TRANSPOSABLE ELEMENT CONTRIBUTION AND BIOLOGICAL CONSEQUENCE OF GENOME SIZE VARIATION AMONG WILD SUNFLOWER SPECIES

by

HANNAH M. TETREAULT

B.S., University of Maine-Machias, 2006

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

Nuclear genome size varies immensely across flowering plants, spanning nearly 2400-fold. The causes and consequences of this vast amount of variation have intrigued biologists since it became clear that nuclear DNA amount did not reflect organismal complexity (the so-called C-value paradox). In my dissertation I utilize wild sunflower species in the genus *Helianthus* to examine the role of transposable elements (TEs), and in particular, long terminal repeat (LTR) retrotransposons, in generating genome size variation and whether variation in genome size influences aspects of plant growth and development across multiple phenotypic levels. The genus *Helianthus* provides an excellent system for studying these questions given four-fold variation in nuclear DNA content among diploid species and well-resolved phylogenetic relationships.

Utilizing short-read Illumina data and sequence information from a diverse panel of *Helianthus annuus* (common sunflower) full-length LTR retrotransposons, I found that nuclear genome size in *Helianthus* species is positively correlated with repetitive DNA, and LTR retrotransposon subtypes generally show similar patterns in genomic abundance across taxa. *Helianthus* species with the largest genomes, however, exhibit large-scale amplification of a small number of LTR retrotransposon subtypes. Measuring aspects of plant growth and development at cell-, organ- and whole plant-levels in a panel of diploid *Helianthus* species that vary 4-fold in nuclear genome size, I found that genome size is negatively correlated with cell production rate, but that this negative correlation does not persist at higher organizational levels.

Taken together, these results provide insights into the mechanisms contributing to genome size evolution in plants and the organizational level at which genome size may impact growth patterns and developmental rates. Genome expansion in wild sunflowers is influenced most significantly by amplification of a small number of TEs and not necessarily by a greater diversity of TEs. Genome size is strongly negatively correlated with cell production rate but this relationship weakens at higher organizational levels, such as that of organ and whole-plant development.

TRANSPOSABLE ELEMENT CONTRIBUTION AND BIOLOGICAL CONSEQUENCE OF GENOME SIZE VARIATION AMONG WILD SUNFLOWER SPECIES

by

HANNAH M. TETREAULT

B.S., University of Maine-Machias, 2006

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Mark C. Ungerer

Copyright

HANNAH M. TETREAULT

2016

Abstract

Nuclear genome size varies immensely across flowering plants, spanning nearly 2400-fold. The causes and consequences of this vast amount of variation have intrigued biologists since it became clear that nuclear DNA amount did not reflect organismal complexity (the so-called C-value paradox). In my dissertation I utilize wild sunflower species in the genus *Helianthus* to examine the role of transposable elements (TEs), and in particular, long terminal repeat (LTR) retrotransposons, in generating genome size variation and whether variation in genome size influences aspects of plant growth and development across multiple organizational levels. The genus *Helianthus* provides an excellent system for studying these questions given four-fold variation in nuclear DNA content among diploid species and well-resolved phylogenetic relationships.

Utilizing short-read Illumina data and sequence information from a diverse panel of *Helianthus annuus* (common sunflower) full-length LTR retrotransposons, I found that nuclear genome size in *Helianthus* species is positively correlated with repetitive DNA, and LTR retrotransposon subtypes generally show similar patterns in genomic abundance across taxa. *Helianthus* species with the largest genomes, however, exhibit large-scale amplification of a small number of LTR retrotransposon subtypes. Measuring aspects of plant growth and development at cell-, organ- and whole plant-levels in a panel of diploid *Helianthus* species that vary 4-fold in nuclear genome size, I found that genome size is negatively correlated with cell production rate, but that this negative correlation does not persist at higher organizatonal levels.

Taken together, these results provide insights into the mechanisms contributing to genome size evolution in plants and the organizational level at which genome size may impact growth patterns and developmental rates. Genome expansion in wild sunflowers is influenced most significantly by amplification of a small number of TEs and not necessarily by a greater diversity of TEs. Genome size is strongly negatively correlated with cell production rate but this relationship weakens at higher organizational levels, such as that of organ and whole-plant development.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I want to thank my advisor, Mark Ungerer, for his intellectual support and meticulous eye for details. I also thank my committee members who have provided helpful feedback and support throughout my graduate career, Susan Brown, Christopher Toomajian and Michael Herman. I want to thank Bernd Friebe, for providing me with an opportunity to collaborate on a project together as well as his mentoring and support throughout graduate school.

I thank the Department of Education-GAANN fellowship for funding and support for most of my graduate career especially the co-PIs that received this funding Brett Sandercock, Anthony Joern, Michael Herman and Loretta Johnson. The Biology Graduate Student Association and the Graduate Student Council for travel awards to go to professional meetings. I want to thank Dr. Dan Boyle for help with microscopy.

I am thankful for all the friendships I have developed during my time in Manhattan, Kansas, without them it would have made life in the Flint Hills less agreeable. I want to thank my parents for their support, especially my father for demonstrating hard-work and perseverance to achieve your goals. Lastly, I am grateful to have the care and support from my husband, Edward Raynor, whose encouragement and support helped make this work possible and for always believing in me.

Preface

The contents of this dissertation are in collaboration with my advisor, Mark C. Ungerer and others who aided in data analysis and interpretation Chapter 2 is published with Mark C. Ungerer and formatted for the journal G3:Genes|Genomes|Genetics. Chapter 4 is published with Mark C. Ungerer, Takeshi Kawakami and Charlotte Levy as co-authors and is published in The American Midland Naturalist.

Chapter 1 - Introduction

Genome size evolution in plants

Plants exhibit a considerable amount of genome size variation, with angiosperms having almost 2400-fold variation in nuclear DNA content (Leitch and Leitch 2013). Understanding factors underlying this vast amount of variation has intrigued biologists since it first became clear that the DNA amount of an organism did not reflect organismal complexity (Vendrely, 1955). In 1971, Thomas coined the term 'C-value paradox' to reflect the lack of correlation between genome size and organismal complexity (Thomas, 1971). This paradox has largely been resolved as it is now well understood that genome size variation does not reflect differences in gene number but differences in the amount of repetitive DNA sequences, most typically transposable elements (TEs) (Flavell et al 1974). Questions remain, however, concerning the diversity of TEs that predominate in genomes of different size, the mechanisms and evolutionary forces driving genome size variation and its consequences on evolution. It is recognized that an increase in genome size arises through polyploidy and amplification of repetitive DNA, especially retrotransposons. Mechanisms of increase are offset by processes such as illegitimate recombination, and/or unequal intrastrand homologous recombination events resulting in genome size reduction (Devos et al. 2002; Vitte and Panaud 2003; Ma et al. 2004; Hawkins et al. 2009). Illegitimate recombination can occur between asymmetric pairing of homologous sequences and occasionally can occur in the absence of sequence homology (Wicker et al 2007b). Understanding TE diversity and abundance underlying genome size variation and the biological consequence of genome size variation have received considerable attention during recent years.

Transposable element contribution to genome size variation

The repetitive and highly abundant sequences in plant genomes consist mainly of Class I TEs, particularly long terminal repeat (LTR) retrotransposons (Kumar and Bennetzen 1999; Wicker et al. 2007a) with the most abundant being superfamilies *gypsy* and *copia.* LTR retrotransposons are capable of proliferating in their host genome via an RNA intermediate which is converted to extrachromosomal DNA prior to reinsertion in another place in the genome. This form of transposition and amplification is often referred to as a "copy-and-paste" mechanism and permits exceptionally high copy numbers of LTR retrotransposons in many plant lineages (Federoff 2012). LTR retrotransposons are characterized by two long terminal repeats at the 5' and 3' ends which flank two major genes, *gag* and *pol*, that encode a number of proteins (Kumar and Bennetzen 1999). LTR retrotransposons *gyspy* and *copia* superfamilies are identified by the order of *pol* protein coding domains (protease, reverse-transcriptase, integrase and RNAseH) and on sequence similarity (Figure 1.1). Sequence similarity is also used to classify sublineages within *gypsy* and *copia* superfamilies (Wicker et al 2007a). Different rates and degrees of amplification between TE superfamily and sublineage allow plant genomes of closely related species to differ considerably in TE content and genome size (Hawkins et al. 2006; Tenaillon et al. 2011). Comparisons of TE content and composition across species can highlight the dynamic role of TEs in plant genome evolution. In this dissertation, I explore the contribution of superfamilies *gypsy* and *copia* LTR retrotransposons to genome size variation among eight diploid sunflower species and one out-group species *Phoebanthus tenuifolius* (Chapter 2). The significance of this study is to inform how TEs contribute to genome size variation within species and among closely related species.

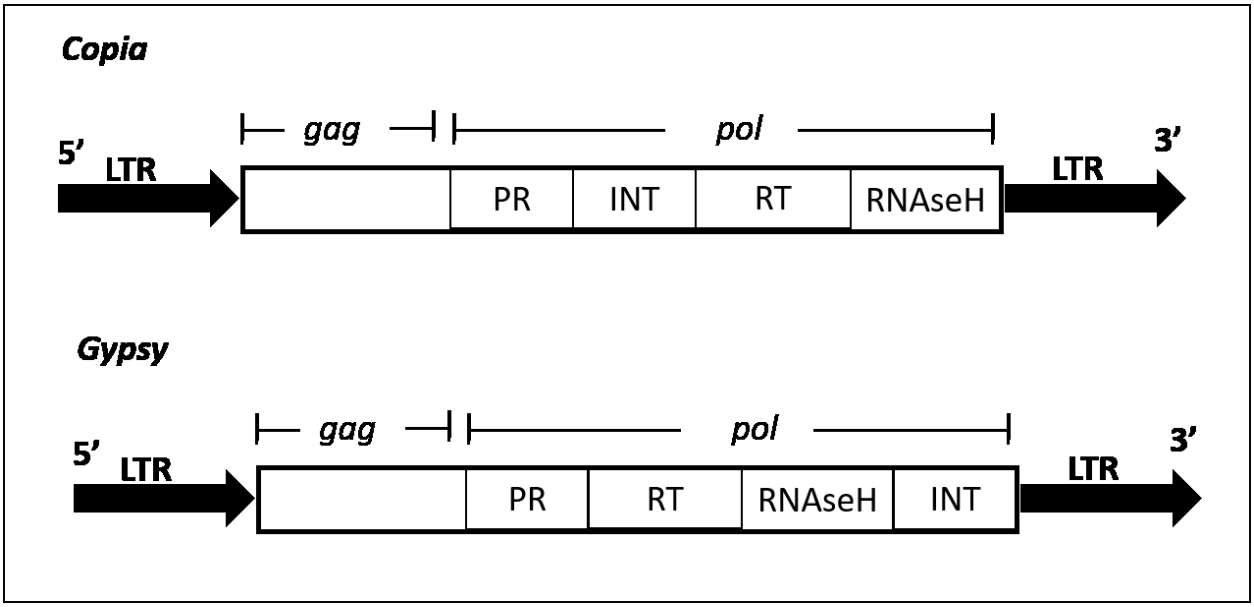

Figure 1.1 General structure of *copia* and *gypsy* LTR retrotransposons. The LTR retrotransposons have long terminal repeats on the 5' and 3' ends that are identical upon integration. The genes within the retrotransposons encode protease (PR), integrase (INT), reverse-transcriptase (RT) and RNAse-H.

Identification of transposable elements

Next generation sequencing (NGS) technologies and analytical approaches offer a novel approach to generate whole-genome sequence data and characterize repetitive DNA (Lam et al. 2012). A major advance from the advent of NGS sequencing is the ability to rapidly generate vast amounts of sequence data across the genome at a moderately low cost (Kelly and Leitch 2011). The amount of data generated from a single NGS sequencing run can be massive, therefore with the complexity and sequence variability of TEs, this substantial component of plant genomes has typically been treated as an inconvenience and discarded from sequence data. However, bioinformatic tools have been developed to aid in the analysis of TEs in genome sequences (Bergman and Quesneville 2007), advancing the analysis of the repetitive component of plant genomes

in genome structure and organization. With better analysis methods developed for the characterization of TE content, even under scenarios of low sequence coverage (Macas et al. 2007; Kurtz et al. 2008; Novak et al. 2010), there is much improvement in knowledge on the genomic TE content both within species and across related taxa (Tenaillon et al. 2011; Bonchev and Parisod 2013).

Genome size and the phenotype

The biological consequence to an organism having a large or small genome size is a current topic of research in plant evolutionary biology specifically for traits and developmental rates. Nuclear DNA content has been found to be positively correlated with cell size and duration of cell division rates (Bennett 1971, 1972, 1977; Cavalier-Smith et al 1985; Bennett 1998; Francis et al. 2008). Observations of genome size effect on cellular traits furthered predictions that larger quantities of DNA require longer cell cycle time thus would impact the growth rates of plant tissue (Beaulieu et al 2007; Bennett 1972) and many have studied the consequence of genome size on phenotypes, such as leaf anatomical traits and growth rate. Genome size has been found to be positively correlated with leaf area and cells per leaf in *Microseris* (Castro-Jimenez et al 1989), leaf width and length among cultivars of soybean (Chung et al 1998), and seed size in 18 *Pinus* species (Wakamiya et al 1993). However, in populations of *Poa annua* genome size positively correlated with stomatal cell size but negatively correlated with growth rate after 6 weeks of growth (Mowforth and Grime 1989). Similarly, in populations of *Dasypyrum villosum* genome size was negatively correlated with leaf length and width while positively correlated with seed weight and flowering interval (Caceres et al 1998). Chapter 3 of

this dissertation tests whether genome size variation effects growth among sunflower species that have a four-fold difference in genome size.

Helianthus as a study system for genome evolution

Helianthus has emerged as an excellent system for genomic, genetic and ecological studies (Kane et al. 2013). *Helianthus* is a diverse group of perennial and annual dicots, with a widespread range throughout the United States (Heiser et al. 1969) and a well resolved phylogeny of the diploid species (Stephens et al. 2015). *H. annuus (*common sunflower) is one of the well-studied diploid species in the genus. Information has been gained on domestication (Burke 2005; Wills and Burke 2007; Baack 2008), genome structure and organization (Kane et al 2011), local adaptation and the evolution of ecotypes within species (Blackman et 2011) utilizing this species. The genome of *H. annuus* is known to be highly repetitive with LTR retrotransposons comprising > 70% of the nuclear DNA (Staton et al 2012; Gill et al 2014). *Helianthus* as a system contains three hybrid species that arose from two progenitors three different evolutionary times and is a model system for speciation (Rieseberg et al 2003). Recent and large-scale proliferation events of LTR retrotransposons are documented among the hybrid species, as well as genome expansion when compared to the progenitors (Ungerer et al. 2006; Ungerer et al. 2009; Kawakami et al. 2010). Wild sunflowers in the genus *Helianthus* exhibit a four-fold difference in genome size among diploid species (Sims and Price 1985) providing an ideal system for studies of TE proliferation dynamics and associated genome evolution (Giordani et al. 2014) as well as biological consequence of genome size variation.

Research Objectives

In this dissertation, I focused on genome size variation among diploid, perennial and annual sunflower species. Chapter 2 explores the diversity of TEs that predominate in sunflower genomes of different size and the mechanisms and evolutionary forces driving genome size variation. Genome size is estimated for eight diploid sunflower species and an out-group species, *Phoebanthus tenuifolius* using flow cytometry; genome size for these species ranged nearly four-fold. Utilizing short read next-generation sequence data for the species under investigation, combined with sequence information from a panel of *H. annuus* full-length LTR retrotransposons in a *de novo* graph-based clustering comparison across species of LTR retrotransposon sublineage identity and abundance were conducted. Genome size and LTR retrotransposon dynamics among sunflower species were studied to: (1) determine the contribution of repetitive DNA to genome size among sunflowers, and (2) evaluate the contribution of *gypsy* and *copia* LTR retrotransposon sublineages to the genomes among and within sunflower species in this study. Chapter 3 tests whether genome size has an effect on aspects of plant growth and development. Nuclear DNA content has been consistently found to be positively correlated with cell size and duration of cell division rates but with growth traits at different levels, i.e., organ and whole plant, relationships between genomes size are less clear. For this chapter early growth and development was measured for 20 diploid perennial and annual sunflower species with genome size estimates, generated in our lab and by collaborators, that ranged over a four-fold difference. Rates of root cell production, shoot growth rate, leaf expansion and biomass accumulation were measured during early stages of plant development. Measurements of growth for species with varying genome sizes with different life cycles were

studied to assess (1) whether genome size has a consequence on growth and development and (2) whether genome size has a different effect on growth at various organizational levels, cell (cell production rate), organ (leaf expansion rate) and whole plant (growth rate and biomass accumulation). Combined, results from these studies demonstrate that (1) genome size is positively correlated with repetitive nuclear DNA among sunflower species, (2) sunflowers generally show similar abundances of different LTR retrotransposon sublineages with amplification of a small number of sublineages in species with the largest genomes, (3) genome size negatively correlates with root cell production, and (4) at higher organizational levels genome size does not correlate with growth traits, instead a signature of life cycle (annual and perennial) may play a role in growth and development.

Chapter 4 of this dissertation addresses variation among natural populations of *Helianthus maximiliani,* a perennial sunflower that spans along a latitudinal gradient from Canada to southwest United States in Texas. Species that are distributed broadly along geographical zones of extreme temperature variation are subject to local adaptation, especially plants with a sessile lifestyle. Three populations (Texas, Kansas and Manitoba, Canada) spanning a 2134 km gradient were examined for (1) variation in freezing tolerance before and after cold acclimation and (2) variation in flowering phenology from herbaria records. This chapter highlights both variation in freezing tolerance and flowering phenology along the latitudinal gradient. This study demonstrates that plant populations from the northernmost latitude that experience colder temperatures and a shorter growing season display the highest freezing tolerance and a history of earlier flowering phenology showing signs of local adaptation.

Literature cited

Beaulieu, J.M., Leitch, I.J., and C.A. Knight. 2007. Genome size evolution in relation to leaf strategy and metabolic rates revisited. Annals of Botany 99:495-505.

Bennett, M.D.. 1971. The duration of meiosis. Proceedings of the Royal Society of London Series B, Biological Sciences 178(1052):277-99.

Bennett, M.D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London Biological Sciences. 181(1063):277-299.

Bennett, M.D.. 1977. The time and duration of meiosis. Philosophical Transactions 277(955):201-26.

Bennett, M.D. 1998. Plant genome values: How much do we know? PNAS 95(5):2011-2016

Bergman, C.M. and H. Quesneville. 2007. Discovering and detecting transposable elements in genome sequences. *Briefings in Bioinformatics* 8(6):382-392.

Biradar, D. P., Bullock, D. G., & Rayburn, A. L. (1994). Nuclear DNA amount, growth, and yield parameters in maize. *Theoretical and Applied Genetics*,*88*(5), 557-560.

Cavalier-Smith, T. 1985. Eukaryotic gene numbers, non-coding DNA and genome size. In: The evolution of genome size. John Wiley & Sons Ltd, Chichester, pp69-103.

Cavallini, A., Natali, L., Cionini, G., & Gennai, D. (1993). Nuclear DNA variability within Pisum sativum (Leguminosae): nucleotypic effects on plant growth. Heredity-London-70, 561-561.

Caceres, M. E., De Pace, C., Mugnozza, G. S., Kotsonis, P., Ceccarelli, M., & Cionini, P. G. (1998). Genome size variations within Dasypyrum villosum: correlations with chromosomal traits, environmental factors and plant phenotypic characteristics and behaviour in reproduction. Theoretical and Applied Genetics, 96(5), 559-567.

Castro-Jimenez, Y., Newton, R., Price, H., & Halliwell, R. (1989). Drought Stress Responses of Microseris Species Differing in Nuclear DNA Content. American Journal of Botany, 76(6), 789-795. Retrieved from http://www.jstor.org/stable/2444535

Chung, J., Lee, J. H., Arumuganathan, K., Graef, G. L., & Specht, J. E. (1998). Relationships between nuclear DNA content and seed and leaf size in soybean. TAG Theoretical and Applied Genetics, 96(8), 1064-1068.

Devos, K.M., J.K. Brown, and J.L. Bennetzen, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res 12 (7):1075-1079.

Federoff, N.V. 2012. Transposable Elements, Epigenetics, and Genome Evolution. Science. 338(6108):758-767.

Flavell RB., Bennett MD, Sith JB, Smith DB. 1974. Genome size and proportion of repeated nucleotide DNA sequence in Plants. Biochemical Genetics. 12:257-269.

Francis, D. Davies, M.S., and P.B. Barlow. 2008. A strong nucleotypic effectof DNA C-value on the cell cycle regardless of ploidy level. Annals of Botany 101(6):747-757.

Giordani, T., A. Cavallini, and L. Natali, 2014 The repetitive component of the sunflower genome. Current Plant Biology 1:45-54.

Grime, J. P., Shacklock, J. M. L., & Band, S. R. (1985). Nuclear DNA contents, shoot phenology and species co-existence in a limestone grassland community. New Phytologist, 100(3), 435-445.

Hawkins, J.S., H. Kim, J.D. Nason, R.A. Wing, and J.F. Wendel, 2006 Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res 16 (10):1252-1261.

Hawkins, J.S., S.R. Proulx, R.A. Rapp, and J.F. Wendel, 2009 Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc Natl Acad Sci U S A 106 (42):17811-17816.

Heiser, C.B., D.M. Smith, S.B. Clevenger, and W.C. Martin, 1969 The North American sunflowers (Helianthus). Memoirs of the Torrey Botanical Club 22:1-218.

Kumar, A., and J.L. Bennetzen, 1999 Plant retrotransposons. Annu Rev Genet 33:479-532.

Leitch, I.J., and A.R. Leitch, 2013 Genome size diversity and evolution in land plants, pp. 307-322 in Plant Genome Diversity Volume 2, edited by I.J.L.e. al. Springer.

Ma, J., K.M. Devos, and J.L. Bennetzen, 2004 Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14 (5):860-869.

Mowforth, M. A., & Grime, J. P. (1989). Intra-population variation in nuclear DNA amount, cell size and growth rate in Poa annua L. Functional ecology, 289-295.

Niklas K. (1992) Plant Allometry (Univ. of Chicago Press, Chicago)

Rieseberg, L.H., 1991 Homoploid Reticulate Evolution in Helianthus (Asteraceae) - Evidence from Ribosomal Genes. American Journal Of Botany 78 (9):1218-1237.

Schilling, E.E., 1997 Phylogenetic analysis of Helianthus (Asteraceae) based on chloroplast DNA restriction site data. Theoretical and Applied Genetics 94 (6-7):925-933.

Schilling, E.E., C.R. Linder, R.D. Noyes, and L.H. Rieseberg, 1998 Phylogenetic relationships in Helianthus (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. Systematic Botany 23 (2):177-187.

Sims, L.E., and H.J. Price, 1985 Nuclear DNA content variation in Helianthus (Asteraceae). American Journal Of Botany 72 (8):1213-1219.

Stephens, J.D., W.L. Rogers, C.M. Mason, L.A. Donovan, and R.L. Malmberg, 2015 Species Tree Estimation of Diploid Helianthus (Asteraceae) Using Target Enrichment. American Journal Of Botany 102 (6):910-920.

Tenaillon, M.I., M.B. Hufford, B.S. Gaut, and J. Ross-Ibarra, 2011 Genome Size and Transposable Element Content as Determined by High-Throughput Sequencing in Maize and Zea luxurians. Genome Biology and Evolution 3:219-229.

Thomas CA. The genetic organization of chromosomes. Annual Review of Genetics 1971;5:237-256.

Timme, R.E., B.B. Simpson, and C.R. Linder, 2007 High-resolution phylogeny for Helianthus (Asteraceae) using the 18s-26s ribosomal DNA external transcribed spacer. American Journal Of Botany 94 (11):1837-1852.

Vendrely R. The deoxyribonucleic acid content of the nucleus. In: Chargaff E, Davidson JN, editors. The nuclei acids. Vol. II. New York, NY: Academic Press; 1955. p. 155-180.

Vitte, C., and O. Panaud, 2003 Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L.

Wakamiya, I., Newton, R. J., Johnston, J. S., and H.J. Price. (1993). Genome size and environmental factors in the genus Pinus. American Journal of Botany, 1235-1241.

Weiss-Schneeweiss H, Leitch AR, McCann J, Jang T-S, Macas J. Employing next generation sequencing to explore the repeat landscape of the plant genome. In: Hörandl E, Appelhans M, editors. Next Generation Sequencing in Plant Systematics. Regnum Vegetabile 157. Königstein, Germany: Koeltz Scientific Books; 2015.

Wicker, T., F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy et al., 2007a A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8 (12):973-982.

Wicker, T., Yahiaoui, N., and B. Keller. 2007b. Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance. The Plant Journal. 51(4):631-641.

Zilberman, D, and S. Henikoff. 2004. Silencing of transposons in plant genomes: kick them when they're down. *Genome Biology.* 5:249.

Chapter 2 - *L*ong *T*erminal *R*epeat retrotransposon content in eight diploid sunflower species inferred from next-generation sequence data

This chapter is formatted for the journal "G3|Genes|Genomes|Genetics"

The citation for this chapter is: Tetreault, H.M. and Ungerer M.C. 2016 *L*ong *T*erminal *R*epeat retrotransposon content in eight diploid sunflower species inferred from next generation sequence data. G3 doi:10.1534/g3.116.029082.

Abstract

The most abundant transposable elements (TEs) in plant genomes are Class I long terminal repeat (LTR) retrotransposons represented by superfamilies *gypsy* and *copia*. Amplification of these superfamilies directly impacts genome structure and contributes to differential patterns of genome size evolution among plant lineages. Utilizing short-read Illumina data and sequence information from a panel of *Helianthus annuus* (sunflower) full-length *gypsy* and *copia* elements, we explore the contribution of these sequences to genome size variation among eight diploid *Helianthus* species and an outgroup taxon, *Phoebanthus tenuifolius.* We also explore transcriptional dynamics of these elements in both leaf and bud tissue via RT-PCR. We demonstrate that most LTR retrotransposon sublineages (i.e., families) display patterns of similar genomic abundance across species. A small number of LTR retrotransposon sublineages exhibit lineage specific amplification, particularly in the genomes of species with larger estimated nuclear DNA content. RT-PCR assays reveal that some LTR retrotransposon sublineages are transcriptionally active across all species and tissue types,

whereas others display species-specific and tissue-specific expression. The species with the largest estimated genome size, *Helianthus agrestis*, has experienced amplification of LTR retrotransposon sublineages, some of which have proliferated independently in other lineages in the *Helianthus* phylogeny.

Introduction

Transposable elements (TEs) are DNA sequences capable of mobilizing within a host genome. Mobilization typically occurs either by physical excision-reinsertion events or through a process of replicative transposition whereby individual elements transcriptionally give rise to new copies that are reverse transcribed prior to insertion at new locations in the genome (Kumar and Bennetzen 1999; Feschotte et al. 2002). TEs that mobilize via replicative transposition (Class I TEs) are a major genomic component of many plant species because their mobilization involves sequence propagation and large-scale copy number increases. Because these events occur independently and at different rates and scales in the genomes of different plant lineages, even closely related species may diverge rapidly in TE content and genome size (Hawkins et al. 2006; Tenaillon et al. 2011).

The most abundant Class I TEs in plants are long terminal repeat (LTR) retrotransposons and are subdivided into superfamilies *gypsy* and *copia* (Kumar and Bennetzen 1999; Wicker et al. 2007). Differential proliferation and abundance of families (or sublineages) within these superfamilies are known to directly impact genome size evolution (Vitte and Panaud 2005; Hawkins et al. 2006; Piegu et al. 2006; Charles et al.

2008; Tenaillon et al. 2011; Piednoel et al. 2012; Kelly et al. 2015). Characterizing particular sublineages within superfamilies that undergo proliferation and determining patterns of proliferation events among related species can be a difficult task given sequence variation among sublineages and the difficulty of accurately estimating copy number abundance of elements within sublineages across taxa.

Advances in next generation sequencing (NGS) approaches have greatly facilitated efforts to generate and characterize whole-genome-level sequence data (Lam et al. 2012) for model and nonmodel organisms alike (Kelly and Leitch 2011). Major impediments of *de novo* genome assembly of NGS data exist, however, on account of short read lengths generated by many NGS platforms and the difficulty of assembling reads derived from genomes with a high repetitive fraction (e.g., with a large TE component). NGS data nonetheless have proved extremely informative for characterization of the genomic TE content both within species and across related taxa (Tenaillon et al. 2011; Bonchev and Parisod 2013), and several analysis methods have been developed for such characterizations, even under scenarios of low sequence coverage (Macas et al. 2007; Kurtz et al. 2008; Novak et al. 2010)

Wild sunflowers in the genus *Helianthus* provide an opportune system for studies of TE proliferation dynamics and associated genome evolution (Giordani et al. 2014). *Helianthus* includes approximately 49 species native to N. America that are collectively widespread throughout the United States, southern Canada and northern Mexico (Heiser et al. 1969) and phylogenetic relationships are well resolved (Rieseberg 1991; Schilling

1997; Schilling et al. 1998; Timme et al. 2007; Stephens et al. 2015). Genomic resources and tools are available for several *Helianthus* species (Kane et al. 2013) and a genome sequencing effort is underway for the cultivated sunflower *H. annuus* (Kane et al. 2011). Multiple ploidy levels are found within the genus (Kane et al. 2013) with genome size varying considerably even among species of the same ploidy (Sims and Price 1985). Genome structure and organization have been best characterized for the diploid species *H. annuus.* The genome of this species is highly repetitive, with LTR retrotransposons and their derivatives comprising > 70% of nuclear DNA (Staton et al. 2012; Gill et al. 2014). Recent insertional activity of these sequences has been documented in *H. annuus* (Buti et al. 2011; Staton et al. 2012) as have patterns of tissue specific expression (Gill et al. 2014). Recent and even larger-scale proliferation of LTR retrotransposons has been documented for three diploid annual *Helianthus* species derived via ancient hybridization events (Ungerer et al. 2006; Staton et al. 2009; Ungerer et al. 2009; Kawakami et al. 2010), with sublineages that proliferated in these species remaining active transcriptionally and expressed at higher levels when compared to the parental species from which the hybrid taxa are derived (Kawakami et al. 2011; Ungerer and Kawakami 2013).

In the current study we explore the contribution of LTR retrotransposons to genome size variation among eight diploid *Helianthus* species representing all four taxonomic sections based on current classification (Schilling and Heiser 1981) and an outgroup species, *Phoebanthus tenuifolius*. These eight *Helianthus* species represent much of the existing variation in diploid genome size, ranging nearly four-fold in

estimated nuclear DNA content (Sims and Price 1985). We combine short read NGS data with sequence information from a panel of *H. annuus* (common sunflower) full-length LTR retrotransposons in a *de novo* graph-based clustering approach that enables meaningful comparisons of LTR retrotransposon sublineage identity and abundances across species. We demonstrate that nuclear genome size is significantly correlated with repetitive DNA content in these species and that the species under investigation generally exhibit similar abundances of different LTR retrotransposon sublineages, suggestive of shared ancestry. We also note signatures of amplification for a small number of LTR retrotransposon sublineages in species with the largest genomes, thus identifying a contributing mechanism of genome size expansion in these species. Lastly, we highlight how graph-based clustering approaches are preferable to read-mapping based approaches in interspecific comparative analyses of TE abundance.

Materials and Methods

Plant materials and DNA sequencing

Seeds of species utilized in this study were obtained from the United States Department of Agriculture (USDA) National Plant Germplasm System (http://www.ars-grin.gov/npgs/) or collected from natural populations (Table 2.1). Seeds were germinated in the dark on moist filter paper in Petri dishes and 2-3 day old seedlings transferred to 8-inch pots with a 2:1 mixture of Metro-mix 350: all-purpose sand. All plants were grown under a 16 h : 8 h, light : dark cycle in the Kansas State University glasshouse facility. Watering was conducted daily or as needed and fertilization with a weak nutrient solution (N:P:K = 15:30:15) was applied weekly.

Young, fully-expanded leaves from one individual per species were collected for DNA extraction and subsequent whole genome shotgun (WGS) sequencing. All harvested tissue was flash-frozen in liquid nitrogen and stored at -80 ˚C until needed. Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). Three micrograms of total DNA per species were utilized for library preparation and WGS sequencing on an Illumina HiSeq2000 platform, generating 2x100 bp paired-end reads. Library preparation was performed following the Tru-Seq standard protocol (Illumina Inc., San Diego, CA, USA) with a library insert size of 350 bp. Libraries were multiplex sequenced on a single lane. Library construction and sequencing were performed at the University of Missouri DNA Core Facility, Columbia, MO, USA (http://biotech.missouri.edu/dnacore/). Raw sequence reads have been submitted to the NCBI Short Read Archive [SRP074507]. Sequence data were trimmed and filtered using Trimmomatic V0.30 (Bolger et al. 2014) according to the following criteria: (1) adapters and barcodes removed, (2) reads <80 bases removed, (3) bases trimmed from read ends if quality <30, and (4) read ends trimmed while mean quality <25 in a 4 bp sliding window. Chloroplast reads were removed by mapping the filtered dataset to the *H. annuus* chloroplast genome (NC_007977.1) using BWA v0.7.6 (Li and Durbin 2009) with default parameters. Genomic coverage for each species was estimated using the equation Coverage=LN/G (Lander and Waterman 1988), where L is average read length, N is number of reads per species and G is genome length. Genome length for each species was calculated utilizing the haploid 1C value, derived from 2C data estimated by flow cytometry (Table 2.1), and the equation 1 pg = 978 Mb (Dolezel et al. 2003).

Genome size determination

Nuclear DNA content (2C genome size) was estimated using a Guava PCA-96 microcapillary flow cytometry system (Guava Technologies, Hayward, CA, USA). Five biological replicates were evaluated per species with a minimum of 5000 events per sample. Sample preparation for flow cytometry followed that of Kawakami et al. (2011). An external standard (*Helianthus petiolaris*) was used along with the internal standard chicken erythrocyte nuclei (CEN; BioSure). Data were analyzed using CytoSoft V 2.5.4 (Guava Technologies, Haywood, CA, USA).

Estimation of genomic repetitive fraction based on short-read sequence data

The genomic repetitive fraction of each species was determined separately using a graph-based clustering approach developed by Novak et al. (2010) and implemented in RepeatExplorer (Novak et al. 2013) on the Galaxy Server (http://www.repeatexplorer.org/). Briefly, approximately 3 M single end (R1) 100 bp reads were randomly sampled from each species (Table 2.1, Table A.1) and clustered based on an all-by-all comparison of sequence similarity (≥90%) and overlap (≥55%). Individual clusters were identified and counted toward the genomic repetitive fraction if they contained ≥ 0.01% of the starting number of sampled sequences (e.g., for 3 M sequences, minimum cluster size = 300 sequences). These parameter values represent default settings of RepeatExplorer. Datasets where fewer than 3M reads were sampled (Table A.1), were automatically reduced by RepeatExplorer based on an initial analysis of a randomly sampled subset of reads and assessment of genome repeat structure as described in the RepeatExplorer manual. To assess potential variation in repetitive

fraction estimates for a given dataset, five separate graph-based clustering analysis runs (each analysis run ≈ 3 M randomly sampled reads) were conducted, with means ±SE presented in the Results.

To assess the strength of association between genome size and repetitive fraction, Pearson product–moment correlation coefficients and phylogenetically adjusted correlation coefficients were performed in Program R (v3.2.2, R Foundation for Statistical Computing, Vienna, Austria). The phylogenetically adjusted correlations were performed using phylogenetic independent contrasts with the 'APE' package in R (Paradis et al. 2001), based on evolutionary relationships presented in Stephens et al. (2015). The phylogeny was truncated using the drop.tip function in APE to consist only of the species under investigation, with the exception of *H. anomalus,* which is of hybrid origin (Rieseberg 2006).

Clustering with full length LTR retrotransposons from the *H. annuus* genome

To aid interpretation of repetitive sequence cluster identity and size across species as they pertain to LTR retrotransposons, graph-based clustering analysis runs were performed with a diverse reference panel of full-length *gypsy* and *copia* LTR retrotransposons derived from the *H. annuus* genome (Buti et al. 2011; Staton et al. 2012) (File S1). Individual elements were extracted and characterized from published BAC sequences utilizing the LTR*harvest* (Ellinghaus et al. 2008) LTR*digest* (Steinbiss et al. 2009) pipeline in *genometools* V 1.4.2. Of 110 full-length elements identified by these methods, 52 (40 *gypsy* + 12 *copia*) were identified as possessing an intact reverse-transcriptase (*RT*) domain and thus retained for phylogenetic analysis based on their *RT*

amino acid sequences (File S2). The majority of these full-length elements represent relatively 'young' copies, with insertion estimates within the last 2 million years (Buti et al. 2011; Staton et al. 2012). Multiple sequence alignment was performed with ClustalW and phylogenetic analysis was conducted using neighbor-joining (NJ) and maximum parsimony (MP) methods in MEGA 4.0.2 (Tamura et al. 2007). The reliability of tree topologies was estimated with bootstrap replication (1000 pseudo-replicates).

Full-length elements subjected to phylogenetic analysis (n=52) were subsequently converted to 100 bp kmers of sliding 85 bp overlap using a custom perl script to standardize sequence length with the Illumina-generated short read dataset. By this method, 281 to 1073 kmers were generated per full-length element (35,488 kmers total). The approximately 3 million Illumina reads per species were analyzed jointly with this collection of 100 bp kmers, which served as genomic 'tracers' enabling meaningful comparisons of the LTR retrotransposon content and abundance of different species' genomes. The decision to use 85 bp overlap for adjacent 100 bp kmers for each full-length element was based on the fact that the resulting similarity (100% shared bases across overlap of 85%) exceeded considerably the RepeatExplorer parameters for sequence clustering (i.e., ≥90% shared bases across overlap of ≥55%).

RT-PCR assays

LTR retrotransposon transcriptional activity was evaluated via RT-PCR in both vegetative (leaf) and reproductive (bud) tissues from a single individual per species. For each sampled plant, leaf tissue representing the eight-leaf stage and the first bud were

harvested and immediately flash frozen in liquid nitrogen. Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) and purified with an RNeasy Mini Kit (Qiagen, Valencia, CA, USA). RNA was treated with RNase-Free DNase (Qiagen, Valencia, CA, USA) to eliminate DNA contamination. Two sublineages of *gypsy* and a single sublineage of *copia* were assayed for transcriptional activity in both tissue types for all species utilizing sublineage specific primers targeting the *Integrase* and *RNASEH* domains of *gypsy* and *copia* elements, respectively (Kawakami et al. 2010; Ungerer and Kawakami 2013). RT-PCR assays were conducted using the ImProm-II Reverse Transcriptase system (Promega, Madison, WI, USA). RT-PCR reactions of the *actin* gene were used as positive controls for all samples. Negative control reactions were performed by withholding the reverse transcriptase enzyme. RT-PCR amplifications were conducted with an initial denaturing step of 94˚C for 2 min, followed by 5 cycles of 94˚C for 15 s, 55˚C (+1.0˚C /cycle) for 15 s, and 72˚C for 15 s, followed by 30 cycles of 94˚C for 15 s, 60˚C for 15 s, and 72˚C for 15 s, with a final incubation step of 72 ˚C for 5 min. Amplification products were size-separated via electrophoresis in 2% agarose gels and stained with ethidium bromide for visualization.

Results

Genome size and repetitive sequence content

Genome size estimates based on flow cytometry (Table 1) are largely consistent with earlier reports for overlapping *Helianthus* species (n=7) obtained by Feulgen-staining (Sims and Price 1985), with the exception of *H. divaricatus*, which was estimated at 2C = 9.41 (±0.08) pg in the current study versus 16.9 pg reported previously (Sims and Price 1985). Intraspecific

ploidy variation in *H. divaricatus* may underlie this observation (E. Baack, pers. comm.), though it is generally thought to be rare in *Helianthus* (Kane et al. 2013). Greater variability in 2C values exists among the sampled *Helianthus* annual species (range = 6.94 - 24.23 pg) versus perennial species (range = 9.32 - 12.91 pg; Table 2.1). With the exception of *H. agrestis,* all *Helianthus* species display 2C values lower than observed for closely related *Phoebanthus tenuifolius* (2C = 13.94pg ± 0.71), a diploid species and outgroup taxon for *Helianthus* (Schilling et al. 1998; Timme et al. 2007; Stephens et al. 2015).

The Illumina Hi-Seq platform generated ~6.7-16.9 M paired end reads (2 × 100 bp), post processing, for each of the eight *Helianthus* species and *P. tenuifolius*, yielding genome coverage estimates ranging from 0.23x to 0.68x (Table 2.1). Based on subsampling of ~2.4-3 M single end reads per species, graph based clustering yielded genomic repetitive fraction estimates between 68.17% and 82.12% (Table 2.1, Figure 2.1) and these estimates are strongly correlated with estimates of genome size (phylogenetic independent contrast analysis: $r = 0.9041$, $P = 0.0052$; Figure 2.1).

LTR retrotransposon contribution to genomic repetitive fraction

To evaluate the contribution of LTR retrotransposons to the repetitive fraction of these genomes, the short-read sequence data for each species were analyzed jointly with a library of 100 bp overlapping kmers derived from 40 full-length *gypsy* and 12 full-length *copia* elements identified previously from the *H. annuus* genome (see Methods). Phylogenetic analyses based on the *reverse transcriptase* (*RT*) domains of these 40 + 12 full-length elements indicate multiple well-supported *gypsy* and *copia* sublineages (Figure 2.2A,B, respectively). Comparisons of these

full-length element derived *RT* amino acid sequences across sublineages for both superfamilies revealed high sequence variability, with average genetic distances ranging from 0.108 to 0.667, and from 0.318 to 0.644 in pairwise comparisons of amino acid sequences from different sublineages within *gypsy* and *copia*, respectively (Table A.3). These elements are highly diverse, and represent a majority of the *gypsy* and *copia* diversity reported previously in sunflower based on sequence survey approaches (Ungerer et al. 2009; Kawakami et al. 2010) and analyses of multiple sequenced *H. annuus* BACs (Buti et al. 2011; Staton et al. 2012). Nomenclature for sublineage designations follows that reported in Ungerer et al. (2009) and Kawakami et al. (2010). Identified sublineages based on phylogenetic analyses presented herein also are largely congruent with family classification described in Staton et al. (2012) [Figure 2.2].

Clustering with these panels of modified full-length LTR retrotransposons allowed, for each species under investigation, assignment of short-read Illumina sequences to distinct *gypsy* and *copia* superfamilies and sublineages within these superfamilies (Figure 2.3A,B). Across species, sequences derived from *gypsy* elements were 3.8 to 5.3-fold more abundant than sequences derived from *copia* elements and together sequences derived from these two superfamilies combine for between 38.3% and 49.2% of all sequences for the species assayed (Table A.1). Sequences from specific *gypsy* sublineages consistently are more abundant within species genomes than others (e.g., sublineages A, B, C, X1 and X2 versus sublineages E', Y1, Y2, Z1, Z2; Figure 2.3A); these more abundant sublineages form a monophyletic group in phylogenetic analysis of *gypsy* sequences (Figure 2.2A).

For certain sublineages, elevated read densities were observed for some species, suggesting species-specific amplification events. For example, *Helianthus agrestis*, the species with the largest estimated genome size and highest genomic repetitive fraction, displayed elevated read densities for two *gypsy* sublineages (A and C), indicating that proliferation of these sublineages may underlie genome expansion in this species. Similar elevated density of reads was observed for sublineage A in *H. anomalus* and sublineage X1 in *Phoebanthus tenuifolius*.

Analogous patterns were observed for sublineages of *copia* elements with respect to relative abundance, with sublineages 1, 3 and 7 contributing disproportionately more, and sublineages 2, 4, 5 and 6 disproportionally less, to the genome repetitive fraction across species. Unlike observations for *gypsy* sublineages, however, the more abundant *copia* lineages are not monophyletic but rather consist of three separate, well supported lineages in the *copia* phylogeny (Figure 2.2B). Elevated density was observed in *copia* sublineage 1 for *H. agrestis,* again suggestive of a role of this sublineage in genome expansion. Elevated density, though to a lesser degree, also was observed in *copia* sublineage 3 for *H. angustifolius*.

Transcriptional activity of LTR retrotransposons in leaf and bud tissue

Transcriptional activity of *gypsy* sublineages A and C and *copia* sublineage 1 (see Figure 2.2A,B) was assayed via RT-PCR in both leaf and bud tissues for all species under investigation. Detection of transcriptional activity was variable across species and tissue types for *gypsy* sublineage A (Figure 2.4A), with transcripts clearly detected in both leaf

and bud tissue for all annual species but less detectable in perennial species; and with more detectable expression signal in bud versus leaf tissue for perennials. In contrast, transcriptional activity of *copia* sublineage 1 was equally detectable across all species and in both tissue types (Figure 2.4B). Transcriptional activity of *gypsy* sublineage C was not detected in any tissue type in any species (data not shown). Positive control reactions targeting *actin* expression yielded no detectable expression differences across tissue types or species (Figure A.1).

Discussion

Nuclear genome size across angiosperms varies dramatically, stretching nearly 2,400 fold between the smallest and largest documented flowering plant genomes (Leitch and Leitch 2013). Differential abundance and proliferation of TEs is now recognized as a significant contributor to genome size variation in plants, with LTR retrotransposons recognized as the most abundant and transpositionally dynamic (Hawkins et al. 2006; Piegu et al. 2006; El Baidouri and Panaud 2013). Evaluating TE diversity in organismal genomes has been revolutionized by next generation sequencing (NGS) technologies that enable rapid and detailed analysis of TE composition both within and among species (Macas et al. 2007; Swaminathan et al. 2007; Wicker et al. 2009; Piednoel et al. 2012; Sveinsson et al. 2013; Diez et al. 2014; Agren et al. 2015; Kelly et al. 2015).

Here we utilized Illumina short read sequence data coupled with sequence information from a panel of full-length *gypsy* and *copia* LTR retrotransposons obtained from the *H. annuus* genome to explore the contribution of LTR retrotransposons to

genome size variation among eight diploid *Helianthus* species representing all four taxonomic sections based on current classification schemes (Schilling and Heiser 1981) and an outgroup species, *Phoebanthus tenuifolius*. The species under investigation consist of both annuals and perennials, vary in genome size by approximately 4-fold, yet all are diploid with a haploid chromosome complement of n=17. Given that other major classes of TEs such as DNA transposons and non-LTR retrotransposons (e.g, LINEs) represent a very small fraction of the sunflower genome (~2% and 0.6%, respectively) (Staton et al. 2012), these other TE categories were not included in the current analyses. In addition, cluster annotation in RepeatExplorer based on the RepeatMasker Viridplantae database indicates that other repeat types (i.e., low complexity repeats, simple repeats, and satellite DNA) generally are rare (<2% combined). This latter category of repeat types was thus also excluded from analysis.

Variation in genome size

With the exception of *H. agrestis*, all *Helianthus* species investigated in the current study possess genome size estimates lower than that for the outgroup species *P. tenuifolius*. It is currently unknown whether this pattern is attributable to a general history of genome size reduction of *Helianthus* lineages, genomic expansion in *P. tenuifolius,* a combination of the two forces, or an artifact of the species sampled. Genome size reduction (DNA loss) can result from processes such as illegitimate recombination and/or unequal intrastrand homologous recombination events at the site of LTRs or interior coding regions of LTR retrotransposons (Devos et al. 2002; Vitte and Panaud 2003; Ma et al. 2004; Hawkins et al. 2009). Hallmarks of these events include the presence in the

genome of truncated elements and solo LTRs. These hallmarks have not been thoroughly investigated in *Helianthus* species or comparatively in *P. tenuifolius* due to a lack of sufficient sequence data. Based on analyses of 21 BAC clone sequences of the common sunflower *H. annuus*, however, truncated elements and solo LTRs do not appear a common feature of the sunflower genome (Staton et al. 2012). Evidence for independent genome expansion in *P. tenuifolius* following divergence from *Helianthus* lineages is suggested by elevated read density for at least one *gypsy* sublineage described in the current study (Figure 2.3A, see also Staton and Burke 2015). Despite these observations, broader trends across the Asteraceae suggest a directional increase in abundance of the more common *gypsy* LTR retrotransposons (and accordingly in genome size) from basal to more derived lineages, the latter of which include *Helianthus* and *Phoebanthus* species (Staton and Burke 2015). As such, *Helianthus* and *Phoebanthus* species genomes should be considered larger and with higher copy numbers of LTR retrotransposons compared with other members of Asteraceae, at least based on the limited sampling to date.

Clustering with panels of full-length LTR retrotransposons

A strong positive correlation was found between genome size and genome repetitive fraction, indicating an important role for repetitive DNA in underlying genome size variation in this group. Combining short read data with sequence information from a panel of full-length LTR retrotransposons in a *de novo* graph based clustering approach enabled meaningful comparisons of LTR retrotransposon presence and relative abundance across species. The majority of elements within this panel have estimated insertion times in the *H. annuus* genome within the last 2 million years (Buti et al. 2011;

Staton et al. 2012). As such, our analyses focus on LTR retrotransposons in *Helianthus* likely to have been active recently; more ancient elements potentially involved in older amplification events may be less well represented. Sequences derived from *gypsy* elements were observed to be 3.8 to 5.3-fold more common than sequences from *copia* elements for these species. These results are consistent with previous analyses of the *H. annuus* genome (Buti et al. 2011; Staton et al. 2012), and consistent with genomic composition analyses in other plant species where similar abundance biases have been observed (International Rice Genome Sequencing 2005; Ming et al. 2008; Paterson et al. 2009).

Our results indicate variation in abundance for different sublineages of *gypsy* and *copia* elements within genomes, but general stability in read density within a sublineage across species. Stability in read density across species is expected if most LTR retrotransposon proliferation activity occurred in the common ancestor of these species, with sequence abundances remaining relatively unchanged following subsequent cladogenesis. Elements from the most abundant sublineages of *gypsy* (i.e., sublineages A, B, C, X1 and X2; Figure 2.3A) represent part of a larger, well supported, monophyletic group (Figure 2.2A), and thus share a common evolutionary history. In contrast, *copia* sublineages with the highest read densities (i.e., sublineages 1, 3 and 7) represent more distantly related and non-monophyletic elements.

While general stability in read density within most sublineages was observed across species, exceptions to this pattern were found, most notably for three *gypsy*

sublineages (sublineages A, C and X1) and a single *copia* sublineage (sublineage 1). These exceptions were marked by higher read densities for species with larger genome size estimates, and were most apparent for *H. agrestis* and *P. tenuifolius*. These patterns likely reflect recent and lineage specific amplifications that have contributed to genome size expansion in these species. Similar patterns have been observed in other plant genera whereby differential abundance of a small number of LTR retrotransposon lineages underlies large genome size differences among species (Hawkins et al. 2006; Piegu et al. 2006; El Baidouri and Panaud 2013). Interestingly, representative elements for two of the abundant *gypsy* sublineages (i.e., RLG-iketas and RLG-wimu) and *copia* sublineage 1 (RLC-amov, RLC-jiliwu) [see Figure 2.2A and B, respectively] also display signatures of recent insertional activity in the common sunflower (*H. annuus*) genome, indicating potential widespread activity throughout the genus.

Developing appropriate methods for meaningful comparisons of TE content and abundance across species genomes has become increasingly necessary as next generation sequencing technologies continue to improve and costs continue to decline. The graph based clustering approach of short read Illumina data combined with sequence information from a TE reference panel proved effective for interspecific analyses of sublineage identification and sequence densities in *Helianthus*, and provides a useful method when TE reference panels are available. A potential complicating factor of this method is that, due to sequence divergence among genomes, fewer sequence reads and/or sublineages might be identified in interspecific comparisons as genetic distance increases from the TE reference panel. To explore this possibility, we tested whether the density of

species-specific Illumina reads clustering with *gypsy* and *copia* tracer sequences decreased with increasing genetic distance from the *H. annuus*-derived TE reference panel. We failed to find such a negative correlation (Figure A.2A). Interestingly, however, a negative correlation was observed when the LTR retrotransposon panel was used as a reference in a read-mapping based approach (Figure A.2B). This negative correlation persisted when mapping stringency was relaxed and greater numbers of mismatches allowed. Interspecific read-mapping to quantify TE abundances has been problematic in other species groups as well (Sveinsson et al. 2013), and generally should be avoided.

Transcriptional activity of gypsy and copia

Transcriptional activity of both *gypsy* and *copia* elements has been documented previously in both cultivated (Vukich et al. 2009; Gill et al. 2014) and wild (Kawakami et al. 2011; Kawakami et al. 2014; Ungerer and Kawakami 2013; Renaut et al. 2014) sunflowers. In the current study we confirmed expression of these elements in two species (*H. annuus* and *H. anomalus*) and demonstrate that transcriptional activity occurs broadly across the genus. Transcriptionally active elements documented in the current study represent the same variants associated with genome expansion events documented in three sunflower homoploid hybrid species (Ungerer et al. 2006; Kawakami et al. 2010).

Transcriptional activity of *gypsy* sequences was readily detectable in both leaf and bud tissue for all annual *Helianthus* species, less detectable in bud tissue of perennial

Helianthus species, and undetectable in leaf tissue of perennial *Helianthus* species. Although the primers used to assay for transcriptional activity were developed from *H. annuus* (an annual species), differential detection for annual versus perennial species is unlikely attributable to sequence divergence with increasing phylogenetic distance from *H. annuus* given that *H. agrestis* is an independently evolved annual species and more distant genetically from *H. annuus* than the remaining *Helianthus* species under investigation [Figure A.2]. It is interesting to note that more detectable transcriptional activity in annual species is consistent with a higher density of reads derived from this same sublineage based on clustering analyses of genomic short-read data (Figure 2.3A), demonstrating a potential link between transcriptional activity and genomic abundance level of element copy number in this group of plants. Quantitative PCR experiments have confirmed such a relationship comparing annual sunflower taxa *H. annuus* and *H. petiolaris* with their hybrid derivative species *H. anomalus*, *H. deserticola*, and *H. paradoxus*, where higher expression was observed in species with higher copy number abundances (Ungerer and Kawakami 2013, Renault et al. 2014, but see Gill et al. 2014). Transcriptional activity of this *gypsy* sublineage was not detected in either leaf or bud tissue of *Phoebanthus tenuifolius*, indicating that expression may be restricted to within *Helianthus*.

In contrast to results for *gypsy* transcriptional activity, expression of *copia* was equally detectable among *Helianthus* annual and perennial species, across tissue types, and in *P. tenuifolius*. Read density of genomic short-read data for this same sublineage generally are comparable across annual and perennial *Helianthus* species with the

exception of *H. agrestis*, for which read density is higher. More quantitative assays of transcriptional activity of both *gypsy* and *copia* elements may yield additional insights into expression dynamics of these elements across the sunflower genus. Transcriptional activity of additional sublineages of *gypsy* and *copia* have been documented previously in *H. annuus* (Gill et al. 2015) via RNA-seq and shown to exhibit tissue specific expression.

Genome expansion in *Helianthus agrestis*

A notable finding of the current study is genomic amplification of LTR retrotransposon sublineages in the genome of *H. agrestis*. *Helianthus agrestis* has a restricted geographical distribution in the southeastern United States, with populations found in central and southern Florida and in a single county in southern Georgia (Heiser et al. 1969). As noted above, this species is an annual, but distantly related from most other *Helianthus* annuals that form a monophyletic group and thus has independently evolved this life history form (Stephens et al. 2015). *Helianthus agrestis* is atypical in being one of only two *Helianthus* species that lack a self-incompatibility system (Heiser et al. 1969). Genome size estimates of *H. agrestis* indicate a nuclear genome ~1.9x - 3.5x larger than any other *Helianthus* species under investigation in the current study and ~1.7x larger than that for the outgroup species *P. tenuifolius.*

Genome expansion in *H. agrestis* is associated with amplification of a small number of LTR retrotransposon sublineages, represented by two different *gypsy* sublineages and a single *copia* sublineage. Sequences from these three sublineages represent approximately 28% of the *H. agrestis* genome based on our estimation

procedures (Table A.1). This observation is consistent with previous findings demonstrating that large interspecific variation in genome size can be attributable to a small number of LTR retrotransposon sublineages (Hawkins et al. 2006; Piegu et al. 2006; Vitte and Bennetzen 2006; El Baidouri and Panaud 2013) but contrasts with results observed for species of plants harboring some of the largest genomes (e.g., *Fritillaria*) where genome composition appears to consist of highly diverse, but relatively low abundance repeat types (Kelly et al. 2015). As noted above, two of the three most abundant sublineages in the *H. agrestis* genome (*gypsy* sublineage A and *copia* sublineage 1) have contributed to major genome expansion events in one or more diploid hybrid *Helianthus* species (Ungerer et al. 2006; Kawakami et al. 2010), and these sublineages remain active transcriptionally across the genus. It is thus noteworthy that the same LTR retrotransposon sublineages have experienced large-scale amplification events and promoted genome expansion independently in different regions of the *Helianthus* phylogeny. The forces governing activation (and repression) of these sublineages in different *Helianthus* species genomes is the focus of ongoing work.

Acknowledgements

We thank Loran Anderson for providing seeds of *Phoebanthus tenuifolius*, Bradley Olson for access to a Guava flow cytometry system, Carolyn Ferguson for helpful advice and two anonymous reviewers for comments that improved the final version. This work was supported by Kansas State University and National Science Foundation grant DEB-0742993 to M.C.U. Hannah Tetreault received support from the Department of Education Graduate Assistance in Area of National Need (GAANN) program. Contribution no. 16-330-J from the Kansas Agricultural Experiment Station.

Literature Cited

Agren, J.A., S. Greiner, M.T.J. Johnson, and S.I. Wright, 2015 No evidence that sex and transposable elements drive genome size variation in evening primroses. Evolution 69 (4):1053-1062.

Bonchev, G., and C. Parisod, 2013 Transposable elements and microevolutionary changes in natural populations. Molecular Ecology Resources 13 (5):765-775.

Buti, M., T. Giordani, F. Cattonaro, R.M. Cossu, L. Pistelli et al., 2011 Temporal dynamics in the evolution of the sunflower genome as revealed by sequencing and annotation of three large genomic regions. Theoretical and Applied Genetics 123 (5):779-791.

Charles, M., H. Belcram, J. Just, C. Huneau, A. Viollet et al., 2008 Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. Genetics 180 (2):1071-1086.

Devos, K.M., J.K. Brown, and J.L. Bennetzen, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res 12 (7):1075-1079.

Diez, C.M., E. Meca, M.I. Tenaillon, and B.S. Gaut, 2014 Three Groups of Transposable Elements with Contrasting Copy Number Dynamics and Host Responses in the Maize (Zea mays ssp mays) Genome. Plos Genetics 10 (4): e1004298.

Dolezel, J., J. Bartos, H. Voglmayr, and J. Greilhuber, 2003 Nuclear DNA content and genome size of trout and human. Cytometry A 51 (2):127-128.

El Baidouri, M., and O. Panaud, 2013 Comparative Genomic Paleontology across Plant Kingdom Reveals the Dynamics of TE-Driven Genome Evolution. Genome Biology and Evolution 5 (5):954-965.

Ellinghaus, D., S. Kurtz, and U. Willhoeft, 2008 LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9:18.

Feschotte, C., N. Jiang, and S.R. Wessler, 2002 Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3 (5):329-341.

Gill, N., M. Buti, N. Kane, A. Bellec, N. Helmstetter et al., 2014 Sequence-Based Analysis of Structural Organization and Composition of the Cultivated Sunflower (Helianthus annuus L.) Genome. Biology (Basel) 3 (2):295-319.

Giordani, T., A. Cavallini, and L. Natali, 2014 The repetitive component of the sunflower genome. Current Plant Biology 1:45-54.

Hawkins, J.S., H. Kim, J.D. Nason, R.A. Wing, and J.F. Wendel, 2006 Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res 16 (10):1252-1261.

Hawkins, J.S., S.R. Proulx, R.A. Rapp, and J.F. Wendel, 2009 Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc Natl Acad Sci U S A 106 (42):17811-17816.

Heiser, C.B., D.M. Smith, S.B. Clevenger, and W.C. Martin, 1969 The North American sunflowers (Helianthus). Memoirs of the Torrey Botanical Club 22:1-218.

International Rice Genome Sequencing, P., 2005 The map-based sequence of the rice genome. Nature 436 (7052):793-800.

Kane, N.C., J.M. Burke, L. Marek, G. Seiler, F. Vear et al., 2013 Sunflower genetic, genomic and ecological resources. Molecular Ecology Resources 13 (1):10-20.

Kane, N.C., N. Gill, M.G. King, J.E. Bowers, H. Berges et al., 2011 Progress towards a reference genome for sunflower. Botany-Botanique 89 (7):429-437.

Kawakami, T., B.J. Darby, and M.C. Ungerer, 2014 Transcriptome resources for the perennial sunflower Helianthus maximiliani obtained from ecologically divergent populations. Molecular Ecology Resources 14 (4):812-819.

Kawakami, T., P. Dhakal, A.N. Katterhenry, C.A. Heatherington, and M.C. Ungerer, 2011 Transposable Element Proliferation and Genome Expansion Are Rare in Contemporary Sunflower Hybrid Populations Despite Widespread Transcriptional Activity of LTR Retrotransposons. Genome Biology and Evolution 3:156-167.

Kawakami, T., S.C. Strakosh, Y. Zhen, and M.C. Ungerer, 2010 Different scales of Ty1/copia-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. Heredity 104 (4):341-350.

Kelly, L.J., and I.J. Leitch, 2011 Exploring giant plant genomes with next-generation sequencing technology. Chromosome Res 19 (7):939-953.

Kelly, L.J., S. Renny-Byfield, J. Pellicer, J. Macas, P. Novak et al., 2015 Analysis of the giant genomes of Fritillaria (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol 208 (2):596-607.

Kumar, A., and J.L. Bennetzen, 1999 Plant retrotransposons. Annu Rev Genet 33:479-532.

Kurtz, S., A. Narechania, J.C. Stein, and D. Ware, 2008 A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. BMC Genomics 9: 517.

Lam, H.Y., M.J. Clark, R. Chen, R. Chen, G. Natsoulis et al., 2012 Performance comparison of whole-genome sequencing platforms. Nat Biotechnol 30 (1):78-82.

Leitch, I.J., and A.R. Leitch, 2013 Genome size diversity and evolution in land plants, pp. 307-322 in Plant Genome Diversity Volume 2, edited by I.J.L.e. al. Springer.

Ma, J., K.M. Devos, and J.L. Bennetzen, 2004 Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14 (5):860-869.

Macas, J., P. Neumann, and A. Navratilova, 2007 Repetitive DNA in the pea (Pisum sativum L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. BMC Genomics 8:427.

Ming, R., S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte et al., 2008 The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452 (7190):991-996.

Novak, P., P. Neumann, and J. Macas, 2010 Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11:378.

Novak, P., P. Neumann, J. Pech, J. Steinhaisl, and J. Macas, 2013 RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29 (6):792-793.

Paterson, A.H., J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood et al., 2009 The Sorghum bicolor genome and the diversification of grasses. Nature 457 (7229):551-556.

Piednoel, M., A.J. Aberer, G.M. Schneeweiss, J. Macas, P. Novak et al., 2012 Next-generation sequencing reveals the impact of repetitive DNA across phylogenetically closely related genomes of Orobanchaceae. Mol Biol Evol 29 (11):3601-3611.

Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Saniyal et al., 2006 Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. Genome Research 16 (10):1262-1269.

Renaut, S., H.C. Rowe, M.C. Ungerer, and L.H. Rieseberg, 2014 Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. Philosophical Transactions of the Royal Society B-Biological Sciences 369 (1648).

Rieseberg, L.H., 1991 Homoploid Reticulate Evolution in Helianthus (Asteraceae) - Evidence from Ribosomal Genes. American Journal Of Botany 78 (9):1218-1237.

Rieseberg, L.H., 2006 Hybrid speciation in wild sunflowers. Annals of the Missouri Botanical Garden 93 (1):34-48.

Schilling, E.E., 1997 Phylogenetic analysis of Helianthus (Asteraceae) based on chloroplast DNA restriction site data. Theoretical and Applied Genetics 94 (6-7):925-933.

Schilling, E.E., and C.B. Heiser, 1981 Infrageneric Classification of Helianthus (Compositae). Taxon 30 (2):393-403.

Schilling, E.E., C.R. Linder, R.D. Noyes, and L.H. Rieseberg, 1998 Phylogenetic relationships in Helianthus (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. Systematic Botany 23 (2):177-187.

Sims, L.E., and H.J. Price, 1985 Nuclear DNA content variation in Helianthus (Asteraceae). American Journal Of Botany 72 (8):1213-1219.

Staton, S.E., B.H. Bakken, B.K. Blackman, M.A. Chapman, N.C. Kane et al., 2012 The sunflower (Helianthus annuus L.) genome reflects a recent history of biased accumulation of transposable elements. Plant J 72 (1):142-153.

Staton, S.E., and J.M. Burke, 2015 Evolutionary transitions in the Asteraceae coincide with marked shifts in transposable element abundance. BMC Genomics 16:623.

Staton, S.E., M.C. Ungerer, and R.C. Moore, 2009 The Genomic Organization of Ty3/Gypsy-Like Retrotransposons in Helianthus (Asteraceae) Homoploid Hybrid Species. American Journal Of Botany 96 (9):1646-1655.

Steinbiss, S., U. Willhoeft, G. Gremme, and S. Kurtz, 2009 Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res 37 (21):7002-7013.

Stephens, J.D., W.L. Rogers, C.M. Mason, L.A. Donovan, and R.L. Malmberg, 2015 Species Tree Estimation of Diploid Helianthus (Asteraceae) Using Target Enrichment. American Journal Of Botany 102 (6):910-920.

Sveinsson, S., N. Gill, N.C. Kane, and Q. Cronk, 2013 Transposon fingerprinting using low coverage whole genome shotgun sequencing in Cacao (Theobroma cacao L.) and related species. BMC Genomics 14:502.

Swaminathan, K., K. Varala, and M.E. Hudson, 2007 Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. BMC Genomics 8:132.

Tamura, K., J. Dudley, M. Nei, and S. Kumar, 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24 (8):1596-1599.

Tenaillon, M.I., M.B. Hufford, B.S. Gaut, and J. Ross-Ibarra, 2011 Genome Size and Transposable Element Content as Determined by High-Throughput Sequencing in Maize and Zea luxurians. Genome Biology and Evolution 3:219-229.

Timme, R.E., B.B. Simpson, and C.R. Linder, 2007 High-resolution phylogeny for Helianthus (Asteraceae) using the 18s-26s ribosomal DNA external transcribed spacer. American Journal Of Botany 94 (11):1837-1852.

Ungerer, M.C., and T. Kawakami, 2013 Transcriptional Dynamics of LTR Retrotransposons in Early Generation and Ancient Sunflower Hybrids. Genome Biology and Evolution 5 (2):329-337.

Ungerer, M.C., S.C. Strakosh, and K.M. Stimpson, 2009 Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. BMC Biology 7:40.

Ungerer, M.C., S.C. Strakosh, and Y. Zhen, 2006 Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16 (20):R872-873.

Vitte, C., and J.L. Bennetzen, 2006 Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proceedings of the National Academy of Sciences of the United States of America 103 (47):17638-17643.

Vitte, C., and O. Panaud, 2003 Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L. Molecular Biology and Evolution 20 (4):528-540.

Vitte, C., and O. Panaud, 2005 LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res 110 (1-4):91-107.

Vukich, M., T. Giordani, L. Natali, and A. Cavallini, 2009 Copia and Gypsy retrotransposons activity in sunflower (Helianthus annuus L.). BMC Plant Biology 9:150.

Wicker, T., F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy et al., 2007 A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8 (12):973-982.

Wicker, T., S. Taudien, A. Houben, B. Keller, A. Graner et al., 2009 A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant Journal 59 (5):712-722.

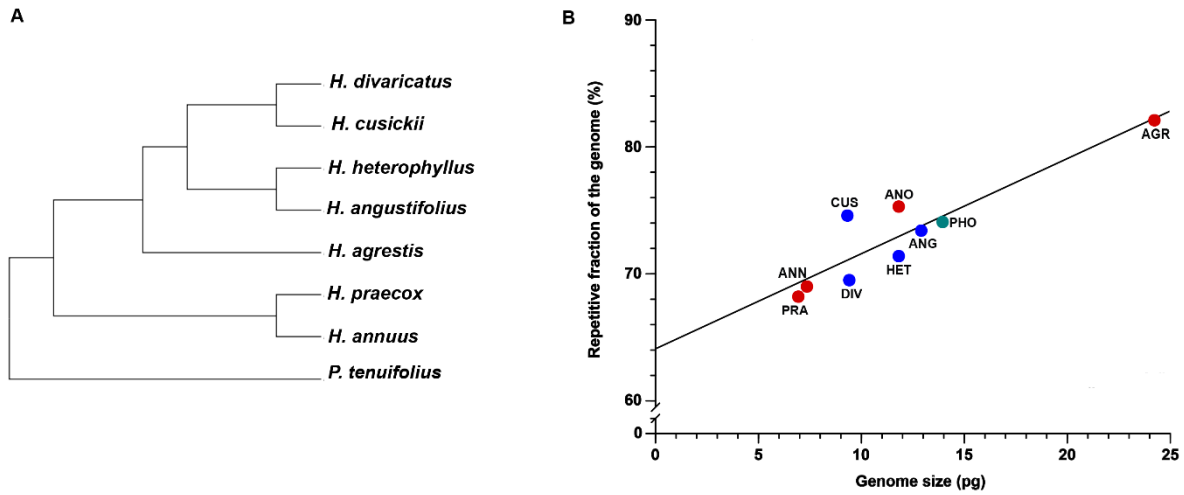

Figure 2.1 Phylogenetic relationships (A) and correlation between genome size and genomic repetitive fraction (B) for species under investigation. Phylogenetic tree is based on relationships presented in Stephens et al. (2015) and does not include H. anomalus, which is of hybrid origin (Rieseberg, 2006). Genome size and genome repetitive fraction are significantly correlated: phylogenetic independent contrast analysis: r = 0.9041, P = 0.0052; unmodified analysis: r = 0.9121, P = 0.0006. Species abbreviations in (B) are as in Table 1. Red = annual, blue = perennial, teal = perennial outgroup. Values (± SE) are provided in Table 1.

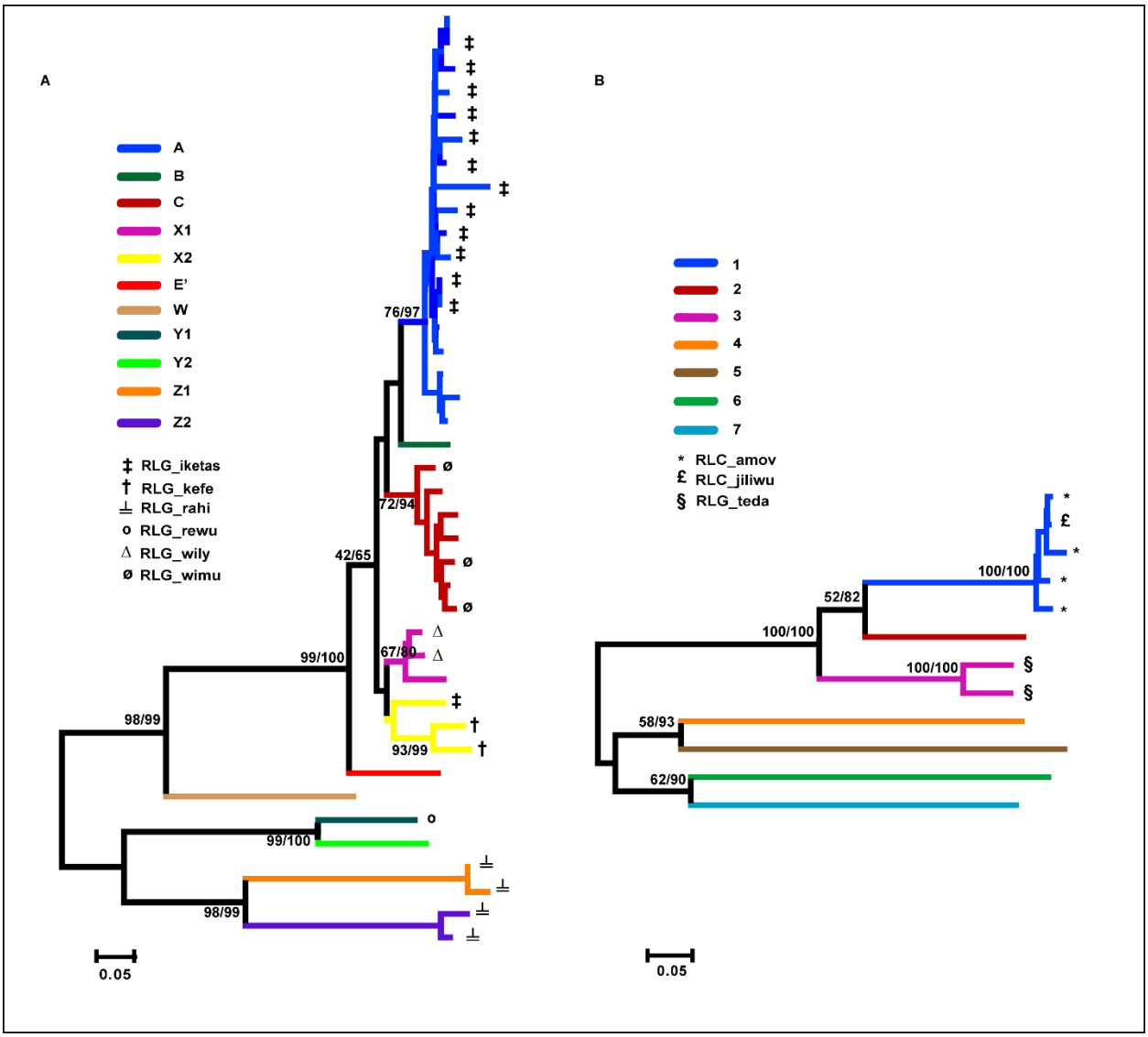

Figure 2.2 Neighbor-joining trees depicting sublineages of gypsy (A) and copia (B) elements based on 129 and 239 amino acid residues of the reverse transcriptase (RT) domain, respectively. Numbers above branches indicate bootstrap support for Maximum Parsimony/Neighbor-joining analyses. Branch colors depict different LTR retrotransposon sublineages and correspond to designations used in Ungerer et al. (2009) and Kawakami et al. (2010). Symbols at branch tips correspond to sunflower LTR retrotransposon families identified as highly abundant in Helianthus in Staton and Burke (2015).

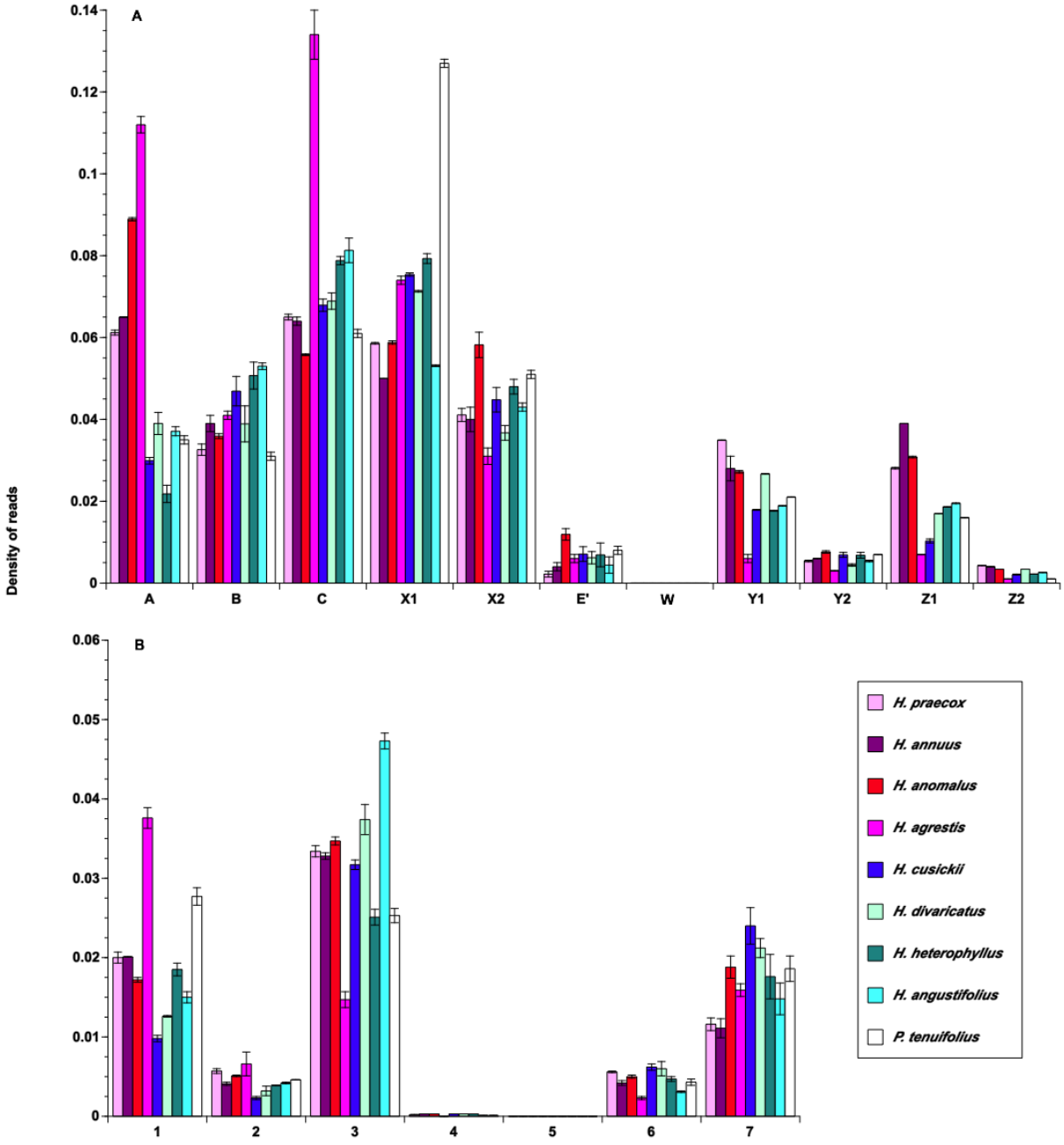

Figure 2.3 Genomic abundance of different sublineages of gypsy (A) and copia (B) elements. Shown are means (±SE) based on five graph-based clustering analysis runs for each dataset. Error bars for some histograms are too small to be seen at the resolution of this figure.

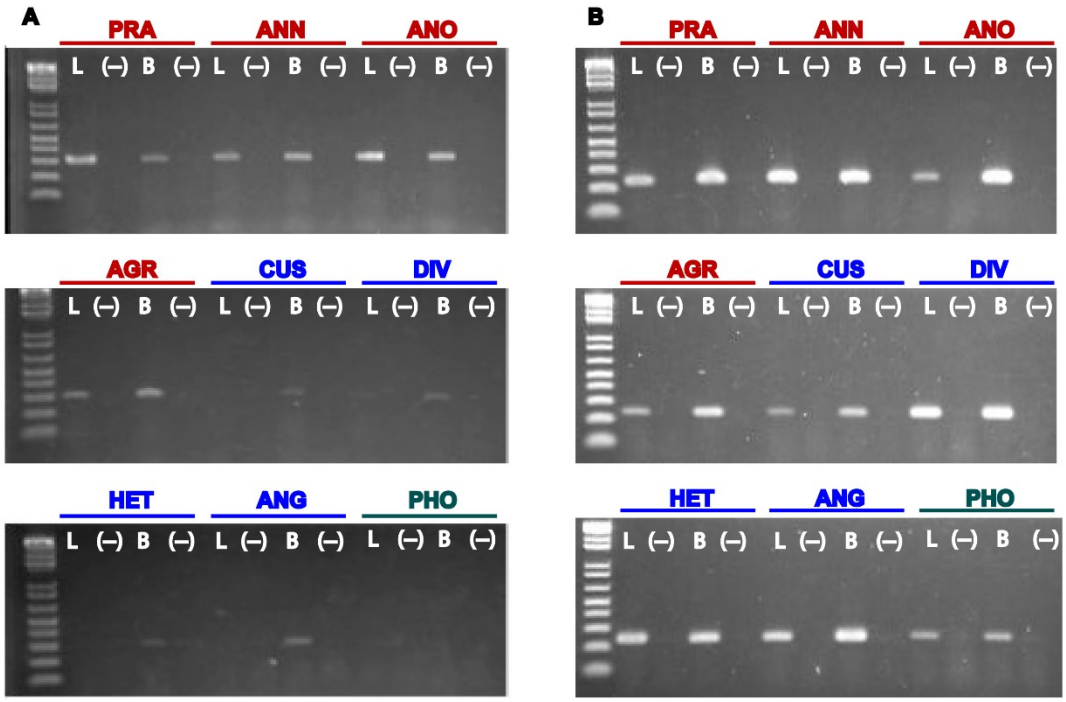

Figure 2.4 RT-PCR assays of gypsy sublineage A (A) and copia sublineage 1 (B) in leaf (L) and bud (B) tissue. Minus signs in parentheses indicate lanes with negative control reactions. Species abbreviations are as in Table 1. Red = annual, blue = perennial, teal = perennial outgroup.

Species	Abbreviation	Life cycle	Accession	Paired-end reads†	2C (pg) (SE)	Genome coverage	Repetitive fraction (%) (SE)
H. praecox	PRA	annual	PI 435847	10,314,126	6.94 (0.10)	0.59	68.17 (0.18)
H. annuus	ANN	annual	PI 468607	12,060,743	7.36 (0.12)	0.67	68.97 (0.21)
H. cusickii	CUS	perennial	PI 649959	11,981,577	9.32 (0.24)	0.51	74.58 (0.18)
H. divaricatus	DIV	perennial	PI 503212	6,752,840	9.41 (0.08)	0.29	69.55 (0.29)
H. anomalus	ANO	annual	PI 468642	12,228,849	11.82 (0.37)	0.41	75.26 (0.19)
H. heterophyllus	HET	perennial	PI 664732	11,753,278	11.82 (0.29)	0.40	71.42 (0.20)
H. angustifolius	ANG	perennial	ANG-MCU[1]	6,837,151	12.91 (0.32)	0.21	73.38 (0.33)
H. agrestis	AGR	annual	PI 468416	16,909,589	24.23 (0.84)	0.28	82.12 (0.15)
P. tenuifolius	PHO	perennial	PHO-LA[2]	10,971,465	13.94 (0.71)	0.31	74.08 (0.16)

Table 2.1 Study species, genome size estimates and associated genomic data

Note: † post-processing; [1] collected in Anson County, N.Carolina (M. Ungerer); [2] collected in Apalachicola National Forest (Loran Anderson)

Chapter 3 - Genome size effect on growth and development at multiple plant organizational levels among wild sunflower species

Abstract

Nuclear genome size varies considerably among plants, mostly due to differences in ploidy level and amount of repetitive DNA. The consequence of nuclear DNA variation among plants is still widely questioned, however for plants the relationship between genome and cell size have been thoroughly documented. Correlations between genome size and cell level traits across many plants established hypotheses on the consequence of genome size variation at larger organizational scales i.e., organ- and whole plant- level. This study evaluates whether genome size has an effect on aspects of plant growth and development among perennial and annual sunflower species. Wild sunflower species are an excellent system to study aspects of genome size variation given a four-fold difference in nuclear DNA content among diploid species. Measurements of growth and development were conducted at different organizational levels, cell production rate (cell-), leaf expansion rate (organ-), growth rate and biomass accumulation (whole plant- level). We demonstrate that genome size is negatively correlated with root cell production rate (cell-level) but at a lesser extent as plant organizational level increases. Instead a signature of life cycle (perennial vs annual) is observed at the organ and whole plant level traits.

Introduction

Nuclear DNA content varies considerably among plants with approximately 2400-fold variation among angiosperms (Leitch and Leitch 2013). These drastic differences in genome size are observed even among closely related species (Hawkins et al. 2006; Vitte et al. 2007).

Genome size variation among plants is a result from events of polyploidization (Wendel 2000) and proliferation of non-genic repetitive DNA such as transposable elements (Flavell et al. 1974; SanMiguel et al. 1996). An interest for plant evolutionary biologists is the biological consequence of genome size variation, especially in context to traits and developmental rates. Early investigators have found that nuclear DNA content affects cell division (Bennett 1977; Bennett 1971) and cell size (Bennett 1972; Francis et al. 2008) by increasing duration and size, respectively. Considering that genome size is strongly correlated with measurements at the cell-level, studies have tested whether genome size has an impact on phenotypes at the organ and whole plant level, such as seed weight, leaf morphology, growth rate, and photosynthetic rates (Bennet 1972; Beaulieu et al. 2007). However, patterns between genome size and traits at these higher plant organizational levels are less robust than observations at the cell level. Nuclear DNA amount and phenotypes at the higher organizational level do not always correlate in the same direction or magnitude across plant species studied thus far.

During their life cycle, plants grow by the addition of modules, which are repeated structures occurring in both root and shoot (Preston and Ackerly, 2004; Lyndon 2012). The apical meristem is the region where cells divide, enlarge and differentiate into tissue type, forming these modules. Growth is measured by a variety of parameters: some of which are weight, seed size, length, area, volume and cell number; growth rate is the change in one of these parameters over a given time. Plant development involves both growth and differentiation and is under control by intra and inter-cellular factors, an example for intra-cellular factors would be enzyme synthesis which differentially control gene expression for genes involved in cell differentiation (Jiang and Clouse 2001), an example for inter-cellular factors can be chemicals

such as growth hormones (Lyndon 2012). Intrinsic factors are a major contributor to plant development but extrinsic factors such as external environment greatly alter developmental and physiological traits (Sultan 2000). With plant growth greatly influenced by extrinsic and intrinsic factors, and nuclear DNA amount strongly correlated with cell level traits there is reason to believe genome size may play an important role on phenotypes at a higher organizational level, e.g., organ and whole plant, on growth and development (Knight et al. 2005).

Plant growth can be regarded as a process across different organization levels, operating from the cellular level to the organ and whole-plant level. The cell is the basic unit of a multicellular organism; therefore, the cell should constitute the most fundamental unit for growth and development of phenotypes across higher organizational levels, the organ and whole plant level (Tsukaya 2003). There is strong evidence that genome size is correlated with cell level measurements and in effect have formed the basis for hypotheses of genome size consequences on phenotypes at higher organizational levels i.e., at the organ and whole-plant level. An alternative to this prediction, however, is that the cost of carrying extra DNA is minimal at the organ and whole plant level with no direct consequence on traits (Knight et al. 2005; Oliver et al. 2007). Studies show contradictory results on the prediction of genome size effect with leaf anatomical traits (organ-level) and growth rates (whole plant-level). For example, genome size is correlated with specific leaf area (SLA), however both the direction and intensity of these correlations vary among plant groups, some have claimed a general trend of positive in angiosperms and negative in gymnosperms (Grotkopp et al. 2004; Knight et al. 2005; Morgan and Westoby 2005). It could be that genome size correlates strongly with traits at the cell level,

such as cell size and division rates, but the relationship decreases at higher organizational levels, such as in leaf and seed mass (Knight and Beaulieu 2008).

The genus *Helianthus* is an excellent system to study aspects of genome size variation given a four-fold difference in nuclear genome content among diploid species (Sims and Price 1985, Kane et al. 2013). Wild sunflowers are a diverse group of species that are widespread throughout North America and reside in many diverse habitats (Heiser et al. 1969). Given the well-resolved phylogeny (Stephens et al. 2015), analyses can be conducted that take into account evolutionary relationships (Felsenstein 1985). *Helianthus* contains both annual and perennial species and vast species-level variation in plant size and growth form (Kane et al. 2013), making this group an excellent system to evaluate the intrinsic factor of genome size on growth.

Here, we assess the effects of genome size on growth and development for plants within the genus *Helianthus*. Many studies have explored the biological consequence of genome size on traits at the cell level (Bennett 1972; Beaulieu et al. 2008), organ level (Castro-Jimenez et al. 1989; Chung et al. 1998) and whole plant level (Wakamiya et al. 1993) but seldom across all levels (but see Knight and Beaulieu 2007). If genome size is negatively correlated with cell production rate (cell-level), will we also detect slower growth rates at higher organizational levels for species with larger genomes? Here we use a phylogenetic comparative approach to examine the effect of genome size on rates of cell production (cell level), leaf expansion (organ level), growth and biomass accumulation (whole plant level) among 20 diploid *Helianthus* species. We found genome size is negatively correlated with cell production rate, consistent with studies on cell level traits, but this negative correlation does not persist with growth and

developmental traits at higher organizational levels. Differences were found between perennial and annual species consistent with theoretical predictions of slower growth for perennial species.

Materials and Methods

Plant material

To examine genome size consequence on growth across *Helianthus*, we selected 20 diploid non-hybrid species. These 20 species consisted of 7 annual and 13 perennial species distributed across the *Helianthus* phylogeny (Figure 3.1). Seeds from these species were collected in the field or obtained from the United States Department of Agriculture (USDA) National Plant Germplasm System (http://www.ars-grin.gov/npgs/)(Table 3.1). Seeds were germinated on moist filter paper in Petri dishes, 2-3 day old seedlings transferred to 2-inch pots with a 2:1 mixture of Metro-mix 350: all-purpose sand and grown under a 16 h: 8 h, light: dark cycle in a growth room for 7 days. After 7 days all seedlings were transferred to 8- inch pots with a 2:1 mixture of Metro-mix 350: all-purpose sand and grown under a 16 h : 8 h, light : dark cycle under ambient temperatures in the Kansas State University greenhouse facility. All species were replicated with 10-20 individuals. Watering was conducted daily or as needed and fertilization with a weak nutrient solution (N:P:K = 15:30:15) was applied weekly. Plants remained in the greenhouse for four weeks.

Cell production rate

Cell production rate measures the rate of increase for a given population of cells, in this case root cells (Baskin 2000) and is significantly correlated with cell cycle duration (Beemster et al. 2002). Six seeds for each species in this study (Table 3.1) were sterilized in a 10% bleach

solution for 5 min, then rinsed with sterilized distilled water five times. Seeds were imbibed on moist filter paper overnight in petri dishes, seed coats removed the following morning, then placed along a line on petri dishes (8 seeds/dish, 18 petri dishes total) containing 0.8% micropropagation agar Type-II (Caisson Laboratories Inc., Smithfield, UT, USA), 100 µg/ml ampicillin and 25 µg/ml gentamicin (Invitrogen Inc., Carlsbad, CA, USA). Plates were stored near vertical in a dark cabinet at 23°C with the line of seeds at the top. After plating, seed and root tip position were marked daily for 5 days. The growth rate (GRr) of individual roots was based on 5 days of growth. On day 5, roots were harvested from plates and fixed with a 10% formaldehyde in 1x PBS solution for 3 hours, rinsed and stored in 1x PBS.

Root cell walls were stained with an orange fluorescent dye, lipophilic carbocyanine 1mg/ml SP-DiIC18 (Invitrogen Inc., Carlsbad, CA, USA) dissolved in 100% ethanol, for 10 min and mounted on a microscope slide. A rectangle of fingernail polish was used as a spacer to reduce pressure from the coverslip. A Zeiss LSM 5 PASCAL (laser-scanning confocal microscope) equipped with a Zeiss Axiocam HR digital camera (Carl Zeiss Inc., Thornwood, NY, USA) was used to image roots. Fluorescence emission of DiIC18 was accomplished using the 543 nm line with the 20x/0.5 objective. Using the PASCAL imaging software, images were captured along the entire root beginning at the root meristem and moving proximally at a field of 450 µm. Mature cells were identified as those consistently producing root hairs (Foreman and Dolan 2001). Cell length (L) measurements were carried out using ImageJ software (Abramoff et al. 2004), lengths of approximately 20 mature cells were measured per root. Cell production rate (P) was calculated according to P = GRr/L (Baskin 2013), where P = cell production rate, GRr =

root growth rate and L= average mature cell length. A total of 4-6 individuals per species were determined for P and averaged together to represent species cell production rate.

Leaf expansion rate

Leaf images were captured every other day for 14 days on 8 randomly selected plants per species, using a Nikon D90 digital SLR camera (Nikon Inc., New York, USA). For each plant used for leaf imaging 2 leaves at the top node were selected to obtain an average leaf expansion rate for a given plant. During image acquisition leaves were gently taped to a white background with a 10-cm ruler fixed in the field of view. Leaf area was determined for each leaf using ImageJ software (Abramoff et al. 2004) with the 10-cm ruler used to set scale in each image. Leaf expansion rate (LER) was determined using the equation $LER = (\ln A_f - \ln A_o) / (\Delta t)$, where A_f is final leaf area, A_o is initial leaf area and t is time (Hunt 1982).

Growth rate

Height was measured from the soil surface to the apical meristem for each plant every other day for four weeks. Growth rate was assessed using linear and nonlinear growth models (Paine et al. 2012) fitted to height measurements, with the most appropriate models selected by evaluation of AIC and R^2. Growth models were implemented using the *gnls* function (Pinheiro et al. 2009) in library *nlme* (R Core Development Team, 2016). The linear model was the best performing model and thus was used to estimate growth rate for each species. Linear mixed effects models were performed using the *nlmer* function (Bates, Maechler, Dai 2008) in library lme4, with individuals accounted for as a random effect term and variance modeled to reduce heteroscedasticity using the power function.

Biomass

After the four weeks of growth in the greenhouse plants were harvested, aboveground and belowground tissue cleaned, separated and dried at 60°C for one week then mass recorded. Biomass was obtained by summing belowground and aboveground biomass. Aboveground and belowground tissue mass was used to determine root to shoot biomass ratio.

Genome size determination

Nuclear DNA content (2C) was estimated using flow cytometry of multiple individuals from 2-6 populations per species or was kindly provided by E. Baack and K. Whitney). Sample preparation for flow cytometry followed the one-step protocol as described in Dolezel et al. (2007). Briefly young, intact tissue of the experimental species was chopped together with an appropriate internal reference standard using the LB01 buffer formula (Dolezel et al. 2007). The sample was filtered through a 30 um nylon mesh then centrifuged to collect nuclei. Nuclei were stained for 2 hours using propidium iodide then analyzed with a Guava PCA-96 microcapillary (Guava Technologies, Hayward, CA, USA) or with an Attune NxT (Thermo Fischer Scientific Inc., Waltham, MA, USA) flow cytometry system. The following species were used as internal reference standards: *Zea mays* L. 'CE-777' (5.43 pg), *Pisum sativum* L. 'Citrad'(9.09 pg) and *Secale cereal* L. 'Dankovske' (16.19 pg) (Dolezel et al. 2007).

Statistical analysis

All statistical analyses were performed within the computing environment R 3.2.4 (R Development Core Team, 2016). All annual sunflowers exhibit an erect growth form, while perennials, exhibit one of two growth forms: erect or basal rosette. To test whether there was a difference among perennial growth form in our growth metrics, an ANOVA was implemented with growth metric as the response variable and perennial growth form as the independent variable. When there was a significant difference among perennial growth form in a growth metric perennial species with the basal rosette growth form were excluded from downstream analysis.

To assess the strength of association between genome size and growth metric, species means for growth rate, leaf expansion rate, total biomass, root to shoot biomass ratio and cell production rate were used in Pearson product-moment correlation coefficients performed using phylogenetic independent contrasts in library "APE" (Paradis et al. 2004) based on evolutionary relationships presented in Stephen et al. (2015). The phylogeny was truncated using the *drop.tip* function in APE to consist only of the species under investigation. Species estimates for growth metrics and genome size were log transformed for correlation analyses.

To assess differences in growth metrics among annual and perennial life cycle, individuals were analyzed by ANOVA using linear mixed effects models with *nlmer* function (Bates, Maechler, Dai 2008) in library lme4, with Species represented by a random effect term. Data were log-transformed to meet the assumption of normal distribution. Least square means were estimated for life cycle using the *lsmeans* function in the library lsmeans (Lenth 2012). Using the trimmed phylogeny, phylogenetic generalized least squares ANOVA using species

growth metric means were performed with the *pgls* function (Freckleton et al. 2002) with library *caper* (Orme 2013). To test for phylogenetic signal in life cycle strategy the *D* statistic, which is appropriate for use with binary traits was implemented using the *phylo.d* function (Pinheiro et al. 2009) utilizing 1000 permutations. An estimated $D > 1$ represents a trait that is randomly distributed at the tips of a given phylogeny, whereas $D < 0$ represents a distribution expected under Brownian motion (i.e. in a random walk with constant trait variance over time [Felsenstein 1985]), *D* greater than 1 is interpreted as over-dispersed while a negative *D* means more phylogenetically clumped than expected (Fritz and Purvis 2010).

Results

Genome size relationship with growth traits

Based on phylogenetic independent contrast analysis root cell production rate estimates were negatively correlated with genome size (r= -0.7052, p= 0.0011, Figure 3.2A). At the whole plant traits, we failed to find evidence of a relationship between genome size and shoot growth rate, leaf expansion rate, biomass and root to shoot biomass ratio (Figure 3.2B-E). The difference between erect and basal rosette growth form was significant for growth metrics of growth rate and biomass ($p > 0.001$ and $p = 0.0253$, respectively) but not for leaf expansion rate and root to shoot biomass ratio (p= 0.234 and p= 0.421, respectively). Therefore, species with the basal rosette growth form, *H. angustifolius, H. occidentalis, H. atrorubens, H. heterophyllus,* and *H. carnosus*, were removed from downstream analyses involving growth rate and biomass.

Growth metric differences between annual vs perennial species

ANOVA results for perennial and annual life cycle showed trait differences for leaf expansion rate (p = 0.0342, Figure 3.3B, Table 3.2) and biomass (p = 0.0492, Figure 3.3E, Table 3.2). Cell production rate, growth rate and root to shoot biomass ratio were not significantly different between perennial and annual life cycle strategies among species in this study (Figure 3.3A, C and D). The distribution of annual and perennial species across *Helianthus* exhibited a high level of phylogenetic signal (D = -7.80). When phylogenetic relationships between annual and perennial is accounted for in the model, however, differences in leaf expansion rate and biomass were marginally non- significant (p = 0.0955 and p = 0.1472, respectively).

Discussion

Sunflowers occupy a broad range of habitats and exhibit considerable phenotype variation related to growth, survival and fitness (Kane et a 2013; Mason and Donovan 2015). Recent focus in sunflower research is centered on phenotype variation among species and populations (Kane et a 2013). With a four-fold difference in genome size among diploid sunflower species (Sims and Price 1985; Kane et al. 2013) determining whether genome size has an important role on growth traits would aid in a better evolution-based understanding of these species. Here, we have sampled perennial and annual species across *Helianthus* to advance our understanding of the role genome size has on growth. The growth metrics that we assessed involved traits during early belowground and aboveground tissue development. At this level, our findings suggest genome size has a greater effect at the cell level but not at the organ or whole plant level of development.

Genome size relationship with growth metrics

Our results indicate cell production rate is negatively correlated with genome size among the diploid *Helianthus* species sampled (Figure 3.2A). Our results at the cell level are consistent with studies that involved sampling broadly across angiosperm species where nuclear DNA amount is positively correlated with cell size (Beaulieu et al. 2008) and cell division duration (Van't and Sparrow 1963; Francis et al. 2008).

The link between cell level traits and genome size is a robust relationship in broad groups of angiosperms and we show a similar pattern among *Helianthus* species in this study. Plant material is composed of cells undergoing division, enlargement and differentiation to tissue type (Steeves & Sussex 1989). A common prediction is at a higher organizational level, i.e., organ and whole-plant, nuclear DNA amount would affect plant material accumulation, either in growth rate, leaf expansion rate or in total biomass accumulation. However, we did not detect such a relationship between genome size and growth metrics at the organ and whole-plant level. Studies including higher level traits generally show variable patterns between genome size and the specific trait: for example, leaf anatomical traits have been positively (Castro-Jimenez et al. 1989; Chung et al. 1998), negatively (Caceres et al. 1998; Grime et al. 1997) and non-significantly (Grime et al. 1997) correlated with genome size. These contradictory patterns are also evident for relative growth rates, where positive (Leishman 1999; Grime et al 1985), negative (Mowforth and Grime 1989; Ceccarelli et al. 1993), and non-significant (Natali et al. 1993; Grime et al. 1997) correlations are observed. It is not uncommon for phenotypes at the higher organizational level to exhibit less association between genome size. Knight and Beaulieu

(2008) observed across a large cross-species comparison that genome size explains less variation at the level of cell density than cell size, for stomatal cells, and even less predictive power at the higher organizational level of whole-plant for traits of photosynthetic rates and biomass. Plant growth is a complex trait, our results at the higher organizational level indicate that the relationship between genome size and phenotype is not always direct and allometric (Balao et al. 2011). Perhaps phenotypes at the higher organizational level are influenced more by other factors than amount of nuclear DNA, factors such as environmental and genetic. Since our study was conducted in a common garden it is more likely that growth metrics at higher organizational levels are due to differences in gene expression (Broz et al. 2009) involved in growth and development.

Growth metric differences in annual and perennial species

Life cycle strategies are extensively studied within plants, specifically how they modify growth to maximize survival and fitness (Bell 1980; Stearns 1992; Westoby et al. 2002). There are two general categories: annuals which grow vegetatively with reproductive growth occurring toward the end of the growing season and perennials that reproduce repeatedly and cycle through vegetative and reproductive phases in their lifetime. Both annuals and perennials undergo a developmental switch to concentrate growth from vegetative to reproductive tissue; however, the timing to achieve this switch varies among species and can vary among closely related taxa (Friedman and Rubin 2015).

On the basis of previous theoretical and experimental work, whole-plant development is found to be associated with life cycle (Young and Augsperger 1991). Our study's main objective

was to focus on genome size, however with the inclusion of both life cycles it was possible to look at differences in growth metrics between annual and perennial. We show there is greater tissue acquisition among annuals than perennials for both above and below ground tissues in leaf expansion rate and biomass, following the prediction for annual life strategy of fast development during a short life cycle (Franco and Silvertown 1996). Seedlings of annuals differ from perennials for a number of leaf traits (Garnier and Laurent 1994; Garnier et 1997) resulting in a higher growth rate for annual species. Studies comparing leaf traits (organ level) between the two life cycles largely support the theory that annual species are characterized by a set of leaf traits enabling high resource capture while perennials are characterized by leaf traits associated with longevity and defense (Garnier 1992; Garnier et al. 1997). We show a phylogenetic signal (D = -7.80) for life cycle among *Helianthus* species, when we account for phylogeny in our analysis the level of significance decreases for leaf expansion rate and biomass and are only marginally significant.

Despite the differences summarized above, root to shoot biomass ratio and growth rate were not significantly different between annual and perennial *Helianthus* species which contradict predictions. An expectation is a greater root to shoot biomass ratio for perennials than annuals, but our lack of difference in root to shoot biomass ratio allocation is not uncommon during the early stages of plant growth (Garnier 1992), if we grew plants for longer than four weeks these differences may emerge.

On the basis of our data, we conclude that the relationship between genome size and growth metrics among *Helianthus* species decreases with traits at higher organizational levels of

plant development. This is somewhat surprising given the strength of the relationship at the cell level. The genome size effect on cellular processes is not unique as other investigators have documented strong relationships between cell division and size. Lastly, though life cycle is confounded in the genus by phylogeny we detect a trend for growth metrics at the whole plant level to be governed by life cycle.

Literature Cited

Ackerly, D.D., Dudley, S.A., Sultan, S.E., Schmitt, J., Coleman, J.S., Linder, C.R. et al. 2000. The Evolution of Plant Ecophysiological Traits: Recent Advances and Future Directions: New research addresses natural selection, genetic constraints and the adaptive evolution of plant ecophysiological traits. BioScience 50(11):979-995.

Balao, F., Herrera, J., and S. Talavera. 2011. Phenotypic consequences of polyploidy and genome size at the microevolutionary scale: a multivariate morphological approach. New Phytologist 192(1):256-265.

Baskin TI. 2000. On the constancy of cell division rate in the root meristem. Plant Molecular Biology 43(5-6):545-54.

Baskin, T.I. 2000. On the constancy of cell division rate in the root meristem. Plant Molecular Biology 43(6):545-54.

Bates D, Maechler M, Dai B. 2008. The lme4 package R package version 2(1):74.

Bates, D., Maechler, M., Dai, B. and L. Matrix. 2007. The lme4 package. R package version. 2(1):14.

Beaulieu, J. M., Leitch, I. J., Patel, S., Pendharkar, A., ad C.A. Knight. 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. New Phytologist 179(4):975-986.

Beaulieu, J.M., Leitch, I.J., and C.A. Knight. 2007. Genome size evolution in relation to leaf strategy and metabolic rates revisited. Annals of Botany 99:495-505.

Beemster, G.T.S., Vusser, K., Tavernier, E., De Bock, K., and D. Inze. 2002. Variation in growth rate between Arabidopsis ecotypes is correlated with cell division and A-type cyclin-dependent kinase activity. Plant Physiology 129(2):854-864.

Bell, G. 1980. The Costs of Reproduction and their Consequences. The American Naturalist 116(1):45-76.

Bennett MD. 1971. The duration of meiosis. Proceedings of the Royal Society of London Series B, Biological Sciences 178(1052):277-99.

Bennett, M. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London B: Biological Sciences 181(1063):109-135.

Bennett MD. 1977. The time and duration of meiosis. Philosophical Transactions 277(955):201-26.

Biradar, D. P., Bullock, D. G., and Rayburn, A. L. 1994. Nuclear DNA amount, growth, and yield parameters in maize. Theoretical and Applied Genetics 88(5):557-560.

Bonser, S.P., and Aarssen, L.W. 2005. Meristem allocation and life- history evolution in herbaceous plants. Canadian Journal of Botany 84(1):143-150.

Broz, A.K., Manter, D.K., Bowman, G., Muller-Scharer, H., and J.M. Vivanco. 2009. Plant origin and ploidy influence gene expression and life cycle characteristics in an invasive weed. BMC Plant Biology 9(33): doi: 10.1186/1471-2229-9-33.

Caceres, M. E., De Pace, C., Mugnozza, G. S., Kotsonis, P., Ceccarelli, M., and Cionini, P. G. 1998. Genome size variations within *Dasypyrum villosum*: correlations with chromosomal traits, environmental factors and plant phenotypic characteristics and behaviour in reproduction. Theoretical and Applied Genetics 96(5):559-567.

Castro-Jimenez, Y., Newton, R., Price, and H., Halliwell, R. 1989. Drought Stress Responses of *Microseris* Species Differing in Nuclear DNA Content. American Journal of Botany 76(6):789-795.

Cavallini, A., Natali, L., Cionini, G., and Gennai, D. 1993. Nuclear DNA variability within *Pisum sativum* (Leguminosae): nucleotypic effects on plant growth. Heredity 70: 561-561.

Ceccarelli, M., Minelli, S., Falcinelli, M., and Cionini, P. G. 1993. Genome size and plant development in hexaploid *Festuca arundinacea*. Heredity 71(6):555-560.

Chung, J., Lee, J. H., Arumuganathan, K., Graef, G. L., and Specht, J. E. 1998. Relationships between nuclear DNA content and seed and leaf size in soybean. Theoretical and Applied Genetics 96(8):1064-1068.

Cornwell, W.K. and Ackerly, D.D. 2009. Community assembly and shifts in the distribution of functional trait values across an environmental gradient in coastal California. Ecological Monographs 79(1):109–126.

Dolezel, J. Greilhuber, J., and J. Suda. 2007. Estimation of nuclear DNA content in plants using flow cytometry. Nature Protocols. 2(9):2233-2244.

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

Flavell, R.B., Bennett, M.D., Sith, J.B., and D.B. Smith. 1974. Genome size and proportion of repeated nucleotide DNA sequence in Plants. Biochemical Genetics 12:257-269.

Foreman J and Dolan L. 2001. Root hairs as a model system for studying plant cell growth. Annals of Botany 88(1):1-7.

Foreman J and L. Dolan. 2001. Root hairs as a model system for studying plant cell growth. Annals of Botany 88(1):1-7

Francis, D., Davies, M.S., and P.B. Barlow. 2008. A strong nucleotypic effect of DNA C-vaue on the cell cycle regardless of ploidy level. Annals of Botany. 101:747-757.

Franco, M., and J. Silvertown. 1996 Life history variation in plants: an exploration of the fast-slow continuum hypothesis.

Freckleton, R. P., Harvey, P. H., and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. The American Naturalist 160(6):712-726.

Fritz, S.A., and A. Purvis. 2010. Selectivity in Mammalian Extinction Risk and Threat Types: A New Measure of Phylogenetic Signal Strength in Binary Traits. Conservation Biology 24(4):1042-1051.

Garnier, E. 1992. Growth analysis of congeneric annual and perennial grass species. Journal of Ecology 665-675.

Garnier, E., and G. Laurent. 1994. Leaf anatomy, specific mass and water content in congeneric annual and perennial grass species. New Phytologist 128(4):725-736.

Garnier, E., Cordonnier, P., Guillerm, J. L., and L. Sonié. 1997. Specific leaf area and leaf nitrogen concentration in annual and perennial grass species growing in Mediterranean old-fields. Oecologia 111(4): 490-498.

Grime, J. P., Shacklock, J. M. L., and S.R. Band. 1985. Nuclear DNA contents, shoot phenology and species co-existence in a limestone grassland community. New Phytologist 100(3):435-445.

Grime, J. P., Thompson, K., Hunt, R., Hodgson, J. G., Cornelissen, J. H. C., Rorison, I. H., and R.E. Booth. 1997. Integrated screening validates primary axes of specialisation in plants. Oikos 259-281.

Grotkopp, E., Rejmanek, M., Sanderson, M.J., and T.L. Rost. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree alayses. Evolution. 58:1705-1729.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., and J.F. Wednel. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Research. 16(10): 1252-1261.

Heiser, C.B., D.M. Smith, S.B. Clevenger, and W.C. Martin. 1969. The North American sunflowers (*Helianthus*). Memoirs of the Torrey Botanical Club 22:1-218.

Hunt, R. 1982. Plant growth curves. the functional approach to plant growth analysis. Plant Growth Curves. Edward Arnold Ltd .

Jiang, J. and S.D, Clouse. 2001. Expression of a plant gene with sequence similarity to animal TDF-β receptor interacting protein is regulated by brassinosteroids and required for hormonal plant development. The Plant Journal 26(1):35-45.

Kane, N.C., J.M. Burke, L. Marek, G. Seiler, F. Vear et al., 2013. Sunflower genetic, genomic and ecological resources. Molecular Ecology Resources 13 (1):10-20.

Kawakami, T., P. Dhakal, A.N. Katterhenry, C.A. Heatherington, and M.C. Ungerer, 2011 Transposable Element Proliferation and Genome Expansion Are Rare in Contemporary Sunflower Hybrid Populations Despite Widespread Transcriptional Activity of LTR Retrotransposons. Genome Biology and Evolution 3:156-167.

Knight, C. A., Molinari, N. A., and D.A. Petrov. 2005. The large genome constraint hypothesis: evolution, ecology and phenotype. Annals of Botany 95(1): 177-190.

Knight, C.A. and J.M. Beaulieu. 2008. Genome size scaling through phenotype space. Annals of Botany 101(6):759-66.

Leitch, I.J., and A.R. Leitch, 2013 Genome size diversity and evolution in land plants, pp. 307-322 in Plant Genome Diversity Volume 2, edited by I.J.L.e. al. Springer.

Lenth, R. V. 2012. Using the lsmeans Package. Pain 50(60): 70.

Lyndon, R. F. 2012. Plant development: the cellular basis. Vol. 3. Springer Science & Business Media.

Mason, C.M. and L.A. Donovan. 2015. Evolution of the leaf economics spectrum in herbs: Evidence from environmental divergences in leaf physiology across *Helianthus* (Asteraceae). Evolution 69:2705-2720.

Morgan, H.D., and M. Westoby. 2005. The relationship between nuclear DNA content and leaf strategy in seed plants. Annals of Botany. 96:1321-1330

Mowforth, M. A., and J.P. Grime. 1989. Intra-population variation in nuclear DNA amount, cell size and growth rate in *Poa annua* L. Functional Ecology 289-295.

Newton, R.J., Price, H.J., and R.S. Halliwell. 1989. Drought stress responses of *Microseris* species differing in nuclear-DNA content. American Journal of Botany 76(6):789-95.

Oliver, M.J., Petrov, D., Ackerly, D.D., Falkowski, P., and O.M.Schofield. 2007. The mode and tempo of genome size evolution in eukaryotes. Genome Research 17:594-601.

Orme, D. 2013. The caper package: comparative analysis of phylogenetics and evolution in R. R package version, 5(2).

Paine, C. E. T., Marthews, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A. and L. A. Turnbull. 2012. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. Methods in Ecology and Evolution, 3:245–256.

Paradis, E., Claude, J., and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289-290.

Pinheiro J, Bates D, DebRoy S, Sarkar D. 2009. nlme: Linear and nonlinear Mixed Effects Models [computer program]. Version R package version 3.1-86. .

Preston, K.A., and D.D. Ackerly. 2004. The evolution of allometry in modular organisms. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Oxford University Press, Oxford 80-106.

R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rieseberg, L.H., 1991. Homoploid Reticulate Evolution in *Helianthus* (Asteraceae) - Evidence from Ribosomal Genes. American Journal of Botany 78 (9):1218-1237.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., and J.L. Bennetzen. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768.

Schilling, E.E., C.R. Linder, R.D. Noyes, and L.H. Rieseberg. 1998. Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. Systematic Botany 23 (2):177-187.

Sims, L.E., and H.J. Price. 1985. Nuclear DNA content variation in Helianthus (Asteraceae). American Journal of Botany 72 (8):1213-1219.

Stearns, S.C. 1992. The evolution of life histories. Vol. 249. Oxford: Oxford University Press.

Steeves, T. A., and I. M. Sussex. 1989. Patterns in plant development. Cambridge University Press.

Stephens, J.D., W.L. Rogers, C.M. Mason, L.A. Donovan, and R.L. Malmberg. 2015. Species Tree Estimation of Diploid *Helianthus* (Asteraceae) Using Target Enrichment. American Journal of Botany 102 (6):910-920.

Sultan, S. E. 2000. Phenotypic plasticity for plant development, function and life history. Trends in plant science 5(12):537-542.

Timme, R.E., B.B. Simpson, and C.R. Linder. 2007. High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18s-26s ribosomal DNA external transcribed spacer. American Journal of Botany 94 (11):1837-1852.

Van't, J., and A.H. Sparrow. 1963. A relationship between DNA content, nuclear volume and minimum mitotic cycle time. Proc. Natl Acad Sci. 49:897-902.

Vitte, C. Panaud, O., and H. Quesneville. 2007. LTR retrotransposons in rice (*Oryza sativa* L.): recent burst amplifications followed by rapid DNA loss. BMC Genomics. 8(218).

Wakamiya, I., Newton, R. J., Johnston, J. S., and H.J. Price. 1993. Genome size and environmental factors in the genus *Pinus.* American Journal of Botany 80(11):1235-1241.

Wendel, J.F. 2000. Genome evolution in polyploids. Plant Molecular Biology. 42:225-249.

Westoby, M. , Falster D. S., Moles, A. T., Vesk, P. A., and I. J. Wright. 2002. Plant ecological strategies: some leading dimensions of variation between species. Annual Review of Ecology and Systematics 33: 125–159

Young, T.P., and C.K. Augspurger. 1991. Ecology and Evolution of Long-lived Semelparous Plants. TREE. 6(9):285-289.

Zanne, A. E., Chave, J.S., Wright, J., Sheremet'ev, S. N., Jactel, H., Baraloto, C., Cerabolini, B. et al. 2016. The global spectrum of plant form and function. Nature 529(7585):167-171.

Zanne, A.E and D.S. Falster. 2010. Plant functional traits- linkages among stem anatomy, plant performance and life history. New Phytologist 185(2):348-351.

Table 3.1 Species included in study, genome size data used for analysis and where it was generated.

Species	Accession	Location	Life cycle
H. petiolaris	PI 468807	New Mexico	annual
H. annuus	PI 468607	Utah	annual
H. praecox	PI 468851	Texas	annual
H. porteri	PI 649918	Georgia	annual
H. debilis	PI 468668	Florida	annual
H. argophyllus	PI 435635	Texas	annual
H. agrestis	PI 468416	Florida	annual
H. salicifolius	PI 664758	Kansas	perennial
H. nuttallii	PI 468796	Colorado	perennial
H. divaricatus	PI 435675	Oklahoma	perennial
H. arizonensis	PI 653549	Arizona	perennial
H. microcephalus	PI 664743	South Carolina	perennial
H. occidentalis	PI 435788	Missouri	perennial
H. giganteus	PI 547184	Illinois	perennial
H. maximiliani	Konza[2]	Kansas	perennial
H. cusickii	PI 649966	California	perennial
H. atrorubens	PI 468658	South Carolina	perennial
H. heterophyllus	PI 673162	Louisiana	perennial
H. carnosus	PI 649956	Florida	perennial
H. angustifolius	PI 673154	Louisiana	perennial

Note: [1]Bitter Lake National Wildlife Refuge Roswell; [2]Konza Biological Research Station, Manhattan, Kansas

Table 3.2 Results from a mixed linear model ANOVA for the effects of life cycle (perennial vs annual) on growth metrics.

Growth metric	df	*F*	*P*
Cell production rate	1, 17.9	0.9907	0.3196
Leaf expansion rate	1, 18.1	4.4857	0.0342
Growth rate	1, 12.6	0.6643	0.586
Root to shoot biomass ratio	1, 18.0	1.5113	0.2189
Biomass	1, 13.1	2.8769	0.0492

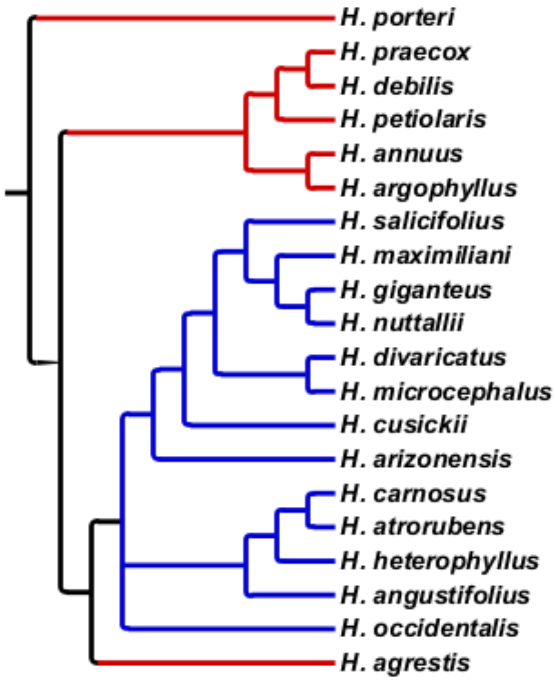

Figure 3.1 Phylogenetic relationships for species in this study. Phylogenetic tree is based on relationships presented in Stephens et al. (2015). Red branches= annual and blue branches = perennial

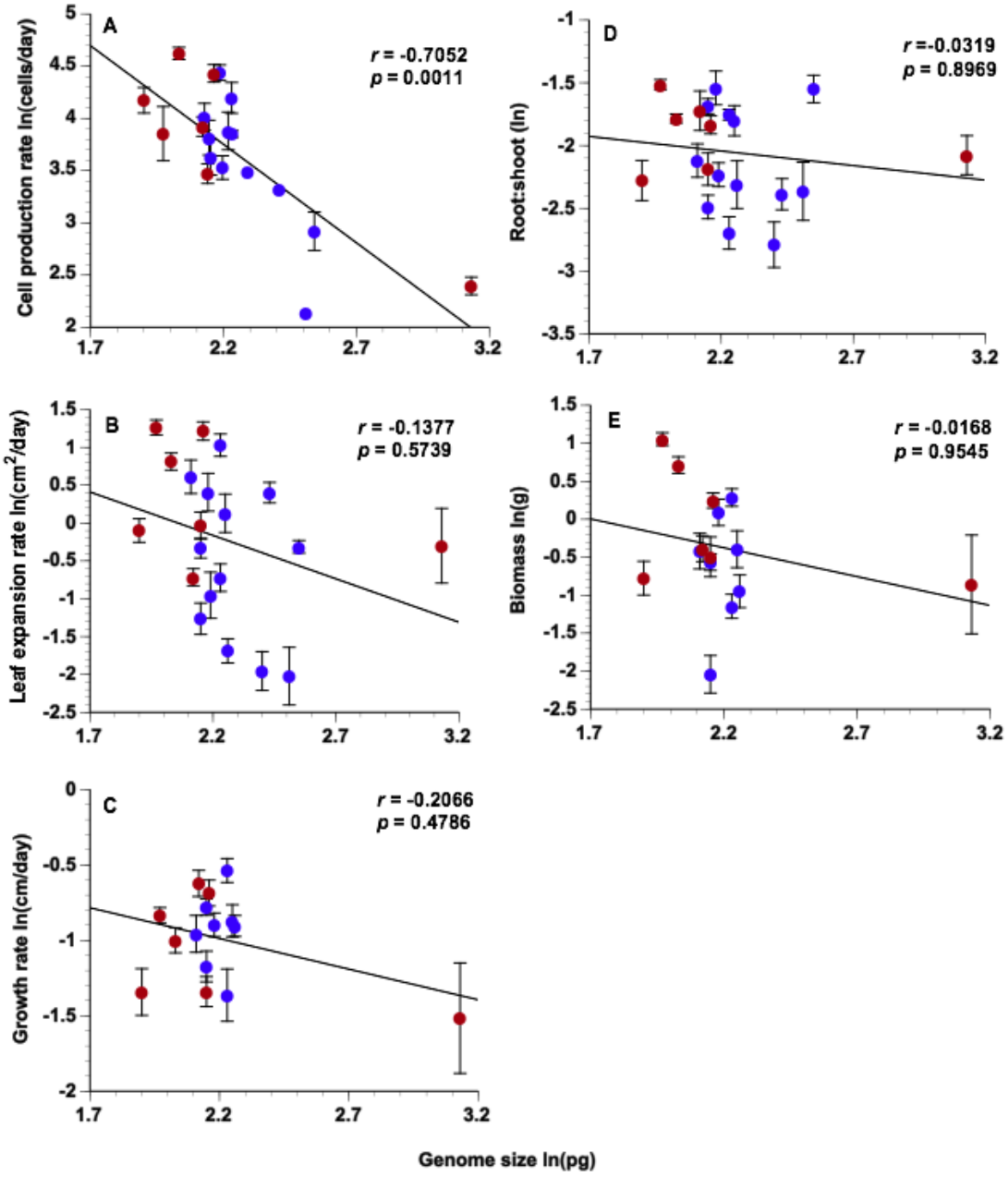

Figure 3.2 Correlations between genome size and (A) cell production rate, (B) leaf expansion rate, (C) growth rate, (D) root: shoot and (E) biomass. Error bars are (± SE). Red = annual and blue = perennial.

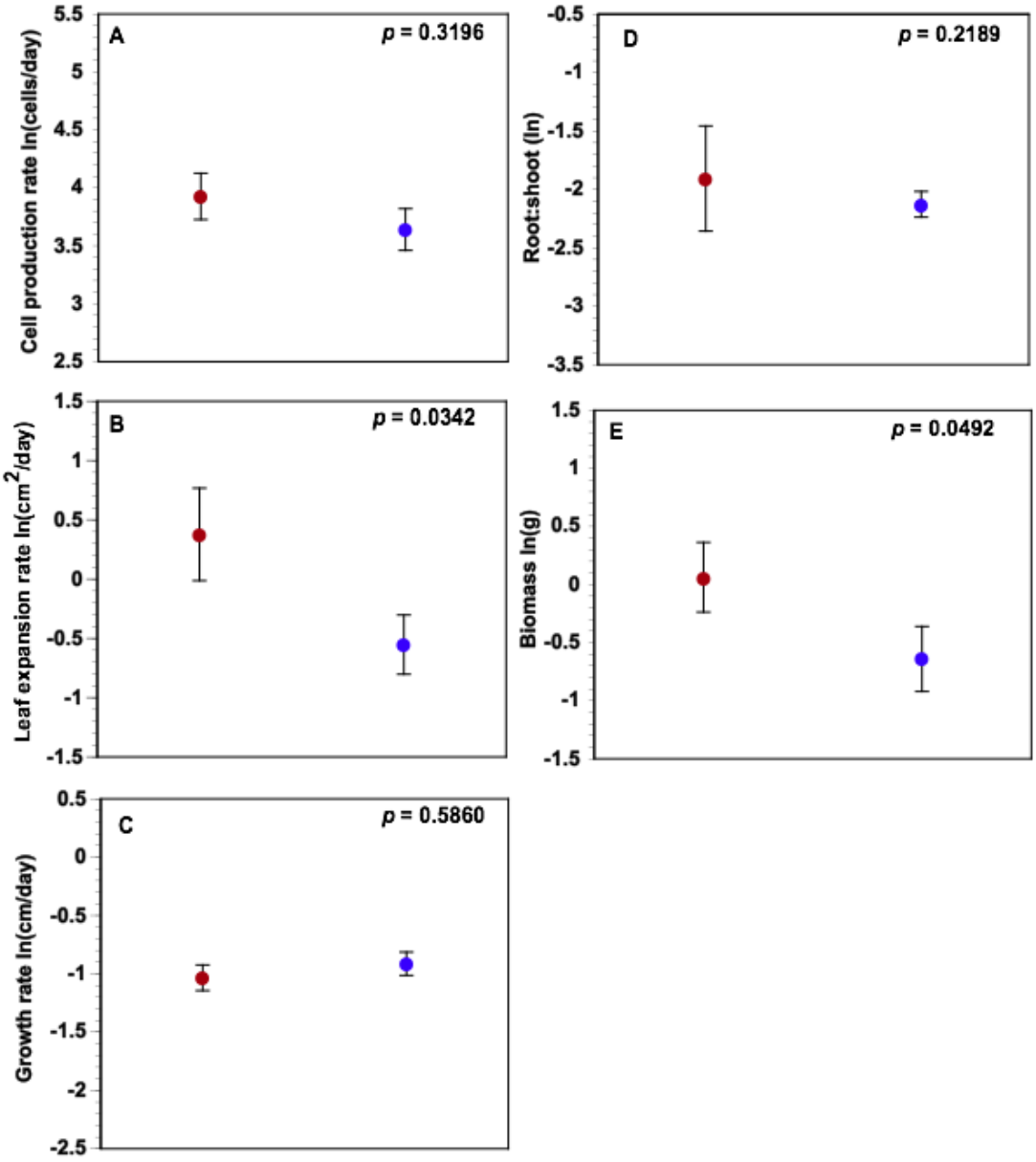

Figure 3.3 Mean annual (red) and perennial (blue) growth metrics for (A) cell production rate, (B) leaf expansion rate, (C) growth rate, (D) root:shoot and (E) biomass. Error bars are (± SE).

Chapter 4 - Low temperature tolerance in the perennial sunflower *Helianthus maximiliani*

This chapter is formatted for the journal "The American Midland Naturalist"

The citation for this chapter is: Tetreault, H.M., Kawakami, T., Levy, C., and M.C. Ungerer. 2016. Low temperature tolerance in the perennial sunflower *Helianthus maximiliani.* The American Midland Naturalist. 175(1):91-102.

Abstract

Species distributed across diverse climate and thermal conditions represent opportune systems for studying tolerance of low temperature stress. We examined variation in cold acclimation capacity and freezing tolerance among three natural populations (Texas, Kansas, and Manitoba) of the perennial sunflower species *Helianthus maximiliani*, originally collected across a 2134 km latitudinal transect in central North America. Tolerance to low temperatures was evaluated through leaf electrolyte leakage assays that quantify loss of cellular electrolytes into an aqueous medium due to plasma membrane damage. Freezing tolerance was highest for plants from the northernmost latitude (Manitoba population) under both non cold- acclimated and cold-acclimated experimental conditions. Individuals from Kansas and Texas populations exhibited lower freezing tolerance compared to Manitoba but did not differ from one another. All populations retain the ability to increase freezing tolerance through cold acclimation, and effects of cold acclimation actually trended greater in populations from warmer regions (Texas and

Kansas). Freezing tolerance of Manitoba X Texas F1 hybrids was statistically indistinguishable from plants from the Texas population, suggesting patterns of genetic dominance for alleles in Texas populations. Analysis of flowering specimens from herbaria records of corresponding regional locations indicates considerable variation in flowering phenology whereby flowering occurs progressively earlier with increasing latitude. This phenological variation may provide an additional mechanism of coping with low temperature stress through temporal avoidance.

KEYWORDS: abiotic stress; cold acclimation; electrolyte leakage; freezing tolerance; phenology; avoidance

Introduction

Freezing temperatures represent an important abiotic stress for plants and limit species distribution patterns and opportunities for dispersal and colonization (Woodward, 1987). Geographical variation in low temperature extremes can drive local adaptation within species as well, especially for taxa distributed broadly (Green, 1969; Casler *et al*., 2004; Saenz-Romero and Tapia-Olivares, 2008; Zhen and Ungerer, 2008a; Lee *et al*., 2012). Because of their sessile lifestyle and inability to escape ambient climate conditions, plants provide a powerful experimental system to examine strategies of coping with low temperature stress, with direct relevance to ecological and evolutionary population dynamics and agriculture.

For many temperate plant species, maximum freezing tolerance is enhanced via acclimation to low but nonfreezing temperatures. This phenomenon, known as cold acclimation, is marked by major cellular biochemical changes enabling plants to withstand temperatures

several degrees colder than non cold- acclimated controls (Guy, 1990; Xin and Browse, 2000; Iba, 2002; Smallwood and Bowles, 2002; Guy *et al.*, 2008). Cold- acclimation represents an inducible change and likely evolved in response to seasonal changes in which low, nonfreezing temperatures portend colder temperatures that are potentially more harmful. Genes and gene pathways regulating plant cold- acclimation have been identified in the model plant species *Arabidopsis thaliana* (Gilmour *et al*., 1998; Liu *et al*., 1998; Xin and Browse, 2000; Chinnusamy *et al*., 2003; Vogel *et al*., 2005; Agarwal et al., 2006; Van Buskirk and Thomashow, 2006; Doherty *et al*., 2009; Thomashow, 2010) with some genes shown to exhibit functional variability among natural *A. thaliana* accessions subjected to different historical selection pressures for freezing tolerance (Zhen and Ungerer, 2008b).

Avoidance mechanisms also represent feasible means through which organisms can cope with abiotic stress. Such mechanisms are relevant to plant species via seasonal growth patterns and reproductive timing events that minimize the probability of encountering stressful environments and/or conditions (Heide, 1994; Bennington and Mcgraw, 1995; Griffith and Watson, 2005; Heschel and Riginos, 2005). Tolerance and avoidance mechanisms of abiotic stress need not be mutually exclusive and both may function in natural plant populations (Geber and Dawson, 1997; Heschel and Riginos, 2005).

Helianthus maximiliani is a diploid perennial sunflower species that flowers in late summer-fall (Heiser *et al*., 1969; Schilling, 2006). Though distributed widely in North America, populations are found in highest concentration in mid-continental regions between Texas, U.S.A. and Manitoba, CA (Schilling, 2006). *Helianthus maximiliani* exhibits steep latitudinal clines in

multiple morphological and life history traits that are hypothesized to be driven by variation in photoperiod and climate (Kawakami *et al*., 2011). A role for natural selection in shaping clinal variation in this species is supported by patterns of phenotypic differentiation among populations that exceed neutral expectations estimated from putatively neutral molecular markers, *i.e*., Q_{st} — F_{st} analysis (Leinonen *et al*., 2013).

In the current report, we test whether *H. maximiliani* populations from three climatically diverse regions in central North America (Texas, Kansas and Manitoba) differ in their tolerance to freezing temperatures and whether effects of cold acclimation treatment differentially impacts freezing tolerance among these groups. We explore basic genetic aspects of cold acclimation in *H. maximiliani* by examining cold acclimation capacity and freezing tolerance of F1 hybrids derived from a cross between Manitoba and Texas plants. We demonstrate plants from the northernmost population (Manitoba) predictably display highest freezing tolerance under most treatment combinations but that cold acclimation effects persist (and are of greater magnitude under the experimental temperatures assayed) for plants from warmer regions (Texas and Kansas). Patterns of freezing tolerance of Manitoba x Texas F1 hybrids more closely resemble plants from Texas, suggesting dominance of alleles in Texas populations. We additionally examine patterns of *H. maximiliani* flowering phenology based on herbaria records in the native regions of these populations in light of corresponding seasonal temperature data. We reveal highly different phenology among populations that also may serve as an avoidance mechanism of low temperature stress.

Methods

Sunflower populations and growing conditions

Seeds from natural populations of *H. maximiliani* were collected in the field or obtained from the USDA National Plant Germplasm System (Table 4.1). Seeds obtained from the USDA are derived directly from wild- collected populations. Manitoba x Texas F1 hybrids were generated with a Manitoba individual serving as maternal parent and assayed for freezing tolerance alongside individuals from the natural populations. All plants were grown in 8 in pots with a 1:1 mixture of Metro- mix 350: all- purpose sand under a 14:10 h light:dark cycle and ambient temperature in the Kansas State University greenhouses. Watering was conducted daily or as needed and fertilization with a weak nutrient solution (N:P:K = 15:30:15) was provided once per week. Plants were randomly positioned across three 1.2 x 2.4 m greenhouse benches. Plants receiving a cold- acclimation treatment prior to assays of freezing tolerance were placed in a 4 C walk- in chamber for 6 d, where they experienced constant low ambient light conditions and were watered as needed.

Electrolyte leakage assay

Damage to plant tissue from exposure to freezing stress can be quantified through loss of cellular electrolytes into an aqueous medium due to damage to cellular plasma membranes (Sukumaran and Weiser, 1972). Terminal 5 cm sections of leaves were excised and placed in individual 50 ml plastic Corning tubes containing 1 ml of ddH2O with the cut side of the leaf sample facing the tube bottom. All tubes were kept on ice during sample collection. Samples were subjected to freezing temperatures in an ESPEC ESU-3CA Platinous series environmental

test chamber (ESPEC North America, Hudsonville, Michigan, USA). Samples from non cold-acclimated plants were subjected to freezing stress at -4 C and -5 C; samples from cold-acclimated plants were subjected to freezing stress at -5 C and -6 C. These temperatures were selected based on preliminary experiments surveying tolerances across different temperature ranges. Assays of non cold- acclimated and cold- acclimated individuals at the same temperature (*i.e.,* -5 C) enabled determination of the effects of cold- acclimation on freezing tolerance. To facilitate ice nucleation during periods of cooling, ice chips were added to Corning tubes when the chamber temperature reached -1 C. Minimum temperatures were maintained for 3 h and the duration of cooling (-1 C to experimental minimum temperature) and warming periods (experimental minimum temperature to 4 C) was 2.5 h and 5 h, respectively. Rates of temperature change thus ranged 1.2 – 2 C h^{-1} during cooling and 1.6 – 2 C h^{-1} during warming.

Following freezing stress treatments, samples were removed from the environmental chamber, ddH2O was added to fully cover the leaf samples, and tubes were placed on a platform shaker at 180 rpm for 24 h. The following day ion conductivity of the solution in each tube was measured using a Mettler Toledo FiveEasy conductivity meter (Mettler-Toledo, Columbus, Ohio, USA). All samples were measured twice and the mean of the two measurements used in subsequent analyses. Following initial measurements, samples were placed at -20 C for 24 hours to fully rupture cells and maximize freezing- based tissue damage. Samples were subsequently thawed and placed on a platform shaker at 180 rpm for 1 h. Ion conductivity was measured again by the same protocol, and the ratio of ion conductivities at the two temperatures (freezing stress assay temperature/ -20C) was used as a metric of relative electrolyte leakage for the sample. Lower and higher ratios of electrolyte leakage thus correspond to higher and lower tissue

freezing tolerance, respectively. For each acclimation temperature combination, 17 biological replicates of each genotype (*i.e.*, population or F1 hybrid) were assayed in two replicate sets (n=9 and n=8, sequentially). For all assays, fully expanded and healthy leaves of similar age were harvested from the central region of the plant stem.

Electrolyte leakage data were analyzed by mixed model analysis of variance (ANOVA) using JMP version 5.0.1a (SAS Institute, Cary, North Carolina, USA). Cold- acclimated and non cold- acclimated plants were tested over different temperature ranges and analyzed separately according to the model $y = u + P + T + P \square T + R[T] + R \square P[T]$, where P is *population*, T is *temperature*, and R is *replicate*. An additional model, $y = u + P + A + P \square A + R[A] + R \square P[A]$, was evaluated examining cold- acclimated and non cold- acclimated plants assayed at -5 C, where terms are the same as those above and where A represents *acclimation treatment*. Square brackets represent nested terms. *Replicate* was treated as a random effect with all other main effects treated as fixed.

Natural phenological and climate data

Historical flowering time data for *H. maximiliani* in the relevant collection locations were obtained from herbarium specimens, which can serve as a proxy for seasonal flowering times and flowering durations for locally collected samples. Data for Texas plants were obtained from The University of Texas at Austin Plant Resources Center (TEX-LL) in electronic format (www.biosci.utexas.edu/prc/databases.html), limiting the query to individuals in anthesis. Data for Kansas and Manitoba plants were collected manually by recording collection dates of herbarium specimens in flower, from the Kansas State University Herbarium (KSC) and the

University of Manitoba Herbarium (WIN), respectively. Only specimens collected within the state or province boundaries of the relevant herbaria were utilized and only one specimen per collector per collection date was retained for analysis. All flowering dates were converted to Julian calendar days.

Seasonal first frost information for Manitoba is based on climate information reported in Environment Canada (http://www.ec.gc.ca/). Corresponding data for Manhattan, Kansas and Austin, Texas were obtained from NOAA Satellite and Information Service (http://www.ncdc.noaa.gov/cdo-web/). These locations are centrally located relative to the collection locations of the specimens analyzed. Seasonal first frost data represent averages from 1951 to 1980, and define 'light freeze' events (29 to 32 F) with 50% possibility of frost occurring before or after.

Results

Freezing tolerance variation among populations

Populations of *H. maximiliani* from locations in Manitoba, Kansas and Texas experience appreciably different temperature conditions during the growing season and likely face different selection pressures for tolerance to low temperature. In assays of freezing-induced leaf electrolyte leakage, ANOVA revealed significant effects of Population and Temperature in analyses of non cold- acclimated plants (Table 4.2) and significant effects of Population and Replicate for cold- acclimated plants (Table 4.3). None of the interaction terms in either statistical model were significant. Significantly lower electrolyte leakage (higher tolerance) was observed for Manitoba plants versus other populations under non cold-acclimated conditions at

both -4 C and -5 C (Figure 4.1A). Under cold- acclimated conditions, significantly lower electrolyte leakage was observed for Manitoba plants versus Texas plants at -6 C, but significant differences were not observed among populations at -5 C (Figure 4.1B). F1 hybrids derived from a Manitoba x Texas inter-population cross displayed electrolyte leakage scores more similar to, and statistically indistinguishable from Texas plants across all acclimation and temperature treatments (Figure 4.1A, B). For all populations and acclimation conditions, colder temperatures resulted in higher electrolyte leakage scores (Figure 4.1A, B).

Effects of cold-acclimation on freezing tolerance

Assays of cold- acclimated and non cold- acclimated plants subjected to freezing stress at -5 C enabled examination of acclimation effects on leaf freezing tolerance and whether differences among populations exist with regard to this inducible response. Analysis of this subset of the data revealed significant effects of Population and Acclimation (Table 4.4), with all other effects nonsignificant. For all populations and the F1 hybrids, cold- acclimation resulted in lower leaf electrolyte leakage scores [higher tolerance] (Figure 4.2). Populations/F1 hybrids did not differ in the effects of cold- acclimation treatment on enhancing freezing tolerance (F = 2.1394; P = 0.1965 for the Population*Acclimation interaction; Table 4.4) although the difference in non cold- acclimated versus cold- acclimated leaf electrolyte leakage trended consistently higher for Kansas and Texas populations and F1 hybrids versus Manitoba (Figure 4.2). Indeed, differences were not observed among populations subjected to -5 C following cold acclimation (Figure 4.1B) whereas Kansas and Texas populations exhibited significantly higher electrolyte leakage values versus Manitoba at this same temperature under non cold acclimated conditions (Figure 4.1A).

Regional phenologies

Similar numbers of *H. maximiliani* records were analyzed from each of three herbaria: University of Manitoba-WIN (n=38), Kansas State University-KSC (n=35), and University of Texas-TEX-LL (n=28). Considerable variation in seasonal phenology was observed among locations, with earliest flowering dates recorded for Manitoba plants (mean = 220±4.9 Julian days, range = 182 to 252), followed by Kansas plants (mean = 251±5.1 Julian days, range = 195 to 299), and Texas plants (mean = 267±5.7 Julian days, range = 98 to 322). Flowering periods largely precede 30 y means of autumn first frost dates within but not across regions (*e.g.*, Texas or Kansas phenology data compared with Manitoba climate data or Texas phenology data compared with Kansas climate data) [Figure 4.3].

Discussion

Population variation in freezing tolerance

Ambient temperatures encountered by natural plant populations vary predictably with latitude and selection pressures for tolerance to low temperature are expected to be stronger for populations from colder climates (Dionne *et al.*, 2001; Shahba *et al*., 2003; Zhen and Ungerer, 2008a). We tested this prediction in the broadly distributed perennial sunflower species *H. maximiliani* and found that both with and without cold-acclimation pretreatments, plants from the highest latitude (Manitoba) exhibited the lowest levels of leaf electrolyte leakage (highest tolerance) compared to plants from Kansas and Texas populations, though not all comparisons were significant in post hoc tests.

No differences in leaf electrolyte leakage were observed between samples from Kansas and Texas populations for any of the experimental temperature/pretreatment combinations, suggesting that tolerance to low temperature for these populations is indistinguishable, at least under the conditions utilized in the current study. Lack of observed differences in this trait between plants from Kansas and Texas populations lies in contrast to large differences across the same transect for life history and size- related traits such as flowering time, dry biomass, and stem diameter (Kawakami *et al*., 2011). These results suggest that physiological, morphological, and life history characters may be subject to different selection pressures across the species range.

While plant freezing tolerance is a classic quantitative trait with a complex genetic basis (Thomashow, 1999), F1 hybrids derived from a cross between plants from Manitoba and Texas populations displayed levels of leaf electrolyte leakage more similar to and statistically indistinguishable from Texas plants, suggesting alleles of large effect from Texas plants exhibit dominance with respect to this phenotype. Similar patterns have been observed in segregating F2 hybrids between Manitoba and Texas *H. maximiliani* plants of the same populations for traits associated with plant architecture and growth rate (Kawakami *et al*., 2011). It is currently unknown whether these traits collectively have a shared genetic basis though genetic correlations for freezing tolerance and life history traits have been observed in other plant species (Agrawal *et al*., 2004).

Effects of cold-acclimation across populations

Cold- acclimation pretreatment of 4 C for 6 days resulted in reduced leaf electrolyte leakage (enhanced freezing tolerance) for all groups under study. Populations/F1 hybrids did not differ in the magnitude of this effect as determined by a nonsignificant Population*Acclimation interaction term ($F = 2.1394$, $P = 0.1965$; Table 4.4) although the difference in leaf electrolyte leakage under these different conditions trended lower for Manitoba plants versus others (Figure 4.2). This is an unexpected and interesting finding given that, in other plant species, individuals from colder environments have been shown to exhibit a stronger cold- acclimation response (Hannah *et al*., 2006). This result should be interpreted with caution, however, as the lesser ability of Manitoba plants to undergo cold- acclimation in the current study could be attributable to lower levels of leaf electrolyte leakage under non cold-acclimated conditions on account of higher intrinsic freezing tolerance for this population (Figure. 4.1A, 4. 2) and thus be an artifact of the particular assay conditions utilized. It is clear, however, that all assayed populations of *H. maximiliani* possess the relevant biochemical machinery to undergo this important physiological change. Populations from southern (warmer) climates have not lost this capacity, despite the potential for relaxed selection on cold-acclimation capacity in warmer climates (Zhen and Ungerer, 2008b).

Results presented here also contrast with previous reports describing an absence of cold acclimation capacity in domesticated varieties of the common sunflower *Helianthus annuus* (Hewezi *et al*., 2006; Allinne *et al*., 2009). It is currently unknown whether wild accessions of *H. annuus* also lack this capacity or if absence of this response in domesticated varieties could be a consequence of reduced genetic variability following domestication (Mandel *et al*., 2011). Given

similar widespread distributions of *H. maximiliani* and *H. annuus* and a relatively young age of the genus as a whole (Kane *et al*., 2013), the latter hypothesis seems more tenable, but further experiments are required to answer this question with certainty.

The genetic basis of cold- acclimation has not been investigated in *H. maximiliani*, although recent advances in next generation sequencing (NGS) have enabled cost- effective generation of genome-level resources for this nonmodel species. A recent analysis of transcriptomes of Manitoba and Texas populations of *H. maximiliani* (Kawakami *et al*., 2014) revealed homologs of several *Arabidopsis thaliana* genes known to be involved in plant cold-acclimation, including important key regulators such as the *C-repeat/dehydration responsive element binding factor 2* (*CBF2*) (Jaglo *et al*., 2001), *calmodulin binding transcription activator 3* (*CAMTA3*) (Doherty *et al*., 2009), and *inducer of CBF expression 1* (*ICE1*) (Chinnusamy *et al*., 2003). These and other identified homologs represent excellent ecological 'candidate loci' for future investigations of the molecular underpinnings of cold- acclimation and freezing tolerance variation among populations of *H. maximiliani.*

Phenological differences as a mechanism of abiotic stress avoidance

Natural history collections provide a useful resource for studying biological patterns and trends in nature (Lavoie and Lachance, 2006; Miller-Rushing *et al*., 2006; Robbirt *et al*., 2011; Panchen *et al*., 2012). Herbaria records of *H. maximiliani* collected from regional locations overlapping with the contemporary populations examined herein provide a rough approximation of historical flowering time information. Examined jointly with data on seasonal temperature change and averaged autumn first frost dates, it is possible to address whether regional

differences in flowering time may serve as an avoidance mechanism of low temperature stress. Earliest flowering in Manitoba populations and progressively later flowering in Kansas and Texas populations is consistent with differences in the length of the growing season in these locations and corroborates experimental flowering time data for this species under common garden conditions (Kawakami *et al*., 2011). For a given location, flowering records always precede historical autumn first frost dates (Figure 4.3). For comparisons across locations, however, this pattern is not upheld. For example, several Kansas specimens and the majority of Texas specimens were collected on dates following the average first frost for Manitoba populations. Similarly, multiple Texas specimens were collected on dates following the average first frost for Kansas populations (Figure 4.3).

While these results are suggestive of reproductive timing events that avoid low temperature stress, such observational data are not without caveats. First, herbarium records, while a rough approximation of the duration of flowering, are not an exact measure of the reproductive period. It is difficult to control and/or standardize among herbaria for factors such as the number of collectors, when those individuals were able to collect, and the number of years over which specimens were obtained. Second, herbarium specimens do not provide information on time requirements for seed maturation and thus production of viable seed sets; although it is noteworthy that a significant time lag exists between median flowering time and first frost dates for each location. Third, the time lag between autumn first frost dates for these locations and subsequent colder, and more damaging temperatures currently is unknown, as is whether differences exist among locations for such time lags.

Flowering phenologies for *H. maximiliani* populations prior to the onset of low, potentially damaging temperatures are consistent with experimental studies documenting phenotypic selection for earlier flowering in response to abiotic stress (Stanton *et al*., 2000; Griffith and Watson, 2005; Heschel and Riginos, 2005). Direct fitness consequences of adaptive (and phenotypically plastic) life history strategies are best studied, however, in reciprocal transplant experiments where genotypes can be directly compared under a range of native and non-native experimental conditions (Niewiarowski and Roosenburg, 1993; Angert and Schemske, 2005; Griffith and Watson, 2005; Miglia *et al*., 2005; Agren and Schemske, 2012; Bennington *et al*., 2012). Such studies clearly are warranted in *H. maximiliani.*

Conclusions

Populations of the perennial sunflower species *H. maximiliani* from across the native latitudinal range of the species differ in their physiological tolerance of low temperature stress as measured by leaf electrolyte leakage assays. Plants from the highest latitude population (Manitoba) exhibited higher freezing tolerance than those from lower latitude populations (Kansas and Texas). Differences in freezing tolerance were not observed between plants from Kansas and Texas populations, although plants from all populations retain the ability to undergo cold-acclimation. Flowering phenology data based on herbarium records analyzed with climate data from corresponding regional locations indicate that reproductive timing events also may serve as an avoidance mechanism of low temperature stress. Elucidation of the genetic underpinnings of the phenotypic variation described herein currently is unknown and will be the focus of future work. Ongoing genomic resource development in wild sunflowers (Kane *et al*., 2011; Kane *et al*., 2013; Kawakami *et al*., 2014) will greatly facilitate this effort.

Acknowledgments

We thank Ruth Timme for providing *H. maximiliani* seeds from the Texas population, Elizabeth Punter for providing data on Manitoba *H. maximiliani* herbarium specimens, and Carolyn Ferguson for comments on an earlier version of this manuscript. This work was supported by Kansas State University and NSF REU grant (DBI-0851835). This is contribution number 15-074-J from the Kansas Agricultural Experiment Station.

Literature Cited

AGARWAL, M., Y.J.HAO, A. KAPOOR, C.H. DONG, H. FUJII, X.W. ZHENG AND J.K. ZHU. 2006. A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem*., **281**:37636-37645.

AGRAWAL, A. A., J. K. CONNER AND J. R. STINCHCOMBE. 2004. Evolution of plant resistance and tolerance to frost damage. *Ecol. Lett.*, **7**:1199-1208.

AGREN, J. AND D.W. SCHEMSKE. 2012. Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytol.*, **194**:1112-1122.

ALLINNE, C., P. MAURY, A. SARRAFI AND P. GRIEU. 2009. Genetic control of physiological traits associated to low temperature growth in sunflower under early sowing conditions. *Plant Sci.*, **177**:349-359.

ANGERT, A.L. AND D.W. SCHEMSKE. 2005. The evolution of species' distributions: Reciprocal transplants across the elevation ranges of *Mimulus cardinalis* and *M. lewisii*. *Evolution,* **59**:1671-1684.

BENNINGTON, C. C., N. FETCHER, M. C. VAVREK, G. R. SHAVER, K. J. CUMMINGS AND J. B. MCGRAW. 2012. Home site advantage in two long-lived arctic plant species: results from two 30-year reciprocal transplant studies. *J. Ecol.*, **100**:841-851.

——— AND J.B. MCGRAW. 1995. Natural-selection and ecotypic differentiation in *Impatiens pallida*. *Ecol. Monogr.,* **65**:303-323.

CASLER, M. D., K. P. VOGEL, C. M. TALIAFERRO AND R. L. WYNIA. 2004. Latitudinal adaptation of switchgrass populations. *Crop Sci.,* **44**:293-303.

CHINNUSAMY, V., M. OHTA, S. KANRAR, B. H. LEE, X. H. HONG, M. AGARWAL AND J. K. ZHU. 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. *Gene Dev.,* **17**:1043-1054.

DIONNE, J., Y. CASTONGUAY, P. NADEAU AND Y. DESJARDINS. 2001. Freezing tolerance and carbohydrate changes during cold acclimation of green-type annual bluegrass (*Poa annua* L.) ecotypes. *Crop Sci.,* **41**:443-451.

DOHERTY, C.J., H. A. VAN BUSKIRK, S. J. MYERS AND M. F. THOMASHOW. 2009. Roles for *Arabidopsis* CAMTA Transcription Factors in Cold-Regulated Gene Expression and Freezing Tolerance. *Plant Cell*, **21**:972-984.

GEBER, M.A. AND T. E. DAWSON. 1997. Genetic variation in stomatal and biochemical limitations to photosynthesis in the annual plant, *Polygonum arenastrum*. *Oecologia*, **109**:535-546.

GILMOUR, S. J., D. G.ZARKA, E. J. STOCKINGER, M. P. SALAZAR, J. M. HOUGHTON AND M. F. THOMASHOW. 1998. Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *Plant J*., **16**:433-442.

GREEN, J.W. 1969. Temperature responses in altidudinal populations of *Eucalyptus pauciflora* Sieb Ex Spreng. *New Phytol.*, **68**:399-410.

GRIFFITH, T. M. AND M. A. WATSON. 2005. Stress avoidance in a common annual: reproductive timing is important for local adaptation and geographic distribution. *J. Evolution. Biol.,* **18**:1601-1612.

GUY C., F. KAPLAN, J. KOPKA, J. SELBIG, AND D.K. HINCHA. 2008. Metabolomics of temperature stress. *Physiol. Plantarum,* **132**:220-235.

———. 1990. Cold-Acclimation and Freezing Stress Tolerance - Role of Protein-Metabolism. *Annu. Rev. Plant Phys.,* **41**:187-223.

HANNAH, M. A., D. WIESE, S. FREUND, O. FIEHN, A. G. HEYER, AND D. K. HINCHA. 2006. Natural genetic variation of freezing tolerance in *Arabidopsis*. *Plant Physiol.,* **142**:98-112.

HEIDE, O. M. 1994. Control of Flowering and Reproduction in Temperate Grasses. *New Phytol.,* **128**:347-362.

HEISER, C. B., D. M. SMITH, S. B. CLEVENGER, AND W. C. MARTIN. 1969. The North American sunflowers (*Helianthus*). *Mem. Torrey Bot. Club,* **22**:1-218.

HESCHEL, M. S. AND C. RIGINOS. 2005. Mechanisms of selection for drought stress tolerance and avoidance in *Impatiens capensis* (Balsaminacea). *Am. J. Bot.***, 92**:37-44.

HEWEZI, T., M. LEGER, W. EL KAYAL, AND L. GENTZBITTEL. 2006. Transcriptional profiling of sunflower plants growing under low temperatures reveals an extensive down-regulation of gene expression associated with chilling sensitivity. *J. Exp. Bot.,* **57**:3109-3122.

IBA, K. 2002. Acclimative response to temperature stress in higher plants: Approaches of gene engineering for temperature tolerance. *Annu. Rev. Plant Biol.*, **53**:225-245.

JAGLO, K. R., S. KLEFF, K. L. AMUNDSEN, X. ZHANG, V. HAAKE, J. Z. ZHANG, T. DEITS, AND M. F. THOMASHOW. 2001. Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiol*., **127**:910-917.

KANE, N. C., J. M. BURKE, L. MAREK, G. SEILER, F. VEAR, G. BAUTE, S. J. KNAPP, P. VINCOURT, AND L. H. RIESEBERG. 2013. Sunflower genetic, genomic and ecological resources. *Mol. Ecol. Resour.,* **13**:10-20.

———, N. GILL, M. G. KING, J. E. BOWERS, H. BERGES, J. GOUZY, E. BACHLAVA, N. B. LANGLADE, Z. LAI, M. STEWART, J. M. BURKE, P. VINCOURT, S. J. KNAPP, AND L. H. RIESEBERG. 2011. Progress towards a reference genome for sunflower. *Botany,* **89**:429-437.

KAWAKAMI, T., B. J. DARBY, AND M. C. UNGERER. 2014. Transcriptome resources for the perennial sunflower *Helianthus maximiliani* obtained from ecologically divergent populations. *Mol. Ecol. Resour.,* **14**:812-819.

———, T. J. MORGAN, J. B. NIPPERT, T. W. OCHELTREE, R. KEITH, P. DHAKAL, AND M. C. UNGERER. 2011. Natural selection drives clinal life history patterns in the perennial sunflower species, *Helianthus maximiliani*. *Mol. Ecol.,* **20**:2318-2328.

LAVOIE, C. AND D. LACHANCE. 2006. A new herbarium-based method for reconstructing the phenology of plant species across large areas. *Am. J. Bot.,* **93**:512-516.

LEE, Y. P., A. BABAKOV, B. DE BOER, E. ZUTHER, AND D. K. HINCHA. 2012. Comparison of freezing tolerance, compatible solutes and polyamines in geographically diverse collections of *Thellungiella* sp. and *Arabidopsis thaliana* accessions. *BMC Plant Biol.,* **12**: 131.

LEINONEN, T., R. J. S. MCCAIRNS, R. B. O'HARA, AND J. MERILA. 2013. Q(ST)-F-ST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.,* **14**:179-190.

LIU, Q., M. KASUGA, Y. SAKUMA, H. ABE, S. MIURA, K. YAMAGUCHI-SHINOZAKI, AND K. SHINOZAKI. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell,* **10**:1391-1406.

MANDEL, J. R., J. M. DECHAINE, L. F. MAREK, AND J. M. BURKE. 2011. Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *Theor. Appl. Genet*., **123**:693-704.

MIGLIA, K. J., E. D. MCARTHUR, W. S. MOORE, H. WANG, J. H. GRAHAM, AND D. C. FREEMAN. 2005. Nine-year reciprocal transplant experiment in the gardens of the basin and mountain big sagebrush (*Artemisia tridentata*: Asteraceae) hybrid zone of Salt Creek Canyon: the importance of multiple-year tracking of fitness. *Biol. J. Linnean. Soc.,* **86**:213-225.

MILLER-RUSHING, A. J., R. B. PRIMACK, D. PRIMACK, AND S. MUKUNDA. 2006. Photographs and herbarium specimens as tools to document phenological changes in response to global warming. *Am. J. Bot.,* **93**:1667-1674.

NIEWIAROWSKI, P. H. AND W. ROOSENBURG. 1993. Reciprocal transplant reveals sources of variation in growth rates of the lizard *Sceloporus undulatus*. *Ecology*, **74**:1992-2002.

PANCHEN, Z. A., R. B. PRIMACK, T. ANISKO, AND R. E. LYONS. 2012. Herbarium Specimens, Photographs, and Field Observations Show Philadelphia Area Plants Are Responding to Climate Change. *Am. J. Bot.,* **99**:751-756.

ROBBIRT, K. M., A. J. DAVY, M. J. HUTCHINGS, AND D. L. ROBERTS. 2011. Validation of biological collections as a source of phenological data for use in climate change studies: a case study with the orchid *Ophrys sphegodes*. *J. Ecol.,* **99**:235-241.

SAENZ-ROMERO, C. AND B. L. TAPIA-OLIVARES. 2008. Genetic variation in frost damage and seed zone delineation within an altitudinal transect of *Pinus devoniana* (*P. michoacana*) in Mexico. *Silvae Genet.,* **57**:165-170.

SCHILLING, E. E. 2006. Helianthus, p. 141-169. *In:* Committee FoNAE (ed.). Flora of North American North of Mexico. 16+ vols. New York and Oxford.

SHAHBA, M. A., Y. L. QIAN, H. G. HUGHES, D. CHRISTENSEN, AND A. J. KOSKI. 2003. Cold hardiness of saltgrass accessions. *Crop Sci.,* **43**:2142-2147.

SMALLWOOD, M. AND D. J. BOWLES. 2002. Plants in a cold climate. *Philos. T. R. Soc. Lond. B.,* **357**:831-847.

STANTON, M. L., B. A. ROY, AND D. A. THIEDE. 2000. Evolution in stressful environments. I. Phenotypic variability, phenotypic selection, and response to selection in five distinct environmental stresses. *Evolution,* **54**:93-111.

SUKUMARAN, N. P. AND C. J. WEISER. 1972. An excised leaflet test for evaluating potato frost tolerance. *HortScience,* **7**:467-468.

THOMASHOW, M. F. 2010. Molecular Basis of Plant Cold Acclimation: Insights Gained from Studying the CBF Cold Response Pathway. *Plant Physiol.,* **154**:571-577.

———. 1999. PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. *Annu. Rev. Plant Phys.,* **50**:571-599.

VAN BUSKIRK, H. A. AND M. F. THOMASHOW. (Blackwell Publishing co-authors). 2006. *Arabidopsis* transcription factors regulating cold acclimation. *Physiol. Plantarum,* **126**:72-80.

VOGEL, J. T., D. G. ZARKA, H. A. VAN BUSKIRK, S. G. FOWLER, AND M. F. THOMASHOW. 2005. Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. *Plant J.,* **41**:195-211.

WOODWARD, F. I. 1987. Climate and Plant Distribution. Cambridge: Cambridge University Press.

XIN, Z. AND BROWSE J. 2000. Cold comfort farm: the acclimation of plants to freezing temperatures. *Plant Cell Environ.,* **23**:893-902.

ZHEN, Y. AND M. C. UNGERER. 2008a. Clinal variation in freezing tolerance among natural accessions of *Arabidopsis thaliana*. *New Phytol.,* **177**:419-427.

——— AND ———. 2008b. Relaxed Selection on the CBF/DREB1 Regulatory Genes and Reduced Freezing Tolerance in the Southern Range of *Arabidopsis thaliana*. *Mol. Biol. Evol.,* **25**:2547-2555.

Table 4.1 Populations assayed in the current study

Population	State/Province	Lat. (N), Long. (W)	Accession
Man-5	Manitoba	49.486, 100.533	PI 592335[1]
KS	Kansas	39.100, 96.580	MCU-KS[2]
TX-2	Texas	30.422, 97.592	RT-TX[2]

Note: [1]USDA-National Plant Germplasm System accession ID (http://www.ars-grin.gov/npgs/); [2]see Kawakami *et al*. (2011)

Table 4.2 ANOVA results for nonacclimated plants assayed for leaf electrolyte leakage at -4 C and -5 C

Source	d.f.	SS	MS	*F*	*P*
Population	3	2.7622	0.9207	17.1067	0.0024
Temperature	1	2.6820	2.6820	18.8207	0.0492
Population*Temperature	3	0.1249	0.0416	0.7737	0.5497
Replicate[Temperature]	2	0.2850	0.1425	2.6476	0.1499
Replicate*Population[Temperature]	6	0.3229	0.0538	0.8102	0.5640
Error	120	7.9717	0.0664		

Table 4.3 ANOVA results for cold- acclimated plants assayed for leaf electrolyte leakage at -5 C and -6 C

Source	d.f.	SS	MS	*F*	*P*
Population	3	0.6966	0.2322	5.1238	0.0430
Temperature	1	1.9210	1.9210	1.9651	0.2960
Population*Temperature	3	0.0949	0.0317	0.6983	0.5864
Replicate [Temperature]	2	1.9552	0.9776	21.5707	0.0018
Replicate *Population[Temperature]	6	0.2719	0.0453	1.1999	0.3110
Error	120	4.5323	0.0378		

Table 4.4 ANOVA results for non acclimated and cold- acclimated plants assayed for leaf electrolyte leakage at -5 C

Source	d.f.	SS	MS	*F*	P
Population	3	1.3301	0.4434	8.7946	0.0129
Acclimation	1	4.2431	4.2431	21.9278	0.0427
Population*Acclimation	3	0.3236	0.1079	2.1394	0.1965
Replicate [Acclimation]	2	0.3870	0.1935	3.8384	0.0844
Replicate *Population[Acclimation]	6	0.3025	0.0504	0.8617	0.5253
Error	120	7.0204	0.0585		

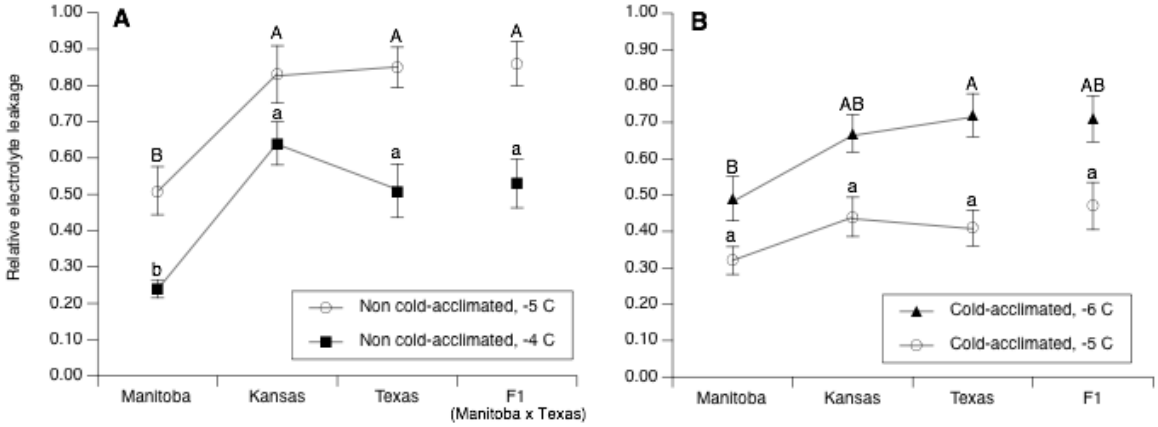

Figure 4.1 Leaf electrolyte leakage scores at -4 C and -5 C under non cold- acclimated conditions (**A**), and at -5 C and -6 C under cold- acclimated conditions (**B**). Within a given temperature treatment, values with different letters are significantly different.

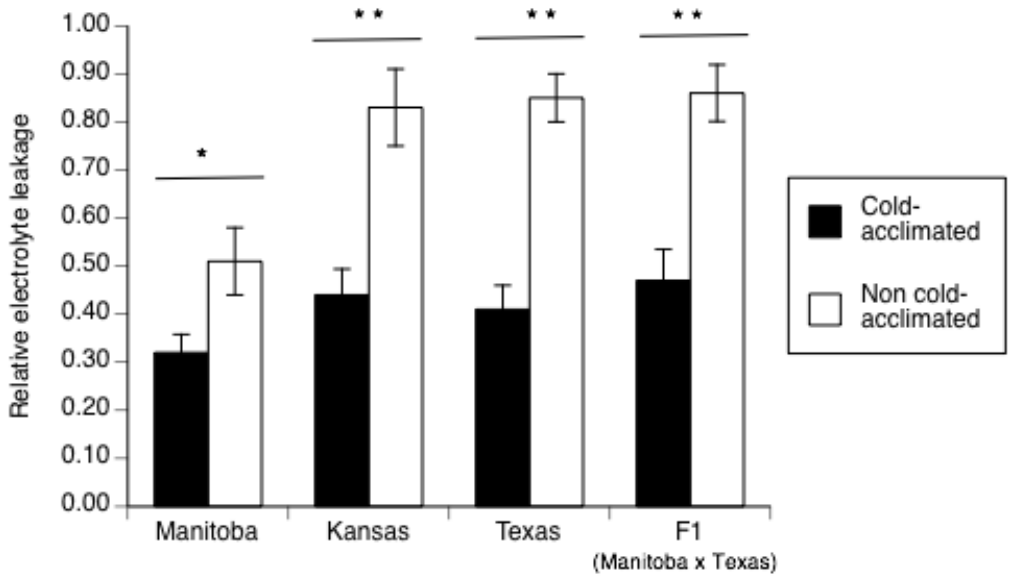

Figure 4.2 Leaf electrolyte leakage scores at -5 C under non cold- acclimated (white bars) and cold- acclimated (black bars) conditions; *, $P < 0.05$; **, $P < 0.01$.

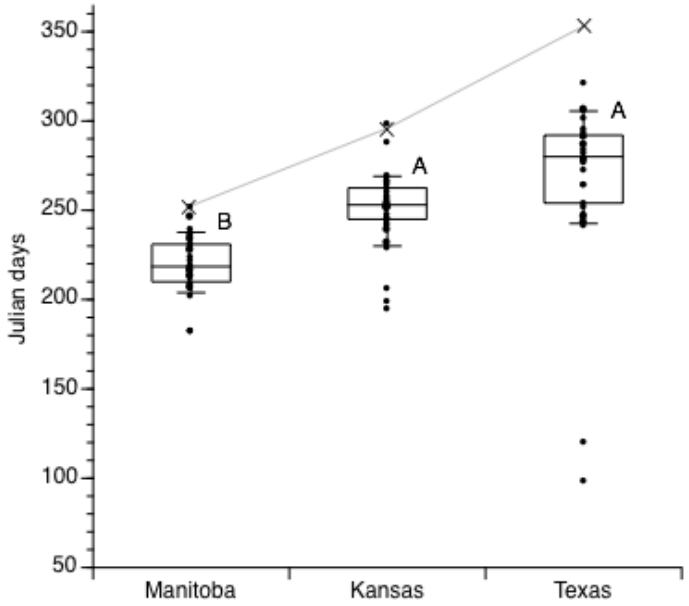

Figure 4.3 Natural flowering times (box plots) and autumn first frost dates (× symbols) for regional locations of populations used in this study. Filled circles represent collection dates of individual specimens.

Chapter 5 - Epilogue

It is well established that transposable elements (TEs) and in particular, long terminal repeat (LTR) retrotransposons, contribute to plant genome size and variation among species. The mobility and amplification of TEs represent a major source for genomes to differ immensely in the number of TE copies, the level of TE activity, and the diversity of TE families and sublineages (Hawkins et al. 2006; Piegu et al. 2006; Feschotte 2008; Bonchev and Parisod 2013). TEs can facilitate rapid genomic and phenotypic changes which potentially lead to both adaptation and species divergence (McClintook 1984; Ungerer et al. 2006; Renaut et al. 2014). The nature of TEs and their complex arrangements in the genome made it difficult to investigate diversity and abundance, especially to a level of sublineage (i.e., variants of superfamily *gypsy* LTR retrotransposon), with previous methods of marker-based, hybridization or PCR techniques. However, with advances in next-generation sequencing (NGS) technology and development of bioinformatic tools, studies on TE evolution are more informative and efficient. For example, Chapter 2 demonstrates how meaningful comparisons on TE dynamics across non-model species can be made utilizing new bioinformatics tools and sequencing technology, even with low genomic coverage of sequence reads (0.23-0.68x).

This work utilizes wild sunflower species in the genus *Helianthus* to examine the role of TEs in generating genome size variation and whether variation in genome size influences aspects of plant growth and development across multiple organizational levels. The genus *Helianthus* provides an excellent system for studying these questions given four-fold variation in nuclear DNA content among diploid species and well-resolved phylogenetic relationships. The primary

motivation for this work is to provide meaningful comparisons of TE content and abundances across species and explore if the existing genome size variation impacts plant growth and development.

Genome size variation is attributable to transposable elements among *Helianthus* species

Differential abundance and proliferation of TEs in flowering plants is identified as a significant contributor to genome size variation among plants, with LTR retrotransposons documented as the most abundant. In chapter 2, I explore the contribution of LTR retrotransposons to genome size variation among eight diploid *Helianthus* species and an outgroup species, *Phoebanthus tenuifolius.* Genome sizes were estimated using flow cytometry and show the species under investigation consist of approximately 4-fold variation. By utilizing genomic short read sequence data combined with sequence information from a panel of full length *gypsy* and *copia* LTR retrotransposons from *H. annuus* in a *de novo* graph based clustering approach, I show the contribution of LTR retrotransposons to genome size variation among species in this study. There is a strong positive correlation between genome size and repetitive fraction, concluding that repetitive DNA plays an important role in the underlying genome size variation in this group. Among these *Helianthus* species, *gypsy* is at greater abundance than *copia* elements, reinforcing previous sunflower work that the LTR retrotransposon superfamily *gypsy* is at greater abundance than *copia* among other sunflower species (Kawakami et al. 2010; Staton et al. 2012; Ungerer and Kawakami 2013). The observed overall stability in abundance for sublineages within each superfamily across *Helianthus* species suggest shared ancestry of *gypsy* and *copia* sublineages. Interestingly, species with larger

genomes, *H. agrestis* and *P. tenuifolius,* have signatures of higher read densities for a small number of sublineages, which likely reflects lineage specific amplifications contributing to genome size expansion. This pattern of a small number of LTR retrotransposon lineages underlying genome size differences is observed in other plant genera, i.e. cotton and rice (Hawkins et al. 2006; Piegu et al. 2006).

This study is informative on how TEs contribute to genome size variation and dynamics of these elements in the genomes of *Helianthus* but warrants further research regarding the mechanisms that allow particular sublineages to proliferate. It also remains unclear what processes are most important in the pattern of genome size variation among *Helianthus* species, i.e., genome size reduction from processes such as illegitimate recombination and or unequal intrastrand homologous recombination (Devos et al. 2002; Vitte and Panaud 2003). Therefore, further research on the mechanisms underlying genome size variation would increase our understanding on how TEs contribute to genome evolution in plants, whether it is genome downsizing (DNA loss), or if it is truly genome expansion through recent proliferation events as suggested here.

Biological consequence of genome size variation among *Helianthus* species

With approximately 2400-fold variation in nuclear DNA content among plants (Leitch and Leitch 2013) an interest for plant evolutionary biologists is the biological consequence of genome size. There is strong evidence supporting genome size effect on cell level traits i.e., cell size and cell division rates, where larger genomes have larger cells and longer durations of cell division (Bennett 1971; Bennett 1972; Bennett 1977; Francis et al.. 2008). In this dissertation, I

assessed whether genome size has an effect on phenotypic traits at multiple levels, cell (cell production rate), organ (leaf expansion rate) and whole-plant (growth rate and biomass accumulation) among diploid perennial and annual sunflowers with a four-fold difference in genome size (Chapter 3).

I found that genome size has a greater effect on cell production rate (cell level) but not at the organ or whole-plant level of development. Genome size among sunflower species is strongly negatively correlated with cell production rate, which is consistent with observations in other studies. There was not a distinct relationship, however, between genome size and growth at higher organizational levels (organ and whole plant). Chapter 3 results suggest that at a certain level genome size does not affect growth. Although life cycle (perennial and annual) has a phylogenetic signal within *Helianthus,* I assessed how the measured traits compared between life cycle and found that there is a trend for life cycle being an important factor on traits at the organ and whole plant level. This study begins to address the impact genome size has on growth and development but warrants further work in deciphering other factors associated with genome size among *Helianthus* species. For example, some studies predict that plants with larger genomes reside in temperate environments with a narrow range (reviewed in Greilhuber and Leitch 2012). *Helianthus* species have an expansive range in North America, with species residing in different climates and ranges. Further work on habitat and genome size could present interesting patterns on genome size distribution. Another element to consider, with habitat playing a key role in plant growth and development, is a study that excludes the environmental factor by examining plant traits and genome size within a region.

Published work on plant genomics has significantly increased with the advances in sequencing technology with most attention on economically important species such as maize, banana, tomato, barley and rice (Michael and Jackson 2013). Crop species tend to have some of the largest plant genomes and contain pronounced amounts of TEs, for example >85% of the maize genome and barley are composed of TEs (Tenallion et al. 2010). Identification and annotation of TE content in these crop species is of interest as TEs are important sources of functional variation in gene regulation (Sturder et al. 2011) and cause major chromosomal rearrangements (Lisch 2013). Work in this dissertation provides insight into the mechanisms contributing to genome size evolution in plants, specifically among *Helianthus* species, but the methods and concepts can be applied to other plant groups to gain a better understanding of genome evolution. Genome expansion in wild sunflowers is influenced largely by amplification of a small number of TEs and not necessarily by a greater diversity of TEs at low abundance (Chapter 2). Phentype traits for both growth and development eventually determine plant performance in terms of biomass and yield (Tsukaya 2003). By measuring aspects of growth and development at cell-, organ- and whole plant-levels I show how genome size impacts phenotypes at these different levels of organization. Briefly, genome size is negatively correlated with cell production rate (cell-level) but this relationship weakens for phenotypes at higher organizational levels, such as that of organ and whole-plant (Chapter 3). An integration of basic knowledge on how genome size impacts growth and development can enhance our ability to improve crop plants for our benefit.

Literature Cited

Bennett, M. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London B: Biological Sciences 181(1063):109-135.

Bennett, M.D. 1971. The duration of meiosis. Proceedings of the Royal Society of London Series B, Biological Sciences 178(1052):277-99.

Bennett, M.D. 1977. The time and duration of meiosis. Philosophical Transactions 277(955):201-26.

Bonchev, G., and C. Parisod, 2013 Transposable elements and microevolutionary changes in natural populations. Molecular Ecology Resources 13 (5):765-775

Devos, K.M., J.K. Brown, and J.L. Bennetzen, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res 12 (7):1075-1079.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nature Rev. Gen. 9, 397-405.

Francis, D., Davies, M.S., and P.B. Barlow. 2008. A strong nucleotypic effect of DNA C-vaue on the cell cycle regardless of ploidy level. Annals of Botany. 101:747-757.

Hawkins JS, Kim H, Nason JD, Wing RA, and Wendel JF. 2006. Differential lineagespecific amplification of transposable elements is responsible for genome size variation in Gossypium. Genome Res. 16(10), 1252-1261.

Hilbrict T, Varotto S, Sgaramella V, Bartels D, Salamini F, Furini A. 2008. Retrotransposons and siRNA have a role in the evolution of desiccation tolerance leading to resurrection of the plant Craterostigma plantagineum. New Phyt. 179, 877-887.

Kawakami, T., Strakosh, S.C., Zhen, Y., and M.C. Ungerer. 2010 Different scales of Ty1/copia-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. Heredity 104:341-350.

Leitch, I.J., and A.R. Leitch, 2013 Genome size diversity and evolution in land plants, pp. 307-322 in Plant Genome Diversity Volume 2, edited by I.J.L.e. al. Springer.

Lisch, D. 2013. How important are transposons for plant evolution? Nature Reviews:Genetics. 14:49-61.

McClintock B. 1984. The significance of responses of the genome to challenge. Science 226(4676), 792-801.

Michael, T.P. and S. Jackson. 2013. The first 50 plant genomes. The Plant Genome 6(2): doi: 10.3835/lantgenome2013.03.0001in

Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, et al. 2006. Doubling genome size without polyploidization: dynamics of retrotransposon mediated genome expansions in *Oryza australensis*, a wild relative of rice. Genome Res. 16, 1262-1269

Renaut, S., Rowe, H.C., Ungerer, M.C., and L.H. Rieseberg. 2014. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of LTR retrotransposons in hybrd sunflower. Phil. Transactions of the Royal Society B 369(1648): 20130345

Studer, A., Zhao, Q., Ross-Ibarra, J., and J. Doebley. 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nature Genetics 43(11):1160-1165.

Tenaillon, M.I., Hollister, J.D. and B.S. Gaut. 2010. A triptych of the evolution of plant transposable elements. Trends in Plant Science 15(8):471-478.

Tsukaya, H. 2003. Organ shape and size: a lesson from studies of leaf morphology. Current Opinions in Plant Biology 6(1):57-62.

Ungerer MC, Strakosh SC, and Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr. Biol. 16, 872-873.

Ungerer, M.C. and T. Kawakami. 2013 Transcriptional dynamics of LTR retrotransposons in early generation and ancient sunflower hybrids. Genome Biology and Evolution 5(2):329-337.

Vitte, C., and O. Panaud, 2003 Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L. Molecular Biology and Evolution 20 (4):528-540.

Appendix A - Supplemental Information for Chapter 2

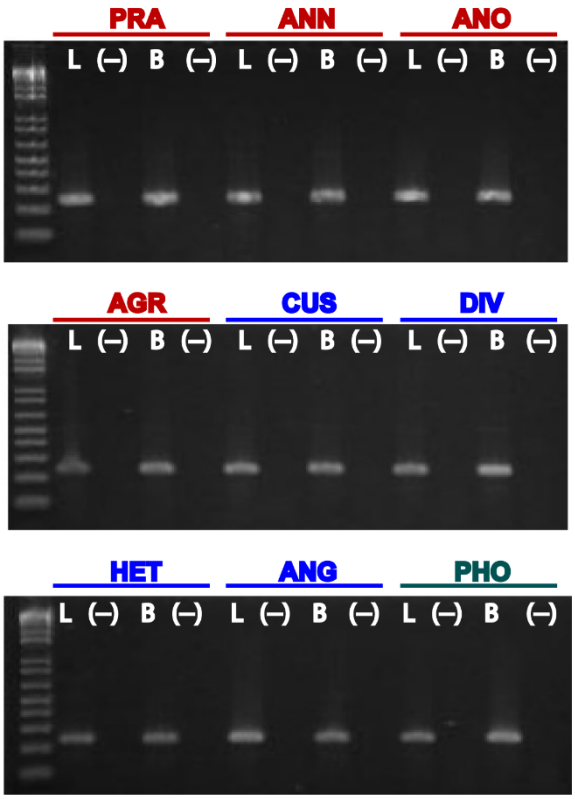

Figure A.5.1 Positive control RT-PCR assays of *Actin* in leaf (L) and bud (B) tissue. Minus signs in parentheses indicate lanes with negative control reactions. Species abbreviations are as in Table 1. Red = annual, blue = perennial, teal = perennial outgroup

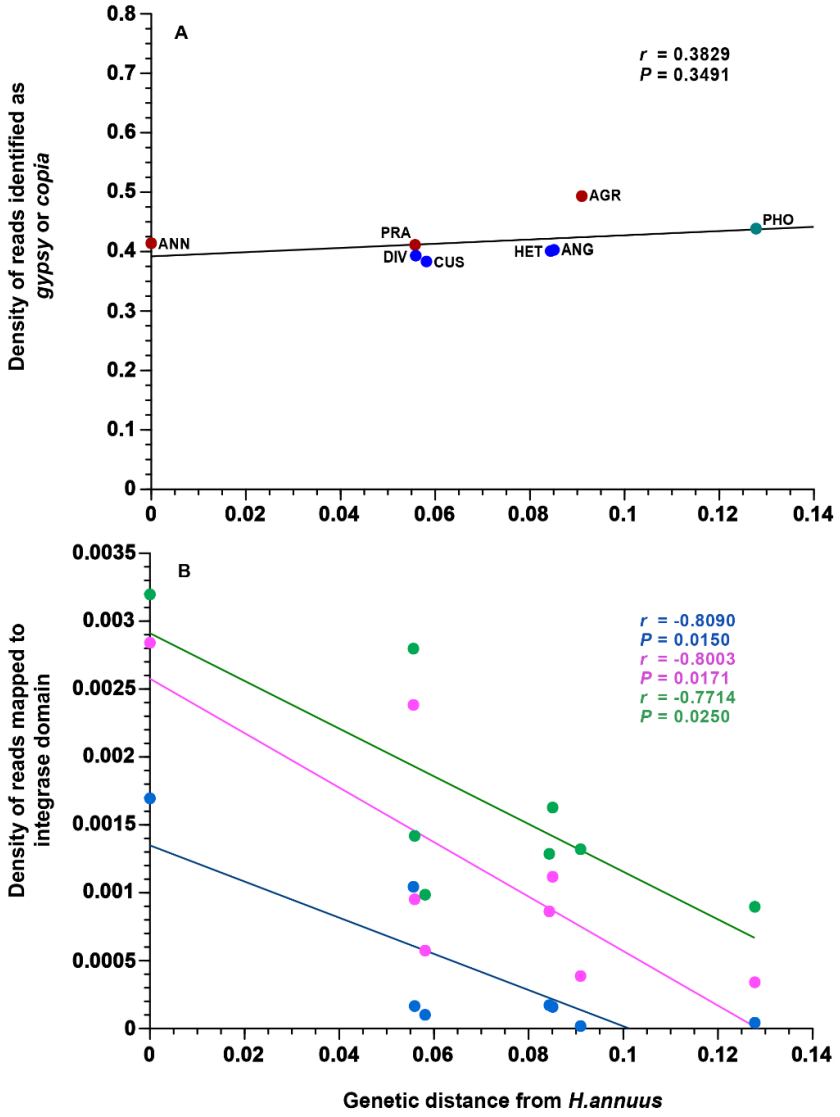

Figure A.5.2 Comparison of graph-based clustering (A) and mapping-based (B) approaches for identifying sequences derived from *gypsy* and *copia* elements across species using an LTR retrotransposon reference panel derived from *H. annuus*. The *x*-axis depicts genetic distance from *H. annuus* for species under investigation, using 15 (randomly selected) of 170 loci used to evaluate phylogenetic relationships among diploid species of *Helianthus* (Stephens et al. 2015). *H. anomalus* was not used in these analyses because of lack of inclusion in Stephens et al. (2015). For panel (A), reads were identified as *gypsy* or *copia* sequences based on methods outlined in main text. For panel (B), 6.7 million reads from each species were mapped to the *integrase* (*INT*) domains of 52 full-length LTR retrotransposons using BWA v0.7.6. Species abbreviations are as in Table 1. For panel (A): Red = annual, blue = perennial, teal = perennial outgroup; For panel (B): blue = 4 mismatches allowed, pink = 10 mismatches allowed, green = 20 mismatches allowed

Table A.5.1 Mean number of reads sampled and mean number of reads identified as belonging to different sublineages of *gypsy* (A) and *copia* (B) LTR retrotransposons based on five graph-based clustering analysis runs per dataset

A

Species	Reads sampled	A	B	C	X1	X2	E'	W	Y1	Y2	Z1	Z2	Total reads (%)
PRA	3,000,000	183,666	97,911	195,123	175,936	123,372	6,630	19	104,742	16,169	84,280	13,038	1,000,886 (33.4)
ANN	3,000,000	196,424	115,854	193,471	149,330	119,033	13,130	13	84,352	19,193	118,365	12,245	1,021,410 (34.0)
ANO	3,000,000	266,677	107,696	167,265	176,321	174,587	35,573	12	81,569	22,902	92,417	10,187	1,135,206 (37.8)
AGR	2,368,072	265,678	96,277	317,877	175,457	72,563	14,629	1	13,164	7,951	17,342	1,445	982,383 (41.5)
CUS	3,000,000	89,677	140,613	203,565	226,226	134,301	21,219	16	53,608	20,719	30,887	6,297	927,128 (30.9)
DIV	2,542,761	99,317	98,812	175,048	181,348	93,304	15,724	6	67,856	11,148	43,100	8,686	794,349 (31.2)
HET	3,000,000	65,433	152,037	236,482	237,938	143,884	20,590	7	53,147	20,482	55,798	6,506	992,304 (33.1)
ANG	3,000,000	111,205	158,886	243,769	159,213	129,122	13,282	11	56,587	16,277	58,418	7,843	954,613 (31.8)
PHO	2,954,112	102,775	90,264	179,378	374,050	151,686	23,661	3	61,673	21,394	46,052	4,066	1,055,002 (35.7)
													8,863,281 (34.3)

B

Species	Reads sampled	1	2	3	4	5	6	7	Total reads (%)
PRA	3,000,000	60,003	16,976	100,322	734	58	16,778	34,890	229,761 (7.7)
ANN	3,000,000	60,431	12,255	98,265	768	127	12,516	33,388	217,750 (7.3)
ANO	3,000,000	51,646	15,197	104,123	847	52	14,938	56,279	243,082 (8.1)
AGR	2,368,072	89,155	15,598	34,831	62	8	5,523	37,631	182,808 (7.7)
CUS	3,000,000	29,298	6,796	95,225	773	30	18,568	72,026	222,716 (7.4)
DIV	2,542,761	32,090	8,074	95,154	666	68	15,191	53,978	205,221 (8.1)
HET	3,000,000	55,516	11,669	75,266	889	57	14,136	52,686	210,219 (7.0)
ANG	3,000,000	45,092	12,680	141,930	340	49	9,260	44,310	253,661 (8.5)
PHO	2,954,112	81,920	13,673	74,526	320	2	12,634	54,640	237,715 (8.0)
									2,002,933 (7.7)

Table A.5.2 Primers utilized in RT-PCR assays

Region	Forward (5'→3')	Reverse (5'→3')
gypsy sublineage A	GRTGCTTTTCCCAGCYGTTG	TCGACTCACCAAGTCTGCAC
gypsy sublineage C	AAGTCAGCKCATTTYYTACCC	TTCCARAAATGWGACGTRTATCTTAGT
copia sublineage 1	TCTCAGAACCTCGGCAATCT	GGCGAGCAAAAGAGAAAATG
actin	AGATTCCGTTGCCCTGAGGT	CTCTCTGGAGGWGCAACCAC

Table A.5.3 Average amino acid divergence of *RT* domains within and between sublineages of *gypsy* (A) and *copia* (B) elements depicted in Figure 2.2.

A

.

	A	B	C	X1	X2	E'	W	Y1	Y2	Z1	Z2
A	0.046										
B	0.118	na									
C	0.152	0.169	0.045								
X1	0.150	0.108	0.169	0.056							
X2	0.188	0.147	0.204	0.146	0.132						
E'	0.223	0.212	0.191	0.217	0.200	na					
W	0.478	0.470	0.444	0.472	0.477	0.470	na				
Y1	0.619	0.614	0.627	0.634	0.629	0.621	0.561	na			
Y2	0.627	0.621	0.628	0.636	0.621	0.614	0.568	0.235	na		
Z1	0.655	0.667	0.653	0.659	0.663	0.660	0.614	0.557	0.606	0.015	
Z2	0.647	0.622	0.638	0.619	0.648	0.629	0.633	0.561	0.572	0.445	0.045

B

	1	2	3	4	5	6	7
1	0.027						
2	0.318	n.a.					
3	0.382	0.325	0.100				
4	0.622	0.615	0.609	n.a.			
5	0.641	0.644	0.609	0.552	n.a.		
6	0.628	0.632	0.617	0.619	0.636	n.a.	
7	0.620	0.611	0.615	0.569	0.607	0.531	n.a.

File S1 Reference panel of full-length *gypsy* and *copia* LTR retrotransposons derived from *H.annuus* genome. Fasta file link is below.
http://www.g3journal.org/content/suppl/2016/05/25/g3.116.029082.DC1/FileS1.zip

File S2 Amino acid sequences for the reverse-transcriptase (*RT*) domain of 52 (40 *gypsy* + 12 *copia*) full length LTR retrotransposon elements used for phylogenetic analysis in Figure 2.
http://www.g3journal.org/content/suppl/2016/05/25/g3.116.029082.DC1/FileS2.zip

Appendix B - Supplemental data for Chapter 3

Table B.1 Data for growth rate (cm/day), root : shoot biomass (g), total biomass (g) and leaf expansion rate(cm2/day) collected from plants grown in the greenhouse.

Species	Life cycle	Individual	Growth rate (cm/day)	Root : shoot biomass (g)	Total biomass (g)	Leaf expansion rate (cm^2/day)
H.agrestis	annual	1	0.18	0.19	0.40	0.68
H.agrestis	annual	2	0.18	0.19	0.40	1.37
H.agrestis	annual	3	0.58	0.12	2.07	3.07
H.agrestis	annual	4	0.10	0.07	0.10	0.10
H.agrestis	annual	5	0.58	0.12	2.72	1.91
H.agrestis	annual	6	0.29	0.15	0.65	1.47
H.agrestis	annual	7	0.34	0.21	0.80	NA
H.agrestis	annual	8	0.04	0.08	0.02	NA
H.angustifolius	perennial	1	0.06	0.30	0.47	0.81
H.angustifolius	perennial	2	0.07	0.11	0.32	0.53
H.angustifolius	perennial	3	0.12	0.29	0.90	1.10
H.angustifolius	perennial	4	0.10	0.21	0.50	0.83
H.angustifolius	perennial	5	0.05	0.17	0.35	0.70
H.angustifolius	perennial	6	0.07	0.16	0.22	0.60
H.angustifolius	perennial	7	0.05	0.22	0.20	0.60
H.angustifolius	perennial	8	0.04	0.29	0.41	0.80
H.angustifolius	perennial	9	0.03	0.61	0.23	NA
H.angustifolius	perennial	10	0.05	0.18	0.52	NA
H.angustifolius	perennial	11	0.06	0.19	0.23	NA
H.angustifolius	perennial	12	0.04	0.31	0.36	NA
H.angustifolius	perennial	13	0.02	0.14	0.11	NA
H.angustifolius	perennial	14	0.08	0.16	0.30	NA
H.angustifolius	perennial	15	0.05	0.15	0.27	NA
H.annuus	annual	1	0.35	0.30	2.04	2.92
H.annuus	annual	2	0.30	0.18	2.86	4.17
H.annuus	annual	3	0.45	0.18	3.01	4.30
H.annuus	annual	4	0.47	0.20	5.68	6.09
H.annuus	annual	5	0.39	0.19	2.66	4.08
H.annuus	annual	6	0.38	0.23	3.06	3.54
H.annuus	annual	7	0.49	0.20	1.39	2.32
H.annuus	annual	8	0.32	0.19	1.33	2.58
H.annuus	annual	9	0.62	0.25	2.45	3.14
H.annuus	annual	10	0.38	0.19	2.50	NA
H.annuus	annual	11	0.53	0.20	3.65	NA
H.annuus	annual	12	0.40	0.31	3.09	NA
H.annuus	annual	13	0.40	0.25	2.24	NA
H.annuus	annual	14	0.49	0.20	3.45	NA
H.annuus	annual	15	0.51	0.17	3.97	NA

H.annuus	annual	16	0.79	0.27	4.98	NA
H.annuus	annual	17	0.41	0.27	4.66	NA
H.annuus	annual	18	0.43	0.27	4.50	NA
H.annuus	annual	19	0.31	0.23	1.81	NA
H.annuus	annual	20	0.49	0.18	2.81	NA
H.annuus	annual	21	0.46	0.23	2.43	NA
H.argophyllus	annual	1	0.32	0.07	1.51	2.87
H.argophyllus	annual	2	0.59	0.14	4.01	4.64
H.argophyllus	annual	3	0.63	0.17	3.88	3.87
H.argophyllus	annual	4	0.74	0.25	6.38	5.69
H.argophyllus	annual	5	0.48	0.19	2.07	2.18
H.argophyllus	annual	6	0.52	0.16	2.28	2.91
H.argophyllus	annual	7	0.56	0.19	2.72	3.06
H.argophyllus	annual	8	0.33	0.17	1.73	2.02
H.argophyllus	annual	9	0.80	0.10	5.08	NA
H.argophyllus	annual	10	0.57	0.14	2.22	NA
H.argophyllus	annual	11	0.44	0.20	4.02	NA
H.argophyllus	annual	12	0.37	0.24	3.23	NA
H.argophyllus	annual	13	0.36	0.11	1.88	NA
H.argophyllus	annual	14	0.61	0.13	3.19	NA
H.argophyllus	annual	15	0.76	0.15	5.10	NA
H.argophyllus	annual	16	0.64	0.19	4.43	NA
H.argophyllus	annual	17	0.23	0.16	1.42	NA
H.argophyllus	annual	18	0.54	0.23	2.61	NA
H.argophyllus	annual	19	0.55	0.18	2.69	NA
H.arizonensis	perennial	1	0.43	0.11	0.13	0.26
H.arizonensis	perennial	2	0.36	0.11	0.11	0.27
H.arizonensis	perennial	3	0.48	0.06	0.23	0.37
H.arizonensis	perennial	4	0.36	0.07	0.13	0.08
H.arizonensis	perennial	5	0.35	0.06	0.14	0.21
H.arizonensis	perennial	6	0.52	0.07	0.36	0.44
H.arizonensis	perennial	7	0.34	0.05	0.25	0.42
H.arizonensis	perennial	8	0.27	0.10	0.16	0.17
H.arizonensis	perennial	9	0.15	0.05	0.02	NA
H.arizonensis	perennial	10	0.60	0.06	0.65	NA
H.arizonensis	perennial	11	0.46	0.17	0.37	NA
H.arizonensis	perennial	12	0.11	0.09	0.01	NA
H.arizonensis	perennial	13	0.37	0.10	0.49	NA
H.arizonensis	perennial	14	0.36	0.13	0.22	NA
H.arizonensis	perennial	15	0.32	0.15	0.14	NA
H.arizonensis	perennial	16	0.16	0.05	0.05	NA
H.arizonensis	perennial	17	0.24	0.05	0.07	NA
H.arizonensis	perennial	18	0.25	0.12	0.11	NA
H.arizonensis	perennial	19	0.24	0.09	0.07	NA
H.atrorubens	perennial	1	0.04	0.10	0.41	1.31

H.atrorubens	perennial	2	0.03	0.11	0.73	1.94
H.atrorubens	perennial	3	0.10	0.14	1.59	3.24
H.atrorubens	perennial	4	0.04	0.15	0.59	1.42
H.atrorubens	perennial	5	0.09	0.14	0.85	1.71
H.atrorubens	perennial	6	0.04	0.04	0.38	0.90
H.atrorubens	perennial	7	0.01	0.11	0.61	1.97
H.atrorubens	perennial	8	0.05	0.05	0.39	1.08
H.atrorubens	perennial	9	0.03	0.09	0.29	NA
H.atrorubens	perennial	10	0.05	0.10	0.70	NA
H.atrorubens	perennial	11	0.02	0.02	0.03	NA
H.atrorubens	perennial	12	0.03	0.06	0.63	NA
H.atrorubens	perennial	13	0.06	0.15	1.43	NA
H.atrorubens	perennial	14	0.03	0.14	0.44	NA
H.atrorubens	perennial	15	0.04	0.14	0.57	NA
H.atrorubens	perennial	16	0.05	0.10	0.57	NA
H.atrorubens	perennial	17	0.09	0.11	1.85	NA
H.atrorubens	perennial	18	0.01	0.06	0.36	NA
H.atrorubens	perennial	19	0.04	0.10	0.42	NA
H.atrorubens	perennial	20	0.04	0.14	0.25	NA
H.carnosus	perennial	1	0.02	0.09	0.21	0.52
H.carnosus	perennial	2	0.03	0.19	0.08	0.05
H.carnosus	perennial	3	0.02	0.17	0.25	0.27
H.carnosus	perennial	4	0.03	0.02	0.03	0.03
H.carnosus	perennial	5	0.02	0.11	0.07	0.09
H.carnosus	perennial	6	0.01	0.22	0.06	NA
H.carnosus	perennial	7	0.03	0.13	0.03	NA
H.carnosus	perennial	8	0.00	0.10	0.01	NA
H.carnosus	perennial	9	0.04	0.18	0.19	NA
H.carnosus	perennial	10	0.01	0.11	0.06	NA
H.carnosus	perennial	11	0.03	0.06	0.08	NA
H.carnosus	perennial	12	0.05	0.02	0.03	NA
H.cusickii	perennial	1	0.28	0.04	0.54	0.72
H.cusickii	perennial	2	0.24	0.11	0.32	0.52
H.cusickii	perennial	3	0.42	0.10	0.58	0.79
H.cusickii	perennial	4	0.30	0.06	0.54	0.72
H.cusickii	perennial	5	0.23	0.05	0.27	0.45
H.cusickii	perennial	6	0.34	0.06	0.23	0.32
H.cusickii	perennial	7	0.32	0.07	0.41	0.60
H.cusickii	perennial	8	0.27	0.04	0.13	0.17
H.cusickii	perennial	9	0.02	0.03	0.05	NA
H.cusickii	perennial	10	0.22	0.12	0.33	NA
H.cusickii	perennial	11	0.29	0.20	0.49	NA
H.cusickii	perennial	12	0.52	0.07	0.37	NA
H.cusickii	perennial	13	0.48	0.04	0.31	NA
H.cusickii	perennial	14	0.25	0.06	0.49	NA

H.cusickii	perennial	15	0.23	0.13	0.30	NA
H.cusickii	perennial	16	0.29	0.04	0.44	NA
H.debilis	annual	1	0.26	0.11	0.68	1.33
H.debilis	annual	2	0.44	0.18	1.78	1.45
H.debilis	annual	3	0.39	0.15	1.37	1.09
H.debilis	annual	4	0.30	0.10	0.95	1.46
H.debilis	annual	5	0.35	0.19	1.07	1.20
H.debilis	annual	6	0.28	0.12	0.56	0.77
H.debilis	annual	7	0.23	0.10	0.45	0.84
H.debilis	annual	8	0.18	0.07	0.36	NA
H.debilis	annual	9	0.15	0.12	0.49	NA
H.debilis	annual	10	0.19	0.05	0.10	NA
H.divaricatus	perennial	1	0.46	0.19	1.77	2.57
H.divaricatus	perennial	2	0.54	0.10	1.28	1.75
H.divaricatus	perennial	3	0.68	0.12	1.74	2.38
H.divaricatus	perennial	4	0.45	0.13	0.55	0.56
H.divaricatus	perennial	5	0.59	0.11	1.22	1.61
H.divaricatus	perennial	6	0.61	0.19	1.54	NA
H.divaricatus	perennial	7	0.32	0.18	0.46	NA
H.divaricatus	perennial	8	0.49	0.19	1.02	NA
H.divaricatus	perennial	9	0.23	0.15	0.38	NA
H.divaricatus	perennial	10	0.79	0.13	2.28	NA
H.divaricatus	perennial	11	0.26	0.06	0.37	NA
H.divaricatus	perennial	12	0.49	0.12	1.66	NA
H.divaricatus	perennial	13	0.34	0.09	0.51	NA
H.divaricatus	perennial	14	0.12	0.22	0.17	NA
H.divaricatus	perennial	15	0.45	0.12	0.55	NA
H.divaricatus	perennial	16	0.24	0.15	0.25	NA
H.giganteus	perennial	1	0.43	0.23	0.95	0.97
H.giganteus	perennial	2	0.54	0.15	1.69	2.09
H.giganteus	perennial	3	0.44	0.16	2.00	2.15
H.giganteus	perennial	4	0.59	0.18	1.84	1.83
H.giganteus	perennial	5	0.74	0.25	3.71	4.78
H.giganteus	perennial	6	0.66	0.26	2.24	1.85
H.giganteus	perennial	7	0.35	0.18	0.52	0.61
H.giganteus	perennial	8	0.40	0.14	0.43	0.61
H.giganteus	perennial	9	0.32	0.30	1.07	NA
H.giganteus	perennial	10	0.31	0.21	1.18	NA
H.giganteus	perennial	11	0.37	1.23	1.24	NA
H.giganteus	perennial	12	0.35	0.25	1.53	NA
H.giganteus	perennial	13	0.47	0.18	1.60	NA
H.giganteus	perennial	14	0.25	0.19	0.30	NA
H.giganteus	perennial	15	0.32	0.17	0.79	NA
H.giganteus	perennial	16	0.28	0.10	0.42	NA
H.heterophyllus	perennial	1	0.05	0.09	0.13	0.07

H.heterophyllus	perennial	2	0.04	0.04	0.12	0.12
H.heterophyllus	perennial	3	0.04	0.04	0.10	0.11
H.heterophyllus	perennial	4	0.03	0.03	0.11	0.27
H.heterophyllus	perennial	5	0.03	0.13	0.09	0.24
H.heterophyllus	perennial	20	0.03	0.07	0.09	NA
H.heterophyllus	perennial	21	0.01	0.09	0.05	NA
H.heterophyllus	perennial	22	0.05	0.05	0.09	NA
H.heterophyllus	perennial	23	0.04	0.10	0.13	NA
H.maximilliani	perennial	1	0.31	0.27	0.40	0.56
H.maximilliani	perennial	2	0.44	0.25	0.71	0.83
H.maximilliani	perennial	3	0.51	0.14	0.87	1.06
H.maximilliani	perennial	4	0.62	0.23	0.78	0.95
H.maximilliani	perennial	5	0.36	0.22	0.93	1.33
H.maximilliani	perennial	6	0.65	0.20	1.12	0.66
H.maximilliani	perennial	7	0.51	0.16	0.66	0.64
H.maximilliani	perennial	8	0.37	0.14	0.33	0.40
H.maximilliani	perennial	9	0.55	0.16	0.36	0.47
H.maximilliani	perennial	10	0.44	0.20	0.88	NA
H.maximilliani	perennial	11	0.45	0.28	0.39	NA
H.maximilliani	perennial	12	0.49	0.26	1.13	NA
H.maximilliani	perennial	13	0.31	0.17	0.33	NA
H.maximilliani	perennial	14	0.42	0.19	0.40	NA
H.maximilliani	perennial	15	0.46	0.14	0.33	NA
H.maximilliani	perennial	16	0.31	0.11	0.26	NA
H.maximilliani	perennial	17	0.62	0.21	0.82	NA
H.maximilliani	perennial	18	0.51	0.16	0.66	NA
H.maximilliani	perennial	19	0.61	0.17	0.57	NA
H.microcephalus	perennial	1	0.50	0.21	1.10	2.34
H.microcephalus	perennial	2	0.40	0.13	0.71	2.00
H.microcephalus	perennial	3	0.91	0.17	3.08	5.39
H.microcephalus	perennial	4	0.66	0.12	1.79	3.74
H.microcephalus	perennial	5	0.66	0.12	1.79	5.31
H.microcephalus	perennial	6	0.87	0.19	2.73	4.49
H.microcephalus	perennial	7	0.49	0.18	1.36	2.20
H.microcephalus	perennial	8	0.62	0.17	1.43	1.98
H.microcephalus	perennial	9	0.27	0.14	0.53	1.49
H.microcephalus	perennial	10	0.55	0.16	1.39	2.04
H.microcephalus	perennial	11	1.03	0.15	2.41	NA
H.microcephalus	perennial	12	0.37	0.16	1.09	NA
H.microcephalus	perennial	13	0.61	0.30	2.44	NA
H.microcephalus	perennial	14	0.54	0.17	0.96	NA
H.microcephalus	perennial	15	0.93	0.19	2.72	NA
H.microcephalus	perennial	16	0.80	0.15	2.19	NA
H.microcephalus	perennial	17	0.58	0.16	1.57	NA
H.microcephalus	perennial	18	1.03	0.17	0.42	NA

H.microcephalus	perennial	19	0.55	0.19	1.17	NA
H.microcephalus	perennial	20	0.28	0.22	0.66	NA
H.microcephalus	perennial	21	0.57	0.22	1.12	NA
H.microcephalus	perennial	22	0.51	0.17	1.10	NA
H.microcephalus	perennial	23	0.61	0.16	1.42	NA
H.nuttalli	perennial	1	0.45	0.35	1.03	1.82
H.nuttalli	perennial	2	0.76	0.25	3.31	2.66
H.nuttalli	perennial	3	0.53	0.18	1.27	1.82
H.nuttalli	perennial	4	0.72	0.21	1.86	1.81
H.nuttalli	perennial	5	0.50	0.20	1.52	1.68
H.nuttalli	perennial	6	0.31	0.17	0.70	0.58
H.nuttalli	perennial	7	0.53	0.14	1.15	1.19
H.nuttalli	perennial	8	0.65	0.20	0.93	0.73
H.nuttalli	perennial	9	0.31	0.14	0.15	0.24
H.nuttalli	perennial	10	0.14	0.05	0.05	NA
H.nuttalli	perennial	11	0.20	0.07	0.09	NA
H.nuttalli	perennial	12	0.16	0.13	0.27	NA
H.nuttalli	perennial	13	0.42	0.26	1.99	NA
H.nuttalli	perennial	14	0.63	0.26	2.03	NA
H.nuttalli	perennial	15	0.86	0.20	2.47	NA
H.nuttalli	perennial	16	0.69	0.19	1.69	NA
H.nuttalli	perennial	17	0.37	0.22	0.50	NA
H.nuttalli	perennial	18	0.44	0.22	0.35	NA
H.nuttalli	perennial	19	0.55	0.19	0.94	NA
H.nuttalli	perennial	20	0.47	0.22	0.70	NA
H.nuttalli	perennial	21	0.24	0.03	0.14	NA
H.nuttalli	perennial	22	0.32	0.18	0.34	NA
H.occidentalis	perennial	1	0.04	0.10	0.28	0.11
H.occidentalis	perennial	2	0.03	0.17	0.11	0.32
H.occidentalis	perennial	3	0.02	0.10	0.13	0.09
H.occidentalis	perennial	4	0.06	0.12	0.63	1.16
H.occidentalis	perennial	5	0.05	0.14	0.59	1.00
H.occidentalis	perennial	6	0.02	0.06	0.35	0.83
H.occidentalis	perennial	7	0.03	0.10	0.20	0.55
H.occidentalis	perennial	8	0.04	0.10	0.18	0.24
H.occidentalis	perennial	9	0.02	0.12	0.19	0.44
H.occidentalis	perennial	10	0.04	0.18	0.26	NA
H.occidentalis	perennial	11	0.03	0.11	0.37	NA
H.occidentalis	perennial	12	0.05	0.21	0.50	NA
H.occidentalis	perennial	13	0.00	0.09	0.09	NA
H.occidentalis	perennial	14	0.03	0.05	0.11	NA
H.occidentalis	perennial	15	0.03	0.09	0.12	NA
H.occidentalis	perennial	16	0.05	0.16	0.12	NA
H.occidentalis	perennial	17	0.01	0.07	0.02	NA
H.petiolaris	annual	1	0.13	0.01	0.24	0.68

H.petiolaris	annual	2	0.57	0.13	1.08	1.66
H.petiolaris	annual	3	0.15	0.10	0.36	0.69
H.petiolaris	annual	4	0.42	0.13	0.57	0.66
H.petiolaris	annual	5	0.39	0.16	0.90	1.22
H.petiolaris	annual	6	0.50	0.08	0.45	0.60
H.petiolaris	annual	7	0.37	0.13	1.25	1.34
H.petiolaris	annual	8	0.21	0.11	0.23	NA
H.petiolaris	annual	9	0.23	0.11	0.73	NA
H.petiolaris	annual	10	0.41	0.13	1.20	NA
H.petiolaris	annual	11	0.33	0.10	0.48	NA
H.petiolaris	annual	12	0.13	0.12	0.21	NA
H.petiolaris	annual	13	0.09	0.11	0.06	NA
H.petiolaris	annual	14	0.27	0.17	0.58	NA
H.porteri	annual	1	0.69	0.29	1.07	0.49
H.porteri	annual	2	0.69	0.26	1.53	0.80
H.porteri	annual	3	0.75	0.18	0.98	0.44
H.porteri	annual	4	0.52	0.25	0.86	0.64
H.porteri	annual	5	0.52	0.25	0.86	0.51
H.porteri	annual	6	0.74	0.23	1.63	0.57
H.porteri	annual	7	0.40	0.02	0.26	0.31
H.porteri	annual	8	0.33	0.12	0.36	0.33
H.porteri	annual	9	0.53	0.21	1.04	NA
H.porteri	annual	10	0.64	0.25	0.97	NA
H.porteri	annual	11	0.40	0.31	0.29	NA
H.porteri	annual	12	0.71	0.19	1.06	NA
H.porteri	annual	13	0.77	0.25	1.26	NA
H.porteri	annual	14	0.49	0.15	0.72	NA
H.porteri	annual	15	0.78	0.19	1.15	NA
H.porteri	annual	16	0.53	0.17	0.38	NA
H.porteri	annual	17	0.22	0.20	0.18	NA
H.porteri	annual	18	0.44	0.16	0.30	NA
H.praecox	annual	1	0.22	0.12	1.08	2.00
H.praecox	annual	2	0.31	0.17	2.72	2.70
H.praecox	annual	3	0.48	0.21	3.19	3.26
H.praecox	annual	4	0.59	0.16	3.54	2.96
H.praecox	annual	5	0.46	0.16	2.99	2.28
H.praecox	annual	6	0.33	0.21	2.44	2.12
H.praecox	annual	7	0.59	0.17	2.30	2.31
H.praecox	annual	8	0.45	0.16	2.49	NA
H.praecox	annual	9	0.23	0.16	1.64	NA
H.praecox	annual	10	0.41	0.15	3.46	NA
H.praecox	annual	11	0.38	0.20	2.63	NA
H.praecox	annual	12	0.43	0.20	3.65	NA
H.praecox	annual	13	0.26	0.19	1.66	NA
H.praecox	annual	14	0.54	0.16	2.83	NA

H.praecox	annual	15	0.51	0.20	1.28	NA
H.praecox	annual	16	0.32	0.14	0.71	NA
H.praecox	annual	17	0.37	0.17	1.71	NA
H.praecox	annual	18	0.15	0.13	1.27	NA
H.praecox	annual	19	0.37	0.16	1.07	NA
H.salicifolius	perennial	1	0.54	0.09	0.52	0.26
H.salicifolius	perennial	2	0.41	0.23	0.96	0.35
H.salicifolius	perennial	3	0.32	0.14	0.26	0.25
H.salicifolius	perennial	4	0.49	0.16	0.41	0.13
H.salicifolius	perennial	5	0.39	0.07	0.42	0.24
H.salicifolius	perennial	6	0.43	0.03	0.31	0.13
H.salicifolius	perennial	7	0.39	0.03	0.14	0.10
H.salicifolius	perennial	8	0.39	0.03	0.14	0.14
H.salicifolius	perennial	9	0.30	0.09	0.26	NA
H.salicifolius	perennial	10	0.56	0.14	0.70	NA
H.salicifolius	perennial	11	0.28	0.14	0.13	NA
H.salicifolius	perennial	12	0.46	0.15	1.20	NA

Appendix C - Copyright

The American Midland Naturalist

Department of Biological
University of Notre Dame
Notre Dame, Indiana
46556 U.S.A.

tel (574)631-7481
fax (574)-631-7413
e-mail: ammidnat@nd.edu
web: www.nd.edu/~ammidnat

The American Midland Naturalist Copyright Owned Request Form
Granting Permission to use published work.

This form confirms *the American Midland Naturalist* ("AMN"), a publication of the University of Notre Dame, hereby grants Iowa State University the non-exclusive right to use the article listed below for personal, educational and/or academic purposes during the full term of the copyright.

Name of article: Low temperature tolerance in the perennial sunflower Helianthus maximiliani

Author(s): Hannah Tetreault, Takeshi Kawakami, Charlotte Levy and Mark Ungerer

Volume and date of first publication in AMN: 175(1) January 2016

How article will be used: This article will be used as a chapter in Hannah Tetreault's PhD dissertation titled, "Transposable element contribution and biological consequence of genome size variation among wild sunflower species."

The right granted herein is non-exclusive, non-assignable and limited in scope to use of the Article in hard copy, other print format or pdf, e-book and languages worldwide. This does not include the right to use the Article solely for financial gain. The University of Notre Dame warrants that it is the sole copyright owner of the publication, that it has complete authority and capacity to grant the rights contained herein, and that no other releases or permissions are necessary.

By granting this non-exclusive right to limited use of the Article, neither the University of Notre Dame or AMN release or revoke their copyright ownership or other interest of any kind or nature whatsoever in the Article.

AMERICAN MIDLAND NATURALIST

By Hope Hollocher

Date 8/10/2016

UNIVERSITY OF NOTRE DAME

Figure 5.3 Copyrights for Chapter 4