LANDSCAPE GENOMICS AND DOMESTICATION STATUS OF MAXIMILIAN SUNFLOWER (*Helianthus maximiliani* Schrad.).

BY

Sean Robert Asselin

A Thesis Submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

Department of Plant Science
University of Manitoba
Winnipeg, Manitoba, Canada

ACKNOWLEDGEMENTS

I would like to acknowledge the following people who have contributed to and supported me throughout my Ph.D. studies.

First and foremost, my Ph.D. co-advisors Dr. Doug Cattani and Dr. Anita Brûlé-Babel for your guidance, support, enthusiasm for research, and at times blunt pragmatism throughout my time at the University of Manitoba. I am incredibly grateful for the innumerable research, career and life skills that I've learned under your guidance, and your commitment to my education.

Dr. Rob Gulden, Dr. Khalid Rashid and Dr. Barb Sharanowski for serving on my advisory committee, providing useful insight, and expanding the scope of my knowledge base.

Dr. David Van Tassel of The Land Institute for sharing his wealth of knowledge on Maximilian sunflower and providing germplasm, technical support, and financial means for conducting my research, Dr. Maria Stoimenova for providing molecular biology support and Dr. Curt McCartney for providing training and advice in the area of genetic mapping and QTL analysis.

The technical staff of the perennial grains research program, Ardelle Slamma and Rob Visser, for your technical expertise and support, and Martha Blouw of the Department of

Plant science for being a reliable source of advice and good stories at times when I needed to get away from my desk.

I would like to thank my fellow graduate students, past and present, from the Department of Plant Science. In particular, I would like to acknowledge Tyler Guerrieri, Roger Watts, Dr. Caroline Halde and Dr. Kevin Baron for sharing your experiences and advice in graduate school and beyond, and Alex Cattini, Calvin Dick, Jon Rosset and Don Sanders for your camaraderie.

My friends Claire and Anthony Tordiffe, Katie Hartle and Alex Miller, Donavan Purdy and Dr. Jen Tanner as well as Keith and Elizabeth Montgomery for your friendship and many great weekends out at the Libau farm and Winnipeg Beach.

My parents Chuck and Carol Asselin as well as my brother Thomas Asselin and his partner Jess Libby for all the unconditional support and understanding that you've provided throughout the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC: Akaike information criterion
AMOVA: Analysis of molecular variance
ANOVA: Analysis of variance
AWF: Among-and-within-family
Bp: Base-pair
CCA: Canonical correlation analysis
CCE: Caclcium carbonate equivalent
CTAB: cetyl-trimethylammonium bromide
CV: Coefficent of variation
DH: Doubled haploid
DNA: Deoxyribonucleic acid
DTPA: Diethylenetriaminepentaacetic acid
EAA: Envrionmental association analysis
FBAM: Family-based-associaton-mapping
FDR: False-dicovery-rate
F_{is}**:** Inbreeding coefficent
F_{it}**:** Overall fixation index
F_{st}**:** Fixation index
G x E: Genotype x environment
G x L: Genotype x location
G x L x Y: Genotype x location x year
G x Y: Genotype x year
Gb: Gigabase
GBS: Genotype-by-sequencing
GC: Genomic control
GLM: General linear model
GM: Genetic modification
GPS: Global positioning system
GWAS: Genome-wide-assocation-study
IBD: Identical by descent
***K*-matrix:** Kinship matrix
LD: Linkage disequilibrium
LG: Linkage group
LOD: Logarithm of odds
MAGIC: Multi-parent-advanced-inter-cross
MAS: Marker-assisted-selection
MLM: Mixed-linear-model
NAM: Nested-asscoation-mapping
NGS: Next-generation-sequencing
PC: Principal component
PCA: Principal component analysis
PCR: Principal component regression
PD: Population differentiation
PPM: Plant preservative mixture
PRESS: Predicted residual sums of squares
QTL: Quantitative trait loci
RIL: Recombinant inbred line
SA: Structured asscoation
SDR: Segregation distoiriton region
SNPs: Single nucleotide polymorphisms
SSRs: Simple sequence repeats
TDT: Transmission disequilibrium test
tGBS: Tunable-genotype-by-sequencing
TLI: The Land Institute
VIP: Variable importance plot

Abstract

Asselin, Sean Robert, Ph.D., The University of Manitoba, February, 2019. Landscape genomics and domestication status of Maximilian sunflower (*Helianthus maximiliani* Schrad.). Major Professors: Dr. Douglas J. Cattani and Dr. Anita L. Brûlé-Babel.

The Canadian prairies are a highly productive agricultural zone with a short growing season and a high proportion of land devoted to production of annual grains. Perennial grains and oilseeds are capable of extending the growing season while providing grain, forage, or biomass. Candidate species for perennial grains and oilseeds lack basic agronomic and genetic characterization to support breeding efforts, limiting the ability to develop perennial grain cropping systems. The purpose of this thesis was to characterize the candidate perennial oilseed species Maximilian sunflower (*Helianthus maximiliani* Schrad.) for the development of a locally adapted perennial oilseed crop for the Canadian prairies. This work consisted of three primary studies examining the phenotypic characteristics of Maximilian sunflower and related perennial *Helianthus* species, the adaptation of germplasm to its environment of origin, and the development of genomic resources for breeding Maximilian sunflower. Local environmental clines influenced population structure and phenotypic differentiation in Maximilian sunflower, including important adaptive characteristics such as timing of anthesis. Genomic analysis revealed a highly heterozygous genome and low levels of population structure. The first reported genetic map of Maximilian sunflower was developed. Variation in important

domestication syndrome traits such as branching architecture and capitulum size were observed in wild sampled and experimental mapping populations. Association and QTL analysis revealed candidate SNPs for multiple agronomic traits and adaptation to local environmental clines to support the neo-domestication of Maximilian sunflower as a perennial oilseed crop. The research contained in this thesis identified variation in key traits in Maximilian sunflower indicating advancement under selection for domestication is possible.

Forward

The thesis includes an introduction, literature review and three research chapters followed by a general discussion and conclusions. The research chapters summarize work conducted at the University of Manitoba from January 2012 to August 2018. The research chapters follow the format of Crop Science and follow the style defined by the Department of Plant Science, University of Manitoba, Winnipeg, MB, CA.

CHAPTER 1.0: Introduction

Perennial grains and oilseeds have been proposed as a class of crops to enhance the sustainability of agricultural systems through the introduction of new functional diversity and associated ecosystem services (Cox et al., 2006; Glover et al., 2010b; Kane et al., 2016). It is estimated that globally, annual crops are grown on 60-80% of total arable cropland (Lobell et al., 2011) and annual cropping systems dominate in Western Canada. In temperate environments, the use of annual crops can result in gaps between cropping cycles where solar energy, heat units, and precipitation may be available to support active plant growth, but no standing crop is present. These gaps may leave land vulnerable to nutrient loss, erosion, and infiltration by weedy species, and represent a potential niche to be exploited for crop production (Glover et al., 2010b; Cattani and Asselin, 2018b). Compared to annuals, perennial species generally have a longer photosynthetically active growing season (Dohleman and Long, 2009; Glover et al., 2010a), deeper rooting depth, and an increased capacity to utilize and retain moisture in their environment (Gordon-Werner and Dörffling, 1988; Toker et al., 2007; Ferchaud et al., 2014). Perennial crops may help mitigate the environmental impact of annual cropping systems through reduced erosion and nutrient leaching (Kort et al., 1998; Entz et al., 2001; Culman et al., 2013), greater carbon sequestration (Lal, 2003; Montagnini and Nair, 2004; Lemus and Lai, 2005; Tilman et al., 2006; Fargione et al., 2008) and improved nutrient cycling and availability (Smaje, 2015; Crews et al., 2016).

Perennial grains are intended to combine characteristics of conventional perennial forage crops and annual grains. Developing perennial grains from adapted plant species may provide benefits through extending the growing season and providing a wider range of end uses for perennial crops in Western Canada as a grain or dual purpose such as grain/forage or grain/bioenergy crops (Bell et al., 2010; Larkin et al., 2014; DeHaan et al., 2016; Ryan et al., 2018).

Candidate species for perennial grain and oilseed systems lack basic characterization in terms of biology, agronomics, and genetic potential for improvement, which limits the ability of plant breeders to develop effective breeding strategies. A number of candidate perennial species have been envisioned through both hybridization with annual crops and identification of wild species with favourable characteristics for neo-domestication (Wagoner and Schaeffer, 1990; DeHaan et al., 2016). Members of the genus *Helianthus* have been identified as candidates for perennial crop development (Cox et al., 2002; Van Tassel et al., 2014; Kantar et al., 2014, 2018).

Maximilian sunflower (*Helianthus maximiliani* Schrad.) is an herbaceous perennial forb, native to western Canada and a candidate perennial oilseed crop. This species is of interest to plant breeders as a crop wild relative of cultivated sunflower, but there is limited characterization of the basic components of its biology. Study of Maximilian sunflower will provide insight into its biology, agronomics and genetic potential for development as a perennial grain oilseed crop. The thesis presents the results of research conducted on the characterization of Maximilian sunflower and related perennial species of the genus *Helianthus*. The main objectives of these experiments were to:

1) Provide a baseline characterization of the phenotypic characteristics of available perennial *Helianthus* germplasm adapted to southern Manitoba, Canada and their potential for advancement under selection (Chapter 3).
2) Examine the landscape genetics and genomics of Maximilian sunflower (Chapters 3 and 4).
3) Develop the first reported genetic map of Maximilian sunflower (Chapter 5).
4) Identify QTL and candidate SNPs associated with phenotypic differentiation, environmental clines and domestication syndrome characteristics in Maximilian sunflower to support breeding efforts (Chapters 4 and 5).

CHAPTER 2.0: Literature Review

2.1 History of perennial grains and oilseeds

2.1.1 Historic use of perennials as harvested grains

Perennial grains have been used throughout history in many parts of the world in conjunction with annual grains to support human diets as minor or wild harvested crops (Wagoner and Schaeffer, 1990). The grains of wild harvested perennial grasses such as Canada wild rye (*Elymus Canadensis* L.), blue grama [*Bouteloua gracilis* (Willd. Ex Kunth) Lag. Ex Griffiths], Indian ricegrass [*Achnatherum hymenoides* (Roem. & Schult.) Barkworth], sand drop seed [*Sporobolus cryptandrus* (Torr.) A. Gray] and prairie junegrass [*Koeleria cristata* (Ledeb.) Schult.], amongst others, were consumed by indigenous peoples of North America and are common in the archeological record (Kuhnlein and Turner, 1991).

Perennial crop wild relatives harvested for seed include the oilseeds Maximilian sunflower (*Helianthus maximiliani* Schrad.), Giant sunflower (*Helianthus giganteus* L.), blue flax (*Linum perenne* L.), and Lewis flax (*Linum lewisii* Pursh.) (Yanovsky, 1936; Kuhnlein and Turner, 1991; Moerman, 2010). Mountain rye (*Secale montanum* L.), a perennial ancestor of cereal rye (*S. cereale* L.), has a widely reported presence in early Neolithic agrarian sites with other annual and perennial grains as reviewed by Wagoner and Schaeffer (1990) and Van Tassel et al. (2010). Sea lyme grass or strand-wheat [*Leymus arenarius* (L.) Hochst.] was an important cereal grain of the Vikings and evidence of its cultivation in North America during pre-Columbian times can be found at L'Anse-aux-Meadows, Newfoundland (Griffin and Rowlett, 1981). In other parts of the

world, there is evidence that weakly perennial varieties of annual crops such as rice (*Oryza sativa* L.), were initially cultivated as a perennial through the practice of ratooning (repeat harvesting of post-harvest regrowth) (Hill, 2010). Similarly, ratooning is practiced in sorghum, [*Sorghum bicolor* (L.) Moench] and pigeonpea [*Cajanus cajan* (L.) Millsp.] to reduce production costs (Paterson et al., 2014), although this species is generally grown as annual crop.

2.1.2 Early breeding efforts

Breeding efforts focused on the improvement of perennial species as grain crops date to the late 19th century, with efforts focused on the hybridization of annual crops with perennial crop wild relatives (Wagoner and Schaeffer, 1990). Interest in developing mixed-use crops for forage and grain to reduce soil erosion and conserve moisture led to breeding efforts focused on wild perennials as well as interspecific hybrids between annual grains and perennial relatives. Throughout the 1920s through to the 1940s, breeding programs focused on the development of perennial crops through interspecific hybridization were established in the Soviet Union (Tsitsin and Lubimova, 1959; Tsitsin, 1965), the United States (Suneson et al., 1963), Germany (Wakar, 1937) and Canada (Armstrong, 1936; Johnson, 1938) with a major focus on perennialized wheat (*Tricicum* x *Thinopyrum spp.* hybrids) and perennial cereal rye (*S. cereal* x *S. montanum*).

Considerable improvements had been made during this period, particularly in perennial wheat, but due to the low fertility of interspecific hybrids (Armstrong, 1936; Tsitsin, 1965), linkage drag between the perennial habit and wild traits, and observed trade-offs between grain yield and regrowth (Suneson et al., 1963), perennial grain

germplasm from these early programs were never established as viable crops and most programs had largely ceased by the 1960s (Suneson et al., 1963; Wagoner and Schaeffer, 1990; Cox et al., 2002). Much of the germplasm developed by these early programs was incorporated into established higher yielding annual cereal breeding programs as sources of disease resistance to pathogens such as stripe, leaf and stem rust, and other important traits (Scheinost et al., 2001; Cox et al., 2002).

2.1.3 Modern breeding efforts

Interest in perennial wheat and the concept of perennial grains for reducing soil erosion and production costs persisted throughout the 1970s and 1980s. During this period, screening of various wild perennial species for production characteristics began at The Land Institute (TLI) and Rodale Institute for use in conservation agriculture (Wagoner and Schaeffer, 1990). In 1984, research into developing hybrids of *S. bicolor* was conducted at TLI and early perennial sorghum germplasm was developed (Piper and Kulakow, 1994). In 1991, researchers at Washington State University began growing hybridized perennial wheat, and by the early 2000s had developed several thousand lines from winter wheat and various *Thinopyrum* species crosses (Scheinost et al., 2001). Concentrated efforts to breed perennial grains at TLI were initiated in 2002, with the establishment of breeding programs for intermediate wheatgrass [*Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey], perennial wheat, perennial sorghum, Maximilian sunflower, perennial sunflower (*Helianthus* hybr.), Illinois bundleflower [*Desmanthus illinoensis* (Michx.) MacMill. Ex B.L. Rob. & Fernald] and perennial rice (*Oryza* hybr.)(Cox et al., 2010). Throughout the 2000s, interest in perennial grains

continued to grow and breeding programs were established at several institutions worldwide with interest in a wide variety of perennial grain candidates (Crews and Cattani, 2018). The first breeding program with a specific focus on perennial grains and oilseeds in Canada was established in 2010 at the University of Manitoba, with a major focus on intermediate wheatgrass, Maximilian sunflower, perennial wheat and the development of perennial grain polycultures.

2.2 Ecosystem services

2.2.1 Ecosystem services provided by perennial crops

A strong argument for the development of perennial grains to date has been the suggestion that they will enhance agricultural sustainability through the incorporation of additional ecosystem services (Cox et al., 2006; Glover et al., 2010b; Kane et al., 2016). Ecosystem services are defined as benefits people obtain from ecosystems, and fall under four broad categories: *provisioning services*, which include products obtained from ecosystems such as food, fiber and fuel; *regulating services*, that include benefits obtained from the regulation of ecosystem processes and include climate, natural hazards, wastes and environmental quality; *supporting services*, which include "services that are necessary for the production of all other ecosystem services" such as soil formation, photosynthesis and nutrient cycling; and *cultural services*, which encompass non-material services that enrich the human experience, including learning, aesthetics and recreation (MEA, 2005).

Individual species may provide both beneficial, detrimental or both beneficial and detrimental services to their ecosystem (Power, 2010). Annual grain crops for instance,

provide provisioning services through the production of grain, but are frequently reliant on tillage, which when overused, can lead to ecosystem disservices such as soil erosion, and environmental pollution in the form of runoff (Tiessen et al., 2010; Crews et al., 2016). Perennial crops have greater root mass and often fibrous root systems, reducing erosion risk and have a tendency to maintain more soil carbon compared to annual crops (Glover et al., 2010b; Kell, 2011). Crop diversification is one strategy used to broaden the range of ecosystem services in a given environment to improve resilience. Resilience is the capacity of an ecosystem to resist and recover from abiotic and biotic disturbances (Lin, 2011).

Annual crops comprise an estimated 60-80% of global cropland and approximately 75% of calories consumed by humans come from four annual grain crops: maize, wheat, rice and soybean (Lobell et al., 2011). Targeting perennial species and integrating them into agroecosystems dominated by annual crops has been suggested as a method of enhancing functional diversity (i.e. the number of functionally disparate species), ecosystem function (Isbell et al., 2011), and ultimately, productivity through the introduction of new ecosystem services (Asbjornsen et al., 2014).

Perennial species occupy a different series of environmental niches than annuals, taking advantage of seasonal resources unavailable to annual crops, and provide a different assemblage of ecosystem services (Crews and Cattani, 2018). A proxy to understanding the potential benefits of perennial grain systems in the context of western Canada may be derived from knowledge of perennial forages (e.g. Lasisi et al., 2018) and no-till annual grain production systems (Janzen et al., 1998; Tiessen et al., 2010), which share some commonalities with perennial grains.

2.2.1.1 Carbon sequestration and greenhouse gas mitigation

Perennial species represent a largely unexploited potential carbon sink for mitigating increasing atmospheric CO_2 through carbon sequestration, and have been suggested for developing carbon neutral agricultural and bioenergy systems (Lal, 2003; Montagnini and Nair, 2004; Lemus and Lai, 2005; Tilman et al., 2006; Fargione et al., 2008). While aboveground biomass contributes to soil carbon deposition, in many instances plant roots and their exudates are the largest contributors to soil carbon (Paustian et al., 1990; Maas et al., 2013). Converting cropland into perennial forage in Canada has been suggested as a method to substantially increase soil carbon sequestration (Bruce et al., 1999; Conant et al., 2001). Several studies based in western Canada have demonstrated this potential with different forage species such as pubescent wheatgrass (*Thinopyrum intermedium*, formerly *Agropyron trichophorum* Richt.) (Bremer et al., 2002), crested wheatgrass (*Agropyron desertorum* L.) (Hutchinson et al., 2007), and alfalfa (*Medicago sativa* L.) (Maas et al., 2013). Evidence from continuously harvested grasslands in Kansas has shown that, compared to fertilized annual crops such as wheat (*Triticum aestivum* L.), harvested perennial grasslands maintained more soil carbon and nitrogen despite a lack of inputs (Glover et al., 2010a; Culman et al., 2010). Studies on the use of perennial forages in crop rotations in the Red River Valley of Manitoba show that their inclusion can provide an immediate reduction in soil N_2O emissions and net increase in CO_2 uptake relative to annual systems (Taylor et al., 2013; Maas et al., 2013). The effects of reduced tillage and biomass accumulation in annual systems on soil carbon sequestration and nutrient retention in western Canada give support to the potential benefits of developing perennial grain cropping systems.

2.2.1.2 Water management and erosion control

Water management and erosion control were early drivers of research into perennial wheat in Canada (Armstrong, 1936). Water management services of perennial crops are well documented. Deep rooting of perennial species has been shown to impart better resistance to drought conditions in some species due to their capacity to capture water from deeper soil layers (Gordon-Werner and Dörffling, 1988; Toker et al., 2007; Ferchaud et al., 2014). Root tunnels produced by perennial crops have also shown a positive effect on soil moisture infiltration and yield benefits in subsequent annual crops (Meek et al., 1992; Ward et al., 2002). Erosion has a negative impact on the waterholding capacity of soils. The deep fibrous, or spreading, roots of many perennial species have been identified as a tool for managing erosion (Kort et al., 1998; Lemus and Lai, 2005) which may enhance the retention of moisture and reduce nutrient leaching.

2.2.1.3 Nutrient leaching, cycling and availability

Perennial crops show a considerable ability to reduce nutrient leaching in some environments (Randall and Mulla, 2001; Crews, 2005; Lasisi et al., 2018). For instance, intermediate wheatgrass, a herbaceous perennial grain candidate, reduced soil moisture to greater depths and reduced total NO_3 leaching by 86% when compared to annual winter wheat (*T. aestivum*) in the Midwest United States (Culman et al., 2013). Alfalfa in rotation with annuals has shown the potential to reduce NO_3 leaching in western Canada (Campbell et al., 1994; Entz et al., 2001), similar to intermediate wheatgrass. Reduced

leaching in these crops is partially attributed to the ability of these crops to reduce water buildup in the soil and reduce seasonal runoff.

Reviewed extensively by Crews and DeHaan (2015) and Crews et al. (2016), perennial cropping systems are expected to have improved synchrony between crop nutrient demand and nutrient availability. Increased organic matter, carbon sequestration, a stable soil microbiome, greater nitrogen retention, and seasonal translocation of photosynthate from aboveground to belowground plant structures are expected to be greater in perennial systems (Crews, 2005; Smaje, 2015; Crews et al., 2016). The Rothamsted continuous hay experiment has shown that over the course of 120 years there have been no consistent changes to total soil N content, nor biomass yields in plots where perennial hay (composed of a complex mixture of grasses, forbs and legumes) has been harvested twice a year, despite the lack of exogenously applied nutrient inputs (Jenkinson et al., 1994, 2004). Continuously harvested perennial grasslands have been shown to yield similar levels of N in dried plant tissues (expressed as kg $ha^{-1}yr^{-1}$ of N) while maintaining higher levels of soil C and N than adjacent regularly fertilized annual wheat fields over the course of 75 years (Glover et al., 2010a). In addition, soils with perennial vegetation have a tendency to maintain greater levels of plant-available forms of phosphorus. Soils in which perennial hay has been sown may show five to ten times the amount of P in their microbial biomass when compared to annual wheat (Crews and Brookes, 2014).

2.2.1.4 Potential ecosystem disservices of perennial-based cropping systems

Perennials occupy land over several growing seasons, which may pose potential risks, particularly if species are grown as a monoculture. Common cultural practices used in annual cropping systems for managing pathogen load such as tillage and annual crop rotation may not be available, posing a risk of soil- and residue-borne pathogens (Cox et al., 2007). In perennial forages, stand productivity tends to decline over time due to the accumulation of weedy species and insect pests (Boelt et al., 2015). Increasing species richness generally reduces the potential for invasion of weeds or other pests through greater competition and reduced resource gaps, and provides resilience to perturbations (Wilsey and Polley, 2002; Picasso et al., 2008). Furthermore, the maintenance of genetic diversity within species, such as in certain outcrossing forage crops which maintain a degree of heterogeneity, can enhance resiliency to insect pests and pathogens through genetic variation in defense and recovery mechanisms (Wilkins and Humphreys, 2003; Uppalapati et al., 2013; Annicchiarico et al., 2014).

High biomass production is a characteristic of many herbaceous perennial crops, including perennial grain candidates, and may provide a mechanism of weed suppression, but may also shift weed communities within rotations (Ominski et al., 1999; Meiss et al., 2010). Similar to pathogen load, shifts in the composition of weed communities and tolerance of different perennial grains species to weed incursion will require further study to establish effective management strategies.

2.2.2 Economic challenges of perennial grain systems

Ecosystem services can maintain agricultural production, but with the exception of provisioning services, such as grain yield, they are largely uncharacterized or undervalued (Daily et al., 2009; Power, 2010; Bommarco et al., 2013). Lack of characterization often leads to the exclusion of ecosystem services from economic decision-making (MEA, 2005; Carpenter et al., 2009). Assigning monetary value to regulating and supporting ecosystem services is challenging due their contextual nature, indirect effects, and the task of measuring the cumulative impact of subtle processes over time and/or diffused over large areas (Daily et al., 2009; Power, 2010; Bommarco et al., 2013). Most often, supporting and regulating ecosystem services are assigned value in the context of how they influence provisioning services, such as the effect of pollination services on fruit and seed set (Bommarco et al., 2012), reduction of disease and pest outbreaks though crop rotation (Cheatham et al., 2009), or the influence of biological nitrogen fixation by legumes on subsequent crops (Crews and Peoples, 2004; St. Luce et al., 2015).

Adoption of best management practices that support ecosystem services may be influenced by multiple factors, including the dissemination of information, environmental awareness, connection to agencies or local farm networks, and the financial capacity to implement support of new technology (Zentner et al., 2002; Baumgart-Getz et al., 2012). In the absence of policy tools and financial mechanisms to support ecosystem services (MEA, 2005; Daily et al., 2009; Asbjornsen et al., 2014), the ability of perennial grains to produce a consistent economic yield to be competitive with existing crops, such as annual

grains, forages, and bioenergy crops, will be a major factor in their adoption (Wagoner and Schaeffer, 1990; Bell et al., 2010; Adebiyi et al., 2016).

Perennial grains are not commercially available, and the time for development has been estimated to be approximately 10 to 25 years (Glover, 2005). Grain yields of early perennial grain cultivars are expected to be low during the initial breeding cycles as germplasm develops (Bell et al., 2008). In their current state, yield and quality of perennial grains fall below that of current crops, posing an economic barrier, with yields ranging from 10 to 70% of related annuals (Jaikumar et al., 2014; Runck et al., 2014). Subsequently, selection for yield, yield stability, and grain quality has been a major focus of perennial grain breeding programs (Cox et al., 2010; Kantar et al., 2014; DeHaan et al., 2016; Nabukalu and Cox, 2016; Zhang et al., 2016).

2.2.3 Production costs in perennial grain systems

Early efforts to develop perennial wheat recognized the economic potential of perennial grains in reducing annual seed, tillage and labour costs (Tsitsin, 1965; Wagoner and Schaeffer, 1990). Perennial grains exhibit characteristics that may reduce input costs and potentially offset the expected lower yields of early release germplasm. Input costs in the form of fertilizer, herbicides, seed and fuel are expected to be lower in perennial grains (Bell et al., 2008, 2010; Pimentel et al., 2012). For instance, in Manitoba, herbicide, fungicide, insecticide and fuel costs have the tendency to be lower in some perennial crops such as timothy (*Phelum pretense* L.), birdsfoot trefoil (*Lotus corniculatus* L.), and meadow fescue [Schedonorus pratensis (Huds.) P. Beauv.] relative to common annual crops such as canola (*Brassica napus* L.), spring wheat and soybean

[*Glycine max* (L.) Merr.] (Cattani and Asselin, 2018b) (Table 2.1). Similar to annual crops, production costs differ between different perennial crops, depending on the competitiveness, pest tolerance, nutrient requirements and ease of establishment. If production costs remain relatively low, similar those in many perennial forage seed crops, a suitable return on investment may be achievable to support the adoption of perennial grains despite lower grain yields than established annual crops.

Fall seeding of perennial crops following annual crop harvests could aid in bypassing establishment challenges in years of excess spring moisture and utilize light, nutrient, and moisture resources during periods in fall and early spring that otherwise may go unused, or be capitalized on by weeds (Cattani and Asselin, 2018b).

Proposed dual-use systems of a primary grain harvest, followed by secondary uses of biomass such as hay or use in grazing, could greatly increase the economic potential of perennial wheat (Bell et al., 2008). Forage qualities, or other secondary uses such as feedstock for ligno-cellulosic biofuels, may provide added value during early perennial grain adoption (Larkin et al., 2014; DeHaan et al., 2016; Ryan et al., 2018). Niche markets that demand a greater return, such as premiums enjoyed in organically produced crops, or the production of high-value specialty products such as unique oils, proteins, or phytochemicals, could also offset costs for perennial grains.

Due to growing global demands for food, fiber and fuel, agricultural production is expected to expand into increasingly marginal landscapes (Lambin and Meyfroidt, 2011), including those at high risk of erosion (Cassman et al., 2003). Perennial grains may protect and restore marginal lands through improvements in ecosystem services relative to annual grains, particularly in systems which are heavily reliant on tillage (Cox et al.,

2006; Glover et al., 2010b; Pimentel et al., 2012). Growing perennial plants on degraded or abandoned land may reduce competition for land between annual and perennial grains (Tilman et al., 2009). Preliminary work in perennial wheat suggests that it is most likely to achieve the grain yields necessary for profitability on lands where cereal grains are not profitable, or higher yields are not achievable (Adebiyi et al., 2016). This may provide a niche for producers to adopt perennial grains.

Table 2.1: Ranking of estimated production costs ha^{-1} of common annual and perennial seed crops in Manitoba for 2016 (Cattani and Asselin, 2018b).

Annual crops	**Seed cost**	**Fertilizer**	**Herbicide**	**Fungicide**	**Insecticide**	**Fuel**	**Total**	**Rank**
Canola[1]	$52.25	$78.99	$13.13	$36.25	$4.73	$16.43	$201.78	13
Spring wheat	$22.00	$61.23	$26.21	$21.31	$0.00	$20.05	$150.80	9
Soybean	$94.38	$11.35	$14.67	$0.00	$0.00	$15.37	$135.77	6
Oats (*Avena sativa* L.)	$18.13	$48.57	$9.50	$10.13	$0.00	$23.33	$109.66	3
Barley (*Hordeum vulgare* L.)	$15.00	$53.48	$24.88	$17.25	$0.00	$21.67	$132.28	5
Grain corn (*Zea mays* L.)	$78.30	$94.42	$18.17	$0.00	$0.00	$23.65	$214.54	15
Winter wheat	$20.00	$66.14	$13.83	$21.31	$0.00	$21.71	$142.99	7
Sunflower-Oilseed (*Helianthus annuus* L.)	$38.00	$53.48	$24.75	$28.75	$0.00	$20.25	$165.23	12
Sunflower-Confectionary	$43.31	$53.48	$43.79	$28.75	$13.23	$19.92	$202.48	14
Perennial seed crops	**Seed cost**[2]	**Fertilizer**	**Herbicide**	**Fungicide**	**Insecticide**	**Fuel**	**Total**	**Rank**
Alfalfa	$24.24	$24.84	$49.00	$36.00	$14.00	$9.77	$157.85	11
Timothy	$21.97	$64.21	$10.00	$0.00	$3.00	$9.34	$108.52	2
Red Clover (*Trifolium pratense* L.)	$99.40	$24.84	$20.00	$0.00	$0.00	$8.21	$152.45	10
Meadow fescue	$26.28	$64.21	$10.00	$0.00	$3.00	$10.18	$113.67	4
Birdsfoot trefoil	$22.88	$24.84	$20.00	$17.00	$14.00	$9.18	$107.90	1
Tall fescue (*Festuca arundinacea* Schreb.)	$31.55	$64.21	$23.00	$17.00	$3.00	$11.55	$150.31	8

[1]Italicised crops are the currently highest production area crops of annual and perennial seed crops in Manitoba.
[2]Seed cost for forage seed crops calculated as (Seed + nurse crop costs – nurse crop revenue)

2.3 Breeding perennial grains

2.3.1 The phenotypic trade-off model

A critique of perennial grain systems is that, development of high yielding perennial grain and oilseed crops, comparable to the yields of annual crops, is unlikely due to biological constraints. In the presence of competition, resource allocation of finite resources between traits such as seed size and seed number, sexual and asexual reproduction, or plant age and plant size at reproduction, trade-offs are expected (van Noordwijk and de Jong, 1986; Worley et al., 2003; Roff and Fairbairn, 2007; Sadras, 2007; Haselhorst et al., 2011). In plants, expected trade-offs between the maximization of population growth (high reproductive rate) (r-strategy), versus maximization of biomass accumulation and persistence (K-strategy), predicts that no plant can theoretically be successfully adapted to both high seed production and vegetative growth (Closset-Kopp et al., 2007; Smaje, 2015). Much of the gains in the proportion of plant biomass allocated to grain in annuals (Wilkins and Humphreys, 2003) would likely compromise longevity, if similar levels of allocation of dry matter partitioning took place in perennial grains (Smaje, 2015). A common recurring theme in evolutionary biology is the assumption that trait evolution is restricted by resource limitations and fitness trade-offs (Roff and Fairbairn, 2007; DeHaan and Van Tassel, 2014). This is the basis of the phenotypic trade-off model in plants, which rests on the assumption that the relationship between two traits is defined as a static, fixed function in which selection moves the mean trait value along a bivariate curve (DeHaan et al., 2005).

The interpretation of strict trade-offs imposed by natural selection may be misleading when applied to systems where artificial selection is occurring, or in environments where limiting resources are managed, such as intensely managed agricultural landscapes (Smaje, 2015). While the presence of resource allocation trade-offs between sexual and vegetative growth is well documented in the ecological literature (Roff and Fairbairn, 2007; Vico et al., 2016), there are reports of herbaceous perennials which exhibit an apparent lack of trade-offs (Pitelka et al., 1985; Horvitz and Schemske, 1988; Karlsson et al., 1990; Jennersten, 1991; Jackson and Dewald, 1994; Cheplick, 1995). There is limited information on the effect that artificial selection has on herbaceous perennials selected for grain production in managed ecosystems (Smaje, 2015; González-Paleo et al., 2016). Perennial grain studies by Piper and Kulakow (1994) found no relationship between rhizome mass and seed mass in perennial sorghum germplasm (*S. bicolor* x *S. halpense* (L.) Pers.), supporting the assumption that selection for both high seed yield and maintenance of rhizome production (perenniating structures) is possible (Piper and Kulakow, 1994). Similarly, DeHaan et al. (2018) observed that seed yield per spike was not negatively correlated with post-harvest regrowth in intermediate wheatgrass, and concluded that selection for seed yield may not impact perennial survival.

2.3.2 The quantitative genetic model of trade-offs

An alternative model to approach trade-offs is the quantitative genetic model, in which relationships between traits are not static and change as a result of selection (Lande, 1982; Roff et al., 2002). Traits are viewed as static bivariate relationships in the

phenotypic trade-off model, while in the quantitative genetic model, trade-offs are dynamic and viewed as multi-factorial in nature, following a multivariate distribution (Lande, 1982; Roff et al., 2002; Gianola and Sorensen, 2004). Trade-offs between traits can be explained as the result of negative genetic covariance, caused by linkage disequilibrium or antagonistic pleiotropy, and may be influenced by genetic drift, mutation, selection, or genetic differences in resource allocation hierarchies (Lande, 1982; Worley and Barrett, 2000; Worley et al., 2003; Roff and Fairbairn, 2007). Under the quantitative genetic model, simultaneous selection for negatively correlated traits is theoretically possible as trait pairs are considered part of a greater network of interconnected traits. Perennial grain breeders have argued that the quantitative genetic model is a more appropriate approach to understanding resource allocation trade-offs and the constraints they may impose on selection (DeHaan et al., 2005). For instance, resource allocation to vegetative growth and seed production is possible if constraining traits such as photosynthetic capacity were to increase, or photosynthates were to be partitioned from energy-expensive adaptive traits, which may not be necessary or are redundant under cultivated conditions (DeHaan et al., 2005).

2.3.3 Trade-offs in the context of production

Resource allocation trade-offs between the development of vegetative perenniating structures and seed production may not be a limitation in perennial grain systems. Under either the phenotypic or quantitative genetic models, what is deemed an acceptable level of grain yield, biomass production, or perenniality is contextual (Smaje, 2015). Perennials with high biomass production and low harvest index may be suitable in

systems where costs are recovered through secondary products such as forage (Bell et al., 2008; Larkin et al., 2014; DeHaan et al., 2016; Ryan et al., 2018). Additionally, these crops may support ecosystem services which may enhance yield regionally (such as pollinator habitat for adjacent crops such as canola or alfalfa (Bommarco et al., 2012), or be recovered in subsequent crops (e.g. nitrogen fixation from perennial legumes (St. Luce et al., 2015)). Alternatively, the presence of trade-offs between vegetative storage and seed yield may prove beneficial for the development of short-lived perennials (3-5 years) implemented into rotations with annuals crops. Excessive rhizome production may result in agronomic challenges such as weediness, as observed in perennial species such as Johnsongrass (*S. halpense*) (Paterson et al., 1995) and Jerusalem artichoke (*Helianthus tuberosus* L.) (Swanton et al., 1992). Resources could potentially be diverted to grain production in the presence of resource allocation trade-offs to reduce weediness and enhance grain yield in candidate species if excessive vegetative spread or persistence is a concern. Furthermore, adaptation to the agricultural environment through selection may provide greater access to resources, minimizing negative effects of trade-offs (DeHaan et al., 2005, 2018).

2.3.4 Candidate perennial grains

Characteristics that make a good perennial grain candidate have been proposed in multiple synthesis papers by perennial grain researchers (e.g. Wagoner and Schaeffer, 1990; DeHaan et al. 2016). Characteristics of perennial grain candidates mainly relate to how perennial grains could be incorporated into existing agricultural production systems through domestication and improvement. Wagoner and Schaeffer's (1990) extensive

review of proposed perennial grain systems outlined 13 important characteristics for candidate perennial grains, including agronomic characteristics, and factors that influence the ease of breeding. Agronomic characteristics include: first and foremost, vigorous perennial growth (maintenance of perenniality), then useable grain (potential for end use), easily threshed grain, manageable grain size (>2.0 mg seed^{-1}), synchronous seed maturation, shattering resistance, non-lodging stems, seed heads held above the level of the foliage, dry-down of inflorescences and stems at maturity and high potential for mechanical harvest. Factors influencing breeding include sufficient variability to make selection, acceptable grain-yield potential for more than two years and meiotic stability (Wagoner and Schaeffer, 1990). DeHaan et al. (2016) put forward a comprehensive framework for the development of perennial grains in a broader context, based upon a review of new crop domestication efforts over the previous 30 years, along with experience and progress made in perennial grain breeding (Table 2.2). Similarly, Kantar et al. (2016) summarized breeding objectives for perennial grains, with a greater focus on how newly emerging technologies may facilitate breeding efforts. No candidate species is expected to meet all of these criteria (and in fact some of these characteristics are contradictory), however, they provide a useful framework for evaluating strengths and weaknesses of perennial grain candidates. Candidate species under consideration are variable in their biology, but can be divided broadly into two categories based upon the origin of the germplasm: interspecific hybrids and wild species. Interspecific hybrids and wild species each face unique challenges as candidate materials for the development of perennial crops.

Table 2.2: Adapted summary of suggested characteristics put forward by DeHaan et al. (2016) for the selection and evaluation of perennial grain candidate species.

Characteristic	Examples
Domestic morphology and phenology	Domestication traits, characteristics outlined by Wagoner and Schaeffer (1990)
Ease of breeding and genetics	Reproductive biology, amenability to breeding techniques, genome composition and ploidy
Easily harvestable	Amenability to mechanical harvest, ability of grain to be recovered from stands
High yield	Necessary biomass to produce a yield worth harvesting, sufficient harvest index
Grain similar to that of current crops	Crop can easily be integrated into existing commodity markets, ability to easily be adopted by existing consumer taste
High-value product	Ability of crop to be marketed for particular high-value characteristics or specialty niche
High nutrition and quality attributes	Food crops which are safe for human consumption, limited processing for industrial crops to yield a useful product
Available genetic resources	Necessary germplasm to support breeding efforts, secondary or tertiary gene pool assets
Broadly adapted or adaptable	Crop species has a broad enough tolerance of different growing conditions to be grown over large areas, ability to crop to be acclimatized to new regions
Low input requirements	Reduced reliance on pesticide, irrigation, tillage, fertilizer and weed control as economic and conservation benefits
Enhanced ecosystem services	Value placed on ecosystem services may benefit the adoption of new crops
Culturally tenable	Careful attention to ethnobotany, equity, intellectual property and benefit-sharing considerations issues of new crops with historic use
Knowledge of the candidates disease and pest risk	Understanding of limiting biotic factors, risk of new crops being hosts to existing pathogens and potential for weediness
Low potential to become invasive or contaminate the gene pool of native species	Invasiveness potential of exotic species, influence of gene flow between crops and wild relatives

2.3.5 Reproductive biology and breeding strategies

The biology of perennial crops differs from that of annuals, and this is reflected in how selections are made during the breeding process and how populations are assembled. Selection cycles in perennial crops are longer than those of annual crops. For instance, in many perennial forage crops, each generation requires a minimum of three years for evaluation of all-around performance, compared to a single year for annual crops (Wilkins and Humphreys, 2003). This is due to differences in performance in newly seeded and established stands, and the requirement that perenniality is maintained. For species such as intermediate wheatgrass and Maximilian sunflower, two-year breeding cycles have been implemented to expedite the perennial grain breeding process (DeHaan et al., 2014; Van Tassel et al., 2014). However, these expedited cycles may limit the ability to assess long-term grain productivity during early rounds of selection (Wagoner and Schaeffer, 1990; Cattani, 2017). Selection strategies for perennial grains are determined by the reproductive biology of the species in question, as self-pollinated and cross-pollinated species respond differently to selection procedures. Perennials tend to experience high rates of intra- and interspecific gene flow, greater inbreeding depression (Byers and Waller, 1999; Morgan, 2001), and self-pollination is relatively rare compared to annual crop species (Morgan et al., 1997; Miller and Gross, 2011). The cross-pollinated nature of many perennial species results in greater heterozygosity, different patterns of population stratification, and linkage disequilibrium compared to self-pollinating species (Flint-Garcia et al., 2003). These factors influence the speed at which alleles are fixed within a population and the number of progeny required to find favourable recombinants when producing segregating populations. New methods such as

marker-assisted-selection (MAS) and genomic selection are becoming available in many species which previously had limited genomic resources (Zhang et al., 2016), and are expected to dramatically reduce the length of the selection cycles for many traits.

2.3.5.1 Self-pollinated species

Self-pollinated crops differ from cross-pollinated crops in that they often exhibit a high level of homogeneity due to systematic inbreeding (Falconer and Mackay, 1996). Under self-pollination, homozygosity increases and heterozygosity decreases by 50% in each generation of selfing, resulting in the rapid fixation of alleles. High levels of homozygosity may be achieved within five to six generations of self-pollination (Falconer and Mackay, 1996). In self-pollinated crops, the focus is often at the level of the individual, as a single individual may act as a founder for an improved population. In cross-pollinated crops, the focus is to shift the frequency of favourable alleles at the population level (Sleper and Poehlman, 2006). Populations of self-pollinated plants may consist of a single homozygous individual, or a mixture of unique homozygous individuals propagated by self-pollination. Self-pollinated crop breeding strategies are common in many annual crops such as wheat, rice, canola, soybean and corn (Sleper and Poehlman, 2006). Qualitative and additive traits can be advanced rapidly in self-pollinated crops as fixation occurs at a faster rate than through cross-pollination.

Common self-pollinated breeding procedures include bulk-population selection, pure-line selection, pedigree selection, single seed descent, and backcross breeding (Sleper and Poehlman, 2006). Doubled haploid production may also be used in both self- and cross-pollinated crops to rapidly generate pure-breeding homozygotes, though it may

be less effective in outcrossing species due to inbreeding depression, unpurged genetic load and loss of heterozygote advantage (Charlesworth and Willis, 2009; Van Tassel et al., 2010). A common self-pollinated breeding procedure used to develop perennial grain germplasm has been backcross breeding. Backcrossing strategies have proved to be effective for qualitative traits in perennial sorghum (Nabukalu and Cox, 2016), perennial rice (Cox et al., 2010) and perennial wheat (Jones et al., 1999). Backcrossing strategies to develop perennial grain germplasm generally involve crossing perennial crop wild relatives with domesticated crops and recurrently crossing the hybrid progeny back to either the annual or perennial parent while selecting for traits of interest such as perenniality. Once a trait has been introduced, uniform lines may be generated through self-pollination.

2.3.5.2 Cross-pollinated species

Cross-pollinated species exhibit higher levels of heterozygosity than self-pollinated species. Mechanisms which discourage inbreeding and promote outcrossing between individuals include the presence of self-incompatibility systems and dioecious (male and female flowers) or dichogamous modes of reproduction (pistils and stems mature at different times). Cross-pollinated crops are more phenotypically variable within populations than self-pollinated crops (Sleper and Poehlman, 2006). Cultivars are selected at the level of the population in cross-pollinated crops, with a greater emphasis on quantitatively inherited traits within populations. Selection is focused on increasing the frequency of favourable alleles while maintaining genetic variability in the population. Over time genetic shifts in the populations may occur due to environmental

selection pressures (Charles, 1961; Steiner et al., 1992). Allelic shifts may be beneficial, but over time a narrow genetic base may be detrimental, resulting in the loss of plasticity, inbreeding depression, or reduced seed set due to self-incompatibility.

Perennial forage crops such as red clover, alfalfa, intermediate wheatgrass, perennial ryegrass (*Lolium perenne* L.) and birdsfoot trefoil are generally bred using cross-pollinated crop breeding procedures. Recurrent phenotypic or genotypic selection, mass-selection, among-and-within-family (AWF) selection, half-sib and full-sib selection procedures with testcross or progeny tests are often used in outcrossing species (Sleper and Poehlman, 2006; Casler and Brummer, 2008; Annicchiarico et al., 2014). Traits such as flowering time are important in cross-pollinated crops to ensure synchronous flowering between genotypes, allowing cross-pollination and maximum seed set to occur. Common cross-pollinated selection procedures in perennial grains have included mass selection and half-sibling evaluation, which has been effective in increasing seed size and yield in intermediate wheatgrass (DeHaan et al., 2014, 2018) and Maximilian sunflower (Van Tassel et al., 2014).

2.3.5.3 Breeding with wide hybrids

Inter- and intra-specific hybridization between annual grain crops and perennial crop relatives is an attractive strategy for perennial grain development as characteristics from established crops such as high yield; quality and ease of management are readily available for exploitation from elite germplasm. Most interspecific hybrids of interest for perennial grain development involve crosses made between domesticated crops and members of their secondary gene pool, and are amenable to self-pollinated breeding

procedures (Table 2.3). Novel variation is generated, but not without challenges, such as; nuclear-cytoplasmic incompatibilities; and chromosomal non-disjunction or antagonistic pleiotropy between loci derived from annual and perennial genomes (Stebbins, 1958; Bomblies and Weigel, 2007).

Nuclear-cytoplasmic incompatibilities resulting in reduced vigor (Jan, 1992; Jan et al., 2014), chlorophyll deficiency (Bogdanova et al., 2012) and cytoplasmic male sterility (Chase, 2007) have been observed in some interspecific hybrids. While some interspecific hybrids exhibit great vegetative vigor, their use as grain crops may be limited by low fertility. Low fertility in some hybrids can increase the occurrence of certain diseases such as ergot (*Claviceps* spp.), which infect sterile florets, as observed in perennial cereal rye (Acharya et al., 2004; Cattani and Asselin, 2018b). Interspecific hybridization may result in the generation of meiotically unstable progeny, which undergo abnormal meiosis and fail to produce viable gametes. Chromosomal non-disjunction, particularly in early generations, has been observed in some *Triticum-Thinopyrum* hybrids (Suneson et al., 1963; Cai et al., 2001), perennial sunflower hybrids (Whelan, 1978; Sujatha, 2006) and perennial cereal rye (Reimann-Philipp, 1986). Meiotic stability is an important objective in perennial grain development for the ease of breeding efforts and ultimately seed production (Wagoner and Schaeffer, 1990; DeHaan et al., 2016).

While the combination of yield and quality characteristics of annuals with a perennial growth habit is a goal of these hybridizations, it is not always the case. Linkage drag or antagonistic pleiotropy, from either the annual parent (loss of perenniality) or wild relatives (loss of yield or quality characteristics) is another challenge facing hybrids.

For instance, perennial wheat, while capable of developing perenniating structures, does not exhibit the overwintering capacity of the perennial donor species such as intermediate wheatgrass in Manitoba (Cattani, personal communication; Hayes et al., 2018) or the ability to enter summer dormancy in Kansas (Cox et al., 2006). Though advancements are being made, annual traits such as yield and baking characteristics of perennial wheat differ from those of annual wheat (Murphy et al., 2009). The perennial growth habit is not a simple Mendelian trait, as perennials and annuals differ in many physiological traits (Garnier, 1992; Vico et al., 2016). Ensuring that hybrids maintain the desirable characteristics of both donor species, while minimizing linkage drag between parental genomes, is the goal of wide-hybridization methods.

Table 2.3: Interspecific hybrid perennial grain candidates.

Species	Common name	Crop type	Mating system	References
A. sativa x *Avena macrostachya* Bal. ex. Cos et Dur.	Perennial oat	Grain	Self-pollination	(Leggett, 1985)
H. annuus x *H. tuberosus*	Perennial sunflower	Oilseed	Outcrossing	(Kantar et al., 2016)
Helianthus divaricatus L. x *H. annuus*	Perennial sunflower	Oilseed	Outcrossing	(Kantar et al., 2016)
H. annuus x Helianthus pauciflorus Nutt.	Perennial sunflower	Oilseed	Outcrossing	(Van Tassel et al., 2014)
H. vulgare x *Hordeum jubatum* L.	Perennial barley	Grain	Outcrossing	(Kantar et al., 2016)
Oryza sativa L. x *Oryza longistaminata* A. Chev. & Roehr.	Perennial rice	Grain	Self-pollination	(Cox et al., 2010)
S. cereal x *S. montanum**	Perennial cereal rye	Grain	Outcrossing	(Reimann-Philipp, 1995)
S. bicolor x *S. halepense*	Perennial sorghum	Grain	Self-pollination	(Piper and Kulakow, 1994)
T. aestivum x *Thinopyrum Intermedium**	Perennial wheat	Grain	Self-pollination	(Jones et al., 1999)
T. aestivum x *Thinopyrum elongatum* (Host) D.R. Dewey	Perennial wheat	Grain	Self-pollination	(Larkin et al., 2014)
Triticum aestivum x *Thinopyrum ponticum* (Podp.) Z.-W. Liu & R.-C. Wang	Perennial wheat	Grain	Self-pollination	(Brasileiro-Vidal et al., 2005)
Z. mays x *Tripsacum dactyloides* (L.) L.	Perennial maize	Grain	Outcrossing	(Kantar et al., 2016)
Z. mays x *Zea diploperennis* Iltis, Doebley & Guzman	Perennial maize	Grain	Self-pollination	(Murray and Jessup, 2014)
Z. mays x Zea perennis (Hitchc.) Reeves & Manglesdorf	Perennial maize	Grain	Outcrossing	(Shaver, 1964)

* Species of interest at The University of Manitoba.

2.3.5.4 Breeding *de novo* domesticates

An alternative approach to using wide crosses to develop perennial grains is to focus breeding efforts on the improvement of wild species. In *de novo* domestication, wild populations of candidate species undergo recurrent cycles of selection for yield and domestication type traits (DeHaan and Van Tassel, 2014). Many domesticated crops have perennial relatives, and knowledge of the domestication process of annual counterparts could benefit breeding efforts (Table 2.4). Characteristics that define the differences between domesticated crops and their wild progenitors are commonly referred to as domestication syndrome traits (Meyer et al., 2012). Common domestication syndrome characteristics include the loss of seed dormancy, increased seed size, reduced shattering, increased apical dominance and changes in secondary metabolites, traits that facilitate the establishment, harvesting and utility of crops (Meyer et al., 2012). In many instances these traits may, or may not, be apparent in wild germplasm. Outcrossing is common in many herbaceous perennial species relative to annuals (Barrett et al., 1996; Morgan et al., 1997; Morgan, 2001) resulting in high levels of heterozygosity and cryptic variation due to the masking of allelic effects through dominance. The heterozygous state masks recessive alleles, which may contribute to genetic load (alleles which contribute negatively to fitness) (Marshall and Ludlam, 1989; Charlesworth et al., 1990; Barrett and Charlesworth, 1991) when inbreeding is induced through selection and a narrowing of the genetic base. Genetic load is purged from outcrossing populations at a slower rate than inbreeding populations (DeHaan and Van Tassel, 2014) as allele fixation can occur at a faster rate through self-pollinated selection procedures (Sleper and Poehlman, 2006). Fortunately, genetic control of some of the major domestication traits is relatively simple.

Quantitative trait loci (QTL) with major effects have been identified in many crop species (Simons, 2005; Gross and Olsen, 2010), though this varies on a species by species basis. Many species lack, or show weak differences between wild and domesticated populations, particularly perennial species such as perennial forage crops and tree crops (Meyer et al., 2012). Other crop species bear striking differences from their wild counterparts (Burke et al., 2002; Wills and Burke, 2007). Reports of repeated, independent, parallel selection on loci conferring domestication syndrome traits in various crop lineages suggests that wild perennial species could also harbor allelic variants favourable for crop development (Simons, 2005; Lin et al., 2012).

Table 2.4: Summary of mating systems of wild species with potential for perennial grain development.

Species	Common name	Crop type	Mating system	References
C. cajun	Pigeon pea	Legume	Outcrossing, facultative self-pollination	(Cumaraswamy and Bawa, 1989)
D. illinoensis	Illinois bundleflower	Legume	Outcrossing, facultative self-pollination	(Latting, 1961)
*E. canadensis**	Canada wild rye	Grain	Self-pollination, facultative outcrossing	(Sanders and Hamrick, 1980)
H. divaricatus	Woodland sunflower	Oilseed	Outcrossing	(Heiser et al., 1969)
*H. maximiliani**	Maximilian sunflower	Oilseed	Outcrossing	(Heiser et al., 1969)
*H. pauciflorus**	Stiff sunflower	Oilseed	Outcrossing	(Heiser et al., 1969)
Hordeum bulbosum L.	Perennial barley	Grain	Outcrossing	(Kakeda et al., 2008)
*L. lewisii**	Perennial flax	Oilseed	Outcrossing, facultative self-pollination	(Kearns and Inouye, 1994)
L. perenne	Perennial flax	Oilseed	Outcrossing	(Ockendon, 1968)
Physaria sp.	Bladderpod	Oilseed	Presumed Outcrossing**	(González-Paleo et al., 2016)
Microlaena stipoides (Labill.) R.Br.	Weeping rice grass	Grain	Self-pollination, facultative outcrossing	(Davies et al., 2005)
Silphium integrifolium Michx.	Rosin weed	Oilseed	Outcrossing	(Kantar et al., 2016)
Silphium laciniatum L.	Compass plant	Oilseed	Outcrossing	(Kantar et al., 2016)
*Th. intermedium**	Kernza (Intermediate Wheatgrass)	Oilseed	Outcrossing	(Knowles, 1977)
Zea diploperennis	Perennial maize	Grain	Outcrossing	(Tiffin and Gaut, 2001)

***** Species of interest at The University of Manitoba

** Some species of *Physaria* are reported as self-compatible, most species are outcrossing. Reproductive biology of the genus has not been studied extensively.

2.3.5.5 Techniques for generating novel germplasm

In instances where variation is not present within the breeding gene pool of a perennial grain candidate, alternative measures exist to generate variation through gene-editing, mutagenesis and genetic modification. Mutagenesis, which involves the exposure of seeds, pollen or ovules to radiation (x-rays, gamma rays) or chemical mutagens (ethyl methanesulfonate, dimethyl sulfate) has been explored in some perennial grain candidates. Shapter et al. (2013) combined mutagenesis with a candidate gene-approach to screen for induced mutations of value to the domestication of weeping ricegrass (*Microlaena stipoides*) with favourable results (Shapter et al., 2013). Genetic modification (GM) via techniques such as *Agrobacterium*-mediated transformation is another approach for the introduction of novel traits into perennials for species that are not recalcitrant to the technique (Casler and Brummer, 2008).

Emerging genome-editing techniques such as clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 systems, transcription activator-like effector nucleases (TALENs) and zinc finger nucleases (ZFNs) allow targeted modification of endogenous genes through insertions, deletions or replacements of nucleotides (Gaj, 2014). These techniques may be useful in the development perennial grains if accepted by the consumer. Classification of these techniques as genetic modification as opposed to conventional breeding techniques may influence the cost and delay commercialization in some countries through differences in regulatory requirements (Hartung and Schiemann, 2014). A number of domestication syndrome and quality traits in annual grains are controlled by alleles of large effect and their relatively simple genetic control is

paralleled across crop lineages. Targeting orthologs of *sh1* genes which contribute to shattering tolerance in sorghum, rice and maize (Lin et al., 2012) or *Q* genes in wheat and barley, which confer the brittle rachis free-threshing trait (Simons, 2005), or fatty acid desaturase genes controlling fatty acid profile in crops such as sunflower (Chapman and Burke, 2012) and canola (*Brassica napus*) (Peng et al., 2010), may be an effective strategy for fixing these traits in perennial crops without linkage drag and/or multiple backcross generations associated with introgression. The regulation of genome-editing techniques has yet to be established in many countries, therefore their potential for use in developing novel perennial grain germplasm, the influence this may have on potential markets and the cost of development remains to be seen (Hartung and Schiemann, 2014; Voytas and Gao, 2014). Under the Canadian framework the use of genome-editing techniques is permitted as the regulatory focus is not on the use of breeding methodology but rather the novelty of the end product. Plants generated through conventional, mutagenesis, and genetic modification breeding techniques are considered novel if they pose the potential for environmental impact and if the introduced trait in question is new to stable, cultivated populations of the species in question in Canada (Canadian Food Inspection Agency, 2019).

2.4 The genus *Helianthus*

2.4.1 Phylogenetic relationships

The genus *Helianthus* contains 52 species comprised of 67 taxa (Heiser et al., 1969; Stebbins, 2013). The various species and sub-species which comprise the genus occupy a variety of environmental niches which range from open plains to salt marshes (Kantar et al., 2015). The genus is noted for its reticulate evolution and has become a model for studying diploid and polyploid hybrid speciation (Timme et al., 2007). Interspecific hybridization and polyploidization events are common within the genus and multiple species are believed to have arisen through these mechanisms (Ungerer et al., 1998; Bock et al., 2014). The genus *Helianthus*, like many of the *Compositae*, has experienced a rapid diversification in the last 200 million years (Barker et al., 2008) and contains a diverse array of annual, perennial, and facultative perennial species (Heiser et al., 1969; Mandel et al., 2013a). The phylogeny of the perennial *Helianthus,* which contains 13 known polyploid species is particularly complex relative to the diploid annuals and is composed of two large polyphyletic clades, *H*. sect. *Divaricati*, and *H*. sect. *Ciliares*. The perennial *Helianthus* clades, which contain diploid (17n), tetraploid (34n) and hexaploid (51n) species, are believed to be ancestral to the two monophyletic annual clades (*H*. sect. *Helianthus*, *H*. sect. *Agrestis*) (Timme et al., 2007; Stephens et al., 2015). The perennial *Helianthus* are primarily members of the secondary or tertiary gene pools of *H. annuus* (Kantar et al., 2015). Recently, a new perennial species, *H. winteri*, has been described within the annual clade *H. sect. Helianthus*, which is morphologically similar, and clusters phylogenitcally with *H. annuus*. *H. winteri* is believed to be an example of a recently derived reversion of an annual *Helianthus* species to the perennial

habit (Stebbins, 2013; Moyers and Rieseberg, 2013), and is the sole reported perennial within the primary gene pool of *H. annuus*. Due to its apparent close relationship with *H. annuus*, and likely shared gene pool (Moyers and Rieseberg, 2013), *H. winteri* may warrant further investigation in perennial sunflower breeding efforts as a donor of perennial characteristics.

2.4.2 *Helianthus* species native to Manitoba

The western Canadian prairie represents the northernmost range of the genus *Helianthus*. The diversity of the genus is relatively unexplored in Canada compared to the United States (Seiler and Brothers, 1999). Seven *Helianthus* species are native to Manitoba (Scoggan, 1957; Rogers et al., 1982), consisting of two annual and five perennial species and hybrids (Table 2.5, Figure 2.1-2.7). The genus is primarily found in the southern quarter of the province, though there are some reports of Maximilian sunflower being observed as far North as Wekusko Lake in the central region of Manitoba (54°N) (Scoggan, 1957). Common habitats include roadside ditches, remnant prairie, and disturbed sites.

Table 2.5: *Helianthus* species described as native to Manitoba by Scoggan (1957) and Rogers et al. (1982).

Species	Common name	Lifecycle	Habitat
H. annuus L.	Common sunflower	Annual	Disturbed roadside ditches
H. petiolaris Nutt.	Prairie sunflower	Annual	Dry, sandy roadside ditches, sand dunes
H. maximiliani Schrader.	Maximilian sunflower	Perennial	Sandy-clay loam roadside ditches, dry to moderately wet soils
H. giganteus L.	Giant sunflower	Perennial	Mostly wet, open areas
H. nuttallii Torr.& A. Gray	Nuttall's sunflower	Perennial	Moist roadside ditches, sandy dry to wet soils
H. pauciflorus Nutt.	Stiff sunflower	Perennial	Dry, sandy roadside ditches, open habitats
H. tuberosus L.	Jerusalem artichoke	Perennial	Moist areas near woodlands

Though the various *Helianthus* species have habitat preferences, species ranges within Manitoba overlap and it is possible for multiple species to cohabit an area. Natural hybridization is noted to occur between several of the species native to Manitoba, though description of natural hybrids is limited to the hexaploid species. Long (1960) suggested that the diploid species *H. giganteus*, *H. grosseserratus* M. Martens, *H. maximiliani* and *H. nuttallii*, may compose a species complex of interfertile ecotypes based upon observations of high inter-fertility. *H. maximiliani* and *H. giganteus* are noted as being capable of producing highly fertile hybrids with normal meiotic behavior (Heiser et al., 1962; Whelan, 1978) and show a degree of morphological similarity, though they are distinguishable at the genetic level (Saftic-Pankovic et al., 2005; Timme et al., 2007). The hexaploid species *H. tuberosus* and *H. pauciflorus* are also known to be interfertile and hybridize to produce a naturally occurring hybrid *H. x laetiflorus* Pers. (pro sp.) (Heiser et al., 1969), which has been sporadically reported in Manitoba (Scoggan, 1957).

Figure 2.1: *Helianthus petiolaris*. (image: D.J. Cattani).

Figure 2.2: ***Helianthus maximiliani*** **(image: D.J. Cattani).**

Figure 2.3: *Helianthus pauciflorus* (image: D.J. Cattani).

Figure 2.4: *Helianthus tuberosus* (image: D.J. Cattani).

Figure 2.5: *Helianthus annuus* (image: D.J. Cattani).

Figure 2.6: *Helianthus nuttallii* (image: D.J. Cattani).

Figure 2.7: *Helianthus giganteus* (image: D.J. Cattani).

2.4.3 Domestication syndrome patterns in *Compositae* oilseeds

The suite of traits which differentiate crops from their wild progenitors are collectively known as the domestication syndrome (Harlan, 1992; Meyer et al., 2012), as described above. As domestication is an ongoing adaptive process, the trait combinations which comprise domestication syndromes differ from species to species and are not universal (Meyer et al., 2012). Parallels in domestication syndromes are often observed between crop lineages that have experienced common selection pressures. Common selection pressures in grains include the ability to successfully germinate and establish from greater burial depths following disturbance (loss of seed dormancy, increased seed size and seedling vigor) and traits which facilitate harvest (changes in branching and

stature, determinant growth habit and loss of shattering) (Zohary, 2004; Purugganan and Fuller, 2009; Lin et al., 2012). Changes in reproductive strategy (loss of self-incompatibility), changes in flowering phenology (loss of photoperiod sensitivity, synchronized flowering) and changes in secondary metabolites (loss of anti-nutritional compounds and changes in seed oil and protein content and composition) are also commonly associated with domestication in many species (Meyer et al., 2012).

The domestication syndrome in the *Compositae* is diverse due to differing end uses of species as oilseeds (*H. annuus*, *Guizotia abyssinica* (L. f.) Cass. ramtilla, *Carthamus tinctorius* L.), tuber and root vegetables (*H. tuberosus*, *Cichorium intybus* L., *Balsamorhiza sagittata* (Pursh) Nutt., *Smallanthus sonchifolius* (Poepp.) H. Rob.), edible leaves (*Cichorium endivia* L., *Cichorium intybus* L., *Cynara cardunculus* L., *Lactuca sativa* L.), phytochemical compounds such as latex (*Parthenium argentatum* A. Gray, *Taraxacum kok-saghyz* L.E.Rodin) or as ornamentals (*Gerbera* spp., *Solidago canadensis* L., *Zinnia* spp.) (Dempewolf et al., 2008). In the genus *Helianthus*, one species, *H. annuus*, is considered domesticated and a second species, *H. tuberosus*, is considered semi-domesticated. The domestication syndrome in *H. annuus* consists of a pronounced increase in apical dominance, resulting in a complete loss of branching, a single large central capitulum, increased achene length, width and weight, reduced seed shattering, reduced seed dormancy, the presence of self-compatibility alleles, and greater synchronization of pollen release and pistil receptivity allowing for self-pollination (Burke et al., 2002; Gandhi et al., 2005), increased seed oil content, changes to fatty acid composition (Chapman and Burke, 2012), and the loss of photoperiod sensitivity (Blackman, 2013).

Relative to other composites, loss of branching is a striking feature of *H. annuus's* domestication syndrome, and results in a considerable increase in seed size (Burke et al., 2002). Increased apical dominance is not observed in other oilseed composites such as safflower (*Carthamus tinctorius*) and noug (*Guizotia abyssinica*), which exhibit a domestication syndrome in the opposite direction of *H. annuus*, with selection favouring increased seed production though increased branching and capitula number relative to the wild type (Pearl et al., 2014; Dempewolf et al., 2015). Recently, restricted branching and unbranched individuals have been discovered in *H. maximiliani* populations, providing a substantial change in plant architecture (Van Tassel et al., 2014). The use of this trait in developing improved materials is under investigation to determine its usefulness and impact on other desirable traits (Van Tassel et al., 2014).

Currently, the only cultivated perennial *Helianthus* crop, *H. tuberosus*, is grown for its inulin-rich tubers and exhibits a different suite of domestication syndrome traits than species which are used for seed. *H. tuberosus* is considered semi-domesticated (Kays and Nottingham, 2007; Dempewolf et al., 2008) and exhibits few characteristics which differentiate domesticated and wild populations. *H. tuberosus* is propagated vegetatively with most clones being favourable selections of wild accessions and not the result of recurrent selection. "Domesticated" *H. tuberosus* clones exhibit reduced rhizome number and seed development (as a function of the loss of sexual fertility), increased leaf number, and increased dry matter allocated to tubers relative to wild individuals (Kays and Nottingham, 2007).

2.5 Biology of Maximilian Sunflower

2.5.1 Biological description

2.5.1.1 Habitat and establishment characteristics

Maximilian sunflower is an herbaceous perennial native to the central plains of North America, ranging from southern Canada to northern Mexico (Kawakami et al., 2011). Common habitats include dry to moist open prairie, disturbed sites along roadsides, railways and ditches. It is adapted to a range soil types ranging from sands to clays, performing best on medium sandy to clayey loams and poorly on gravel, dense clay and excessively saline soils (Texas Agric. Exp. Stn., 1979, Dietz et al., 1992). Seed propagation is enhanced in weed-free seedbeds (Texas Agric. Exp. Stn., 1979).

H. maximiliani tolerates moderate shade and drought, but is sensitive to heavy grazing (Dietz et al., 1992). Once rhizomes have developed and plants have entered dormancy, fire tolerance is good and it vigorously re-establishes following fire events. *H. maximiliani* is a competitive species and following disturbance can establish as a dominant or subdominant species (Wayne Polley et al., 2007; Dickson and Busby, 2009; Mangan et al., 2011). Various reports suggest *H. maximiliani*, like other sunflowers such as *H. annuus* and *H. tuberosus*, may produce allelopathic compounds, which may contribute to its competitiveness (Leather, 1983; Piper, 1998; Tesio et al., 2010, 2011). In dense vegetative cover, establishment by rhizomes is more effective than by seed (Call and Owens, 1986).

2.5.1.2 Phenology

Seed dormancy in *H. maximiliani* is the result of both physiological dormancy and the physical structure of the seed coat, similar to other wild *Helianthus* species (Heiser et al., 1969; Gay et al., 1991; Seiler, 1998). Physiological dormancy is variable and largely uncharacterized, but influenced both by environmental factors during seed development, and storage treatment following harvest (Corbineau et al., 1988). Dormancy is easily overcome by cold-stratification for approximately six weeks at 5C° (Bratcher et al., 1993). Scarification followed by treatment with gibberellic acid (Chandler and Jan, 1985) and seed coat removal is also effective for breaking seed dormancy.

Plants often develop a single, highly branched stem bearing many capitula and are reproductive in the first year of growth when established from seed. Plants established from rhizomes may exhibit several stems emerging per rhizome. Branching architecture is plastic, similar to other perennial sunflowers such as *H. tuberosus* (Kays and Nottingham, 2007). Plants that have undergone thinning exhibit a compensatory increase in branching and a reduction in stem height (Jackson, 1990). Maximilian sunflower is known to exhibit clinal variation in many size and life history traits across its native range. Heiser (1969) recognized a Northern and Southern ecotype of the species with populations from Manitoba (Northern) flowering in July, while populations from Texas (Southern) flowered in September-October. Plants continue to flower for an extended period that may continue until the first seasonal frost. Several weeks following anthesis, the capitula dry-down and shatter, releasing the seed. Seeds undergo cold stratification

over winter and may begin to emerge early in the growing season. New seedlings may emerge in either fall or early spring following cold stratification (Bratcher et al., 1993).

2.5.1.3 Description

Species descriptions of Maximilian sunflower are variable (Heiser et al., 1969; Rogers et al., 1982). Maximilian sunflower is described as containing well-developed rhizomes, with a variable number of light green to light red stems ranging in height from 0.5 - 3m tall. Leaves are numerous, sessile, or borne on short (2 cm) petioles, opposite at lower nodes, but mostly alternate. Leaves are lanceolate shaped, 10-30 cm long and 2.0-5.5 cm wide, light green to gray-green in colour and with entire to serrate margins. Leaves contain a characteristic single prominent central vein, are somewhat falcate, and conduplicate with scabrous to scabrous-hispidulous upper and lower leaf surfaces. Capitula are few to numerous, 1.6-2.8 cm in diameter and contain large showy ray flowers (2.5-4 cm) and bear achenes of 3-4 mm long. Bracts are loose, narrowly lance-shaped (1.5 mm wide) exceeding the captitula and with varying hairiness on margins (Heiser et al., 1969; Rogers et al., 1982). Variation between northern and southern populations appears to be clinal and continuous across the latitudinal gradient, with northern plants being shorter, flowering earlier, exhibiting smaller capitula, thinner stems, reduced plant biomass, faster growth rate, and greater numbers of capitula per unit of biomass than their southern counterparts (Heiser, 1969; Kawakami et al., 2011). Northern and southern populations show noted differences in cold acclimatization and freezing tolerances, with Manitoban populations exhibiting greater tolerance to freezing injury than Texas populations (Kawakami et al., 2014; Tetreault et al., 2016).

2.5.2 End uses

2.5.2.1 Oil content and quality

As a crop wild relative of annual sunflower, *H. maximiliani* provides a genetic resource for the improvement of the oil content and fatty acid composition of annual sunflower. As such, the oil content and fatty acid composition of Maximilian sunflower has been investigated by multiple researchers (Dorrell and Whelan, 1978; Seiler, 1994; Seiler and Brothers, 1999). Records in the USDA germplasm resources information network database report total seed oil content as a percentage of dry weight varying between 12-40% for this species (www.ars-grin.gov). Seiler (1999) found that wild harvested seed of *H. maximiliani* collected from the Canadian prairies had an oil content of 31.1%, exceeding the levels found in wild annual accessions of *H. annuus* (26.9%) and all other perennials species surveyed. Oil of Maximilian sunflower is primarily composed of the unsaturated fatty acids linoleic acid (C18:2) (58-87%) and oleic acid (C18:1) (8-28%) with low amounts of the saturated fatty acids palmitic acid (C16:0) (4-9%), stearic acid (C18:0)(2-4%), and trace amounts of behenic acid (C22:0) (>1%) (Dorrell and Whelan, 1978; Seiler, 1994; Seiler and Brothers, 1999; USDA-NCRS, 2017).

In a survey of 36 species (257 accessions), Maximilian sunflower was found to have the highest tocopherol content with 673 mg kg^{-1} of seed (Velasco et al., 2004). A-tocopherol, which composes the majority of the tocopherol profile (99.4%), is a form of vitamin E that is preferentially absorbed in human diets and among the most powerful natural fat-soluble anti-oxidants (Demurin et al., 1996). A-tocopherol has been targeted in sunflower breeding for its ability to increase the oxidative stability of sunflower oil (Demurin et al., 1996).

Oil concentration and fatty acid composition have been correlated with latitude of origin in *Helianthus* and is influenced by a combination of genetic and environmental effects (Seiler, 1994; Seiler and Brothers, 1999; Linder, 2000). Selection on genes involved in fatty acid biosynthesis has been important in the domestication and improvement of *H. annuus* (Chapman and Burke, 2012). Temperature during achene development and maturation has a large effect on final oil profile composition (Seiler, 1985), particularly when it comes to the ratio of oleic to linoleic acids. Oleic acid is the precursor to linoleic acid and their relative concentrations are correlated in sunflower (Chapman and Burke, 2012). This relationship is inversely proportional and influenced by growing temperature, as observed in other oilseeds such as rapeseed/canola (*Brassica napus*) and flax (*Linum usitatissimum* L.) (Canvin, 1965; Seiler, 1983a).

2.5.2.2 Biofuel feedstock

The potential for various *Helianthus* species to be used as a source of biofuels has been recognized and may be applicable to *H. maximiliani*. The primary storage carbohydrate in the root, rhizomes and tubers of sunflowers and other composite species is inulin (Dempewolf et al., 2008), which has been studied extensively in the tuber crop Jerusalem artichoke (Chubey and Dorrell, 1974; Kays and Nottingham, 2007; Gunnarsson et al., 2014). Maximilian sunflower produces an edible rhizome which has historically consumed as a root vegetable (Yanovsky, 1936; Kuhnlein and Turner, 1991; Moerman, 2010). The aboveground biomass of sunflowers (primarily *H. annuus* and *H. tuberosus*) have been investigated as a source of feedstock for lignocellulosic biofuel use and genetic variation for biofuel production traits such as lignin content and composition

have been noted (Ziebell et al., 2013). Maximilian sunflower is also capable of producing a high amount of lignified biomass under cultivated conditions, which could potentially serve as a feedstock for ligonocellulosic biofuels (Van Tassel pers. comm.).

2.5.2.3 Forage and feedstock value

Maximilian sunflower has been used for wildlife feed and cover, and is readily consumed by deer and livestock. The species is described as a desirable plant for livestock feeding due to its palatability and for revegetation due to its prolific seed production (Texas Agric. Exp. Stn., 1979). Maximilian sunflower seed is consumed by a wide range of bird and mammal species and its biomass may be used as cover (Dietz et al., 1992). Late season flowering provides pollinator species with pollen and nectar into the fall when other food sources are no longer available (Dietz et al., 1992). Forage quality of wild sunflowers was assessed by Seiler (Seiler, 1983b), who found that the crude protein content of the leaves and stem of *H. maximiliani* is lower than that of *H. annuus*, but comparable or greater than that of *H. tuberosus* during nearly all life stages. Seed crude protein content was greater in *H. maximiliani* (284.4 g kg^{-1}) than in wild *H. annuus* (180.0 g kg^{-1}) and a *H. annuus* commercial hybrid check (189.0 g kg^{-1}). *Helianthus maximiliani* seed ranks high amongst native plant materials and select crops for crude fat, gross energy and crude protein (Applegate, 2015).

2.5.3 Use as a genetic resource in sunflower breeding

Domesticated sunflower readily hybridizes with many of its wild relatives, which are important contributors to disease and pest resistance, abiotic stress tolerance and quality traits in modern cultivars of cultivated sunflower. Maximilian sunflower is a tertiary member of the *H. annuus* genepool (Kantar et al., 2015) and considerable research has been conducted into screening of traits to introgress from *H. maximiliani* to *H. annuus* (Table 2.6). Resistance to major pathogens of *H. annuus* such as sclerotinia [*Sclerotinia sclerotiorum* (Lib.) de Bary], the causative agent of white mold, sunflower stalk and head rot, has been a major focus of introgression efforts (Henn et al., 1998; Ronicke et al., 2004).

H. annuus and *H. maximiliani* are both diploid species (2n = 17) capable of hybridizing, but seed set and pollen viability in subsequent F_1 progeny is variable, ranging from >5 - 93.2% (Dorrell and Whelan, 1978; Atlagić et al., 1995). *H. annuus* x *H. maximiliani* F_1 hybrids cytologically exhibit a paracentric inversion and at least three translocations, which differentiate parental chromosomes (Dorrell and Whelan, 1978). Backcrossing with *H. annuus* rapidly restores pollen fertility and few multivalents are observed following the first backcross generation (Whelan and Dorrell, 1980).

Dorrell and Whelan (1978) suggested possible genotypic differences in the ability of *H. maximiliani* to hybridize with *H. annuus* and produce F_1 seed. This was supported by later studies showing genotypic differences in hybridization potential between the two species (Atlagić et al., 1995; Breton, 2010). In general, hand pollination results in few seeds, (with reports ranging from 0-122 seeds, depending on the genetic background (Breton, 2010), but techniques such as embryo rescue can improve hybridization success

(Espinasse et al., 1991; Breton, 2010). *Helianthus maximiliani* is amenable to tissue culture based techniques such as micropropagation via shoot apex culture (Dragana et al., 2001), cell suspension, protoplast culture (Polgár and Krasnyanski, 1992) and secondary somatic embryogenesis (Vasic et al., 2001). Interspecific hybridization between *H. maximiliani* and *H. annuus* has been achieved via protoplast fusion (Henn et al., 1998; Ronicke et al., 2004) in efforts to introduce greater Sclerotinia resistance into *H. annuus* (Cerboncini et al., 2002; Taski-Ajdukovic et al., 2006; Jan et al., 2007). Cytogenetic analyses of the resulting progeny show few meiotic abnormalities, making it favourable for the development of *H. annuus* x *H. maximiliani* hybrids (Binsfeld et al., 2001). The ability to transfer traits between *H. maximiliani* and *H. annuus* suggests that introgression in the opposite direction, from *H. annuus* to *H. maximiliani*, may also be a possible route to develop *H. maximiliani* as a perennial grain, similar to recent efforts being made to improve *H. pauciflorus*, another perennial sunflower candidate (Van Tassel et al., 2014).

Table 2.6: Traits of *H. maximiliani* of interest in the improvement of *H. annuus.*

Trait	Type	Reference
Resistance to sclerotinia (*Sclerotinia sclerotiorum*)	Disease resistance	(Cerboncini et al., 2002; Taski-Ajdukovic et al., 2006; Jan et al., 2007)
Resistance to leaf rust (*Puccinia helianthi* Schwein.)	Disease resistance	(Zimmer and Rehder, 1976)
Resistance to alternaria blight [*Alternaria helianthi* (Hansf.) Tubaki & Mishih.]	Disease resistance	(Sujatha and Prabakaran, 1997)
Resistance to downy mildew [*Plasmopara halstedii* (Farl.) Berl. & De Toni]	Disease resistance	(Vear, 2010)
Resistance to phomopsis (*Diaporthe helianthi* Munt.-Cvetk., Mihaljč. & M. Petrov)	Disease resistance	(Tikhomirov and Chiryaev, 2005)
Resistance to phoma (*Phoma macdonaldii* Boerema)	Disease resistance	(Škorić, 1992)
Resistance to charcoal rot [*Macrophomina phaseolina* (Tassi) Goid.]	Disease resistance	(Seiler, 2010)
Resistance to broomrape (*Orobanche* spp.)	Resistance to hemiparasites	(Jan and Fernández-Martínez, 2002)
Resistance to sunflower moth (*Homoeosoma nebulella* Denis & Schiffermüller)	Insect resistance	(Gershenzon and Mabry, 1984)
Resistance to stem weevil (*Cylindrocopturus adspersus* LeConte)	Insect resistance	(Charlet and Brewer, 1995)
Resistance to tobacco caterpillar (*Spodoptera litura* Fabricius)	Insect resistance	(Sujatha, 2006)
MAX1 cytoplasm	Male sterile cytoplasm	(Hahn and Friedt, 1994)
MAX2 cytoplasm	Male sterile cytoplasm	(Jan and Zhang, 1994)
MAX3 cytoplasm	Male sterile cytoplasm	(Liu et al., 2014)
RMAX1	Fertility restoration gene	(Miller and Wolf, 1991)
Rf 4	Fertility restoration gene	(Feng and Jan, 2008)
Water use efficiency	Ecophysiological trait	(Škorić, 1992)
Low temperature tolerance	Abiotic stress tolerance	(Tetreault et al., 2016)
Tolerance to erratic temperature	Abiotic stress tolerance	(Kantar et al., 2015)
High linoleic acid content	Seed oil fatty acid composition	(Seiler and Brothers, 1999)
Seed tocopherol concentration	Secondary metabolites	(Velasco et al., 2004)

2.5.4 Breeding efforts as a perennial crop

In comparison to cereal perennial grain candidates such as perennial wheat and intermediate wheatgrass, which have a history as forages, breeding Maximilian sunflower as a perennial oilseed crop is relatively recent. Efforts to develop Maximilian sunflower cultivars for conservation and pasture uses were conducted throughout the late 1960s and into 1970s with two open pollinated cultivars, 'Aztec' and 'Prairie Gold' released for commercial use by the United States Department of Agriculture Natural Resources Conservation Service in 1978 (Texas Agric. Exp. Stn., 1979; USDA-NCRS, 2017). 'Aztec' was developed for wildlife feed, livestock forage cover, use as a natural hedge, filter-strip, and an ornamental landscape plant. 'Prairie Gold' was released for landscape reclamation and wildlife feed plantings. Both cultivars were selected for vigor and stand establishment in regions of Oklahoma and Texas (Aztec) or Kansas and further north (Prairie Gold). Agronomic research on Maximilian sunflower as a perennial grain candidate began at The Land Institute in the 1980s (Jackson, 1990), eventually culminating in the launch of the first breeding program focused on developing Maximilian sunflower in 2002 (Cox et al., 2002).

Selection on seed size was effective in this species and by 2012; following three rounds of recurrent selection, average seed size had increased by 240 % (Van Tassel et al., 2014). One of the early agronomic challenges observed in Maximilian sunflower was that due to profuse branching, capitula were produced at various heights and the high density of lignified stems made mechanical harvesting difficult. Restricted branching genotypes have been recently identified (Van Tassel et al., 2014) and may alleviate

harvest challenges, as proposed in other crops facing similar challenges (Baldanzi et al., 2003; Galwey et al., 2003).

2.6 Genotype x environment interactions in plant breeding

Breeding crops with appropriate adaptation to their growth environment is a critical factor in the success of plant breeding programs. Genotype x environment (G x E) interactions are important indicators of genotype performance, the stability of traits across a variety of growing conditions and can inform the selection of appropriate genotypes in different growth environments (Allard and Bradshaw, 1964; Lin and Binns, 1988). Genotype x environment interactions are typically determined through the analysis of variance components of multi-environment trials (Annicchiarico, 2002). Strong G x E interactions are indicative of differential plant performance amongst test environments (local adaptation), while weak G X E interactions are indicative of consistent plant performance (broad adaptation). Genotype x environment interactions can be subdivided into predictable and non-predictable components of variation. Predictable components include genotype x location (G x L) interactions, while non-predictable components include genotype x year (G x Y) and genotype x location x year interactions (G x L x Y). Estimates of G x Y and G x L x Y interactions give an indication of genotypic performance at a given point in time, while G x L interactions give an indication of future performance (Lin and Binns, 1988, Annicchiarico, 2002).

Estimates of G x L interactions are a useful indicator of adaptation to local growing conditions such as soil properties or stable climatic factors and provide an indication of trait stability across locations (Annicchiarico, 2002). Strong G x L

interactions may result in inconsistent performance amongst growth environments and may be indicative of maladaptive phenology, such as initiation of flowering period, timing of fall dormancy or spring emergence out of synch with optimal growing conditions (Casler et al., 2007; Des Marais et al., 2013). Genotype x location interactions may be exploited by growing locally adapted germplasm, such as materials adapted to specific soil conditions, and which may outperform broadly adapted materials under those conditions. Alternatively, G x L interactions may be minimized by growing widely adapted materials which perform consistently amongst environments (Annicchiarico, 2002).

2.7 Genomic and breeding resources for Maximilian Sunflower

2.7.1 Next-generation sequencing

Genomic-assisted breeding is a multifaceted approach to the improvement of crop species using DNA based molecular markers and can greatly increase the speed at which crop domestication and improvement occurs (Xu and Crouch, 2008; Varshney et al., 2009).The emergence of next-generation sequencing (NGS) technologies has resulted in the rapid and cost-effective development of genomic resources for many wild species (Narum et al., 2013), which have traditionally lacked resources for their domestication and improvement (Henry, 2012; Sedbrook et al., 2014). Next generation sequencing has been applied to characterize the transcriptome of Maximilian sunflower, identify single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) between ecologically divergent populations for use as molecular markers (Kawakami et al., 2014).

Surveys of SNP diversity in public seed bank accessions has been used to clarify phylogenetic relationships in *Helianthus* (Baute et al., 2016).

Genotype-by-sequencing (GBS) is a molecular marker platform for developing reduced representation SNP libraries for species with high diversity and large complex genomes using NGS (Elshire et al., 2011; Poland et al., 2012). The major advantage of approaches such as GBS is that they may allow plant breeders to gauge population stratification, linkage disequilibrium, perform trait mapping, and conduct genomic selection in novel germplasm or species without prior knowledge of the genome (Elshire et al., 2011; Zhang et al., 2016). Genotype-by-sequencing employs methylation sensitive restriction enzymes to reduce genome complexity, particularly highly methylated repetitive regions, which are indicative of inactive DNA prior to sequencing (Elshire et al., 2011; Poland and Rife, 2012). Following digestion, individual samples are ligated to barcoded adaptors and then pooled into highly multiplexed libraries for sequencing on NGS platforms using a shotgun approach. Sequencing reads can be aligned *de novo* (Kharabian-Masouleh et al., 2011; Poland and Rife, 2012) or through the use of a reference genome of the species (Stanton-Geddes et al., 2013; Mamidi et al., 2014) or of a homeologous relative (Guajardo et al., 2015; Baute et al., 2016), making this technique suitable for species that lack specific genomic information. The number of markers derived using GBS relies on the diversity of the sampled genotypes, restriction enzyme choice, the number of unique tags generated in the digestion process, the depth of sequencing coverage, and the ability to accurately align reads to each other, or to a reference genome. Marker numbers can range from several thousand to hundreds of thousands and are highly scalable to project end-use (Elshire et al., 2011; Poland and

Rife, 2012). Depth of coverage in GBS is generally low, but can be modulated by the selection of the restriction enzyme and level of multiplexing, as the depth a coverage is negatively correlated with the multiplexing pool size (Elshire et al., 2011). Low cost per sample, choice of marker number, coverage and number of individuals sequenced per library makes GBS flexible for multiple applications (Poland and Rife, 2012; Zhang et al., 2016).

2.7.2 Reference species

H. annuus offers a wealth of genomic tools and resources for the development of useful markers in sunflower wild relatives, including a reference genome (Kane et al., 2011; Badouin et al., 2017), transcriptome resources (Bowers et al., 2012a), high-density genetic maps (Talukder et al., 2014) and marker arrays (Bowers et al., 2012b). Recently the reference genome of *H. annuus* (Kane et al., 2011; Badouin et al., 2017) (www.sunflowergenome.org) has been employed to align GBS reads from wild sunflowers with success, including the diploid perennials *H. maximiliani*, *H. divaricatus*, *H. giganteus*, *H. grosseserratus, H. hirsutus* Raf., *H. nuttallii* and *H. winteri*, the perennial polyploid *H. tuberosus* (Baute et al., 2016) and diploid annuals *H. anomalus* S.F. Blake, *H. argophyllus* Torr. & A. Gray, *H. bolanderi* A. Gray, *H. debilis* Nutt., *H. decapetalus* L., *H. neglectus* Heiser, *H. niveus* (Benth.) Brandegee, *H. paradoxus* Heiser, *H. petiolaris, H. praecox* Engelm. & A. Gray and *H. strumosus* L. (Owens et al., 2016; Baute et al., 2016), providing useful genomic insights into these species.

Genetic control of the domestication syndrome of sunflower and a variety of agronomic and quality traits have been explored through bi-parental mapping providing a

useful guide for approaching the domestication and improvement of *H. maximiliani.* Candidate genes for domestication and quality traits have been identified in *H. annuus* and can serve as potential tools for the development of *H. maximiliani* as a perennial grain (Chapman et al., 2008b; Bachlava et al., 2010; Blackman et al., 2011; Chapman and Burke, 2012; Henry et al., 2014; Nambeesan et al., 2015).

2.7.3 Genetic mapping concepts

Genetic mapping involves the association between a phenotypic (physical), biochemical or DNA-based marker and the inheritance of a trait of interest (Sax, 1923; Tanksley, 1983; Mohan et al., 1997). Different genetic mapping approaches have been developed to suit a variety of self- and cross-pollinated crops based on their reproductive biology and respective forms of population stratification. The basis of all genetic mapping approaches are the Mendelian concepts of independent segregation and independent assortment of genes and alleles during meiosis (Sax, 1923; Tanksley, 1983). Considerations need to be made when dealing with factors that result in deviations from expected Mendelian ratios, such as population stratification and linkage disequilibrium. Modern genetic mapping approaches can be broadly defined as; A) linkage-based methods; and B) association mapping based approaches.

2.7.3.1 Population stratification

Population stratification is an important concept in genetic mapping that influences the ability to detect true marker-phenotype relationships and reject spurious

long-distance or unlinked associations (Pritchard et al., 2000a; Flint-Garcia et al., 2003). Population stratification (also referred to as population structure) is the presence of systematic differences in allele frequencies between populations as the result of neutral and non-neutral processes. Population stratification is measured as the average proportion of alleles which are identical by descent (IBD) between two individuals. Structure is generated through various processes that disrupt Hardy-Weinberg equilibrium, including neutral processes such as non-random mating, isolation by distance and random genetic drift which impact linkage disequilibrium (LD) throughout the genome (Soto-Cerda and Cloutier, 2012). Non-neutral processes, which contribute to population stratification in crops, include selection for different end-uses and pedigree selection, and the loss of allelic diversity due to population bottlenecks which contribute to patterns of population stratification (Somers et al., 2007; Mandel et al., 2013b; Soto-Cerda et al., 2014). In wild *Helianthus*, population stratification is low, in part due to the obligate outcrossing nature of the genus, large effective population sizes, and high genetic diversity of many species (Mandel et al., 2013a; Owens et al., 2016). High population stratification in the self-compatible domesticated sunflower is largely attributed to breeding history, resulting in distinct clusters between unbranched maintainer lines and branched restorer lines, reflecting the history of breeding inbred lines for hybrid seed production (Mandel et al., 2013b). Accounting for LD is critical for management of type I and type II error rates in genetic mapping (Yu et al., 2006). In linkage based mapping, population stratification is accounted for by pedigree, but in association studies, where the demographic history of a population is unclear, stratification influences the ability to detect marker-phenotype associations.

2.7.3.2 Linkage disequilibrium

Linkage disequilibrium (LD), also known as gametic phase disequilibrium or allelic association, is the non-random association between two alleles at different loci within a genome (Flint-Garcia et al., 2003). Linkage disequilibrium is not the same as the concept of linkage. Linkage refers to the estimated physical distance between two loci on the same chromosome assessed through recombination fraction (i.e.: the ratio of the number of recombined gametes relative to the total gametes produced). Linkage disequilibrium is the measurement of the deviation of two alleles at different loci regardless of physical location within the genome from expected independent assortment within a population as expected under Hardy-Weinberg equilibrium. While population stratification influences allelic frequencies at the level of the genome, LD influences patterns at the level of genetic loci (Soto-Cerda and Cloutier, 2012). Linkage disequilibrium is influenced by both allelic diversity within a population and recombination points between loci, and as such, it is influenced by most population genetic processes (Flint-Garcia et al., 2003). Linkage disequilibrium may result in spurious marker-phenotype associations through non-random correlations between causative loci and other linked or unlinked genomic regions. Linkage disequilibrium is generally greater in self-pollinated than cross-pollinated crops, as in autogamous crops the detection of recombination is less effective due to higher homozygosity and fewer cross-pollination events to produce allelic diversity (Flint-Garcia et al., 2003). As LD is affected by recombination frequency, it is not distributed evenly throughout genomes, and may present itself as blocks of low recombination (i.e. haplotype or linkage blocks) (Fisher, 1954). The extent to which linkage blocks persist is of interest in plant breeding,

which is dependent of the generation of favourable genetic combinations between parental materials (Hanson, 1959).

In terms of marker coverage and population size in genetic mapping, fewer genetic markers are required to cover a given genomic region when linkage blocks are large, as blocks are inherited as a unit until disrupted by recombination (Fisher, 1954). Conversely, the presence of linkage blocks can reduce mapping resolution due to reduced recombination frequency and the ability to produce segregating progeny for loci of interest. Knowledge of LD in a given species can help determine optimum numbers of individuals and marker saturation for effective genetic mapping (Liu et al., 2013). In wild out-crossing *H. annuus* and other wild sunflowers, linkage disequilibrium decays rapidly (negligible at ~ 200bp) relative to its self-compatible domesticated counterpart (~negligible at ~ 1100bp) (Liu et al., 2006). Linkage disequilibrium is variable across the genome of *H. annuus*, partially due to the effects of selection for domestication syndrome traits with islands of elevated LD noted in regions of the genome which harbor QTL for traits such as branching (Mandel et al., 2013b).

2.7.4 Linkage and QTL mapping

Quantitative trait loci (QTL) are a type of marker-phenotype associations (Flint-Garcia et al., 2003). The purpose of QTL mapping is to locate markers in LD with the genetic control of quantitative traits. Common QTL mapping techniques include single marker regression (single factor, analysis of variance based) (Edwards et al., 1987), simple interval mapping (single factor, likelihood based) (Lander and Botstein, 1989), composite interval mapping (multiple factors, multiple-regression based) (Zeng, 1994),

multiple interval mapping (multiple factors, likelihood based) (Kao et al., 1999), inclusive composite interval mapping (stepwise regression of factors, likelihood based) (Li et al., 2008), Bayesian interval mapping (multiple factors, Bayesian approach) (Banerjee et al., 2008) and variations thereof. While a genetic linkage map with ordered molecular markers are generally used for QTL mapping, it is not strictly necessary for all types of QTL mapping, as QTL mapping as a technique predates molecular markers (Sax, 1923).

In QTL mapping studies, population structure is controlled through the generation of mapping populations in which total alleles are limited, and expected allele frequencies are highly predictable in progeny. Mapping studies based upon QTL analysis are common in self-pollinating crops, but can be challenging in outcrossing species due to self-incompatibility influencing the ability to easily generate informative populations and the need for accurate parental linkage phasing (Wu et al., 2002). Common types of populations for QTL mapping in inbreeding crops include doubled haploid (DH), recombinant inbred lines (RIL), and F_2 populations designed to segregate for particular characters of interest. The level of LD can be manipulated by the generation of recombinants during the population development process and desired level of mapping resolution. Applying QTL analysis to self-pollinating perennial grain candidates such as perennial rice (Hu et al., 2001), sorghum (Paterson et al., 1995), and maize (Westerbergh and Doebley, 2004) has been successful in investigating the genetic control of perennial growth habit.

Applying QTL analysis to cross-pollinated crops is possible, but due to traits such as self-incompatibility and inbreeding depression, development of inbred populations is

limited. Due to outcrossing, parents in mapping populations are highly heterozygous, with as many as four alleles segregating at a given locus in diploids and any given marker may segregate in two (1:1), three (1:2:1) or four (1:1:1:1) genotypic classes (Warnke et al., 2004). Options for linkage mapping in outcrossing species involve the use of various types of F_1 and full or half sib intercross-derived populations (Song et al., 1999). A common approach in outcrossing species is to apply the double "pseudo-testcross" approach in which markers which are identified as segregating in a 1:1 pattern from each parent are mapped in both parents independently, in a similar fashion to backcrossed populations in self-pollinated populations (Margarido et al., 2007). This approach produces two linkage maps, one corresponding to each parent of the cross. This method complicates analysis as additional anchoring markers are required to join the maps. It is only suitable when all markers follow the same segregation patterns and may exclude large portions of the genome. Pseudo-testcross populations, which are F_1 populations derived from two highly heterozygous parents, are often employed in outcrossing species such as perennial forage and turf crops as well as woody perennials (Pearl et al., 2004; Wang et al., 2011, 2014; Honig et al., 2014). An alternate approach is to employ a multi-point maximum likelihood-based approach for the estimation of parental linkage phase and recombination fractions as proposed by Wu (2002). Maximum likelihood approaches allow for a combination of marker types to be employed as *a priori* estimation of parental linkage phase is not required (Wu et al., 2002; Margarido et al., 2007). This approach has successfully been applied in outcrossing plant species such as perennial ryegrass (Do Canto et al., 2018), sugarcane (Balsalobre et al., 2017) and loblolly pine (Xiong et al., 2016) amongst others.

Limitations of QTL mapping approaches are their ability to handle allelic diversity, limitations on recombination (and resulting mapping resolution), over-estimation of allelic effects, and instability of QTL in different genetic backgrounds (Holland, 2007). Advancements in sequencing technology have allowed large marker arrays to be developed rapidly and at low cost, allowing for powerful association mapping techniques to support and enhance genetic mapping efforts.

2.7.5 Association mapping

Association mapping, also known as linkage disequilibrium based mapping, differs from linkage-based QTL mapping approaches in that it does not depend on highly structured populations with few allelic variants at each locus, but instead uses populations with ideally limited population structure, and multiple alleles present at each locus to maximize diversity. Association mapping is used to identify significant allele-frequency differences between unrelated individuals (Pritchard et al., 2000b). The underlying assumption in association mapping is that loci under divergent selection for traits will show differing patterns of LD than selectively neutral loci. Association mapping differs from linkage mapping techniques in that it is reliant on historical recombination to break down linkage disequilibrium between loci, and is suitable for out-crossing species where linkage disequilibrium decay occurs rapidly. As linkage blocks are expected to be smaller is association-mapping studies, marker saturation needs to be greater to capture recombination events and achieve higher quality mapping resolution (Liu et al., 2013). With the emergence of NGS, association mapping based techniques are becoming more popular in plant breeding studies (Soto-Cerda and Cloutier, 2012). Population size in

association mapping often is larger than in linkage mapping to capture recombination events, greater allelic diversity, and the presence of rare or low frequency alleles that may go undetected in smaller mapping panels (Liu et al., 2013). A variety of association mapping approaches have been developed which include different strategies to control population structure and expected LD. Structured populations are employed in multi-parent advanced generation intercrosses (MAGIC), nested association mapping populations (NAM), *a priori* case and control panels and transmission disequilibrium tests (TDT) approaches may be used to control for population stratification and patterns of LD by design. In genomic control (GC) approaches, random sets of markers are employed to assess population structure, the magnitude of which is used as a uniform adjustment to the critical value for significance tests of candidate loci (Devlin and Roeder, 1999; Pritchard et al., 2000b). In structured association (SA) approaches, population structure and LD are assessed and accounted for using linear statistical models (Soto-Cerda and Cloutier, 2012). Structured association uses random markers to estimate population structure (*Q*) and assigns individuals to population clusters, using programs such as STRUCTURE (Pritchard et al., 2000a). This method uses a Bayesian clustering approach and a user defined number of clusters to assign individuals to subpopulations. Subpopulation information can subsequently be integrated into a general linear model (GLM) to aid in correcting for false associations (Pritchard et al., 2000a). Elaborations of this approach include the use of mixed linear models (MLM) and the incorporation of pairwise family relatedness (kinship) within population clusters (*K*-matrix); though using *Q* in conjunction with *K* may result in an over-correction of relatedness and loss of association detection power (Würschum, 2012). An alternate approach is to use principal

component analysis (PCA) on whole genome marker data to define population stratification (Price et al., 2006) and incorporate this information into the statistical model as a covariate. Association mapping offers advantages over linkage-based mapping approaches, such as greater potential mapping resolution due to greater historic recombination, the assessment a greater number of alleles, and reduced applicability of QTL to a wider range of genetic backgrounds than bi-parental mapping populations (Holland, 2007).

2.7.6 Environmental association analyses

Environmental association analysis (EAA) is an approach related to association mapping to identify signatures of adaptive genetic variation and relate them to environmental variation (Rellstab et al., 2015). This method differs from linkage and association mapping in that in place of phenotypes, environmental variables are employed to discover candidate genes or patterns of population stratification/LD associated with environmental differences. Linkage and association mapping are top-down approaches where phenotypes are first identified, followed by genetic characterization, while the strength of EAA is that it is a bottom-up 'genotype-first' approach to detecting candidate regions for adaptive differences between populations without the need for *a priori* phenotypic characterization (Joost et al., 2013). Abiotic data is becoming widely available through public databases such as WorldClim (Hijmans et al., 2005) and ISRIC (Hengl et al., 2014) for use by researchers in combination with NGS to make this technique possible. Similar to association mapping, EAA rests on the concept that loci under selection will exhibit greater LD with the environments they are

adapted to than neutral loci. This technique has been employed in natural populations of plant and animal species, which are cumbersome, technically challenging, or costly to phenotype intensely to identify loci of interest (Sork et al., 2013). Recently EAA has proven to be a useful tool to identify candidate genes and genomic regions conferring abiotic stress tolerance in wild species (Eckert et al., 2010; Hamlin and Arnold, 2015), model species (Yoder et al., 2014; Friesen et al., 2014) and crop wild relatives in crops such as barley (Fang et al., 2014), soybean (Anderson et al., 2016), wheat (Brunazzi et al., 2018) and *Brassica* species (Zulliger et al., 2013). Combining EAA with common garden experiments can allow for the simultaneous detection of both genetic and environmental contributions to adaptive variation and explore the complex relationships between phenotype, genotype and environment (De Villemereuil et al., 2015).

A variety of methods have been utilized to detect allele frequency differences indicative of population differentiation in EAA. Common techniques include population differentiation based *Fst* outlier tests, Mantel tests, logistic regression, general or mixed linear model methods and canonical correlation analysis (CCA) (Foll and Gaggiotti, 2008; Narum and Hess, 2011). In *Fst* outlier tests, loci are tested for deviations in genetic distance relative to neutral expectations. Mantel tests, and partial Mantel tests estimate the strength of correlations between distance matrices (neutral population structure, environmental distance and differentiation at particular loci) of either two (Mantel test) or more distances matrices (partial Mantel test). Logistic regression tests whether an environmental factor is associated with the presence or absence of a given allele or allele frequencies at a given loci. Univariate or multivariate general linear models include incorporating allele frequencies as a response variable and environmental variables as a

fixed effect (GLM) or mixed linear models (MLM) in which allele frequencies are treated as the response variable, environmental factors as fixed effects, and neutral genetic structure as a random effect. Multivariate techniques such as canonical correlation analysis (CCA) are also employed in some instances, which finds the maximum correlations between combinations of marker data and environmental variables (De Mita et al., 2013; Rellstab et al., 2015).

Similar to association mapping, the challenge in EAA is differentiating between neutral and non-neutral patterns of population stratification, correcting for population stratification due to neutral processes (e.g.: genetic drift, isolation by distance, gene flow, mutation) and relating these patterns to environmental variation in place of a given phenotype (Sork et al., 2013; Rellstab et al., 2015). Neutral population stratification may be accounted for in EAA through the incorporation of spatial data in regression based models (coordinates of origin, pairwise Euclidean distances between samples), the use of selectively neutral markers acting as controls (synonymous sites, non-coding regions, non-outlier regions), and estimation or whole-genome estimates of population stratification using many markers, under the assumption that loci following neutral processes greatly outnumber those under selection. Similar to association mapping, the application of PCA or the calculation of a *Q*-matrix to categorize population structure and *K*-matrices to control for kinship can be conducted to account for neutral differentiation when incorporated into GLM or MLM approaches.

2.7.7 Management of error rates in genetic mapping and genomic association studies

The analysis of marker-trait associations necessitates the testing of a large number of hypothesis, which may rapidly inflate the family-wise error rate of a given study. This results in an increased probability of producing a false positive result (type I error). While the use of GLM or MLM with appropriate covariates may reduce the potential for type I or type II errors, these models do not explicitly protect against increases in the family-wise error rate (Bradbury et al., 2007).

In QTL studies, a minimum logarithm of odds (LOD) score between 2 and 3, corresponding to a type I error rate of 5% is often employed as a baseline to declare QTL (Lander and Botstein, 1989). To correct for multiple hypothesis testing permutation analyses are often employed at either the chromosome or genome-wide level to calculate an appropriate LOD threshold to declare significance (Churchill and Doerge, 1994).

To account for multiple testing in genome-wide association studies, such as association mapping or EAA, permutation testing or a *p*-value adjustment such as the Benjamini and Hochberg (Benjamini and Hochberg, 1995) false-discovery rate criteria, Storey's *Q*-value approximation (Storey and Tibshirani, 2003) or a Bonferroni adjustment may be conducted. Though considered a gold standard for determining significance in GWAS studies employing GLM, permutation testing is computationally intensive and is not always a feasible option when running a large number of markers. Methodology for performing permutation testing for MLM has not been developed (Bradbury et al., 2007) and currently there is no widely accepted protocol for the control of false-discovery rate in MLM. *P*-value adjustments are conservative and may result in

false-negative results, particularly in the presence of linkage disequilibrium between adjacent markers. *P*-value adjustment tests are best suited for studies in which linkage disequilibrium is low, as these methods necessitate independence between hypotheses. In MLM alternative approaches such as the selection of an arbitrary number of candidate SNPs based upon a chosen probability value have been employed to manage type II error rates in the presence of linkage disequilibrium as a less stringent alternative to identify candidate regions of interest (Stanton-Geddes et al., 2013; Kang et al., 2015; Henning et al., 2017; Sakiroglu and Brummer, 2017).

2.8 Statement on the synthesis of relevant literature

To date literature on both breeding perennial grains and oilseeds and perennial *Helianthus* is limited. Research efforts for the neo-domestication of these species are reliant on the synthesis of relevant literature in the areas of plant breeding, genetics, agronomy and ecology to achieve research objectives. The purpose of this review was to summarize the current state of knowledge available for the development of Maximilian sunflower as a perennial oilseed crop. Existing knowledge of Maximilian sunflower and related species provide a framework for how to approach the challenges of breeding this species as a perennial crop. Significant knowledge gaps exist in regards to how to approach breeding efforts in this species, and perennial grains cropping systems in general. Characterization of the genetics, diversity, agronomic characteristics, and ecology of Maximilian sunflower will provide answers to how to approach breeding this crop wild relative of cultivated sunflower. The research undertaken in this thesis was

designed to characterize Maximilian sunflower and apply modern breeding technologies to address the strengths, limitations and challenges associated with breeding this species.

CHAPTER 3.0: Phenotypic variation and clinal differentiation of perennial *Helianthus* germplasm in Manitoba

3.1 Abstract

The Canadian prairies are a highly productive agricultural zone with primarily annual grain production and a short growing season. Heat units and precipitation prior to sowing and following harvest of annual crops represent potential resources that could be utilized to increase productivity. Perennial crops may take advantage of seasonal resources that would otherwise not be utilized to their full extent in annual cropping systems. In this study, candidate perennial species from the genus *Helianthus* were characterized for phenotypic variation under replicated field trials to determine the presence of genetic variation that may be useful in the development of a perennial oilseed crop from locally sourced germplasm. Clonally replicated common gardens were established in Carman and Winnipeg, Manitoba to evaluate local germplasm over the course of two growing seasons following the year of establishment. Clinal variation in timing of anthesis across latitudinal and temperature gradients in southern Manitoba were uncovered in all species examined, and may inform future plant performance and selection for either broadly or locally adapted materials. The presence of genetic variation for timing of anthesis, capitulum size, timing of shattering, and lodging indicate that the local gene pool of perennial *Helianthus* has sufficient diversity to make selections for important phenological and agronomic characteristics for production in southern Manitoba. Given these results, the development of a perennial sunflower crop to extend

the growing season on the Canadian prairies appears favourable due to the existing standing genetic variation in locally adapted germplasm.

3.2 Introduction

The Canadian prairie is a highly productive agricultural region with annual grain crops comprising the majority of seeded acres. Agricultural production in this region is limited by the frost-free period, availability of heat units, and precipitation throughout the growing season (Fowler, 2012). The frost-free period for much of the Canadian prairies is approximately 90-120 days in most years, imposing restrictions on crop production (Nadler and Bullock, 2011). The length of the growing season limits crop choice to early maturing varieties and cold-hardy species that are capable of producing acceptable yields under these extreme conditions. Adverse conditions early in the growing season, such as excess moisture, can further limit growing season potential for annual crops by delaying seeding or preventing adequate stand establishment, and may provide opportunities for weedy species to establish. Gaps in active plant growth between cropping cycles, such as during the pre-seeding and post-harvest periods in annual crops results in potential unexploited heat units and precipitation that may be available for agricultural production (Cattani and Asselin, 2018b).

Use of fall-seeded winter annual crops or perennial crops can extend the growing season in western Canada by taking advantage of available heat units earlier in the growing season and following the primary crop harvest (Larsen et al., 2018; Cattani and Asselin, 2018b). Living ground cover and active plant growth during these periods can provide agronomic benefits, such as suppression of weeds (Moyer et al., 2000;

Blackshaw et al., 2008), reduced nutrient leaching (Crews, 2005; Culman et al., 2013), pollinator habitat (Hopwood, 2008; Saunders et al., 2013) and carbon sequestration (Conant et al., 2001; Tilman, 2007).

Cold tolerance has been a considerable challenge in fall-seeded winter cereals such as winter wheat (*T. aestivum* L.) and fall-rye (*Secale cereale* L.) (Fowler, 2012; Larsen et al., 2018), which is a limiting factor in the adoption of these crops. Perennial and biennial forage species improve ground-cover, but their use is limited by lack of adapted materials and concerns of potential invasiveness, weedy characteristics and reversion to ferality of some introduced species in native habitats. Concerns have been raised over introduced species such of bromegrasses (*Bromus spp.*) (Upadhyaya et al., 1986; Otfinowski et al., 2007; Otfinowski and Kenkel, 2008), sweet-clovers (*Melilotus* spp.) (Turkington et al., 1978), alfalfa (*Medicago sativa* L.) (Bagavathiannan et al., 2010) and crested wheatgrass [*Agropyron cristatum* (L.) Gaertn.] (Henderson and Naeth, 2005) and their impact on native plant populations in unmanaged habitats.

Conversely, native plant species adapted to these growing conditions are plentiful, but their use in agricultural production is limited. Use of native species in agriculture on the Canadian prairies has been restricted to forage or biomass production, with limited breeding efforts (Applegate, 2015; Jefferson et al., 2013; Biligetu et al., 2014). Interest in the agronomic benefits of perennial crops and performance under different environmental conditions has driven interest in incorporating native perennial species into agricultural production.

Native plant species may be better suited for dealing with climatic variation than non-native species due to adaptation to local conditions (Willms et al., 2005). Additional

benefits of incorporating native species into agricultural production include habitat preservation and reclamation in western Canada. In Alberta, Saskatchewan and Manitoba, 61, 89 and 99%, respectively, of native mixed-grass prairie has been displaced, mainly by the cultivation of annual grain crops (Samson and Knopf, 1994). Incorporating native plant species into agricultural production may mitigate habitat loss for pollinators and other beneficial species. In some situations, native forage species can outperform improved introduced species under unfertilized conditions (Johnston et al., 1968; Knowles, 1987). Some native species exhibit yields and quality comparable to introduced species later in the growing season that may be useful for late-season or stockpile grazing (Applegate et al., 2015; Biligetu et al., 2014; Schellenberg et al., 2017).

Developing perennial grains from native plant species may provide benefits over conventional annual cereal and perennial forage crops. Perennial grains with acceptable grain yield, or mixed-use perennial grain/forage, or grain/bioenergy crops from adapted native species, could enhance not only cropping system diversity, but end-use potential as well (Bell et al., 2010; Kantar et al., 2016; Cattani and Asselin, 2018b, Ryan et al., 2018). A large number of native species produce edible grains such as Canada wild rye (*Elymus Canadensis* L.), blue grama [*Bouteloua gracilis* (Willd. Ex Kunth) Lag. Ex Griffiths], Indian ricegrass [*Achnatherum hymenoides* (Roem. & Schult.) Barkworth], sand dropseed [*Sporobolus cryptandrus* (Torr.) A. Gray], prairie junegrass [*Koeleria cristata* (Ledeb.) Schult.] and members of the genus *Linum* and *Helianthus* (Yanovsky, 1936; Kuhnlein and Turner, 1991; Moerman, 2010). Achenes of perennial species of *Helianthus* are edible and contain a fatty acid composition primarily composed of linoleic and oleic acid. The western Canadian prairies represent the northernmost range for the genus and the

various species present are an unexploited source of crop germplasm (Seiler and Brothers, 1999). Members of the genus *Helianthus* have the potential for use in agriculture and their potential as perennial oilseed crops is an active area of research (Van Tassel et al., 2014; Kantar et al., 2014, 2018; Cattani and Asselin, 2018b).

The objective of this study was to characterize phenotypic variation in perennial *Helianthus* species collected across southern Manitoba, and to determine the presence of potential ecotypes associated with environmental factors to inform the selection of either locally or broadly adapted materials and future plant performance. Clonally replicated genotypes were assessed in multi-year field trials under common garden conditions for agronomic characteristics to determine the presence of genetic variation for crop development. Correlations and principal component regression analyses were employed to determine relationships between phenotype and environmental factors for the identification of factors influencing clinal differentiation in the genus *Helianthus* in Manitoba.

3.3 Materials and methods

3.3.1 Plant sampling

Accessions were collected throughout southern Manitoba and from Agriculture and Agri-Food Canada's perennial sunflower nursery located in Morden, Manitoba (herein referred to as the Morden collection) in 2011 and 2012. The GPS locations of samples collected from southern Manitoba were recorded and mapped (Figure 3.1). Genotypes were assigned to a species group using a combination of visual identification of diagnostic features described by Heiser et al. (1969) and Rogers et al. (1982), and seed

morphological characters (data not shown). The initial collection consisted of accessions of the diploid species *H. giganteus* L., *H. maximiliani* Schrad., *H. nuttallii* Torr. & A. Gray and hexaploid species *H. pauciflorus* Nutt. and *H. tuberosus* L..

Accessions were propagated vegetatively, through stem or rhizome cuttings, under growth chamber conditions to provide adequate replicates for field evaluation. As few accessions of the hexaploid species were amenable to vegetative cloning, *H. tuberosus* and *H. pauciflorus* were omitted from analysis and the diploid *H. giganteus*, *H. maximiliani* and *H. nuttallii* were analyzed.

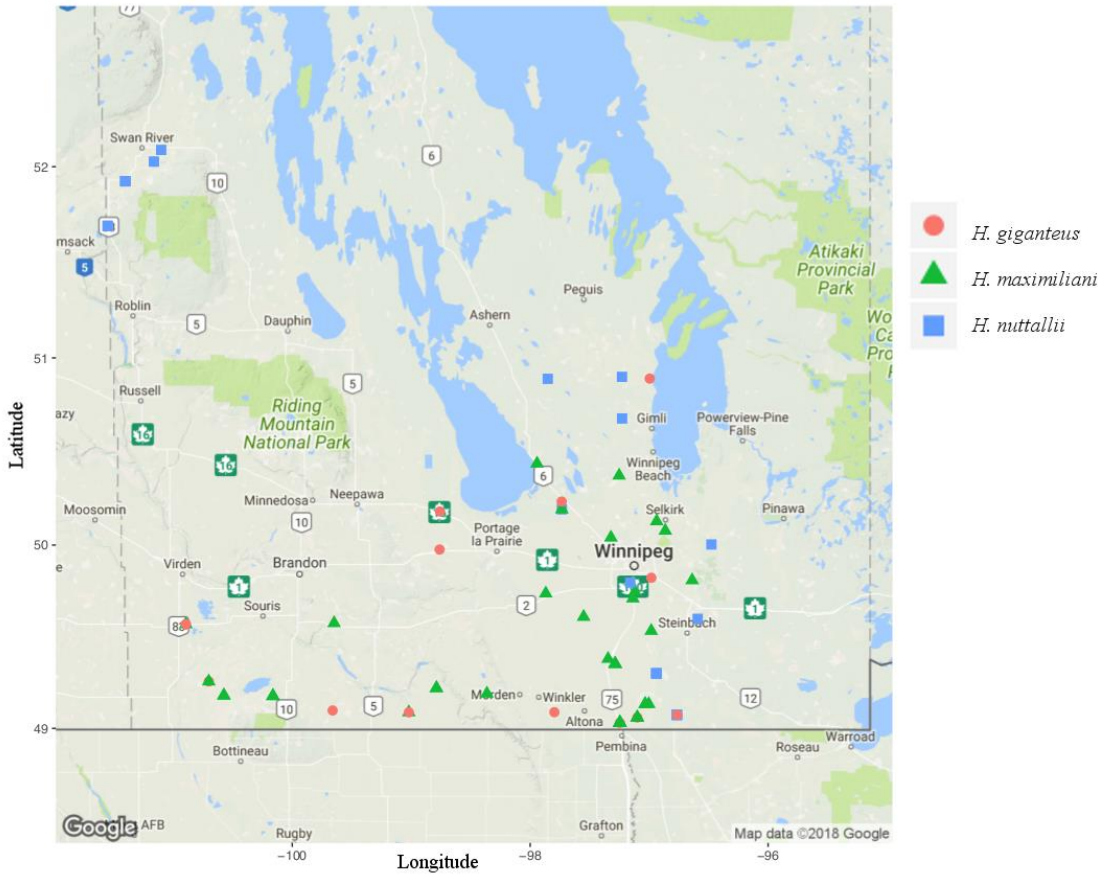


Figure 3.1: Collection sites for 78 diploid *Helianthus* accessions analyzed under replicated field conditions in Carman and Winnipeg in 2013 and 2014.

3.3.2 Phenotypic characterization

In 2013 and 2014 plants were assessed under field conditions at sites in Winnipeg (49º48′55.08º N, 97°7′14.54° W) and Carman (49°29′46.70° N, 98°2′43.64° W), Manitoba. Each site contained four replicates of each genotype, established between June and September 2012. A total of 127 genotypes were evaluated. These included 77 accessions obtained from collections in southern Manitoba and 50 accessions donated from the Agriculture and Agri-Food Canada perennial sunflower collection in Morden, Manitoba. The genotypes consisted of 31 accessions of *H. giganteus*, 62 accessions of *H. maximiliani* and 34 accessions of *H. nuttallii*.

Days to first anthesis, days to fifth anthesis, flowering synchronicity, average capitulum diameter, timing of shattering, branching architecture and lodging score were evaluated in 2013 and 2014 in Carman and Winnipeg on each clone individually. Days to first anthesis was scored as the Julian date the first capitulum exhibited pollen dehiscence, days to fifth anthesis was measured as the Julian date at which five capitula on the same plant had reached anthesis, while flowering synchronicity was measured as the interval in days between first and fifth anthesis. Timing of shattering was measured by randomly marking a capitulum at anthesis and measuring the number of days until achenes could visibly be observed being lost while shaking the stem gently. Branching architecture and lodging were assessed at the end of the growing season in October. Branching was scored on a categorical basis as one of five general branching patterns (Supplemental Figure 3.1), while lodging was assessed by the degree at which plants deviated from an upright position relative to the soil surface. Plants were scored for lodging on a six-point scale with plants standing at a 75°-90° angle from the soil surface

being classified as a 1, 60°-75° as 2, 45°-60° as 3, 30°-45° as 4, 15°-30° as 5 and 0°-15° as 6.

3.3.3 Statistical analysis of field data

3.3.3.1 Field experiment

All statistical analyses were conducted using SAS 9.3 (SAS Institute, 2011). Phenotypic data were examined for normality and homogeneity of residuals to meet the assumptions of analysis of variance (ANOVA) using the SAS UNIVARIATE and GLIMMIX procedures with the COVTEST HOMOGENEITY option employed to aid in determining the best model fit. Pearson's correlation coefficients were calculated between continuous phenotypic traits using the SAS CORR procedure by species. All data with the exception of branching score were examined using the SAS MIXED procedure with the METHOD=REML option. Differences between species were first tested with a model consisting of species, location, year, and location x year as fixed effects and genotype(species), rep(location x year), species x location, species x year and species x location x year were included as random effects in the model. To explore within species effects data were analyzed by species with a model consisting of genotype, location, year and location x year as a fixed effects with rep(location x year), genotype x location, genotype x year and genotype x location x year as random effects with the ddfm=KR option invoked and lsmeans were generated. Location, year and their interaction were considered fixed effects as per recommendations of Yang (2010) as fewer than 10 levels were present. For continuous variables the METHOD=TYPE 3

option was also run in SAS MIXED to generate a partitioning variance components to estimate the relative contribution of fixed and random effects. Branching score was examined using the same models in SAS GLIMMIX with the DIST=MULT option invoked allowing for the analysis of nominal data. In instances where heterogeneity of variances was suspected, the GROUP option within the RANDOM statement in SAS MIXED procedure was employed and the model with the lowest AIC score was selected. *Post-hoc* pairwise separation of lsmeans in both the MIXED and GLIMMIX procedures were conducted using the ADJUST=TUKEY option in the LSMEANS statement to calculate Tukey's honest significance test (alpha=0.05) to account for multiple comparisons.

3.3.3.2 Environmental correlations and principal component regression analysis

Spearman's ranked correlation coefficients were calculated between environmental variables and lsmeans of continuous traits showing significant genotypic differences when run using SAS PROC CORR by species. Environmental data were collected from two sources; the first was a summary of public Environment Canada weather data for Manitoba compiled by Nadler (2007); and the second was from the public global bioclimatic repository WorldClim (Hijmans et al., 2005).

The Environment Canada data consisted of ten variables compiled from 30 years of daily climate data from 1971 through 2000 and included information of monthly temperature norms, precipitation, and soil properties. Mean values for each weather station were compiled and analyzed initially by Nadler (2007). Each accession studied in

the field was assigned to the nearest Environment Canada weather station to the accession collection site of origin. WorldClim data consisted of 30 environmental variables interpolated from over 50 years of global weather station data at a 1 km^2 resolution. WorldClim variables for the coordinates of each accession were extracted using the R package *Raster* (Hijmans et al., 2016). The forty environmental variables employed with their full descriptions are listed in Table 3.1. Environmental variables were screened for influential outliers using the SAS UNVIARIATE procedure and log-transformed when necessary. The Morden collection was omitted from this analysis due to the lack of collection site origin data, leaving 77 genotypes for environmental correlation analysis. SAS UNIVARIATE procedure plots and SAS CORR procedure scatterplots were examined for the presence of outliers, which were removed when necessary prior to principal component regression.

Principal component regression (PCR) was run using the SAS PLS procedure to assess environmental variables which best explained phenotypic means amongst collection sites by species. As the traits days to first anthesis and days to fifth anthesis were highly correlated in all three species examined, PCR was run only on days to first anthesis. These analyses were conducted using the Method=PCR option on centered and scaled data. To avoid the potential over-fitting of models, the appropriate number of factors to retain was determined by examining van der Voet's statistic (van der Voet, 1994) and the absolute minimum predicted residual sum of squares (PRESS) statistics throughout model iterations. The cross validation option "CV=RANDOM" was employed and was run with 10,000 iterations. Traits which could be explained by one or more factors had their models pruned iteratively until no variables could be removed by

examining variable importance plots (VIP) and removing factors below a VIP threshold of 0.8.

Table 3.1: Bioclimatic variables collected from WorldClim or Nadler (2007) for use in correlation and principal component regression analysis.

Variable	Description	Source
AWHC	Mean soil water holding capacity (AWHC)	Nadler (2007)
BIO1	Mean annual mean temperature (°C)	WorldClim
BIO10	Mean temperature of warmest quarter* (°C)	WorldClim
BIO11	Mean temperature of coldest quarter (°C)	WorldClim
BIO12	Mean annual precipitation (mm)	WorldClim
BIO13	Mean total precipitation of the wettest month (mm)	WorldClim
BIO14	Mean total precipitation of driest month (mm)	WorldClim
BIO15	Mean total precipitation seasonality (coefficient of variation) (mm)	WorldClim
BIO16	Mean total precipitation of the wettest quarter (mm)	WorldClim
BIO17	Mean total precipitation of the driest quarter (mm)	WorldClim
BIO18	Mean total precipitation of the warmest quarter (mm)	WorldClim
BIO19	Mean total precipitation of the coldest quarter (mm)	WorldClim
BIO2	Mean diurnal temperature range (mean of monthly (max temp - min temp)) (°C)	WorldClim
BIO3	Isothermality (BIO2/BIO7) (* 100) (°C)	WorldClim
BIO4	Mean temperature seasonality (standard deviation *100) (°C)	WorldClim
BIO5	Mean maximum temperature of warmest month (°C)	WorldClim
BIO6	Mean minimum temperature of the coldest month (°C)	WorldClim
BIO7	Mean annual temperature range (BIO5-BIO6) (°C)	WorldClim
BIO8	Mean temperature of wettest quarter (°C)	WorldClim
BIO9	Mean temperature of driest quarter (°C)	WorldClim
ELV	Elevation (m)	WorldClim
FFD0	Mean number of frost free days above 0°C (days)	Nadler (2007)
FFDNEG2	Mean number of frost free days above -2°C (days)	Nadler (2007)
FFF0	Mean Julian date of first fall frost below 0°C (days)	Nadler (2007)
FFFNEG2	Mean Julian date of first fall frost below -2°C (days)	Nadler (2007)
FRAIN	Mean growing season rainfall for forage crops (mm)	Nadler (2007)
FWDEM	Mean crop water demand for forage crops (mm)	Nadler (2007)
GDDFOR	Mean growing degree days for forage crops (GDD base 5°C)	Nadler (2007)

Variable	Description	Source
LAT	Collection site latitude (°)	GPS
LOG	Collection site longitude (°)	GPS
LSF0	Mean Julian date of last spring frost below 0°C (days)	Nadler (2007)
LSFNEG2	Mean Julian date of last spring frost below -2°C (days)	Nadler (2007)
TMEAN10	Mean daily temperature for the month of October (°C)	WorldClim
TMEAN11	Mean daily temperature for the month of November (°C)	WorldClim
TMEAN4	Mean daily temperature for the month of April (°C)	WorldClim
TMEAN5	Mean daily temperature for the month of May (°C)	WorldClim
TMEAN6	Mean daily temperature for the month of June (°C)	WorldClim
TMEAN7	Mean daily temperature for the month of July (°C)	WorldClim
TMEAN8	Mean daily temperature for the month of August (°C)	WorldClim
TMEAN9	Mean daily temperature for the month of September (°C)	WorldClim

* Note: quarter = Period of three successive months within a year.

3.4 Results

3.4.1: Environmental conditions in common garden experiment

Average daily temperature and precipitation patterns varied by year and were relatively consistent among sites. Average daily temperatures from May through to September inclusive were greater in 2013 (Carman: 16.17°C, Winnipeg: 16.43°C) than 2014 (Carman: 15.57°C, Winnipeg: 15.9°C) as evidenced by growing degree days, while cumulative precipitation was greater at both sites in 2014 (Figure 3.2-3.3). The influence of year was most notable for precipitation early in the growing season, around the first week of June in 2013 and late in the growing season in 2014, from late August and throughout September.

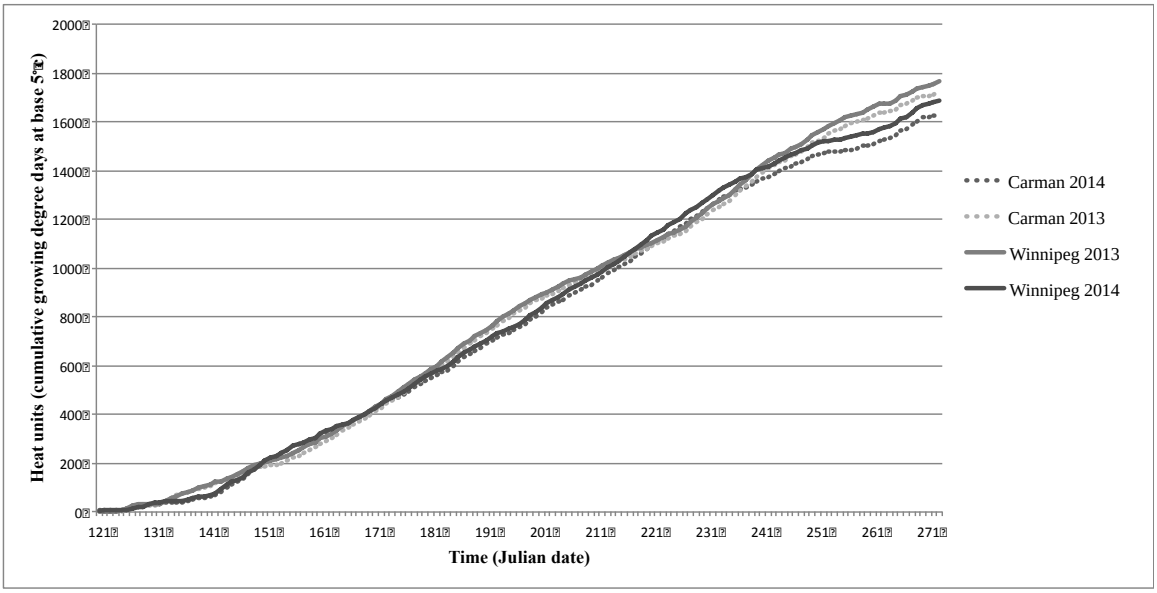


Figure 3.2: Cumulative growing degree days calculated at a base temperature of 5°C for common garden sites from the months of May through September inclusive in Carman and Winnipeg in 2013 and 2014.

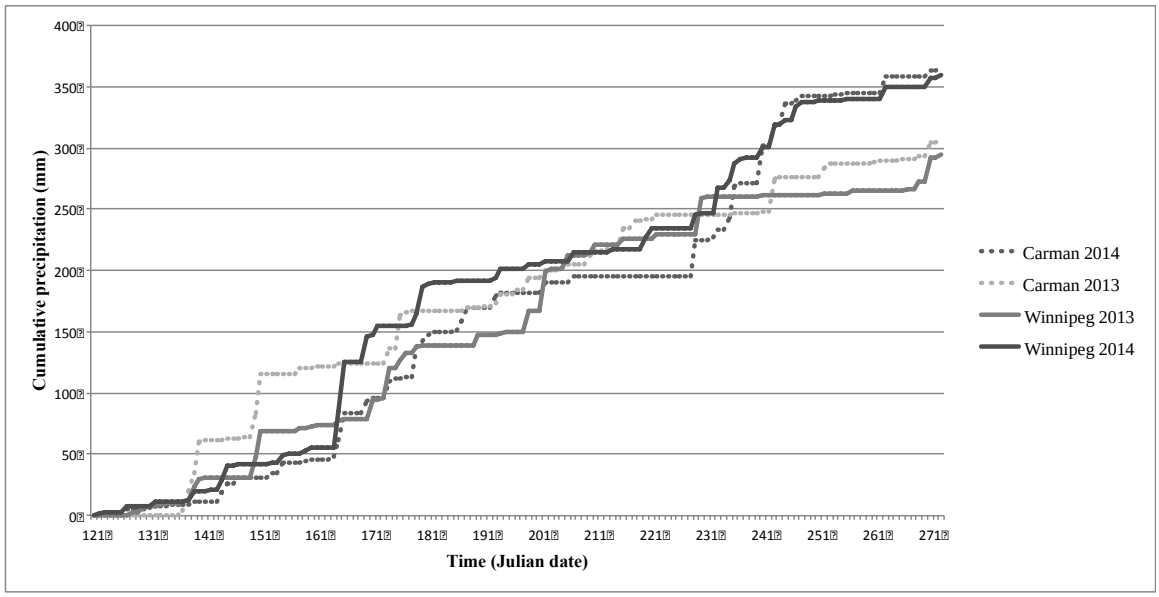


Figure 3.3: Cumulative precipitation in mm at common garden sites from the months of May through September inclusive in Carman and Winnipeg in 2013 and 2014.

3.4.2 Phenotypic characterization

Phenotypic ranges overlapped for all three species examined (Supplemental Figure 3.2-3.7), and while significant differences were detected amongst genotypes within species, few differences were observed between species. Significant differences between species were detected for the traits days of first anthesis (p=0.015) and days to fifth anthesis (p=0.037), but not for any other measured trait. *Helianthus nuttallii* exhibited slightly earlier mean first and fifth anthesis than either *H. giganteus* or *H. maximiliani*. No significant differences were detected between *H. giganteus* and *H. maximiliani* for any measured trait. The effect of genotype was significant within all species for the traits days to first anthesis, days to fifth anthesis, average capitulum diameter, timing of shatter and plant lodging (Table 3.2) indicating a heritable component to the variation, and that these species may respond to directional selection. Genotypic

differences in mean timing of anthesis was considerable for all three species, with the earliest flowering genotypes reaching anthesis approximately three to four weeks earlier than the latest flowering genotypes (Table 3.2). These differences in initiation of anthesis represent a significant portion of the growing season in Southern Manitoba in which the frost-free period is limited to 90-120 days. The distribution of means is indicative of primarily a quantitative mode of inheritance for all traits examined. Peaks indicative of potential major genes of influence were observed for the traits days to first and fifth anthesis and average capitulum diameter in all species (Supplemental Figure 3.2-3.4). Species and genotypes within species were not significantly different for branching score, with all species showing a score range of 1-5 and a mode of 2 (primarily apically branched).

The coefficient of variation (CV) for each continuous trait was relatively consistent amongst site years and species (Supplemental Table 3.1), likely due to the relative proximity of testing sites, and similar phenotypic ranges amongst species. Days to first anthesis and days to fifth anthesis showed consistent, low CVs ranging from 2.8%-7.5% and 2.6%-7.1% respectively, while average capitulum diameter was moderately larger, ranging from 11.2%-17.9% amongst species. The CVs for lodging score ranged from 21.6%-31.6% among site years and species. Timing of shatter showed the largest CVs and a greater range amongst site years by species, ranging from 29.5%-39.9%, 30.5%-44.0% and 22.3%-47.3% for *H. giganteus*, *H. maximiliani* and *H. nuttallii*, respectively. Flowering synchronicity showed the highest CVs ranging from 39.4% to 63.7% amongst site years and species.

Significant correlations between phenotypic traits were detected within each species, though few of these correlations showed a strong relationship. Days to first anthesis was strongly correlated with days to fifth anthesis with $r > 0.87$ for all species. All other correlations were relatively low, falling below $r = 0.5$ (Supplemental Table 3.2-3.4). The low correlations observed between the measured phenotypic traits indicate a lack of direct trade-offs between most traits, and simultaneous selection for or against trait pairs should be possible with relative ease.

Table 3.2: Observed ranges of genotypic means and significance of genotypes as assessed in PROC MIXED by species for the six continuous traits measured under field conditions in Carman and Winnipeg in 2013 and 2014.

	Observed genotypic range		Results of Type III tests of fixed effects for genotype			
	Min	Max	Numerator DF†	Denominator DF	*F* value	Pr>*F*
H. giganteus						
Days to anthesis (mean date)	July 17th	August 7th	30	24.5	11.67	<0.001*
Days to fifth anthesis (mean date)	July 23rd	August 13th	30	2236.0	11.11	<0.001*
Average capitulum diameter (mm)	9.9 mm	16.1 mm	30	30.7	8.69	<0.001*
Timing of shatter (days)	15.1 days	28.9 days	30	28.2	3.34	<0.001*
Flowering synchronicity (days)	4.3 days	8.2 days	30	21.2	1.80	0.082
Lodging score	2.7	4.0	30	18.0	2.69	0.015*
H. maximiliani						
Days to anthesis (mean date)	July 24th	August 17th	61	63.4	4.91	<0.001*
Days to fifth anthesis (mean date)	July 30th	August 21st	61	468.0	6.09	<0.001*
Average capitulum diameter (mm)	11.2 mm	18.4 mm	61	56.1	8.78	<0.001*
Timing of shatter (days)	13.5 days	26.2 days	61	94.2	2.77	<0.001*
Flowering synchronicity (days)	2.6 days	7.6 days	61	44.8	0.72	0.882
Lodging score	2.3	4.4	61	57.4	2.43	<0.001*
H. nuttallii						
Days to anthesis (mean date)	July 10th	August 6th	33	317.0	13.42	<0.001*
Days to fifth anthesis (mean date)	July 14th	August 13th	33	249.0	14.88	<0.001*
Average capitulum diameter (mm)	9.6 mm	15.1 mm	33	30.1	2.63	0.009*
Timing of shatter (days)	15.5 days	24.1 days	33	305.0	4.05	<0.001*
Flowering synchronicity (days)	3.8 days	7.6 days	33	16.7	0.97	0.548
Lodging score	2.6	3.6	33	76.2	2.80	<0.001*

*=Significant effect of genotype at alpha=0.05 as per PROC MIXED METHOD=REML

†=Note: DF=degrees of freedom

3.4.3 Genotype x environment interactions

The partitioning of variances revealed that the effect of genotype consumed the largest portion of the sums of squares for the response variables days to anthesis, days to fifth anthesis and average capitulum diameter for all species (Table 3.3-3.5, Supplemental Table 3.5-3.13). Genotype explained 56.4%, 50.0% and 54.8% of the variation in days to first anthesis, 58.7%, 46.5% and 64.0% of the variation in days to fifth anthesis and 66.0%. 53.6% and 42.9% of the variation in average capitulum diameter observed in *H. giganteus*, *H. maximiliani* and *H. nuttallii*, respectively. This pattern may be due to the fact that growing conditions were similar amongst sites and years therefore genotypes performed consistently. Genotype was the second largest source of variation for timing of shattering following residual effects, with genotype contributing 27.5%, 20.8% and 20.2% of the variation in *H. giganteus*, *H. maximiliani* and *H. nuttallii*, respectively (Tables 3.3-3.5, Supplemental Table 3.14-3.16). Lodging score showed a similar pattern with 24.8% (*H. giganteus*), 18.9% (*H. maximiliani*) and 19.3% (*H. nuttallii*) of the of the sums of squares being attributed to genotype but with residual effects accounting for the largest proportion of the sums of squares (Tables 3.3-3.5, Supplemental Table 3.14). For flowering synchronicity, unexplained residual effects accounted for the greatest amount of variation, followed by the effect of year, which explained 30.3%, 24.1% and 23.4% of the observed variation in *H. giganteus*, *H. maximiliani* and *H. nuttallii*, respectively (3-3.5, Supplemental Table 3.17-3.22).

Interactions with location and year were low and most non-significant for all three species (Tables 3.3-3.5, Supplemental Table 3.5-3.22). Genotype x location interactions were low for all observed traits, falling below 10% for all species (Table 3.3). This

interaction was only significant for timing of first anthesis in *H. giganteus* and *H. maximiliani* (Supplemental Table 3.5-3.6) and capitulum size in *H. giganteus* and *H. nuttallii* (Supplemental Table 3.11 & 3.13) and explained little of the variation in these traits. The low interaction between genotype and location indicates that the genotypes were broadly adapted to the growing conditions found in Winnipeg and Carman in 2013 and 2014. Similarly, genotype x year interactions were low for all traits observed, which is not surprising as similar heat units and precipitation were accumulated in both test years. The remaining sources of variation: location, location x year, rep(location x year) and genotype x location x year contributions were either non-significant or represented small proportions to the overall variation (Tables 3.3-3.5, Supplemental Table 3.5-3.22).

Table 3.3: Percentage of sums of squares contributed to genotypic and environment effects for combined phenotypic data for six traits collected in Carman and Winnipeg in the years 2013 and 2014 as calculated using method=TYPE 3 in SAS MIXED for *H. giganteus.*

Source	Days to anthesis	Days to fifth anthesis	Average Capitulum Diameter	%	Timing of Shattering	Flowering Synchronicity	Lodging Score
Genotype	56.4	58.7	66.0		27.5	15.5	24.8
Location	3.1	3.5	2.9		7.9	1.2	5.1
Year	1.3	0.8	0.4		3.1	30.3	1.2
L x Y	1.6	0.3	4.0		6.0	1.0	0.6
Rep(L x Y)	3.9	2.1	1.2		2.5	0.8	4.0
G x L x Y	3.6	3.2	2.1		6.9	5.4	4.7
G x Y	5.3	5.3	5.0		9.4	4.4	7.3
G x L	7.9	7.1	4.3		4.8	7.3	7.5
Residual	16.8	18.9	14.2		31.9	34.1	44.7

Table 3.4: Percentage of sums of squares contributed to genotypic and environment effects for combined phenotypic data for six traits collected in Carman and Winnipeg in the years 2013 and 2014 as calculated using method=TYPE 3 in SAS MIXED for *H. maximiliani.*

Source	Days to anthesis	Days to fifth anthesis	Average Capitulum Diameter	%	Timing of Shattering	Flowering Synchronicity	Lodging Score
Genotype	50.0	46.5	53.7		20.8	10.5	18.9
Location	1.1	0.4	1.2		3.4	>0.01	4.3
Year	0.1	4.4	2.0		2.2	24.1	2.6
L x Y	0.5	0.2	2.4		4.6	>0.01	0.5
Rep(L x Y)	1.2	1.3	1.2		2.0	1.1	1.6
G x L x Y	5.2	4.4	3.8		8.0	5.0	6.9
G x Y	11.3	9.4	6.2		9.9	14.0	9.3
G x L	9.3	6.9	4.4		6.7	6.6	6.9
Residual	21.3	27.3	25.1		39.9	38.6	49.0

Table 3.5: Percentage of sums of squares contributed to genotypic and environment effects for combined phenotypic data for six traits collected in Carman and Winnipeg in the years 2013 and 2014 as calculated using method=TYPE 3 in SAS MIXED for *H. nuttallii*.

Source	Days to anthesis	Days to fifth anthesis	Average Capitulum Diameter	%	Timing of Shattering	Flowering Synchronicity	Lodging Score
Genotype	54.8	64.0	42.9		20.2	9.0	19.3
Location	4.2	3.0	1.4		4.0	1.8	3.9
Year	0.4	0.1	0.5		15.1	23.4	3.6
L x Y	0.7	0.2	2.9		0.8	0.1	>0.01
Rep(L x Y)	2.4	2.0	2.6		2.2	0.6	1.2
G x L x Y	5.3	2.7	3.8		6.1	4.2	6.7
G x Y	8.1	5.9	13.5		6.9	9.0	7.7
G x L	7.7	4.0	8.6		4.8	6.1	5.4
Residual	18.5	17.9	23.6		40.0	45.9	52.0

3.4.4 Correlation and principal component regression analysis

3.4.4.1 Correlation and PCR results for *H. giganteus*

Correlation and principal component regression analysis revealed relationships between annual and monthly temperature norms and days to anthesis, lodging score and average capitulum diameter for *H. giganteus* (Table 3.6, Supplemental Table 3.23). Correlations between latitude and days to first (r = -0.58, p = 0.022) and days to fifth anthesis (r = -0.60, p = 0.019) were observed as well as correlations with measurements of average annual and monthly temperature norms (Supplemental Table 3.23). The strongest correlations include a positive relationship between monthly average temperature in April for days to first anthesis (r = 0.70, p = 0.003) and days to fifth anthesis (r = 0.77, p = <0.001). Similarly, average capitulum diameter was also positively correlated with various measurements of average annual and monthly temperature norms, the strongest being average daily temperature in May (r = 0.61, p = 0.015) followed by daily average temperature in April (r = 0.60, p = 0.018). Significant correlations were not observed between environmental variables and lodging score.

Principal component regression (PCR) models were fit and factors extracted for the traits days to first anthesis, average capitulum diameter and lodging score *for H. giganteus* (Table 3.6). As the traits days to first anthesis and days to fifth anthesis were highly correlated (r = 0.90, p = <0.001), PCR models for days to first anthesis are presented herein. The PCR models for days to first anthesis, average capitulum diameter and lodging score explained 94.9%, 80.4% and 99.7% of the variation in the explanatory variables, respectively. The PCR models explained 52.7% of the variation in days to first anthesis, 29.1% of the variation in average capitulum diameter and 60.0% of the variation

in lodging score. For all three traits mean monthly temperatures showed loadings of similar magnitude and explained the most variation between genotypes, indicating temperature norms are a major driver of differentiation in *H. giganteus*. Following temperature norms, forage water demand was also identified as an important variable explaining the variation observed between genotypes, particularly for lodging score. Latitude showed lower magnitude loadings than monthly temperature norms for days to first anthesis and average capitulum diameter, indicating in the same fashion as the correlation analysis, it has less of an influence than monthly temperature norms.

Table 3.6: Factor loadings for retained environmental variables contributing to model effects and percent variation explained by factor for dependent phenotypic variables measured in Carman and Winnipeg in the years 2013 and 2014 as calculated using principal component regression for *H. giganteus.*

	Days to first anthesis				Lodging score					Av. capitulum diameter
Factors	1	2	3	4	1	2	3	4	5	1
% Variation in Y explained	35.9	0.02	8.77	8.08	19.26	2.55	6.32	17.71	15.15	29.12
Variable loadings										
AWHC*	0.031	0.316	-0.272	-0.666						
BIO1	0.227	-0.083	0.090	-0.029						0.234
BIO5	0.198	-0.135	0.230	-0.002						0.205
BIO6	-0.188	0.206	-0.082	0.070						-0.189
BIO7					-0.316	0.592	-0.504	0.519	-0.141	
BIO8	0.226	0.001	0.159	-0.007						0.238
BIO10	0.226	0.001	0.159	-0.007						0.238
BIO11	-0.199	0.224	-0.080	0.130						-0.199
BIO14	-0.124	0.283	0.459	-0.107	-0.294	0.608	0.619	-0.174	0.361	
BIO15	0.174	-0.283	-0.285	-0.028						
BIO17	-0.162	0.341	0.256	-0.033						
BIO19	-0.162	0.341	0.256	-0.033						
FFD0	0.191	0.232	-0.109	0.209						0.207
FFDNEG2	0.173	0.189	-0.065	0.488						0.187
FFF0	0.169	0.284	-0.241	0.149						0.183
FWDEM	0.212	0.062	-0.240	-0.140	0.463	0.116	-0.360	0.000	0.794	0.218
GDDFOR	0.180	0.303	-0.103	0.154	0.396	0.484	-0.233	-0.608	-0.376	0.198
Lat	-0.202	-0.143	0.130	0.327						-0.213
LSF0	0.169	0.284	-0.241	0.149						0.183
TMEAN5	0.221	0.011	0.181	-0.013	0.478	0.144	0.291	0.229	-0.275	0.234
TMEAN6	0.226	0.045	0.111	0.024						0.240
TMEAN7	0.221	-0.003	0.181	0.021						0.233
TMEAN8	0.225	-0.006	0.153	-0.074						0.236
TMEAN9	0.215	0.124	0.071	-0.092						0.230
TMEAN10	0.220	0.011	0.110	0.079						0.232
TMEAN11	-0.225	-0.061	-0.082	0.067						-0.238
TMEAN4	0.216	-0.042	0.188	-0.132	0.462	0.112	0.306	0.528	-0.056	0.227

* Note: Abbreviations for environmental variables are fully described in Table 3.1.

3.4.4.2 Correlation and PCR results for *H. maximiliani*

Correlation and principal component regression results for *H. maximiliani* differ slightly from those to *H. giganteus*, with latitude, temperature norms and precipitation norms explaining differences between genotypes. Similar to *H. giganteus*, days to first anthesis and days to fifth anthesis were highly correlated (r = 0.87, p = <0.001). Significant environmental correlations were observed for days to first and fifth anthesis, average capitulum diameter, timing of shatter and average capitulum diameter in *H. maximiliani* (Supplemental Table 3.24). Days to first and fifth anthesis were correlated with latitude, precipitation accumulation during the forage growing season, and various measurements of monthly temperature averages. The strongest correlations were observed between latitude and days to first anthesis (r = -0.56, p = <0.001), latitude and days to fifth anthesis (r = -0.61, p = <0.001) with all other correlations falling below r= 0.5.

Days to anthesis was the sole trait for which significant factors could be extracted using PCR as per PRESS cross-validation results (Table 3.7). Three factors were extracted, explaining 89.44% of the variation in the explanatory variables and 38.14% of the variation in days to anthesis. The most important factor explaining the variation between genotypes for days to anthesis in *H. maximiliani* was latitude, followed by the annual mean temperature range which dominate the first extracted factor. Precipitation norms were also identified as important variables but contribute to factors that explained considerably less variation in days to first anthesis than the first factor. The first factor explained 25.17% of the variation in days to first anthesis was positively loaded by latitude (0.611), followed by annual temperature range (0.556). The second factor

explaining 6.00% of the variations in days to first anthesis was negatively loaded by mean precipitation during the forage growing season (-0.753) and loaded by average daily temperatures in July. The third factor explaining 6.97% of the variation in days to first anthesis was positive loaded by precipitation in the coldest quarter. A general pattern of temperature norms as being leading factors influencing phenotypic differentiation in *H. maximiliani* was observed in a similar fashion to *H. giganteus.*

Table 3.7: Factor loadings for retained environmental variables contributing to model effects and percent variation explained by factor for dependent phenotypic variables measured in Carman and Winnipeg in the years 2013 and 2014 as calculated using principal component regression for *H. maximiliani*.

	Days to anthesis		
Factors	1	2	3
% Variation explained in Y	25.17	6.00	6.97
Variable loadings			
Lat*	0.611	0.133	-0.041
FRAIN	-0.163	-0.753	0.430
BIO7	0.556	-0.055	-0.199
BIO19	0.451	0.036	0.753
TMEAN7	-0.295	0.641	0.455

* Note: Abbreviations for environmental variables are fully described in Table 3.1

3.4.4.3 Correlation and PCR results for *H. nuttallii*

Multiple significant environmental correlations were observed for days to first and fifth anthesis, lodging score, timing of shatter and average capitulum diameter for *H. nuttallii* Supplemental Table 3.25). Similar to the other species, days to first anthesis and days to fifth anthesis were highly correlated (r = 0.94, p = <0.001) and all other phenotypic correlations exhibited a weak relationship, falling below r = 0.5. Days to first and fifth anthesis were correlated most strongly with precipitation in the warmest and

wettest quarters with both variables exhibiting $r = 0.73$, $p = <0.001$ for days to first anthesis and $r = 0.79$, $p = <0.001$ for days to fifth anthesis, respectively. Average capitulum diameter was most strongly correlated with frost free days above -2°C ($r = 0.52$, $p = 0.013$). Lodging score was correlated with the average daily minimum temperature of the coldest month ($r = -0.61$, $p = 0.002$), while timing of shatter was most strongly correlated with latitude ($r = -0.83$, $p = <0.001$) and precipitation in the warmest and wettest quarters (both $r = 0.82$, $p = <0.001$).

Significant PCR models were fit for days to first anthesis, timing of shattering and lodging score for *H. nuttallii* (Table 3.8). Variables describing mean monthly temperatures or temperature seasonality explained the most variation in days to first anthesis, lodging score and timing of anthesis. Three factors were extracted for days to anthesis explaining 92.18% of the variation in the explanatory variables and 58.74% of the variation in days to first anthesis. The first factor explaining 30.76% of the variation in days to first anthesis was loaded positively by average daily temperatures during the months of June, July, August and September and negatively loaded by average daily temperatures for the month of November. The second factor, explaining 13.35% of the variation in timing of anthesis was also loaded by temperature variables, namely the average annual temperature range (0.297). The third factor extracted, explaining 14.62% of the variation in days to first anthesis was dominated by factors describing the length of the growing season and include the timing of last spring frost below -2°C (0.455) and first fall frost below -2°C (0.455).

Two factors were extracted for lodging score explaining 89.47% of the variation in the environmental variables and 38.14% of the variation in lodging score. Both factors

were loaded primarily by mean average temperature variables, with the first factor positively loaded by average daily temperatures for the months of June, July, August and September and negatively loaded by average daily temperatures for the month of November. The second factor was also loaded primarily by temperature variables, namely mean annual temperature range (0.302), followed by temperature seasonality (0.298).

Three factors were extracted for timing of shattering, explaining 95.94% of the variation in environmental variables and 73.32% of the variation in timing of shatter. The first factor explained 66.69% of the variation and was dominated by average monthly temperature variables, being negatively loaded by average daily temperatures for the month of November (-0.218) and positively loaded by average daily temperatures for the months of September (0.217) and July (0.216). The second and third factors explaining 2.47% and 4.15% of the variation, respectively, in timing of shatter were positively loaded by elevation and the average number of frost free days above 0°C respectively. A similar pattern of temperature norms being a common latent factor influencing phenotypic differentiation was noted in *H. nuttallii* in a similar fashion to *H. giganteus* and *H. maximiliani* suggesting these species undergo common selection pressures which coincide with the limitations of the growing season in southern Manitoba.

Table 3.8: Factor loadings for retained environmental variables contributing to model effects and percent variation explained by factor for dependent phenotypic variables measured in Carman and Winnipeg in the years 2013 and 2014 as calculated using principal component regression for *H. nuttallii*.

	Days to anthesis			Lodging score		Timing of shatter		
Factors	1	2	2	1	2	1	2	3
% Variation explained in Y	30.76	13.35	14.62	17.17	20.97	66.69	2.47	4.15
Variable loadings								
Lat	-0.196	-0.087	0.021	-0.210	-0.080	-0.212	0.102	0.048
Long	-0.159	-0.186	0.111	-0.175	-0.181	-0.185	0.255	0.256
Elevation	-0.140	-0.228	-0.039	-0.151	-0.226	-0.166	0.346	0.033
AWHC	0.114	0.217	0.126	0.120	0.217			
LSF0	0.160	0.119	0.222	0.166	0.116	0.175	-0.200	0.409
LSFNEG2	0.063	-0.012	0.455					
FFF0	0.160	0.119	0.222	0.166	0.116	0.175	-0.200	0.409
FFFNEG2	0.063	-0.012	0.455					
FFD0	0.147	0.037	0.332			0.150	-0.097	0.600
FFDNEG2	0.057	-0.079	0.440					
GDDFOR	0.192	0.022	0.146			0.199	-0.026	0.275
FRAIN	0.198	-0.043	-0.015	0.211	-0.051	0.202	0.089	0.049
FWDEM	0.185	-0.042	-0.066			0.189	0.120	-0.057
BIO1	0.190	-0.034	-0.069	0.206	-0.043	0.196	0.202	-0.047
BIO2	-0.060	0.260	0.097	-0.061	0.264			
BIO3	-0.070	0.252	0.009	-0.070	0.255			
BIO4	-0.033	0.295	-0.026	-0.028	0.298			
BIO5	0.194	0.055	-0.001	0.209	0.047	0.208	0.040	0.011
BIO6	-0.148	0.216	0.062	-0.156	0.225			
BIO7	-0.037	0.297	0.072	-0.037	0.302			
BIO8	0.206	-0.014	-0.060	0.221	-0.023	0.214	0.130	-0.034
BIO9	-0.159	0.089	0.135					
BIO10	0.206	-0.014	-0.060	0.221	-0.023	0.214	0.130	-0.034
BIO11	-0.191	0.124	0.047	-0.203	0.133	-0.187	-0.280	-0.013
BIO12	0.120	0.247	-0.080	0.134	0.244	0.150	-0.382	-0.236
BIO13	0.167	0.164	-0.054	0.182	0.159	0.190	-0.258	-0.147
BIO14	-0.053	0.256	-0.087	-0.048	0.259			
BIO15	0.124	-0.238	0.042	0.126	-0.244			
BIO16	0.160	0.177	-0.067	0.175	0.172	0.185	-0.280	-0.148
BIO17	-0.082	0.280	-0.072	-0.080	0.285			
BIO18	0.160	0.177	-0.067	0.175	0.172	0.185	-0.280	-0.148
BIO19	-0.082	0.280	-0.072	-0.080	0.285			
TMEAN4	0.202	-0.043	-0.050	0.217	-0.052	0.208	0.182	-0.015
TMEAN5	0.198	-0.063	-0.057	0.213	-0.073	0.201	0.215	-0.013
TMEAN6	0.205	-0.026	-0.049	0.220	-0.035	0.212	0.152	-0.013
TMEAN7	0.205	0.007	-0.054	0.221	-0.001	0.216	0.096	-0.038
TMEAN8	0.204	-0.019	-0.085	0.220	-0.028	0.213	0.135	-0.071

	Days to anthesis			Lodging score		Timing of shatter		
Factors	1	2	2	1	2	1	2	3
TMEAN9	0.206	0.006	-0.072	0.223	-0.002	0.217	0.095	-0.073
TMEAN10	0.200	-0.025	-0.097	0.216	-0.034	0.207	0.164	-0.093
TMEAN11	-0.206	-0.023	0.073	-0.222	-0.015	-0.218	-0.051	0.092

* Note: Abbreviations for environmental variables are fully described in Table 3.1

3.5 Discussion

3.5.1 Phenotypic variation in local germplasm

The significant effect of genotype indicates that there is suitable genetic variation in local germplasm to make selections on the traits days to anthesis, average capitulum diameter, timing of shatter and lodging score in *H. giganteus*, *H. maximiliani* and *H. nuttallii*. The presence of two peaks amongst collection sites observed for days to first anthesis and average capitulum diameter could be indicative of diversifying or disruptive selection for these traits across an environmental gradient (Mather, 1955; Lande and Arnold, 1983), with bimodal phenotypic classes exhibiting higher fitness over intermediate phenotypes. Presence of a major gene exhibiting complete dominance influencing either trait is also possible. As genotypes were sampled across the landscape as they were observed and their demographic histories are unknown, gaps in the observed phenotype distribution could be a reflection of sampling bias in the wild collected materials. Multiple peaks were observed in the distribution of means for average capitulum diameter in *H. nuttallii* indicating the possibility of 3-5 phenotypic classes (Supplemental Figure 3.4).

Days to first and fifth anthesis and average capitulum diameter were influenced primarily by the effects of genotype as opposed to environment, indicating a strong genetic component to their expression. Previous studies have established traits such as that branching architecture and capitulum size respond readily to selection pressures in *H. maximiliani* (Van Tassel et al. 2014). The strong genetic component of these traits and considerable range amongst genotypes for all species indicates that flowering time and capitulum size will likely respond to directional selection readily. The strong effect of genotype may have contributed to the low genotype x environment interaction due to the relative similarity of environmental variables at test locations and amongst years resulting in similar genotypic performance. Under these conditions genotypes are expected to perform consistently and further environments are required to determine the stability of these traits on a larger scale (Annicchiarico, 2002).

Timing of shatter appears to be continuous and commences relatively quickly in all species, ranging from approximately two to five weeks following anthesis. The rapid onset of shattering is not outside the range of other wild *Helianthus* species (Gutierrez et al., 2010) and is a reflection of the undomesticated status of perennial sunflower germplasm. Shattering in *Helianthus* is a genetically complex trait, involving factors such as capitulum size and shape (Burke et al., 2002). In wild sunflower, continued growth of the capitulum throughout seed maturation results in a convex shape increasing the depth: width ratio and encouraging seed dispersal. Domesticated sunflower exhibits a flat capitulum which retains achenes at maturity (Burke et al., 2002). As flowering is indeterminate in *H. giganteus*, *H. maximiliani* and *H. nuttallii*, increasing the interval between anthesis and timing of shattering could enhance harvest uniformity and yield.

Complete loss of mature seeds appears to be a gradual process influenced by capitulum morphology and most likely by environmental factors such as daily temperatures and precipitation. It was observed that shattering was less apparent following rain events and a tendency for later shattering was observed towards the end of the growing season when average daily temperatures lowered. A high proportion of the variation observed for this trait was explained by residual effects, which may represent weather events not accounted for by site and year effects, such as changes in precipitation patterns throughout the growing season. Precipitation patterns influence soil and plant moisture status, which in turn may influence capitulum dry-down and the accuracy of phenotyping for this trait. Genotypic effects contributed the most to the observed variation following residual effects, indicating an underlying genetic component to this trait, which may be useful for crop development. Further characterization is needed for this trait to determine if underlying genetic effects are the result of differences in initiation of flowering or other characteristics such as capitulum morphology, and what impact the initiation of shatter has on yield losses.

Primarily residual effects, followed by genotype for all species, explained lodging score. Lodging score showed weak, but significantly negative correlations to flowering synchronicity and average capitulum diameter, and is likely related to plant size. In sunflower, factors influencing lodging include plant area loaded by wind gusts, stem strength, and size of the root plate (Sposaro et al., 2010). Lodging was observed to be greater in taller plants with a high degree of branching, which may be a function of greater exposure to wind. The use of appropriate genotypes and agronomic practices can alleviate losses due to lodging. Reducing plant height through selection has been one

strategy to improve lodging resistance in the perennial sunflower *H. tuberosus* (Kays and Nottingham, 2007) and may be useful in other perennial sunflower species to improve harvest manageability. In experimental plots, Maximilian sunflower shows considerable phenotypic plasticity and under higher planting densities *H. maximiliani* has been observed to exhibit taller stature as plant densities increase (Kois, 1985). *H. maximiliani* plants surveyed in their natural setting of high competition, low nutrient status showed greatly reduced total plant biomass, total number of stems, plant height, total branches, capitulum number, and total branches $stem^{-1}$ relative to plants grown under cultivated settings (data not shown). Selection for smaller, more compact plants and growing plants at appropriate plant densities could reduce loading forces of wind on a per stem basis, reduce lodging, and facilitate mechanical harvest in perennial sunflower species.

Branching pattern did not differ amongst genotypes for all species with plants exhibiting a primarily apical branching pattern amongst all species in the collection. The apparent lack of basal branching may be indicative of effects of density and light competition between stems, suppressing basal branching under nursery conditions in which a large number of stems emerge. Branching in perennial sunflowers is strongly plastic, with a great deal of variation depending on factors such as planting density, growing conditions, photoperiod and availability of resources (Kois, 1985; Kays and Nottingham, 2007). A large range in total branch number was observed, but distinct differences in general branching classes were not apparent in the local germplasm collection. In wild populations of annual sunflower, branching is believed to be controlled by numerous loci exhibiting small effects, while few loci of major effect control apical and basal branching in domesticated annual sunflower (Nambeesan et al.,

2015). Genetic variation has been observed in branching pattern for *H. maximiliani.* Unbranched plants, which exhibit a single central capitulum akin to the domesticated *H. annuus*, have been developed at The Land Institute (Van Tassel et al., 2014). This trait is likely to be recessive and controlled by several genes in *H. maximiliani*, with the emergence of unbranched phenotypes requiring several generations of inbreeding before the phenotype emerged in outbred breeding populations (Van Tassel, pers. comm.). If this trait is present in Manitoban germplasm it may be masked due to its likely recessive nature and the high degree of heterozygosity expected in an outcrossing species, and may require several generations of inbreeding to recover. Introducing variation in branching may be accomplished through crossing Manitoban germplasm with unbranched TLI materials and adapting these materials to northern latitudes.

Year effects primarily influenced flowering synchronicity, which may be due to differences in climatic conditions amongst years or changes in stem number or rhizomatous spread as plants became more established. Flowering synchronicity was measured on the basis of timing between the first and fifth capitula to reach anthesis regardless of stem, therefore differences in stem number, emergence and growth may influence this trait. Genotype composed a relatively small, non-significant proportion of the observed variation for flowering synchrony. Resource limitations have been shown to greatly reduce rhizome development and subsequent stem emergence in perennial sunflower species such as *H. tuberosus* (Kays and Nottingham, 2008). As plants were established throughout the summer of 2012 they did not have an entire growing season (spring through fall) to develop rhizomes prior to evaluation in 2013 as they did in 2014, therefore differences in flowering synchronicity amongst years may be in part due to the

timing of plant establishment and the development of rhizome resources and resulting stem emergence.

A great deal of variability was noted in total stem number, stem diameter and degree of rhizomatous spread amongst accessions. A large range in total stem number, ranging from two to three stems to upwards of over 100 stems, were observed within the nurseries, with stem diameter, number of capitula produced and total branches showing a similar degree of variation. These observations indicate genetic variation underlying these traits, which may be useful in optimizing plant density of perennial sunflower stands for agronomic production.

Analysis of seed and biomass yield and their components has yet to be conducted fully in these species and requires further investigation under field level agronomic conditions in which producers would be likely to sow and harvest perennial sunflower. Sufficient genetic variation within locally available sunflower germplasm appears to exist to allow for selection on traits that contribute to seed yield in perennial sunflowers, such as timing of transition from vegetative to reproductive growth, capitulum size, degree of plant lodging, as well as timing of shattering.

3.5.2 Genotype x environment interactions

The relatively low (>10%) contribution of G x L interactions observed for all traits indicate that, while there are significant differences amongst genotypes based upon their collection site, the materials appear to be broadly adapted to the common garden testing sites in Carman and Winnipeg and performed predictably across both locations. The high contribution of the effect of genotype for days to first and fifth anthesis and

average capitulum diameter indicates a strong underlying genetic component to this variation.

Mean monthly temperatures were identified as factors influencing timing of anthesis amongst all species. This is indicative of a potential G x L interaction resulting in selection for earlier flowering in colder growing regions. Mean monthly temperatures have been identified as important factors associated with plant size in this study such as lodging in *H. giganteus* and *H. nuttallii*, a trait often associated with plant height, and capitulum size in *H. giganteus.* While not observed in the current study, similar patterns have been observed in studies of *H. maximiliani* (Tetreault et al., 2016; Chapter 4). Based upon an initial examination, days to anthesis and capitulum size are relatively stable, while traits such as lodging score and timing of shatter are more complex and further characterization is needed to improve the accuracy in phenotyping these traits to minimize residual effects. Improvements in phenotyping methodology will likely decrease the proportion of variance consumed by residual effects and determine the contributions of genetic and environmental components of phenotypic variation with greater accuracy in future studies.

3.5.3 Adaptation of *Helianthus* to local environmental clines

Latitudinal clines play an important role in population differentiation in many plant species. Differentiation based upon timing of vegetative and reproductive growth is considered an adaptive response to abiotic stress imposed by factors such as frost or heat (Engelmann and Purugganan, 2006; Kawakami et al., 2011; Tetreault et al., 2016). Clinal variation has been observed in response to latitude in *H. maximiliani* in which Northern

and Southern ecotypes show phenotypic and genotypic divergence across a latitudinal gradient spanning 2500 km of North America (Kawakami et al., 2011; Tetreault et al., 2016). Higher latitudes are often associated with earlier flowering times and lower overall plant biomass as trade-offs are observed between biomass accumulation and timing of reproduction (Egli, 2011). As latitude encompasses daylength and therefore photoperiod, it is often highly correlated with latitudinal temperature gradients effects can be difficult to separate. Latitude and average monthly heat unit accumulation were both identified as important factors influencing days to first anthesis amongst all three species in this study and lodging score in *H. giganteus* and *H. nuttallii*, average capitulum diameter in *H. giganteus* and timing of shattering in *H. nuttallii*. These patterns are in line with the observed differences in cold tolerance in *H. maximiliani* across a latitudinal transect of North America, and suggests adaptation to shorter, cooler growing conditions (Tetreault et al., 2016). Principal component regression did not strongly separate latitude and average monthly temperatures as separate factors, with latitude showing loadings of similar magnitude to average temperature norms along the first principal component for all species. Latitude as an independent factor was not observed amongst lower order components, indicating that temperature norms, and not daylength, may be the major environmental factor driving differences in days to anthesis observed within all three species for the sampled geographic area. Given the geography of southern Manitoba, heat unit distribution is not entirely equal across latitudes, therefore latitude as a separate factor would be expected if it were an independent factor contributing to the phenotypic variation (Nadler 2007, Nadler and Bullock, 2011).

In *Helianthus*, flowering has a strong genetic basis and influenced by both photoperiod and temperature cues (Leon et al., 2001), with major genes known to underlie these traits (Blackman et al., 2010, 2011, 2013, Henry et al., 2014). As latitude/day length were not identified as separate factors in Manitoba populations, major genes controlling photoperiod or temperature response may be fixed within the examined populations and influenced by minor genes giving a more continuous distribution of days to first anthesis.

3.5.4 Timing of anthesis and selection for locally adapted materials

The initiation of reproduction and timing of anthesis is an important phenological event in plant growth and development as it signals the transition from vegetative to reproductive growth. Latitudinal trends in plant size are noted in *H. maximiliani* (Kawakami et al., 2011; Chapter 4) with total plant biomass and timing of first anthesis decreasing along an increasing latitudinal gradient. Kawakami (2011) noted greater production of capitula per unit of aboveground biomass with increasing latitude in this species. This could be indicative of a greater emphasis on sexual as opposed to asexual reproduction in Northern populations, and potentially a greater harvest index if these patterns translate to total seed yield. The trend of later flowering in regions with higher average monthly temperatures suggests plants from colder growing regions have experienced selection for earlier flowering, while plants from warmer regions flower later in the growing season.

The relationship between average monthly temperatures and timing of anthesis may indicate selection, either for avoidance of abiotic stress (such as frost or heat) during

reproduction, or growth limitations imposed by the environment (resource limitations). In *Helianthus*, the stem operates as a temporary sink organ for carbon, which is eventually repartitioned synchronously into seed and perenniating structures (Sadras et al., 1993; Kays and Nottingham, 2007). Selection for plants that enter reproduction at an earlier developmental stage, requiring less biomass to reach reproductive maturity, may be adaptive if heat units are a limiting environmental factor supporting the growth and development necessary for seed and rhizome production (overwintering potential). Genotypes from colder growing locations appear to perform consistently in the warmer testing locations of Carman and Winnipeg for days to anthesis and average capitulum diameter. They do not appear to be limited by the warmer growing conditions and appear broadly adapted. Given the current information it appears that average monthly temperatures are the primary limiting environmental factor influencing these species. Therefore, genotypes from warmer growing regions may be limited in colder regions due to local adaptation. This pattern has been observed in the acclimatization of multiple crops to the Canadian prairies, such as winter wheat (Fowler, 2012) and the candidate perennial grain crops intermediate wheatgrass and perennial wheat (Hayes et al., 2018) in which concentrated efforts to improve adaptation have been undertaken. As both testing locations were located in relatively warm growing regions of Manitoba, further research is needed to determine if genotypes from warmer regions are adequately adapted to colder growing conditions. Reciprocal transplants of extreme phenotypes into their contrasting environments and an examination of dry matter partitioning and yield components would be needed to determine if clinal differences affect overall plant performance, persistence, and survival in a cultivated setting.

3.6 Conclusions

The results of this study indicates that there is suitable genetic variation within local germplasm to make advancements under selection in native perennial *Helianthus* for agronomic characters such as timing of anthesis, capitulum size, timing of shatter and to a lesser degree, lodging resistance. Days to anthesis and average capitulum diameter appear to be influenced primarily by genetic effects and should respond to selection. Based upon testing locations in Winnipeg and Carman, G x L interactions do not appear to influence genotype performance to a large degree, indicating that available germplasm is broadly adapted to these growing regions. This result is not entirely unexpected given the relative proximity of the testing sites and limited test locations. In all species, native plant collections from regions with lower average monthly temperatures tended to exhibit an earlier initiation of anthesis. The presence of this cline suggests the possibility of developing ecotypes based on this environmental factor to best suit agricultural production and the strengths and limitations of different growing regions. Average monthly temperatures may be a limiting factor influencing timing of anthesis and possibly related traits such as biomass accumulation. Further testing is needed to determine how genotypes from warmer regions perform in colder average monthly temperatures and if these differences contribute to differences in agronomic performance. Determining how manipulating timing of anthesis impacts plant biomass and dry matter partitioning to seed production and rhizome development will give insights into how these characters can be used to optimize perennial sunflowers for cultivation. The results of this study suggest that the timing of the transition from vegetative to reproductive growth varies at a local scale and may be manipulated through selection to give further

insights into the development of an appropriate crop ideotype for perennial sunflower under cultivated settings. An earlier flowering ideotype may decrease biomass accumulation through a reduced vegetative period and provide a longer reproductive period, which in turn may increase the number of capitula produced per stem, two traits which in conjunction with a necessary level shattering resistance, may ultimately increase harvest index. Given these lines of evidence, the development of a perennial sunflower crop to extend the growing season in western Canada appears favourable through exploiting the existing standing genetic variation within adapted *Helianthus* germplasm sources.

CHAPTER 4.0: Genetic Characterization of Maximilian Sunflower for the Development of a Locally Adapted Perennial Grain Oilseed

Sean R. Asselin[A], Anita L. Brûlé-Babel[A] and Douglas J. Cattani[A]

[A]Department of Plant Science, University of Manitoba, 222 Agriculture Building, Winnipeg, MB, R3T 2N2, Canada.

Article has been published in *Crop Science*, DOI: 10.2135/cropsci2017.10.0611

In this research chapter Sean R. Asselin contributed to the phenotypic, genotypic and environmental data collection, data analysis, writing and editing of the manuscript.

4.1 Abstract

Maximilian sunflower (*Helianthus maximiliani* Schrad.), a wild relative of domesticated sunflower (*Helianthus annuus* L.), has been identified as a species of interest for the development of a perennial oilseed crop. Knowledge of the diversity, the potential for crop development, and genomic resources of this crop wild relative is limited. To facilitate its use in breeding programs, a baseline characterization of locally adapted germplasm is required. Individuals were collected from nine sites in southern Manitoba, Canada, and characterized for phenotypic and genotypic divergence to estimate traits of interest for the implementation of a breeding program in Maximilian sunflower. Genotype-by-sequencing was used to characterize population genetic parameters and identify candidate single nucleotide polymorphisms (SNPs) associated with phenotypic divergence and environmental differences amongst collection sites. Candidate SNPs associated with frost-free period, temperature during the primary vegetative growth period, elevation, soil $CaCO_3$ equivalent, days to anthesis and capitulum size were identified and may be useful for the improvement of *H. maximiliani* and crop species related to cultivated sunflower. Associations between temperature, population structure, and overall plant size were also identified, suggesting phenotypic divergence across a local temperature gradient. The sampled Maximilian sunflower populations exhibited a high degree of polymorphism, low levels of inbreeding, and a highly heterozygous genome at the local scale. These traits favour the establishment of locally adapted germplasm pools. There is sufficient variation to make selections for agronomic traits in local germplasm of Maximilian sunflower to support its development as a perennial oilseed.

4.2 Introduction

Perennial grains have the potential to introduce new ecosystem services to landscapes dominated by annual cropping systems and are a useful tool to improve sustainability (DeHaan et al., 2005; Glover et al., 2010b; Pimentel et al., 2012). Increasing the functional diversity and composition of ecosystems affects their function (Tilman et al., 1997) and supports ecosystem services (Zavaleta et al., 2010; Isbell et al., 2011; Asbjornsen et al., 2014). Perennial species that are not heavily reliant on tillage have been shown to improve water quality, reduce soil erosion, and nutrient leaching (Randall and Mulla, 2001; Crews, 2005; Culman et al., 2013), and are less reliant on external inputs such as fertilizers (Jenkinson et al., 1994, 2004; Glover et al., 2010a; Crews and Brookes, 2014). Perennial species can access resources such as heat units, light, and moisture available prior to seeding and after harvest of most annual species (Jaikumar et al., 2016), similar to cold-hardy winter annual species (Kell, 2011; Jaikumar et al., 2013). Maximilian sunflower (*Helianthus maximiliani* Schrad.), a perennial crop wild relative of cultivated sunflower (*Helianthus annuus* L.), has been suggested as a candidate species for perennial oilseed crop development (Van Tassel et al., 2014). Crop wild relatives are an important resource for improvement of crop species and serve as a resource for crop diversification through novel genetic variation. Diversity present in many crop wild relatives has the potential to extend the adaptive range of current crops by serving as a source of genetic variation for resistance to biotic and abiotic stress. Climatic variability, the loss of/or marginalization of agricultural land, and growing global demands for food, fiber, and fuel necessitate the use of wild relatives for genetic improvement of existing crops and exploration of the potential to develop new crops that

provide different ecosystem services to the environment. In addition, there has been recent interest in several perennial wild relatives of grain and oilseed crops for development of new crop types such as perennial grains (Cox et al., 2002, 2006; DeHaan et al., 2016).

Maximilian sunflower is a common species widely distributed across the North American Prairie (USDA-NRCS, 2017). As a crop wild relative in the tertiary gene pool of cultivated sunflower (Kantar et al., 2015), *H. maximiliani* is of interest to plant breeders as a source of novel traits for crop improvement, particularly for resistance to the major economically damaging pathogens Sclerotinia rot [*Sclerotinia sclerotiorum* (Lib.) de Bary.] (Taski-Ajdukovic et al., 2006; Liu et al., 2011) and leaf rust (*Puccinia helianthi* Schwein.) (Zimmer and Rehder, 1976), among other diseases (Seiler et al., 2017), and as a source of novel cytoplasm and restorer genes in cytoplasmic male sterility systems (Whelan, 1980; Jan and Zhang, 1994; Feng and Jan, 2008). Like many crop wild relatives of sunflower, *H. maximiliani* is a relatively untapped resource, and genetic characterization will further its utility in plant breeding programs.

As one of the most widely distributed perennial species in the genus *Helianthus*, *H. maximiliani* exhibits regional variants, with distinct northern and southern ecotypes (Heiser et al., 1969). Differences in abiotic stress tolerance and life history traits have been reported across its range, which stretches from the Canadian Prairies through much of the continental United States to northern Mexico (Kawakami et al., 2011, 2014; Tetreault et al., 2016). The wide distribution of *H. maximiliani* disease resistance traits, along with abundant seed production and a fatty acid profile suitable for human and animal consumption, makes the species a candidate for development as a perennial

oilseed crop (Dorrell and Whelan, 1978; Seiler, 1994; Seiler and Brothers, 1999; Cox et al., 2002; Van Tassel et al., 2014).

Helianthus maximiliani, like many crop wild relatives, is largely uncharacterized as a resource for crop development and improvement that has limited its potential use. The species has been identified as a high-priority species in need of urgent collection and characterization as a crop genetic resource, in part due to its tolerance for low and erratic temperature fluctuations (Kantar et al., 2015). Next-generation sequencing approaches are removing barriers to the genetic characterization of crop wild relatives, allowing for greater integration into crop breeding programs (Kantar et al., 2015; Dempewolf et al., 2017; Seiler et al., 2017). The extensive range of Maximilian sunflower suggests the species is capable of adapting to a wide range of growing conditions and likely harbors useful traits as a genetic resource. Local adaptation is critical for the introduction of new crop cultivars into habitats that may differ in abiotic factors such as daylength, elevation, heat unit and moisture availability, frost-free period, soil type, nutrient status, or biotic stresses such as pathogens and other pest populations (Allard and Bradshaw, 1964; Fowler, 2012). The decreasing cost of genotyping has enabled increased evaluation and use of crop wild relatives (Pyhäjärvi et al., 2013; Fang et al., 2014; Anderson et al., 2016), and the development of new crops such as perennial grains (Zhang et al., 2016).

Landscape genomics approaches attempt to bridge patterns of genetic variation with environmental features that underlie local adaptation (Joost et al., 2013). Incorporating knowledge from landscape genomics is a potential tool for identifying traits associated with local adaptation in plant breeding programs. Recent studies have shown its potential use for crop improvement through the characterization of crop wild

relatives (Fang et al., 2014; Anderson et al., 2016; Dempewolf et al., 2017). Common landscape genomics approaches include population differentiation (PD) and environmental association analysis (EAA) (Rellstab et al., 2015). Population differentiation methods are based on inferred genetic structure of populations and patterns of allelic diversity relative to a genetic model. Single-locus estimates of differentiation are compared with either a null model of neutral evolution or model that incorporates population structure. Population differentiation methods are based on the analysis of outliers from an expected distribution and are well suited for the detection of major loci under strong selection (Narum and Hess, 2011; Leinonen et al., 2013).

Environmental association analysis incorporates environmental data and looks for relationships between allele frequencies and environmental variables in an approach similar to association mapping (De Mita et al., 2013; Rellstab et al., 2015). The EAA approaches have an advantage when looking at clinal distributions of alleles in continuous populations that may not show strong patterns of fixation, but gradual changes in allele frequencies (François et al., 2016). Weak genetic sweeps may be detected through correlations between allele frequencies and environmental variables in these approaches (François et al., 2016; Stephan, 2016). Quantitative traits may show more subtle patterns of differentiation than monogenetic traits due to the influences of many loci and may be overlooked if employing overly conservative PD approaches. The recent release of the cultivated sunflower reference genome has provided new resources for the characterization of crop wild relatives in the genus *Helianthus*, facilitating the acceleration of breeding efforts (Badouin et al., 2017) and application of novel tools such as landscape genomics.

The objective of this study was to apply a landscape genomics approach to characterize local variation and detect candidate single nucleotide polymorphisms (SNPs) underlying traits of interest for breeding efforts in Maximilian sunflower as a perennial grain oilseed and as a resource for the improvement of cultivated sunflower using next-generation sequencing.

4.3 Materials and methods

4.3.1 Collection site characterization and plant sampling

Starting in summer 2012, nine sites within southern Manitoba where *H. maximiliani* had been observed were selected for sample collection following a north–south and east–west transect of southern Manitoba (Table 4.1, Supplemental Figure 4.1). Sites were selected based on the presence of a minimum of 50 to 100 flowering *H. maximiliani* individuals and no known demographic history. For site characterization, four soil cores from each sampling site were collected at the 0- to 15- and 15- to 60-cm depths and were bulked by depth to determine soil properties at the collection sites. After collection, and prior to analyses, all samples were frozen at −20°C. Soil quality analyses for soil nutrient status, soil pH, total organic matter, and soluble salts were conducted by AgVise Laboratories (Northwood, ND) for each collection site at both depths. Environmental data were collected from two sources for each collection site, the first being a summary of available public Environment Canada weather data consisting of average monthly temperature and precipitation norms from the years 1971 to 2000 compiled by Nadler (2007). The nearest weather stations to the collection sites were

selected to represent their respective environmental norms. The second source consisted of data collected from the public global bioclimatic repository WorldClim (Hijmans et al., 2005). Collection site coordinates were used to extract environmental data from WorldClim at a 1 km^2 resolution using the R package *Raster* (Hijmans et al., 2016). A full listing of environmental variables employed for downstream analyses are listed in Supplemental Table 4.1.

Table 4.1: Collection site descriptions for nine Maximilian sunflower field collection sites sampled in 2012 in Southern Manitoba employed in phenotypic evaluation, environmental association, and correlation analyses.

Environmental variable*	Woodlands	Emerson	Portage La Prairie	Brunkild	Oak Bluff	Dunnottar	Birds Hill	St. Pierre Jolys	Brandon
Latitude (°)	50.22	49.03	49.98	49.59	49.77	50.46	50.04	49.53	50.01
Longitude (°)	−97.63	−97.25	−98.23	−97.57	−97.21	−96.95	−96.91	−96.98	−99.94
Elevation (m)	259	238	254	237	229	221	233	239	469
Soil water-holding (AWHC)	246	314	246	314	314	314	314	174	314
Corn heat units (CHU)	2567	2441	2513	2401	2470	2288	2369	2443	2378
Frost-free days above 0°C	126.8	128	130.8	109.5	121.8	125.1	112.2	112.1	109
Soil pH	8.4	8.1	8.1	8.1	8.1	7.7	8	8.3	8
Soil organic matter (%)	3.2	10	7.6	5.3	5.4	8.1	10.9	3	6
Soil N content (kg ha^{-1})	8	29.1	11.2	16.8	10.0	11.2	40.3	12.3	5.6
Soil P content (mg kg^{-1})	2	7	8	3	6	5	5	2	5
Soil K content (mg kg^{-1})	77	270	231	455	353	455	431	180	206
Soil Ca content (mg kg^{-1})	4769	6397	4954	5588	5884	5296	7586	4847	4678
Soil Mg content (mg kg^{-1})	714	983	2204	2040	1586	1535	2976	1127	1673
Soil Zn content (mg kg^{-1})	2.07	1.66	1.57	0.58	1.04	1.37	1.51	0.77	8.82
Soil Cl content (mg kg^{-1})	492	104	2876	64	1092	152	52	144	612
Soil Cu content (mg kg^{-1})	1.06	1.81	3.24	2.52	3.84	1.95	1.75	3.75	3.49
Soil B content (mg kg^{-1})	0.89	2.75	2.98	2.05	1.03	0.87	1.5	0.8	1.83
Soil Fe content (mg kg^{-1})	32.4	18.5	28.5	17.2	29	30.3	20	29.7	53.8
Soil Mn content (mg kg^{-1})	5.26	6.56	8.45	1.88	1.9	3.01	1.25	3.72	8.49
Soil Na content (mg kg^{-1})	116	23	266	119	288	46	73	126	95
Cation exchange capacity (mEq)	30.5	40.97	44.89	46.62	44.79	40.64	64.15	34.64	38.27
$CaCO_3$ equivalent (%)	16.1	5.7	5.2	7.3	5.3	1	3.6	9	6.3
Annual mean temperature (°C)	1.7	2.9	2.6	2.6	2.2	1.4	2	2.4	1.7
Annual mean precipitation (mm)	513	507	470	506	517	532	520	513	469
Avg. monthly temp. in June (°C)	16.3	17.1	16.9	17	16.5	15.8	16.3	16.4	15.8

Note: *Refer to Supplemental Table 4.1 for detailed descriptions of environmental variables

Individually labeled leaf, seed, and rhizome samples were taken at all nine collection sites from 16 individuals were randomly sampled at 10-m intervals to prevent repeated sampling of potential clones. Multiple leaf samples were taken from individual stems, marked with an identity number and stored on ice during transport. Leaf samples were frozen in liquid N within 6 h of collection prior to lyophilization and storage at room temperature. Rhizomes were collected by sampling individual stems and potted in 1-L pots for transport prior to being transplanted to the common garden sites. Capitulum samples were taken from single stems and air dried at room temperature prior to hand threshing to release seed that was then stored at room temperature.

4.3.2 Phenotypic characterization

4.3.2.1 Growth chamber studies

Open-pollinated seed from 16 individuals from each of the nine collection sites were collected in August 2012. Four open-pollinated maternal seed families with adequate sample sizes were selected by site for growth chamber analysis. Seeds were surface sterilized using a 70% ethanol solution for 10 minutes, allowed to air dry, and cold stratified under dark conditions at 4°C for 6 weeks in Petri dishes containing filter paper moistened with distilled water to break seed dormancy. All seedlings were started in Sunshine #4 soilless potting mix (SunGro Horticulture) and transferred to 1-L pots containing a 2:2:1 ratio of soil/sand/peat by volume once they had reached the three- to four-true-leaf stage. Plants were grown in growth chambers following a 23°C 16-h

day/18°C 8-h night cycle to simulate early summer growing conditions in southern Manitoba. Plant positions were assigned randomly and rotated on a weekly basis to account for potential differences in airflow, humidity, and light intensity amongst benches within the growth chamber. Phenotyping was conducted using four runs in a common growth chamber which each collection site being represented in each run. A total of 404 plants from across the nine collection sites were phenotyped for days to anthesis, stem diameter, plant height, total number of capitula, total number of branches, total number of nodes, size of the first capitulum, and average capitula diameter of the first five capitula to reach anthesis. With the exception of days to anthesis and diameter of the first capitulum, which were measured on the day of first anthesis, all plant measurements were taken one week after first anthesis. Due to differences in seed viability amongst the wild sampled capitula, growth chamber runs were bulked for analysis with 20 to 65 plants representing each collection site.

Data were examined for outliers visually using the SAS 9.3 software (SAS Institute, 2011) UNIVARIATE procedure and analyzed using the mixed-model MIXED procedure as described in Littell et al. (2007) to determine if differences observed amongst collection sites were significant. As replicates were grown in more than one growth chamber, growth chamber was included as a blocking term to account for potential confounding effects. Collection site and growth chamber served as fixed effects, whereas collection site x growth chamber served as a random effect with the ddfm = Satterthwaite option invoked to test for the significance of collection site on all metric traits. Homogeneity of residual variances was tested using the Levene's test protocol for mixed models. In instances where variances were suspected to be heterogeneous amongst

sites, the “group” option in the random statement was set to growth chamber as an adjustment, and Akaike’s information criterion (AIC) values were compared to determine the best model fit (Littell et al., 2007). To examine trait relationships Spearman’s ranked correlations were calculated between collection site means using the CORR procedure in SAS 9.3. The best linear unbiased predictions of the traits that exhibited significant differences amongst collection sites were tested for normality using Shapiro–Wilk test (Shapiro and Wilk, 1965) prior to use in association analysis.

4.3.2.2 Common garden studies

Rhizomes from 16 individuals collected in August 2012 from each sampling location were established in common garden field plots in August and September 2012 at sites in Winnipeg and Carman, MB. Plants were organized in a randomized complete block design, and each site contained four blocks with two individuals representing each collection site in each block giving a total of 16 plants to represent each collection site in each test year (2 plants x 2 sites x 4 blocks). Plants were phenotyped in 2013 and 2014 for the Julian date of first anthesis, Julian date of fifth anthesis, days to shatter (based on the Julian date of first shattering observation − first observation date of anthesis), and average capitulum diameter of the first five capitula after anthesis. Due to poor establishment in two of the blocks at the Winnipeg site, only the two full blocks were used in the analyses, whereas all four blocks were analyzed from the Carman site. Data were analyzed following a mixed model approach in SAS 9.3 following the MIXED procedure. Year, block and collection site served as fixed effects. Year was considered

fixed effects in this instance as fewer than 10 levels were present, as per recommendations of Yang (2010), while block, which did not change between years, served as a proxy for test location. Collection site x year, collection site x block, year x block, and collection site x year x block served as random effects with the ddfm = Satterthwaite option invoked.

In instances were heterogeneity of variances was suspected, the "group" option in the random statement was set to year or replicate and AIC values were examined to determine the best model fit. A summary of traits measured under the common garden and growth chamber conditions is presented in Table 4.2.

Table 4.2: List of phenotypic traits assessed under growth chamber conditions and under common garden conditions in 2013 and 2014 in Carman and Winnipeg.

Growth chambers	Common gardens
Plant height (cm)	Average capitulum diameter (mm)
Average capitulum diameter (mm)	Days to anthesis (Julian date)
Capitulum count	Days to fifth anthesis (Julian date)
Days to first anthesis (days)	Timing of Shatter (days)
Diameter of the first capitulum (mm)	
Lowest branching node	
Stem diameter (mm)	
Total branches	
Total nodes	

4.3.3 DNA extraction

Genomic DNA was extracted from wild collected lyophilized leaf tissue for ten of the 16 individuals from each collection site using a modified single-tube cetyltrimethylammonium bromide (CTAB) extraction protocol (Doyle, 1987; Li et al.,

2007) with 1μL of 10 mg mL^{-1} solution of proteinase K added to the initial CTAB incubation step (Promega, Madison, Wisconsin, USA). DNA quantity was determined using a dsDNA broad Range Fluorescence Assay kit and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, California, USA) following the manufacturers' instructions for 1μL sample sizes. DNA quality was assessed using 260/280 and 260/230 nm wavelength absorbance ratios measured using a Nanodrop 1000 spectrophotometer (Fischer Scientific, Waltham, Massachusetts, USA). Samples that fell below a minimum DNA concentration of 50 μg/μL or absorbance ratios below 1.7 for either 260/230 or 260/280 nm were discarded and extractions were repeated to meet quality requirements.

4.3.4 Genotype-by-sequencing and SNP calling

Genomic DNA from 10 samples from each of the 9 collection sites were submitted to Data2bio (Ames, IA) for GBS and SNP calling. A tunable GBS (tGBS) protocol was used, which differs from conventional GBS through the use of a greater selective genome reduction procedure (Ott et al., 2017). Fewer sites are sequenced in tGBS relative to the conventional GBS, resulting in greater read depth, more effective calling of heterozygous genotypes, reduced false homozygote calls, and a lower level of missing data per site.

Sequencing via tGBS was performed using a three selective base (TGT) protocol as described in (Ott et al., 2017) for genome reduction. Fragments were sequenced via eight runs on an Ion Proton sequencer (Thermo Fisher Scientific). The conservative three selective base approach was chosen because *H. maximiliani* is an obligate outcrossing

species expected to exhibit a high degree of heterozygosity. Each read was scanned for low-quality regions and bases with Phred quality scores (Ewing et al., 1998) of <15 out of 40 being removed. Trimmed reads subsequently were aligned to the *H. annuus* reference genome HA412.v.1.1.bronze.20142015 (www.sunflowergenome.org) using GSNAP (Wu and Nacu, 2010), with polymorphisms within the first and last 3 bp for the read being ignored (Ott et al., 2017). Polymorphisms with base calls with a Phred score <20 were removed from the analyses. Single nucleotide polymorphisms were called as homozygous if the most common allele was supported by a minimum of 80% aligned reads, whereas heterozygous SNPs were called if the two most common alleles were supported by a minimum of 30% of all aligned reads and the sum of the two most common alleles accounted for a minimum of 80% of the aligned reads. Single nucleotide polymorphisms that could be genotyped in a minimum of half of the samples and that contained an allele number of 2, a minor allele frequency >1%, and a heterozygous rate of <70% were maintained for downstream analysis.

4.3.5 SNP filtering

Missing data and the degree of linkage disequilibrium between SNPs can result in bias in population structure and association analyses; therefore, prior to analyses linkage disequilibrium values expressed as the coefficient of determination (r^2) were calculated between all SNP pairs and SNPs were filtered into three datasets using PLINK version 1.9 (Purcell et al., 2007). To maximize potential loci for testing, the first dataset (referred to as Filter20) consisted of 5,414 SNPs with no more than 20% missing data and a minor

allele frequency of <5% for outlier and association analysis. The second dataset (Filter20C) consisted of the same SNPs as the Filter20 set, minus 549 SNPs that were aligned to HA0_73NS, an unordered collection of contigs with no assigned linkage group (LG) in the cultivated *H. annuus* bronze reference genome to reduce bias in linkage decay estimates. The third set (Filter10) consisted of 1,009 SNPs with a stricter missing data threshold (10%), a minor allele frequency of <5%, and a maximum r^2 value between markers of 0.2 to account for potential bias in population structure estimates due to linkage between loci.

4.3.6 Population structure analysis

Population structure was explored using two methods, the Bayesian, model-based clustering algorithm implemented in the software package ParallelStructure (Pritchard et al., 2000a; Besnier and Glover, 2013) and principal component analysis (PCA), a model-free approach, implemented in the R package *pcadapt* (Luu et al., 2017). ParallelStructure was run on XSEDE via the CIPRES science gateway (Miller et al., 2010) using the Filter10 dataset. Prior population information (collection site) was incorporated using the LOCPRIOR option to facilitate population assignment and population admixture was assumed. Analysis was run with 20 independent replicates using K = 1 to 10 population genetic clusters and p-values were averaged across runs. The initial burn-in period was set to 500,000 runs followed by 500,000 Markov chain Monte Carlo iterations. Structure analysis was conducted assuming correlated allele frequencies as per the recommendation of Pritchard et al. (2000a) as default. The most

likely number of clusters was determined using the $\hat{\lambda}K$ method of Evanno et al. (2005) and visualized using STRUCTURE HARVESTER (Earl and von Holdt, 2012). A neighbor-joining cladogram was generated in TASSEL using the Filter10 dataset and visualized in the R package *ggtree* (Yu et al., 2017).

Principal component analysis was run on the Filter10 dataset using the R package *pcadapt* on centered and scaled data and compared with results produced in TASSEL. To account for missing data, distances between individuals were calculated only using scored SNPs; missing data between individual pairs were ignored. The optimal value of *K* denoting population structure was determined using a graphical approach by examining the point at which a scree plot of the principal components' (PCs') eigenvalues reached a plateau following Cattell's rule (Cattell, 1966; Jackson, 1993). An analysis of molecular variance (AMOVA) and calculation of the *F* statistics F_{st}, F_{it}, and F_{is} was run on the Filter10 dataset to determine the proportion of allelic variation which could be attributed to differences amongst collection sites, amongst individuals within collection site, and within individuals, respectively. The AMOVA (run using 9999 permutations) and calculation of population genetic parameters, observed heterozygosity, and inbreeding coefficients were calculated by collection site using GenAlEx version 6.5 (Peakall and Smouse, 2012). To determine if environmental variables or phenotype were associated with population structure, Spearman's ranked correlations were run between the population structure estimates using the mean of ancestry coefficients (*Q*) and PCA scores 1 to 3 for each collection site to their environmental characteristics and phenotypic means using the SAS CORR procedure.

4.3.7 Linkage disequilibrium analysis

Linkage disequilibrium between SNPs expressed as the r^2 was calculated using Plink 1.9 (Purcell et al., 2007) using the Filter20C dataset. Linkage disequilibrium was estimated between pairs of SNPs located on the same LG based on the cultivated *H. annuus* LGs. To estimate syntenic linkage decay, SNP–SNP r^2 values were plotted against their genetic distance in base pairs estimated from the *H. annuus* reference genome. The decay of r^2 by distance was fitted using the expectation of r^2 between adjacent sites calculated using the Hill and Weir formula method (Hill and Weir, 1988) using a R script described in Marroni et al. (2011). To determine the average effective distance where linkage decays, the decay curve was compared with a minimum threshold point of $r^2 = 0.2$, a common threshold reported in plant studies (Zhu et al., 2008).

4.3.8 Detection of outlier loci using population differentiation

Combining multiple outlier tests is recommended to reduce the false discovery rate (FDR) when using PD approaches to genome scans (Rellstab et al., 2015). Two F_{st}–outlier tests and a multivariate outlier test using PCA were run independently on the Filter20 dataset to reduce the potential for spurious associations. The programs BayeScan (Foll and Gaggiotti, 2008; Foll, 2012), Lositan (Antao et al., 2008), and the R package *pcadapt* (Luu et al., 2017) were run separately to detect outlier loci. Data were converted to an appropriate file format using PGDSpider 2.1.1.0 (Lischer and Excoffier, 2011) prior to analyses. To further reduce the level of false positives that may arise through multiple testing, a FDR procedure was used to generate *q* values (*q*FDR) for the interpretation of

p-values for all three tests. The SNPs that exhibited a *q*FDR value <0.05 in two of the three tests were considered under putative selection.

BayeScan, which uses a Bayesian approach to estimating F_{st} coefficients, was run with a burn-in of 100,000 iterations with 20 pilot runs each with a length of 5000 iterations followed by running 500,000 iterations with a sampling size of 5000 and a thinning interval of 100 iterations. Prior odds for a neutral model at 1:10, 1:100, and 1:1,000 scales were run with prior odds of 1:10 being retained. In BayeScan, *q*FDR values are generated by default and loci with a *q*FDR value <0.05 were considered non-neutral.

Lositan uses the Beaumont and Nichols (1996) approach to identify outlier loci based on the distribution of heterozygosity and F_{st} using an island model of migration. Population structure estimated using the Filter10 dataset was used to define groups prior to analysis. The "neutral mean F_{st}" and "force mean F_{st}" options were invoked to remove F_{st} outliers prior to calculation of the null F_{st} distribution, and 1 x 10^6 simulations were run on the dataset. Loci exceeding the 95th upper confidence area were considered non-neutral, and those remaining after the application of a FDR threshold of 0.05 in Lositan were considered candidates for selection.

Pcadpadt uses PCA to ascertain population structure; therefore, it is not reliant on predefined groups such as collection site. It is well suited for populations that show continuous population structure, such as populations that have undergone range expansion and exhibit a high degree of admixture (Luu et al., 2017). The *p*-values generated from *pcadapt* were converted to *q*FDR using the R package *qvalue* (Bass et al., 2016), and loci exhibiting a *q*FDR value <0.05 were retained.

4.3.9 Detection of SNP–environment associations

Environmental association analysis of the SNP dataset using an ecological association approach was performed using TASSEL 5.0 (Bradbury et al., 2007) on the Filter20 dataset to maximize the potential number of loci for testing. Five models were run for each environmental and morphological trait: two general linear models (GLMs) and three mixed linear models (MLMs). The GLM models consisted of the environmental or morphological trait in question as the response variable, the individual genotypes (SNP call) as a fixed effect, and either the *Q* matrix (GLM *Q* model) or PCA score (GLM PCA model) as a fixed-effect covariate (Price et al., 2006). The mixed models included kinship (*K*) as a random effect, and genotype either as the sole fixed effect (MLM *K* model) or with the inclusion of the *Q* matrix (MLM *Q* + *K* model) or PCA score (MLM PCA + *K* model) as fixed effects. Kinship, *Q* matrix, and PCA scores used were derived from the Filter10 dataset to reduce the potential bias of missing data on population structure. The best model for each environmental and morphological trait was determined by examining quantile-quantile plots of the observed *p*-values for each SNP plotted against expected values. Coefficient of determination (r^2) values were determined between expected and observed *p*-value distributions to help determine best model fit. The model that showed the closest distribution to the expected values was selected as the most appropriate for each trait (Bradbury et al., 2011). Estimates of allelic effects were calculated in TASSEL using an additive plus dominance gene effects model, with homozygous and heterozygous states being treated as separate classes. Coefficient of determination (r^2) values were generated by default for each marker using each model. To protect against type I errors, a FDR correction (Benjamini and Hochberg, 1995) was

applied using the R program *qvalue* (Bass et al., 2016), and SNPs with *q*FDR values <0.05 were considered candidates for selection.

4.3.10 SNP annotation

To infer candidate genes and provide functional annotation of the candidate SNPs generated from PD and EAA approaches, *H. maximiliani* SNP reads were compared with the cultivated *H. annuus* reference genome HA412.v.1.1.bronze.20142015 assembly using Jbrowse (Skinner et al., 2009), available via the HeliaGene bioinformatics portal (Carrere et al., 2008), to identify putative sunflower functional genes. *Helianthus annuus* gene identity, interpro annotations, and top BLASTx hits to the *Arabidopsis thaliana* (L.) Heynh. TAIR10 protein database were gathered to support putative *H. maximiliani* gene function. The SNPs found in intergenic regions had their flanking regions searched for the presence of the nearest annotated gene within the cultivated *H. annuus* reference genome. The flanking sequences were then compared with the UniProtKB database (UniProt Consortium, 2017) using BLASTx to provide putative protein function using an *E* value cutoff $<10^{-5}$ via the program Blast2GO (www.blast2go.com). The SNPs that were found to be in close proximity (<100 bp) in the *H. annuus* reference genome and in high linkage disequilibrium (>0.99) were considered as a single entry for annotation.

4.4 Results

4.4.1 Genotype-by-sequencing and SNP calling

A total of 24,957 SNPs called by Data2Bio met the minimum quality requirements of being genotyped in at least 50% of the 90 samples and having a genotype number of two or more, an allele number two, a minor allele frequency of ≥1%, and a heterozygous rate of ≤70%. The distributions of SNPs called by LG are listed in Supplemental Table 4.2. The unordered contig HA0_73NS, a composite of all unordered sequences in the cultivated *H. annuus* reference genome, contained the most *H. maximiliani* SNPs. The LG with the greatest number of called SNPs was LG 10 with 2075, 439, and 82 SNPs in the unfiltered, Filter20/Filter20C, and Filter10 datasets, respectively. The fewest numbers of SNPs were called on LGs 6 and 7 (Supplemental Table 4.2).

4.4.2 Population structure analysis

Results of the AMOVA show that the majority of genetic variation present in the samples is attributed to variation within individuals (61%), followed by differences among individuals within collection site (37%) and amongst collection sites (2%) (Table 4.3). Inbreeding coefficients were significant for individuals relative to total population (F_{it}, $p < 0.0001$), individuals relative to subpopulation (F_i, $p < 0.0001$), and subpopulation relative to total population (F_{st}, $p < 0.0001$). Average heterozygosity ranged between

0.191 to 0.227, whereas F_{st} ranged from −0.047 to 0.008 and F_{is} ranged from −0.050 to 0.024 (Table 4.4).

Table 4.3: Results of analysis of molecular variance calculated in GenAlEx for 90 wild diploid samples of Maximilian sunflower plants from nine collection sites in southern Manitoba genotyped using genotype-by-sequencing.

Source	DF	Sum of squares	Mean square	Estimated variability	Proportion of estimated of variance	*F*-value	*P*-value
Among collection sites	8	2,230.67	278.83	3.23	0.02	0.02	< 0.0001***
Among individuals within collection sites	81	17,338.35	214.05	58.90	0.37	0.38	<0.0001***
Within individuals	90	8,662.50	96.25	96.25	0.61	0.39	<0.0001***
Total	179	28,231.52	–	158.39	1.00	–	–

Note: *** significant at alpha 0.0001 probability level.

Table 4.4: Heterozygosity and inbreeding coefficients of nine collection sites sampled in southern Manitoba in 2012 and analyzed using genotype-by-sequencing.

Collection site	H_o†	F_{st}‡	F_{is}§
Woodlands	0.209	−0.025	−0.029
Emerson	0.197	−0.023	−0.006
Portage La Prairie	0.197	−0.001	0.008
Brunkild	0.200	−0.021	−0.010
Oak Bluff	0.191	0.008	0.020
Dunnottar	0.202	−0.034	−0.045
Birds Hill	0.206	0.007	0.024
St. Pierre Jolys	0.213	−0.010	−0.005
Brandon	0.227	−0.047	−0.050
Mean	0.205	−0.016	−0.010

Note:† H_o, observed heterozygosity.

‡ F_{st}, inbreeding coefficient within collection site relative to population.

§ F_{is}, inbreeding coefficient within individuals relative to collection site.

Structure analysis using 1,009 bi-allelic SNPs identified three population groups with varying degrees of admixture based on the $\hat{\lambda}K$ approach (Supplemental Figure 4.2-4.3). Group 1 (G1) consisted of sites Woodlands, Emerson, Portage La Prairie, Brunkild, St. Pierre Jolys, and Oak Bluff. Group 2 (G2) consisted of the Brandon and Birds Hill sites, whereas Group 3 (G3) consists of the Dunnottar site (Figures 4.1-4.2, Supplemental Figure 4.4). Evidence of admixture amongst groups was present in all three groupings, particularly the St. Pierre Jolys site, which had an average *Q* of 0.507 for G1 and 0.480 for G2, indicating nearly equal group membership.

The scree plot of the PCA shows a plateau at a value of $K = 3$ (Supplemental Figure 4.5), indicating that the first two PCs explaining 5.31% of the total variation in the dataset adequately describe the presence of weak population structure and likely genetic admixture amongst groups. Subsequent PCs did not reveal separation patterns as clear as those of the first two components (Figure 4.1). The drop off in explained variance following the first two PCs followed by a plateau in the scree plot (Supplemental Figure 4.5) is indicative of random noise as opposed to genetic structure. The result of the PCA is largely in line with the groups suggested by structure analysis with separation of G1, G2, and G3 groups, with the St. Pierre Jolys site falling between G1 and G2 (Figure 4.1 A). The neighbor-joining tree for the most part supported the results of the PC and STRUCTURE analysis, indicating the presence of the G2 and G3 groups as separate branches composed of the Brandon/Birds Hill and Dunnottar sites, respectively (Figure 4.2). Few individuals from G2 and G3 grouped into the large G1 group, while several individuals from St. Pierre Jolys grouped with G1 and G2, which may be due to the

presence of admixture. The G1 group showed several branches that were not supported as separate groups by PC or STRUCTURE analysis and did not correspond to clear patterns such as collection site of origin.

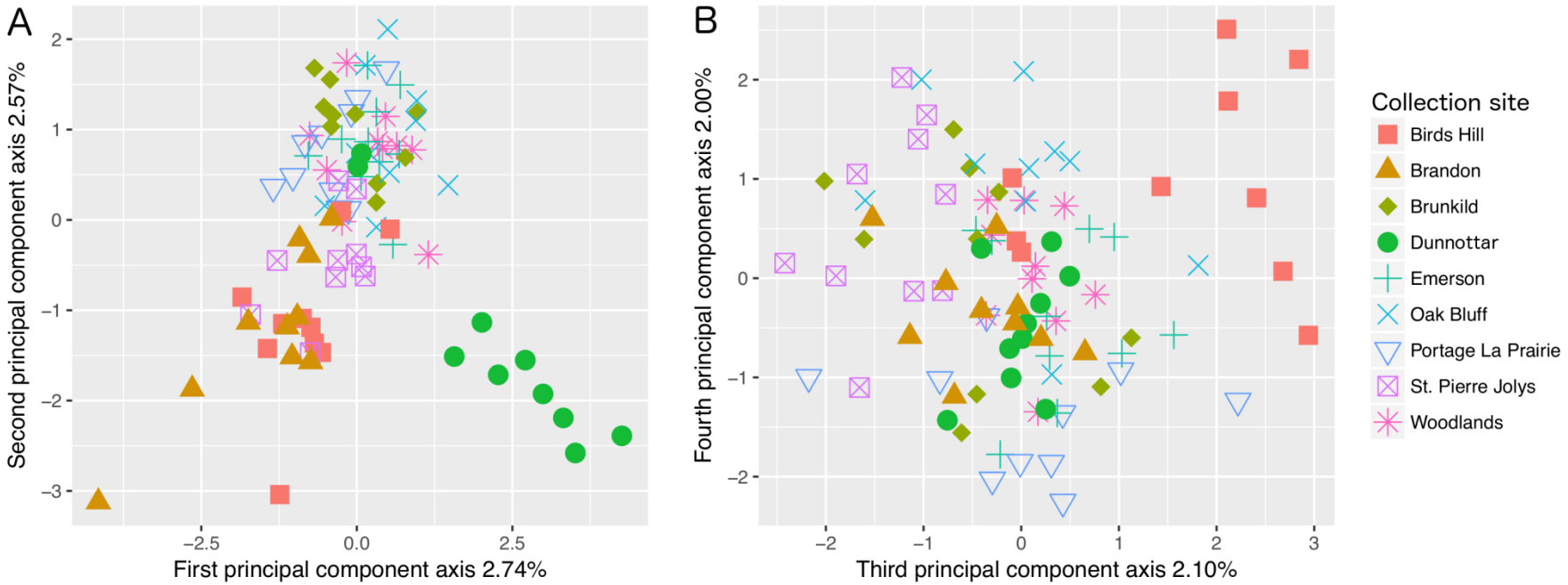


Figure 4.1: Percentage of variance explained and distribution of individuals by collection site along the a) first two principal component axes and b) third and fourth principal component axes as generated in *pcadapt* for 90 *H. maximiliani* individuals collected from Southern Manitoba.

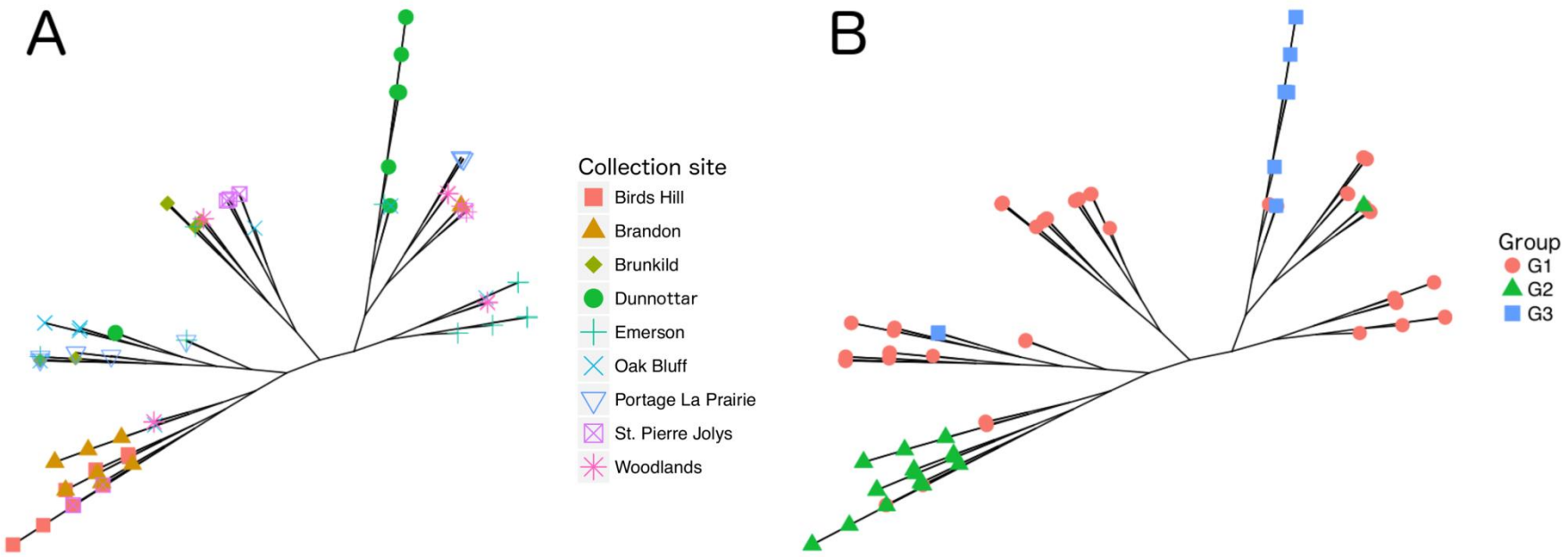


Figure 4.2: Neighbour-joining trees of ten Maximilian sunflower plants from each of the nine Manitoban collections a) generated in TASSEL with collection site identity superimposed and b) generated in TASSEL with suggested ancestry groupings generated by STRUCTURE superimposed. It should be noted that each terminal node represents a single individual.

Correlation analysis revealed that population structure estimates were correlated with environmental variables, indicating a relationship between temperature, precipitation and population structure. Average ancestry coefficient for G1 was correlated with annual temperature ($r = 0.72$, $p = 0.028$) and average daily temperature in June ($r = 0.87$, $p = 0.002$), whereas G3 was correlated with annual precipitation ($r = 0.70$, $p = 0.035$) and soil B content ($r = -0.68$, $p = 0.042$) (Average collection site principal component scores explaining population structure were not correlated to any environmental variable with the exception of the first PC and average daily June temperatures ($r = 0.79$, $p = 0.011$).

In the growth chamber experiment population structure measurements were correlated to phenotypic differences between collection sites. The first PC (PC1) was positively correlated with average capitulum diameter ($r = 0.70$, $p = 0.036$), diameter of the first capitulum ($r = 0.68$, $p = 0.042$), total plant nodes ($r = 0.88$, $p = 0.002$), and plant height ($r = 0.67$, $p = 0.050$). The second PC was not correlated with any measured phenotypic trait. Ancestry coefficient for G1 was correlated with the same phenotypic characters as PC1, average capitulum diameter ($r = 0.70$, $p = 0.036$), first capitulum diameter ($r = 0.67$, $p = 0.050$), total plant nodes ($r = 0.88$, $p = 0.005$) and plant height ($r = 0.70$, $p = 0.036$) in the growth chamber. Ancestry coefficients for G2 and G3 were not correlated with any measured phenotypic traits (Supplemental Tables 4.3-4.4).

Linkage decay was estimated across the 17 *H. maximiliani* LGs using 4865 SNPs in the Filter20C dataset. Linkage decay declined sharply with genetic distance across all sites, falling below $r^2 < 0.2$ at 93 bp and decreasing to $r^2 < 0.1$ at 271 bp (Supplemental

Figure 4.6-4.7). A low proportion of marker pairs were found to be in disequilibrium, with 0.58 and 2.25% of SNP pairs having r^2 values <0.2.

4.4.3 Phenotypic variation

Phenotypic differences were observed amongst collections sites indicative of differentiation across the sampling range. Significant differences in days to first anthesis and days to fifth anthesis were detected between plants from common garden plots, whereas average head diameter and timing of shattering did not show significant differences (Table 4.5& 4.6). Plant height, total nodes, days to first anthesis, diameter of the first capitulum, average capitulum diameter, and stem diameter showed significant differences in the growth chamber plants, whereas lowest branching node, branch number, and total capitulum count did not differ between plants from different collection sites under these conditions (Tables 4.5 & 4.7).

Table 4.5: Mean and SE trait values by collection site for traits assessed in the growth chambers and under common garden conditions in Carman and Winnipeg in 2013 and 2014.

Trait	*P*-value	Woodlands	Emerson	Portage La Prairie	Brunkild
Common garden					
Average capitulum diameter (mm)	0.203	13.18 ± 0.70	12.80 ± 0.57	12.92 ± 0.64	14.48 ± 0.66
Days to anthesis (days)	0.0015*	211.91 ± 2.34	220.88 ± 1.33	216.41 ± 1.80	212.96 ± 1.96
Days to fifth anthesis (days)	0.0011*	217.26 ± 2.09	224.62 ± 1.15	221.85 ± 1.67	219.27 ± 1.73
Timing of shattering (days)	0.167	15.99 ± 1.72	18.90 ± 1.91	21.19 ± 2.10	19.12 ± 2.57
Growth chamber					
Average capitulum diameter (mm)	0.0046*	7.78 ± 0.41	9.08 ± 0.42	9.58 ± 0.40	10.39 ± 0.38
Capitulum count	0.0745	15.96 ± 0.89	20.44 ± 0.96	16.51 ± 0.86	15.04 ± 0.79
Days to anthesis (days)	0.0153*	61.39 ± 1.81	72.20 ± 1.84	60.73 ± 1.79	63.05 ± 1.77
Diameter of the first capitulum (mm)	0.0287*	9.49 ± 0.46	10.39 ± 0.47	11.26 ± 0.45	11.63 ± 0.43
Lowest branching node	0.1428	4.00 ± 0.65	6.29 ± 0.81	5.09 ± 0.67	7.10 ± 0.65
Plant height (cm)	0.0029*	81.31 ± 3.34	113.45 ± 3.52	93.92 ± 3.25	106.20 ± 2.99
Stem diameter (mm)	<0.0001*	6.23 ± 0.16	7.21 ± 0.18	7.44 ± 0.15	7.21 ± 0.13
Total branches	0.1147	15.53 ± 0.70	14.12 ± 0.74	13.67 ± 0.67	11.81 ± 0.62
Total nodes	0.0068*	17.12 ± 0.73	21.73 ± 0.77	16.69 ± 0.72	19.23 ± 0.67

Note: * Significant at the 0.05 probability level.

Table 4.5 continued: Mean and SE trait values by collection site for traits assessed in the growth chambers and under common garden conditions in Carman and Winnipeg in 2013 and 2014.

Trait	Oak Bluff	Dunnottar	Birds Hill	St. Pierre Jolys	Brandon
Average capitulum diameter (mm)	14.62 ± 0.60	12.84 ± 0.62	12.46 ± 0.62	13.35 ± 0.62	13.90 ± 0.59
Days to anthesis (days)	219.50 ± 1.59	213.15 ± 1.80	214.00 ± 1.66	220.29 ± 1.74	217.44 ± 1.53
Days to fifth anthesis (days)	225.60 ± 1.40	218.09 ± 1.59	219.59 ± 1.44	225.30 ± 1.56	223.64 ± 1.33
Timing of shattering (days)	17.95 ± 1.80	15.96 ± 1.36	18.78 ± 1.96	24.80 ± 2.60	20.17 ± 1.57
Average capitulum diameter (mm)	10.86 ± 0.40	7.78 ± 0.41	7.96 ± 0.48	8.96 ± 0.40	8.80 ± 0.39
Capitulum count	17.53 ± 0.85	12.91 ± 0.93	12.24 ± 1.16	14.06 ± 0.85	15.30 ± 0.81
Days to anthesis (days)	63.86 ± 1.77	55.81 ± 1.82	62.31 ± 2.06	63.06 ± 1.77	61.09 ± 1.75
Diameter of the first capitulum (mm)	12.28 ± 0.44	9.75 ± 0.46	10.15 ± 0.53	10.28 ± 0.44	10.34 ± 0.44
Lowest branching node	6.59 ± 0.71	4.29 ± 0.63	5.43 ± 0.90	5.63 ± 0.60	5.35 ± 0.65
Plant height (cm)	106.26 ± 3.15	86.66 ± 3.48	87.50 ± 4.28	103.52 ± 3.21	90.39 ± 3.09
Stem diameter (mm)	7.29 ± 0.15	5.97 ± 0.17	6.40 ± 0.23	6.77 ± 0.15	6.68 ± 0.13
Total branches	12.88 ± 0.66	14.43 ± 0.73	13.20 ± 0.93	13.22 ± 0.67	12.65 ± 0.64
Total nodes	20.42 ± 0.71	15.71 ± 0.76	15.84 ± 0.91	17.60 ± 0.71	16.14 ± 0.69

Note: * Significant at the 0.05 probability level.

Table 4.6: Significance of fixed effects for traits assessed under common garden conditions in 2013 and 2014 in Carman and Winnipeg sites examined using mixed linear model analysis.

Trait	Effect	Numerator DF	Denominator DF	*F* value	Pr > *F*
Days to anthesis	Collection site	8	37.2	4.06	0.002*
	Block	5	34.9	0.48	0.789
	Year	1	97.2	36.27	<0.001*
Days to fifth anthesis	Collection site	8	35.1	4.28	0.001*
	Block	5	32.6	0.64	0.674
	Year	1	95.5	76.44	<0.001*
Avg. capitulum diameter	Collection site	8	13.3	1.64	0.203
	Block	5	10.5	0.65	0.669
	Year	1	9.01	3.7	0.087
Timing of shatter	Collection site	8	12.5	1.81	0.167
	Block	5	26.3	0.86	0.522
	Year	1	5.92	29.66	0.002*

Note: * Significant at the 0.05 probability level.

Table 4.7: Significance of fixed effects for traits assessed under growth chamber conditions examined using mixed linear model analysis.

Trait	Effect	Numerator DF	Denominator DF	*F* value	Pr > *F*
Plant height	Population	8	6.90	10.64	0.003*
	Growth chamber	1	7.94	48.73	<0.001*
Total nodes	Population	8	6.82	8.04	0.007*
	Growth chamber	1	7.04	10.43	0.014*
Capitulum count	Population	8	3.01	6.56	0.075
	Growth chamber	1	6.74	50.67	<0.001*
Total branches	Population	8	6.72	2.63	0.115
	Growth chamber	1	7.98	196.99	<0.001*
Lowest branching node	Population	8	6.96	2.32	0.143
	Growth chamber	1	7.36	50.11	<0.001*
Days to anthesis	Population	8	7.58	5.48	0.015*
	Growth chamber	1	7.62	481.75	<0.001*
Diameter of the first capitulum	Population	8	8.44	4.08	0.0287*

Trait	Effect	Numerator DF	Denominator DF	*F* value	Pr > *F*
Avg. capitulum diameter	Growth chamber	1	8.58	1.82	0.212
	Population	8	8.46	7.27	0.005*
Stem diameter	Growth chamber	1	8.51	0.23	0.644
	Population	8	357	9.86	<0.001*
	Growth chamber	1	357	90.55	<0.001*

Note: * Significant at the 0.05 probability level.

A number of phenotypic traits were correlated with environmental variables with several traits being correlated to latitude and average daily temperature in the month of June (Supplemental Table 4.5). As might be expected, population collection site latitude was correlated with days to first anthesis in both the common garden ($r = -0.70$, $p = 0.036$) and growth chamber ($r = -0.80$, $p = 0.010$) plants, whereas average capitulum diameter ($r = -0.72$, $p = 0.030$), total nodes ($r = -0.83$, $p = 0.005$), stem diameter ($r = -0.72$, $p = 0.030$), and plant height ($r = -0.92$, $p = 0.001$) were correlated with latitude only for growth chamber plants (Supplemental Table 4.5). In the growth chamber plants average annual temperature was correlated to capitulum diameter ($r = -0.72$, $p = 0.028$), total nodes ($r = 0.73$, $p = 0.025$), stem diameter ($r = 0.82$, $p = 0.007$), and plant height ($r = 0.82$, $p = 0.007$). Average daily June temperature was also correlated with average capitulum diameter ($r = 0.74$, $p = 0.022$), diameter of the first capitulum ($r = 0.68$, $p = 0.043$), plant height ($r = 0.81$, $p = 0.008$), total nodes ($r = 0.82$, $p = 0.007$), and stem diameter ($r = 0.79$, $p = 0.011$) in the growth chamber experiment. Soil Cu content at the collection sites was correlated with days to fifth anthesis in the common garden plants ($r = 0.75$, $p = 0.02$). Traits associated with annual temperature and average daily June

temperature at the original collection location were correlated to one another in growth chamber plants and were associated primarily with overall plant size characteristics such as plant height (Table 4.8). This pattern is indicative that timing of anthesis and the transition from vegetative to reproductive growth (which in turn may influence traits such as plant height) is influenced by these environmental factors.

Table 4.8: Spearman's ranked correlation coefficients between population means (n = 9) for traits assessed under growth chamber conditions.

Trait		Average capitulum diameter	Capitulum count	Days to anthesis	Diameter of the first capitulum	Lowest branching node	Plant height	Stem diameter	Total branches
Capitulum count	*r*	0.54							
	p	0.137							
Days to anthesis	*r*	0.54	0.42						
	p	0.13	0.267						
Diameter of the first capitulum	*r*	0.98	0.53	0.45					
	p	<0.001*	0.132	0.224					
Lowest branching node	*r*	0.8	0.2	0.78	0.77				
	p	0.01*	0.606	0.013*	0.016*				
Plant height	*r*	0.86	0.57	0.78	0.83	0.87			
	p	0.002*	0.112	0.013*	0.01*	<0.001*			
Stem diameter	*r*	0.93	0.64	0.43	0.9	0.59	0.79		
	p	<0.001*	0.061	0.252	<0.001*	0.097	0.012*		
Total branches	*r*	−0.59	0.12	−0.25	−0.65	−0.70	−0.43	−0.40	
	p	0.091	0.785	0.517	0.058	0.036*	0.244	0.283	
Total nodes	*r*	0.7	0.72	0.87	0.62	0.7	0.83	0.64	−0.17
	p	0.034*	0.03*	0.002*	0.077	0.036*	0.001*	0.065	0.668

Note: * Significant at the 0.05 probability level.

† *r* denotes Spearman's ranked correlation coefficient.

‡ *p* denotes *p*-values.

4.4.4 Outlier loci detection and environmental association analysis

After the application of a FDR of 5%, deviations from neutral expectations were observed for 39, 99, and 170 loci for BayeScan, *pcadapt*, and Lositan, respectively. A total of 43 loci were observed to have *q*FDR <0.05 for two of the three outlier tests, and seven loci had a *q*FDR <0.01 for all three tests (Supplemental Table 4.6). Models fit in the EAA differed slightly between traits (Supplemental Tables 4.7-4.8), but overall, most models did not deviate greatly from expected values, indicating that population structure and kinship were adequately controlled. TASSEL detected a total of 12,651 significant ($p < 0.05$) SNP-trait associations between the 5,414 markers and the 25 environmental variables. After the application of *qFDR* correction, 200, 32, 11, and 6 associations remained at a FDR of 15, 10, 5, and 1%, respectively. At a FDR of 5%, seven SNPs exhibited significant correlations with environmental traits, one locus on LG 16 was associated with percentage $CaCO_3$ equivalent (CCE), two separate SNPs on LG 10 and LG 17 were associated with frost-free days above 0°C, and two additional SNPs on the same LGs were associated with elevation. Two closely linked SNPs on LG 9 were associated with average June daily temperature (Table 4.9). Four SNPs were associated with phenotypic traits: one SNP on LG 7 and one SNP on LG 13 were associated with diameter of the first capitulum, while one SNP on LG 11 and one SNP on LG 13 were associated with days to anthesis in the growth chamber. No SNPs were found to be associated with common garden assessed traits.

The SNP locus *HA16-220344187*, associated with CCE, showed the strongest SNP-environment or SNP-phenotype association (Table 4.9). The T allele is associated

with a 10.21% greater CCE relative to the heterozygous condition, whereas the C allele is associated with a −1.78% lower CCE. The SNP loci *HA11-188999507* and *HA13-187158373* were associated with earlier and later days to anthesis, respectively, in growth chamber plants, but were not associated with days to anthesis in the common garden plants. The *HA11-188999507* C allele was associated with later days to flower (+6.58 d), whereas the *HA13-187158373* T allele was associated with earlier days to anthesis (−6.28 d) relative to the heterozygote classes. *HA7-101023330* and *HA13-81101147* were associated with the size of the first capitulum to emerge.

Table 4.9: Allelic effects of single nucleotide polymorphisms (SNPs) associated with either environmental variables or phenotypic traits assessed in the growth chamber or under common garden conditions in Carman and Winnipeg in 2013 and 2014 using TASSEL.

SNP marker	LG	Position (bp)	Trait	*P*-value	Marker r^2	Allele	Additive effect
HA7-101023330	7	101,023,330	Diameter (mm) of first capitulum (growth chamber)	9.46×10^{-7}	0.240	C/C	−1.252
						C/T	0
HA9-206499835	9	206,499,835	Avg. daily temperature (°C) in June	1.19×10^{-6}	0.167	C/C	+3.53
						G/G	−0.67
						C/G	0
HA9-206499851	9	206,499,851	Avg. daily temperature (°C) in June	5.09×10^{-6}	0.163	C/C	+3.60
						G/G	−0.71
						C/G	0
HA10-53871327	10	53,871,327	Frost-free days above 0°C	1.22×10^{-5}	0.218	T/T	−9.376
						C/T	0
HA10-218772737	10	218,772,737	Elevation (m)	2.91×10^{-6}	0.164	A/A	−1.042
						T/T	+5.890
						A/T	0
HA11-188999507	11	188,999,507	Days to anthesis (growth chamber)	2.41×10^{-6}	0.264	C/C	+6.587
						C/G	0
HA13-81101147	13	81,101,147	Diameter (mm) of first capitulum (growth chamber)	3.13×10^{-6}	0.274	C/C	−0.672
						T/T	+1.184
						C/T	0
HA13-187158373	13	187,158,373	Days to anthesis (growth chamber)	1.69×10^{-6}	0.245	T/T	−6.284
						C/T	0
HA16-220344187	16	220,344,187	$CaCO_3$ equivalent (%)	1.79×10^{-7}	0.497	C/C	−1.787
						T/T	10.213
						C/T	0
HA17-216837251	17	216,837,251	Frost-free days above 0°C	1.36×10^{-5}	0.267	A/A	−1.792
						G/G	+8.801
						A/G	0
HA17-265653117	17	265,653,117	Elevation (m)	7.78×10^{-6}	0.160	C/C	+20.651
						T/T	−6.750
						C/T	0

The *HA7-101023330* C allele negatively on average affected capitulum diameter relative to the heterozygote class (−1.25 mm). The C allele of *HA13-81101147* also negatively affected the first capitulum diameter (−0.67 mm), whereas the T allele was associated with increased average capitulum diameter (+1.18 mm) relative to the heterozygote class. The *HA10-53871327* and *HA17-216837251* loci were associated with frost-free days above 0°C. The T allele of *HA10-53871327* was associated with a shorter frost-free period (−9.37 d), as was the *HA17-216837251* A allele (−1.79 d) relative to the heterozygote class. The G allele of *HA17-216837251* was associated with a longer frost-free period (+8.80 d). The loci *HA10-218772737* and *HA17-265653117*, which were associated with collection site elevation, had the smallest associations with environment or phenotype. The C *HA17-265653117* allele was associated with higher elevation (+20.65 m), whereas the T allele was associated with lower elevation (−6.75 m) relative to the heterozygous state.

The SNPs *HA9-206499835* and *HA9-206499851* were both associated with daily average June temperature. The C allele of *HA9-206499835* was associated with an average temperature of 3.53°C greater than the heterozygote class, whereas the G allele was 0.67°C lower. *HA9-206499851* showed a similar association, with the C allele associated with a 3.6°C greater average temperature.

4.4.5 Candidate gene annotation

SNP reads were compared with the cultivated *H. annuus* HA412.v.1.1.bronze.20142015 sequence assembly to identify putative genes and protein annotations. Fourteen SNPs out of 39 (35%) loci identified by outlier analysis were found within either known or predicted genes in the *H. annuus* or *A. thaliana* reference genomes (Supplemental Table 4.9). Two pairs of SNPs were found to be within the same gene leaving 12 unique candidate genes, of which 11 had associated Interpro and GO annotations. None of the SNPs associated with either phenotype or environment had known annotations in either reference genome, and BLASTx results returned hits for seven of the outlier SNPs and two of the TASSEL SNPs (Supplemental Table 4.10). All BLASTx results returned predicted uncharacterized proteins and did not contain associated GO terms (Supplemental Table 4.11).

4.5 Discussion

4.5.1 Phenotypic variation

Phenotypic differences were observed in the common garden and in the growth chamber experiments which has implications for the adaptation of Maximilian sunflower to different growing locations. Difference in mean days to anthesis between the earliest and latest collection site was ~8 days in the common garden plants and 16 days in the growth chamber plants. These differences represent a significant portion of the growing season in Manitoba, with averages ranging from a total of 109 to 131 days among the

collection locations. Imposing selection on this characteristic may prove important in the adaptation of Maximilian sunflower to regions that differ in growing season length. Timing of anthesis is a major characteristic in the adaptation of *H. annuus* and other crops to new growth environments to take advantage of available resources in regions with shorter or longer growing seasons (Craufurd and Wheeler, 2009; Blackman et al., 2011). Cultivated sunflower adapted to Canadian growing conditions has undergone selection for shorter, earlier-maturing cultivars to coincide with limitations in the growing season length. Differences were observed in average plant height amongst collection sites under growth chamber conditions, suggesting that this trait may be selected for to suit agronomic production.

Selection for shorter plants with smaller capitula that could be grown at higher densities and harvested earlier in the growing season has been suggested as a route to increase yields in cultivated *H. annuus* (Vear, 2016). Earlier maturing plants with a shorter stature capable of being grown at high densities similar to other oilseed crops such as canola (*Brassica napus* L.) may be a potential ideotype for Maximilian sunflower given its ability to produce multiple stems bearing many capitula. Improvements in capitulum size and uniformity in the timing of maturation in Maximilian sunflower would likely be needed under this scenario to produce acceptable yields. Selection for uniformity between stems would likely be necessary under this scenario to ensure uniform development and maturity. Stem emergence may not be entirely synchronized in Maximilian sunflower, likely due to differences in rhizome size and depth of planting. These differences in timing of emergence may affect growth and development, resulting in stems that differ in size. Mean capitulum size did not differ amongst plants from the

different collection sites in the common garden plots. Variability introduced through the measurement of multiple stems per plant and the potential differences in emergence or resource capture amongst stems may contribute to these differences. Average capitulum diameter was positively correlated with stem height and diameter in the growth chamber, indicating that stem size can influence this trait. In the growth chamber, the influence of competition for resources such as light between stems was not a factor due to the establishment of single-flowering stems with predictable stem spacing from seeds in the growth chamber. Significant differences amongst collection sites were detected for both average capitulum diameter and the first capitulum to reach anthesis, suggesting variation in these traits at a local scale.

Timing of shattering did not differ amongst collection sites and was primarily influenced by year effects. Shattering was observed to occur 16 to 25 d after anthesis, which was sooner than expected, although achenes were collected for further studies we did not have the necessary labor to determine if they were unfilled or viable. Further work is needed to understand the shattering characteristics of Maximilian sunflower, particularly the influence of environmental effects such as plant moisture status. For example, if shattering is significant, is the loss of potential yield due to inadequate pollination, or to some other factor causing premature dehiscence? If development of shorter, early-maturing plants capable of being produced at high densities is a suitable ideotype for this species, factors influencing the synchronization of emergence, flowering, maturity, and timing of shattering will need to be explored further to develop appropriate agronomic practices for harvesting. Currently, the focus of programs working in Maximilian sunflower has been on improving seed production and developing a

perennial crop that could be seeded in a similar fashion to other small-seeded oilseeds. Further research is needed to develop appropriate agronomic practices that will help define the appropriate ideotype. Given the phenotypic range observed in local materials within this study as well as Chapter 3, and wide geographic distribution of *H. maximiliani* (Heiser, 1969; Kawakami et al., 2011), the genetic potential to select for materials with adapted phenology for different growing regions useful yield contributing characteristics appears favourable in this species.

4.5.2 Population structure

The AMOVA revealed significant but weak population structure in wild populations of Maximilian sunflower sampled in Manitoba. This suggests that while the sampled populations approach expectations of random mating under Hardy-Weinberg equilibrium they exhibit a degree of differentiation. The greatest source of variation in the SNP dataset was at the level of the individual plant (61%), followed by variation among individuals within the collection sites (37%). The lowest molecular variation was observed among collection sites (2%). Despite the ability of *H. maximiliani* to spread rhizomatously, collection sites do not appear to be dominated by a few widely dispersed clones as observed in other perennial sunflower species such as Jerusalem artichoke (*Helianthus tuberosus* L.), which spreads aggressively through vegetative rhizomes and tubers while producing little seed (Swanton et al., 1992). Individuals within collection sites showed low levels of inbreeding (F_{is}), indicative of random mating. Inbreeding within subpopulations relative to the region as a whole (F_{st}) is similarly low, indicative of

low genetic differentiation between populations. This was supported by the STRUCTURE analysis and PCA, which shows a majority (six of nine) of the collection sites belonging to a single group, G1, as well as the presence of admixtures within G2 and G3 from G1. This result is somewhat expected due to the relative proximity of collection sites (with a maximum distance among the collection sites of only 219 km) and relative abundance of Maximilian sunflower populations in some regions of southern Manitoba. Population structure in outcrossing species is often weak due to the rapid decay of linkage, greater recombination, the maintenance of heterozygosity, and large effective population sizes relative to self-pollinated species (Flint-Garcia et al., 2003; Gupta et al., 2005). In parts of southern Manitoba, Maximilian sunflower is commonly found in roadside ditches near agricultural fields, forming long stretches of semi-continuous populations. This commonly occurring habitat could potentially serve as a corridor for gene flow, explaining the low level of differentiation amongst some of the collection sites.

The low inbreeding coefficient values of F_{st} and F_{is} is indicative that mating both within and amongst collection sites is mostly random. The slight negative values indicate an excess of heterozygosity, suggesting either negative assortative mating or selection in favour of heterozygotes (Hartl and Clark, 1997). Negative assortative mating is expected in wild *Helianthus*, an obligate outcrosser, due to the presence of sporophytic self-incompatibility, which limits mating between close relatives (Gandhi et al., 2005; Liu and Burke, 2006). The Brandon site, which was the most western site, exhibited the greatest heterozygosity and polymorphic loci and the lowest F_{st} and F_{is} values, indicating that it is the most genetically diverse collection site. The Oak Bluff site exhibited the lowest level

of heterozygosity, the highest F_{st} value, and the second highest F_{is} value, indicative of a degree of inbreeding in this population. Genetic load, the accumulation of deleterious recessive alleles, is anticipated in outcrossing species as a function of the maintenance of heterozygosity and a reduced capacity for selection to act on recessive phenotypes (Barrett and Charlesworth, 1991). Interestingly, plants from the Oak Bluff site grown in the growth chamber from seed were noted to exhibit petal color and morphology not observed at the collection location in field surveys or in the common garden plots. One particular maternal seed family segregated (5 of 18 plants) for a "lemon petal" trait in which ray petals were light yellow as opposed to the common golden type, a trait that is known to be recessive and controlled by one to two genes in cultivated *H. annuus* (Yue et al., 2008). This trait was also observed in a single plant grown from seed from the Birds Hill site, which exhibits similar F_{is} and F_{it} values to Oak Bluff, but greater heterozygosity and percentage of polymorphic loci.

A small number of plants grown from seed from the Oak Bluff site also exhibited tubular ray flowers, a recessive condition in which ray flowers are radialized, forming a tubular-like ray and not the common flat zygomorphic presentation. This trait is also described as recessive in *H. annuus* and likely controlled by a single locus, *HaCYC2c* (Chapman et al., 2012). Although the presence of these recessive traits is not indicative of inbreeding per se, they may represent a fitness cost, which keeps these traits at low frequencies. Alterations of ray flower morphology or color may be detrimental through alterations in pollinator attraction and subsequent seed set and may be actively selected against under pollinator-mediated selection (Stuessy et al., 1986). Tolerance of Maximilian sunflower to inbreeding depression is currently unknown, but the low levels

of inbreeding are indicative of a strong selection for the maintenance of heterozygosity. Future selection efforts will require the characterization of genetic load to determine inbreeding tolerance and the potential to purge genetic load.

4.5.3 SNP yield and linkage disequilibrium

The 24,957 SNPs called in this study provide several hundred SNPs per LG, which may be used in future mapping studies in Maximilian sunflower, a species with limited genomic characterization. Kawakami et al. (2014) identified 2277 SSR and 2062 polymorphic SNPs within the species through comparing individuals collected from Manitoba and Texas populations. These SNPs add to the growing resources for the perennial crop wild relatives of sunflower (Kawakami et al., 2014; Baute et al., 2016). Although the genus *Helianthus* has undergone a recent and rapid radiation (Timme et al., 2007), the sequence divergence of annual and perennial clades of *Helianthus* is considerable (Bock et al., 2014; Baute et al., 2016) and the complex perennial clades have limited genetic characterization at the population level. The potential for ascertainment bias through the use of the cultivated *H. annuus* reference genome cannot be ruled out, and a reference-free *de novo* approach may yield additional SNPs for regions of the genome that may differ between annual and perennial clades, as well as identify SNPs that are linked to unique features of *H. maximiliani*. The rapid linkage decay we observed is in line with previous estimates from wild *Helianthus* species (Liu and Burke, 2006; Mandel et al., 2011) and is largely a function of the obligate outcrossing nature of the genus. Previous studies have reported that linkage decay occurs

rapidly, at ~200 bp in wild *Helianthus* species and at 1100 bp in cultivated *H. annuus* lines (Liu and Burke, 2006; Mandel et al., 2011). The rate of linkage decay based on marker order in the cultivated *H. annuus* reference genome (~93 bp) and the large genome size of diploid *Helianthus* species (~3.6 Gb) indicates that similar to other wild *Helianthus* species, marker density for genome wide association studies or genomic selection in Maximilian sunflower would need to be high to capture historic recombination events in wild collected germplasm. Although several wild annual *Helianthus* species are reported as being highly syntenic with *H. annuus* (Barb et al., 2014), the degree of structural chromosomal rearrangements between cultivated *H. annuus* and the perennial *H. maximiliani* is currently uncharacterized, which may influence the inference of linkage decay. Further research is needed to confirm marker order and chromosomal structure in this species and other perennial *Helianthus*. Recent efforts using association mapping and genome-wide association approaches in sunflower have employed marker densities ranging from ~5,300 to ~450,000 SNPs (Mandel et al., 2013b; Nambeesan et al., 2015; Badouin et al., 2017) with linkage decays within 1100 bp (Liu and Burke, 2006). Given the rapid linkage decay, approaches in which linkage blocks are maintained to a greater degree through consanguinity, such as through family-based association mapping approaches (Guo et al., 2013), or pseudo-testcross approaches, such as those used in other outcrossing species (Zhang et al., 2016; Covarrubias-Pazaran et al., 2016), may be appropriate in Maximilian sunflower.

4.5.4 Environmental associations and implications for selecting for regional adaptation

Optimal resource partitioning between vegetative and reproductive growth is critical for successful reproduction in plants. Trade-offs are often observed between biomass accumulation and timing of reproduction (Egli, 2011). In addition to growing season and photoperiod gradients, many plant species have the tendency to produce less biomass and transition from vegetative to reproductive growth earlier in the growing season when moving north across latitudinal gradients. In *Helianthus*, flowering is strongly influenced by both photoperiod and temperature cues (Leon et al., 2001). Days to anthesis has been noted to decrease with increasing latitude in Maximilian sunflower and other wild *Helianthus* species (Kawakami et al., 2011; Henry et al., 2014), and a latitudinal effect on flowering was observed in populations collected for this study of plants grown both in common garden plots and growth chambers. Single nucleotide polymorphisms associated with temperature (frost-free days, average daily June temperature, and elevation), days to anthesis and capitulum diameters were detected. The lack of SNPs detected for phenotypic traits measured in the common gardens may be due to environmental effects increasing the level of noise in the data, reducing the ability to differentiate the phenotypic divergence amongst collection sites relative to the growth chamber. Factors such as timing of emergence, the initial condition of the rhizomes on transplantation from the wild and overall establishment and growth may have played a role. Establishing the common garden plots by seed or rhizomes grown under common conditions may reduce potential noise in future studies.

Maximilian sunflower is known to exhibit striking clinal variation across its native range, with distinct northern and southern ecotypes. Previous studies have established that along a latitudinal transect of North America, plant size-related traits (plant height, capitulum diameter, stem diameter, and leaf mass) decrease with increasing latitude in Maximilian sunflower (Kawakami et al., 2011). Similarly, in this study, days to anthesis and average capitulum diameter measured in both common garden plots and in the growth chamber, as well as stem diameter, total nodes, and plant height only measured in the growth chamber, were correlated with the original collection site latitudes, showing a similar pattern of differentiation across a latitudinal gradient.

Tetreault et al. (2016) suggested that phenological differences between Manitoba, Kansas, and Texas populations of Maximilian sunflower might serve as a mechanism of avoiding local abiotic stress, particularly low temperatures. Ancestry coefficient G1, average capitulum diameter in the common garden plots and in the growth chamber, days to first anthesis, first capitulum diameter, stem diameter, total nodes and plant height were positively correlated with increasing annual temperature and with average daily temperature in June. Several plant traits related to size (plant height, capitulum size, stem diameter, and total nodes) were associated with annual temperature and average daily June temperature, and may represent an adaptation to heat unit availability and overall biomass accumulation. The month of June is when northern populations of Maximilian sunflower experience rapid growth and accumulation of biomass prior to anthesis. In our experiments, the average date of first anthesis observed under common garden conditions in southern Manitoba ranged from July 29th to August 7th, with some individuals flowering as early as July 14th.

Given this information, selection for earlier flowering plants of smaller stature may be a suitable strategy for growing regions with fewer available heat units, or shorter frost-free periods, as found in more northern latitudes. Larger plant size as determined by plant height, stem diameter, total nodes, and capitulum diameter might be better suited for regions with greater heat unit availability. There appears to be sufficient phenotypic variation amongst collection sites to select for earlier or later days to anthesis and examine the effects on correlated traits such as plant height. Further research is needed to determine resource partitioning patterns within Maximilian sunflower and the influence of number of days to anthesis and seed yield characteristics on traits such as plant biomass production, the development of overwintering structures, and resulting stand longevity.

4.5.5 Candidate gene detection

Fourteen SNPs at 12 loci identified using outlier analyses were correlated with known genes in the cultivated *H. annuus* or *A. thaliana* reference genomes. Several of the identified genes were associated with potential functions of interest related to adaptation and survival and warrant further investigation. Two SNPs on LG 5, *HA5-152646835* and *HA5-152646876*, have sequence similarity with the *Gnk2-homologous domain* in the cultivated *H. annuus* reference genome. *Gnk2* produces an antifungal protein Ginkobilobin-2 found in the endosperm of *Ginkgo* seeds, which inhibits the growth of phytopathogenic fungi such as *Fusarium* spp. (Miyakawa et al., 2014). *HA7-911263*, found on LG 7, has similarity to an armadillo-type fold protein. McAssey et al. (2016)

previously observed a SNP outlier corresponding to an ARM repeat protein in *H. annuus* on the same LG (19.29 cM position), which co-localized with quantitative trait loci for flowering time, plant height, leaf number, and susceptibility to head herbivory by insects in previous mapping studies (Burke et al., 2002; Dechaine et al., 2009). *HA10-86839089* on LG 10 was annotated as similar to the known gene *MALE GAMETOPHYTE DEFECTIVE 1* (*MGP1-AT2G21870*) in *Arabidopsis,* involved in the mitochondrial F1FO-ATP synthase necessary for viable pollen formation (Li et al., 2010). *HA11-5185166* on LG 11 has similarity to a DnaJ domain, a type of heat shock protein that has been identified as a protein of interest in *Helianthus*, associated with abiotic stress tolerance (Kane and Rieseberg, 2008; Scascitelli et al., 2010; Prunier et al., 2013). The remaining outlier loci showing sequence similarity to known genes include functions such as DNA methylation and translation, intracellular signaling, protein phosphorylation, protein deubiquitination and carbohydrate binding.

A number of loci were identified within close proximity to annotated genes in the cultivated *H. annuus* reference genome. These include *HA5-148744834*, which was identified as being 608 bp upstream of a gene involved with amidase activity, and three SNPs, *HA17-53277130*, *HA17-53277133*, and *HA17-53277164*, which were 377, 380, and 411 bp downstream of a ribonuclease H domain associated with DNA repair. None of the 11 SNPs identified using TASSEL at a FDR correction rate of 5% fell within described genes in the *H. annuus* reference genome or matched characterized proteins in the Uniprot database. One locus, *HA17-265653117*, associated with elevation was found to be in close proximity downstream to an F-box domain (206 bp downstream), a class of proteins known to play essential roles in plant growth and development (Bu et al., 2014).

Collectively, the SNPs identified by TASSEL and outlier analysis include several candidate sequences that are potentially in biotic and abiotic stress tolerance genes in Maximilian sunflower and may prove useful in crop breeding. Through marker-assisted selection, these loci may be useful in efforts to develop *H. maximiliani* as a perennial oilseed crop. As with other outcrossing species, allele fixation occurs at a slower rate under cross-pollinated breeding systems than in self-pollinated species in which a single favourable individual may give rise to a useful breeding population. Marker-assisted selection allows for a more efficient selection of favourable offspring in segregating populations and is of particular use in open-pollinated populations where allele fixation occurs slowly due to the presence of multiple parents in crosses, the maintenance of heterozygosity over time, and allelic effects that may be masked through dominance effects throughout multiple cycles of selection.

4.5.6 Landscape genomics as a tool in sunflower breeding

The application of landscape genomics is an attractive tool in expanding the adaptive range of crops by connecting the underlying mechanisms driving phenotypic divergence to landscape features, and uncovering causative loci influencing adaptation. These approaches have the distinct advantage in that specific environments can be targeted without *a priori* knowledge of the phenotypic traits underlying adaptation. Using a landscape genomics approach, we identified several candidate SNPs associated with potential sources of abiotic stress and biotic stress that could be useful in expanding the current range of cultivated sunflower to more northern growing regions. Through EAA,

SNPs associated average daily temperatures, frost-free days above 0°C, elevation and soil properties such CCE (which may influence soil pH and nutrient availability) were identified and could (on validation) be applied to enhance abiotic stress tolerance in Maximilian sunflower and related species such as cultivated sunflower. Population differentiation identified SNPs under putative selections in southern Manitoba, with flanking sequences exhibiting similarities to annotated genes with potential antifungal and heat shock protein functions that could be further investigated to determine potential stress factors and targets for selection. Given the limited number of studies that have applied landscape genomics in crop breeding compared with applications in the areas of conservation and forestry (Sork et al., 2013, Rellstab et al., 2015), these techniques may have only scratched the surface in terms of their potential to enhance field crops. The correlations observed in this study between phenotype differentiation, population structure, and environment give weight to the use of landscape genomics to develop locally adapted crops. Cultivated sunflower benefits from a wealth of crop wild relatives that persist in a wide range of marginal habitats beyond the current adapted range of the crop. It is estimated that much of the primary gene pool of cultivated sunflower exists in extreme environments (high temperature fluctuations, low annual precipitation and short growing season environments), leaving the potential for adaptive traits to be, in theory, introgressed with relative ease (Kantar et al., 2015). The ubiquity of wild relatives in the genus *Helianthus* suggests untapped potential to use landscape genomics as a tool for determining and harnessing meaningful traits from exotic germplasm. Incorporating traits from crop wild relatives into cultivated sunflower is a potential route for the development of locally adapted crops and adjustment to limitations imposed by climate change

(Dempewolf et al., 2014). With the growth of next-generation sequencing technologies and the release of the sunflower reference genome, and by applying landscape genomics to target underutilized environments, it should be possible to identify useful traits for adaptation to new growing environments. To date, the use of crop wild relatives to enhance abiotic stress tolerance in cultivated sunflower is considered relatively untapped (Kantar et al., 2015). Employing landscape genomics has the potential to uncover useful traits which connect genotypes to abiotic stressors of different landscapes. Harnessing these resources may expand the range of cultivated sunflower and facilitate the development of novel crops, such as perennial oilseeds or those adapted to a wider range of environmental conditions than existing crops.

4.6 Conclusions

This study revealed the presence of weak, but significant, population structure associated with a temperature gradient and phenotypic divergence in local populations of Maximilian sunflower. High levels of historical recombination and heterozygosity are indicative of low levels of inbreeding, suggesting that cross-pollinated breeding strategies are likely the best approach for initial improvement in this species. Traits of interest for agronomic production such as days to anthesis and correlated plant size traits show favourable differences among the nine populations sampled from a relatively small geographic area, indicating that local populations harbor the necessary diversity to make advancements under selection. Cross-pollinated breeding strategies such as half-sibling evaluation with test crossing or progeny testing may be best suited to avoid inbreeding

depression and may be enhanced through the use of marker-assisted selection. Given the high levels of heterozygosity and rapid linkage decay in this species, future efforts employing association-based approaches should focus on increasing marker yield by using a less stringent level of genome reduction to increase mapping resolution. Alternatively, phenotypic selection for traits of interest may increase the size of stable linkage blocks and reduce linkage decay, which could be useful for establishing marker-trait associations in breeding populations as fewer markers may be necessary to capture recombination events.

Maximilian sunflower, like many crop wild relatives of cultivated sunflower, remains an untapped resource for the improvement of sunflower and the development of novel crops. Evidence for adaptation at the local scale indicates that Maximilian sunflower may be mined for useful agronomic traits from the different environments across its extensive native range in North America using a landscape genomics approach. A targeted approach of characterizing crop wild relatives of cultivated sunflower from environments of interest will likely yield further discoveries and expand the current range of the crop. Candidate SNPs associated with days to anthesis and capitulum size, as well as for abiotic and biotic stresses, were identified in the diverse cold-hardy species *H. maximiliani* and may be useful for future marker-assisted selection in Maximilian sunflower and in related species such as the cultivated *H. annuus*. The identification of environmental factors that are influencing population structure and plant characteristics in *H. maximiliani* provides a pathway for developing suitable materials targeted towards different growing environments in cultivated sunflower and the development of Maximilian sunflower into a perennial grain oilseed.

5.0 Genetic analysis of domestication syndrome traits in Maximilian sunflower (*Helianthus maximiliani* Schrad.).

5.1 Abstract

Parallels exist between the domestication of and the improvement of many plant species through selection on traits which favour the sowing, harvest and retention of yield potential in crops and directed efforts to improve the agronomics, disease resistance and quality characteristics. Common selection pressures may result in the parallel selection of orthologs underlying these traits and homologies between crop species are often exploited by plant breeders to improve germplasm. Perennial grains and oilseeds are a class of proposed crops for improving the diversity and sustainability of agricultural systems. Maximilian sunflower (*Helianthus maximiliani* Schrad.) is a perennial crop wild relative of sunflower (*Helianthus annuus* L.) and a candidate perennial oilseed species. Understanding parallels between cultivated *H. annuus* and *H. maximiliani* may provide new tools for the development of Maximilian sunflower as a crop. F_2 populations of Maximilian sunflower segregating for traits including branching architecture, capitulum morphology and flowering time were developed to investigate parallels between *H. maximiliani* and *H. annuus*. Genotype-by-sequencing (GBS) was employed to develop a genetic map of Maximilian sunflower using existing genomic resources from cultivated sunflower and to perform quantitative-trait-loci (QTL) analysis and association mapping. The results of QTL and association mapping revealed 7 QTL and 20 candidate SNPs which correspond to linkage groups known to harbor domestication syndrome associated traits in previous studies in *H. annuus.* This suggests the potential to exploit orthologs for

neo-domestication of *H. maximiliani* for traits such as branching architecture, timing of anthesis and capitulum size and morphology for the development of a perennial oilseed crop. Additional QTL and SNP associations unique to *H. maximiliani* were discovered which favour the development of marker-assisted-selection based breeding strategies for the improvement of this species as a perennial oilseed.

5.2 Introduction

Maximilian sunflower (*Helianthus maximiliani* Schrad.) is an herbaceous forb that is native to much of the interior plains of North America, with a range extending from southern Canada to northern Mexico (Kawakami et al., 2011, 2014; Tetreault et al., 2016). Maximilian sunflower has been used in range seeding mixtures for high quality livestock forage, as a perennial filter strip to reduce agricultural run-off, and as a source of wildlife food and habitat (Dietz et al., 1992; USDA-NCRS, 2017).The natural range of Maximilian sunflower, along with its phenotypic divergence (Kawakami et al., 2011), high seed production potential, and documented resistance to known common pathogens of annual sunflower (*Helianthus annuus* L.) such as Sclerotinia rot [*Sclerotinia sclerotiorum* (Lib.) de Bary] (Taski-Ajdukovic et al., 2006; Liu et al., 2011), have attracted attention to the species for crop development. Maximilian sunflower is a candidate species for the development of dual-use perennial crop for grain, forage and bioenergy applications and is targeted for domestication and improvement (Cox et al., 2002; Van Tassel et al., 2014; Chapter 4).

Efforts to improve of Maximilian sunflower were focused on use for conservation and pasture applications or as a trait donor for *H. annuus* until recently. Two open

pollinated commercialized cultivars, ‘Aztec’ and ‘Prairie Gold’ were released by the United States Department of Agriculture Natural Resources Conservation Service (USDA-NRCS) in 1978 (Texas Agric. Exp. Stn., 1979; Dietz et al., 1992; USDA-NCRS, 2017). Aztec was developed for wildlife feed, livestock forage cover, use as a natural hedge, filter-strips, and as an ornamental landscape plant. Prairie Gold was released for landscape reclamation and wildlife food plantings. Both cultivars were selected for vigour and stand establishment in Oklahoma and Texas (Aztec), or Kansas and further north (Prairie Gold). Agronomic research on Maximilian sunflower as a perennial grain candidate began at The Land Institute in the 1980s (Jackson, 1990). The first breeding program focused on developing Maximilian sunflower as a perennial grain was launched in 2002 (Cox et al., 2002). Selection for seed size and apical dominance has been effective and following three rounds of recurrent selection, the average seed size was increased by 240% and plants exhibiting highly restricted branching were developed by 2012 (Van Tassel et al., 2014).

As a tertiary member of the cultivated sunflower (*H. annuus*) genepool, screening *H. maximiliani* for traits to introgress into *H. annuus* has been conducted. Use of *H. maximiliani* in *H. annuus* breeding programs has focused on the introgression of favourable disease resistance characteristics and as a source of cytoplasm and restorer genes for hybrid production (Taski-Ajdukovic et al., 2006; Feng and Jan, 2008). Characterization of low temperature tolerance and development of transcriptome resources in *H. maximiliani* for the improvement of abiotic stress tolerance in *H. annuus* has been recently conducted (Kawakami et al., 2014; Tetreault et al., 2016) and may be utilized to expand the adaptive range of cultivated sunflower (Kantar et al., 2015).

Genome characterization of *H. maximiliani* and other perennial *Helianthus* species is limited. The genomes of annual *Helianthus* species are believed to be highly syntenic which may be partially attributed to the rapid and recent radiation of different species within the genus (Timme et al., 2007; Barb et al., 2014). However, numerous rearrangements are reported between species. Similar to *H. annuus*, *H. maximiliani* is a diploid species (n=17). The two species have the capacity for hybridization, but pollen grain viability in subsequent F_1 progeny is variable, ranging from 5% to 93.2% (Whelan, 1978; Atlagić et al., 1995) and seed set is poor. *H. annuus* x *H. maximiliani* F_1 hybrids cytologically exhibit a paracentric inversion and at least three translocations (Whelan, 1978), suggesting that the genomes are similar, but structural differences pose a reproductive barrier. Embryo rescue is often required to produce viable progeny (Espinasse et al., 1991; Breton, 2010). Early generation introgression lines often exhibit poor agronomics, fertility and quality characteristics, and may require several generations of backcrossing to recover acceptable quality (Whelan and Dorrell, 1980; Atlagić et al., 1995; Jan et al., 2014).

Crop domestication shows many parallels between species. In grains and oilseeds such as sunflower, this includes selection for a more determinate growth habit, larger inflorescence size and increased grain size (Zohary, 2004; Purugganan and Fuller, 2009; Lin et al., 2012). A number of domestication syndrome and quality traits in annual grains are controlled by alleles with large effects and relatively simple genetic control. Domestication has led to the selection of orthologs such as *sh1* genes which contribute to shattering tolerance in sorghum, rice and maize (Lin et al., 2012); *Q* genes in wheat (*T. aestivum* L.) and barley (*Hordeum vulgare* L.) which confer the brittle rachis/free-

threshing trait (Simons, 2005); and variations in fatty acid desaturase genes in oilseeds such as canola (*Brassica napus* L.) and sunflower (Peng et al., 2010; Chapman and Burke, 2012). Strong apical dominance and restricted branching were key traits identified for the domestication of *H. annuus*. Wild type Maximilian sunflower has many similarities to wild type *H. annuus*, which exhibits profuse branching, small achenes, indeterminate flowering and lacks shattering resistance (Burke et al., 2002; Wills and Burke, 2007). Genomic resources available in annual sunflower open the possibility of using it as a model for the domestication and improvement of Maximilian sunflower through marker-assisted-selection (MAS) for domestication orthologs and the development of genomic selection. Marker-assisted breeding methodologies can rapidly fix existing or introduced traits in populations, and reduce the cost of genotyping and the length of breeding cycles (Zhang et al., 2016). Utilizing available genetic variation within Maximilian sunflower may by-pass the challenges associated with fertility, cytoplasmic-nuclear genome incompatibilities and linkage drag noted when crossing *H. maximiliani* with its domesticated counterpart *H. annuus* (Dorrell and Whelan, 1978; Jan et al., 1992; Atlagić et al., 1995; Jan et al., 2014).

Wild members of the genus *Helianthus* exhibit sporophytic self-incompatibility (Heiser, 1954; Heiser et al., 1969), limiting inbred-line development and pedigree selection strategies. Obligate out-crossing reduces the rate of fixation of desired alleles in populations. In addition, the requirements of multiple years of assessment are often necessary for each round of selection in perennial crops. Identification of alleles associated with improved agronomic traits and the use of genotypic selection could

increase the rate of genetic improvement in heterogeneous populations of Maximilian sunflower.

Genotype-by-sequencing has successfully been utilized for species in which heterologous reference genomes are available for SNP discovery such as for wild crop relatives of soybean (Chang et al., 2014), wheat (Edae et al., 2016) and sunflower (Baute et al., 2016). A genetic map for Maximilian sunflower would assist the development of the species as a perennial grain crop and facilitate improvement of *H. annuus*. *H. maximiliani* currently lacks a linkage map and the development of segregating mapping populations in this species could provide insights into the genetic control of traits of interest for neo-domestication and parallels between *H. maximiliani* and *H. annuus*. Incorporating *H. maximiliani* SNP markers into a genetic map would provide for the analysis of quantitative trait loci (QTL) to achieve these objectives.

Genome-wide association analyses are typically applied to unstructured populations exhibiting a high degree of linkage disequilibrium (LD), but can serve as useful tools to identify trait associations in bi-parental populations (Henning et al., 2016, 2017). Accounting for population structure is a substantial challenge in the control of false positive and false negative results (Yu et al., 2006; Bradbury et al., 2007). One method of accounting for population structure is to employ half or full-sibling families of known parentage to account for relatedness (family-based association mapping or FBAM) (Guo et al., 2013). Association mapping does not require the development of a linkage map, in which marker order influences the ability to detect significant marker-trait associations, and is not limited to markers in which the parental phase is known allowing for a greater number of markers to be employed.

The goal of this research was to develop the first linkage map of Maximilian sunflower using the annual sunflower reference genome as a template and to perform both QTL analysis and association mapping for traits of interest for the neo-domestication of this species.

5.3 Materials and methods

5.3.1 Population development

The mapping populations described in this study were produced through crossing highly-branched wild-type *H. maximiliani* plants to advanced unbranched breeding lines developed by The Land Institute (TLI), Salina, Kansas. Full-sib families were developed for mapping. Materials from TLI were screened under controlled environments and one accession was selected as a parent for the development of the mapping population as it exhibited a complete suppression of branching, golden coloured ray florets and a single large central capitulum (Supplemental Figure 5.1). The Manitoba parent exhibited extensive branching, lemon coloured ray petals and a large number of small capitula (Supplemental Figure 5.1). The resulting F_1 plants were clonally propagated from rhizome cuttings to produce materials for sib-crossing. Three F_1 individuals (herein denoted as F1A, F1B and F1C) were intercrossed through reciprocal sib-mating to generate two F_2 populations, herein denoted as crosses AB (F1A/F1B) and BC (F1B/F1C).

5.3.2 F_2 propagation

Seeds of the F_2 were surface sterilized using a 70% ethanol solution for 10 minutes, allowed to air dry, and seed coats were manually removed to break seed dormancy. Seeds were placed in 9 cm petri dishes on filter paper moistened with distilled water containing a 0.1% solution of plant preservative mixture (PPM™, Plant Cell Technology, Washington, D.C., U.S.A.). Petri plates were placed in the dark at room temperature for 48 hours to germinate. Germinated seeds were transplanted once radicles had emerged. All seedlings were started in 6-well 122.9 mL volume seeding trays filled with Sunshine #4 ™ soilless potting mix (SunGro Horticulture Ltd., Agawam, MA, USA) and transferred to 1L pots containing a 2:2:1 ratio of soil:sand:peat by volume once they had reached the three to four true-leaf stage. Plants were grown in a growth chamber under a 23°C 16-hour day/ 18°C 8-hour night cycle. Plant positions were assigned randomly and rotated on a weekly basis to account for potential differences in airflow, humidity and light intensity across benches within the growth chamber. Phenotypic traits measured can be broadly classified as capitulum size traits, traits relating to anthesis, traits associated with branching architecture, traits related to total plant size and traits related to petal morphology, colour and arrangement. Seed yield and size traits were not examined in this study due to the constraints of the testing location and ability to cross-pollinate all F_2 progeny examined.

5.3.3 Phenotypic analysis

Twenty traits were phenotyped on 341 F_2 plants (Table 5.1). As the parental and F_1 plants were grown under different conditions than the F_2 populations, complete phenotypic analysis was conducted solely on the F_2 generation. Phenotyping of the appearance of reproductive buds and date of first capitulum anthesis through fifth capitulum anthesis were conducted three times a week, while all other traits were assessed on the date of anthesis of the fifth capitulum. Branching characteristics were characterized as a percentage based upon total node count. All nodes were counted starting at the cotyledonary node moving upwards along the main stem. Percentage of branched nodes was measured as the total number of branches greater than 2cm in length relative to total node number. The highest branching node percentage was measured as the total number of unbranched nodes which were found above the highest branch on the apical portion of the stem and expressed as a percentage of total nodes. Lowest branching node percentage was the percentage of unbranched nodes relative to total node count on the basal region of the stem between the cotyledonary node and lowest branching node point. Shattering in wild *H. annuus* is in part due to the continued growth of the capitulum resulting in a convex shape (increased depth:width ratio). Domesticated *H. annuus* exhibits a low depth:width ratio, resulting in a flatter capitulum, less prone to shattering, with the depth:width ratio employed as a proxy for this trait (Burke et al., 2002). Floral variation for traits known in ornamental sunflowers were present and also examined in this study. Petal colour was rated on a 1-5 scale, ranging from near white (1) to deep gold in colour (5) on a progressive scale. The presence of multiple petal whorls was classified into three classes: normal presentation (one row); weakly doubled; and

doubled or chrysanthemum-type petal arrangement patterns (Supplemental Figure 5.2). Petal morphology was classified into three types:; normal (flat); mixed normal-tubular; and tubular (Supplemental Figure 5.3).

Data were analyzed for normality using the SAS UNIVARIATE procedure and base 10 logarithmic transformations were conducted when necessary to meet assumptions of normality for downstream analysis. Principal component analysis was conducted to explore trait relationships on centered and scaled data using the R function *PRCOMP*. Factor loadings were squared to calculate the percentage of variance the various components contributed to in the phenotypic data. The number of principal components to retain was determined by examining a breakdown of eigenvectors following Cattell's rule (Cattell, 1966) and Horn's parallel analysis (Horn, 1965) was run using 1,000 permutations in the R package *paran* (Dinno, 2012).

Table 5.1: Descriptions of 21 phenotypic traits measured under growth chamber conditions for 341 *H. maximiliani* F_2 plants under growth chamber conditions following a 23°C 16-hour day/ 18°C 8-hour night cycle.

Capitulum traits	**Description**
Diameter of capitulum 1-5	Average diameter of the first five capitulum to reach anthesis in mm
Depth of capitulum 1-5	Average depth of the first five capitula to reach anthesis in mm
Depth:width ratio	Average capitulum depth/width in mm
Size of central capitulum	Size of central capitulum in mm
Total capitula	Total number of capitula above a minimum size of 2 mm
Branching	
Length of branches 1-5	Average length of the first five apical branches above 2cm in in length measured in cm
Highest branching node	Percentage of apical branched nodes on the main stem with a branch length greater than 2 cm in length relative to total nodes
Lowest branching node	Percentage of basal branched nodes of the main stem with a branch length greater than 2 cm in length relative to total nodes
Percentage of branch bearing nodes	Percentage of branched nodes greater than 2 cm in length relative to total nodes
Plant size	
Total nodes	Total plant nodes apparent on the main stem
Stem diameter	Stem diameter measured at the first basal node in mm at maturity
Plant height	Plant height in cm
Length of leaf 1-5 at maturity	Length of first five leaves below the central capitulum in cm
Anthesis	
Emergence of apical buds	Number of days to the first appearance of reproductive buds
First anthesis	Number of days to first anthesis
Average anthesis for the first 5 capitula	Average days to first anthesis for the first five capitula
Petal characteristics	
Petal colour	Colour rating of petal colour from white to gold on a 1-5 scale
Petal length	Average length of five randomly sampled ray petals in mm
Petal morphology	Score of petal shape from 1-3: flat (1), mixed (2) or tubular (*tub*-like) (3)
Petal whorling	Score of ray flower arrangement from 1-3: single row (1), multiple row (weak *dbl*) (2) or chrysanthemum-like (*dbl*) (3)

5.3.4 DNA extraction

Leaf tissues from all plants grown under growth chamber conditions was sampled for DNA extraction. Samples from each plant were labeled with an identity number and frozen in liquid N within 1 h of collection prior to lyophilization and storage at room temperature. Genomic DNA was extracted from the parental, F_1 and a subset consisting of 190 F_2 individuals following the same protocol as described in Chapter 4. A modified single-tube cetyltrimethylammonium bromide (CTAB) extraction protocol (Doyle, 1987; Li et al., 2007) with 1μL of 10 mg mL^{-1} solution of proteinase K added to the initial CTAB incubation step (Promega). DNA quantity was determined using a dsDNA broad Range Fluorescence Assay kit and a Qubit 2.0 Fluorometer (Life Technologies) following the manufacturers' instructions for 1μL sample sizes. DNA quality was assessed using 260/280 and 260/230 nm wavelength absorbance ratios measured using a Nanodrop 1000 spectrophotometer (Fischer Scientific). Samples that fell below a minimum DNA concentration of 50 μg/μL or absorbance ratios below 1.7 for either 260/230 or 260/280 nm were discarded and extractions were repeated to meet quality requirements.

5.3.5 Genotype-by-sequencing and SNP calling

One-hundred and ninety-five DNA samples consisting of the Manitoba and Kansas parentals, F_1 parents and 190 F_2 plants were submitted to Data2bio LLC (Ames, Iowa, USA) for genotype-by-sequencing (GBS) and single nucleotide polymorphisms

(SNPs) calling. A tunable genotype-by-sequencing (tGBS) protocol was employed, which differs from conventional GBS through the use of a greater selective genome reduction procedure (Ott et al., 2017). Fewer sites are sequenced in tGBS relative to the conventional GBS, resulting in greater read depth. The advantage of this approach is that greater depth of coverage allows for more effective calling of heterozygous genotypes and reduces false homozygote calls, leading to a lower level of missing data per site.

Sequencing via tGBS was carried out using a three selective base (TGT) protocol as described in Ott et al. (2017) for genome reduction. Fragments were sequenced via eight runs on an Ion Proton^{TM} sequencer (Thermo Fisher Scientific, Waltham, MA). The conservative three selective base approach was chosen as *H. maximiliani* is an obligate outcrossing species likely to exhibit a high degree of heterozygosity. Trimmed reads were aligned to the cultivated *H. annuus* reference genome, HA412.v.1.1.bronze.20142015 (www.sunflowergenome.org), using GSNAP (Wu and Nacu, 2010) with polymorphisms within the first and last 3bp for the read being ignored (Ott et al., 2017). Polymorphisms with base calls with a PHRED score below 20 were removed from the analyses. The SNPs were called as homozygous if the most common allele was supported by a minimum of 80% aligned reads, while heterozygous SNPs were called if the two most common alleles were supported by a minimum of 30% of all aligned reads and the sum of the two most common alleles accounted for a minimum of 80% of the aligned reads. Single nucleotide polymorphisms that could be genotyped in a minimum of half of the samples, that contained an allele number of 2, a minor allele frequency >1% and a heterozygous rate of <70% were maintained for downstream analysis.

5.3.6 SNP filtering

The AB and BC crosses were examined jointly as a single population employing markers segregating in the same fashion between the two crosses. Common segregation patterns between the crosses were determined using the F_1 genotypes and observed segregation patterns in the F_2 generation to allow joint analysis. To allow for common coding of the crosses, the common parent F1B was designated as parent 1, while F1A and F1C were considered jointly as parent 2. Two SNP datasets were generated for linkage and association analysis separately and are herein referred to as the SNPLG (linkage mapping) and SNPAM (association mapping) datasets. To generate the SNPAM dataset SNPs were initially filtered to remove loci with the minor allele state represented in fewer than five individuals or a minor allele frequency below 10% and exhibiting no more than 50% missing data.

Most members of the family *Compositae* are believed to be derived from a paleohexaploid progenitor and the genus *Helianthus* is known to have undergone a genus specific whole-genome duplication event approximately 29 million years ago (Barker et al., 2008). The presence of paralogs may result in false linkages ("pseudo-linkages") between ancestrally homologous sequences, complicating linkage mapping efforts and reduces the power to detect marker-trait associations. To account for potential SNPs called from paralogous sequences as opposed to allelic differences, SNP flanking regions were compared to the sunflower reference genome using Blast2Go (www.blast2go.com). A high-stringency BLASTn analysis was used to determine to what degree called SNPs aligned to either a single position or to multiple positions in the *H. annuus* HA412.v.1.1.bronze.20142015 reference genome. The SNPs which aligned to more than

one position within the reference genome at an *E*-value cutoff of $<10^{-100}$ were removed from the dataset. The SNPLG dataset was then generated through selecting markers from the filtered SNPAM dataset in which F_1 genotypes were available to allow marker classification for linkage mapping.

5.3.7 Linkage map construction

Maximilian sunflower is an obligate-outcrossing species, therefore heterozygosity is present in the parental individuals and F_1 populations may show a mixture of homozygous and heterozygous markers. Multiple methods using both maximum likelihood-based and pseudo-testcross-based strategies were examined for linkage map construction including using *ASMap* (Taylor and Butler, 2017), r/*QTL* (Broman et al., 2003), *Onemap* (Margarido et al., 2007), MapDisto (Lorieux, 2012) and TMAP (Cartwright et al., 2007) using both reference-genome guided and LOD score based approaches. Final maps were constructed using a pseudo-testcross approach (Grattapaglia and Sederoff, 1994) in *ASMap*. Markers were classified into three segregation types based upon either two-way pseudo-test backcross (1:1) or F_2-like (1:2:1) segregation patterns following the nomenclature outlined by Wu et al. (2002) for outbred pedigrees which takes into account parental allele patterns and null (presence/absence) alleles. The letter (A, B, C or D) denotes the observed segregation pattern, followed by the first number (1-3) that denotes the null allele pattern. The third number which follows the period (1-18) corresponds to allelic composition of the parents contributing to the cross. As bi-alleleic SNPs were generated using tGBS, three segregation patterns were observed corresponding to D2.15, D1.10 and B3.7 patterns.

D2.15 markers consisted of markers corresponding to a backcross to F_1 parent 1 (*aa* x *ab*), D1.10 corresponded to backcross to F_1 parent 2 (*ab* x *aa*) and B3.7 corresponds to markers exhibiting a F_2-like or intercross marker pattern in which both F_1 parents are heterozygous (*ab* x *ab*). The pseudo-testcross mapping approach required the analysis of backcross-like markers segregating as separate datasets using an approach akin to map construction in backcross or doubled-haploid populations. This resulted in separate genetic maps corresponding to segregation observed in either the first parent F1B (D1.10) or "second parent" F1A and F1C (D2.15).

To infer linkage phase and detect repulsion phase linkages, the D1.10 and D2.15 backcross-like markers were duplicated and recoded as "mirror image" markers (homozygous calls being recoded as heterozygous and vice versa) and combined with the original markers (Grattapaglia and Sederoff, 1994; Lewis and Sink, 1996; Nishio et al., 2013; Barb et al., 2014; van Heerden et al., 2014). This method produces two identical maps for each linkage group corresponding to markers in opposite phases from which marker position in then inferred. Homozygous allele states accounting less than 10% of the observed total calls were considered likely genotyping errors and set to missing data prior to analysis. B3.7 (*ab* x *ab*) class markers and those in which the either the parental genotype could not be accurately determined or were non-informative (*aa* x *aa*, *aa x bb*, *bb* x *aa*, *bb* x *bb*), were removed. The remaining markers were tested for distorted segregation using a chi-square goodness-of-fit test through the test segregation function of *ASMap* at $p < 0.05$ for downstream analysis. Due to the presence of a high degree of segregation distortion, markers exhibiting distortion were retained within the dataset and grouped based upon their assigned linkage groups in the cultivated *H. annuus* reference

genome for ordering. Segregation distortion has been reported as having little effect on marker order (Zhang et al., 2010), but erroneously called genotypes, which may result in the appearance of highly-distorted markers, may cause pseudo-linkages between biologically unlinked sequences (Ronin et al., 2010). Patterns of segregation distortion were examined along the length of the maps and single markers that showed strong segregation distortion relative to neighboring markers, indicative of possible genotyping errors, were removed and maps lengths were recalculated.

Linkage maps were constructed separately for the D1.10 and D2.15 marker classes using the minimum spanning tree approach (Wu et al., 2008) implemented in the R program *ASMap*. Markers were initially ordered within linkage groups based upon a minimum LOD score of 3 and a maximum recombination fraction of 0.3 with the maximum likelihood option invoked for marker ordering. Groups consisting of greater than 20 markers were considered as separate linkage groups for imputation. The initial maps were exported to Maskov 1.01 (Ward et al., 2013) for map-based imputation using parameters described in Ward et al. (2013) (max errors=5, threshold=5). Maps distances were calculated again in *ASMap* following imputation using Kosambi's method (Kosambi, 1943). Imputed markers were ordered using a minimum LOD score of 3, again using the maximum likelihood approach. The detectBadData function of *ASMap* was used to set remaining putative genotyping errors to missing data. Heatmaps were examined to determine the presence of putative translocations between the reference genome linkage groups and to combine small "fragment" linkage groups to existing groups.

5.3.8 QTL analysis

QTL analysis was conducted at 1 cM intervals using an interval mapping approach in the R program r/*QTL* on the D1.10 and D2.15 maps separately. One-way genome scans were run by trait using the *scanone* function to identify putative QTL employing the Haley-Knott regression (Haley and Knott, 1992). The QTL analysis was conducted using three models, a naïve model employing no covariates, one in which the F_1 maternal genotype served as an additive covariate and one in which the F_1 maternal genotype was considered as both an additive and interactive covariate. To account for multiple testing and control of family-wise-error rate for type I errors, whole-genome and chromosome-specific LOD thresholds were calculated for each trait independently (Churchill and Doerge, 1994). Appropriate LOD threshold values were estimated using 10,000 permutations and an alpha threshold of 0.05 at both the whole-genome and chromosome levels. Quantitative trait loci falling above a minimum LOD of 3 and surpassing either the chromosome or genome-wide level LOD thresholds were considered significant QTL. Putative QTL that fell below a LOD of 3 but above the minimum LOD thresholds at either the chromosome or genome-wide levels were considered as suggestive QTL.

LOD scores were examined visually by trait amongst linkage groups to determine the presence of secondary LOD peaks that may inflate QTL confidence intervals. Markers corresponding to the secondary LOD peak positions were included as a covariate when necessary to refine QTL position in instances of inflation. Analysis of variance was conducted for each trait QTL via the *fitqtl* function and single QTL models were fit. In instances where maternal genotype, or the interaction of maternal genotype with

phenotype, were non-significant they were dropped as co-factors from the model and the percent variance contributed to QTL effects were calculated. Confidence intervals were calculated using the one LOD drop-off method which approximates a 96.8% confidence interval of QTL position (Lander and Botstein, 1989) via the *lodint* function of r/*QTL*.

5.3.9 Association analysis

Single marker-trait association analysis was conducted between the phenotypic dataset and the ALLSNP dataset in TASSEL 5.0 (Bradbury et al., 2007) to take advantage of markers that could not be incorporated into the linkage maps. A benefit of this approach is that all marker classes including intercross markers (class B3.7) may be analyzed allowing for the detection of QTL exhibiting recessive inheritance patterns. Five models were run and *q*-*q* plots were examined to select the model exhibiting the closest fit to expected *p*-value distributions. Models included three general linear models (GLM) and two mixed linear models (MLM). The GLM models included a model with no co-factors, a GLM incorporating F_1 family as a fixed effect, and a GLM incorporating F_1 and kinship (*K*) scores as fixed effects. The MLM included a model incorporating F_1 family as a fixed effect and kinship as a random effect, and an MLM incorporating solely kinship as a random effect. Kinship values were calculated using a subset of markers exhibiting no more than 10% missing data in TASSEL prior to analysis. A threshold of $p<0.001$, corresponding to a LOD score of 3, a common threshold in genetic mapping studies for ensuring an overall false positive (type I error) of 5% (Lander and Botstein, 1989) was selected to denote significant associations for the MLMs and to further guard against the risk of false positives. Putative QTL were declared when supported by a

minimum of two SNPs within close proximity (>200 bp) in the cultivated *H. annuus* reference genome. Permutations were run for the GLMs using 10,000 permutations in TASSEL at both the chromosome and genome wide levels. Estimates of the proportion of phenotypic variance explained were calculated in TASSEL for each SNP-trait association and expressed as r^2 values.

5.3.10 SNP annotation

To infer candidate genes and provide functional annotation of the candidate SNPs generated from the linkage and association mapping approaches, *H. maximiliani* SNP reads were compared with the cultivated *H. annuus* reference genome HA412.v.1.1.bronze.20142015 assembly using Jbrowse (Skinner et al., 2009), available via the HeliaGene bioinformatics portal (Carrere et al., 2008). The SNP flanking regions were compared to *Arabidopsis thaliana* (L.) Heynh. TAIR10 gene model database available via the *Arabidopsis* genome database (www.arabidopsis.org) using BLASTn with an *E*-value cutoff $<10^{-5}$. *H. annuus* and *A. thaliana* gene identity and interpro annotations were gathered using Jbrowse to support putative *H. maximiliani* gene function. Flanking sequences of SNPs with no known gene in the *H. annuus* or *A. thaliana reference* genomes were examined to provide putative protein function by searching the UniProtKB database (The UniProt Consortium, 2017) using BLASTx an *E*-value cutoff $<10^{-5}$ via the program Blast2GO (www.blast2go.com).

5.4 Results

5.4.1 Genotype-by-sequencing and SNP calling

A total of 33,608 SNPs called by Data2Bio met the minimum quality requirements of being genotyped in at least 50% of the 195 samples, and having a genotype number of ≥2, an allele number equal to 2, a minor allele frequency of ≥1% and a heterozygous rate of ≤ 70%. The distributions of SNPs called by linkage group are listed in Supplemental Table 5.1. Following filtering (section 5.3.6 SNP filtering), 10,144 SNPs remained in the SNPAM dataset while 4,755 SNPs were retained in the SNPLG dataset for analysis. The unordered contig of *H. annuus* Ha0_73Ns contained the largest number of SNPs in the unfiltered dataset (3,378) while the fewest SNPs were aligned to linkage groups 6 (679) and 7 (889) relative to the other linkage groups, which contained between 1,440-2,876 SNPs (Supplemental Table 5.1).

5.4.2 Phenotypic analysis

The range of phenotypes observed amongst both crosses was extensive and indicates that the majority of examined traits are quantitative in nature (Table 5.2). The unbranched phenotype observed in the Kansas parent was not recovered in the F_1 or F_2 generations though a number of plants with highly restricted branching were observed in the F_2 generation. Branch number ranged from one to 21 branches and the percentage of branch bearing nodes ranged from 2.38 to 48.97%. Timing of first anthesis range was from 61 to 198 days. Plant height ranged from 90 cm to 192 cm and diameter of the central capitulum ranged from 6.2 mm to 25.6 mm, approaching approximate values

observed in the parental materials (Supplemental Table 5.2, Supplemental Figure 5.1). The total number of capitula showed the highest coefficient of variation, indicative of a high degree of plasticity relative to the other examined traits (Table 5.2). A range of petal colours were observed, ranging from off-white to a deep golden colour (non-parental). Variation in petal arrangement and morphology was not observed in the parental materials or the F_1 generation, but became apparent in the F_2 generation, including the appearance of tubular ray florets (Supplemental Figure 5.3) and both doubled-flowered and chrysanthemum type petal whorling patterns (Supplemental Figure 5.2).

Table 5.2: Phenotypic mean, standard deviations (SD), range of values and coefficient of variation (CV) observed amongst 341 F_2 *H. maximiliani* plants phenotyped for 21 traits under growth chamber conditions following a 23°C 16-hour day/ 18°C 8-hour night cycle.

Capitulum traits	Mean	SD	Min	Max	Range	CV
Diameter of capitula 1-5 (mm)	12.60	2.54	5.80	19.98	14.18	20%
Depth of capitula 1-5 (mm)	11.80	1.81	5.08	17.12	12.04	15%
Capitulum depth:width ratio	0.95	0.10	0.71	1.62	0.90	11%
Size of central capitulum (mm)	13.89	3.38	6.20	25.60	19.40	24%
Total capitula count	39.64	22.36	4.00	126.00	122.00	56%
Branching						
Average length of branches 1-5 (cm)	29.41	9.67	3.80	61.60	57.80	33%
Highest branching node (%)	5.60	2.24	1.00	16.00	15.00	40%
Lowest branching node (%)	36.28	11.01	5.00	73.00	68.00	30%
Percentage of branch bearing nodes (%)	16.95	6.55	2.38	51.35	48.97	39%
Plant size						
Total nodes	50.96	10.74	25.00	92.00	67.00	21%
Stem diameter (mm)	9.53	1.31	6.10	14.60	8.50	14%
Plant height (cm)	142.11	20.61	90.00	192.00	102.00	15%
Length of leaf 1-5 at maturity (cm)	12.53	3.89	0.00	22.10	22.10	31%
Anthesis						
Emergence of reproductive buds (days)	67.44	14.28	30.00	112.00	82.00	21%
First anthesis (days)	118.37	24.69	61.00	198.00	137.00	21%
Average anthesis for the first 5 capitula (days)	133.00	24.53	66.40	201.20	134.80	18%
Petal characteristics						
Petal colour (1-5 scale)	3.42	1.15	1.00	5.00	4.00	34%
Petal length (mm)	36.35	9.09	10.20	58.30	48.10	25%
Petal shape (1-3 scale)	1.77	0.58	1.00	3.00	3.00	33%
Petal whirling (1-3 scale)	1.61	0.74	1.00	3.00	3.00	46%

Note: comparison of F_2 populations and Manitoba parental population is available in Supplemental Table 5.2.

Principal component analysis revealed positive correlations along the first principal component for timing of anthesis and capitulum size related traits and negative correlations for the percentage of branched nodes, capitulum depth:width ratio and total capitula count. The opposing directions of these traits along the first principal component axis show negative associations between these traits indicative of potential tradeoffs (Figure 5.1, Supplemental Table 5.3). Tradeoffs were also observed between the

percentage of branched nodes and total capitula count which were negatively loaded and traits related to capitulum size and timing of anthesis, which were positively loaded. The second axis revealed different relationships to the first for some trait pairs, with a negative relationship between timing of anthesis and capitulum size traits and between the percentage of branched nodes and total capitula count.

Relationships between trait pairs and the amount of variation explained for each trait differed between principal components indicating different trade-off patterns. The first principal component axis which accounted for 26.77% of the variation in the dataset explained 11.1% of the variation in average capitulum size, 8.0% of the variation in central capitulum size, 8.5% of the variation in the percentage of branched nodes, 3.8% of the variation in total capitulum count and 10.2%, 5%, and 5.6% of the variation in the timing of reproductive budding, first anthesis, and average anthesis, respectively. The second principal component axis explained 16.83% of the variation in the dataset. The second axis explained 16.5% and 16.2% of the variation in first and average timing of anthesis, 10.5% of the variation in central capitulum size, 7.9% of the variation in average capitulum size and 1.4% of the variation in the percentage of branched nodes.

Associations between the traits first anthesis, average anthesis and timing of reproductive budding were observed along the first two principal component axes, as were associations between the traits central capitulum size, central capitulum depth and average capitulum size (Figure 5.1, Supplemental Table 5.3). Patterns indicative of trade-offs were supported by the first two principal components with negative associations between capitulum size traits and the capitulum depth:width ratio and total capitulum count. Jointly, the first two principal component axes explain 20.1%, 16.7% and 19% of

the variation in central capitulum size, central capitulum depth and average capitulum size and 10.2% and 9.5% of the variation in total capitulum count and capitulum depth:width ratio, respectively, indicating that the observed patterns are not 1:1 trade-off.

The remaining significant principal components each explained less than 10% of the total variation. While the percentage of total variation explained by these components is low, the amount of variation they explain for different traits is notable, and could indicate major genes. The third through fifth significant principal components explained 8.25%, 7.37% and 6.29% of the variation in the dataset respectively. Notably, the third principal component explained 41.8% of the variation in the position of the highest branching node while the fourth principal component explained 25.1% of the variation in stem diameter (Supplemental Table 5.3).

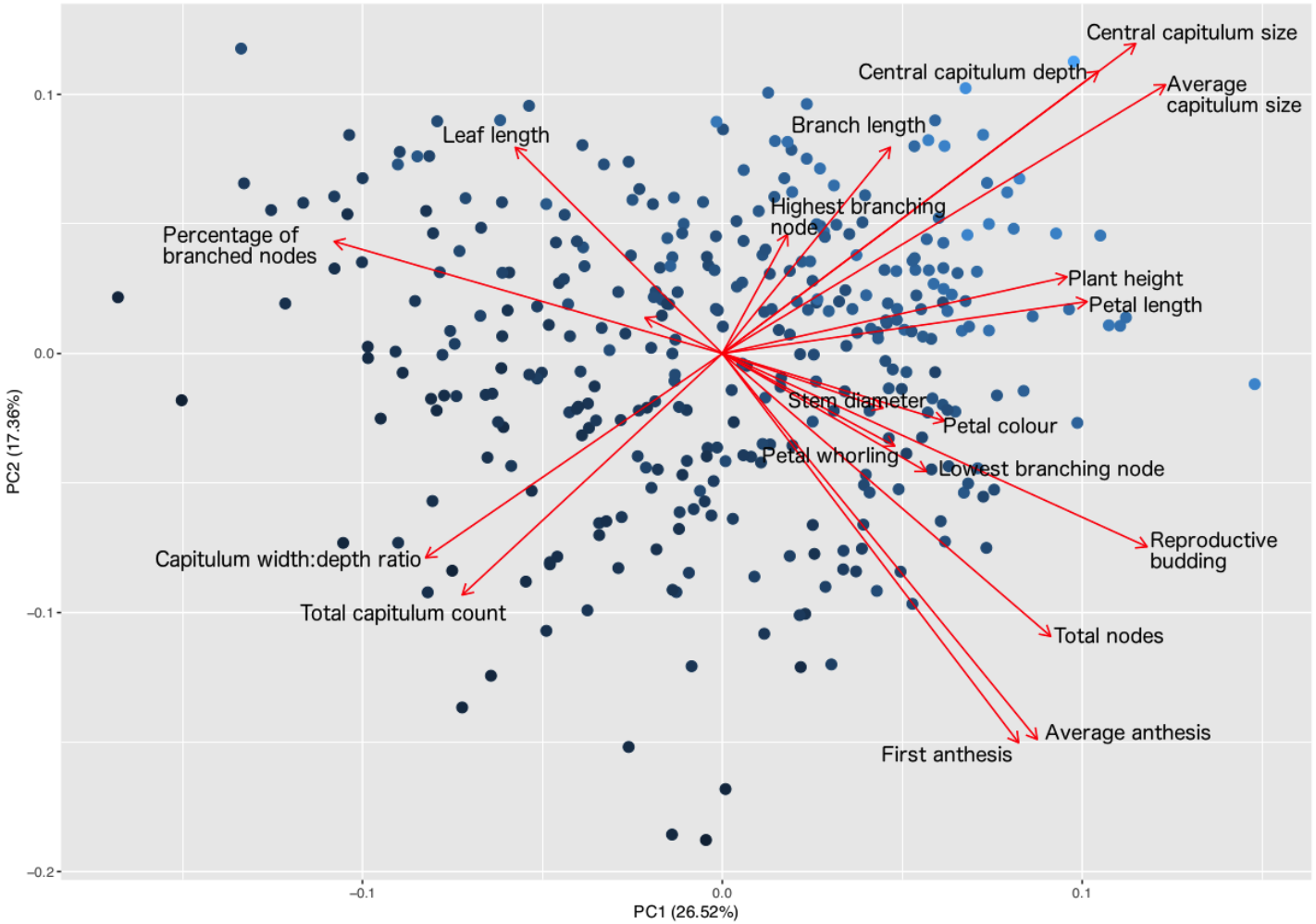


Figure 5.1: First and second principal component axes of 20 phenotypic measurements taken on 341 *H. maximiliani* plants under growth chamber conditions following a 23°C 16-hour day/ 18°C 8-hour night cycle.

5.4.3 Linkage map of Maximilian sunflower using *H. annuus* as a reference guide

Maps generated through the sole use of recombination fractions and LOD score information incorporated few markers (~800) relative to the maps generated using the *H. annuus* reference genome linkage groups as a guide (~2,000), therefore maps generated using the reference genome approach were retained. Attempts to combine D1.10, D2.15 and B3.7 marker types using a maximum-likelihood approach in *Onemap* resulted in maps with considerable regions of inflation (data not shown), therefore maps developed in *ASMap* were selected for QTL analysis and reporting. The D1.10 map (Figure 5.2) spanned a total of 4,530.0 cM with an average marker density of 2.23 SNPs per cM while the D2.15 map (Figure 5.3) spanned 4,109.85 cM with an average marker density of 2.35 SNPs per cM (Table 5.3). A high degree of segregation distortion was observed in both linkage maps with multiple smooth peaks being observed amongst all linkage groups (data not shown). Distortion patterns along all chromosomes is indicative of systematic factors such as consanguinity and resulting linkage disequilibrium as opposed to major loci of influence are resulting in the segregation distortion. In both maps, the ratio of homozygous to heterozygous alleles favoured the homozygous state, particularly the D1.10 map in which 68% of the scored alleles were homozygous, while 55.7% of the alleles in the D2.15 map were homozygous.

Table 5.3: Summary statistics by linkage group for linkage maps generated using D1.10 and D2.15 marker classes segregating amongst 190 F_2 *H. maximiliani* individuals.

	D1.10 map			D2.15 map		
Linkage group	cM	Total SNPs	SNPs/cM	cM	Total SNPs	SNPs/cM
1	224.84	86	2.61	276.55	124	2.23
2	369.68	154	2.40	287.04	115	2.50
3	295.13	132	2.24	222.64	99	2.25
4	260.45	137	1.90	145.29	78	1.86
5	448.05	184	2.44	316.28	123	2.57
6	107.9	44	2.45	83.47	41	2.04
7	95.66	49	1.95	101.32	50	2.03
8	327.61	145	2.26	314.78	125	2.52
9	110.11	81	1.36	414.3	157	2.64
10	216.54	121	1.79	498.08	184	2.71
11	299.7	122	2.46	177.33	89	1.99
12	220.87	98	2.25	211.52	80	2.64
13	368.83	145	2.54	176.11	87	2.02
14	348.35	145	2.40	161.57	73	2.21
15	363.61	141	2.58	178.69	92	1.94
16	165.68	85	1.95	226.52	83	2.73
17	306.99	163	1.88	318.36	152	2.09
Total	4,530	2,032	2.23	4,109.85	1,752	2.35

5.4.4 Association analysis

The MLM and GLM approaches produced similar *p*-value distributions, both indicating adequate control of false positive associations (data not shown). As the MLM approaches underfit the data for some traits (indicative of overly stringent control and the potential for false negative results), for the purposes of identifying candidate SNPs underlying traits of interest, the results of both the MLM + K + F_1 family and GLM + K +F_1 family models were selected for reporting. In the GLM, 41 SNPs corresponding to 34 regions in the cultivated *H. annuus* reference genome showed significant trait associations at the chromosome level **(**Table 5.4). Three SNPs were significant at the

genome-wide level: *Ha0-283128806*, which explained 14% of the variation in timing of average anthesis, *Ha15-4372179* and *Ha1-31234022* which explained 22.3% and 15.5% of the variation, respectively, in the position of the highest branching node. In the MLM analysis a total of 147 SNP-trait associations were found to be significant at $p>0.001$, of which phenotypic associations in 23 regions were supported by a minimum of 2 SNPs within close proximity (Table 5.5). Twenty-one of the SNP-trait associations were found to be significant in both GLM and MLM models, showing SNP associations for the traits timing of reproductive budding, average time timing of anthesis, percentage of branched nodes, total nodes and petal morphology.

Table 5.4: SNPs exhibiting significant associations to phenotype using association mapping (GLM) following permutation testing at the chromosome level.

Trait	Marker	*H. annuus* LG	Position (bp)	*P*-value	Perm. *p*-value	Marker r^2
Average Anthesis	*Ha0-283128806*	HA0‡	283,128,806	3.15×10^{-06}*	0.003	0.140
	Ha12-6455535	HA12	6,455,535	4.31×10^{-05}	0.016	0.153
	Ha14-158966313†	HA14	158,966,313	6.23×10^{-05}	0.022	0.132
	Ha14-158966327†	HA14	158,966,327	5.54×10^{-05}	0.020	0.134
	Ha15-6924663	HA15	6,924,663	2.46×10^{-05}	0.009	0.119
Branch Length	*Ha0-72910708*	HA0	72,910,708	6.63×10^{-05}	0.049	0.093
First Anthesis	*Ha0-283128806*	HA0	283,128,806	7.12×10^{-06}	0.006	0.131
	Ha12-6455535	HA12	6,455,535	1.39×10^{-04}	0.048	0.134
	Ha15-6924663	HA15	6,924,663	5.50×10^{-05}	0.020	0.110
Highest branching node	*Ha1-31234022*	HA1	31,234,022	9.50×10^{-07}*	0.001	0.155
	Ha15-4372179	HA15	4,372,179	3.80×10^{-06}*	0.002	0.223
	Ha5-145356067†	HA5	145,356,067	6.17×10^{-05}	0.035	0.167
Lowest branching node	*Ha17-193706040*	HA17	193,706,040	8.69×10^{-06}	0.005	0.191
	Ha4-165430726†	HA4	165,430,726	7.63×10^{-05}	0.029	0.131
Leaf Length	*Ha15-4372229*†	HA15	4,372,229	2.17×10^{-05}	0.008	0.130
Percent Branched	*Ha1-107557612*†	HA1	107,557,612	1.45×10^{-04}	0.047	0.120
	Ha1-107557623†	HA1	107,557,623	1.37×10^{-04}	0.044	0.119
	Ha13-54522410†	HA13	54,522,410	6.04×10^{-05}	0.028	0.130
	Ha14-63973234	HA14	63,973,234	1.24×10^{-04}	0.043	0.160
	Ha15-4372275†	HA15	4,372,275	1.15×10^{-04}	0.041	0.092

Trait	Marker	*H. annuus* LG	Position (bp)	*P*-value	Perm. *p*-value	Marker r^2
Petal Colour	*Ha13-178942791*†	HA13	178,942,791	8.61×10^{-06}	0.004	0.123
	Ha16-19120005	HA16	19,120,005	2.37×10^{-05}	0.009	0.144
	Ha8-6754280	HA8	6,754,280	8.98×10^{-05}	0.038	0.133
Petal Morphology	*Ha5-132920307*	HA5	132,920,307	4.95×10^{-05}	0.027	0.116
	Ha5-150754552†	HA5	150,754,552	4.81×10^{-05}	0.026	0.155
	Ha5-150754563†	HA5	150,754,563	5.29×10^{-05}	0.029	0.152
	Ha5-150754597†	HA5	150,754,597	5.42×10^{-05}	0.030	0.152
	Ha5-150754605†	HA5	150,754,605	7.38×10^{-05}	0.041	0.147
	Ha5-150754617†	HA5	150,754,617	5.42×10^{-05}	0.030	0.152
	Ha5-155001315	HA5	155,001,315	5.96×10^{-05}	0.033	0.163
	Ha5-160952967†	HA5	160,952,967	2.19×10^{-05}	0.012	0.141
	Ha5-160953020†	HA5	160,953,020	2.19×10^{-05}	0.012	0.141
	Ha5-160953053†	HA5	160,953,053	2.49×10^{-05}	0.014	0.140
	Ha5-160953056†	HA5	160,953,056	2.19×10^{-05}	0.012	0.141
Petal Whorling	*Ha1-120468303*	HA1	120,468,303	9.60×10^{-05}	0.029	0.119
Reproductive Budding	*Ha13-24088571*†	HA13	24,088,571	2.41×10^{-05}	0.011	0.129
	Ha13-24088583†	HA13	24,088,583	9.94×10^{-05}	0.045	0.107
Depth:width Ratio	*Ha17-261238188*	HA17	261,238,188	5.52×10^{-05}	0.029	0.117
Total Nodes	*Ha14-34533074*†	HA14	34,533,074	5.94×10^{-05}	0.021	0.171
	Ha14-34533101†	HA14	34,533,101	4.08×10^{-05}	0.014	0.160
	Ha4-430240	HA4	430,240	2.63×10^{-05}	0.010	0.134

Note: * Denotes significance at the genome-wide permutation threshold, alpha=0.05

†: Significant SNP-trait association supported in both GLM and MLM analysis (Table 5.5).

‡HA0 refers to HA0_73Ns, the unordered contig of the *H. annuus* HA412.v.1.1.bronze.20142015 reference genome.

Table 5.5: SNPs exhibiting significant associations at $p > 0.001$ to phenotype using association mapping (MLM) and supported by a minimum of two SNPs within close proximity (>200bp).

Trait	Marker	*H. annuus* LG	Position (bp)	*P*-value	Marker r^2
Average Anthesis	*Ha14-158966298*	HA14	158,966,298	6.63×10^{-04}	0.112
	Ha14-158966313†	HA14	158,966,313	3.75×10^{-04}	0.123
	Ha14-158966327†	HA14	158,966,327	3.07×10^{-04}	0.127
Central Capitulum Size	*Ha13-62416803*	HA13	62,416,803	7.63×10^{-04}	0.070
	Ha13-62416862	HA13	62,416,862	3.75×10^{-04}	0.076
	Ha5-160670593	HA5	160,670,593	8.64×10^{-04}	0.107
	Ha5-160670596	HA5	160,670,596	5.97×10^{-04}	0.122
Highest branching node	*Ha5-145356033*	HA5	145,356,033	4.23×10^{-04}	0.135
	Ha5-145356067†	HA5	145,356,067	3.28×10^{-04}	0.167
Lowest branching node	*Ha15-195617056*	HA15	195,617,056	5.04×10^{-04}	0.121
	Ha15-195617065	HA15	195,617,065	5.04×10^{-04}	0.121
	Ha4-165430726†	HA4	165,430,726	1.74×10^{-04}	0.120
	Ha4-165430836	HA4	165,430,836	6.84×10^{-04}	0.100
	Ha9-3116414	HA9	3,116,414	8.48×10^{-04}	0.140
	Ha9-3116423	HA9	3,116,423	7.90×10^{-04}	0.142
Leaf Length	*Ha15-4372229*†	HA15	4,372,229	1.00×10^{-04}	0.133
	Ha15-4372265	HA15	4,372,265	6.51×10^{-04}	0.099
	Ha15-6924673	HA15	6,924,673	9.22×10^{-04}	0.080
	Ha15-6924688	HA15	6,924,688	7.57×10^{-04}	0.083
	Ha4-177393919	HA4	177,393,919	3.76×10^{-04}	0.100
	Ha4-177393987	HA4	177,393,987	5.94×10^{-04}	0.095
	Ha5-198744253	HA5	198,744,253	9.56×10^{-04}	0.073
	Ha5-198744277	HA5	198,744,277	9.56×10^{-04}	0.073
Percent Branched	*Ha1-107557612*†	HA1	107,557,612	2.65×10^{-04}	0.117

Trait	Marker	*H. annuus* LG	Position (bp)	*P*-value	Marker r^2
	Ha1-107557623†	HA1	107,557,623	2.78×10^{-04}	0.115
	Ha13-54522410†	HA13	54,522,410	1.14×10^{-04}	0.126
	Ha13-54522440	HA13	54,522,440	3.97×10^{-04}	0.102
	Ha15-4372179	HA15	4,372,179	1.86×10^{-04}	0.161
	Ha15-4372229	HA15	4,372,229	2.99×10^{-04}	0.110
	Ha15-4372275†	HA15	4,372,275	1.37×10^{-04}	0.090
	Ha8-92972191	HA8	92,972,191	4.90×10^{-04}	0.090
	Ha8-92972203	HA8	92,972,203	4.88×10^{-04}	0.088
Petal Colour	*Ha13-178942784*	HA13	178,942,784	6.11×10^{-04}	0.074
	Ha13-178942791†	HA13	178,942,791	5.04×10^{-05}	0.109
Petal Morphology	*Ha5-150754539*	HA5	150,754,539	6.15×10^{-04}	0.142
	Ha5-150754552†	HA5	150,754,552	3.37×10^{-04}	0.155
	Ha5-150754563†	HA5	150,754,563	3.68×10^{-04}	0.152
Petal Morphology	*Ha5-150754597*†	HA5	150,754,597	3.79×10^{-04}	0.152
	Ha5-150754605†	HA5	150,754,605	4.64×10^{-04}	0.147
	Ha5-150754617†	HA5	150,754,617	3.79×10^{-04}	0.152
	Ha5-160952967†	HA5	160,952,967	7.63×10^{-05}	0.141
	Ha5-160953020†	HA5	160,953,020	7.63×10^{-05}	0.141
	Ha5-160953053†	HA5	160,953,053	9.18×10^{-05}	0.140
	Ha5-160953056†	HA5	160,953,056	7.63×10^{-05}	0.141
Reproductive Budding	*Ha13-24088571*†	HA13	24,088,571	8.42×10^{-05}	0.132
	Ha13-24088583†	HA13	24,088,583	2.63×10^{-04}	0.107
Depth:width Ratio	*Ha17-212455484*	HA17	212,455,484	4.17×10^{-04}	0.104
	Ha17-212455507	HA17	212,455,507	7.63×10^{-04}	0.084
Total Nodes	*Ha0_181488072*	HA0	181,488,072	6.31×10^{-04}	0.080
	Ha0_181488079	HA0	181,488,079	9.95×10^{-04}	0.076

Trait	Marker	*H. annuus* LG	Position (bp)	*P*-value	Marker r^2
	Ha14-34533074†	HA14	34,533,074	1.08×10^{-04}	0.173
	Ha14-34533077	HA14	34,533,077	4.12×10^{-04}	0.139
	Ha14-34533101†	HA14	34,533,101	1.04×10^{-04}	0.160
	Ha14-158966313	HA14	158,966,313	7.61×10^{-04}	0.088
	Ha14-158966327	HA14	158,966,327	5.66×10^{-04}	0.093

†: SNP-trait association supported by both GLM (Table 5.4) and MLM analysis.

‡HA0 refers to HA0_73Ns, the unordered contig of the *H. annuus* HA412.v.1.1.bronze.20142015 reference genome

5.4.5 QTL analysis

QTL analysis of the 190 F_2 individuals identified 29 and 3 QTL at the chromosome and genome-wide significance levels, respectively. The number of QTL per trait ranged from 1 to 4. In the D1.10 map, QTL were uncovered for the traits branch length, capitulum depth:width ratio, central capitulum depth, highest branching node, percentage of branched nodes, petal length, total capitulum count and total nodes (Table 5.6, figure 5.2). In the D2.15 map, QTL were found for the traits branch length, capitulum depth:width ratio, central capitulum depth, highest branching node, leaf length, lowest branching node, percentage of branched nodes, petal colour, reproductive budding, stem diameter and total capitulum count (Table 5.7, figure 5.3). An additional 18 putative QTL with LOD scores <3 were identified as significant at the chromosome level (Supplemental Tables 5.4-5.5). No QTL with a LOD score >3 were detected for the traits first anthesis, average anthesis, petal morphology, petal whorling, central capitulum size, average capitulum size and plant height. Few QTL were significant at the genome-wide significance threshold of $P<0.05$. The QTL *CWD1*, explaining 8.94% of the variation in the capitulum depth:width ratio and *TN1* explaining 13.32% of the variation in total node count, were significant at the genome-wide level in the D1.10 map. The QTL *RB1*, explaining 8.76% of the variation in the timing of reproductive budding, was significant at the genome-wide level in the D2.15 map. The QTL that explained the most variation (14.74%) was *LBN1* for the position of the lowest branching node, found on linkage group 5 of the D2.15 map. The QTL *PB1*, located on linkage group 14 of the D1.10 map, explained 13.95% of the variation in the percentage of branched nodes.

LBN2, found on linkage group 6 of the D2.15 map, explained 13.82% of the variation in the position of the lowest branching node.

A number of overlapping QTL were declared suggesting common loci influencing multiple traits. Overlapping QTL positions were detected in the D1.10 map include QTL for total nodes (*TN2*) and percentage of branched nodes (*PB3*) on linkage group 8 and central capitulum depth (*CCD2*) and branch length (*BL3*) on linkage group 13. In the D2.15 map overlapping QTL for stem diameter (*SD2*), total capitula count (*TCC4*) and average leaf length (*LL1*) were found on linkage group 12, while overlapping QTL for branch length (*BL2*) and stem diameter (*SD1*) were observed on linkage group 16.

Table 5.6: QTL supported by a LOD score >3 and surpassing an alpha threshold of 0.05 at the chromosome level in the D1.10 *H. maximiliani* linkage map.

Trait	QTL name	Nearest Marker	Linkage group	Peak (cM) †	CI (cM) ‡	%Var. §	LOD
Branch length	*BL1*	Ha15_184996368	15	334.0	333.0-335.0	8.87	4.08
Branch length	*BL3*	Ha13_45947027	13	223.0	215.0-229.0	7.62	3.27
Capitulum depth:width ratio	*CWD1*	Ha13_151824966	13	155.0	147.0-169.0	8.94	3.87*
Central capitulum depth	*CCD2*	Ha13_168139407	13	215.0	214.0-217.0	7.54	3.24
Highest branching node	*HBN1*	Ha8_164660018	8	263.0	262.0-264.39	11.39	4.79
Highest branching node	*HBN2*	Ha17_153743998	17	245.0	244.0-248.0	10.42	4.54
Percent branched	*PB1*	Ha14_189902425	14	141.0	140.52-142.0	13.95	6.32
Percent branched	*PB3*	Ha8_128549448	8	75.0	74.14-76.0	9.64	4.19
Percent branched	*PB4*	Ha1_57733568	1	190.0	178.1-196.0	7.37	3.16
Petal Length	*PL1*	Ha16_140663319	16	74.0	69.0-81.0	7.01	3.0
Total capitulum count	*TCC1*	Ha2_70725956	2	266.0	254.0-273.0	10.46	4.68
Total capitulum count	*TCC2*	Ha5_139647411	5	140.0	138.27-142.0	8.8	4.0
Total capitulum count	*TCC3*	Ha11_37687444	11	51.0	46.0-56.0	8.18	3.75
Total nodes	*TN1*	Ha3_138518462	3	195.0	194.0-197.0	13.32	5.9*
Total nodes	*TN2*	Ha8_41103619	8	88.0	60.0-89.0	7.64	3.28

Note: * Significant at alpha=0.05 at the genome wide level

† cM =Centimorgan

‡CI = LOD-1 confidence interval

§ %Var. =proportion of phenotypic variance explained by QTL (equivalent to marker r^2 x 100)

Table 5.7: QTL supported by a LOD score >3 and surpassing an alpha threshold of 0.05 at the chromosome level in the D2.15 *H. maximiliani* linkage map.

Trait	QTL name	Nearest Marker	Linkage group	Peak (cM) †	CI (cM) ‡	%Var. §	LOD
Branch length	*BL2*	Ha16_198498442	16	100.0	99.0-101.0	7.23	3.29
Capitulum depth:width ratio	*CWD2*	Ha2_120253566	2	145.0	144.0-148.0	8.62	3.74
Central capitulum depth	*CCD1*	Ha9_90319903	9	136.0	131.0-141.0	7.46	3.35
Highest branching node	*HBN3*	Ha1_82619781	1	243.0	242.0-244.36	8.06	3.47
Leaf length	*LL1*	Ha11_203269770	11	138.0	109.0-143.0	8.86	4.0
Lowest branching node	*LBN1*	Ha5_72184303	5	63.0	62.36-64.0	14.75	6.68
Lowest branching node	*LBN2*	Ha6_49969337	6	60.0	58.0-61.0	13.82	6.22
Lowest branching node	*LBN3*	Ha8_25299468	8	115.0	109.0-117.0	8.4	3.62
Percent branched	*PB2*	Ha9_124866422	9	337.0	336.0-337.7	11.56	5.08
Petal colour	*PC1*	Ha17_60676293	17	211.0	210.0-211.0	5.94	3.99
Reproductive budding	*RB1*	Ha2_68205646	2	53.1	51.0-62.33	8.76	3.78*
Stem diameter	*SD1*	Ha16_198498440	16	110.0	101.0-111.0	7.84	3.38
Stem diameter	*SD2*	Ha11_61988152	11	112.0	98.0-113.0	7.33	3.15
Total capitula count	*TCC4*	Ha11_61988152	11	112.0	100.38-113.0	8.35	3.66

Note: * Significant at alpha=0.05 at the genome wide level

† cM =Centimorgan

‡CI = LOD-1 confidence interval

§ %Var. =proportion of phenotypic variance explained by QTL (equivalent to marker r^2 x 100)

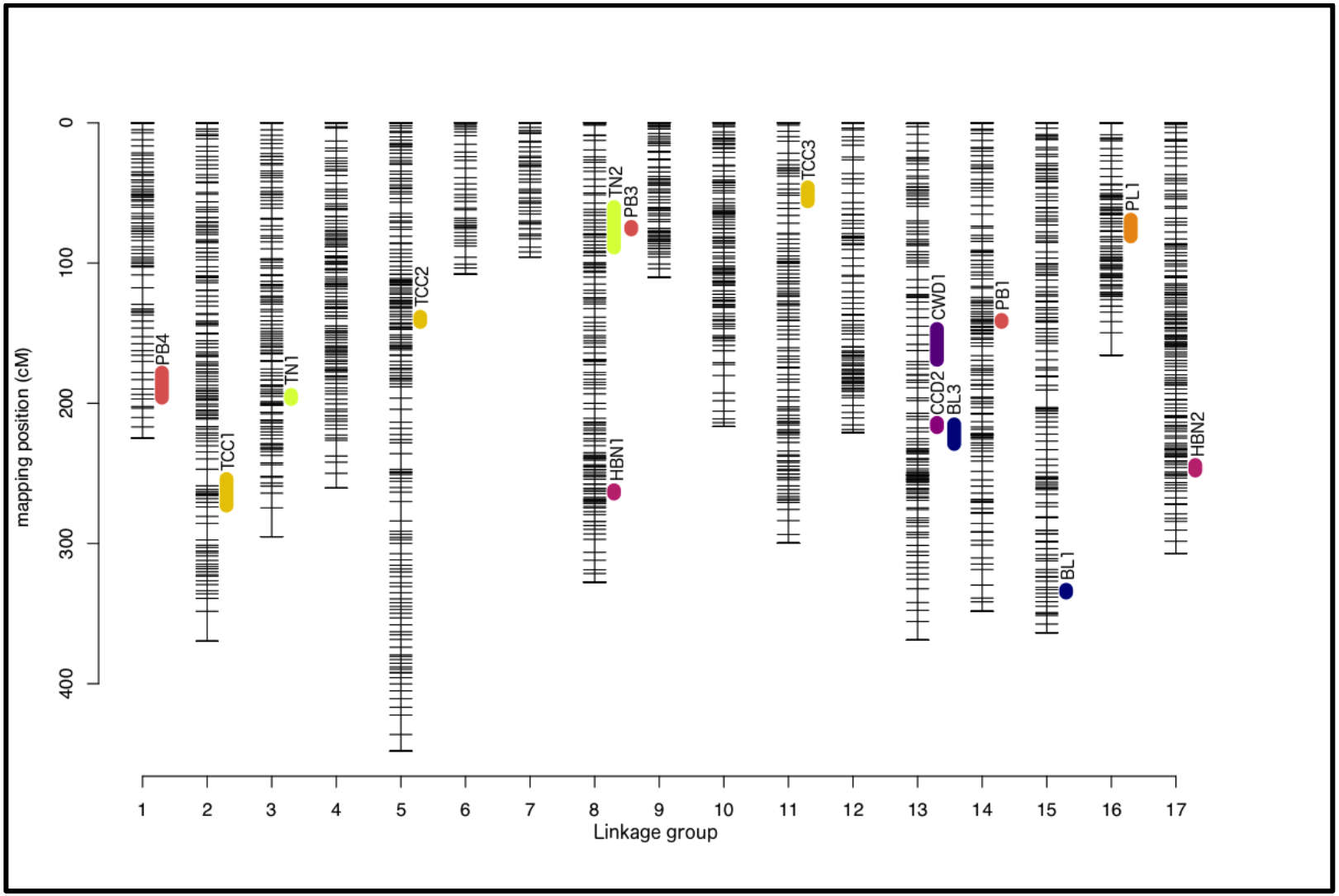


Figure 5.2: *H. maximiliani* genetic map produced using D1.10 class markers overlaid with QTL significant at the chromosome level.

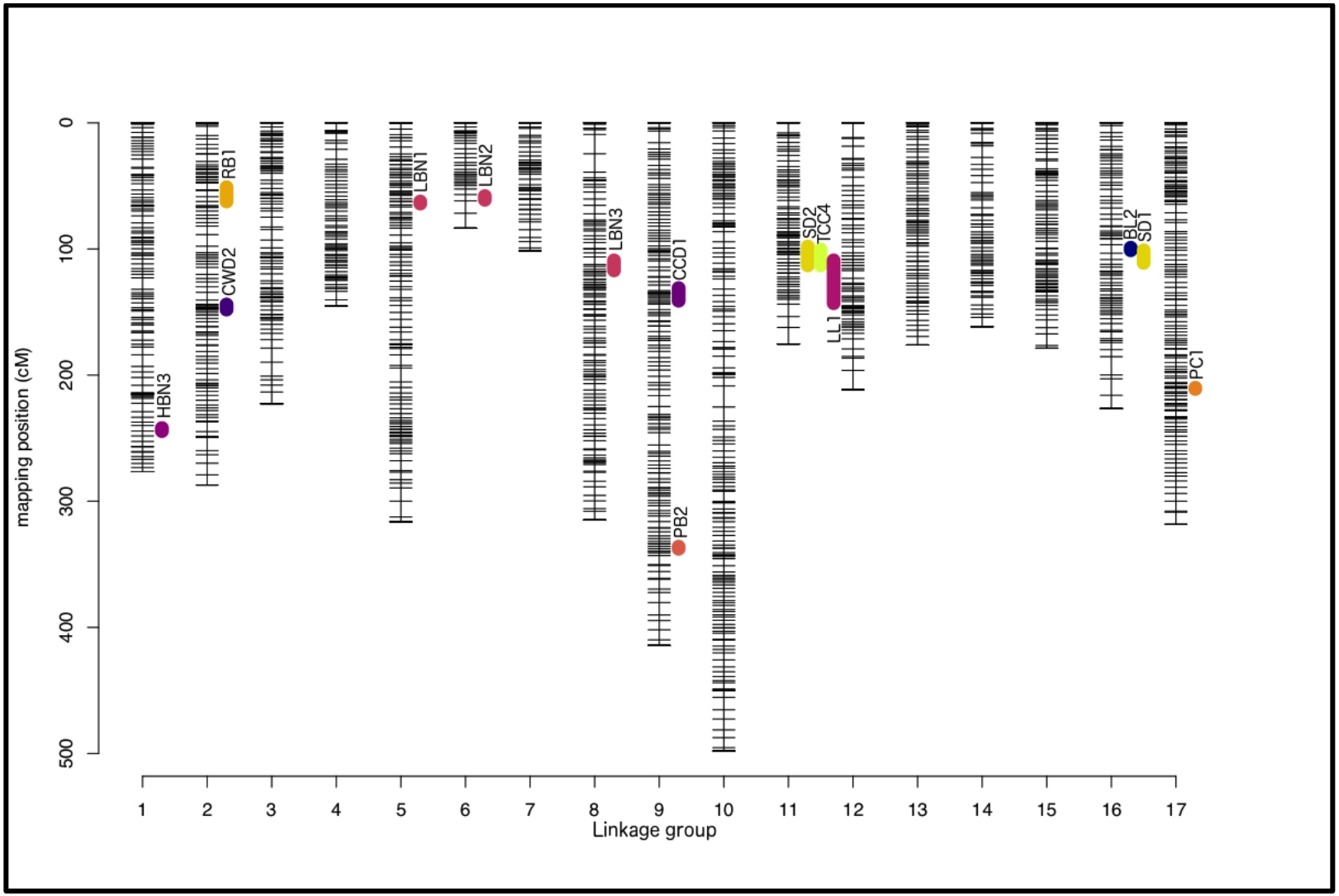


Figure 5.3: *H. maximiliani* genetic map produced using D2.15 class markers overlaid with QTL significant at the chromosome level.

5.4.6 SNP annotations

BLASTn analysis annotated nine of the candidate SNPs to genes of known functions within the *H. annuus* and *A. thaliana* references genomes, with the remainder being annotated to genes of unknown function. Notable SNPs found within known genes include the *H. maximiliani* SNPs *Ha1_107557612* and *Ha1_107557623*, which were found within the *H. annuus* gene *Ha412v1r1_01g025490* annotated as a terpenoid cyclases/protein prenyltransferase alpha-alpha toroid (IPR008930) and terpenoid synthase (IPR008949) gene; *Ha14_131232381* found within the gene *Ha412v1r1_14g023720* which is described as a nicotinamide N-methyltransferase-like gene (IPR019410); *Ha4_177393919* and *Ha4_177393987*, which fall within the gene *Ha412v1r1_04g042290* which is annotated as coding a Glyoxalase-like domain (IPR025870); and *Ha16_198498440*, found within the *H. annuus* gene *Ha412v1r1_16g048540*, annotated as a *SPX*, N-terminal (IPR004331) and EXS, C-terminal (IPR004342) coding gene. *Ha14_34533074*, *Ha16_215897889* and *Ha4_430240* were found within gypsy-like retrotransposon family genes in the *A. thaliana* reference genome. BLASTx analysis did provide additional matches.

5.5 Discussion

5.5.1 Phenotypic variation and trait associations

The phenotypic variation within the F_2 populations was extensive due to the wide divergence between the parental phenotypes selected for the initial crosses. In diploid species if parents contributing to a cross are heterozygous for unique alleles at each locus, (i.e. AB x CD) there are four potential allelic combinations in the F_1 generation (AC, AD, BC, BD) and if mating is random in the F_1 generation, there are 10 allelic combinations in the F_2 generation (AA, BB, CC, DD, AB, AC, AD, BC, BD, CD). In this study, traits that were not observed in the parental or F_1 generations (the presence of tubular rays and double-flowered/chrysanthemum-like phenotypes) may be attributed to this type of phenomena or to the masking effects of dominance or epistasis. Branching architecture, timing of anthesis, plant height and capitulum size showed a wide range, indicative of primarily quantitative modes of inheritance.

The quantitative genetic model of trade-offs views trait relationships as dynamic, multi-factorial in nature and capable of changing under selection (Lande, 1982; Roff et al., 2002; DeHaan et al., 2005). Therefore, it is possible that trait relationships may simultaneously be both positively and negatively correlated as the components composing the traits in question may interact in different fashions. For example, in this study the relationships between traits such as timing of anthesis and central capitulum size exhibited a positive relationship along the first principal component axis and a negative relationship along the second principal component axis. This is indicative of that while there may be a trade-off between these traits, it is not absolute and may be explained by underlying phenomena such as linkage disequilibrium.

Several researchers have studied the influence of selection on branching architecture and a single capitulum in cultivated sunflower. Branching locus *B* (Fick and Miller, 1997), located on linkage group 10, is one of several loci that has undergone selection to produce the unbranched plant architecture observed in cultivated sunflower. Branching locus *B*, while not the sole locus responsible for reduced apical branching, is known to pleiotropically affect branching architecture, capitulum and seed size, pericarp thickness and seed oil content (Bachlava et al., 2010). This relationship has been cited as an example of a resource allocation trade-off between seed size and seed number (Sadras, 2007). Branching in cultivated sunflower increases the number of capitula, but typically decreases capitulum diameter and seed weight (Fick et al., 1974; Dedio, 1980; Tang et al., 2006). Selection for restricted branching in Maximilian sunflower has been suggested as a possible cause for a greater range of capitulum sizes observed in unbranched plants (Van Tassel pers. comm.). Seed yield capitulum^{-1} in unbranched plants can exceed those of branched Maximilian sunflower plants (Van Tassel et al., 2014), indicating parallel trade-offs in both species under selection.

5.5.2 Linkage map development

To date the genomes of perennial *Helianthus* species are poorly characterized and the level of syntenty between perennial and annual genomes is unresolved. This study is the first report of a linkage map for Maximilian sunflower and starting point for understanding *H. maximiliani* and related species. This study will serve as a framework for future genetic studies in this species, and builds towards understanding the relationships between annual and perennial *Helianthus* species. The D1.10 and D2.15

analyses show a degree of map inflation, similar to other recent studies employing genotype-by-sequencing (Shirasawa et al., 2017; Konar et al., 2017). Mapping studies in annual cultivated *H. annuus* involving multiple crosses have produced maps with an estimated size of approximately 1,310-1,443.83 cM (Bowers et al., 2012b; Talukder et al., 2014) compared to the D1.10 (4,530 cM) and D2.15 (4,109.85 cM) maps. Due to the nature of the population under examination, technical challenges in developing the genetic map and potential ascertainment bias of using a homeologous reference genome, it was not possible to accurately detect putative translocations which have been suggested to differentiate the *H. maximiliani* and *H. annuus* genomes (Whelan, 1978). The development of genetic resources to support breeding efforts in crops is a continuous process and builds on established research. This study presents a starting point for establishing genetic resources for breeding Maximilian sunflower, which will lead to future refinements to the genetic map and ability to identify QTL and candidate genes for traits of interest.

5.5.3 Domestication syndrome characteristics in Maximilian sunflower

Common adaptations to human cultivation in crop species includes the selection for plants exhibiting larger grains that are retained until harvest, manipulation of flowering time to coincide with the length of the growing season, increases in harvest index, reduction in plant height and lodging and loss of seed dormancy. There appears to be genetic variation for a number of these traits in Maximilian sunflower though some characteristics appear to be either pleiotropic or genetically linked within the F_2 populations examined. Relatively few QTL or GLM associations were significant at the

genome wide-level, which may reflect the nature of the population in this study and patterns of segregation distortion along the linkage groups. The influence of phenotypic covariates may also confound the power to detect QTL and mask the effects of causative loci. Caution must be taken in the use of covariates in QTL analysis as the strength of relationships between covariates can greatly increase (low correlation) or decrease (high correlation) the power to detect QTL (Zeegers et al., 2004). Doerge and Churchill (1994) cautioned that, in the presence of segregation distortion, permutation analysis may result in overly conservative thresholds. Therefore, permutations should be run for regions with similar marker segregation patterns. The presence of segregation distortion peaks along the linkage groups may have also contributed to a reduced power to detect QTL at the genome wide-level through inflation of threshold values. The MLM analysis and chromosome-level permutation testing revealed a number of QTL and GLM associations on linkage groups known to harbor domestication syndrome QTL in cultivated sunflower (Supplemental Table 5.6).

Timing of anthesis is one of the few traits believed to be controlled by relatively few loci in the domestication of *H. annuus* (Burke et al., 2002; Chapman et al., 2008b; Blackman et al., 2010; Wills et al., 2010). A single QTL for timing of reproductive budding, *RB1*, was detected in the D1.10 map and candidate SNPs were detected in both the MLM and GLM analysis corresponding to linkage group 13 of *H. annuus*. A series of SNPs were associated with both timing of first anthesis and average anthesis in the GLM association analysis, corresponding to linkage groups 12, 15 and the unordered contig Ha0_73Ns of *H. annuus*. Further SNPs corresponding to linkage groups 13 and 14 were supported in both the MLM and GLM analysis for timing of average and first anthesis,

respectively. Previous studies of crosses between cultivated *H. annuus*, early landraces and wild populations have established flowering time QTL on several linkage groups, with major QTL being found on linkage groups 6 and 15 (Burke et al., 2002; Will and Burke, 2008; Blackman et al., 2010). The GLM analysis supported an association between a SNP corresponding to linkage group 15 of *H. annuus* and both first and average anthesis.

The presence of multiple QTLs for branching characteristics agrees with previous studies in crosses between wild and cultivated *H. annuus*. Separate loci controlling basal, apical, and complete branching suggest branching is a genetically complex trait influenced by numerous small effect loci spread throughout the genome (Burke et al., 2002; Wills and Burke, 2007; Dechaine et al., 2009; Nambeesan et al., 2015). A number of branched nodes corresponded to linkage groups 1, 8, 9 and 14 and association analysis supported SNP associations for linkage groups 1, 13, 14 and 15 in the GLM and 1, 8, 13 and 15 in the MLM analysis, respectively. Previous QTL studies examining branching in *H. annuus* have detected QTLs on chromosomes 6, 7, 10, 13, 16 and 17 (Burke et al., 2002; Tang et al., 2006; Wills and Burke, 2007). Association mapping studies have detected candidate SNPs on linkage groups 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 15, 16 and 17 controlling various aspects of branching architecture in *H. annuus* (Mandel et al., 2013b; Nambeesan et al., 2015). Similarly, QTL and associations were detected on several linkage groups for the percentage of branched nodes as well as general branching patterns which coincide with previous studies in *H. annuus* (Supplemental Table 5.6). Quantitative trait loci for branching position were identified on linkage groups 1 and 17 for the highest branching node and 5, 6 and 8 for lowest branching node. Association

analysis revealed further associations corresponding to linkage groups 5, 13 and 15 in the GLM analysis and linkage group 5 in the MLM analysis for the highest branching node. Associations for lowest branching node were detected for linkage groups 4 and 17 for the GLM analysis and 4, 9 and 15 for the MLM analysis. Both the QTL and association analysis supported associations between linkage groups 1, 8 and 14 and the percentage of branched nodes. In addition, linkage group 8 harbored a QTL for the highest branching node (*HBN1*) in the D1.10 map and lowest branching node (*LBN3*) in the D2.15 map, while GLM analysis supported an association for highest branching node on linkage group 1, giving further weight to these linkage groups harboring loci contributing to branching architecture. The QTL on chromosome 8 associated with percentage of branched nodes (*PB3*) overlaps with a QTL for total nodes (*TN2*), which may have influenced branching measurements, as they were calculated as a percentage of total nodes.

A candidate gene potentially involved in the control of branching architecture was uncovered in the association analysis, and may prove useful in the selection for unbranched plant architecture in *H. maximiliani*. The SNPs *Ha1_107557612* and *Ha1_107557623*, which were associated with the percentage of branched nodes in both the GLM and MLM analysis, were annotated to a *H. annuus* gene *Ha412v1r1_01g025490*, involved with terpenoid cyclases/protein prenyltransferase alpha-alpha toroid and terpenoid synthase. Three phytohormones, auxin, cytokinin and strigolactones and genes involved in their homeostasis and signalling are thought to be responsible for the regulation of branching (Umehara et al., 2008; Ferguson and Beveridge, 2009; Nambeesan et al., 2015). Terpenoid derived compounds including

strigolactones are novel phytohormones regulating the suppression of shoot branching in the *MAX/RMS/D* pathway in both dicots and monocots (Umehara et al., 2008). The annotations of *Ha412v1r1_01g025490* in combination with the presence of a QTL on the same linkage group influencing the percentage of branched nodes (*PB4*) and supporting studies in *H. annuus* suggesting the presence of branching QTL on this linkage group (Nambeesan et al., 2015) present *Ha412v1r1_01g025490* as a strong candidate gene influencing branching.

The relative number of nodes influences branching and leaf number as the nodes are where leaves and axillary buds emerge. Two QTL for total nodes were detected on linkage groups 3 and 8 of the D1.10 map and three SNP associations were found corresponding to the unordered contig Ha0_73Ns and linkage groups 4 and 14. Total nodes were correlated with plant height, timing of anthesis, and capitulum size traits to a degree in this study, similar to previous studies in Maximilian sunflower (Chapter 4). The identified QTL may be important for direct or indirect selection for traits related to harvest index. No QTL were detected for plant height, which may be the result of the confounding effect of timing of the switch from vegetative to reproductive growth. Plants with a longer vegetative period tended to be taller with greater biomass than plants with a shorter vegetative period (Kawakami et al., 2011; Chapter 4). Controlling for the transition from vegetative to reproductive growth may allow for further growth related QTLs, independent of the length of the vegetative period, to be detected.

No QTL were detected for central capitulum size, although two regions corresponding to chromosomes 5 and 13 of *H. annuus* were detected in the MLM association analysis. Burke et al. (2002) described QTL on both of these linkage groups

in *H. annuus* and characterized them as partially recessive and partially dominant, respectively. Similarly, association analysis revealed SNP associations in both the MLM and GLM corresponding to linkage group 17 of cultivated sunflower for the capitulum depth:width ratio, the same linkage group Burke et al. (2002) previously described as a QTL for this trait which they associated with shattering resistance. The presence of QTL and SNP associations for this trait is promising as shattering resistance is a major domestication syndrome trait in many crops and a common target trait identified for the improvement of perennial grains and oilseeds (Wagoner and Schaeffer, 1990; DeHaan et al., 2016).

Association analysis revealed a cluster of SNPs associated with petal morphology corresponding to linkage group 5 of *H. annuus*. Members of the CYCLOIDEA(*CYC*)/ TEOSINTE-BRANCHED1 (*TB1*) family of transcription factors are known to play a role in the floral development in the *Compositae*, resulting in tubular-rayed (*tub*) and double-flowered/chrysanthemum-like (*dbl*) mutants (Chapman et al., 2012). Though this trait is mapped to linkage group 9 of *H. annuus*, associations were found for petal morphology and whorling corresponding to linkage groups 5 and 1, respectively. Chapman et al. (2012) suggested *tub* and *dbl* mutations may be allelic or tightly linked in *H. annuus* due to the lack of individuals exhibiting both traits in their study. Individuals exhibiting both the *tub* and *dbl* trait were observed in this study (Supplemental Figure 5.4) indicating the presence of multiple loci or co-dominance influencing this trait in Maximilian sunflower. Though petal morphology is not a domestication syndrome characteristic in field cultivated sunflower, these traits are of interest for the development of ornamental sunflowers. The presence of traits associated with mutations in the *CYC/TB1* gene family

are of particular interest given the role this gene family in known to play in the domestication of several crop species. *TB1* family of transcription factors and related orthologs have been implicated in the domestication of crops such as maize (*Zea mays* L.)(Doebley et al., 1997), pearl millet [Pennisetum glaucum (L.) R.Br.](Remigereau et al., 2011) and barley (Ramsay et al., 2011). Previous studies have established that members of this gene family overlap with QTL for branching architecture in domesticate x wild crosses of *H. annuus,* suggesting a potential link between these traits (Chapman et al., 2008a). Though association was not observed in the present study, the presence of this trait in a population segregating for branching architecture is of interest given its known role in other species. The role of this gene family may warrant further investigation in both branching architecture and potential horticultural applications of *H. maximiliani*.

5.6 Conclusions

This study uncovered candidate regions for the control traits such as branching architecture, timing of anthesis, and capitulum morphology of interest for neo-domestication of Maximilian sunflower, a crop wild relative of annual sunflower. Some of the challenges and prospects in breeding this species as a perennial oilseed crop are identified. Several QTL and candidate SNPs were found on the same linkage groups as previously described domestication syndrome traits in *H. annuus,* including important traits such as timing of anthesis, branching architecture and shattering potential. This indicates that orthologs may influence these traits in a parallel fashion between species.

Other novel candidate loci were found to influence previously described trait relationships and pathways underlying important domestication syndrome characteristics in annual sunflower, suggesting independent paths may be taken to the domestication of this species. Candidate traits described in this study provide an initial step in understanding the genetic control of the domestication process in Maximilian sunflower and in the development of molecular markers for genomic assisted breeding methodologies such as marker-assisted and genomic selection. The application of marker-assisted selection is particularly beneficial in outcrossing species such as Maximilian sunflower as greater genetic heterogeneity and the effects of dominance may slow the rate at which traits are fixed under direct phenotypic selection. Applying MAS in Maximilian sunflower may also allow for the development of a wider pool of germplasm to study the influence of genetic background and introduce traits of interest into new populations. Given the observed variation in *H. maximiliani*, MAS or genomic selection may be employed to capitalize on standing variation within the species to make genetic advancements without resorting to interspecific introgression and associated challenges.

Further research is required to examine the exact nature of segregation distortion observed in this study and how it may affect the development of this species as a crop. Improvements in bioinformatic approaches to studying out-crossing species will likely improve the quality of the genetic maps and provide better control of paralogous loci. Improvements to the genetic map will provide a greater understanding of trait relationships, greater accuracy in detecting QTL and the degree of genetic divergence between *H. annuus* and *H. maximiliani*. Studies directly comparing the domestication of

H. annuus and neo-domestication of *H. maximiliani* will give further insight to the degree in which parallel paths to domestication may be exploited for crop development. Understanding trait relationships, how they may be selected for or against and how this translates to performance as a crop will aid in defining the neo-domestication of Maximilian sunflower. The present study provides a baseline for future investigations into breeding Maximilian sunflower and how modern breeding techniques may be applied to develop this and other species as novel perennial crops. This research supports the idea the parallels between *H. maximiliani* and *H. annuus* may be utilized to bring Maximilian sunflower or other perennial oilseed candidates into domestication.

CHAPTER 6.0: General discussion and conclusions

This Ph.D. thesis provides further knowledge to support breeding efforts in Maximilian sunflower and related perennial *Helianthus* species for the development of a locally adapted perennial oilseed crop. This research provided the first report of the population genetics, landscape genomics and adaptation of Maximilian sunflower at the local scale. Breeding resources in the form of phenotypic characterization, the first reported genetic map of Maximilian sunflower, QTL for potential domestication syndrome traits and SNP associations with phenotypic variation and environmental characteristics were revealed to further breeding efforts. The overall objectives of these research studies were to:

1) Provide a baseline characterization of the phenotypic characteristics of available perennial *Helianthus* germplasm adapted to southern Manitoba, Canada and the potential for advancement under selection (Chapter 3).
2) Examine the landscape genetics and genomics of Maximilian sunflower (Chapters 3 and 4).
3) Develop the first reported genetic map of Maximilian sunflower (Chapter 5).
4) Identify QTL and candidate SNPs associated with phenotypic differentiation, environmental clines and domestication syndrome characteristics in Maximilian sunflower to support breeding efforts (Chapters 4 and 5).

The following general discussion and conclusions synthesizes the pertinent findings from this thesis that may be used to inform effective breeding strategies for the development of

Maximilian sunflower as a perennial oilseed crop. Future avenues for research in Maximilian sunflower and the area of perennial grains and oilseed development are discussed.

6.1 Defining the domestication ideotype of Maximilian sunflower

Phenotypic variation for a number of important domestication syndrome traits of grain and oilseed crops was observed in Maximilian sunflower and related perennial *Helianthus* species. There is sufficient variation in the initiation of anthesis, average capitulum size, plant height, branching architecture and traits associated with shattering to make advancements under selection. Maximilian sunflower lacks a definitive crop ideotype and research into the performance of various phenotypes under real-world agronomic conditions has yet to be examined. While selection for a single large capitulum is a defining domestication syndrome characteristic of cultivated annual sunflower (*Helianthus annuus*), the path Maximilian sunflower will take ultimately depends on the targeted end-product (seed or biomass) and how various traits contribute yield. In this thesis a baseline characterization of Maximilian sunflower and related species shows that there is the necessary diversity to select for different ecotypes from local germplasm and that traits such as restricted branching architecture may be recovered readily when crossed with local materials. This research sets the stage for future studies examining the yield components and agronomic characteristics of different Maximilian sunflower branching ideotypes in an appropriate genetic background.

Branching pattern directly influences light capture, water transport, mechanical

support and wind resistance (Farnsworth and Niklas, 1995). In sunflower, branching interacts with capitulum size, total capitula number, seed weight and oil content amongst other traits (Fick et al., 1974; Dedio, 1980; Tang et al., 2006; Bachlava et al., 2010). While there are inherent benefits to restricting branching, such as the facilitation of mechanical harvest or reducing competition between branches and uniform maturity, restricting branching may also limit yield potential if selected for too stringently. Under both growth chamber (Chapter 4-5) and common garden conditions (Chapter 3, data not shown), branched Maximilian sunflower was capable of producing a large number of capitula per $stem^{-1}$ (>100) while completely unbranched plants exhibited a single, central capitulum. The relationship between capitulum size and capitula number was not a 1:1 ratio (Chapter 5). However, plants with restricted branching had the tendency to exhibit larger capitula (Chapter 5; Van Tassel et al., 2014) and an increase in captilulum size did not appear to fully compensate for the loss of capitulum number. For instance, the negative relationship observed between capitulum size and number accounted for 19% of the variation in average capitulum size and only 10.2% of the variation in capitulum number (Chapter 5). Selection for a single central capitulum, akin to cultivated sunflower, would decrease harvest index if the loss of capitula number is not compensated for by a reciprocal increase in capitulum size.

Helianthus annuus was initially domesticated for its edible seed, pigments and medicinal compounds (Heiser, 1951) with selection for oil content and composition occurring more recently (Burke et al., 2005; Chapman and Burke, 2012). Therefore, the defining characteristics of its domestication may not necessarily apply to the neo-domestication of Maximilian sunflower as an oilseed crop. Selection for a Maximilian

sunflower ideotype as an oilseed may parallel other composite and/or small-seeded oilseeds which have retained branching. The *Compositae* oilseeds noug (*Guizotia abyssinica*) and safflower (*Carthamus tinctorius*) both exhibit contrasting domestication syndromes to annual sunflower (Pearl et al., 2014; Dempewolf et al., 2015), having retained a branched architecture. Similarly, successful oilseed crops in Western Canada such as canola (*Brassica napus* L.) retain a branched architecture, small seed size, and are still prone to shattering under certain conditions (Gulden et al., 2003; Cavalieri et al., 2016). Despite these characteristics, these crops are capable of sustaining economic yields supporting their use as crops.

Selection for restricted, but not unbranched plant architecture may enhance harvestable yield indirectly through greater synchronicity of flowering and maturity of capitula and through changes in capitulum morphology, such as an increased capitulum depth:width ratio (Chapter 5). Restricted branching may also alleviate lodging through a reduced plant size and biomass (Chapter 3). While increased branching and seed number may constrain seed size, in *H. annuus* smaller seeds tend to bear a higher concentration of oil. Estimates of 6.7-8.5% greater oil concentration in the seed of branched individuals relative to unbranched individuals in segregating populations have been observed, presumably due to a thinner pericarp (Tang et al., 2006; Bachlava et al., 2010). Therefore, manipulating branching and seed number may prove beneficial in increasing oil yield per unit area. Selection for early flowering may reduce plant biomass through reductions in correlated traits such as plant height (Chapter 4) and potentially increase the number of capitula produced per unit of biomass. Increased harvest index may be achieved through optimizing the restriction of branching to reduce non-productive lower-order branches if

selection for greater reproductive output per unit area is maintained, likely through increases in capitulum size or number. Furthermore, increasing capitulum size has the benefit of potentially increasing the capitulum depth:width ratio (Chapter 5) which may alleviate shattering through changes in capitulum morphology. Increasing the harvest index of Maximilian sunflower through conventional as well as marker-assisted breeding approaches appears to be possible through multiple pathways given the variation in traits observed within both the wild sampled accessions (Chapters 3 and 4) and F_2 populations characterized (Chapter 5).

6.2 Examination of trade-offs in Maximilian sunflower

Understanding trade-offs and how they influence plant traits is important in predicting response to selection in different populations. Trait relationships may differ between populations, or amongst genotypes within populations as well as growth environments. Different patterns were observed between trait pairs in various chapters of this thesis. For instance, in Chapter 4 no relationship was observed between days to first anthesis and branch number while in chapter 5 a negative relationship was observed between days to first anthesis and the percentage of branched nodes. This is an example of how differences in populations and linkage disequilibrium may influence the appearance of trait trade-offs and couple traits to one another. The populations studied in Chapter 4 were genetically diverse, wild sampled plants with low population structure which have not undergone selection. This differed from the populations studied in Chapter 5, which were derived from crosses between wild plants and plants that had

undergone selection for restricted branching. As the materials employed Chapter 5 were of the F_2 generation, a certain degree of linkage disequilibrium is expected to influence trait relationships due to limited generations of recombination. Associations between later flowering and a lower percentage of branched nodes revealed along the first and second axes of the principal component analysis may reflect linkage disequilibrium due to the fact that the Kansas parent exhibited these characteristics, while the earlier flowering Manitoba parent exhibited branching. While these patterns may occur to chance, the SNPs associated with the percentage of branched nodes and reproductive budding, as well as percentage of branched nodes and average timing of anthesis, corresponded to linkage groups 13 and 14 of *H. annuus,* respectively. The presence of SNP associations on the same linkage group indicates a possible physical linkage between loci controlling these traits. Further generations of recombination, a larger population size and the analysis of a wider scope of populations segregating for these traits may remove the confounding effect of linkage disequilibrium and decouple these traits.

Trade-off patterns between quantitative traits may be multi-factorial and exhibit both positive and negative associations simultaneously as observed in Chapter 5. Capitulum size traits were positively associated with timing of anthesis reflecting the parental types along the first principal component axes, but showed the opposite relationship along the second axes. The differences in relationships between the first and second principal component axes may represent potential recombination between linked loci, the breakdown of linkage disequilibrium, or that relationships between the various loci influencing these traits differ amongst individuals within the F_2 populations.

Lastly, the influence of trait trade-offs is also dependent on external factors that interact with genotype, resulting in phenotypic plasticity. Resource availability, (genotype x environment) and management practices (genotype x management) are important factors that may interact with genotype and influence both the appearance and relative importance of trade-offs though their influence on phenotype. Ultimately, in Maximilian sunflower understanding the appearance of trade-offs in the context of production will dictates their importance and utility in plant breeding efforts.

6.3 Breeding strategies for developing Maximilian sunflower as a perennial oilseed crop

Maximilian sunflower is a highly heterozygous species (Chapter 4) and would benefit by the application of cross-pollinated breeding strategies. While the development of partially-inbred materials through sibling mating is possible in this species (Chapter 5) and may be useful in the development of parental materials for polycrosses, overreliance on sib-mating may result in segregation distortion due to the effects of self-incompatibility and inbreeding depression. Segregation distortion and linkage disequilibrium can decrease the effectiveness of selection through a reduction in favourable recombinants between parental traits for loci in distorted regions of the genome, particularly on linkage groups associated with self-incompatibility genes (Anhalt et al., 2008; Do Canto et al., 2016, 2018).

The research conducted in this thesis provides a framework for development of marker-assisted-selection and training of genomic selection models for Maximilian sunflower using next-generation sequencing technology. Application of maker-assisted-

selection in polycross populations of Maximilian sunflower of sufficient size may be an effective strategy to fix favourable traits without the sacrificing genetic diversity. Furthermore, in genetically heterogeneous populations, the maintenance of diversity affords a degree of plasticity to abiotic and biotic stress through the buffering effects of multiple genotypes with potentially different mechanisms of resistance (Wilkins and Humphreys, 2003; Uppalapati et al., 2013; Annicchiarico et al., 2014). Methods that maximize genetic gain, while maintaining population diversity, such as among-and-within-family selection methods (Casler and Brummer, 2008), are likely the best approach to developing Maximilian sunflower given its outcrossing nature.

Applying marker assisted and genomic selection techniques to the development of perennial grains could greatly increase the speed at which selection cycles are completed. The length of selection cycles in perennial crops are approximately 3 to 5 years, as often an establishment year is required, followed by a minimum of two years of evaluation (Wilkins and Humphreys, 2003; Casler and Brummer, 2008; Resende et al., 2013). In addition, selection for recessive or epistatic traits that cannot be selected for as effectively in cross-pollinated populations using standard phenotypic selection will greatly benefit from the development of molecular markers and genomic prediction tools. Numerous QTL and SNP associations found on linkage groups corresponding with previous genetic mapping studies in *H. annuus* (Chapters 4 & 5). Targeting the increasing number of characterized candidate genes in *Arabidopsis* and *Helianthus* genomes may provide the opportunity to exploit homologies between these species and *H. maximiliani*. Exploiting these homologies though marker-assisted selection may prove to be an effective strategy to develop a domestication ideotype for Maximilian sunflower without having to resort to

interspecific introgression and associated challenges of linkage drag between annual and perennial *Helianthus* genomes.

6.4 Future research

6.4.1 Field cultivation of Maximilian sunflower

There appears to be the necessary genetic potential to make advancements under selection in locally adapted germplasm of Maximilian sunflower. Understanding the performance of the species under cultivated settings is critical to identify of an appropriate ideotype for selection and to study the impact of genotype x environment interactions on yield and yield components. As discussed in Chapters 3 and 4, Maximilian sunflower shows a degree of population differentiation in response to mean monthly temperatures, particularly in regards to timing of anthesis, indicative of adaptation to local growing conditions. Further investigation of genotype x environment interactions influencing plant yield, productivity and biomass production is required. The research contained within this thesis sets the stage for future agronomic studies for perennial sunflowers.

The influence of plant spatial arrangement and competition will likely influence characteristics such as branching architecture, total capitula count, rhizome development, stem density, and emergence in perennial sunflowers. This is supported by that fact that plants growing under high competition settings have a greatly reduced size compared to plants grown under common garden conditions. Agronomic studies examining the

interaction between plant morphology and agronomic practices (genotype x management interactions) will inform the direction the domestication ideotype of Maximilian sunflower. The performance of Maximilian sunflower under sward conditions remains uncharacterized. Early investigations conducted at The Land Institute using unimproved germplasm showed that final yield did not differ across planting densities in the year of establishment, while branching and capitula number became more restricted and plant height increased as stand density increased (Riley, 1984; Calsbeek, 1984; Kois, 1985), indicative of a high potential for compensatory growth. The influence of rhizomatous spread and number of stems produced per plant, both traits which showed a strong genetic component and influenced by stand density in the related *H. tuberosus* (Gallard, 1985; Lemercier, 1987; Kay and Nottingham, 2008), require further investigation into how these traits may influence yield components under different stand densities.

Due to the outcrossing nature of Maximilian sunflower, studying important yield contributing traits such as seed size/weight and total seed yield is constrained in environments were adequate cross-pollination is limited. Future studies under field conditions with adequate potential for cross-pollination will reveal true determinants of yield and important agronomic characteristics necessary to achieve yield potential. Full yield component analysis of different plant morphotypes and characterization of oil and protein content and quality will provide information of the economic potential of Maximilian sunflower. Understanding these characteristics will help uncover potential markets for Maximilian sunflower which will drive the development of appropriate agronomic packages to support yield.

6.4.2 Refining the genetic map and exploiting genomic resources of Maximilian sunflower

The presence of segregation distortion and pseudo-linkage in the mapping population (Chapter 5) limited the ability to develop a genetic map independently of the reference genome of *H. annuus*. Developing a linkage map from larger F_1 populations in place of a F_2 population may provide further insights into the structure of the Maximilian sunflower genome and potential chromosomal rearrangements that have occurred between the divergence of *H. annuus* and *H. maximiliani*. Additionally, *de novo* SNP assembly and the establishment of a pseudo-reference genome of Maximilian sunflower may recover regions of the Maximilian sunflower genome which do not correspond to the *H. annuus* reference genome and may provide insights into the genetic control of characteristics which differentiate these species, including the nature of perenniality in the genus *Helianthus*. Few markers were aligned to linkage groups 6 and 7 of *H. annuus* in both the association panel (Chapter 4) and F_2 mapping population (Chapter 5). Linkage group 6 is known to harbor a number of QTL for traits of interest for domestication and improvement of sunflower, including major clusters of QTL for flowering time (Blackman et al., 2010), achene width and weight (Burke et al., 2005) and oil-related traits. Linkage group 7 is believed to harbor QTL for traits such as flowering time, plant height, stem diameter, branch and capitulum number (Chapman et al., 2008b) in crosses between various wild, land race and improved sunflower lines. Further characterization of these linkage groups may yield further QTL for these and other important traits for breeding Maximilian sunflower.

Refined bioinformatic approaches to calling SNPs in highly repetitive genomes

and improvements to genetic mapping algorithms to handle error rates in next-generation-sequencing data with complex pedigrees will be key to further refine the genetic map of this species (N'Diaye et al., 2017; Paris et al., 2017; Bilton et al., 2018; McKinney et al., 2018). The development of specialized populations which are phenotypically informative, such as bi-parental mapping and association panel populations, will also yield further candidate SNPs and expand the genomic resources available to this species. These approaches will support the application of modern breeding techniques such as marker-assisted and genomic selection techniques in Maximilian sunflower.

Maps incorporating the D1.10 and D2.15 markers using B3.7 markers as a bridge generated in Onemap showed considerable regions of map inflation and overall poor integration. As many of the B3.7 markers were heterozygous in both the parental and F_1 generations, accurate phasing was not possible for many of the markers. In addition, the presence of sequencing errors or sequence reads representing paralogs incorrectly called as single loci that map to multiple regions of the *H. maximiliani* genome may give the appearance of F_2-like segregation patterns. This phenomena may also have contributed to map inflation. While BLASTn analysis allowed for the removal of loci which corresponded to multiple locations in the *H. annuus* reference genome, this analysis does not guard against paralogous regions of the *H. maximiliani* genome which may have diverged.

Maps produced using the MSTmap algorithm via *ASMap* showed considerably shorter map lengths, but this approach is limited to phase-known markers and is not designed to handle outcrossed pedigrees containing a mixture of marker patterns.

Separating the D1.10 and D2.15 markers into separate genetic maps and not including B3.7 markers potentially compromised the power to detect QTL, particularly traits masked by the effects of dominance. This phenomenon may explain why some traits, which became apparent in the F_2 generation were detected by the association analysis, and not in the QTL analysis. Recently developed software adapting the MSTmap algorithm to outbred populations such as LEP-MAP (Rastas, 2017) and *GUSmap* (Bilton et al., 2018) showed promise to reduce map inflation while incorporating multiple marker types when working with GBS-derived datasets in outcrossing species, and may be employed to improve the quality of the genetic map moving forward.

The genetic maps produced in this study exhibit inflation, as stated above. The structure of the crosses likely contributed to map inflation due to the presence of a high degree of marker segregation distortion. Retention of markers showing segregation distortion was necessary, as their removal resulted in many small fragmented linkage groups, which were non-conducive to linkage mapping. The populations were developed though mating full-siblings in an attempt to generate phenotypically informative materials, necessitating inbreeding of a normally cross-pollinated species. *H. maximiliani* is an obligate outcrossing species exhibiting sporophytic self-incompatibility (Heiser et al., 1969) which may contribute to regions of segregation distortion (SDR) through the presence of a pre-zygotic reproductive barrier (Bodénès et al., 2016). While segregation distortion may contribute to genetic map inflation and the loss of power in QTL detection, it does not strongly impact marker order or accuracy of QTL position in backcross-like populations of sufficient size and marker density under most

circumstances. Segregation distortion can negatively affect the power to detect QTL with dominance effects in F_2-like populations (Xu, 2008; Zhang et al., 2010).

The population dynamics of Maximilian sunflower may also contribute to the presence of SDR through the accumulation of unfavourable alleles (genetic load). In its natural setting *H. maximiliani* may form large populations and exhibit a high degree of outcrossing (Chapter 4), which facilitates the accumulation of genetic load due to the masking effects of heterozygosity. Genetic load contributes to SDR through post-zygotic selection against individuals carrying deleterious alleles. The analysis of F_1 populations generated from unrelated parental materials may provide a workaround for generating a genetic map in Maximilian sunflower with reduced segregation distortion.

The presence of co-segregating markers can rapidly inflate genetic maps and is a substantial challenge when working with next-generation sequencing technologies where the number of markers can easily exceed the resolution of recombination for a given population size (N'Diaye et al., 2017). Sequencing errors also contribute to map inflation. It is estimated that for every 1% of error in a given marker, the corresponding map increases in length by 2cM (Hackett and Broadfoot, 2003). While software packages such as Maskov are capable of imputing missing data and correcting for erroneous genotyping calls, they are dependent on a defined reference genome or linkage map order and known linkage phase (Ward et al., 2013).

Ascertainment bias can also complicate linkage and association mapping efforts through uneven genome coverage and the under-representation of parts of the genome. As reported above, relatively few SNPs were discovered corresponding to linkage groups 6 and 7 of *H. annuus* compared to other linkage groups. This result may be due to natural

variation in restriction enzyme cut sites and choice of restriction enzyme for generating reduced presentation libraries for GBS or ascertainment bias through the use of a potentially homeologous reference genome (Paris et al., 2017). Further SNPs may be recovered through the use of *de novo* SNP calling and may represent portions of the genome that have ancestrally diverged from *H. annuus.*

6.4.3 Agroecology of perennial grain systems

Perennial grain cropping systems are in their infancy and while they are currently not available for commercial scale production, cultivars of candidate species such as Kernza™ /intermediate wheatgrass (*Thinopyrum intermedium*) are anticipated to be available by the year 2019 (www.landinstitute.org). The management of perennial crops differs substantially from perennial forage and annual crops, and the development of agronomic packages to support their use as crops will be necessary to ensure yield potentials are achieved (Cattani and Asselin, 2018a; b). Perennial crops require different weed, pest and nutrient management practices than annual species, as do grain crops versus forage crops. Different cultural management practices may be necessary for the management of perennial grain cropping systems as some conventional tools for managing pathogen, pest or weed populations such as annual tillage and crop rotation are not applicable to perennial crops. Knowledge from both annual grain and perennial forage systems will aid in developing appropriate management practices. There are many proposed ecological benefits and services provided by growing perennial crops (Cox et al., 2006; Glover et al., 2010b; Kane et al., 2016), but less attention has been paid to

potential disservices (Cox et al., 2007; Pimentel et al., 2012). These will need to be understood to address the limitations of different candidate species. A more complete understanding of the strengths and limitations of perennial grain crops will provide answers to if and how they may be incorporated into existing crop rotations and to define the best environments for different practices.

6.5 Importance of the Ph.D. Work for the Advancement of Science

The development of perennial grain and oilseed cropping systems requires a multifaceted approach, being in part an exercise in the concepts of plant breeding and genetics, agronomy and ecology but also policy and economics (DeHaan et al., 2005, 2016; Bell et al., 2010; Pimentel et al., 2012). Few studies have examined the biology, agronomics or genetics of Maximilian sunflower and the research presented in this thesis provides a baseline for future investigations of Maximilian sunflower and related species as perennial oilseed crops. The variation and genomic resources identified in this study will aid in developing populations of Maximilian sunflower which differ in traits such as flowering time, branching, architecture, and capitulum size, and help define breeding pools for different environments. This thesis builds on previously established research of annual and perennial crop breeding and incorporates aspects of ecology and genomics to the challenge of breeding Maximilian sunflower as a perennial oilseed crop.

CHAPTER 7.0: REFERENCE MATTER

7.1 Literature Cited

Acharya, S.N., Z. Mir, and J.R. Moyer. 2004. ACE-1 perennial cereal rye. Can. J. Plant Sci. 84(3): 819–821.

Adebiyi, J., L.S. Olabisi, and S. Snapp. 2016. Understanding perennial wheat adoption as a transformative technology: evidence from the literature and farmers. Renew. Agric. Food Syst. 31(2): 101–110.

Allard, R.W., and A.D. Bradshaw. 1964. Implications of Genotype-Environmental Interactions in Applied Plant Breeding. Crop Sci. (4): 503–508.

Anderson, J.E., T.J. Kono, R.M. Stupar, M.B. Kantar, and P.L. Morrell. 2016. Environmental association analyses identify candidates for abiotic stress tolerance in *Glycine soja*, the wild progenitor of cultivated soybeans. G3 (Bethesda) 6:835–843. doi:10.1534/g3.116.026914

Anhalt, U.C.M., P. (J. S.) Heslop-Harrison, S. Byrne, A. Guillard, and S. Barth. 2008. Segregation distortion in *Lolium*: evidence for genetic effects. Theor. Appl. Genet. 117(2): 297–306. doi: 10.1007/s00122-008-0774-7.

Annicchiarico, P., B. Barrett, E.C. Brummer, B. Julier, and A.H. Marshall. 2014. Achievements and Challenges in Improving Temperate Perennial Forage Legumes. CRC. Crit. Rev. Plant Sci. 34(1–3): 327–380.

Antao, T., A. Lopes, R.J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on a FST-outlier method. BMC Bioinformatics 9:323. doi:10.1186/1471-2105-9-323

Applegate, R.D. 2015. Native plants provide equal or better nutrition than crop plants in wildlife plantings. Nativ. Plants J. 16(1): 28–36.

Armstrong, J.M. 1936. Hybridization of *Triticum* and *Agropyron*: I. Crossing results and description of the first generation hybrids. Can. J. Res. 14(5): 190–202.

Asbjornsen, H., V. Hernandez-Santan, M. Liebman, J. Bayala, J. Chen, M. Helmers, C.K. Ong and L.A. Schulte 2014. Targeting perennial vegetation in agricultural landscapes for enhancing ecosystem services. Renew. Agric. Food Syst. 29:101–125. doi:10.1017/S1742170512000385

Atlagić, J., B. Dozet, and D. Škorić. 1995. Meiosis and pollen grain viability in *Helianthus mollis*, *Helianthus salicifolius*, *Helianthus maximiliani* and their F1

hybrids with cultivated sunflower. Euphytica 81(3): 259–263. doi: 10.1007/BF00025615.

Baack, E.J., Y. Sapir, M.A. Chapman, J.M. Burke and L.H. Rieseberg. 2008. Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. Mol. Ecol. 17(2): 666-677.

Bachlava, E., S. Tang, G. Pizarro, G.F. Schuppert, R.K. Brunick, D. Draeger, A. Leon, V. Hahn, and S.J. Knapp. 2010. Pleiotropy of the branching locus (B) masks linked and unlinked quantitative trait loci affecting seed traits in sunflower. Theor. Appl. Genet. 120(4): 829–842.

Badouin, H., J. Gouzy, C.J. Grassa, F. Murat, S.E. Staton, L. Cottret, C. Lelandais-Brière, G.L. Owens, S. Carrere, B. Mayjonade, L. Legrand, N. Gill, N.C. Kane, J.E. Bowers, S. Hubner, A. Bellec, A. Bérard, H. Bergès, N. Blanchet, M.-C. Boniface, D. Brunel, O. Catrice, N. Chaidir, C. Claudel, C. Donnadieu, T. Faraut, G. Fievet, N. Helmstetter, M. King, S.J. Knapp, Z. Lai, M.-C. Le Paslier, Y. Lippi, L. Lorenzon, J.R. Mandel, G. Marage, G. Marchand, E. Marquand, E. Bret-Mestries, E. Morien, S. Nambeesan, T. Nguyen, P. Pegot-Espagnet, N. Pouilly, F. Raftis, E. Sallet, T. Schiex, J. Thomas, C. Vandecasteele, D. Varès, F. Vear, S. Vautrin, M. Crespi, B. Mangin, J.M. Burke, J. Salse, S. Muños, P. Vincourt, L.H. Rieseberg, and N.B. Langlade. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature 546(7656): 148–152.

Bagavathiannan, M. V., R.H. Gulden, G.S. Begg, and R.C. van Acker. 2010. The demography of feral alfalfa (*Medicago sativa* L.) populations occurring in roadside habitats in Southern Manitoba, Canada: Implications for novel trait confinement. Environ. Sci. Pollut. Res. 17(8): 1448–1459. doi: 10.1007/s11356-010-0330-2.

Baldanzi, M., M. Fambrini, and C. Pugliesi. 2003. Redesign of the castorbean plant body plan for optimal combine harvesting. Ann. Appl. Biol. 142(3): 299–306.

Balsalobre, T.W.A., G. da Silva Pereira, G.R.A. Margarido, R. Gazaffi, F.Z. Barreto, C.O. Anoni, C.B. Cardoso-Silva, E.A. Costa, M.C. Mancini, H.P. Hoffmann, A.P. de Souza, A.A.F. Garcia, and M.S. Carneiro. 2017. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. BMC Genomics 18(1): 72. doi: 10.1186/s12864-016-3383-x.

Banerjee, S., B.S. Yandell, and N. Yi. 2008. Bayesian Quantitative Trait Loci Mapping for Multiple Traits. Genetics 179(4): 2275–2289.

Barb, J.G., J.E. Bowers, S. Renaut, J.I. Rey, S.J. Knapp, L.H. Rieseberg, and J.M. Burke. 2014. Chromosomal evolution and patterns of introgression in *Helianthus*. Genetics 197:969–979. doi:10.1534/genetics.114.165548

Barker, M.S., N.C. Kane, M. Matvienko, A. Kozik, R.W. Michelmore, S.J. Knapp, and L.H. Rieseberg. 2008. Multiple paleopolyploidizations during the evolution of the *compositae* reveal parallel patterns of duplicate gene retention after millions of years. Mol. Biol. Evol. 25(11): 2445–2455. doi: 10.1093/molbev/msn187.

Barrett, S.C., and D. Charlesworth. 1991. Effects of a change in the level of inbreeding on the genetic load. Nature 352(6335): 522–524.

Barrett, S.C.H., L.D. Harder, and A.C. Worley. 1996. The Comparative Biology of Pollination and Mating in Flowering Plants. Philos. Trans. R. Soc. London B Biol. Sci. 351(1345): 1271–1280.

Bass, J., A. Dabney, and D. Robinson. 2016. qvalue: Q-value estimation for false discovery rate control. R package version 2.6.0. GitHub. http://github.com/jdstorey/qvalue (accessed 18 June 2018).

Baumgart-Getz, A., L.S. Prokopy, and K. Floress. 2012. Why farmers adopt best management practice in the United States: a meta-analysis of the adoption literature. J. Environ. Manage. 96(1): 17–25.

Baute, G.J., G.L. Owens, D.G. Bock, and L.H. Rieseberg. 2016. Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. Am. J. Bot. 103(12): 2170–2177.

Beaumont, M.A., and R.A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. Biol. Sci. 263:1619–1626. doi:10.1098/rspb.1996.0237

Bell, L.W., F. Byrne, M.A. Ewing, and L.J. Wade. 2008. A preliminary whole-farm economic analysis of perennial wheat in an Australian dryland farming system. Agric. Syst. 96(1–3): 166–174. doi: 10.1016/j.agsy.2007.07.007.

Bell, L.W., L.J. Wade, and M.A. Ewing. 2010. Perennial wheat: a review of environmental and agronomic prospects for development in Australia. Crop Pasture Sci. 61(9): 679–690.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57(1): 289–300. doi: 10.2307/2346101.

Besnier, F., and K.A. Glover. 2013. ParallelStructure: AR package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. PLoS One 8:e70651. doi:10.1371/journal.pone.0070651

Biligetu, B., P.G. Jefferson, R. Muri, and M.P. Schellenberg. 2014. Late summer forage yield, nutritive value, compatibility of warm- and cool-season grasses seeded with legumes in western Canada. Can. J. Plant Sci. 94(7): 1139–1148. doi: 10.4141/cjps2013-269.

Bilton, T.P., M.R. Schofield, M.A. Black, D. Chagné, P.L. Wilcox, and K.G. Dodds. 2018. Accounting for errors in low coverage high-throughput sequencing data when constructing genetic maps using Biparental outcrossed populations. Genetics 209(1): 65–76. doi: 10.1534/genetics.117.300627.

Binsfeld P.C., R. Wingender, and H. Schnabl. 2001. Cytogenetic analysis of interspecific sunflower hybrids and molecular evaluation of their progeny. Theor. Appl. Genet. 102(8): 1280–1285.

Blackman, B.K., J.L. Strasburg, A.R. Raduski, S.D. Michaels, and L.H. Rieseberg. 2010. The role of recently derived FT paralogs in sunflower domestication. Curr. Biol. 20(7): 629–635.

Blackman, B.K., D.A. Rasmussen, J.L. Strasburg, A.R. Raduski, J.M. Burke, S.J. Knapp, S.D. Michaels, and L.H. Rieseberg. 2011. Contributions of flowering time genes to sunflower domestication and improvement. Genetics 187(1): 271–287.

Blackman, B.K. 2013. Interacting duplications, fluctuating selection, and convergence: The complex dynamics of flowering time evolution during sunflower domestication. J. Exp. Bot. 64(2): 421–431. doi: 10.1093/jxb/ers359.

Blackshaw, R., K. Harker, J. O'donovan, and H.J. Beckie. 2008. Ongoing development of integrated weed management systems on the Canadian prairies. Weed Sci. 56(1): 146–150.

Bock, D.G., N.C. Kane, D.P. Ebert, and L.H. Rieseberg. 2014. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: Neither from Jerusalem nor an artichoke. New Phytol. 201:1021–1030. doi:10.1111/nph.12560

Bodénès, C., E. Chancerel, F. Ehrenmann, A. Kremer, and C. Plomion. 2016. High-density linkage mapping and distribution of segregation distortion regions in the oak genome. DNA Res. 23(2): 115–124. doi: 10.1093/dnares/dsw001.

Boelt, B., Julier, B., Karagić, Đ. And J. Hampton. 2015. Legume seed production meeting market requirements and economic impacts. Crit. Rev. Plant Sci. 34(1-3): 412-427.

Bogdanova, V.S., E.R. Galieva, A.K. Yadrikhinskiy, and O.E. Kosterin. 2012. Inheritance and genetic mapping of two nuclear genes involved in nuclear-cytoplasmic incompatibility in peas (*Pisum sativum* L.). Theor Appl Genet 124(8): 1503–1512. doi: 10.1007/s00122-012-1804-z.

Bomblies, K., and D. Weigel. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nat. Rev. Genet. 8(5): 382–393.

Bommarco, R., L. Marini, and B.E. Vaissière. 2012. Insect pollination enhances seed yield, quality, and market value in oilseed rape. Oecologia 169(4): 1025–1032.

Bommarco, R., D. Kleijn, and S.G. Potts. 2013. Ecological intensification: Harnessing ecosystem services for food security. Trends Ecol. Evol. 28(4): 230–238. doi: 10.1016/j.tree.2012.10.012.

Bowers, J.E., E. Bachlava, R.L. Brunick, L.H. Rieseberg, S.J. Knapp, and J.M. Burke. 2012a. Development of a 10,000 Locus Genetic Map of the Sunflower Genome Based on Multiple Crosses. G3 (Bethesda) 2(7): 721–729. doi: 10.1534/g3.112.002659.

Bowers, J.E., S. Nambeesan, J. Corbi, M.S. Barker, L.H. Rieseberg, S.J. Knapp, and J.M. Burke. 2012b. Development of an Ultra-Dense Genetic Map of the Sunflower Genome Based on Single-Feature Polymorphisms. PLoS One 7(12): e51360.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and S.E. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308

Bradbury, P., T. Parker, M.T. Hamblin, and J.L. Jannink. 2011. Assessment of power and false discovery rate in genome-wide association studies using the BarleyCAP germplasm. Crop Sci. 51:52–59. doi:10.2135/cropsci2010.02.0064

Brasileiro-Vidal, A.C., S. Brammer, M.J. Puertas, A.C. Zanatta, A. Prestes, M.I.B. Moraes-Fernandes, and M. Guerra. 2005. Mitotic instability in wheat x *Thinopyrum ponticum* derivatives revealed by chromosome counting, nuclear DNA content and histone H3 phosphorylation pattern. Plant Cell Rep. 24(3): 172–178. doi: 10.1007/s00299-005-0913-4.

Bratcher, C., J. Dole, and J. Cole. 1993. Stratification Improves Seed Germination of Five Native Wildflower Species. HortScience 28(9): 899–901.

Bremer, E., H.H. Janzen, and R.H. McKenzie. 2002. Short-term impact of fallow frequency and perennial grass on soil organic carbon in a Brown Chernozem in southern Alberta. Can. J. Soil Sci. 82(4): 481–488. doi: 10.4141/S02-007.

Breton. 2010. Gene Transfer From Wild Helianthus. Oléagineux, Corps gras, Lipides 17(2): 104–114.

Broman, K.W., H. Wu, and G.A. Churchill. 2003. R / qtl : QTL mapping in experimental crosses. Bioinformatics 19(7): 889–890. doi: 10.1093/bioinformatics/btg112.

Bruce, J. P., Frome, M., Haites, E., Janzen, H., Lal, R. and K. Pausitan. 1999. Carbon sequestration in soils. J. Soil Water Conserv. 54(1), 383-389.

Brunazzi, A., D. Scaglione, R.F. Talini, M. Miculan, F. Magni, J. Poland, M. Enrico Pè, A. Brandolini, and M. Dell'Acqua. 2018. Molecular diversity and landscape genomics of the crop wild relative *Triticum urartu* across the Fertile Crescent. Plant J. 94(4): 670–684. doi: 10.1111/tpj.13888.

Bu, Q., T. Lv, H. Shen, P. Luong, J. Wang, Z. Wang, Z. Huang, L. Xiao, C. Engineer, T.H. Kim and J.I. Schroeder. 2014. Regulation of drought tolerance by the F-box protein MAX2 in *Arabidopsis*. Plant Physiol. 164:424–439. doi:10.1104/pp.113.226837

Burke, J.M., S. Tang, S.J. Knapp, and L.H. Rieseberg. 2002. Genetic analysis of sunflower domestication. Genetics 161(3): 1257–1267.

Burke, J.M., S.J. Knapp, and L.H. Rieseberg. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. Genetics 171(4): 1933–1940.

Byers, D.L., and D.M. Waller. 1999. Do plant populations purge their genetic load? effects of population size and mating history on inbreeding depression. Annu. Rev. Ecol. Syst. 30(1): 479–513.

Cai, X., S.S. Jones, T.D. Murray, and W.E. Weber. 2001. Molecular cytogenetic characterization of *Thinopyrum* genomes conferring perennial growth habit in wheat-*Thinopyrum* amphiploids. Plant Breed. 120(1): 21–26.

Call, C.A., and D.W. Owens. 1986. Establishment of *Helianthus maximiliani* (*Asteraceae*) in Abandoned Cropland in the Post Oak Savannah of Texas. Southwest. Nat. 31(3): 367.

Calsbeek, J. 1984. Seed yield comparison of *Helianthus grossesseratus* x *H. maximiliani*. L. Rep. Research Suppl. 1: 19–20.

Campbell, C.A., G.P. Lafond, R.P. Zentner, and Y.W. Jame. 1994. Nitrate Leaching in a Udic Haploboroll as Influenced by Fertilization and Legumes. J. Environ. Qual. 23(1): 195–201.

Canvin, D.T. 1965. The effect of temperature on the oil content and fatty acid composition of the oils from several oil seed crops. Can. J. Bot. 43(1): 63–69.

Carpenter, S.R., H.A. Mooney, J. Agard, D. Capistrano, R.S. DeFries, S. Diaz, T. Dietz, A.K. Duraiappah, A. Oteng-Yeboah, H.M. Pereira, C. Perrings, W. V Reid, J.

Sarukhan, R.J. Scholes, and A. Whyte. 2009. Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. Proc. Natl. Acad. Sci. 106(5): 1305–1312. doi: 10.1073/pnas.0808772106.

Carrere, S., J. Gouzy, N. Langlade, P. Gamas, and P. Vincourt. 2008. HELIAGENE, a bioinformatics portal for *Helianthus* sp genomics. In: Proceeding of the 17th International Sunflower Conference, Cordoba, Spain. 8–12 June 2008. Int. Sunflower Assoc., Paris. p. 611–615.

Cartwright, D.A., M. Troggio, R. Velasco, and A. Gutin. 2007. Genetic mapping in the presence of genotyping errors. Genetics 176(4): 2521–2527. doi: 10.1534/genetics.106.063982.

Casler, M.D., K.P. Vogel, C.M. Taliaferro, N.J. Ehlke, J.D. Berdahl, E.C. Brummer, R.L. Kallenbach, C.P. West, and R.B. Mitchell. 2007. Latitudinal and longitudinal adaptation of switchgrass populations. Crop Sci. 47(6): 2249-2260. doi: 10.2135/cropsci2006.12.0780.

Casler, M.E., and E.C. Brummer. 2008. Theoretical Expected Genetic Gains for Among-and-Within-Family Selection Methods in Perennial Forage Crops. Crop Sci. 48(3): 890–902.

Cassman, K.G., A. Dobermann, D.T. Walters, and H. Yang. 2003. Meeting Cereals Demand While Protecting Natural Resources and Improving Environmental Quality. Annu. Rev. Environ. Resour. 28(1): 315–358. doi: 10.1146/annurev.energy.28.040202.122858.

Cattani, D.J. 2016. Selection of a perennial grain for seed productivity across years: Intermediate wheatgrass as a test species. Can. J. Plant Sci. 97(3): 516–524. doi: 10.1139/CJPS-2016-0280.

Cattani, D.J., and S.R. Asselin. 2018a. Has selection for grain yield altered intermediate wheatgrass? Sustainability 10(3):688-704. doi: 10.3390/su10030688.

Cattani, D.J., and S.R. Asselin. 2018b. Extending the growing season: forage seed production and perennial grains. Can. J. plant Sci. 98(2): 235–246. doi: 10.1139/cjps-2017-0212.

Cattell, R.B. 1966. The scree test for the number of factors. Multivariate Behav. Res. 1:245–276. doi:10.1207/s15327906mbr0102_10

Cavalieri, A., K.N. Harker, L.M. Hall, C.J. Willenborg, T.A. Haile, S.J. Shirtliffe, and R.H. Gulden. 2016. Evaluation of the causes of on-farm harvest losses in canola in the northern great plains. Crop Sci. 56(4): 2005–2015. doi: 10.2135/cropsci2016.01.0014.

Cerboncini, C., G. Beine, Binsfeld P.C., B. Dresen, H. Peisker, A. Zerwas, and H. Schnabl. 2002. Sources of resistance to *Sclerotinia sclerotiorum* (Lib) de Bary in a natural *Helianthus* gene pool. Helia 25(36): 167–176.

Canadian Food Inspection Agency. 2019. "Novelty" and Plant with Novel Traits. http://www.inspection.gc.ca/plants/plants-with-novel-traits/general-public/novelty/eng/1338181110010/1338181243773. Retrieved January 24th, 2019.

Chandler, J.M., and C.C. Jan. 1985. Comparison Of Germination Techniques for Wild *Helianthus* Seeds. Crop Sci. 25(2): 356–358.

Chang, S., C.S. Thurber, P.J. Brown, G.L. Hartman, K.N. Lambert, and L.L. Domier. 2014. Comparative Mapping of the Wild Perennial *Glycine latifolia* and Soybean (*G. max*) Reveals Extensive Chromosome Rearrangements in the Genus *Glycine*. PLoS One 9(6): e99427. doi: 10.1371/journal.pone.0099427.

Chapman, M.A., J.H. Leebens-Mack, and J.M. Burke. 2008a. Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. Mol. Biol. Evol. 25(7): 1260–1273.

Chapman, M.A., C.H. Pashley, J. Wenzler, J. Hvala, S. Tang, S.J. Knapp, and J.M. Burke. 2008b. A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). Plant Cell 20(11): 2931–2945.

Chapman, M.A., and J.M. Burke. 2012. Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower. Theor Appl Genet 125: 897–907. doi: 10.1007/s00122-012-1881-z.

Chapman, M.A., S. Tang, D. Draeger, S. Nambeesan, H. Shaffer, J.G. Barb, S.J. Knapp, and J.M. Burke. 2012. Genetic analysis of floral symmetry in Van Gogh's sunflowers reveals independent recruitment of CYCLOIDEA genes in the *Asteraceae*. PLoS Genet. 8(3): e1002628.

Charles, A.H. 1961. Differential survival of cultivars of *Lolium*, *Dactylis* and *Phleum*. Grass Forage Sci. 16(1): 69–75. doi: 10.1111/j.1365-2494.1961.tb00214.x.

Charlesworth, D., M.T. Morgan, and B. Charlesworth. 1990. Inbreeding Depression, Genetic Load, and the Evolution of Outcrossing Rates in a Multilocus System with No Linkage. Evolution (N. Y). 44(6): 1469.

Charlesworth, D. and J.H. Willis. The genetics of inbreeding depression. Nature Reviews Genetics. 10(11). 783-796.

Charlet, L.D., and G.J. Brewer. 1995. Resistance of Native Sunflowers (*Asterales: Asteraceae*) to the Banded Sunflower Moth (*Lepidoptera: Cochylidae*). Environ. Entomol. 24(5): 1224–1228.

Chase, C.D. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. 23(2): 81–90. doi: 10.1016/j.tig.2006.12.004.

Cheatham, M.R., M.N. Rouse, P.D. Esker, S. Ignacio, W. Pradel, R. Raymundo, A.H. Sparks, G.A. Forbes, T.R. Gordon, and K.A. Garrett. 2009. Beyond yield: plant disease in the context of ecosystem services. Phytopathology 99(11): 1228–1236.

Cheplick, G.P. 1995. Life history trade-offs in *Amphibromus scabrivalvis* (*Poaceae*): allocation to clonal growth, storage, and cleistogamous reproduction. Am. J. Bot. 82(5): 621–629.

Chubey, B.B., and D.G. Dorrell. 1974. Jerusalem Artichoke, a Potential Fructose Crop for the Prairies. Can. Inst. Food Sci. Technol. J. 7(2): 98–100. doi: 10.1016/S0315-5463(74)73870-6.

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold value for quantitative trait mapping. Genetics 138(3): 963–971. doi: 10.1534/genetics.107.080101.

Closset-Kopp, D., O. Chabrerie, B. Valentin, H. Delachapelle, and G. Decocq. 2007. When Oskar meets Alice: Does a lack of trade-off in r/K-strategies make *Prunus serotina* a successful invader of European forests? For. Ecol. Manage. 247(1–3): 120–130.

Conant, R.T., K. Paustian, and E.T. Elliott. 2001. Grassland Management and Conversion into Grassland: Effects on Soil Carbon. Ecol. Appl. 11(2): 343–355.

Corbineau, F., R.M. Rudnicki, and D. Come. 1988. Induction of secondary dormancy in sunflower seeds by high temperature. Possible involvement of ethylene biosynthesis. Physiol. Plant. 73(3): 368–373. doi: 10.1111/j.1399-3054.1988.tb00612.x.

Covarrubias-Pazaran, G., L. Diaz-Garcia, B. Schlautman, J. Deutsch, W. Salazar, M. Hernandez-Ochoa,E. Grygleski, S. Steffan, M. Iorizzo, J. Polasshock and N. Vorsa. 2016. Exploiting genotyping by sequencing to characterize the genomic structure of the American cranberry through high-density linkage mapping. BMC Genomics 17:451. doi:10.1186/s12864-016-2802-3

Cox, T.S., M. Bender, C. Picone, D.L. Van Tassel, J.B. Holland, E.C. Brummer, B.E. Zoeller, A.H. Paterson, and W. Jackson. 2002. Breeding Perennial Grain Crops. CRC. Crit. Rev. Plant Sci. 21(2): 59–91.

Cox, T.S., J.D. Glover, D.L. Van Tassel, C.M. Cox, and L.R. DeHaan. 2006. Prospects for Developing Perennial Grain Crops. Bioscience 56(8): 649–659. doi: 10.1641/0006-3568(2006)56[649:PFDPGC]2.0.CO;2.

Cox, C.M., K.A. Garrett, and W.W. Bockus. 2007. Meeting the challenge of disease management in perennial grain cropping systems. Renew. Agric. Food Syst. 20(1): 15–24.

Cox, T.S., D.L. Van Tassel, C.M. Cox, and L.R. Dehaan. 2010. Progress in breeding perennial grains. Crop Pasture Sci. 61(7): 513–521. doi: 10.1071/CP09201.

Craufurd, P.Q., and T.R. Wheeler. 2009. Climate change and the flowering time of annual crops. J. Exp. Bot. 60:2529–2539. doi:10.1093/jxb/erp196

Crews, T.E., and M.B. Peoples. 2004. Legume versus fertilizer sources of nitrogen: Ecological tradeoffs and human needs. Agric. Ecosyst. Environ. 102(3): 279–297. doi: 10.1016/j.agee.2003.09.018.

Crews, T.E. 2005. Perennial crops and endogenous nutrient supplies. Renew. Agric. Food Syst. 20:25–37. doi:10.1079/RAF200497

Crews, T.E., and P.C. Brookes. 2014. Changes in soil phosphorus forms through time in perennial versus annual agroecosystems. Agric. Ecosyst. Environ. 184:168–181. doi:10.1016/j.agee.2013.11.022.

Crews, T.E. and L.R. DeHaan, 2015. The strong perennial vision: A response. Agroecol. Sust. Food. 39(5):500-515.

Crews, T.E., J. Blesh, S.W. Culman, R.C. Hayes, E.S. Jensen, M.C. Mack, M.B. Peoples, and M.E. Schipanski. 2016. Going where no grains have gone before: From early to mid-succession. Agric. Ecosyst. Environ. 223: 223–238. doi: 10.1016/j.agee.2016.03.012.

Crews, T.E., and D.J. Cattani. 2018. Strategies, advances, and challenges in breeding perennial grain crops. Sustain. 10(7). doi: 10.3390/su10072192.

Culman, S.W., S.T. DuPont, J.D. Glover, D.H. Buckley, G. Fick, H. Ferris, and C. T.E. 2010. Long-term impacts of high-input annual cropping and unfertilized perennial grass production on soil properties and belowground food webs in Kansas, USA. Agric. Ecosyst. Environ. 137(1–2): 13–24.

Culman, S.W., S.S. Snapp, M. Ollenburger, B. Basso, and L.R. DeHaan. 2013. Soil and water quality rapidly responds to the perennial grain Kernza wheatgrass. Agron. J. 105(3): 735–744. doi: 10.2134/agronj2012.0273.

Cumaraswamy, A., and K.S. Bawa. 1989. Sex allocation and mating systems in pigeonpea, *Cajanus cajan* (*Fabaceae*). Plant Syst. Evol. 168(1–2): 59–69.

Daily, G.C., T.H. Ricketts, S. Polasky, J. Salzman, J. Goldstein, P.M. Kareiva, and R. Shallenberger. 2009. Ecosystem services in decision making: time to deliver. Front Ecol Env. 7: 21–28. doi: 10.1890/080025.

Davies, C.L., D.L. Waugh, and E.C. Lefroy. 2005. Variation in seed yield and its components in the Australian native grass *Microlaena stipoides* as a guide to its potential as a perennial grain crop. Aust. J. Agric. Res. 56(3): 308–309.

De Mita, S., A.C. Thuillet, L. Gay, N. Ahmadi, S. Manel, J. Ronfort, and Y. Vigouroux. 2013. Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol. Ecol. 22:1383–1399. doi:10.1111/mec.12182

De Villemereuil, P., O. Gaggiotti, M. Mouterde, and I. Till-Bottraud. 2015. Common garden experiments in the genomic era: new perspectives and opportunities. Heredity (Edinb). 116(3): 249–254. doi: 10.1038/hdy.2015.93.

Dechaine, J.M., J.C. Burger, M.A. Chapman, G.J. Seiler, R. Brunick, S.J. Knapp, and J.M. Burke. 2009. Fitness effects and genetic architecture of plant–herbivore interactions in sunflower crop–wild hybrids. New Phytol. 184:828–841. doi:10.1111/j.1469-8137.2009.02964.x

Dedio, W. 1980. Comparison of achene characteristics and combining ability of branching and non-branching near-isogenic sunflower restorer lines. Crop Sci. 20: 189–190. doi: 10.2135/cropsci1980.0011183X002000020010x.

DeHaan, L.R., D.L. Van Tassel, and T.S. Cox. 2005. Perennial grain crops: A synthesis of ecology and plant breeding. Renew. Agric. Food Syst. 20:5–14. doi:10.1079/RAF200496

DeHaan, L.R., and D.L. Van Tassel. 2014. Useful insights from evolutionary biology for developing perennial grain crops. Am. J. Bot. 101(10): 1801–1819.

DeHaan, L.R., S. Wang, S.R. Larson, D.J. Cattani, X. Zhang, and T. Kantarski. 2014. Current efforts to develop perennial wheat and domesticate *Thinopyrum intermedium* as a perennial grain. p. 72–89. *In* Batello, C., Wade, L., Cox, S., Pogna, N., Bozzini, A., Choptiany, J. (eds.), Perennial Crops for Food Security; Proceedings of the FAO Expert Workshop. 28-30 Aug. 2013. FAO, Rome, Italy.

DeHaan, L.R., D.L. Van Tassel, J.A. Anderson, S.R. Asselin, R. Barnes, G.J. Baute, D.J. Cattani, S.W. Culman, K.M. Dorn, B.S. Hulke, M. Kantar, S. Larson, M. David Marks, A.J. Miller, J. Poland, D.A. Ravetta, E. Rude, M.R. Ryan, D. Wyse, and X.

Zhang. 2016. A pipeline strategy for grain crop domestication. Crop Sci. 56(3): 917–930. doi: 10.2135/cropsci2015.06.0356.

DeHaan, L., M. Christians, J. Crain, and J.A. Poland. 2018. Development and Evolution of an Intermediate Wheatgrass Domestication Program. Sustainability 10(5). doi: 10.3390/su10051499.

Dempewolf, H., L.H. Rieseberg, and Q.C. Cronk. 2008. Crop domestication in the *Compositae*: a family-wide trait assessment. Genet. Resour. Crop Evol. 55(8): 1141–1157.

Dempewolf, H., R.J. Eastwood, L. Guarino, C.K. Khoury, J.V. Müller and J. Toll. 2014. Adapting agriculture to climate change: a global initiative to collect, conserve and use crop wild relatives. Agroecol. Sust. Food. 38(4):369-377. doi:10.1080/21683565.2013.870629

Dempewolf, H., M. Tesfaye, A. Teshome, A.D. Bjorkman, R.L. Andrew, M. Scascitelli, S. Black, E. Bekele, J.M.M. Engels, Q.C.B. Cronk, and L.H. Rieseberg. 2015. Patterns of domestication in the Ethiopian oil-seed crop noug (*Guizotia abyssinica*). Evol. Appl. 8(5): 464–475.

Dempewolf, H., G. Baute, J. Anderson, B. Kilian, C. Smith, and L. Guarino. 2017. Past and future use of wild relatives in crop breeding. Crop Sci. 57:1070–1082. doi:10.2135/cropsci2016.10.0885

Demurin, Y., D. Škorić, and D. Karlovic. 1996. Genetic variability of tocopherol composition in sunflower seeds as a basis of breeding for improved oil quality. Plant Breed. 115(1): 33–36.

Des Marais, D.L., K.M. Hernandez, and T.E. Juenger. 2013. Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. Annu. Rev. Ecol. Evol. Syst. 44(1): 5–29. doi: 10.1146/annurev-ecolsys-110512-135806.

Devlin, B., and K. Roeder. 1999. Genomic Control for Association Studies. Biometrics 55(4): 997–1004.

Dickson, T.L., and W.H. Busby. 2009. Forb Species Establishment Increases with Decreased Grass Seeding Density and with Increased Forb Seeding Density in a Northeast Kansas, U.S.A., Experimental Prairie Restoration. Restor. Ecol. 17(5): 597–605.

Dietz, D.R., C.H. Wasser, P.L. Dittberner, and C.O. Martin. 1992. Defense Natural Resources Program: Maximilian Sunflower (*Helianthus maximiliani*), Section 7.4. 3, US Army Corps of Engineers Wildlife Resources Management.

Dinno, A. 2012. paran: Horn's test of principal components/factors. R package version 1.1.

Do Canto, J., B. Studer, and T. Lubberstedt. 2016. Overcoming self-incompatibility in grasses: a pathway to hybrid breeding. Theor. Appl. Genet. 129(10): 1815–1829. doi: 10.1007/s00122-016-2775-2.

Do Canto, J., B. Studer, U. Frei, and T. Lübberstedt. 2018. Fine mapping a self-fertility locus in perennial ryegrass. Theor. Appl. Genet. 131(4): 817–827. doi: 10.1007/s00122-017-3038-6.

Doebley, J., A. Stec, and L. Hubbard. 1997. The evolution of apical dominance in maize. Nature 386: 485–488.

Dohleman, F.G., and S.P. Long. 2009. More Productive Than Maize in the Midwest: How Does *Miscanthus* Do It? PLANT Physiol. 150(4): 2104–2115. doi: 10.1104/pp.109.139162.

Dorrell, D.G., and E.D.P. Whelan. 1978. Chemical and morphological characteristics of seeds of some sunflower species. Crop Sci. 18:969–971. doi:10.2135/cropsci1978.0011183X001800060015x

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

Dragana, V., Š. Dragan, G. Alibert, and M. Vladimir. 2001. Micropropagation of *Helianthus maximiliani* (Schrader) by shoot apex culture. Helia 24(34): 63–68.

Earl, Dent A. and von Holdt, Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources vol. 4 (2):359-361 doi: 10.1007/s12686-011-9548-7

Eckert, A.J., J. van Heerwaarden, J.L. Wegrzyn, C.D. Nelson, J. Ross-Ibarra, S.C. González-Martínez, and D.B. Neale. 2010. Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., *Pinaceae*). Genetics 185(3): 969–982.

Edae, E.A., P.D. Olivera, Y. Jin, J.A. Poland, and M.N. Rouse. 2016. Genotype-by-sequencing facilitates genetic mapping of a stem rust resistance locus in *Aegilops umbellulata*, a wild relative of cultivated wheat. BMC Genomics 17(1): 1039. doi: 10.1186/s12864-016-3370-2.

Edwards, M.D., C.W. Stuber, and J.F. Wendel. 1987. Molecular-marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116(1): 113–125.

Egli, D.B. 2011. Time and the productivity of agronomic crops and cropping systems. Agron. J. 103(3): 743–750. doi: 10.2134/agronj2010.0508.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5): e19379.

Engelmann, K., and M. Purugganan. 2006. The molecular evolutionary ecology of plant development: Flowering time in *Arabidopsis thaliana*. Adv. Bot. Res. Inc. Adv. Plant Pathol. 44(6): 507–526. doi: 10.1016/s0065-2296(06)44013-1.

Entz, M.H., W.J. Bullied, D.A. Forster, R. Gulden, and J.K. Vessey. 2001. Extraction of subsoil nitrogen by alfalfa, alfalfa-wheat, and perennial grass systems. Agron. J. 93(3): 495–503. doi: 10.2134/agronj2001.933495x.

Espinasse, A., J. Volin, C.D. Dybing, and C. Lay. 1991. Embryo Rescue through in Ovulo Culture in *Helianthus*. Crop Sci. 31(1): 102–108.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Ewing, B., L. Hillier, M.C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. 8:175–185. doi:10.1101/gr.8.3.175

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4th ed. Longmans Green, Harlow, UK.

Fang, Z., A.M. Gonzales, M.T. Clegg, K.P. Smith, G.J. Muehlbauer, B.J. Steffenson, and P.L. Morrell. 2014. Two genomic regions contribute disproportionately to geographic differentiation in wild barley. G3 (Bethesda) 4:1193–1203. doi:10.1534/g3.114.010561

Fargione, J., J. Hill, D. Tilman, S. Polasky, and P. Hawthorne. 2008. Land Clearing and the Biofuel Carbon Debt. Science. 319(5867): 1235–1238.

Farnsworth, K., and K. Niklas. 1995. Theories of optimization, form and function in branching architecture in plants. Funct. Ecol. (United Kingdom) 9(3): 355–363. doi: 10.2307/2389997.

Feng, J., and C.C. Jan. 2008. Introgression and molecular tagging of *Rf 4*, a new male fertility restoration gene from wild sunflower *Helianthus maximiliani* L. Theor. Appl. Genet. 117:241-249. doi:10.1007/s00122-008-0769-4

Ferchaud, F., G. Vitte, F. Bornet, L. Strullu, and B. Mary. 2014. Soil water uptake and root distribution of different perennial and annual bioenergy crops. Plant Soil 388(1–2): 307–322.

Ferguson, B.J., and C.A. Beveridge. 2009. Roles for Auxin, Cytokinin, and Strigolactone in Regulating Shoot Branching. Plant Physiol. 149(4): 1929–1944. doi: 10.1104/pp.109.135475.

Fick, G.N., D.E. Zimmer, and D.C. Zimmerman. 1974. Correlation of seed oil content in sunflowers with other plant and seed characteristics. Crop Sci. 14: 755–757.

Fick, G.N., and J.F. Miller. 1997. Sunflower Breeding. p. 395–439. *In* Schneider, A. (ed.), Sunflower technology and production. ASA, Madison, WI.

Fisher, R.A. 1954. A fuller theory of "junctions" in inbreeding. Heredity (Edinb). 8(2): 187–197. http://dx.doi.org/10.1038/hdy.1954.17.

Flint-Garcia, S.A., J.M. Thornsberry, and E.S. Buckler, IV. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54:357–374. doi:10.1146/annurev.arplant.54.031902.134907

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977–993. doi:10.1534/genetics.108.092221

Foll, M. 2012. BayeScan v2. 1 user manual. Ecology 20:1450–1462.

Fowler, D.B. 2012. Wheat production in the high winter stress climate of the great plains of North America: An experiment in crop adaptation. Crop Sci. 52:11–20. doi:10.2135/cropsci2011.05.0279

François, O., H. Martins, K. Caye, and S.D. Schoville. 2016. Controlling false discoveries in genome scans for selection. Mol. Ecol. 25:454–469. doi:10.1111/mec.13513

Friesen, M.L., E.J.B. von Wettberg, M. Badri, K.S. Moriuchi, F. Barhoumi, P.L. Chang, S. Cuellar-Ortiz, M.A. Cordeiro, W.T. Vu, S. Arraouadi, N. Djébali, K. Zribi, Y. Badri, S.S. Porter, M.E. Aouani, D.R. Cook, S.Y. Strauss, and S. V Nuzhdin. 2014. The ecological genomic basis of salinity adaptation in Tunisian *Medicago truncatula*. BMC Genomics 15(1): 1160.

Gaj, T. 2014. ZFN, TALEN and CRISPR/Cas based methods for genome engineering. 2013 31(7): 397–405. doi: 10.1016/j.tibtech.2013.04.004.ZFN.

Gallard, C. 1985. Contribution a l'etude de la tuberisation chez le topinambour (*Helianthus tuberosus* L.). *In* Memoire de D.E.A. Sciences Agronomiques. Universite Rennes I, Rennes, France.

Galwey, N.W., K. Adhikari, M. Dracup, and R. Thomson. 2003. Agronomic potential of genetically diverse narrow-leafed lupins (*Lupinus angustifolius* L.) with restricted branching. Crop Pasture Sci. 54(7): 649–661.

Gandhi, S.D., A.F. Heesacker, C.A. Freeman, J. Argyris, K. Bradford, and S.J. Knapp. 2005. The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. Theor. Appl. Genet. 111(4): 619–629.

Garnier, E. 1992. Growth Analysis of Congeneric Annual and Perennial Grass Species. J. Ecol. 80(4): 665.

Gay, C., F. Corbineau, and D. Côme. 1991. Effects of temperature and oxygen on seed germination and seedling growth in sunflower (*Helianthus annuus* L.). Environ. Exp. Bot. 31(2): 193–200.

Gershenzon, J., and T.J. Mabry. 1984. Sesquiterpene lactones from a Texas population of *Helianthus maximiliani*. Phytochemistry 23(9): 1959–1966.

Gianola, D., and D. Sorensen. 2004. Quantitative Genetic Models for Describing Simultaneous and Recursive Relationships Between Phenotypes. Genetics 167(3): 1407–1424.

Glover, J.D. 2005. The necessity and possibility of perennial grain production systems. Renew. Agric. Food Syst. 20(01): 1–4. doi: 10.1079/RAF200499.

Glover, J.D., S.W. Culman, S.T. DuPont, W. Broussard, L. Young, M.E. Mangan,J.G. Mai, T.E. Crews, L.R. DeHaan, D.H. Buckley and H. Ferris. 2010a. Harvested perennial grasslands provide ecological benchmarks for agricultural sustainability. Agric. Ecosyst. Environ. 137:3–12. doi:10.1016/j.agee.2009.11.001

Glover, J.D., J.P. Reganold, L.W. Bell, J. Borevitz, E.C. Brummer, E.S. Buckler, C.M. Cox, T.S. Cox, T.E. Crews, S.W. Culman, and L.R. DeHaan 2010b. Increased food and ecosystem security via perennial grains. Science 328:1638–1639. doi:10.1126/science.1188761

González-Paleo, L., A.E. Vilela, and D.A. Ravetta. 2016. Back to perennials: Does selection enhance tradeoffs between yield and longevity? Ind. Crops Prod. 91: 272–278.

Gordon-Werner, E., and K. Dörffling. 1988. Morphological and Physiological Studies Concerning the Drought Tolerance of the *Secale cereale* × *Secale montanum* Cross

Permontra. J. Agron. Crop Sci. 160(4): 277–285. doi: 10.1111/j.1439-037X.1988.tb00330.x.

Grattapaglia, D., and R. Sederoff. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137(4): 1121–37. doi: 10.1007/s11033-010-0612-2.

Griffin, L.C., and R.M. Rowlett. 1981. A "lost" Viking cereal grain. J. Ethnobiol. 1(2): 200–207.

Gross, B.L., and K.M. Olsen. 2010. Genetic perspectives on crop domestication. Trends Plant Sci. 15(9): 529–537.

Guajardo, V., S. Solís, B. Sagredo, F. Gainza, and C. Mu. 2015. Construction of High Density Sweet Cherry (*Prunus avium* L .) Linkage Maps Using Microsatellite Markers and SNPs Detected by Genotyping-by-Sequencing (GBS). PLoS One 10(5): 1–18. doi: 10.1371/journal.pone.0127750.

Gulden, R.H., S.J. Shirtliffe, and A.G. Thomas. 2003. Harvest losses of canola (*Brassica napus*) cause large seedbank inputs. Weed Sci. 51(1): 83–86. doi: 10.1614/0043-1745(2003)051[0083:HLOCBN]2.0.CO;2.

Gunnarsson, I.B., S.E. Svensson, E. Johansson, D. Karakashev, and I. Angelidaki. 2014. Potential of Jerusalem artichoke (*Helianthus tuberosus* L.) as a biorefinery crop. Ind. Crops Prod. 56: 231–240.

Guo, B., D. Wang, Z. Guo, and W.D. Beavis. 2013. Family-based association mapping in crop species. Theor. Appl. Genet. 126:1419–1430. doi:10.1007/s00122-013-2100-2

Gupta, P.K., S. Rustgi, and P.L. Kulwal. 2005. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol. Biol. 57(4):461–485. doi:10.1007/s11103-005-0257-z

Gutierrez, A., A. Carrera, J. Basualdo, R. Rodriguez, M. Cantamutto, and M. Poverene. 2010. Gene flow between cultivated sunflower and *Helianthus petiolaris* (*Asteraceae*). Euphytica 172(1): 67–76. doi: 10.1007/s10681-009-0045-y.

Hackett, C.A., and L.B. Broadfoot. 2003. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity (Edinb). 90(1): 33–38. doi: 10.1038/sj.hdy.6800173.

Hahn, V., and W. Friedt. 1994. Molecular analysis of the cms-inducing MAX1 cytoplasm in sunflower. Theor. Appl. Genet. 89–89(2–3).

Haley, C.S., and S.A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity (Edinb). 69(4): 315–324. doi: 10.1038/hdy.1992.131.

Hamlin, J.A.P., and M.L. Arnold. 2015. Neutral and Selective Processes Drive Population Differentiation for *Iris hexagona*. J. Hered. 106(5): 628–636.

Hanson, W.D. 1959. The Breakup of Initial Linkage Blocks under Selected Mating Systems. Genetics 44(5): 857–868.

Harlan, J.R. 1992. Crops and man. Madison: American Society of Agronomy, Inc. Crop Science Society of America.

Hartl, D.L., and A.G. Clark. 1997. Principles of population genetics. Vol. 116. Sinauer Assoc., Sunderland, MA.

Hartung, F., and J. Schiemann. 2014. Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU. Plant J. 78(5): 742–752.

Haselhorst, M.S.H., C.E. Edwards, M.J. Rubin, and C. Weinig. 2011. Genetic architecture of life history traits and environment-specific trade-offs. Mol. Ecol. 20(19): 4042–4058.

Hayes, R.C., S. Wang, M.T. Newell, K. Turner, J. Larsen, L. Gazza, J.A. Anderson, L.W. Bell, D.J. Cattani, K. Frels, E. Galassi, A.I. Morgounov, C.K. Revell, D.B. Thapa, E.J. Sacks, M. Sameri, L.J. Wade, A. Westerbergh, V. Shamanin, A. Amanov, and G.D. Li. 2018. The performance of early-generation perennial winter cereals at 21 sites across four continents. Sustain. 10(4). doi: 10.3390/su10041124.

Heiser, C.B. 1951. The sunflower among the North American Indians. Proc. Am. Philos. Soc. 95(4): 432–448.

Heiser, C.B. 1954. Variation and subspeciation in the common sunflower *Helianthus annuus*. Am. Midl. Nat. 51(1): 287–305.

Heiser, C.B., W.C. Martin, and D.M. Smith. 1962. Species crosses in *Helianthus*: I. Diploid species. Brittonia 14(2): 137–147.

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.

Henderson, D.C., and M.A. Naeth. 2005. Multi-scale impacts of crested wheatgrass invasion in mixed-grass prairie. Biol. Invasions 7(4): 639–650.

Hengl, T., J.M. de Jesus, R.A. MacMillan, N.H. Batjes, G.B.M. Heuvelink, E. Ribeiro,

A. Samuel-Rosa, B. Kempen, J.G.B. Leenaars, M.G. Walsh, and M.R. Gonzalez. 2014. SoilGrids1km — Global Soil Information Based on Automated Mapping. PLoS One 9(8): e105992. doi: 10.1371/journal.pone.0105992.

Henn, H.J., R. Wingender, and H. Schnabl. 1998. Regeneration of fertile interspecific hybrids from protoplast fusions between *Helianthus annuus* L. and wild *Helianthus* species. Plant Cell Rep. 18(3–4): 220–224.

Henning, J.A., D.H. Gent, M.C. Twomey, M.S. Townsend, N.J. Pitra, and P.D. Matthews. 2016. Genotyping-by-sequencing of a bi-parental mapping population segregating for downy mildew resistance in hop (*Humulus lupulus* L.). Euphytica 208(3): 545–559. doi: 10.1007/s10681-015-1600-3.

Henning, J., S. Hill, P. Darby, and D. Hendrix. 2017. QTL examination of a bi-parental mapping population segregating for "short-stature" in hop (*Humulus lupulus* L.). Euphytica 213(3). doi: 10.1007/s10681-017-1848-x.

Henry, R.J. 2012. Next-generation sequencing for understanding and accelerating crop domestication. Brief. Funct. Genomics 11(1): 51–56.

Henry, L.P., R.H.B. Watson, and B.K. Blackman. 2014. Transitions in photoperiodic flowering are common and involve few loci in wild sunflowers (*Helianthus*; *Asteraceae*). Am. J. Bot. 101(10): 1748–1758.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25(15): 1965–1978.

Hijmans, R.J., J. van Etten, and J. Cheng. 2016. Package "raster". R Found. Stat. Comput. https://cran.r-project.org/web/packages/raster/index.html (accessed 18 June 2018).

Hill, W.G., and B.S. Weir. 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33:54–78. doi:10.1016/0040-5809(88)90004-4

Hill, R.D. 2010. The cultivation of perennial rice, an early phase in Southeast Asian agriculture? J. Hist. Geogr. 36(2): 215–223.

Holland, J.B. 2007. Genetic architecture of complex traits in plants. Curr. Opin. Plant Biol. 10(2): 156–161.

Honig, J.A., C. Kubik, M. Majewski, C. Poulsen, E. Weibel, K. Amundsen, S.E. Warnke, W.A. Meyer, and S.A. Bonos. 2014. A PCR-based linkage map of *Agrostis stolonifera* and identification of QTL markers for dollar spot resistance. Mol. Breed. 34(1): 185–203.

Hopwood, J.L. 2008. The contribution of roadside grassland restorations to native bee conservation. Biol. Conserv. 141(10): 2632–2640.

Horn, J.L. 1965. A rationale and test for the number of factors in factor analysis. Psychometrika. 30(2):179-185. doi: 10.1007/BF02289447.

Horvitz, C.C., and D.W. Schemske. 1988. Demographic Cost of Reproduction in a Neotropical Herb: An Experimental Field Study. Ecology 69(6): 1741–1745.

Hu, F., D. Tao, P. Xu, J. Li, and Y. Yang. 2001. Two dominant complementary genes controlling rhizomatous expression in *Oryza longistaminata*. Rice Genet. Newsl. 18: 34–36. http://www.gramene.org/newsletters/rice_genetics/rgn18/c14.html (accessed 25 September 2018).

Hutchinson, J.J., C.A. Campbell, and R.L. Desjardins. 2007. Some perspectives on carbon sequestration in agriculture. Agric. For. Meteorol. 142(2–4): 288–302. doi: 10.1016/j.agrformet.2006.03.030.

Isbell, F., V. Calcagno, A. Hector, J. Connolly, W.S. Harpole, P.B. Reich, M. Scherer-Lorenzen, B. Schmid, D. Tilman, J. van Ruijven, A. Weigelt, B.J. Wilsey, E.S. Zavaleta, and M. Loreau. 2011. High plant diversity is needed to maintain ecosystem services. Nature 477(7363): 199–202.

Jackson, W. 1990. Agriculture with nature as analogy. p. 381–422. *In* Francis, C.A., Butler, C., King, L.D. (eds.), Sustainable agriculture in Temperate Zones. John Wiley & Sons, Inc., New York, NY.

Jackson, D.A. 1993. Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. Ecology 74:2204–2214. doi:10.2307/1939574

Jackson, L.L., and C.L. Dewald. 1994. Predicting Evolutionary Consequences of Greater Reproductive Effort in *Tripsacum Dactyloides*, a Perennial Grass. Ecology 75(3): 627–641.

Jaikumar, N.S., S.S. Snapp, and T.D. Sharkey. 2013. Life history and resource acquisition: Photosynthetic traits in selected accessions of three perennial cereal species compared with annual wheat and rye. Am. J. Bot. 100:2468–2477. doi:10.3732/ajb.1300122

Jaikumar, N.S., S.S. Snapp, J.A. Flore, and W. Loescher. 2014. Photosynthetic Responses in Annual Rye, Perennial Wheat, and Perennial Rye Subjected to Modest Source:Sink Ratio Changes. Crop Sci. 54(1): 274–283.

Jaikumar, N.S., S.S. Snapp, and T.D. Sharkey. 2016. Older *Thinopyrum intermedium* (*Poaceae*) plants exhibit superior photosynthetic tolerance to cold stress and greater

increases in two photosynthetic enzymes under freezing stress compared with young plants. J. Exp. Bot. 67:4743–4753. doi:10.1093/jxb/erw253

Jan, C.C. 1992. Cytoplasmic-Nuclear Gene Interaction for Plant Vigor in *Helianthus* Species. Crop Sci. 32(2): 320–323.

Jan, C.C., and T.X. Zhang. 1994. Fertility restoration of a cytoplasmic male-sterile plant identified in a *Helianthus maximiliani* Schrad. Population. In: Agronomy abstracts. ASA, Madison, WI. p. 129.

Jan, C.C., and J.M. Fernández-Martínez. 2002. Interspecific hybridization, gene transfer, and the development of resistance to the broomrape race F in Spain. Helia 25(36): 123–136.

Jan, C.C., J. Feng, G.J. Seiler, and K.Y. Rashid. 2007. Development of Sclerotinia Head Rot Resistant Germplasm Utilizing *H. maximiliani* and *H. nuttallii*. *In* 29th Sunflower Research Workshop, January 10-11, 2007. Fargo, ND.

Jan, C.C., G.J. Seiler, and J.J. Hammond. 2014. Effect of wild *Helianthus* cytoplasms on agronomic and oil characteristics of cultivated sunflower (*Helianthus annuus* L.). Plant Breed.

Janzen, H.H., C.A. Campbell, R.C. Izaurralde, B.H. Ellert, N. Juma, W.B. McGill, and R.P. Zentner. 1998. Management effects on soil C storage on the Canadian prairies. Soil Tillage Res. 47(3–4): 181–195.

Jefferson, R.G., A.D. Iwaasa, M.R. Schellenberg, and J.G. Mcleod. 2013. Re-evaluation of native plant species for seeding and grazing by livestock on the semiarid prairie of western Canada. Prairie Forum 38: 275–304.

Jenkinson, D.S., J.M. Potts, J.N. Perry, V. Barnett, K. Coleman, and A.E. Johnston. 1994. Trends in herbage yields over the last century on the Rothamsted Long-term Continuous Hay Experiment. J. Agric. Sci. 122(3): 365–374.

Jenkinson, D.S., P.R. Poulton, A.E. Johnston, and D.S. Powlson. 2004. Turnover of nitrogen-15-labeled fertilizer in old grassland. Soil Sci. Soc. Am. J. 68:865–875. doi:10.2136/sssaj2004.8650

Jennersten, O. 1991. Cost of Reproduction in *Viscaria vulgaris* (*Caryophyllaceae*): A Field Experiment. Oikos 61(2): 197.

Johnson, L.P. V. 1938. Hybridization of *Triticum* and *Agropyron*: IV. Further crossing results and studies on the F 1 hybrids. Can. J. Res. 16(10):417-444.

Johnston, A., A.D. Smith, L.E. Lutwick, and S. Smoliak. 1968. Fertilizer response of native and seeded ranges. Can. J. Plant Sci. 48(5): 467–472. doi: 10.4141/cjps68-092.

Jones, T.A., X.Y. Zhang, and R.R.C. Wang. 1999. Genome characterization of MT-2 perennial and OK-906 annual wheat x intermediate wheatgrass hybrids. Crop Sci. 39(4): 1041–1043. doi: 10.2135/cropsci1999.0011183X003900040013x.

Joost, S., S. Vuilleumier, J.D. Jensen, S. Schoville, K. Leempoel, S. Stucki, I. Widmer, C. Melodelima, J. Rolland, and S. Manel. 2013. Uncovering the genetic basis of adaptive change: On the intersection of landscape genomics and theoretical population genetics. Mol. Ecol. 22(14): 3659–3665. doi: 10.1111/mec.12352.

Kakeda, K., T. Ibuki, J. Suzuki, H. Tadano, Y. Kurita, Y. Hanai, and Y. Kowyama. 2008. Molecular and genetic characterization of the S locus in *Hordeum bulbosum* L., a wild self-incompatible species related to cultivated barley. Mol. Genet. Genomics 280(6): 509–519.

Kane, D.A., P. Rogé, and S.S. Snapp. 2016. A Systematic Review of Perennial Staple Crops Literature Using Topic Modeling and Bibliometric Analysis. PLoS One 11(5): e0155788.

Kane, N.C., and L.H. Rieseberg. 2008. Genetics and evolution of weedy *Helianthus annuus* populations: Adaptation of an agricultural weed. Mol. Ecol. 17:384–394. doi:10.1111/j.1365-294X.2007.03467.x

Kane, N.C., 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(7): 429–437. doi: 10.1139/b11-032.

Kang, Y., M. Sakiroglu, N. Krom, J. Stanton-Geddes, M. Wang, Y.C. Lee, N.D. Young, and M. Udvardi. 2015. Genome-wide association of drought-related and biomass traits with HapMap SNPs in *Medicago truncatula*. Plant, Cell Environ. 38(10): 1997–2011. doi: 10.1111/pce.12520.

Kantar, M.B., K. Betts, J.M. Michno, J.J. Luby, P.L. Morrell, B.S. Hulke, R.M. Stupar, and D.L. Wyse. 2014. Evaluating an interspecific *Helianthus annuus* × *Helianthus tuberosus* population for use in a perennial sunflower breeding program. F. Crop. Res. 155: 254–264. doi: 10.1016/j.fcr.2013.04.018.

Kantar, M.B., C.C. Sosa, C.K. Khoury, N.P. Castañeda-Álvarez, H.A. Achicanoy, V. Bernau, N.C. Kane, L. Marek, G. Seiler, and L.H. Rieseberg. 2015. Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* L.). Front. Plant Sci. 6: 20132980. doi: 10.3389/fpls.2015.00841.

Kantar, M.B., C.E. Tyl, K.M. Dorn, X. Zhang, J.M. Jungers, J.M. Kaser, R.R. Schendel, J.O. Eckberg, B.C. Runck, M. Bunzel, N.R. Jordan, R.M. Stupar, M.D. Marks, J.A. Anderson, G.A. Johnson, C.C. Sheaffer, T.C. Schoenfuss, B. Ismail, G.E. Heimpel, and D.L. Wyse. 2016. Perennial Grain and Oilseed Crops. Annu. Rev. Plant Biol. 67: 703–729.

Kantar, M., S. Hüber, A. Herman, D. Bock, G. Baute, K. Betts, M. Ott, Y. Brandvain, D. Wyse, R. Stupar, and L. Rieseberg. 2018. Neo-Domestication of an Interspecific Tetraploid *Helianthus annuus* × *Helianthus tuberous* Population That Segregates for Perennial Habit. Genes (Basel). 9(9): 422. doi: 10.3390/genes9090422.

Kao, C.H., Z.B. Zeng, and R.D. Teasdale. 1999. Multiple Interval Mapping for Quantitative Trait Loci. Genetics 152(3): 1203–1216.

Karlsson, P.S., B.M. Svensson, K.O. Carlsson Bengt Åand Nordell, and B.A. Carlsson. 1990. Resource Investment in Reproduction and Its Consequences in Three *Pinguicula* Species. Oikos 59(3): 393.

Kawakami, T., 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. doi:10.1111/j.1365-294X.2011.05105.x

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. doi:10.1111/1755-0998.12227

Kays, S.J., and S.F. Nottingham. 2007. Biology and Chemistry of Jerusalem Artichoke. CRC Press, Boca Raton, FL.

Kearns, C.A., and D.W. Inouye. 1994. Fly Pollination of *Linum lewisii* (*Linaceae*). Am. J. Bot. 81(9): 1091.

Kell, D.B. 2011. Breeding crop plants with deep roots: their role in sustainable carbon, nutrient and water sequestration. Ann. Bot. 108(3): 407–418.

Kharabian-Masouleh, A., D.L.E. Waters, R.F. Reinke, and R.J. Henry. 2011. Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. Plant Biotechnol. J. 9(9): 1074–1085.

Knowles, R.P. 1977. Recurrent Mass Selection for Improved Seed Yields in Intermediate Wheatgrass. Crop Sci. 17(1): 51–54.

Knowles, R.P. 1987. Productivity of Grass Species in the Dark Brown Soil Zone of Saskatchewan. Can. J. Plant Sci. 67(3): 719–725. doi: 10.4141/cjps87-099.

Kois, J. 1985. Effect of Density on Seed Yield and Vegetative Spread in First-Year Maximilian Sunflowers (*Helianthus maximiliani*). L. Rep. Research Suppl.3: 1–3.

Konar, A., O. Choudhury, R. Bullis, L. Fiedler, J.M. Kruser, M.T. Stephens, O. Gailing, S. Schlarbaum, M. V. Coggeshall, M.E. Staton, J.E. Carlson, S. Emrich, and J. Romero-Severson. 2017. High-quality genetic mapping with ddRADseq in the non-model tree *Quercus rubra*. BMC Genomics 18(1): 417. doi: 10.1186/s12864-017-3765-8.

Kort, J., M. Collins, and D. Ditsch. 1998. A review of soil erosion potential associated with biomass crops. Biomass and Bioenergy 14(4): 351–359.

Kosambi, D.D. 1943. The estimation of map distances from recombination values. Ann. Eugen. 12(1): 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x.

Kuhnlein, H.V.H., and N.N.J. Turner. 1991. Traditional Plant Foods of Canadian Indigenous Peoples: Nutrition, Botany, and Use (Vol. 8). Taylor & Francis.

Lal, R. 2003. Global Potential of Soil Carbon Sequestration to Mitigate the Greenhouse Effect. CRC. Crit. Rev. Plant Sci. 22(2): 151–184.

Lambin, E.F., and P. Meyfroidt. 2011. Global land use change, economic globalization, and the looming land scarcity. Proc. Natl. Acad. Sci. U. S. A. 108(9): 3465–3472.

Lande, R. 1982. A Quantitative Genetic Theory of Life History Evolution. Ecology 63(3): 607–615. doi: 10.2307/1936778.

Lande, R., and S.J. Arnold. 1983. The Measurement of Selection on Correlated Characters. Evolution (N. Y). 37(6): 1210. doi: 10.2307/2408842.

Lander, E.S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 136(2): 705.

Larkin, P.J., M.T. Newell, R.C. Hayes, J. Aktar, M.R. Norton, S.J. Moroni, and L.J. Wade. 2014. Progress in developing perennial wheats for grain and grazing. Crop Pasture Sci. 65(11): 1147–1164.

Larsen, J., B.L. Beres, R. Blackshaw, and R.J. Graf. 2018. Extending the Growing Season-Winter Cereals in Western Canada. Can. J. Plant Sci. 98(2): 267-277. doi: 10.1139/CJPS-2017-0278.

Lasisi, A.A., O.O. Akinremi, M. Tenuta, and D. Cattani. 2018. Below-ground plant biomass and nitrogen uptake of perennial forage grasses and annual crops fertilized with pig manures. Agric. Ecosyst. Environ. 268: 1–7. doi: 10.1016/j.agee.2018.08.006.

Latting, J. 1961. The Biology of *Desmanthus Illinoensis*. Ecology 42(3): 487–493.

Leather, G.R. 1983. Sunflowers (*Helianthus annuus*) are allelopathic to weeds. Weed Sci. 31(1): 37–42.

Leggett, J.M. 1985. Interspecific hybrids involving the perennial oat species *Avena macrostachya*. Can. J. Genet. Cytol. 27(1): 29–32.

Leinonen, T., R.S. McCairns, R.B. O'Hara, and J. Merilä. 2013. QST–FST comparisons: Evolutionary and ecological insights from genomic heterogeneity. Nat. Rev. Genet. 14:179. doi:10.1038/nrg3395

Lemercier, E. 1987. Elaboration de la biomasse chez le topinambour et bilan glucidique a different stades du developpement. Thesis. Universite Rennes I, Rennes, France.

Lemus, R., and R. Lai. 2005. Bioenergy Crops and Carbon Sequestration. CRC. Crit. Rev. Plant Sci. 24(1): 1–21. doi: 10.1080/0735268059091039.1.

Leon, A.J., M. Lee, and F.H. Andrade. 2001. Quantitative trait loci for growing degree days to flowering and photoperiod response in sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. 102(4): 497–503. doi: 10.1007/s001220051673.

Lewis, M.E., and K.C. Sink. 1996. RFLP linkage map of asparagus. Genome 39(4): 622–627. doi: 10.1139/g96-079.

Li, J.T., J. Yang, D.C. Chen, X.L. Zhang, and Z.S. Tang. 2007. An optimized mini-preparation method to obtain high-quality genomic DNA from mature leaves of sunflower. Genet. Mol. Res. 6(4): 1064–1071.

Li, H., J.-M. Ribaut, Z. Li, and J. Wang. 2008. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theor. Appl. Genet. 116(2): 243–260.

Li, W.Q., X.Q. Zhang, C. Xia, Y. Deng, and D. Ye. 2010. MALE GAMETOPHYTE DEFECTIVE 1, encoding the FAd subunit of mitochondrial F1F0-ATP synthase, is essential for pollen formation in *Arabidopsis thaliana*. Plant Cell Physiol. 51:923–935. doi:10.1093/pcp/pcq066

Lin, B.B. 2011. Resilience in Agriculture through Crop Diversification: Adaptive Management for Environmental Change. Bioscience 61(3): 183–193.

Lin, C.S., and M.R. Binns. 1988. A method of analyzing cultivar x location x year experiments: a new stability parameter. Theor. Appl. Genet. 76(3). doi: 10.1007/BF00265344.

Lin, Z., X. Li, L.M. Shannon, C.-T. Yeh, M.L. Wang, G. Bai, Z. Peng, J. Li, H.N. Trick, T.E. Clemente, J. Doebley, P.S. Schnable, M.R. Tuinstra, T.T. Tesso, F. White, and J. Yu. 2012. Parallel domestication of the Shattering1 genes in cereals. Nat. Genet. 44(6): 720–724.

Linder, C.R. 2000. Adaptive Evolution of Seed Oils in Plants: Accounting for the Biogeographic Distribution of Saturated and Unsaturated Fatty Acids in Seed Oils. Am. Nat. 156(4): 442–458.

Lischer, H.E., and L. Excoffier. 2011. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299. doi:10.1093/bioinformatics/btr642

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2007. SAS for mixed models. SAS Inst., Cary, NC.

Liu, A., and J.M. Burke. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics 173:321–330. doi:10.1534/genetics.105.051110.

Liu, Z., X. Cai, G.J. Seiler, T.J. Gulya, K.Y. Rashid, and C.C. Jan. 2011. Transferring Sclerotinia stalk rot resistance genes from wild *Helianthus* species into cultivated sunflower. In: Proceedings of the 33rd Sunflower Research Forum. Fargo, ND. 12-13
Jan. 2011. Natl. Sunflower Assoc., Bismark, ND. http://www.sunflowernsa.com/uploads/research/581/liu_sclerotinia_11.pdf (accessed 19 September. 2018).

Liu, W., H.P. Maurer, J.C. Reif, A.E. Melchinger, H.F. Utz, M.R. Tucker, N. Ranc, G. Della Porta, and T. Würschum. 2013. Optimum design of family structure and allocation of resources in association mapping with lines from multiple crosses. Heredity (Edinb). 110(1): 71–79. doi: 10.1038/hdy.2012.63.

Liu, Z., X. Cai, G.J. Seiler, and C.-C. Jan. 2014. Interspecific amphiploid-derived alloplasmic male sterility with defective anthers, narrow disc florets and small ray flowers in sunflower. Plant Breed. 133(6): 742–747.

Lobell, D.B., W. Schlenker, and J. Costa-Roberts. 2011. Climate trends and global crop production since 1980. Science. 333:616-620.

Long, R.W. 1960. Biosystematics of Two Perennial Species of *Helianthus* (*Compositae*). I. Crossing Relationships and Transplant Studies. Am. J. Bot. 47(9): 729–735.

Lorieux, M. 2012. MapDisto: fast and efficient computation of genetic linkage maps. Mol. Breed. 30(2): 1231–1235.

Luu, K., E. Bazin, and M.G. Blum. 2017. pcadapt: An R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. 17:67–77. doi:10.1111/1755-0998.12592

Maas, S.E., A.J. Glenn, M. Tenuta, and B.D. Amiro. 2013. Net CO2 and N2O exchange during perennial forage establishment in an annual crop rotation in the Red River Valley, Manitoba. Can. J. Soil Sci. 93(5): 639–652.

Mamidi, S., R.K. Lee, J.R. Goos, and P.E. McClean. 2014. Genome-Wide Association Studies Identifies Seven Major Regions Responsible for Iron Deficiency Chlorosis in Soybean (*Glycine max*). PLoS One 9(9): e107469.

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. doi:10.1007/s00122-011-1619-3

Mandel, J.R., E.F. Milton, L.A. Donovan, S.J. Knapp, and J.M. Burke. 2013a. Genetic diversity and population structure in the rare Algodones sunflower (*Helianthus niveus* ssp. *tephrodes*). Conserv. Genet. 14(1): 31–40. doi: 10.1007/s10592-012-0421-3.

Mandel, J.R., S. Nambeesan, J.E. Bowers, L.F. Marek, D. Ebert, L.H. Rieseberg, S.J. Knapp, and J.M. Burke. 2013b. Association Mapping and the Genomic Consequences of Selection in Sunflower. PLoS Genet. 9(3): e1003378. doi: 10.1371/journal.pgen.1003378.

Mangan, M.E., C. Sheaffer, D.L. Wyse, N.J. Ehlke, and P.B. Reich. 2011. Native Perennial Grassland Species for Bioenergy: Establishment and Biomass Productivity. Agron. J. 103(2): 509–519.

Margarido, G.R.A., A.P. Souza, and A.A.F. Garcia. 2007. OneMap: software for genetic mapping in outcrossing species. Hereditas 144(3): 78–79.

Marroni, F., S. Pinosio, G. Zaina, F. Fogolari, N. Felice, F. Cattonaro, and M. Morgante. 2011. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. Tree Genet. Genomes 7:1011–1023. doi:10.1007/s11295-011-0391-5

Marshall, C., and D. Ludlam. 1989. The Pattern of Abortion of Developing Seeds in *Lolium perenne* L. Ann. Bot. 63(1): 19–28.

Mather, K. 1955. Polymorphism as an Outcome of Disruptive Selection. Evolution (N. Y). 9(1): 52. doi: 10.2307/2405357.

McAssey, E.V., J. Corbi, and J.M. Burke. 2016. Range-wide phenotypic and genetic differentiation in wild sunflower. BMC Plant Biol. 16:249. doi:10.1186/s12870-016-0937-7

McKinney, G.J., R.K. Waples, C.E. Pascal, L.W. Seeb, and J.E. Seeb. 2018. Resolving allele dosage in duplicated loci using genotyping-by-sequencing data: A path forward for population genetic analysis. Mol. Ecol. Resour. 18(3): 570–579. doi: 10.1111/1755-0998.12763.

MEA. 2005. Millenium Ecosystem Assesment. Ecosystems and Human Well-being: Synthesis. Island Press, Washington, DC.

Meek, B.D., E.R. Rechel, L.M. Carter, W.R. DeTar, and A.L. Urie. 1992. Infiltration Rate of a Sandy Loam Soil: Effects of Traffic, Tillage, and Plant Roots. Soil Sci. Soc. Am. J. 56(3): 908–913.

Meiss, H., S. Mediene, R. Waldhardt, J. Caneill, and N. Munier-Jolain. 2010. Contrasting weed species composition in perennial alfalfas and six annual crops: implications for integrated weed management. Agron. Sustain. Dev. 30(3): 657–666.

Meyer, R.S., A.E. DuVal, and H.R. Jensen. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytol. 196(1): 29–48.

Miller, J., and S. Wolf. 1991. Registration of three cytoplasmic male-sterile and three restorer sunflower germplasm lines. Crop Sci. 31(2): 500–500.

Miller, M.A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees In: Proceedings of the Gateway Computing Environments Workshop. 14 Nov. 2010. Inst., Elect. Electron. Eng., Piscataway, NJ. p. 1–8, doi:10.1109/GCE.2010.5676129

Miller, A.J., and B.L. Gross. 2011. From forest to field: Perennial fruit crop domestication. Am. J. Bot. 98(9): 1389–1414.

Miyakawa, T., K.I. Hatano, Y. Miyauchi, Y.I. Suwa, Y. Sawano, and M. Tanokura. 2014. A secreted protein with plant-specific cysteine-rich motif functions as a mannose-binding lectin that exhibits antifungal activity. Plant Physiol. 166:766–778. doi:10.1104/pp.114.242636

Moerman, D.E. 2010. Native American Food Plants- an ethnobotanical dictionary. Timber Press.

Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M. Yano, C.R. Bhatia, and T. Sasaki. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol. Breed. 3(2): 87–103.

Montagnini, F., and P.K.R. Nair. 2004. Carbon sequestration: An underexploited environmental benefit of agroforestry systems. Agrofor. Syst. 61–62(1–3): 281–295.

Morgan, M.T., D.J. Schoen, and T.M. Bataillon. 1997. The Evolution of Self-Fertilization in Perennials. Am. Nat. 150(5): 618–638.

Morgan, M.T. 2001. Consequences of life history for inbreeding depression and mating system evolution in plants. Proc. R. Soc. London B Biol. Sci. 268(1478): 1817–1824.

Moyer, J.R., R.E. Blackshaw, E.G. Smith, and S.M. McGinn. 2000. Cereal cover crops for weed suppression in a summer fallow-wheat cropping sequence. Can. J. plant Sci. 80(2): 441–449.

Moyers, B.T., and L.H. Rieseberg. 2013. Divergence in Gene Expression Is Uncoupled from Divergence in Coding Sequence in a Secondarily Woody Sunflower. Int. J. Plant Sci. 174(7): 1079–1089.

Murphy, K.M., L.A. Hoagland, P.G. Reeves, B.-K. Baik, and S.S. Jones. 2009. Nutritional and quality characteristics expressed in 31 perennial wheat breeding lines. Renew. Agric. Food Syst. 24(04): 285.

Murray, S.C., and R.W. Jessup. 2014. Breeding and genetics of perennial maize: progress, oppertunities and challenges. p. 20–30. *In* Batello, C., Wade, L., Cox, S., Pogna, N., Bozzini, A., Choptiany, J. (eds.), Perennial Crops for Food Security; Proceedings of the FAO Expert Workshop. FAO, 28-30 Aug. 2013. FAO, Rome, Italy.

N'Diaye, A., J.K. Haile, D.B. Fowler, K. Ammar, and C.J. Pozniak. 2017. Effect of Co-segregating Markers on High-Density Genetic Maps and Prediction of Map Expansion Using Machine Learning Algorithms. Front. Plant Sci. 8:1434. doi: 10.3389/fpls.2017.01434.

Nabukalu, P., and T.S. Cox. 2016. Response to selection in the initial stages of a perennial sorghum breeding program. Euphytica 209(1): 103–111.

Nadler, A. 2007. An agroclimatic risk assessment of crop production on the Canadian prairies. M.Sc. Thesis, Uiversity of Manitoba. http://mspace.lib.umanitoba.ca/handle/1993/2829 (accessed 26 September 2018).

Nadler, A.J., and P.R. Bullock. 2011. Long-term changes in heat and moisture related to corn production on the Canadian Prairies. Clim. Change 104(2): 339–352. doi: 10.1007/s10584-010-9881-y.

Nambeesan, S.U., J.R. Mandel, J.E. Bowers, L.F. Marek, D. Ebert, J. Corbi, L.H. Rieseberg, S.J. Knapp, and J.M. Burke. 2015. Association mapping in sunflower (*Helianthus annuus* L.) reveals independent control of apical vs. basal branching. BMC Plant Biol. 15(1): 1.

Narum, S.R., and J.E. Hess. 2011. Comparison of *FST* outlier tests for SNP loci under selection. Mol. Ecol. Resour. 11(1): 184–194. doi: 10.1111/j.1755-0998.2011.02987.x.

Narum, S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, and P.A. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Mol. Ecol. 22(11): 2841–2847. doi: 10.1111/mec.12350.

Nishio, S., N. Takada, T. Yamamoto, S. Terakami, T. Hayashi, Y. Sawamura, and T. Saito. 2013. Mapping and pedigree analysis of the gene that controls the easy peel pellicle trait in Japanese chestnut (*Castanea crenata* Sieb. et Zucc.). Tree Genet. Genomes 9(3): 723–730. doi: 10.1007/s11295-012-0587-3.

Norland, J., S. Fasching, C. Dixon, K. Askerooth, K. Kelsey, and G. Wang. 2013. Reduced Establishment of Canada Thistle (*Cirsium arvense*) Using Functionally Similar Native Forbs. Ecol. Restor. 31(2): 143–146.

Ockendon, D.J. 1968. Biosystematic studies in the *Linum perenne* group. New Phytol. 67(4): 787-813.

Ominski, P.D., M.H.H. Entz, and N. Kenkel. 1999. Weed suppression by *Medicago sativa* in subsequent cereal crops: a comparative survey. Weed Sci. 47(3): 282–290.

Otfinowski, R., N.C. Kenkel, and P.M. Catling. 2007. The biology of Canadian weeds. 134. *Bromus inermis* Leyss. Can. J. Plant Sci. 87(1982): 183–198. doi: doi:10.4141/P06-071.

Otfinowski, R., and N.C. Kenkel. 2008. Clonal integration facilitates the proliferation of smooth brome clones invading northern fescue prairies. Plant Ecol. 199(2): 235–242. doi: 10.1007/s11258-008-9428-8.

Ott, A., S. Liu, J.C. Schnable, C.-T. Yeh, C. Wang, and P.S. Schnable. 2017. Tunable Genotyping-By-Sequencing (tGBS®) Enables Reliable Genotyping of Heterozygous Loci. bioRxiv: 100461.

Owens, G.L., G.J. Baute, and L.H. Rieseberg. 2016. Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. Mol. Ecol. 25(11): 2630–2643. doi: 10.1111/mec.13569.

Paris, J.R., J.R. Stevens, and J.M. Catchen. 2017. Lost in parameter space: a road map for stacks. Methods Ecol. Evol. 8(10): 1360–1373. doi: 10.1111/2041-210X.12775.

Paterson, A.H., K.F. Schertz, Y.R. Lin, S.C. Liu, and Y.L. Chang. 1995. The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. Proc. Natl. Acad. Sci. 92(13): 6127–6131.

Paterson, A.H., T.S. Cox, and W. Kong. 2014. Multiple-Harvest sorghum towards improved food security. p. 90–102. *In* Batello, C., Wade, L., Cox, S., Pagna, N., Bozzini, A., J., C. (eds.), Perennial Crops for Food Security; Proceedings of the FAO Expert Workshop. 28-30 Aug. 2013. FAO, Rome, Italy.

Paustian, K., O. Andren, M. Clarholm, A.C. Hansson, G. Johansson, J. Lagerlof, T. Lindberg, R. Pettersson, and B. Sohlenius. 1990. Carbon and Nitrogen Budgets of Four Agro-Ecosystems With Annual and Perennial Crops, With and Without N Fertilization. J. Appl. Ecol. 27(1): 60.

Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. doi:10.1093/bioinformatics/bts460

Pearl, H.M., C. Nagai, P.H. Moore, D.L. Steiger, R. V Osgood, and R. Ming. 2004. Construction of a genetic map for arabica coffee. Theor. Appl. Genet. 108(5): 829–835.

Pearl, S.A., J.E. Bowers, S. Reyes-Chin-Wo, R.W. Michelmore, and J.M. Burke. 2014. Genetic analysis of safflower domestication. BMC Plant Biol. 14(1):43. doi: 10.1186/1471-2229-14-43.

Peng, Q., Y. Hu, R. Wei, Y. Zhang, C. Guan, Y. Ruan, and C. Liu. 2010. Simultaneous silencing of FAD2 and FAE1 genes affects both oleic acid and erucic acid contents in *Brassica napus* seeds. Plant Cell Rep. 29(4): 317–325.

Picasso, V.D., E.C. Brummer, M. Liebman, P.M. Dixon, and B.J. Wilsey. 2008. Crop Species Diversity Affects Productivity and Weed Suppression in Perennial Polycultures under Two Management Strategies. Crop Sci. 48(1): 331.

Pimentel, D., D. Cerasale, R.C. Stanley, R. Perlman, E.M. Newman, L.C. Brent, A. Mullan, and D.T.I. Chang. 2012. Annual vs. perennial grain production. Agric. Ecosyst. Environ. 161: 1–9. doi: 10.1016/j.agee.2012.05.025.

Piper, J.K., and P.A. Kulakow. 1994. Seed yield and biomass allocation in Sorghum bicolor and F 1 and backcross generations of *S . bicolor* × *S . halepense* hybrids. Can. J. Bot. 72(4): 468–474. doi: 10.1139/b94-062.

Piper, J.K. 1998. Growth and seed yield of three perennial grains within monocultures and mixed stands. Agric. Ecosyst. Environ. 68(1–2): 1–11.

Pitelka, L.F., S.B. Hansen, and J.W. Ashmun. 1985. Population Biology of *Clintonia Borealis*: I. Ramet and Patch Dynamics. J. Ecol. 73(1): 169.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome J. 5(3): 103.

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome J. 5(3): 92–102.

Polgár, Z., and S. Krasnyanski. 1992. Plant regeneration from cell suspension and mesophyll protoplasts of *Helianthus maximiliani* (Schrad.). Plant Sci. 87(2): 191–197. doi: 10.1016/0168-9452(92)90150-K.

Power, A.G. 2010. Ecosystem services and agriculture: tradeoffs and synergies. Philos. Trans. R. Soc. B Biol. Sci. 365(1554): 2959–2971.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904. doi:10.1038/ng1847

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000a. Inference of Population Structure Using Multilocus Genotype Data. Genetics 155(2): 945–959.

Pritchard, J.K., M. Stephens, N.A. Rosenberg, and P. Donnelly. 2000b. Association Mapping in Structured Populations. Am. J. Hum. Genet. 67(1): 170–181.

Prunier, J., B. Pelgas, F. Gagnon, M. Desponts, N. Isabel, J. Beaulieu, and J. Bousquet. 2013. The genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. BMC Genomics 14:368. doi:10.1186/1471-2164-14-368

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. De Baker, M.J. Daly and P.C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. doi:10.1086/519795

Purugganan, M.D., and D.Q. Fuller. 2009. The nature of selection during plant domestication. Nature 457(7231): 843–848.

Pyhäjärvi, T., M.B. Hufford, S. Mezmouk, and J. Ross-Ibarra. 2013. Complex patterns of local adaptation in teosinte. Genome Biol. Evol. 5:1594–1609. doi:10.1093/gbe/evt109

Ramsay, L., J. Comadran, A. Druka, D.F. Marshall, W.T.B. Thomas, M. MacAulay, K. MacKenzie, C. Simpson, J. Fuller, N. Bonar, P.M. Hayes, U. Lundqvist, J.D. Franckowiak, T.J. Close, G.J. Muehlbauer, and R. Waugh. 2011. INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. Nat. Genet. 43(2): 169–172. doi: 10.1038/ng.745.

Randall, G.W., and D.J. Mulla. 2001. Nitrate nitrogen in surface waters as influenced by climatic conditions and agricultural practices. J. Environ. Qual. 30:337–344. doi:10.2134/jeq2001.302337x

Rastas, P. 2017. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 33(23): 3726–3732. doi: 10.1093/bioinformatics/btx494.

Reimann‐Philipp, R. 1986. Perennial Spring Rye as a Crop Alternative. J. Agron. Crop Sci. 157(4): 281–285. doi: 10.1111/j.1439-037X.1986.tb00077.x.

Reimann-Philipp, R. 1995. Breeding Perennial Rye. Plant Breed. Rev. 13: 265–292. doi: 10.1002/9780470650059.ch8.

Rellstab, C., F. Gugerli, A.J. Eckert, A.M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. Mol. Ecol. 24(17): 4348–4370. doi: 10.1111/mec.13322.

Remigereau, M.-S., G. Lakis, S. Rekima, M. Leveugle, M.C. Fontaine, T. Langin, A. Sarr, and T. Robert. 2011. Cereal Domestication and Evolution of Branching: Evidence for Soft Selection in the Tb1 Orthologue of Pearl Millet (*Pennisetum glaucum* [L.] R. Br.) (PK Ingvarsson, Ed.). PLoS One 6(7): e22404. doi: 10.1371/journal.pone.0022404.

Resende, R.M.S., M.D. Casler, and M.D.V. de Resende. 2013. Selection methods in forage breeding: A quantitative appraisal. Crop Sci. 53(5): 1925–1936. doi: 10.2135/cropsci2013.03.0143.

Riley, K. 1984. *Helianthus maximiliani* density trial. L. Rep. Reseach Suppl.: 6–8.

Roff, D.A., S. Mostowy, and D.J. Fairbairn. 2002. The evolution of trade-offs: testing predicitons on reponse to selection and envrionmental variation. J. Evol. Biol. 56(1): 84–95.

Roff, D.A., and D.J. Fairbairn. 2007. The evolution of trade-offs: Where are we? J. Evol. Biol. 20(2): 433–447. doi: 10.1111/j.1420-9101.2006.01255.x.

Rogers, C., T.E. Thompson, and G.J. Seiler. 1982. Sunflower species of the United States.

Ronicke, S., V. Hahn, R. Horn, I. Grone, L. Brahm, H. Schnabl, and W. Friedt. 2004. Interspecific hybrids of sunflower as a source of Sclerotinia resistance. Plant Breed. 123(2): 152–157.

Ronin, Y., D. Mester, D. Minkov, and A. Korol. 2010. Building reliable genetic maps: different mapping strategies may result in different maps. Nat. Sci. 02(06): 576–589. doi: 10.4236/ns.2010.26073.

Runck, B.C., M.B. Kantar, N.R. Jordan, J.A. Anderson, D.L. Wyse, J.O. Eckberg, R.J. Barnes, C.L. Lehman, L.R. DeHaan, R.M. Stupar, C.C. Sheaffer, and P.M. Porter. 2014. The Reflective Plant Breeding Paradigm: A Robust System of Germplasm Development to Support Strategic Diversification of Agroecosystems. Crop Sci. 54(5): 1939–1948.

Ryan, M., T. Crews, S. Culman, and L. DeHaan. 2018. Managing for Multifunctionality in Perennial Grain Crops. Bioscience 68(4): 294–304.

Sadras, V.O. 2007. Evolutionary aspects of the trade-off between seed size and number in crops. F. Crop. Res. 100: 125–138.

Sadras, V.O., D.J. Connor, and D.M. Whitfield. 1993. Yield, yield components and source-sink relationships in water-stressed sunflower. F. Crop. Res. 31(1–2): 27–39. doi: 10.1016/0378-4290(93)90048-R.

Saftic-Pankovic, D., J. Atlagić, T. Miljanovic G., and N. Radovanovic. 2005. Morphological and molecular variability of *Helianthus giganteus* L. and *Helianthus maximiliani* Sch. species. Genetika 37(2): 121–130.

Sakiroglu, M., and E.C. Brummer. 2017. Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. Theor. Appl. Genet. 130(2): 261–268. doi: 10.1007/s00122-016-2782-3.

Samson, F., and F. Knopf. 1994. Prairie Conservation in North America. Bioscience 44(6): 418–421.

Sanders, T.B. and J.L.H. 1980. Variation in the breeding system of *Elymus canadensis*. Evolution (N. Y). 34: 117–123.

SAS Institute. 2011. SAS version 9.3. SAS Inst., Cary, NC.

Saunders, M.E., G.W. Luck, and M.M. Mayfield. 2013. Almond orchards with living ground cover host more wild insect pollinators. J. Insect Conserv. 17(5): 1011–1025. doi: 10.1007/s10841-013-9584-6.

Sax, K. 1923. The Association of Size Differences with Seed-Coat Pattern and Pigmentation in *Phaseolus vulgaris*. Genetics 8(6): 552.

Scascitelli, M., K.D. Whitney, R.A. Randell, M. King, C.A. Buerkle, and L.H. Rieseberg. 2010. Genome scan of hybridizing sunflowers from Texas *(Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. Mol. Ecol. 19:521–541. doi:10.1111/j.1365-294X.2009.04504.x

Scheinost, P.L., D.L. Lammer, X. Cai, T.D. Murray, and S.S. Jones. 2001. Perennial wheat: The development of a sustainable cropping system for the U.S. Pacific Northwest. Am. J. Altern. Agric. 16(04): 147–151.

Schellenberg, M.P., B. Biligetu, G.J. McLeod, and Z. Wang. 2012. Phenotypic variation of side-oats grama grass [*Bouteloua curtipendula* (Michx.) Torr.] collections from the Canadian prairie. Can. J. Plant Sci. 92(6): 1043–1048. doi: 10.4141/cjps2011-142.

Scoggan, H.J. 1957. Flora of Manitoba. Vol. 47. Minister of Northern Affairs and National Resources.

Sedbrook, J.C., W.B. Phippen, and M.D. Marks. 2014. New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: Example pennycress (*Thlaspi arvense* L.). Plant Sci. 227: 122–132.

Seiler, G.J. 1983a. Effect of Genotype, Flowering Date, and Environment on Oil Content and Oil Quality of Wild Sunflower Seed 1. Crop Sci. 23(6): 1063. doi: 10.2135/cropsci1983.0011183X002300060010x.

Seller, G.J. 1983b. Protein and Mineral Concentrations of Selected Wild Sunflower Species 1. Agron. J. 76(2): 289–294.

Seiler, G.J. 1985. Evaluation of Seeds of Sunflower Species for Several Chemical and Morphological Characteristics. Crop Sci. 25(1): 183–187.

Seiler, G.J. 1994. Oil concentration and fatty acid composition of achenes of North American *Helianthus* (*Asteraceae*) species. Econ. Bot. 48:271–279. doi:10.1007/BF02862328

Seiler, G.J. 1998. Seed maturity, storage time and temperature, and media treatment effects on germination of two wild sunflowers. Agron. J. 90(2): 221–226.

Seiler, G.J. 2010. Utilization of wild *Helianthus* species in breeding for disease resistance. p. 36–50. *In* International Symposium “Sunflower Breeding on Resistance to Diseases".

Seiler, G.J., and M.E. Brothers. 1999. Oil concentration and fatty acid composition of achenes of Helianthus species (*Asteraceae*) from Canada. Econ. Bot. 53:273–280. doi:10.1007/BF02866637

Seiler, G.J., L.L. Qi, and L.F. Marek. 2017. Utilization of sunflower crop wild relatives for cultivated sunflower improvement. Crop Sci. 57(3):1083–1101. doi:10.2135/cropsci2016.10.0856

Shapiro, S.S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). Biometrika 52:591–611. doi:10.1093/biomet/52.3-4.591

Shapter, F.M., M. Cross, G. Ablett, S. Malory, I.H. Chivers, G.J. King, and R.J. Henry. 2013. High-Throughput Sequencing and Mutagenesis to Accelerate the Domestication of *Microlaena stipoides* as a New Food Crop. PLoS One 8(12): e82641.

Shaver, D.L. 1964. Perennialism in *Zea*. Genetics 50: 393–406.

Shirasawa, K., M. Tanaka, Y. Takahata, D. Ma, Q. Cao, Q. Liu, H. Zhai, S.S. Kwak, J. Cheol Jeong, U.H. Yoon, H.U. Lee, H. Hirakawa, and S. Isobe. 2017. A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). Sci. Rep. 7:44207 doi: 10.1038/srep44207.

Simons, K.J. 2005. Molecular Characterization of the Major Wheat Domestication Gene Q. Genetics 172(1): 547–555.

Skinner, M.E., A.V. Uzilov, L.D. Stein, C.J. Mungall, and I.H. Holmes. 2009. JBrowse: A next-generation genome browser. Genome Res. 19:1630–1638. doi:10.1101/gr.094607.109

Škorić, D. 1992. Achievements and future directions of sunflower breeding. F. Crop. Res. 30(3–4): 231–270.

Sleper, D.A., and J.M. Poehlman. 2006. Breeding Field Crops. 5th ed. Blackwell publishing.

Smaje, C. 2015. The Strong Perennial Vision: A Critical Review The Strong Perennial Vision: A Critical Review. Agroecol. Sustain. Food Syst. 3565(October). doi: 10.1080/21683565.2015.1007200.

Somers, D.J., T. Banks, R. DePauw, S. Fox, J. Clarke, C. Pozniak, and C. McCartney. 2007. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50(6): 557–567. doi: 10.1139/G07-031.

Song, J.Z., M. Soller, and A. Genizi. 1999. The full-sib intercross line (FSIL): a QTL mapping design for outcrossing species. Genet. Res. 73(01): 61–73.

Sork, V.L., S.N. Aitken, R.J. Dyer, A.J. Eckert, P. Legendre, and D.B. Neale. 2013. Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. Tree Genet. Genomes 9(4): 901–911. doi: 10.1007/s11295-013-0596-x.

Soto-Cerda, B.J., and S. Cloutier. 2012. Association mapping in plant genomes. *In* Genetic diversity in plants. Ed. M. Caliskan, 29-54. InTech.

Soto-Cerda, B.J., A. Diederichsen, S. Duguid, H. Booker, G. Rowland, and S. Cloutier. 2014. The potential of pale flax as a source of useful genetic variation for cultivated flax revealed through molecular diversity and association analyses. Mol. Breed. 34(4): 2091–2107.

Sposaro, M.M., P.M. Berry, M. Sterling, A.J. Hall, and C.A. Chimenti. 2010. Modelling root and stem lodging in sunflower. F. Crop. Res. 119(1): 125–134. doi: 10.1016/j.fcr.2010.06.021.

St. Luce, M., C.A. Grant, B.J. Zebarth, N. Ziadi, J.T. O'Donovan, R.E. Blackshaw, K.N. Harker, E.N. Johnson, Y. Gan, G.P. Lafond, W.E. May, M. Khakbazan, and E.G. Smith. 2015. Legumes can reduce economic optimum nitrogen rates and increase yields in a wheat-canola cropping sequence in western canada. F. Crop. Res. 179: 12–25. doi: 10.1016/j.fcr.2015.04.003.

Stanton-Geddes, J., T. Paape, B. Epstein, R. Briskine, J. Yoder, J. Mudge, A.K. Bharti, A.D. Farmer, P. Zhou, R. Denny, G.D. May, S. Erlandson, M. Yakub, M. Sugawara, M.J. Sadowsky, N.D. Young, and P. Tiffin. 2013. Candidate Genes and Genetic Architecture of Symbiotic and Agronomic Traits Revealed by Whole-Genome, Sequence-Based Association Genetics in *Medicago truncatula*. PLoS One 8(5): e65688.

Stebbins, G. 1958. The inviability, weakness, and sterility of interspecific hybrids. p. 147–215. *In* Advances in genetics. Vol. 9. Academic Press.

Stebbins, J. 2013. *Helianthus winteri* (Asteraceae), a New Perennial Species From the Southern Sierra Nevada Foothills, California. Aliso 31(1): 19–23.

Steiner, J.J., P.R. Beuselinck, R.N. Peaden, W.P. Kojis, and E.T. Bingham. 1992. Pollinator effects on crossing and genetic shift in a three-flower-color alfalfa population. Crop Sci. 32: 73–77. doi: 10.2135/cropsci1992.0011183X003200010016x.

Stephan, W. 2016. Signatures of positive selection: From selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Mol. Ecol. 25:79–88. doi:10.1111/mec.13288

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. Am. J. Bot. 102(6): 910–920. doi: 10.3732/ajb.1500031.

Storey, J.D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100(16): 9440–9445. doi: 10.1073/pnas.1530509100.

Stuessy, T.F., D.M. Spooner, and K.A. Evans. 1986. Adaptive significance of ray corollas in *Helianthus grosseserratus* (*Compositae*). Am. Midl. Nat. 115:191–197. doi:10.2307/2425849

Sujatha, M., and A.J. Prabakaran. 1997. Reaction of wild sunflowers and certain interspecies hybrids to *Alternaria helianthi*. Helia (20): 107–114.

Sujatha, M. 2006. Wild *Helianthus* species used for broadening the genetic base of cultivated sunflower in India. Helia 29(44): 77–86.

Suneson, C.A., E. Sharkawy, and W.E. Hall. 1963. Progress in 25 years of perennial wheat development. Crop Sci. 3: 437–439.

Swanton, C.J., D.R. Clements, M.J. Moore, and P.B. Cavers. 1992. The biology of Canadian weeds. 101. *Helianthus tuberosus* L. Can. J. Plant Sci. 72:1367–1382. doi:10.4141/cjps92-169

Talukder, Z.I., L. Gong, B.S. Hulke, V. Pegadaraju, Q. Song, Q. Schultz, and L. Qi. 2014. A High-Density SNP Map of Sunflower Derived from RAD-Sequencing Facilitating Fine-Mapping of the Rust Resistance Gene R 12. PLoS One 9(7): e98628.

Tang, S.X., A. Leon, W.C. Bridges, and S.J. Knapp. 2006. Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. Crop Sci. 46(2): 721–734.

Tanksley, S.D. 1983. Molecular markers in plant breeding. Plant Mol. Biol. Report. 1(1): 3–8.

Taski-Ajdukovic, K., D. Vasic, and N. Nagl. 2006. Regeneration of interspecific somatic hybrids between *Helianthus annuus* L. and *Helianthus maximiliani* (Schrader) via protoplast electrofusion. Plant Cell Rep. 25:698–704. doi:10.1007/s00299-006-0134-5

Taylor, A.M., B.D. Amiro, and T.J. Fraser. 2013. Net CO2 exchange and carbon budgets of a three-year crop rotation following conversion of perennial lands to annual cropping in Manitoba, Canada. Agric. For. Meteorol. 182–183: 67–75. doi: 10.1016/j.agrformet.2013.07.008.

Taylor, J., and D. Butler. 2017. R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis. J. Stat. Softw. 79(6): 1–29. doi: 10.18637/jss.v079.i06.

Tesio, F., L.A. Weston, F. Vidotto, and A. Ferrero. 2010. Potential Allelopathic Effects of Jerusalem Artichoke (*Helianthus tuberosus*) Leaf Tissues. Weed Technol. 24(3): 378–385.

Tesio, F., L.A. Weston, and A. Ferrero. 2011. Allelochemicals identified from Jerusalem artichoke (*Helianthus tuberosus* L.) residues and their potential inhibitory activity in the field and laboratory. Sci. Hortic. (Amsterdam). 129(3): 361–368.

Tetreault, H.M., T. Kawakami, M.C. Ungerer, and C. Levy. 2016. Low Temperature Tolerance in the Perennial Sunflower *Helianthus maximiliani*. Am. Midl. Nat. 175(1): 91–102. doi: 10.1674/amid-175-01-91-102.1.

Texas Agricultural Experiment Station. 1979. Aztec Maximilian sunflower. Leaflet 1775. 1-4.

Tiessen, K.H.D., J.A. Elliott, J. Yarotski, D.A. Lobb, D.N. Flaten, and N.E. Glozier. 2010. Conventional and Conservation Tillage: Influence on Seasonal Runoff, Sediment, and Nutrient Losses in the Canadian Prairies. J. Environ. Qual. 39(3): 964–980.

Tiffin, P., and B.S. Gaut. 2001. Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: Insights from four nuclear loci. Genetics 158(1): 401–412.

Tikhomirov, V.T., and P. V Chiryaev. 2005. Sources of resistance to diseases in original material of sunflower. Helia 28(42): 101–106.

Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, and E. Siemann. 1997. The influence of functional diversity and composition on ecosystem processes. Science 277:1300–1302. doi:10.1126/science.277.5330.1300

Tilman, D., J. Hill, and C. Lehman. 2006. Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass. Science. 314(5805): 1598–1600.

Tilman, D. 2007. Resource competition and plant traits: a response to Craine et al. 2005. J. Ecol. 95(2): 231–234.

Tilman, D., R. Socolow, J.A. Foley, J. Hill, E. Larson, L. Lynd, S. Pacala, J. Reilly, T. Searchinger, C. Somerville, and R. Williams. 2009. Beneficial biofuels: the food, energy, and environment trilemma. Science 325(5938): 270–1. doi: 10.1126/science.1177970.

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. Am. J. Bot. 94:1837–1852. doi:10.3732/ajb.94.11.1837.

Toker, C., H. Canci, and T. Yildirim. 2007. Evaluation of perennial wild *Cicer* species for drought resistance. Genet. Resour. Crop Evol. 54(8): 1781–1786.

Tsitsin, N. V, and V.F. Lubimova. 1959. New Species and Forms of Cereals Derived from Hybridization between Wheat and Couch Grass. Am. Nat. 93(870): 181–191.

Tsitsin, N. V. 1965. Remote hybridization as a method of creating new species and varieties of plants. Euphytica 14(3): 326–330.

Turkington, R.A., P.B. Cavers, and E. Rempel. 1978. The biology of Canadian weeds.: 29. *Melilotus alba* Desr. and *M .officinalis* (L.) Lam. Can. J. Plant Sci. 58(2): 523–537. doi: 10.4141/cjps78-078.

Umehara, M., A. Hanada, S. Yoshida, K. Akiyama, T. Arite, N. Takeda-Kamiya, H. Magome, Y. Kamiya, K. Shirasu, K. Yoneyama, J. Kyozuka, and S. Yamaguchi. 2008. Inhibition of shoot branching by new terpenoid plant hormones. Nature 455(7210): 195–200. doi: 10.1038/nature07272.

Ungerer, M.C., S.J. Baird, J. Pan, and L.H. Rieseberg. 1998. Rapid hybrid speciation in wild sunflowers. Proc. Natl. Acad. Sci. 95(20): 11757–11762.

UniProt Consortium. 2017. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 45:D158–D169. doi:10.1093/nar/gkw1099

Upadhyaya, M.K., R. Turkington, and D. McIlvride. 1986. The biology of Canadian weeds. 75. *Bromus tectorum* L. Can. J. Plant Sci. 66(3): 689–709. doi: 10.4141/cjps86-091.

Uppalapati, S.R., D.D. Serba, Y. Ishiga, L.J. Szabo, S. Mittal, H.S. Bhandari, J.H. Bouton, K.S. Mysore, and M.C. Saha. 2013. Characterization of the Rust Fungus, *Puccinia emaculata*, and Evaluation of Genetic Variability for Rust Resistance in Switchgrass Populations. BioEnergy Res. 6(2): 458–468. doi: 10.1007/s12155-012-9263-6.

USDA-NRCS. 2017. The PLANTS database. Natl. Plant Data Team, Greensboro, NC. http://plants.usda.gov (accessed 13 Oct. 2017).

van der Voet, H. 1994. Comparing the predictive accuracy of models using a simple randomization test. Chemom. Intell. Lab. Syst. 25(2): 313-322. doi: 10.1016/0169-7439(94)85050-X.

van Heerden, C.J., P. Burger, A. Vermeulen, and R. Prins. 2014. Detection of downy and powdery mildew resistance QTL in a 'Regent' × 'RedGlobe' population. Euphytica 200(2): 281–295. doi: 10.1007/s10681-014-1167-4.

van Noordwijk, A.J., and G. de Jong. 1986. Acquisition and Allocation of Resources: Their Influence on Variation in Life History Tactics. Am. Nat. 128(1): 137–142. doi: 10.1086/284547.

Van Tassel, D.L., L.R. DeHaan, and T.S. Cox. 2010. Missing domesticated plant forms: can artificial selection fill the gap? Evol. Appl. 3(5–6): 434–452.

Van Tassel, D.L., S.R. Asselin, S.A. Cox, G. Sideli, and D.J. Cattani. 2014. Evaluating perennial candidates for domestication: lessons from wild sunflower relatives. p. 112–140. *In* Batello, C., Wade, L., Cox, S., Pagna, N., Bozzini, A., J., C. (eds.), Perennial Crops for Food Security; Proceedings of the FAO Expert Workshop. 28-30 Aug. 2013. FAO, Rome, Italy.

Varshney, R.K., S.N. Nayak, G.D. May, and S.A. Jackson. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol. 27(9): 522–530.

Vasic, D., G. Alibert, and D. Škorić. 2001. Protocols for efficient repetitive and secondary somatic embryogenesis in *Helianthus maximiliani* (Schrader). Plant Cell Rep. 20(2): 121–125.

Vear, F. 2010. Wild Crop Relatives: Genomic and Breeding Resources (C Kole, Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg.

Vear, F. 2016. Changes in sunflower breeding over the last fifty years. Oilseeds Fats Crops Lipids 23:D202. doi:10.1051/ocl/2016006

Velasco, L., B. Pérez-Vich, and J.M. Fernández-Martínez. 2004. Evaluation of wild sunflower species for tocopherol content and composition. Helia 27(40): 107–112.

Vico, G., S. Manzoni, L. Nkurunziza, K. Murphy, and M. Weih. 2016. Trade-offs between seed output and life span - a quantitative comparison of traits between annual and perennial congeneric species. New Phytol. 209(1): 104–114.

Voytas, D.F., and C. Gao. 2014. Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biol. 12(6): e1001877.

Wagoner, P., and J.R. Schaeffer. 1990. Perennial grain development: Past efforts and potential for the future. CRC. Crit. Rev. Plant Sci. 9(5): 381–408. doi: 10.1080/07352689009382298.

Wakar, B.A. 1937. Cytologische Untersuchung der selbstfertilen ersten Generation der Weizen-Queckengras Bastarde. Cytologia (Tokyo). 8(1): 67–90. doi: 10.1508/cytologia.8.67.

Wang, J., R.C. Baillie, N.O. Cogan, N.M. McFarlane, M.P. Dupal, K.F. Smith, and J.W. Forster. 2011. Molecular genetic marker-based analysis of species-differentiated phenotypic characters in an interspecific ryegrass mapping population. Crop Pasture Sci. 62(10): 892–902.

Wang, H., J.A. Walla, V.A. Magnusson, S. Zhong, and W. Dai. 2014. Construction of genetic linkage maps and QTL mapping for X-disease resistance in tetraploid chokecherry (*Prunus virginiana* L.) using SSR and AFLP markers. Mol. Breed. 34(1): 143–157.

Ward, P.R., F.X. Dunin, and S.F. Micin. 2002. Water use and root growth by annual and perennial pastures and subsequent crops in a phase rotation. Agric. Water Manag. 53(1–3): 83–97. doi: 10.1016/S0378-3774(01)00157-3.

Ward, J.A., J. Bhangoo, F. Fernández-Fernández, P. Moore, J.D. Swanson, R. Viola, R. Velasco, N. Bassil, C.A. Weber, and D.J. Sargent. 2013. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics 14(1): 2.

Warnke, S.E., R.E. Barker, G. Jung, S.-C. Sim, M.A. Rouf Mian, M.C. Saha, L.A. Brilman, M.P. Dupal, and J.W. Forster. 2004. Genetic linkage mapping of an annual x perennial ryegrass population. Theor. Appl. Genet. 109(2): 294–304. doi: 10.1007/s00122-004-1647-3.

Wayne Polley, H., B.J. Wilsey, and J.D. Derner. 2007. Dominant species constrain effects of species diversity on temporal variability in biomass production of tallgrass prairie. Oikos 116(12): 2044–2052.

Westerbergh, A., and J. Doebley. 2004. Quantitative trait loci controlling phenotypes related to the perennial versus annual habit in wild relatives of maize. Theor. Appl. Genet. 109(7): 1544–1553.

Whelan, E.D.P. 1978. Hybridization between annual and perennial diploid species of *Helianthus*. Can. J. Genet. Cytol. 20(4): 523–530.

Whelan, E.D. 1980. A new source of cytoplasmic male sterility in sunflower. Euphytica 29:33–46. doi:10.1007/BF00037247

Whelan, E.D.P., and D.G. Dorrell. 1980. Interspecific hybrids between *Helianthus maximiliani* Schrad. and *H. annuus* L.: effect of backcrossing on meiosis, anther morphology, and seed characteristics. Crop Sci. 20(1): 29–34.

Wilkins, P.W., and M.O. Humphreys. 2003. Progress in breeding perennial forage grasses for temperate agriculture. J. Agric. Sci. 140(02): 129–150.

Willms, W.D., B.H. Ellert, H.H. Janzen, and H. Douwes. 2005. Evaluation of native and introduced grasses for reclamation and production. Rangel. Ecol. Manag. 58(2): 177-183. doi: 10.2111/1551-5028(2005)58<177:EONAIG>2.0.CO;2.

Wills, D.M., and J.M. Burke. 2007. Quantitative trait locus analysis of the early domestication of sunflower. Genetics 176(4): 2589–2599.

Wills, D.M., H. Abdel-Haleem, S.J. Knapp, and J.M. Burke. 2010. Genetic Architecture of Novel Traits in the Hopi Sunflower. J. Hered. 101(6): 727–736.

Wilsey, B.J., and H.W. Polley. 2002. Reductions in grassland species evenness increase dicot seedling invasion and spittle bug infestation. Ecol. Lett. 5(5): 676–684.

Worley, A.C. and S.C.H. Barrett. 2000. Evolution of floral display in *Eichhornia paniculata* (*Pontederiaceae)*: direct and correlated responses to selection on flower size and number. Evolution. 54(5):1533-1545.

Worley, A.C., D. Houle, and S.C.H. Barrett. 2003. Consequences of hierarchical allocation for the evolution of life-history traits. Am. Nat. 161(1): 153–167.

Wu, R., C.-X. Ma, I. Painter, and Z.-B. Zeng. 2002. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theor. Popul. Biol. 61(3): 349–363.

Wu, Y., P.R. Bhat, T.J. Close, and S. Lonardi. 2008. Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph (L Kruglyak, Ed.). PLoS Genet. 4(10): e1000212. doi: 10.1371/journal.pgen.1000212.

Wu, T.D., and S. Nacu. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873–881. doi:10.1093/bioinformatics/btq057

Würschum, T. 2012. Mapping QTL for agronomic traits in breeding populations. Theor. Appl. Genet. 125(2): 201–210.

Xiong, J.S., S.E. McKeand, F. Isik, J. Wegrzyn, D.B. Neale, Z.B. Zeng, L. da Costa e Silva, and R.W. Whetten. 2016. Quantitative trait loci influencing forking defects in an outbred pedigree of loblolly pine. BMC Genet. 17(1): 138. doi: 10.1186/s12863-016-0446-6.

Xu, S. 2008. QTL Mapping Can Benefit from Segregation Distortion. Genet. Soc Am. (951): 1–31.

Xu, Y., and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. Crop Sci. 48(2): 391. doi: 10.2135/cropsci2007.04.0191.

Yang, R.-C. 2010. Towards understanding and use of mixed-model analysis of agricultural experiments. Can. J. Plant Sci. 90(5): 605-627.

Yanovsky, E. 1936. Food plants of the North American Indians. 237th ed. US Department of Agriculture.

Yoder, J.B., J. Stanton-Geddes, P. Zhou, R. Briskine, N.D. Young, and P. Tiffin. 2014. Genomic Signature of Adaptation to Climate in *Medicago truncatula*. Genetics 196(4): 1263–1275.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38(2): 203–208.

Yu, G., D.K. Smith, H. Zhu, Y. Guan, and T.T.Y. Lam. 2017. ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8:28–36. doi:10.1111/2041-210X.12628

Yue, B., B.A. Vick, W. Yuan, and J. Hu. 2008. Mapping one of the 2 genes controlling lemon ray flower color in sunflower (*Helianthus annuus* L.). J. Hered. 99:564–567. doi:10.1093/jhered/esn033

Zavaleta, E.S., J.R. Pasari, K.B. Hulvey, and G.D. Tilman. 2010. Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. Proc. Natl. Acad. Sci. USA 107:1443–1446. doi:10.1073/pnas.0906829107

Zeegers, M., F. Rijsdijk, and P. Sham. 2004. Adjusting for Covariates in Variance Components QTL Linkage Analysis. Behav. Genet. 34(2): 127–133. doi: 10.1023/B:BEGE.0000013726.65708.c2.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136(4): 1457–1468.

Zentner, R.P., D.D. Wall, C.N. Nagy, E.G. Smith, D.L. Young, P.R. Miller, C.A. Campbell, B.G. McConkey, S.A. Brandt, G.P. Lafond, A.M. Johnston, and D.A. Derksen. 2002. Economics of Crop Diversification and Soil Tillage Opportunities in the Canadian Prairies. Agron. J. 94(2): 216–230.

Zhang, L., S. Wang, H. Li, Q. Deng, A. Zheng, S. Li, P. Li, Z. Li, and J. Wang. 2010. Effects of missing marker and segregation distortion on QTL mapping in F2 populations. Theor. Appl. Genet. 121(6): 1071–1082. doi: 10.1007/s00122-010-1372-z.

Zhang, X., A. Sallam, L. Gao, T. Kantarski, J. Poland, L.R. DeHaan, D.L. Wyse, and J.A. Anderson. 2016. Establishment and Optimization of Genomic Selection to Accelerate the Domestication and Improvement of Intermediate Wheatgrass. Plant Genome 9(1). doi: 10.3835/plantgenome2015.07.0059.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1:5–16. doi:10.3835/plantgenome2008.02.0089

Ziebell, A.L., J.G. Barb, S. Sandhu, B.T. Moyers, R.W. Sykes, C. Doeppke, K.L. Gracom, M. Carlile, L.F. Marek, M.F. Davis, S.J. Knapp, and J.M. Burke. 2013. Sunflower as a biofuels crop: An analysis of lignocellulosic chemical properties. Biomass and Bioenergy 59: 208–217.

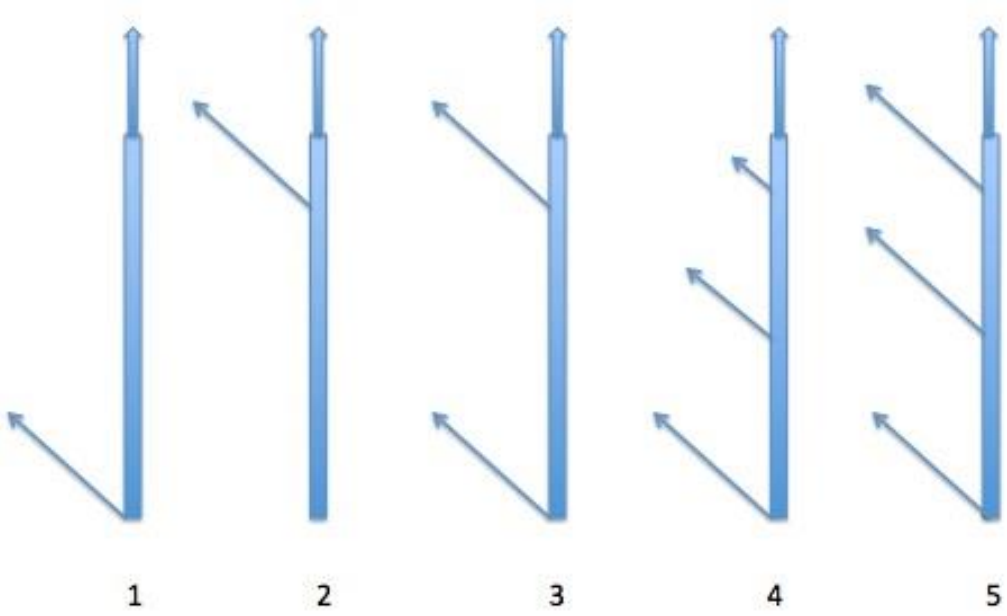
Zimmer, D.E., and D. Rehder. 1976. Rust resistance of wild *Helianthus* species of the north central United States. Phytopathology 66:208–211. doi:10.1094/Phyto-66-208.

Zohary, D. 2004. Unconscious Selection and the Evolution of Domesticated Plants. Econ. Bot. 58(1): 5–10.

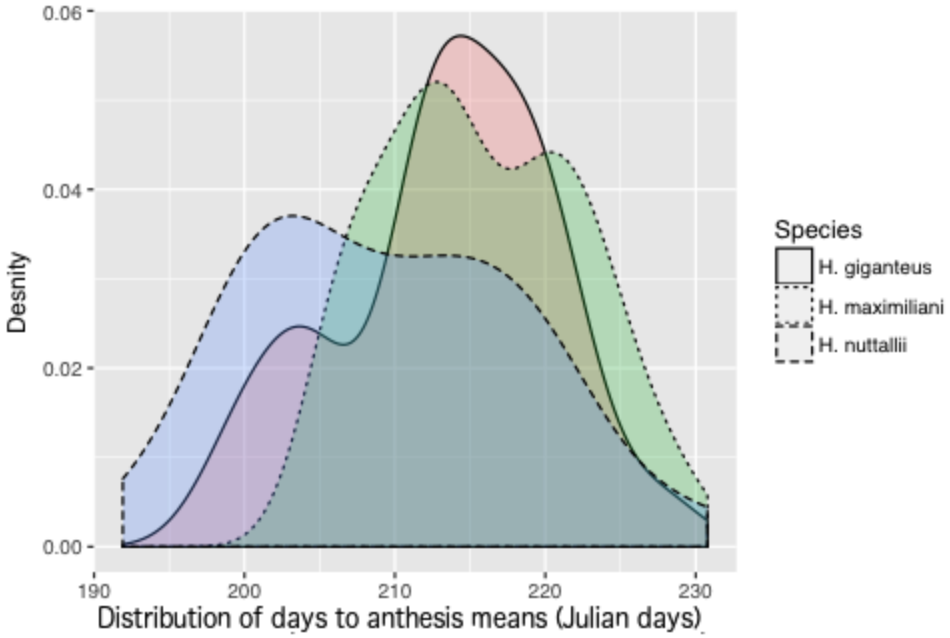
Zulliger, D., E. Schnyder, and F. Gugerli. 2013. Are adaptive loci transferable across genomes of related species? Outlier and environmental association analyses in Alpine *Brassicaceae* species. Mol. Ecol. 22(6): 1626–1639.

7.2 Appendices

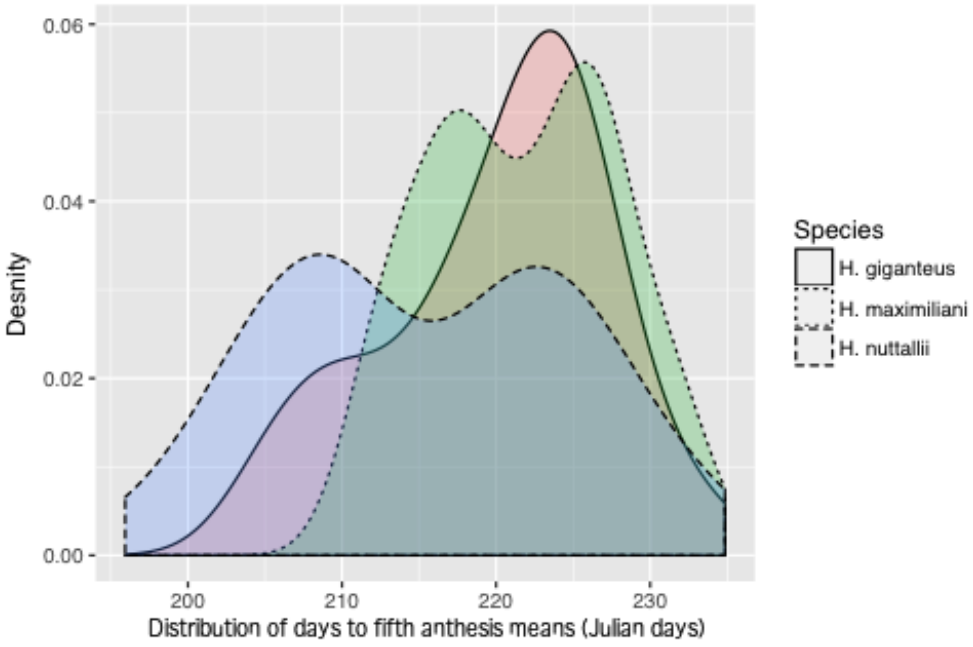
7.2.1 Appendix for Chapter 3.0



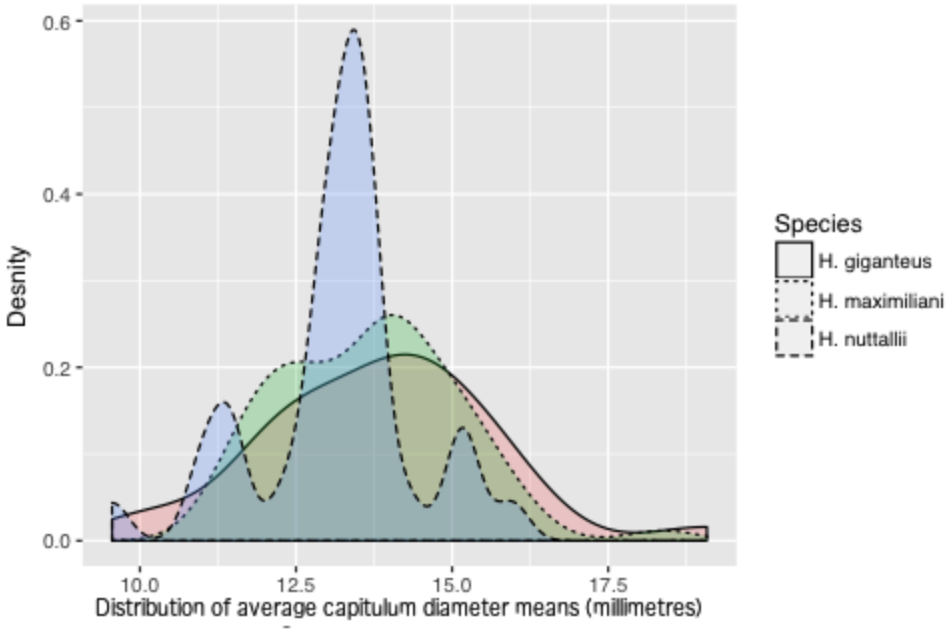
Supplemental Figure 3.1: Branching score classes: 1 = primarily basal branching only, 2= primarily apical branching, 3=apical and basal branching lacking mid-plant branching, 4= entire plant branching with apical dominance, 5 entire plant branching with no apical dominance.



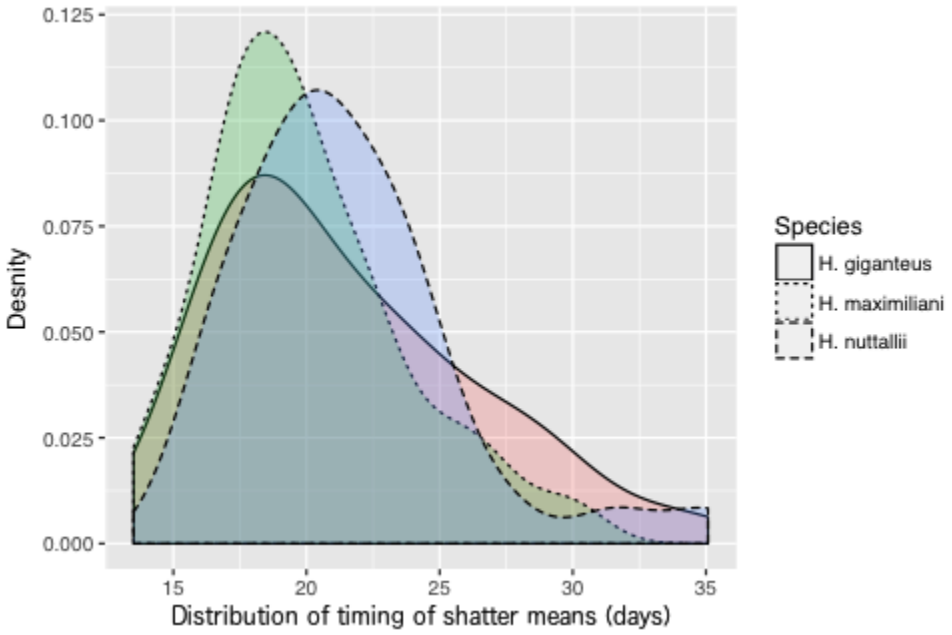
Supplemental Figure 3.2: Distribution of days to first anthesis across both wild collected and Morden collections in Julian days for *H. giganteus*, *H. maximiliani* and *H. nuttallii* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014.



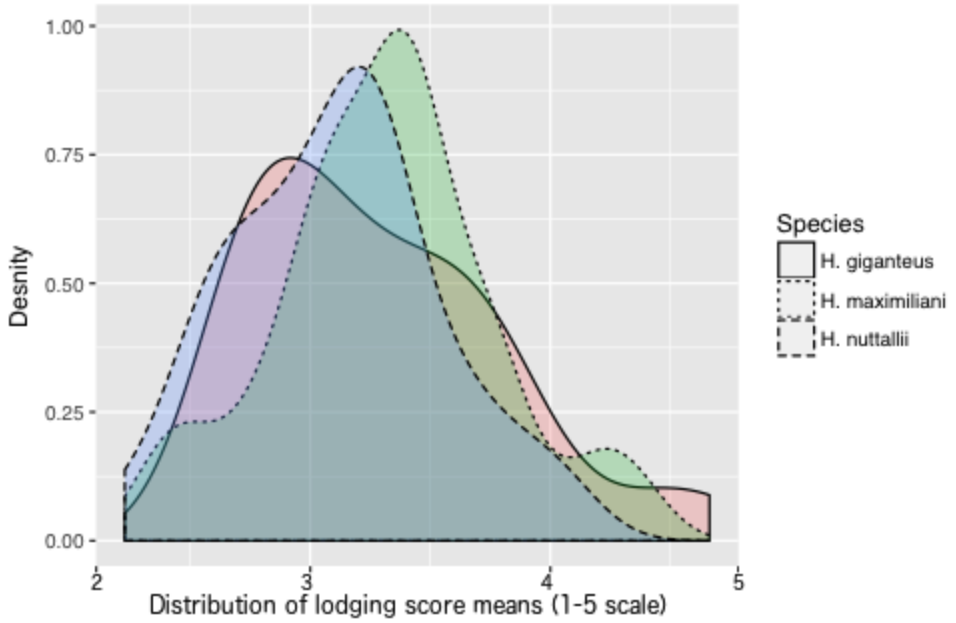
Supplemental Figure 3.3: Distribution of days to fifth anthesis across both wild collected and Morden collections in Julian days for *H. giganteus*, *H. maximiliani* and *H. nuttallii* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014.



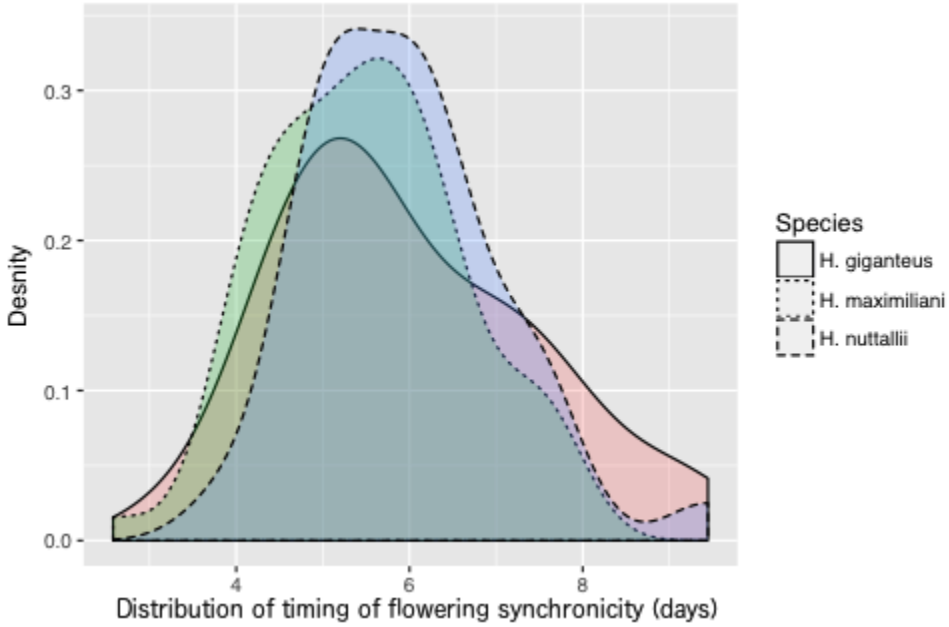
Supplemental Figure 3.4: Distribution of average capitulum diameter in millimeters across both wild collected and Morden collections for *H. giganteus*, *H. maximiliani* and *H. nuttallii* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014.



Supplemental Figure 3.5: Distribution of timing of shattering in days across both wild collected and Morden collections for *H. giganteus*, *H. maximiliani* and *H. nuttallii* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014.



Supplemental Figure 3.6: Distribution of lodging score across both wild collected and Morden collections for *H. giganteus*, *H. maximiliani* and *H. nuttallii* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014.



Supplemental Figure 3.7: Distribution of timing of flowering synchronicity in days across both wild collected and Morden collections for *H. giganteus*, *H. maximiliani* and *H. nuttallii* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014.

Supplemental Table 3.1: Mean, standard deviation and coefficient of variation for continuous traits measured in Carman and Winnipeg in 2013 and 2014 for *H. giganteus, H. maximiliani* and *H. nuttallii.*

	Carman 2013			Carman 2014			Winnipeg 2013			Winnipeg 2014		
	Mean	SD	CV%	Mean	SD	CV%	Mean	SD	CV%	Mean	SD	CV%
H. giganteus												
Days to anthesis	209.58	10.08	4.8%	213.98	7.26	3.4%	215.42	13.23	6.1%	214.87	7.66	3.6%
Days to fifth anthesis	218.77	10.59	4.8%	217.8	6.89	3.2%	222.84	12.21	5.9%	220.21	7.49	3.4%
Av. cap. diameter (mm)*	14.58	2.23	15.3%	13.95	2.06	14.8%	12.89	2.31	17.9%	14.13	2.12	15.0%
Timing of shatter (days)	23.68	8.32	35.1%	17.00	7.01	41.2%	23.88	7.05	29.5%	26.28	10.50	39.9%
Flowering synchronicity (days)	9.19	3.92	42.7%	4.01	2.04	50.9%	9.03	4.58	50.7%	5.43	3.46	63.7%
Lodging score (1-5)	3.61	1.01	28.0%	3.67	0.90	24.5%	3.03	0.95	31.4%	3.32	0.81	24.4%
H. maximiliani												
Days to anthesis	213.7	10.51	4.9%	215.6	6.05	2.8%	217.32	11.76	5.4%	216.42	7.34	3.4%
Days to fifth anthesis	222.57	9.72	4.4%	219.71	5.85	2.7%	224.68	10.57	4.7%	220.22	6.37	2.9%
Av. cap. diameter (mm)	13.99	2.05	14.7%	13.90	1.56	11.2%	12.95	2.03	15.7%	14.23	1.73	12.1%
Timing of shatter (days)	22.52	7.96	35.4%	16.43	6.74	41.0%	21.9	6.68	30.5%	24.13	10.62	44.0%
Flowering synchronicity (days)	8.70	4.74	54.5%	4.34	2.31	53.2%	8.67	4.04	46.6%	4.45	2.45	55.1%
Lodging score (1-5)	3.59	1.03	28.7%	3.75	0.94	25.1%	3.00	0.88	29.3%	3.43	0.74	21.6%
H. nuttallii												
Days to anthesis	205.77	12.18	5.9%	208.61	9.05	4.3%	210.8	15.97	7.6%	211.73	8.55	4.0%
Days to fifth anthesis	213.56	12.63	5.9%	213.11	8.99	4.2%	217.8	15.53	7.1%	216.55	8.97	4.1%
Av. cap. diameter (mm)	13.66	2.16	15.8%	13.16	1.62	12.3%	12.8	2.11	16.5%	13.36	1.53	11.5%
Timing of shatter (days)	24.83	8.29	33.4%	16.90	6.87	40.7%	27.52	6.14	22.3%	22.49	10.65	47.4%
Flowering synchronicity (days)	8.44	4.24	50.2%	4.43	2.19	49.4%	9.77	4.90	50.2%	5.02	1.98	39.4%
Lodging score (1-5)	3.29	1.04	31.6%	3.59	0.83	23.1%	2.98	0.86	28.9%	3.14	0.68	21.7%

* Note: Av. Cap. diameter= Average capitulum diameter

Supplemental Table 3.2: Correlations between continuous phenotypic traits as assessed under field conditions in Carman and Winnipeg in 2013 and 2014 for *H. giganteus* (n= 344-350).

		Days to first anthesis	Days to fifth anthesis	Flowering synchronicity	Average capitulum diameter	Timing of shatter
Days to fifth anthesis	*r*	0.90				
	p	<.0001				
Flowering synchronicity	*r*	-0.26	0.15			
	p	<.0001	0.005			
Average capitulum	*r*	0.36	0.36	-0.10		
diameter	*p*	<.0001	<.0001	0.053		
Timing of shatter	*r*	0.29	0.42	0.31	0.04	
	p	<.0001	<.0001	<.0001	0.053	
Lodging score	*r*	0.11	0.07	-0.18	0.21	0.09
	p	0.038	0.177	0.001	<.0001	0.077

Supplemental Table 3.3: Correlations between continuous phenotypic traits as assessed under field conditions in Carman and Winnipeg in 2013 and 2014 for *H. maximiliani* (n= 642-655).

		Days to first anthesis	Days to fifth anthesis	Flowering synchronicity	Average capitulum diameter	Timing of shatter
Days to fifth anthesis	*r*	0.87				
	p	<.0001				
Flowering synchronicity	*r*	-0.38	0.07			
	p	<.0001	0.070			
Average capitulum	*r*	-0.01	-0.03	-0.09		
diameter	*p*	0.710	0.323	0.029		
Timing of shatter	*r*	0.22	0.35	0.27	-0.03	
	p	<.0001	<.0001	<.0001	0.413	
Lodging score	*r*	-0.04	-0.07	-0.09	0.18	-0.06
	p	0.287	0.055	0.025	<.0001	0.099

Supplemental Table 3.4: Correlations between continuous phenotypic traits as assessed under field conditions in Carman and Winnipeg in 2013 and 2014 for *H. nuttallii* (n= 359-367).

		Days to first anthesis	Days to fifth anthesis	Flowering synchronicity	Average capitulum diameter	Timing of shatter
Days to fifth anthesis	*r*	0.94				
	p	<.0001				
Flowering synchronicity	*r*	-0.11	0.22			
	p	0.035	<.0001			
Average capitulum diameter	*r*	-0.19	-0.16	-0.01		
	p	<.0001	0.003	0.712		
Timing of shatter	*r*	0.30	0.43	0.44	-0.04	
	p	<.0001	<.0001	<.0001	0.416	
Lodging score	*r*	-0.05	-0.01	-0.12	0.16	-0.13
	p	0.302	0.807	0.022	0.003	0.013

Supplemental Table 3.5: Analysis of variance for days to anthesis in *H. giganteus* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	30	0.0719	0.0024	24.94	5.87	<.0001*
Location	1	0.0039	0.0039	14.89	4.68	0.0472*
Year	1	0.0017	0.0017	11.78	2.29	0.1568
Location x year	1	0.0021	0.0021	9.85	3.12	0.1083
Rep(location x year)	8	0.005	0.0006	222.00	6.50	<.0001*
Genotype x location	29	0.0101	0.0003	28.42	2.13	0.0241*
Genotype x year	29	0.0068	0.0002	27.37	1.41	0.1858
Genotype x location x year	28	0.0046	0.0002	222.00	1.71	0.018*
Residual	222	0.0213	0.0001	.	.	.

Note:*=Significant at alpha 0.05

Supplemental Table 3.6: Analysis of variance for days to anthesis in *H. maximiliani* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	61	0.0967	0.0016	51.07	3.28	<.0001*
Location	1	0.0022	0.0022	22.26	4.70	0.0411*
Year	1	0.0002	0.0002	25.15	0.47	0.5001
Location x year	1	0.001	0.001	13.89	2.77	0.1185
Rep(location x year)	8	0.0023	0.0003	407.00	2.88	0.004*
Genotype x location	58	0.0180	0.0003	54.86	1.68	0.028*
Genotype x year	61	0.0218	0.0004	51.51	1.88	0.0108*
Genotype x location x year	54	0.0101	0.0002	407.00	1.84	0.0005*
Residual	407	0.0412	0.0001	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.7: Analysis of variance for days to anthesis in *H. nuttallii* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	33	0.1064	0.0032	16.14	5.24	0.0005*
Location	1	0.0082	0.0082	20.19	9.46	0.0059*
Year	1	0.0008	0.0008	18.11	0.98	0.3343
Location x year	1	0.0013	0.0013	15.97	1.74	0.2051
Rep(location x year)	8	0.0047	0.0006	233.00	4.25	<.0001*
Genotype x location	29	0.0149	0.0005	27.48	1.38	0.1995
Genotype x year	33	0.0158	0.0005	25.33	1.19	0.3277
Genotype x location x year	27	0.0102	0.0004	233.00	2.75	<.0001*
Residual	233	0.032	0.0001	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.8: Analysis of variance for days to fifth anthesis in *H. giganteus* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	30	0.0642	0.0021	25.71	6.46	<.0001*
Location	1	0.0038	0.0038	17.75	8.74	0.0085*
Year	1	0.0009	0.0009	13.67	2.42	0.1425
Location x year	1	0.0003	0.0003	10.03	1.15	0.3096
Rep(location x year)	8	0.0023	0.0003	217.00	3.03	0.003*
Genotype x location	29	0.0078	0.0003	28.02	2.10	0.0264*
Genotype x year	29	0.0058	0.0002	25.43	1.52	0.144
Genotype x location x year	27	0.0035	0.0001	217.00	1.35	0.1226
Residual	217	0.0207	0.0001	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.9: Analysis of variance for days to fifth anthesis in *H. maximiliani* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	61	0.0737	0.0012	45.72	4.00	<.0001*
Location	1	0.0007	0.0007	15.37	2.30	0.1494
Year	1	0.007	0.0070	18.91	18.99	0.0003*
Location x year	1	0.0003	0.0003	10.89	1.07	0.3242
Rep(location x year)	8	0.002	0.0003	399	2.41	0.0151*
Genotype x location	58	0.011	0.0002	55.24	1.44	0.086
Genotype x year	61	0.0150	0.0002	47.44	1.83	0.0157*
Genotype x location x year	53	0.007	0.0001	399.00	1.26	0.1166
Residual	399	0.0419	0.0001	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.10: Analysis of variance for days to fifth anthesis in *H. nuttallii* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	33	0.1117	0.0034	16.60	9.60	<.0001*
Location	1	0.0053	0.0053	13.64	11.14	0.005*
Year	1	0.0002	0.0002	15.48	0.35	0.5615
Location x year	1	0.0003	0.0003	11.78	0.73	0.4093
Rep(location x year)	8	0.0036	0.0005	228.00	3.30	0.0014*
Genotype x location	29	0.007	0.0002	24.97	1.22	0.306
Genotype x year	33	0.0104	0.0003	19.79	1.51	0.168
Genotype x location x year	24	0.0048	0.0002	228.00	1.45	0.0879
Residual	228	0.0313	0.0001	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.11: Analysis of variance for average capitulum diameter in *H. giganteus* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	30	1.0199	0.034	31.31	9.12	<.0001*
Location	1	0.0453	0.0453	16.07	13.27	0.0022*
Year	1	0.0055	0.0055	18.17	1.40	0.2523
Location x year	1	0.0614	0.0614	8.84	26.31	0.0007*
Rep(location x year)	8	0.0183	0.0023	220.00	2.30	0.0221*
Genotype x location	29	0.066	0.0023	28.62	2.01	0.0331*
Genotype x year	29	0.0775	0.0027	27.02	2.35	0.0142*
Genotype x location x year	28	0.0317	0.0011	220.00	1.14	0.293
Residual	220	0.2186	0.001	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.12: Analysis of variance for average capitulum diameter in *H. maximiliani* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	61	1.1724	0.0192	30.19	8.24	<.0001*
Location	1	0.0258	0.0258	10.35	7.59	0.0197*
Year	1	0.0438	0.0438	12.96	10.86	0.0058*
Location x year	1	0.0519	0.0519	9.72	15.74	0.0028*
Rep(location x year)	8	0.0272	0.0034	407.00	2.52	0.0109*
Genotype x location	58	0.096	0.0017	55.30	1.08	0.3879
Genotype x year	61	0.1355	0.0022	48.35	1.44	0.0972
Genotype x location x year	54	0.0829	0.0015	407.00	1.14	0.2441
Residual	407	0.5492	0.0013	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.13: Analysis of variance for average capitulum diameter in *H. nuttallii* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	33	0.6004	0.0182	35.12	2.27	0.0092*
Location	1	0.0191	0.0191	19.00	2.90	0.1049
Year	1	0.0074	0.0074	23.38	0.89	0.3538
Location x year	1	0.0404	0.0404	10.84	8.87	0.0127*
Rep(location x year)	8	0.0371	0.0046	229.00	3.21	0.0018*
Genotype x location	29	0.121	0.0042	27.93	2.12	0.0254*
Genotype x year	33	0.1899	0.0058	23.71	2.83	0.0052*
Genotype x location x year	27	0.0536	0.002	229.00	1.37	0.1104
Residual	229	0.3306	0.0014	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.14: Analysis of variance for timing of shattering in *H. giganteus* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	30	2.479	0.0826	7.45	4.06	0.0262*
Location	1	0.7092	0.7092	9.02	25.49	0.0007*
Year	1	0.2775	0.2775	16.00	6.56	0.0209*
Location x year	1	0.5418	0.5418	13.28	15.23	0.0017*
Rep(location x year)	8	0.2237	0.028	213.00	2.07	0.0404*
Genotype x location	29	0.4328	0.0149	27.26	0.65	0.8756
Genotype x year	29	0.8511	0.0293	25.66	1.24	0.2915
Genotype x location x year	27	0.6267	0.0232	213.00	1.71	0.0193*
Residual	213	2.8834	0.0135	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.15: Analysis of variance for timing of shattering in *H. maximiliani* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	61	3.7987	0.0623	12.99	2.74	0.024*
Location	1	0.6166	0.6166	9.24	8.32	0.0176*
Year	1	0.5916	0.5916	10.71	7.05	0.0228*
Location x year	1	0.8443	0.8443	10.68	10.64	0.0079*
Rep(location x year)	8	0.6367	0.0796	398.00	4.34	<.0001*
Genotype x location	57	1.2272	0.0215	53.91	0.78	0.8206
Genotype x year	61	1.8032	0.0296	48.02	1.04	0.4468*
Genotype x location x year	53	1.4671	0.0277	398.00	1.51	0.0159
Residual	398	7.2972	0.0183	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.16: Analysis of variance for timing of shattering in *H. nuttallii* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	33	1.8939	0.0574	2.93	4.94	0.1099
Location	1	0.3723	0.3723	9.18	15.72	0.0032*
Year	1	1.4182	1.4182	10.54	51.40	<.0001*
Location x year	1	0.0708	0.0708	12.56	2.32	0.1527
Rep(location x year)	8	0.2082	0.026	233.00	1.62	0.1208
Genotype x location	29	0.4536	0.0156	26.33	0.69	0.8317
Genotype x year	33	0.6465	0.0196	21.91	0.83	0.6875
Genotype x location x year	25	0.5703	0.0228	233.00	1.42	0.0962
Residual	233	3.7503	0.0161	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.17: Analysis of variance for lodging score in *H. giganteus* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	30	1.347	0.0449	22.08	2.44	0.0169*
Location	1	0.2758	0.2758	9.65	9.61	0.0117*
Year	1	0.0692	0.0692	9.31	2.38	0.156
Location x year	1	0.0301	0.0301	6.98	1.25	0.301
Rep(location x year)	8	0.2165	0.0271	219.00	2.45	0.0149*
Genotype x location	29	0.4065	0.014	28.86	1.53	0.1284
Genotype x year	29	0.3983	0.0137	26.65	1.51	0.1426
Genotype x location x year	28	0.2553	0.0091	219.00	0.82	0.7226
Residual	219	2.4241	0.0111	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.18: Analysis of variance for lodging score in *H. maximiliani* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	61	1.8687	0.0306	19.54	2.19	0.0281*
Location	1	0.4298	0.4298	8.91	24.11	0.0009*
Year	1	0.256	0.256	11.36	12.02	0.005*
Location x year	1	0.0457	0.0457	9.91	2.44	0.1498
Rep(location x year)	8	0.1535	0.0192	402.00	1.59	0.1267
Genotype x location	58	0.6814	0.0117	53.95	0.92	0.6295
Genotype x year	61	0.9257	0.0152	45.78	1.18	0.2843
Genotype x location x year	53	0.6803	0.0128	402.00	1.06	0.366
Residual	402	4.862	0.0121	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.19: Analysis of variance for lodging score in *H. nuttallii* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	33	1.0878	0.033	5.01	3.47	0.0834
Location	1	0.2192	0.2192	5.59	29.20	0.0021*
Year	1	0.2001	0.2001	7.56	20.63	0.0022*
Location x year	1	0.0005	0.0005	8.37	0.04	0.8385
Rep(location x year)	8	0.0691	0.0086	234.00	0.69	0.6995
Genotype x location	29	0.3057	0.0105	28.21	0.76	0.7697
Genotype x year	33	0.4353	0.0132	23.01	0.94	0.5767
Genotype x location x year	27	0.3767	0.014	234.00	1.12	0.3211
Residual	234	2.9237	0.0125	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.20: Analysis of variance for flowering synchronicity in *H. giganteus* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	30	3.8936	0.1298	10.17	2.49	0.0627
Location	1	0.3129	0.3129	12.51	6.25	0.0272*
Year	1	7.5973	7.5973	4.61	305.95	<.0001*
Location x year	1	0.258	0.258	8.10	7.12	0.0281*
Rep(location x year)	8	0.2127	0.0266	213	0.66	0.7237
Genotype x location	29	1.8354	0.0633	28.10	1.30	0.2421
Genotype x year	29	1.0918	0.0376	25.18	0.77	0.7562
Genotype x location x year	27	1.316	0.0487	213.00	1.22	0.222
Residual	213	8.541	0.0401	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.21: Analysis of variance for flowering synchronicity in *H. maximiliani* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	61	4.4375	0.0727	47.52	0.69	0.9175
Location	1	0.001	0.001	11.29	0.02	0.9024
Year	1	10.1933	10.1933	24.71	93.52	<.0001*
Location x year	1	0.001	0.001	8.87	0.02	0.8948
Rep(location x year)	8	0.4721	0.059	384.00	1.39	0.1998
Genotype x location	57	2.7995	0.0491	53.95	1.21	0.2367
Genotype x year	61	5.9504	0.0975	43.36	2.43	0.0013*
Genotype x location x year	52	2.1007	0.0404	384.00	0.95	0.575
Residual	384	16.325	0.0425	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.22: Analysis of variance for flowering synchronicity in *H. nuttallii* as assessed in Carman and Winnipeg in 2013 and 2014 using SAS MIXED METHOD=TYPE 3.

Source	DF	Sum of Square	Mean Square	Error DF	*F* value	Pr>*F*
Genotype	33	1.8402	0.0558	17.05	0.87	0.6485
Location	1	0.3686	0.3686	6.57	16.23	0.0057*
Year	1	4.7998	4.7998	11.05	144.06	<.0001*
Location x year	1	0.0184	0.0184	3.75	1.13	0.3514
Rep(location x year)	8	0.129	0.0161	221.00	0.38	0.9314
Genotype x location	29	1.2499	0.0431	25.67	1.20	0.3211
Genotype x year	33	1.8493	0.056	16.77	1.62	0.1478
Genotype x location x year	24	0.8566	0.0357	221.00	0.84	0.6865
Residual	221	9.4179	0.0426	.	.	.

Note: *=Significant at alpha 0.05

Supplemental Table 3.23: Summary of environmental variables showing at least one significant Spearman ranked correlation to phenotypic means in *H. giganteus* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014 (n=15).

Environmental variables	Days to first anthesis		Days to fifth anthesis		Average Capitulum diameter	
	r†	*p*‡	*r*	*p*	*r*	*p*
Lat	-0.58	0.023*	-0.60	0.019*	-0.49	0.064
BIO1	0.66	0.007*	0.68	0.006*	0.56	0.030*
BIO5	0.56	0.044*	0.61	0.015*	0.36	0.188
BIO6	-0.56	0.030*	-0.59	0.022*	-0.57	0.026*
BIO7	-0.38	0.152	-0.35	0.205	-0.54	0.037*
BIO8	0.61	0.015*	0.64	0.011*	0.52	0.048*
BIO10	0.61	0.015*	0.64	0.011*	0.52	0.048*
BIO11	-0.55	0.033*	-0.60	0.017*	-0.49	0.066
TMEAN4	0.70	0.004*	0.77	0.001*	0.60	0.019*
TMEAN5	0.70	0.004*	0.73	0.002*	0.61	0.015*
TMEAN6	0.64	0.010*	0.64	0.010*	0.54	0.038*
TMEAN7	0.57	0.027*	0.60	0.019*	0.46	0.087
TMEAN8	0.64	0.010*	0.66	0.007*	0.55	0.032*
TMEAN9	0.61	0.016*	0.59	0.022*	0.44	0.094
TMEAN10	0.63	0.012*	0.60	0.018*	0.51	0.053
TMEAN11	-0.62	0.013*	-0.62	0.013*	-0.48	0.066

Note: * Significant at the 0.05 probability level.

† *r* denotes Spearman's ranked correlation coefficient.

‡ *p* denotes *p*-values.

Supplemental Table 3.24: Summary of environmental variables showing at least one significant Spearman ranked correlation to phenotypic means in *H. maximiliani* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014 (n=40).

Environmental variables	Days to first anthesis		Days to fifth anthesis		Average Capitulum diameter	
	r †	*p*‡	*r*	*p*	*r*	*p*
Lat	-0.56	<.0001*	-0.61	<.0001*	-0.15	0.344
FRAIN	0.41	0.007*	0.33	0.040*	-0.27	0.088
FWDEM	0.41	0.008*	0.45	0.003*	0.19	0.220
BIO1	0.35	0.026*	0.38	0.017*	0.20	0.196
BIO6	-0.37	0.020*	-0.39	0.014*	-0.25	0.133
BIO7	-0.45	0.004*	-0.51	0.001*	-0.11	0.483
BIO16	0.15	0.356	0.24	0.137	-0.32	0.042*
BIO18	0.15	0.356	0.24	0.137	-0.32	0.042*
TMEAN4	0.34	0.033*	0.36	0.024*	0.24	0.129
TMEAN5	0.44	0.005*	0.47	0.002*	0.18	0.258
TMEAN6	0.41	0.009*	0.43	0.005*	0.15	0.339
TMEAN8	0.30	0.057	0.34	0.032*	0.20	0.211
TMEAN9	0.36	0.024*	0.42	0.008*	0.07	0.642
TMEAN10	0.25	0.109	0.32	0.046*	0.08	0.582
TMEAN11	-0.38	0.016*	-0.43	0.005*	-0.17	0.267

Note: * Significant at the 0.05 probability level.

† *r* denotes Spearman's ranked correlation coefficient.

‡ *p* denotes *p*-values.

Supplemental Table 3.25: Summary of environmental variables showing at least one significant Spearman ranked correlation to phenotypic means for *H. nuttallii* as assessed under field conditions in Carman and Winnipeg in 2013 and 2014 (n=22).

Environmental variables	Days to first anthesis		Days to fifth anthesis		Lodging score		Timing of shatter		Average Capitulum diameter	
	r†	*p*‡	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Lat	-0.67	0.001*	-0.71	<.0001*	-0.30	0.168	-0.83	<.0001	0.36	0.093
Long	-0.69	<.0001*	-0.75	<.0001*	-0.06	0.764	-0.79	<.0001*	0.44	0.043*
Elevation	-0.55	0.008*	-0.61	0.003*	0.06	0.774	-0.67	0.001*	0.21	0.342
AWHC	0.38	0.077	0.42	0.050*	-0.18	0.412	0.49	0.020*	-0.07	0.731
LSF0	0.40	0.062	0.44	0.039*	0.09	0.669	0.62	0.002*	0.03	0.867
LSFNEG2	-0.24	0.268	-0.23	0.300	0.11	0.620	-0.03	0.885	0.48	0.023*
FFF0	0.40	0.062	0.44	0.039*	0.09	0.669	0.62	0.002*	0.03	0.867
FFFNEG2	-0.24	0.268	-0.23	0300	0.11	0.620	-0.03	0.885	0.48	0.023*
FFDNEG2	-0.24	0.268	-0.30	0.180	0.18	0.408	-0.09	0.65	0.52	0.013*
GDDFOR	0.47	0.027*	0.450	0.019*	0.37	0.08	0.68	0.001*	-0.12	0.566
FRAIN	0.53	0.012*	0.57	0.006*	0.52	0.013*	0.71	<0.001*	-0.17	0.426
FWDEM	0.63	0.002*	0.63	0.002*	0.50	0.017*	0.74	<.0001*	-0.33	0.130
BIO1	0.44	0.038*	0.45	0.035*	0.43	0.048*	0.63	0.002*	-0.22	0.306
BIO2	0.07	0.177	0.07	0.752	-0.49	0.021*	0.01	0.968	-0.07	0.745
BIO4	0.30	0.161	0.34	0.125	-0.48	0.023*	0.23	0.301	-0.17	0.439
BIO5	0.57	0.005*	0.58	0.005*	0.33	0.125	0.77	<.0001*	-0.11	0.605
BIO6	-0.19	0.394	-0.17	0.436	-0.61	0.002*	-0.36	0.099	-0.29	0.179
BIO7	0.18	0.412	0.21	0.347	-0.49	0.020*	0.12	0.778	0.09	0.660
BIO8	0.57	0.006*	0.58	0.004*	0.45	0.035*	0.78	<.0001*	-0.10	0.664
BIO9	-0.27	0.225	-0.32	0.143	-0.51	0.016*	-0.42	0.050*	-0.29	0.188
BIO10	0.57	0.006*	0.58	0.004*	0.45	0.035*	0.78	<.0001*	-0.29	0.188
BIO11	-0.40	0.067	-0.40	0.067	-0.57	0.006*	-0.60	0.003*	0.23	0.300
BIO12	0.65	0.001*	0.72	<0.001*	-0.08	0.710	0.72	<.0001*	-0.37	0.136
BIO13	0.72	<.0001*	0.77	<.0001*	0.10	0.643	0.81	<.0001*	-0.36	0.090
BIO14	0.27	0.231	0.30	0.168	-0.43	0.044*	0.15	0.494	0.07	0.752

Environmental variables	Days to first anthesis		Days to fifth anthesis		Lodging score		Timing of shatter		Average Capitulum diameter	
BIO15	0.10	0.655	0.08	0.705	0.58	0.005*	0.22	0.305	-0.08	0.720
BIO16	0.73	<.0001*	0.79	<.0001*	0.09	0.681	0.82	<.0001*	-0.29	0.183
BIO17	0.10	0.654	0.14	0.510	-0.59	0.008*	-0.02	0.907	<0.01	0.985
BIO18	0.73	<.0001*	0.79	<.0001*	0.09	0.681	0.82	<.0001*	-0.29	0.183
BIO19	0.10	0.659	0.14	0.510	-0.59	0.008*	-0.02	0.906	<-0.01	0.985
TMEAN4	0.52	0.013*	0.53	0.011*	0.48	0.024*	0.71	<.0001*	-0.23	0.300
TMEAN5	0.51	0.017*	0.51	0.015*	0.51	0.017*	0.69	<.0001*	-0.24	0.280
TMEAN6	0.57	0.006*	0.57	0.006*	0.45	0.038*	0.75	<.0001*	-0.28	0.210
TMEAN7	0.56	0.008*	0.58	0.005*	0.41	0.054	0.78	<.0001*	-0.27	0.224
TMEAN8	0.60	0.003*	0.61	0.003*	0.45	0.037*	0.79	<.0001*	-0.33	0.132
TMEAN9	0.60	0.003*	0.62	0.002*	0.41	0.054	0.79	<.0001*	-0.32	0.146
TMEAN10	0.53	0.012*	0.54	0.009*	0.45	0.037*	0.73	<.0001*	-0.34	0.117
TMEAN11	-0.60	0.004*	-0.62	0.002*	-0.40	0.064	-0.80	<.0001*	0.35	0.110

Note: * Significant at the 0.05 probability level.

† *r* denotes Spearman's ranked correlation coefficient.

‡ *p* denotes *p*-values.

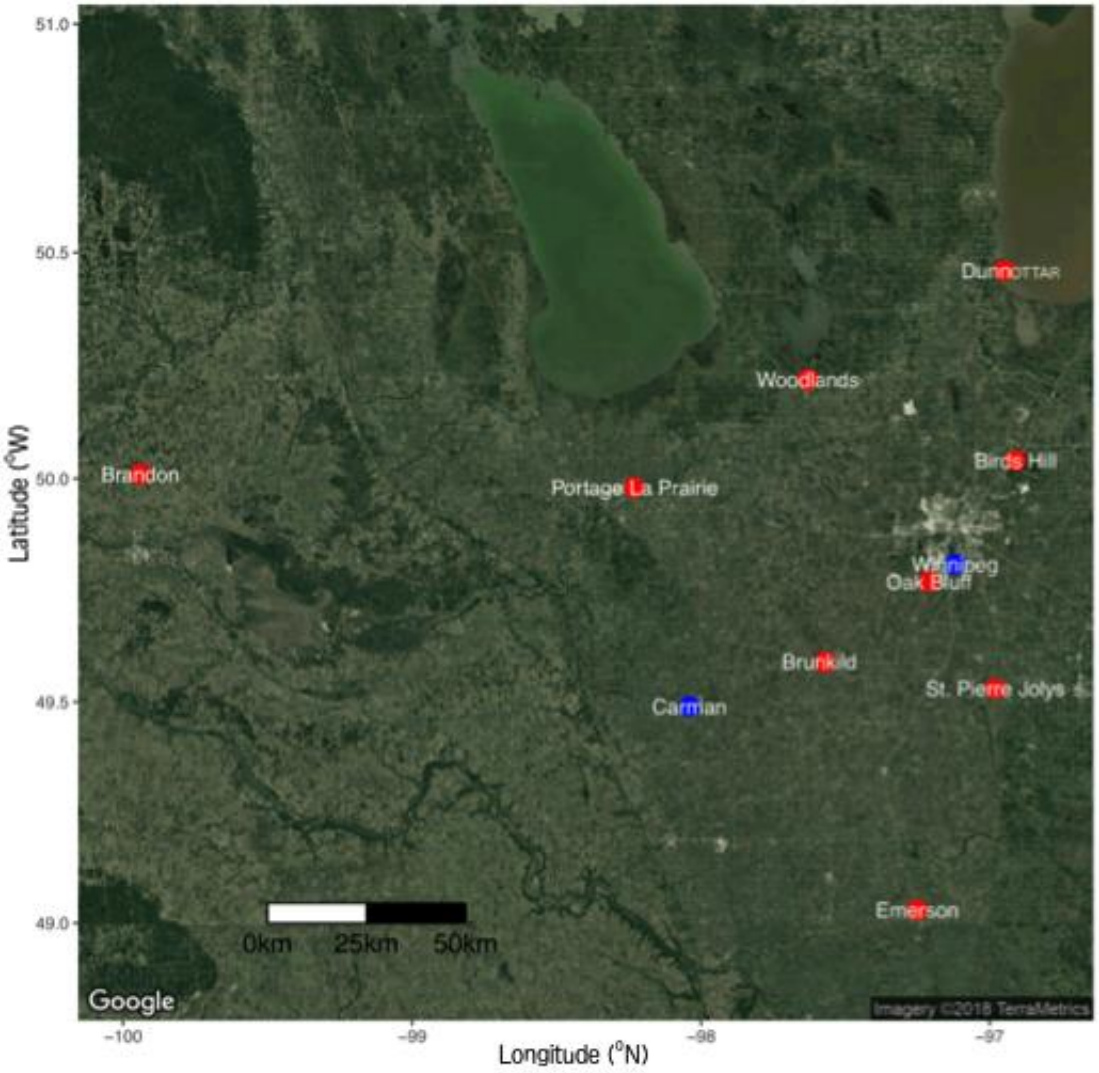
7.2.2 Appendix for Chapter 4.0

Supplemental Table 4.1: Description and source of environmental variables employed in environmental association and correlation analyses

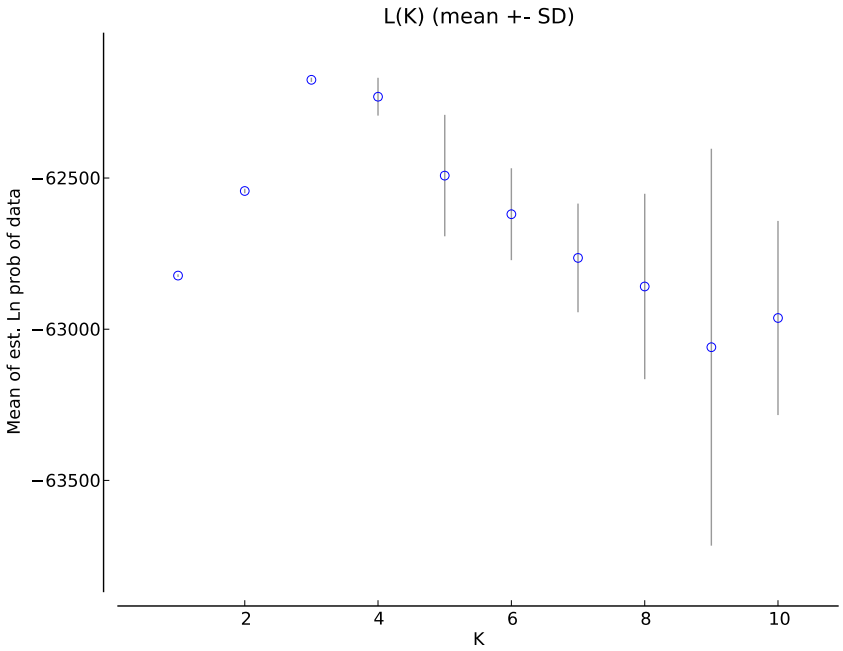
Environmental variable	Description	Source
Latitude (°)	Site latitude	GPS
Longitude(°)	Site longitude	GPS
Elevation (m)	Site elevation	GPS
Soil water-holding capacity	–	Nadler (2007)
Corn heat units (CHU)	Annual corn heat units averaged over 30 yr (1971–2000)	Nadler (2007)
Frost-free days (0□C)	Avg. frost-free period above 0□C averaged over 30 yr (1971–2000)	Nadler (2007)
Soil pH	Hydrogen ion concentration using 1:1 soil water method	AgVise
Soil organic matter (%)	Percentage or organic matter in soil at 0–15 cm using loss on ignition method	AgVise
Soil N content (kg ha^{-1})	Soil N content in kg ha^{-1}at 15–60 cm using 0.2 M KCl extraction method using Cd reduction	AgVise
Soil P content (mg kg^{-1})	Olsen-bicarbonate P test, P in mg kg^{-1} at 0–15 cm	AgVise
Soil K content (mg kg^{-1})	Ammonium acetate exchangeable K test, soil K content in mg kg^{-1} at 0–15 cm	AgVise
Soil Ca content (mg kg^{-1})	Soil Ca content in mg kg^{-1} at 0–15 cm (DTPA* method)	AgVise
Soil Mg content (mg kg^{-1})	Soil Mg in mg kg^{-1} at 0–15 cm (DTPA method)	AgVise
Soil Zn content (mg kg^{-1})	Soil Zn content in mg kg^{-1} at 0–15 cm (DTPA method)	AgVise
Soil Cl content (mg kg^{-1})	Soil Cl content in mg kg^{-1} at 15–60 cm (DTPA method)	AgVise
Soil Cu content (mg kg^{-1})	Soil Cu content in mg kg^{-1} at 0–15 cm (DTPA method)	AgVise
Soil B content (mg kg^{-1})	Soil B content in mg kg^{-1} at 0–15 cm (DTPA method)	AgVise
Soil Fe content (mg kg^{-1})	Soil Fe content in mg kg^{-1} at 0–15 cm (DTPA method)	AgVise
Soil Mn content (mg kg^{-1})	Soil Mn in mg kg^{-1} at 0–15 cm (DTPA method)	AgVise
Soil Na content (mg kg^{-1})	Soil Na content in mg kg^{-1} at 0–15 cm (DTPA method)	AgVise
Cation exchange capacity (meq)	Index of a soils ability to hold all cations (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, H^{+}) calculated using summation method	AgVise

Environmental variable	Description	Source
$CaCO_3$ equivalent (%)	Total $CaCO_3$ and $MgCO_3$ present in soil as precipitated soil or crystal determined using a modified Williams method using a pressure transducer	AgVise
Annual mean temperature (°C)	–	WorldClim
Annual mean precipitation (°C)	–	WorldClim
Avg. monthly temperature in June (°C)	–	WorldClim

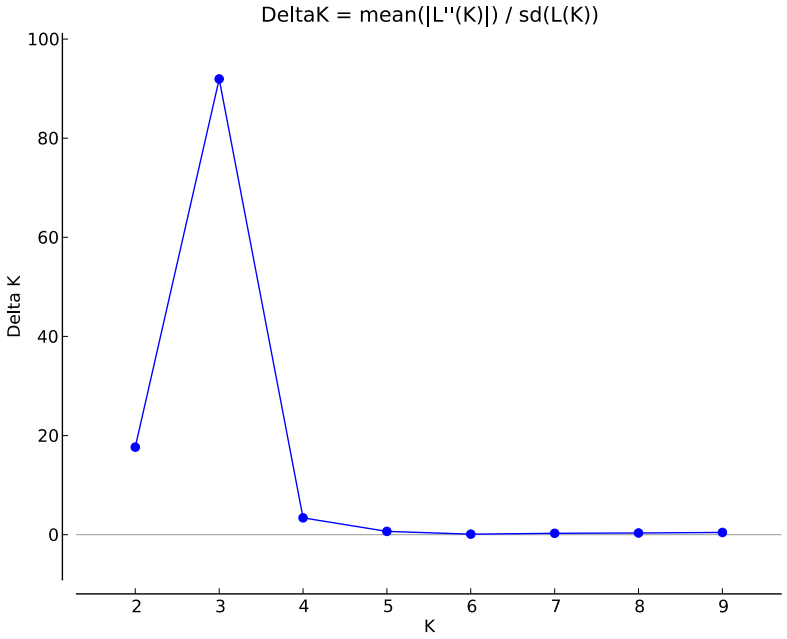
* DTPA= Diethylenetriaminepentaacetic acid



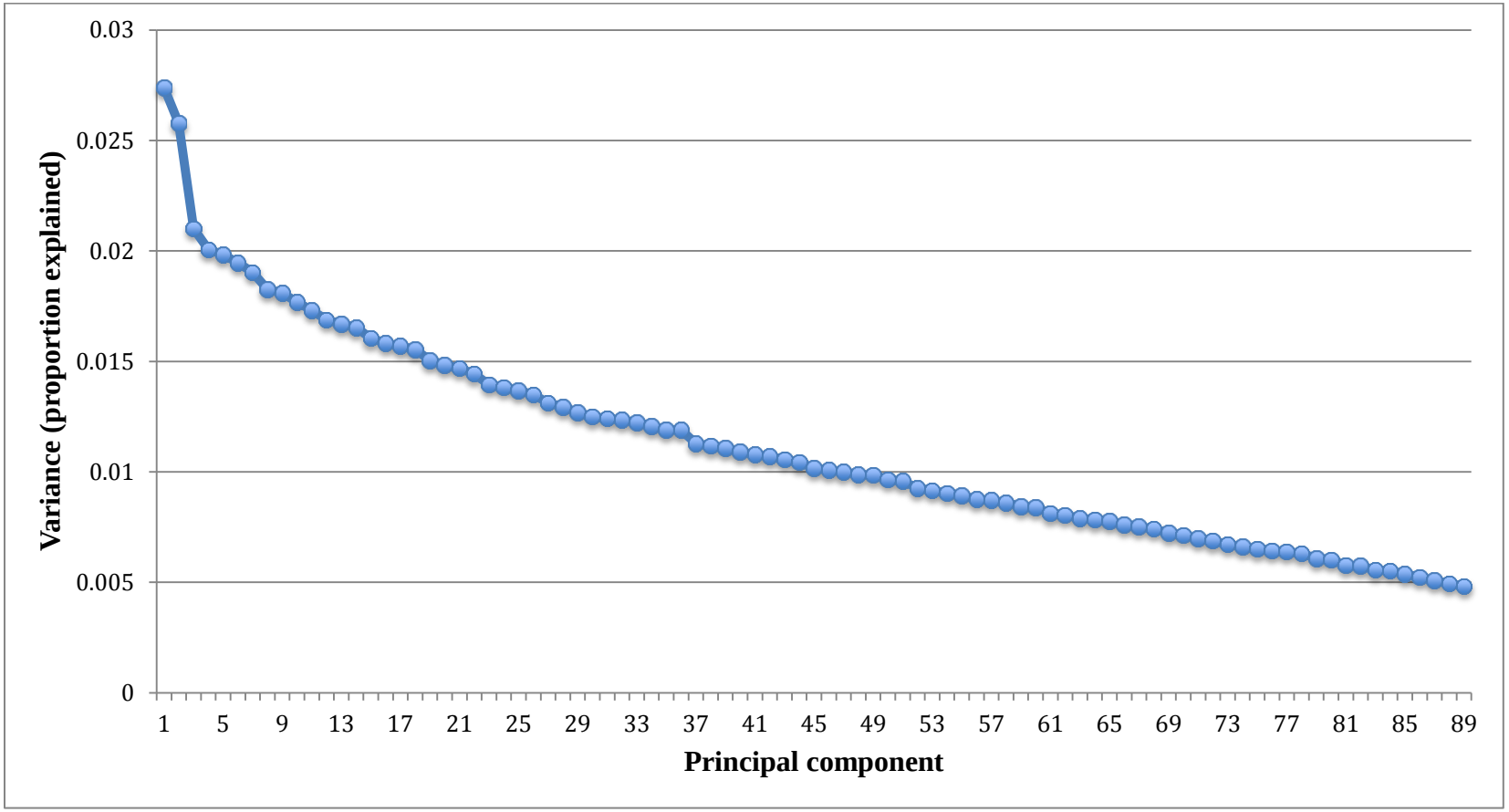
Supplemental Figure 4.1: The location of the Maximilian sunflower collection sites sampled across southern Manitoba in 2012 are indicated by red dots and the location of the two common garden sites used for evaluation in 2013 and 2014 are indicated by blue dots.



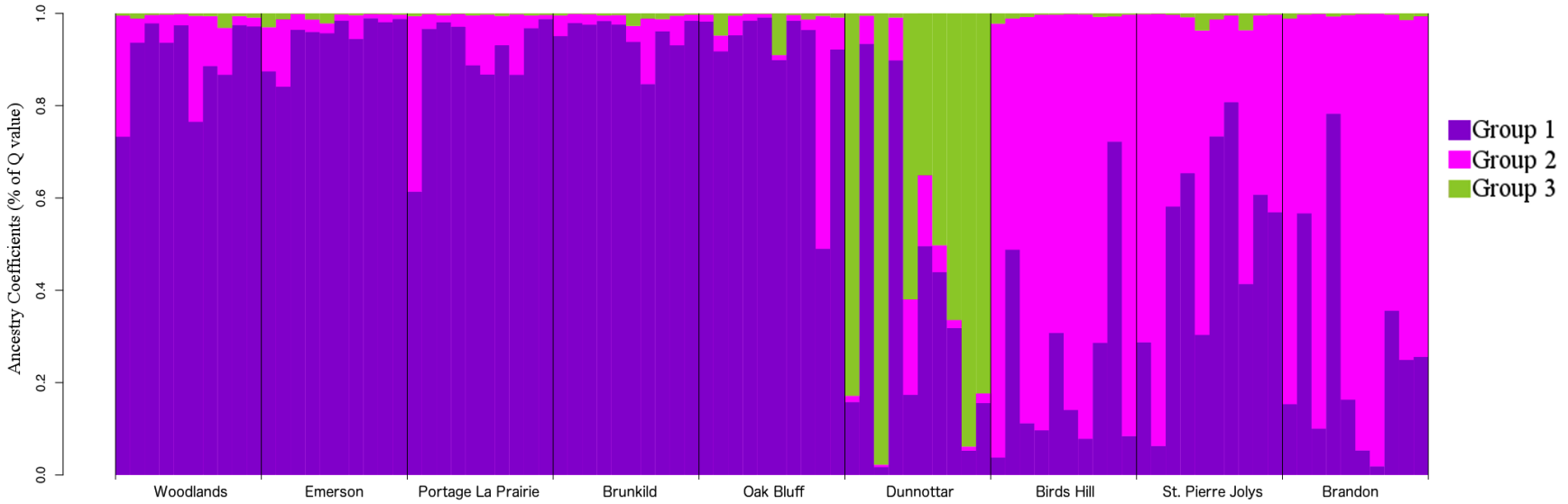
Supplemental Figure 4.2: Mean likelihood of models K=1-10 +/- 1 standard deviation as produced in STRUCTURE for 90 Maximilian sunflower samples examined using genotype-by-sequencing.



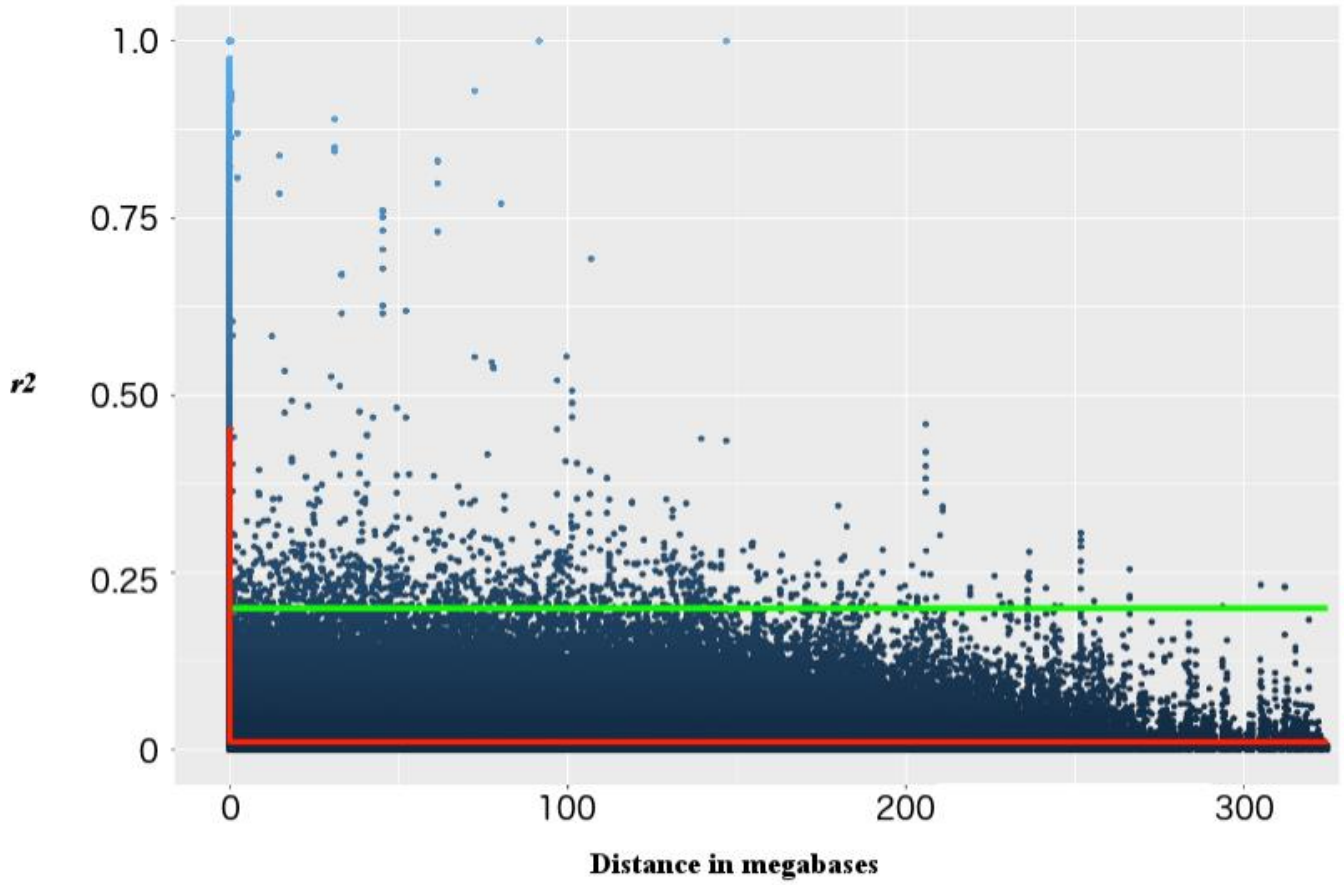
Supplemental Figure 4.3: ƗK likelihood of models K=1-10 as produced in STRUCTURE following Evanno's method for 90 Maximilian sunflower samples examined using genotype-by-sequencing.



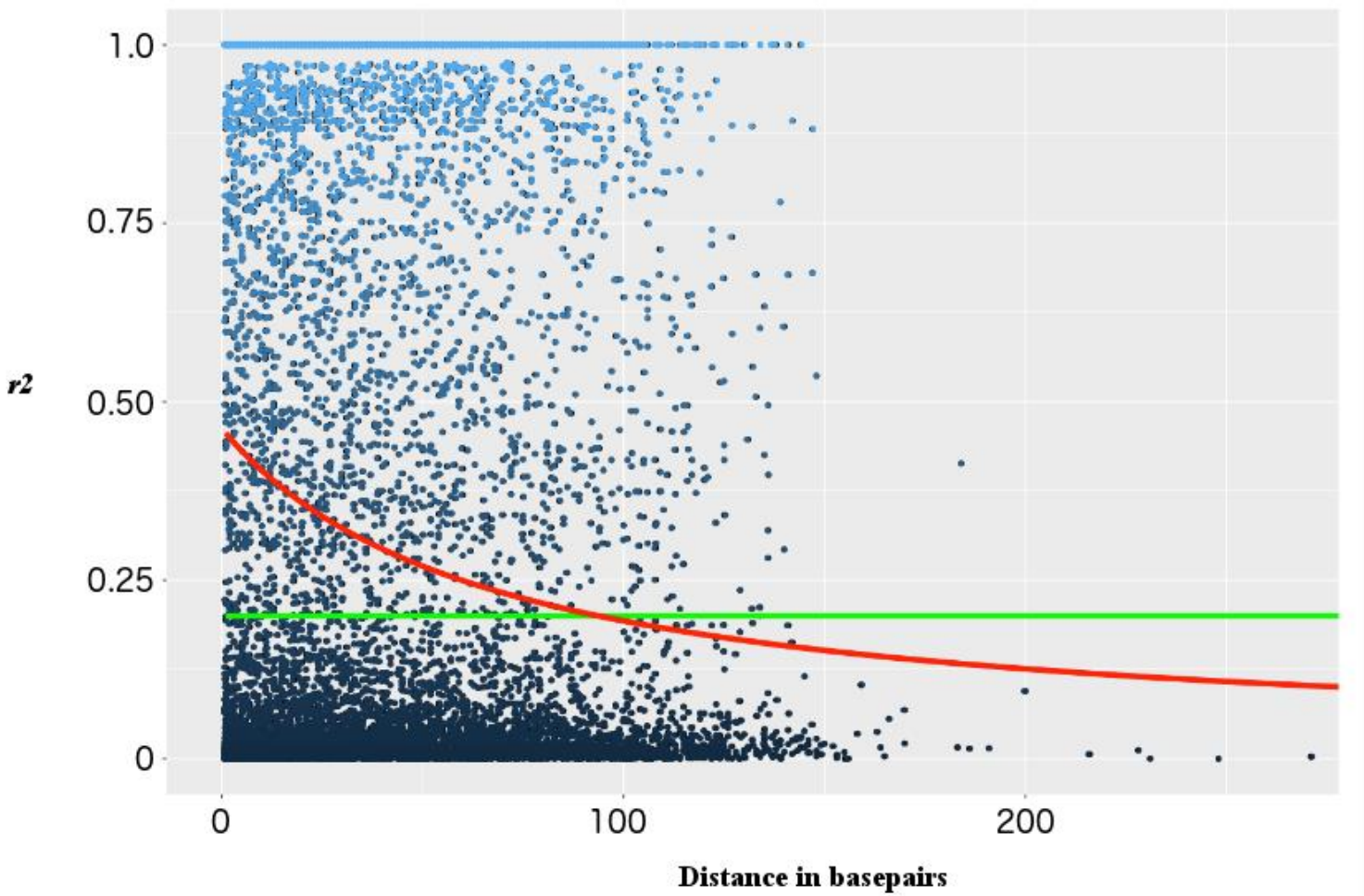
Supplemental Figure 4.4: Scree plot showing the breakdown of principal components describing population structure for the filter10 dataset.



Supplemental Figure 4.5: STRUCTURE plot of 90 *H. maximiliani* individuals organized by collection site showing the proportion of ancestry coefficients (*Q*) for *K*=3 population clusters. Note: Each bar represents the ancestry coefficients of a single individual.



Supplemental Figure 4.6: Linkage disequilibrium spanning a ~350 MB distance between SNPs. Note: the green line denotes an r^2 value of 0.2, while the red line denotes the expected decay of linkage as per Hill and Weir (1988).



Supplemental Figure 4.7: Linkage disequilibrium spanning a ~275 BP distance between SNPs. Note: the green line denotes an r^2 value of 0.2, while the red line denotes the expected decay of linkage as per Hill and Weir (1988).

Supplemental Table 4.2: Total marker distribution by linkage group in unfiltered and filtered *H. maximiliani* SNP datasets called to the *H. annuus* reference genome HA412.v.1.1.bronze.20142015 assembly.

	Dataset†			
Linkage groups	Unfiltered	Filter20	Filter20c	Filter10
1	1,132	226	226	43
2	1,417	281	281	46
3	1,324	309	309	71
4	1,190	267	267	61
5	1,830	395	395	61
6	585	155	155	24
7	650	147	147	35
8	1,468	302	302	61
9	1,516	290	290	45
10	2,075	439	439	82
11	1,160	266	266	62
12	923	195	195	48
13	1,703	400	400	69
14	1,295	286	286	56
15	1,243	269	269	58
16	1,306	291	291	55
17	1,620	347	347	53
HA0_73Ns‡	2,520	549	-	79
Total	29,957	5,414	4,865	1,009

Note: †The Filter20 and Filter10 datasets contain no more than 20% and 10% missing data respectively, while the Filter20c dataset contains the same data as Filter20 but with the exclusion of HA0_73Ns.
‡Unordered *H. annuus* contig.

Supplemental Table 4.3: Spearman ranked correlation coefficients between population structure (PCA/*Q*) as assessed in *pcadapt* and STRUCTURE and environmental variables collected from Environment Canada, WorldClim and soil sampling analysis datasets used in environmental association analysis (n=9).

		Principal Components		**Ancestry Coefficients**		
		PC1	**PC2**	**G1**	**G2**	**G3**
Topography	**Statistic**					
Latitude	*r*†	-0.62	0.30	-0.63	0.07	-0.03
	p‡	0.077	0.433	0.067	0.865	0.932
Longitude	*r*	-0.35	0.25	-0.28	-0.08	0.60
	p	0.356	0.517	0.46	0.831	0.088
Elevation	*r*	0.05	-0.62	-0.08	0.50	-0.63
	p	0.898	0.077	0.831	0.171	0.067
Temperature and precipitation	**Statistic**					
Corn Heat Units	*r*	0.57	0.08	0.47	0.07	-0.12
	p	0.112	0.831	0.205	0.865	0.765
Frost free days above 0°C	*r*	0.08	0.42	0.28	-0.28	0.07
	p	0.831	0.265	0.46	0.46	0.865
Annual temperature	*r*	0.65	-0.20	0.72	-0.12	-0.22
	p	0.06	0.603	0.028*	0.78	0.572
Annual precipitation	*r*	-0.35	0.54	-0.31	-0.20	0.70
	p	0.354	0.13	0.418	0.604	0.035*
Average daily June temperature	*r*	0.79	0.07	0.87	-0.31	-0.08
	p	0.011*	0.863	0.002*	0.415	0.846
Soil characteristics	**Statistic**					
Soil water holding capacity	*r*	0.03	0.25	0.03	-0.3	0.10
	p	0.939	0.518	0.939	0.435	0.799
Soil pH	*r*	0.60	-0.01	0.48	0.04	0.04
	p	0.091	0.982	0.19	0.929	0.929
Percent organic matter	*r*	-0.40	<0.001	-0.27	0.03	-0.17
	p	0.286	1	0.488	0.932	0.668
N	*r*	0.03	-0.11	0.14	-0.04	0.02
	p	0.932	0.781	0.715	0.915	0.966
P	*r*	0.07	0.06	0.21	-0.09	-0.20
	p	0.862	0.879	0.583	0.828	0.598
K	*r*	<0.001	0.44	0.15	-0.51	0.23
	p	1	0.232	0.699	0.16	0.559
Ca	*r*	0.18	0.27	0.23	-0.20	0.20
	p	0.637	0.488	0.546	0.606	0.606
Mg	*r*	-0.17	-0.32	-0.22	0.32	-0.60
	p	0.668	0.406	0.576	0.332	0.088

		Principal Components		Ancestry Coefficients		
		PC1	**PC2**	**G1**	**G2**	**G3**
Zn	*r*	-0.27	-0.27	-0.33	0.35	-0.43
	p	0.488	0.488	0.381	0.356	0.244
Cl	*r*	0.05	0.12	0.05	0.03	-0.10
	p	0.898	0.765	0.898	0.932	0.798
Cu	*r*	0.17	-0.20	0.13	0.08	0.13
	p	0.668	0.606	0.732	0.831	0.732
B	*r*	0.38	-0.22	0.43	<0.001	-0.68
	p	0.309	0.576	0.244	1	0.042*
Fe	*r*	-0.50	-0.07	-0.58	0.25	0.12
	p	0.171	0.865	0.099	0.517	0.765
Mn	*r*	-0.08	-0.37	-0.02	0.17	-0.35
	p	0.831	0.332	0.966	0.668	0.356
Na	*r*	0.43	<0.001	0.30	0.15	-0.10
	p	0.244	1	0.433	0.7	0.798
Cation exchange capacity	*r*	0.17	0.02	0.18	<0.001	-0.33
	p	0.668	0.966	0.637	1	0.381
$CaCO_3$ equivalent	*r*	0.48	-0.23	0.30	0.10	-0.07
	p	0.188	0.546	0.433	0.798	0.865

Note: * Significant at the 0.05 probability level.

† *r* denotes Spearman's ranked correlation coefficient.

‡ *p* denotes *p*-values.

Supplemental Table 4.4: Spearman ranked correlations between population structure (PCA/*Q*) assessed in *pcadapt* and STRUCTURE and phenotypic data collected in both the growth chamber and common garden conditions in Carman and Winnipeg in 2013 and 2014 (n=9).

		Principal Components		**Ancestry Coefficients**		
		PC1	**PC2**	**G1**	**G2**	**G3**
Common gardens	**Statistic**					
Days to anthesis	*r*†	0.07	-0.40	0.10	0.20	0.17
	p‡	0.865	0.286	0.798	0.606	0.668
Days to fifth anthesis	*r*	0.23	-0.35	0.17	0.25	0.15
	p	0.546	0.356	0.668	0.517	0.7
Growth chamber	**Statistic**					
Average capitulum diameter	*r*	0.70	0.02	0.70	-0.20	-0.05
	p	0.036*	0.966	0.036*	0.606	0.898
Days to anthesis	*r*	0.63	-0.05	0.52	-0.05	0.30
	p	0.067	0.898	0.154	0.898	0.433
Stem diameter	*r*	0.63	-0.13	0.65	-0.02	-0.25
	p	0.067	0.732	0.058	0.966	0.517
Diameter of first capitulum	*r*	0.68	<0.001	0.67	-0.18	-0.13
	p	0.042*	1	0.050*	0.637	0.732
Total nodes	*r*	0.88	0.12	0.83	-0.30	0.22
	p	0.002*	0.765	0.005*	0.433	0.576
Plant height	*r*	0.67	-0.07	0.70	-0.23	0.13
	p	0.050*	0.865	0.036*	0.546	0.732

Note: * Significant at the 0.05 probability level.

† *r* denotes Spearman's ranked correlation coefficient.

‡ *p* denotes *p*-values.

Supplemental Table 4.5: Spearman ranked correlation coefficients between the environmental dataset assembled from Environment Canada, WorldClim and soil sampling analysis datasets and phenotypic data collected in the growth chamber and under common garden conditions in Winnipeg and Carman in 2013 and 2014 (n=9).

		Common gardens		**Growth chamber**					
		Days to anthesis (days)	**Days to fifth anthesis (days)**	**Average capitulum diameter (mm)**	**Days to anthesis (days)**	**Stem diameter (mm)**	**Diameter of first capitulum (mm)**	**Total nodes**	**Plant height (cm)**
Topography	**Statistic**								
Latitude	*r*†	-0.70	-0.70	-0.72	-0.80	-0.72	-0.65	-0.83	-0.92
	p‡	0.036*	0.036*	0.030*	0.010*	0.030*	0.058	0.005*	0.001*
Longitude	*r*	0.08	0.05	-0.17	0.22	-0.33	-0.27	-0.17	-0.03
	p	0.831	0.898	0.668	0.576	0.381	0.488	0.668	0.932
Elevation	*r*	0.03	-0.02	-0.18	-0.15	0.08	-0.13	0.03	-0.17
	p	0.932	0.966	0.637	0.7	0.831	0.732	0.932	0.668
Temperature and precipitation	**Statistic**								
Corn Heat Units	*r*	0.05	0.17	0.25	0.20	0.48	0.20	0.52	0.15
	p	0.898	0.668	0.517	0.606	0.188	0.606	0.154	0.7
Frost free days above 0°C	*r*	0.03	-0.13	-0.05	-0.08	0.25	-0.08	0.12	-0.02
	p	0.932	0.732	0.898	0.831	0.517	0.831	0.765	0.966
Mean annual temperature	*r*	0.47	0.45	0.72	0.61	0.82	0.66	0.73	0.82
	p	0.201	0.23	0.028*	0.079	0.007*	0.051	0.025*	0.007*
Mean annual precipitation	*r*	-0.16	-0.16	-0.35	0.03	-0.49	-0.43	-0.26	-0.29
	p	0.683	0.683	0.354	0.949	0.177	0.252	0.5	0.444
Average daily June temperature	*r*	0.34	0.4	0.739	0.65	0.79	0.68	0.82	0.81
	p	0.378	0.363	0.022*	0.059	0.011*	0.043*	0.007*	0.008*

		Common gardens		Growth chamber					
		Days to anthesis (days)	**Days to fifth anthesis (days)**	**Average capitulum diameter (mm)**	**Days to anthesis (days)**	**Stem diameter (mm)**	**Diameter of first capitulum (mm)**	**Total nodes**	**Plant height (cm)**
Soil characteristics	**Statistic**								
Soil water holding capacity	*r*	-0.01	-0.01	0.16	0.10	-0.08	0.27	-0.03	0.21
	p	0.98	0.98	0.682	0.799	0.839	0.484	0.939	0.589
Soil pH	*r*	0.04	0.12	0.15	0.45	0.29	0.05	0.62	0.19
	p	0.911	0.754	0.703	0.229	0.452	0.893	0.074	0.62
Percent organic matter	*r*	0.13	-0.07	-0.12	-0.2	-0.03	-0.05	-0.33	-0.02
	p	0.732	0.865	0.765	0.7	0.932	0.898	0.381	0.966
N	*r*	0.17	0.04	0.16	0.41	0.12	0.08	0.16	0.34
	p	0.667	0.915	0.683	0.273	0.764	0.847	0.683	0.366
P	*r*	0.43	0.37	0.46	0.01	0.62	0.52	0.11	0.42
	p	0.243	0.333	0.213	0.983	0.074	0.152	0.777	0.264
K	*r*	-0.21	-0.18	0.27	-0.04	-0.03	0.29	-0.13	0.18
	p	0.589	0.651	0.486	0.915	0.949	0.458	0.731	0.651
Ca	*r*	0.18	0.17	0.33	0.48	0.23	0.30	0.25	0.43
	p	0.637	0.668	0.381	0.188	0.546	0.433	0.517	0.244
Mg	*r*	-0.15	<0.001	0.32	-0.28	0.25	0.37	-0.38	<0.001
	p	0.7	1	0.406	0.46	0.517	0.332	0.309	1
Zn	*r*	0.02	-0.17	-0.48	-0.30	-0.18	-0.37	-0.23	-0.37
	p	0.966	0.668	0.188	0.433	0.637	0.332	0.546	0.332
Cl	*r*	0.10	0.22	0.18	-0.35	0.35	0.23	-0.03	-0.07
	p	0.798	0.576	0.637	0.356	0.356	0.546	0.932	0.865
Cu	*r*	0.53	0.75	0.65	0.18	0.53	0.63	0.25	0.50
	p	0.139	0.020*	0.058	0.637	0.139	0.067	0.517	0.171

		Common gardens		Growth chamber					
		Days to anthesis (days)	**Days to fifth anthesis (days)**	**Average capitulum diameter (mm)**	**Days to anthesis (days)**	**Stem diameter (mm)**	**Diameter of first capitulum (mm)**	**Total nodes**	**Plant height (cm)**
B	*r*	0.12	0.07	0.48	0.05	0.63	0.57	0.25	0.42
	p	0.765	0.865	0.188	0.898	0.067	0.112	0.517	0.265
Fe	*r*	-0.12	-0.12	-0.58	-0.50	-0.52	-0.53	-0.47	-0.62
	p	0.765	0.765	0.099	0.171	0.154	0.139	0.205	0.077
Mn	*r*	0.32	0.15	-0.10	-0.25	0.20	-0.03	0.02	-0.02
	p	0.406	0.7	0.798	0.517	0.606	0.932	0.966	0.966
Na	*r*	0.03	0.37	0.58	0.1	0.570	0.53	0.28	0.23
	p	0.932	0.332	0.099	0.798	0.112	0.139	0.46	0.546
Cation exchange capacity	*r*	-0.07	0.03	0.50	0.10	0.40	0.52	<0.001	0.32
	p	0.865	0.932	0.171	0.798	0.286	0.154	1	0.406
$CaCO_3$ equivalent	*r*	-0.05	0.02	-0.02	0.35	<0.001	-0.05	0.48	0.08
	p	0.898	0.966	0.966	0.356	1	0.898	0.188	0.831

Note: * Significant at the 0.05 probability level.

† *r* denotes Spearman's ranked correlation coefficient.

‡ *p* denotes *p*-values.

Supplemental Table 4.6: *Pcadapt* and BayeScan *p*-values and loci exceeding the 99th confidence interval in Lostian for SNPs showing signs of being outliers in two of three outlier tests performed on the filter20 dataset.

Linkage group	Position (BP)	*pcadapt*		BayeScan		Lositan	
1	18,649,099	0.323		0.014	*	1	†
2	160,735,807	1		0.011	*	1	†
2	181,484,751	0.428		0.044	*	0.999	†
3	39,682,355	0.016	*	0.701		1	†
4	12,487,955	0.99		0.036	*	1	†
5	136,497,829	0.002	**	0.807		1	†
5	148,744,834	0.387		0.04	*	1	†
5	152,646,835	0.193		0.004	**	1	†
5	152,646,876	0.007	**	0.012	*	1	†
7	911,263	0	***	0.699		1	†
7	44,670,153	1		0.028	*	1	†
8	4,217,892	0.771		0.03	*	0.999	†
8	120,360,369	0.983		0	***	1	†
9	78,370,880	0.912		0.001	***	1	†
9	119,577,919	0.001	***	0	***	1	†
9	165,067,563	0.004	**	0.001	***	0.999	†
9	165,067,587	0.004	**	0.002	**	1	†
9	204,408,439	0.66		0	***	1	†
9	204,408,453	0.289		0	***	1	†
9	204,408,543	0.205		0	***	1	†
9	204,408,581	0.169		0	***	1	†
9	206,499,835	0.862		0.016	*	0.999	†
9	206,499,851	0.842		0.005	**	1	†
10	86,839,089	0.008	**	0.878		1	†
11	5,185,166	0.016	*	0.592		1	†
11	41,575,005	0.007	**	0.213		1	†
11	122,272,027	0.304		0.009	*	0.999	†
12	183,026,270	0.783		0.021	*	1	†
13	35,154,665	0.035	*	0.006	**	1	†
13	35,154,705	0.035	*	0.001	**	1	†
13	81,101,147	0.723		0	***	1	†
14	114,246,647	0.983		0.003	**	1	†
14	218,818,856	0.001	**	0.377		0.999	†
16	46,246,344	1		0	***	1	†
16	46,246,362	1		0	***	1	†
17	53,277,130	1		0.007	**	1	†
17	53,277,133	1		0.006	**	1	†
17	53,277,164	1		0	***	1	†
17	216,837,251	0	***	0.051		1	†
17	265,653,117	0.996		0.008	**	1	†
HA0_73NS	163,961,183	0.951		0.048	*	0.999	†
HA0_73NS	171,702,369	0.041	*	0	***	1	†
HA0_73NS	343,807,996	0.041	*	0.795		1	†

Note: * Significant at the 0.05 probability level. ** Significant at the 0.01 probability level.*** Significant at the 0.001 probability level. †= Exceeds the 99% confidence interval.

Supplemental Table 4.7: Coefficient of determination (r^2) between expected and observed *p*-value distributions following GLM *Q*, GLM *K*, MLM *K*, MLM PCA + *K* and MLM *Q* + *K* models in TASSEL for phenotypic variables collected in the growth chamber and under common garden conditions in Carman and Winnipeg in 2013 and 2014.

	Model				
	GLM *Q*	GLM PCA	MLM *K*	MLM PCA + *K*	MLM *Q* + *K*
Common gardens					
Days to anthesis	0.9911	0.9947	0.9989	0.9991	0.9989
Days to fifth anthesis	0.9939	0.9978	0.9967	0.997	0.9967
Growth chamber					
Average capitulum diameter	0.9942	0.9986	0.9946	0.9962	0.9949
Days to anthesis	0.9976	0.9962	0.989	0.9954	0.989
Diameter of the first capitulum	0.9972	0.9964	0.9695	0.9671	0.9695
Plant height	0.9981	0.9952	0.9994	0.9989	0.9994
Stem diameter	0.9944	0.9956	0.9976	0.9979	0.9976
Total nodes	0.9989	0.9983	0.9979	0.9985	0.9979

Supplemental Table 4.8: Coefficient of determination (r^2) between expected and observed *p*-value distributions following GLM *Q*, GLM *K*, MLM *K*, MLM PCA + *K* and MLM *Q* + *K* models in TASSEL for environmental variables collected from Environment Canada, WorldClim and soil sampling analyses.

	Model				
	GLM *Q*	GLM PCA	MLM *K*	MLM PCA + *K*	MLM *Q* + *K*
Topography					
Latitude	0.9982	0.9975	0.9985	0.9977	0.9985
Longitude	0.9759	0.9973	0.9726	0.9863	0.9726
Elevation (m)	0.9273	0.9899	0.9663	0.9809	0.9663
Temperature and precipitation					
Corn Heat Units	0.9914	0.9982	0.9692	0.9961	0.9692
Frost free days above 0°C	0.9985	0.9975	0.9844	0.994	0.9844
Mean annual temperature	0.9968	0.9953	0.9968	0.9993	0.9953
Mean precipitation	0.9992	0.9965	0.9992	0.9988	0.9992
Average daily June temperature	0.992	0.9994	0.9933	0.9981	0.9933
Soil characteristics					
Soil water holding capacity	0.9974	0.999	0.9019	0.9201	0.9019
Soil pH	0.9861	0.9974	0.7149	0.7185	0.7149
Percent organic matter	0.9987	0.9989	0.9983	0.997	0.9983
N	0.9989	0.9972	0.9954	0.9965	0.9954
P	0.9988	0.9987	0.9045	0.9181	0.9045
K	0.9993	0.9985	0.9983	0.9981	0.9983
Ca	0.9943	0.9948	0.9964	0.9988	0.9964
Mg	0.998	0.9933	0.9933	0.9992	0.9933
Zn	0.9848	0.9985	0.9984	0.9992	0.9984
Cl	0.9955	0.9956	0.9994	0.9993	0.9994
Cu	0.9984	0.9993	0.9983	0.998	0.9983
B	0.9995	0.9994	0.9993	0.9994	0.9993
Fe	0.9865	0.9987	0.9991	0.9989	0.9991
Mn	0.999	0.9992	0.9954	0.9951	0.9954
Na	0.998	0.9965	0.9549	0.9595	0.9549
Cation exchange capacity	0.9896	0.9939	0.9939	0.9965	0.9939
$CaCO_3$ equivalent	0.9785	0.9827	0.981	0.987	0.981

Supplemental Table 4.9: Description of *H. maximiliani* SNPs of interest identified by outlier analysis within known genes within the *H. annuus* or *A. thaliana* reference genomes.

Linkage group	Position (BP)	*H. annuus* reference gene	*H. annuus* reference description	TAIR10 reference genome	TAIR10 reference description
1	18,649,099	Ha412v1r1_01g004940	IPR011009:Protein kinase-like domain;	-	-
2	181,484,751	Ha412v1r1_02g035950	IPR021720:Malectin;	-	-
5	152,646,835 152,646,876	Ha412v1r1_05g025940	IPR002902:Gnk2-homologous domain;	PDLP8-AT3G60720	Encodes a plasmodesmal protein.
7	911,263	Ha412v1r1_07g000110	IPR016024:Armadillo-type fold;	-	-
8	120,360,369	Ha412v1r1_08g031730	IPR000630:Ribosomal protein S8;	AT5G59850	Ribosomal protein S8 family protein;
9	165,067,563 165,067,587	Ha412v1r1_09g029900	IPR012234:Tyrosine-protein kinase, non-receptor SYK/ZAP-70; IPR013320:Concanavalin A-like lectin/glucanase, subgroup	AT3G26700	Protein kinase superfamily protein
10	86,839,089	Ha412v1r1_10g021310	Uncharacterized protein	MGP1-AT2G21870	Encodes the FAd subunit of mitochondrial F1F0-ATP synthase. Essential for pollen formation.
11	5,185,166	Ha412v1r1_11g002230	IPR001623:DnaJ domain;	AT5G49580	Chaperone DnaJ-domain superfamily protein
12	183,026,270	Ha412v1r1_12g043050	IPR000308:14-3-3 protein; IPR023409:14-3-3 protein,	AT1G26480	14-3-3 protein GF14iota (grf12)

Linkage group	Position (BP)	*H. annuus* reference gene	*H. annuus* reference description	TAIR10 reference genome	TAIR10 reference description
			conserved site; IPR023410:14-3-3 domain;		
14	218,818,856	Ha412v1r1_14g048510	IPR004159:Putative S-adenosyl-L-methionine-dependent methyltransferase;	-	-
HA0_73NS	163,961,183	Ha412v1r1_00g064010	Uncharacterized protein	-	-
HA0_73NS	171,702,369	Ha412v1r1_00g066310	IPR001394:Ubiquitin carboxyl-terminal hydrolases family 2;	-	-

Supplemental Table 4.9 Continued: Description of *H. maximiliani* SNPs of interest identified by outlier analysis within known genes within the *H. annuus* or *A. thaliana* reference genomes.

Linkage group	Position (BP)	GO terms		
		Cellular component	**Molecular function**	**Biological process**
1	18,649,099	-	-	-
2	181,484,751	-	-	-
5	152,646,835 152,646,876	-	-	-
7	911,263	-	GO:0005488 binding	-
8	120,360,369	GO:0005840 ribosome	GO:0003735 structural constituent of ribosome	GO:0006412 translation
9	165,067,563 165,067,587	GO:0005737 cytoplasm	GO:0005524 ATP binding, GO:0004715 non-membrane spanning protein tyrosine	GO:0035556 intracellular signal transduction, GO:0006468 protein phosphorylation

Linkage group	Position (BP)	GO terms		
		Cellular component	Molecular function	Biological process
			kinase activity	
10	86,839,089	-	-	-
11	5,185,166	-	-	-
12	183,026,270	-	GO:0019904 protein domain specific binding	-
14	218,818,856	-	GO:0008168 methyltransferase activity	-
HA0_73NS	163,961,183	-	-	-
HA0_73NS	171,702,369	-	GO:0036459 thiol-dependent ubiquitinyl hydrolase activity	GO:0016579 protein deubiquitination

Supplemental Table 4.10: Proximity of *H. maximiliani* SNPs of interest identified by TASSEL or outlier analysis to know described genes within the *H. annuus* reference genome.

Linkage group	Position (BP)	TASSEL Associations	Closest *annotated H. annuus gene*	Distance from SNP (BP)	Description
2	160,735,807	-	Ha412v1r1_02g030740	7,308	AutoIPR: IPR001133:NADH-ubiquinone oxidoreductase chain 4L/K;
3	39,682,355	-	Ha412v1r1_03g007030	23,810	AutoIPR: IPR015429:Cyclin C/H/T/L;
4	12,487,955	-	Ha412v1r1_04g004310	9,015	AutoIPR: IPR011598:Myc-type, basic helix-loop-helix (bHLH) domain;
5	136,497,829	-	Ha412v1r1_05g023630	51,141	AutoIPR: IPR014001:Helicase, superfamily 1/2, ATP-binding

Linkage group	Position (BP)	TASSEL Associations	Closest *annotated H. annuus gene*	Distance from SNP (BP)	Description
					domain; IPR027417:P-loop containing nucleoside triphosphate hydrolase;
5	148,744,834	-	Ha412v1r1_05g025350	608	AutoIPR: IPR000120:Amidase; IPR021183:N-terminal acetyltransferase A, auxiliary subunit; IPR023631:Amidase signature domain;
7	101,023,330	Diameter of first capitulum (growth chamber)	Ha412v1r1_07g026530	1,358	AutoIPR: IPR011990:Tetratricopeptide-like helical;
8	4,217,892	-	Ha412v1r1_08g001540	1,388	AutoIPR: IPR004314:Domain of unknown function DUF239;
9	78,370,880	-	Ha412v1r1_09g013950	31,016	AutoIPR: IPR002913:START domain; IPR011993:Pleckstrin homology-like domain; IPR023393:START-like domain;
9	119,577,919	-	Ha412v1r1_09g021330	31,163	AutoIPR: IPR013057:Amino acid transporter, transmembrane;
9	204,408,439 204,408,453 204,408,543 204,408,581	-	Ha412v1r1_09g041410	76,019 76,005 75,915 75,877	AutoIPR: IPR003441:NAC domain;
9	206,499,835 206,499,851	Average daily temperature in June	Ha412v1r1_09g042140	2,183 2,199	AutoIPR: IPR008479:Protein of unknown function DUF760;
10	53,871,327	Frost free days	Ha412v1r1_10g013280	17,058	AutoIPR: IPR009136:Vascular

Linkage group	Position (BP)	TASSEL Associations	Closest *annotated H. annuus gene*	Distance from SNP (BP)	Description
		above 0°C			endothelial growth factor receptor 2 (VEGFR2); IPR013320:Concanavalin A-like lectin/glucanase, subgroup; IPR013320:Concanavalin A-like lectin/glucanase, subgroup;
10	218,772,737	Elevation	Ha412v1r1_10g045250	47,731	AutoIPR: IPR013210:Leucine-rich repeat-containing N-terminal, type 2;
11	41,575,005	-	Ha412v1r1_11g012630	57,982	AutoIPR: IPR021168:Bifunctional polymyxin resistance protein, ArnA;
11	188,999,507	Days to anthesis (growth chamber)	Ha412v1r1_11g043390	69,816	AutoIPR: IPR002290:Serine/threonine- / dual specificity protein kinase, catalytic domain; IPR016243:Tyrosine-protein kinase, CSF-1/PDGF receptor;
13	81,101,147	Diameter of first capitulum (growth chamber)	Ha412v1r1_13g014650	5,150	AutoIPR: IPR016040:NAD(P)-binding domain;
13	187,158,373	Days to anthesis (growth chamber)	Ha412v1r1_13g040450	45,563	AutoIPR: IPR001353:Proteasome, subunit alpha/beta;
13	35,154,665 35,154,705	-	Ha412v1r1_13g006230	12,017 12,057	AutoIPR: IPR000767:Disease resistance protein; IPR027417:P-

Linkage group	Position (BP)	TASSEL Associations	Closest *annotated H. annuus gene*	Distance from SNP (BP)	Description
					loop containing nucleoside triphosphate hydrolase;
14	114,246,647	-	Ha412v1r1_14g019070	74,161	AutoIPR: IPR011598:Myc-type, basic helix-loop-helix (bHLH) domain;
16	220,344,187	CaCO3 equivalent (%)	Ha412v1r1_16g053670	5,470	AutoIPR: IPR005835:Nucleotidyl transferase;
16	46,246,344 46,246,362	-	Ha412v1r1_16g011170	1,532 1,550	AutoIPR: IPR001683:Phox homologous domain;
17	216,837,251	Frost free days above 0°C	Ha412v1r1_17g044000	56,153	AutoIPR: IPR021151:GINS complex;
17	265,653,117	Elevation	Ha412v1r1_17g056740	206	AutoIPR: IPR001810:F-box domain;
17	53,277,130 53,277,133 53,277,164	-	Ha412v1r1_17g014690	377 380 411	AutoIPR: IPR002156:Ribonuclease H domain; IPR009027:Ribosomal protein L9/RNase H1, N-terminal; IPR017290:Ribonuclease H, bacteria
HA0_73NS	343,807,996	-	Ha412v1r1_00g122810	7,655	AutoIPR: IPR016035:Acyl transferase/acyl hydrolase/lysophospholipase;

Supplemental Table 4.10 Continued: Proximity of *H. maximiliani* SNPs of interest identified by TASSEL or outlier analysis to know described genes within the *H. annuus* reference genome.

Linkage group	Position (BP)	Cellular component	Molecular function	Biological process
2	160,735,807	-	GO:0016651 oxidoreductase activity, acting on NAD(P)H	GO:0042773 ATP synthesis coupled electron transport, GO:0055114 oxidation-reduction process
3	39,682,355	-	GO:0019901 protein kinase binding	GO:0000079 regulation of cyclin-dependent protein serine/threonine kinase activity, GO:0006355 regulation of transcription, DNA-templated
4	12,487,955	-	GO:0046983 protein dimerization activity	-
5	136,497,829	-	-	-
5	148,744,834	-	GO:0004040 amidase activity	-
7	101,023,330	-	GO:0005515 protein binding	-
8	4,217,892	-	-	-
9	78,370,880	-	GO:0008289 lipid binding	-
9	119,577,919	-	-	-
9	204,408,439 204,408,453 204,408,543 204,408,581	-	GO:0003677 DNA binding	GO:0006355 regulation of transcription, DNA-templated
9	206,499,835 206,499,851	-	-	-
10	53,871,327	GO:0016021 integral component of	GO:0005524 ATP binding, GO:0019838 growth factor binding, GO:0005021 vascular	GO:0006468 protein phosphorylation, GO:0048010 vascular endothelial growth factor

Linkage group	Position (BP)	Cellular component	Molecular function	Biological process
		membrane	endothelial growth factor-activated receptor activity	receptor signalling pathway
10	218,772,737	-	-	-
11	41,575,005	-	GO:0016742 hydroxymethyl-, formyl- and related transferase activity	GO:0055114 oxidation-reduction process
11	188,999,507	-	GO:0004714 transmembrane receptor protein tyrosine kinase activity	GO:0007169 transmembrane receptor protein tyrosine kinase signalling pathway
13	81,101,147	-	-	-
13	187,158,373	GO:0005839 proteasome core complex	GO:0004298 threonine-type endopeptidase activity	GO:0051603 proteolysis involved in cellular protein catabolic process
13	35,154,665 35,154,705	-	-	-
14	114,246,647	-	GO:0046983 protein dimerization activity	-
16	220,344,187	-	GO:0016779 nucleotidyltransferase activity	GO:0009058 biosynthetic process
16	46,246,344 46,246,362	-	GO:0035091 phosphatidylinositol binding	-
17	216,837,251	-	-	-
17	265,653,117	-	GO:0005515 protein binding	-
17	53,277,130 53,277,133 53,277,164	-	GO:0004523 RNA-DNA hybrid ribonuclease activity, GO:0003676 nucleic acid binding	-
HA0_73NS	343,807,996	-	-	GO:0008152 metabolic process

Supplemental Table 4.11: BLASTx results to the UniProKB database of *H. maximiliani* SNPs with no known gene in the *H. annuus* or *A. thaliana* reference genomes.

Linkage group	Position (BP)	TASSEL Associations	BLASTx-Hit (UniProtKB database)	Gene Identifier	Alignment length	E-value	Similarity %
2	160,735,807	-	-	-	-	-	-
3	39,682,355	-	-	-	-	-	-
4	12,487,955	-	-	-	-	-	-
5	136,497,829	-	-	-	-	-	-
5	148,744,834	-	-	-	-	-	-
7	101,023,330	Diameter of first capitulum (growth chamber)	-	-	-	-	-
8	4,217,892	-	PREDICTED: uncharacterized protein LOC104712604 isoform X2 [*Camelina sativa*]	1,109,160,880	58	3.21E-12	69.00%
9	78,370,880	-	-	-	-	-	-
9	119,577,919	-	PREDICTED: uncharacterized protein LOC107766570 [*Nicotiana tabacum*]	1,025,415,770	66	7.65E-27	83.30%
9	204,408,439	-	PREDICTED: uncharacterized mitochondrial protein AtMg00810-like [*Oryza sativa* Japonica Group]	1,002,303,726	36	4.84E-10	75.00%
9	204,408,453	-	PREDICTED: uncharacterized protein LOC105353428 [*Fragaria vesca* subsp. vesca]	764,641,197	26	3.24E-09	92.30%

Linkage group	Position (BP)	TASSEL Associations	BLASTx-Hit (UniProtKB database)	Gene Identifier	Alignment length	E-value	Similarity %
9	204,408,543	-	PREDICTED: uncharacterized mitochondrial protein AtMg00810-like [*Citrus sinensis*]	985,462,561	53	6.44E-12	66.00%
9	204,408,581	-	-	-	-	-	-
9	206,499,835	Average daily temperature in June	-	-	-	-	-
9	206,499,851	Average daily temperature in June	-	-	-	-	-
10	53,871,327	Frost free days above 0°C	-	-	-	-	-
10	218,772,737	Elevation	-	-	-	-	-
11	41,575,005	-	PREDICTED: uncharacterized protein LOC109163743 [*Ipomoea nil*]	1,109,287,040	66	2.47E-11	63.60%
11	188,999,507	Days to anthesis (growth chamber)	-	-	-	-	-
13	35,154,665	-	-	-	-	-	-
13	35,154,705	-	-	-	-	-	-
13	81,101,147	Diameter of first capitulum (growth chamber)	PREDICTED: uncharacterized protein LOC109134898 [*Beta vulgaris* subsp. vulgaris]	1,108,916,611	31	1.78E-05	80.60%
13	187,158,373	Days to anthesis (growth chamber)	-	-	-	-	-

Linkage group	Position (BP)	TASSEL Associations	BLASTx-Hit (UniProtKB database)	Gene Identifier	Alignment length	E-value	Similarity %
14	114,246,647	-	-	-	-	-	-
16	46,246,344	-	-	-	-	-	-
16	46,246,362	-	-	-	-	-	-
16	220,344,187	CaCO3 equivalent (%)	-	-	-	-	-
17	53,277,130	-	-	-	-	-	-
17	53,277,133	-	-	-	-	-	-
17	53,277,164	-	-	-	-	-	-
17	216,837,251	Frost free days above 0°C	PREDICTED: uncharacterized protein LOC107633185 [*Arachis ipaensis*]	1,021,498,286	66	5.12E-21	69.70%
17	265,653,117	Elevation	-	-	-	-	-
HA0_73NS	343,807,996	-	PREDICTED: uncharacterized protein LOC108201116 [*Daucus carota* subsp. sativus]	1,040,925,712	67	7.72E-10	65.70%

7.2.3 Appendix for Chapter 5.0

Supplemental Table 5.1: Number of SNPs aligned to the *H. annuus* reference genome in unfiltered, association mapping (SNPAM) and linkage mapping (SNPLG) datasets.

	Dataset†		
Linkage group	Unfiltered	SNPAM	SNPLG
1	1,440	458	252
2	1,878	510	304
3	1,605	517	250
4	1,801	558	261
5	2,337	719	376
6	679	173	95
7	889	258	125
8	1,830	623	324
9	2,015	576	349
10	2,876	811	460
11	1,767	524	290
12	1,514	470	204
13	2,184	604	274
14	1,568	464	255
15	1,684	532	291
16	1,654	537	233
17	2,507	766	412
Ha0_73Ns‡	3,378	1,044	-
Total	33,606	10,144	4,755

†Note: The SNPAM dataset contains SNPs with no greater than 50% missing data and a minor allele state of >5, SNPLG datasets contains SNPs which could be classified into D1.10 and D2.15 marker categories and does not contain Ha0_73Ns SNPs.
‡Unordered *H. annuus* contig.

Supplemental Table 5.2: Phenotypic characteristics of the Manitoban parental population employed to generate bi-parental mapping populations relative to F_2 populations examined under growth chamber conditions.

Trait	Manitoba parent population*			F_2 Population		
	Mean	**Range**	**SD**	**Mean**	**Range**	**SD**
Diameter of capitulum 1-5 (mm)	10.47	4.92	1.32	12.6	14.18	2.54
Size of central capitulum (mm)	11.95	6.2	1.6	13.89	19.4	3.38
Total capitulum count	13.8	20	1.86	39.64	122	22.36
Lowest branching node (%)	48.25	55.16	13.95	36.28	68	11.01
Percentage of branch bearing nodes (%)	48.25	55.16	13.92	16.95	48.97	6.55
Total nodes	18.6	11	3.13	50.96	67	10.74
Stem diameter (mm)	6.67	3.7	3.13	9.53	8.5	1.31
Plant height (cm)	93.76	52	14.31	142.11	102	60.61
First anthesis (days)	76.96	26	7.37	118.37	137	24.69

Note: * Values derived from a sample of 25 individuals collected from Oak Bluff, Manitoba in 2012 described in Chapter 4. Full phenotypic information on the TLI Kansas parental population is not available due to the use of photoperiod induction to induce premature flowering.

Supplemental Table 5.3: Loadings of the first five principal components extracted from 20 traits measured on 341 F_2 *H. maximiliani* plants grown under growth chamber conditions.

Trait	PC1	PC2	PC3	PC4	PC5
Average anthesis	0.237	-0.403	-0.099	-0.008	0.023
Average capitulum size	0.333	0.281	0.020	0.135	-0.051
Branch length	0.126	0.215	-0.211	0.170	-0.148
Capitulum depth:width ratio	-0.223	-0.213	-0.040	0.022	-0.062
Central capitulum depth	0.283	0.295	0.001	0.124	-0.075
Central capitulum size	0.311	0.324	0.038	0.073	-0.020
First anthesis	0.223	-0.406	-0.002	0.023	-0.024
Highest branching node	0.049	0.124	-0.647	-0.277	0.003
Leaf length	-0.156	0.215	0.517	0.039	-0.042
Lowest branching node	0.153	-0.123	0.269	0.116	0.301
Percent branched	-0.292	0.117	0.166	0.040	-0.337
Petal colour	0.167	-0.070	0.211	-0.387	-0.322
Petal length	0.274	0.054	0.045	-0.304	-0.225
Petal morphology	-0.058	0.037	-0.086	-0.019	-0.423
Petal whorling	0.130	-0.097	0.202	-0.326	-0.393
Plant height	0.259	0.080	0.151	0.262	0.092
Reproductive budding	0.319	-0.202	0.060	-0.125	0.096
Stem diameter	0.121	-0.057	-0.170	0.501	-0.397
Total capitulum count	-0.195	-0.252	-0.035	0.323	-0.294
Total nodes	0.247	-0.295	0.077	0.222	-0.098
Variance explained (%)	26.77%	16.83%	8.25%	7.37%	6.29%
Cumulative variance explained (%)	26.77%	43.61%	51.86%	59.19%	65.49%

Supplemental Table 5.4: QTL supported by a LOD score <3 and surpass an alpha threshold of 0.05 at the chromosome level in the D1.10 map *H. maximiliani* genetic map.

Trait	QTL name	Nearest Marker	Linkage group	Peak (cM) †	CI (cM) ‡	%Var. §	LOD
Branch length	*BL4*	Ha11_107467238	11	1.6	0.0-21.0	5.99	2.71
Capitulum depth:width ratio	*CWD3*	Ha11_108059224	11	272.0	270.0-273.0	6.21	2.65
Capitulum depth:width ratio	*CWD4*	Ha1_20875749	1	122.0	118.0-124.0	6.22	2.65
Central capitulum depth	*CCD4*	Ha6_80026224	6	50.0	49.0-52.0	5.14	2.27
Petal whorling	*PW1*	Ha7_78392672	7	34.1	21.0-52.0	5.71	2.43
Petal whorling	*PW2*	Ha10_115598916	10	15.6	14.53-20.0	5.40	2.39
Reproductive budding	*RB2*	Ha16_215897889	16	110.0	107.26-122.0	5.82	2.57
Total capitula count	*TCC5*	Ha6_35231733	6	53.3	42.0-58.0	6.87	2.94
Total capitula count	*TCC6*	Ha13_121481818	13	188.0	168.0-207.0	6.22	2.65

Note: † cM =Centimorgan

‡CI = LOD-1 confidence interval

§ %Var. =proportion of phenotypic variance explained by QTL (equivalent to marker r^2 x 100)

Supplemental Table 5.5: QTL supported by a LOD score <3 and surpass an alpha threshold of 0.05 at the chromosome level in the D2.15 H. *maximiliani* genetic map.

Trait	QTL name	Nearest Marker	Linkage group	Peak (cM) †	CI (cM) ‡	%Var. §	LOD
Average anthesis	*AA1*	Ha14_131232381	14	85.0	81.0-91.0	5.74	2.44
Central capitulum depth	*CCD3*	Ha2_54803449	2	72.0	71.0-85.0	6.17	2.63
First anthesis	*FA1*	Ha14_131232381	14	85.0	80.26-81.0	6.72	2.87
Highest branching node	*HBN4*	Ha7_25287910	7	74.0	67.0-75.0	6.25	2.66
Highest branching node	*HBN5*	Ha12_2410843	12	199.0	175.0-203.0	5.61	2.38
Petal morphology	*PM1*	Ha15_122250909	15	40.3	36.0-43.0	6.74	2.9
Stem diameter	*SD3*	Ha14_199552757	14	138.0	136.0-143.0	5.65	2.4
Total nodes	*TN3*	Ha11_44990227	11	108.0	72.0-113.0	6.85	2.97

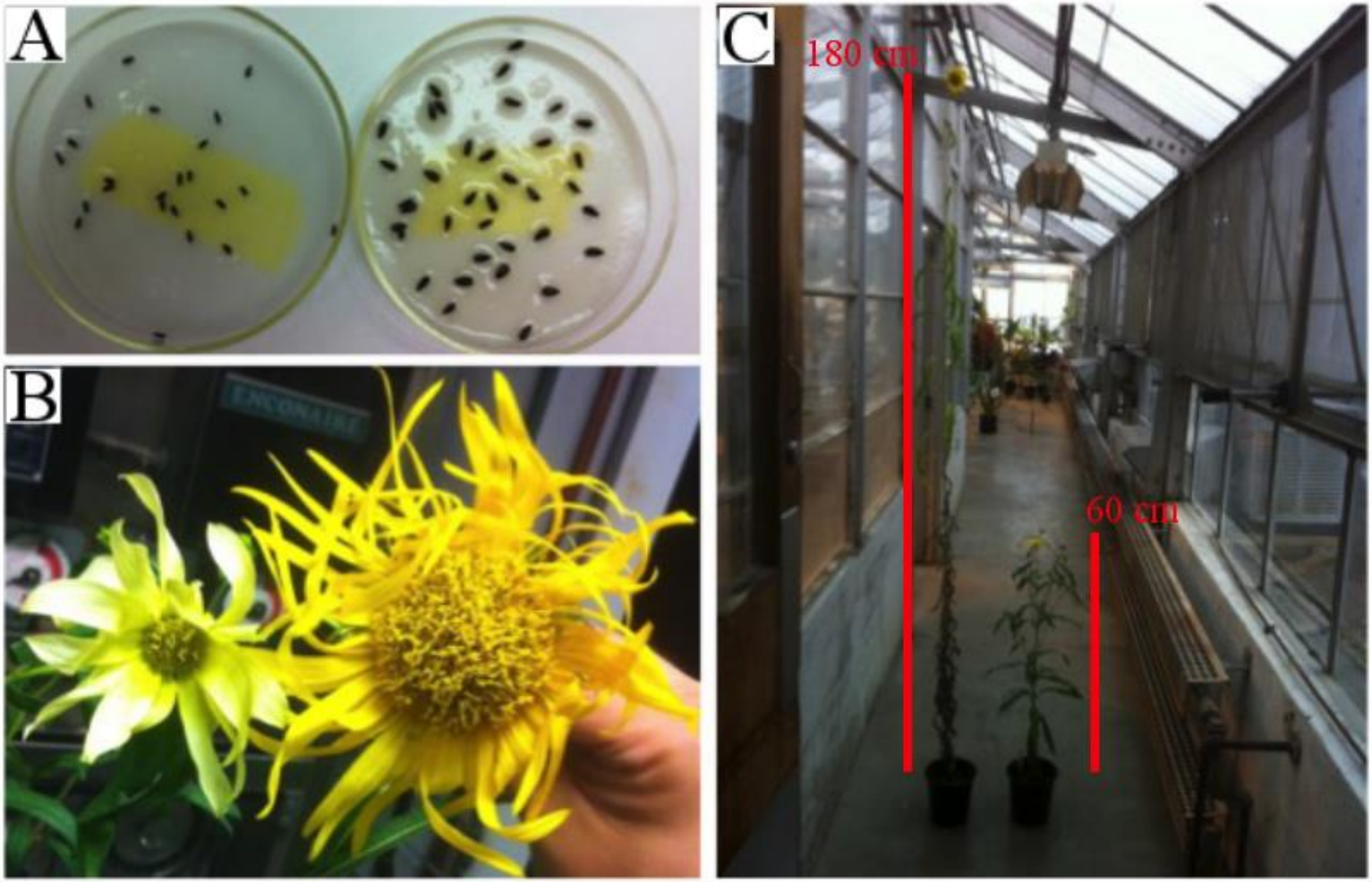
Note: † cM =Centimorgan

‡CI = LOD-1 confidence interval

§ %Var. =proportion of phenotypic variance explained by QTL (equivalent to marker r^2 x 100)

Supplemental Table 5.6: Linkage groups in which QTL and GLM or MLM associations have been detected for domestication syndrome type traits in *H. maximiliani* which correspond to previously established studies in *H. annuus.*

Trait	*H. maximiliani* linkage group			Analogous trait from *H. annuus* literature	Reference
	QTL	GLM	MLM		
Size of central capitulum			5	Capitulum diameter	Burke et al., 2002
			13	Capitulum diameter	Burke et al., 2002; Mandel et al., 2013b
Highest branching node	1	1		Basal branching	Nambeesan et al., 2015
		5	5	Mid-Basal branching	Nambeesan et al., 2015
	8			Basal/mid-basal branching	Nambeesan et al., 2015
		15		Apical branching	Nambeesan et al., 2015
Lowest branching node		4	4	Basal/mid-basal branching	Nambeesan et al., 2015
	6			Apical/mid-basal branching	Nambeesan et al., 2015
		7		Number of branches	Chapman et al., 2008b
		8		Basal/mid-basal branching	Nambeesan et al., 2015
			9	Mid-Basal branching	Nambeesan et al., 2015
			15	Apical branching	Nambeesan et al., 2015
Percentage of branch bearing nodes	1	1	1	Basal branching	Nambeesan et al., 2015
	8		8	Basal/mid-basal branching	Nambeesan et al., 2015
	9			Branch number; apical/mid-Basal branching	Dechaine et al., 2009;Nambeesan et al., 2015
		13	13	Number of branches, apical/mid-apical branching	Burke et al., 2002;Wills and Burke, 2007;Nambeesan et al., 2015
		15	15	Apical branching	Nambeesan et al., 2015
Stem diameter	11			Stem diameter	Burke et al., 2002
Leaf length			4	Leaf size	Burke et al., 2002
			5	Leaf size	Burke et al., 2002
First/Average anthesis		14	14	Days to flower	Baack et al., 2008
		15		Days to flower	Wills and Burke, 2007
Capitulum depth:width ratio		17	17	Shattering	Burke et al., 2002

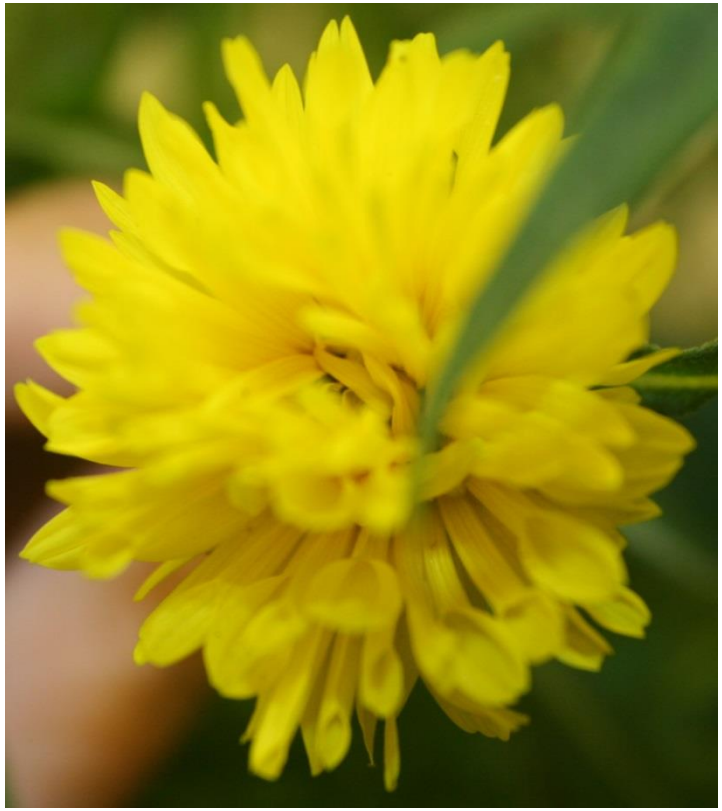


Supplemental Figure 5.1: Phenotypic characteristics of wild-harvested Manitoban and improved TLI Kansas parental populations employed to generate bi-parental mapping populations. A) Left: wild-type seed size of Manitoba parental population. Right: TLI parental population exhibiting increased seed size. B) Left: Manitoba parent exhibiting small (~11mm) capitulum size and lemon coloured ray florets. Right: TLI parent exhibiting increased capitulum size (~40mm) and golden

ray florets. C) Right: TLI parent exhibiting unbranched plant architecture and a single, large central capitulum and increased stature. Left: Manitoba parent exhibiting branched plant architecture with multiple small capitula present and reduced plant height (image: S.R. Asselin).

Supplemental Figure 5.2: Variation in petal whorling observed segregating in 190 F_2 Maximilian sunflower plants grown under growth chamber conditions. Left: Normal petal whorling presentation, middle: weakly doubled whorling presentation, right: doubled/chrysanthemum-type petal presentation (image S.R. Asselin).

Supplemental Figure 5.3: Variation in petal morphology observed segregating in 190 F_2 Maximilian sunflower plants grown under growth chamber conditions. Left: an individual exhibiting normal petal morphology presentation, right: an individual exhibiting tubular ray florets (image S.R. Asselin).



Supplemental Figure 5.4: Example of an F_2 individual grown under growth chamber conditions exhibiting both doubled/chrysanthemum-like phenotype and tubular ray florets (image: S.R. Asselin).