

**THE ROLE OF ADAPTIVE EVOLUTION IN THE SUCCESS OF
AN AGRICULTURAL WEED, *HELIANTHUS ANNUUS***

by

Emily Barbara McKenzie Drummond

B.Sc., The University of Toronto, 2006

M.Sc., The University of British Columbia, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2018

© Emily Barbara McKenzie Drummond, 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

The Role of Adaptive Evolution in the Success of an Agricultural Weed, Helianthus annuus

submitted by Emily Barbara McKenzie Drummond in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

in Botany

Examining Committee:

Loren Rieseberg, Botany
Supervisor

Sarah Otto, Zoology
Supervisory Committee Member

Mark Vellend, Botany
Supervisory Committee Member

Keith Adams, Botany
University Examiner

Eric Taylor, Zoology
University Examiner

Additional Supervisory Committee Members:

Quentin Cronk, Botany
Supervisory Committee Member

Abstract

If a weed is defined as a plant that is “growing where it is not wanted”, then agricultural weeds, or plants that invade and persist in cultivated fields, might be the epitome of weeds. Agricultural weeds have arisen repeatedly from wild plant species, often undergoing rapid evolution to escape eradication. While agricultural weeds thus represent an attractive opportunity to study evolutionary processes operating over short timescales, the genetic basis of local adaptation and, in cases of multiple independent weed origins, the factors influencing parallel evolution, they remain understudied.

For my thesis work, I asked whether populations of common sunflower (*Helianthus annuus*) growing as agricultural weeds have adapted to the unique challenges posed by cultivated fields. In a common garden, I compared paired weedy and wild (i.e., non-agricultural) populations, collected over a latitudinal transect from Canada to Kansas, USA. Weedy populations grew faster and flowered earlier than wild populations, suggesting an evolutionary shift in life history strategy to prioritize growth and reproduction. One wild population from a wetland site showed the same pattern, indicating that wild sunflowers may face similar selection pressures in certain contexts.

I then used whole genome resequencing to investigate the extent of parallel genetic differentiation between weedy and wild populations. Using two different metrics, a “cluster separation score” based on genetic distance matrices and F_{ST} , I identified a list of 148 differentiated genomic regions, though our analysis lacked power to distinguish true positives after correction for multiple testing, and therefore these regions are only suggestively linked to adaptation to the agricultural environment. Genes overlapping these regions were varied and included those involved in plant stress responses, flowering time genes and transporter genes linked to herbicide resistance.

To connect phenotype to genotype, I conducted a genome-wide association analysis of glyphosate resistance, a trait likely critical for the success of weedy populations. At a glyphosate application rate of 0.5 kg a.e. ha⁻¹, or half the rate typically applied by a farmer, resistance segregated in the mapping population, with surviving plants (78.5%) showing a variety of symptoms. Mapping identified 68 SNPs suggestively associated with resistance, and three transporter proteins, among other genes.

Lay Summary

Since the inception of agriculture, wild plant species have been invading cultivated fields, competing with crops and decreasing yields. These agricultural weeds cost the economy \$33 billion annually in the USA alone, and further research is needed in order to develop better management plans. One way weeds may become more successful over time, and better able to escape eradication, is via evolution. Evolution describes the process whereby in each successive generation, the genes of plants that leave more descendants become dominant in the population. Weeds may evolve to have life cycles that better match those of short-season crops, for example, or to be resistant to herbicides such as Roundup. In my dissertation, I discovered that evolution has helped common sunflower succeed as an agricultural weed, with weeds showing faster growth, earlier flowering and Roundup resistance. Using modern sequencing technology, I identified DNA regions that may underlie these changes.

Preface

Work for Chapter 2 was conducted in collaboration with Loren Rieseberg. Together, we conceived the idea for the experiment, which I then executed and analyzed with his feedback. I wrote the manuscript, with input from Loren Rieseberg.

The whole genome resequencing data that I used in Chapter 3 ($n = 16$ individuals) was part of a larger dataset ($n = 321$ individuals) that I then used in Chapter 4. This dataset was created by a team of collaborators in Loren Rieseberg's lab. To create this dataset, I first grew the plants, harvested leaf tissue and extracted DNA. Marco Todesco then prepared the sequencing-ready libraries for generation of the whole genome sequencing data. The raw reads were then cleaned and aligned to the *Helianthus annuus* reference genome by Sarel Hubner; to produce the final SNP dataset, Greg Owens then called variants and performed further filtering of the dataset. Using a subset of these SNP data, I then conducted the analyses, as conceived by myself and Loren Rieseberg, who also provided feedback on the manuscript that I wrote.

The idea and experimental design for Chapter 4 was developed by Loren Rieseberg and myself. As described for Chapter 3, the SNP dataset used in Chapter 4 was generated through the work of myself, Marco Todesco, Sarel Hubner and Greg Owens. I additionally performed all the greenhouse trials of glyphosate resistance, and assessed plant performance. Greg Owens performed the statistical analysis for the association mapping, with feedback from myself, Marco Todesco and Loren Rieseberg. I then finalized the list of candidate regions and identified overlapping genes of interest. The manuscript was written by myself, with the assistance of Loren Rieseberg.

Table of Contents

Abstract	iii
Lay Summary	iv
Preface	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
List of Abbreviations	xi
Acknowledgements	xii
Dedication	xiv
Chapter 1 : Introduction	1
1.1 General Introduction.....	1
1.2 Research Objectives.....	3
1.3 Introduction to the Study System	3
1.4 Prior Research in This System	5
1.5 Summary of Studies	7
Chapter 2 : Common Garden Experiment Examining Life History Differences in Weedy Versus Wild Sunflower Populations	13
2.1 Introduction	13
2.2 Materials and Methods.....	16
2.2.1 Study Populations and Sample Collection	16
2.2.2 Seedling Common Garden and Growth Study	19
2.2.3 Field Common Garden and Flowering Study	21
2.2.4 Maternal Effects Experiment	23
2.2.5 Statistical Analyses of Common Garden Data.....	25
2.3 Results.....	26
2.3.1 Source Populations were Well-Matched for Climate and Phenotypically Diverse.....	26
2.3.2 Weedy Sunflowers Exhibit Faster Growth and Earlier Reproduction	27
2.3.3 Seed Source Influences Seedling Growth but Not Time to Flower.....	29

2.4 Discussion.....	30
2.4.1 Evidence for Faster Growth and Earlier Reproduction in Weedy Sunflowers.....	31
2.4.2 Effects of seed provisioning and seed source.....	34
2.4.3 Conclusions and Future Directions	36
Chapter 3 : Parallel Genomic Signatures of Divergence between Agricultural-Weed and Non-Agricultural, Wild Populations of Sunflowers	51
3.1 Introduction	51
3.2 Materials and Methods.....	53
3.2.1 Study Populations and Sample Collection	53
3.2.2 DNA Extraction and Library Preparation.....	54
3.2.2 Bioinformatics Pipeline	55
3.2.3 Analysis of Linkage Disequilibrium.....	56
3.2.4 Sliding Window Analysis of Weedy versus Wild Divergence	57
3.2.5 Variable-Sized Distinct Window Analysis of Divergence.....	59
3.2.6 Outlier Windows and Candidate Genes.....	60
3.3 Results.....	61
3.3.1 Linkage Disequilibrium Decays Rapidly.....	61
3.3.2 Regions of Genetic Differentiation are Small and Scattered Over the Genome	61
3.3.3 Candidate Adaptive Genes.....	63
3.4 Discussion.....	65
3.4.1 Parallel Adaptation to the Agricultural Environment Proceeds from Standing Variation	65
3.4.2 Multiple Genes of Small Effect Contribute to Phenotypic Differences.....	67
3.4.3 Conclusions	69
Chapter 4 : Genome-Wide Association Analysis of Glyphosate Resistance in Wild Sunflowers	80
4.1 Introduction	80
4.2 Materials and Methods.....	84
4.2.1 Plant Materials.....	84
4.2.2 Determination of Glyphosate Resistance Phenotype.....	85
4.2.3 DNA Extraction and Library Preparation.....	87
4.2.4 Bioinformatics Pipeline	88
4.2.5 Genome-Wide Association Study	89
4.2.6 Significance Testing and Genes of Interest.....	91
4.3 Results.....	92

4.3.1 Glyphosate Treatment Produced a Variety of Phenotypic Effects	92
4.3.2 GWAS Identified SNPs Suggestively Associated with Glyphosate Resistance	94
4.3.3 Genes of Interest for Glyphosate Resistance.....	96
4.4 Discussion.....	98
4.4.1 Glyphosate Resistance in Wild Sunflower, <i>Helianthus annuus</i>	99
4.4.2 Glyphosate Resistance in Sunflower Likely Involves Non-Target-Site Mechanisms.....	101
4.4.3 Conclusions	104
Chapter 5 : Conclusion	116
5.1 Summary	116
5.2 General Conclusions.....	118
5.2.1 Rapid and Repeated Evolution of Weediness in Sunflower.....	118
5.2.2 Widespread Glyphosate Resistance in Sunflower Suggests Multiple Origins	120
5.3 Future Directions	121
5.3.1 Extending the Common Garden Work to Compare Weedy-Wild Fitness Differences	121
5.3.2 Accounting for the Paired Nature of the Weedy-Wild Data in the Genome Scans.....	122
5.3.3 Greater Precision in Mapping Glyphosate Resistance	122
Bibliography	124
Appendices.....	147
Appendix A: Gene List for Genome Scan Results.....	147
Appendix B: Gene List for GWAS Results	150

List of Tables

Table 2.1: Description of collection site details for wild sunflower populations	38
Table 2.2: Mean annual climate for the study populations	39
Table 2.3: Variability among harvested seedlings in height, leaf number and leaf dimensions	40
Table 2.4: Description of the plant materials generated in the 2012 common garden	41
Table 3.1: CSS analysis summary statistics describing the distribution of SNPs per window	71
Table 3.2: F_{ST} analysis summary statistics describing window size (in SNPs and bp)	72
Table 4.1: Overview of plant materials used in the genome-wide association study.....	106

List of Figures

Figure 1.1: Range of <i>Helianthus annuus</i> in Canada and the United States	10
Figure 1.2: Photographs of representative domesticated and wild sunflower	11
Figure 1.3: Photographs of weedy sunflower populations infesting crop fields	12
Figure 2.1: Map of sampling locations for twenty focal sunflower populations.....	42
Figure 2.2: Relationship between seedling biomass and four non-destructive measurements.....	43
Figure 2.3: PCA of annual climate variables for study populations	44
Figure 2.4: Box-plots of plant height and number of inflorescences in study populations.....	45
Figure 2.5: Biomass (untransformed and log-transformed) over time for four representative seedlings	46
Figure 2.6: Growth curves by location and population type (weedy or wild) in 2012	47
Figure 2.7: Days until first flower by location and population type in 2012	48
Figure 2.8: Growth curves by location, population type and seed source in 2013	49
Figure 2.9: Days until first flower by location, population type and seed source in 2013	50
Figure 3.1: Map of sampling locations for populations of sequenced sunflowers.....	73
Figure 3.2: Example PCA figures for positive, neutral and negative values of CSS	74
Figure 3.3: Hypothetical illustration of the cubic spline smoothing method	75
Figure 3.4: Sample curve of linkage disequilibrium decay for chromosome 14.....	76
Figure 3.5: Graph of F_{ST} versus CSS for 10,000 bp windows across the genome	77
Figure 3.6: Genome-wide distribution of CSS and W -Statistic scores	78
Figure 3.7: Location of CSS windows with $p < 0.01$ on chromosomes 1, 7 and 17.....	79
Figure 4.1: Increase in global cases of herbicide resistance over time	107
Figure 4.2: Map of sampling locations for sunflower populations included in the study	108
Figure 4.3: Sample photos of sunflower seedlings before and after glyphosate treatment.....	109
Figure 4.4: Photo collage of glyphosate-damaged sunflower seedlings	110
Figure 4.5: Violin plot of seedling biomass versus herbicide resistance score.....	111
Figure 4.6: First two axes of a PCA of the genetic (SNP) data for the study populations	112
Figure 4.7: Normal QQ plot for the genome-wide association study.....	113
Figure 4.8: Manhattan plot of the genome-wide association study	114
Figure 4.9: Zoomed in Manhattan plots for chromosomes 1, 9 and 16.....	115

List of Abbreviations

bp	Base pairs
CSS	Cluster separation score
EST	Expressed sequence tag
FDR	False discovery rate
GAM	Generalized additive model
GBS	Genotyping-by-sequencing
GLS	Generalized least squares
GWAS	Genome-wide association study
GxE	Genotype-by-environment
LD	Linkage disequilibrium
LRT	Likelihood ratio test
MAF	Minor allele frequency
NGS	Next-generation sequencing
NPGS	National plant germplasm system
NTSR	Non-target-site resistance
PCA	Principal component analysis
QQ	Quantile-quantile
QTL	Quantitative trait loci
RAD	Restriction-site associated DNA
REML	Restricted maximum likelihood
RGR	Relative growth rate
RIL	Recombinant inbred line
SNP	Single nucleotide polymorphisms
SVD	Singular value decomposition
TE	Transposable element
TSR	Target-site resistance
USDA	United States Department of Agriculture
USGS	United States Geological Survey
WGS	Whole genome shotgun

Acknowledgements

My thanks go first and foremost to my supervisor, Dr. Loren Rieseberg, for granting me the opportunity to do my doctoral work in his lab. The experience has been life changing, both personally and professionally. I am so tremendously grateful for the number and diversity of talented scientists in the lab that I have had the benefit of learning from over the years. The ability to hone my wet lab skills, establish a bioinformatics toolkit and think deeply about the topics in evolutionary biology that I am most interested in has been such a gift. Thank-you Loren, for all the time, energy and mentoring you have invested in me!

My research has benefited from the wisdom and guidance of my committee members: Dr. Quentin Cronk, Dr. Sally Otto and Dr. Mark Vellend. I greatly appreciated their constructive feedback on every aspect of my work, from inception to analysis, and their mentoring in other aspects of academic life. Thank-you committee, for your investment in nurturing my growth as a scientist!

This work would not have been possible without the army of colleagues, family, friends, students and volunteers who spent countless hours in the lab, field and greenhouse with me. No matter the weather, whitefly abundance in the greenhouse, or lab disaster, this crew worked tirelessly to help me achieve my goals. My talented assistants included (in alphabetical order): Emily Beeson, Gwylim Blackburn, Nadia Chaidir, Anna Crofts, Cindy Dai, Teale Dunsford, Julie de Vriendt, Cassandra Konecny, David Liou, Jessica Liu, Maria Luyten, Kaehn Mok Rong, and Ada Roman. Friends and colleagues who helped me at various stages of the work include: Eric Baack, Greg Baute, Anne Bjorkman, Dylan Burge, Céline Caseys, Winnie Cheung, Chris Grassa, Kay Hodgins, Heather Kharouba, Matt King, Danika Kleiber, Maya Mayrose, Jenny McCune, Evan Morien, Brook Moyers, Thuy Nguyen, Kristin Nurkowski, Kate Ostevik, Greg Owens, Sébastien Renaut, Evan Staton, Megan Stewart and Marco Todesco. I am lucky to have you all!

I am indebted to the competent staff at various locations who facilitated my work. In the greenhouse, David Kaplan and Melina Biron were both wizards of horticulture, who also were kind enough to allow me to perform herbicide treatments on the premises; also Meylin Zink Yi at the Botany greenhouse. At the UBC Farm, I thank Véronik Campbell, Tim Carter and crew for their hard work. And Totem Field would not have been the same without the oversight of Seane Trehearne.

For funding, I thank NSERC and the Faculty of Science at the University of British Columbia. Thanks also to my colleagues in the Botany Department and Biodiversity Research Centre, who have

supported me, and provided an intellectually stimulating and nurturing environment in which to do my doctoral work.

Last but not least, thank-you to my family for putting up with me (and my absences) these past few years! Thank-you for your love, encouragement and the vacations you've given up to help me instead! It's been an adventure and I certainly could not have done it without your support. There are no words to express my gratitude, especially to my husband Daniel... who virtually deserves his own doctorate at this point. Love to you all!

Dedication

In loving memory of my grandmother, Barbara Ruth McKenzie (1928-2016) – a staunch supporter of women’s education, who attained a Bachelor’s degree herself at the age of 54.

Chapter 1: Introduction

1.1 General Introduction

We live in a world of introduced species. Sometimes intentionally, for example in the case of domesticated species and garden ornamentals, often accidentally, as in the case of seed contaminants and other stowaway species, humans have facilitated the movement of species around the globe, greatly accelerating the rate of biotic exchange (Vitousek et al. 1997). Often translocated individuals will fail to establish, but sometimes self-sustaining populations occur, and a small fraction of introduced species will even go on to become invasive (Williamson 1996; Jeschke and Strayer 2006). Invasive species are those that exhibit a high rate of spread, typically overwhelming and displacing existing native species in the habitats they take over. While invaders are typically thought of exclusively as introduced, non-native species, the continued and drastic alteration of natural habitats by humankind has led to “invasive” behaviour on the part of some native species (Mooney and Hobbs 2000). Examples include the current population explosion in white-tailed deer (*Odocoileus virginianus* Zimmerman, 1780) caused by decimation of their native predators, making this a nuisance species (Côté et al. 2004), and the proliferation and expansion of native plants such as Canada fleabane (*Conyza canadensis* (L.) Cronq.; Dauer et al. 2007) and common ragweed (*Ambrosia artemisiifolia* L.; MacKay and Kotanen 2008) in disturbed areas. While many species are experiencing a decline in numbers owing to human activities, a select few are benefitting from the transition to a more human-dominated landscape.

Invasive species have significant ecological costs for impacted native communities (reviewed in Elton 1958 and Sakai et al. 2001), and are considered to be a major driver of global change and the accelerated rate of species extinction (Sala et al. 2000). In addition to direct detrimental effects on the native biota, invasive species may also influence key ecosystem functions and processes via alteration of the physical environment (e.g. soil properties, disturbance regimes; Vitousek et al. 1997), sometimes with direct costs to humans (e.g. via altered fire regimes; D’Antonio and Vitousek 1992). Similarly, weedy species, where weeds are broadly defined as plants “growing where they are not desired” or “plants out of place” (Monaco et al. 2002), and more specifically as plant pests that interfere with human activities (Ellstrand et al. 2010), may have direct costs in the form of reduced agricultural yields. The economic costs of biological invasions, whether of cultivated fields or natural ecosystems, are staggering, with billions of dollars spent annually in control efforts and in lost agricultural productivity (e.g. ~\$33 billion per year in North America; Pimentel et al. 2005).

Understanding how invasive species and weeds arise and what traits mediate their success may be critical for mitigating their effects. One pertinent question is whether invaders arrive “pre-adapted” for their new habitats (or at least with enough phenotypic plasticity to enable survival), or if adaptive evolution might play a role in enabling success in novel environments. As invaders face new biotic and abiotic conditions, whether in a site outside of the native range or a novel, local environment such as a cultivated field, there is the potential for mismatch between phenotypic traits and the environment. Such a mismatch should lead to a host of novel selective pressures. Indeed, a number of recent studies have revealed trait evolution after introduction, such as the evolution of latitudinal clines in key morphological and life-history traits in St. John’s wort (*Hypericum perforatum* L.) in its invaded range (Maron et al. 2004), or the increased susceptibility of introduced populations of bladder campion (*Silene latifolia* Poir.) to herbivores and fungal pathogens (Blair and Wolfe 2004; Wolfe et al. 2004). This has prompted the assertion that rapid evolutionary changes during invasion may be common (Prentis et al. 2008; Whitney and Gabler 2008; Buswell et al. 2011), though the importance of such changes to invasion success remains under debate. What is clear is that many invasive species and weeds provide a remarkable opportunity to study evolution in action and the process of local adaptation.

Agricultural weeds in particular may often serve as interesting case studies (Vigueira et al. 2013). Not only can weeds show startlingly rapid evolutionary change, responding to many crop management practices that are recent in origin, but weeds may also present exciting cases of parallel evolution, where multiple populations or species show similar phenotypic changes in adapting to the cultivated environment. Weeds growing in crop fields are exposed to a variety of human interventions, including the use of chemicals and irrigation, as well as regular disturbance from cropping techniques such as cultivation, harvesting and ploughing. Such disturbances may impose strong selection on weed populations (Barrett 1988), as seen, for example, in the recent evolution of herbicide resistance in many species (Heap 2014). In order to evade removal, weeds may evolve highly specific adaptations; crop mimicry, where weeds come to physically resemble the crop species they infest, provides one such example. More generally, adaptation to herbicides and the timing of disturbances may be crucial for survival (e.g. Tranel and Horvath 2009; Vigueira et al. 2013; Kuester et al. 2016). Annual weeds and weeds of short-season crops may experience selection for rapid development and precocious reproduction (Barrett 1983), in order to maximize fitness prior to harvest. As agricultural fields tend to represent resource-rich environments (Mohler 2001), weeds may come to prosper by evolving to maximize their growth rates. Faster growth could allow weeds to take advantage of transient favorable

conditions before disturbance occurs; however, enhanced growth may come at the expense of stress tolerance traits (Mayrose et al. 2011), if trade-offs in growth and tolerance traits exist.

1.2 Research Objectives

For my dissertation, I study a widespread agricultural weed, *Helianthus annuus* L. or common sunflower, to look for adaptation to cultivated environments and to characterize the genetic architecture of weedy traits. Are weedy populations of this naturally disturbance-adapted species pre-wired for success in agricultural settings, or, despite ongoing gene flow with local wild populations of *H. annuus* (growing in more natural habitats), has adaptive evolution been important for their success? In Chapter 2 of my thesis, I use common garden experiments to look for genetically-based changes in sunflower growth and phenology in weedy populations. If similar changes are seen across multiple weedy populations, this provides strong evidence for a role for natural selection, as stochastic processes such as drift would be unlikely to result in such concerted change. The genetic basis of weediness, or what makes a weed a weed, generally remains poorly understood (Basu et al. 2004; Stewart et al. 2009), making this a critical issue for weed research. Hence, in Chapter 3, I look for differences between paired weedy and wild sunflower populations at the molecular level. Using whole genome shotgun resequencing data, I seek to identify genetically differentiated regions in the weed genome, and ask if the same changes have occurred in parallel across weedy populations. Finally, in Chapter 4, I connect phenotype to genotype, implementing genome-wide association mapping on a dense SNP dataset, in order to uncover the genetic basis of glyphosate resistance in sunflower.

1.3 Introduction to the Study System

The common sunflower, *Helianthus annuus* L., is a member of the family Asteraceae, the largest family of named flowering plants (Funk et al. 2005). This diverse family is most easily recognized by their composite inflorescences, which are often mistaken for single large flowers, but are actually composed of many individual ray and/or disk florets that cluster together, forming a capitulum at the top of the stem. The Asteraceae has a worldwide distribution and includes numerous economically important species (Kesseli and Michelmore 1997; Dempewolf et al. 2008). In the family are food plants, such as lettuce (*Lactuca sativa* L.), globe artichoke (*Cynara cardunculus* L.), endive and chicory (*Cichorium* spp.), as well as medicinal species and ornamentals (e.g. chrysanthemums, cosmos, dahlias, gerberas and marigolds). Many species are also problematic weeds (Hodgins et al. 2015), with over 100 species on the U.S. Federal and State noxious weed lists (USDA, NRCS 2017), for example.

Helianthus annuus is an annual, outcrossing, diploid ($n = 17$ chromosomes) native to North America, where it is but one member of a diverse genus of wild sunflowers, containing some 52 species (Schilling 2006; Stebbins et al. 2013). The genus contains both annual and perennial species with varied ecologies and habitats (Kane et al. 2013), and includes species occupying extreme environments such as sand dunes (e.g. *H. anomalus* Blake and *H. neglectus* Heiser), desert floors (*H. deserticola* Heiser), serpentine soils (*H. exilis* A. Gray) and salt marshes (*H. paradoxus* Heiser). In evolutionary biology, the *Helianthus* genus has long-served as a model system for the study of adaptive introgression and hybrid speciation (Rieseberg et al. 1995), as a result of the propensity for gene flow between sympatric species (Heiser et al. 1969). Common sunflower (*H. annuus*) also has considerable economic value, owing to its domestication as an oil-seed crop, with over 25 million hectares grown worldwide in 2014 (FAOSTAT 2014).

Originally domesticated in eastern North America over 4,000 years ago (Harter et al. 2004), sunflower has been spread around the globe for use in agriculture and as an ornamental. Compared to the wild progenitor, domesticated sunflower has a higher seed oil content and altered flowering time, and has lost traits such as branching, self-incompatibility, seed dormancy and seed shattering (Snow et al. 1998; Burke et al. 2002). Cultivated sunflower has served as a model for understanding the genetics of domestication (e.g. Burke et al. 2005; Baack et al. 2008; Baute et al. 2015) and in identifying genes underlying agronomic traits of interest (Blackman et al. 2010, 2011; Chapman et al. 2012). As described by Kane et al. (2013), ample genetic resources are available for sunflower, facilitating not only the study of domestication genetics, but also of adaptation in wild sunflower species. The recently completed, high-quality XRQ reference genome assembly for *H. annuus* (where XRQ is an inbred genotype developed by the French National Institute for Agricultural Research, or INRA), mapping over 80% of the 3.6 Mbp genome and 97% of the gene content to 17 pseudo-chromosomes (Badouin et al. 2017), has paved the way for future ecological and evolutionary studies in the *Helianthus* genus.

The current distribution of wild *H. annuus*, the progenitor of the domesticated sunflower, extends over much of the United States (Figure 1.1), as well as parts of southern Canada and northern Mexico. However, *H. annuus* is likely indigenous to the central USA, with the hypothesized range prior to human colonization comprising a narrow column from North Dakota south to Texas (see Figure 1 in Whitney et al. 2010). In addition to its native distribution, *H. annuus* is now abundant in parts of Australia (Dry and Burdon 1986; Seiler et al. 2008), Europe (Bervillé et al. 2005; Muller et al. 2009) and South America (Poverene et al. 2009; Cantamutto et al. 2010; Casquero and Cantamutto 2016), where it

frequently acts as a weed. In contrast to the domesticated sunflower, wild *H. annuus* has multiple inflorescences with small achenes, grows indeterminately and is highly branched (Figure 1.2), though there is tremendous morphological diversity across the native range (McAssey et al. 2016). Preferring heavy, clay soils and open grasslands (Heiser et al. 1969), wild populations of *H. annuus* may be found growing in a range of open habitats that experience frequent disturbance (Heiser 1954), such as along roads and railway lines, in vacant lots and waste places, and in crop fields. This heliophilic species has been postulated to have originated as a colonizer of natural disturbances (Asche 1993), especially those created by bison, which may also have acted as a dispersal agent for achenes trapped in their fur. Wild and cultivated *H. annuus* remain interfertile, and gene flow between the two is common across the landscape (Linder et al. 1998), with crop-wild hybrids frequently reported in the native range (Arias and Rieseberg 1994; Whitton et al. 1997).

Sunflower commonly acts as an agricultural weed, infesting crop fields and their margins, in both North America (where it is native) and parts of Australia, Europe and South America (where it is not) (e.g. Al-Khatib et al. 1998; Muller et al. 2009; Casquero and Cantamutto 2016). In the USA, it has been listed as a noxious weed in several states (Iowa, Minnesota, Alaska: USDA, NRCS 2017), as it may decrease crop yields significantly in agricultural fields (Figure 1.3). For example, for corn (*Zea mays* L.), soybean (*Glycine max* (L.) Merr.) and sugar beet (*Beta vulgaris* L.) fields, a heavy infestation of weedy sunflowers can reduce crop productivity by up to 64%, 97% and 73%, respectively (Schweizer and Bridge 1982; Geier et al. 1996; Deines et al. 2004). In North America, weedy populations likely originated as wild sunflowers that colonized agricultural fields, as weedy populations tend to be more closely related to nearby wild populations (occurring in more natural areas) than other weedy populations (Kane and Rieseberg 2008); the role, if any, of crop alleles in contributing to the success of weeds remains unknown. Meanwhile, in other parts of the globe, weedy sunflowers have crop-wild hybrid origins, perhaps originating as seed contaminants in sunflower crop fields. For example, in France and Spain, Muller et al. (2011) found that, while all weeds retained a mitochondrial crop-specific marker, they also possessed a number of alleles not present in the cultivated pool; additionally, the low population structure and high marker diversity found were consistent with multiple introduction events.

1.4 Prior Research in This System

While sunflower is a disturbance-adapted species, and may therefore naturally act as an agricultural weed, both genetic and phenotypic studies of weeds have indicated that adaptation to the cultivated environment has occurred. Common garden experiments comparing agricultural weed

populations to wild populations (from non-agricultural habitats) have found evidence of trade-offs between growth and stress tolerance, with weeds favoring faster growth. For example, in a greenhouse study including four U.S. and three European weedy populations, Mayrose et al. (2011) found that weedy versus wild individuals were more susceptible to drought, dying earlier when water was withheld; under well-watered conditions, there was a marginal trend to a higher growth rate (as change in height over time) in weeds. Similarly, Koziol et al. (2012) found that weedy populations (five from Australia and four from the USA) had more wilted leaves under drought stress, and drought stress was ameliorated to a greater degree by inoculation with arbuscular mycorrhizal (AM) fungi. Differences in growth rate (higher in weeds) and root architecture (reduced fine root structure in weeds) explained 39% and 60%, respectively, of the variation in drought tolerance seen across populations. Lastly, in a comparison of Argentinean *H. annuus* biotypes, Presotto et al. (2017) found higher growth (in aboveground biomass and height) in the agricultural weedy versus wild types in an irrigated, outdoor common garden, but under drought stress, fitness was reduced to a greater extent in the weeds; however, only three populations were tested in total (one weedy, two wild). As cultivated fields represent resource-rich environments, where water may be more accessible than in natural habitats, weedy sunflowers may have traded-off tolerance of drought (and potentially other abiotic stresses) in favour of higher growth rates, to better compete with crops and complete their life cycle before the harvest period.

Weedy and wild sunflower populations are also differentiated at the molecular level. As a follow-up to their work on growth-tolerance trade-offs, Mayrose et al. (2011) investigated the expression of several candidate genes during the drought response. For two candidates, there was a significant correlation between mRNA expression level and the number of days until plant death, as well as an effect of plant type (weedy, wild or domestic); the kinetics of the HD-Zip transcription factor *Athb-8* differed between weedy and wild plants. Expression of the *Athb-8* gene was downregulated over the course of the drought response, but to a greater extent in weeds. Previous research has shown *Athb-8* to be responsive to a range of environmental cues and stressors (Baima et al. 1995), and to regulate vascular development. Also examining gene expression, Lai et al. (2008) used a sunflower cDNA microarray to characterize differences in expression between two wild and four weedy sunflower populations from the USA. When grown under standard conditions in a growth chamber, over 165 uni-genes, or roughly 5% of the array, showed differential expression between at least one weedy population and the pooled wild populations. Among these uni-genes, abiotic and biotic stimulus-response and stress-related proteins were over-represented, again suggesting that stress tolerance is an

important differentiator of weedy versus wild populations. Finally, in an analysis of 106 microsatellites in six wild and four weedy populations, Kane and Rieseberg (2008) found that between 1% and 6% of loci were outliers showing reduced genetic variability (according to $\ln RV$ and $\ln RH$ tests) in the weedy populations, suggesting that they may be under selection. There was no global reduction in variation across the genome, however, indicating no recent population bottlenecks during weed evolution. In conclusion, while nearby weedy and wild populations are genetically similar, and may share local adaptations, they are differentiated at a small proportion of key loci.

As the use of herbicides in agriculture rises year by year, an increasingly important adaptation for agricultural weeds is herbicide resistance (Powles and Yu 2010). The number of weed species with resistance to one or more herbicides continues to grow over time (Heap 2014), and weedy sunflower is no exception. On the International Survey of Herbicide-Resistant Weeds (ISHRW 2017: www.weedscience.org; retrieved on 13 Dec 2017), resistance to acetolactate synthase (ALS) inhibiting herbicides was reported for *H. annuus* beginning in 1996 for several U.S. states (Iowa, Kansas, Missouri and South Dakota). Acetolactate synthase, also known as acetohydroxyacid synthase (AHAS), is a required enzyme in the biosynthesis of branched-chain amino acids (i.e., isoleucine, leucine and valine), with ALS inhibition leading to plant starvation (Tranel and Wright 2002). Resistance to ALS-inhibitors evolves relatively easily compared to other types of herbicides, with ALS inhibitor-resistant weeds currently accounting for nearly a third of all cases of resistance (Heap 2014). As there are separate herbicide binding and catalytic domains on the ALS enzyme (Duggleby et al. 2008), many resistance mutations do not affect catalytic activity, likely explaining why resistance is common. In sunflower, resistance has evolved to many different ALS-inhibiting herbicides (e.g. imazethapyr, imidazolinone and sulfonylurea) and has also been recently reported in France in 2009 (ISHRW 2017). Interestingly, in some of these cases, despite high herbicide resistance frequencies in weed populations, resistance is uncommon in nearby unmanaged, wild populations (Massinga et al. 2003). Hence, for traits under strong selection, such as herbicide resistance, there is potential for differentiation between weedy and wild populations, even allowing for moderate levels of gene flow. More recently, resistance to glyphosate, the active ingredient in the immensely popular herbicide Roundup, has been reported for *H. annuus* populations infesting corn crops in Texas, USA (ISHRW 2017).

1.5 Summary of Studies

To characterize genetically-based trait differences between weedy and wild sunflower populations, and to elucidate potential weed adaptations, I first grew sunflowers of both types in a

common garden. As described in Chapter 2, I collected agricultural weed and non-agricultural wild sunflower populations in pairs; weedy-wild pairs were located as close as possible on the landscape, to isolate changes due to weediness per se from other local adaptations. To control for possible maternal effects resulting from the use of field-collected seed, I used seed weight as a covariate in all analyses. In 2012, I grew seedlings from nine population pairs under standard conditions in a glasshouse and estimated seedling growth as the change in aboveground biomass over time. Biomass was estimated from non-destructive measurements of seedling height and leaf number, according to a previously established linear model. After roughly six weeks, I transplanted seedlings ($n = 5$ per maternal family) into a prepared field, following a randomized complete block design. Plants were monitored daily in order to capture the date of first flower and other developmental milestones. To obtain seeds from mothers grown in a common environment, I generated within population crosses in the field in 2012. The following year, I grew sunflowers from three weedy-wild pairs in a repeat common garden, replicating the experiment twice by seed source (field or 2012 common garden); I then made the same measurements as in 2012, to examine the influence of seed source on the results.

While natural selection acts on phenotypes, trait changes are mediated by the evolution of the genes underlying these traits. For the most part, the genes underlying the evolution of weedy and invasive traits remain poorly characterized, perhaps as a result of meager availability of genomic tools and resources for many weed species (Stewart et al. 2009). In Chapter 3, I used whole genome resequencing to investigate the extent of genetic differentiation between weedy and wild sunflowers. Data were pooled across individuals from different population types (i.e., weedy or wild), meaning that any observed genetic differences must be common to multiple weedy populations, and therefore likely the result of selection and not drift. As the common garden work in Chapter 2 was carried out in Vancouver, Canada, well outside the *H. annuus* range, there was the potential for novel genotype-by-environment interactions to obscure true trait differences; using genetic data allowed me to look for signatures of selection more directly. I calculated two metrics of genetic differentiation, a “cluster separation score” (CSS: based on genetic distance matrices)(Jones et al. 2012) and F_{ST} , in windows across the genome. For CSS, I used sliding windows of 10,000 bp in size, while for F_{ST} I used a novel method (Beissinger et al. 2015) to delineate distinct, variable-sized windows on the basis of inflection points in a cubic smoothing spline of F_{ST} . Taking the consensus of both methods, I identified potential regions of divergence across the sunflower genome and then identified genes falling within these regions.

In Chapter 4, I extended the comparison of weedy and wild sunflowers on a molecular level to investigate the genetic basis of a critical weedy trait, herbicide resistance. Prior greenhouse and field trials (unpublished results) established the segregation of glyphosate resistance in my study populations, with some populations showing resistance at roughly half to two-thirds the rate typically applied by a farmer ($1.0 \text{ kg a.e. ha}^{-1}$). I grew sunflowers from 28 populations (both weedy and wild) in the glasshouse and then treated them with glyphosate at the four- to eight-leaf stage. Glyphosate was applied as a foliar spray at a rate of $0.5 \text{ kg a.e. ha}^{-1}$, or half the field rate. Survival was assessed daily, and I additionally created a metric to classify survivors according to the amount of herbicide damage they sustained. Using SNP data obtained from whole genome resequencing of the surveyed individuals and the herbicide resistance scores, I performed a genome-wide association (GWA) mapping to look for genotype-phenotype associations. Looking for candidate SNPs clustering in peaks in the Manhattan plot, I identified a set of top SNPs potentially linked to glyphosate resistance, and then identified overlapping or nearby genes, where possible.

Figure 1.1: Range of wild *Helianthus annuus* in Canada and the United States, based on Rogers et al. (1982). The range also extends southwards into Mexico.

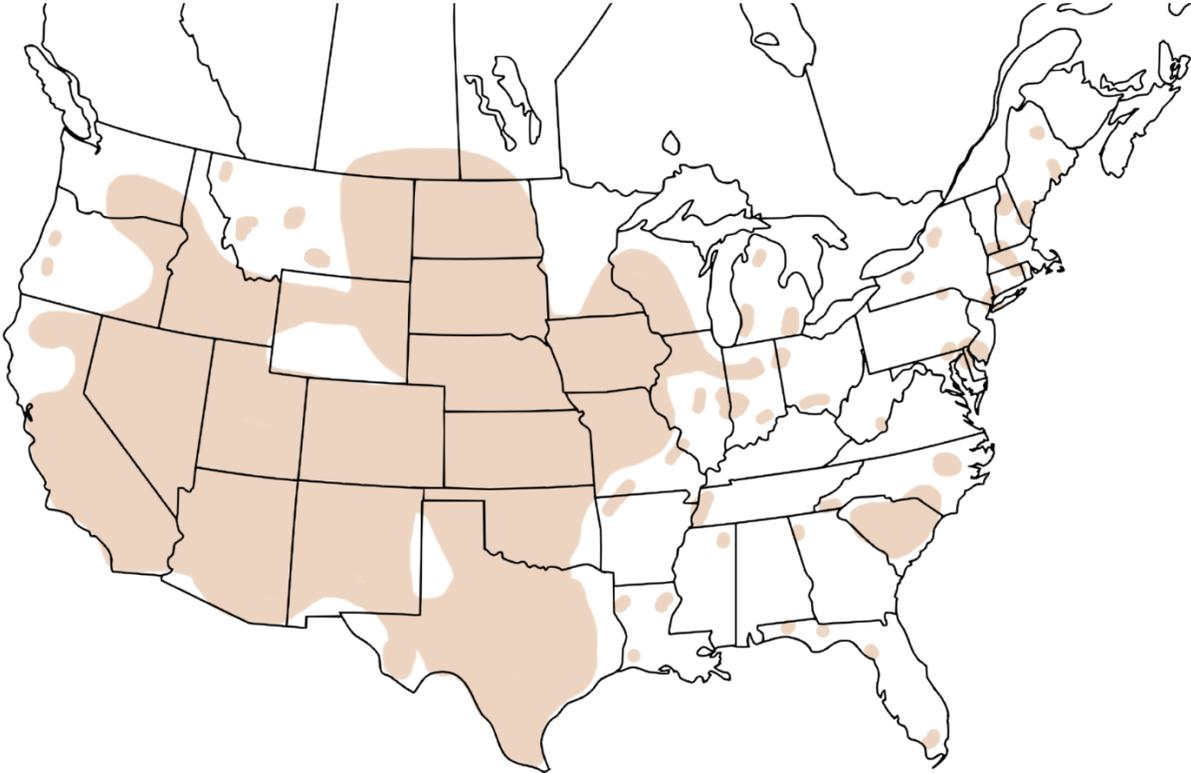


Figure 1.2: Photographs of representative domesticated (a) and wild (b) *Helianthus annuus*, taken in an outdoor common garden located on the University of British Columbia campus in Vancouver, Canada in the year 2013. The scale is approximately equal in (a) and (b). Photographs are the authors.

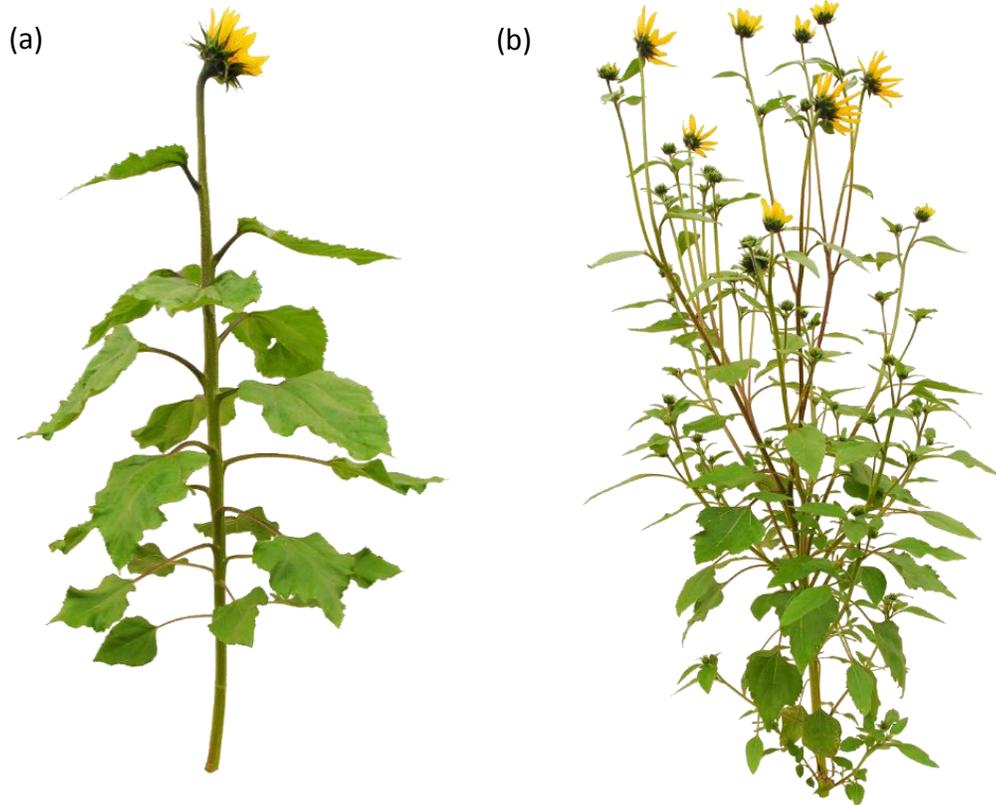
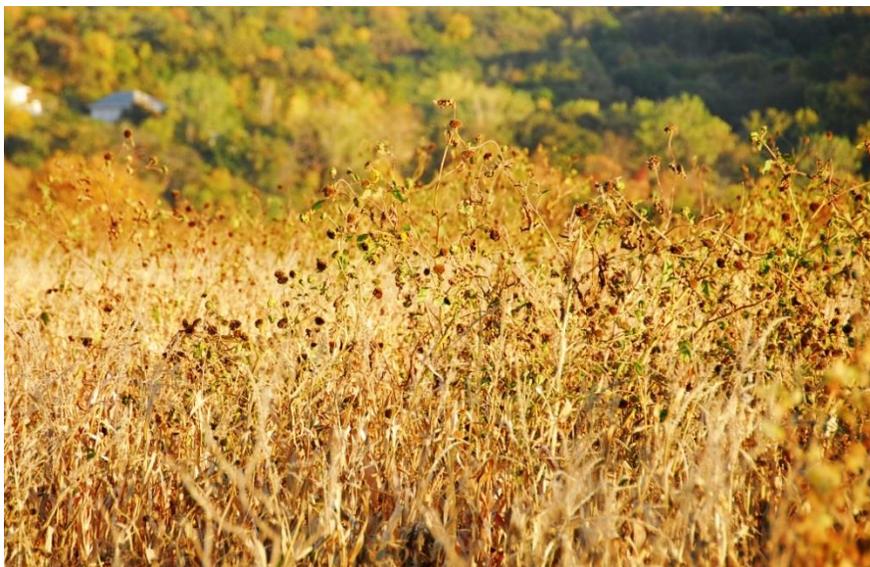


Figure 1.3: Photographs of a weedy *Helianthus annuus* population infesting a sorghum crop in Kansas (a) and weedy population competing with a corn crop in South Dakota (b). Both photos were taken towards the end of the growing season in 2011, when sunflowers had completed flowering (note the brown, dry inflorescences). Weedy sunflowers occurred throughout the crop field in both cases, competing directly with the crop plants and partially shading them. Photographs are the authors.

(a)



(b)



Chapter 2 : Common Garden Experiment Examining Life History Differences in Weedy Versus Wild Sunflower Populations

2.1 Introduction

Weeds have been colonizing agroecosystems since shortly after the dawn of agriculture, some 12,000 years ago (Doebley et al. 2006). For example, archaeologists identified 35 weed species among crop plant remains from a 9,000 year old coastal site, Atlit-Yam, in Israel (Hartmann-Shenkman et al. 2014). Agricultural weeds (also known as “agrestals”), or plants that invade crop and range lands, represent a tremendous threat to crop productivity, lowering global yields by roughly 10% annually (Oerke 2006) and costing the economy \$33 billion in the United States alone (Pimentel et al. 2005). Understanding how such weeds arise and adapt to the agricultural environment may be critical for achieving successful long-term weed control (Liebman et al. 2001). Apart from practical concerns, however, agricultural weeds also represent excellent case studies of adaptation to human-mediated selection occurring on a contemporary time-scale (Baker 1974). Weeds growing in crop fields are exposed to chemicals, such as fertilizers, herbicides and other pesticides, as well as irrigation and regular disturbance from cropping techniques such as cultivation, harvesting and ploughing. Such disturbances are often highly predictable and may impose strong selection on weed populations (Barrett 1988). For example, the selection of dwarf forms of fool’s parsley (*Aethusa cynapium* L.) and erect hedge-parsley (*Torilis japonica* (Houtt.) DC.) in cereal crops following the introduction of the reaper was one of the earliest documented cases of weed evolution in an agricultural setting (Salisbury 1962). Cases of crop mimicry, where the weed comes to physically resemble the crop in order to evade detection and removal, have also evolved, with well-documented cases including barnyard grass (*Echinochloa crus-galli* (L.) Beauv., formerly *E. oryzicola* Vasing.: Guo et al. 2017) in cultivated rice (Barrett 1983) and common vetch (*Vicia sativa* L.) as a seed mimic in lentil crops (Gould 1991). More recently, the rapid evolution of herbicide resistance has enabled the success of species such as Canada fleabane (*Conyza canadensis* (L.) Cronq.), the pigweeds (*Amaranthus* spp.) and rigid ryegrass (*Lolium rigidum* Gaud.) as agricultural weeds (Heap 2014).

Modern cultivated fields represent a unique environment and one that has only come into being in recent times, following World War II and the Green Revolution (Gould 1991), when technologies such as herbicides and mechanization, as well as planting in monocultures, became common practice. Conditions in the field are closely controlled by the farmer and are often less complex than in natural habitats (Snaydon 1980), even those such as open, disturbed sites. The goal is typically to reduce

environmental heterogeneity to produce uniform growing conditions ideal for maximizing crop yield. For weeds invading agricultural fields, adaptation to herbicides and the timing of disturbances may be crucial for survival (e.g. Tranel and Horvath 2009; Vigueira et al. 2013; Kuester et al. 2016). For example, annual weeds (especially those occurring within the crop field and not just on the fringes) must complete their lifecycle and produce seed before the crop is harvested and the field tilled. Weeds therefore need to time their phenology to closely coincide with both crop sowing (for which the field is prepared and weeds removed) and harvesting. This may result in selection for rapid development and precocious reproduction in weeds (Barrett 1983), especially those of short season crops. Early flowering has been documented, for example, in aquatic weeds of seasonally inundated habitats, such as rice (*Oryza sativa* L.) fields. In Southeast Asia, pickerel weed (*Monochoria vaginalis* (Burm. f.) C. Presl. ex Kunth) is a problematic weed of rice fields, and the short juvenile period of some annual varieties allows weedy individuals to flower before the drying period at harvest time (Steenis 1955); the phenology of populations of barnyard grass has also evolved to closely match that of the rice crops infested (Smith 1988). Similarly, in Californian rice fields, weedy arrowhead (*Sagittaria montevidensis* Cham. & Schltld.) flowers a mere month after flooding and matures seeds prior to herbicide treatment of the fields (Barrett and Seaman 1980). Such phenological shifts may be important for the success of many agrestal species but remain understudied.

Agricultural fields represent relatively benign, low-stress environments for plants, as pests are typically controlled, vegetation density low, and nutrients and water provided in excess (Mohler 2001). Weeds may come to prosper in these environments by evolving to maximize their growth rates, taking advantage of resources before they are monopolized by crops or before disturbance occurs. Indeed, in Grime & Hunt's (1975) survey of relative growth rates (RGRs) in 132 plant species, annual agricultural weeds had some of the highest RGRs. However, the high RGRs necessary for exploiting temporarily favorable conditions may come at the expense of abiotic stress-tolerance and competitive traits, according to life history theory (Grime 1977). Plants face intrinsic trade-offs in biomass allocation to defense, growth, maintenance, reproduction and storage (e.g. Coley et al. 1985; Bazzaz et al. 1987; Herms and Mattson 1992), with a higher investment in growth potentially reducing the resources available for other functions. In the field of invasion biology, it has been commonly observed that introduced species tend to be larger in the invaded versus native range (Crawley 1987). Increased growth in introduced versus native populations has also been described for a variety of species, including common ragweed (*Ambrosia artemisiifolia* L.: Hodgins and Rieseberg 2011), purple loosestrife (*Lythrum salicaria* L.: Blossey and Notzold 1995), smooth cordgrass (*Spartina alterniflora* Loisel.: Daehler

and Strong 1997), and white campion (*Silene latifolia* Poir.: Wolfe et al. 2004), and there is evidence in some cases that improved growth was correlated with lower abiotic stress tolerance. For example, Hodgins and Rieseberg (2011) found a trade-off between rapid growth and drought tolerance in ragweed, with introduced plants experiencing higher mortality under drought conditions. However, in a re-analysis of data from 32 common garden studies comparing natives and invaders, Colautti et al. (2009) demonstrated that differences in latitude between the native and invaded range may confound the results of some studies. Thus, while plant species may evolve faster growth under advantageous conditions, other factors, such as latitude, may also influence growth rates and must be accounted for.

There are three main ways in which agricultural weeds appear to arise (De Wet and Harlan 1975). First, disturbance-adapted wild species may colonize agricultural fields; these may already have many of the traits needed for survival as agrestals, but evolution may also frequently play a role in their success (Stewart et al. 2009; Vigueira et al. 2013). Secondly, hybridization between wild species and domesticated crops can produce new weeds of admixed origin, with crop genes facilitating success in the agricultural environment. Lastly, feral populations of crops that have escaped cultivation may evolve into agricultural weeds, following a process of “de-domestication” in which traits such as non-shattering (i.e., seed retention) are lost. Examples of at least the first two pathways to weediness can be found in common, annual sunflower (*Helianthus annuus* L.). *Helianthus annuus* is native to North America, where it was domesticated at least 4,000 years ago (Harter et al. 2004); wild and domesticated sunflowers remain interfertile, despite morphological differences in the crop (e.g. loss of branching, seed dormancy and self-incompatibility, as well as larger achenes) (Snow et al. 1998; Burke et al. 2002). Sunflower commonly acts as an agricultural weed in North America (where it is native), as well as in Australia, Europe and South America (where it is not), causing significant crop yield losses where infestations occur (e.g. Geier et al. 1996; Deines et al. 2004; Muller et al. 2009; Gerstein et al. 2015; Casquero and Cantamutto 2016). In North America, agricultural weed populations have been shown to be closely related to nearby wild populations occurring in more natural areas (Kane and Rieseberg 2008), suggesting that weediness has arisen independently multiple times from wild *H. annuus*. Although crop sunflower frequently exchanges pollen with wild sunflower populations on the landscape (Linder et al. 1998), and crop-wild hybrids have been reported (Arias and Rieseberg 1994; Whitton et al. 1997), crop alleles have not been hypothesized to contribute to the evolution of weed populations. In contrast, European weed populations are likely of crop-wild origin, as a crop-specific maternally inherited genetic marker was present in all weeds surveyed (Muller et al. 2011); as sunflower seed grown in Europe is often sourced from the United States, these hybrids likely originated from the unintentional pollination

of maternal lines in U.S. seed production fields by wild sunflower. Similarly, a recently described agrestal biotype of *H. annuus* in Argentina that is intermediate in its morphology between crop and wild sunflower (Casquero et al. 2013) is believed to be a crop-wild hybrid (Presotto et al. 2017).

Here, we used common garden experiments to quantify life history differences between weedy and wild populations of *H. annuus* collected across the southern Canadian prairies and the US Midwest. Previous greenhouse work has revealed that weedy sunflowers tend to exhibit faster growth than wild sunflowers when grown as seedlings under benign conditions (Mayrose et al. 2011; Koziol et al. 2012; Presotto et al. 2017); however, weedy seedlings were more susceptible to drought and fungal infection, as well as more palatable to a generalist insect herbivore. These studies suggest that weedy sunflowers from North America may have evolved faster growth at the expense of defense and stress-tolerance traits. However, studies did not account for potential differences in maternal effects on seed provisioning among populations (and seed was collected from very different habitats); also, weedy and wild populations were not matched for climate and spanned a large geographical transect. Thus, environmental differences among source populations may influence the results. To address these issues, in this study, we used seed collected from weedy-wild population pairs, located as close as possible to one another on the landscape; this paired design should control for climatic differences and allow us to isolate changes due to weediness per se from other local adaptations. We used seed weight as a covariate in all analyses to account for maternal effects, and additionally conducted a separate follow-up study to directly compare the growth of individuals from two seed sources: field-collected seed (different maternal environments) and common-garden generated seed (same maternal environments). In both experiments, we measured seedling growth and then followed individuals to flowering. If weeds have experienced selection for a shortened life cycle as a result of agricultural disturbance, or to take advantage of relatively benign crop-field conditions, weedy populations may show increased growth and earlier flowering than wild populations.

2.2 Materials and Methods

2.2.1 Study Populations and Sample Collection

In the fall of 2011, I collected seeds from each of twenty populations of wild sunflower, *Helianthus annuus*, over a latitudinal gradient across the Midwestern United States and Southern Canada (Figure 2.1). This region consists of rich farming land with fertile soils, and accordingly agriculture drives the local economy (Hatfield 2012), with extensive areas on the landscape planted in monocultures. Maize (*Zea mays* L.) and soybean (*Glycine max* (L.) Merr.) are the two most common

crops (United States Department of Agriculture, National Agricultural Statistics Service: www.nass.usda.gov; retrieved on 21 Oct 2017). Across the study region, *H. annuus* shows tremendous variation in plant size, architecture and phenology; to isolate changes due to agricultural weediness per se from other local adaptations, I used a paired design, matching local agricultural weed and non-agricultural populations. Paired populations were located as close to one another as possible; the mean distance between pairs was $28.5 \text{ km} \pm 7.6 \text{ km}$ (mean \pm standard error). Additionally, to minimize the impact of outside gene flow, all selected populations were large, averaging $\sim 2,000$ individuals (range = 250 to 10,000+ individuals). Finally, to avoid gene flow from cultivated sunflower, selected populations were a minimum of 1 km from any crop sunflower fields in 2011 (though populations may have received crop pollen in previous years).

I targeted agricultural weed populations (hereafter “weedy” populations) that had a long-term presence in a given area, as established by collection records from our lab, lab alumni and the United States Department of Agriculture (USDA); wherever possible, I also spoke to landowners to confirm local land-use history and the long-term presence of weedy sunflowers. Agricultural weed populations were found competing directly with a cultivated crop of either corn ($n = 3$), soybean ($n = 3$), wheat (*Triticum* spp.) ($n = 2$) or sorghum (*Sorghum bicolor* (L.) Moench) ($n = 2$) in areas of high-intensity agricultural use. Each weedy population was paired with a nearby non-agricultural wild population (hereafter “wild” populations) at roughly the same latitude. Wild populations occurred in grasslands, fallow lands, roadside ditches, wetlands and other unmanaged lands. Finding wild populations was challenging given the dominance of agriculture across the landscape; in the end, we proceeded with the analysis of only nine weedy-wild population pairs ($n = 18$ populations), as the remaining pair (Kansas 1; Figure 2.1) did not meet our selection criteria. In Kansas 1, I was not able to locate a true wild population on non-agricultural land, only a weedy-wild hybrid population growing mostly on rangeland, but also nearby fallow lands and the margins of a sorghum field. While the sunflowers were not subject to removal efforts, a large subset of the population was certainly influenced by agricultural chemicals, fertilizers, and irrigation, as well as disturbance by beef cattle. Upon careful consideration, we decided that the physical attributes of the rangeland were too similar to those of crop fields, and we excluded this population and its match from the weedy-wild analysis, though Kansas 1 does appear in overall descriptions of the study populations.

For each population, I recorded site characteristics such as the elevation, distance to the nearest road and any associated physiographic features, as well as population characteristics such as flowering

stage, size and spatial pattern (Table 2.1). I collected seeds from a representative sample of 40 focal maternal plants, which I selected by first establishing a transect through each population along its longest axis, then walking the transect and selecting the closest plant every ~10 plants (depending on population size). For each focal plant, I measured the height (as the length of the main stem), height to the lowest branch, number of branches, branching degree, stem diameter (at the ground) and number of inflorescences. The branching degree is the number of branches on branches, e.g. a main stem with only primary branches would have a degree of one, a main stem with primary branches that had secondary branches (growing on the primary branches) a degree of two, and so on. For one weedy population, Kansas 2, most individuals had an unusual phenotype: the main stem had been severed early in the plant's growth (perhaps by mowing) and there were one or two basal branches growing in place of the main stem. For these individuals, I selected the largest basal branch and treated this as the main stem. Seeds collected from different maternal families were kept separate.

Summary statistics for the maternal plant traits were calculated using base R (version 3.3.0: R Core Team 2017) and plots created using the R packages *ggpubr* and *ggplot2* (Wickham 2009). To explore the range of phenotypic variation among populations for plant size, architecture and fitness, I performed a two-way Analysis of Variance (ANOVA) for each quantitative trait with location (i.e. the ten population pairs) and population type (weedy or wild) as independent variables, using the *stats* and *car* packages in R (Fox and Weisberg 2011); Tukey's HSD tests were performed post-ANOVA to compare factor means. Dependent variables were log or square root transformed as necessary to meet model assumptions of normality in the residuals and homogeneity of variances between groups. As branching degree can be thought of as an ordinal variable, I used chi-squared tests to examine the relationship between degree and population location or type.

To explore the variation in local climate for my study populations, and to confirm that paired populations experience similar conditions, I generated annual climate data (mean annual temperature, precipitation, etc.) for the twenty populations using the web-based version of ClimateNA (v5.21: Wang et al. 2016), available at www.climatewna.com (retrieved on 21 Oct 2017). Historical data were averaged for the 1961-1990 normal period. The ClimateNA software is unique in that it locally downscales climate data layers into scale-free estimates of climate values, using a combination of bilinear interpolation and elevation adjustments (details in Wang et al. 2016). Raw data for the 23 climate variables obtained for each population are presented in Table 2.2. Relationships among the climate variables and study populations were investigated using Principal Components Analysis (PCA), as

implemented in the *stats* package in base R. Climate variables were each normalized to have a mean of zero and unit variance prior to PCA to control for differences in magnitude among variables. I created the PCA plot using the packages *ggbiplot* and *ggplot2* (Wickham 2009).

Germination trials performed in January 2012 revealed that one of the weedy populations (from Manitoba) had seeds that, while appearing viable, did not germinate. We replaced this population with an accession obtained from the USDA (National Plant Germplasm System PI 592327) that was collected along the edge of a harvested wheat field in 1994, in roughly the same location as the population to be replaced. The replacement can thus be considered an historical version of Manitoba weedy. The USDA provides bulk seed and so the family structure is unknown.

2.2.2 Seedling Common Garden and Growth Study

Wild sunflower seeds possess strong seed dormancy that must be broken in order for germination to proceed (Chandler and Jan 1985). As such, direct-field seeding typically results in very low germination rates, and we instead germinated seeds indoors in a glasshouse at the University of British Columbia (UBC) in Vancouver, Canada to produce seedlings to plant in the field. Our goal for the common garden was to include five individuals from each of five maternal families per population ($n = 25$ per population). As germination trials revealed less than perfect germination rates for some populations, with rates ranging from ~66% to 90%, I started seeds for an average of ten maternal families per population (range = 7 to 17) and a surplus of seeds (ten or more) per family. The maternal families were randomly selected from the 40 available per population, though families that completely failed to germinate in trials were excluded. For the replacement Manitoba weedy population, germination rates were unknown and I started a total of 35 seeds. Prior to starting, seeds were weighed for all maternal families ($n = 30$ seeds per family) to obtain the mean seed weight as an estimate of maternal provisioning.

Seeds were scarified on two separate dates one week apart. Germination trials had indicated that seeds from more southern populations (Iowa, Kansas, Missouri, and South Dakota) did not germinate as readily as those from northern populations (Manitoba, North Dakota, Saskatchewan), so the goal was to synchronize germination. Southern population pairs were scarified on March 27th, 2012 and northern population pairs on April 3rd, 2012; matched weedy and wild populations for a given location were always scarified on the same day. Scarification involved cutting off the blunt (widest) end of the cypsela, removing up to a quarter of the husk (i.e. pericarp). Seeds were then placed on moist filter paper in petri dishes to imbibe overnight. Seed coats were fully removed on the following day to

enhance germination. I changed filter paper daily and watered with a 1% solution of the biocide plant preservative mixture, or PPM (Plant Cell Technology, Washington, DC, USA), to reduce microbial contamination, though some seeds still became contaminated; these were removed daily and the remaining seeds moved to a new dish. Seeds were germinated in the dark, with dishes moved into the light once seeds began producing chlorophyll. A seed was considered “germinated” and suitable for planting once the primary root was at least 1 cm long and ideally some secondary roots had also appeared (this did not always happen prior to planting, depending on the population).

Perhaps owing to enhanced natural light availability in the spring, or the longer amount of time the seeds had had to ripen, southern populations now germinated readily, making it necessary to begin planting just before scarifying the northern populations, on April 2nd, 2012. For each population, I waited until all or the majority of families had germinated before randomly selecting five families to plant. My goal was to not bias the selection of families on the basis of ease of germination. Seven seedlings were planted per family (to have extras), with each planted into a 5 cm diameter conical “Deepot” (Stuewe & Sons, Inc.) prefilled with moistened, standard potting soil mix (75% peat with 25% perlite). Deepots were placed into support trays that held 50 Deepots each, for a seedling density of 269 per m². Support trays rested on a glasshouse bench that flooded twice daily with weak fertilizer solution prepared on-site; at the outset of the experiment, seedlings were also misted frequently to keep the top layers of soil moist. Seedlings received 16 hours of supplemental lighting per day, delivered by 600 W high-pressure sodium lights. On April 13th, 2012, when all seedlings were planted, I fully randomized the experiment. Not all seedlings survived the transition into soil and there were also a number of seedlings with growth abnormalities (e.g. fused first true leaves, lack of apical meristem, etc.) that only became apparent after planting; I replaced these individuals (~10% of the experiment) as needed using retained seedlings on petri dishes or, if these were no longer available, newly scarified seeds. Scarification dates and planting dates were recorded individually.

To assess potential early life history differences between weedy and wild sunflowers, I documented seedling growth beginning on April 6th, 2012. Non-destructive measurements were taken every two days throughout the month of April, before transferring seedlings outdoors to harden-off in May. Growth slowed at this time and measurements were taken roughly every four days from April 27th until May 10th, 2012. For each seedling, I recorded the height (i.e. length of the main stem) and the number of fully expanded true leaves on each date after planting. To characterize the relationship between these non-destructive measurements and seedling biomass (i.e. dry weight), I planted extra

germinants (n = 70) not needed for the main experiment in early April. Every few days, a number of seedlings were measured (height, number of leaves and largest leaf dimensions) and then harvested. Plant materials were dried in an oven at 60°C for three days before weighing.

Using multiple linear regression in R, I determined that biomass was highly predictable based on measurements of seedling height and leaf characteristics. To improve model predictions, I included an additional 157 individuals in the analysis (n = 227 total) harvested from other greenhouse experiments (not reported) using the same populations. These seedlings were grown in the same glasshouse, Deepots, etc. though experiments took place at different times of the year and included a broader diversity of maternal families. Seedlings ranged greatly in size (Table 2.3). Exploratory graphs (Figure 2.2) showed strong relationships between all measured variables and biomass. Biomass was transformed using a Box-Cox power transformation (*MASS* package: Venables and Ripley 2002) to achieve normality in the model residuals. The best model, not including interactions, was:

$$Biomass = \left(\frac{Height}{100} + \frac{Leaf\ Half\ Width}{700} + 0.44 \right)^{4.55}$$

which had an R^2 of 92% ($F_{2,224} = 1321, p < 0.001$). The number of leaves, leaf length and leaf half width were essentially interchangeable in the model. The model including leaf number had an R^2 of 90% ($F_{2,224} = 983.5, p < 0.001$):

$$Biomass = \left(\frac{Height}{100} + \frac{Leaf\ Number}{200} + 0.39 \right)^{4.55}$$

Including interactions was not found to greatly improve the amount of variation explained by the model (best model with interactions, $R^2 = 93\%$), and so simpler models, without interactions, were preferred. Seedling measurements from the common garden were converted into biomass estimates using height and leaf number.

2.2.3 Field Common Garden and Flowering Study

We transplanted seedlings into a common garden at the UBC Farm to assess differences among populations in reproductive traits. While it is possible to grow annual sunflowers to maturity in pots, in our experience this leads to dramatically different plants compared to those grown in the ground. Branching, for example, is reduced, plants are spindlier, and flowering time may be affected as pot-bound plants can flower prematurely due to stress. We therefore planted the seedlings at the UBC Farm (ubcfarm.ubc.ca; retrieved on 23 Oct 2017), a 24-hectare integrated teaching and research space on the

UBC campus with predominantly sandy-loam soils (Dennis and Kou 2014). The UBC Farm follows organic practices and fields are regularly amended with compost to maintain soil fertility. The field used for the common garden had a gentle slope with a south-west aspect; soils also became sandier in texture from north to south.

The common garden followed a randomized complete block design, with one replicate of each population per block, to control for observed differences in soils across the field. The longest axis of the field ran from approximately north-east to south-west, and so blocks ($n = 25$) were configured as parallel rows, running across this axis (i.e. from the north-west to south-east side). Blocks, or rows, were spaced 1.5 m apart, and plants within rows were 1 m apart. Without assigning specific individuals, maternal families for each population were randomly assigned to rows. Within a row, the order of the populations was then randomized with the constraint that populations be distributed approximately evenly across the shorter axis of the field, so e.g. all individuals from a given population could not occur only on the north-west side. To achieve this effect, I divided the rows into three subsections (seven, six and seven individuals wide) and required that each population occur at least seven times per subsection across all rows. To reduce edge effects, the common garden included a single-plant border around the exterior, following the same spacing as within the garden.

Seedlings were hardened-off outdoors at the UBC Farm from May 2nd to 11th, 2012. For the first five days, they were placed in a sheltered, shady area before moving them into the sun for the next five days. On May 11th, 2012 the seedlings were planted into a freshly tilled and prepared field, randomly selecting an individual from the maternal family assigned to each position. The border was planted on May 15th, 2012 using left-over seedlings placed in random order. Irrigation was provided by sprinklers on 1 m risers, which were run daily for the first two weeks, and then twice a week (for ~1 hr each time) as needed for the duration of the summer. Conditions were drought-like in July and August, and so irrigation was necessary to avoid significant stress to the plants; riser height was increased to 3 m in July to allow for better water distribution. Eight plants died within the first three weeks of the experiment, most as a result of breakage after a high wind event. Each was replaced with another seedling from the same population and maternal family, with the exception of a wild Manitoba individual where no seedlings from family 4 were available (family 16 was substituted). Germination was very poor for Manitoba wild and, in planting the field experiment, I also had to substitute an individual from family 14 for family 8; similarly, for Missouri weedy, an individual from family 23 was substituted for 33. With the

exception of these two populations, I was otherwise successful in planting five individuals from five different maternal families per population.

We surveyed sunflowers daily to record the date of first flower for all individuals and track the progression of flowering. An inflorescence was considered to have “flowered” on the date when stamens first emerged. More than a quarter of plants became infected with *Sclerotinia sclerotiorum* (Lib.) de Bary, a generalist fungal pathogen causing stem rot disease, which can be fatal if the main stem is girdled. Onset of disease systems was typically concurrent with the initiation of flowering, though this varied, and four individuals died pre-flowering; these were excluded from the analysis. To obtain seeds for each population from mothers growing in a common environment, I made within-population crosses during peak flowering. Plants from different maternal families within a population were paired based on proximity and overlap in flowering, and pollen exchanged reciprocally to create new maternal families of full-sibs. Focal inflorescences were enclosed in organza bags while still buds; the organza allows air flow while preventing pollinator access to pollen. Pollen was exchanged every second day between the inflorescences of matched individuals using a paper towel; pollination success was near 100%. Seeds were harvested when completely ripe (at the stage when the inflorescence dries and turns completely brown), threshed and stored on silica gel.

2.2.4 Maternal Effects Experiment

Seed weight is commonly used as a metric of maternal investment (e.g. Jonas and Geber 1999; Sambatti and Rice 2006; Dlugosch and Parker 2008), though cotyledon length may also be used (e.g. Parker et al. 2003; Angert et al. 2008). Including seed weight as a covariate in models may therefore account for maternal effects, and authors then often attribute any remaining differences among individuals to genetics. However, it is important to note that the maternal environment may affect factors other than seed provisioning, for example the seed oil or protein content (Roach and Wulff 1987), which may not be accounted for by a seed weight covariate. Here, we had the opportunity to compare the growth and reproduction of individuals obtained from different seed sources directly, owing to the within-population crosses we made in 2012. In the field, weedy and wild sunflower populations occupied different habitats which potentially led to differences in maternal provisioning between the two types. In the 2012 common garden, conditions were much more uniform and environmental influences on seed provisioning should be the same for weedy and wild maternal plants.

Seeds from two sources were used to initiate a second common garden in 2013 to examine in more detail the role of maternal effects on traits measured in the 2012 common garden. This

experiment included a subset of three locations, North Dakota and South Dakota 1 and 2, for which we had good seed yields; we were unable to harvest mature seeds for the majority of population pairs owing to cold, rainy fall conditions that arrested seed development prematurely. In the 2013 common garden, I compared growth rates and flowering time for individuals from field-collected seed (obtained from wild populations in 2011) and common garden seed (generated in 2012 experimental crosses). The same five maternal families included in the 2012 common garden were used here for the field-collected seed, with the exception of family 30 from South Dakota 2 wild; as there were no more seeds available, family 4 was used instead. In selecting families to use from the common-garden generated seed, I attempted to maximize diversity by including materials from crosses made using all five maternal families as both mothers and fathers (Table 2.4), though this was not always possible in the case of the fathers.

Seeds were scarified on April 2nd, 2013 using the protocol established for the 2012 common garden. We germinated ten seedlings per maternal family, except for some field-collected families from South Dakota 2 that had poor germination rates ($n = 12$ instead), and planted five individuals per family in the glasshouse ($n = 300$ total). Seedlings were maintained under identical conditions to the previous year, and the experiment was fully randomized after planting all materials on April 9th, 2013. Only four seedlings died following planting, and these were not replaced. Seedling height and the number of leaves was recorded every second day from April 8th to May 7th, 2013. Seedlings were hardened-off in a protected area adjacent to the glasshouse for one week before planting into a prepared field at the Totem Field Research Station, a 12 hectare research facility located on the UBC campus. Owing to the heavy infestation of *S. sclerotiorum* in soils at the field site used in 2012, we elected to plant the maternal effects common garden at a different location. We planted three randomly-selected seedlings per maternal family (i.e. $n = 15$ per population per seed source, $n = 180$ total), with one replicate from each population and seed source per experimental block. Again, maternal families from a given population and seed source were randomly assigned to each block and position within a block. Blocks were arranged in five rows running west-to-east, with 1.5 m spacing between rows and 1 m spacing between adjacent plants within rows. A single plant border composed of left-over seedlings was used to reduce edge effects. We surveyed plants daily to record the first flower date and counted the total number of inflorescences produced by each individual at the end of the season.

2.2.5 Statistical Analyses of Common Garden Data

For seedlings in both common garden experiments, we were interested in analyzing differences in growth between sunflower types (i.e. weedy vs wild) over the period from germination to planting in the field. I analyzed growth curves using linear mixed effects models with the biomass of each individual as the response variable. Fixed effects included the time of measurement, the type of sunflower and seed weight (taken as the familial average), as well as interactions between time and type. Seed weight was used as a covariate to represent possible differences in maternal provisioning. To account for non-linear patterns in seedling growth, I also included a quadratic term for time of measurement. Random effects of individual (to account for the repeated measures), location (i.e. which population pair) and maternal family on model intercepts were also included. The full model, using R syntax for the random effects, was:

$$\text{Biomass} = (\text{Type} * \text{Time}) + (\text{Type} * \text{Time}^2) + \text{Seed Weight} + (1|\text{Location}/\text{Maternal Family}) + (1|\text{Individual})$$

Biomass was first zeroed (by subtracting the minimum value from all values) then log-transformed (adding a one first) to meet model assumptions. The significance of each term was determined by comparing nested models using likelihood ratio tests (LRTs), dropping non-significant effects one at a time (cut-off p -value = 0.05). We decided to test the significance of random effects (again using LRTs and comparing to the full model), though there is debate in the literature regarding whether such tests should be performed (see Hurlbert 1984; Pinheiro and Bates 2000). For the maternal effects common garden, data from different seed sources (i.e. field versus common garden) were analyzed separately.

To look for differences in first flower date between weedy and wild sunflowers in both common garden experiments, I again used linear mixed effects models with random intercepts. Here, fixed effects included the type of sunflower and seed weight covariate, and random effects included location, maternal family and experimental block. The full model in 2012 was:

$$\text{Days to First Flower} = \text{Type} + \text{Weight} + (1|\text{Location}/\text{Maternal Family}) + (1|\text{Block})$$

Note that in the 2012 common garden, we discovered a planting mistake mid-season in block 2; owing to a smudged label, Iowa 1 wild family 4 (IA1W4) was planted in place of Iowa 2 wild family 4 (IA2W4) and the mistake was not noted in initial checks. Thus, block 2 was incomplete in the analyses (no IA2W), but this was taken into account by the model. For the 2013 common garden, an additional fixed effect of seed source and its interaction with type was included:

$$\text{Days to First Flower} = (\text{Type} * \text{Source}) + \text{Weight} + (1|\text{Location}/\text{Maternal Family}) + (1|\text{Block})$$

The number of days to first flower (day zero was the unique scarification date for each individual) was transformed to better meet model assumptions of normality and homogeneity of variances in the residuals. A square root transformation was used for the days to first flower in 2013; for 2012, a stronger transformation was needed, and I took the square of the reciprocal of days to first flower. Significance testing was again achieved comparing nested models with likelihood ratio tests, and all analyses were performed using the *lme4* package in R (Bates et al. 2014).

2.3 Results

2.3.1 Source Populations were Well-Matched for Climate and Phenotypically Diverse

Population pairs were well matched climatically (Figure 2.3), confirming that we selected weedy-wild pairs that experienced similar conditions apart from habitat. In the PCA plot, most pairs clustered together, though the wild population of Kansas 2 was closer to the Kansas 1 populations (later excluded from analysis) than the Kansas 2 weedy population. The first principal component (PC1) separated locations roughly on the basis of latitude. More northern locations (such as the Canadian populations and North Dakota) had higher PC1 values, a greater temperature differential between the warmest and coldest months (TD), and lower mean annual temperatures (e.g. MAT, MWMT) overall. Meanwhile, southern locations (such as Kansas and Missouri) had lower PC1 values, lower temperature differentials and higher mean annual temperatures. The second PC axis (PC2) further separated locations on the basis of moisture availability, with higher PC2 values associated with greater annual and seasonal precipitation (e.g. MAP, MSP), higher relative humidity (RH) and a lower heat to moisture index (AHM, SHM). Hot, dry locations such as Kansas had greater values of the heat to moisture index than cooler, wetter ones such as southwestern Missouri. Overall, locations spanned a diverse range of climatic conditions, as expected given the latitudinal gradient over which we sampled.

Sunflower populations showed substantial phenotypic variation across the collection region (Figure 2.4). Though we not able to score phenological variables (owing to arriving at the end of the season in order to collect seeds), we noted differences in plant size, architecture and fitness. For all measured continuous variables (height, height to the lowest branch, stem diameter, number of branches and number of heads), use of ANOVA revealed an interaction between the location and type of population ($p < 0.0001$), indicating that weedy-wild differences were not consistent across locations. For example, weedy sunflowers were taller than wild ones on average for some locations (e.g. Iowa 2 and Missouri), but the reverse was true in other locations (e.g. Kansas) (Figure 2.4a). Though main effects cannot be interpreted in the presence of a significant interaction, the ANOVA sum-of-squares

was larger for location than for the location-by-type interaction for all variables; in the case of height, height to the lowest branch and stem diameter, it was almost double. Thus location accounted for more of the phenotypic variation than the location-by-type interaction or type itself (which explained the smallest proportion of variation in the data). Looking at plant fitness (Figure 2.4b), wild sunflowers produced more inflorescences than weedy ones overall, though this trend was driven by three locations (Kansas, Iowa and North Dakota). Lastly, for branching degree, wild sunflowers were more branched than weedy ones ($\chi^2 = 15.3$, $df = 5$, $p = 0.009$), and branching degree also varied by location ($\chi^2 = 242.2$, $df = 45$, $p < 0.001$) with no obvious pattern. In conclusion, sunflowers varied greatly in measured traits across locations and, while weedy-wild pairs often differed within locations, type differences were not consistent in direction across locations.

2.3.2 Weedy Sunflowers Exhibit Faster Growth and Earlier Reproduction

Sunflower seedling biomass in the 2012 common garden increased over time (Figure 2.5a), with a slight slowdown visible in the last week (day 23 onwards) at the beginning of the hardening-off period. Though growth followed a similar curve for all seedlings, some seedlings grew much faster than others and achieved relatively larger sizes by the end of the experiment (note the different y-axes in the figure). On a log-scale (Figure 2.5b), the slope of the growth curve at any time point illustrates the relative growth rate (RGR: $\text{mg mg}^{-1} \text{day}^{-1}$) at that time. If seedling growth were exponential, the RGR would be constant and growth curves would follow a straight line on the log-scale (Paine et al. 2012); here, the RGR decreases at later measurement dates and the curves deviate from a straight line, confirming that growth slowed towards the end of the measurement period.

Trend lines for weedy versus wild seedling growth (in untransformed biomass) are presented separately for each location in Figure 2.6. Visually, there were large differences in seedling biomass among locations, with Missouri having some of the largest seedlings and Saskatchewan the smallest. For eight out of nine locations, weedy seedlings grew faster than wild ones, achieving a higher biomass by the end of the experiment. The only exception was Iowa 1, where the trend was reversed; also, in South Dakota 1, growth curves were similar for the two types, and though the weedy sunflowers had a higher mean biomass at the last few measurement dates, confidence intervals for the two types were overlapping. Maternal family was an important source of variation in growth within some populations (e.g. in Iowa 1) but not others (e.g. Saskatchewan: no appreciable difference among families).

To examine differences in growth for weedy versus wild sunflowers statistically, I implemented a mixed effects model of individual biomass over time, accounting for the non-linearity of growth curves

on the log-scale by incorporating a quadratic term for time ($p < 0.001$). Neither the type \times time ($p = 0.61$) nor type \times time² ($p = 0.59$) interaction were significant, indicating that patterns of growth (i.e. shape of the curves) were similar for weedy and wild sunflowers. However, there was an effect of type ($p = 0.013$) such that, on average, weedy sunflowers were larger than wild sunflowers. Seed weight also had a positive effect on growth ($p = 0.010$), with heavier seeds producing larger sunflower seedlings. Random effects of individual, location and maternal family were included in all models when testing the significance of the fixed effects. Differences among individuals accounted for the most variance (44%), followed by maternal family (39%) and lastly location (9%). It made sense to include these random effects whether they were significant or not, as both individual and maternal family represent sources of pseudoreplication, and the data are naturally paired on the basis of location. However, all were significant according to LRTs: $p < 0.001$ for individual and maternal family, and $p = 0.034$ for location. Thus, seedlings from a given location or maternal family were more similar to one another than seedlings from different locations or maternal families. Removing Iowa 1 (the discordant location) from the analysis and re-running the models, the interaction terms for type \times time remained non-significant (both $p > 0.05$), but the p -value for type decreased slightly from $p = 0.013$ to $p = 0.010$; interestingly, the effect of seed weight was no longer significant ($p = 0.12$). Importantly, the same results were obtained when fitting a three-parameter logistic model to the biomass data, confirming the results of the quadratic model.

To examine if faster weedy seedling growth also led to earlier reproduction, I used a linear mixed model to compare the time to first flower across locations and sunflower types (Figure 2.7). Weedy populations flowered earlier on average for most locations: in Iowa 1 and North Dakota the wild population flowered earlier. Model results differed depending on whether or not Iowa 1 (which also showed a reversed pattern for seedling growth) was included in the analysis. Including Iowa 1, the effect of type on flowering date was only marginally significant ($p = 0.051$) and seed weight had no effect ($p = 0.64$). The random effects of location and maternal family were both important for explaining variation in flowering time ($p < 0.001$ for both), though location explained the majority of the variance (89% vs 4%), while experimental block did not influence flowering date ($p = 0.19$; variance explained $< 1\%$). Excluding Iowa 1, the effect of type became significant ($p < 0.001$; weedy flowering earlier than wild), while weight remained unimportant ($p = 0.21$); in the random effects, block was marginally significant ($p = 0.072$), though it explained much less variation ($< 1\%$) in flowering time than location (91%) or maternal family (2%). Flowering time was roughly correlated with latitude, with more northern populations flowering earlier (Figure 2.7), and there was also less variability in flowering date for

northern populations. Similarly, differences between weedy and wild were slighter at higher latitudes; for example, in Saskatchewan, weedy individuals flowered an average of 4.6 days earlier than wild individuals, while in Kansas 2, the weedy population flowered 13.7 days earlier. In contrast to seedling growth, the proportion of variance in flowering time explained by maternal families (within locations) was less than for location itself. Over all locations, faster seedling growth generally led to earlier flowering, with the only exception the North Dakota weedy population.

2.3.3 Seed Source Influences Seedling Growth but Not Time to Flower

In the 2013 maternal effects common garden, sunflower seedling biomass again increased over time (Figure 2.8), following a curve similar to that seen in the previous year (Figure 2.6). However, seedlings grew slightly faster, without any visible slow-down during hardening-off, and so the greenhouse measurement period was shorter (29 versus 34 days). For both field and common garden sourced seed, weedy-wild differences were less exaggerated than in the previous year. The effect of sunflower type differed depending on seed source. Analyzing only individuals derived from field-sourced seed in a quadratic mixed model, the type x time and type x time² interactions were not significant ($p > 0.05$), nor was the effect of type ($p = 0.27$), confirming that there was no consistent difference between weedy and wild, as seen in Figure 2.8b, and in contrast to the 2012 results. As in 2012, however, the quadratic term time² was significant ($p < 0.001$), seed mass had a positive effect on growth ($p = 0.0062$), and differences among both individuals and maternal families explained a significant proportion of the variation in growth (33% and 29%, respectively; $p < 0.001$ for both). Here, location did not explain any additional variation after accounting for maternal family ($p > 0.1$) and there was a large fraction of residual variance not accounted for by the model (37%). Meanwhile, analyzing only data for individuals from common-garden sourced seed (Figure 2.8a), there was a type x time interaction ($p < 0.001$; for the quadratic interaction type x time², $p = 0.11$), so the main effect of type was not investigated. Again, the quadratic term was significant ($p < 0.001$), larger seeds resulted in larger seedlings ($p = 0.026$) and individual, location and maternal family each explained a significant proportion of the variation in growth (17% and $p < 0.001$; 31% and $p = 0.011$; and 19% and $p < 0.001$, respectively). Thus, while the field-collected seed did not recapture weedy-wild differences observed in the 2012 common garden, weedy seedlings consistently grew faster than wild ones for common garden-sourced seed.

Compared to 2012, flowering in the 2013 common garden began earlier in the calendar year. Averaging across seed sources and population type, and correcting for different scarification dates, North Dakota flowered roughly thirteen calendar days earlier in 2013 versus 2012, South Dakota 1

flowered seven days earlier, and South Dakota 2 flowered nine days earlier. Flowering time was earliest for North Dakota in 2013 (Figure 2.9), followed by South Dakota 1 and then 2, in order of decreasing latitude. However, confidence intervals overlapped for all weedy-wild population pairs for a given location and seed source, with the exception of common-garden sourced seed for North Dakota, for which the wild population flowered earlier. Statistical analysis using a linear mixed effects model confirmed that there was no interaction between type and seed source ($p = 0.69$), and no main effects of either source ($p = 0.66$) or type ($p = 0.25$). As found in the previous year, seed weight did not have an effect on flowering date ($p = 0.37$). Of the random effects, location explained 71% of the variance, maternal family 3% and the experimental block virtually none; only location explained a significant proportion of the variance according to likelihood ratio testing ($p < 0.001$). Thus, in 2013, the field-collected seed did not show the same pattern of weedy-wild differences as in 2012, although location differences were recaptured. For the common-garden generated seed, North Dakota and South Dakota 2 did show the same patterns in 2013 versus 2012 (compare Figures 2.7 and 2.9), but South Dakota 1 did not. Here, faster weedy seedling growth did not lead to an earlier flowering date.

2.4 Discussion

In this study, we investigated life history differences between *Helianthus annuus* populations found directly competing with crops in agricultural fields (i.e., “weedy” populations) and paired wild populations found growing in non-agricultural, unmanaged areas (i.e., “wild” populations). With collection locations spanning a latitudinal gradient from Manitoba, Canada to Kansas, USA, the paired design allowed us to isolate changes due to agricultural weediness per se from other local adaptations. We grew paired populations in a common garden in Vancouver, Canada in order to compare seedling growth rates and the time to first flower; under standardized environmental conditions, any differences observed among individuals should be due to genetics and not phenotypic plasticity, though non-genetic influences of maternal environment (“maternal effects”) may also influence phenotypes. After controlling for seed weight (a proxy for maternal effects) in our analyses, we consider among-population differences to be genetically based, although we cannot entirely exclude the influence of maternal or epigenetic effects on offspring phenotype. In the 2012 common garden, weedy seedlings grew faster than wild seedlings and flowered earlier for most locations, suggesting an adaptive shift in life history strategy for weedy populations. A follow-up common garden in 2013, which compared the growth and phenology of individuals from two seed sources (field and 2012 common garden) to test for maternal effects more directly, revealed an effect of seed source - but not in the expected direction. Original weedy-wild growth rate differences were not recaptured with the field collected seed, but were

recaptured for common garden sourced seed; flowering time differences between weedy and wild were also not observed, in contrast to the previous year.

2.4.1 Evidence for Faster Growth and Earlier Reproduction in Weedy Sunflowers

To better capitalize on the highly favourable, but temporary, conditions in agricultural fields, weedy sunflowers may have traded-off stress tolerance in favour of growth. In support of this hypothesis, we observed that weedy seedlings grew faster than wild seedlings in the 2012 common garden. For the majority of population pairs, weedy seedlings achieved a larger biomass on average than wild seedlings by the final measurement date (Figure 2.6). This finding is consistent with previous research on the evolution of weediness in *H. annuus*. Growth-defense trade-off experiments performed by Mayrose et al. (2011) revealed that, in comparison to wild populations, weedy *H. annuus* populations (four from the USA and three from Europe) were more susceptible to drought, wilting earlier and surviving fewer days after water was withheld. Weedy individuals were also more palatable to a generalist herbivore (*Trichoplusia ni* (Hübner, 1803)) in leaf choice bioassays and had higher rates of secondary and tertiary infection with a fungal pathogen (*Botrytis cinerea* Pers.), though the difference in fungal infection was not significant. When seedlings were grown under benign greenhouse conditions, weedy seedlings showed a higher growth rate, but this was not statistically distinguishable from that of wild seedlings. However, growth was assessed only at the final time point and as the change in height in over time; in our experience, sunflower seedlings of the same height may have very different biomasses, and so height comparisons may miss much of the actual size variation. In a similar study of both Australian ($n = 5$) and U.S. ($n = 4$) weedy *H. annuus* populations, Koziol et al. (2012) also found a marginally significant trend ($p = 0.05$) to higher aboveground biomass (as final size) in weedy versus wild individuals under well-watered conditions. Weedy plants also had coarser roots, with a smaller specific root length, and were more sensitive to drought than wild plants, wilting to a greater extent when water was withheld. Across populations, drought tolerance and growth rate (as either plant biomass or mean root diameter) were negatively correlated. Finally, in a comparison of Argentinean *H. annuus* biotypes, Presotto et al. (2017) found a trade-off between growth and defoliation tolerance, and also drought tolerance. There was only a single agrestal population tested in an irrigated, outdoor common garden, but according to a three-parameter logistic model, the growth rate (in aboveground biomass or height) for this population was higher than for the two ruderal (i.e., wild) populations. However, agrestals suffered a greater reduction in fitness components (e.g. inflorescence number, seeds per inflorescence, seed weight, etc.) compared to ruderals upon simulated herbivory or water limitation. Thus, previous

research has suggested that growth rates may be higher for weedy sunflowers as a result of growth-tolerance trade-offs, but maternal effects, source population latitude and other potentially confounding effects were not accounted for. After controlling for these factors in this study, we can confirm that weedy North American sunflowers do indeed show a shift to faster growth.

The one exception to this finding was Iowa 1; at this location (i.e., for this population pair), the trend was reversed and wild seedlings grew faster than weedy ones. For Iowa 1, weedy seedlings grew very slowly for four out of five maternal families, leading to a low growth rate overall, quite different from that observed for other weedy populations at similar latitudes (compare to, e.g., Iowa 2 or South Dakota 2 in Figure 2.6). This may have been due to poor seed quality; though seeds were not overly small, germination rates were low and seedlings were also difficult to transplant successfully. In contrast, wild seedlings from Iowa 1 grew faster than those from any of the other surveyed wild populations and achieved an average final biomass similar to that of the largest seedlings (i.e., those from weedy populations from Missouri and South Dakota 2). Comparing the site characteristics of the source populations, Iowa 1 was the only wild population located in a wetland (the Wolf Creek Wetland Management Area). The other wild sites were typically xeric by the end of the growing season; I observed dry, dusty conditions and cracked soils at most locations. Thus, soil moisture was likely much more abundant at Iowa 1 compared to the other wild sites. This may be an important difference: under the trade-off hypothesis, agricultural weeds are able to maximize their growth rates owing to the relatively benign conditions in crop fields (Mayrose et al. 2011), where competition for nutrients and water are limited compared to wild habitats. It may be that under conditions of heightened resource availability, such as when water or nutrients are no longer limiting, wild sunflower populations also experience selection for faster growth rates. In the literature, resource rich environments, such as agroecosystems, are often recognized as favouring plant species with the potential for rapid growth (e.g. Chapin 1980; Coley et al. 1985; Lambers and Poorter 1992). Wetlands can also be very seasonal environments, with alternate periods of flooding and drying (e.g. Casanova and Brock 2000; Brock et al. 2003; Van Der Valk 2005), and intense competition among plant species is common, especially during the drying period (Merlin et al. 2015). Thus, in the Iowa 1 wetland site, wild sunflowers that are better able to complete their life cycle quickly, while conditions are most favourable, may be at a selective advantage. In conclusion, Iowa 1 presents an interesting exception: it suggests that, under the right circumstances, wild sunflower populations may also experience selection for rapid growth.

Faster seedling growth was generally associated with an earlier transition to flowering, and this translated into an effect of sunflower type (i.e., weedy versus wild) when Iowa 1 was excluded from the analysis. Weedy populations flowered earlier than paired wild populations for seven out of nine locations; Iowa 1 and North Dakota were the exceptions (Figure 2.7). As discussed, Iowa 1 represents an unusual case as the wild population was uniquely located in a wetland; though the weedy-wild trend was reversed for this location, faster seedling growth was correlated with earlier flowering as found at other locations. Earlier flowering may be beneficial for plants growing in habitats experiencing regular disturbance, such as agroecosystems and wetlands, if precocious reproduction allows for greater seed production before disturbance occurs (Barrett 1988; Basu et al. 2004). In North Dakota, the wild population initiated flowering before the weedy population, despite faster weedy seedling growth. The North Dakota weedy population was phenotypically distinct among all sampled populations and closely resembled crop sunflower, with plants having minimal branching, shorter stature and fewer, larger inflorescences with much heavier seeds. Seeds averaged $16 \text{ mg} \pm 2 \text{ mg}$ (SE) for this population versus a global mean (across all populations except North Dakota) of $7 \text{ mg} \pm 0.4 \text{ mg}$. Given our selection criteria for sampled populations (no crop sunflower within a 1 km radius), offspring from this population should not be F1 crop-wild hybrids, but they do seem likely to represent a recent hybridization event. We grew two European weedy populations (of mixed crop-wild origins) in the 2012 common garden for comparison purposes, and both had larger achenes, averaging $21 \text{ mg} \pm 3 \text{ mg}$ for Italy and 108 mg (single bulk measure, no SE calculated) for France, as well as reduced branching. In wild sunflower, the main stem inflorescence usually flowers first, but it is typically no larger than the inflorescences that later develop on branches. In contrast, in weedy North Dakota individuals, the main stem inflorescence was much larger than any subsequent inflorescences. Development may have taken longer for these large inflorescences, leading to a delay in flowering for weedy North Dakota individuals. Though we did not track bud development in detail, main stem inflorescences did take longer to progress through flowering for weedy sunflower ($44 \text{ days} \pm 2 \text{ days}$) than wild sunflower ($33 \text{ days} \pm 2 \text{ days}$) in North Dakota, likely owing to larger inflorescence and/or seed size. Apart from in North Dakota, which may have been unique owing to a high proportion of cultivar alleles in the weedy population, faster growth was typically associated with earlier flowering.

Sunflower showed tremendous phenotypic variation across the sampled range. Much of the morphological variation we observed in field populations (e.g. Figure 2.4) was likely due to phenotypic plasticity, owing to environmental variability among sites (Figure 2.3). However, some differences, such as latitudinal clines in size and flowering time, persisted in the common garden. The number of days

until flowering decreased with increasing latitude (Figure 2.7), and in the mixed model, location explained the majority of variance in flowering time. Similar latitudinal clines have been previously reported for *H. annuus* (Blackman et al. 2011; McAssey et al. 2016), among other species (e.g. Koornneef et al. 2004; Bohlenius et al. 2006; Zhang et al. 2008), and represent an adaptive response to growing season length, which shortens with increasing latitude, necessitating an earlier transition to flowering (often at a smaller size). In a growth chamber experiment that manipulated day length, Blackman et al. (2011) compared *H. annuus* populations collected across a latitudinal transect from Manitoba, Canada to Texas, USA. The interaction of day length treatment and latitude explained over 80% of the variation in flowering time, with three distinct photoperiod responses observed (day neutral, long- and short-day); however, there was also variability within populations under specific treatments (see Figure 1b in Blackman et al. 2011), and this variability decreased with increasing latitude. We observed a similar pattern here, with greater variation among individuals in days to first flower for southern versus northern populations; weedy-wild differences were also more pronounced for southern populations. This suggests that northern populations, which are more constrained by season length, may have less flexibility in the onset of flowering. With a longer growing season, there may be more flexibility to adapt flowering time to other environmental cues (apart from photoperiod), such as light intensity or quality, and temperature (Koornneef et al. 2004), as well as to habitat type. Thus, despite strong latitudinal constraints on flowering, weeds evolved earlier flowering at many study locations, potentially enhancing their reproductive success in short-season crops.

2.4.2 Effects of seed provisioning and seed source

Utilizing field collected seed, such as we did in the 2012 common garden, raises the possibility that observed differences among individuals may not be genetically-based, but rather due to maternal effects. For example, differences in the quality of the maternal environment may impact the provisioning of seeds (Larios and Venable 2015), with subsequent effects on seedling germination, growth and establishment (Zas et al. 2013). In both of our common garden experiments, seed weight varied among populations: in 2012, average seed weight ranged from 5.7 mg (Kansas wild) to 17 mg (North Dakota weedy), but most populations fell between 6.5 mg and 8 mg. In 2013, common garden sourced seeds were slightly larger than field collected ones (10.6 mg \pm 0.6 mg versus 8.8 mg \pm 0.4 mg); again, the weedy North Dakota population had by far the largest seeds. For most locations and across both experiments, weedy seeds weighed slightly more than wild seeds, potentially confounding the weedy-wild comparison had we not accounted for seed weight as a covariate. As previously reported for a wide variety of plant species (e.g. Houssard and Escarré 1991; Susko and Lovett-Doust 2000; Parker et

al. 2006; Zas et al. 2013), we found that higher seed mass enhanced germinant growth, with larger sunflower seeds producing larger seedlings in both experiments. Importantly, in our models, the effects of seed weight and sunflower type both influenced growth trajectories; thus, in sunflower, both genetics and environment (of both maternal plants and their offspring) can affect seedling growth. In contrast, seed weight did not explain any variation in flowering time in either experiment, implying that maternal effects may be limited to early in the life cycle. This makes sense intuitively, as the growth of young seedlings often depends on resources stored in the seeds, while as time passes, plant performance becomes increasingly dependent on other factors (e.g. current environmental conditions). In the literature (e.g. see for discussion Roach and Wulff 1987), it has been generally observed that seed mass can affect growth rates early on, but that this effect often dissipates with time, although there are exceptions; for example, in an annual dune plant (*Cakile edentula* (Bigelow) Hook.), seedling growth rate was proportional to initial seed mass, but flower bud formation and fruit maturation also occurred earlier for larger seeds (Zhang 1996). In conclusion, we found evidence for effects of seed size on sunflower seedling growth, but not flowering time; weedy-wild differences in growth persisted after accounting for variation in seed size and are therefore likely genetically based.

In a follow-up common garden in 2013, we also compared the effects of seed source directly, as we had the opportunity to generate seed via within-population crosses in the 2012 common garden. However, the follow-up experiment included only three mid-transect locations. In retrospect, these locations were not ideal choices, as weedy-wild differences in 2012 were weaker in these population pairs: there was no effect of population type on seedling growth in South Dakota 1 (Figure 2.8), and muted (South Dakota 2) or reverse (North Dakota) effects of type on flowering time (Figure 2.9). Regardless, patterns from 2012 (in growth and flowering time) were generally not recaptured in 2013 for field collected seed. Although seedlings were grown in the same glasshouse chamber in both years, environmental variables such as light availability and temperature likely varied between years; field experiments were carried out at different sites on the UBC campus, where soil type, nutrients and moisture availability also differed. If genotypes varied in their response to these environmental factors, these genotype-by-environment (GxE) interactions could explain the different results between years. In studies that have multiple common gardens, either across space or time, it is quite common to observe variability in treatment responses across gardens. For example, in a meta-analysis comparing common garden studies of native and introduced plant populations, Colautti et al. (2009) observed near universal GxE effects of maternal family, population and range (i.e., native versus introduced), as well as for experimental abiotic or biotic stress treatments, in studies with replicated gardens. Thus, GxE effects

represent a plausible explanation for the inconsistent results across years, especially in flowering time, though other possibilities exist.

One such possibility is that the relatively smaller sample of individuals from each population taken in 2013 ($n = 15$ individuals versus $n = 25$ in 2012) was insufficient to consistently recapture population patterns, given the substantial variation seen within populations in 2012. Another is a non-genetic effect of seed age: field seeds were one year older than common garden seeds, and so seed source and seed age are confounded. If maternal effects had been driving weedy-wild differences in the 2012 garden, we would expect to recapture 2012 patterns with the field seed but not the common garden seed in 2013, when in fact we saw the opposite (common garden seed more closely recaptured 2012 results). Although GxE effects might explain the differing patterns between years, it is less clear how they may explain the inconsistencies between seed sources in seedling growth, as the same populations and families were used. An effect of seed age could explain this pattern, however, if seed quality was affected differently by aging in weedy versus wild populations. Sunflower seed quality is known to decline over time in storage, with decreasing oil content for example, even if humidity is kept low (e.g. El-Maarouf-Bouteau et al. 2011; Abreu et al. 2013). While all of our seeds were stored in airtight containers with silica gel as a desiccant, germinability and physiological quality of the seeds are both expected to decline over time. However, under similar storage conditions, Seiler (2010) observed high viability in seeds stored for 20 years, suggesting there should be little difference in seeds stored for a single year and we did not observe differences in germination between seed sources, though we forced seeds by breaking dormancy via scarification. Thus, any effects of seed age would be mediated through quality factors; if weedy seeds declined more in quality than wild seeds during storage, then this could explain their slower growth rates. In contrast to Baker's (1974) "ideal weed", which should have high seed longevity, this would imply that weedy sunflowers somehow invest in lower quality seeds than wild sunflowers, that are more prone to degradation, perhaps akin to adult plants investing less in stress tolerance traits, though this is highly speculative. Therefore, while the conflicting results in our two common gardens are difficult to explain, some combination of GxE effects and seed age effects may be involved, though further work would be needed to validate an effect of seed age.

2.4.3 Conclusions and Future Directions

In our study of paired weedy and wild sunflower populations collected along a latitudinal transect, we found faster growth in weedy seedlings under favourable greenhouse conditions. Compared to earlier work in sunflower on life history trade-offs, our study had the advantage of working

with paired populations (to control for other environmental differences not related to weediness) and using seed weight as a covariate (to control for maternal effects). Thus, we believe growth differences represent an adaptive response to the unique challenges of agroecosystems; weedy sunflowers may achieve higher growth rates through decreasing abiotic and biotic stress tolerance traits, as suggested by previous work. We followed seedlings to reproductive maturity in order to ascertain if the initiation of flowering had also been accelerated in weeds, as expected given the short growing season of many modern crops. Overall, faster growth was correlated with earlier reproduction. We were unable to determine, however, if this advancement in phenology led to enhanced plant fitness, as many individuals did not complete flowering. This was for two reasons: first, most southern populations were mismatched to the climate in Vancouver and were only part-way through flowering when fall weather (colder temperatures and heavy rain) arrived. Secondly, more than a quarter of plants became infected with *Sclerotinia sclerotiorum* at the onset of flowering. Further work is therefore needed to better understand how phenological differences affect overall plant performance, as well as to disentangle effects of seed age from complex genotype-by-environment interactions in our second common garden, which manipulated seed source to test more directly for maternal effects. Lastly, a single wild population from our collections was located in a wetland; showing faster growth and earlier reproduction than the other wild populations (which came from drier sites), the wetland population behaved more akin to the weedy sites. This suggests that, apart from agricultural fields, there may also be conditions in the wild (e.g. high resource availability) that select for similar life history shifts.

Table 2.1: Description of wild sunflower populations selected for study in the common garden, including the date sampled, elevation of the site and distance to the closest road, crop infested (if weedy), spatial pattern, estimates of population size and approximate flowering stage.

Location	Type	Date Sampled (2011)	Elevation (m)	Crop	Distance to Nearest Road (m)	Spatial Pattern	Population Size ($\times 10^3$ m ²)	Population Size (# individuals)	Flowering Stage
Iowa 1	weedy	Oct 2	328	soybean	20	intermittent	18	1,000	long finished
	wild	Oct 2	329	<i>na</i>	0	mosaic	30	500	recently finished
Iowa 2	weedy	Oct 3	299	corn	10	continuous	1,200	7,500	recently finished
	wild	Oct 3	290	<i>na</i>	100	intermittent	350	1,000	mid-flowering
Kansas 1	weedy	Oct 5	599	sorghum	20	mosaic	10	2,500	end-of-flowering
	wild	Oct 6	550	<i>na</i>	5	intermittent	10	1,750	end-of-flowering
Kansas 2	weedy	Oct 7	828	sorghum	5	continuous	3.5	500	recently finished
	wild	Oct 6	709	<i>na</i>	5	continuous	160	700	end-of-flowering
Manitoba	weedy	Sept 25	511	wheat	10	continuous	500	1,000	long finished
	wild	Sept 26	507	<i>na</i>	5	intermittent	3	200	long finished
Missouri	weedy	Oct 4	319	soybean	35	mosaic	640	6,000	end-of-flowering
	wild	Oct 4	324	<i>na</i>	5	mosaic	32	1,000	mid-flowering
North Dakota	weedy	Sept 28	605	corn	20	mosaic	21	2,000	long finished
	wild	Sept 28	516	<i>na</i>	0	intermittent	15	33,000	recently finished
South Dakota 1	weedy	Sept 30	577	corn & soybean	5	mosaic	80	700	recently finished
	wild	Sept 29	549	<i>na</i>	15	mosaic	50	800	recently finished
South Dakota 2	weedy	Oct 1	419	corn	20	intermittent	30	250	long finished
	wild	Oct 1	381	<i>na</i>	60	continuous	40	1,500	recently finished
Saskatchewan	weedy	Sept 27	626	wheat	10	mosaic	1	200+	long finished
	wild	Sept 27	585	<i>na</i>	5	mosaic	160	2,000	long finished

Table 2.2: Annual climatic data for the study populations , as obtained from the software ClimateNA (Wang et al. 2016), as scale-free estimates for the normal period 1961-1990.

Location	Type	MAT*	MWMT	MCMT	TD	MAP	MSP	AHM	SHM	DD_0	DD5	DD_18	DD18	NFFD	bFFP	eFFP	FFP	PAS	EMT	EXT	Eref	CMD	MAR	RH
IA 1	A ^x	9.5	24.5	-7.3	31.8	722	463	27	52.9	629	2862	3642	560	197	120	282	162	50	-34.9	41.6	936	276	14.2	57
	W	9.4	24.4	-7.4	31.8	725	467	26.8	52.3	634	2854	3653	556	196	120	282	161	50	-34.9	41.6	934	271	14.1	57
IA 2	A	10.8	25.3	-5.5	30.8	804	512	25.9	49.5	462	3110	3277	675	212	113	286	173	36	-32.5	41.9	963	228	14.1	59
	W	10.9	25.3	-5.3	30.7	810	521	25.8	48.6	455	3128	3256	682	212	113	287	173	34	-32.4	42.1	972	227	14.3	58
KS 1	A	11.7	26.4	-3	29.4	590	389	36.8	67.7	315	3240	3032	759	212	114	285	171	16	-32.4	43.5	1077	527	15.6	53
	W	11.3	26.2	-3.8	30.1	612	401	34.9	65.4	361	3171	3145	734	208	115	283	167	20	-33.3	43.5	1067	500	15.6	52
KS 2	A	11	25.5	-3	28.6	537	368	39.2	69.3	333	3046	3170	656	197	119	279	160	16	-33.8	42.9	1079	576	15.9	50
	W	10.9	25.5	-3.6	29.1	594	408	35.2	62.7	359	3037	3209	653	202	118	281	163	17	-33.6	42.6	1051	487	15.9	52
MB	A	2.9	19.5	-16.8	36.4	466	313	27.6	62.3	1702	1761	5639	155	157	140	258	118	96	-43.9	39.4	681	305	12.7	55
	W	2.7	19.4	-17	36.5	467	315	27.2	61.7	1725	1741	5680	149	157	141	258	117	97	-43.9	39.2	675	297	12.8	55
MO	A	11.1	25.1	-4.4	29.5	922	571	22.9	44	393	3134	3147	664	212	115	288	174	34	-31.4	41.6	966	161	14.6	59
	W	11.1	25.1	-4.4	29.5	923	571	22.9	43.9	393	3132	3148	663	212	115	288	174	34	-31.4	41.6	965	161	14.7	59
ND	A	4.5	20.9	-13.7	34.5	423	299	34.2	69.7	1338	1946	5121	227	165	138	263	125	59	-41.5	40.4	723	358	13.1	54
	W	5.2	21.6	-13	34.6	428	300	35.5	72	1240	2100	4907	278	168	136	266	130	54	-41	41.2	757	386	13.1	54
SD 1	A	7.2	23.4	-10.3	33.7	495	327	34.7	71.5	940	2445	4333	418	172	131	272	141	53	-38.8	42.7	864	433	14.1	52
	W	7.6	24	-9.7	33.6	460	306	38.2	78.3	876	2521	4221	463	177	129	273	144	50	-38.5	43.1	873	478	14.2	52
SD 2	A	8.3	23.8	-9	32.8	646	419	28.3	56.8	788	2638	3995	474	182	125	277	152	56	-36.6	41.4	899	302	14.1	56
	W	8.9	24.4	-8.3	32.8	646	422	29.2	57.9	716	2774	3824	534	188	122	279	157	49	-36	42.1	931	332	14.2	56
SK	A	3.6	19.9	-15	34.9	432	286	31.5	69.7	1496	1808	5387	173	166	136	261	126	94	-43.5	38.8	673	332	13.1	57
	W	3.7	20.1	-15	35.1	416	276	32.8	72.6	1494	1825	5372	180	166	136	261	125	89	-43.4	38.9	678	346	13.1	57

*MAT = mean annual temperature (°C), MWMT = mean warmest month temperature (°C), MCMT = mean coldest month temperature (°C), TD = continentality (°C), MAP = mean annual precipitation (mm), MSP = mean summer (May to Sep) precipitation (mm), AHM = annual heat to moisture index, SHM = summer heat to moisture index, DD_0 = degree days below 0°C, DD5 = degree days above 5°C, DD_18 = degree days below 18°C, DD18 = degree days above 18°C, NFFD = number of frost-free days, bFFP = Julian date on which the frost-free period begins, eFFP = Julian date on which the frost-free period ends, FFP = length of frost-free period, PAS = precipitation as snow (mm), EMT = extreme minimum temperature (°C), EXT = extreme maximum temperature (°C), Eref = Hargreaves reference evaporation, CMD = Hargreaves climatic moisture deficit, MAR = mean annual solar radiation (MJ m⁻² d⁻¹), RH = monthly average relative humidity (%)

*A = agricultural weed population, W = wild population

Table 2.3: Size variation among seedlings (n = 227) used to establish the relationship between non-destructive measurements and biomass.

	Height (cm)	Number of Leaves	Length of Largest Leaf (cm)	Half With of Largest Leaf (cm)	Biomass (g)
Maximum	96.5	25	17.3	7.6	18.54
Minimum	1	2	0.4	0.1	0.005
Mean	41.37	9.45	8.09	2.47	2.99
Standard Deviation	25.45	4.26	3.96	1.66	3.63
Standard Error	1.69	0.28	0.26	0.11	0.24

Table 2.4: Description of the plant materials generated in the 2012 common garden that were included in the 2013 maternal effects common garden, including the identity of the (field-collected) maternal families of the maternal and paternal individuals used for each cross.

Location	Type	Family	Mother's Family	Father's Family
North Dakota	weedy	117	33	40
		122	11	16
		125	16	11
		126	26	16
		140	40	33
North Dakota	wild	144	38	20
		145	20	38
		148	33	38
		153	6	20
		156	27	38
South Dakota 1	weedy	53	34	28
		56	28	34
		59	37	40
		62	40	38
		69	38	28
		100	20	4
South Dakota 1	wild	106	6	20
		108	21	6
		113	11	4
		115	4	11
		117	4	11
South Dakota 2	weedy	1	4	7
		3	7	4
		7	32	27
		8	27	32
		24	30	7
		26	27	38
South Dakota 2	wild	31	23	27
		36	14	27
		40	38	23
		42	30	14
		44	30	14

Figure 2.1: North American range of *Helianthus annuus* (based on Rogers et al. 1982) and collection locations of populations included in this study. Each agricultural-weed population was paired with a nearby population of wild sunflowers in a non-agricultural area, to isolate changes due to weediness per se from other local adaptations. Location names are for the pair. Note that the Kansas 1 pair was excluded from analysis.

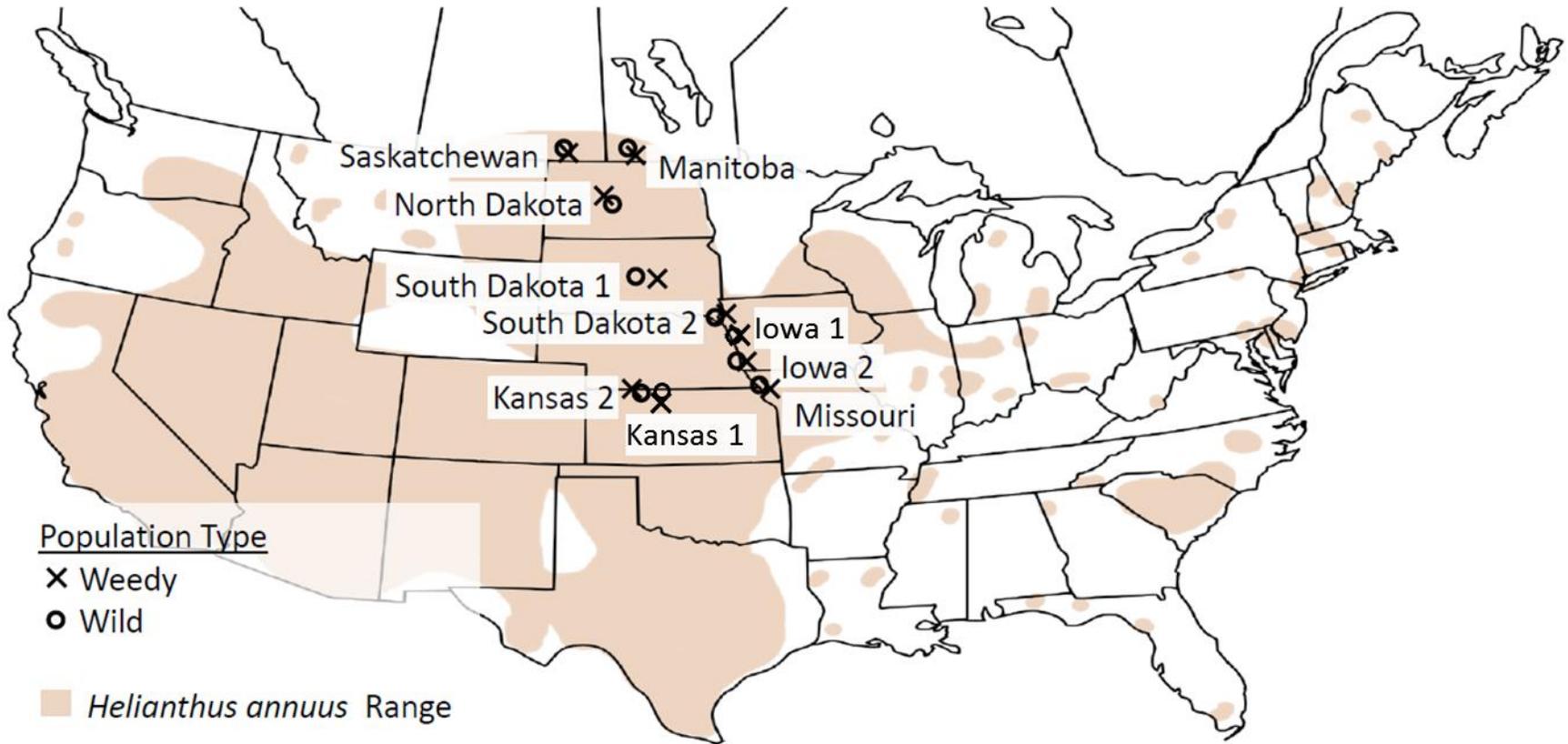


Figure 2.2: Relationship between sunflower seedling biomass and four non-destructive measurements of plant size : height (a), the number of leaves (b), half width of the largest leaf (c) and length of the largest leaf (d). Seedlings (n = 227) represented all study populations and a broad range of maternal families within each. Though seedlings were harvested at different times, all were grown under standardized conditions in the same glasshouse on the UBC campus.

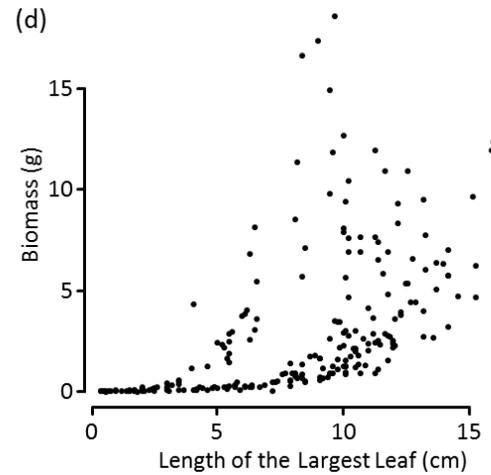
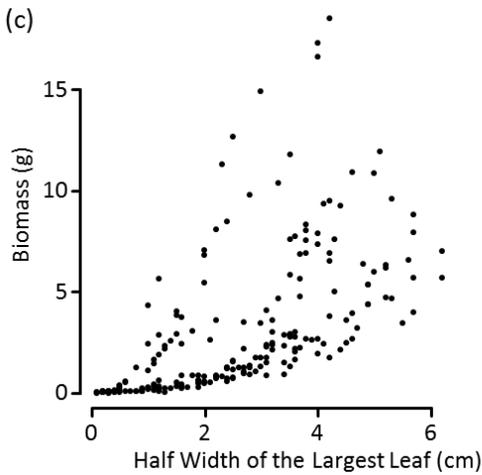
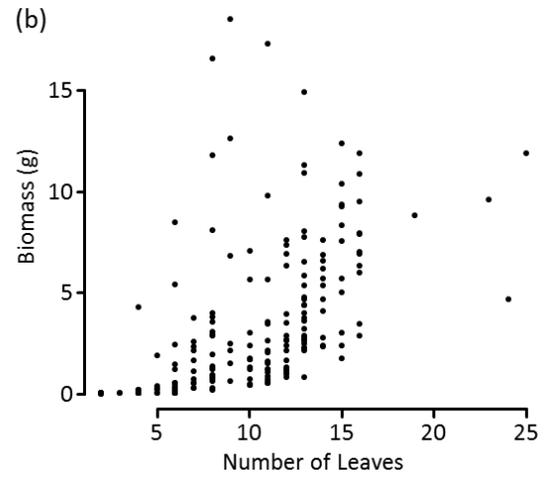
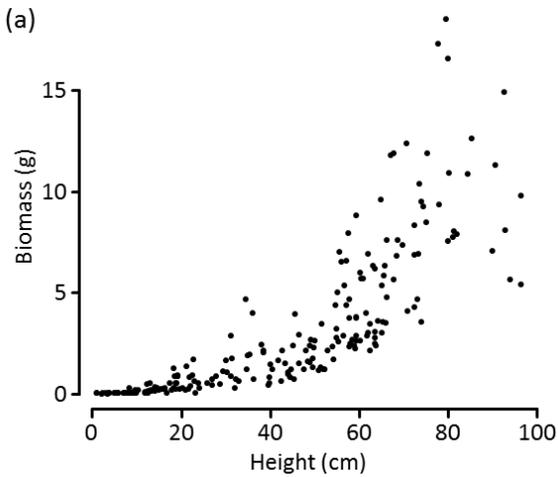
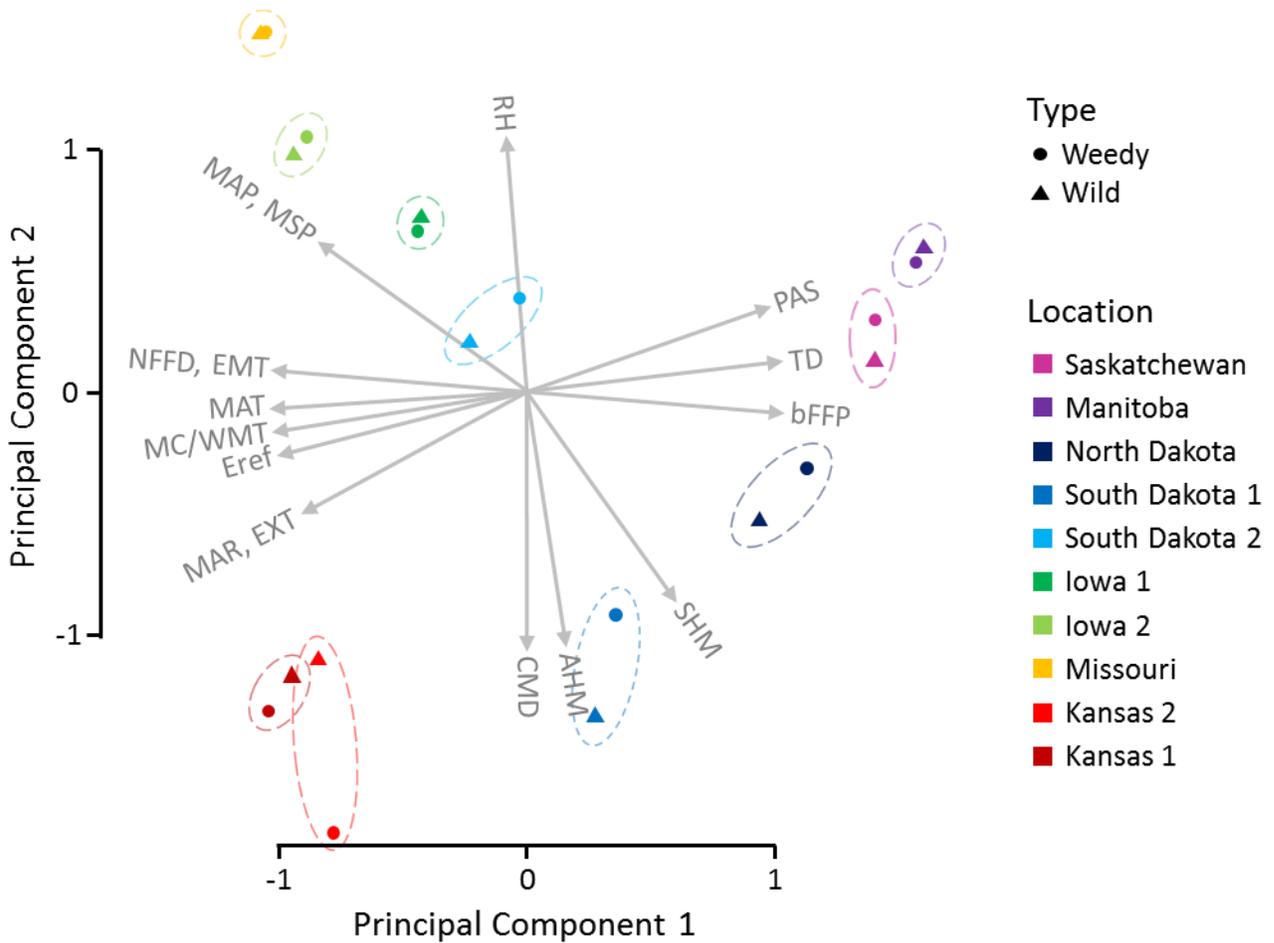


Figure 2.3: Principal components analysis (PCA) of 23 annual climate variables obtained as scale-free point estimates for each population for the 1961-1990 normal period. The first PC axis explained 75.8% of the variation in the data and was correlated with temperature, with more northern populations having higher PC1 values; the second axis explained 21.4% of the variation and was correlated with moisture availability. Populations are coloured by location, while shapes indicate population type; dashed ellipses were drawn by hand to visualize pairings. Overlapping arrows for climate variables have been removed for the sake of clarity.



Climate variables: MAT = mean annual temperature (°C), MWMT = mean warmest month temperature (°C), MCMT = mean coldest month temperature (°C), TD = continentality (°C), MAP = mean annual precipitation (mm), MSP = mean summer (May to Sep) precipitation (mm), AHM = annual heat to moisture index, SHM = summer heat to moisture index, NFFD = number of frost-free days, bFFP = Julian date on which the frost-free period begins, PAS = precipitation as snow (mm), EMT = extreme minimum temperature (°C), EXT = extreme maximum temperature (°C), Eref = Hargreaves reference evaporation, CMD = Hargreaves climatic moisture deficit, MAR = mean annual solar radiation ($\text{MJ m}^{-2} \text{d}^{-1}$), RH = monthly average relative humidity (%).

Figure 2.4: Box-plots of plant height and number of inflorescences for individual sunflowers sampled in the field ($n = 40$ per population) summarized by collection location (i.e. the name of a population pair) and population type (weedy or wild). Locations are arranged from left-to-right in order of increasing latitude. Shading indicates significant differences between weedy and wild for a given location, according to Tukey HSD post-hoc tests following an ANOVA.

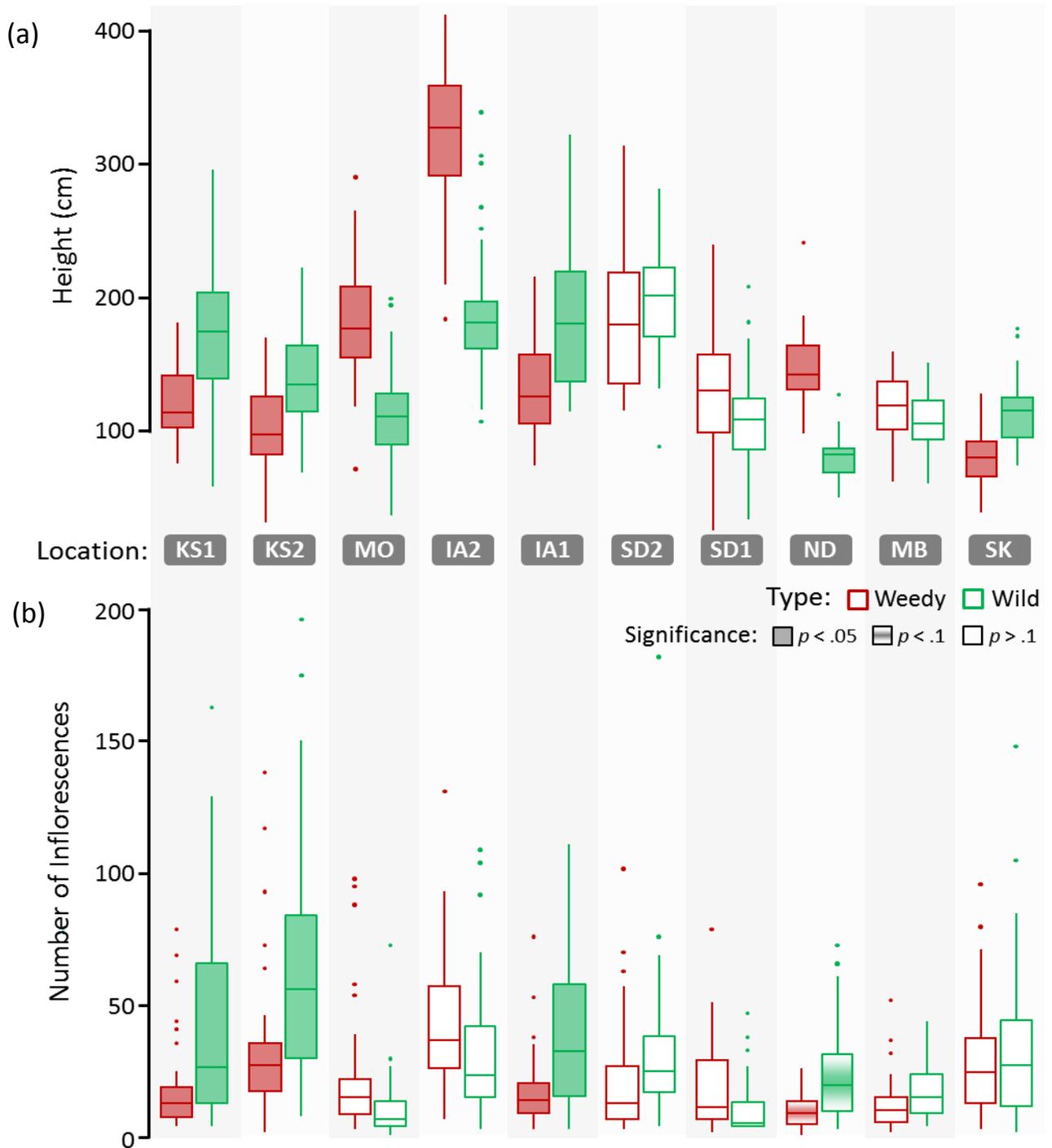


Figure 2.5: Growth of four representative sunflower seedlings in the 2012 common garden over time in untransformed biomass (a) and log-transformed, zeroed biomass (b). Biomass was obtained from measurements of seedling height and the number of leaves using a previously established linear model ($R^2 = 0.9$, $n = 227$ plants). Note that y-axes differ, owing to the large variation among individuals in size. Time is measured in days since initiation of the greenhouse experiment, beginning at day zero; seedlings germinated at different times, hence first measurements occur on different days for different seedlings.

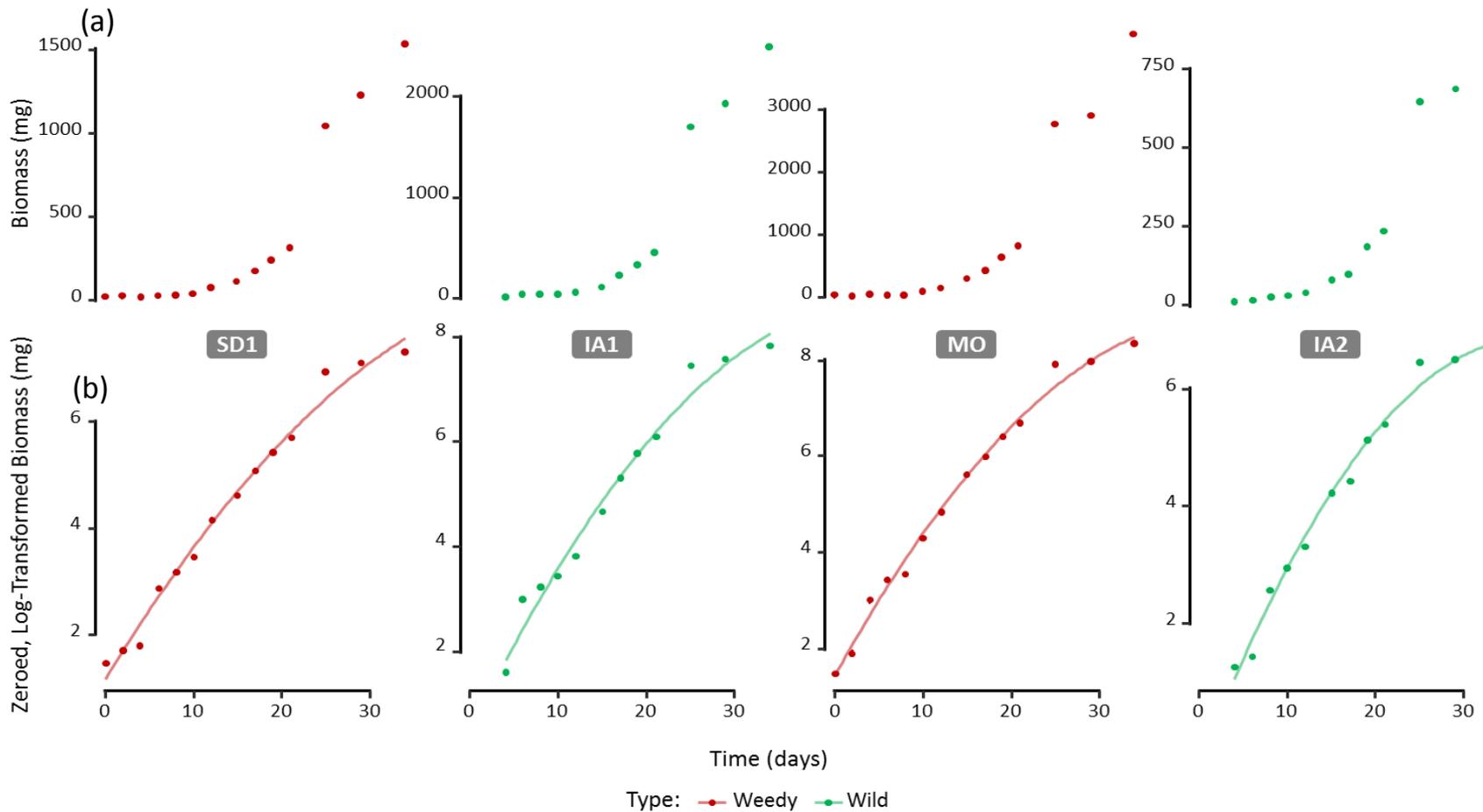


Figure 2.6: Change in biomass over time by location and population type (weedy or wild) for the 2012 common garden. Biomass was obtained from measurements of seedling height and the number of leaves using a previously established linear model ($R^2 = 0.9$, $n = 227$ plants). In order to better visualize within-location trends, y-axes differ across locations. Trend lines were obtained using local regression (LOESS); shading illustrates 95% confidence intervals. Points represent mean seedling biomass at a given time point, with five maternal families and an average of 33 individuals (range: 26 to 35) per population.

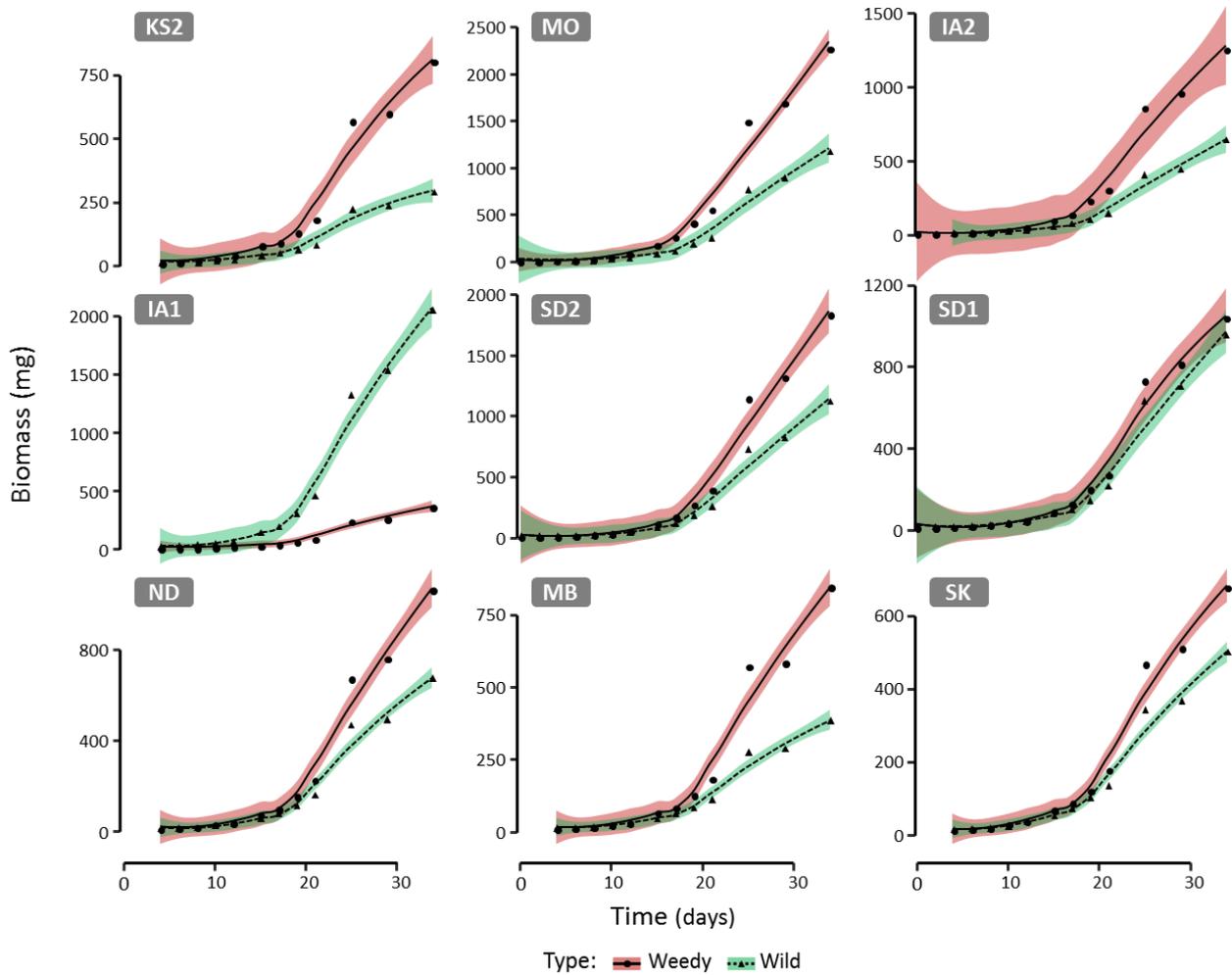


Figure 2.7: Days until first flower by location and population type (weedy or wild) for the 2012 common garden. Data points represent means \pm standard errors, with 25 individuals per population (five from each of five maternal families); two populations (Iowa 2 and Iowa 1) had fewer individuals (22 and 24, respectively) as some plants did not flower. Note that the y-axis does not begin from zero in order to better visualize within-location trends. Day zero was set as the scarification date for each individual.

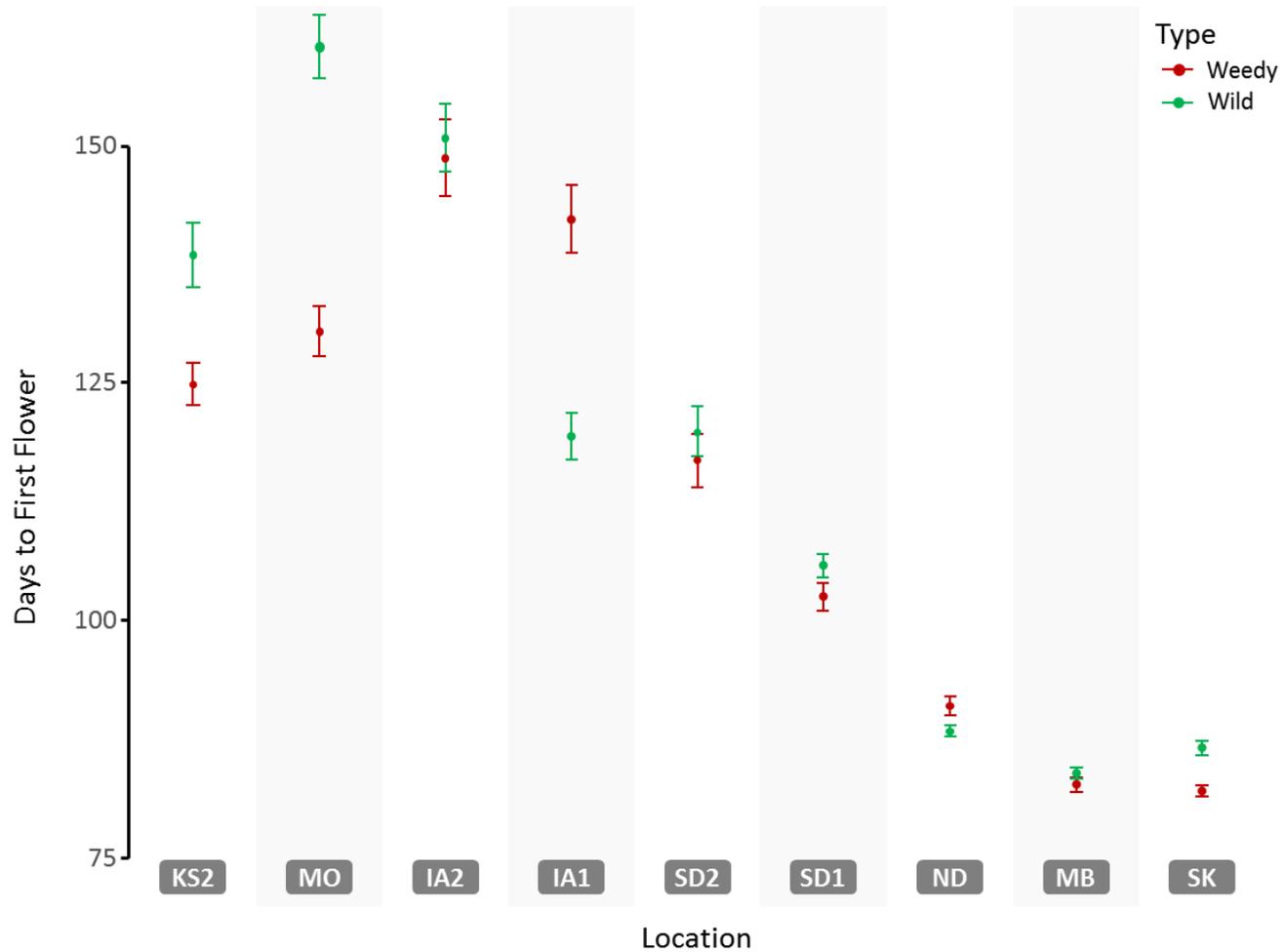


Figure 2.8: Change in biomass over time for seedlings derived from field collected seed (a) or from seed produced in the 2012 common garden (b) in the follow-up 2013 common garden. Separate growth curves are presented for locations and population types (weedy or wild). Biomass was obtained from measurements of seedling height and the number of leaves using a previously established linear model ($R^2 = 0.9$, $n = 227$ plants). In order to better visualize within-location trends, y-axes differ across locations. Trend lines were obtained using local regression (LOESS); shading illustrates 95% confidence intervals. Points represent mean seedling biomass at a given time point, with five maternal families and 24-25 individuals per unique combination of population and seed source.

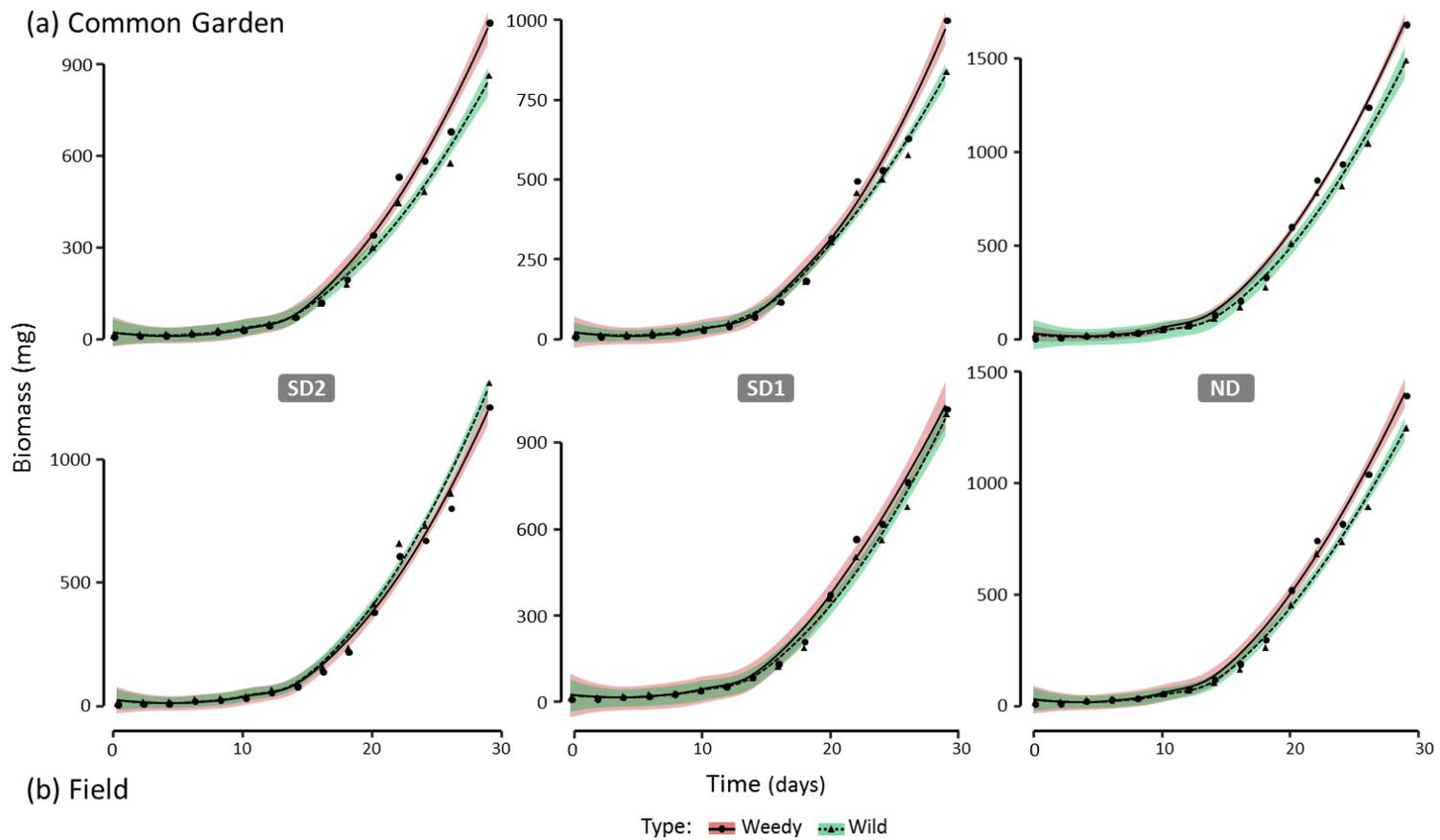
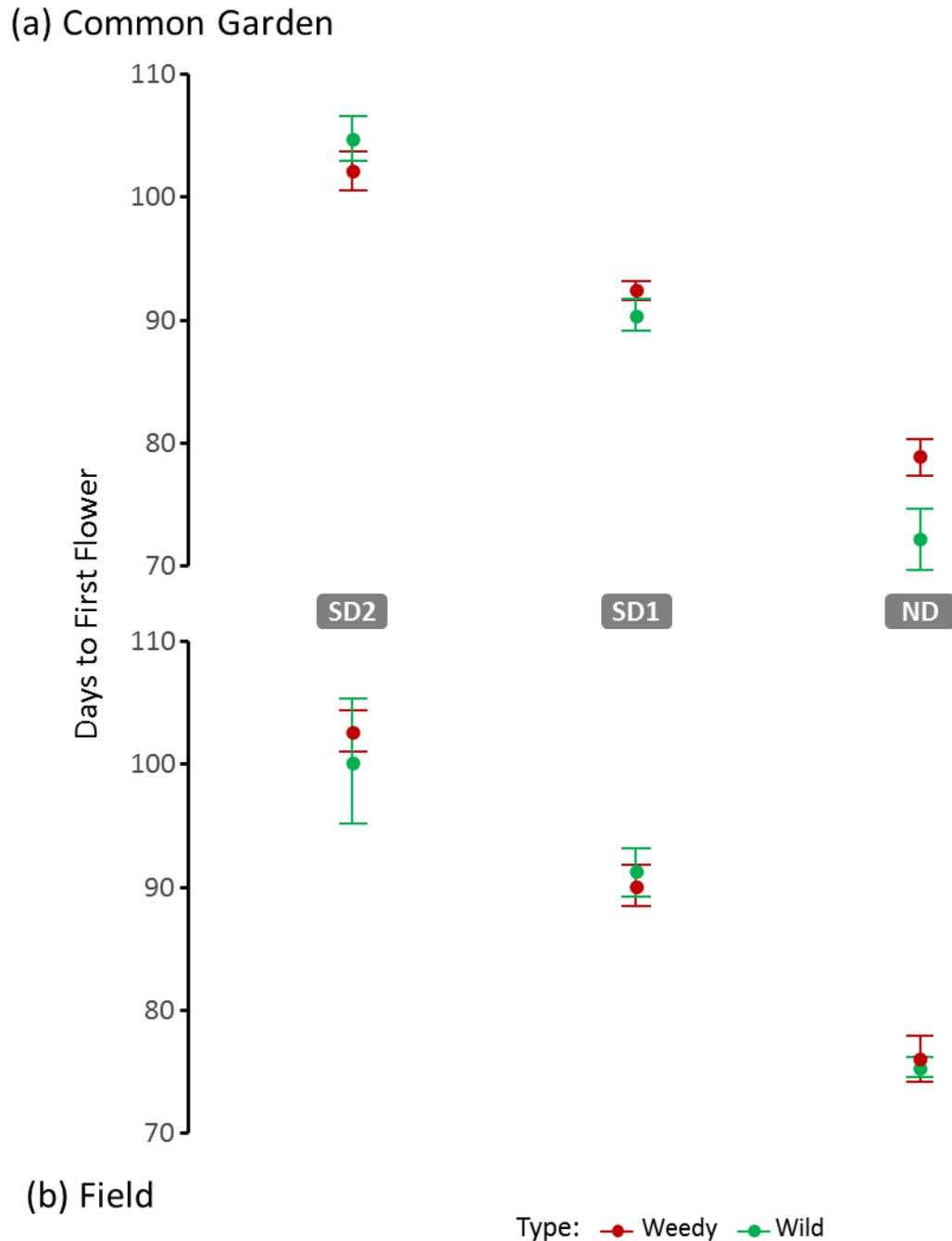


Figure 2.9: Days until first flower by location and population type (weedy or wild) for the 2013 common garden. Results are presented separately for individuals derived from seed generated in the 2012 common garden (a) and from field collected seed (b). Data points represent means \pm standard errors, with 15 individuals per unique combination of population and seed source (i.e., three plants from each of five maternal families). Note that the y-axis does not begin from zero in order to better visualize within-location trends. Day zero was set as the scarification date for each individual.



Chapter 3 : Parallel Genomic Signatures of Divergence between Agricultural-Weed and Non-Agricultural, Wild Populations of Sunflowers

3.1 Introduction

It is now clear that evolutionary change can happen very rapidly, occurring at timescales fast enough to impact even the ecological dynamics of species (e.g. Holt 2005; Carroll et al. 2007). Rapid evolutionary change may be most common under the strong selective regimes invoked by changing environmental conditions (Neuhauser et al. 2003) or when an organism colonizes a novel habitat. For example, in a survey of the literature on introduced plant and animal species, Buswell et al. (2011) found that a majority of species (70%) showed a change in at least one morphological trait over time. Additionally, when the same novel environment is colonized by multiple populations or species, parallel evolution may result in similar phenotypic changes (see e.g. Wood et al. 2005; Arendt and Reznick 2008; Elmer and Meyer 2011; Conte et al. 2012). Such cases of parallel divergence, when occurring in geographical isolation, provide strong support for the role of natural selection, as neutral evolutionary processes such as genetic drift are unlikely to be responsible for repeated adaptive phenotypes (Elmer and Meyer 2011). Thus, invasive or weedy species in which multiple populations invade a novel range or habitat type can provide excellent test cases for studying not only the role of evolution (versus pre-adaptation or phenotypic plasticity) in the response to environmental change, but also the repeatability of evolution and its rapidity.

Wild, annual sunflowers (*Helianthus annuus* L.) may provide one such test case. *Helianthus annuus* is an outcrossing diploid ($n = 17$) native to North America, which grows indeterminately and is highly branched. Its typical habitat is heavy-soiled, undisturbed, open grassland (Heiser et al. 1969); however, the species is now more commonly found as an inhabitant of waste places, crop fields, and other human-dominated and disturbed areas (e.g. Smith 1989), suggesting it is disturbance-adapted. A weedy form of *H. annuus* is considered problematic in the Central USA, Southern Canada, and parts of Europe, Asia and Australia (Al-Khatib et al. 1998). This form has been listed as a noxious weed in several US states (e.g. Iowa, Minnesota, Alaska) and may decrease crop yields significantly in agricultural fields. For example, for corn (*Zea mays* L.) and soybean (*Glycine max* (L.) Merr.) fields, infestation with weedy sunflowers has reduced crop productivity by up to 64% and 97%, respectively (Geier et al. 1996; Deines et al. 2004). Common garden experiments indicate that the agricultural weed grows faster than non-agricultural, wild plants under benign conditions, but this comes at a cost of a decreased ability to tolerate drought conditions, greater palatability to insect predators, and lower tolerance of crowding

(Chapter 2, Mayrose et al. 2011). Hence, weedy types may have shifted their life history strategy to one of faster growth at the expense of stress tolerance and herbivore resistance traits. Weedy populations were found to be more closely related (as determined by F_{ST} and $(\delta\mu)^2$) to nearby non-agricultural populations than to other weedy populations (Kane and Rieseberg 2008), implying that sunflowers have colonized and adapted to agricultural environments repeatedly.

Adaptations seen in agricultural-weed populations of sunflower may therefore represent parallel phenotypic evolution, though further work is needed to confirm this finding. Classic examples of parallel phenotypic evolution include: independent eye reduction in cave-dwelling amphipods *Gammarus minus* Say, 1818 (Jones et al. 1992); parallel latitudinal clines in *Drosophila subobscura* Collin, 1936 wing length (Huey 2000); multiple origins of ecomorphs in island-colonizing *Anolis* lizards (Losos 1998); repeated shifts in life-history traits at high versus low elevation sites in guppies, *Poecilia reticulata* Peters, 1859 (Reznick et al. 1996); and the repeated evolution of a freshwater form from marine threespine sticklebacks, *Gasterosteus aculeatus* L., 1758 (Foster and Baker 2004). In each of these fascinating cases, the same trait or suite of traits has evolved with similar environmental transitions, indicating the same solution to a common problem in each case. Parallel phenotypic changes may have a similar genetic basis or may rely on different genetic changes that produce the same phenotypic results. Both possibilities are seen, for example, in the transition from marine to freshwater living in threespine stickleback, *G. aculeatus* (Foster and Baker 2004). Lateral plating and pelvic spines are reduced in freshwater environments, due to lower predation rates (Barrett and Schluter 2008; Marchinko 2009). Reductions in plate number have been linked to allelic variation at a single locus, *Ectodysplasin-A (Eda)*, and the same low-plate *Eda* alleles have been implicated in most populations (Colosimo 2005). In contrast, reductions in pelvic spines have been linked to different deletions in the *Pituitary homeobox 1 (Pitx1)* gene and to unidentified mutations in non-homologous genes in different populations (Chan et al. 2010). Thus, the repeated parallel phenotypic adaptation of marine sticklebacks to freshwater involves a combination of repeated recruitment of standing genetic variation, different mutations at a homologous gene, and also the involvement of non-homologous genes.

A primary goal of evolutionary biology is to understand how natural selection shapes the genome and the genomic architecture underlying ecologically important traits (Yeaman 2013). Recent advances in high-throughput, next-generation sequencing (NGS) technologies have the potential to elucidate the genetic basis of adaptive traits, especially in non-model organisms (Stapley et al. 2010). Thus far, studies of wild populations have mostly relied on reduced-representation genome sequencing

methods (e.g. Hohenlohe et al. 2010), such as Restriction-site Associated DNA sequencing (RADseq: Miller et al. 2007; Baird et al. 2008) and Genotyping-By-Sequencing (GBS: Elshire et al. 2011), which use restriction-enzymes to reduce genome complexity. While these methods are cost-effective, in comparison with whole genome shotgun (WGS) resequencing, they provide much more limited data and so adaptive loci not in linkage disequilibrium with study markers may be missed. Furthermore, reduced-representation methods come with certain errors and biases, for example genotyping errors due to allele dropout (when a polymorphism at a restriction enzyme cut-site results in a cut failure: Andrews et al. 2016). In order to gain a complete understanding of the number and distribution of genes underlying adaptive divergence, WGS sequencing may therefore be required.

Here, I describe the results of a genome-wide analysis, based on WGS resequencing data, examining the genetic basis of adaptation to contemporary, high-intensity agricultural environments in weedy sunflowers. Complementary to previous common garden work, which identified phenotypic differences in life history in agricultural-weeds, this analysis should permit us to identify any genetic changes that have occurred in parallel among weed populations. Importantly, by looking at replicated agricultural-weed and non-agricultural population pairs, we can isolate changes due to weediness per se, rather than other local adaptations. Furthermore, while genetic drift or population bottlenecks can produce false signatures of selection between populations (Excoffier and Ray 2008), our replicated design should identify only genomic regions that have diverged in parallel, and hence are more likely to be the result of similar selection pressures across populations. My goal was to discover what proportion of the genome may be involved in weed adaptation, the number and location of differentiated regions, and to identify candidate genes potentially underlying weedy trait differences within these regions.

3.2 Materials and Methods

3.2.1 Study Populations and Sample Collection

I collected seeds in the fall of 2011 from each of twenty populations of wild sunflower (*Helianthus annuus*) over a latitudinal gradient (Figure 3.1). Ten populations were located in areas of high-intensity agricultural use and were found competing directly with a cultivated crop of either corn, soybean, wheat or sorghum. Each agricultural-weed (henceforth “weedy”) population was paired with a non-agricultural wild (henceforth “wild”) population collected from a more natural habitat (e.g. grassland, fallow area, etc.) located nearby (mean distance = 28.5 ± 7.6 km), at roughly equal latitudes, to ensure that population pairs experienced similar environmental conditions, such as temperature, rainfall, day length and seasonality. Populations were large (mean $\approx 2,000$ individuals, range = 250 to

10,000+ individuals) and located at least 1 km distant from any cultivated sunflower with which gene flow could occur.

One of the weedy populations (MB1A) was found to have sterile seeds in germination trials. Seeds from the National Plant Germplasm System (NPGS) collections of the United States Department of Agriculture (USDA) were obtained to replace this population; the accession I selected (PI 592327) was collected in 1994 from along the edge of a harvested wheat field located in the same geographic area as the population to be replaced. For this study, I used phenotypic comparisons made in a common garden (Chapter 2), as well as careful evaluation of the collection sites themselves, to guide selection of a subset of representative weedy populations and their respective wild pairs. In the two omitted pairs, one wild population represented a weed/non-weed hybrid, while another occurred in a much more mesic environment than all other populations. Hence, sixteen populations were included in total (eight pairs).

While my initial goal was to use sequence data generated from phenotyped individuals in the common garden (Chapter 2) that were representative of average trait values for each selected population, this was not possible. DNA from these samples were sequenced on older sequencing technology, and some samples did not have enough markers across the genome to conduct genome scans. Therefore, I randomly selected a single sequenced individual from a later mapping study (Chapter 4) from one of the original maternal families evaluated in the common garden for each population for analysis in this study.

We grew seedlings for the mapping study in a glasshouse at the University of British Columbia (UBC) in Vancouver, Canada in summer 2014. We scarified and then germinated seeds on moist filter paper in petri dishes, watering with a 1% solution of plant preservative media (PPM) to reduce microbial contamination. Seedlings were planted as they germinated into 5 cm diameter cones in standard potting soil placed on a glasshouse bench. The soil was thoroughly watered prior to planting and kept moist by misting during seedling establishment. Supplemental lighting was available 16 hours a day.

3.2.2 DNA Extraction and Library Preparation

We collected leaf tissue from newly expanded leaves when seedlings were at the four- to eight-leaf stage. Tissue was placed on dry ice shortly after collection and then stored at -80 °C in the lab. For the majority of samples, I extracted genomic DNA from frozen leaf tissue using a modified CTAB protocol based on Murray and Thompson (1980). Samples that failed to extract cleanly using the CTAB protocol, I extracted with the QIAGEN® DNeasy Plant Mini Kit or a DNeasy 96 Plant Kit (QIAGEN®, Hilden, Germany)

which both use a silica-based approach. Quality checks of DNA samples were performed using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham MA, USA) and by running out samples on an EtBr-agarose gel. DNA was quantified using a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham MA, USA).

High-quality DNA samples were sheared to an average fragment size of 350 bp using a Covaris M220 Focused-ultrasonicator (Covaris, Woburn MA, USA), and 750 ng of sheared DNA was then used to create a paired-end whole-genome shotgun Illumina library for each sample. Library preparation used a custom lab protocol based largely on Rowan et al. (2015), the TruSeq DNA Sample Preparation Guide from Illumina (Illumina, San Diego CA, USA) and Rohland and Reich (2012); importantly, completed adapters were identical to the Illumina TruSeq adapters. As the sunflower genome contains a substantial fraction of highly repetitive sequences derived from the expansion of two retrotransposon families (Staton et al. 2012), we performed a depletion step to reduce the representation of repetitive sequences: Enriched libraries were treated with Duplex Specific Nuclease (DSN; Evrogen, Moscow, Russia) following a modified method based on Shagina et al. (2010) and Matvienko et al. (2013).

Enriched, depleted libraries were purified twice with 1.6 volumes of a solution of paramagnetic SPRI beads (prepared according to Rohland and Reich 2012) before quantification on the Qubit®. We ran libraries on a 2100 Bioanalyzer instrument using a High Sensitivity DNA Analysis Kit (Agilent Technologies, Santa Clarita CA, USA) to determine average fragment size, and assessed molarity on an iQ5 Real Time PCR Detection System (Bio-Rad, Hercules CA, USA). Groups of ten barcoded, randomly-selected libraries from the mapping study (Chapter 4) were then pooled at equal molarity to be sequenced on a single Illumina lane. All libraries were sequenced at the Genome Québec Innovation Center on either an Illumina HiSeq X or HiSeq 2500 instrument (Illumina, San Diego CA, USA).

3.2.2 Bioinformatics Pipeline

Samples for this study were analyzed as part of a larger group (of 323 wild *H. annuus* samples) for the mapping study presented in Chapter 4. We used Trimmomatic (version 0.36: Bolger et al. 2014) to clean the raw data, removing adapters and low-quality bases from each read, before aligning sequencing reads to the *H. annuus* reference genome (XRQ assembly: Badouin et al. 2017) using the BWA-MEM aligner (version 0.7.9a: Li and Durbin 2010) with default parameters. PCR duplicates were marked with Picard (version 2.5, <http://broadinstitute.github.io/picard/>; retrieved on 25 Sept 2017), while potential indels were realigned using GATK (version 3.6: McKenna et al. 2010), again using default parameters. Single nucleotide polymorphisms (SNPs) were identified using FreeBayes (version 1.1.0:

Garrison and Marth 2012), a Bayesian genetic variant detector; as the method implemented in FreeBayes is haplotype-based, working on the literal sequences of reads, it is more robust to local alignment issues than site-based variant detection methods (i.e. samtools).

Low-quality SNPs were removed using cut-offs derived from comparing the distribution of quality metrics obtained for this dataset with that observed for a set of validated SNPs from a SNP-chip (Mandel et al. 2013). To do this, regions surrounding SNP-chip variants were aligned to the XRQ genome using BWA-MEM. Regions with the maximum observed mapping quality (60) were kept to reduce comparison of paralogs. Using base counts derived from samtools mpileup and custom perl scripts, exact variant sites were extracted from SNP-chip regions. These variants were compared to observed SNPs to confirm that the same change was observed in each. SNP-chip variants that were observed in the SNP data were considered validated. We visually compared the distribution of site quality metrics between validated and unvalidated SNPs to determine appropriate cut-offs. The following filters were used: mapping quality score > 20, mean mapping quality of observed alternate alleles > 40, and allele balance at heterozygous sites between 0.4 and 0.6. (Allele balance ranges from 0 to 1, and represents the ratio of reference allele reads to all reads for heterozygous individuals.) Additionally, for the divergence analysis presented here, I excluded loci with missing data for any of the sixteen focal individuals or with a read depth of < 4. Loci on unmapped scaffolds (i.e. those not placed on chromosomes in the reference genome assembly) were also not considered. I used BCFtools (version 1.3: Li et al. 2009a) to filter by read depth and remove unmapped scaffolds, and a custom python script to remove loci with missing data as well as any sites that were invariant across the focal individuals.

3.2.3 Analysis of Linkage Disequilibrium

To inform window-based outlier analyses of divergence between weedy and wild populations, I examined how linkage disequilibrium (LD), i.e. the non-random association of alleles at different loci, decayed with the physical distance between SNPs located on each chromosome. My goal was to enable a more meaningful, data-driven selection of window size for analysis, as this choice is often only loosely justified (e.g. Burke et al. 2010; Rubin et al. 2010) or somewhat arbitrary (e.g. Myles et al. 2008; Turner et al. 2011) in the literature. Recombination rates and LD can vary considerably across the genome (Hey et al. 2004; see Mandel et al. 2013 for patterns in sunflower), but estimating how quickly LD decays on average provides guidance on the minimum window size possible.

Here, I utilized the larger dataset of wild sunflower genomes available from the mapping study (Chapter 4), excluding only 20 hybrid wild-cultivar individuals. An additional 20 low-coverage wild

genomes from an early sequencing project (described in Section 3.2.1) were included as well, for a total of 343 individuals. To remove loci with low heterozygosity, I used VCFtools (version 0.1.15: Danecek et al. 2011) to filter out SNPs with a minor allele frequency (MAF) of < 10%, which is a common cut-off in analyses of linkage disequilibrium (e.g. Mandel et al. 2013). Each chromosome was then analyzed separately, using a random subset of 10,000 SNPs (per chromosome) that I selected using a custom python script.

As our data were unphased (i.e. haplotypes were unknown), I calculated LD as the squared correlation coefficient between genotypes, or r^2 , using VCFtools. Only bi-allelic sites are considered with this method, and thus genotypes were coded as either 0, 1 or 2 for the number of reference alleles in each individual at a given locus. This approach, based on genotypic allele counts, is not identical to the familiar r^2 that is estimated from haplotype frequencies, but is typically very similar, as under random mating genotypic frequencies should be a product of gametic frequencies (Weir 2008). Even when the assumption of non-random mating is relaxed, genotypic estimates of r^2 have been shown to remain fairly accurate (Rogers and Huff 2009). As my goal was to examine LD only between relatively close SNPs, I performed the analysis within a 100 kb window (i.e. only pairs of SNPs within 100 kb of each other were considered).

I used the statistical software R (version 3.3.3: R Core Team 2017) to calculate summary statistics and to fit a cubic spline, a type of generalized additive model (GAM), with normal errors to the LD decay curve using the *mgvc* package (Wood 2011).

3.2.4 Sliding Window Analysis of Weedy versus Wild Divergence

To identify regions of the sunflower genome that have differentiated in parallel between agricultural-weed populations and wild populations from more natural habitats, I generated a modified cluster separation score (CSS) between individuals from different population types in windows across the genome. The CSS approach (Jones et al. 2012) uses genetic distance matrices to identify divergent regions of the genome for isolated populations adapting to the same ecological conditions. As the weedy populations studied here are separated geographically, with limited gene flow among some population pairs, they do not collectively form a distinct population (separate from the wild populations) and therefore F_{ST} is likely not an appropriate measure of genetic differentiation (Bhatia et al. 2013).

Here, I modified the CSS approach following Miller (2016) to use principal components analysis (PCA) to calculate genetic distances rather than multi-dimensional scaling, which is more

computationally intensive. The method utilizes only bi-allelic loci and so I removed multi-allelic loci from the dataset using a custom python script. For each locus, a numeric value was assigned to each individual's genotype as follows: two reference alleles (0/0) = 0, one reference allele (0/1) = 0.5, and two alternate alleles (1/1) = 1. Each chromosome was then analyzed separately, first dividing each into sliding windows of 10000 bp in length with an overlap of 5000 bp. A PCA of the covariance matrix was calculated for each window using the *pcaMethods* package (Stacklies et al. 2007); the "svd" algorithm was selected to compute the PCA scores using singular value decomposition (SVD).

The first two principal components were retained in each window and used to calculate the Euclidean distance matrix for all individuals. The CSS score was then calculated as the mean pairwise distance between individuals from different groups (i.e. weedy versus wild) minus the average pairwise distance among individuals within groups, according to the following equation:

$$CSS = \frac{\sum_{i=1}^s \sum_{j=1}^n D_{i,j}}{sn} - \left(\frac{1}{s+n} \right) \left(\frac{\sum_{i=1}^{s-1} D_{i,i} + 1}{\frac{s-1}{2}} + \frac{\sum_{j=1}^{n-1} D_{j,j} + 1}{\frac{n-1}{2}} \right)$$

where D is the Euclidean distance between two individuals, i an individual weedy sunflower and j an individual wild sunflower, and s and n the number of weedy and wild individuals respectively. Windows that contained fewer SNPs than the total number of individuals (i.e. sixteen) were discarded. Both positive and negative values of CSS are possible. A positive value indicates a greater mean distance among sunflowers from different groups versus those from the same group, while a negative value indicates the opposite (more variation within versus among groups). Figure 3.2 provides a hypothetical example of how the PCA might look for different values of CSS. Importantly, the CSS approach was designed as a metric of parallel divergence among groups (personal communication from Dolph Schluter, also see Jones et al. 2012), with higher CSS values indicating both stronger and more parallel divergence.

Permutation testing was performed to assess the statistical significance of CSS values in each window. Group membership was shuffled 10,000 times without replacement (and keeping the same number of individuals per population type) and a CSS score calculated for each permutation. Counting the number of permuted CSS scores with absolute values greater than that of the observed CSS score (i.e., the number of scores with extreme positive or negative values) for a given window and dividing through by 10,000 gave a two-tailed p -value for the window. (While I am primarily interested in windows with high, positive CSS scores, there was signal in both directions in the dataset, and so a one-

tailed test would be statistically inappropriate). A one was added to both the numerator and denominator in this calculation to avoid obtaining values of $p = 0$.

All analyses were performed in R, and the *variantAnnotation* package (Obenchain et al. 2014) was used to load vcf files into R.

3.2.5 Variable-Sized Distinct Window Analysis of Divergence

Analyses of high-density genomic data (such as SNPs) typically proceed by pooling data over windows of adjacent markers, in order to reduce sampling noise, increase statistical power and simplify analysis. When windows overlap, as in the CSS sliding window analysis, correlated statistics are generated, but while non-overlapping, distinct windows avoid this issue, they have the disadvantage of potentially missing outliers of interest that occur at window boundaries. Here, in a complementary analysis to the CSS approach, I implemented the smoothing spline technique of Beissinger et al. (2015) to statistically identify breakpoints in the data and generate variable-sized distinct windows for an outlier analysis of weedy-wild genetic differences.

The smoothing spline method first fits a cubic smoothing spline to the raw data, then identifies inflection points of the spline to use as window boundaries (see Figure 3.3 for an illustration of the technique). Observations for each marker may be treated as estimates of an underlying continuous function f that specifies the true value of a metric (e.g. F_{ST}) at every position t_i , where t_i is the chromosomal position in bp of marker i . The cubic smoothing spline estimate \hat{f} of the function f over a range x is then obtained by minimizing $S(f)$:

$$S(f) = \sum\{Y_i - f(t_i)\}^2 + \lambda \int f''(x)^2 dx$$

where Y_i is the observed value of the metric for a marker, λ a smoothing parameter and \hat{f} is restricted to be a twice-differentiable function. Inflection points of the fitted spline occur where $\hat{f}'' = 0$.

As the smoothing spline approach requires a per-marker metric of divergence (rather than a window-based metric such as CSS), I estimated Weir and Cockerham's F_{ST} (Weir and Cockerham 1984) for each SNP using VCFtools. While F_{ST} may not be the most appropriate measure of genetic differentiation for this dataset, as discussed, it is widely used in the literature, and Miller (2016) reported a concordance between CSS and F_{ST} , so it is of interest for comparison purposes. The smoothing spline method was implemented using the *GenWin* package (Beissinger et al. 2015) in R; each chromosome was analyzed separately.

To allow for comparison among windows of different sizes (i.e. with different numbers of SNPs and hence sampling errors), a *t*-test like statistic, W , was calculated for each window in *GenWin* as follows:

$$W = \frac{(\bar{X} - \mu)}{\sqrt{s^2/n}}$$

where \bar{X} is the mean F_{ST} over the window, μ the mean F_{ST} across the chromosome, s^2 the variance in F_{ST} across the chromosome, and n the number of SNPs in the window.

3.2.6 Outlier Windows and Candidate Genes

To identify windows that represent regions of significant differentiation between weedy and wild populations of sunflowers, I took a multi-step approach. Identification of true positives remains a challenging issue in genomic analyses using high-density SNPs, given the sheer number of statistical tests typically performed (Abramovich and Benjamini 2005). It is important to recognize that, through chance alone, a number of *p*-values falling below the chosen confidence threshold α will be obtained when performing multiple tests (Noble 2009). How to correct *p*-values for multiple testing without sacrificing too much statistical power remains an open question, however methods to control the false discovery rate (FDR: Benjamini and Hochberg 1995) have become popular (Storey and Tibshirani 2003).

In the CSS analysis, 5.2% of tested windows were found to have *p*-values of < 0.05 in permutation testing, and this proportion was significantly higher than expected by chance, according to a binomial test ($p = 0.017$). Using the *qvalue* package (Storey et al. 2015) in R, I also found that values of π_0 , or the overall estimate of the proportion of true null hypotheses, averaged 0.95, with values ranging from 0.88 to 1 across chromosomes; this again suggests a small proportion of true alternative hypotheses exists for most chromosomes. However, in this low power analysis, the *q*-value approach was not able to explicitly identify which specific tests were truly significant, and FDR correction obliterated all signal in the data.

As established methods for correcting *p*-values under multiple testing were too conservative to identify true outlier windows here, I reasoned that candidate windows might be found by comparing the results from the CSS and F_{ST} analyses for overlap, as true positives may be more likely than false ones to co-occur (if false positives occur randomly across the genome in both analyses). For the CSS analysis, I considered windows with $p < 0.001$ to be regions that may potentially reflect selective differences between weedy and wild populations, worthy of further investigation. Meanwhile in the F_{ST} analysis, I

followed other authors (e.g. Choi et al. 2016; Ramu et al. 2017) in classifying windows with W -statistic values above the ninety-ninth percentile across each chromosome as regions of interest in weedy-wild divergence. For each dataset separately, I used bedtools (version 2.25.0: Quinlan and Hall 2010) to first merge any adjacent windows in the list into single entries, before taking the intersection, or overlap, of the two lists. To further focus the list of regions, I retained only intersected windows with top percentile CSS scores (i.e. mean CSS ≥ 0.3).

I looked at the number and identity of genes within this set of regions. Using bedtools, I queried the list of genomic regions against the annotated XRQ assembly (version 1.0: Badouin et al. 2017) of the *H. annuus* reference genome. When GenBank GenInfo Identifier (gi) numbers were available, records were pulled from the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/genbank/>; retrieved on 5 Oct 2017) database using the NCBI REST API to obtain protein names.

3.3 Results

3.3.1 Linkage Disequilibrium Decays Rapidly

Linkage disequilibrium, as estimated for each chromosome by genotypic r^2 over a subset of loci with MAF $\geq 10\%$, decreased with increasing physical distance between pairs of SNPs; this decline was initially rapid, but became more gradual after the distance between SNPs approached roughly 10,000 to 15,000 bp. The pattern was similar across all chromosomes and so is illustrated by a single representative chromosome (14) in Figure 3.4. To create Figure 3.4, I binned pairs of SNPs based on the distance between them at every 1 kb and took the average r^2 of all pairs for each bin. Across all chromosomes, mean (i.e. binned) r^2 values averaged 0.075 and ranged from a minimum of 0.033 to a maximum of 0.25. Given the rapid drop-off in LD after $\sim 10,000$ bp, I selected a small sliding window size (i.e. 10,000 bp) for the CSS analysis, in order not to miss small regions of differentiation between the genomes of weedy and wild populations.

3.3.2 Regions of Genetic Differentiation are Small and Scattered Over the Genome

Our bioinformatics pipeline and additional filtering steps resulted in a final dataset of 2,795,422 SNPs, of which 2,741,044 were biallelic. Given the 3.0 Gb size of the *Helianthus annuus* genome assembly (Badouin et al. 2017), this should translate to roughly one SNP every ~ 1090 bp, although SNPs will not be distributed evenly. In actuality, SNPs averaged one per 1072 bp ± 22 bp (standard error) across all chromosomes. The distribution of distances between adjacent SNPs was right-skewed with a

very long tail; the maximum distance between adjacent SNPs was highly variable among chromosomes, ranging from 132 kbp (chromosome 9) to 1149 kbp (chromosome 8). This suggests that there are regions across the genome, such as those enriched for LTR retrotransposons or other repeated sequences, for which few SNPs were called.

To identify genomic regions that have diverged in parallel between weedy and wild populations of sunflower, I performed genome scans on two different metrics of differentiation: a genetic distance-based cluster separation score (CSS) for biallelic SNPs and the fixation index, F_{ST} , for multiallelic SNPs. For the CSS analysis, there were sufficient data for 115,478 windows (each of 10,000 bp in size) across the genome, representing 19% of the total possible windows. Many windows were excluded for having fewer than sixteen SNPs; these mostly occurred in regions with no reported SNPs, and on average 91.5% of the available SNPs were used per chromosome in the analysis. Table 3.1 provides summary statistics on the distribution of SNPs per CSS window, for windows where the metric was calculated; on average there were 41 SNPs per window. For the F_{ST} analysis, all data were used genome-wide and the minimum window size possible was set to 100 bp. Table 3.2 provides summary statistics on the size of the variable-sized windows, both in base pairs and number of SNPs; on average, windows were larger than in the CSS analysis, averaging 49,456 bp per window.

There was a strong, positive correlation between values of CSS and average F_{ST} ($r = 0.85$) (Figure 3.5). Note that for each CSS sliding window, window-averaged values of F_{ST} were obtained as the “ratio of averages” (as recommended by Bhatia et al. 2013), i.e. by summing separately the numerators and denominators of all per-SNP F_{ST} estimates before taking the ratio of the sums, rather than taking a simple average of all per-SNP F_{ST} values (the “average of ratios”). The concordance between CSS and F_{ST} values can also be seen in the close correspondence of the genome scan results for each metric (Figure 3.6). Although the metrics are highly similar, CSS tends to show more exaggerated peaks, and there are peaks in CSS not seen in the W -statistic scores (summarized F_{ST} values), for example on chromosomes 3 and 15.

In considering which windows in the tails of the CSS distribution to investigate further, there were a total of 1184 CSS windows with $p < 0.01$. Although windows occurred on each of the seventeen chromosomes, they were not distributed evenly, as determined by a chi-squared test ($\chi^2 = 2209.2$, $df = 16$, $p < 10^{-16}$) that accounted for chromosome size. For example, chromosomes 1, 7 and 17 showed elevated divergence between weedy and wild populations, whereas chromosomes 11 through 13 showed little divergence (Figure 3.6). Merging adjacent windows with $p < 0.01$ produced 793 potential

regions of interest. Looking at the F_{ST} analysis, there were a total of 619 windows with W -statistic values in the top percentile, and merging adjacent windows produced 576 regions. Intersecting the two lists of regions from each analysis produced a list of 236 regions, while imposing the condition of mean CSS ≥ 3 reduced the final list to 148 regions; taken together, these regions account for less than 1% of the genome. As an example, Figure 3.7 illustrates where these regions are located on chromosomes 1, 7 and 17.

3.3.3 Candidate Adaptive Genes

Overlapping with the 148 regions of interest in weedy-wild divergence, we found 267 genes in the *H. annuus* XRQ genome assembly annotations; this equated to roughly 1.8 ± 0.13 genes per region. The majority of the genes ($n = 153$) were matches to expressed sequence tags in sunflower for which the gene products show no matches to protein or signature databases. Hence, these matches represent transcribed regions with unknown functions. Additionally, 36 genes were described as producing “uncharacterized proteins” and thus also have unknown functions to date. Finally, a total of 77 genes (presented in Appendix A) were linked to a probable protein match, and I will focus on these here. Note that 22 of these genes do not yet have a gene identifier assigned for the sunflower genome, but a probable protein match could be made owing to close sequence homology to identified proteins in other plant species.

Collectively, identified proteins were a diverse group, ranging from structural proteins (e.g. 60S ribosomal protein L38), to catalytic enzymes (e.g. hexokinase-2, isopentyltransferase-5, phospholipase A1-II 1), to signalling molecules and receptors (e.g. cysteine-rich RLK 33) and finally to transcription factors (e.g. floral homeotic protein APETALA 2-like, histone-lysine N-methyltransferase Tr). In many cases, functional information was limited. For example, a gene might be specified as coding for a particular type of domain (e.g. a flavoprotein-like domain) without knowledge of which protein it belonged to; similarly, a protein family or superfamily might be identified rather than a specific protein. As protein families may include members with diverse and varied functions, we are limited in what we can say regarding the potential role of these genes in the divergence between the weedy and wild types.

Many of the identified proteins have been implicated in the response to abiotic or biotic stress, and these may be candidates for the phenotypic differences seen in agricultural weeds. Some examples of proteins linked to abiotic stress tolerance include: nuclease HARBI1, which is upregulated under salt-stress in *Reaumuria trigyna* Maxim. (Dang et al. 2014); ribonuclease H protein At1g65750, downregulated in nickel-resistant white birch (*Betula papyrifera* Marshall: Theriault and Nkongolo

2017); DREB2A-interacting protein 1, enhances drought and high temperature stress tolerance when overexpressed in *Arabidopsis thaliana* (L.) Heynh. (Sakuma et al. 2006); luminal-binding protein 5, linked to increased drought tolerance in both soybean (*G. max*) and tobacco (*Nicotiana tabacum* L.: Valente et al. 2009); and ascorbate peroxidase 3, a hydrogen peroxide scavenger that detoxifies H₂O₂ (a type of reactive oxygen species, or ROS) produced under adverse environmental conditions (Caverzan et al. 2012). Several of the protein families identified here also have links to stress tolerance, including the integrase-type, DNA-binding superfamily (role in submergence and hypoxia tolerance in *Arabidopsis*: Seok et al. 2014), late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family (LEAs accumulate under drought stress in legumes: Battaglia and Covarrubias 2013), and proteins with F-box domains (can play a role in stress response via the ubiquitin pathway, e.g. in wheat (*Triticum* spp.), *TaFBA1* enhances tolerance to oxidative stress and drought: Zhou et al. 2015). Considering biotic stressors such as pathogens and herbivores, GnK2 is an anti-fungal protein expressed in the endosperm of *Ginkgo biloba* L. seeds (Miyakawa et al. 2009), and the LRR receptor-like serine/threonine-protein kinase EFR is a pattern-recognition receptor for the elongation factor Ef-Tu, a potent activator of the anti-pathogen defense response (Zipfel et al. 2006); meanwhile, proteins from the glycoside hydrolase family 1 play a role in the creation of chemical defenses against herbivory (Xu et al. 2004), and VAMP family proteins function in plant disease-resistance pathways (e.g. to powdery mildew: Yun et al. 2016). Conversely, other outliers such as BONZAI 3-like are known to repress plant immunity, facilitating growth and development (Li et al. 2009b).

Interestingly, we found two flowering-related genes in the windows of interest. Considering that only 35 regions were found to affect flowering in sunflower in a recent GWAS study (Badouin et al. 2017), it is unlikely that these would be found by chance alone. The first related protein, floral homeotic protein APETALA 2-like (AP2), is a transcriptional activator that promotes early floral meristem identity, and is necessary for the transition of an inflorescence meristem to a floral meristem (Jofuku et al. 1994). The second, AGAMOUS-like 24 (AGL24), was associated with a strong peak on chromosome 15, and is a transcription factor that mediates crosstalk between the flowering time genes *FT*, *SOC1* and *LEAFY*. Overexpression of AGL24 in *Arabidopsis thaliana* leads to precocious flowering, while loss of AGL24 results in late flowering (Yu et al. 2002). Finally, genes coding for an ABC-transporter and two cytochrome P450s were identified. Both families, though associated with diverse functions, have been implicated in non-target site herbicide resistance (Yuan et al. 2007; Nol et al. 2012).

3.4 Discussion

In this study, we compared collections of wild sunflowers infesting crop fields (agricultural-weed or “weedy” populations) to those collected from more natural environments (non-agricultural or “wild” populations). In common garden work, weedy sunflowers have been shown to demonstrate a shift in life-history strategy (see Chapter 2 & Mayrose et al. 2011), trading off stress tolerance in favour of growth and reproduction. Here, we examined the genomes of sixteen individual sunflowers (eight weedy and eight wild) to look for regions of parallel differentiation between the two types. Using two complementary approaches, a sliding window analysis of cluster separation scores (a metric based on genetic distances: Jones et al. 2012) and a distinct, variable-sized window analysis of F_{ST} between groups (following the methods of Beissinger et al. 2015), we found a number of potential small regions of genetic differentiation scattered across the genome. Within these regions, there are candidate genes for observed phenotypic differences, including genes linked to plant growth, immunity and abiotic stress tolerance, as well as two genes (*AP2* and *AGL24*) with known roles in the transition to and timing of flowering.

3.4.1 Parallel Adaptation to the Agricultural Environment Proceeds from Standing Variation

Despite ongoing gene flow among sunflower populations of different types, we were able to identify regions of genetic differentiation between weedy and wild individuals, though these did not pass strict false-discovery rate correction (FDR). However, careful evaluation of the proportion of true null hypotheses (which ranged from $\pi_0 = 0.88$ to 1 across chromosomes) revealed a meaningful fraction of true alternative hypotheses (given as $\pi_0 - 1$) for most chromosomes. Thus, although the FDR method was not able to identify which specific windows were true outliers, we decided to look within top-ranked windows (mean CSS ≥ 0.3 , CSS $p < 0.01$ and W -statistic in the top percentile) for candidate genes. These windows should be approached with caution as not all represent true outliers, though as a group, they indicate regions of interest. It is possible that including a larger number of individuals in our analysis may strengthen our ability to detect outlier windows. However, this may not be the case, given that there is likely variability among individual weeds from different populations (i.e. genetic differences are not fixed between types), leading to small effect sizes. With a larger sample size, it would however be possible to better account for the paired nature of the data (i.e., weedy-wild populations are paired by latitude), which would likely strengthen our findings, as strong genetic differences due to adaptation to climate, for example, may be masking the relatively weaker signal of weedy-wild divergence.

We observed a total of 148 potential regions of parallel differentiation, varying in size, that in total accounted for a small fraction of the sunflower genome (less than 1%). Given that our methodology, i.e., pooling individuals from each population type, only identifies regions that have diverged in parallel across all agricultural-weed populations, this represents a conservative estimate of the proportion of the genome implicated in the evolution of weediness. Regions (and loci) that are differentiated in only one or a few population pairs will go undetected here, as they will not achieve high CSS values (Dolph Schluter, personal communication), though any such unique adaptations may be important locally. Nonetheless, long-standing weedy populations (i.e., that have been crop weeds consistently for decades) do show some potential parallel genetic changes. This result is perhaps unexpected. Kane and Rieseberg (2008) found that weedy sunflowers (from a different region of the USA) are more closely related to nearby non-agricultural, wild populations than to other weedy populations, implying that weediness has arisen independently. Given the large geographic distances between many weed populations, this could allow for adaptation to proceed individually and idiosyncratically among weed populations. *Helianthus annuus* shows a pattern of isolation-by-distance across the Midwest region where our populations were collected from (see Chapter 4), without significant population structure; this is suggestive of independent evolution of the weedy populations in our study. While it is possible that beneficial weedy alleles have been shared across the range, given the short timespan over which agriculture has intensified (beginning with the Green Revolution less than 100 years ago), adaptation from standing genetic variation seems more plausible, i.e. soft selective sweeps (Hermisson and Pennings 2005).

For each region of interest, we identified an average of 1.8 genes. The proportion of potential differentiated regions was less than 1% of the genome, within the range found by other whole genome resequencing studies of divergence between contrasting environments, although these controlled for FDR (and we did not). For example, Jones et al. (2012) identified a genome-wide set of loci associated with the divergence of marine-freshwater stickleback, *G. aculeatus*, using two different methods; outlier regions accounted for from 0.2% of the stickleback genome (when only consensus regions were considered) to 0.5% (considering regions identified by either method). In contrast, Miller (2016) found a higher proportion of the stickleback genome (1.7%) had differentiated in response to predation by sculpin. Both studies identified standing genetic variation as playing an important role in the repeated evolution they observed, as with weedy sunflowers. Moving to plant species, Steane et al. (2017) compared populations of three Eucalypts (*Eucalyptus* spp.) growing across a rainfall gradient in Southern Australia; the proportion of outlier loci detected ranged from 0.12-0.2% (*E. salubris* F. Muell.), to 1.4%

(*E. loxopleba* Benth.), to 2.6% (*E. tricarpa* (L.A.S. Johnson) L.A.S. Johnson & K.D. Hill), but the authors attributed the low percentage for *E. salubris* to be due to small sample size. In a comparison of *Arabidopsis lyrata* (L.) O’Kane & Al-Shehbaz populations growing on serpentine versus non-serpentine soils, Turner et al. (2010) identified 96 loci associated with soil type out of 8.4 million SNPs analyzed. Also in serpentine soils, Porter et al. (2017) found up to 3% of genes were associated with divergence between wild *Mesorhizobium* strains found in nickel-enriched serpentine soils versus low-nickel soils. In combination with these findings, our results suggest that differentiation due to adaptation to contrasting environments occurs only for a small fraction of the genome, though this depends to some extent on the stringency of the methods used to detect locally adapted regions.

Regions of divergence between weedy and wild populations were not evenly distributed across the genome, but rather followed a more clumped distribution. This finding is consistent with that of many previous studies (e.g. Nosil et al. 2009; Nadeau et al. 2012; Renaut et al. 2012; Delmore et al. 2015), which have shown that the loci involved in local adaptation often group together. As adaptation often proceeds with ongoing maladaptive gene flow, theory predicts that adaptive alleles may be located close together (Yeaman 2013), as less recombination events will then occur to separate co-adapted alleles. For the same reason, these clusters of adaptive loci may often occur in genomic regions of low recombination (Samuk et al. 2017). Given the close proximity of many wild sunflower populations on the landscape, and the ability of sunflower pollen to travel some distance (1 km or more: Arias and Rieseberg 1994), weedy populations almost certainly receive a constant influx of pollen from nearby wild populations (as well as from crop sunflower). Hence, the combined forces of divergent selection and homogenizing gene flow may have produced the clumped distribution seen here, although other explanations are possible (see Cruickshank and Hahn 2014).

3.4.2 Multiple Genes of Small Effect Contribute to Phenotypic Differences

Regions of genetic differentiation between the weedy and wild populations were found to overlap with 267 genes in the *H. annuus* XRQ reference genome annotations, suggesting that agricultural environments result in selection on a substantial number of genes. As multiple genes overlapped with the same region in some cases, it is possible that not all genes are selected loci however. In the case of tight linkage between two loci within a region, for example, only a single loci may be the target of selection. Neutral or even slightly deleterious alleles may “hitch-hike” (Barton 2000) along with selected loci when in tight linkage. Hence, the outlier list should be approached with

some caution, and requires further study to explore and validate the candidates. Candidate genes were diverse, coding for many types of proteins, but many were implicated in plant stress response pathways.

Our findings are consistent with previous work on weedy sunflower populations infesting corn fields that identified between 1% to 6% of tested loci (from a total of 106 microsatellites), “a small, but not insignificant fraction of the genome” (Kane and Rieseberg 2008 p. 384), as outliers in comparisons with wild populations. Similar to Kane and Rieseberg (2008), we also identified transcription factors, membrane-bound transporters and heat-shock proteins as top candidate genes, though it is unknown if the two studies identified any of the same genes. Working with a subset of the same populations as Kane and Rieseberg (2008), Lai et al. (2008) examined gene expression using a sunflower cDNA microarray. Growing individuals in a common growth chamber, gene expression was found to differ between weedy and wild populations at 165 uni-genes (~5% of the total). Interestingly, uni-genes were enriched for abiotic/biotic stimulus and stress response proteins, as suggested by our work, although we did not carry out a formal GO enrichment analysis.

Unlike mapping studies, genome scans for regions of divergence do not provide information on the effect sizes of candidate genes. Thus, we cannot rank our candidates in terms of their importance for weed adaptation. However, given the number of loci identified as potential candidates, it seems likely that weed evolution has involved many genes that are of small individual effect. Agricultural-weed sunflowers differ from wild sunflowers for multiple phenotypic traits (e.g. seedling growth rate, flowering time, drought tolerance, etc.) and this may necessitate the involvement of multiple genes and pathways with different functions. Similarly, many of the traits distinguishing agricultural weeds represent complex, quantitative phenotypes, again perhaps precluding a simple genetic basis. For example, drought tolerance has been well studied in crop plants such as wheat, where the genetic basis has been found to be multi-genic, with low trait heritability and high G×E interactions (Fleury et al. 2010).

Many candidate genes (e.g. ascorbate peroxidase 3, DREB2A-interacting protein 1, luminal-binding protein 5) were linked to abiotic or biotic stress response pathways in plants, though we did not test statistically for enrichment of this gene category. As part of the life-history shift observed for weedy sunflowers, stress tolerance has been traded-off in favour of faster growth and development. This may be achieved through the same genes linked to the decreased stress response. For example, the candidate *BON3* (producing protein BONZAI 3-like) has been shown to negatively regulate disease resistance (*R*) genes in *Arabidopsis* (Li et al. 2009b). Pathogen resistance responses can impose fitness

costs even in the absence of pathogens (Tian et al. 2003) and mutants with constitutive expression of R proteins often exhibit growth defects (e.g. Shirano et al. 2002); therefore, control of R gene expression, such as that achieved via *BON3*, is necessary to maximize plant growth.

Agricultural weeds may frequently evolve herbicide resistance (Heap 2014) as an important adaptation to infesting modern agricultural fields. The International Survey of Herbicide-Resistant Weeds (www.weedscience.org; retrieved on 22 Sept 2017) reports that 210 agricultural weed species worldwide have evolved resistance to a least one type of herbicide, with evolved resistance reported for 152 different herbicides. Thus, it seems likely that this may also be an important source of strong selection on weedy sunflowers, and indeed resistance to common herbicides (e.g. imazethapyr, imidazolinone, and sulfonyleurea) has been reported for sunflower in several regions of the US (Massinga et al. 2003). In this study, we identified an ABC-transporter and two cytochrome P450 monooxygenases as candidate genes. Both represent large protein families (Kang et al. 2011), with P450s being the third largest gene family in *Arabidopsis* (Mizutani 2012), participating in a variety of biochemical pathways. Interestingly, both have been implicated as playing key roles in herbicide resistance achieved via detoxification mechanisms (see Yuan et al. 2007 for a review). For example, Nol et al. (2012) found that two ABC-transporters (M10 and M11) were upregulated in *Conyza canadensis* (L.) Cronquist individuals treated with glyphosate, supporting a role for these genes in glyphosate resistance via reduced translocation. While we do not know if the candidate genes identified here are responsible for herbicide resistance, or if they were simply picked up by chance in the analysis, the possibility is intriguing.

3.4.3 Conclusions

In our study of parallel adaptation in agricultural-weed populations of wild sunflower, we found a number of small regions of potential genetic differentiation between weedy and wild populations. The regions identified here were discovered on the basis of parallel changes across multiple weedy populations, and hence are likely to be the result of adaptive evolution. While neutral processes, such as genetic drift, may result in allele frequency differences between populations, it is unlikely that the same changes would occur repeatedly across populations due to chance alone (Elmer and Meyer 2011). Given that agriculture originated less than 12,000 years ago (Doebley et al. 2006), and that agricultural intensification in the US and Canada has occurred in only the last 100 years, evolutionary changes seen here have taken place in a relatively short timespan. Adaptation to an agricultural context has also occurred in the face of gene flow from wild sunflower populations, as shown by both other theoretical work and empirical work in other species (e.g. see Nosil 2012; Westram et al. 2014). Our results

represent one of a growing number of studies using whole genome resequencing to study ecological differentiation and illustrate the power of the approach to detect candidate loci and regions of divergence.

Table 3.1: Summary statistics by chromosome for the CSS analysis , including the number of windows with sufficient data for analysis, the mean and maximum number of SNPs per window, and the number of windows with $p < 0.01$.

Chromosome	Chromosome Length (Mb)	Chromosome Length (SNPs)	Number of Windows	Mean Number of SNPs	Maximum Number of SNPs	Number of Windows with $p < 0.01$
1	153.9	144,953	6,199	40.3	209	121
2	180.6	136,030	5,890	38.2	157	48
3	168.5	158,953	6,545	42.5	191	50
4	178.9	153,726	6,668	39.3	175	54
5	219.0	175,721	7,429	40.0	187	71
6	103.8	106,435	4,448	41.6	176	46
7	103.8	99,389	4,141	41.7	234	57
8	153.2	148,232	6,216	41.3	222	76
9	209.8	206,635	8,375	43.5	222	77
10	246.3	215,859	9,082	40.8	177	108
11	168.4	137,508	5,860	39.6	214	54
12	166.4	155,423	6,518	40.3	188	31
13	197.2	185,970	7,984	40.3	191	36
14	174.4	167,239	6,751	43.3	184	70
15	171.2	181,039	7,723	40.8	212	95
16	188.6	182,372	7,809	40.4	176	95
17	214.7	185,560	7,840	40.5	191	95

Table 3.2: Summary statistics by chromosome for the F_{ST} analysis using the cubic smoothing spine technique of Beissinger et al. (2015) to identify distinct window boundaries; windows may be of variable size.

Chromosome	Number of Windows	Mean Window Size (bp)	Maximum Window Size (bp)	Mean Window Size (SNPs)	Maximum Window Size (SNPs)	Number of Windows in Top 1%
1	3,382	45,503	553,600	43.7	390	34
2	3,013	59,886	813,400	46.0	311	31
3	3,388	49,606	701,800	47.9	292	34
4	3,849	46,420	714,000	40.8	306	39
5	3,130	70,002	518,500	57.2	418	32
6	2,663	38,988	386,500	40.8	257	27
7	2,028	51,206	579,800	50.1	316	21
8	3,554	43,089	653,600	42.5	361	36
9	3,966	52,869	729,600	53.2	363	40
10	4,627	53,215	666,200	47.5	476	47
11	3,150	53,481	576,900	44.6	369	32
12	3,341	49,817	1,147,400	47.4	382	34
13	4,631	42,582	743,100	40.9	299	47
14	4,166	41,863	628,300	41.0	324	42
15	4,086	41,863	881,100	45.1	305	41
16	4,113	45,831	737,600	45.2	323	42
17	3,937	54,537	774,100	48.1	432	40

Figure 3.1: North American range of *Helianthus annuus* (based on Rogers et al. 1982) and collection locations of populations included in this study. Each agricultural-weed population was paired with a nearby non-agricultural population of wild sunflowers, and location names are for the pair. Note that two of the population pairs included in the phenotypic comparisons of Chapter 2 (Iowa 1 and Kansas 1) are not included in the genetic analysis.

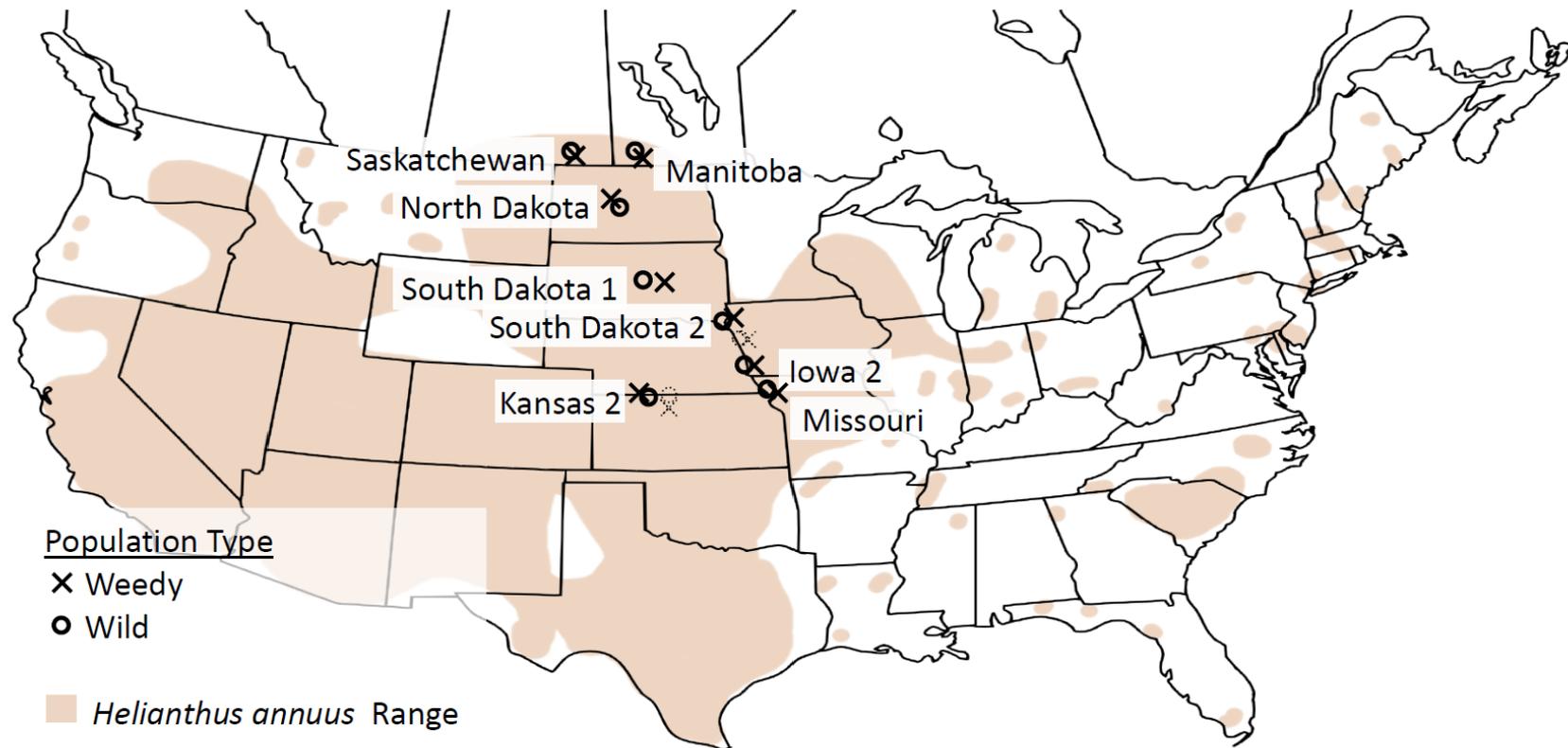


Figure 3.2: Hypothetical graphs illustrating the first two components (PC1 and PC2) of a principal components analysis (PCA) conducted on genotype data where the CSS metric is (a) positive ($CSS > 1$), (b) neutral ($CSS = 0$) and (c) negative ($CSS < 0$). Each “x” represents an individual that is an agricultural-weed, while each “o” is a wild, non-agricultural individual. In this analysis, we are primarily interested in cases where CSS is positive, and there is greater genetic distance between individuals from different groups than among individuals of the same group.

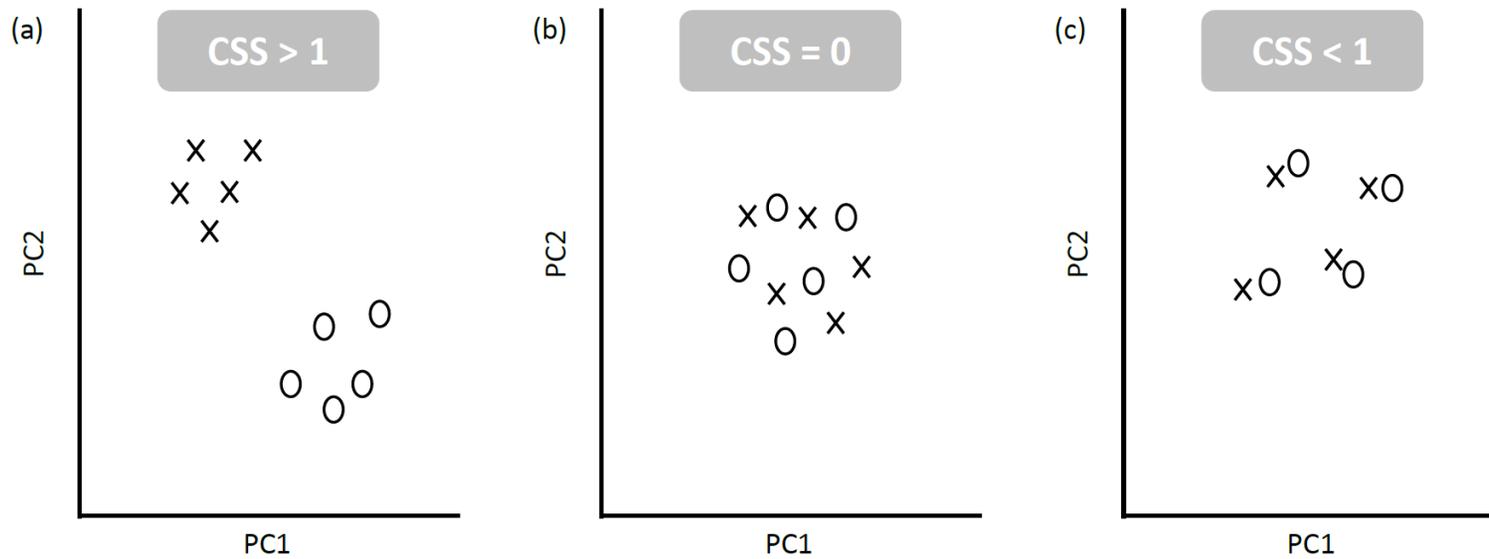


Figure 3.3: Depiction of the cubic smoothing spline method developed by Beissinger et al. (2015) for a hypothetical chromosomal segment. (a) First, a cubic smoothing spline (shown in red) is fitted to the raw data: F_{ST} values for each SNP in this case. (b) Inflection points, indicated by the dashed lines, are identified where the second derivative of the spline equals zero. (c) Using the inflection points to define distinct window boundaries, data are summarized for each window using a W -statistic, for example.

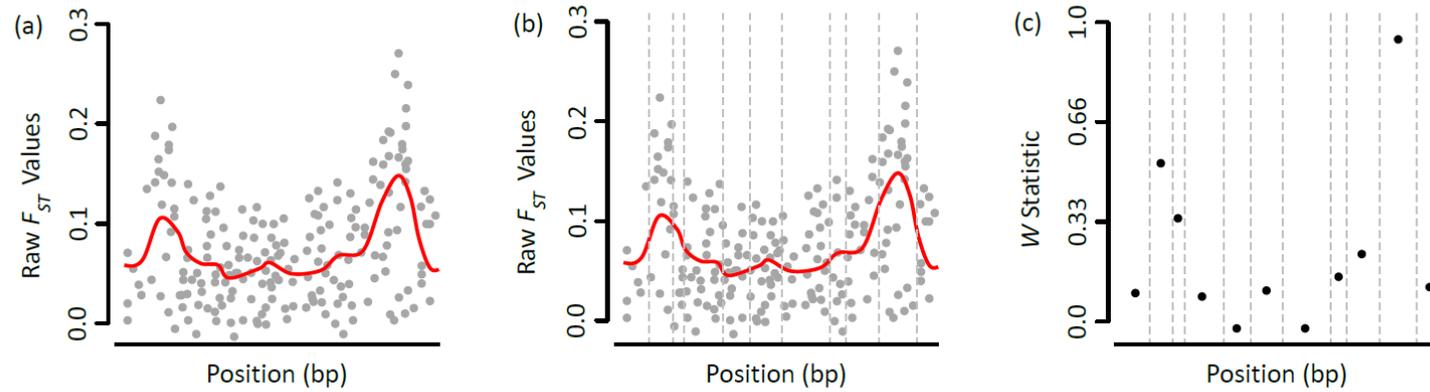


Figure 3.4: Pattern of decay in linkage disequilibrium (LD), as estimated by the squared correlation coefficient among genotypes r^2 , for chromosome 14. Data points represent the mean LD for pairs of SNPs located in each 1 kb bin, with error bars showing standard errors. The red line depicts a cubic spline fit using a generalized-additive model (GAM) with normal errors (adjusted $R^2 = 89.6\%$).

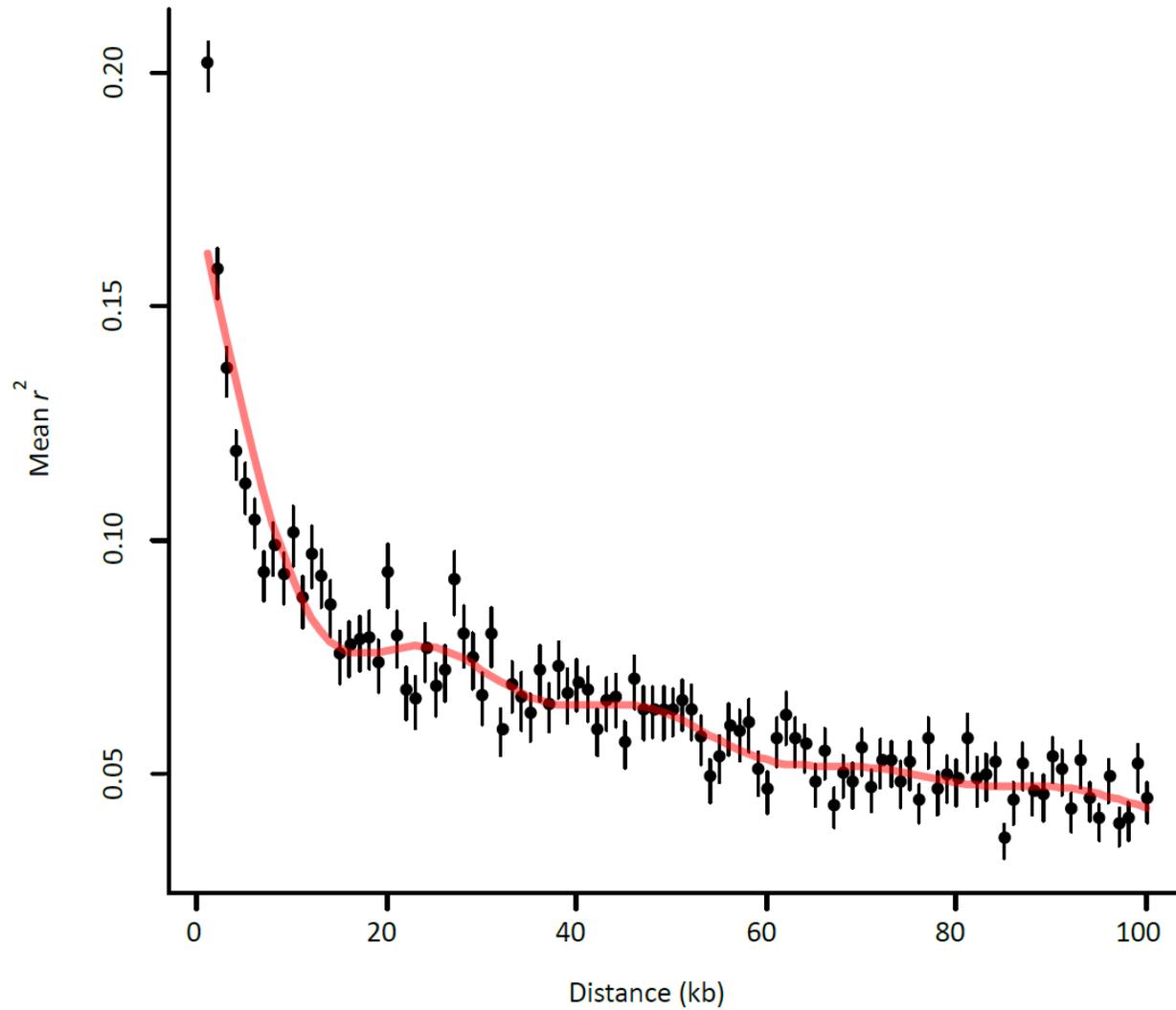


Figure 3.5: Plot of weighted F_{ST} calculated as the “ratio of averages” for all SNPs in a window versus the cluster separation score (CSS) for 10,000 bp windows located across the genome. Both metrics assess genetic differentiation between agricultural-weed populations and non-agricultural populations of sunflower. Points coloured red indicate windows where $p < 0.01$ in the CSS analysis.

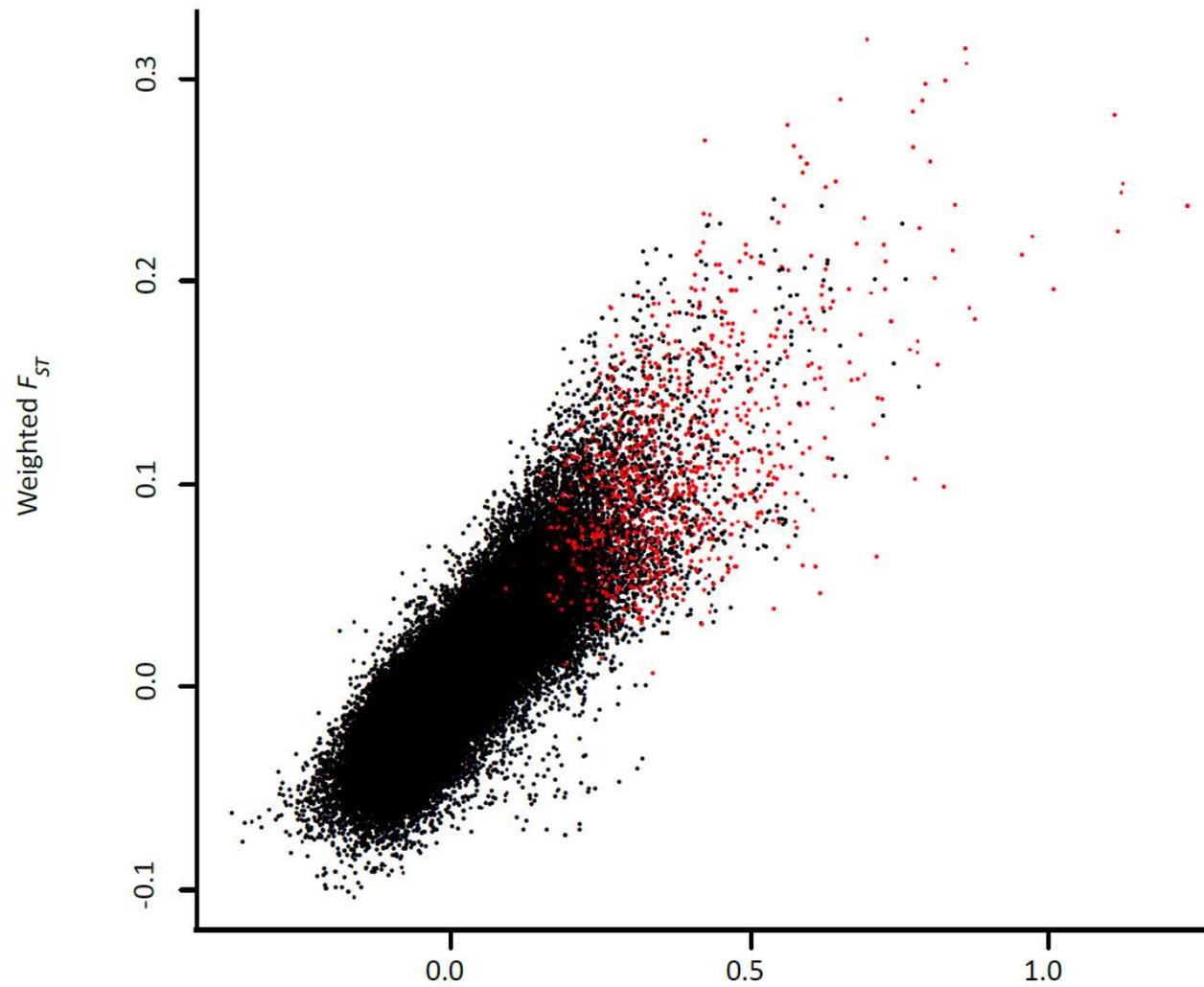


Figure 3.6: Genome-wide distribution of CSS and W -statistic scores (based on F_{ST} data), which are both presented as ten-window rolling averages. All chromosomes are plotted on the same scale. Values of CSS are shown in black with a corresponding axis on the left-hand side; W -statistics are shown in red with a corresponding axis on the right-hand side. Both metrics show strong concordance in their patterns, with many small regions of genomic differentiation between agricultural-weed and non-agricultural sunflowers scattered across the genome.

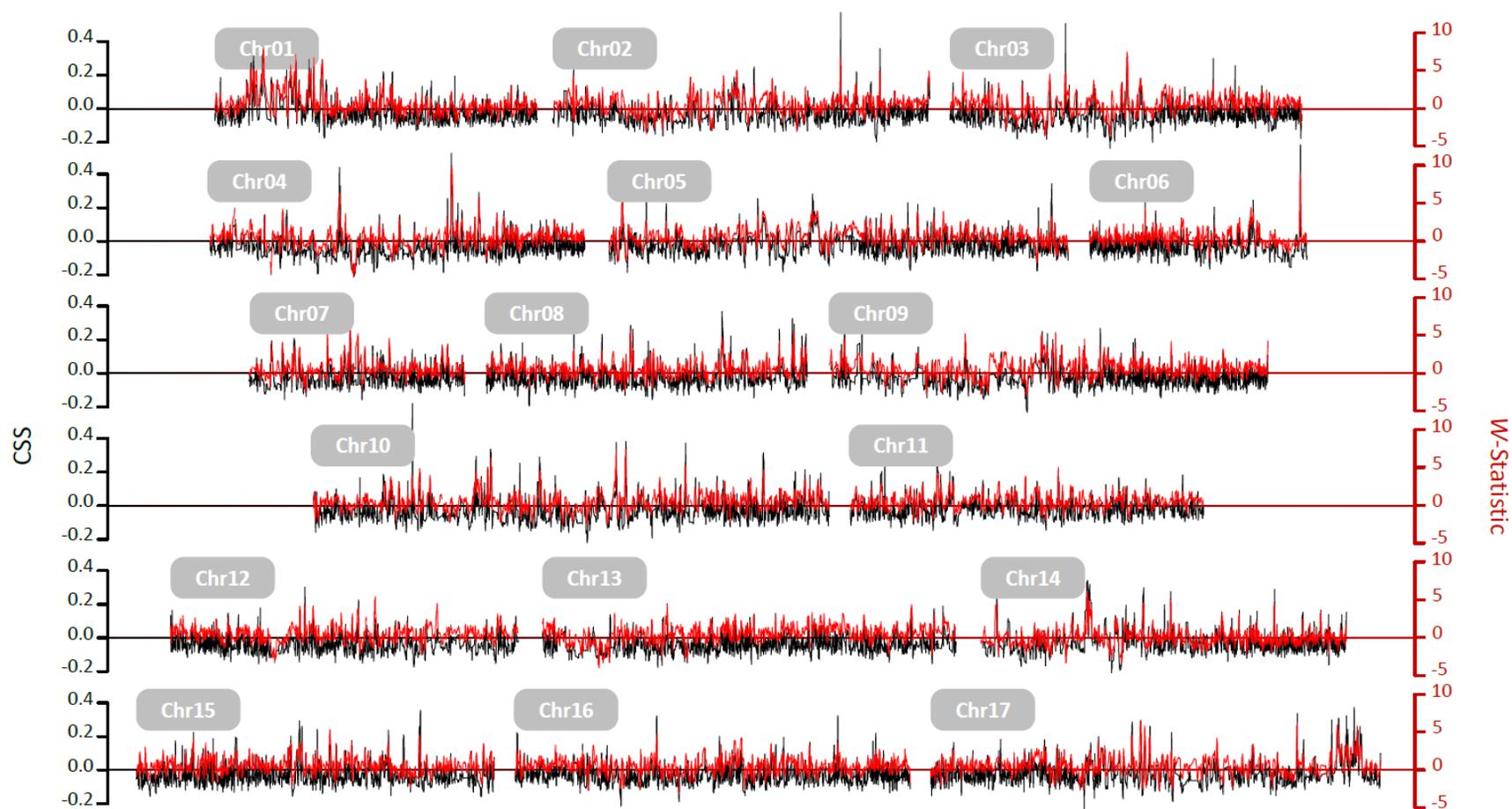
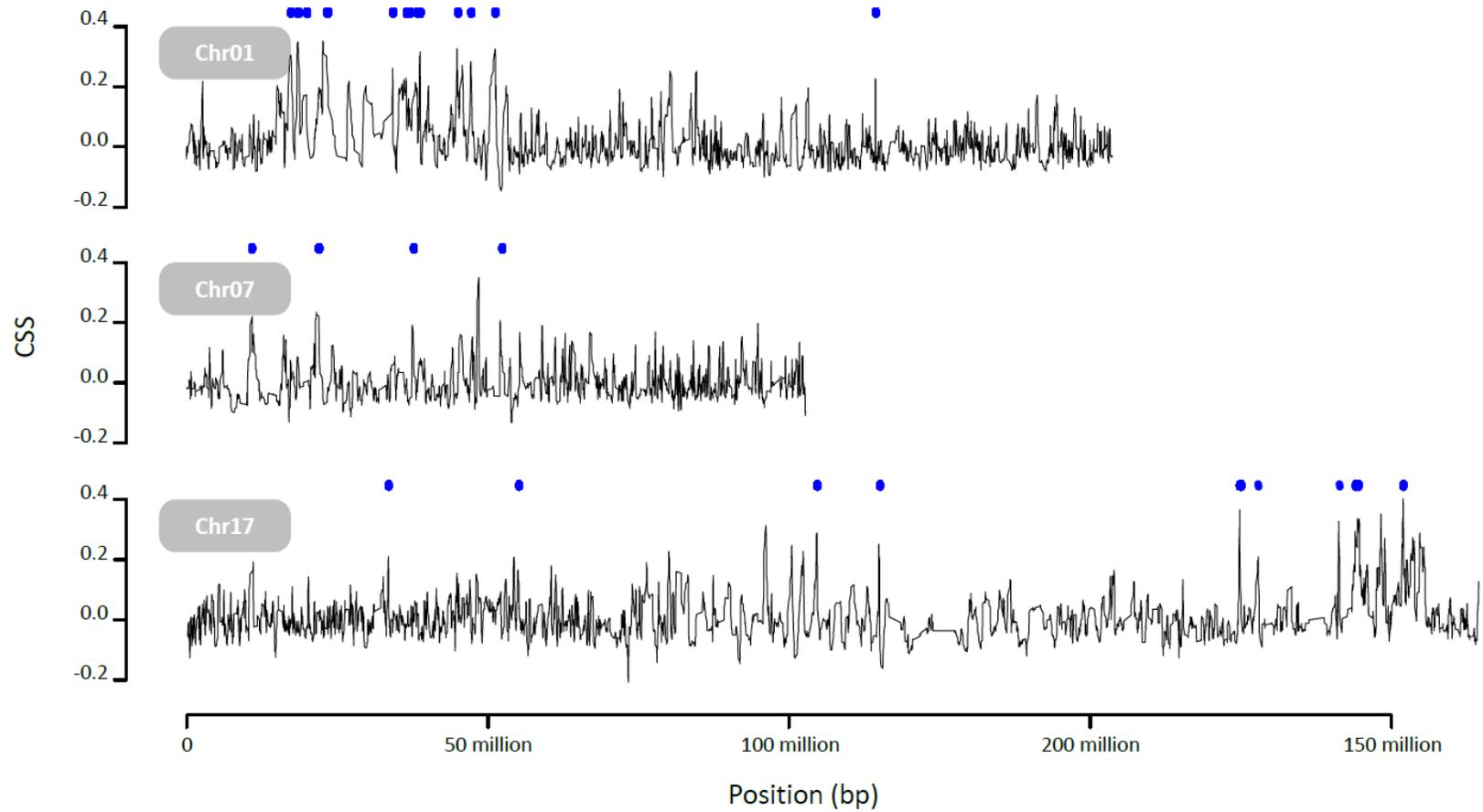


Figure 3.7: Ten-window rolling average of the CSS metric between agricultural-weed and non-agricultural sunflower populations for chromosomes 1 (a), 7 (b) and 17 (c). All chromosomes are plotted on the same scale. Blue line segments indicate the presence of outlier windows, which often, but not always, correspond to the highest peaks in CSS.



Chapter 4 : Genome-Wide Association Analysis of Glyphosate Resistance in Wild Sunflowers

4.1 Introduction

Herbicide resistance represents a striking example of rapid evolution. While humans have battled to control weeds reducing crop productivity since the inception of agriculture, herbicides have only become an integral part of this battle in the last half century or so (Vats 2015), meaning that resistance has evolved in just a few short decades, with the exact timeframe depending on the herbicide. Globally, the evolution of herbicide resistance has been reported in an ever-increasing number of weed species annually (see Figure 4.1), with resistance to at least one herbicide reported in 252 species to date, according to the International Survey of Herbicide Resistant Weeds (Heap 2017). Weeds have evolved resistance to 163 herbicides, with 23 unique sites of action (where a site of action is the specific protein inhibited by an herbicide), with some species resistant to more than one herbicide and/or site of action. Furthermore, distantly related species may evolve resistance to the same herbicide or site of action, presenting interesting cases of convergent evolution (Baucom 2016). Thus, from the perspective of evolutionary biology, evolved herbicide resistance (as opposed to the natural tolerance, or innate resistance, some species can show to certain herbicides: Nandula 2010) represents an exciting opportunity to understand the genetics of adaptation as it proceeds in “real time”. What is the role of standing genetic variation versus *de novo* mutation in the genetics of resistance? Are structural (i.e., protein changing) or regulatory (i.e., gene expression changing) mutations more commonly involved? In cases of convergence seen across genera, are the same or different genetic mechanisms involved? And, finally, what are the fitness consequences of resistance and its dynamics over space and time? Meanwhile, from a practical standpoint, the economic costs of weed control are huge (Pimentel et al. 2005). Achieving weed control in the long-term requires better management practices, which depend on knowledge of the evolutionary dynamics and mechanisms of resistance (Neve et al. 2009).

The majority of herbicides inhibit specific plant enzymes, known as “target sites”, that are essential to plant metabolism (Powles and Yu 2010). As such, there are two main types of mechanisms by which herbicide resistance can evolve: target-site and non-target-site resistance. Target-site resistance (TSR) results from alterations to the target enzyme such that the herbicide can no longer effectively inhibit enzyme action. For example, resistance to the herbicide triazine is mediated by a single mutation that has evolved independently in weedy species worldwide (Fuerst and Norman 1991); here, a point mutation in the chloroplastic *psbA* gene encoding the targeted enzyme (photosystem two

protein D1) causes a single amino acid shift (Ser-264-Gly) in the binding site of D1 (Oettmeier 1999), which is sufficient to prevent triazine binding. The Ser-264-Gly mutation still allows for normal enzyme function, but at a reduced level (Gronwald 1997). This is a common feature of TSR involving binding site changes in targeted enzymes: plant fitness may be lowered in the absence of the herbicide as a result of reduced efficiency of the enzyme (Vila-Aiub et al. 2009). Alternatively, TSR may involve enhanced expression of the targeted enzyme (via gene amplification or alterations to the promoter), which can allow critical reactions to proceed via overproduction of the enzyme. For example, in Palmer amaranth (*Amaranthus palmeri* S. Watson), genotypes resistant to the herbicide glyphosate possess a greater number of copies of the *EPSPS* gene, and this translates into greater EPSPS protein levels and enzymatic activity (Gaines et al. 2011). The mechanism underlying this gene amplification remains unknown, but transposon-mediated amplification seems likely (Gaines et al. 2013). By nature, TSR is typically monogenic (Mithila and Godar 2013) and can exhibit varying levels of dominance (see Powles and Shaner 2001 for examples).

Non-target-site resistance (NTSR) encompasses various non-exclusive means of preventing an herbicide from reaching its target site (Powles and Yu 2010), including decreased foliar penetration or translocation of the herbicide within the plant, as well as increased sequestration or metabolic detoxification of the herbicide. Over evolutionary time, plants have developed sophisticated detoxification systems for harmful chemicals (Vaahtera and Brosché 2011), and these may be co-opted in new ways to remove toxic herbicide molecules. Many plant detoxifying enzymes and transporters may be involved in NTSR (Délye 2013), and overall our understanding of the physiological basis of most cases of NTSR is poor (Ghanizadeh and Harrington 2017b). Participation in NTSR has been confirmed for only four gene families to date (Yuan et al. 2007): cytochrome P450 mono-oxygenases, glutathione S-transferases, glycosyltransferases and ABC transporters. For example, in velvetleaf (*Abutilon theophrasti* Medik.), glutathione S-transferases (GSTs) have been reported to mediate triazine resistance via detoxification by glutathione conjugation (Anderson and Gronwald 1991); resistance is achieved via increased GST activity that results from enhanced catalytic capacity (Plaisance et al. 1999), suggesting there has been a mutation that enhances herbicide binding. A unique feature of NTSR is that, unlike TSR which is very herbicide specific, NTSR can unpredictably confer resistance to multiple herbicides, including herbicides with different sites of action or even novel products not yet marketed (e.g. Preston et al. 1996; Cummins et al. 1999; Petit et al. 2010). Thus, NTSR represents a clear threat to weed management, and more information is urgently needed on this widespread, but poorly characterized, form of resistance (Délye 2013). Given the complexity and diversity of NTSR mechanisms, which often

involve a coordinated cellular response, it is believed that NTSR is typically a quantitative trait that evolves through the accumulation of different resistance alleles, though there are very limited data on the genetics of NTSR in existing weeds (Yuan et al. 2007).

Dubbed a “once-in-a-century” herbicide (Duke and Powles 2008) for its versatility, efficacy and low environmental toxicity, *N*-(phosphonomethyl)glycine, or glyphosate, has risen in prominence since its introduction in 1974. Commercialization of genetically engineered glyphosate-resistant crops by Monsanto in the late 1990s greatly contributed to the utility of glyphosate. Glyphosate is currently the most used herbicide in the world, with particularly high levels of consumption in the Americas (Powles and Preston 2006; Benbrook 2016), largely due to the extensive adoption of glyphosate resistant, Roundup Ready corn (*Zea mays* L.), cotton (*Gossypium* spp.) and soybean (*Glycine max* (L.) Merr.) in the U.S.. The global glyphosate market has been predicted to exceed 1.1 million metric tonnes by 2022 (Global Industry Analysts, Inc.), based on the growing popularity of Roundup Ready crops, low- and no-till systems and an expected increase in biofuel projects. The evolution of glyphosate resistance in weeds was initially thought to be very unlikely for several reasons (Bradshaw et al. 1997), including the difficulty in developing resistance in crops (which ultimately required the use of a bacterial transgene, as site-directed mutants had only low levels of resistance), the fact that many tested plant species did not readily metabolize glyphosate, and the lack of resistance observed in weeds in the first 20 years of commercial use. However, resistance was eventually reported in 1996 and its prevalence continues to increase (Heap 2014). There are currently 38 glyphosate resistant weed species, as reported by The International Survey of Herbicide Resistant Weeds (Heap 2017); species include both monocots, such as *Lolium rigidum* Gaudin (the first species to have documented resistance: Powles et al. 1998) and *Eleusine indica* (L.) Gaertn. (Lee and Ngim 2000), and dicots, such as *Ambrosia artemisiifolia* L. (Brewer and Oliver 2009) and *Conyza canadensis* (L.) Cronquist (VanGessel 2001). Many of these species are now under study to determine the mechanism(s) and genetics of resistance, and more recently genomic approaches have been initiated.

Glyphosate exerts its toxic effects by binding to and consequently inhibiting the function of the chloroplastic enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) (Steinrücken and Amrhein 1980). This enzyme plays a key role in the shikimic acid pathway, catalyzing the reaction of shikimate-3-phosphate (S3P) and phosphoenol pyruvate to form 5-enolpyruvyl-shikimate-3-phosphate (ESP). The shikimic acid pathway converts simple carbohydrate precursors into aromatic amino acids (i.e., phenylalanine, tyrosine and tryptophan), which may be transformed into plant hormones, vitamins and

other metabolites essential for plant function (Herrmann and Weaver 1999). When glyphosate is sprayed on a plant, it is initially and quickly absorbed through foliar surfaces and then translocated within the plant to growth sites such as meristem and roots where it will bind EPSPS (Duke and Powles 2008). As the reaction is blocked, shikimic acid begins to build up (leading to carbon shortages) and aromatic amino acids stop being produced; the plant eventually dies from starvation. The EPSPS enzyme active site is highly conserved across higher plant families, enabling glyphosate to affect a broad spectrum of weedy plant species (Garg et al. 2014). Several weed species have independently evolved TSR to glyphosate via a single amino acid change in EPSPS position 106 (see e.g. Christoffers and Varanasi 2008 for a review) that reduces the binding affinity of glyphosate but does not eliminate it completely. Thus, the resistance conferred by binding site changes in EPSPS is relatively weak (~2-4 times that of susceptible plants); in contrast, recently documented cases of NTSR confer higher levels of resistance (~8-12 fold) (Perez-Jones et al. 2007). To date, reported NTSR mechanisms are varied and include reduced translocation (e.g. Koger and Reddy 2005; Nandula et al. 2008), active vacuolar sequestration (Ge et al. 2010), limited cellular uptake (Ge 2013) and a rapid necrosis response (Robertson 2010). No alleles involved in NTSR have been identified yet (Délye 2013), but transmembrane transporter proteins are suspected to play a role in reduced translocation and sequestration.

Glyphosate resistance was first identified in wild, annual sunflower (*Helianthus annuus* L.) populations from the US Midwest and Canadian prairies in 2012 glasshouse trials by our lab group (unpublished results), and confirmed in field trials by an industry collaborator, DuPont-Pioneer, at roughly half to two-thirds the 1 kg a.e. ha⁻¹ rate typically applied by a farmer. Sunflower populations from Texas have also been reported to have glyphosate resistance (as of 2015) on the International Survey of Herbicide Resistant Weeds (www.weedscience.org, Heap 2017; retrieved on 10 Oct 2017). Wild sunflower is known to have evolved resistance to other common herbicides, including imazethapyr, imidazolinone, and sulfonyleurea (Massinga et al. 2003). Here, we investigated the genetic basis of glyphosate resistance segregating in our sunflower populations using next-generation sequencing and genome-wide association mapping, a statistical technique that examines phenotype-genotype correlations for each genetic marker. Genomics offers powerful opportunities for identifying the loci underlying evolved herbicide resistance (Stewart et al. 2009), as well as the inheritance, occurrence and movement across the landscape of herbicide-resistance genes. Especially in the case of NTSR, the lack of progress in elucidating underlying mechanisms and loci may be due in part to limited availability of genomic information (Yuan et al. 2007), a situation that will hopefully improve as sequencing continues

to decrease in cost. With the recent availability of a high-quality reference for the sunflower genome (Badouin et al. 2017) created using long reads generated by PacBio sequencing, sunflower mapping studies based on high-density single-nucleotide polymorphism (SNP) data obtained from resequencing are now possible. To our knowledge, ours is the first study to date to utilize whole genome shotgun (WGS) resequencing data to investigate the loci and alleles underlying glyphosate resistance.

4.2 Materials and Methods

4.2.1 Plant Materials

The association mapping population included *Helianthus annuus* accessions from three different sources (Table 4.1): the twenty paired agricultural-weed and non-agricultural wild sunflower populations collected across the Midwest in 2011 (described in Chapter 2), eight weedy populations obtained from Dr. Matt King (DuPont-Pioneer, Iowa) based on targeted collections in 2013 (hereafter the “Matt King [MK]” populations), and pre-breeding materials provided by an industry collaborator. The MK collections were made in the Central USA (Figure 4.2) and represented populations highly likely to possess glyphosate resistance. These populations were found directly infesting fields of Roundup Ready crops in intensively-managed agricultural areas. All eight MK populations were large, containing 1,000 individuals or more. Within a population, seeds were collected from a number of mother plants in a random fashion, for an average of 33 maternal families per population (range = 27 to 55).

In 2013, we shared seed materials with an industry collaborator, DuPont-Pioneer, for field validation of our preliminary glyphosate testing results. The shared materials included seeds from six weedy populations collected for my thesis (IA1A, IA2A, KS1A, KS2A, MO1A, and SD1A), as well as F1 seed from each of three crosses I made in 2012 between a glyphosate-resistant, weedy father (i.e., pollen-donor) and a susceptible cultivated mother. Two fathers were from Kansas (KS2A families 22 and 6) and one from North Dakota (ND1A family 4). To facilitate crosses, all maternal plants (i.e., pollen recipients) were cytoplasmic male sterile (cms) HA 412 individuals; HA 412 is a maintainer germplasm line available from the United States Department of Agriculture (USDA) National Plant Germplasm System (NPGS) collections (PI 603993). After screening all donated materials for glyphosate resistance at a rate of $\frac{2}{3}$ kg a.e. (i.e., acid equivalent) of glyphosate per hectare (i.e., two-thirds the field application rate), collaborators at DuPont-Pioneer allowed the survivors to freely inter-mate. The DuPont-Pioneer pre-breeding materials used here are seeds produced by this process. Note that individuals were of mixed origin and some may represent crop-wild hybrids.

Our goal was to include as diverse material as possible in the association mapping population. With this in mind, we tried to maximize the inclusion of maternal families for all source materials. For the twenty thesis populations, with the exception of “Man” (which was bulked seed; see Table 4.1), we germinated seeds from at least ten maternal families per population (more for populations with lower germination fractions in the common garden, Chapter 2), with the goal of obtaining at least ten individuals per population (i.e., $n = 200$ total), each from a different maternal family. Similarly, we germinated seeds from twelve maternal families for each MK population, with the goal of inclusion of at least ten individuals from unique maternal families in the mapping population ($n = 80$ total). From the DuPont-Pioneer pre-breeding materials, seeds from twenty accessions (i.e., maternal families) were germinated, with the goal of including one individual for each ($n = 20$). Hence, the anticipated size of the mapping population would be 300 individuals or more.

4.2.2 Determination of Glyphosate Resistance Phenotype

We grew seedlings for the mapping study in a glasshouse at the University of British Columbia (UBC) in Vancouver, Canada in summer 2014. Seeds for the populations and maternal families described above (4.2.1 Plant Materials) were scarified on July 10th, 2014 and placed on moist filter paper in petri dishes to imbibe overnight. Seed coats were fully removed on the following day to enhance germination. We watered with a 1% solution of plant preservative media to reduce microbial contamination and changed filter paper daily. Seeds were germinated in the dark; once greening (i.e., chlorophyll production) of any seeds in a given dish occurred, the dish was transferred to the light for the completion of germination. Germination dates were recorded at the level of the family (i.e., we did not keep track of individual seeds), and a seed was considered “germinated” once the primary root was at least 1 cm long and some secondary roots had appeared.

We planted seedlings as they germinated beginning on July 14th, 2014. Seedlings were planted into 5 cm diameter conical “Deepots” (Stuewe & Sons, Inc.), which we had pre-filled with potting soil (a mix of 75% peat with 25% perlite) in sequential order. Planting date was recorded individually. Seedlings were kept on a mist bench initially after planting, to avoid dehydration. Once all seedlings were planted, we fully randomized the experiment on July 20th, 2014, with individual Deepots placed in support trays with 50 slots each (i.e., a seedling density of 269 per m²). Trays were laid out in a large rectangle on a glasshouse flood bench receiving 16 hours of supplemental lighting per day, delivered by 600 W high-pressure sodium lights. We continued to mist by hand twice daily, until seedling roots were long enough

to obtain water from the flood bench. The bench was flooded twice daily with a solution of weak fertilizer water prepared on-site.

We took twice-weekly non-destructive measurements of seedling growth, which included the height (i.e., length of the main stem), number of true leaves and dimensions of the largest leaf. Owing to the diversity of the plant materials included, we observed significant variation in growth rates (as time to reach the four-leaf stage). Hence, we decided to split the experiment into two groups, to be treated with glyphosate on different dates. While not ideal, this approach reduced the variation in leaf number among individuals in each spray group; studies in other species have shown a strong effect of plant size or growth stage at the time of herbicide treatment on the level of resistance observed (see e.g. Shaner 2010; Chauhan and Abugho 2012; Dennis et al. 2016). As seedlings approached the four-leaf stage (our desired target for the time of spraying), they also became crowded and overlapping. As this could have an impact on the amount of glyphosate received (when delivered as a foliar spray), we also decided to increase the spacing at this time (to 134 plants per m²); plants were kept in the same order, but each tray was expanded into two, with Deepots placed in every second available slot.

The first spray group, with seedlings ranging in size from four- to eight-true leaves, was treated with glyphosate on July 31st, 2014, after taking pre-spray photographs of every seedling. Similarly, group two seedlings were photographed and then treated a week later, on August 6th, 2014. Each plant was individually sprayed with 1 mL of glyphosate solution, delivered from a pre-calibrated spray bottle positioned at a constant distance of 20 cm above all plants. We used a commercial formula available locally: *Roundup Concentrated Grass and Weed Control, 1-L, rainproof in 2 hours* (143 g a.e. L⁻¹). As the ingredient list for this product is proprietary, we unfortunately do not know which chemical surfactants (used to enhance foliar penetration, coverage and overall herbicide effectiveness) are included in the formula. The concentration of the solution was calibrated to deliver a rate of 0.5 kg a.e. ha⁻¹ (or half the field rate), calculated based on the size of the Deepot trays. Because some plant leaves spread past the edges of the Deepots, we also delivered 1 mL of spray above each empty slot in a tray, and an additional 34 mL in a perimeter around each tray, to ensure even coverage. Note that seedlings do not receive identical doses with a spray application, however, as plants have different surface areas and angles of leaves (which can cause more or less of the herbicide to drip off). Cultivated sunflowers (HA 89, NPGS PI 599773) were used as positive controls (n = 15), while extra seedlings (from a variety of populations, n = 46) not included in the mapping served as negative controls being sprayed with 1 mL of distilled water only.

Post-spray photos were taken at one week post-treatment and deaths resulting from the glyphosate treatment were recorded daily. Meanwhile, surviving individuals were assessed quantitatively at each of the first, second and third week after treatment. For each week, the regular non-destructive growth measurements were taken (e.g. height, etc.), but additionally the numbers of dead and deformed leaves were recorded, and the presence of any wilting. At three weeks post-treatment, survivors were ranked visually from one to five, with a rank of one indicating very little remaining living tissue, while a rank of five indicated the individual appeared virtually untouched by the treatment; dead plants received a score of zero. The severity of the deformation in new leaves was also ranked at this time from one to four (1 = leaves < 25% of normal size, 2 = 25% < leaf size < 50%, 3 = 50% < leaf size < 75%, and 4 = leaf size >75% of normal size).

I used the statistical software R (version 3.3.0: R Core Team 2017) to calculate summary statistics for the phenotypic data, and to evaluate the relationship between glyphosate resistance and each of plant size, spray group and spray tray, using the *stats* package. Linear models or *t*-tests were used in the case of one categorical and one quantitative variable, as appropriate; quantitative variables were log-transformed as needed to achieve normality in the model residuals. In the case of two categorical variables, a chi-squared test was used. Seedling biomass was determined by inputting non-destructive growth measurements (of seedling height and leaf number) into a previously established linear model (see regression details in Chapter 2: $R^2 = 0.9$, $n = 227$ plants).

4.2.3 DNA Extraction and Library Preparation

Prior to glyphosate treatment in each spray group, we collected leaf tissue from newly expanded leaves when seedlings were at the four- to eight-leaf stage. Tissue was harvested quickly, placed on dry ice shortly after collection and then stored at -80 °C in the lab. For the majority of samples ($n = 255$), I was able to extract genomic DNA cleanly using a modified CTAB protocol based on Murray and Thompson (1980). The remaining samples I extracted using either a QIAGEN DNeasy Plant Mini Kit ($n = 51$) or a DNeasy 96 Plant Kit ($n = 15$) (QIAGEN, Hilden, Germany) which both use silica-based columns. Unfortunately, a side effect of the column-based methods was that genomic DNA was fragmented, as seen by running out samples on an EtBr-agarose gel. A NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham MA, USA) was used to assess sample purity, while DNA was quantified with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham MA, USA) using a Qubit dsDNA Broad Range (BR) Assay Kit (Invitrogen, Carlsbad CA, USA).

High-purity DNA samples were sheared to an average fragment size of 350 bp using a Covaris M220 Focused-ultrasonicator (Covaris, Woburn MA, USA); for column-extracted DNAs that were degraded, the shearing protocol was adjusted following manufacturer recommendations to account for the smaller initial size of DNA fragments. A total of 750 ng of sheared DNA was then used as starting material to prepare a paired-end whole-genome shotgun Illumina library for each sample, using a custom lab protocol based on Rowan et al. (2015), the TruSeq DNA Sample Preparation Guide from Illumina (Illumina, San Diego CA, USA) and Rohland and Reich (2012). Importantly, completed adapters were identical to the Illumina TruSeq adapters. We performed a depletion step on the enriched libraries to reduce the representation of repetitive sequences; libraries were treated with Duplex-Specific Nuclease (DSN; Evrogen, Moscow, Russia) following the methods of Shagina et al. (2010) and Matvienko et al. (2013). As the sunflower genome contains a large fraction of highly repetitive sequences (over 81% of the genome is composed of transposable elements: Staton et al. 2012), this step was important to maximize the capture of useful sequence data. Libraries were re-amplified after the depletion step and six-bp indexes were added at this time (to the P7 adapter) to allow for sample multiplexing.

Enriched, depleted libraries were purified twice with 1.6 volumes of a solution of paramagnetic SPRI beads (prepared according to Rohland and Reich 2012) to clean up any free primers or adapters in our libraries. We quantified libraries using the Qubit BR Assay Kit, and ran them on a 2100 Bioanalyzer instrument using a High Sensitivity DNA Analysis Kit (Agilent Technologies, Santa Clarita CA, USA) to determine average fragment size. Library molarity was assessed on an iQ5 Real Time PCR Detection System (Bio-Rad, Hercules CA, USA). Groups of ten barcoded, randomly-selected libraries were then pooled at equal molarity to be sequenced on a single Illumina lane. All libraries were sequenced at the Genome Québec Innovation Center on either an Illumina HiSeq X or HiSeq 2500 instrument (Illumina, San Diego CA, USA).

4.2.4 Bioinformatics Pipeline

To obtain a set of high-confidence variants (i.e., single-nucleotide polymorphisms, or SNPs) to use in the association analysis, we performed a number of data processing steps. Raw sequencing reads were cleaned using Trimmomatic (version 0.36: Bolger et al. 2014); this step is necessary to both remove technical sequences (i.e., Illumina adapters and barcodes) and low-quality bases, such as those that typically occur at the 3' end of a read. Filtered reads were then aligned to the *H. annuus* reference genome (XRQ assembly: Badouin et al. 2017) using the BWA-MEM aligner (version 0.7.12 : Li and Durbin 2010) with default parameters. Picard (version 2.5, <http://broadinstitute.github.io/picard/>; retrieved 25

Sept 2017) was used to mark duplicate reads, which can result from sequencing the same DNA fragment more than once; duplicate DNA fragments arise during the PCR amplification step of library construction and are usually removed to avoid biases in variant calling (Ebbert et al. 2016). Potential indels were identified and realigned using GATK (version 3.6: McKenna et al. 2010) to remove alignment artifacts. A number of libraries were sequenced twice by Genome Québec, resulting in two alignment files per individual; this occurred either because a flow cell under-performed in the first run and libraries were re-sequenced (on the HiSeq 2500, n = 69), or because the same pool of libraries was sequenced in different flow cells at the same time (as a common operating practice on the HiSeq X10, n = 79). For these individuals, data were combined prior to variant calling.

We used a haplotype-based method, as implemented by the Bayesian genetic variant detector FreeBayes (version 1.1.0: Garrison and Marth 2012), to identify single nucleotide polymorphisms. Variants that were located in transposable elements (TEs) were filtered out by referring to an annotation of repeated elements within the reference *Helianthus annuus* XRQ assembly (Badouin et al. 2017). Low-quality SNPs were also removed on the basis of quality metric cut-offs obtained using a set of validated SNPs from a SNP-chip (Mandel et al. 2013), as described previously in Chapter 3. Briefly, SNP-chip variants observed in the SNP data were considered validated, and the distribution of site quality metrics was compared between validated and unvalidated SNPs to determine meaningful cut-offs, with the following filters used: mapping quality score > 20, mean mapping quality of observed alternate alleles > 40, and allele balance at heterozygous sites between 0.4 and 0.6. Additionally, only biallelic SNPs on mapped scaffolds were considered for this study, and we required a minor allele frequency (MAF) > 0.05 for each locus included in the association analysis. Filtering on MAF is standard in genome-wide association studies, as when MAF is very small most individuals have two copies of the major allele, and this results in low power to detect an effect of the SNP on the trait of interest (Reed et al. 2015).

4.2.5 Genome-Wide Association Study

To identify SNPs correlated with glyphosate resistance we performed a genome-wide association study (GWAS). Association mapping is one of two commonly used methods for identifying quantitative trait loci (QTL) underlying a phenotypic trait of interest; the other is QTL-mapping. Unlike QTL-mapping, which relies on recombination events in controlled crosses, association mapping takes advantage of natural recombination events (that have occurred over evolutionary history) and does not use controlled crosses (Yu and Buckler 2006). Thus, association mapping represents a quicker approach,

as only a single generation of plants is needed, and may also capture a wider diversity of loci and alleles due to the variety of plant germplasm included; a QTL-mapping population, in contrast, is initiated with a cross between a single pair of individuals divergent in the phenotype of interest (Hall et al. 2010). Additionally, linkage disequilibrium (LD), or the non-random association of alleles between genetic loci, will be lower in association mapping than QTL-mapping as linkage blocks will be smaller, allowing QTL to be mapped with greater resolution (Myles et al. 2009).

While GWAS has many advantages, it is also highly susceptible to spurious genotype-phenotype associations resulting from population structure (Bouaziz et al. 2011), such as that which occurs when subgroups within the mapping population possess systematic differences in allele frequencies. False positives occur because both patterns of genetic relatedness among individuals in the mapping population and genetic structure among mapping subpopulations can create linkage disequilibrium (LD) between unlinked loci (Lander and Schork 1994). Many methods have been developed to deal with these confounding effects, including those of “genomic control” (Devlin and Roeder 1999), structured association (Pritchard et al. 2000) and the use of principal components analysis (PCA: Price et al. 2006) to capture information on structure. Recently, however, approaches based on linear mixed models have gained popularity (Eu-ahsunthornwattana et al. 2014) as they are better able to reduce the false-positive rate while maintaining statistical power (e.g. see Zhao et al. 2007; Myles et al. 2008). To control for population structure and genetic relatedness in our GWAS, we combined the use of PCA with a mixed-model approach.

We first performed a PCA on our SNP data for use in accounting for genetic structure in the GWAS. Principal components analysis is a technique for reducing the dimensionality of a dataset that has long been used to infer population structure in genetic data (Price et al. 2010). In PCA, a new set of axes through the data are created such that the first principal components axis explains the most variation in the data, while subsequent axes (which must all be orthogonal, i.e., uncorrelated) explain progressively less variation; the total number of components equals the number of dimensions in the data, i.e., the number of individuals. Including the top principal components (PCs) in GWAS has been shown to effectively correct for population structure (e.g. Price et al. 2006; Hinrichs et al. 2009; Bouaziz et al. 2011). We used the R package *SNPRelate* (Zheng et al. 2012) to conduct a PCA, as its optimized algorithm is many times faster than competing implementations. Prior to PCA, we pruned the dataset heavily to use only a subset of SNPs in approximate linkage equilibrium, following best practice recommendations to avoid influential SNP clusters in the PCA (Laurie et al. 2010). The *SNPRelate*

function “snpGdsLDpruning” was used for pruning with the $ld.threshold = 0.2$; this function works on one chromosome at a time to recursively remove SNPs within a sliding window based on pairwise genotypic correlations. The PCA was then performed using the function “snpGdsPCA” and results plotted using the *ggplot2* package (Wickham 2009).

As PCA may not always capture the full complexity of genetic relationships in a mapping population, we additionally calculated a Balding-Nichols kinship matrix for inclusion in our GWAS, to provide additional correction for hidden relatedness. Especially as our mapping population included not only materials from widely-separated geographic locations, but also replicates from within populations (which may share varying degrees of ancestry) and, in some cases, multiple individuals from the same maternal family (which are at least half-sibs), we expected patterns of genetic relatedness to be complex; there are also a subset of individuals that may be part cultivar. Mixed-model approaches, as pioneered by Yu et al. (2006) for use in GWAS, are able to incorporate a kinship matrix (K) as a random effect in the model, allowing for a more detailed description of the relatedness of individuals. We calculated kinship coefficients between all pairs of individuals on the basis of the full SNP dataset (filtered for $MAF > 0.05$), using the Efficient Mixed-Model Association eXpedited (EMMAX) software (beta version: Kang et al. 2010).

To identify loci associated with glyphosate resistance as determined by visual scores (assigned three weeks post-treatment), we fit a linear mixed model for each SNP across the genome in turn using the EMMAX software. The model included fixed effects of the alleles at a given locus (coded numerically as the minor allele count for a given individual, whether 0, 1 or 2) and of population structure as captured by the first two PCs, which were included as covariates; K was included as a random effect. As fitting the full linear mixed model genome-wide would be computationally intractable, EMMAX makes use of several approximations (Kang et al. 2010). In essence, the contribution of genetic structure to the phenotype is estimated only once using a variance component model, based on restricted-maximum likelihood (REML). This produces a phenotypic covariance matrix incorporating the effect of genetic relatedness, which is then used globally across all markers. The model thus reduces to a generalized least square (GLS) F -test for each marker. Results were visualized using the *qqman* package (Turner 2014) in R.

4.2.6 Significance Testing and Genes of Interest

After performing the GWAS in EMMAX, we evaluated the resulting p -values for each SNP to look for outliers. The question of what strength of evidence to consider significant in GWAS, and how to

account for the massive number of tests performed, remains an issue of active research lacking in clear directives (Dudbridge and Gusnanto 2008; Johnson et al. 2010). For example, using a significance cut-off of $p = 0.05$, roughly 5% of all SNPs will randomly appear to be associated with the trait of interest, which represents hundreds of thousands of false positives when considering millions of SNPs. However, traditional Bonferroni correction, in which the p -value is adjusted by dividing by the number of tests, assumes that tests are independent, but this is not the case for dense SNP datasets as SNPs in close proximity may be linked; this results in testing being overly conservative (Duggal et al. 2008). Instead, the total number of SNPs might be replaced with the “effective number of independent SNPs” (M_e) in the Bonferroni correction (e.g. Gao et al. 2010; Li et al. 2012), but methods for determining M_e are still in development. Methods to instead control the false discovery rate (FDR: Benjamini and Hochberg 1995) have also been suggested, and we implemented this approach for our data using the *qvalue* package (Storey et al. 2015) in R. While π_0 , or the overall estimate of the proportion of true null hypotheses, averaged 0.996, suggesting a very small proportion (100% - 99.6% = 0.4%) of SNPs truly associated with glyphosate resistance, FDR correction resulted in no significant p -values.

True genotype-phenotype associations may not reach statistical significance in GWAS for a variety of reasons (e.g. Schork et al. 2013; Sham and Purcell 2014; Shin and Lee 2015). As there were suggestive “peaks”, i.e., columns of nearby SNPs that all show the same signal, on several chromosomes (see section 4.3.2), we decided to investigate SNPs falling above a “suggestive” p -value cut-off of $p < 1 \times 10^{-5}$, a common cut-off used in other GWAS studies (Stranger et al. 2011). For these SNPs, we focused on cases where multiple correlated SNPs had suggestive p -values, rather than singleton SNPs, which may represent genotyping errors (Reed et al. 2015). We then identified all genes within 100 kbp of these SNPs using bedtools (version 2.25.0: Quinlan and Hall 2010) to query each region against the annotated XRQ assembly of the *H. annuus* reference genome (Badouin et al. 2017). When GenBank GenInfo Identifier (gi) numbers were available, records were pulled from the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/genbank/>; retrieved on 14 Oct 2017) database using the NCBI REST API to obtain protein names.

4.3 Results

4.3.1 Glyphosate Treatment Produced a Variety of Phenotypic Effects

The association mapping population included a total of 321 individuals (Table 4.1): 206 from the 2011 thesis collections, 95 from the 2013 targeted collections, and 20 from the DuPont-Pioneer pre-breeding materials. Owing to poor germination in some thesis collection populations, we were not

always able to get ten individuals from unique maternal families. In these cases, we included a second individual from families with extra germinants. Thus, all populations included at least ten individuals, with the exception of the Manitoba weedy population (“Man” from the USDA) that had very low germination. Owing to differences in growth rate, individuals were treated with glyphosate on two different dates (see section 4.2.2). The first spray group contained the majority of individuals ($n = 274$), while the second spray group was much smaller ($n = 47$).

Our glyphosate treatment of the genome-wide association mapping population was effective in producing segregation in resistance (see e.g. Figure 4.3); this was independently validated in field trials by collaborator DuPont-Pioneer at higher glyphosate application rates. In our greenhouse study, at the applied rate of $0.5 \text{ kg a.e. ha}^{-1}$, not only did positive controls (cultivars) die, but also a proportion of the individuals in the mapping population. Meanwhile, negative controls sprayed with distilled water survived with no visible damage, implying that, apart from the herbicide treatment itself, there were no other causes of damage to experimental plants such as disease, pests, or under- or over-watering or fertilizing. A total of 69 individuals died within the month after treatment, or 21.5% of the mapping population. Looking at source materials separately, 15% of DuPont-Pioneer individuals died, 20% of MK individuals and 23% of thesis individuals. The effect of source on survival was not significant, however, as determined by a chi-squared test ($\chi^2 = 0.84$, $df = 2$, $p = 0.66$), though the test may be unreliable as a result of the low numbers of DuPont-Pioneer samples. Similarly, we did not find significant differences in survival among populations ($\chi^2 = 35.8$, $df = 28$, $p = 0.15$), despite much greater variation in death rates (from 0% in MK1 & MO1A, to 44% in SD1W). Note that this study had low replication (only ~11 individuals/population), however, which makes looking at inter-population differences difficult.

Among the surviving individuals, glyphosate resistance proved very challenging to phenotype quantitatively owing to the complex and varying responses of individual plants (Figure 4.3 and Figure 4.4). While some individuals appeared virtually untouched by the herbicide, others showed some combination of the following symptoms: wilting, bleaching, leaf die-off, necrotic patches, death of the apical meristem, deformation/stunting of new leaves, premature budding and excessive branch production. No buds or branching were seen in negative control seedlings; typically, both appear later in the sunflower life cycle. Symptoms also developed over time, with seedlings often on disparate time courses. While for example, some seedlings appeared fine initially (e.g. in the one week post-spray photos) they later wilted and died; in contrast, other seedlings that experienced massive leaf die-off early on, later produced new healthy leaves and recovered. As such, we focused on assessments made

of the phenotype after a full three weeks; by this point, plants that were severely damaged were no longer capable of recovering, whereas plants that were recovering only continued to improve. This was confirmed by following seedlings for an additional few weeks. Visual assessment, and ranking of the survivors in terms of their recovery at three weeks post-spray, was deemed to be the most robust way of assessing overall herbicide resistance; plants with low scores (1) had very little living tissue, whereas intermediate scores (2 or 3) had both severe deformation as well as leaf loss/necrosis, and finally high scores (4 or 5) had mostly intact, normal-sized leaves. Scores assigned by different observers were largely concordant, and any disagreements were individually re-assessed and resolved.

We evaluated the influence of potential confounding variables in our mapping population, including the effects of plant size at the time of glyphosate treatment, experimental tray and spray group. Despite attempts to standardize as much as possible, wild materials were diverse and grew at different rates, translating into variability in plant size at the time of spraying. Leaf number ranged from 2 to 13.5 (mean = 6), height from 6.9 cm to 31.5cm (mean = 18.6 cm), and estimated biomass ranged from 0.06 g to 1.52 g (mean = 0.29 g). However, plant size, whether measured as leaf number (LN), height (H) or biomass (B), was not related to herbicide resistance as determined by survival (LN: $p = 0.06$, H: $p = 0.17$, B: $p = 0.47$) or herbicide score (LN: $p = 0.94$, H: $p = 0.92$, B: $p = 0.91$) (Figure 4.5). Note that the few individuals with the most leaves (LN > 8, $n = 8$), which were not necessarily the tallest plants, all survived, and this may be the cause of the marginally significant p -value for the LN ~ Survival model. There was also no effect of the experimental tray on survival ($\chi^2 = 19.1$, $df = 14$, $p = 0.14$), indicating that the glyphosate application was consistent across trays (which were treated one at a time). Lastly, survival was higher in the second spray group compared to the first ($\chi^2 = 4.6$, $df = 1$, $p = 0.03$), and plants were also less damaged ($\chi^2 = 19.4$, $df = 9$, $p = 0.02$). Plants in the second spray group were also smaller than those in the first at the time of spraying (0.16 g of estimated biomass versus 0.31 g: means differ according to a t -test, $p < 0.0001$), though they had the same number of leaves on average ($p = 0.93$).

4.3.2 GWAS Identified SNPs Suggestively Associated with Glyphosate Resistance

Our bioinformatics pipeline and additional filtering steps resulted in a final dataset of 7,556,414 biallelic SNPs. Given the 3.0 Gb size of the *Helianthus annuus* genome assembly (Badouin et al. 2017), this should translate into roughly one SNP per 397 bp. In actuality, biallelic SNPs averaged one per 396 bp \pm 4.5 bp (standard error) across all chromosomes, but as the distribution of distances between SNPs was highly right-skewed, the median = 44 bp provides a better idea of the spacing between adjacent

SNPs. The maximum distance between SNPs ranged from 137 kbp (chromosome 7) to 438 kbp (chromosome 12), suggesting that there are large regions of the genome where few SNPs were called, likely as they contained highly repetitive sequences.

To correct for the effects of genetic relatedness and population structure in the GWAS, we calculated pairwise Balding-Nichols kinship coefficients among all individuals using the SNP data and also performed a PCA to characterize genetic structure (Figure 4.6). The PCA included 422,090 SNPs after trimming for LD, or about 5.6% of our total dataset. This implies that many SNPs in the full dataset are linked and hence not independent in the GWAS. The amount of variance explained by each individual PC axis was low, with the first four PCs, for example, explaining only 1.08%, 0.81%, 0.75% and 0.71% of the variance, respectively. The first PC axis correlates roughly with latitude (Kendalls' $\tau = 0.52$, $p < 0.001$), while PC2 separates the four populations from Kansas from the rest, though there are a few individuals from other U.S. states in the Kansas group. Thus, there is some evidence for subtle isolation-by-distance across our sunflower populations to account for in the GWAS.

Examination of a quantile-quantile plot (QQ plot: Figure 4.7) and of the genomic inflation factor (λ) indicate that genetic structure was adequately corrected for in the GWAS. The QQ plot illustrates the relationship between the expected (x -axis) and observed (y -axis) distributions of the $-\log(p\text{-value})$ s for all SNPs. The data generally fall on the $y = x$ line indicating that there is no systemic bias. If the points were to be shifted up from the line, for example, this could indicate inflation by population structure or familial relatedness (Reed et al. 2015). A number of points in the upper right tail do deviate from the reference line, and these suggest crudely the presence of true associations in the data. The genomic inflation factor λ , defined as the median of the chi-squared test statistics (as calculated from the p -values) divided by the expected median of the chi-squared distribution (when $p = 0.5$: Power et al. 2016), provides a more formal way of measuring any deviation from the $y = x$ line by the data. If the observed data follow the chi-squared distribution, then it is expected that $\lambda = 1$ (Lo et al. 2016). Here, $\lambda = 1.004587$, further confirming that population structure and relatedness were accounted for in the GWAS. Values of $\lambda < 1.05$ are generally considered benign, though inflation is proportional to sample size (Price et al. 2006).

Results of the GWAS are presented as a Manhattan plot in Figure 4.8; the negative log of the p -value for each SNP is plotted against its genomic position. Hence, high values on the negative log scale indicate low p -values. Evaluation of the GWAS results indicated that no SNPs were significantly associated with glyphosate resistance, according to false-discovery rate (FDR) testing. However, the FDR

testing revealed that a small proportion of SNPs ($1 - \pi_0$ or 0.4%) were expected to have true associations with glyphosate resistance, though the analysis was not able to identify which SNPs. To explore this possibility, we considered SNPs with p -values below a widely-used “suggestive significance” cut-off of $p < 1 \times 10^{-5}$. Of these 64 SNPs, 12 SNPs occurred within 50 kbp of another SNP below the cut-off (mean pairwise distance = 13,231 bp, se = $\pm 6,104$ bp); these SNPs were found in five clusters located on chromosomes 1, 9, 12 and 16 (Figure 4.8 and Figure 4.9). There was a sixth potential cluster on chromosome 12 with SNPs spaced 150 kbp apart that we also considered. Finally, in a separate category, we examined singleton SNPs that similarly formed a cluster (or column in the Manhattan plot) with nearby SNPs, only in these cases the surrounding SNPs did not fall below the p -value cut-off ($n = 14$). SNPs in this category were located on chromosomes 3, 8, 12, 13, 15, 16 and 17.

4.3.3 Genes of Interest for Glyphosate Resistance

For all 28 SNPs of interest (12 + 2 + 14), we obtained a list of potential genes of interest for glyphosate resistance, falling within 100 kbp of one of the SNPs. There were a total of 143 genes identified from the *H. annuus* XRQ genome assembly annotations, or roughly five per SNP. There was only one SNP on chromosome 13 for which no genes were located within 100 kbp. The majority of the genes ($n = 69$) were matches to expressed sequence tags (ESTs) in sunflower, representing transcribed regions with unknown functions. For one EST located on chromosome 15 (7,333,500 bp to 7,342,805 bp), GeneInfo Identifier (gi) numbers for GenBank were provided in the annotation; this sequence had homology to late embryogenesis abundant protein LEA5 in other species. Otherwise, ESTs had no matches to protein or signature databases. Additionally, 52 genes were described as producing “uncharacterized proteins” and thus also have no known function. Searching through gi numbers available for some of these entries, I was able to obtain a putative function through homology to identified proteins in other plant species ($n = 21$); note these are not validated for sunflower. Finally, a total of 22 genes were linked to a probable protein match, or category of protein. I will focus on these here, as well as the uncharacterized proteins with homology to known proteins in other species. The gene list is available as Appendix B.

Within the gene list (Appendix B), entries for the 14 SNPs of greatest interest (falling in clusters with other suggestive SNPs) are highlighted in grey. For the suggestive peak on chromosome 1, we identified two nearby transcription factors, one an auxin response factor (ARF). While plant ARFs may regulate diverse biological processes (Baranwal et al. 2017), many have been implicated in plant stress responses, with a number of ARFs being up-regulated under drought stress in soybean, for example (Ha

et al. 2013). For the peak on chromosome 9, we found an overlapping ABC transporter; ABC-transporters are one of the few gene families that have been definitely linked to non-target-site resistance (NTSR) to herbicides (Yuan et al. 2007; Nol et al. 2012). Near to the three suggestive SNPs in this peak (859 bp to the closest SNP), we also identified BTB/POZ domain-containing protein FBL11, which mediates the ubiquitination and subsequent proteasomal degradation of target proteins (Zulet et al. 2013). Considering the three suggestive peaks on chromosome 12, the first contained cysteine-rich receptor-like kinase 8 (overlapping a SNP in the peak), DNA polymerase and a ribonuclease H protein; the second another ABC transporter (overlapping both SNPs in the peak); and the third an alanine:glyosylate aminotransferase and an unidentified protein belonging to the alpha/beta-hydrolase superfamily. Cysteine-rich receptor-like kinases (CRKs), constitute a large subfamily of receptor-like protein kinases, with 44 CRKs in *Arabidopsis thaliana* (L.) Heynh., for example (Burdiak et al. 2015); CRKs play essential roles in signal transduction in plants. Finally, for the last suggestive peak on chromosome 16, only ESTs were identified in the vicinity (n = 8) and no known proteins. It is important to note that ESTs and uncharacterized proteins (ucprot) were also found in the vicinity of the other SNP clusters (Chr01: 3 ucprot; Chr09: 1 ucprot, 3 ESTs; Chr12: 7 ucprot, 12 ESTs).

Considering the remaining suggestive SNPs, i.e., those not in the six primary clusters, there were a number of other genes of interest found overlapping or in the vicinity of the SNPs. These included additional CRKs, which have been suggested to play important roles in the regulation of pathogen defenses (Wrzaczek et al. 2010), but also to be upregulated under osmotic stress in *Arabidopsis* seedlings (Skiryycz et al. 2011). Some CRKs have also been implicated in signalling pathways responsive to herbicides; for example, CRK21 and CRK42 were upregulated in *Arabidopsis* upon treatment with paraquat (Han et al. 2014). Both erf domain (chromosome 16) and C2H2 zinc finger protein (ZFP) (chromosome 3) transcription factors may be reactive to environmental stressors such as drought, salt, herbicides and oxidative stress (Rashid et al. 2012; Liu et al. 2015). For example, erf domain transcription factors were upregulated upon treatment of *Arabidopsis* seedlings with glyphosate (Abdeen and Miki 2009). Transmembrane transporters belonging to the major facilitator family (chromosome 12) have been implicated in herbicide resistance due to altered translocation patterns; in bacteria, high-level glyphosate resistance can be achieved via overexpression of the *yhhS* gene encoding a major facilitator transporter involved in drug efflux (Staub et al. 2012). Several identified proteins, in addition to FBL11 (discussed above), participate in the ubiquitin-mediated proteolysis (i.e., protein degradation) pathway, including: DNA damage-inducible protein 1-like (chromosome 12: Nowicka et al. 2015), E3 ubiquitin ligase RBR (chromosome 3: Chen et al. 2014), and phosphatidylinositol 3-/4-kinase (chromosome 12:

Galvão et al. 2008). This pathway may be critical for abiotic stress responses in plants (Guo et al. 2013), and recently, an ubiquitin ligase was identified as playing an important role in the metabolism of ALS and ACCase herbicides in *Lolium multiflorum* Lam. (Mahmood et al. 2016). Further examples of identified proteins linked to abiotic stress tolerance include: nuclease HARBI1, which is upregulated under salt-stress in *Reaumuria trigyna* Maxim. (Dang et al. 2014); ribonuclease H protein At1g65750, downregulated in nickel-resistant white birch, *Betula papyrifera* Marshall (Theriat and Nkongolo 2017); LEA5, expression induced under drought, heat or salt stress in citrus seedlings (Naot et al. 1995); and glycine-rich proteins, which can function to enhance stress tolerance, especially under drought (Kim et al. 2008).

Interestingly, no suggestive peaks or singleton SNPs were found to overlap with either of the copies of the gene targeted by glyphosate (*EPSPS*) in the sunflower reference genome. Work by our lab group has previously identified two complete copies of *EPSPS*: one on chromosome 4 (gene ID = HanXRQChr04g0122731: 162,864,991 bp to 162,872,274 bp) and a second on chromosome 16 (gene ID = HanXRQChr16g0520001: 131,884,753 bp to 131,890,330 bp). While suggestive SNPs were identified on chromosome 4 (see Figure 4.8), these were discounted as singletons: nonetheless, the closest SNP to *EPSPS* was located at a distance of 5.4 Mb. Similarly, on chromosome 16, the closest SNP identified by GWAS was located at a distance of 6.3 Mb. Thus, linkage between either of the *EPSPS* copies and a suggestive SNP identified by GWAS seems unlikely, as LD tends to decay rapidly in sunflower (see Chapter 3 and Mandel et al. 2013).

4.4 Discussion

In this study, we investigated the genetic basis of glyphosate resistance segregating in wild sunflower populations collected from the U.S. Midwest and the Canadian Prairies. We used whole genome shotgun (WGS) resequencing to obtain SNP data for 321 individuals from 28 populations. Seedlings at roughly the four- to eight-leaf stage were phenotyped for glyphosate resistance at a rate of 0.5 kg a.e. ha⁻¹, and assigned a score based on their recovery post-herbicide application. We used genome-wide association mapping to look for genotype-phenotype associations. To our knowledge, ours is the first study to implement a GWAS approach to investigate glyphosate resistance using genome-wide SNP data. We found 64 SNPs with suggestive associations ($p < 1 \times 10^{-5}$) to glyphosate resistance, with a subset of 28 of these SNPs occurring in columns in the Manhattan plot (with other nearby SNPs). Examining these 28 SNPs, we identified nearby genes potentially linked to resistance, including two ABC transporters and a major facilitator transporter, several transcription factors (zinc

fingers of C2H2 and GATA types), signaling molecules (e.g. CRKs and other protein kinases) and enzymes involved in the ubiquitin-mediated proteolysis pathway (e.g. ubiquitin ligase RBR), among others. Further work is needed to validate if and how these genes of interest may mediate glyphosate resistance in sunflower, but non-target-site mechanisms of resistance, such as altered translocation patterns or sequestration of the herbicide, seem likely. Interestingly, no suggestive SNPs were located near either of the two copies of *EPSPS* in the sunflower genome, though further work is needed to analyse the *EPSPS* sequences in each individual before target-site resistance can be ruled out.

4.4.1 Glyphosate Resistance in Wild Sunflower, *Helianthus annuus*

Our 2012 finding of widespread glyphosate resistance in wild sunflower populations was novel, if perhaps not unexpected. The regions of the USA from which I collected populations in 2011 represent areas of intense agricultural use, with large areas planted in monocultures and extensive use of agricultural chemicals. For example, across my sampling transect, agricultural lands were treated with an average of 88 lbs of glyphosate per square mile or more in 2011, according to county-level estimates made by the United States Geological Survey (USGS, <http://water.usgs.gov>; retrieved on 14 Oct 2017). These values fall into the highest category recognized by the USGS, and represent more than a quadrupling of the amount of glyphosate applied 20 years ago. Truly, glyphosate has single-handedly replaced much of the former diversity of herbicides used on the landscape (Benbrook 2016). Speaking to landowners during sample collection revealed that glyphosate may also be applied to non-agricultural lands, such as ditches, fallow areas, roadsides and other waste places, to remove weedy species, such as sunflower, that may compete with crops. In the state of Iowa, *H. annuus* is considered a noxious weed under state law (despite its status as a native species) and must be destroyed when found, even on private lands (Iowa Department of Agriculture and Land Stewardship, <http://www.weeds.iastate.edu/reference/weedlaw.htm>; retrieved on 14 Oct 2017). Hence, there may be strong selection pressure for glyphosate resistance across the sampling transect, for both crop weed populations of *H. annuus* and wild populations growing in more natural areas.

Here, most sunflowers in the mapping population (78.5%) survived a glyphosate application of 0.5 kg a.e. ha⁻¹, or half of the rate typically applied on fields of Roundup Ready crops. Additional glasshouse trials (unpublished) found that a majority of tested seedlings also survived a higher application rate of two-thirds the field rate ($\frac{2}{3}x$), but few seedlings survived a second application at $1\frac{1}{2}x$. Note that many factors affect glyphosate efficacy, such as the light intensity, air temperature, and relative humidity at the time of spraying (Waltz et al. 2004), as well as the growth stage, nutrient and

water status of the plants (Shaner 2010), and hence independent trials may not be strictly comparable. Additionally, glasshouse and field trials may also find different results; for example, in Palmer amaranth (*Amaranthus palmeri*), the LD₅₀ (i.e., dose that kills 50% of a population) was twice as high in the field versus greenhouse (Culpepper et al. 2006). Hence, our testing results may not translate into equivalent resistance in a field setting. However, the level of glyphosate resistance observed was not trivial, as all tested controls (cultivated sunflowers) died, even when tested at lower application rates (e.g. ½x). Hence, many tested sunflower populations had, on average, resistance at a level higher than expected from spray drift (typically ~10% the field rate, e.g. Hensley et al. 2013). We were not able to detect significant differences in survival among study populations, implying no strong geographic patterns in the incidence of herbicide resistance. However, survival varied considerably among populations (from 66-100%), suggesting that with greater sample size, differences in resistance may become evident among populations. Survival was significantly higher in the second spray group compared to the first; this may be explained by differences in ambient conditions on the spray dates. Plants in the second spray group also grew more slowly than those in the first, so another possibility is that slower growth translated to lower susceptibility to the herbicide.

Overall, we did not find an effect of sunflower seedling size on glyphosate resistance, when measured as either survival or herbicide score. This contrasts with work in other species that has found a consistent relationship between plant size and glyphosate resistance. In common ragweed (*Ambrosia artemisiifolia*), taller plants at the time of spraying were more resistant than shorter ones (personal communication from K. Hodgins). Similarly, Shrestha et al. (2007) found that the level of resistance observed for horseweed (*Conyza canadensis*) seedlings increased with the number of true leaves. Finally, similar observations have been made for hairy fleabane (*Conyza bonariensis* (L.) Cronquist: Dinelli et al. 2008), Johnsongrass (*Sorghum halepense* (L.) Pers.: Vila-Aiub et al. 2007) and lambsquarters (*Chenopodium album* L.: Schuster et al. 2007), illustrating the need to standardize growth stages in testing for herbicide resistance. We attempted to standardize leaf stage at the time of spraying in this study, though this was challenging owing to the diversity of wild materials (collected across a latitudinal gradient) included in the mapping population. Seedlings were sprayed at a size of approximately four to eight true leaves, consistent with the methodology of other studies (Shaner 2010), and with the anticipated size of treated weedy sunflowers in agricultural fields. Thus, perhaps seedlings were similar enough in size at the time of testing that there was no detectable impact of size on resistance. The observation that the few seedlings with more than eight leaves in the mapping population all survived the glyphosate treatment is consistent with this hypothesis. Spray applications also tend to deposit

spray proportionally to the leaf area of the seedling, meaning that larger plants receive larger doses of glyphosate, and this may also have acted to mitigate any effects of size in our study. However, once sunflower seedlings get much beyond the eight leaf stage, lower leaves begin to be shielded from the spray by upper leaves, and hence plants of very different heights or leaf numbers may actually receive similar doses; the plants with the most leaves in our study would thus have received a proportionately smaller dose for their size.

Fitness costs of resistance may act to limit the level of glyphosate resistance seen in natural populations, perhaps explaining why the levels of resistance we observed were moderate, with very few seedlings surviving applications greater than the field rate. There is strong evidence that some herbicide resistance alleles are associated with pleiotropic effects that act to lower plant fitness in the absence of the herbicide (Vila-Aiub et al. 2009). For example, amino acid substitutions in targeted enzymes (i.e., TSR) may often, but not always, act to reduce enzyme efficiency, at the cost of reduced metabolism (Powles and Yu 2010); however, there are no studies of the fitness costs of the altered *EPSPS* allele in any weed species to date, though work is currently underway in *Eleusine indica* (Yu et al. 2015). Fitness costs can also arise from NTSR mechanisms because of fundamental trade-offs in resource allocation between plant growth, reproduction and defense (Vila-Aiub et al. 2009). As resources are diverted away from other plant functions, resistant individuals are often less competitive than susceptible ones in the absence of the herbicide. Fitness costs have been noted for the reduced translocation mechanisms of glyphosate resistance. In naturally glyphosate-tolerant tall morning glory (*Ipomoea purpurea* (L.) Roth), fewer seeds were produced by more resistant biotypes when grown under benign, herbicide-free conditions (Baucom and Mauricio 2004). Similarly, resistant ryegrass (*Lolium rigidum*) biotypes became less common over time in a mixed population (including susceptible biotypes) when no glyphosate was applied (Wakelin and Preston 2006), and when both resistant and susceptible biotypes were competed against a wheat crop, resistant types produced fewer seeds (Pedersen et al. 2007).

4.4.2 Glyphosate Resistance in Sunflower Likely Involves Non-Target-Site Mechanisms

Our GWAS was unable to detect any significant genotype-phenotype associations when using false discovery rate correction for multiple testing. However, a total of 64 SNPs were considered suggestive of an association with glyphosate resistance using a cut-off of $p < 1 \times 10^{-5}$ (Figure 4.8). A striking feature of the analysis was that no suggestive SNPs occurred near either copy of the *EPSPS* gene in the sunflower genome, suggesting that amino acid changes in the target enzyme may not play a major role in sunflower glyphosate resistance, at least for the populations we surveyed. Before TSR can be

completely ruled out, however, examination of the EPSPS sequences to look for amino acid changes and quantitative PCR to look for *EPSPS* copy number variation would be needed. Amino acid changes in EPSPS have been reported for a total of 7 plant species to date (Gaines and Heap 2017), with the most common mutation occurring at position 106, from proline to either alanine, leucine, serine, or threonine. However, the Pro-106 mutations confer only weak glyphosate resistance (Christoffers and Varanasi 2008), and hence NTSR has generally been considered to be of greater importance for glyphosate resistance in weeds (Powles and Yu 2010). However, in a wild population of *E. indica*, a double amino acid substitution in EPSPS was recently discovered (Yu et al. 2015), combining a mutation at position 102, from threonine to isoleucine, with the Pro-106-Ser mutation (i.e., Thr-102-Ile and Pro-106-Ser, or TIPS). The TIPS mutation confers high glyphosate resistance (600 times that of Pro-106-Ser in *E. indica* in vitro testing) and may evolve sequentially in weeds subject to strong glyphosate selection. But the mutation substantially decreases the catalytic efficiency of EPSPS, and hence likely decreases plant fitness if glyphosate selection is relaxed, meaning that TIPS may be maintained only rarely in nature. Other forms of TSR not explored here include gene duplication and enhanced expression of EPSPS (see Sammons and Gaines 2014a for a review); it is unknown if either plays a role in resistance in our populations.

While TSR to glyphosate has been the focus of research to date, likely due to the greater ease of study, NTSR mechanisms are believed to also play an important role in weed resistance (Powles and Yu 2010). Here, we found a total of 28 SNPs both suggestive of an association with glyphosate resistance and also occurring near to other SNPs showing the same pattern. Together the SNPs were part of 17 clusters, or peaks in the Manhattan plot (Figure 4.8), though four peaks on chromosome 12 were roughly adjacent and so may not be independent; six of the clusters contained multiple suggestive SNPs ($n = 15$ SNPs total). While these peaks are merely suggestive and require further study, it is interesting to note that multiple genomic regions may be implicated in glyphosate resistance. This suggests a quantitative, polygenic basis to resistance. Investigations of NTSR for other herbicides have largely revealed that the genetic basis may be complex, and that individuals can accumulate many resistance alleles, especially in cross-pollinated species (Délye 2013). According to the “allele stacking” model of the evolution of NTSR proposed by Délye (2013), NTSR evolves gradually over multiple generations with different parental resistance alleles accumulating in individuals over time; as alleles continue to accumulate, herbicide sensitivity declines. Under this model, non-lethal applications of herbicide (such as those occurring from spray drift) enable a broader range of NTSR alleles to be selected for, as alleles that would not enable survival at a full dose are now maintained. This model has been experimentally

demonstrated for both ACCase inhibitors (Neve and Powles 2005; Busi et al. 2013) and glyphosate (Busi and Powles 2009). In their experiment, Busi and Powles (2009) subjected a susceptible population of out-crossing *L. rigidum* to recurrent low-dose glyphosate selection; after only three generations, the estimated LD₅₀ had doubled and 33% of individuals were able to survive a glyphosate application at the label rate (1x). The results were consistent with progressive enrichment for minor genes contributing to glyphosate resistance. Biochemical studies of NTSR have also shown that multiple mechanisms may operate, with minor genes playing a role in the NTSR observed (see review in Ghanizadeh and Harrington 2017); at least seven NTSR loci were identified in a single *Alopecurus myosuroides* Huds. plant resistant to ACCase and ALS inhibitors (Petit et al. 2010), for example.

Considering the genetics of NTSR in glyphosate specifically, data are limited. Reduced translocation of glyphosate to meristematic tissues has been found in populations of horseweed (*C. canadensis*) and rigid ryegrass (*L. rigidum*) (Lorraine-Colwill et al. 2002; Koger and Reddy 2005). In both cases, classical studies of inheritance found that resistance is due to an incompletely dominant single nuclear gene (with dominance varying from high to moderate in crossing studies). In tall waterhemp (*Amaranthus tuberculatus* (Moq.) Sauer), in contrast, multiple genes are likely to be involved in glyphosate resistance, as indicated by the variability in herbicide responses seen among individuals after a period of recurrent selection (Zelaya and Owen 2005). Similarly, in field bindweed (*Convolvulus arvensis* L.), the existence of minor genes influencing susceptibility to glyphosate was confirmed using a diallel cross (Duncan and Weller 1987). Finally, some populations of resistant rigid ryegrass have both TSR and also reduced translocation patterns (Preston et al. 2009); these mechanisms combine additively to enhance overall glyphosate resistance. In conclusion, glyphosate resistance may often rely on many genes, especially when multiple modes of resistance are involved, but this is not universal. In sunflower, our GWAS results indicate that six or more NTSR loci are involved in glyphosate resistance. As genomics approaches become more widely employed to fine map loci involved in glyphosate resistance (e.g. Peng et al. 2010), hopefully the loci underlying cases of NTSR will finally be identified.

In the majority of cases of NTSR to glyphosate for which a putative mechanism has been identified, altered translocation patterns have been implicated, although other mechanisms do exist (Preston et al. 2009). For example, in Chilean ryegrass (*Lolium multiflorum*) resistant plants had lower spray retention and reduced absorption of glyphosate through the abaxial leaf surface (Michitte et al. 2007). The EPSPS enzyme targeted by glyphosate preferentially accumulates in meristematic tissues (Shaner 2009), and thus glyphosate must translocate to sites of active growth within a plant to exert its

toxic effects; in susceptible plants, glyphosate is rapidly translocated via the phloem, following the same source to sink pattern as photoassimilates (Perez-Jones and Mallory-Smith 2010). In resistant individuals of *C. bonariensis* (Dinelli et al. 2008), *C. canadensis* (Koger and Reddy 2005), *L. multiflorum* (e.g. Perez-Jones et al. 2007; Nandula et al. 2008) and *L. rigidum* (e.g. Lorraine-Colwill et al. 2002), glyphosate may instead become trapped, e.g. in leaf tissue. Experiments (reviewed in Sammons and Gaines 2014b) have revealed that this reduced translocation is due to the rapid vacuolar sequestration of glyphosate via a transporter mechanism. Characterization of the horseweed transcriptome has identified several putative transporter proteins including a tonoplast intrinsic protein (TIP) and several ABC transporters (Yuan et al. 2010); an expression analysis of ABC transporters found that, for thirteen out of seventeen tested genes, expression increased in resistant individuals sprayed with glyphosate (Peng et al. 2010). The ATP-binding cassette (ABC) transporters are transmembrane proteins localized in most extra- and intracellular membranes (e.g. plasma membrane, chloroplasts, etc.) that actively transport a wide variety of substrates, including drugs, lipids, metals, and metabolites, using ATP hydrolysis (Kang et al. 2011; Lane et al. 2016). As reviewed by Yuan et al. (2007), ABC transporter activity toward herbicides and their metabolites has been well established for crops and model species, but more research is needed in terms of their role in NTSR in weedy species. In this study, we identified two ABC transporters of the pleiotropic drug resistance (PDR) family as well as a major facilitator transporter, and these may represent our best candidates for glyphosate resistance. We also identified a number of proteins linked to detoxification processes, suggesting a potential role of metabolism, as well as several transcription factors that may be needed to coordinate the cellular response upon treatment with glyphosate (Délye 2013).

4.4.3 Conclusions

In our genome-wide association study of glyphosate resistance in populations of wild sunflower ($n = 321$), we detected no significant associations with glyphosate resistance after correcting for the false discovery rate, but 64 SNPs had suggestive associations ($p < 1 \times 10^{-5}$). Peaks in the Manhattan plot, especially those containing multiple suggestive SNPs ($n = 6$ peaks), were explored to identify nearby genes. A variety of transcription factors, signalling molecules and detoxification enzymes implicated in plant abiotic stress pathways were found: these may be co-opted for glyphosate resistance in sunflower. Furthermore, we identified three transporter genes (two ABC transporters and a major facilitator transporter) from families clearly implicated in herbicide resistance in other species. The physiological mechanisms and genetic loci underlying non-target-site resistance (NTSR) to herbicides remain poorly characterized, especially for glyphosate. Thus, our work represents an important step towards

elucidating NTSR loci and is the first that we are aware of to use whole-genome resequencing data for this purpose. It is unfortunate that the GWAS was underpowered and not able to identify significant SNPs, perhaps due to the highly polygenic nature of NTSR in sunflower and the smaller sample size of the study. If different sunflower populations possess different resistance alleles, as seems likely, then GWAS could also fail to identify significant associations (Myles et al. 2009). Furthermore, the 0.5 kg a.e. ha⁻¹ dose we used in screening the mapping population resulted in a ~80% survival rate, much higher than we expected; thus, we were forced to visually assess and rank survivors, rather than simply segregate individuals on the basis of survival. It may be that variability among plants (with a similar resistance level) in the symptoms expressed introduced further variation into the association mapping. Nonetheless, our results implicate varied NTSR mechanisms acting to confer glyphosate resistance in sunflower, on the basis of the number of suggestive peaks in the GWAS and types of genes found overlapping these peaks.

Table 4.1: Description of plant germplasm used in the genome-wide association analysis of glyphosate resistance in wild populations of *Helianthus annuus*, including population identifiers, state/province of origin, crop species infested (for weeds) and number of maternal families and individuals used.

Material Source	Population ID	State of Origin	Crop Infested	Number of Maternal Families	Number of Individuals
Original 2011 Weedy & Wild Thesis Collections (Emily Drummond)	IA1A	Iowa	Soybean	7	11
	IA1W	Iowa	<i>na</i>	10	12
	IA2A	Iowa	Corn	9	12
	IA2W	Iowa	<i>na</i>	10	12
	KS1A	Kansas	Sorghum	10	11
	KS1W	Kansas	<i>na</i>	11	12
	KS2A	Kansas	Sorghum	10	12
	KS2W	Kansas	<i>na</i>	9	12
	Man*	Manitoba (Canada)	Wheat	<i>na</i>	8
	MB1W	Manitoba (Canada)	<i>na</i>	9	11
	MO1A	Missouri	Soybean	10	11
	MO1W	Missouri	<i>na</i>	10	11
	ND1A	North Dakota	Corn	9	11
	ND1W	North Dakota	<i>na</i>	10	11
	SD1A	South Dakota	Corn, Soybean	9	10
	SD1W	South Dakota	<i>na</i>	9	10
	SD2A	South Dakota	Corn	12	13
	SD2W	South Dakota	<i>na</i>	12	13
	SK1A	Saskatchewan (Canada)	Wheat	8	11
	SK1W	Saskatchewan (Canada)	<i>na</i>	8	11
New 2013 Targeted Collections (Matt King)	MK1	Iowa	Corn, Soybean	12	12
	MK2	South Dakota	Soybean	12	12
	MK3	Nebraska	Soybean	11	11
	MK4	Nebraska	Corn	12	14
	MK5	Iowa	Corn	11	11
	MK6	Nebraska	Soybean	11	11
	MK7	South Dakota	Soybean	12	12
	MK8	Iowa	Corn, Soybean	12	12
Pre-breeding Materials (DuPont-Pioneer)	<i>na</i>	mixed	mixed	20	20

* Replacement for original MB1A population whose seeds were sterile. Replacement obtained from the USDA NPGS (PI 592327).

Figure 4.1: Chronological increase in the reported number of cases of herbicide resistance worldwide for eight herbicide sites of action. The site of action refers to the specific process in plants that is disrupted to affect growth and development. All herbicides may be categorized by their site of action, with 26 sites of action currently recognized by the Weed Science Society of America. Note that different types of herbicides have different propensities to select for resistance. Data for the creation of this figure were obtained from publicly available records kept by the International Survey of Herbicide Resistant Weeds (www.weedscience.org) accessed on September 30th, 2017.

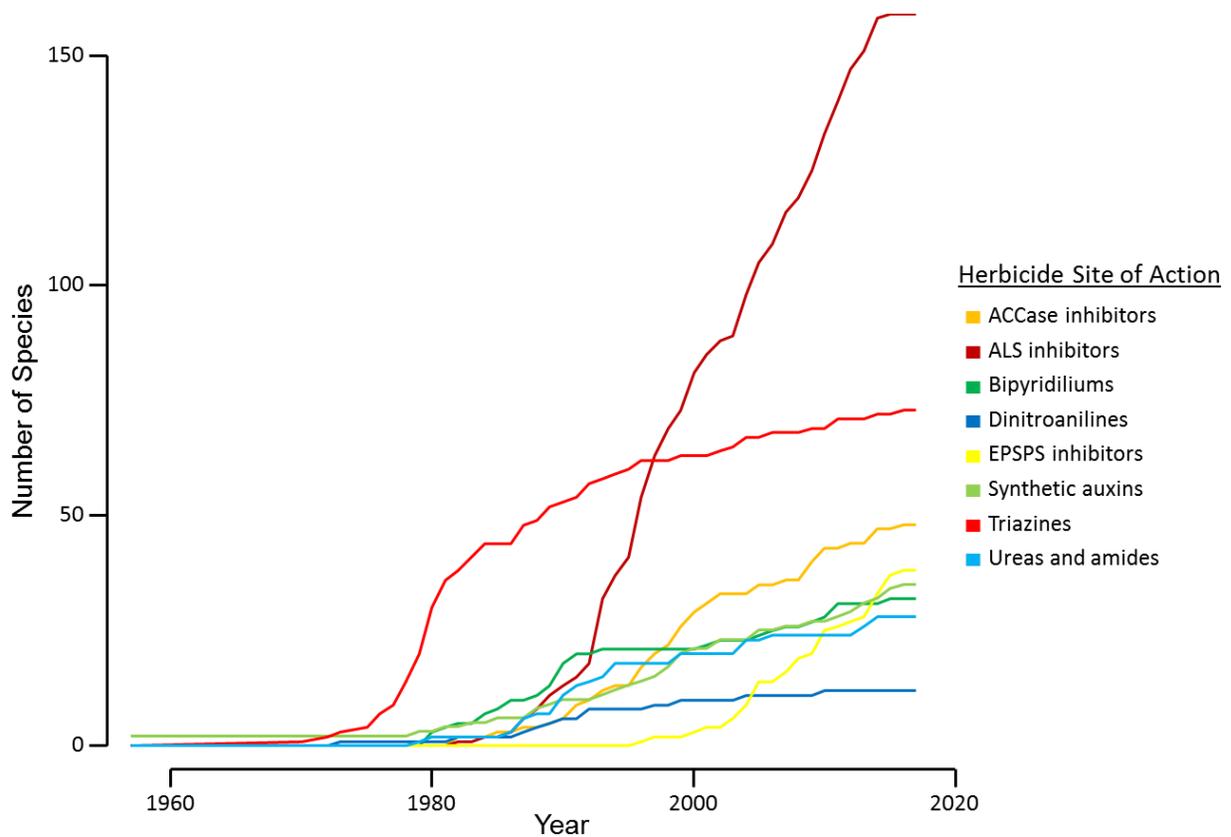


Figure 4.2: North American range of *Helianthus annuus* (based on Rogers et al. 1982) and collection locations of populations included in this study. For the twenty thesis populations, each agricultural-weed population was paired with a nearby population of wild sunflowers in a non-agricultural area, and location names are for the pair. The targeted collections of agricultural-weed populations infesting Roundup Ready crop fields (shown in red) were made by Dr. Matt King and are represented by his initials (“MK”).

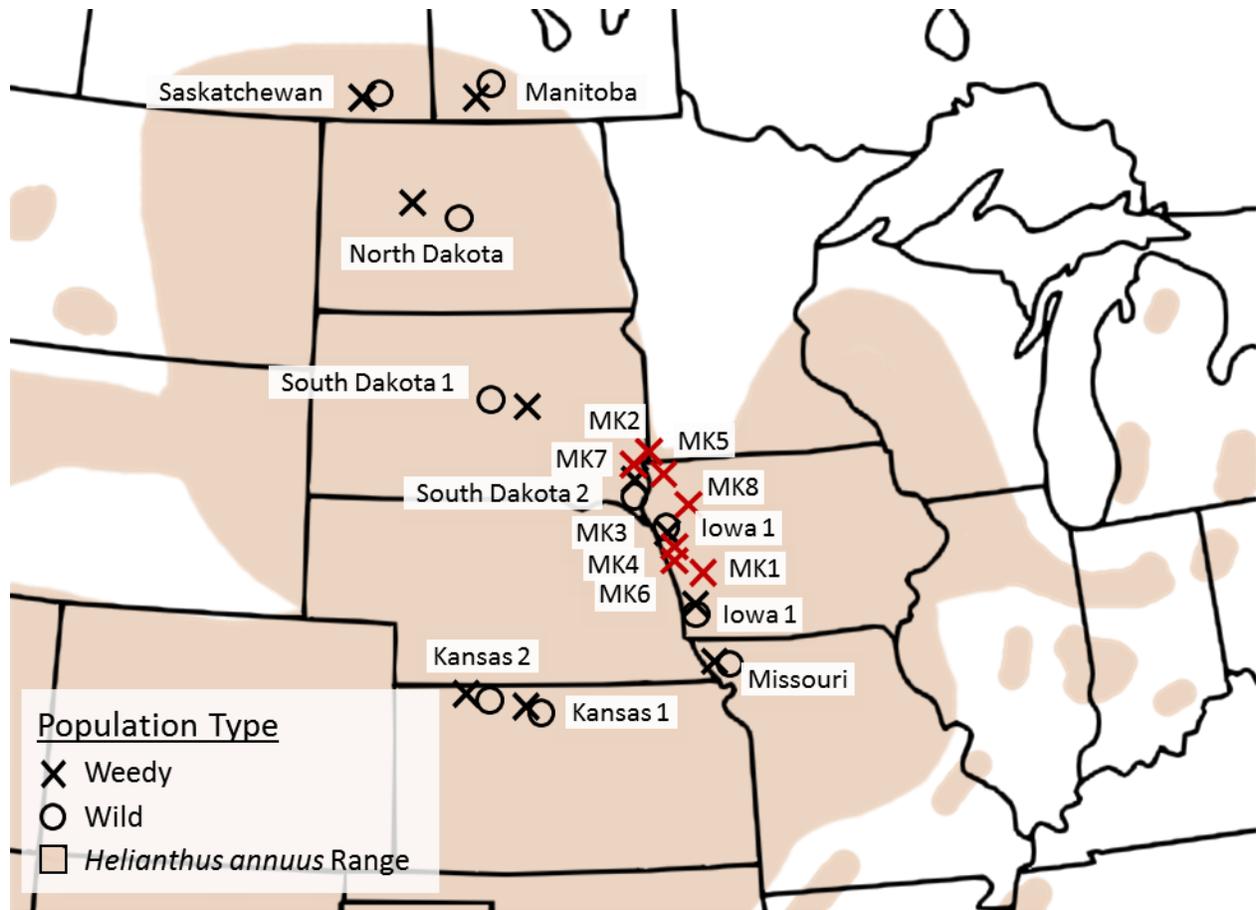


Figure 4.3: Example photos of sunflowers taken (a) immediately prior to glyphosate application and (b) one-week after glyphosate application . Though all individuals pictured here were at the four- or six-leaf stage, size differences are apparent among individuals at a given developmental stage. Damaged leaves (i.e. with clipped edges or hole punches) were the result of tissue collection for the study. Most individuals show some degree of wilting after treatment, necrotic patches and an overall yellowing, especially of younger tissues, which is a hallmark of glyphosate treatment. While some individuals have died by one week post-treatment, others appear virtually undamaged.



Figure 4.4: Photo collage of sunflower seedlings treated with 0.5 kg a.e. ha⁻¹ of glyphosate at varying times after treatment, from this study and others performed by the author. Individuals respond to herbicide treatment in a variety of ways, including wilting, bleaching of growing tissues, leaf die-off and formation of necrotic patches, premature budding and stunting of new leaves. In some individuals where the apical meristem died as a result of treatment, large branches were produced, even in very small seedlings.



Figure 4.5: Violin plots of seedling biomass, as measured just prior to glyphosate application, and the herbicide score assigned by visual assessment three weeks post-treatment. Seedling biomass was determined by inputting non-destructive measurements into a previously determined linear model (see regression details in Chapter 2: $R^2 = 0.9$, $n = 227$ plants). A higher herbicide score equals greater resistance. Horizontal lines represent the mean and 95% confidence interval. There was no relationship between seedling size and herbicide score ($p = 0.91$).

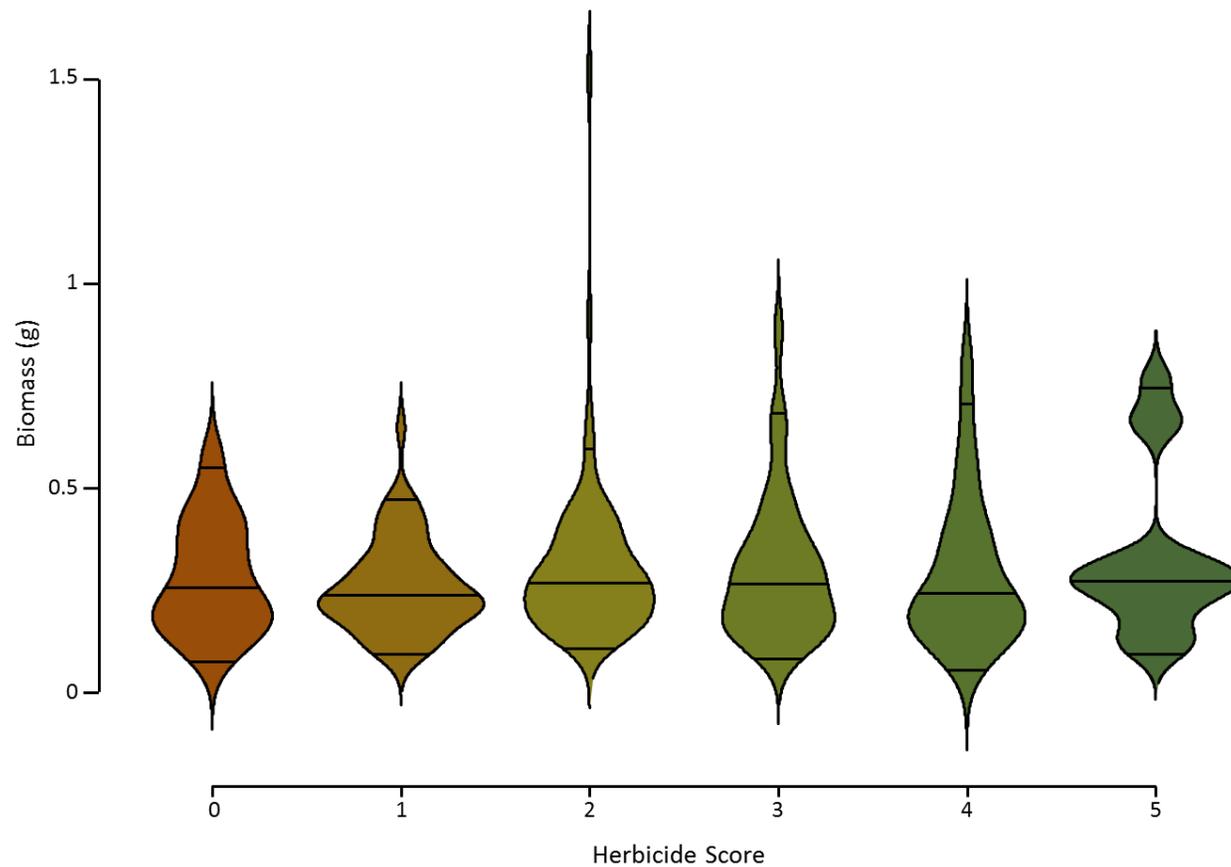


Figure 4.6: First two axes of a principal components analysis (PCA) of the genetic data . Pruning the dataset for linkage disequilibrium (LD threshold of 0.2) resulted in a total of 422,090 SNPs for use in the PCA. Individuals are coloured by population; populations are sorted in order of increasing latitude. Note that individuals of mixed ancestry (DuPont-Pioneer pre-breeding materials) and the replacement population (“Man”) for the weedy MB1A population are not pictured here. The first component roughly separates populations by latitude, while the second separates individuals from Kansas from the rest.

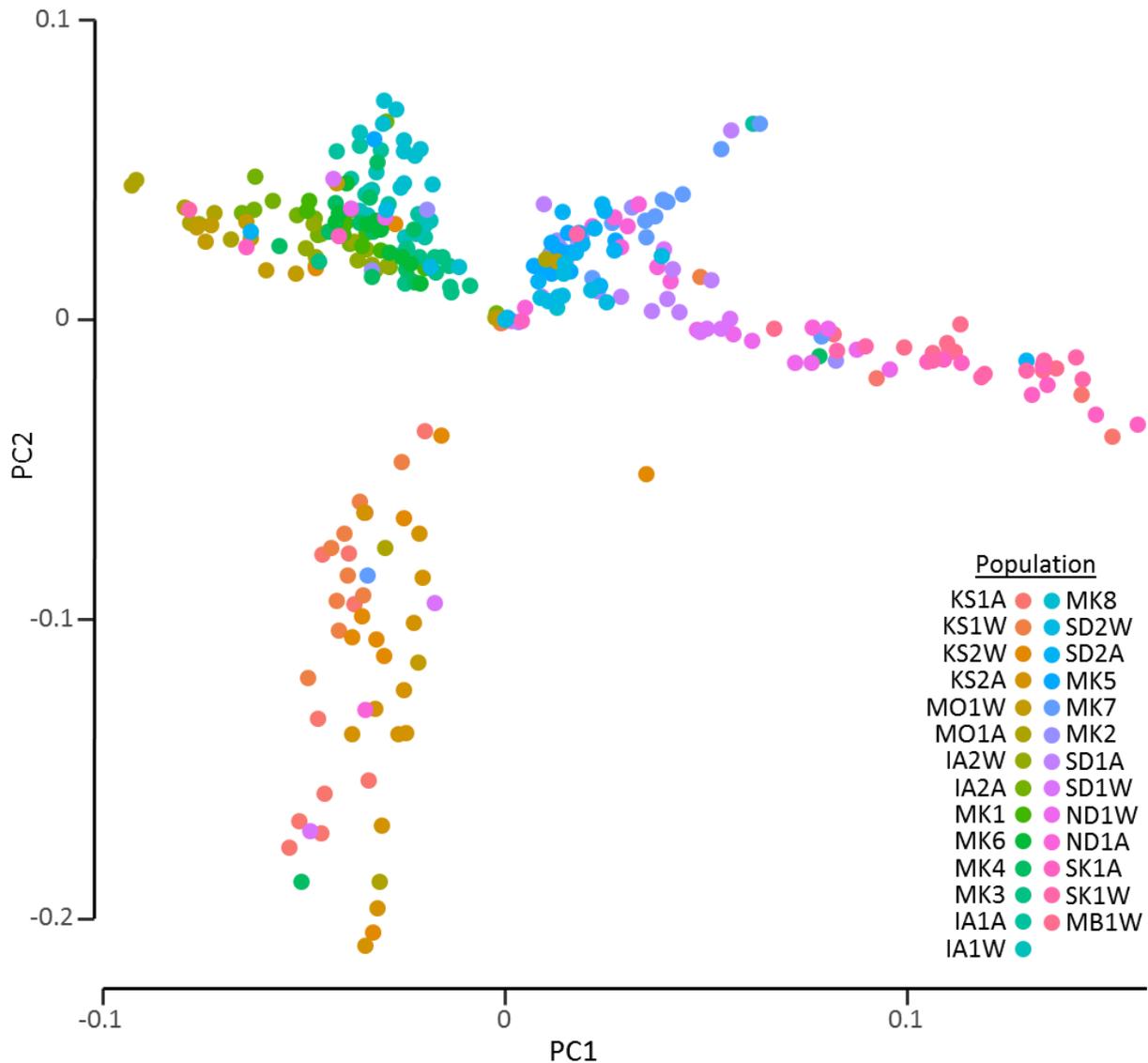


Figure 4.7: Quantile-quantile (QQ) plot for the genome-wide association study (GWAS) showing the relationship between the observed distribution of $-\log(p\text{-value})$ s calculated for each SNP ($n = 422,090$) and the expected distribution under the null hypothesis. The GWAS used the first two principal components and a kinship matrix to account for population structure and relatedness. Departure of observed values from the $y = x$ reference line (shown in red) can reflect systematic inflation in the test statistics owing to population structure; however, polygenicity can also result in departures from the line, but only for those SNPs with high $-\log(p\text{-value})$ s. The data closely follow the reference line, suggesting that genetic structure is not affecting our results.

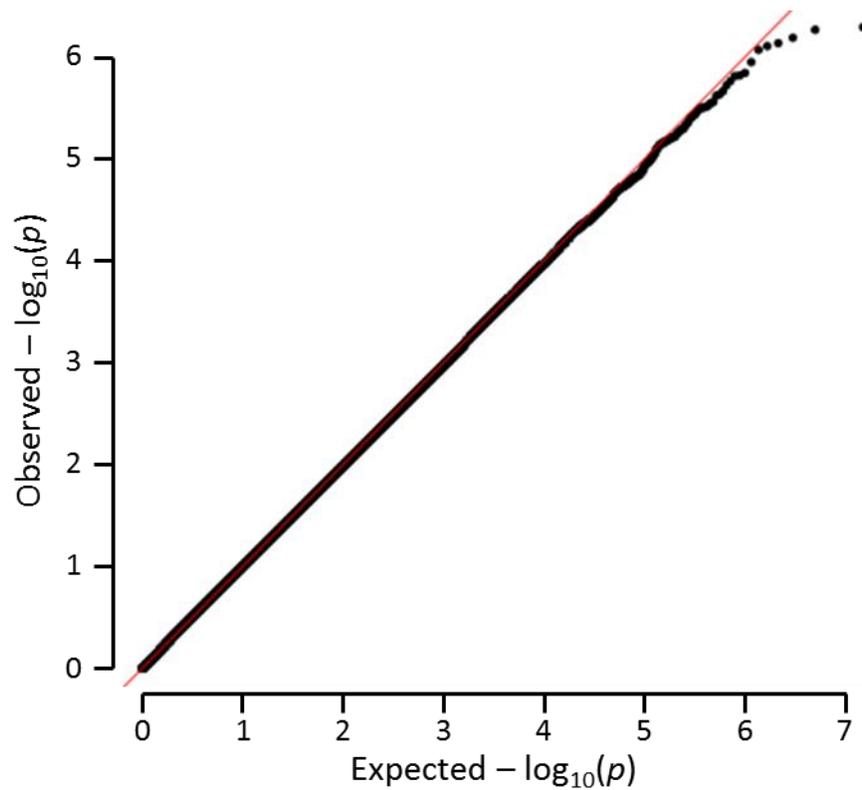


Figure 4.8: Manhattan plot of the genome-wide association study (GWAS) showing the negative of the log p -value of each variant (i.e., SNP, $n = 7,556,414$) against its genomic position. Note that large values on the y -axis correspond to small p -values. The blue horizontal line represents a “suggestive significance” threshold at $p = 0.00001$, with SNPs falling above this line suggestive of an association. While no SNPs were significant according to strict Bonferroni correction ($p < 5 \times 10^{-8}$), 64 SNPs located across the genome fell above the suggestive line.

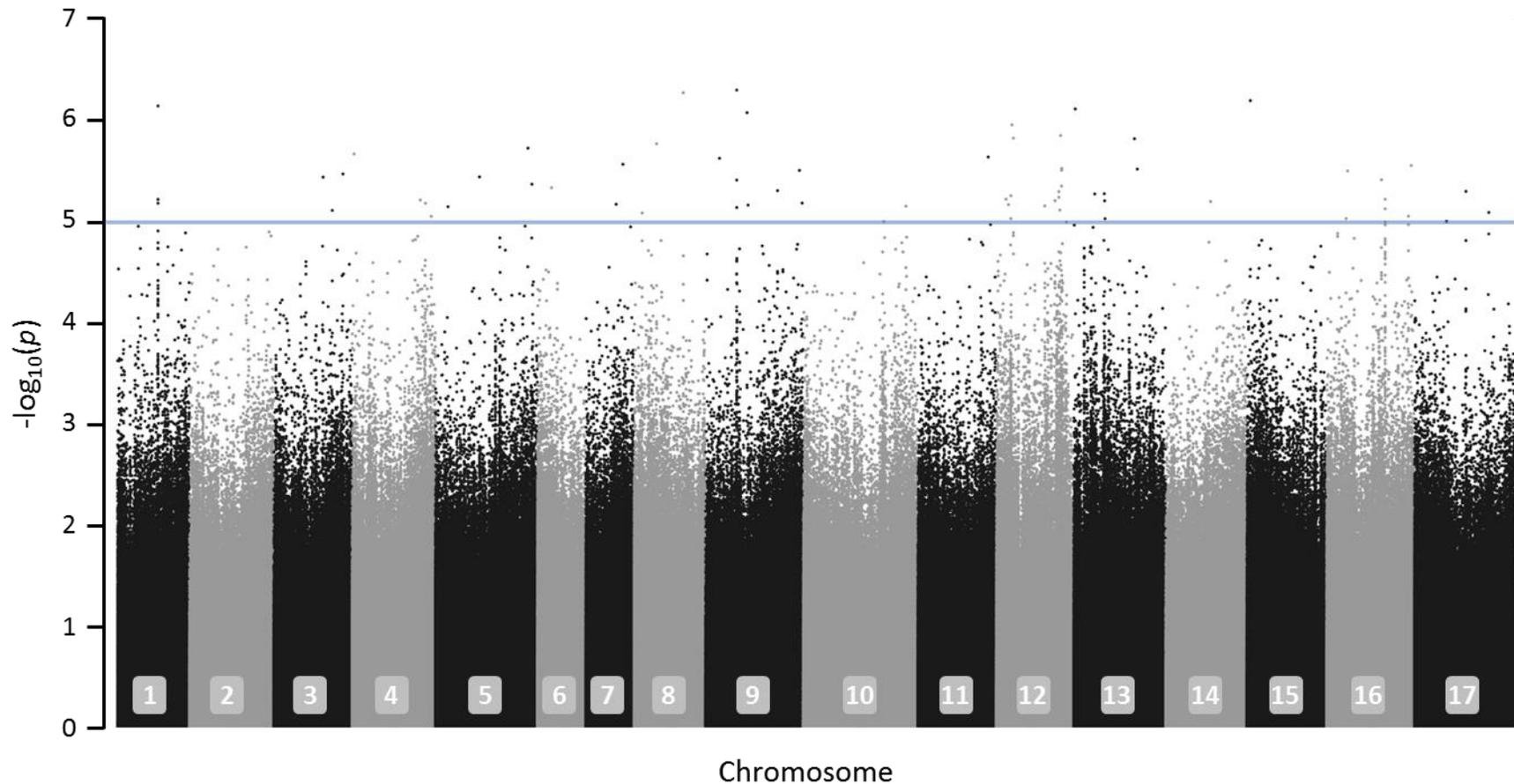
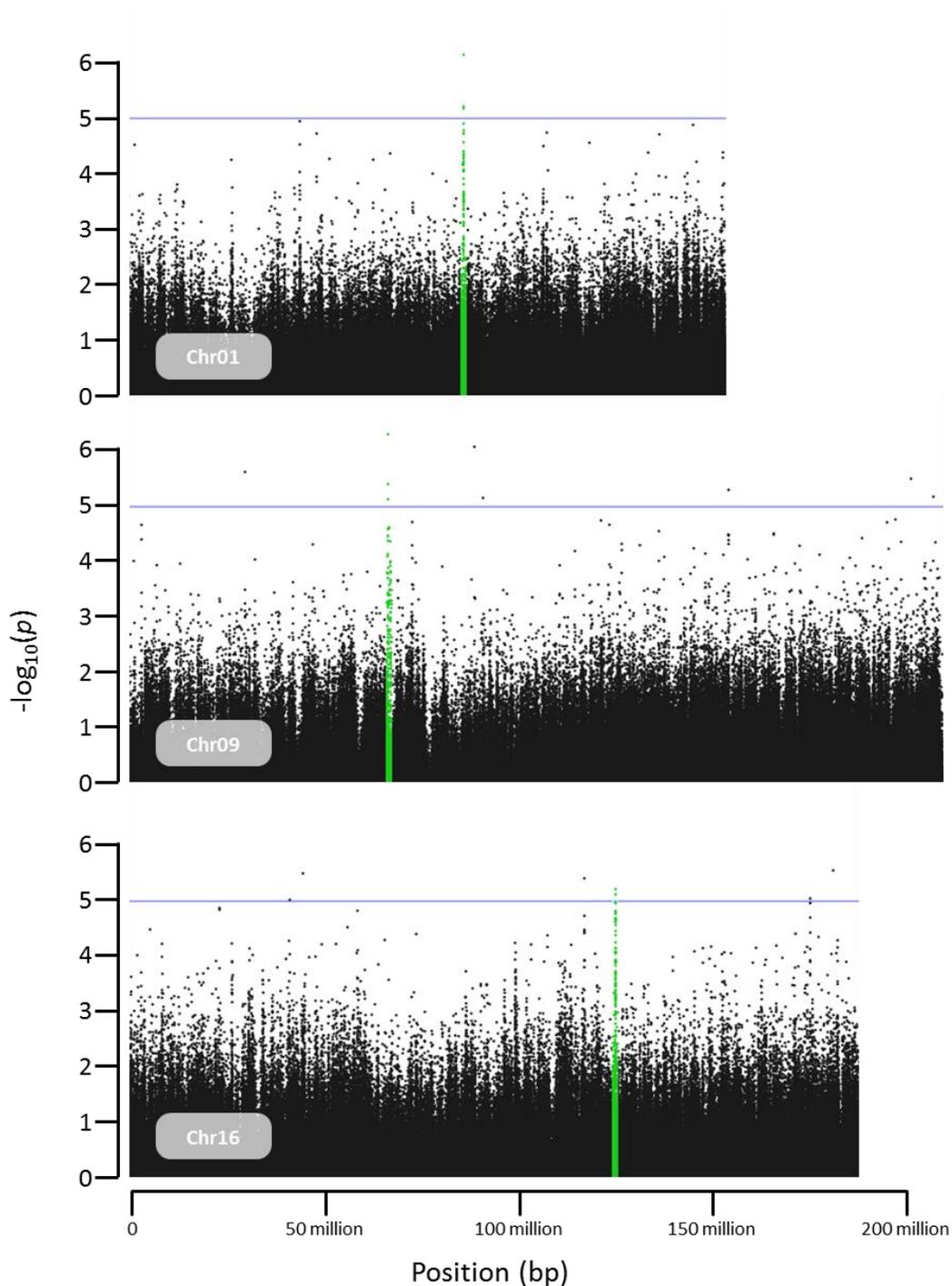


Figure 4.9: Zoomed in Manhattan plots for chromosomes 1, 9 and 16, showing the negative value of the $\log p$ -value of each variant versus its chromosomal position. All chromosomes are plotted on the same scale. The blue horizontal line represents a “suggestive significance” threshold at $p = 0.00001$. Columns of SNPs such as those shown in green are more suggestive of a true association than singletons, which may represent genotyping errors.



Chapter 5 : Conclusion

As one of the greatest pests of agriculture, weeds significantly decrease crop yields causing billions of dollars in losses each year (Pimentel et al. 2005). Agricultural management practices have changed drastically over time, with cropping practices intensifying since the Green Revolution (Gould 1991), and weeds have adapted rapidly to these progressive changes. From the perspective of evolutionary biology, weeds can be excellent subjects for the study of adaptation (Harper 1960; Baker 1974; De Wet and Harlan 1975; Vigueira et al. 2013). Not only do they show rapid evolution in the face of human-mediated selection, but when replicate weed populations or weed species invade agricultural environments, evolution may also happen in parallel, allowing for tests of repeatability. Practically, the history of weeds on the landscape may be documented, and weed species also tend to be abundant, easy to grow and fast to reproduce, lending themselves to experimental manipulations. Despite these advantages, as mostly non-model organisms, the genetics of adaptation remain understudied in weeds (Stewart et al. 2009). In this thesis, I sought to improve our understanding of the evolution of weediness, including its genetic basis, in a common North American weed, annual sunflower (*Helianthus annuus* L.). In this final chapter, I will summarize findings from phenotypic and genotypic comparisons of paired populations of weedy sunflowers infesting crop fields and wild sunflowers collected from more natural habitats, as well as a genetic mapping study of herbicide resistance. Synthesizing the results, I will discuss the broader implications of my findings, while also acknowledging limitations. As a final step, I will identify interesting future avenues for follow-up research, while discussing the merits of agricultural weeds as model systems for understanding local adaptation.

5.1 Summary

For my thesis work, I asked whether populations of common sunflower growing as agricultural weeds show adaptations to the unique conditions of cultivated fields. Compared to more natural environments, modern monocultures are highly simplified environments, typically resource-rich, but experiencing frequent disturbance (Barrett 1988). Annual weeds may face selection to accelerate their development, in order to better compete with crop plants, or to produce seed prior to harvesting of the crop. This is indeed what I discovered in the case of sunflower, as described in Chapter 2.

Comparing populations of sunflower that had long acted as agricultural weeds (according to collection records) to populations obtained from non-agricultural environments (everything from construction sites and waste places, to prairie preserves and wetlands), weedy seedlings showed faster growth in the 2012 common garden. Faster growth in turn led to earlier flowering, suggesting a shift in

life history strategy to prioritize growth and reproduction. These trait differences were most likely genetically based, as I controlled for variation in seed provisioning (i.e., maternal effects). The only wild population collected from a mesic site (a wetland) also showed accelerated growth and reproduction. Wetlands are also relatively resource-rich and prone to disturbance (i.e., flooding and drying cycles), illustrating that selection for accelerated development is not unique to agriculture fields.

Following up on the initial common garden work, I performed a second common garden comparison the subsequent year to further explore the role of maternal effects in this system. Using seed generated from within-population crosses the previous year, I grew plants from both seed sources (field and common garden) for a subset of weedy-wild pairs, again measuring growth and flowering time. If effects of the maternal environment, and not genetic differences, were responsible for the results in 2012, we would expect to recapture the 2012 findings with the field-derived individuals, but not the common garden derived ones (where all mothers shared the same environment). For seedling growth, the results were exactly the opposite. This implies that, firstly, differences in growth seen in 2012 were genetically-based (as they were also observed in the next generation), and secondly that seed age may also influence growth rates. Flowering time did not differ by seed source or population type in 2013, indicating that genotype-by-environment interactions may influence reproductive traits.

In Chapters 3 and 4, I explored the extent of genetic differentiation between weedy and wild sunflowers, and investigated the genomic architecture of an important weed trait, resistance to the herbicide glyphosate. To date, investigations into the genetic basis of adaptive traits in wild species (weeds or otherwise) have largely relied on either marker data or reduced-representation genome sequencing methods (e.g. Genotyping-By-Sequencing; Elshire et al. 2011). To better elucidate the effects of selection across the entire weed genome, I used whole genome resequencing to obtain dense SNP datasets for both chapters: 2.7 million SNPs for sixteen focal individuals (eight of each population type) in Chapter 3, and 7.6 million SNPs for 321 mapping population individuals in Chapter 4. To our knowledge, our study represents one of the first to use whole genome shotgun (WGS) data to investigate the genetics of weed adaptation, but see Li et al. (2017).

For the weedy-wild comparison, I used two complementary approaches, a sliding window analysis of cluster separation scores (a metric based on genetic distances: Jones et al. 2012) and a distinct, variable-sized window analysis of F_{ST} between groups (following the methods of Beissinger et al. 2015), to identify a number of potential small regions of genetic differentiation, on the basis of consensus between the methods. In total, the analysis identified 148 regions, unequally distributed

across the genome and accounting for much less than 1% of the genome overall. However, as the analysis lacked power to detect true positives while controlling the false-discovery rate, these regions must be approached cautiously, as some will be false positives. An average of 1.8 genes overlapped each region, and the gene set included plant stress response proteins, flowering time factors and transporter genes linked to herbicide resistance.

To connect genotype to phenotype more directly, I used genome-wide association mapping (GWAS) to investigate the genes responsible for glyphosate resistance in sunflower, with the additional goal of providing insight on the mechanism(s) underlying resistance. Herbicide resistance may evolve via target-site changes (i.e., changes to the plant enzyme affected by the herbicide) or non-target-site changes, which encompass a variety of non-exclusive mechanisms to reduce the effects of the herbicide. While the genetics of target-site resistance are typically well characterized, being linked to a single (often known) gene, the genetic basis of non-target-site resistance can be considered the “dark side” of resistance research (Délye 2013). After treating sunflower seedlings with glyphosate at a rate of 0.5 kg a.e. ha⁻¹, I assessed survival and assigned a resistance score to each individual. Interpreting the results of the GWAS, I identified a total of 28 SNPs with suggestive associations ($p < 1 \times 10^{-5}$) to glyphosate resistance as candidates. Overlapping genes included two ABC transporters and a major facilitator transporter, both previously implicated in non-target-site resistance in other species (Yuan et al. 2007; Staub et al. 2012), but did not include either gene copy of the enzyme targeted by glyphosate.

5.2 General Conclusions

5.2.1 Rapid and Repeated Evolution of Weediness in Sunflower

The results presented in my thesis contribute to our understanding of what adaptations may be important for agricultural weeds in contemporary, high-intensity agroecosystems, and how these may evolve. There are several key take-home messages. Firstly, in accordance with a growing body of work on weedy and invasive species (e.g. see reviews in Buswell et al. 2011; Vigueira et al. 2013), I found that successful weedy sunflower populations had evolved in response to colonizing cultivated fields. Though *Helianthus annuus* is a generalist ruderal species that prefers open areas, and can show broad environmental tolerances and phenotypic plasticity (Heiser et al. 1969), weedy populations nonetheless showed genetic changes that are likely adaptive for the agricultural environment. The paired design utilized in my common garden work allowed me to identify changes due to weediness itself, a very important element for such a broad-ranged species. Previous work has identified strong phenotypic and genotypic geographical, and particularly latitudinal, differentiation in *H. annuus* (Cantamutto et al. 2010;

Blackman et al. 2011; McAssey et al. 2016); I also observed substantial among-population variation (e.g. Figure 2.4) with latitudinal patterns in traits such as plant size and flowering time. Populations were also sorted according to latitude in a principal component analysis (PCA) of the genetic data. If we were to compare weedy and wild populations from different latitudes blindly, weediness would be confounded with other local adaptations. Sorting loci linked to weediness from loci that had undergone unrelated local and regional selective sweeps was a challenge in earlier studies of this system (Kane and Rieseberg 2008; Lai et al. 2008).

As I focussed on phenotypic and genotypic differences that were common across multiple weedy-wild population pairs, this also increases the probability that observed differences were adaptive. Neutral processes, such as gene flow, may cause populations to diverge randomly (Elmer and Meyer 2011). In contrast, parallel shifts between replicate population pairs are most likely a consequence of natural selection, as it is unlikely that the same changes would occur repeatedly across populations due to chance alone. Given that agricultural intensification in the Midwest has occurred only within the last 100 years or so, weedy sunflower evolution has been rapid. This agrees with a recent consensus in the literature that evolutionary change can happen very quickly (e.g. Holt 2005; Carroll et al. 2007), to the extent that even the ecological dynamics of species may be affected. Such rapid change may be more common when environmental conditions suddenly shift (Neuhauser et al. 2003) or a species colonizes a novel habitat, as seen in this work. Here, I found that sunflowers have evolved both life history differences and resistance to the herbicide glyphosate as a consequence of selection in agroecosystems. Weed evolution also proceeded despite ongoing gene flow from wild sunflower populations, again suggesting that selection is strong enough to offset maladaptive gene flow.

Overall, there was little genetic structure among the study populations, suggesting that levels of gene flow are fairly high across the landscape in sunflower. The average genome-wide F_{ST} between weedy and wild individuals in Chapter 3 was close to zero ($F_{ST} < 0.01$), implying that, while weedy and wild populations may differ at a small proportion of key loci, most of the genome is not differentiated. In an analysis of genetic structure based on microsatellite data in U.S. sunflower populations, Kane and Rieseberg (2008) also found that weedy ecotypes were not highly differentiated from geographically proximal wild populations; additionally, in a neighbour-joining tree, their populations clustered geographically. In my thesis populations, there was a signal of isolation-by-distance in the PCA in Chapter 4 (Figure 4.6); however, the first two PC axes explained less than 2% of the variation in the data. Altogether, these results imply that weediness has likely evolved multiple times across the *H. annuus*

range, presenting an interesting case of parallel evolution. Weedy populations have not only evolved the same phenotypic adaptations (e.g. faster seedling growth), but the genome scans also revealed some amount of parallel genetic differentiation. Adaptation to cultivated fields therefore most likely involves standing variation as opposed to new mutations, given the short time span. However, the evidence was weaker for genetic versus phenotypic parallelism. This suggests that the alleles and loci underlying weediness may vary among populations, diluting the overall signal of divergence. Similarly, Qi et al. (2015) found little evidence for a shared genetic basis to weediness in different strains of weedy rice (*Oryza sativa* L.), and there may be ample genetic variation in both rice and sunflower to allow the evolution of weedy traits via multiple genetic mechanisms.

5.2.2 Widespread Glyphosate Resistance in Sunflower Suggests Multiple Origins

In agricultural systems, the strong selection imposed by the use of novel herbicides can lead to a particularly rapid evolutionary response (Powles and Yu 2010). For example, in a selection experiment for diclofop-methyl resistance in rigid ryegrass (*Lolium rigidum* Gaudin), Neve and Powles (Neve and Powles 2005) found that even a single cycle of high-dose herbicide selection significantly increased resistance across populations. In my thesis, I documented the widespread evolution of glyphosate resistance in North American sunflowers. Glyphosate was first introduced as a herbicide in 1974, but became widely used only after the commercialization of glyphosate-resistant crops by Monsanto in the 1990s (Duke and Powles 2008). All cases of glyphosate resistance in weeds have evolved since then, with 39 reports to date, according to the International Survey of Herbicide-Resistant Weeds (ISHRW 2017, www.weedscience.org; retrieved on 11 Dec 2017). The first reported case occurred in rigid ryegrass in Australia in 1996 (Powles et al. 1998), showing just how quickly resistance can evolve, even for an herbicide for which evolving resistance can be challenging (Bradshaw et al. 1997).

In comparison to herbicides with other target sites, such as acetolactate synthase (ALS) inhibitors, the evolution of resistance to glyphosate can be challenging as changes in the amino acid sequence of the target enzyme (5-enolpyruvylshikimate-3-phosphate synthase, or EPSPS) that inhibit glyphosate binding also decrease enzyme function (Powles and Yu 2010). In contrast, amino acid changes conferring resistance to ALS inhibitors, for example, tend not to inhibit enzyme function (Duggleby et al. 2008) and, perhaps as a result, target-site ALS resistance has evolved easily in many species, with 159 cases reported to date (ISHRW 2017). The alternative to target-site resistance is a variety of non-mutually exclusive mechanisms to reduce herbicide absorption and translocation, sequester the herbicide or detoxify it, collectively called non-target site resistance (NTSR) mechanisms

(Ghanizadeh and Harrington 2017a). While we can't exclude target-site resistance in sunflower without further study of the *EPSPS* sequences, results from the GWAS are indicative of NTSR, with several loci on multiple chromosomes involved. Three candidate SNPs were found in genes for transporter proteins, implicating a role for altered glyphosate translocation patterns or herbicide sequestration as mechanisms of resistance. Thus, though much more complex to study than target-site resistance, the potential for NTSR to play an important role in the response to novel herbicides should not be ignored.

An interesting model of the evolution of NTSR was recently proposed by Délye (2013), referred to as “allele stacking”. In the model, NTSR increases gradually over time, as different alleles conferring resistance accumulate in the population, “stacking” within individuals in outcrossing species. Sub-lethal herbicide applications can exacerbate the accumulation of resistance alleles in weeds, by allowing the persistence of weak resistance alleles that may later combine. Here, peaks in the GWAS were merely suggestive of association with glyphosate resistance, not passing stringent correction for multiple testing. It therefore seems likely that different populations, separated geographically, may have stacked different sets of resistance alleles and loci, each evolving glyphosate resistance in a unique way. Idiosyncrasy in the precise NTSR mechanisms involved and/or genes could lead to greater noise in the GWAS and weaker associations. Interestingly, as wild sunflower is often subject to removal efforts across the landscape (owing to its role as an agricultural weed), glyphosate resistance may not be a weed-specific adaptation. In the GWAS mapping population, resistance also segregated in wild populations, with some wild individuals among the best survivors. Interestingly, there was no overlap in terms of either genes or candidate regions between the analyses in Chapter 3 and Chapter 4; comparing the two studies, the three closest matches occurred on chromosomes nine, thirteen and sixteen and were spaced roughly one million base pairs apart. If weedy and wild populations are not differentiated in terms of glyphosate resistance, this may explain why candidate loci identified by the GWAS were not recaptured in the genome scan.

5.3 Future Directions

5.3.1 Extending the Common Garden Work to Compare Weedy-Wild Fitness Differences

While the weedy-wild differences in growth rate and flowering time seen in the common garden are intriguing, suggesting an adaptive shift in life history strategy, the common garden work could be extended to better elucidate the whole story. Genotype-by-environment interactions are extremely common (Des Marais et al. 2013), and this was evident here when comparing the effects of two replicate experiments performed in two different years. Growing sunflowers well outside of the native

range and climate normals they typically experience could obscure true weedy-wild differences, and a repeat common garden in the native range may be advised to confirm trait differences seen in Vancouver, Canada. It is also unclear how common garden work in a benign environment, which is neither an agricultural field nor a typical resource-poor wild site, informs how plants may perform in these habitats. A reciprocal transplant experiment would allow for quantifying weedy and wild plant performance in both habitats. As we could not quantify plant fitness, the effects of differences in flowering time on seed production, for example, are unknown. Do earlier flowering plants necessarily have an advantage? Quantifying fitness in meaningful habitats, such as agricultural fields, could reveal the importance of putative weed adaptations identified here.

5.3.2 Accounting for the Paired Nature of the Weedy-Wild Data in the Genome Scans

One particular challenge with the genome scan of weedy-wild genetic divergence was the small sample size, because I could not account for the paired nature of the data in the analysis. In the common garden, accounting for the fact that weedy-wild populations were paired was critical to uncover trait differences between the population types. As sunflowers show dramatic latitudinal variation, this must first be accounted for in order to then isolate the effect of type. Putative differentiated genomic regions in Chapter 3 could not be confirmed owing to low power in the analysis, likely caused by this same issue: much larger genetic differences owing to local adaptation to other factors (e.g. climate) obscuring the smaller effects of population type. To rectify this issue, we are planning a follow-up analysis including ten individuals per population. This should permit modifications of the randomization testing in order to better reflect the nature of the data, as well as allow us to make among population comparisons. Comparing individual weedy-wild pairs and then looking for overlap would be an alternative way to identify genomic regions of divergence between population types.

5.3.3 Greater Precision in Mapping Glyphosate Resistance

Non-target-site resistance may have a complex genetic basis (Yuan et al. 2007). In the sunflower populations I studied here, the loci and alleles underlying NTSR to glyphosate also likely varied among populations, creating additional complexity. Finally, as our selected rate of glyphosate application only resulted in a ~20% death rate, I had to score survivors qualitatively to assign a herbicide resistance score to be used in the GWAS; however, survivor phenotypes were immensely varied and it was sometimes unclear what differences in say, leaf die-off, meant in terms of a plant's potential future fitness. In the GWAS, these complexities added noise to the analysis, and SNPs with potential associations to glyphosate resistance did not pass corrections for multiple testing. Genetic mapping of NTSR involving

multiple loci will always be challenging, but we might improve our chances by reducing the complexity. Including only a single population, or subset of populations from a small geographical area, might be one option to reduce the number of loci involved. Similarly, one could select a particularly resistant accession and work to create a mapping population from just this line, with an extreme example being the creation of recombinant inbred lines (RILs). A powerful approach for mapping the loci underlying complex traits (Pollard 2012), RILs are created via the repeated sibling mating (or selfing, in self-compatible species) of offspring from an initial cross between two parents differing in a phenotype of interest, such as herbicide resistance.

Bibliography

- Abdeen, A., and B. Miki. 2009. The pleiotropic effects of the bar gene and glufosinate on the *Arabidopsis* transcriptome. *Plant Biotechnol. J.* 7:266–282.
- Abramovich, F., and Y. Benjamini. 2005. False Discovery Rate. *Encycl. Stat. Sci.* 2240–2243. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Abreu, L. A. de S., M. L. M. de Carvalho, C. A. G. Pinto, V. Y. Kataoka, and T. T. de A. Silva. 2013. Deterioration of sunflower seeds during storage. *J. Seed Sci.* 35:240–247.
- Al-Khatib, K., J. R. Baumgartner, D. E. Peterson, and R. S. Currie. 1998. Imazethapyr resistance in common sunflower (*Helianthus annuus*). *Weed Sci.* 46:403–407.
- Anderson, M. P., and J. W. Gronwald. 1991. Atrazine resistance in a velvetleaf (*Abutilon theophrasti*) biotype due to enhanced glutathione S-transferase activity. *Plant Physiol* 96:104–109.
- Andrews, K. R., J. M. Good, M. R. Miller, G. Luikart, and P. A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17:81–92.
- Angert, A. L., H. D. Bradshaw, and D. W. Schemske. 2008. Using experimental evolution to investigate geographic range limits in monkeyflowers. *Evolution* 62:2260–2675.
- Arendt, J., and D. Reznick. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* 23:26–32.
- Arias, D. M., and L. H. Rieseberg. 1994. Gene flow between cultivated and wild sunflowers. *Theor. Appl. Genet.* 89:655–660.
- Asche, D.L. 1993. Common sunflower (*Helianthus annuus* L.): The pathway toward domestication. *Proceedings of the 58th Annual Meeting of the Society for American Archaeology*, St. Louis, MO.
- 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.* 17:666–677.
- Badouin, H., J. Gouzy, C. J. Grassa, F. Murat, S. E. Staton, L. Cottret, C. Lelandais-Brière, G. L. Owens, S. Carrère, B. Mayjonade, L. Legrand, N. Gill, N. C. Kane, J. E. Bowers, S. Hubner, A. Bellec, A. Bérard, H. Bergès, N. Blanchet, M. Boniface, D. Brunel, O. Catrice, N. Chaidir, C. Claudel, C. Donnadiou, 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. Muñoz, P. Vincourt, L. H. Rieseberg, and N. B. Langlade. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546:148–152.
- Baima, S., F. Nobili, G. Sessa, S. Lucchetti, I. Ruberti, and G. Morelli. 1995. The expression of the *Athb-8* homeobox gene is restricted to provascular cells in *Arabidopsis thaliana*. *Development* 121:4171–4182.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi: 10.1371/journal.pone.0003376
- Baker, H. G. 1974. The evolution of weeds. *Annu. Rev. Ecol. Syst.* 5:1–24.

- Baranwal, V. K., N. Negi, and P. Khurana. 2017. Auxin response factor genes repertoire in mulberry: identification, and structural, functional and evolutionary analyses. *Genes* 8:202.
- Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23:38-44.
- Barrett, S. C. H. 1983. Crop mimicry in weeds. *Econ. Bot.* 37:255–282.
- Barrett, S. C. H. 1988. Genetics and evolution of agricultural weeds. Pp. 57–75 *in* M. A. Altieri and M. Liebman, eds. *Weed Management in Agroecosystems: Ecological Approaches*. CRC Press, Boca Raton, Florida.
- Barrett, S. C. H., and D. E. Seaman. 1980. The weed flora of californian rice fields. *Aquat. Bot.* 9:351–376.
- Barton, N. H. 2000. Genetic hitchhiking. *Philos. Trans. R. Soc. B Biol. Sci.* 355:1553–1562.
- Basu, C., M. D. Halfhill, T. C. Mueller, and C. N. Stewart. 2004. Weed genomics: New tools to understand weed biology. *Trends Plant Sci.* 9:391–398.
- Bates, D., M. Mächler, B. Bolker, and S. Walker. 2014. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1–48.
- Battaglia, M., and A. A. Covarrubias. 2013. Late embryogenesis abundant (LEA) proteins in legumes. *Front Plant Sci* 4:190.
- Baucom, R. S. 2016. The remarkable repeated evolution of herbicide resistance. *Am. J. Bot.* 103:181–183.
- Baucom, R. S., and R. Mauricio. 2004. Fitness costs and benefits of novel herbicide tolerance in a noxious weed. *Proc. Natl. Acad. Sci. U. S. A.* 101:13386–90.
- Baute, G. J., N. C. Kane, C. J. Grassa, Z. Lai, and L. H. Rieseberg. 2015. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. *New Phytol.* 206:830–838.
- Bazzaz, F. A., N. R. Chiariello, P. D. Coley, and L. F. Pitelka. 1987. Allocating resources to reproduction and defense. *Bioscience* 37:58–67.
- Beissinger, T. M., G. J. Rosa, S. M. Kaeppler, D. Gianola, and N. de Leon. 2015. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genet. Sel. Evol.* 47:30.
- Benbrook, C. M. 2016. Trends in glyphosate herbicide use in the United States and globally. *Environ. Sci. Eur.* 28:3.
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.* 57:289-300.
- Bervillé, A., 2002. Perennial sunflower in breeding for broomrape resistance. Pp. 14-16 *in* *Parasitic Plant Management in Sustainable Agriculture Joint Meeting of COST Action 849*. Sofia, Bulgaria.
- Bhatia, G., N. Patterson, S. Sankararaman, and A. L. Price. 2013. Estimating and interpreting FST: The impact of rare variants. *Genome Res.* 23:1514–1521.
- Blackman, B. K., S. D. Michaels, and L. H. Rieseberg. 2011. Connecting the sun to flowering in sunflower adaptation. *Mol. Ecol.* 20:3503–3512.

- 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.* 20:629–635.
- Blair, A. C., and L. M. Wolfe. 2004. The evolution of an invasive plant: An experimental study with *Silene latifolia*. *Ecology* 85:3035–3042.
- Blossey, B., and R. Notzold. 1995. Evolution of increased competitive ability in invasive nonindigenous plants: A hypothesis. *J. Ecol.* 83:887.
- Bohlenius, H., T. Huang, L. Charbonnel-Campaa, A. B. Brunner, S. Jansson, S. H. Strauss, and O. Nilsson. 2006. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* 312:1040–1043.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–20.
- Bouaziz, M., C. Ambroise, and M. Guedj. 2011. Accounting for population stratification in practice: A comparison of the main strategies dedicated to genome-wide association studies. *PLoS One* 6:e28845. doi: 10.1371/journal.pone.0028845
- Bradshaw, L. D., S. R. Padgett, S. L. Kimball, and B. H. Wells. 1997. Perspectives on glyphosate resistance. *Weed Technol.* 11:189–198.
- Brewer, C. E., and L. R. Oliver. 2009. Confirmation and resistance mechanisms in glyphosate-resistant common ragweed (*Ambrosia artemisiifolia*) in Arkansas. *Weed Sci.* 57:567–573.
- Brock, M. A., D. L. Nielsen, R. J. Shiel, J. D. Green, and J. D. Langley. 2003. Drought and aquatic community resilience: The role of eggs and seeds in sediments of temporary wetlands. *Freshw. Biol.* 48:1207–1218.
- Burdiak, P., A. Rusaczonek, D. Witoń, D. Głów, and S. Karpiński. 2015. Cysteine-rich receptor-like kinase CRK5 as a regulator of growth, development, and ultraviolet radiation responses in *Arabidopsis thaliana*. *J. Exp. Bot.* 66:3325–3337.
- Burke, J. M., S. J. Knapp, and L. H. Rieseberg. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* 171:1933–1940.
- Burke, J. M., S. Tang, S. J. Knapp, and L. H. Rieseberg. 2002. Genetic analysis of sunflower domestication. *Genetics* 161:1257–1267.
- Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, and A. D. Long. 2010. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467:587–590.
- Busi, R., P. Neve, and S. Powles. 2013. Evolved polygenic herbicide resistance in *Lolium rigidum* by low-dose herbicide selection within standing genetic variation. *Evol. Appl.* 6:231–242.
- Busi, R., and S. B. Powles. 2009. Evolution of glyphosate resistance in a *Lolium rigidum* population by glyphosate selection at sublethal doses. *Heredity* 103:318–325.
- Buswell, J. M., A. T. Moles, and S. Hartley. 2011. Is rapid evolution common in introduced plant species? *J. Ecol.* 99:214–224.
- Cantamutto, M., A. Presotto, I. Fernandez Moroni, D. Alvarez, M. Poverene, and G. Seiler. 2010. High infraspecific diversity of wild sunflowers (*Helianthus annuus* L.) naturally developed in central

- Argentina. *Flora Morphol. Distrib. Funct. Ecol. Plants* 205:306–312.
- Carroll, S. P., A. P. Hendry, D. N. Reznick, and C. W. Fox. 2007. Evolution on ecological time-scales. *Funct. Ecol.* 21:387-393.
- Casanova, M. T., and M. A. Brock. 2000. How do depth, duration and frequency of flooding influence the establishment of wetland plant communities? *Plant Ecol.* 147:237–250.
- Casquero, M., and M. Cantamutto. 2016. Interference of the agrestal *Helianthus annuus* biotype with sunflower growth. *Weed Res.* 56:229–236.
- Casquero, M., A. Presotto, and M. Cantamutto. 2013. Exoferality in sunflower (*Helianthus annuus* L.): A case study of intraspecific/interbiotype interference promoted by human activity. *F. Crop. Res.* 142:95–101.
- Caverzan, A., G. Passaia, S. B. Rosa, C. W. Ribeiro, F. Lazzarotto, and M. Margis-Pinheiro. 2012. Plant responses to stresses: Role of ascorbate peroxidase in the antioxidant protection. *Genet. Mol. Biol.* 35:1011-1019.
- Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jonsson, D. Schluter, M. A. Bell, and D. M. Kingsley. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327:302–305.
- Chandler, J. M., and C. C. Jan. 1985. Comparison of germination techniques for wild *Helianthus* seeds. *Crop Sci.* 25:356–358.
- Chapin, F. S. 1980. The mineral nutrition of wild plants. *Annu. Rev. Ecol. Syst.* 11:233–260.
- 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:e1002628. doi: 10.1371/journal.pgen.1002628
- Chauhan, B. S., and S. B. Abugho. 2012. Effect of growth stage on the efficacy of postemergence herbicides on four weed species of direct-seeded rice. *ScientificWorldJournal.* 2012:123071. doi: 10.1100/2012/123071
- Chen, P., X. Zhang, T. Zhao, Y. Li, and J. Gai. 2014. Genome-wide identification and characterization of RBR ubiquitin ligase genes in soybean. *PLoS One* 9:e87282. doi: 10.1371/journal.pone.0087282
- Choi, Y.-J., R. Tyagi, S. N. McNulty, B. A. Rosa, P. Ozersky, J. Martin, K. Hallsworth-Pepin, T. R. Unnasch, C. T. Norice, T. B. Nutman, G. J. Weil, P. U. Fischer, and M. Mitreva. 2016. Genomic diversity in *Onchocerca volvulus* and its *Wolbachia* endosymbiont. *Nat. Microbiol.* 2:16207.
- Christoffers, M. J., and A. V Varanasi. 2010. Glyphosate resistance: genetic basis in weeds. P. 141–148 in V. K. Nandula, ed. *Glyphosate Resistance in Crops and Weeds: History, Development, and Management.* John Wiley & Sons, Inc., Hoboken.
- Colautti, R. I., J. L. Maron, and S. C. H. Barrett. 2009. Common garden comparisons of native and introduced plant populations: Latitudinal clines can obscure evolutionary inferences. *Evol. Appl.* 2:187–199.
- Coley, P. D., J. P. Bryant, and F. S. Chapin. 1985. Resource availability and plant antiherbivore defense.

Science 230:895–899.

- Colosimo, P. F. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307:1928–1933.
- Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B Biol. Sci.* 279:5039–5047.
- Côté, S. D., T. P. Rooney, J.-P. Tremblay, C. Dussault, and D. M. Waller. 2004. Ecological impacts of deer overabundance. *Annu. Rev. Ecol. Evol. Syst.* 35:113–147.
- Crawley, M. J. 1987. What makes a community invasible? Pp. 429–453 in A. J. Gray, M. J. Crawley, and P. J. Edwards, eds. *Colonization, succession, and stability: The 26th Symposium of the British Ecological Society Held Jointly with the Linnean Society of London*. Blackwell Scientific Publications, Oxford.
- Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23:3133–3157.
- Culpepper, A. S., T. L. Grey, W. K. Vencill, J. M. Kichler, T. M. Webster, S. M. Brown, A. C. York, J. W. Davis, and W. W. Hanna. 2006. Glyphosate-resistant Palmer amaranth (*Amaranthus palmeri*) confirmed in Georgia. *Weed Sci.* 54:620–626.
- Cummins, I., D. J. Cole, and R. Edwards. 1999. A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black-grass. *Plant J.* 18:285–292.
- D’Antonio, C. M., and P. M. Vitousek. 1992. Biological invasions by exotic grasses, the grass/fire cycle, and global change. *Annu. Rev. Ecol. Syst.* 23:63–87.
- Daehler, C. C., and D. R. Strong. 1997. Reduced herbivore resistance in introduced smooth cordgrass (*Spartina alterniflora*) after a century of herbivore-free growth. *Oecologia* 110:99–108.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
- Dang, Z. H., Q. Qi, H. R. Zhang, H. Y. Li, S. B. Wu, and Y. C. Wang. 2014. Identification of salt-stress-induced genes from the RNA-Seq data of *Reaumuria trigyna* using differential-display reverse transcription PCR. *Int. J. Genomics* 2014: 381501.
- Dauer, J. T., D. A. Mortensen, and M. J. Vangessel. 2007. Temporal and spatial dynamics of long-distance *Conyza canadensis* seed dispersal. *J. Appl. Ecol.* 44:105–114.
- De Wet, J. M. J., and J. R. Harlan. 1975. Weeds and domesticates: Evolution in the man-made habitat. *Econ. Bot.* 29:99–108.
- Deines, S. R., J. A. Dille, E. L. Blinka, D. L. Regehr, and S. A. Staggenborg. 2004. Common sunflower (*Helianthus annuus*) and shattercane (*Sorghum bicolor*) interference in corn. *Weed Sci.* 52:976–983.
- Delmore, K. E., S. Hübner, N. C. Kane, R. Schuster, R. L. Andrew, F. Câmara, R. Guigó, and D. E. Irwin. 2015. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Mol. Ecol.* 24:1873–1888.

- Délye, C. 2013. Unravelling the genetic bases of non-target-site-based resistance (NTSR) to herbicides: A major challenge for weed science in the forthcoming decade. *Pest Manag. Sci.* 69:176–187.
- Dempewolf, H., L. H. Rieseberg, and Q. C. Cronk. 2008. Crop domestication in the Compositae: A family-wide trait assessment. *Genet. Resour. Crop Evol.* 55:1141–1157.
- Dennis, J., and K. C. S. Kou. 2014. Evaluating the agronomic benefits of biochar amended soils in an organic system : results from a field study at the UBC Farm, Vancouver. *Cent. Sustain. Food Syst. UBC Farm.*
- Dennis, M., K. J. Hembree, J. T. Bushoven, and A. Shrestha. 2016. Growth stage, temperature, and time of year affects the control of glyphosate-resistant and glyphosate-paraquat resistant *Conyza bonariensis* with saflufenacil. *Crop Prot.* 81:129–137.
- 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.* 44:5–29.
- Devlin, B., and K. Roeder. 1999. Genomic control for association studies. *Biometrics* 55:997–1004.
- Dinelli, G., I. Marotti, A. Bonetti, P. Catizone, J. M. Urbano, and J. Barnes. 2008. Physiological and molecular bases of glyphosate resistance in *Conyza bonariensis* biotypes from Spain. *Weed Res.* 48:257–265.
- Dlugosch, K. M., and I. M. Parker. 2008. Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecol. Lett.* 11:701–709.
- Doebley, J. F., B. S. Gaut, and B. D. Smith. 2006. The molecular genetics of crop domestication. *Cell* 127:1309-1321.
- Dry, P., and J. Burdon. 1986. Genetic structure of natural populations of wild sunflowers (*Helianthus annuus* L.) in Australia. *Aust. J. Biol. Sci.* 39:255–270.
- Dudbridge, F., and A. Gusnanto. 2008. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* 32:227–34.
- Duggal, P., E. M. Gillanders, T. N. Holmes, and J. E. Bailey-Wilson. 2008. Establishing an adjusted *p*-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 9:516. doi: 10.1186/1471-2164-9-516
- Duggleby, R. G., J. A. McCourt, and L. W. Guddat. 2008. Structure and mechanism of inhibition of plant acetohydroxyacid synthase. *Plant Physiol. Biochem.* 46:309-324.
- Duke, S. O., and S. B. Powles. 2008. Glyphosate: A once-in-a-century herbicide. *Pest Manag. Sci.* 64:319-325.
- Duncan, C. N., and S. C. Weller. 1987. Heritability of glyphosate susceptibility among biotypes of field bindweed. *J. Hered.* 78:257–260.
- Ebbert, M. T. W., M. E. Wadsworth, L. A. Staley, K. L. Hoyt, B. Pickett, J. Miller, J. Duce, J. S. K. Kauwe, and P. G. Ridge. 2016. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics* 17:239. doi: 10.1186/s12859-016-1097-3

- El-Maarouf-Bouteau, H., C. Mazuy, F. Corbineau, and C. Bailly. 2011. DNA alteration and programmed cell death during ageing of sunflower seed. *J. Exp. Bot.* 62:5003–5011.
- Ellstrand, N. C., S. M. Heredia, J. A. Leak-Garcia, J. M. Heraty, J. C. Burger, L. Yao, S. Nohzadeh-Malakshah, and C. E. Ridley. 2010. Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. *Evol. Appl.* 3:494–504.
- Elmer, K. R., and A. Meyer. 2011. Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends Ecol. Evol.* 26:298–306.
- 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:e19379. doi: 10.1371/journal.pone.0019379
- Elton, C. S. 1958. *The Ecology of Invasions by Animals and Plants*. University of Chicago Press, Chicago.
- Eu-ahsunthornwattana, J., E. N. Miller, M. Fakiola, S. M. B. Jeronimo, J. M. Blackwell, and H. J. Cordell. 2014. Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genet.* 10:e1004445. doi: 10.1371/journal.pgen.1004445
- Excoffier, L., and N. Ray. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.* 23:347–351.
- FAOSTAT. 2014. Food and Agriculture Organization of the United Nations Statistics Division. <http://faostat.fao.org>, retrieved on 10 Dec 2017.
- Fleury, D., S. Jefferies, H. Kuchel, and P. Langridge. 2010. Genetic and genomic tools to improve drought tolerance in wheat. Oxford University Press, Oxford.
- Foster, S. A., and J. A. Baker. 2004. Evolution in parallel: New insights from a classic system. *Trends Ecol. Evol.* 19:456–459.
- Fox, J., and S. Weisberg. 2011. An R Companion to Applied Regression. R package version 2.1-3.
- Fuerst, E. P., and M. A. Norman. 1991. Interactions of herbicides with photosynthetic electron transport interactions of herbicides with photosynthetic electron transport. *Weed Sci.* 39:458–464.
- Funk, V. A., R. J. Bayer, S. Keeley, R. Chan, L. Watson, B. Gemeinholzer, E. E. Schilling, J. L. Panero, B. G. Baldwin, N. Garcia-Jacas, A. Susanna, and R. K. Jansen. 2005. Everywhere but Antarctica: Using a supertree to understand the diversity and distribution of the Compositae. *K. Danske Vidensk. Selsk. Biol. Skr.* 55:343–373.
- Gaines, T. A., D. L. Shaner, S. M. Ward, J. E. Leach, C. Preston, and P. Westra. 2011. Mechanism of resistance of evolved glyphosate-resistant palmer amaranth (*Amaranthus palmeri*). *J. Agric. Food Chem.* 59:5886–5889.
- Gaines, T. A., A. A. Wright, W. T. Molin, L. Lorentz, C. W. Riggins, P. J. Tranel, R. Beffa, P. Westra, and S. B. Powles. 2013. Identification of genetic elements associated with *EPSPS* gene amplification. *PLoS One* 8:1–10. doi: 10.1371/journal.pone.0065819
- Galvão, R. M., U. Kota, E. J. Soderblom, M. B. Goshe, and W. F. Boss. 2008. Characterization of a new family of protein kinases from *Arabidopsis* containing phosphoinositide 3/4-kinase and ubiquitin-like domains. *Biochem. J.* 409:117–27.

- Gao, X., L. C. Becker, D. M. Becker, J. D. Starmer, and M. A. Province. 2010. Avoiding the high bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.* 34:100–105.
- Garg, B., N. Vaid, and N. Tuteja. 2014. In-silico analysis and expression profiling implicate diverse role of *EPSPS* family genes in regulating developmental and metabolic processes. *BMC Res. Notes* 7:58. doi: 10.1186/1756-0500-7-58
- Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. doi: arXiv:1207.3907 [q-bio.GN].
- Ge, X., D. André d'Avignon, J. J. H. Ackerman, and R. Douglas Sammons. 2010. Rapid vacuolar sequestration: The horseweed glyphosate resistance mechanism. *Pest Manag. Sci.* 66:345–348.
- Geier, P. W., L. D. Maddux, L. J. Moshier, and P. W. Stahlman. 1996. Common sunflower (*Helianthus annuus*) interference in soybean (*Glycine max*). *Weed Technol.* 10:317–321.
- Gerstein, A. C., J. Ono, D. S. Lo, M. L. Campbell, A. Kuzmin, and S. P. Otto. 2015. Too much of a good thing: the unique and repeated paths toward copper adaptation. *Genetics* 199:555–71.
- Ghanizadeh, H., and K. C. Harrington. 2017a. Non-target site mechanisms of resistance to herbicides. *Crit. Rev. Plant Sci.* 36:24–34.
- Ghanizadeh, H., and K. C. Harrington. 2017b. Perspectives on non-target site mechanisms of herbicide resistance in weedy plant species using evolutionary physiology. *AoB Plants* 9:plx035. doi: 10.1093/aobpla/plx035
- Gould, F. 1991. The evolutionary potential of crop pests. *Am. Sci.* 79:496–507.
- Grime, J., and R. Hunt. 1975. Relative growth-rate: Its range and adaptive significance in a local flora. *J. Ecol.* 63:393–422.
- Grime, J. P. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Am. Nat.* 111:1169–1194.
- Gronwald, J. W. 1997. Resistance to PS II inhibitor herbicides. Pp. 27–60 in S. B. Powles and J. A. M. Holtum, eds. *Herbicide Resistance in Plants: Biology and Biochemistry*. Springer Netherlands, Dordrecht.
- Guo, L., C. D. Nezames, L. Sheng, X. Deng, and N. Wei. 2013. Cullin-RING ubiquitin ligase family in plant abiotic stress pathways. *J. Integr. Plant Biol.* 55:21-30.
- Guo, L., J. Qiu, C. Ye, G. Jin, L. Mao, H. Zhang, X. Yang, Q. Peng, Y. Wang, L. Jia, Z. Lin, G. Li, F. Fu, C. Liu, L. Chen, E. Shen, W. Wang, Q. Chu, D. Wu, S. Wu, C. Xia, Y. Zhang, X. Zhou, L. Wang, L. Wu, W. Song, Y. Wang, Q. Shu, D. Aoki, E. Yumoto, T. Yokota, K. Miyamoto, K. Okada, D. S. Kim, D. Cai, C. Zhang, Y. Lou, Q. Qian, H. Yamaguchi, H. Yamane, C. H. Kong, M. P. Timko, L. Bai, and L. Fan. 2017. *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat. Commun.* 8:1031. doi: 10.1038/s41467-017-01067-5
- Ha, C. Van, D. T. Le, R. Nishiyama, Y. Watanabe, S. Sulieman, U. T. Tran, K. Mochida, N. Van Dong, K. Yamaguchi-Shinozaki, K. Shinozaki, and L.-S. P. Tran. 2013. The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress. *DNA Res.* 20:511–24.
- Hall, D., C. Tegström, and P. K. Ingvarsson. 2010. Using association mapping to dissect the genetic basis

- of complex traits in plants. *Briefings Funct. Genomics Proteomics* 9:157–165.
- Han, H. J., R. H. Peng, B. Zhu, X. Y. Fu, W. Zhao, B. Shi, and Q. H. Yao. 2014. Gene expression profiles of *Arabidopsis* under the stress of methyl viologen: a microarray analysis. *Mol. Biol. Rep.* 41:7089–7102.
- Harper, J. L. 1960. *The biology of weeds*. A symposium of the British Ecological Society, Oxford 2-4 April, 1959. Blackwell Scientific Publications Ltd, Oxford.
- Harter, A. V., K. A. Gardner, D. Falush, D. L. Lentz, R. A. Bye, and L. H. Rieseberg. 2004. Origin of extant domesticated sunflowers in eastern North America. *Nature* 430:201–205.
- Hartmann-Shenkman, A., M. E. Kislev, E. Galili, Y. Melamed, and E. Weiss. 2014. Invading a new niche: obligatory weeds at Neolithic Atlit-Yam, Israel. *Veg. Hist. Archaeobot.* 24:9–18.
- Hatfield, J. 2012. *Agriculture in the Midwest*. U.S. Natl. Clim. Assess. Midwest Tech. Input Rep. 1–8.
- Heap, I. 2014. Herbicide resistant weeds. Pp. 281–301 in D. Pimentel and R. Peshin, eds. *Integrated Pest Management: Pesticide Problems*, Vol.3. Springer Netherlands, Dordrecht.
- Heap, I. 2017. *The International Survey of Herbicide Resistant Weeds*. www.weedscience.org, retrieved on 10 Dec 2017.
- Heiser, C. B. 1954. Variation and subspeciation in the common sunflower, *Helianthus annuus*. *Am. Midl. Nat.* 51:287–305.
- Heiser, C. B., D. M. Smith, S. B. Clevenger, and W. C. Martin. 1969. The North American Sunflowers (*Helianthus*). *Memoirs of the Torrey Botanical Club* 22:1-218.
- Hensley, J. B., E. P. Webster, D. C. Blouin, D. L. Harrell, and J. A. Bond. 2013. Response of rice to drift rates of glyphosate applied at low carrier volumes. *Weed Technol.* 27:257–262.
- Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–52.
- Herms, D. A., and W. J. Mattson. 1992. The Dilemma of plants: To grow or defend. *Q. Rev. Biol.* 67:283–335.
- Herrmann, K. M., and L. M. Weaver. 1999. The shikimate pathway. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50:473–503.
- Hey, J., N. Kleckner, K. Zhang, R. Chakraborty, and L. Jin. 2004. What's so hot about recombination hotspots? *PLoS Biol.* 2:e190. doi: 10.1371/journal.pbio.0020190
- Hinrichs, A. L., E. K. Larkin, and B. K. Suarez. 2009. Population stratification and patterns of linkage disequilibrium. *Genetic Epidemiology* 33:S88-S92.
- Hodgins, K. A., D. G. Bock, M. A. Hahn, S. M. Heredia, K. G. Turner, and L. H. Rieseberg. 2015. Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa. *Mol. Ecol.* 24:2226–2240.
- Hodgins, K. A., and L. Rieseberg. 2011. Genetic differentiation in life-history traits of introduced and native common ragweed (*Ambrosia artemisiifolia*) populations. *J. Evol. Biol.* 24:2731–2749.
- Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population

- genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6: e1000862. doi: 10.1371/journal.pgen.1000862
- Holt, R. 2005. On the integration of community ecology and evolutionary biology: Historical perspectives and current prospects. Pp. 235–271 in K. Cuddington and B. Beisner, eds. *Ecological Paradigms Lost*. Elsevier.
- Houssard, C., and J. Escarré. 1991. The effects of seed weight on growth and competitive ability of *Rumex acetosella* from two successional old-fields. *Oecologia* 86:236–242.
- Huey, R. B. 2000. Rapid Evolution of a Geographic Cline in Size in an Introduced Fly. *Science* 287:308–309.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54:187–212.
- Jeschke, J. M., and D. L. Strayer. 2006. Determinants of vertebrate invasion success in Europe and North America. *Glob. Chang. Biol.* 12:1608–1619.
- Jofuku, K. D., B. G. den Boer, M. Van Montagu, and J. K. Okamoto. 1994. Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* 6:1211–1225.
- Johnson, R. C., G. W. Nelson, J. L. Troyer, J. A. Lautenberger, B. D. Kessing, C. A. Winkler, and S. J. O'Brien. 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics* 11:724. doi: 10.1186/1471-2164-11-724
- Jonas, C. S., and M. A. Geber. 1999. Variation among populations of *Clarkia unguiculata* (Onagraceae) along altitudinal and latitudinal gradients. *Am. J. Bot.* 86:333–343.
- Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
- Jones, R., D. C. Culver, and T. C. Kane. 1992. Are parallel morphologies of cave organisms the result of similar selection pressures? *Evolution* 46:353–365.
- Kane, N. C., J. M. Burke, L. Marek, G. Seiler, F. Vear, G. Baute, S. J. Knapp, P. Vincourt, and L. H. Rieseberg. 2013. Sunflower genetic, genomic and ecological resources. *Mol. Ecol. Resour.* 13:10–20.
- 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.
- Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S. Kong, N. B. Freimer, C. Sabatti, and E. Eskin. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42:348–354.
- Kang, J., J. Park, H. Choi, B. Burla, T. Kretschmar, Y. Lee, and E. Martinoia. 2011. Plant ABC Transporters. *Arab. B.* 9:e0153. doi: 10.1199/tab.0153
- Kesseli, R. and R. Michelmore. 1997. The Compositae: Systematically fascinating but specifically neglected. Pp. 181–194 in A. H. Paterson and R. G. Landes, eds. *Genome Mapping in Plants*. Landes

Press, Georgetown, TX.

- Kim, J. S., H. J. Jung, H. J. Lee, K. A. Kim, C. H. Goh, Y. Woo, S. H. Oh, Y. S. Han, and H. Kang. 2008. Glycine-rich RNA-binding protein7 affects abiotic stress responses by regulating stomata opening and closing in *Arabidopsis thaliana*. *Plant J.* 55:455–466.
- Koger, C. H., and K. N. Reddy. 2005. Role of absorption and translocation in the mechanism of glyphosate resistance in horseweed (*Conyza canadensis*). *Weed Sci.* 53:84–89.
- Koornneef, M., C. Alonso-Blanco, and D. Vreugdenhil. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* 55:141–72.
- Kozioł, L., L. H. Rieseberg, N. Kane, and J. D. Bever. 2012. Reduced drought tolerance during domestication and the evolution of weediness results from tolerance-growth trade-offs. *Evolution* 66:3803–3814.
- Kuester, A., A. Wilson, S. M. Chang, and R. S. Baucom. 2016. A resurrection experiment finds evidence of both reduced genetic diversity and potential adaptive evolution in the agricultural weed *Ipomoea purpurea*. *Mol. Ecol.* 25:4508–4520.
- Lai, Z., N. C. Kane, Y. Zou, and L. H. Rieseberg. 2008. Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. *Genetics* 179:1881–1890.
- Lambers, H., and H. Poorter. 1992. Advances in ecological research: Classic papers. *Adv. Ecol. Res.* 34:187–261.
- Lander, E., and N. Schork. 1994. Genetic dissection of complex traits. *Science* 265:2037–2048.
- Lane, T. S., C. S. Rempe, J. Davitt, M. E. Staton, Y. Peng, D. E. Soltis, M. Melkonian, M. Deyholos, J. H. Leebens-Mack, M. Chase, C. J. Rothfels, D. Stevenson, S. W. Graham, J. Yu, T. Liu, J. C. Pires, P. P. Edger, Y. Zhang, Y. Xie, Y. Zhu, E. Carpenter, G. K.-S. Wong, and C. N. Stewart. 2016. Diversity of ABC transporter genes across the plant kingdom and their potential utility in biotechnology. *BMC Biotechnol.* 16:47. doi: 10.1186/s12896-016-0277-6
- Larios, E., and D. L. Venable. 2015. Maternal adjustment of offspring provisioning and the consequences for dispersal. *Ecology* 96:2771–2780.
- Laurie, C. C., K. F. Doheny, D. B. Mirel, E. W. Pugh, L. J. Bierut, T. Bhangale, F. Boehm, N. E. Caporaso, M. C. Cornelis, H. J. Edenberg, S. B. Gabriel, E. L. Harris, F. B. Hu, K. B. Jacobs, P. Kraft, M. T. Landi, T. Lumley, T. A. Manolio, C. McHugh, I. Painter, J. Paschall, J. P. Rice, K. M. Rice, X. Zheng, and B. S. Weir. 2010. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* 34:591–602.
- Lee, L. J., and J. Ngim. 2000. A first report of glyphosate-resistant goosegrass (*Eleusine indica* (L) Gaertn) in Malaysia. *Pest Manag. Sci.* 56:336–339.
- Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li, L.-F., Y.-L. Li, Y. Jia, A. L. Caicedo, and K. M. Olsen. 2017. Signatures of adaptation in the weedy rice genome. *Nat. Genet.* 49:811–814.

- Li, M. X., J. M. Y. Yeung, S. S. Cherny, and P. C. Sham. 2012. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131:747–756.
- Li, Y., B. O. Pennington, and J. Hua. 2009b. Multiple R-like genes are negatively regulated by *BON1* and *BON3* in *Arabidopsis*. *Mol. Plant. Microbe. Interact.* 22:840–848.
- Liebman, M., C. L. Mohler, and C. P. Staver. 2001. *Ecological Management of Agricultural Weeds*. Cambridge University Press, Cambridge.
- Linder, C. R., I. Taha, G. J. Seiler, A. A. Snow, and L. H. Rieseberg. 1998. Long-term introgression of crop genes into wild sunflower populations. *Theor. Appl. Genet.* 96:339–347.
- Liu, Q., Z. Wang, X. Xu, H. Zhang, and C. Li. 2015. Genome-wide analysis of C2H2 zinc-finger family transcription factors and their responses to abiotic stresses in poplar (*Populus trichocarpa*). *PLoS One* 10:e0134753. doi: 10.1371/journal.pone.0134753
- Lo, M.-T., D. A. Hinds, J. Y. Tung, C. Franz, C.-C. Fan, Y. Wang, O. B. Smeland, A. Schork, D. Holland, K. Kauppi, N. Sanyal, V. Escott-Price, D. J. Smith, M. O'Donovan, H. Stefansson, G. Bjornsdottir, T. E. Thorgeirsson, K. Stefansson, L. K. McEvoy, A. M. Dale, O. A. Andreassen, and C.-H. Chen. 2016. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* 49:152–156.
- Lorraine-Colwill, D. F., S. B. Powles, T. R. Hawkes, P. H. Hollinshead, S. A. J. Warner, and C. Preston. 2002. Investigations into the mechanism of glyphosate resistance in *Lolium rigidum*. *Pestic. Biochem. Physiol.* 74:62–72.
- Losos, J. B. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279:2115–2118.
- MacKay, J., and P. M. Kotanen. 2008. Local escape of an invasive plant, common ragweed (*Ambrosia artemisiifolia* L.), from above-ground and below-ground enemies in its native area. *J. Ecol.* 96:1152–1161.
- Mahmood, K., S. K. Mathiassen, M. Kristensen, and P. Kudsk. 2016. Multiple herbicide resistance in *Lolium multiflorum* and identification of conserved regulatory elements of herbicide resistance Genes. *Front. Plant Sci.* 7:1160. doi: 10.3389/fpls.2016.01160
- Mandel, J. R., S. Nambeesan, J. E. Bowers, L. F. Marek, D. Ebert, L. H. Rieseberg, S. J. Knapp, and J. M. Burke. 2013. Association mapping and the genomic consequences of selection in sunflower. *PLoS Genet.* 9:e1003378. doi: 10.1371/journal.pgen.1003378
- Marchinko, K. B. 2009. Predation's role in repeated phenotypic and genetic divergence of armor in threespine stickleback. *Evolution* 63:127–138.
- Maron, J. L., M. Vilà, and J. Arnason. 2004. Loss of enemy resistance among introduced populations of St. John's Wort (*Hypericum perforatum*). *Ecology* 85:3243–3253.
- Massinga, R. A., K. Al-Khatib, P. St. Amand, and J. F. Miller. 2003. Gene flow from imidazolinone-resistant domesticated sunflower to wild relatives. *Weed Sci.* 51:854–862.
- Matvienko, M., A. Kozik, L. Froenicke, D. Lavelle, B. Martineau, B. Perroud, and R. Michelmore. 2013. Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a

- duplex-specific nuclease and tetramethylammonium chloride. PLoS One 8:e55913. doi: 10.1371/journal.pone.0055913
- Mayrose, M., N. C. Kane, I. Mayrose, K. M. Dlugosch, and L. H. Rieseberg. 2011. Increased growth in sunflower correlates with reduced defences and altered gene expression in response to biotic and abiotic stress. *Mol. Ecol.* 20:4683–4694.
- 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
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.
- Merlin, A., A. Bonis, C. F. Damgaard, and F. Mesléard. 2015. Competition is a strong driving factor in wetlands, peaking during drying out periods. PLoS One 10:e0130152. doi: 10.1371/journal.pone.0130152
- Michitte, P., R. De Prado, N. Espinoza, J. P. Ruiz-Santaella, and C. Gauvrit. 2007. Mechanisms of resistance to glyphosate (*Lolium multiflorum*) biotype from Chile. *Weed Sci.* 55:435–440.
- Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17:240–248.
- Miller, S. E. 2016. Intraguild predation is a mechanism of divergent selection in the threespine stickleback. (Doctoral dissertation). Retrieved from: UBC cIRcle on 22 Sept 2017.
- Mithila, J., and A. S. Godar. 2013. Understanding genetics of herbicide resistance in weeds: Implications for weed management. *Adv. Crop Sci. Technol.* 1:115. doi: 10.4172/2329-8863.1000115
- Miyakawa, T., K. I. Miyazono, Y. Sawano, K. I. Hatano, and M. Tanokura. 2009. Crystal structure of ginkbilobin-2 with homology to the extracellular domain of plant cysteine-rich receptor-like kinases. *Proteins Struct. Funct. Bioinforma.* 77:247–251.
- Mizutani, M. 2012. Impacts of diversification of cytochrome P450 on plant metabolism. *Biol. Pharm. Bull.* 35:824–832.
- Mohler, C. L. 2001. Weed life history: identifying vulnerabilities. Pp. 40–98 in M. Liebman, C. L. Mohler, and C. P. Staver, eds. *Ecological Management of Agricultural Weeds*. Cambridge University Press, Cambridge.
- Monaco, T. J., S. C. Weller, F. M. Ashton, and F. M. Ashton. 2002. *Weed science : principles and practices*. John Wiley & Sons, Inc., Hoboken.
- Mooney, H. A., and R. J. Hobbs. 2000. *Invasive species in a changing world*. Island Press, Washington.
- Muller, M. H., F. Délieux, J. M. Fernández-Martínez, B. Garric, V. Lecomte, G. Anglade, M. Leflon, C. Motard, and R. Segura. 2009. Occurrence, distribution and distinctive morphological traits of weedy *Helianthus annuus* L. populations in Spain and France. *Genet. Resour. Crop Evol.* 56:869–877.
- Muller, M. H., M. Latreille, and C. Tollon. 2011. The origin and evolution of a recent agricultural weed: Population genetic diversity of weedy populations of sunflower (*Helianthus annuus* L.) in Spain and

- France. *Evol. Appl.* 4:499–514.
- Murray, M. G., and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8:4321–5.
- Myles, S., J. Peiffer, P. J. Brown, E. S. Ersoz, Z. Zhang, D. E. Costich, and E. S. Buckler. 2009. Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *Plant Cell* 21:2194–2202.
- Myles, S., K. Tang, M. Somel, R. E. Green, J. Kelso, and M. Stoneking. 2008. Identification and analysis of genomic regions with large between-population differentiation in humans. *Ann. Hum. Genet.* 72:99–110.
- Nadeau, N. J., A. Whibley, R. T. Jones, J. W. Davey, K. K. Dasmahapatra, S. W. Baxter, M. A. Quail, M. Joron, R. H. Ffrench-Constant, M. L. Blaxter, J. Mallet, and C. D. Jiggins. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* 367:343–353.
- Nandula, V. K. 2010. Herbicide resistance: Definitions and concepts. Pp. 35–43 in V. K. Nandula, ed. *Glyphosate Resistance in Crops and Weeds: History, Development, and Management*. John Wiley & Sons, Inc., Hoboken.
- Nandula, V. K., K. N. Reddy, D. H. Poston, A. M. Rimando, and S. O. Duke. 2008. Glyphosate tolerance mechanism in Italian ryegrass (*Lolium multiflorum*) from Mississippi. *Weed Sci.* 56:344–349.
- Naot, D., G. Ben-Hayyim, Y. Eshdat, and D. Holland. 1995. Drought, heat and salt stress induce the expression of a citrus homologue of an atypical late-embryogenesis *Lea5* gene. *Plant Mol. Biol.* 27:619–622.
- Neuhauser, C., D. A. Andow, G. E. Heimper, G. May, R. G. Shaw, and S. Wagenius. 2003. Community genetics: expanding the synthesis of ecology and genetics. *Ecology* 84:545–558.
- Neve, P., and S. Powles. 2005. High survival frequencies at low herbicide use rates in populations of *Lolium rigidum* result in rapid evolution of herbicide resistance. *Heredity* 95:485–492.
- Neve, P., M. Vila-Aiub, and F. Roux. 2009. Evolutionary-thinking in agricultural weed management. *New Phytol.* 184:783–793.
- Noble, W. S. 2009. How does multiple testing correction work? *Nat. Biotechnol.* 27:1135–1137.
- Nol, N., D. Tsikou, M. Eid, I. C. Livieratos, and C. N. Giannopolitis. 2012. Shikimate leaf disc assay for early detection of glyphosate resistance in *Conyza canadensis* and relative transcript levels of EPSPS and ABC transporter genes. *Weed Res.* 52:233–241.
- Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18:375–402.
- Nosil, P. P. 2012. *Ecological speciation*. Oxford University Press, Oxford.
- Nowicka, U., D. Zhang, O. Walker, D. Krutauz, C. A. Castañeda, A. Chaturvedi, T. Y. Chen, N. Reis, M. H. Glickman, and D. Fushman. 2015. DNA-damage-inducible 1 protein (Ddi1) contains an uncharacteristic ubiquitin-like domain that binds ubiquitin. *Structure* 23:542–557.
- Obenchain, V., M. Lawrence, V. Carey, S. Gogarten, P. Shannon, and M. Morgan. 2014.

- VariantAnnotation: A Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* 30:2076–2078.
- Oerke, E.-C. 2006. Crop losses to pests. *J. Agric. Sci.* 144:31. doi: 10.1017/S0021859605005708
- Oettmeier, W. 1999. Herbicide resistance and supersensitivity in photosystem II. *Cell Mol. Life Sci.* 15:1255-1277.
- Paine, C. E. T., T. R. Marthews, D. R. Vogt, D. Purves, M. Rees, A. Hector, and L. A. Turnbull. 2012. How to fit nonlinear plant growth models and calculate growth rates: An update for ecologists. *Methods Ecol. Evol.* 3:245–256.
- Parker, I. M., J. Rodriguez, and M. E. Loik. 2003. An evolutionary approach to understanding the biology of invasions: Local adaptation and general-purpose genotypes in the weed *Verbascum thapsus*. *Conserv. Biol.* 17:59–72.
- Parker, W. C., T. L. Noland, and A. E. Morneau. 2006. The effects of seed mass on germination, seedling emergence, and early seedling growth of eastern white pine (*Pinus strobus* L.). *New For.* 32:33–49.
- Pedersen, B. P., P. Neve, C. Andreasen, and S. B. Powles. 2007. Ecological fitness of a glyphosate-resistant *Lolium rigidum* population: Growth and seed production along a competition gradient. *Basic Appl. Ecol.* 8:258–268.
- Peng, Y., L. L. G. Abercrombie, J. S. Yuan, C. W. Riggins, R. D. Sammons, P. J. Tranel, and C. N. Stewart. 2010. Characterization of the horseweed (*Conyza canadensis*) transcriptome using GS-FLX 454 pyrosequencing and its application for expression analysis of candidate non-target herbicide resistance genes. *Pest Manag. Sci.* 66:1053–1062.
- Perez-Jones, A., and C. Mallory-Smith. 2010. Molecular basis of evolved glyphosate resistance. Pp. 119–140 in V. K. Nandula, ed. *Glyphosate Resistance in Crops and Weeds: History, Development, and Management*. John Wiley & Sons, Inc., Hoboken.
- Perez-Jones, A., K. W. Park, N. Polge, J. Colquhoun, and C. A. Mallory-Smith. 2007. Investigating the mechanisms of glyphosate resistance in *Lolium multiflorum*. *Planta* 226:395–404.
- Petit, C., G. Bay, F. Pernin, and C. Délye. 2010. Prevalence of cross- or multiple resistance to the acetyl-coenzyme a carboxylase inhibitors fenoxaprop, clodinafop and pinoxaden in black-grass (*Alopecurus myosuroides* Huds.) in France. *Pest Manag. Sci.* 66:168–177.
- Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52:273-288.
- Pinheiro, J. C., and D. M. Bates. 2000. *Mixed-effects models in S and S-PLUS*. Springer, New York.
- Plaisance, K. L., K. L. Plaisance, J. W. Gronwald, and J. W. Gronwald. 1999. Enhanced catalytic constant for glutathione s-transferase (atrazine) activity in an atrazine-resistant *Abutilon theophrasti* biotype. *Pest. Biochem. Physiol.* 63:34–49.
- Pollard, D. A. 2012. Design and construction of recombinant inbred lines. *Methods Mol. Biol.* 871:31–39.
- Porter, S. S., P. L. Chang, C. A. Conow, J. P. Dunham, and M. L. Friesen. 2017. Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J.* 11:248–262.

- Poverene, M., M. Cantamutto, and G. J. Seiler. 2009. Ecological characterization of wild *Helianthus annuus* and *Helianthus petiolaris* germplasm in Argentina. *Plant Genet. Resour.* 7:42–49.
- Power, R. A., J. Parkhill, and T. de Oliveira. 2016. Microbial genome-wide association studies: lessons from human GWAS. *Nat. Rev. Genet.* 18:41–50.
- Powles, S. B., D. F. Lorraine-Colwill, J. J. Dellow, and C. Preston. 1998. Evolved resistance to glyphosate in rigid ryegrass (*Lolium rigidum*) in Australia. *Weed Sci.* 46:604–607.
- Powles, S. B., and C. Preston. 2006. Evolved glyphosate resistance in plants: Biochemical and genetic basis of resistance. *Weed Technol.* 20:282–289.
- Powles, S. B., and D. L. Shaner. 2001. *Herbicide Resistance and World Grains*. CRC Press, Boca Raton.
- Powles, S. B., and Q. Yu. 2010. Evolution in action: plants resistant to herbicides. *Annu. Rev. Plant Biol.* 61:317–347.
- Prentis, P. J., J. R. U. Wilson, E. E. Dormontt, D. M. Richardson, and A. J. Lowe. 2008. Adaptive evolution in invasive species. *Trends Plant Sci.* 13:288–294.
- Presotto, A., F. Hernández, M. Díaz, I. Fernández-Moroni, C. Pandolfo, J. Basualdo, S. Cuppari, M. Cantamutto, and M. Poverene. 2017. Crop-wild sunflower hybridization can mediate weediness throughout growth-stress tolerance trade-offs. *Agric. Ecosyst. Environ.* 249:12–21.
- Preston, C., F. J. Tardif, J. T. Christopher, and S. B. Powles. 1996. Multiple resistance to dissimilar herbicide chemistries in a biotype of *Lolium rigidum* due to enhanced activity of several herbicide degrading enzymes. *Pestic. Biochem. Physiol.* 134:123–134.
- Preston, C., A. M. Wakelin, F. C. Dolman, Y. Bostamam, and P. Boutsalis. 2009. A decade of glyphosate-resistant *Lolium* around the world: mechanisms, genes, fitness, and agronomic management. *Weed Sci.* 57:435–441.
- 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.* 38:904–909.
- Price, A. L., N. A. Zaitlen, D. Reich, and N. Patterson. 2010. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* 11:459–463.
- Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. Association Mapping in Structured Populations. *Am. J. Hum. Genet.* 67:170–181.
- Qi, X., Y. Liu, C. C. Vigueira, N. D. Young, A. L. Caicedo, Y. Jia, D. R. Gealy, and K. M. Olsen. 2015. More than one way to evolve a weed: Parallel evolution of US weedy rice through independent genetic mechanisms. *Mol. Ecol.* 24:3329–3344.
- Quinlan, A. R., and I. M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- R Development Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>, retrieved on 25 Sept 2017.
- Ramu, P., W. Esuma, R. Kawuki, I. Y. Rabbi, C. Egesi, J. V Bredeson, R. S. Bart, J. Verma, E. S. Buckler, and

- F. Lu. 2017. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* 49:959–963.
- Rashid, M., H. Guangyuan, Y. Guangxiao, J. Hussain, and Y. Xu. 2012. AP2/ERF transcription factor in rice: Genome-wide canvas and syntenic relationships between monocots and eudicots. *Evol. Bioinforma.* 2012:321–355.
- Reed, E., S. Nunez, D. Kulp, J. Qian, M. P. Reilly, and A. S. Foulkes. 2015. A guide to genome-wide association analysis and post-analytic interrogation. *Stat. Med.* 34:3769–3792.
- Renaut, S., N. Maillet, E. Normandeau, C. Sauvage, N. Derome, S. M. Rogers, and L. Bernatchez. 2012. Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philos. Trans. R. Soc. B Biol. Sci.* 367:354–363.
- Reznick, D. N., F. H. Rodd, and M. Cadenas. 1996. Life-history evolution in guppies (*Poecilia reticulata*: Poeciliidae). IV. Parallelism in life-history phenotypes. *Am. Nat.* 147:319–338.
- Rieseberg, L. H., C. Van Fossen, and A. M. Desrochers. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375:313–316.
- Roach, D. A., and R. D. Wulff. 1987. Maternal effects in plants. *Annu. Rev. Ecol. Syst.* 18:209–235.
- Robertson, R. R. 2010. Physiological and biochemical characterization of glyphosate resistant *Ambrosia trifida* L. (Master's thesis). Retrieved from: Purdue University on 10 Dec 2017.
- Rogers, A. R., and C. Huff. 2009. Linkage disequilibrium between loci with unknown phase. *Genetics* 182:839–844.
- Rogers, C. E., T. E. Thompson, and G. J. Seiler. 1982. Sunflower Species of the United States. National Sunflower Association, Fargo, ND.
- Rohland, N., and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* 22:939–946.
- Rowan, B. A., V. Patel, D. Weigel, and K. Schneeberger. 2015. Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3* 5:385–398.
- Rubin, C.-J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallböök, F. Besnier, Ö. Carlborg, B. Bed'hom, M. Tixier-Boichard, P. Jensen, P. Siegel, K. Lindblad-Toh, and L. Andersson. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464:587–591.
- Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, N. C. Ellstrand, D. E. McCauley, P. O'Neil, I. M. Parker, J. N. Thompson, and S. G. Weller. 2001. The population biology of invasive species. *Annu. Rev. Ecol. Syst.* 32:305–332.
- Sakuma, Y., K. Maruyama, F. Qin, Y. Osakabe, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2006. Dual function of an Arabidopsis transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc. Natl. Acad. Sci.* 103:18822–18827.
- Sala, O. E., F. S. Chapin Iii, J. J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L. F. Hueneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. Poff, M. T. Sykes, B. H. Walker, M. Walker, and D. H. Wall. 2000. Global biodiversity scenarios for

- the year 2100. *Science* 287:1770–1774.
- Salisbury, S. E. 1962. *Weeds and Aliens*. Collins, London.
- Sambatti, J. B. M., and K. J. Rice. 2006. Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution* 60:696–710.
- Sammons, R. D., and T. A. Gaines. 2014a. Glyphosate resistance: State of knowledge. *Pest Manag. Sci.* 70:1367–1377.
- Sammons, R. D., and T. A. Gaines. 2014b. *Glyphosate resistance: State of knowledge*. John Wiley & Sons, Inc., Hoboken.
- Samuk, K., G. L. Owens, K. E. Delmore, S. E. Miller, D. J. Rennison, and D. Schluter. 2017. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Mol. Ecol.* 26:4378–4390.
- Schilling, E.E., 2006. *Helianthus*. Pp. 141-169 in *Flora of North America North of Mexico*. Oxford University Press, New York and Oxford.
- Schork, A. J., W. K. Thompson, P. Pham, A. Torkamani, J. C. Roddey, P. F. Sullivan, J. R. Kelsoe, M. C. O'Donovan, H. Furberg, N. J. Schork, O. A. Andreassen, and A. M. Dale. 2013. All SNPs are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* 9:e1003449. doi: 10.1371/journal.pgen.1003449
- Schuster, C. L., D. E. Shoup, and K. Al-Khatib. 2007. Response of common lambsquarters (*Chenopodium album*) to glyphosate as affected by growth stage. *Weed Sci.* 55:147–151.
- Schweizer, E. E., and L. D. Bridge. 1982. Sunflower (*Helianthus annuus*) and velvetleaf (*Abutilon theophrasti*) interference in sugarbeets (*Beta vulgaris*). *Weed Sci.* 30:514–519.
- Seiler, G. J., T. J. Gulya, G. Kong, S. Thompson, and J. Mitchell. 2008. Collection of wild naturalized sunflowers from the land down under. Pp. 10–11 in *Proceedings of the 30th sunflower research workshop*. National Sunflower Association, Fargo, ND.
- Seiler, G. J. 2010. Germination and viability of wild sunflower species achenes stored at room temperature for 20 years. *Seed Sci. Technol.* 38:786–791.
- Seok, H.-Y., V. N. Tarte, S.-Y. Lee, H.-Y. Park, and Y.-H. Moon. 2014. *Arabidopsis* HRE1 α , a splicing variant of AtERF73/HRE1, functions as a nuclear transcription activator in hypoxia response and root development. *Plant Cell Rep.* 33:1255–1262.
- Shagina, I., E. Bogdanova, I. Mamedov, Y. Lebedev, S. Lukyanov, and D. Shagin. 2010. Normalization of genomic DNA using duplex-specific nuclease. *Biotechniques* 48:455–459.
- Sham, P. C., and S. M. Purcell. 2014. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15:335–346.
- Shaner, D. L. 2009. Role of translocation as a mechanism of resistance to glyphosate. *Weed Sci.* 57:118–123.
- Shaner, D. L. 2010. Testing methods for glyphosate resistance. Pp. 93-118 in *Glyphosate Resistance in Crops and Weeds: History, Development and Management*. John Wiley & Sons, Inc., Hoboken.
- Shin, J., and C. Lee. 2015. Statistical power for identifying nucleotide markers associated with

- quantitative traits in genome-wide association analysis using a mixed model. *Genomics* 105:1–4.
- Shirano, Y., P. Kachroo, J. Shah, and D. F. Klessig. 2002. A gain-of-function mutation in an *Arabidopsis* Toll Interleukin-1 Receptor-Nucleotide Binding Site-Leucine-Rich Repeat type R gene triggers defense responses and results in enhanced disease resistance. *Plant Cell* 14:3149–3162.
- Shrestha, A., K. J. Hembree, and N. Va. 2007. Growth stage influences level of resistance in glyphosate-resistant horseweed. *Calif. Agric.* 61:67–70.
- Skiryycz, A., H. Claeys, S. De Bodt, A. Oikawa, S. Shinoda, M. Andriankaja, K. Maleux, N. B. Eloy, F. Coppens, S.-D. Yoo, K. Saito, and D. Inzé. 2011. Pause-and-stop: The effects of osmotic stress on cell proliferation during early leaf development in *Arabidopsis* and a role for ethylene signaling in cell cycle arrest. *Plant Cell* 23:1876–1888.
- Smith, B. D. 1989. Origins of agriculture in eastern north america. *Science* 246:1566-1571.
- Smith, R. J. J. 1988. Weed thresholds in Southern U.S. rice, *Oryza sativa*. *Weed Technol.* 2:232–241.
- Snaydon, R. W. 1980. Plant demography in agricultural systems. Pp. 131–160 in O. T. Solbrig, ed. *Demography and Evolution in Plant Populations*. The Alden Press, Oxford.
- Snow, A. A., P. Moran-Palma, L. H. Rieseberg, A. Wszelaki, and G. J. Seiler. 1998. Fecundity, phenology, and seed dormancy of F1 wild-crop hybrids in sunflower (*Helianthus annuus*, Asteraceae). *Am. J. Bot.* 85:794–801.
- Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig. 2007. pcaMethods - A bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23:1164–1167.
- Stapley, J., J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, and J. Slate. 2010. Adaptation genomics: The next generation. *Trends Ecol. Evol.* 15:705-712.
- Staton, S. E., B. H. Bakken, B. K. Blackman, M. A. Chapman, N. C. Kane, S. Tang, M. C. Ungerer, S. J. Knapp, L. H. Rieseberg, and J. M. Burke. 2012. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* 72:142-153.
- Staub, J. M., L. Brand, M. Tran, Y. Kong, and S. G. Rogers. 2012. Bacterial glyphosate resistance conferred by overexpression of an *E. coli* membrane eZux transporter. *J. Ind. Microbiol. Biotechnol.* 39:641–647.
- Steane, D. A., B. M. Potts, E. H. McLean, L. Collins, B. R. Holland, S. M. Prober, W. D. Stock, R. E. Vaillancourt, and M. Byrne. 2017. Genomic scans across three eucalypts suggest that adaptation to aridity is a genome-wide phenomenon. *Genome Biol. Evol.* 9:253–265.
- Stebbins, J. C., C. J. Winchell, and J. V. H. Constable. 2013. *Helianthus winteri* (Asteraceae), a new perennial species from the southern Sierra Nevada foothills, California. *Aliso* 31:19–24.
- Steenis, C. G. G. J. van. 1955. Specific and infraspecific delimitation. Pp. 167–234 in *Flora Malesiana-Series 1, Spermatophyta*. National Herbarium of the Netherlands, Leiden.
- Steinrücken, H. C., and N. Amrhein. 1980. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. *Biochem. Biophys. Res. Commun.* 94:1207–1212.
- Stewart, C. N., P. J. Tranel, D. P. Horvath, J. V Anderson, L. H. Rieseberg, J. H. Westwood, C. A. Mallory-

- Smith, M. L. Zapiola, and K. M. Dlugosch. 2009. Evolution of weediness and invasiveness: Charting the course for weed genomics. *Weed Sci.* 57:451–462.
- Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100:9440–5.
- Stranger, B. E., E. A. Stahl, and T. Raj. 2011. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187:367–383.
- Susko, D. J., and L. Lovett-Doust. 2000. Patterns of seed mass variation and their effects on seedling traits in *Alliaria petiolata* (Brassicaceae). *Am. J. Bot.* 87:56–66.
- Theriault, G., and K. K. Nkongolo. 2017. Evidence of prokaryote like protein associated with nickel resistance in higher plants: horizontal transfer of TonB-dependent receptor/protein in *Betula* genus or de novo mechanisms? *Heredity* 118:358–365.
- Tian, D., M. B. Traw, J. Q. Chen, M. Kreitman, and J. Bergelson. 2003. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423:74–77.
- Tranel, P. J., and D. P. Horvath. 2009. Molecular biology and genomics: New tools for weed science. *Bioscience* 59:207–215.
- Tranel, P. J., and T. R. Wright. 2002. Resistance of weeds to ALS-inhibiting herbicides: what have we learned? *Weed Sci.* 50:700–712.
- Turner, S. D. 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv 005165; <https://doi.org/10.1101/005165>, retrieved on 29 Sept 2017.
- Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42:260–263.
- Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genet.* 7:e1001336. doi: 10.1371/journal.pgen.1001336
- USDA, NRCS. 2017. The PLANTS Database National Plant Data Team, Greensboro, North Carolina.
- Vaahtera, L., and M. Brosché. 2011. More than the sum of its parts - How to achieve a specific transcriptional response to abiotic stress. *Plant Sci.* 180:421–430.
- Valente, M. A. S., J. A. Q. A. Faria, J. R. L. Soares-Ramos, P. A. B. Reis, G. L. Pinheiro, N. D. Piovesan, A. T. Morais, C. C. Menezes, M. A. O. Cano, L. G. Fietto, M. E. Loureiro, F. J. L. Aragão, and E. P. B. Fontes. 2009. The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. *J. Exp. Bot.* 60:533–46.
- Van Der Valk, A. G. 2005. Water-level fluctuations in North American prairie wetlands. *Hydrobiologia* 539:171–188.
- VanGessel, M. J. 2001. Glyphosate-resistant horseweed from Delaware. *Weed Sci.* 49:703–705.
- Vats, S. 2015. Herbicides: History, classification and genetic manipulation of plants for herbicide resistance. Pp. 153–192 in E. Lichtfouse, ed. *Sustainable Agriculture Reviews*. Springer, New York.
- Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Springer, New York.

- Vigueira, C. C., K. M. Olsen, and A. L. Caicedo. 2013. The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity* 110:303–311.
- Vila-Aiub, M. M., M. C. Balbi, P. E. Gundel, C. M. Ghersa, and S. B. Powles. 2007. Evolution of glyphosate-resistant Johnsongrass (*Sorghum halepense*) in glyphosate-resistant soybean. *Weed Sci.* 55:566–571.
- Vila-Aiub, M. M., P. Neve, and S. B. Powles. 2009. Fitness costs associated with evolved herbicide resistance genes in plants. *New Phytol.* 184:751–767.
- Vitousek, P. M., H. A. Mooney, J. Lubchenko, and J. M. Melillo. 1997. Human domination of earth's ecosystems. *Science* 277:494–499.
- Wakelin, A. M., and C. Preston. 2006. The cost of glyphosate resistance: is there a fitness penalty associated with glyphosate resistance in annual ryegrass? 15th Aust. Weeds Conf. Pap. Proceedings 17:515–518.
- Waltz, A. L., A. R. Martin, F. W. Roeth, and J. L. Lindquist. 2004. Glyphosate efficacy on velvetleaf varies with application time of day. *Weed Technol.* 18:931–939.
- Wang, T., A. Hamann, D. Spittlehouse, and C. Carroll. 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS One* 11. doi: 10.1371/journal.pone.0156720
- Weir, B. S. 2008. Linkage disequilibrium and association mapping. *Annu. Rev. Genomics Hum. Genet.* 9:129–142.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Westram, A. M., J. Galindo, M. Alm Rosenblad, J. W. Grahame, M. Panova, and R. K. Butlin. 2014. Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Mol. Ecol.* 23:4603–4616.
- Whitney, K. D., and C. A. Gabler. 2008. Rapid evolution in introduced species, “invasive traits” and recipient communities: Challenges for predicting invasive potential. *Divers. Distrib.* 14:569–580.
- Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytol.* 187:230–239.
- Whitton, J., D. E. Wolf, D. M. Arias, A. A. Snow, and L. H. Rieseberg. 1997. The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization. *Theor. Appl. Genet.* 95:33–40.
- Wickham, H. 2009. *ggplot2* Elegant Graphics for Data Analysis. Springer, New York.
- Williamson, M. 1996. *Biological invasions*. Chapman & Hall, London, UK.
- Wolfe, L. M., J. A. Elzinga, and A. Biere. 2004. Increased susceptibility to enemies following introduction in the invasive plant *Silene latifolia*. *Ecol. Lett.* 7:813–820.
- Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73:3–36.
- Wood, T. E., J. M. Burke, and L. H. Rieseberg. 2005. Parallel genotypic adaptation: When evolution

repeats itself. *Genetica* 123:157–170.

- Wrzaczek, M., M. Brosché, J. Salojärvi, S. Kangasjärvi, N. Idänheimo, S. Mersmann, S. Robatzek, S. Karpiński, B. Karpińska, and J. Kangasjärvi. 2010. Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in *Arabidopsis*. *BMC Plant Biol.* 10:95. doi: 10.1186/1471-2229-10-95
- Xu, Z., L. Escamilla-Treviño, L. Zeng, M. Lalgondar, D. Bevan, B. Winkel, A. Mohamed, C.-L. Cheng, M.-C. Shih, J. Poulton, and A. Esen. 2004. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* 55:343–367.
- Yeaman, S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci.* 110:E1743–E1751. doi: 10.1073/pnas.1219381110
- Yu, H., Y. Xu, E. L. Tan, and P. P. Kumar. 2002. AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. *Proc. Natl. Acad. Sci. U. S. A.* 99:16336–16341.
- Yu, J., and E. S. Buckler. 2006. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17:155-160.
- Yu, J., G. Pressoir, W. H. Briggs, I. Vroh 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.* 38:203–208.
- Yu, Q., A. Jalaludin, H. Han, M. Chen, R. D. Sammons, and S. B. Powles. 2015. Evolution of a double amino acid substitution in the 5-Enolpyruvylshikimate-3-Phosphate Synthase in *Eleusine indica* conferring high-level glyphosate resistance. *Plant Physiol.* 167:1440–1447.
- Yuan, J. S., L. L. G. Abercrombie, Y. Cao, M. D. Halfhill, X. Zhou, Y. Peng, J. Hu, M. R. Rao, G. R. Heck, T. J. Larosa, R. D. Sammons, X. Wang, P. Ranjan, D. H. Johnson, P. A. Wadl, B. E. Scheffler, T. A. Rinehart, R. N. Trigiano, and C. N. Stewart. 2010. Functional genomics analysis of horseweed (*Conyza canadensis*) with special reference to the evolution of non-target-site glyphosate resistance. *Weed Sci.* 58:109–117.
- Yuan, J. S., P. J. Tranel, and C. N. Stewart. 2007. Non-target-site herbicide resistance: a family business. *Trends Plant Sci.* 12:6–13.
- Yun, H. S., B. G. Kang, and C. Kwon. 2016. *Arabidopsis* immune secretory pathways to powdery mildew fungi. *Plant Signal. Behav.* 11:e1226456. doi: 10.1080/15592324.2016.1226456
- Zas, R., C. Cendán, and L. Sampedro. 2013. Mediation of seed provisioning in the transmission of environmental maternal effects in Maritime pine (*Pinus pinaster* Aiton). *Heredity* 111:248–255.
- Zelaya, I. A., and M. D. Owen. 2005. Differential response of *Amaranthus tuberculatus* (Moq ex DC) JD Sauer to glyphosate. *Pest Manag. Sci.* 61:936-950.
- Zhang, J. 1996. Seed mass effects across environments in an annual dune plant. *Ann. Bot.* 77:555–563.
- Zhang, Q., H. Li, R. Li, R. Hu, C. Fan, F. Chen, Z. Wang, X. Liu, Y. Fu, and C. Lin. 2008. Association of the circadian rhythmic expression of GmCRY1a with a latitudinal cline in photoperiodic flowering of soybean. *Proc. Natl. Acad. Sci.* 105:21028–21033.
- Zhao, K., M. J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram, and M. Nordborg. 2007. An *Arabidopsis* example of association mapping in structured

samples. PLoS Genet. 3:0071–0082. doi: 10.1371/journal.pgen.0030004

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328.

Zhou, S.-M., X.-Z. Kong, H.-H. Kang, X.-D. Sun, and W. Wang. 2015. The involvement of wheat F-box protein gene *TaFBA1* in the oxidative stress tolerance of plants. *PLoS One* 10:e0122117. doi: 10.1371/journal.pone.0122117

Zipfel, C., G. Kunze, D. Chinchilla, A. Caniard, J. D. G. Jones, T. Boller, and G. Felix. 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts agrobacterium-mediated transformation. *Cell* 125:749–760.

Zulet, A., M. Gil-Monreal, J. G. Villamor, A. Zabalza, R. A. L. van der Hoorn, and M. Royuela. 2013. Proteolytic pathways induced by herbicides that inhibit amino acid biosynthesis. *PLoS One* 8:e73847. doi: 10.1371/journal.pone.0073847

Appendices

Appendix A: Gene List for Genome Scan Results

Chromosome	Start Position (bp)	Gene ID	Protein Name
HanXRQChr01	17330075	HanXRQChr01g0003741	Nucleotide-binding alpha-beta plait domain
HanXRQChr01	17332821	HanXRQChr01g0003751	S-locus lectin protein kinase family protein
HanXRQChr01	18417984	HanXRQChr01g0003821	Dormancy-associated protein-like 1
HanXRQChr01	18479858	HanXRQChr01g0003841	DEHYDRATION-INDUCED 19 homolog 3
HanXRQChr01	18796866	<i>not assigned</i>	ATP-dependent DNA helicase PIF1
HanXRQChr01	18800133	HanXRQChr01g0003891	Phospholipase A1-II 1
HanXRQChr01	19967814	HanXRQChr01g0004001	Protein kinase superfamily protein
HanXRQChr01	20037080	HanXRQChr01g0004011	Cellulose synthase 1
HanXRQChr01	23290490	HanXRQChr01g0004401	Gnk2-homologous
HanXRQChr01	23554582	HanXRQChr01g0004421	SKP1/BTB/POZ domain; NPH3 domain
HanXRQChr01	34310431	HanXRQChr01g0005071	NAD(P)H dehydrogenase B2
HanXRQChr01	36544046	HanXRQChr01g0005331	DREB2A-interacting protein 1
HanXRQChr01	38347516	HanXRQChr01g0005501	NAD(P)-binding Rossmann-fold superfamily protein
HanXRQChr01	38758339	HanXRQChr01g0005661	ABC transporter; P-loop containing nucleoside triphosphate hydrolase
HanXRQChr01	38759244	HanXRQChr01g0005671	Beige/BEACH domain; WD domain, G-beta repeat protein
HanXRQChr01	38849478	HanXRQChr01g0005691	Sec23/Sec24 protein transport family protein
HanXRQChr01	45063685	HanXRQChr01g0006491	Serine/threonine/dual specificity protein kinase, catalytic domain
HanXRQChr01	114488596	HanXRQChr01g0019491	Peptidase S28; Alpha/Beta hydrolase fold
HanXRQChr02	10158511	HanXRQChr02g0033041	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein
HanXRQChr02	10164572	HanXRQChr02g0033051	Mce/MlaD
HanXRQChr02	156230966	HanXRQChr02g0053201	Aluminium activated malate transporter family protein
HanXRQChr03	17145104	<i>not assigned</i>	Transmembrane protein
HanXRQChr03	85176295	HanXRQChr03g0072551	Isopentenyltransferase 5
HanXRQChr03	114444888	HanXRQChr03g0077231	Cyclic nucleotide-gated cation channel 4
HanXRQChr03	125903755	<i>not assigned</i>	Nuclease HARBI1
HanXRQChr03	125917681	HanXRQChr03g0079661	Glycoside hydrolase family 1
HanXRQChr04	62271069	HanXRQChr04g0106061	Alpha/beta-Hydrolases superfamily protein
HanXRQChr04	115231934	HanXRQChr04g0112531	Integrase-type DNA-binding superfamily protein
HanXRQChr04	165740641	<i>not assigned</i>	Ribonuclease H protein At1g65750-like
HanXRQChr04	165742494	HanXRQChr04g0123491	Oxoglutarate/iron-dependent dioxygenase; Non-haem dioxygenase N-terminal domain
HanXRQChr05	17874237	HanXRQChr05g0132481	Cysteine-rich RLK (RECEPTOR-like protein kinase) 33
HanXRQChr05	204755331	HanXRQChr05g0158991	Flavoprotein-like domain
HanXRQChr06	67605734	HanXRQChr06g0181091	Peroxin4
HanXRQChr06	77964366	HanXRQChr06g0182221	SCO1 homolog, mitochondrial

Chromosome	Start Position (bp)	Gene ID	Protein Name
HanXRQChr06	100505079	HanXRQChr06g0184771	Hexokinase-2, chloroplastic
HanXRQChr06	100509474	<i>not assigned</i>	ATP-dependent DNA helicase PIF1
HanXRQChr06	100566760	<i>not assigned</i>	60S ribosomal protein L38
HanXRQChr06	100568100	<i>not assigned</i>	Nuclease HARBI1
HanXRQChr06	100612968	HanXRQChr06g0184781	F-box domain; F-box associated interaction domain
HanXRQChr07	10904213	HanXRQChr07g0187811	Thioredoxin-like fold
HanXRQChr08	41842349	HanXRQChr08g0219811	Serine/threonine/dual specificity protein kinase, catalytic domain
HanXRQChr08	70008833	<i>not assigned</i>	Nuclease HARBI1
HanXRQChr08	77313850	HanXRQChr08g0226021	Pectin lyase-like superfamily protein
HanXRQChr08	78435674	HanXRQChr08g0226151	GYF domain; Histone-lysine N-methyltransferase Trr
HanXRQChr08	112841373	<i>not assigned</i>	BONZAI 3-like
HanXRQChr08	112847826	HanXRQChr08g0231961	Extensin 3
HanXRQChr09	15879229	HanXRQChr09g0239981	NC domain-containing protein-related
HanXRQChr10	47386245	<i>not assigned</i>	Floral homeotic protein APETALA 2-like
HanXRQChr10	51064546	<i>not assigned</i>	Sulfoquinovosyldiacylglycerol 2
HanXRQChr10	77673557	<i>not assigned</i>	ATP-dependent DNA helicase PIF1-like isoform X1
HanXRQChr10	84839370	<i>not assigned</i>	Ribonuclease H protein At1g65750-like
HanXRQChr10	85297950	HanXRQChr10g0291451	LRR receptor-like serine/threonine-protein kinase EFR
HanXRQChr10	155024465	HanXRQChr10g0300601	Tyrosine-protein kinase; receptor ROR
HanXRQChr10	177672770	HanXRQChr10g0303601	Bifunctional riboflavin biosynthesis protein RIBA
HanXRQChr10	214895610	HanXRQChr10g0310551	Rab5-interacting protein family
HanXRQChr11	158353603	<i>not assigned</i>	Adenosine-5'-phosphosulfate (APS) kinase 4
HanXRQChr12	42968218	HanXRQChr12g0366451	Cytochrome P450
HanXRQChr12	64164437	HanXRQChr12g0370371	CDP-alcohol phosphatidyltransferase
HanXRQChr13	62236872	HanXRQChr13g0395661	Plant VAMP (vesicle-associated membrane protein) family protein
HanXRQChr13	99378265	HanXRQChr13g0403571	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
HanXRQChr13	99380953	<i>not assigned</i>	Transmembrane protein
HanXRQChr13	100991585	HanXRQChr13g0403971	Leucine-rich repeat-containing N-terminal, plant-type
HanXRQChr13	186782972	<i>not assigned</i>	Cytochrome P450
HanXRQChr13	186791506	HanXRQChr13g0423381	Alpha/beta-Hydrolases superfamily protein
HanXRQChr15	79111145	<i>not assigned</i>	Luminal-binding protein 5
HanXRQChr15	79115849	HanXRQChr15g0483691	Zinc finger, RING/FYVE/PHD-type
HanXRQChr15	81098919	HanXRQChr15g0483901	Ascorbate peroxidase 3
HanXRQChr15	86385343	HanXRQChr15g0484901	AGAMOUS-like 24
HanXRQChr16	1271327	HanXRQChr16g0498001	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protein
HanXRQChr16	153871742	<i>not assigned</i>	Carotenoid 9,10(9',10')-cleavage dioxygenase 1-like isoform X1
HanXRQChr16	171951880	<i>not assigned</i>	GDSL esterase/lipase At3g48460-like

Chromosome	Start Position (bp)	Gene ID	Protein Name
HanXRQChr16	171958254	HanXRQChr16g0529381	SGNH hydrolase-type esterase domain
HanXRQChr17	174929190	<i>not assigned</i>	Ribonuclease H protein At1g65750-like
HanXRQChr17	174934473	<i>not assigned</i>	Proline-rich receptor-like protein kinase PERK8
HanXRQChr17	177995992	<i>not assigned</i>	Chlorophyll A-B binding protein
HanXRQChr17	177999628	HanXRQChr17g0564351	F-box domain; Leucine-rich repeat domain
HanXRQChr17	191494042	HanXRQChr17g0566361	RNI-like superfamily protein

Appendix B: Gene List for GWAS Results

Chromosome	Start Position (bp)	Gene ID	Protein Name
HanXRQChr01	86424884	HanXRQChr01g0014391	Auxin response factor ⁺
HanXRQChr01	86438342	HanXRQChr01g0014421	Nucleic acid-binding, OB-fold-like protein
HanXRQChr03	106213474	HanXRQChr03g0075421	Glycine-rich protein
HanXRQChr03	106241973	HanXRQChr03g0075431	Zinc finger, C2H2-like
HanXRQChr03	106263234	<i>not assigned</i>	Serine/threonine-protein kinase TOUSLED-like*
HanXRQChr03	106264767	HanXRQChr03g0075461	Calcium load-activated calcium channel*
HanXRQChr03	106266534	HanXRQChr03g0075471	IBR domain; E3 ubiquitin ligase RBR family
HanXRQChr03	106267708	HanXRQChr03g0075481	Adenine phosphoribosyl transferase
HanXRQChr03	106278046	<i>not assigned</i>	Ribonuclease H protein At1g65750-like*
HanXRQChr03	106294441	HanXRQChr03g0075491	Transposase, gypsy type
HanXRQChr08	105564979	HanXRQChr08g0230781	Mlo-related protein
HanXRQChr08	105615964	HanXRQChr08g0230811	RPA-interacting protein A*
HanXRQChr09	66975439	HanXRQChr09g0247281	BTB/POZ domain-containing protein FBL11
HanXRQChr09	66996355	HanXRQChr09g0247291	Plant PDR ABC-transporter associated
HanXRQChr12	31224957	<i>not assigned</i>	Ribonuclease H protein At1g65750-like*
HanXRQChr12	31361738	<i>not assigned</i>	Cysteine-rich RLK 8*
HanXRQChr12	31389482	<i>not assigned</i>	RNA-directed DNA polymerase*
HanXRQChr12	36686159	HanXRQChr12g0365101	Phosphatidylinositol 3-/4-kinase, catalytic domain
HanXRQChr12	36690789	<i>not assigned</i>	Cysteine-rich RLK 8*
HanXRQChr12	36705847	HanXRQChr12g0365111	DNA-directed RNA polymerase 1B, mitochondrial
HanXRQChr12	36723906	HanXRQChr12g0365131	Myc-type, basic helix-loop-helix (bHLH) domain
HanXRQChr12	104115414	HanXRQChr12g0377241	Major facilitator superfamily, sugar transporter-like
HanXRQChr12	134373597	HanXRQChr12g0380931	Plant PDR ABC-transporter associated
HanXRQChr12	138034586	<i>not assigned</i>	Photosystem I P700 chlorophyll a apoprotein A1-like*
HanXRQChr12	138054149	<i>not assigned</i>	Chromatin remodeling protein EBS*
HanXRQChr12	138084542	<i>not assigned</i>	Chromatin remodeling protein EBS*
HanXRQChr12	138097104	<i>not assigned</i>	DNA damage-inducible protein 1-like*
HanXRQChr12	140112626	HanXRQChr12g0381201	Alanine:glyoxylate aminotransferase*
HanXRQChr12	140510824	HanXRQChr12g0381221	Alpha/beta-Hydrolases superfamily protein*
HanXRQChr13	2165748	<i>not assigned</i>	Protein trichome birefringence-like 10*
HanXRQChr13	2169026	HanXRQChr13g0386871	SecY protein transport family protein
HanXRQChr13	65390194	<i>not assigned</i>	Nuclease HARBI1*
HanXRQChr13	67088450	HanXRQChr13g0396151	Zinc-finger protein 1
HanXRQChr15	7307152	HanXRQChr15g0465561	Alpha/beta-Hydrolases superfamily protein
HanXRQChr15	7316921	<i>not assigned</i>	60S ribosomal protein L18-2-like*
HanXRQChr15	7370878	<i>not assigned</i>	ATP-dependent DNA helicase PIF1*
HanXRQChr16	117538989	HanXRQChr16g0517611	Erf domain protein 9

Chromosome	Start Position (bp)	Gene ID	Protein Name
HanXRQChr16	175652942	<i>not assigned</i>	Serine/threonine-protein kinase DDB*
HanXRQChr16	175653909	<i>not assigned</i>	Mitochondrial protein AtMg00810-like*
HanXRQChr16	175679027	<i>not assigned</i>	Cysteine-rich RLK 8*
HanXRQChr16	175715161	<i>not assigned</i>	Glucan endo-1,3-beta-glucosidase 9-like*
HanXRQChr16	175727613	HanXRQChr16g0530461	Zinc finger, NHR/GATA-type

*Indicates that a putative function was assigned based on homology to proteins in other plant species, using available GenBank gene identifier numbers.

†Light grey shading indicates genes associated with SNPs in the top six suggestive peaks.