POPULATION GENETICS OF HERBICIDE RESISTANCE IN WATERHEMP

BY

KATE ANNE THINGLUM

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Master's Committee:

Professor Patrick J. Tranel, Chair
Assistant Professor Adam S. Davis
Assistant Professor Yoshie Hanzawa

ABSTRACT

Waterhemp populations resistant to herbicides that inhibit protoporphyrinogen oxidase (PPO) have been identified throughout the Midwest and the mechanism of resistance in populations from Illinois was previously determined to result from a 3 base pair deletion in the protoporphyrinogen oxidase (PPO)-encoding gene, *PPX2*. The deletion results in the loss of glycine at position 210 (ΔG210). While resistance is becoming more widespread and waterhemp is becoming more problematic in corn and soybean fields, little is known about waterhemp's population genetics. The general purpose of this thesis was to gain insight into the population genetics of waterhemp resistant to PPO-inhibiting herbicides. Do all resistant populations have the same mechanism for resistance to PPO-inhibiting herbicides? How did resistance spread? Chapter One reviews the literature surrounding these important issues. Chapter Two discusses research that was done to determine if ΔG210, the previously identified mutation, is responsible for resistance in waterhemp populations outside of Illinois and examines the possible mechanisms for the spread of resistance. These studies revealed that ΔG210 is the only resistance mechanism identified thus far in waterhemp resistant to PPO-inhibitors and the mutation correlates highly with resistance. However, the mechanism of the spread of resistance is not as clear-cut. Genetic analysis did not distinguish whether independent evolution or gene flow was responsible for the spread of resistance, possible due to recombination within the *PPX2* gene. Chapter Three discusses the development of a set of microsatellite markers for use in waterhemp population genetics. Although these markers were not useful in determining the spread of herbicide

resistant waterhemp in Illinois, they could be useful in a variety of other population studies. The fourth and final chapter discusses concluding remarks, future research, and drawbacks.

To Mom and Dad

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Patrick Tranel for his continued guidance and support. Without him, and his patience and immense knowledge, this thesis would not have been possible. I would also like to thank my committee members: Dr. Adam Davis, for sparking my interest in weed science and for always being willing to share his knowledge with me, I express my sincerest gratitude; and to Dr. Yoshie Hanzawa, for always being available to answer my questions and have a good chat, I am very thankful.

It is a pleasure to thank my fellow lab members, past and present. Michael Bell, I am not sure I would have kept my sanity throughout graduate school without our long talks at Espresso Royal. To Sukhvinder Singh, thank you for always lending support and a good story to make me laugh. To Stephanie Rousonelos, I am grateful for your friendship and advice. To Dr. Ryan Lee, thank you for guiding me at the beginning of my research career and giving me my very first lab notebook. To Dr. Jianyang Liu, I appreciate your kindness and willingness to help. Thank you to Dr. Chance Riggins for taking the time to teach me about phylogenetics.

I am extremely grateful to the University of Illinois for providing me with the best education possible and countless invaluable memories. I could not think of another place on earth that I would have rather spent these past 6 years.

Finally, I owe my deepest gratitude to my parents and my sister. Not only have they always encouraged me to continue my education, they have taught me the importance of hard work and keeping a positive attitude. To them I am forever grateful.

TABLE OF CONTENTS

CHAPTER 1 1
INTRODUCTION 1
1.1 Biology of Waterhemp 1
1.2 Mode of Action of Protoporphyrinogen Oxidase-Inhibiting Herbicides 3
1.3 Waterhemp's Unique Mechanism of Resistance to Protoporphyrinogen Oxidase-Inhibiting Herbicides 4
1.4 Glyphosate Resistant Waterhemp 5
1.5 Molecular Markers for Population Genetics Studies 7
1.6 DNA Sequencing Technologies 7
1.7 Research Objectives 9
1.8 Attributions 11
1.9 Literature Cited 11

CHAPTER 2 17
INVESTIGATION INTO THE PROTOPORPHYRINOGEN OXIDASE GLYCINE DELETION AS WATERHEMP'S RESISTANCE MECHANISM TO PPO-INHIBITING HERBICIDES 17
2.1 Abstract 17
2.2 Introduction 18
2.3 Materials and Methods 19
2.4 Results and Discussion 26
2.5 Sources of Materials 32
2.6 Literature Cited 33
2.7 Figures and Tables 37

CHAPTER 3 46
USE OF SIMPLE SEQUENCE REPEAT MARKERS FOR WATERHEMP POPULATION GENETICS 46
3.1 Abstract 46
3.2 Introduction 47
3.3 Materials and Methods 48
3.4 Results and Discussion 52
3.5 Sources of Materials 54
3.6 Literature Cited 55
3.7 Figures and Tables 57

CHAPTER 4 65
CONCLUDING REMARKS 65
4.1 Conclusions and Future Directions 65
4.2 Literature Cited 69

CHAPTER 1

INTRODUCTION

1.1 Biology of Waterhemp

Waterhemp (*Amaranthus tuberculatus*) is a broadleaf weed that is indigenous to Illinois. Along with other pigweeds, it is a member of the Amaranth family. It is a dioecious summer annual that has caused problems for farmers in Midwestern soybean and corn fields for years, specifically since the 1990s with a change in farming practices that favored waterhemp's germination and growth (Hager et al. 2002; Buhler 1992). In fact, a survey done by Hager and Sprague (2002) found that waterhemp, along with giant foxtail and giant ragweed, was most frequently ranked as the most prevalent weed species by Illinois farmers. Furthermore, the same survey found that waterhemp was one of the three most common weeds to escape herbicide treatment in both soybean and corn fields in Illinois (Hager and Sprague 2002).

Many factors contribute to waterhemp's ability to invade agronomic fields and compete with the crop. First of all, each waterhemp plant has the ability to produce over 300,000 small seeds even under unfavorable conditions. In favorable environments, each plant can produce up to 5 million small seeds (Hartzler 2003). Not only does waterhemp produce a large amount of seeds, but these seeds are long-lived, have a high germination rate, and possess the ability to germinate throughout the growing season, requiring season-long weed control. These small seeds favored the switch to no-till farming systems because they were no longer pushed deep into the soil; rather they were able to

remain near the surface where germination was favorable (Buhler 1992). Even waterhemp seeds that germinate later in the growing season pose a threat to the crop due to waterhemp's rapid growth rate (Horak and Loughin 2000). With its ability to reach heights of 2 to 3 m in height, it is a tough competitor in agronomic fields (Bensch et al. 2003). Finally, since waterhemp is dioecious and requires cross-pollination, there is a high level of variability. Therefore, favorable traits, including herbicide resistance traits, can spread rather quickly over short distances from plant to plant (Hatzler 2003). Furthermore, it is thought that on a windy day waterhemp pollen can travel over a half mile from the source plant, allowing longer distance means of pollen dispersal (Costea et al. 2005). Waterhemp's favorable traits can also be spread via seed. Longer distance means of seed dispersal include the transportation of waterhemp's small seeds via vectors such as animals, waterways, or farm equipment.

The aforementioned factors have allowed waterhemp to acquire resistance to several commonly used herbicides. With repeated use of single herbicide modes of action, waterhemp has evolved resistance to several different modes of action. And because waterhemp outcrosses with other waterhemp plants, herbicide resistance traits are able to spread quickly (Patzoldt et al. 2005). However, herbicide resistant waterhemp was not reported until the 1990s, when the weed became much more prevalent in the no-till farming systems. Waterhemp's first case of herbicide resistance was reported in 1990 in southeast Nebraska, where triazine resistant waterhemp was confirmed (Anderson et al. 1996). In 1993, waterhemp resistant to acetolactate synthase (ALS)-inhibiting herbicides was confirmed in Kansas (Horak and Peterson 1995). Waterhemp resistant to PPO-inhibiting herbicides was reported in Kansas in 2001 and waterhemp in Missouri

was reported to be resistant to glyphosate in 2005 (Li et al. 2004; www.weedscience.org). The focus of this thesis is on PPO-inhibiting herbicide resistance in waterhemp, so the next sections will discuss PPO-inhibiting herbicides and waterhemp's unique mechanism of resistance to this herbicide mode of action.

1.2 Mode of Action of Protoporphyrinogen Oxidase-Inhibiting Herbicides

Protoporphyrinogen oxidase-inhibiting herbicides, or PPO-inhibitors, are post-applied, contact burners that are used to control broadleaf weeds. They act on the last common step of the heme and chlorophyll biosynthesis pathway by inhibiting the protoporphyrinogen oxidase (PPO) enzyme in plastid membranes (Beale and Weinstein 1990). This inhibition causes protoporphyrinogen IX to leak out of the plastid membrane and into the cytoplasm, where it is converted to protoporphyrin IX, a photosynthetically active chemical that produces toxic singlet oxygen species (Jacobs and Jacobs 1984; Lee and Duke 1994). The singlet oxygen species cause lipid peroxidation and damage to the cell membrane, ultimately leading to plant death (Duke et al. 1991). Rapid metabolic detoxification of the herbicide is one method responsible for crop selectivity to PPO-inhibiting herbicides. A double mutant protoporphyrinogen oxidase enzyme from *Arabidopsis thaliana* was responsible for the first PPO-resistant corn line, although it was not commercialized. Other methods of crop selectivity to PPO-inhibitors include overexpression of protoporphyrinogen oxidase enzyme from *Bacillus subtilis* that is naturally resistant to PPO-inhibiting herbicides and increasing gene copy number (Green 2009). It has been reported that PPO enzyme mutations reduce enzyme activity, which

could explain why few weeds are naturally resistant to PPO-inhibiting herbicides, even though these herbicides have been used for over 30 years (Li et al. 2003). However, both waterhemp and common ragweed have acquired resistance to this herbicide mode of action due to target site alterations (Lee et al. 2008; Rousonelos et al. 2009).

PPO-inhibiting herbicide families include diphenylethers, N-phenylphthalamides, and triazolinones (Hartzler 1998). Among these herbicide chemistries, waterhemp has acquired resistance to three of the active ingredients in the diphenylether family, acifluorfen (5-(2-chloro-α,α,α-trifluoro-*p*-tolyloxy)-2-nitrobenzoic acid), fomesafen (5-(2-chloro-α,α,α-trifluoro-*p*-tolyloxy)-*N*-mesyl-2-nitrobenzamide), and lactofen (ethyl *O*-[5-(2-chloro-α,α,α-trifluoro-*p*-tolyloxy)-2-nitrobenzoyl]-DL-lactate, two of the active ingredients in the *N*-phenyl-phthalimides family, flumiclorac ([2-chloro-5-(cyclohex-1-ene-1,2-dicarboximido)-4-fluororophenoxy]acetic acid) and flumioxazin (*N*-(7-fluoro-3,4-dihydro-3-oxo-4-prop-2-ynyl-2*H*-1,4-benzoxazin-6-yl)cyclohex-1-ene-1,2-dicarboxamide), and one of the active ingredients in the triazolinone family, sulfentrazone (2′,4′-dichloro-5′-(4-difluoromethyl-4,5-dihydro-3-methyl-5-oxo-1*H*-1,2,4-triazol-1-yl)methanesulfonanilide) (Wood 2010; Patzoldt 2005).

1.3 Waterhemp's Unique Mechanism of Resistance to Protoporphyrinogen Oxidase-Inhibiting Herbicides

Although PPO-inhibiting herbicides have been used to control weeds for over 30 years, resistance was not identified until 2001, when waterhemp from Kansas evolved resistance to PPO-inhibiting herbicides (Li et al. 2004). Shortly after, resistance to these

herbicides was also confirmed in waterhemp from Illinois and Missouri (www.weedsciene.com). The resistance mechanism in waterhemp from Illinois was previously identified as a 3 base pair deletion at position 210 in exon 9 of the *PPX2* gene, resulting in the loss of a glycine. This mutation is referred to as ΔG210 (Patzoldt et al. 2006). Position 210 is located near the herbicide-binding site. The mutation has been found to confer resistance to active ingredients in all three families of PPO-inhibiting herbicides. This resistance mechanism is unusual, in that it involves an amino acid deletion, rather than a substitution, which is the more common target-site resistance mechanism. Despite being an unusual mutation, Lee et al. (2008) found that the resistance mechanism is common in Illinois waterhemp populations with PPO resistance; all six out of six resistant populations tested contained the deletion conferring resistance. While waterhemp resistant to PPO-inhibiting herbicides was identified nearly ten years ago and is fairly widespread throughout the Midwest, glyphosate-resistant waterhemp was not identified until 2005 and resistance is not yet as prevalent. Therefore, studying the population genetics of waterhemp resistant to PPO-inhibiting herbicides could provide insight into the development and spread of glyphosate-resistant waterhemp. The next section will provide information about glyphosate and waterhemp's resistance to this herbicide.

1.4 Glyphosate Resistant Waterhemp

Glyphosate is a nonselective, systemic herbicide that effectively controls most broadleaf weeds and grasses. It is a competitive inhibitor of EPSP (5-enolpyruvyl-

shikimate-3-phosphate) synthase, as it competes with phosphoenolpyruvate (PEP) for the binding site (Sikorski and Gruys 1997). As EPSP synthase is inhibited, the plant runs out of aromatic amino acids and cannot synthesize new proteins or make secondary compounds derived from the shikimate pathway. In healthy plants, regulation of the shikimate pathway occurs via feedback inhibition. However, since the herbicide causes a depletion of the pathway's end products, the shikimate pathway is unregulated, causing a drain of carbon through the pathway. Ultimately the plant dies due to starvation of essential compounds. Additionally, with inhibition of EPSP synthase, two toxic substrates accumulate, shikimate and shikimate-3-P, which contribute to plant death (Baylis 2000).

Seventeen weed species have evolved resistance to glyphosate, including two in the *Amaranthus* family, waterhemp and Palmer amaranth (*Amaranthus palmeri*). Glyphosate resistant waterhemp was first reported in Missouri in 2005, followed by Illinois (2005), Kansas (2006), Minnesota (2007), and Iowa (2009) (www.weedscience.org). While the mechanism of waterhemp resistance to glyphosate is unknown, it has been speculated that increased gene amplification of 5-enolpyruvylshikimate 3-phosphate **(**EPSP) synthase may be responsible for conferring resistance (Bell and Tranel, unpublished data). With an increase in weeds resistant to both PPO-inhibiting herbicides and glyphosate, the question is raised as to whether gene flow or independent selection contributes to the spread of herbicide resistant weeds. Molecular markers serve as an important tool to track the spread of alleles. Therefore, molecular markers could provide insight into the spread of herbicide resistance.

1.5 Molecular Markers for Population Genetics Studies

Molecular markers are useful in a variety of genetic experiments, including linkage analysis, population genetics, and phylogeny. Simple sequence repeats (SSRs) are regions of nucleotide repeats, usually one to four base pairs in length. SSRs are highly polymorphic due to their high mutation rate. Therefore, only a small number of markers are needed to resolve relatedness. This makes them particularly useful in population genetics studies within species (Mallory et al. 2008; Ellwood et al. 2006). SSRs are codominant, so SSR markers can identify both homozygotes and heterozygotes (Ellwood et al. 2006). Since waterhemp has become such a prevalent weed and is quickly evolving resistance to multiple herbicides, it is useful to have molecular markers for this plant so that it is possible to track its evolution and spread of weedy traits.

Traditionally, SSRs are identified from genomic DNA libraries. These genomic libraries consist of short fragments of an organism's DNA. The DNA library is then screened; one way to do this is using digoxigenin-labeled probes that bind to regions containing repeats. The regions that hybridize to the probe are then suspected to contain SSRs and these regions are isolated and sequenced using traditional PCR procedures. However, this approach is both time-consuming and expensive (Billote et al. 2001; He et al. 2003)

1.6 DNA Sequencing Technologies

New breakthroughs in sequencing technology have allowed for the rapid and accurate identification of SSRs throughout an organism's genome. Sanger sequencing,

developed in 1974, uses a DNA template and repeatedly replicates that template starting with the first base and ending with a random base downstream from the starting base. Each terminating base is labeled fluorescently. These fluorescently-labeled strands of varying lengths are then separated via electrophoresis and the sequence can be determined based on the distance each strand moves during electrophoresis, with the distance being inversely proportional to the length of the fragment (Sanger et al. 1997). Although this method has been used for years, it is expensive and timely (Wheeler et al. 2008). On the other hand, Roche 454 sequencing technology was developed in the 1990s and has a 100-fold higher throughput, is more ideal for larger sequencing experiments, and is less costly (Margulies et al. 2005; Wheeler et al. 2008). Roche 454 sequencing begins with a library of DNA fragments that are ligated to adapters. Each fragment is fixed to a bead, and the beads are then isolated and PCR-amplified. After amplification, each bead contains copies of the DNA template it was bound to. The beads are then placed in wells of a fiber-optic slide, where the 454 sequencing device sequences each fragment by detecting light production whenever a nucleotide is incorporated (Margulies et al. 2005). Therefore, while Sanger sequencing relies on electrophoresis to separate fragments, Roche 454 sequencing relies on the emission of light whenever a new nucleotide is incorporated. This gives Roche 454 sequencing an advantage by creating more accuracy and allowing smaller-sized fragments to be sequenced (Rothberg and Leaman 2008). Additionally, Roche 454 sequencing is ideal for large sequencing projects due to its lower cost. In fact, Wheeler et al. (2008) sequenced the entire human genome using Roche 454 sequencing for less than 1 million US dollars, whereas when Levy et al.

(2007) sequenced an entire human genome via Sanger sequencing, it reportedly cost 100 million US dollars.

Roche 454 sequencing techniques allow for the identification of SSRs throughout an organism's genome. This is done by assembling the reads from the Roche 454 sequencing into a sequence set using GS Assembler (Newbler) software. A Basic Local Alignment Search Tool (BLAST) is then used to annotate the sequences (Altschul et al. 1990). The annotated contiguous consensus sequences (contigs) are run through Repeat-Masker to search for known repeats (http://www.repeatmasker.org). Finally, SSR Finder searches through the contigs to identify SSRs (http://www.maizemap.org/bioinformatics/SSRFINDER/SSR_ Finder_Download.html).

1.7 Research Objectives

While several waterhemp populations with resistance to PPO-inhibiting herbicides in Illinois were shown to have the resistance-encoding glycine deletion, waterhemp populations outside of Illinois had not yet been screened for this resistance mechanism. Several resistance mechanisms exist in plants, including increased metabolic degradation of the herbicide, increased translocation of the herbicide, or target site alterations, such as insertions or deletions. Research was conducted to determine if waterhemp populations outside of Illinois conferring resistance to PPO-inhibiting herbicides also have the ΔG210 mechanism of resistance.

There are two ways for allele movement to occur: maternally (through seed movement) or paternally (through pollen movement). Many vectors may be able to

spread waterhemp seeds and pollen throughout the Midwest. The small seeds may travel long distances via farm equipment or barges. Acetyl CoA carboxylase (ACCase)-resistant oat (*Avena fatua*) was found to have spread due to farm machinery contaminated with resistant seeds (Andrews et al. 1998). Humans and animals could unknowingly carry the seed from one location to another. Also, pollen movement followed by fertilization could be responsible for the spread of waterhemp resistant to PPO-inhibiting herbicides. Costea et al. (2005) calculated that on a windy day, waterhemp pollen could travel a half-mile or farther from the host plant. Pollen flow was previously found to be responsible for the spread of imidazolinone resistance from domesticated sunflower to common sunflower (*Helianthus annuus*) and prairie sunflower (*Helianthus petiolaris*) (Massinga et al. 2003). Therefore, a second objective of this project was to determine if the allele conferring resistance to the herbicide in waterhemp was independently selected or if the allele was selected for once and subsequently able to spread throughout the Midwest via pollen or seed flow. While PPO-inhibiting herbicide resistance in waterhemp has evolved and spread in the past 9 nine years, glyphosate resistance in waterhemp was not seen until 2005 and, therefore, is fairly new. This study may be helpful in modeling glyphosate resistance and preventing its spread.

Molecular markers are very useful in a variety of genetic experiments, including linkage analysis, population genetics, and phylogeny. SSR markers are especially beneficial because they are codominant, highly polymorphic, and multiallelic (Mallory et al. 2008). A third objective of this project was to develop a set of SSR markers for use in waterhemp population genetics.

1.8 Attributions

Chapter 2 will be reformatted and submitted for publication in *Weed Science* and was a collaboration with Kassim Al-Khatib at Kansas State University and Kevin Bradley at The University of Missouri.

A portion of chapter 3 was published in Weed Science 2009, issue 57, pages 463-469 in the article Sampling the Waterhemp (*Amaranthus tuberculatus*) Genome Using Pyrosenquencing Technology by Ryan M. Lee, Jyothi Thimmapuram, Kate A. Thinglum, George Gong, Alvaro G. Hernandez, Chris L. Wright, Ryan W. Kim, Mark A. Mikel, and Patrick J. Tranel. The STRUCTURE analysis was done by Kate Waselkov in Kenneth M. Olsen's lab at Washington University.

1.9 Literature Cited

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, D. J. Lipman. 1990.Basic local alignment search tool. *J Mol Biol.* 215:403–410.

Anderson, D. D., F. W. Roeth, and A. R. Martin. 1996. Occurrence and control of triazine-resistant common waterhemp (*Amaranthus rudis*) in field corn (*Zea mays*). *Weed Technol.* 10:570–575.

Andrews, T. S., I. N. Morrison, and G. A. Penner. 1998. Monitoring the spread of ACCase inhibitor resistance among wild oat (*Avena fatua*) patches using AFLP analysis. *Weed Sci.* 46:196–199.

Baylis, A. D. 2000. Why glyphosate is a global herbicide: strength, weaknesses and prospects. *Pest Manag. Sci.* 56:299-308.

Beale, S. I. and J. D. Weinstein. 1990. *Tetrapyrrole Metabolism in Photosynthetic Organisms*, ed. Dailey, H. A. McGraw–Hill, New York

Bensch, C. N., M. J. Horak, and D. E. Peterson. 2003. Interference of redroot pigweed, Palmer amaranth, and common waterhemp in soybean. *Weed Sci.* 51:37–43.

Billotte, N., A. M. Risterucci, E. Barcelos, J. L. Noyer, P. Amblard, and F. C. Baurens. 2001. Development, characterization, and across-taxa utility of oil palm (*Elaeis guineensis Jacq*.) microsatellite markers. *Genome* 44:413–425.

Buhler, D. D.1992. Population dynamics and control of annual weeds in corn (Zea mays) as influenced by tillage systems. *Weed Sci*.40:241- 248.

Costea, M., S. E.Weaver, and F. J. Tardif. 2005. The Biology of Invasive Alien Plants in Canada. 3. *Amaranthus tuberculatus* (Moq.) Sauer var. *rudis* (Sauer). *Can. J. Plant Sci.* 85:507-522.

Duke, S. O., J. Lydon,, J. M. Becerril, T. D. Sherman, L. P. Lehnen, and H. Matsumoto. 1991. Protoporphyrinogen Oxidase-Inhibiting Herbicides. *Weed Sci.* 39: 465– 473.

Ellwood S.R., N.K. D' Souza, L.G. Kamphuis, T.I. Burgess, R.M. Nair, R.P. Oliver. 2006. SSR analysis of the Medicagotruncatula SARDI core collection reveals

substantial diversity and unusual genotype dispersal throughout the Mediterranean basin. Theor Appl Genet 112: 977–983.

Green, J. M. 2009. Evolution of Glyphosate-Resistant Crop Technology. *Weed Sci.* 57:108-117

Hager, A. G. and C. L. Sprague. 2002. Weeds on the horizon. University of Illinois–Urbana. *The Pest Management and Crop Development Bull* 2:6.

Hager, A. G., L. M. Wax, G. A. Bollero and F. W. Simmons. 2002. Common waterhemp (*Amaranthus rudis* Sauer) management with soil-applied herbicides in soybean [*Glycine max* (L.) Merr.]. *Crop Prot.* 21:277–283.

Hartzler, B. 1998. Herbicide Site of Action. Iowa State University. http://www.weeds.iastate.edu/reference/siteofaction.htm.

Hartzler, B. Dec. 18, 2003. Waterhemp - the perfect weed? Iowa State University. http://www.weeds.iastate.edu/mgmt/2003/symposium.shtml

He, G., R. Meng, M. Newman, G. Gao, R. N. Pittman, and C. S. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 3: 3.

Heap, I. The International Survey of Herbicide Resistant Weeds. Online. Internet. April 18, 2010. Available at www.weedscience.org.

Horak, M. J. and T. M. Loughin. 2000. Growth analysis of four Amaranthus species. *Weed Sci.* 48:347–355.

Horak, M. J. and D. E. Peterson. 1995. Biotypes of Palmer amaranth (*Amaranthus palmeri*) and common waterhemp (*Amaranthus rudis*) are resistant to imazethapyr and thifensulfuron. *Weed Technol*. 9:192–195.

Jacobs, J. M. and N. J. Jacobs. 1984. Protoporphyrinogen oxidation, an enzymatic step in heme and chlorophyll synthesis: Partial characterization of the reaction in plant organelles and comparison with mammalian and bacterial systems. *Arch. Biochem. Biophys.* 229, 312–319.

Lee, H. J., and S. O. Duke. 1994. Protoporphyrinogen IX-oxidizing activities involved in the mode of action of peroxidizing Herbicides. *J. Agric. Food Chem.* 42, 2610 – 2618.

Lee, R. M., A. G. Hager, and P. J. Tranel. 2008. Prevalence of a novel resistance mechanism to PPO-inhibiting herbicides in waterhemp (Amaranthus tuberculatus). *Weed Sci.* 56:371-375.

Levy, S. et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol.* 5, e254.

Li, J., R. J. Smeda, K. A. Nelson, and F. E. Dayan. 2004. Physiological basis for resistance to diphenyl ether herbicides in common waterhemp (*Amaranthus rudis*). *Weed Sci* 52: 333-338 (2004).

Li, X., S. L. Volrath, D. B. G. Nicholl, C. E. Chilcott, M. A. Johnson, E. R. Ward, and M. D. Law. 2003. Development of protoporphyrinogen oxidase as an efficient selection marker for *Agrobacterium tumefaciens*-mediated transformation of maize. *Plant Physiol.* 133:736–747.

Mallory, M. A., R. V. Hall, A. R. McNabb, D. B. Pratt, E. N. Jellen, and P. J. Maughan. 2008. Development and characterization of microsatellite markers for the grain amaranths. *Crop Sci.* 48:1098–1106.

Margulies et al. 2005. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature* 437(7057):376-380.

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.

Patzoldt, W. L., A. G. Hager, J. S. McCormick, and P. J. Tranel. 2006. A codon deletion confers resistance to herbicides inhibiting protoporphyrinogen oxidase. *Proc. Natl. Acad. Sci. U.S.A* 103:12,329–12,334.

Patzoldt, W. L., P. J. Tranel, and A. G. Hager. 2005. A waterhemp (*Amaranthus tuberculatus*) biotype with multiple resistance across three herbicide sites of action. *Weed Sci.* 53:30-36.

Rothberg, J. M. and J. H. Leamon. 2008. The development and impact of 454 sequencing. *Nature Biotechnology* 26:1117-1124.

Rousonelos, S. L., R. M. Lee, and P. J. Tranel. 2009. An Alternative to the Glycine Deletion: Why R98L was Selected in Common Ragweed Protoporphyrinogen Oxidase. Proc. North Cent. Weed Sci. Soc. 64:98.

Sanger F., S. Nicklen, and A. Coulson. 1997. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74:5463–5467.

Sikorski J. A. and K.J. Gruys. 1997. Understanding glyphosate's molecular mode of action with EPSP synthase: evidence favoring an allosteric inhibitor model. *Acc Chem Res* 30: 2-8

Wheeler, D. A. et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876.

Wood, A. 2010. Compendium of Pesticide Common Names. http://www.alanwood.net/pesticides/index.html

CHAPTER 2

INVESTIGATION INTO THE PROTOPORPHYRINOGEN OXIDASE GLYCINE DELETION AS WATERHEMP'S RESISTANCE MECHANISM TO PPO-INHIBITING HERBICIDES

2.1 Abstract

Resistance in waterhemp to herbicides that inhibit protoporphyrinogen oxidase (PPO) previously was shown to result from the deletion of a glycine codon at position 210 (ΔG210) in the PPO-encoding gene, *PPX2*. Research was conducted to determine if this same mechanism accounted for resistance in geographically separated populations – from Illinois, Kansas, and Missouri – and, if so, to determine if the resistance mutation was independently selected. A dose-response study with lactofen indicated that the resistant populations had different levels of resistance. These differences, however, could be accounted for by different frequencies of resistant individuals within populations and, therefore, the dose-response data were consistent with the hypothesis that the populations contained the same resistance mechanism. Direct evidence in support of this hypothesis was provided by DNA sequencing, which showed that nearly all resistant plants evaluated contained the ΔG210 mutation. A variable region of the *PPX2* gene was sequenced and resulting sequences were aligned and organized into a phylogenetic tree. In general, *PPX2* sequences clustered by geographic origin, suggesting independent selection of the resistance mutation. However, a single selection event, followed by gene flow and extensive recombination within the *PPX2* gene cannot be ruled out.

2.2 Introduction

Waterhemp, indigenous to Illinois, has caused problems for farmers in soybean and corn fields for years (Hager et al. 2002), specifically since the 1990s with a change in farming practices that favored waterhemp germination and growth. In particular, waterhemp's small seeds make this species well adapted to no-till farming systems (Buhler 1992). Waterhemp may produce over 300,000 small seeds per plant (Hartzler 2003). These seeds persist in soil seed banks and have a high level of dormancy (Buhler and Hartzler 2001). Furthermore, waterhemp is a dioecious plant, which ensures outcrossing, allowing herbicide resistance to spread rapidly (Foes et al. 1998). The long-distance flow of pollen may have allowed herbicide resistance in waterhemp to spread long distances. The aforementioned factors allow waterhemp to quickly adapt to its environment. With repeated use of single herbicide modes of action, waterhemp was quick to evolve herbicide resistance (Patzoldt et al. 2005).

PPO-inhibiting herbicides have been used to control weeds for the past 30 years, however, resistance was not identified until 2001 (Li et al. 2004) when waterhemp from Kansas evolved resistance to PPO-inhibiting herbicides. Shortly after, waterhemp in Illinois and Missouri were also confirmed as resistant to these herbicides (www.weedscience.com). The resistance mechanism in waterhemp from Illinois was previously identified as a single codon deletion in the *PPX2* gene, resulting in the loss of a glycine, ΔG210 (Patzoldt et al. 2006). The resistance mechanism is unusual, in that it involves an amino acid deletion, rather than a substitution. Despite being an unusual mutation, it was previously found to account for resistance in all six out of six waterhemp populations tested in Illinois (Lee et al. 2008).

The objective of this study was 1) to expand on the Lee et al. study to determine if the ΔG210 mutation also accounted for resistance in populations outside of Illinois and 2) to determine if the allele conferring resistance to the herbicide in waterhemp was independently selected in multiple populations.

2.3 Materials and Methods

Waterhemp populations used included one resistant population each from Illinois (Adams County Resistant, or ACR) and Kansas (KS_R), three resistant populations from Missouri (Trenton, Meadville, and MO1), and one sensitive population from Missouri (Bradford). In all greenhouse trials, 0.25 g of seed from each respective waterhemp population was broadcast into 25- by 50-cm plastic greenhouse flats containing a 3:1 mixture of commercial potting medium to sand. This same mixture was used to cover the seedbed at a thickness of approximately 6 mm. After emergence, waterhemp seedlings at the one to two leaf stage were transplanted into 15-cm diameter pots containing the same 3:1 mixture of commercial potting medium to sand. Each pot contained one waterhemp plant and all plants were maintained in a greenhouse at 25 to 30 C, watered and fertilized as needed, and provided with artificial lighting from metal halide lamps (600 μmol photon m^{-2} s^{-1}) simulating a 16-h photoperiod day. All waterhemp plants were sprayed with herbicides when average plant height reached 10 cm. Herbicide treatments were applied with a compressed air laboratory spray chamber equipped with an even flat-fan spray nozzle delivering 220 L/ha at 234 kPa. Rates of lactofen used on all plants for the dose-response were 0, 10.0, 31.6, 100, 316, 1000, and 3160 g ai ha^{-1} plus .25% COC.

Two additional rates of lactofen, 0.1, 0.32 g ai ha^{-1}, were used on plants from the Bradford population and two additional rates were used on plants from the 5 resistant populations, 1000, and 3160 g ai ha^{-1}. All spray mixtures included .25% COC. At two weeks after treatment (WAT), aboveground biomass of each waterhemp plant was harvested, dried at 70 C for 5 d, and weighed. Dry weights were converted to a percentage of the untreated control using the untreated plants from each respective waterhemp population. All greenhouse experiments were arranged in a completely randomized design with six replications and each experiment was conducted twice.

Percent dry weights were calculated to account for any difference in plant growth between populations (% dry weight= dry weight/average dry weight of the control x100). R software version 2.10.1[1] with the dose response curve (drc) package was used to analyze the dose-response data (Ritz and Streibig 2005). The 4 parameter log-logistic function used to generate dose-response curves is expressed as:

$$f(x) = c + \frac{d-c}{1+\exp\left(b\left(\log(x)-\log(e)\right)\right)} \quad \text{[Equation 1]}$$

where e=dose resulting in a response halfway between D and C (ED_{50}), b=curvature parameter, c=lower asymptote, and d=upper asymptote (Seber 1989). To determine if fitted curves were different from one another, a comparison of parameters was used for analysis. ED_{50} values were obtained from the fit of the log-logistic model and the difference between the ED_{50} value and the standard error of each population was calculated. These differences were compared in a multiple comparison to determine whether the curves were different from each other.

In order to determine the frequency of resistant plants in each population tested in the dose-response, approximately 100 waterhemp seeds from 7 populations (the same resistant populations used in the dose-response, plus additional sensitive controls from Kansas and Illinois) were planted in flats containing 1:1:1 mixture of soil:peat:torpedo sand. At the two-leaf seedling stage, plants were transplanted into pots with a 12-cm diameter and grown until average plant height reached 10 to 12 cm. At this time, DNA was extracted from new leaf tissue using the hexadecyltrimethyl-ammonium bromide (CTAB) method (Doyle and Doyle, 1990). Treated plants were sprayed with lactofen at 110 g ai ha^{-1} plus 1% crop oil concentrate and controls were sprayed with 1% COC. Lactofen at this dose has been shown to discriminate between resistant and susceptible plants (Patzoldt et al. 2006). Herbicide treatments were applied using a compressed-air, moving-nozzle laboratory sprayer equipped with an 80° flat fan nozzle[2] (Teejet; Spraying Systems Co., Wheaton, IL) delivering 187 liters·ha^{-1} of water at 207 kPa. The nozzle was maintained ≈45 cm above the plant canopy. Herbicide resistance was evaluated 2 weeks after treatment (WAT) and plants were scored as either dead or alive. The experiment was arranged in a completely randomized design and conducted twice. Percentage of resistant plants was calculated.

To examine the resistance mechanism of waterhemp populations outside of Illinois that are resistant to PPO-inhibiting herbicides, waterhemp DNA extracted before the single-rate screen was used. A subset of the resistant and sensitive plants were randomly chosen. To assess the presence or absence of ΔG210, primers amplifying the region surrounding the deletion were used to amplify and sequence the selected samples. Primers AmPPX2LwalkDR, 5'- acagcctccagaaaatgttg-3', and AmPPX2LwalkDF, 5'-

gagaaaacacaatgctactgaa-3', were used. Amplification reactions totaled 20 µl and contained 100 ng DNA, 0.3 µl Taq polymerase[3], 1X reaction buffer, 25 mM $MgCl_2$, 0.2 mM deoxyribonucleotide triphosphates, and 300nM of the forward and reverse primers. PCR protocol was as follows: incubation for 2 min at 95 C, followed by 38 cycles of 0.5 min at 95 C, 0.5 min at 50 C, and 0.75 min at 72C. A portion of the PCR product (5 µl) was then fractionated on a 1.5% agarose[4] gel plus 1% ethidium bromide at 80 volts. PCR products were visualized using ultraviolet light to confirm that amplification of the correct band occurred. The remaining 15 µl of product was purified using QIAquick PCR Purification Kit[5]. The purified product was sequenced using AB BigDye Terminator v3.1 Cycle Sequencing Kit[6] using the forward primer, AmPPX2LwalkDF. The sequenced products were then analyzed by the W.M. Keck Center for Comparative and Functional Genomics High Throughput Sequencing Lab[7] using an ABI 3730xl Capillary DNA Analyzer[8]. The resulting sequence data were analyzed using Sequencher 4.7 software[9]. Sequences were aligned using Multalin (Corpet 1988) and evaluated for the presence or absence of the ΔG210 codon. In order to determine the relationship between biotype (resistant or sensitive to PPO-inhibitors) and deletion frequency, this data was organized into a contingency table for Pearson X^2 analysis. X^2 values were calculated using R based on the equation:

$$X^2 = \sum \frac{(o-e)^2}{e} \quad \text{[Equation 2]}$$

where o=observed data and e=expected data. Under the null hypothesis the distribution of the observed data is equal to the distribution of the expected data.

The sequence alignment of the region surrounding ΔG210 was also analyzed visually to determine number of polymorphic sites. A high number of polymorphic sites would make this region a good candidate for a phylogenetic tree analysis.

A nested PCR procedure was developed to selectively amplify exons two through four of the resistant *PPX2* allele. This region was chosen for its variability among plants. To selectively amplify the resistant allele, the first PCR procedure included a reverse primer, specific for ΔG210. The primers used were ARMS9Rb-R, 5'-tctgcacatgtacccgcaac-3', and *PPX2L*-F2, 5'-aattcaatccattacccacc-3'. Amplification reactions totaled 20 µl and contained 100 ng DNA, 0.3 µl Taq polymerase, 1X reaction buffer, 25 mM $MgCl_2$, 0.2 mM deoxyribonucleotide triphosphates, and 300nM of the forward and reverse primers. PCR protocol was as follows: incubation for 2 min at 95 C, followed by 38 cycles of 0.5 min at 95 C, 0.5 min at 56 C, and 0.75 min at 72 C. A portion of the PCR product (5 µl) was then fractionated and visualized on an agarose gel, as described above. The remaining 15 µl of product was purified using QIAquick PCR purification kit. The purified products were then subjected to a second PCR reaction, using primers PPX2Le2-F1, 5'-cctctgctaaaagggttgctg-3', and PPX2L-R1, 5'-tcaaactcgagacctctgcc-3'. PCR reaction and protocol was the same as previously listed. Known sensitive plants were only subjected to the second PCR of the nested PCR procedure. PCR products were fractionated on agarose gel, purified and sequenced, as described previously. Sequencing was done using the forward primer (PPX2Le2-F1) for all reactions and the reverse primer (PPX2L-R1) for reactions in which the forward primer did not provide clean sequence data. The sequenced products were then analyzed as previously described. To assess the reliability of the nested PCR procedure, a PCR was

performed that contained DNA from both a known resistant plant and a known sensitive plant. These same resistant and sensitive plants were also PCR-amplified separately in two different reactions. The three PCR products were sequenced and the resulting sequence alignments were compared to examine whether the nested PCR selectively amplified the allele from the resistant plant.

Aligned sequences were further analyzed using Molecular Evolutionary Genetics Analysis (MEGA v4.1) software (Tamura et al. 2007). The Neighbor-Joining method was used to infer the evolutionary history, since the sequence data contained both protein-coding data (exons) and introns (Tamura et al. 2004; Saitou and Nei 1987). The Maximum Composite Likelihood method was used to determine evolutionary distances and a pairwise deletion sequence comparison was done on the 1st, 2nd, 3rd, and noncoding positions. Confidence in the Neighbor-Joining tree method was estimated by 1000 bootstrap replicates.

The sequence data were further analyzed using Tajima's D (Tajima 1989), which tests the neutral theory of molecular evolution, which assumes constant population size and no selection (Kimura 1983). Tajima's D can be expressed as:

$$D = \frac{\Pi - \Theta}{\sqrt{V}} \quad \text{[Equation 3]}$$

where Π= average number of pairwise differences; Θ=total number of segregating sites; and V=sampling variance of the difference between Π and Θ. Under the null hypothesis, Π and Θ should be equal in value, meaning that the evolution of mutations is neutral in a population in equilibrium.

DNAsp v5 (Librado and Rozas 2009) was used to calculate Tajima's D based on the neutral model, which assumes that the number of segregating sites and the average number of nucleotide differences are correlated (Tajima 1989). Nucleotide diversity, or Π, was also calculated by DNAsp v5. Nucleotide diversity is an estimate of the average number of nucleotide differences per site between two randomly chosen sequences and can be expressed as:

$$\Pi = \sum_{ij} x_i x_j \Pi_{ij} \quad \text{[Equation 4]}$$

where Π_{ij}=proportion of different nucleotides between the ith and j^{th} DNA sequences; and χ_i and χ_j are, respectively, the frequencies of the ith and j^{th} alleles. Each sequence is randomly paired with every other sequence, without repetition and the sum is calculated (Nei 1987).

Sequences were analyzed for traces of recombination using DNAsp v5 (Librado and Rozas 2009) which computes the recombination parameter (R) as:

$$R=4Nr \quad \text{[Equation 5]}$$

where N=population size and r=recombination rate between adjacent sites (Hudson et al. 1987). The R parameter was used to estimate the minimum number of recombination events (R_M) in the history of the samples (Hudson and Kaplan 1985).

In order to determine the relationship between geographic distance and genetic distance, pairwise distances were calculated in (MEGA v4.1) software (Tamura et al. 2007) using the sequence data from exons 2 through 4. General locations of the 7 populations were used to determine the latitude and longitude of each source location.

Latitude and longitude values were converted to pairwise geographic distances using spatstat package in R (Baddeley and Turner 2005). The relationship between population location and genetic distance was examined by plotting genetic distance vs. geographic distance using a mixed model and looking at changes in slope and intercept. To determine if mean genetic distance varied for resistant vs. sensitive plants, a categorical comparison was done.

2.4 Results and Discussion

The fitted dose-response curves showed that 4 of the resistant populations (Trenton, Meadville, MO1, KS, and IL) have differing dose-response curves when compared to Bradford, the sensitive population from Missouri (Figure 2.1). The MO1 population shows a lower level of resistance than the other 4 resistant populations. A multiple comparison based on ED_{50} levels showed that IL, KS, Trenton, and Meadville had similar curves, 52.79 (SE=56.48), 20.21 (SE=26.03), 96.92 (SE=103.75), and 32.73 (SE=50.14), respectively. On the other hand, the MO1 and Bradford dose-response curves each were not similar to any other population, with ED_{50} values of 1.93 (SE=1.06) and 0.46 (SE=0.28), respectively (Figure 2.2). Different levels of resistance among the 5 resistant populations could be explained by one of two hypotheses: the different populations comprise either different biotypes (i.e. different mechanisms of resistance) or they vary in the frequencies of a common resistant biotype.

To determine if the differences in the dose-response curves were due to different frequencies of resistant plants, the percentage of resistant plants was calculated for each

resistant population, based on the single-rate screen with lactofen that was used for sequence comparisons. The frequency of resistant plants in the MO1 population was much lower than the frequencies of resistant plants in the other 4 resistant populations. In fact, the percentage of resistant plants in MO1 was only 11.3%, compared to 65.7% for Meadville, 82.4% for KS_R, 82.9% for Trenton, and 100% for IL_R (Table 2.1). MO1's decreased frequency of resistant plants, rather than having a different mechanism of resistance, may account for its differing dose-response curve. These results favor the hypothesis that different frequencies of resistant plants are associated with the varying dose-response curves, suggesting that all resistant plants may share the same resistance mechanism.

The sequence analysis of the ΔG210 site from both resistant and sensitive plants was examined to determine if ΔG210 is correlated to PPO-inhibiting herbicide resistance. Based on sequence comparisons, ΔG210 is strongly associated with resistance in plants from all 3 states (Figure 2.3). Of the resistant plants screened, 4 out of 4 from IL_R, 10 out of 10 from KS_R, 4 out of 5 from MO1, and 8 out of 8 from each Trenton and Meadville showed to have the 3 base pair deletion at ΔG210. Moreover, of the sensitive plants tested, 3 out of 3 from IL_S, 3 out of 4 from KS_R, 2 out of 2 from KS_S, 11 out of 12 from MO1, 4 out of 4 from Trenton, and 5 out of 6 from Meadville did not have the 3 base pair deletion at ΔG210. Based on the contingency table analysis, Pearson X^2 test with Yates' continuity correction was equal to 47.567 with 1 degree of freedom and a P-value of 5.316e-12 (Table 2.2). The X^2 value greatly exceeds the X^2 critical value (1 degree of freedom, α=0.05) of 3.841. Therefore, it can be concluded that the biotype (whether the plants were resistant or sensitive to lactofen) significantly impacts the presence of the

deletion, regardless of the source population. This shows that even though ΔG210 is a rare resistance mechanism, it is the mechanism for resistance to PPO-inhibiting herbicides that was selected in the Midwest waterhemp populations and it is waterhemp's only resistance mechanism to PPO-inhibiting herbicides that has been identified thus far. The few discrepancies in data can most likely be attributed to technical errors. Resistant plants may have appeared sensitive if they died for reasons not related to herbicide damage, due to factors such as water stress or limited growing space. Errors could have also been due to DNA contamination or sequence errors.

The sequence analysis of the region surrounding the ΔG210 site did not prove to be useful for a phylogenetic study. Each sequence was approximately 750 base pairs in length, however polymorphic sites (besides at ΔG210) were not abundant. In fact, of the few polymorphisms that were identified, it was unsure if these polymorphisms were truly polymorphisms or if they were simply a result of ambiguous chromatogram data. The fact that the sequence surrounding ΔG210 is highly conserved may suggest that this is a critical region for the function of the gene. Since a highly polymorphic region is best to infer a phylogenetic tree, the region surrounding ΔG210 would not be ideal for analysis. For this reason, a more highly polymorphic region, exons 2 through 4 of the *PPX2* allele, was chosen for phylogenetic analysis.

Dioecious plants can have a high level of heterozygosity. To combat this problem, a nested PCR procedure was developed to selectively amplify exons 2 through 4 of the resistant *PPX2* allele (Figure 2.4). Due to the sequence of the sensitive allele at the ΔG210 site, designing an allele specific primer to selectively amplify the sensitive allele was not possible. Therefore, heterozygotic sites in sensitive alleles and in resistant plants

that were heterozygous for the resistant allele were scored using the IUPAC ambiguity code (Tipton 1994). Sequence analysis of the three nested PCR products, the sensitive plant alone, the resistant plant alone, and the resistant and sensitive plant together, revealed that the nested PCR only amplified the allele from the known resistant plant (Figure 2.5). This affirms that the nested PCR procedure will only amplify the resistant allele in heterozygous plants.

PPX2 sequences were aligned and organized into a phylogenetic tree. The tree was constructed from a 418 base pair sequence with 27 polymorphisms amongst 24 taxa. The sum of branch lengths (SBL) was calculated to be 0.06877850 under the Neighbor-Joining method. This is a small value, which is ideal because the smallest SBL values indicate the best tree. While the phylogenetic tree does not show separation based solely on geographic location, some population structure based on location can be seen as plants from Kansas and Missouri segregated together regardless of whether they were resistant or sensitive to lactofen (Figure 2.6). Had the tree separated the plants based on whether they were resistant or sensitive, it would have been concluded that the deletion conferring resistance to PPO-inhibiting herbicides occurred one time in one area and spread to all the other areas. However, the tree shows a small amount of separation based on geography, and not between resistant and susceptible plants, suggesting that PPO herbicide resistance evolved independently, or separately, in each geographic area. The weeds may have been able to separately evolve resistance due to the extreme selection pressure against them that is caused by repeated herbicide applications. With the selection pressure on weeds, they evolve much more quickly than other plants (Gressel 2009). Therefore, herbicide resistant plants will evolve and be selected for rather quickly.

However, since multiple alleles were seen in a single population and the structure of the tree is not solely based on geographic location, another possibility is that waterhemp's obligate outcrossing allows for a high amount of recombination. For example, seed from a resistant waterhemp plant from Illinois could be transferred to another state, for example Missouri, and the resistant plant from Illinois could pollinate a susceptible waterhemp plant from Missouri. Normally, each progeny would receive one *PPX2* allele from each parent, therefore receiving one sensitive and one resistant allele; however, if recombination occurred somewhere along the *PPX2* gene, a resistant allele from Illinois (with the ΔG210 resistance mechanism) could partially resemble the sensitive allele from Missouri (Figure 2.7). Therefore, a high amount of recombination would make tracking alleles difficult.

Aligned sequences were analyzed using Tajima's D. The D statistic was calculated to equal -0.46064, which is slightly skewed towards negative values. A negative D statistic suggests either population expansion or purifying selection. A positive D statistic would have suggested a recent population bottleneck or recent balancing selection. However, with a P-value greater than 0.10, the statistic is not significant and this means that no conclusions regarding population expansion can be drawn from these results. DNAsp calculated Π_S to equal 0.01473 and Θ_S to equal 0.01708. Π_S is a measure of the average number of nucleotide differences per site between any two randomly chosen DNA sequences. The Π_S value for exons 2 through 4 of *PPX2* is higher than that of wild maize (*Zea maize*) relatives (0.012-0.013; Tiffin and Gaut 2001), *Arabidopsis thaliana* (0.0083; Schmid et al. 2005), and soybean (*Glycine max*) (0.0015; Zhu et al. 2003). Comparing the Π_S value of exons 2 through 4 to the Π_S

values of other plants suggests that this region of the *PPX2* allele has a high level of nucleotide diversity.

Further DNAsp analysis was done to determine if any recombination could have occurred throughout exons 2 through 4 on *PPX2*. Recombination analysis this region showed that the estimated minimum number of recombination events (R_M) was equal to 3, meaning that there are an estimated 3 sites between exons 2 and 4 where recombination may have occurred. This estimated recombination may explain the high level of diversity within populations and could also account for why the phylogenetic tree analysis was unable to give a tree organized solely based on either geography (if independent evolution was responsible for the spread of the resistant allele) or resistant vs. sensitive (if gene flow was responsible for the spread of the resistant allele).

The plot of genetic distance vs. geographic distance revealed a slightly negative relationship, but with a p-value greater than 0.05, the relationship was not significant. Therefore, we can conclude that that among the 7 populations examined in this study, geographic distance did influence genetic distance; these 7 populations are not genetically isolated from one another. A categorical comparison for resistant and sensitive plants revealed that mean genetic distance did not vary. Had resistant plants shown a smaller genetic distance, it could have been concluded that the resistance mutation event was one event that spread quickly throughout the Midwest. However, resistant and sensitive plants had similar mean genetic distances (p-value=0.15), which suggests that the mutation event was not a single event.

In conclusion, although the data adds evidence that the glycine deletion is the predominant, if not the only, mechanism of resistance, further genetic analysis failed to determine whether the spread of resistance was due to independent evolution or gene flow. The dioecious nature of this species allows for genetic mixing of waterhemp in the Midwest, making it difficult to evaluate evolutionary dynamics of this species.

2.5 Sources of Materials

[1] R software version 2.10.1, The R Project For Statistical Computing.

[2] Compressed-air, moving-nozzle laboratory sprayer equipped with an 80° flat fan nozzle, Teejet, Spraying Systems Co., P.O. Box 7900 Wheaton, IL 60187

[3] Taq polymerase, New England Biolabs, 240 County Road, Ipswich, MA 01938.

[4] Agarose, Bio-Rad Laboratories, 1000 Alfred Nobel Drive, Hercules, CA 94547.

[5] QUIAquick PCR Purification Kit, 27220 Turnberry Lane Suite 200, Valencia, California 91355.

[6] AB Big Dye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems Inc, 850 Lincoln Centre Drive, Foster City, CA 94404.

[7] W.M. Keck Center for Comparative and Functional Genomics, University of Illinois Biotechnology Center, 340 Edward R. Madigan Laboratory, 1201 W. Gregory Drive, Urbana, IL 61801.

[8] ABI 3730xl Capillary DNA Analyzer, Applied Biosystems Inc, 850 Lincoln Centre Drive, Foster City, CA 94404.

[9] Sequencher™ 4.7 software, Gene Codes Corporation, 775 Technology Drive Suite 100A Ann Arbor, MI 48108 USA.

2.6 Literature Cited

A. Baddeley and R. Turner. 2005. Spatstat: an *R* package for analyzing spatial point patterns. *J. of Statist. Software* 12:1-42.

Buhler, D. D.1992. Population dynamics and control of annual weeds in corn (*Zea mays*) as influenced by tillage systems. *Weed Sci*. 40:241- 248.

Buhler, D. D. and R. G. Hartzler. 2001. Emergence and persistence of seed of velvetleaf, common waterhemp, wooly cupgrass, and giant foxtail. *Weed Sci.* 49:230-235.

Corpet, F. 1998. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res*.,16(22):0881-10890.

Doyle, J. J. and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13–15.

Foes, M. J., L. Liu, P. J. Tranel, L. M. Wax, and E. W. Stoller. 1998. A biotype of common waterhemp (*Amaranthus rudis*) resistant to triazine and ALS herbicides. *Weed Sci.* 46:514–520.

Gressel, J. 2009. Evolving understanding of the evolution of herbicide resistance. *Pest Management Sci.* 65(11):1164-73.

Hager, A. G., L. M. Wax, G. A. Bollero and F. W. Simmons. 2002. Common waterhemp (*Amaranthus rudis* Sauer) management with soil-applied herbicides in soybean [*Glycine max* (L.) Merr.]. *Crop Prot.* 21:277–283.

Hartzler, Robert G. 2003. Waterhemp-the perfect weed? www.weeds.iastate.edu/mgmt/2003/symposium.shtml. WSSA-North American Weed Management Association Invasive Plant Species Workshop, Kansas City, MO. Feb. 12-13.

Hudson, R. R. and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147-164.

Hudson, R. R., Kreitman, M., and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153-159.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, MA.

Lee, R. M., A. G. Hager, and P. J. Tranel. 2008. Prevalence of a novel resistance mechanism to PPO-inhibiting herbicides in waterhemp (Amaranthus tuberculatus). *Weed Sci.* 56:371-375.

Li, J., R. J. Smeda, K. A. Nelson, and F. E. Dayan. 2004. Physiological basis for resistance to diphenyl ether herbicides in common waterhemp (*Amaranthus rudis*). *Weed Sci.* 52:333-338.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Nei, M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York, NY.

Patzoldt, W. L., P. J. Tranel, and A. G. Hager. 2005. A waterhemp (*Amaranthus tuberculatus*) biotype with multiple resistance across three herbicide sites of action. *Weed Sci.* 53:30–36.

Patzoldt, W. L., A. G. Hager, J. S. McCormick, and P.J. Tranel. 2006. A codon deletion confers resistance to herbicides inhibiting protoporphyrinogen oxidase. *Proc. Natl. Acad. Sci. USA* 103:12329–12334.

Ritz, C. and J. C. Streibig. 2005. *Bioassay analyses using R.* J. Stat. Softw. 12:1–22.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. and Evol.* 4:406-425.

Schmid, K. J., S. Ramos-Onsins, H. Ringys-Bechstein, B. Weisshaar, and T. Mitchell-Olds. 2005. A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* 169:1601-1615.

Seber, G. A. F. and Wild, C. J. 1989. *Nonlinear Regression.* New York: Wiley and Sons.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.

Tamura K., M. Nei, and S. Kumar. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA* 101:11030-11035.

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

Tiffin, P. and B.S. Gaut. 2001. Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: Insights from four nuclear loci. *Genetics* 158:401-412.

Tipton, K. F. 1994. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme nomenclature. Recommendations 1992. Supplement: corrections and additions. *Eur. J. Biochem.* 223(1):1-5.

Zhu, Y. L., et al. 2003. Single-nucleotide polymorphisms in soybean. *Genetics* 163:1123-1134.

2.7 Figures and Tables

Figure 2.1. Dose responses of five resistant waterhemp populations from Illinois (IL), Kansas (KS), and Missouri (MO1, Trenton, Meadville) and a sensitive population from Missouri (Bradford) to lactofen. Dose-response curves were fitted using R software and a four parameter, log-logistic function.

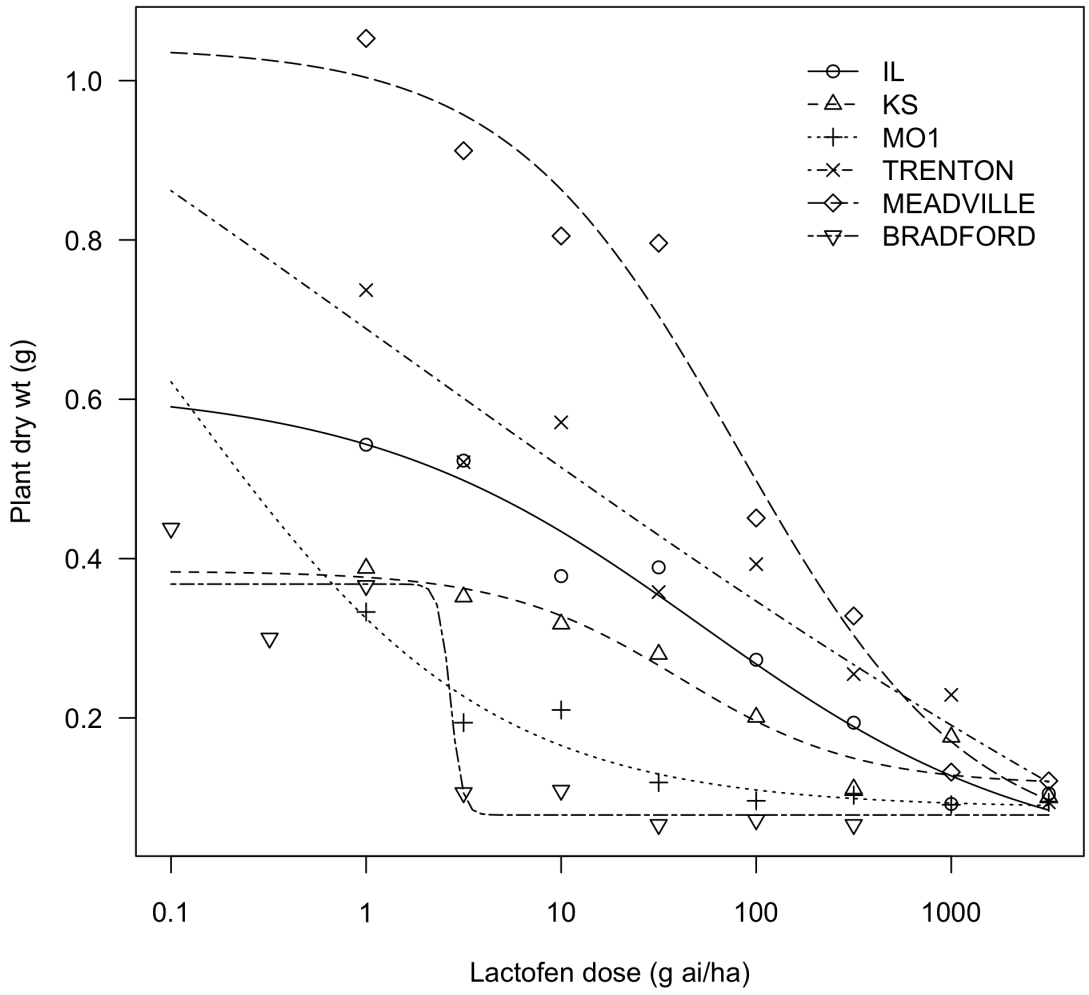


Figure 2.2. ED_{50} values were calculated based on the dose-response data of five resistant waterhemp populations from Illinois (IL), Kansas (KS), and Missouri (MO1, Trenton, Meadville) and a sensitive population from Missouri (Bradford) to lactofen. A multiple comparison was done based on standard errors of the ED_{50} values. The multiple comparison test showed that IL, KS, Trenton, and Meadville had similar curves, 52.79 (SE=56.48), 20.21 (SE=26.03), 96.92 (SE=103.75), and 32.73 (SE=50.14), respectively. On the other hand, the MO1 and Bradford dose-response curves each were not similar to any other population, with ED_{50} values of 1.93 (SE=1.06) and 0.46 (SE=0.28), respectively.

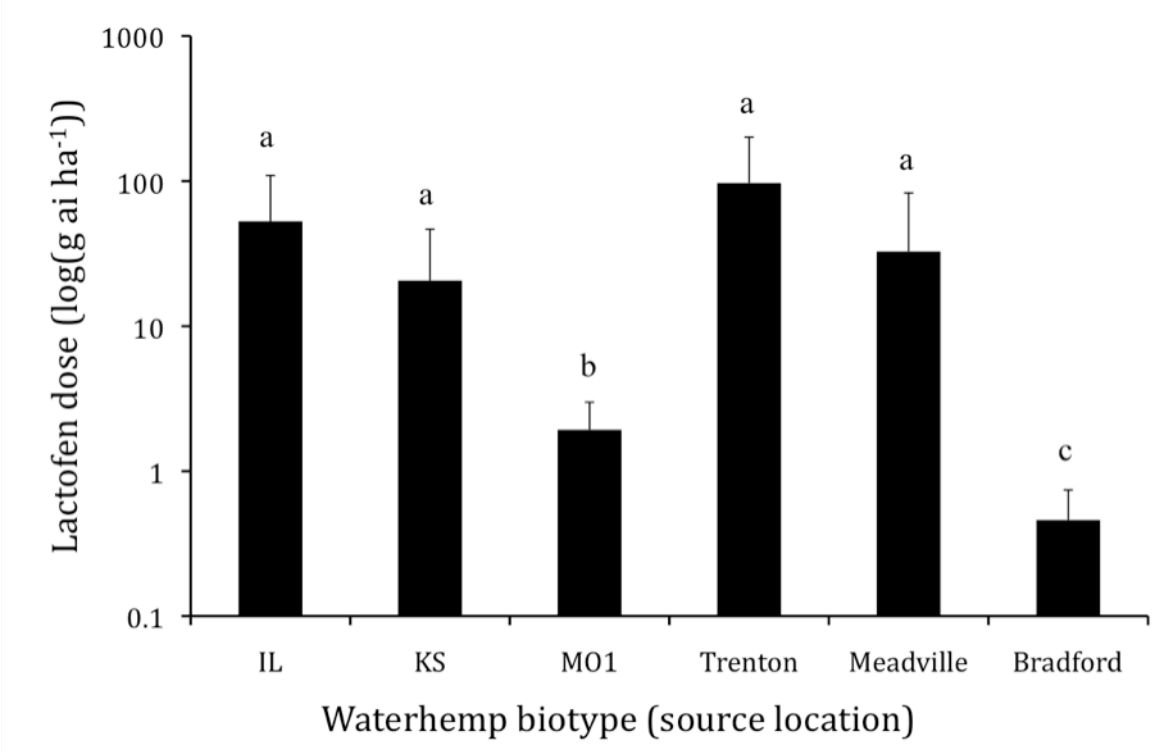


Figure 2.3. Sequence alignment of the ΔG210 deletion. Plants with the deletion are deemed resistant to PPO-inhibiting herbicides, while plants without the deletion are sensitive to PPO-inhibiting herbicides.

```
 IL-R TGTGG---AGA
 IL-R TGTGG---AGA
 IL-S TGTGGTGGAGA
 IL-S TGTGGTGGAGA
MOM-R TGTGG---AGA
MOM-R TGTGG---AGA
MOM-R TGTGG---AGA
MOM-S TGTGGTGGAGA
MOT-R TGTGG---AGA
MOT-R TGTGG---AGA
MOT-S TGTGGTGGAGA
 MO-S TGTGGTGGAGA
 MO-S TGTGGTGGAGA
 MO-S TGTGGTGGAGA
 MO-S TGTGGTGGAGA
 MO-S TGTGGTGGAGA
 MO-R TGTGG---AGA
KSS-S TGTGGTGGAGA
KSS-S TGTGGTGGAGA
KSR-R TGTGG---AGA
KSR-R TGTGG---AGA
KSR-R TGTGG---AGA
KSR-S TGTGGTGGAGA
```

Figure 2.4. Due to a high amount of heterozygosity at the *PPX2* allele, a nested PCR procedure was developed to selectively amplify exons 2 through 4 of the resistant *PPX2* allele. A first PCR uses a reverse primer that specifically anneals to the resistant allele at the ΔG210 site and a forward primer that anneals to exon 1 of *PPX2*. The resulting PCR product is then subjected to a second PCR, with a forward primer in exon 2 and a reverse primer in exon 4 of *PPX2*. The resulting PCR product should contain only exons 2-4 of the resistant *PPX2* allele.

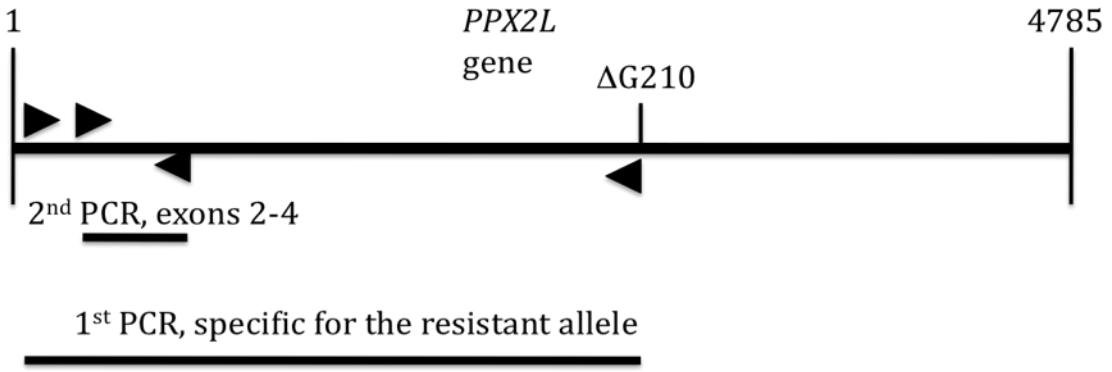


Figure 2.5. To eliminate heterozygosity in plants with both a resistant and sensitive PPX2 allele, a nested PCR was developed to selectively amplify exons two through four of the resistant allele. To determine whether the PCR selectively amplified only the resistant allele, DNA from a known resistant place (ACR) and a known sensitive plant (WCS), both previously determined to lack heterozygous base pairs, were together amplified in a single nested PCR reaction. Sequence data from the ACR+WCS amplification was aligned with sequence data from separate ACR and WCS amplifications. Alignment showed that the nested PCR selectively amplified the *PPX2* allele from ACR, meaning that only the resistant allele was amplified.

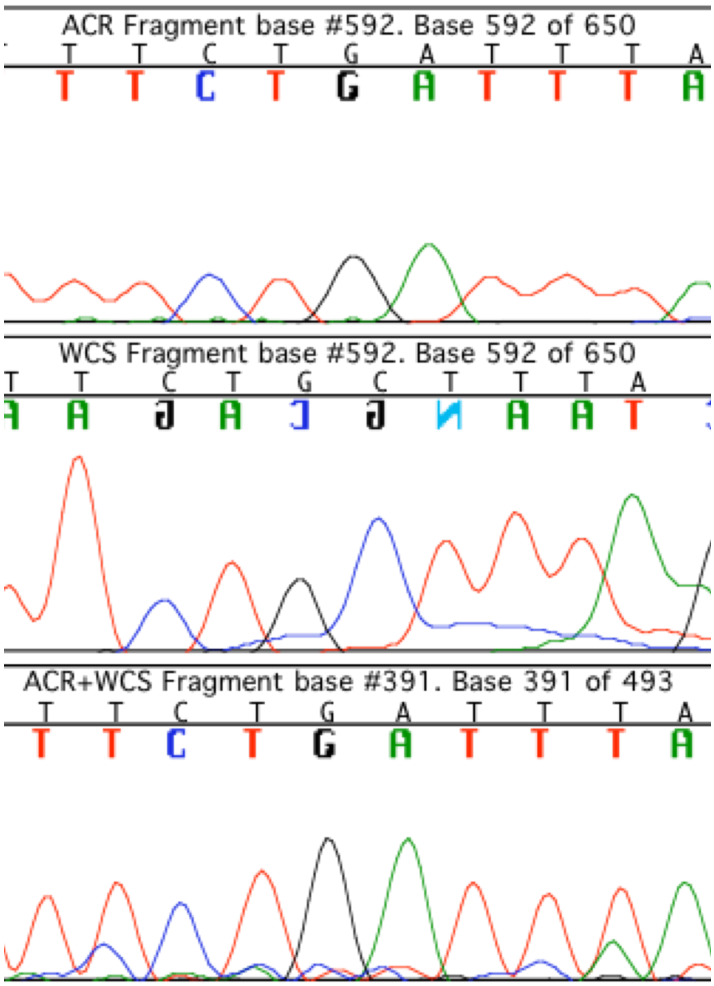


Figure 2.6. Using MEGA, a neighbor-joining tree was created to gain insight into the evolutionary history of the *PPX2* allele in waterhemp throughout the Midwest. Analysis options included the Maximum Composite Likelihood method and the pairwise deletion option. The IL-S population is from a known PPO-susceptible population in IL, while the IL-R population is from a known PPO-resistant population in IL. The Missouri population is a mostly PPO-susceptible population, while the Meadville population (also from Missouri) is a population that has newly-acquired PPO-resistance. The KS-R population is from a mostly PPO-resistant population in Kansas and the KS-S population is from a sensitive population from Kansas. Each taxon on the tree represents one plant (or one allele for the samples analyzed using the nested PCR).

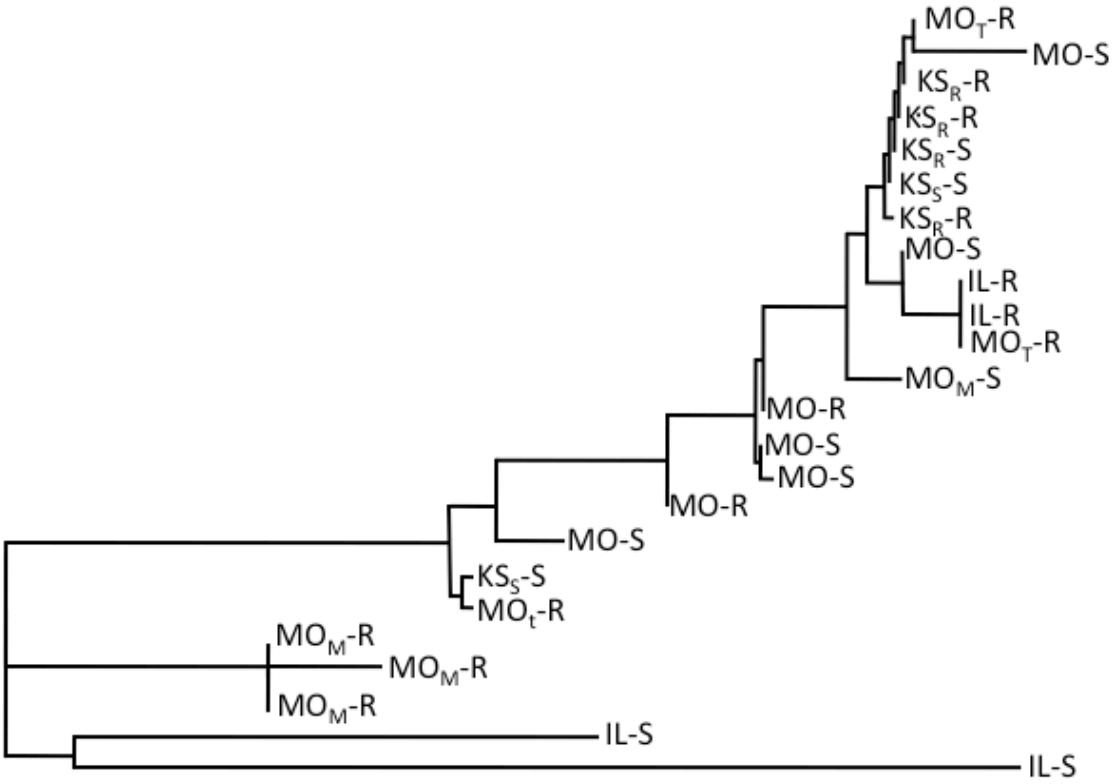


Figure 2.7 Recombination along the PPX2 allele would make tracking alleles difficult.

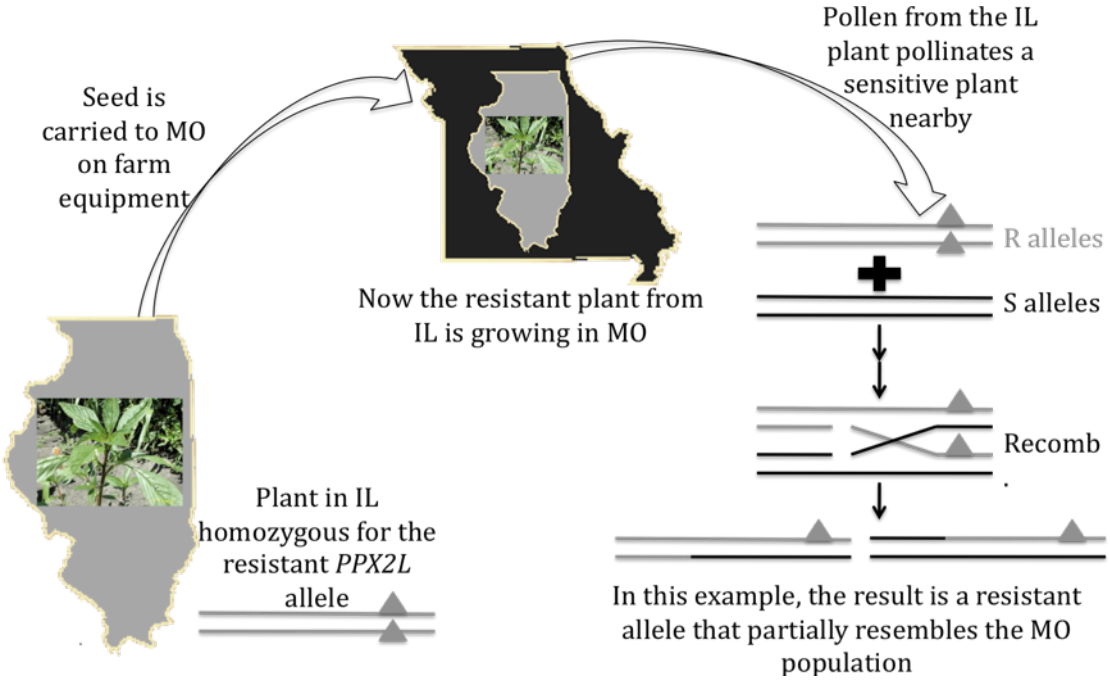


Table 2.1 Results from the single-rate screen with lactofen at 110 g ai ha^{-1} and screen for ΔG210.

Population	% Resistant	# Resistant Plants Screened for ΔG210	# Resistant Plants with ΔG210	# Sensitive Plants Screened for ΔG210	# Sensitive Plants without ΔG210
IL_R	100	4	4	0	NA
IL_S	0	0	NA	3	3
KS_R	82.4	10	10	4	3
KS_S	0	0	NA	2	2
MO1	11.3	5	4	12	11
Trenton	82.9	8	8	4	4
Meadville	65.7	8	8	6	5
Total	NA	35	34	31	28

Table 2.2 Contingency table analyses indicated that the ΔG210 resistance mechanism strongly correlated with resistance, regardless of the source population. Plants were first screened for resistance to the PPO-inhibiting herbicide lactofen and then each was assayed for the presence or absence of the resistance-encoding ΔG210. The X^2 value (47.567) greatly exceeded the X^2 critical value (based on 1 degree of freedom and α=0.05, critical value was equal to 3.841), indicating that biotype strongly associated with the presence or absence of ΔG210.

Biotype	ΔG210 absent	ΔG210 present	*Total*
Resistant	1	34	35
Sensitive	28	3	31
Total	**29**	**37**	**66**

Pearson's Chi-squared test with Yates' continuity correction

X-squared = 47.567, df = 1, p-value = 5.316e-12

CHAPTER 3

USE OF SIMPLE SEQUENCE REPEAT MARKERS FOR WATERHEMP POPULATION GENETICS

3.1 Abstract

Simple Sequence Repeat, or SSR, markers are useful tools in population genetics studies because they amplify highly variable regions of the genome. These markers can be easily identified through the PCR process using primers that flank the repeating sequence and visualization of the PCR product on an agarose gel. The development of a set of SSR markers for waterhemp population genetics studies is important since waterhemp is a troublesome weed in the Midwest that has developed resistance to several herbicide modes of action. Therefore, insight into waterhemp's population structure and how gene flow occurs would be helpful in examining the spread of herbicide resistance traits. In this study, Roche 454 sequencing technologies were used to identify SSRs throughout waterhemp's genome and primers were developed that flank the SSRs. These primer pairs were initially screened using four Midwestern waterhemp populations to develop a working set of SSR markers. Then, a wider sample set from six waterhemp populations from Illinois that had been screened for resistance to PPO-inhibiting herbicides were screened using the SSR markers and STRUCTURE analysis revealed a lack of population structure among these six populations.

3.2 Introduction

Although waterhemp is a troublesome weed throughout Midwestern agronomic fields, little is known about its molecular biology. Due to these restrictions, waterhemp population genetics studies have been limited. Recently, however, Roche 454 sequencing technology has created the opportunity for genome-wide sequencing experiments on waterhemp to become available. In spite of this, several of waterhemp's "weedy" characteristics make population genetics studies difficult. Specifically, a high level of intraspecific variation exists within waterhemp and there is a high level of gene flow between natural habitats and agronomic fields. Also, waterhemp is capable of adaptive plasticity, allowing it to quickly adapt to new environments (Hatzler 2003). These characteristics, combined with the lack of molecular information that was not available until recently, have made waterhemp population studies almost nonexistent until now.

Roche 454 sequencing has allowed for breakthroughs in sequencing technologies by increasing throughput, while decreasing cost (Rothberg and Leamon 2008). This technology has allowed for the identification of molecular markers throughout an organism's genome. Molecular markers are useful in a variety of genetic experiments, including linkage analysis, population genetics, and phylogeny. Simple Sequence Repeats (SSRs) are regions of nucleotide repeats, usually one to four base pairs in length. Benefits of SSR markers are that they amplify highly polymorphic regions, so fewer markers are needed to show relatedness, and they are codominant, meaning they can amplify both homozygotes and heterozygotes (Mallory et al. 2008; Ellwood et al. 2006). This makes them useful in waterhemp population genetics studies, since dioecious plants have a high level of heterozygosity.

STRUCTURE is a free software package that is useful in analyzing multi-locus genotype data to investigate population structure. Its uses include inferring the presence of distinct populations, assigning individuals to populations, studying hybrid zones, identifying migrants and admixed individuals, and estimating allele frequencies in populations with a large number of migrants or admixture (Pritchard et al. 2000). These inferences can be applied to most genetic markers, including single nucleotide polymorphisms (SNPs), restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs), and SSRs. It uses a Markov Chain Monte Carlo (MCMC) algorithm, which is useful in sampling from a probability distribution. The MCMC algorithm is based on a Markov chain, which is a sequence of random variables that represent past, present, and future events (Pritchard et al. 2000).

Since waterhemp has become such a prevalent weed and is quickly evolving resistance to multiple herbicides, it is useful to have molecular markers for this plant so that we can track its evolution and the spread of its weedy traits. The goal of this study was to develop a set of SSR markers that would be useful for tracking allele movement among waterhemp plants in the Midwest.

3.3 Materials and Methods

Roche 454 sequencing data came from an Illinois waterhemp population from Adams County. This population was the first to be identified as having three-way herbicide resistance, as it is resistant to three herbicide modes of action: PPO-inhibitors, acetolactate synthase (ALS) inhibitors, and photosystem II inhibitors. The sequencing

data were assembled to form contiguous consensus sequence (contig) data that were all greater than 100 nt in length and a Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990) was used to separate the contigs into chloroplast and nuclear sequences. To identify the SSR markers in waterhemp, the nuclear sequences were analyzed with SSR Finder software (http://www.maizemap.org/bioinformatics/SSRFINDER/SSR_Finder_Download.html). This software designs primers to amplify regions with SSRs, filtering out those primers with low complexity or primer sequences that were repeated elsewhere in the genome. Primers were further evaluated to select those that had an optimum annealing temperature around 63 C and that amplified fragments under 200 bp in length so that small differences in fragment lengths could be identified.

DNA isolated from four waterhemp populations (ACR plus an additional population from Illinois and one each from Kansas and Missouri) was used for amplification. Amplification of the SSR fragments was conducted from PCRs containing approximately 100 ng DNA, 0.3 µl Taq polymerase[1], 1× reaction buffer, 25 mM $MgCl_2$, 0.2 mM deoxyribonucleotide triphosphates, and 300 nM each of forward and reverse primer in a final volume of 20 µl. Reactions were incubated for 2 min at 95 C; then underwent 38 cycles of 0.5 min at 95 C, 0.5 min at 58 C, 45 s at 72 C, followed by 4 min at 72 C. The PCR products were fractionated on a 4% Low Range Ultra agarose gel[2] containing 0.5 µg ml^{-1} ethidium bromide and visualized with ultraviolet light. A subset of amplified fragments of the expected size were selected for DNA sequencing. Fragments were isolated from the gel and sequenced using AB BigDye Terminator v3.1 Cycle Sequencing Kit[3]. The resulting sequence data were analyzed using Sequencher 4.7

software[4]. Using Multalign (Corpet 1988), sequences were aligned and compared to the known sequence data obtained from Roche 454 sequencing. Primers that consistently amplified fragments of the expected size were analyzed further.

In order to determine if there was any population structure among resistant and sensitive populations in Illinois, plants from 4 different counties in Illinois (Schuyler, Grundy, Christian, and Henderson) were used. Plants from a resistant and sensitive population each from Grundy County and Christian County, a resistant population from Henderson County, and a sensitive population from Schuyler County were all screened for resistance to lactofen. These counties were chosen because each has reported waterhemp that is resistant to PPO-inhibiting herbicides, yet for the most part they are geographically spread across the state of Illinois. Henderson County is located on the western edge of Illinois and Schuyler County is southeast of Henderson County; the two counties are separated by McDonough County. Christian County is located in Central Illinois, while Grundy County is near the northeast corner of Illinois. About 100 seeds from each population were planted in flats containing 1:1:1 mixture of soil:peat:torpedo sand. At the two-leaf seedling stage, plants were transplanted into pots with a 12-cm diameter and grown until average plant height reached 10 to 12 cm. At this time, DNA was extracted from new leaf tissue using the hexadecyltrimethyl-ammonium bromide (CTAB) method (Doyle and Doyle, 1990). Treated plants were sprayed with lactofen at 110 g ai ha^{-1}, a dose shown to discriminate between resistant and sensitive plants (Patzoldt 2006), plus 1% crop oil concentrate (COC) and controls were sprayed with 1% COC using a compressed-air, moving-nozzle laboratory sprayer equipped with an 80° flat fan nozzle[5] delivering 187 liters·ha^{-1} of water at 207 kPa. The nozzle was maintained ≈45

cm above the plant canopy. Herbicide resistance was evaluated 14 days after treatment (DAT) and plants were scored as either dead or alive. Sensitive plants from Schuyler County, the sensitive Grundy County population, and the sensitive Christian County population, plus resistant plants from the resistant Grundy County and Christian County populations and Henderson County were used for further analysis. The extracted DNA was then subjected to PCR using all the primer pairs that were previously identified to amplify polymorphic SSRs and the PCR products were visualized on a 4% agarose gel, as described previously. The gels were analyzed and scored based on band size. This data was formatted to be analyzed using STRUCTURE (Pritchard et al. 2000).

Gel data was analyzed using STRUCTURE, where it was run under several different sets of initial parameters. The data was run using 10,000 MCMC repetitions for the burn-in period and 10,000 repetitions after the burn-in. To stabilize the likelihood and LnP(D) data plots, the burn-in was increased to 1,000,000 for the burn-in and to 1,000,000 for after the burn-in. For each set of parameters, one run was done with K (# of assumed populations) equal to 1 to estimate lambda, and then set this value of lambda for the subsequent runs for K = 2 through K = 6.

Different parameters tried included: admixture model and correlated allele frequencies between populations assumed (estimated lambda = 0.8272); admixture model and independent allele frequencies between populations assumed (estimated lambda = 0.8272) admixture model, correlated allele frequencies, and population IDs as sampling location information priors (estimated lambda = 0.8322); and no admixture model, correlated allele frequencies (estimated lambda = 0.8336).

3.4 Results and Discussion

The SSR Finder program identified 382 primer pairs that met the design criteria. Of 136 primer pairs tested on four different waterhemp populations, 54 amplified a fragment of the expected size in each of the four populations (Figure 3.1). Many of the other primer pairs amplified the expected-size fragment, but were discounted for further use for a variety of reasons including inconsistent amplification among the four populations, nonspecific amplification, and faint band intensity. Nevertheless, these primer pairs might be useful if more time was spent optimizing the PCR conditions. From the 54 primer pairs that successfully amplified the correct fragment size in all populations, the fragment from one non-ACR population from each of 23 selected primer pairs was sequenced. In each case, the resulting sequence contained the expected SSR, confirming that the correct fragment was amplified (Table 3.1). Thirteen of the 54 successful primer pairs also yielded PCR products that appeared polymorphic in size when analyzed by gel electrophoresis. Four of the polymorphic PCR products were further evaluated by DNA sequencing and in all four cases the sequence data confirmed the presence of differing numbers of repeats. An example of one such polymorphic marker is shown in Figure 3.2.

The lactofen screen yielded the following: 13 out of 14 plants from the Schuyler County were susceptible to lactofen treatment, 6 out of 9 plants from the resistant Grundy County were resistant to treatment, 3 out of 3 plants from the sensitive Grundy County were susceptible to herbicide treatment, 1 out of 4 plants from the resistant Christian County population were resistant to treatment, 11 out of 14 plants from the susceptible Christian County population were sensitive to herbicide treatment, and 10 out of 14 plants from Henderson County were resistant to herbicide treatment. Sensitive plants

from the susceptible populations and resistant plants from the resistant populations were used for molecular marker analysis.

DNA from the 4 Illinois counties was amplified using the 13 primer pairs that previously amplified bands of varying sizes. Of these 13 primer pairs, all 13 appeared to amplify PCR products of varying sizes when visualized by gel electrophoresis (Figure 2.3). Gels were scored

based on whether or not a band appeared at each expected size (Table 3.2). These data were used for further analysis using STRUCTURE.

STRUCTURE analysis revealed a lack of population structure. LnP(D) is the probability that determines what K value fits the best to the given data; when LnP(D) drops suddenly (becomes more positive), this is the real value of K. However, the data never yielded a steep drop with any parameters or value of K. Furthermore, LnP(D) values often differed for the same K value, which should not occur when population structure is present. Bar plots, which represent how well each individual fits into each population for K number of populations, were analyzed. Bar plots appeared different for different runs (Figure 3.4) at the same K value and tended to show that each individual in the data set belonged equally to all K populations (Figure 3.5). In addition, Fst and alpha values stabilized over the course of each run on STRUCTURE, which is indicative of a lack of population structure.

With such a large amount of variation and many opportunities for gene flow to occur, tracking alleles proves to be a difficult task. Although above populations could not be distinguished using the SSR markers developed in this study, these same markers may

be useful in other waterhemp population genetics studies, such as for looking at waterhemp populations that are separated by greater geographical distance or examining sympatric waterhemp populations in natural habitats.

Furthermore, the fact that no population structure was seen within these 6 waterhemp populations further attests to the high amount of gene flow in waterhemp. Given waterhemp's high amount of variation and lack of population structure (possibly due to gene flow), waterhemp population studies appear to be difficult.

3.5 Sources of Materials

[1] Taq polymerase, New England Biolabs, 240 County Road, Ipswich, MA 01938.

[2] Low Range Ultra agarose gel, Bio-Rad Laboratories, 1000 Alfred Nobel Drive, Hercules, CA 94547.

[3] AB Big Dye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems Inc, 850 Lincoln Centre Drive, Foster City, CA 94404.

[4] Sequencher™ 4.7 software, Gene Codes Corporation, 775 Technology Drive Suite 100A Ann Arbor, MI 48108 USA.

[5] Compressed-air, moving-nozzle laboratory sprayer equipped with an 80° flat fan nozzle, Teejet, Spraying Systems Co., P.O. Box 7900 Wheaton, IL 60187

3.6 Literature Cited

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* 16(22):10881-10890

Doyle, J. J. and J. L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13–15.

Ellwood S.R., N.K. D' Souza, L.G. Kamphuis, T.I. Burgess, R.M. Nair, R.P. Oliver. 2006. SSR analysis of the Medicagotruncatula SARDI core collection reveals substantial diversity and unusual genotype dispersal throughout the Mediterranean basin. *Theor. Appl. Genet.* 112:977–983.

Hartzler, B. 2003. Waterhemp - the perfect weed? Iowa State University. http://www.weeds.iastate.edu/mgmt/2003/symposium.shtml

Mallory, M. A., R. V. Hall, A. R. McNabb, D. B. Pratt, E. N. Jellen, and P. J. Maughan. 2008. Development and characterization of microsatellite markers for the grain amaranths. *Crop Sci.* 48:1098–1106.

Patzoldt, W. L., A. G. Hager, J. S. McCormick, and P.J. Tranel. 2006. A codon deletion confers resistance to herbicides inhibiting protoporphyrinogen oxidase. *Proc. Natl. Acad. Sci. USA* 103:12329–12334.

Pritchard, Jonathan K., Matthew Stephens and Peter Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155:945–959.

Rothberg, J. M. and J. H. Leamon. 2008. The development and impact of 454 sequencing. *Nat. Biotechnol.* 26:1117-1124.

3.7 Figures and Tables

Figure 3.1. Examples of primer pairs that amplified simple sequence repeat (SSR)-containing DNA fragments of the expected sizes. Different primer pairs were used in polymerase chain reactions (PCRs) with DNA from the Adams County, IL biotype as the template. Numbers above the lanes indicate the expected sizes (in base pairs) of the PCR products. A DNA marker is shown with the 100 and 200 base pair fragments indicated.

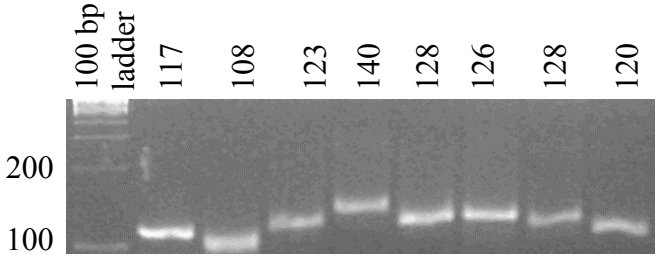
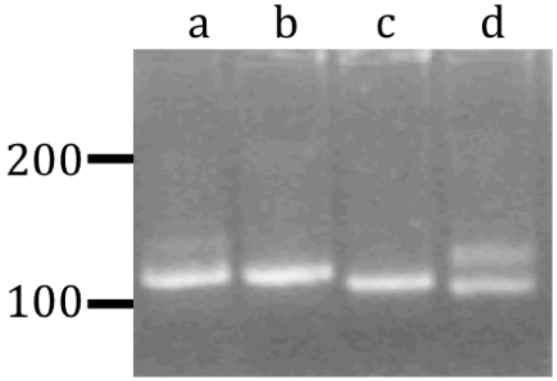


Figure 3.2. (Panel A) Gel analysis of a simple sequence repeat (SSR) marker amplified by a primer pair flanking a CAT repeat from four different populations. Lanes 1 through 4 (labeled a, b, c, and d) show the polymerase chain reaction (PCR) products from plants obtained from Adams Country, IL (ACR), Wayne County, IL (WCS), Missouri (MO), or Kansas (KS) population, respectively. DNA size markers (in base pairs) are indicated to the left. The ACR-derived PCR product was expected to be 114 base pairs in length. Gel analysis indicated that the MO-derived PCR product was smaller in size, and thus contained fewer repeats. This was confirmed by DNA sequencing, with partial sequence data shown in Panel B. As shown in lane 4 of Panel A, the KS plant apparently was heterozygous at the SSR locus analyzed.

A.



B.

```
ACR GAATCATCATCATCATCATCATCATTATCATGACAT
MO  GAATCATCATCATCAT---------TATCATGACAT
```

Figure 3.3. Gel analysis of a simple sequence repeat (SSR) markers amplified by a primer pair flanking an AAC repeat (Panel A) and TTG repeat (Panel B). In Panel A, the expected fragment size, based on ACR pyrosequencing data, was 107 base pairs in length, while the expected fragment size for Panel B was 123 base pairs in length. Waterhemp DNA samples from four different Illinois counties (sensitive plants from Schuyler County, resistant plants from Henderson County, sensitive and resistant plants from each Christian County and Grundy County) were PCR amplified and visualized by gel electrophoresis, as shown. The gel was scored based on band size. Four different-sized fragments were identified in Panel A, as seen on the gel, and this gel was one that was more difficult to score. Panel B shows amplification of 3 different-sized fragments and this gel was easier to score. On Panel B, Schuyler-S 5 was scored from a separate gel (not shown). Results from these 2 gels, plus 11 others were compiled into a table for further analysis (Table 2).

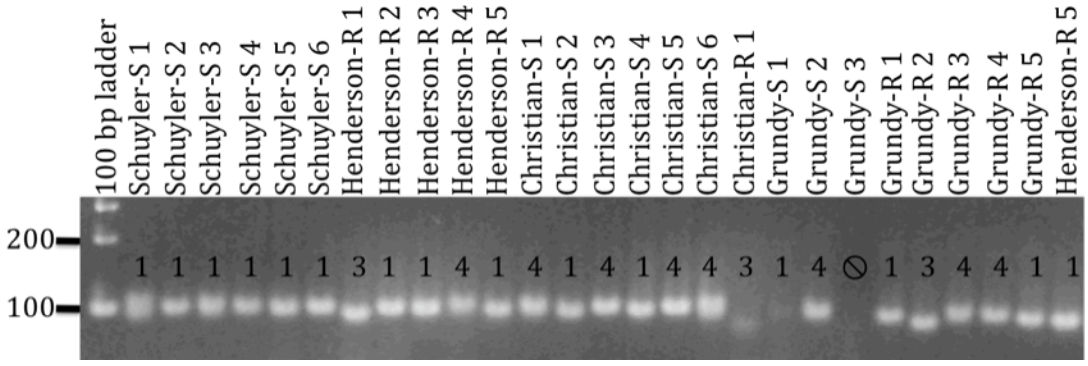


Figure 3.3 (cont.)

B.

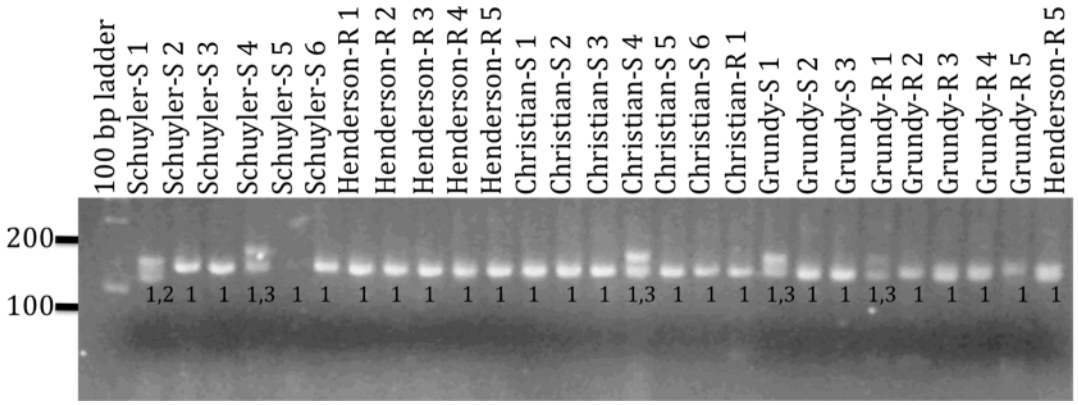


Figure 3.4. Waterhemp plants from four Illinois counties (a resistant and sensitive population each from Grundy County and Christian County, a resistant population from Henderson County, and a sensitive population from Schuyler County) were screened using 13 SSR markers and results were analyzed using STRUCTURE. Two STRUCTURE runs, using the admixture model and independent allele frequencies between populations assumed, with the same K-value (K=3) look different.

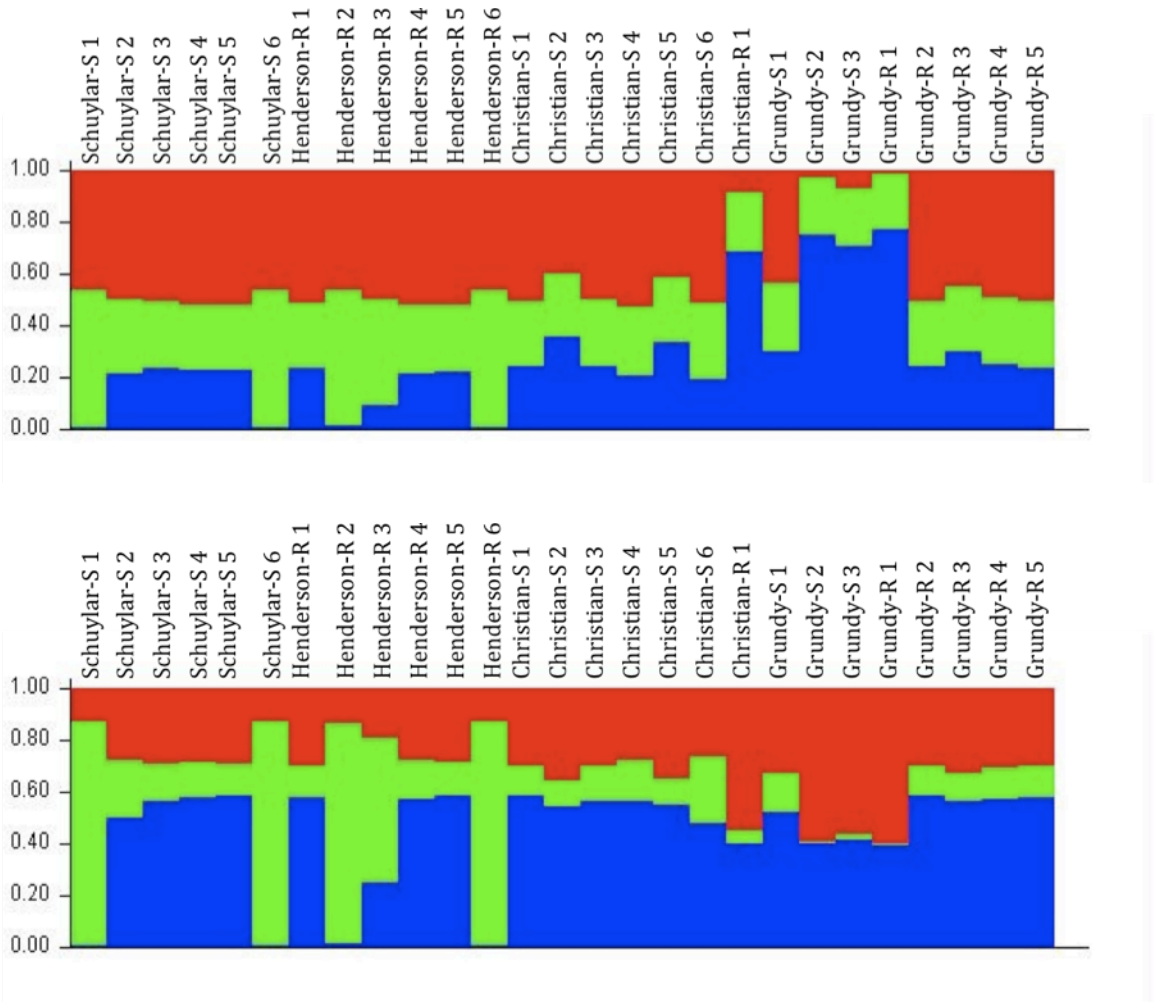


Figure 3.5. Waterhemp plants from four Illinois counties (a resistant and sensitive population each from Grundy County and Christian County, a resistant population from Henderson County, and a sensitive population from Schuyler County) were screened using 13 SSR markers and results were analyzed using STRUCTURE. Using the admixture model, independent allele frequencies between populations assumed, and K=6, STRUCTURE analysis suggested that each individual belongs to each of the K populations equally.

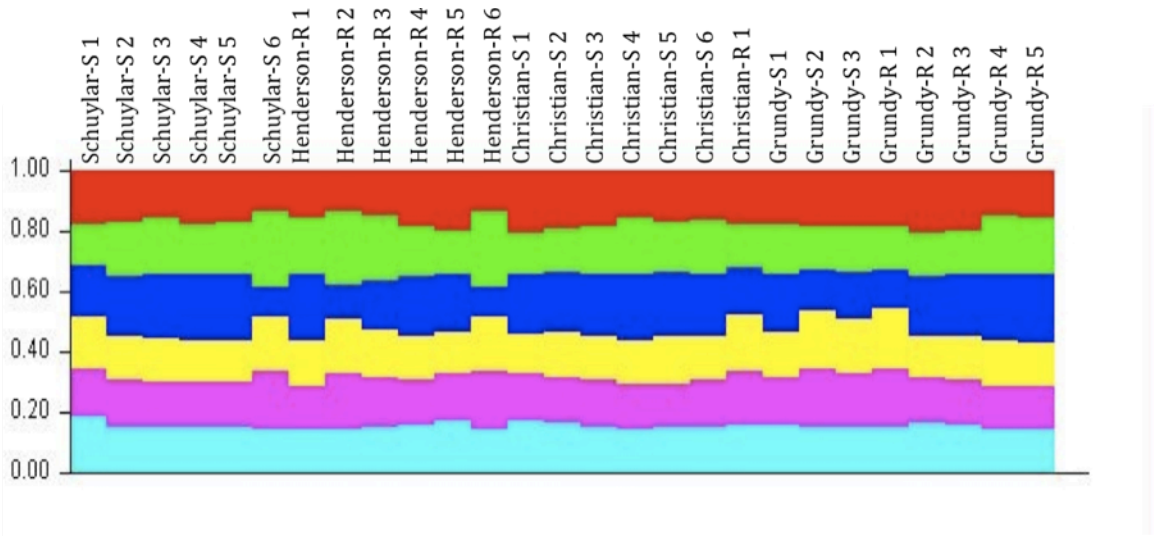


Table 3.1. Leads for 382 simple sequence repeats (SSRs) and corresponding primers to amplify each SSR were obtained from 454 pyrosequencing data using DNA from a waterhemp plant from Adams County, IL (ACR). Of 136 primer pairs tested, 54 amplified a fragment of the expected size in all four of four waterhemp populations (two from Illinois, one from Kansas, and one from Missouri) based off of gel data. Of the 54 working primer pairs, one fragment from one non-ACR population from each of 23 selected primer pairs was sequenced. Sequences were compared to the known sequence from ACR, which was obtained from 454 sequencing. All 23 resulting sequences contained the expected SSR, confirming that the correct fragment was amplified, and 13 proved to be polymorphic at the SSR as compared to ACR.

Forward primer	Reverse primer	Amplicon length (nucleotides)	Repeat sequence	Polymorphic
5'-gtcgctgaattgttttagcttggt	5'-tgggaattctctcttgtgacacagt	119	tag	yes
5'-ggacaatttggtcgaataggagtt	5'-gacggtctcaagtaagaatttgtgtt	188	aaag	yes
5'-tgttaaaggtccatcaaaaacaatctc	5'-tgctattaaatgactccctaggcaaa	108	tta	yes
5'-cccaccaaggatgatcatttagac	5'-tcatcattatttgttggcgttgac	107	aac	yes
5'-gtctcatatgtcatcataggcgga	5'-tggaactgtatcatcattcatggc	123	ttt	yes
5'-acaaggcacgaaagtaataaggct	5'-tgcggaatacactgggtagg	138	tca	yes
5'-caaacttctctcttctttaaatggaatc	5'-gatgatgatacaaggggtcaaaaa	114	cat	yes
5'-accaaatttcccaaatcttccatt	5'-tcaaattccttgaaattaggtggg	117	att	yes
5'-aatatgggaacaaatgaatggtcc	5'-ttgtcatctattggttcttgcagc	131	at	yes
5'-cttattgcttctgtcctgaccgtt	5'-cacattgctaccagtctcagagga	123	ttg	yes
5'-tagccatttcaaccttacgaggaa	5'-accgttgattgattttatggcatc	128	atc	yes
5'-tgagtcgcgatgatacttttctagg	5'-cgctttttctctccaccattctaa	140	tcc	yes
5'-attggatgatgtgcagtgtaatga	5'-tcaagttgacttctacattatgtggttt	125	aaa	yes
5'-accacatcagctggttgtgttagt	5'-gaataccaaacctgttaaccaaagc	117	atc	no
5'-cacactaaatgtacatggatcaaccaa	5'-gcgggataatgggagtgtcttatt	124	tat	no
5'-caatagcaggagcagaatcagtca	5'-ggataaggactatgatccgacgaa	116	cat	no
5'-ccagtccaaaatattcaatgatctgt	5'-ttttagtgcattgttgttctctaagtca	161	aag	no
5'-ccccttaacttgttcgtaaacct	5'-aaccatcaaaagttcctaccacga	130	tct	no
5'-atcatgcaacactctcctttaccc	5'-ggcctaaggtttgggtcacttaat	113	atc	no
5'-caagctcatctcccaactgaaagt	5'-aaacaccaaggaagatccaaatca	122	at	no
5'-caccaatgatgagcaagtttgaag	5'-tgcaataagcagaaaagaatggaa	117	ttt	no
5'-tggttttcaatattacattgtttgttgg	5'-tacatcacctccatgaccaagcta	126	att	no
5'- caagctcatctcccaactgaaagt	5'- aaacaccaaggaagatccaaatca	122	At	no

Table 3.2. Waterhemp DNA from four counties in Illinois (sensitive plants from Schuyler, Christian, and Grundy counties and resistant plants from Henderson, Christian, and Grundy counties) was PCR amplified using 13 different SSR markers. PCR products were run on a 4% agarose gel and scored based on whether or not a band appeared at each expected size. A score of '1' indicates that a band was seen at that given location, whereas a score of '0' indicates that no band was seen.

		sample																										
		Schuyler-S						Henderson-R						Christian-S						Christian-R	Grundy-S			Grundy-R				
		Plant #						Plant #						Plant #						Plant #	Plant #			Plant #				
Repeat amplified	Band #	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	1	2	3	1	2	3	4	5
tag	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
aaat	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
aac	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	1	1	0
aaa	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	1	1	1	1	0	1	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
	3	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
atc	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
aaa	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0
	2	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1	1	1	1	0	1	1	1
ttt	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ttg	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0
att	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	0	0	1	1	1	1	1	1
	2	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0
at	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
	3	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
tca	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	1	1
	2	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	1	1	1	0	1	1	1	1	1	0	0
tta	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	1	1	1
	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
aaa	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	1	1	1	1	0	1	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
	3	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0

CHAPTER 4

CONCLUDING REMARKS

4.1 Conclusions and Future Directions

Advances in scientific technology have opened the door for molecular research in agricultural sciences. While field studies still prove to be important in weed science, molecular work has helped the agricultural community gain insight into important issues, such as herbicide mode of action, herbicide resistance, and weed phylogeny, just to name a few.

The overall purpose of this thesis was to gain insight on Midwestern waterhemp resistant to PPO-inhibiting herbicides. Information surrounding the spread of resistant biotypes could provide a model for the spread of herbicide resistant waterhemp in the future and aid in population studies.

Prior to the studies discussed in this dissertation, molecular work allowed for the identification of waterhemp's mode of resistance to protoporphyrinogen oxidase (PPO)-inhibiting herbicides. A 3-base pair deletion (ΔG210) in the *PPX2* allele proved to be responsible for resistance to PPO-inhibiting herbicides in 6 waterhemp populations from Illinois (Lee et al. 2008). Since PPO-inhibiting herbicide resistant waterhemp is prevalent outside of Illinois, research efforts presented in this dissertation focused on determining whether these other resistant populations (from Missouri and Kansas) shared the same deletion mutation as the Illinois waterhemp populations. Chapter Two discusses how

PCR amplification and sequencing of a portion of the *PPX2* allele revealed that ΔG210 was strongly correlated with resistance in waterhemp from three Midwestern states (Illinois, Kansas, and Missouri). In total, out of 35 plants that survived treatment with lactofen at 100g ai ha^{-1} (therefore deemed resistant) that were screened for the deletion, 34 had the deletion. Conversely, out of 31 plants that were scored as sensitive after lactofen treatment that were screened for the deletion, 28 did not have the deletion. This data suggests that ΔG210 is waterhemp's only mechanism for resistance to PPO-inhibiting herbicides. This is significant because it allows for the use of a single molecular marker to test for resistance. Waterhemp suspected to be resistant to PPO-inhibiting herbicides can quickly and easily be screened with a primer that is specific for the resistant allele (Lee et al. 2008). This simple assay for resistance allows farmers to determine whether they have PPO-inhibiting herbicide resistant waterhemp in their field, and based on these results, herbicides can be applied accordingly.

This mechanism of resistance in waterhemp could have spread in two ways: either the resistant allele was selected for once and subsequently able to spread throughout the Midwest via pollen or seed flow, or the allele was independently selected for. Chapter Two discusses experiments that were performed in an attempt to answer this question. After sequence and phylogenetic tree analysis of waterhemp plants from Illinois, Kansas, and Missouri using a variable region of the *PPX2* allele, the inferred tree did not show enough structure to make any definite conclusions about the spread of PPO-inhibiting herbicide resistance among Midwestern waterhemp plants.

One possible explanation for why analysis of exons 2 through 4 did not resolve the evolutionary history of the *PPX2* allele could be because there is a high frequency of

recombination at this allele. To determine the role that recombination plays, the *PPX2* allele from multiple generations of waterhemp plants could be sequenced. These sequences could be aligned and analyzed using a program such as DNAsp to determine the recombination frequency of this allele.

One way to look at the spread of the resistant *PPX2* allele would be to sequence more of the allele and compile a sequence consisting solely of exon sequences. Since introns are non-coding regions and experience no pressure to maintain a conserved sequence, mutations in these regions, as seen in waterhemp's *PPX2* allele, appear to be abundant. Therefore, since mutations in introns may be occurring too frequently to track, the intron sequence may not be a good candidate to track allele movement. Instead of analyzing sequence data containing both introns and exons, an option would be to compare only the exon sequences. Since exons are coding regions, mutations in these regions may be more informational and may help gain insight into the spread of the resistant *PPX2* allele.

Another option to track the movement of the resistant *PPX2* allele would be to examine waterhemp ribosomal RNA (rRNA). Ribosomal RNA is highly conserved and is frequently used to determine evolutionary relationships among taxa (Smit et al. 2007). Although rRNA is extremely conserved, waterhemp is a quickly evolving species. Therefore, it would be interesting to see if waterhemp's rRNA provides any insight into the evolution and spread of the *PPX2* allele.

The hypothesis of the deletion mutation being a series of independent selection events seems more logical due to the fact that the three base pair deletion at position 210

is favored because of the surrounding sequence. At this position on the *PPX2* allele, there is a TGG repeat, which allows for a deletion of either GTG or TGG to result in the net loss of a glycine (TGG) (Patzoldt et al. 2006). Therefore, with intense herbicide selection pressure, this deletion seems like it could evolve and be selected for rather easily.

Waterhemp's affinity for ΔG210 can be contrasted to that of common ragweed (*Ambrosia artimisiifolia*), whose resistance mechanism to PPO-inhibiting herbicides is an amino acid substitution of arginine to leucine at position 98 (R98L) on the *PPX2* allele. Sequence analysis of the *PPX2* allele in common ragweed showed that R98L is a made possible by requiring only a single base pair substitution. This resistance mechanism would not be probable for waterhemp because waterhemp would require a change in two base pairs to achieve this same mutation. On the other hand, the ΔG210 resistance mechanism would not be a likely occurrence for common ragweed because the mutation-facilitating TGG repeat is not present (Rousonelos et al. 2009).

Chapter Three discusses the development of a set of simple sequence repeat (SSR) markers for waterhemp population studies. While these markers were not useful in tracking the spread of PPO-inhibiting herbicide resistant waterhemp in Illinois, the markers might prove to be useful in other experiments. For example, these SSR markers could be explored further by looking at samples separated by a greater geographic distance. Further geographic separation of samples would ensure more genetic variability and it is possible that population structure could be seen.

Additionally, since SSRs have an increased rate of mutation without the risk of lethal mutations, SSR markers are useful for detecting changes in population structure

over time. Since waterhemp is a quickly evolving species, it would be interesting to use these SSR markers to examine the amount of evolutionary change occurring in waterhemp over the span of multiple generations. This could possibly provide insight into how quickly herbicide resistance traits, and other evolutionarily favored traits, evolve. Furthermore, the 13 polymorphic SSR markers may be useful in creating a linkage map to gain insight into waterhemp's population structure. For the most accurate linkage map, it would be helpful to identify additional SSR markers to use along with the 13 described here. A study such as this could provide information on recombination frequencies. It may also be possible to run these same SSR markers on other members of the *Amaranthus* genus and compare linkage maps of different *Amaranthus* species.

4.2 Literature Cited

Lee, R. M., A. G. Hager, and P. J. Tranel. 2008. Prevalence of a novel resistance mechanism to PPO-inhibiting herbicides in waterhemp (Amaranthus tuberculatus). *Weed Sci.* 56:371-375.

Patzoldt, W. L., A. G. Hager, J. S. McCormick, and P.J. Tranel. 2006. A codon deletion confers resistance to herbicides inhibiting protoporphyrinogen oxidase. *Proc. Natl. Acad. Sci. USA* 103:12329–12334.

Rousonelos, S. L., R. M. Lee, and P. J. Tranel. 2009. An Alternative to the Glycine Deletion: Why R98L was Selected in Common Ragweed Protoporphyrinogen Oxidase. Proc. North Cent. Weed Sci. Soc. 64:98.

Smit, S., J.Widmann, and R. Knight. 2007. Evolutionary rates vary among rRNA structural elements. Nucleic Acids Res. 35(10): 3339–3354.