genomic DNA can return reliable, robust organellar and nuclear ribosomal sequences in a variety of plant lineages.

High coverage plastomes are not dependent on genome size or amount of plastid present in the total genomic DNA template or availability of closely related reference sequences, allowing our methods to be implemented broadly in plants.

Methods

Taxon selection

We selected two independent lineages of monocotyledonous plants to test our methodology. The grass family (Poaceae) is comprised of many agriculturally and ecologically important herbaceous species, for which complete genome sequences have been published or are in progress for many taxa. We resequenced six grass taxa to test our ability to assemble organellar genomes from Illumina data. Three taxa (*Oryza sativa* ssp. *japonica* cv. Nipponbare, *Sorghum bicolor* cv. B Tx642, and *Zea mays* ssp. *mays* cv. B73, hereafter *Oryza*, *Sorghum*, and *Zea* B73) have substantial genomic information, including complete cytotype-specific plastomes, available through GenBank. These taxa were sequenced because the wealth of available genomic information allows them to serve as controls for the efficacy of our sequence and assembly methods, especially in the presence of structural variation [i.e., plastomes in Poales, 14]. We sequenced an additional maize inbred line (*Z. m.* ssp. *mays* va. CIMMYT Maize Inbred Line 52) and two maize wild relatives (*Z. m.* ssp. *mexicana* and *Z. m.* ssp. *parviglumis*) to examine the consistency of our methods between closely related species (hereafter, *Z. m.* CML52, *Z. m. mexicana*, and *Z. m. parviglumis*, respectively).

The monocot order Asparagales comprises three families including a broad variety of plants important to horticulture and agriculture (e.g., asparagus, onion and agave); these taxa possess quite evolutionarily labile genome sizes [15]. We sequenced 48 Asparagales taxa to test our ability to assemble contigs lineages with genome sizes that vary widely between taxa. We obtained genome size estimates for our Asparagales taxa via flow cytometry at the Benaroya Research Institute at Virginia Mason in Seattle, Washington using a protocol modified from Arumuganathan and Earle [see Supplemental Methods, 16]. When fresh leaf material from the exact accession was not available, we averaged genome sizes from individuals of the same species or used values reported from the RBG Kew Angiosperm DNA C-values database [17].

Illumina sequencing

Methods for Illumina sequencing are explained briefly here with details in Supplemental methods. We extracted total genomic DNA from ca. 20 mg silica dried or an equivalent amount of fresh leaf tissue using a Qiagen DNeasy Plant Mini Kit. For Asparagales taxa, we performed real-time (RT)-PCR to obtain a Ct (cycle threshold) value, or number of cycles required to reach the florescence threshold (indicating a signal stronger than background fluorescence). In our case, smaller Ct values indicate more plastome present in total genomic DNA. All taxa except *Asparagus asparagoides* exhibited a Ct value less than 21.0.

For Illumina library preparation, we performed end repair on sheared genomic DNA prior to ligating barcoding adapters for multiplexing. We size selected samples for ~300 bp and enriched these fragments using PCR. We sent the final product to the University of

Missouri DNA Core for quantitation, fragment size verification, and sequencing on the Illumina Genome Analyzer. All samples ran on one sixth of an Illumina lane with single-end 80 or 120 bp reads.

Sequence assembly, annotation and analysis

Processing raw reads. We parsed raw reads from sequencing of a single Illumina lane into six bins (one for each taxon in the lane) and removed barcoding adaptor tags using custom perl scripts. The same scripts also deleted sequences containing more than five ambiguous states (represented in raw sequence data as "N"). We employed a reference-based assembly strategy to mine GSS for desired sequences using YASRA (Yet Another Short Read Assembler, http://www.bx.psu.edu/miller_lab/), a reference based assembly algorithm designed for assembly of short reads into organellar genomes [18]. We used high quality sequences from closely related taxa as references (Tables 1 and 6) to assemble target sequences using the medium threshold parameter in YASRA.

Poaceae plastome assembly, annotation, and summary statistics. For grasses, we assembled plastomes using the published sequence for each taxon, which should be identical to the assembly. We reported values from the first complete YASRA assembly for Poaceae, and indicate the total number of contigs generated per assembly as a measure of the difficulty of assembling that target genome. Fewer and longer contigs are preferable for ease of assembly and annotation. We also tested the effect of phylogenetic distance of the reference from the target taxon on assembly quality by reassembling each of the grass genomes with eleven different reference sequences, ranging from closely related grasses to a distantly related cycad (Table 2). The final step of YASRA reports the percent sequence

identity (similarity) between the reference and target sequences, which provides a crude estimate of phylogenetic distance.

We evaluated how relative size of the target and reference plastomes affect plastome assembly in Poaceae using the genome length ratio (GLR), the ratio of the size (length in bp) of the target taxon to the reference taxon. We interpret this ratio as follows: GLR=1 indicates target and reference plastomes are nearly equal in length, GLR>1 indicates the target taxon plastome is larger than the reference, and GLR<1 indicates the target plastome is smaller than the reference.

We considered two possible sources of variation when evaluating quality of assembly for Illumina data from the three grass species. First, we compared sequences obtained from YASRA assemblies using different reference sequences by examining MAFFT alignments [19] in MEGA [20] to calculate the number of variable sites and insertion/deletion polymorphisms (indels). Second, we assembled sequences of each of the three grasses *de novo* using a combination of the NextGENe software package (Softgenetics, State College, PA, USA) and CAP3 analysis [21]. Detailed assembly parameters are available in Supplemental Methods.

mtDNA assemblies in Poaceae. The lability of size and structure in plant mitogenomes makes assembly difficult, especially given the paucity of available reference sequences. Furthermore, reference-based assemblies for entire mitogenomes in monocots are computationally intensive and generate hundreds or even thousands of contigs (Hertweck, data not shown), making them suboptimal for large scale phylogenetic studies. Our strategy for evaluating the presence of mitogenomic sequences in Illumina GSS was to

perform reference-based assemblies in YASRA using single mitochondrial gene sequences. We selected two genes, atp1/atpA (alpha subunit for ATP synthase) and cox3 (cytochrome oxidase) commonly used the mitochondrial genome in molecular phylogenetic studies [22,23] and extracted genic regions from published, annotated grass mitogenomes for each of three Poaceae taxa. These were run as reference sequences in YASRA using the same parameters as plastomes. We compared assemblies to both the original sequences and, because mitogenomic sequences diverge so rapidly, we performed BLAST [24] on each contig.

nrDNA assemblies in Poaceae. We performed a single YASRA run to assemble nuclear ribosomal sequences in grasses. We again tested the effects of reference sequences on assembly quality by reassembling each target genome with six reference sequences; we only used a single grass reference sequence because of the relative conservation of ribosomal genes. Prior to assembly, we aligned the raw reference sequences and trimmed them to the length of the shortest sequence on each end. This method allowed us to test the robustness of YASRA to building a longer assembly from a truncated or partial reference sequence.

Asparagales plastome assembly and annotation. The final goal of plastome assembly is to obtain a single contig representing all portions of the plastid genome, including the Inverted Repeat (IR), Large Single Copy region (LSC), and Small Single Copy region (SSC). We used an iterative process to extend the flanking regions of contigs to join them together into a single sequence for Asparagales. We input the initial result from YASRA containing multiple contigs into Geneious v5.3 [25] to align overlapping regions to each other. The

resulting sequence was fed back into YASRA as the reference sequence and run against the entire compliment of Illumina reads from that sample. This process was repeated as many times as was necessary to obtain a complete plastome. The last step was to input the complete plastid sequence into YASRA as the reference to obtain accurate summary statistics for that taxon. We recorded summary statistics for each taxon from the final iteration of the summary file output by YASRA. The percent plastome reported here is the percent of reads saved and integrated into the assembly from the full complement of Illumina reads, while plastome coverage indicates the average depth of coverage (i.e., 50X coverage of 120,000 bp template). We annotated all Asparagales plastomes using the automatic annotation program DOGMA [26]; annotated plastomes are described in Steele et. al [27]. We conducted power analysis for Asparagales plastome data using Java Applets for Power and Sample Size (from http://www.stat.uiowa.edu/~rlenth/Power).

Results

Reference tests in Poaceae. For the six Poaceae taxa, the number of reads from one sample (representing one sixth of an Illumina lane) varied from 1.82 million (*Zea* CML52) to almost 5.46 million (*Sorghum*, Table 1). The percentage of Illumina reads used in plastome reference-based assembly ranged from 0.56 (*Zea* B73) to 4.37% (*Sorghum*). The average depth of coverage for the plastome ranged from 14.6 (*Zea* CML52) to 196.5X (*Sorghum*). The largest GLR resulted from assembling *Sorghum* as a target with the *Oryza* genome (1.21, target longer than reference sequence, Table 2). The smallest GLR resulted from assembling *Oryza* with *Cycas* as the reference (0.82, target shorter than reference). Each

grass target assembled with a reference sequence from the same species resulted in identity over 99%. The lowest percent identity (94.1%) between the reference and assembled target was *Sorghum* (target) and *Cycas* (reference). *Oryza* and *Sorghum* targets assembled with their control reference sequences both resulted in a single contig spanning the entire range of the reference. The highest number of contigs (70) resulted from assembling *Oryza* with *Amborella*.

We tested for correlations between variables for each of three Poaceae taxa separately. As there were no *a priori* reasons to assume nonlinearity, all correlations presented are linear. In some comparisons R^2 improved with exponential curves, but these modifications do not change the interpretation of our results (data not shown). As percent identity between the reference and target taxon increased, both percent plastome and plastome coverage increased (Fig. 1A and 1B). As percent plastome and plastome coverage increased, the number of contigs decreased (Fig. 1C and 1D). There was no relationship between either percent plastome or plastome coverage) and the relative size of the target and reference genomes (GLR, Fig. 1E and 1F). As percent identity increased, the number of contigs decreased (Fig. 1G). Finally, GLR was weakly and positively correlated with percent identity (Fig. 1H), indicating for taxa sharing sequence identity, reference and target genomes tended to be of similar sizes.

Quality assessment of plastome assembly in Poaceae. De novo assemblies resulted in similar percentage of plastome reads and depth of coverage as reference based methods (Table 1). *Oryza* and *Sorghum* resulted in a single contig from *de novo* methods, but lower depth of coverage across the plastome in *Zea* B73 yielded a large number of contigs.

Assembled sequences may differ from published plastomes because of sequencing/assembly error and/or natural variation in plant genomes. Large numbers of contigs preclude accurate comparisons between assemblies and reference genomes, especially in tests between reference sequences (Table 2), but there are several trends concerning the nature of sequence variation. Sequences of plastome assemblies were generally consistent regardless of the assembly method or reference sequence used. Variation in the number of single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms (indels) between assemblies accounted for less than 0.05% of the plastome (data not shown). Indels generally involved single nucleotides, except in the case of a few large indels in *Oryza*. In this case, we found that Illumina reads are too short to assemble over large indels (>50 bp) relative to reference sequences. SNPs indicated expected levels of variation within taxa relative to other published studies of intraspecific taxon variation in grasses [5].

Structural changes in the plastome between species can complicate sequence analysis, but results of reference-based assembly can reflect such rearrangements. Analysis of the *Typha* plastome indicates a number of rearrangements relative to Poaceae plastomes [14]. For all three test grasses, the number of contigs from assemblies using references within Poaceae ranged from one to 14. The number of contigs from assemblies using Typha as a reference, however, ranged from 22 to 59. While rearrangements are not the only reason for breakpoints in the assembly, here reflected by number of contigs, the sudden increase in the number of contigs suggests some structural differences.

mtDNA results in Poaceae. Mitochondrial gene assemblies returned a single contig for both genes in all three grass taxa except for atp1 in *Zea* B73 (Table 3). This result is not surprising given the frequency with which sections of the mitochondrial genome are transferred to the nuclear genome [28]. Top BLAST results for both genes in all three taxa were the same mitogenomic sequences as the reference, except for *Oryza*. In this case, the top BLAST match was *Oryza sativa* ssp. *indica*, while the target taxon was *O. s.* ssp. *japonica*. We interpret this result to mean the plant from which we isolated DNA contains the mitochondrial haplotype of *O. sativa* ssp. *indica*.

nrDNA results in Poaceae. Trimmed 18S ribosomal gene sequences were ~1675 bp in length; some references contained internal indels. The percentage of Illumina reads used to assemble 18S rDNA from the grass reference was below 0.4%, but average depth of coverage was very high (e.g., 1072.5X in *Zea* B73, Table 4). A single contig resulted from all YASRA assemblies of rDNA, except for *Sorghum* assembled with the *Dioscorea* reference. In this case, one of two resulting contigs appeared to be an artifact as the other contig was comparable to the other assemblies for that taxon. Assemblies for each grass taxon from different reference sequences were identical (contained no SNPs or indels). From the initial ~1675 bp reference, YASRA returned contigs ranging from 1889 (*Zea* B73 assembled with *Phoenix*) to 4147 bp (*Sorghum* assembled with *Dioscorea*). However, alignments between assemblies of each grass taxon revealed variation in their terminal portions. We posit that this variation is artifactual and occurs because of the high copy number of 18S rDNA in the nuclear genome; highly variable flanking regions represent problematic sequences to align without a reliable reference. Regardless, we were able to obtain the entire 18S rDNA gene

(ca. 1750 bp) from a truncated reference in all three grasses. In the case of *Sorghum*, we obtained a reliable assembly from all references spanning a great deal of the flanking regions as well (nearly 4000 bp).

Genome size in Asparagales. Genome sizes are represented as pg/2C, or mass of DNA in a diploid (somatic) cell. In Asparagales these values ranged from 1.3 pg/2C in Aphyllanthes to 50.9 pg/2C in Amaryllis; the average genome size for the 43 taxa for which data were available was 16.9 pg/2C (SD=±13.8).

Ct values in Asparagales. Our samples had a Ct value of 21.0 or below with the exception *Asparagus asparagoides* (Ct=24.1), as we were unable to obtain a DNA sample with a Ct value within the desirable range. The lowest Ct value for our samples was 14.2 in *Trichopetalum*, and the average Ct value was 17.5 (SD=±1.8).

Plastome assembly relationships with genome size and Ct value in Asparagales. For the 48 Asparagales taxa, the number of reads ranged from 1.28 million (*Agapanthus africanus*) to 6.86 million (*Brodiaea californica*, Table 6**5**. The percent of Illumina reads assembling into plastomes in Asparagales ranged from 0.51-10.55% (*Scadoxus* and *Asphodeline*, respectively), while average plastome depth of sequence ranged from 12.5-482.8X (*Eucharis* and *Cordyline*). For the 48 Asparagales taxa sampled, the average plastome coverage was 80X (SD=±75.9) and percentage of plastome reads averaged 3.8% (SD=±2.8).

Plastome coverage generally increased as percent plastome increased **(**Fig. 2A, power=1**)**, but we tested both genome size and Ct value against each variable for confirmation. Ct value was unrelated to genome size (Fig. 2B, power=0.47). Removing an outlier (*Asparagus asparagoides*, with a Ct value higher than our desired threshold) had

little impact on the relationship. As genome size increased, both percent plastome and plastome coverage decreased, although relationships were weak (Fig. 2C, power=0.59 and 2D, power=0.66). Finally, there was no correlation between Ct value and either percent plastome or plastome coverage (Fig 2E, power=0.73 and 2F, power=0.42). Our power to detect relationships between these variables is admittedly weak, especially given the samples are not completely independent (some clusters of phylogenetic relatedness).

Discussion

We used an easy, low-cost approach to sequencing plastomes from total genomic DNA by barcoding six taxa per Illumina lane. The resulting sequence data is a low-redundancy set of genome survey sequences (GSS) from which not only full plastome sequences, but also nrDNA and limited mitogenomic gene sequences, can be assembled using reference-based methods. We evaluated the efficacy of our assembly methods using six Poaceae taxa. We also tested whether these methods could provide similar quality data for another monocot lineage, order Asparagales. Our results indicate these methods yield sequence data from all three genomic partitions in plants, and we recommend appropriate quality-control measures for ensuring reliability of resulting data.

Taxon selection for GSS. Previous plastome sequencing from total genomic DNA highlighted the necessity of selecting particular taxa (and subsequent DNA extractions) based on genome size and relative amount of chloroplast in the DNA sample [here represented as Ct value, 5]. Our results suggest that these two criteria are not applicable in Illumina GSS; the percentage of total reads (and as a result, assembly coverage) from the

plastome is not dependent on either Ct value or genome size. Selection of taxa for Illumina GSS need not be constrained by genome size; genomic characteristics like ploidy level need not necessarily exclude a taxon from GSS. While larger genomes are generally thought to complicate plastome sequencing from total genomic DNA, our results agree with knowledge about cellular alterations that accompany genome size changes. Because cell size increases with genome size, the number of organelles per cell increases. Thus, the relative number of chloroplasts likely increases, too.

Furthermore, it is unnecessary to perform chloroplast isolations for such sequencing; total genomic DNA provides sufficient sequence data to assemble plastomes. Stochastic variation in library preparation resulted in some taxa with much deeper sequencing than expected. *Sorghum* sequencing, for example, generated 25% more reads than *Oryza*, and the robustness of sequence assembly reflects a higher depth of coverage (Table 1). Even taxa of the same species (e.g., *Zea mays* ssp. *Mays* accessions we sampled) vary widely in depth of sequencing, suggesting these differences may result from stochastic variation in library preparation. Proportion of plastome sequences in GSS also likely varies based on physiological differences between taxa (or inbred lines), as well as growing conditions. Finally, problematic assembly of the mitogenome due to its larger size indicates that size of the organellar genome itself can decrease overall depth of coverage. These complicating factors make sequencing of some taxa more difficult, but such concerns could be alleviated by decreasing the number of taxa per lane.

Sequence assembly of GSS. As the number and public availability of sequenced organellar and nuclear genomes increases, the task of assembling additional genomes is

simplified. Even if a genome is assembled *de novo*, comparison to a reference afterwards can target areas where mistakes in assembly may have occurred. Furthermore, genome assembly and annotation of any type is a continual process. Deeper sequencing, optimized parameters, and sequencing of additional accessions of the same species or closely related taxa can all illuminate novel features of a species' genome sequence.

Our results indicate that reference sequences from closely related taxa are not necessary to obtain at least partial sequence information from GSS. However, decreased similarity (and therefore, phylogenetic distance) can complicate attempts to assemble large contigs. Breakpoints in assemblies, illustrated by increased numbers of contigs, result from rearrangements relative to the reference sequence, as well as areas of decreased depth of sequencing coverage. While *de novo* assembly methods can alleviate the first issue, our results from *Zea* B73 plastome assemblies indicate that the second issue is exacerbated. We contend that reference-based assemblies are an appropriate application for systematic studies, because they capitalize on the nature of Illumina GSS to reliably construct coding regions useful in phylogenetic reconstructions.

Like any other sequencing method, Illumina technology inherently contains biases [29] and types of error [30] that can inhibit robust reconstructions of genomic sequences, especially in organisms with large genomes [31]. We present here different methods for *a priori* quality control for trimming reads, a variety of methods for sequence assembly, and ways to compare resulting assemblies. Most important are quality control measures to ensure the assemblies from any method are reliable, repeatable, and not artifacts of the

assembly process. Errors occur in all sequencing and assembly procedures, and checking for consistency of results is essential, especially when working in under-studied systems.

Finally, this paper presents results of assembly for plastome, mitogenome, and nuclear ribosomal sequences in plants, but these data still only account for, at most, 10% of Illumina GSS reads. The majority of reads are presumably from the nuclear genome, and further work should investigate the feasibility of assembling repetitive elements (REs) from these data. For example, deeper Illumina GSS sequences have been applied effectively in barley to characterize REs in a genome [32]. Further research should explore the the effectiveness of very low coverage GSS to recover REs in non-model systems, or where the RE compliment is unknown.

Applications. We have shown the feasibility of obtaining large amounts of both coding and non-coding DNA sequence data from three genomic compartments, which allows phylogenetic reconstruction between even problematic groups with recent divergence [33]. Our method of Illumina GSS is especially attractive for systematic studies, where large numbers of taxa and many genes are optimal for phylogeny estimation. Ideally, databases for plastomes, mitogenomes, and nuclear ribosomal repeats should be prioritized for systematists, as well as support for online tools that make assembly and annotation easier. Consolidation and standardization of these types of analysis will allow broader applications for both taxonomy and molecular evolution. Plastomes, for example, have potential as a single-locus DNA barcode for identification of plants [5], and we contend that mitogenomes and nuclear ribosomal loci have similar potential for confirming problematic taxa [27,34]. Similarly, mitogenomes may serve as a DNA barcode in animals and can be

gleaned from GSS in animals just as easily as plastomes in plants (Pires, J. C., unpub. data). Furthermore, a broader sampling of plastomes from across the plant kingdom will help inform the relevance and frequency of structural changes in organellar genomes and provide a framework for comparative biology of organellar evolution. The promise of mining lllumina GSS for plastome, mitogenomic, and ribosomal nuclear elements makes developing genomic tools across diverse organisms possible.

Acknowledgements

I would like to thank my collaborators and co-authors on the publication resulting from this chapter: Pamela R. Steele, Dustin Mayfield, and J. Chris Pires. This research was funded by the National Science Foundation (DEB 0829849).

Literature Cited

1. Green P (2007) 2x Genomes - Does depth matter? Genome Research 17: 1547-1549.

2. Wernersson R, Schierup MH, Jorgensen FG, Gorodkin J, Panitz F, et al. (2005) Pigs in sequence space: A 0.66X coverage pig genome survey based on shotgun sequencing. Bmc Genomics 6.

3. Rasmussen DA, Noor MAF (2009) What can you do with 0.1x genome coverage? A case study based on a genome survey of the scuttle fly Megaselia scalaris (Phoridae). Bmc Genomics 10.

4. Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in Drosophila: Diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genetics 5.

5. Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, et al. (2010) Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnology Journal.

6. Atherton R, McComish B, Shepherd L, Berry L, Albert N, et al. (2010) Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. Plant Methods 6: 22.

7. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, et al. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92: 142-166.

8. Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal Of Botany 94: 275-288.

9. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, et al. (2010) Insights into the Evolution of Mitochondrial Genome Size from Complete Sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). Molecular Biology and Evolution 27: 1436-1448.

10. Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD (2002) Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proceedings of the National Academy of Sciences, USA 99: 9905-9912.

11. Steele PR, Pires JC (2011) Biodiversity assessment: State-of-the-art techniques in phylogenomics and species identification. American Journal Of Botany 98: 415-425.

12. Rabinowicz PD, Bennetzen JL (2006) The maize genome as a model for efficient sequence analysis of large plant genomes. Current Opinion in Plant Biology 9: 149-156.

13. Bennett MD, Leitch IJ (2011) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Annals Of Botany 107: 467-590.

14. Guisinger M, Chumley T, Kuehl J, Boore J, Jansen R (2010) Implications of the Plastid Genome Sequence of *Typha* (Typhaceae, Poales) for Understanding Genome Evolution in Poaceae. Journal of Molecular Evolution 70: 149-166.

15. Pires JC, Maureira IJ, Givnish TJ, Sytsma KJ, Seberg O, et al. (2006) Phylogeny, genome size, and chromosome evolution of Asparagales. Aliso 22: 285-302.

16. Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9: 208-218.

17. Bennett MD, Leitch IJ (2010) Angiosperm DNA C-values database. http://www.kew.org/cvalues.

18. Ratan A (2009) Assembly algorithms for next generation sequence data. State College, PN: Pennsylvania State University.

19. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30: 3059 - 3066.

20. Kumar S, Tamura K, Nei M (1994) MEGA: Molecular evolutionary genetics analysis software for microcomputers. Computer Applications in the Biosciences 10: 189-191.

21. Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Research 9: 868-877.

22. Davis JI, Petersen G, Seberg O, Stevenson DW, Hardy CR, et al. (2006) Are mitochondrial genes useful for the analysis of monocot relationships? Taxon 55: 857-870.

23. Duminil J, Pemonge MH, Petit RJ (2002) A set of 35 consensus primer pairs amplifying genes and introns of plant mitochondrial DNA. Molecular Ecology Notes 2: 428-430.

24. McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucl Acids Res 32: W20-25.

25. Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, et al. (2010) Geneious. 5.3 ed.

26. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252-3255.

27. Steele PR, Hertweck KL, T.Docktor, Pires. JC (in prep) Molecular phylogenomics using massively parallel sequencing: an example in core Asparagales.

28. Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Molecular Phylogenetics and Evolution 29: 380-395.

29. Hansen KD, Brenner SE, Dudoit S (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Research 38: e131.

30. Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Research 36: e105.

31. Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. Genome Research 20: 1165-1173.

32. Wicker T, Narechania A, Sabot F, Stein J, Vu GTH, et al. (2008) Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. Bmc Genomics 9.

33. Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biology 7: 84.

34. Steele PR, Hertweck KL, Mayfield D, Pflug J, Pires JC (in prep) Species identification using evidence from total genomic data.

35. APGII (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Botanical Journal of the Linnean Society 141: 399.

36. APGIII (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal Of The Linnean Society 161: 105-121.

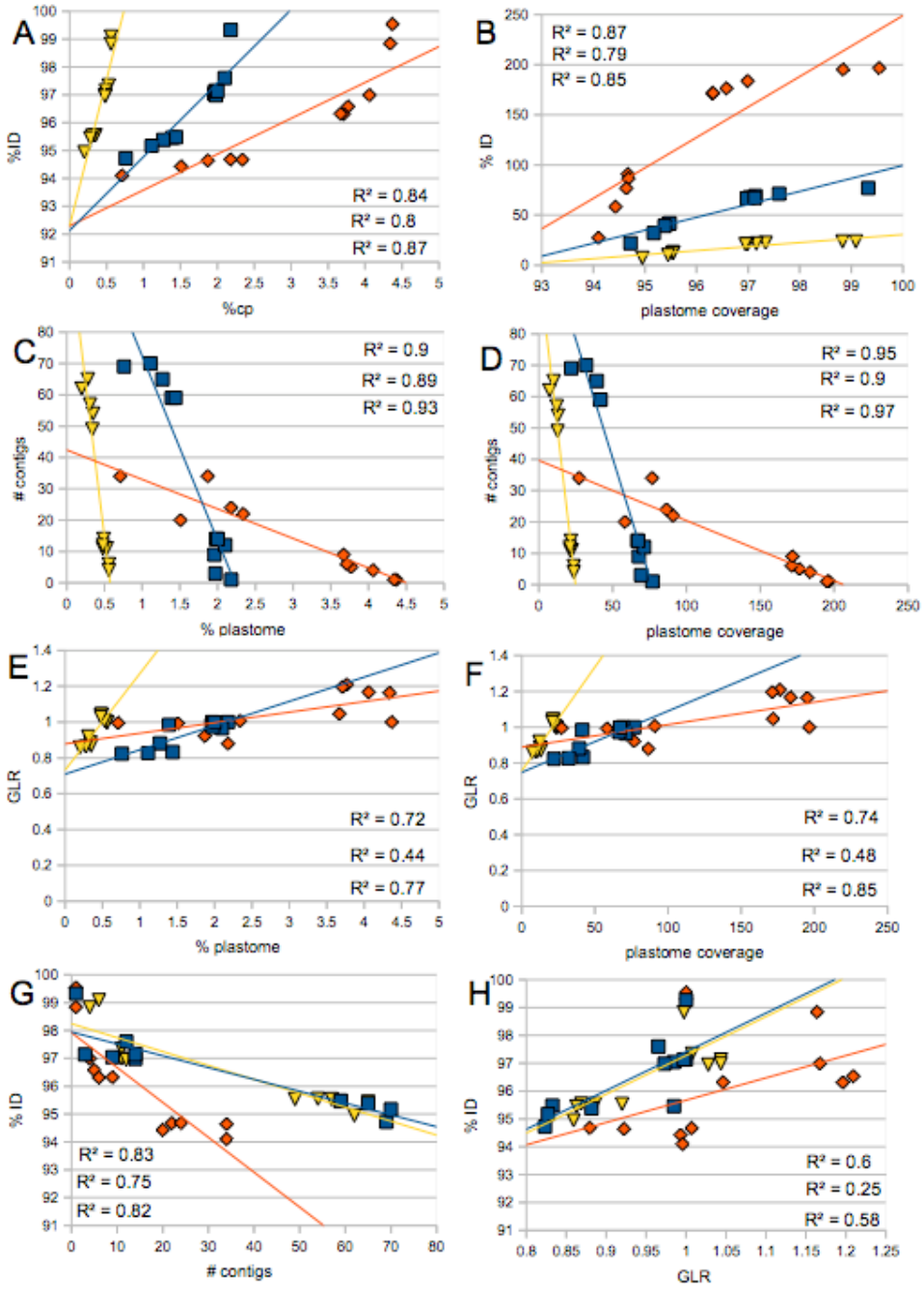
A
%ID
%cp
R² = 0.84
R² = 0.8
R² = 0.87
B
% ID
plastome coverage
R² = 0.87
R² = 0.79
R² = 0.85
C
contigs
% plastome
R² = 0.9
R² = 0.89
R² = 0.93
D
contigs
plastome coverage
R² = 0.95
R² = 0.9
R² = 0.97
E
GLR
% plastome
R² = 0.72
R² = 0.44
R² = 0.77
F
GLR
plastome coverage
R² = 0.74
R² = 0.48
R² = 0.85
G
% ID
# contigs
R² = 0.83
R² = 0.75
R² = 0.82
H
% ID
GLR
R² = 0.6
R² = 0.25
R² = 0.58

Figure 1. Effect of phylogenetic distance between target and reference taxa on plastome assembly in Poaceae. All relationships reported are linear. Blue is *Oryza*, red is *Sorghum*, and yellow is *Zea*. R^2 values are from *Oryza*, *Sorghum*, and *Zea* listed from top to bottom.

A. Percentage of Illumina reads from the plastome and percent identity between reference and target genomes.

B. Average depth of coverage in plastome assembly and percent identity between reference and target genomes.

C. Percentage of Illumina reads from the plastome and number of contigs resulting from first YASRA assembly.

D. Average depth of coverage in plastome assembly and number of contigs resulting from first YASRA assembly.

E. Percentage of Illumina reads from the plastome and ratio of target to reference genome length.

F. Average depth of coverage in plastome assembly and ratio of target to reference genome length.

G. Number of contigs resulting from first YASRA assembly and percent identity between reference and target genomes.

H. Ratio of target to reference genome length and percent identity between reference and target genomes.

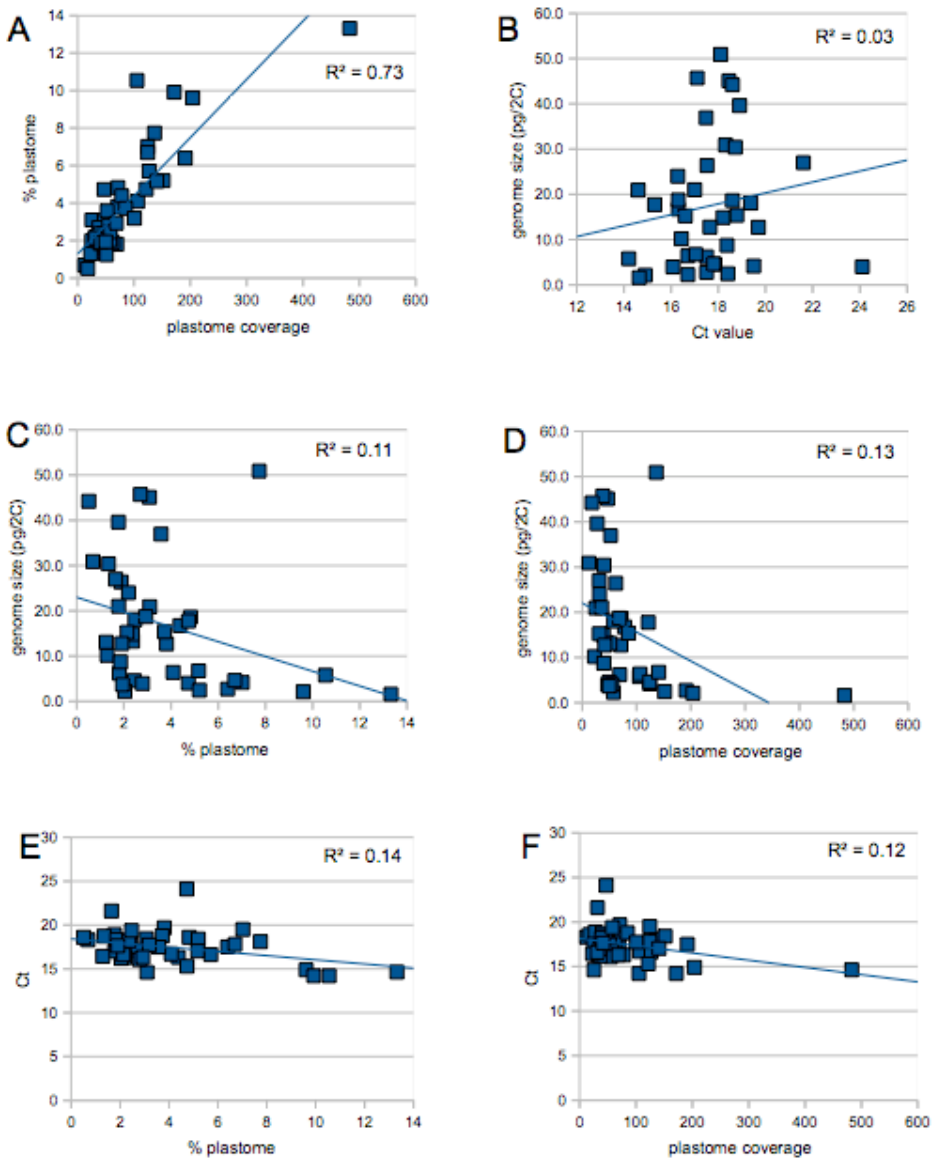
A
R² = 0.73
% plastome
plastome coverage
B
R² = 0.03
genome size (pg/2C)
Ct value
C
R² = 0.11
genome size (pg/2C)
% plastome
D
R² = 0.13
genome size (pg/2C)
plastome coverage
E
R² = 0.14
Ct
% plastome
F
R² = 0.12
Ct
plastome coverage

Figure 2. Effect of Ct value and genome size on plastome assembly in Asparagales.

A. Average depth of coverage in plastome assembly and percentage of Illumina reads from the plastome; removal of (*Cordyline australis*) does not change relationship (R^2=0.72).

B. Ct value (and genome size; power; removal of outlier (*Asparagus asparagoides*) does not change strength of relationship (R^2=0.09).

C. Percentage of Illumina reads from the plastome and genome size; power; removal of outlier (*Amaryllis belladona*) slightly strengthens the relationship (R^2=0.32).

D. Average depth of coverage in plastome assembly and genome size; power, removal of outliers (*Cordyline australis* and *Amaryllis belladona*) slightly strengthens the relationship (R^2=0.4)

E. Percentage of Illumina reads from the plastome and Ct value, removal of outlier (*Asparagus asparagoides*) strengthens the realtionship (R^2=0.25).

F. Average depth of coverage in plastome assembly and Ct value; removal of outlier (*Cordyline australis*) decreases the strength of the relationship (R^2=0.08).

Table 1. Summary information for Poaceae taxa used in this study and both reference-based and *de novo* plastome assemblies. All reads are 120 bp single-end.

Taxon (voucher)	Abbreviation	Number of reads	% plastome (coverage)	Reference (Genbank Accession)
Oryza sativa ssp. *japonica* cv. Nipponbarre	*Oryza*	4095296	2.18 (76.9X)	*Oryza sativa* ssp. *japonica* (X15901.1)
			2.29 (65X)	*de novo*, 1 contig
Sorghum bicolor cv. B Tx642	*Sorghum*	5457273	4.37 (196.5)	*Sorghum bicolor* (EF115542.1)
			4.41 (177X)	*de novo*, 1 contig
Zea mays ssp. *mays* cv. B73	*Zea* B73	5158725	0.56 (23.7X)	*Zea mays* (X86563.2)
			0.53 (27X)	*de novo*, 97 contigs
Zea mays ssp. *mays* va. CIMMYT Maize Inbred Line 52	*Zea* CML52	1820080	0.98 (14.6X)	*Zea mays* (X86563.2)
Zea mays ssp. *mexicana*	*Z. m. mexicana*	4707250	2.11 (82.1X)	*Zea mays* (X86563.2)
Zea mays ssp. *parviglumis*	*Z. m. parviglumis*	4917582	0.94 (38X)	*Zea mays* (X86563.2)

Table 2. Effect of reference sequence on assembly quality for three target Poaceae taxa. All reads are 120 bp single-end.

		Oryza (Ehrhartoideae)				**Sorghum (**Panicoideae)				**Zea (**Panicoideae)			
Reference taxon Genbank Accession		**% plastome (coverage)**	**% identity**	**GLR**	**# contigs**	**% plastome (coverage)**	**% identity**	**GLR**	**# contigs**	**% plastome (coverage)**	**% identity**	**GLR**	**# contigs**
Poaceae (Ehrhartoideae)	***Oryza*** X15901.1	2.18 (76.9X)	99.27	1	1	3.75 (175.8X)	96.53	1.21	5	0.49 (21.7X)	97.13	1.04	14
Poaceae (Pooidaea)	***Triticum*** AB042240.3	1.97 (69.2X)	97.14	1	3	3.67 (171.8X)	96.32	1.05	9	0.48 (21.2X)	96.99	1.04	11
Poaceae (Aristidoideae)	***Agrostis*** EF115543.1	1.96 (67.7X)	97.04	0.98	9	3.71 (171.X)	96.31	1.2	6	0.48 (21.1)	96.95	1.03	12
Poaceae (Bambusoideae)	***Bambusa*** FJ970915.1	2.1 (71.3X)	97.6	0.97	12	4.06 (183.7X)	96.99	1.17	4	0.53 (22.4)	97.34	1.01	11
Poaceae (Panicoideae)	***Zea*** X86563.2	1.98 (66.7X)	96.98	0.97	14	4.34 (195.2X)	98.84	1	1	0.56 (23.7X)	99.09	1	6
Poaceae (Panicoideae)	***Sorghum*** EF115542.1	2 (67.2X)	97.13	1	14	4.37 (196.5X)	99.54	1	1	0.56 (23.7X)	98.83	1	4
Typhaceae	***Typha*** NC013823	1.44 (41.9X)	95.43	0.83	59	2.34 (90.9X)	94.67	1.01	22	0.35 (13X)	95.55	0.87	54
Arecales	***Phoenix*** GU811709.2	1.39 (41.4X)	95.46	0.98	59	2.18 (86.5X)	94.68	0.88	24	0.34 (12.7X)	95.54	0.89	49
Dioscoreales	***Dioscorea*** EF380353.1	1.27 (39.3X)	95.38	0.88	65	1.87 (76.8X)	94.64	0.92	34	0.31 (12.2X)	95.54	0.92	57
Amborellales	***Amborella*** AJ506156.2	1.11 (32.1X)	95.17	0.83	70	1.51 (58.3X)	94.43	0.99	20	0.28 (10.1X)	95.45	0.86	65
Cycads	***Cycas*** AP009339.1	0.76 (22X)	94.72	0.82	69	0.71 (27.3X)	94.1	1	34	0.2 (7.3X)	94.95	0.86	62

Table 3. Mitochondrial gene assembly in Poaceae using YASRA.

	Reads in contigs	coverage	# contigs	% identity	Genbank mitogenome, bases for gene
atp1					
Oryza	216	16.1X	1	99.25	NC_011033.1, 352379-353908
Sorghum	283	21.1X	1	99.57	NC_008360.1, 13551-15092
Zea B73	82	6.2X	2	99.1	AY506529.1, 454351-455877
cox3					
Oryza	42	5.8X	1	99.04	NC_011033.1, 17226-18068
Sorghum	72	10.3X	1	99.61	NC_008360.1, 119088-119885
Zea B73	8	1.1X	1	98.78	AY506529.1, 441570-442367

Table 4. Nuclear ribosomal DNA sequences (nrDNA) assembled with *Zea mays* 18S small subunit ribosomal RNA reference sequence (AF168884.1) 1670 bp in length. All assemblies resulted in a single contig.

Target taxon	% reads	coverage	% identity	Consistent assembly length
Oryza	0.39	1120.7X	97.83	1720
Sorghum bicolor	0.22	842.8X	97.88	3665
Zea B73	0.33	1170.2X	98.67	2722
Z. m. CML52	0.31	392.1X	98.74	1923
Z. m. mexicana	0.24	797.9X	98.68	1909
Z. m. parviglumis	0.12	422.2X	98.68	1745

Table 5. Summary information for Asparagales taxa used in this study. Voucher and GenBank accession numbers are available in Steele et. al [34]. Family assignations noted are from APGIII/APGII[35,36]. Genome size notations: *average, #previously published, ^taxon substituted. Number of reads notations: & 2-pass Illumina run.

Lineage	Taxon	Genome size (pg/2C)	Ct value	Length of reads	Number of reads	% plastome (coverage)
Asparagales (Amaryllidaceae/Agapanthaceae)	*Agapanthus africanus*	20.95	14.6	80	1281941&	3.1 (25.3X)
Asparagales (Asparagaceae/Agavaceae)	*Anemarrhena asphodeloides*	6.21	17.5	80	6425759	1.82 (69.3X)
Asparagales (Asparagaceae/Agavaceae)	*Echeandia sp.*	18.63	18.6	80	2368193	4.83 (71.3X)
Asparagales (Asparagaceae/Agavaceae)	*Manfreda virginica*	12.71	19.7	80	3055209	3.81 (71.3X)
Asparagales (Asparagaceae/Agavaceae)	*Polianthes* sp.	4.58*#	17.85	80	3274771	2.44 (49.4X)
Asparagales (Asparagaceae/Alliaceae)	*Allium cepa*	16.8#	16.3	80	2795386	4.38 (78.2X)
Asparagales (Asparagaceae/Alliaceae)	*Allium fistulosum*	26.4	17.52	80	0	1.89 (62.1X)
Asparagales (Amaryllidaceae/Alliaceae)	*Gillesia graminea*	N/A	17.25	80	2915826	1.91 (35.5X)
Asparagales (Asparagaceae/Alliaceae)	*Tulbaghia violacea*	45.1	18.45	80	2381172	3.08 (46.8X)
Asparagales (Amaryllidaceae)	*Amaryllis belladonna*	50.9*	18.1	80	2972595	7.47 (136.3)
Asparagales (Amaryllidaceae)	*Crinum asiaticum*	45.7*	17.1	80	2230364	2.7 (38.1X)
Asparagales (Amaryllidaceae)	*Eucharis grandiflora*	30.9*	18.3	80	2745718	0.69 (11.7X)
Asparagales (Amaryllidaceae)	*Scadoxus cinabaria*	44.2#^	18.6	120	5942909	0.51 (18X)

Lineage	Taxon	Genome size (pg/2C)	Ct value	Length of reads	Number of reads	% plastome (coverage)
Asparagales (Asparagaceae)	*Asparagus officinalis*	2.73	17.5	120	4996275	6.37 (190.3X)
Asparagales (Asparagaceae)	*Hemiphylacus alatostylus*	4.18#	17.5	80	2876326	7.02 (124.5X)
Asparagales (Xanthorrhoeaceae/Asphodelaceae)	*Aloe vera*	39.6	18.9	80	2451314	1.77 (27.7X)
Asparagales (Xanthorrhoeaceae/Asphodelaceae)	*Asphodeline damascena*	5.8*	14.2	80	1608643	10.55 (105.3X)
Asparagales (Xanthorrhoeaceae/Asphodelaceae)	*Kniphofia linearfolia*	27	21.6	80	3078437	1.65 (31.6X)
Asparagales (Xanthorrhoeaceae/Hemerocallidaceae)	*Doryanthes palmeri*	6.4	16.7	120	4446830	4.09 (106.5X)
Asparagales (Xanthorrhoeaceae/Hemerocallidaceae)	*Geitonoplesium cymosum*	N/A	16.63	80	3644530	5.71 (126.9X)
Asparagales (Xanthorrhoeaceae/Hemerocallidaceae)	*Phormium tenax*	2.1	14.9	80	3424451	9.61 (202.9X)
Asparagales (Asparagaceae/Hyacinthaceae)	*Bowiea volubilis*	4.6#	17.8	80	2965244	6.7 (124.3X)
Asparagales (Asparagaceae/Hyacinthaceae)	*Drimia altissima*	15.4*	18.78	80	3670644	3.72 (84.9X)
Asparagales (Asparagaceae/Hyacinthaceae)	*Ledebouria* cf. *cordifolia*	17.7	15.3	80	4137538	4.74 (121.8X)
Asparagales (Asparagaceae/Hyacinthaceae)	*Ornithogalum tenuifolium*	36.9	17.48	80	2374018&	3.58 (52.4X)
Asparagales (Asparagaceae/Hyacinthaceae)	*Oziroe biflora*	N/A	17.5	80	1996258	1.99 (25X)
Asparagales (Iridaceae)	*Iris tenax*	N/A	17.74	80	4917819	3.19 (100.7X)
Asparagales (Asparagaceae/Laxmanniaceae)	*Lomandra longifolium*	2.3	16.7	80	4465309&	2.04 (57.6X)

Lineage	Taxon	Genome size (pg/2C)	Ct value	Length of reads	Number of reads	% plastome (coverage)
Asparagales (Asparagaceae/Laxmanniaceae)	*Trichopetalum plumosum*	N/A	14.2	80	2753011	9.92 (171.4X)
Asparagales (Asparagaceae/Ruscaceae)	*Calibanus hookeri*	24	16.28	80	2417131	2.2 (31.9X)
Asparagales (Asparagaceae/Ruscaceae)	*Dasylirion wheeleri*	4	16.07	80	3116974	2.79 (55X)
Asparagales (Asparagaceae/Ruscaceae)	*Eriospermum cervicorne*	N/A	16.2	120	3037618	2.05 (37.2X)
Asparagales (Asparagaceae/Ruscaceae)	*Liriope spicata*	21	17	120	3321934	1.78 (35.5X)
Asparagales (Asparagaceae/Ruscaceae)	*Ophiopogon japonicus*	10.2	16.43	80	2942473	1.29 (22.9X)
Asparagales (Asparagaceae/Ruscaceae)	*Ruscus aculeata*	8.8#^	18.37	80	3352547	1.86 (39.7X)
Asparagales (Asparagaceae/Ruscaceae)	*Sanseveria trifasciata*	2.5	18.4	120	4865400	5.1 (148.4X)
Asparagales (Asparagaceae/Ruscaceae)	*Smilacina stellata*	13.3#	N/A	120	3171872	2.37 (45.1X)
Asparagales (Asparagaceae/Themidaceae)	*Androstephium caeruleum*	14.9	18.2	80	2633504	2.36 (39X)
Asparagales (Asparagaceae/Themidaceae)	*Dichelostemma capitatum*	18.1	19.37	120	3915145	2.47 (58.2X)
Asparagales (Asparagaceae/Themidaceae)	*Dichelostemma congestum*	15.3*	16.6	120	2492563	2.14 (31.6X)
Asparagales (Asparagaceae/Themidaceae)	*Dichelostemma ida-maia*	18.7*	16.3	120	3933031	2.93 (68.9X)
Asparagales (Asparagaceae/Themidaceae)	*Tritileia hyacinthia*	12.8	17.64	120	3559280	1.91 (41.5X)
Asparagales (Xanthorrhoeaceae)	*Xeronema callistemon*	6.8	17.04	120	4506941	5.17 (140.8X)

SUPPLEMENTAL METHODS

Genome sizing

Flow cytometric procedures to estimate nuclear DNA content in plant cells was modified from Arumuganathan and Earle (1991). Values for nuclear DNA content were estimated by comparing fluorescence intensities of the nuclei of the test population with those of an appropriate internal standard. We used chicken red blood cells (CRBC, 2.5 pg/2C) or *Nicotiana tabacum* (ca. 8.4 pg/2C, calibrated from CRBC for each sample) as the internal standard for small and large genomes, respectively. Fifty milligrams of fresh leaf tissue was placed on ice in a sterile 35 x 10 mm plastric petri dish and was sliced into 0.25 mm to 1 mm segments in a solution containing 10 mM MgSO4.7H2O, 50mM KCl, 5 mM Hepes, pH 8.0, 3 mM dithiothreitol, 0.1 mg/mL propidium iodide, 1.5 mg/mL DNAse free RNAse (Roche, Indianapolis, IN) and 0.25% Triton X-100. Suspended nuclei were withdrawn using a pipettor, filtered through 30-µm nylon mesh, and incubated at 37 °C for 30 min. Suspensions of sample nuclei were spiked with suspension of standard nuclei (prepared in above solution) and analyzed with a FACScalibur flow cytometer (Becton-Dickinson, San Jose, CA). For each sample, propidium iodide fluorescence area signals (FL2-A) from 1000 nuclei were collected and analyzed by CellQuest Pro software (Becton-Dickinson, San Jose, CA). The mean position of the G0/G1 (Nuclei) peak of the sample and the internal standard were compared and the mean nuclear DNA content of each sample was reported as mass per diploid (somatic) cell (pg/2C).

RT-PCR to obtain Ct values

We estimated Ct values using real-time PCR (RT-PCR) and quantified the presence of the plastid locus rbcL using Fermentas Maxima SYBR Green qPCR Master Mix with an Asparagus BAC isolation as the positive control and standard. We performed 20 uL reactions (8 uL ddH2O, 10 uL SYBR green mastermix, 0.5 uL of each primer [rbcL-F: TGG CAG CAT TYC GAG TAA CT, rbcL-R: ACG ATC AAG RCT GGT AAG TC], and 1 uL of DNA at 2.5 ng/uL) and ran them in an Opticon Monitor3 (Bio-Rad Laboratories) using the following parameters: 50C for 2 min, 95C for 10 min, and 45 cycles of 95C for 15 sec, 58C for 15 sec, 68C for 20 sec. The melting curve read every 0.2C from 72 to 95C. We exported our resulting data from the Opticon Monitor3 software into LinRegPCR v11.3 [1] to calculate the Ct threshold using our standard (control value=12.0). We input these results back into Opticon Monitor3 to calculate the standardized Ct values for our samples.

Library preparation for Illumina sequencing

Shearing genomic DNA. We prepared total genomic DNA for Illumina sequencing by sonicating 5 ug (diluted to 6.25 ng/uL) in a Bioruptor for 24 minutes, inverting the tubes at 12 minutes. We purified using QIAquick PCR purification kits (Qiagen) and eluted with 37.5 uL EB buffer + 37.5 uL ddH2O in the final step. We ran 200ng of the sheared DNA on an agarose gel to verify shearing. We prepared libraries for Illumina sequencing using NEBNext© DNA Sample Prep Kits for Illumina (New England BioLabs); all reagents that follow are a part of this kit.

End repair. We performed end repair at 100 uL volume (75 uL eluted DNA, 10 uL phosphorylation buffer, 4 uL dNTP mix, 5 uL T4 DNA polymerase, 1 uL Klenow DNA polymerase, and 5 uL T4 PNK), incubated these reactions at 20C for 30 minutes, and purified with QIAquick PCR Purification kits (32 uL EB buffer for final elution).

Adapter ligation. We prepared fragments for adapter ligation using a total reaction volume of 50 uL (32 uL eluted DNA, 5 uL NEBuffer 2, 10 uL dATP, 3 uL 3' to 5' exo-Klenow) and incubating for 30 minutes at 37C. We purified these reactions with a Qiagen MinElute PCR Purification kit and eluted to 10 uL. We ligated adapters to fragments in a 50 uL reaction (10 uL eluted DNA, 25 uL 2X Quick ligation buffer, 10 uL adapter/water mix, 5 uL Quick T4 DNA ligase) and incubated at room temperature for 20 minutes followed by purification (QIAquick PCR Purification, elute with 20 uL EB buffer). We ran these reactions on a 2% low-melt gel (100 bp ladder for comparison) and excised 300 bp products for purification (QIAquick Gel Extraction, elute with 30 uL EB buffer).

Enrich fragments. We enriched the selected fragments in duplicate for each sample using 50 uL PCR reactions (3 uL ligation DNA, 20 uL H2O, 25 uL Phusion Flash High Fidelity PCR 2x mastermix, and 1 uL each of enrichment primers at 25uM; PCR parameters: 98C for 30 sec, 15 cycles of [98C for 10 sec, 65C for 30 sec, 72C for 30C], 72C for 5 min). We combined duplicate reactions for each sample prior to purification (QIAquick PCR Purification, elute with 20 uL EB buffer). We ran all products on 2% low-melt gel (100 bp ladder), excised all products, and purified (QIAquick Gel Extraction, elute with 30 uL EB buffer).

Adapter tags (12 pairs)

AD1_ACGT /5Phos/CGT AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_ACGT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ACG* T
AD1_CGTT /5Phos/ACG AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_CGTT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CGT* T
AD1_GTAT /5Phos/TAC AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_GTAT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GTA* T
AD1_TACT /5Phos/GTA AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_TACT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TAC* T
AD1_AGCT /5Phos/GCT AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_AGCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT AGC* T
AD1_CTGT /5Phos/CAG AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_CTGT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CTG* T
AD1_GATT /5Phos/ATC AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_GATT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GAT* T
AD1_TCAT /5Phos/TGA AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_TCAT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TCA* T
AD1_GCTT /5Phos/AGC AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_GCTT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GCT* T
AD1_TGCT /5Phos/GCA AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_TGCT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT TGC* T
AD1_CACT /5Phos/GTG AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_CACT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT CAC* T
AD1_ATGT /5Phos/CAT AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
AD2_ATGT ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT ATG* T

Enrichment primers

PCR 1: AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC* T
PCR 2: CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC* T

De novo sequence assembly

We quality trimmed Illumina sequences based on Phred quality scores included in FASTQ format. For each read, the median score threshold was >=20, the maximum number of uncalled bases was <=3, the minimum bases called were >=25, and the read was trimmed

when >=3 bases had phred scores <=16. We used NextGENe software (Softgenetics, State College, PA, USA) for *de novo* assembly using 5 cycles of condensation (see parameters below). NextGENe uses the maximum overlap for Illumina data. We further assembled sequences longer than 61bp using CAP3 [2] using the following parameters: (-a 20 -b 20 -c 12 -d 200 -e 30 -f 11 -g 6 -h 100 -m 2 -n -5 -o 60 -p 98 -r 1 -s 900 -u 3 -v 2 -y 250 -z 3). We screened contigs screened for sequence similarity to previously published plastid genomes (Table 1) with nucleotide BLAST [BLASTn,3] using default parameters. Contigs that had high similarity were truncated on each end by 200bp, and we mapped the original Illumina reads (see parameters below) to these truncated contigs. The unmatched reads were used to help extend the contigs of interest in another round of de novo assembly. This process of de novo assembly of unmatched reads, followed by further assembly with CAP3 was continued until contig length failed to increase. We aligned contigs to reference genomes for comparison purposes using Geneious v5.3.4 [4].

NextGENe Mapping Parameters:

Alignment:
Matching Requirement: >=40 Bases and >=97%
Do not check “Allow Ambiguous Mapping,” “Remove Ambiguously Mapped Reads,” “Detect Large Indels,” or “Rigorous Alignment”
Sample Trim:
Do not check “Select Sequence Range” or “Hide Unmatched Ends”
Mutation Filter:
Mutation Percentage<=0
SNP Allele <=0 Counts
Coverage <=0
Do not check “Use Original,” “Allow Software to Delete Mutations,” or “Delete 1bp Indels”
File Type:
Do not check “Load Assembled Result Files” or “Load Paired Reads”

Do not check “Save Matched Reads,” “Highlight Anchor Sequence,” or “Detect Structural Variations”

NextGENe Condensation Parameters:

Cycle1:
Minimum Read Length for Condensation: 56
Range in Read to Index: 1
Bases to Length minus 16 Bases
Reads Required for Each Group in One Direction: 3-60000
Reads Required for Each Group in Each Direction: 2-60000
Bridge Reads Required for Each Subgroup: 3 and 1
Total Reads Required for Each Subgroup: 5 and 0.2
Flexible Sequence Length: 18,16,14
Start Index at 3 Homopolymers
Check “AT,GC,ATT,… Complements”
Remove Low Quality Ends when Score <=10
Cycle2:
Minimum Read Length for Condensation: 56
Range in Read to Index: 6
Bases to Length minus 6 Bases
Reads Required for Each Group in One Direction: 5-60000
Reads Required for Each Group in Each Direction: -1-60000
Bridge Reads Required for Each Subgroup: -1 and -1
Total Reads Required for Each Subgroup: 5 and 0.2
Flexible Sequence Length: 20,18,16
Start Index at 3 Homopolymers
Check “AT,GC,ATT,… Complements”
Remove Low Quality Ends when Score <=10
Require Bridge Read Covering Middle 70%
Cycle3:
Minimum Read Length for Condensation: 56
Range in Read to Index: 6
Bases to Length minus 6 Bases
Reads Required for Each Group in One Direction: 5-60000
Reads Required for Each Group in Each Direction: -1-60000
Bridge Reads Required for Each Subgroup: -1 and -1
Total Reads Required for Each Subgroup: 5 and 0.2
Flexible Sequence Length: 22,20,18
Start Index at 3 Homopolymers
Check “AT,GC,ATT,… Complements”
Remove Low Quality Ends when Score <=10
Require Bridge Read Covering Middle 70%

Cycle4:
Minimum Read Length for Condensation: 56
Range in Read to Index: 6
Bases to Length minus 6 Bases
Reads Required for Each Group in One Direction: 5-60000
Reads Required for Each Group in Each Direction: -1-60000
Bridge Reads Required for Each Subgroup: -1 and -1
Total Reads Required for Each Subgroup: 5 and 0.2
Flexible Sequence Length: 24,22,20
Start Index at 3 Homopolymers
Check “AT,GC,ATT,… Complements”
Remove Low Quality Ends when Score <=10
Require Bridge Read Covering Middle 70%
Cycle5:
Minimum Read Length for Condensation: 56
Range in Read to Index: 6
Bases to Length minus 6 Bases
Reads Required for Each Group in One Direction: 5-60000
Reads Required for Each Group in Each Direction: -1-60000
Bridge Reads Required for Each Subgroup: -1 and -1
Total Reads Required for Each Subgroup: 5 and 0.2
Flexible Sequence Length: 26,24,22
Start Index at 3 Homopolymers
Check “AT,GC,ATT,… Complements”
Remove Low Quality Ends when Score <=10
Require Bridge Read Covering Middle 70%

Supplemental References

1. Ramakers C, Ruijter JM, Deprez RHL, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neuroscience Letters 339: 62-66.
2. Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Research 9: 868-877.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.
4. Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, et al. (2010) Geneious. 5.3 ed.

CHAPTER 5

CONCLUSION

The preceding chapters span the breadth of methodological and theoretical issues relevant to evolutionary analysis. Methodologically, the vignettes differ according to taxonomic level; Chapter 2 evaluates patterns across monocots, Chapter 3 analyzes effects within orders, and Chapter 4 describes relationships among species and genera. Moreover, the type of data used in each chapter of my dissertation varies. For Chapter 2 I sampled sequence data from across all three genomes and combined these with fossil and species number data to infer patterns. For Chapter 3 I used relatively little sequence data, but organismal characteristics (life history, biogeography, and genome size) to evaluate patterns of diversification. Finally, in Chapter 4 I constructed whole plastomes, as well as smaller sets of sequences from across the other genomic partitions, and evaluated these data in the context of genome sizes and the monocot phylogeny. Each of these taxonomic levels and types of data carry concomitant types of error. The deep divergence and fossil data of Chapter 2 generates relatively large confidence intervals despite a well resolved phylogeny. Chapter 3 highlights problems associated with low levels of divergence between taxa, and Chapter 4 suggests the amount of variation possible in molecular data from sequence and assembly error.

These methodological issues emphasize the importance of how we manage data and interpret results, especially given the convergence of biological themes in the theory behind each chapter. Life history traits, for example, are relevant to both chapters 2 and 3. While these characters are explicitly incorporated into analysis for the *Tradescantia* alliance, the monocot-wide phylogeny requires some knowledge of the herbaceous life history of monocots to interpret correctly (see Chapter 2). Similarly, organismal diversification is a theme for both chapters 2 and 3. In this case, Chapter 2 directly evaluates diversification rates across the monocot phylogeny, but Chapter 3 addresses the theme in the context of trait evolution. Genome size is a vital component in evolutionary analysis in Chapter 3, but is also necessary to develop sequencing methods in Chapter 4. Finally, all chapters require some knowledge of molecular evolution, although the breadth and depth of information required varies greatly. While molecular models are used to infer evolutionary rates, and thus phylogenies, for each analysis, Chapter 4 requires a deeper understanding of sequence structure and evolution.

In a broader sense, this dissertation exploits both historical and cutting-edge research methods in evolutionary biology. The systematic treatment of the *Tradescantia* alliance (Chapter 3) touches on classical molecular systematics, in which a phylogeny is used as a tool to evaluate taxonomic classification. The ancestral character state and correlational analyses begin to explore some of the *a posteriori* uses of phylogenetic trees, but the primary goal of the paper is to inform classification and taxon sampling primarily accommodates this goal. The monocot diversification analysis (Chapter 2) has a foundation in the same questions about classification. However, methodical taxon sampling allows

more elegant analyses modeling evolution across the clade, and provides the context of divergence times to ask additional evolutionary questions. The Illumina methods development analysis (Chapter 4) represents the cutting edge of evolutionary biology research, as it proposes the sampling of entire genomes for many taxa. The availability of such data will revolutionize our ability to test questions related to evolutionary rates, processes of character evolution, and organismal diversification.

Phylogenetics is the backbone of evolutionary biology. Leaves are being placed on the tree of life at an increasingly rapid rate, and observational systematics is gradually being overshadowed by hypothesis-driven research exploring processes of evolution. The three approaches of my dissertation research begin to address the two broad questions about plant diversification I highlighted in the introduction (Chapter 1). First, what is the historical context for evolution of particular plant lineages? Chapter 2 suggests that major monocot lineages diversified in the late Cretaceous, near the same time as the eudicot lineages that would eventually form angiosperm-dominated forests. Several monocot orders continued to diversify with animal lineages relevant to their pollination and dispersal mechanisms. These broad scale patterns in diversification are relevant to the shared characteristics of monocots, which occur in prairies and understories of forests. Chapter 3 highlights characteristics of a smaller group of monocots. The *Tradescantia* alliance exhibits morphological and life history lability that allowed species to diversify into new habits and geographic areas. Ancestral reconstructions suggest they were introduced into South America and dispersed northward, adapting characteristics suitable for northern climates (e.g., an erect habit which can inhabit edges of prairies and forests more easily than a

creeping habit). Both of these chapters indicate the life history of monocots is especially imporant in shaping their evolutionary history. Second, how do genomic characterisitics affect plant evolution and adaptation? I attempt in Chapter 3 to find a relationship between genome size and biogeographic spread in the *Tradescantia* alliance, but detect no correlation. Similarly, Chapter 4 relates how conserved organellar genomes are across the order Asparagales, and that nuclear genome size does not affect cell composition to the same extent as expected. Contrary to my expectations, my research does not support plant diversification as a result of genome-wide changes.

In the future, I am particularly interested in pursuing the intersection between genomic and organismal evolution. Availability of genomic sequences from a wide variety of taxa reveal intriguing patterns in genomic evolution, including gene content and chromosomal structure. One of the most striking and variable contrasts between genomes arises when comparing the repetitive element compliment of genomes. A large proportion of eukaryotic genomes is comprised of widely variable but repetitive centromeric, telomeric, and transposable elements (TEs). Evidence from several evolutionary lineages suggests TEs contribute to changes in genome structure and function by altering genome size, gene expression and the rate and placement of recombination. These genomic changes, in turn, result in corresponding changes to morphology and life history traits. Knowledge gained from both systematic and genomic science are reaching a critical point at which such relationships can be explicitly tested, and perhaps even experimentally manipulated. I hope to capitalize on the convergence of these themes, and provide a

synthetic mindset to fuse the theoretical foundation of both organismal and genomic science.

VITA

Kate Hertweck was born on September 4, 1983 in Evansville, Indiana to John and Judy Hertweck. She lived in southern Indiana until after graduation from F. J. Reitz High School in 2001, after which she attended Western Kentucky University in Bowling Green, Kentucky. She initially intended to minor in biology, but switched to a major after beginning undergraduate research in Dr. Larry Alice's molecular systematics lab. Attendance as a summer undergraduate intern at Savannah River Ecology Lab near Aiken, South Carolina expanded her research experience with work on sexual selection in salamanders with Dean Croshaw and Dr. Travis Glenn. Kate graduated from WKU in May 2005 with additional minors in communication and history; her senior honors thesis involved molecular systematics in mints.

Participation in undergraduate research allowed Kate to attend and present her research at numerous regional and national conferences during her time at WKU. She met J. Chris Pires, who would become her mentor for graduate research, at the Evolution conference in summer 2004, and began attending the University of Missouri as his first graduate student in August 2005. She completed her dissertation and graduated in summer 2011, and plans to pursue academic research in genomics.