ABSTRACT

Title of Document: MOVING TOWARD AN OPTIMUM: THE ADAPTATION GENETICS OF *ARABIDOPSIS THALIANA*.

Frank Warren Stearns Jr., Doctor of Philosophy, 2015

Directed By: Professor Charles B. Fenster, Department of Biology

Adaptation accounts for many of the interesting characteristics of biodiversity. Despite this, the genetic mechanisms underlying the process of adaptation in nature are largely unknown. While general principles are emerging, important questions remain. Although experimental evidence has corroborated theoretical predictions, very few studies have tested macroorganisms in nature, where adaptation is most relevant.

My dissertation addresses several important questions in adaptation genetics in the context of fitness landscapes, primarily using the model plant *Arabidopsis thaliana*. Fitness landscapes are used to visualize the relationship between genetics and fitness (evolutionary success of an individual). Although fitness landscapes have been considered metaphorical, recent work (and this dissertation) suggests they may approximate reality, providing testable predictions. I first assess pleiotropy (when one

gene has multiple effects), an important component of fitness landscape models. I examine this concept in historical context and suggest future directions for research. Next I evaluate how well genetic relatedness corresponds to climate adaptation across the native range of *A. thaliana* and find support for parallel evolution (identical but independent genetic changes), suggesting that fitness landscapes are complex. In my next chapter, using a combination of natural and artificial conditions, I examine how induced mutations impact traits that are fitness indicators as compared to general traits. I find that new mutations tend to reduce fitness, whereas their effect on general traits is bidirectional. This result is more pronounced under stressful field conditions. Finally, I evaluate a mathematical model of adaptation in the field using induced mutations in *A. thaliana*. I find support for a previous result from laboratory studies - that lineages that are less well adapted to an environment are more likely to benefit from new mutations whereas lineages that are well adapted are more likely to be disrupted by new mutations - and extend that to the wild.

Throughout I explore the importance of contingency in evolution, sometimes underscoring how it leads to unpredictable adaptation (chapters one and two), yet also demonstrating that the actions of mutations can be fit to simplifying assumptions (chapters three and four). These studies therefore significantly contribute to the emerging scholarship on adaptation genetics.

MOVING TOWARD AN OPTIMUM: THE ADAPTATION GENETICS OF *ARABIDOPSIS THALIANA*.

By

Frank Warren Stearns, Jr.

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Charles B. Fenster, Chair
Professor Michele R. Dudash
Professor David J. Hawthorne
Professor Thomas D. Kocher
Professor Stephen M. Mount

Dedication

I dedicate this dissertation to my family - Caroline, Henry and Ruby.

Acknowledgements

First and foremost I want to thank my adviser, Charles Fenster. I could not have completed this project without his guidance and support. I could always count on his enthusiasm for my good ideas and his honest criticism of my bad ones. I would also like to thank the members of my committee, Michele Dudash, David Hawthorne, Thomas Kocher, Stephen Mount and up until the bitter end, Michael Cummings. My committee meetings always felt more like collaborations than inquisitions and my research benefited from discussion with each one of them.

I have been fortunate to have had a series of great mentors along the way toward my degree. Lawrence Hurd and John Knox shepherded me through my undergraduate education and provided an opportunity for me to "get my hands dirty" in original research in the Shenandoah Valley. The lessons I learned from Thomas Wood, Reginald Cocroft and Douglas Tallamy during my Master's Degree left and indelible mark on my current research. And I may never have gotten interested in adaptive landscapes at all if it were not for the time I spent in the labs of Thomas Juenger and Michael Singer at the University of Texas, Austin.

I would be remiss if I did not thank my lab mates, Kevin Barry, Jason Berg, Carolina Diller, Sara Konkel, Abigail Kula, Richard Reynolds, Clark Rushing, Nicola Seitz, Andrew Simpson, Calandra Stanley, Alyssa Stewart, and Juannan Zhou. They not only helped in the field and with statistical analyses, but also with friendly reviews of

manuscripts and presentations. They were invaluable for discussion as I formulated my ideas and interpreted my results.

Of course my family has shown a great deal of patience as they supported me throughout my education; my father, Frank Stearns, and my mother Judith Stearns in particular but also my brothers, Brian and Joseph. They may not have always understood what I was doing or why, but they (usually) pretended to be interested.

Last but not least, I need to thank my lovely wife Caroline and my son Henry – my editor and a partner in "sciencing plants" respectively. They made the bad times good and the good times better.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

The field of adaptation genetics as it currently stands has seen the bulk of its development in the last two decades (Orr 2005a) through both important experimental studies (e.g., Burch an Chao 1999; Rokyta et al. 2005; Barrett et al. 2008; Wiser et al. 2013) and theoretical work, notably by H. Allen Orr (1998, 2000, 2003, 2005b) but also others (Dieckman and Doebeli 1999; Gavrilets 2004; Martin and Lenormand, 2006a, 2006b, 2008). Although the study of adaptation has been important to evolutionary biology (and biology in general) as far back as Darwin (1859), the field of adaptation genetics arguably started just prior to the Modern Synthesis with the work of Sewall Wright (1931, 1932) and R. A. Fisher (1918, 1930). While both of these giants of evolution made great strides resolving Mendelian genetics with Darwinian selection, they are well known to have disagreed (often in press) on a number of fundamental issues. Many of the points they disagreed with persist today and exist as some of the foundations of adaptation genetics research.

The contrasting views of the process of adaptation held by both Fisher and Wright were colored by some of their disagreements in the way they viewed certain aspects of evolution (Provine 1971). The two "worldviews" of evolution were captured in the metaphors that each developed to explain adaptation. Fisher viewed populations as immensely large, such that epistatic interactions could be essentially ignored and natural selection was sufficient to explain evolutionary change. He developed what

has become known as “Fisher’s Geometric Model of Adaptation” (FGM) or sometimes, “The Fisher-Orr Model” due to the significant extensions made by H. Allen Orr (Fisher 1930; Orr 1998). This model posits a field of all possible combinations of phenotypic traits, usually presented in two or three dimensions, but really n-dimensional. For each species in a given habit there is a single optimum combination. The challenge of adaptation that a population faces is moving from their current spot in phenotypic space (their current combination of traits) toward the optimum through the action of random mutations. While populations are often at or near the optimum, dispersal to new habitats or environmental changes can cause them to be displaced.

Wright, on the other hand, felt populations were smaller and epistasis was a significant effect. Because a species is not panmictic, according to Wright, then an allele will not be screened against all possible genetic backgrounds, and so the allele's ability to combine with specific alleles at other loci, a.k.a, epistasis, becomes the object of selection (Fenster et al. 1997) He developed a metaphor that he called, “The Surface of Adaptive Value” more commonly known as “The Fitness Landscape” (Wright 1932). This, like FGM, is n-dimensional but usually presented in two or three dimensions. At least one dimension is fitness, while the rest describe genotype or allele frequency space. Due to the action of epistasis, the Fitness Landscape has many fitness optima with varying fitness values. This results in peaks of high fitness surrounded by valleys of low fitness. For comparison, FGM can be thought of as a fitness landscape with a single fitness peak. The challenge for adaptation of a

population in the context of the Fitness Landscape is moving from one local peak through a low fitness valley (against the action of selection) to reach a higher fitness peak. Therefore adaption on a Wrightian Fitness Landscape involved the action of not just natural selection, but also drift and migration (Wright 1932). Wright proposed his "Shifting Balance Theory" as a solution (Wright 1931), but one that has been met with a great deal of criticism (Coyne et al. 1997; Whitlock and Phillips 2000). Importantly, while only Wright included epistasis in his view of adaptation, pleiotropy figured prominently in the metaphors of both he and Fisher (Stearns 2010).

Despite many decades having passed, students of adaptation genetics still face many of these same questions (Orr 2005a). How important is genetic drift as an evolutionary force (Coyne and Orr 2004)? What is the magnitude and typical number of mutations contributing to adaptation (Via and Hawthorne 1998; Barrett et al. 2006)? How important are the complex genetics of pleiotropy and epistasis, and how do they impact adaptation (Cordell 2002; Stearns 2010)? Are adaptive landscapes typically rough (Wrightian) or smooth (Fisherian), or even more than just metaphors (Provine 1971)? In what way and to what extent do new mutations or standing genetic variation contribute to adaptation (Barrett and Schluter 2008)? Is there a "cost of complexity" (Orr 2000)? How common is parallel adaptation (Orr 2005b)?

Some answers are beginning to emerge. New beneficial mutations contributing to adaptation seem to follow a pattern of diminishing returns in magnitude – intermediate in size in the early steps of an "adaptive walk," and increasingly smaller

as an optimum is approached (Orr 1998; Rokyta et al. 2005; Perfeito et al. 2014). This is corroborated by quantitative trait locus (QTL) studies that regularly detect a few genes of intermediate effect and many more of small effect, consistent that adaptation involves the fixation of alleles of both large and small effect. New mutations and standing genetic variation both seem to be able to contribute to adaptation (Barrett and Schluter 2008) especially as studies began to build support for the idea that beneficial mutations are more common than once believed (Shaw et al. 2000; MacKenzie 2005; Hall and Joseph 2010; Rutter et al. 2010). Furthermore, parallel adaptation (identical changes that arise independently in two or more lineages), once considered uncommon, is finding increasing support in the molecular age (Wichman et al. 1999; Orr 2005a, 2005b; Stearns and Fenster 2013). The high likelihood of parallel adaptation is both derived from adaptive landscape models (Orr 2005b) and supports the existence of relatively stable topology that restricts the paths to an adaptive peak (Wichman et al. 1999). Yet in other cases conflicting results are found. Experimental work in microorganism (phage and bacteria) seems to reveal the signature of both rough (Burch and Chao 1999; Melnyk and Kassen 2011) and smooth (MacLean et al. 2010; Trindade et al. 2010; Perfeito et al. 2014).

My dissertation aims to address several of these major issues through historical analysis, computational methods, and direct experimentation.

The first chapter, “One Hundred Years of Pleiotropy: A Retrospective” was published in *Genetics* (Stearns 2010). It is a historical treatment of the concept of pleiotropy

from the point where the term was coined (1910 by Ludwig Plate) through the Modern Synthesis and into the molecular age. Both Fisher and Wright considered pleiotropy as an important component of their views of adaptation. In fact, they relied on the concept of "Universal Pleiotropy" where every locus impacts all other traits to a greater or lesser extent (Wright 1968). During my investigation I found that a split occurred shortly after the Modern Synthesis, primarily between developmental biologists and ecological evolutionary biologists, causing a disconnect in the way the term is used by different biologists. I conclude that recent molecular work has begun to uncover the source of pleiotropy (for example, isolating pleiotropic effects to a single nucleotide), but that major questions of importance to genetics remain unanswered. In particular, how "universal" is pleiotropy? Is pleiotropy an evolved (and evolvable) trait, or simply an unavoidable byproduct of biochemistry and physiology? This paper has already been extensively cited (n = 78, as of Spring 2015) and I hope that it will generate new avenues of research; or invigorate old ones that are not yet settled.

The second chapter of my dissertation, "Evidence of parallel adaptation to climate across the natural range of *Arabidopsis thaliana*" was published in the journal *Ecology & Evolution* (Stearns and Fenster 2013). Taking advantage of publicly available large datasets of worldwide climate factors and genetic markers from *A. thaliana*, we compare genetic distance among populations to habitat climate distance to detect the level of concordance between the two. Using two different approaches we determined that there was some concordance between climate distance and

genetic distance, but not very much. We interpret this to mean that lineages of the species *A. thaliana* are labile with regard to climate space adaptation; that is, distantly related individuals were able to colonize similar habitat spaces. We suggest this is most likely a signature of parallel adaptation, corroborating the recent molecular data on parallel adaptation. This may also support a rough (Wrightian) landscape at the whole species level, as distantly related lineages could occupy similar climate space indicating that there are multiple genetic "solutions" to a particular climate "problem."

In the third chapter, we investigated the differences in the distribution of the effects of new mutations on fitness under natural (field) and artificial (growth chamber) conditions, as well as the differences between the distribution of mutation effects of new mutations on traits that closely related to fitness and those traits that are less closely related to fitness. We used the chemical ethylmethane sulfonate to induce mutations in the Columbia strain of *Arabidopsis thaliana* to generate 20 mutant lines. We planted the mutant lines alongside the founder lines in the Fall 2013 at the Beltsville Experimental Agricultural Station (Beltsville, MD). We kept additional lines in a growth chamber in the Biology-Psychology Building at the University of Maryland (College Park, MD) during Spring of 2014. We found that new mutations increase variance in a fitness component as well as in traits less closely related to fitness under growth chamber conditions. We were not able to show that new mutations were more likely to decrease the fitness component relative to the founder than they were for other traits under artificial conditions. However, we were able to

show that new mutations decreased fitness components relative to the founder under field conditions. This highlights the importance of studying the effect of new mutations on fitness under natural conditions.

Finally, the fourth chapter tests a fundamental prediction of Fisher's Geometric Model – that the distance from a fitness optimum will affect the average outcome of random mutations on fitness. The expectation from FGM is that lineages further from an optimum will benefit more from new mutations than lineages closer to an optimum, where most mutations will move the lineage away from the highest fitness. We treated 18 different accessions of *Arabidopsis thaliana* from across the native range with the chemical mutagen ethylmethane sulfonate (EMS) and planted the premutation founders alongside with their mutant lines in two different field habitats (Blandy Experimental Farm, Boyce, VA and Beltsville Experimental Agricultural Station, Beltsville, MD) over two seasons, Fall and Spring (Fall 2011, 2012 and Spring 2013), for a total of three studies. In total 8,511 plants were studies under field conditions. *Arabidopsis thaliana* has two distinct life history strategies due to its broad range of climate habits (Nordborg and Bergelson 1999) therefore planting in Fall and Spring provided two different experimental regimes that the ecotypes were expected to respond to differently, based on their native locality (Rutter and Fenster 2007). We used the average fitness for each founder to estimate distance from the optimum (i.e., lower fitness indicates more distance) and the difference between the average fitness of a founder from the average fitness of the mutant lines derived from that founder as an estimate of the average amount of improvement.

We found a negative linear relationship between our distance estimate and our improvement estimate, supporting the expectation from FGM. This result tells us two things: first that the adaptive landscape seems to be more than just a metaphor. It is able to make qualitative predictions about the outcome of new mutations, even when complicated by environmental noise in the wild. Second, it lends support to FGM over Wright's Adaptive Landscape model. The latter may be idiosyncratic to intraspecific studies over brief ecological time. To our knowledge, this is the first study to show this relationship in a macroorganism under field conditions. This chapter has been submitted to the journal *Evolution*.

Taken as a whole, my dissertation addresses some of the major questions currently challenging adaptation genetics. Most significantly, I extend results that have previously been tested only in microorganisms under laboratory conditions by finding support for the same predictions under natural conditions in a macroorganism. This suggests that the same principles are meaningful in a general way, and can be brought to bear on realistic situations in the wild, where research on adaptation and evolution are most relevant. These results suggest that fitness landscapes are more than just metaphors, and can inform us about the process of adaptation. On a species wide scale, parallel adaptation indicates that the landscape may be Wrightian and rough, whereas in a local population scale lineages seem to face the challenge of a single peaks and the FGM is able to provide an elegant way to make qualitative predictions of the outcome of new mutations on fitness. In this sense, the process of adaptation

may not be as constrained as was once believed, and the goals of adaptation genetics may be well within our grasp.

1 ONE HUNDRED YEARS OF PLEIOTROPY: A RETROSPECTIVE

1.1 Introduction

"Pleiotropy" refers to the phenomenon in which a single locus affects two or more apparently unrelated phenotypic traits, often identified as a single mutation that affects two more wild type traits. The study of pleiotropic genes has typically involved evaluation of segregation patterns or more recently the mapping of mutant phenotypic traits to a single mutant locus. The concept of pleiotropy has played a prominent role in theories of aging (Williams 1957; Zwaan 1999; Moorad and Promislow 2009), facilitation and constraints of the direction of selection (Hawthorne and Via 2001; Reusch and Wood 2007; Latta and Gardner 2009), models of adaptation (Fisher 1930; Orr 2000), speciation (Maynard Smith 1966; Tauber and Tauber 1989) and human diseases (Pyeritz 1989; MacKay and Anholt 2006; Wilkins 2010). Although at times obvious, pleiotropy can sometimes be difficult to demonstrate. It is often challenging to distinguish between close physical linkage of two distinct genes and actual pleiotropy (Flint and Mackay 2009). This can be further complicated in cases where traits are not well defined. A major goal in genetics is to determine when pleiotropy is caused by a single locus with multiple primary products and when a single gene product is incorporated in many different ways (He and Zhang 2006).

Mendel described an early case of pleiotropy in his classic 1866 paper (Mendel 1866). His character number 3 for *Pisum* displays either a brown seed coat, violet flowers and axial spots or a white seed coat, white flowers, and lack of spots. He states that the three characters that are attributed to each strain are always found together and considers them to be correlated and under the control of a single factor. Whether or not the three characters (seed coat color, flower color, and axial spots) are due to a single gene or not are unknown, but the fact that Mendel believed them to be shows that he considered this sort of inheritance, albeit in a rather cursory manner.

The recognition of pleiotropic traits goes back even further than Mendel, as many medical syndromes were known to have multiple distinct symptoms and a simple "familial" component (Eckman 1788; Weil 1981; Pyeritz1989). However, pleiotropy as a term was not formally described and defined until 1910 by the German geneticist Ludwig Plate. Consequently this is the 100th year since pleiotropy was formally introduced into the scientific literature. In this article, I intend to provide a historical perspective on the progression of pleiotropy, as well as establish some of the more important considerations related to its study.

1.2 The Beginning – Ludwig Plate (1910)

The term "pleiotropie" was coined by the geneticist Ludwig Plate in a festschrift to Richard Hertwigs published in 1910. Plate was a prominent German developmental geneticist in the early part of the century. He began his career as a student under Ernst Haeckel, taking over his position as the Director of the Institute of Zoology at

the University of Jena. As soon as he took over, he removed Haeckel from the museum, beginning a very public feud that resulted in legal proceedings. This was just one of many professional conflicts in which he was involved. He was a member of the Nazi party and a misogynist, openly opposing the advancement of several Jewish and female colleagues (Levit and Hossfeld 2006).

Plate's main interest in genetics was as a means to understand evolution. Like many German geneticists of his time, he attempted to resolve Lamarckian ideas with Darwinian natural selection (Levit and Hossfeld 2006). He synthesized what he considered the important components of evolution and genetics into a program he called "Old-Darwinism." The main structure of Old-Darwinism was a combination of Lamarkian evolution, orthogenesis, and natural selection, studied in light of genetic heredity. Although he ascribed primary importance to natural selection, he felt that some adaptations were only explicable by his particular interpretation of Lamarkian evolution. He clung to these ideas throughout his life. These ideas combined with his personal conflicts severely damaged his reputation as a scientist, and it can be argued that the concept of pleiotropy is his major legacy.

To support his concept of Old-Darwinism, Plate studied the genetics of a variety of organisms. During the course of his own studies and through the results of others, Plate noticed that some distinct phenotypes were only explicable by the mechanism of a single gene. His original definition of pleiotropy is as follows: "I will call a unit of inheritance pleiotropic if several characteristics are dependent on it; these

characteristics will then always appear together and may thus be correlated." (Plate 1910 quoted from McKusick 1976) This original definition is still used today to describe the basic mode of action of pleiotropy. The same mechanism was described under the name "polyphean" in 1925 by Haecker, but by then pleiotropy had received enough attention to be established in the literature (Caspari 1952).

Plate further commented on the ubiquity of pleiotropy, stating that, "The more research into Mendelian factors advances, the more examples become known which can be explained only under the assumption of pleiotropy." His assertion of the extent and importance of pleiotropy has been a central theme that has been challenged and strengthened throughout the last hundred years as the way that we study pleiotropy has changed.

1.3 Development of Pleiotropic Research

One of the first experimental studies of the mechanism of pleiotropy (Gruneberg 1938) came during the Modern Synthesis. Gruneberg was a young German biologist who captured the attention of J.B.S. Haldane. In 1933, he was invited to come to University College London by Haldane on recommendation of Hermann Muller and Richard Goldschmidt (Lewis and Hunt 1984). Haldane immediately suggested he begin studying rat developmental genetics. Gruneberg published a paper on this topic in 1938. His major contribution was to divide pleiotropy into "genuine" and "spurious" pleiotropy. He asserted that genuine pleiotropy was characterized by two distinct primary products arising from a single locus; whereas spurious pleiotropy

involved a single primary product that was utilized in different ways. He also considered a second form of spurious pleiotropy, when one primary product initiated a cascade of events with different consequences on the phenotype. He approached this distinction through the study of a particular genetic skeletal abnormality in rats. The pathology was the result of a new mutation discovered in laboratory colonies of Marthe Vogt and had multitudinous effects on skeletal development. Without molecular evidence, Gruneberg relied on breeding experiments and physiological investigations. By careful study of the anatomy of afflicted rats, Gruneberg was able to create a chart depicting the relationships of the various aspects of the phenotype. He concluded that while both types of spurious pleiotropy were represented, this did not constitute an illustration of genuine pleiotropy. This idea was further established by support for the "one gene one enzyme" hypothesis by Beadle and Tatum (1941, 1945) published only a few years later. "Spurious" was therefore a bad choice of terms, as the majority of investigations into pleiotropy that followed focused on different mechanisms whereby a single gene product could be used in multiple ways. The term "spurious pleiotropy" subsequently fell into disuse. Although "genuine pleiotropy" continued to appear, it was only to suggest that it was unlikely. Despite Gruneberg's feeling that mechanisms involving a single gene product were not true pleiotropy, he was to spend the rest of his career studying these genetic correlations in rats (Pyeritz 1989).

In 1941, Beadle and Tatum published a paper providing support for the "one gene one enzyme" hypothesis, an idea originally introduced (but not pursued) by Cuenot

(1903). The essence of this hypothesis was that a single gene codes for a single protein. The developmental and physiological action of this single protein may be complex as it is incorporated into metabolic pathways, but the genetics were not. Their study on *Neurospora* fungus was fundamental to understanding how genes influenced phenotypic traits and proved to be widely influential to physiological genetics. However, it provided a limited view of gene action that was later expanded by molecular biology. This hypothesis left no room for mechanisms of genuine pleiotropy. Subsequently, emphasis shifted away from the distinction between genuine and spurious pleiotropy and focused more on different mechanisms by which a single gene product could produce multiple phenotypic traits. More about the history of this line of research can be found in Horowitz (1995) and Hickman and Cairns (2003).

A surge of interest in defining the developmental mechanisms of pleiotropy occurred in the mid 1950s. Although this was shortly after the discovery of the structure of DNA, molecular techniques did not advance enough to shed light on pleiotropic action until the early 1980s. In particular two German geneticists played a prominent role in the renewed interest in pleiotropic mechanisms. Richard Goldschmidt (1955) and Ernst Hadorn (1945, 1961) more or less simultaneously used their knowledge of developmental physiology and genetics to elaborate on the various ways by which a single gene product could have multiple uses. Although they addressed the old mechanism of genuine pleiotropy, both authors perpetuated the belief that it was nonexistent and that a single gene was only capable of producing a single primary

gene product. Hadorn had a particularly useful distinction between two types of pleiotropy that he referred to as "mosaic" and "relational." Mosaic pleiotropy describes instances where a single locus directly affects two phenotypic traits. Relational pleiotropy is where the action of a single locus indirectly initiates a cascade of events that impact multiple independent traits. Although these terms are no longer in use, this distinction remains an important one in the study of pleiotropy (Wilkins 1993).

At the same time that the physiological geneticists were struggling with the mechanisms of pleiotropy, population geneticists and ecological geneticists were running productive research programs around largely ignoring the details of pleiotropic gene action. As Sewall Wright stated in the first volume of his four volume treatise on evolutionary genetics, "Pleiotropy has a broader meaning in population genetics than in physiological genetics."(Wright 1968). Although population geneticists acknowledged the physiological genetic assertion that genes produced only a single primary product (one gene one enzyme), they felt that the important factor was how traits were correlated and what the effects of recombination would be on uncoupling phenotypic traits. This viewpoint led to a broader view of pleiotropy in ecology and evolution. This view was so broad that Wright and others asserted that there was "universal pleiotropy." That is, a mutation at any locus had the potential to affect almost all traits through direct and indirect influence. Universal pleiotropy was central to Ernst Mayr's emphasis on coadapted gene complexes (1963) and implicit in Fisher's geometric model of adaptation (Fisher 1930; Orr

2000). A contrasting view that has emerged more recently is the idea that organisms can be broken up into modules and that pleiotropy is restricted to action within these modules (Welch and Waxman 2003). Although pleiotropy is prevalent under the modular hypothesis, it is more restricted than universal pleiotropy would suggest.

The continued study of pleiotropy in ecology and evolution proved very fruitful and led to some major research programs. In particular, G.C. Williams's hypothesis for senescence through antagonistic pleiotropy has proved to be one of the most well know applications of pleiotropy in evolution and medicine. Following a suggestion by Medawar (1952), Williams suggested that genes with antagonistic effects at different life stages could contribute to aging in a way that natural selection could not alter (Williams 1957). That is, genes with beneficial effects prior to reproduction but negative effects after reproduction would be favored by natural selection over those that increased longevity but were less favorable to reproduction and survival to reproductive age. Although Medawar suggested this effect could occur if the genes were pleiotropic or closely linked, Williams emphasized that close linkage would not be sufficient. If natural selection could separate the effects before and after reproduction then effects beneficial early in life and longevity could be maintained. However, if the genes were truly pleiotropic then longevity would never be favored and senescence would be inevitable. This hypothesis has given rise to numerous research programs on aging from a human health perspective as well as senescence as a component of evolution and ecological biology.

1.4 The Molecular Age

It was not until the advent of sequencing in the late 1970s that molecular techniques became refined enough to shed light on genuine pleiotropic mechanisms. It was quickly discovered that a single locus could produce different primary products. These different primary products were found to occur at all levels of gene expression and protein processing. Good reviews of molecular mechanisms of pleiotropy can be found in Pyeritz (1989) and Hodgkins (1998).

Shortly after the first sequencing, it was found that a locus could have multiple or overlapping reading frames (Barrell et al. 1976, Sanger et al. 1977). That is, a strand could be read starting at different points such that a single locus could produce different mRNAs and different proteins. This finding has proved to be fairly common in bacterial genomes. Although the alternate reading frames are sometimes referred to as different genes, the fact that the information for two primary products is contained in one locus and the two products cannot be separated through recombination arguably fits the criteria for pleiotropy.

There are two ways in which alternate transcripts can be produced from a single locus, alternative splicing and alternate start/stop codons. They were discovered a brief time after multiple reading frames and provided a mechanism for pleiotropy at the mRNA processing level. Alternate start/stop codons exist within a locus and transcription of these can result in truncated forms of proteins with altered function. Alternative splicing allows for different exons to be selected from a single locus

(Weber et al. 1977, Donoto et al. 1981). It is known that mRNA strands must go through a processing stage before they can produce a protein. Introns must be spliced out, leaving only the exons (Berget et al. 1977). The whole strand is then given a cap and a tail (Furuichi et al. 1975). However, the splicing stage for mRNA from a single locus could be spliced in different ways to produce different processed mRNA strands. Each of these alternative splicing routes would lead to a different protein. Through RNA processing, a cell can produce multiple proteins from one DNA locus. Alternative splicing plays a role in many aspects of cell maintenance and development and is ubiquitous in higher eukaryotes (Black 2003; Reddy 2007).

The transcribed RNA can be further modified through mRNA editing, first described in 1986 (Benne et al. 1986). Through this process the cell is able to make actual nucleotide substitutions in the mRNA, leading to amino acid differences that can affect protein function. Although these changes can be slight, the affect on the function of the protein can be significant. Even a single substitution can impact amino acid composition, RNA secondary structure, or other forms of transcript processing (Maydanovych and Beal 2006). Editing occurs in different tissues or during differential expression and may play an overlooked role in adaptation (Gommans et al. 2009).

Multifunctional proteins are a final example of molecular mechanisms of pleiotropy. In these cases a single gene product is used for two or more functions or has different functions in different tissues. These mechanisms are reviewed and classified by

Hodgkins (1998). A special class of multifunctional proteins ("moonlighting" proteins) has recently received much attention. The classic example is lens crystallins, which not only serve a structural function in eye lenses but also have enzymatic properties. This example is found under Hodgkins "adoptive pleiotropy." However, in a recent review (Huberts and Van der Klei 2010) the authors state that moonlighting proteins should not be considered pleiotropic as they are defined as multifunctional proteins with independent functions such that a mutation in the coding region for the protein will NOT affect more than one function. Given this as the current definition, I would not include protein moonlighting with other multifunctional proteins as a mechanism of pleiotropy.

1.5 Current Research

More recent work has continued to explore the two major questions of pleiotropy: how extensive is pleiotropy in the genome (universal pleiotropy vs. modular pleiotropy) and how do common mechanisms of pleiotropy work. The genomic age and the accessibility of more advanced molecular techniques have provided insight into these questions from a variety of different angles.

Many of the early architects of the modern synthesis implicitly (Fisher 1930; Mayr 1963) or explicitly (Wright 1968) invoked universal pleiotropy. That is, the assertion that any one gene in a genome has the potential to affect all traits in some way. This assumption was included in many models of evolutionary process. Although not all of these models have been formally tested, those that have provide useful and

biologically relevant results for evolutionary studies. However, experiments in gene manipulation conducted by the early physiological geneticists and more recently by molecular geneticists have suggested something quite different. In their studies disruption of a single locus has limited and measureable phenotypic effects. In order to rationalize the utility of extensive universal pleiotropy with the experimental findings of limited pleiotropy, models have been constructed suggesting that the genome is modular (Welch and Waxman 2003). Genes may have extensive pleiotropic effects on phenotypes within their module but are limited with regard to the organism as a whole. This is a more restrictive view than that of universal pleiotropy. Several recent approaches have been taken to evaluate the extent of pleiotropy as more universal in nature or more modular.

It has been suggested that network theory may be a useful way to study the extent of pleiotropy through computation (Featherstone and Broadie 2002). Early research has suggested that gene expression networks follow small world and scale free dynamics (Featherstone and Broadie 2002). That is, a few genes have extensive pleiotropic effects; whereas most genes are more limited in their effects on the phenotype. But nearly all genes have some degree of pleiotropy. In order to test between the extent of pleiotropy in a genome, Li et al. (2006) analyzed the protein interaction networks of *Saccharomyces cerevisiae, Drosophila melanogaster,* and *Caenorhabditis elegans.* They addressed several aspects of network properties including the "diameter" of the network. This is the mean shortest path length, or how many traits a given gene will affect, on average. They determined that the diameter was about 4-5. In other words,

each gene in the three genomes affects on average four or five proteins. This supports the assertion that pleiotropic effects are more modular than universal.

Another study addressed this same question using comparative techniques (Su et al. 2009). Using 321 genes from eight vertebrate species, the researchers were able to estimate the number of traits affected by each gene in their sample using comparative data from protein sequence and microarray analysis in conjunction with mathematical modeling. They found that the number of traits affected per gene was about six to seven. This closely approximates the results from network analysis and further supports the modular pleiotropy hypothesis.

A more direct study was conducted by Wagner *et al.* (2008). This research used quantitative trait analysis to further expand upon Gruneberg's work on rodent skeletal genetics. The study aimed to identify the magnitude of gene effects as well as the extent of pleiotropy through genotype-phenotype associations in mice. Interestingly, the results closely approximated those from computational and comparative approaches. The number of phenotypic traits per locus was found to be 7.8. This was a small degree higher than previous studies but far short of universal pleiotropy. Therefore, the conclusion from current studies is that pleiotropic effects of genes involve a small number of traits. Although there are no direct experimental results, the strong agreement among these studies is compelling.

A second line of study has been to dissect the action of a single gene. This approach is much like that of Gruneberg with his rats but with the added data from actual gene sequence. In some cases changes in multiple phenotypic traits can be traced to a change in a single nucleotide of a gene. Such mechanistic studies are informative in determining how often a single gene product is used for multiple purposes and how often multiple products arise from a single gene.

A particular mutant strain of yeast is characterized by a change from brown colonies to rust colored colonies when grown in the presence of copper, as well as sensitivity to a range of drug compounds. These two traits segregate together and have been traced to a single amino acid polymorphism in the protein cystathione β-synthase (CYS4) (Kim and Fay 2007). CYS4 plays a role in the pathway converting hemicysteine to cysteine. Disruption of this pathway was biologically likely to affect both drug sensitivity and colony color changes. Although further investigation indicated that the gene network involved may be far more complicated, this is an excellent example of pleiotropy being investigated at the nucleotide level.

Knight *et al.* (2006) looked at a single nucleotide change that allows *Pseudomonas fluorescens* to occupy a novel niche at the air broth interface in laboratory colonies. Previous work has shown that this nucleotide change produces a large number of pleiotropic effects (MacLean et al. 2004) and that it is necessary and sufficient for the habitat shift. The investigators in this study were able to show that the mutation affected the regulation of an entire gene network (involving over 50 protein species)

by “rewiring” it. Some of the genes in the network were upregulated, and some were downregulated. Both synergistic and antagonistic interactions were discovered. Further, changes involved several modules, indicating a more universal pleiotropy. This is one of the most compelling examples of pleiotropy associated with a single nucleotide. A separate study on a gene in the dopamine synthesis pathway (*Catsup*) associated individual traits with separate nucleotides (Carbone et al. 2006). The authors of this study concluded that *Catsup* is pleiotropic at the gene but not nucleotide level. This raises interesting questions as to the unit of pleiotropic action that is relevant.

Whole genome data have also proven useful in studying mechanisms of pleiotropy. He and Zhang (2006) took advantage of the genomic sequence data of *Saccharomyces cerevisiae* to evaluate general patterns of pleiotropic action. They estimated the level of pleiotropy for 4494 genes under 21 different lab conditions. They compared the level of pleiotropy to the number of protein domains per gene, the number of molecular functions, the number of biological processes in which each gene was involved, and the number of protein-protein interactions. High pleiotropy was correlated with a high degree of protein interactions and biological processes but not with the number of molecular functions or the number of proteins per gene. The authors interpreted this to suggest that pleiotropic genes more often produced single multifunctional products.

An additional area that has generated some recent interest concerns the maintenance of pleiotropy. In particular, when pleiotropic action is antagonistic with regard to fitness it would seem that gene duplication and subfunctionalization would allow for an escape from fitness constraints. Waxman and Peck (1998) used mathematical modeling to suggest that pleiotropic traits under stabilizing selection are more likely to reach an optimum genetic sequence. This is in contrast to earlier models that did not allow for pleiotropy. In these earlier models slightly suboptimal sequences tended to predominate. This suggests that pleiotropic traits are more likely to be favored by natural selection. However, two more recent studies have found evidence for subdivision of pleiotropic traits through gene duplication. In one, QTL analysis of two paralogous regulatory genes in maize (*zfl1* and *zfl2*) indicated that both genes were associated with several disjunct traits (Bomblies and Doebley 2006). Although both genes were associated with the same suite of traits, the data further indicated that each gene was more strongly associated with some traits than others and that the traits they were most strongly associated with was different for each paralog. The authors cautioned that this data is indirect and that further studies were necessary, but they also suggested that this may be a case of subfunctionalization that allows escape from pleiotropic effects that are antagonistic under agricultural conditions. More recently, Des Marais and Rausher (2008) used a combination of comparative methods, sequencing, and enzyme assays to examine a pleiotropic gene that had duplicated in some lineages from the Convolvulaceae but not in others. These analyses indicated that duplication in the gene (dihydroflavonol-4-reductase) was more consistent with an adaptive escape from pleiotropic constraints than a case of neofunctionalization.

Taken together these latter two examples suggest that it may be possible for gene duplication to provide and escape from constraints imposed by pleiotropic action, but more work in this area is surely needed.

1.6 Conclusions

The concept of pleiotropy has developed since its introduction into the literature 100 years ago, yet it still has the potential to develop further in the current genomic age. Major questions that were raised during the Modern Synthesis have yet to be settled. How universal is pleiotropy? How often do genes produce multiple products with disparate functions? Both of these questions have significant implications for evolutionary theory.

The ubiquity of pleiotropy as well as the interaction among affected traits impacts the tempo of adaptation to novel environmental input. Extensive pleiotropy, particularly when antagonistic, will often constrain adaptation, whereas synergistic pleiotropy confined to single phenotypic modules may allow populations to rapidly evolve phenotypic novelties that produce new solutions to environmental puzzles. General trends in the extent of pleiotropy and the effects on adaptation are particularly important in light of rapid anthropogenic environmental impacts (Reusch and Wood 2007).

Similarly, the mechanism of pleiotropy may respond to evolutionary dynamics in different ways. A single gene product with multiple effects will be strengthened or

weakened by different processes than those that will impact a single gene that can produce multiple products. Multiple reading frames and alternative transcripts may be more difficult for evolution to disrupt than a single product incorporated into different pathways. Regulatory genes and their far-reaching pleiotropic effects can be considered a special case of pleiotropy that may have extensive consequences (Knight et al. 2006).

The evolutionary outcomes of pleiotropy are only half the story. What is the evolutionary origin of pleiotropic systems? Is pleiotropy an evolved trait, or is it simply a byproduct of biochemical and genetic constraints? Answers to these questions will increase our understanding of how organisms can adapt and what generates the wide range of biodiversity we observe around us. As well, insight into the origin of genetic diseases and disorders will in some cases facilitate their treatment (Cheverud 1996).

In the history of physiological genetics, pleiotropy has often been overlooked and even discounted as an artifact of incomplete understanding of developmental processes. However, evolution and ecology studies of pleiotropy have provided rich interpretations of the evolutionary process. The molecular age has produced evidence that single genes are able to produce multiple products with pervasive effects on the phenotype. Even after one hundred years, studies of pleiotropy have a great deal to tell us, both in ecology and evolution as well as physiology and medicine.

2 EVIDENCE FOR PARALLEL ADAPTATION TO CLIMATE ACROSS THE NATURAL RANGE OF *ARABIDOPSIS THALIANA*

2.1 Introduction

Climate is one of the most important factors determining the distribution of plants (Walther 2003) and therefore adaptation to climate should be a major selective force. Furthermore the ability to adapt to climate heterogeneity can facilitate or constrain the dispersal of organisms, affecting species range (Angert et al. 2011) and climate adaptation may even play an important role in speciation (Keller and Seehausen 2012). Although historically local climates have been known to fluctuate across space and time at an ecological scale, human impacts are accelerating climate change and this has already affected the survival and distribution of some organisms (Parmesan and Yohe 2003; Parmesan 2006). The effects of climate change are expected to increase in the future (Hancock et al. 2011). Thus the ability to adapt to different climate regimes will likely be an important factor in the persistence of populations and species. This is especially true of plants, which are sessile and less able to disperse to more favorable climates as climate change occurs.

Of particular interest is how labile populations are with respect to climate adaptation. That is, how easily are they able to expand their range into novel climate space, and how readily are they able to respond to climate shifts in their own range? The ability to predict the evolutionary dynamics that will result from widespread climate change

will inform both conservation efforts and basic evolutionary theory (Bradshaw and Holzapfel 2001; Olsen et al. 2004; Teplitsky et al. 2008; Kearney et al. 2009; Hoffmann and Sgro 2011; Hansen et al. 2012).

Studies of the effect of climate on species ranges have a long history in plant ecology and evolution (see e.g., Darwin 1859, chapter 11). Furthermore, there is extensive evidence for ecotypic variation within species that contributes to climate adaptation (e.g., Clausen 1926; Clausen et al. 1947; Lowry and Willis 2010). Although plasticity does play a role (Nicotra et al. 2010) the overall picture is that there is a significant genetic contribution to climate adaptation.

Large, publicly available datasets provide a wealth of information for genetic studies. Climate data are also widely available. Given that climate is a significant selective pressure, when populations have resided in a locality for a considerable time (number of generations), it is reasonable to assume they have adapted to the local conditions. Therefore, combining such large-scale datasets allows researchers to estimate adaptation to climate on a greater scale than would be possible using experimental methods (Banta et al. 2012).

The mouse-eared cress *Arabidopsis thaliana* is an ideal candidate for such a study. *Arabidopsis thaliana* exhibits an annual life history strategy with a cosmopolitan distribution across a wide range of habitat types. As a model organism for genetic studies, *A. thaliana* strains from many different climate regimes have been

extensively genotyped (Shindo et al. 2007). Climate is known to be an important feature affecting fitness of *A. thaliana* (Wilczek et al. 2009; Fournier-Level et al. 2011). Climate regimes have been experimentally shown to predict performance under common garden conditions (Hoffmann et al. 2005; Rutter and Fenster 2007). Finally, climate has been shown to be an important factor limiting the distribution of *A. thaliana* (Hoffman 2002). Although only a few loci contributing to climate adaptation have been well studied, the emerging picture is that climate adaptation in *A. thaliana* is affected by a vast network of genes affecting traits such as tolerance to temperature (Westerman 1971) and drought (Mckay et al. 2003). Loci related to climate adaptation have been found to be widespread throughout the genome by a genome scan (Hancock et al. 2011) and a recent study found a correlation between climate and particular nonsynoymous substitutions at the genomic level (Lasky et al. 2012). Despite this, few studies have empirically examined adaptation to climate in natural *A. thaliana* populations due in large part to the difficulty in conducting field studies across a large sample of populations (but see Agren and Schemske 2012). When environmental factors can be correlated to fitness, relying on publicly available environmental and genetic data allows for more comprehensive studies.

Here we quantify whether the genotype of an ecotype is a useful predictor of the climate habitat it occupies. Based on earlier studies (Wilczek et al. 2009; Lasky et al. 2012) we expect the relationship between genetic distance and climate distance to be positive. However, how strong shared evolutionary lineage determines the ability to invade climate space is key to our understanding the lability of populations to adapt to

climate. If there is a weak relationship between genetic relatedness and occupied climate space then it would suggest that there are multiple ways that a lineage can adapt to a particular climate regime, indicating high lability in the ability of this organism to adapt to climate. This question is highly relevant given the current state of drastic anthropogenic climate change. If the relationship between climate space and genotype space is limited, then it bodes ill for organisms like plants that may be restricted in their ability to escape unsuitable habitat.

2.2 Methods

We used a large genetic dataset from *Arabidopsis thaliana* and a worldwide climate database to examine the relationship between genetic relatedness and occupied climate space. To compile data on a substantial number of ecotypes and to generate a genetic distance matrix, we took advantage of publically available data from a large-scale genotyping study (Borevitz lab: http://www.naturalvariation.org/hapmap). The *Arabidopsis thaliana* accessions that we used were taken from 853 lines characterized at 149 single nucleotide polymorphisms (SNPs). It was important to have evidence that the accessions had experienced the local climate for long enough to adapt to their collection climate locality. Thus we attempted to only use accessions that were collected from less anthropogenically disturbed habitats (i.e., not roadsides) typical of *A. thaliana*'s natural habitat where they were more likely to have a relatively long history, and consequently enough time to adapt to local climatic conditions. Such habitats include steep rocky slopes, open areas near forest (but not in understory), and open habitats with sandy or limited soil. We included these habitats as well as

habitats that reflect some human disturbance including fallow fields, rocky walls, cemeteries with sandy soil etc. In response to reviewers suggestions we added a further 67 accessions that from the habitat descriptions appeared to be from more anthropogenically disturbed sites including roadsides, tourist parks, fields under active cultivation, railway ballasts etc. (Table 1). However, the vast majority of lines had no habitat data and were omitted immediately. Of the rest, several were collected outside of the native range of *A. thaliana* (e.g. in North America) and several more were not genotyped at the majority of the SNPs. In addition, we excluded such habitats as "Botanic Garden" or any university associated sites, as those plants may represent escaped accessions adapted to other localities. This resulted in a dataset consisting of 60 accessions (Table 2) that derived from what we consider the least anthropogenically disturbed habitats and 67 more accessions from sites that might reflect higher anthropogenic disturbance (Table 1).

To generate a climate distance matrix among the 60 and 67 accessions, we compiled climate data for each locality from a database consisting of nine different climate factors recorded every 10 degree minutes worldwide. The closest recorded point to the collection site of each *A. thaliana* line was used for this study. In some cases this created overlap in the site data for certain accessions. While this dataset does not capture what may be important microclimate variation, it was the most precise data available to us. Given that previous studies have demonstrated a genetic contribution to climate adaptation, we believe that this will provide a conservative measure of the lability of genetic adaptation to climate, as genotypes from populations in the same

climate regime would be expected to increase the correlation between genotype and habitat climate. Data for eight of the climate factors was collected monthly. These were precipitation (pre), number of wet days (wet), mean temperature (tmp), mean diurnal temperature range (dtr), relative humidity (reh), sunshine (sunp), ground frost (frs), and 10m wind speed (wnd). The ninth was elevation and consisted of a single measure for each location. We included all available climate factors to avoid any a priori assumptions about which factors were most important. We ran additional analyses on a selected subset of the data (mean temperature from November to June and precipitation from June to August). These factors were selected using Banta et al. (2012) as a guide. This reduced the partial Mantel correlation slightly and to a non-significant degree. We therefore included all climate factors in the final analysis. The data are from New et al. (2002) and can be downloaded from Climate Research Unit website (http://www.cru.uea.ac.uk/cru/data/tmc.htm).

To compare climate distance to genetic distance we calculated distance matrices for both genotype (SNPs) and climate. The genetic distance matrix was calculated using DNADIST from the PHYLIP package (Felsenstein 1989) using the F84 substitution model. The climate distance matrix was calculated using PROC DISTANCE METHOD=DGOWERS in SAS 9.1.3 (SAS Institute 2004). This is Gower's environmental distance metric (Gower 1971).

To compare the genetic distance matrix to the climate distance matrix using tree based methods, we estimated neighbor joining trees for each distance matrix using the

program NEIGHBOR in the PHYLIP package (Felsenstein 1989). We then calculated the Robinson-Foulds tree distance metric using TREEDIST from the PHYLIP package (Felsenstein 1989). This metric measures the dissimilarity among the overall topology of two or more unrooted trees (Robinson and Foulds 1981). Smaller numbers indicate higher similarity among topologies. The scale of the metric ranges from ranges from 0 (total concordance) to 2n-6, where n is the number of terminal nodes. In the situation where we used the 60 accessions, then the maximum Robinson-Foulds index would be 2(60)-6 = 114. We then used the program topd/fMtS (Puigbo et al. 2007) to calculate a Robinson-Foulds metric among a set of randomized trees. This number is expected to reflect low concordance due to random topologies.

We expected that geographic distance could inflate the relationship between genetic and climate distance because closely related genotypes are expected to share geographic locales and hence similar climates (Beck et al. 2008). Thus to remove the confounding influence of geographic proximity, we calculated an additional matrix of geographic great circle distance using package SP in R (R Development Core Team 2011). We used the VEGAN (Oksanen et al. 2011) package in R to calculate the partial Mantel correlation (Mantel 1967) between the genetic distance matrix and the climate distance matrix controlling for the geographic distance matrix for both the 60 least disturbed and 67 moderately disturbed accessions separately and together. Mantel and partial Mantel tests are commonly used in ecology to study the relationship between ecological factors and genetic distance (Smouse et al. 1986).

VEGAN calculates three correlation measures: Pearson's product moment correlation, Spearman's rank correlation and Kendall's rank correlation. All R scripts can be found in Appendix A.

2.3 Results

The genetic distance matrix (SNP) and the climate distance matrix for the 60 accessions collected from less disturbed sites are both represented as neighbor joining trees (Figure 1). The Robinson-Foulds distance metric for the two neighbor joining trees was 110, indicating very low concordance between trees. The least possible concordance is 2n-6, 114 in this case. The calculated Robinson-Foulds for a set of 100 randomized topologies for this dataset is 113 with a 95% confidence interval of +/- 0.2. Therefore, although the concordance between these two trees is very low, there is a small signal of lineage on occupied climate space.

The results of the partial Mantel tests are presented in table 3 for the 60 accessions collected from less disturbed sites that in our opinion more likely reflect native habitat. The partial Mantel correlations comparing the genetic distance matrix to the climate distance matrix and controlling for geographic distance were positive but low. Including the 67 moderately disturbed localities with the less disturbed (a total of 127 lineages) did not affect the ranked correlations from the partial Mantel test but reduced the Pearson correlation by about two thirds (from r= 0.23 to r= 0.07). When a partial Mantel test was conducted with the 67 lineages from the moderately disturbed localities alone the Pearson correlation between genetic distance and climate distance

was not significantly different from 0 (r=0.02, p=0.285) Therefore, we decided to base our conclusions only on the 60 accessions collected from less disturbed localities.

As an additional visualization we include a scatterplot of pairwise climate distance by pairwise genetic distance for the 60 accessions that shows a positive but low correlation, with the majority of the points reflecting high genetic distance coupled with low climate distance (Figure 2). We acknowledge that pseudoreplication is a concern with this presentation and we do not base any of our formal analyses on this figure. It is included solely for illustrative purposes.

2.4 Discussion

We demonstrate positive but low concordance between genetic relatedness in *Arabidopsis thaliana* populations and the climate space that those populations inhabit. Both the partial Mantel tests and the Robinson-Foulds index indicate that genetic relatedness has little explanatory power in predicting the climate in which a genotype will be found. We interpret our results to mean that similar *A. thaliana* genotypes are able to occupy different climate regimes and that different genotypes have access or the ability to evolve to similar climate regimes. Therefore, access to different climate spaces appears to be relatively unconstrained by the *A. thaliana* genotype. This is also seen in Figure 2 where a number of genotypes have zero genetic distance based on the survey of 149 SNP's and yet occupy a wide range of climate regimes.

Hoffmann (2005) examined the evolution of climate adaptation in the genus *Arabidopsis* using phylogenetic reconstruction with climate space as a character. The analyses determined the core climate space (the climate space where all studied taxa coexist) and the realized climate niche (the intersection of taxa distribution ranges and climate data) of the genus. Hoffmann concluded that there was a high degree of parallel evolution to climate across the genus. Here we demonstrate this same phenomenon within a species.

The ability of different genotypes to access similar climate habitats, and vice versa may help explain how *A. thaliana* has been able to achieve a cosmopolitan distribution in such a short period (e.g., across North America in approximately the last 150 to 200 years) (Vander Zwan et al. 2000; Jorgensen and Mauricio 2004). Given its annual life history it is possible that populations have adapted to climate on the order of 10's to 100's of generations. Recent range expansion in *A. thaliana* is certainly a result of human interference, but it is believed that the dispersal of self-fertilizing seed colonists has been the most important force in the history of the species. The ability of these annual colonists to rapidly adapt to novel climate habitats likely facilitated this process (Samis et al. 2012). Genetic adaptation to climate may exhibit the pattern we found due to parallel evolution or convergent evolution. In the former, the same genetic changes occur independently. In the latter, different genetic changes occur but the end result is the same. Several greenhouse and field studies have found a high beneficial mutation rate in *A. thaliana*, with as many as 50% of new mutations conferring a fitness advantage (Shaw et al. 2000; Mackenzie et al.

2005; Rutter et al. 2010, 2012). Thus the independent adaptation to similar climates by genotypes that are not directly related may be due to the contribution of new beneficial mutations.

We foresee two potential concerns for our interpretation of independent adaptation to climate space. The first is the possibility that *A. thaliana* is phenotypically plastic with regard to climate space. Although this likely plays some role, we believe that our study also reflects genetic adaptation to climate. A reciprocal transplant study demonstrated genetic adaptation to climate between two European populations (Agren and Schemske 2012). Likewise, a common garden experiment has shown that the success of an accession of *A. thaliana* in a particular habitat can be accurately predicted by the similarity of that habitat to the native habitat of the ecotype (Rutter and Fenster 2007), consistent with adaptive differentiation to climate. Finally, analysis of the 67 accessions from localities we deemed moderately disturbed showed no correlation between climate distance and genetic distance. We therefore feel our criteria for selecting localities was sufficiently conservative and reflects accessions that are likely to be locally adapted to their climate regime.

The second concern is that our study may not have captured variation in the loci that are involved in climate adaptation. The 149 SNP markers that we used to construct the genetic distance matrix were not intended to be used to identify loci associated with climate adaptation, although it is likely that some were given that linkage disequilibrium estimates vary from 10 – 250 kb (Nordberg et al. 2002, 2005; Kim et

al. 2007). Rather our SNP based genetic distance tree clearly demonstrates that genetic relatedness is not a strong predictor of the climate inhabited by the genotype. We do know that climate adaptation in *A. thaliana* involves the interaction of a large number of loci distributed throughout its genome (Wilczek et al. 2009). In a recent study using 214,051 SNPs and 1003 accessions of *A. thaliana*, 15.7% of the genetic variation was found to be associated with climate (Lasky et al. 2012). Therefore one would not expect all the same loci to be involved in the evolution to similar climates given the large number of loci so far identified to be associated with climate adaptation.

Based on our results, it seems likely that parallel evolution is common not just among species of *Arabidopsis* (Hoffmann 2005) but also within *A. thaliana*. An intriguing generality from adaptation genetics studies is that, at the sequence level, parallel evolution may be more common than once believed (Orr 2005a, 2005b). If this is true, then species may not be as genetically constrained with regard to potential habitats, or adaptation to changing habitats. Our data indicate that *A. thaliana* is unconstrained with regard to climate adaptation within the range of climate in which it is found and this may account for its rapid cosmopolitan range expansion. Similar results were reported by Banta et al. (2012). These authors found that later flowering time restricted the niche breadth (measured by climate variables) of *A. thaliana* accessions as compared to earlier flowering accessions which were relatively unconstrained. Flowering time in their study could be controlled by any one of 12

different loci. That is, adaptation to climate could be affected by at least 12 different genetic pathways.

Our findings have important implications to conservation efforts that are responding to anthropogenic change (Hoffmann and Sgro 2011), suggesting that it may not be easy to predict which populations have the ability to adapt to new climate regimes. Using information from ecological and genetic databases in conjunction with smaller scale field studies may therefore be a useful way to generate results from a much larger number of populations than is feasible using experimental methods, and may help shed light on the genetics of climate adaptation.

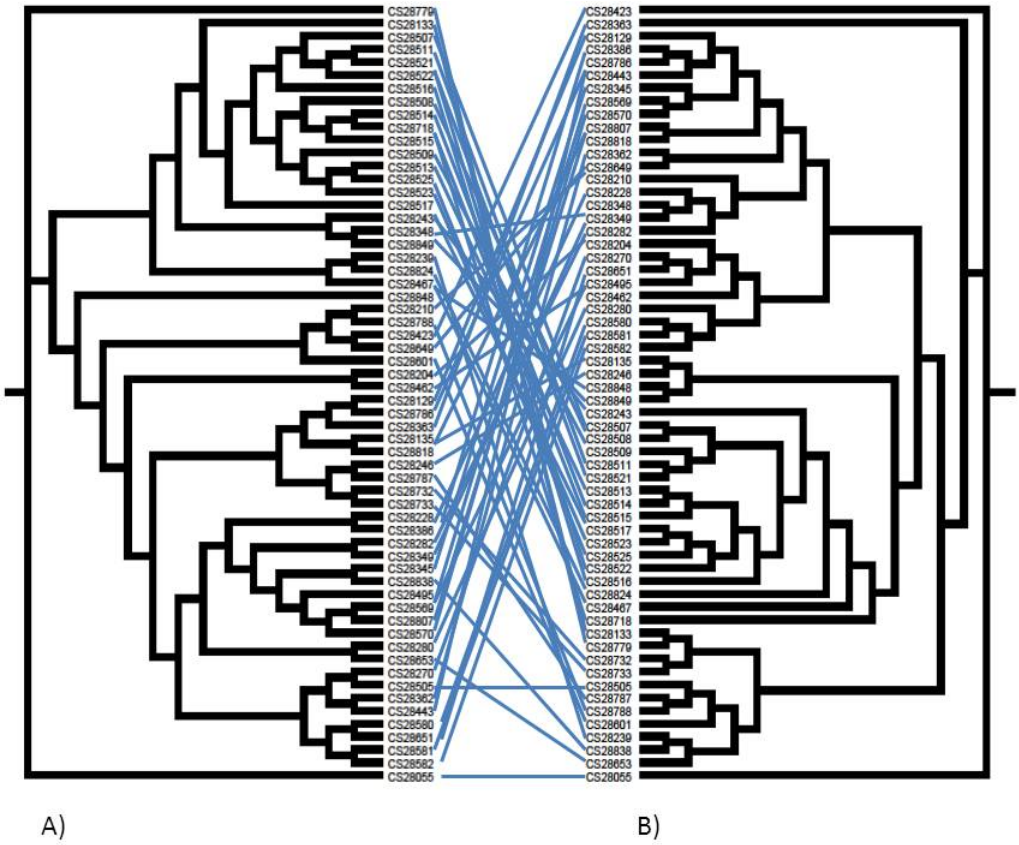

Figure 1. Neighbor joining trees from NEIGHBOR in PHYLIP. (A) Tree reconstructed from a genetic distance matrix for *Arabidopsis thaliana* lines from across Europe and Asia. Pairwise genetic distance was calculated from 149 SNPs for 60 *A. thaliana* lines. (B) Tree reconstructed from climate data matrix for the habitat of each *A. thaliana* line. Pairwise climate Gower's distance was calculated from nine climate factors for 60 collection localities. Lines joining the two trees indicate which genotype (A) inhabits which climate space (B). There was overlap in collection sites for the 60 *A. thaliana* lines.

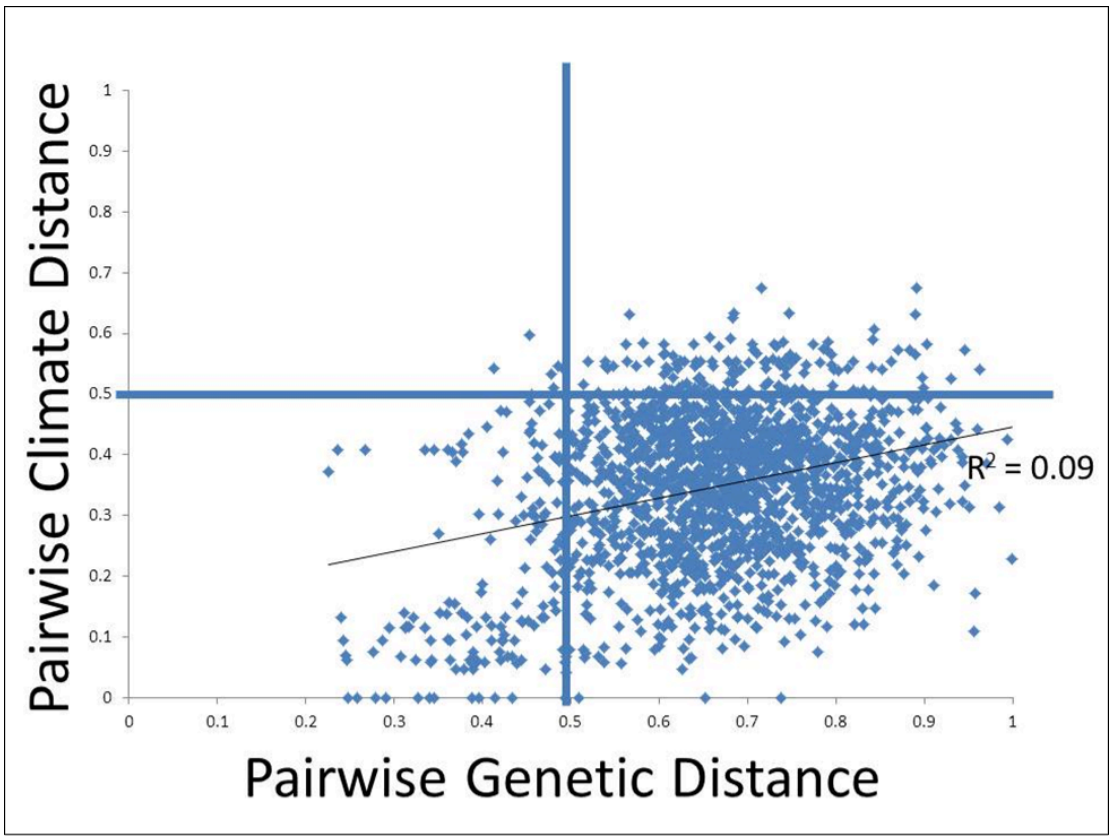

Figure 2. Scatterplot of pairwise climate data (Gower's distance) versus pairwise genetic data for 60 *Arabidopsis thaliana* accessions collected across its native range. As this data suffers from pseudoreplication we did not include it in our formal analyses and only include it for illustrative purposes. The figure is divided into four quadrants representing general relationships between climate distance and genetic distance. They are A) high climate distance and low genetic distance B) high climate distance and high genetic distance C) low climate distance and low genetic distance and D) low climate distance and low genetic distance. The majority of the points reflect a correlation between low climate distance and high genetic distance.

Table 1. The 67 additional *Arabidopsis thaliana* accessions from moderately disturbed habitats. The city and country where they were collected as well as the latitude and longitude and habitat data are listed.

Table 1

Accession	City	Country	Latitude	Longitude	Habitat
CS28007	Aun/Rhon	Germany	50.63544	10.11601	field border
CS28009	Argentat	France	45.09336	1.93755	railway ballast
CS28011	Achkarren/Frieberg	Germany	48.06781	7.62644	vineyard
CS28013	Alston	United Kingdom	54.81217	-2.43868	deserted garden
CS28014	Ameland	Netherlands	53.44056	5.65877	on dunes near firehouse
CS28049	Annecy	France	45.89925	6.12938	garden
CS28050	Appeltern	Netherlands	51.83212	5.58465	parking lot near show gardens
CS28051	Arby	Sweden	59.38194	16.52056	country road
CS28053	Blackmount	United Kingdom	55.97291	-3.70978	roadside 300m
CS28054	Baarlo	Netherlands	51.32836	6.0876	road side
CS28058	Buchen/Lauenberg	Germany	53.48325	10.61413	deposited sand
CS28059	Buchen/Lauenberg	Germany	53.48325	10.61413	deposited sand
CS28060	Buchen/Lauenberg	Germany	53.48325	10.61413	deposited sand
CS28064	Bennekom	Netherlands	51.9991	5.67475	roadside
CS28090	Bulhary	Czechoslovakia	48.83147	16.74874	Distr. Breclav (3 km E), left riverside of the Thaya, grassy place, outside wharf
CS28091	Boot, Eskdale	United Kingdom	54.39966	-3.27095	parking lot pub
CS28100	Buchschlag/ Frankfurt am Main	Germany	50.01475	8.70092	near a rail line
CS28130	Canary Islands	Spain	28.29156	-16.62913	LasPalmas/Mirador
CS28134	Champex	Switzerland	46.02786	7.1165	Alpine garden

CS28216	Durham	United Kingdom	54.77525	-1.58485	near cathedral
CS28217	Ede	Netherlands	52.04361	5.66667	parking lot railway station
CS28234	Enkheim/Frankfurt	Germany	50.14222	8.75269	field border
CS28240	Eringsboda	Sweden	56.43902	15.37709	in tourist park
CS28244	Estland	Russia	59.90436	23.78079	railway slope near Pinsa
CS28269	Frankfurt/Niederrad	Germany	50.08833	8.64361	roadside/river Main
CS28275	Gudow	Germany	53.55637	10.77171	roadside
CS28279	Geleen	Netherlands	50.96912	5.82289	park
CS28283	Goettingen	Germany	51.53835	9.92969	near a highway
CS28344	Heythuysen	Netherlands	51.24748	5.90143	garden
CS28347	Holtesen	Germany	51.56354	9.88978	field of winter rye
CS28366	Jedburgh	United Kingdom	55.47772	-2.55494	country road
CS28441	Lanark	United Kingdom	55.67386	-3.78214	railway ballast
CS28453	Limburg	Germany	50.3986	8.07958	near a lock
CS28454	Limburg	Germany	50.3986	8.07958	near a lock
CS28455	Limburg	Germany	50.3986	8.07958	near a lock
CS28457	Limburg	Germany	50.3986	8.07958	roadside to Dietkirchen
CS28458	Limburg	Germany	50.3986	8.07958	roadside to Dietkirchen
CS28459	Limburg	Germany	50.3986	8.07958	railway embankment
CS28460	Limburg	Germany	50.3986	8.07958	railway embankment
CS28461	Limburg	Germany	50.3986	8.07958	thrown up earth
CS28466	Lindisfarne	United Kingdom	55.68077	-1.80086	road side
CS28473	Le Mans	France	48.00611	0.19956	wheat field
CS28490	Mickles Fell	United Kingdom	54.61572	-2.30183	Rocky ground on limestone
CS28510	Solomennoye	Russia	56.48347	31.6686	road to Botanical Garden and soccer field
CS28518	Zarevichi	Russia	61.5	34	in the village near the chapel

CS28520	Konchezero	Russia	62.04679	34.11207	in the village
CS28524	Petrozavodsk	Russia	61.78886	34.35972	Segejskaya street
CS28563	Niederlauken	Germany	50.34472	8.43278	roadside, loam
CS28572	Nieps/Salzwedel	Germany	52.69655	10.98148	railway embankment, sand
CS28589	Otterburn	United Kingdom	55.23106	-2.17652	parking lot shop
CS28638	Pitztal/Tirol	Austria	47.11667	10.78333	roadside
CS28645	Pontivy	France	48.06615	-2.96706	roadside
CS28667	Ravensglas	United Kingdom	54.35627	-3.40582	parking lot railwaystation
CS28669	Ravenscar	United Kingdom	54.40191	-0.49084	country road near pub
CS28672	Renkum	Netherlands	51.97609	5.73409	garden
CS28685	Rhenen	Netherlands	51.96214	5.57112	roadside
CS28691	Rome	Italy	41.90151	12.46077	cerveteri monument
CS28692	Rouen	France	49.44323	1.09997	roadside
CS28739	Siegen	Germany	50.88385	8.02096	roadside to Hermesbach
CS28758	The Hague	Netherlands	52.0705	4.3007	street near railwaystation
CS28759	Tingsryd	Sweden	56.52475	14.97853	along main road in village
CS28760	Tivoli	Italy	41.95982	12.80223	near ruins
CS28789	Umkirch	Germany	48.03446	7.76357	sewage field
CS28800	Veenendaal	Netherlands	52.02344	5.55025	parking lot railwaystation
CS28801	Veenendaal	Netherlands	52.02344	5.55025	parking lot near house
CS28803	Vindolanda	United Kingdom	54.99224	-2.35527	near Vindolanda restaurant
CS28808	Wageningen-Asserpark	Netherlands	51.96919	5.66539	tree nusery near Asserpark
CS28819	Wilna/Litauen	Russia	54.68716	25.27965	near Towniskaia
CS28821	Wilna/Litauen	Russia	54.68716	25.27965	near Towniskaya

CS28850	Timisoara City	Romania	45.75554	21.2375	Plane; continental and mediterranean climate. Roundhouse area of the main Timisoara railway station. Average annual temperature= 10.9C. Average annual precipitation=630mm.

Table 2. The 60 *Arabidopsis thaliana* accessions used in this study. The city and country where they were collected as well as the latitude and longitude are listed. Habitat data are also listed and were used in an attempt to restrict the study to less disturbed localities where local adaptation to climate has occurred.

Table 2

Accession	City	Country	Latitude	Longitude	Habitat
CS28055	Bayreuth	Germany	49.941598	11.571146	fallow land
CS28129	Calver	United Kingdom	53.269942	-1.642924	rocky limestone slope
CS28133	Champex	Switzerland	46.313743	6.940206	dry loam
CS28135	Chateaudun	France	48.069624	1.329393	country road
CS28204	Dombachtal	Germany	50.091647	8.239811	stony roadside
CS28210	Donsbach	Germany	50.722356	8.237199	sunny, rocky soil
CS28228	Ellershausen	Germany	51.51052	9.682644	limestone, south side
CS28239	Erlangen	Germany	49.599937	11.0063	dry, sandy way
CS28243	Estland	Russia	58.595272	25.013607	sandy hill
CS28246	Etraygues	France	44.644709	2.564473	rocky slope
CS28270	Frankfurt	Germany	50.111512	8.680506	fallow land, house garden
CS28280	Gieben	Germany	50.584007	8.678247	edge of the forest
CS28282	Goettingen	Germany	51.532638	9.92816	sunny slope
CS28345	Hohenlieth	Germany	54.268266	9.332247	field
CS28348	Holtensen	Germany	51.809947	9.800169	field
CS28349	Holtensen	Germany	51.809947	9.800169	field
CS28362	Isenburg	Germany	51.834324	8.397892	field
CS28364	Jena	Germany	50.926999	11.587011	shaded new red sandstone
CS28386	Killean	United Kingdom	56.042425	-4.368315	rocks on mica schist
CS28423	Krottensee	Germany	49.631206	11.572221	rock outcrop
CS28443	Loch Ness	United Kingdom	57.322858	-4.424382	rock ledges, moine schist
CS28462	Limburg	Germany	50.374069	8.122167	fallow land
CS28467	Lipowiec	Poland	53.465086	21.136739	loamy soil/limestone
CS28495	Mainz	Germany	49.995123	8.267426	sandy soil, cemetery
CS28505	Merzhausen	Germany	47.965954	7.828732	roadside, stony loam

CS28507		Russia	61.27	34.56	quarry
CS28508		Russia	61.27	34.56	dry meadow on the rocks
CS28509		Russia	61.37	34.38	rocks near the road
CS28511		Russia	61.83	34.4	stool on the rocks
CS28513		Russia	61.88	34.55	between two rocks
CS28514		Russia	61.97	34.58	dry meadow on the rocks
CS28515		Russia	61.97	34.58	a small sandy hole
CS28516		Russia	61.97	34.2	rock near ranger station
CS28517		Russia	62.02	34.12	near the Lake Konchezero
CS28521		Russia	61.5	34	mountain
CS28522		Russia	62.2	34.27	rocks near town
CS28523		Russia	62.02	34.12	rocks after the village
CS28525		Russia	62.02	34.12	rocks near the road
CS28569	Noordwijk	Netherlands	52.234393	4.448311	dune sand
CS28570	Noordwijk	Netherlands	52.234393	4.448311	dune sand
CS28580	Oberursel	Germany	50.203323	8.576922	sandy stony wall
CS28581	Oberursel	Germany	50.203323	8.576922	sandy stony wall
CS28582	Oberursel	Germany	50.203323	8.576922	sandy loam
CS28601	Pfrondorf	Germany	48.547803	9.110747	sunny field
CS28649	Poppelsdorf	Germany	50.722039	7.088521	sandy ground
CS28651	Praunheim	Germany	50.144782	8.607063	loamy soil
CS28653	Poetrau	Germany	48.649	12.325969	sandy fallow land
CS28718	Rubezhnoe	Ukraine	49.010799	38.381321	near lake
CS28732	St.Georgen	Germany	48.122787	8.333986	fallow land
CS28733	St.Georgen	Germany	48.122787	8.333986	fallow land
CS28779	Tsagguns	Austria	47.077727	9.901948	camping site
CS28786	Taynuilt	United Kingdom	56.428904	-5.239063	rock ledges on basalt
CS28787	Umkirch	Germany	48.031687	7.761354	embankment/river Dreisam
CS28788	Umkirch	Germany	48.031687	7.761354	embankment/river Dreisam

CS28807	Wageningen	Netherlands	51.964641	5.662361	papenpad in woods
CS28818	Weerseloo	Netherlands	51.358578	5.308936	path in woods
CS28824	Wassilewskija	Russia	54.37773	19.433775	sandy rye field
CS28838	Wu	Germany	49.632257	9.945612	sandy soil
CS28848	Orsova	Rumania	44.714211	22.408039	Hill very close to Danube river
CS28849	Orsova	Romania	44.714211	22.408039	Hill very close to Danube river

Table 3. The results of partial Mantel test estimating the correlation between genetic distance from 60 *Arabidopsis thaliana* lines and climate distance from their collection localities across Europe and Asia and controlling for geographic distance. Pairwise genetic distance was calculated from 149 SNPs. Pairwise climate Gower's distance was calculated from nine climate factors for each collection locality. There was overlap in some localities. The results are presented for Pearson correlation, Spearmen rank correlation, and Kendall rank correlation.

Table 3

	Partial Mantel Test	**p**
Pearson Correlation (r)	0.2389	0.001
Spearman Rank Correlation (ρ)	0.07039	0.029
Kendall Rank Correlation (τ)	0.06944	0.003

3 THE EFFECT OF INDUCED MUTATIONS ON QUANTITATIVE TRAITS IN ARABIDOPSIS THALIANA: NATURAL VERSUS ARTIFICIAL CONDITIONS

3.1 Introduction

Mutations are the ultimate source of all genetic variation. Because of this, the study of adaptation genetics historically modeled the process of adaptation as occurring due to novel beneficial mutations (Fisher 1930). However, adaptation can also occur by acting on standing genetic variation from accumulated mutations (Barrett and Schluter 2008; Karasov et al. 2010). The question of great importance is whether adaptation is mutation limited or selection limited. At heart is the belief that the waiting time for new beneficial mutations is too long and their effects too small for them to adequately contribute to adaptation to environmental changes under relatively short ecological time scales. Additionally, new mutations of traits that are closely associated with fitness are typically believed to be deleterious far more often than beneficial (Camara et al. 2000; Keightley and Lynch 2003).

There have traditionally been two major experimental approaches to studying the effects of mutations on fitness (van Harten 1998; Kondrashov and Kondrashov 2010). Mutation accumulation studies (MA) reduce the strength of selection on experimental populations, allowing spontaneous mutations to accumulate by drift (Mukai 1964; Halligan and Keightley 2009). Since selection is reduced, all but the most strongly deleterious mutations will be represented. This approach has the advantage of

investigating spontaneous mutations, preventing biases in the mutations that do not exist under natural conditions. The second approach is chemical or radiation mutagenesis (Auerbach 1949; Singer and Kusmierek 1982; Jambhulkar 2007). Although this approach does bias the spectrum of mutation types dependent on the mutagenizing agent, mutagenesis has the advantage of being much quicker than MA approaches based on spontaneous naturally occurring mutations.

Recent work in *Arabidopsis thaliana* (Shaw et al. 2000; MacKenzie et al. 2005; Rutter et al. 2010) and *Saccharomyces cerevisiae* (Hall and Joseph 2010) has indicated that mutations may be more beneficial more often in these two organisms than in previous studies of the distribution of mutation effects on fitness. In particular, mutant lines in *A. thaliana* were found to increase fitness components relative to a premutated founder nearly half the time under both greenhouse (Shaw et al. 2000; MacKenzie et al. 2005) and field conditions (Rutter et al. 2010). This suggests that new mutations may be able to contribute to adaptation more quickly than previously assumed, however more study is needed.

Arabidopsis thaliana is an ideal organism for the study of adaptation and mutation. It was the first plant genome sequenced and has been the primary model for plant genetics for several decades (Weigel and Glazebrook 2002). The genome size is small with only 5 chromosomes and about 125 Mbp (The Arabidopsis Initiative 2000). Mutation protocols are well established for the Columbia accession (Camara and Pigliucci 1999; Salinas and Sanchez-Serrano 2006). The mutation rates of several

important traits are known, and many mutations have been characterized and are commercially available (Arabidopsis Biological Research Center 2015). Isogenic mutant lines can be readily established through selfing, the most common reproductive method in *A. thaliana* (Agren et al. 2013). Hundreds of seed can then be produced from a single plant to propagate these mutant lines. Greenhouse and lab maintenance is relatively easy and in many cases the entire life cycle can occur in several months (Weigel and Glazebrook 2002). Additionally, *A. thaliana* has a wide natural distribution over a large range of habitat and climate regimes (Hoffmann 2002, 2005; Hancock et al. 2011; Banta et al. 2012; Stearns and Fenster 2013). A large number of these accessions have genotypic data in the form of markers and gene sequence as well as physiological and morphological data that may be related to their native habitat requirements (Rutter and Fenster 2007; Horton et al. 2012; Lasky et al. 2012; Stearns and Fenster 2013; Lasky et al. 2014; Hamilton et al. 2015).

The effects of new mutations on fitness are expected to differ in the wild than in artificial environments. Under natural conditions many more mutations will affect fitness than in more controlled environments (Rutter et al., 2010). In field experiments the expectation is that more mutations will be involved in fitness, and they will have a greater effect on fitness. Stressful conditions are also known to increase the variance in mutations effects on fitness (Martin and Lenormand 2006) and to result in mutations being more deleterious on average (Kondrashov and Houle 1994 but see Chang and Shaw 2003). For example, inbreeding depression is generally greater in more stressful versus less stressful environments (Dudash 1990; Frankham

2015; but see Agrawal and Whitlock 2011). Furthermore, interactions among the experimental mutations and between the mutations and the environment will be more complex and it is possible that higher environmental variance may mask the effects of individual mutations (Jaenike 1982).

Here we use chemical mutagenesis to generate mutant lines of *A. thaliana* and plant them under field conditions along side the premutation founder. We also planted the same lines under artificial (growth room) conditions. In this way we were able to gauge the magnitude of fitness changes through mutation in the field and the growth room, two environments with contrasting survivorship (growth room >> field), suggesting different degrees of stress. We also measured traits less closely related to fitness in the growth room experiment to compare the distribution of mutation effects for fitness components versus quantitative traits that are not as closely related to fitness in artificial conditions. There is an expectation that new mutations will affect such traits differently, increasing variance in either direction as opposed to being skewed towards a decrease in fitness (Keightley and Lynch 2003).

Here we address three questions: 1) What is the distribution of fitness effects of mutations derived from mutagenesis? 2) Does the distribution of fitness effects differ between field and relatively benign laboratory environment conditions? 3) Does the distribution of mutation effects for fitness proxies differ from the distribution for traits that are less likely to be related to fitness? To date, three labs have used *A. thaliana* lines and have demonstrated a much higher frequency of beneficial

mutations (Shaw et al. 2000; MacKenzie et al. 2005; Rutter et al. 2010) than expected (Keightley and Lynch 2003; Keightley and Eyre-Walker 2010). However, because of the much longer timespan between generations, *A. thaliana* mutation accumulation experiments represent ½ to an order of magnitude fewer generations than mutation accumulation experiments conducted with shorter-lived organisms (Keightley et al. 2009; Halligan and Keightley 2009; Hall and Joseph 2010; Latta et al. 2015; Katju et al. 2015). Thus by using mutagenic approaches we hoped to not only add another field based estimate of mutation effects on fitness, but also to ask what the cumulative fitness effects of mutations might be when many more are generated, corresponding to other mutation accumulation experiments with shorter-lived organisms.

3.2 Methods

3.2.1 Mutagenesis

Mutations were induced in the Columbia accession of *Arabidopsis thaliana* using ethylmethane sulfonate (EMS). EMS is an alkylating agent that most commonly results in G:C to A:T substitutions (transitions) (Greene et al. 2003). Overall, EMS induces a spectrum of point mutations similar to spontaneous mutations, although it does not produce indels (Greene et al. 2003). Mutation rate was evaluated by exposing the lines to 20, 30, 40 and 50 μM solutions and estimating the number of siliques with albino seeds from a sample of each treatment, a common measure of mutation rate in *A. thaliana* (Camara et al. 2000). The 20 μM dosage was chosen in order to generate enough mutations for a measurable effect on fitness.

We estimate 25 mutations per cell in coding regions per genome, which are expected to affect a wide range of quantitative traits (Brock 1976; Camara et al. 2000). This estimate is based on the measured rate of mutations induced by EMS, which is approximately 3.7×10^{-6} locus^{-1} • cell^{-1} • μM^{-1} • hr^{-1} (Korneef et al. 1982; Camara et al. 2000). *Arabidopsis thaliana* is estimated to have about 28,000 loci (Redei and Koncz 1992). This mutation rate is about three to four fold greater than the frequency of nonsynonymous mutations in protein coding sequence as quantified from direct sequencing (Ossowski et al. 2010) of a subset of the Columbia mutation accumulation lines (representing spontaneous mutations) tested by Rutter et al. (2010, 2012). Using a 20 μM dosage was considered ideal because it generated visible mutations, but not so many mutations that seed would not germinate and complete their life-cycle. Furthermore, the dosage resulted in meeting our criteria of generating many more mutations than were quantified in previous *A. thaliana* MA line experiments conducted in the field.

After the 20 μM solution of EMS was determined as the optimal concentration, seeds were washed with a 0.1% solution of Tween-20 and then soaked for 12 hours in EMS with rotation. The seeds were next washed and soaked in distilled water for 6 hours with rotation. After washing, the seeds were sown directly onto soil in individual pots (Figure 3). Seeds from the premutated founder were sown out at the same time. These seeds and seeds sown from the pre-mutated founder were cold treated at 2º C for two weeks with no light. They were then allowed to germinate on benches in a growth

room at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light for two weeks. After this time, individual seedlings were transplanted to their own pots to generate the founder and the 20 mutant lines. This was the M1 generation. These plants were grown in the growth room under the same conditions used for germination (20º C with 24 hours of fluorescent light and 8 hours of incandescent light). Seed was collected from these plants and sown out as before, with 2 weeks at 2º C with no light and then moved to a growth room to germinate at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light for two weeks as before. Individual seedlings were isolated into pots, 20 individuals per mutation line and premutant founder. This was the M2 generation. These plants were again allowed to grow in the growth room under the same conditions used for germination (20º C with 24 hours of fluorescent light and 8 hours of incandescent light). Seed was collected from these plants. This seed was treated as before (two weeks at 2º C with no light followed by two weeks for germination in the growth room with 20º C with 24 hours of fluorescent light and 8 hours of incandescent light). At transplant, each of the 20 replicates from the 20 mutant lines and the founder were split into four sublines by transplanting seedlings into individual pots. This was the M3 generation. These plants were again grown in the growth room at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light. Seed collected from these individuals (two sublines out of the four per replicate) was the M4 generation. Individuals from this generation were used for fitness assessment in the field. Based on our earlier estimate of 25 mutations per cell, we estimate that 11 of the induced mutations will be homozygous and three will be heterozygous for the mutation, with the other 11 mutations lost.

3.2.2 Fitness Assessment – Field Conditions

Pre mutated and mutated seeds from among the founder and mutant lines, respectively, were planted during Fall of 2013 at the Beltsville Experimental Agricultural Station (UMD) in Beltsville MD (N 39.05378 W -76.95387). Weather data records for the Beltsville Experimental Agriculture Station can be found here: http://www.ba.ars.usda.gov/weather/ba-weather-2.html. Prior to planting, seeds were cold treated for two weeks at 2º C and allowed to germinate on benches at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light for two weeks, as above. Seedlings were then transplanted in random order into plug trays and allowed two more weeks to establish; 20º C with 24 hours of fluorescent light on a bench and 8 hours of incandescent light for one week and then at 2º C with 8 hours of incandescent light for another week in order to cold acclimate them. Plug trays were transported to the field and then the seedlings were transplanted into shallow holes with their soil plugs at the field site. Seedlings were spaced 10 cm apart, maintaining the spatial orientation of seedlings in the plug trays. In Fall of 2013, 2 sublines for each of the 20 mutant lines and the founder were planted with 60 plants per subline, for a total of 2520 plants. When planted, the seedlings were at the 2 - 4 leaf stage. The plots were initially watered at planting to facilitate establishment but otherwise were exposed to natural weather conditions, pathogens and predation from herbivores and competition to other plant species. All experimental plants were harvested above ground at the end of May 2014, when plants were in the senescent phase. Harvested plants were dried in heat chambers. Above ground dry mass and total fruit number were measured from a sample for each experiment, and above ground dry mass was

found to be a good predictor of fruit number (r^2 = 0.89, n=15). We therefore used dry mass of survivors multiplied by proportion of plants surviving to harvest as our measure of fitness for each mutant line and pre-mutant founder. This is a reasonable proxy of fitness for a selfing annual plant (Shaw et al. 2000).

3.2.3 Growth Room

Plants from the same lines were grown in a walk-in growth room at the University of Maryland. The same experimental design was sowed into plug trays in Spring 2014. Seeds were cold treated for two weeks at 2º C and allowed to germinate on benches at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light for two weeks, as above. These plants were cultivated in the growth room at 20º C with 24 hours of fluorescent light. Seedlings were then transplanted in random order into plug trays and allowed two more weeks to establish; 20º C with 24 hours of fluorescent light on a bench and 8 hours of incandescent light for one week, as above. 15-25 plants from each subline and the founder were grown in 3 blocks. A power outage exposed the seedlings to very high heat stress, however the plants used in this experiment did not seem to be affected (no death due to the heat stress). Those plants were left to grow in their plug trays and measured for four traits: silique number (=number of fruit), julian day of first flower, number of trichomes per mm of midrib length, and the ratio of side branch mass to main branch mass. They were allowed to grow from April 2014 to December 2014. All plants were harvested when they went to seed unless they had not bolted by December 2014, at which point all remaining plants were harvested. Julian day of first flower was assessed daily. The largest leaf at first flower was collected and used to determine trichome number per midrib length.

When plants went to seed they were collected and dried. Silique number and the mass of the main branch and side branches were measured on the dried samples. We consider silique number to be a fitness proxy.

3.2.4 Statistical Analyses

Field

Statistical analyses were conducted in R (R Core Team 2014) using the package STATS. Fitness (survivorship x plant mass) was examined for significant mutant line effects. Because fitness had a great deal of zeros, the non-parametric Kruskal-Wallis test was used, a conservative test for testing line effects. We used the model Fitness = Line + Error. We used Mann-Whitney U (Whitlock and Schluter 2009) tests to determine if the any of the mutant lines differed significantly from the founder. Again, because of low survivorship, we used a nonparametric test to determine if the founder phenotype or performance differed significantly from the mean of the mutant lines. Thus we used a sign test (Whitlock and Schluter 2009) to determine if the rank of the founder differed from the null expectation of no difference, i.e., the rank order of the founder = the median of the mutant lines.

Growth Room

Statistical analyses were conducted in R (R Core Team 2014) using the package STATS except for the negative binomial GLM which used the package MASS (see Appendix A for R codes). All traits (Julian day of flowering, silique number,

trichome number per midrib length and proportion of side branch/main branch mass were examined for significant mutant line effects. Different analyses were used depending on the character. Flowering time was normally distributed and a one way ANOVA was performed. Silique number was examined with a negative binomial general linearized model (GLM). Trichome number per midrib length was gamma distributed but contained many zeroes, so was also analyzed with a Kruskal-Wallis test. Side branch/main branch mass is a continuous trait with a high number of zeros, so the non-parametric Kruskal-Wallis test was used here as well. All tests used the model Trait = Line + Error. To determine if any of the mutant lines differed from the founder we used a Dunnett's test in R package MULTCOMP (Hothorn et al., 2008) for flowering time and Mann-Whitney U tests for the remainder of the traits due to non-normality. To determine if the founder phenotype or performance differed significantly from the mean of the mutant lines, sign tests were used for all traits to determine if the rank of the founder differed from the null expectation of no difference, i.e., the rank order of the founder = the median of the mutant lines. To determine if the cumulative effects of mutations were correlated among the four measured traits, i.e., pleiotropic, we quantified Pearson product moment correlations among the traits, using each mutant line as a replicate. Since multiple tests were conducted with the same data set derived from the growth room experiments, all results from the growth room were sequentially Bonferroni corrected (Whitlock and Schluter 2009) to provide an appropriate Type I error rate.

Because of overall low survivorship in the field and limited replication for the growth room experiments, we collapsed subline effects to line effects. The generation of the sublines occurred in random locations in the growth room, and so by pooling sublines we control for the effect of specific location of maternal plant on seed quality. By not utilizing sublines in our analyses of the performance of seed from these sublines we likely inflate the environmental variance, making any finding of mutant line effects more robust. In other words, by pooling sublines we control for micro-environmental or maternal effects on seed quality but do not remove these effects from the effect of line and therefore likely increase the error component contributing to performance or trait expression.

3.3 Results

3.3.1 Field

Under field conditions, the Columbia founder line was found to have the second highest fitness (and the highest mass when survivorship was not included in the data) (Figure 4). There was a significant mutation line effect on fitness (X^2= 85.11, df=19, $p<0.001$), indicating that induced mutations contributed to among line variance. None of the individual lines differed significantly from the founder line (all $p> 0.204$). A sign test indicated that the probability that the founder had significantly higher fitness than the median of the mutant lines ($p<0.001$, n=21).

3.3.2 Growth Room

There was a significant mutant line effect for day of first flowering (df=19, $p<0.001$) (Figure 5) siliques (df=19, $p<0.001$) (Figure 6), and trichome number (Kruskal-Wallis $X^2 = 32.09$, df=19, p=0.031) (Figure 7) indicating that among line variance in these traits could be attributed to induced mutations. There was no significant line effect for side branch/main branch mass (Kruskal-Wallis $X^2 = 22.45$, df=19, p=0.26) (Figure 8). After sequential Bonferroni correction there was an effect for day of first flower and for silique number, but trichome number only trended towards significant ($p = 0.06$). Three lines differed from the founder (lines 6, 7 and 19 for trichome number), but after sequential Bonferroni, none of the lines differed significantly from the founder for any of the traits (all $p>0.03$ before correction). A sign test only indicated a significant deviation from the median value of the MA lines for the founder for number of trichomes per midrib length (p=0.04, n=21), the founder had more trichomes than 15 of the 20 MA lines, but this was not significant after sequential Bonferroni correction. The rank of all other traits was not significant. In other words, there is little evidence of the mean character state of the MA lines differing from the founder for the four traits measured in the growth room.

There were no significant correlations between pairs of traits (all values $p>0.29$, n=42) except for flowering with side branch/main branch mass (p=0.018, n=42) and siliques with side branch/main branch mass (p=0.02, n=42). As this latter trait did not have a significant line effect we did not consider those correlations to be informative.

3.4 Discussion

Mutations are expected to have different effects on quantitative traits depending on how much they affect fitness (Camara and Pigliucci 1999; Keightley and Lynch 2003). Mutations are expected to decrease fitness more often, but to have bidirectional or symmetrical effects on traits that are not closely related to fitness (Camara and Pigliucci 1999; Keightley et al. 2000). Using mutant lines of the Columbia strain of *Arabidopsis thaliana* derived from EMS mutagenesis and measuring four traits in a growth room, as well as fitness characters in the field we found that mutations were more likely to decrease fitness in the field, than unrelated characters under artificial growth conditions. Although the mutant lines did not differ significantly from the founder for any of the traits after controlling for experiment wide Type I error, significant mutant line effects suggest that measured differences in these traits were due to induced mutations. These results are consistent with the notion that most mutations are deleterious (Keightley and Lynch 2003) but in conflict with recent results in studies of *A. thaliana* mutant lines (Shaw et al. 2000; MacKenzie et al. 2005; Rutter et al. 2010).

3.4.1 Fitness

Earlier work on *A. thaliana* mutation accumulation lines, as well as on mutation accumulation lines from other species, has demonstrated that mutations can be beneficial more often than previously believed (Shaw et al. 2000; MacKenzie et al. 2005; Rutter et al. 2010; Hall and Joseph 2010). In *A. thaliana*, mutation accumulation lines have been shown to have increased fitness over the premutation

founder as much as 1/3 to 1/2 the time (Shaw et al. 2000; MacKenzie et al. 2005; Rutter et al. 2010). However, here we found that most mutant lines (19/20) had reduced fitness relative to the premutation founder under field conditions. In the more benign conditions of the growth room, the difference between founder and MA lines was much less pronounced. Previous work on an earlier generation of these mutants lines derived from EMS mutagenesis found a similar result: in two out of three field plantings mutations reduced fitness on average relative to the founder (Stearns and Fenster submitted).

The difference between the mutation accumulation studies and our chemical mutagenesis study may be explained in two ways. First, the estimate of the number of mutations induced in coding regions, and hence potentially affecting fitness, was far greater than the number of mutations accumulated in the studies of spontaneous mutations (25 here via mutagenesis vs. 4 per line via spontaneous mutations in Rutter et al 2012 as determined by direct sequence). The increase in the number of mutations potentially affecting fitness increases the probability that a large magnitude deleterious mutation will affect the line (Camara et al. 2000) and lead to a genetic death (Muller 1950; Crow 1997, 2000). That is, if mutations that have a strong deleterious effect on fitness occur with some regularity, increasing the number of mutations increases the chance that a line will get one of these mutations, and it will swamp the effects of any slightly beneficial mutations, dragging the fitness of the line down. This has been corroborated experimentally (Davies et al. 1999) and most recently by Heilbron et al. (2014) who found that 42.3% of the decrease in fitness in

Pseudomonas aeruginosa mutation accumulation lines was explained by only 4.5% of the mutational steps that had a highly deleterious effect on fitness, only 0.5% of all mutations fixed. Previous work with an earlier generation of these mutant lines (and the mutant lines of several other *A. thaliana* ecotypes) suggested that beneficial mutations occurred, but that the magnitude of deleterious mutations was greater (Stearns and Fenster submitted, Chapter 4: Figure 4). This study when considered in the context of the previous *A. thaliana* mutation accumulation studies, suggests that high magnitude deleterious mutations are more common than high magnitude beneficial mutations, and that adaptation likely occurs due to small or intermediate effect beneficial mutations, as suggested by Fisher (1930) and Kimura (1983) and supported experimentally (Barrett et al. 2006; Sousa et al. 2012; Heilbron et al. 2014). The large deleterious mutations that may be affecting these lines would likely be removed from the population via selection and would not contribute significantly to standing genetic variation.

While the spectrum of mutations due to EMS is similar to that from spontaneous mutation (Greene et al. 2003, Ossowski et al. 2010), EMS does not result in indels and is therefore only a subset of natural spontaneous mutations. It is possible that mutations resulting in indels may have a different distribution of effects on fitness than the point mutations induced by EMS. It is difficult to conclude that indels are more likely to be beneficial. Thus we favor the hypothesis that we observed generally greater deleterious cumulative effects of mutations because we generated many more mutations and deleterious mutations are more likely to have large negative effects on

fitness than beneficial mutations having large positive effects on fitness. This may be important to the process of adaptation. If deleterious mutations are more often strongly deleterious, then they are likely to be lost quickly due to selection. The smaller magnitude beneficial mutations, while not as common, are therefore more likely to contribute to standing genetic variation. It is important to note that we attempted to address this issue of variable effects by testing whether lines that were displaced further from the founder in any of the quantitative traits was correlated to lower fitness in these lines under field or greenhouse conditions, but we were unable to detect any relationship (analyses not shown).

In comparing fitness from field assays (mass (=reproduction) x survivorship) to that under growth room conditions (silique number) we see that there is a significant line effect for both. However, the rank of the founder differs in both. Under field conditions all but one line performed worse than the founder (4). Under artificial grow room conditions six out of the 20 mutant lines outperformed the founder (Figure 6). We believe this reflects the fact that there are more perturbations under field conditions and that more of the induced mutations are affecting fitness. Under the relatively benign artificial conditions fewer mutations affect fitness (in fact survivorship was not even a factor) and variation due to mutations in the fitness component (siliques) presented were similar to the quantitative traits that are not as closely tied to fitness. This highlights the importance of investigating the fitness effects of new mutations under more natural conditions, as the difference between that and artificial conditions can be nontrivial. Previous work has suggested that

mutations are more likely to be harmful under more stressful conditions (Kondrashov and Houle 1994). Likewise, the genetic load may be more detrimental under more stressful conditions (Frankham 2015). It is therefore not surprising that we find a significant number of lines with reduced fitness under field conditions and not under artificial conditions.

3.4.2 Other Quantitative Traits

Traits that are not as closely related to fitness are expected to be affected by new mutations in a bidirectional fashion. Three other traits were investigated under artificial conditions (Julian day of first flower, number of trichomes per midrib length, and side branch/main branch mass). These traits are likely less closely tied to fitness with the exception of branching (Lortie and Aarssen 2000), particularly under artificial conditions. Branching for an annual plant will be associated with more reproductive meristems and therefore more flowers and fruit. Significant mutant line effects were found for all the traits except side branch/main branch mass, although the number of trichomes per midrib length was not significant after sequential Bonferroni correction. Although mutations produced a significant line effect in two of the traits that were not as related to fitness (day of first flowering, number of trichomes per midrib length) they were no more likely to decrease the trait value than increase it, as expected. This was also true of the trait most closely related to fitness (silique number), despite predictions about the effects of new mutations on fitness components, and contrary to the results from the field study.

While the traits flowering time and the presence or absence of trichomes have been shown to be under major gene control (Johanson et al. 2000; Shindo et al. 2005 and Marks 1997; Karkkainen and Agren 2002 respectively), our results corroborate that these traits also have a polygenic component (Symonds et al. 2005; Wilczek et al. 2009; Samis et al. 2012). For example, flowering time has evolved across an east-west gradient in North American invasive *A. thaliana* independent of the alleles at the major flowering time loci (Samis et al. 2012). Trichome density has been tied to at least nine QTL (Symonds et al. 2005). Thus, we conclude that despite some traits evolving through substitutions at major loci, mutation effects on polygenic loci may also contribute to standing genetic variation for these traits, and hence, also to a selection response. Furthermore, the lack of correlation in the effects of mutations on the traits suggests that these traits may be able to evolve independently, although we cannot rule out a failure to detect correlations due to low sample size. This lack of correlation in expression of the traits across the mutant lines also supports the notion that the EMS approach led to mutations throughout the genome.

3.4.3 Conclusion

The results of this study are congruent with the mainstream view of the effect of mutations on fitness in that most of the lines decreased fitness relative to a premutation founder, and that effect was more pronounced under field conditions. This is contrasted to previous recent results from *Arabidopsis thaliana*, but strays from those results in an explicable way. Increasing the number of mutations by such a high degree increases the likelihood that a strongly deleterious mutation will occur and counter the effects of slightly beneficial mutations. Fisher used the analogy of

focusing a microscope. If a specimen is reasonably near focus, then large movements with the course adjustment are more likely to reduce focus than small changes with the fine focus (Fisher 1930). It is likely that induced mutation experiments such as this one may be misleading with regard to how adaptation actually occurs, due to the high number of mutations. Still, our study can inform us about the ability of new mutations to contribute to adaptation. This study confirms the widely held belief that new mutations are able to contribute more to quantitative traits that are not close to fitness than they are to traits more directly related to fitness, particularly under field conditions. It would therefore seem that new mutations do contribute to standing genetic variation.

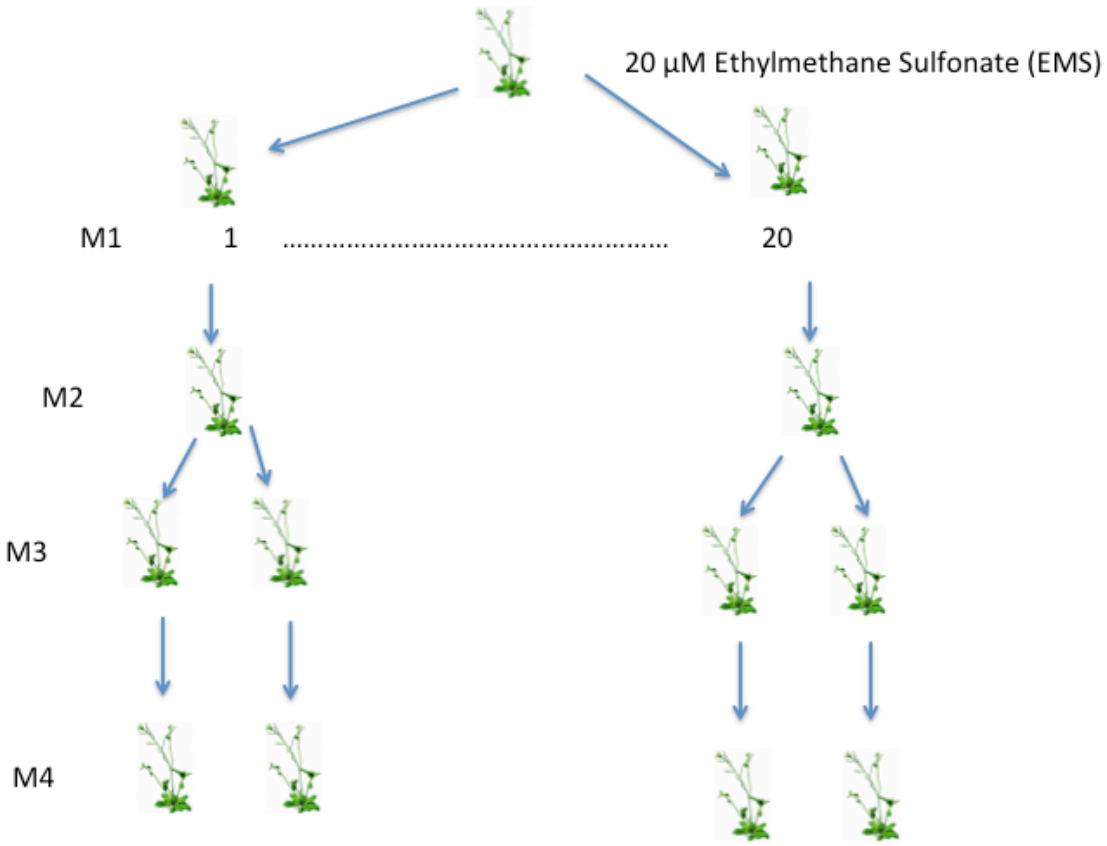

Figure 3. A schematic representation showing the process of generating mutation sublines from *Arabidopsis thaliana*. Seed were collected from a single Columbia founder and treated with 20 µM ethylmethane sulfonate for 12 hours. Twenty mutant lines were derived from this treatment. The lines were then split into two sublines each (M3 generation) and seed from these sublines (via selfing) were planted in the field (Beltsville Experimental Agricultural Station (UMD) in Beltsville MD (N 39.05378 W -76.95387) or in the growth room at the University of Maryland College Park campus.

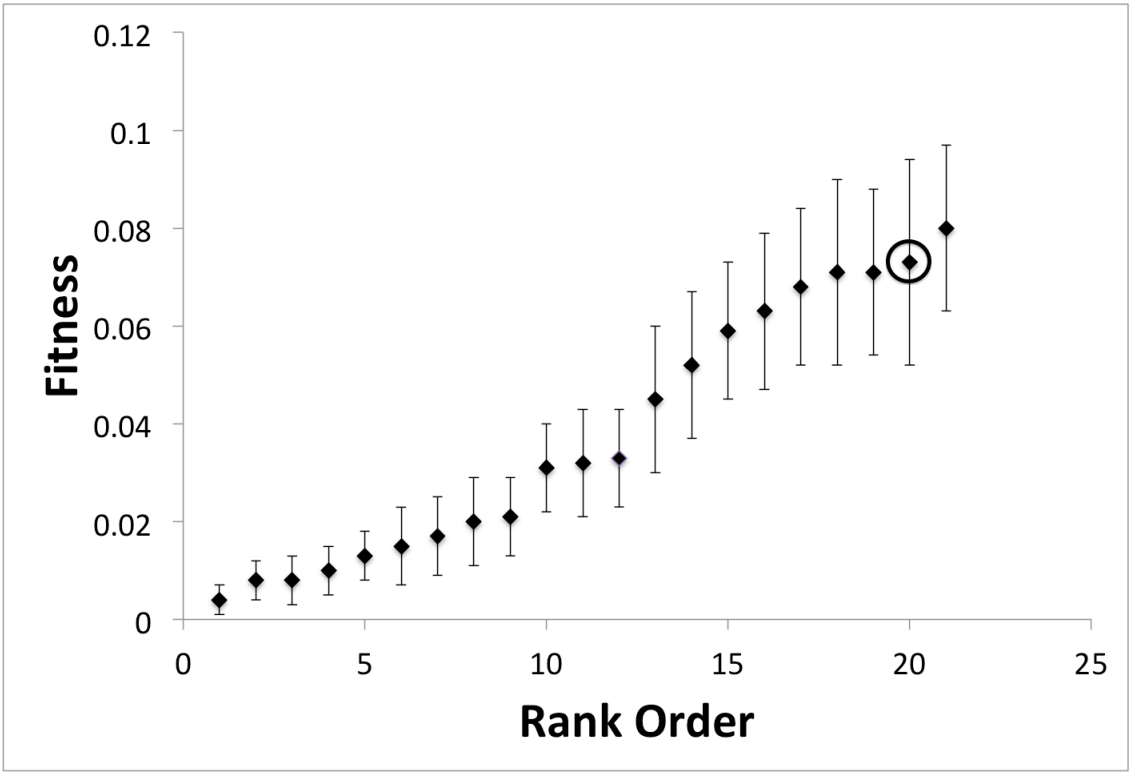

Figure 4. The mean fitness (mass x survivorship) for *Arabidopsis thaliana* under field conditions at the Beltsville Experimental Agricultural Station (UMD) in Beltsville MD (N 39.05378 W -76.95387) listed in rank order for the premutation founder and 20 mutant lines generated from ethylmethane sulfonate (20 μM) mutagenesis (planted: n = 120; weighed: mutant line average sample size n = 12.5, σ of sample size = 6.57, founder n = 17). A mutant line effect was detected ($p<0.001$). The premutation founder is circled. All but one mutant line showed reduced fitness compared to the founder ($p<0.001$). Error bars indicate one standard error from the mean.

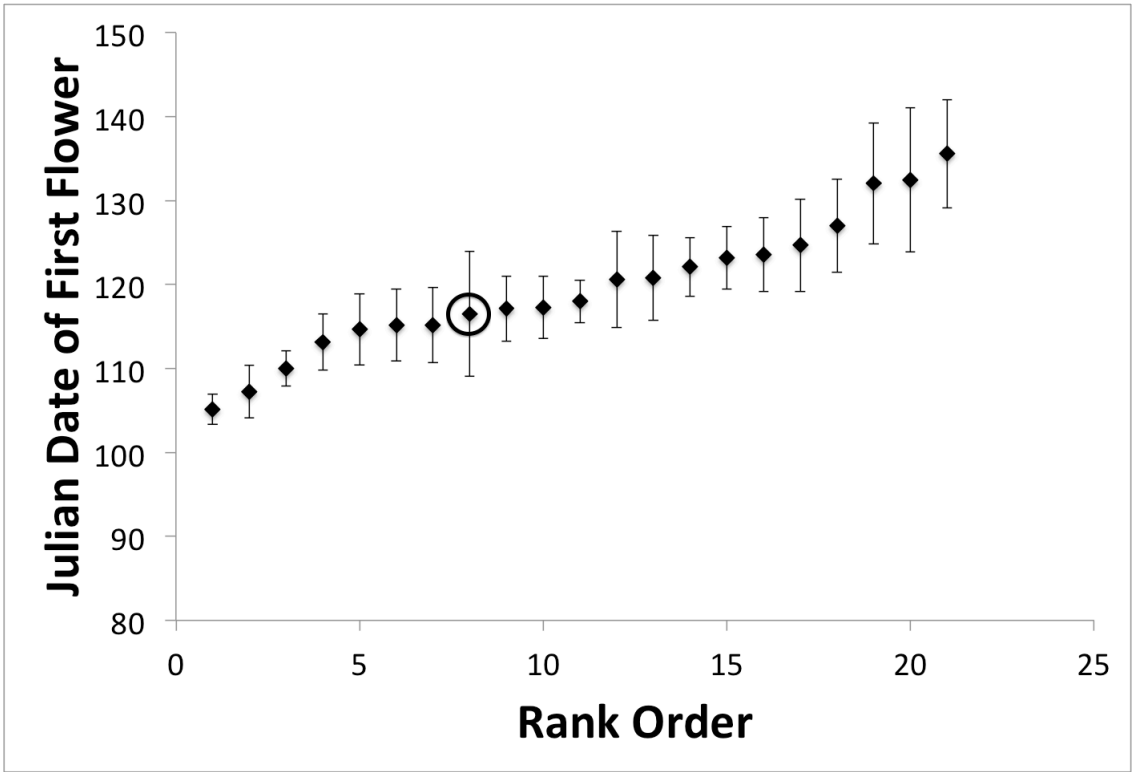

Figure 5. The mean julian day of first flower for *Arabidopsis thaliana* under growth room conditions (20^{o}C, 8 hours of incandescent light) on the University of Maryland College Park campus. Means are presented in rank order for the premutation founder and 20 mutant lines generated from ethylmethane sulfonate (20 μM) mutagenesis (mutant line mean sample size n = 15.21, σ of sample size = 3.84; founder n = 6). A mutant line effect was detected ($p<0.001$). The premutation founder is circled. Error bars indicate one standard error from the mean.

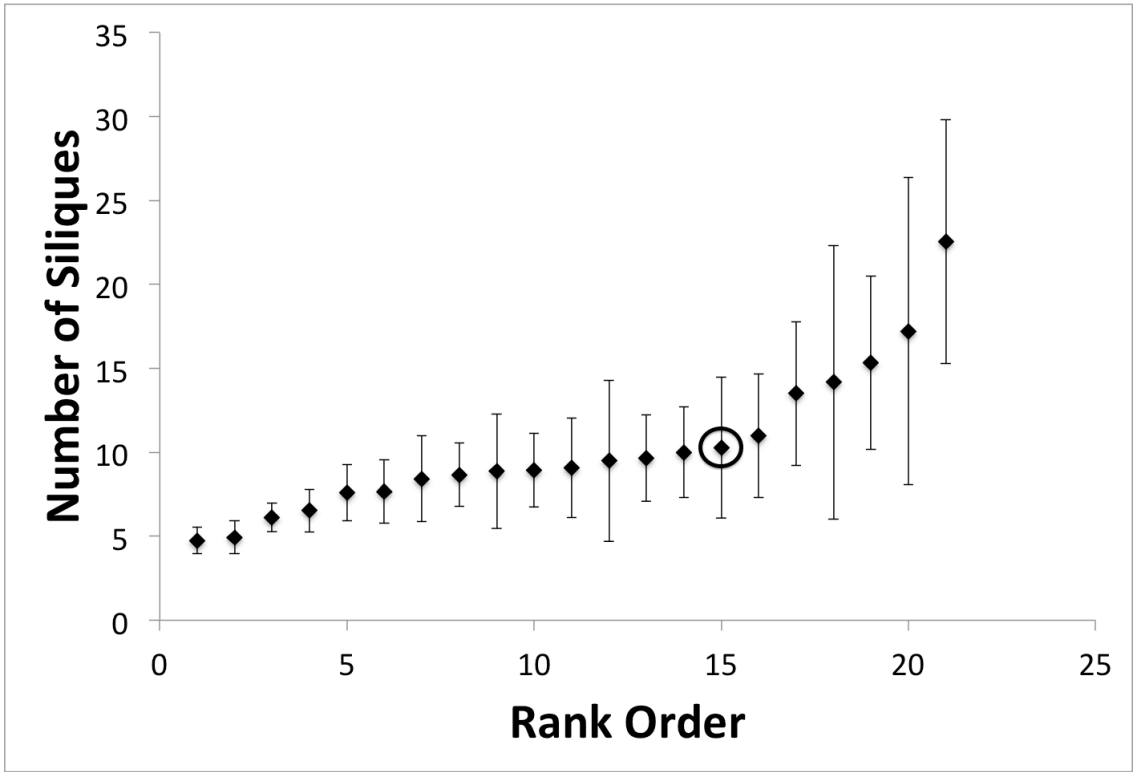

Figure 6. The mean number of siliques for *Arabidopsis thaliana* under growth room conditions (20^{o}C, 8 hours of incandescent light) on the University of Maryland College Park campus. Means are presented in rank order for the premutation founder and 20 mutant lines generated from ethylmethane sulfonate (20 μM) mutagenesis (mutant line mean sample size n = 14.80, σ of sample size = 4.12; founder n = 7). A mutant line effect was detected ($p<0.001$). The premutation founder is circled. Error bars indicate one standard error from the mean.

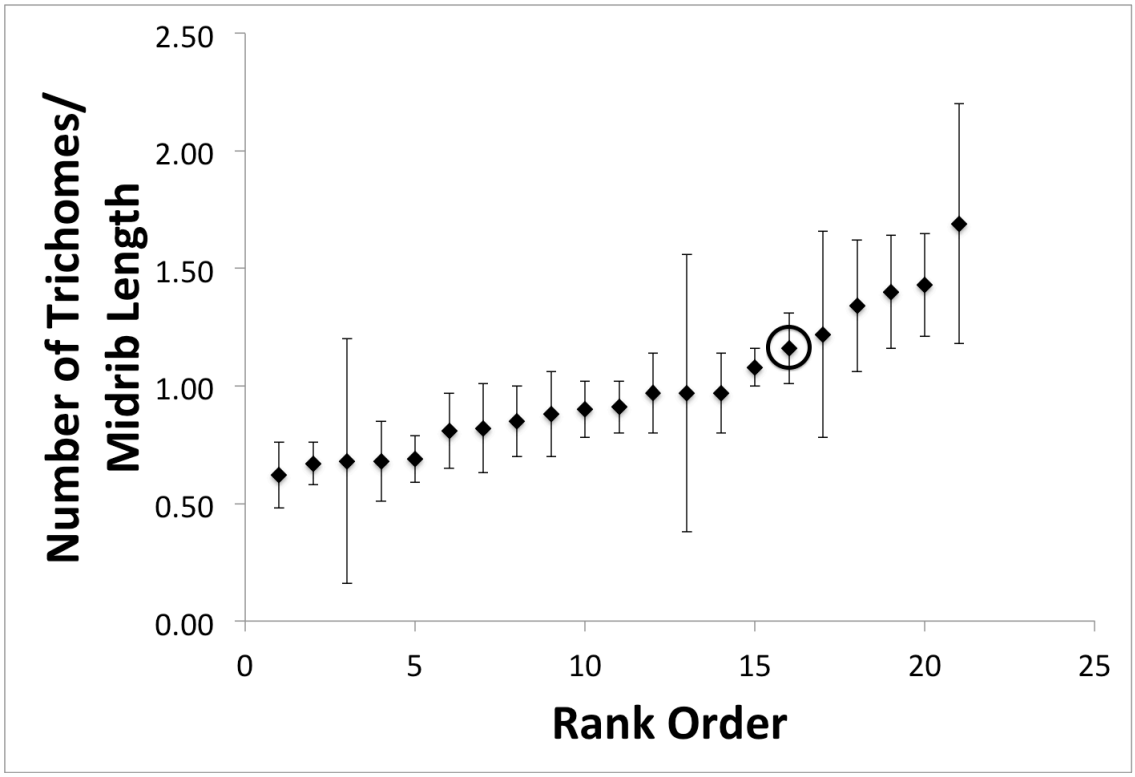

Figure 7. The mean number of trichomes per midrib length (mm) for *Arabidopsis thaliana* under growth room conditions (20^{o}C, 8 hours of incandescent light) on the University of Maryland College Park campus. Means, in rank order, are presented for the premutation founder and 20 mutant lines generated from ethylmethane sulfonate (20 μM) mutagenesis (mutant line mean sample size n = 10.65, σ of sample size 3.73; founder n = 7). The effect of mutant line on the trait was not significant after sequential Bonferroni correction (p=0.025, α=0.0125). The premutation founder is circled. Error bars indicate one standard error from the mean.

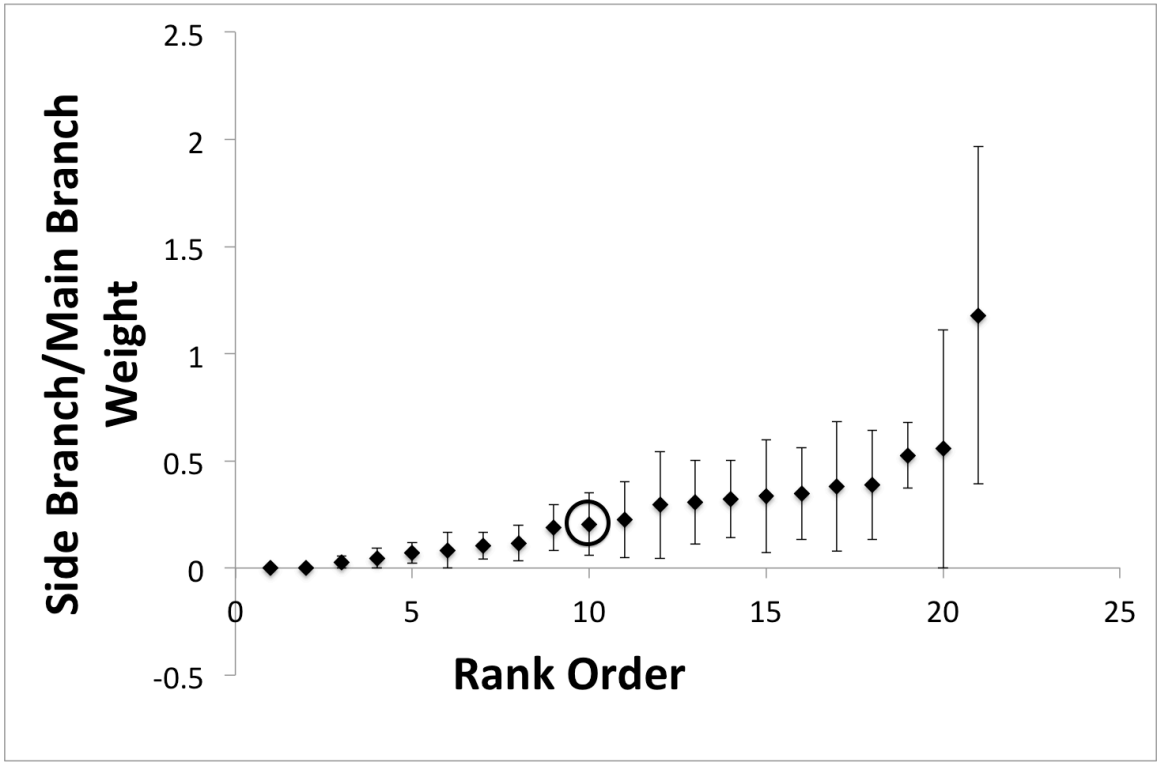

Figure 8. The mean proportion of side branches/main branch mass for *Arabidopsis thaliana* under growth room conditions (20^{o}C, 8 hours of incandescent light) on the University of Maryland College Park campus. Means are presented in rank order for the premutation founder and 20 mutant lines generated from ethylmethane sulfonate (20 μM) mutagenesis (mutant line mean sample size n= 12.40, σ of sample size 2.89; founder n = 7). Plants without side branches were given a side branch mass of 0. There was no significant effect of mutant line on the trait (p=0.31). The premutation founder is circled. Error bars indicate one standard error from the mean.

4 FISHER'S GEOMETRIC MODEL PREDICTS THE EFFECTS OF RANDOM MUTATIONS WHEN TESTED IN THE WILD

4.1 Introduction

Recent and historical studies of adaptation have focused on elucidating the genetic mechanisms that explain the pattern of past adaptation and accurately predict the outcome of the process of adaptation (Orr 2005; Bull and Otto 2005; Blows and Walsh 2009; Anderson et al. 2011; Zuellig et al. 2014). One of the pillars of the conceptual foundation of evolutionary genetics that is focused on quantifying adaptive evolution is Fisher's Geometric Model (FGM) (Burch and Chao 1999; Orr 2005; Martin and Lenormand 2006a; Blows and Walsh 2009). Fisher posited a multidimensional field of phenotypic trait combinations with a single fitness optimum; that is a single combination of traits with the highest local fitness (Fisher 1930). If a lineage is positioned somewhere in that field away from the optimum, a multidimensional sphere maybe be drawn with the optimum at the center and the lineage at the periphery (or circumference if a circle). Any movement to points inside the sphere will increase the fitness of the lineage (i.e., will be beneficial) and any movement outside of the sphere will reduce fitness (i.e., be deleterious). Movement through the field of phenotypic combinations is typically described as an adaptive walk where the steps are new beneficial mutations (Orr 1998; Orr 2005; Bull and Otto 2005; Seetharam and Jain 2014).

Fisher's Geometric Model has generated several predictions about the course of adaptation including that the genetic basis of the evolution of adaptive traits reflects the accumulation of numerous mutations each of small individual effect (mirroring the gradualism of evolution as posited by Darwin) (Orr 2005) and that the probability of improvement (P_i) through new mutations is proportional to the distance from the optimum (Fisher 1930; Joseph and Hall 2004; Agrawal and Whitlock 2010). The sphere of beneficial space will be smaller when a lineage is close to an optimum and there will be fewer beneficial points in the entire field of phenotypic combinations. While the first prediction of FGM has been challenged both theoretically and empirically (Kimura 1983; Orr 1998; Martin and Lenormand 2006a; Trindade et al. 2012), the second prediction has been corroborated but only through empirical studies on microorganisms under laboratory condition (MacLean et al. 2010; Khan et al. 2011; Perfeito et al. 2014; Kryazhimsky et al. 2104).

A logical next step in this area of research is to determine whether or not FGM will have the same predictive power for macro-organisms under more realistic field conditions. In particular, it is important to know if FGM can predict the effect of new mutations on fitness when fitness is a result of interactions between genotypes and the greater degree of environmental variance that results from more complex environmental stresses (Agrawal and Whitlock 2010; Anderson et al. 2011) and when the mutations themselves are random and their phenotypic effects unknown. These properties of FGM can be unambiguously tested by examining the change in

phenotypic fitness components even without knowing the specific nucleotide sequence changes.

Here, we use 19 founders, mutate them with a mutagen, EMS, and quantify the performance of the EMS derived mutants relative to their respective founders. We ask whether FGM is able to predict the effects of new mutations given the relative fitness of the pre mutation genotype in the model plant *Arabidopsis thaliana* under field conditions, as predicted by FGM.

4.2 Methods

4.2.1 Mutagenesis

Nineteen accessions were chosen from across the natural range of *Arabidopsis thaliana* (Table 4). All accessions were mutated using a 20 μM solution of ethylmethane sulfonate (EMS). EMS is an alkylating agent that most commonly results in G:C to A:T substitutions (transitions) (Greene et al. 2003). These mutations are expected to affect a wide range of quantitative traits (Brock 1976; Camara et al. 2000). Mutation rate was evaluated by exposing the Columbia accession to 20, 30, 40 and 50 μM solutions and estimating the number of siliques with albino seeds from a sample of each treatment, a common measure of mutation rate in *A. thaliana* (Camara et al. 2000). The consistency of the dosage was assayed by exposing a subset of the 19 accessions to 20 μM EMS and estimating as above (20-30% of fruit with albino seeds). Seeds were washed with a 0.1% solution of Tween-20 and then soaked for 12 hours in EMS with rotation. Based on this dosage and exposure, we estimate 25

mutations were induced per line. This estimate is based on the measured rate of mutations induced by EMS, which is approximately 3.7×10^{-6} locus^{-1} • cell^{-1} • μM^{-1} • hr^{-1} (Korneef et al. 1982; Camara et al. 2000). *Arabidopsis thaliana* is estimated to have about 28,000 loci (Redei and Koncz, 1992).

The seeds were next washed and soaked in distilled water for 6 hours with rotation. After washing, the seeds were sown directly onto soil in individual pots. These seeds and seeds sown from the pre-mutated founders for each accession were cold treated at 2º C for two weeks. They were then allowed to germinate on benches at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light for two weeks. After this time, individual seedlings (10 per treatment) were transplanted to their own pots. Seed collected from these individuals were used for fitness assessment in the field. Based on the estimate of 25 induced mutations per line, each of these individuals has approximately six of these mutations in the homozygous state, with another 12 induced mutations remaining as heterozygous. The remainder of the induced mutations would be lost and revert to the wild type.

4.2.2 Fitness Assessment

Mutated and pre-mutated seeds from among the same 19 accessions were planted under three different experimental field conditions: at Blandy Experimental Farm (UVA) in Boyce VA (N 39.06261 W -78.06222) during Fall of 2011 and at the Beltsville Experimental Agricultural Station (UMD) in Beltsville MD (N 39.05378 W -76.95387) in Fall of 2012 and Spring of 2013. Weather data records for Blandy Experimental Farm can be found here:

http://www.virginia.edu/blandy/blandy_web/all_blandy/blandy_weather_history.php and for the Beltsville Experimental Agriculture Station here: http://www.ba.ars.usda.gov/weather/ba-weather-2.html

Seeds were cold treated for two weeks at 2º C and allowed to germinate on benches at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light for two weeks, as above. Seedlings were then transplanted in random order into plug trays and allowed two more weeks to establish; on a bench at 20º C with 24 hours of fluorescent light and 8 hours of incandescent light for the spring planting and at 2º C with 8 hours of incandescent light for the two fall plantings in order to cold acclimate them. Plug trays were transplanted into the field spaced 10 cm apart, maintaining the spatial orientation of seedlings in the plug trays. In Fall of 2011 seven accessions were planted with 140 plants per pre-mutant treatment and 70 plants from the mutant lines for each accession, for a total of 1470 plants. In Fall of 2012, 12 accessions were planted with150 plants per treatment for a total of 3600 plants. In Spring of 2013, 15 accessions were planted with 150 plants per treatment, for a total of 4500 plants. When planted, the seedlings were at the 2 - 4 leaf stage. The plots were initially watered at planting to facilitate establishment but otherwise were exposed to natural weather conditions and pathogens and predation from herbivores and competition to other plant species. All experimental plants (from both spring and fall plantings) were harvested at the end of May and dried in heat chambers. Dry weight and total fruit number were measured from a sample for each experiment, and found to be highly correlated (Fall 2011, $r^2 = 0.89$; Fall 2012, $r^2 = 0.89$; Spring 2013, $r^2 = 0.90$). We therefore used average dry weight of survivors multiplied by proportion of plants

surviving to harvest as our measure of fitness for each mutant and pre-mutant treatment. This is a reasonable proxy of fitness for a selfing annual plant (Shaw et al. 2000). These experimental plantings were expected to represent three different optima; and indeed a spearman rank correlation between two of the experimental conditions with adequate representation of accessions indicated that the same accessions responded differently to each planting (r^2=0.004). This result allows us to make inferences about the distance from the optimum on the fitness effects of the new mutations separate from the effects due to the genetic background of any particular accession. We used the difference between the average fitness of the mutant lines and the average fitness of the founder for each accession to estimate the probability of improvement via mutation. As fitness is inversely proportional to the distance from the optimum (low fitness lineages are further from the optimum) FGM predicts a negative linear relationship between founder fitness and the average improvement (Fisher 1930; Joseph and Hall 2004; Agrawal and Whitlock 2010).

4.2.3 Statistical Analysis

We analyzed the data using a linear regression analysis of the pre-mutation found fitness (distance from the optimum) by the difference between the mutant fitness and pre-mutation founder fitness for each accession (probability of improvement). The data from each experimental plot were analyzed separately using the *lm* function from the package STATS in R (R Core Team 2014). In order to investigate more general results of all three optima, we also standardized all data points using z-scores and analyzed these as one group using the *lm* function in R. We used the latter for our main conclusions. We were interested in two measures, r^2 (which indicates how well

founder fitness predicts the probability of improvement) and the slope of the line and its significance from zero (which provides information about the general prediction from FGM). We used Cook's D test (performed in R, package CAR) to identify potential outliers and found three (Fox and Weisberg, 2011). We ran the analysis with and without these points and did not find any qualitative differences. Removal of these did not change the overall pattern (r^2 of 0.12 and slope of -0.26, p=0.026, 95% CI: -0.49 to -0.03). Since we had no reason to doubt the accuracy of these points, they were left in the final results. All R codes can be found in Appendix A.

4.3 Results

We plot the founder fitness (distance from the optimum) against the difference between the fitness of the mutant lines and the pre-mutant founder lines (probability of improvement on average) for each accession (Figure 9). The slope of the line is -0.45 and differs significantly from zero ($p<<0.0001$; 95% CI: -0.64 to -0.25). The r^2 is 0.39. While most of the mutant lines were nearly neutral or deleterious with respect to fitness, that is, the mutant lines performed worse on average than the premutation genotypes from which they were derived, there were those where mutations increased fitness on average (Figure 10). The effects of deleterious mutations on fitness are not larger, on average, $\bar{x}$ = -0.70 (all mutant lines having lower fitness than their respective premutation genotypes, n = 15 mutant-premutation comparisons), than the effects of beneficial mutations, $\bar{x}$ = 0.43 (all mutant lines having higher fitness than their respective premutation genotypes, n = 20 mutant-premutation comparisons) (t = 1.33, df = 32, p = -0.19).

4.4 Discussion

We found that when *A. thaliana* accessions are exposed to field conditions, the fitness of pre-mutation founders can accurately predict whether a set of random mutations on average will be beneficial or deleterious. Our results support one of the predictions from Fisher's Geometric Model of Adaptation, the assertions that lineages sitting more distant from a local fitness optimum are more likely to experience an improvement in fitness through new mutations (Fisher 1930; Joseph and Hall 2004; Agrawal and Whitlock 2010). As well, the relatively high r^2 suggests that the degree to which fitness changes will occur is predictable.

To our knowledge, this is the first time that this prediction of FGM has been demonstrated in a macroorganism under natural field conditions. Despite the fact that the mutations were unknown and environmental variance was high, our results corroborate those of previous studies using microorganisms under laboratory conditions (Khan et al. 2011; Trindade et al. 2012; Perfeito et al. 2014; Kryazhimsky et al. 2014). A pattern of diminishing returns epistasis has been found in organisms as diverse as *Escherichia coli* (Khan et al. 2011), *Pseudomonas aeruginosa* (MacLean et al. 2010) and *Saccharomyces cerevisiae* (Kryashimsky et al. 2014), where the magnitude of beneficial mutations decreases over the course of adaptation as a lineage approaches a fitness peak. Another study in *E. coli* (Perfeito et al. 2014) found similar results as we have, where there is a negative linear relationship between fitness of a lineage and the net fitness effect of new mutations. They attribute this to

epistatic effects between new mutations and genetic background. Either of these (diminishing returns or sign epistasis) may account for our findings, and both are consistent with FGM.

Two studies report results inconsistent with our study. When exposed to two different host environments, mutations in the bacteriophage ϕX174 were found to have the same or similar effects on fitness (Vale et al. 2012). In contrast, we observed independent effects of mutations on fitness across environments as mutation effects were conditional on the performance of the founders in any particular environment. In another study, distance from a hypothesized fitness optimum was found to effect the variance in the effect of new mutations on fitness, but not their average effect (Trindade et al. 2012). This last result is consistent with a theoretical extension of FGM (Martin and Lenormand 2006b) demonstrating that the mean effect of new mutations on fitness would not be impacted from the distance from a fitness optimum, but the variance of mutation effects on fitness would increase with increasing distance. Our results suggest otherwise, i.e., average effect of mutations are dependent on the distance to the optima, and seem more in line with a traditional interpretation of the FGM (Fisher 1930; Joseph and Hall 2004).

Although not a main focus of this project, we found that the distribution of the average effects of new mutations on fitness from our study to be consistent with the results from previous mutation accumulation research on *Arabidopsis thaliana* (Shaw et al. 2000; MacKenzie et al. 2005; Rutter et al. 2010). The small number of founders

and derived mutant lines prevents strong statements on the distribution of mutations other than the lower the fitness of the premutation founder the more likely beneficial mutations will occur. However, the patterns revealed in both figures are consistent with previous studies with *A. thaliana*, i.e., the summed effects of most mutations have little detectable effects on fitness, and there is a high frequency of beneficial mutations.

As mutation accumulation experiments detect the effects of spontaneous mutations, the distribution of mutation effects on fitness from our experiment, although chemically induced, is a reasonable estimate of the effect of naturally occurring point mutations. This result is also consistent with FGM, where the breakdown of phenotypic space that a mutation can access is likely more deleterious than beneficial (Fisher 1930). Ethylmethane sulfonate induced mutations are primarily G:C -> A:T transitions (Greene et al. 2003). While spontaneous mutations have been found to consist primarily of these same transitions, they also have approximately 15% insertion-deletions (including long deletions) (Ossowski et al. 2010). Given the expectation that insertion-deletions are likely to have greater phenotypic effect, our results likely underestimate the full spectrum of effects of mutations on fitness. Further, the simple linear relationship and the high number of beneficial mutations suggests that adaptation of *A. thaliana* in this habitat is due to a small number of dimensions (and a consistent number among all accessions) as predicted by the "Cost of Complexity" (Orr 2000).

4.5 Conclusion

If adaptation in natural populations follows such straightforward rules as documented here than we should be able to anticipate the course of adaptation to novel environmental perturbations via new mutations. Our results have at least three implications. First, since we have been able to observe a considerable frequency of beneficial mutations through randomized mutagenesis, also seen in other recent studies including those quantifying the effects of spontaneous mutations (Shaw et al. 2000; MacKenzie et al. 2005; Perfeito et al. 2007; Hall et al. 2008; Rutter et al. 2010; Zhang et al. 2011), it is clear that new mutations can play a large role in the evolution of adaptation. Second our findings are consistent with recent refinements of FGM in that genotypes far from the optimum can adaptively evolve through a wider set of mutations (including those with relatively large effect) but as the genotype evolves to be closer to the optimum, fewer mutations, likely of smaller effect, will contribute to the continuity of the process. Third, our results may have considerable practical applications for conservation genetics. As more and more species find themselves exposed to environments where they are likely displaced from a fitness optimum due to limits imposed by human driven fragmentation of the landscape and anthropomorphic climate change, mutations, although contributing to a population's demise under certain conditions (Lande 1994) may also contribute to its persistence.

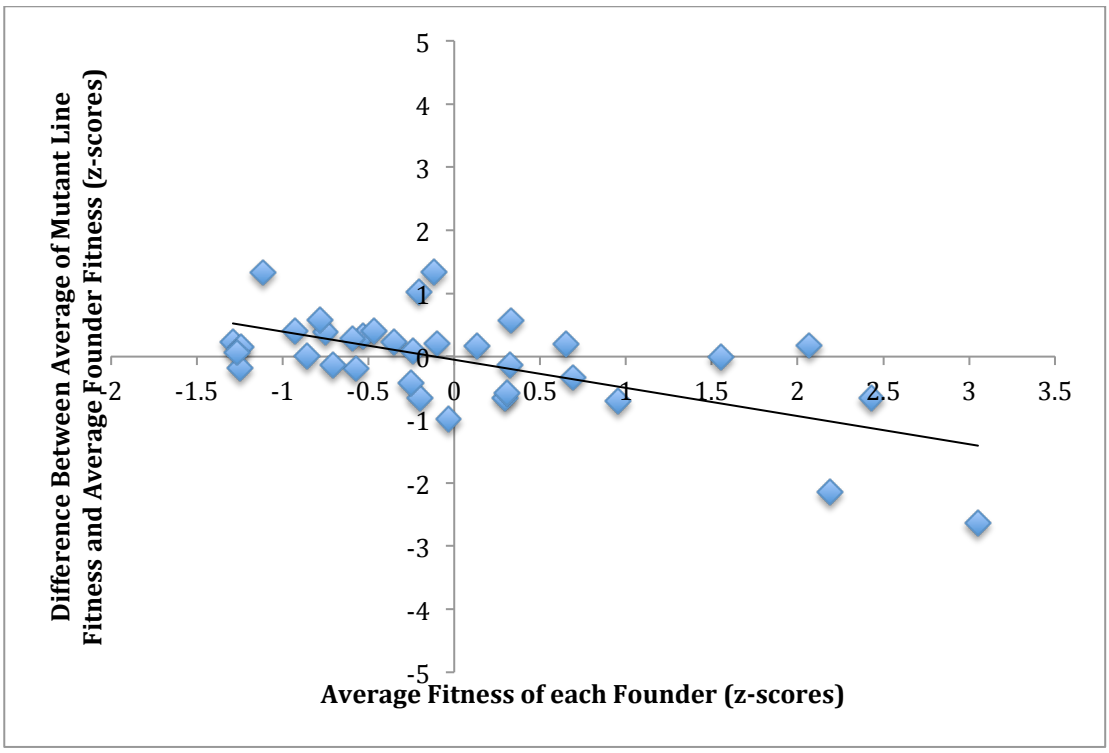

Figure 9. Linear relationship between pre-mutant founder fitness and the difference between average EMS induced mutant line fitness and pre-mutant founder fitness for *Arabidopsis thaliana* studied in the field (planted: mean n = 108.8, σ = 36.2; weighed: mutant lines mean n = 18.7,σ = 14.4, founder lines mean n = 24.5,σ = 19.4). This graph reflects a summary of all three studies (n=35), with data points converted to z-scores. The x-axis is the average fitness of the pre-mutant founder (reflecting distance from the optimum) and the y-axis is the difference between the average of the mutant line and the average of the pre-mutant founders (representing the probability of improvement). The slope of the line is -0.45 ($p<<0.0001$, 95% CI: -0.49 to -0.03) with an r^2 of 0.39.

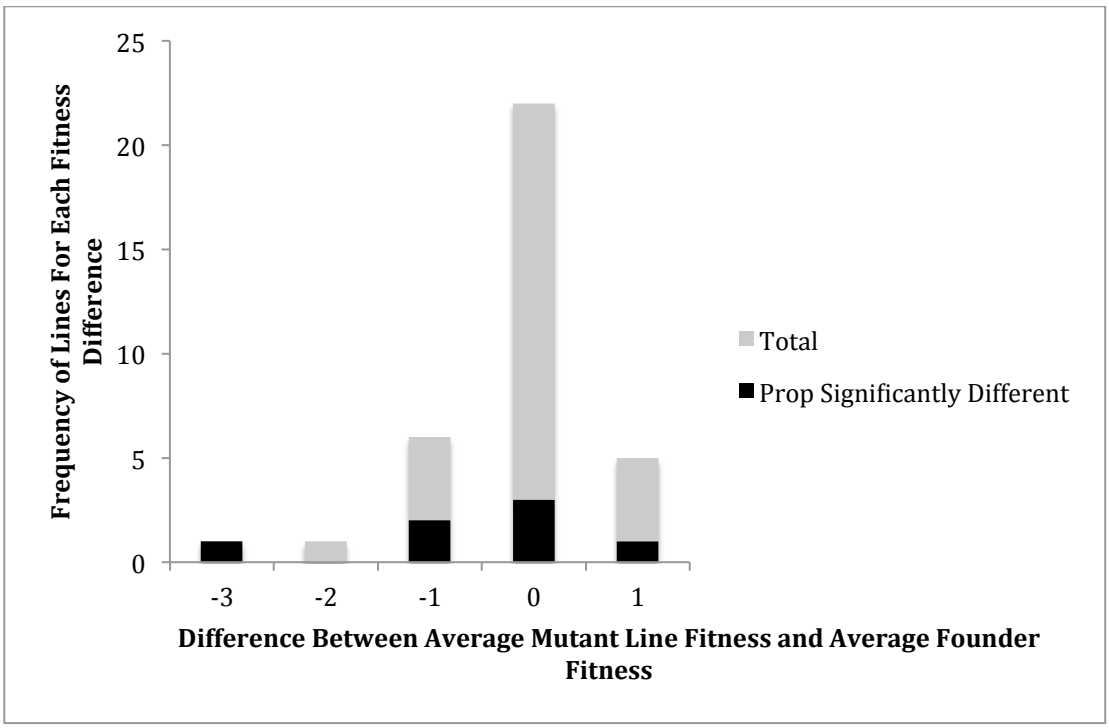

Figure 10. Distribution of average EMS induced mutation affects on fitness for *Arabidopsis thaliana* studied in the field. Presented is a frequency histogram of the difference between the average of the mutation lines and the average of the pre-mutant founder lines for each accession (n=35). Fitness differences were rounded to the nearest whole number. The proportion of each class where the mutant lines differed significantly from the pre-mutant founder lines (calculated by Mann-Whitney U test) is indicated in black.

Table 4. Nineteen *Arabidopsis thaliana* accessions used in the experiments. Not all accessions were used in each experiment, the planting season is indicated in the last column (F11 = Fall 2011 at Blandy Experimental Farm, UVA; F12 = Fall 2012 at the Beltsville Farm, UMD; S13 = Spring 2013 at the Beltsville Farm, UMD).

Table 4

Accession	**Locality**	**Latitude**	**Longitude**	**Season**
CS902	Cape Verde Islands,	16.0N	24.0W	F12, S13
CS911	Estland, Germany	53.8N	10.6E	F11
CS926	Petergof, Russia	60.0N	30.0E	F11, F12, S13
CS1096	Cape Verde Islands	16.0N	24.0W	S13
CS1122	Edinburgh, UK	56.0N	3.2W	F11
CS6180	Shakdara, Tadjikstan	37.0N	71.0E	F11, F12, S13
CS28051	Arby, Sweden	59.3N	16.5E	F12, S13
CS28200	Darmstadt, Germany	50.0N	8.7E	F12, S13
CS28364	Jena, Germany	50.9N	11.6E	F12, S13
CS28375	Karnten, Austria	46.7N	14.2E	S13
CS28510	Solomennoye, Russia	56.5N	31.7E	F12, S13
CS28779	Tsagguns, Austria	47.1N	10.0E	F12, S13
CS76116	Cape Verde Islands	16.0N	24.0W	S13
CS76197	Niederzenz, Germany	50.0N	8.0E	F12, S13
ARRIGAS	Arrigas, France	44.0N	3.5E	F12, S13
CNRS1	Montpellier, France	43.6N	3.9E	F11, F12, S13
COL	Missouri, USA	39.0N	92.3E	F11, F12, S13
KATA MTN	Golan Heights, Israel	31.1N	34.9E	F11
SKATVAL	Skatval, Norway	63.5N	10.9E	S13

Appendix A: R Codes

Chapter 2:

Geographic Distance Matrix:

```
x.sp <- SpatialPoints(x)

dist <- spDists(x.sp, longlat = TRUE)
```

Mantel Test:

```
mantel(y, x, method="pearson", permutations=999)
```

Partial Mantel Test:

```
mantel.partial(z, y, x, method = "pearson", permutations = 999)
```

Chapter 3:

Kruskal-Wallis:

```
kruskal.test(formula, data, subset, na.action, ...)
```

One-way ANOVA:

```
aov(formula, data = NULL, projections = FALSE, qr = TRUE,
   contrasts = NULL, ...)
```

Generalized Linear Model (GLM) for a negative binomial distribution:

```
glm.nb(formula, data, weights, subset, na.action, start = NULL,
       etastart, mustart, control = glm.control(...),
       method = "glm.fit", model = TRUE, x = FALSE, y = TRUE,
       contrasts = NULL, ..., init.theta, link = log)
```

Pearson Correlation:

```
cor(data, use="all.obs", method="pearson")
```

Dunnet's Test:

```
data.aov<-aov(Trait ~ Line, data=d)
```

data.dunnetts<-glht(data.aov, linfct=mcp(Line="Dunnett"))

Linear Model:

lm(formula, data, subset, weights, na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, ...)

Chapter 4:

Linear Model:

lm(formula, data, subset, weights, na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, ...)

Cook's D:

```
plot(cooks.distance(model))
cutoff <-4/model$df.residual
abline(h=cutoff, lty=2, col=c("orange", "red"))
identify(cooks.distance(model))
```

Bibliography

Agrawal, A. F., and M. C. Whitlock. 2010. Environmental duress and epistasis: how does stress affect the strength of selection on new mutations? *Trends Ecol. Evo.* **25**:450-458.

Agren, J., C. G. Oakley, J. K. McKay, J. T. Lovell, and D. W. Schemske. 2013. Genetic mapping of adaptation reveals fitness trade-offs in *Arabidopsis thaliana*. *Proc. Nat. Acad. Sci. USA* **110**:21077-21082.

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

Anderson, J. T., Willis, J. H., and T. Mitchell-Olds. 2011. Evolutionary genetics of plant adaptation. *Trends Gen.* **27**:258-266.

Angert, A. L., L. G. Crozie, L. J. Rissler, S. E. Gilman, J. J. Tewksbury, and A. J. Chunco. 2011. Do species' traits predict recent shifts at expanding range edges? *Ecol. Lett.* **14**:677-689.

Arabidopsis Biological Research Center. 2015. See https://abrc.osu.edu/

The Arabidopsis Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**:796-815.

Auerbach, C. 1949. Chemical Mutagenesis. *Biol. Rev. Cambridge Phil. Soc*. **24:**355-391.

Banta, J. A., I. M. Ehrenreich, S. Gerard, l. Chou, A. Wilczek, J. Schmitt, P. X. Kover, and M. Purugganan. 2012. Climate envelope modeling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in *Arabidopsis thaliana*. *Ecol. Lett.* **15**:769-777.

Barrell, B. G., G. M. Air, and C. A. Hutchison. 1976. Overlapping genes in bacteriophage ϕX174. *Nature* **264**:34-41.

Barrett, R. W., R. C. MacLean, and G. Bell. 2006. Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. *Biol. Lett.* **2**:236-238.

Barrett, R. W., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* **322**:255-257.

Barrett, R. D., and D. Schluter. 2008. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**:38-44.

Beadle, G. W., and E. L. Tatum. 1941. Genetic control of biochemical reactions in *Neurospora*. *Proc. Nat. Acad. Sci. USA* **27:**499-506.

Beadle, G. W., and E. L. Tatum. 1945. *Neurospora* .2. Methods of Producing and Detecting Mutations Concerned with Nutritional Requirements. *Am. J. Bot.* **32:**678-686.

Beck, J. B., H. Schmuths, B. A. Schaal. 2008. Native range genetic variation in Arabidopsis thaliana is strongly geographically structured and reflects Pleistocene glacial dynamics. *Mol. Ecol.* **17**:902-915.

Benne, R., J. Van den Burg, J. P. J. Brakenhoff, P. Sloof, J. H. Van Boom, and M. C. Tromp. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46:**819-826.

Berget, S. M., C. Moore, and P. A. Sharp. 1977. Spliced Segments at 5' Terminus of Adenovirus 2 Late Messenger-RNA. *Proc. Nat. Acad. Sci. USA* **74:**3171-3175.

Black, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Ann.Rev. Biochem.* **72:**291-336.

Bomblies, K., and J. F. Doebley. 2006. Pleiotropic effects of the duplicate maize FLORICAULA/LEAFY genes *zfl1* and *zfl2* on traits under selection during maize domestication. *Genetics* **172**:519-531.

Blows, M., and B. Walsh. 2009. Spherical cows grazing in flatland: Constraints to selection and adaptation. Pp. 83-101 *in* J. VanderWerf, H. U. Graser, R. Frankham, and C. Gondro, eds. *Adaptation and Fitness in Animal Populations*. Springer, Netherlands.

Bradshaw, W. F., C. M. Holzapfel. 2001. Genetic shift in photoperiod response correlated with global warming. *Proc.Nat.l Acad.Sci. USA* **98**:14509-14511.

Brock, R. D. 1976. Quantitatively inherited variation in *Arabidopsis thaliana*, induced by chemical mutagens. *Environ. Exp. Bot.* **16**:241-253.

Bull, J. J., and S. P. Otto. 2005. The first steps in adaptive evolution. *Nat. Gen*. **37**:342-343.

Burch, C. L., and L. Chao. 1999. Evolution by small steps and rugged landscapes in the RNA virus ϕ6. *Genetics* **151**:921-927.

Camara, M. D., and M. Pigliucci. 1999. Mutational contributions to genetic variance-covariance matrices: An experimental approach using induced mutations in *Arabidopsis thaliana*. *Evolution* **53**:1692-1703.

Camara, M.D., C. A. Ancell, and M. Pigliucci. 2000. Induced mutations: A novel tool to study phenotypic integration and evolutionary constraints in *Arabidopsis thaliana*. *Evol. Ecol. Res.* **2**:1009-1029.

Carbone, M. A., K. W. Jordan, R. F. Lyman, S. T. Harbison, J. Leips, T. J. Morgan, and T. MacKay. 2006. Phenotypic variation and natural selection at *Catsup*, a pleiotropic quantitative trait gene in *Drosophila*. *Curr. Biol.* **16**:912-919.

Caspari, E. 1952. Pleiotropic Gene Action. *Evolution* **6:**1-18.

Chang, S., and R. G. Shaw. 2003. The contribution of spontaneous mutation to variation in environmental response in *Arabidopsis thaliana*: Responses to nutrients. *Evolution* **57**:984-994.

Cheverud, J. M. 1996. Developmental integration and the evolution of pleiotropy. *Am. Zool.* **36:**44-50.

Clausen, J. 1926. Genetical and cytological investigations on *Viola tricolor* L. and *V. arvensis* Murr. *Hereditas* **8**:1-156.

Clausen, J., D. D. Keck, and W. M. Hiesey. 1947. Heredity of geographically and ecologically isolated races. *Am. Nat.* **81:**114-133.

Cordell, H. J. 2002. What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11**:2463-2468.

Coyne, J. A., N. H. Barton, and M. Turelli. 1997. A critique of Sewall Wright's Shifting Balance Theory of Evolution. *Evolution* **51**:643-671.

Coyne, J.A., H. A. Orr. 2004. *Speciation*. Sinauer Associates Inc., Sunderland, Mass.

Crow, J. F. 1997. The high spontaneous mutation rate: Is it a health risk? *Proc. Nat. Acad. Sci. USA* **94**:8380-8386.

Crow, J. F. 2000. The origins, patterns, and implications of human spontaneous mutation. *Nat. Rev. Gen.* **1**:40-47.

Cuenot, L. 1903. L'hérédité de la pigmentation chez les Souris. *Archives de Zoologie Experimentale et Generale* **4:**xxxiii-xli.

Darwin ,C. 1859. *On the origin of species.* John Murrary, London, UK.

Davies, E. K., A. D. Peters, and P. D. Keightley. 1999. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**:1748-1751.

Des Marais, D. L., and M. D. Rausher. 200.8 Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**:762-766.

Denoto, F. M., D. D. Moore, and H. M. Goodman. 1981. Human Growth-Hormone DNA-Sequence and Messenger-Rna Structure - Possible Alternative Splicing. *Nucl. Acids Res.* **9:**3719-3730.

Dieckmann, U., and M. Doebel. 1999. On the origin of species by sympatric speciation. *Nature* **400**:354-357.

Dudash, M. R. 1990. Relative fitness of selfed and outcrossed progeny in a self-compatible, protandrous species, Sabatia angularis L. (Gentianaceae): A comparison on three environments. *Evolution* **44**:1129-1139.

Eckman, O. J. 1788. Dissertatio medica descriptionem et casus aliquot osteomalaciae sistens, Upsala.

Featherstone, D. E., and K. Broadie. 2002. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays* **24:**267-274.

Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**:164-166.

Fenster, C. B., L. F. Galloway, and L. Chao. 1997. Epistasis and its consequences for the evolution of natural populations. *Trends Eco. Evol.* **12**:282-286.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Phil. Trans. Roy. Soc. Edinburgh* **52**:399-433.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford, UK.

Flint, J., and T. F. C. Mackay. 2009. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19:**723-733.

Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, and A. M. Wilczek. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**:86-89.

Fox, J., and S. Weisberg. 2011. *An {R} Companion to Applied Regression,* Second Edition. Thousand Oaks CA: Sage.
URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Frankham, R. 2015. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Mol. Ecol*. doi: 10.1111/mec.13139.

Furuichi, Y., M. Morgan, S. Muthukrishnan, and A. J. Shatkin. 1975. Reovirus Messenger-Rna Contains a Methylated, Blocked 5'-Terminal Structure - M7g(5')Ppp(5')Gmpcp. *Proc. Nat. Acad. Sci. USA* **72:**362-366.

Gavrilets, S. 2004. *Fitness landscapes and the origin of species.* Princeton University Press, Princeton, NJ.

Goldschmidt, R. B. 1955. *Theoretical Genetics*. University of California Press, Berkeley, CA.

Gommans, W. M., S. P. Mullen, and S. Maas. 2009. RNA editing: a driving force for adaptive evolution? *Bioessays* **31:**1137-1145.

Gower, J. C. 1971. General coefficient of similarity and some of its properties. *Biometrics* **27**:857-871.

Greene, E .A., C. A. Codomo, N. E. Taylor, J. G. Henikoff, B. J. Till, S. H. Reynolds, L. C. Enns, C. Burtner, J. E. Johnson, A. R. Odden, L. Comai, and S. Henikoff. 2003. Spectrum of chemically induced mutations from a large scale reverse-genetic screen in *Arabidopsis*. *Genetics* **164**:731-740.

Gruneberg, H. 1938. An analysis of the "pleiotropic" effects of a new lethal mutation in the rat (*Mus norvegicus*). *Proc. Roy. Soc. London B* **125:**123-144.

Hadorn, E. 1945. Zur Pleiotropie der Genwirkung. Archives Julius Klaus Stiftung **20:**81-95.

Hadorn, E. 1961. *Developmental Genetics and Lethal Factors*. Methuen and Co Ltd., London, UK.

Haeker, V. 1925. Aufgaben und Ergebnisse der Phaenogenetik. *Bibliographia Genetica* **1:**1-314.

Hall, D. W., and S. B. Joseph. 2010. A high frequency of beneficial mutations across multiple fitness components in *Saccharomyces cerevisiae*. *Genetics* **185**:1397-1409.

Halligan, D. L., and P.D. Keightley. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Ann. Rev. Ecol. Evol. System.* **40**:151-172.

Hamilton, J. A., M. Okada, T. Korves, and J. Schmitt. 2015. The role of climate adaptation in colonization success in *Arabidopsis thaliana. Mol. Ecol*. doi:10.1111/mec.13099.

Hancock, A .M., B. Brachi, M. Fauer, M. W. Horton, L. B. Jarymowycz, F. G. Sperone, C. Toomajian, F. Roux, and J. Bergelson. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**:83-86

Hansen, M. M., I. Olivieri, D. M. Waller, E. E. Nielsen, The GeM Working Group. 2012. Monitoring adaptive genetic responses to environmental change. *Mol. Ecol.* **21**:1311-1329.

Hawthorne, D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature* **412:**904-907.

He, X. L., and J. Z. Zhang. 2006. Toward a molecular understanding of pleiotropy. *Genetics* **173:**1885-1891.

Heilbron, K., M. Toll-Reira, M. Kojadinovic, and R. C. MacLean. 2014. Fitness is strongly influenced by rare mutations of large effect in a microbial mutation accumulation experiment. *Genetics* **197**:981-990.

Hickman, M., and J. Cairns. 2003. The centenary of the one-gene one-enzyme hypothesis. *Genetics* **163:**39-841.

Hodgkin, J. 1998. Seven types of pleiotropy. *Int. J. Dev. Biol.* **42:**501-505.

Hoffman, A. A., and C. M. Sgro. 2011. Climate change and evolutionary adaptation. *Nature* **470**:479-485.

Hoffmann, M. H. 2002. Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). *J. Biogeography* **29**:125-134.

Hoffmann, M. H. 2005. Evolution of the realized climate niche in the genus *Arabidopsis* (Brassicaceae). *Evolution* **59**:1425-1436.

Hoffmann, M. H., J. Tomiuk, H. Schmuths, C. Koch, and K. Bachmann. 2005. Phenological and morphological responses to different temperature treatments differ among a world-wide sample of accessions of *Arabidopsis thaliana*. *Acta Oecologica* **28**:181-187.

Horowitz, N. H. 1995. One-Gene One-Enzyme - Remembering Biochemical Genetics. *Protein Sci*. **4:**1017-1019.

Horton, M. W., A. M. Hancock, Y. S. Huang, C. Toomajian,, S. Atwell, A. Auton, N. Wayan Muliyati, A. Patt, F. G. Sperone, B. J. Vilhjalmsson, M. Nordborg, J. O. Borevitz, and J. Bergleson. 2012. Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat. Gen.* **44**:212-216.

Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous Inference in General Parametric Models. *Biomet. J.* **50**:346--363.

Huberts, D. H., and I. J. Van Der Klei. 2010. Moonlighting proteins: An intriguing mode of multitasking. *Biochimica et Biophysica Acta* **1803:**520-525.

Jaenike, J. 1982. On the relation between genetic and environmental variability in animals. *J. Molec. Evol*. **18**:310-314.

Jambhulkar, S. J. 2007. Mutagenesis: Generation and evaluation of induced mutations. *Adv. Bot. Res.: Inc. adv. plant path.* **45**:417-434.

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino, and C. Dean. 2000. Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**:344-347.

Jorgensen, S., and R. Mauricio. 2004. Neutral genetic variation among wild North American populations of the weedy plant *Arabidopsis thaliana* is not geographically structured. *Molec. Ecol.* **13**:3403-3413.

Joseph, S. B., and D. W. Hall. 2004. Spontaneos mutations in *Saccharomyces cerevisiae*: More beneficial than expected. *Genetics* **168**:1817-1825.

Karasov, T., P. W. Messer, and D. A. Petrov. 2010. Evidence that adaptation in *Drosophila* is not limited by mutations at single sites. *PLoS One,* doi: 10.1371/journal.pgen.1000924.

Karkkainen, K., and J. Agren. 2002. Genetic basis of trichome production in *Arabidopsis lyrata*. *Hereditas* **136**:219-226.

Katju, V., L. B. Packard, L. J. Bu, P. D. Keightley, and U. Bergthorsson. 2015. Fitness decline in spontaneous mutation accumulation lines of *Caenorhabditis elegans* with varying effective population size. *Evolution* **69**:104-116.

Kearney, M., W. P. Porter, C. Williams, S. Ritchie, and A. A. Hoffmann. 2009. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Func. Ecol.* **23**:528-538.

Keightley, P. D., E. K. Davies, A. D. Peters, and R. G. Shaw. 2000. Properties of ethylmethane sulfonate-induced mutations affecting life history traits in *Caenorhabditis elegans* and inferences about bivariate distributions of mutation effects. *Genetics* **156**:143-154.

Keightley, P.D., and A. Eyre-Walker. 2010. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Phil. Trans. Roy. Soc. B* **365**:1187-1193.

Keightley, P.D., and M. Lynch. 2003. Toward a realistic model of mutations affecting fitness. *Evolution* **57**:683-685.

Keightley, P.D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar, and M. L. Blaxter. 2009. Analysis of the genome sequence of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Gen. Res.* **19**:1195-1201.

Keller, I., and O. Seehausen. 2012. Thermal adaptation and ecological speciation. *Molec. Ecol.* **21**:782-799.

Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**:1193–1196.

Kim, H. S., and J. C. Fay. 2007. Genetic variation in the cysteine biosynthesis pathway causes sensitivity to pharmacological compounds. *Proc. Nat. Acad. Sci. USA* **104:**19387-19391.

Kim, S., V. Plagnol, T. T. Hu, C. Toomajian, R. M. Clark, S. Ossowski, J. E. Ecker, D. Weigel, and M. Nordbord. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Gen.* **39**:1151-1155.

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

Knight, C. G., N. Zitzmann, S. Prabhakar, R. Antrobus, R. Dwek, H. Hebestreit, and P. B. Rainey. 2006. Unraveling adaptive evolution: how a single point mutation affects the protein coregulation network. *Nat. Gen.* **38:**1015-1022.

Kondrashov, A. S., and D. Houle. 1994. Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proc. Roy. Soc. B* **258**:221-227.

Kondrashov, F. A., and A. S. Kondrashov. 2010. Measurements of spontaneous rates of mutations in the recent past and the near future. *Phil. Trans. Roy. Soc. B* **365**:1169-1176.

Korneef, M., M. L. Jorna, D. L. C. Brinkhorst-van der Swan, and C. M. Karssen. 1982. The isolation of abscisic acid (ABA) deficient mutants by selection of induced revertants in non-germinating gibberellin sensitive lines of *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* **61**:385-393.

Kryazhimskiy, S., Rice, D. P., Jerison, E. R., and M. M. Desai. 2014. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**:1519-1522.

Lasky, J. R., D. L. Des Marais, J. K. Mckay, J. H. Richards, T. E. Juenger, and T. H. Keitt, T.H. 2012. Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molec. Ecol.* **21**:5512-5529.

Lasky, J.R., D. L. Des Marais, D. B. Lowry, I. Provolotskaya, J. K. McKay, and J. H. Richards. 2014. Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. *Molec. Biol. Evol.* **31**:2283-2296.

Lande, R. 1994. Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**:1460-1469.

Latta, R. G., and K. M. Gardner. 2009. Natural Selection on Pleiotropic Quantitative Trait Loci Affecting a Life-History Trade-Off in *Avena Barbata*. *Evolution* **63:**2153-2163.

Latta, L.C., M. Peacock, D. J. Civitello, J. L. Dudycha, J. M. Meik, and S. Schaak. 2015. The phenotypic effects of spontaneous mutations in different environments. *Am. Nat.* **185**:243-252.

Lewis, D., and D. M. Hunt. 1984. Hans Gruneberg. *Biograph. Mem. Fellows. Roy. Soc.* **30:**227-247.

Li, R. H., S. W. Tsaih, K. Shockley, I. M. Stylianou, J. Wergedal, B. Paigen, G. A. Churchill. 2006. Structural model analysis of multiple quantitative traits. *PLOS Genetics* **2:**1046-1057.

Lortie, C.J., and L. W. Aarssen. 2000. Fitness consequences of branching in *Verbascum Thapsus* (Scrophulariaceae). *Am. J. Bot.* **87**:1793-1796.

Lowry, D. B., and J. H. Willis. 2010. A widespread chromosomal inversion polymorphism contributes to a major life history transition, local adaptation, and reproductive isolation. *PLOS Biology* **8**:1-14.

Mackay, T. F. C., and R. R. H. Anholt. 2006. Of flies and man: Drosophila as a model for human complex traits. *Ann. Rev. Genomics Hum. Gen.* **7:**339-367.

MacKenzie, J. L., F. E. Saade, Q. H. Le, T. E. Bureau, D. J. Schoen. 2005. Genomic mutation in lines of *Arabidopsis thaliana* exposed to ultraviolet-B radiation. *Genetics* **171**:715-723.

Maclean, R. C., G. Bell, and P. B. Rainey. 2004. The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. *Proc. Nat. Acad. Sci. USA* **101:**8072-8077.

MacLean, R. C., G. G. Perron, and A. Gardner. 2010. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* **186**:1345-54.

Mantel, N. 1967. Detection of disease clustering and a generalized regression approach. *Cancer Research* **27**:209-220.

Marks, M.D. 1997. Molecular genetic analysis of trichome development in *Arabidopsis*. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **48**:137-163.

Martin, G., and T. Lenormand. 2006a. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**:893-907.

Martin, G., and T. Lenormand. 2006b. The fitness effects of mutations across environments: A survey in light of fitness landscape. *Evolution* **60**: 2413-2427.

Martin, G., and T. Lenormand. 2008. The distribution of beneficial and fixed mutation fitness effects close to an optimum. *Genetics* **179**:907-916.

Maydanovych, O., and P. A. Beal. 2006. Breaking the central dogma by RNA editing. *Chem. Rev.* **106:**3397-3411.

Maynard Smith, J. 1966. Sympatric Speciation. *Am. Nat.* **100:**637-650.

Mayr, E. 1963. *Animal Species and Evolution*. The Bellknap Press of Harvard University Press, Cambridge, MA.

McKay, J. K., J. H. Richards, and T. Mitchell-Olds. 2003. Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. *Molec. Ecol.* **12**:1137-1151.

McKusick, V. A. 1976. Pleiotropism. *Am. J.Hum. Gen.* **28:**301-302.

Melnyk, A. H., and R. Kassen. 2011. Adaptive landscapes in evolving populations of *Pseudomonas fluorescens*. *Evolution* **65**:3048-3059.

Mendel, J. G. 1866. Versuche uber Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brunn* **4:**3-47.

Moorad, J. A., and D. E. L. Promislow. 2009. What can genetic variation tell us about the evolution of senescence? *Proc. Roy. Soc. B-Biol. Sci*. **276:**2271-2278.

Mukai, T. 1964. Genetic structure of natural populations of *Drosophila melanogaster* 1: Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**:1-19.

Muller, H. J. 1950. Our load of mutations. *Am. J. Hum. Gen.* **2**:111-176.

New, M., D. Lister, M. Hulme, and I. Makin. 2002. A high-resolution data set of surface climate over global land areas. *Climate Res.* **21**:1-25.

Nicotra, A. B., O. K. Atkin, S. P. Bonser, A. M. Davidson, E. J. Finnegan, U. Mathesius, P. Poot, M. D. Puruggnnan, C. L. Richards, F. Valladares, and M. van Kleunen. 2010. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* **15**:684-692.

Nordborg, M., and J. Bergelson. 1999. The effect of seed and rosette cold treatment on germination and flowering time in some *Arabidopsis thaliana* (Brassicaceae) ecotypes. *Am. J. Bot.* **86**:470-475.

Nordborg, M., J. O. Borevitz, J. Bergelson, C. C. Berry, J. Chory, J. Hagenblad, M. Kreitman, J. N. Maloof, T. Noyes, P. J. Oefner, E. A. Stahl, and D. Weigel. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Gen.* **30**:190-193.

Nordborg, M., T. T. Hu, Y. Ishino, J Jhaveri, C. Toomajian, H. Zheng, E. Bakker, P. Calabrese, J. Gladstone, R. Goyal, M. Jakobsson, S. Kim, Y. Morozov, P. Padhukasahasram, V. Plagnol, N. A. Rosenberg, C. Shah, J. D. Wall, J. Wang, K. Zhao, T. Kalbfleisch, V. Schulz, M. Kreitman, and J. Bergelson. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLOS Biol.* **3**:289-1299.

Oksanen J., F. Guillaume Blanchet, R. Kindt, L. Pierre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2011. vegan: Community Ecology Package. R package version 2.0-1. Retrieved from: http://CRAN.R-project.org/package=vegan

Olsen, E. M., M. Heino, G. R. Lilly, M. J. Morgan, J. Brattey, B. Ernande, and U. Dieckmann. 2004. Maturation trends indicative of rapid evolution preceded the collapse of northern cod. *Nature* **428**:932-935.

Orr, H. A.1998. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**:935-949.

Orr, H. A. 2000. Adaptation and the cost of complexity. *Evolution* **54**:13-20.

Orr, H. A. 2003. The distribution of fitness effects among beneficial mutations. *Genetics* **163**:1519-1526.

Orr, H. A. 2005a. The probability of parallel evolution. *Evolution* **59**:216-220.

Orr, H. A. 2005b. The genetic theory of adaptation: A brief history. *Nat. Rev. Gen*. **6**:119-127.

Ossowski, S., K. Schneeberger, J. I. Lucas-Lledo, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, and M. Lynch. 2010. Rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**:92-94.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. *Ann. Rev. Ecol. Evol. System.* **37**:637-669.

Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**:37-42.

Perfeito, L., L. Fernandes, C. Mota, and I. Gordo. 2007. Adaptive mutations in bacteria: High rate and small effects. *Science* **317**:813-815.

Perfeito, L., A. Sousa, T. Bataillon, and I. Gordo. 2014. Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution* **68**:150-162.

Plate, L. 1910. Vererbungslehre und Deszendenztheorie, pp. 537 in *Festschrift zum sechzigsten Geburtstag Richard Hertwigs*. Fischer, Jena.

Provine, W. B. 1971. *On the origin of theoretical population genetics*. University of Chicago Press, Chicago, Il.

Puigbo, P., S. Garcia-Vallve, and J. O. McInterney. 2007. TOPD/FMTS: A new software to compare phylogenetic trees. *Bioinformatics* **23**:1556-1558.

Pyeritz, R. E. 1989. Pleiotropy Revisited - Molecular Explanations of a Classic Concept. *Am. J. Med. Gen.* **34:**124-134.

R Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: http://www.R-project.org

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: http://www.R-project.org/

Reddy, A. S. N. 2007. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Ann. Rev. Plant Biol.* **58:**267-294.

Redei, G. P., and C. Koncz. 1992. Classical mutagenesis. Pp. 16-82 *in* C. Koncz, N-H Chua, and J. Schell, eds. *Methods in* Arabidopsis *Research*, World Scientic, Singapore.

Reusch, T. B. H., and T. E. Wood. 2007. Molecular ecology of global change. *Mol. Ecol.* **16:**3973-3992.

Rieger, R., A. Michaelis, and M. M. Green. 1981. *Glossary of Genetics and Cytogenetics*. Springer-Verlag, Berlin.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Math. Biosci.* **53**:131-147.

Rokyta, D. R., P. Joyce, S. B. Caudle, and H. A. Wichman. 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. *Nat. Gen.* **37**:441-444.

Rutter, M. T., and C. B. Fenster. 2007. Testing for adaptation to climate in *Arabidopsis thaliana*: A calibrated common garden approach. *Ann. Bot.* **99**:529-536.

Rutter, M. T., F. H. Shaw, and C. B. Fenster. 2010. Spontaneous mutation parameters for *Arabidopsis thaliana* measured in the wild. *Evolution* **64**:1825-1835.

Rutter, M. A., J. Roles, R. Conner, R. G. Shaw, F. H. Shaw, K. Schneeberger, S. Ossowski, D. Weigel, And C. B. Fenster. 2012. Brief Communication: Fitness of *Arabidopsis thaliana* mutation accumulation lines whose spontaneous mutations are known. *Evolution* **66**:2335-2339.

Salinas, J., and J. J. Sanchez-Serrano. 2006. *Arabidopsis Protocols*, 2nd edn. Humana Press, NY.

Samis, K. E., C. J. Murren, O. Bossdorf, K. Donohue, C. B. Fenster, R. L. Malmberg, M. D. Purugganan, and J. R. Stinchcombe. 2012. Longitudinal trends in climate drive flowering time clines in North American *Arabidopsis thaliana*. *Ecol. & Evol*. **2**:1162-1180.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, and M. Smith. 1977. Nucleotide-Sequence of Bacteriophage Phichi174 DNA. *Nature* **265:**687-695.

SAS Institute. 2004. *SAS/STAT, Version 8.1*. SAS Institute, Cary NC.

Seetharam, S., and K. Jain. 2014. Adaptive walks and distribution of beneficial fitness effects. *Evolution* **68**:965-975.

Shaw, R. G., D. L. Byers, and E. Darmo. 2000. Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics* **155**:369-378.

Shindo, C., M. J. Aranzana, C. Lister, C. Baxter, C. Nicholls, M. Nordborg, and C. Dean. 2005. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time in *Arabidopsis*. *Plant Physiol*. **138**:1163-1173.

Shindo, C., G. Bernasconi, and C. S. Hardtke. 2007. Natural genetic variation in *Arabidopsis*: Tools, traits and prospects for evolutionary ecology. *Ann. Bot.* **99**:1043-1054.

Singer, B., and J. T. Kusmierek. 1982. Chemical Mutagenesis. *Ann. Rev. Biochem*. **51**:655-693.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple-regression and correlation extensions of the Mantel test of matrix correspondence. *System. Zool.* **35**:627-632.

Sokal, R. R., and F. J. Rohlf. 1995. *Biometry*. W. H. Freeman and Company, NY.

Sousa, A., S. Magalhaes, and I. Gordo. 2012. Cost of antibiotic resistance and the geometry of adaptation. *Mol. Biol. Evol.* **29**:1417-1428.

Stearns, F. W. 2010. One hundred years of pleiotropy: A retrospective. *Genetics* **186**:767-773.

Stearns, F. W., and C. B. Fenster. 2013. Evidence for parallel adaptation to climate across the natural range of *Arabidopsis thaliana*. *Ecol. & Evol.* **3**:2241-2247.

Su, Z., Y. Zeng, and X. Gu. 2009. A preliminary analysis of gene pleiotropy estimated from protein sequences. *J. Exp. Zool*. **312B:**1-10.

Symonds, V.V., A. V. Godoy, T. Alconada, J. F. Botto, T. E. Juenger, J. J. Casal, and A. M. Lloyd. 2005. Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. *Genetics* **169**:1649-1658.

Tauber, C. A., and M. J. Tauber. 1989. Sympatric speciation in insects: Perceptions and perspective. Pp. 307-344 *in* D. Otte and J. A. Endler, eds. *Speciation and Its Consequences*, Sinauer Associates, Inc., Sunderland, MA.

Teplitsky, C., J. A. Mills, J. S. Alho, J. W. Yarrall, and J. Merila. 2008. Bergmann's rule and climate change revisited: disentangling environmental and genetic responses in a wild bird population. *Proc. Nat. Acad. Sci. USA* **105**:13492-13496.

Trindade, S., A. Sousa, and I. Gordo. 2012. Antibiotic resistance and stress in the light of Fisher's model. *Evolution* **66**:3815-3824.

Vale, P. F., M. Choisy, R. Froissart, R. Sanjuan, and S. Gandon. 2012. The distribution of mutational fitness effects of phage **ϕ**X174 on different hosts. *Evolution* **66**:3495-3507.

Vander Zwan, C., S. A. Brodie, and J. J. Campanella. 2000. The intraspecific phylogenetics of
Arabidopsis thaliana in worldwide populations. *System. Bot.* **25**:47-59.

van Harten, A.M. 1998. *Mutation Breeding: Theory and Practical Applications*. Cambridge University Press, Cambridge, UK.

Via, S., and D. J. Hawthorne. 1998. The genetics of speciation: Promises and prospects of quantitative trait locus mapping. *In* D. J. Howard, S. H. Berlocher, eds. *Endless Forms: Species and Speciation,* Oxford University Press, New York, NY

Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman, and J. M. Cheverud. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature* **452:**470-U479.

Walther, G-R. 2003. Plants in a warmer world. *Perspec. Plant Ecol. Evol. System.* **6**:169-185.

Waxman, D., and J. R. Peck. 1998. Pleiotropy and the preservation of perfection. *Science* **279**:1210-1213.

Weber, J., W. Jelinek, and J. E. Darnell. 1977. Definition of a large viral transcription unit late in Ad2 infection of HeLa-cells: Mapping of nascent RNA molecules labeled in isolated nuclei. *Cell* **10**:611-616.

Weigel, D., and J. Glazebrook. 2002. *Arabidopsis: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Weil, U. H. 1981. Osteogenesis Imperfecta - Historical Background. *Clin. Orthopaedics Rel. Res*.**:** 6-10.

Welch, J. J., and D. Waxman. 2003. Modularity and the cost of complexity. *Evolution* **57:**1723-1734.

Westerman, J. M. 1971. Genotype-environment interaction and developmental regulation in *Arabidopsis thaliana*. II. Inbred lines: Analysis. *Heredity* **26**:93-106.

Whitlock, M. C., and P. C. Phillips. 2000. The equisite corpse: A shifting view of the shifting balance. *Trends Eco. Evol.* **15**:347-348

Whitlock, M.C., and D. Schluter. 2009. *The Analysis of Biological Data*. Roberts and Company Publishers, Greenwood Village, CO

Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne, and J. J. Bull. 1999. Different trajectories of parallel evolution during viral adaptation. *Science* **285**:422-424.

Wilczek, A. M., J. L. Roe, M. C. Knapp, M. D. Cooper, C. Lopez-Gallego, L. J. Martin, L.J., C. D. Muir, S. Sim, A. Walker, J. Anderson, J. F. Egan, B. T. Moyers, R. Petipas, A. Giakountis, E. Charbit, G. Coupland, S. M. Welch, and J. Schmitt. 2009. Effects of genetic perturbation on seasonal life history plasticity. *Science* **323**:930-934

Wilkins, J. F. 2010. Antagonistic Coevolution of Two Imprinted Loci with Pleiotropic Effects. *Evolution* **64:**142-151.

Wilkins, A. S. 1993. *Genetic Analysis of Animal Development*. Wiley-Liss, New York.

Williams, G. C., 1957 Pleiotropy, Natural-Selection, and the Evolution of Senescence. *Evolution* **11:**398-411.

Wiser, M. J., N. Ribeck, and R. E. Lenski. 2013. Long-term dynamics of adaptation in asexual populations. *Science* **342**:1364-1367.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97-159.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection on evolution. *Proc. Sixth Int. Cong. Genet.* **1**:356-366.

Wright, S. 1968. *Genetic and Biometric Foundations*. The University of Chicago Press, Chicago, IL.

Zhang, M., P. Azad, and R. C. Woodruff. 2011. Adaptation of *Drosophila melanogaster* to increased NaCl concentration due to dominant beneficial mutations. *Genetica* **139**:177-186.

Zuellig, M. P., A. M. Kenney, and A. L. Sweigart. 2014. Evolutionary genetics of plant adaptation: insights from new model systems. *Curr. Opin. Plant Biol.* **18**:44-50.

Zwaan, B. J. 1999. The evolutionary genetics of ageing and longevity. *Heredity* **82:**589-597.