5-8-2024

Understanding mechanisms of local adaptation in white clover (Trifolium repens L.): insights from a phenotypic cline and population genomics

Wen-Hsi Kuo
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Ecology and Evolutionary Biology

Dissertation Examination Committee:
Kenneth Olsen, Chair
Christine Edwards
Michael Landis
Allan Larson
Ben Mansfeld
David Queller

Understanding Mechanisms of Local Adaptation in White Clover (*Trifolium repens* L.): Insights from a Phenotypic Cline and Population Genomics

By
Wen-Hsi Kuo

A dissertation presented to
Washington University in St. Louis
In partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2024
St. Louis, Missouri

Table of Content

List of Figures

List of Table

List of Supplementary Figures

List of Supplementary Tables

Acknowledgements

First and foremost, I extend my deepest gratitude to my advisor, Dr. Kenneth Olsen, whose passion for plant research sparked my own curiosity from our very first email communication. Dr. Olsen generously introduced me to the fascinating world of crop domestication and the evolution of cyanogenesis polymorphism in white clover within his laboratory. His approach combines tremendous freedom for exploration with timely and essential guidance, helping us navigate through moments of uncertainty. Notably supportive in grant proposal writing and manuscript preparation, Dr. Olsen's patience and precision in language have been especially instructive for me as a non-native English speaker. Beyond academia, he has been a friendly mentor, always welcoming with fragrant Chinese tea and an intriguing collection of live plants in his office. During a particularly challenging personal time, his support was unwavering. Dr. Olsen's dedication to teaching is evident in his meticulous preparation for the Evolution course at WashU, where it has been both a pleasure and an honor to assist and learn from him.

The journey through PhD research is rife with challenges, particularly when embarking on scarcely trodden paths. My thesis committee— Dr. Christine Edwards, Dr. Michael Landis, Dr. Allan Larsen, Dr. Ben Mansfield, and Dr. David Queller —has been instrumental in guiding this path. Their persistent, constructive feedback and reminders have kept me focused and inspired. I am profoundly thankful for their dedication and insightful contributions.

Special thanks are due to Linda Small, the Olsen lab technician, whose welcoming demeanor and patient guidance at the wet bench were invaluable. Linda's assistance with the rigorous demands of molecular experiments and her tireless support in the greenhouse were crucial to my research.

Acknowledging the indispensable support of the greenhouse staff, Michael Dyer, Michael Stephan, and Hammy Sorkin, is especially important. Their meticulous care of my clover plants and their encouragement during challenging times played a key role in my experimental work.

I am also grateful to the exceptional team of WashU undergraduates—Eimear Cunningham, Keiko Farah, Nick Ho, Grace Li, Angel Lu, Chinh Mach, Erin Reardon, Emily Shen, Emily Talkow, Simon Traub-Epstein, and my friends Chao-Cheng Kuo and Renee Wu. Their assistance during harvesting and replanting seasons was invaluable.

During my PhD training, offering mentorship proved to be one of the most rewarding experiences. It deepened my understanding of interpersonal interactions and enhanced my ability to identify and respond to the needs of others. I am grateful to Cheyenne Anderson for her dedication to the rotation project on RNA-seq analysis. My thanks also go to Eimear Cunningham and Emily Guo for their meticulous work on leaf mark and thickness documentation in their internship. Additionally, I appreciate Grace Li's adaptability and diligence during her remote internship, where she skillfully learned geoinformatics analyses amid the pandemic.

A fantastic graduate school experience is incomplete without the support of colleagues, cohorts, and friends. Within the Olsen lab, I have had the privilege of working alongside several intelligent and engaging individuals who have enriched my academic and personal journey. Not only have they contributed to stimulating discussions, but they have also provided invaluable support through the initial confusions and challenges of graduate school. I am grateful to Sara Wright, David Goad, Jordan Brock, Marshall Wedger, Brock Mashburn, and Limei Zhong for their companionship and insights.

Beyond the confines of the Olsen lab, I have been fortunate to meet and collaborate with many wonderful individuals including Philippa Tanford, Israt Jahan, Rhiannon Vargas, David Henderson, Mahal Bugay, Justin Baldwin, and James Lucas. Their friendship and support have been pillars of my graduate experience.

Additionally, I cherish the memories created with a great group of friends who shared countless hours playing badminton, hiking, camping, and enjoying other activities, making my time in St. Louis truly memorable.

Conducting research can be financially demanding, and I am immensely grateful for the generous support provided by my funding sources. I would like to express my deepest appreciation to the Division of Biology and Biomedical Sciences at Washington University and the William H. Danforth Plant Science Graduate Research Fellowship for their financial support. Additionally, I am thankful for the funding from the Taiwan Ministry of Education, which has been crucial in facilitating my research endeavors. The work presented in this dissertation also

greatly benefited from the US National Science Foundation grant IOS-1557770, awarded to Dr. Kenneth Olsen.

Last but certainly not least, my heartfelt thanks go to my family in Taiwan. Despite the physical distance, their emotional support has been unwavering. I am deeply grateful to my parents for encouraging me to step out of my comfort zone and pursue a challenging PhD in a foreign country. Their belief in my potential has been a constant source of strength. I also extend my gratitude to my sister, grandmother, and uncle for their warm welcomes and genuine care every time I return home. Their love and encouragement have been fundamental to my success.

Each of you has left an indelible mark on my personal and professional growth, and for this, I am eternally grateful. Thank you all for making this journey memorable and successful.

Wen-Hsi Kuo

2024-05-01

St. Louis, Missouri, USA

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Understanding mechanisms of local adaptation in white clover (*Trifolium repens* L.): insights from a phenotypic cline and population genomics

Division of Biology and Biomedical Sciences
Ecology and Evolutionary Biology

By
Wen-Hsi Kuo

Washington University in St. Louis, 2024
Professor Kenneth M. Olsen, Chair

Rapid global climate change threatens food production and natural ecosystems. Knowledge of the mechanisms by which species adapt to local climates can inform future agriculture and conservation efforts and help to ensure sustainability and resilience of both cultivated and wild species. The overarching goal of this dissertation is to investigate the broad range of adaptive traits and their underlying genetics in wild populations of a widely distributed and polyploid forage plant, white clover. White clover is an ideal system because of its genetic diversity, widespread distribution, and its ability adapt to various environmental stresses through mechanisms that include its well-studied polymorphism for cyanogenesis (HCN release with tissue damage). These qualities make white clover a valuable model for understanding plant adaptation and agricultural sustainability in the face of climate change. In **Chapter 1**, I investigated whether the products of cyanogenesis can be recycled into primary metabolism, functioning as additional carbon and nitrogen sources. The gene expression data suggest that cyanogenic white clover can actively recycle the product of cyanogenesis (HCN) without

herbivore-triggering reactions, supporting a beyond-herbivore-defense hypothesis. In **Chapter 2**, I documented the existence and the adaptive functions of copy number variations (CNVs) at the cyanogenesis loci in white clover based on a pangenome perspective. A key advance is *de novo* assembling a haplotype-phased genome and the integration of PacBio HiFi seq, Omni-C, linkage maps. Additionally, I documented the trajectory of karyotype evolution in the clover genus *Trifolium* and related genera in the "IRLC clade" of the legume family (Fabaceae). In **Chapter 3**, I investigated the contribution of the cyanogenesis polymorphism to the overall signatures of local adaptation in white clover. Utilizing landscape genomic approaches, I found that the cyanogenesis polymorphism is not one of the most significant features in local adaptation. Instead, phenological regulatory genes are likely to play some more important roles. In **Chapter 4**, I mapped the *V* locus and six modifier loci for the white leaf mark polymorphism in white clover, revealing a complex basis for this structure-based variegation. The apparent absence of compromised photosynthesis in variegated leaves suggests that factors other than simple fitness trade-offs may maintain this leaf mark polymorphism. In **Chapter 5**, I investigated the genetic mechanisms that underlie drought adaptation in white clover, and documented the relative contributions of cyanogenesis. I found that the cyanogenesis did not provide additional fitness benefits for white clover in drought-prone environments. Instead, the QTL mapping detected a locus showing antagonistic pleiotropy, at which the same allele could promote the flowering under drought but repress vegetative growth in water-sufficient environments. The energy allocation between vegetative and reproductive growth is likely to be the adaptive mechanism for white clover at locations of divergent levels of drought stress.

Introduction

Overall Significance

We are now faced with unprecedently rapid global climate change, which could catastrophically affect our food production and natural ecosystems if ignored (Dai, 2011; Anderson *et al.*, 2012; Dai, 2013). Numerous studies have documented increased drought and heat in East Asia, North America, Africa, and Australia in the last twenty years (Lyon & DeWitt, 2012; McGrath *et al.*, 2012; Cook *et al.*, 2014). Long-standing landraces and cultivars of many crop species may not successfully tolerate the climate of the imminent future. In addition, modern agriculture extensively relies on very few elite cultivars and large-scaled monoculture production systems, which are vulnerable to herbivores and pathogens and require extensive chemical-dependent pest control (Bianchi *et al.*, 2006). To achieve a sustainable and eco-friendly agricultural system, it is important to explore and integrate adaptations from the natural variation found in wild species, including drought and heat tolerance and herbivore pest resistance traits (Ashraf, 2010; Mitchell *et al.*, 2016; Nuccio *et al.*, 2018). Humans have demonstrated our ability to take advantage of existing variation and modify plant species to obtain desired traits that will help them withstand environmental stresses. However, the modern breeding system has greatly reduced the genetic diversity of crops and the available trait variation is limited (Keneni *et al.*, 2012). In contrast, crop wild relatives and other wild species, because they have evolved adaptations to variable environments (local adaptation), often possess high genetic diversity, which is useful for developing more productive, herbivore-

resistant and stress-tolerant crop varieties (Castañeda-Álvarez *et al.*, 2016; Dempewolf *et al.*, 2017; Zhang *et al.*, 2017; Khoury *et al.*, 2020).

In addition, wild species are as threatened by climate change, as is agricultural production is. Understanding functional trait variation and its underlying genetics in wild plant populations can play a critical role in future plant conservation. For example, by assisting migration of selected genotypes to increase the adaptability of local populations, genetically-informed conservation strategies can help wild species withstand environmental stress. Recent landscape genomics approaches leverage next-generation sequencing technologies and widely available environmental data to calculate the association between allelic compositions and local environments (Balkenhol *et al.*, 2019).

The Values of Studying Local Adaptation

Local adaptation is the fitness tradeoff of populations in different environments, where genotypes in their native habitats exhibit higher fitness compared to foreign genotypes (Kawecki & Ebert, 2004). Natural selection that promotes local adaptation will increase population differentiation while gene flow could hinder local adaptation by homogenizing the populations. Among the different study systems that can potentially be used to study local adaptation, species that have repeatedly and independently evolved similar phenotypic clines across an environmental gradient in different parts of their range have been held up as a "best-case scenario" for detecting locally-adapted traits and their underlying genetic basis (Lasky *et al.*, 2022). One of the classic examples is the pigmentation clines in the old field mouse

(*Peromyscus polionotus*), showing strong correlations between fur pigmentation and soil reflectance despite a high level of gene flow and low population structure (Mullen & Hoekstra, 2008). Similar pigmentation-based clines have also been documented widely, such as European barn owl (Antoniazza *et al.*, 2010; Burri *et al.*, 2016), fruit fly (Wittkopp *et al.*, 2011), British peppered moth (Hof *et al.*, 2016), American bellflower (Koski & Galloway, 2018), and a marine snail (Gefaell *et al.*, 2023). Finally, one interesting cline can be found in white clover, where a latitudinal cline in cyanogenesis has evolved independently in both its native and introduced range (Kooyers & Olsen, 2012; Kooyers & Olsen, 2013). In such species, the repeatedly-evolved clinal variation is more likely to reflect a genuine signal of environmental adaptation than to be an artifact of neutral population structure or demographic history.

Understanding the phenotypic and genetic bases of local adaptation has emerged as a major focus in evolutionary biology (Kawecki & Ebert, 2004; Savolainen *et al.*, 2013; Lascoux *et al.*, 2016; Lasky *et al.*, 2022). One common approach is to employ reciprocal transplants and/or common garden experiments to evaluate fitness trade-offs between individuals originating from different environments (Stinchcombe *et al.*, 2004; Agrena *et al.*, 2013; Wright *et al.*, 2018; Wright *et al.*, 2022). When combined with genotypic information, such as quantitative trait locus (QTL) mapping of fitness traits, this approach can elucidate the genetic architecture of locally adaptive phenotypes, such as antagonistic pleiotropy (Agrena *et al.*, 2013; Grillo *et al.*, 2013; Wright *et al.*, 2022). With advances in genome sequencing technologies and the availability of global climatic data, a complementary approach for studying local adaptation is to employ landscape genomic methods to identify genomic regions with signatures of local

adaptation; these may include loci that show evidence of adaptive differentiation between environments (e.g., F_{ST} outliers) or strong environmental correlations as detected in genotype-environment association (GEA) analyses (Guerrero *et al.*, 2018; Gugger *et al.*, 2018; reviewed in Lasky *et al.*, 2022; Battlay *et al.*, 2023; Wang, Y *et al.*, 2023). Thus, through the integration of classic evolutionary principles with cutting-edge genomic technologies, research on local adaptation not only sheds light on the complex mechanisms driving natural variation but also underscores the importance of evolutionary processes in shaping the intricate relationships between organisms and their ever-changing environments.

Polyploidization as a Resource of Evolvability in Local Adaptation

Polyploidization (whole genome duplication) has been proposed to confer an extra genetic toolkit that can facilitate trait diversification and adaptations (Van de Peer *et al.*, 2017). In plants, ancient polyploidization events may have led to a dramatic increase in species richness in several angiosperm lineages, including the Poaceae, Solanaceae, Fabaceae, and Brassicaceae (Soltis *et al.*, 2009). During the worldwide emergence of agriculture ~10,000 years ago, humans may have uncovered the high evolvability of polyploid plant species and intentionally selected for agriculturally desirable traits. This unconscious association may be the reason why a large proportion of modern crops are polyploid (reviewed in Udall & Wendel, 2006). Examples include wheat (Dubcovsky & Dvorak, 2007), sweet potato (Yang *et al.*, 2017), peanut (2n = 4x = 40) (Zhuang *et al.*, 2019), sugarcane (2n = 8x = 80) (Zhang *et al.*, 2018), some mustard crops (*Brassica napus*, *B. juncea* and *B. carinata*, 2n = 4x = 34-38), and cotton (2n = 4x = 52) (Jiang *et al.*, 1998). In addition to agricultural systems, polyploidization has been well

documented in natural environments as a way to instantaneously establish reproductive isolation and speciation from the parental species (Wood *et al.*, 2009). Polyploidy-driven speciation is often associated with niche expansion or shifts (Blaine Marchant *et al.*, 2016; Molina-Henao & Hopkins, 2019), which is expected to occur in the very early stages of polyploidization to mitigate the competition between the incipient polyploid species and its parental species (Levin, 1975). For example, *Arabidopsis arenosa* is an autopolyploid species that shows niche expansion compared to its diploid progenitor (Molina-Henao & Hopkins, 2019). This fast niche evolution parallels the rapid modification of crop species in domestication, which both present high evolvabilities. The high adaptability could be attributed to having greater phenotypic plasticity (Hahn *et al.*, 2012; Wei *et al.*, 2019), and/or a greater genetic potential to evolve adaptations in response to local selective pressures (Selmecki *et al.*, 2015; Godfree *et al.*, 2017). Several genetic mechanisms have been proposed to explain polyploid advantage, including relaxation of purifying selection (Baduel *et al.*, 2019) and additional interactions between the homeologous genes (Santantonio *et al.*, 2019). Understanding the molecular mechanisms by which polyploidy confers enhanced evolvability is important in future crop improvement, plant conservation, and advancement of biological knowledge.

Cyanogenesis as a Defense System and Beyond

Cyanogenesis – where a plant can actively release HCN when the tissue is ruptured – is an important herbivore defense trait in over 3,000 plant species across 130 flowering families, including almond, apple, cassava, lima bean, and sorghum (Gleadow & Moller, 2014). In

cyanogenic species, tissue damage that causes cell rupture brings together two cyanogenic precursors, cyanogenic glucosides (CNglcs) and their hydrolyzing enzyme (cyanogenic β-glucosidases), which triggers the liberation of HCN. Cyanogenesis is well documented as a mechanism for generalist herbivore deterrence. However, CNglcs can also serve as sources of reduced nitrogen that can be metabolized through non-cyanogenic pathways (i.e., primary metabolism), and they may be especially valuable in drought prone environments, as water stress limits nitrogen uptake from soil (Machingura *et al.*, 2016). Recent studies have shown that accumulation of CNglcs is associated with drought tolerance in sorghum (Burke *et al.*, 2013) and tobacco (Liang, 2003). HCN may also function as a systemic signal molecule to environmental stress. The cyanide molecule could directly modify protein activity through *S*-cyanylation (García *et al.*, 2019) or indirectly interact with alternative oxidase (AOX) to regulate ROS production in mitochondria (Xu *et al.*, 2012).

In plants, β-cyanoalanine synthase (β-CAS) is the major HCN recycle (detoxification) enzyme; it catalyzes the assimilation of β-cyanoalanine from HCN and cysteine (Blumenthal-Goldschmidt *et al.*, 1963). β-cyanoalanine can then be processed to asparagine and to aspartate, thereby returning the nitrogen to the pool of primary metabolites (Machingura *et al.*, 2016). Unlike β-CAS, a second enzyme, mitochondrial alternative oxidase (AOX), is not directly involved in HCN recycle. However, it may provide a critical, temporary tolerance to a surge of intracellular HCN levels. While HCN is impairing mitochondrial cytochrome c oxidase, AOX can mitigate cyanide intoxication syndrome by providing an alternative route to the electron transport chain, which directly relaxes the over-reduced ETC although at the cost of a greatly

reduced electrochemical proton gradient and ATP turnover rate. Despite these energetic shortcomings, AOX can provide a short-term fix by maintaining metabolic homeostasis for plant cells during oxidative stress, especially when the cytochrome pathway is temporarily unavailable (reviewed inMcDonald, 2008). AOX has been shown to work synergistically with β-CAS to withstand oxidative stress in tobacco (Yu *et al.*, 2020).

Study System – White Clover

White clover, historically dubbed the "agricultural equivalent of coal" for its pivotal role in nitrogen fixation on farms, stands as a cornerstone forage legume across Europe, North America, New Zealand, and East Asia (Daday, 1958; Kjærgaard, 2003). In addition to its direct economic value, white clover has several important features that make it an excellent model system for probing adaptive traits crucial for both agricultural enhancement and the conservation of wild plant species.

Firstly, this species has been shown to rapidly adapt to an extremely wide range of environments. Originally native to southern Europe, white clover has successfully established itself in a diverse range of global climates, from the cold reaches of the boreal zones to the warm subtropics. This widespread adaptability is likely bolstered by its obligate-outcrossing mating system and short generation time, which together enable rapid local adaptation (Wright *et al.*, 2018; Wright *et al.*, 2022).

Secondly, white clover is characterized by a well-studied adaptive polymorphism for the herbivore defense mechanism – cyanogenesis. This defense mechanism, which involves the release of HCN upon tissue damage, presents a fascinating system for studying the balancing selection between varying herbivore pressure and energy cost (Daday, 1965; Kakes, 1989). The genetic basis for this trait involves a two-locus system controlling the production of cyanogenic glucosides and its hydrolyzing enzyme (described below), a setup that exemplifies classical Mendelian genetics.

Thirdly, white clover's status as an allotetraploid species, derived from the hybrid of two distinct diploid progenitors, enriches its value as a model organism. By combining the adaptive advantages from both parental species, this polyploid complexity has likely facilitated white clover's ability to thrive in a vast array of environments, far surpassing the ecological niches of its ancestors (Griffiths *et al.*, 2019). White clover is thus well suited as a system for expanding our knowledge of how subgenomes within polyploids contribute to the ability of species to adapt to a wide range of environments.

Finally, white clover is readily crossed by hand-pollination and can also be easily propagated vegetatively to create clonal replicates for experiments. It is also genetically tractable, with linkage maps and high-quality reference genomes available (Griffiths *et al.*, 2019; Santangelo *et al.*, 2023). In conclusion, white clover stands at the forefront of species suited for in-depth genetic and ecological studies, offering direct implications for agricultural practices and plant conservation strategies.

Genetic Mechanism of the Cyanogenesis Polymorphism in White Clover

The cyanogenic polymorphism in white clover is controlled by two unlinked genetic polymorphisms which are responsible for the presence/absence of the two required biochemical components. The first locus (*Ac*/*ac*) contains a three-gene metabolic cluster required for cyanogenic glucoside synthesis (specifically, linamarin and lotaustralin), which are stored in the vacuoles of photosynthetic tissue (Olsen & Small, 2018). The second, unlinked locus (*Li*/*li*) encodes their hydrolyzing enzyme, linamarase (a cyanogenic β-glucosidase), which is present in the apoplast and can hydrolyze the CNglcs and release toxic HCN. Homozygote recessive genotypes (*acac* or *lili*) are acyanogenic. They carry gene deletions at the *Ac* or *Li* loci, respectively, and therefore lack the ability to synthesize the necessary cyanogenic precursors (Olsen *et al.*, 2007; Olsen *et al.*, 2008). Thus, only the cyanogenesis phenotypes (or "cyanotypes") that possess at least one functional allele at both loci ("AcLi" plants) are cyanogenic. In total, four different cyanotypes occur in nature: the cyanogenic form ("AcLi", with both cyanogenic precursors present), and three acyanogenic forms ("Acli", lacking linamarase; "acLi", lacking cyanogenic glucosides; and "acli", lacking both components).

Chapters of the Dissertation

The overarching aim of this dissertation is to investigate the broad range of adaptive traits and their corresponding genetics in wild populations of a widely distributed and polyploid forage plant, white clover. In addition to providing insights into the genomic basis of adaptation

in plants, I provide implications in future agricultural improvements and conservational practices.

Chapter 1: White clover is characterized by an adaptive polymorphism in cyanogenesis and thus serves as an excellent system to address the question of whether cyanogenic plants can actively recycle CNglcs and/or HCN through a detoxification pathway. I investigated the hypothesis that cyanogenic white clover plants should have a higher β-CAS activity than non-cyanogenic plants to recycle the internal CNglc and/or HCN they possess. The recycling pathway may also induce AOX activity in order to increase cellular HCN tolerance in case HCN leaking occurs during the detoxification process. I used white clover to test the following specific hypotheses: Cyanogenic (AcLi) white clover has higher expression of β-*CAS* and *AOX* genes than acyanogenic white clover, even in the absence of tissue damage. This hypothesis aligns with the hypothesis that CNglcs and HCN have major functions not just in herbivore defense but also as primary metabolic components that are potentially used as additional carbon and nitrogen sources. *Alternatively,* cyanogenic white clover does not show higher expression of β-*CAS* and *AOX* in the absence of tissue damage, and instead expression increases only after tissue damage. This hypothesis aligns with the hypothesis that CNglcs and HCN function primarily or entirely in herbivore defense, and that HCN detoxification occurs only under conditions where the chemical defense has been triggered. My findings suggest a heightened constitutive role for HCN detoxification (via elevated β-*CAS* expression) and HCN-toxicity mitigation (via elevated *Aox2a* expression) in plants that are capable of cyanogenesis.

Chapter 2: White clover's importance both in agriculture and as an eco-evolutionary model system has spurred two recent independent genome sequencing projects that have utilized long-read sequencing technologies to overcome the scaffolding difficulties associated with polyploid genome assembly (Santangelo *et al.*, 2023; Wang, H *et al.*, 2023). While these two high quality genomes are a major step in white clover studies, the recent growth of plant pangenome projects (e.g. Alonge *et al.*, 2020; Zmienko *et al.*, 2020; Qin *et al.*, 2021) has demonstrated clearly that *multiple* high quality reference genomes are required to properly understand the genomic structure and intraspecific diversity of a species; this is especially true when patterns of genomic structural variation contribute to natural phenotypic variation.

Among the many types of structural variation that can occur in a genome, copy number variation (CNV), defined as the variable repetition of specific sequence motifs ranging from 50 bp to several Mbp, is a major contributor to both genetic and phenotypic variability across eukaryotes (Wright *et al.*, 2009; Cook *et al.*, 2012; Wang *et al.*, 2015; Pos *et al.*, 2021; Stalder *et al.*, 2023). Based on the prior knowledge of white clover and the genetic basis of adaptation in this species (Olsen *et al.*, 2007; Olsen *et al.*, 2008; Olsen & Small, 2018), I hypothesized that local environmental adaptation in wild populations could arise, in part, through CNVs, particularly at the cyanogenesis loci. I generated and *de novo* assembled a new chromosome-scale and haplotype-resolved genome by PacBio HiFi, Omni-C and two linkage maps. Structural variation at the cyanogenesis loci among three high quality white clover genomes were then compared (this study; Santangelo *et al.*, 2023; Wang, H *et al.*, 2023). Next, I assessed *Ac* and *Li* CNV occurrence and distributions across the North American species range. In a

complementary analysis, I examined the contribution of CNVs to natural phenotypic variation by assessing the relationship between *Ac* CNVs, gene expression, and cyanogenic glucoside content. I further tested whether CNVs at the cyanogenesis loci contribute to adaptation in nature by conducting association analyses between the CNVs and the local environments of the sampled wild populations. Finally, I leveraged the high-quality genome to address questions concerning karyotype evolution in the clover genus *Trifolium* and related genera in the "IRLC clade" of the legume family (Fabaceae). A key finding of this chapter is that CNVs at the cyanogenesis loci contribute to quantitative variation in the cyanogenic phenotype and to local adaptation across wild North American populations.

Chapter 3: A clear indicator of white clover's adaptive ability is that populations worldwide have evolved climate-associated cyanogenesis clines. Despite the apparent strong selection favoring the evolution of climate-associated cyanogenesis clines, previous studies have not implicated the cyanogenesis polymorphism or its underlying genes as significant contributors to local climatic adaptation (Wright *et al.*, 2018; Wright *et al.*, 2022). This chapter utilizes white clover's high-quality genome from the second chapter to conduct genotype-environment association (GEA) analysis and genomic differentiation scans to test for signatures of local adaptation and the extent to which they do or do not involve cyanogenesis loci in wild populations.

Using a geographically representative sample of white clover populations across North America (415 accessions from 43 locations), I determined that the genetic differences across

populations were more strongly associated with isolation-by-environment (IBE) rather than isolation-by-distance (IBD). As this pattern implicates environmental selection as a factor shaping the distribution of genotypes (Wang & Bradburd, 2014), I then examined whether the variation in cyanogenesis—a chemical defense mechanism—across populations mirrored the broader patterns of genome-wide population differentiation, or if it displayed unique, locus-specific variation. I also explored which environmental variables most accurately predicted cyanogenesis frequencies, and whether Genome-Environment Association (GEA) studies and genetic differentiation scans pinpointed the cyanogenesis loci as significant genetic factors contributing to local environmental adaptation. Then, I compared the overlaps between fitness QTL for local adaptation identified in a previous study (Wright *et al.*, 2022) and genomic regions detected by GEA and genetic differentiation scans. Finally, I examined the candidate genes of known functions near the detected signals. I find clear evidence of local adaptation, with temperature-related climatic variables best describing genome-wide differentiation between sampling locations. The same climatic variables are also strongly correlated with cyanogenesis frequencies and gene copy number variations (CNVs) at cyanogenesis loci. However, landscape genomic analyses indicate no significant contribution of cyanogenesis loci to local adaptation. Instead, several genomic regions containing promising candidate genes for local climatic adaptation are identified — some of which are shared with previously-identified QTL for locally-adaptive fitness traits in North American white clover.

Chapter 4: Polymorphism in leaf variegation, characterized by heterogeneous color patterns on leaves, is observed across diverse natural and cultivated settings. White clover is naturally

polymorphic for several leaf variegation traits, including white leaf mark, red fleck, and red midrib (Brewbaker, 1955; Carnahan *et al.*, 1955; Brewbaker & Carnahan, 1956; Corkill, 1971). In this chapter, I documented the white leaf mark polymorphism using a quantitative measure of the phenotype, accounting for the plasticity of developmental stages and intra-individual variation in an intercrossed mapping population with a homogeneous genetic background. Since the leaf mark has been shown to be structural, arising from enlarged intercellular or sub-epidermal spaces (Carnahan *et al.*, 1955), I also investigated the relationship between leaf thickness and the strength of the leaf mark phenotype. The hypothesis is that greater leaf thickness may correlate with a more pronounced leaf mark, as it could facilitate the development of larger intercellular or subepidermal spaces (e.g., Chen *et al.*, 2017). Then, I mapped the QTLs for leaf mark and leaf thickness. If the presence of leaf mark is associated with higher leaf thickness, I would expect that the QTLs for the two phenotypes would overlap to some degree. Finally, I tested for any evidence of fitness drawbacks associated with the presence of leaf mark, since in principle the leaf mark could compromise photosynthetic efficiency due to a reduction of mesophyll cell density or chlorophyll content within the leaf mark. I successfully mapped the major locus governing the white leaf mark in white clover and several modifier loci, revealing a complex basis for this structure-based variegation. The apparent absence of compromised photosynthesis in variegated leaves suggests that factors other than fitness tradeoffs may maintain the leaf mark polymorphism in white clover.

Chapter 5 (Appendix): Cyanogenesis may confer fitness advantages in drought-prone environments in addition to serving as an herbivore defense (Burke et al., 2013; Kooyers et al.,

2014; Rosati et al., 2019; Shehab et al., 2020; Myrans et al., 2021; Hayes et al., 2023). The HCN and /or its chemical precursor, CNglcs, can serve as a reservoir of reduced-state nitrogen, which can be especially beneficial to plants during periods of drought stress (Møller, 2010; Machingura et al., 2016; Nielsen et al., 2016; McMahon et al., 2021). In this chapter, I investigated the genetic mechanisms that underlie drought adaptation in white clover and the contribution of cyanogenesis variation to this phenotype. Using two geographically distinct wild genotypes as parental lines, I generated an intercrossed F_3 mapping population to establish a homogeneous genomic background. The phenotypes of the parental accessions, including morphological, physiological, and gene expression traits were documented in order to quantify differences in how they respond to drought vs. control environmental conditions. I then applied a non-lethal and progressive drought stress treatment to the mapping population in controlled greenhouse conditions, quantified fitness-related traits, and compared the performance of plants with different cyanotypes. Quantitative trait locus (QTL) mapping was conducted in order to investigate the genetic architecture of drought-stress responses, and whether the drought responsive QTLs overlap with the cyanogenesis loci. Finally, I compared the cyanogenesis loci and the drought-responsive QTLs to the signals of genotype-environmental association analyses (GEA) of a wide sampling in North America representing sites exhibiting a broad range of aridity index (the samplings are shared with Chapter 2). The key finding of this chapter is that cyanogenesis did not provide additional fitness benefits for white clover in drought-prone environments. Instead, the QTL mapping showed antagonistic pleiotropy of a QTL not involved in cyanogenesis, at which the same allele promoted flowering under drought but repressed vegetative growth in water-sufficient environments.

Reference

Agrena J, Oakley CG, McKay JK, Lovell JT, Schemske DW. 2013. Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **110**(52): 21077-21082.

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**(1): 145-161 e123.

Anderson JT, Panetta AM, Mitchell-Olds T. 2012. Evolutionary and ecological responses to anthropogenic climate change. *Update on Anthropogenic Climate Change* **160**(4): 1728-1740.

Antoniazza S, Burri R, Fumagalli L, Goudet J, Roulin A. 2010. Local adaptation maintains clinal variation in melanin-based coloration of european barn owls (*Tyto alba*). *Evolution* **64**(7): 1944-1954.

Ashraf M. 2010. Inducing drought tolerance in plants: recent advances. *Biotechnology Advances* **28**(1): 169-183.

Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nature Communications* **10**(1): 5818.

Balkenhol N, Dudaniec RY, Krutovsky KV, Johnson JS, Cairns DM, Segelbacher G, Selkoe KA, von der Heyden S, Wang IJ, Selmoni O, et al. 2019. Landscape genomics: Understanding relationships between environmental heterogeneity and genomic characteristics of populations. In: Rajora OP ed. *Population Genomics: Concepts, Approaches and Applications*. Cham: Springer International Publishing, 261-322.

Battlay P, Wilson J, Bieker VC, Lee C, Prapas D, Petersen B, Craig S, van Boheemen L, Scalone R, de Silva NP, et al. 2023. Large haploblocks underlie rapid adaptation in the invasive weed *Ambrosia artemisiifolia*. *Nature Communications* **14**(1): 1717.

Bianchi FJ, Booij C, Tscharntke T. 2006. Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control. *Proceedings of the Royal Society B: Biological Sciences* **273**(1595): 1715-1727.

Blaine Marchant D, Soltis DE, Soltis PS. 2016. Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytologist* **212**(3): 708-718.

Blumenthal-Goldschmidt S, Butler GW, Conn EE. 1963. Incorporation of hydrocyanic acid labelled with carbon-14 into asparagine in seedlings. *Nature* **197**(4868): 718-719.

Brewbaker JL. 1955. V-leaf markings of white clover. *Journal of Heredity* **46**(3): 115-123.

Brewbaker JL, Carnahan HL. 1956. Leaf marking alleles in white clover: Uniform nomenclature. *Journal of Heredity* **47**(2): 103-104.

Burke JJ, Chen J, Burow G, Mechref Y, Rosenow D, Payton P, Xin Z, Hayes CM. 2013. Leaf Dhurrin Content is a Quantitative Measure of the Level of Pre- and Postflowering Drought Tolerance in Sorghum. *Crop Science* **0**(0).

Burri R, Antoniazza S, Gaigher A, Ducrest A-L, Simon C, Network TEBO, Fumagalli L, Goudet J, Roulin A. 2016. The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution* **70**(1): 140-153.

Carnahan hL, Hill hd, Hanson aA, Brown kG. 1955. Inheritance and frequencies of leaf markings in white clover. *Journal of Heredity* **46**(3): 109-114.

Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H, Eastwood RJ, Guarino L, Harker RH, Jarvis A, Maxted N, et al. 2016. Global conservation priorities for crop wild relatives. *Nature Plants* **2**(4): 16022.

Chen Y-S, Chesson P, Wu H-W, Pao S-H, Liu J-W, Chien L-F, Yong JWH, Sheue C-R. 2017. Leaf structure affects a plant's appearance: combined multiple-mechanisms intensify remarkable foliar variegation. *Journal of Plant Research* **130**(2): 311-325.

Cook BI, Smerdon JE, Seager R, Coats S. 2014. Global warming and 21st century drying. *Climate Dynamics* **43**(9): 2607-2627.

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, et al. 2012. Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* **338**(6111): 1206-1209.

Corkill L. 1971. Leaf markings in white clover. *Journal of Heredity* **62**(5): 307-310.

Daday H. 1958. Gene frequencies in wild populations of *Trifolium repens* L. III. World distribution. *Heredity* **12**(2): 169-184.

Daday H. 1965. Gene frequencies in wild populations of *Trifolium repens* L. IV. Mechanism of natural selection. *Heredity* **20**(3): 355-365.

Dai A. 2011. Characteristics and trends in various forms of the palmer drought severity index during 1900–2008. *Journal of Geophysical Research: Atmospheres* **116**(D12).

Dai A. 2013. Increasing drought under global warming in observations and models. *Nature Climate Change* **3**(1): 52-58.

Dempewolf H, Baute G, Anderson J, Kilian B, Smith C, Guarino L. 2017. Past and future use of wild relatives in crop breeding. *Crop Science* **57**: 1070-1082.

Dubcovsky J, Dvorak J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* **316**(5833): 1862-1866.

García I, Arenas-Alfonseca L, Moreno I, Gotor C, Romero LC. 2019. HCN regulates cellular processes through posttranslational modification of proteins by *S*-cyanylation. *Plant Physiology* **179**(1): 107-123.

Gefaell J, Vigo R, González-Vázquez AH, Galindo J, Rolán-Alvarez E. 2023. Temporal stability and directional change in a color cline of a marine snail from NW Spain. *Current Zoology*.

Gleadow RM, Moller BL. 2014. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annual Review of Plant Biology* **65**: 155-185.

Godfree RC, Marshall DJ, Young AG, Miller CH, Mathews S. 2017. Empirical evidence of fixed and homeostatic patterns of polyploid advantage in a keystone grass exposed to drought and heat stress. *Royal Society Open Science* **4**(11): 170934.

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019. Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31**(7): 1466-1487.

Grillo MA, Li C, Hammond M, Wang L, Schemske DW. 2013. Genetic architecture of flowering time differentiation between locally adapted populations of *Arabidopsis thaliana*. *New Phytologist* **197**(4): 1321-1331.

Guerrero J, Andrello M, Burgarella C, Manel S. 2018. Soil environment is a key driver of adaptation in *Medicago truncatula:* new insights from landscape genomics. *New Phytologist* **219**(1): 378-390.

Gugger PF, Liang CT, Sork VL, Hodgskiss P, Wright JW. 2018. Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary Applications* **11**(2): 231-242.

Hahn MA, van Kleunen M, Muller-Scharer H. 2012. Increased phenotypic plasticity to climate may have boosted the invasion success of polyploid *Centaurea stoebe*. *PLoS One* **7**(11): e50284.

Hayes C, Emendack Y, Sanchez J, Burke J, Pugh NA, Xin Z, Rooney W. 2023. Evaluation of Diverse Sorghum for Leaf Dhurrin Content and Post-Anthesis (Stay-Green) Drought Tolerance. *Crops* **3**(3): 241-250.

Hof AEvt, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**(7605): 102-105.

Jiang C-x, Wright RJ, El-Zik KM, Paterson AH. 1998. Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proceedings of the National Academy of Sciences of the United States of America* **95**(8): 4419-4424.

Kakes P. 1989. An analysis of the costs and benefits of the cyanogenic system in *Trifolium repens* L. *Theoretical and Applied Genetics* **77**(1): 111-118.

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

Keneni G, Bekele E, Imtiaz M, Dagne K. 2012. Genetic vulnerability of modern crop cultivars: causes, mechanism and remedies. *International Journal of Plant Research* **2**(3): 69-79.

Khoury CK, Carver D, Greene SL, Williams KA, Achicanoy HA, Schori M, Leon B, Wiersema JH, Frances A. 2020. Crop wild relatives of the United States require urgent conservation action. *Proceedings of the National Academy of Sciences of the United States of America* **117**(52): 33351-33357.

Kjærgaard T. 2003. A plant that changed the world: The rise and fall of clover 1000-2000. *Landscape Research* **28**(1): 41-49.

Kooyers NJ, Gage LR, Al-Lozi A, Olsen KM. 2014. Aridity shapes cyanogenesis cline evolution in white clover (*Trifolium repens* L.). *Molecular Ecology* **23**(5): 1053-1070.

Kooyers NJ, Olsen KM. 2012. Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.). *Molecular Ecology* **21**(10): 2455-2468.

Kooyers NJ, Olsen KM. 2013. Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover (*Trifolium repens* L.). *Heredity* **111**(6): 495-504.

Koski MH, Galloway LF. 2018. Geographic variation in pollen color is associated with temperature stress. *New Phytologist* **218**(1): 370-379.

Lascoux M, Glémin S, Savolainen O 2016. Local adaptation in plants. *Encyclopedia of Life Sciences*, 1-7.

Lasky JR, Josephs EB, Morris GP. 2022. Genotype–environment associations to reveal the molecular basis of environmental adaptation. *The Plant Cell* **35**(1): 125-138.

Levin DA. 1975. Minority cytotype exclusion in local plant populations. *Taxon* **24**(1): 35-43.

Liang W-S. 2003. Drought stress increases both cyanogenesis and β-cyanoalanine synthase activity in tobacco. *Plant Science* **165**(5): 1109-1115.

Lyon B, DeWitt DG. 2012. A recent and abrupt decline in the East African long rains. *Geophysical Research Letters* **39**(2).

Machingura M, Salomon E, Jez JM, Ebbs SD. 2016. The beta-cyanoalanine synthase pathway: Beyond cyanide detoxification. *Plant, Cell & Environment* **39**(10): 2329-2341.

McDonald AE. 2008. Alternative oxidase: an inter-kingdom perspective on the function and regulation of this broadly distributed 'cyanide-resistant' terminal oxidase. *Functional Plant Biology* **35**(7): 535-552.

McGrath GS, Sadler R, Fleming K, Tregoning P, Hinz C, Veneklaas EJ. 2012. Tropical cyclones and the ecohydrology of Australia's recent continental-scale drought. *Geophysical Research Letters* **39**(3).

McMahon J, Sayre R, Zidenga T. 2021. Cyanogenesis in cassava and its molecular manipulation for crop improvement. *Journal of Experimental Botany* **73**(7): 1853-1867.

Mitchell C, Brennan RM, Graham J, Karley AJ. 2016. Plant Defense against Herbivorous Pests: Exploiting Resistance and Tolerance Traits for Sustainable Crop Protection. *Frontiers in Plant Science* **7**: 1132.

Molina-Henao YF, Hopkins R. 2019. Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. *American Journal of Botany* **106**(1): 61-70.

Møller BL. 2010. Functional diversifications of cyanogenic glucosides. *Current Opinion in Plant Biology* **13**(3): 337-346.

Mullen LM, Hoekstra HE. 2008. Natural selection along an environmental gradient: A classic cline in mouse pigmentation. *Evolution* **62**(7): 1555-1570.

Myrans H, Vandegeer RK, Henry RJ, Gleadow RM. 2021. Nitrogen availability and allocation in sorghum and its wild relatives: Divergent roles for cyanogenic glucosides. *Journal of Plant Physiology* **258-259**: 153393.

Nielsen LJ, Stuart P, Picmanova M, Rasmussen S, Olsen CE, Harholt J, Moller BL, Bjarnholt N. 2016. Dhurrin metabolism in the developing grain of *Sorghum bicolor* (L.) Moench investigated by metabolite profiling and novel clustering analyses of time-resolved transcriptomic data. *BMC Genomics* **17**(1): 1021.

Nuccio ML, Paul M, Bate NJ, Cohn J, Cutler SR. 2018. Where are the drought tolerant crops? An assessment of more than two decades of plant biotechnology effort in crop improvement. *Plant Science* **273**: 110-119.

Olsen KM, Hsu SC, Small LL. 2008. Evidence on the molecular basis of the *Ac*/*ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* **179**(1): 517-526.

Olsen KM, Small LL. 2018. Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytologist* **219**(2): 757-766.

Olsen KM, Sutherland BL, Small LL. 2007. Molecular evolution of the *Li*/*li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Molecular Ecology* **16**(19): 4180-4193.

Pos O, Radvanszky J, Buglyo G, Pos Z, Rusnakova D, Nagy B, Szemes T. 2021. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal* **44**(5): 548-559.

Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, He Q, Ou S, Zhang H, Li X. 2021. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**(13): 3542-3558. e3516.

Rosati VC, Blomstedt CK, Moller BL, Garnett T, Gleadow R. 2019. The interplay between water limitation, dhurrin, and nitrate in the low-cyanogenic sorghum mutant adult cyanide deficient class 1. *Frontiers in Plant Science* **10**: 1458.

Santangelo JS, Battlay P, Hendrickson BT, Kuo WH, Olsen KM, Kooyers NJ, Johnson MTJ, Hodgins KA, Ness RW. 2023. Haplotype-resolved, chromosome-level assembly of white clover (*Trifolium repens* L., Fabaceae). *Genome Biology and Evolution* **15**(8): 2023.2006.2006.543960.

Santantonio N, Jannink JL, Sorrells M. 2019. Homeologous epistasis in wheat: The search for an immortal hybrid. *Genetics* **211**(3): 1105-1122.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**(11): 807-820.

Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, De S, Kishony R, Michor F, Dowell R, et al. 2015. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**(7543): 349-352.

Shehab AA, Yao L, Wei L, Wang D, Li Y, Zhang X, Guo Y. 2020. The increased hydrocyanic acid in drought-stressed sorghums could be alleviated by plant growth regulators. *Crop and Pasture Science* **71**(5).

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**(1): 336-348.

Stalder L, Oggenfuss U, Mohd-Assaad N, Croll D. 2023. The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Molecular Ecology* **32**(10): 2443-2460.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences of the United States of America* **101**(13): 4712-4717.

Udall JA, Wendel JF. 2006. Polyploidy and crop improvement. *Crop Science* **46**(Supplement_1): S-3-S-14.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18**(7): 411-424.

Wang H, Wu Y, He Y, Li G, Ma L, Li S, Huang J, Yang G. 2023. High-quality chromosome-level de novo assembly of the *Trifolium repens*. *BMC Genomics* **24**(1): 326.

Wang IJ, Bradburd GS. 2014. Isolation by environment. *Molecular Ecology* **23**(23): 5649-5662.

Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J. 2015. Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nature genetics* **47**(8): 944-948.

Wang Y, Zhang L, Zhou Y, Ma W, Li M, Guo P, Feng L, Fu C. 2023. Using landscape genomics to assess local adaptation and genomic vulnerability of a perennial herb *Tetrastigma hemsleyanum* (Vitaceae) in subtropical China. *Frontiers in Genetics* **14**: 1150704.

Wei N, Cronn R, Liston A, Ashman TL. 2019. Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. *New Phytologist* **221**(4): 2286-2297.

Wittkopp PJ, Smith-Winberry G, Arnold LL, Thompson EM, Cooley AM, Yuan DC, Song Q, McAllister BF. 2011. Local adaptation for body color in *Drosophila americana*. *Heredity* **106**(4): 592-602.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**(33): 13875-13879.

Wright D, Boije H, Meadows JR, Bed'hom B, Gourichon D, Vieaud A, Tixier-Boichard M, Rubin CJ, Imsland F, Hallbook F, et al. 2009. Copy number variation in intron 1 of *SOX5* causes the Pea-comb phenotype in chickens. *PLOS Genetics* **5**(6): e1000512.

Wright SJ, Cui Zhou D, Kuhle A, Olsen KM. 2018. Continent-wide climatic variation drives local adaptation in North American white clover. *Journal of Heredity* **109**(1): 78-89.

Wright SJ, Goad DM, Gross BL, Muñoz PR, Olsen KM. 2022. Genetic trade-offs underlie divergent life history strategies for local adaptation in white clover. *Molecular Ecology* **31**(14): 3742-3760.

Xu F, Zhang D-W, Zhu F, Tang H, Lv X, Cheng J, Xie H-F, Lin H-H. 2012. A novel role for cyanide in the control of cucumber (*Cucumis sativus* L.) seedlings response to environmental stress. *Plant, Cell & Environment* **35**(11): 1983-1997.

Yang J, Moeinzadeh MH, Kuhl H, Helmuth J, Xiao P, Haas S, Liu G, Zheng J, Sun Z, Fan W, et al. 2017. Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nature Plants* **3**(9): 696-703.

Yu L-L, Liu Y, Liu C-J, Zhu F, He Z-Q, Xu F. 2020. Overexpressed β-cyanoalanine synthase functions with alternative oxidase to improve tobacco resistance to salt stress by alleviating oxidative damage. *FEBS Letters* **n/a**(n/a).

Zhang H, Mittal N, Leamy LJ, Barazani O, Song B-H. 2017. Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications* **10**(1): 5-24.

Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, et al. 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics* **50**(11): 1565-1573.

Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, et al. 2019. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics* **51**(5): 865-876.

Zmienko A, Marszalek-Zenczak M, Wojciechowski P, Samelak-Czajka A, Luczak M, Kozlowski P, Karlowski WM, Figlerowicz M. 2020. AthCNV: A map of DNA copy number variations in the Arabidopsis genome. *Plant Cell* **32**(6): 1797-1819.

Chapter 1 Variable expression of cyanide detoxification and tolerance genes in cyanogenic and acyanogenic white clover (*Trifolium repens* L.)

1.1 Authorship & Affiliations

Wen-Hsi Kuo[1], Linda L. Small[1], and Kenneth M. Olsen[1]

[1]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130

This chapter has been peer reviewed and published in:

American Journal of Botany 2023 – 110(10):e16233. https://doi.org/10.1002/ajb2.16233

1.2 Summary

- **Premise of the study:** β-cyanoalanine synthase (β-CAS) and alternative oxidase (AOX) play important roles in the ability of plants to detoxify and tolerate hydrogen cyanide (HCN) stress. These functions are critical for all plants, as HCN is produced at low levels during basic metabolic processes, but are likely to be especially important in cyanogenic species, which release high levels of HCN following tissue damage. However, their expression has not been examined in cyanogenic species, nor has it been compared between cyanogenic and acyanogenic genotypes within a species.
- **Methods:** We used a natural polymorphism for cyanogenesis in white clover to examine β-*CAS* and *Aox* gene expression in relation to cyanogenesis-associated HCN exposure. We identified all β-*CAS* and *Aox* gene copies present in the genome, including members of the *Aox1*, *Aox2a* and *Aox2d* subfamilies previously reported in legumes. Expression levels were compared between cyanogenic and acyanogenic genotypes, and under conditions of leaf tissue damage compared to undamaged tissue.
- **Key results:** Results indicate that β-*CAS* and *Aox2a* expression are differentially elevated in cyanogenic genotypes, and that tissue damage is not required to induce this increased expression. *Aox2d*, in contrast, appears to be upregulated as a generalized wounding response.
- **Conclusions:** These findings suggest a heightened constitutive role for both HCN detoxification (via elevated β-*CAS* expression) and HCN-toxicity mitigation (via elevated *Aox2a* expression) in plants that are capable of cyanogenesis. As such, freezing-induced cyanide autotoxicity is unlikely to be the primary selective factor in the evolution of climate-associated cyanogenesis clines.

Keywords: alternative oxidase (AOX); autotoxicity; balanced polymorphism; β-cyanoalanine synthase (β-*CAS*); cline; cyanogenesis; nitrogen assimilation; white clover (*Trifolium repens* L.)

1.3 Introduction

Cyanide (HCN or CN^-) is produced by all plants during ethylene biosynthesis and other developmental processes (Peiser *et al.*, 1984). Cyanide is also highly toxic, as it binds irreversibly to the terminal cytochrome c oxidase in the mitochondrial electron transport chain (ETC), shutting down cellular respiration (Antonini *et al.*, 1971; Jones *et al.*, 1984) and resulting in the production of highly damaging reactive oxygen species (ROS) and reactive nitrogen species (RNS) (Amirsadeghi *et al.*, 2006; Giraud *et al.*, 2008). Plants must therefore possess mechanisms to detoxify cyanide and prevent the autotoxicity that would otherwise occur in its presence. Research over the last half-century has established that the primary mechanism for cyanide detoxification in plants occurs through the β-cyanoalanine synthase (β-CAS) pathway (first confirmed by Tschiersch, 1964; recently reviewed by Gleadow & Moller, 2014; Machingura *et al.*, 2016). The first step is catalyzed by β-CAS, which substitutes cyanide for the sulfur in the amino acid cysteine to form β-cyanoalanine plus H_2S; subsequent steps convert the β-cyanoalanine into asparagine, which can then be metabolized to form aspartate and ammonia. The β-CAS pathway thereby serves to couple the process of cyanide detoxification with nitrogen assimilation for primary metabolism (Gleadow & Moller, 2014; Machingura *et al.*, 2016).

In addition to cyanide detoxification via β-CAS, members of the alternative oxidase (AOX) family may also play important roles related to cyanide exposure in plants. AOX enzymes do not function directly in cyanide detoxification, but provide an alternative pathway to cytochrome c oxidase-mediated electron transport; while generating less ATP, this pathway

allows for respiration to continue while cyanide is present (reviewed by Vanlerberghe, 2013). More generally, upregulation of AOX enzymes appears to provide beneficial functions under conditions of environmental and chemical stress, including cyanide exposure (reviewed by Selinski *et al.*, 2018; Sweetman *et al.*, 2019). For example, AOX has been shown to work synergistically with β-CAS to withstand oxidative stress in tobacco (Yu *et al.*, 2020), and transgenic up-regulation of *Aox* in cassava can prevent the ROS-mediated post-harvest tissue damage that otherwise occurs in high-cyanide tissues (Zidenga *et al.*, 2012). AOX enzymes could thus potentially operate in multiple ways to mitigate cyanide toxicity within plants. The β-CAS and AOX pathways in plants are summarized in **Fig. 1.1**.

At the genetic level, β-*CAS* commonly occurs as a single-copy gene in plants, with sequences forming a clade within the larger β-substituted alanine synthase (BSAS) family (Yi & Jez, 2012; Yi *et al.*, 2012). In contrast*, Aox* occurs as a multigene family in all plants examined to date, with dicots characterized by the *Aox1* and *Aox2* subfamilies (Costa *et al.*, 2014). Within the legumes (Fabaceae), *Aox2* has further diversified into the *Aox2a* and *Aox2d* gene classes (Sweetman *et al.*, 2019). The different AOX isoforms appear to have diverged in function and in tissue-specific patterns of activity, as well as in the extent to which they are expressed constitutively or under specific stress conditions (McDonald, 2008; Sweetman *et al.*, 2019). *Aox1* expression has been found to be highly responsive to abiotic stress in Arabidopsis (Giraud *et al.*, 2008; Smith *et al.*, 2009); in legumes, *Aox2* genes have been found to show development- and tissue-dependent expression, even in the absence of environmental stress (Sweetman *et al.*, 2019).

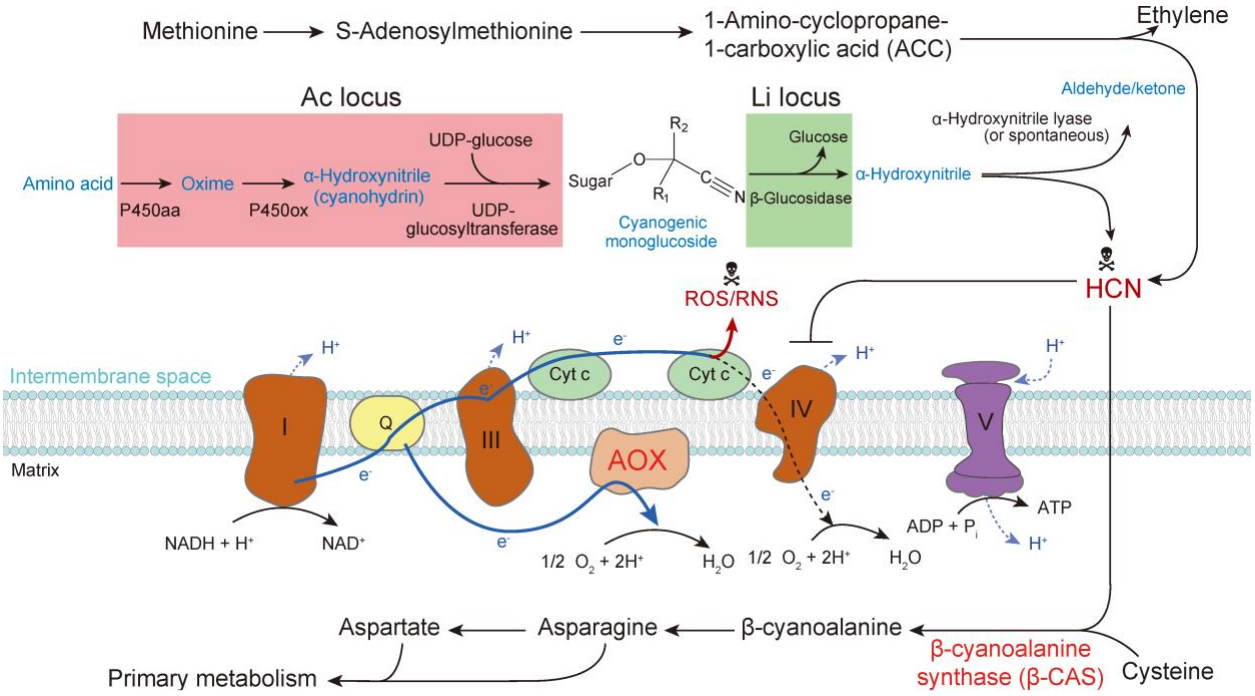


Figure 1.1 Cyanide detoxification and tolerance pathways in plant mitochondria.

The mitochondrial inner membrane with electron transport chain (ETC) and protein complexes are shown. Q, ubiquinone pool. AOX, alternative oxidase.

Despite the toxic burden imposed by cyanide production, more than 3,000 plant species actively synthesize cyanogenic glycosides, which, when hydrolyzed, release high levels of cyanide within plant tissue (Gleadow & Moller, 2014). This cyanogenic reaction typically occurs following tissue damage that brings the cyanogenic glycosides in contact with their hydrolyzing cyanogenic β-glucosidases. As this "cyanide bomb" can be triggered by herbivores, cyanogenesis is generally believed to have evolved as an anti-herbivore chemical defense (Møller, 2010; Mithöfer & Boland, 2012). The fact that cyanogenesis has evolved multiple times independently across the plant kingdom further supports its apparent adaptive value (Gleadow & Moller, 2014). For cyanogenic plants, autotoxicity from cyanide exposure could occur either incidentally during the synthesis of cyanogenic glycosides or with triggering of the cyanogenic response. Notably, cyanogenic glucosides may constitute as much as 30% of the leaf dry matter in some cyanogenic plants (Halkier & Moller, 1989); the need to rapidly and efficiently mitigate cyanide toxicity is therefore likely to be especially acute in cyanogenic species.

Consistent with this prediction, research to date suggests that the β-CAS pathway plays a critical role in cyanide detoxification in cyanogenic plants. Studies in several cyanogenic species, including cassava, rubber tree, almond, Japanese plum, and barley, have indicated that β-CAS enzyme activity tends to be higher in tissues with the greatest concentrations of cyanogenic glycosides, with the highest rate of glycoside synthesis, and/or high levels of cyanogenic β-glucosidase activity (Selmar *et al.*, 1988; Mizutani *et al.*, 1992; Nambisan & Sundaresan, 1994; Elias *et al.*, 1997; Forslund & Jonsson, 1997; Sánchez-Pérez *et al.*, 2008;

Kongsawadworakul *et al.*, 2009). However, much remains unresolved about the mechanisms of cyanide detoxification and tolerance in cyanogenic plants. To the extent that β-CAS has been compared between cyanogenic and non-cyanogenic plants, this has been restricted to comparisons of enzymatic activity between cyanogenic and acyanogenic species (Miller & Conn, 1980); there have been no direct comparisons of cyanogenic and acyanogenic genotypes within species or comparisons of cyanogenic species within a single genus. In addition, relatively little is known about β-*CAS* at the genetic level, as studies to date have focused on enzymatic activity. While the gene can be identified based on specific amino acid motifs that distinguish it from the closely related O-acetylserine sulfhydrylase (*OASS*) genes within the BSAS family (with three amino acid differences collectively conferring OASS vs. CAS substrate specificity; see below) (Yi *et al.*, 2012), no studies to date have examined β-*CAS* gene expression as related to cyanide detoxification. Another knowledge gap relates to *Aox* in cyanogenic species, and whether the different members of this multigene family might serve particular functions related to cyanogenesis-mediated cyanide exposure.

Species that are naturally polymorphic for cyanogenesis can provide highly useful study systems for examining questions on cyanogenesis-related cyanide metabolism. White clover (*Trifolium repens*, Fabaceae) is one such species that has long served as a model for studying cyanogenesis (e.g. Mirande, 1912; Hughes, 1991; Santangelo *et al.*, 2022). Wild populations of this species contain both cyanogenic and acyanogenic plants, and studies spanning more than a century have examined this polymorphism — determining its biochemical basis and inheritance (Coop, 1940; Melville & Doak, 1940; Corkill, 1942), establishing the molecular genetic basis and

molecular evolutionary origins (Olsen *et al.*, 2007; Olsen & Ungerer, 2008; Olsen *et al.*, 2013; Olsen & Small, 2018; Olsen *et al.*, 2021; Wright *et al.*, 2022), and assessing the climatic and other ecological factors that maintain the polymorphism in nature (Daday, 1954b; Daday, 1954a; Kooyers & Olsen, 2012; Santangelo *et al.*, 2022; Wright *et al.*, 2022).

The genetic basis of the white clover cyanogenesis polymorphism makes it particularly well suited for comparing mechanisms of cyanide metabolism in cyanogenic and acyanogenic genotypes. The polymorphism is controlled by two independently-segregating simple Mendelian polymorphisms for the presence/absence of the two required cyanogenic precursors: cyanogenic glucosides (linamarin and lotaustralin), which are stored in the vacuoles of photosynthetic tissue, and their hydrolyzing β-glucosidase (linamarase), which is present in the apoplast (reviewed by Hughes, 1991). For each biochemical polymorphism, dominant (functional) alleles confer the presence of the precursor: *Ac/ac* determines the presence/absence of the cyanogenic glucosides, and *Li/li* determines presence/absence of linamarase. At the molecular level, the *Ac* locus corresponds to the three-gene metabolic cluster comprising the cyanogenic glucoside biosynthetic pathway (Olsen & Small, 2018), and *Li* encodes the linamarase enzyme (Olsen *et al.*, 2007); for both loci, recessive alleles correspond to genomic deletions of the component genes (Olsen *et al.*, 2013; Olsen & Small, 2018; Olsen *et al.*, 2021).

Because the *Ac/ac* and *Li/li* polymorphisms are independently segregating, four different cyanogenesis phenotypes (or 'cyanotypes') occur in nature: the cyanogenic form

('AcLi', with both cyanogenic precursors present), and three acyanogenic forms ('Acli', lacking linamarase; 'acLi', lacking cyanogenic glucosides; and 'acli', lacking both components). For cyanide mitigation, the AcLi plants alone would be at risk of cyanogenesis-triggered cyanide exposure, whereas both the AcLi and Acli plants would be subject to any cyanide exposure related to cyanogenic glucoside biosynthesis. Comparisons of β-*CAS* and *Aox* gene expression between the AcLi, Acli and acli cyanotypes can thus provide a way to test whether these genes function either in the mitigation of cyanide exposure associated with cyanogenesis, with cyanide released incidentally during the synthesis of cyanogenic precursors, with both of these processes, or with neither.

In this study, we used the white clover cyanogenesis polymorphism to test whether β-*CAS* and *Aox* genes show evidence of elevated expression associated with the cyanogenic response and/or synthesis of cyanogenic precursors. Taking advantage of this natural intraspecific variation, along with genomic resources available for this system (Griffiths *et al.*, 2019; Olsen *et al.*, 2021), we identified all β-*CAS* and *Aox* genes present in the white clover genome and then tested for variation in gene expression as related to cyanotype variation and tissue damage-induced cyanogenesis. We specifically addressed the following questions: 1) What β-*CAS* and *Aox* genes are present in the white clover genome? 2) Is there evidence of elevated gene expression following tissue damage? If so, does it occur in a pattern consistent with cyanogenesis-mediated cyanide exposure (detected in the AcLi genotypes only), or a more general tissue wounding stress response (detected in all cyanotypes)? 3) Which genes, if any, show cyanotype-specific expression variation in undamaged tissue (potentially suggesting

cyanide mitigation not directly tied to wounding-induced cyanogenesis)? Our results indicate that β-*CAS* and *Aox* expression are elevated specifically in cyanogenic (AcLi) plants, and that this elevated expression occurs regardless of whether the cyanogenic response is triggered by tissue damage. This finding suggests that the cyanide detoxification and mitigation response functions differentially in plants that are capable of cyanogenesis, and that tissue damage-induced HCN release is not required to induce this function.

1.4 Materials and Methods

Sampling of plant materials and cyanogenesis characterization —

All plant samples were collected from wild *Trifolium repens* populations in North America and propagated in the greenhouse of Washington University in St. Louis. Localities and voucher information are shown in **Supplementary Table S1.1 & Fig. S1.1**. Cyanotypes were determined using fresh leaf tissue from greenhouse-grown plants and a modified Feigl-Anger HCN assay (Feigl & Anger, 1966; details in Olsen *et al.*, 2007; Olsen *et al.*, 2008). In short, fresh leaf tissue was collected in the greenhouse and immediately frozen in 48-well polystyrene cell culture plates in a -80°C freezer to induce cell rupture. The tissue was thawed at room temperature and pulverized with a pipet tip to fully mix the cyanogenic glucosides (linamarin and lotaustralin) and hydrolyzing β-glucosidase (linamarase) if present. The plate was covered with Feigl-Anger test paper and the cell plate cover, then clamped along the edges with large binder clips; a blue color change in the test paper covering a leaf tissue sample indicates HCN liberation and a cyanogenic (AcLi) genotype. For acyanogenic tissue samples, additional leaf

tissue samples were tested with either exogenously added linamarin or linamarase solutions to distinguish between Acli, acLi and acli genotypes. The corresponding genotypes for presence/absence of cyanogenic glucosides (*Ac/ac*) and linamarase (*Li/li*) were confirmed by PCR genotyping for the underlying gene presence/absence variation (Olsen *et al.*, 2007; Olsen *et al.*, 2008). All plants used in the study were consistent between their PCR genotype and cyanogenesis phenotype.

Gene isolations, sequencing and phylogenetic analyses —

Genes encoding β-*CAS* and *AOX* were identified in the white clover reference genome and related eudicot species by BLAST-based searches in OrthoFinder (Emms & Kelly, 2019). The white clover primary protein sequences were translated from the reference genome with the annotation file (Griffiths *et al.*, 2019). For related eudicot species, the primary protein sequences were download from Phytozome; these included *Arabidopsis thaliana* TAIR10, *Cicer arietinum* (chickpea) v1.0, *Glycine max* (soybean) Wm82.a2.v1, *Lotus japonicus* Lj1.0v1, *Medicago truncatula* Mt4.0v1, *Phaseolus vulgaris* (common bean) v2.1, *Trifolium pratense* (red clover) v2, and *Vigna unguiculata* (cowpea) v1.2. The output "orthogroups" were identified by manual BLASTs using the published sequences. To ensure correct identification of β-*CAS* orthologs, both the β-*CAS* and the closely related *OASS* genes were identified using *Glycine max CAS* (NM_001352889.1) and *OASS1* (AF452451.1) sequences as queries. Site-specific mutagenesis experiments in soybean have established that three amino acid residues within two active site domains are collectively responsible for CAS *vs*. OASS substrate activity (Yi *et al.*,

2012); these amino acid differences are thus diagnostic of the two gene subfamilies. For *AOX*, we used *Medicago sativa Aox1* (KC685557.1), *Aox2a* (KC685558.1), *Aox2d1* (KC685559.1) and *Aox2d2* (KC685560.1) genes (isolated by Cavalcanti *et al.*, 2013) as queries in BLAST searches.

Because white clover is a recently evolved allotetraploid species (2n = 4x = 32), descended from two extant diploid congeners, *Trifolium occidentale* (2n = 16) and *T. pallescens* (2n = 16) (Williams *et al.*, 2012), much of the genome is expected to show two-fold gene duplications compared to diploid relatives. We categorized homeologous copies of the β-*CAS* and *Aox* genes as belonging to either the *T. occidentale* (to) or *T. pallescens* (tp) subgenome based on chromosomal location. Any putative homeologous copies were further validated beyond the published reference genome by mapping to whole-genome PacBio HiFi long reads (NCBI SRA PRJNA953427). Similarly, in cases where a homeolog was not detected, its absence from the subgenome was also confirmed by the read mapping technique.

For each targeted gene or gene subfamily, a maximum likelihood (ML) tree was constructed based on protein sequences using RAxML with automatic searching for the best protein substitution mode (Stamatakis, 2014). Twenty ML trees were reconstructed independently, and topological consistency among trees was assessed by visual inspection. The bootstrap value of each branch was assessed as the percentage of consensus to the best tree with 1000 replicates. For the β-*CAS* tree, gene sequences from spinach (LOC110789773) and apple (DQ471308.1, DQ471309.1) were added to stabilize the topology of the focal species in the ML output.

Wounding treatment, RNA extraction and reverse transcription —

To examine whether β-*CAS* and *Aox* genes are inducible by cyanide liberation within damaged leaf tissue, we mimicked herbivore-induced cyanogenesis by piercing leaflets of cyanogenic genotypes with a syringe needle; 20-25 piercings were applied to each of the three leaflets per leaf. After the wounding treatment, two leaflets per genotype were sampled and frozen in liquid nitrogen at four designated time points: 5, 30, 60 and 120 minutes after wounding. All the wounding treatments were conducted and sampled between 11 am – 2 pm for a homogeneous circadian background. Frozen leaf samples were homogenized in 2.0 mL round bottom microcentrifuge tubes with steel beads using a homogenizer. Total RNA was extracted using PureLink™ Plant RNA Reagent. Extraction quantity and quality were assessed using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, and Qubit™ RNA HS Assay Kits. First-strand cDNA was synthesized by SuperScript™ IV First-Strand Synthesis System with a poly-A primer. The cDNA was diluted to 20 ng/μL as working solution based on the initial RNA concentration.

Real-time qPCR —

To design qPCR primers for expression analyses, targeted genes were PCR amplified in 6-10 white clover accessions and Sanger sequenced to identify conserved sequences (available in **Supplementary Table S1.2**). Cloning of PCR products was conducted by pGEM®-T Easy Vector System if necessary. Primer design followed the BIO-RAD primer design guide (http://www.bio-rad.com/en-us/applications-technologies/qpcr-assay-design-optimization); primer sequences

are provided in **Supplementary Table S1.2**. Six serial dilutions of a standard cDNA (50, 10, 2, 0.4, 0.08, 0.016 ng/reaction) were used to test the primer efficiency (>95%), specificity (single peak in melting curve plot) and the lowest amount of starting materials (linear relationship between log[cDNA amount] and Ct value). Because of the very high coding sequence similarity of between homeologous copies of a given gene (>97% sequence identity), qPCR did not distinguish expression differences between individual gene copies within a given gene class (β-*CAS*, *Aox1*, *Aox2a*, or *Aox2d*; see Results).

All qPCR protocols followed MIQE guidelines (Bustin *et al.*, 2009), including for numbers of biological and technical replicates and reaction optimization protocols. Gene expression level was quantified using the CFX96 Real-Time PCR Detection System (BIO-RAD, Hercules, California, USA). For all genes, reactions included 1× iTaq™ Universal SYBR® Green Supermix, 0.5 µM (0.3 µM for β-*CAS*) forward primer, 0.5 µM (0.3 µM for β-*CAS*) reverse primer and 4-10 ng cDNA in a 20 µL reaction volume. Ef1α was used as the reference gene (Olsen & Small, 2018). PCR conditions followed the standard SYBR® Green cycle in CFX96 Real-Time PCR Detection System with slight modifications to maximize reaction efficiency: 95°C for 30 s, then [95°C for 5 s, 60°C (63°C for β-*CAS*) for 30 s] × 40 cycles. Three technical replicates were applied to each cDNA of the sample. The data were analyzed by the $\Delta\Delta C_T$ method with efficiency calibration in CFX Maestro™ Software (BIO-RAD, Hercules, California, USA) according to the manufacturer's instructions. For the gene expression assessments after the tissue wounding treatment, expression levels were normalized to *Ef1α* and then scaled to the control treatment (i.e., the expression level before wounding treatment) for each individual accession, to control

for variation in baseline expression among genotypes. For the other analyses, the relative expression levels were normalized to the *Ef1α* gene and then globally scaled to the minimum expression of *Aox1* genes to enable the comparison of expression levels among genes. Except for *Aox2d*, the samples were pooled among wound treatments since there were no detectable differences between them. For *Aox2d*, only the control samples were included. The ANOVA of one or two factors was conducted in R v.4.2.0 by using the aov function.

1.5 Results

Gene identifications and phylogenetic inferences —

Based on unique amino acid motifs that distinguish β-*CAS* from the closely-related *OASS* genes within the *BSAS* gene family (Yi *et al.*, 2012), β-*CAS* gene copies could be identified in the white clover reference genome and their sequences aligned to orthologs of related legumes and other eudicot species (**Supplementary Fig. S1.2**). Consistent with white clover's evolutionary origin as a recently-evolved allotetraploid species, its genome carries two highly similar β-*CAS* copies corresponding to its two diploid progenitor subgenomes, *Trifolium occidentale* (*TrCAS_to*) and *T. pallescens* (*TrCAS_tp*); these are located at 18-20 Mbp on homeologous chromosomes 7 and 15, respectively (**Supplementary Fig. S1.4**). PCR amplification and Sanger sequencing confirmed β-*CAS* gene identity in amplified products. As previously reported for β-*CAS* orthologs in other species, an approximately 50-amino mitochondrial targeting peptide was present in the newly identified isolates. The white clover

β-*CAS* and *OASS* gene sequences are available in GenBank (accessions OP437418-OP437421) (**Supplementary Table S1.3**).

Using *OASS* genes as the outgroup, a maximum likelihood (ML) β-*CAS* gene phylogeny was reconstructed based on the gene sequence alignment. Although bootstrap support is low, the most frequently obtained gene tree is congruent with phylogenetic relationships of the represented taxa (**Fig. 1.2**). The two white clover homeologs form a clade and are sister to the single gene copy present in the closely related legume *Medicago truncatula* (*MtCAS*), and then chickpea (*CaCAS*). Other more distantly related papilionoid legumes form a sister clade to these gene sequences; these include *Lotus japonicus* (*LjCAS*), common bean (*PvCAS*), cowpea (*VuCAS*), and soybean, which as an ancient allotetraploid also carries two homeologous copies (*GmCAS_chr09*, *GmCAS_chr18*). Similarly, outside the legumes, the gene tree corresponds to phylogenetic relationships for apple (*MdCAS1*, *MdCAS2*), *Arabidopsis* (*AtCAS*) and spinach (*SoCAS*).

Compared to β-*CAS*, *Aox* gene copy variation in white clover is more complex, with the *T. occidentale* and *T. pallescens* subgenomes contributing asymmetrically to the numbers of homeologs. We isolated a total of seven *Aox* genes, all of which were confirmed to share four diagnostic conserved motifs with published AOX proteins (Berthold *et al.*, 2000) (**Supplementary Fig. S1.3**). Their phylogenetic relationships and locations in the white clover genome indicate that there are two homeologous *Aox1* genes (*TrAox1_to*, *TrAox1_tp*), four *Aox2a* genes (*TrAox2a_to_1*, *TrAox2a_to_2*, *TrAox2a_to_3*, *TrAox2a_tp*), and a single *Aox2d*

gene (*TrAox2d_tp*) in white clover. *TrAox1_to* and *TrAox1_tp* are at 13-14 Mbp of the homeologous chromosomes 5 and 13, respectively. For the *Aox2a* genes, *TrAox2a_to_1*, *TrAox2a_to_2*, and *TrAox2a_to_3* represent a proximal gene triplication at 47-48 Mbp on chromosome 4, while the one copy of *TrAox2a_tp* is at 44-45 Mbp on the homeologous chromosome 12. The single *Aox2d* copy (*TrAox2d_tp*) is located at 47 Mbp of chromosome 13 within the *T. pallescens* subgenome (**Supplementary Fig. S1.4**). Similar to β-*CAS*, the phylogenetic relationships among orthologous *Aox* genes reflect the overall phylogeny of the legume family. For both *Aox1* and *Aox2*, the white clover gene copies are most closely related to the genes present in its congener red clover (*TpAox*), followed by other legume species (**Fig. 1.3**). Newly identified white clover *Aox* gene sequences are available in GenBank (accessions OP437422-OP437427, OP456086) (**Supplementary Table S1.3**).

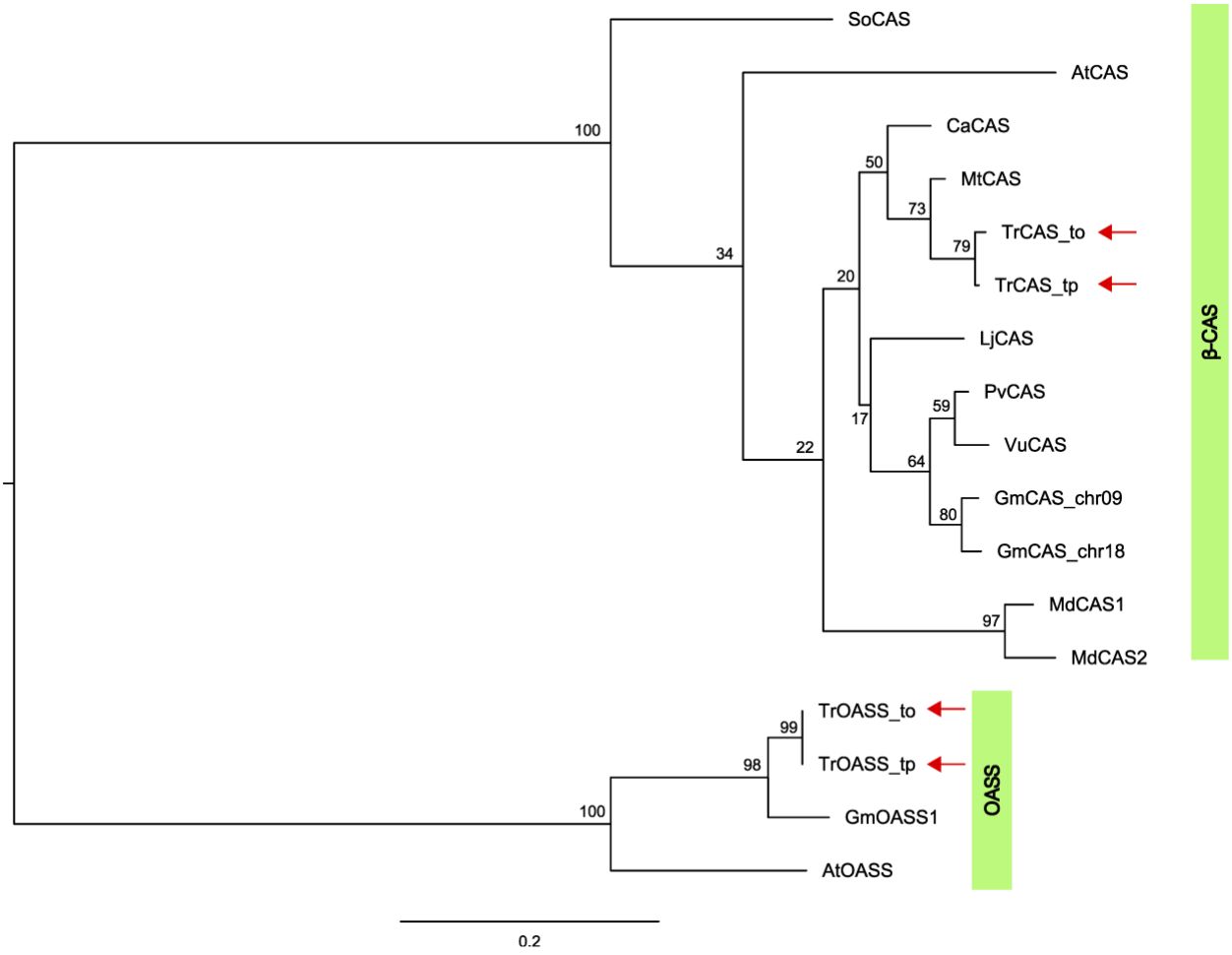


Figure 1.2 Maximum likelihood phylogeny of β-substituted alanine synthase (BSAS) family gene sequences.

White clover genes isolated in this study are highlighted by the red arrows. Numbers on branches indicate percent bootstrap support (1000 replicates). The first two letters of gene labels indicate species as follows: *Arabidopsis thaliana* (At), *Cicer arietinum* (Ca), *Glycine max* (Gm), *Lotus japonica* (Lj), *Malus domestica* (Md), *Medicago truncatula* (Mt), *Phaseolus vulgaris* (Pv), *Spinacea oleracea* (So), *Trifolium repens* (Tr), *Vigna unguiculata* (Vu).

β-CAS and Aox expression with tissue wounding —

Because β-*CAS* and *Aox* genes are known to have functions in plant responses to environmental stress (McDonald, 2008; Sweetman *et al.*, 2019; Yu *et al.*, 2020), we hypothesized that the expression level of some gene classes could be upregulated in white clover in response to leaf tissue damage, regardless of whether a plant genotype was cyanogenic or acyanogenic. To test this hypothesis, we measured changes in gene expression level after leaflet wounding treatments at 5, 30, 60 and 120 minutes. Due to the high coding sequence similarity between the white clover homeologous gene copies (>97% identity), qPCR primers were designed to amplify all homeologous copies of a given target gene class (i.e., *TrCAS*, *TrAox1*, *TrAox2a*, or *TrAox2d*); the qPCR results thereby provided an assessment of total gene expression by a given gene class, without information on gene copy-specific expression. As shown in **Fig. 1.4 d, h**, only *Aox2d* showed significantly higher expression levels after the wounding treatment; this response occurred in both cyanogenic (AcLi, $F_{3,20} = 3.60$, $P = 0.032$) and acyanogenic (acli, $F_{3,24} = 3.55$, $P = 0.029$) cyanotypes, with the greatest increase in expression occurring between 60-120 minutes. The final mean expression level was 4.5 times higher than the control level, and there was no significant difference in degree of response between cyanotypes. The occurrence of elevated *Aox2d* expression following tissue damage in all tested genotypes, regardless of cyanotype, suggests that this is a general wounding response and is unrelated to cyanide exposure (no interaction in the ANOVA of the cyanotypes and the treatments. $F_{3,44} = 0.14$, $P = 0.94$).

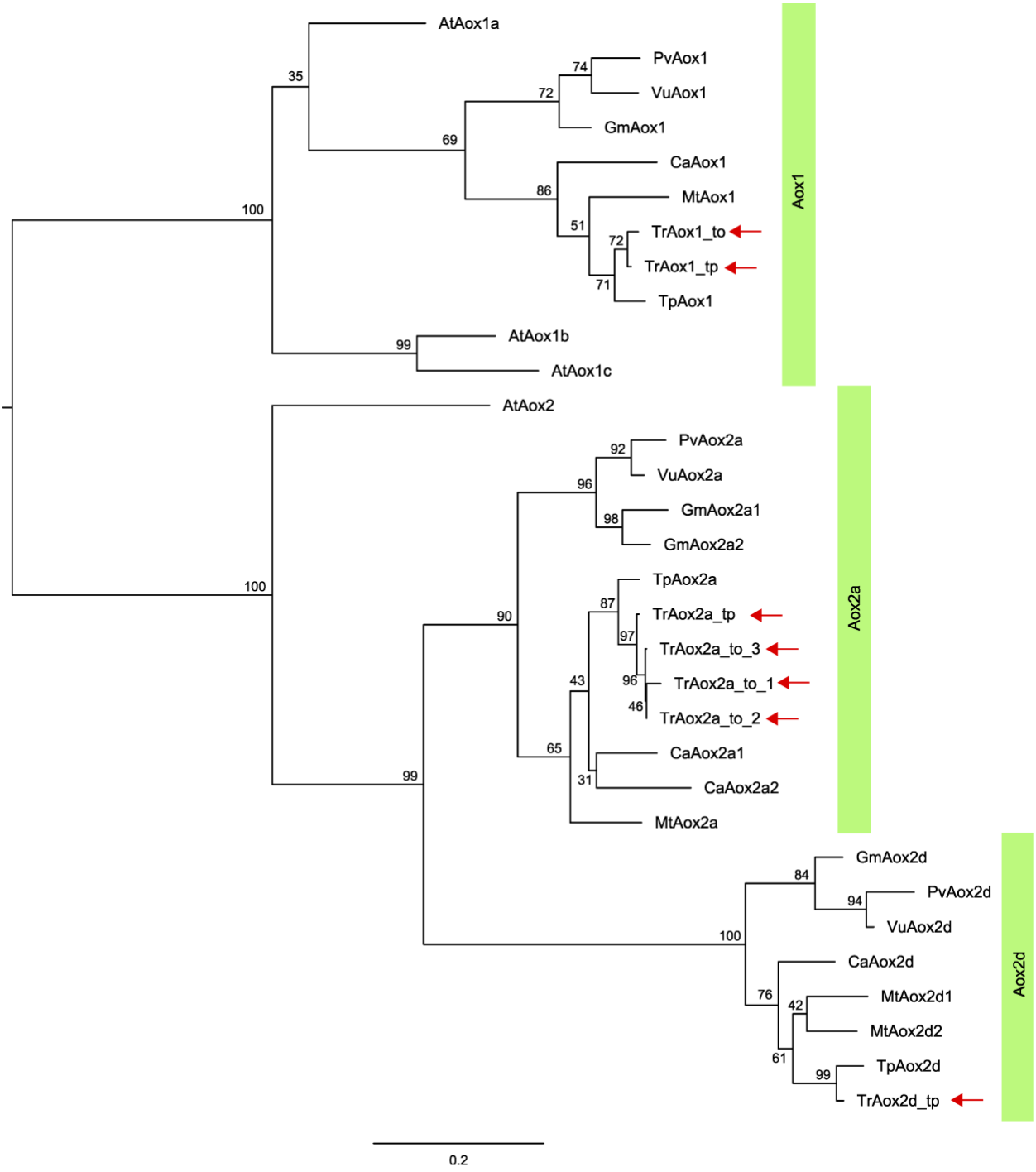


Figure 1.3 Maximum likelihood phylogeny of *Aox* gene family.

Numbers on branches indicate bootstrap support (1000 replicates). The first two letters of gene labels indicate species as follows: *Arabidopsis thaliana* (At), *Cicer arietinum* (Ca), *Glycine max* (Gm), *Medicago truncatula* (Mt), *Phaseolus vulgaris* (Pv), *Trifolium pratense* (Tp), *Trifolium repens* (Tr), *Vigna unguiculata* (Vu).

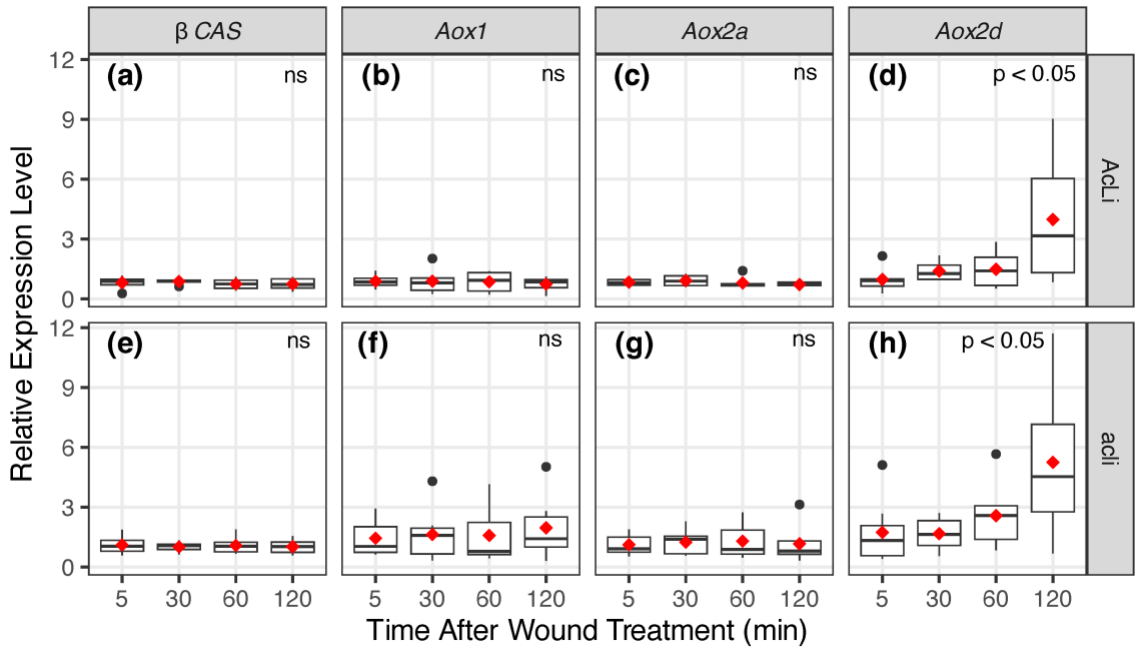


Figure 1.4 Relative expression level of β-*CAS* (a,e) and *Aox* genes (b-d,f-h) after leaf tissue wounding treatment.

Sample sizes per time point are as follows: N_{AcLi} = 6-9 biological replications ($n_{control}$ = 9, n_5 = 6, n_{30} = 6, n_{60} = 6, n_{120} = 6); N_{acli} = 7-10 biological replications ($n_{control}$ = 10, n_5 = 7, n_{30} = 7, n_{60} = 7, n_{120} = 7) (see **Supplementary Table S1.1**). AcLi, plants with cyanogenic glucosides and glucosidase (cyanogenic); acli, the plants with neither cyanogenic component (acyanogenic). Red diamond, mean value of each treatment. The *p* values were calculated by one-way ANOVA for the difference between the treatments in each panel.

Cyanogenic genotype shows elevated expression of β-*CAS* and *Aox2a* without wounding —

Whereas the wounding treatment did not induce cyanotype-specific expression changes, significant differences in expression were detected for β-*CAS* and *Aox2a* between cyanogenic and acyanogenic cyanotypes when tissue was intact; specifically, relative expression was significantly higher in the cyanogenic (AcLi) cyanotype than in Acli and acli cyanotypes (**Fig. 1.5 a, c**. For β-*CAS*, $F_{2,26} = 9.46$, P = 0.00082. For *Aox2a*, $F_{2,26} = 5.73$, P = 0.0087). For the β-*CAS* gene, the expression level of the AcLi cyanotype was approximately two times higher than for the Acli and acli cyanotypes. No significant difference was detected between the expression levels of the two acyanogenic cyanotypes, although Acli showed slightly higher median and mean values than acli. Similarly, for *Aox2a,* expression in the AcLi cyanotype was approximately double those of the Acli and acli cyanotypes. For *Aox1* and *Aox2d* genes, expression levels were not significantly different between cyanotypes (**Fig. 1.5**. For *Aox1*, $F_{2,26} = 1.45$, P = 0.25. For *Aox2d*, $F_{2,26} = 1.96$, P = 0.16), although the mean and median of *Aox2d* was marginally higher in the AcLi cyanotype than in the Acli and acli cyanotypes. Together these results suggest that in white clover, upregulation of β-*CAS* and *Aox2a* activity occurs specifically in plants that produce both chemical precursors required for cyanogenesis, and that this upregulation does not require wounding-induced cyanide release in order to be induced.

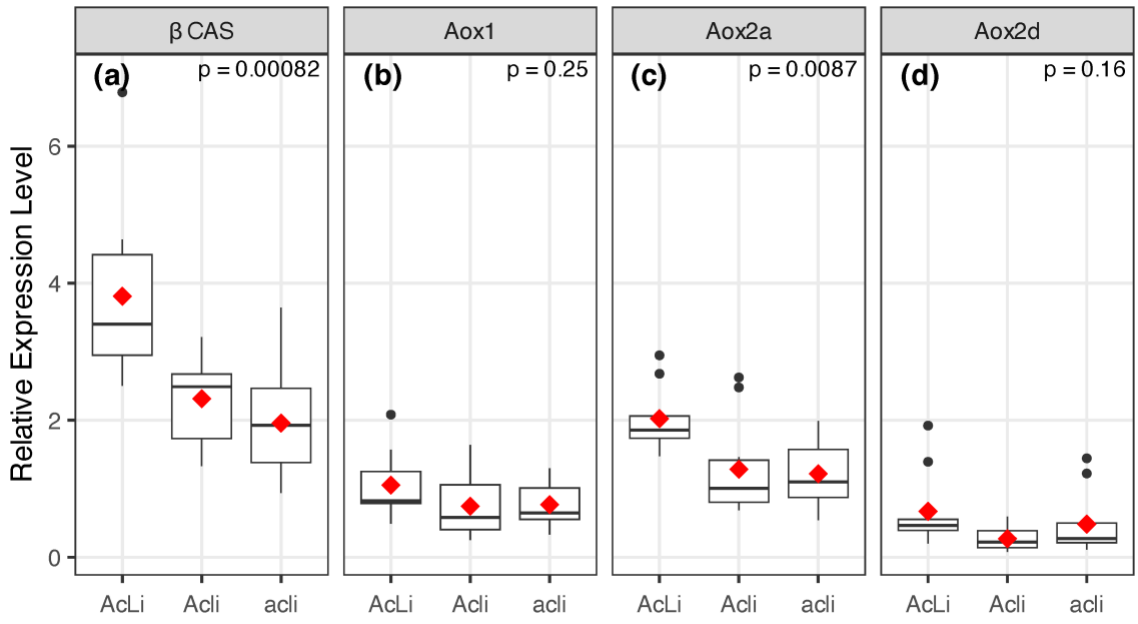


Figure 1.5 Relative expression level of β-*CAS* (a) and *Aox* genes (b-d) in different cyanotypes. Sample sizes are as follows: N_{AcLi} = 9 accessions, N_{Acli} = 10 accessions, N_{acli} = 10 accessions. The mean expression level of an individual was calculated from 3-10 leaf samples. Red diamond, mean value of each treatment. AcLi, plants with cyanogenic glucosides and glucosidase (cyanogenic); Acli, the plants with only cyanogenic glucosides (acyanogenic); acli, the plants with neither cyanogenic component (acyanogenic). Red diamond, mean value of each treatment. The *p* values were calculated by one-way ANOVA for the difference between the treatments in each panel.

1.6 Discussion

This study provides the first comprehensive identification and transcriptional characterization of the key genes controlling cyanide detoxification (β-*CAS*) and tolerance (*Aox*) pathways in a cyanogenic plant species. We first identified all members of these gene families in the white clover genome; we found two homologous β-*CAS* copies corresponding to the two diploid progenitor subgenomes (*TrCAS_to, TrCAS_tp*) and seven *Aox* genes that are distributed asymmetrically between the subgenomes and among the *Aox* subfamilies (*TrAox1_to*, *TrAox1_tp*; *TrAox2a_to_1*, *TrAox2a_to_2*, *TrAox2a_to_3*, *TrAox2a_tp*; and *TrAox2d_tp*) (**Supplementary Fig. S1.4**). The sharing of conserved motifs with known orthologs from other species (**Supplementary Figs. S1.2, S1.3**) and the concordance of gene trees with organismal phylogenetic relationships (**Figs. 1.2, 1.3**) strongly support that the inferred gene identities are correct and that they show conservation of protein function. With these genes identified, we could then exploit the natural cyanogenesis polymorphism in white clover to investigate their activity in cyanogenic vs. acyanogenic genotypes, and with or without the leaf tissue damage that induces HCN release in cyanogenic genotypes. As discussed below, our findings suggest that members of both the β-*CAS* and *Aox* families play roles in reducing and mitigating cyanide toxicity in plants that synthesize the two required cyanogenic precursors, although we did not detect further upregulation in response to wounding-induced cyanogenesis. We also found evidence for *Aox2d* expression as a general wounding response unrelated to cyanogenesis.

β-CAS-mediated cyanide metabolism is the major cyanide detoxification pathway in plants (reviewed by Gleadow & Moller, 2014; Machingura *et al.*, 2016). In white clover, we

found that β-*CAS* shows a higher expression level in cyanogenic (AcLi) genotypes than in acyanogenic genotypes (*Acli* and *acli*) in undamaged leaf tissue, with no apparent upregulation following tissue damage (**Figs. 1.4 a,e & Fig. 1.5 a**). Notably, this elevated expression was significant for AcLi genotypes only, which produce both of the required cyanogenic precursors, and not Acli genotypes, which synthesize cyanogenic glucosides but not their hydrolyzing β-glucosidase. Thus, we did not find evidence for upregulation of the cyanide detoxification pathway in association with the synthesis or metabolic turnover of cyanogenic glucosides alone. Rather, the detoxification pathway appears to be upregulated in tissues where their rapid hydrolysis to liberate free cyanide is possible.

Given this result, it is perhaps surprising that we did not detect any further increase in β-*CAS* expression in AcLi plants following the tissue wounding treatment, which was implemented specifically to induce the cyanogenic response. One potential explanation could be that the leaf-piercing protocol we used to induce tissue damage was insufficient to trigger cyanogenesis; this seems unlikely, however, as cyanide production was detectable through Feigl-Anger assays of AcLi leaf tissue that was subjected to the same leaf-piercing treatment as in the wounding experiment (and was not detected in undamaged leaf tissue, which served as a negative control) (data not shown). Another potential explanation is that in plant tissues with both cyanogenic glucosides and linamarase present, a low level of hydrolysis is occurring in undamaged tissue despite their compartmentalization, triggering β-*CAS* upregulation. In addition to this possibility, other unknown factors might also act to maintain elevated β-*CAS* expression in AcLi plants. In principle, it would be to the benefit of cyanogenic genotypes to

have the β-CAS enzyme already present before the cyanogenic response is triggered, as acute cyanide intoxication could otherwise shut down the main energy source required for activation of the detoxification pathway (Antonini *et al.*, 1971; Jones *et al.*, 1984). Additional experiments, such as those involving manipulations with exogenous HCN (e.g., Zidenga *et al.*, 2012) and/or direct assessments of enzyme activity (e.g., Miller & Conn, 1980), would be useful in further examining the cyanotype-specific pattern of β-*CAS* expression we observed.

Members of the *Aox* gene family, while sharing a conserved function as alternative electron acceptors in mitochondria, show a diverse range of functions in plants (McDonald, 2008; Sweetman *et al.*, 2019). In many plant species, *Aox1* genes have been shown to be responsive to abiotic stress, such as low temperature (Djajanegara *et al.*, 2002), drought (Bartoli *et al.*, 2005; Giraud *et al.*, 2008) or direct blocking of mitochondrial electron transportation chain (Saisho *et al.*, 1997; Polidoros *et al.*, 2005). In the present study, we did not observe any upregulation of *Aox1* following the tissue wounding treatment (**Fig. 1.4 b,f**). However, we did observe such a response in *Aox2d*, which showed a significant increase in expression in both cyanogenic and acyanogenic genotypes after tissue wounding, particularly after 60 minutes **(Fig. 1.4 d,h)**. This finding is consistent with observations from several other legume species (including soybean, chickpea, alfalfa, *Medicago truncatula*, *Lotus japonicus*, and wild peanut), where *Aox2d* has been found to be stress-inducible, showing responses to cold, drought and salinity stress (reviewed in Sweetman *et al.*, 2019).

In contrast to *Aox2d*, *Aox2a* showed upregulation in undamaged cyanogenic plants in a pattern similar to β-*CAS* (**Fig. 1.5 c).** Recent work on the *Aox* genes and their protein isoforms has indicated that the gene duplication events in the *Aox2* subfamily within legumes might have resulted in complex patterns of regulatory and functional differentiation (Sweetman *et al.*, 2019). The *Aox2a* genes typically show constitutive expression in shoot tissue in this plant family and are underrepresented in roots, which suggests that *Aox2a* may be related to photosynthetic metabolism. The similarity of expression we found between *Aox2a* and β-*CAS* raises the possibility that this particular AOX isoform could be functioning to maintain cellular respiration under the same elevated-cyanide conditions where the cyanide detoxification pathway is upregulated. As with the cyanotype-specific expression we observed for β-*CAS*, further investigations, including direct assessments of protein activity, would be useful to further examine this possibility.

While this study provides a comprehensive assessment of β-*CAS* and *Aox* gene family and subfamily representation in the white clover genome and a first investigation of their transcriptional activity in this cyanogenic species, some important caveats and limitations should be noted. We sampled multiple wild white clover genotypes representing the different cyanotypes from 19 wild populations (supplementary Table S1.) to assess how cyanide detoxification and tolerance pathways are regulated at a species level. Although the cyanotype distributions of our samples are largely overlapping geographically (**Supplementary Fig. S1.1**), there are more AcLi and Acli cyanotypes from lower latitudes, and more acli cyanotype from higher latitude, which reflects the natural abundance of the cyanotypes at those locations

(Kooyers & Olsen, 2012). We therefore cannot exclude the possibility that climate-associated genotypic differences that co-vary with cyanotype variation could affect the observed gene expression variation. However, given the largely geographically overlapping sampling of the white clover cyanotypes used in the study, and the overall lack of population structure in this species more generally (Kooyers & Olsen, 2012), we believe that any such effects are unlikely to be the primary factor in determining the observed results.

Another limitation is that qPCR primers used for gene expression analyses were not designed to be homeolog-specific, due to the high coding sequence similarity between gene copies. As a result, the expression analyses do not provide insights into subgenome dominance or other homeolog-specific patterns of gene activity that might exist. Gene subfunctionalization, involving either the partitioning of function between duplicated gene copies or divergence in expression pattern, is common after polyploidization events. Differential homeologous expression can be detected as early following a genome duplication as in the F_1 hybrid generation, as has been observed in synthetic allopolyploid Arabidopsis (Akama *et al.*, 2014); it can also persist for millions of years, such as in the paleopolyploid maize (Walsh *et al.*, 2020). White clover, as a very recently-evolved allotetraploid species (with an estimated origin of 15-28 kyr, Griffiths *et al.* (2019)) could provide a useful test case for examining the fates of duplicated genes in an evolutionarily young but stable polyploid genome. The comprehensive cataloging of β-*CAS* and *Aox* gene copies provided by the present study provides the basis for future work on this topic.

Third, the gene expression alone was examined in this study, without direct assessment of β-CAS or AOX protein activity. Both of these proteins function as dimers and are subject to post-translational regulation, so that transcript abundance alone does necessarily predict protein activity. For AOX isoforms in particular, it is well established that levels of protein activity *in vivo* involve multiple levels of regulation, both transcriptional and post-transcriptional (reviewed by Vanlerberghe *et al.*, 2020; Sweetman *et al.*, 2022). This means that protein abundance at a given time may not directly reflect gene expression (potentially reflecting the long half-life of some AOX proteins), and protein activity is not always directly predicted by protein abundance. Thus, while the cyanotype-specific expression we detected for β-*CAS* and *Aox2a* suggests that the corresponding β-CAS and AOX enzymes have elevated activity in the cyanogenic genotype, direct measures of protein function *in vivo* should be employed to test this prediction.

Implications for the white clover cyanogenesis polymorphism —

Following the discovery and characterization of the white clover cyanogenesis polymorphism in the early 20th century, classic ecological genetic studies by Daday (Daday, 1954a; Daday, 1954b; Daday, 1958) and later researchers documented the widespread occurrence of climate-associated cyanogenesis clines, with cyanogenic plants occurring at higher frequencies in warmer climates worldwide (reviewed by Kooyers & Olsen, 2012; Kooyers *et al.*, 2018). To account for the low frequency of cyanogenic plants in cold climates, Daday (1965) originally proposed that in frost-prone locations, cyanide toxicity resulting from f

reezing-induced tissue damage could impose a differential physiological cost on cyanogenic genotypes. This "autotoxicity" hypothesis predated knowledge of the β-CAS and AOX pathways, and subsequent experiments in white clover have not established a causal relationship between freezing-induced cyanogenesis and physiological damage, although there do appear to be differential physiological costs under cold stress for plants that invest in production of the cyanogenic precursors (Kooyers *et al.*, 2018; Fadoul *et al.*, 2023). The present study, by documenting differential expression of β-*CAS* and *Aox* genes in cyanogenic white clover genotypes, adds to the evidence that cyanide autotoxicity in cold climates is unlikely to be the primary selective factor in the evolution of white clover cyanogenesis clines. Rather, climate-associated variation in herbivore pressure may be the more important selective force, given the extensive evidence that cyanogenesis serves as an effective but energetically costly anti-herbivore defense in this species (discussed by Kooyers *et al.*, 2018; Wright *et al.*, 2022).

In conclusion, we provide the first gene regulatory study of the cyanide detoxification and tolerance pathways in a species that is polymorphic for cyanogenesis. We found that β-*CAS* and *Aox2a* genes show higher gene expression levels in cyanogenic white clover, even in the absence of tissue damage — suggesting that it is in plants where both cyanogenic precursors are present that upregulation of the cyanide detoxification and mitigation pathways comes into play. Our finding of functional β-CAS and AOX pathways in cyanogenic white clover genotypes further adds to evidence that freezing-induced cyanide autotoxicity is unlikely to be the major selective force favoring acyanogenic genotypes in colder climates.

1.7 Acknowledgements

The authors thank Joe Jez (Washington University in St Louis) for discussion of the experimental design and comments on the manuscript. We also thank the Washington University greenhouse staff, especially Michael Dyer and Hammy Sorkin, for care of plants used in the study. Wen-hsi Kuo was funded through the William H. Danforth Plant Science Graduate Research Fellowship in the Division of Biology and Biomedical Sciences at Washington University, and by a scholarship from the Taiwan Ministry of Education. Additional funding for the project was provided by US National Science Foundation grant IOS-1557770 to Kenneth Olsen. Finally, we thank two anonymous reviewers for their suggestions and comments.

1.8 Competing Interests

The authors declare no competing interest.

1.9 Author Contributions

WHK conducted experiments, statistical analyses, data visualization and manuscript preparation. LLS conducted a portion of the experiments and provided troubleshooting on laboratory techniques. KMO conceived the project, designed the experiments, and contributed to data interpretation and manuscript preparation.

1.10 Data Availability

All sequences used in this study have been uploaded to NCBI GenBank (**Supplementary Table S1.2 & S1.3**) and the whole-genome PacBio HiFi long reads has been uploaded to NCBI SRA PRJNA953427. The raw qPCR output is available at DRYAD (DOI: 10.5061/dryad.cvdncjt8q).

1.11 Reference

Akama S, Shimizu-Inatsugi R, Shimizu KK, Sese J. 2014. Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid Arabidopsis. *Nucleic Acids Research* **42**(6): e46.

Amirsadeghi S, Robson CA, McDonald AE, Vanlerberghe GC. 2006. Changes in plant mitochondrial electron transport alter cellular levels of reactive oxygen species and susceptibility to cell death signaling molecules. *Plant and Cell Physiology* **47**(11): 1509-1519.

Antonini E, Brunori M, Rotilio GC, Greenwood C, MalmstrOm BG. 1971. The interaction of cyanide with cytochrome oxidase. *European Journal of Biochemistry* **23**(2): 396-400.

Bartoli CG, Gomez F, Gergoff G, Guiamet JJ, Puntarulo S. 2005. Up-regulation of the mitochondrial alternative oxidase pathway enhances photosynthetic electron transport under drought conditions. *Journal of Experimental Botany* **56**(415): 1269-1276.

Berthold DA, Andersson ME, Nordlund P. 2000. New insight into the structure and function of the alternative oxidase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1460**(2-3): 241-254.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**(4): 611-622.

Cavalcanti JH, Oliveira GM, Saraiva KD, Torquato JP, Maia IG, de Melo DF, Costa JH. 2013. Identification of duplicated and stress-inducible *Aox2b* gene co-expressed with *Aox1* in species of the *Medicago* genus reveals a regulation linked to gene rearrangement in leguminous genomes. *Journal of Plant Physiology* **170**(18): 1609-1619.

Coop I. 1940. Cyanogenesis in white clover (*Trifolium repens* L.). III. A study of linamarase, the enzyme which hydrolyses lotaustralin. *The New Zealand Journal of Science and Technology. B, General section* **22**: 71-83.

Corkill L. 1942. Cyanogenesis in white clover (*Trifolium repens* L.) V. The inheritance of cyanogenesis. *The New Zealand Journal of Science and Technology. B, General section* **23**: 178-193.

Costa JH, McDonald AE, Arnholdt-Schmitt B, Fernandes de Melo D. 2014. A classification scheme for alternative oxidases reveals the taxonomic distribution and evolutionary history of the enzyme in angiosperms. *Mitochondrion* **19, Part B**: 172-183.

Daday H. 1954a. Gene frequencies in wild populations of *Trifolium repens* L. Distribution by latitude. *Heredity* **8**(1): 61-78.

Daday H. 1954b. Gene frequencies in wild populations of *Trifolium repens*. II. Distribution by altitude. *Heredity* **8**(3): 377-384.

Daday H. 1958. Gene frequencies in wild populations of *Trifolium repens* L. III. World distribution. *Heredity* **12**(2): 169-184.

Daday H. 1965. Gene frequencies in wild populations of *Trifolium repens* L. IV. Mechanism of natural selection. *Heredity* **20**(3): 355-365.

Djajanegara I, Finnegan PM, Mathieu C, McCabe T, Whelan J, Day DA. 2002. Regulation of alternative oxidase gene expression in soybean. *Plant Molecular Biology* **50**(4): 735-742.

Elias M, Sudhakaran PR, Nambisan B. 1997. Purification and characterisation of β-cyanoalanine synthase from cassava tissues. *Phytochemistry* **46**(3): 469-472.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**(1): 238.

Fadoul HE, Albano LJ, Bergman ME, Phillips MA, Johnson MTJ. 2023. Assessing the benefits and costs of the hydrogen cyanide antiherbivore defense in *Trifolium repens*. *Plants* **12**(6): 1213.

Feigl F, Anger V. 1966. Replacement of benzidine by copper ethylacetoacetate and tetra base as spot-test reagent for hydrogen cyanide and cyanogen. *Analyst* **91**(1081): 282-284.

Forslund K, Jonsson L. 1997. Cyanogenic glycosides and their metabolic enzymes in barley, in relation to nitrogen levels. *Physiologia Plantarum* **101**(2): 367-372.

Giraud E, Ho LH, Clifton R, Carroll A, Estavillo G, Tan YF, Howell KA, Ivanova A, Pogson BJ, Millar AH, et al. 2008. The absence of ALTERNATIVE OXIDASE1a in *Arabidopsis* results in acute sensitivity to combined light and drought stress. *Plant Physiology* **147**(2): 595-610.

Gleadow RM, Moller BL. 2014. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annual Review of Plant Biology* **65**: 155-185.

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019. Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31**(7): 1466-1487.

Halkier BA, Moller BL. 1989. Biosynthesis of the cyanogenic glucoside dhurrin in seedlings of *Sorghum bicolo*r (L.) Moench and partial purification of the enzyme system involved. *Plant Physiology* **90**(4): 1552-1559.

Hughes MA. 1991. The cyanogenic polymorphism in *Trifolium repens* L. (white clover). *Heredity* **66**(1): 105-115.

Jones MG, Bickar D, Wilson MT, Brunori M, Colosimo A, Sarti P. 1984. A re-examination of the reactions of cyanide with cytochrome c oxidase. *Biochemical Journal* **220**(1): 57-66.

Kongsawadworakul P, Viboonjun U, Romruensukharom P, Chantuma P, Ruderman S, Chrestin H. 2009. The leaf, inner bark and latex cyanide potential of *Hevea brasiliensis*: Evidence for involvement of cyanogenic glucosides in rubber yield. *Phytochemistry* **70**(6): 730-739.

Kooyers NJ, Hartman Bakken B, Ungerer MC, Olsen KM. 2018. Freeze-induced cyanide toxicity does not maintain the cyanogenesis polymorphism in white clover (*Trifolium repens*). *American Journal of Botany* **105**(7): 1224-1231.

Kooyers NJ, Olsen KM. 2012. Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.). *Molecular Ecology* **21**(10): 2455-2468.

Machingura M, Salomon E, Jez JM, Ebbs SD. 2016. The beta-cyanoalanine synthase pathway: Beyond cyanide detoxification. *Plant, Cell & Environment* **39**(10): 2329-2341.

McDonald AE. 2008. Alternative oxidase: an inter-kingdom perspective on the function and regulation of this broadly distributed 'cyanide-resistant' terminal oxidase. *Functional Plant Biology* **35**(7): 535-552.

Melville J, Doak B. 1940. Cyanogenesis in white clover (*Trifolium repens* L.). II. isolation of the gluco-sidal constituents. *The New Zealand Journal of Science and Technology. B, General section* **22**: 67-71.

Miller JM, Conn EE. 1980. Metabolism of hydrogen cyanide by higher plants. *Plant Physiology* **65**(6): 1199-1202.

Mirande M. 1912. Sur la présence de l'acide cyanhydrique dans le trèfle rampant (*Trifolium repens* L.). *Comptes rendus de l'Académie des Sciences* **155**: 651-653.

Mithöfer A, Boland W. 2012. Plant defense against herbivores: chemical aspects. *Annual Review of Plant Biology* **63**: 431-450.

Mizutani F, Yamanaka Y, Amano S, Hino A, Kadoya K. 1992. Effects of ethylene and hydrogen cyanide on β-cyanoalanine synthase activity in satsuma mandarin (*Citrus unshiu* Marc.) fruit. *Scientia Horticulturae* **49**(3-4): 223-231.

Møller BL. 2010. Functional diversifications of cyanogenic glucosides. *Current Opinion in Plant Biology* **13**(3): 337-346.

Nambisan B, Sundaresan S. 1994. Distribution of linamarin and its metabolising enzymes in cassava tissues. *Journal of the Science of Food and Agriculture* **66**(4): 503-507.

Olsen KM, Goad DM, Wright SJ, Dutta ML, Myers SR, Small LL, Li LF. 2021. Dual-species origin of an adaptive chemical defense polymorphism. *New Phytologist* **232**(3): 1477-1487.

Olsen KM, Hsu SC, Small LL. 2008. Evidence on the molecular basis of the *Ac*/*ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* **179**(1): 517-526.

Olsen KM, Kooyers NJ, Small LL. 2013. Recurrent gene deletions and the evolution of adaptive cyanogenesis polymorphisms in white clover (*Trifolium repens* L.). *Molecular Ecology* **22**(3): 724-738.

Olsen KM, Small LL. 2018. Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytologist* **219**(2): 757-766.

Olsen KM, Sutherland BL, Small LL. 2007. Molecular evolution of the *Li*/*li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Molecular Ecology* **16**(19): 4180-4193.

Olsen KM, Ungerer MC. 2008. Freezing Tolerance and Cyanogenesis in White Clover (*Trifolium repens* L. Fabaceae). *International Journal of Plant Sciences* **169**(9): 1141-1147.

Peiser GD, Wang T-T, Hoffman NE, Yang SF, Liu H-W, Walsh CT. 1984. Formation of cyanide from carbon 1 of 1-aminocyclopropane-1-carboxylic acid during its conversion to ethylene. *Proceedings of the National Academy of Sciences of the United States of America* **81**(10): 3059-3063.

Polidoros AN, Mylona PV, Pasentsis K, Scandalios JG, Tsaftaris AS. 2005. The maize alternative oxidase 1a (*Aox1a*) gene is regulated by signals related to oxidative stress. *Redox Report* **10**(2): 71-78.

Saisho D, Nambara E, Naito S, Tsutsumi N, Hirai A, Nakazono M. 1997. Characterization of the gene family for alternative oxidase from *Arabidopsis thaliana*. *Plant Molecular Biology* **35**(5): 585-596.

Sánchez-Pérez R, Jørgensen K, Olsen CE, Dicenta F, Møller BL. 2008. Bitterness in almonds. *Plant Physiology* **146**(3): 1040-1052.

Santangelo JS, Ness RW, Cohan B, Fitzpatrick CR, Innes SG, Koch S, Miles LS, Munim S, Peres-Neto PR, Prashad C, et al. 2022. Global urban environmental change drives adaptation in white clover. *Science* **375**(6586): 1275-1281.

Selinski J, Scheibe R, Day DA, Whelan J. 2018. Alternative oxidase is positive for plant performance. *Trends in Plant Science* **23**(7): 588-597.

Selmar D, Lieberei R, Biehl Bl. 1988. Mobilization and utilization of cyanogenic glycosides: the linustatin pathway. *Plant Physiology* **86**(3): 711-716.

Smith CA, Melino VJ, Sweetman C, Soole KL. 2009. Manipulation of alternative oxidase can influence salt tolerance in *Arabidopsis thaliana*. *Physiologia Plantarum* **137**(4): 459-472.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9): 1312-1313.

Sweetman C, Selinski J, Miller TK, Whelan J, Day DA. 2022. Legume alternative oxidase isoforms show differential sensitivity to pyruvate activation. *Frontiers in Plant Science* **12**.

Sweetman C, Soole KL, Jenkins CLD, Day DA. 2019. Genomic structure and expression of alternative oxidase genes in legumes. *Plant, Cell & Environment* **42**(1): 71-84.

Tschiersch B. 1964. Metabolism of hydrocyanic acid-III: Assimilation of H14CN by *Lathyrus odoratus* L., *Vicia sativa* L., and *Ricinus communis* L. *Phytochemistry* **3**(3): 365-367.

Vanlerberghe GC. 2013. Alternative oxidase: A mitochondrial respiratory pathway to maintain metabolic and signaling homeostasis during abiotic and biotic stress in plants. *International Journal of Molecular Sciences* **14**(4): 6805-6847.

Vanlerberghe GC, Dahal K, Alber NA, Chadee A. 2020. Photosynthesis, respiration and growth: A carbon and energy balancing act for alternative oxidase. *Mitochondrion* **52**: 197-211.

Walsh JR, Woodhouse MR, Andorf CM, Sen TZ. 2020. Tissue-specific gene expression and protein abundance patterns are associated with fractionation bias in maize. *BMC Plant Biology* **20**: 1-11.

Williams WM, Ellison NW, Ansari HA, Verry IM, Hussain SW. 2012. Experimental evidence for the ancestry of allotetraploid *Trifolium repens* and creation of synthetic forms with value for plant breeding. *BMC Plant Biology* **12**: 55.

Wright SJ, Goad DM, Gross BL, Muñoz PR, Olsen KM. 2022. Genetic trade-offs underlie divergent life history strategies for local adaptation in white clover. *Molecular Ecology* **31**(14): 3742-3760.

Yi H, Jez JM. 2012. Assessing functional diversity in the soybean β-substituted alanine synthase enzyme family. *Phytochemistry* **83**: 15-24.

Yi H, Juergens M, Jez JM. 2012. Structure of soybean beta-cyanoalanine synthase and the molecular basis for cyanide detoxification in plants. *Plant Cell* **24**(6): 2696-2706.

Yu LL, Liu Y, Liu CJ, Zhu F, He ZQ, Xu F. 2020. Overexpressed β-cyanoalanine synthase functions with alternative oxidase to improve tobacco resistance to salt stress by alleviating oxidative damage. *FEBS Letters* **594**(8): 1284-1295.

Zidenga T, Leyva-Guerrero E, Moon H, Siritunga D, Sayre R. 2012. Extending cassava root shelf life via reduction of reactive oxygen species production. *Plant Physiology* **159**(4): 1396-1407.

1.12 Supporting Information

Supplementary Table S1.1 Plant accessions

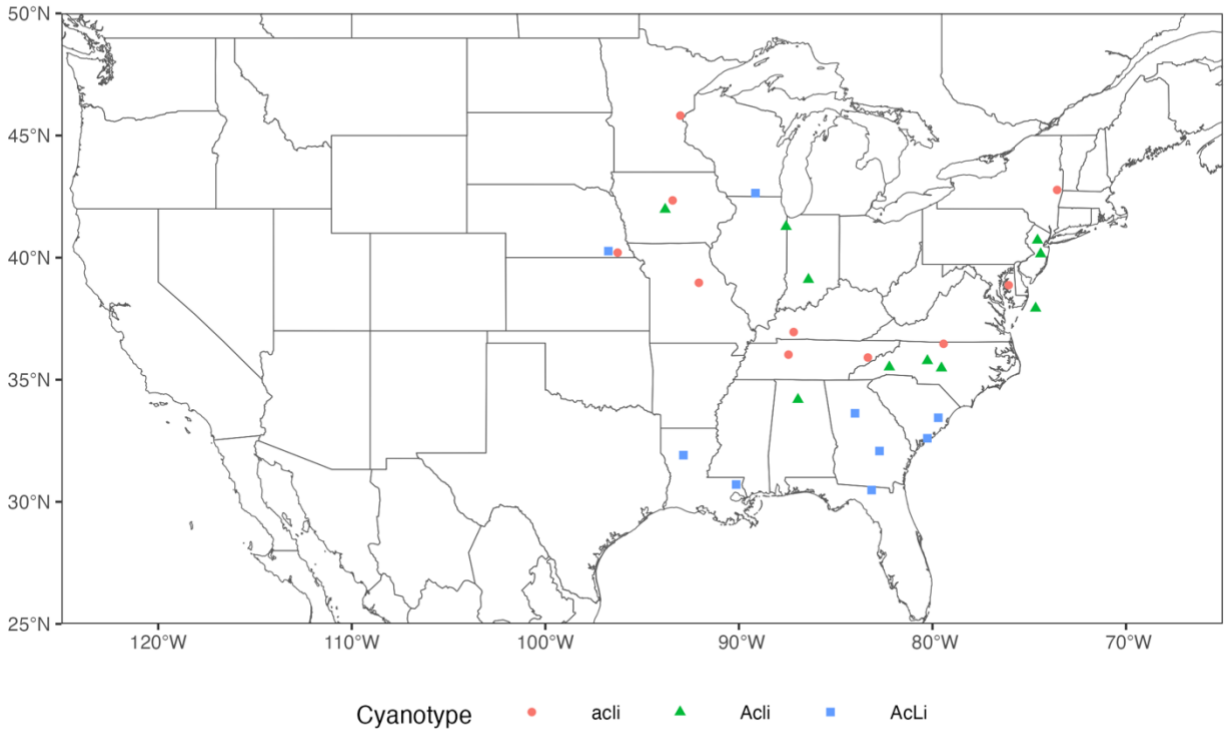
Accession	Latitude	Longitude	Nearest City	Collector/provider	Used in tissue wounding experiment	Used in experiment without tissue wounding
			Cyanotype: *AcLi*			
CSC_006	32.795278	-80.080278	Charleston, SC	Sandra Branham	YES	YES
CSC_010	32.793611	-80.109722	Charleston, SC	Sandra Branham		YES
DGA_006	34.124712	-83.215489	Danielsville, GA	Ed McAssey	YES	YES
DWI_009	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	YES	YES
LNE_002	40.835631	-96.648226	Lincoln, NE	Trisha Spanbauer		YES
MLA_007	32.513056	-92.119722	Monroe, LA	John McIntire	YES	YES
NLA_003	29.928889	-90.078056	New Orleans, LA	Sara Wright	YES	YES
TGA_008	31.458046	-83.517106	Tifton, GA	Patricio Munoz		YES
TGA_011	31.458046	-83.517106	Tifton, GA	Patricio Munoz	YES	YES
			Cyanotype: *Acli*			
AIA_007	42.046667	-93.621111	Ames, IA	Sara Wright		YES
ANC_002	35.56833	-82.568333	Asheville, NC	Rebecca Thomson		YES
BAL_020	33.487966	-86.817766	Birmingham, AL	Nicole Riddle		YES
CIL_009	41.963611	-87.631389	Chicago, IL	Sara Wright		YES
HDE_012	38.725248	-75.283731	Harbeson, DE	John McDonald		YES
LKY_007	38.283056	-85.618056	Louisville, KY	Sunita Crittenden		YES
NYC_009	40.782182	-73.965913	New York City, NY	Dan Cui Zhou		YES
NYC_014	40.782182	-73.965913	New York City, NY	Dan Cui Zhou		YES
WNC_005	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira		YES
WNC_010	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira		YES
			Cyanotype: *acli*			
AIA_009	42.046667	-93.621111	Ames, IA	Sara Wright	YES	YES
ANC_008	35.566389	-82.573333	Asheville, NC	Rebecca Thomson	YES	YES
CMO_010	38.962766	-92.25064	Columbia, MO	Sara Wright	YES	YES
HDE_006	38.725248	-75.283731	Harbeson, DE	John McDonald	YES	YES
LNE_004	40.836309	-96.651577	Lincoln, NE	Trisha Spanbauer		YES
MMN_001	45.081522	-93.199453	Minneapolis, MN	Mary Lyman-Onkka		YES
NTN_004	36.184444	-86.758333	Nashville, TN	Sara Wright		YES
NTN_014	36.148775	-86.812494	Nashville, TN	Sara Wright	YES	YES
SCT_003	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	YES	YES
WNC_004	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	YES	YES
			Non-cyanogenic species (As a negative control)			
T. stoloniferum				USDA accession PI 631732	YES	YES

Supplementary Table S1.2 Genes and primers used in the study

Gene	Accession	Targeted region	Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon size (bp)
TrCAS	MN872236	Exon5/Exon6	GGTCCTGAGATATGGGAGG	ACTTTCAGATGGCTCCACTC	147
TrAox1	MN872239	Exon3/ Exon3	GGTTGGTTACCTTGAGGAAGAAG	ATTGCCTTTATCAAGCTCCTTG	70
TrAox2a	MN872241	Exon2/Exon3	GAAGCTTCTCAGAATTCCCAC	ATCGCACGACAACCATACTG	60
TrAox2d	MN872243	Exon3/ Exon3	GGTTATTTGGAAGAAGAAGCTGTGA	ACCACTTTCAATTGCATTTAAATGT	63

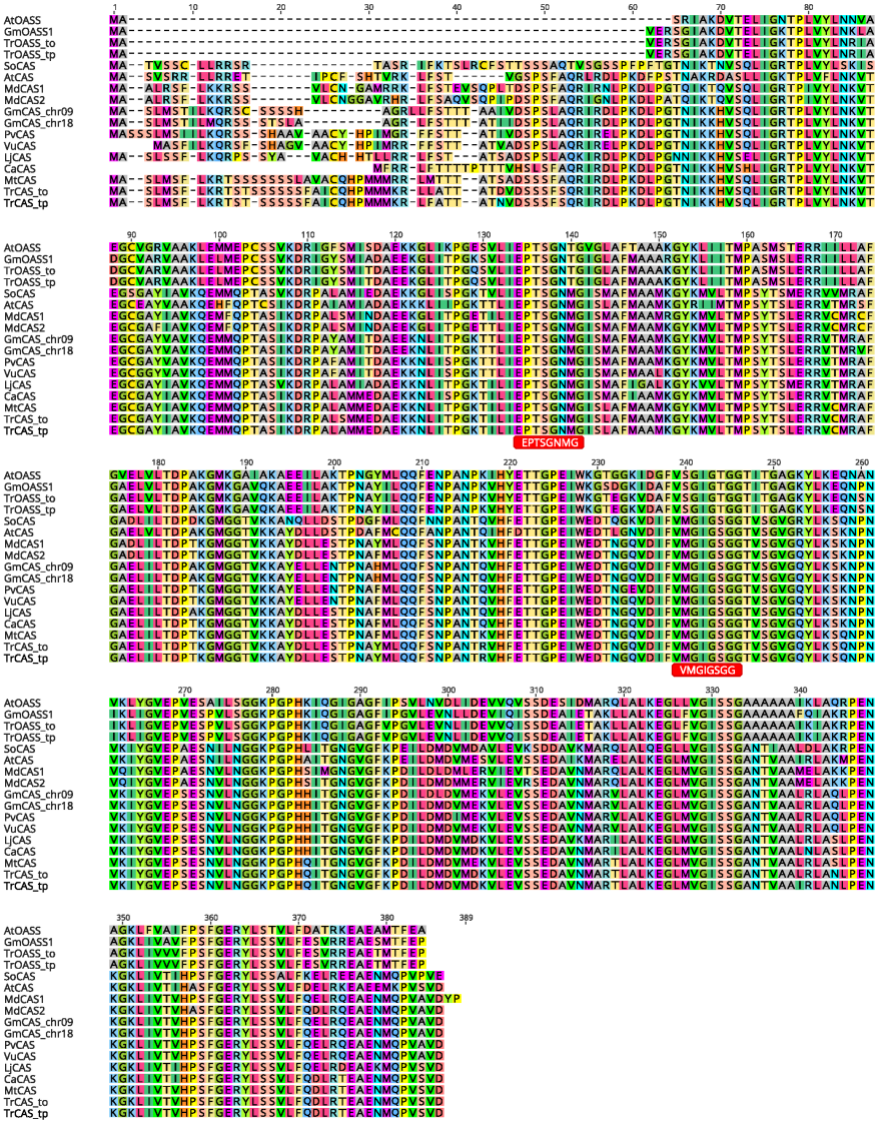
Supplementary Table S1.3 Compete gene sequences isolated in this study

Gene name	NCBI accession
TrOASS_tp	OP437418
TrOASS_to	OP437419
TrCAS_tp	OP437420
TrCAS_to	OP437421
TrAox2d_tp	OP437422
TrAox1_tp	OP437423
TrAox1_to	OP437424
TrAox2a_to_3	OP437425
TrAox2a_to_2	OP437426
TrAox2a_to_1	OP437427
TrAox2a_tp	OP456086



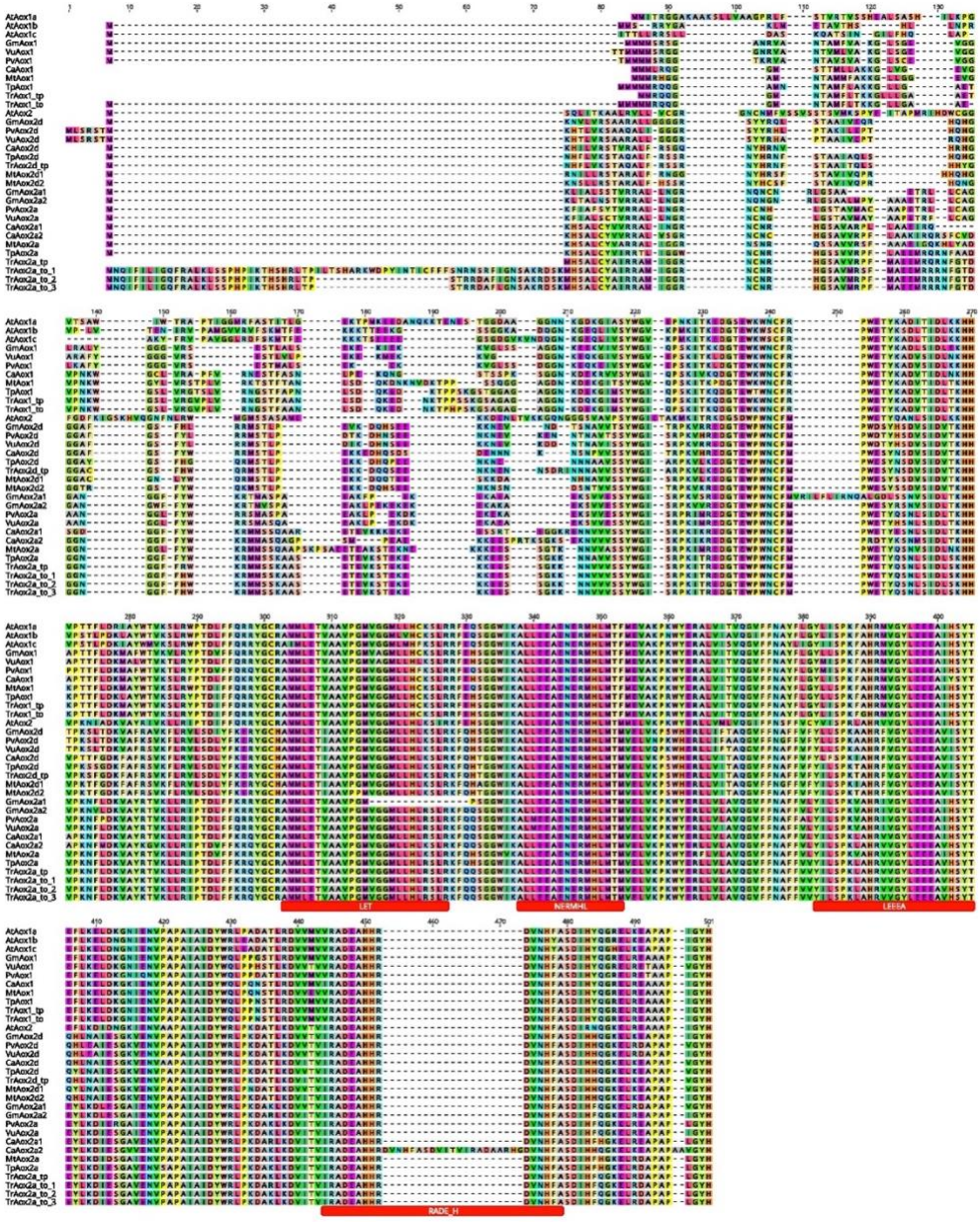
Supplementary Figure S1.1 Map of the plant accessions.

The points were jittered to avoid the overlaps of the close geographic locations.



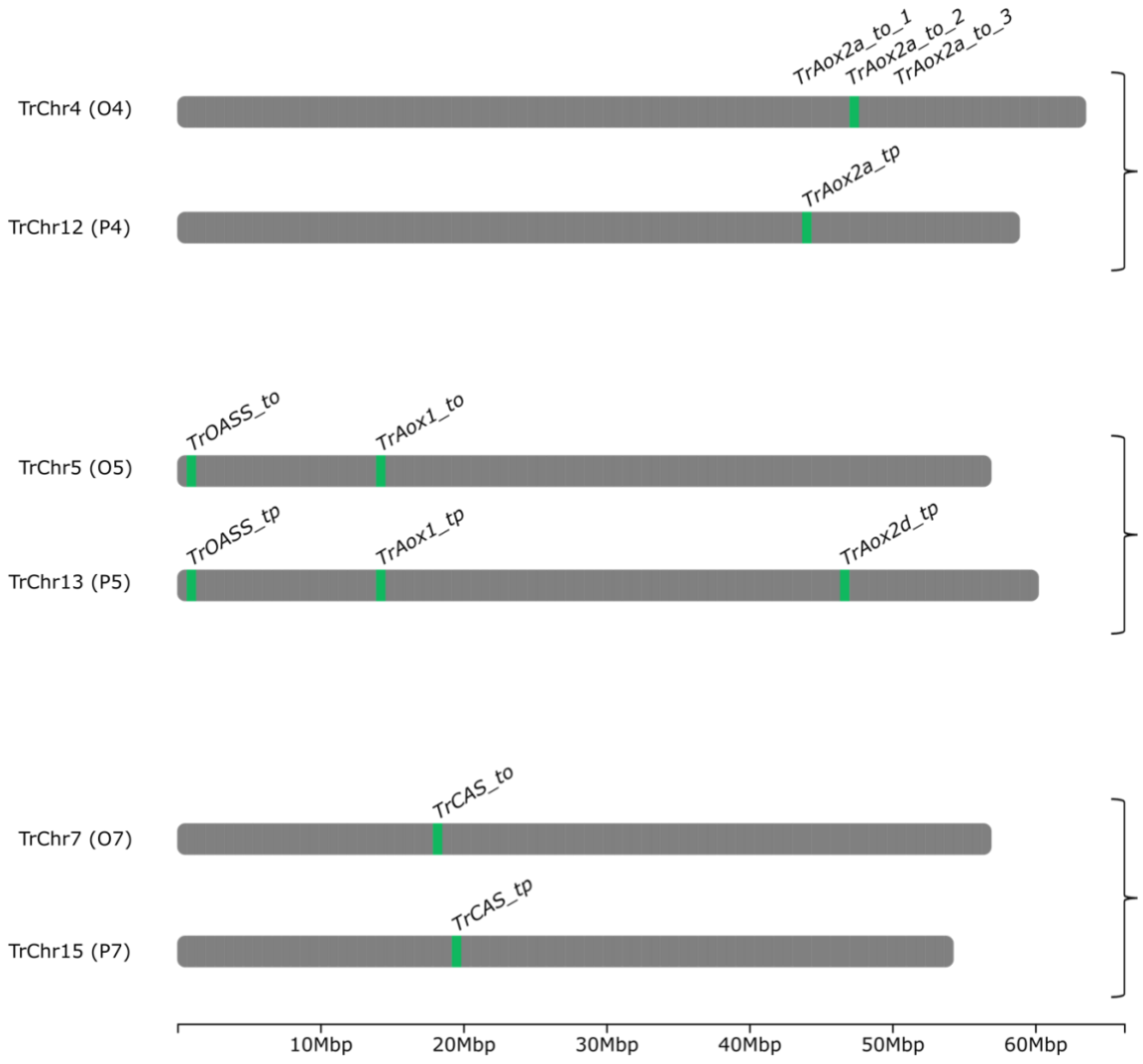
Supplementary Figure S1.2 Protein sequence alignment of the β-substituted alanine synthase (BSAS) family.

The two red tiles indicate the core reaction centers which confer OASS vs. β-CAS substrate specificity (specifically residues T140M, S240M, and T244S) (Yi *et al.*, 2012). The first ~50 amino acid sequence in the alignment is a mitochondrial localization peptide unique to β-CAS peptides. The first two letters of sequence labels indicate species as follows: *Arabidopsis thaliana* (At), *Glycine max* (Gm), *Trifolium repens* (Tr), *Spinacea oleracea* (So), *Malus domestica* (Md), *Phaseolus vulgaris* (Pv), *Vigna unguiculata* (Vu), *Lotus japonica* (Lj), *Cicer arietinum* (Ca), *Medicago truncatula* (Mt).



Supplementary Figure S1.3 Alignment of mitochondrial alternative oxidase (AOX) protein sequences.

Red tiles indicate four highly conserved motifs characteristic of AOX sequences (Berthold *et al.*, 2000). *Tp* indicates *Trifolium pratense* (red clover); see Fig. S2 for all other species labels.



Supplementary Figure S1.4 Physical locations of the β-*CAS* and *Aox* genes in the white clover genome.

Homeologous chromosomes are grouped in pairs, where chromosomes 1 to 8 correspond to the *T. occidentale* (O) subgenome, and chromosomes 9-16 correspond to the *T. pallescens* (P) subgenome.

1.13 Supplementary Reference

Berthold DA, Andersson ME, Nordlund P. 2000. New insight into the structure and function of the alternative oxidase. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1460**(2-3): 241-254.

Yi H, Juergens M, Jez JM. 2012. Structure of soybean beta-cyanoalanine synthase and the molecular basis for cyanide detoxification in plants. *Plant Cell* **24**(6): 2696-2706.

Chapter 2 *De novo* genome assembly of white clover (*Trifolium repens* L.) reveals the role of Copy Number Variation in rapid environmental adaptation

2.1 Authorship & Affiliations

Wen-Hsi Kuo[1], Sara J. Wright[1,2], Linda L. Small[1], and Kenneth M. Olsen[1]

[1]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130 USA

[2]Present address: Department of Biological and Biomedical Sciences, Rowan University, Glassboro, NJ 08028 USA

This chapter is currently under revision in *BMC Biology*.

2.2 Summary

Background: White clover (*Trifolium repens*) is a globally important perennial forage legume. This species also serves as an eco-evolutionary model system for studying within-species chemical defense variation; it features a well-studied polymorphism for cyanogenesis (HCN release following tissue damage), with higher frequencies of cyanogenic plants favored in warmer locations worldwide. Using a newly-generated haplotype-resolved genome and two other long-read assemblies, we tested the hypothesis that copy number variants (CNVs) at cyanogenesis genes play a role in the ability of white clover to rapidly adapt to local environments. We also examined questions on subgenome evolution in this recently evolved allotetraploid species and on chromosomal rearrangements in the broader IRLC legume clade.

Results: Integration of PacBio HiFi, Omni-C, Illumina and linkage map data yielded the first completely *de novo* genome assembly for white clover (created without *a priori* sequence assignment to subgenomes). We find that white clover has undergone extensive transposon diversification since its origin but otherwise shows highly conserved genome organization and composition with its diploid progenitors; unlike some other clover species, its chromosomal structure is conserved with other IRLC legumes. We further find extensive evidence of CNVs at the major cyanogenesis loci; these contribute to quantitative variation in the cyanogenic phenotype and to local adaptation across wild North American populations.

Conclusions: This work provides a case study documenting the role of CNVs in local adaptation in a plant species, and it highlights the value of pan-genome data for identifying contributions of structural variants to adaptation in nature.

Keywords

Allopolyploid; copy number variation (CNV); cyanogenesis; haplotype-resolved genome; IRLC legumes; karyotype; local adaptation; pangenome; subgenome; white clover (*Trifolium repens*)

2.3 Introduction

White clover (*Trifolium repens* L., 2n = 4x = 32) is the most widely grown temperate forage legume worldwide due to its superior nitrogen-fixing ability, perenniality, and resilience in frequently disturbed habitats (Zeven, 1991; Taylor, 2008; Abberton & Marshall, 2010). Prior to the 20th century invention of synthetic nitrogen fertilizer, this species was so important as a source of soil nitrogen that it was considered the agricultural equivalent of coal in fueling agricultural expansion during the Industrial Revolution (Kjærgaard, 2003). In addition to its longstanding agricultural importance, white clover is also a well-known eco-evolutionary model system for studying how selection maintains within-species chemical defense variation. The species is characterized by a genetically well characterized polymorphism for cyanogenesis (HCN release following tissue damage); multiple studies over the last 80 years have established that climate-associated clines in cyanogenesis evolve rapidly across latitudinal and other environmental gradients worldwide, with cyanogenic plants generally predominating in warmer locations (Daday, 1954a; Daday, 1954b; Kooyers & Olsen, 2013; Santangelo *et al.*, 2022). While native to Europe, white clover has become widely naturalized in mesic environments worldwide (from subtropical to boreal climates), and recent genomic research has linked its ability to become so widely adapted to its allopolyploid origin from two ecologically distinct diploid progenitor species (Griffiths *et al.*, 2019).

White clover's importance both in agriculture and as an eco-evolutionary model system has spurred recent efforts to develop high-quality genetic and genomic resources for the species (Ravagnani *et al.*, 2012; Griffiths *et al.*, 2019; Olsen *et al.*, 2021; Santangelo *et al.*, 2023;

Wang *et al.*, 2023). Most recently, two independent genome sequencing projects have utilized long-read sequencing technologies to overcome the scaffolding difficulties associated with polyploid genome assembly (Santangelo *et al.*, 2023; Wang *et al.*, 2023). While these two high quality genomes are a major step in white clover studies, the recent growth of plant pangenome projects (e.g. Alonge *et al.*, 2020; Zmienko *et al.*, 2020; Qin *et al.*, 2021) has demonstrated clearly that multiple high quality reference genomes are required to properly understand the genomic structure and intraspecific diversity of a species; this is especially true when patterns of genomic structural variation contribute to natural phenotypic variation. In addition, high-quality reference genomes can facilitate investigations of genome evolution on a macroevolutionary scale. For white clover specifically, genome sequence comparisons between this species and related legume genera could fill a current knowledge gap concerning karyotype evolution in the "IRLC clade" of the legume family, which contains multiple economically important crop species and is characterized by extensive genome rearrangements (Zhuang *et al.*, 2019).

Among the many types of structural variation that can occur in a genome, copy number variation (CNV), defined as the variable repetition of specific sequence motifs ranging from 50 bp to several Mbp, is a major contributor to both genetic and phenotypic variability across eukaryotes (Wright *et al.*, 2009; Cook *et al.*, 2012; Wang *et al.*, 2015; Pos *et al.*, 2021; Stalder *et al.*, 2023). Despite abundant evidence that CNVs are pervasive in genomes and that they contribute to phenotypic variation, there are remarkably few documented cases where CNVs have been shown to underlie adaptation in wild species (Ishikawa *et al.*, 2019; Ishikawa *et al.*,

2021; Kreiner *et al.*, 2023). To the best of our knowledge, the only one example in plant species is the flowering time regulation in common waterhemp; CNVs of the gene set of ATP synthesis pathway is related to the days require to flower. However, the causal function of the CNV and the mechanism of flowering time regulation in adaptation are still ambiguous. Consequently, our knowledge about the role of CNVs in shaping local adaptation in wild populations is scarce, particularly for plant species. Even in the model plant Arabidopsis, studies of CNVs have been restricted to either documenting genome-wide CNV distributions without knowledge of associated phenotypes (e.g., Zmienko *et al.*, 2020), or to functional characterization of the phenotypic impact of CNVs without data from natural populations to assess their role, if any, in adaptation (e.g., DeBolt, 2010).

Based on our prior knowledge of white clover and the genetic basis of adaptation in this species, we hypothesized that local environmental adaptation in wild populations could arise, in part, through CNVs, particularly at the loci known to control the well-documented cyanogenesis polymorphism. This chemical defense polymorphism is controlled by two independently segregating simple Mendelian genetic polymorphisms that determine the presence or absence of two cyanogenic chemical precursors, both of which must be present for a plant to produce the cyanogenic phenotype: 1) *Ac/ac*, controlling the presence/absence of cyanogenic glucosides; and 2) *Li/li*, controlling the presence/absence of their hydrolyzing cyanogenic β-glucosidase enzyme, linamarase (reviewed by Hughes, 1991). At the molecular level, *Ac* is a 3-gene metabolic cluster on chromosome 2 that comprises the cyanogenic glucoside biosynthetic pathway, while *Li* is a single gene located on chromosome 12 that encodes linamarase (Olsen *et*

al., 2007; Olsen *et al.*, 2008; Olsen & Small, 2018; Olsen *et al.*, 2021). For both loci, the recessive (nonfunctional) alleles are the result of gene deletions, meaning that both the *Ac/ac* and *Li/li* polymorphisms are gene presence/absence variations (PAVs) (Olsen *et al.*, 2007; Olsen *et al.*, 2014; Olsen & Small, 2018). However, while the recessive *ac* and *li* alleles have consistently been shown to be gene deletions (Olsen *et al.*, 2013; Kooyers & Olsen, 2014), our observations of *Ac* and *Li* dominant allele inheritance in greenhouse pedigree populations suggested the occurrence of CNVs at both loci (see Results). Given that variability in the cyanogenesis phenotype is known to contribute to white clover's adaptation across climatic gradients (Hughes, 1991; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013; Santangelo *et al.*, 2022), we therefore hypothesized that CNVs at the *Ac* and *Li* loci could play an important role in this adaptive response.

In this study we tested the hypothesis that CNVs at the *Ac* and *Li* cyanogenesis loci contribute to natural adaptive variation in white clover. We used PacBio HiFi, Omni-C and two linkage maps to generate and *de novo* assemble a new chromosome-scale and haplotype-resolved genome. We then compared structural variation at the cyanogenesis loci among three high quality white clover genomes (this study; Santangelo *et al.*, 2023; Wang *et al.*, 2023). Next, we assessed *Ac* and *Li* CNV occurrence and distributions across wild populations spanning much of the North American species range (419 accessions across 43 locations). In a complementary analysis, we examined the contribution of CNVs to natural phenotypic variation by assessing the relationship between *Ac* CNVs, gene expression, and cyanogenic glucoside content. We further tested whether CNVs at the cyanogenesis loci contribute to adaptation in nature by

conducting association analyses between the CNVs and the local environments of the sampled wild populations. Finally, we leveraged our high-quality genome to address questions concerning karyotype evolution in the clover genus *Trifolium* and related genera in the "IRLC clade" of the legume family (Fabaceae). To our knowledge, this study is the first to document the contribution of CNVs to local environmental adaptation in a plant species, and it also provides new insights into the history of genome rearrangements in an economically important clade of legume species.

2.4 Methods and Methods

De novo genome assembly

Young leaf tissue of a wild North American white clover accession "GFL_007" (grown from seed collected in Gainesville, Florida; SAMN37329216) was dark-treated for 48 hours before DNA isolation. High molecular weight (HMW) DNA was isolated by Polar Genomics (Ithaca, NY) using a CsCl gradient to remove mitochondrial and chloroplast genomes. The HMW DNA was then submitted to the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign, Urbana, IL 61801) for PacBio HiFi sequencing. The HMW DNA was sheared with a Megaruptor 3 to a size of ~15 Kbp. Sheared gDNA fragments were converted to a library with the SMRTBell Express Template Prep kit 2.0. The library was sequenced on three SMRTcell 8M on a PacBio Sequel II using the circular consensus sequencing (CCS) sequencing mode and a 30-hour movie time. CCS analysis was performed using SMRTLink V8.0 using the following

parameters: ccs --min-length 1000 --max-length 50000 --min-passes 3 --min-rq 0.99. By this approach, PacBio HiFi generated 7.3 million reads with average length 11.6 Kbp. The expected mean coverage is 77x for a collapsed haploid genome based on the sequencing depth and previously reported genome size (Griffiths *et al.*, 2019). In order to phase the diploid genome and scaffold, Omni-C was used to capture the chromatin proximity interactions. The Omni-C library and sequencing services were also provided by the Roy J. Carver Biotechnology Center. One hundred million read pairs (2×150 bp) were sequenced by using NovaSeq 6000 System with SP flow cell (Illumina). The raw reads were demultiplexed and the adaptors were trimmed using the Illumina standard software.

Although extant white clover genome is likely to have minimal inter-subgenome translocations (e.g., homeologous recombination) (Griffiths *et al.*, 2019), we cannot exclude that translocations might have occurred following polyploidization. Therefore, unlike the most recent previous assembly (Santangelo *et al.*, 2023), which categorized the PacBio HiFi reads into subgenomes before final assembly, we conducted a complete *de novo* assembly from the raw PacBio HiFi reads. Hifiasm v0.16.1 was used for the *de novo* genome assembly with Hi-C integrated mode (Cheng *et al.*, 2021). The output of the fully phased contigs of the primary and alternative haplotigs were used as the template for Omni-C scaffolding independently. The adaptor-trimmed Omni-C reads were mapped to the two haplotigs independently according to the mapping pipeline provided by Arima Genomics (2019). In short, the pipeline first maps the reads to the reference in single-end mode. Then the chimeric mapped reads (if present) were

processed to retain only the 5'-end of the sequence (on the assumption that the 3'-end was a result of proximal ligation). Finally, the paired reads were joined and sorted in the output.

The genome scaffolding was conducted in two steps. First, the Hi-C scaffolding was performed by SALSA2 with the default settings (Ghurye *et al.*, 2019). Second, two linkage maps created from two F_2 mapping populations (see linkage map construction below) were used to place the Hi-C scaffolds into pseudochromosomes. The two linkage maps were merged based on the physical locations of the markers in the corresponding contigs, and the genetic distances (cM) were re-estimated using either the mean genetic distance of the two linkage maps or the only available value if the marker was only present in one linkage map. Chromonomer was used to integrate the information of the Hi-C scaffolding and the linkage map and then generate the final genome assembly information file (AGP file) (Catchen *et al.*, 2020). Then, agptools was used to update the final assembly (Rice). The chromosome numbers were labeled by the BLAST against the diploid progenitors' genome (see subgenome characterization) and the nucleotide alignment to *Medicago truncatula* (see gene synteny and nucleotide alignment). The complete chloroplast and the partial mitochondrial genomes were assembled from the Illumina reads in GetOrganelle (Jin *et al.*, 2020). Then, the chloroplast and mitochondrial genome, plus the NCBI prokaryotic RefSeq genomes were used as the reference in BLAST to remove the contaminations from the nuclear genome.

k-mer based genome quality assessment

To estimate the reference-free genome size and heterozygosity before assembly, an independent Illumina short read (2×150 bp) dataset of the same accession (GFL_007; SAMN37329216) was used to generate a histogram of k-mer frequencies in jellyfish with default settings (k-mer length of 21) (Marçais & Kingsford, 2011). Then, the estimation was performed by GenomeScope (Vurture *et al.*, 2017). To assess the assembly quality, the same Illumina dataset was used and the assessment was performed by Merqury (Rhie *et al.*, 2020).

Linkage map construction

Two linkage maps were used to scaffold and assess the scaffolding quality of the new assembly. These were generated from two F_2 mapping populations derived from biparental crosses of GFL_007 and two other white clover genotypes [DG population from a cross of accessions DMN_010 (SAMN34157026) and GFL_007 (SAMN37329216); GS population from accessions GFL_007 (SAMN37329216) and STL_0701 (SAMN34157027)] (Olsen *et al.*, 2021; Wright *et al.*, 2022). Protocols for generation of genotyping-by-sequencing (GBS) marker data used in linkage map construction are described in (Olsen *et al.*, 2021). For the present study, *de novo* linkage map construction was performed by mapping raw GBS reads to the two Omni-C scaffolded haplotigs; SNPs were called independently following the GATK best practice workflow (Poplin *et al.*, 2017). In brief, the reads were mapped to the reference by Bowtie2 (bowtie2 --sensitive --no-mixed --no-discordant --minins 100 --maxins 1000). Then, the SNPs were called by HaplotypeCaller (-ERC GVCF) and combined by GenotypeGVCFs. The output in vcf format was first hard-filtered (bcftools filter -e 'QD < 0.5 || FS > 200.0 || MQ < 20.0 ||

MQRankSum < -12.5 || MQRankSum > 12.5 || ReadPosRankSum < -8.0 || SOR > 8 || INFO/DP < 2000'), then individuals with missing genotype (SNP site) >0.25 were removed, then filtered for minor allele frequency >0.35, max number of alleles <=2, p-value >1e-20 on a genotype frequency test (indicating no significant deviation from 1:2:1 segregation in the F_2 populations), minimum mean depth >15, and missing data <0.25; finally, only genotypes (SNP sites) that were homozygous in the parents were preserved. The linkage maps were then constructed by GUSMap package with the standard protocol with the minimum LOD threshold that can form 16 linkage groups (Bilton *et al.*, 2018). The quality of the final phased genome assembly (primary haplotig) was further validated by serving as the reference for mapping the raw GBS reads. All procedures for raw read mapping and follow-up steps were the same as for linkage map construction, except for that all the individuals of the mapping populations were kept. The final vcf files were converted to the "cross" format and imported to the R/qtl package (Broman *et al.*, 2003). Then the heatmaps of the LOD scores and the recombination fractions between the pairwise markers were visualized.

Subgenome categorization

In order to characterize the subgenome identities of either *T. occidentale* or *T. pallescens* origin, the final assembly (primary and alternative haplotigs) was used as the query to BLAST (which allows multiple alignment hits to ambiguous genomic regions) against published genomes of the two diploid progenitors, *T. occidentale* and *T. pallescens* (Griffiths *et al.*, 2019). Only the hits with alignment >1000 bp and identity >95% were kept. Then, a customized index was calculated based on: $\sum(To\ alignment\ length)_i\ -$

$\sum(Tp\ alignment\ length)_k\ /Bin\ size\ (0.5\ Mb)$, where *i* and *k* are the filtered hits to either the *T. occidentale* (To) or *T. pallescens* (Tp) subgenome, respectively. This measure creates an index where a value of -1 is completely *T. pallescens*-like and +1 is completely *T. occidentale*-like.

Gene prediction and functional annotation

The two haplotigs of the diploid assembly were independently annotated with the BRAKER2 pipeline (Brůna *et al.*, 2021). The haploid genome was first softmasked for the repeated regions by RepeatModeler v2.0.4 and RepeatMasker v4.1.4. Then, the gene model was separately predicted with RNA-seq and protein data. The results were then combined by TSEBRA as suggested (Gabriel *et al.*, 2021). The RNA-seq data were derived from leaf tissue samples of the three accessions used to create the linkage maps (DMN_010, GFL_007, STL_0701); tissue was collected from plants grown in the greenhouse under control and drought treatments (N = 18, including three accessions and three clonal replications of each accession). The RNA-seq reads were mapped to the reference in a relaxed mode (--outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0) by STAR v2.7.10a (Dobin *et al.*, 2012). The protein data were from the translated sequences of the previous published white clover genome (Griffiths *et al.*, 2019) and OrthoDB v10 'Fabales' database. For functional annotation, the primary transcripts of each gene model were translated into protein sequences and analyzed using the blastp algorithm against the NCBI RefSeq genomes of Arabidopsis, soybean (*Glycine max*) and *Medicago truncatula*. The BLAST result was then imported in

Blast2GO software. After combining the results from InterProScan, the GO terms were mapped and annotated following standard protocols (Conesa *et al.*, 2005).

Gene synteny and nucleotide alignment

Interspecific gene synteny analysis was conducted for related legume species by GENESPACE v0.9.3 (Lovell *et al.*, 2022), which integrates the results from OrthoFinder (Emms & Kelly, 2019) and MCscanX (Wang *et al.*, 2012). In short, the primary transcripts of *Trifolium pratense* (Bickhart *et al.*, 2022), *Trifolium subterraneum* (Shirasawa *et al.*, 2023), *Medicago truncatula* (Pecrix *et al.*, 2018), *Cicer arietinum* (Varshney *et al.*, 2013), *Lotus japonicus* (Li *et al.*, 2020), *Phaseolus vulgaris* (Schmutz *et al.*, 2014) were used as the input in OrthoFinder in nucleotide mode. Then, the gene synteny was generated by MCscanX with the settings: blkSize = 20, nGaps = 25. Nucleotide alignment was conducted by Mummer 4.0.0 with the function: nucmer (Marçais *et al.*, 2018). Only the alignment hits with more than 1000 bp were kept. The gene synteny and nucleotide alignment were then visualized using package gggenomes v0.9.5 in R.

Synonymous mutation rate (K_S)

To assess the divergence among orthologs across species and examine subgenome-specific patterns of molecular evolution within white clover (e.g., relaxed purifying selection), synonymous mutation rate (K_S) and nonsynonymous mutation rate (K_A) was calculated based on the single copy orthologous genes that are universally available in all the taxa in the synteny analysis (except for *Trifolium subterraneum*). After filtering, 3884 genes were kept. The

orthologous gene sequences were aligned and guided by their translated protein sequences in MACSE v2 (Ranwez *et al.*, 2018). The low-quality alignment was removed following the OMM_MACSE pipeline. The cleaned alignments were then used for mutation rate calculation in KaKs_Calculator 2.0 with YN method (Wang *et al.*, 2010).

Cyanogenesis gene copy number variation, gene expression and cyanide quantification

Plant materials used to characterize cyanogenesis variation were collected as seeds or stolons from North American wild populations and cultivated in the greenhouse of Washington University in St. Louis (**Supplementary Table S2.7**). The cyanogenic phenotype, including the presence/absence of the cyanogenic glucosides and the substrate-specific glucosidase, was assessed biochemically by Feigl-Anger cyanogenesis assay and genetically by PCR-genotyping of *CYP79D15* (the first gene in the *Ac* metabolic cluster) and *Li* (Olsen *et al.*, 2007; Olsen *et al.*, 2008). Plants that produced one or both cyanogenic precursors were further examined by qPCR with gDNA to estimate the gene copy number information, and with cDNA to estimate the relative gene expression level. The qPCR primers of *CYP79D15*, *CYP736A187*, and *UGT85K17* (together comprising the *Ac* locus), and *Li* were designed based on the sequences in our previous publications (Olsen *et al.*, 2007; Olsen & Small, 2018), and are available in **Supplementary Table S2.8**. The qPCR was conducted in the CFX96 Real-Time PCR Detection System (BIO-RAD), and the data were analyzed by $\Delta\Delta C_T$ method with efficiency calibration in CFX Maestro™ Software (BIO-RAD) according to the manufacturer's instructions. At least three independent qPCR replicates were conducted for each gene. In each independent replicate, three technical replicates were included. The C_T value was first normalized to *Ef1α* gene and

then scaled to the normalized C_T value of the single-copy genotype. To provide a reference genotype for quantification of gene copy numbers, accessions possessing a single copy of each assayed gene were identified in pedigree populations created by crossing cyanogenic and acyanogenic parents (**Supplementary Table S2.3**).

The cyanogenic glucoside content per unit weight of the fresh leaf tissue was measured by a colorimetric method. Leaf tissue (1 g) was homogenized in liquid nitrogen, and resuspended in 2 mL of Na_2HPO_4 buffer (0.067 M, pH 6.0). Then, the sample was centrifuged at 4°C for 20 minutes at 13,000 ×g twice. After each centrifugation, the supernatant was transferred to a new tube. Then, an excess amount of linamarase (100 μL bulk protein extract from cassava latex, (Haque & Bradbury, 1999)) was added to the sample (50 μL supernatant + 450 μL Na_2HPO_4 buffer) and incubated 30 minutes at 37°C to release cyanide from cyanogenic glucosides (linamarin and lotaustralin). After incubation, 600 μL NaOH solution (0.2 M) was added to terminate the reaction. Then, 500 μL of the reaction mix was used for cyanide quantification by Spectroquant® Cyanide Test kit (Millipore cat. no. 1.09701) following the standard protocol. The standard cyanide solution was prepared by the same procedure but replaced the supernatant with the commercial cyanide standard (Millipore cat. no. 119533). The colorimetric reaction had the maximum absorbance at 607 nm. We used the absorbances at 700 nm and 775 nm to estimate the background absorbance at 607 nm by simple linear regression. Finally, the absorbance of 607 nm was subtracted by the background, and then compared to the standard curve for concentration.

Environmental association

To investigate the association between the cyanogenesis gene copy number variation and environmental factors, climatic information at 1-km resolution was downloaded from WorldClim 2 (Fick & Hijmans, 2017). We extracted the environmental information for collection locations of the sampled wild clover populations as the mean value from the 1000-m-buffered sample location. Then, the gene copy number was correlated to *Mean Temperature of Coldest Quarter* (MTCQ) and visualized in R.

2.5 Results

Genome assembly statistics

We isolated DNA from a wild North American white clover accession (GFL_007) to assemble a haplotype-resolved white clover genome using a combination of PacBio HiFi, Omni-C, and linkage map technologies. The total length of the primary haplotig was 995,597,458 bp and contained 1,033 contigs (N_{50} = 17,751,673 bp). After Omni-C and linkage map scaffolding, 94.529% of the haplotig sequences were placed into 16 linkage groups, corresponding to the 16 chromosomes in white clover (N_{50} = 57,173,913 bp). The total length of the alternative haplotig was of similar size (1,002,960,897 bp) and contained 584 contigs (N_{50} = 15,995,289 bp); after scaffolding (performed independently of the primary haplotig), 95.546% of the alternative haplotig sequences were placed into 16 chromosome scale linkage groups (N_{50} = 60,046,767 bp). An independent Illumina-sequenced dataset (2 $\times$ 150 bp, 185 million read pairs; SRR27541471) of the same plant accession was used for genome size estimation and quality assessment. The estimated short-read haploid genome size was similar to the PacBio HiFi

assembly (1,025,512,829 bp), with 1.46% heterozygous sites (k-mer length = 21) (**Supplementary Fig. S2.1**). The consensus quality (QV) of the diploid genome was 51.3697, indicating >99.99% accuracy in the consensus base calls. The k-mer completeness was 98.4765%.

Approximately 60% of the genome assembly was made up of repeated elements; half of these (30% of the genome) were retroelements, which closely matches retroelement proportions in the two recently published long-read assemblies (Santangelo *et al.*, 2023; Wang *et al.*, 2023). The retroelements did not show any biased distribution between the two white clover subgenomes. Surprisingly, however, counts of both major retroelement classes (*Copia*, *Gypsy*) were more than ten times greater in white clover than the summed total inferred from published genome sequences of its two diploid progenitor species (*T. occidentale* and *T. pallescens*) (Griffiths *et al.*, 2019) (**Supplementary Fig. S2.2**). Further, our phylogenetic analyses revealed several major, recent retroelement diversification events that are unique to the white clover genome (**Supplementary Fig. S2.3, S2.4, S2.5, S2.6**). These patterns are consistent with transposable element (TE) proliferation following the allopolyploid speciation origin of white clover, a phenomenon that has been reported in other allopolyploid systems (Parisod *et al.*, 2010).

BUSCO assessments indicated that 98% of the single-copy orthologous genes of the fabales_odb10 database (5366 genes) were covered in our assembly. Of these, 82% were found to be duplicated, which is consistent with the allotetraploid origin of white clover. More

detailed genome assembly statistics are provided in **Supplementary Table S2.1**. The quality of the separation of the two haplotigs was assessed by comparing the k-mer distributions between the raw-reads, the primary haplotig and the alternative haplotig (**Fig. 2.1C & D**). The k-mer multiplicity distributions of the primary and alternative haplotigs show nearly complete overlap (indicated by purple shading in **Fig. 2.1C**), suggesting a highly accurate separation of the genome's two haplotigs. The k-mer distribution of the diploid genome indicates that most k-mers were found only once (**Fig. 2.1D**), consistent with the high heterozygosity of the genome. The much lower peaks of the 3- or 4-hit distributions compared to 1- or 2-hit distributions suggest that while the genome is of allotetraploid origin, it behaves meiotically as a diploid species.

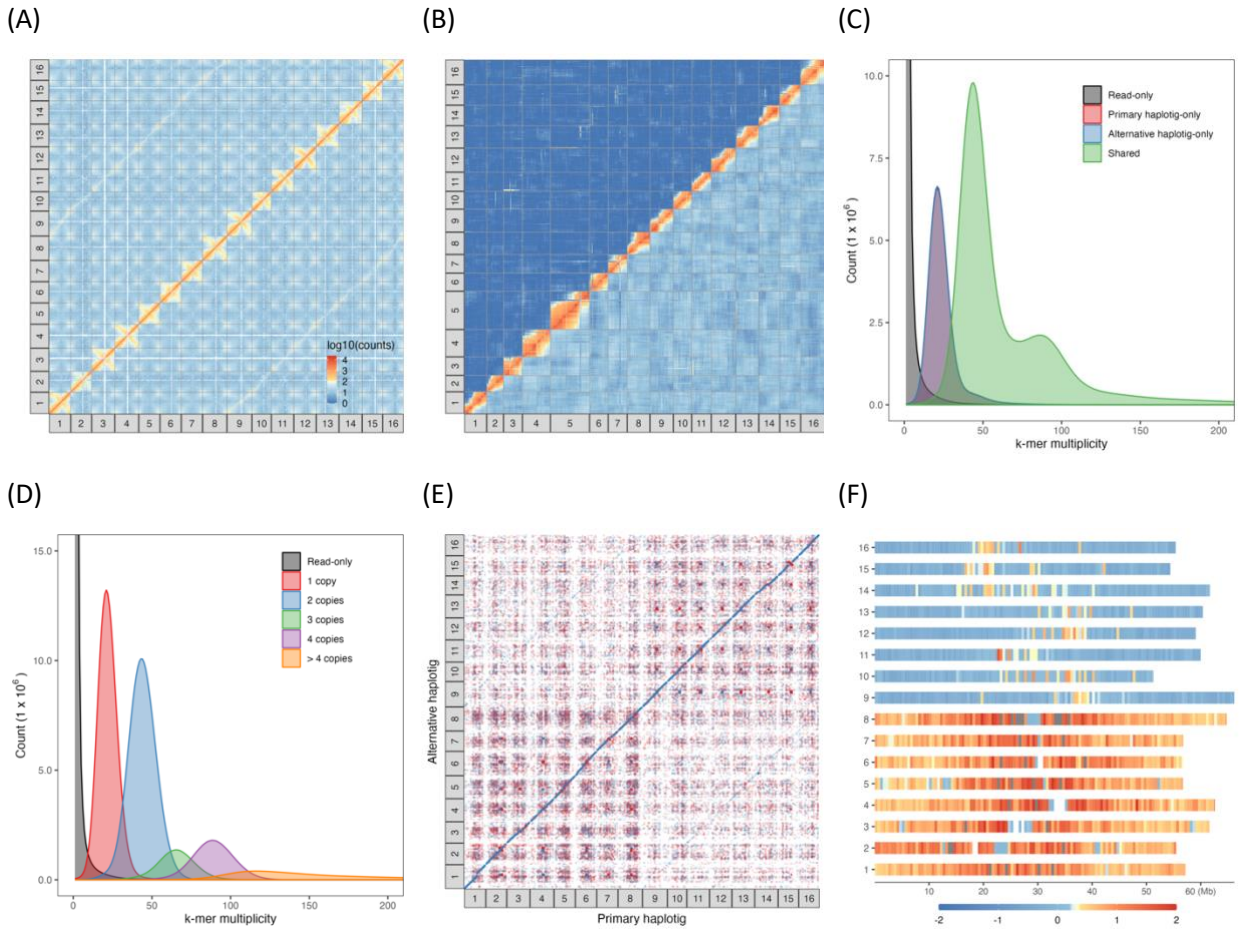


Figure 2.1 Genome quality assessment and subgenome identity characterization.

(A), Omni-C heatmap visualized by KR normalization method with window size at 1 Mbp. **(B)**, linkage map visualized by calculating the LOD (upper triangle) and the recombination fraction (RF, lower triangle) of the genetic markers of the DG F_2 mapping population at their physical location. Only the primary haplotig is presented here. **(C)**, k-mer multiplicity distribution between the primary and alternative haplotigs. Purple color indicates the overlap between the primary and alternative haplotigs, reflecting their nearly identical distributions. **(D)**, the number of the times of the k-mers are found in the diploid genome. **(E)**, whole genome nucleotide alignment of the primary and alternative haplotigs (identity >95%, query length >10,000 bp, hit length >10,000 bp). **(F)**, chromosomes categorized into *T. occidentale* (To) or *T. pallescens* (Tp) subgenomes. The gradient color is the mapping rate to either diploid progenitor, where -1 is ideally completely *T. pallescens*-like and +1 is ideally completely *T. occidentale*-like; values exceeding that range indicate mapping to multiple locations in the diploid progenitors' genomes.

Omni-C and linkage map scaffolding and subgenome characterization

The two haplotigs of the contiguous genome were first scaffolded by Omni-C linked reads, then scaffolded by our two previously published linkage maps (DG and GS F_2 mapping populations; 6,173 unique GBS SNP markers after combining the two mapping populations) (Olsen *et al.*, 2021). Where there were conflictions between the Omni-C and linkage map scaffolding results, we prioritized the results from the linkage maps (see Discussion). The read-count distribution of the chromatin contact map of the scaffolded genome supports our finding that our 16 scaffolds (pseudomolecules) are high-quality representatives of the 16 white clover chromosomes (**Fig. 2.1A**). Although we do see weak signals between the homeologous chromosomes (e.g., between chr 1 and chr 9), we believe this result can be attributed to the highly similar gene sequences (see BUSCO analyses in **Supplementary Table S2.1**) and inaccurate read-mapping of Omni-C short-reads (2×150 bp). In addition, the inter-homeologous signals are not evident in the recombination map (**Fig. 2.1B**), as most of the GBS markers used in linkage map construction are in the non-genic regions, which are more divergent between the homeologous chromosomes.

The subgenome identities of the 16 scaffolds/chromosomes were assessed using BLAST against the two diploid progenitors' genomes. Across the board, there were more hits to the *T. occidentale* genome than to the *T. pallescens* genome **(Fig. 2.1F).** This is consistent with prior observations that the *T. occidentale*-like subgenome of white clover is more closely related to extant populations of *T. occidentale* than the *T. pallescens*-like subgenome is to present-day representatives of *T. pallescens* (Hand *et al.*, 2008). When setting the mapping threshold at

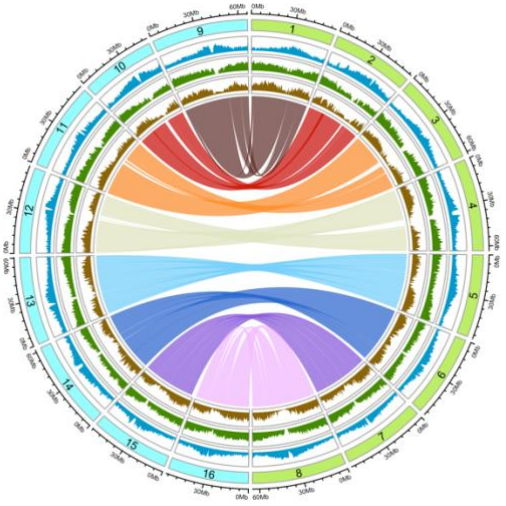
0.45 (where -1 is completely *T. pallescens*-like and +1 is completely *T. occidentale*-like; see Methods), we were able to categorize the 16 chromosomes into eight of *T. occidentale* origin (chr 1 - chr 8) and eight of *T. pallescens* origin (chr 9 - chr 16) (**Fig. 2.1F & Supplementary Fig. S2.7**). We detected no obvious inter-chromosomal translocation signal, except for at the center of the chromosomes where the signals were ambiguous; this is likely because these regions include centromeric long tandem repeats, which can interfere with the mapping process. However, we cannot validate this conclusion due to the limited resolution of the diploid progenitors' genomes (Griffiths *et al.*, 2019).

As predicted, our assessments of haplotig similarity using our whole genome nucleotide alignment revealed clear alignment hits between the primary and alternative haplotigs (indicated by the major blue diagonal line in **Fig. 2.1E**). In contrast, hits between homeologous loci are much less evident (the two weak blue diagonal lines in the top-left and bottom-right quadrants of **Fig. 2.1E**). We also found that there are generally more intra-subgenome hits than inter-subgenome hits (**Fig. 2.1E & Supplementary Table S2.2**). Together these observations suggest that inter-subgenome recombination (i.e., homeologous translocation) was very limited after the polyploidization event that created white clover, and that the integrity of the two parental species' subgenomes has been preserved, in agreement with the same conclusion inferred from the k-mer analysis above.

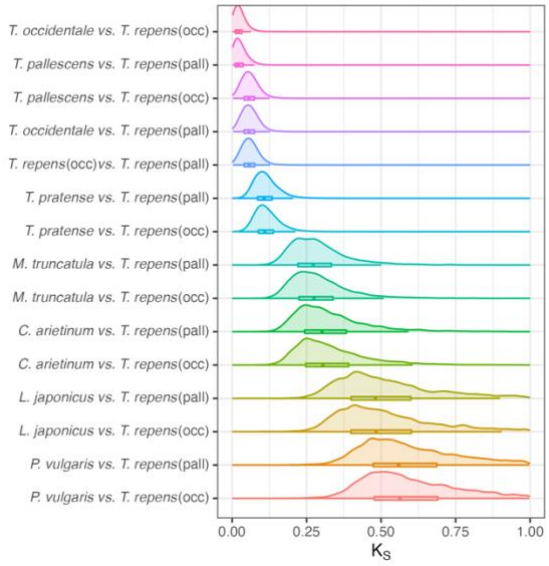
Gene synteny and synonymous mutation rate

We annotated 96,293 protein coding genes in the primary haplotig, which is slightly more than double the number of genes in the diploid genome of the related legume *Medicago truncatula* (44,295); this finding is consistent with white clover's tetraploid genome. A total of 61,352 of the annotated genes were successfully mapped by BLAST against the NCBI RefSeq database and functionally annotated with the GO database. The median lengths of the concatenated coding sequence (CDS) and intron sequence per gene were 927 bp and 1693 bp, respectively, which is comparable to closely related legume species (**Supplementary Fig. S2.8**). The gene synteny between the two subgenomes was especially high (**Fig. 2.2A**), suggesting the lack of any large-scale genome rearrangements in white clover or its closest relatives either before or after the polyploidization event that produced the species. The orthologous synonymous rate (K_S) analysis showed a neutral genetic distance to the other species, reflecting the species-level phylogenetic relationships (**Fig. 2.2B**). Interestingly, the K_A/K_S distribution for the two white clover subgenomes [*T. repens* (occ) vs. *T. repens* (pall)] was statistically indistinguishable from the K_A/K_S of the two diploid progenitor species [*T. pallescens* vs. *T. occidentale*] (**Supplementary Fig. S2.9**). This pattern suggests that, following the polyploidization event that created white clover, the homeologous gene copies within the tetraploid genome have exhibited no detectable sign of relaxed purifying selection, despite the high likelihood of functional redundancy among the many gene duplicates.

(A)



(B)


T. occidentale vs. T. repens(occ)
T. pallescens vs. T. repens(pall)
T. pallescens vs. T. repens(occ)
T. occidentale vs. T. repens(pall)
T. repens(occ)vs. T. repens(pall)
T. pratense vs. T. repens(pall)
T. pratense vs. T. repens(occ)
M. truncatula vs. T. repens(pall)
M. truncatula vs. T. repens(occ)
C. arietinum vs. T. repens(pall)
C. arietinum vs. T. repens(occ)
L. japonicus vs. T. repens(pall)
L. japonicus vs. T. repens(occ)
P. vulgaris vs. T. repens(pall)
P. vulgaris vs. T. repens(occ)
0.00
0.25
0.50
0.75
1.00
Ks

(C)

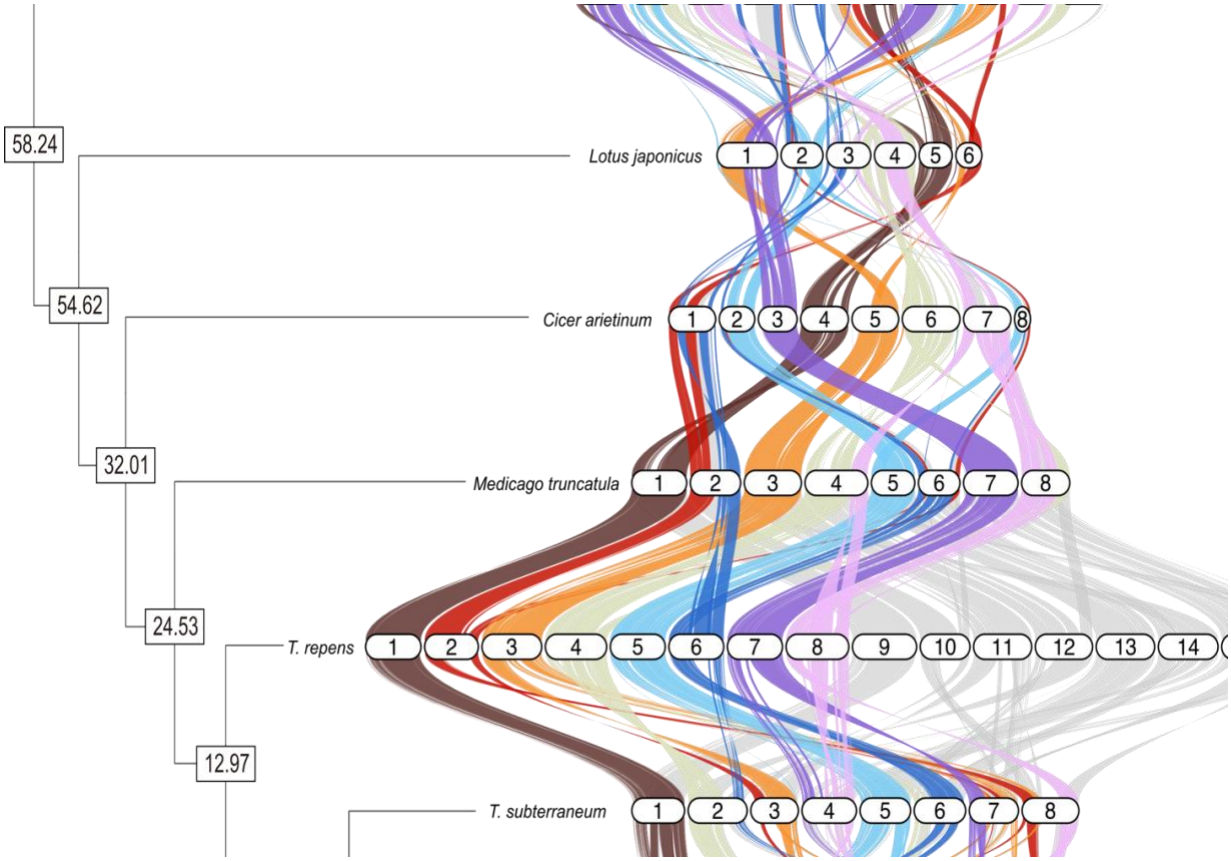

58.24
54.62
32.01
24.53
12.97
Lotus japonicus
Cicer arietinum
Medicago truncatula
T. repens
T. subterraneum

Figure 2.2 Genome assembly of white clover.

(A), circular view of genomic feature distribution along the 16 chromosomes. The blue, green, brown tracks stand for the GC content, gene density and repetitive element density, respectively. The links show the gene synteny between the two subgenomes, where chromosomes 1-8 (green) and chromosomes 9-16 (blue) belong to the *T. occidentale* (To) and *T. pallescens* (Tp) subgenomes, respectively. **(B)**, pairwise synonymous mutation rate (K_S) between the subgenomes and closely related species in the legume family (Fabaceae, subfamily Papilionoideae). **(C)**, gene synteny plot and consensus gene tree of IRLC clade within subfamily Papilionoideae. The node numbers show the crown ages in Myr, which are adapted from (Zhao *et al.*, 2021). For visualization purpose, only the synteny of the chromosome 1-8 was plotted. The plot of the chromosome 9-16 is highly similar.

Interspecific gene synteny analysis indicated that the white clover genome is highly syntenic to the closely related species *Medicago truncatula* (24.53 MYr divergence) and moderately syntenic to the more distantly related *Cicer arietinum* (32.01 MYr divergence) (**Fig. 2.2C**). There were only two chromosome-scale translocations between *M. truncatula* and white clover: between *Mt* chr 2 and *Tr* chr 6, and between *Mt* chr 4 and *Tr* chr 8 (**Fig. 2.2C & Supplementary Fig. S2.10E, F**). Given the relatively high synteny between these species in different genera, we were surprised to detect multiple genome rearrangements within the genus *Trifolium* when comparing white clover with its two congener species with reference genomes, *T. subterraneum* and *T. pratense* (12.97 MYr divergence). *Trifolium subterraneum* has the same basal chromosome number as white clover (n = 8); however, this lineage appears to have undergone at least 2 or 3 chromosome-scale translocations in chr 2, chr 3, chr 4, chr 6, chr7 and chr 8 (**Fig. 2.2C & Supplementary Fig. S2.10A, B**). *Trifolium pratense*, in comparison, appears to have lost one chromosome after divergence from the *T. subterraneum* lineage, and it shows more complex patterns of genomic rearrangements (**Fig. 2.2C & Supplementary Fig. S2.10C, D**). Based on the phylogeny and the gene synteny analyses, we therefore infer that among the three *Trifolium* species with reference genomes, white clover preserves the more ancestral genome structure, and that there have been extensive rearrangements among its diploid congeners.

Cyanogenesis genes and copy number variation

Genome sequence analyses. We hypothesized that CNVs at the *Ac* and *Li* loci contribute to the phenotypic variation found in white clover populations across environmental gradients.

This hypothesis was initially based on patterns of SNP inheritance in greenhouse pedigree populations, where Sanger sequencing of PCR amplicons revealed that *Acac* and *Lili* hemizygotes could transmit dominant alleles with SNPs that appeared heterozygous (**Supplementary Table S2.3**). To assess this hypothesis with our whole-genome data, we first tested for evidence of gene copy number variation (CNV) between the primary and alternate haplotigs in our newly-generated genome. We found that our long-read based assembly unambiguously confirmed the presence of CNVs at both loci. We aligned the sequences around the *Ac* gene cluster (comprising the three cyanogenic glucoside biosynthesis genes: *CYP79D15*, *CYP736A187* and *UGT85K17*) and the *Li* gene and performed BLAST analyses between haplotigs (**Fig. 2.3**). For *Ac*, there are two complete gene clusters in the primary haplotig and three in the alternative haplotig (**Fig. 2.3A**; **Supplementary Table S2.4**). The gene order inside the *Ac* gene cluster is consistent with previous findings (Olsen & Small, 2018) (**Supplementary Fig. S2.11**). For *Li*, there are two gene copies in both haplotigs (**Fig. 2.3A; Supplementary Table S2.4**). To determine the span of the *Ac* and *Li* CNV repeat motif, we plotted the gene synteny in the chromosomal region surrounding each locus (**Fig. 2.3C & D**). A single copy of the *Ac* motif is approximately 255-320 Kbp (**Supplementary Table S2.5**), while a single copy of the *Li* motif is approximately 24 Kbp (**Supplementary Table S2.6**). Although the CNV motifs are clearly replicated, the non-coding regions between the copies show sequence divergence (**Fig. 2.3C & D**). Multiple repetitive sequences are evident in the *Li* gene region, including other glucosidase genes that are predicted to have different substrate specificities than linamarase (Kongsaeree *et al.*, 2010) (**Fig. 2.3D & Supplementary Fig. S2.12**). The presence of these repetitive

sequences could contribute to genome instability that underlies the structural variation of this genomic region.

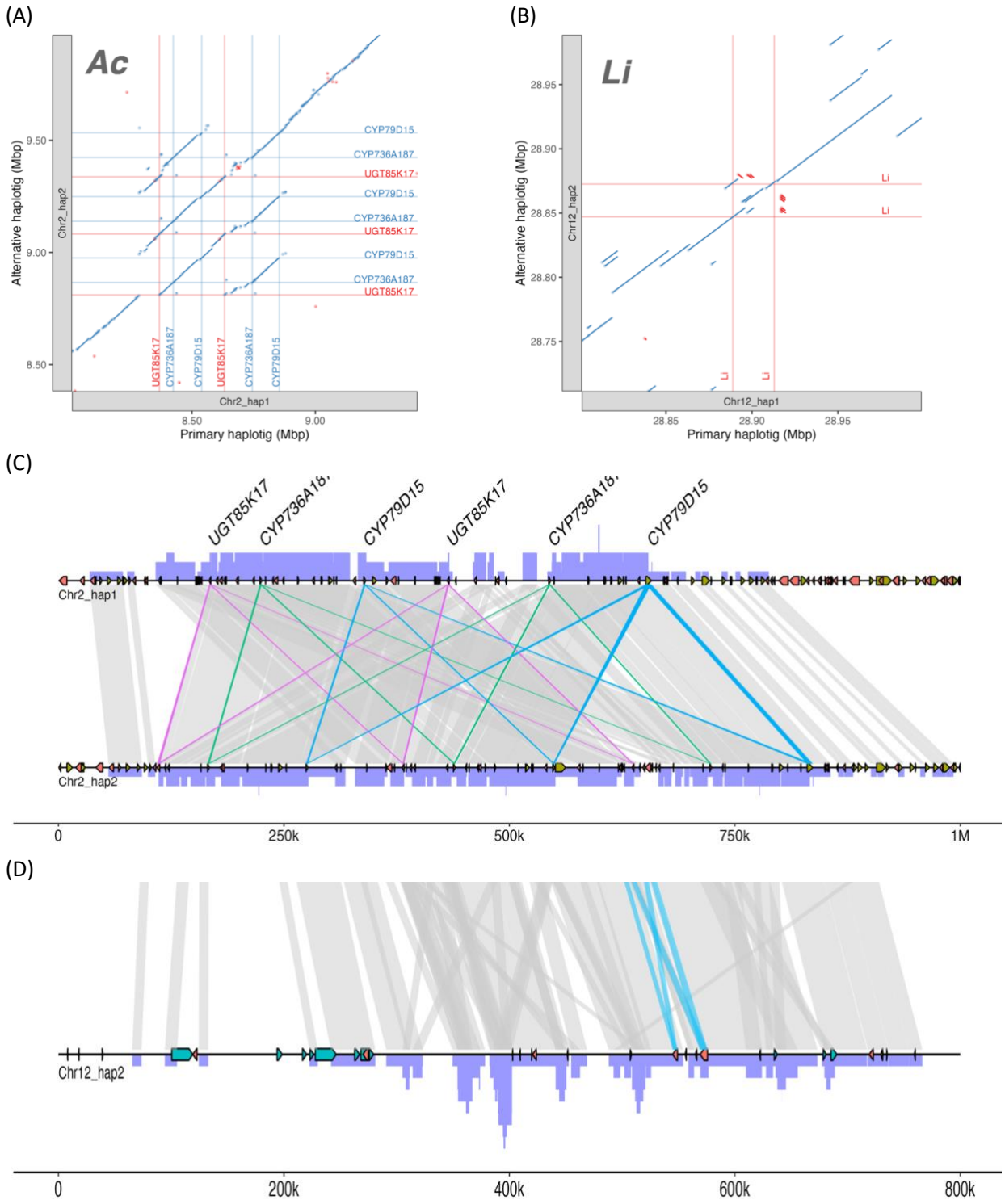

(A)
Ac
Alternative haplotig (Mbp)
Primary haplotig (Mbp)
Chr2_hap1
Chr2_hap2
CYP79D15
CYP736A187
UGT85K17
(B)
Li
Chr12_hap1
Chr12_hap2
(C)
UGT85K17
CYP736A187
CYP79D15
0
250k
500k
750k
1M
(D)
0
200k
400k
600k
800k

Figure 2.3 Nucleotide alignment and gene synteny of the white clover cyanogenesis genes. **(A-B)**, nucleotide alignment of the regions around the *Ac* and the *Li* loci. Blue, the pairwise alignment is in the same direction. Red, the pairwise alignment is in the opposite directions. **(C-D)**, gene synteny of the regions around the *Ac* (C) and the *Li* (D) loci. Top sequence, the chromosome in the primary haplotig. Bottom sequence, the corresponding homeologous chromosome in the alternative haplotig. Gray links, nucleotide alignments [same as in (A) and (B)]. Purple stacks, the number of nucleotide alignment hits in the homeologous chromosome.

In addition to comparing CNVs in our newly generated genome, we performed similar analyses with the two other published white clover genomes. In the haplotig-resolved genome of Santangelo *et al*. (Santangelo *et al.*, 2023), we found two copies of the *Ac* gene cluster in the primary haplotig and a deletion of the entire *Ac* gene cluster in the alternative haplotig (a ~500 Kbp deletion compared to the primary haplotig in the present study) (**Supplementary Fig. S2.13A & B**). For the *Li* gene, we found two copies per haplotig, similar to our new genome assembly (**Supplementary Fig. S2.13C & D**). In the genome of Wang *et al*. (Wang *et al.*, 2023), which does not provide haplotig resolution, we found three copies of the *Ac* gene cluster and two copies of the *Li* gene (**Supplementary Fig. S2.14**). In none of the three white clover whole genomes did we find a haplotig with a complete deletion of the *Li* locus. However, since *Li* deletions are known to be common in nature (Olsen *et al.*, 2007; Olsen *et al.*, 2013), we wanted to estimate of the size of the *Li* genomic deletion in plants where it occurs. To do this we mapped the short read genome sequence of an accession lacking linamarase activity (DMN_010, *lili* genotype) to the primary haplotig of our novel genome. In the *lili* accession, we found a low coverage window of ~500 Kbp (**Supplementary Fig. S2.15**); this result suggests a much larger genomic deletion than the span of the *Li* CNV motif (24 Kbp) alone.

Gene expression and phenotypic data. After demonstrating the existence of both *Ac* and *Li* CNVs in all three white clover whole genomes, we sought to test the hypothesis that CNVs at the *Ac* and *Li* loci directly affect the cyanogenesis phenotype. To do this, we focused on the *Ac* locus because synthesis of cyanogenic glucosides, mediated by the *Ac* gene cluster, is predicted to be directly correlated with the cyanogenic response (McMahon *et al.*, 2021). We performed

qPCR of genomic DNA and mRNA to estimate the correlation between CNV counts and expression in the *Ac* gene cluster (*CYP79D15*, *CYP736A187*, *UGT85K17*). First, we found that the inferred copy numbers of the three genes in the *Ac* cluster consistently occur in a 1:1:1 ratio (**Supplementary Fig. S2.16**); this is consistent with our findings above that the entire *Ac* gene cluster is duplicated as a unit. In addition, we found that cyanogenic glucoside content is significantly positively correlated with copy number for each of the three genes in the *Ac* cluster (**Fig. 2.4A-C)**. In comparison, no such correlation was observed for the *Li* gene, whose function is unrelated to cyanogenic glucoside biosynthesis (**Fig. 2.4D**). We further detected that cyanogenic glucoside content is significantly positively correlated with the expression of *CYP79D15* and *CYP736A187*, which correspond to the first two steps in the cyanogenic glucoside biosynthetic pathway (**Supplementary Fig. S2.17**). Surprisingly, however, no significant correlation was observed between the number of gene copies and levels of gene expression (**Supplementary Fig. S2.18**); this could be an artifact of inadequate time point sampling of tissue, as regulation of cyanogenic glucoside synthesis is known to vary diurnally in other cyanogenic plant species (Schmidt *et al.*, 2018).

Wild populations and environmental associations. Finally, to assess whether CNVs at the *Ac* and *Li* cyanogenesis loci contribute to environmental adaptation in wild populations of white clover, we sampled 419 wild accessions from 43 white clover populations distributed across North America, estimated CNVs at the *Ac* and *Li* loci using genomic DNA qPCR, and tested for associations with climatic variables of the collection locales using the WorldClim 2 dataset (Fick & Hijmans, 2017). CNVs for the wild population samples ranged from 0-8 per

accession for *Ac* and 0-6 for *Li* (**Supplementary Fig. S2.19; Supplementary Table S2.7**). Because previous observations of white clover cyanogenesis clines have detected strongest associations between cyanogenesis frequencies and minimum winter temperature (Kooyers & Olsen, 2012; Kooyers & Olsen, 2013), we focused on the climatic variable *Mean Temperature of Coldest Quarter* (MTCQ) for CNV association tests. Notably, both *CYP79D15* (the first gene in the *Ac* cluster) and *Li* are significantly correlated with MTCQ (**Fig. 2.4E, F & H**). Importantly, these correlations remained significant after removing accessions with zero copies of the genes (corresponding to the *acac* and *lili* recessive genotypes); this indicates that the significant correlations between CNVs and climate are not simply reflecting the previously described PAVs at the *Ac* and *Li* loci, which are well known to contribute to white clover cyanogenesis clines (Kooyers & Olsen, 2012; Kooyers & Olsen, 2013) (**Supplementary Figs. S2.20A & B, S2.21A & B**).

Because the cyanogenic phenotype arises through an epistatic interaction between the two cyanogenesis loci, we also examined the cumulative effects of CNVs at both loci by assessing MTCQ associations with the product of *CYP79D15* and *Li* gene copy numbers. As with the individual loci, this assessment revealed a significant positive correlation (**Figs. 2.4G, Supplementary Figs. S2.20C, S2.21C**). This finding further suggests that adaptive clinal variation at the white clover cyanogenesis loci arises through selection acting on the *Ac* and *Li* CNVs.

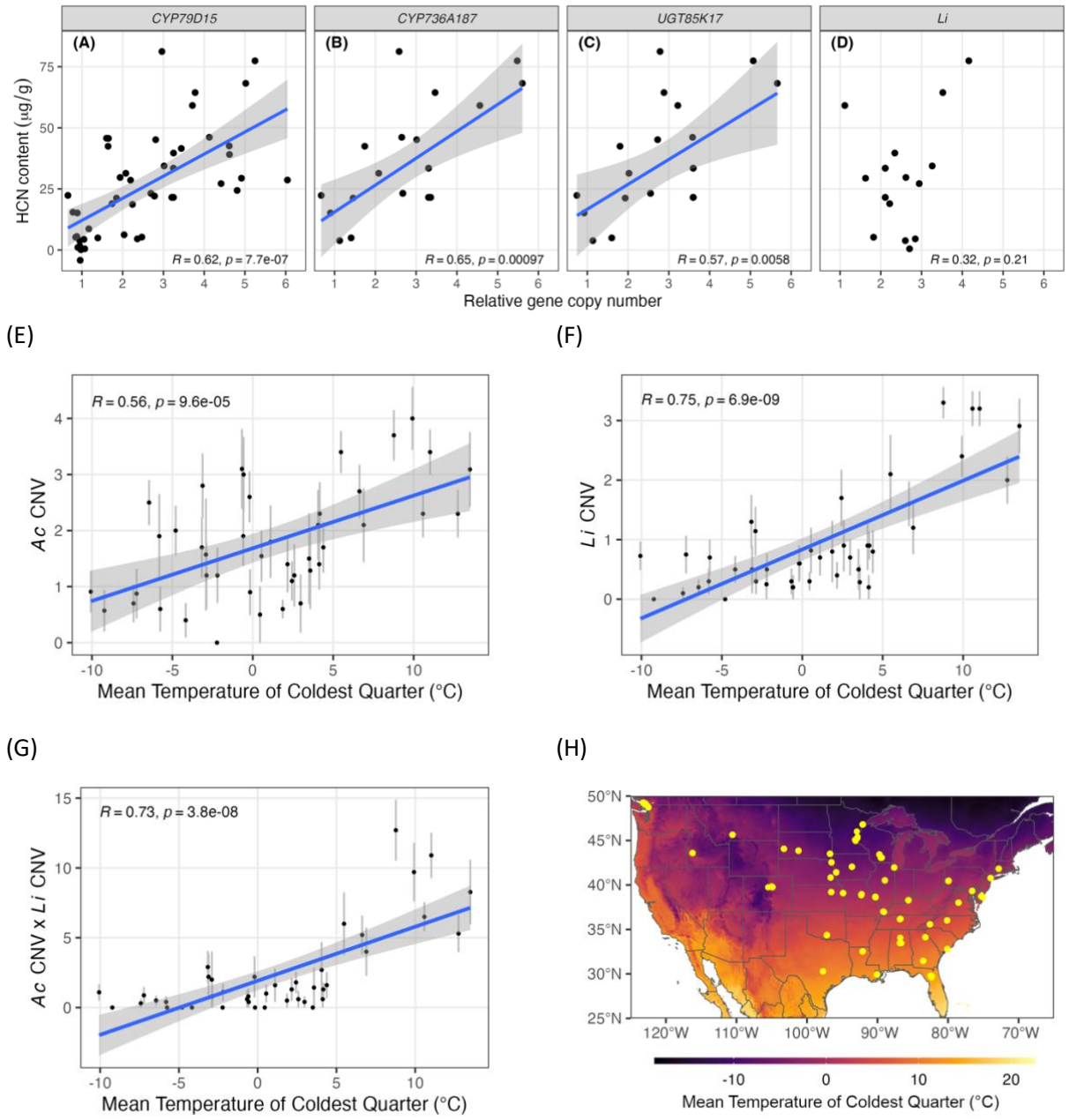


Figure 2.4 Cyanogenesis gene copy number variation and associations with cyanogenic potential and climate.

(A-D), Gene copy number variation and cyanide (HCN) content (per gram fresh leaf tissue). Linear regression line and standard error are plotted if the slope is significantly different from zero ($p < 0.05$). Each point stands for the mean value of 3-4 repeated measurements. **(E-H)**,

Association between cyanogenesis gene copy number variation (CNV) and mean temperature of coldest quarter (MTCQ). **(E)**, *Ac* CNV. **(F)**, *Li* CNV. **(G)**, the product of (*Ac* CNV)*(*Li* CNV) within an accession. **(H)**, collection locations of 419 sampled accessions (43 populations) across North America. The map is shaded by MTCQ.

2.6 Discussion

PacBio HiFi, Omni-C, and linkage map bridge the gap to allotetraploid genome assembly

The increased availability of high-quality long-read genomic data has contributed to our understanding of the importance of structural variation, including CNVs, as contributors to phenotypic variation. However, despite the fact that CNVs are prevalent in nature, very few studies to date (and to our knowledge, no previous studies in plants) have documented the distribution of CNVs in the wild and correlated these with known adaptive phenotypes along an environmental gradient. In this study, we bridge this knowledge gap by revisiting the molecular mechanism underlying the well-known cyanogenesis polymorphism in white clover using three newly-available high-quality genomes generated with long-read based sequencing technology (this study; Santangelo *et al.*, 2023; Wang *et al.*, 2023). The genome assembly presented here is the second haplotig-resolved genome of white clover and was generated by incorporating high-depth (84×) PacBio HiFi sequencing, Omni-C chromatin interaction, and linkage map information. The quality of our new genome assembly highlights the value of high-density linkage map data for complementing Omni-C data in scaffolding, especially when joining two long-range scaffolds in their correct orientation (e.g., across centromeres). The ability to accurately scaffold chromosomes is often limited when using Omni-C data alone (see, e.g., comparisons of the three white clover genomes with and without linkage map assisted scaffolding; **Supplementary Fig. S2.22**, **Fig. S2.23**). Our genome also represents the first true *de novo* genome assembly for white clover — created without first assigning sequences to a subgenome before attempting the assembly. The resulting assembly thereby provides the most

unbiased and reliable chromosome-level and haplotig-resolved genome to date for this economically important allotetraploid legume species.

Unexpectedly little genome reorganization in white clover

Polyploidization is characterized by an immediate multiplication in chromosome number through whole a genome duplication event, which can produce a major shock to the genome that may cause a series of unpredictable but discernible downstream changes (McClintock, 1984). These may include but are not restricted to bursts of transposable element (TE) proliferation (Parisod *et al.*, 2010), large scale genome rearrangements (Chen & Ni, 2006), fractionation (Cheng *et al.*, 2018) and the accumulation of deleterious mutations due to duplicated gene redundancy (Otto, 2007). In our study, we found that the number of LTR retrotransposons (LTR-RTs), the dominant type of TEs in *Trifolium*, is ten times higher in white clover than in the genomes of its two diploid progenitor species combined (**Supplementary Fig. S2.2**). In addition, the short branches of the LTR-RT phylogenetic trees further suggest a recent multiplication of LTR-RT elements within the white clover genome following the origin of this species (**Supplementary Fig. S2.3**, **Fig. S2.4**). It is possible that the high number of TEs in the white clover genome has facilitated the diploidization process by accumulating new mutations that contributed to subgenome divergence, thereby repressing recombination between homeologous chromosomes (Parisod *et al.*, 2010). Further investigation of the TE family expansions, including in the diploid progenitor species, would be useful to test this hypothesis.

The genus *Trifolium* is characterized by extensive chromosome number variation between species and occasional polyploid speciation events (Ellison *et al.*, 2006). In our genomic comparisons with the closely related legume *Medicago truncatula*, white clover shows the most conserved chromosome structure of the three *Trifolium* species with reference genomes, the other two of which are diploid; *T. pratense* (2n = 14) and *T. subterraneum* (2n = 16) each possess a minimum of 2-3 chromosomal rearrangements in a majority of their chromosomes (**Fig. 2.2C**). This finding suggests that polyploidization by itself does not necessarily lead to large-scale genome reorganizations, and conversely, that extensive genome reorganization can occur in the absence of polyploidization.

Beyond the genus *Trifolium,* inclusion of the white clover genome in legume genome sequence comparisons is also pivotal in accurately reconstructing the karyotype evolution of the IRLC legume clade. This economically important group contains the majority of agriculturally important legumes, including chickpea, lentil, pea, vetch, alfalfa, and clovers. Notably, without the white clover genome sequence comparison, it would be impossible to tell that the basal chromosomal organization in *Trifolium* is conserved with related legume genera (**Fig. 2.2C**).

Evolutionary theory predicts that species may experience a relaxation of purifying selection after polyploidization because the redundant gene copies (homeologs) are capable of masking the effect of deleterious recessive mutations (Otto, 2007). We therefore expected to detect more mutations in homeologous gene pairs than in pairwise comparisons of homologous

(orthologous) genes of the extant progenitors. However, no significant difference in the ratio of nonsynonymous versus synonymous mutation rate (K_A/K_S) was found (**Supplementary Fig. S2.9**). Moreover, we did not detect any major genome fractionation in either subgenome (**Fig. 2.2A**) (see also Griffiths *et al.*, 2019). These observations together suggest that white clover has experienced little relaxation of purifying selection after the polyploidization event that created this species. Although this finding is contrary to the classic expectations for homeologous gene copies, it is possible that selection to maintain dosage balance could be preventing any such relaxation of purifying selection (Cheng *et al.*, 2018). It is also possible that white clover's very recent evolutionary origin (estimated at 15,000-28,000 years ago) (Griffiths *et al.*, 2019) is simply too short of a time period to allow for the detectable accumulation of deleterious mutations.

Revisiting the evolution of the adaptive cyanogenesis polymorphism

Understanding the genetic and biochemical bases of the white clover cyanogenesis polymorphism, and the ecological factors that shape its evolution, has been a century-long research endeavor. Initial characterizations of the biochemical components and the polymorphism's inheritance patterns predated the molecular era (Armstrong *et al.*, 1913; Ware, 1925; Coop, 1940; Melville & Doak, 1940; Corkill, 1942). In the 1950s, a series of classic studies by Daday surveyed cyanogenesis frequencies in wild populations worldwide; this work revealed the widespread occurrence of latitudinal and elevational cyanogenesis clines throughout the native and introduced species range, with cyanogenic plants consistently present at higher frequencies in warmer climates (Daday, 1954a; Daday, 1954b; Daday, 1958). In subsequent

decades, numerous studies have examined the potential environmental factors that maintain the polymorphism and drive the evolution of climate-associated cyanogenesis clines (reviewed by Hughes, 1991; Kooyers & Olsen, 2012; Kooyers *et al.*, 2014; Kooyers *et al.*, 2018; Santangelo *et al.*, 2022). Key selective factors likely include chemical defense in areas of high herbivore abundance, fitness tradeoffs associated with the energetic costs of producing cyanogenic components, and abiotic stresses including cold and drought. Complementary studies have focused on the molecular basis of the polymorphism and its evolutionary origins. This work has revealed that the *Ac/ac* and *Li/li* polymorphisms arise through two unlinked PAVs (Olsen *et al.*, 2007; Olsen *et al.*, 2008; Olsen & Small, 2018); that the recessive (deletion) alleles have evolved repeatedly in white clover as well as in related *Trifolium* species (Olsen *et al.*, 2013; Olsen *et al.*, 2014; Olsen & Small, 2018); and that the *Ac* and *Li* loci of white clover are derived from different diploid progenitors, with *Ac* derived from *T. occidentale* and *Li* derived from the *T. pallescens* lineage (Olsen *et al.*, 2021). As the body of white clover cyanogenesis research has progressed, this system has come to be regarded as a textbook example of a balanced polymorphism maintained by geographically heterogeneous selective pressures (Briggs & Walters, 2016; Futuyma & Kirkpatrick, 2017).

By identifying the occurrence of CNVs for the dominant (gene-presence) alleles of the *Ac* and *Li* loci, and by documenting their associations with climatic variation in wild populations across the North American species range, the present study adds new depth to our understanding of the white clover cyanogenesis polymorphism and its role in adaptive evolution. Analysis of the three long-read genome assemblies (this study; Santangelo *et al.*,

2023; Wang *et al.*, 2023) revealed CNV counts of 0, 2, and 3 for the *Ac* locus and 2 copies for the *Li* locus in the haploid genomes (**Fig. 2.3**, **Supplementary Fig. S2.13**, **Fig. S2.14**). With expanded sampling to include 419 wild North American accessions, estimated CNV counts ranged from 0-8 per accession for *Ac* and 0-6 for *Li* (**Supplementary Fig. S2.19; Table S2.7**). The mechanism responsible for the producing these CNVs remains unknown; however, our genome sequence analyses suggest that it is unlikely to be due to mobile element activity. TEs are not especially abundant in the genomic regions of the *Ac* and *Li* loci, and those elements that are present do not show any detectable patterns associated with CNV counts (**Supplementary Fig. S2.24**). This may suggest that other mechanisms, such as nonallelic homologous recombination (NAHR), are responsible for the observed structural variation.

The physical chromosome map produced in the present study corroborates our previous findings, based on QTL mapping of the cyanogenesis phenotype, on the genomic locations of the *Ac* and *Li* loci (Olsen *et al.*, 2021). Interestingly, unlike most genes in the polyploid white clover genome, no functional homeologs of either *Ac* or *Li* were detected in BLASTs against all gene models (data not shown), despite the highly conserved gene synteny between the two subgenomes overall. These findings lend further support to our prior conclusions that the *Ac* and *Li* genes of white clover are uniquely derived from its two different diploid progenitors (Olsen *et al.*, 2021).

Another key finding of our study is the positive correlation between the number of CNVs at the *Ac* locus and cyanogenic glucoside production (**Fig. 2.4A-C**). To our knowledge, this is the

first reported evidence in any cyanogenic species of quantitative variation in cyanogenic glucoside production due to CNVs of the underlying biosynthetic pathway. In contrast to white clover, the cyanogenic glucoside content in cassava (Ogbonna *et al.*, 2021) and *Lotus corniculatus* (Chen *et al.*, 2023) has been reported to be regulated by cell membrane transporter genes and stress responsive genes, respectively. It thus appears that quantitative control of cyanogenic glucoside accumulation in different plant lineages has evolved independently through different molecular mechanisms. The correlation between *Ac* CNV dosage and cyanogenic glucoside production also illustrates the importance of multiplication of the entire gene cluster. We found that each *Ac* CNV unit includes all three component genes (*CYP79D15*, *CYP736A187*, *UGT85K17*) of the biosynthetic pathway (Olsen & Small, 2018) and that the gene order is conserved across CNV copies (**Fig. 2.3A, C; Supplementary Fig. S2.13A, Fig. S2.14A**). If there were multiplication of some but not all of the three genes, the imbalance and resulting accumulation of unstable chemical intermediates could potentially be maladaptive or even toxic (Kristensen *et al.*, 2005; Takos & Rook, 2012; Olsen & Small, 2018). To our knowledge, only one other study has documented the role of variable copies of an entire gene cluster (soybean nematode resistance locus *Rhg1*) in contributing to phenotypic variation (Cook *et al.*, 2012).

Finally, our results demonstrate that CNVs contribute to local adaptation and the evolution of climate-associated adaptive clines in wild populations of a plant species. As noted above, the white clover cyanogenesis polymorphism has long served as a model for understanding the forces that generate and maintain adaptive polymorphisms within species.

Our findings here that CNVs in wild white clover populations are significantly correlated with local climate (**Fig. 2.4E-H**), together with the evidence that CNVs directly affect the cyanogenic phenotype (**Fig. 2.4A-C**), provide a new level of resolution to our understanding of how natural selection interacts with the genome to shape this adaptive polymorphism. To our knowledge, these results provide the first documented case of CNV-mediated local environmental adaptation in a plant species. Equally importantly, they highlight the value of long-read sequencing technologies in accurately capturing the full spectrum of genomic variants that contribute to adaptation. The last decade has witnessed a rapid growth in studies using whole genome sequencing to assess molecular basis of adaptation in plants and many other species (Stapley *et al.*, 2010; Song *et al.*, 2023). However, reliance on short-read data and alignments to a single reference genome has meant that discoveries are heavily biased towards SNP variants, with little knowledge of the role that CNVs or other structural variants (SVs) may play in adaptation. While population-scale pangenome sequencing is still economically unfeasible for nearly all eukaryotic systems, a hybrid approach such as employed in our study (i.e., using a low-sample pangenome to discover and validate SVs, followed by qPCR to conduct population-scale investigations), could be an affordable way to uncover the role of SVs in adaptation.

2.7 Conclusions

Through analyses based on a newly-generated, *de novo* assembled haplotype-resolved genome for white clover, we are able to draw several key inferences about the genomic composition and evolution of this economically important species: 1) The allotetraploid white clover genome, while characterized by elevated TE activity compared to its diploid progenitors,

retains highly conserved genome organization and gene composition, with no evidence of relaxed purifying selection in homeologs since its origin; 2) Unlike its two diploid congeners with reference genomes (*T. pratense and T. subterraneum*), both of which have undergone extensive chromosomal reorganizations, white clover has preserved a more ancestral genome structure that is shared with other members of the IRLC legume clade; 3) The *Ac* and *Li* cyanogenesis loci are both characterized by CNVs, which contribute to quantitative variation in the cyanogenic phenotype and to local climatic adaptation in wild populations spanning the North American species range. Finally, this study highlights the value of integrating PacBio long-read sequencing, Omni-C chromatin interaction, and linkage map data in polyploid genome assembly, and the power of pangenome perspectives in identifying the functional effects of structural variation and its role in local adaptation in nature.

2.8 Acknowledgements

We thank the high school Biology teachers, plant scientists and others who participated in the crowd-sourced collection of wild white clover accessions across North America, including Miranda Colletta, Sunita Crittenden, Felix Fritschi, Briana Gross, Nic Kooyers, Lisa Limeri, Mary Lyman-Onkka, Ed McAssey, John McDonald, John McIntire, Maria Monteros, Jim Mullen, Patricio Munoz, Sandra Pelc, Anne Puzzo, Nicole Riddle, Trisha Spanbauer, Daniel Tabor, Rebecca Thomson, Kathryn Turner, Cynthia Vigueira, Kate Waselkov, Marshall Wedger, and Dan Cui Zhou. We especially thank Alexander Mahmoud (undergraduate student at Washington University in St. Louis) for processing wild accessions and maintaining them in greenhouse. We also thank Mike Dyer and other staffs of the Washington University in St. Louis greenhouse

facility for providing plant care. Finally, we thank Brock Mashburn and other members of the Olsen lab group who provided valuable comments on earlier versions of the manuscript.

2.9 Authors' contributions

W-HK designed the experiments, analyzed all data, interpreted the results, and wrote the manuscript. SJW created the mapping populations, collected and maintained the wild accessions, prepared and sequenced GBS library, and phenotyped the cyanogenesis traits. LLS maintained the plants in greenhouse, prepared GBS library, phenotyped the cyanogenesis traits, and quantified the cyanogenic glucoside content per plant. KMO conducted the greenhouse pedigree crossings, sequenced the cyanogenesis genes, conceived the project, interpreted the results, and wrote the manuscript. All authors read and approved the final manuscript.

2.10 Availability of data and materials

The *de novo* genome assembling datasets generated during the current study have been deposited to BioProject accession number PRJNA953427 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/). It includes the PacBio HiFi raw reads (SRR24107384), the RNA-seq (SRR24147954 - SRR24147971), the GBS adapter-trimmed reads of the GS F_2 mapping population (SRR24400943 - SRR24401442), the GBS adapter-trimmed reads of the DG F_2 mapping population (SRR24330286 - SRR24330790), Omni-C raw reads (SRR24133794), the assembled genome and annotation (GCA_032173615.1;

JAVQLX000000000). The shotgun resequencing datasets generated during the current study have been deposited to BioProject accession number PRJNA1064563.

2.11 Funding

W-HK was funded through the William H. Danforth Plant Science Graduate Research Fellowship in the Division of Biology and Biomedical Sciences at Washington University, and by a scholarship from the Taiwan Ministry of Education. Additional funding for the project was provided by US National Science Foundation grants IOS-1557770 to KMO, and DEB-1601641 and DGE-1143954 to SJW.

2.13 Reference

2019. Arima-HiC Mapping Pipeline: Arima Genomics, Inc.

Abberton MT, Marshall AH 2010. White clover. In: Boller B, Posselt UK, Veronesi F eds. *Fodder Crops and Amenity Grasses*. New York, NY: Springer New York, 457-476.

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**(1): 145-161 e123.

Armstrong HE, Armstrong EF, Horton E. 1913. Herbage Studies. II.-Variation in *Lotus corniculatus* and *Trifolium repens* (Cyanophoric plants). *Proceedings of the Royal Society B: Biological Sciences* **86**(587): 262-269.

Bickhart DM, Koch LM, Smith TPL, Riday H, Sullivan ML. 2022. Chromosome-scale assembly of the highly heterozygous genome of red clover (*Trifolium pratense* L.), an allogamous forage crop species. *Gigabyte* **2022**: 1-13.

Bilton TP, Schofield MR, Black MA, Chagné D, Wilcox PL, Dodds KG. 2018. Accounting for errors in low coverage high-throughput sequencing data when constructing genetic maps using biparental outcrossed populations. *Genetics* **209**(1): 65-76.

Briggs D, Walters SM. 2016. *Plant variation and evolution*. Cambridge: Cambridge University Press.

Broman KW, Wu H, Sen Ś, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**(7): 889-890.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* **3**(1).

Catchen J, Amores A, Bassham S. 2020. Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3 Genes|Genomes|Genetics* **10**(11): 4115-4128.

Chen C, Zhang K, Liu F, Wang X, Yao Y, Niu X, He Y, Hong J, Liu F, Gao Q, et al. 2023. Resequencing of global *Lotus corniculatus* accessions reveals population distribution and genetic loci, associated with cyanogenic glycosides accumulation and growth traits. *BMC Biology* **21**(1): 176.

Chen ZJ, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. *BioEssays* **28**(3): 240-252.

Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X. 2018. Gene retention, fractionation and subgenome differences in polyploid plants. *Nature Plants* **4**(5): 258-268.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**(2): 170-175.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**(18): 3674-3676.

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, et al. 2012. Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* **338**(6111): 1206-1209.

Coop I. 1940. Cyanogenesis in white clover (*Trifolium repens* L.). III. A study of linamarase, the enzyme which hydrolyses lotaustralin. *The New Zealand Journal of Science and Technology. B, General section* **22**: 71-83.

Corkill L. 1942. Cyanogenesis in white clover (*Trifolium repens* L.) V. The inheritance of cyanogenesis. *The New Zealand Journal of Science and Technology. B, General section* **23**: 178-193.

Daday H. 1954a. Gene frequencies in wild populations of *Trifolium repens* L. Distribution by latitude. *Heredity* **8**(1): 61-78.

Daday H. 1954b. Gene frequencies in wild populations of *Trifolium repens*. II. Distribution by altitude. *Heredity* **8**(3): 377-384.

Daday H. 1958. Gene frequencies in wild populations of *Trifolium repens* L. III. World distribution. *Heredity* **12**(2): 169-184.

DeBolt S. 2010. Copy number variation shapes genome diversity in arabidopsis over immediate family generational scales. *Genome Biology and Evolution* **2**: 441-453.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1): 15-21.

Ellison NW, Liston A, Steiner JJ, Williams WM, Taylor NL. 2006. Molecular phylogenetics of the clover genus (*Trifolium*—Leguminosae). *Molecular Phylogenetics and Evolution* **39**(3): 688-705.

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**(1): 238.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**(12): 4302-4315.

Futuyma DJ, Kirkpatrick M. 2017. *Evolution*. Sunderland, Massachusetts: Sinauer Associates, Inc., Publishers Sunderland, Massachusetts.

Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. 2021. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* **22**(1).

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLOS Computational Biology* **15**(8): e1007273.

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019. Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31**(7): 1466-1487.

Hand ML, Ponting RC, Drayton MC, Lawless KA, Cogan NO, Charles Brummer E, Sawbridge TI, Spangenberg GC, Smith KF, Forster JW. 2008. Identification of homologous, homoeologous and paralogous sequence variants in an outbreeding allopolyploid

species based on comparison with progenitor taxa. *Molecular Genetics and Genomics* **280**(4): 293-304.

Haque MR, Bradbury JH. 1999. Preparation of linamarase solution from cassava latex for use in the cassava cyanide kit. *Food Chemistry* **67**(3): 305-309.

Hughes MA. 1991. The cyanogenic polymorphism in *Trifolium repens* L. (white clover). *Heredity* **66**(1): 105-115.

Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, et al. 2019. A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* **364**(6443): 886-889.

Ishikawa A, Stuart YE, Bolnick DI, Kitano J. 2021. Copy number variation of a fatty acid desaturase gene *Fads2* associated with ecological divergence in freshwater stickleback populations. *Biology Letters* **17**(8): 20210204.

Jin J-J, Yu W-B, Yang J-B, Song Y, Depamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biology* **21**(1).

Kjærgaard T. 2003. A plant that changed the world: The rise and fall of clover 1000-2000. *Landscape Research* **28**(1): 41-49.

Kongsaeree PT, Ratananikom K, Choengpanya K, Tongtubtim N, Sujiwattanarat P, Porncharoennop C, Onpium A, Svasti J. 2010. Substrate specificity in hydrolysis and transglucosylation by family 1 β-glucosidases from cassava and Thai rosewood. *Journal of Molecular Catalysis B: Enzymatic* **67**(3): 257-265.

Kooyers NJ, Gage LR, Al-Lozi A, Olsen KM. 2014. Aridity shapes cyanogenesis cline evolution in white clover (*Trifolium repens* L.). *Molecular Ecology* **23**(5): 1053-1070.

Kooyers NJ, Hartman Bakken B, Ungerer MC, Olsen KM. 2018. Freeze-induced cyanide toxicity does not maintain the cyanogenesis polymorphism in white clover (*Trifolium repens*). *American Journal of Botany* **105**(7): 1224-1231.

Kooyers NJ, Olsen KM. 2012. Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.). *Molecular Ecology* **21**(10): 2455-2468.

Kooyers NJ, Olsen KM. 2013. Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover (*Trifolium repens* L.). *Heredity* **111**(6): 495-504.

Kooyers NJ, Olsen KM. 2014. Adaptive cyanogenesis clines evolve recurrently through geographical sorting of existing gene deletions. *Journal of Evolutionary Biology* **27**(11): 2554-2558.

Kreiner JM, Hnatovska S, Stinchcombe JR, Wright SI. 2023. Quantifying the role of genome size and repeat content in adaptive variation and the architecture of flowering time in *Amaranthus tuberculatus*. *PLOS Genetics* **19**(12): e1010865.

Kristensen C, Morant M, Olsen CE, Ekstrom CT, Galbraith DW, Moller BL, Bak S. 2005. Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. *Proceedings of the National Academy of Sciences of the United States of America* **102**(5): 1779-1784.

Li H, Jiang F, Wu P, Wang K, Cao Y. 2020. A high-quality genome sequence of model legume *Lotus japonicus* (MG-20) provides insights into the evolution of root nodule symbiosis. *Genes* **11**(5): 483.

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology* **14**(1): e1005944.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**(6): 764-770.

McClintock B. 1984. The Significance of Responses of the Genome to Challenge. *Science* **226**(4676): 792-801.

McMahon J, Sayre R, Zidenga T. 2021. Cyanogenesis in cassava and its molecular manipulation for crop improvement. *Journal of Experimental Botany* **73**(7): 1853-1867.

Melville J, Doak B. 1940. Cyanogenesis in white clover (*Trifolium repens* L.). II. isolation of the gluco-sidal constituents. *The New Zealand Journal of Science and Technology. B, General section* **22**: 67-71.

Ogbonna AC, Braatz de Andrade LR, Rabbi IY, Mueller LA, Jorge de Oliveira E, Bauchet GJ. 2021. Large-scale genome-wide association study, using historical data, identifies conserved genetic architecture of cyanogenic glucoside content in cassava (*Manihot esculenta* Crantz) root. *The Plant Journal* **105**(3): 754-770.

Olsen KM, Goad DM, Wright SJ, Dutta ML, Myers SR, Small LL, Li LF. 2021. Dual-species origin of an adaptive chemical defense polymorphism. *New Phytologist* **232**(3): 1477-1487.

Olsen KM, Hsu SC, Small LL. 2008. Evidence on the molecular basis of the *Ac*/*ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* **179**(1): 517-526.

Olsen KM, Kooyers NJ, Small LL. 2013. Recurrent gene deletions and the evolution of adaptive cyanogenesis polymorphisms in white clover (*Trifolium repens* L.). *Molecular Ecology* **22**(3): 724-738.

Olsen KM, Kooyers NJ, Small LL. 2014. Adaptive gains through repeated gene loss: parallel evolution of cyanogenesis polymorphisms in the genus *Trifolium* (Fabaceae). *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**(1648).

Olsen KM, Small LL. 2018. Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytologist* **219**(2): 757-766.

Olsen KM, Sutherland BL, Small LL. 2007. Molecular evolution of the *Li*/*li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Molecular Ecology* **16**(19): 4180-4193.

Otto SP. 2007. The Evolutionary Consequences of Polyploidy. *Cell* **131**(3): 452-462.

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C, Ainouche M, Chalhoub B, Grandbastien M-A. 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytologist* **186**(1): 37-45.

Pecrix Y, Staton SE, Sallet E, Lelandais-Brière C, Moreau S, Carrère S, Blein T, Jardinaud M-F, Latrasse D, Zouine M, et al. 2018. Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nature Plants* **4**(12): 1017-1025.

Poplin R, Ruano-Rubio V, Depristo MA, Fennell TJ, Carneiro MO, Van Der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*.

Pos O, Radvanszky J, Buglyo G, Pos Z, Rusnakova D, Nagy B, Szemes T. 2021. DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal* **44**(5): 548-559.

Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, He Q, Ou S, Zhang H, Li X. 2021. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**(13): 3542-3558. e3516.

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution* **35**(10): 2582-2584.

Ravagnani A, Abberton MT, Skøt L. 2012. Development of genomic resources in the species of *Trifolium* L. and its application in forage legume breeding. *Agronomy* **2**(2): 116-131.

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**(1).

Rice E agptools: Tools for working with agp files.

Santangelo JS, Battlay P, Hendrickson BT, Kuo WH, Olsen KM, Kooyers NJ, Johnson MTJ, Hodgins KA, Ness RW. 2023. Haplotype-resolved, chromosome-level assembly of white clover (*Trifolium repens* L., Fabaceae). *Genome Biology and Evolution* **15**(8): 2023.2006.2006.543960.

Santangelo JS, Ness RW, Cohan B, Fitzpatrick CR, Innes SG, Koch S, Miles LS, Munim S, Peres-Neto PR, Prashad C, et al. 2022. Global urban environmental change drives adaptation in white clover. *Science* **375**(6586): 1275-1281.

Schmidt FB, Cho SK, Olsen CE, Yang SW, Møller BL, Jørgensen K. 2018. Diurnal regulation of cyanogenic glucoside biosynthesis and endogenous turnover in cassava. *Plant Direct* **2**(2): e00038.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, et al. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics* **46**(7): 707-713.

Shirasawa K, Moraga R, Ghelfi A, Hirakawa H, Nagasaki H, Ghamkhar K, Barrett BA, Griffiths AG, Isobe SN. 2023. An improved reference genome for *Trifolium subterraneum* L. provides insight into molecular diversity and intra-specific phylogeny. *Frontiers in Plant Science* **14**: 1103857.

Song B, Ning W, Wei D, Jiang M, Zhu K, Wang X, Edwards D, Odeny DA, Cheng S. 2023. Plant genome resequencing and population genomics: Current status and future prospects. *Molecular Plant* **16**(8): 1252-1268.

Stalder L, Oggenfuss U, Mohd-Assaad N, Croll D. 2023. The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Molecular Ecology* **32**(10): 2443-2460.

Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**(12): 705-712.

Takos AM, Rook F. 2012. Why biosynthetic genes for chemical defense compounds cluster. *Trends in Plant Science* **17**(7): 383-388.

Taylor NL. 2008. A century of clover breeding developments in the united states. *Crop Science* **48**(1): 1-13.

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'An B, et al. 2013. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology* **31**(3): 240-246.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**(14): 2202-2204.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. Kaks_calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics* **8**(1): 77-80.

Wang H, Wu Y, He Y, Li G, Ma L, Li S, Huang J, Yang G. 2023. High-quality chromosome-level de novo assembly of the *Trifolium repens*. *BMC Genomics* **24**(1): 326.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**(7): e49-e49.

Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J. 2015. Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nature genetics* **47**(8): 944-948.

Ware WM. 1925. Experiments and observations on forms and strains of *Trifolium repens* L. *The Journal of Agricultural Science* **15**(1): 47-67.

Wright D, Boije H, Meadows JR, Bed'hom B, Gourichon D, Vieaud A, Tixier-Boichard M, Rubin CJ, Imsland F, Hallbook F, et al. 2009. Copy number variation in intron 1 of *SOX5* causes the Pea-comb phenotype in chickens. *PLOS Genetics* **5**(6): e1000512.

Wright SJ, Goad DM, Gross BL, Muñoz PR, Olsen KM. 2022. Genetic trade-offs underlie divergent life history strategies for local adaptation in white clover. *Molecular Ecology* **31**(14): 3742-3760.

Zeven AC. 1991. Four hundred years of cultivation of Dutch white clover landraces. *Euphytica* **54**(1): 93-99.

Zhao Y, Zhang R, Jiang K-W, Qi J, Hu Y, Guo J, Zhu R, Zhang T, Egan AN, Yi T-S, et al. 2021. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Molecular Plant* **14**(5): 748-773.

Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, et al. 2019. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics* **51**(5): 865-876.

Zmienko A, Marszalek-Zenczak M, Wojciechowski P, Samelak-Czajka A, Luczak M, Kozlowski P, Karlowski WM, Figlerowicz M. 2020. AthCNV: A map of DNA copy number variations in the Arabidopsis genome. *Plant Cell* **32**(6): 1797-1819.

2.14 Supporting Information

Chapter 1

Chapter 2

Supplementary Table S2.1 Genome assembly summary

	Primary*	Alternative*
HiFi Read coverage	Homozygous regions: 84, Heterozygous regions: 43	
Number of contigs	1,033	584
Contig N50 (bp)	17,751,673	15,995,289
Longest contigs	45,963,080	53,361,982
Number of scaffolds	970	497
Scaffold N50 (bp)	57,173,913	60,046,767
Size of final assembly (bp)	995,597,458	1,002,960,897
Base pair QV	50.004	53.3544
Base pair QV (both)	51.3697	
k-mer completeness	79.8197%	80.0837%
k-mer completeness (both)	98.4765%	
BUSCO eukaryota_odb10 (255 genes)	C: 98.4% [S: 7.8%, D: 90.6%], F: 0.8%, M: 0.8%	C: 98.1% [S: 7.1%, D: 91.0%], F: 0.8%, M: 1.1%
BUSCO embryophyta_odb10 (1614 genes)	C: 99.1% [S: 11.0%, D: 88.1%], F: 0.2%, M: 0.7%	C: 99.3% [S: 12.3%, D: 87.0%], F: 0.1%, M: 0.6%
BUSCO fabales_odb10 (5366 genes)	C: 97.6% [S: 15.6%, D: 82.0%], F: 0.2%, M: 2.2%	C: 97.7% [S: 16.2%, D: 81.5%], F: 0.2%, M:2.1%
Assigned chromosomes	94.529%	95.546%
# N's	26800	27500
# N's per 100 kbp	2.69	2.74
GC level	33.91 %	33.98 %
Repeated elements	60.51 %	60.18 %
Retroelements	29.07 %	30.40 %
LTR elements (Ty1/Copia)	13.32 %	13.58 %
LTR elements (Gypsy/DIRS1)	7.69 %	9.73 %
DNA transposons	1.11 %	1.38 %

*Representing the two haploid genomes of a diploid organism

Supplementary Table S2.2. Total alignment length (identity >95%, query length >10,000 bp, hit length >10,000 bp)

Alternative haplotig	Primary haplotig: 1-8 (to)	Primary haplotig: 9-16 (tp)
25-32 (tp)	3,811,295,141	6,158,361,485
17-24 (to)	5,817,545,634	3,764,652,819

Supplementary Table S2.3 SNPs detected by direct Sanger sequencing of PCR products of *Ac* and *Li* genes in parents (bold) and F1 progeny (two-letter codes) in greenhouse crossing experiments

See separate Excel file.

Supplementary Table S2.4 Genomic locations of cyanogenesis genes

Gene	Primary (this study)	Alternative (this study)	Santangelo et. al., 2023	Wang et. al., 2023
CYP79D15	drTriRepe4Chr2g112600.1 drTriRepe4Chr2g115700.1	drTriRepe4Chr18g122900.1 drTriRepe4Chr18g125500.1 drTriRepe4Chr18g128300.1	TrR_v6_g12417.t1 TrR_v6_g12456.t1	Chr10.g56734.m1 Chr10.g56705.m1
CYP736A187	drTriRepe4Chr2g111200.1 drTriRepe4Chr2g114400.1	drTriRepe4Chr18g121900.1 drTriRepe4Chr18g124100.1 drTriRepe4Chr18g127300.1	TrR_v6_g12398.t1 TrR_v6_g12430.t1	Chr10.g56752.m1 Chr10.g56786.m1 Chr10.g56724.m1
UGT85K17	drTriRepe4Chr2g110400.1 drTriRepe4Chr2g113800.1	drTriRepe4Chr18g121500.1 drTriRepe4Chr18g123600.1 drTriRepe4Chr18g126400.1	TrR_v6_g12398.t1 TrR_v6_g12430.t1	Chr10.g56793.m1 Chr10.g56732.m1 Chr10.g56760.m1
Li	drTriRepe4Chr12g338400.1 drTriRepe4Chr12g338600.1	drTriRepe4Chr28g340600.1 drTriRepe4Chr28g340900.1	TrR_v6_g38299.t1 TrR_v6_g38301.t1	Chr07.g39283.m2 Chr07.g39287.m1

Supplementary Table S2.5. Genomic locations of Ac gene cluster

Gene	chromosome	Transcript	Start	End	Strand
UGT85K17	drTriRepe4Chr2	drTriRepe4Chr2g110400.1	8,367,129	8,369,048	-
CYP736A187	drTriRepe4Chr2	drTriRepe4Chr2g111200.1	8,423,330	8,425,360	+
CYP79D15	drTriRepe4Chr2	drTriRepe4Chr2g112600.1	8,538,499	8,540,592	+
UGT85K17	drTriRepe4Chr2	drTriRepe4Chr2g113800.1	8,631,125	8,633,208	-
CYP736A187	drTriRepe4Chr2	drTriRepe4Chr2g114400.1	8,743,594	8,745,624	+
CYP79D15	drTriRepe4Chr2	drTriRepe4Chr2g115700.1	8,851,286	8,857,454	+
UGT85K17	drTriRepe4Chr18	drTriRepe4Chr18g121500.1	8,809,396	8,812,326	-
CYP736A187	drTriRepe4Chr18	drTriRepe4Chr18g121900.1	8,865,280	8,867,310	+
CYP79D15	drTriRepe4Chr18	drTriRepe4Chr18g122900.1	8,974,416	8,976,497	+
UGT85K17	drTriRepe4Chr18	drTriRepe4Chr18g123600.1	9,081,205	9,083,126	-
CYP736A187	drTriRepe4Chr18	drTriRepe4Chr18g124100.1	9,137,923	9,139,953	+
CYP79D15	drTriRepe4Chr18	drTriRepe4Chr18g125500.1	9,248,143	9,250,216	+
UGT85K17	drTriRepe4Chr18	drTriRepe4Chr18g126400.1	9,336,113	9,338,193	-
CYP736A187	drTriRepe4Chr18	drTriRepe4Chr18g127300.1	9,422,195	9,424,222	+
CYP79D15	drTriRepe4Chr18	drTriRepe4Chr18g128300.1	9,530,323	9,536,398	+

Supplementary Table S2.6 Genomic locations of Li genes

Gene	chromosome	Transcript	Start	End	Strand
Li	drTriRepe4Chr12	drTriRepe4Chr12g338400.1	28,886,861	28,890,925	-
Li	drTriRepe4Chr12	drTriRepe4Chr12g338600.1	28,910,546	28,915,335	-
Li	drTriRepe4Chr28	drTriRepe4Chr28g340600.1	28,844,916	28,849,166	-
Li	drTriRepe4Chr28	drTriRepe4Chr28g340900.1	28,869,253	28,875,856	-

Supplementary Table S2.7 Accession information for characterizing cyanogenesis variation

See separate Excel file.

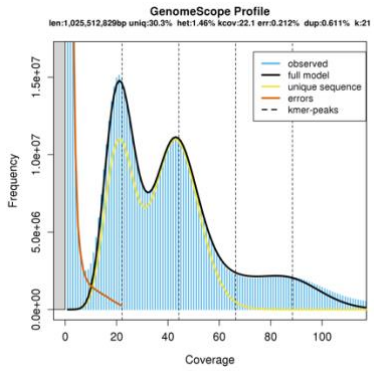
Supplementary Table S2.8 Primer list

Name	Sequence	Amplicon length	Note
CYP79D15_q03Fa (exon2)	CAA TGT GGT TGG CAA AGA GAG	72	For CNV and gene expression
CYP79D15_q03R (exon2)	GCG CAA GCT CTA ACA AAT GG		
CYP736_q05F (exon1)	GCT TGT TCA TGA AGT CCT GC	61+39	For CNV and gene expression
CYP736_q05R (exon2)	AAA TTC TTT GTG CGC TTT CTT G		
CYP736_q06F (exon1)	ATG AAG TCC TGC ACC TCA TG	53+39	For CNV and gene expression
CYP736_q05R (exon2)	AAA TTC TTT GTG CGC TTT CTT G		
CYP736_q02F (intron)	ACG TAC GCA ATA AAC TTC TTT CC	135	For CNV
CYP736_q01R (intron)	TGT CCT TGG TAT AAG GTC TTT CAC A		
UGT85_q02F (exon1)	GTG GCT TCG TCG GAT ATT TG	63+37	For CNV and gene expression
UGT85_q02R (exon2)	GTT GTA TCC AAG GTG CCA TC		
UGT85_q03F (exon1)	TGC AAT TCG ATG AAC TTG CC	45+37	For CNV and gene expression
UGT85_q02R (exon2)	GTT GTA TCC AAG GTG CCA TC		
UGT85_q01F (intron)	TTG CTC AAT CTA TTT CGG ATG TC	110	For CNV
UGT85_q01R (exon2)	GTC GCA TAT TTT TCA AGC CAG AG		
Li_q04Fa (intron)	GCA AGT TCC ATT CAT TTT ATT TTT AGT T	183	For CNV
Li_q04Rb (exon3)	CTT GTA CCG ATG ATA TTC GTC AAT		
Ef1-alpha_01F	TCG AGA AGG AAG CTG CTG AAA		Abeynayake *et al.* (2011)
Ef1-alpha _01R	CCC AGG CAT ACT TGA ATG ACC T		

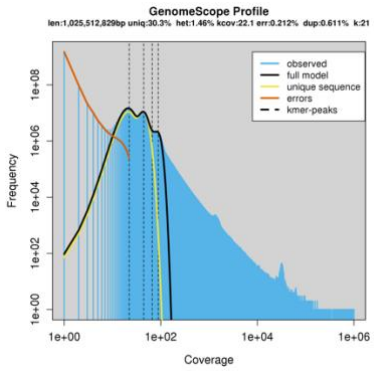
Chapter 1

Chapter 2

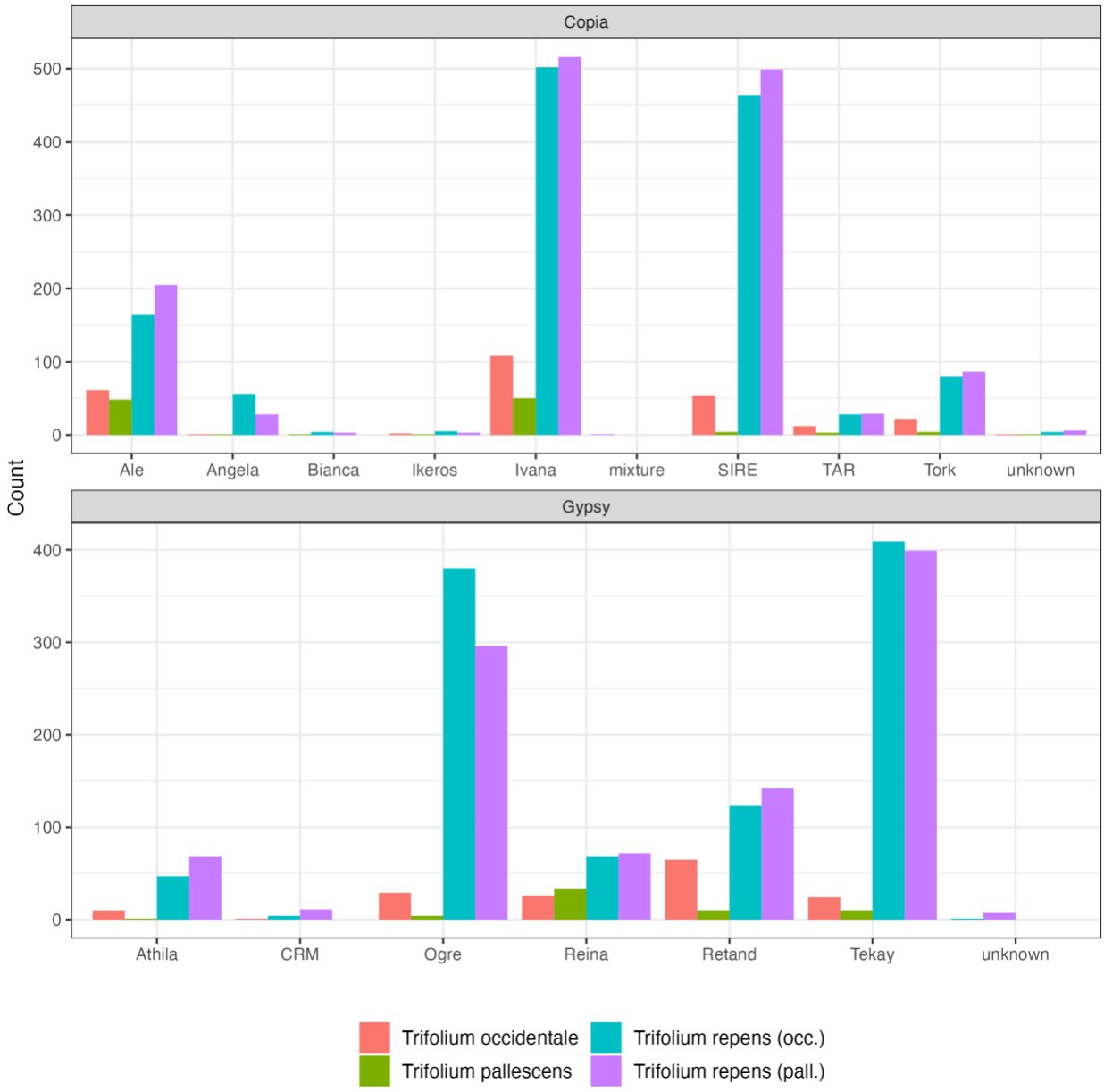
(A)



(B)

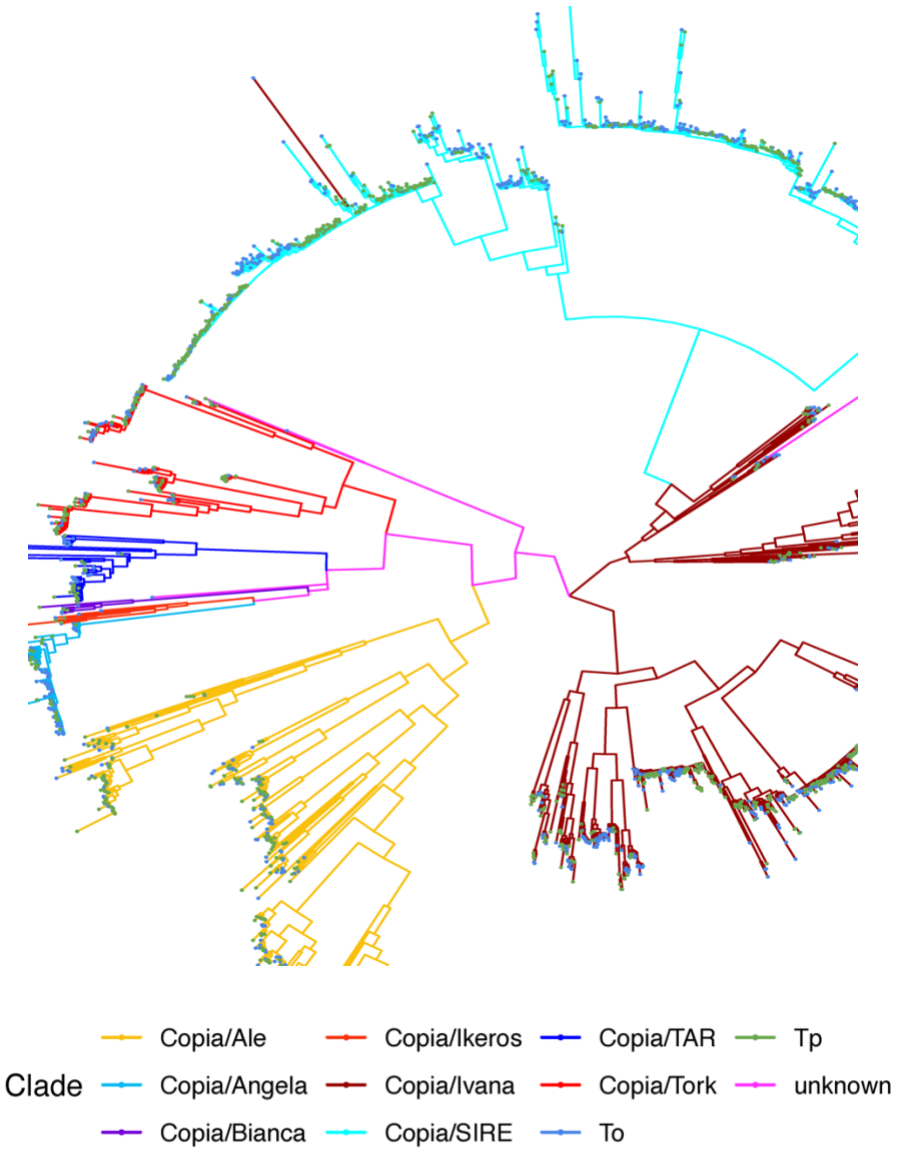


Supplementary Figure S 2.1 GenomeScope genome size and heterozygosity assessment (A-B).



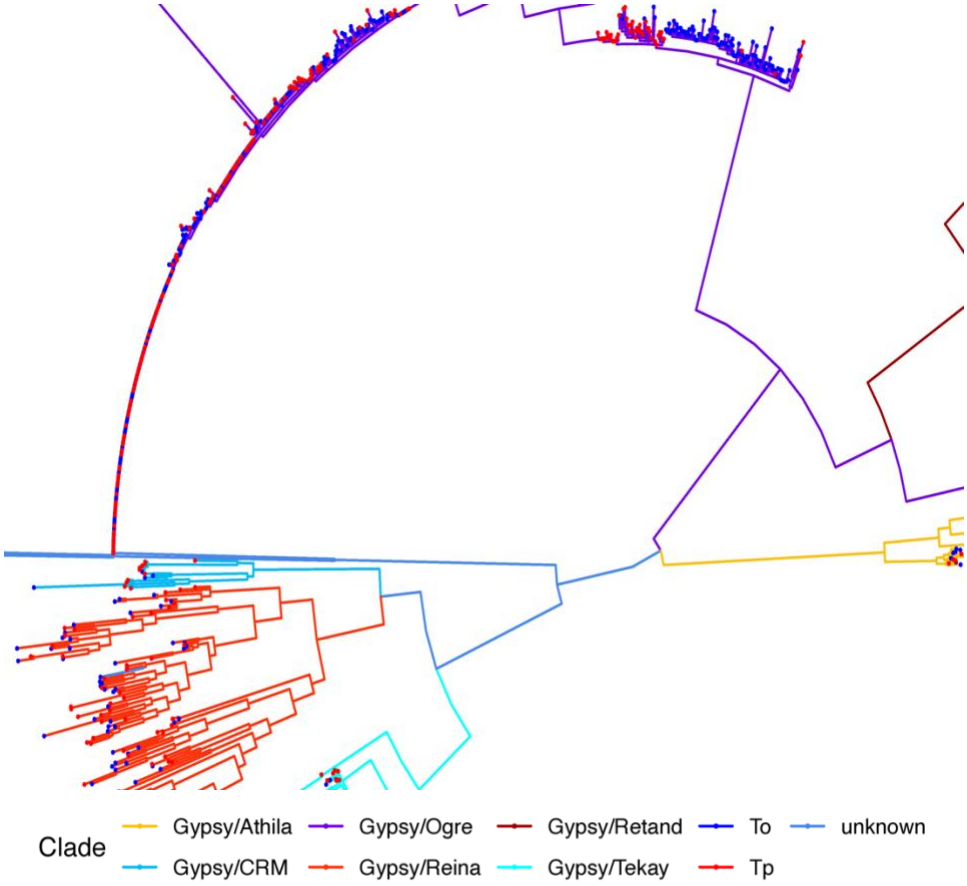
Supplementary Figure S2.2 Categorization of retroelements (LTR-RT).

The *Trifolium repens* genome (occ. & pall.) is from the haplotig1 of this study. The two extent diploid progenitors' genomes are from Griffiths *et al.* (2019).



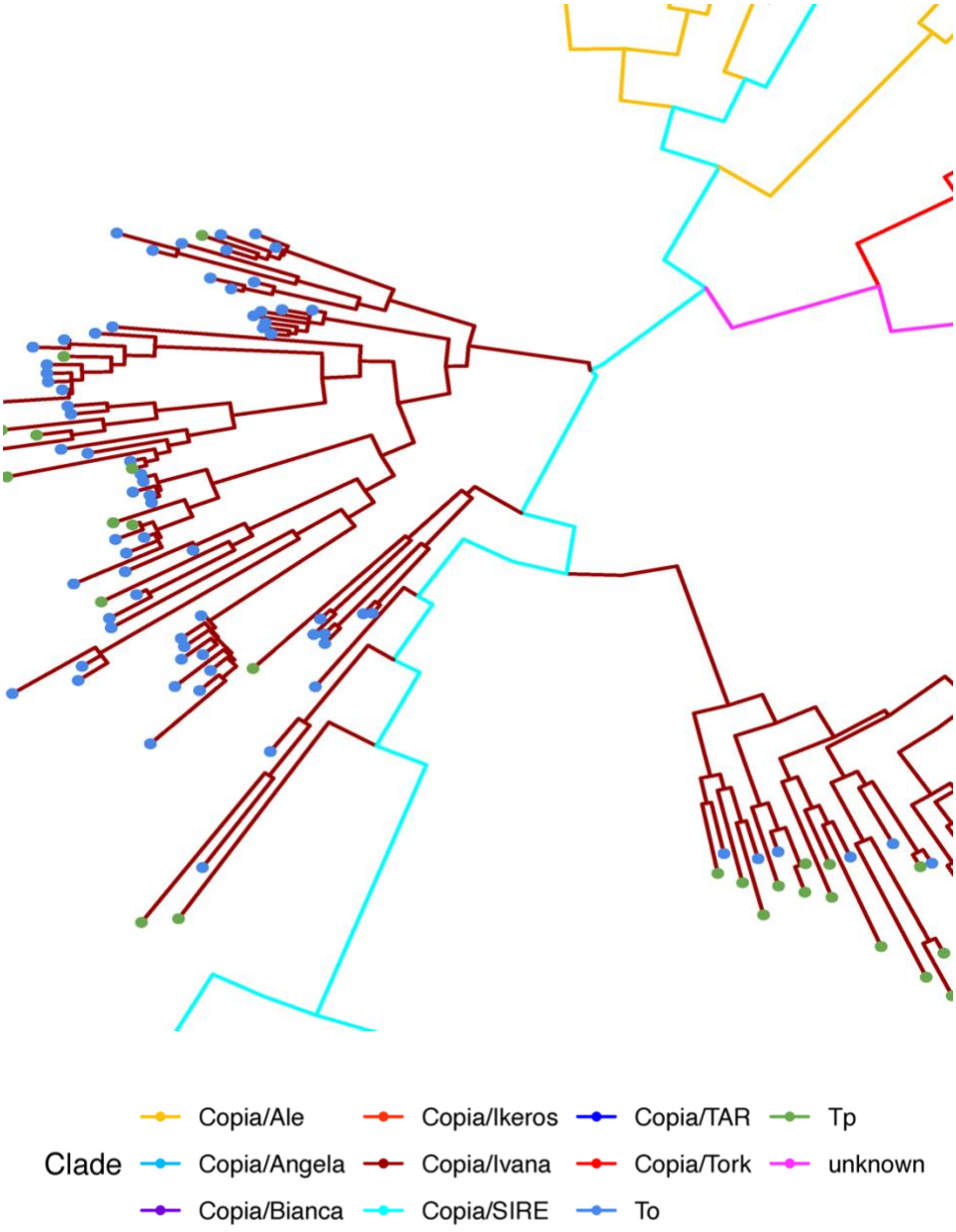
Supplementary Figure S2.3 Phylogeny of the *Copia* super-family LTR-RT in the white clover genome.

Phylogeny of the reverse transcriptase (RT) amino acid sequences of the *Copia* super-family in the white clover genome (hap1) by maximum likelihood method. The tree branches are colored by clades. The tree tips are color by the locations of the LTR-RT in Trifolium repens (occ., To in the legend) or Trifolium repens (pall., Tp in the legend) subgenome.



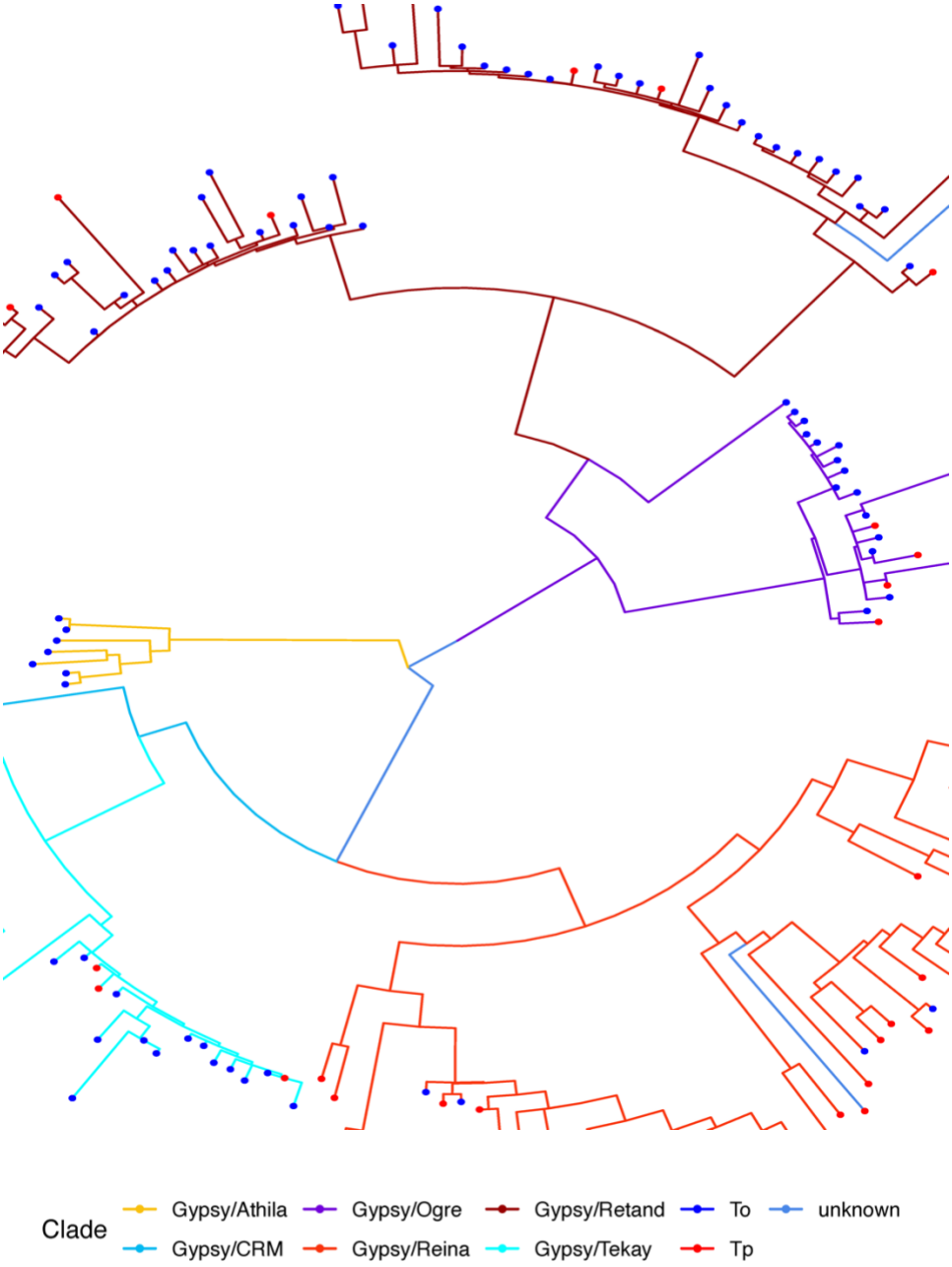
Supplementary Figure S2.4. Phylogeny of the *Gypsy* super-family LTR-RT in the white clover genome.

Phylogeny of the reverse transcriptase (RT) amino acid sequences of the *Gypsy* super-family in the white clover genome (hap1) by maximum likelihood method. The tree branches are colored by clades. The tree tips are color by the locations of the LTR-RT in *Trifolium repens* (occ., To in the legend) or *Trifolium repens* (pall., Tp in the legend) subgenome.



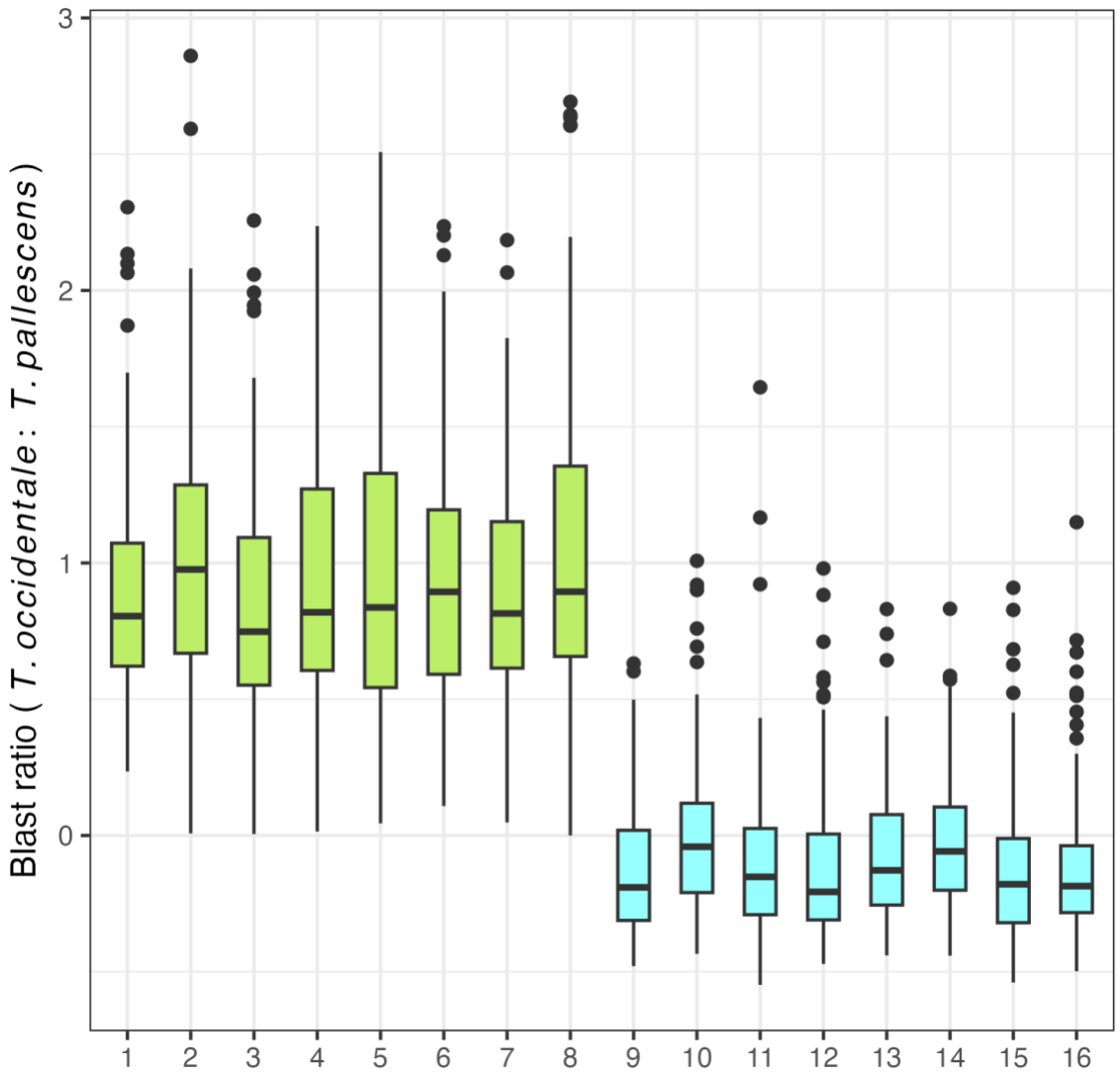
Supplementary Figure S2.5 Phylogeny of the *Copia* super-family LTR-RT in *T. occidentale* (To) and *T. pallescens* (Tp) genomes.

Phylogeny of the reverse transcriptase (RT) amino acid sequences of the *Copia* super-family in *T. occidentale* (To) and *T. pallescens* (Tp) genomes by maximum likelihood method. The tree branches are colored by clades. The tree tips are color by the locations of the LTR-RT in *T. occidentale* (To) or *T. pallescens* (Tp) species.



Supplementary Figure S2.6 Phylogeny of the *Gypsy* super-family LTR-RT in *T. occidentale* (To) and *T. pallescens* (Tp) genomes.

Phylogeny of the reverse transcriptase (RT) amino acid sequences of the *Gypsy* super-family in *T. occidentale* (To) and *T. pallescens* (Tp) genomes by maximum likelihood method. The tree branches are colored by clades. The tree tips are color by the locations of the LTR-RT in *T. occidentale* (To) or *T. pallescens* (Tp) species.

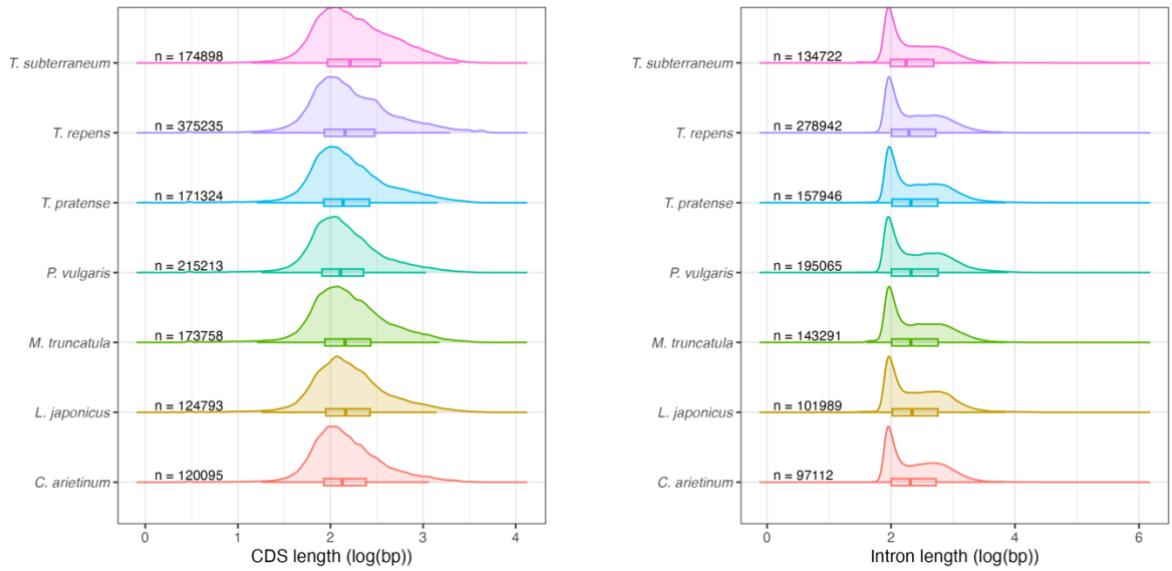


Supplementary Figure S2.7 Chromosome categorized into *T. occidentale* (to) or *T. pallescens* (tp) subgenome.

The Blast ratio is defined as $\sum(To\ alignment\ length)_i - \sum(Tp\ alignment\ length)_k\ /\ Bin\ size\ (0.5\ Mb)$, where i and k are the filtered hits to either the T. occidentale (To) or T. pallescens (Tp) subgenome, respectively. This measure creates an index where a value of -1 is completely T. pallescens-like and +1 is completely T. occidentale-like.

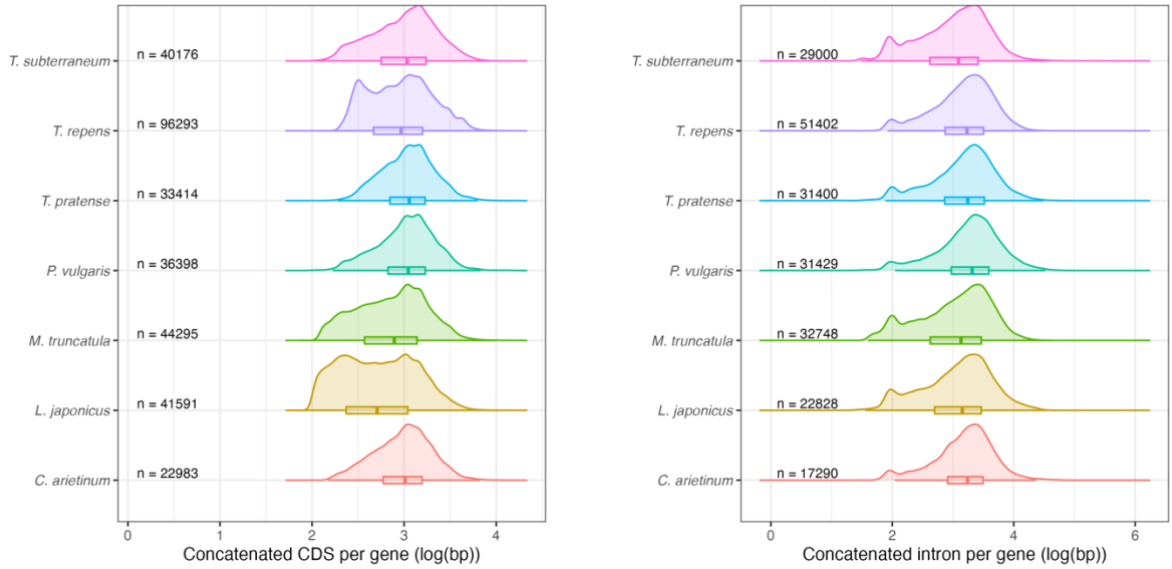
(A)

(B)



(C)

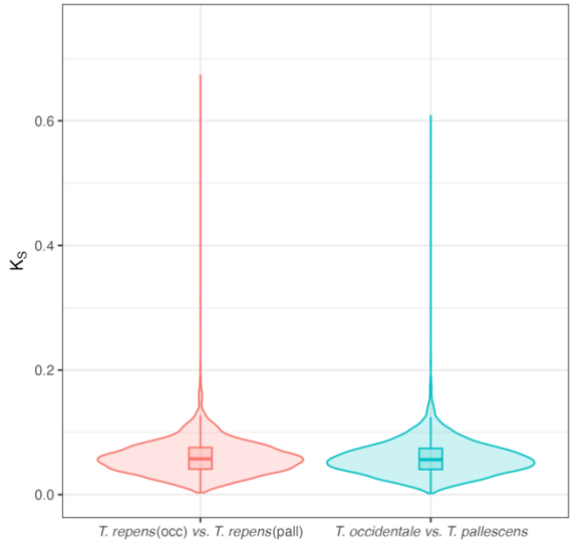
(D)



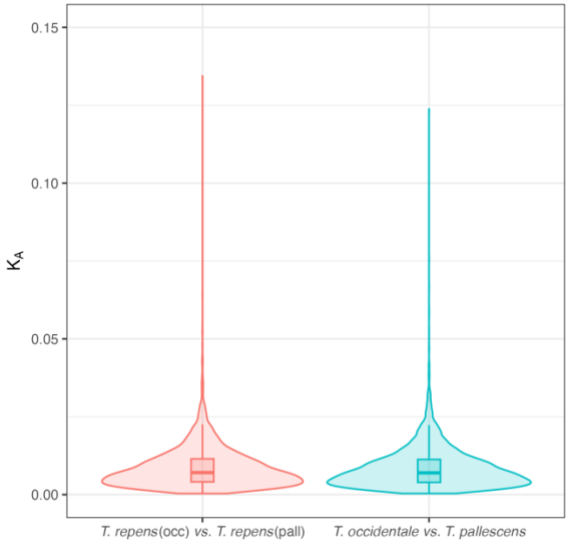
Supplementary Figure S2.8 Coding sequence (CDS) length and intron length distribution.

(A-B), the CDS or the introns are pooled and counted individually. **(C-D)**, the CDS or the introns are concatenated per gene before counting.

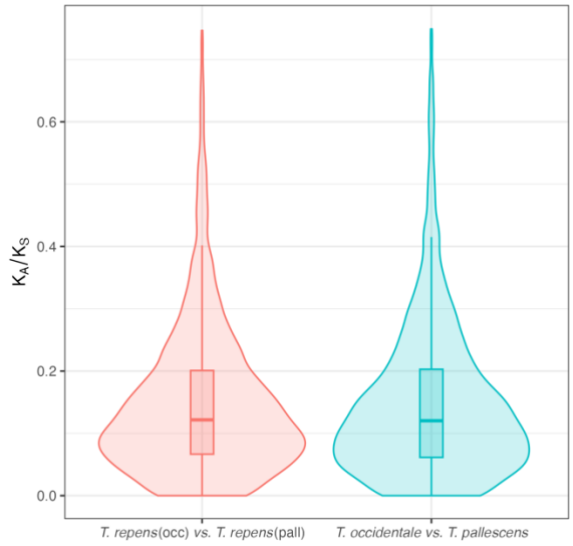
(A)



(B)



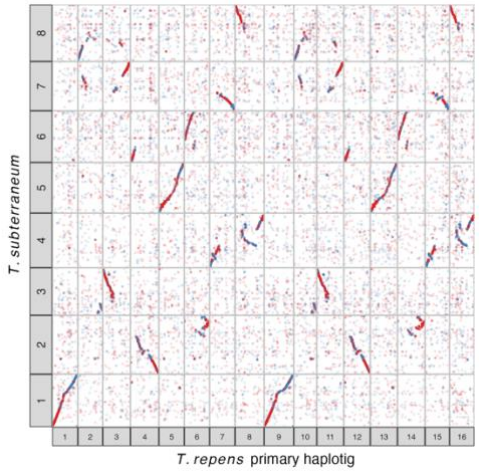
(C)



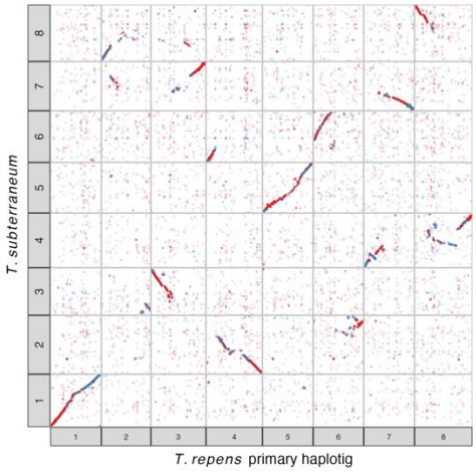
Supplementary Figure S2.9 Pairwise nonsynonymous mutation rate (K_A) and synonymous mutation rate (K_S) between the white clover subgenomes and the diploid progenitors' genomes.

No significant differences were detected by Student's t test.

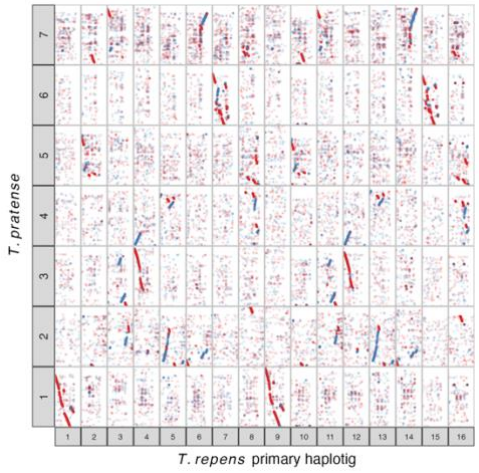
(A)


T. subterraneum
T. repens primary haplotig

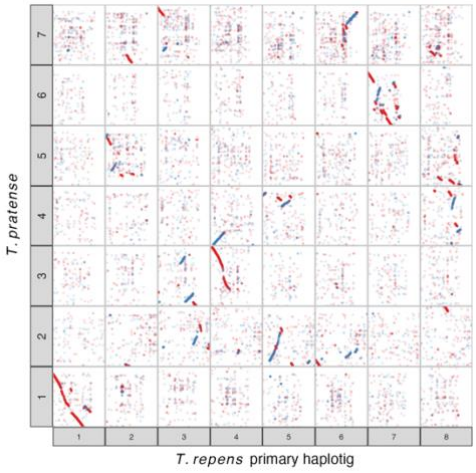
(B)


T. subterraneum
T. repens primary haplotig

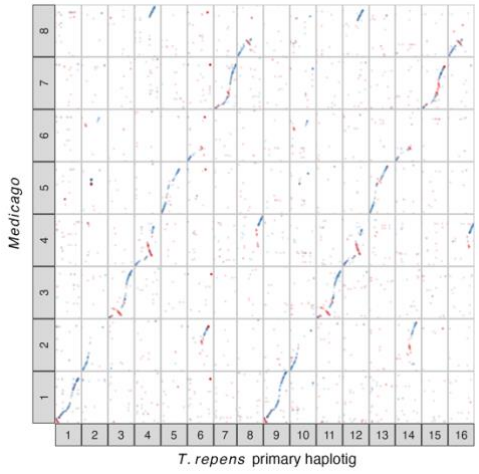
(C)


T. pratense
T. repens primary haplotig

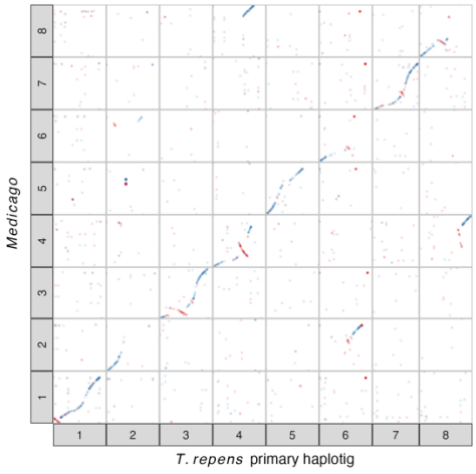
(D)


T. pratense
T. repens primary haplotig

(E)

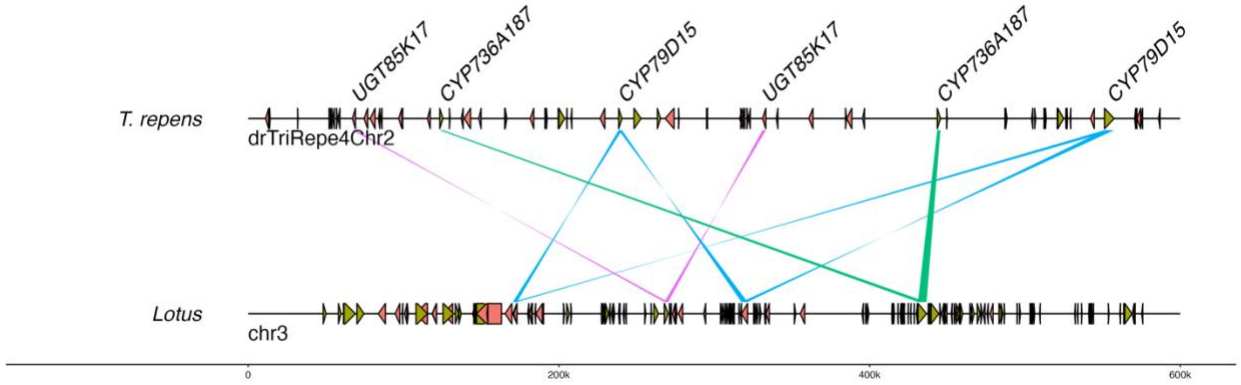

Medicago
T. repens primary haplotig

(F)

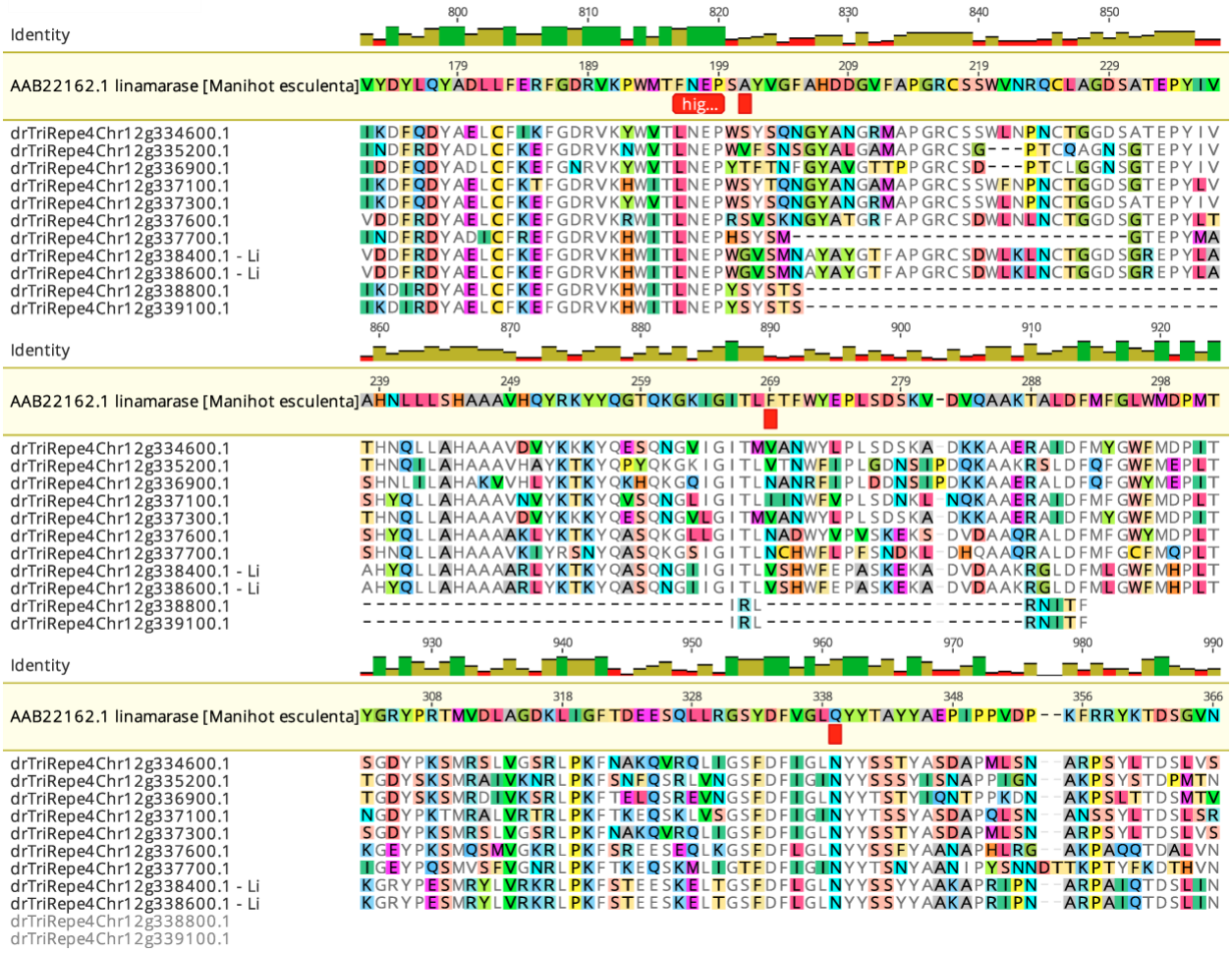

Medicago
T. repens primary haplotig

Supplementary Figure S2.10 Whole genome nucleotide alignment of the white clover (this study) and related species.

Blue, the pairwise alignment is in the same direction. Red, the pairwise alignment is in the opposite directions. (B), (D), (F), only the chr 1 – chr 8 are shown. Only the alignments of the length >1000 bp and the identity >90% are shown.



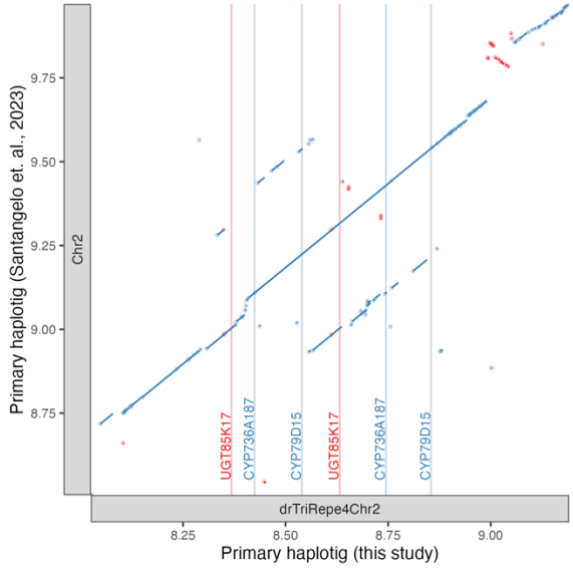
Supplementary Figure S2.11 *Ac* gene cluster. Gene synteny between white clover chr 2 (primary haplotig) and *Lotus japonica* chr 3.



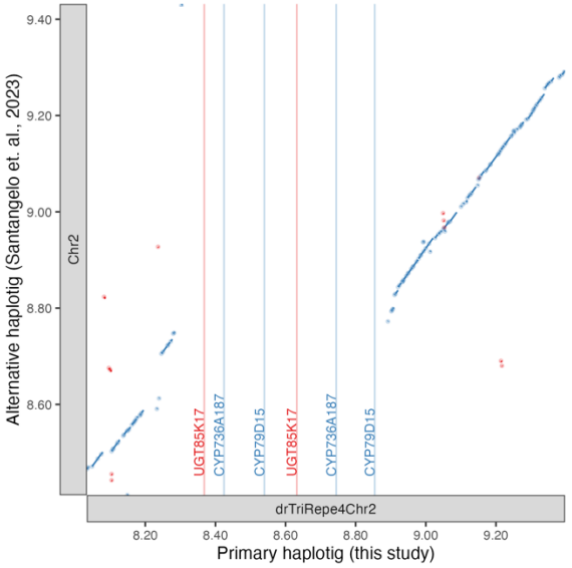
Supplementary Figure S2.12 Nucleotide alignment of the *Li* gene (cyanogenic glucosidase) and other glucosidase gene.

The other glucosidase gene are predicted to have different substrate specificities than linamarase, at the *Li* locus. Only partial alignment, including the enzyme active center (the red annotation in the cassava linamarase gene), is shown. The amino acid composition around the active center is predicted to affect the substrate specificity (Kongsaeree *et al.*, 2010).

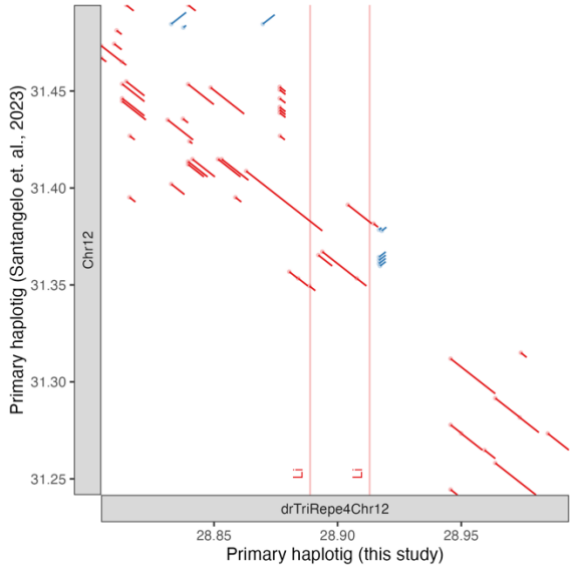
(A)


Primary haplotig (Santangelo et. al., 2023)
Chr2
9.75
9.50
9.25
9.00
8.75
UGT85K17
CYP736A187
CYP79D15
UGT85K17
CYP736A187
CYP79D15
drTriRepe4Chr2
8.25
8.50
8.75
9.00
Primary haplotig (this study)

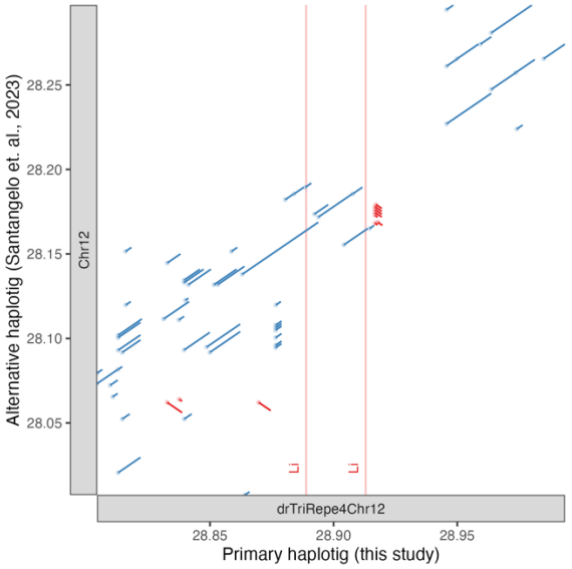
(B)


Alternative haplotig (Santangelo et. al., 2023)
Chr2
9.40
9.20
9.00
8.80
8.60
UGT85K17
CYP736A187
CYP79D15
UGT85K17
CYP736A187
CYP79D15
drTriRepe4Chr2
8.20
8.40
8.60
8.80
9.00
9.20
Primary haplotig (this study)

(C)


Primary haplotig (Santangelo et. al., 2023)
Chr12
31.45
31.40
31.35
31.30
31.25
Li
Li
drTriRepe4Chr12
28.85
28.90
28.95
Primary haplotig (this study)

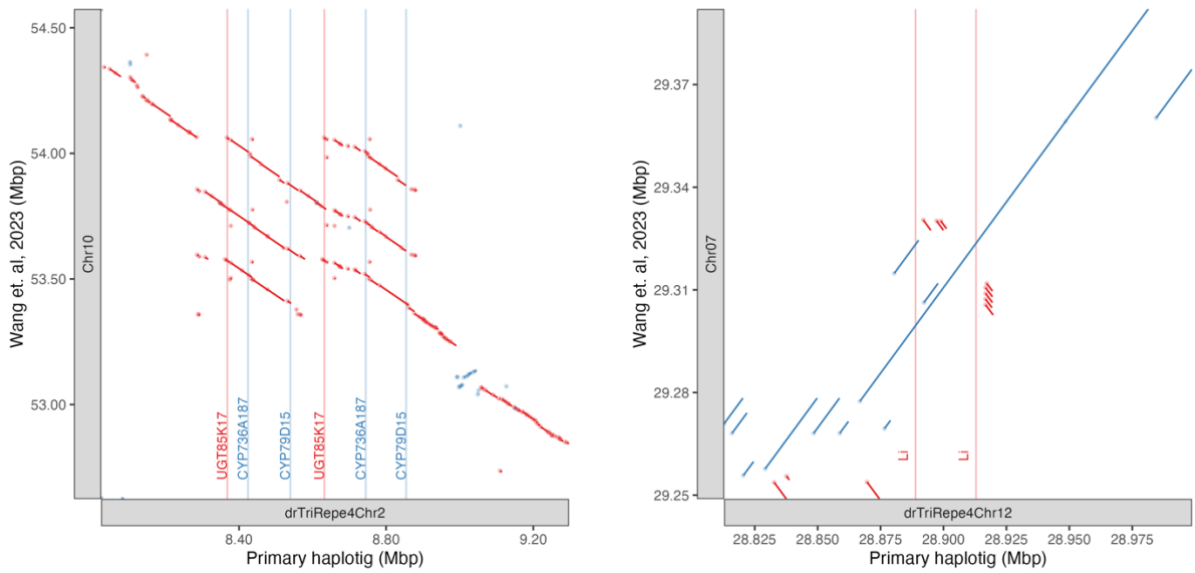
(D)


Alternative haplotig (Santangelo et. al., 2023)
Chr12
28.25
28.20
28.15
28.10
28.05
Li
Li
drTriRepe4Chr12
28.85
28.90
28.95
Primary haplotig (this study)

Supplementary Figure S2.13 Nucleotide alignment of the cyanogenesis gene copy number variation regions between the white clover reference genome of the present study and that of Santangelo *et al.* (2023).

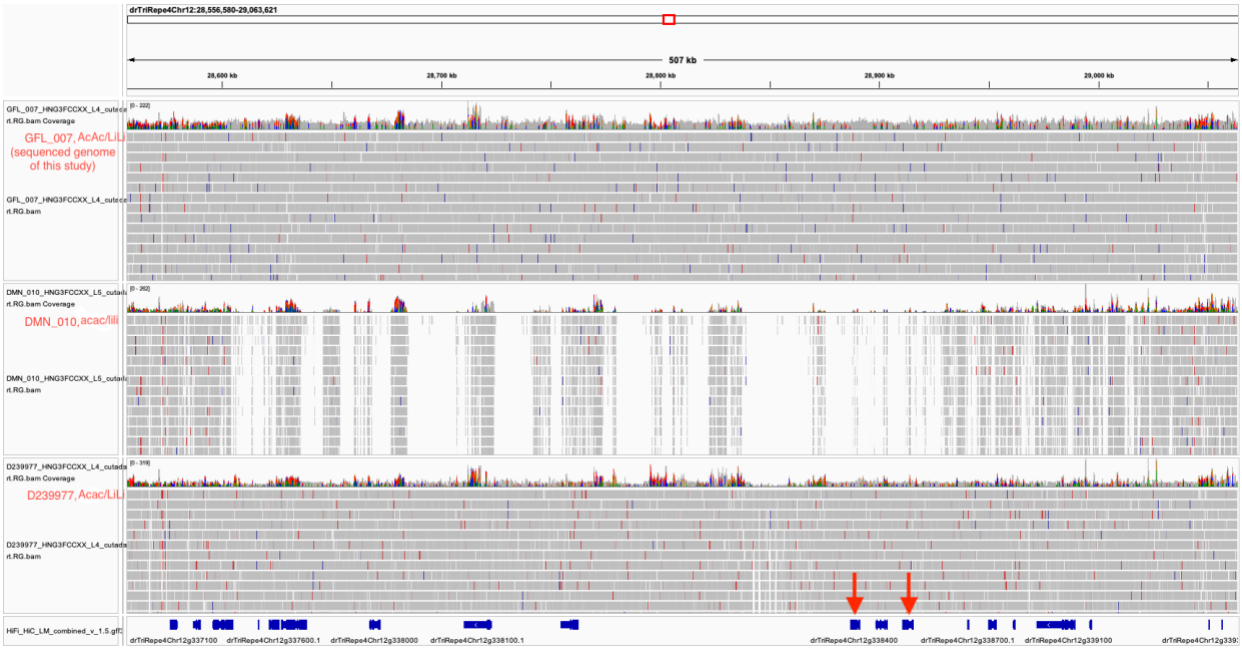
(A), there are two *Ac* gene clusters in the primary haplotig of this study and of Santangelo *et al.* (2023) (the existence of the genes has been confirmed by BLAST). **(B)**, whole *Ac* gene cluster deletion in the alternative haplotig of Santangelo *et al.* (2023). **(C)-(D)**, there are two copies of *Li* gene in each haplotig of Santangelo *et al.* (2023).

(A) (B)



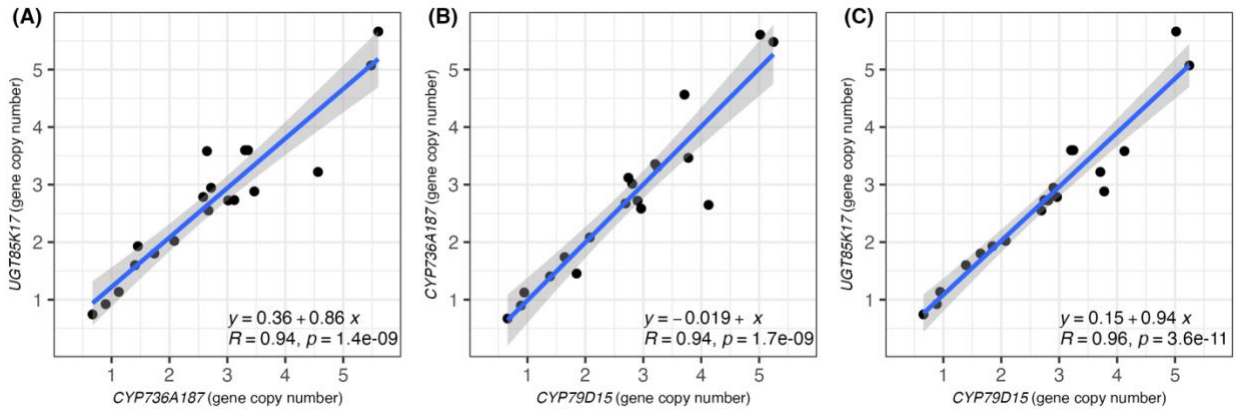
Supplementary Figure S2.14 Nucleotide alignment of the cyanogenesis gene copy number variation regions between the existing white clover reference genome and that of Wang *et al.* (2023).

(A), there are two *Ac* gene clusters in the primary haplotig of this study and three *Ac* gene clusters in the genome of Wang *et al.* (2023). However, only two *CYP79D15* genes can be BLAST found (See **Supplementary Table S2.4**). (B), there are two copies of *Li* gene in the primary haplotig of this study and in the genome of Wang *et al.* (2023).

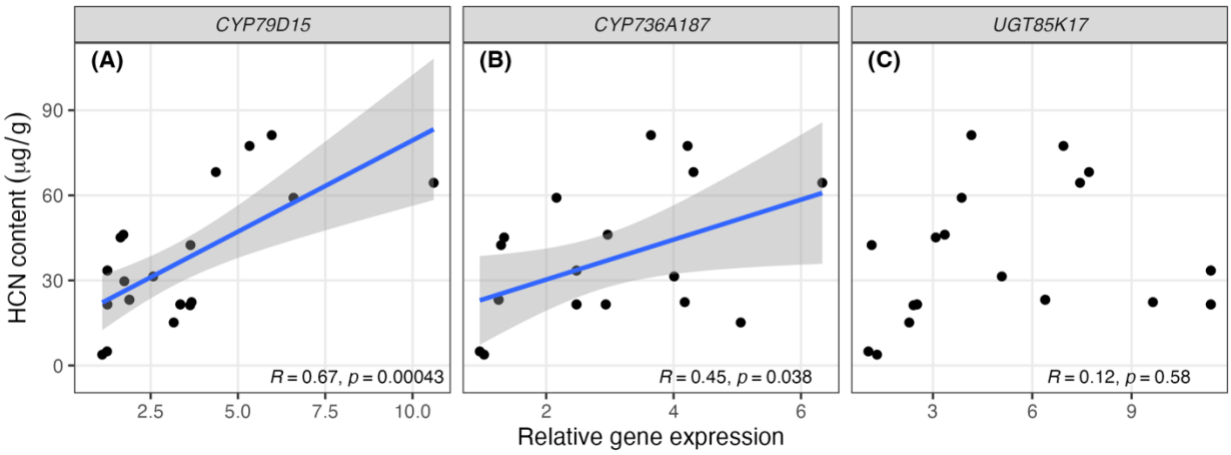


Supplementary Figure S2.15 Mapping result at the *Li* locus.

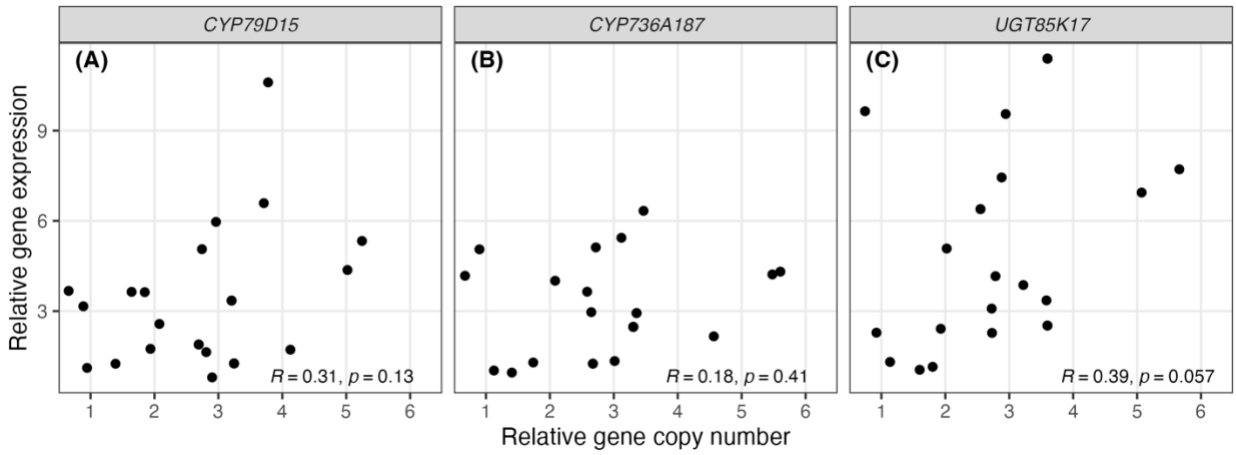
Mapping result of the whole genome re-sequencing of the accessions GFL_007 (NCBI BioSample: SAMN37329216) (*AcAc*/*LiLi*, sequenced genome of this study), DMN_010 (NCBI BioSample: SAMN34157026) (*acac*/*lili*), and PI239977 (NCBI BioSample: SAMNXXXXXXXX) (*Acac*/*LiLi*). Note that in the accession that is homozygous for the *Li* locus deletion (*lili*), there is a low mapping window up to 500 kbp.



Supplementary Figure S2.16 Pairwise gene copy number correlation of the genes in *Ac* gene cluster.

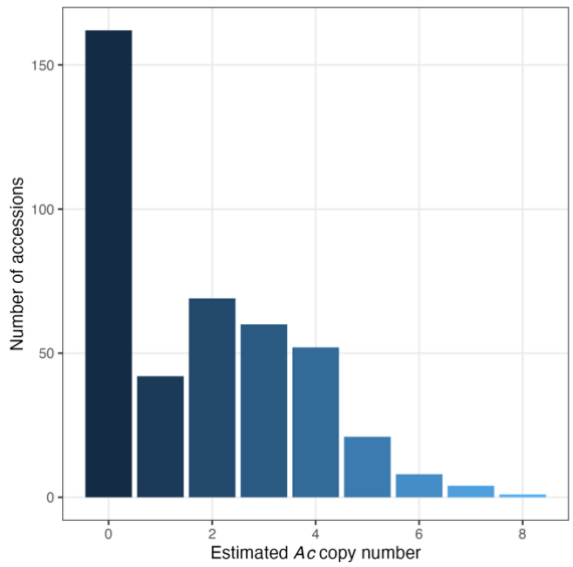


Supplementary Figure S2.17 Relative gene expression and cyanide (HCN) content.

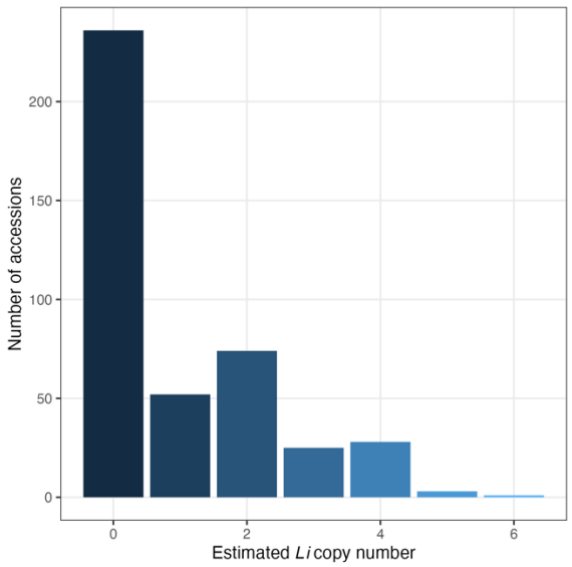


Supplementary Figure S2.18 Gene copy number variation and relative gene expression. (A-C), *Ac* gene cluster.

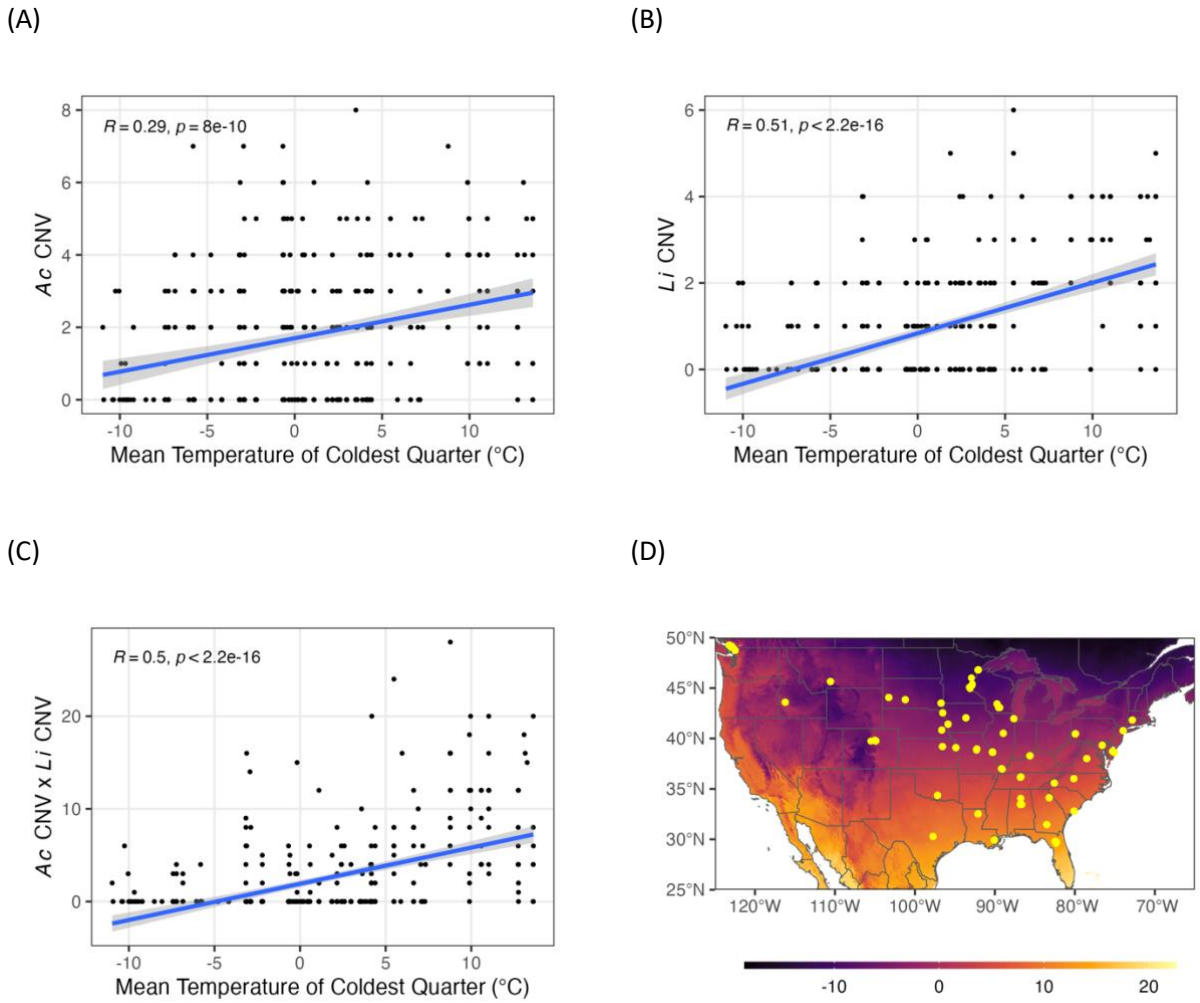
(A)



(B)



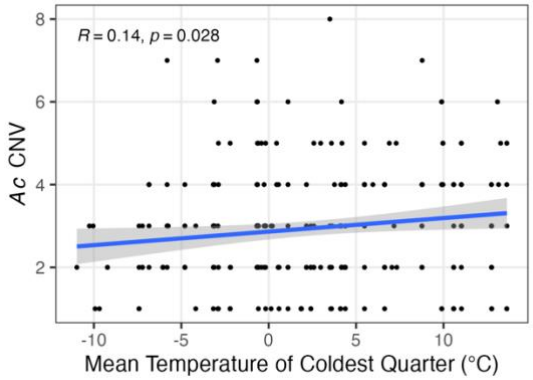
Supplementary Figure S2.19 Estimated *Ac* (A) and *Li* (B) copy number of 419 wild accessions by genomic qPCR.



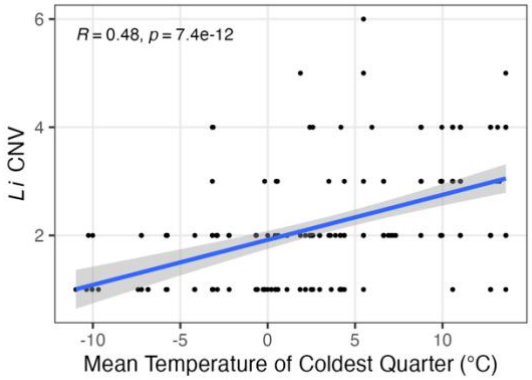
Supplementary Figure S2.20 Association between cyanogenesis gene copy number variation (CNV) and mean temperature of coldest quarter (MTCQ).

(A), *Ac* CNV. (B), *Li* CNV. (C), the product of (*Ac* CNV)*(*Li* CNV). (D), map of 419 accessions (43 populations) across North America. The map is shaded by MTCQ.

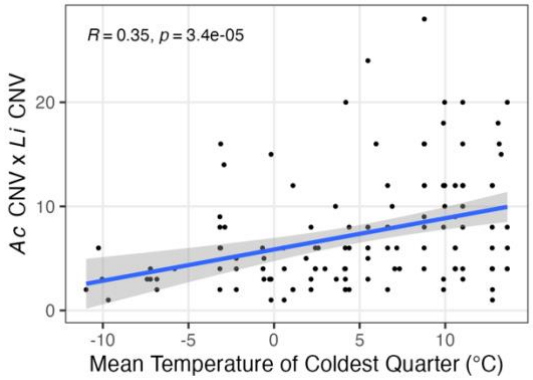
(A)



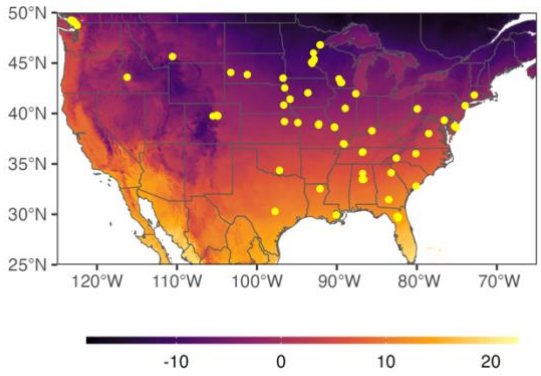
(B)



(C)

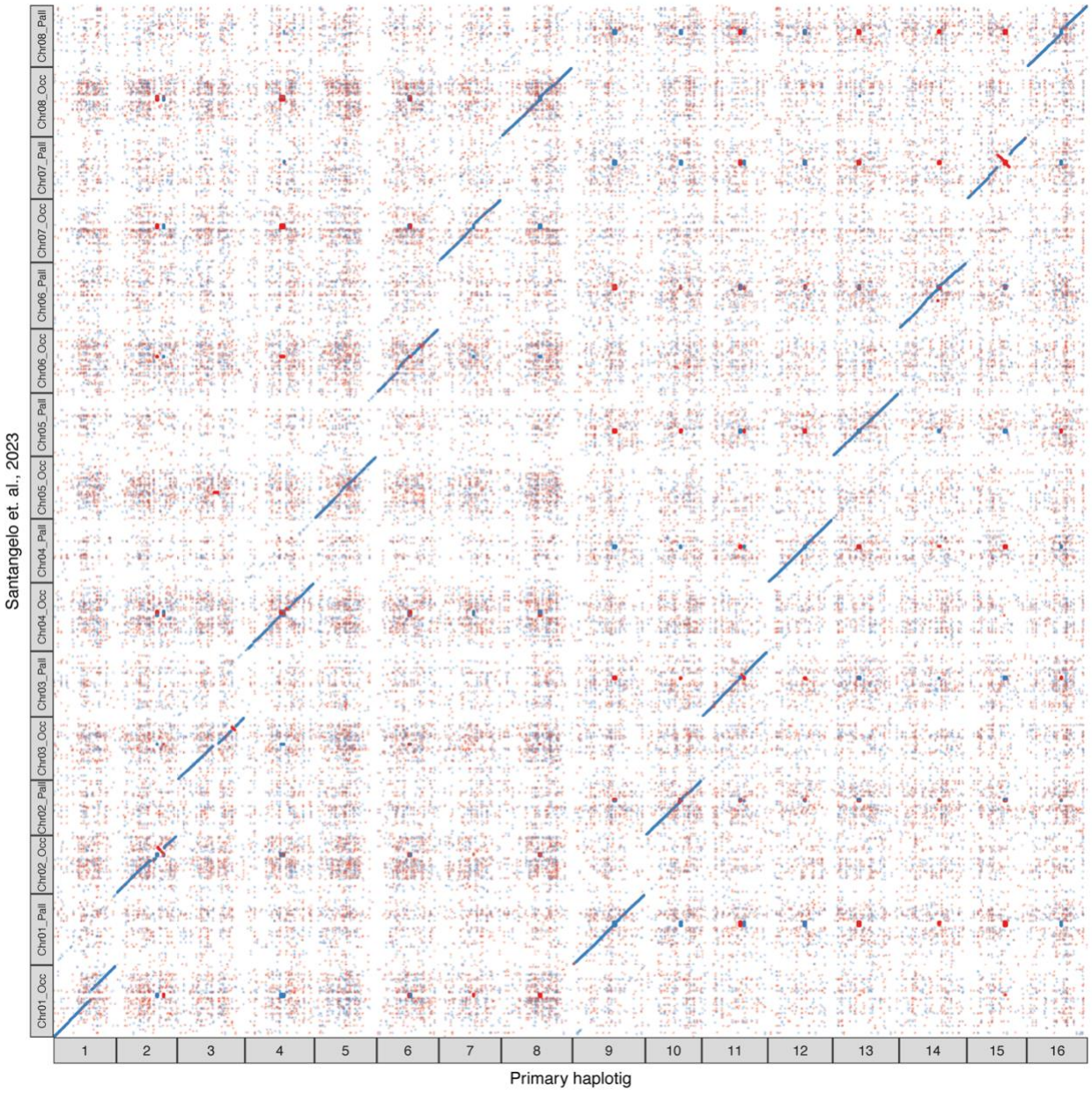


(D)



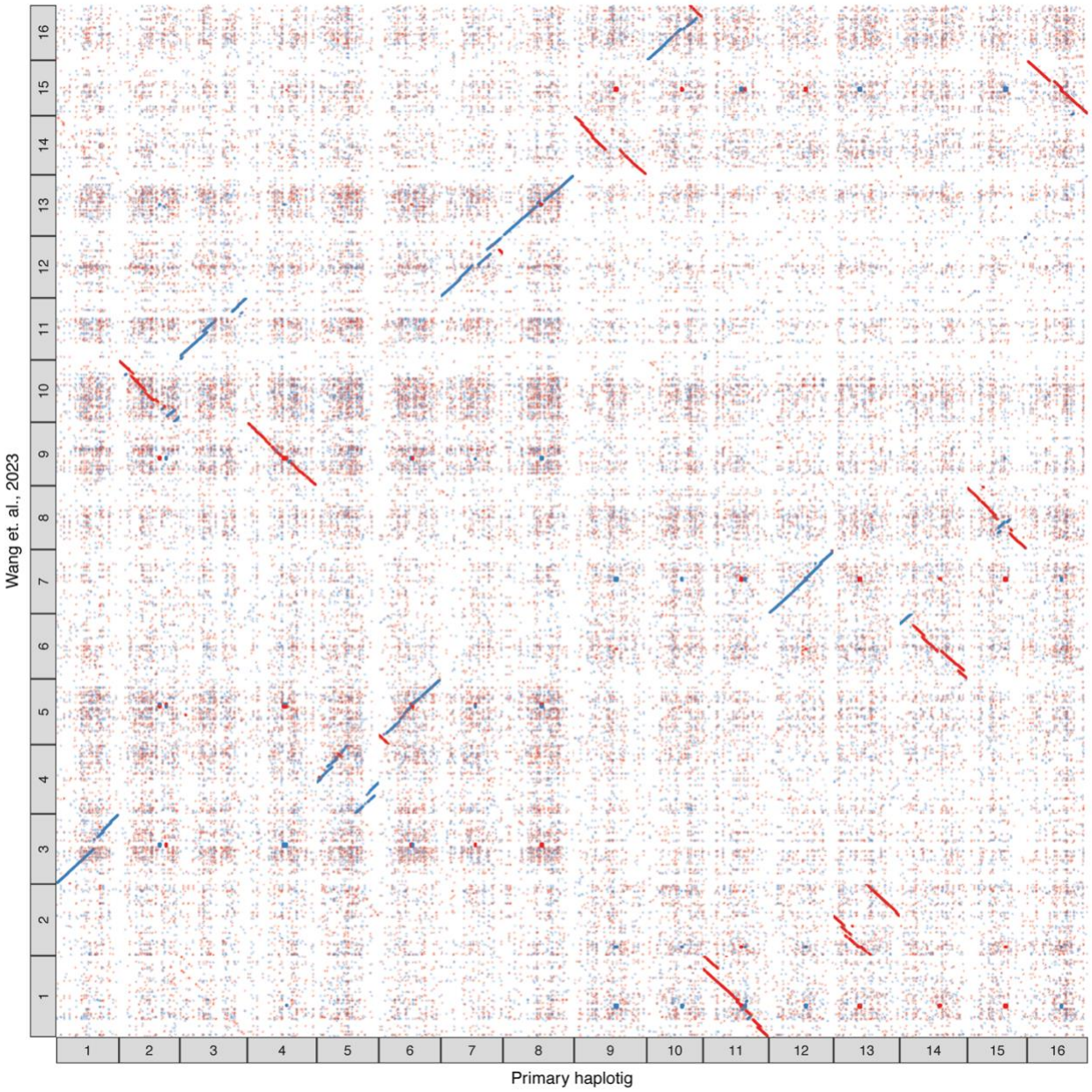
Supplementary Figure S2.21 Association between cyanogenesis gene copy number variation (CNV) and mean temperature of coldest quarter (MTCQ). Only the data with CNV ≥ 1 are shown.

(A), *Ac* CNV. (B), *Li* CNV. (C), the product of (*Ac* CNV)*(*Li* CNV). (D), map of 419 accessions (43 populations) across North America. The map is shaded by MTCQ.



Supplementary Figure S2.22 Whole genome nucleotide alignment between the primary haplotig (this study) and Santangelo *et al.* (2023).

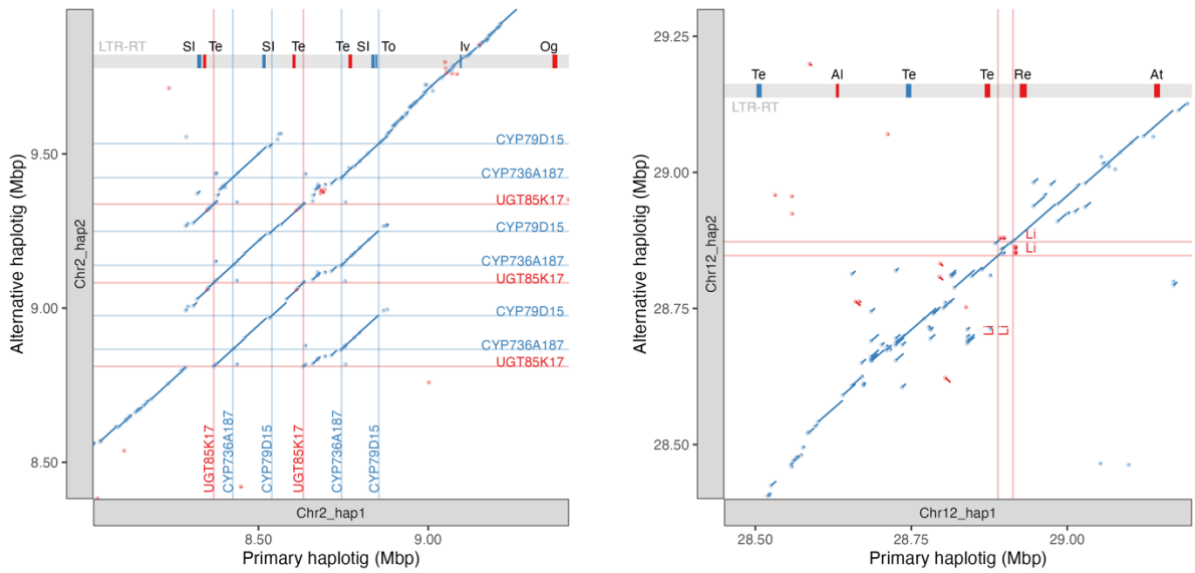
Whole genome nucleotide alignment (identity >95%, query length >10,000 bp, hit length >10,000 bp). Linkage map assisted scaffolding was conducted in both genomes. Note that the two genomes are highly colinear to each other, except for a few scaffolds around the centromeric regions.



Supplementary Figure S2.23 Whole genome nucleotide alignment between the primary haplotig (this study) and. Wang *et al.* (2023).

Whole genome nucleotide alignment (identity >95%, query length >10,000 bp, hit length >10,000 bp). Linkage map assisted scaffolding was only conducted in the primary haplotig (this study). Note that Wang *et al.* (Wang *et al.*, 2023) shows chromosome inversions or duplications at the corresponding chr 2, chr 5, chr 6, chr 10, chr 11, chr 13, chr 14 of the primary haplotig (this study).

(A) (B)



Supplementary Figure S2.24 Nucleotide alignment and gene synteny of the white clover cyanogenesis genes with LTR-RT labeled.

(A-B), nucleotide alignment of the regions around the *Ac* and the *Li* loci. Blue, the pairwise alignment is in the same direction. Red, the pairwise alignment is in the opposite directions. SI, *Copia*/*SIRE*. Te, *Gypsy*/*Tekay*. To, *Copia*/*Tork*. Iv, *Copia*/*Ivana*. Og, *Gypsy*/*Ogre*. Al, *Copia*/*Angela*. Re, *Gypsy*/*Reina*. At, *Gypsy*/*Athila*.

2.15 Supplementary Reference

Abeynayake SW, Panter S, Chapman R, Webster T, Rochfort S, Mouradov A, Spangenberg G. 2011. Biosynthesis of proanthocyanidins in white clover flowers: Cross talk within the flavonoid pathway *Plant Physiology* **158**(2): 666-678.

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019. Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31**(7): 1466-1487.

Kongsaeree PT, Ratananikom K, Choengpanya K, Tongtubtim N, Sujiwattanarat P, Porncharoennop C, Onpium A, Svasti J. 2010. Substrate specificity in hydrolysis and transglucosylation by family 1 β-glucosidases from cassava and Thai rosewood. *Journal of Molecular Catalysis B: Enzymatic* **67**(3): 257-265.

Santangelo JS, Battlay P, Hendrickson BT, Kuo WH, Olsen KM, Kooyers NJ, Johnson MTJ, Hodgins KA, Ness RW. 2023. Haplotype-resolved, chromosome-level assembly of white clover (*Trifolium repens* L., Fabaceae). *Genome Biology and Evolution* **15**(8): 2023.2006.2006.543960.

Wang H, Wu Y, He Y, Li G, Ma L, Li S, Huang J, Yang G. 2023. High-quality chromosome-level de novo assembly of the *Trifolium repens*. *BMC Genomics* **24**(1): 326.

Chapter 3 Temperature gradient as major driver of environmental adaptation in North American white clover (*Trifolium repens* L.)

3.1 Authorship and Affiliations

Wen-Hsi Kuo[1], Limei Zhong[1,2], Sara J. Wright[1,3], David M. Goad[1], Kenneth M. Olsen[1]

[1]Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA

[2]Present address: Key Laboratory of Molecular Biology and Gene Engineering in Jiangxi, School of Life Sciences, Nanchang University, Nanchang, China

[3]Present address: Department of Biological and Biomedical Sciences, Rowan University, Glassboro, New Jersey, USA

This chapter is under revision in *Molecular Ecology*

3.2 Abstract

Species that repeatedly evolve phenotypic clines across environmental gradients have been highlighted as ideal systems for characterizing the genomic basis of local climatic adaptation. However, few studies have assessed the importance of clinally-varying phenotypes for local adaptation: traits that vary clinally may not necessarily be the most critical in determining local fitness. The present study was designed to fill this gap, using a plant species characterized by repeatedly-evolved adaptive phenotypic clines. White clover is naturally polymorphic for the chemical defense cyanogenesis (HCN release with tissue damage); climate-associated cyanogenesis clines have evolved throughout its native and introduced range worldwide. We performed landscape genomic analyses on 415 wild genotypes from 43 locations spanning much of the North American species range to assess the relative importance of cyanogenesis loci vs. other genomic factors in local climatic adaptation. We find clear evidence of local adaptation, with temperature-related climatic variables best describing genome-wide differentiation between sampling locations. The same climatic variables are also strongly correlated with cyanogenesis frequencies and gene copy number variations (CNVs) at cyanogenesis loci. However, landscape genomic analyses indicate no significant contribution of cyanogenesis loci to local adaptation. Instead, several genomic regions containing promising candidate genes for plant response to seasonal cues are identified — some of which are shared with previously-identified QTLs for locally-adaptive fitness traits in North American white clover. Our findings suggest that local adaptation in white clover is likely determined primarily by genes controlling the timing of growth and flowering in response to local seasonal cues.

More generally, this work suggests that caution is warranted when considering the importance of phenotypic clines for local adaptation.

Keywords
Local adaptation; cline; cyanogenesis; genotype-environment association, landscape genomics; temperature; white clover (*Trifolium repens*)

3.3 Introduction

Understanding the phenotypic and genetic bases of local adaptation, where populations in their native habitats exhibit higher fitness compared to foreign populations, has emerged as a major focus in evolutionary biology (Kawecki & Ebert, 2004; Savolainen *et al.*, 2013; Lascoux *et al.*, 2016; Lasky *et al.*, 2022). One common approach is to employ reciprocal transplants and/or common garden experiments to evaluate fitness trade-offs between individuals originating from different environments (Stinchcombe *et al.*, 2004; Agrena *et al.*, 2013; Wright *et al.*, 2018; Wright *et al.*, 2022). When combined with genotypic information, such as quantitative trait locus (QTL) mapping of fitness traits, this approach can elucidate the genetic architecture of locally adaptive phenotypes (Agrena *et al.*, 2013; Grillo *et al.*, 2013; Wright *et al.*, 2022). A complementary approach for studying local adaptation is to employ landscape genomic methods to identify genomic regions with signatures of local adaptation; these may include loci that show evidence of adaptive differentiation between environments (e.g., F_{ST} outliers) or strong environmental correlations as detected in genotype-environment association (GEA) analyses (Guerrero *et al.*, 2018; Gugger *et al.*, 2018; reviewed in Lasky *et al.*, 2022; Battlay *et al.*, 2023; Wang, Y *et al.*, 2023).

Among the different study systems that can potentially be used to study local adaptation, species that have repeatedly evolved phenotypic clines across an environmental gradient in different parts of their range have been held up as a "best-case scenario" for successfully detecting locally-adapted traits and their underlying genetic basis (Lasky *et al.*, 2022). In such species, the clinal variation is more likely to be reflecting a genuine signal of

environmental adaptation than to be an artifact of neutral population structure or demographic history. However, even for such "best-case" study systems, it is important to recognize the distinction between traits that show clinal variation and traits that are most crucial for local adaptation: just because natural selection repeatedly favors the evolution of a phenotypic cline in a species, it does not necessarily follow that the clinally-varying phenotype is most critical for determining fitness in local environments.

Nonetheless, very few studies have explicitly assessed the relative importance of traits that show clinal variation for local adaptation. In species with repeatedly-evolved phenotypic clines, research has tended to focus on measuring the fitness effects of the known phenotypes, without knowledge of fitness contributions from other traits or loci (e.g., Mullen & Hoekstra, 2008; Linnen *et al.*, 2009; Antoniazza *et al.*, 2010; Wittkopp *et al.*, 2011; Hof *et al.*, 2016; Koski & Galloway, 2018; Gefaell *et al.*, 2023). Conversely, studies using landscape genomic methods to study local adaptation have generally not considered adaptive phenotypic clines (e.g., Fournier-Level *et al.*, 2011; Hancock *et al.*, 2011; Kujala *et al.*, 2017; Price *et al.*, 2018; Battlay *et al.*, 2023). To address this gap, the present study applies landscape genomic analyses to study the genetic determinants of local adaptation in a geographically widespread plant species that features repeatedly-evolved phenotypic clines. By incorporating previously-collected data on fitness traits and QTLs for local environmental adaptation (Wright *et al.*, 2018; Wright *et al.*, 2022), we integrate insights from adaptive phenotypic clines, locally-adaptive fitness QTLs, and genomic signatures of local adaptation to holistically assess the genomic and phenotypic contributions to local environmental adaptation.

Our focal species, white clover (*Trifolium repens*, 2n = 4x = 32), is a geographically widespread herbaceous perennial legume that can be found in mesic regions worldwide. A native species of southern Europe, it was introduced globally as a forage and rotation crop and as a source of soil nitrogen, and it has since become widely naturalized as a component of lawns, roadsides and other mowed or grazed areas (Zeven, 1991; Kjærgaard, 2003). In both its native and introduced range, white clover populations can be found in abundance across a wide climatic range, spanning subtropical to boreal ecozones; this wide environmental adaptability has been attributed to white clover's allopolyploid origin from two ecologically distinct diploid progenitor species (Griffiths *et al.*, 2019).

One clear indicator of white clover's adaptive ability is that populations worldwide have evolved climate-associated clines in the chemical defense trait cyanogenesis (HCN release following tissue damage). The species is characterized by a genetically determined polymorphism for this defense (described below), with higher frequencies of cyanogenic plants generally present in warmer locations. Latitudinal and elevational cyanogenesis clines have been documented worldwide in numerous studies since the 1950s (Daday, 1954b; Daday, 1954a; Daday, 1958; Hughes, 1991; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013; Kuo *et al.*, 2024); more recently, urban-to-rural cyanogenesis clines have also been described (Thompson *et al.*, 2016; Santangelo *et al.*, 2022). While the selective factors that favor the rapid evolution of cyanogenesis clines have not been definitively determined, they likely involve fitness trade-offs in areas of high and low herbivore abundance, energetic costs associated with production of the required chemical precursors, and abiotic stress factors that could include freezing and

drought (reviewed by Olsen *et al.*, 2013; Kooyers *et al.*, 2014; Kooyers *et al.*, 2018). Regardless of the specific selective factors at play, the fact that climate-associated cyanogenesis clines have evolved repeatedly throughout the species range suggests that selection on the polymorphism is strong, and that this trait could be an important determinant of local environmental adaptation.

Despite the apparent strong selection favoring the repeated evolution of climate-associated cyanogenesis clines, our previous studies have not implicated the cyanogenesis polymorphism or its underlying genes as significant contributors to local climatic adaptation (Wright *et al.*, 2018; Wright *et al.*, 2022). In a common garden experiment using 161 wild North American genotypes sampled from 15 locations spanning the North American species range, vegetative and reproductive fitness in a central US location were found to be strongly positively correlated with the climatic similarity of a genotype's location-of-origin and the common garden site; however, cyanogenesis variation was only a negligible predictor of fitness in the common garden (Wright et al. 2018). In a second study, QTL mapping of fitness traits was performed using clonally-replicated F_2 progeny of two biparental crosses between plants originating from the northern, central and southern US, with fitness measured in reciprocal common gardens in the locations of both parental genotypes. While genetic mapping of fitness traits revealed clear evidence of local adaptation and allelic trade-offs between environments, there was again little evidence that the cyanogenesis polymorphism contributes significantly to local adaptation (Wright et al. 2022). However, both of these studies were necessarily limited by experimental design (see reviews by Savolainen *et al.*, 2013; Germino *et*

al., 2019) — most notably, a single common garden site in the study of Wright *et al.* (2018), and just three parental genotypes in the QTL mapping study of Wright *et al.* (2022). This leaves open the possibility that the lack of evidence for the cyanogenesis polymorphism as a contributor to local adaptation is an artifact of limited resolution in studies conducted to date.

To complement the previous common garden and QTL mapping approaches in white clover, the present study leverages white clover's recently published high-quality genome assemblies (Santangelo *et al.*, 2023; Wang, H *et al.*, 2023; Kuo *et al.*, 2024) to conduct genotype environment association (GEA) analysis and genomic differentiation scans to test for signatures of local adaptation and the extent to which they do or do not involve cyanogenesis loci. Using a geographically representative sample of white clover genotypes from across North America (415 accessions from 43 locations), we sought to address four key questions: **1)** Is genome-wide genetic differentiation across the sampled range better explained by isolation-by-distance (IBD), which would be consistent with neutral population structure, or by isolation-by-environment (IBE), which would implicate environmental selection as a factor shaping the distribution of genotypes (Wang & Bradburd, 2014)? **2)** Does cyanogenesis clinal variation across the sampled range parallel genome-wide patterns of population differentiation, or are locus-specific patterns apparent for the genes controlling this chemical defense polymorphism? As a corollary, which environmental variable(s) can best predict cyanogenesis frequencies across the sampled populations? **3)** Do GEA and genomic differentiation scans identify the cyanogenesis loci as genetic contributors to local environmental adaptation? **4)** To what extent do genomic regions detected by the GEA and genomic differentiation scans overlap with fitness

QTLs for local adaptation identified in our previous study (Wright *et al.*, 2022)? Are there candidate genes of known functions near the detected signals? To our knowledge, this study marks the first application of landscape genomic approaches to study local adaptation in a system characterized by adaptive phenotypic clines, allowing a comprehensive assessment of the relative importance of the clinal variation for local adaptation.

3.4 Materials and Methods

Sample Collection, DNA Extraction, and GBS Library Preparation

Using a nationwide network of K-12 science teachers, citizen scientists and colleagues, we obtained mature seeds or stolon cuttings for 419 wild white clover accessions across 43 locations in North America during the growing seasons of 2014-2017 (**Figure 3.1**; **Supplementary Table S3.1**). Each location was represented by 6 to 11 accessions (individual genotypes), and latitude and longitude were recorded for each sample. Seeds and stolon cuttings were cultivated in the greenhouse at Washington University in St. Louis under standard greenhouse conditions (see Wright *et al*. 2022). Genomic DNA was extracted from young leaves using a standard DNA extraction protocol (Whitlock *et al.*, 2008). Extraction quantity and quality were assessed using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, and Qubit™ dsDNA HS Assay Kits.

Genotyping-by-sequencing (GBS) libraries were prepared following Elshire *et al.* (2011) with *ApeK*I methylation sensitive restriction enzyme. Barcoding and protocol modifications are

described in Olsen *et al.* (2021). Paired-end sequencing (150-bp reads) was performed using the Illumina Hi-Seq 2500 platform (Novogene Corp., Chula Vista, CA, USA).

Read Mapping, SNP Calling, and SNP Filtering

Raw GBS reads were demultiplexed with SABRE (https://github.com/najoshi/sabre) and adaptor-trimmed with CUTADAPT (Martin, 2011). The processed reads were mapped back to our white clover reference genome (Kuo *et al.*, 2024) using BWA (Li & Durbin, 2009) with default paired-end settings. SNP calling from the alignments followed GATK best practices, with omission of the duplicated-read removal step as recommended for GBS data (Poplin *et al.*, 2017). The output SNP dataset in vcf format underwent hard filtering (bcftools filter -e 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || INFO/DP < 2500'). It was then filtered for sites with missing accessions < 0.25, missing sites < 0.35, and minor allele frequency > 0.05. Missing genotypes were imputed with Beagle v5.4 (Browning *et al.*, 2018). A relaxed Hardy-Weinberg filter was applied to remove sites with skewed heterozygosity suggesting sequencing error ($p < 1 \times 10^{-50}$). This dataset was used for GWAS and genome-wide environmental association (GEA). For population structure analyses and genomic differentiation scans, three different levels of LD-pruning were performed using PLINK2 (--indep-pairwise 100kb 0.8; --indep-pairwise 200kb 0.5; --indep-pairwise 500kb 0.2) before analysis (Chang *et al.*, 2015).

Population structure analysis and its association with local environmental variables

Population structure was analyzed based on the LD-pruned datasets. Ancestry estimation was performed by ADMIXTURE v1.3.0 with K = 1-8 and cross-validation mode (Alexander *et al.*, 2009). Principal component analysis (PCA) and pairwise F_{ST} estimation (between the 43 sampling locations) were conducted by PLINK2 (Chang *et al.*, 2015). To assess distributions of total genetic variation within and across population locations, an analysis of molecular variance (AMOVA) and its subsequent randomization test of significance (permutation = 1000) were carried out with the 'poppr' package in R (Kamvar *et al.*, 2014).

To explore the correlations between genetic differentiation, geographical distance, and environmental distance, a Mantel test and partial Mantel test (10,000 permutations apiece) were conducted with the 'Vegan' package in R (Dixon, 2003). Prior to the tests, F_{ST} was linearized as $F_{ST}/(1-F_{ST})$. Geographical distance was calculated as the great circle distance between sampling locations (using the mean value of geographical coordinates for all accessions in a given location) with the 'sf' package in R (Pebesma, 2018). The environmental distance was calculated as the Euclidean distance between the mean values of pairwise locations. Environmental values of sampling locations were derived from the 1000-m-buffered mean of WorldClim2 (Fick & Hijmans, 2017) and ENVIREM (Title & Bemmels, 2018) databases.

Cyanogenesis polymorphism and its association with local environmental variables

The white clover cyanogenesis polymorphism is controlled by two independently-segregating simple Mendelian polymorphisms that control the presence/absence of two

cyanogenic precursors, both of which must be present for a plant to be cyanogenic: *Ac/ac* controls the presence/absence of cyanogenic glucosides (stored in the vacuoles of photosynthetic tissue); and *Li/li* controls the presence/absence of their hydrolyzing enzyme, linamarase (present in the apoplast). In cyanogenic plants, tissue damage that causes cell rupture brings the two precursors together, leading to the liberation of HCN. At the molecular level, both the *Ac/ac* and *Li/li* polymorphisms correspond to gene presence/absence variation (PAVs), where recessive alleles correspond to genomic deletions (Olsen *et al.*, 2007; Olsen *et al.*, 2008; Olsen & Small, 2018). Additionally, the dominant (functional) *Ac* and *Li* alleles both show tandem gene copy number variations (CNVs), which contribute to quantitative variation in the cyanogenic response (Kuo *et al.*, 2024). In wild North American populations, the CNVs at both loci show clinal variation in patterns that parallel climate-associated clines for the *Ac/ac* and *Li/li* PAVs (Kuo *et al.*, 2024).

Phenotyping and genotyping of the cyanogenesis polymorphism for the sampled white clover accessions are described in Kuo *et al.* (2024). In brief, the presence/absence of cyanogenic glucosides and linamarase were assessed by Feigl-Anger cyanogenesis assays following previously described protocols (Olsen *et al.*, 2007). For plants lacking given precursor, PAVs were confirmed by PCR-genotyping of *CYP79D15* (the first gene in the *Ac* gene cluster) and *Li* (Olsen *et al.*, 2007; Olsen *et al.*, 2008; Olsen & Small, 2018). To test whether epistatic selection on the *Ac/ac* and *Li/li* polymorphisms affects genotype frequencies in nature (a prediction given that cyanogenesis requires the co-occurrence of dominant alleles at both loci) (Ennos, 1982; Kooyers & Olsen, 2012), we applied a Pearson's Chi-square test to compare

observed frequencies of the two-locus genotypes with frequencies expected for unlinked genetic polymorphisms. Plants producing one or both cyanogenic precursors underwent further examination via genomic DNA qPCR to estimate CNVs (see Kuo *et al.*, 2024 for methodological details). The mean copy number of the *Ac* or *Li* locus was calculated for each of the 43 sampling locations before conducting the association analysis.

Scanning for genome-wide environmental association

To assess the capability of our GBS dataset to capture adaptive signals, the dataset was first used in a genome-wide association study (GWAS) to map the loci that control the cyanogenesis polymorphism and confirm that the mapped locations correspond to the known genomic locations of *Ac* and *Li* (Olsen et al. 2021; Kuo et al. 2024). The GWAS was conducted using PLINK2, with the first ten principal components serving as the covariates.

GEA analysis. To identify the SNPs significantly associated with local environmental variables within the context of population structure, we employed a latent factor mixed model (LFMM) to examine the relationship between genetic and the environmental datasets (Caye *et al.*, 2019). LFMM requires no prior population information, and covariates (i.e., population structure or relatedness matrix) are calculated at the individual level, aligning well with our dataset's lack of discrete population structure (see Results). We set the number of latent factors (K) in the LFMM at three, determined as the best estimate from the ADMIXTURE analysis. Additionally, results from different K values (1-3) were compared. Output p-values

were then transformed into q-values for false discovery rate (FDR) calibration using the 'qvalue' package in R.

Differentiation scan. To identify the SNPs significantly differentiated in the context of admixed population structure, we utilized PCadapt v4.3.5 to calculate the distance of a SNP to the overall genetic background (Privé *et al.*, 2020). The K value for PCadapt (the number of principal components) was set at two, following Cattell's graphical rule as suggested in the original paper. A conventional p-value threshold was set at 5×10^{-8} based on the distribution of *p* values in a Q-Q plot (**Supplementary Fig. S3.1**).

All annotated genes in our reference genome (Kuo *et al.*, 2024) located within 25 Kbp of significant SNPs in the GEA and PCadapt analyses were manually examined based on descriptions in the UniProt database. Genes related to light response, developmental processes, stress, or pathogen response were retained as potential candidate genes for local environmental adaptation. Genomic regions identified in the landscape genomic analyses were also compared to locations of previously-identified fitness QTL for local adaptation in white clover (Wright et al. 2022) to assess the extent of overlap.

3.5 Results

Geographical distributions of SNP markers and environmental variation

After filtering to remove low-quality individuals and SNPs, the final dataset comprised 415 accessions and 345,762 biallelic SNPs. Three levels of pruning to remove SNPs in LD for population structure analysis resulted in the removal of 104,391 (mild pruning), 156,662 (medium pruning), and 226,663 (strong pruning) SNPs, respectively. The different degrees of LD pruning did not alter population structure inferences (**Supplementary Fig. S3.2**); therefore, the dataset subjected to mild LD pruning was selected for subsequent analyses. Across the genome, the majority of SNPs with significant LD were within 25 Kbp, decreasing to background levels beyond 200 Kbp (**Supplementary Fig. S3.3A**); this rapid LD decay is consistent with white clover's obligately outcrossing mating system as a self-incompatible species.

In ADMIXTURE analyses with assumed ancestral population numbers ranging from K = 1 to 8, the lowest cross-validation error was observed at K = 3 (**Supplementary Fig. S3.4**). Most of the 43 sampling locations (see **Fig. 3.1**) exhibited admixed compositions at all assumed K values greater than 2, with the relative contributions of inferred ancestral populations varying by latitude (**Fig. 3.2A, Supplementary Fig. S3.5, S3.6).** The five southernmost locations (GFL, MLA, NLA, ATX, TGA) differed from the rest of the range in maintaining a single ancestral population composition up to K=6. In line with the ADMIXTURE results, pairwise F_{ST} values between populations were very low overall (median $F_{ST} = 0.034$), with somewhat greater differentiation detected for comparisons to the five southernmost populations (GFL, MLA, NLA, ATX, TGA; median $F_{ST} = 0.078$) (**Supplementary Fig. S3.7**)**.** In the AMOVA, significant genetic

differentiation was detected between the 43 sampling locations; however, the percentage of total genetic variation explained was low (2.89%) (**Supplementary Table S2; Supplementary Fig. S8**). Instead, most variation was distributed within locations — either between individuals (25.55%) or within them as heterozygosity (71.56%). Together these population structure assessments indicate that the 43 sampling locations do not represent distinct genetic populations, and that there is substantial genetic admixture throughout the range. Subsequent analyses were therefore based on this overall admixed population structure.

To assess environmental divergence among the 43 sampling locations, we performed a PCA of environmental variables. PC1 (explaining 53.78% of the total environmental variance) was strongly correlated with latitude, except for sampling locations in Washington state (BWA) and nearby Vancouver, British Columbia (VBC) which formed a distinct cluster along PC2 (explaining 19.41% of the total environmental variance) (**Supplementary Fig. S9**). This environmental divergence is likely attributable to the moderate oceanic climate of the Pacific Northwest compared to inland locations of similar latitude.

To complement the population structure and environmental divergence analyses, we utilized a genetic PCA to visualize the population structure in relation to environmental variation. **Figure 3.2B** illustrates individual accessions plotted in PC1 and PC2 space. Most samples from a given location clustered together in genetic PCA space, indicating high genetic similarity among individuals within locations (**Supplementary Fig. S3.10**). Consistent with the ADMIXTURE results, samples from southernmost locations (GFL, MLA, NLA, ATX, TGA)

constituted a distinct genetic cluster (top-left corner of **Fig. 3.2B**). Notably, in a pattern that parallels the PCA of environmental data, PC1 for the genetic data also exhibited high collinearity with latitude, although this axis accounted for only a small portion of the total genetic variance (2.62%) (**Fig. 3.2C**). Considering that environmental variables are correlated with the latitude of sampling locations (**Supplementary Fig. S3.9B**), this pattern suggests that latitude-associated environmental gradients could play some role in shaping and maintaining white clover population structure in North America.

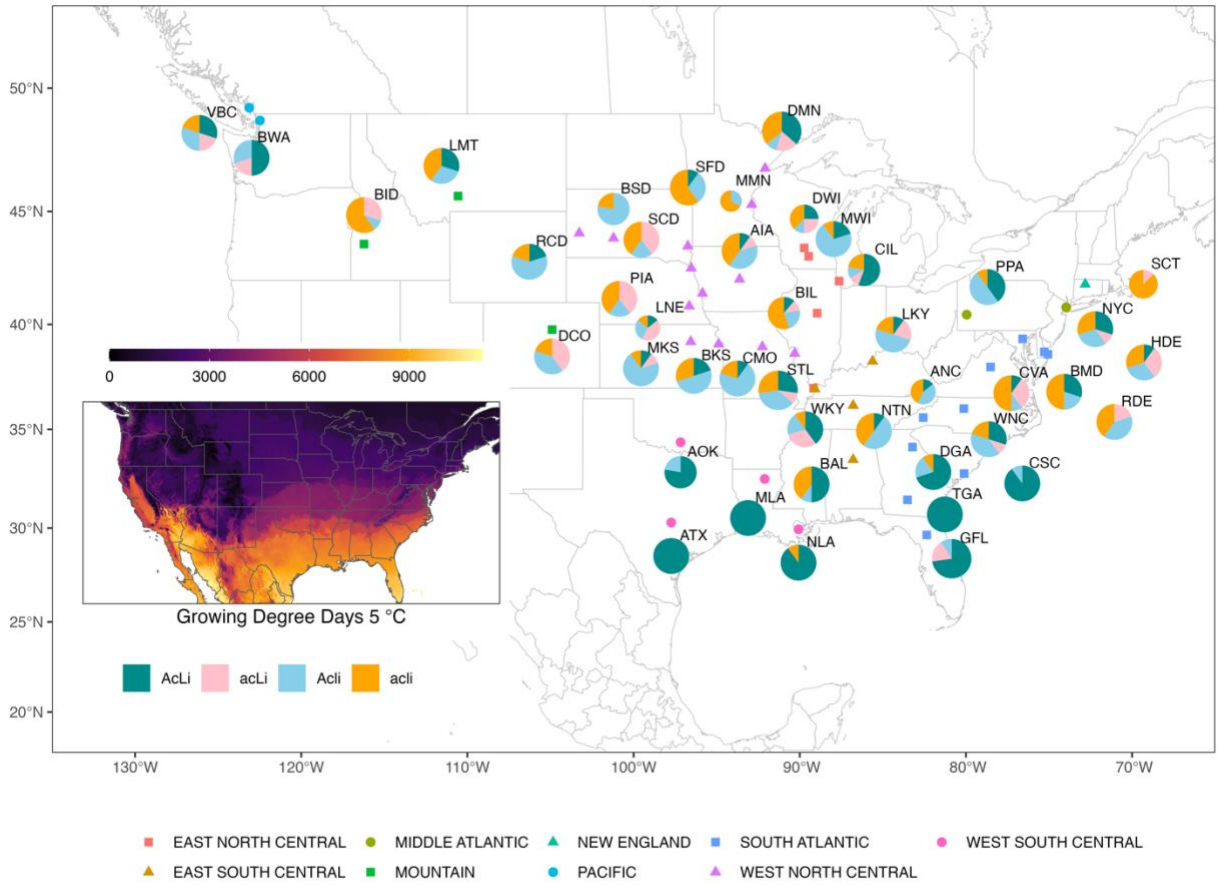


Figure 3.1 North American white clover collections used in the study.

The 415 accessions were collected from 43 locations; broader geographical regions are indicated by different colors and shapes as shown below the x-axis. The pie chart next to a location shows its cyanotype composition. The size of a pie chart is proportional to the number of accessions from the location. The inset figure shows the latitudinal gradient in *Growing Degree Days (> 5*°C*)* across the sampled geographical range. Source of labels for geographical regions: https://www.census.gov/geographies/reference-maps/2010/geo/2010-census-regions-and-divisions-of-the-united-states.html .

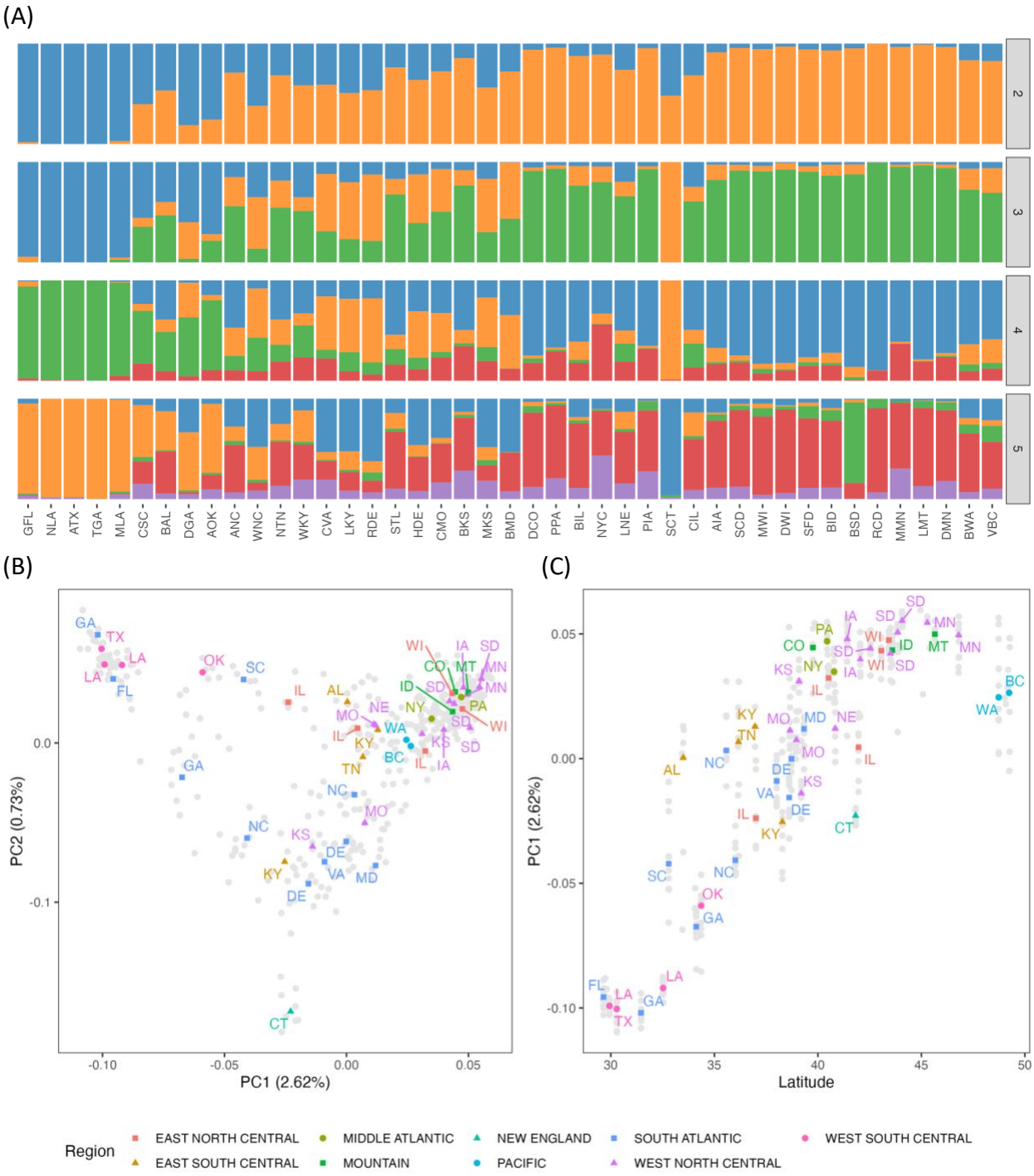


Figure 3.2 Population structure analyses.

(A), ADMIXTURE output at K = 2-5. The minimum cross validation error occurs at K = 3. Locations are ordered along the x-axis by latitude, from lowest (GFL) to highest (VBC). The ancestral composition of each location is based on all accessions of a location. **(B)**, genetic

principal component analysis (PCA). The gray dots indicate the 415 accessions; the colored dots indicate the median values for each sampling location; the state abbreviations of the sampling locations are labeled with color corresponding to the broader geographical regions in **Fig. 3.1**. **(C)**, Association between PC1 and the latitude for the sampling locations.

Associations between genetic differentiation, geographical distance, and environmental divergence

If the geographical distribution of genotypes is shaped by spatial environmental heterogeneity, we would predict that environmental differences between locations would account for more genetic differentiation between locations than geographical distance alone. To test this hypothesis, we assessed pairwise F_{ST} between the 43 sampling locations as a function of geographical distance or environmental divergence. As depicted in **Fig. 3.3A**, genetic differentiation is significantly correlated with geographical distance between locations (Mantel $r = 0.21$, $p = 0.023$); this pattern indicates that there is genetic isolation by distance (IBD) across the sampled range. Notably, however, genetic differentiation exhibits stronger associations with temperature-related environmental variables than with geographical distance alone (**Fig. 3.3B, Supplementary Fig. S3.11**). This result indicates a stronger signal of isolation by environment (IBE) than IBD. Among the temperature-related variables, *Growing Degree Days (> 5℃)*, a common measure of thermal accumulation in the growing season (see **Fig. 3.1** inset), is the most effective in explaining genetic differentiation (Mantel $r = 0.78$, $p < 0.001$) (**Fig. 3.3B, Supplementary Fig. S3.11**). Importantly, the associations with temperature-related variables remained significant even after accounting for the effect of geographical distance (partial Mantel test, **Supplementary Fig. S3.12**). Together these results indicate that environmental distance is a better predictor of genetic differentiation between the sampling locations than geographical distance, providing positive evidence for the existence of IBE and local adaptation in North American white clover.

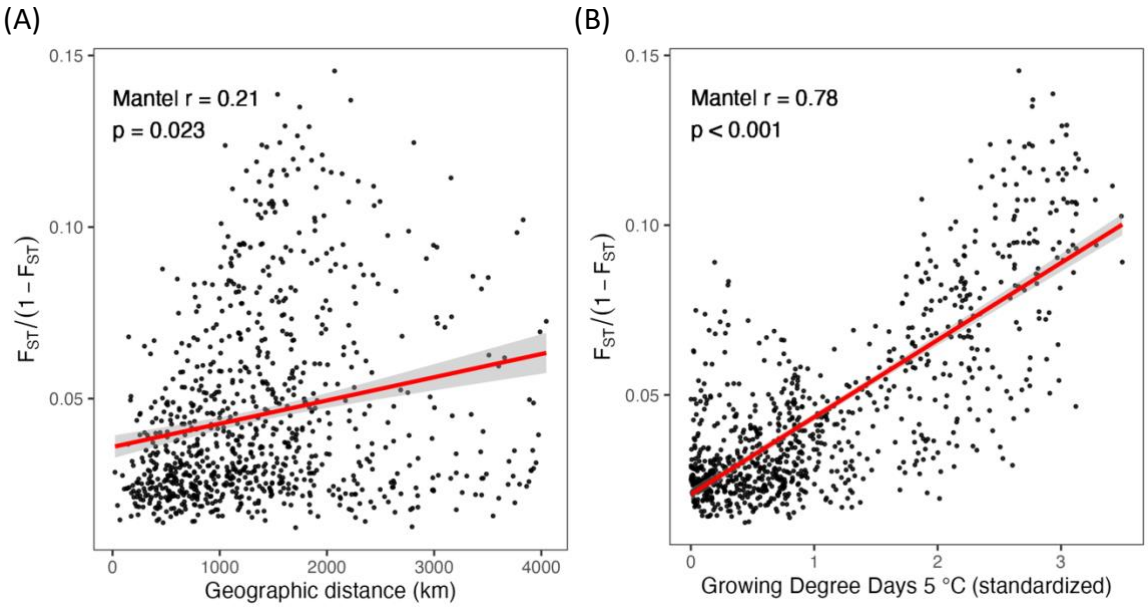


Figure 3.3 Geographical and environmental correlations with genomic differentiation for the sampled locations (n = 43).

(A), Pairwise linearized genetic differentiation between locations ($F_{ST}/1 - F_{ST}$) is significantly associated with geographical distance and **(B)**, Euclidean distance of *Growing Degree Days (> 5°C).*

Cyanogenesis distributions and environmental associations

The white clover cyanogenesis polymorphism is determined through the presence/absence of cyanogenic glucosides (controlled by the *Ac/ac* genetic polymorphism) and their hydrolyzing enzyme linamarase (controlled by the unlinked *Li/li* polymorphism) (see Methods). Therefore, four different cyanogenesis phenotypes (or "cyanotypes") occur in nature: cyanogenic plants ("AcLi"); and acyanogenic plants that lack either cyanogenic glucosides ("acLi"), linamarase ("Acli") or both components ("acli"). The 415 sampled accessions collectively included a total of 134 AcLi (32.29%), 121 Acli (29.16%), 49 acLi (11.81%), and 111 acli (26.75%) cyanotypes. As illustrated in **Fig. 3.1**, the frequency of cyanogenic (AcLi) cyanotypes is highest in the southernmost sampled locations, consistent with previously reported latitudinal cyanogenesis clines in North America (Daday, 1958; Kooyers & Olsen, 2012; Innes *et al.*, 2022).

The cyanogenic phenotype (AcLi) requires the production of both cyanogenic glucosides and linamarase. This creates the possibility that epistatic selection could be operating in nature, favoring plants that are either AcLi (cyanogenic) or acli (acyanogenic and not bearing any energetic costs of producing cyanogenic precursors), and disfavoring Acli and acLi plants (which bear some energetic costs of cyanogenesis but without the benefit of the chemical defense) (Ennos 1982; Kooyers & Olsen 2012). To test for evidence of epistatic selection, we compared our observed cyanotype frequencies to null expectations for two unlinked, independently-assorting polymorphisms. Consistent with the pattern expected under epistatic selection, the observed cyanotype frequencies showed a significant excess of "AcLi" and "acli" cyanotypes

and deficits of “acLi” and “Acli” cyanotypes (χ^2 = 18.29, df = 1, p = 1.897×10^{-5}) (**Fig. 3.4C**). Interestingly, this non-random association between the *Ac* and *Li* loci was evident not only from cyanotype frequencies, but also in the form of significantly elevated interchromosomal LD between SNPs linked to each cyanogenesis locus compared to background levels (**Supplementary Fig. S3.3B**). The locus-specific pattern of this interchromosomal LD rules out the possibility that the nonrandom association could be an artifact of population structure. Our inference of epistatic selection at the cyanogenesis loci is in line with some but not all previous studies in white clover (Ennos 1982; reviewed by Kooyers & Olsen 2012).

The repeated evolution of climate-associated cyanogenesis clines in white clover is a well-documented indicator of rapid environmental adaptation in this species (Daday, 1958; Kooyers & Olsen, 2012). In a recent study, we determined that the North American latitudinal cyanogenesis cline not only reflects population frequencies of cyanogenic genotypes (corresponding to PAVs at the *Ac* and *Li* loci), but that both *Ac* and *Li* alleles also show clinal variation in tandem gene copy number, with these CNVs contributing quantitatively to the strength of the cyanogenic phenotype (Kuo *et al.*, 2024). Based on this knowledge of CNV clines, we sought to determine which of the IBE-associated environmental variables identified above could best explain *Ac* and *Li* CNV distributions across the sampled locations. Taking into account both PAVs and CNVs, the cyanogenesis clines are best explained by *Growing Degree Days (> 5 ℃)* (**Fig. 3.4A, B**), with other temperature-related variables following closely (**Supplementary Fig. S3.13**). In contrast, precipitation-related variables (which have previously been reported to show associations with cyanogenesis clines; Kooyers & Olsen 2013; Kooyers et

al. 2014), exhibited considerably lower associations. Notably, *Growing Degree Days (> 5 ℃)* is the same environmental variable that best explains genome-wide IBE (**Supplementary Fig. S3.11; Fig. S3.13**); this suggests a very close association between the cyanogenesis cline distribution and overall patterns of environmental adaptation in white clover.

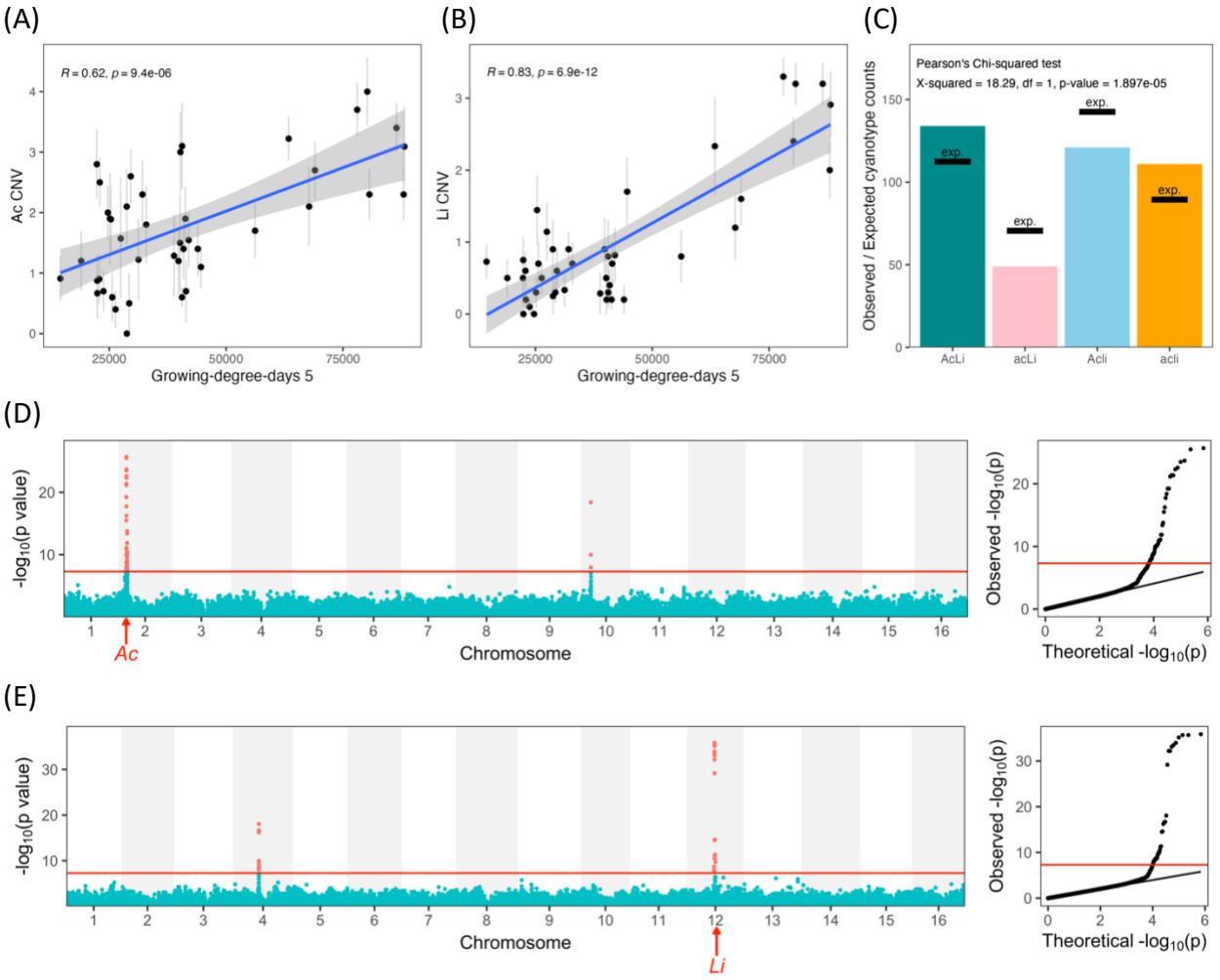


Figure 3.4 Cyanogenesis polymorphism associations with environmental and genomic variation.

(A-B), Associations between the copy number variation (CNV) of the cyanogenesis genes (*Ac*, *Li*) and *Growing Degree Days (> 5 ℃)*. **(C)**, observed and expected counts of the four cyanotypes among the 415 accessions. **(D)**, GWAS of the Ac phenotype (cyanogenic glucoside). **(E)**, GWAS of the Li phenotype (linamarase). Chromosomes 1-8 and 9-16 correspond to the two subgenomes derived from white clover's two diploid progenitors. The red lines show the *p* value threshold at 5×10^{-8}. Red arrows indicate the genomic locations of the *Ac* and *Li* loci.

Genome-wide scans for selection signals

The ability to pinpoint the genetic locus underlying trait variation is a crucial prerequisite for association analyses. To establish that our SNP dataset could be successfully applied in environmental association mapping, we first used the cyanogenesis phenotype data from our samples to map the underlying *Ac* and *Li* loci. Consistent with their known genomic locations (Olsen et al. 2021; Kuo et al. 2024), the *Ac* locus was successfully mapped to Chr. 2 at 8.5-9.0 Mbp, and the *Li* locus was mapped to Chr. 12 at 28 Mbp (**Fig 3.4D, E**; **Supplementary Fig. S3.14**). Additional peaks were detected in the homeologous positions of each cyanogenesis gene, which we attribute to highly similar homeologous sequences in these subgenome regions (see Kuo et al. 2024).

Having established the feasibility of association mapping with our dataset, we next performed landscape genomic analyses to identify genomic regions associated with local environmental adaptation. We focused on *Growing Degree Days (> 5°C)* based on the IBE results above. In a GEA analysis controlling for population structure (K = 3 in LFMM, based on ADMIXTURE results), we identified 442 out of 345,762 SNPs (0.1278%) that exhibited significant associations ($q < 0.05$) (**Fig. 3.5B**). In the genome-wide differentiation scan to identify SNPs that significantly diverge in the context of population structure (PCadapt), we detected 213 out of 241,371 SNPs (0.08825%) as significant outliers (**Fig. 3.5C**). Genomic regions jointly identified in the two landscape genomic analyses are located at the bottom of Chr. 1, bottom of Chr. 7, middle of Chr. 13, and bottom of Chr. 15 (discussed below).

Since the GWAS for the cyanogenesis phenotype and the GEA shared the same underlying SNP dataset, we could directly compare them to test for signal overlaps at the cyanogenesis loci. Notably, neither *Ac* nor *Li* showed significant associations in the GEA, although SNPs linked to the *Ac* locus on Chr. 2 exhibited a detectable peak that marginally approached significance (**Fig. 3.5B**). This result potentially suggests that cyanogenesis variation is not a major determinant of local adaptation. However, it is important to recognize that the lack of statistical significance at the cyanogenesis loci could be partially attributable to the strong collinearity between the cyanogenesis cline and population structure (**Fig. 3.1, Supplementary Fig. S3.15**); by controlling for population structure (setting K = 3 in LFMM), the GEA would have diminished power to detect these collinear adaptive loci (Lotterhos & Whitlock, 2015; Hoban *et al.*, 2016; Yoder & Tiffin, 2017). Therefore, to assess this possibility, we re-ran the GEA without any population structure control (K = 1) and at K=2. Both the *Ac* and *Li* loci exhibited clear peaks in these re-analyses; however, neither locus reached the significance threshold whereas multiple other genomic regions were highly significant (**Supplementary Fig. S3.16**). These results remained unchanged even after excluding of samples from Washington state (BWA) and Vancouver (VBC), where the environmentally distinct climate could potentially dampen signal detection (**Supplementary Fig. S3.17**). Together these findings suggest that the cyanogenesis polymorphism, while spatially distributed in a pattern consistent with local adaptation, is not a primary determinant of local adaptation in white clover.

If outlier SNPs in landscape genomic scans reflect true signals of local adaptation, many of these SNPs would be expected to occur within causal genes or to be linked to them. To

assess these possibilities, we manually investigated all protein-coding genes located within 25 Kbp of the outlier SNPs from the GEA and PCadapt output based on gene annotations in the reference genome (Kuo *et al.*, 2024). In the significant outlier intervals of the GEA and the genome-wide differentiation scan, there were 491 genes and 321 genes, respectively. After filtering for potential functions related to light response, developmental processes, stress, or pathogen response, 21 genes and 11 genes remained, respectively (**Fig. 3.5**). Among these, 8 genes were identified in both analyses; these include several with functions related to flowering time and other developmental and stress response processes that are plausible candidates for local adaptation. Detailed gene information is provided in the supplementary information (**Supplementary Table S3.3)**.

Finally, we compared the results of the GEA and the genomic differentiation scan to the previously identified QTLs for local adaptation in white clover, which were identified in reciprocal common garden experiments conducted in the northern, central and southern United States (Wright *et al.*, 2022). The fitness-related QTLs were categorized into vegetative growth traits (leaf area), survival (mortality) and reproductive output (flowering duration, floral count). We found that a subset of the regions identified through the landscape genomic analyses overlap these previously-identified fitness QTLs (**Fig. 3.5**). Among the overlapping regions, some contain candidate genes that could be playing functional roles in local adaptation. For example, a region at the bottom of Chr. 15 overlaps a major QTL for reproductive fitness traits and also contains candidate genes for flowering time and other

developmental processes that were identified in both the GEA and differentiation scan analysis (e.g., *MYB13*, *FER*; see **Supplementary Table S3.3**).

Notably, the three analyses presented in **Fig. 3.5** all have the shared goal of searching for adaptive genetic variation, but are based on different underlying hypotheses and empirical data. Candidate loci at the intersection of the three analyses may therefore be of particular interest for characterizing the molecular basis of local adaptation in white clover. Loci identified in only one of the analyses could also be important contributors to local adaptation. By comparison, the *Ac* and *Li* cyanogenesis loci, which are not present in any of these genomic regions of interest, appear to play a secondary role at best in determining local adaptation in the sampled white clover populations.

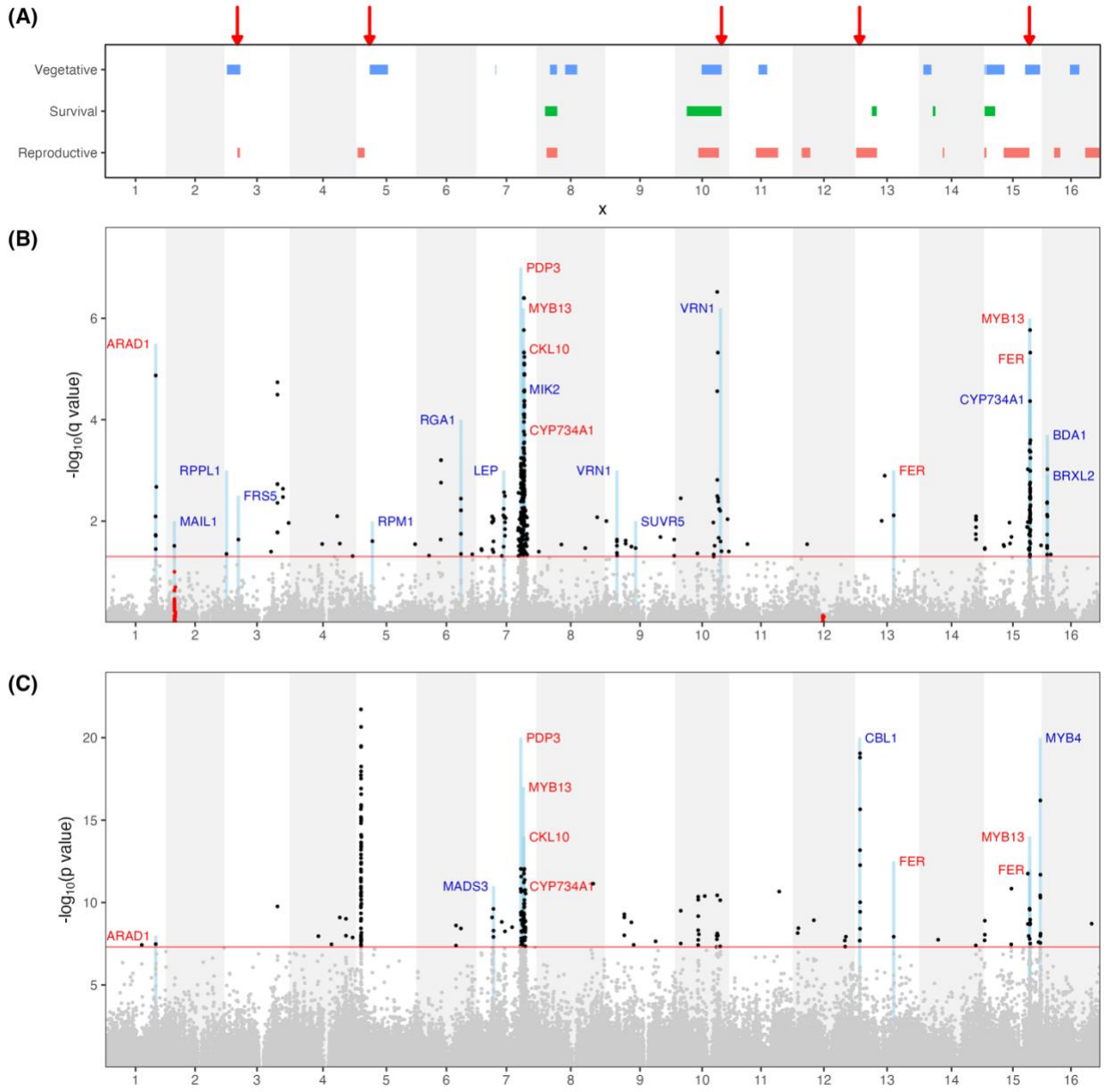


Figure 3.5 Genome-wide scans for selection and relationship to previously identified fitness QTLs.

(A), QTL mapping of the fitness traits from a reciprocal common garden experiment (Wright *et al.*, 2022). QTL intervals correspond to the 1-LOD drop. The traits are colored by fitness categories. The arrows on the top indicate that the QTL intervals overlap with the candidate genes detected in the LFMM or the PCadapt analyses. **(B)**, LFMM test for association with *Growing Degree Days (> 5 ℃)*. The confounding effect of the population structure was controlled by setting the latent factor k = 3. SNPs linked to the *Ac* and *Li* loci are colored in red

(Chr. 2 and 12, respectively). **(C)**, PCadapt for genetic differentiation outlier scan. The first two PCs were included in the analysis. The *Ac* and *Li* SNPs were not labeled because PCadapt used a LD pruned dataset, which is different from the one used in GWAS. The candidate genes (related to light response, developmental process, stress or pathogen response) within 25 Kbp of the outlier SNPs are labeled. If the genes are shown in both LFMM and PCadapt results, the color is red (otherwise, blue). Note that some genes are duplicated at the proximal physical locations but only labeled once, including *RPPL1* on Chr. 3, *VRN1* on Chr. 9, *CYP734A1* on Chr. 15, and *BDA1* on Chr. 16. See **Supplementary Table S3.3** for detailed candidate gene information.

3.6 Discussion

Repeatedly-evolved clines across environmental gradients are considered a hallmark of local adaptation. However, the relative importance of clinally-varying phenotypes, compared to other potential contributors to local adaptation, is seldom evaluated. The present study was undertaken with the goal of bridging this gap, utilizing the well-studied cyanogenesis polymorphism in white clover, which has repeatedly evolved climate-associated cyanogenesis clines worldwide. In our analyses of wild samples covering much of the North American species range, we find clear evidence for local environmental adaptation. Genome-wide differentiation across the sampled locations is better explained by temperature-related climatic variables than by geographic distance alone, with *Growing Degree Days (> 5°C)* identified as the most closely correlated environmental variable (**Fig. 3.3A, B**; **Supplementary Fig. S3.11**). This same climatic variable is also strongly correlated with cyanogenesis clinal variation (**Supplementary Fig. S3.13**). However, landscape genomic analyses indicate only a negligible contribution for cyanogenesis variation and its underlying genes in local adaptation; instead, multiple other genomic regions are implicated (**Fig. 3.5**). This finding aligns with our earlier results from fitness QTL mapping and common garden experiments (Wright *et al.*, 2018; Wright *et al.*, 2022), where we also found no significant support for cyanogenesis variation as a determinant of local climatic adaptation. Below we explore these results further, considering them in context of the white clover system and their broader implications for understanding the genetic basis of local adaptation.

Genomic insights on local adaptation in North American white clover

White clover, widely grown as a forage and rotation crop in temperate climates because of its remarkable nitrogen fixation capabilities, was introduced worldwide from its native range in Europe beginning in the 17th century; it is now a common cultivated and naturalized species throughout much of the world (Zeven, 1991; Kjærgaard, 2003; Taylor, 2008; Santangelo *et al.*, 2022). In North America, where its presence dates to European colonization (and its abundance was already noted by the mid-18th century; Andrae & Hancock, 2016), studies of wild populations have revealed clear evidence of local adaptation following its introduction; these include multiple studies documenting latitudinal and elevational cyanogenesis clines that are correlated with temperature gradients (Ganders, 1990; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013), as well as studies that have directly measured fitness traits in common garden experiments (Wright *et al.*, 2018; Wright *et al.*, 2022). The present study, by providing a complementary landscape genomics perspective, allows for several additional insights into the factors shaping the population structure and local adaptation in these North American populations.

Genetic admixture across North American populations. First, with respect to population structure, North American white clover shows minimal evidence of geographical structuring into discrete genetic subpopulations. Instead, genetic admixture predominates across the sampled locations (**Fig. 3.2A**), with the vast majority of the total genetic variation (>97%) distributed within sampling locations rather than between them (**Supplementary Table S3.2**). Consistent with these patterns, the level of genetic differentiation among sampling locations is

very low overall (median pairwise F_{ST} = 0.034). Together these observations suggest that barriers to gene flow are low across the sampled range. This inference is consistent with white clover's obligately-outcrossing mating system and its abundance in lawns and other mowed and grazed areas, which creates a near-continuous distribution across much of the continent (U.S. Department of Agriculture, 2024).

The one region where we detect some evidence of population substructure is in the southernmost sampling locations (Texas, Louisiana, Florida and southern Georgia). These samples form a homogeneous genetic subpopulation at the optimal K=3 in ADMIXTURE analyses and at higher assumed K values (**Fig. 3.2A**; **Supplementary Fig. S3.5**); they also form a distinct cluster in the PCA (although on an axis that explains < 3% of the total genetic variance) (**Fig. 3.2B**), and they collectively show higher levels of genetic differentiation from other samples in F_{ST} analyses (**Supplementary Fig. S3.7**). The slight but detectable genetic distinctness of these southernmost samples could reflect the remnants of a genetically distinct ancestral subpopulation in the native species range that contributed differentially to the establishment southern US populations. Comparative population structure analyses with samples from native European range could be useful to test this hypothesis.

Genetic differentiation reflects Isolation-by-environment (IBE). While genome-wide SNPs reveal a clear pattern of geographical isolation-by-distance (IBD) across the sampled locations (**Fig. 3.3A**), genetic differentiation is better explained as a function of temperature-related climatic variables than geography alone. This suggests that latitudinally related climatic

variation acts as stronger barrier to the movement of some alleles across the landscape than geographical distance. Systems characterized by extensive gene flow across a large-scale, continuous environmental gradient provide ideal conditions for the emergence of local adaptation (Wang & Bradburd, 2014). This process of local adaptation can in principle involve either new mutations arising *in situ*, with limited gene flow between climatically distinct locations (Orsini *et al.*, 2013), or selection acting on standing genetic variation that then becomes sorted geographically (Fay & Wu, 2000). As an introduced species with a history of <500 years in North America, the latter mechanism is likely most important for white clover. White clover populations worldwide are characterized by high genetic diversity, reflecting its history of intentional, repeated introductions globally as a nitrogen-fixing forage and lawn plant (Santangelo *et al.*, 2022). This demographic history would facilitate the introduction of standing variants which could then become sorted by local environment. Consistent with this selective scenario, our previous analyses of the *Ac/ac* and *Li/li* cyanogenesis polymorphisms have shown that cyanogenesis clines in introduced world regions have arisen through selection on pre-existing alleles that are found in the native species range (Olsen *et al.*, 2013; Kooyers & Olsen, 2014).

The specific climatic variable that we find best describes IBE in white clover, *Growing Degree Days (> 5°C)*, is a common measure of the point at which the cumulative temperature in spring is warm enough for the start of the growing season. This suggests that phenological response to local seasonal cues likely plays a critical role in determining locally-adaptive fitness in white clover. Notably, this finding mirrors our previous inferences from fitness data collected

in common garden experiments in the northern, central and southern US (Wright *et al.*, 2018; Wright *et al.*, 2022), which also point to phenological timing as a key factor in local adaptation (discussed below).

Insights on cyanogenesis cline evolution. Climate-associated cyanogenesis clines have been extensively documented and studied in native and introduced populations of white clover for more than 70 years (Daday, 1954b; Daday, 1954a; Daday, 1958; Ennos, 1982; Ganders, 1990; Kooyers & Olsen, 2012; Kooyers & Olsen, 2013). This body of work has generated a wealth of information on the evolution and ecology of this adaptive chemical defense polymorphism. While the findings of the present study do not implicate the cyanogenesis polymorphism as a key factor in local climatic adaptation in white clover, they do nonetheless provide two important new insights into the cyanogenesis polymorphism and cline evolution.

First, taking into account the full range of molecular variation that underlies the cyanogenesis polymorphism (both the PAVs that underlie the classic *Ac/ac* and *Li/li* polymorphisms, and the tandem CNVs present within 'dominant' alleles at each locus; **Supplementary Table S3.1**; Kuo et al. 2024), we find that the geographical distribution of cyanogenesis clines is best described by the same environmental variable that best describes genome-wide differentiation, i.e., *Growing Degree Days (>5°C)* (**Supplementary Tables S3.10-S3.12**). Previous studies of cyanogenesis clines have generally reported correlations with measures related to winter temperature (e.g., minimum winter temperature [MWT]; mean temperature of the coldest quarter [MTCQ]) (e.g., Kooyers & Olsen, 2012; Kooyers & Olsen,

2013; Innes *et al.*, 2022; Kuo *et al.*, 2024), a period of dormancy and potential cold-related winter mortality. Our findings here suggest that rather than winter cold stress, it is instead selective factors related to the active growing season that underlie cyanogenesis cline evolution. *Growing Degree Days* (> 5°C) is a common estimate of insect pest life cycle (Bonhomme, 2000; Cayton *et al.*, 2015); thus, it is possible that this variable captures the level of herbivore pressure in a local environment that would favor the cyanogenesis chemical defense — especially in the early growing season, when young shoots are most vulnerable (Wright *et al.*, 2018). Our inference that cyanogenesis clines reflect selection during the warmer growing season rather than winter cold stress is also in line with recent work that has found no support for an older 'autotoxicity' hypothesis, which had proposed that freezing-induced cyanide toxicity might differentially select against cyanogenic genotypes in colder climates (Daday, 1965; Kooyers *et al.*, 2018; Kuo *et al.*, 2023)

A second inference on the cyanogenesis polymorphism is that we find clear evidence for epistatic selection as a factor shaping the cyanogenesis polymorphism in natural populations. Whereas the cyanogenic phenotype requires the presence of both cyanogenic glucosides and linamarase within a plant, neither of the individual components is known to serve any function other than in cyanogenesis (although cyanogenic glucosides could function as N storage compounds) (see Gleadow & Moller, 2014); at the same time, production of both cyanogenic components is likely energetically costly (Daday, 1965; Kakes, 1989; Kooyers *et al.*, 2018). This creates selective scenario whereby plants that produce only one of the two components could be maladaptive, since they carry some of the energetic costs of cyanogenesis but without any

benefits of the chemical defense. Previous studies of white clover populations in North America and elsewhere have found mixed evidence for epistatic selection (Ennos, 1982) (reviewed by Kooyers & Olsen, 2012). Here, with a large, representative sample of North American genotypes (N=415), we observe significant deficits in frequencies of Acli and acLi plants, and a corresponding excess of AcLi and acli plants (**Fig. 3.3C**), consistent with a model of balancing epistatic selection that favors either cyanogenic plants or those that lack both components. Unlike previous studies, which relied solely on cyanotype frequencies to test for epistatic selection, we have further documented LD between the *Ac* and *Li* loci using genome-wide SNP data — which reveal significant interchromosomal LD specifically for the neutral SNPs that are linked to *Ac* and *Li* (**Supplementary Fig. S3.3B**). These data not only provide compelling evidence for a model of epistatic balancing selection at the cyanogenesis loci (and rule out any possibility that the nonrandom association between *Ac* and *Li* could be an artifact of population structure), but they also add white clover to a growing list of species with genomic evidence of interchromosomal LD resulting from epistatic selection (e.g., Hench *et al.*, 2019; Gupta *et al.*, 2023).

Landscape genomic insights on local adaptation in white clover. Our GEA and genome-wide differentiation scans do not indicate that selection on the cyanogenesis polymorphism plays a significant role in local climatic adaptation. While *Ac*-linked SNPs did show a detectable peak below the significance threshold in the GEA , there was no visible signal at *Li* locus **(Fig. 3.5B)**; moreover, neither cyanogenesis locus reached significance even after removing corrections for population structure, which can undermine the power of detection in a GEA

(Lotterhos & Whitlock, 2015; Hoban *et al.*, 2016; Yoder & Tiffin, 2017) (**Supplementary Fig. S3.16, Fig. S3.17**). This lack of evidence that cyanogenesis plays a key role in local climatic adaptation is consistent with our previous inferences based on common garden fitness measurements using wild ecotypes and F_2 mapping populations (Wright *et al.*, 2018; Wright *et al.*, 2022).

While not implicating cyanogenesis as critical for local adaptation, the landscape genomic analyses do provide valuable insights into the genomic mechanisms of local adaptation in this climatically widespread species. First, it is likely that a relatively few loci are driving this process. For species such as white clover, characterized by outcrossing mating systems, large population sizes and extensive gene flow, local adaptation is expected to evolve though selection on a relatively few genes of major effect (Yeaman *et al.*, 2023). In line with this theoretical prediction, less than one-half of 1% of the total SNPs in the GEA scan (0.1278%) and less than one-tenth of 1% of SNPs in the PCadapt analysis (0.08825%) exhibited significant signatures of local adaptation.

As for the identity of the targets of selection, we hypothesize that selection for local climatic adaptation has acted on genes with functions related to the phenological response to local seasonal cues during the active growing season. Candidate genes within the significantly associated genomic regions include, for example, the photoperiod regulatory genes ***FRS5*** (Lin & Wang, 2004; Li *et al.*, 2011; Ma & Li, 2018) and ***CKL10*** (Tan *et al.*, 2013), as well the temperature sensing gene ***VRN1***, which functions in vernalization response (Sheldon *et al.*,

2000; Levy *et al.*, 2002). Additionally, the candidates include genes that function upstream of the flowering time regulator ***FLC***, including ***PDP3*** (Zhou *et al.*, 2018) and ***FER*** (Wang *et al.*, 2020) (**Supplementary Table S3.3**). Consistent with this hypothesis, previous results from our common garden and fitness QTL mapping studies indicated that local climatic adaptation across the US is closely tied to loci controlling the relative timing of vegetative growth *vs*. flowering over successive seasons (Wright *et al.*, 2018; Wright *et al.*, 2022). Common garden observations indicated that southern US populations experience favorable growing conditions very early in the spring, followed by a period of dormancy or mortality during the hot summer months; northern climates, in contrast, experience a later initiation of favorable growth conditions but are able to continue vegetative growth and flowering later into the summer. These climatic differences appear to favor genotypes with early investment in flowering in southern climates, and those with later flowering (after an extended period of vegetative growth) further north (Wright *et al.*, 2022). Similar to these observations for white clover, research in diverse other plant species has also pointed to connections between latitudinal or elevational temperature gradients and phenology (most notably flowering time) in the process of local adaptation, which can potentially lead to population structure and even the emergence of reproductive isolating barriers among adaptively diverged populations (Stinchcombe *et al.*, 2004; De Frenne *et al.*, 2013; Toftegaard *et al.*, 2015; Bucher *et al.*, 2018).

While our study provides important insights into the genomic basis of local adaptation in white clover, some limitations in the data should be noted. First, the genotyping-by-sequencing (GBS) protocol, while cost-effective, is a non-randomized and reduced

representative whole-genome re-sequencing approach. Although it provides an efficient method for genotypic information in non-model organisms, it compromises resolution for detecting causal variations in coding or regulatory sequences and is prone to high missing rates in variant callings among samples. As population-level pangenome analyses become cost-effective in the future, clearer insights on the causal basis of local adaptation in this species will be possible. Similarly, expanded phenotypic analyses beyond cyanogenesis (the sole phenotype examined in the present study) would substantially enhance the depth and resolution of our understanding of the key traits involved, such as growth rate, flowering time, and stress responses. Future investigations in white clover and other widespread plant species will benefit from the incorporation of both of these approaches.

3.7 Acknowledgements

We thank the high school Biology teachers, plant scientists and others who participated in the crowd-sourced collection of wild white clover accessions across North America, including Miranda Colletta, Sunita Crittenden, Felix Fritschi, Briana Gross, Nic Kooyers, Lisa Limeri, Mary Lyman-Onkka, Ed McAssey, John McDonald, John McIntire, Maria Monteros, Jim Mullen, Patricio Munoz, Sandra Pelc, Anne Puzzo, Nicole Riddle, Trisha Spanbauer, Daniel Tabor, Rebecca Thomson, Kathryn Turner, Cynthia Vigueira, Kate Waselkov, Marshall Wedger, and Dan Cui Zhou. We especially thank Alexander Mahmoud (undergraduate student at Washington University in St. Louis) for processing wild accessions and maintaining them in greenhouse. We also thank Mike Dyer and other staff of the Washington University in St. Louis greenhouse

facility for providing plant care. Finally, we thank the members of the Olsen lab group who provided valuable comments on earlier versions of the manuscript.

3.8 Author Contribution

Wen-Hsi Kuo and Limei Zhong analyzed the data, interpreted the results, and wrote the manuscript. Wen-Hsi Kuo and Limei Zhong contributed equally to this work. Sara J. Wright designed the experiments, collected and maintained the wild accessions in greenhouse, phenotyped the cyanogenesis trait, and prepared and sequenced the GBS library. David M. Goad helped with the bioinformatic analyses and provided suggestions at the early stage of the project. Kenneth M. Olsen conceived and designed the project, interpreted the results, and edited the manuscript. All authors read and approved the final manuscript.

3.9 Data Accessibility

Raw GBS sequence reads are deposited in NCBI SRA (BioProject: PRJNA1014483).

3.10 Reference

Agrena J, Oakley CG, McKay JK, Lovell JT, Schemske DW. 2013. Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **110**(52): 21077-21082.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**(9): 1655-1664.

Andrae J, Hancock D. 2016. University of Georgia. College of A, Environmental S, eds. White clover establishment and management guide: University of Georgia.

Antoniazza S, Burri R, Fumagalli L, Goudet J, Roulin A. 2010. Local adaptation maintains clinal variation in melanin-based coloration of european barn owls (*Tyto alba*). *Evolution* **64**(7): 1944-1954.

Battlay P, Wilson J, Bieker VC, Lee C, Prapas D, Petersen B, Craig S, van Boheemen L, Scalone R, de Silva NP, et al. 2023. Large haploblocks underlie rapid adaptation in the invasive weed *Ambrosia artemisiifolia*. *Nature Communications* **14**(1): 1717.

Bonhomme R. 2000. Bases and limits to using 'degree.day' units. *European Journal of Agronomy* **13**(1): 1-10.

Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics* **103**(3): 338-348.

Bucher SF, Konig P, Menzel A, Migliavacca M, Ewald J, Romermann C. 2018. Traits and climate are associated with first flowering day in herbaceous species along elevational gradients. *Ecology and Evolution* **8**(2): 1147-1158.

Caye K, Jumentier B, Lepeule J, Francois O. 2019. LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution* **36**(4): 852-860.

Cayton HL, Haddad NM, Gross K, Diamond SE, Ries L. 2015. Do growing degree days predict phenology across butterfly species? *Ecology* **96**(6): 1473-1479.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**: 7.

Daday H. 1954a. Gene frequencies in wild populations of *Trifolium repens* L. Distribution by latitude. *Heredity* **8**(1): 61-78.

Daday H. 1954b. Gene frequencies in wild populations of *Trifolium repens*. II. Distribution by altitude. *Heredity* **8**(3): 377-384.

Daday H. 1958. Gene frequencies in wild populations of *Trifolium repens* L. III. World distribution. *Heredity* **12**(2): 169-184.

Daday H. 1965. Gene frequencies in wild populations of *Trifolium repens* L. IV. Mechanism of natural selection. *Heredity* **20**(3): 355-365.

De Frenne P, Graae BJ, Rodríguez-Sánchez F, Kolb A, Chabrerie O, Decocq G, De Kort H, De Schrijver A, Diekmann M, Eriksson O, et al. 2013. Latitudinal gradients as natural laboratories to infer species' responses to temperature. *Journal of Ecology* **101**(3): 784-795.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**(6): 927-930.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**(5): e19379.

Ennos RA. 1982. Association of the cyanogenic loci in white clover. *Genetics Research* **40**: 65-72.

Fay JC, Wu C-I. 2000. Hitchhiking under positive darwinian selection. *Genetics* **155**(3): 1405-1413.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**(12): 4302-4315.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**(6052): 86-89.

Ganders FR. 1990. Altitudinal clines for cyanogenesis in introduced populations of white clover near Vancouver, Canada. *Heredity* **64**(3): 387-390.

Gefaell J, Vigo R, González-Vázquez AH, Galindo J, Rolán-Alvarez E. 2023. Temporal stability and directional change in a color cline of a marine snail from NW Spain. *Current Zoology*.

Germino MJ, Moser AM, Sands AR. 2019. Adaptive variation, including local adaptation, requires decades to become evident in common gardens. *Ecological Applications* **29**(2): e01842.

Gleadow RM, Moller BL. 2014. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annual Review of Plant Biology* **65**: 155-185.

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019. Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31**(7): 1466-1487.

Grillo MA, Li C, Hammond M, Wang L, Schemske DW. 2013. Genetic architecture of flowering time differentiation between locally adapted populations of *Arabidopsis thaliana*. *New Phytologist* **197**(4): 1321-1331.

Guerrero J, Andrello M, Burgarella C, Manel S. 2018. Soil environment is a key driver of adaptation in *Medicago truncatula:* new insights from landscape genomics. *New Phytologist* **219**(1): 378-390.

Gugger PF, Liang CT, Sork VL, Hodgskiss P, Wright JW. 2018. Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary Applications* **11**(2): 231-242.

Gupta S, Harkess A, Soble A, Van Etten M, Leebens-Mack J, Baucom RS. 2023. Interchromosomal linkage disequilibrium and linked fitness cost loci associated with selection for herbicide resistance. *New Phytologist*.

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

Hench K, Vargas M, Hoppner MP, McMillan WO, Puebla O. 2019. Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology and Evolution* **3**(4): 657-667.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. 2016. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist* **188**(4): 379-397.

Hof AEvt, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**(7605): 102-105.

Hughes MA. 1991. The cyanogenic polymorphism in *Trifolium repens* L. (white clover). *Heredity* **66**(1): 105-115.

Innes SG, Santangelo JS, Kooyers NJ, Olsen KM, Johnson MTJ. 2022. Evolution in response to climate in the native and introduced ranges of a globally distributed plant. *Evolution* **76**(7): 1495-1511.

Kakes P. 1989. An analysis of the costs and benefits of the cyanogenic system in *Trifolium repens* L. *Theoretical and Applied Genetics* **77**(1): 111-118.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**: e281.

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

Kjærgaard T. 2003. A plant that changed the world: The rise and fall of clover 1000-2000. *Landscape Research* **28**(1): 41-49.

Kooyers NJ, Gage LR, Al-Lozi A, Olsen KM. 2014. Aridity shapes cyanogenesis cline evolution in white clover (*Trifolium repens* L.). *Molecular Ecology* **23**(5): 1053-1070.

Kooyers NJ, Hartman Bakken B, Ungerer MC, Olsen KM. 2018. Freeze-induced cyanide toxicity does not maintain the cyanogenesis polymorphism in white clover (*Trifolium repens*). *American Journal of Botany* **105**(7): 1224-1231.

Kooyers NJ, Olsen KM. 2012. Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.). *Molecular Ecology* **21**(10): 2455-2468.

Kooyers NJ, Olsen KM. 2013. Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover (*Trifolium repens* L.). *Heredity* **111**(6): 495-504.

Kooyers NJ, Olsen KM. 2014. Adaptive cyanogenesis clines evolve recurrently through geographical sorting of existing gene deletions. *Journal of Evolutionary Biology* **27**(11): 2554-2558.

Koski MH, Galloway LF. 2018. Geographic variation in pollen color is associated with temperature stress. *New Phytologist* **218**(1): 370-379.

Kujala ST, Knürr T, Kärkkäinen K, Neale DB, Sillanpää MJ, Savolainen O. 2017. Genetic heterogeneity underlying variation in a locally adaptive clinal trait in *Pinus sylvestris* revealed by a Bayesian multipopulation analysis. *Heredity* **118**(5): 413-423.

Kuo W-H, Small LL, Olsen KM. 2023. Variable expression of cyanide detoxification and tolerance genes in cyanogenic and acyanogenic white clover (*Trifolium repens*). *American Journal of Botany* **110**(10): e16233.

Kuo W-H, Wright SJ, Small LL, Olsen KM. 2024. *De novo* genome assembly of white clover (*Trifolium repens* L.) reveals the role of copy number variation in rapid environmental adaptation. *Research Square*.

Lascoux M, Glémin S, Savolainen O 2016. Local adaptation in plants. *Encyclopedia of Life Sciences*, 1-7.

Lasky JR, Josephs EB, Morris GP. 2022. Genotype–environment associations to reveal the molecular basis of environmental adaptation. *The Plant Cell* **35**(1): 125-138.

Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. 2002. Multiple Roles of Arabidopsis *VRN1* in Vernalization and Flowering Time Control. *Science* **297**(5579): 243-246.

Li G, Siddiqui H, Teng Y, Lin R, Wan XY, Li J, Lau OS, Ouyang X, Dai M, Wan J, et al. 2011. Coordinated transcriptional regulation underlying the circadian clock in Arabidopsis. *Nature Cell Biology* **13**(5): 616-622.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Lin R, Wang H. 2004. Arabidopsis *FHY3*/*FAR1* gene family and distinct roles of its members in light control of Arabidopsis development. *Plant Physiology* **136**(4): 4010-4022.

Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009. On the origin and spread of an adaptive allele in deer mice. *Science* **325**(5944): 1095-1098.

Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* **24**(5): 1031-1046.

Ma L, Li G. 2018. FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) family proteins in Arabidopsis growth and development. *Frontiers in Plant Science* **9**.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**(1): 3.

Mullen LM, Hoekstra HE. 2008. Natural selection along an environmental gradient: A classic cline in mouse pigmentation. *Evolution* **62**(7): 1555-1570.

Olsen KM, Goad DM, Wright SJ, Dutta ML, Myers SR, Small LL, Li LF. 2021. Dual-species origin of an adaptive chemical defense polymorphism. *New Phytologist* **232**(3): 1477-1487.

Olsen KM, Hsu SC, Small LL. 2008. Evidence on the molecular basis of the *Ac*/*ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* **179**(1): 517-526.

Olsen KM, Kooyers NJ, Small LL. 2013. Recurrent gene deletions and the evolution of adaptive cyanogenesis polymorphisms in white clover (*Trifolium repens* L.). *Molecular Ecology* **22**(3): 724-738.

Olsen KM, Small LL. 2018. Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytologist* **219**(2): 757-766.

Olsen KM, Sutherland BL, Small LL. 2007. Molecular evolution of the *Li*/*li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Molecular Ecology* **16**(19): 4180-4193.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* **22**(24): 5983-5999.

Pebesma EJ. 2018. Simple features for R: standardized support for spatial vector data. *The R Journal* **10**(1): 439.

Poplin R, Ruano-Rubio V, Depristo MA, Fennell TJ, Carneiro MO, Van Der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*.

Price N, Moyers BT, Lopez L, Lasky JR, Monroe JG, Mullen JL, Oakley CG, Lin J, Ågren J, Schrider DR, et al. 2018. Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **115**(19): 5028-5033.

Privé F, Luu K, Vilhjálmsson BJ, Blum MGB. 2020. Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution* **37**(7): 2153-2154.

Santangelo JS, Battlay P, Hendrickson BT, Kuo W-H, Olsen KM, Kooyers NJ, Johnson MTJ, Hodgins KA, Ness RW. 2023. Haplotype-resolved, chromosome-level assembly of white clover (*Trifolium repens* L., Fabaceae). *Genome Biology and Evolution* **15**(8).

Santangelo JS, Ness RW, Cohan B, Fitzpatrick CR, Innes SG, Koch S, Miles LS, Munim S, Peres-Neto PR, Prashad C, et al. 2022. Global urban environmental change drives adaptation in white clover. *Science* **375**(6586): 1275-1281.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**(11): 807-820.

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES. 2000. The molecular basis of vernalization: The central role of *FLOWERING LOCUS C* (*FLC*). *Proceedings of the National Academy of Sciences of the United States of America* **97**(7): 3753-3758.

Stinchcombe JR, Weinig C, Ungerer M, Olsen KM, Mays C, Halldorsdottir SS, Purugganan MD, Schmitt J. 2004. A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proceedings of the National Academy of Sciences of the United States of America* **101**(13): 4712-4717.

Tan ST, Dai C, Liu HT, Xue HW. 2013. Arabidopsis casein kinase1 proteins CK1.3 and CK1.4 phosphorylate cryptochrome2 to regulate blue light signaling. *The Plant Cell* **25**(7): 2618-2632.

Taylor NL. 2008. A century of clover breeding developments in the united states. *Crop Science* **48**(1): 1-13.

Thompson KA, Renaudin M, Johnson MT. 2016. Urbanization drives the evolution of parallel clines in plant populations. *Proceedings of the Royal Society B: Biological Sciences* **283**(1845).

Title PO, Bemmels JB. 2018. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* **41**(2): 291-307.

Toftegaard T, Posledovich D, Navarro-Cano JA, Wiklund C, Gotthard K, Ehrlén J. 2015. Variation in plant thermal reaction norms along a latitudinal gradient – more than adaptation to season length. *Oikos* **125**(5): 622-628.

U.S. Department of Agriculture NRCS 2024. *Trifolium repens* L.

Wang H, Wu Y, He Y, Li G, Ma L, Li S, Huang J, Yang G. 2023. High-quality chromosome-level de novo assembly of the *Trifolium repens*. *BMC Genomics* **24**(1): 326.

Wang IJ, Bradburd GS. 2014. Isolation by environment. *Molecular Ecology* **23**(23): 5649-5662.

Wang L, Yang T, Lin Q, Wang B, Li X, Luan S, Yu F. 2020. Receptor kinase FERONIA regulates flowering time in Arabidopsis. *BMC Plant Biol* **20**(1): 26.

Wang Y, Zhang L, Zhou Y, Ma W, Li M, Guo P, Feng L, Fu C. 2023. Using landscape genomics to assess local adaptation and genomic vulnerability of a perennial herb *Tetrastigma hemsleyanum* (Vitaceae) in subtropical China. *Frontiers in Genetics* **14**: 1150704.

Whitlock R, Hipperson H, Mannarelli M, Burke T. 2008. A high-throughput protocol for extracting high-purity genomic DNA from plants and animals. *Molecular Ecology Resources* **8**(4): 736-741.

Wittkopp PJ, Smith-Winberry G, Arnold LL, Thompson EM, Cooley AM, Yuan DC, Song Q, McAllister BF. 2011. Local adaptation for body color in *Drosophila americana*. *Heredity* **106**(4): 592-602.

Wright SJ, Cui Zhou D, Kuhle A, Olsen KM. 2018. Continent-wide climatic variation drives local adaptation in North American white clover. *Journal of Heredity* **109**(1): 78-89.

Wright SJ, Goad DM, Gross BL, Muñoz PR, Olsen KM. 2022. Genetic trade-offs underlie divergent life history strategies for local adaptation in white clover. *Molecular Ecology* **31**(14): 3742-3760.

Yeaman S, Whiting J, Booker T, Rougeux C, Lind B, Singh P, Lu M, Huang K, Whitlock M, Aitken S. 2023. Core genes driving climate adaptation in plants. *Research Square*.

Yoder JB, Tiffin P. 2017. Effects of gene action, marker density, and timing of selection on the performance of landscape genomic scans of local adaptation. *Journal of Heredity* **109**(1): 16-28.

Zeven AC. 1991. Four hundred years of cultivation of Dutch white clover landraces. *Euphytica* **54**(1): 93-99.

Zhou JX, Liu ZW, Li YQ, Li L, Wang B, Chen S, He XJ. 2018. Arabidopsis PWWP domain proteins mediate H3K27 trimethylation on FLC and regulate flowering time. *Journal of Integrative Plant Biology* **60**(5): 362-368.

3.11 Supporting Information

Supplementary Table S3.1 Accession information

Accession	Date	Location	Latitude	Longitude	Nearest City	Collector	Ac CNV †	Li CNV †
AIA_001	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	2	2
AIA_002	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	7	0
AIA_003	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	0	0
AIA_004	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	0	0
AIA_007	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	4	0
AIA_008	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	3	0
AIA_009	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	0	0
AIA_010	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	3	0
AIA_011	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	0	1
AIA_012	2015-10	AIA	42.046667	-93.621111	Ames, IA	Sara Wright	0	0
ANC_001	2015-07	ANC	35.566944	-82.570833	Asheville, NC	Rebecca Thomson	5	2
ANC_002	2015-07	ANC	35.56833	-82.568333	Asheville, NC	Rebecca Thomson	2	0
ANC_004	2015-07	ANC	35.578056	-82.581667	Asheville, NC	Rebecca Thomson	0	0
ANC_005	2015-07	ANC	35.578056	-82.581944	Asheville, NC	Rebecca Thomson	1	0
ANC_006	2015-07	ANC	35.577778	-82.561389	Asheville, NC	Rebecca Thomson	1	0
ANC_007	2015-07	ANC	35.5775	-82.55422	Asheville, NC	Rebecca Thomson	0	0
ANC_008	2015-07	ANC	35.566389	-82.573333	Asheville, NC	Rebecca Thomson	0	0
ATX_001	2017-04	ATX	30.285951	-97.734002	Austin, TX	Sara Wright	4	3
ATX_002	2017-04	ATX	30.285951	-97.734002	Austin, TX	Sara Wright	4	4
ATX_004	2017-04	ATX	30.285951	-97.734002	Austin, TX	Sara Wright	4	2
ATX_005	2017-04	ATX	30.285951	-97.734002	Austin, TX	Sara Wright	5	2
ATX_006	2017-04	ATX	30.285951	-97.734002	Austin, TX	Sara Wright	2	2
ATX_007	2017-04	ATX	30.285951	-97.734002	Austin, TX	Sara Wright	5	4
ATX_008	2017-04	ATX	30.285951	-97.734002	Austin, TX	Sara Wright	3	4
ATX_009	2017-04	ATX	30.285706	-97.730922	Austin, TX	Sara Wright	3	4
ATX_015	2017-04	ATX	30.285175	-97.729543	Austin, TX	Sara Wright	3	4
ATX_016	2017-04	ATX	30.285175	-97.729543	Austin, TX	Sara Wright	1	3
BAL_005	2015-07	BAL	34.056278	-86.762583	Hanceville, AL	Nicole Riddle	4	4
BAL_013	2015-07	BAL	34.0725	-86.781944	Hanceville, AL	Nicole Riddle	0	0
BAL_018	2015-07	BAL	33.501772	-86.774156	Homewood, AL	Nicole Riddle	0	0
BAL_019	2015-07	BAL	33.4760152	-86.819773	Birmingham, AL	Nicole Riddle	0	0
BAL_020	2015-07	BAL	33.487966	-86.817766	Birmingham, AL	Nicole Riddle	5	0

BAL_021	2015-07	BAL	33.4931613	-86.796101	Birmingham, AL	Nicole Riddle	3	2
BAL_024	2015-07	BAL	33.438553	-86.837503	Birmingham, AL	Nicole Riddle	2	2
BAL_027	2015-07	BAL	33.51721	-86.766378	Birmingham, AL	Nicole Riddle	0	0
BAL_031	2015-07	BAL	33.500044	-86.807627	Birmingham, AL	Nicole Riddle	2	2
BAL_043	2015-07	BAL	33.455556	-86.628056	Birmingham, AL	Nicole Riddle	5	2
BID_001	2014-09	BID	43.596111	-116.22222	Boise, ID	Sara Wright	0	0
BID_002	2014-09	BID	43.596111	-116.22222	Boise, ID	Sara Wright	0	1
BID_003	2014-09	BID	43.596111	-116.22222	Boise, ID	Sara Wright	0	1
BID_004	2014-09	BID	43.596111	-116.22222	Boise, ID	Sara Wright	5	0
BID_005	2014-09	BID	43.596111	-116.22222	Boise, ID	Sara Wright	0	0
BID_006-2	2014-09	BID	43.596111	-116.22222	Boise, ID	Sara Wright	0	0
BID_007-2	2014-09	BID	43.595556	-116.22222	Boise, ID	Sara Wright	0	0
BID_008-2	2014-09	BID	43.595556	-116.22222	Boise, ID	Sara Wright	0	0
BID_010	2014-09	BID	43.595556	-116.22222	Boise, ID	Sara Wright	0	1
BID_011	2014-09	BID	43.595556	-116.22222	Boise, ID	Sara Wright	0	0
BIL_002	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	2	0
BIL_003	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	0	0
BIL_004	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	0	0
BIL_005	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	4	2
BIL_007	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	5	0
BIL_008	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	0	1
BIL_009	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	0	0
BIL_011	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	0	0
BIL_012	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	1	0
BIL_016	2015-08	BIL	40.517778	-88.953056	Bloomington-Normal, IL	Sara Wright	0	0
BKS_001	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	2	0
BKS_003	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	0	0
BKS_004	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	5	1
BKS_005	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	2	0
BKS_006	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	3	0
BKS_007	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	0	0
BKS_008	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	3	0

BKS_009	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	1	0
BKS_010	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	0	0
BKS_012	2015-11	BKS	39.088611	-94.880556	Bonner Springs, KS	Sara Wright	3	1
BMD_001	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	0	0
BMD_002	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	0	0
BMD_003	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	0	0
BMD_005	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	4	2
BMD_006	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	0	0
BMD_007	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	3	0
BMD_008	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	2	1
BMD_009	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	0	0
BMD_011	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	3	1
BMD_012	2014-09	BMD	39.3313	-76.58768	Baltimore, MD	John McDonald	2	0
BSD_001	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	4	0
BSD_002	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	2	0
BSD_003	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	3	0
BSD_004	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	2	0
BSD_005	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	3	0
BSD_006	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	2	0
BSD_007	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	0	0
BSD_008	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	0	0
BSD_009	2014-09	BSD	43.853611	-101.20361	Belvidere, SD	Sara Wright	2	0
BWA_001	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	3	1
BWA_002	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	0	2
BWA_005	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	2	1
BWA_006	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	0	1
BWA_007	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	2	0
BWA_008	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	4	0
BWA_009	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	6	0
BWA_010	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	2	2
BWA_011	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	2	1
BWA_012	2015-06	BWA	48.734	-122.486	Bellingham, WA	Jim Mullen	2	1
CIL_001	2014-08	CIL	41.962222	-87.633333	Chicago, IL	Sara Wright	0	0
CIL_003-2	2014-08	CIL	41.963056	-87.634444	Chicago, IL	Sara Wright	2	1
CIL_004	2014-08	CIL	41.963056	-87.635278	Chicago, IL	Sara Wright	0	4
CIL_005	2014-08	CIL	41.963056	-87.632778	Chicago, IL	Sara Wright	3	3
CIL_006	2014-08	CIL	41.963333	-87.631667	Chicago, IL	Sara Wright	0	0
CIL_007	2014-08	CIL	41.963333	-87.634167	Chicago, IL	Sara Wright	0	0

CIL_008	2014-08	CIL	41.963333	-87.634167	Chicago, IL	Sara Wright	3	2
CIL_009	2014-08	CIL	41.963611	-87.631389	Chicago, IL	Sara Wright	1	0
CIL_010	2014-08	CIL	41.963889	-87.632778	Chicago, IL	Sara Wright	4	1
CIL_011	2014-08	CIL	41.963889	-87.633611	Chicago, IL	Sara Wright	4	2
CMO_004	2015-11	CMO	38.962766	-92.25064	Columbia, MO	Sara Wright	4	0
CMO_005	2015-11	CMO	38.962766	-92.25064	Columbia, MO	Sara Wright	5	0
CMO_006	2015-11	CMO	38.962766	-92.25064	Columbia, MO	Sara Wright	2	2
CMO_007	2015-11	CMO	38.962766	-92.25064	Columbia, MO	Sara Wright	6	0
CMO_008	2015-11	CMO	38.962766	-92.25064	Columbia, MO	Sara Wright	2	0
CMO_010	2015-11	CMO	38.962766	-92.25064	Columbia, MO	Sara Wright	0	0
CMO_012	2015-11	CMO	38.962766	-92.25064	Columbia, MO	Sara Wright	4	0
CMO_013	2016	CMO	38.9379	-92.295	Columbia, MO	Felix Fritschi	5	0
CMO_014	2016	CMO	38.8397	-92.295	Columbia, MO	Felix Fritschi	0	0
CMO_015	2016	CMO	38.9395	-92.2954	Columbia, MO	Felix Fritschi	2	0
CSC_001	2015-06	CSC	32.795278	-80.108889	Charleston, SC	Sandra Pelc	4	2
CSC_005	2015-06	CSC	32.7925	-80.1125	Charleston, SC	Sandra Pelc	4	0
CSC_006	2015-06	CSC	32.795278	-80.080278	Charleston, SC	Sandra Pelc	4	3
CSC_007	2015-06	CSC	32.795556	-80.081944	Charleston, SC	Sandra Pelc	4	3
CSC_008	2015-06	CSC	32.796111	-80.109444	Charleston, SC	Sandra Pelc	6	2
CSC_009	2015-06	CSC	32.793056	-80.108056	Charleston, SC	Sandra Pelc	1	3
CSC_010	2015-06	CSC	32.793611	-80.109722	Charleston, SC	Sandra Pelc	1	2
CSC_011	2015-06	CSC	32.793333	-80.111944	Charleston, SC	Sandra Pelc	6	3
CSC_012	2015-06	CSC	32.795833	-80.0825	Charleston, SC	Sandra Pelc	5	4
CSC_014	2015-06	CSC	32.795278	-80.081944	Charleston, SC	Sandra Pelc	5	2
CVA_001	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	2	2
CVA_002	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	2
CVA_003	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	0
CVA_004	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	0
CVA_005	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	5	0
CVA_006	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	0
CVA_007	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	1
CVA_010	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	0
CVA_011	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	2
CVA_012	2014-09	CVA	38.010224	-78.5171	Charlottesville, VA	Nic Kooyers	0	0
DCO_002-2	2014-09	DCO	39.754444	-104.90306	Denver, CO	Rebecca Thomson	0	0

DCO_003	2014-09	DCO	39.754444	-104.90306	Denver, CO	Rebecca Thomson	0	0
DCO_005	2014-09	DCO	39.764444	-104.92222	Denver, CO	Rebecca Thomson	3	0
DCO_006	2014-09	DCO	39.841944	-105.00472	Denver, CO	Rebecca Thomson	0	2
DCO_008	2014-09	DCO	39.764167	-104.92056	Denver, CO	Rebecca Thomson	0	1
DCO_009	2014-09	DCO	39.784444	-104.89639	Denver, CO	Rebecca Thomson	0	1
DCO_010	2014-09	DCO	39.784444	-104.89639	Denver, CO	Rebecca Thomson	0	2
DCO_011	2014-09	DCO	39.742488	-105.51361	Denver, CO	Rebecca Thomson	2	0
DCO_012	2014-09	DCO	39.754086	-104.89274	Denver, CO	Rebecca Thomson	1	0
DCO_013	2014-09	DCO	39.754086	-104.89274	Denver, CO	Rebecca Thomson	3	0
DGA_001	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	0	0
DGA_002	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	4	2
DGA_005	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	1	2
DGA_006	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	4	2
DGA_007	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	2	0
DGA_008	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	4	3
DGA_009	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	2	0
DGA_010	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	4	2
DGA_011	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	4	2
DGA_012	2015-07	DGA	34.124712	-83.215489	Danielsville, GA	Ed McAssey	2	3
D03-04	2015-06	DMN	46.817778	-92.079444	Duluth, MN	Marshall Wedger	3	2
D05-03	2015-06	DMN	46.806667	-92.0625	Duluth, MN	Marshall Wedger	0	0
D07-08	2015-06	DMN	46.786667	-92.095278	Duluth, MN	Marshall Wedger	0	0
D08-04	2015-06	DMN	46.799722	-92.1075	Duluth, MN	Marshall Wedger	0	1
D09-10	2015-06	DMN	46.803899	-92.134167	Duluth, MN	Marshall Wedger	2	1
DMN_100	2015-06	DMN	46.800278	-92.090556	Duluth, MN	Marshall Wedger	1	0
DMN_130	2015-06	DMN	46.811389	-92.080556	Duluth, MN	Marshall Wedger	3	1
DMN_155	2015-06	DMN	46.796667	-92.0825	Duluth, MN	Marshall Wedger	1	1
DMN_014	2014	DMN	46.797961	-92.094026	Duluth, MN	Briana Gross	0	2
DMN_016	2014	DMN	46.805066	-92.074884	Duluth, MN	Briana Gross	0	0
DWI_001	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	0	0
DWI_002	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	3	1
DWI_003	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	0	0
DWI_005	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	0	2
DWI_006	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	0	0
DWI_008	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	0	1
DWI_009	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	2	2
DWI_010	2015-08	DWI	43.43078	-89.733112	Devil's Lake, WI	Miranda Colletta	2	0
G02-14	2015-06	GFL	29.803349	-82.416056	Gainesville, FL	Patricio Munoz	6	3
G03-07	2015-06	GFL	29.738727	-82.276693	Gainesville, FL	Patricio Munoz	4	4

G06-14	2015-06	GFL	29.643707	-82.355531	Gainesville, FL	Patricio Munoz	5	0
G07-01	2015-06	GFL	29.643707	-82.355531	Gainesville, FL	Patricio Munoz	0	5
G08-07	2015-06	GFL	29.643707	-82.355531	Gainesville, FL	Patricio Munoz	0	1
G09-21	2015-06	GFL	29.643707	-82.355531	Gainesville, FL	Patricio Munoz	1	4
GFL_6001	2015-06	GFL	29.633528	-82.355694	Gainesville, FL	Patricio Munoz	3	2
GFL_6035	2015-06	GFL	29.794477	-82.493449	Gainesville, FL	Patricio Munoz	5	3
GFL_6041	2015-06	GFL	29.643707	-82.355531	Gainesville, FL	Patricio Munoz	1	4
GFL_6096	2015-06	GFL	29.643707	-82.355531	Gainesville, FL	Patricio Munoz	4	2
HDE_001	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	2	0
HDE_006	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	0	0
HDE_007	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	0	1
HDE_008	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	0	4
HDE_009	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	3	2
HDE_010	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	0	2
HDE_011	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	5	0
HDE_012	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	2	0
HDE_013	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	0	0
HDE_014	2014-09	HDE	38.725248	-75.283731	Harbeson, DE	John McDonald	0	0
LKY_001	2014-09	LKY	38.279444	-85.620278	Louisville, KY	Sunita Crittenden	0	0
LKY_002	2014-09	LKY	38.279444	-85.620556	Louisville, KY	Sunita Crittenden	1	5
LKY_003	2014-09	LKY	38.279444	-85.621111	Louisville, KY	Sunita Crittenden	0	2
LKY_004	2014-09	LKY	38.279444	-85.621944	Louisville, KY	Sunita Crittenden	1	0
LKY_005	2014-09	LKY	38.282222	-85.618333	Louisville, KY	Sunita Crittenden	0	1
LKY_006	2014-09	LKY	38.283056	-85.618056	Louisville, KY	Sunita Crittenden	1	0
LKY_007	2014-09	LKY	38.283056	-85.618056	Louisville, KY	Sunita Crittenden	1	0
LKY_008	2014-09	LKY	38.283056	-85.618611	Louisville, KY	Sunita Crittenden	0	0
LKY_010	2014-09	LKY	38.283611	-85.617778	Louisville, KY	Sunita Crittenden	1	0
LKY_011	2014-09	LKY	38.283889	-85.6175	Louisville, KY	Sunita Crittenden	1	0
LMT_001	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	0	0
LMT_002	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	5	1
LMT_004	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	2	2
LMT_005	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	1	2
LMT_008	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	1	0
LMT_010	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	1	0
LMT_011	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	0	0
LMT_012	2014-09	LMT	45.650278	-110.5625	Livingston, MT	Sara Wright	0	0
LMT_013	2014-09	LMT	45.6536	-110.5598	Livingston, MT	Sara Wright	2	0
LMT_014	2014-09	LMT	45.6536	-110.5598	Livingston, MT	Sara Wright	0	0
LNE_001	2015	LNE	40.836625	-96.651273	Lincoln, NE	Trisha Spanbauer	0	2

LNE_002	2015	LNE	40.835631	-96.648226	Lincoln, NE	Trisha Spanbauer	7	2
LNE_004	2015	LNE	40.836309	-96.651577	Lincoln, NE	Trisha Spanbauer	0	0
LNE_005	2015	LNE	40.836253	-96.651337	Lincoln, NE	Trisha Spanbauer	3	0
LNE_007	2015	LNE	40.836349	-96.649343	Lincoln, NE	Trisha Spanbauer	1	0
LNE_008	2015	LNE	40.836349	-96.649343	Lincoln, NE	Trisha Spanbauer	0	2
LNE_010	2015	LNE	40.8363	-96.648664	Lincoln, NE	Trisha Spanbauer	0	2
MKS_001	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	2	0
MKS_002	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	0	1
MKS_003-2	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	6	0
MKS_004-2	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	3	2
MKS_005	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	3	0
MKS_006	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	3	0
MKS_007	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	0	0
MKS_008	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	7	0
MKS_009-2	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	4	0
MKS_010	2014-09	MKS	39.197791	-96.55872	Manhattan, KS	Kate Waselkov	3	0
MLA_002	2015-06	MLA	32.525278	-92.132778	Monroe, LA	John McIntire	4	3
MLA_003	2015-06	MLA	32.5425	-92.095833	Monroe, LA	John McIntire	4	4
MLA_004	2015-06	MLA	32.5425	-92.095833	Monroe, LA	John McIntire	4	3
MLA_005	2015-06	MLA	32.545525	-92.090831	Monroe, LA	John McIntire	4	4
MLA_006	2015-06	MLA	32.545525	-92.090831	Monroe, LA	John McIntire	2	2
MLA_007	2015-06	MLA	32.513056	-92.119722	Monroe, LA	John McIntire	3	3
MLA_008	2015-06	MLA	32.513056	-92.119722	Monroe, LA	John McIntire	3	2
MLA_010	2015-06	MLA	32.514722	-92.114722	Monroe, LA	John McIntire	2	4
MLA_011	2015-06	MLA	32.522222	-92.130556	Monroe, LA	John McIntire	7	4
MLA_012	2015-06	MLA	32.518333	-92.108611	Monroe, LA	John McIntire	4	4
MMN_001	NA	MMN	45.081522	-93.199453	Minneapolis, MN	Mary Lyman-Onkka	0	0
MMN_002	NA	MMN	44.949541	-93.106725	St. Paul, MN	Trisha Spanbauer	0	0
MMN_007	NA	MMN	45.405833	-92.822778	Chisago City, MN	Marshall Wedger	0	0
MMN_010	NA	MMN	46.010556	-92.930833	Hinckley, MN	Marshall Wedger	0	0
MMN_012	NA	MMN	45.288226	-92.848073	Scandia, MN	Marshall Wedger	2	0
MMN_013	NA	MMN	45.288276	-92.849752	Scandia, MN	Marshall Wedger	2	0
MMN_014	NA	MMN	45.288276	-92.849752	Scandia, MN	Marshall Wedger	0	0
MWI_001	2015-07	MWI	43.069185	-89.401679	Madison, WI	Daniel Tabor	0	0
MWI_002	2015-07	MWI	43.069185	-89.401679	Madison, WI	Daniel Tabor	2	0

MWI_003	2015-07	MWI	43.069185	-89.401679	Madison, WI	Daniel Tabor	2	0
MWI_004	2015-07	MWI	43.069185	-89.401679	Madison, WI	Daniel Tabor	2	0
MWI_005	2015-07	MWI	43.069185	-89.401679	Madison, WI	Daniel Tabor	2	0
MWI_006	2015-07	MWI	43.054264	-89.526083	Madison, WI	Daniel Tabor	4	0
MWI_008	2015-07	MWI	43.054264	-89.526083	Madison, WI	Daniel Tabor	4	0
MWI_009	2015-07	MWI	43.054264	-89.526083	Madison, WI	Daniel Tabor	3	1
MWI_010	2015-07	MWI	43.054264	-89.526083	Madison, WI	Daniel Tabor	4	0
MWI_011	2015-07	MWI	43.054264	-89.526083	Madison, WI	Daniel Tabor	2	1
NFWC_04-3	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	4	0
NFWC_04-5	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	5	0
NFWC_04-7	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	4	2
NFWC_04-8	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	4	2
NFWC_05-118	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	1	5
NFWC_05-120	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	4	0
NFWC_05-122	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	4	6
NFWC_05-124	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	3	1
NFWC_05-128	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	2	3
NFWC_05-132	NA	AOK	34.36086	-97.171325	Ardmore, OK	Maria Monteros	3	2
NLA_001	2016	NLA	29.928611	-90.078056	New Orleans, LA	Sara Wright	2	2
NLA_002	2016	NLA	29.928611	-90.078056	New Orleans, LA	Sara Wright	1	2
NLA_003	2016	NLA	29.928889	-90.078056	New Orleans, LA	Sara Wright	3	4
NLA_005	2016	NLA	29.928889	-90.078056	New Orleans, LA	Sara Wright	0	0
NLA_006	2016	NLA	29.928889	-90.078056	New Orleans, LA	Sara Wright	1	1
NLA_008	2016	NLA	29.936389	-90.0725	New Orleans, LA	Sara Wright	4	1
NLA_009	2016	NLA	29.936389	-90.0725	New Orleans, LA	Sara Wright	4	2
NLA_010	2016	NLA	29.936389	-90.0725	New Orleans, LA	Sara Wright	3	2
NLA_011	2016	NLA	29.936389	-90.0725	New Orleans, LA	Sara Wright	3	4
NLA_012	2016	NLA	29.936389	-90.0725	New Orleans, LA	Sara Wright	2	2
NTN_001	2015-08	NTN	36.181944	-86.749167	Nashville, TN	Sara Wright	0	0
NTN_002	2015-08	NTN	36.184444	-86.758333	Nashville, TN	Sara Wright	3	0
NTN_004	2015-08	NTN	36.184444	-86.758333	Nashville, TN	Sara Wright	0	0
NTN_005	2015-08	NTN	36.184444	-86.758333	Nashville, TN	Sara Wright	3	2
NTN_006	2015-08	NTN	36.184444	-86.758333	Nashville, TN	Sara Wright	0	0
NTN_010	2015-08	NTN	36.148775	-86.812494	Nashville, TN	Sara Wright	1	0
NTN_011	2015-08	NTN	36.148775	-86.812494	Nashville, TN	Sara Wright	1	0

NTN_012	2015-08	NTN	36.148775	-86.812494	Nashville, TN	Sara Wright	3	0
NTN_013	2015-08	NTN	36.148775	-86.812494	Nashville, TN	Sara Wright	3	0
NTN_014	2015-08	NTN	36.148775	-86.812494	Nashville, TN	Sara Wright	0	0
NYC_004	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	0	0
NYC_005	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	0	2
NYC_007	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	2	1
NYC_008	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	0	0
NYC_009	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	4	0
NYC_010	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	2	0
NYC_011	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	1	2
NYC_012	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	6	2
NYC_013	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	0	0
NYC_014	2016	NYC	40.782182	-73.965913	New York City, NY	Dan Cui Zhou	3	0
PIA_001	2014-09	PIA	41.416389	-95.853611	Pottawattamie, IA	Sara Wright	3	0
PIA_002	2014-09	PIA	41.416389	-95.853611	Pottawattamie, IA	Sara Wright	0	0
PIA_005	2014-09	PIA	41.416389	-95.853611	Pottawattamie, IA	Sara Wright	0	0
PIA_006	2014-09	PIA	41.416389	-95.853611	Pottawattamie, IA	Sara Wright	0	1
PIA_007	2014-09	PIA	41.416389	-95.853611	Pottawattamie, IA	Sara Wright	0	1
PIA_008	2014-09	PIA	41.416389	-95.853611	Pottawattamie, IA	Sara Wright	0	1
PIA_010	2014-09	PIA	41.416389	-95.853611	Pottawattamie, IA	Sara Wright	1	0
PIA_011	2014-09	PIA	41.414722	-95.855833	Pottawattamie, IA	Sara Wright	0	0
PIA_012	2014-09	PIA	41.414722	-95.855833	Pottawattamie, IA	Sara Wright	0	0
PIA_013	2014-09	PIA	41.414722	-95.855833	Pottawattamie, IA	Sara Wright	0	2
PPA_001	2014-08	PPA	40.446664	-79.95409	Pittsburgh, PA	Lisa Limeri	5	3
PPA_003	2014-08	PPA	40.446664	-79.95409	Pittsburgh, PA	Lisa Limeri	3	1
PPA_004	2014-08	PPA	40.446664	-79.95409	Pittsburgh, PA	Lisa Limeri	3	0
PPA_006	2014-08	PPA	40.446664	-79.95409	Pittsburgh, PA	Lisa Limeri	1	1
PPA_007	2014-08	PPA	40.446664	-79.95409	Pittsburgh, PA	Lisa Limeri	0	0
PPA_008	2014-08	PPA	40.446664	-79.95409	Pittsburgh, PA	Lisa Limeri	2	0
PPA_009	2014-08	PPA	40.446664	-79.95409	Pittsburgh, PA	Lisa Limeri	3	0
PPA_010	2014-08	PPA	40.479275	-79.918644	Pittsburgh, PA	Lisa Limeri	3	1
PPA_011	2014-08	PPA	40.479275	-79.918644	Pittsburgh, PA	Lisa Limeri	4	0

PPA_012	2014-08	PPA	40.479275	-79.918644	Pittsburgh, PA	Lisa Limeri	2	0
RCD_001	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	3	0
RCD_002	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	3	0
RCD_004	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	0	0
RCD_005-2	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	3	0
RCD_006	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	2	0
RCD_007-2	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	3	0
RCD_008-2	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	4	4
RCD_009-2	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	4	0
RCD_010-2	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	0	0
RCD_013	2014-09	RCD	44.071389	-103.27194	Rapid City, SD	Sara Wright	6	1
RDE_001	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	2	0
RDE_002	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	2	0
RDE_003	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	3	0
RDE_004	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	0	0
RDE_005	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	0	3
RDE_007	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	0	0
RDE_008	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	0	0
RDE_009	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	8	0
RDE_010	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	0	0
RDE_011	2014-09	RDE	38.61383	-75.06528	Dewey Beach, DE	John McDonald	0	2
SCD_001	2014-09	SCD	42.540556	-96.526944	Sioux City, SD	Sara Wright	3	0
SCD_003	2014-09	SCD	42.540556	-96.526944	Sioux City, SD	Sara Wright	0	0
SCD_004	2014-09	SCD	42.540556	-96.526944	Sioux City, SD	Sara Wright	0	2
SCD_005	2014-09	SCD	42.540556	-96.526944	Sioux City, SD	Sara Wright	0	2
SCD_006-2	2014-09	SCD	42.538889	-96.526111	Sioux City, SD	Sara Wright	0	1
SCD_007	2014-09	SCD	42.538889	-96.526111	Sioux City, SD	Sara Wright	3	0
SCD_008	2014-09	SCD	42.538889	-96.526111	Sioux City, SD	Sara Wright	0	0
SCD_009	2014-09	SCD	42.538889	-96.526111	Sioux City, SD	Sara Wright	0	0
SCD_010	2014-09	SCD	42.540278	-96.527222	Sioux City, SD	Sara Wright	0	2
SCD_011	2014-09	SCD	42.540278	-96.527222	Sioux City, SD	Sara Wright	0	0
SCT_001	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	0
SCT_002	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	0

SCT_003	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	0
SCT_004	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	0
SCT_005	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	0
SCT_006	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	2
SCT_007	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	0
SCT_008	2015-06	SCT	41.822875	-72.820475	Simsbury, CT	Anne Puzzo	0	0
SFD_002	2014-09	SFD	43.5075	-96.733611	Sioux Falls, SD	Sara Wright	0	0
SFD_003	2014-09	SFD	43.5075	-96.733611	Sioux Falls, SD	Sara Wright	1	0
SFD_005	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	0	0
SFD_006	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	1	0
SFD_007	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	2	0
SFD_008-2	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	0	0
SFD_009	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	0	0
SFD_010	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	0	0
SFD_011	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	0	0
SFD_012	2014-09	SFD	43.508056	-96.7325	Sioux Falls, SD	Sara Wright	3	1
S01-08	2015-05	STL	38.650556	-90.2975	St. Louis, MO	Sara Wright	1	1
S03-03	2015-05	STL	38.648333	-90.288056	St. Louis, MO	Sara Wright	2	2
S04-04	2015-05	STL	38.673333	-90.318056	St. Louis, MO	Sara Wright	4	0
S05-03	2015-05	STL	38.661389	-90.304444	St. Louis, MO	Sara Wright	2	3
S07-08	2015-05	STL	38.635556	-90.284444	St. Louis, MO	Sara Wright	0	0
S11-06	2015-05	STL	38.663889	-90.322222	St. Louis, MO	Sara Wright	0	3
S13-06	2015-05	STL	38.656389	-90.310833	St. Louis, MO	Sara Wright	2	0
STL_6013	2015-05	STL	38.649722	-90.260833	St. Louis, MO	Sara Wright	2	0
STL_6049	2015-05	STL	38.633889	-90.302778	St. Louis, MO	Sara Wright	4	0
STL_6101	2015-05	STL	38.627891	-90.319764	St. Louis, MO	Sara Wright	0	0
TGA_001	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	1	4
TGA_002	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	2	3
TGA_003	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	1	4
TGA_004	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	1	4
TGA_005	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	3	3
TGA_006	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	1	4
TGA_008	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	4	3
TGA_009	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	2	3
TGA_011	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	4	3
TGA_012	2015-06	TGA	31.458046	-83.517106	Tifton, GA	Patricio Munoz	4	1
VBC_001	2014-09	VBC	49.264999	-123.24768	Vancouver, BC	Kathryn Turner	5	4
VBC_002	2014-09	VBC	49.251094	-123.23439	Vancouver, BC	Kathryn Turner	0	0
VBC_004	2014-09	VBC	49.033333	-122.78333	Vancouver, BC	Kathryn Turner	3	1

VBC_008	2014-09	VBC	49.268999	-123.15273	Vancouver, BC	Kathryn Turner	0	1
VBC_009	2014-09	VBC	49.1964	-123.10332	Vancouver, BC	Kathryn Turner	2	0
VBC_010	2014-09	VBC	49.266667	-123.24167	Vancouver, BC	Kathryn Turner	4	1
VBC_012	2014-09	VBC	49.261689	-123.25236	Vancouver, BC	Kathryn Turner	0	0
VBC_014	2014-09	VBC	49.225248	-123.07315	Vancouver, BC	Kathryn Turner	3	0
VBC_015	2014-09	VBC	49.087056	-123.04275	Vancouver, BC	Kathryn Turner	0	2
VBC_016	2014-09	VBC	49.171	-122.883	Vancouver, BC	Kathryn Turner	4	0
WKY_003	2015-08	WKY	36.965278	-89.089167	Wickliffe, KY	Sara Wright	0	2
WKY_004	2015-08	WKY	36.965278	-89.089167	Wickliffe, KY	Sara Wright	0	2
WKY_005	2015-08	WKY	36.965278	-89.089167	Wickliffe, KY	Sara Wright	0	1
WKY_006	2015-08	WKY	37.002778	-89.168889	Cairo, IL	Sara Wright	2	0
WKY_007	2015-08	WKY	37.002778	-89.168889	Cairo, IL	Sara Wright	1	4
WKY_008	2015-08	WKY	37.002778	-89.168889	Cairo, IL	Sara Wright	3	2
WKY_009	2015-08	WKY	37.002778	-89.168889	Cairo, IL	Sara Wright	0	0
WKY_010	2015-08	WKY	37.002778	-89.168889	Cairo, IL	Sara Wright	1	4
WKY_011	2015-08	WKY	37.002778	-89.168889	Cairo, IL	Sara Wright	2	0
WKY_012	2015-08	WKY	37.002778	-89.168889	Cairo, IL	Sara Wright	2	2
WNC_001	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	3	0
WNC_002	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	2	0
WNC_003	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	2	1
WNC_004	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	0	0
WNC_005	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	3	0
WNC_006	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	0	2
WNC_007	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	4	2
WNC_008	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	0	0
WNC_010	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	1	0
WNC_012	2015-06	WNC	36.014791	-80.128158	Winston-Salem, NC	Cynthia Vigueira	2	3
DMN_010	2014	DMN	46.8185	-92.085	Duluth, MN	Sara Wright	0	0
STL_0701	2014	STL	38.6051	-90.2673	St. Louis, MO	Sara Wright	0	0
GFL_007	2015-06	GFL	29.6407	-82.3623	Gainesville, FL	Sara Wright	5	4

†CNV: Copy number variation

Supplementary Table S3.2 AMOVA

	df	Sum Sq	Mean Sq	Sigma	%	p
Between Pop (location)	42	2565993.38	61095.0805	988.610496	2.88614474	<0.001
Between samples Within Pop (location)	372	15630533.7	42017.5637	8752.50413	25.5520186	<0.001
Within samples	415	10172710.5	24512.5554	24512.5554	71.5618367	<0.001
Total	829	28369237.6	34221.0345	34253.6701	100	

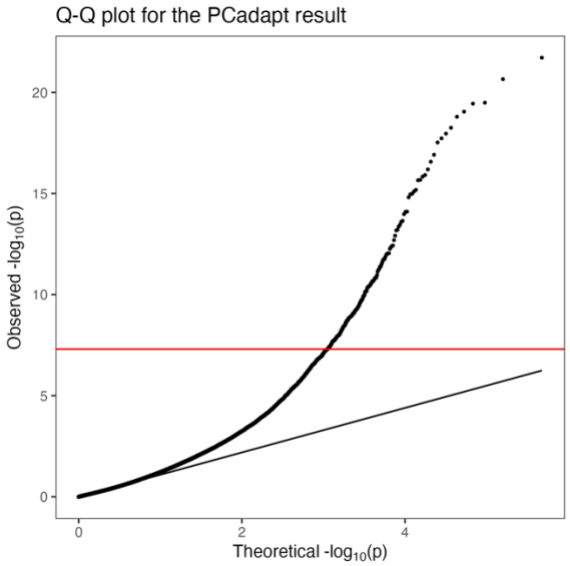
Supplementary Table S3.3 Candidate genes

Chr	Gene†	Protein‡	Description	Detected by	Reference
1	ARAD1	putative arabinosyltransferase ARAD1	Modulator of the flexibility of cell walls and thigmomorphogenesis response by regulating arabinan biosynthesis	LFMM PCadapt	Verhertbrugge n *et al.* (2013)
2	MAIL1	MAIN-LIKE 1	Regulator of primary **root development**	LFMM	Uhlken *et al.* (2014)
3	RPPL1	putative disease resistance RPP13-like protein 1	Related to downy mildew resistance. This gene has been found to have elevated evolutionary rate between species, which is likely to have adaptive values.	LFMM	Bittner-Eddy *et al.* (2000); Wang *et al.* (2011)
3	RPPL1	putative disease resistance RPP13-like protein 1	Related to downy mildew resistance. This gene has been found to have elevated evolutionary rate between species, which is likely to have adaptive values. (Paralogous gene on the same chromosome)	LFMM	See above
3	FRS5	FAR1-RELATED SEQUENCE 5-like	Light control development, including **flowering time regulation**	LFMM QTL	Lin and Wang (2004); Li *et al.* (2011); Ma and Li (2018)
5	RPM1	disease resistance protein RPM1	Pathogen recognition gene, which has been reported to have high genetic polymorphism in Arabidopsis	LFMM QTL	Stahl *et al.* (1999)
6	RGA1	putative disease resistance protein RGA1	A resistance gene analogue (RGA), which confers late blight resistance in Solanaceae species	LFMM	van der Vossen *et al.* (2003)
7	LEP	ethylene-responsive transcription factor	A AP2 transcription factor in subgroup VIIIb, regulating leaf development and GA-regulated seed germination. The genes in the same subgroup (VIIIb) also count for **regulating the transition of vegetative meristem to floral meristem**.	LFMM	Chandler (2018)
7	PDP3	Tudor/PWWP/MBT domain-containing protein	A PWWP domain proteins (PDPs), interacting with FVE and MSI5 to suppress *FLC* expression and thereby **promote flowering**	LFMM PCadapt	Zhou *et al.* (2018)
7	MYB13	transcription factor MYB13	A diverse regulator of multiple developmental processes and stress responses	LFMM PCadapt	Wang *et al.* (2021)

7	CKL10	casein kinase 1-like protein 10	Kinase for cryptochromes (CRYs) phosphorylation, which is a blue light sensor related to several developmental process regulations, including photoperiodic **flowering**	LFMM PCadapt	Tan *et al.* (2013)
7	MIK2	MDIS1-interacting receptor like kinase 2	A leucine-rich repeat receptor kinase for elicits various immune responses for pathogen resistance and herbivore	LFMM	Hou *et al.* (2021); Stahl *et al.* (2022)
7	CYP734 A1	cytochrome P450 734A1	A brassinosteroid-inactivating enzymes, which is important for steroid-mediated signal transduction, involving in multiple developmental processes and stress responses	LFMM PCadapt	Thornton *et al.* (2010)
9	VRN1	B3 domain-containing transcription factor VRN1-like	Modulator of vernalization-dependent **flowering time regulation** through the FLC pathway	LFMM	Sheldon *et al.* (2000); Levy *et al.* (2002)
9	VRN1	B3 domain-containing transcription factor VRN1-like	Modulator of vernalization-dependent **flowering time regulation** through the FLC pathway (Paralogous gene on the same chromosome)	LFMM	See above
9	SUVR5	histone-lysine N-methyltransferase SUVR5	A histone methyltransferase and a gene expression regulator. It is a regulator of **flower timing** in Arabidopsis	LFMM	Caro *et al.* (2012)
10	VRN1	B3 domain-containing transcription factor VRN1-like	Modulator of vernalization-dependent **flowering time regulation** through the FLC pathway	LFMM QTL	See above
13	CBL1	calcineurin B-like protein 1	A calcium sensor of diverse roles in development and plant defense in addition to abiotic stress tolerance	PCadapt QTL	Meena *et al.* (2015)
13	FER	receptor-like protein kinase FERONIA	A regulator of the transcript accumulation and mRNA alternative splicing (AS) of some important **flowering** genes, including *FLC*. It plays important roles in floral transition	LFMM PCadapt	Wang *et al.* (2020)
15	MYB13	transcription factor MYB13	A diverse regulator of multiple developmental processes and stress responses	LFMM PCadapt QTL	See above
15	FER	receptor-like protein kinase FERONIA	A regulator of the transcript accumulation and mRNA alternative splicing (AS) of some important **flowering** genes, including *FLC*. It plays important roles in floral transition	LFMM PCadapt QTL	See above

15	CYP734A1	cytochrome P450 734A1	A brassinosteroid-inactivating enzymes, which is important for steroid-mediated signal	LFMM	See above
15	CYP734A1	cytochrome P450 734A1	A brassinosteroid-inactivating enzymes, which is important for steroid-mediated signal (Paralogous gene on the same chromosome)	LFMM	See above
15	MYB4	transcription factor MYB4	Controlling sinapate ester formation and in the development of sunscreens, offering protection to UV-B light	PCadapt	Jin *et al.* (2000)
16	BDA1	ankyrin repeat-containing protein BDA1	A critical signaling component to regulate plant immunity	LFMM	Yang *et al.* (2012)
16	BDA1	ankyrin repeat-containing protein BDA1	A critical signaling component to regulate plant immunity (Paralogous gene on the same chromosome)	LFMM	See above
16	BRXL2	protein Brevis radix-like 2	A probable modulator of **root growth**	LFMM	Briggs *et al.* (2006)

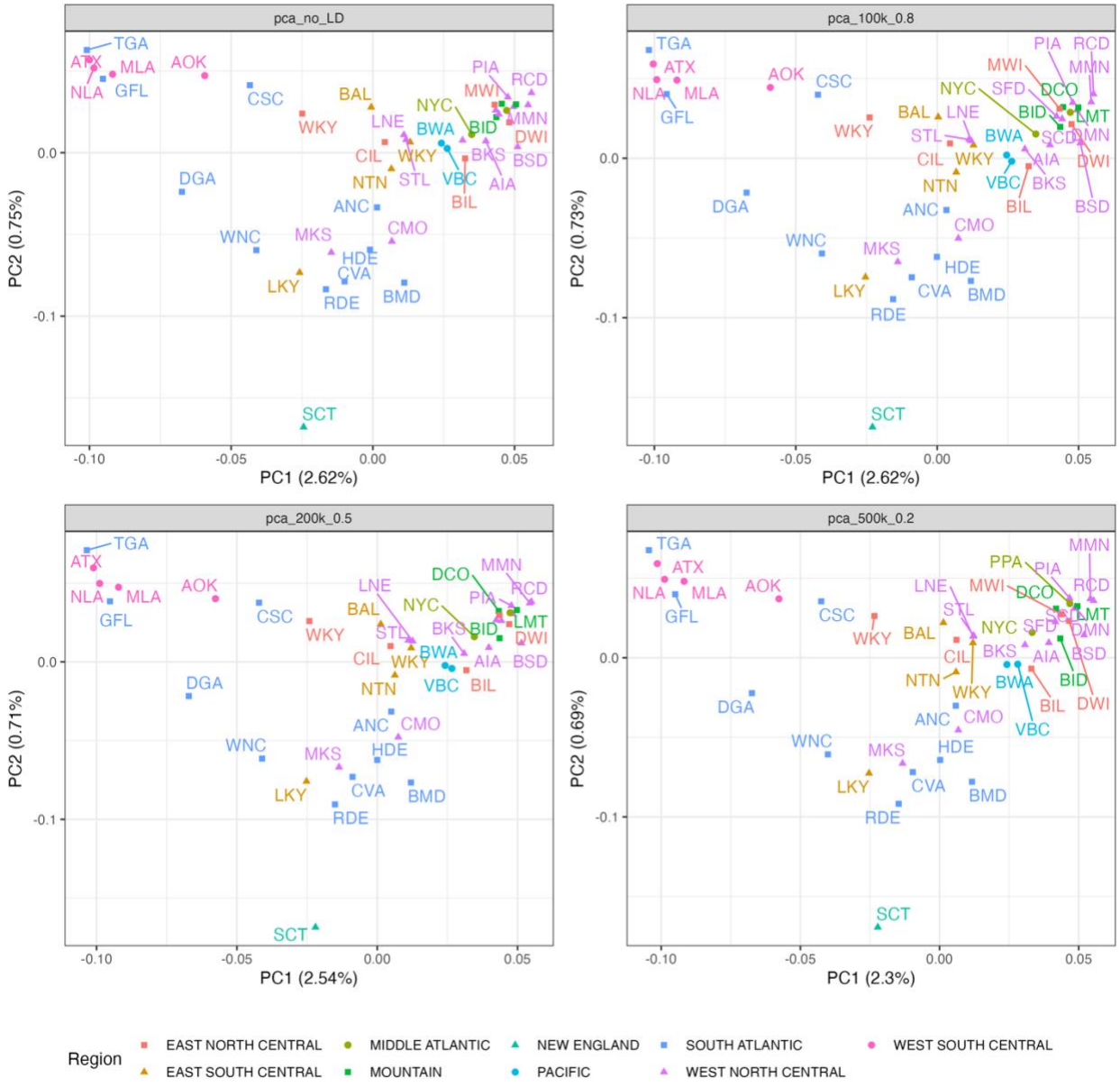
†Hyperlink to UniPort database for homologous genes in model species
‡Hyperlink to NCBI GenBank database for the genes detected in the current study.



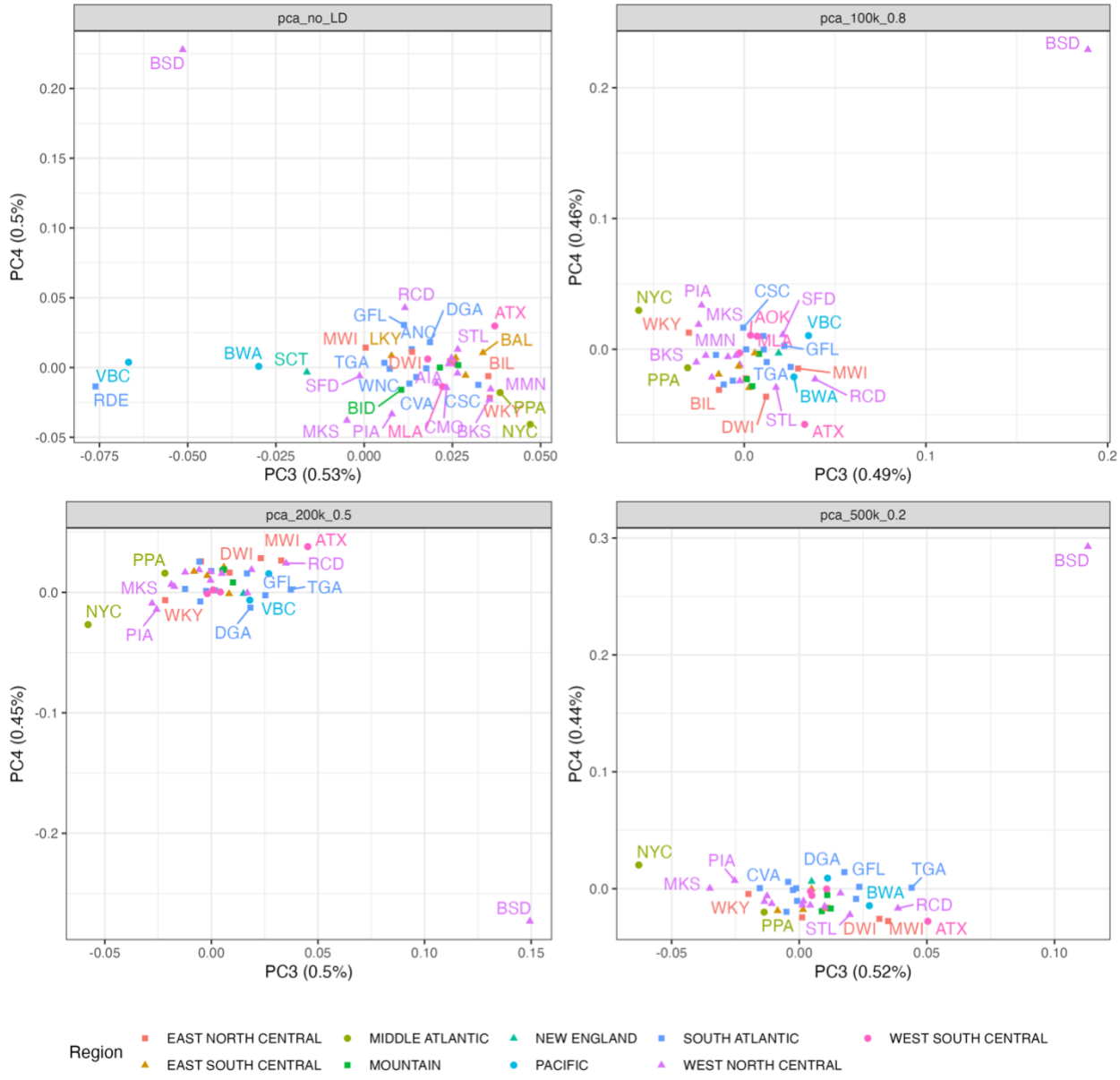
Supplementary Figure S3.1. Q-Q plot for the PCadapt result.

A conventional p-value threshold was set at 5×10^{-8} (red line).

(A)

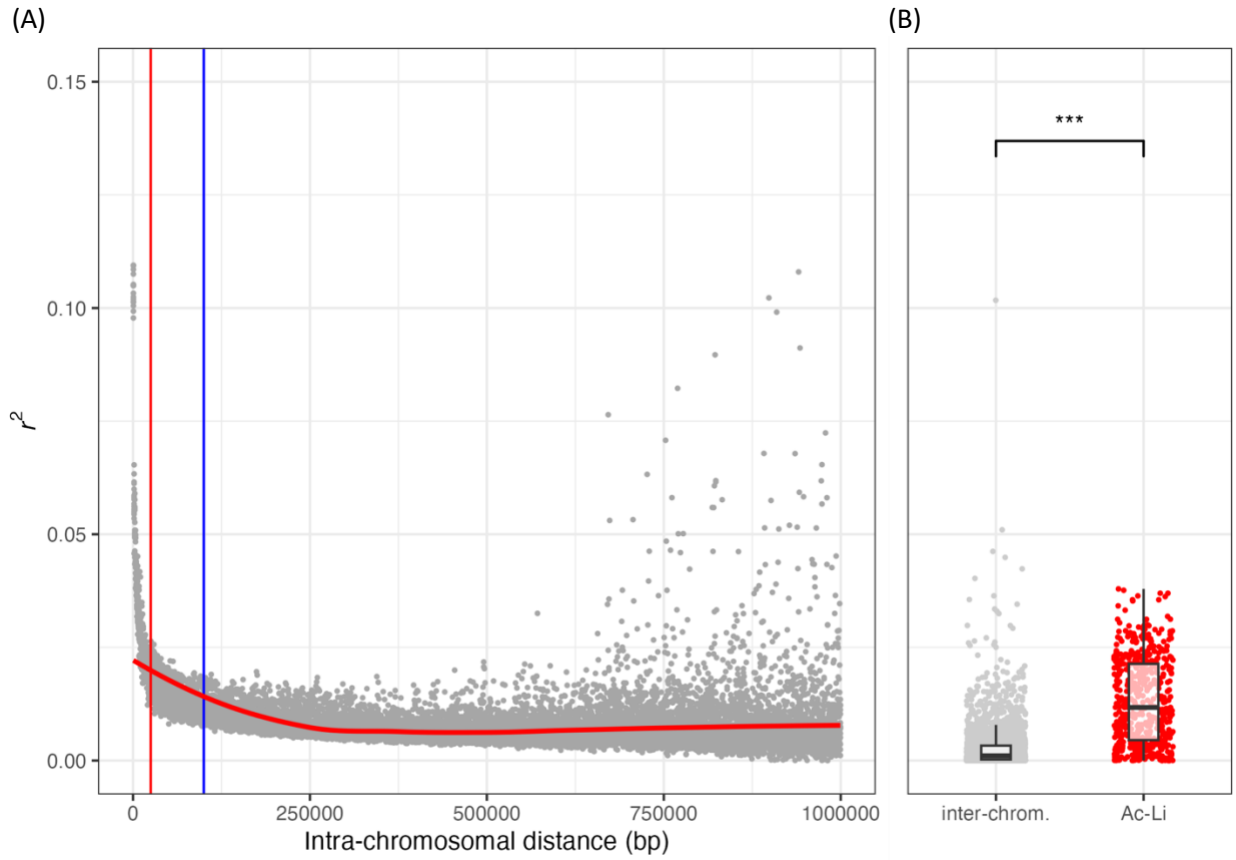

pca_no_LD
pca_100k_0.8
pca_200k_0.5
pca_500k_0.2
PC1 (2.62%)
PC2 (0.75%)
PC2 (0.73%)
PC1 (2.54%)
PC2 (0.71%)
PC1 (2.3%)
PC2 (0.69%)
Region
EAST NORTH CENTRAL
EAST SOUTH CENTRAL
MIDDLE ATLANTIC
MOUNTAIN
NEW ENGLAND
PACIFIC
SOUTH ATLANTIC
WEST NORTH CENTRAL
WEST SOUTH CENTRAL

(B)



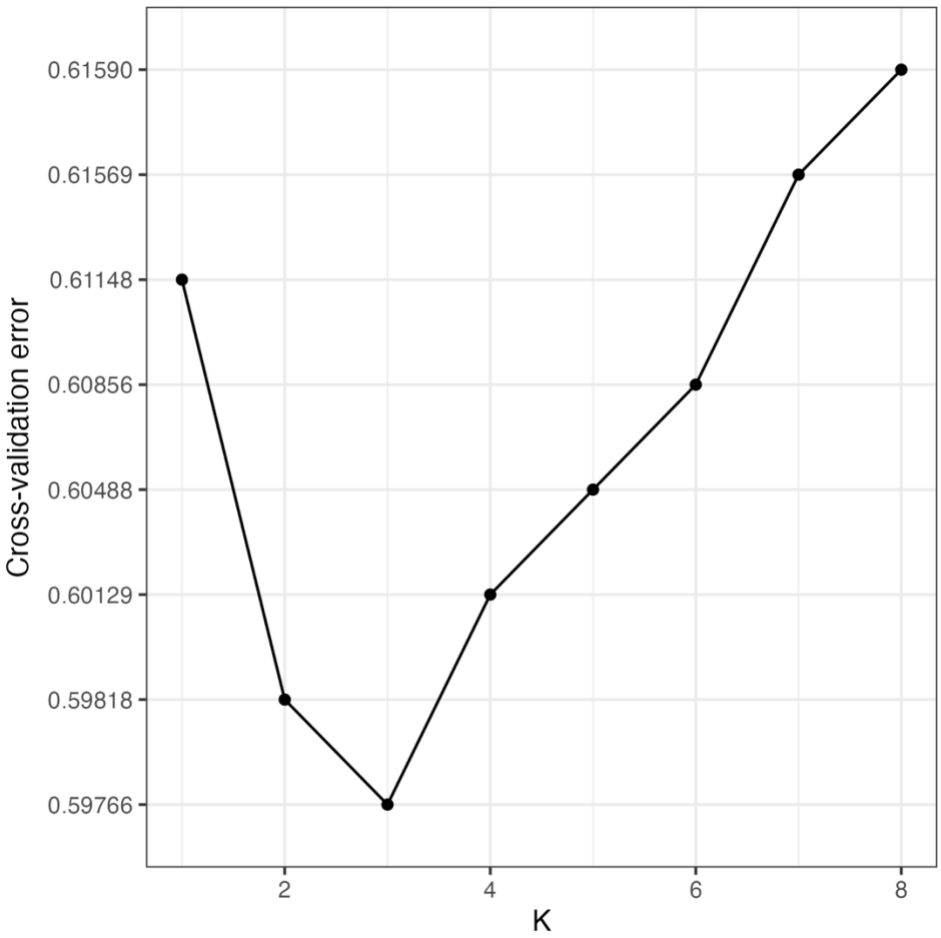
Supplementary Figure S3.2 Population structure assessment under different degrees of LD pruning.

(A), PC1 and PC2. (B), PC3 and PC4. Top left, original dataset. Top right, mild LD pruning (--indep-pairwise 100kb 0.8). Bottom left, medium LD pruning (--indep-pairwise 200kb 0.5). Bottom right, strong LD pruning (--indep-pairwise 500kb 0.2).



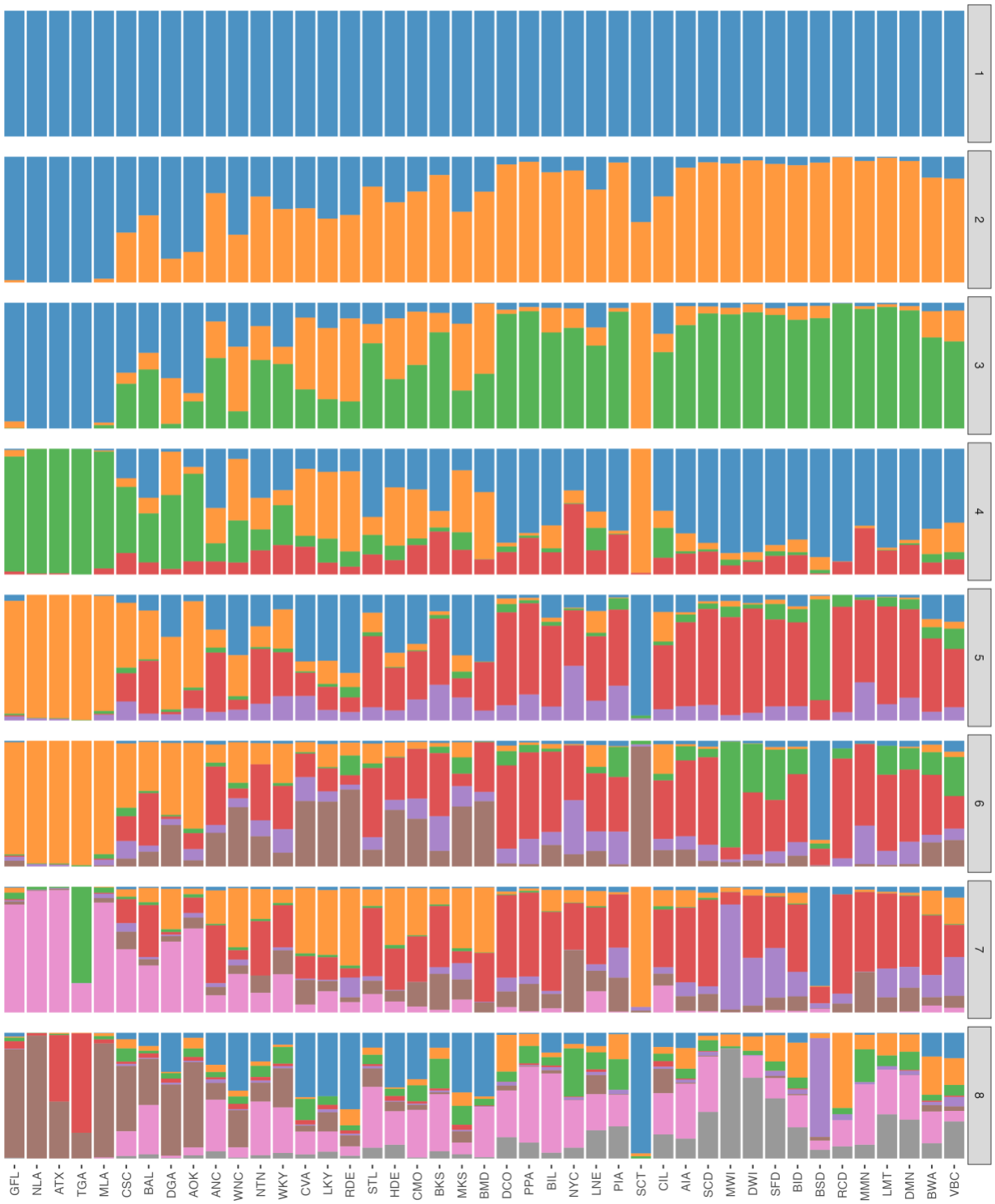
Supplementary Figure S3.3 LD decay of the mild-LD-pruned dataset (241,371 SNPs) and interchromosomal LD.

(A), the LDs of the pairwise SNPs between over 100 bp and under 1000 kbp were calculated in PLINK 1.9. Then, average LD per incremental 500 bp window was plotted. Red vertical line, 25 kbp. Blue vertical line, 100 kbp. (B), the interchromosomal LDs were from 3,000 randomly selected SNP pairs with minor allele frequency > 0.2 after excluding the SNPs linked to *Ac* and *Li* loci. The interchromosomal SNPs linked to *Ac* and *Li* loci were retrieved from the GWAS result, and then filtered with minor allele frequency > 0.2. The difference between the two distributions was statistically tested by Kolmogorov-Smirnov (KS) test (***p <0.001).



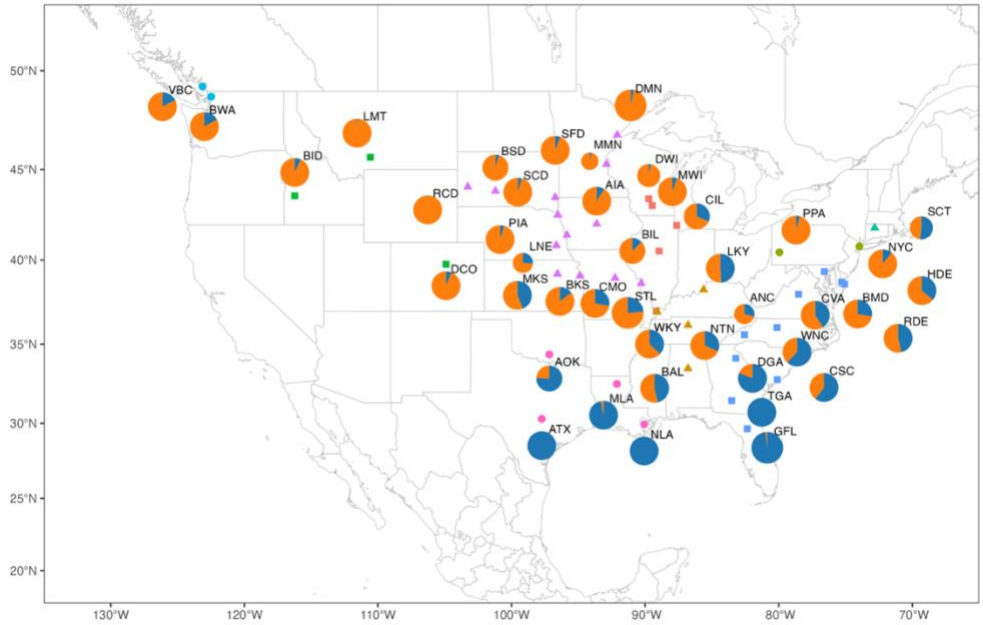
Supplementary Figure S3.4 Cross validation error of the ADMIXTURE analysis.

The minimum cross validation error happens at K = 3.

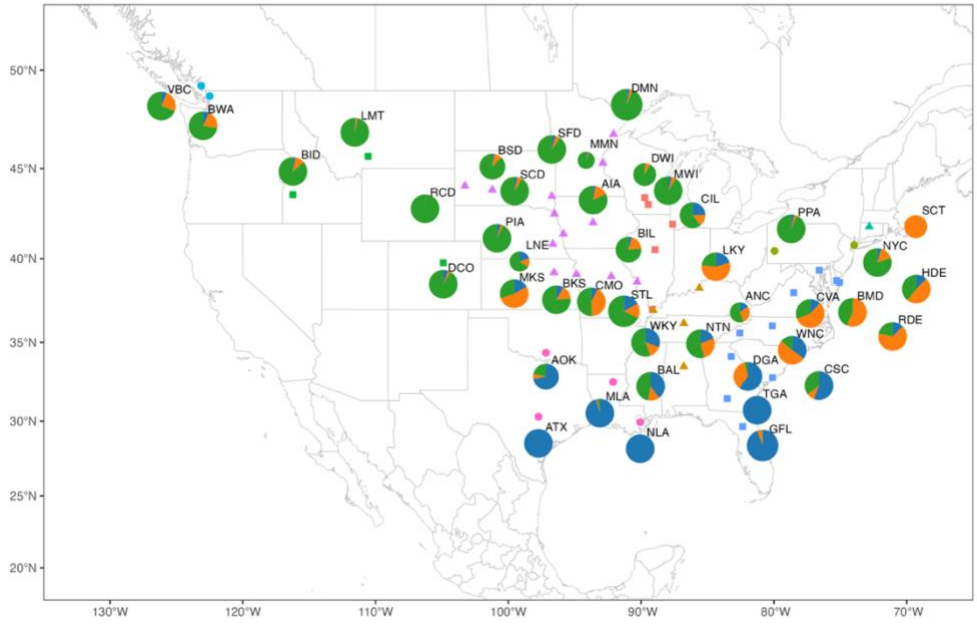


Supplementary Figure S3.5 ADMIXTURE analysis (K = 1-8).

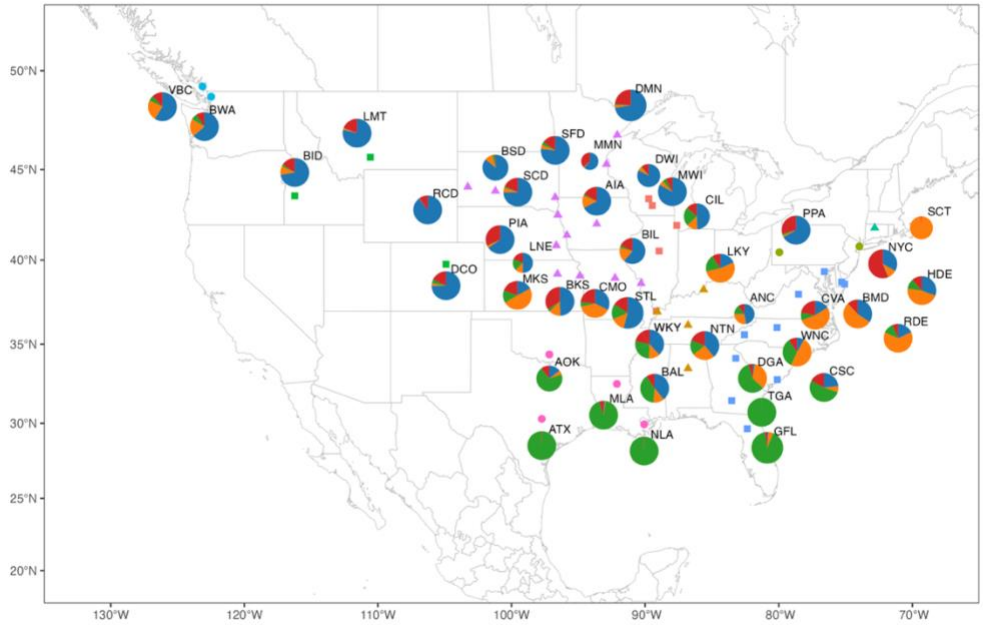
(A)


50°N
45°N
40°N
35°N
30°N
25°N
20°N
130°W
120°W
110°W
100°W
90°W
80°W
70°W
VBC
BWA
LMT
BID
DMN
SFD
MMN
BSD
SCD
DWI
MWI
RCD
AIA
CIL
PIA
PPA
SCT
BIL
LNE
LKY
NYC
DCO
MKS
BKS
CMO
STL
ANC
CVA
BMD
HDE
RDE
WKY
NTN
WNC
AOK
DGA
CSC
BAL
TGA
MLA
ATX
NLA
GFL

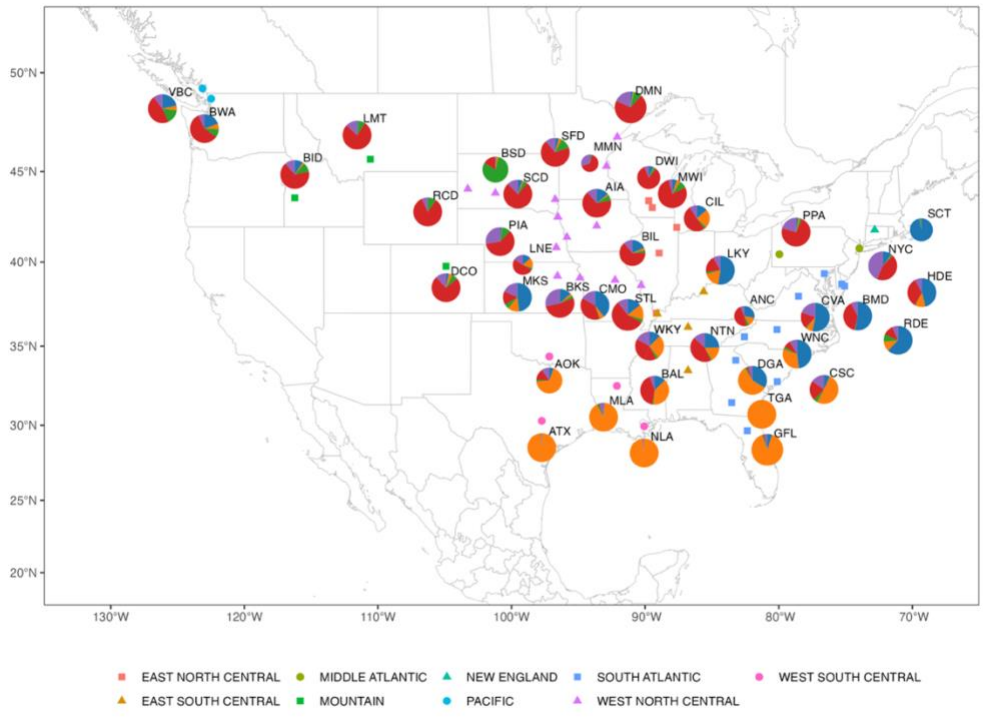
(B)


50°N
45°N
40°N
35°N
30°N
25°N
20°N
130°W
120°W
110°W
100°W
90°W
80°W
70°W
VBC
BWA
LMT
BID
DMN
SFD
MMN
BSD
SCD
DWI
MWI
RCD
AIA
CIL
PIA
PPA
SCT
BIL
LNE
LKY
NYC
DCO
MKS
BKS
CMO
STL
ANC
CVA
BMD
HDE
RDE
WKY
NTN
WNC
AOK
DGA
CSC
BAL
TGA
MLA
ATX
NLA
GFL

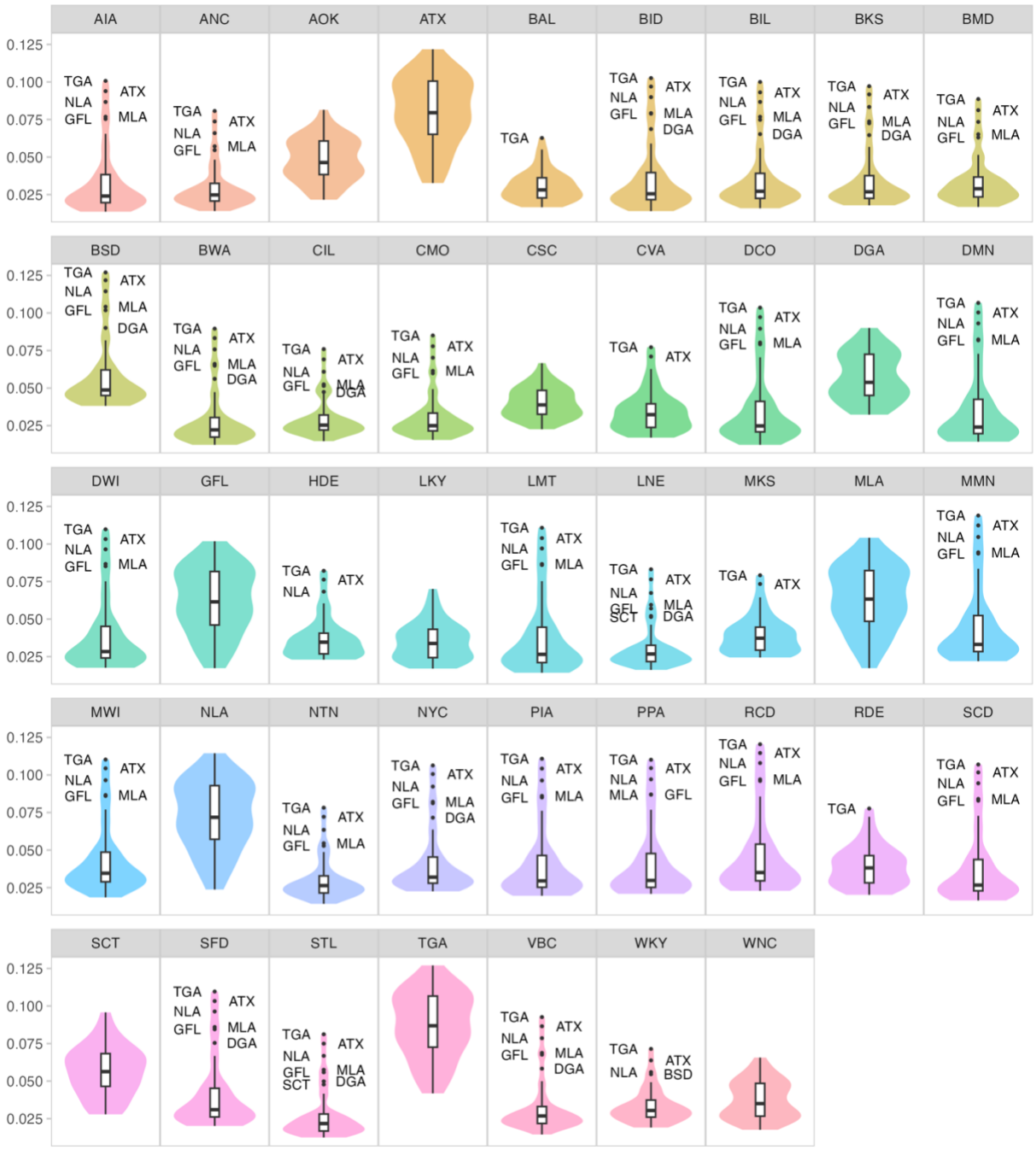
(C)



(D)

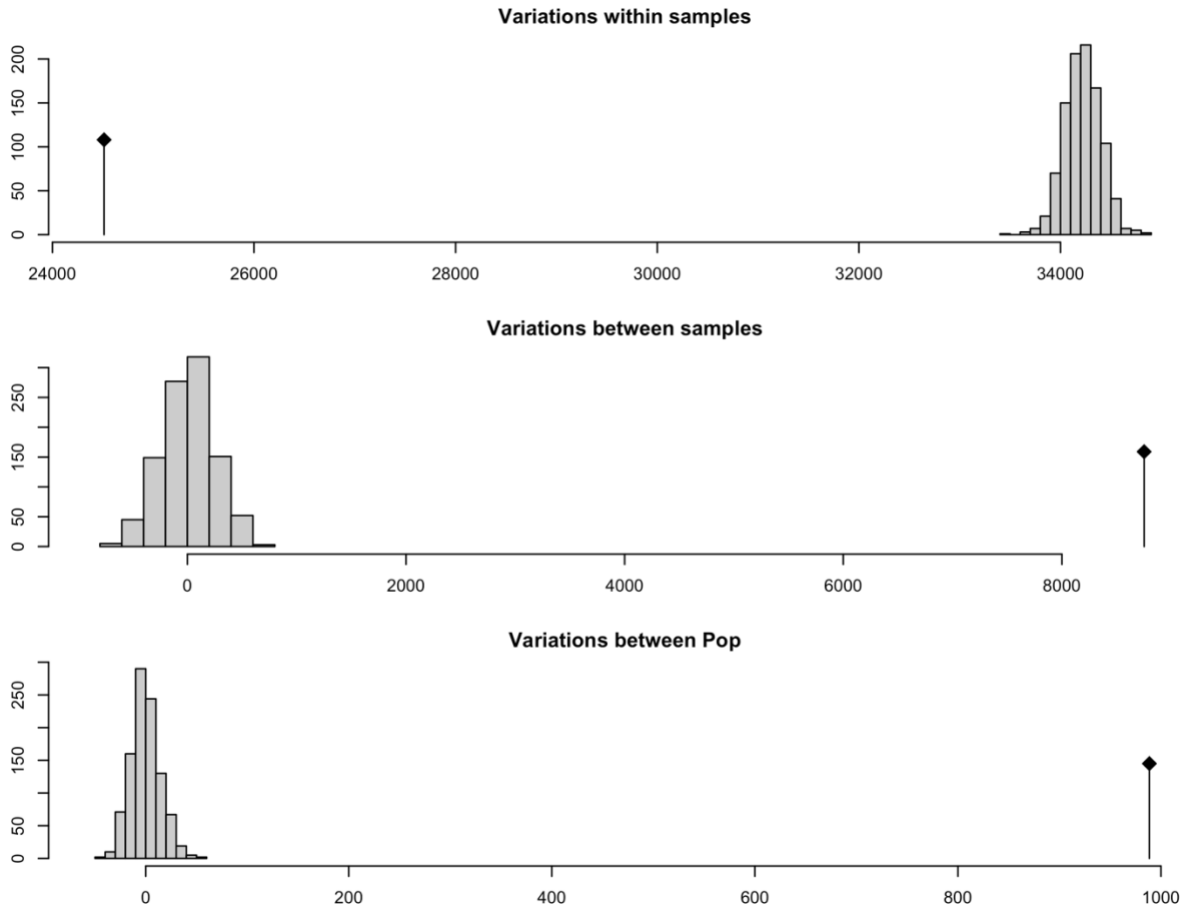


Supplementary Figure S3.6 Map of the ADMIXTURE result. (A), k = 2. (B), k = 3. (C), k = 4. (D), k = 5.

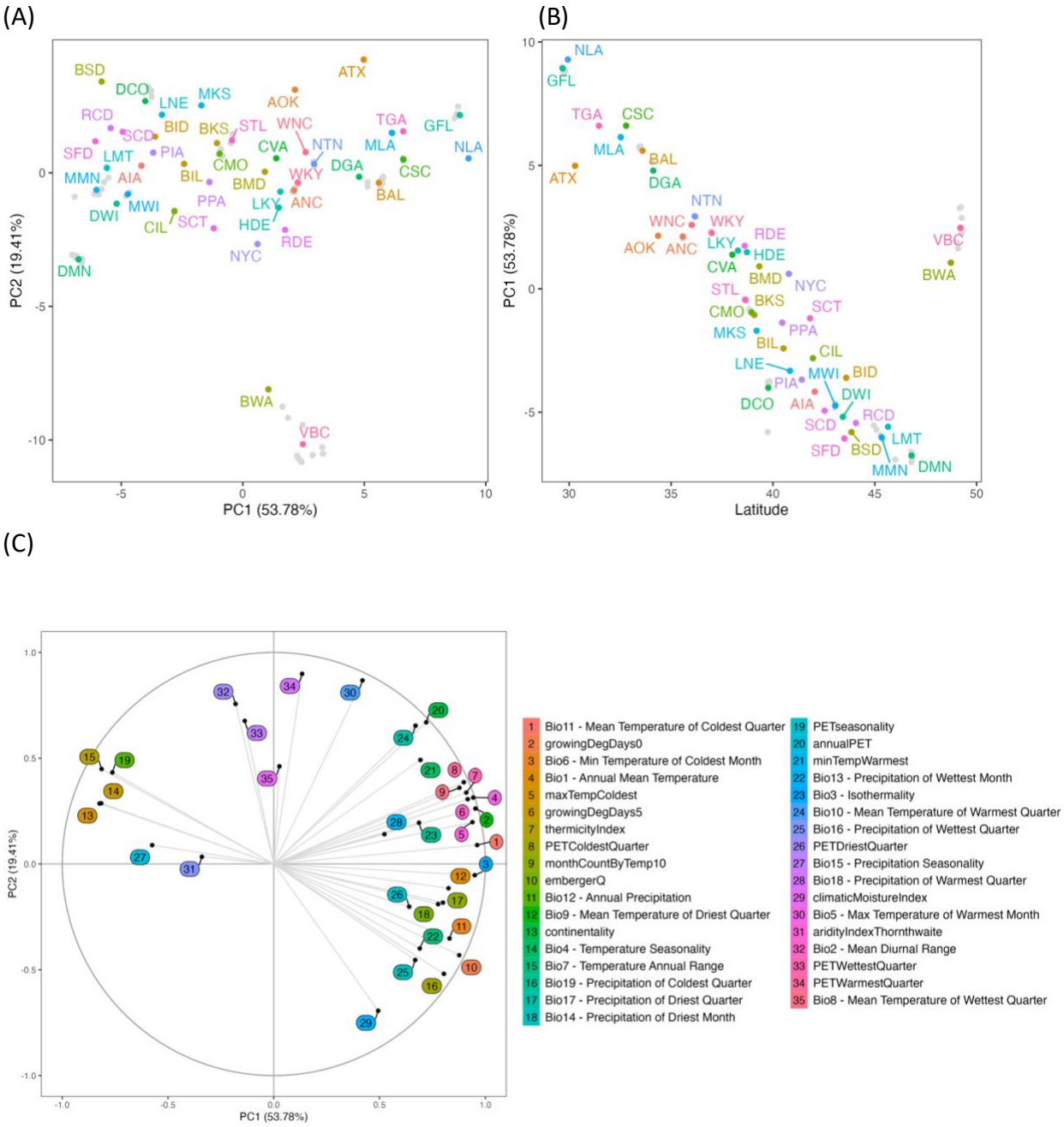


Supplementary Figure S3.7 Pairwise F_{ST} distributions between the 43 sampling locations.

The reference location is shown above each panel. The distribution below is the other 42 pairwise F_{ST} to the reference location. The location names are labeled if the F_{ST} value is characterized as the outliers in the distribution [> 75% quantile + 1.5*(75% quantile – 25% quantile)]. Note that the five south locations from Florida (GFL), Louisiana (MLA, NLA), Texas (ATX), Georgia (TGA) are frequently shown as the outliers.

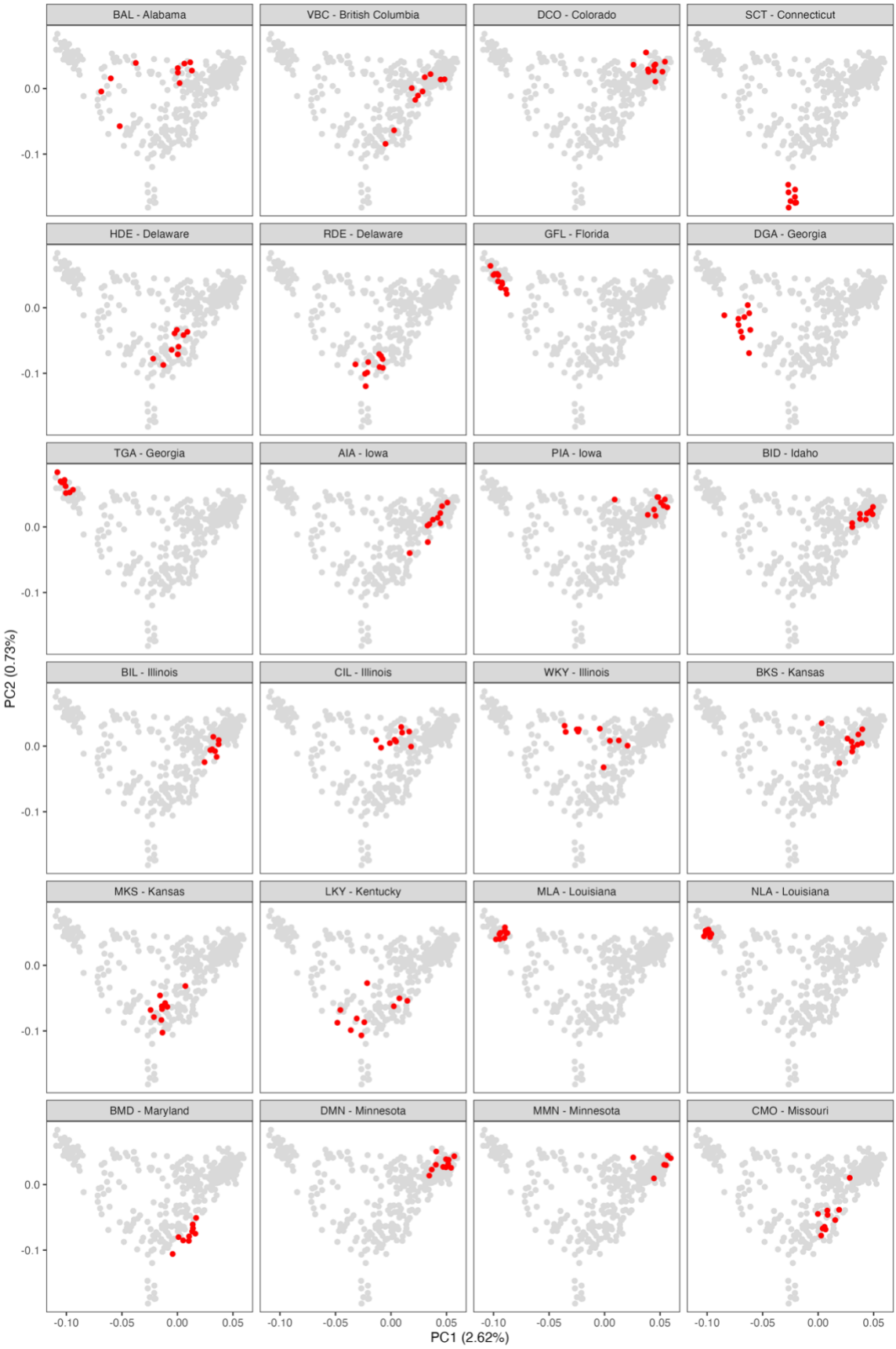


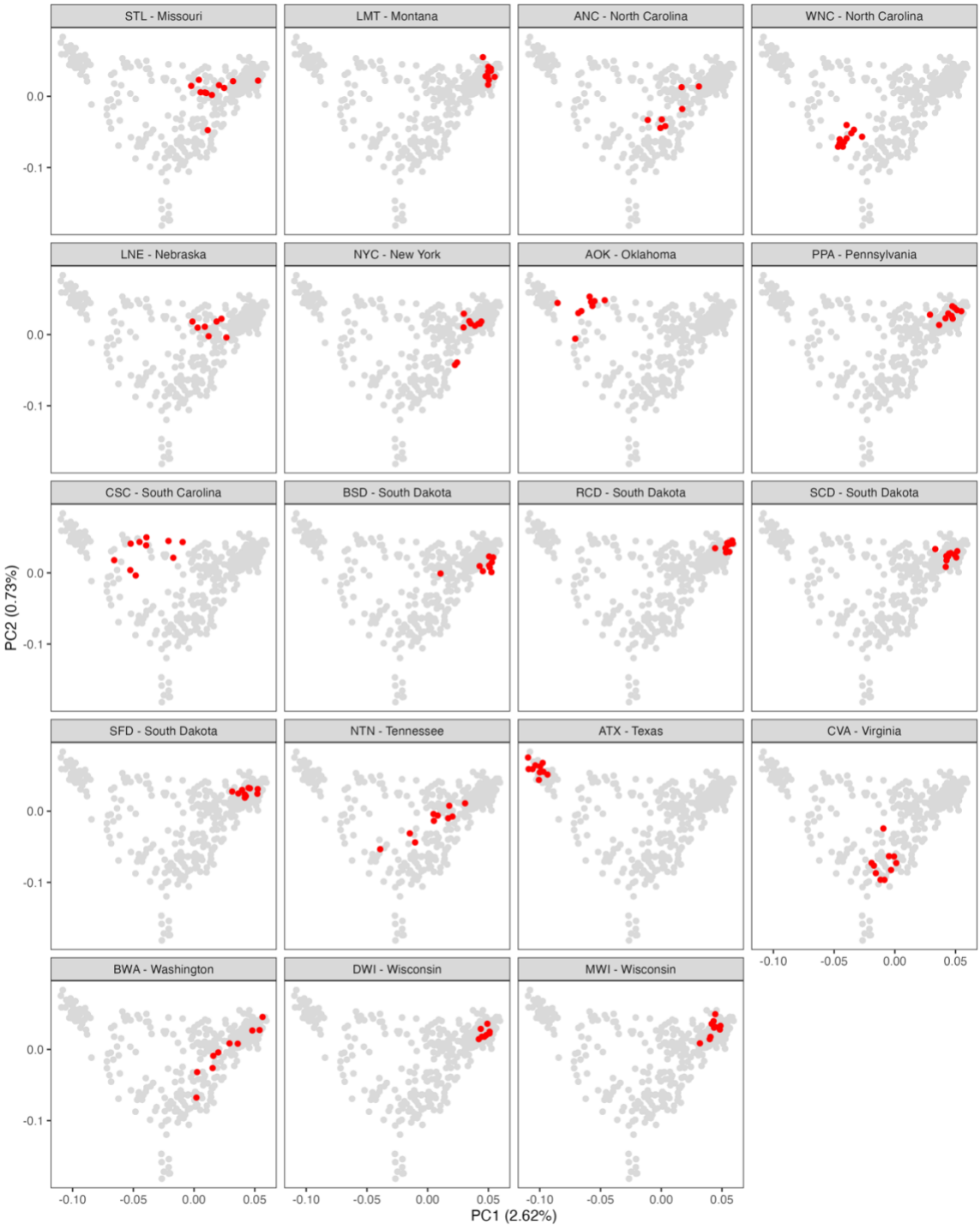
Supplementary Figure S3.8 Permutation test of AMOVA.

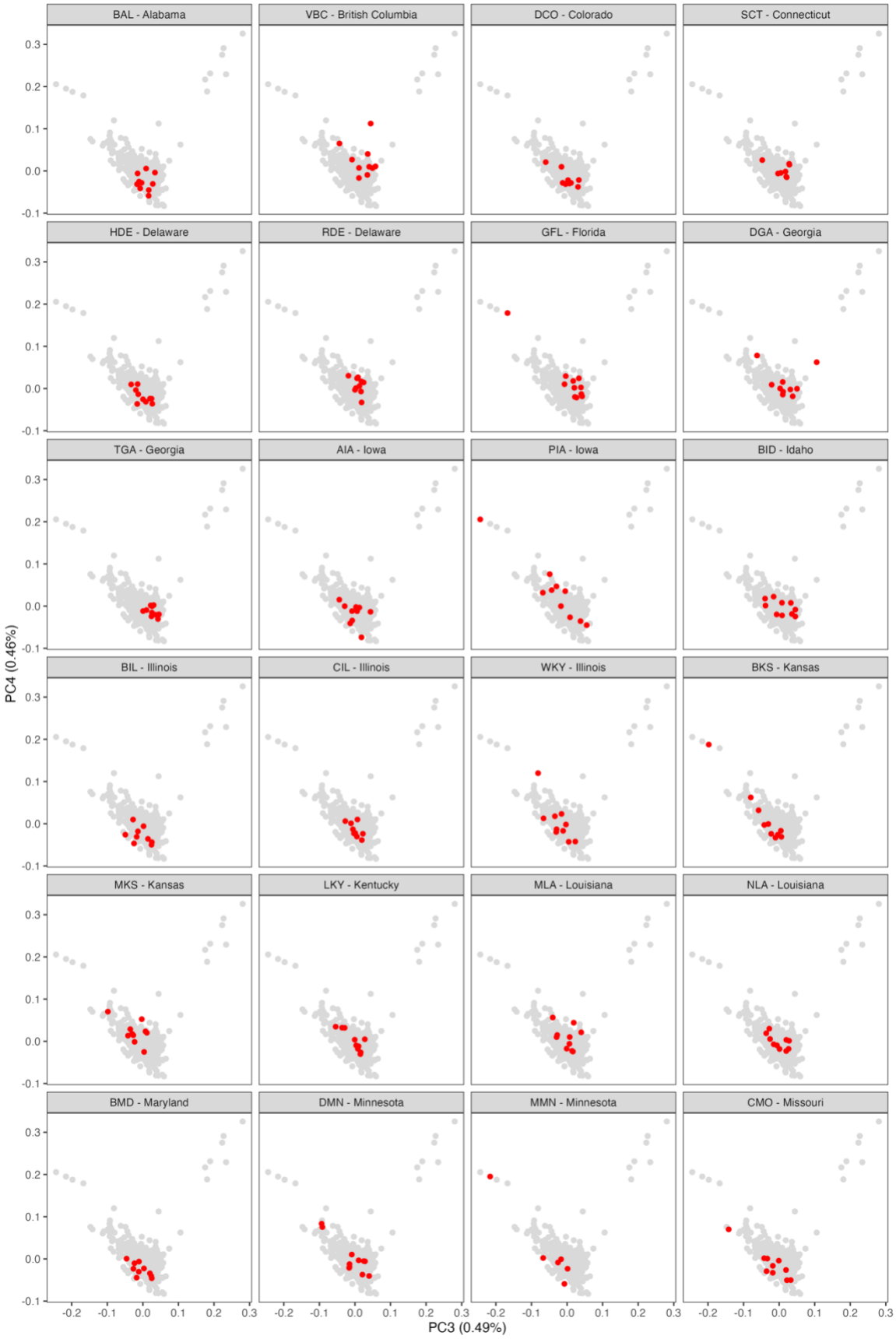


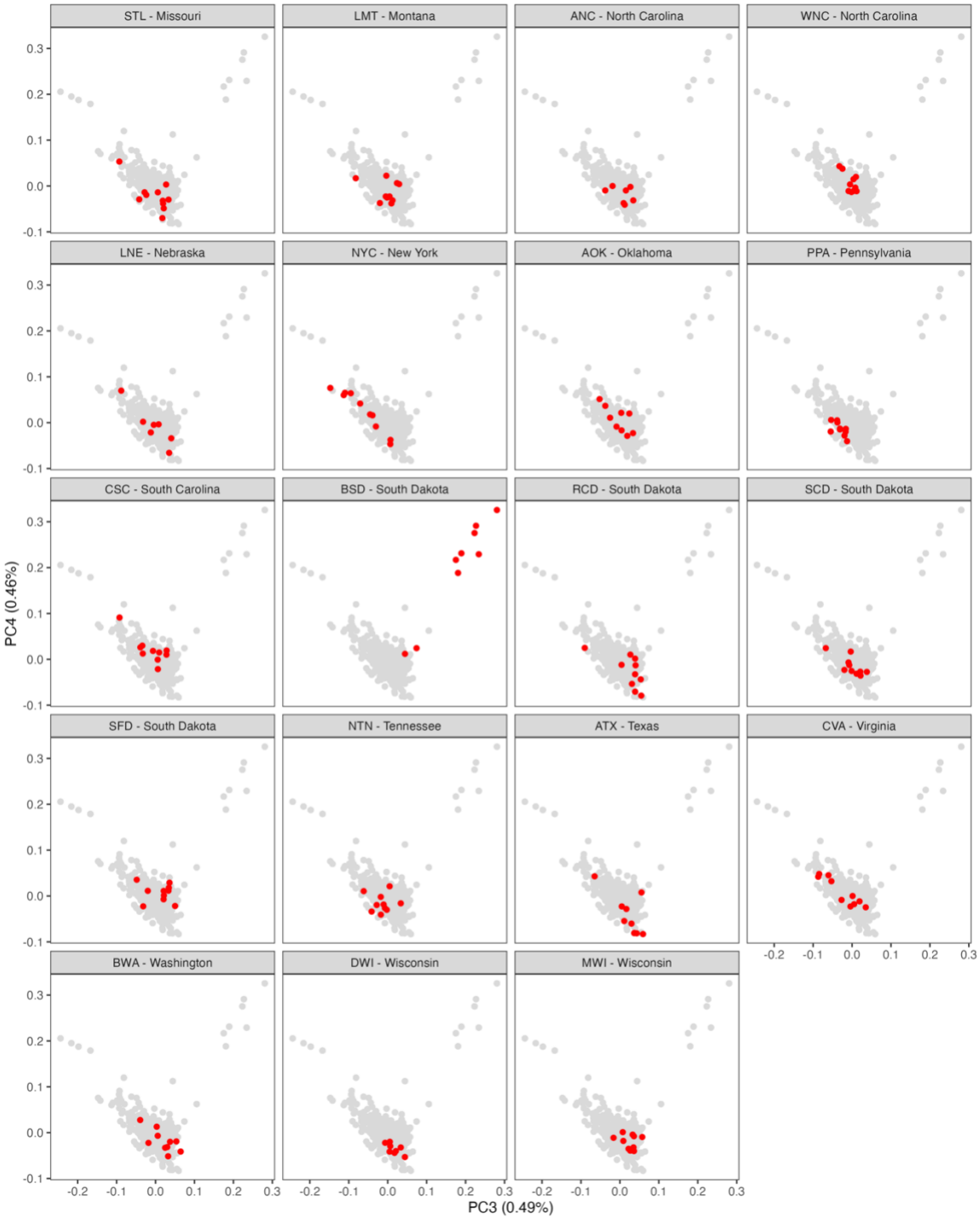
Supplementary Figure S3.9 Environmental PCA.

(A), Projection of the environmental variables of the 415 accessions in PC1 and PC2 space. **(B)**, Association between the climate PC1 and the latitude. **(C)**, Contribution of each environmental variables to the PC1 and PC2 space. The per sample environmental variables (n = 415) were analyzed in PCA (gray dots). Then, the mean values were calculated based on the 43 sampling locations (colored dots).


BAL - Alabama
VBC - British Columbia
DCO - Colorado
SCT - Connecticut
HDE - Delaware
RDE - Delaware
GFL - Florida
DGA - Georgia
TGA - Georgia
AIA - Iowa
PIA - Iowa
BID - Idaho
BIL - Illinois
CIL - Illinois
WKY - Illinois
BKS - Kansas
MKS - Kansas
LKY - Kentucky
MLA - Louisiana
NLA - Louisiana
BMD - Maryland
DMN - Minnesota
MMN - Minnesota
CMO - Missouri
PC2 (0.73%)
PC1 (2.62%)
0.0
-0.1
-0.10
-0.05
0.00
0.05

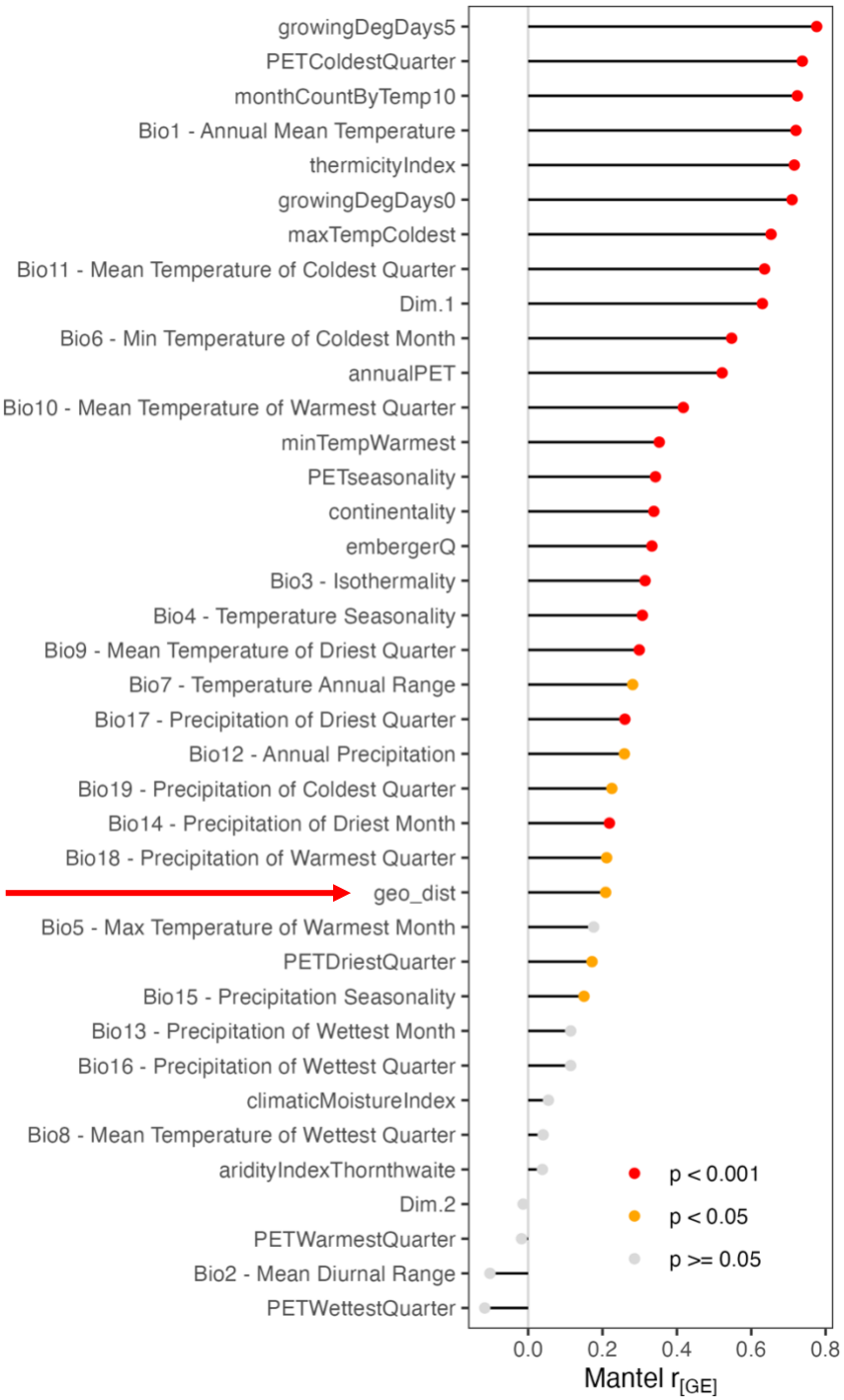

STL - Missouri
LMT - Montana
ANC - North Carolina
WNC - North Carolina
LNE - Nebraska
NYC - New York
AOK - Oklahoma
PPA - Pennsylvania
CSC - South Carolina
BSD - South Dakota
RCD - South Dakota
SCD - South Dakota
SFD - South Dakota
NTN - Tennessee
ATX - Texas
CVA - Virginia
BWA - Washington
DWI - Wisconsin
MWI - Wisconsin
PC2 (0.73%)
PC1 (2.62%)
0.0
-0.1
-0.10
-0.05
0.00
0.05


BAL - Alabama
VBC - British Columbia
DCO - Colorado
SCT - Connecticut
HDE - Delaware
RDE - Delaware
GFL - Florida
DGA - Georgia
TGA - Georgia
AIA - Iowa
PIA - Iowa
BID - Idaho
BIL - Illinois
CIL - Illinois
WKY - Illinois
BKS - Kansas
MKS - Kansas
LKY - Kentucky
MLA - Louisiana
NLA - Louisiana
BMD - Maryland
DMN - Minnesota
MMN - Minnesota
CMO - Missouri
PC4 (0.46%)
PC3 (0.49%)
0.3
0.2
0.1
0.0
-0.1
-0.2



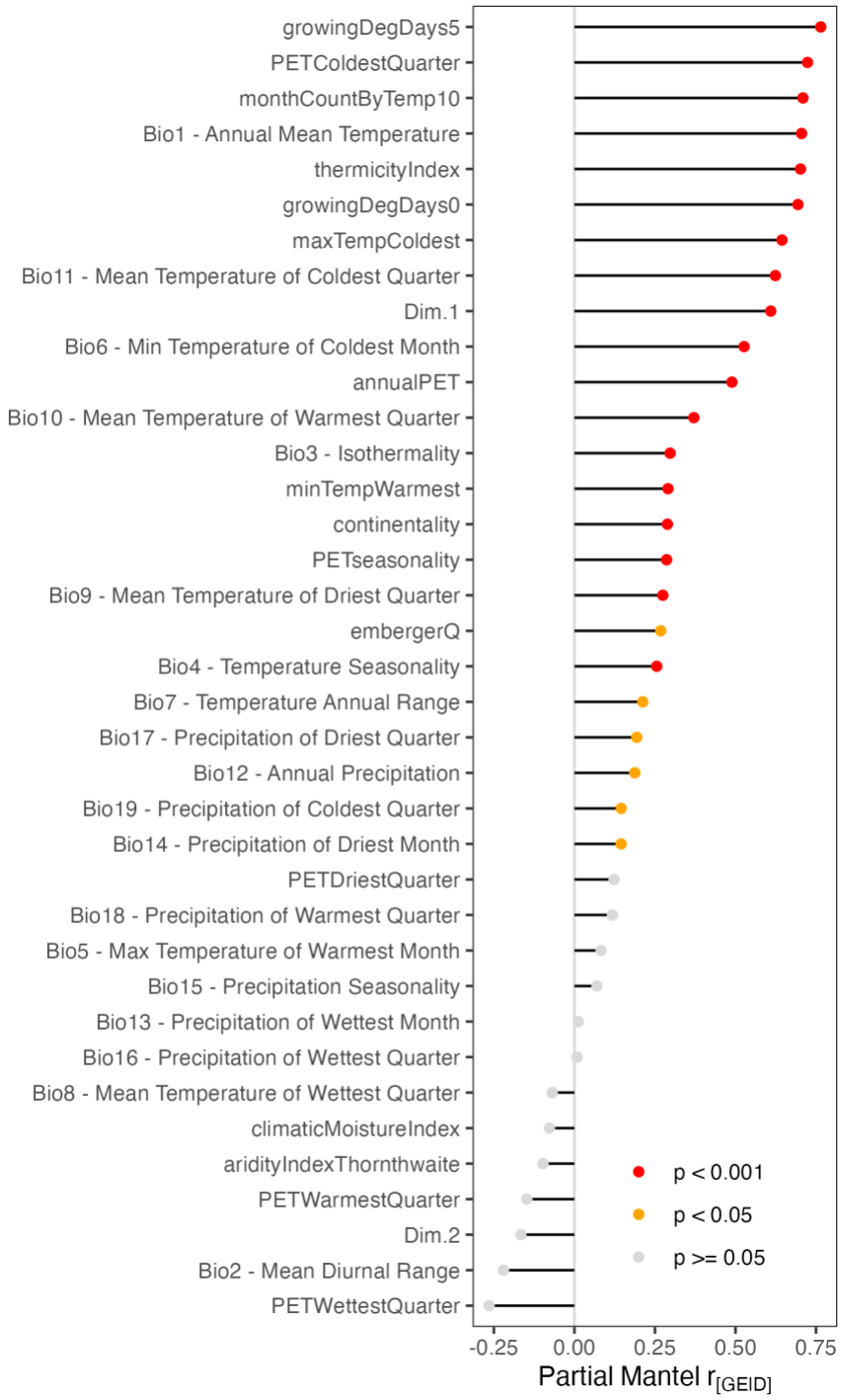
Supplementary Figure S3.10 Genetic principal component analysis (PCA).

For each of the 43 sampling locations (individual panels) accessions from that location are colored in red.



Supplementary Figure S3.11 Mantel test for the correlations between the genetic distance ($F_{ST}/1-F_{ST}$) and the environmental variables.

The p value was estimated from permutation test of 10,000 times. G, genetic distance matrix. E, environmental distance matrix. The geographical distance is highligthed by a red arrow.



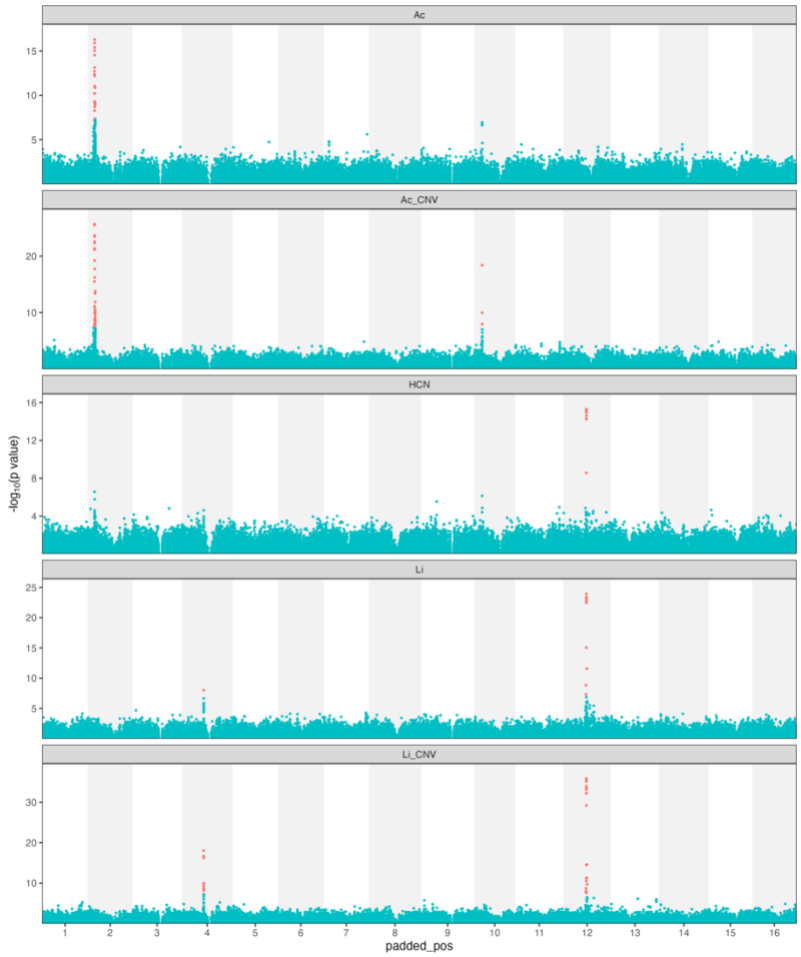
Supplementary Figure S3.12 Partial Mantel test for the correlations between the genetic distance ($F_{ST}/1-F_{ST}$) and the environmental variables with the geographical distance as the control matrix.

The p value was estimated from permutation test of 10,000 times. G, genetic distance matrix. E, environmental distance matrix. D, geographic distance matrix.

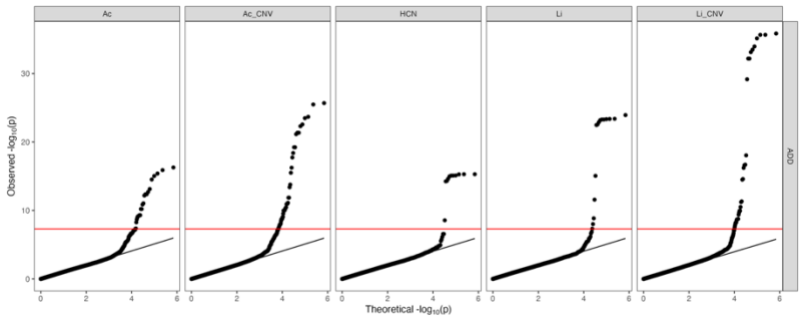
Supplementary Figure S3.13 Association between cyanogenesis trait and environmental variables.

Simple linear model was used for copy number variation (CNV). Logistic model was used for a combined phenotype of the presence of both *Ac* (>1) and *Li* (>1) – being cyanogenic. The environmental variables are ordered by their significances. The significance indicates how the slope is different from 0.

(A)

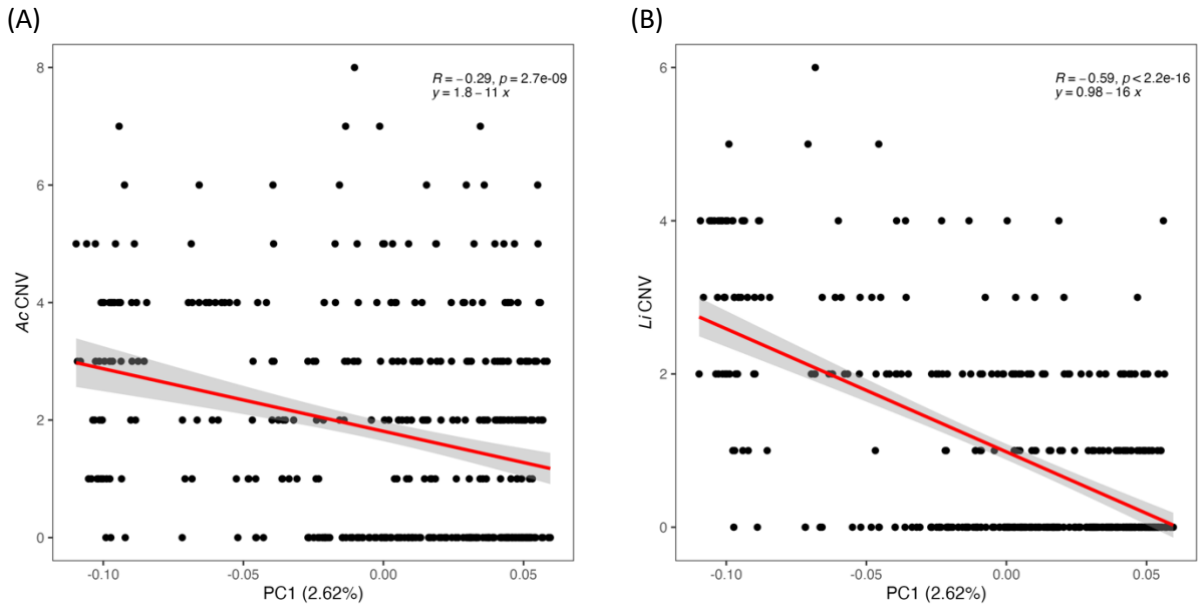


(B)



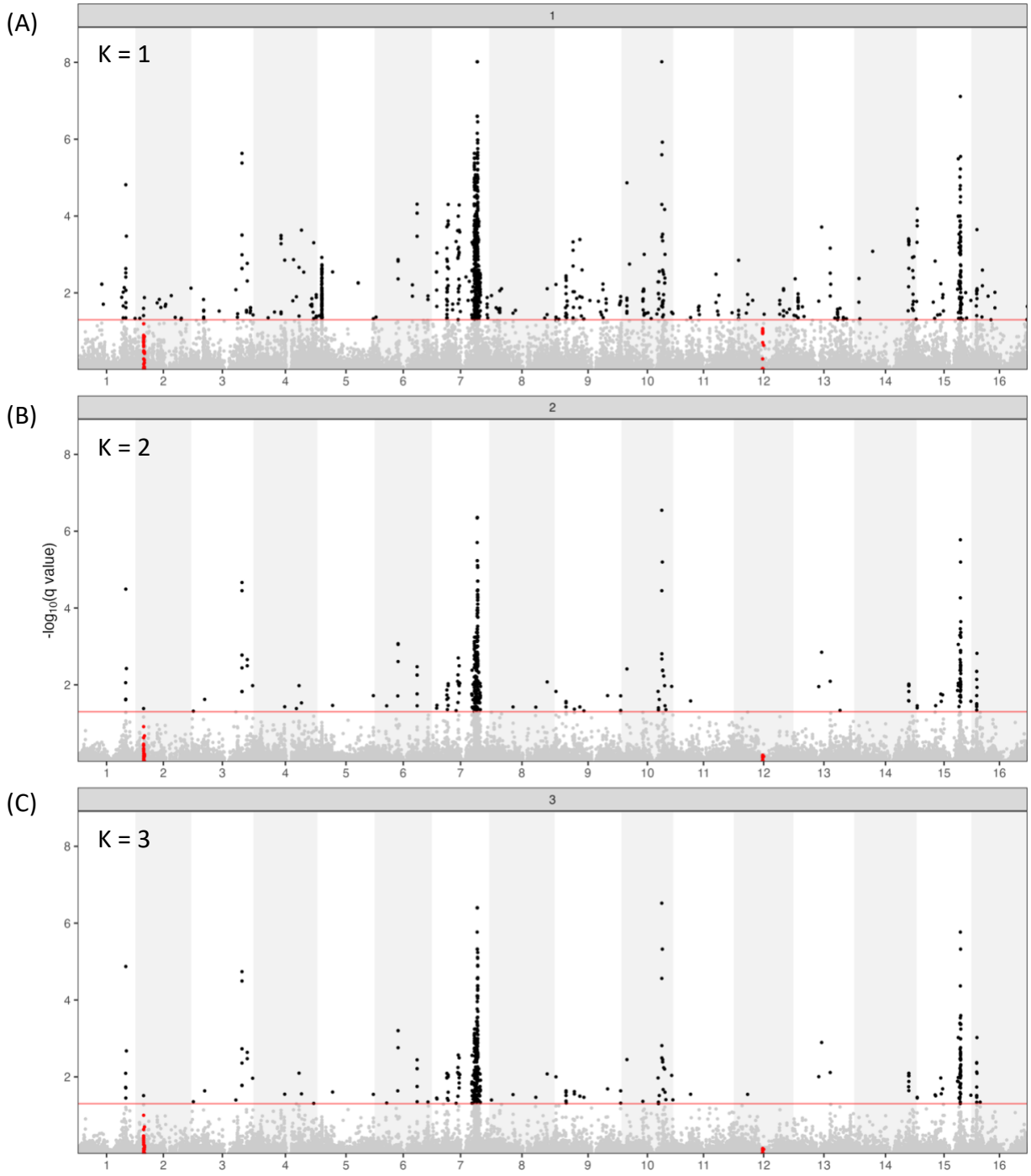
Supplementary Figure S3.14 GWAS of the cyanogenesis traits.

(A), GWAS results for Ac (binary), Ac CNV (numeric), HCN (Ac + Li; binary), Li (binary), Li CNV (numeric). The *p* values lower than the threshold at 5×10^{-8} are colored in red. **(B)**, Q-Q plot for the GWAS results. The red lines show the *p* value threshold at 5×10^{-8}. Ac, the presence of cyanogenic glucosides. Li, the presence of linamarase (cyanogenic glucosidase).



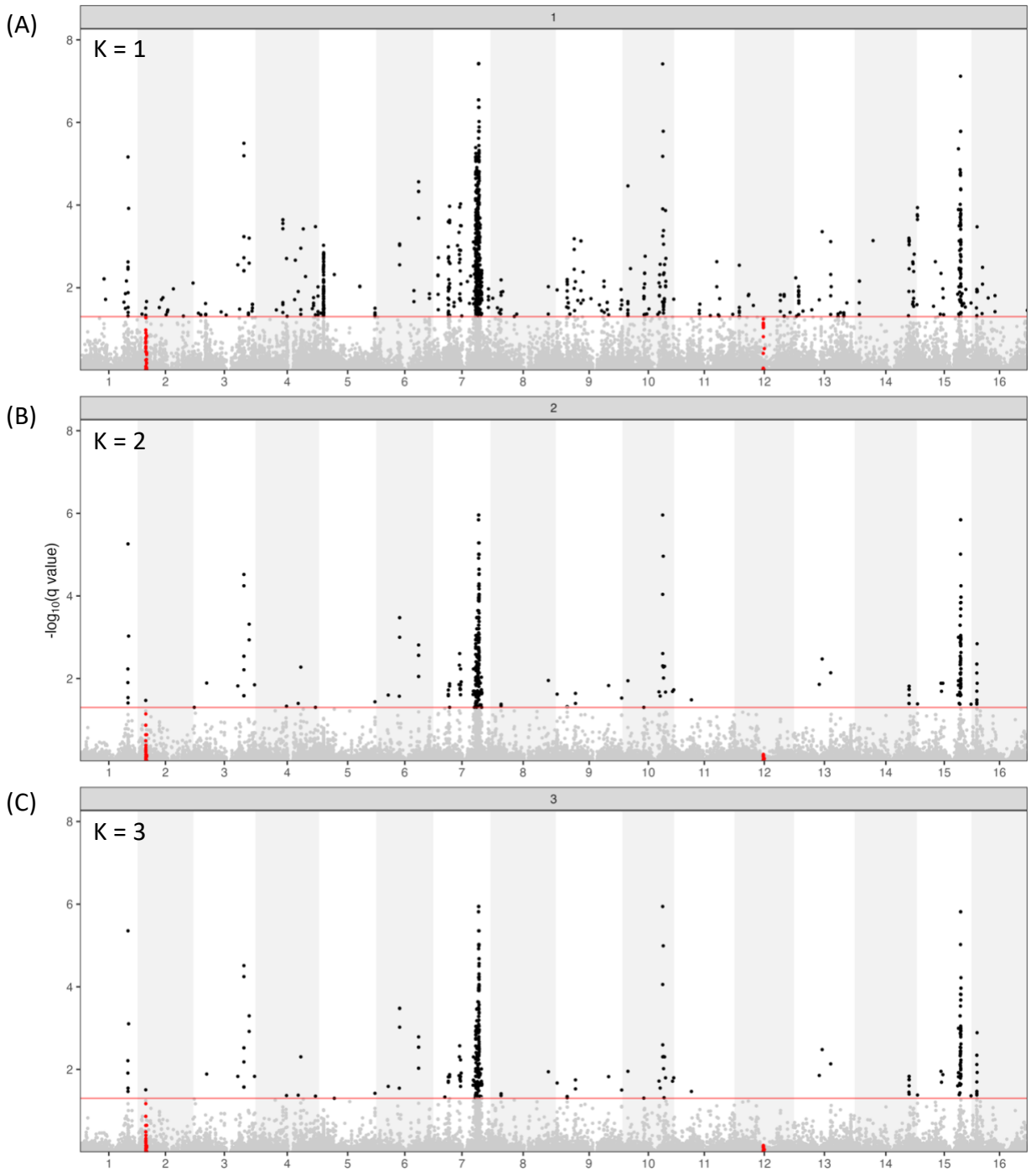
Supplementary Figure S3.15 Correlations between the population structure (PC1) and cyanogenesis CNV.

(A), *Ac* CNV. **(B)**, *Li* CNV.



Supplementary Figure S3.16 LFMM test for association with Growing Degree Days (> 5°C).

The confounding effect of the population structure was controlled by setting the latent factor K = 1, 2, 3. The same SNPs captured for Ac and Li phenotype by GWAS are colored in red.



Supplementary Figure S3.17 Repeat of the LFMM test without BWA and VBC locations (accessions) for association with Growing Degree Days (> 5°C).

The BWA and VBC locations are environmental PCA outliers, which may affect the LFMM output. The confounding effect of the population structure was controlled by setting the latent factor k = 1, 2, 3. The same SNPs captured for Ac and Li phenotype by GWAS are colored in red.

3.12 Supplementary Reference

Bittner-Eddy PD, Crute IR, Holub EB, Beynon JL. 2000. RPP13 is a simple locus in Arabidopsis thaliana for alleles that specify downy mildew resistance to different avirulence determinants in Peronospora parasitica. *The Plant Journal* **21**(2): 177-188.

Briggs GC, Mouchel ClF, Hardtke CS. 2006. Characterization of the Plant-Specific BREVIS RADIX Gene Family Reveals Limited Genetic Redundancy Despite High Sequence Conservation. *Plant Physiology* **140**(4): 1306-1316.

Caro E, Stroud H, Greenberg MVC, Bernatavichute YV, Feng S, Groth M, Vashisht AA, Wohlschlegel J, Jacobsen SE. 2012. The SET-Domain Protein SUVR5 Mediates H3K9me2 Deposition and Silencing at Stimulus Response Genes in a DNA Methylation–Independent Manner. *PLOS Genetics* **8**(10): e1002995.

Chandler JW. 2018. Class VIIIb APETALA2 Ethylene Response Factors in Plant Development. *Trends Plant Sci* **23**(2): 151-162.

Hou S, Liu D, Huang S, Luo D, Liu Z, Xiang Q, Wang P, Mu R, Han Z, Chen S, et al. 2021. The Arabidopsis MIK2 receptor elicits immunity by sensing a conserved signature from phytocytokines and microbes. *Nature Communications* **12**(1): 5494.

Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C. 2000. Transcriptional repression by AtMYB4 controls production of UV‐protecting sunscreens in Arabidopsis. *The EMBO Journal* **19**(22): 6150-6161.

Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. 2002. Multiple Roles of Arabidopsis *VRN1* in Vernalization and Flowering Time Control. *Science* **297**(5579): 243-246.

Li G, Siddiqui H, Teng Y, Lin R, Wan XY, Li J, Lau OS, Ouyang X, Dai M, Wan J, et al. 2011. Coordinated transcriptional regulation underlying the circadian clock in Arabidopsis. *Nature Cell Biology* **13**(5): 616-622.

Lin R, Wang H. 2004. Arabidopsis *FHY3*/*FAR1* gene family and distinct roles of its members in light control of Arabidopsis development. *Plant Physiology* **136**(4): 4010-4022.

Ma L, Li G. 2018. FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) family proteins in Arabidopsis growth and development. *Frontiers in Plant Science* **9**.

Meena MK, Ghawana S, Sardar A, Dwivedi V, Khandal H, Roy R, Chattopadhyay D. 2015. Investigation of genes encoding calcineurin B-like protein family in legumes and their expression analyses in chickpea (Cicer arietinum L.). *PLoS One* **10**(4): e0123640.

Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES. 2000. The molecular basis of vernalization: The central role of *FLOWERING LOCUS C* (*FLC*). *Proceedings of the National Academy of Sciences of the United States of America* **97**(7): 3753-3758.

Stahl E, Fernandez Martin A, Glauser G, Guillou M-C, Aubourg S, Renou J-P, Reymond P. 2022. The MIK2/SCOOP Signaling System Contributes to Arabidopsis Resistance Against Herbivory by Modulating Jasmonate and Indole Glucosinolate Biosynthesis. *Frontiers in Plant Science* **13**.

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999. Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature* **400**(6745): 667-671.

Tan ST, Dai C, Liu HT, Xue HW. 2013. Arabidopsis casein kinase1 proteins CK1.3 and CK1.4 phosphorylate cryptochrome2 to regulate blue light signaling. *The Plant Cell* **25**(7): 2618-2632.

Thornton LE, Rupasinghe SG, Peng H, Schuler MA, Neff MM. 2010. Arabidopsis CYP72C1 is an atypical cytochrome P450 that inactivates brassinosteroids. *Plant Molecular Biology* **74**(1): 167-181.

Uhlken C, Horvath B, Stadler R, Sauer N, Weingartner M. 2014. MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in Arabidopsis thaliana. *Plant J* **78**(1): 107-120.

van der Vossen E, Sikkema A, Hekkert B, Gros J, Stevens P, Muskens M, Wouters D, Pereira A, Stiekema W, Allefs S. 2003. An ancient R gene from the wild potato species Solanum bulbocastanum confers broad-spectrum resistance to Phytophthora infestans in cultivated potato and tomato. *Plant J* **36**(6): 867-882.

Verhertbruggen Y, Marcus SE, Chen J, Knox JP. 2013. Cell wall pectic arabinans influence the mechanical properties of Arabidopsis thaliana inflorescence stems and their response to mechanical stress. *Plant Cell Physiol* **54**(8): 1278-1288.

Wang J, Zhang L, Li J, Lawton-Rauh A, Tian D. 2011. Unusual signatures of highly adaptable R-loci in closely-related Arabidopsis species. *Gene* **482**(1-2): 24-33.

Wang L, Yang T, Lin Q, Wang B, Li X, Luan S, Yu F. 2020. Receptor kinase FERONIA regulates flowering time in Arabidopsis. *BMC Plant Biol* **20**(1): 26.

Wang X, Niu Y, Zheng Y. 2021. Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. *Int J Mol Sci* **22**(11).

Yang Y, Zhang Y, Ding P, Johnson K, Li X, Zhang Y. 2012. The Ankyrin-Repeat Transmembrane Protein BDA1 Functions Downstream of the Receptor-Like Protein SNC2 to Regulate Plant Immunity. *Plant Physiology* **159**(4): 1857-1865.

Zhou JX, Liu ZW, Li YQ, Li L, Wang B, Chen S, He XJ. 2018. Arabidopsis PWWP domain proteins mediate H3K27 trimethylation on FLC and regulate flowering time. *Journal of Integrative Plant Biology* **60**(5): 362-368.

Chapter 4 Genetics and plasticity of white leaf mark variegation in white clover (*Trifolium repens* L.)

4.1 Authorship and Affiliations

Wen-Hsi Kuo[1], Eimear Cunningham[1], Emily Guo[1], Kenneth M. Olsen[1*]

[1]Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA

This chapter is under review in *Annals of Botany*

4.2 Abstract

Background and Aims

Leaf variegation is common in plants and confers diverse adaptive functions. However, its genetic underpinnings remain largely unresolved; this is particularly true for variegation that arises through variation in leaf tissue structure that affects light reflection. White clover is naturally polymorphic for structure-based white leaf mark variegation. It therefore provides a useful system to examine the genetic basis of this phenotype, and to assess potential costs to photosynthetic efficiency resulting from modified leaf structure. This study sought to map the loci controlling the white leaf mark in white clover and evaluate the relationship between white leaf mark, leaf thickness, and photosynthetic efficiency.

• Methods

We generated a high-density genetic linkage map from an F_3 mapping population, employing reference genome-based SNP markers. White leaf mark was quantified through detailed phenotypic evaluations alongside leaf thickness to test how tissue thickness may affect the variegation phenotype. Quantitative trait locus (QTL) mapping was performed to characterize their genetic bases. Photosynthetic efficiency measurements were conducted to test for physiological trade-offs between variegation and photosynthetic output.

• Key Results

The *V* locus, a major gene responsible for the white leaf mark polymorphism, was mapped to chromosome 5, and several modifier loci were also mapped that contribute

additively to variegation intensity. The presence of white leaf mark was correlated with increased leaf thickness; however, increased variegation did not detectably affect photosynthetic efficiency.

• Conclusions

We have successfully mapped the major locus governing the white leaf mark in white clover, along with several modifier loci, revealing a complex basis for this structure-based variegation. The apparent absence of compromised photosynthesis in variegated leaves challenges the notion that variegation creates fitness trade-offs between photosynthetic performance and other adaptive functions. This finding suggests that other factors may maintain the white leaf mark polymorphism in white clover.

Keywords

Leaf variegation; white clover (*Trifolium repens* L.); leaf mark polymorphism; leaf thickness; plasticity; *V* locus; photosynthetic efficiency; quantitative trait locus (QTL) mapping

4.3 Introduction

Polymorphism in leaf variegation, characterized by heterogeneous color patterns on leaves, is observed across diverse natural and cultivated settings. Leaf variegation may arise from pigment-related factors, such as anthocyanin accumulation, or from structural attributes of tissue organization, such as the expansion of intercellular air spaces which can alter light reflection (Zhang *et al.*, 2020). Beyond its horticultural appeal, various adaptive functions and mechanisms underlying leaf variegation polymorphism have been suggested. The presence of leaf variegation could facilitate herbivore deterrence (Charles, 1968; Cahn & Harper, 1976b; Campitelli *et al.*, 2008; Soltau *et al.*, 2008; Lev-Yadun, 2014), aid in thermoregulation (Ganders *et al.*, 1980; Shelef *et al.*, 2019), provide UV protection (Koski & Ashman, 2015), and/or enhance pollinator attraction (Song *et al.*, 2018). Conversely, the presence of leaf variegation could compromise photosynthetic efficiency due to a reduction in light absorption or reduced chloroplast content within tissue (reviewed by Menzies *et al.*, 2015). Despite its potential roles in adaptation, the genetic mechanisms underlying naturally occurring leaf variegation, especially structure-based coloration, are largely unknown (Zhang *et al.*, 2020).

White clover (*Trifolium repens* L.) serves as an ideal system to investigate the genetic basis of leaf variegation and its potential adaptive value. This species is naturally polymorphic for several leaf variegation traits, including white leaf mark, red fleck, and red midrib (Brewbaker, 1955; Carnahan *et al.*, 1955; Brewbaker & Carnahan, 1956; Corkill, 1971). In addition, these polymorphisms are found not only in white clover but also in several other *Trifolium* species (Attila, 1987; Tan & Collins, 1987; Abdi *et al.*, 2020), as well as the closely

related genus *Medicago* (Mccomb, 1974). The study of leaf variegation in white clover could therefore have generalizability beyond this single species.

Classical genetics studies in white clover over the last 70 years have revealed that the white leaf mark polymorphism, a structure-based variegation pattern, is governed primarily by a single locus (the *V* locus); the pigment-based color polymorphisms, including red fleck and red midrib, are governed by an unlinked locus (*R* locus). The presence of either variegation pattern is dominant to its absence, with multiple dominant alleles at both loci conferring slightly different variegation patterns (Brewbaker, 1955; Carnahan *et al.*, 1955; Corkill, 1971). Although the *R* locus has been successfully mapped (Barrett *et al.*, 2004; Tashiro *et al.*, 2010), the genomic location of the white mark *V* locus has remained unknown despite recent mapping attempts (Tashiro *et al.*, 2010). The lack of success in mapping the *V* locus thus far could be attributable to the limited resolution of the microsatellite-based linkage map markers used until recently in white clover (Barrett *et al.*, 2004; Tashiro *et al.*, 2010; Isobe *et al.*, 2012; Griffiths *et al.*, 2013), as well as the plasticity of expression of the leaf mark phenotype (Postma & Agren, 2016), which can challenge accurate trait scoring. As a structure-based variegation trait, the intensity of the white leaf mark may be specifically related to leaf thickness, which can affect the extent to which there are enlarged intercellular or sub-epidermal spaces (Carnahan *et al.*, 1955).

Leveraging recent advances in genomic resource availability for white clover (Santangelo *et al.*, 2023; Kuo *et al.*, 2024), we revisited this question using a combination of genome-wide

high-density SNP markers and comprehensive phenotyping data across developmental stages in an F_3 mapping population. Our study had three objectives. **Firstly**, we aimed to document the white leaf mark polymorphism using a quantitative measure of the phenotype, accounting for the plasticity associated with leaf developmental stages and intra-individual variation, and assessing its potential relationship to leaf thickness. **Secondly**, we aimed to map the QTLs for white leaf mark and leaf thickness. If the presence of white leaf mark is associated with greater leaf thickness, we would expect QTLs for the two phenotypes to show some degree of overlap. **Thirdly**, we aimed to test for any evidence of fitness impacts associated with the presence of white leaf mark, since in principle the white leaf mark could compromise photosynthetic efficiency due to a reduction of mesophyll cell density or chlorophyll content within the white leaf mark.

4.4 Materials and Methods

Mapping population

An F_3 mapping population ('DG', n = 500) was constructed from a cross of two wild North American white clover samples, DMN_010 (BioSample: SAMN34157026) and GFL_007 (BioSample: SAMN37329216, SAMN34116011). DMN_010 has no white leaf marks on any leaves at any developmental stages. In contrast, GFL_007 has white leaf marks on all leaves at all developmental stages. The two parents were cross-pollinated bidirectionally by hand. An F_2 population was constructed by intercrossing the F_1 generation (N=50-100 plants) with solitary bees (Crown Bees, Woodinville, WA) in enclosed cages in the greenhouse (Olsen *et al.*, 2021;

Wright *et al.*, 2022). The F_3 population (N=500) was constructed by randomly intercrossing the F_2 population by hand. The gametophytic self-incompatibility feature of white clover prevented plants from self-pollination while allowing some sib-crossing to occur. All plants were cultivated in the Washington University greenhouse facilities under standard conditions (16 h: 8 h, light: dark photoperiod with 140 µmol supplemental lighting as needed; diurnal temperature range of approximately 18-28°C).

Genotyping

Library preparation and sequencing

Genomic DNA was extracted from 100 mg young leaf tissue samples from F_3 plants. Liquid nitrogen-frozen leaf samples were homogenized with steel beads using a homogenizer. Then, the DNA was extracted using a protocol modified from Whitlock *et al.* (2008). Genotyping-by-sequencing (GBS) library preparation followed a modification of the protocol of Elshire *et al.* (2011). Specifically, the DNA (100 ng) was digested with *ApeK*I and ligated to our customized barcoding adaptors (Olsen *et al.*, 2021). The ligation product was cleaned and size-selected by AMPure XP beads (1.2× volume beads were added to the ligation product) (Beckman Coulter, Brea, CA, USA). Then, each sample was individually used in PCR amplification. Each amplified product was quantified by Qubit dsDNA HS Assay Kits (Thermo Fisher Scientific, Waltham, MA, USA), and samples were adjusted to the same concentration before pooling. Then, a final clean-up and size selection were conducted by 0.8× volume AMPure XP beads. Paired-end sequencing (150-bp reads) was performed with each library in a

unique lane using the Illumina Hi-Seq 2500 platform (Novogene Corp., Chula Vista, CA, USA). The detailed step-by-step GBS protocol is provided in **Supplementary file 1**.

SNP calling and filtering

The GBS SNP dataset was generated by the fast-GBS v2.0 pipeline (Torkamaneh *et al.*, 2020). A minimum of six reads was set as the required minimum for a SNP call. The raw output was pre-filtered to exclude all SNPs with >0.5 missing rate. Remaining missing data were imputed by Beagle 5.4 with default settings (Browning *et al.*, 2021). Then, the imputed output was filtered for homozygosity in the parental accessions, minor allele frequency >0.2, and *p*-value >0.01 in a Hardy-Weinberg genotype frequency test (indicating no significant deviation from 1:2:1 segregation in the F_3 populations).

Phenotyping

Leaf mark phenotyping

The white leaf mark trait was independently phenotyped three times for each accession (summer 2020, spring 2021 and summer 2022) by two different people in order to eliminate potential biases caused by different developmental stages, intra-individual variation, and human error. The leaf mark trait was categorized into 5 ranks: “0”, no observable leaf mark on any leaves at any time; “1”, weakly visible leaf mark on some or all leaves in at least one phenotyping trial; “2”, weak and medium-intensity leaf marks on some or all leaves in all trials; “3”, medium and strongly visible leaf marks on some or all leaves in all trials; “4”, strongly visible leaf mark on all leaves in all trials. See **Fig. 1A** for representative photos of leaf mark

variation among individual leaves; see **Supplementary Fig. S1** for the examples of leaf mark ranks for different F_3 plants. Any data incongruencies between the measurements were rechecked to minimize phenotyping error.

Leaf thickness phenotyping

To assess the correlation between leaf thickness and white leaf mark intensity, leaf thickness was measured after leaf saturation in distilled water overnight, a point at which cells contained their maximum water content. The protocol was modified from Afzal *et al.* (2017). Four healthy and mature leaves were cut from each plant, and the petioles with the cut surfaces were immediately placed in distilled water. The leaflets were kept above the water to avoid water infiltration into intercellular spaces. The leaves and the distilled water were contained in a 50 mL centrifuge tube with lid, and the sealed tubes were kept in the dark at 4°C overnight for complete saturation. The next day, the leaves were blotted dry with tissue paper. The lamina thickness of the top leaflet of a trifoliate leaf was then measured using a Thickness Gauge (CODE NO. 547-526S; Mitutoyo, Japan). The midrib was avoided for leaf thickness measurements.

Leaf physiology

To investigate the physiological effects of different levels of white leaf mark, we randomly sampled 33 individuals from the DG F_3 mapping population (the same population used in QTL mapping) and 34 individuals from a second F_3 mapping population, 'GS'. The GS F_3 mapping population was created using the same crossing protocol as DG but with the parental

genotypes GFL_007 and STL_0701 (BioSample: SAMN34157027). Like, DMN_010, the parent STL_0701 has no leaf marks on any leaves at any developmental stages. At least 3 leaves per individual were measured for all phenotyped accessions. In total, measurements were made for 102 individuals from the DG F_3 population and 98 individuals from the GS F_3 population. Measurements, including net photosynthesis rate, stomatal conductance, intercellular CO_2 concentration, and water use efficiency (WUE), were using a LICOR-6400XT system (LI-COR Biosciences, Lincoln, NE, USA) in summer 2021 at 10 am - 3 pm with LED light source (1100 µmol m^{-2} s^{-1}) and controlled CO_2 flow (400 µmol s^{-1}). The block temperature (23 - 30°C) and reference relative humidity (40 – 60%) were set to the ambient environment. After the LICOR-6400XT measurement, the leaf was cut from the plant and the leaf thickness was measured immediately. Then, the leaf water potential was measured by a PMS pressure chamber (Model 1000; Albany, OR, USA). The leaf area was measured by scanned images. All raw data for photosynthesis-related measurements are available in **Supplementary file 2**.

QTL mapping

Prior to QTL mapping, the genetic distance between all SNP markers was estimated by ASMap (Taylor & Butler, 2017), and any markers showing abnormal recombination rates (>150 cM) in juxtaposition were manually removed. The chromosomal locations and physical order of the markers were based on our recently generated high-quality reference genome (Kuo *et al.*, 2024). QTL mapping was conducted in R/qtl package using the normal model and the EM method with 1,000 permutations (Broman *et al.*, 2003). The genetic map with phenotypes in "cross" object (CSV format) is available in **Supplementary file 3.**

4.5 Results

Leaf phenotypes

Based on the phenotyping protocol, the DG parental accession GFL_007 was classified for white leaf mark as rank 3 (medium and strong leaf marks on some or all leaves at all developmental stages) while the other parental accession, DMN_010, was classified as rank 0 (no observable leaf mark on any leaves at any time) (**Fig. 4.1A**). In the F_3 population, there were 197 plants scored as rank 0 and 303 plants showing different degrees of the white leaf mark ranging from rank 1 to 4 (**Fig. 4.1B**). We found that individuals with the leaf mark present (rank > 0) varied in their expressivity of the phenotype (i.e., the extent to which the white leaf mark was consistently expressed across all leaves at all developmental stages in a plant). In general, plants at early developmental stages, when the young leaves were rarely covered by other leaves, had the most clearly observable white leaf mark patterns. In the later developmental stages, white leaf mark usually became less observable, but it could be restored by vigorous pruning, which exposed the re-growing young leaves to the light again (**Supplementary Fig. S4.2**). This observation suggests that expressivity of white leaf mark is dependent on light exposure during leaf development.

We examined the leaf thickness of the plants with different white leaf mark levels to test the hypothesis that the intensity of leaf mark is associated with greater leaf thickness. We measured the leaf thickness from 396 plants of the DG F_3 population and the two parental accessions. We found that the leaf thickness of the parental accession GFL_007 (179.33 ± 9.45 μm; mean $\pm$ SD; n =15) is significantly greater than that of the DMN_010 parent (133.93 ± 9.67

μm; n = 15) (t = -13.01, df = 28, p = 2.19 × 10^{-13}). In their F_3 progeny, leaf thickness has a high broad-sense heritability (H^2 = 0.61) and is significantly different between the plants with different white leaf mark ranks ($F_{(2, 393)}$ = 6.17, p = 2.31 × 10^{-3}). Specifically, plants with a more intense white leaf mark have greater leaf thickness (rank 4, 175.52 ± 17.31 μm) than the plants with medium-intensity (rank 1 + 2 +3, 164.47 ± 17.25 μm) or no leaf mark (rank 0, 165.05 ± 15.91 μm).

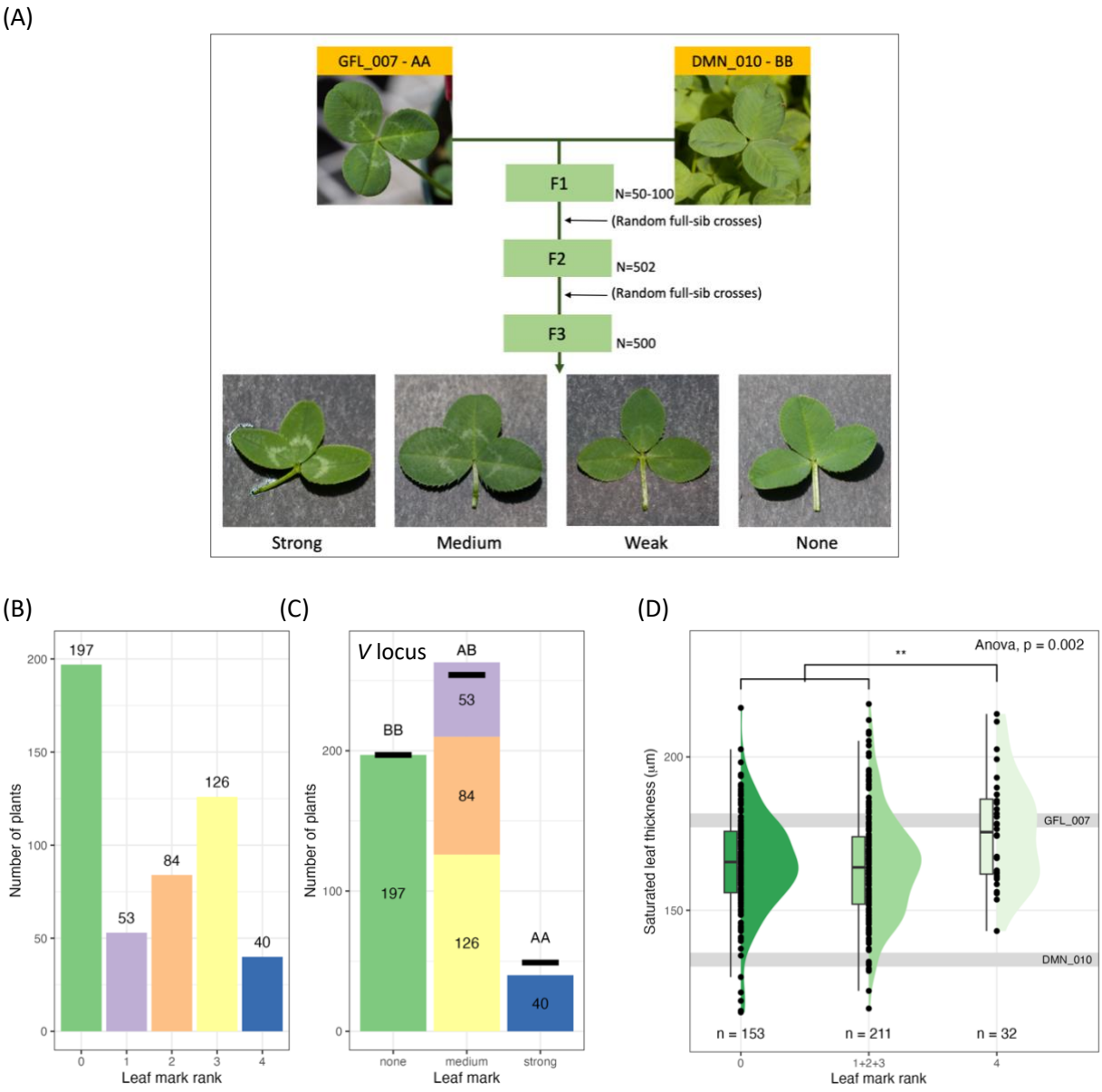


Figure 4.1 Phenotypic variation in white leaf mark and leaf thickness.

(A), DG mapping population design and examples of different intensities of the white leaf mark. **(B)**, Numbers of plants of the different leaf mark ranks. **(C)**, Observed and expected numbers of plants of different leaf mark ranks based on the genotype of the *V* locus (chr_05.48106478), where the A allele is from GFL_007 and the B allele is from DMN_010. Horizontal black bars indicate observed counts for each *V* locus genotype. **(D)**, Saturated leaf thickness. Each point is the mean value of four leaves from a single plant. The mean leaf thickness of the parents is shown by gray lines.

QTL mapping

Using the F_3 progeny of the DG cross, we generated a high-density linkage map with 6343 segregating SNP markers (28,893.88 cM, 4.55 cM/marker). For the white leaf mark polymorphism, QTL mapping detected a major QTL on Chr. 5 [chr_05:48106478, LOD = 134.87, Percentage of variance explained (PVE) = 70.94%]. We infer that this major QTL corresponds to the previously described but unmapped *V* leaf mark locus (Brewbaker, 1955; Carnahan et al., 1955; Brewbaker & Carnahan, 1956; Corkill, 1971). In addition, we detected six minor-effect modifier QTLs on different chromosomes (LOD = 5.00-8.12, PVE = 4.50 - 7.21%). The seven QTLs can jointly explain 72.04% of the total variance. For all seven QTLs, the GFL_007 parental allele (designated the "A" allele) increased the intensity of leaf mark phenotype (**Fig. 4.2A**). We found a high consistency between the leaf mark phenotype and the genotype of the *V* locus; notably, when merging the leaf mark ranks into three categories [no mark (rank 0), medium (rank 1, 2, 3), strong (rank 4)], the number of plants in each category is highly consistent with counts of "AA", "AB", and "BB" genotypes at the *V* locus ($\chi^2 = 6$, $df = 4$, $p = 0.20$) (**Fig. 4.1B**). There were 197 plants classified as no (rank 0) leaf mark, which is exactly the same number as the 197 plants with the "BB" genotype at the *V* locus; there are 263 plants classified as medium (rank 1, 2, 3) leaf mark, which is slightly higher than the 254 plants with the "AB" genotype; and there are 40 plants classified as strong (rank 4) leaf mark, which is slightly lower than the 49 plants with the "AA" genotype. Although there are a few incongruencies between the leaf mark phenotype and the corresponding genotype at the *V* locus (**Supplementary Table S4.1**), this overall pattern demonstrates a clear pattern of incomplete dominance at the *V* locus, and it

confirms the QTL mapping results that the *V* locus alone can account for much of the observed leaf mark variation.

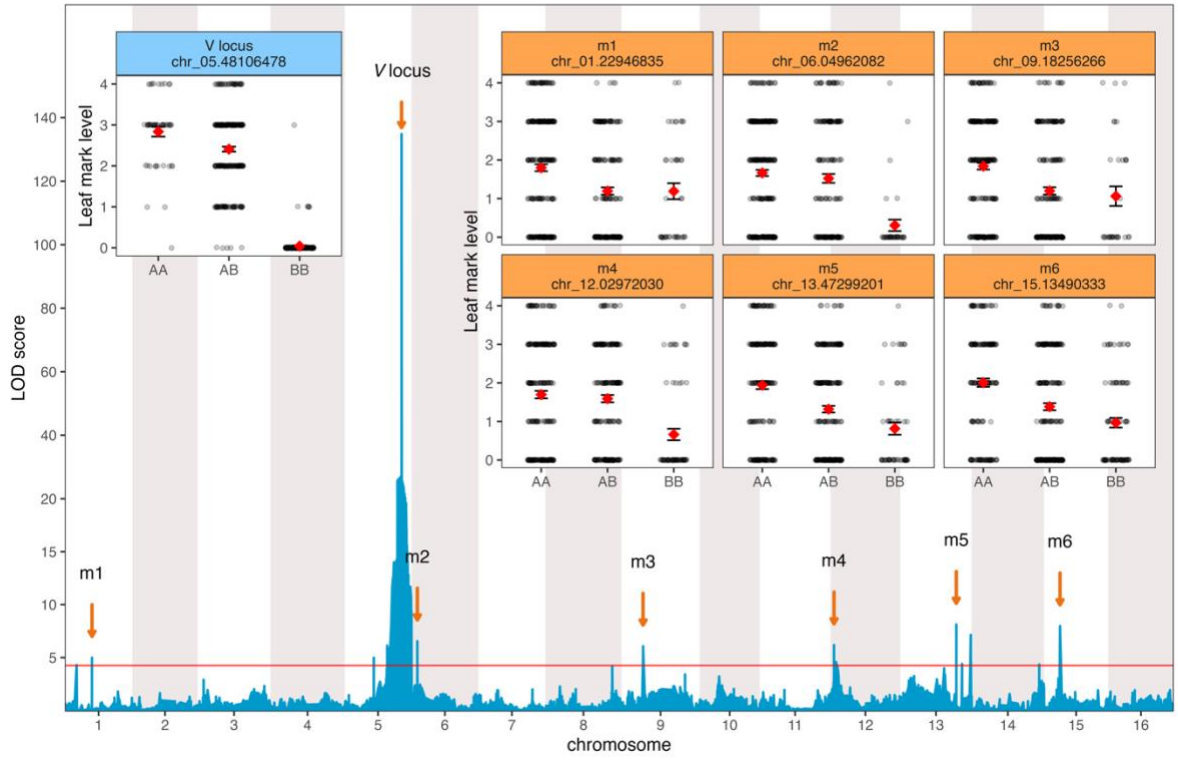
Taking into consideration the phenotypic contributions of the *V* locus plus the six modifier loci, we found that the leaf mark intensity is directly correlated with the number of the "A" alleles across all seven loci; this indicates that the loci contribute additively to the leaf mark trait (**Fig. 4.2B**). This pattern of additivity is especially clear when the *V* locus is heterozygous (**Fig. 4.2C**), suggesting a pattern of incomplete dominance at the major-effect *V* locus, and with the modifier loci collectively contributing to the intensity of the leaf mark trait to a lesser degree.

Based on our finding that the intensity of the leaf mark phenotype is associated with greater leaf thickness (**Fig. 4.1E**), we tested for overlap between the leaf mark QTLs and the leaf thickness QTLs. However, despite a clear difference in the leaf thickness of the parents (**Fig. 4.1E**) and the high broad-sense heritability of this trait in the mapping population ($H^2 = 0.61$), we failed to detect any significant QTLs related to the leaf thickness variation. It is possible that leaf thickness is a polygenic trait, with multiple QTLs contributing a small effect that falls below the significance threshold for detection.

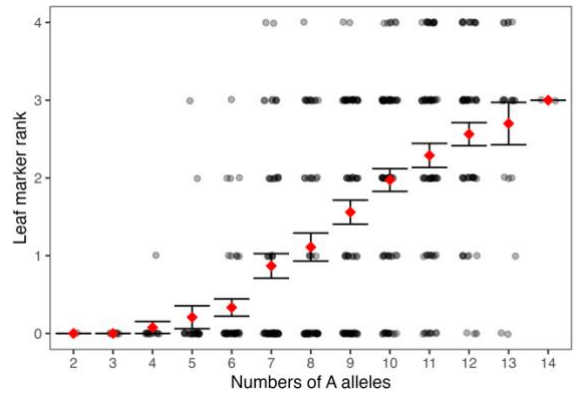
Since we could not directly compare leaf thickness QTLs to the leaf mark QTLs, we instead tested for differences in leaf thickness among the different genotypes of the seven leaf mark QTLs. For the *V* locus, "AA" plants had significantly greater leaf thickness (172.61 ± 15.45 μm) than "BB" plants (165.21 ± 15.55 μm) (**Fig. 4.2D)** ($F_{(2, 393)} = 3.1434$, $p = 4.42 \times 10^{-2}$).

Similarly, the different genotypes of the modifier loci showed significant correlations with leaf thickness, but only when the *V* locus was heterozygous (**Fig. 4.2E**). These findings indicate that there is a correlation between leaf mark intensity and leaf thickness, and that the effect of the modifier loci on the leaf thickness variation is *V* locus-dependent.

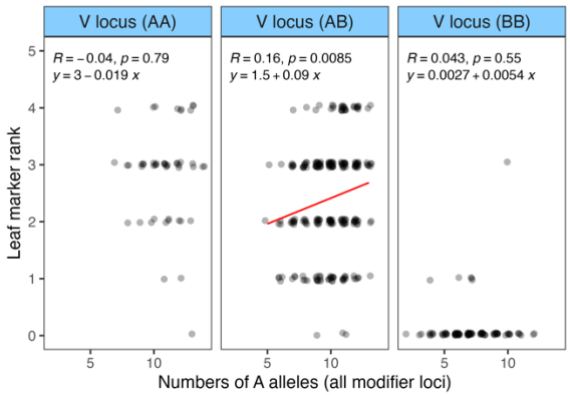
(A)


V locus
chr_05.48106478
Leaf mark level
AA
AB
BB
V locus
m1
chr_01.22946835
m2
chr_06.04962082
m3
chr_09.18256266
m4
chr_12.02972030
m5
chr_13.47299201
m6
chr_15.13490333
LOD score
m1
m2
m3
m4
m5
m6
chromosome

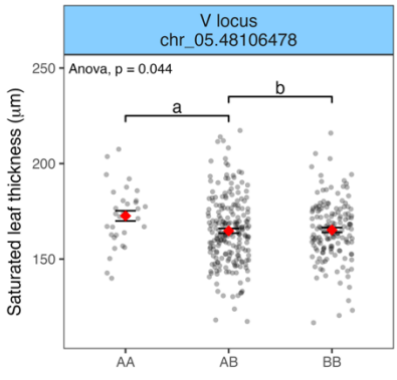
(B)


Leaf marker rank
Numbers of A alleles

(C)


V locus (AA)
R = −0.04, p = 0.79
y = 3 − 0.019 x
V locus (AB)
R = 0.16, p = 0.0085
y = 1.5 + 0.09 x
V locus (BB)
R = 0.043, p = 0.55
y = 0.0027 + 0.0054 x
Leaf marker rank
Numbers of A alleles (all modifier loci)

(D)


V locus
chr_05.48106478
Anova, p = 0.044
a
b
Saturated leaf thickness (μm)
AA
AB
BB

(E)

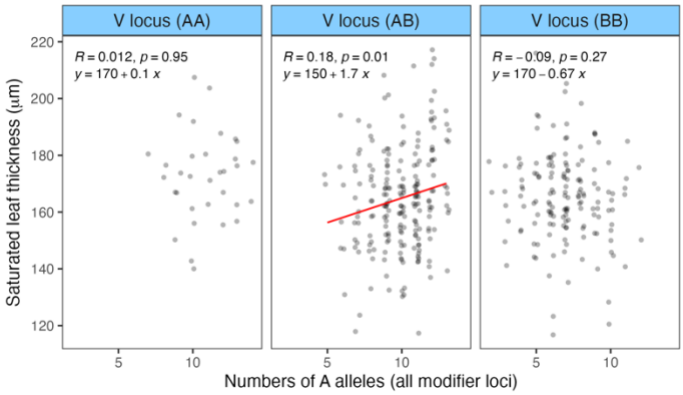

V locus (AA)
R = 0.012, p = 0.95
y = 170 + 0.1 x
V locus (AB)
R = 0.18, p = 0.01
y = 150 + 1.7 x
V locus (BB)
R = −0.09, p = 0.27
y = 170 − 0.67 x
Saturated leaf thickness (μm)
Numbers of A alleles (all modifier loci)

Figure 4.2 Genetics of white leaf mark and leaf thickness polymorphisms.

(A), LOD scan for the white leaf mark polymorphism. In addition to the major peak on Chr. 5 (chr_05.48106478, LOD = 134.87), which corresponds to the *V* locus, there are six minor peaks on different chromosomes with LOD scores greater than the significance threshold (LOD > 4.48, $p < 0.05$). Only one peak of the highest LOD score per chromosome is labeled. Inset panels show the phenotypic effects of the *V* locus genotypes (blue label, upper left) and the modifier loci (orange label, right). **(B)**, Additive effect of "A" alleles (pooled for the V locus and all six modifier loci) on leaf mark rank. **(C)**, Additive effect of the six modifier loci on white leaf mark in the context of the different genotypes of the *V* locus. **(D),** Saturated leaf thickness by *V* locus genotype. Tukey multiple comparison was conducted at the alpha of 0.05. (E), Saturated leaf thickness, additive effect of the six modifier loci in the context of the different genotypes of the *V* locus.

Photosynthetic efficiency

In addition to the association with greater leaf thickness, we hypothesized that the presence of white leaf mark could compromise the photosynthetic efficiency due to an increase in intercellular spaces or reduced chloroplast content. However, we observed no decline of net photosynthesis rate (P_n) with increasing leaf mark intensity. Instead, we found a nonsignificant trend in the opposite direction, with the P_n of the strong leaf mark leaves (22.72 ± 6.74 μmol $m^{-2}s^{-2}$) greater than the medium rank (21.88 ± 5.83 μmol $m^{-2}s^{-2}$) and no leaf mark (20.99 ± 5.39 μmol $m^{-2}s^{-2}$) leaves (ANOVA, $F_{(2, 194)} = 1.32$, $p = 0.27$; **Fig. 4.3A**). Similar to P_n, we did not detect any significant differences of stomatal conductance (G_s, $F_{(2, 194)} = 0.79$, $p = 0.46$), intercellular CO_2 concentration (C_i, $F_{(2, 197)} = 0.87$, $p = 0.42$), water use efficiency (WUE, $F_{(2, 197)} = 0.29$, $p = 0.75$), or water potential ($-\Psi_W$, $F_{(2, 195)} = 2.03$, $p = 0.13$) between different levels of leaf mark intensity (**Supplementary Fig. S4.3**).

Because the intensity of the white leaf mark was found to be positively correlated with leaf thickness (**Fig. 4.1D**), we also tested for correlations between leaf thickness and net photosynthesis rate. We found a significantly positive association for these two traits (P_n, $F_{(1, 195)} = 30.87$, $p = 8.97\times10^{-8}$) (**Fig. 4.3B**). We investigated this relationship further by examining leaf stomatal conductance; this revealed that leaf thickness has no effect on stomatal conductance (G_s, $F_{(1, 195)} = 1.90$, $p = 0.17$) (**Fig. 3C**) but has a significantly negative effect on intercellular CO_2 concentration (C_i, $F_{(1, 195)} = 7.73$, $p = 5.97\times10^{-3}$) (**Fig. 4.3D**). Since the leaf thickness has no detectable effect on the water status of the tested leaves ($-\Psi_W$, $F_{(1, 193)} = 1.60$, $p = 0.21$) (**Fig. 4.3E**), these findings together suggest that the leaf thickness is positively associated with WUE

($F_{(1, 195)}$ = 11.37, p = 8.98×10^{-4}) (**Fig. 4.3F**). They further suggest that at a comparable cellular water level, the leaves with greater thickness, possibly associated with loose arrangement or additional layers of cells in a leaf, can fix more CO_2 per unit of water use.

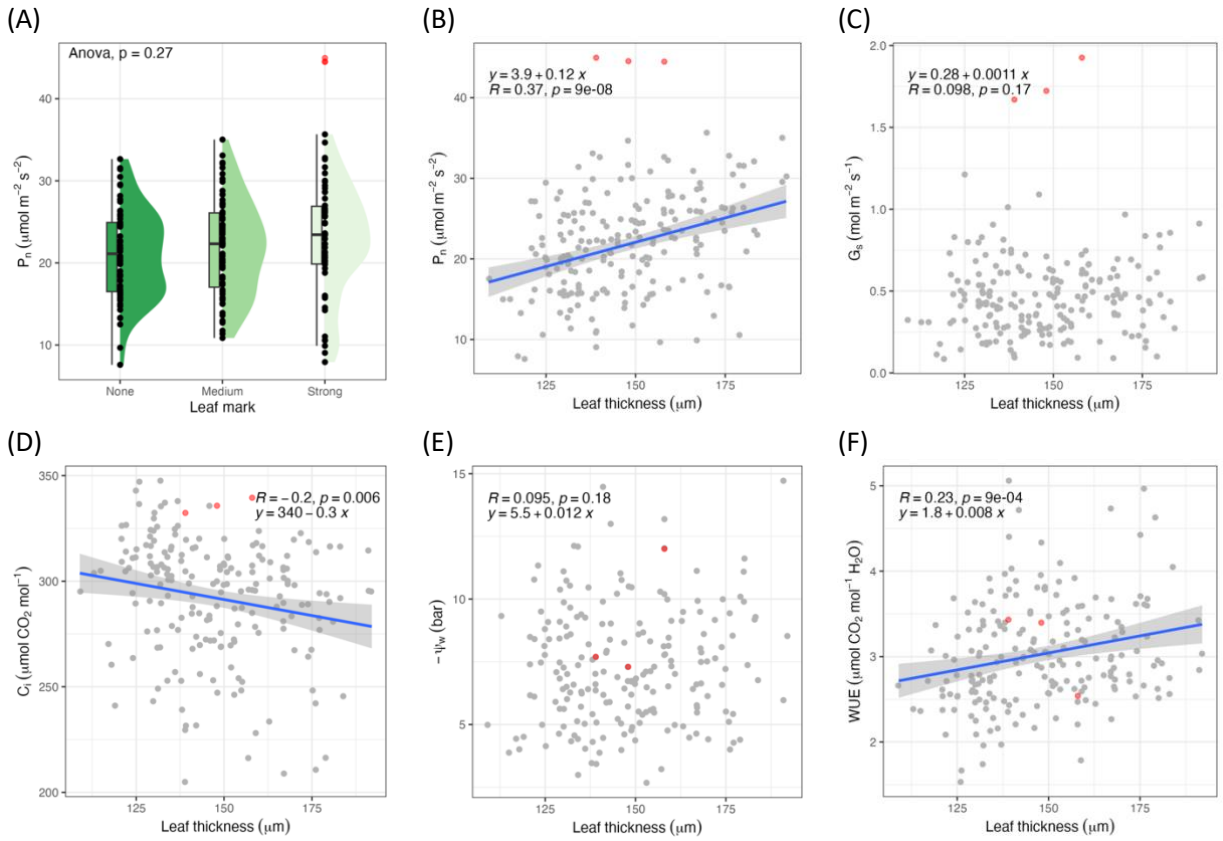


Figure 4.3 Physiological parameters of the leaves with different levels of white leaf mark and leaf thickness.

(A), Net photosynthesis rate of leaves with different white leaf mark levels. **(B-F)**, Associations with leaf thickness: **(B)**, Net photosynthesis rate. **(C)**, Stomatal conductance. **(D)**, Intercellular CO_2 concentration. **(E)**, Water potential. **(F)**, Water use efficiency. The three red data points showing abnormally high net photosynthesis rate were excluded from all but water potential statistical analyses.

4.6 Discussion

The inheritance and genetics of leaf mark variation in white clover have been studied for many decades (Brown, 1947) but the characteristic white leaf mark polymorphism (also referred to as the V mark or crescent mark) has remained largely uncharacterized with respect to its genetic basis and potential effects on other aspects of the leaf phenotype. Here we have successfully mapped the loci that control the white leaf mark polymorphism, and we have investigated its relationship to other aspects of the plant phenotype, including leaf morphology (specifically, leaf thickness) and physiology (photosynthetic efficiency). Consistent with classical genetics studies (Brewbaker, 1955; Carnahan *et al.*, 1955; Corkill, 1971), we find that the white leaf mark polymorphism is primarily controlled by a single major-effect locus (the *V* locus, which we map to Chr. 5), with the leaf mark-conferring allele ("A") showing incomplete dominance over the no-leaf-mark allele ("B") in our mapping population. Additionally, we discovered that the modifier loci contribute additively to the intensity of the leaf mark, particularly when the *V* locus is heterozygous. Moreover, the presence and greater intensity of the leaf mark are associated with greater leaf thickness, with "A" alleles at the *V* locus and modifier loci positively associated with this phenotype as well. With respect to physiology, we found no evidence that the white leaf mark compromises photosynthesis. Instead, leaves with greater thickness were correlated with a higher net photosynthesis rate (P_n) due to enhanced water use efficiency (WUE), irrespective of the presence of the leaf mark trait.

Unlike many model species, where structure-based leaf variation has been studied using chimeric albino mutants or other genotypes with unstable chloroplast development (Sakamoto,

2003), the white leaf mark in white clover has long been known to be a nuclear-encoded stable and non-chimeric polymorphism (Brewbaker, 1955; Carnahan *et al.*, 1955; Corkill, 1971). However, despite over 60 years of research into the genetics of the white leaf mark polymorphism (Brewbaker, 1955; Carnahan *et al.*, 1955; Corkill, 1971; Tashiro *et al.*, 2010), the corresponding *V* locus has remained unmapped until this study. The success of the present study can be attributed to the use of sequencing-based genetic markers (Olsen *et al.*, 2021; Kuo *et al.*, 2024), which offer greater density and precision than previously-used microsatellite markers (Tashiro *et al.*, 2010), and a repeated phenotyping process at multiple stages of leaf development to account for the phenotypic plasticity, which has not been addressed in previous studies of the white leaf mark polymorphism.

The *V* locus, mapped to Chr. 5, can explain 70.94% of the white leaf mark variation in DG F_3 progeny of our mapping population. Both this locus and the six identified modifier loci contribute additively to leaf mark intensity (**Fig. 4.2B**); the varying intensity of the white leaf mark is significantly correlated with the number of the "A" alleles at the modifier loci, especially when the *V* locus is heterozygous — a reflection of incomplete dominance at the *V* locus (**Fig. 4.2C**). These findings together suggest that the presence of white leaf mark requires at least one copy of the functional ("A") allele at the *V* locus to develop, which is consistent with findings of previous studies (Brewbaker, 1955; Carnahan *et al.*, 1955; Corkill, 1971). Building upon this foundation, this study extends our understanding that the intensity of white leaf mark is likely dependent on an intricate interaction among other genetic loci, which is likely to be associated with leaf structural development.

In addition to the heritable component, we also observed extensive plasticity in the white leaf mark phenotype. Based on our observations of leaf phenotypes at different developmental stages, the white leaf mark is most evident at the early developmental stages. As plants matured, the white leaf mark became less observable in some genotypes (**Supplementary Fig. S4.2**). However, the evident leaf mark could reappear in the following growth season or after vigorous pruning, suggesting that this plasticity is light-exposure dependent. A similar phenomenon in white clover appears to have been indirectly documented in field observations by Cahn and Harper (1976a), who reported that the frequency of leaf mark bearing genotypes decreased with increasing grass length in a mixed planting (although the authors attributed this phenotypic shift to differential herbivory pressure rather than developmental plasticity).

Our finding that the intensity of leaf mark is correlated with leaf thickness suggests that leaf thickness has a major effect on the intensity of the white leaf mark, especially when the *V* locus is heterozygous. Although no QTLs were detected for leaf thickness, we found that the genotypes of the *V* locus and leaf mark modifier loci were associated with leaf thickness variation. The effect direction was consistent with that of the leaf mark, with the "A" allele contributing to the increase of leaf mark and leaf thickness (**Fig. 4.2D, E**). This correlation illustrates the phenotypic interactions between leaf mark and leaf thickness development and suggests potential pleiotropic effects of the *V* locus and modifier loci in controlling multiple aspects of leaf development.

Various balancing selection hypotheses for the maintenance of the white leaf mark polymorphism have been proposed, most of them related to potential fitness trade-offs between the effectiveness of the leaf mark in herbivore avoidance and the costs of compromised photosynthetic efficiency (Esteban *et al.*, 2008; Menzies *et al.*, 2015). Herbivore avoidance mechanisms could include the leaf mark serving as aposematic coloration to warn herbivores of cyanogenic toxicity (Lev-Yadun, 2014; Lev-Yadun, 2021), or as a dazzling effect that could hinder herbivores' ability to locate the leaves (Lev-Yadun, 2014), or as a disruptive coloration to make the leaves look small or broken to lower the foraging preferences (Cahn & Harper, 1976b). Conversely, the presence of the leaf mark could be associated with compromised photosynthetic efficiency due to the presence of atypical palisade cells, increased intercellular and subepidermal spaces (Carnahan *et al.*, 1955; Esteban *et al.*, 2008).

In our analyses of the white leaf mark polymorphism in white clover, we explicitly tested for fitness trade-offs by examining the photosynthetic efficiency between different intensities of white leaf mark in our mapping population. We observed no decrease in photosynthetic efficiency between leaves with and without white leaf mark, suggesting a minimal fitness cost for a white leaf mark bearing genotype. Similar results of uncompromised photosynthetic efficiency have been reported in natural variegated *Begonia* spp. (Sheue *et al.*, 2012) and multiple horticultural plants (Konoplyova *et al.*, 2008; Zhang *et al.*, 2019). These findings together suggest that structure-based color patterning is a highly efficient way for plants to develop variegation patterns, with negligible costs for photosynthetic output. They further suggest that the presence of the white leaf mark polymorphism in white clover involves

mechanisms more complex than a simple fitness trade-off between herbivore avoidance and photosynthesis efficiency.

4.7 Conclusion

The elucidation of the genetic mechanisms underlying natural leaf variegation in plants, a largely uncharted territory, marks the overarching significance of this study. Important advances of this study include mapping the white clover *V* locus and detailed documentation of associations between the white leaf mark, leaf thickness, and photosynthetic efficiency. In addition, we not only documented the plasticity of the leaf mark intensity in white leaf mark bearing genotypes, but also documented the additive effects of the *V* locus and the six modifier loci, which are potentially pleiotropic in also controlling leaf thickness development. Physiologically, we found no evidence that the white leaf mark compromised photosynthetic efficiency. Instead, the significant positive correlations we detected between the white leaf mark intensity, leaf thickness and photosynthetic efficiency highlight that selective explanations based on simple fitness trade-offs cannot fully account for the observed trait variation.

4.8 Funding

This work was supported by the William H. Danforth Plant Science Graduate Research Fellowship in the Division of Biology and Biomedical Sciences at Washington University to W.-H.K.; by the scholarship of Taiwan Ministry of Education to W.-H.K.; and by the U.S. National Science Foundation (grant number IOS-1557770) to K.M.O.'

4.9 Supplementary Information

Supplementary file 4.1 – Step-by-step protocol of GBS library preparation. **Supplementary file 4.2** – Raw data of photosynthetic efficiency measurement. **Supplementary file 4.3** – Linkage map and all phenotypic data (a "cross" for R/qtl package). **Supplementary Figure S4.1 –** Examples of white leaf mark rank classification. **Supplementary Figure S4.2 –** Variable intensity of the white leaf mark in a single plant. **Supplementary Figure S4.3 –** Physiological parameters of leaves with different white leaf mark intensity.

4.10 Acknowledgements

We thank the Washington University greenhouse staff, especially Michael Dyer, Michael Stephan, and Hammy Sorkin, for care of plants used in the study. We thank David Goad for providing suggestions in the QTL mapping and result interpretations. We thank anonymous reviewers for their suggestions and comments. W.-H. K. designed and conducted the experiment, analyzed the data, interpreted the results, and wrote the manuscript. E. C. grew and maintained the plants, recorded the white leaf mark and photosynthesis data. E. G. grew and maintained the plants and recorded the leaf thickness data. E. C. and E. G. were undergraduate students in Washington University's Bio 500 course (Undergraduate Independent Research). K. M. O. conceived the project, interpreted the results, and edited the manuscript. All authors read and approved the final manuscript.

4.11 Data availability

Raw GBS reads for DG F_3 linkage map construction may be found under NCBI SRA BioProject: PRJNA1027808.

4.12 Reference

Abdi AI, Nichols PGH, Kaur P, Wintle BJ, Erskine W. 2020. Morphological diversity within a core collection of subterranean clover (*Trifolium subterraneum* L.): Lessons in pasture adaptation from the wild. *PLoS One* **15**(1): e0223699.

Afzal A, Duiker SW, Watson JE. 2017. Leaf thickness to predict plant water status. *Biosystems Engineering* **156**: 148-156.

Attila TS. 1987. V-leaf marking in a *Trifolium ambiguum* Bieb. germplasm collection. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* **17**(1).

Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, et al. 2004. A microsatellite map of white clover. *Theoretical and Applied Genetics* **109**(3): 596-608.

Brewbaker JL. 1955. V-leaf markings of white clover. *Journal of Heredity* **46**(3): 115-123.

Brewbaker JL, Carnahan HL. 1956. Leaf marking alleles in white clover: Uniform nomenclature. *Journal of Heredity* **47**(2): 103-104.

Broman KW, Wu H, Sen Ś, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**(7): 889-890.

Brown K 1947. Inheritance of patterned leaf marking in white clover. *Meeting of American Society of Agronomy*.

Browning BL, Tian X, Zhou Y, Browning SR. 2021. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* **108**(10): 1880-1890.

Cahn MG, Harper JL. 1976a. The biology of the leaf mark polymorphism in *Trifolium repens* L. - 1. distribution of phenotypes at a local scale. *Heredity* **37**(3): 309-325.

Cahn MG, Harper JL. 1976b. The biology of the leaf mark polymorphism in *Trifolium repens* L. - 2. evidence for the selection of leaf marks by rumen fistulated sheep. *Heredity* **37**(3): 327-333.

Campitelli BE, Stehlik I, Stinchcombe JR. 2008. Leaf variegation is associated with reduced herbivore damage in *Hydrophyllum virginianum*. *Botany* **86**(3): 306-313.

Carnahan hL, Hill hd, Hanson aA, Brown kG. 1955. Inheritance and frequencies of leaf markings in white clover. *Journal of Heredity* **46**(3): 109-114.

Charles AH. 1968. Some selective effects operating on white- and red-clover in swards. *Grass and Forage Science* **23**(1): 20-25.

Corkill L. 1971. Leaf markings in white clover. *Journal of Heredity* **62**(5): 307-310.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**(5): e19379.

Esteban R, Fernandez-Marin B, Becerril JM, Garcia-Plazaola JI. 2008. Photoprotective implications of leaf variegation in *E. dens-canis* L. and *P. officinalis* L. *Journal of Plant Physiology* **165**(12): 1255-1263.

Ganders FR, Griffiths AJF, Carey K. 1980. Natural selection for spotted leaves: parallel morph ratio variation in three species of annual plants. *Canadian Journal of Botany* **58**(6): 689-693.

Griffiths AG, Barrett BA, Simon D, Khan AK, Bickerstaff P, Anderson CB, Franzmayr BK, Hancock KR, Jones CS. 2013. An integrated genetic linkage map for white clover (*Trifolium repens* L.) with alignment to *Medicago*. *BMC Genomics* **14**: 388.

Isobe SN, Hisano H, Sato S, Hirakawa H, Okumura K, Shirasawa K, Sasamoto S, Watanabe A, Wada T, Kishida Y, et al. 2012. Comparative genetic mapping and discovery of linkage disequilibrium across linkage groups in white clover (*Trifolium repens* L.). *G3: Genes, Genomes, Genetics* **2**(5): 607-617.

Konoplyova A, Petropoulou Y, Yiotis C, Psaras GK, Manetas Y. 2008. The fine structure and photosynthetic cost of structural leaf variegation. *Flora* **203**(8): 653-662.

Koski MH, Ashman TL. 2015. Floral pigmentation patterns provide an example of Gloger's rule in plants. *Nature Plants* **1**: 14007.

Kuo W-H, Wright SJ, Small LL, Olsen KM. 2024. *De novo* genome assembly of white clover (*Trifolium repens* L.) reveals the role of copy number variation in rapid environmental adaptation. *Research Square*.

Lev-Yadun S. 2014. Potential defence from herbivory by 'dazzle effects' and 'trickery coloration' of leaf variegation. *Biological Journal of the Linnean Society* **111**(3): 692-697.

Lev-Yadun S. 2021. Avoiding rather than resisting herbivore attacks is often the first line of plant defence. *Biological Journal of the Linnean Society* **134**(4): 775-802.

Mccomb J. 1974. Leaf marks in *Medicago*, with special reference to their inheritance in *Medicago truncatula*. *Australian Journal of Botany* **22**(1): 67-80.

Menzies IJ, Youard LW, Lord JM, Carpenter KL, van Klink JW, Perry NB, Schaefer HM, Gould KS, Heil M. 2015. Leaf colour polymorphisms: a balance between plant defence and photosynthesis. *Journal of Ecology* **104**(1): 104-113.

Olsen KM, Goad DM, Wright SJ, Dutta ML, Myers SR, Small LL, Li LF. 2021. Dual-species origin of an adaptive chemical defense polymorphism. *New Phytologist* **232**(3): 1477-1487.

Postma FM, Agren J. 2016. Early life stages contribute strongly to local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **113**(27): 7590-7595.

Sakamoto W. 2003. Leaf-variegated mutations and their responsible genes in *Arabidopsis thaliana*. *Genes & Genetic Systems* **78**(1): 1-9.

Santangelo JS, Battlay P, Hendrickson BT, Kuo WH, Olsen KM, Kooyers NJ, Johnson MTJ, Hodgins KA, Ness RW. 2023. Haplotype-resolved, chromosome-level assembly of white

clover (*Trifolium repens* L., Fabaceae). *Genome Biology and Evolution* **15**(8): 2023.2006.2006.543960.

Shelef O, Summerfield L, Lev-Yadun S, Villamarin-Cortez S, Sadeh R, Herrmann I, Rachmilevitch S. 2019. Thermal benefits from white variegation of *Silybum marianum* leaves. *Frontiers in Plant Science* **10**: 688.

Sheue CR, Pao SH, Chien LF, Chesson P, Peng CI. 2012. Natural foliar variegation without costs? The case of *Begonia*. *Annals of Botany* **109**(6): 1065-1074.

Soltau U, Dötterl S, Liede-Schumann S. 2008. Leaf variegation in *Caladium steudneriifolium* (Araceae): a case of mimicry? *Evolutionary Ecology* **23**(4): 503-512.

Song B, Stocklin J, Armbruster WS, Gao Y, Peng D, Sun H. 2018. Reversible colour change in leaves enhances pollinator attraction and reproductive success in *Saururus chinensis* (Saururaceae). *Annals of Botany* **121**(4): 641-650.

Tan BH, Collins WJ. 1987. Multi-allelic nature of the locus controlling leaf marking in subterranean clover. *Australian Journal of Agricultural Research* **38**(3): 547-558.

Tashiro RM, Han Y, Monteros MJ, Bouton JH, Parrott WA. 2010. Leaf trait coloration in white clover and molecular mapping of the red midrib and leaflet number traits. *Crop Science* **50**(4): 1260-1268.

Taylor J, Butler D. 2017. R Package ASMap: Efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software* **79**(6): 1 - 29.

Torkamaneh D, Laroche J, Belzile F. 2020. Fast-GBS v2.0: an analysis toolkit for genotyping-by-sequencing data. *Genome* **63**(11): 577-581.

Whitlock R, Hipperson H, Mannarelli M, Burke T. 2008. A high-throughput protocol for extracting high-purity genomic DNA from plants and animals. *Molecular Ecology Resources* **8**(4): 736-741.

Wright SJ, Goad DM, Gross BL, Muñoz PR, Olsen KM. 2022. Genetic trade-offs underlie divergent life history strategies for local adaptation in white clover. *Molecular Ecology* **31**(14): 3742-3760.

Zhang J-H, Zeng J-C, Wang X-M, Chen S-F, Albach DC, Li H-Q. 2020. A revised classification of leaf variegation types. *Flora* **272**.

Zhang Z, Liu Z, Song H, Chen M, Cheng S. 2019. Protective role of leaf variegation in *Pittosporum tobira* under low temperature: Insights into the physio-biochemical and molecular mechanisms. *International Journal of Molecular Sciences* **20**(19).

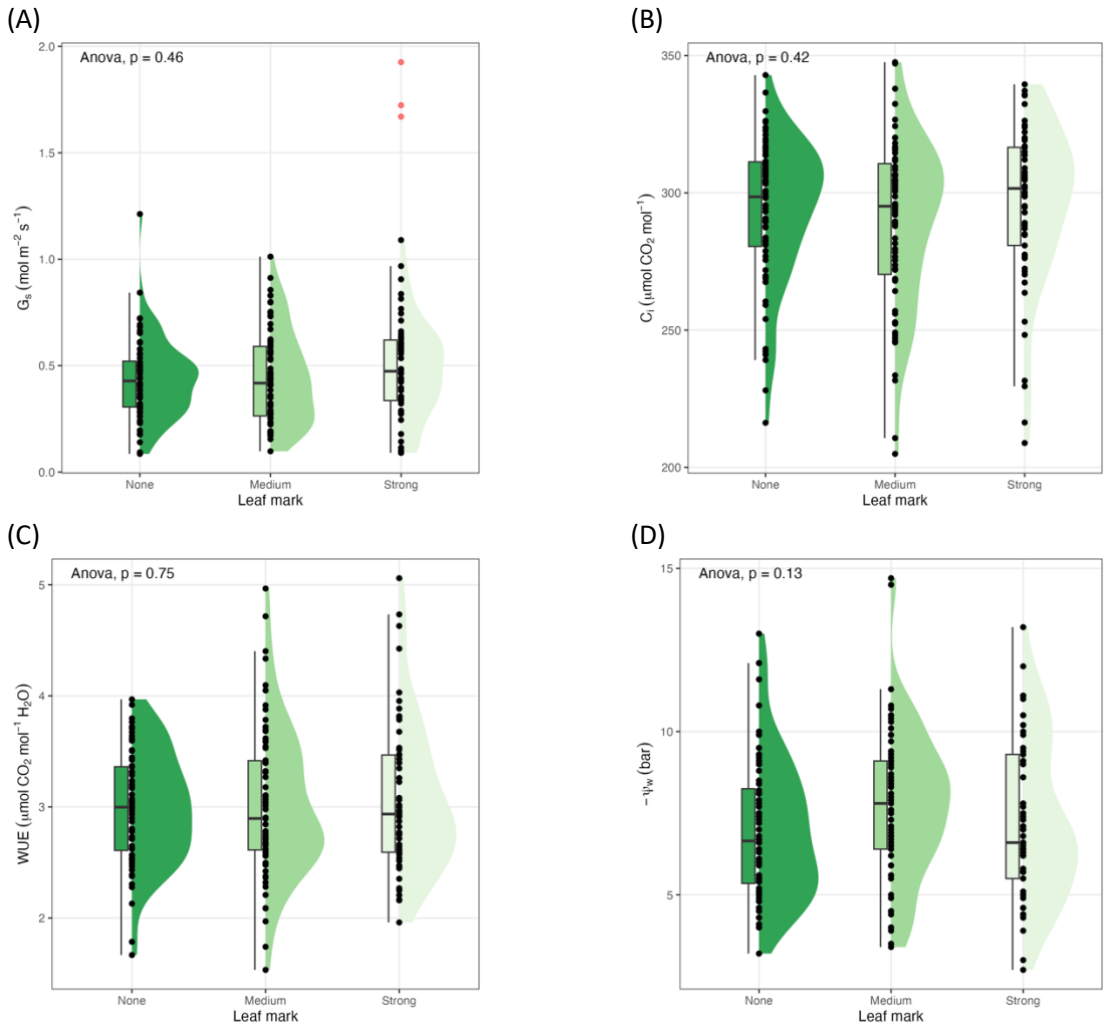
(A)

(B)

(C)

Supplementary Figure S4.1 Variable expressivity of white leaf mark in a single plant.

(A), plant accession number DG_F3_186. (B), plant accession number DG_F3_180. (C), plant accession number DG_F3_387.



Supplementary Figure S4.2 Physiological parameters of leaves with different white leaf mark levels.

(A), stomata conductance. **(B)**, intercellular CO_2 concentration. **(C)**, water use efficiency. **(D)**, water potential.

Chapter 5 Phenotypic and genetic bases of variable drought stress responses in a widely adapted allotetraploid species (Appendix)

5.1 Authorship and Affiliations

Wen-Hsi Kuo[1], Linda Small, Kenneth M. Olsen[1]

[1]Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA

5.2 Introduction

Local adaptation, the process by which a local population evolves to optimize its fitness in response to its specific environmental conditions, is especially crucial for plant species with wide distributions. This phenomenon has been well-documented in various studies (e.g., Agrena *et al.*, 2013; Postma & Agren, 2016; Price *et al.*, 2018; van Boheemen *et al.*, 2019; McGoey *et al.*, 2020; Bieker *et al.*, 2022; Kreiner *et al.*, 2022; Lee *et al.*, 2024), highlighting its importance across the broad range of conditions that widely distributed plant species often face, including distinct temperature, precipitation, and photoperiod regimes. Such 'resilience' adaptations are essential for plants to either escape or tolerate extreme seasonal stresses — such as freezing winters, periods of drought, and summer heat (e.g., Fernández-Pascual *et al.*, 2013; Burghardt *et al.*, 2016; Postma & Agren, 2016; Price *et al.*, 2018) — and to maximize vegetative and reproductive growth when conditions are favorable (Gould & Stinchcombe, 2017; Hämälä *et al.*, 2018; McGoey *et al.*, 2020; Yan *et al.*, 2021; Kreiner *et al.*, 2023; Wunder *et al.*, 2023). The study of these adaptations, including their genetic underpinnings, is valuable for future agricultural improvements (e.g., Guan *et al.*, 2022) and biodiversity conservation (e.g., Wang *et al.*, 2023), particularly in the face of rapid climate change (Park *et al.*, 2018; Chan *et al.*, 2024). However, most research to date has focused on a limited number of model species, often overlooking those with complex genetic and demographic backgrounds, such as species with polyploid genomes or those introduced widely by humans.

White clover (*Trifolium repens* L., 2n = 4x = 32), historically dubbed the "agricultural equivalent of coal" for its pivotal role in nitrogen fixation before the advent of synthetic

fertilizers, is a key forage legume originally native to Europe. It has since become naturalized across a broad spectrum of climates worldwide, including throughout North America (Zeven, 1991; Kjærgaard, 2003). In addition, white clover is an allopolyploid species, descended from two extant diploid species. Its two parental species have very restricted distributions, either on coastal cliffs of western Europe (*T. occidentale*) or high-elevation meadows of the Alps (*T. pallescens*). It is hypothesized that white clover has acquired its drought and salt tolerance from the coastal ancestor, and cold and freeze tolerance from the alpine ancestor; thus, allopolyploidization may significantly contribute to the ability of white clover to adapt to a wide range of climates (Griffiths et al., 2019).

White clover serves as a textbook example of a plant species that rapidly evolves local adaptation (Briggs & Walters, 2016; Futuyma & Kirkpatrick, 2017). It is particularly noted for its repeatedly evolved clines in the chemical defense cyanogenesis (HCN release upon tissue damage). White clover is polymorphic for this antiherbivore defense, with both cyanogenic and acyanogenic plants occurring commonly in nature; higher frequencies of the cyanogenic form are consistently found in warmer locations throughout the native and introduced species range (Daday, 1958; Hughes, 1991; Olsen *et al.*, 2007; Olsen *et al.*, 2008; Olsen *et al.*, 2021; Kuo *et al.*, 2024). The rapid evolution of these climate-associated cyanogenesis clines is thought to be an evolved response to local herbivore pressure, with higher frequencies of cyanogenesis favored in areas where herbivore activity is more intense (Daday, 1965; Hughes, 1991). Besides its role in herbivore defense, the chemical precursor to HCN, cyanogenic glucosides (CNglcs), is believed to act as a form of nitrogen storage during drought stress, offering an additional

survival advantage (Møller, 2010; Machingura *et al.*, 2016; Nielsen *et al.*, 2016; McMahon *et al.*, 2021). Evidence suggests that white clover plants are capable of recycling these CNglcs, even without tissue damage, supporting the theory of its multifaceted role in the plant's adaptive strategy (Kuo *et al.*, 2023).

Clines is cyanogenesis and/or CNglcs have been documented across a wide range of temperature (Kooyers & Olsen, 2012; Kuo *et al.*, 2024) and aridity gradients (Kooyers *et al.*, 2014), including throughout most of species distribution in North America. However, our recent studies of North American white clover, incorporating diverse methodologies such as common garden experiments, quantitative trait locus (QTL) mapping, and genotype-environment association (GEA) studies, indicate that cyanogenesis might not be the main adaptive trait for coping with regional variations in temperature or aridity (Wright *et al.*, 2018; Wright *et al.*, 2022)(Kuo *et. al.*, 2024). Instead, the GEA and QTL mapping suggest that selection for local adaptation acts on traits associated with responses to local seasonal cues for the timing of vegetative growth and flowering. This insight is pivotal, as it shifts the perspective on white clover's wide environmental adaptability beyond cyanogenesis, which has historically been the primary focus of adaptive mechanisms in this species (Wright *et al.*, 2022)(Kuo *et. al.*, 2024).

Expanding on these insights, the present study examines white clover plants representing three climatically distinct locations along a latitudinal transect in the United States: Duluth (Minnesota), St. Louis (Missouri), and Gainesville (Florida). These sites were chosen for their distinct positions across the North American species range — northernmost,

middle, and southernmost — providing a comprehensive view of the climatic variation experienced across this transect. Each location exhibits unique temperature and precipitation patterns throughout the year, which together dictate the length and conditions of the growing season. Specifically, the temperature regime defines the start and end of the growing period, while the interaction between temperature and precipitation informs the likelihood and severity of drought stress within this timeframe.

White clover populations in Duluth experience up to a six-month growing season (where the growing season is broadly defined by temperatures exceeding 5°C and sufficient precipitation to support growth). The coincidence of rainfall with favorable growing temperatures in Duluth confers relatively mesic environment, with local populations experiencing minimal drought stress during its growing season (**Fig. 5.1A**). In comparison, the white clover in St. Louis is subject to a longer growing season of up to nine months and a relatively constant precipitation throughout the year (**Fig. 5.1B**). Despite this extended growing period, the St. Louis population is posited to endure periods of drought stress coincident with summer heat, occurring shortly after the onset of its growing season and lasting about four months. Conversely, the Gainesville white clover benefits from a year-round growing season due to its warmer climate, which is conducive to rapid growth and enhanced fitness. However, this advantage is undermined by two annual drought periods in April and October, during which a lack of synchronization between precipitation and temperature (highlighted by red arrows in **Fig. 5.1C**) introduces significant challenges to the plants' fitness in Gainesville.

Drawing on the insights from our earlier research (Wright *et al.*, 2018; Wright *et al.*, 2022) and the monthly climatic data for the three focal regions (**Fig. 5.1**), the present study was designed to test whether water availability is a key factor influencing life history and flowering time in white clover across its climatic range, and which traits and genes confer adaptation to drought stress. Through an experiment the simulated progressive and sub-lethal drought stress conditions in a greenhouse setting, we examined the performance of white clover genotypes from Duluth, St. Louis, and Gainesville, as well as a Quantitative Trait Locus (QTL) mapping population derived from a bi-parental cross between Duluth and Gainesville genotypes, aiming to uncover how variation in water availability across different latitudes affect plants' adaptive strategies. The study addressed five questions: **Firstly**, what are the morphological, physiological, and gene expression responses of the three white clover genotypes (from Duluth, St. Louis, and Gainesville) when subjected to drought conditions? This involves comprehensive documentation of how drought stress impacts the plants at various levels. **Secondly**, utilizing the Duluth x Gainesville QTL mapping population, what QTLs underlie the observed phenotypic variations under drought stress? This question encompasses an examination of trait heritability, interactions between genotype and treatment, how energy is allocated between vegetative growth and reproductive output, and the presence of allelic trade-offs (i.e., antagonistic pleiotropy), among the identified loci. **Thirdly**, to what extent does genetic variation in the production of cyanogenic glucosides (CNglcs) correlate with the plant's response to drought? Specifically, we test whether the QTLs associated with drought response overlap with loci controlling the cyanogenesis polymorphism. **Fourthly**, do the QTLs identified in this study align with the genetic variations associated with drought stress adaptation observed in natural

populations, as determined by our previous genotype-environment association (GEA) analysis (Kuo *et. al.*, 2024)? **Finally**, in light of the allopolyploidization and niche expansion hypothesis (Griffiths *et al.*, 2019), do the drought stress-induced differentially expressed genes (DEGs) and QTLs show a biased distribution on either subgenome? Specifically, does genetic material from the salt-exposed coastal ancestor (*T. occidentale*) contribute more to these DEGs and QTLs compared to that from the alpine ancestor (*T. pallescens*)?

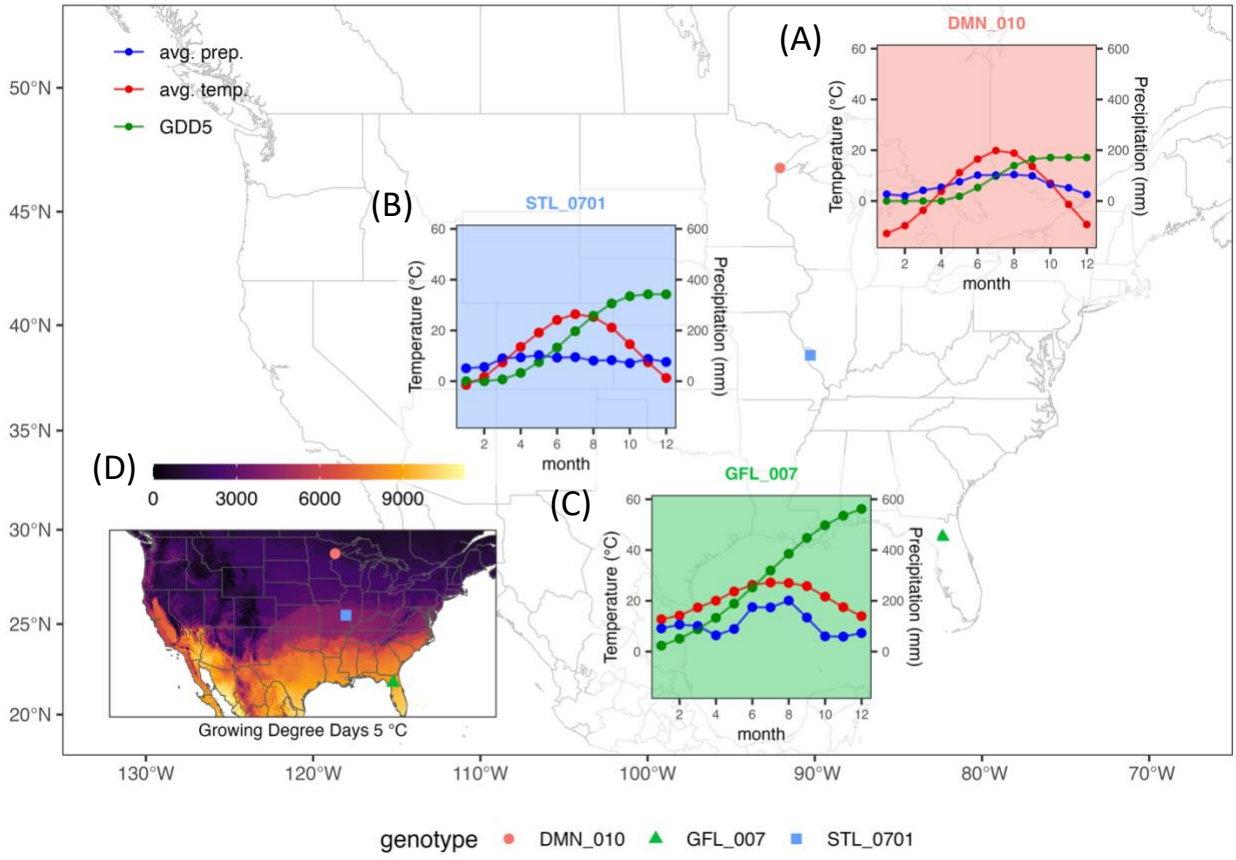


Figure 5.1 Local climates of the parental accessions (genotypes).

The panel **(A, B, C)** colors are corresponded to the local climates of each accession, respectively. Avg. prep., average monthly precipitation. Avg. temp., average monthly temperature. GDD5, Growing Degree Days (> 5°C); the value is divided by 100 for visualization (see left-side axis). **(D)**, panel shows Growing Degree Days (> 5°C) on the map. The two arrows in the GFL_007 panel (green) indicate two drought periods during the growth season.

5.3 Materials and Methods

Generation of F_3 mapping population

For generating a QTL population, the three parent (F_0) plants were selected from environments which differ in cyanogenesis frequency. They are all from wild populations in North America: Duluth, MN (DMN_010, *ac*/*ac*, *li*/*li*; BioSample: SAMN34157026), St. Louis, MO (STL_0701, *ac*/*ac*, *li*/*li*), and Gainesville, FL (GFL_007, (*Ac*/*Ac*, *Li*/*Li*; BioSample: SAMN37329216, SAMN34116011). Duluth sits near the U.S.-Canadian border and experiences some of the coldest winter temperatures. Gainesville locates near the transition from temperate to subtropical climate, near the southern limit of naturalized U.S. white clover populations. St. Louis is centrally located in the U.S., midway latitudinally between Duluth and Gainesville, and experiences a continental climate marked by cold winters and hot summers. The two parents were cross-pollinated bidirectionally by hand. An F_2 population was constructed by intercrossing the F_1 generation (N=50-100 plants) with solitary bees (Crown Bees, Woodinville, WA) in enclosed cages in the greenhouse (Olsen *et al.*, 2021; Wright *et al.*, 2022). The F_3 population was constructed by randomly intercrossing the F_2 population by hand. The gametophytic self-incompatibility feature of white clover prevented plants from self-pollination. Cyanotypes for all plants were determined using Feigl-Anger cyanogenesis assays and by PCR-genotyping the *Ac/ac* and *Li/li* polymorphisms (Olsen *et al.*, 2007; Olsen *et al.*, 2008). All plants were cultivated in the Washington University glasshouse facilities under standard conditions (16 h: 8 h, light: dark photoperiod with 140 µmol supplemental lighting as needed; 18-28°C).

Characterization of cyanotypes

The cyanogenic polymorphism in white clover is controlled by two unlinked genetic loci which are responsible for the presence/absence of two biochemical components, both of which must be present for a plant to be cyanogenic. The first locus (*Ac*/*ac*) contains a three-gene metabolic cluster required for cyanogenic glucoside (CNglcs) synthesis (Olsen & Small, 2018). The second, unlinked locus (*Li*/*li*) encodes their hydrolyzing enzyme, linamarase (a cyanogenic β-glucosidase), which can hydrolyze the CNglcs and release toxic HCN. Homozygote recessive genotypes (*acac* or *lili*) are acyanogenic. They carry gene deletions at *Ac* or *Li* loci, respectively, and lack the ability to synthesize the necessary cyanogenic precursors (Olsen *et al.*, 2007; Olsen *et al.*, 2008). Because of the two underlying genetic polymorphisms, four different cyanogenesis phenotypes (or "cyanotypes") occur in nature: the cyanogenic form ("AcLi", with both cyanogenic precursors present), and three acyanogenic forms ("Acli", lacking linamarase; "acLi", lacking cyanogenic glucosides; and "acli", lacking both components). In white clover, the ratio of cyanogenic to acyanogenic genotypes in wild populations is positively correlated with winter temperature and growing degree days, which is a common estimate of accumulation of temperature during the growing season (Kooyers & Olsen, 2012; Kooyers & Olsen, 2013; Kuo *et al.*, 2024).

Drought stress experiment

A consistent drought environment was established and maintain in Washington University in St. Louis greenhouse, following the instruction in Marchin *et al.* (2019) with modifications. Ceramic coasters as water resistors were added between the commercial porous foam (OASIS® DELUXE FLORAL FOAM MAXLIFE, Smithers-Oasis, Kent, OH, USA) and customized PVC pots with 0.2 μm porous nylon mesh at the bottom (3.5 inch in diameter and 4 inch in height). The ceramic coasters were coated by water impermeable paint except for the center of 1-inch diameter for water to pass. The controlled environment was the same setting but without the ceramic coasters. The drought treatment was applied after a week of transplanting. The volumetric water content (VWC) for the control was maintained between 20-30%. The VWC for the drought decreased to 0-5% gradually a week after the treatment started. The VWC was checked every other day. If a pot in drought treatment had a VWC higher than 5%, the pot was manually removed from the porous foam until the VWC decreased below 3%.

Genotyping

DNA extraction, genotyping-by-sequencing (GBS) library preparation, and sequencing were exactly described in Kuo et. al., 2024. After sequencing, the single nucleotide polymorphism (SNP) and indel were called following the GATK best practice workflow (Poplin et al., 2017). In brief, the GBS reads of the F_3 mapping population (NCBI SRA BioProject: PRJNA1027808) and the whole genome resequencing reads for the parental accessions (NCBI SRA BioProject: PRJNA1064563) were mapped to the reference (Kuo *et al.*, 2024) by BWA-MEM (Li, 2013). Then, the SNPs were called by HaplotypeCaller (-ERC GVCF) and combined by

GenotypeGVCFs. The output in vcf format was undergone a progressive filtering pipeline: pre-filter (vcftools --max-missing 0.5), hard-filter (bcftools filter -e 'QD < 0.5 || FS > 200.0 || MQ < 20.0 || MQRankSum < -12.5 || MQRankSum > 12.5 || ReadPosRankSum < -8.0 || SOR > 8'), mid-filter (vcftools --maf 0.01), imputation (Beagle 5.4), homozygotes in both parental accessions (vcftools --positions), and post-filter (vcftools --maf 0.2 --hwe 0.01). The remaining biallelic SNPs and indels were used in linkage map construction and QTL analyses.

Phenotyping

Morphological data

Leaf areas were assessed weekly through top-view photographic techniques. The green areas' pixels were captured using a customized Python script and subsequently converted into square millimeters (mm^2). The number of inflorescences was cumulatively counted, with only mature inflorescences at anthesis being considered for weekly counts. To prevent duplicate counts, labeled stickers were affixed to the counted inflorescences. Subsequently, the thickness of the top leaflet of a trifoliate leaf was measured using a thickness gauge (Model No. 547-526S; Mitutoyo, Japan), carefully avoiding the midrib during the measurement process. At the end of the experiment, plants were harvested, and their soil was rinsed off under tap water. The shoots and roots were then separated and subjected to drying in an oven at 50°C to facilitate the measurement of dry biomass.

Physiological data

In order to investigate the physiological responses to drought stress, a minimum of three leaves per specimen were randomly selected for measurement of net photosynthesis rate, stomatal conductance, and water potential. These assessments were performed using a LICOR-6400XT (LI-COR Biosciences, Lincoln, NE, USA) during the summer of 2021, specifically between 10 am and 3 pm. Measurements were conducted under a controlled environment with an LED light source providing an intensity of 1100 µmol m^{-2} s^{-1}, and a steady CO2 flow of 400 µmol s^{-1}. The experiment maintained a block temperature ranging from 23 to 30°C and a reference relative humidity between 40% and 60%, mirroring the ambient conditions. Following the LICOR-6400XT procedure, the leaf was detached from the plant for immediate thickness measurement. Subsequently, the leaf water potential was determined using a PMS pressure chamber (Model 1000; Albany, OR, USA). Finally, leaf area was quantified through scanned images.

Gene expression

The RNA-seq data were derived from leaf tissue samples of the three accessions used to create the linkage maps (DMN_010, GFL_007, STL_0701); tissue was collected from plants grown in greenhouse at midday after eight weeks of controlled and drought treatments (N = 18, including three accessions and three clonal replications of each accession). RNA was extracted by TRIzol™ Reagent (Thermo Fisher Scientific, Waltham, MA, USA). The libraries were prepared with Kapa Hyper Stranded mRNA library kit (Roche, Basel, Switzerland). The libraries were pooled; quantitated by qPCR and sequenced on one SP lane for 151 cycles from one end of the

fragments on a NovaSeq 6000 (Illumina, San Diego, CA, USA). Fastq files were generated and demultiplexed with the bcl2fastq v2.20 Conversion Software (Illumina). Adaptors have been trimmed and reads were mapped to the reference genome in with an annotation file by STAR v2.7.10a (--twopassMode Basic --outFilterScoreMinOverLread 0.1 --outFilterMatchNminOverLread 0.1) (Dobin *et al.*, 2012). The raw counts per gene were generated by featureCounts (-C -p --countReadPairs -B -P -d 50 -D 1000000 -Q 3 -t exon -g gene_id -F GTF -s 2) (Liao *et al.*, 2014). Differential expression of genes was analyzed by DEseq2 in R (Love *et al.*, 2014).

Statistical analyses

In order to test the combined effect of cyanotype and treatments on fitness trait, a mixed effect linear model was fitted to the data:

$$\text{Phenotype}_{ij} = \beta_0 + \beta_1 \text{Cyanotype}_{ij} + \beta_2 \text{Treatment}_{ij} + \beta_3 (\text{Cyanotype} \times \text{Treatment})_{ij} + u_i + \epsilon_{ij}$$

Cyanotype_{ij} and Treatment_{ij} are fixed effects representing the cyanotype (“AcLi”, “Acli”, “acLi”, “acli”) and treatment (control and drought) conditions, respectively, and their interaction term $(\text{Cyanotype} \times \text{Treatment})_{ij}$ captures the combined effect of these conditions on the fitness trait. Block (u_i) is a random effect accounting for variability associated with different blocks (replications), acknowledging that measurements within a block may be more similar to each other than to measurements from different blocks, possible due to the seasonal changes in greenhouse. We utilized the ‘lmer’ and 'lmerTest' function to fit the linear model and

employed Type-II ANOVA to test for the significance of the main effects and the interaction term.

In order to calculate the broad-sense heritability within each treatment, a mixed effect linear model was fitted to the data:

$$\text{Phenotype}_{ij} = \beta_0 + u_i^{Block} + v_j^{Genotype} + \epsilon_{ij}$$

$u_i^{Replication}$ is the random effect for the *i*th replication, capturing the variation in phenotype attributable to different blocks (replications). $v_j^{Genotype}$ is the random effect for the *j*th genotype, capturing the variation in phenotype attributable to genetic differences. The broad-sense heritability was calculated based on the formula:

$$H^2 = \frac{\sigma^2_{Genotype}}{\sigma^2_{Total}} = \frac{\sigma^2_{Genotype}}{\sigma^2_{Genotype} + \sigma^2_{Residual}}$$

The same mixed effect linear model was also used to generate Best Linear Unbiased Predictor (BLUP) in order to accurately estimate the genetic values of genotypes within each treatment. The estimated deviations of each genotype's effect from the overall mean, adjusted for the block effect, was calculated by 'ranef' function.

The correlations between the genotypic effects (BLUPs) and different phenotypes within each treatment were calculated by 'rcorr' via Pearson linear correlations. The correlations ($r_{G\times E}$) between the genotypic effects (BLUPs) and different treatments within each phenotype were

also calculated by the same method. The Fisher Z test was used to test if the correlation coefficient is significantly different from zero. All statistical analyses were conducted using R statistical software (R Core Team, 2017).

QTL mapping

Prior to QTL mapping, the genetic distance between all SNP markers was estimated by ASMap (Taylor & Butler, 2017), and any markers showing abnormal recombination rates (>150 cM) in juxtaposition were manually removed. The chromosomal locations and physical order of the markers were based on our recently generated high-quality reference genome (Kuo *et al.*, 2024). The QTL mapping based on the BLUPs was conducted in R/qtl package using the normal model and the Haley-Knott regression method with 1,000 permutations (Broman *et al.*, 2003). After a significant QTL was found, the QTL×Treatment (Q×E) interaction was tested by a simple linear model with the QTL, Treatment and their interaction modeled as fixed effects. Subsequently, the significance of the interaction was tested by type-III ANOVA.

5.4 Results

Phenotypic responses of the parental accessions between treatments

1. We simulated a progressive drought stress in greenhouse in order to document the phenotypic responses between different the accessions from distinct habitats (**Supplementary Fig. S5.1**).
2. We found that DMN_010, STL_0701, and GFL_007 have distinct vegetative and reproductive outputs in the controlled and drought treatments.
3. STL_0701 produced the greatest shoot biomass in the control but not in the drought

treatment (**Fig. 5.2A**).

4. STL_0701 produced the greatest root biomass in both control and drought treatments (**Fig. 5.2B**).
5. Jointly, STL_0701 produced the greatest total biomass in both control and drought treatments (**Fig. 5.2C**).
6. Accessions and treatments did not have any effects on the shoot/root biomass (**Fig. 5.2D**).
7. STL_0701 generated significantly greater root/shoot biomass than the other two accessions (**Fig. 5.2E**). It is likely a result of adaption to its prolong-drought environment in summer.
8. DMN_010 had the minimum drop of water potential at the midday of drought treatment (**Fig. 5.2F**), suggesting DMN_010 was experiencing a milder drought stress than the other two accession.
9. Three accessions had distinct leaf thickness. GFL_007 had the greatest leaf thickness; DMN had the least leaf thickness; STL_0701 had the leaf thickness between the other two accessions. The treatments did not affect the distribution of leaf thickness (**Fig. 5.2G**).
10. STL_0701 had the greatest average growth rate in the control treatment. In contrast, the average growth rate is globally low in the drought treatment (**Fig. 5.2H**; **Supplementary Fig. S5.1A**)
11. GFL_007 produced the greatest number of inflorescences in both control and drought environment, which is an eight-week interval. STL_0701 barely produced any inflorescences. DMN_010 is in between (**Fig. 5.2I**; **Supplementary Fig. S5.1B**).
12. GFL_007 had the highest net photosynthesis rate (**Fig. 5.3A**) and stomatal conductance (Fig. 3B) than other two accessions in the control treatment. There is no difference between net photosynthesis rate in the drought treatment.
13. In the drought treatment, both GFL_007 and STL_0701 show enhanced water use efficiency (WUE); However, DMN_010 had the same WUE between the treatment (**Fig. 5.3C**). The unchanged WUE of DMN_010 could be attributed to the milder drought stress it was

experiencing.

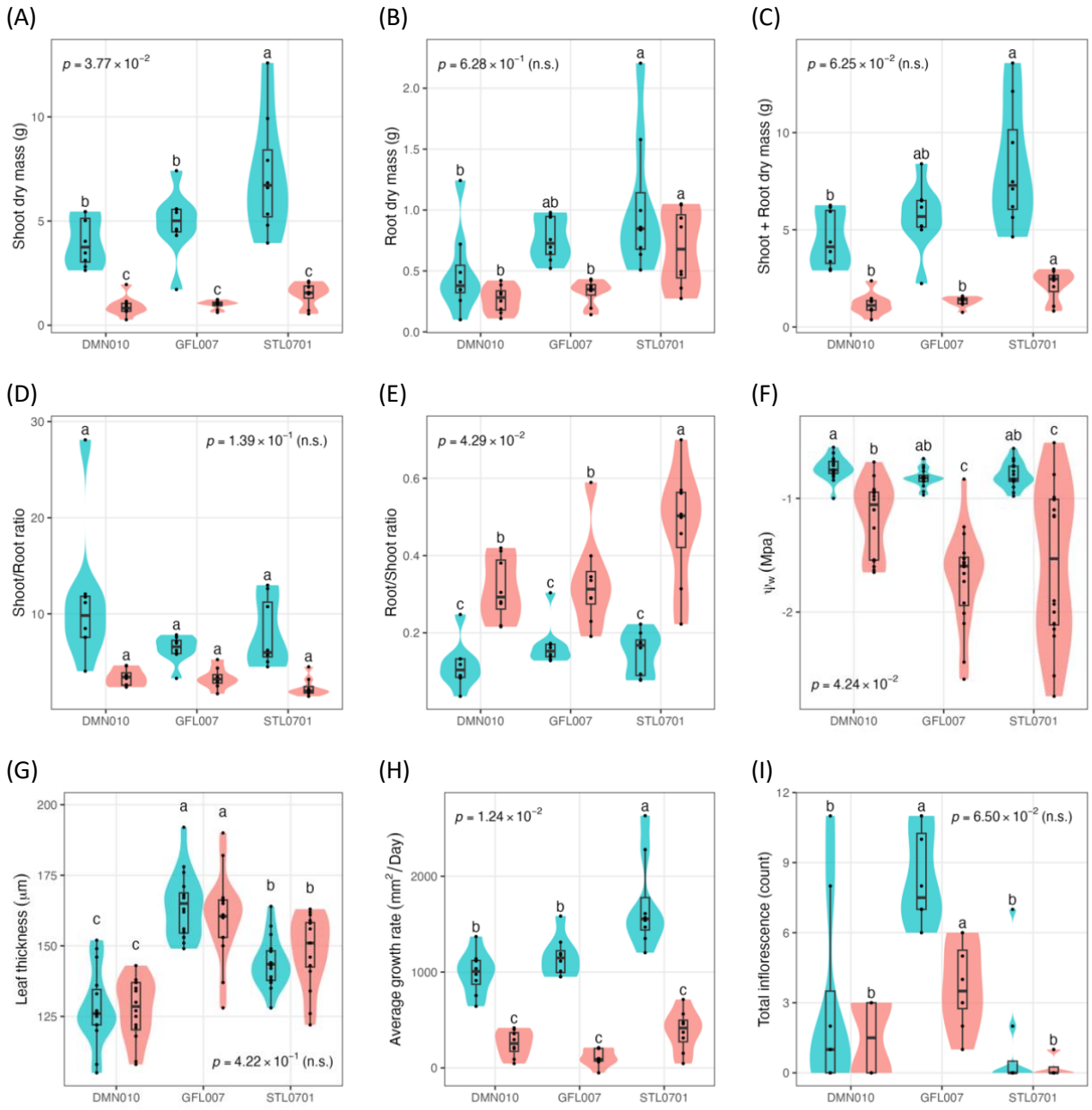


Figure 5.2 Morphological responses of the parental accessions in the control (blue, unlimited water supply) and drought (red, limited water supply) treatments.

The treatment was applied from 2021-08-09 to 2021-09-22 (end of experiment). **(A-E)**, the biomass was measured at the end of the experiment. **(F-G)**, the water potential and leaf thickness were measured at the midday from 2021-09-16 to 2021-09-22. **(H)**, average growth rate was the mean of the weekly vegetative (green) area increase from 2021-08-09 to 2021-09-

12. **(I)**, total inflorescence is the sum of inflorescence produced from 2021-08-09 to 2021-09-12. The interaction between the accessions and treatments was tested by type-II ANOVA. If the interaction is significant ($p < 0.05$), the Tukey multiple comparisons were used to group the accessions and treatment jointly. Otherwise, the comparisons were conducted within each treatment.

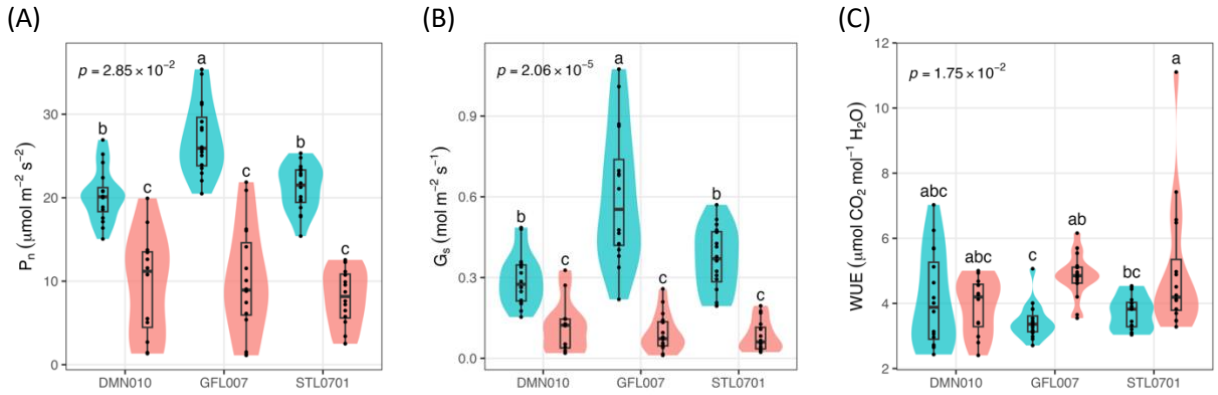


Figure 5.3 Photosynthetic physiology of the parental accessions in the control (Blue, unlimited water supply) and drought (red, limited water supply) treatments.

The measurement was conducted at the midday from 2021-09-16 to 2021-09-22. **(A)**, Net photosynthesis rate. **(B)**, stomata conductance. **(C)**, water use efficiency (WUE).

Subgenome-specific transcriptomics

1. At molecular level, we documented the differential gene expression patterns between the accessions and between the treatments. In order to test our hypothesis that the diploid progenitors' genome (subgenome) have unequal contributions to white clover's adaptation in drought-prone environments, we separated the gene expression data into *T. occidentale* subgenome and *T. pallescens* subgenome based on the physical locations of the gene models.
2. At total transcriptomic level, we found that the genotypes have a major effect on clustering (**Fig. 5.4A, B**). The three genotypes show clear cluster on the PCA plots. Within the genotype cluster, the treatments separate the samples into two sub-clusters. STL_007 shows the most distinct gene expression patterns between the treatments while DMn_010 shows the least. This finding is paralleled to the morphological responses to the drought treatment, in which STL had particular more investment into root growth.
3. The total transcriptomic patterns of the two subgenomes are almost identical, suggesting the lack of subgenome biased gene expression patterns.
4. When we filtered for the genes that were showing significant differential expressions between the treatments, PC1 becomes the major axis separating the treatment effect, and the genotypes are no longer the major clustering factor on the PCA plot (**Fig. 5.4C, D**).
5. The gene expression patterns of the controls of DMN_010 are more similar to those of the controls of STL_0701 than the droughts of DMN_010 itself, *vice versa*. PC2 does not separate DMN_010 and STL_0701.
6. In contrast, GFL_007 shows distinct patterns than the other two genotypes, which is separated by PC2.
7. The two subgenomes still have highly similar gene expression pattern after the filtering; however, it is more visually distinguishable than the gene expressions at the total transcriptomic level. It suggests that subgenome biased gene regulations still exist, especially for the genes that are responsive to drought stress.

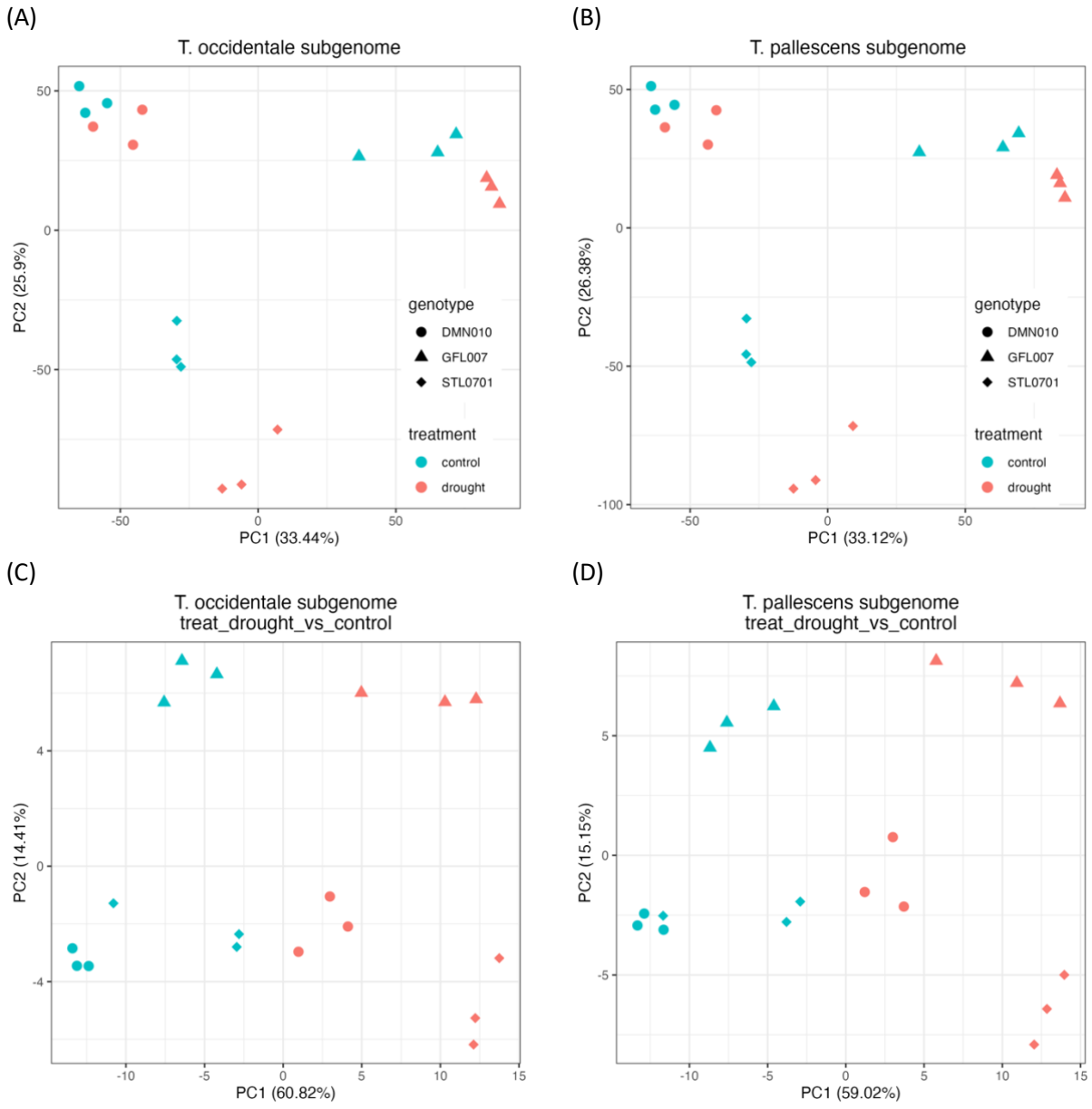


Figure 5.4 PCA of the gene expressions between the subgenomes of the parental accessions in the control (unlimited water supply) and drought (limited water supply) treatments.

(A-B), subgenome expressions of all genes. **(C-D)**, subgenome expressions of the genes that show differential expressions between the control and the drought treatments ($p < 0.05$).

Cyanogenesis in drought stress

1. The effect of cyanogenesis on white clover's fitness under drought stress was investigated by mixed effect linear models, which can reduce the block's effects (e.g., seasonality) (**Supplementary Fig. 5.2-5**).
2. We measured the shoot and root biomass, weekly leaf area, and cumulative inflorescence among the F3 progeny of the intercrossed population of DMN_010 and STL_0701.
3. The F3 progeny included 92 "AcLi" plants, 81 "Acli" plants, 91 "acLi" plants, and 36 "acli" plants.
4. We found the application of the drought treatment significantly affects all measured phenotypes; however, there is no difference between the cyanotypes, nor between the cyanotype×treatment interactions (**Table 5.1**).
5. In contrast to our hypothesis, we did not detect any fitness contributions from being cyanogenic in white clover.

Table 5.1 Type II Analysis of Variance Table with Satterthwaite's method

Trait	Factor	Sum Sq	Mean Sq	NumDF	DenDF	F value	P value
Shoot dry mass							
	Cyanotype	2.8703	0.9568	3	1787	0.84	0.47
	Trt	2063.7685	2063.7685	1	1787	1817.50	1.43E-274
	Cyanotype:Trt	1.4401	0.4800	3	1787	0.42	0.74
Root dry mass							
	Cyanotype	0.0169	0.0056	3	1785	0.12	0.95
	Trt	41.1626	41.1626	1	1785	859.97	1.25E-154
	Cyanotype:Trt	0.1148	0.0383	3	1785	0.80	0.49
Leaf area w2							
	Cyanotype	5686865.37	1895621.79	3	1787	0.55	0.65
	Trt	234311780.94	234311780.94	1	1787	68.20	2.84E-16***
	Cyanotype:Trt	5355360.24	1785120.08	3	1787	0.52	0.67
Leaf area w3							
	Cyanotype	61608986.54	20536328.85	3	1787	1.40	0.24
	Trt	14679948171.93	14679948171.93	1	1787	1002.47	4.99E-175***
	Cyanotype:Trt	28379656.39	9459885.46	3	1787	0.65	0.59
Leaf area w4							
	Cyanotype	111724994	37241664.8	3	1787	0.68	0.56
	Trt	1.4438E+11	1.4438E+11	1	1787	2639.99	0.00***
	Cyanotype:Trt	61222223.1	20407407.7	3	1787	0.37	0.77
Leaf area w5							
	Cyanotype	128883785	42961261.7	3	1787	0.34	0.80
	Trt	5.0061E+11	5.0061E+11	1	1787	3904.63	0.00***
	Cyanotype:Trt	131177764	43725921.4	3	1787	0.34	0.80
Inflorescence count							
	Cyanotype	59.04	19.68	3	1787	1.82	0.14
	Trt	174.22	174.22	1	1787	16.13	6.17E-05***
	Cyanotype:Trt	2.20	0.73	3	1787	0.07	0.98

Heritability and environmental interactions

1. In order to document comprehensive view of molecular mechanisms under drought stress, we generated a set of 6,491 high-density and sequencing based markers for investigating molecular responses to drought treatment and their genotype × environment interactions, which (if any) could be the underlying mechanisms of local adaptation.
2. The broad-sense heritability (H^2) was calculated for all measured phenotypes within each treatment (**Supplementary Table S5.1**). The phenotypes from the control treatment generally have higher H^2 than the phenotypes from the drought treatment. Except for the number of cumulative inflorescences, which $H^2_{Drought}$ (0.28) is higher than $H^2_{Control}$ (0.24), the other phenotypes have H^2 ranges from 0.20 to 0.40 in the control treatment, and have H^2 ranges from 0.07 to 0.16 in the drought treatment.
3. In addition to the heritability, we also tested the effect of treatments and treatment × genotype interactions on the measured phenotypes (**Supplementary Table S5.1**). We found that the drought treatment has effects on all measured phenotypes except for the number of cumulative inflorescences, which is likely to be genetically determined. In contrast, only the leaf area measured at the fourth week of the treatments shows significant treatment × genotype interaction.
4. We calculated the Best Linear Unbiased Predictor (BLUP) of each measured phenotype in order to estimate the genotypic effects within each treatment. Subsequently, the BLUPs were used to calculate the correlations ($r_{G\times E}$) between the genotypic effects (BLUPs) and treatments.
5. Positive values were observed in all phenotypes, indicating the prevalence of genetic effects. Consistent with the previous finding that the treatments have undetectable effects on the number of cumulative inflorescences, it has the highest $r_{G\times E}$ value (0.76) compared to other phenotypes ($r_{G\times E}$ = 0.34-0.62).
6. We also analyzed the correlations between different phenotypes of the same treatment because it is possible to detect some allocation trade-offs between phenotypes when the

disposable energy is limited.

7. We found significant correlations between all vegetative phenotypes; however, there is no correlations between vegetative and reproductive phenotypes except for the two correlations in the drought treatment (**Supplementary Fig. S5.6)**. The number of cumulative inflorescences is positively correlated with the shoot biomass, and is negatively correlated with the root biomass in the drought treatment. It suggests that the reproductive output is dependent on vegetative output only when water is limited. Different genotypes have divergent strategies to invest either root biomass or early flowering when the water is limited.

QTL mapping

1. We mapped the QTLs that are significantly associated with the variations of the BLUPs (**Fig. 5.5**, raw data-based QTL mapping is available at **Supplementary Fig. S5.7-9**).
2. Three QTLs are detected on Chr. 4, Chr. 7, and Chr. 8, respectively. The DMN_010 alleles ("B" alleles) are generally associated with the lower phenotypic values, except for the QTL on Chr. 4. The QTL on Chr. 4 is associated with all vegetative phenotypes in the control treatment and the reproductive phenotype in the drought treatment. The "B" allele had opposite effects on this QTL: The "B" allele at this QTL could promote flowering in the drought treatment but could repress the vegetative growth in the control treatment.
3. The QTL×Treatment (Q×E) interaction was tested at the three detected QTLs.
4. For the QTL on Chr. 4, the interactions are significant for the leaf area, shoot biomass, and root biomass, in which the "B" alleles of the QTLs only had negative effects in the control treatment but have no effect in the drought treatment (**Fig. 5.6B, C, D**). In contrast, the "B" allele of this QTL could promote flowering in regardless of the treatments (**Fig. 5.6A**).
5. For the QTLs on Chr. 7 and Chr. 8, the "B" allele of these QTLs could repress the vegetative growth (leaf area) and flowering, respectively (**Fig. 5.6E, F**). No interaction is detected.

Landscape genomics

1. We compared the detected QTLs of this study with the natural genetic variations that are particularly associated with Aridity index (an estimate for drought stress) of North American white clover populations (**Fig. 5.7**).
2. Although we did not find a perfect overlap between the two approaches, we found there is a strong signal on Chr. 4, which is overlapped with the 1-drop LOD range of the QTL corresponding for the flowering in the drought treatment and the signal is also close to other vegetative QTLs.
3. The aridity-associated signal on Chr. 4 is spanning up to 5Mbp, which could be a structural variant that is associated with the aridity gradient of North America.

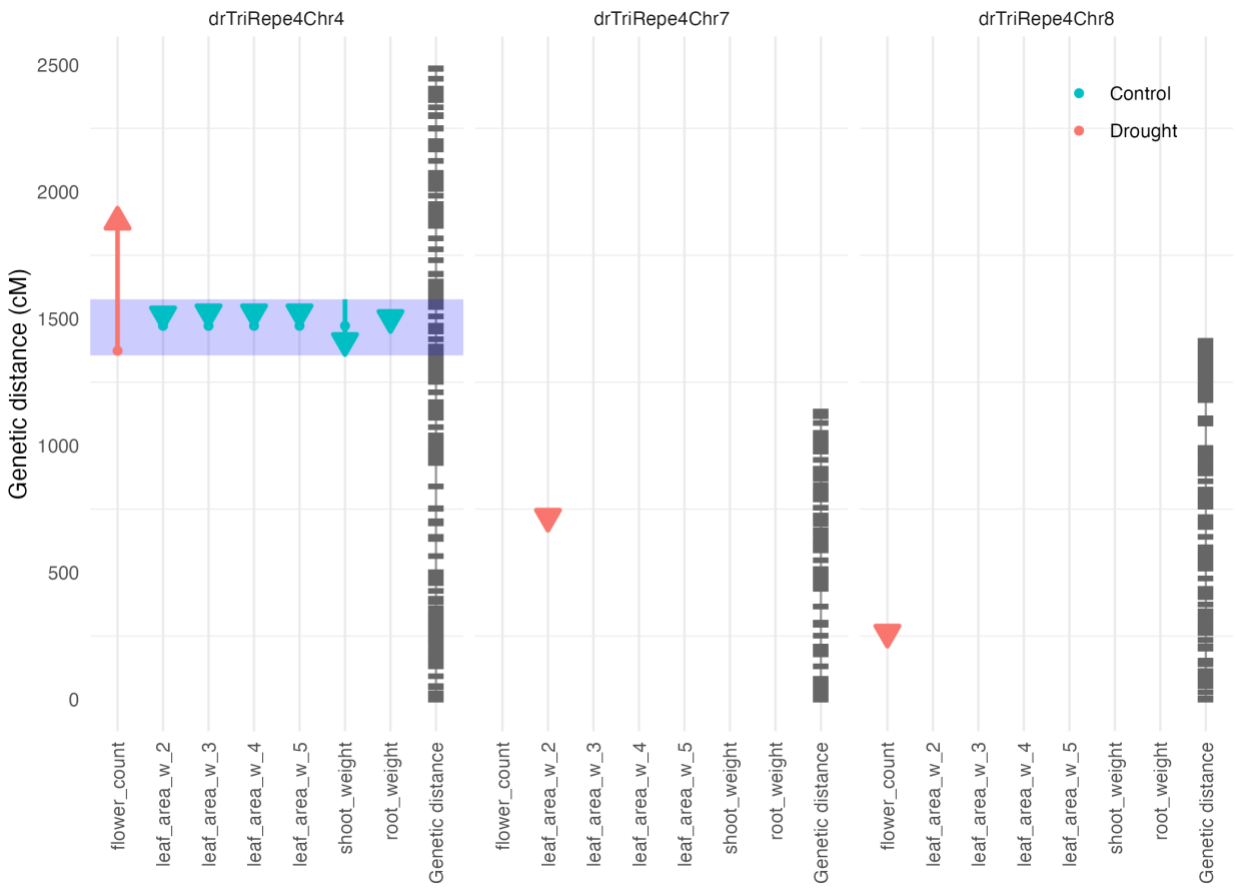
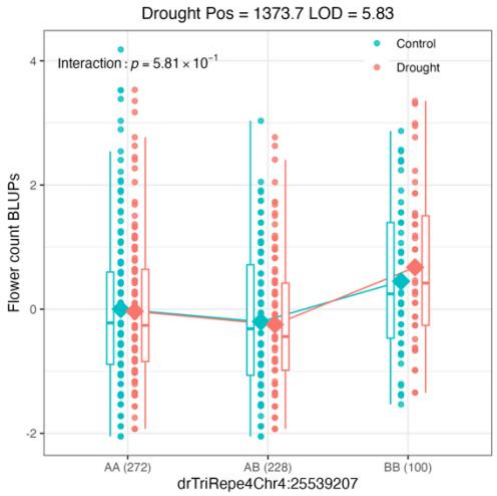


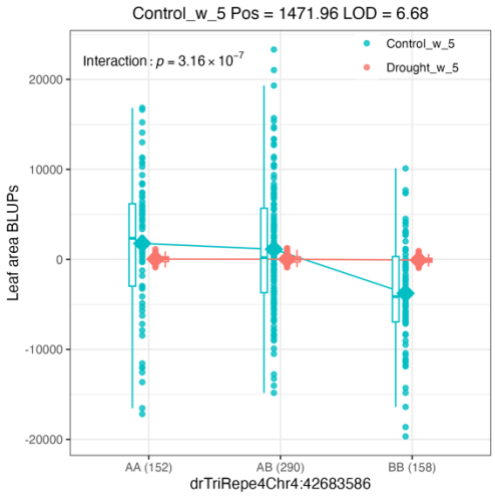
Figure 5.5 Fitness related QTLs in the control and drought treatments.

Flower count, number of inflorescences produced during the six-week experiment period. Leaf area, relative leaf area changes compared to "w_1". The number after the "w" indicates the week of the experiment. The experiment started at "w_0". Shoot/root weight, dry mass of shoot/root after six-week experiment. Best linear unbiased prediction (BLUP) was used to estimate the genotypic effect across all experimental blocks and replications. The QTLs are shown if they are statistically significant after 1,000 times permutation test (p value < 0.1). The dots indicate the centers of the QTLs. The segments indicate the 1-drop LOD ranges. The arrowheads indicate the additive effects of the DMN010 allele ("B") compared to the GFL007 allele ("A"). The blue tile highlights the fitness trade-off (antagonistic pleiotropy) between the control and drought treatments.

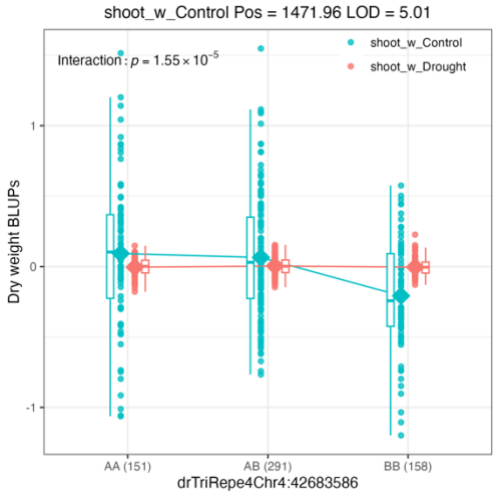
(A)


Drought Pos = 1373.7 LOD = 5.83
Interaction : $p = 5.81 \times 10^{-1}$
Control
Drought
Flower count BLUPs
AA (272)
AB (228)
BB (100)
drTriRepe4Chr4:25539207

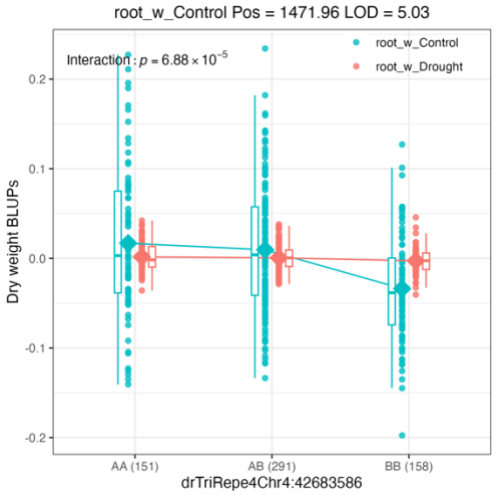
(B)***


Control_w_5 Pos = 1471.96 LOD = 6.68
Interaction : $p = 3.16 \times 10^{-7}$
Control_w_5
Drought_w_5
Leaf area BLUPs
AA (152)
AB (290)
BB (158)
drTriRepe4Chr4:42683586

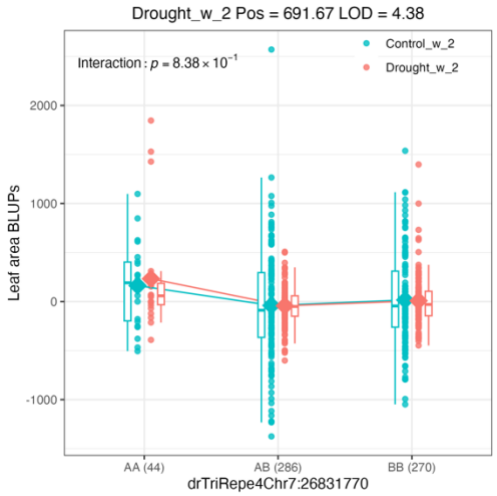
(C)***


shoot_w_Control Pos = 1471.96 LOD = 5.01
Interaction : $p = 1.55 \times 10^{-5}$
shoot_w_Control
shoot_w_Drought
Dry weight BLUPs
AA (151)
AB (291)
BB (158)
drTriRepe4Chr4:42683586

(D)***


root_w_Control Pos = 1471.96 LOD = 5.03
Interaction : $p = 6.88 \times 10^{-5}$
root_w_Control
root_w_Drought
Dry weight BLUPs
AA (151)
AB (291)
BB (158)
drTriRepe4Chr4:42683586

(E)


Drought_w_2 Pos = 691.67 LOD = 4.38
Interaction : $p = 8.38 \times 10^{-1}$
Control_w_2
Drought_w_2
Leaf area BLUPs
AA (44)
AB (286)
BB (270)
drTriRepe4Chr7:26831770

(F)

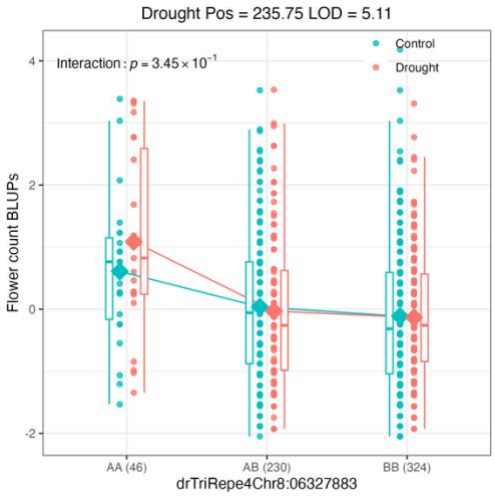

Drought Pos = 235.75 LOD = 5.11
Interaction : $p = 3.45 \times 10^{-1}$
Control
Drought
Flower count BLUPs
AA (46)
AB (230)
BB (324)
drTriRepe4Chr8:06327883

Figure 5.6 Genotype effects of the significant fitness related QTLs between treatments (QTL × Environment interaction).

(A-D), QTLs on drTriRepe4Chr4. **(E)**, QTL on drTriRepe4Chr7. **(F)**, drTriRepe4Chr8. The title of each panel includes the QTL position and the LOD score. The parental line GFL007 has "AA" genotype. The parental line DMN010 has "BB" genotype. The interaction between the genotype and the treatment was test by type-III ANOVA by using package "car" in R. The *p* values lower than 0.05 (*), 0.01 (**), 0.001 (***) are labeled.

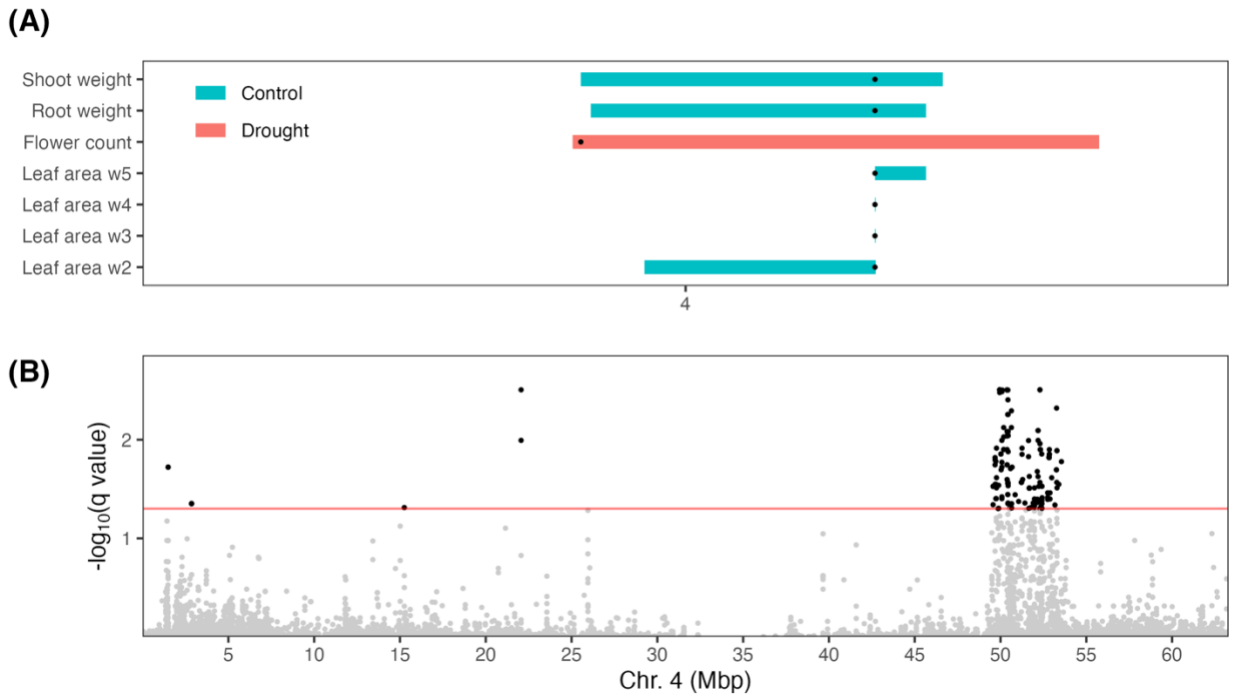


Figure 5.7 Overlap between the detected QTLs from the mapping population and the SNPs detected from genotype-environmental association (GEA) of Aridity index.

(A), the same 1-drop QTL ranges as Fig. 4 but plotted in the physical locations of the corresponding markers. **(B)**, GEA result by using Aridity index as the predictor for the genetic variation of the 415 North American wild accessions. The p values have been converted to q value for false discovery rate (FDR) correction. The threshold is set at FDR = 0.05.

5.5 Funding

This work was supported by the William H. Danforth Plant Science Graduate Research Fellowship in the Division of Biology and Biomedical Sciences at Washington University to W.-H.K.; by the scholarship of Taiwan Ministry of Education to W.-H.K.; and by the U.S. National Science Foundation (grant number IOS-1557770) to K.M.O.'

5.6 Acknowledgements

We thank the Washington University greenhouse staffs, especially Michael Dyer, Michael Stephan and Hammy Sorkin, for care of plants used in the study. We thank an undergraduate team of Washington University, including Eimear Cunningham, Keiko Farah, Nick Ho, Grace Li, Angel Lu, Chinh Mach, Erin Reardon, Emily Shen, Emily Talkow, Simon Traub-Epstein, and W.-H. K.'s friends, including Chao-Cheng Kuo and Renee Wu, helping the phenotyping process in the greenhouse. We thank anonymous reviewers for their suggestions and comments. W.-H. K. designed and conducted the experiment, analyzed the data, interpreted the results, and wrote the manuscript. S. L. conducted part of the genotyping experiments and assisted the works in the greenhouse. K. M. O. conceived the project, interpreted the results, and edited the manuscript. All authors read and approved the final manuscript.

5.7 Reference

Agrena J, Oakley CG, McKay JK, Lovell JT, Schemske DW. 2013. Genetic mapping of adaptation reveals fitness tradeoffs in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **110**(52): 21077-21082.

Bieker VC, Battlay P, Petersen B, Sun X, Wilson J, Brealey JC, Bretagnolle F, Nurkowski K, Lee C, Barreiro FS, et al. 2022. Uncovering the genomic basis of an extraordinary plant invasion. *Sci Adv* **8**(34): eabo5115.

Briggs D, Walters SM. 2016. *Plant variation and evolution*. Cambridge: Cambridge University Press.

Broman KW, Wu H, Sen Ś, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**(7): 889-890.

Burghardt LT, Metcalf CJE, Donohue K. 2016. A cline in seed dormancy helps conserve the environment experienced during reproduction across the range of Arabidopsis thaliana. *American Journal of Botany* **103**(1): 47-59.

Chan W-P, Lenoir J, Mai G-S, Kuo H-C, Chen IC, Shen S-F. 2024. Climate velocities and species tracking in global mountain regions. *Nature*.

Daday H. 1958. Gene frequencies in wild populations of *Trifolium repens* L. III. World distribution. *Heredity* **12**(2): 169-184.

Daday H. 1965. Gene frequencies in wild populations of *Trifolium repens* L. IV. Mechanism of natural selection. *Heredity* **20**(3): 355-365.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1): 15-21.

Fernández-Pascual E, Jiménez-Alfaro B, Caujapé-Castells J, Jaén-Molina R, Díaz TE. 2013. A local dormancy cline is related to the seed maturation environment, population genetic composition and climate. *Annals of Botany* **112**(5): 937-945.

Futuyma DJ, Kirkpatrick M. 2017. *Evolution*. Sunderland, Massachusetts: Sinauer Associates, Inc., Publishers Sunderland, Massachusetts.

Gould BA, Stinchcombe JR. 2017. Population genomic scans suggest novel genes underlie convergent flowering time evolution in the introduced range of *Arabidopsis thaliana*. *Molecular Ecology* **26**(1): 92-106.

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, et al. 2019. Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. *The Plant Cell* **31**(7): 1466-1487.

Guan J, Zhang J, Gong D, Zhang Z, Yu Y, Luo G, Somta P, Hu Z, Wang S, Yuan X, et al. 2022. Genomic analyses of rice bean landraces reveal adaptation and yield related loci to accelerate breeding. *Nature Communications* **13**(1): 5707.

Hämälä T, Mattila TM, Savolainen O. 2018. Local adaptation and ecological differentiation under selection, migration, and drift in Arabidopsis lyrata*. *Evolution* **72**(7): 1373-1386.

Hughes MA. 1991. The cyanogenic polymorphism in *Trifolium repens* L. (white clover). *Heredity* **66**(1): 105-115.

Kjærgaard T. 2003. A plant that changed the world: The rise and fall of clover 1000-2000. *Landscape Research* **28**(1): 41-49.

Kooyers NJ, Gage LR, Al-Lozi A, Olsen KM. 2014. Aridity shapes cyanogenesis cline evolution in white clover (*Trifolium repens* L.). *Molecular Ecology* **23**(5): 1053-1070.

Kooyers NJ, Olsen KM. 2012. Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.). *Molecular Ecology* **21**(10): 2455-2468.

Kooyers NJ, Olsen KM. 2013. Searching for the bull's eye: agents and targets of selection vary among geographically disparate cyanogenesis clines in white clover (*Trifolium repens* L.). *Heredity* **111**(6): 495-504.

Kreiner JM, Hnatovska S, Stinchcombe JR, Wright SI. 2023. Quantifying the role of genome size and repeat content in adaptive variation and the architecture of flowering time in *Amaranthus tuberculatus*. *PLOS Genetics* **19**(12): e1010865.

Kreiner JM, Latorre SM, Burbano HA, Stinchcombe JR, Otto SP, Weigel D, Wright SI. 2022. Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science* **378**(6624): 1079-1085.

Kuo W-H, Small LL, Olsen KM. 2023. Variable expression of cyanide detoxification and tolerance genes in cyanogenic and acyanogenic white clover (*Trifolium repens*). *American Journal of Botany* **110**(10): e16233.

Kuo W-H, Wright SJ, Small LL, Olsen KM. 2024. *De novo* genome assembly of white clover (*Trifolium repens* L.) reveals the role of copy number variation in rapid environmental adaptation. *Research Square*.

Lee G, Sanderson BJ, Ellis TJ, Dilkes BP, McKay JK, Agren J, Oakley CG. 2024. A large-effect fitness trade-off across environments is explained by a single mutation affecting cold acclimation. *Proc Natl Acad Sci U S A* **121**(6): e2317461121.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**(7): 923-930.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**(12): 550.

Machingura M, Salomon E, Jez JM, Ebbs SD. 2016. The beta-cyanoalanine synthase pathway: Beyond cyanide detoxification. *Plant, Cell & Environment* **39**(10): 2329-2341.

Marchin RM, Ossola A, Leishman MR, Ellsworth DS. 2019. A simple method for simulating drought effects on plants. *Front Plant Sci* **10**: 1715.

McGoey BV, Hodgins KA, Stinchcombe JR. 2020. Parallel flowering time clines in native and introduced ragweed populations are likely due to adaptation. *Ecology and Evolution* **10**(11): 4595-4608.

McMahon J, Sayre R, Zidenga T. 2021. Cyanogenesis in cassava and its molecular manipulation for crop improvement. *Journal of Experimental Botany* **73**(7): 1853-1867.

Møller BL. 2010. Functional diversifications of cyanogenic glucosides. *Current Opinion in Plant Biology* **13**(3): 337-346.

Nielsen LJ, Stuart P, Picmanova M, Rasmussen S, Olsen CE, Harholt J, Moller BL, Bjarnholt N. 2016. Dhurrin metabolism in the developing grain of *Sorghum bicolor* (L.) Moench investigated by metabolite profiling and novel clustering analyses of time-resolved transcriptomic data. *BMC Genomics* **17**(1): 1021.

Olsen KM, Goad DM, Wright SJ, Dutta ML, Myers SR, Small LL, Li LF. 2021. Dual-species origin of an adaptive chemical defense polymorphism. *New Phytologist* **232**(3): 1477-1487.

Olsen KM, Hsu SC, Small LL. 2008. Evidence on the molecular basis of the *Ac*/*ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* **179**(1): 517-526.

Olsen KM, Small LL. 2018. Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytologist* **219**(2): 757-766.

Olsen KM, Sutherland BL, Small LL. 2007. Molecular evolution of the *Li*/*li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Molecular Ecology* **16**(19): 4180-4193.

Park DS, Breckheimer I, Williams AC, Law E, Ellison AM, Davis CC. 2018. Herbarium specimens reveal substantial and unexpected variation in phenological sensitivity across the eastern United States. *Philos Trans R Soc Lond B Biol Sci* **374**(1763).

Poplin R, Ruano-Rubio V, Depristo MA, Fennell TJ, Carneiro MO, Van Der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*.

Postma FM, Agren J. 2016. Early life stages contribute strongly to local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **113**(27): 7590-7595.

Price N, Moyers BT, Lopez L, Lasky JR, Monroe JG, Mullen JL, Oakley CG, Lin J, Ågren J, Schrider DR, et al. 2018. Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **115**(19): 5028-5033.

R Core Team 2017. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Taylor J, Butler D. 2017. R Package ASMap: Efficient genetic linkage map construction and diagnosis. *Journal of Statistical Software* **79**(6): 1 - 29.

van Boheemen LA, Atwater DZ, Hodgins KA. 2019. Rapid and repeated local adaptation to climate in an invasive plant. *New Phytol* **222**(1): 614-627.

Wang Y, Zhang L, Zhou Y, Ma W, Li M, Guo P, Feng L, Fu C. 2023. Using landscape genomics to assess local adaptation and genomic vulnerability of a perennial herb *Tetrastigma hemsleyanum* (Vitaceae) in subtropical China. *Frontiers in Genetics* **14**: 1150704.

Wright SJ, Cui Zhou D, Kuhle A, Olsen KM. 2018. Continent-wide climatic variation drives local adaptation in North American white clover. *Journal of Heredity* **109**(1): 78-89.

Wright SJ, Goad DM, Gross BL, Muñoz PR, Olsen KM. 2022. Genetic trade-offs underlie divergent life history strategies for local adaptation in white clover. *Molecular Ecology* **31**(14): 3742-3760.

Wunder J, Fulgione A, Torang P, Wotzel S, Herzog M, Obeso JR, Kourmpetis Y, van Ham R, Odong T, Bink M, et al. 2023. Adaptation of perennial flowering phenology across the European range of Arabis alpina. *Proc Biol Sci* **290**(2011): 20231401.

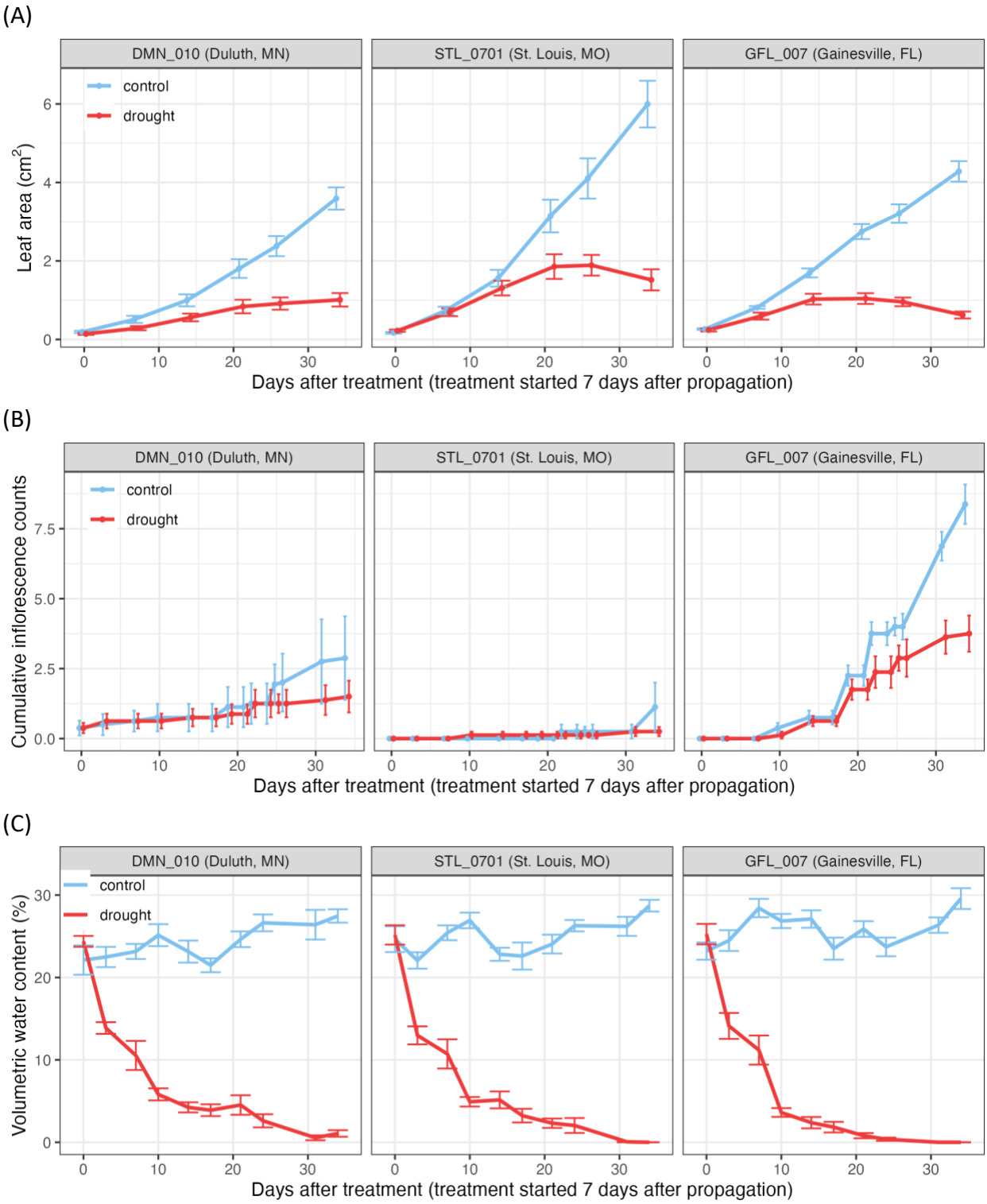
Yan W, Wang B, Chan E, Mitchell-Olds T. 2021. Genetic architecture and adaptation of flowering time among environments. *New Phytol* **230**(3): 1214-1227.

Zeven AC. 1991. Four hundred years of cultivation of Dutch white clover landraces. *Euphytica* **54**(1): 93-99.

5.8 Supplementary Information

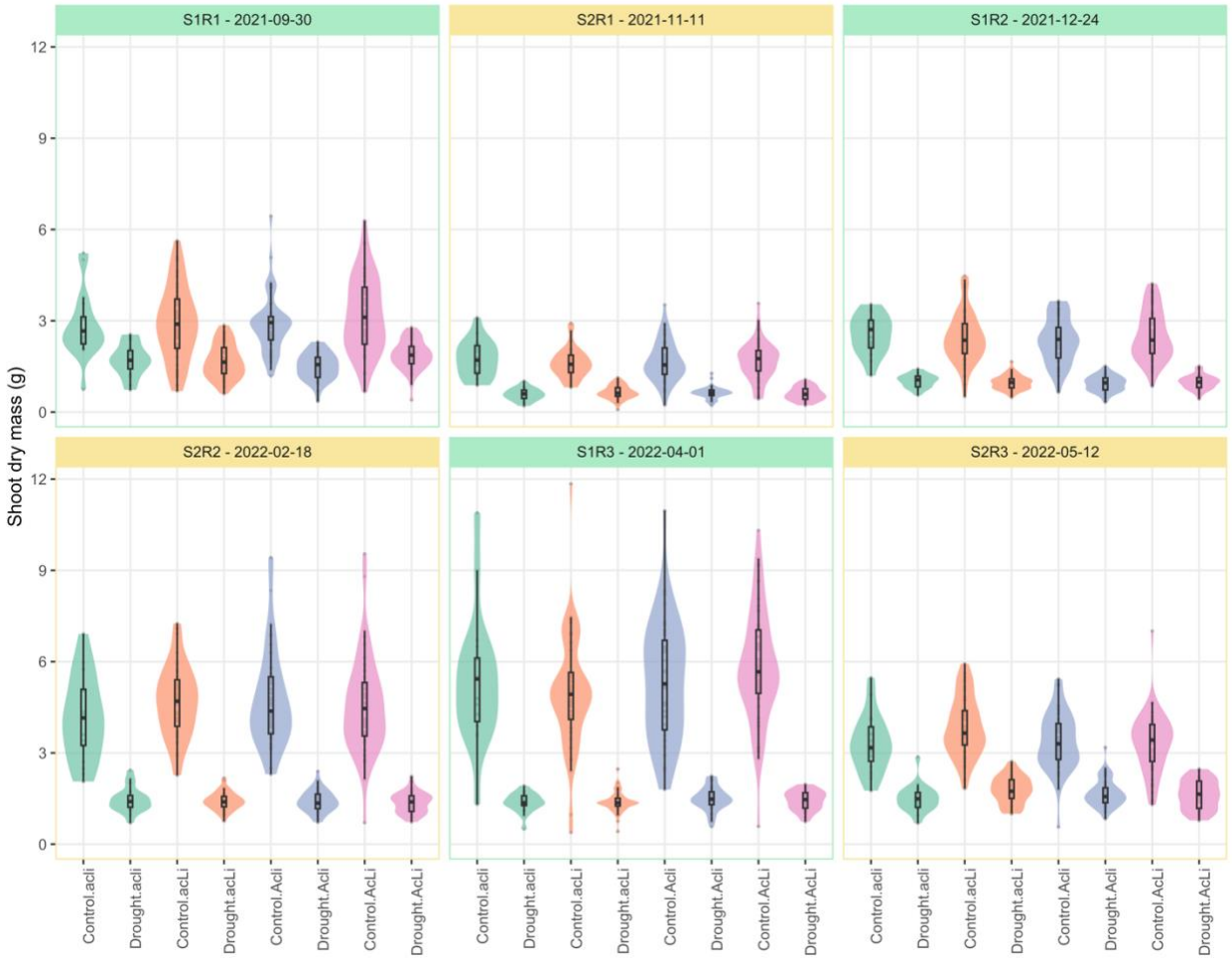
Supplementary Table S5.1 Heritability within each treatment, treatment effects, and genotype × environment interactions.

				Treatment					Treatment:Genotype						
Trait	H^2 control	H^2 drought		NumDF	DenDF	F value	Pr(>F)		NumDF	DenDF	F value	Pr(>F)		$r_{G\times E}$	p_value
flower count	0.24	0.28		1	1196	0.19	0.67		598	835	0.35	1		0.76	0
leaf_area_w_2	0.21	0.09		1	1196	9.33	**0.0023**		598	1179	0.47	1		0.40	6.25E-13
leaf_area_w_3	0.27	0.16		1	1196	24.44	**8.77E-07**		598	2	0.66	0.78		0.62	0
leaf_area_w_4	0.37	0.08		1	1196	28.52	**1.11E-07**		598	927	1.28	**3.59E-04**		0.50	0
leaf_area_w_5	0.40	0.07		1	1196	21.54	**3.84E-06**		598	2	1.48	0.49		0.44	8.88E-16
shoot_weight	0.26	0.11		1	1196	12.24	**4.84E-04**		598	2	0.90	0.67		0.42	2.04E-14
root_weight	0.20	0.08		1	1195	12.22	**4.91E-04**		598	2	0.81	0.71		0.34	1.70E-09



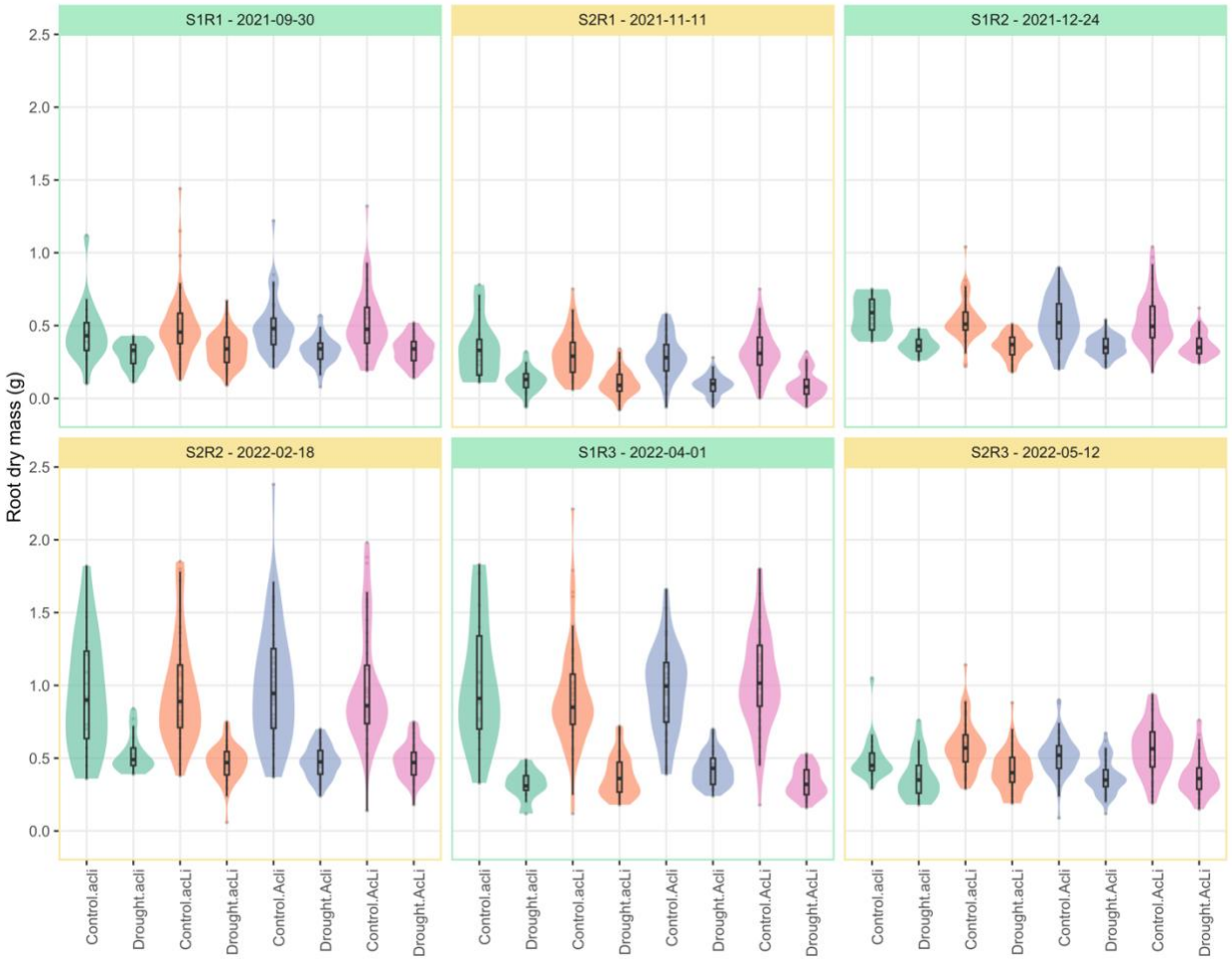
Supplementary Figure S5.1 Life history comparison between accessions under unlimited water (control) and limited water treatments in greenhouse.

(A), leaf area. (B), Cumulative inflorescence counts.



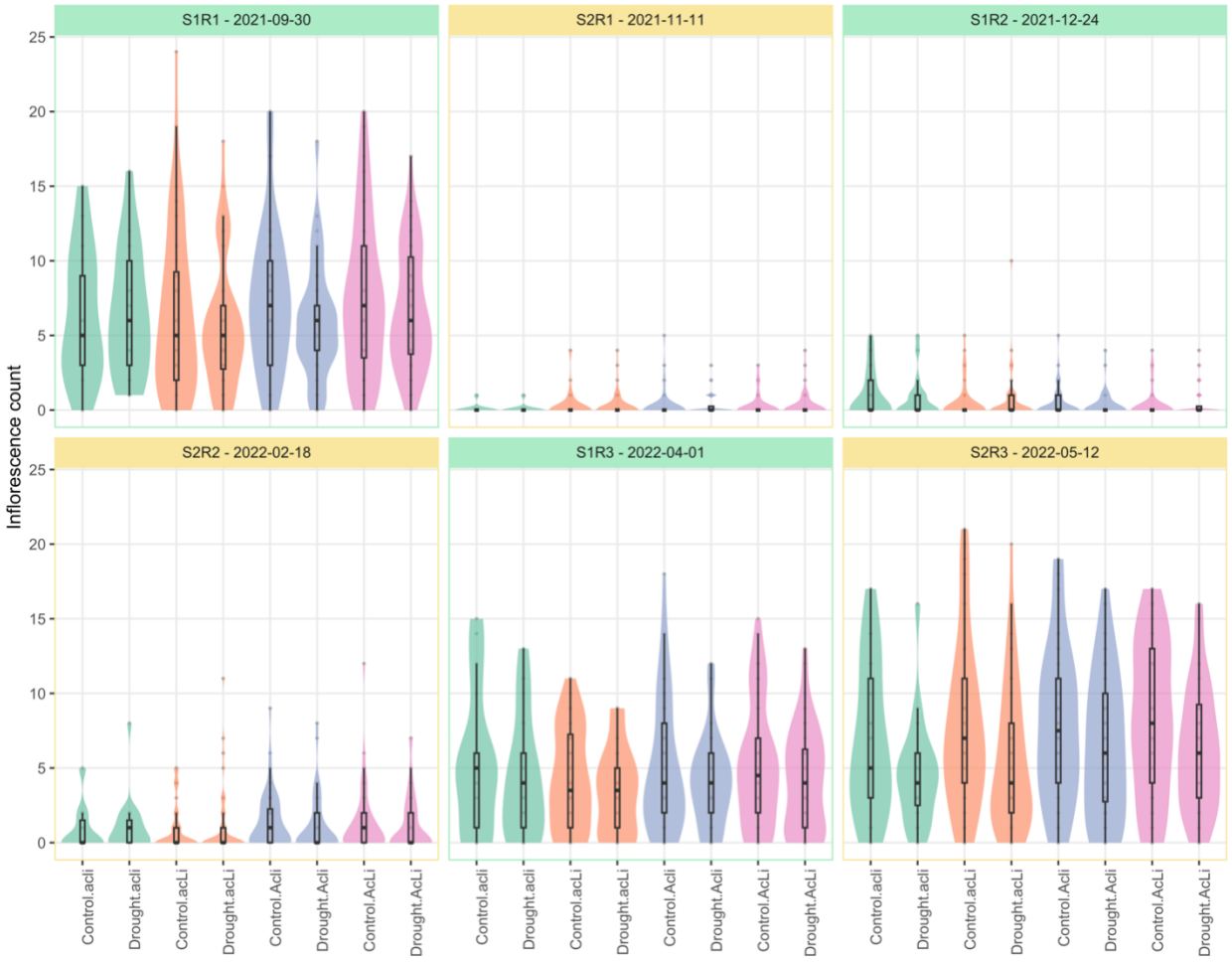
Supplementary Figure S5.2 Shoot dry mass between different cyanotypes.

The facet stripe shows the information of the genotype block (S), the replication (R), and the date of the experiment. Each of the two genotype blocks contained 150 non-overlapping genotypes. The three replications were nested in each genotype block. The date is the first week of the experiment. The linear mixed model: (Shoot dry mass) ~ 1 + Cyanotype + Treatment + Cyanotype: Treatment + (1 | Genotype block) + (1 | Genotype block: Replication). Type II ANOVA with Satterthwaite's method was used to test the significances of the fixed effect variables. Cyanotype, $F_{(3,1787)} = 0.8426$, $p = 0.4705$. Treatment, $F_{(1,1787)} = 1817.5027$, $p = 2e\text{-}16$. Cyanotype: Treatment, $F_{(3, 1787)} = 0.4228$, $p = 0.7367$.



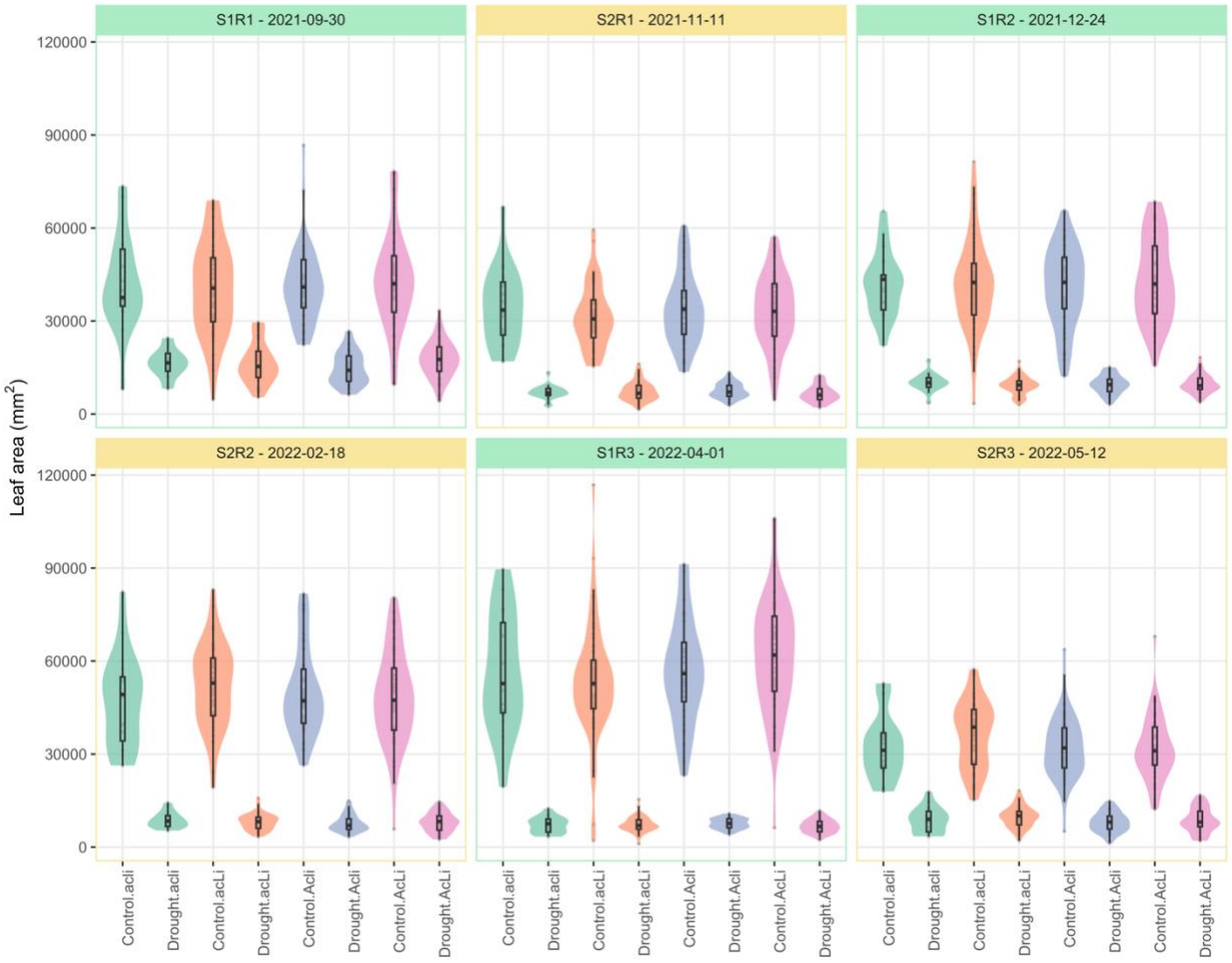
Supplementary Figure S5.3 Root dry mass between different cyanotypes.

The facet stripe shows the information of the genotype block (S), the replication (R), and the date of the experiment. Each of the two genotype blocks contained 150 non-overlapping genotypes. The three replications were nested in each genotype block. The date is the first week of the experiment. The linear mixed model: (Root dry mass) ~ 1 + Cyanotype + Treatment + Cyanotype: Treatment + (1 | Genotype block) + (1 | Genotype block: Replication). Type II ANOVA with Satterthwaite's method was used to test the significances of the fixed effect variables. Cyanotype, $F_{(3,1785)} = 0.118$, $p = 0.9496$. Treatment, $F_{(1,1785)} = 859.9665$, $p = 2e\text{-}16$. Cyanotype: Treatment, $F_{(3,\ 1785)} = 0.7992$, $p = 0.4942$.



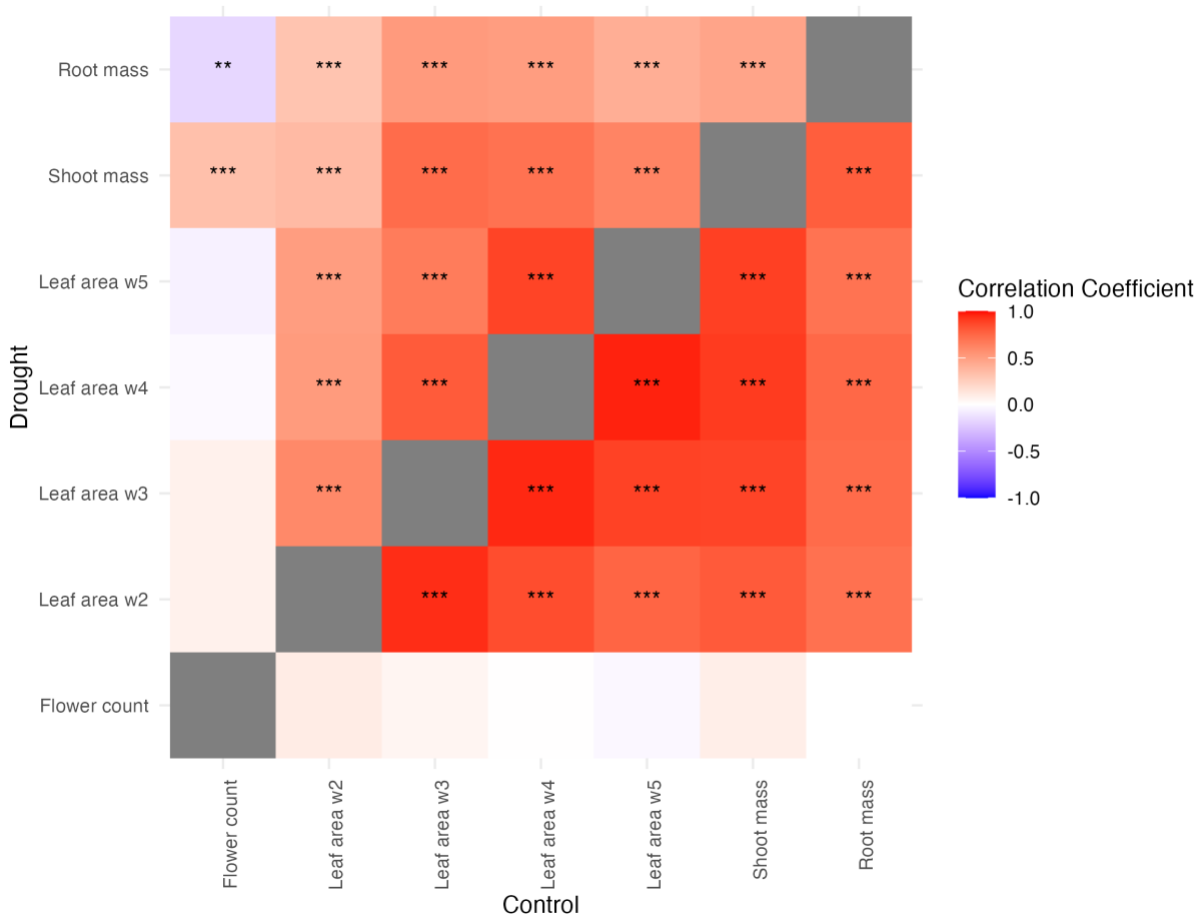
Supplementary Figure S5.4 Inflorescence counts between different cyanotypes.

The facet stripe shows the information of the genotype block (S), the replication (R), and the date of the experiment. Each of the two genotype blocks contained 150 non-overlapping genotypes. The three replications were nested in each genotype block. The date is the first week of the experiment. The linear mixed model: (Inflorescence counts) ~ 1 + Cyanotype + Treatment + Cyanotype: Treatment + (1 | Genotype block) + (1 | Genotype block: Replication). Type II ANOVA with Satterthwaite's method was used to test the significances of the fixed effect variables. Cyanotype, $F_{(3,1787)} = 1.8218$, $p = 0.1411$. Treatment, $F_{(1,1787)} = 16.167$, $p = 6.167e\text{-}5$. Cyanotype: Treatment, $F_{(3,\ 1787)} = 0.0678$, $p = 0.9770$.



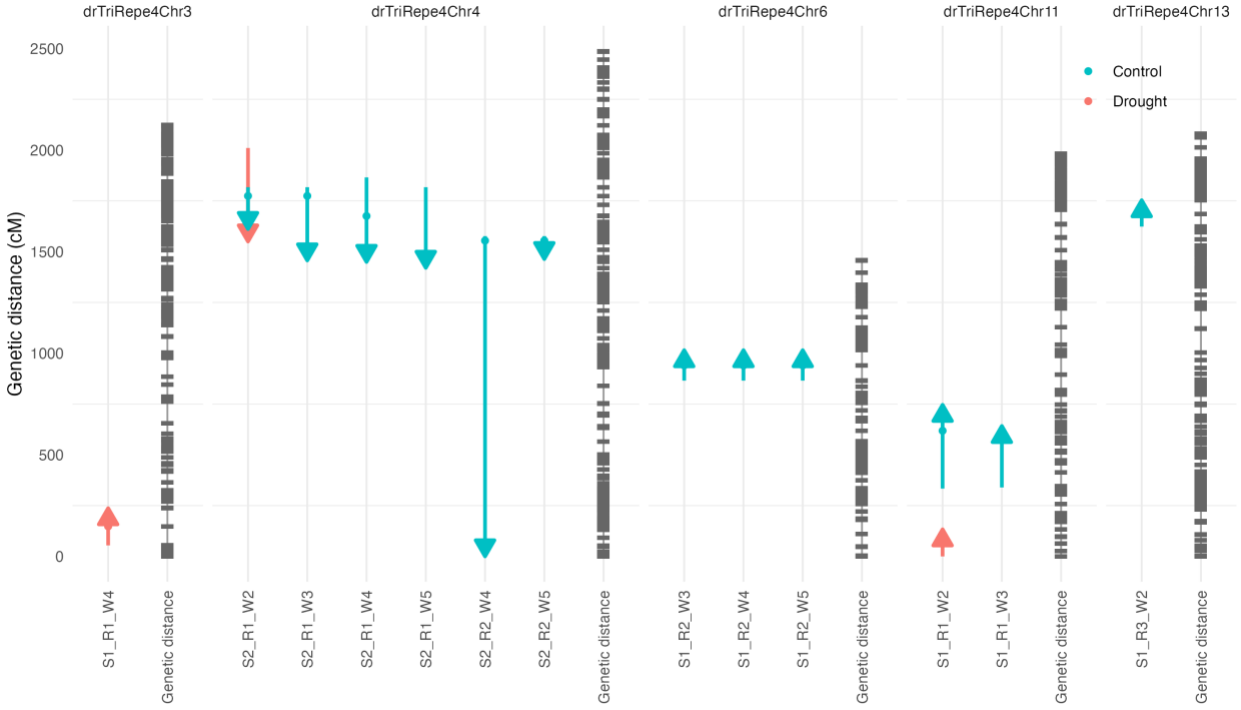
Supplementary Figure S5.5 Relative leaf area between different cyanotypes.

The relative leaf area is the difference between the leaf areas of the first week and the sixth week plants. The facet stripe shows the information of the genotype block (S), the replication (R), and the date of the experiment. Each of the two genotype blocks contained 150 non-overlapping genotypes. The three replications were nested in each genotype block. The date is the first week of the experiment. The linear mixed model: (Inflorescence counts) ~ 1 + Cyanotype + Treatment + Cyanotype: Treatment + (1 | Genotype block) + (1 | Genotype block: Replication). Type II ANOVA with Satterthwaite's method was used to test the significances of the fixed effect variables. Cyanotype, $F_{(3,1787)} = 0.3351$, $p = 0.8$. Treatment, $F_{(1,1787)} = 3904.6273$, $p < 2e\text{-}16$. Cyanotype: Treatment, $F_{(3, 1787)} = 0.3411$, $p = 0.7956$.



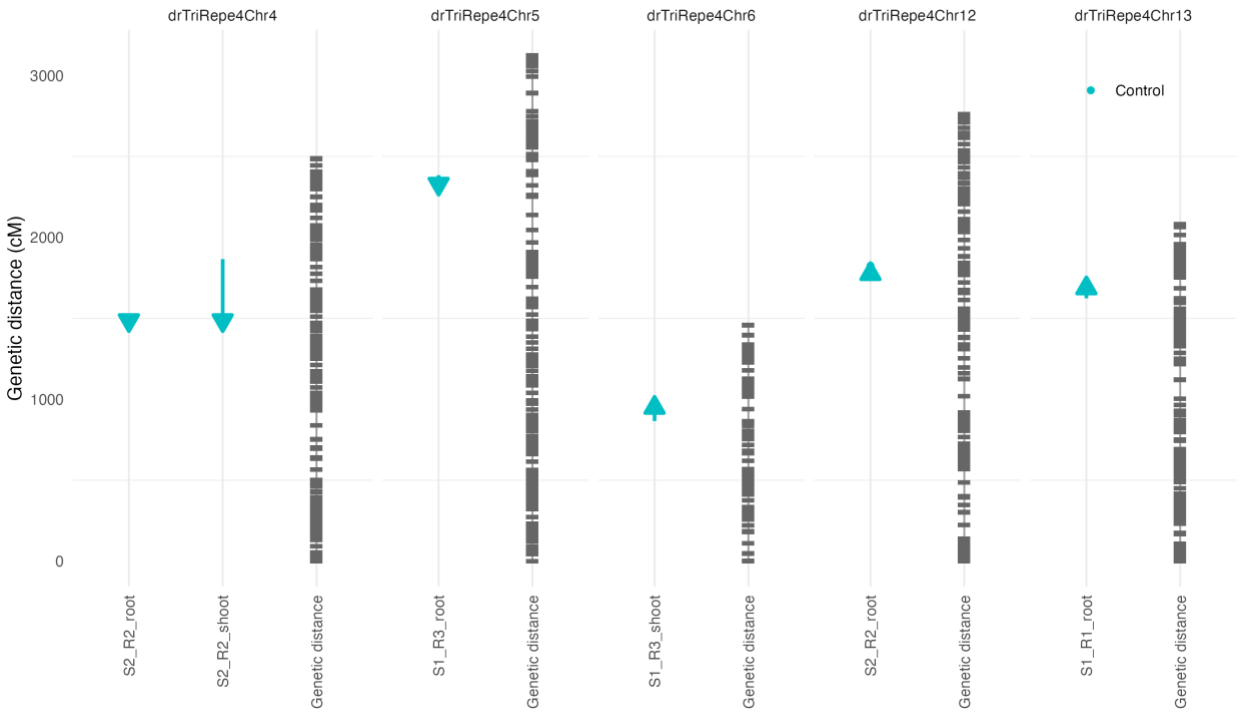
Supplementary Figure S5.6 Pearson correlation coefficients (*r*) between the different phenotypes within each treatment.

The calculation of the *r* is based on the BLUPs within each treatment. The significance of the r value is tested by Fisher's *Z* test for being different from zero. The *p* values below 0.05 (*), 0.01 (**), and 0.001 (***) are labeled in the grids. The top-left triangle is the drought treatment. The bottom-right is the control treatment.



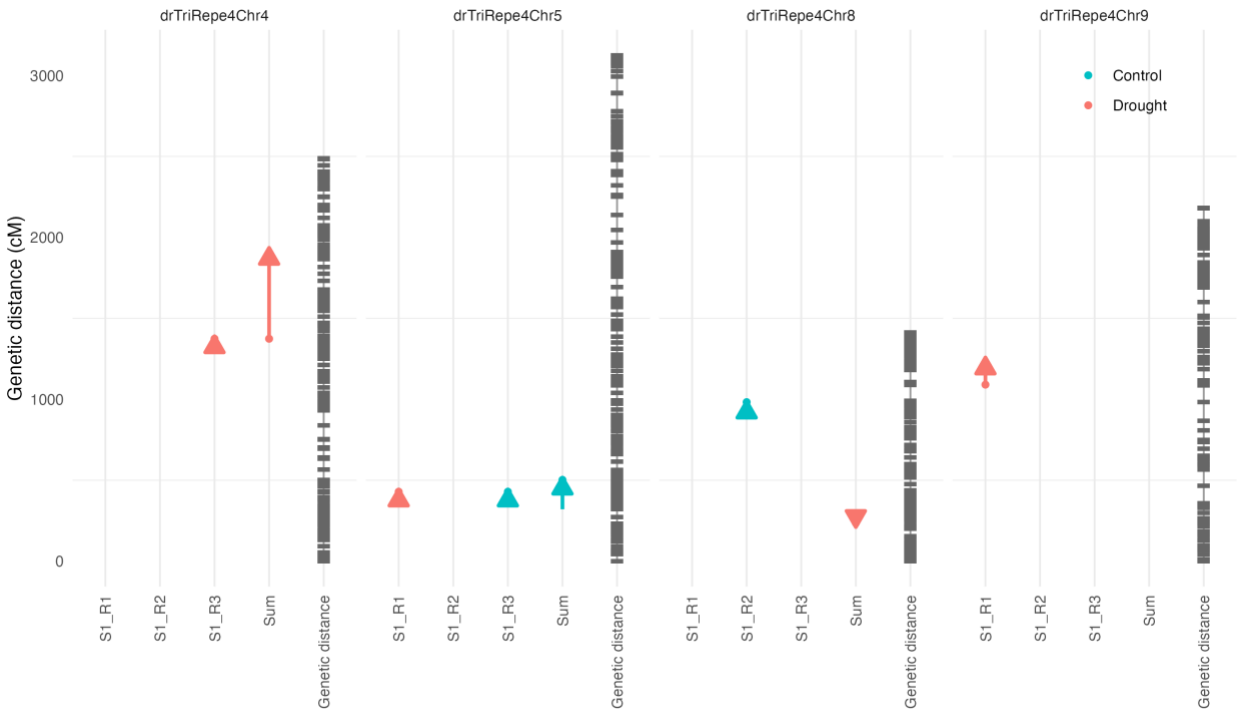
Supplementary Figure S5.7 Raw leaf area QTLs in the control and drought treatments.

The x-axis labels are composed of genotype block (S), the replication (R), and the week (W) of the experiment. The QTLs are shown if they are statistically significant after 1,000 times permutation test (*p* value < 0.1). The dots indicate the centers of the QTLs. The segments indicate the 1-drop LOD ranges. The arrowheads indicate the additive effects of the DMN010 allele ("B") compared to the GFL007 allele ("A").



Supplementary Figure S5.8 Raw dry weight QTLs in the control and drought treatments.

The x-axis labels are composed of genotype block (S), the replication (R) of the experiment. The QTLs are shown if they are statistically significant after 1,000 times permutation test (*p* value < 0.1). The dots indicate the centers of the QTLs. The segments indicate the 1-drop LOD ranges. The arrowheads indicate the additive effects of the DMN010 allele ("B") compared to the GFL007 allele ("A").



Supplementary Figure S5.9 Raw flower count QTLs in the control and drought treatments.

The x-axis labels are composed of genotype block (S), the replication (R) of the experiment. The QTLs are shown if they are statistically significant after 1,000 times permutation test (*p* value < 0.1). The dots indicate the centers of the QTLs. The segments indicate the 1-drop LOD ranges. The arrowheads indicate the additive effects of the DMN010 allele ("B") compared to the GFL007 allele ("A").