5-2014

Can Very Rapid Adaptation Arise Without Ancestral Variation? Insight from the Molecular Evolution of Herbicide Resistance in Genus Amaranthus

Kristin Elizabeth Beard
Clemson University

CAN VERY RAPID ADAPTATION ARISE WITHOUT ANCESTRAL VARIATION? INSIGHT FROM THE MOLECULAR EVOLUTION OF HERBICIDE RESISTANCE IN GENUS *Amaranthus*

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Genetics

---

by
Kristin Elizabeth Beard
May 2014

Accepted by:
Dr. Amy Lawton-Rauh, Committee Chair
Dr. Kerry Smith
Dr. Alex Feltus
Dr. Hong Luo

ABSTRACT

In this dissertation I investigate the source and spread of adaptive resistance to the herbicide glyphosate by the weedy species *Amaranthus palmeri*. This is an ideal system for furthering scientific understanding of the dynamics of evolution because the adaptation is in response to a well understood selection pressure and is happening on an extremely short time scale.

The plant genus *Amaranthus* contains several agriculturally important weeds, but is not closely related to any current model systems—the closest model system is *Beta vulgaris* (sugar beet). In the first part of this work I seek to determine the relationship between extant species of *Amaranthus*, particularly the relationship between weedy and non-weedy species, using Bayesian phylogenetic analysis of independent genomic loci. This phylogeny will provide context for investigating the dynamics of adaptation to glyphosate stress.

The second chapter is an investigation of the sequence constraints and selection pressures acting on the gene that codes for 5-enolpyruvalshikimate-3-phosphate synthase (*EPSPS*) in the genus *Amaranthus.* The first population of *A. palmeri* verified as resistant to glyphosate was identified in Macon, Georgia in 2004. The mechanism of resistance was found to be proliferation in copy number of the gene encoding the enzyme target of glyphosate toxicity, *EPSPS*. The proliferation of genomic copies of the gene encoding a target enzyme is unique among mechanisms documented for herbicide resistance—

though it has been observed as a resistance mechanism in other systems, including human cancer resistance to chemotherapy. Understanding *EPSPS* DNA sequence constraint will allow a better understanding of the evolutionary processes that led to a unique mechanism of herbicide resistance.

In the final chapter I seek to determine if the same *EPSPS* copy number proliferation mechanism is responsible for glyphosate resistance in North Carolina *A. palmeri* and address the question of the source of the *EPSPS* copy number proliferation genotype. I investigate the potential of parallel evolution from ancestral variation as an explanation for observed spread of resistance by looking for genus-wide variation in *EPSPS* copy number and analyze population structure to determine the most probable number of adaptive events.

Understanding the constraints on the EPSPS gene and protein that may have led to the observed resistance mechanism, how many times the mechanism evolved independently, and how it spread through the population(s) improves our understanding of how genomes are changed by adaptation to environmental stress. It also has the potential to provide important insights about the dynamics of herbicide resistance adaptation that can help growers make the best possible choices in weed management for protecting our food supply and our environment.

ACKNOWLEDGMENTS

There are quite a few people who I would like to thank: people who share credit for my completion of this degree. I would like first thank my family, or maybe blame them. Without their unfailing encouragement of critical thinking and exploration as a girl I might have never started this degree, and without their love and support throughout it I would not have finished. I also want to thank my husband; we met just before I chose a lab for my thesis work, and he has been there to support me for the last five years as my shoulder to cry on when experiments failed. Also, Liz Crouse and my fellow Genetics and Biochemistry grad students—particularly Krystal Cadle—for being the people to go have a drink with after I was done crying about failed experiments.

I would like to thank Amy Lawton-Rauh for the opportunity to further my education and for her enthusiasm, her encouragement, and her resolute dedication my success. She helped me find a work ethic that I did not know existed and her tireless support and help will always be appreciated. I would also like to thank Kerry Smith, Alex Feltus and Hong Luo for serving on my graduate committee. I would also like to thank Jim Burton and Nilda Burgos for their assistance with the weed science parts of this work, David Brenner at the USDA for germplasm, and Billy Bridges for all the advice about statistics. I am also thankful for the hard work and advice given to me by Brad Rauh, and all the undergraduate interns who helped along the way: Steven Vensko, Thomas Lane, Caroline Thomas, Sarah Barfield, and Katerina Lay. And lastly, I owe much appreciation to Cotton Inc., Clemson University, and Wade Stackhouse for generous financial support.

TABLE OF CONTENTS

Table of Contents (Continued)

Page

LIST OF TABLES

Page

Table

List of Tables (continued)

Page

Table

LIST OF FIGURES

Page

Figure

List of Figures (continued)

Page

Figure

PREFACE

This thesis has three primary objectives which are laid out in each of the next three chapters. These three objectives each address an important facet of the larger question considered in this work: how do adaptive mutations arise, spread, and persist in populations? Understanding the dynamics of adaptive change is an important part of understanding evolution and thus biology.

The first chapter presents the phylogenetic relationships amongst *Amaranthus* species. Amaranths are not closely related to any current model genomic systems and are currently an understudied group—though that is rapidly changing in the face of spreading glyphosate resistance. Currently the most extensive investigation of the taxonomic relationship between *Amaranthus* species is the one presented in the Flora of North America and is based on morphological character alone (Mosyakin and Robertson. 1993). A few studies have looked at a small subset of amaranths and determined their relationship based on DNA sequence: in 2001 Xu and Sun described the relationship between the cultivated *Amaranthus* (Xu and Sun. 2001), and another study described the relationship between weedy species *A. palmeri, A. powellii, A. retroflexus, A. arenicola, A. hybridus, A. spinosus, A. albus,* and *A. tuberculatus* (Wassom and Tranel. 2005). Amaranth have also been included in larger investigations of the plant family *Amaranthaceae,* these studies do not investigate the relationships between *Amaranthus* species, but instead the relationship of the genus to other genera in the family. Results of

these studies are not consistent due to differences in sampling, but the genera most closely related to *Amaranthus* are likely *Celosia, Chamissoa,* and *Pleuropetalum* (Kadereit *et al*. 2003; Sage *et al*. 2007).To date no large DNA sequence based phylogeny has been constructed for *Amaranthus*. In order to better understand the glyphosate resistance spreading in *A. palmeri* as well as *A. spinosus* and *A. tuberculatus* such an understanding of species relationships is needed.

The second chapter tests for the constraints on selection in the *EPSPS* gene that may cause the unique adaptive mechanism observed. While gene copy number proliferation is seen in some cases of insecticide resistance (Devonshire and Field. 1991), the mechanism identified by Gaines *et al.* for glyphosate resistance in *A. palmeri* is unique compared to mechanisms conferring resistance to other herbicides (Beckie and Tardif. 2012; Gaines. 2010; Gaines *et al*. 2011; Gaines *et al*. 2010). The goal of this work is to better understand the evolutionary dynamics of *EPSPS* in the genus *Amaranthus* as a way to understand the potentially unique evolutionary forces that encourage rapid adaptation to glyphosate stress via a unique, independent mechanism.

The third chapter focuses on investigating the source and spread of glyphosate resistance in *A. palmeri* growing in North Carolina. The first identified glyphosate resistant *A. palmeri* was in Georgia in 2004; in 2005 the first individuals were documented in North Carolina. Resistance has now spread throughout North Carolina as well as throughout other states (Heap. 2013). It is not currently known if the subsequent instances of

glyphosate resistance represent spreading of the original adaptive event, new independent mutations, or parallel evolution. This work seeks to address that question by focusing on *A. palmeri* collected from NC in 2010 due to this collection being one of the most rigorous done in response to glyphosate resistance (Culpepper *et al*. 2008; Whitaker *et al*. 2013; Whitaker. 2009).

References

Beckie, H. J., and F. J. Tardif, 2012 Herbicide cross resistance in weeds. Crop Protection **35:** 15-28.

Culpepper, A., J. Whitaker, A. MacRae and A. York, 2008 Distribution of glyphosate-resistant palmer amaranth (*Amaranthus palmeri*) in Georgia and North Carolina during 2005 and 2006. Journal of Cotton Science **12:** 306-310.

Devonshire, A. L., and L. M. Field, 1991 Gene amplification and insecticide resistance. Annual Reviews in Entomology **36:** 1-21.

Gaines, T. A., 2010 *Molecular Genetics of Glyphosate Resistance in Palmer Amaranth (Amaranthus Palmeri L.)*. Colorado State University.

Gaines, T. A., D. L. Shaner, S. M. Ward, J. E. Leach, C. Preston *et al*, 2011 Mechanism of resistance of evolved glyphosate-resistant palmer amaranth (*Amaranthus palmeri*). Journal of Agricultural Food Chemistry .

Gaines, T. A., A. A. Wright, W. T. Molin, L. Lorentz, C. W. Riggins *et al*, 2013 Identification of genetic elements associated with EPSPS gene amplification. PloS One **8:** e65819.

Gaines, T. A., W. Zhang, D. Wang, B. Bukun, S. T. Chisholm *et al*, 2010 Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proceedings of the National Academy of Sciences **107:** 1029-1034.

Heap, I. M., *International Survey of Herbicide Resistant Weeds*. Online. Internet. Wednesday, November 06, 2013 . Available www.weedscience.com

Kadereit, T. Borsch, K. Weising and H. Freitag, 2003 Phylogeny of *amaranthaceae* and *chenopodiaceae* and the evolution of C4 photosynthesis. International Journal of Plant Science **164:** 959-986.

Sage, R. F., T. L. Sage, R. W. Pearcy and T. Borsch, 2007 The taxonomic distribution of C4 photosynthesis in amaranthaceae sensu stricto. American Journal of Botany **94:** 1992-2003.

Wassom, J. J., and P. J. Tranel, 2005 Amplified fragment length polymorphism-based genetic relationships among weedy *Amaranthus* species. Journal of Heredity **96:** 410-416.

Whitaker, J. R., J. D. Burton, A. C. York, D. L. Jordan and A. Chandi, 2013 Physiology of glyphosate-resistant and glyphosate-susceptible palmer amaranth (*Amaranthus palmeri*) biotypes collected from North Carolina. International Journal of Agronomy **2013**.

Xu, F., and M. Sun, 2001 Comparative analysis of phylogenetic relationships of grain amaranths and their wild relatives (*Amaranthus*; *Amaranthaceae*) using internal transcribed spacer, amplified fragment length polymorphism, and double-primer fluorescent intersimple sequence repeat markers. Molecular Phylogenetics and Evolution **21:** 372-387.

INTRODUCTION

HERBICIDE RESISTANCE IN *Amaranthus* AS A SYSTEM TO STUDY ADAPTIVE EVOLUTION: A REVIEW OF CURRENT LITURATURE

K. E. Beard

Literature Review

The adaptation of weedy amaranth species to glyphosate represents a threat to agriculture and a unique opportunity to understand the source and spread of adaptive genetic variation during natural selection. The first part of this literature review will examine the history of scientific understanding of variation in natural populations with particular focus on theories about the maintenance of observed levels of variation. From there, this review will describe the current research in herbicide resistance and why herbicide resistance is an ideal and important system in which to study adaptive evolution.

Genetic Sources of Adaptive Variation

Mutation, including both point mutations and genomic rearangments (sequence transpositions, duplications, and deletions), in the DNA sequence of germ-line cells is the fundamental source of heritable diversity. Most mutations are caused by DNA replication or repair errors. Rates of point mutation are estimated to be in the range of 10^{-4} to 10^{-11} nucleotides per replication cycle (Drake *et al*. 1998), which is a large range influenced by many factors. Many of these factors also influence the rate of rearrangement mutations. Mutation rate can be related to sequence context; methylated CpG nucleotides have an average of 10-fold higher mutation rate than nucleotides with different sequence contexts or methylation status (Hodgkinson and Eyre-Walker. 2011). Another factor is position within a genome. Sequence near the telomeres of chromosomes have an approximately 20% higher mutation rate (Tyekucheva *et al*. 2008). The entire length of the Y chromosome has a higher mutation rate than other chromosomes in primates (Makova

and Li. 2002). In plants the silent substitution rate (a proxy for overall mutation rate) in mitochondrial DNA (mtDNA) is less than one-third that in chloroplast DNA (cpDNA), which in turn evolves only half as fast as plant nuclear DNA (Wolfe *et al*. 1987). While in animals mtDNA mutates an order of magnitude faster than nuclear DNA (Wolfe *et al*. 1987). This also points to a third factor influencing mutation rate—lineage. Lineage dependent difference in mutation rate, such as the mutation rate of the mitochondrial genomes of plants versus animals, can be related to differences in DNA repair pathways (Nei *et al*. 2010), as well as the biology of the organism. Genetic recombination can introduce or correct mutations, and thus the rate of recombination can influence mutation rate. Additionally, among sexually reproducing organisms, species with more DNA replication cycles during gametogenesis will have an apparent higher mutation rate (Makova and Li. 2002).

In addition to mutation rate, level of sequence variation is dependent on persistence of mutations in the population. Different mutations will persist at different rates amongst lineages and across loci depending on forces such as drift and natural selection (Nei *et al*. 2010). The relative influences of selection versus random drift in shaping variation among individuals has been a long ranging debate since it was first noted by Darwin that phenotypic variation exists in natural populations (Nei *et al*. 2010; Darwin. 2009).

The first theory about mutation persistence was described by Fisher and Haldane in the 1920s, and is called the Selectionist Theory. It is essentially Darwin's theory of natural

selection for phenotypes extended to non-phenotypic diversity (such as isozyme or nucleotide sequence variation) (Fisher. 1930; Haldane. 1937; Darwin. 2009). Fisher's work first uncovered the idea of the quantitative trait: that the continuous variation measured by biometricians could be produced by multiple independent loci, and that natural selection could change allele frequencies in a population, resulting in evolution (Fisher. 1930). J.B.S. Haldane applied these ideas through statistical analysis to real-world examples, such as the proliferation of the dark morph of peppered moths in response to industrial pollution (Haldane. 1937). He showed that natural selection changed allele frequencies and population fitness at an even faster rate than Fisher predicted (Haldane. 1937). In summary, the Selectionist Theory is that mutation rate and strength of natural selection are the only forces driving observed diversity.

As the ability to interrogate organisms for diversity deepened via methods such as RFLP analysis (and later DNA sequencing) levels of diversity were observed that could not be explained predominantly by natural selection (Nei. 2010). In 1968 Motoo Kimura proposed a new model of evolutionary change at the molecular level: the Neutral Theory (Kimura. 1984). This theory asserts that nearly all DNA-level mutations are selectively neutral, neither increasing nor decreasing fitness. This means that observed nucleotide diversity is shaped by rate of mutation and level of genetic drift (i.e. population size). While this theory is still used as the statistical null hypothesis, the theory is incomplete: the connection between DNA sequence variation and phenotypic variation is missing.

Kimura accepted that natural selection did act on phenotypic variation but if genetic variation is selectively neutral, what is the source of the heritable phenotypic variation?

In 1973, Tomoko Ohta, a student of Kimura, introduced the Nearly Neutral Theory. This theory is a compromise between Selectionism and Neutralism; it asserts that most DNA-level mutations are very slightly deleterious or advantageous (Ohta. 1973; Ohta and Gillespie. 1996). This means that nucleotide diversity is driven differently depending on population size: large populations are less influenced by drift and more by natural selection while small populations are influenced more by genetic drift.

When a population experiences a selective stress any alleles that confer an advantage in the face of the stress will increase in frequency. While fundamentally these alleles must come from mutation events or heritable changes in epigenetic status, there is another perspective on the question of where the adaptive alleles come from that has important consequences for the dynamics of adaptation. Adaptive alleles can be thought of as having three sources: introgression from outside the population, standing variation within the population, or *de novo* mutation within the population (Hermisson and Pennings. 2005; Feldman *et al*. 2009).

Adaptation via *de novo* mutation is the simplest case. As mutations accumulate from generation to generation an advantageous allele can arise and spread in the stressed population. Because the probability of an advantageous allele arising through new

mutation in any generation is low, this mechanism is predicted to respond to stress slowly (Hermisson and Pennings. 2005). *De novo* mutations are also predicted to result in few alleles of large effect with dominant inheritance (Hermisson and Pennings. 2005). This is related to the principal of Haldane's sieve (Haldane. 1927). If the new allele is completely recessive then natural selection cannot act to increase its frequency until a homozygous individual arises, this could take many generations in an out-crossing species, and in the mean time genetic drift could remove the allele from the population through random loss. Similarly, if the allele is of small effect drift may be the more powerful evolutionary force. If the allele spreads through the population via natural selection the result is a hard selective sweep, which is an area around the selected site that is depleted of ancestral variation due to hitchhiking of neutral variants linked to the advantageous mutation and accumulating new rare alleles (Fay and Wu. 2000). Much of the research done on the sources of adaptive variants has uncovered *de novo* mutation as the most likely mechanism. Even though the mechanism is predicted to supply slower adaptation, it may be easier for research to find because of its resulting Mendelian inheritance and obvious genetic signatures. Examples include coat coloration among cats (Eizirik *et al*. 2003), prey toxicity tolerance in garter snakes (Feldman *et al*. 2009), as well as others in fish, birds, and mammals (Dowling *et al*. 2002; Mundy *et al*. 2004; Steiner *et al*. 2007; Strecker *et al*. 2003; Theron *et al*. 2001).

Another source of adaptive alleles is ancestral or standing variation. This is due to increased fitness in a new environment for a previously neutral (or even deleterious)

allele. These alleles are predicted to lead to a rapid response to stress (Barrett and Schluter. 2008; Innan and Kim. 2004), and more likely to result in adaptive alleles with smaller fitness gains and non-dominant inheritance relative to *de novo* mutation due to a higher initial allele frequency (Barrett and Schluter. 2008). If the standing variation is old enough, predating divergence between populations or species, it can lead to parallel evolution (convergent adaptation) in separate populations under similar stress (Schluter *et al*. 2004). Because the pre-adaptive allele has been segregating in the population for many generations before becoming adaptive the genetic signature of such an event is a soft sweep, which is harder to statistically detect than a hard sweep (Hermisson and Pennings. 2005; Pennings and Hermisson. 2006a; Pennings and Hermisson. 2006b). Soft sweep refers to the fact that the linkage with neutral variants has likely been broken up by recombination over the generations the allele was neutral so several similar haplotypes are part of the sweep. This obscures the expected pattern of an excess of rare haplotypes as there may be several haplotypes that become common via the sweep. The genetic footprints of soft sweeps have come under intense focus recently as it has been realized that this mechanism may be much more common than previously realized (Barrett and Schluter. 2008; Hermisson and Pennings. 2005; Olson-Manning *et al*. 2012; Pennings and Hermisson. 2006a; Pennings and Hermisson. 2006b). Examples from literature include the insertion of an *Accord*-like element into *Cyp6g1* in *Drosophilla melanogaster* (fruit fly) and its association with resistance to the insecticide DDT. Analysis of global patterns of this insertion shows that the insertion event predates the use of DDT, and patterns of variation are consistent with a soft sweep of the *Accord* insertion genotype in

areas with strong DDT stress (Catania *et al*. 2004). Another example is armor loss in three spine stickleback fish *(Gasterosteus aculeatus)* living in freshwater environments (Colosimo *et al*. 2004; Cresko *et al*. 2004). Research has revealed that many of the freshwater populations are fixed for a parallel mutation that segregates at low frequency in the marine populations. Other examples in have been found in plants and mammals (Steiner *et al.* 2007; Shimizu *et al*. 2004).

A third source of adaptive alleles is introgression or gene flow: a process where adaptive alleles are introduced to the stressed population via immigration (of individuals or gametes). Such events are considered introgression when the immigrant is of a different (sub-) species and gene flow is when the immigrant is of a different population of the same species (Hedrick. 2013). The increase in population fitness is predicted to be slow, like in the case of *de novo* mutations, as the adaptive allele is not already segregating in the population. The initial allele frequency will usually be low, like *de novo* mutations, unless multiple introgression events have occurred (Hedrick. 2013). Introgression events can potentially introduce both large and small effect alleles (Hedrick. 2013), but to spread the benefits of the new allele must to outweigh any cost of hybridization. This is generally a larger concern with introgression than with gene flow, although neighboring populations may have other traits that would be maladaptive in the environment they are introduced to. The genetic signatures of introgression and gene flow are similar to a sweep, but the linked region has—instead of reduced ancestral diversity—similarity to the source population and high divergence from the rest of the genome (Alcala *et al*.

2013; Hedrick. 2013). Examples of adaptation via introgression include warfarin resistance in house mice (*Mus musculus domesticus*), which is the result of an introgression event of the gene *vkorc1* (*vitamin K epoxide reductase subcomponent 1*) from *M. spretus* (Algerian mouse) (Song et al. 2011). Another example can be found with wolves; populations have shared alleles for coat color with dogs and body size with coyotes (Anderson et al. 2009; Kays et al. 2010). Other examples in humans and plants have also been found (Evans *et al*. 2006; Kenneth D. Whitney *et al*. 2006; Martin *et al*. 2006).

Herbicide Resistance as a Model of Adaptation

Herbicide resistance is a good model for adaptive evolution for both scientific and practical reasons. The purpose of herbicides is to remove unwanted plants—generally referred to as weeds—from areas of planned cultivation. The first chemical herbicide introduced was 2,4-Dichlorophenoxyacetic acid (2,4-D), a synthetic auxin developed in the 1940s. This was followed by the development of triazine herbicides and then many others: currently there are over two hundred different chemicals used as herbicides in the United States (Hawks. 2013). These herbicides can be grouped into classes that correspond to mode of action. Some classes of herbicide, such as G (glycines), have only a single member herbicide (i.e. there is only one chemical compound with this mode of herbicide action); other classes have many member herbicides, such as class A (acetyl-CoA carboxylase (ACCase) inhibitors) with twenty. It was not long after herbicides came into use that cases of weeds adapting and developing herbicide resistance were first

documented. The first cases were *Daucus carota* (wild carrot or Queen Anne's lace) in Canada and *Commelina diffusa* (climbing dayflower) in Hawaii with resistance to 2,4-D in 1957 (Heap. 2013). Since then the problem of herbicide resistance has only increased. As of writing there are 404 documented cases of herbicide resistance with some documented populations resistant to multiple herbicides (Heap. 2013).

Herbicide resistance is a threat to agriculture, human health, and the environment. As weeds become harder to kill growers must sacrifice yield, apply more chemicals, or increase the amount of physical tilling preformed (Powles and Yu. 2010; Price *et al*. July/August 2011; Rowland *et al*. 1999). Tilling buries the weed seeds, suppressing their growth, but also damages the soil and consumes more fuel (Powles and Yu. 2010). Applying more chemicals costs money and increases the levels of herbicide run off into our water supply and the ecosystem, and sacrificing yield results in economic loss through less product and lower quality product. The spread of herbicide resistant weeds already costs our economy millions—an estimated $200 million for 2011 in Tennessee alone—and the problem gets worse every season (Hembree. 2011).

Herbicide resistance is a valuable model of adaptive evolution because the selection pressure—herbicide application—is tractable and predictable and adaptation has been extremely rapid. Also, differences in crops, herbicide regulations, and geography lead to growers making different choices about herbicide application. This allows one to find both near-replicates and fields with known controlled differences to compare how

different herbicide application strategies influence weed adaptation (Gressel. 2011). Additionally, the rate of increase in the number of herbicide resistant weedy species suggests that this is an ongoing process with different populations at different points in the process of fixing adaptive mutations or genomic changes, permitting the opportunity to observe evolution at multiple points in a very short time scale. As new cases of herbicide resistance have been investigated several types of resistance mechanisms have been described.

Target Site Gene Mutation

Target site gene mutation is a process where the enzyme target of an herbicide is modified at the DNA sequence level causing a change in amino acid sequence and consequently protein structure that reduces the ability of the herbicide to inhibit its target. This mechanism plays a major role in resistance to most classes of herbicides, and is the predominate mechanism observed in classes B (acetolactate synthase (ALS) inhibitors), A (ACCase inhibitors), K (mitosis inhibitors), C (Photosystem II inhibitors), E (PPO inhibitors), and F1 (carotenoid biosynthesis inhibitors): many examples for each class can be found (Class B: Beckie *et al*. 2007; Boutsalis *et al*. 1999; Cruz-Hipolito *et al*. 2009; Cui *et al*. 2008; Cui *et al*. 2011; Delye and Boucansaud. 2008; Delye *et al*. 2009; Duran-Prado *et al*. 2004; Guttieri *et al*. 1992; Guttieri *et al*. 1995; Imaizumi *et al*. 2008; Intanon *et al*. 2011; Intanon *et al*. 2011; Jin *et al*. 2011; Kaloumenos *et al*. 2009; Kaloumenos *et al*. 2011; Kolkman *et al*. 2004; Krysiak *et al*. 2011; Liu *et al*. 2007; Marshall and Moss. 2008; Marshall *et al*. 2010; Massa *et al*. 2011; Ohsako and Tominaga. 2007; Park and

Mallory-Smith. 2004; Preston *et al*. 2006; Scarabel *et al*. 2004; Scarabel *et al*. 2010; Sibony *et al*. 2001; Sibony and Rubin. 2003; Tan and Medd. 2002; Uchino and Watanabe. 2002; Uchino *et al*. 2007; Wang *et al*. 2004; Warwick *et al*. 2008; Warwick *et al*. 2005; Warwick *et al*. 2010; Yu *et al*. 2003; Yu *et al*. 2007; Yu *et al*. 2008; Zheng *et al*. 2011), (Class A: Beckie *et al*. 2012; Beckie *et al*. 2012; Beckie *et al*. 2012; Beckie *et al*. 2012; Brown *et al*. 2002; Christoffers *et al*. 2002; Collavo *et al*. 2011; Cruz-Hipolito *et al*. 2011; Cruz-Hipolito *et al*. 2011; Delye *et al*. 2002; Délye *et al*. 2005; Délye *et al*. 2003; Hochberg *et al*. 2009; Hochberg *et al*. 2009; Kaundun. 2010; Liu *et al*. 2007; Liu *et al*. 2007; Liu *et al*. 2007; Petit *et al*. 2010; White *et al*. 2005; Yu *et al*. 2007b; Yu *et al*. 2007b; Zagnitko *et al*. 2001; Zhang and Powles. 2006), (Class K: Anthony *et al*. 1998; Beckie *et al*. 2012; Delye *et al*. 2004; Delye *et al*. 2004; Morrison *et al*. 1989; Smeda *et al*. 1992; Vaughn *et al*. 1987; Yamamoto *et al*. 1998), (Class C: Beckie and Tardif. 2012; Devine and Shukla. 2000; McCloskey and Holt. 1990; Park and Mallory-Smith. 2006), (Class E: Dayan *et al*. 2010; Devine and Shukla. 2000; Patzoldt *et al*. 2006; Randolph-Anderson *et al*. 1998), and (Class F1: Arias *et al*. 2006; Beckie *et al*. 2012; Michel *et al*. 2004; Puri *et al*. 2007). However, there is a large amount of variation in the number of resistant populations between herbicide classes. Some of this disparity is due to some herbicides having been on the market longer and thus weeds have had more generations to adapt. Additionally, some of this disparity is likely related to the relative sequence constraint on the genes encoding the proteins directly involved in herbicide toxicity. There might not be any target protein mutations that confer high levels of resistance without a major fitness cost for some herbicides (Menchari *et al*. 2008).

Altered Translocation / Absorption

Changes in patterns of herbicide translocation and absorption have been shown to be important mechanisms for class D herbicides (photosystem I inhibitors) such as Paraquat (*N*,*N*′-dimethyl-4,4′-bipyridinium dichloride) (Soar *et al*. 2003). The exact mechanism of this altered translocation is not known. It is known that this resistance is not as consistent as other mechanisms. In *Hordeum leporinum* (false barley) it has been shown that the altered translocation confers resistance to Paraquat only in cool weather (Purba *et al*. 1995). Resistance to auxin mimics (class O) is also sometimes due to altered translocation. MCPA (2-methyl-4-chlorophenoxyacetic acid, class O) and Fluroxypyr (4-amino-3,5-dichloro-6-fluoro-2-pyridinyl oxy acetic acid, class O) resistant *Galeopsis tetrahit* (brittle stem hemp nettle) have decreased translocation of the herbicide up towards the apical meristem (Weinberg *et al*. 2006). A mechanism similar to altered translocation has been described in *Avena fatua* (common wild oat). Enhanced gibberellins confer resistance to triallate (2,3,3-Trichloroallyl N,N-diisopropylthiocarbamate ,class N) and Difenzoquat (1,2-dimethyl-3,5-diphenylpyrazolium , class Z) by causing rapid shoot growth that prevents toxic levels of herbicides from reaching the meristem (Kern *et al*. 2002; Odonovan *et al*. 1994). A population of Fenoxaprop-p ((2R)-2-[4-(6-Chloro-2-benzoxazolyloxy)phenoxy]propionic acid, class A) resistant *Echinochloa phyllopogon* (rice barnyard grass) that has moderately reduced herbicide absorption has been reported (Bakkali *et al*. 2007). There is far less evidence for weeds exhibiting resistance via altered translocation than there is for

altered target protein gene sequences. This could be because translocation studies are not frequently performed on resistant plant populations in which another resistance mechanism has already been identified. However, there are several studies, particularly of ALS inhibitor resistant populations, where altered translocation has been explicitly ruled out (Al-Khatib *et al*. 1998; Saari *et al*. 1990). So it is not unwarranted to consider the specifics of herbicide biochemistry that makes altered translocation a common mechanism for photosystem I inhibitors and glycines but an uncommon or non-existent mechanism for other classes of herbicides.

Altered Herbicide Metabolism

Altered herbicide metabolism is, as a response to selection pressure, more complicated than target gene mutation. Two major gene families have been found to be responsible for most herbicide metabolism, the *cytochrome p450 monooxygenases* (*P450*) and the *glutathione S-transferases* (*GST*). Depending on which gene family member is mutated multiple patterns of multi-herbicide resistance can be observed. It is even possible for selection by a single herbicide to result in a population of weeds that is resistant to that herbicide and others from different herbicide classes to which the population has never been exposed. With both *P450* and *GST*, it is common for the selecting herbicide to be a member of classes A (ACCase inhibitors), B (ALS inhibitors), C (photo-system II inhibitors), or D (photo-system I disruptor) and to confer resistance to other herbicides in these same classes (A-D) as well as dinitroanilines (class K1) (Bakkali *et al*. 2007; Beckie *et al*. 2012; Beckie *et al*. 2012; Bravin *et al*. 2001; Burnet *et al*. 1993a; Burnet *et*

al. 1993b; Christopher *et al*. 1991; Cummins *et al*. 1997; Cummins *et al*. 1999; GimenezEspinosa *et al*. 1996; Kemp *et al*. 1990; Letouze and Gasquez. 2001; Letouze and Gasquez. 2003; Maneechote *et al*. 1997; Menendez and DePrado. 1996; Preston. 2004; Singh *et al*. 1998; Tardif and Powles. 1999; Yun *et al*. 2005). Resistance through herbicide metabolism, while fairly common, is almost exclusively found in monocots; the reason for this is not known (Beckie *et al*. 2012).

Target Gene Duplication

All herbicides have at least one target; the target is a protein that is inhibited or misregulated as a consequence of herbicide exposure. Generally this inhibition or misregulation results in the breakdown of a critical metabolic pathway leading to plant death. In just the last few years it has been shown that target site gene duplication is a major mechanism for resistance to glyphosate (glycine, class G) (Gaines *et al.* 2010). This mechanism has never been documented with any other class of herbicide, although it has been found as the mechanism of resistance to pesticides in insects (Devonshire and Field. 1991). Because of how recently this mechanism has been discovered in weedy plants, research may have not yet uncovered many of the populations taking advantage of this mechanism. There are many populations of weeds with unknown mechanisms of resistance. Another possibility is that this mechanism contributes to resistance in populations where another mechanism has been identified and presumed to be 100% responsible. It was shown in 1984 that cell cultures of alfalfa (*Medicago sativa*) exposed to glufosinate ammonium (class H) would develop high levels of resistance by gene

duplication (Donn *et al*. 1984) but no naturally occurring cases of this type of resistance have been documented.

Glyphosate Resistance as a Model of Herbicide Resistance

Herbicide resistance is an excellent model to study adaptive evolution. Herbicide resistance mechanism and ecology have been extensively studied due to obvious importance in industry. This work focuses specifically on the herbicide glyphosate (marketed as RoundUp™ by Monsanto). Focusing on glyphosate has two major advantages from both a scientific and practical perspective. Glyphosate is the sole member of herbicide class G; this means that there are no concerns about the intra-class cross resistance that is common in other, larger, herbicide classes. Also, there are no documented cases of cross resistance like what is observed in resistance to classes A-D and K1 via *GST* or *P450* catalyzed herbicide detoxification. There is also a huge amount of acreage that has been under a glyphosate dominated control regime for about twenty seasons. After growers adopted RoundUp Ready™ (i.e. glyphosate tolerant) crops, which were first introduced in the mid-1990s (Padgette *et al*. 1996; Padgette *et al*. 1995), glyphosate was sprayed post crop emergence and use of other herbicides was halted. This makes the selection pressure in these fields clearly timed, precisely located spatially, and very strong. On the practical side, there is a desparate need to better understand the evolutionary genetics of glyphosate resistance. Glyphosate is favored by growers because it is reportedly one of the safest herbicides, it does not contaminate water supplies, and it is not toxic to animals or humans (Franz *et al*. 1997). Due to these favorable traits, loss of utility of glyphosate threatens hundreds of thousands of acres of farmland with weedy

and invasive species infestations that cannot be controlled in an as environmentally friendly and sustainable manner (Price. 2011).

Glyphosate Mode of Action

The shikimate pathway (Figure I.1) starts with the non-hydrolytic addition of phosphoenolpyruvate (PEP) to erythrose 4-phosphate and ends with the formation of chorismate, the precursor to quinones, folates, and the aromatic amino acids phenylalanine, tyrosine, and tryptophan. This pathway is only found in plants and bacteria, animals must get their aromatic amino acids from their diet (Weaver and Herrmann. 1997). While bacteria use the pathway almost exclusively for the synthesis of amino acids, plants use chorismate as a precursor for various pigments, defense compounds, and lignin (Weaver and Herrmann. 1997). This critical pathway has seven steps catalyzed by at least 8 enzymes: two 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP) synthases, 3-dehydroquinate synthase, 3-dehydroquinate dehydrogenase, shikimate dehydrogenase, shikimate kinase, 5-enolpyrovylshikimate 3-phosphate (EPSP) synthase, and chorismate synthase. The second to last enzyme of this pathway, EPSP synthase (EPSPS), is the best studied of pathway; it catalyzes the reversible formation of EPSP from shikimate 3-phosphate and PEP. The reason EPSPS is the most studied enzyme of the shikimate pathway is that it is the target of the broad spectrum herbicide glyphosate. Glyphosate competitively inhibits EPSPS by binding to the EPSPS-shikimate-3-phosphate complex in place of PEP. Even though glyphosate is a structural analog of PEP it has not been shown to inhibit any other enzyme that uses PEP as a

substrate; this may be related to the fact that glyphosate only binds to the complex of EPSPS and shikimate-3-phosphate (Schönbrunn *et al*. 2001; Steinrücken and Amrhein. 1980; Weaver and Herrmann. 1997).

RoundUp Ready™ Mechanism

Many crop species have been genetically engineered to be tolerant to glyphosate. The most common method of conferring this resistance is through the introduction of a glyphosate insensitive EPSPS (Funke *et al*. 2006). The most common glyphosate insensitive EPSPS used was isolated from a strain of the bacterium *Agrobacterium tumefaciens*. This insensitive EPSPS was identified from screening the whole cell extracts of microorganisms in search of an EPSPS enzyme that maintained high catalytic efficiency in the presence of glyphosate. A strain of *Agrobacterium* known as CP4 (Padgette *et al*. 1995 was isolated from a glyphosate waste treatment facility (US patent 5633435). This strain contained what is known as a class II EPSPS; it has very low sequence homology to the class I EPSPS found in plants and *E. coli*. A pBLAST comparison of the *Agrobacterium* and *E. coli* EPSPS have 38% maximum identity. The very first RoundUp Ready™ crop was soy (*Glycine max*); it was developed by fusing the *Agrobacterium* CP4 *EPSPS* to the chloroplast transit peptide from petunia *EPSPS* and driving it with the *E35S* promoter then transforming this cassette into the soy cultivar A5403 (Padgette *et al*. 1996; Padgette *et al*. 1995). At the time RoundUp Ready™ crops were first developed it was not known why the CP4 EPSPS was insensitive to glyphosate, only that is was. Subsequent research showed that the main source of resistance was an

alanine in the active site (Ala-100) that made the PEP binding pocket slightly smaller. This smaller pocket could only be bound by glyphosate in a higher energy shortened configuration achieved by rotating a C-N bond. This shortened configuration was not only higher energy but also causes the glyphosate to clash with another active site residue, a glutamic acid (Glu-354). In short, the CP4 active site has a size and shape such that PEP fits but glyphosate does not fit well enough to be an efficient inhibitor (IC_{50} is 11mM vs. 2.5uM for *E. coli* EPSPS) (Funke *et al*. 2006). After the introduction of RoundUp Ready™ soy several other crops were introduced using the same gene to confer resistance. Corn (*Zea mays*) is somewhat different; there are several strains of glyphosate resistant corn and at least one of the strains, GA21, has resistance that is not conferred by insertion of the CP4 *EPSPS*. In this strain of corn the resistance is conferred via a altered version of the endogenous *EPSPS*. The *Z. mays EPSPS* was substituted at two points (TI02I and P106S) and then added to an expression cassette that would drive expression with an exogenous promoter, this was then transformed into the corn. This corn line contains two different *EPSPS* genes, the wild type glyphosate susceptible *EPSPS* and the resistant *EPSPS* that was derived from the wild type (US patent 6040497). Subsequent investigation of this glyphosate resistant *EPSPS* (called TIPS *EPSPS*) showed that these two substitutions are located in an alpha helix of the N-terminal globular domain, not in the active site. These two substitutions cause a change in the positioning of Gly-96 which narrows the binding site and creates steric hindrance for glyphosate but not PEP (Funke *et al*. 2009). In addition to insensitive EPSPS, resistance has also been conferred onto crops by introducing a glyphosate metabolizing gene, *glyphosate oxidoreductase* (*GOX*),

to cleave glyphosate into aminomethyl phosphonic acid and glyoxylate. One example of a cultivar using an introduced *gox* gene is oil rape seed (*Brassica rapa*) cultivar GT73; however, glyphosate metabolism is not the sole mechanisms of resistance, this cultivar also contains the CP4 *EPSPS* (US patent 5633448). The *GOX* gene used to confer glyphosate resistance came from the gram negative bacteria *Ochrobactrum anthropi* strain LBAA but was then modified with codon optimization and a chloroplast transporter peptide. While GOX does confer resistance to glyphosate, it alone does not confer enough resistance to be of use to growers, at least not in rapeseed, and must be combined with CP4 *EPSPS*. This is probably why most glyphosate resistant crops just have the CP4 *EPSPS* gene insertion (McVetty and Zelmer. 2007).

Gene Duplication

Some of both the newest and oldest research on the evolution of glyphosate resistance has involved gene duplication. As early as the 1980s researchers at Monsanto were experimenting with artificial selection in response to glyphosate stress in cell cultures. In these experiments, suspended plant cell cultures were exposed to increasing concentrations of glyphosate to select for resistance and then the mechanisms of resistance was determined. In experiments using petunia (cell line MP4), carrot (*Daucus carota*), alfalfa (*Medicago sativa*), soy (*Glycine max*), and tobacco (*Nicotiana tabacum*) cell cultures the response to glyphosate stress was increasing genomic copy number of *EPSPS* (Nafziger *et al*. 1984; Shah *et al*. 1986; Widholm *et al*. 2001). When further studying the tobacco cell lines, it was determined that they adapted by increasing steady

state EPSPS mRNA levels by increasing genomic copy number (Goldsbrough *et al.* 1990). Researchers then tried to regenerate whole glyphosate resistant tobacco plants, but when they were regenerated the plants were not resistant and many lacked vigor (Singer and McDaniel. June 1985). Eventually Monsanto was successful in creating glyphosate resistant crops using a different mechanism (see above) and this work was forgotten about until the late 2000s when a population of glyphosate resistant *Amaranthus palmeri* was found growing in a field of RoundUp Ready™ cotton in Georgia. Gaines *et al.* determined the mechanism of glyphosate resistance in this population to be massive *EPSPS* gene copy proliferation (40-100 copies) (Gaines. 2010; Gaines *et al*. 2011). They showed a linear relationship between the number of copies and the level of resistance, and that the extra copies were distributed across the genome, which suggests transposable elements may be involved. Another example of this type of resistance was found in an Arkansas population of Italian ryegrass (*Lolium multiflorum*). These plants, first identified in 2008, were shown to have up to 25 copies of *EPSPS* in their genomes (Salas *et al*. 2012). Likely in part due to the fact that this mechanism of resistance has been so recently rediscovered, there have been no studies of the potential fitness costs to the plants containing these massive levels of gene duplication. However, if we presume that transposable elements are responsible for the duplications it is not hard to imagine possible fitness costs for the plants. In order for these high levels of duplication to be achieved the transposons would be very active and if the transposon inserted itself into a critical gene in a seed or pollen precursor cell those seeds/pollen would be unviable and so plants with such high transposon activity could have lower fertility. Or if the disrupted

gene was less important the seed might be viable and produce a plant but the plant might lack vigor and not compete as well. Both of these only consider the effect of high transposon activity and not the *EPSPS* duplications themselves. Even if all the *EPSPS* copies are inserted in genomic regions that are accessible to transcription but not interrupting gene function there is still the fitness cost of producing 40-100 times as much EPSPS protein as a wild type plant. Clearly, under selection through glyphosate application, this will prove beneficial but in a glyphosate free environment these plants would likely be much less metabolically efficient and be out-competed by plants that were not wasting resources to make so much extra EPSPS protein. In addition to simple inefficiency it is also possible that having all this extra EPSPS protein might disrupt metabolism by changing the equilibrium concentrations of product and substrate (Bentley and Haslam. 1990).

Target Site Mutation

A target site mutation is a mutation to the gene sequence encoding an herbicide target protein that changes the amino acid sequence and thus changes the inhibition efficiency of the herbicide. Target site mutations to *EPSPS* conferring resistance to glyphosate were some of the first investigated when glyphosate tolerant crops were being developed, and eventually the TIPS mutant was developed and used in *Z. mays* (corn) to confer resistance (see above). However, as this requires two mutation steps, and only making one step will reduce catalytic efficiency with only a marginal increase in glyphosate resistance, it was not originally considered a likely source of weed resistance (Funke *et*

al. 2009). But when glyphosate resistant weeds started showing up in fields the *EPSPS* sequence was one of the first places that researchers looked. In some populations, particularly those with low level resistance, mutations in *EPSPS* have been found. All of the *EPSPS* mutations found to date in glyphosate resistant weeds have been mutations of Pro-106. As discussed in the section on RoundUp Ready™ crops, this residue is not directly involved in glyphosate binding. To better understand how the Pro-106 substitution conferred resistance in weeds when it was shown to be too inefficient to use in crops researchers produced *E. coli* EPSPS enzymes with glycine, alanine, serine, or leucine substituted for proline. These variant enzymes were analyzed by steady-state kinetics, and the crystal structures of the enzyme+shikimate-3-phosphate (S3P) and enzyme+S3P+glyphosate ternary complexes of P101S and P101L EPSPS (Pro-101 in *E. coli* is analogous to Pro-106 in plants) were determined. They showed that residues smaller than leucine may be substituted for proline without significantly decreasing catalytic efficiency, and that any substitution at this site results in a structural change in the glyphosate/PEP-binding site, shifting Thr-96 and Gly-97 into the active site (Healy-Fried *et al*. 2007). In the *E coli* model all the P101(S/G/A/L) alterations resulted in a decrease in V_{max} (from 50 to 22, 28, 31, and 8 u/mg respectively) (Healy-Fried *et al*. 2007). A more explicit study of fitness cost has never been done. Examples of populations that have become glyphosate resistant due to a mutation of Pro-106 can be found in several species. Sequencing of *EPSPS* in glyphosate resistant *Eleusine indica* (wire grass) populations growing in Malaysia showed multiple substitutions including two glyphosate resistance conferring amino acid substitutions: Pro-106 to serine and

threonine (Baerson *et al*. 2002; Ng *et al*. 2003). Several populations of *Lolium rigidum* (annual rye grass) have also been found with Pro-106 substitutions; one South African population has a P106A substitution (Yu *et al*. 2007a), a second has a P106L substitution that confers 1.7X resistance (Kaundun *et al*. 2011), a Chilean population has P106S substitution (Perez-Jones *et al*. 2007), and an Australian population has a P106T substitution (Wakelin and Preston. 2006).

Altered Absorption and Translocation

The other common mechanism for glyphosate resistance is change in translocation or absorption of the herbicide. Based on current literature this would seem to be the most commonly identified mechanism leading to high levels of resistance. It is possible that the literature represents a biased view of biological reality since it is only recently that gene duplication has been rediscovered as a mechanism for achieving high levels of glyphosate resistance in weeds. While there have been many studies documenting this as the mechanisms for resistance, the specific biochemistry of this mechanism has yet to be described. What is known is that in some populations plants do not absorb the herbicide as well, which reduces the level of plant injury, but it is not known how the plants prevent the absorption of glyphosate. It has been suggested that a simple change in the angle the leaf grows could be a contributing factor (Michitte *et al*. 2007). Another explanation is that the resistant weeds have thicker waxy coatings on their leaves (Cruz-Hipolito *et al*.), but the study supporting this mechanism compared resistant and susceptible populations of two different species, which means that other factors may be

contributing such as wax chemistry. Another study showed that there is not a simple linear relationship between waxiness and absorption of glyphosate (Norsworthy *et al*. 2001). With respect to altered translocation, the biochemistry of the mechanism is also not yet fully understood, even though many populations have been found taking advantage of this mechanisms. It has been shown that the altered translocation involves the glyphosate not moving from the sprayed leaf to the areas of new growth, keeping the fast growing tissue safe from aromatic amino acid synthesis inhibition. Many times the glyphosate is, instead of being loaded into the phloem and moved up, sequestered into the tips of the leaves that have been sprayed (Lorraine-Colwill *et al*. 2002; Michitte *et al*. 2005; Perez-Jones *et al*. 2007). Specifically it has been shown in resistant *Conyza canadensis* (Canadian horse weed) that there is much slower movement of glyphosate to the apoplast and then phloem (Feng *et al*. 2009). There have been efforts to understand the molecular and genetic basis of this resistance mechanism, but there has been little success to date. It is a difficult question because there are many genes and proteins involved in the regulation of intra and inter-cellular transport; including what compounds are loaded into the phloem. To date there has been one study on the transcriptomes of resistant (via altered translocation) and susceptible *C. canadensis*. The study showed that many transport genes, particularly four ABC transporters were significantly up-regulated in resistant individuals exposed to glyphosate (Yuan *et al*. 2010). The year before a review by Shaner also suggested that the accumulated evidence about altered translocation suggests an altered ABC transporter as the genetic mechanisms of resistance to glyphosate via altered translocation (Shaner. 2009). However, a specific

molecular mechanism has not been fully established. Some specific examples of weeds utilizing altered translocation or absorption to achieve glyphosate resistance include *Lolium rigidum* (annual rye grass) (Lorraine-Colwill *et al*. 2002; Wakelin *et al*. 2004; Yu *et al*. 2007a), *L. multiflorum* (Italian rye grass) (Michitte *et al*. 2007; Michitte *et al*. 2005; Perez-Jones *et al*. 2007), *Clitoria ternatea* (butterfly pea), *Neonotonia wrightii* (perennial soy bean) (Cruz-Hipolito *et al*.), *Conyza canadensis* (Canadian horse weed) (Feng *et al*. 2009), *Digitaria isularis* (sour grass) (Carvalho *et al*. 2011), and *Ipomoea lacunosa* L. (white star potato) (Norsworthy *et al*. 2001). The specific fitness costs have not been assayed in most of these populations. One of the populations of resistant *Lolium rigidum* (annual rye grass) was shown to be unlikely to have a fitness cost for its altered glyphosate trasport: the plants were shown to produce fewer seeds, but the seeds that were produced were larger and germinated at a higher rate. When the plants were in competition with crop wheat (*Triticum aestiva*), the resistant plants continued to produce larger seeds and the advantage the susceptible plants had in seed count diminished (Pedersen *et al*. 2007). This suggests an interesting and likely complex metabolic change as the result of changing the chemical transport regulation of the plant. Unfortunately, this study was done using only a single population, without replication, thus it is unclear if changes to the seeds are truly a side effect of altered translocation or if there are other mechanisms unrelated to glyphosate resistance at work in this population that increases competitiveness.

Metabolism

When glyphosate leaches into the soil, it is metabolized as a carbon and nitrogen source by soil microbes. The major metabolite of this glyphosate use is aminomethylphosphonic acid (AMPA). The metabolizing enzyme (a glyphosate oxidase) needs to hydrolyze a secondary amine at one of the C-N bonds (the other metabolite, sacrinose, also requires the same type of reaction but on the other side of the N) (McVetty and Zelmer. 2007). To date there have been no documented cases of glyphosate resistance due solely to glyphosate metabolism. Monsanto tried to optimize a glyphosate oxidase to confer resistance (see above) but were never able to get useful levels of resistance (McVetty and Zelmer. 2007). One population of *Digitaria isularis* (sour grass) growing in Brazil was found to have glyphosate metabolism as a contribution to overall observed resistance; however, this population also had a P106T target site mutation and altered glyphosate translocation (de Carvalho *et al*. 2011). Another population of *Conyza canadensis* (Canadian horse weed) has glyphosate metabolism pathways to AMPA and to sacrinose along with altered translocation (González-Torralva *et al*. 2012). In other populations where metabolism has been investigated, no differences between susceptible and resistant plants have been found (Cruz-Hipolito *et al*.; Feng *et al*. 1999; Lorraine-Colwill *et al*. 2002).

Amaranths as a Model of Glyphosate Resistance

Amaranthus is a large genus in the family *Amaranthaceae*, the largest within the order *Caryophyllales*. The genus contains approximately seventy species; most are native to

tropical and subtropical zones of Central and South America with many species present nearly world-wide as introduced ruderals or weeds. They are well adapted to early colonization of disturbed habitats (including freshly plowed agricultural fields). They are highly competitive and grow well at high temperatures and high light levels due to their C4 carbon metabolism (G. Kadereit *et al*. 2003). Female plants produce large numbers of seeds —over 500,000 for some species—that can disperse long distances and remain dormant for many years (Stevens. 1957); male plants produce pollen that is similarly capable of wide dispersal (Sosnoskie *et al*. 2009). While a few species are minor vegetable, pseudo-cereal crops, and ornamentals (Brenner *et al*. 2000; Mosyakin and Robertson. 1993), many more species are considered weeds. Worldwide seventeen species are classified as weeds and twelve of these weedy species have at least one population with resistance to at least one herbicide (Heap. 2013; WSSA Standardized Plant Names Subcommittee. 2010). Three species, *A. palmeri, A. spinosus,* and *A. tuberculatus,* have multiple populations documented as resistant to glyphosate. These three species are also responsible for most of the amaranth infestations in the US; in the Southeastern US *A. palmeri* represents the majority of the glyphosate resistant amaranth problem. Herbicide resistant *A. palmeri* costs growers in the US billions in increased management costs and lost yield (Burke *et al*. 2007; Gaylon D. Morgan *et al*. 2001; Klingaman and Oliver. 1994; Massinga *et al*. 2001; Rowland *et al*. 1999).

In addition to the practical concerns regarding glyphosate resistance in *Amaranthus,* the system has several advantages to study from a scientific perspective. Most importantly,

amaranths are not congeneric with the crops they infest. Amaranth infests many crops in the US: peanut, corn, cotton, and soy (Burke *et al*. 2007; Gaylon D. Morgan *et al*. 2001; Klingaman and Oliver. 1994; Massinga *et al*. 2001; Rowland *et al*. 1999), but they cannot hybridize with any of these species. There are no major crops in *Amaranthoideae*; the closest crop relatives are spinach (*Spinacia oleracea*), sugar beet (*Beta vulgaris*), and quinoa (*Chenopodium quinoa*), none of which are major crops in the US. This is important because in systems with congeneric crops and weeds (i.e. *Oryza*) adaptive processes are complicated by the potential of the weed to hybridize with the crop and share adaptive alleles.

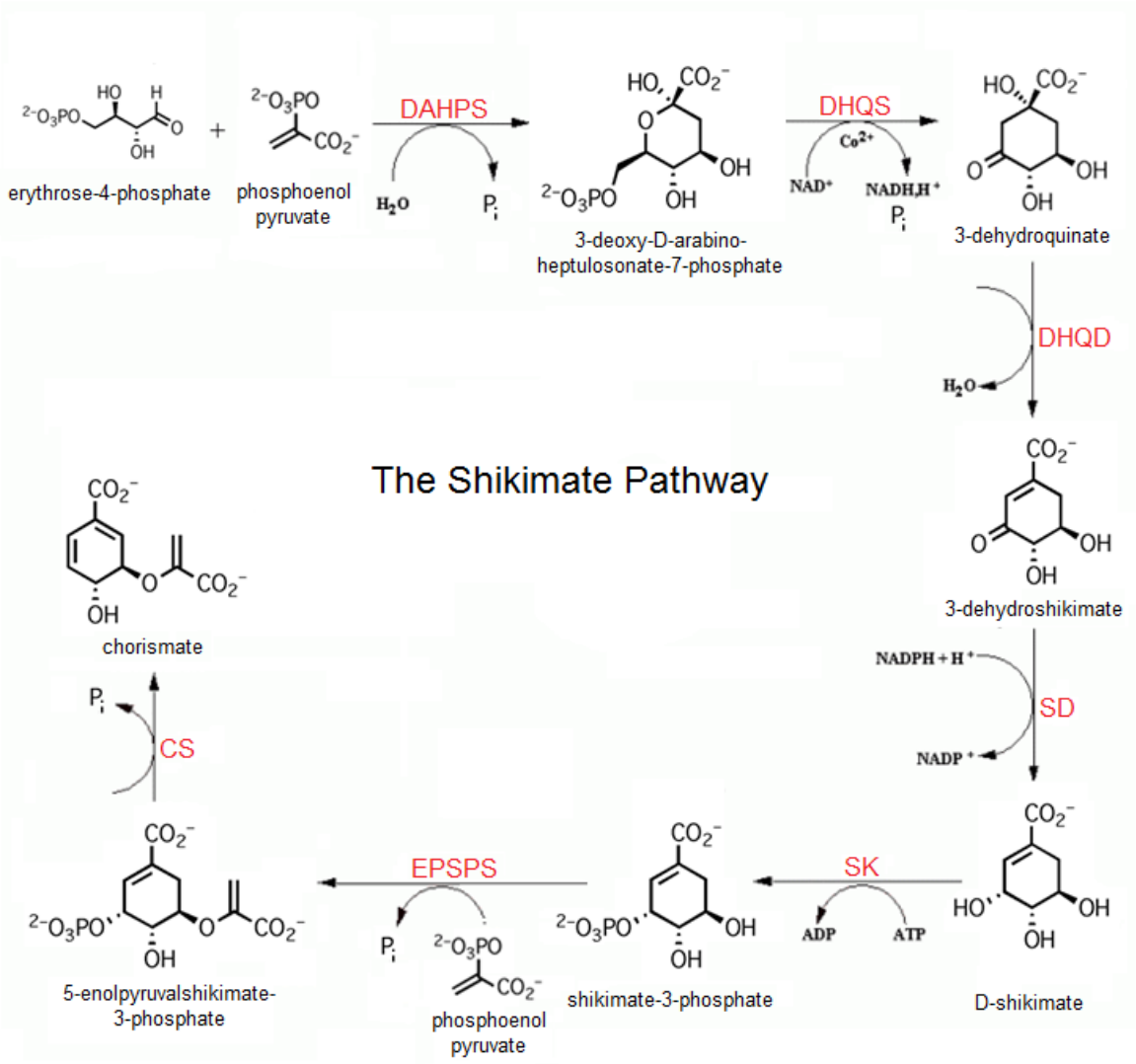
Another advantage of amaranths is that they have moderate size diploid genomes with 2n = 32 or 34 chromosomes (Costea *et al*. 2004). The genome sizes range is 937.65Mbp to 1.36Gbp (Jeschke *et al*. 2009; Rayburn *et al*. 2005). For comparison, *Arabidopsis thaliana* has a 157Mbp genome and the mouse has a 2.7Gbp genome. This makes the genome size comparable to other frequently studied systems. Additionally, the recent debut and rapid spread of resistance suggests that the adaptations to glyphosate stress are very recent, which permits a unique opportunity to investigate ongoing adaptive evolution.

The first glyphosate resistant amaranth population, *Amaranthus palmeri* growing in Macon, GA, was documented in 2004 (Culpepper *et al*. 2006). The mechanism of resistance was determined to be increases in *EPSPS* copy number (Gaines. 2010; Gaines

et al. 2011; Gaines *et al*. 2013). From there resistance has spread: 2005 North Carolina; 2006 Arkansas and Tennessee; 2007 New Mexico; 2008 Alabama and Missouri; 2010 Louisiana, Illinois, and Ohio; 2011 Michigan, Virginia, and Kansas; 2012 California, Arizona, and Delaware (Heap. 2013). Resistant *A. palmeri* are now found in over 150,000 sq mi of the Southeastern United States alone (William Vencill, personal communication) plus more spread throughout the country.

This rapid spread of glyphosate resistant weeds presents an expensive problem for agriculture and a unique opportunity for increased scientific understanding of evolution. *A. palmeri* resistance to glyphosate represents a system that is very quickly adapting to a strong, well understood selection pressure in the natural environment. The origin of resistance in *A. palmeri*, the most prolific weed currently in the Southern US, is also extremely important because of the burgeoning resistance in other *Amaranthus* species through independent events or shared sources of adaptive variation. A better understanding of the dynamics at work in this case will contribute greatly to an overall better understanding of adaptive evolution, and potentially inform weed management practices that will preserve the utility of herbicides, ensure food security, and help minimize the negative impact of herbicides on human health and the environment.

Figure I.1: The shikimate pathway. The shikimate pathway is a seven step pathway that catalyzes the formation of chorismate from erythrose-4-phosphate and phosphoenol pyruvate. The seven enzymes are, in order, 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase [EC 4.1.2.15], 3-dehydroquinate (DHQ) synthase [EC 4.2.3.4], DHQ dehydratase [EC 4.2.1.10], shikimate dehydrogenase [EC 1.1.1.25], shikimate kinase [EC 2.7.1.71], 5-enolpyruvalshikimate-3-phosphate (EPSP) synthase [EC 2.5.1.19], and chorismate synthase [EC 4.2.3.5]. Chorismate is a major branch point in carbon metabolism and a precursor to the aromatic amino acids (phenylalnine, tryptophan, and tyrosine), salicylic acid, indole derivatives and alkaloids (plant defense compounds).


The Shikimate Pathway
erythrose-4-phosphate
phosphoenol pyruvate
DAHPS
3-deoxy-D-arabino-heptulosonate-7-phosphate
DHQS
3-dehydroquinate
DHQD
3-dehydroshikimate
SD
D-shikimate
SK
shikimate-3-phosphate
EPSPS
5-enolpyruvalshikimate-3-phosphate
CS
chorismate

References

Alcala, N., D. Streit, J. Goudet and S. Vuilleumier, 2013 Peak and persistent excess of genetic diversity following an abrupt migration increase. Genetics **193:** 953-971.

Al-Khatib, K., J. R. Baumgartner, D. E. Peterson and R. S. Currie, 1998 Imazethapyr resistance in common sunflower (*Helianthus annuus*). Weed Science **46:** pp. 403-407.

Anderson, T. M., S. I. Candille, M. Musiani, C. Greco, D. R. Stahler *et al*, 2009 Molecular and evolutionary history of melanism in north American gray wolves. Science **323:** 1339-1343.

Anthony, R. G., T. R. Waldin, J. A. Ray, S. W. J. Bright and P. J. Hussey, 1998 Herbicide resistance caused by spontaneous mutation of the cytoskeletal protein tubulin. Nature **393:** 260-263.

Arias, R. S., F. E. Dayan, A. Michel, J. Howell and B. E. Scheffler, 2006 Characterization of a higher plant herbicide-resistant phytoene desaturase and its use as a selectable marker. Plant Biotechnology Journal **4:** 263-273.

Baerson, S. R., D. J. Rodriguez, M. Tran, Y. Feng, N. A. Biest *et al*, 2002 Glyphosate-resistant goosegrass. identification of a mutation in the target enzyme 5-enolpyruvylshikimate-3-phosphate synthase. Plant Physiology **129:** 1265-1275.

Bakkali, Y., J. P. Ruiz-Santaella, M. D. Osuna, J. Wagner, A. J. Fischer *et al*, 2007 Late watergrass (*Echinochloa phyllopogon*): Mechanisms involved in the resistance to fenoxaprop-p-ethyl. Journal of Agricultural Food Chemistry **55:** 4052-4058.

Barrett, R. D. H., and D. Schluter, 2008 Adaptation from standing genetic variation. Trends in Ecology & Evolution **23:** 38-44.

Beckie, H. J., and F. J. Tardif, 2012 Herbicide cross resistance in weeds. Crop Protection **35:** 15-28.

Beckie, H. J., S. I. Warwick and C. A. Sauder, 2012 Basis for herbicide resistance in canadian populations of wild oat (*Avena fatua*). Weed Science **60:** 10-18.

Beckie, H. J., L. M. Hall, F. J. Tardif and G. Seguin-Swartz, 2007 Acetolactate synthase inhibitor-resistant stinkweed (*Thlaspi arvense* L.) in Alberta. Canadian Journal of Plant Science **87:** 965-972.

Bentley, R., and E. Haslam, 1990 The shikimate pathway-A metabolic tree with many branches. Critical Reviews in Biochemistry and Molecular Biology **25:** 307-384.

Boutsalis, P., J. Karotam and S. B. Powles, 1999 Molecular basis of resistance to acetolactate synthase-inhibiting herbicides in sisymbrium orientale and *Brassica tournefortii.* Pesticide Science **55:** 507-516.

Bravin, F., G. Zanin and C. Preston, 2001 Resistance to diclofop-methyl in two *Lolium* spp. populations from italy: Studies on the mechanism of resistance. Weed Research **41:** 461-473.

Brenner, D. M., D. D. Baltensperger, P. A. Kulakow, J. W. Lehmann, R. L. Myers *et al*, 2000 Genetic resources and breeding of Amaranthus*. Plant Breeding Research **19:** 227-285.

Brown, A. C., S. R. Moss, Z. A. Wilson and L. M. Field, 2002 An isoleucine to leucine substitution in the ACCase of *Alopecurus myosuroides* (black-grass) is associated with resistance to the herbicide sethoxydim. Pesticide Biochemistry and Physiology **72:** 160-168.

Burke, I. C., M. Schroeder, W. E. Thomas and J. W. Wilcut, 2007 Palmer amaranth interference and seed production in peanut. Weed Technology **21:** 367-371.

Burnet, M. W. M., B. R. Loveys, J. A. M. Holtum and S. B. Powles, 1993a Increased detoxification is a mechanism of simazine resistance in *Lolium rigidum*. Pesticide Biochemistry and Physiology **46:** 207-218.

Burnet, M. W. M., B. R. Loveys, J. A. M. Holtum and S. B. Powles, 1993b A mechanism of chlorotoluron resistance in *Lolium rigidum*. Planta **190:** 182-189.

Carvalho, L. B., F. González-Torralva, H. Cruz Hipolito, A. M. Rojano-Delgado, J. Gil-Humanes *et al*, 2011 A pool of resistance mechanisms to glyphosate in *Digitaria insularis*. Journal of Agricultural Food Chemistry .

Catania, F., M. O. Kauer, P. J. Daborn, J. L. Yen, R. H. French-Constant *et al*, 2004 World-wide survey of an accord insertion and its association with DDT resistance in *Drosophila melanogaster*. Molecular Ecology **13:** 2491-2504.

Christoffers, M. J., M. L. Berg and C. G. Messersmith, 2002 An isoleucine to leucine mutation in acetyl-CoA carboxylase confers herbicide resistance in wild oat. Genome **45:** 1049-1056.

Christopher, J. T., S. B. Powles, D. R. Liljegren and J. A. M. Holtum, 1991 Cross-resistance to herbicides in annual ryegrass (*Lolium rigidum*) .2. chlorsulfuron resistance involves a wheat-like detoxification system. Plant Physiology **95:** 1036-1043.

Collavo, A., S. Panozzo, G. Lucchesi, L. Scarabel and M. Sattin, 2011 Characterisation and management of *Phalaris paradoxa* resistant to ACCase-inhibitors. Crop Protection **30:** 293-299.

Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro *et al*, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS biology **2:** e109.

Costea, M., S. E. Weaver and F. J. Tardif, 2004 The biology of canadian weeds. 130. *Amaranthus retroflexus* L., *A. powellii* S. watson and *A. hybridus* L. Canadian journal of plant science **84:** 631-668.

Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al*, 2004 Parallel genetic basis for repeated evolution of armor loss in alaskan threespine stickleback populations. Proceedings of the National Academy of Sciences of the United States of America **101:** 6050-6055.

Cruz-Hipolito, H., M. D. Osuna, J. A. Domínguez-Valenzuela, N. Espinoza and R. De Prado, 2011 Mechanism of resistance to ACCase-inhibiting herbicides in wild oat (*Avena fatua*) from latin america. J. Agricultural Food Chemistry.

Cruz-Hipolito, H., A. Rojano-Delgado, J. A. Domínguez-Valenzuela, A. Heredia, M. D. L. de Castro *et al*, Glyphosate tolerance by *Clitoria ternatea* and *Neonotonia wightii* plants involves differential absorption and translocation of the herbicide. Plant Soil 1-10.

Cruz-Hipolito, H., M. D. Osuna, R. A. Vidal and R. De Prado, 2009 Resistance mechanism to bensulfuron-methyl in biotypes of *Scirpus mucronatus* L. collected in chilean rice fields. Journal of Agricultural Food Chemistry **57:** 4273-4278.

Cui, H. L., C. X. Zhang, S. H. Wei, H. J. Zhang, X. J. Li *et al*, 2011 Acetolactate synthase gene pro line (197) mutations confer tribenuron-methyl resistance in flixweed (*Descurainia sophia*) populations from china. Weed Science **59:** 376-379.

Cui, H. L., C. X. Zhang, H. J. Zhang, X. Liu, Y. Liu *et al*, 2008 Confirmation of flixweed (*Descurainia sophia*) resistance to tribenuron in china. Weed Science **56:** 775-779.

Culpepper, A., J. Whitaker, A. MacRae and A. York, 2008 Distribution of glyphosate-resistant palmer amaranth (*Amaranthus palmeri*) in georgia and north carolina during 2005 and 2006. Journal of Cotton Science **12:** 306-310.

Culpepper, A. S., T. L. Grey, W. K. Vencill, J. M. Kichler, T. M. Webster *et al*, 2006 Glyphosate-resistant palmer amaranth (*Amaranthus palmeri*) confirmed in georgia. Weed Science **54:** 620-626.

Cummins, I., D. J. Cole and R. Edwards, 1999 A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black-grass. Plant Journal **18:** 285-292.

Cummins, I., S. Moss, D. J. Cole and R. Edwards, 1997 Glutathione transferases in herbicide-resistant and herbicide-susceptible black-grass (*Alopecurus myosuroides*). Pesticide Science **51:** 244-250.

Darwin, Charles. *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*. Ed. William F. Bynum. AL Burt, 2009.

Dayan, F. E., P. R. Daga, S. O. Duke, R. M. Lee, P. J. Tranel *et al*, 2010 Biochemical and structural consequences of a glycine deletion in the alpha-8 helix of protoporphyrinogen oxidase. Biochimica et Biophysica Acta-Proteins and Proteomics **1804:** 1548-1556.

de Carvalho, L. B., H. Cruz-Hipolito, F. González-Torralva, da Costa Aguiar Alves, P.L., P. J. Christoffoleti *et al*, 2011 Detection of sourgrass (*Digitaria insularis*) biotypes resistant to glyphosate in brazil. Weed Science **59:** 171-176.

Delye, C., T. Wang and H. Darmency, 2002 An isoleucine-leucine substitution in chloroplastic acetyl-CoA carboxylase from green foxtail (*Setaria viridis* L. beauv.) is responsible for resistance to the cyclohexanedione herbicide sethoxydim. Planta **214:** 421-427.

Délye, C., X. Q. Zhang, S. Michel, A. Matéjicek and S. B. Powles, 2005 Molecular bases for sensitivity to acetyl-coenzyme A carboxylase inhibitors in black-grass. Plant Physiology **137:** 794-806.

Delye, C., and K. Boucansaud, 2008 A molecular assay for the proactive detection of target site-based resistance to herbicides inhibiting acetolactate synthase in *Alopecurus myosuroides*. Weed Research **48:** 97-101.

Delye, C., K. Boucansaud, F. Pernin and V. Le Corre, 2009 Variation in the gene encoding acetolactate-synthase in Lolium species and proactive detection of mutant, herbicide-resistant alleles. Weed Research **49:** 326-336.

Delye, C., Y. Menchari, S. Michel and H. Darmency, 2004 Molecular bases for sensitivity to tubulin-binding herbicides in green foxtail. Plant Physiology **136:** 3920-3932.

Délye, C., X. Zhang, C. Chalopin, S. Michel and S. B. Powles, 2003 An isoleucine residue within the carboxyl-transferase domain of multidomain acetyl-coenzyme A carboxylase is a major determinant of sensitivity to aryloxyphenoxypropionate but not to cyclohexanedione inhibitors. Plant Physiology **132:** 1716-1723.

Devine, M. D., and A. Shukla, 2000 Altered target sites as a mechanism of herbicide resistance. Crop Protection **19:** 881-889.

Devonshire, A. L., and L. M. Field, 1991 Gene amplification and insecticide resistance. Annual Reviews in Entomology **36:** 1-21.

Donn, G., E. Tischer, J. Smith and H. Goodman, 1984 Herbicide-resistant alfalfa cells: An example of gene amplification in plants. Journal of Molecular and Applied Genetics **2:** 621.

Dowling, T. E., D. P. Martasian and W. R. Jeffery, 2002 Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. Molecular Biology and Evolution **19:** 446-455.

Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow, 1998 Rates of spontaneous mutation. Genetics **148:** 1667-1686.

Duran-Prado, M., M. D. Osuna, R. De Prado and A. R. Franco, 2004 Molecular basis of resistance to sulfonylureas in *Papaver rhoeas*. Pesticide Biochemistry and Physiology **79:** 10-17.

Eizirik, E., N. Yuhki, W. E. Johnson, M. Menotti-Raymond, S. S. Hannah *et al*, 2003 Molecular genetics and evolution of melanism in the cat family. Current Biology **13:** 448-453.

Evans, P. D., N. Mekel-Bobrov, E. J. Vallender, R. R. Hudson and B. T. Lahn, 2006 Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic homo lineage. Proceedings of the National Academy of Sciences **103:** 18178-18183.

Fay, J. C., and C. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405-1413.

Feldman, C. R., E. D. Brodie, E. D. Brodie and M. E. Pfrender, 2009 The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. Proceedings of the National Academy of Sciences **106:** 13415-13420.

Feng, P. C. C., M. Tran, T. Chiu, R. D. Sammons, G. R. Heck *et al*, 2009 Investigations into glyphosate-resistant horseweed (*Conyza canadensis*): Retention, uptake, translocation, and metabolism. Weed Science: **52**: 498-505.

Feng, P. C. C., J. E. Pratley and J. A. Bohn, 1999 Resistance to glyphosate in *Lolium rigidum*. II. uptake, translocation, and metabolism. Weed Science **47:** 412-415.

Fisher, R., 1930 The genetical the ory of natural selection. Clarendon, Oxford 19.

Franz, J. E., M. K. Mao and J. A. Sikorski, 1997 *Glyphosate: A Unique Global Herbicide.* American Chemical Society.

Funke, T., H. Han, M. L. Healy-Fried, M. Fischer and E. Schönbrunn, 2006 Molecular basis for the herbicide resistance of roundup ready crops. Proceedings of the National Academy of Sciences **103:** 13010-13015.

Funke, T., Y. Yang, H. Han, M. Healy-Fried, S. Olesen *et al*, 2009 Structural basis of glyphosate resistance resulting from the double mutation Thr97 → ile and Pro101 → ser in 5-enolpyruvylshikimate-3-phosphate synthase from *Escherichia coli.* Journal of Biological Chemistry **284:** 9854-9860.

G. Kadereit, T. Borsch, K. Weising and H. Freitag, 2003 Phylogeny of *amaranthaceae* and *chenopodiaceae* and the evolution of C4 photosynthesis. International Journal of Plant Science **164:** 959-986.

Gaines, T. A., 2010 *Molecular Genetics of Glyphosate Resistance in Palmer Amaranth (Amaranthus Palmeri L.)*. Colorado State University.

Gaines, T. A., D. L. Shaner, S. M. Ward, J. E. Leach, C. Preston *et al*, 2011 Mechanism of resistance of evolved glyphosate-resistant palmer amaranth (*Amaranthus palmeri*). Journal of Agricultural Food Chemistry **59:** 5886-5889.

Gaines, T. A., A. A. Wright, W. T. Molin, L. Lorentz, C. W. Riggins *et al*, 2013 Identification of genetic elements associated with EPSPS gene amplification. PloS One **8:** e65819.

Gaines, T. A., W. Zhang, D. Wang, B. Bukun, S. T. Chisholm *et al*, 2010 Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proceedings of the National Academy of Sciences **107:** 1029-1034.

Gaylon D. Morgan, P. A. Baumann and J. M. Chandler, 2001 Competitive impact of palmer amaranth (*Amaranthus palmeri*) on cotton (*Gossypium hirsutum*) development and yield. Weed Technology **15:** 408-412.

GimenezEspinosa, R., E. Romera, M. Tena and R. DePrado, 1996 Fate of atrazine in treated and pristine accessions of three *Setaria* species. Pesticide Biochemistry and Physiology **56:** 196-207.

Goldsbrough, P. B., E. M. Hatch, B. Huang, W. G. Kosinski, W. E. Dyer *et al*, 1990 Gene amplification in glyphosate tolerant tobacco cells. Plant Science **72:** 53-62.

González-Torralva, F., A. M. Rojano-Delgado, M. D. Luque de Castro, N. Mülleder and R. De Prado, 2012 Two non-target mechanisms are involved in glyphosate-resistant horseweed (*Conyza canadensis* L. cronq.) biotypes. Journal of Plant Physiology 1673.

Gressel, J., 2011 Low pesticide rates may hasten the evolution of resistance by increasing mutation frequencies. Pest Management Science **67:** 253-257.

Guttieri, M. J., C. V. Eberlein and D. C. Thill, 1995 Diverse mutations in the acetolactate synthase gene confer chlorsulfuron resistance in kochia (*Kochia scoparia*) biotypes. Weed Science **43:** 175-178.

Guttieri, M. J., C. V. Eberlein, C. A. Mallorysmith, D. C. Thill and D. L. Hoffman, 1992 Dna-sequence variation in domain a of the acetolactate synthase genes of herbicide-resistant and herbicide-susceptible weed biotypes. Weed Science **40:** 670-676.

Haldane, J., 1937 The effect of variation of fitness. The American Naturalist 337-349.

Haldane, J. B. S., 1927 *A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation*. Cambridge University Press.

Hawks, D., Y. Xia, and E. Zurek, *National AgriculturalStatistics Service (NASS): Agricultural Chemical Use Database*. Online. Internet. Wednesday, November 06, 2013. Available http://www.pestmanagement.info/nass/

Healy-Fried, M. L., T. Funke, M. A. Priestman, H. Han and E. Schönbrunn, 2007 Structural basis of glyphosate tolerance resulting from mutations of Pro101 in escherichia coli 5-enolpyruvylshikimate-3-phosphate synthase. Journal of Biological Chemistry **282:** 32949-32955.

Heap, I. M., *International Survey of Herbicide Resistant Weeds*. Online. Internet. Wednesday, November 06, 2013 . Available www.weedscience.com

Hedrick, P. W., 2013 Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology .

Hembree, B., 2/25/2011 *Herbicide Resistant Weeds Costing Farmers Millions in Lost Yield, Increased Expense*.

Hermisson, J., and P. S. Pennings, 2005 Soft sweeps molecular population genetics of adaptation from standing genetic variation. Genetics **169:** 2335-2352.

Hochberg, O., M. Sibony and B. Rubin, 2009 The response of ACCase-resistant *Phalaris paradoxa* populations involves two different target site mutations. Weed Research **49:** 37-46.

Hodgkinson, A., and A. Eyre-Walker, 2011 Variation in the mutation rate across mammalian genomes. Nature Reviews Genetics **12:** 756-766.

Imaizumi, T., G. X. Wang, T. Ohsako and T. Tominaga, 2008 Genetic diversity of sulfonylurea-resistant and -susceptible monochoria vaginalis populations in japan. Weed Research **48:** 187-196.

Innan, H., and Y. Kim, 2004 Pattern of polymorphism after strong artificial selection in a domestication event. Proceedings of the National Academy of Sciences of the United States of America **101:** 10667-10672.

Intanon, S., A. Perez-Jones, A. G. Hulting and C. A. Mallory-Smith, 2011 Multiple pro(197) ALS substitutions endow resistance to ALS inhibitors within and among mayweed chamomile populations. Weed Science **59:** 431-437.

Jeschke, M. R., P. J. Tranel and A. L. Rayburn, 2009 DNA content analysis of smooth pigweed (*Amaranthus hybridus*) and tall waterhemp (*A. tuberculatus*): Implications for hybrid detection.

Jin, T., J. Liu, Z. Huan, C. Wu, Y. Bi *et al*, 2011 Molecular basis for resistance to tribenuron in shepherd's purse (*Capsella bursa-pastoris* (L.) medik.). Pesticide Biochemistry and Physiology **100:** 160-164.

Kaloumenos, N. S., V. N. Adamouli, C. A. Dordas and I. G. Eleftherohorinos, 2011 Corn poppy (*Papaver rhoeas*) cross-resistance to ALS-inhibiting herbicides. Pest Management Science **67:** 574-585.

Kaloumenos, N. S., C. A. Dordas, G. C. Diamantidis and I. G. Eleftherohorinos, 2009 Multiple pro(197) substitutions in the acetolactate synthase of corn poppy (papaver rhoeas) confer resistance to tribenuron. Weed Science **57:** 362-368.

Kaundun, S. S., 2010 An aspartate to glycine change in the carboxyl transferase domain of acetyl CoA carboxylase and non-target-site mechanism (s) confer resistance to ACCase inhibitor herbicides in a Lolium multiflorum population. Pest Management Science **66:** 1249-1256.

Kaundun, S. S., R. P. Dale, I. A. Zelaya, G. Dinelli, I. Marotti *et al*, 2011 A novel P106L mutation in EPSPS and an unknown mechanism (s) act additively to confer resistance to glyphosate in a south african Lolium rigidum population. Journal of Agricultural Food Chemistry .

Kays, R., A. Curtis and J. J. Kirchman, 2010 Rapid adaptive evolution of northeastern coyotes via hybridization with wolves. Biology Letters **6:** 89-93.

Kemp, M. S., S. R. Moss and T. H. Thomas, 1990 Herbicide resistance in *alopecurus myosuroides*, pp. 376-393 in *Managing Resistance to Agrochemicals*. American Chemical Society.

Kenneth D. Whitney, Rebecca A. Randell and Loren H. Rieseberg, 2006 Adaptive introgression of herbivore resistance traits in the weedy sunflower helianthus annuus. The American Naturalist **167:** 794-807.

Kern, A. J., T. M. Myers, M. Jasieniuk, B. G. Murray, B. D. Maxwell *et al*, 2002 Two recessive gene inheritance for triallate resistance in avena fatua L. Journal of Heredity **93:** 48-50.

Kimura, M., 1984 *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Klingaman, T. E., and L. R. Oliver, 1994 Palmer amaranth (*Amaranthus palmeri*) interference in soybeans (*Glycine max*). Weed Science **42:** 523-527.

Kolkman, J. M., M. B. Slabaugh, J. M. Bruniard, S. Berry, B. S. Bushman *et al*, 2004 Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. Theor. Appl. Genet. **109:** 1147-1159.

Krysiak, M., S. W. Gawroński, K. Adamczewski and R. Kierzek, 2011 ALS gene mutations in apera spica-venti confer broad-range resistance to herbicides. Journal of Plant Protection Research **51:** 261-267.

Letouze, A., and J. Gasquez, 2003 Enhanced activity of several herbicide-degrading enzymes: A suggested mechanism responsible for multiple resistance in blackgrass (*alopecurus myosuroides* huds.). Agronomie **23:** 601-608.

Letouze, A., and J. Gasquez, 2001 Inheritance of fenoxaprop-P-ethyl resistance in a blackgrass (*alopecurus myosuroides* huds.) population. Theor. Appl. Genet. **103:** 288-296.

Liu, W., D. K. Harrison, D. Chalupska, P. Gornicki, C. C. O'Donnell *et al*, 2007 Single-site mutations in the carboxyltransferase domain of plastid acetyl-CoA carboxylase confer resistance to grass-specific herbicides. Proceedings of the National Academy of Sciences **104:** 3627-3632.

Lorraine-Colwill, D. F., S. B. Powles, T. R. Hawkes, P. H. Hollinshead, S. A. J. Warner *et al*, 2002 Investigations into the mechanism of glyphosate resistance in Lolium rigidum. Pesticide Biochemistry and Physiology **74:** 62-72.

Makova, K. D., and W. Li, 2002 Strong male-driven evolution of DNA sequences in humans and apes. Nature **416:** 624-626.

Maneechote, C., C. Preston and S. B. Powles, 1997 A diclofop-methyl-resistant avena sterilis biotype with a herbicide-resistant acetyl-coenzyme A carboxylase and enhanced metabolism of diclofop-methyl. Pesticide Science **49:** 105-114.

Marshall, R., and S. R. Moss, 2008 Characterisation and molecular basis of ALS inhibitor resistance in the grass weed alopecurus myosuroides. Weed Research **48:** 439-447.

Marshall, R., R. Hull and S. R. Moss, 2010 Target site resistance to ALS inhibiting herbicides in papaver rhoeas and stellaria media biotypes from the UK. Weed Research **50:** 621-630.

Martin, N. H., A. C. Bouck and M. L. Arnold, 2006 Detecting adaptive trait introgression between *iris fulva* and *I. brevicaulis* in highly selective field conditions. Genetics **172:** 2481-2489.

Massa, D., B. Krenz and R. Gerhards, 2011 Target-site resistance to ALS-inhibiting herbicides in *apera spica-venti* populations is conferred by documented and previously unknown mutations. Weed Research **51:** 294-303.

Massinga, R. A., R. S. Currie, M. J. Horak and J. Boyer Jr., 2001 Interference of palmer amaranth in corn. Weed Science **49:** 202-208.

McCloskey, W. B., and J. S. Holt, 1990 Triazine resistance in *Senecio vulgaris* parental and nearly isonuclear backcrossed biotypes is correlated with reduced productivity. Plant Physiology **92:** 954.

McVetty, P. B. E., and C. D. Zelmer, 2007 Breeding Herbicide-Tolerant oilseed rape cultivars. Rapeseed breeding **45:** 233.

Menchari, Y., B. Chauvel, H. Darmency and C. Delye, 2008 Fitness costs associated with three mutant acetyl-coenzyme A carboxylase alleles endowing herbicide resistance in black-grass *alopecurus myosuroides*. J. Appl. Ecol. **45:** 939-947.

Menendez, J., and R. DePrado, 1996 Diclofop-methyl cross-resistance in a chlorotoluron-resistant biotype of *alopecurus myosuroides*. Pesticide Biochemistry and Physiology **56:** 123-133.

Michel, A., R. S. Arias, B. E. Scheffler, S. O. Duke, M. Netherland *et al*, 2004 Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (*hydrilla verticillata*). Molecular Ecology **13:** 3229-3237.

Michitte, P., R. De Prado, N. Espinoza, J. Pedro Ruiz-Santaella and C. Gauvrit, 2007 Mechanisms of resistance to glyphosate in a ryegrass (*Lolium multiflorum*) biotype from chile. Weed Science **55:** 435-440.

Michitte, P., R. De Prado, N. Espinosa and C. Gauvrit, 2005 Glyphosate resistance in a chilean *Lolium multiflorum*. Communications in Agcultral and Applied Biological Sciences **70:** 507-513.

Morrison, I. N., B. G. Todd and K. M. Nawolsky, 1989 Confirmation of trifluralin-resistant green foxtail (*setaria-viridis*) in Manitoba. Weed Technology **3:** 544-551.

Mosyakin, S. L., and K. R. Robertson, 1993 Amaranthus, pp. 410-435 in *Flora of North America North of Mexico*, edited by Flora of North America Editorial Committee. New York and Oxford.

Mundy, N. I., N. S. Badcock, T. Hart, K. Scribner, K. Janssen *et al*, 2004 Conserved genetic basis of a quantitative plumage trait involved in mate choice. Science **303:** 1870-1873.

Nafziger, E. D., J. M. Widholm, H. C. Steinrücken and J. L. Killmer, 1984 Selection and characterization of a carrot cell line tolerant to glyphosate. Plant Physiology **76:** 571.

Nei, M., Y. Suzuki and M. Nozawa, 2010 The neutral theory of molecular evolution in the genomic era. Annual review of genomics and human genetics **11:** 265-289.

Ng, C. H., R. Wickneswari, S. Salmijah, Y. T. Teng and B. S. Ismail, 2003 Gene polymorphisms in glyphosate-resistant and -susceptible biotypes of eleusine indica from malaysia. Weed Research **43:** 108-115.

Norsworthy, J. K., N. R. Burgos and L. R. Oliver, 2001 Differences in weed tolerance to glyphosate involve different mechanisms 1. Weed Technology **15:** 725-731.

Odonovan, J. T., M. P. Sharma, K. N. Harker, D. Maurice, M. N. Baig *et al*, 1994 Wild oat (*avena-fatua*) populations resistant to triallate are also resistant to difenzoquat. Weed Science **42:** 195-199.

Ohsako, T., and T. Tominaga, 2007 Nucleotide substitutions in the acetolactate synthase genes of sulfonylurea-resistant biotypes of *monochoria vaginalis* (pontederiaceae). Genes Genet. Syst. **82:** 207-215.

Ohta, T., 1973 Slightly deleterious mutant substitutions in evolution. Nature **246:** 96-98.

Ohta, T., and J. H. Gillespie, 1996 Development of neutral and nearly neutral theories. Theor. Popul. Biol. **49:** 128-142.

Olson-Manning, C. F., M. R. Wagner and T. Mitchell-Olds, 2012 Adaptive evolution: Evaluating empirical support for theoretical predictions. Nature Reviews Genetics **13:** 867-877.

Padgette, S. R., D. B. Re, G. F. Barry, D. E. Eichholtz, X. Delannay *et al*, 1996 New weed control opportunities: Development of soybeans with a roundup ready gene. Herbicide Resistant Crops 53-84.

Padgette, S. R., K. H. Kolacz, X. Delannay, D. B. Re, B. J. LaVallee *et al*, 1995 Development, identification, and characterization of a glyphosate-tolerant soybean line. Crop Science **35:** 1451-1461.

Park, K. W., and C. A. Mallory-Smith, 2004 Physiological and molecular basis for ALS inhibitor resistance in *bromus tectorum* biotypes. Weed Research **44:** 71-77.

Park, K. W., and C. A. Mallory-Smith, 2006 psbA mutation (asn(266) to thr) in *senecio vulgaris* L. confers resistance to several PSII-inhibiting herbicides. Pest Management Science **62:** 880-885.

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. Science U. S. A. **103:** 12329-12334.

Pedersen, B. P., P. Neve, C. Andreasen and S. B. Powles, 2007 Ecological fitness of a glyphosate-resistant *Lolium rigidum* population: Growth and seed production along a competition gradient. Basic Appl. Ecol. **8:** 258-268.

Pennings, P. S., and J. Hermisson, 2006a Soft sweeps II—molecular population genetics of adaptation from recurrent mutation or migration. Molecular Biology and Evolution **23:** 1076-1084.

Pennings, P. S., and J. Hermisson, 2006b Soft sweeps III: The signature of positive selection from recurrent mutation. PLoS genetics **2:** e186.

Perez-Jones, A., K. Park, N. Polge, J. Colquhoun and C. Mallory-Smith, 2007 Investigating the mechanisms of glyphosate resistance in *Lolium multiflorum*. Planta **226:** 395-404.

Petit, C., G. Bay, F. Pernin and C. Delye, 2010 Prevalence of cross-or multiple resistance to the acetyl-coenzyme A carboxylase inhibitors fenoxaprop, clodinafop and pinoxaden in black-grass (*alopecurus myosuroides* huds.) in france. Pest Management Science **66:** 168-177.

Powles, S. B., and Q. Yu, 2010 Evolution in action: Plants resistant to herbicides. Annual Review of Plant Biology **61:** 317-347.

Preston, C., 2004 Herbicide resistance in weeds endowed by enhanced detoxification: Complications for management. Weed Science **52:** 448-453.

Preston, C., L. M. Stone, M. A. Rieger and J. Baker, 2006 Multiple effects of a naturally occurring proline to threonine substitution within acetolactate synthase in two herbicide-resistant populations of *lactuca serriola*. Pesticide Biochemistry and Physiology **84:** 227-235.

Price, A. J., K. S. Balkcom, S. A. Culpepper, J. A. Kelton, R. L. Nichols *et al*, July/August 2011 Glyphosate-resistant palmer amaranth: A threat to conservation tillage. Journal of Soil and Water Conservation **66:** 265-275.

Purba, E., C. Preston and S. B. Powles, 1995 The mechanism of resistance to Paraquat is strongly temperature dependent in resistant *hordeum leporinum* link and H. glaucum steud. Planta **196:** 464-468.

Puri, A., G. E. MacDonald, F. Altpeter and W. T. Haller, 2007 Mutations in phytoene desaturase gend in fluridone-resistant hydrilla (*hydrilla verticillata*) biotypes in florida. Weed Science **55:** 412-420.

Randolph-Anderson, B. L., R. Sato, A. M. Johnson, E. H. Harris, C. R. Hauser *et al*, 1998 Isolation and characterization of a mutant protoporphyrinogen oxidase gene from *chlamydomonas reinhardtii* conferring resistance to porphyric herbicides. Plant Mol. Biol. **38:** 839-859.

Rayburn, A. L., R. McCloskey, T. C. Tatum, G. A. Bollero, M. R. Jeschke *et al*, 2005 Genome size analysis of weedy *Amaranthus* species. Crop Science **45:** 2557-2562.

Rowland, M. W., Don S. Murray and L. M. Verhalen, 1999 Full-season palmer amaranth (*Amaranthus palmeri*) interference with cotton (gossypium hirsutum). Weed Science **47:** 305-309.

Saari, L. L., J. C. Cotterman and M. M. Primiani, 1990 Mechanism of sulfonylurea herbicide resistance in the broadleaf weed, *kochia scoparia*. Plant Physiology **93:** 55.

Sage, R. F., T. L. Sage, R. W. Pearcy and T. Borsch, 2007 The taxonomic distribution of C4 photosynthesis in amaranthaceae sensu stricto. American Journal of Botany **94:** 1992-2003.

Salas, R. A., F. E. Dayan, Z. Pan, S. B. Watson, J. W. Dickson *et al*, 2012 EPSPS gene amplification in glyphosate-resistant italian ryegrass (*Lolium perenne ssp multiflorum*) from Arkansas, USA. Pest Management Science .

Scarabel, L., N. Carraro, M. Sattin and S. Varotto, 2004 Molecular basis and genetic characterisation of evolved resistance to ALS-inhibitors in *papaver rhoeas*. Plant Science **166:** 703-709.

Scarabel, L., A. Locascio, A. Furini, M. Sattin and S. Varotto, 2010 Characterisation of ALS genes in the polyploid species *schoenoplectus mucronatus* and implications for resistance management. Pest Management Science **66:** 337-344.

Schluter, D., E. A. Clifford, M. Nemethy and J. S. McKinnon, 2004 Parallel evolution and inheritance of quantitative traits. The American Naturalist **163:** 809-822.

Schönbrunn, E., S. Eschenburg, W. A. Shuttleworth, J. V. Schloss, N. Amrhein *et al*, 2001 Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proceedings of the National Academy of Sciences of the United States of America **98:** 1376-1380.

Shah, D. M., R. B. Horsch, H. J. Klee, G. M. Kishore, J. A. Winter *et al*, 1986 Engineering herbicide tolerance in transgenic plants. Science **233:** 478-481.

Shaner, D. L., 2009 Role of translocation as a mechanism of resistance to glyphosate. Weed Science **57:** 118-123.

Sibony, M., and B. Rubin, 2003 Molecular basis for multiple resistance to acetolactate synthase-inhibiting herbicides and atrazine in *Amaranthus blitoides* (prostrate pigweed). Planta **216:** 1022-1027.

Sibony, M., A. Michel, H. U. Haas, B. Rubin and K. Hurle, 2001 Sulfometuron-resistant *Amaranthus retroflexus*: Cross-resistance and molecular basis for resistance to acetolactate synthase-inhibiting herbicides. Weed Research **41:** 509-522.

Singer, S. R., and C. N. McDaniel, June 1985 Selection of glyphosate-tolerant tobacco calli and the expression of this tolerance in regenerated plants. Plant Physiology **78:** 411-416.

Singh, S., R. C. Kirkwood and G. Marshall, 1998 Effect of the monooxygenase inhibitor piperonyl butoxide on the herbicidal activity and metabolism of isoproturon in herbicide resistant and susceptible biotypes of *phalaris minor* and wheat. Pesticide Biochemistry and Physiology **59:** 143-153.

Smeda, R. J., K. C. Vaughn and I. N. Morrison, 1992 A novel pattern of herbicide cross-resistance in a trifluralin-resistant biotype of green foxtail [*setaria-viridis* (L) beauv]. Pesticide Biochemistry and Physiology **42:** 227-241.

Soar, C. J., J. Karotam, C. Preston and S. B. Powles, 2003 Reduced Paraquat translocation in Paraquat resistant *Arctotheca calendula* (L.) levyns is a consequence of the primary resistance mechanism, not the cause. Pesticide Biochemistry and Physiology **76:** 91-98.

Song, Y., S. Endepols, N. Klemann, D. Richter, F. Matuschka *et al*, 2011 Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Current Biology **21:** 1296-1301.

Sosnoskie, L. M., T. M. Webster, D. Dales, G. C. Rains, T. L. Grey *et al*, 2009 Pollen grain size, density, and settling velocity for palmer amaranth (*Amaranthus palmeri*). Weed Science **57:** 404-409.

Steiner, C. C., J. N. Weber and H. E. Hoekstra, 2007 Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS biology **5:** e219.

Steinrücken, H. C., and N. Amrhein, 1980 The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. Biochemical and Biophysical Research Communications **94:** 1207-1212.

Stevens, O. A., 1957 Weights of seeds and numbers per plant. Weeds **5:** 46-55.

Strecker, U., L. Bernatchez and H. Wilkens, 2003 Genetic divergence between cave and surface populations of *astyanax* in Mexico (characidae, teleostei). Molecular Ecology **12:** 699-710.

Tan, M. K., and R. W. Medd, 2002 Characterisation of the acetolactate synthase (ALS) gene of *raphanus raphanistrum* L. and the molecular assay of mutations associated with herbicide resistance. Plant Science **163:** 195-205.

Tardif, F. J., and S. B. Powles, 1999 Effect of malathion on resistance to soil-applied herbicides in a population of rigid ryegrass (*Lolium rigidum*). Weed Science **47:** 258-261.

Theron, E., K. Hawkins, E. Bermingham, R. E. Ricklefs and N. I. Mundy, 2001 The molecular basis of an avian plumage polymorphism in the wild: A melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *coereba flaveola*. Current Biology **11:** 550-557.

Tyekucheva, S., K. D. Makova, J. E. Karro, R. C. Hardison, W. Miller *et al*, 2008 Human-macaque comparisons illuminate variation in neutral substitution rates. Genome Biol. **9:** R76.

Uchino, A., and H. Watanabe, 2002 Mutations in the acetolactate synthase genes of sulfonylurea-resistant biotypes of *lindernia* spp. Weed Biology and Management **2:** 104-109.

Uchino, A., S. Ogata, H. Kohara, S. Yoshida, T. Yoshioka *et al*, 2007 Molecular basis of diverse responses to acetolactate synthase-inhibiting herbicides in sulfonylurea-resistant biotypes of *Schoenoplectus juncoides*. Weed Biology and Management **7:** 89-96.

Vaughn, K. C., M. D. Marks and D. P. Weeks, 1987 A dinitroaniline-resistant mutant of *Eleusine indica* exhibits cross-resistance and supersensitivity to antimicrotubule herbicides and drugs. Plant Physiology **83:** 956-964.

Wakelin, A. M., D. F. Lorraine-Colwill and C. Preston, 2004 Glyphosate resistance in four different populations of *Lolium rigidum* is associated with reduced translocation of glyphosate to meristematic zones. Weed Research **44:** 453-459.

Wakelin, A., and C. Preston, 2006 A target-site mutation is present in a glyphosate-resistant *Lolium rigidum* population. Weed Research **46:** 432-440.

Wang, G. X., Y. Lin, W. Li, M. Ito and K. Itoh, 2004 A mutation confers *Monochoria vaginalis* resistance to sulfonylureas that target acetolactate synthase. Pesticide Biochemistry and Physiology **80:** 43-46.

Warwick, S. I., R. Xu, C. Sauder and H. J. Beckie, 2008 Acetolactate synthase target-site mutations and single nucleotide polymorphism genotyping in ALS-resistant kochia (*Kochia scoparia*). Weed Science **56:** 797.

Warwick, S. I., C. Sander and H. J. Beckie, 2005 Resistance in canadian biotypes of wild mustard (*Sinapis arvensis*) to acetolactate synthase inhibiting herbicides. Weed Science **53:** 631-639.

Warwick, S. I., C. A. Sauder and H. J. Beckie, 2010 Acetolactate synthase (ALS) target-site mutations in ALS inhibitor-resistant russian thistle (*Salsola tragus*). Weed Science **58:** 244-251.

Wassom, J. J., and P. J. Tranel, 2005 Amplified fragment length polymorphism-based genetic relationships among weedy *Amaranthus* species. Journal of Heredity **96:** 410-416.

Weaver, L. M., and K. M. Herrmann, 1997 Dynamics of the shikimate pathway in plants. Trends Plant Science **2:** 346-351.

Weinberg, T., G. R. Stephenson, M. D. McLean and J. C. Hall, 2006 MCPA (4-chloro-2-ethylphenoxyacetate) resistance in hemp-nettle (*Galeopsis tetrahit* L.). Journal of Agricultural Food Chemistry **54:** 9126-9134.

Whitaker, J. R., J. D. Burton, A. C. York, D. L. Jordan and A. Chandi, 2013 Physiology of glyphosate-resistant and glyphosate-susceptible palmer amaranth (*Amaranthus palmeri*) biotypes collected from North Carolina. International Journal of Agronomy **2013:** .

Whitaker, J. R., 2009 *Distribution, Biology, and Management of Glyphosate-Resistant Palmer Amaranth in North Carolina*. ProQuest.

White, G., S. Moss and A. Karp, 2005 Differences in the molecular basis of resistance to the cyclohexanedione herbicide sethoxydim in *Lolium multiflorum*. Weed Research **45:** 440-448.

Widholm, J. M., A. R. Chinnala, J. Ryu, H. Song, T. Eggett *et al*, 2001 Glyphosate selection of gene amplification in suspension cultures of 3 plant species. Physiol. Plantarum **112:** 540-545.

Wolfe, K. H., W. Li and P. M. Sharp, 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences **84:** 9054-9058.

WSSA Standardized Plant Names Subcommittee, 2010 *January 2010 Revision of the Composite List of Weeds*. Weed Science Society of America.

Xu, F., and M. Sun, 2001 Comparative analysis of phylogenetic relationships of grain amaranths and their wild relatives (*Amaranthus*; *Amaranthaceae*) using internal transcribed spacer, amplified fragment length polymorphism, and double-primer fluorescent intersimple sequence repeat markers. Molecular Phylogenetics and Evolution **21:** 372-387.

Yamamoto, E., L. H. Zeng and W. V. Baird, 1998 Alpha-tubulin missense mutations correlate with antimicrotubule drug resistance in *Eleusine indica*. Plant Cell **10:** 297-308.

Yu, Q., A. Cairns and S. Powles, 2007a Glyphosate, Paraquat and ACCase multiple herbicide resistance evolved in a *Lolium rigidum* biotype. Planta **225:** 499-513.

Yu, Q., A. Collavo, M. Q. Zheng, M. Owen, M. Sattin *et al*, 2007b Diversity of acetyl-coenzyme A carboxylase mutations in resistant *Lolium populations*: Evaluation using clethodim. Plant Physiology **145:** 547-558.

Yu, Q., X. Q. Zhang, A. Hashem, M. J. Walsh and S. B. Powles, 2003 ALS gene proline (197) mutations confer ALS herbicide resistance in eight separated wild radish (*Raphanus raphanistrum*) populations. Weed Science **51:** 831-838.

Yu, Q., H. Han and S. B. Powles, 2008 Mutations of the ALS gene endowing resistance to ALS-inhibiting herbicides in *Lolium rigidum* populations. Pest Management Science **64:** 1229-1236.

Yu, Q., J. K. Nelson, M. Q. Zheng, M. Jackson and S. B. Powles, 2007 Molecular characterisation of resistance to ALS-inhibiting herbicides in *Hordeum leporinum* biotypes. Pest Management Science **63:** 918-927.

Yuan, J. S., L. L. G. Abercrombie, Y. Cao, M. D. Halfhill, X. Zhou *et al*, 2010 Functional genomics analysis of horseweed (*Conyza canadensis*) with special reference to the evolution of non-target-site glyphosate resistance. Weed Science **58:** 109-117.

Yun, M., Y. Yogo, R. Miura, Y. Yamasue and A. J. Fischer, 2005 Cytochrome P-450 monooxygenase activity in herbicide-resistant and -susceptible late watergrass (*Echinochloa phyllopogon*). Pesticide Biochemistry and Physiology **83:** 107-114.

Zagnitko, O., J. Jelenska, G. Tevzadze, R. Haselkorn and P. Gornicki, 2001 An isoleucine/leucine residue in the carboxyltransferase domain of acetyl-CoA carboxylase is critical for interaction with aryloxyphenoxypropionate and cyclohexanedione inhibitors. Proceedings of the National Academy of Sciences **98:** 6617-6622.

Zhang, X. Q., and S. B. Powles, 2006 Six amino acid substitutions in the carboxyl-transferase domain of the plastidic acetyl-CoA carboxylase gene are linked with resistance to herbicides in a Lolium rigidum population. New Phytol. **172:** 636-645.

Zheng, D., G. R. Kruger, S. Singh, V. M. Davis, P. J. Tranel *et al*, 2011 Cross-resistance of horseweed (*Conyza canadensis*) populations with three different ALS mutations. Pest Management Science **67:** 1486-1492.

CHAPTER I

THE PHYLOGENY OF *Amaranthus* (*Amaranthaceae*) WITH A FOCUS ON THE DISTRIBUTION OF AGRONOMICALLY IMPORTANT TRAITS

K. E. Beard, B. Rauh, N. Burgos, J. Burton, and A. Lawton-Rauh

Abstract

Premise of the study

The genus *Amaranthus* (family *Amaranthaceae*) contains some 75 species, with at least 17 identified as agricultural weeds. Many populations of weedy amaranth have evolved high levels of resistance to multiple herbicides. These herbicide resistant weeds have a large economic impact on agriculture and a large environmental impact as growers increase herbicide application rates to compensate. In spite of this economic importance, the relationship between these weedy amaranths is not currently well understood.

Methods

Genomic DNA was collected from 53 amaranth representing 31 species; from this DNA we sequenced 5 nuclear loci and the plastid gene *Maturase K* (*MatK*). These sequences were analyzed using coalescent and Bayesian methods to estimate the best-fit species phylogeny.

Key Results

The Bayesian and coalescent estimations of the phylogeny show similar results that suggest the genus may be best described as containing four sub-genera (one of which is further divided into two sections) that rapidly radiated from the common ancestor. In addition the Bayesian phylogeny shows the weedy amaranths are distributed across all four sub-genera.

Conclusions

The genus *Amaranthus* should be reorganized into four sub-genera, instead of the three currently described. Also the distribution of weedy amaranths and the habitat descriptions of the non-weedy species strongly suggest that other species may also prove to be threat to agriculture as weediness seems to evolve readily from the ruderal habit common to the genus.

Introduction

Amaranthus is a large genus in the family *Amaranthaceae*, the largest within the order *Caryophyllales*. The genus contains approximately seventy species; most species of *Amaranthus* are native to tropical and subtropical zones of Central and South America with many species present nearly world-wide as introduced ruderals or weeds. A few species are minor vegetable, pseudo-cereal crops, and ornamentals; however, many of the species of *Amaranthus* are weeds of considerable agricultural concern (Brenner *et al*. 2000; Mosyakin and Robertson. 1993). In the Southeastern United States, for example, one *A. palmeri* per meter of row results in a 50% reduction in cotton yield (Gaylon D. Morgan *et al*. 2001; Rowland *et al*. 1999). In spite of the economic importance of this genus there has been, to date, relatively little investigation of the genus as a whole. This work seeks to remedy this by using DNA sequence data to understand the relationship between *Amaranthus* species and put this phylogeny in the context of economically important traits, particularly weediness.

As of writing, seventeen species of amaranth are considered weeds, and of those twelve have populations identified as resistant to one or more herbicides (Heap. 2013; WSSA Standardized Plant Names Subcommittee. 2010). Herbicide resistance costs growers millions of dollars, in both increased spending on weed management and lost crops, annually. Because there are so many resistant species of weeds in the genus it is of vital importance to understand the relationship between these species so that we may better understand the emergence of resistance in the genus and potentially develop techniques for slowing it.

Currently the most extensive investigation of the taxonomic relationship between *Amaranthus* species is the one presented in the Flora of North America and is based on morphological character. It separates the genus into three subgenera: Acnida, Amaranthus, and Albersia. The defining traits for each of these groups respectively is dioecy, well developed terminal inflorescences, and not being a member of Acnida or Amaranthus respectively (Mosyakin and Robertson. 1993; Mosyakin and Robertson. 1996).

This phylogeny will be the first multi-gene based phylogeny of *Amaranthus*; previous phylogenies have either been of wider scope and only included one or two species of amaranth or have been based on non-sequence data. In 2001 Xu and Sun described the relationship between the cultivated *Amaranthus* (*A. caudatus, A. cruentus, A. hypochondriacus*, and *A. tricolor*) using the gene sequence of the internal transcribed

spacer (ITS), amplified fragment length polymorphism (AFLP), and inter-simple sequence repeat (ISSR) data (Xu and Sun. 2001). In 1997 a larger study with 23 species was conducted using isozyme and random amplification of polymorphic DNA (RAPD) data (Chan and Sun. 1997). In both of these studies the focus was on understanding the origin of the cultivated pseudo-cereal amaranths. It was shown that the pseudo-cereal species were monophyletic to each other and their closest wild relative/putative progenitor, *A. hybridus*. A study on weedy amaranth species used AFLP to determine the relationship of eight species (*A. palmeri, A. powellii, A. retroflexus, A. arenicola, A. hybridus, A. spinosus, A. albus,* and *A. tuberculatus*). It showed that *A. palmeri*, a dioecious species, is most closely related to *A. spinosus*, a monoecious species (Wassom and Tranel. 2005). Taken with the AFLP based phylogeny of pseudo-cereal amaranths, we also see that the pseudo-cereals group together with *A. hybridus* and *A. powellii* separate from the other weedy amaranth (Wassom and Tranel. 2005; Xu and Sun. 2001).

Amaranth have also been included in larger investigations of *Amaranthaceae*, these studies do not investigate the relationships between *Amaranthus* species, but instead the relationship of the genus to other genera in the family. Results of these studies are not consistent due to differences in sampling, but the genera most closely related to *Amaranthus* are likely *Celosia, Chamissoa,* and *Pleuropetalum* (G. Kadereit *et al*. 2003; Sage *et al*. 2007).

The purpose of this study is to investigate the genetic relationships between members of the genus *Amaranthus,* particularly with respect to agriculturally relevant traits of weediness, utilization as a crop, and dioecy.

Materials and Methods

Sampling

All samples were obtained from the USDA germplasm repository. We sampled individuals from 53 accessions representing 31 *Amaranthus* species plus 4 accessions representing 2 out-group species (Table 1.1). Seeds were planted out and grown in the greenhouse under ambient lighting until they had at least four true leaves, at which point leaf tissue was collected for DNA extraction. Tissue was frozen at -80°C and then either used directly or lyophilized for storage.

DNA extraction, amplification, and sequencing

DNA was extracted from leaf tissue in one of two forms: either lyophilized or frozen (-80°C). If lyophilized tissue was used, about 40mg was used for the extraction protocol, if frozen then about 100mg. The tissue was ground in a mixer mill and then DNA was purified using the standard protocol of the Nucleospin Plant II DNA extraction kit (Macherey Nagel, Düren, Germany) was followed. DNA was amplified using GoTaq Flexi (Promega, Madison, Wisconsin) at six loci using a unique primers pair for each locus (Table 1.2).

The cycling conditions used for all loci were as follows: 5min initial denaturation at 94°C then 35 cycles of touch-down PCR with 30sec denaturation at 94°C, 30sec annealing at 60-50°C (first cycle 60°, then each subsequent cycle 1°C lower than the previous until the cycle with a 51°C annealing temperature. Then 25 cycles each with a 50°C annealing temperature), and 3min extension at 72°C, lastly a 10min final extension at 72°C.

PCR products were electrophoresed on a 1% agarose gel to ensure quality amplification. Successful amplifications were sequenced at the Clemson University Genomics Institute using the same primers used for PCR. Before submission, PCR reactions were cleaned using an ExoAP treatment: to each 1μL of DNA we added 0.2 Units of exonuclease I (New England Biolabs, Ipswich, Massachusetts), 0.5 Units of Antarctic phosphotase (New England Biolabs, Ipswich, Massachusetts), and water to 2μL; then samples were incubated for 30min at 37°C and heat treated for 15min at 80°C.

Sequences were Phred-Phraped to merge forward and reverse sequencing reads and viewed in Biolign (Ewing and Green. 1998; Ewing *et al*. 1998; Hall. 2001). After sequencing, several sites had two base calls, suggesting heterozygosity. Most of these heterozygous sites were single nucleotide polymorphisms between the two alleles and were coded into the sequence as the appropriate IUPAC code. One sequence (*A36* of *Celosia trigina* PI649298) had an indel heterozygous site. This site was resolved by cloning the gene with the TOPO system (Life Technologies, Carlsbad, California) after PCR amplification with the high fidelity polymerase *Pfu* Ultra II (Agilent Technologies, Santa Clara, California), sequencing the clones, and selecting one sequence to use in the

alignment. The final alignment used for phylogeny estimation was created manually in Bioedit and has all six genes concatenated end-to-end (Hall. 1999).

"A" Primer Design

The four "A" primers used in this experiment were designed in our lab from 454 sequencing data generated by the Burton and Trannel labs (Lee *et al*. 2009 and unpublished data). Two separate data sets were used, a 454 sequencing of *A. palmeri* and one of *A. tuberculatus*. These two datasets were compared to each other using local BLASTn (Altschul *et al*. 1990). Then the 48 hits with the best e-value were chosen and primer pairs were designed with Primer3 to amplify 1kb regions contained within those hits (Rozen and Skaletsky. 2000). The pairs were then tested on extracted *A. palmeri* genomic DNA. Of the 48 pairs, the four chosen for the phylogeny analysis had the best amplification across species, no indel heterozygosity, and high levels of inter-specific polymorphism. An attempt was made using NCBI BLAST to identify the genes these primer pairs amplified, but all hits were to putative or predicted proteins (Table 1.2). None of these genes are suspected to be targets of herbicide selection based on the BLAST results.

Model Selection

We used MEGA 5 (Tamura *et al*. 2011) to choose the best model. MEGA estimated the best fit model under both the automatic neighbor joining tree and a user defined tree. In order to use the same mutational model for all analyses we restricted choices to models available in MrBayes, BEST, and MEGA 5. We also wanted to use the same model across all partitions since we change partitioning schemes between these three analysis programs. To choose this best all-around model we ran the model test on the whole concatenated dataset, each individual gene, and the three codon based partitions (1+2, 3, and intron). We then ranked the models by both Akaike Information Criterion (Akaike. 1974) and Bayesian Information Criterion values (Akaike. 1981). We chose GTR+G (general time reversible with a gamma distribution of rates, (Tavaré. 1986)) because it consistently ranked in the top 3 models. This is in contrast to other models, such as HKY (Hasegawa, Kishino and Yano (Hasegawa et al. 1985)) which in some data sets was one of the best-fit models but in other data sets was one of the worst fit.

Phylogeny Analysis I

The Bayesian inference phylogeny of *Amaranthus* was constructed in MrBayes (Ronquist and Huelsenbeck. 2003). The sequence data of all six genes concatenated was split into 3 partitions: 1st and 2nd codon position, 3rd codon position, and non-coding (ITS1 and ITS2). The 5.8s between ITS1 and 2 was excluded because it is neither coding nor intron and including it with the introns would skew the parameter estimates. Each partition had an independent estimate of the GTR+G rate parameters and rate multiplier. The partitions

shared a tree height and topology. The tree was run for 5,000,000 generations. After running the parameter estimates were viewed in Tracer v1.5 to determine if the run had converged, and how much data needed to be discarded as burn-in.

This same protocol was followed to generate individual gene trees except the data was partitioned by gene instead of codon position and tree height and topology were allowed to be independent between partitions to generate six gene trees.

A maximum likelihood tree was also estimated using MEGA 5 (Tamura *et al*. 2011) using 1000 bootstraps replicates to determine confidence. The six concatenated genes were used together with the 5.8s region again excluded. The GTR+G model was again used and missing data was handled with full deletion, meaning that any site where any one individual had missing or ambiguous data was ignored in the analysis.

Any bifurcations of the phylogenies with less than 75% posterior or bootstrap support were collapsed and sister nodes representing the same species were collapsed and labeled with the species name using the program Archaeopteryx (Han and Zmasek. 2009).

Phylogeny Analysis II

Based on the results of the above analyses we were able to see that *Amaranthus* formed five groups and that membership in these groups was well supported. However, the basal relationships between the groups had low support. To try to clarify this relationship we

estimated a coalescent tree using BEST (Liu. 2008) on a reduced dataset. The reduced dataset contained 21 sequences. Two sequences each for *A. powellii, A. caudataus, A. palmeri, A. spinosus, A. acanthochiton, A. tuberculatus, A. viridis, A. tricolor, A. crassipies,* and *Celosia trigyna*; and then one sequence to be used as the outgroup from *Acathocharis bidentata.* Each pair of sequences from the same species was grouped into an appropriate taxset and the concatenated sequence was partitioned into 6 genes (*ITS1*, *5.8s*, and *ITS2* were kept combined as *ITS*). All the partitions were assigned GTR+G as their mutation model, *MatK* was assigned as haploid, and all partitions were allowed to have independent model parameters.

Results

Phylogentic Analysis I

The full concatenated data set contains the six gene sequences (*MatK*, *ITS*, *A07*, *A36*, *A37*, and *A40*) for each of the 56 accessions (*Acathacaryes bidentata* was not included) listed in table 1.1. The total concatenated and aligned sequence was 4533 base pairs long, with 817 sites considered parsimony informative By DnaSP v5.10 (Rozas and Rozas. 1999). The phylogeny was calculated in both MrBayes 3.2.1 (Figure 1.1) and MEGA 5. The maximum likelihood tree estimated in MEGA is not shown. The maximum likelihood tree was very similar to the one estimated by MrBayes; the only major difference is the genetic distances between the *Amaranthus* species. The distance from *Amaranthus* to *Celosia* is very similar between the two estimates. The distances between *Amaranthus* species in the maximum likelihood tree are about one third that of the

Bayesian tree. However, based on the handling of ambiguous sites each estimation method uses we were not surprised by this difference and we feel that the Bayesian tree is the more accurate representation of the genetic distance. Our data had some heterozygous sites coded using IUPAC ambiguity codes, when MrBayes encounters this it considers either nucleotide equally likely in the analysis. When MEGA encounters this it deletes the whole position from the dataset (even for individuals that are not heterozygous); this results in many potentially informative sites being removed and the apparent genetic distance being reduced.

From the results of the phylogenetic analysis (Figure 1.1) we see that the *Amaranthus* species form five groups. Based on the naming scheme used by the Flora of North America (Mosyakin and Robertson. 1993; Mosyakin and Robertson. 1996) we refer to these groups Acnida, Amaranthus A, Amaranthus B, Albersia A, and Albersia B. However, the relationship between the groups is unclear as all the basal bifurcations between the groups were collapsed due to low support.

The six individual gene trees were also estimated (Appendix B figures B.1-6). They support the grouping of *Amaranthus* species into five groups but suffer from poor resolution. This is likely due to each gene not having enough data to inform a well-supported phylogeny with so many operational taxonomic units. The trees also suffered from branch length expansion; this is a problem often caused by low data density and the resulting flat posterior probability for tree height (Brown *et al*. 2010; Marshall. 2010).

Phylogenetic Analysis II

Because neither the Bayesian inference tree (Figure 1.1) nor the maximum likelihood tree were able to give a clear picture of the basal relationships between the groups we made a coalescent tree in BEST using a subset of the original data set (Table 1.1, bolded accessions). The results of the BEST analysis (Figure 1.2), however, are similar: the deep bifurcations still have low support. The more shallow bifurcations are very well supported, confirming the validity of the five groupings within *Amaranthus*.

Phylogeny of Amaranthus *Based on Bayesian and Coalescent Analysis*

All three analyses (Bayesian, Maximum likelihood, and coalescent) resulted in similar topologies (Figures 1.1 & 1.2). They show the genus divided into five subgenera that are each well supported as monophyletic in all analyses; the branch leading to Acnida has 99% posterior probability in the Bayesian estimate (PPB), 95% bootstrap support in the maximum likelihood analysis (BS), and 100% posterior in the coalescent analysis (PPC). The support for the other groupings in similarly high: Albersia A has 100% PPB, 94% BS, and 81% PPC; Albersia B has 99% PPB, and 77% BS, there is no value for the coalescent because only one representative of that clade was included; Amaranthus has 100% PPB, 100% BS, and 97% PPC. Amaranthus then splits into Amaranthus A and B; these groupings are well supported. The branch leading to Amaranthus A has 100% PPB, 99% BS, and 97% PPC; and the branch leading to Amaranthus B has 100% PPB, 98% BS, and 98% PPC. However, the relationships between Amaranthus, Acnida, Albersia A, and Albersia B are not well resolved; they were collapsed due to low support in the

Bayesian tree (Figure 1.1) and they are similarly poorly supported in the coalescent tree (the Albersia A&B / Acnida split has 34% posterior probability). This may suggest an initial rapid radiation after *Amaranthus* diverged from the other members of *Amaranthaceae*. Alternatively, we may just not have data from a locus with slow enough evolution to resolve the deep branches of the *Amaranthus* phylogeny. However, these phylogenies included both a chloroplast gene (*MatK*) and the *ITS* region and many other phylogenies of plants have successfully resolved branches of similar depth using these two genes; this leads us to favor the theory of rapid radiation.

Differences between Observed Phylogeny and Expected Species Relationships

If the results of our phylogenic analyses (Figures 1.1 and 1.2) are compared to the groupings of *Amaranthus* presented in the Flora of North America (FNA), the most rigorous treatment of the relationships within *Amaranthus* currently published, a few notable changes are seen (Mosyakin and Robertson. 1993; Mosyakin and Robertson. 1996). The first is that in the FNA the sub-genus Acnida is further subdivided into Acnida, Saueranthus, and Acanthochiton; our phylogeny does not support further division of the sub-genus Acnida thus we do not recognize the sections Acanthochiton, Acnida, or Saueranthus. In contrast, we have evidence to divide the sub-genus Amaranthus into two sections that are not recognized by the FNA. We propose sections A and B within the sub-genus Amaranthus. Lastly, the sub-genus Albersia in the FNA contains all remaining

species, based on our phylogeny this sub-genus should be split into two sub-genera, which we refer to as Albersia A and Albersia B.

In addition to these changes to the sub-divisions within the genus *Amaranthus* we also propose some changes to which of these groups some species belong to. Most of the species remain in the same group described in the FNA with most of the Albersia species in Albersia A and most of the Amaranthus species in Amaranthus A (Mosyakin and Robertson. 1993; Mosyakin and Robertson. 1996). However, *A. spinosus* is a member of Amaranthus B. *A. palmeri* is a member of Amaranthus B, not Acnida (this was also suggested by Wassom and Tranel (2005)). *A. greggii* is moved from Acnida to Albersia A. *A. crassipies* and *A californicus* are moved from Albersia to Albersia B, and *A. tamaulipensis* is moved from Amaranthus to Albersia B. Lastly, *Amaranthus graecizans silvestris* and *A. graecizans aschersonianus* should be separate species instead of sub species (both in Albersia A). For the final species assignments proposed, see figure 1.1.

After making these rearrangements there are some unexpected observations of the phylogeny (Figure 1.1) regarding individuals that were listed in the USDA germplasm as being of the same species / sub-species but are not sister tips in our phylogeny. One of the *A. tuberculatus* accessions (PI553086) groups with *A. floridanus* and the other (PI603881) with *A. arenicola*. The one grouping with *A. floridanus* had its name changed 5-Feb-2002 from *A. rudis* to *A. tuberculatus* ssp. *rudis*. This may suggest that the original nomenclature was more correct. Additionally our phylogeny includes two accessions of

A. hypochondriacus that are not grouping together. The individual PI477917 has a long branch length and is not grouping with its sister accession, Ames5689. *A. hypochondriacus* is a crop so that might be part of the explanation. There may have been different hybridizations involved in making each of these cultivars and they are not actually the same species, but rather morphologically similar domestic amaranths of different origins.

Discussion

Dioecy versus Bayesian Inference Phylogeny

Most of the dioecious plants are part of the Acnida subgenera; in fact, that was originally the defining characteristic for that group (Mosyakin and Robertson. 1993; Mosyakin and Robertson. 1996). However, our results place two dioecious species in different sub-genera. *A. palmeri* groups with *A. spinosus* in Amaranthus B, and *A. greggii* is in Albersia A (Figure 3). This suggests that dioecy had to evolve at least three times. AFLP based genetic relationships among weedy *Amaranthus* species (Wassom and Tranel. 2005) corroborates the grouping of *A. palmeri* with *A. spinosus*, instead of the other dioecious species. Unfortunately there are no other gene-based phylogenies that include *A. greggii* at this time.

A study on the inheritance of mating system type and sex determination in Amaranth was done in 1940. Based on the resulting progeny from crosses between monoecious and dioecious species the dioecious species were determined to have an XY system of sex

determination with most monoecious plants being genetically female (XX). Most of the monoecious *Amaranthus* included in this study have both male and female flowers in their inflorescences (*A. hybridus*, *A. caudatus*, *A. retroflexus*, *A. powellii*, *A. tuberculatus* (labeled in this study as *Acnida tuberculata* and *Acnida tamariscina*), and *A. australis* (*Acnida cuspidate*)). The male flowers are at the tip and then the rest of the flowers are female; Murray called this phenotype "type I monoecious". *Amaranthus spinosus* is different from these other monoecious amaranth with respect to flower arrangement and was labeled as "type II" by Murray (Murray. 1940). The female flowers develop in the axils of the branches and at the very base of the terminal inflorescences, the rest of the flowers (making up the bulk of the terminal inflorescences) are male.

This "type I" and "type II" binary is not really a complete view of floral phenotype in *Amaranthus*, but there have been many taxonomic rearrangements since the 1940s. Murray fairly accurately covers the dioecious amaranth and the amaranth from the sub-genus Amaranthus (Murray. 1940). However, he does not include Albersia A and B. Most of these species have flower morphology that is somewhere between "type I" and "type II". The Albersia amaranths all have axillary clusters and some also have terminal spikes. For the species with spikes, the male flowers tend to be at the tip of the spike, otherwise the two flower types are mixed together (Mosyakin and Robertson. 1993; Mosyakin and Robertson. 1996).

Murray did several crosses to try to understand the segregation of the sex determining trait in *Amaranthus*. He showed that in a cross of an Amaranthus A and an Acnida with the Acnida as the pollen donor a 1:1 male to female offspring is observed. In the reciprocal cross only female and the occasional sterile offspring were observed. This suggests that in Acnida the males are the heterogametic sex and that Amaranth A individuals are all "female". But in a cross between an Acnida and *A. spinosus* predominantly male offspring and a few monoecious offspring with a very small number of female flowers were observed. This seems to suggest that *A. spinosus* is not "female" like the species in Amaranthus A (Murray. 1940). Unfortunately, Murray's study did not include any Albersia individuals and no further work on this topic has been published.

The ability of the dioecious and monoecious amaranth to cross and form fertile hybrids (either of the monoecious or dioecious type) combines with the results of this phylogeny to suggest that dioecy is not a reproductively isolating trait that strongly separates some amaranth from others, that instead it is a more variable trait that has switched multiple times in the evolution of the genus. Better understanding the mating systems and level of reproductive isolation in *Amaranthus* is important not only from a taxonomic stand-point—as dioecy was originally considered a key determining factor in deciding which amaranth were closely related. It is also important to understanding the potential for spread of traits and gene-flow, which may not be as hindered by differences in mating system as originally presumed. This is important due to the fact that many amaranths are agricultural weeds, and many of these weeds have adapted to be resistant to herbicides.

Weediness in Amaranthus

There are species considered weeds in all of the *Amaranthus* sub-genera except Albersia B. The weeds with herbicide resistance in the US are *A. palmeri, A. retroflexus, A. hybridus, A. tuberculatus, A. powellii,* and *A. blitum* (Heap. 2013). Most of these are from Amaranthus A (*A. palmeri* and *A. spinosus* – Am B, *A. tuberculatus* – Acnida, and *A. blitum* – Albersia A), but they are not more closely related than that (Figure 1.3). Species identified as weeds were done so based on the January 2010 revision of the Composite List of Weeds as compiled by the Standardized Plant Names subcommittee (WSSA Standardized Plant Names Subcommittee. 2010).

Weediness is not a specific suite of traits. A weed is defined as a plant that grows unwanted in a plant community under human cultivation. However, there are many traits that make a species more successful as a weed (Table 1.3). One of the most important traits relating to weediness is adaptation to growing in disturbed habitats. For a true weed the disturbed habitat is the agricultural field (which is tilled, planted, and harvested yearly or more often), but plants can also be adapted to growing in waste areas such as railroad tracks or road-side ditches; these habitats are also frequently disturbed by human activity. This trait, adaptation to growing in disturbed habitats, is a trait that is found in almost all members of *Amaranthus*. Of the thirty-eight species of amaranth described in the Flora of North America twenty-six are specifically described as growing in "disturbed habitats", "waste areas", or "agricultural fields". Of the remaining eight species three are cultivated and the remaining five are described as growing in habitats that could easily be

considered disturbed by humans or water including canals, ditches, ballasts, sand dunes, and tidal flats (Mosyakin and Robertson. 1993).

An important part of being adapted to growing in disturbed habitats is being competitive in said habitat. Amaranths use C4 photosynthesis; a trait that appears to be a recent switch by the genus, as closely related genera *Celosia, Chamissoa,* and *Pleuropetalum* are all C3 (G. Kadereit *et al*. 2003; Sage *et al*. 2007). C4 photosynthesis is a molecular pathway that improves the efficiency of photosynthesis by reducing the amount of photorespiration. This improved efficiency is most marked under conditions of drought, high temperature, and high light level (Sage and Monson. 1999). The increased photosynthesis efficiency of C4 plants improves their competitiveness in their environment, and that the improvement is most marked in high light suggests that this is even more beneficial to plants growing in disturbed habitats that would not have large established plants creating shade.

Reproductive strategy can also influence success as a weed. As was discussed in detail in the preceding section, most amaranths are monoecious and able to both outcross and self-pollinate. The ability to employ both strategies confers individuals with a competitive advantage in their environment. Out-crossing allows for more genetic variation and thus the potential for new beneficial phenotypes to arise. Particularly in weeds, variation itself can be an advantage—such as variation in dormancy resulting in discontinuous germination. At the same time, there is a risk with being an obligate out-crossing species

in that there may be no potential mates available. The ability to employ both strategies optimizes reproductive success. Not all amaranth are monoecious, some species, particularly the weed *A. palmeri*, are dioecious and thus obligate out-crossers. This could potentially have a negative impact on reproduction if there are no suitable mates; however, in *A. palmeri* it has been shown that female plants are capable of generating seeds through facultative apopmixis (agamospermy) which alleviates this limitation to reproductive success (Trucco *et al*. 2007).

These traits taken together suggest that weediness or invasiveness is the default state for members of this genus. Traits favoring weedy and invasive character were likely found in the common ancestor of all *Amaranthus*. Since many species of amaranth are adapted to growing in sand dunes and tidal flats it is possible that the common ancestor adapted to grow in a similarly naturally disturbed habitat. Being adapted to grow in habitats that were regularly disturbed by water made these species able to quickly adapt to the new human disturbed habitats that began spreading over the Americas. The species that are not currently considered weeds are likely at a high risk of becoming weeds, needing only to become more competitive with popular crops to gain a foothold in agricultural habitats. Gaining this competitiveness may be the result of adaptation by the amaranth or changing field management practices by growers. Then considering the high percentage of weedy amaranth that have some level of herbicide resistance it seems likely that any amaranths that move into agricultural habitats will be at high risk of developing herbicide resistance either through gene flow with existing resistant populations or through novel mutations.

Cultivated Amaranth

Many of the traits that make a plant well adapted to weedy or invasiveness would also be beneficial in the context of a domestic crop species. Both weeds and crops must be adapted to growing in a disturbed habitat, and traits such as fast growth, and high competitiveness would be beneficial in a crop species. The unsurprising result of this is that many important crop species have congeneric weeds. The weedy relatives may or may not infest the crop species themselves. Examples include rice (*Oryza sativa*), oats (*Avena sativa*), sorghum (*Sorghum bicolor*) and sweet potatoes (*Ipomea batatas*) (Warwick and Stewart. 2005). *Amaranthus* can also be counted, as there are three species of cultivated amaranth. Cultivated amaranths are grown both as a grain/pseudo-cereal (*A. caudatus, A. cruentus* and *A. hypochondriacus*) and as a vegetable (*A. tricolor* and *A. graecizans aschersoniasis*). While none of these species are major crops they have begun to attract more attention due to their robust growth—they are fast growing and highly drought tolerant—and nutritional superiority to many more widely cultivated grains. These traits make amaranth a potentially valuable food source, especially in areas where food security is an issue due to harsh or unpredictable climate that may make the use of more traditional crops problematic.

With respect to the findings of our phylogeny (Figures 1.1 and 1.2), the grain crops are sister species in Amaranthus A, this may be due to both cultivar species being domesticated from the same ancestral species of amaranth, likely one with larger (relative to the genus) seeds. The two cultivated vegetable crops are both from Albersia A,

although not sister species. Since they are not sister species and all *Amaranthus* species are edible as vegetables it is not certain if the relationship between these two species is incidental or the result of some beneficial traits common to the Albersia A sub-genus. As all amaranth are edible, there are also many species that have semi-cultivated land-race populations but these populations are spread out among all the sub-genera and do not seem to have any relationship to each other or to particular traits.

Table 1.1: Taxa sampled, including USDA germplasm repository numbers, for this study. Bold taxa and accession numbers indicate individuals included in smaller dataset used for coalescent analysis

Species	**Accession number(s)**
Amaranthus acanthochiton	**PI 632239, PI 632238**
Amaranthus acutilobus	PI 633578, PI 633579
Amaranthus arenicola	PI 599673, PI 607459
Amaranthus asplundii	PI 604196
Amaranthus australis	PI 553076
Amaranthus blitoides	PI 612387
Amaranthus blitum	PI 610262, PI 632245, PI 652433
Amaranthus californicus	PI 595319
Amaranthus caudatus	**Ames 13860, PI 553073**
Amaranthus crassipes	**PI 642743, PI 649302**
Amaranthus crispus	PI 633582
Amaranthus cruentus	Ames 2093, PI 477913
Amaranthus fimbriatus	PI 605738
Amaranthus floridanus	PI 553078
Amaranthus graecizans	Ames 24671, Ames 5387, PI 271465
Amaranthus greggii	PI 632240
Amaranthus hybrid	Ames 5688
Amaranthus hybridus	PI 603886
Amaranthus hypochondriacus	Ames 5689, PI 477917
Amaranthus muricatus	PI 633583
Amaranthus palmeri	**Ames 15298, PI 549158,** PI 604557
Amaranthus powellii	**PI 572260, PI 572261,** PI 632241
Amaranthus quitensis	PI 511745, PI 652419, PI 652421
Amaranthus retroflexus	PI 603845, PI 607447
Amaranthus spinosus	**PI 632248, PI 642740**
Amaranthus standleyanus	PI 605739
Amaranthus tamaulipensis	PI 642738
Amaranthus tricolor	**Ames 15326, PI 477918**
Amaranthus tuberculatus	**PI 553086, PI 603881**
Amaranthus viridis	**PI 536439, PI 652434**
Amaranthus wrightii	PI 632242
Celosia trigyna	**PI 649298,** PI 649299, **PI 482244**
Acathacaryes bidentata	**PI 613015**

Table 1.2: Primer pairs used to PCR amplify and then Sanger sequence the six genes used in the phylogeny

Gene targeted for amplification	Source of Primers	5’ to 3’ Sequence (Forward/reverse)
ITS (internal transcribed spacers 1 and 2)	White *et al*. 1990	TCCTCCGCTTATTGATATGC / GGAAGTAAAAGTCGTAACAAGG
MatK (maturase K)	Our lab	CGATCTATTCATTCAATATTTC / TCTAGCACACGAAAGTCGAAGT
A07 (Endosomal P24A protein precursor, putative)	Our lab	GGAAGCTTGTTGTGGGTGAT / AATGGCTGAAACAGGTCCAC
A36 (DEAD box RNA helicase, putative)	Our lab	TGGTTATCCGTGCCTTTCTC / CAGGACCTGGATTCTTTCCA
A37 (serine-type endopeptidase, putative)	Our lab	CACTGAAGCCTACGGAGARG / GATTGGGCTGGTCACTSTGT
A40 (glutaredoxin, putative)	Our lab	GGTGAGCTTATCGGTGGGTG / TCCGAAAGGGTTGATTTRAG

Table 1.3: Traits associated with weediness based on Warwick and Stewart 2005, and an assessment how well these traits represent *Amaranthus* based on current literature.

Traits associated with weediness (Warwick and Stewart. 2005)	Are amaranths like this?
Seeds that are easily dispersed long distances	Likely, seed size varies between species but all are between 0.7 and 1.6mm in diameter (Mosyakin and Robertson. 1993)
Seed dormancy	Yes, varies by species but all tested species show some level of dormancy (Cristaudo *et al*. 2007)
Discontinuous germination	Likely, under optimal conditions tested species germinated 2-8 days after planting (Steckel *et al*. 2004)
Ability to germinate under a wide range of conditions	Yes, tested species showed some level of germination under all testing conditions <10°C (Cristaudo *et al*. 2007; Steckel *et al*. 2004)
Long lived seeds	No, within 3 years 80% of *A. palmeri* seeds are dead (Sosnoskie *et al*. 2011)
Rapid growth to flower (annual)	Yes, all amaranths are annuals (Mosyakin and Robertson. 1993)
Continuous (non-determinate) seed production	Unlikely, most species have coordinated flowering and terminal inflorescences (Mosyakin and Robertson. 1993)
High seed output	Yes, *A. retroflexus* can produce up to 250,000 seeds / plant (and it is not exceptional for the genus) (Stevens. 1957)
Seed produced in wide range of conditions	Likely, plants under stress will flower early. However, different parental flowering conditions affects seed dormancy (Kigel *et al*. 1977)
Seed shattering	Yes (Fitterer *et al*. 1996)
Plasticity of growth	Likely, planting date has been shown to affect the growth rate of *A. palmeri* (Keeley *et al*. 1987)
Highly competitive	Likely, all species are C4 photosynthesizers (G. Kadereit *et al*. 2003; Sage *et al*. 2007), and *A. palmeri* is highly competitive in agricultural fields (Burke *et al*. 2007; Gaylon D. Morgan *et al*. 2001; Klingaman and Oliver. 1994; Massinga *et al*. 2001; Rowland *et al*. 1999)
Not an obligate selfer	Yes, most amaranth can both self and outcross (Mosyakin and Robertson. 1993). Dioecious species have documented facultative apopmixis (Whitaker *et al*. 2013).
Unspecialized pollinators	Yes, amaranth are wind pollinated (Sosnoskie *et al*. 2009)
Presence of bitter substances in seed/fruit to increase pest resistance	No, amaranths are edible (Mosyakin and Robertson. 1993)

Figure 1.1: Phylogeny of *Amaranthus*. The evolutionary history was inferred using the Bayesian estimation method based on the General time reversible model with a discrete Gamma distribution used to model evolutionary rate differences among sites (4 categories (+*G*)) and separate partitions for $1^{st}/2^{nd}$ codon position, 3^{rd} codon position, and noncoding. Each of the partitions had independent estimates of the model parameters and rate multiplier but shared the estimate for tree topology and branch lengths. Branches corresponding to partitions reproduced with less than 75% posterior probability support are collapsed. Also any partitions with sister OTUs of the same species were collapsed for readability. Initial tree(s) for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 56 nucleotide sequences. There were a total of 4533 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.

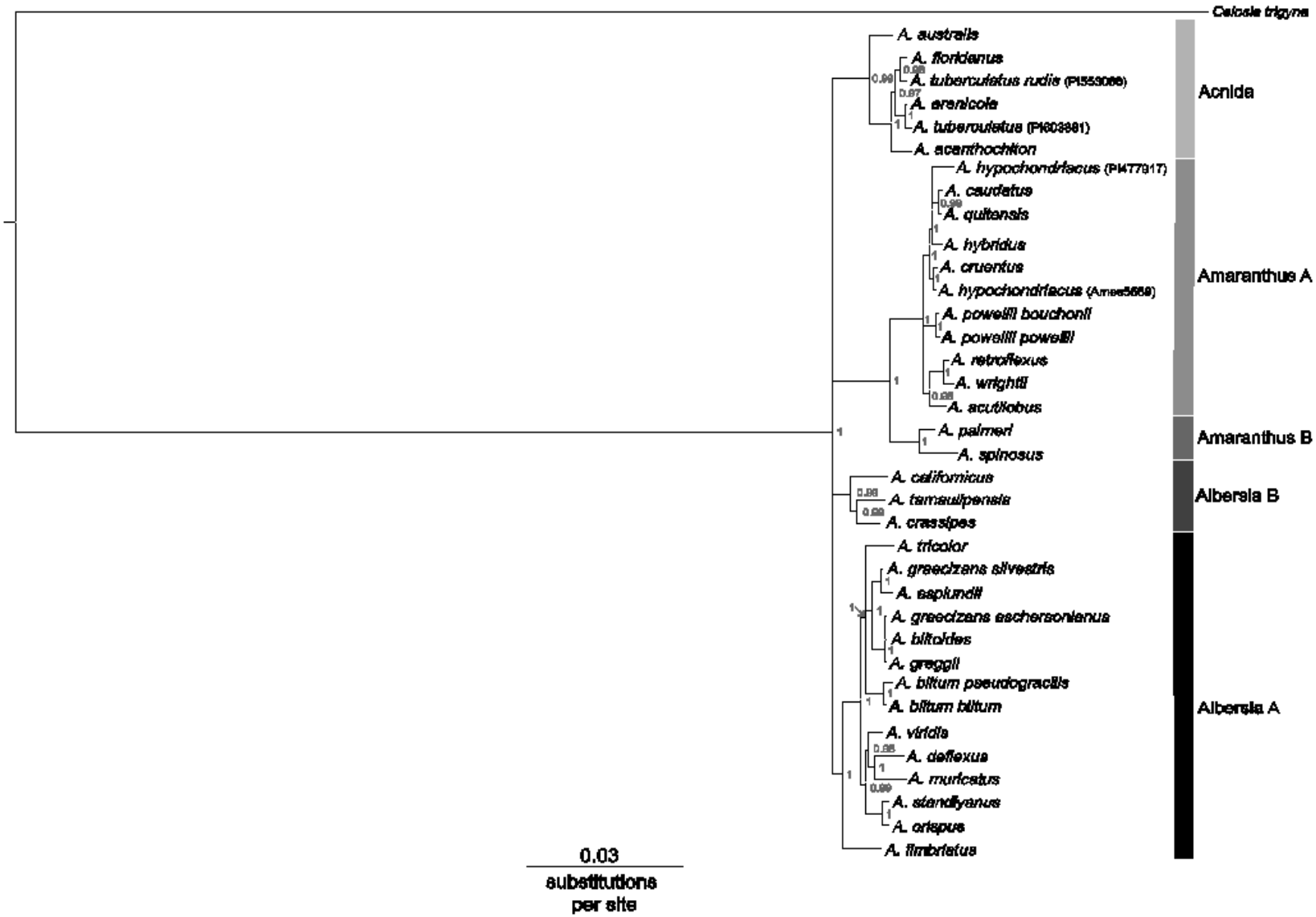

Celosia trigyna
A. australis
A. floridanus
A. tuberculatus rudis (PI553066)
A. arenicola
A. tuberculatus (PI603981)
A. acanthochiton
A. hypochondriacus (PI477917)
A. caudatus
A. quitensis
A. hybridus
A. cruentus
A. hypochondriacus (Ames5668)
A. powellii bouchonii
A. powellii powellii
A. retroflexus
A. wrightii
A. acutilobus
A. palmeri
A. spinosus
A. californicus
A. tamaulipensis
A. crassipes
A. tricolor
A. graecizans silvestris
A. aspundii
A. graecizans eschersonianus
A. blitoides
A. greggii
A. bitum pseudogracilis
A. bitum bitum
A. viridis
A. deflexus
A. muricatus
A. standleyanus
A. crispus
A. fimbriatus
Acnida
Amaranthus A
Amaranthus B
Albersia B
Albersia A
0.03
substitutions
per site

Figure 1.2: Coalescent inference of basal relationships. The evolutionary history was inferred by using BEST (Bayesian estimation of species trees) method based on the general time reversible model. The consensus tree is taken to represent the evolutionary history of the taxa analyzed; the numbers to the right of the branches indicate the posterior probability of that bifurcation. Initial tree(s) for the heuristic search were obtained automatically as follows: BIONJ method with MCL distance matrix was used. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+*G*)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 21 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 4533 positions in the final dataset. Evolutionary analyses were conducted in BEST v2.3.

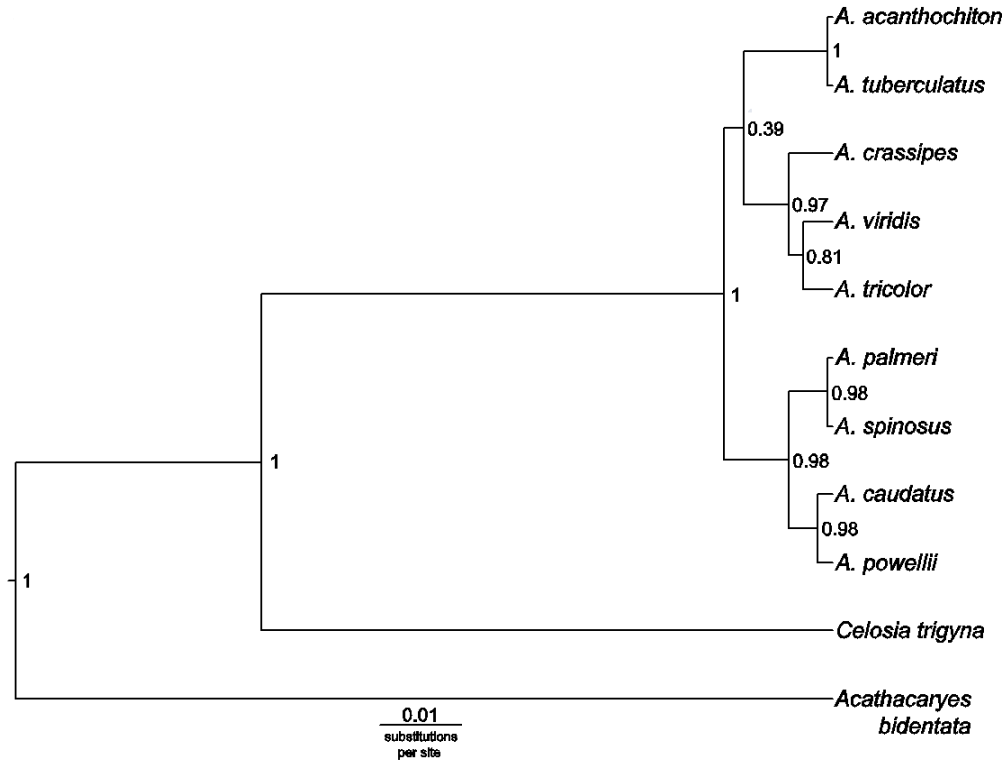
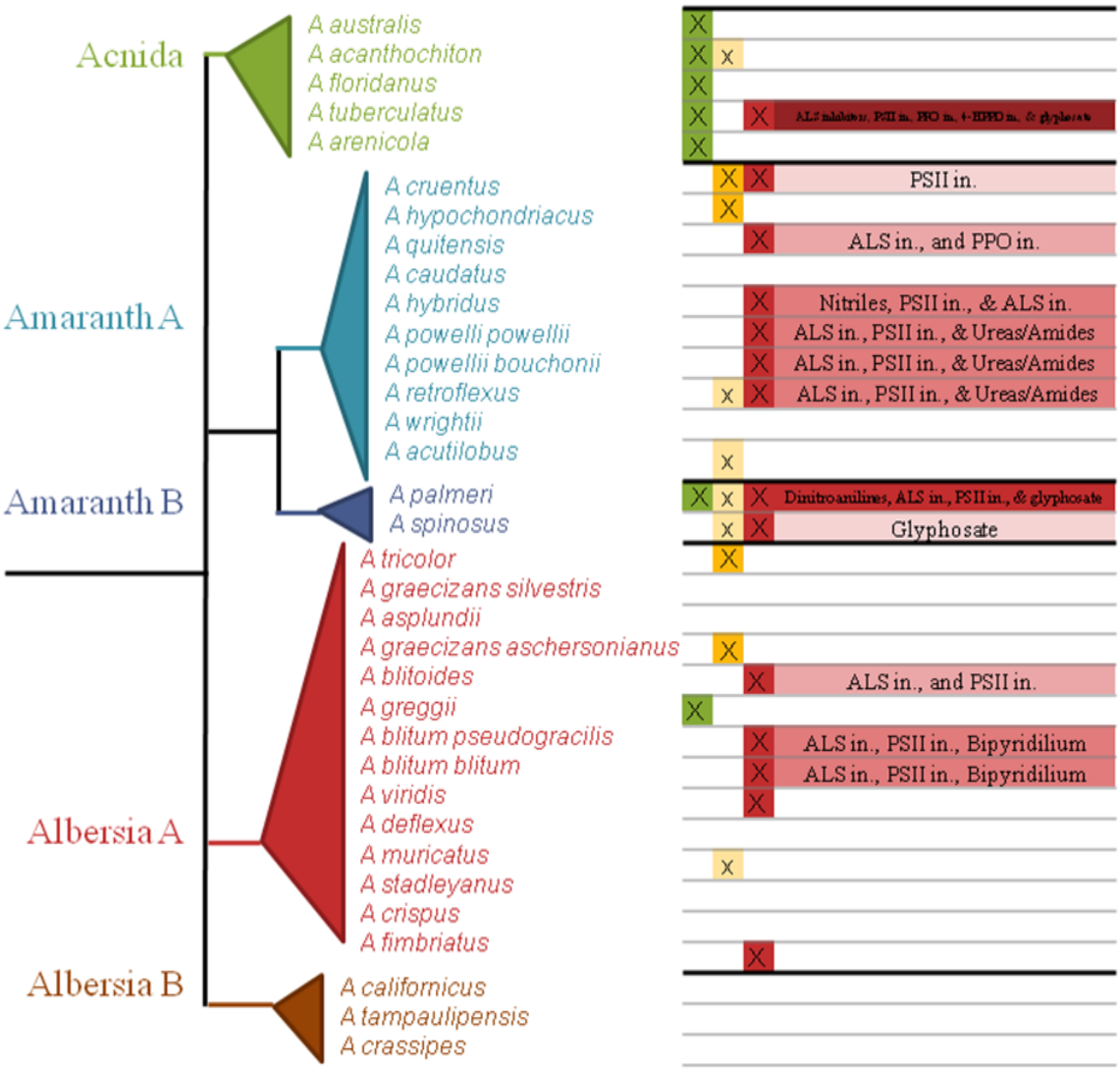

A. acanthochiton
1
A. tuberculatus
0.39
A. crassipes
0.97
A. viridis
0.81
A. tricolor
1
A. palmeri
0.98
A. spinosus
0.98
A. caudatus
0.98
A. powellii
1
1
Celosia trigyna
Acathacaryes bidentata
0.01
substitutions per site

Figure 1.3: Distribution of agrinomically relevant traits in *Amaranthus*. A comparison between the five clades identified in my phylogenetic analysis and the distribution of agronomically important traits. Column 1 (green) indicates dioecy; column 2 (yellow) indicates utilization as a food source, dark yellow indicates a cultivar while light yellow indicates a landrace; column 3 (red) indicates weediness, and column 4 lists the herbicides or classes of herbicides to which populations of the species have become resistant, color depth increases as the number of distinct herbicide resistances increases. ALS – acetolactate synthase inhibitors (HRAC group B), PSII – photosystems II inhibitors (group C1), PPO – protoporphyrinogen oxidase inhibitors (group E), 4-HPPD – 4-hydroxyphenyl-pyruvate-dioxygenase inhibitors (group F2), PSI – photosystem I disruptors (group D), Ureas – ureas and amides (group C2), nitriles (group C3), dinitroanilines (group K1), and glyphosate (group G)


Acnida
A australis
A acanthochiton
A floridanus
A tuberculatus
A arenicola
Amaranth A
A cruentus
A hypochondriacus
A quitensis
A caudatus
A hybridus
A powelli powellii
A powellii bouchonii
A retroflexus
A wrightii
A acutilobus
Amaranth B
A palmeri
A spinosus
Albersia A
A tricolor
A graecizans silvestris
A asplundii
A graecizans aschersonianus
A blitoides
A greggii
A blitum pseudogracilis
A blitum blitum
A viridis
A deflexus
A muricatus
A stadleyanus
A crispus
A fimbriatus
Albersia B
A californicus
A tampaulipensis
A crassipes
PSII in.
ALS in., and PPO in.
Nitriles, PSII in., & ALS in.
ALS in., PSII in., & Ureas/Amides
ALS in., PSII in., & Ureas/Amides
ALS in., PSII in., & Ureas/Amides
Dinitroanilines, ALS in., PSII in., & glyphosate
Glyphosate
ALS in., and PSII in.
ALS in., PSII in., Bipyridilium
ALS in., PSII in., Bipyridilium

References

Akaike, H., 1981 Likelihood of a model and information criteria. Journal of econometrics **16:** 3-14.

Akaike, H., 1974 A new look at the statistical model identification. Automatic Control, IEEE Transactions on **19:** 716-723.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. Journal of Molecular Biology **215:** 403-410.

Brenner, D. M., D. D. Baltensperger, P. A. Kulakow, J. W. Lehmann, R. L. Myers *et al*, 2000 Genetic resources and breeding of *Amaranthus**. Plant Breeding Research **19:** 227-285.

Brown, J. M., S. M. Hedtke, A. R. Lemmon and E. M. Lemmon, 2010 When trees grow too long: Investigating the causes of highly inaccurate bayesian branch-length estimates. Systematic Biology **59:** 145-161.

Burke, I. C., M. Schroeder, W. E. Thomas and J. W. Wilcut, 2007 Palmer amaranth interference and seed production in peanut. Weed Technology **21:** 367-371.

Chan, K. F., and M. Sun, 1997 Genetic diversity and relationships detected by isozyme and RAPD analysis of crop and wild species of *Amaranthus*. Theoretical and Applied Genetics **95:** 865-873.

Cristaudo, A., F. Gresta, F. Luciani and A. Restuccia, 2007 Effects of after-harvest period and environmental factors on seed dormancy of *Amaranthus* species. Weed Research **47:** 327-334.

Ewing, B., and P. Green, 1998 Base-calling of automated sequencer traces using Phred. II. Error Probabilities. Genome Research **8:** 186-194.

Ewing, B., L. Hillier, M. C. Wendl and P. Green, 1998 Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. Genome Research **8:** 175-185.

Fitterer, S., B. Johnson and A. Schneiter, 1996 Grain amaranth harvest timeliness in eastern North Dakota. Progress in New Crops.ASHS Press, Alexandria 220-223.

G. Kadereit, T. Borsch, K. Weising and H. Freitag, 2003 Phylogeny of *amaranthaceae* and *chenopodiaceae* and the evolution of C4 photosynthesis. International Journal of Plant Science **164:** 959-986.

Gaylon D. Morgan, P. A. Baumann and J. M. Chandler, 2001 Competitive impact of palmer amaranth (*Amaranthus palmeri*) on cotton (*Gossypium hirsutum*) development and yield. Weed Technology **15:** 408-412.

Hall, T., 2001 Biolign alignment and multiple contig editor. North Carolina State University, Department of Microbiology, Raleigh, North Carolina.

Hall, T. A., 1999 *BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT*.

Han, M., and C. Zmasek, 2009 phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics **10:** 356.

Hasegawa, M., H. Kishino and T. Yano, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution **22:** 160-174.

Heap, I. M., *International Survey of Herbicide Resistant Weeds*. Online. Internet. Wednesday, November 06, 2013 . Available www.weedscience.com

Keeley, P. E., C. H. Carter and R. J. Thullen, 1987 Influence of planting date on growth of palmer amaranth (*Amaranthus palmeri*). Weed Science **35:** pp. 199-204.

Kigel, J., M. Ofir and D. Koller, 1977 Control of the germination responses of *Amaranthus retroflexus* L. seeds by their parental photothermal environment. Journal of Experimental Botany **28:** 1125-1136.

Klingaman, T. E., and L. R. Oliver, 1994 Palmer amaranth (*Amaranthus palmeri*) interference in soybeans (glycine max). Weed Science **42:** 523-527.

Lee, R. M., J. Thimmapuram, K. A. Thinglum, G. Gong, A. G. Hernandez *et al*, 2009 Sampling the waterhemp (*Amaranthus tuberculatus*) genome using pyrosequencing technology. Weed Science **57:** 463-469.

Liu, L., 2008 BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics **24:** 2542-2543.

Marshall, D. C., 2010 Cryptic failure of partitioned bayesian phylogenetic analyses: Lost in the land of long trees. Systematic Biology **59:** 108-117.

Massinga, R. A., R. S. Currie, M. J. Horak and J. Boyer Jr., 2001 Interference of palmer amaranth in corn. Weed Science **49:** 202-208.

Mosyakin, S. L., and K. R. Robertson, 1996 *New Infrageneric Taxa and Combinations in Amaranthus (Amaranthaceae)*. Helsinki: Societas Biologica Fennica Vanamo, 1964.

Mosyakin, S. L., and K. R. Robertson, 1993 Amaranthus, pp. 410-435 in *Flora of North America North of Mexico*, edited by Flora of North America Editorial Committee. New York and Oxford.

Murray, M. J., 1940 The genetics of sex determination in the family *amaranthaceae*. Genetics **25:** 409.

Ronquist, F., and J. P. Huelsenbeck, 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:** 1572-1574.

Rowland, M. W., Don S. Murray and L. M. Verhalen, 1999 Full-season palmer amaranth (*Amaranthus palmeri*) interference with cotton (*Gossypium hirsutum*). Weed Science **47:** 305-309.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174-175.

Rozen, S., and H. J. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers., pp. 365-368 in , edited by K. S. Misener. Humana Press, Totowa, NJ.

Sage, R. F., and R. K. Monson, 1999 *C4 Plant Biology*. San Diego; Toronto: Academic Press.

Sage, R. F., T. L. Sage, R. W. Pearcy and T. Borsch, 2007 The taxonomic distribution of C4 photosynthesis in *amaranthaceae* sensu stricto. American Journal of Botany **94:** 1992-2003.

Sosnoskie, L. M., T. M. Webster, D. Dales, G. C. Rains, T. L. Grey *et al*, 2009 Pollen grain size, density, and settling velocity for palmer amaranth (*Amaranthus palmeri*). Weed Science **57:** 404-409.

Sosnoskie, L. M., T. M. Webster, A. S. Culpepper and J. M. Kichler, 2011 *The Biology and Ecology of Palmer Amaranth: Implications for Control*. University of Georgia Cooperative Extension, Athens, GA.

Steckel, L. E., C. L. Sprague, E. W. Stoller and L. M. Wax, 2004 Temperature effects on germination of nine *Amaranthus* species. Weed Science **52:** 217-221.

Stevens, O. A., 1957 Weights of seeds and numbers per plant. Weeds **5:** 46-55.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei *et al*, 2011 MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution **28:** 2731-2739.

Tavaré, S., 1986 Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathmatics in the Life Sciences **17:** 57-86.

Trucco, F., D. Zheng, A. J. Woodyard, J. R. Walter, T. C. Tatum *et al*, 2007 Nonhybrid progeny from crosses of dioecious amaranths: Implications for gene-flow research. Weed Science **55:** 119-122.

Warwick, S. I., and C. N. Stewart, 2005 Crops come from wild plants -- how domestication, transgenes, and linkages together shape ferality, pp. 9 in *Crop Ferality and Volunteerism*, edited by J. Gressel. CRC, Boca Raton, FL.

Wassom, J. J., and P. J. Tranel, 2005 Amplified fragment length polymorphism-based genetic relationships among weedy *Amaranthus* species. Journal of Heredity **96:** 410-416.

Whitaker, J. R., J. D. Burton, A. C. York, D. L. Jordan and A. Chandi, 2013 Physiology of glyphosate-resistant and glyphosate-susceptible palmer amaranth (*Amaranthus palmeri*) biotypes collected from north carolina. International Journal of Agronomy **2013**.

WSSA Standardized Plant Names Subcommittee, 2010 *January 2010 Revision of the Composite List of Weeds*. Weed Science Society of America.

Xu, F., and M. Sun, 2001 Comparative analysis of phylogenetic relationships of grain amaranths and their wild relatives (*Amaranthus*; amaranthaceae) using internal transcribed spacer, amplified fragment length polymorphism, and double-primer fluorescent intersimple sequence repeat markers. Molecular Phylogenetics and Evolution **21:** 372-387.

CHAPTER II

A UNIQUE SOLUTION: HOW HAS SEQUENCE CONTEXT AND CONSTRAINT INFLUENCED THE ADAPTATION OF *A. palmeri* TO THE HERBICIDE GLYPHOSATE?

K. E. Beard and A. Lawton-Rauh

Abstract

Premise of the study

Glyphosate resistance is unique among herbicide resistances because it can result from genomic copy number variation of the target enzyme to confer resistance. This study investigates the selection pressures acting on the *EPSPS* gene in the plant genus *Amaranthus* to better understand the evolutionary history that led to this unique herbicide resistance mechanism.

Methods

EPSPS was sequenced from 16 individuals representing eight *Amaranthus* species. These sequences were used to estimate synonymous and non-synonymous mutation accumulation rates in *EPSPS*. Mutation accumulation in *EPSPS* was compared to four other non-herbicide-target nuclear genes to better understand how patterns in *EPSPS* are caused by unique selection forces specific to the genomic region containing the *EPSPS* gene or genome-wide demographic forces.

Key Results

Analyses suggest that *EPSPS* is under strong purifying selection relative to other genes in *Amaranthus*. This purifying selection is not due to a lower than average rate of non-synonymous mutation accumulation, but rather from a higher than average rate of synonymous mutation accumulation found in other examples of purifying selection. This elevated synonymous mutation accumulation rate could be related to the observed

prevalence of CpG clusters in the coding sequence of *EPSPS*, as methylated CpGs are known to have a higher C to T and G to A mutation rate than surrounding sequence.

Conclusions

The non-synonymous mutation accumulation, a proxy for sequence constraint, is at a typical expected level in *EPSPS*, so extreme sequence constraint is likely not related to the lack of high-glyphosate-resistance conferring *EPSPS* point mutations. The synonymous mutation rate is higher than typical in *EPSPS*, and this is potentially caused by a high density of methylated CpG dinucleotides. Based on these findings we suggest a potential hypothesis regarding *EPSPS* gene amplification in response to glyphosate stress. Under glyphosate stress there could be methylation changes to *EPSPS* that cause *EPSPS* to become a site of replication initiation— unmethylated CpG islands are known to be involved in the initiation of replication (Delgado *et al*. 1998). This could result in fragments of *EPSPS* genomic sequence in the nucleus (Delgado *et al*. 1998), which could potentially be incorporated into the genome through non-homologous end joining. This would result in increased *EPSPS* copy number under herbicide stress without traditional transposon mediation. This represents an alternative explanation versus the current model of transposable element origins of gene duplication and proliferation.

Introduction

5-enolpyruval shikimate 3-phosphate synthase (EPSPS) is the penultimate enzyme of the shikimate pathway and the target of inhibition by the herbicide glyphosate (marketed as

RoundUp™ by Monsanto). The seven-step shikimate pathway starts with the non-hydrolytic addition of phosphoenolpyruvate (PEP) to erythrose 4-phosphate and ends with the formation of chorismate, the precursor to quinones, folates, and the aromatic amino acids phenylalanine, tyrosine, and tryptophan. EPSPS, the best studied enzyme of the pathway, catalyzes the reversible formation of 5-enolpyruval shikimate 3-phosphate (EPSP) from shikimate 3-phosphate and phosphoenolpyruvate (PEP) (Weaver and Herrmann. 1997). Glyphosate competitively inhibits EPSPS by binding to the EPSPS-shikimate 3-phosphate complex in place of PEP. Even though glyphosate is a structural analog of PEP it has not been shown to inhibit any other enzyme that uses PEP as a substrate; this may be related to the fact that glyphosate only binds to the complex of EPSPS and shikimate-3-phosphate (Schönbrunn *et al*. 2001; Steinrücken and Amrhein. 1980; Weaver and Herrmann. 1997).

In the United States, several weedy species controlled by glyphosate are members of the genus *Amaranthus*. *Amaranthus* is a large genus in *Amaranthaceae*, the largest family within the order *Caryophyllales* (Beard *et al.*, submitted). The genus contains approximately seventy species; most of which are native to tropical and subtropical zones of Central and South America with many species present nearly world-wide as introduced ruderals or weeds. Seventeen species are considered weeds and twelve of these weedy species have at least one population with resistance to at least one herbicide (Heap. 2013; WSSA Standardized Plant Names Subcommittee. 2010). Three species, *A. palmeri*, *A. spinosus*, and *A. tuberculatus*, have multiple populations documented as resistant to

glyphosate; these three species also responsible for most of the amaranth infestations in the US. Glyphosate resistance in *Amaranth* has been spreading since the first identified population in Macon, GA in 2004. There are now resistant *A. palmeri* found in over 150,000 sq mi of the South Eastern United States (William Vencill, personal communication). This costs growers in the US in increased management costs and lost yield. For example, the estimated cost of glyphosate resistant *A. palmeri* to growers due to increased management costs and decreased yield per season in Tennessee alone was over $200 million (Hembree. 2011; Gaylon D. Morgan *et al*. 2001; Rowland *et al*. 1999).

Glyphosate is used to control weeds in many different crops, particularly since the introduction of RoundUp Ready™, glyphosate tolerant, crops. Among its advantages are that it is relatively non-toxic, particularly to animals and humans, and that it works on many different weedy species. Before the first populations of glyphosate resistant weeds were identified there was much investigation into possible molecular mechanisms of glyphosate resistance. This work was primarily done with the goal of developing the glyphosate tolerant lines of crops to allow glyphosate to be used post-emergence. The final mechanism applied to crops was transgenic insertion of a glyphosate tolerant bacterial *AroA* (homologous to plant *EPSPS*) gene. The most common glyphosate insensitive EPSPS used was isolated from a strain of *Agrobacterium tumefaciens* known as CP4 (Padgette *et al*. 1995), which had been isolated from a glyphosate waste treatment facility (US patent 5633435). This strain contains what is known as a class II EPSPS; it has very low sequence homology to the class I EPSPS found in plants and *E. coli.*

Before discovering the CP4 EPSPS, target site mutations conferring resistance to glyphosate were some of the first looked into when glyphosate tolerant crops were being developed. However, the only mutation identified as conferring sufficient resistance without excessive loss of catalytic efficiency requires two mutation steps, and so it was not originally considered a likely source of weed resistance. However, when glyphosate resistant weeds started showing up in fields the *EPSPS* sequence was one of the first places that researchers looked. In some populations, particularly those with low level resistance (2-5 fold), mutations in *EPSPS* have been found. All of the *EPSPS* mutations found to date in glyphosate resistant weeds have been substitutions of proline 106 (Healy-Fried *et al*. 2007).

Another mechanism identified for glyphosate resistance in weeds is changes in translocation or absorption of the herbicide. Based on the current literature this would seem to be the most commonly identified mechanisms leading to high levels of resistance. The glyphosate is, instead of being loaded into the phloem, moved up, and sequestered into the tips of the leaves that have been sprayed (Lorraine-Colwill *et al*. 2002; Michitte *et al*. 2005; Perez-Jones *et al*. 2007). There have been efforts to understand the molecular and genetic basis of this resistance mechanism, but there has been little success to date. It is a difficult question because there are many genes and proteins involved in the regulation of intra and inter-cellular transport.

In 2010 the mechanism of glyphosate resistance in *A. palmeri* growing in Georgia (USA) was found to be massive gene duplication (40-100 copies) of *EPSPS* (Gaines. 2010; Gaines *et al*. 2011). A linear relationship between the number of copies and the level of resistance was seen, and the extra copies were distributed across the genome, which suggested transposable elements may be involved. Another example of this type of resistance was found in an Arkansas (USA) population of Italian ryegrass *(Lolium multiflorum*). These plants, first identified in 2008, were shown to have up to 25 copies of EPSPS in their genomes (Salas *et al*. 2012). Only a few other examples of this have been documented to date (Baerson *et al*. 2002; Ribeiro *et al*. 2013).

There are two naturally occurring mechanisms that have evolved to confer resistance to glyphosate: altered translocation and increased EPSPS protein level via increased *EPSPS* genomic copy number. This is in notable contrast to the adaptive strategies to other herbicides, where point mutations changing the amino acid sequence of the target protein are more common and confer high levels of resistance. Increase in genomic copy number of the gene encoding a pesticide target has been described in other systems. Examples include the naphthyl esterase gene in aphids and mosquitoes, and dihydrofolate reductase gene in cancer tissues treated with methotrexate (Devonshire and Field. 1991). However, target gene copy number proliferation is unique among documented herbicide resistances. The purpose of this work is to better understand the dynamics of mutation accumulation in the *EPSPS* gene, particularly with respect to selective constraint. Specifically in *Amaranthus*, a genus that is both highly problematic from an economic standpoint as well

as interesting for taking advantage of this unique amplification mechanism. This will potentially allow us to better understand the lack of a high-glyphosate-resistance conferring point mutation common in other herbicide resistances and the evolutionary context that lead to the genomic copy number proliferation mechanism resistance.

Materials and Methods

Sampling

All samples were obtained from the USDA germplasm repository. We sampled individuals from 16 accessions representing eight *Amaranthus* species (Table 2.1). Seeds were planted and grown in the greenhouse under ambient lighting until they had at least four true leaves, at which point leaf tissue was collected for DNA and RNA extraction.

DNA was extracted from 100mg of -80°C frozen leaf tissue. The tissue was ground in a Quiagen Mixer Mill (Qiagen, Germantown, MD, USA) and then DNA was purified following the standard protocol of the Nucleospin Plant II DNA extraction kit (Macherey Nagel, Düren, Germany). RNA was extracted from 100mg of fresh leaf tissue using the plant tissue protocol of the RNAeasy kit (Qiagen, Germantown, MD, USA), then cDNA was made with random priming using the iScript system (BioRad, Hercules, CA, USA).

Sequencing

DNA (or cDNA) was amplified using GoTaq Flexi (Promega, Madison, Wisconsin). For most individuals, we used EPSPS_F and EPSPS_R on cDNA. For individuals where

good quality sequences could not be obtained from cDNA we used exon specific primers on genomic DNA (Table 2.2).

The cycling conditions used for all primer pairs were as follows: 5min initial denaturation at 94°C then 35 cycles of touch-down PCR with 30sec denaturation at 94°C, 30sec annealing at 60-50°C (first cycle 60°, then each subsequent cycle 1°C lower than the previous until the cycle with a 51°C annealing temperature. Then 25 cycles each with a 50°C annealing temperature), and 3min extension at 72°C, and a final 10min extension at 72°C.

PCR products were electrophoresed on a 1% agarose gel to ensure quality amplification. Successful amplifications were sequenced at the Clemson University Genomics Institute using the same primers used for PCR. Before submission, PCR reactions were cleaned using an ExoAP treatment: to each 1µL of DNA we added 0.2 Units of exonuclease I (New England Biolabs, Ipswich, Massachusetts), 0.5 Units of Antarctic phosphotase (New England Biolabs, Ipswich, Massachusetts), and water to 2µL; then samples were incubated for 30min at 37°C and heat treated for 15min at 80°C.
Sequences were Phred-Phraped to merge forward and reverse sequencing reads and viewed in Biolign (Ewing and Green. 1998; Ewing *et al*. 1998; Hall. 2001). *EPSPS* genomic sequence was also identified from the reference genome of *Beta vulgaris* using BLAST and used as the out group. The final alignment used for analysis was created manually in bioedit (Hall. 1999).

Computational Analysis of DNA Sequences

We used MEGA v.5.05 (Tamura *et al*. 2011) to choose the best mutation model for the *EPSPS* sequence. MEGA 5 estimated the best fit model under both the automatic neighbor joining tree and a user defined tree. For consistency between analyses we restricted choices to models available in MEGA 5, MrBayes, and Hyphy/Data Monkey. Based on Akaike Information Criterion (Akaike. 1974) and Bayesian Information Criterion values (Akaike. 1981) we chose HKY for all subsequent analyses.

The Bayesian inference gene tree of *Amaranthus EPSPS* was estimated in MrBayes (Ronquist and Huelsenbeck. 2003) using a single data partition and the HKY model of sequence evolution. The tree was run for 5,000,000 generations. After running the parameter estimates were viewed in Tracer v1.5 to determine if the run had converged, and how much data needed to be discarded as burn-in (Rambaut and Drummond. 2007). Fast, Unconstrained Bayesian AppRoximation (FUBAR) implemented in Data Monkey was employed to detect sites evolving under the influence of pervasive diversifying and purifying selection pressures (Murrell *et al.* 2013). For each site in *EPSPS* FUBAR estimated the best fit values of α (synonymous mutation rate) and β (non-synonymous mutation rate) and calculated the posterior probability that ω (β/α) was not equal to zero. Estimates were fit with the gene tree calculated in Mr. Bayes with the out-group *Beta vulgaris* both included and excluded. For comparison to the rest of the *Amaranthus* genome we also conducted FUBAR analysis on four non-target nuclear loci (see chapter

I) using a neighbor joining tree estimated by Data Monkey. We used JMP (SAS institute inc., Cary, NC, USA) to compare the distribution of α, β, and ω values in each gene and look for statistically significant differences in average value and variance.

To visualize the location of variable and conserved amino acids in the EPSPS protein we used ConSerf (Ashkenazy *et al*. 2010). Two separate alignments were submitted, one with the translated amino acid sequence of eight *Amaranthus* sequences—one per species, randomly chosen; and one with amino acid sequence from several different eudicots downloaded from Genbank (Table 2.3), translated *Beta vulgaris* sequence, and three translated *Amaranthus* sequences (*A. palmeri* Ames15298, *A. tuberculatus* PI553086, and *A. spinosus* PI632248). In both cases ConSerf calculated the tree and chose the protein to model the amino acid sequences too, with both alignments ConSerf chose the wild type *E. coli* EPSPS (pdb 1G6S).

To investigate CpG methylation as a potential source of variation in mutation accumulation rate CpG Islands, part of the bioinformatics organization sequence manipulation suite, was implemented (Plot. 2000). CpG Islands reports potential CpG island regions using the method described by Gardiner-Garden and Frommer (1987). The calculation is performed using a 200bp window moving across the sequence at 1bp intervals. CpG islands are defined as sequence ranges where the Observed/Expected number of CpG dimers is greater than 0.6 and the GC content is greater than 50%. For our purposes all identified CpG islands that has a start site within 15bp of the next island

upstream were merged into a single island. This was done on all *Amaranthus EPSPS* sequences, all *Amaranthus* non-target sequences, and several *EPSPS* sequences downloaded from Genbank (Table 2.4)

Results

Sequence Alignment

The amino sequence of EPSPS (excluding the leader peptide) of *E. coli* and several Eudicot species from Genbank was compared to *Amaranthus* sequences (Figure 2.1). With the exception of site 336 all active site residues were completely conserved in all individuals. Also the proline at position 106 that has often been implicated in glyphosate resistance is conserved in all individuals examined. However, there was variation in many non-active site residues even within *Amaranthus* and between individuals representing the same species. This suggests that while active site residues were highly conserved such strong purifying selection was not consistent across the whole gene.

Structural Context of Conserved and Variable Amino Acids

ConSerf was used to model amino acid conservation with respect to structure (Ashkenazy *et al*. 2010). Within *Amranthus* (Figure 2.2) we found that the majority of sites were highly conserved, including those in the active site. The variable sites were all located on the outside surface of the protein consistent with their being less involved in structure and function.

When the model was expanded to include more distantly related plant species' amino acid sequences (Figure 2.3) the basic pattern remained: the inner amino acids were all highly conserved with variable amino acids on the outside on the protein. While yellow sites are colored as such to indicate that the ConSerf algorithm had difficulty calculating the conservation of these sites a comparison with the MSA suggests that these sites are highly variable.

Bayesian Estimate of Site-by-Site Mutation Rate

When *Beta vulgaris* was included in the analysis 154/518 codons were estimated to be under pervasive purifying selection with a posterior probability of at least 90%. There were also two sites under pervasive positive selection. When *Beta vulgaris* was excluded from the analysis and only *Amaranthus* individuals were considered 42/518 codons were estimated to be under pervasive purifying selection (posterior probability greater than or equal to 90%) and no sites were identified as having positive selection. All the sites identified as being under pervasive negative selection with *Beta vulgaris* excluded were also estimated to be under negative selection when it was included. When these results were compared with information about the putative location of active site residues (based on the *E. coli* EPSPS protein structure), the sites under negative selection did not correlate with the sites identified as being part of the active site.

Site-by-site Mutation Rate Comparison between Genes

A statistical comparisons of the site-by-site estimates of α and β was generated by FUBAR analysis of the *Amaranthus EPSPS* with *Beta vulgaris* excluded (Table 2.5). Similar results were also found when *Beta vulgaris* was included; these results are not shown.

The comparison of ω showed a statistically significantly lower average ω for *EPSPS* compared to the four non-target genes (Table 2.5). The mean value of ω for *EPSPS* was 0.444918 (standard error 0.0149, 95% CI 0.41569 - 0.47415). The mean values for *A07*, *A36*, and *A40* were all statistically significantly higher. The comparison of β showed a statistically similar average β for *EPSPS* compared to the four non-target genes (Table 2.5). The mean value of β for *EPSPS* was 0.148375 (standard. error 0.00444, 95% CI 0.13966 - 0.15709). This value was not statistically different from *A07*, *A37*, or *A40*. It was statistically significantly lower than *A36*. *EPSPS* did not appear to have more stringent sequence constraint than the non-target genes based on the rate of accumulation of non-synonymous mutations. The comparison of α showed a higher average α for *EPSPS* compared to the four non-target genes (Table 2.5). The mean value of α for *EPSPS* was 0.611964 (standard. error 0.04032, 95% CI 0.53287 - 0.69106). This was the highest of the five values and was statistically significantly higher than *A07* and *A40*. The mean value of α for *EPSPS* was statistically significantly higher than the mean value for the four non-target loci together. Analysis of means for variance of α showed that *EPSPS* has statistically more site-by-site variation in α than average among the five loci (Figure

2.4). Taken together, the analysis of α and β suggest that the significantly lower ω was not due to stronger sequence constraint on *EPSPS* but rather due to a higher than average rate of synonymous substitution accumulation, or more specifically to the presence of a few sites with very high synonymous mutation rates.

CpG Cluster Detection

The GC content of each gene was studied to understand the elevated rate of synonymous substitutions in the *EPSPS* gene. Each of the five loci had an average GC content of 44-46%. However, the *EPSPS* gene was unique among the five loci studied in its long stretches of sequence that were statistically enriched for CpG dinucleotides. Loci *A36* and *A07* had no CpG clusters, locus *A37* had a CpG cluster in the first half of the gene, and in locus *A40* half of individuals had a small (200bp) cluster near the end of the gene.

To see if the high concentration of CpG clusters in *EPSPS* was unique to *Amaranthus* we interrogated other plant *EPSPS* sequences from Genbank. Approximately half of the species examined contained CpG clusters: *Plantago lanceolata* (Ribwort Plantain), *Erigeron annus* (Daisy Fleabane), *Capsicum annuum* (hot pepper), *Convolvulus arvensis (*field bindweed*), Brassica rapa subsp. oleifera* (biennial turnip rape), *Vitis vinifera* (wine grape), *Calystegia hederacea* (Japanese False Bindweed), *Helianthus salicifolius* (willow leaf sunflower), *Conyza bonariensis* (hairy fleabane), *Arabidopsis thaliana* (mouseear cress), *Arabidopsis lyrata* (northern rock cress), and *Triticum aestivum* (domestic wheat). Approximately half of the species did not contain clusters*: Sarracenia purpurea* (purple

pitcher plant), *Camptotheca acuminate* (happy tree), *Populus tricocarpa* (black cotton wood), *Gossypium hirsutum* (domestic cotton), *Orychophragmus violaceus* (violet orychophtagmus), *Fagus sylvatica* (common beach), *Conyza Canadensis* (Canadian horse weed), *Ricinus communis* (castorbean), and *Solanum lycopersicum* (tomato) (Table 2.6). Among the species with clusters there was variation in the amount of sequences identified as being in a putative CpG cluster; some sequences were as CpG rich as the *Amaranthus EPSPS*, while others only had small regions identified (Table 2.6). The presence of CpG clusters does not appear to be related to the taxonomy of the species, nor to its status as a cultivar.

Discussion

The active site residues of EPSPS were very highly conserved; only one active site residue (336) was different between *E. coli* (asparagine) and all plants (serine) investigated. No active site residues were variable among plants. The proline-106 implicated in glyphosate resistance was not mutated in any of the *Amaranthus* species investigated. This pattern suggests that there was strong conservation of the active site in this gene even when looking at species as distantly related as *E. coli* and *Arabidopsis*. This is consistent with our current understanding of "important genes"; in general, genes that are highly expressed or have a strong negative phenotype—including lethality—when knocked out evolve slowly (Jordan *et al*. 2002; Pál *et al*. 2001; Wall *et al*. 2005; Wolf. 2006; Zhang and He. 2005). This variation in mutation rate is likely to be even more fine-grained than gene-by-gene; it is reasonable to extrapolate that functionally

important regions of genes will accumulate mutations more slowly than less critical regions. Again, this was what we observed in *EPSPS*; the sites that are variable, within *Amaranthus* and within eudicots, are found on the outer surface of the enzyme. These positions are likely to be less critical for enzyme conformation and catalysis and thus under less stringent purifying selection.

Based on analysis of sequence with the FUBAR method there were many sites identified as being under pervasive purifying selection. Surprisingly, most (16/20) of the putative active site residues were not among the sites identified. This is particularly unexpected considering the amino acid sequence alignment and the relative importance of the active site residues of EPSPS. However, careful investigation of the values estimated for α and β suggest that this may have been be due to the active site residues having low values for α which keeps ω from being small enough to achieve statistical significance. This is consistent with other observations that important residues change slowly (Jordan *et al*. 2002; Pál *et al*. 2001; Wall *et al*. 2005; Wolf. 2006; Zhang and He. 2005).

While several regions of *EPSPS* gene accumulate mutations at a slow rate, FUBAR analysis showed that on average *EPSPS* accumulated synonymous mutations at a higher rate than the other four loci investigated and that there was more site-by-site variance in the synonymous mutation accumulation rate (α). Based on this observation and the 10-fold higher mutation rate of methlyated CpGs, we investigated *Amaranthus EPSPS* for CpG clusters (Hodgkinson and Eyre-Walker. 2011). Plants are known to use methylated

CpGs for transcriptional regulation. In plants, full methylation of the upstream promoter region of a gene will effectively silence the gene; however, intermediate levels of methylation in the exonic sequence are associated with high levels of transcription—higher than either fully methylated or fully unmethylated genes (Antequera and Bird. 1988). It is reasonable to postulate that *EPSPS* may have a large number of methylated CpGs in its exonic sequence, given that it is an important and highly expressed gene. Analysis of CpG clusters suggests that *EPSPS* is rich in CpG clusters, if a portion of these CpGs are methylated then these highly mutable methyl-CpG sites may be the cause of *EPSPS*'s elevated rate of synonymous mutation accumulation. The rate of non-synonymous mutation accumulation is unaffected, likely because natural selection removes many of these mutations.

Natural selection likely does not favor the CpG richness of the *EPSPS* gene *per se.* An among-species comparison of available *EPSPS* sequence uploaded to Genbank showed that many species have lost CpGs to the point of no longer meeting the threshold to be considered a potential CpG island (at least 60% the expected number of CpG dinucleotides relative to random sequence and a 40% or higher GC content). The presence of CpG clusters in the *EPSPS* gene suggests an interesting potential hypothesis regarding the *EPSPS* copy number proliferation observed in glyphosate resistant amaranth (Gaines *et al*. 2010). Plants change their methylation patterns in response to stress (Lukens and Zhan. 2007; Madlung and Comai. 2004), and unmethylated CpG islands often function as initiation sites for genome replication (Delgado *et al*. 1998).

Small DNA fragments that match to replication initiation sites are often found in the nucleus due to aborted replication (Delgado *et al*. 1998). Under stress a cell could use non-homologous end joining to incorporate these fragments into the genome. If stress, such as that caused by glyphosate application, causes *EPSPS* to become a replication initiation site via demethlyation of CpG sites this could provide a mechanism for copy number proliferation that is independent of transposable elements. This would be consistent with the finding that the *EPSPS* insertion region does not contain signatures of recent transposable element activity (Gaines *et al*. 2013). Clearly more investigation into the possibility of this mechanism is needed; future work should focus on looking for methylation patterns of *EPSPS* in glyphosate stressed and unstressed plants using bisulfate sequencing and looking for the nuclear DNA fragments suggestive of *EPSPS* as a replication initiation site via Illumina sequencing of extra-chromosomal DNA.

If aborted replication fragment mediated gene duplication is identified as the mechanisms of gene amplification in glyphosate resistant *Amaranthus* it would have profound repercussion for the field of genetics. Such a mechanism has never before been described in any system. Particularly such a mechanism could have a significant impact of how we understand adaptation to stresses such as herbicides, insecticides, other pesticides (such as Warfarin), and drugs (such as chemotherapy or antibiotics), as stress in and of itself could be increasing the genomic instability in stressed populations and increasing the rate of evolution.

Table 2.1: Taxa sampled, including USDA germplasm repository accession numbers, for this study.

Species	Accession	
A. palmeri	Ames15298	PI549158
A. caudatus	Ames13860	PI553073
A. tricolor	Ames15326	PI477918
A. powellii	PI572260	PI572261
A. tuberculatus	PI603881	PI553086
A. spinosus	PI632248	PI642740
A. crassipes	PI642743	PI649302
A. viridis	PI536439	PI652434

Table 2.2: Primer pairs used to PCR amplify and then Sanger sequence EPSPS

Primer name	5'→ 3' Sequence
cDNA	
EPSPS_F	GCCAAGAACACAAAGCGAAATTCAGA
EPSPS_1749R	TCAAATCAAAACCTTCGRCGTA
Exon 1	
EPSPS_F	GCCAAGAACACAAAGCGAAATTCAGA
EPSPS_ex1R	CACCCAAAACAGAATCACGA
Exon 2	
EPSPS_ex2F	ATTGTCCCTGCTTTCACGTC
EPSPS_ex2R	ATTTCAGGGGTACGGCTTCT
Exon 3	
EPSPS_ex3F	TGTGTTCCTTTGGGGTCATT
EPSPS_g2450R	AGCTCATATCCCGGGTTTCT
Exon 4-6	
EPSPS_ex4-6F	GGAGGTAAAGTTGCATGTTGG
EPSPS_ex4-6R	CATTGGGGACAGCAAAAATC
Exon 7-8	
EPSPS_ex7-8F	ACTTTCGGAATGAGGAAGCA
EPSPS_1749R	TCAAATCAAAACCTTCGRCGTA

Table 2.3: Genbank sequences included in ConSef analysis.

Species	Accession Code
Arabidopsis thaliana	NP_182055.1
Arabidopsis lyrata	XP_002880170.1
Vitis vinifera	NP_001268176.1
Ricinus communis	XP_002511692.1
Populus trichocarpa	XP_002301279.1
Solanum lycopersicum	XP_004229803.1

Table 2.4: Genbank sequences analyzed for CpG content.

Species	Accession Code	Sequence length
Sorghum halepense	HQ436354.1	1,335
Lolium multiflorum	DQ153168.2	1,316
Triticum aestivum	EU977181.1	1,789
Ricinus communis	XM_002511646	1557
Populus trichocarpa	XM_002301243.1	1,908
Fagus sylvatica	DQ166525.1	2,085
Gossypium hirsutum	EU194952.1	3,344
Orychophragmus violaceus	AF440389.1	1,758
Brassica rapa	AY512663.1	1,726
Arabidopsis thaliana	NM_103780	1,886
Arabidopsis lyrata	XM_002880124	1,831
Vitis vinifera	GU060646.2	1,566
Plantago lanceolata	AY545665.1	795
Calystegia hederacea	EU526078.1	1,753
Convolvulus arvensis	EU698030.1	1,563
Solanum lycopersicum	M21071.1	2,045
Capsicum annuum	JN160845.1	1,750
Helianthus salicifolius	AY545661.1	792
Erigeron annuus	AY545659.1	1,074
Conyza canadensis	AY545668.1	1,452
Conyza bonariensis	EF200074.1	1,338
Camptotheca acuminata	AY639815.1	1,748
Sarracenia purpurea	AY545663.1	795

Table 2.5: Statistical analysis of site-by-site estimates of α, β, and ω within *Amaranthus* by FUBAR. Alpha is the estimate of synonymous divergence among the individuals investigated, β is the estimate of non-synonymous divergence, and ω is the ratio of β/α. Connecting letters reports show which values are statistically similar (they share a letter) and which are statistically distinct (they have no common letters).

Means for Oneway ANOVA on α

Level	Number	Mean	Std Error	Lower 95% CL	Upper 95% CL	Connecting letters report		
A07	247	0.479595	0.05659	0.36860	0.59059	A	B	
A36	269	0.555539	0.05423	0.44918	0.66190	A	B	
A37	285	0.507368	0.05268	0.40403	0.61070	A	B	
A40	234	0.433333	0.05814	0.31929	0.54737		B	
EPSPS	443	0.604228	0.03908	0.52758	0.68088	A		

Means for Oneway ANOVA on β

Level	Number	Mean	Std Error	Lower 95% CL	Upper 95% CL			
A07	247	0.149676	0.00595	0.13801	0.16134		B	C
A36	269	0.203234	0.00570	0.19205	0.21441	A		
A37	285	0.137719	0.00554	0.12686	0.14858			C
A40	234	0.159786	0.00611	0.14780	0.17177		B	
EPSPS	443	0.148375	0.00444	0.13966	0.15709		B	C

Means for Oneway ANOVA on ω

Level	Number	Mean	Std Error	Lower 95% CL	Upper 95% CL			
A07	247	0.525225	0.01996	0.48608	0.56437		B	
A36	269	0.532598	0.01912	0.49509	0.57011		B	
A37	285	0.489136	0.01858	0.45269	0.52558		B	C
A40	234	0.593459	0.02050	0.55324	0.63368	A		
EPSPS	443	0.444918	0.01490	0.41569	0.47415			C

Table 2.6: The location and size of potential CpG clusters identified in EPSPS sequences. Columns roughly represent 1-400bp, 400-600, 500-800, 800-1200, 1100-1300, 1200-1400, and 1300-1500. For each individual the specific location of predicted clusters in given in the best fit column to show the similarities and differences in cluster location between individuals. Clusters were calculated using the method of Gardiner-Garden and Frommer (1987); clusters larger than 200bp are the result of merging adjacent clusters with gaps of 10bp or less.

Species	**Sequence length**	**CpG clusters**						
A. palmeri Ames15298	1,554	426 - 646		993 - 1203	1109 - 1309	1144 - 1347	1234 - 1477	1301 - 1531
A. palmeri PI549158	1,555					1147 - 1346	1234 - 1440	1296 - 1531
A. caudatus Ames13860	1,556			990 - 1206	1108 - 1317		1278 - 1477	1301 - 1531
A. caudatus PI553073	1,557			990 - 1211	1108 - 1330		1278 - 1477	1301 - 1531
A. tricolor Ames15326	1,558						1234 - 1477	
A. tricolor PI477918	1,559						1234 - 1477	
A. powellii PI572260	1,560			990 - 1205	1109 - 1317		1241 - 1477	1301 - 1531
A. powellii PI572261	1,561	415 - 645		990 - 1211	1108 - 1330		1241 - 1477	1301 - 1531
A. tuberculatus PI603881	1,562	417 - 624		993 - 1202	1109 - 1308		1234 - 1537	
A. tuberculatus PI553086	1,563	426 - 653	500 - 724				1234 - 1524	
A. spinosus PI632248	1,564	415 - 659	500 - 729			1147 - 1346	1234 - 1440	1296 - 1531
A. spinosus PI642740	1,565	415 - 659	500 - 729			1147 - 1346	1234 - 1440	1296 - 1531
A. crassipes PI642743	1,566				1108 - 1329			1306 - 1531
A. crassipes PI649302	1,567			960 - 1179		1158 - 1383		
A. viridis PI536439	1,568				1104 - 1342		1240 - 1454	
A. viridis PI652434	1,569				1104 - 1342			1306 - 1531

Species	Sequence length	CpG clusters					
Ricinus communis	1,557						none
Populus trichocarpa	1,908						none
Fagus sylvatica	2,085						none
Gossypium hirsutum	3,344						none
Orychophragmus violaceus	1,758						none
Brassica rapa	1,726	1 - 298	384 - 601				
Arabidopsis thaliana	1,886			514 - 740			
Arabidopsis lyrata	1,831		433 - 632	495 - 706			
Vitis vinifera	1,566	35 - 435		508 - 743			
Plantago lanceolata	795	8 - 209					
Calystegia hederacea	1,753	93 - 445			969 - 1168	1026 - 1225	
Convolvulus arvensis	1,563	23 - 405					
Solanum lycopersicum	2,045						none
Capsicum annuum	1,750				947 - 1197		
Helianthus salicifolius	792	123 - 322		511 - 792			
Erigeron annuus	1,074	156 - 404		516 - 757			
Conyza canadensis	1,452						none
Conyza bonariensis	1,338		302 - 518				
Camptotheca acuminata	1,748						none
Sarracenia purpurea	795						none
Sorghum halepense	1,335	1 - 215	240- 559		885 - 1335		
Lolium multiflorum	1,316	1 - 503			963 - 1316		
Triticum aestivum	1,789	1 - 468	414 - 652			1103 - 1533	

Figure 2.1: Multiple sequence alignment of EPSPS amino acid sequence. Highlighting in *E. coli* sequence: lime – active site residues; dark blue – hinge residues (these are involved in the conformation change upon substrait binding); pink, tan, and blue – the three structural domains of EPSPS. Highlighting in plant sequences: dark gray – putative active site residues, light gray – putative hinge residues. Dots in leu of single letter amino acid code indicate homology with the *Amaranthus palmeri* Ames15298 sequence. While there is sequence divergence between the included individuals, there are no amino acid differences between any individuals—including *E. coli*—at the active site and hinge residues. All abbreviated species names are members of genus *Amaranthus*.

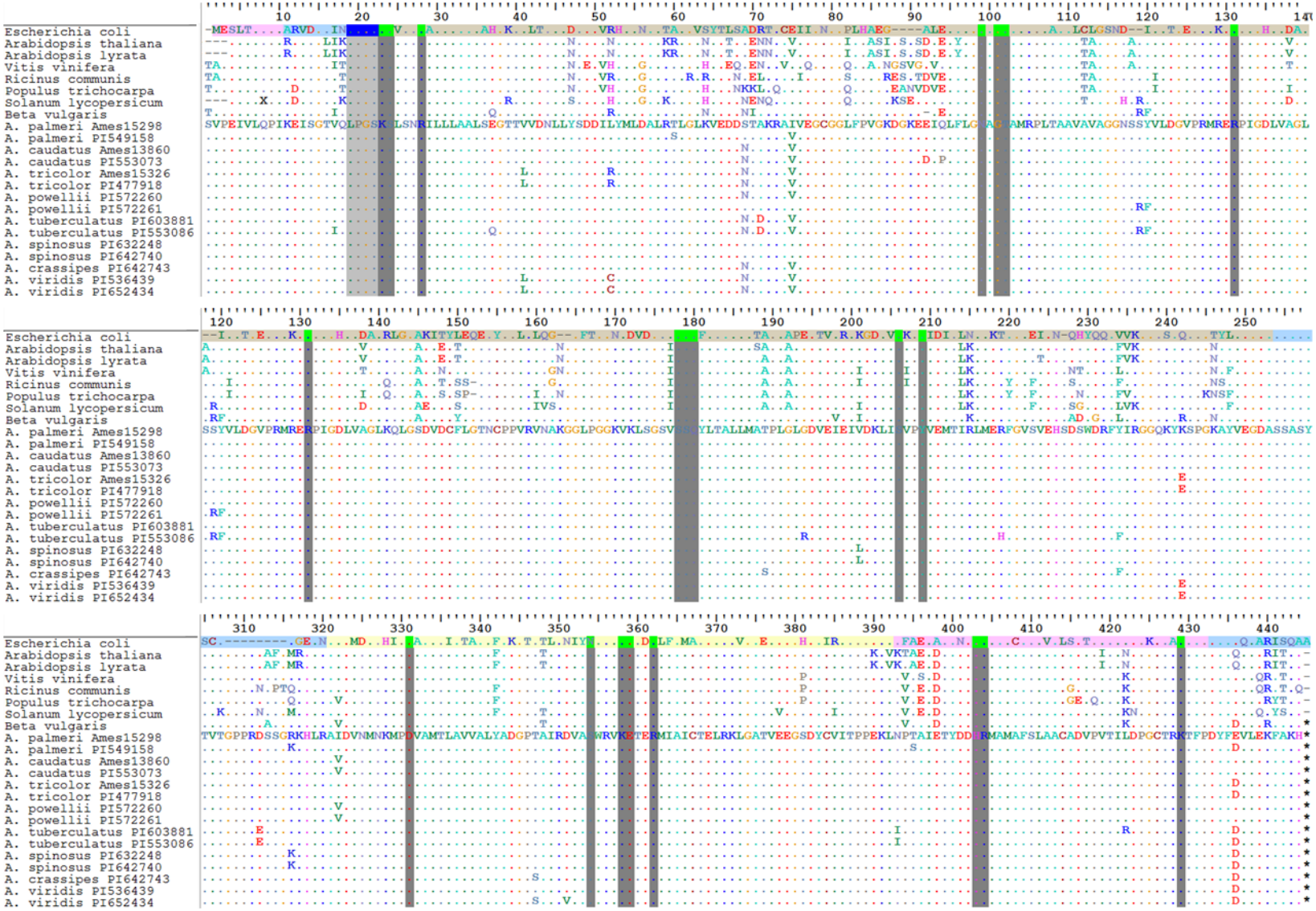

Escherichia coli
Arabidopsis thaliana
Arabidopsis lyrata
Vitis vinifera
Ricinus communis
Populus trichocarpa
Solanum lycopersicum
Beta vulgaris
A. palmeri Ames15298
A. palmeri PI549158
A. caudatus Ames13860
A. caudatus PI553073
A. tricolor Ames15326
A. tricolor PI477918
A. powellii PI572260
A. powellii PI572261
A. tuberculatus PI603881
A. tuberculatus PI553086
A. spinosus PI632248
A. spinosus PI642740
A. crassipes PI642743
A. viridis PI536439
A. viridis PI652434

Figure 2.2: Consensus protein structure model of *Amaranthus* EPSPS based on *E. coli* EPSPS structure bound to glyphosate and shikimate-3-phosphate (PDB 1G6S) and colored based on ConSerf score. Figure 2.2a is a space filling model and figure 2.2b is wire frame to allow internal residues to be more visible. ConSerf scores are calculated based on an amino acid alignment are range from 1 (highly variable) to 9 (highly conserved). In *Amaranthus* EPSPS the majority of residues are conserved, and all internal residues are conserved; the only variable residues are located on the outer surface of the enzyme.

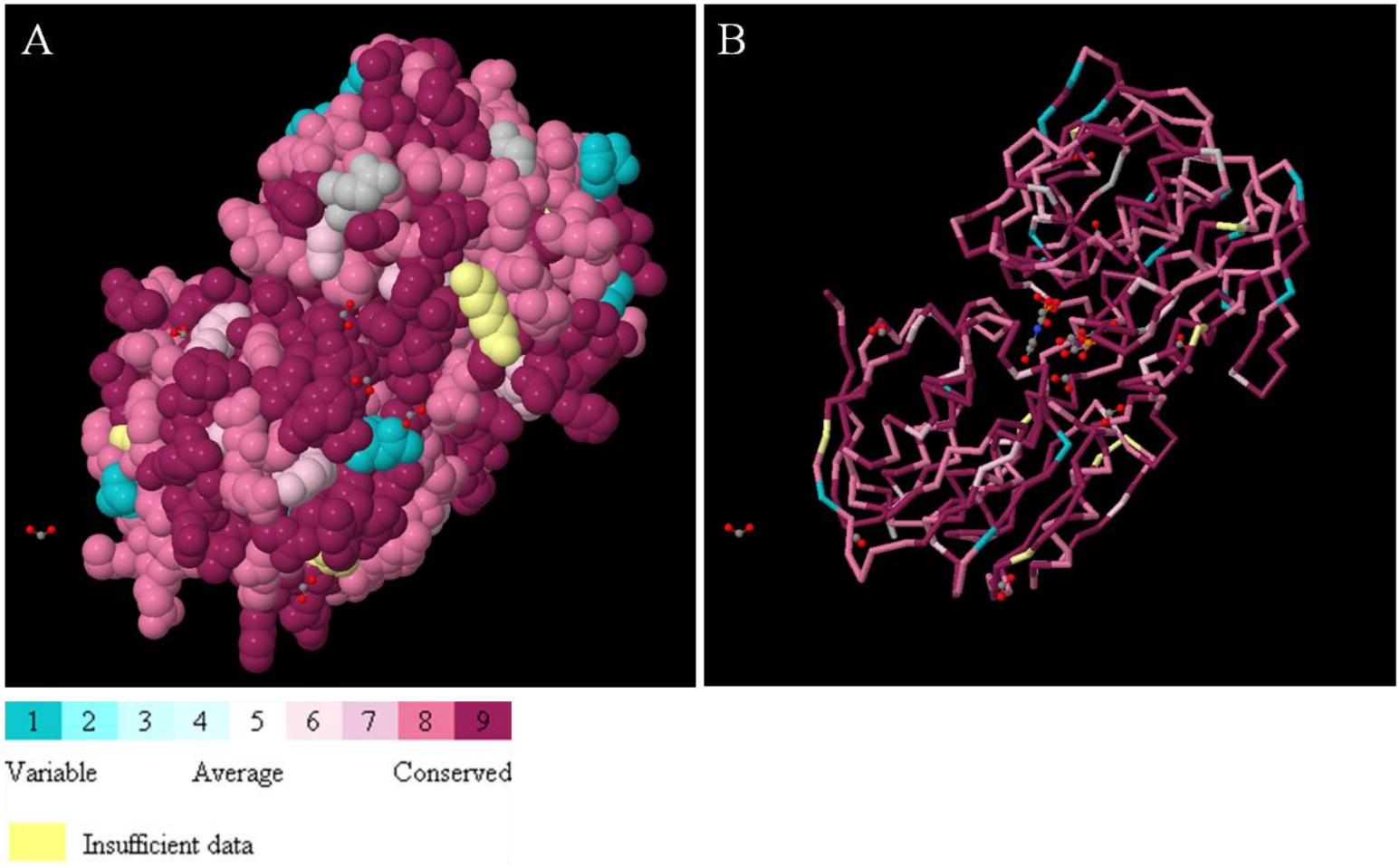

A
B
1 2 3 4 5 6 7 8 9
Variable
Average
Conserved
Insufficient data

Figure 2.3: Consensus protein structure model of EPSPS based on *E. coli* EPSPS structure bound to glyphosate and shikimate-3-phosphate (PDB 1G6S) and colored based on ConSerf score. Alignment includes the six sequences listed in Table 2.3 and two *Amaranthus* sequences. Figure 2.3a is a space filling model and figure 2.3b is wire frame to allow internal residues to be more visible. ConSerf scores are calculated based on an amino acid alignment are range from 1 (highly variable) to 9 (highly conserved). Yellow residues are determined by ConSerf to have insufficient data to calculate conservation, when the amino acid alignment is inspected it can be seen that these are highly variable sites. In this EPSPS alignment the majority of internal residues are conserved, and the variable residues are located primarily on the outer surface of the enzyme.

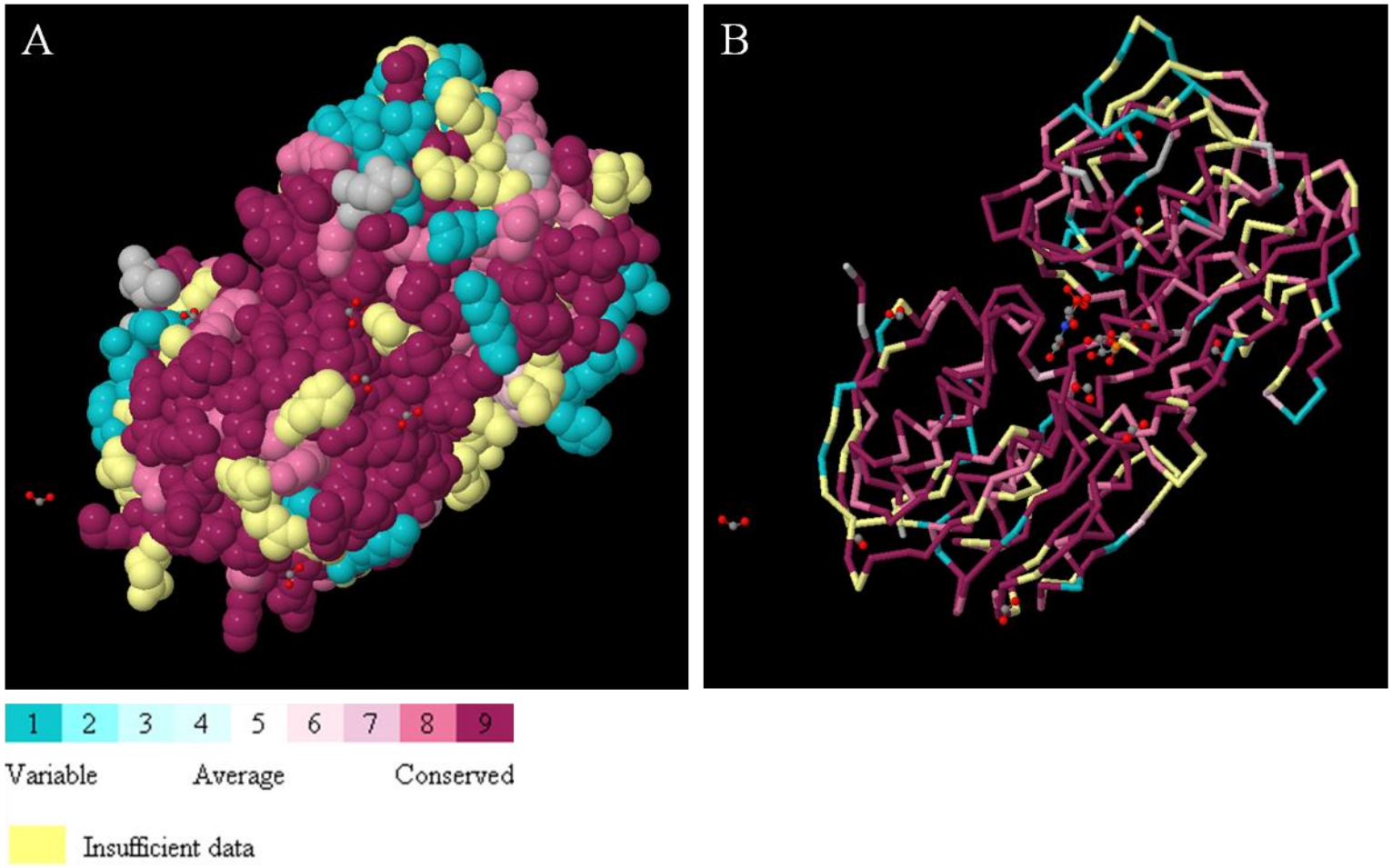
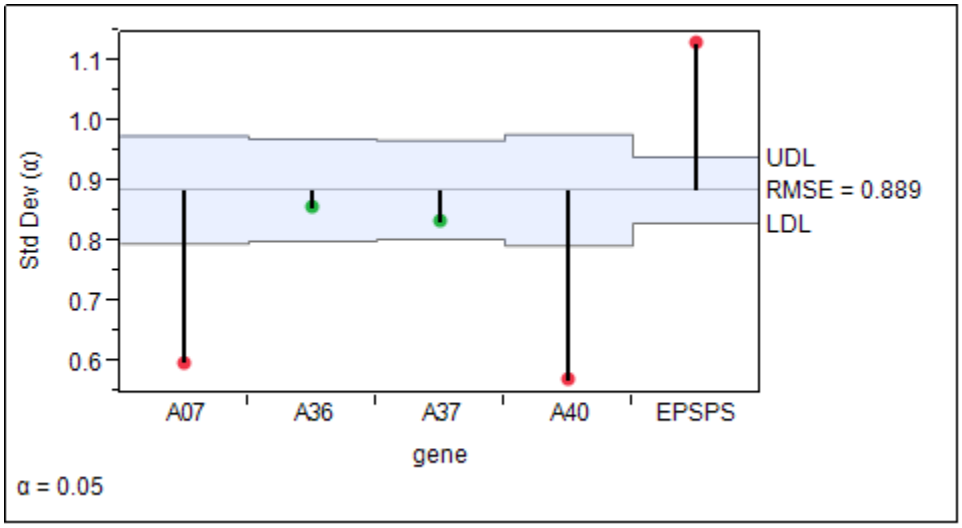

A
B
1 2 3 4 5 6 7 8 9
Variable
Average
Conserved
Insufficient data

Figure 2.4: Analysis of means for variance of α calculated in JMP. The RMSE (root mean square error) is the overall variance in α for all sites included in the analysis; the UDL and LDL (upper and lower decision limit) are marked by the bounds of the gray boxes and vary between loci based on the number of sites. Each point represents the RMSE of α for only the sites that are part of the given locus. Red indicates that the RMSE of the subgroup is statistically significantly different than the overall RMSE. EPSPS's α has a statistically significantly higher RMSE, indicating higher variance in observed values of α. (Note the α in the bottom corner indicates this analysis was done at the 95% confidence level)



References

Akaike, H., 1981 Likelihood of a model and information criteria. Journal of econometrics **16:** 3-14.

Akaike, H., 1974 A new look at the statistical model identification. Automatic Control, IEEE Transactions on **19:** 716-723.

Antequera, F., and A. P. Bird, 1988 Unmethylated CpG islands associated with genes in higher plant DNA. The EMBO Journal **7:** 2295.

Ashkenazy, H., E. Erez, E. Martz, T. Pupko and N. Ben-Tal, 2010 ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research **38:** W529-W533.

Baerson, S. R., D. J. Rodriguez, N. A. Biest, M. Tran, J. You *et al*, 2002 Investigating the mechanism of glyphosate resistance in rigid ryegrass (*Lolium ridigum*). Weed Science **50:** 721-730.

Brenner, D. M., D. D. Baltensperger, P. A. Kulakow, J. W. Lehmann, R. L. Myers *et al*, 2000 Genetic resources and breeding of *Amaranthus**. Plant Breeding Research **19:** 227-285.

Delgado, S., M. Gómez, A. Bird and F. Antequera, 1998 Initiation of DNA replication at CpG islands in mammalian chromosomes. The EMBO Journal **17:** 2426-2435.

Devonshire, A. L., and L. M. Field, 1991 Gene amplification and insecticide resistance. Annual Reviews in Entomology **36:** 1-21.

Ewing, B., and P. Green, 1998 Base-calling of automated sequencer traces UsingPhred. II. Error Probabilities. Genome Research **8:** 186-194.

Ewing, B., L. Hillier, M. C. Wendl and P. Green, 1998 Base-calling of automated sequencer traces UsingPhred. I. Accuracy Assessment. Genome Research **8:** 175-185.

Gaines, T. A., 2010 *Molecular Genetics of Glyphosate Resistance in Palmer Amaranth (Amaranthus Palmeri L.)*. Colorado State University.

Gaines, T. A., D. L. Shaner, S. M. Ward, J. E. Leach, C. Preston *et al*, 2011 Mechanism of resistance of evolved glyphosate-resistant palmer amaranth (*Amaranthus palmeri*). Journal of Agricultural Food Chemistry.

Gaines, T. A., A. A. Wright, W. T. Molin, L. Lorentz, C. W. Riggins *et al*, 2013 Identification of genetic elements associated with EPSPS gene amplification. PloS One **8:** e65819.

Gaines, T. A., W. Zhang, D. Wang, B. Bukun, S. T. Chisholm *et al*, 2010 Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proceedings of the National Academy of Sciences **107:** 1029-1034.

Gaylon D. Morgan, P. A. Baumann and J. M. Chandler, 2001 Competitive impact of palmer amaranth (*Amaranthus palmeri*) on cotton (*Gossypium hirsutum*) development and yield. Weed Technology **15:** 408-412.

Hall, T., 2001 Biolign alignment and multiple contig editor. North Carolina State University, Department of Microbiology, Raleigh, North Carolina.

Hall, T. A., 1999 *BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT*.

Heap, I. M., *International Survey of Herbicide Resistant Weeds*. Online. Internet. Wednesday, November 06, 2013 . Available www.weedscience.com

Hodgkinson, A., and A. Eyre-Walker, 2011 Variation in the mutation rate across mammalian genomes. Nature Reviews Genetics **12:** 756-766.

Jordan, I. K., I. B. Rogozin, Y. I. Wolf and E. V. Koonin, 2002 Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Research **12:** 962-968.

Lorraine-Colwill, D. F., S. B. Powles, T. R. Hawkes, P. H. Hollinshead, S. A. J. Warner *et al*, 2002 Investigations into the mechanism of glyphosate resistance in *Lolium rigidum*. Pesticide Biochemistry and Physiology **74:** 62-72.

Lukens, L. N., and S. Zhan, 2007 The plant genome's methylation status and response to stress: Implications for plant improvement. Current Opinions in Plant Biology **10:** 317-322.

Madlung, A., and L. Comai, 2004 The effect of stress on genome regulation and structure. Annals of Botany **94:** 481-495.

Michitte, P., R. De Prado, N. Espinosa and C. Gauvrit, 2005 Glyphosate resistance in a chilean *Lolium multiflorum*. Communications in Agcultral and Applied Biological Sciences **70:** 507-513.

Mosyakin, S. L., and K. R. Robertson, 1993 Amaranthus, pp. 410-435 in *Flora of North America North of Mexico*, edited by Flora of North America Editorial Committee. New York and Oxford.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Pond, S. L. K., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian appRoximation for inferring selection. Molecular Biology and Evolution, **30:** 1196-1205.

Padgette, S. R., K. H. Kolacz, X. Delannay, D. B. Re, B. J. LaVallee *et al*, 1995 Development, identification, and characterization of a glyphosate-tolerant soybean line. Crop Science **35:** 1451-1461.

Pál, C., B. Papp and L. D. Hurst, 2001 Highly expressed genes in yeast evolve slowly. Genetics **158:** 927-931.

Perez-Jones, A., K. Park, N. Polge, J. Colquhoun and C. Mallory-Smith, 2007 Investigating the mechanisms of glyphosate resistance in *Lolium multiflorum*. Planta **226:** 395-404.

Plot, C., 2000 The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques **28:** .

Rambaut, A., and A. J. Drummond, 2007 *Tracer v1.4*. Available from http://beast.bio.ed.ac.uk/Tracer.

Ribeiro, D. N., Z. Pan, S. O. Duke, V. K. Nandula, B. S. Baldwin *et al*, 2013 Involvement of facultative apomixis in inheritance of EPSPS gene amplification in glyphosate-resistant *Amaranthus palmeri*. Planta 1-14.

Ronquist, F., and J. P. Huelsenbeck, 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:** 1572-1574.

Rowland, M. W., Don S. Murray and L. M. Verhalen, 1999 Full-season palmer amaranth (*Amaranthus palmeri*) interference with cotton (*Gossypium hirsutum*). Weed Science **47:** 305-309.

Salas, R. A., F. E. Dayan, Z. Pan, S. B. Watson, J. W. Dickson *et al*, 2012 EPSPS gene amplification in glyphosate-resistant italian ryegrass (*Lolium perenne* ssp multiflorum) from arkansas, USA. Pest Management Science.

Schönbrunn, E., S. Eschenburg, W. A. Shuttleworth, J. V. Schloss, N. Amrhein *et al*, 2001 Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proceedings of the National Academy of Sciences of the United States of America **98:** 1376-1380.

Steinrücken, H. C., and N. Amrhein, 1980 The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. Biochemical and Biophysical Research Communications **94:** 1207-1212.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei *et al*, 2011 MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution **28:** 2731-2739.

Wall, D. P., A. E. Hirsh, H. B. Fraser, J. Kumm, G. Giaever *et al*, 2005 Functional genomic analysis of the rates of protein evolution. Proceedings of the National Academy of Sciences of the United States of America **102:** 5483-5488.

Weaver, L. M., and K. M. Herrmann, 1997 Dynamics of the shikimate pathway in plants. Trends Plant Science **2:** 346-351.

Wolf, Y. I., 2006 Coping with the quantitative genomics 'elephant': The correlation between the gene dispensability and evolution rate. Trends in Genetics **22:** 354-357.

WSSA Standardized Plant Names Subcommittee, 2010 *January 2010 Revision of the Composite List of Weeds*. Weed Science Society of America.

Zhang, J., and X. He, 2005 Significant impact of protein dispensability on the instantaneous rate of protein evolution. Molecular Biology and Evolution **22:** 1147-1155.

CHAPTER III

GENETIC SIGNATURES OF POPULATION STRUCTURE SHOW EVIDENCE OF PARALLEL ADAPTATION OF NORTH CAROLINA *A. palmeri* TO GLYPHOSATE STRESS

K. E. Beard, K. S. Lay, J. D. Burton, N. Burgos, and A. L. Lawton-Rauh

Abstract

Premise of the Study

After glyphosate resistant *A. palmeri* was first identified in Macon, GA massive proliferation in copy number of the *EPSPS* gene was identified as the most likely mechanism of resistance (Gaines *et al*. 2011; Gaines *et al*. 2010). This study investigated whether the same mechanism was responsible for glyphosate resistance in North Carolina *A. palmeri* and tests whether the source of adaptation is independent or shared amongst individuals.

Methods

Genomic DNA was collected from *A. palmeri* seeds representing 31 localities in NC, and 29 accessions representing 14 species of amaranth from the USDA germplasm repository with no history of glyphosate exposure. This DNA was then used for quantitative real-time PCR to determine genomic copy number of *EPSPS*, and to sequence four genomic loci for population structure analysis. Copy numbers were then compared with glyphosate resistance levels of the sampled locations/accessions.

Key Results

Many of the NC *A. palmeri* individuals included for study have multiple copies of the *EPSPS* gene. When individuals are grouped based on documented resistance to injury or death from glyphosate, a statistically significant association between copy number and resistance level is found. None of the investigated amaranth accessions show evidence of

EPSPS copy number proliferation. Estimation of the best fit number of populations (population structure analysis) for the NC *A. palmeri* suggests five clusters, four of which are statistically significantly associated with a resistance phenotype.

Conclusions

Palmer amaranths with elevated *EPSPS* gene copy number are present in NC; it is likely that this elevated copy number is responsible for documented glyphosate resistance of this species in the area. There is no evidence that the observed increase *EPSPS* gene copy number is part of standing variation in the *Amaranthus* genus. However, the fact that four of the five identified population clusters in NC are statistically associated with increased glyphosate resistance suggests that more than one adaptive event may be responsible for the observed resistance in NC. Current work cannot determine if these represent multiple *de novo* events in NC, multiple introgression events, or some combination. Further work to compare NC genotypes to those found in other parts of the southeast is needed to fully understand the dynamics of the proliferation of glyphosate resistance via increased *EPSPS* gene copy number in *A. palmeri*.

Introduction

One of the most fundamental questions of genetics regards the mechanism and evolutionary dynamics underlying adaptation. While the primary source of variation is always mutation, the dynamics of mutation accumulation and retention can have important impacts on the observed trajectories of adaptation. This question has been

investigated in many systems—including mammals, fish, birds, insects, and plants—and examples have been found that fit each of the three sources of adaptive mutations: standing variation (Barrett and Schluter. 2008; Catania *et al*. 2004; Schlenke and Begun. 2004; Shimizu *et al*. 2004; Steiner *et al*. 2007), introgression (Evans *et al*. 2006; Kays *et al*. 2010; Kenneth D. Whitney *et al*. 2006), and *de novo* mutation in the stressed population (Colosimo *et al*. 2004; Cresko *et al*. 2004; Dowling *et al*. 2002; Eizirik *et al*. 2003; Feldman *et al*. 2009; Mundy *et al*. 2004; Steiner *et al*. 2007; Strecker *et al*. 2003; Theron *et al*. 2001). Of these three sources standing ancestral variation is expected to be the source of the most rapid adaptation, because populations should be able to respond to novel stress quickly when the allele is already present in the population at a relatively high frequency—compared to the frequency it would have as a *de novo* mutation or recent immigrant (Barrett and Schluter. 2008; Innan and Kim. 2004). This is also expected to be the primary source of adaptive alleles of minor affect or recessive inheritance (Barrett and Schluter. 2008).

To investigate the sources of adaptive variation, we used a system with strong clear selection pressure, rapid adaptation, and considerable economic importance: resistance of weedy *Amaranthus palmeri* to the herbicide glyphosate (primary active ingredient of Monsanto's RoundUp™). Glyphosate was introduced to the market in the 1970s. It competitively inhibits 5-enolpyruval shikimate 3-phosphate synthase (EPSPS) by binding to the EPSPS-shikimate 3-phosphate complex in place of phosphoenolpyruvate (PEP). The shikimate pathway starts with erythrose 4-phosphate and ends with the formation of

chorismate, the precursor to quinones, folates, and the aromatic amino acids phenylalanine, tyrosine, and tryptophan. This pathway is only found in plants and bacteria. While bacteria use the pathway almost exclusively for the synthesis of amino acids, plants use chorismate as a precursor for various pigments, defense compounds, and lignin (Weaver and Herrmann. 1997).

RoundUp Ready™ (genetically engineered glyphosate tolerant) crops were introduced to the market in the late 1990s (U.S. Patent 5633435), at that point there were no documented cases of glyphosate resistance. However, within just a few seasons of the introduction of glyphosate tolerant crops and a switch by growers to a glyphosate dominated weed control regime resistant populations began to emerge (Heap. 2013). As of 2013, glyphosate resistance has been confirmed in 24 species (Heap. 2013). In weed species that have evolved glyphosate resistance, three primary mechanisms have thus far been identified. One mechanism is changes in translocation or absorption of the herbicide (Carvalho *et al*. 2011; Cruz-Hipolito *et al*. ; Feng *et al*. 2009; Lorraine-Colwill *et al*. 2002; Michitte *et al*. 2007; Michitte *et al*. 2005; Norsworthy *et al*. 2001; Perez-Jones *et al*. 2007; Wakelin *et al*. 2004; Yu *et al*. 2007). Another mechanism is mutation in the *EPSPS* gene coding sequence (Baerson *et al*. 2002a; Kaundun *et al*. 2011; Ng *et al*. 2003; Perez-Jones *et al*. 2007; Wakelin and Preston. 2006; Yu *et al*. 2007). Finally, the most recently identified mechanism is proliferation of genomic EPSPS gene copy number. This was first identified in a population of resistant *A. palmeri* from Macon, GA (Gaines. 2010; Gaines *et al*. 2011). Since then a few more populations using this mechanism for

resistance have been identified (Baerson *et al*. 2002b; Salas *et al*. 2012). There is also evidence of glyphosate metabolism as a source of resistance, but to date no populations have been identified that rely solely on that mechanism (de Carvalho *et al*. 2011; González-Torralva *et al*. 2012).

Among the twenty-four species identified as glyphosate resistant, three are members of the genus *Amaranthus*. Herbicide resistance in *Amaranthus* is an expensive and growing problem in the United States. In the Southeastern US, for example, an infestation density of just one *A. palmeri* per meter of row results in a 50% reduction in cotton yield (Gaylon D. Morgan *et al*. 2001; Rowland *et al*. 1999). In 2004 the first case of glyphosate resistance in *Amaranthus* was identified in Georgia. As of 2013 glyphosate resistant amaranth are found in seventeen states (Heap, I. 2013).

It was shown in 2010 that the glyphosate resistance observed in *A. palmeri* growing in Macon, GA was caused by massive copy number proliferation of the *EPSPS* gene (Gaines. 2010; Gaines *et al*. 2011). We investigated the source and spread of glyphosate resistance in Southeastern US amaranth. We first identified the pattern of shared versus independently derived mechanism of resistance in glyphosate resistant NC amaranth. Do they share the *EPSPS* gene copy number proliferation mechanism with the GA population? A shared mechanism suggests common ancestry or convergent independent adaptation. To distinguish between these we investigated whether *EPSPS* copy number variation originated from standing neutral variation in the genus *Amaranthus*. We also

analyzed population structure to determine if the resistant NC amaranths represent a single population or if there is evidence of multiple adaptive events. Taken together these data allow a better understanding of the source of glyphosate stress adaptation in amaranths: selection upon ancestral variation, a single spreading adaptive event, or multiple adaptive events. Not only will a better understanding about the spread of glyphosate resistant amaranth potentially help inform agricultural practice it will improve our understanding of the dynamics of adaptation and how it factors into divergence, speciation, and evolution.

Materials and Methods

Sampling of North Carolina A. palmeri

Initial sampling was performed in North Carolina by Alan York and David Jordan in 2010. From that collection seeds from 31 locations (Table 3.1 and figure 3.1) were used. Seeds were planted in flats and grown in a greenhouse under ambient light supplemented to achieve 16h days until they had at least four true leaves, at which point leaf tissue was collected for DNA extraction from 8-10 plants representing each location.

Sampling of USDA Amaranth Accessions

Seed samples were obtained from the USDA germplasm repository. We sampled 56 individuals from 29 accessions representing 14 *Amaranthus* species plus four individuals from two accessions representing one out-group species (Table 3.2). Seeds were planted out and grown in the greenhouse under ambient lighting supplemented to achieve 16h

days until they had at least four true leaves, at which point leaf tissue was collected and frozen at -80°C.

Assessment of Glyphosate Resistance

Glyphosate resistance of NC *A. palmeri* populations was assessed at North Carolina State University by whole plant response to glyphosate application using the method described by J.R. Whitaker (Whitaker. 2009) and provided by J. Burton. Glyphosate resistance of USDA amaranth samples was assessed by the Burgos lab at the University of Arkansas using the same whole plant response protocol.

DNA extraction

Approximately 100mg of tissue previously frozen at -80°C was ground in a mixer mill and then DNA was purified by the standard protocol of the Nucleospin Plant II DNA extraction kit (Macherey Nagel, Düren, Germany) was followed. DNA concentration, 260/280 ratio, and 260/230 ratio of the DNA was measured with a NanoDrop 1000. A successful DNA extraction was determined to have a concentration of at least 10 ng/µL, a 260/280 greater than 2.0, and a 260/230 ratio greater than 1.8. This protocol was repeated as needed to get all samples to meet this quality standard.

FTA card preparation

Of the 31 locations included in the study a sub-set of 15 locations that represent multiple levels of resistance as measured by both control/injury and plant mortality and were

spread across the state were chosen for *EPSPS* gene copy number analysis (Table 3.1 and Figure 3.1, bolded locations). Genomic DNA was extracted from eight individuals representing each location using FTA elute cards (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA). One to two healthy leaves were taken from the same meristem of each plant and pressed onto the FTA card. The cards were allowed to dry overnight and were stored at room temperature in plastic freezer storage bags with desiccant pouches (Fisher scientific) until they were used for analysis.

To elute the DNA from the FTA card the manufacturer's suggested protocol was used. A 3.0 mm disk was taken from the card, washed with sterile water, and then incubated with 30 µL of sterile water at 98°C for 30 minutes. The eluted DNA was either used immediately or stored at 4°C.

Quantitative PCR

The relative *EPSPS* gene copy number was determined using a real-time quantitative polymerase chain reaction. We elected to use two reference genes, *A36* and *ALS*. *A36* is a putative RNA dead-box helicase. This locus was initially developed as part of the study on the phylogeny of *Amaranthus* (see chapter I). Putative helicase function was based on the results of a BLAST comparison to all eudicot sequences in the non-redundant nucleotide database. The *ALS* gene codes for *acetolactate synthase*. Two reference genes were used for two reasons; first, some of the populations included are resistant to ALS inhibitors and while copy number proliferation of the *ALS* gene has never been

implicated as a resistance mechanism it remained a possibility that could complicate the interpretation of the data. However, we chose to still include *ALS* because it is the reference gene in the other published studies of *EPSPS* copy number (Gaines *et al*. 2011; Gaines *et al*. 2010; Ribeiro *et al*. 2013; Salas *et al*. 2012). The second reason including two reference genes was that when the organism being studied has few available genomic resources, it may be difficult to choose with certainty a reference gene that does not exhibit copy number variation. Using multiple reference genes avoids this problem (Ginzinger. 2002).

The control individual was determined using a standard curve of the three primer sets. From these curves and initial known concentration as determined by the NanoDrop 1000, the EPSPS, ALS, and A36 copy numbers for PI477918B (*A. tricolor*) were absolutely quantified and found to be matching copy number and presumed to be single copy. This accession was used as the control individual for all qPCR experiments.

For analysis of NC *A. palmeri* on FTA elute cards each 25 µL reaction was prepared using 12.5 µL of SYBR green super-mix, 2.5 µL of FTA eluted DNA, and 1uL each of 10uM forward and reverse primer. The primer sets for the *EPSPS* gene, EPSF1 (5′-ATGTTGGACGCTCTCAGAACTCTTGGT-3′) x EPSR8 (5′-TGAATTTCCTCCAGCAACGGCAA-3′) were designed based on sequences from *A. palmeri* (Gaines *et al*. 2010). The *ALS* primer sets, ALSF2 (5′-AGCTCTGGAACGTGAAGGC-3′) x ALSR2 (5′-TCAATTAAAACCGGTCCGGG-3′) were made from sequences of common cocklebur and ragweed (Tranel *et al*. 2004). The

A36 primers sets were designed using BioEdit and Primer3 software from sequences of the *Amaranthus* genus: A36_F244 (5'-TTGGAACTGTCAGAGCAACC-3') x A36_R363 (5'-GAACCCACTTCCACCAAAAC-3') (Hall. 1999; Rozen and Skaletsky. 2000). Each sample was run for each primer set in triplicate. All runs were performed by the same operator (KSL). The thermoprofile used, proposed by Chandi *et al* (2012), was 3 minute denaturation at 95 C, 40 cycles of 95 C for 30 seconds, and 60 C for one minute. This program was then followed by a melt curve analysis of 81 cycles of 55°C for 30 seconds. The cycle threshold (C_t) values were calculated using BioRad iQ5 Thermocycler software. All runs were completed on the iQ5 real time PCR thermocycler (BioRad).

For analysis of USDA amaranth accessions with spin column extracted DNA the 25 µL reactions were prepared using SYBR green super-mix and 1ng of genomic DNA template. All other parameters were identical to those used on the NC *A. palmeri.*

Statistical Analyses of qPCR Data

The relative EPSPS copy number was determined using SAS software (SAS institute Inc., Cary, NC, USA). The program ran a modified ANCOVA test based on a program by Yuan *et al.* (Yuan *et al*. 2006) which calculated point estimates, 95% confidence intervals, and P-values for the C_t data after applying the following formulas:

$$\Delta C_t = C_t \text{ of target (EPSPS or ALS)} - C_t \text{ of reference (A36)}$$

$$\Delta\Delta C_t = \Delta C_t \text{ of Treatment Individual} - \Delta C_t \text{ of Control (PI477918B)}$$

Because the program yielded $\Delta\Delta C_t$ values, the relative copy number was determined by taking the values and applying them to the following formula:

$$\text{Copy number} = 2^{-\Delta\Delta Ct}$$

The point estimates of the copy number for all populations were then analyzed by a second SAS program. This program determined the minimum, maximum, and quartile values for each population.

The statistical association between copy number and documented resistance was determined using SAS (SAS institute inc., Cary, NC, USA). The program for calculating $\Delta\Delta C_t$ values was expanded to average the values for individuals sharing resistance categories and to calculate the difference in those average values.

Southern Blot

Genomic DNA was re-isolated from eight plants representing two localities and a variety of copy numbers as determined by quantitative PCR of genomic DNA. Isolation was performed using a cetyltrimethyl ammonium bromide (CTAB) method (Luo *et al.* 1995). The isolated DNA was then digested with EcoRI according to the supplier's instructions (New England Biolabs, Beverly, MA, USA). Digested DNA was electrophoresed on a 0.8% agarose gel at 50V for 16hrs, transferred onto nylon membranes (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA), and hybridized to 32P-labled DNA probes of exon 1 of *EPSPS*. Hybridization was carried out in ExpressHyb buffer (Clontech, Mountain View, CA, USA) at 65°C following the manufacturer's suggested protocol.

Hybridizing fragments were detected by exposure of the membrane on a phosphor screen overnight at room temperature, and scanning on a Typhoon 9400 phosphor-imager.

DNA Amplification and Sequencing

Genomic DNA from all NC individuals was amplified using GoTaq Flexi (Promega, Madison, Wisconsin) at four loci using a unique primers pair for each locus (Table 3.3). The cycling conditions and Sanger sequencing protocol used for all loci were the same as described in chapter one. Sequences were Phred-Phraped to merge forward and reverse sequencing reads and viewed in Biolign (Ewing and Green. 1998; Ewing *et al*. 1998; Hall. 2001). After sequencing, several sites had two base calls suggesting heterozygosity. To resolve heterozygosity and estimate best haplotypes phase assignment the sequences were run through PHASE 2.1 (Stephens *et al*. 2001). The parameters used for haplotypes phasing were a run length of 100 generations after discarding 100 generations as burn-in, and an output acceptance threshold of 90%. Subsequent analyses were preformed on the sequences identified as the best pairs by analysis in PHASE.

Population Structure Estimation

The haplotype pairs identified by PHASE of the four loci for all NC *A. palmeri* individuals sequenced were input into Structure 2.3.4 (Hubisz *et al*. 2009; Pritchard *et al*. 2000). The analyses were conducted to test potential clustering with K (number of population clusters) from 1 to 10. The Admixture model was implemented with glyphosate resistance level as the clustering prior, and a single value for α was inferred

and shared between clusters. For each K value the clustering was estimated independently 10 times. Each estimate was from a run with 600,000 generations with the first 100,000 discarded as burn-in. Clustering patterns were then compared to other information about the collection locations including documented glyphosate resistance with JMP (SAS institute Inc., Cary, NC, USA).

Results

Copy number of control individual

To confirm the validity of using individual PI477918B as the control individual its *EPSPS, ALS,* and *A36* gene copy numbers needed to be determined using absolute instead of relative quantification. The primer efficiencies were all similar and sufficiently high (Table 3.4). The C_t values generated for each gene were fit to the *A36* standard curve to ensure they agreed about the starting quantity of DNA, as seen in figure 3.2, which was observed. PI477918B was used in all subsequent experiments as the control individual.

Copy number of EPSPS and ALS in NC A. palmeri

Figures 3.3 and 3.4 are box plot summaries of the gene copy numbers determined by qPCR. The upper and lower points represent the maximum and minimum observed copy number and the bounds of the boxes represent the first and third quartiles. Median values are given in Table 3.5. With the exception of a few individuals, there was no observed proliferation of *ALS* gene copy number. In EPSPS, however, considerable variation in estimated copy number was found. Values between 1 and 180 were observed with mean

value of 25.14 *EPSPS* copies across all individuals. To corroborate this observation DNA was analyzed by Southern Blot.

The single copy-number individual (3E, figure 3.5 lane 4) had one band present at about 5kb, all other individuals showed multiple bands of varying intensity. Comparing the copy numbers estimated by qPCR (Table 3.6) and the Southern Blot results (Figure 3.5) showed that all of the high copy-number individuals have a major band at ~3kb that varies in intensity between individuals. All the high-copy number individuals also have additional bands at different fragment sizes, including some as small as 150bp—the same size as the PCR product positive control. None of the high copy-number individuals had the 5kb band found in the normal copy-number individual (Figure 3.5).

Gene Copy Number vs. Documented Resistance

Plants from locations identified as susceptible based on plant injury or control had, on average, statistically significantly fewer copies of the *EPSPS* gene than plants from locations identified as resistant. There was no statistical significance to the difference in average EPSPS copy number between low and high levels of resistance based on plant injury or control (Figure 3.6 and table 3.7). The average EPSPS copy number for plants from locations described as susceptible was 8.378 (95% CI 7.258 to 9.671), and the averages for locations described as low and high level resistant were 34.672 (95% CI 28.290 to 42.494) and 37.402 (95% CI 29.590 to 47.275) respectively.

Plants from locations identified as resistant based on mortality had, on average, statistically significantly more copies of EPSPS than plants from locations identified as susceptible. Highly resistant locations also had statistically significantly more *EPSPS* copies than mildly resistant locations (Figure 3.7 and table 3.8). The average *EPSPS* copy number for plants from locations described as susceptible was 2.500 (95% CI 1.975 to 3.831), and the averages for locations described as low and high level resistant were 16.5783 (95% CI 12.648 to 21.730) and 41.449 (95% CI 36.199 to 47.462) respectively.

A summary of *EPSPS* copy number distribution broken apart by documented resistance both to injury and death is shown in figure 3.8. Each point represents the point estimate of *EPSPS* copy number of an individual plant. There is not a simple relationship between copy number of *EPSPS* for an individual and resistance category. However, the highest copy number individuals are members of populations that were documented to be highly resistant to both control and death. Among plants that are from populations documented as highly resistant based on mortality there are no individuals with more than 100 copies of *EPSPS* in the locations that are susceptible to injury, only one individual with more than 100 copies of *EPSPS* in the populations that have low resistance to injury, and three individuals with more than 100 copies of *EPSPS* in the populations that have high resistance to injury.

Relative Copy Number of ALS and EPSPS Genes in Other Amaranth Species

Most of the accessions investigated were single copy for both *ALS* and *EPSPS* (Figure 3.9). The only accession with evidence for increased copy number was *A. tuberculatus* (tall waterhemp) PI553086, this accession shows evidence of increased *ALS* copy number but not increased *EPSPS* copy number. Four studied accessions were removed from this figure: *A. tamaulipensis* PI642738, *A. blitum* (purple amaranth) PI652433, and *Celosia trigyna* (wool flower) PI482244 and PI649298. These four accessions show poor amplification efficiency of the A36 locus and thus report spuriously high copy numbers of *EPSPS* and *ALS*. Additionally the two *Celosia* accessions also have poor amplification efficiency of the *ALS* locus. This makes the relative quantification of *EPSPS* copy number impossible. *A. tamaulipensis* and *A. blitum* both had acceptable *ALS* and *EPSPS* amplification efficiencies, so *EPSPS* copy number was estimated relative to *ALS* and was seen to be normal (Table 3.9).

Once problematic amplification efficiencies were taken into account, there was very little variation in estimated *EPSPS* copy number (Table 3.10). For the accessions sampled the minimum, first quartile, median, third quartile, and maximum estimated *EPSPS* copy numbers were 0.575, 0.724, 0.981, 1.159, and 1.659 respectively. While about half of the point estimates were below 1, all point estimates below 1 also contained 1 within their 95% confidence interval and were considered to be single copy.

Tolerance of other amaranth to glyphosate

To investigate the possibility of a relationship between glyphosate resistance and *EPSPS* copy number similar to that seen in the NC populations the tolerance of individuals to glyphosate was assayed. Only three accessions showed any tolerance to full dose glyphosate, *A. tricolor* (PI477918 and Ames15326) and *Celosia trigyna* (PI482244) (Table 3.11). There was tolerance to glyphosate in the accession used as the single copy control, though results above (Figure 3.2) show the individual used was normal copy number. There was considerable variation in the level of tolerance to 0.5X glyphosate, with only 4 accessions fully susceptible. Two accessions from weedy species, *A. tuberculatus* and *A. palmeri*, were included in those fully susceptible accessions.

Haplotype Diversity of NC A. palmeri

Phase analysis identified the most likely genotypes of the 248 individuals phased. In several cases the best pair estimate had a low posterior probability, this was determined to be the result of unique mutations that phase could not place with the rest of the estimated haplotypes (i.e. was the new mutation a mutation of the A allele or the B allele). This was a common occurrence, observed in at least one locus of four in 100 of the individuals. The high level of genetic diversity also showed in the number of unique haplotypes estimated by phase—which will be a conservative estimate given how phase estimates haplotypes; table 3.12 shows that for each loci there were in excess of 100 unique haplotypes observed in 248 individuals.

Population Structure of A. palmeri in North Carolina

After an initial analysis, individuals from location 20 were determined to not be *A. palmeri*, but rather a potential hybrid with close genetic relationship to *A. caudatus* (Appendix B figure B.7). These individuals were excluded from subsequent analyses.

Based on the ln”(K), the second derivative of the function of average likelihood of K=x and K (Evanno *et al*. 2005), the best fit number of populations found in our NC *A. palmeri* data set was 5. We refer to these five populations using the colors assigned by structure (see figure 3.10): red, blue, green, yellow, and purple. When we examined other potential K values, we found the next best fit was K=4, at K=4 the purple and yellow clusters merge into a single population but other assignments are unchanged. At K=3 the purple, yellow, and blue clusters merge into a single population. Green is the most diverged population, and remains distinct at K=2.

To better understand the potential biological relevance of the five populations identified by Structure, population membership was statistically fit to other traits: collection location, field measured glyphosate control level, and field measured glyphosate mortality level. When population membership was compared to field tested glyphosate resistance the red cluster was associated with susceptibility, the green cluster with high control and mortality resistance, the blue cluster with low mortality resistance, the yellow cluster high mortality resistance (but not high control resistance), and the purple cluster with low control and mortality resistance (Table 3.12).

Analysis of means was used to compare the expected distribution of population admix to the observed distribution to look for collection location that had a significant association with a specific population, these results (not shown) suggested possible centers of origin for each of the populations other than red. A representation of the relative admixture contributions of each population in each collection location was represented as a pie chart centered over the approximate GPS coordinates of the collection site (Figure 3.11). Members of the red cluster were found evenly distributed across the state. The yellow and purple clusters did not seem to have a strong pattern of a center of origin with spread; individuals with yellow or purple ancestry were distributed across the state. Members of the blue cluster are found following the coast line of NC about 50-75mi inland, with the greatest prevalence suggesting an origin near the Virginia state line in Gates County and then spread south. The green cluster had the greatest prevalence in Anson County near the South Carolina state line and extends east, decreasing in prevalence as distance to Anson county increases. A band of highly admixed locations was observed almost following the I-95 highway corridor, with the most admixed locations—60 and 62—fifty miles south of Raleigh at the I-95 / I-40 highway junction. This suggested that some of the genetic diversity observed in NC *A. palmeri* could be the result of new genotypes being introduced by vehicle movement.

Discussion

Copy Number in EPSPS and ALS genes in NC A. palmeri

There is a wide range of *EPSPS* gene copy number amongst individuals: many individuals had normal copy number (one) and one had an estimated 180 copies. When individuals were grouped by collection location considerable diversity of copy number in many of these locations was seen (Figure 3.4). Most locations that were described as having any level of glyphosate resistance showed a large amount of variation in estimated *EPSPS* gene copy number, while the locations described as susceptible were comprised predominantly of single copy individuals. The fact that susceptible locations did contain high-copy number individuals at low frequencies could be a sign that if this study were repeated at the same locations in 2013 it would be found that many of the formerly susceptible locations are now resistant to glyphosate. The variation in copy number in resistant populations was also interesting, one possibility, described in Gaines *et al.* 2011 (Gaines *et al*. 2011), is that copy number was not stably inherited and that gain and loss of copies through the proliferation mechanism is ongoing. Another, not mutually exclusive, explanation is that high copy number plants were breeding with low copy number plants from locations with less glyphosate selection pressure (such as road-side ditches) and this gene flow contributed to the observed variation.

The Southern blot showed clear differences between high and normal copy number individuals and corroborated the findings of the quantitative PCR. However, the blot did not show precise copy number; this was likely due to band intensity in a Southern blot

being only semi-quantitative. Other work in highly amplified genomic elements have shown that Southern blots can not accurately assess the precise copy number when it is more than about 25 copies (Va *et al*. 1998). It was unexpected that the wild-type band (5kb) was not present in any of the high copy number individuals, but there was a high intensity major band (approx. 3kb) running below the wild-type band. Since something had to be responsible for the copy number proliferation, it is possible that this genetic element interferes with the wild-type cut site. Recent work by Gaines *et al.* (Gaines *et al*. 2013) showed that the amplified fragment in high *EPSPS* copy number *A. palmeri* from Georgia and Mississippi may be as large as 30kb and includes MITE-like sequences. It is possible that the major band in the high copy number individuals was the result of a MITE introducing an EcoRI site downstream of the nearest one in the wild-type. If this is the case, however, it leaves the question of where the other sized hybridizing fragments came from. Previous work by our lab and the Burgos lab showed that *EPSPS* may exists in two copies in wild type susceptible individuals (Burgos *et al*. 2008). The fact that this is not reflected in figure 3.3 is consistent with Gaines *et al*. who showed that two different versions as two bands can be seen on the Southern, but not when cut with EcoRI and probed for exon 1: the two versions co-migrate under these conditions. While Gaines *et al.* only had evidence that one of these *EPSPS* copies was participating in the copy number proliferation it is possible that the second *EPSPS* was also participating at a lower level with a smaller fragment that does not include a second EcoRI cut site, and thus insertions with different genomic contexts migrated differently on the gel because a cut sites was not part of the insertion cassette. How this reconciles with quantitative PCR

showing *EPSPS* in single copy is interesting and potentially problematic. One explanation is that *A. palmeri* are tetraploids and have two copies of all three genes, but previous research has shown *Amaranthus* to be diploid (Costea *et al*. 2004). Based on the fact that we found high copy number individuals in susceptible populations, the most likely explanation is that Burgos *et al.* studied a few susceptible individuals with extra copies, but did not know it because they did not assay for copy number. Based on finding susceptible individuals with more than two *EPSPS* alleles they concluded that the other susceptible individuals were similarly high copy number but less heterozygous.

While almost all of the plants assayed for copy number had normal ALS copy number, three plants did not, and all three were ALS resistant. This was a surprising and concerning results because other studies have been done that use qPCR referenced to *ALS* to determine *EPSPS* copy number (Gaines *et al*. 2010; Ribeiro *et al*. 2013; Salas *et al.* 2012) and any samples with elevated copy number of the reference gene would show an incorrectly low *EPSPS* copy number. However, *ALS* copy number was not the point of this study. These results should be considered during future work in ALS inhibitor resistance and quantitative PCR assessment of genomic copy number.

Copy Number and Resistance in North Carolina A. palmeri

Statistically significant associations between the estimated copy number for an individual and the resistance category it belonged to based on the resistance assessment of its collection location were found. Resistance category assignment was made based on how

multiple independent samples of each location responded in aggregate to different doses of herbicide. This method is not ideal for measuring the variation of resistance in a population because scores are given to the collection location as a whole instead of to individual plants. So the fact that these statistical associations hold up was impressive. It is also why there was so much variance.

What these findings suggest is that plants that are not killed by glyphosate application have elevated copy number and that plants that are not injured by glyphosate have even higher copy number. Location samples where >50% of individuals survived the low (280g) dose of glyphosate had on average seven copies but location samples that were not injured by the same dose had an average of 35. This is compared to highly resistant location samples where >50% of individuals withstood the 560g and 840g dose of glyphosate; locations where at least half survived these doses had on average 42 copies, and locations where at least half avoided injury had on average 37 copies. These data suggest that these values could represent threshold copy numbers for specific resistance phenotypes in *A. palmeri*. These results are also interesting in light of recent work by Zulet *et al.* (Zulet *et al*. 2013) regarding proteolysis as a metabolic response to amino acid synthesis disruption via glyphosate stress. Our results suggest that glyphosate induced injury is the result of proteolysis, which could still be induced by glyphosate in moderate copy number individuals, while the mortality comes from more complete blockage of the shikimate pathway. In a moderate copy number individual, glyphosate could still cause enough stress to trigger proteolysis and thus plant injury but not enough

stress to kill the plant. While a high copy number individual does not experience sufficient stress from glyphosate to induce proteolysis and thus injury is avoided. Future work to understand the specifics of the relationship between *EPSPS* gene copy number and glyphosate response would help explain the observation that the number of copies needed to avoid injury is very similar (35 and 37) for low and high dose, this may have to do with how proteolysis is induced through glyphosate stress.

Copy Number of ALS and EPSPS in Other Amaranthus Species

Quantitative PCR using the primers described is a robust way to estimate copy number in amaranth. Gene *A36* had the most variation in primer efficiency; this low efficiency in some species was unexpected because it is the only primer that was designed using multiple *Amaranthus* species as reference. Future work should involve careful selection of qPCR primers with consistent amplification efficiency across the range of species to be studied. This also highlights the importance of using multiple reference genes when working in species with few genomic resources. Without the second *ALS* reference gene, it would have been more difficult to diagnose the spurious copy number estimates in the four accessions with low *A36* gene amplification efficiency.

There is no evidence of copy number variation as standing neutral variation in the genus *Amaranthus*. However, due to sample size limitations we cannot completely eliminate the possibility.

Tolerance to Glyphosate in Other Amaranthus Species

Three species showed some level of tolerance to 1X glyphosate. Included in those species was *A. tricolor*, the accession that used as the control individual. This is concerning because if the control individual were to have elevated *EPSPS* copy number this would make the estimation of all other species' *EPSPS* copy number incorrectly low. However, absolute quantification of DNA concentration in our samples from each gene showed that all three genes estimated the same 0.8ng of starting DNA, which is very close to the target 1ng of DNA that was loaded. If *EPSPS* had been at a higher copy number then quantifying with the *EPSPS* locus and the *A36* standard curve would have given a fold over estimation of starting DNA quantity related to the fold copy number difference between *A36* and *EPSPS*. It is possible that individuals PI477918 A and B were single copy, but elevated *EPSPS* copy numbers do segregate in *A. tricolor*. The other possibility is that *A. tricolor* takes advantage of a different mechanism to resist glyphosate toxicity.

It is also interesting that low-levels of glyphosate tolerance segregate in two of the three *A. palmeri* accessions tested. None the accession show any tolerance to full dose glyphosate. At the same time none of the individuals assayed for *EPSPS* copy number showed any copy number elevation. Similar to observations in *A. tricolor*, possible explanations include these accessions using a different resistance mechanism and sample bias. However, since PI604557 only had 11% mortality the odds of choosing two normal copy-number plants if copy number variation causes the observed glyphosate resistance were low.

A. tricolor is relatively distantly related to the weedy amaranth species (see chapter I), so *A. tricolor* may have a private allele that confers resistance that evolved without the presence of selection pressure by glyphosate. As for the other plants, there seems to be ancestral tolerance as most accessions investigated have some level of tolerance to the half-dose. We propose three potential hypotheses regarding evidence of ancestral variation in tolerance but not in *EPSPS* copy number while *EPSPS* copy number is strongly associated with glyphosate resistance in North Carolina, Mississippi, and Georgia *A. palmeri*. The first is that there are two mechanisms functioning in most resistant *A. palmeri* and the presence of the second mechanism is not being detected due to the strong effect of copy number proliferation. Additionally the mechanism at work could be one that cannot confer sufficient tolerance to make a plant glyphosate resistant in the agricultural setting such as metabolism. To date evidence has been found that glyphosate can be metabolized by some glyphosate resistant weeds but the mechanism has never been found to be solely responsible for observed resistance levels (Carvalho *et al*. 2011; de Carvalho *et al*. 2011; González-Torralva *et al*. 2012). Thirdly, it is possible that the founder population of *A. palmeri* that now infests the Southeast contained no tolerant individuals and thus adaptation could not rely on ancestral variants that had been removed during the population bottleneck.

Analysis of Population Structure in North Carolina A. palmeri

Population structure analysis of the NC *A. palmeri* suggested that the individuals investigated cluster into 5 populations, with the red cluster being the largest of the five (Figure 3.10). All the locations with individuals described as susceptible to glyphosate were estimated to have almost 100% red ancestry. This suggested that the red cluster may represent the original susceptible population of North Carolina before resistance first arose. Ancestry in the blue cluster was statistically associated with low levels of mortality resistance and seemed to be centered on the eastern edge of the Virginia-North Carolina boarder and then spreading south. This pattern could have been caused by introgression of low glyphosate resistant genotypes from the north that then spread down, or by a *de novo* adaptive event in the northeastern part of North Carolina that then spread. Ancestry in the green cluster was associated with a high level of glyphosate resistance and with Anson country NC. The green cluster seemed to be almost completely confined to Anson County (location 55), with a small number of individuals to the northeast showing green ancestry. One explanation is a very recent introgression of the green cluster into Anson County that had not yet mixed with the other *A. palmeri*. Another possible explanation is that an individual within NC with high resistance founded the Anson County population which structure now identifies as separate from the original source population. This was corroborated by an approximately 20% lower heterozygosity estimated by structure for the green population (0.7884 vs. 0.9797, 0.9475, 0.9441, and 0.9247), potentially suggesting a recent population bottleneck. With respect to the origin of the yellow and purple population clusters, both of these populations were only seen in the highly

admixed locations (47, 48, 58, 60, and 62) all of which were near interstate highways, but were not all close to each other or on the same interstate highway. Locations 47 and 48 were near Hwy 74, which connects I-26 to I-85; the other three locations were on I-95. And while the locations were highly admixed, several individuals have predominantly yellow or purple cluster ancestry. This may suggest that these populations existed outside of NC and were very recent immigrants brought in repeatedly though vehicle movement. The recent immigration would not have given the individuals a chance to mix with the pre-existing genotype but still share a growing location. A final important observation regarding the red population cluster: while all the susceptible locations had nearly 100% red ancestry, not all locations with predominantly red ancestry were susceptible. This suggests that our estimate of 5 clusters and 4 adaptive events may be overly conservative with deeper levels of structure or older adaptive events not detected due to the relatively low number of loci used. The final conclusion to draw from this is that there is evidence to suggest that multiple adaptive events were responsible for the pattern of glyphosate resistance observed in 2010 among the *A. palmeri* growing in North Carolina. The current data did not allow us to determine the origin of these events—*de novo* in NC or originating elsewhere and introgression into NC. Future work with the focus of making this distinction will give us a clearer understanding of just how easily *A. palmeri* gain this elevated *EPSPS* gene copy number in response to glyphosate stress (i.e. were there four events in total or in North Carolina alone). To understand that, future work will require comparing the genotypes found in NC to those found in other states with glyphosate resistant *A. palmeri*. However, the large number of genetically distinct events considered

with the lack of genus-wide EPSPS copy number variation does suggest that multiple *de novo* events have occurred. This has many potentially important impacts for both weed management and our understanding of adaptation, not only has the same phenotype arisen multiple times, but it has arisen via the same mechanism and in an incredibly short period of time.

Considering the lack of evidence for ancestral variation of *EPSPS* gene copy number in the genus *Amaranthus* the rapid adaptation to glyphosate stress seems to contrast with existing hypotheses regarding adaptive evolution. When adaptive alleles are not already segregating in a population due to ancestral variation, adaptation is expected to be slow (Hermisson and Pennings. 2005); *Amaranthus* adapted to glyphosate stress very quickly. Also, there is evidence for multiple adaptive events that converge on the same genetic change without common ancestry—not just between *A. palmeri* populations, this mechanism has also been found in *Lolium rigdum* (Salas *et al*. 2012). Improved understanding of this dynamic—rapid adaptation without ancestral variation—could have far ranging impacts for scientific understanding of evolution, and important practical applications for pest, pathogen, and cancer management through a better ability to predict and prevent the adaptation of pesticide/drug resistance.

Table 3.1: Sampling of Palmer amaranth (*Amaranthus palmeri*). The 31 population included in this study with GPS coordinates and glyphosate and ALS inhibitor resistance levels documented by the Burton Lab at NCSU (HR – highly resistant, LR – low resistance, S – susceptible, ratings were made based on the methods described in Whitaker, 2009) Locations that share a map symbol also share a glyphosate resistance profile. Bolded rows indicate populations included in EPSPS copy number analysis.

Map symbol	Location code	Longitude	Latitude	Crop	Glyphosate resistance (Control)	Glyphosate resistance (mortality)	ALS inhibitor resistance (Control)	ALS inhibitor resistance (mortality)
Diamond	**3**	**36.415748**	**-78.17158**	**Soybean**	**HR**	**HR**	**LR**	**HR**
Star	9	-77.5519	36.349234	Cotton	S	S	LR	HR
Star	10	-77.5371	36.19869	Cotton	S	S	LR	LR
Square	12	-76.7709	36.50205	Cotton	S	LR	S	LR
Diamond	**13**	**36.347794**	**-76.88039**	**Soybean**	**HR**	**HR**	**S**	**LR**
Square	**14**	**36.186534**	**-77.08629**	**Soybean**	**S**	**LR**	**LR**	**HR**
Diamond	17	-76.6305	36.138638	Cotton	HR	HR	S	LR
Star	**19**	**36.02**	**-80.9725**	**Soybean**	**S**	**S**	**HR**	**HR**
Star	20	36.00167	-80.78833	Soybean	S	S	S	LR
Square	22	-80.7547	35.65283	Soybean	S	LR	HR	HR
Diamond	24	-79.6341	35.92914	Soybean	HR	HR	HR	HR
Triangle	25	-79.6461	35.8481	Soybean	LR	LR	HR	HR
Star	**33**	**35.829737**	**-77.92322**	**Cotton**	**S**	**S**	**HR**	**HR**
Triangle	**47**	**35.32166**	**-81.6719**	**Soybean**	**LR**	**HR**	**LR**	**HR**
House	**48**	**35.31199**	**-81.73936**	**Cotton**	**LR**	**LR**	**LR**	**HR**
Diamond	**55**	**35.144**	**-80.10255**	**Cotton**	**HR**	**HR**	**HR**	**HR**
Circle	**58**	**35.46398**	**-78.82922**	**Cotton**	**S**	**HR**	**HR**	**HR**
Circle	**60**	**35.34374**	**-78.5777**	**Soybean**	**S**	**HR**	**LR**	**HR**
Circle	**62**	**35.1422**	**-78.59029**	**Soybean**	**S**	**HR**	**LR**	**HR**
Square	**71**	**35.31988**	**-77.93745**	**Soybean**	**S**	**LR**	**HR**	**HR**
Square	75	-77.5328	35.48921	Soybean	S	LR	HR	HR
Square	81	-77.2402	35.1473	Soybean	S	LR	HR	HR
Triangle	**85**	**35.301155**	**-76.80207**	**Soybean**	**LR**	**HR**	**LR**	**HR**
Diamond	88	-80.3809	34.86468	Soybean	HR	HR	LR	HR
Triangle	**90**	**34.98358**	**-79.73276**	**Soybean**	**LR**	**HR**	**HR**	**HR**
Star	100	-79.0697	34.6268	Soybean	S	S	LR	HR
Square	101	-78.8976	34.87389	Soybean	S	LR	LR	HR
Star	**103**	**34.77485**	**78.84609**	**Soybean**	**S**	**S**	**HR**	**HR**
Star	109	-77.9647	34.98332	Soybean	S	S	HR	HR
Diamond	116	-79.0175	34.29125	Soybean	HR	HR	HR	HR
Diamond	122	-78.4076	34.4591	Cotton	HR	HR	HR	HR

Table 3.2: USDA germplasm accession numbers for species included in genus-wide *EPSPS* gene copy number assessment

Species	Accession Numbers			
A. caudatus	Ames3860	PI553073		
A. powellii	PI572260	PI577261	PI632241	
A. retroflexus	PI603845	PI607447		
A. palmeri	Ames5298	PI549158	PI604557	PI607454
A. spinosus	PI632248	PI642740		
A. tricolor	Ames15326	PI477918		
A. viridis	PI536439	PI652434		
A. blitum	PI632245	PI652433		
A. fimbriatus	PI605738	PI662285		
A. crassipes	PI642743	PI649302		
A. californicus	PI595319			
A. tamaulipensis	PI642738			
A. tuberculatus	PI553086	PI603881		
A. arenicola	PI599673	PI607459		
Celosia trigyna	PI482244	PI649298		

Table 3.3: Primer pairs used to PCR amplify and then Sanger sequence the four genes used for structure analysis (see chapter one for more details)

Gene targeted for amplification	5’ to 3’ Sequence (Forward/reverse)
A07 (Endosomal P24A protein precursor, putative)	GGAAGCTTGTTGTGGGTGAT / AATGGCTGAAACAGGTCCAC
A36 (DEAD box RNA helicase, putative)	TGGTTATCCGTGCCTTTCTC / CAGGACCTGGATTCTTTCCA
A37 (serine-type endopeptidase, putative)	CACTGAAGCCTACGGAGARG / GATTGGGCTGGTCACTSTGT
A40 (glutaredoxin, putative)	GGTGAGCTTATCGGTGGGTG / TCCGAAAGGGTTGATTTRAG

Table 3.4: Primer efficiency and 5' → 3' sequence of three primer pairs used for EPSPS copy number quantification. Reference genes used are ALS (acetolactate synthase gene) and A36 (putative DEAD box RNA helicase).

Locus	Primer efficiency	Forward Sequence	Reverse Sequence
A36	96.10%	TTGGAACTGTCAGAGCAACC	GAACCCACTTCACCAAAAC
ALS	97.50%	AGCTCTGGAACGTGAAGGC	TCAATTAAAACCGGTCCGGG
EPSPS	97.00%	ATGTTGGACGCTCTCAGAACTCTTGGT	TGAATTTCCTCCAGCAACGGCAA

Table 3.5: Copy number estimates by collection location. Point estimates were made for each individual assayed and then individuals were grouped by collection location. These values represent the minimum, median, maximum, and quartile values observed for the individuals within each collection location.

	ALS					EPSPS				
Location	**Min**	**Q25**	**Median**	**Q75**	**Max**	**Min**	**Q25**	**Median**	**Q75**	**Max**
3	1.0151	1.0626	1.0718	1.1615	1.4726	1.3503	50.7041	65.7237	75.9690	125.3658
13	0.8716	0.9703	1.0895	1.3535	1.8298	11.1967	15.5569	34.9395	46.0190	52.4062
14	0.9277	1.1096	1.1514	1.1886	1.3180	0.7873	0.9555	1.0487	8.1223	53.6312
19	0.9406	1.0169	1.0640	1.1566	1.3692	0.9604	1.0811	1.5591	2.2457	4.7185
33	0.9514	1.1117	1.2004	1.5133	4.2338	0.8782	1.2539	3.1201	35.2031	63.3913
47	1.1920	1.2494	1.3108	1.4340	1.4540	17.7942	40.1938	69.6653	87.4003	182.7001
48	0.8399	1.0183	1.3744	1.8772	2.3054	0.8536	11.0574	37.2203	52.7416	83.3822
55	0.8645	0.9950	1.2283	1.5732	2.4284	16.9514	23.2372	57.0175	65.1947	166.9571
58	0.8094	0.9466	1.0271	1.0619	1.1303	22.8111	39.0771	46.4009	51.6356	65.4203
60	0.7955	0.8729	0.9266	0.9483	1.0000	29.2426	34.8290	41.9326	46.5589	52.8318
62	0.8151	0.9659	1.0012	1.1555	1.8213	1.1264	42.9901	53.6155	62.1618	66.1040
71	0.9384	1.0632	1.1267	1.1709	1.2938	1.1728	1.3772	14.0600	23.6896	69.2304
85	0.9320	1.0467	1.1161	1.1959	1.3241	13.8646	28.2139	33.0372	45.6189	74.8882
90	0.8878	1.0847	1.1500	1.1729	1.2585	26.7846	28.8643	45.1939	51.8642	74.1993
130	0.6643	0.8684	0.9332	1.1000	1.1594	0.9287	0.9900	1.1920	4.7830	60.2685

Table 3.6: Individuals assayed for copy number via Southern Blot. The 8 plants used for Southern blot analysis confirmation of copy number elevation and their EPSPS genomic copy number as determined by real-time PCR analysis including point estimate and the 95% confidence interval (CI) as determined by ANOVA in SAS.

Plant	Copy Number	Upper 95% CI	Lower 95% CI
3A	55.72	81.49	38.49
3C	99.04	144.86	67.72
3E	1.35	1.98	0.92
3G	125.37	183.36	85.71
47B	67.46	124.68	36.54
47C	17.79	32.87	9.63
47E	84.79	152.60	47.11
48A	39.62	64.74	24.25

Table 3.7: Mean values for the ddC_t (not log transformed into relative copy number) of each injury resistance category and $d(ddC_t)$ for each comparison between resistance categories. *P*-values and confidence limits (CL) are a statistical comparison against the null hypothesis that $ddC_t = 0$ (individuals have, on average, a single *EPSPS* gene copy) or $d(ddC_t) = 0$ (there is statistically no difference in the average copy numbers between injury resistance categories)

Control/Injury Resistance	**Mean ddCt**	**Std Error**	***P*-value**	**Upper 95% CL**	**Lower 95% CL**
Susceptible	-3.06662	0.105415	<.0001	-3.27363	-2.85961
Low resistance	-5.1157	0.149458	<.0001	-5.40919	-4.8222
High resistance	-5.22503	0.172109	<.0001	-5.56301	-4.88705
	d(mean ddCt)	**Std Error**	***P*-value**	**Upper 95% CL**	**Lower 95% CL**
Low resistance v. Susceptible	-2.04908	0.182893	<.0001	-2.40824	-1.68992
High resistance v. Susceptible	-2.15841	0.201826	<.0001	-2.55475	-1.76207
High resistance v. Low res.	-0.10933	0.227945	0.6316	-0.55696	0.338296

Table 3.8: Mean values for the ddC_t (not log transformed into relative copy number) of each mortality resistance category and $d(ddC_t)$ for each comparison between resistance categories. *P*-values and confidence limits (CL) are a statistical comparison against the null hypothesis that $ddC_t = 0$ (individuals have, on average, a single *EPSPS* gene copy) or $d(ddC_t) = 0$ (there is statistically no difference in the average copy numbers between mortality resistance categories)

Mortality Resistance	Mean ddCt	Std Error	*P*-value	Upper 95% CL	Lower 95% CL
Susceptible	-1.32206	0.172109	<.0001	-1.66003	-0.98408
Low resistance	-2.78169	0.172109	<.0001	-3.11967	-2.44372
High resistance	-5.37328	0.099505	<.0001	-5.56869	-5.17788

	d(mean ddCt)	Std Error	*P*-value	Upper 95% CL	Lower 95% CL
Low resistance v. Susceptible	-1.45964	0.243398	<.0001	-1.93761	-0.98166
High resistance v. Susceptible	-4.05123	0.198803	<.0001	-4.44163	-3.66083
High resistance v. Low res.	-2.59159	0.198803	<.0001	-2.98199	-2.20119

Table 3.9: Copy number estimates by USDA accession. Estimates were made for each accession by pooling data for the two individuals representing the accession. These values represent the point estimate and 95% confidence limits of *ALS* or *EPSPS* gene copy number relative to A36 copy number, and then the *P*-value for the statistical test with the null hypothesis being the estimate value = 1.

	ALS				**EPSPS**			
Species - Accession	**Estimate**	***P*-value**	**95% Confidence limits**		**Estimate**	***P*-value**	**95% Confidence limits**	
***A. arenicola*-PI607459**	0.805452	0.4709	1.45135	0.446999	0.807525	0.4758	1.454457	0.448343
***A. californicus*-PI595319**	0.876913	0.6697	1.605012	0.479109	0.691911	0.2316	1.265883	0.378187
***A. caudatus*-PI553073**	0.839656	0.7221	2.202772	0.320061	0.732844	0.5269	1.922053	0.27942
***A. crassipes*-PI649302**	0.776219	0.2674	1.215147	0.495838	0.677217	0.0878	1.059559	0.432842
***A. fimbriatus*-PI605732**	1.031352	0.918	1.858401	0.572366	0.958098	0.8864	1.725659	0.531942
***A. palmeri*-Ames15298**	1.262916	0.4366	2.275659	0.700877	1.053296	0.8625	1.897124	0.584797
***A. palmeri*-PI549158**	0.7962	0.6428	2.088769	0.303497	0.574978	0.2601	1.508012	0.219229
***A. palmeri*-PI604557**	1.293431	0.3912	2.330644	0.717812	1.239154	0.4745	2.231878	0.687987
***A. powellii*-PI572260**	0.880722	0.578	1.378742	0.562593	0.827036	0.4051	1.293964	0.5286
***A. powellii*-PI572261**	1.269743	0.4261	2.287961	0.704666	0.962535	0.8986	1.733652	0.534406
***A. powellii*-PI632241**	0.470969	0.0049	0.794772	0.279088	1.043607	0.8727	1.760258	0.618725
***A. retroflexus*-PI603845**	1.173813	0.654	2.368175	0.581814	1.178195	0.6464	2.376154	0.584197
***A. retroflexus*-PI607447**	1.240595	0.5636	2.581521	0.59619	1.152479	0.7037	2.397354	0.554031
***A. spinosus*-PI632248**	1.681667	0.0362	2.734895	1.034045	1.199362	0.4627	1.949508	0.737862
***A. spinosus*-PI642740**	0.873292	0.7048	1.761872	0.432857	0.69572	0.3102	1.40311	0.344966
***A. tricolor*-Ames15326**	1.184712	0.5721	2.134743	0.657476	1.287803	0.3989	2.3195	0.714997
***A. tuberculatus*-PI553086**	2.504524	0.0023	4.512923	1.389928	1.659176	0.0916	2.988392	0.921186
***A. tuberculatus*-PI603881**	1.113921	0.7191	2.007185	0.61819	1.084578	0.7865	1.953466	0.602165
***A. viridis*-PI536439**	0.702808	0.3241	1.41792	0.348355	0.64838	0.2256	1.307637	0.321493
***A. arenicola*-PI607459**	0.805452	0.4709	1.45135	0.446999	0.807525	0.4758	1.454457	0.448343

Table 3.10: Gene copy number estimates by USDA accession. Estimates were made for each accession by pooling data for the two individuals representing the accession. These values represent the point estimate and 95% confidence limits of *EPSPS* gene copy number relative to *ALS* gene copy number for two accessions with poor amplification efficiency of the *A36* gene, and then the *P*-value for the statistical test with the null hypothesis being the estimate value = 1.

	EPSPS* vs. *ALS			
Species-Accession	**Estimate**	***P*-value**	**95% Confidence Limits**	
***A. tamaulipensis*-PI642738**	0.76	0.4487	1.806168	0.263246
***A. blitum*-PI652433**	1.8	0.0716	1.048605	0.324597

Table 3.11: Glyphosate tolerance across *Amaranthus*. Results of the Burgos lab analysis of glyphosate tolerance among USDA accessions of amaranth. The percentages of plants injured or killed at two doses (0.5X = 420g/ha and 1X = 840g/ha) of glyphosate are shown here. Due to poor germination *A. fimbriatus* was only tested at the higher dose. Accessions *A. tamaulipensis* (PI642738), *Celosia trigyna* (PI649298), and *A. californicus* (PI595319) were not included due to no germination.

			0.5X		**1X**	
Species	Accession	No. plants/ treatment	Injury %	Mortality %	Injury %	Mortality %
A. tuberculatus	PI553086	2	60	0	100	100
A. retroflexus	PI607447	7	100	100	100	100
A. spinosus	PI642740	14	40	0	100	100
A. tuberculatus	PI603881	5	100	100	100	100
A. tricolor	PI477918	20	35	0	60	25
Celosia trigyna	PI482244	16	40	0	90	50
A. palmeri	Ames15298	4	100	100	100	100
A. fimbriatus	PI605738	3*			100	100
A. tricolor	Ames15326	12	75	0	95	0
A. arenicola	PI607459	7	50	0	100	100
A. crassipes	PI649304	7	80	0	100	100
A. viridis	PI536439	15	90	40	100	100
A. powellii	PI572260	12	90	75	100	100
A. blitum	PI652433	4	100	100	100	100
A. powellii	PI632241	13	20	0	100	100
A. powellii	PI572261	9	100	100	100	100
A. spinosus	PI632248	19	40	0	100	100
A. caudatus	PI553073	20	20	20	100	100
A. palmeri	PI604557	9	40	11	100	100
A. retroflexus	PI603845	12	60	0	100	100
A. palmeri	PI549158	6	80	0	100	100

Table 3.12: Statistical association between STRUCTURE cluster assignment and documented glyphosate resistance. JMP was used to fit structure cluster assignment to documented location resistance level for both injury and mortality. Student's T-test was then used ($\alpha = 0.05$) to determine the statistical significance of associations. Results are represented by a connecting letters report; values within a group with shared letters are not statistically significantly different at a *P*-value of 0.05.

Injury Resistance level	mean cluster assignment	conecting letters report	Mortality Resistance level	mean cluster assignment	conecting letters report
Red Cluster			**Red Cluster**		
High	0.72	A	High	0.58	B
Low	0.64	A	Low	0.39	C
Susceptible	0.56	A	Susceptible	0.98	A
Green Cluster			**Green Cluster**		
High	0.2	A	High	0.15	A
Low	0.04	B	Low	0.02	B
Susceptible	0.04	B	Susceptible	0.02	B
Blue Cluster			**Blue Cluster**		
High	0.004	B	High	0.04	B
Low	0.13	A	Low	0.38	A
Susceptible	0.16	A	Susceptible	0.007	B
Purple Cluster			**Purple Cluster**		
High	0.04	B	High	0.09	A
Low	0.16	A	Low	0.14	A
Susceptible	0.07	B	Susceptible	0.004	B
Yellow Cluster			**Yellow Cluster**		
High	0.025	A	High	0.13	A
Low	0.1	A	Low	0.06	B
Susceptible	0.09	A	Susceptible	0.004	B

Table 3.13: The number of unique haplotypes as estimated by PHASE. For each of the four loci 248 diploid individuals were sequenced.

Locus	Number of estimated haplotypes
A07	146
A36	123
A37	282
A40	257

Figure 3.1: Sampling of Palmer amaranth (*Amaranthus palmeri*). The 31 population included in this study are shown on a map of North Carolina. Locations that share a map symbol also share a glyphosate resistance profile; see table 3.1 for details regarding resistances, exact GPS coordinates, and symbol definitions. Blackened points indicate populations included in *EPSPS* gene copy number analysis.

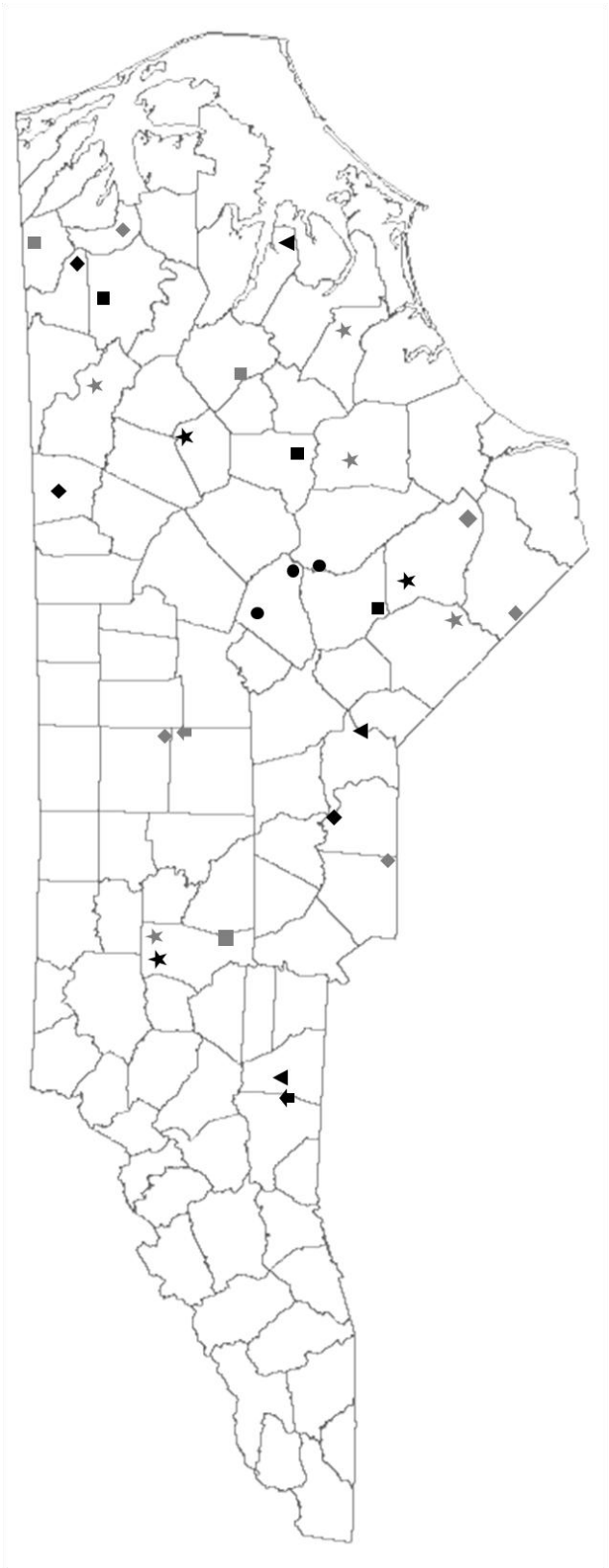


Figure 3.2: Absolute quantification of genomic DNA from *A. tricolor* PI477918. Quantity is estimated at each of the three loci (EPSPS and the two reference loci A36 and ALS). Statistically indistinguishable starting quantities indicate that this individual has the same number of genomic copies of each locus and is an acceptable reference individual.

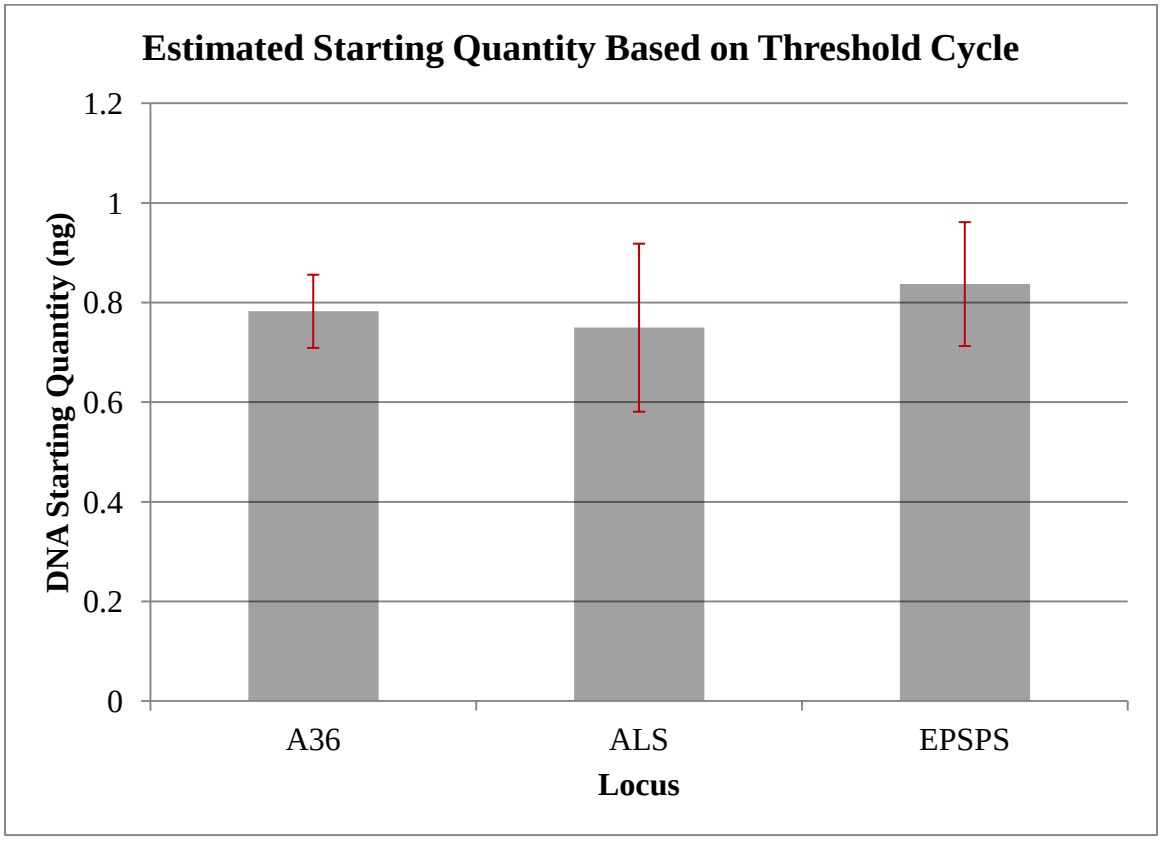


Figure 3.3: *ALS* gene copy number as estimated by qPCR averaged among individuals sharing a collection location (Mean of ANCOVA-based point estimates). Upper and lower stems of the box plots represent the maximum and minimum values observed at that collection location, upper and lower bounds of the box represent the first and third quartile, and the dash in the center represents the median value.

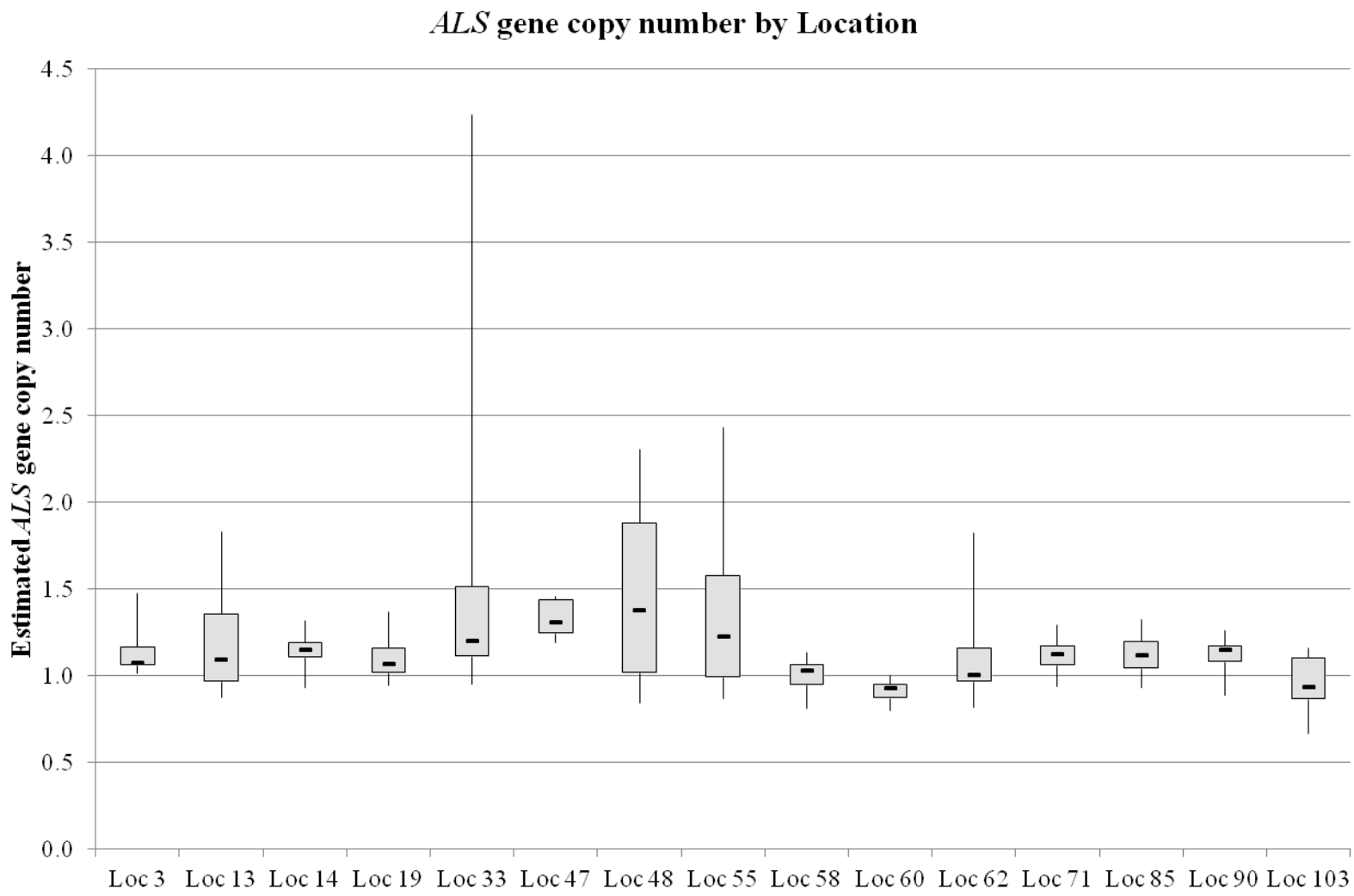

ALS gene copy number by Location
Estimated *ALS* gene copy number
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Loc 3
Loc 13
Loc 14
Loc 19
Loc 33
Loc 47
Loc 48
Loc 55
Loc 58
Loc 60
Loc 62
Loc 71
Loc 85
Loc 90
Loc 103

Figure 3.4: *EPSPS* gene copy number as estimated by qPCR averaged among individuals sharing a collection location. Upper and lower stems of the box plots represent the maximum and minimum values observed at that collection location, upper and lower bounds of the box represent the first and third quartile, and the dash in the center represents the median value. (Note difference in vertical scale compared to figure 3.3)

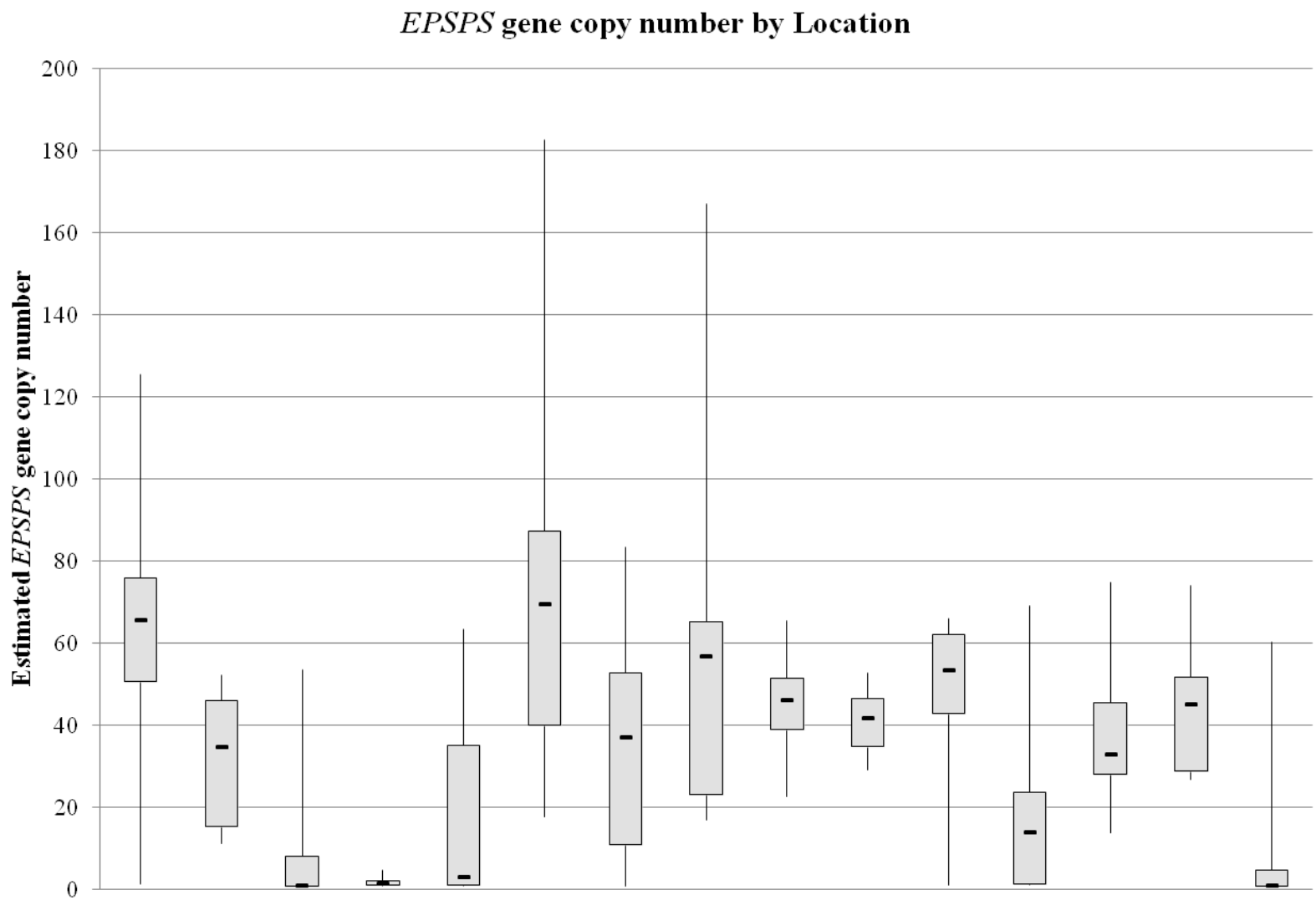

EPSPS gene copy number by Location
Estimated *EPSPS* gene copy number
200
180
160
140
120
100
80
60
40
20
0

Figure 3.5: Southern blot of EcoRI digested genomic DNA probed for *EPSPS* exon 1. Lanes from left to right: marker (visible bands represent 20kb and 5kb), 3A, 3C, 3E, 3G, blank, 47B, 47C, 47E, 48A, *EPSPS* exon 1 PCR amplicon (positive control), and *EPSPS* exon 2 amplicon (negative control).

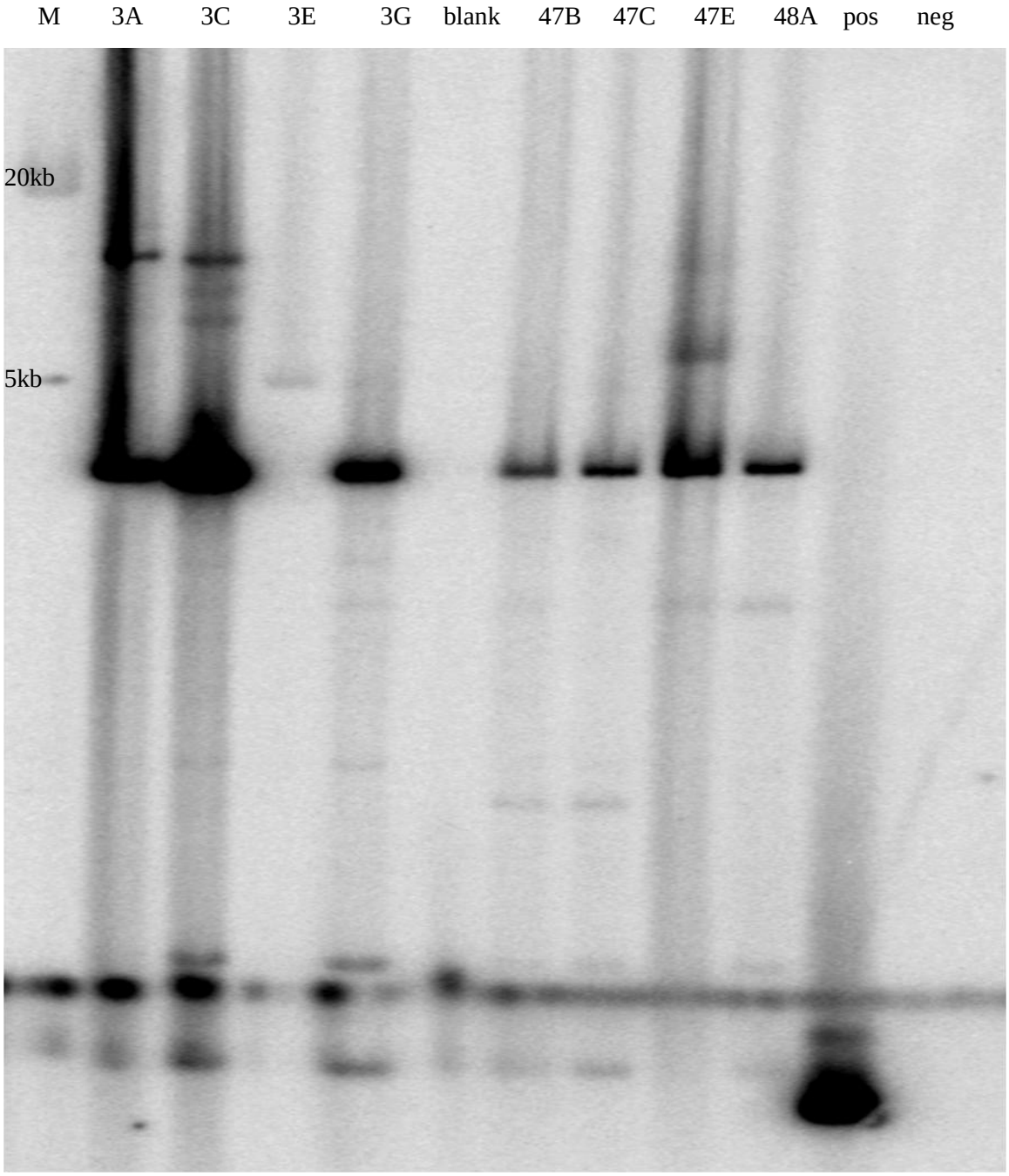


Figure 3.6: Mean values for the relative gene copy number (log transformed from ddC_t values shown in table 3.7) of each injury resistance category. Error bars represent the 95% confidence intervals of each estimate; non-overlapping error bars (i.e. confidence intervals) represent statistically significantly different mean values.

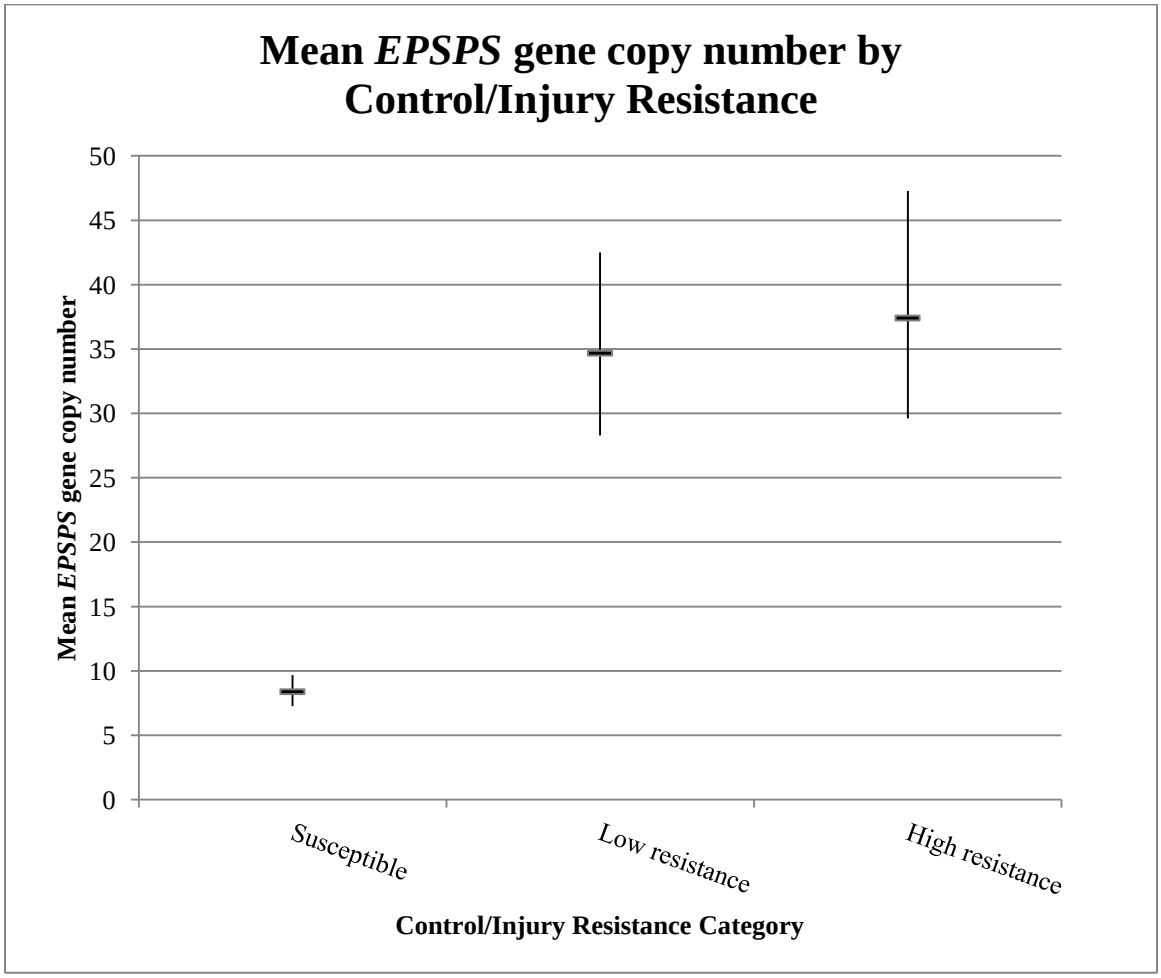


Figure 3.7: Mean values for the relative gene copy number (log transformed from ddC_t values shown in table 3.8) of each mortality resistance category. Error bars represent the 95% confidence intervals of each estimate; non-overlapping error bars (i.e. confidence intervals) represent statistically significantly different mean values.

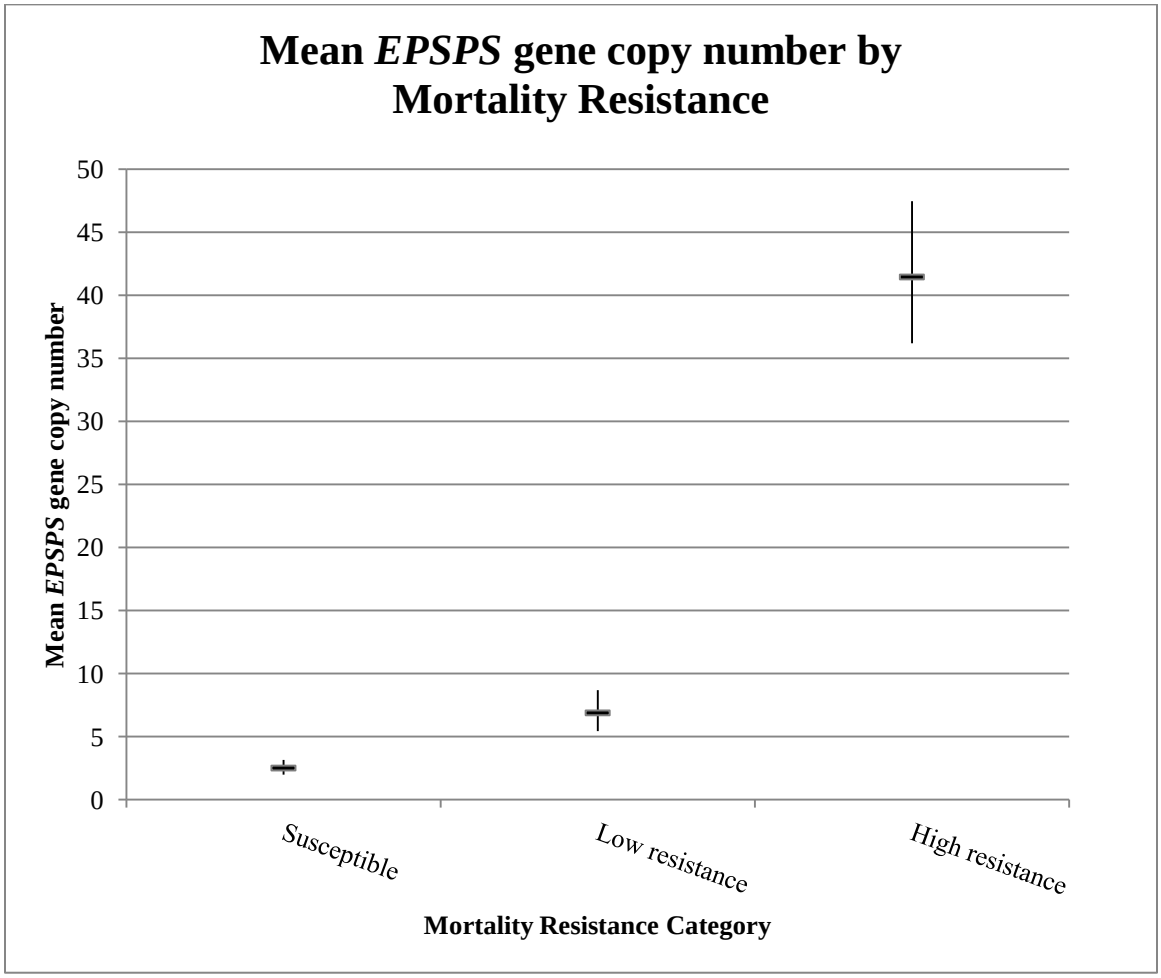


Figure 3.8: Point estimates for *EPSPS* gene copy number are shown grouped by resistance category. Each point represents the estimated *EPSPS* gene copy number for each individual (relative to the *A36* gene); resistance category is based on a collection-location scale assignment.

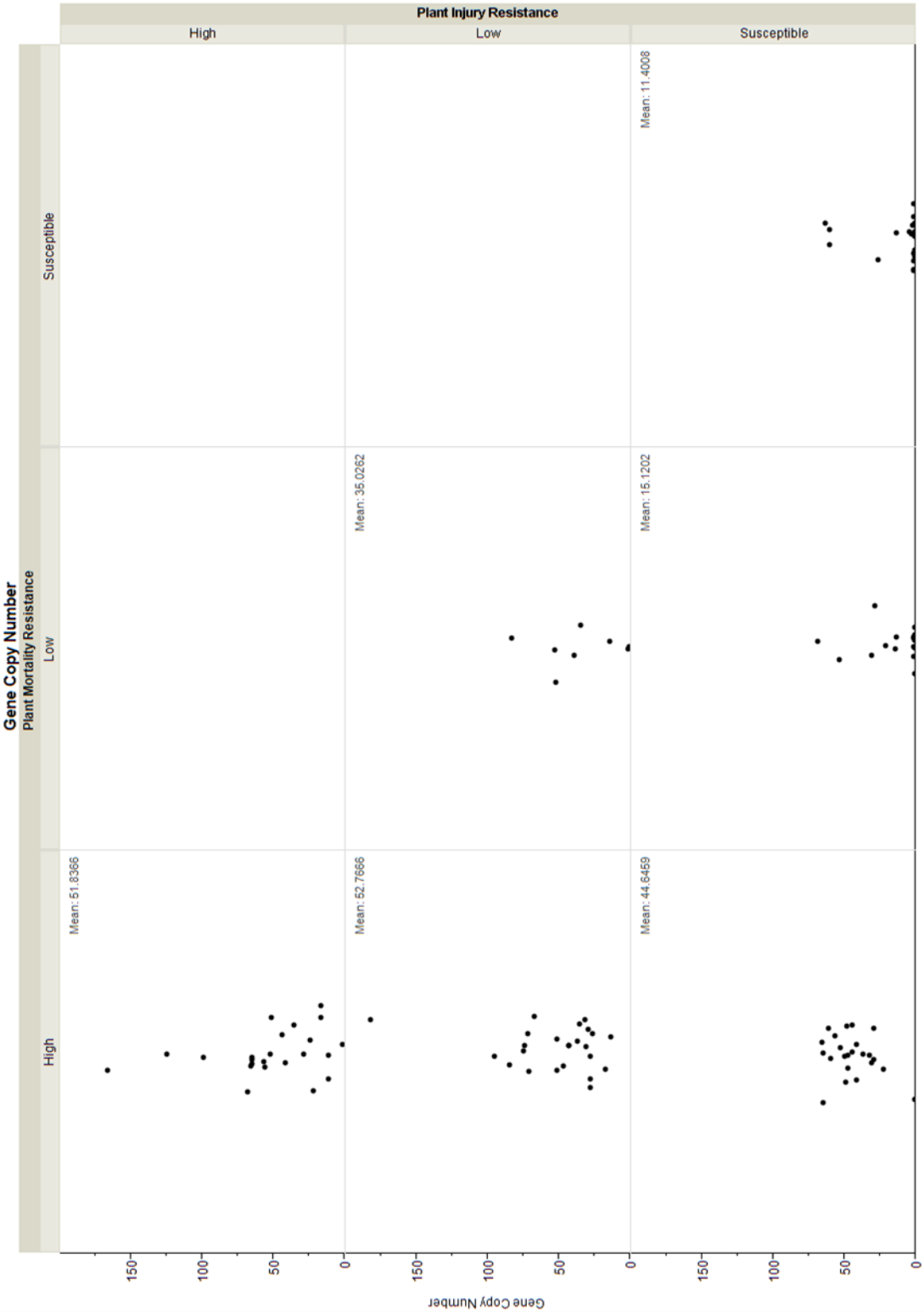

Plant Injury Resistance
High
Low
Susceptible
Gene Copy Number
Plant Mortality Resistance
Susceptible
Low
High
Mean: 11.4008
Mean: 35.0262
Mean: 15.1202
Mean: 51.8366
Mean: 52.7666
Mean: 44.6459
150
100
50
0
Gene Copy Number

Figure 3.9: *EPSPS* and *ALS* gene copy number estimated relative to the A36 gene for the USDA accessions included in this investigation. Error bars represent the 95% confidence interval on the estimates.

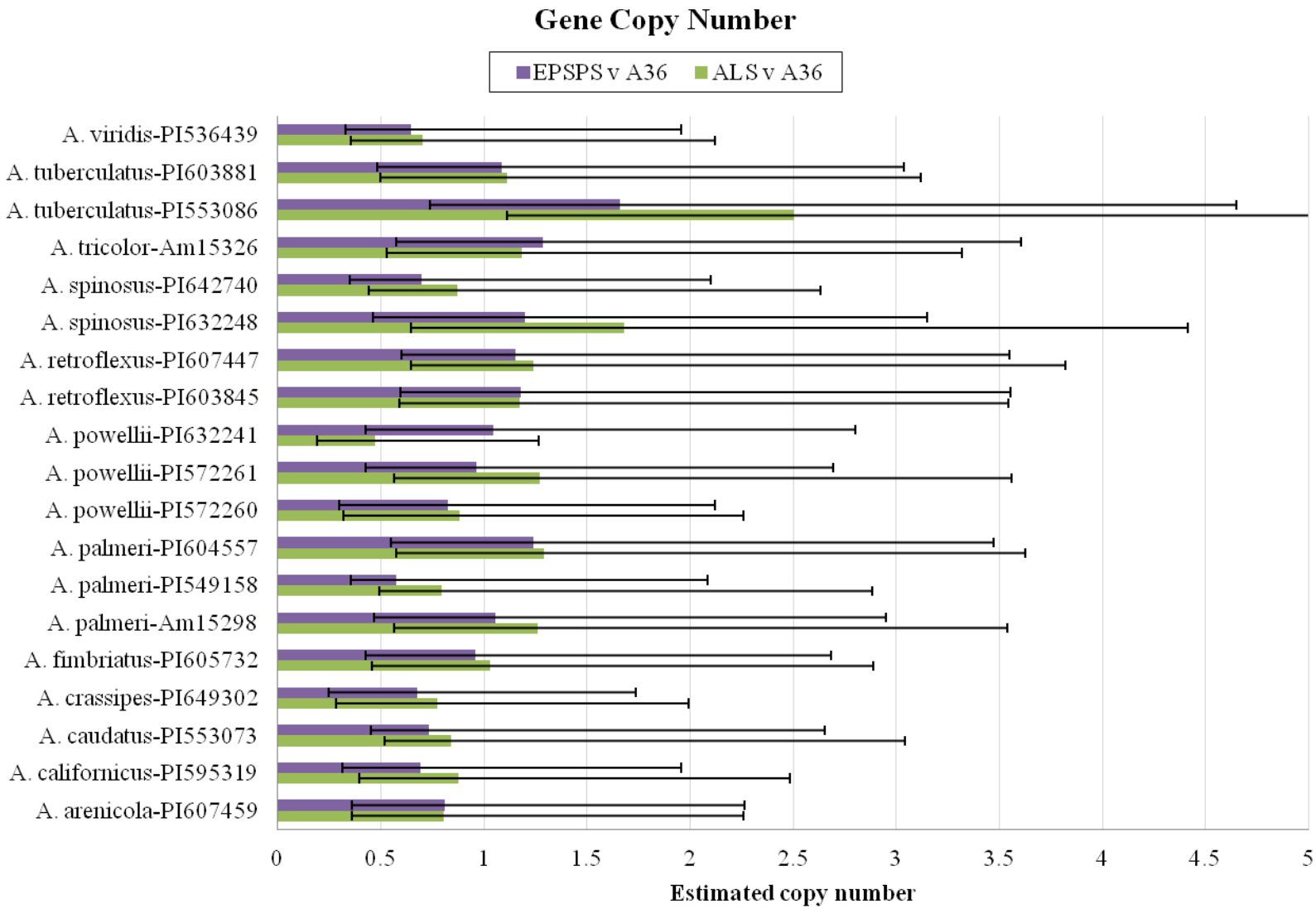

Gene Copy Number
EPSPS v A36
ALS v A36
A. viridis-PI536439
A. tuberculatus-PI603881
A. tuberculatus-PI553086
A. tricolor-Am15326
A. spinosus-PI642740
A. spinosus-PI632248
A. retroflexus-PI607447
A. retroflexus-PI603845
A. powellii-PI632241
A. powellii-PI572261
A. powellii-PI572260
A. palmeri-PI604557
A. palmeri-PI549158
A. palmeri-Am15298
A. fimbriatus-PI605732
A. crassipes-PI649302
A. caudatus-PI553073
A. californicus-PI595319
A. arenicola-PI607459
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Estimated copy number

Figure 3.10: Population structure as inferred from the best fit number of clusters in STRUCTURE. Best fit model has K=5 populations. Each color represents one of the five STRUCTURE estimated populations. Individuals are shown as each vertical line, multiple colors in an individual's assignment represent admixture. Boxes group individuals by collection location (labeled on the bottom).

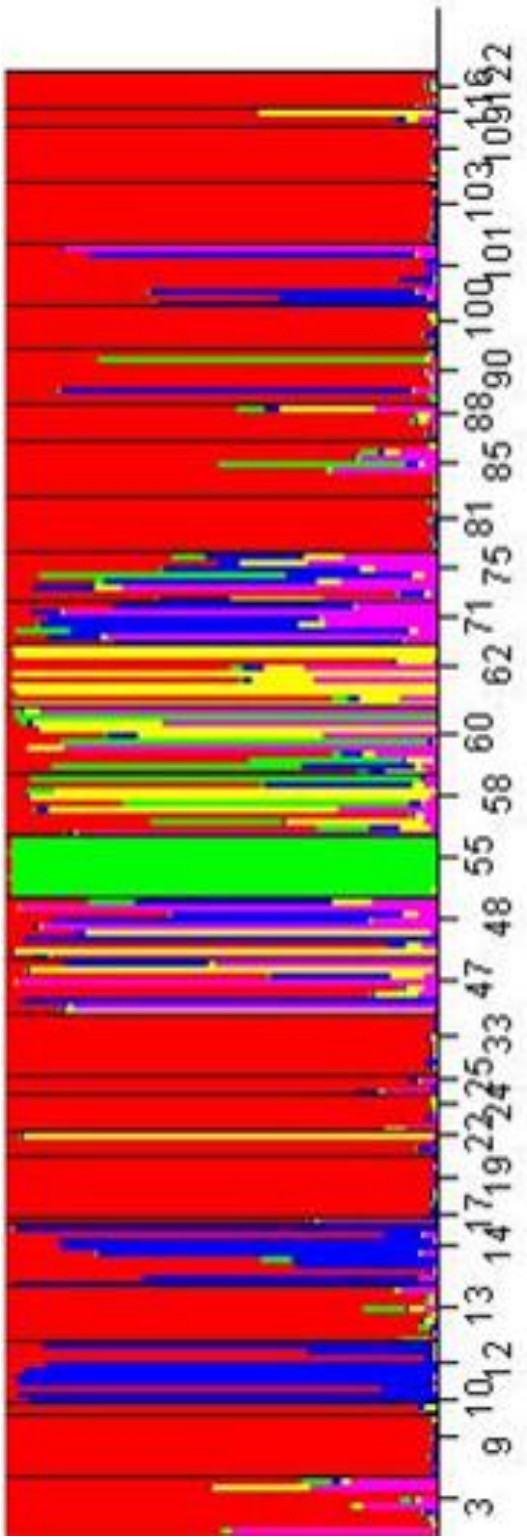
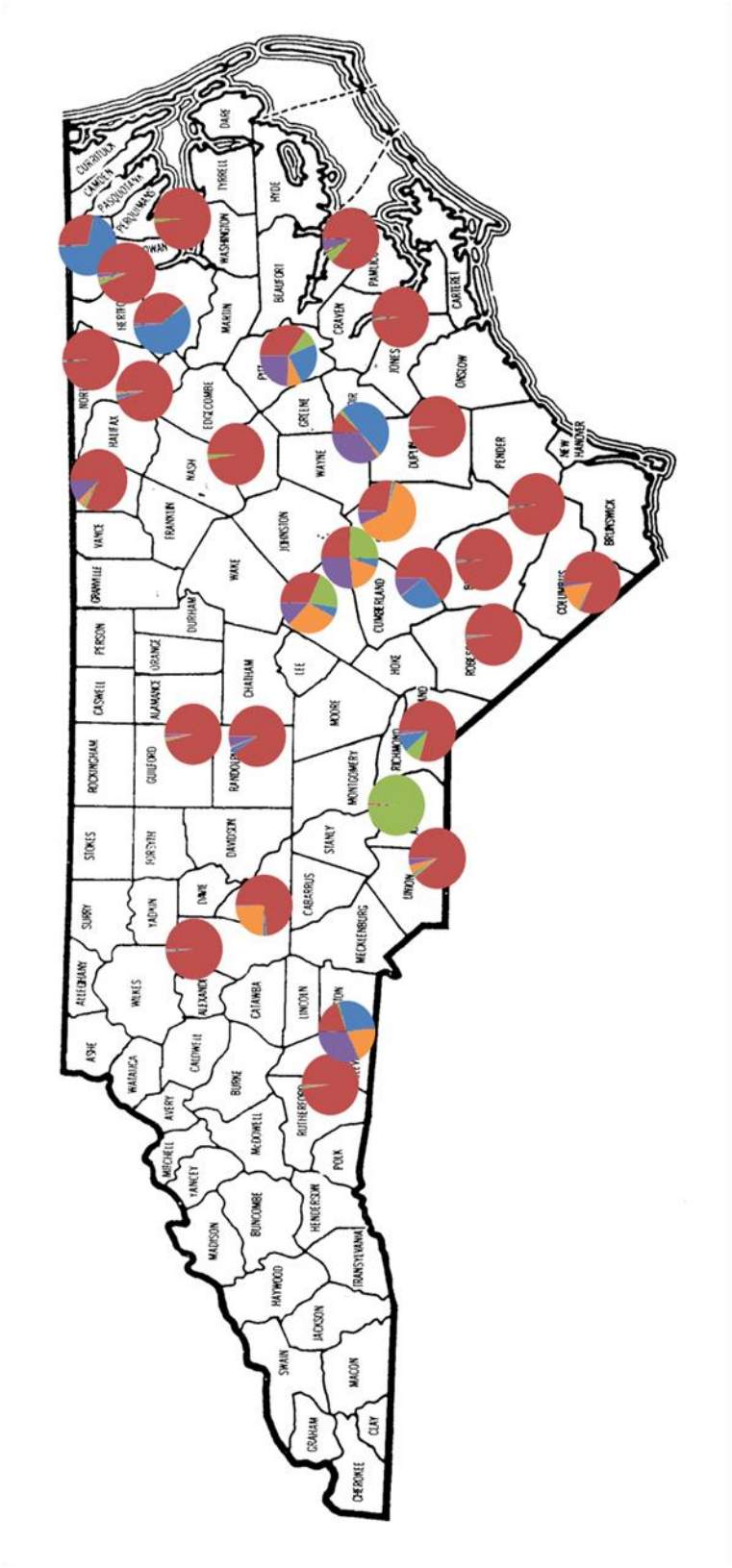

3
9
10
12
13
14
17
19
22
24
25
33
47
48
55
58
60
62
71
75
81
85
88
90
100
101
103
109
116
122

Figure 3.11: The admixture found in different collection locations as estimated by Structure 2.3.4. Each pie chart is placed over the approximate GPS coordinates of the collection location represented. Percentages of population (color) assignment are based on Structure's estimates for admixture of the overall collection location (i.e. a location that is 50% red and 50% blue may be comprised of individuals with any level of admix between red and blue that averages, among individuals, to 50% red and 50% blue)


CURRITUCK
CAMDEN
PASQUOTANK
PERQUIMANS
TYRRELL
DARE
HYDE
WASHINGTON
MARTIN
BEAUFORT
PITT
CRAVEN
PAMLICO
CARTERET
JONES
ONSLOW
GREENE
LENOIR
WAYNE
DUPLIN
PENDER
NEW HANOVER
BRUNSWICK
EDGECOMBE
NASH
HALIFAX
WARREN
FRANKLIN
VANCE
GRANVILLE
JOHNSTON
WAKE
CUMBERLAND
COLUMBUS
DURHAM
PERSON
ORANGE
CASWELL
ALAMANCE
CHATHAM
HOKE
ROBESON
ROCKINGHAM
GUILFORD
RANDOLPH
MOORE
MONTGOMERY
RICHMOND
STOKES
FORSYTH
DAVIDSON
STANLY
UNION
SURRY
YADKIN
DAVIE
CABARRUS
MECKLENBURG
ALLEGHANY
WILKES
ALEXANDER
CATAWBA
LINCOLN
ASHE
WATAUGA
CALDWELL
BURKE
RUTHERFORD
AVERY
MCDOWELL
POLK
MITCHELL
YANCEY
BUNCOMBE
HENDERSON
MADISON
TRANSYLVANIA
HAYWOOD
JACKSON
SWAIN
MACON
GRAHAM
CLAY
CHEROKEE

References

Baerson, S. R., D. J. Rodriguez, M. Tran, Y. Feng, N. A. Biest *et al*, 2002a Glyphosate-resistant goosegrass. identification of a mutation in the target enzyme 5-enolpyruvylshikimate-3-phosphate synthase. Plant Physiology **129:** 1265-1275.

Baerson, S. R., D. J. Rodriguez, N. A. Biest, M. Tran, J. You *et al*, 2002b Investigating the mechanism of glyphosate resistance in rigid ryegrass (*Lolium ridigum*). Weed Science **50:** 721-730.

Barrett, R. D. H., and D. Schluter, 2008 Adaptation from standing genetic variation. Trends in Ecology & Evolution **23:** 38-44.

Burgos, N. R., A. L. Alcober, A. Lawton-Rauh, B. Rauh, L. Estorninos Jr *et al*, 2008 The spread and population genetics of glyphosate-resistant palmer amaranth in arkansas. Summaries of Arkansas Cotton Research **2008**: 103.

Carvalho, L. B., F. González-Torralva, H. Cruz Hipolito, A. M. Rojano-Delgado, J. Gil-Humanes *et al*, 2011 A pool of resistance mechanisms to glyphosate in *Digitaria insularis*. Journal of Agricultural Food Chemistry.

Catania, F., M. O. Kauer, P. J. Daborn, J. L. Yen, R. H. Ffrench-Constant *et al*, 2004 World-wide survey of an accord insertion and its association with DDT resistance in drosophila melanogaster. Molecular Ecology **13:** 2491-2504.

Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro *et al*, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS biology **2:** e109.

Cresko, W. A., A. Amores, C. Wilson, J. Murphy, M. Currey *et al*, 2004 Parallel genetic basis for repeated evolution of armor loss in alaskan threespine stickleback populations. Proceedings of the National Academy of Sciences of the United States of America **101:** 6050-6055.

Cruz-Hipolito, H., A. Rojano-Delgado, J. A. Domínguez-Valenzuela, A. Heredia, M. D. L. de Castro *et al*, Glyphosate tolerance by *Clitoria ternatea* and *Neonotonia wightii* plants involves differential absorption and translocation of the herbicide. Plant Soil: 1-10.

de Carvalho, L. B., H. Cruz-Hipolito, F. González-Torralva, da Costa Aguiar Alves, P.L., P. J. Christoffoleti *et al*, 2011 Detection of sourgrass (*Digitaria insularis*) biotypes resistant to glyphosate in brazil. Weed Science **59:** 171-176.

Dowling, T. E., D. P. Martasian and W. R. Jeffery, 2002 Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. Molecular Biology and Evolution **19:** 446-455.

Eizirik, E., N. Yuhki, W. E. Johnson, M. Menotti-Raymond, S. S. Hannah *et al*, 2003 Molecular genetics and evolution of melanism in the cat family. Current Biology **13:** 448-453.

Evanno, G., S. Regnaut and J. Goudet, 2005 Detecting the number of clusters of individuals using the software structure: A simulation study. Molecular Ecology **14:** 2611-2620.

Evans, P. D., N. Mekel-Bobrov, E. J. Vallender, R. R. Hudson and B. T. Lahn, 2006 Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic homo lineage. Proceedings of the National Academy of Sciences **103:** 18178-18183.

Ewing, B., and P. Green, 1998 Base-calling of automated sequencer traces UsingPhred. II. Error Probabilities. Genome Research **8:** 186-194.

Ewing, B., L. Hillier, M. C. Wendl and P. Green, 1998 Base-calling of automated sequencer traces UsingPhred. I. Accuracy Assessment. Genome Research **8:** 175-185.

Feldman, C. R., E. D. Brodie, E. D. Brodie and M. E. Pfrender, 2009 The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. Proceedings of the National Academy of Sciences **106:** 13415-13420.

Feng, P. C. C., M. Tran, T. Chiu, R. D. Sammons, G. R. Heck *et al*, 2009 Investigations into glyphosate-resistant horseweed (*Conyza canadensis*): Retention, uptake, translocation, and metabolism.

Gaines, T. A., 2010 *Molecular Genetics of Glyphosate Resistance in Palmer Amaranth (Amaranthus Palmeri L.)*. Colorado State University.

Gaines, T. A., D. L. Shaner, S. M. Ward, J. E. Leach, C. Preston *et al*, 2011 Mechanism of resistance of evolved glyphosate-resistant palmer amaranth (*Amaranthus palmeri*). Journal of Agricultural Food Chemistry.

Gaines, T. A., A. A. Wright, W. T. Molin, L. Lorentz, C. W. Riggins *et al*, 2013 Identification of genetic elements associated with EPSPS gene amplification. PloS One **8:** e65819.

Gaines, T. A., W. Zhang, D. Wang, B. Bukun, S. T. Chisholm *et al*, 2010 Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proceedings of the National Academy of Sciences **107:** 1029-1034.

Gaylon D. Morgan, P. A. Baumann and J. M. Chandler, 2001 Competitive impact of palmer amaranth (*Amaranthus palmeri*) on cotton (*Gossypium hirsutum*) development and yield. Weed Technology **15:** 408-412.

Ginzinger, D. G., 2002 Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. Exp. Hematol. **30:** 503-512.

González-Torralva, F., A. M. Rojano-Delgado, M. D. Luque de Castro, N. Mülleder and R. De Prado, 2012 Two non-target mechanisms are involved in glyphosate-resistant horseweed (*Conyza canadensis* L. cronq.) biotypes. Journal of Plant Physiology: 1673.

Hall, T., 2001 Biolign alignment and multiple contig editor. North Carolina State University, Department of Microbiology, Raleigh, North Carolina.

Hall, T. A., 1999 *BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT*.

Heap, I. M., *International Survey of Herbicide Resistant Weeds*. Online. Internet. Wednesday, November 06, 2013. Available www.weedscience.com

Hubisz, M. J., D. Falush, M. Stephens and J. K. Pritchard, 2009 Inferring weak population structure with the assistance of sample group information. Molecular ecology resources **9:** 1322-1332.

Kaundun, S. S., R. P. Dale, I. A. Zelaya, G. Dinelli, I. Marotti *et al*, 2011 A novel P106L mutation in EPSPS and an unknown mechanism(s) act additively to confer resistance to glyphosate in a south African *Lolium rigidum* population. Journal of Agricultural Food Chemistry.

Kays, R., A. Curtis and J. J. Kirchman, 2010 Rapid adaptive evolution of northeastern coyotes via hybridization with wolves. Biology Letters **6:** 89-93.

Kenneth D. Whitney, Rebecca A. Randell and Loren H. Rieseberg, 2006 Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. The American Naturalist **167:** 794-807.

Lorraine-Colwill, D. F., S. B. Powles, T. R. Hawkes, P. H. Hollinshead, S. A. J. Warner *et al*, 2002 Investigations into the mechanism of glyphosate resistance in *Lolium rigidum*. Pesticide Biochemistry and Physiology **74:** 62-72.

Michitte, P., R. De Prado, N. Espinoza, J. Pedro Ruiz-Santaella and C. Gauvrit, 2007 Mechanisms of resistance to glyphosate in a ryegrass (*Lolium multiflorum*) biotype from chile. Weed Science **55:** 435-440.

Michitte, P., R. De Prado, N. Espinosa and C. Gauvrit, 2005 Glyphosate resistance in a Chilean *Lolium multiflorum*. Communications in Agcultral and Applied Biological Sciences **70:** 507-513.

Mundy, N. I., N. S. Badcock, T. Hart, K. Scribner, K. Janssen *et al*, 2004 Conserved genetic basis of a quantitative plumage trait involved in mate choice. Science **303:** 1870-1873.

Ng, C. H., R. Wickneswari, S. Salmijah, Y. T. Teng and B. S. Ismail, 2003 Gene polymorphisms in glyphosate-resistant and -susceptible biotypes of *Eleusine indica* from Malaysia. Weed Research **43:** 108-115.

Norsworthy, J. K., N. R. Burgos and L. R. Oliver, 2001 Differences in weed tolerance to glyphosate involve different mechanisms 1. Weed Technology **15:** 725-731.

Perez-Jones, A., K. Park, N. Polge, J. Colquhoun and C. Mallory-Smith, 2007 Investigating the mechanisms of glyphosate resistance in *Lolium multiflorum*. Planta **226:** 395-404.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics **155:** 945-959.

Ribeiro, D. N., Z. Pan, S. O. Duke, V. K. Nandula, B. S. Baldwin *et al*, 2013 Involvement of facultative apomixis in inheritance of EPSPS gene amplification in glyphosate-resistant *Amaranthus palmeri*. Planta 1-14.

Rowland, M. W., Don S. Murray and L. M. Verhalen, 1999 Full-season palmer amaranth (*Amaranthus palmeri*) interference with cotton (*Gossypium hirsutum*). Weed Science **47:** 305-309.

Rozen, S., and H. J. Skaletsky, 2000 Primer3 on the WWW for general users and for biologist programmers., pp. 365-368 in , edited by K. S. Misener. Humana Press, Totowa, NJ.

Salas, R. A., F. E. Dayan, Z. Pan, S. B. Watson, J. W. Dickson *et al*, 2012 EPSPS gene amplification in glyphosate-resistant Italian ryegrass (*Lolium perenne* ssp multiflorum) from arkansas, USA. Pest Management Science.

Schlenke, T. A., and D. J. Begun, 2004 Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proceedings of the National Academy of Sciences of the United States of America **101:** 1626-1631.

Shimizu, K. K., J. M. Cork, A. L. Caicedo, C. A. Mays, R. C. Moore *et al*, 2004 Darwinian selection on a selfing locus. Science **306:** 2081-2084.

Steiner, C. C., J. N. Weber and H. E. Hoekstra, 2007 Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biology **5:** e219.

Stephens, M., N. J. Smith and P. Donnelly, 2001 A new statistical method for haplotype reconstruction from population data. The American Journal of Human Genetics **68:** 978-989.

Strecker, U., L. Bernatchez and H. Wilkens, 2003 Genetic divergence between cave and surface populations of astyanax in Mexico (characidae, teleostei). Molecular Ecology **12:** 699-710.

Theron, E., K. Hawkins, E. Bermingham, R. E. Ricklefs and N. I. Mundy, 2001 The molecular basis of an avian plumage polymorphism in the wild: A melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, coereba flaveola. Current Biology **11:** 550-557.

Va, D., N. De, N. Broeck, T. Maes, M. Sauer *et al*, 1998 Transposon display identifies individual transposable elements in high copy number lines. The Plant Journal **13:** 121-129.

Wakelin, A. M., D. F. Lorraine-Colwill and C. Preston, 2004 Glyphosate resistance in four different populations of *Lolium rigidum* is associated with reduced translocation of glyphosate to meristematic zones. Weed Research **44:** 453-459.

WAKELIN, A., and C. PRESTON, 2006 A target-site mutation is present in a glyphosate-resistant *Lolium rigidum* population. Weed Research **46:** 432-440.

Weaver, L. M., and K. M. Herrmann, 1997 Dynamics of the shikimate pathway in plants. Trends Plant Science **2:** 346-351.

Whitaker, J. R., 2009 *Distribution, Biology, and Management of Glyphosate-Resistant Palmer Amaranth in North Carolina*. ProQuest.

Yu, Q., A. Cairns and S. Powles, 2007 Glyphosate, Paraquat and ACCase multiple herbicide resistance evolved in a Lolium rigidum biotype. Planta **225:** 499-513.

Yuan, J. S., A. Reed, F. Chen and C. N. Stewart, 2006 Statistical analysis of real-time PCR data. BMC Bioinformatics **7:** 85.

Zulet, A., M. Gil-Monreal, J. G. Villamor, A. Zabalza, van der Hoorn, Renier AL *et al*, 2013 Proteolytic pathways induced by herbicides that inhibit amino acid biosynthesis. PloS One **8:** e73847.

CONCLUSIONS

This dissertation applies the theories of molecular evolution to better understand the dynamics of adaptation in the glyphosate resistant weed *Amaranthus palmeri*. In the introduction I review literature on adaptive evolution and herbicide resistance of weeds in agro-ecosystems as a model thereof. In chapter one, I established a foundation for understanding the dynamics of adaptation by using Bayes' theorem to estimate the best fit phylogeny for the genus *Amaranthus*. This showed that the distribution of weedy amaranths and the habitat descriptions of the non-weedy species strongly suggest that weediness readily evolves from the ruderal habit common to the genus.

In chapter two, I investigated the sequence constraints on *EPSPS* and how that could impact adaptation. The non-synonymous mutation accumulation, a proxy for sequence constraint, is average in EPSPS, so extreme sequence constraint is likely not related to the lack of high-glyphosate-resistance conferring EPSPS point mutations. I found evidence for a high density of methylated CpG dinucleotides. A hypothesis suggested by this is that under glyphosate stress there are methylation changes to EPSPS that cause EPSPS to become a site of replication initiation— unmethylated CpG islands are known to be involved in the initiation of replication in vertebrates. This would result in fragments of EPSPS genomic sequence in the nucleus that could potentially be incorporated into the genome through non-homologous end joining. This would result in increased EPSPS copy number under herbicide stress without traditional transposon mediation. This would also potentially explain the small *EPSPS* DNA fragments found in the Southern blot

analysis (Figure 3.5). Future experiments could determine the veracity of this hypothesis. This would involve looking for *EPSPS* gene containing fragments of DNA not incorporated into chromosomes in the nuclei of amaranth or other plant cells under glyphosate stress. This would have the potential to be a very important finding as such a mechanism of genome rearrangement has never before been described and would greatly contribute to our understanding of how genomes change through time, particularly regarding changes to gene family size.

In chapter three, I show that palmer amaranths with elevated *EPSPS* gene copy number are growing in North Carolina. There is no evidence that the observed increase *EPSPS* copy number is part of standing variation in *Amaranthus*. However, the fact that four of the five identified population clusters in NC are statistically associated with a glyphosate resistance phenotype suggests that more than one adaptive event is responsible for the observed resistance in NC. Current work cannot determine if these represent multiple *de novo* events in NC, multiple introgression events, or some combination. This indicates another important direction for future work. Expanding the study to include more locations in the Southeastern US would allow more context for these findings and address the questions left at the end of the chapter. Do these represent *de novo* events in NC or are they events from other parts of the US that introgressed into NC? Answering these questions with analyses of population structure that include more locations would give better insight into the dynamics of the spread of glyphosate resistance. If the events are *de novo* in North Carolina and not related to resistant individuals found elsewhere then this

suggests that *EPSPS* gene copy number proliferation has evolved multiple times in a very short span of time, this has important consequences for herbicide management. If the events are introgressions from a much smaller number of adaptive events in the US then there are very different conclusions regarding the spread of resistance; rather than being something inherent about the species or the herbicide that is causing rapid adaptation it is a matter of uncontrolled spread of seed and pollen, and given the findings in chapter two regarding high levels of admixture near interstates this spread could be human mediated. Following these or other lines of further research into glyphosate resistance in *Amaranthus* will have benefits for our environment and our understanding of adaptive evolution. An understanding of the specific dynamics at work in this system could potentially lead to improved management practices that slow adaptation of weeds to herbicides and thus protect our food supply and our environment from the consequences of increased chemical pesticide application or soil erosion caused by deep tillage. Understanding this system will also contribute to a broader understanding of the underlying forces of molecular evolution and the dynamics of adaptation. This system represents a unique opportunity to observe a defined selection pressure effect changes in the phenotypes (and thus genomes) of a large number of complex out-crossing organisms as it happens.

APPENDICIES

APENDIX A

LIST OF ABBREVIATIONS

A – Adenine

AA – Amino acid

ABC – AMP binding cassette

ACCase – Acetyl Coenzyme A Carboxylase

AFLP – Amplified fragment length polymorphism

ALS – Acetolactate synthase

AMP – Adenosine monophosphate

AMPA – aminomethyl-phosphonic acid

ANOVA – Analysis of variation

ANCOVA – Analysis of co-variation

BEST – Bayesian estimation of species tree (software)

BLAST – Basic Local Alignment Search Tool

BS – Bootstrap

C – Cytosine

CA – California

cDNA – Copy DNA

cpDNA – Chloroplast DNA

CI – Confidence interval

CL – Confidence limit

C_t – Threshold Cycle

DAHP – 3-deoxy-D-arabinoheptulosonate 7-phosphate

DDT – Dichlorodiphenyltrichloroethane

DNA – Deoxyribonucleaic Acid

EPSP – 5-enolpyruvylshikimate-3-phosphate

EPSPS – 5-enolpyruvylshikimate-3-phosphate synthase

FNA – Flora of North America

FUBAR – Fast Unconstrained Bayesian Approximation of Selection (analysis)

GA – Georgia

Gbp – Gigabase pair (1,000,000,000bp)

GM – Genetically Modified

GPS – Global Positioning Satellite

GST – Glutathione S-Transferase

GTR(+G) – General Time Reversible (+ gamma)

HKY – Hasegawa, Kishino and Yano (independent trasition and transversion rates)

HPPD – *p*-hydroxyphenylpyruvate dioxygenase

HRAC – Herbicide Resistance Action Committee

ITS – Internal transcribed spacer

IUPAC – International Union of Pure and Applied Chemists

LDL – Lower Decision Limit

Mbp – Mega Base Pair (1,000,000bp)

MD – Maryland

MEGA – Molecular Evolutionary Genetic Analysis (software)

MITE – Miniature Inverted-Repeat Transposable Elements

MSA – Multiple Sequence Alignment

mtDNA – Mitochondrial DNA

NC – North Carolina

NCSU – North Carolina State University

nDNA – Nuclear DNA

OTU – Operational Taxonomic Unit

PCR – Polymerase Chain Reaction

PEP – Phosphoenol Pyruvate

PPB – Posterior Probability (Bayesian analysis)

PPC – Posterior Probability (Coalescent analysis)

RFLP – Restriction Fragment Length Polymorphism

RMSE – Root Mean Square Error

RNA – Ribonucleaic Acid

T – Thiamine

UDL – Upper Decision Limit

US – United States

USA – United States of America

USDA – United States' Department of Agriculture

WSSA – Weed Science Society of America

WT – Wild type

APENDIX B

SUPPLEMENTAL FIGURES

Figure B.1: Gene tree of the *A07* (putative endosomal P24A protein) locus. The evolutionary history was inferred using the Bayesian estimation method based on the General time reversible model with a discrete Gamma distribution used to model evolutionary rate differences among sites (4 categories (+*G*)). All three codon positions shared estimates of the evolutionary model parameters, tree topology, and branch lengths. Initial trees for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Posterior probability of each branch bifurcation is indicated above the branch. The analysis involved 56 nucleotide sequences. There were a total of 747 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.

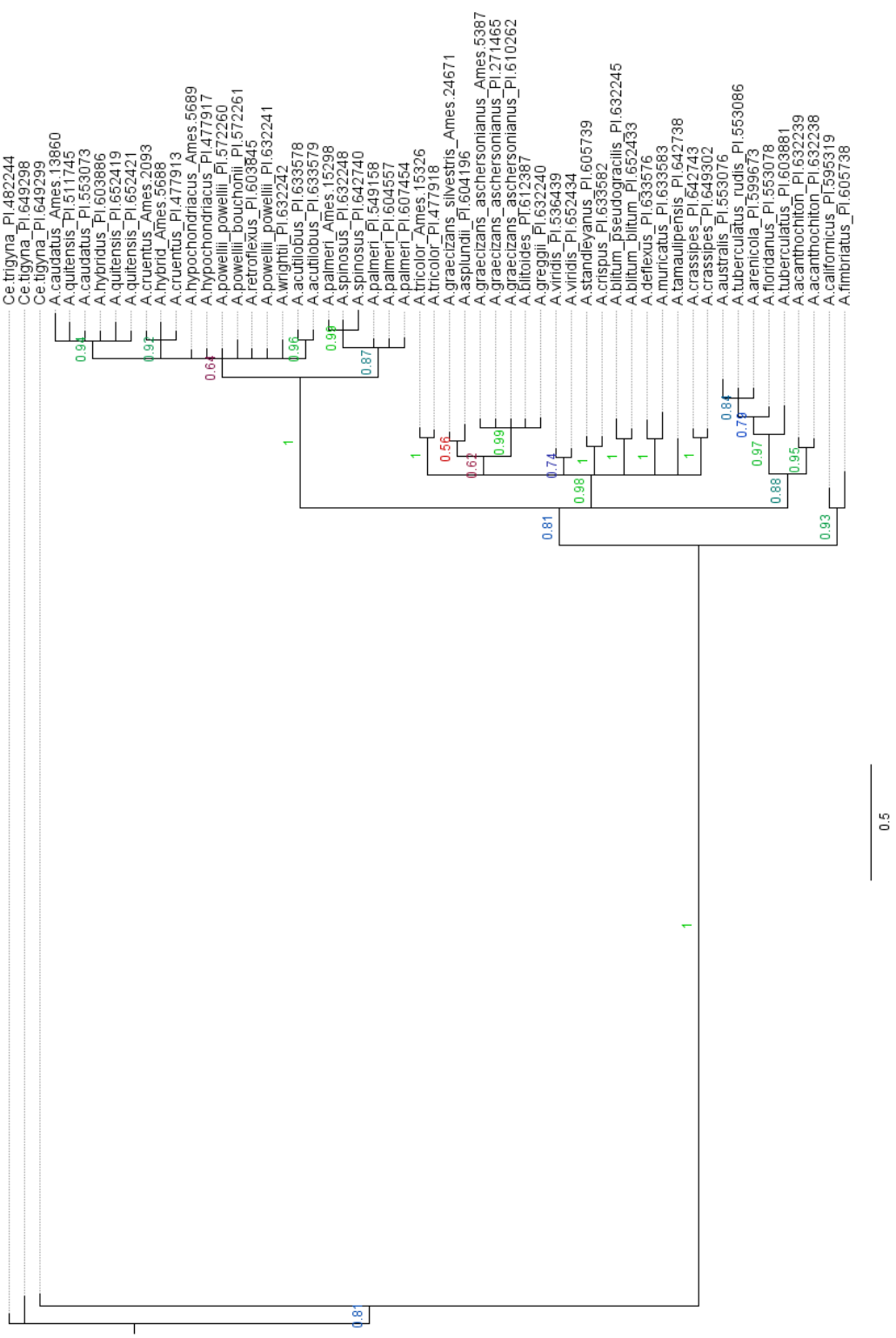

Ce.trigyna_PI.482244
Ce.tigyna_PI.649298
Ce.tigyna_PI.649299
A.caudatus_Ames.13860
A.quitensis_PI.511745
A.caudatus_PI.553073
A.hybridus_PI.603886
A.quitensis_PI.652419
A.quitensis_PI.652421
A.cruentus_Ames.2093
A.hybrid_Ames.5688
A.cruentus_PI.477913
A.hypochondriacus_Ames.5689
A.hypochondriacus_PI.477917
A.powellii_powellii_PI.572260
A.powellii_bouchonii_PI.572261
A.retroflexus_PI.603845
A.powellii_powellii_PI.632241
A.wrightii_PI.632242
A.acutilobus_PI.633578
A.acutilobus_PI.633579
A.palmeri_Ames.15298
A.spinosus_PI.632248
A.spinosus_PI.642740
A.palmeri_PI.549158
A.palmeri_PI.604557
A.palmeri_PI.607454
A.tricolor_Ames.15326
A.tricolor_PI.477918
A.graecizans_silvestris_Ames.24671
A.asplundii_PI.604196
A.graecizans_aschersonianus_Ames.5387
A.graecizans_aschersonianus_PI.271465
A.graecizans_aschersonianus_PI.610262
A.blitoides_PI.612387
A.greggii_PI.632240
A.viridis_PI.536439
A.viridis_PI.652434
A.standleyanus_PI.605739
A.crispus_PI.633582
A.blitum_pseudogracilis_PI.632245
A.blitum_blitum_PI.652433
A.deflexus_PI.633576
A.muricatus_PI.633583
A.tamaulipensis_PI.642738
A.crassipes_PI.642743
A.crassipes_PI.649302
A.australis_PI.553076
A.tuberculatus_rudis_PI.553086
A.arenicola_PI.599673
A.floridanus_PI.553078
A.tuberculatus_PI.603881
A.acanthochiton_PI.632239
A.acanthochiton_PI.632238
A.californicus_PI.595319
A.fimbriatus_PI.605738
0.5

Figure B.2: Gene tree of the *A36* (putative RNA DEAD box helicase) locus. The evolutionary history was inferred using the Bayesian estimation method based on the General time reversible model with a discrete Gamma distribution used to model evolutionary rate differences among sites (4 categories (+*G*)). All three codon positions shared estimates of the evolutionary model parameters, tree topology, and branch lengths. Initial trees for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Posterior probability of each branch bifurcation is indicated above the branch. The analysis involved 56 nucleotide sequences. There were a total of 788 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.

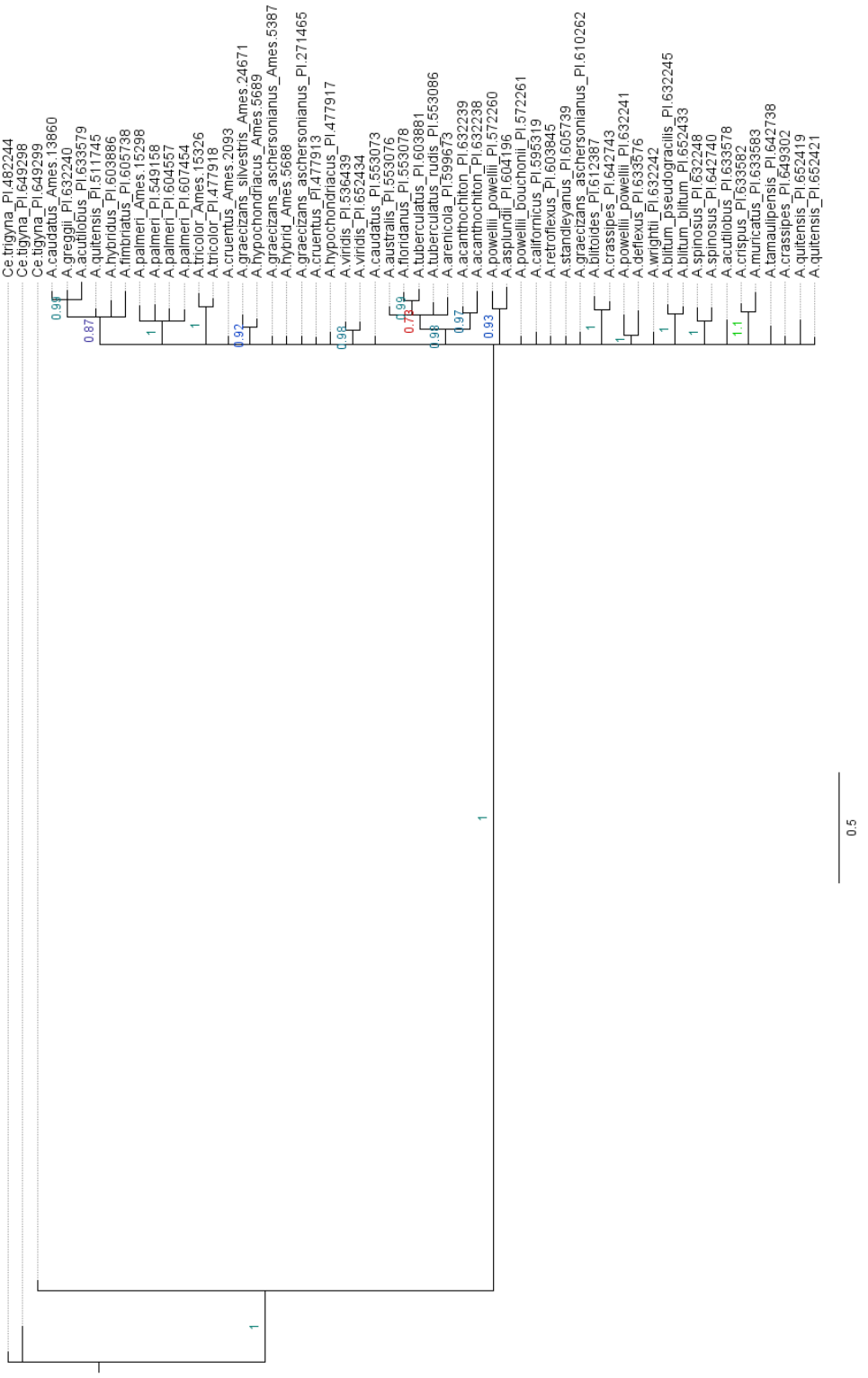

Ce.trigyna_PI.482244
Ce.trigyna_PI.649298
Ce.trigyna_PI.649299
A.caudatus_Ames.13860
A.greggii_PI.632240
A.acutilobus_PI.633579
A.quitensis_PI.511745
A.hybridus_PI.603886
A.fimbriatus_PI.605738
A.palmeri_Ames.15298
A.palmeri_PI.549158
A.palmeri_PI.604557
A.palmeri_PI.607454
A.tricolor_Ames.15326
A.tricolor_PI.477918
A.cruentus_Ames.2093
A.graecizans_silvestris_Ames.24671
A.hypochondriacus_Ames.5689
A.graecizans_aschersonianus_Ames.5387
A.hybrid_Ames.5688
A.graecizans_aschersonianus_PI.271465
A.cruentus_PI.477913
A.hypochondriacus_PI.477917
A.viridis_PI.536439
A.viridis_PI.652434
A.caudatus_PI.553073
A.australis_PI.553076
A.floridanus_PI.553078
A.tuberculatus_PI.603881
A.tuberculatus_rudis_PI.553086
A.arenicola_PI.599673
A.acanthochiton_PI.632239
A.acanthochiton_PI.632238
A.powellii_powellii_PI.572260
A.aspundii_PI.604196
A.powellii_bouchonii_PI.572261
A.californicus_PI.595319
A.retroflexus_PI.603845
A.standleyanus_PI.605739
A.graecizans_aschersonianus_PI.610262
A.blitoides_PI.612387
A.crassipes_PI.642743
A.powellii_powellii_PI.632241
A.deflexus_PI.633576
A.wrightii_PI.632242
A.bitum_pseudogracilis_PI.632245
A.bitum_bitum_PI.652433
A.spinosus_PI.632248
A.spinosus_PI.642740
A.acutilobus_PI.633578
A.crispus_PI.633582
A.muricatus_PI.633583
A.tamaulipensis_PI.642738
A.crassipes_PI.649302
A.quitensis_PI.652419
A.quitensis_PI.652421
0.5

Figure B.3: Gene tree of the *A37* (putative serine-type endopeptidase) locus. The evolutionary history was inferred using the Bayesian estimation method based on the General time reversible model with a discrete Gamma distribution used to model evolutionary rate differences among sites (4 categories (+*G*)). All three codon positions shared estimates of the evolutionary model parameters, tree topology, and branch lengths. Initial trees for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Posterior probability of each branch bifurcation is indicated above the branch. The analysis involved 56 nucleotide sequences. There were a total of 857 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.

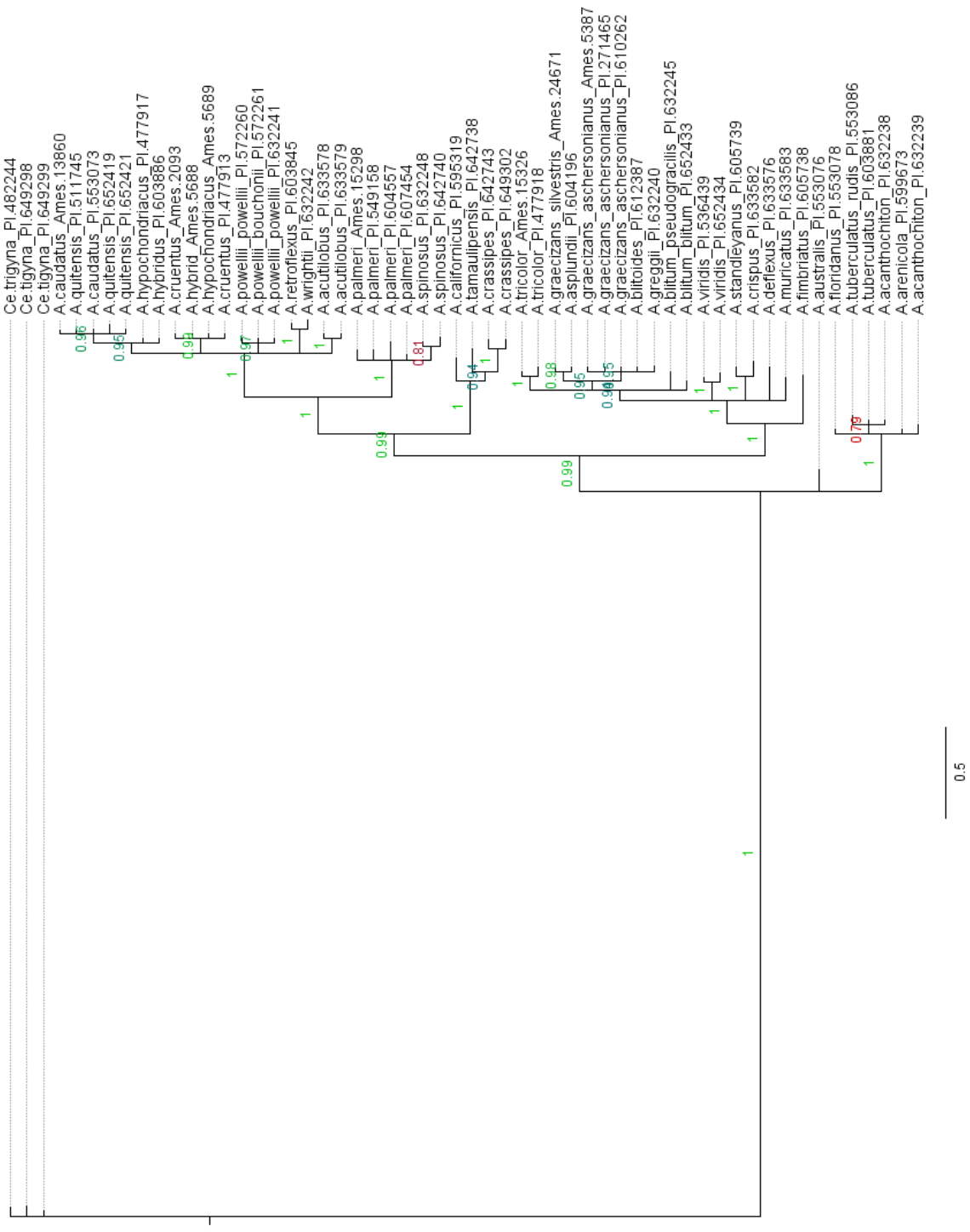

Ce.trigyna_PI.482244
Ce.trigyna_PI.649298
Ce.trigyna_PI.649299
A.caudatus_Ames.13860
A.quitensis_PI.511745
A.caudatus_PI.563073
A.quitensis_PI.652419
A.quitensis_PI.652421
A.hypochondriacus_PI.477917
A.hybridus_PI.603886
A.cruentus_Ames.2093
A.hybrid_Ames.5688
A.hypochondriacus_Ames.5689
A.cruentus_PI.477913
A.powellii_powellii_PI.572260
A.powellii_bouchonii_PI.572261
A.powellii_powellii_PI.632241
A.retroflexus_PI.603845
A.wrightii_PI.632242
A.acutiobus_PI.633578
A.acutiobus_PI.633579
A.palmeri_Ames.15298
A.palmeri_PI.549158
A.palmeri_PI.604557
A.palmeri_PI.607454
A.spinosus_PI.632248
A.spinosus_PI.642740
A.californicus_PI.595319
A.tamaulipensis_PI.642738
A.crassipes_PI.642743
A.crassipes_PI.649302
A.tricolor_Ames.15326
A.tricolor_PI.477918
A.graecizans_silvestris_Ames.24671
A.asplundii_PI.604196
A.graecizans_aschersonianus_Ames.5387
A.graecizans_aschersonianus_PI.271465
A.graecizans_aschersonianus_PI.610262
A.blitoides_PI.612387
A.greggii_PI.632240
A.blitum_pseudogracilis_PI.632245
A.blitum_blitum_PI.652433
A.viridis_PI.536439
A.viridis_PI.652434
A.standleyanus_PI.605739
A.crispus_PI.633582
A.deflexus_PI.633576
A.muricatus_PI.633583
A.fimbriatus_PI.605738
A.australis_PI.553076
A.floridanus_PI.553078
A.tuberculatus_rudis_PI.553086
A.tuberculatus_PI.603881
A.acanthochiton_PI.632238
A.arenicola_PI.599673
A.acanthochiton_PI.632239
0.5

Figure B.4: Gene tree of the *A40* (putative glutaredoxin) locus. The evolutionary history was inferred using the Bayesian estimation method based on the General time reversible model with a discrete Gamma distribution used to model evolutionary rate differences among sites (4 categories (+*G*)). All three codon positions shared estimates of the evolutionary model parameters, tree topology, and branch lengths. Initial trees for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Posterior probability of each branch bifurcation is indicated above the branch. The analysis involved 56 nucleotide sequences. There were a total of 701 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.

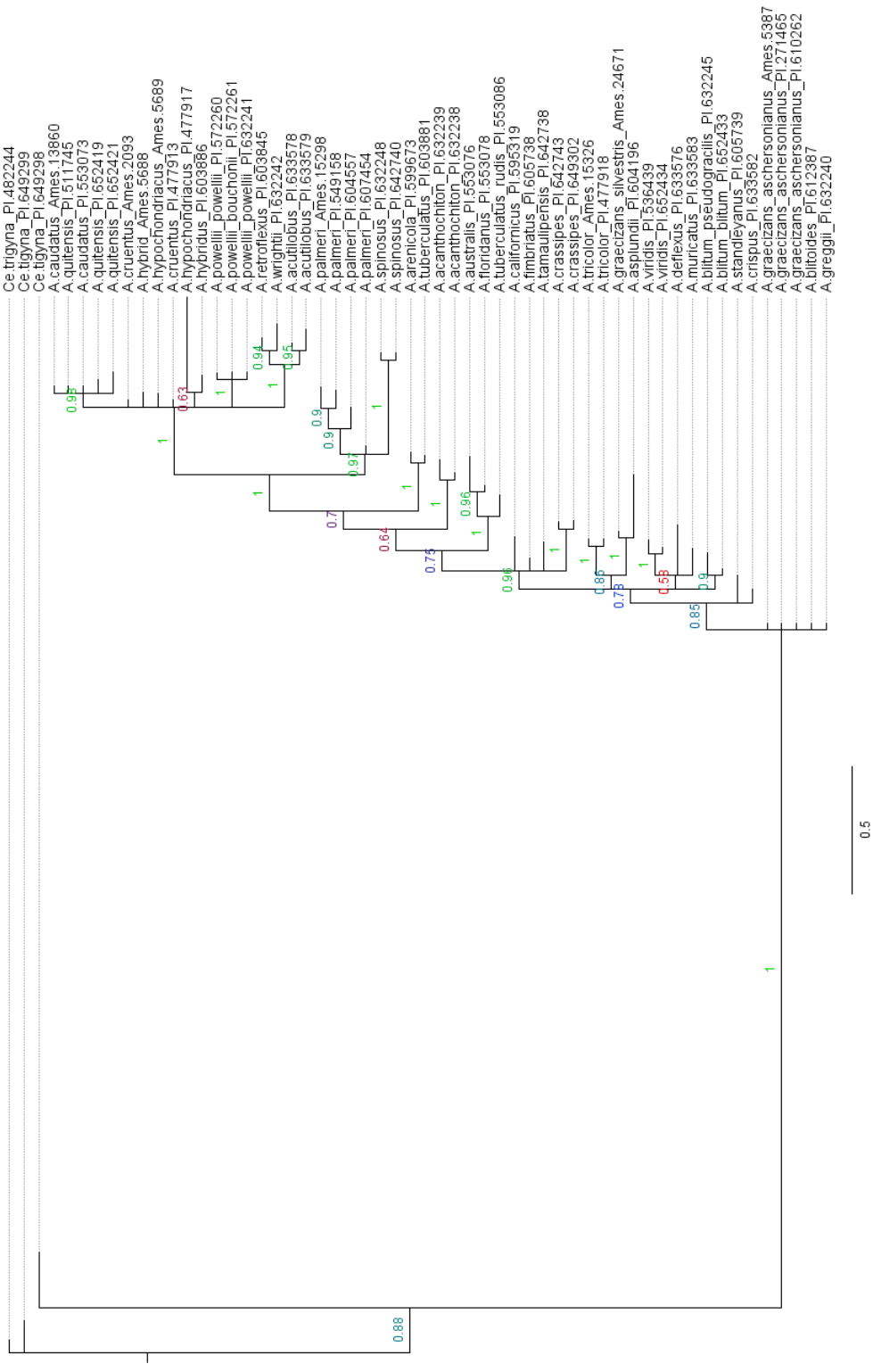

Ce.trigyna_PI.482244
Ce.trigyna_PI.649299
Ce.trigyna_PI.649298
A.caudatus_Ames.13860
A.quitensis_PI.511745
A.caudatus_PI.553073
A.quitensis_PI.652419
A.quitensis_PI.652421
A.cruentus_Ames.2093
A.hybrid_Ames.5688
A.hypochondriacus_Ames.5689
A.cruentus_PI.477913
A.hypochondriacus_PI.477917
A.hybridus_PI.603886
A.powellii_powellii_PI.572260
A.powellii_bouchonii_PI.572261
A.powellii_powellii_PI.632241
A.retroflexus_PI.603845
A.wrightii_PI.632242
A.acutilobus_PI.633578
A.acutilobus_PI.633579
A.palmeri_Ames.15298
A.palmeri_PI.549158
A.palmeri_PI.604557
A.palmeri_PI.607454
A.spinosus_PI.632248
A.spinosus_PI.642740
A.arenicola_PI.599673
A.tuberculatus_PI.603881
A.acanthochiton_PI.632239
A.acanthochiton_PI.632238
A.australis_PI.553076
A.floridanus_PI.553078
A.tuberculatus_rudis_PI.553086
A.californicus_PI.595319
A.fimbriatus_PI.605738
A.tamaulipensis_PI.642738
A.crassipes_PI.642743
A.crassipes_PI.649302
A.tricolor_Ames.15326
A.tricolor_PI.477918
A.graecizans_silvestris_Ames.24671
A.asplundii_PI.604196
A.viridis_PI.536439
A.viridis_PI.652434
A.deflexus_PI.633576
A.muricatus_PI.633583
A.blitum_pseudogracilis_PI.632245
A.blitum_blitum_PI.652433
A.standleyanus_PI.605739
A.crispus_PI.633582
A.graecizans_aschersonianus_Ames.5387
A.graecizans_aschersonianus_PI.271465
A.graecizans_aschersonianus_PI.610262
A.blitoides_PI.612387
A.greggii_PI.632240
0.5

Figure B.5: Gene tree of the *internal transcriped spacer* (*ITS*) *1* and *2* loci. The evolutionary history was inferred using the Bayesian estimation method based on the General time reversible model with a discrete Gamma distribution used to model evolutionary rate differences among sites (4 categories (+*G*)). Both ITS loci shared estimates of the evolutionary model parameters, tree topology, and branch lengths. Initial trees for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Posterior probability of each branch bifurcation is indicated above the branch. The analysis involved 56 nucleotide sequences. There were a total of 460 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.

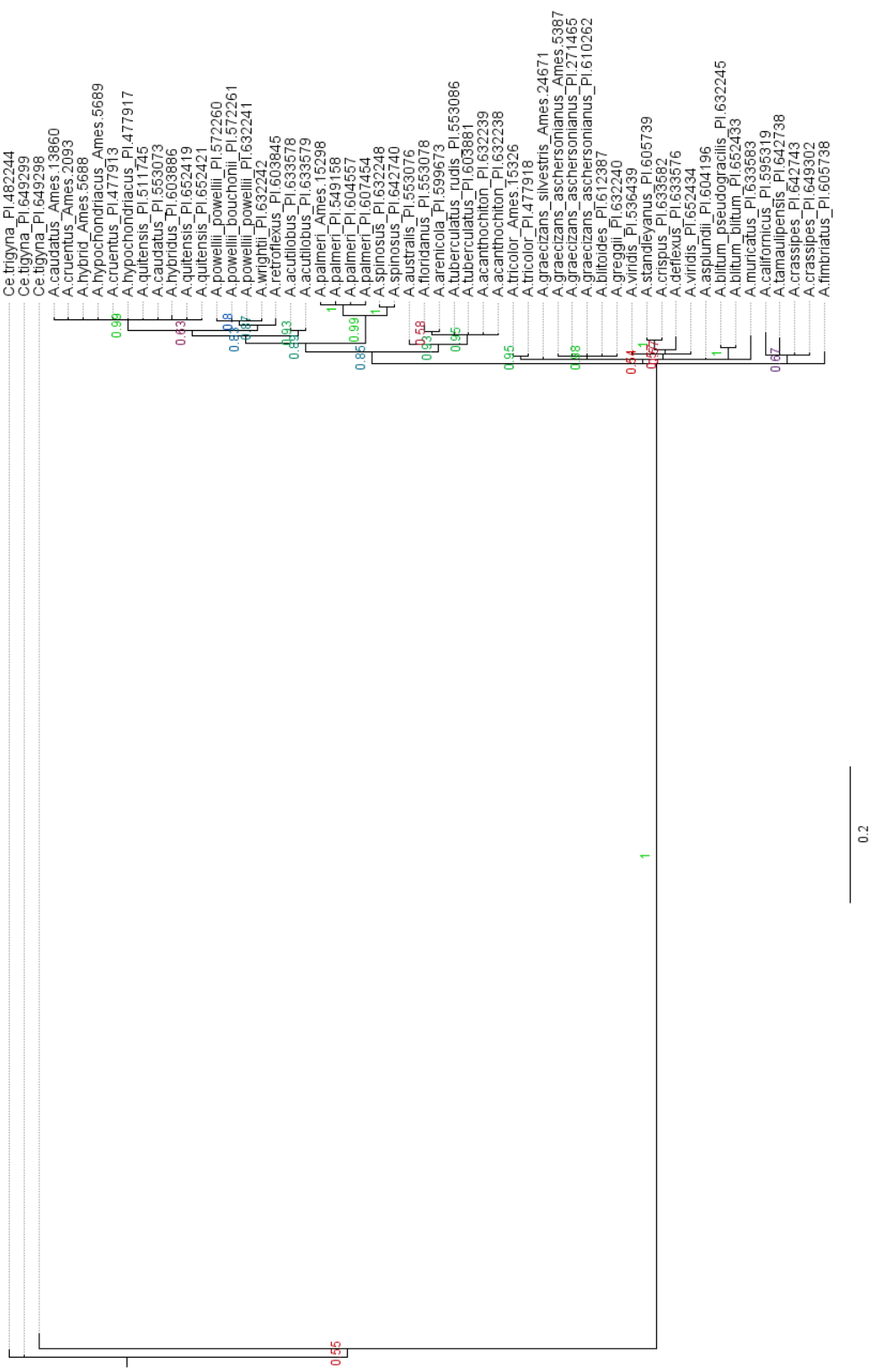

Ce.trigyna_PI.482244
Ce.trigyna_PI.649299
Ce.trigyna_PI.649298
A.caudatus_Ames.13860
A.cruentus_Ames.2093
A.hybrid_Ames.5688
A.hypochondriacus_Ames.5689
A.cruentus_PI.477913
A.hypochondriacus_PI.477917
A.quitensis_PI.511745
A.caudatus_PI.553073
A.hybridus_PI.603886
A.quitensis_PI.652419
A.quitensis_PI.652421
A.powellii_powellii_PI.572260
A.powellii_bouchonii_PI.572261
A.powellii_powellii_PI.632241
A.wrightii_PI.632242
A.retroflexus_PI.603845
A.acutilobus_PI.633578
A.acutilobus_PI.633579
A.palmeri_Ames.15298
A.palmeri_PI.549158
A.palmeri_PI.604557
A.palmeri_PI.607454
A.spinosus_PI.632248
A.spinosus_PI.642740
A.australis_PI.553076
A.floridanus_PI.553078
A.arenicola_PI.599673
A.tuberculatus_rudis_PI.553086
A.tuberculatus_PI.603881
A.acanthochiton_PI.632239
A.acanthochiton_PI.632238
A.tricolor_Ames.15326
A.tricolor_PI.477918
A.graecizans_silvestris_Ames.24671
A.graecizans_aschersonianus_Ames.5387
A.graecizans_aschersonianus_PI.271465
A.graecizans_aschersonianus_PI.610262
A.blitoides_PI.612387
A.greggii_PI.632240
A.viridis_PI.536439
A.standleyanus_PI.605739
A.crispus_PI.633582
A.deflexus_PI.633576
A.viridis_PI.652434
A.asplundii_PI.604196
A.blitum_pseudogracilis_PI.632245
A.blitum_blitum_PI.652433
A.muricatus_PI.633583
A.californicus_PI.595319
A.tamaulipensis_PI.642738
A.crassipes_PI.642743
A.crassipes_PI.649302
A.fimbriatus_PI.605738
0.2

Figure B.6: Gene tree of the *MatK* (Maturase K) locus. The evolutionary history was inferred using the Bayesian estimation method based on the General time reversible model with a discrete Gamma distribution used to model evolutionary rate differences among sites (4 categories (+*G*)). All three codon positions shared estimates of the evolutionary model parameters, tree topology, and branch lengths. Initial trees for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Posterior probability of each branch bifurcation is indicated above the branch. The analysis involved 56 nucleotide sequences. There were a total of 815 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.

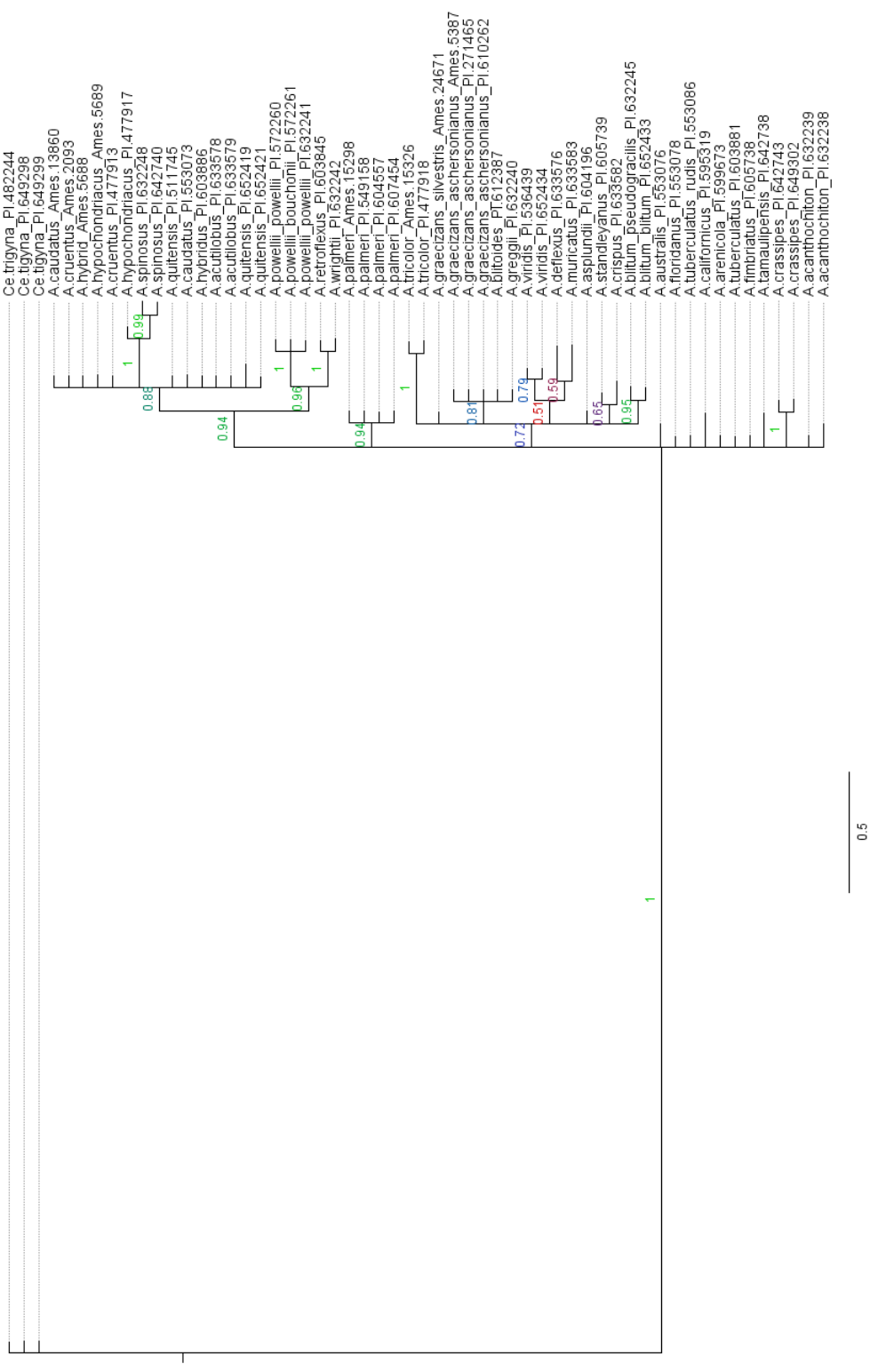
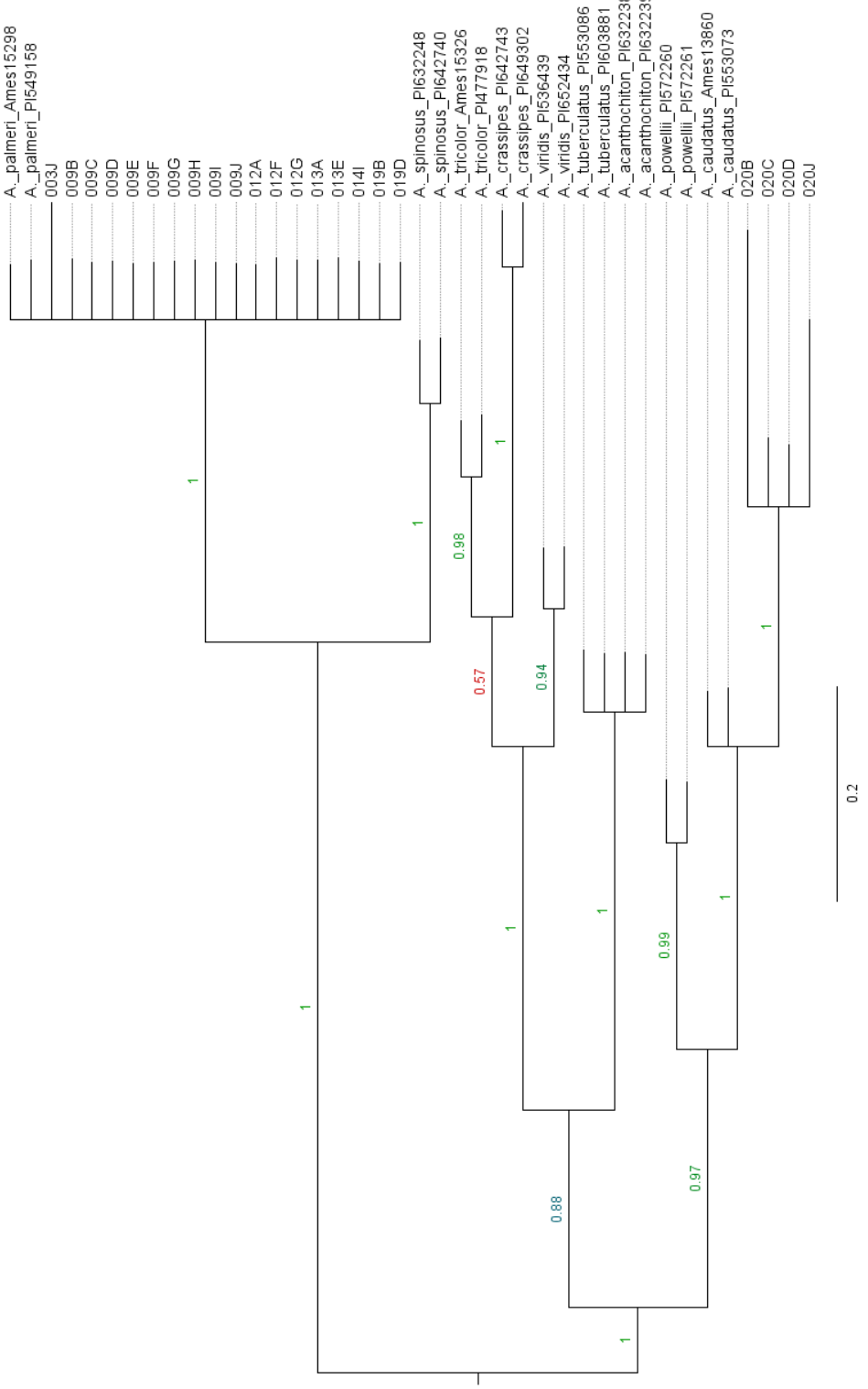

Ce.trigyna_PI.482244
Ce.trigyna_PI.649298
Ce.trigyna_PI.649299
A.caudatus_Ames.13860
A.cruentus_Ames.2093
A.hybrid_Ames.5688
A.hypochondriacus_Ames.5689
A.cruentus_PI.477913
A.hypochondriacus_PI.477917
A.spinosus_PI.632248
A.spinosus_PI.642740
A.quitensis_PI.511745
A.caudatus_PI.553073
A.hybridus_PI.603886
A.acutilobus_PI.633578
A.acutilobus_PI.633579
A.quitensis_PI.652419
A.quitensis_PI.652421
A.powellii_powellii_PI.572260
A.powellii_bouchonii_PI.572261
A.powellii_powellii_PI.632241
A.retroflexus_PI.603845
A.wrightii_PI.632242
A.palmeri_Ames.15298
A.palmeri_PI.549158
A.palmeri_PI.604557
A.palmeri_PI.607454
A.tricolor_Ames.15326
A.tricolor_PI.477918
A.graecizans_silvestris_Ames.24671
A.graecizans_aschersonianus_Ames.5387
A.graecizans_aschersonianus_PI.271465
A.graecizans_aschersonianus_PI.610262
A.blitoides_PI.612387
A.greggii_PI.632240
A.viridis_PI.536439
A.viridis_PI.652434
A.deflexus_PI.633576
A.muricatus_PI.633583
A.asplundii_PI.604196
A.standleyanus_PI.605739
A.crispus_PI.633582
A.blitum_pseudogracilis_PI.632245
A.blitum_blitum_PI.652433
A.australis_PI.553076
A.floridanus_PI.553078
A.tuberculatus_rudis_PI.553086
A.californicus_PI.595319
A.arenicola_PI.599673
A.tuberculatus_PI.603881
A.fimbriatus_PI.605738
A.tamaulipensis_PI.642738
A.crassipes_PI.642743
A.crassipes_PI.649302
A.acanthochiton_PI.632239
A.acanthochiton_PI.632238
0.5

Figure B.7: Placement of *Amaranthus* individuals from location 20 in gene tree of the *internal transcriped spacer* (*ITS*) *1* and *2*. Based on *ITS* sequence, individuals from location 20 are not *A. palmeri*; they are more closely related to *A. caudatus*. The gene tree of *ITS* was inferred using the Bayesian estimation method based on the Kimura 2 parameter model with a discrete Gamma distribution and a proportion of invariant sites used to model evolutionary rate differences among sites (4 categories (+*G* +I)). Both *ITS* loci shared estimates of the evolutionary model parameters, tree topology, and branch lengths. Initial trees for the heuristic search were obtained automatically. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Posterior probability of each branch bifurcation is indicated above the branch. The analysis involved 40 nucleotide sequences. There were a total of 460 positions in the final dataset. Evolutionary analyses were conducted in MrBayes v3.1.2.


A._palmeri_Ames15298
A._palmeri_PI549158
003J
009B
009C
009D
009E
009F
009G
009H
009I
009J
012A
012F
012G
013A
013E
014I
019B
019D
A._spinosus_PI632248
A._spinosus_PI642740
A._tricolor_Ames15326
A._tricolor_PI477918
A._crassipes_PI642743
A._crassipes_PI649302
A._viridis_PI536439
A._viridis_PI652434
A._tuberculatus_PI553086
A._tuberculatus_PI603881
A._acanthochiton_PI632238
A._acanthochiton_PI632239
A._powellii_PI572260
A._powellii_PI572261
A._caudatus_Ames13860
A._caudatus_PI553073
020B
020C
020D
020J
1
1
1
0.98
1
0.57
0.94
1
1
0.99
1
0.97
0.88
1
0.2