DISSERTATION

CAUSES AND CONSEQUENCES OF PLANT CLIMATE ADAPTATION

Submitted by

John Grey Monroe

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2019

Doctoral Committee:

Advisor: John McKay

Ruth Hufbauer
Cameron Ghalambor
Dan Sloan
Dave Des Marais

ABSTRACT

CAUSES AND CONSEQUENCES OF PLANT CLIMATE ADAPTATION

Climatic conditions such as temperature and drought can sources of strong selection on natural populations. In plants, whose sessile nature forces them to adapt to local climate conditions, extensive evidence of local adaptation has been observed. However, the consequences of this adaptation on ecosystem processes such as carbon cycling remain poorly understood. Additionally, the molecular basis of adaptation is often unresolved and the specific climatic factors that drive adaptive evolution unclear.

Addressing these knowledge gaps has become increasingly urgent as climate change threatens to rapidly alter selection regimes. Fortunately, conceptual and technical advances provide new opportunities to characterize and integrate environments, phenotypes, and genes, and thus advance our understanding of the causes and consequences of climate adaptation.

In Chapter 2 of this dissertation, I consider the consequences of climate adaptation through the lens of ecoevolutionary dynamics. Integrating environments and phenotypes by considering ecosystem impacts of adaptive evolution, I review empirical evidence that contemporary climate adaptation could significantly alter the carbon cycle.

In Chapter 3, I investigate the molecular basis of adaptation to winter temperatures in the model plant *Arabidopsis thaliana* by integrating genes and environments through the framework of landscape and population genetics. Specifically, I address the hypothesis that loss-of-function in a family of transcription factors contributes to adaptation to warmer climates.

In Chapter 4, I develop methods combining whole genome sequence data, long term remote sensing, and reverse genetics to study drought as an agent of selection on flowering time and identify loss-of-function variants contributing to this evolution in *Arabidopsis thaliana*.

Together, this work has inspired my interest in combining conceptual, computational, experimental innovations into an integrated research program to understand climate adaptation.

ACKNOWLEDGMENTS

There are many people whose support and contributions made this dissertation possible. First, I could not have asked for a better advisor. To John McKay, I am thankful for the generosity, intellectual challenges, and patience that made my time as a graduate student a great experience and taught me what excellent mentorship looks like. To all of the members of the McKay lab, I thankful for making science so much fun and for their innumerable contributions to this research. To other mentors during graduate school, particularly Dan Sloan, Ruth Hufbauer, Cameron Ghalambor, Amelia Henry and Dave Des Marais I am grateful for welcoming me as an adopted member of their labs in various ways over the years. My experiences with each of these groups have had tremendously positive impacts on my development as a scientist. To my previous mentor, Matt Estep, I owe a great deal for inspiring me to pursue scientific research. To my fellow graduate students, especially those who contributed to the work on ecoevolutionary dynamics of carbon cycling, I am grateful for amazing friendships and support. To my family, I am thankful for unending encouragement to pursue my passions. I dedicate this dissertation to my dad, who taught me to love nature.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Background

Plant climate adaptation: causes and consequences

Climate adaptation is a pervasive feature of natural populations (Clausen *et al*. 1940, Hancock *et al.* 2010). Adaptation to different climate regimes leads to genetic and phenotypic divergence between populations from environments that differ in components of climate such as temperature and drought. This has been conclusively demonstrated in cases where reciprocal transplant experiments have been performed, and evidence of local adaptation in the form of genotype x environment interactions for fitness have been observed (Ågren & Schemske 2012, Anderson *et al.* 2013). Such findings have been particularly compelling for plant species, whose sessile nature subjects them to local climate conditions throughout the year and makes them amenable for study (Harper 1977).

The need to understand climate adaptation of natural populations has grown with the revelation that Earth's climates are rapidly changing, largely because of anthropogenic increases in greenhouse gases (Pachauri *et al*. 2014). Such changes pose existential threats to the stability of agricultural production and the persistence of natural populations (Thomas *et al.* 2004, Müller *et al.* 2010). However, by understanding the causes of climate adaptation, that is, the climatic factors that explain selection and the phenotypic and molecular basis of responses to that selection, solutions to the challenges posed by climate change can be pursued. For example, phenotypes and alleles responsible for climate adaptation in natural species can inform breeders working to mimic natural climate adaptation strategies in crops. Furthermore, by characterizing the climatic factors that explain adaptation, predictions can be made to match crops with target

environments. The research in this dissertation, particularly Chapters 2 and 3, reflects a belief that understanding the causes of climate adaptation in natural plant populations can address fundamental questions while also yielding insights to solve practical problems.

Because climate is itself largely the product of ecosystem processes such as photosynthesis and respiration (IPCC 2013), it is also important to understand the ecosystem consequences of climate adaptation. Darwin introduced the concept of ecoevolutionary dynamics when he considered how organisms alters their environment (Darwin 1881). If this perspective is extended to climate adaptation, the need to consider how evolutionary responses might in turn alter climate becomes apparent if the dynamics of climate adaptation are to be fully understood. Indeed, evolutionary processes have been cited as the cause of variation in the Earth's climate has throughout geological history (Beerling & Berner 2005). Today we are faced with the task of predicting the future climates of a rapidly changing planet. My dissertation research, particularly Chapter 1, is motivated by the belief that evolutionary processes should not be overlooked in this endeavor.

Integrating environments, phenotypes, and genetics

Climate adaptation is a complex process involving environments, phenotypes, and genetics. Thus, the study of climate adaptation calls for integrating these components. For example, a central thesis of local adaptation is that differences in environments drive differences in selection pressure on phenotypes and causal genes. As such, the study of local adaptation often includes investigating the relationships between environmental parameters and phenotypes, as well as environment and genotypes (Kawecki & Ebert 2004). The concept of ecoevolutionary dynamics also calls for integrating environments, phenotypes, and genes through the notion that heritable phenotypic variation alters the environment (Pelletier *et al*. 2009). While local adaptation and ecoevolutionary dynamics represent conceptual tools to integrate environments,

phenotypes and genes, technical advances also present new opportunities to better understand climate adaptation.

Emerging technological tools provide new ways to study the components of climate adaptation: environments, phenotypes, and genes. For example, while climates have often been studied using weather station data, satellite based remote sensing presents new tools to characterize variation in climate across landscape at improved spatial and temporal scales (AghaKouchak *et al.* 2015). Similarly, high throughput phenotyping methods have been developed to replace manual measurement methods (Araus & Cairns 2014). And for genes, the age of affordable DNA sequencing and editing has dramatically changed the scale and molecular resolution at which causal variation can be studied (Mardis 2008). When combined with the conceptual tools to integrate environments, phenotypes and genes that date back to Darwin and the modern synthesis, the opportunities to advance our understanding of climate adaptation have never been more exciting.

Summary of subsequent dissertation chapters

Chapter 2: Eco-evolutionary Dynamics of Carbon Cycling in the Anthropocene

In this chapter, which was a collaborative project I led with several other graduate students, I examine the consequences of climate adaptation on the carbon cycle. By considering the potential for such changes to alter atmospheric carbon dioxide, recent empirical work is reviewed through the framework of ecoevolutionary feedbacks. The resulting paper was published in Trends in Ecology and Evolution (Monroe *et al.* 2018a).

Chapter 3: Adaptation to warmer climates by parallel functional evolution of CBF genes in Arabidopsis thaliana

In this chapter, I address the hypothesis that evolution of a family of cold tolerance transcription factors, CBFs, is involved in adaptation to warmer climates in *Arabidopsis*

thaliana. Consistent with previous work suggesting that CBF2 has experienced locally adaptive loss-of-function in southern climates, this work revealed a significant excess of non-synonymous, missense, and frameshift mutations in populations from climates with warmer temperatures. The resulting paper was published in Molecular Ecology (Monroe *et al.* 2016).

Chapter 4: Drought adaptation in Arabidopsis thaliana by extensive genetic loss-of-function

In this chapter, I develop new approaches using whole genome sequences, satellite detected drought, and transgenic experiments to identify loss-of-function alleles involved in drought adaptation in *Arabidopsis thaliana*. The results provide new insights into the importance of seasonal timing for characterizing drought as a climatic agent of selection. They also support the hypothesis that loss-of-function can be a common genetic mechanism of climate adaptation. The resulting paper was published in eLife (Monroe *et al* 2018b).

REFERENCES

AghaKouchak A, Farahmand A, Melton FS, Teixeira J, Anderson MC, Wardlow BD, Hain CR (2015) Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics*, *53*(2), 452-480.

Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. *New Phytologist*, *194*(4), 1112-1122.

Anderson JT, LEE CR, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, *22*(3), 699-708.

Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends in plant science*, *19*(1), 52-61.

Beerling DJ, Berner RA (2005) Feedbacks and the coevolution of plants and atmospheric CO2. *Proceedings of the National Academy of Sciences*, *102*(5), 1302-1305.

Clausen JC, Keck DD, Hiesey WM (1940) Effect of varied environments on western North American plants Carnegie Institution.

Darwin C (1881) The Formation of vegetable mold through the action of worms, with observations on their habits, John Murray.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Bergelson J (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, *334*(6052), 83-86.

IPCC (2013) Climate change 2013: the physical science basis. In Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Stocker, T.F. et al., eds), pp. 33–118, Cambridge University Press.

Kawecki TJ, Ebert D. (2004). Conceptual issues in local adaptation. *Ecology letters*, *7*(12), 1225-1241.

Mardis ER, Elaine R (2008) The impact of next-generation sequencing technology on genetics. *Trends in genetics,* 24 (3), 133-141.

Monroe JG, Markman DW, Beck WS, Felton AJ, Vahsen ML, Pressler Y (2018) Ecoevolutionary dynamics of carbon cycling in the anthropocene. *Trends in ecology & evolution*, *33*(3), 213-225.

Monroe JG, McGovern C, Lasky JR, Grogan K, Beck J, McKay JK (2016) Adaptation to warmer climates by parallel functional evolution of CBF genes in *Arabidopsis thaliana*. *Molecular ecology*, *25*(15), 3632-3644.

Monroe JG, Powell T, Price N, Mullen JL, Howard A, Evans K, McKay JK (2018) Drought adaptation in *Arabidopsis thaliana* by extensive genetic loss-of-function. *eLife*, *7*, e41038.

Müller C, Bondeau A, Popp A, Waha K, Fader M (2010) Climate change impacts on agricultural yields.

Pachauri RK, Allen MR, Barros VR, Broome J, Cramer W, Christ R, Dubash NK (2014) *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change* (p. 151). IPCC.

Pelletier F, Garant D, Hendry AP (2009). Eco-evolutionary dynamics.

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Hughes L (2004) Extinction risk from climate change. *Nature*, *427*(6970), 145.

CHAPTER 2

ECOEVOLUTIONARY DYNAMICS OF CARBON CYCLING IN THE ANTHROPOCENE

Summary

Climate change is altering natural selection globally, which could shift the evolutionary trajectories of traits central to the carbon (C) cycle. Here we examine the components necessary for the evolution of C cycling traits to substantially influence global C cycling and integrate these components through the framework of eco-evolutionary feedback loops. Recent evidence points to the potential for adaptation in the Anthropocene to lead to significant positive and negative feedback loops affecting atmospheric CO_2. We identify directions for further collaboration between evolutionary, ecosystem, and climate scientists to study these feedbacks and determine whether this evolution will ultimately accelerate or decelerate the current trend in rising atmospheric CO_2.

"I don't want to be a product of my environment. I want my environment to be a product of me." – Frank Costello, The Departed.

Evolution and C cycling on a changing planet

The Earth has entered a new geological epoch, the Anthropocene, marked by rapidly increasing atmospheric CO_2 concentrations and changing environments (Crutzen 2006, IPCC 2013). For many species, these changes will impose strong natural selection and persistence will require adaptation, directly or indirectly affecting the evolution of functional traits - those that characterize an organism's ecological role and effect on its environment [Hoffmann 2011, Shaw 2012, Franks & Hoffmann 2012, Merilä 2016, de Bello). Understanding these evolutionary trajectories is critical, as there is accumulating evidence that trait evolution in response to climate

change could alter key ecosystem processes, including the carbon (C) cycle (De Deyn 2008, Bailey 2009, Matthews 2011, Rudman 2017, Pastor 2017). Describing how the C cycle is affected by trait evolution is particularly important because organism traits ultimately influence whether C is stored or released into the atmosphere (Figure 1, Figure 2). In addition to asking, "*will species persist via evolutionary adaptation to changing environments?"* it is also imperative to ask, *"how will organism C cycling traits evolve once species adapt?"*

Here we examine the potential for contemporary evolution to impact global C cycling and storage through shifts in C cycling traits. We first examine how C cycling is mediated by organism traits and evolution. We then discuss natural selection pressures resulting from climate change and the extent of genetic variation for C cycling traits on which this selection can act. Next, we review potential for rapid evolution of these traits to occur during the Anthropocene. Finally, we review evidence and approaches to study the direction and magnitude of contemporary evolution on C cycling traits and impact on the global C cycle. The body of work presented here reveals progress toward a synthesis between ecosystem science and evolutionary biology and suggests valuable directions for further research (Box 1). We propose that the evolution of C cycling traits presents a particularly important eco-evolutionary feedback system with implications for climate change dynamics (Figure 3) that provides a useful framework to study future global C cycling and storage (Box 2).

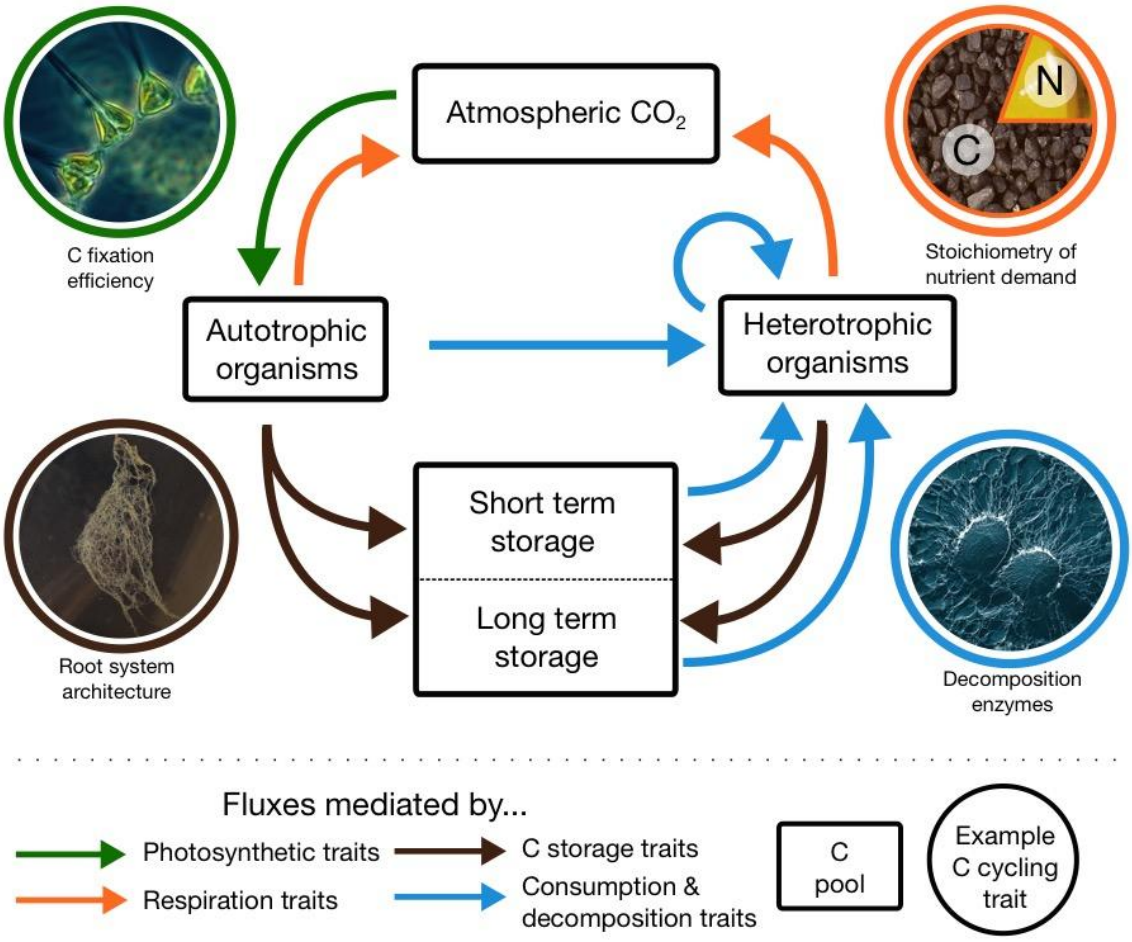

Figure 1. Trait mediated processes in the C cycle. Organisms are responsible for the largest fluxes (arrows) of C between biotic and abiotic pools (boxes) in the global C cycle (IPCC 2013). These fluxes are trait-mediated and can be influenced by photosynthesis, respiration, carbon storage, or consumption and decomposition traits (arrows). Trait-based ecosystem perspectives have revealed the integral relationship between specific organism traits (circles) and the fluxes of C. By studying how these C cycling traits will evolve in response to climate change, we can consider how adaptation might affect carbon cycling and future atmospheric CO_2 concentrations.

Box 1: Progress and future objectives

Recent research has led to significant progress in understanding the potential for eco-evolutionary feedbacks involving C cycling. Here we outline the important components that make such feedbacks possible. For each, we highlight a statement that has received emerging empirical support (bolded) and a challenge where future research is needed.

1. **Organism traits mediate C fluxes.**

   Challenge: It is critical that the relationships between specific traits, ecosystem C fluxes and ultimately atmospheric CO_2 be described quantitatively.

2. **Climate change will impact selection pressures.**

   Challenge: More research is needed to predict how different components of climate change (eg. drought and temperature) will interact to alter evolutionary trajectories of C cycling traits.

3. **Organisms harbor significant genetic variation for C cycling traits.**

   Challenge: Genetic variation in C cycling traits has not been explicitly quantified in many organisms, leaving a gap in our ability to predict evolutionary responses.

4. **Evolution can occur at rapid timescales.**

   Challenge: More studies are needed to understand the limits of this evolution (e.g. in organisms with long generation times or small populations). Studies that go beyond demonstrating that rapid evolution *can* occur toward determining whether rapid evolution most often *does* occur will strengthen our ability to consider eco-evolutionary feedbacks in the Anthropocene global C cycle.

5. **Ongoing environmental changes can lead to rapid evolution of C cycling traits.**

   Challenge: More research is needed to draw generalizable conclusions about the direction of this evolution across species and environments, and how evolution within native populations and communities can scale to impacts on ecosystem-level carbon cycling. Disparate evidence suggests criteria 1-4 exist in nature, yet studies that integrate these components are rare. Additionally, determining the sign and magnitude of evolution on global C cycling and atmospheric CO_2 remains a pressing research objective at the intersection of modern evolutionary biology and ecosystem science.

Box 2: Tools toward a synthesis

Understanding the eco-evolutionary dynamics of C cycling will require the integration of evolutionary biology, ecosystem ecology and climate science (Figure 3). Fortunately each of these disciplines is largely prepared with the tools needed to achieve such a synthesis.

Evolutionary biology

Evolutionary biology is equipped with the tools to quantify the capacity for evolutionary change and direction of selection on C cycling traits. Quantitative genetics delivers a robust statistical framework to accomplish this, wherein measures of genetic variation (i.e. V_a, and h^2) and the strength (S) and response (R) to selection on C cycling traits can be estimated (Shaw 2012, Hattich, et al. 2017, Kumar, et al. 2017). Evolutionary responses of C cycling traits can also be directly observed through experiments or by comparison of populations diverged through time or along chronosequences. Emerging tools such as landscape genomics and genomic prediction (Gienapp, et al. 2017) of C cycling traits might also prove useful in predicting evolutionary responses of C cycling traits (Darwin 1881). Estimating the direction and magnitude of rapid evolutionary responses in these traits and consequences for ecosystem processes remains an important challenge to parameterize C cycling models.

Ecosystem ecology

Ecosystem ecology provides models that describe the relationships between organisms and environments through processes such as the cycling of nutrients between pools and fluxes, and is well suited for trait-based approaches. Key predictions of C cycling models include processes such as net ecosystem productivity (*NEP*), which is fundamentally defined as the difference between total photosynthesis (*GPP*) and ecosystem respiration (*ER*) (Cramer, et al. 1999), two processes that are driven by a suite of traits. Progress in trait-based ecosystem ecology has led to large-scale community and ecosystem models of the C cycle that include more

specific, typically static trait values. Yet as trait-based approaches continue to develop, movement away from static trait values to those that can change as a result of evolutionary processes can add realism, and even constrain model predictions with more realistic trait values. Additional community and ecosystem models including experimentally-derived expected trait value shifts could help inform large-scale earth system models, allowing better tracking of the magnitude and direction of C eco-evolutionary feedback loops on atmospheric CO_2.

Climate science

Climate science, while not the focus of this paper, has considerably advanced in recent years in the ability to model how environmental conditions will change as a function of atmospheric CO_2. These models have moved well beyond predicted increases in mean temperatures, and are able to model changes in climatic variability and extreme events [e.g. (Nagy, et al. (2015)], providing improve accuracy and important dimensions by which future selective environments can be characterized.

The organism-mediated C cycle

Earth's C cycle is dominated by organism-mediated processes (Cornwell, et al. 2008, Allison 2012, Litchman, et al. 2015) (Figure 1). Photosynthesis and respiration are the largest fluxes in the global C cycle (IPCC 2013, Beer, et al. 2010) with photosynthesis moving 20 times more C than all anthropogenic sources combined (Figure 2). If all other fluxes were held constant, an increase or decrease in global photosynthesis by just 2% would either completely offset or double the current rate of rising atmospheric CO_2 (IPCC 2013). Therefore, the global C cycle would be significantly affected changes in photosynthesis and/or respiration, which are affected by traits such as C fixation efficiency (Bradford 2013, Paulus, et al. 2013, Allison 2014),

growth rate (Scheinin, et al. 2015), nutrient stoichiometry (Sterner & Elser 2002, Schmitz 2013) and metabolism (Braakman, et al. 2015).

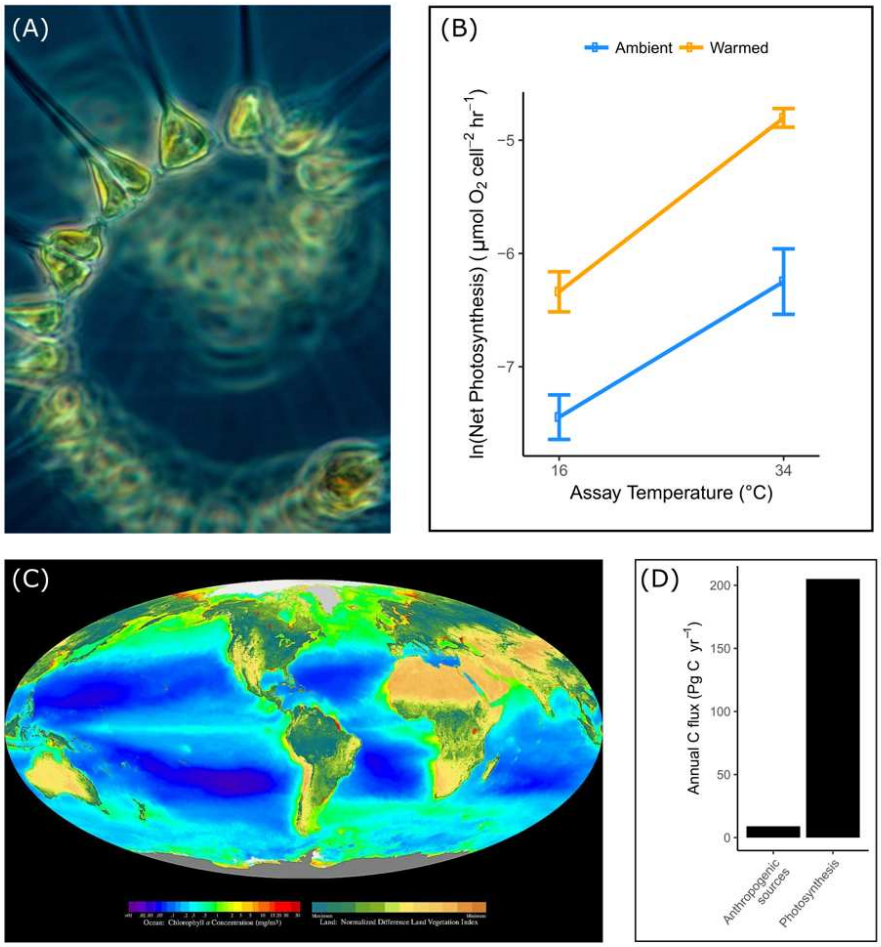

Figure 2. (A) Algae and other phytoplankton contribute to approximately 40 per cent of global photosynthesis (Falkowski 2014) and provide useful models for studying the evolution of C cycling traits in response to climate change related selection pressures experimentally (image courtesy of NOAA). (B) A decade long mesocosm experiment found that under increased temperatures, populations of the algae *Chlamydomonas reinhardtii* evolved 3.5 fold greater net photosynthesis compared to populations evolved under ambient temperatures (data from (Schaum, et al. 2017). (C) Abundances of aquatic and terrestrial photosynthetic organisms as observed from space (image courtesy of NASA). (D) Photosynthesis is the largest flux in the global C cycle, moving more than 20 times the C released by anthropogenic sources (data based on (IPCC 2013 and Allison 2012)). The potential impact of even minor changes in photosynthesis or other organism mediated C cycling fluxes is great. Earth system models that parameterize genotypic differentiation in C cycling traits might lead to better predictions of changing atmospheric CO_2 concentrations.

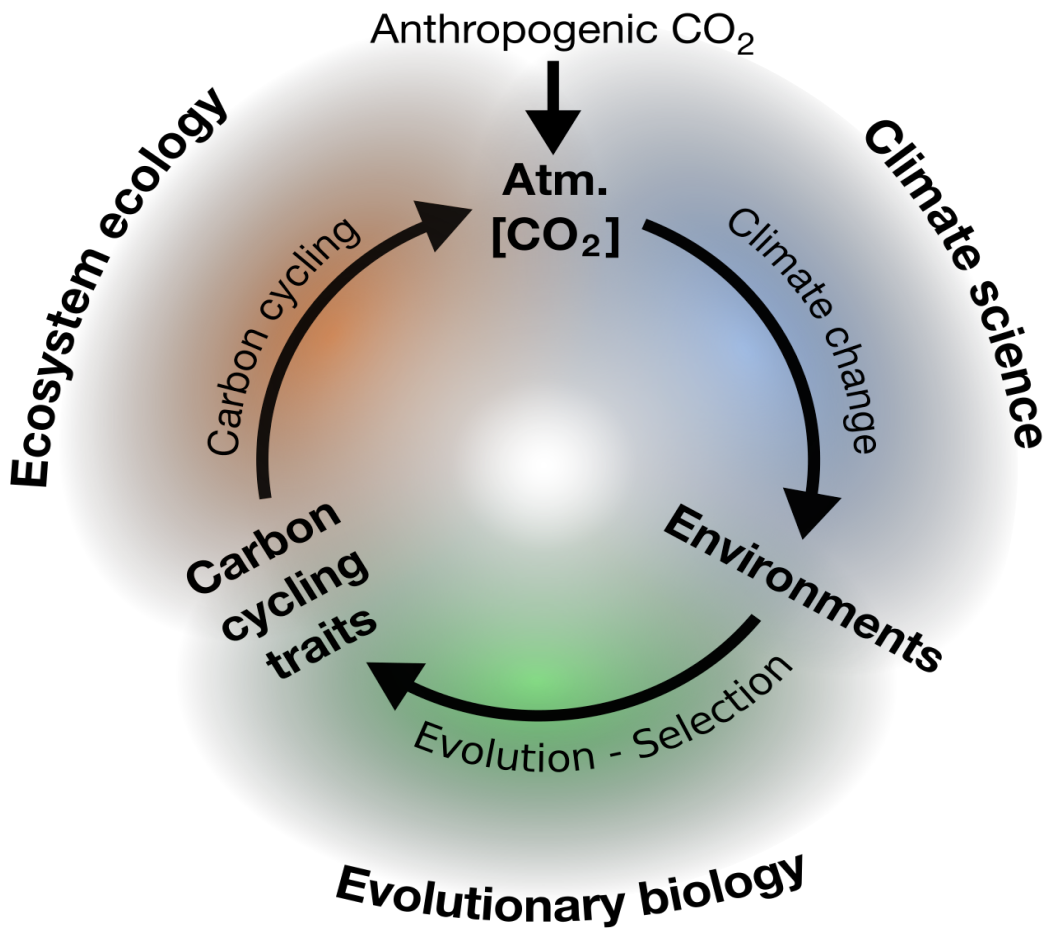

Figure 3. An eco-evolutionary feedback loop perpetuated by the evolution of C cycling traits in response to climate change. Climate change caused by anthropogenic CO_2 increases has changed biotic and abiotic environments and thus altered natural selection experienced by many organisms at a global scale, potentially shifting the evolutionary trajectories of C cycling traits. Evolution of these C cycling traits might in turn lead to changes in atmospheric CO_2 concentrations, which would impact climate change and alter natural selection pressures. Thus, positive feedback loops will arise when trait evolution exacerbates the current trend toward increased atmospheric CO_2 concentrations, whereas negative feedback loops would occur when trait evolution leads to increased C storage. Understanding the eco-evolutionary dynamics of C cycling will require the integration of climate science, evolutionary biology, and ecosystem ecology.

The concentration of atmospheric CO_2 is the net result of C cycling processes happening simultaneously so if photosynthesis and respiration evolve similarly, the cumulative effect could be zero. However if photosynthesis and respiration evolve at different rates the effects could be substantial. A recent study on *Chlamydomonas reinhardtii* (a member of the group of aquatic

atrophic organisms known as phytoplankton that are responsible for almost half of global photosynthesis (Falkowski 1994) has shed some light on this issue. During a decade long outdoor mesocosm experiment under increased temperatures, *C. reinhardtii* evolved a 3.5 fold increase in net photosynthesis, the difference between photosynthesis and respiration (Figure 2) (Schaum, et al. 2017). This experiment suggests trait evolution resulting from adaptation to warming climates to increase photosynthesis relative to respiration, potentially leading to a net decrease in atmospheric CO_2 and a negative feedback (Figure 3).

In addition to directly moving C to and from the atmosphere, organism traits affect C accumulation in short (months - decades) and long-term (decades-millennia) storage. Soils hold approximately 2 to 3 times more C than the atmosphere, and C pools in many soils are considered unsaturated (IPCC 2013, Kell 2012). Plant root traits, such as depth, influence the deposition of C in deep soils (Kell 2012, Kell 2011, Laliberté 2016) and it was recently estimated that a 1 m increase in plant root depth across just 3.9% of arable land would completely offset all CO_2 produced by annual fossil fuel emissions (or vice versa) (Kell 2011), indicating the potential for the evolution of root length to produce a negative or positive feedback with atmospheric CO_2.

Similarly, deposition of biomass to the deep ocean can result in long-term C storage. A recent survey of C fluxes to subsurface ocean layers observed an average flux of 47.5 mg particulate organic and inorganic C m^{-2} d^{-1} (Durkin, et al. 2016). If roughly extrapolated across Earth's ocean surface area this rate equals approximately 6.3 Pg y^{-1}, which is comparable to annual fossil fuel emissions (7.8 Pg y^{-1}) (IPCC 2013). Traits impacting fecal material in zooplankton and the mineral composition of phytoplankton the flux of C to the deep ocean (Durkin, et al. 2016, Armstrong, et al. 2002, Schmitz, et al. 2008, Miklasz & Denny 2010). Several studies have reported evolution of traits (e.g. 26% increase in calcification (Benner, et al. 2013) and 30% increase in PIC:POC (Schluter, et al. 2014) that could to increase the flux of

biomass to deep oceans in phytoplankton adapted to increased temperature or acidification (Benner, et al. 2013, Lohbeck, et al. 2012, Schluter, et al. 2014) which could increase C storage. Other examples of C cycling traits affecting the effective storage of C include those influencing the flammability (Bond & Keeley 2005) and of land plants, and the temperature sensitivity of microbial decomposition enzymes (Alster, et al. 2016). These traits contribute to the storage or release of C from biotic and storage pools and further investigation of the direction of their evolution will be valuable toward predicting potential effects on the global C cycle.

There is little doubt that trait evolution has impacted the global C cycle throughout Earth's history (Cornelissen & Cornwell 2014). The evolution of photosynthetic traits likely caused historic fluctuations in atmospheric CO_2 (Sage 2004, Beerling & Berner 2005, Beerling 2012), the evolution of cellulose decomposing enzymes in fungi increased the flux out of major terrestrial C pools (Floudas, et al. 2012, Floudas, et al. 2015, Nagy, et al. 2015), and the evolution of deep rooted trees and angiosperms has been linked to historic 10 to 20 fold reductions in atmospheric CO_2 (Taylor, et al. 2009). Thus, it is critical to recognize that changes in Earth's atmospheric CO_2 are not simply the product of geological or other abiotic processes, but rather can be largely driven by global evolutionary trends. To understand the effect of evolution on C cycling in the Anthropocene, it is important to consider this phenomenon in the context of global selection pressures resulting from climate change.

Natural selection in the Anthropocene

Increasing atmospheric CO_2 is changing environments and selection pressures globally (IPCC 2013). In addition to the direct effects of increased CO_2, organisms are faced with ocean acidification (Feely, et al. 2004), increased temperatures (Collins, et al. 2013), precipitation changes (Dai 2012, Schlaepfer, et al. 2017), and shifts in ecological communities. There are notable challenges in conclusively showing adaptive evolution in response to climate change

(Merilä & Jendry 2014) but recent studies provide experimental evidence for adaptation in response to such selection associated with traits that can influence C cycling. For example, traits that experiencing selection under increasing CO_2 (atmospheric or oceanic), include carbon fixation rates in phytoplankton (Jin, et al. 2013), nitrogen-fixation and growth rates in cyanobacteria (Hutchins, et al. 2015, Walworth, et al. 2016), and stomatal conductance in plants (Grossman & Rice 2014). Adaptation to changes in temperature has also been shown for microbial heterotrophs (Killeen, et al. 2017) and marine animals (Geerts, et al. 2015). While the evidence of evolution under experimentally manipulated conditions is compelling, it is not complete to conclude that adaptation of C cycling traits will be realized under natural conditions (Merilä and Jendry 2014).

Evidence of local adaptation to climate regimes such as temperature in land plants (Hancock, et al. 2011, Monroe, et al. 2016, Yeaman, et al. 2016), phytoplankton (Thomas, et al. 2012, Irwin, et al. 2015), heterotrophic microbes (Oliverio, et al. 2017) and other consumers (Yampolsky, et al. 2014) provides additional evidence that global climate change will alter natural selection. Additionally, drought-induced selection as a result of climate change might affect the evolution of plant root traits, which are important to C storage in soils (Kell, 2011). For example, altered drought regimes may select for genotypes with greater allocation to root biomass for increased water acquisition or may select for genotypes exhibiting a drought escape strategy wherein plants reproduce quickly and produce less belowground biomass.

Predicting the direction of selection for C cycling traits is complicated by the interaction between selection pressures. The ecological effect of CO_2, temperature, and precipitation (or other climate drivers) on organisms is often non-additive (Dieleman, et al. 2012) but is difficult to manipulate multiple drivers in a fully factorial design. This is important, however, as an experiment with *Emiliana huxleyi*, showed that responses to increased CO_2 can be negated if

temperature is also increasing (Benner, et al. 2013). Thus, in order to make accurate predictions of how populations will respond to selection, the interactions between selection pressures should be studied.

Genetic variation of C cycling traits

Adaptive evolution of C cycling traits requires standing genetic variation on which selection pressures can act (Fisher 1930). This variation can be in the form of mean differences between genotypes or heritable variation in plastic responses to environmental conditions (Via & Lande 1985). For example, the model grass *Brachypodium distachyon* exhibits significant heritable variation in C cycling traits including root biomass ($0.20 < H^2 < 0.36$), as well as genetic variation (genotype x environment interaction) in plastic responses to drought in the nutrient stoichiometry of leaves (Des Marais, et al. 2016). Indeed, there appears to be heritable variation across many plant species in traits such as root structure and leaf lability (Bandau, et al. 2017, Donovan, et al. 2014). A review of leaf lability traits indicates that they tend to exhibit significant estimates of heritability (h^2 as high as 0.9), although in some cases heritability for C cycling traits has not been detected (Donovan, et al. 2011). Laboratory studies also have revealed significant standing genetic variation within phytoplankton species in the responses of C cycling traits like growth rate to elevated CO_2, ocean acidification and increased temperature (Kremp, et al. 2012, Collins, et al. 2014, Listmann, et al. 2016, Hattich, et al. 2017). These studies find support for the presence of significant genetic variation in natural plant and phytoplankton populations and thus capacity to respond to selection of C cycling traits in response to climate change.

We know relatively less about the genetic variation of C cycling traits in non-primary producers. For example, heritable variation in microbial C cycling traits is rarely measured, as most studies report community rather than intraspecific variation (Wallenstein & Hall 2012).

However, recent reports confirm the presence of genetic variation among soil microbes in the temperature sensitivity and efficiency of decomposition enzymes (Alster, et al. 2016, Trivedi, et al. 2016). Heritable variation in such C cycling traits might be common but quantification of this variation via traditional quantitative genetics, genomic prediction, or comparison of microbial isolates is still often lacking (Box 1). Quantifying this variation and avoiding publication bias to report positive results is important for predicting the capacity for organisms to respond to selection. Additionally, describing genetic correlations between C cycling and other traits can provide insight into constraints on adaptive evolution (Etterson & Shaw 2001).

Evolution at ecological timescales

There is growing appreciation for the prevalence of rapid evolution occurring on timescales comparable with ecological processes (e.g. Hairston, et al. 2005). However, organism-environment feedbacks with atmospheric CO_2 remain mostly studied with respect plastic (i.e. physiological) and ecological (i.e. species composition) responses to climate change (Luo 2007). Although these are certainly important drivers of ongoing C cycling dynamics, they will occur in tandem with contemporary and now apparent rapid evolution of populations as well. Adaptation to increased temperature in the keystone zooplankton *Daphnia magna* occurred in as little as two years resulting in an increase in thermal tolerance of 3.5°C (Geerts, et al. 2015) and the evolution of metabolism related genes (Jansen, et al. 2017). Similar rapid adaptation to increased temperature involving metabolic genes or traits has been documented in the phytoplankton *Chlorella vulgaris* in as little as 100 generations (Schaum, et al. 2017, Padfield, et al. 2016). Adaptation to drought has also been observed in less than a decade in a population of the annual plant *Brassica rapa* (Franks, et al. 2007). In response to several years of late season droughts, these populations evolved earlier flowering time (up to 8 days earlier) (Franks, et al. 2007), which is genetically correlated with lower root biomass in the close relative *Brassica*

napus (Fletcher, et al. 2015). Such rapid evolutionary shifts in annual plant species could result in lower soil C storage in environments predicted to experience more frequent summer droughts as a result of climate change, contributing to positive feedback interactions with atmospheric CO_2.

These examples illustrate the potential for evolution at speeds comparable to plastic and ecological responses to climate change. Interestingly, adaptive evolution can also occur in the opposite direction of plastic responses (Ghalambor, et al. 2015). For example, under ocean acidification, phytoplankton often plastically decrease photosynthetic rates, while evolution occurs in the opposite direction to increase photosynthetic rates and therefore the probability of long-term C sequestration (Collins, et al. 2014). Accordingly, predictions of future trait values should not be based on plastic responses alone. However, not all seemingly strong selection pressures will engender rapid evolutionary responses. For example, *Arabidopsis thaliana* plant populations grown under elevated CO_2 in a field setting showed increased growth and fruit production but no evolutionary effect, suggesting a plastic rather than evolutionary response (Lau, et al. 2006).

Direction of selection on C cycling traits

The evolution of C cycling traits could increase, decrease or stabilize atmospheric CO_2 concentrations, depending on the direction of natural selection on C cycling traits. Experimental evolution provides one approach to study the direction of this evolution. By allowing populations to adapt to environments simulating future conditions, the response to selection in C cycling traits can be measured empirically. This approach has proven especially insightful for studying the evolution of phytoplankton. These experiments have observed evolutionary responses to increased CO_2 (and associated ocean acidification) (Scheinin, et al. 2015, Bond & Keeley 2005, Jin, et al. 2013, Hutchins, et al. 2015, Walworth, et al. 2016, Collins, et al. 2014, Lohbeck, et al.

2013, Li, et al. 2016), temperature (Schaum, et al. 2017, Padfield, et al. 2016), or both (Benner, et al. 2013, Schluter, et al. 2014) and have provided valuable information toward resolving the direction of selection on C cycling traits. *C. reinhardtii* adapted to a decade of elevated temperature in outdoor mesocosms evolved increase in net photosynthesis (Figure 2) (Benner, et al. 2013) and in a separate experiment *Chorella vulgaris* down-regulated respiration relative to photosynthesis (Padfield, et al. 2016). Additionally, populations of the calcifying phytoplankton *Emiliania huxleyi* adapted to elevated temperature exhibited shifts in ballasting traits in directions predicted to increase sinking speed and thus the flux of C to deep oceans (Schluter, et al. 2014). Such studies suggest that selection caused by increased temperature might act in directions that increase the net flux of C from the atmosphere to biotic and abiotic storage pools and to produce a negative feedback loop with atmospheric CO_2.

However, studies examining phytoplankton evolution in response to acidification highlight that further work is needed to draw generalizable conclusions about the direction of selection caused by acidification of the world's oceans. For example, under experimentally increased CO_2 and acidification, *Gephyrocapsa oceanica* evolved higher photosynthetic C fixation and growth rate (Jin, et al. 2013), whereas *Phaeodactylum tricornutum* evolved reduced photosynthesis, respiration, and growth rate (Li, et al. 2016). To what extent such discrepancies are caused by differences between species' evolutionary responses or experimental conditions remain unclear. A valuable direction for future research would be to compare the evolutionary trajectories of C cycling traits in multiple species (including heterotrophs such as *Daphnia*), ideally in factorial designs with multiple selection pressures.

Studying evolution under ecologically realistic conditions is also important. For example, recent (< 50 yrs) woody encroachment into grasslands has altered community composition and C balances within these systems (Barger 2011). Such shifts in community composition can

influence evolution by altering interactions within and across trophic levels (Figure 1). Because mesocosm approaches will not capture the more complex influences of community-level change, manipulative and observational studies to assess climate-change driven evolutionary dynamics in real communities will be valuable.

There are several alternatives to mesocosm experiments to study the direction of evolution of C cycling traits in natural populations. The strength of selection on C cycling traits could be measured with a quantitative genetic approach in a pedigreed population grown in a common environment reflecting future conditions (Arnold & Lande 1983). Long-term monitoring of populations in experimentally modified native habitats can also be useful. For example, the a ten-year precipitation manipulation experiment within native grassland communities found that increased intra-annual precipitation variability selected for a genotype of a dominant grass with greater allocation to root biomass (Avolio & Smith 2013). Such a trait shift might lead to more belowground C storage, yet how this population evolution ultimately scaled to ecosystem-level carbon cycling and atmospheric feedbacks is uncertain.

Naturally occurring CO_2, pH, temperature, or precipitation clines substitute space for time, representing a chronosequence of environmental change. Genetic differentiation along these clines can be studied to infer the direction of selection in response to climate change. Populations adapted to terrestrial (Watson-Lazowski, et al. 2016) and aquatic CO_2-emitting vents (Kumar, et al. 2017, Kenkel, et al. 2017) have experienced to environments with higher ambient CO_2 and/or lower pH. A recent study of *Plantago lanceolata* plants adapted to such elevated CO_2 habitats found genetic evidence of an evolved increase in photosynthetic capacity and respiration rate although the effect on net photosynthesis is unclear (Watson-Lazowski, et al. 2016). Similarly, the direction of selection on C cycling traits in response to increased temperature and aridity can be inferred from populations diverged along natural temperature and precipitation

clines. Populations of a keystone snail herbivore, *Radix balthica*, adapted to warmer springs have evolved a higher metabolic, respiration, and consumption rates (Schaum *et al.* 2017, unpublished). This is consistent with theoretical predictions that increased temperature leads to shifts in the stoichiometry of consumers' nutrient needs (increased C:N ratio), increased metabolic demand, higher respiration rates, and greater prey consumption (Schmitz 2013, Schmitz 2008). These results indicate the potential for heterotrophic evolution in response to increasing temperatures to perpetuate a positive feedback on atmospheric CO_2. An exciting direction for future work will be to assess whether such evolution will counter negative feedbacks occurring at lower trophic levels.

It is also possible to measure the recent direction of evolution in C cycling traits *in situ*. 'Resurrection' studies use seed banks or other sources of dormant organisms to measure recent evolution in plants (Etterson, et al. 2016), zooplankton (Geerts, et al. 2015) and microbes (Orsini, et al. 2013). Resurrection studies have provided convincing evidence that rapid evolution can occur across different organisms (Franks, et al. 2017) and present a promising opportunity to reconstruct the direction of recent evolution of C cycling traits. Past, ongoing, and potential for future evolution of C cycling traits could also be studied using genomic sequence data to predict C cycling phenotypes in preserved specimens [98] or recent collections of natural populations (Bay, et al. 2017).

Evolution of C cycling traits as an eco-evolutionary feedback

The idea that organisms alter their selective environment originated with Darwin (Darwin 1881) and can be conceptualized under the framework of eco-evolutionary feedbacks (Hendry 2016). Eco-evolutionary feedbacks occur when the evolution of traits modifies environmental conditions in response to natural selection. This environmental modification then feeds back to influence future evolutionary change. Here we have discussed why organisms are expected to

experience strong natural selection as a result of climate change. We have also seen the potential for this selection to lead to the evolution of C cycling traits. Yet to date, few if any studies have integrated these criteria into a complete eco-evolutionary feedback cycle. Because the primary fluxes that interact with atmospheric CO_2 are those mediated by organisms (IPCC 2013), evolutionary changes in C cycling traits are expected to influence the balance between atmospheric C and other pools in the C cycle. Thus, the evolution of C cycling traits might either accelerate or decelerate the current trend toward increasing atmospheric CO_2 and climate change, altering future selection regimes and completing the eco-evolutionary feedback loop. Therefore, the model of eco-evolutionary feedback loops provides a useful framework to conceptualize the components underlying the evolutionary dynamics of global C cycling (Figure 3). Integrating all of the steps of such feedbacks to characterize the contemporary eco-evolutionary dynamics of C cycling will require an unprecedented yet achievable synthesis, both among multiple disciplines and across multiple taxa (Box 2).

Concluding remarks

Here we have discussed recent evidence for the components of eco-evolutionary feedbacks involving C cycling traits (Box 1, Figure 3). There are also a growing number of studies that have begun to integrate these components. For example, recent work generally suggests that C cycling traits in phytoplankton will evolve in response to elevated temperatures in directions likely to have a dampening effect on atmospheric CO_2, establishing a negative feedback loop. In contrast, theory and recent empirical work indicates that metabolism and nutrient stoichiometry in heterotrophs will evolve in directions that establish positive feedback loops with atmospheric CO_2. Such dynamics speak to the need to understand the potential tradeoffs among different organisms and traits (as well as across ecosystems and communities) as they evolve in response to climate change. A more complete understanding of the

evolutionary trajectories of C cycling traits within individual ecosystems is needed because the net effect of evolution at a global scale will be impacted by pressure from both positive and negative feedbacks.

A critical question is whether the magnitude of evolutionary changes in C cycling traits will be large enough to significantly affect atmospheric CO_2 concentrations. It is important to remember that evolution has demonstrated the capacity to affect global C cycling throughout Earth's history and consider that climate change is altering selection on organisms globally. It is also clear that even small changes to the global C cycle can have large impacts, given that anthropogenic CO_2 sources (burning fossil fuels, land-use change, etc) account for less than 1.5% of global C fluxes yet are driving major environmental changes across the planet (IPCC 2013). It is too early to conclude that evolution will dramatically alter future atmospheric CO_2 but there are clear reasons to believe that the potential is real.

Identifying the feedback potential of different taxa will be useful for management. The evolution of traits that increase C storage represents a particularly valuable 'evosystem service' (Rudman, et al. 2017) and organisms predicted to do so may be prioritized for conservation and agricultural use. For example, *C. reinhardtii* are expected to evolve increased net photosynthesis (Figure 2) and lakes they inhabit might be targeted for protection. While the C storage potential of organisms will never be the only factor influencing management, understanding organisms' roles in eco-evolutionary dynamics of C cycling could lead to more purposeful decision making.

Studying the evolution of C cycling traits also presents a meaningful research objective at a more fundamental level, as it necessarily integrates different levels of biological and ecological organization (Levin 1992). Progress will not only provide insight into important evolutionary processes but also help to better understand the relative importance of eco-evolutionary feedbacks in natural ecosystems (Laland, et al. 2014). Evolutionary biologists and ecosystem

ecologists should work together toward the common research goal of quantifying the impact of organism evolution in natural ecosystems on global C cycling.

REFERENCES

Allison SD (2012) A trait-based approach for modelling microbial litter decomposition. Ecol Lett, 15, 1058–1070.

Allison SD (2014) Modeling adaptation of carbon use efficiency in microbial communities. Front Microbiol, 5.

Alster CJ, et al. (2016) Temperature sensitivity as a microbial trait using parameters from macromolecular rate theory. Front Microbiol, 7, 1821.

Armstrong RA, et al. (2002) A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. Deep Res Part II Top Stud Oceanogr, 49, 219–236.

Arnold SJ, Lande R (1983) The measurement of selection on correlated characters. Evolution, 37, 1210–1226.

Avolio ML, Smith MD (2013) Mechanisms of selection: Phenotypic differences among genotypes explain patterns of selection in a dominant species. Ecology, 94, 953–965.

Bailey JK, et al. (2009) From genes to ecosystems: a synthesis of the effects of plant genetic factors across levels of organization. Philos Trans R Soc Lond B Biol Sci, 364, 1607–1616.

Bandau F, et al. (2017) Genotypic variability in *Populus tremula L*. affects how anthropogenic nitrogen enrichment influences litter decomposition. Plant Soil, 410, 467–481.

Barger NN (2011) Woody plant proliferation in North American drylands: A synthesis of impacts on ecosystem carbon balance. Journal Geophys Res, 116, 2873-2893.

Bay RA, et al. (2017) Predicting responses to contemporary environmental change using evolutionary response architectures. Am Nat, 189, 463-473.

Beer C, et al. (2010) Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate. Science, 329, 834–838.

Beerling DJ (2012) Atmospheric carbon dioxide: a driver of photosynthetic eukaryote evolution for over a billion years? Philos Trans R Soc B Biol Sci, 367, 477–482.

Beerling DJ, Berner R (2005) Feedbacks and the coevolution of plants and atmospheric CO2. Proc Natl Acad Sci USA, 102, 1302–1305.

Benner I, et al. (2013) *Emiliania huxleyi* increases calcification but not expression of calcification-related genes in long-term exposure to elevated temperature and pCO_2. Philos Trans R Soc B Biol Sci, 368.

Bond WJ, Keeley JE (2005) Fire as a global "herbivore": the ecology and evolution of flammable ecosystems. Trends Ecol Evol, 20, 387–394.

Braakman R, et al. (2015) Metabolic evolution and the self-organization of ecosystems. PNAS, 114, E3091–E3100.

Bradford MA, Crowther TW (2013) Carbon use efficiency and storage in terrestrial ecosystems. New Phytol, 199, 7–9.

Collins M, et al. (2013) Long-term Climate change: projections, commitments and irreversibility. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Joussaume, S. et al., eds), 1029–1136.

Collins S, et al. (2014) Evolutionary potential of marine phytoplankton under ocean acidification. Evol Appl, 7, 140–155.

Cornelissen JHC, Cornwell WK (2014) The Tree of Life in ecosystems: Evolution of plant effects on carbon and nutrient cycling. J Ecol, 102, 269–274.

Cornwell WK, et al. (2008) Plant species traits are the predominant control on litter decomposition rates within biomes worldwide. Ecol Lett, 11, 1065–1071.

Cramer W, et al. (1999) Comparing global models of terrestrial net primary productivity (NPP): overview and key results. Glob Chang Biol, 5, 1–15.

Crutzen PJ (2006) The "Anthropocene." In Earth System Science in the Anthropocene (Ehlers, E. and Krafft, T., eds), pp. 13–18, Springer.

Dai A (2012) Increasing drought under global warming in observations and models. Nat Clim Chang, 3, 52–58.

Darwin C (1881) The formation of vegetable mould through the action of worms, with observations on their habits, John Murray.

de Bello F, et al. (2010) Towards an assessment of multiple ecosystem processes and services via functional traits. Biodivers Conserv, 19, 2873-2893.

De Deyn GB, et al. (2008) Plant functional traits and soil carbon sequestration in contrasting biomes. Ecol Lett, 11, 516–531.

Des Marais DL, et al. (2016) Interactive effects of water limitation and elevated temperature on the physiology, development and fitness of diverse accessions of *Brachypodium distachyon*. New Phytol, 214, 132–144.

Dieleman WIJ, et al. (2012) Simple additive effects are rare: a quantitative review of plant biomass and soil process responses to combined manipulations of CO_2 and temperature. Glob Chang Biol, 18, 2681-2693.

Donovan LA, et al. (2011) The evolution of the worldwide leaf economics spectrum. TREE, 26, 89-95.

Donovan LA, et al. (2014) Ecological and evolutionary lability of plant traits affecting carbon and nutrient cycling. J Ecol, 102, 302–314.

Durkin DA, et al. (2016) Sinking phytoplankton associated with carbon flux in the Atlantic Ocean. Limnol Oceanogr, 61, 1172–1187.

Etterson J, Shaw R (2001) Constraint to adaptive evolution in response to global warming. Science, 351, 151-154.

Etterson JR, et al. (2016) Project baseline: An unprecedented resource to study plant evolution across space and time. Am J Bot, 103, 164–173.

Falkowski PG (1994) The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth Res, 39, 235–258.

Feely RA, et al. (2004) Impact of anthropogenic CO2 on the CaCO3 system in the oceans. Science, 305, 362–366.

Fisher RA (1930) The Genetical Theory of Natural Selection, Clarendon Press.

Fletcher RS, et al. (2015) QTL analysis of root morphology, flowering time, and yield reveals trade-offs in response to drought in Brassica napus. J Exp Bot, 66, 245–256.

Floudas D, et al. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science, 336, 1715–1719.

Floudas D, et al. (2015) Evolution of novel wood decay mechanisms in *Agaricales* revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*. Fungal Genet Biol, 76, 78–92.

Franks SJ, et al. (2007) Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proc Natl Acad Sci USA, 104, 1278–1282.

Franks SJ, et al. (2017) Using the resurrection approach to study contemporary evolution in changing environments. Evol Appl, DOI: 10.1111/eva.12528.

Franks, S.J. and Hoffmann, A.A. (2012) Genetics of climate change adaptation. Annu Rev Genet, 46, 185–208.

Geerts, AN, et al. (2015) Rapid evolution of thermal tolerance in the water flea Daphnia. Nat Clim Chang, 5, 665–668.

Ghalambor CK, et al. (2015) Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature, 525, 372–375.

Gienapp P, et al. (2017) Genomic Quantitative Genetics to Study Evolution in the Wild. TREE, DOI: 10.1016/j.tree.2017.09.004.

Grossman JD, Rice KJ (2014) Contemporary evolution of an invasive grass in response to elevated atmospheric CO_2 at a Mojave Desert FACE site. Ecol Lett, 17, 710–716.

Hairston NG, et al. (2005) Rapid evolution and the convergence of ecological and evolutionary time. Ecol Lett, 8, 1114–1127.

Hancock AM, et al. (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. Science, 334, 83–86.

Hattich G, et al. (2017) Experiment on inter- and intra-specific phenotypic plasticity of three phytoplankton species in response to ocean acidification. Biol Lett, 13, 20160774.

Hendry AP (2016) Eco-Evolutionary Dynamics, Princeton University Press.

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. Nature, 470, 479–485.

Hutchins DA, et al. (2015) Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. Nat Commun, 6, 1–7.

IPCC (2013) Climate change 2013: the physical science basis. In Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Irwin AJ, et al. (2015) Phytoplankton adapt to changing ocean environments. Proc Natl Acad Sci USA, 112, 5762–5766.

Jansen M, et al. (2017) Thermal tolerance in the keystone species *Daphnia magna* - a candidate gene and an outlier analysis approach. Mol Ecol, 26, 2291–2305.

Jin P, et al. (2013) Evolutionary responses of a coccolithophorid *Gephyrocapsa oceanica* to ocean acidification. Evolution, 67, 1869–1878.

Kell DB (2012) Large-scale sequestration of atmospheric carbon via plant roots in natural and agricultural ecosystems: why and how. Philos Trans R Soc B Biol Sci, 367, 1589–1597.

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

Kenkel C, et al. (2017) Functional genomic analysis of corals from natural CO*2*-seeps reveals core molecular responses involved in acclimatization to ocean acidification. Global Change Biology, DOI: 10.1111/gcb.13833.

Killeen, J. et al. (2017) Evolutionary rescue and local adaptation under different rates of temperature increase: a combined analysis of changes in phenotype expression and genotype frequency in Paramecium microcosms. Mol. Ecol. 26, 1734–1746.

Kremp A, et al. (2012) Intraspecific variability in the response of bloom-forming marine microalgae to changed climate conditions. Ecol Evol, 2, 1195–1207.

Kumar A, et al. (2017) Molecular response of *Sargassum vulgare* to acidification at volcanic CO2 vents - insights from de novo transcriptomic analysis. Mol Ecol, 26, 2276-2290.

Laland K, et al. (2014) Does evolutionary theory need a rethink? Nature, 514, 161–164.

Laliberté E (2016) Below-ground frontiers in trait-based plant ecology. New Phytol, 213, 1597–1603.

Langer G, et al. (2009) Strain-specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry. Biogeosciences Discuss, 6, 2637–2646.

Lau JA, et al. (2006) Strong ecological but weak evolutionary effects of elevated CO_2 on *Arabidopsis thaliana*. New Phytol, 175, 351-362.

Levin SA (1992) The problem of pattern and scale in ecology. Ecology, 73, 1943–1967.

Li F, et al. (2016) Decreased photosynthesis and growth with reduced respiration in the model diatom *Phaeodactylum tricornutum* grown under elevated CO_2 over 1800 generations. Glob Chang Biol, 23, 127-137.

Listmann L, et al. (2016) Swift thermal reaction norm evolution in a key marine phytoplankton species. Evol Appl, 9, 1156–1164.

Litchman E, et al. (2015) Global biogeochemical impacts of phytoplankton: A trait-based perspective. J Ecol, 103, 1384–1396.

Lohbeck KT, et al. (2012) Adaptive evolution of a key phytoplankton species to ocean acidification. Nat Geosci, 5, 346–351.

Lohbeck KT, et al. (2013) Functional genetic divergence in high CO2 adapted *Emiliania huxleyi* populations. Evolution, 67, 1892–1900.

Luo Y (2007) Terrestrial carbon–cycle feedback to climate warming. Annu Rev Ecol Evol Syst, 38, 683–712.

Matthews B, et al. (2011) Toward an integration of evolutionary biology and ecosystem science. Ecol Lett, 14, 690–701.

Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. Evol Appl, 7, 1–14.

Merilä J, Hoffmann AA (2016) Evolutionary impacts of climate change. Oxford Research Encyclopedia of Environmental Science, Oxford University Press.

Miklasz KA, Denny MW (2010) Diatom sinking speeds: Improved predictions and insight from a modified Stokes' law. Limnol Oceanogr, 55, 2513–2525.

Monroe JG, et al. (2016) Adaptation to warmer climates by parallel functional evolution of CBF genes in Arabidopsis thaliana. Mol Ecol, 25, 3632–3644.

Nagy LG, et al. (2015) Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. Mol Biol Evol, 33, 959-970.

Oliverio AM, et al. (2017) Identifying the microbial taxa which consistently respond to soil warming across time and space. Glob Chang Biol, 23, 2117–2129.

Orsini L, et al. (2013) The evolutionary time machine: Using dormant propagules to forecast how populations can adapt to changing environments. Trends Ecol Evol, 28, 274–282.

Padfield D, et al. (2016) Rapid evolution of metabolic traits explains thermal adaptation in phytoplankton. Ecol Lett, 19, 133–142.

Pastor J (2017) Ecosystem ecology and evolutionary biology, a new frontier for experiments and models. Ecosystems, 20, 245–252.

Paulus JK, et al. (2013) Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. Nat Commun, 4, 1518.

Rudman SM, et al. (2017) Evosystem services : Rapid evolution and the provision of ecosystem services. Trends Ecol Evol, 32, 403–415.

Sage RF (2004) The evolution of C4 photosynthesis. New Phytol, 161, 341–370.

Schaum CE et al. (2017) Adaptation of phytoplankton to a decade of experimental warming linked to increased photosynthesis. Nat Ecol Evol, 1, 94.

Scheinin M, et al. (2015) Experimental evolution gone wild. J R Soc Interface, 12.

Schlaepfer DR, et al. (2017) Climate change reduces extent of temperate drylands and intensifies drought in deep soils. Nat Commun, 8, 14196.

Schluter L, et al. (2014) Adaptation of a globally important coccolithophore to ocean warming and acidification. Nat Clim Chang 4, 1024–1030.

Schmitz OJ (2013) Global climate change and the evolutionary ecology of ecosystem functioning. Ann N Y Acad Sci, 1297, 61–72.

Schmitz OJ, et al. (2008) From individuals to ecosystem function: toward an integration of evolutionary and ecosystem ecology. Ecology, 89, 2436–2445.

Shaw RG, Etterson JR (2012) Rapid climate change and the rate of adaptation: Insight from experimental quantitative genetics. New Phytol, 195, 752–765.

Sterner R, Elser J (2002) Ecological Stoichiometry: the Biology of Elements from Molecules to the Biosphere, Princeton University Press.

Swarts K, et al. (2017) Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science, 357, 512-515.

Taylor LL, et al. (2009) Biological weathering and the long-term carbon cycle: Integrating mycorrhizal evolution and function into the current paradigm. Geobiology, 7, 171-191.

Thomas MK, et al. (2012) A global pattern of thermal adaptation in marine phytoplankton. Science, 338, 1085–1088.

Trivedi P, et al. (2016) Microbial regulation of the soil carbon cycle: evidence from gene–enzyme relationships. ISME J, 10, 2593–2604.

Via S, Lande R (1985) Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution, 39, 505–522.

Wallenstein MD and Hall EK (2012) A trait-based framework for predicting when and where microbial adaptation to climate change will affect ecosystem functioning. Biogeochemistry, 109, 35–47.

Walworth NG, et al. (2016) Molecular and physiological evidence of genetic assimilation to high CO2 in the marine nitrogen fixer *Trichodesmium*. Proc Natl Acad Sci USA, 113, E7367–E7374.

Watson-Lazowski A, et al. (2016) Plant adaptation or acclimation to rising CO_2? Insight from first multigenerational RNA-Seq transcriptome. Glob Chang Biol, 22, 3760–3773.

Yampolsky LY, et al. (2014) Adaptive phenotypic plasticity and local adaptation for temperature tolerance in freshwater zooplankton. Proc R Soc, B 281.

Yeaman S, et al. (2016) Convergent local adaptation to climate in distantly related conifers. Science, 353, 1431–1433.

CHAPTER 3

ADAPTATION TO WARMER CLIMATES BY PARALLEL FUNCTIONAL EVOLUTION OF *CBF* GENES IN *ARABIDOPSIS THALIANA*

Summary

The evolutionary processes and genetics underlying local adaptation at a species wide level are largely unknown. Recent work has indicated that a frameshift mutation in a member of a family of transcription factors, C-repeat binding factors or *CBF*s, underlies local adaptation and freezing tolerance divergence between two European populations of *Arabidopsis thaliana.* To ask whether the species-wide evolution of *CBF* genes in *Arabidopsis* is consistent with local adaptation, we surveyed *CBF* variation from 477 wild accessions collected across the species' range. We found that *CBF* sequence variation is strongly associated with winter temperature variables. Looking specifically at the minimum temperature experienced during the coldest month, we found that *Arabidopsis* from warmer climates exhibit a significant excess of non-synonymous polymorphisms in *CBF* genes and revealed a *CBF* haplotype network whose structure points to multiple independent transitions to warmer climates. We also identified a number of newly described mutations of significant functional effect in *CBF* genes, similar to the frameshift mutation previously indicated to be locally adaptive in Italy, and find that they are significantly associated with warm winters. Lastly, we uncover relationships between climate and the position of significant functional effect mutations between and within *CBF* paralogs suggesting variation in adaptive function of different mutations. Cumulatively, these findings support the hypothesis that disruption of *CBF* gene function is adaptive in warmer climates, and illustrate how parallel evolution in a transcription factor can underlie adaptation to climate.

Introduction

Evidence for adaptation to local environments in nature is widespread (Leimu & Fischer 2008; Hereford 2009; Fournier-Level *et al*. 2011) and adaptive loci have been identified (Shapiro *et al.* 2004; Hoekstra *et al.* 2006; Uga *et al.* 2013; Des Marais *et al.* 2014; Pardo-Diaz *et al.* 2015). However, long-standing questions remain about the processes underlying adaptive evolution. What components of the environment are important selective drivers of divergent adaptive evolution? (Levins 1968; Endler 1986; Mullen & Hoekstra 2008; Lasky *et al.* 2012). At the species level, does molecular evolution proceed by the fixation of a single adaptive allele, or in parallel, through mutations appearing independently in different populations? (Maynard Smith & Haigh 1974; Pennings & Hermisson 2006; Remington 2015; Ralph & Coop 2015).

The flowering plant *Arabidopsis thaliana* (L.) Heynh. (hereafter referred to as *Arabidopsis*) is a well-established model for both functional and ecological genetics (Mitchell-Olds 2001) and shows local adaptation across its range (Fournier-Level *et al.* 2011; Ågren & Schemske *et al.* 2012). *Arabidopsis* populations, having expanded north following the end of the last Pleistocene glaciation (Beck *et al.* 2008), are found in environments that differ greatly in thermal regime, creating a gradient across which populations may have become locally adapted. For example, a reciprocal transplant between *Arabidopsis* accessions originating from the northern and southern portions of the species range, Sweden and Italy respectively, revealed strong local adaptation and pointed to freezing tolerance as an important adaptive trait (Ågren & Schemske 2012). In a subsequent experiment, Ågren and colleagues (2013) carried out a long-term reciprocal transplant of a mapping population to uncover the genetic basis of local adaptation in the Italian and Swedish ecotypes. In this experiment, fitness (quantified by both survival and fruit production) was measured in an Italy × Sweden derived mapping population of recombinant inbred lines (RILs) planted in Italian and Swedish environments for 3 consecutive years. Genomic regions explaining

differences in fitness between individuals in Italian and Swedish locations were identified by quantitative trait loci (QTL) mapping. They identified 15 QTL underlying variation in fitness in Italian and Swedish environments, one of which localized to a genomic region containing 3 transcription factors known to function in freezing tolerance called C-repeat binding factors (*CBF*s). Notably, this QTL exhibited a genetic trade off; in Italy, the Italian allele in this region was adaptive while the Swedish allele was deleterious, with the opposite being true in Sweden. Cumulatively, these findings suggest that a fitness trade off associated with freezing tolerance genes is driving local adaptation between *Arabidopsis* populations diverged along a thermal gradient (Ågren & Schemske 2012; Ågren *et al.* 2013).

Cold acclimation, the processes whereby plants respond to low but non-freezing temperatures and achieve tolerance to future freezing conditions, has been extensively studied. Plants have evolved strategies to prepare for freezing conditions in order to prevent tissue damage caused by cellular freezing. In brief, a drop in temperature toward 0°C cues plants to activate a suite of genes that make cellular osmotic adjustments conferring freezing tolerance to anticipated sub-0°C temperatures (Zarka *et al* 2003; Hannah *et al.* 2006). *Arabidopsis* has served as a model to elucidate the genetic and molecular basis of freezing tolerance. In particular, the role of *CBF*s in the regulation of genes involved in freezing tolerance has been well characterized (Thomashow *et al.* 2001; Thomashow 2010; Park *et al.* 2015). The *CBF*s are a paralogous family of transcription factors, three of which are located in a tandem array on chromosome 4. Each encodes a protein containing an AP2 DNA binding domain that recognizes and binds CRT/DRE promoter regions of target genes (Stockinger *et al.* 1997; Gilmour *et al.* 1998), and a C-terminal activation domain responsible for the recruitment of transcriptional machinery (Stockinger 1997; Wang *et al.* 2005; Kang *et al.* 2013). An increase in expression of the *CBF* genes occurs in a quantitative manner as the environment in which plants are growing approaches low but non-freezing temperatures

(Medina *et al.* 1999, Zarka *et al* 2003). This increase in CBF expression results in the activation of over 100 gene targets, known as the *CBF* regulon (Thomashow 2001; Park *et al* 2015).

Using the same mapping population used to find fitness QTL by Ågren *et al.* (2013), the genomic region containing the *CBF* genes was again detected by QTL mapping as explaining divergence in freezing tolerance between the Italian and Swedish ecotypes (Oakley *et al.* 2014). More recently, Gehan *et al.* (2015) analyzed these Italian and Swedish *CBF* genes and discovered that the Italian allele contains a 13 base pair deletion causing a frameshift mutation that disrupts the activation domain of *CBF2*. With transgenic experiments, they confirmed that this polymorphism disrupts CBF2 function and is responsible for divergence of freezing tolerance between the Italian and Swedish *Arabidopsis*. In light of the results of Ågren *et al.* (2013), the findings of Gehan *et al.* (2015) suggest that the frameshift mutation in *CBF2* confers an adaptive loss of freezing tolerance in the Italian ecotype. The effect that this mutation has on freezing tolerance is consistent with knowledge of the central role that *CBF* genes play in cold response. Indeed, variation in *CBF* genes has been indicated in underlying divergence in freezing tolerance between other *Arabidopsis* accessions (Alonso-Blanco *et al.* 2005; Kang *et al.* 2013). More intriguing is the adaptive value this frameshift mutation in *CBF2* appears to have in Italy, which suggests that there is a fitness penalty for freezing tolerance in Italy that is avoided by a significant functional mutation in *CBF2*. While freezing tolerance has been shown to vary along a latitudinal gradient, decreasing in populations from lower latitudes (Hannah *et al.* 2006; Zhen & Ungerer 2008; Zuther *et al.* 2012), the adaptive significance of this pattern has been unclear. It is possible that unnecessary *CBF* induced cold responses negatively impact phenotypes. Indeed, transgenic *Arabidopsis* lines overexpressing *CBF* perform poorly under above-freezing temperatures (Thomashow 2010 and references therein). The recent studies of locally adapted Italian and Swedish ecotypes (Ågren *et al.* 2013; Oakley *et al.* 2014; Gehan *et al.* 2015) experimentally

connect increased fitness, measured in the field in Italy, and a frameshift mutation in a *CBF* gene that causes a significant reduction in freezing tolerance. This background lays forth a framework and opportunity to investigate how locally adaptive molecular evolution proceeds at a continental scale, by considering *a priori* that a *CBF* frameshift mutation is selectively favored in at least one southern environment. What remains unknown is the role of *CBF* genes in local adaptation in *Arabidopsis* populations elsewhere across the species range.

To address the hypothesis that adaptation to warmer climates is associated with variation in *CBF* across the range of *Arabidopsis,* we asked four questions. Is variation in *CBF* genes associated with a particular environmental gradient? Are *CBF* genes evolving neutrally along this gradient? If not, is there evidence for a hard selective sweep of a single adaptive allele or parallel evolution of multiple independently derived alleles? Is the geographical distribution of frameshift and premature stop codon alleles in *CBF* genes non-random with respect to the environment? Here, we generated or obtained sequences for *CBF1, CBF2,* and *CBF3* from 477 accessions originating from across the species' range, comparing these sequence data with variables characterizing the environments of the accession's locations to assess the relationship between climate and molecular variation in *CBF* genes.

Materials and Methods

Plant and genetic material

Extracted DNAs from 136 *Arabidopsis* accessions collected across Eurasia (Beck *et al.* 2008) were included. Additionally, we extracted DNA from ecotypes Castelnuovo (IT), Rödåsen (SW), Kas, and Tsu-1 (McKay *et al.* 2008; Ågren and Schemske 2012, Ågren *et al*. 2013;). We also downloaded sequence data for a subset of the *CBF* gene family from geo-referenced accessions of the MPICWang2013 collection made available through the *Arabidopsis* 1001 genomes project (see Acknowledgements and Heyndrickx *et al.* 2014). Specifically, we analyzed

CBF1, *CBF2* and *CBF3* coding region sequences (337 total), which are found in a tandem array on chromosome 4 and include the paralog containing a functional mutation in an Italian ecotype (*CBF2*, Ågren *et al*. 2013). We also downloaded the reference genome for *Arabidopsis lyrata* (L.) O'Kane & Al-Shehbaz (Hu *et al.* 2011).

CBF Sequences

For accessions collected from the field (Beck *et al.* 2008) and IT, SW, Kas, and Tsu-1, the coding region of *CBF1*, *CBF2* and *CBF3* paralogs were PCR-amplified with the following primers: *CBF3*, forward primer sequence TTT TCC ACT CGT TTC TAC AAC A, *CBF3* reverse primer sequence CTA CTT AAA CCT TAT CCA GTT T, *CBF1* forward primer sequence TCA ATT TAA TTT ACA CTC GTT T, *CBF1* reverse primer sequence TTT CAG CAA ACC ATA CCA ACA, CBF2 forward primer sequence ACA TTC GTT TCT CAC AAC CAA, and *CBF2* reverse primer sequence TCT CAT AAA CCT TAT CCA GTT T. All amplicons were Sanger–sequenced on an ABI 3130xL Genetic Analyzer using samples prepared with ABI's BigDye® Terminator v3.1 kit. The *CBF* coding regions from the MPICWang2013 accessions were obtained digitally with a custom python script. This script is available for download at https://bitbucket.org/greymonroe/scripts_CBF. *CBF* sequences used in this study will be made available on Data Dryad.

Sequence alignment and polymorphism characterization

The coding regions of *CBF1*, *CBF2* and *CBF3* were concatenated into a single sequence. This concatenated sequence was aligned by MUSCLE using Mega v6.06 (Tamura *et al.* 2013). A single nucleotide polymorphism (SNP) matrix was then generated by scoring individuals at each polymorphic locus as follows using custom R scripts made available for download at https://bitbucket.org/greymonroe/scripts_CBF. At each biallelic polymorphic site an accession was scored '0' if the allelic state was shared by *A. lyrata* or '1' if the allelic state differed from *A.*

lyrata. Sites containing more than one allelic state and indels were ignored in the generation of the SNP matrix. To analyze the role of significant functional mutations, frameshift and premature stop codon mutations were identified by examining translated protein sequences of each accession using Columbia gene models (TAIR 10).

Climate variables and analysis

We wanted to estimate the degree to which variation in SNPs in *CBF* genes was explained by different parameters of local environments. We compiled 116 environmental variables describing the known collection location for each accession. Specifically we obtained data from WorldClim (Hijmans et al 2005), CGIAR-CSI Global-Aridity database (Zomer *et al.* 2008), vapor pressure deficit using Climate Research Unit (CRU) data (New *et al.* 2002), NCEP reanalysis data (Kalnay *et al.* 1996), SRB data (https://eosweb.larc.nasa.gov/project/srb/srb_table), soil water capacity data (Dunne & Wilmott, http://daac.ornl.gov/SOILS/guides/DunneSoil.html), a groundwater dataset (Fan et al 2013), a global soil aluminum toxicity dataset (Sanchez *et al.* 2003), and a soil pH dataset (Nachtergaele & Batjes 2012). Details regarding these datasets are explained in Lasky *et al.* (2012). Additionally, we calculated the frequency of drought and favorable conditions using the vegetative health index (VHI) which uses historic normalized differential vegetative index (NDVI) and thermal condition index (TCI) to calculate drought (VHI<40) and favorable conditions (VHI>60) (Kogan *et al.* 2004). We chose such a large number of environmental variables so that our analysis could identify variables potentially causative of *CBF* variation. To assess the explanatory contribution of different aspects of climate on polymorphisms in *CBF* genes across landscapes we used redundancy analysis, a multiple regression approach which analyzes the relationship between independent and dependent variables that are each multivariate (Legendre and Legendre 1998). Here those are the SNP matrix and climate data matrix, respectively. In our redundancy analysis we excluded 53 of the 477 sequences that had

incomplete climatic data. We performed this analysis in R (R Core Team 2015) using the package 'vegan' (Okasanen *et al.* 2015). We calculated the explanatory contribution (P_x) of each climate variable as described in Lasky *et al.* (2012).

We wanted to determine what types of climate variables have the strongest association with CBF polymorphism. To do this, we tested for significant differences in the proportion of CBF polymorphism variation explained between groups of highly correlated climatic variables. First we performed a hierarchical clustering of climate variables using absolute correlations to calculate Euclidean distance between variables. This clustering analysis revealed four distinct cluster groups of climate variables. Then, with P_x for each individual climate variable determined by the redundancy analysis, we compared the mean P_x between each of the four climate variable groups using Tukey adjusted 95% confidence intervals.

Molecular Evolution Analysis

We identified a strong relationship between winter temperatures and CBF polymorphism through climate variable clustering and redundancy analysis. We wanted to further investigate the evolution of CBF genes along an ecologically significant climatic gradient related to winter temperature. Minimum temperature of the coldest month (Min.Tmp.Cld.M) indicates the potential severity of cold induced stress and was used for the remaining analyses. We performed a binomial logistical regression to assess the relationship between the minimum temperature of coldest month and the probability of an accession having a *CBF* allele containing a frameshift or premature stop codon. Additionally, in order to compare molecular evolution of *CBF* genes between different climates, we grouped accessions into 5°C bins by the minimum temperature of the coldest month. We then performed a McDonald-Kreitman test on concatenated *CBF* sequences in each group using *A. lyrata CBF* sequences as an out-group. The test was carried out using the online tool developed by Egea *et al.* (2008, http://mkt.uab.es/mkt/). Here, Jukes-Cantor corrected divergence

values representing synonymous and non-synonymous nucleotide divergence from *A. lyrata* fixed in the *A. thaliana* samples were compared with synonymous and non-synonymous nucleotide polymorphisms within the samples in a 2x2 χ^2 contingency analysis, testing for significant divergence from the neutral expectation that these ratios are equal.

Haplotype networks

We concatenated all three *CBF* gene sequences for the identification of haplotypes. Median-joining networks (Bandelt *et al.* 1999) of all haplotypes with an allele frequency greater than 1% (at least four accessions) were generated with the software PopArt (http://popart.otago.ac.nz). Because PopArt cannot consider indels, we manually added two haplotypes that were distinguished by a single indel mutational event. The minimum temperature experienced in the coldest month was averaged across accessions with each haplotype in this network.

Results

Climate variable clusters

Four distinct clusters were identified in the dendrogram produced by hierarchically clustering climatic variables by their correlation to each other. An inspection of the variables within each cluster reveals that they generally group both by type of variable (temperature / precipitation) and season of variable ("Winter" / "Summer") . Thus, the four cluster groups contain variables related to or correlated with winter temperatures, summer temperatures, winter precipitation or summer precipitation.

Climate and CBF polymorphisms

To explore the relationship between climate and *CBF* variation we performed a redundancy analysis to determine the contribution (P_x) of each climatic variable in explaining genetic variation in *CBF* genes across the range of *Arabidopsis*. We then compared mean P_x of climate variables

between each climate variable cluster. 'Winter Temperature' variables had significantly greater P_x values than all other variable clusters, and 'Summer Temperature' variables had significantly greater P_x values than 'Summer Precipitation' (Figure 1). These results show that CBF polymorphism is not random with respect to climatic variables. Temperatures during cold months of the year are the most significant predictors of *CBF* molecular polymorphisms.

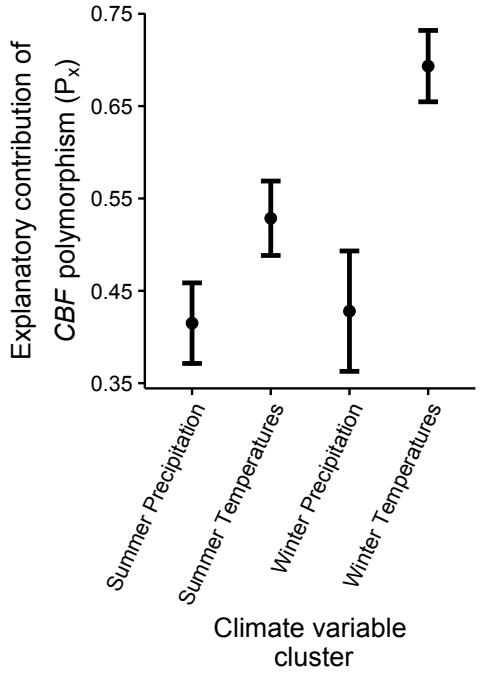

Figure 4. Means and 95% confidence intervals for variance in CBF polymorphisms explained (P_x) by different clusters of climate variables. Variance explained by each variable was determined using redundancy analysis and climate variable groups were formed by hierarchical clustering based on correlations between variables.

McDonald-Kreitman test

The results of the McDonald-Kreitman tests show that the *CBF* sequences in accessions that experience minimum temperatures between -5°C to 0°C (X^2 = 7.789, p-value = 0.005) and 0°C to 5°C (X^2 = 4.279, p-value = 0.038) are significantly enriched for non-synonymous polymorphisms. *CBF* sequences in accessions in other minimum temperature bins did not deviate from neutral expectations of sequence polymorphism (Table 1).

Table 1. McDonald – Kreitman test of molecular evolution *CBF1-3* in *Arabidopsis thaliana* accessions grouped by minimum temperature of the coldest month. *CBF* sequences from *A. lyrata* were used to calculate between species molecular divergence. Polymorphisms reflect within-species variation in the *Arabidopsis thaliana* samples.

Minimum Temperature Coldest Month	**Synonymous polymorphism**	**Non-synonymous polymorphism**	**Synonymous divergence (Jukes-Cantor adjusted)**	**Non-synonymous divergence (Jukes-Cantor adjusted)**	**X^2**	**p**
5°C to 10°C	17	19	77.74	89.73	0.007	0.93
0°C to 5°C	**24**	**51**	**75.34**	**87.61**	**4.279**	**0.038***
-5°C to 0°C	**33**	**69**	**80.53**	**81.08**	**7.789**	**0.005***
-10°C to -5°C	19	34	75.19	87.55	1.742	0.186
-15°C to -10°C	4	11	80.26	95.18	2.039	0.153
-20°C to -15°C	8	7	80.28	96.28	0.344	0.557
-25°C to -20°C	5	4	85.46	98.47	0.284	0.593

CBF haplotype network

We identified 177 unique haplotypes across the three concatenated *CBF* genes. The network displaying haplotypes with a frequency greater than 1% is displayed in Figure 5. This network suggests multiple, evolutionarily independent haplotypes associated with warmer minimum winter temperatures, including haplotypes containing premature stop codon and frameshift mutations (Figure 5).

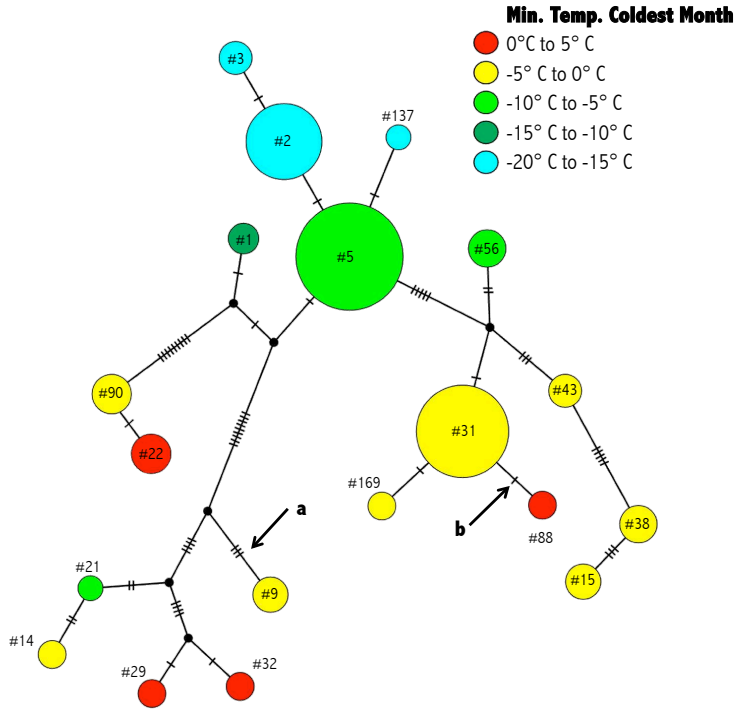

Figure 5. Median joining tree haplotype network of concatenated *CBF* genes showing haplotypes with an allele frequency greater than 1%. The size of the circles is scaled to represent haplotype frequency. Each is colored to indicate the minimum temperature of coldest month averaged across individuals exhibiting that haplotype. Solid black dots represent inferred but unsampled haplotypes. Ticks marks along the lines connecting haplotype circles represent SNP and indel mutations. Two notable mutations, unique to haplotypes also appearing in Figure 6 and labeled on the map in Figure 7, are highlighted here; a.) Premature stop codon in *CBF3* b.) Frameshift mutation in *CBF2*.

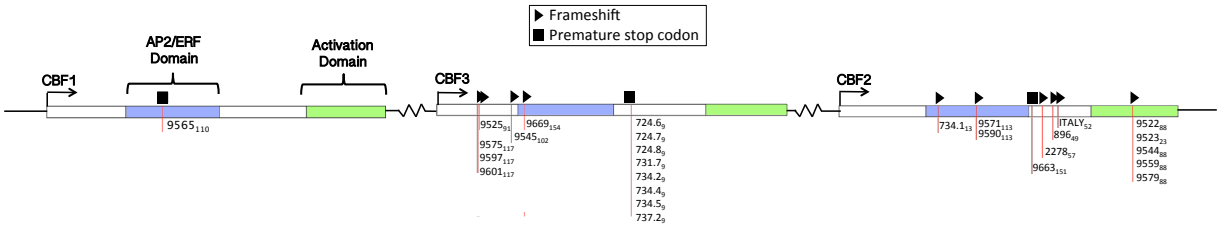

Figure 6. Locations of frameshift and premature stop codon mutations in *CBF1*, *CBF2*, and *CBF3*. Arrows show the direction of transcription. The AP2/ERF domain, which binds C-repeat promoter regions of target genes is noted in blue; the activation domain which recruits transcriptional machinery is appears in green. The subscript in the accession name indicates the *CBF* haplotype which that accession contains.

CBF frameshift and premature stop codon mutations

A frameshift mutation causing a premature stop codon in *CBF2* appears to underlie a genetic tradeoff between locally adapted populations, selectively favored in Italy (Agren *et al.* 2013; Oakley *et al.* 2014, Gehan *et al.* 2015). We found 10 additional frameshift mutations and 3 unique premature stop codon mutations (Figure 6). These mutations, which presumably have a large effect on *CBF* function, were particularly interesting given the adaptive value that a similar mutation appears to have in Italy.

Several patterns emerge from the distribution of these mutations between and within *CBF* genes. Only 1 accession with a premature stop codon was observed in *CBF1*. In contrast, 14 accessions with 5 different mutations in *CBF3*, and 12 accessions with 7 mutations were observed in *CBF2*. Interestingly, we also found that the number of accessions containing these mutations that were downstream of the AP2/ERF DNA binding domain was 70% higher (17 vs. 10) than the number of accessions with these mutations affecting the protein sequence of this domain. The affected *CBF* proteins in accessions with mutations downstream of the DNA binding domain may retain the capacity to bind promoter regions of target genes but lack the domain responsible for recruitment of transcriptional machinery for gene activation (Park *et al.* 2015).

Climate and CBF frameshift and premature stop codon mutations

The results of the redundancy analysis identified variables describing winter temperatures as having a strong association with *CBF* allelic variation. We wanted to assess the role that winter temperatures, specifically minimum temperature of coldest month, may play in driving functional changes in *CBF* genes by analyzing the climate distribution of frameshift or premature stop codon mutations. The map in Figure 7 shows the geographical distribution in Eurasia of these mutations across landscapes differing in the minimum temperature of the coldest month. Accessions grouped by 5°C bins of minimum temperature of coldest month show an increasing relationship in the frequency of individuals within each bin having one of these mutations (Figure 8). Additionally, a logistical regression reveals a significant positive relationship between minimum temperature of coldest month and the probability of an accession having a frameshift or premature stop codon mutation in one of its *CBF* genes (β=0.017721 t=3.281 p-value=0.00103)

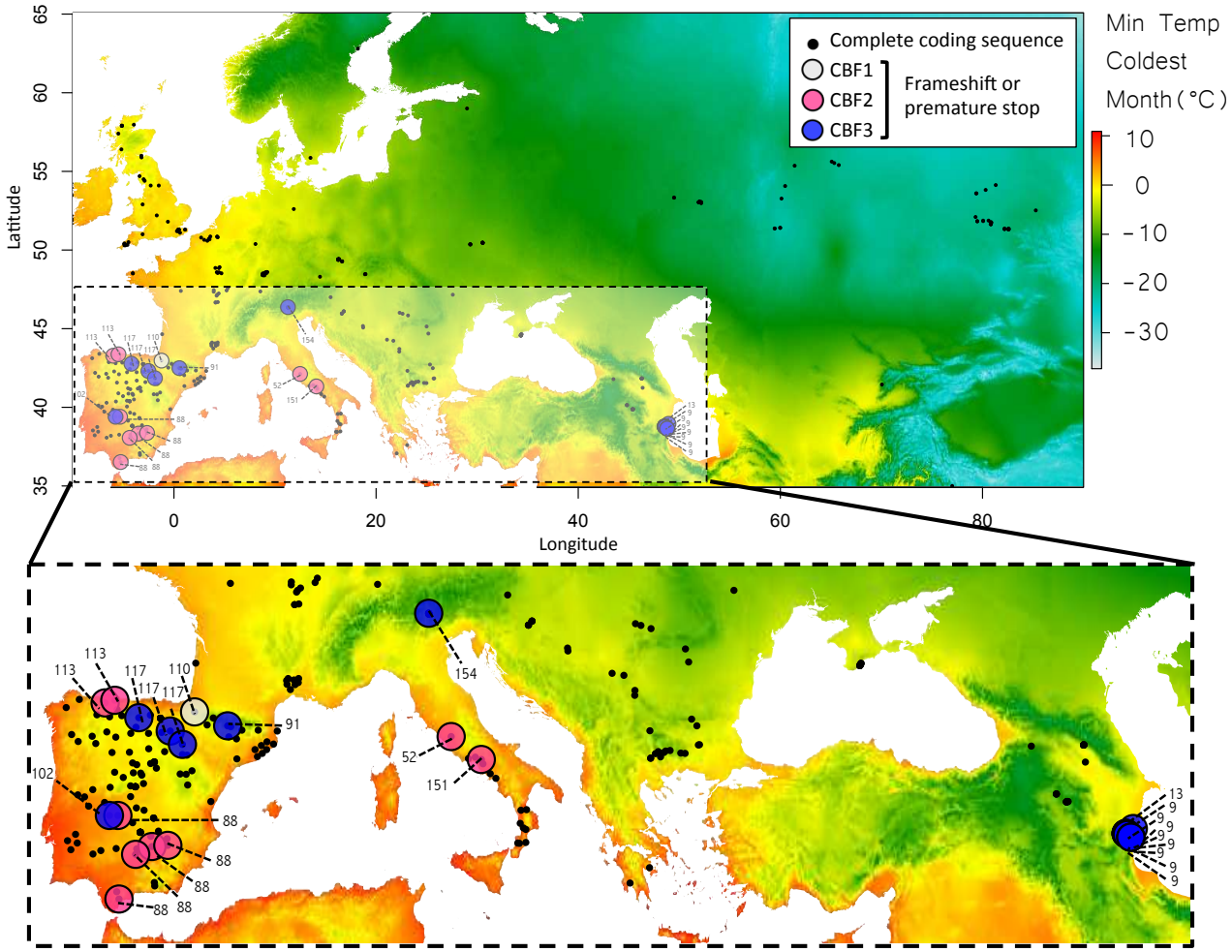

Figure 7. Map showing locations of Eurasian *Arabidopsis* with complete *CBF* coding sequences and those with *CBF* alleles containing frameshift or premature stop codon mutations. Accessions with complete *CBF* coding sequenced are marked with black dots. Accessions with a frameshift or premature stop codon mutation (see Figure 6) are colored by the *CBF* paralog in which this mutation is found and labeled by their *CBF* haplotype number. The landscape is colored according to the minimum temperature of coldest month. For the sake of visualization, only Eurasia is mapped, excluding 23 accessions from North America and Eastern Asia.

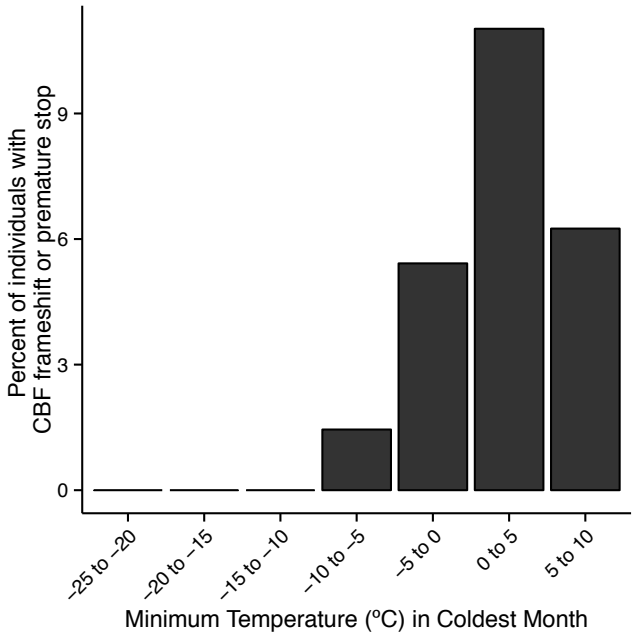

Figure 8. Percent of accessions (binned by minimum temperature of coldest month in 5°C increments) that exhibit *CBF* sequences with frameshift or premature stop codon mutations. The frequency of such mutations is significantly positively correlated with the minimum temperature of coldest month (logistic regression, **β**=0.017721 t=3.28 p-value=0.00103).

We tested for differences in climate experienced by accessions with frameshift and premature stop codon mutations differentially distributed between and within *CBF* paralogs. We compared the minimum temperature of coldest month between accessions with these mutations in *CBF2* and *CBF3* genes (Figure 9). A Mann-Whitney test was performed because a Shapiro-Wilk test revealed non-normality in minimum temperatures of *CBF2* mutated accessions (W= 0.82979, p-value = 0.02085). The Mann-Whitney test results demonstrate that accessions with *CBF2* mutations (median = 1.44°C) come from locations experiencing significantly warmer winters than those with mutations in *CBF3* (median = -0.592°C) (W = 134, p-value = 0.005347). We also tested to see if accessions containing frameshift or premature stop codon mutations affecting one versus both functional domains of *CBF* genes differed significantly in their minimum temperature of coldest month. While accessions with retained AP2 DNA binding domain were on average from

environments with warmer winters (mean = 0.724°C) compared to accessions having both domains disrupted (mean = -0.367°C) the difference was not significant (Students t-test t = 0.98882, df = 22.117, p-value = 0.1667).

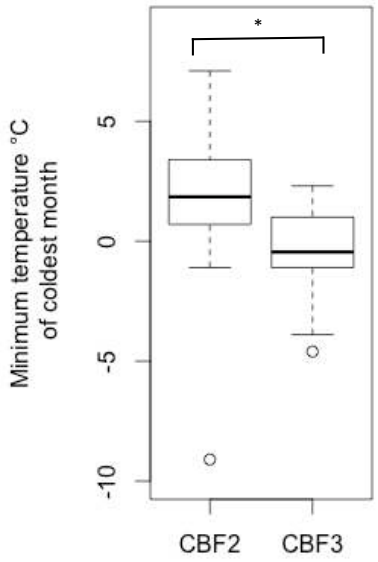

Figure 9. Comparison of the minimum temperature of coldest month experienced by accessions with *CBF2* verses *CBF3* frameshift of premature stop codons. Mann-Whitney test (W = 134, p-value = 0.005347)

Discussion

The environmental agents of selection

Here we examined polymorphism across the species range in the *CBF* transcription factors, which have been shown to play a central role in freezing tolerance and appear to underlie local adaptation in *Arabidopsis* populations at the northern and southern range limits. We found that *CBF* variation is strongly associated with variables describing winter temperatures. Looking specifically at minimum temperature of coldest month, we found that accessions from warmer climates exhibit a significant excess of non-synonymous polymorphisms and reveal a *CBF* haplotype structure consistent with multiple independent transitions to warmer climates. Furthermore, we identified a number of newly discovered *CBF* alleles containing mutations

causing major protein modification, similar to the locally adaptive frameshift mutation in Italy. We found that these alleles are significantly associated with warmer minimum winter temperatures. Lastly, we uncovered relationships between climate and the position of large functional effect mutations between and within *CBF* paralogs suggesting variation in adaptive function of different mutations.

We found that variables describing winter temperatures are predictive of polymorphism in *CBF* genes. *Arabidopsis* plants exhibiting spring, summer, or fall annual life cycles (i.e. completing life cycle within a single growing season) do not experience winter months in a vegetative state. However, a recent review of life history strategies observed in European *Arabidopsis* populations suggests that a winter annual life cycle in which plants over-winter as a vegetative rosette is the most common (Burghardt *et al.* 2015 and references therein). Accordingly, most *Arabidopsis* plants likely experience the coldest month of the year during the vegetative stage. The finding that winter temperatures are the strongest predictors of CBF variation is consistent with the hypothesis that winter temperatures are driving the evolution of *CBF* genes and logical given the known biological role of this gene in conferring freezing tolerance.

Parallel evolution and landscape heterogeneity

A central question in evolutionary biology is whether adaptation proceeds by the fixation of a single advantageous allele or in parallel by alleles with independent mutational origins (Maynard Smith & Haigh 1974; Pennings & Hermisson 2006; Chan *et al.* 2010; Haasl & Payseur 2015; Remington 2015; Ralph & Coop 2015). The McDonald-Kreitman tests reveal a significant excess of non-synonymous polymorphisms in *CBF* genes in accessions from warmer climates (Table 1). This excess of non-synonymous polymorphisms is consistent with local selection favoring loss of function alleles, leading to the retention of multiple adaptive alleles of independent evolutionary origin in warm climates. The haplotype network (Figure 5) provides a visual

demonstration that the transition to warm environments is accompanied by multiple independent trajectories of molecular evolution in *CBF* genes. These results indicate that *CBF* sequences in accessions from warmer climates are evolving non-neutrally and in parallel, supporting the hypothesis that mutations disrupting ancestral *CBF* function are locally favored environments experiencing warmer minimum temperatures. This is consistent with the observed reduced freezing tolerance in *Arabidopsis* from southern populations (Zhen and Ungerer 2008) and other examples of the evolution of paralogous transcription factor families leading to diversification of morphology (Rensing 2014) and stress responses within species (Lehti-Shiu *et al.* 2015; Des Marias *et al.* 2015). Although this study focuses only on coding sequences, a loss of function phenotype could also be achieved through mutation in regulatory regions (Alonso-Blanco *et al.* 2005).

Given that a functional polymorphism, specifically a frameshift mutation in *CBF2* is contained in a QTL explaining fitness in at least one warm climate (Ågren *et al.* 2013) and responsible for loss of freezing tolerance (Gehan *et al.* 2015), we set out to address whether this allele shows evidence of moving to fixation in warm environments. However, the 13 bp deletion that was indicated in conferring a fitness advantage in Italy was found in no other accessions in this study. Instead, we found 12 additional alleles representing other functionally disruptive mutations such as frameshift and premature stop codons in *CBF* genes, and show that they tend to occur where minimum temperatures are warmer. This is consistent with parallel adaptive evolution in which multiple functionally disruptive mutations appear independently and are locally favored by selection in warm climates.

Two important factors may influence the prevalence of hard selective sweep verses parallel evolutionary pathways: the genotype-phenotype map (Remington 2015) and the degree to which landscapes are heterogeneous with respect to selection pressures (Ralph & Coop 2015). For

example, the molecular target size for adaptive mutations may differ such that an adaptive phenotype is produced only by a particular amino acid substitution, a mutational event that occurs at low frequency. Alternatively, an adaptive phenotype may be produced by any mutation that disrupts protein function, a mutational event that occurs at a much higher frequency (Pennings & Hermisson 2006). If the mutational target size is large, such as when an adaptive allele is generated anytime a mutation causes an amino acid change that produces a protein with disrupted function, the probability of a locally favored allele being generated by mutation is greater. This effective increase in mutation rate in turn increases the probability of parallel molecular evolution in a heterogeneous landscape (Ralph & Coop 2015). Thus patchy selection pressures in a heterogeneous environment may result in adaptive evolutionary trajectories involving parallel loss of function.

It is important to recall that the 13 base pair deletion appearing to be adaptive in Italy exhibited a genetic trade off, that is; it was maladaptive in a cold Swedish climate (Ågren *et al.* 2013). Geographic heterogeneity may limit the flow of warm-adaptive alleles between populations separated by landscape features such as mountain ranges where such alleles may be deleterious during cold alpine winters. Indeed, the natural range of *Arabidopsis* is considerably heterogeneous with respect to minimum temperature of coldest month, and different loss of function mutations appear locally restricted within warm patches (Figure 7).

These findings have implications for techniques currently used to find loci under selection. Genome wide scans for selection (GWSS) continue to be a popular technique for detecting loci undergoing positive selection in nature (Haasl & Payseur 2015). Many of the methods used to scan for signatures of selection are based on models that consider the selective sweep of a single allele arising from one mutational event (Pennings and Hermisson 2006). We find support for the hypothesis that disruption of *CBF* gene function is advantageous for *Arabidopsis* accessions in

warmer climates and that this is achieved through parallel evolution rather than a hard selective sweep. While it appears that *CBF* genes are experiencing selection in nature, they would be unlikely to be detected in a genome wide scan for selection based solely on outlier SNPs. However, sliding window approaches and association with climate determined by redundancy analysis may be able to detect evidence of selection in genes evolving similarly to *CBF* (Sasaki *et al.* 2015; Nielsen *et al.* 2005; Forester *et al.* 2016). From this perspective we advise caution when interpreting negative results produced by some methods of genomic scans for selection.

CBF functional mutations

We found that the probability of an accession exhibiting a frameshift or premature stop codon in *CBF* genes increases significantly where minimum winter temperatures are warmer. Given previous reports indicating that a *CBF2* frameshift mutation is selectively advantageous in a warm Italian environment (Ågren *et al.* 2013; Gehan *et al.* 2015), the present findings may indicate that similar loss of function mutations are selectively advantageous in other warm environments within the species range. Thus, previously reported observations of major functional mutations in *CBF* genes from southern *Arabidopsis* populations (Zhen & Ungerer 2008; Kang *et al.* 2013) are corroborated here by a statistically significant link between climate and functional polymorphisms in *CBF* genes. We acknowledge the difficulty in distinguishing local or positive selection from relaxed purifying selection. Indeed, it is theoretically possible that the significant increase in *CBF* loss of function mutations in environments with warmer winters could be due to relaxed selection on *CBF* function or drift. However, field experiments suggest a significant fitness advantage conferred by a *CBF2* frameshift mutation in the field in Italy (Ågren *et al.* 2013, Gehan *et al.* 2015). Additionally, the results of the McDonald-Kreitman test performed here show that molecular evolution of *CBF* genes in warmer climate deviate from the expectation if these genes are evolving neutrally. In light of this, we feel it is likely that the significantly greater frequency

of major functional *CBF* gene mutations in *Arabidopsis* from warmer climates revealed in the present study is driven in large part by climate mediated selective pressure for *CBF* loss of function alleles.

A more detailed look at the frequency of these mutations within climate bins may yield a more nuanced understanding of the evolutionary forces acting on this gene family. Accessions frequently experiencing temperatures low enough to induce *CBF* gene expression (Zarka *et al* 2003), but rarely experiencing freezing temperatures could be most affected by any fitness penalties associated with unnecessary *CBF* induced freezing tolerance. Indeed, when we binned the accessions by 5°C increments and calculated the percent of individuals with a frameshift or premature stop codon in *CBF* genes (Figure 7) a notable peak appears in the 0°C to 5°C group. Additionally, the average (across years) minimum temperature experienced in the Italian site where a *CBF2* frameshift mutation is confirmed to be selectively advantageous falls within this range (2.7°C). Conversely, in the group of accessions experiencing minimum winter temperatures from 5°C to 10°C we observe a decline in the trend toward increasing frequency of *CBF* frameshift and premature stop codon mutations. Temperatures above 5°C may not induce *CBF* expression. Researchers studying *CBF* function have been long aware of the thermometer-like sensitivity of *CBF* expression to temperature, and temperatures below 5°C are the methodological standard for inducing *CBF* gene expression and cold acclimation experimentally (eg. Gilmour *et al.* 1998; Zarka *et al* 2003, Kang *et al.* 2013). Because accessions where minimum temperatures are greater than 5°C infrequently face temperatures low enough to induce *CBF* expression and unnecessary *CBF* induced freezing tolerance, they may experience a reduction in the selective pressure for functional disruption in *CBF* genes. This may also explain why accessions experiencing minimum winter temperatures from -5°C to 5°C show a significant excess of non-synonymous polymorphisms whereas accessions within the 5°C to 10°C group do not (Table 2). These results

suggest that the effects of loss of function on fitness may be non-linear with respect to the minimum temperature of coldest month. Disruptive functional mutations in *CBF* genes may play the greatest role in adaptation to environments where *CBF* genes are frequently induced by temperatures just above freezing, but freezing stress is rare. This supports the hypothesis that there is a fitness penalty associated with unnecessary *CBF* induced freezing tolerance, possibly due to a correlated response in an ecologically important trait such as growth rate.

The *CBF2* frameshift appearing to be adaptive in Italian *Arabidopsis* has a considerable effect on *CBF* function by eliminating its transactivation domain (Gehan *et al.* 2015). Here we found other frameshift and premature stop codon mutations are associated with warmer winter temperatures suggesting that significant function altering mutations may underlie adaptation in these genes. We also found interesting patterns in the distribution of these functional mutations between and within *CBF* paralogs. First, in contrast to frameshift and premature stop codon mutations that cause disruption of both *CBF* functional domains, we found that considerably more accessions (70% more) contained such mutations that disrupted the transactivation domain alone (Figure 6). Second, we found that the winter temperatures experienced by accessions with significant functional mutations in *CBF2* were significantly higher than those experienced by accessions with these mutations in *CBF3* (Figure 9). Furthermore, in contrast to *CBF2* and *CBF3*, we found almost no frameshift or premature stop codons in *CBF1*. This suggests that mutations in these different *CBF* paralogs, which are often assumed to be functionally redundant, may not be functionally or selectively equivalent along a climatic temperature gradient. This highlights the important role that evolutionary studies can play in elucidating gene functional variation to guide future research. A potential explanation of our findings is that significant functional mutations in *CBF2* may be most adaptive in the warmest climates, followed by functional mutations in *CBF3*, and that disruption of *CBF1* may be maladaptive across the entire species range. Additionally, the

observation that 70% more accessions with these functional mutations have intact DNA binding domains compared to accessions with loss of both domains could indicate that these mutations differentially impact phenotypes. Future empirical work will be needed to connect these findings to biology at the organismal level.

This work supports the hypothesis that divergent selection pressures in environments that differ in winter temperatures is driving adaptive functional changes in *CBF* genes. In particular, these results indicate that *CBF* genes in *Arabidopsis* from warmer winter temperatures have undergone parallel adaptive evolution involving disruption of function. This reveals how loss of function alleles generated via parallel molecular evolution may play an important role in local adaptation to rapidly changing climates.

REFERENCES

Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. *New Phytologist,* 194, 1112-1122.

Ågren J, Oakley CG, McKay JK, Lovell JT, Schemske DW (2013) Genetic mapping of local adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the USA*, 110, 21077-21082.

Alonso-Blanco C, Gomez-Mena C, Llorente F, Koorneef M, Salinas J, Martìnez-Zapater J (2005) Genetic and molecular analyses of natural variation indicate *CBF2* as a candidate gene for underlying a freezing tolerance quantitative trait locus in *Arabidopsis*. *Plant Physiology*, 139, 1304-1312.

Bandelt H, Forster P, Röhl A (1999) Median-Joining Networks for Inferring Intraspecific Phylogenies. *Molecular Biology*, 16, 37-48.

Beck JB, Schmuths H, Schaal BA (2008) Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Molecular Ecology*, 17, 902-915.

Burghardt L, Metcalf C, Wilczek A, Schmitt J, Donohue, K (2015) Modeling the Influence of Genetic and Environmental Variation on the Expression of Plant Life Cycles across Landscapes. *The American Naturalist*, 185, 212-227.

Chan YF, Marks ME, Jones FC, et al (2010) Adaptive Evolution o f Pelvic Reduction in Sticklebacks by Recurrent Deletion of a *Pitx1* Enhancer. *Science*, 327, 302-305.

Dunne KA, Willmott CJ (Dunne) Global distribution of plant-extractable water capacity of soil. http://daac.ornl.gov/SOILS/guides/DunneSoil.html. 68.

Des Marais DL, Auchincloss LC, Sukamtoh E, McKay JK, Logan T, Richards JH, Juenger TE (2014) Variation in *MPK12* affects water use efficiency in *Arabidopsis* and reveals a pleiotropic link between guard cell size and ABA response. *Proceedings of the National Academy of Science of the USA* 111: 2836-2841.

Des Marais DL, Skillern WD, Juenger TE (2015) Deeply Diverged Alleles in the *Arabidopsis* AREB1 Transcription Factor Drive Genome-Wide Differences in Transcriptional Response to the Environment. *Molecular Biology and Evolution*, 32, 956-969.

Egea R, Casillas S, and Barbadilla A (2008) Standard and Generalized McDonald-Kreitman Test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res.* 36 (Web Server issue): W157-W162.

Endler JA (1986) Natural selection in the wild. Princeton University Press

Fan Y, Li H, Miguez-Macho G (2013) Global patterns of groundwater table depth. *Science.* 339, 940–943.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, 25, 104–120.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, 334, 86-90.

Gehan MA, Park S, Gilmour SJ, An C, Lee C, Thomashow MF (2015) Natural variation in the C-Repeat Binding Factor (CBF) cold response pathway correlates with local adaptation of *Arabidopsis* ecotypes. *The Plant Journal,* DOI: 10.1111/tpj.13027

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *The Plant Journal*, 16, 433-442

Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, Hincha DK (2006) Natural genetic variation of freezing tolerance in *Arabidopsis*. *Plant Physiology,* 142, 98-112.

Haasl & Paysuer (2015) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, doi: 10.1111/mec.13339

Hereford J (2009) A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *The American Naturalist*, 173, 579-588.

Heyndrickx KS, Van de Velde J, Wang C, Weigel D, Vandepoele K (2014) A Functional and Evolutionary Perspective on Transcription Factor Binding in *Arabidopsis thaliana*. *The Plant Cell*, 26, 3894-3910.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101-104.

Hu TT, Pattyn P, Bakker EG et al (2011) The *Arabidopsis* lyrata genome sequence and the basis of rapid genome size change. *Nature Genetics*, 43, 476-481.

Kalnay E, Kanamitsu M, Kistler R et al (1996) The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* 77,437–471.

Kang J, Zhang H, Sun T, Shi Y, Wang J, Zhang B, Wang Z, Zhou Y, Gu H (2013) Natural variation of C-repeat-binding factor (CBFs) genes is a major cause of divergence in freezing tolerance among a group of *Arabidopsis thaliana* populations along the Yangtze River in China. *New Phytologist*, 199, 1069-1080.

Kogan F, Stark R, Gitelson A, Jargalsaikhan L, Dugrajav C, Tsooj S (2004) Derivation of pasture biomass in Mongolia from AVHRR-based vegetation health indices. *International Journal of Remote Sensing,* 25, 2889-2896.

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology*, 21, 5512-5529.

Legendre P, Legendre L (1998) Numerical Ecology, 2nd edn. Elsevier, New York.

Lehti-Shiu MD, Uygun S, Moghe GD, Panchy N, Fang L, Hufnagel DE, Jasicki HL, Feig M, Shiu S (2015) Molecular Evidence for Functional Divergence and Decay of a Transcription Factor Derived from Whole-Genome Duplication in *Arabidopsis thaliana*. *Plant Physiology,* 168, 1717-1734

Leimu R, Fischer M (2008) A Meta-Analysis of Local Adaptation in Plants. *PLoS ONE,* 3, DOI: 10.1371/journal.pone.0004010

Levins, R (1968) *Evolution in Changing Environments.* Princeton University Press, Princeton, NJ.

Maynard Smith J, Haigh J (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, 23, 23-35.

McKay JK, Richards JH, Nemali KS, Sen S, Mitchell-Olds T, Boles S, Stahl EA, Wayne T, Juenger TE (2008) Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, KAS-1 x TSU-1. *Evolution*, 62, 3014-3026.

Medina J, Bargues M, Terol J, Pérez-Alonso, Salinas J (1999) The *Arabidopsis CBF* gene family is composed of three genes encoding AP2 domain-containing proteins whose expression Is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiology,* 119, 463-470.

Mitchell-Olds T (2001) *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends in Ecology & Evolution*, 16, 693-700.

Mullen LM, Heoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, 62, 1555-1570.

Nachtergaele F, Batjes N (2012) Harmonized World Soil Database (FAO, Rome, 2012); http://www. fao.org/nr/water/docs/harm-world-soil-dbv7cv.Pdf

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. *Genome Research*, 15, 1566-1575.

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

Oakley CG, Ågren J, Atchinson RA, Schemske DW (2014) QTL mapping of freezing tolerance: Links to fitness and adaptive trade-offs. *Molecular Ecology*, 23, 4304-4315.

Okasanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens HH, Wagner H (2015) vegan: Community Ecology Package. R package version 2.3-0. http://CRAN.R-project.org/package=vegan

Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution* 6, 445-464.

Park S, Lee C, Doherty CJ, Gilmour SJ, Kim Y, Thomashow MF (2015) Regulation of the *Arabidopsis CBF* regulon by a complex low-temperature regulatory network. *The Plant Journal*, 82, 193-207.

Pennings PS, Hermisson J (2006) Soft sweeps II - Molecular population genetics of adaptation from recurrent mutation or migration. *Molecular Biology and Evolution*, 23, 1076-1084.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ralph PL & Coop G (2015) Convergent Evolution During Local Adaptation to Patchy Landscapes, *PLOS Genetics*, DOI:10.1371/journal.pgen.1005630

Remington DL (2015) Alleles versus mutations: Understanding the evolution of genetic architecture requires a molecular perspective on allelic origins. *Evolution,* doi: 10.1111/evo.12775

Rensing S (2014) Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology*, 17, 43-48.

Sanchez PA, Palm CA, Buol SW (2003) Fertility capability soil classification: A tool to help assess soil quality in the tropics. *Geoderma*, 114, 157–185.

Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M (2015) "Missing" G x E Variation Controls Flowering Time in *Arabidopsis thaliana*. *PLOS Genetics*, DOI:10.1371/journal.pgen. 1005597.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428, 717-723.

Stockinger EJ, Gilmour SJ, Thomashow MF (1997) *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proceedings of the National Academy of Sciences of the USA,* 94, 1035-1040.

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729.

Thomashow MF (2001) So what's new in the field of plant cold acclimation? Lots! *Plant Physiology,* 125, 89-93.

Thomashow MF (2010) Molecular basis of plant cold acclimation: insights gained from studying the *CBF* cold response pathway. *Plant Physiology*, 154, 571-577.

Uga Y, Yamamoto E, Kanno N, Kawai S, Mizubayashi T, Fukuoka S (2013) A major QTL controlling deep rooting on rice chromosome 4. *Scientific Reports* 3.

Wang Z, Triezenberg SJ, Thomashow MF, Stockinger EJ (2005) Multiple hydrophobic motifs in *Arabidopsis CBF1* COOH-terminus provide functional redundancy in trans-activation. *Plant Molecular Biology*, 58, 543-559.

Xu F, Liu Z, Xie H, Zhu J, Zhang J, Kraus J, Blaschnig T, Nehls R, Wang H (2014) Increased Drought Tolerance through the Suppression of ESKMO1 Gene and Overexpression of CBF-Related Genes in *Arabidopsis*. *PLoS ONE*, 9, e106509.

Zarka DG, Vogel JT, Cook D, Thomashow MF (2003) Cold induction of Arabidopsis CBF genes involves multiple ICE (inducer of CBF expression) promoter elements and a cold-regulatory circuit that is desensitized by low temperature. *Plant Physiology*, 133, 910-918.

Zhen Y, Dhakal P, Ungerer MC (2011) Fitness benefits and costs of cold acclimation in *Arabidopsis thaliana*. *The American Naturalist*, 178, 44-52.

Zhen Y, Ungerer MC (2008) Relaxed selection on the *CBF*/DREB1 regulatory genes and reduced freezing tolerance in the southern range of *Arabidopsis thaliana*. *Molecular Biology and Evolution*, 25, 2547-2555.

Zomer RJ, Trabucco A, Bossio DA, Verchot LV (2008) Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and re-forestation. *Agric Ecosyst Environ*, 126, 67–80.

Zuther E, Schulz E, Childs LH, Hincha DK (2012) Clinal variation in the non-acclimated and cold-acclimated freezing tolerance of *Arabidopsis thaliana* accessions. *Plant, Cell and Environment*, 35, 1860-1878

CHAPTER 4

DROUGHT ADAPTATION IN *ARABIDOPSIS THALIANA* BY EXTENSIVE GENETIC LOSS-OF-FUNCTION

Summary

Interdisciplinary syntheses are needed to scale up discovery of the environmental drivers and molecular basis of adaptation in nature. Here we integrated novel approaches using whole genome sequences, satellite remote sensing, and transgenic experiments to study natural loss-of-function alleles associated with drought histories in wild *Arabidopsis thaliana*. The genes we identified exhibit population genetic signatures of parallel molecular evolution, selection for loss-of-function, and shared associations with flowering time phenotypes in directions consistent with longstanding adaptive hypotheses 7 times more often than expected by chance. We then confirmed predicted phenotypes experimentally in transgenic knockout lines. These findings reveal the importance of drought timing to explain the evolution of alternative drought tolerance strategies and further challenge popular assumptions about the adaptive value of genetic loss-of-function in nature. These results also motivate improved species-wide sequencing efforts to better identify loss-of-function variants and inspire new opportunities for engineering climate resilience in crops.

Introduction

Discovering the environmental drivers and functional genetics of adaptation in nature is a key goal of evolutionary biology and valuable to advance applied genetics in agriculture. Understanding the genetics of drought adaptation in plants is particularly important as crop losses resulting from droughts affect billions of people each year, posing the greatest threat to global food stability. Because droughts also impose strong selection on natural plant populations, investigating

drought adaptation in wild species is both useful for addressing fundamental questions of evolutionary biology, such as determining whether adaptation proceeds by few or many alleles, and informative for efforts to reverse engineer drought tolerance in crops (Mickelbart et al. 2015). Such an evolutionary research program is motivated by the need to understand adaptive drought tolerance strategies for different types of drought conditions, which can vary in severity and timing (Tardieu 2011). Furthermore, previous limitations of single gene approaches have reinforced the necessity of developing methods to identify beneficial alleles at genomic scales and functional molecular resolutions (Dean & Thornton 2007, Passioura 2010).

Drought stress can occur throughout the year and drought timing is forecast to change over the next century (Trenberth et al. 2014). While dramatic evolutionary responses to drought events have been documented, (e.g. (Franks et al. 2007)), little is known about the relationship between drought timing and adaptation. However, the observation both in nature and agriculture that plants are particularly susceptible to drought while flowering (Nam et al. 2001, Dietrich & Smith 2016) has contributed to the longstanding hypothesis that adaptive flowering time should reflect patterns in the seasonal timing of drought events (Passioura 1996). Detailed studies of life history also reveal that locally adapted *Arabidopsis thaliana* (*Arabidopsis* hereafter) populations begin flowering in their home environments just prior to and after periods of increased historical drought frequency (Mojica et al. 2016).

Flowering time in *Arabidopsis* is correlated with other drought tolerance traits such as water use efficiency and can serve as a proxy for alternative drought tolerance strategies, with early flowering genotypes being associated with low water use efficiency (drought escape strategy) and late flowering genotypes with high water use efficiency (dehydration avoidance strategy) (McKay et al. 2003, Lovell et al. 2013, Kenney et al. 2014). Thus, the historical timing of drought experienced by locally adapted populations may explain the evolution of these strategies and the

distribution of alleles responsible for natural flowering time variation. This hypothesis motivated our investigation to identify alleles associated with drought timing and test the prediction that they contribute to adaptive flowering time evolution.

Identifying functionally relevant genetic variation contributing to adaptation is needed to understand fundamental evolutionary processes. In contrast to early theoretical predictions and popular assumptions, loss-of-function (LoF) alleles, those that eliminate or 'knockout' a gene's molecular function, are overrepresented among alleles reported as responsible for crop improvement and often produce adaptive phenotypes in wild species (Hoekstra et al. 2006, Rausher 2008, Olsen & Wendel 2013, Alonso-Blanco & Mendez-Vigo 2014, Weigel & Nordborg 2015b, Tokhamaneh et al. 2018). Indeed, a number of individual genes exhibiting evidence of locally adaptive loss-of-function have been documented in *Arabidopsis* (Grant et al. 1998, Johanson et al. 2000, Kliebenstein et al. 2001, Kroymann et al. 2003, Mouchel et al. 2004, Aukerman et al. 1997, Hauser et al. 2001, Mauricio et al. 2003, Alonso-Blanco et al. 2005, Werner et al. 2005, Barboza et al. 2013, Xiang et al. 2014).

Discovering adaptive LoF alleles is particularly valuable for inspiring targeted molecular breeding because functionally similar mutations can be mined from the breeding pool or generated directly by non-transgenic native gene editing. Unfortunately, traditional genome-wide association scans based on the one-locus two-allele model perform poorly at detecting adaptive LoF alleles, which because of the large number of mutations that can create them, are likely to arise through parallel molecular evolution (Pennings & Jermission 2005, Barboza et al. 2013, Kerdaffrec et al. 2016). Species-wide whole genome sequences however, present the opportunity to advance beyond previous mapping and scanning methods that relied on linked polymorphisms by instead characterizing and contrasting functionally defined alleles.

Here, we combined long-term satellite-detected drought histories, whole genome sequence scans based on allele function, and transgenic knockout experiments in *Arabidopsis* to test historical predictions about how drought timing shapes the evolution of flowering time and outline a broadly scalable approach for discovering loss-of-function gene variants contributing to plant climate adaptation.

Results and Discussion

To study global seasonal drought timing, satellite-detected measurements offer a valuable historical record. One such measurement, the Vegetative Health Index (VHI) has been used for decades to monitor drought, including in many places across the natural range of *Arabidopsis* (Kogan 1997). Though primarily used as a tool to predict crop productivity, by quantifying drought induced vegetative stress this index also provides a resource for evolutionary ecologists to study seasonal patterns in drought-related episodes of natural selection. We analyzed 34 years of VHI data to characterize drought regimens at the home environments of *Arabidopsis* ecotypes (Figure 10). We found that drought frequency during the spring ($ß = 50.016$, $P < 2 \times 10^{-16}$) and summer ($ß = -28.035$, $P = 4.4 \times 10^{-7}$) significantly predict flowering time among *Arabidopsis* ecotypes. We then generated a drought-timing index that quantifies the relative frequency of drought at between spring and summer over the typical reproductive growing season and observed substantial differences in drought timing experienced by ecotypes. This environmental variation presented a useful cline to address classical hypotheses about the evolution of flowering time in relation to drought timing and identify LoF alleles potentially contributing to this evolution.

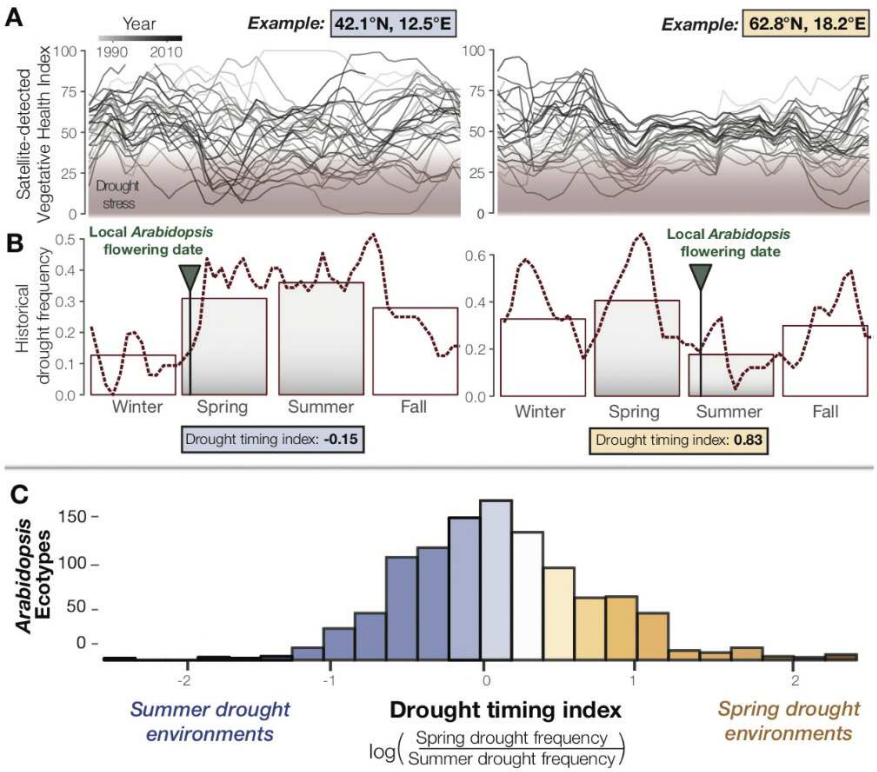

Figure 10. Seasonal drought timing varies across the *Arabidopsis* species range. (A) Examples of home environments for two well-studied *Arabidopsis* ecotypes (Mojica et al. 2016) from Italy and Sweden, left and right plots respectively, showing historical drought conditions detected using the VHI and (B) drought frequency (VHI<40, NOAA drought classification) by week (line) and season (bars). Arrows mark locally observed flowering dates (Mojica et al. 2016) and gray bars highlight the typical reproductive growing season used to quantify a drought-timing index. (C) Variation in historical drought timing experienced at the home environments of *Arabidopsis* ecotypes across the species range. Large values indicate environments where spring droughts occur more frequently than summer drought (i.e. where the frequency of drought decreases over the course of the typical reproductive growing season) and vice versa.

To identify candidate LoF alleles underlying drought adaptation and flowering time evolution, we analyzed whole genome sequences in *Arabidopsis*. We first surveyed the genomes of 1135 ecotypes (Alonso-Blanco et al. 2016) for LoF alleles in protein coding genes predicted to encode truncated amino acid sequences. To overcome the likely parallel evolutionary origins of LoF alleles that would have challenged previous methods, we classified alleles based functional allele state rather than individual polymorphisms for association testing. After filtering to reduce the likelihood of false positives (see materials and methods), we thus tested 2088 genes for LoF

allele associations with drought timing (Figure 11A) and flowering time (Figure 11B). These analyses identified 247 genes in which LoF alleles are significantly associated with drought timing and/or flowering time after accounting for population structure and multiple testing. In contrast, when we performed these analyses on a permuted LoF genotype matrix, we found no genes that were significantly associated with drought timing or flowering time.

It should be noted that the 2088 genes tested for associations to flowering time and drought timing are not a complete representation of LoF alleles in *Arabidopsis*. In some cases, previously studied LoF alleles did not pass filtering steps. This was primarily because the frequency or quality of LoF allele calls in these genes fell below our filtering requirements (see materials and methods). In other cases, the Col-0 reference genome already has a documented LOF allele. Finally, we expect LoF alleles to be undetectable if they are the product of large insertions or deletions which cannot be properly identified with currently available resequencing data. Thus, while the methods used here are designed to minimize false positives (alleles classified as LoF, but which are actually functional), the likely occurrence of false negatives (undetected LoF alleles) in available data motivates the need for more sophisticated species wide genome sequencing efforts including a greater diversity of de-novo quality genomes for comprehensive detection of functionally relevant genetic variation across the species.

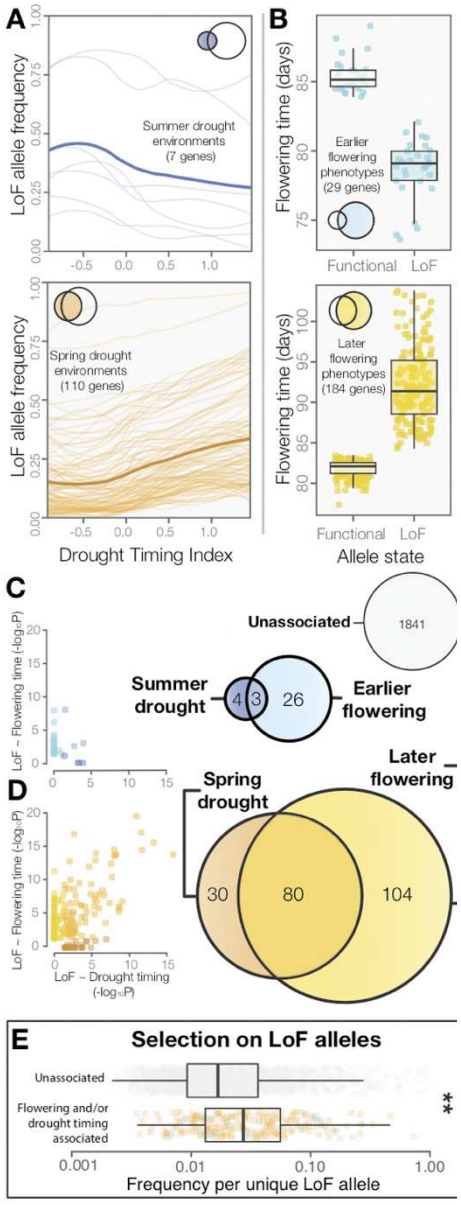

Figure 11. LoF alleles share associations between drought timing and flowering time, exhibit evidence of positive selection. (A) Visualization of the frequency of LoF alleles across environments in genes associated to summer (upper) or spring drought environments (lower). Darker lines indicate the mean across genes. (B) Contrasting flowering times between ecotypes with functional versus LoF alleles in genes associated with earlier (upper) or later (lower) flowering time phenotypes. (C) Overlap and relationships between the strength of LoF allele associations in genes associated with summer drought and earlier flowering, and (D) spring drought and later flowering. (E) Increased frequencies of independent LoF alleles in genes associated with drought timing and/or flowering time compared to genes without detected associations (t-test, $P=3.4\times10^{-7}$), a signature of recurrent mutation accompanied by positive selection (Pennings & Hermisson 2005).

Associations to drought timing predicted associations of LoF alleles to flowering time directly. Together, summer drought and earlier flowering associated genes (Figure 11C), and spring drought and later flowering associated genes (Figure 11D) overlapped 7 times more often than expected by chance (χ^2=492, P<2 $\times 10^{-16}$) and no shared associations were observed in the opposite direction. The strengths of the associations between LoF alleles and drought timing (P values) was also strongly correlated with the strengths of the associations to flowering time (r^2 = 0.48. Figure 11, Figure 11C,D). This result is comparable to overlapping peaks in a "Manhattan plot" generated from a traditional genome wide association scan (e.g. (Bosse et al. 2017)). In contrast, these associations were weakly correlated when genotypes were permuted (r^2 = 0.01 Figure 11), indicating that the result is not simply explained as an artifact of allele frequencies or by the relationship between drought timing and flowering time. Thus, satellite-detected drought histories and a functional genome-wide scanning approach prove useful for predicting the direction and molecular targets of phenotypic evolution. Similar investigations with ecologically meaningful environmental variation could be valuable for discovering candidates underlying other important traits that are especially difficult to measure.

These results further support the classical hypothesis that the relationship between phenology and drought timing is the most important feature of plant drought tolerance (Passioura 1996), indicating the evolution of "drought escape" through earlier flowering in summer drought environments, and "dehydration avoidance" by later flowering genotypes in spring drought environments. Because most *Arabidopsis* populations appear to exhibit a winter annual life habit, germinating in the fall and overwintering as a rosette (Ratcliffe 1961, Thompson 1994, Burghardt et al. 2015), late flowering genotypes in spring drought environments are expected to still encounter drought conditions. However, delayed flowering may ensure that droughts co-occur with vegetative growth rather than during the drought sensitive reproductive phase. This pattern is

also consistent with hypotheses explaining the more water conservative water use and stomatal traits observed in late flowering genotypes (McKay et al. 2003, Lovell et al. 2013, Kenney et al. 2014, Kooyers 2015) and those from spring drought environments (Dittberner et al. 2018). Future experimental work will be valuable to identify other plant physiological traits affected by the LoF alleles associated with drought timing.

These results provide new insight into the ecology and genetics of *Arabidopsis* life history evolution, but the complex ecological reality of these processes is undoubtedly beyond the scope of this study. We found that drought timing remains a significant predictor of allele associations to flowering time when controlling for allele associations with latitude and minimum temperature (slope estimate in multiple linear regression, $P < 2 \times 10^{-16}$). However, other unknown climatic variables or environmental interactions and non-linearities likely contribute to the flowering time adaptation as well. Flowering time is only one component of phenology and other adaptive life history transitions such a germination timing (Donohue 2002) may also be influenced by drought timing and could change how drought timing affects the evolution of flowering time, a hypothesis that warrants further investigation. Furthermore, measuring flowering time in other environments, such alternate light regimes, may yield a different set of candidate genes using similar approaches.

Signatures of selection in the genes identified differ from the genome average and neutral expectations. As expected for genes harboring LoF alleles, these show parallel evolution of LoF and accelerated amino acid sequence evolution among *Arabidopsis* ecotypes (Figure 11). We also found evidence of positive selection for LoF alleles in genes associated with drought timing and/or flowering time. While these genes have similar global frequencies of LoF alleles compared to genes not showing associations with drought timing and/or flowering time (Figure 11), they tend to have significantly fewer unique LoF alleles (Figure 11) and greater frequencies of each independent LoF allele (Figure 11E). This pattern is consistent with theoretical predictions and

results from simulations of adaptation by parallel molecular evolution involving recurrent mutation combined with more rapid local fixation of alleles experiencing positive selection (Pennings & Hermisson 2005). In cases where adaptation proceeds through the fixation of a single adaptive allele, traditional genome scanning approaches may be sufficient to detect causal loci. However, when genetic variation consists of multiple independent alleles, as is often the case for the genes examined here (Figure 11), classifying alleles functionally before testing for associations is likely necessary.

The extent of LoF responsible for adaptive phenotypic evolution is much greater than once assumed (Smith 1970, Albalat & Canestro 2016). LoF alleles identified were overwhelmingly associated with spring drought or later flowering rather than summer drought or earlier flowering ($\chi^2 = 132$, $P < 2\times10^{-16}$, Figure 11). Because the reference genome and gene models are from an early flowering *Arabidopsis* line, Col-0, this is consistent with the hypothesis that LoF alleles are particularly important in the evolution of phenotypic divergence (Rausher 2008). This result also highlights the need to develop functional genomics resources informed by multiple de-novo quality reference genomes. We found that flowering time is strongly predicted by the accumulation of LoF alleles across the 214 candidate genes associated to spring drought and/or later flowering time (Figure 12A-E), estimating a 1-day increase for every 3 additional LoF alleles across these candidate genes (Figure 12F). This relationship is best represented as a simple linear regression; the addition of a non-linear quadratic predictor variable did not significantly improve the fit of the model ($F= 0.7005$, $P = 0.4028$). Importantly, we did not find a broader overabundance of LoF alleles in later flowering ecotypes or those from spring drought environments that would explain this relationship (e.g. Figure 11). Rather, these findings support a model of climate-associated evolution in complex traits that includes a substantial contribution from widespread genetic LoF and give promise to targeted LoF for directed phenotypic engineering.

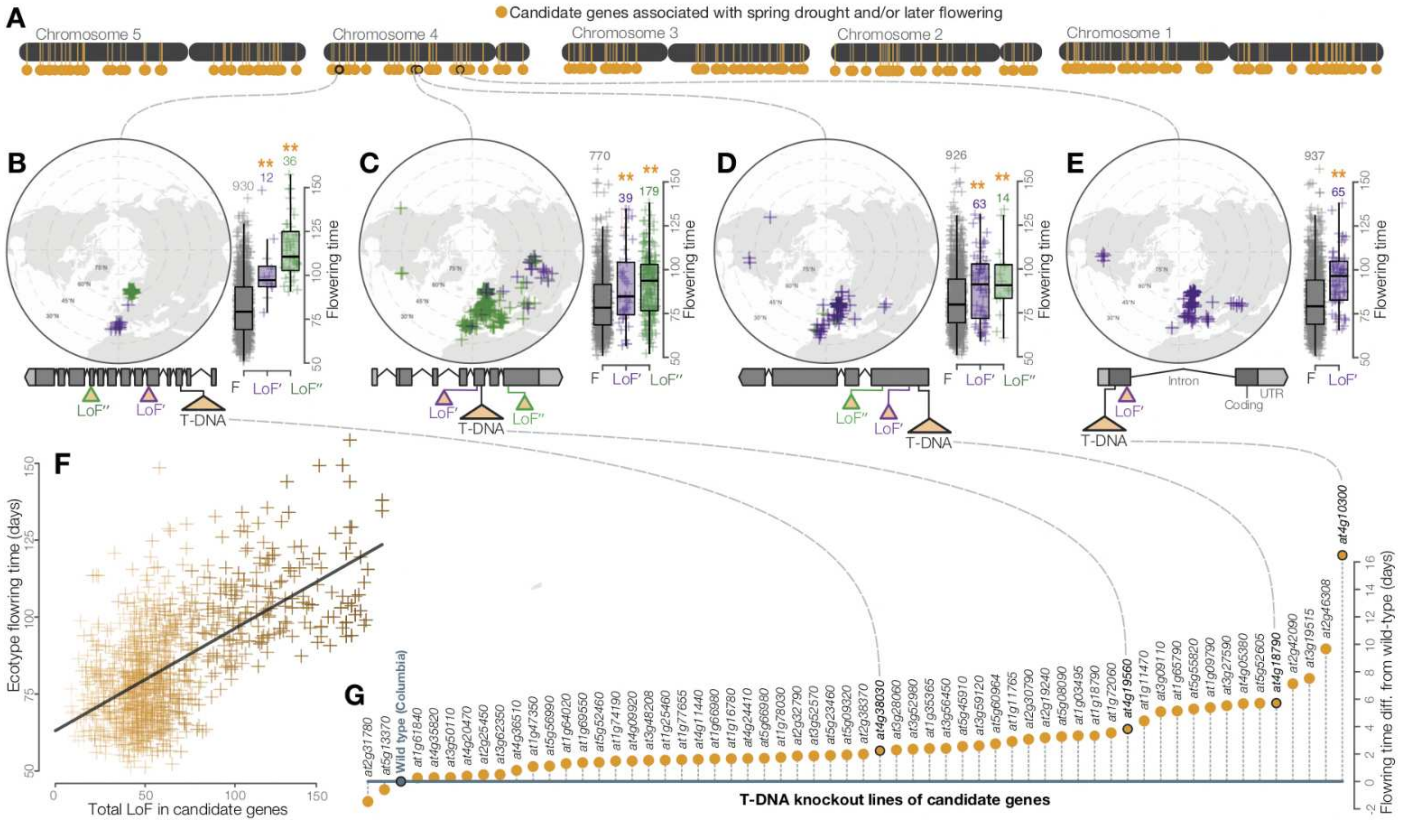

Figure 12. Widespread LoF contributing to later flowering time evolution. (A) Genomic map of 214 candidate genes with associations between LoF alleles and spring drought environments and/or later flowering time phenotypes. (B-E) Examples of the geography and flowering times among *Arabidopsis* ecotypes of LoF alleles in candidate genes including; (B) a previously unstudied rhamnogalacturonate lyase, (C) a cyclin linked to later flowering in prior knockout experiments (Cui et al. 2007), (D) members of the drought-responsive Nramp2 (Qin et al. 2017) (E) and RmlC-like cupin (Aghdasi et al. 2012) protein families. (F) Later flowering time in ecotypes predicted by the accumulation of LoF alleles across all candidate genes. The line shows the best fitting model, a simple linear regression. Color scale of points reflects proportion of total LoF in ecotypes that are candidate genes (darker points = greater proportion) (G) Experimental validation of hypothesized later flowering time in T-DNA knockout lines of candidate genes compared to the wild type genotype

Experimental knockout lines confirmed the later flowering times predicted from natural allele associations. To test phenotypic effects, we screened a panel of confirmed T-DNA insertion mutants representing a sample of candidate LoF alleles associated with spring drought and/or later flowering. As predicted by variation among *Arabidopsis* ecotypes (Figure 11D), the vast majority of knockout lines in these candidate genes (57 of 59, $\chi^2 = 51$, P = 8.045e-13) flowered later on average than the wild type genotype (Figure 12G). LoF alleles identified through these analyses and experiments include those previously linked to flowering time (Cui et al. 2007) and drought responses (Aghdasi et al. 2012, Qin et al. 2017). Implementing a functional genome-wide association scan, we find that allele associations with ecologically meaningful environmental variation (drought timing) accurately predict associations with adaptive phenotypes directly (flowering time).

Together with validation in transgenic lines, these findings outline a scalable model for gaining deeper insights into the functional genomics of climate adaptation in nature. Combining large-scale knockout experiments with functional genome wide association scans may be a valuable approach for future research to quantify the power to predict LoF allele effects. These results also further challenge historical assumptions about molecular adaptation that have implications for influencing evolutionary theory and public attitudes toward emerging molecular breeding approaches.

Groundbreaking yield increases during the green revolution of the 1960s were largely attributable to semi-dwarf phenotypes caused by LoF alleles in both rice and barley (Spielmeyer et al. 2002, Jia et al. 2009). Later it was found that natural LoF alleles of the same gene in wild *Arabidopsis* produce similar phenotypes (Barboza et al. 2013), suggesting the potential to mine ecological species for information directly useful for crop improvement. Visions of a second green revolution powered and informed by such natural variation call for discoveries in evolutionary

functional genomics at scales that have now become possible. The genes identified here could inspire future molecular breeding of climate resilient crops and this work more broadly highlights the value of integrating diverse disciplines to scale up the discovery of the climatic drivers of adaptation and functionally significant genetic variation at molecular resolutions.

Materials and Methods

Satellite-Detected Drought Histories of Arabidopsis

To study patterns in historical drought, the remotely sensed Vegetative Health Index (VHI) was used, a satellite-detected drought measurement tool whose advantage is that it includes information about vegetative impacts of drought (Passioura 1996, AghaKouchak et al. 2015). This index is based on multiple data sources from NOAA satellites, combining deviations from historic climatic (Temperature Condition Index derived from AVHRR-based observations in thermal bands) and vegetative conditions (Vegetative Condition Index derived from NDVI) to detect periods of ecological drought conditions and distinguish between other sources of vegetative stress such as cold (Kogan 1997, Kogan et al. 2005, Rojas et al. 2011). VHI was collected weekly since 1981 at 16 km^2 resolution on a scale from 0 to 100, where values below 40 reflect drought conditions (Kogan 1997) (Figure 10A). The frequencies of observing drought conditions during photoperiodic spring (quarter surrounding spring equinox), summer (quarter surrounding summer solstice), fall (quarter surrounding fall equinox), and winter (quarter surrounding winter solstice) were calculated globally from 1981 to 2015 (Figure 10B) in R (Team 2017) using the *raster* package (Hijmans 2016).

After removing ecotypes with missing location data or locations falling within pixels classified as water, seasonal drought frequencies and drought timing were calculated at the location of origin for 1,097 *Arabidopsis* ecotypes that were included as part of the 1001 Genomes Project (Alonso-Blanco et al. 2016) (Figure 10C). Up to date global map files of seasonal drought

frequency and the drought-timing index used here are available on Dryad and greymonroe.github.io/data alongside a brief tutorial showing how to extract data for points of interest in R. We tested whether seasonal drought frequencies significantly predicted with flowering time (flowering time described in subsequent section regarding LoF associations) by multiple linear regression.

To characterize the seasonal timing of droughts during an important period of *Arabidopsis'* life history, a univariate drought-timing index was generated that quantifies whether the historical frequency of drought increases or decreases over the course of the typical *Arabidopsis* reproductive growing season (Ratcliffe 1961, Thompson 1994, Burghardt et al. 2015). Specifically, this index is equal to the natural log transformed ratio between spring and summer drought frequency. More negative values reflect environments where drought frequency increases from spring to summer and are referred to here as "summer drought environments," (e.g. Figure 10B left). Conversely, more positive values reflect environments where drought frequency decreases from spring to summer and are referred to here as 'spring drought environments,' (e.g. Figure 10B right).

Loss-of-Function (LoF) Alleles in Arabidopsis Genomes

To identify functionally definitive gene variants (Hoekstra & Coyne 2007, Weigel & Nordborg 2015a, Byers et al. 2017), LoF alleles (Albalat & Canestro 2016) were identified from whole genome sequence data of 1,135 *Arabidopsis* accessions (Olson 1999, Cutter & Jovelin 2015, Alonso-Blanco et al. 2016) using R scripts. First, genes were filtered to those containing at least 5% frequency of predicted frameshift or premature stop mutations and less than 5% missing allele calls from results generated by the 1,001 Genomes Consortium (Alonso-Blanco et al. 2016) using '*SnpEff'* (Cingolani et al 2012). To reduce instances where exon skipping might ameliorate LoF mutations (Gan et al. 2011), genes were filtered to those with a single predicted gene model (Lamesch et al. 2011). Additionally, to preclude false LoF calls for cases where compensatory

mutations restore gene function or in which an insignificant portion of the final protein product is affected by putative LoF mutations (MacArthur et al. 2012), coding regions were translated into predicted amino acid sequences from which lengths from start to stop codon were calculated in R. LoF alleles were defined as those producing protein products with at least 10% lost because of late start codons and/or prematurely truncated translation. Allelic heterogeneity expected to mask these genes from traditional GWAS (Remington 2015, Monroe et al. 2016, Flood & Hancock 2017) was corrected for by classifying all alleles as either functional (0) or non-functional (1). A final frequency filter was re-applied (5% global LoF allele frequency), resulting in 2088 genes for downstream association analyses.

LoF Associations to Drought Timing and Flowering Time

To identify candidate LoF alleles responsible for climate adaptation and phenotypic evolution, the relationships between functional allele state and drought timing and between functional allele state and flowering time were evaluated for each of the 2088 genes that passed preceding filtering steps. Specifically, the association between functional allele state among *Arabidopsis* ecotypes and historical drought timing at their locations of origin was tested by logistic regression in a generalized linear model in R (Team 2017). This association study differs from traditional GWAS in several respects. First, because the alleles studied here are functionally defined, they are expected to be more likely to have a phenotypic impact than random SNPs. Second, the scope of our analyses were restricted to a subset of the genome - 2088 genes with high confidence LoF allele calls that passed previous filtering steps, rather than tens of thousands to millions of SNPs. Finally, in contrast to traditional GWAS, which is designed to identify associated chromosomal regions rather than functionally definitive genetic variations, our approach is motivated by the ability to identify alleles at molecular resolutions whose functional relevance can be tested empirically. Thus, the balance of opportunity costs related to trade-offs between false

positive and false negative associations that generally challenge GWAS are shifted to reduce false negatives rather than minimizing false positives. For these reasons, we implemented analyses based on (Price et al. 2006a) to balance false positives and false negatives. Population structure was accounted for by performing a principal component analysis on the kinship matrix among all ecotypes and including in each model the first three resulting principal components, which explain >75% of variance in relatedness between ecotypes (Price et al. 2006b). The P-values ($P_{\text{drought timing}}$) of the slope estimates ($\beta_{\text{drought timing}}$) for drought timing in these models were adjusted to account for multiple tests by a Bonferroni correction to identify those significantly associated.

Summer drought genes were identified as those in which LoF alleles are found in ecotypes that experience a significantly ($\beta_{\text{drought timing}} < 0$ & $P_{\text{drought timing}} < 0.05$) more negative drought-timing index (summer drought environments where drought frequency increases over the course of the reproductive growing season, Figure 10B left and Figure 11A top). Conversely, spring drought genes were identified as those in which LoF alleles are found in ecotypes that experience a significantly ($\beta_{\text{drought timing}} > 0$ & $P_{\text{drought timing}} < 0.05$) more positive drought-timing index (spring drought environments where drought frequency decreases over the course of the reproductive growing season, Figure 10B right and Figure 11A bottom).

The above analytical approach was repeated to test whether functional allele state is associated with the reported common garden flowering times of *Arabidopsis* ecotypes (Alonso-Blanco &Mendez-Vigo 2014). See Alonso-Blanco *et al.* (Alonso-Blanco &Mendez-Vigo 2014) for details, but in brief, flowering time was measured in growth chambers at 10°C (considerably less missing data than experiment at 16°C) under 16 hour days. Earlier flowering genes were identified as those in which LoF alleles are found in ecotypes that flower significantly ($\beta_{\text{flowering time}} < 0$ & $P_{\text{flowering time}} < 0.05$) earlier than ecotypes with a functional allele (Figure 11B top). Later flowering genes were identified as those in which LoF alleles are found in ecotypes that flower

significantly ($\beta_{\text{flowering time}} > 0$ & $P_{\text{flowering time}} < 0.05$) later than ecotypes with a functional allele (Figure 11B bottom). The preceding analyses revealed considerable overlap between genes associated with both drought timing and flowering time. To assess whether this result was an artifact of the binary LoF allele calls, we randomly permuted the genotype matrix and repeated the analyses described above, testing for significant associations between allele states and drought timing and/or flowering time. Quantile-quantile plots of P values were visualized using qqPlot in the GWAS Tools package in R (Gogarten et al. 2012) (Figure 11).

To address the longstanding hypothesis that flowering time reflects adaptation to drought timing (Fox 1990, Passioura 1996, Kooyers 2015), and the test the corresponding prediction that alleles associated with drought timing are also associated with flowering time, the groups of genes identified with significant associations to drought timing or flowering time were compared (Figure 2=11C,D). Deviation from the null hypothesis of independent associations to drought timing and flowering time was evaluated by a chi-squared test (Expected number of co-associated genes = 12, Observed = 83, χ^2=492, P=2x10^{-16}).

The magnitude of P-values have historically served as the basis of selecting candidate loci for further examination toward their contribution to environmental adaptation or phenotypic evolution in quantitative trait locus mapping and genome wide association scans [e.g. (Bosse et al. 2017)]. To test whether associations to environment (drought timing) can be used to identify loci associated with phenotypes (flowering time) directly, the correlation between log transformed P-values describing allele associations with drought timing ($P_{\text{drought timing}}$) and with flowering time ($P_{\text{flowering time}}$) was calculated (Figure 11, r^2=0.48,) and visualized separately for genes associated to summer drought/earlier flowering (Fig. 2C) and to spring drought/later flowering (Fig. 2D). To control for the possibility that allele frequencies or the relationship between drought timing and flowering time explained these observations, we also tested whether allele associations were

correlated when generated from association analyses using a matrix of randomly permuted genotypes with the same allele frequencies (Figure 11, $r^2 = 0.01$).

Finally, to control for the possibility that correlated LoF allele associations were explained by confounding environmental variables we tested whether the LoF allele associations to drought timing remained predictive while accounting for LoF allele associations with latitude and minimum temperature of the coldest month (Hijmans et al. 2005) using a multiple linear regression in R. To do so, we repeated the association analyses described in the previous section but instead tested for LoF allele associations with latitude and minimum temperatures. We then included these P values in a multiple linear regression where the strength of the association to flowering time was predicted by the associations to drought timing, latitude, and minimum temperature simultaneously.

Signatures of Selection

To assess whether histories of selection for genes identified differ from the genome wide expectation, measures of amino acid sequence evolution were evaluated for 122 genes in which loss-of-function is associated with drought timing or flowering time and for which there are orthologs identified between *A. lyrata* and *A. thaliana* (Goodstein et al. 2011). For each gene, sequences were aligned using MAFFT (Katoh & Standley 2013), codons with gaps removed, and the number of non-synonymous and synonymous polymorphisms among *A. thaliana* accessions (P_N and P_S) as well as synonymous and non-synonymous divergence (D_N and D_S) from *A. lyrata* were measured using mkTest.rb (https://github.com/kern-lab/). The ratios P_N/P_S and D_N/D_S were then calculated to measure the proportion of variants predicted to affect amino acid sequences that are segregating among ecotypes and diverged from *A. lyrata*, respectively. These calculations were also performed for genes not associated to drought timing or flowering time (n=912) and the remaining genes across the *A. thaliana* genome (n=20373) with orthologs between *A. lyrata* and

A. thaliana. To test whether genes identified show evidence of accelerated protein sequence evolution, comparisons were made to genes associated with drought timing or flowering time for both P_N/P_S (Figure 11) and D_N/D_S ((Figure 11) by two-sided students t-tests (α=0.05) in R (TEAM 2017).

Because theory predicts adaptation by loss-of-function to proceed through multiple independent alleles, but to exhibit a fewer number of different alleles than in neutral loci at similar LoF allele frequencies (Pennings & Hermisson 2005, Ralph &Coop 2010, Ralph & Coop 2015), the number of unique LoF alleles was estimated by protein length in the genes that passed preceding filtering steps. To address the hypothesis that genes in which LoF alleles are associated to drought history or flowering time are likely to reflect positive selection compared to genes in which LoF are random with respect to drought history or flowering time, the total number of unique LoF alleles between these groups was compared using a two-sided students t-test ($\log_{10}$ transformed, $P=5.8x10^{-7}$, (Figure 11). To control for the possibility that this result in an artifact of reduced frequency of LoF alleles in genes identified, the global frequency of LoF was also compared between these groups ($\log_{10}$ transformed, two-sided students t-test, P=0.11, (Figure 11). Finally, to further test the prediction that LoF alleles in genes identified have increased in frequency because of more positive selection, the frequency per specific LoF allele was compared between groups ($\log_{10}$ transformed, two-sided students t-test, $P= 3.4x10^{-7}$, Figure 11E).

Candidate Genes Contributing to Later Flowering Time by Widespread LoF

The significance of the tendency for LoF associations to spring drought/later flowering time (Figure 11D) was tested by X-squared tests (spring drought vs. summer drought, $P<2x10^{-16}$; later vs. earlier flowering, $P<2x10^{-16}$, spring drought/later flowering vs. summer drought/earlier flowering, $P<2x10^{-16}$). The chromosomal locations of candidate genes (those associated to spring drought/later flowering time) were mapped onto the *Arabidopsis* genome (Lamesch et al. 2011)

(Figure 12A). To address the hypothesis that widespread LoF contributes to later flowering time phenotypes, the total number of LoF in candidate genes for each ecotype was calculated and the correlation between this value and flowering time evaluated (Figure 12F, r^2=0.39, P<$2x10^{-16}$). We also tested whether a model which included a non-linear predictor (squared value of the total number of LoF in candidate genes) was a better fit than the simple linear model by an analysis of variance (F= 0.7005, P = 0.4028).

Experimental Testing of Predicted Phenotypes in Gene Knockout Lines

The preceding analyses provided compelling evidence of LoF in candidate genes as important in the evolution of later flowering time phenotypes. To test the prediction that non-functionalization of these genes causes increased flowering time, phenotypes were measured in transgenic lines in a subsample of candidate genes showing a significant association between loss-of-function and spring drought environments and/or later flowering time. Motivated by the general need to develop a high throughput approach of studying naturally adaptive LoF, knockout lines from the Arabidopsis Biological Resource Center were chosen from a collection created by the SALK Institute in which a T-DNA insertion in an exon of candidate genes has already been identified and confirmed to be homozygous (O'Malley & Ecker 2010, Rutter et al. 2017). These T-DNA knockout lines were generated by the SALK institute and exist in a common genetic background (Columbia) (Alonso et al 2003). Seeds were planted in 2" pots containing wet potting soil and stratified for 5 days at 4°C. Seedlings were thinned to a single plant per pot one week after stratification. Plants were grown (59 T-DNA knockout lines, 10 reps of each line and 30 reps Columbia) in a stratified (by shelf), randomized design in growth chambers (Conviron ATC60, Controlled Environments, Winnipeg, MB) under 16 hours of light at 20°C. Flowering time was measured as days after planting to the emergence of the first open flower, based on the definition of flowering time used by the 1,001 Genomes Consortium (Alonso-Blanco et al. 2016). We

calculated the least squares mean (lsmean from 'lsmeans' package in R) flowering time for each line from a mixed model where shelf and tray were included as random effects. We tested the prediction that knockout lines would flower later (have higher lsmean flowering time estimates) than the wild type Columbia genotype by a X-squared test ($P=8.1x10^{-13}$).

REFERENCES

AghaKouchak A, Farahmand A, Melton FS, Teixeira J, Anderson MC, *et al.* (2015) Remote sensing of drought: Progress, challenges and opportunities. Reviews of Geophysics, 53**,** 452-480.

Aghdasi M, Fazli F, Bagherieh MB (2012) Cloning and expression analysis of Arabidopsis TRR14 gene under salt and drought stress. Journal of Cell and Molecular Research, 4**,** 1-10.

Albalat R, Canestro C (2016) Evolution by gene loss. Nat Rev Genet, 17**,** 379-391.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, *et al.* (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science, 301**,** 653-657.

Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, *et al.* (2016) 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. Cell 166**,** 481-491.

Alonso-Blanco C, Gomez-Mena C, Llorente F, Koornneef M, Salinas J, *et al.* (2005) Genetic and molecular analyses of natural variation indicate CBF2 as a candidate gene for underlying a freezing tolerance quantitative trait locus in Arabidopsis. Plant Physiol, 139**,** 1304-1312.

Alonso-Blanco C, Méndez-Vigo B (2014) Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. Current Opinion in Plant Biology, 18**,** 37-43.

Amiguet-Vercher A, Santuari L, Gonzalez-Guzman M, Depuydt S, Rodriguez PL, *et al.* (2015) The IBO germination quantitative trait locus encodes a phosphatase 2C-related variant with a nonsynonymous amino acid change that interferes with abscisic acid signaling. New Phytologist, 205**,** 1076-1082.

Aukerman MJ, Hirschfeld M, Wester L, Weaver M, Clack T, *et al.* (1997) A deletion in the PHYD gene of the *Arabidopsis Wassilewskija* ecotype defines a role for phytochrome D in red/far-red light sensing. The Plant Cell, 9**,** 1317-1326.

Barboza L, Effgen S, Alonso-Blanco C, Kooke R, Keurentjes JJ, *et al.* (2013) *Arabidopsis* semidwarfs evolved from independent mutations in GA20ox1, ortholog to green revolution dwarf alleles in rice and barley. Proc Natl Acad Sci, USA, 110**,** 15818-15823.

Bloomer RH, Juenger TE, Symonds VV (2012) Natural variation in GL1 and its effects on trichome density in *Arabidopsis thaliana*. Mol Ecol 21**,** 3501-3515.

Bosse M, Spurgin LG, Laine VN, Cole EF, Firth JA, *et al.* (2017) Recent natural selection causes adaptive evolution of an avian polygenic trait. Science 358**,** 365-368.

Burghardt LT, Metcalf CJ, Wilczek AM, Schmitt J, Donohue K (2015) Modeling the influence of genetic and environmental variation on the expression of plant life cycles across landscapes. Am Nat, 185**,** 212-227.

Byers KJ, Xu S, Schluter PM (2017) Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? Mol Ecol, 26**,** 277-290.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly, 6**,** 80-92.

Cui X, Fan B, Scholz J, Chen Z (2007) Roles of Arabidopsis cyclin-dependent kinase C complexes in cauliflower mosaic virus infection, plant growth, and development. The Plant Cell, 19**,** 1388-1402.

Cutter AD, Jovelin R (2015) When natural selection gives gene function the cold shoulder. Bioessays, 37**,** 1169-1173.

Dean AM, Thornton JW (2007) Mechanistic approaches to the study of evolution: the functional

synthesis. Nature Reviews Genetics, 8**,** 675.

Dietrich JD, Smith MD (2016) The effect of timing of growing season drought on flowering of a dominant C4 grass. Oecologia, 181**,** 391-399.

Dittberner HA, Korte T, Mettler-Altmann APM, Weber G, Monroe, *et al.* (2018) Natural variation in stomata size contributes to the local adaptation of water-use efficiency in *Arabidopsis thaliana*. Mol Ecol, 27**,** 4052-4065.

Donohue K (2002) Germination timing influences natural selection on life-history characters in *Arabidopsis thaliana*. Ecology, 83**,** 1006-1016.

Flood PJ, Hancock AM (2017) The genomic basis of adaptation in plants. Curr Opin Plant Biol, 36**,** 88-94.

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD (2009) Population genomics of the *Arabidopsis thaliana* flowering time gene network. Mol Biol Evol, 26**,** 2475-2486.

Fox GA (1990) Drought and the evolution of flowering time in desert annuals. American Journal of Botany**,** 1508-1518.

Franks SJ, Sim S, Weis AE (2007) Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proceedings of the National Academy of Sciences, 104**,** 1278-1282.

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, *et al.* (2011) Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature, 477**:** 419-423.

Gogarten S, Bhangale T, Conomos M, Laurie C, McHugh C, *et al* (2012) GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics, 28**,** 3329-3331.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, *et al.* (2011) Phytozome: a comparative platform for green plant genomics. Nucleic acids research, 40**,** D1178-

D1186.

Grant MR, Mcdowell MJ, Sharpe AG, Zabala M, Lydiate DL, *et al.* (1998) Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. PNAS 95**:** 15843–15848.

Gujas B, Alonso-Blanco C, Hardtke CS (2012) Natural *Arabidopsis* brx loss-of-function alleles confer root adaptation to acidic soil. Curr Biol, 22**,** 1962-1968.

Hauser M-T, Harr B, Schlötterer C (2001) Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene GLABROUS1. Molecular Biology and Evolution, 18**,** 1754-1763.

Hijmans RJ (2016) Raster: Geographic Data Analysis and Modeling.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25**,** 1965-1978.

Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. Evolution, 61**,** 995-1016.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. Science, 313**,** 101-104.

Jia Q, Zhang J, Westcott S, Zhang XQ, Bellgard M, *et al*. (2009) GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. Functional & integrative genomics, 9**,** 255-262.

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

Kang J, Zhang H, Sun T, Shi Y, Wang J *et al.* (2013) Natural variation of C-repeat-binding

factor (CBFs) genes is a major cause of divergence in freezing tolerance among a group of *Arabidopsis thaliana* populations along the Yangtze River in China. New Phytol, 199**,** 1069-1080.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30**,** 772-780.

Kenney AM, McKay JK, Richards JH, Juenger TE (2014) Direct and indirect selection on flowering time, water-use efficiency (WUE, delta (13)C), and WUE plasticity to drought in *Arabidopsis thaliana*. Ecol Evol, 4**,** 4505-4521.

Kerdaffrec E, Filiault DL, Korte A, Sasaki E, Nizhynska V (2016) Multiple alleles at a single locus control seed dormancy in Swedish *Arabidopsis*. eLife, 5**,** e22502.

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene Duplication in the Diversification of Secondary Metabolism: Tandem 2-Oxoglutarate–Dependent Dioxygenases Control Glucosinolate Biosynthesis in *Arabidopsis*. The Plant Cell, 13**,** 681-693.

Kogan F, Yang B, Wei G, Zhiyuan P, Xianfeng J (2005) Modelling corn production in China using AVHRR-based vegetation health indices. International Journal of Remote Sensing, 26**,** 2325-2336.

Kogan FN (1997) Global drought watch from space. Bulletin of the American Meteorological Society, 78**,** 621-636.

Kooyers NJ (2015) The evolution of drought escape and avoidance in natural herbaceous populations. Plant Sci, 234**,** 155-162.

Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proc Natl Acad Sci USA, 100,

Suppl 2**,** 14587-14592.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, *et al.* (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research, 40**,** D1202-D1210.

Le Corre V, Roux F, Reboud X (2002) DNA polymorphism at the FRIGIDA gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. Molecular biology and evolution, 19**,** 1261-1271.

Lovell JT, Juenger TE, Michaels SD, Lasky JR, Platt A, *et al.* (2013) Pleiotropy of FRIGIDA enhances the potential for multivariate adaptation. Proc Biol Sci, 280**,** 20131043.

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, *et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. Science, 335, 823-828.

Mauricio R, Stahl EA, Korves T, Tian D, Kreitman M *et al.* (2003) Natural selection for polymorphism in the disease resistance gene Rps2 of *Arabidopsis thaliana*. Genetics, 163**,** 735-746.

McKay JK, Richards JH, Mitchell-Olds T (2003) Genetics of drought adaptation in *Arabidopsis thaliana*: I. Pleiotropy contributes to genetic correlations among ecological traits. Molecular Ecology, 12**,** 1137-1151.

Mendez-Vigo B, Pico FX, Ramiro M, Martinez-Zapater JM, Alonso-Blanco C (2011) Altitudinal and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in *Arabidopsis.* Plant Physiol, 157**,** 1942-1955.

Mickelbart MV, Hasegawa PM, Bailey-Serres J (2015) Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nat Rev Genet, 16**,** 237-251.

Mojica JP, Mullen J, Lovell JT, Monroe JG, Paul JR, *et al.* (2016) Genetics of water use

physiology in locally adapted *Arabidopsis thaliana*. Plant Sci, 251**,** 12-22.

Monroe JG, McGovern C, Lasky JR, Grogan K, Beck J, *et al.* (2016) Adaptation to warmer climates by parallel functional evolution of CBF genes in *Arabidopsis thaliana*. Mol Ecol, 25**,** 3632-3644.

Mouchel CF, Briggs GC, Hardtke CS (2004) Natural genetic variation in *Arabidopsis* identifies BREVIS RADIX, a novel regulator of cell proliferation and elongation in the root. Genes Dev, 18**,** 700-714.

Nam N, Chauhan Y, Johansen C (2001) Effect of timing of drought stress on growth and grain yield of extra-short-duration pigeonpea lines. The Journal of Agricultural Science, 136**,** 179-189.

O'Malley RC, Ecker JR (2010) Linking genotype to phenotype using the *Arabidopsis* unimutant collection. The Plant Journal, 61**,** 928-940.

Olsen KM, Wendel JF (2013) A bountiful harvest: genomic insights into crop domestication phenotypes. Annu Rev Plant Biol, 64**,** 47-70.

Olson MV (1999) When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet, 64**,** 18-23.

Passioura J (1996) Drought and drought tolerance. Plant Growth Regulation, 20**,** 79-83.

Passioura J (2010) Scaling up: the essence of effective agricultural research. Functional Plant Biology, 37**,** 585-591.

Pennings P, Hermisson J (2005) Soft Sweeps III - The signature of positive selection from recurrent mutation. PLoS Genetics preprint.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, *et al.* (2006a) Principal components analysis corrects for stratification in genome-wide association studies. Nat

Genet, 38**,** 904-909.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, *et al.* (2006b) Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics, 38**,** 904.

Qin L, Han P, Chen L, Walk TC, Li Y, *et al.* (2017) Genome-Wide Identification and Expression Analysis of NRAMP Family Genes in Soybean (Glycine Max L.). Frontiers in Plant Science, 8**,** 1436.

Ralph P, Coop G (2010) Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics, 186**,** 647-668.

Ralph PL, Coop G (2015) Convergent evolution during local adaptation to patchy landscapes. PLoS Genet, 11**,** e1005630.

Ratcliffe D (1961) Adaptation to habitat in a group of annual plants. The Journal of Ecology**,** 187-203.

Rausher MD (2008) Evolutionary transitions in floral color. International Journal of Plant Sciences 169**:** 7-21.

Remington DL (2015) Alleles versus mutations: Understanding the evolution of genetic architecture requires a molecular perspective on allelic origins. Evolution, 69**,** 3025-3038.

Rojas O, Vrieling A, Rembold F (2011) Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery. Remote Sensing of Environment, 115**,** 343-352.

Rose L, Atwell S, Grant M, Holub EB (2012) Parallel Loss-of-Function at the RPM1 Bacterial Resistance Locus in Arabidopsis thaliana. Front Plant Sci 3**:** 287.

Rutter MT, Wieckowski YM, Murren CJ, Strand AE (2017) Fitness effects of mutation: testing genetic redundancy in *Arabidopsis thaliana*. J Evol Biol, 30**,** 1124-1135.

Shindo C, Aranzana MJ, Lister C, Baxter C, Nicholls C, *et al.* (2005) Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of *Arabidopsis*. Plant Physiol, 138**,** 1163-1173.

Shindo C, Bernasconi G, Hardtke CS (2008) Intraspecific competition reveals conditional fitness effects of single gene polymorphism at the *Arabidopsis* root growth regulator BRX. New Phytol, 180**,** 71-80.

Smith JM (1970) Natural selection and the concept of a protein space. Nature, 225**,** 563.

Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (sd-1),"green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proceedings of the National Academy of Sciences, 99**,** 9043-9048.

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. Nature, 400**,** 667.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, *et al.* (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. Proceedings of the National Academy of Sciences, 101**,** 4712-4717.

Tardieu, F., 2011 Any trait or trait-related allele can confer drought tolerance: just design the right drought scenario. Journal of experimental botany 63**:** 25-31.

Team, T. R. C., 2017 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Thompson L (1994) The spatiotemporal effects of nitrogen and litter on the population dynamics of *Arabidopsis thaliana*. Journal of Ecology**,** 63-68.

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. Nature, 423**,** 74.

Torkamaneh D, Laroche J, Rajcan I, Belzile F (2018) Identification of candidate domestication-related genes with a systematic survey of loss-of-function mutations. Plant J.

Trenberth KE, Dai A, Van Der Schrier G, Jones PD, Barichivich J, *et al.* (2014) Global warming and changes in drought. Nature Climate Change, 4**,** 17.

Weigel D, Nordborg M (2015a) Population genomics for understanding adaptation in wild plant species. Annual review of genetics, 49**,** 315-338.

Weigel D, Nordborg M (2015b) Population Genomics for Understanding Adaptation in Wild Plant Species. Annu Rev Genet, 49**,** 315-338.

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, *et al.* (2005) Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. Proceedings of the National Academy of Sciences, 102**,** 2460-2465.

Xiang Y, Nakabayashi K, Ding J, He F, Bentsink L, *et al.* (2014) Reduced Dormancy5 encodes a protein phosphatase 2C that is required for seed dormancy in *Arabidopsis*. The Plant Cell**,** tpc,114.132811.

Xiang Y, Song B, Nee G, Kramer K, Finkemeier I, *et al.* (2016) Sequence polymorphisms at the rEDUCED dORMANCY5 pseudophosphatase underlie natural variation in *Arabidopsis* dormancy. Plant Physiol, 171**,** 2659-2670.

Zhen Y, Ungerer MC (2008) Relaxed selection on the CBF/DREB1 regulatory genes and reduced freezing tolerance in the southern range of *Arabidopsis thaliana*. Mol Biol Evol, 25**,** 2547-2555.

Zhu W, Ausin I, Seleznev A, Mendez-Vigo B, Pico FX, *et al.* (2015) Natural variation identifies ICARUS1, a universal gene required for cell proliferation and growth at high temperatures in *Arabidopsis thaliana*. PLoS Genet, 11, e1005085.