Local adaptation and microevolution along environmental gradients

Keric Slattery Lamb
B.A., University of Colorado – Boulder

A dissertation presented to the graduate faculty of the University of Virginia in Candidacy for the degree of Doctor of Philosophy

Department of Biology

University of Virginia
July 2024

ABSTRACT

Species ranges are frequently constrained by local adaptation near their boundaries. Yet, range expansions typically occur along continuous environmental gradients, where conditions change gradually without abrupt transitions. This general observation raises a fundamental question: how do limitations to local adaptation and range expansion arise when abrupt transitions are absent? In this dissertation, I used the wildflower *Campanula americana* to explore how environmental gradients of differing steepness interact with microevolutionary factors, like genetic drift and natural selection, to influence local adaptation and range expansion. I investigated these patterns using natural populations of *C. americana* found along a steep environmental gradient, driven by elevation in the Appalachian Mountains, and along a shallow environmental gradient, driven by latitude, in the North American Midwest. In Chapter 1, I tested patterns of local adaptation by planting common gardens and modeling adaptive genetic differentiation over the steep elevational and shallow latitudinal gradient. I found that local adaptation is limited along both gradients, though patterns of fitness differ among gradients. Along the steep gradient, fitness in range-edge populations is poor, while fitness is generally high among populations along the shallow gradient. In Chapter 2, I explored patterns of phenotypic and genetic selection along gradients. I found that differences in the strength and direction of selection on reproductive phenology were strong along the steep gradient. Additionally, adaptive alleles shifted in frequency more slowly along the shallower latitudinal gradient than along the steep elevational gradient. In Chapter 3, I evaluated genetic load and found it tended to increase toward range limits. Furthermore, along a shallow environmental gradient, fitness costs associated with genetic load were increased in stressful environments. In Chapter 4, I investigated patterns of gene flow and drivers of genetic differentiation along each gradient. I found asymmetric gene flow along the steep environmental

gradient and strong signatures of isolation-by-distance. Along the shallow environmental gradient, gene flow among populations was symmetric and genetic differentiation among populations was not associated with either geographic (i.e., drift) or environment distance (i.e., selection). Along both gradients, effective population size declined toward range limits. Together, these results paint a picture of ecological specialization and local adaptation constrained by genetic drift and gene flow along steep environmental gradients; and ecological generalization along a shallow environmental gradient, where further range expansion is likely constrained by interactions of genetic drift and environmental stress. Finally, in Chapter 5, I assessed how dynamics of postglacial range expansion influence outcomes of contact between intraspecific lineages. I found that all contact zones arose under a model of lineage divergence in parapatry, but gene flow was significantly greater in a southern contact zone near the species' rear edge. At the northern and mid-latitude contact zones, gene flow was minimal and strongly asymmetric. These results suggest that historic range dynamics can strongly influence outcomes of contact across a species range. Together, my dissertation presents a cohesive narrative demonstrating how the rate of environmental change along gradients influences patterns of genetic drift and natural selection. In turn, these patterns shape interactions among populations, influencing local adaptation, range expansion, and speciation potential.

ACKNOWLEDGEMENTS

It takes a village to raise a dissertation, and this one is no different. First, I would like to thank my advisor, Laura Galloway. She provided patient guidance on the art and execution of science, and her Herculean efforts to tame my writing are present throughout this dissertation. She has been especially patient while I explored new methods and questions that I never would've imagined proposing three years ago. I would also like to thank my committee of Butch Brodie, Alan Bergland, Mandy Gibson, and Max Castorani. Over the years, they have provided excellent feedback and helped me stay on track throughout my dissertation while I chased down loose ends and followed tangents down long rabbit holes. I would also like to thank Drew Schield who became an honorary 6th member of my committee and has provided many insightful comments about honing methodologies and interpreting results.

Many members of the Galloway Lab have left indelible signatures on my development as a scientist, as well as my dissertation. In particular, I would like to thank Hanna Makowski, Antoine Perrier, Alfredo Lopez, Joaquin Nunez, and Catherine Debban. Hanna welcomed me into the Galloway Lab in a year that would rapidly devolve into the COVID-19 pandemic. During the summer that followed the outbreak of COVID, we spent a couple months at MLBS driving around on back roads, building a geodesic boat, and occasionally looking through binoculars at flowers that were only 5ft away. When all of our plants were taken in the fateful Rapture of 2022, it was at least some consolation that, despite needing to restart the experiments that the plants were needed for, we were in the same boat together among friends. The expertise and camaraderie of Hanna, Antoine, and Alfredo provided troves of experience and thought-provoking conversation to mine for the development of my chapters. Antoine especially provided an excellent sounding board for my occasionally batty ideas. Antoine also helped prod along

genomic efforts by performing many DNA extractions, coordinating with the sequencer, and working on the assembly and archiving of genetic data. Joaquin came to Galloway Lab meetings and helped me learn many fundamentals of HPC programming and how to think like a bioinformatician. Catherine Debban graciously let me work on her data, expand and reframe her analyses, and guide one of her chapters to publication.

I would also like to thank all the people who supported me in non-academic ways throughout my dissertation, both kith and kin alike. My family, and especially my parents, Michelle Slattery and Ted Lamb, provided excellent guidance and support throughout. Thank you to my extended family, especially Ethan and Jess Slattery, Jeff and Julie Slattery, and Jim Crain, for helping me find seeds and letting me crash into their lives for several chaotic days at a time during field seasons. Thank you to Anna Jaunarajs for the constant support, companionship, and the reminder that life is not on pause during a dissertation but happening whether you do something with it or not. I'd also like to thank my roommates Thaddeus Weigel and Matt Van Houten, who helped keep me sane and connected during the pandemic with 'roommate dinners', as well as the legion of people I climbed with over the years—climbing isn't half as fun without wonderful people to spend the time with. Thanks especially to Mark Horton, Corey Williams, Chris Robinson, Dawson Payne, Johann Miller, and many others.

Finally, I would like to thank the armada of undergraduates, universities, state and national park personnel, and land managers who made this work possible. The undergraduates Yimfoong Ho, Sydney Cox, Emily Scott, Austin Kim, and Mason Gower-Fici helped make my dissertation possible by performing crosses, collecting fruits, managing plants, and working on experiments. Common gardens would not have happened without the aid of many iNaturalist denizens collecting seeds, as well as numerous land managers, research scientists, and natural

areas. Thanks especially to Andrew Renfro and the Highlands-Cashier Land Trust, Dave Carr and Blandy Experimental Farm, Brooks Saville and Virginia Tech's Kentland Farm, Jonathon Schramm and Goshen College's Merry Lea Environmental Learning Center, John Taylor and Ball State University's Cooper Farm, Wes Braker and St. Olaf's College, Vince Eckhart and Grinnell College's Conard Environmental Research Area, Matthew Dykstra and the Pierce Cedar Creek Institute, Gerald Zuercher and the University of Dubuque's Wolter Woods, Eric Nagy, Butch Brodie, and Mountain Lake Biological Station, the University of Minnesota's Cedar Creek Ecosystem Science Reserve, Grayson Highlands State Park, Monongahela National Forest, Sky Meadows State Park, Mount Jefferson State Natural Area, Hungry Mother State Park, New River State Park, and Shenandoah National Park. Lastly, thank you to my funding, through Laura Galloway, the Department of Biology, and the NSF EXPAND grant.

TABLE OF CONTENTS

INTRODUCTION

Range limits often reflect limits to the species' fundamental niche, particularly where environments change rapidly in space (Hargreaves et al. 2014). Where range limits coincide with niche limits, range expansion requires adaptation to novel habitat beyond the current range. Yet, species ranges often exist along continuums of environmental conditions, i.e., environmental gradients, and climate variables which predict species range limits do not necessarily exhibit discrete cutoffs near range limits. Because of this, theoretical models and empirical studies have frequently queried whether continuous environmental variation can produce static range limits, and how the steepness of the environmental gradients can affect this dynamic. While numerous studies have explored patterns of local adaptation along environmental gradients of differing steepness (Hargreaves et al. 2014), few empirical tests exist that robustly assess how environmental gradients interact with the microevolutionary and genomic landscapes to produce different patterns of adaptation along gradients of varying steepness. In my dissertation, I aimed to evaluate how the steepness of environmental gradients and microevolutionary pressures, namely selection, gene flow, and genetic drift, interact to constrain or promote adaptation near range limits. I explored these relationships using two extremes of environmental gradients—a shallow environmental gradient following latitude, and a steep environmental gradient following elevation in the Appalachian Mountains. In this introduction, I provide brief background and methodological context to the overarching research program, including discussion of my research aims and their significance. Finally, results across chapters are synthesized and discussed together.

BACKGROUND

Local adaptation is less frequent than commonly assumed (Leimu and Fischer 2008), raising questions about what factors may facilitate or constrain its development. Theoretical models posit that adaptation may break down in populations along steep environmental gradients, where environmental conditions change rapidly over space, and near range edges because of interpopulation connectivity and the relative strength of selection and genetic drift (Polechová and Barton 2015; Polechová 2018). At least one large meta-analysis has found that local adaptation tends to constrain range expansion and fitness along steep elevational environmental gradients, with local adaptation breaking down near range limits, while dispersal ability tends to constrain range expansion along shallower latitudinal environmental gradients (Hargreaves et al. 2014). the strength and direction of selection between populations increases, genetic variance in locally adapted loci is reduced (Christiansen 1974; Polechová and Barton 2015), enabling adaptive differentiation among populations along steep environmental gradients, even when well-connected. Empirical studies have demonstrated that genetic diversity and effective population size can also be sharply reduced by strong selection (Falk et al. 2012; Oakley 2013). These effects are often compounded by losses in genetic diversity associated with serial founder effects during range expansion (Slatkin and Excoffier 2012; Willi et al. 2018; Willi and Van Buskirk 2019; Perrier et al. 2022). In turn, these reductions are expected to degrade adaptive potential and expose populations colonizing range limits along steep environmental gradients near range limits to strong genetic drift.

In contrast to range expansion along steep environmental gradients, expansion along shallow gradients can occur rapidly, resulting in strong signatures of genetic drift in the direction of range expansion (Willi et al. 2018; Koski et al. 2019b). Serial founder effects associated with

rapid range expansion can lead to the gradual fixation and accumulation of unconditionally deleterious genetic load (Excoffier et al. 2009; Slatkin and Excoffier 2012; Willi 2013). Greater genetic load associated with range expansion can restrict population growth rates and reduce population fitness (Willi et al. 2018), impeding local response to selection and potentially curtailing further expansion. Importantly, environmental conditions can modify fitness consequences of genetic load (Parsons 1971; Perrier et al. 2020). If genetic load reduces fitness in stressful environments, then it may limit expansion into ecologically marginal habitat. Yet, most empirical tests that have found significant relationships between the expression of load and environment have tested only singular environmental variables, like temperature stress (Parsons 1959). Studies explicitly testing *sensu lato* environmental conditions have found much weaker relationships (Perrier et al. 2022), calling into question whether differences in the expression of load dependent on environmental conditions is relevant for populations.

Gene flow along the gradient can ameliorate signatures of genetic load and potentially increase local adaptation in ecologically marginal populations. In theoretical work, modest levels of gene flow among populations has been shown to enhance local adaptation (Alleaume-Benharira et al. 2006). Similar theoretical models further indicate that differences in the strength and direction of selection among populations are often weak along shallow gradients, reducing costs associated with dispersal and gene flow along the gradient (Polechová and Barton 2015). Consequently, local adaptation may be impeded along shallow environmental gradients where selection is too weak to filter maladaptive alleles imported via gene flow (Spichtig and Kawecki 2004; Tusso et al. 2021). Under these conditions, the evolution of generalist phenotypes may be favored, allowing rapid range expansion constrained by dispersal. However, empirical tests of these predictions are largely lacking (but see: Tusso et al. 2021).

When range expansion along environmental gradients is seeded by multiple refugia or ancestral lineages, populations near range limits may come into contact with other closely related but genetically differentiated lineages. Where enough genetic change has occurred to create reproductive isolation, contact between lineages may produce fitness costs associated with hybridization, reinforcing nascent divergence and driving speciation (Butlin and Smadja 2018). However, differences in the accumulation of reproductive isolation among lineages and geographic regions may produce variable outcomes of contact (Cutter 2015; Mandeville et al. 2015), especially when lineages have complex histories of paleoclimatic range dynamics (Hewitt 2011). For example, during postglacial range expansion, populations near leading range edges often experience substantial genetic drift induced via serial founder effects (Slatkin and Excoffier 2012; Willi et al. 2018; Koski et al. 2019b). Theory suggests that genetic drift associated with founder effects can positively interact with selection during range expansion to enhance adaptive differentiation and facilitate speciation (Templeton 2008). On the other hand, populations near relictual rear edges may harbor extensive genetic diversity among populations due to increased isolation and habitat fragmentation (Hampe and Petit 2005). Thus, the evolutionary history of lineages may play an important role in determining whether contact between divergent lineages degrades or reinforces genetic differentiation.

STUDY SYSTEM

Campanula americana is a monocarpic tetraploid wildflower native to North America. *C. americana* is generally outcrossing and insect-pollinated (Galloway et al. 2003; Koski et al. 2019a). In fall adult plants senesce, fruits dehisce, and seeds germinate into small rosettes. Rosettes overwinter, then bolt in spring. Where present, *C. americana* is often locally abundant, forming

patchy population distributions (unpublished field observations). Previous work has identified three primary genetic lineages (Barnard-Kubow et al. 2015), though most of the work discussed here is limited to the two larger lineages: the Appalachian lineage and the Western lineage.

The Appalachian and Western lineages are genetically distinct, having both nuclear-nuclear and cytonuclear reproductive barriers inhibiting introgression (Barnard-Kubow et al. 2015, 2017; Debban 2019). During the last glacial maximum, the Appalachian lineage likely resided in multiple microrefugia scattered throughout the Appalachian Mountains (Barnard-Kubow et al. 2015). After the last glacial maximum, the Appalachian lineage expanded toward contemporary elevational range limits within the Appalachian Mountains along a steep elevational environmental gradient (Fig. 1; 500m – 1400m).

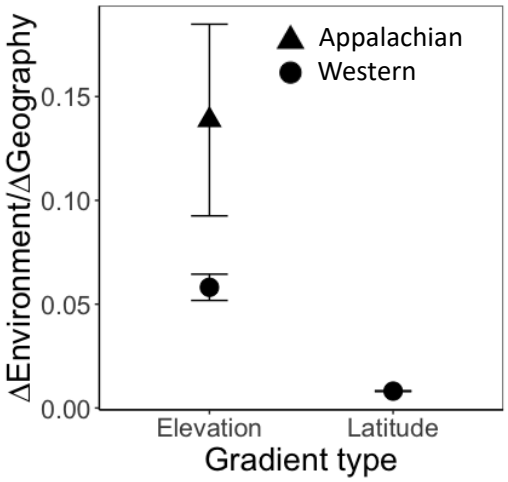


Figure (1) Environmental gradients along elevations and latitudes by lineage (adapted from Chapter 1). The steepness of environmental gradients is defined as the difference in environmental conditions, measured in Euclidean PCA space, divided by the geographic distance among populations.

The Western lineage likely weathered the last glacial maximum in both southerly refugia near the Gulf of Mexico (Barnard-Kubow et al. 2015), and in a mid-latitude refugium near Kentucky (Koski et al. 2019b). The Western lineage expanded toward northern latitudinal range limits from the mid-latitude refugium (Prior et al. 2020), spanning a wide breadth of latitude along a shallow latitudinal environmental gradient (Fig. 1; 38°N-45°N). The Western lineage also expanded along an elevational gradient in the Great Smokey Mountains from relictual southern habitat (Fig. 1; 200m – 1400m; Barnard-Kubow et al. 2015). The Western lineage has accumulated

significant genetic load during its latitudinal range expansion toward the Upper Midwest (Koski et al. 2019b). In contrast to the Western latitudinal range, little is known about patterns of expansion and adaptation along the elevational gradient in the Appalachian lineage, though reproductive phenology is known to differ across elevations (Haggerty and Galloway 2011).

AIMS AND SIGNIFICANCE

My dissertation work deepens our understanding of how the dynamics between microevolutionary processes and environments dictate patterns of adaptation during range expansions. Specifically, my dissertation addresses the questions: (1) what patterns of local adaptation are present along environmental gradients of differing steepness? (2) what patterns of selection drive local adaptation across these environmental gradients? (3) how is the accumulation of genetic load along environmental gradients influenced by range expansion and what effects does it have on fitness? (4) how does gene flow affect dynamics of selection and drift along environmental gradients? And finally, (5) how have range dynamics shaped intraspecific divergence among lineages? I address these questions using populations distributed along two extremes of environmental gradients—a shallow environmental gradient along latitude, and a steep environmental gradient along elevation in the Appalachian Mountains. These gradients are concurrent with the direction of postglacial range expansion in each region, allowing direct comparisons between microevolutionary pressures and environment, and how they influence patterns of local adaptation and range expansion.

Chapters

In Chapter 1, I evaluated patterns of local adaptation in *C. americana* along a steep elevational environmental gradient and a shallower latitudinal environmental gradient at the level of the phenotype, the genotype, and fitness. At the level of the phenotype, I found frost resistance was countergradient with environmental variance along the latitudinal gradient, while cold tolerance was co-gradient. Neither differed across elevations. At the level of the genotype, adaptive genotypic differentiation was found along the latitudinal gradient and plateaued toward range limits, but adaptive differentiation was not associated with elevational environmental gradients. Finally, data from 13 common gardens in replicate transects found minimal local adaptation with generally high fitness across the latitudinal gradient, except for range-edge populations planted in range edge common gardens. High elevation range-edge populations performed better at their home sites but were less fit than range-core populations across the entire elevational gradient, suggesting modest local adaptation exists across the elevational gradient.

Across all tests, I found local adaptation was infrequent along both environmental gradients, while differences in fitness among populations were more common along the steep elevational gradient. Along the shallow latitudinal environmental gradient, the absence of local adaptation, yet strong overall population fitness suggests the evolution of generalist phenotypes which perform well across environmental conditions. Generalist strategies may be favored along shallow environmental gradients if selection against gene flow is weak and the costs of dispersal along the environmental gradient are low (Spichtig and Kawecki 2004; Tusso et al. 2021). Interestingly, range-edge populations had equal fitness to populations near the range core, suggesting that poor fitness in range-edge common gardens reflects conditional fitness associated with other factors, like genetic drift. Along the steep elevational environmental gradients, weak

patterns of home vs. away local adaptation suggest that specialist adaptive strategies may be favored. However, poor fitness of populations near elevational range limits suggests that range-edge populations have likely experienced strong genetic drift. Together, these findings suggest that adaptive potential is limited near range edges, regardless of the steepness of the environmental gradient. Further, these findings demonstrate how interactions of range expansion and gene flow can affect patterns of adaptation (i.e., specialist or generalist) depending on the steepness of environmental gradients.

In Chapter 2, I explored selection underlying local adaptation to climate. First, I explored phenotypic selection on reproductive phenology and functional traits along a steep environmental gradient. I found that selection on reproductive phenology differed in strength and direction among populations from differing elevations, favoring earlier flowering in low elevation populations and later flowering in high elevation populations. Strength and direction of selection on functional traits did not differ among elevations. Next, I explored genomic patterns of selection. I found that loci underlying climate adaptation were driven to fixation along steep environmental gradients, while intermediate frequencies of adaptive alleles were more frequent along shallow environmental gradients. Fixation of adaptive loci along steep gradients suggests strong selection for climate adaptation and against gene flow from other populations along the gradient. These patterns suggest that strong differential selection exists along the steep elevational gradient, though only weak patterns of local adaptation were found in common garden tests (Chapter 1). Along the shallow latitudinal gradient, weaker selection for adaptive genetic differentiation accords with findings of minimal local adaptation.

In Chapter 3, I investigated patterns of genetic load across latitudinal and elevational gradients near range limits, and whether the fitness effects of genetic load depend on

environmental conditions. To do this, I first generated estimates of heterosis using hybrids between populations, I then estimated environmental displacement and site suitability to determine the habitat suitability of common garden locations. I found that lifetime genetic load, measured via heterosis, was marginally associated with range position along the shallow latitudinal gradient and non-significant along the steeper elevational gradient. Along the shallow latitudinal environmental gradient, fitness costs of genetic load were greater in stressful environments than in less stressful environments. These data suggest that genetic load may impede fitness of populations expanding into novel, ecologically marginal habitat, potentially halting range expansion.

In Chapter 4, I evaluated potential mechanisms of genetic differentiation along environmental gradients. In particular, I explored microevolutionary drivers of genetic differentiation along each gradient and analyzed how effective population size and the directionality of migration change with proximity to range limits. To investigate these questions, I performed isolation-by-distance and isolation-by-environment tests using genetic data, then modeled evolutionary history of populations along each gradient. Along both environmental gradients, genetic differentiation showed weak isolation-by-distance. Along the shallow latitudinal gradient, effective population size did not differ among populations, nor was migration significantly asymmetric. In contrast, effective population size was smaller near range limits along the steep environmental gradient, and migration was consistently asymmetric, with more migration coming from the range core toward the range edge than vice versa. Such patterns accord with strong signatures of genetic load along the gradient (Chapter 3) and suggest source-sink dynamics near range limits.

In Chapter 5, I analyzed patterns of demographic and range expansion history in the contexts of lineage divergence in *Campanula americana*. To do so, I used genomic SNPs to

evaluate population structure, simulate demographic history of divergence and migration under different models of between-lineage contact, and estimated contemporary patterns of gene flow. Contemporary gene flow between lineages is high, and historically migration was symmetric in a southern rear-edge contact zone. In a northern leading-edge contact zone and a separate mid-range contact zone in Virginia, contemporary gene flow is low, and historic patterns of migration are strongly asymmetric. At these contact zones, migration asymmetries matched the direction of a known cytonuclear incompatibility (Barnard-Kubow et al. 2016), which favors introgression from Western and Eastern populations into Appalachian, but not vice versa. These patterns suggest that range expansions may help accelerate the evolution and fixation of barrier loci underlying reproductive isolation, such that reinforcement may be much more likely to occur near leading range edges than trailing range edges.

SYNTHESIS & IMPLICATIONS

Together, my dissertation research yields original insights into adaptation along environmental gradients and contributes empirical perspective and grounding to an extensive body of theory. Along steep environmental gradients, asymmetric gene flow and strong divergent selection is degrading fitness in populations near elevational range limits (Chapters 1, 2, 4). In turn, effective population size is reduced within range-edge populations and genetic load increases (Chapters 3, 4). Reductions in effective population size can impede further adaptive genetic differentiation, hampering additional range expansion and reducing the adaptive potential of range-edge populations to weather environmental disturbance or displacement. This accords with predictions from theoretical models of limited local adaptation along steep environmental gradients (Polechová and Barton 2015; Gilbert et al. 2017; Polechová 2018; Bridle et al. 2019), and suggests

that differences in effective population size may explain variation in empirical support of maladaptive gene flow weakening local adaptation along steep environmental gradients (Bridle et al. 2009; Halbritter et al. 2015; Sexton et al. 2016; Zhang et al. 2019; Bachmann et al. 2020; Kottler et al. 2021).

Adaptation also appears to be limited along the shallower latitudinal gradient due to gene flow (Chapter 1). Instead, generalist adaptation may be favored over local adaptation due to low fitness costs associated with dispersal along the gradient (Chapter 1). Rather, accumulation of genetic load near range limits results in reduced fitness associated with environmental stress (Chapter 3). This finding was unique to the latitudinal gradient, suggesting that fitness consequences of genetic load may be integral to understanding when and where range expansion and local adaptation are likely to occur along shallow environmental gradients.

Together, these results provide a cohesive theory of range expansion along the latitudinal and elevational gradients of *Campanula americana*. More generally, these results also provide empirical backing for theoretical predictions of maladaptive gene flow and genetic load constraining local adaptation along environmental gradients, while helping to explain why findings of local adaptation have been less common than expected in empirical studies (Leimu and Fischer 2008).

LITERATURE CITED

Alleaume-Benharira, M., I. R. Pen, and O. Ronce. 2006. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. J. Evol. Biol. 19:203–215.

Bachmann, J. C., A. Jansen Van Rensburg, M. Cortazar-Chinarro, A. Laurila, and J. Van Buskirk. 2020. Gene Flow Limits Adaptation along Steep Environmental Gradients. Am. Nat. 195:E67–E86.

Barnard-Kubow, K. B., C. L. Debban, and L. F. Galloway. 2015. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am. J. Bot. 102:1842–1853.

Barnard-Kubow, K. B., M. A. McCoy, and L. F. Galloway. 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol. 213:1466–1476.

Barnard-Kubow, K. B., N. So, and L. F. Galloway. 2016. Cytonuclear incompatibility contributes to the early stages of speciation. Evolution 70:2752–2766.

Bridle, J. R., S. Gavaz, and W. J. Kennington. 2009. Testing limits to adaptation along altitudinal gradients in rainforest *Drosophila*. Proc. R. Soc. B Biol. Sci. 276:1507–1515.

Bridle, J. R., M. Kawata, and R. K. Butlin. 2019. Local adaptation stops where ecological gradients steepen or are interrupted. Evol. Appl. 12:1449–1462.

Butlin, R. K., and C. M. Smadja. 2018. Coupling, Reinforcement, and Speciation. Am. Nat. 191:155–172.

Christiansen, F. B. 1974. Sufficient Conditions for Protected Polymorphism in a Subdivided Population. Am. Nat. 108:157–166.

Cutter, A. D. 2015. Repeatability, ephemerality and inconvenient truths in the speciation process. Mol. Ecol. 24:1643–1644.

Debban, C. L. 2019. Reproductive isolation and gene flow vary among contact zones between incipient species.

Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic Consequences of Range Expansions. Annu. Rev. Ecol. Evol. Syst. 40:481–501.

Falk, J. J., C. E. Parent, D. Agashe, and D. I. Bolnick. 2012. Drift and selection entwined: asymmetric reproductive isolation in an experimental niche shift. Evol. Ecol. Res. 14:403–423.

Galloway, L. F., J. R. Etterson, and J. L. Hamrick. 2003. Outcrossing rate and inbreeding depression in the herbaceous autotetraploid, Campanula americana. Heredity 90:308–315.

Gilbert, K. J., N. P. Sharp, A. L. Angert, G. L. Conte, J. A. Draghi, F. Guillaume, A. L. Hargreaves, R. Matthey-Doret, and M. C. Whitlock. 2017. Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. Am. Nat. 189:368–380.

Haggerty, B. P., and L. F. Galloway. 2011. Response of individual components of reproductive phenology to growing season length in a monocarpic herb. J. Ecol. 99:242–253.

Halbritter, A. H., R. Billeter, P. J. Edwards, and J. M. Alexander. 2015. Local adaptation at range edges: comparing elevation and latitudinal gradients. J. Evol. Biol. 28:1849–1860.

Hampe, A., and R. J. Petit. 2005. Conserving biodiversity under climate change: the rear edge matters. Ecol. Lett. 8:461–467.

Hargreaves, A. L., K. E. Samis, and C. G. Eckert. 2014. Are Species' Range Limits Simply Niche Limits Writ Large? A Review of Transplant Experiments beyond the Range. Am. Nat. 183:157–173.

Hewitt, G. M. 2011. Quaternary phylogeography: the roots of hybrid zones. Genetica 139:617–638.

Koski, M. H., L. F. Galloway, and J. W. Busch. 2019a. Pollen limitation and autonomous selfing ability interact to shape variation in outcrossing rate across a species range. Am. J. Bot. 106:1240–1247.

Koski, M. H., N. C. Layman, C. J. Prior, J. W. Busch, and L. F. Galloway. 2019b. Selfing ability and drift load evolve with range expansion. Evol. Lett. 3:500–512.

Kottler, E. J., E. E. Dickman, J. P. Sexton, N. C. Emery, and S. J. Franks. 2021. Draining the Swamping Hypothesis: Little Evidence that Gene Flow Reduces Fitness at Range Edges. Trends Ecol. Evol. 36:533–544.

Leimu, R., and M. Fischer. 2008. A Meta-Analysis of Local Adaptation in Plants. PLoS ONE 3:e4010.

Mandeville, E. G., T. L. Parchman, D. B. McDonald, and C. A. Buerkle. 2015. Highly variable reproductive isolation among pairs of *Catostomus* species. Mol. Ecol. 24:1856–1872.

Oakley, C. G. 2013. Small effective size limits performance in a novel environment. Evol. Appl. 6:823–831.

Parsons, P. A. 1971. Extreme-environment heterosis and genetic loads. Heredity 26:479–483.

Parsons, P. A. 1959. Genotypic-environmental interactions for various temperatures in Drosophila melanogaster. Genetics 44:1325–1333.

Perrier, A., D. Sánchez-Castro, and Y. Willi. 2022. Environment dependence of the expression of mutational load and species' range limits. J. Evol. Biol. 35:731–741.

Perrier, A., D. Sánchez-Castro, and Y. Willi. 2020. Expressed mutational load increases toward the edge of a species' geographic range. Evolution 74:1711–1723.

Polechová, J. 2018. Is the sky the limit? On the expansion threshold of a species' range. PLOS Biol. 16:e2005372.

Polechová, J., and N. H. Barton. 2015. Limits to adaptation along environmental gradients. Proc. Natl. Acad. Sci. 112:6401–6406.

Prior, C. J., N. C. Layman, M. H. Koski, L. F. Galloway, and J. W. Busch. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. Mol. Ecol. 29:4473–4486.

Sexton, J. P., M. B. Hufford, A. C.Bateman, D. B. Lowry, H. Meimberg, S. Y. Strauss, and K. J. Rice. 2016. Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. Mol. Ecol. 25:911–928.

Slatkin, M., and L. Excoffier. 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics 191:171–181.

Spichtig, M., and T. J. Kawecki. 2004. The Maintenance (or Not) of Polygenic Variation by Soft Selection in Heterogeneous Environments. Am. Nat. 164:70–84.

Templeton, A. R. 2008. The reality and importance of founder speciation in evolution. BioEssays 30:470–479.

Tusso, S., B. P. S. Nieuwenhuis, B. Weissensteiner, S. Immler, and J. B. W. Wolf. 2021. Experimental evolution of adaptive divergence under varying degrees of gene flow. Nat. Ecol. Evol. 5:338–349.

Willi, Y. 2013. Mutational meltdown in selfing Arabidopsis lyrata. Evolution 67:806–815.

Willi, Y., M. Fracassetti, S. Zoller, and J. Van Buskirk. 2018. Accumulation of Mutational Load at the Edges of a Species Range. Mol. Biol. Evol. 35:781–791.

Willi, Y., and J. Van Buskirk. 2019. A Practical Guide to the Study of Distribution Limits. Am. Nat. 193:773–785.

Zhang, M., H. Suren, and J. A. Holliday. 2019. Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa. Genome Biol. Evol. 11:2256–2272.

CHAPTER 1:

Local adaptation limits fitness along steeper elevational gradients more strongly than shallower latitudinal ones.

ABSTRACT

How patterns of adaptation along environmental gradients relate to range expansion and whether they depend on the steepness of the gradient remains under studied. I characterize adaptation at the level of phenotype, genotype, and fitness along a steep elevational environmental gradient and a shallow latitudinal environmental gradient. I found no evidence of adaptive phenotypic differentiation along the steep elevational gradient. Populations from elevational range limits were generally unfit compared to range core populations across environments though they performed better at home garden sites than away, suggesting modest local adaptation. Adaptive genetic differentiation among populations was present, though varied strongly among transects. In contrast, along the shallow latitudinal environmental gradient, there was countergradient variation associated with frost resistance and modest adaptive genetic differentiation that leveled off near range edges. Common garden tests indicated that populations along the latitudinal gradient were not strongly locally adapted, and their fitness did not differ except near range limits, suggesting ecological generalization. My results suggest that adaptation along the steep elevational gradient is limited by maladaptive gene flow and genetic drift. Whereas along the latitudinal gradient, fitness losses near range limits were environment-dependent and absent when plants were grown in more southern, less stressful environments. Together, my results suggest that evolutionary potential near range limits may be strongly limited by postglacial range expansion and dynamics of genetic drift. As climate change pushes ideal environmental conditions northward, populations near contemporary range limits may be increasingly taxed, curtailing range expansion near leading range edges.

INTRODUCTION

Species ranges are frequently constrained by ecological niche limits near range edges (Hargreaves et al. 2014), requiring local adaptation to novel environmental conditions before further range expansion can occur. Yet, range expansions frequently occur along environmental gradients, where environments change gradually in space. These do not impose strict environmental boundaries at range limits. Whether adaptation is constrained near range edges along environmental gradients likely depends on the rate of environmental change in space (Hargreaves et al. 2014; Polechová and Barton 2015; Gilbert et al. 2017; Bachmann et al. 2020). Where environmental gradients are steepest, theory predicts that the fitness cost of dispersal to novel habitat is greater (Polechová and Barton 2015). Increased costs of dispersal and fitness tradeoffs associated with local adaptation may indicate greater potential for maladaptive gene flow among populations along the gradient (Gilbert et al. 2017; Bachmann et al. 2020). Consequently, adaptation and fitness may be weakened by connectivity among populations along steep environmental gradients, like elevation. I examined these predictions by investigating post-glacial range expansion along a latitudinal and elevational gradient in the wildflower, *Campanula americana.* I determined how environmental gradients of differing steepness influence the evolution of local adaptation at multiple levels of biological organization (phenotypes, genotypes, and population fitness), and evaluated the degree of adaptive genetic differentiation along each gradient.

Theoretical models posit that local adaptation can be constrained by the steepness of the environmental gradient (Polechová and Barton 2015; Gilbert et al. 2017; Bridle et al. 2019). Yet, experimental evidence of local adaptation impeded by steep environmental gradients is mixed. Work in *Plantago lanceolata* found stronger signals of local adaptation at high elevation sites than high latitude sites (Halbritter et al. 2015), while work in *Populus trichocarpa* found strong

evidence of climate adaptation near elevational and latitudinal range limits (Zhang et al. 2019). In contrast, a large meta-analysis of transplant experiments found that latitudinal range limits are more constrained by dispersal limitations, while elevational range limits are more constrained by local adaptation (Hargreaves et al. 2014). Exploring how adaptation varies along and across environmental gradients can provide insight into how limitations of adaptation develop near range limits.

Local adaptation along environmental gradients can be impeded by maladaptive gene flow among adjacent populations (Bachmann et al. 2020). While the likelihood of maladaptive gene flow disrupting local adaptation is disputed (Kottler et al. 2021), reductions in fitness associated with maladaptive gene flow are found more frequently along steep environmental gradients (Bridle et al. 2009; Bachmann et al. 2020). Thus, while gene flow can reduce genetic differentiation among any set of neighboring populations, fitness consequences of gene flow depend on the steepness of the gradient (Polechová and Barton 2015). For example, a recent experimental evolution study predicted that strong gene flow and small differences in the strength and direction of selection among populations along shallow environmental gradients can promote global adaptation, where fitness is high across the breadth of the range without ecological specialization (Spichtig and Kawecki 2004; Tusso et al. 2021); while strong selection and weak gene flow along steep environmental gradients, where costs of dispersal along the gradient are highest, may promote local adaptation (Alleaume-Benharira et al. 2006; Polechová and Barton 2015). When gene flow and selection are both strong, local adaptation can erode (Christiansen 1974; Kawecki and Ebert 2004), allowing source-sink dynamics, where local adaptation collapses and population growth rates fall below replacement, to emerge (Ronce and Kirkpatrick 2001; Alleaume-Benharira et al. 2006; Furrer and Pasinelli 2016).

To explore whether distinct patterns of adaptative differentiation are present along environmental gradients that differ in steepness, I addressed the questions: (1) What patterns of adaptation are present along environmental gradients that differ in steepness? (2) How sensitive are contemporary patterns of adaptation along different environmental gradients to gene flow? Operationally, I examined physiological traits linked to adaptation, *in situ* local adaptation and fitness, and adaptive genetic differentiation near the latitudinal and elevational range limits of *Campanula americana*, a native wildflower. First, I explored physiological traits linked to overwintering survival. Physiological traits that facilitate adaptation to cold stress have been shown to exhibit clinal variation (Zhen and Ungerer 2008), and can provide insight into how patterns of fitness across a species range are translated from, and are produced by, variance in phenotypes. I analyzed phenotypes associated with cold stress by experimentally assessing chlorophyll content and frost resistance during significant cold stress and relate patterns to individual survival and population ecology. Second, I assessed the magnitude of local adaptation near elevational and latitudinal range limits using common gardens. Finally, I explored genomic components of climate adaptation to better understand how adaptive genetic differentiation emerges across gradients and how sensitive such patterns may be to gene flow.

METHODS

Study system

Campanula americana is a monocarpic tetraploid herb native to the eastern United States. *C. americana* is primarily outcrossing and insect pollinated (Galloway et al. 2003; Koski et al. 2019a). Where present, individuals are locally common, producing patchy population distributions across its native range. *C. americana* overwinters prior to flowering. Seeds in many populations germinate in fall, overwinter as small rosettes, then flower and fruit during summer.

C. americana is generally divided into three genetic lineages (Barnard-Kubow et al. 2015), the largest of which are the Western and Appalachian lineages. The third, the Eastern lineage, is derived from and closely related to the Western lineage (Barnard-Kubow et al. 2015), so is folded into the Western lineage for analyses here. The Western lineage is predominantly located in eastern North America, spanning from the Gulf of Mexico toward the Great Lakes. After the last glacial maximum, the lineage expanded northwesterly from a mid-latitude glacial refugium (Barnard-Kubow et al. 2015; Koski et al. 2019b) through serial expansion events (Prior et al. 2020). In the Southeast, the Western lineage expanded toward elevational range limits in the southern Appalachian Mountains. In contrast, the Appalachian lineage likely resided through the last glacial maximum in multiple microrefugia located proximally to, and distributed throughout, the Appalachian Mountains (Barnard-Kubow et al. 2015).

The Western lineage crosses a broad swath of habitat along a latitudinal gradient (28°N - 46°N; Fig. 1), and a steep elevational gradient near the Great Smoky Mountains (234m - 1098m). Local adaptation is found along the latitudinal gradient (Naciuk 2015), though at large spatial scales that cross phylogeographic clusters (Perrier, *unpublished data*). The Appalachian lineage is located within the Appalachian Mountains between North Carolina and Pennsylvania and traverses a steep elevational gradient (143m - 1616m; Fig. 1). In the Appalachian lineage, some home vs. away local adaptation has been observed in high elevation, range-edge populations in southwestern Virginia. However, range-edge populations did not generally have higher fitness than foreign, range core populations in common gardens near range limits (Haggerty and Galloway 2011). Across analyses, populations distributed along latitudinal and elevational gradients are treated separately. For genomic analyses, I analyzed populations separately by lineage (*Western/Appalachian*) and gradient (*elevation/latitude*). I use environmental change over the

latitudinal gradient in the Western lineage and the elevational gradient in the Appalachian and Western lineages to explore how differences in the rate of change of environmental conditions influence patterns of local adaptation (Fig. 1C; SI Fig. 1).

Population selection and greenhouse crosses

I selected populations of *C. americana* from the steep elevational and shallow latitudinal environmental gradients located near their respective range limits (Fig. 1A). I selected populations along both gradients to encompass environmental conditions from near the range core to near range limits. For the latitudinal gradient, I categorized populations as range core, mid-range, and range edge. I defined range core sites as those that reside closest to the Western lineage's latitudinal median, and range edge populations and gardens as those that reside closest to the current range edge. Mid-range sites are those between the range core and range edge. For elevational gradients, I categorized sites and populations as range core, i.e., low elevation, or range edge, i.e., high elevation. For both gradients, populations were nested within transects for sequence analysis and common gardens (Fig. 1A). Each transect contained one population from each range position (core/mid/edge), thereby replicating environmental gradients. In total, I selected 10 populations from the northern latitudinal gradient nested within three transects, and 13 populations from the elevational environmental gradient in the Appalachian Mountains (SI Table 1). For sequence analysis of elevational populations, I used five transects of core-edge pairs of populations. For common garden analysis of elevational populations, I used three transects of core-edge pairs of populations. Specific populations used varied by experiment (Fig. 1A).

I collected seeds from native populations between 2020 and 2022. Experiments either used seeds from within-population crosses performed in the University of Virginia greenhouses (2022)

to control for maternal environment (here, greenhouse seeds), or field-collected seeds, where availability of greenhouse seeds was limited (see: SI Table 2). I mapped range limits using the maximum extent of research-grade observations from iNaturalist as of May 2022 (n=7,781; iNaturalist Community n.d.) and a minimum concave polygon, calculating in R (4.2.0) using the rangemap package (*https:://github.com/marlonecobos/rangemap*).

1. Phenotypes of cold adaptation

I explored how adaptations to acute cold stress vary across environmental gradients using two key physiological traits. First, I assessed leaf chlorophyll content during vernalization and its response to an acute cold stress. Vernalization exposes plants to cold temperatures to simulate winter, providing a phenological cue for many plants from temperate regions (Chouard 1960). In other plant species, chlorophyll content is strongly correlated with photosynthetic rate (Buttery and Buzzell 1977), and cold stress is known to decrease both leaf chlorophyll content (Koç et al. 2010; Talebzadeh and Valeo 2022) and photosynthetic activity in photosystem II (Liang et al. 2007). Next, I assessed how the freezing temperature of leaf tissue differed between populations during and after vernalization. To perform these experiments, I selected 15 populations (Fig. 1; SI Table 1): 10 populations from the latitudinal gradient and five from the elevational gradient. I used greenhouse produced seeds.

To describe environmental conditions across the latitudinal and elevational gradients, I performed a PCA. I computed the PCA using 4km resolution, daily climate data from PRISM for the years of 2017-2022 (PRISM Climate Group, Oregon State University n.d.). I included all available environmental variables: precipitation, maximum and minimum temperature, mean temperature, mean dew point temperature, and maximum and minimum vapor pressure. I extracted

the first principal component (PC1; 62.3% variance explained) as an index of environmental conditions across gradients for all experiments related to cold stress.

Cold tolerance (Chlorophyll content and acute cold stress survival)

I measured chlorophyll content during an acute cold exposure to examine physiological components of cold tolerance. In brief, I measured chlorophyll content via chlorophyll fluorimetry (Krause and Weis 1991) using a handheld fluorometer (Minolta SPAD-502). I used 15 unique maternal families per population where available, else I used repeats of maternal families. I germinated seeds in growth chambers for 6 weeks following standard protocols (SI Table 2) and moved the small rosettes to a cold room at 4°C for 6 weeks prior to experimentation to allow plants to acclimate to the cold. I divided plants into 15 blocks, with 13 populations per block and 26 plants divided into equal and identical sets that were either exposed to cold ('cold exposure') or not ('no cold exposure'; 13 individuals/treatment/block). In total, I used 390 plants, with 195 in the 'no cold exposure' group and 195 in the 'cold exposure' group. For tests of chlorophyll content, I removed blocks from the cold room (4°C) and placed them into a dark environmental chamber at -20°C for 12h, imitating an extreme overnight cold stress that wild plants could experience while overwintering as rosettes. I insulated plants in Styrofoam to slow cooling. I measured chlorophyll content before cold exposure. Once the cold treatment had concluded, I removed blocks from the cold and immediately measured chlorophyll content again. I measured the control group that was not exposed to the acute cold stress in like fashion at the same time. I evaluated survival as a binary trait two weeks after the acute cold stress.

Frost resistance

Frost resistance enables plants to avoid cellular damage associated with freezing temperatures by reducing their freezing point to temperatures below 0°C. To investigate frost resistance, I collected leaf tissue from plants and subjected it to extreme cold (-20° C). I measured leaf freezing temperature, also called the supercooling temperature (Reyes-Diaz et al. 2006), by assessing the thermal inflection point of ice formation in leaf tissue. Because ice formation is thermogenic, a significant temperature spike is observable when solutes within a leaf arrive at their freezing temperature and ice formation begins (Reyes-Diaz et al. 2006).

I germinated seeds in growth chambers for 6 weeks following standard protocols (SI Table 2). In total, I used 132 plants from 13 populations (~10 individuals/population). I tested plants at two time points, once while vernalizing, i.e. winter like conditions, and once two weeks after vernalizing, when plants had been allowed to acclimate to greenhouse and initiated growth, i.e. spring like conditions. I randomly divided plants into sets of four to be tested at the same time. To measure the leaf freezing temperature, I attached leaf tissue to a type K thermocouple probe, then insulated the tissue inside a 1.5mL Eppendorf tube encased in Styrofoam. I placed the rig into an environmental chamber at approximately -20° C. For tests conducted during vernalization, I harvested leaf tissue from plants held at 4°C temperatures for 6 weeks. For post-vernalization tests, I harvested leaf tissue from the same plants 2 weeks after the vernalization period had concluded and plants had been allowed to acclimate to greenhouse conditions (23° C daytime/19° C nighttime ±5° C).

Statistical analysis

I modeled how patterns of chlorophyll content vary during an cold stress across environmental gradients using Gaussian linear models. I examined elevational and latitudinal gradients independently from each other. Briefly, I regressed the population average chlorophyll content per treatment against a binary categorical predictor of whether plants were exposed to the acute cold stress, and a continuous predictor of climate (*chlorophyll content = PC1 + cold stress treatment + PC1 * cold stress treatment*). I computed models using R 4.2.0 (R Core Team 2022) and calculated type three ANOVAs to determine significance of model terms using the car package (Fox and Weisberg 2019). I modeled variation in frost resistance across environmental gradients in similar fashion, using linear models and population averages. For overwintering (vernalizing) and spring (acclimating) experiments, I performed analyses separately but in identical fashion—I computed univariate linear models for each gradient by regressing mean population leaf freezing temperature against continuous predictors of climate PC1 (*freezing temperature = PC1*). Residuals were visually assessed for conformity to model assumptions. Across models, residuals were normally distributed and homoscedastic.

I also assessed relationships between environment, and survival through the cold tolerance experiment. Because plants from elevational populations did not survive cold tolerance tests, I assessed these relationships only among populations along the latitudinal gradient. For environment-survival models, I regressed the population's average proportion survival against a continuous predictor climate PC1 using beta regressions (*survival = PC1*). I determined the significance of model terms using type III ANOVAs. Residuals followed model assumptions.

2. Fitness across environmental gradients

Common gardens

I planted common gardens to explore spatial patterns of local adaptation and population fitness across and within environmental gradients. In total, I established eight common gardens along the shallow latitudinal environmental gradient and four across the steeper elevational environmental gradient using 21 populations (latitude n=9; elevation n=6; Fig. 1; SI Tables 1, 3). I selected common garden sites near latitudinal and elevational range limits for each respective gradient. I categorized gardens, like populations, as range core, mid-range, or range edge. Populations and gardens were organized into replicate transects (Fig. 1). Each transect contained one population per range position (latitude: core/mid/edge; elevation: core/edge; Fig. 1A) and one garden per range position (latitude: core/mid/edge; elevation: core/edge; Fig. 1B). In only one case, transect 2 for latitudinal common gardens, did a transect contain an incomplete set of gardens (mid/edge; no core). Thus, for latitudinal populations there were three replicate populations for each range position (core/mid/edge), and two replicate core gardens, three replicate mid-range gardens, and three replicate range edge gardens. For elevational populations, there were three replicate populations for each range position (core/edge), and two replicate gardens for each range position (core/edge). Across both gradients, I selected garden locations to be as close as possible to populations of the same transect and range position. Home garden for populations was assigned as the nearest garden to the population. While I assigned every garden a home population, not every population had a home garden site.

To detect local adaptation while allowing for the influence of maternal environment, I planted common gardens in May 2022 across the elevational gradient using field-collected seeds

(see: Supplemental Methods 1.1). Survival was measured in June and August. Because of generally poor flowering and survival, fitness related to flowering was not modeled for 2022 gardens.

To understand patterns of local adaptation and fitness across gradients that differ in steepness while controlling for maternal environment, I repeated the common garden experiment in 2023 across the latitudinal and elevational gradients (Fig. 1; SI Table 1, 3). For elevational gardens (n=4), I transplanted two individuals from each of the 6 populations into each of 10 blocks (n=120/garden; n=2 *replicates/population/block*), following identical methods and garden locations as in 2022 (Supplementary Methods 1.1). For latitudinal gardens (n=8), I transplanted two individuals from each of the nine populations into each of 10 blocks (n=180/garden; n=2 *replicates/population/block*) using the same methods (SI Table 3). I planted elevational gardens between 4/18-4/26 and latitudinal gardens between 5/1-5/9.

I collected bolting, survival, stem diameter, and bud/flower/fruit count data from gardens in late summer (Latitudinal: 8/21-8/25; Elevational: 9/4-9/8) prior to fruit dehiscence. On average, plants at elevational gardens spent 138 days in the field and plants at latitudinal gardens spent 110 days in the field. Stem diameter at the stem-soil interface provides a measure of plant size that allowed me to incorporate plants that lost fruit set due to herbivory (Pearson correlation of stem diameter and number of fruits and flowers for plants that did not experience herbivory: elevation $R=0.65$, latitude $R=0.75$). I estimated total reproduction by summing fruit, flower, and mature bud counts for each plant. To account for potential lost reproduction due to latitudinal populations spending less time in common gardens, I added immature bud counts to total fruit and flower counts as an estimate of total reproduction theoretically possible during the growing season ('total theoretic reproduction').

Statistical analysis

To assess how the magnitude of local adaptation differed between environmental gradients, I calculated Hedge's D effect sizes of fitness within and between gardens following Leimu & Fischer (2008). Briefly, for each garden, I divided the difference between mean fitness of the home population (the population closest to the garden site) and the mean fitness across all foreign populations divided by the standard deviation of the fitness for all plants at the garden. This metric assessed local vs. foreign performance within gardens. I also wanted to examine how populations perform across the gardens. To explore performance in home and away environments, I calculated Hedge's D in like fashion, assessing population fitness in common gardens near their home sites compared to garden sites away from their respective homes (i.e., mean fitness of the home population minus the mean fitness of the population when away from the home site, divided by the standard deviation of fitness of the population across all garden sites).

To further explore how patterns of adaptation were shaped by range position and the steepness of environmental gradients, I assessed the relationship between population range position and population fitness at common gardens. I first examined individual fitness in absolute terms. To do this, I calculated individual fitness as the product of survival (0|1) and reproduction (≥0). I then computed Gaussian linear mixed effect ANOVAs for each gradient independently. Models treated individual fitness as a function of categorical population range position (*core/mid/edge*) and garden range position (*core/mid/edge*). I square-root transformed individual fitness data to account for variance within the individual fitness metric that caused scalar issues in model fitting (elevation fitness range: 0-762 (σ=66.75); latitude fitness range: 0-138 (σ=22.37)). Blocks were nested within gardens and garden range position. Populations were nested within population range position (*fitness = population range position + garden range position +*

population range position * *garden range position* + *population*(*population range position*) + *block*(*garden*(*garden range position*)). I also relativized metrics of reproduction by garden to control for differences in mean performance across sites and to differentiate between patterns of divergent selection and suitability of the garden site. To relativize fitness, I divided individual fitness by the mean fitness within the common garden. I analyzed metrics of relative reproduction (i.e., relative total reproduction (*fruits* + *flowers*), relative theoretical reproduction (*fruits* + *flowers* + *immature buds*), and relative stem diameter) using identical structure as absolute Gaussian models described above. Residuals from all models were visually assessed as both normal and homoscedastic. Mixed effect models were computed using the lme4 package (Bates et al. 2014)

In common gardens along elevational gradients, many high elevation, range edge population plants failed to bolt at both high and low elevation garden sites. To account for failure to bolt, I employed zero-truncated and binomial survival models. Briefly, I split the analysis into three additional generalized models: binary survival, which assessed whether plants survived or not; binary reproduction, which assessed whether plants that survived also reproduced; and reproduction ≥ 1, which assessed total reproduction of plants that flowered or fruited. Generalized model syntax followed identical structure as relative fitness models. For survival and binary reproduction models, I employed a generalized binomial model with a logit link function. For reproduction ≥ 1, I used a generalized Poisson model with a log link function. I analyzed data from 2022 and 2023 elevational common gardens separately but in identical fashion with one exception: absolute individual fitness. For 2022 gardens, square-root transformation caused convergence issues in mixed models, so I employed a Poisson GLM with a log-link function on the raw values instead. Across generalized models, residuals were homoscedastic.

3. Genomic signals of adaptative differentiation

Methodological framework

Response to divergent patterns of selection over environmental gradients is expected to produce clines of adaptive genetic differentiation coincident with the gradient. I use an F_{ST}-based approach to describe patterns of adaptive genetic differentiation, i.e. differentiation that is associated with bioclimatic factors across latitudinal and elevational gradients. In brief, I first RAD-sequenced 197 individuals from the 22 sampled populations (~9 *individuals/population*) and identified loci associated with environmental adaptation across each lineage and environmental gradient using multiple genome-scanning approaches (*PCAdapt, BayeScEnv*). After, I computed F_{ST} of adaptive loci for each lineage and environmental gradient. I then compared differences in genetic turnover of adaptive differentiation between populations near the range core (*core*) and populations closer to the range edge (*mid/edge*) along transects (genetic offset). Similar metrics have been employed to explore which populations face greater extinction risk under climate change (Capblancq et al. 2020; Van Daele et al. 2022). Here, I used genetic offset to explore how sensitive patterns of adaptive genetic differentiation are likely to be to gene flow from the range core to elsewhere in the range.

Sequencing and genetic data preparation

I assessed genomic patterns of adaptation and differentiation across the range using SNPs identified via RAD-Sequencing from 22 sampled populations. Seeds were collected from natural populations and germinated according to protocols outlined in SI Table 2 I collected fresh leaf tissue from seedlings and sent samples to Floragenex for RAD library prep and Illumina sequencing using SbfI as the restriction enzyme. I assembled raw reads using a reference sequences

for *C. americana* and STACKS v2 (Rochette et al. 2019). Further details regarding extractions, sequencing, and assembly are outlined in Supplementary Methods 1.2. In total, I sequenced ~1.15% of the genome. After assembly, I filtered SNPs to retain only loci of high quality. Specifically, I removed loci that were not biallelic, had minimum allele frequencies of less than 0.05, and mean read depths of less than 10 or greater than 53.5 ($coverage + 4*\sqrt{coverage}$), following Li (2014), using Stacks and VCFTools (Danecek et al. 2011). I then applied a missing threshold of 80% to remove sites that were predominated by missing data. The final SNP set totaled 6,359 loci, with coverage averaging 31.2 reads per site. I computed a PCA of the final SNP set to check quality and evaluate whether patterns of phylogenetic and population structure were captured (SI Fig. 2). After, I partitioned loci to lineages and regions, and removed loci that became monomorphic through the subsetting process (SNP: Appalachian Elevation n=5,003; Western Elevation n=5,052; Western Latitude n=5,724).

Modeling adaptive differentiation

Environmental gradients may drive different clinal patterns of genetic differentiation for loci associated with climate adaptations. I performed two gene-by-environment association tests to identify loci under putative selection from environmental factors. First, I identified SNPs putatively associated with local adaptation in each lineage and region (Western latitude, Western elevation, Appalachian elevation) using PCAdapt (Privé et al. 2020). For each grouping of lineage and region, I visually assessed the number of principal components that spatially partitioned populations. In total, I retained eight PC dimensions for Western latitudinal populations, four for Western elevational populations, and six for Appalachian elevational populations. I then corrected

for false discovery by subsetting outlier loci to those with significant q-values ($q \leq 0.05$) following the Benjamini-Hochberg procedure.

PCAdapt identifies loci underlying adaptation through analysis of population structure. Because of this approach, the program may identify loci underlying adaptation not associated with environmental conditions, such as adaptations to soil type and biotic competition. To confirm locus-selection from PCAdapt and identify loci specifically related to climate adaptations across environmental gradients, I used BayeScEnv (De Villemereuil and Gaggiotti 2015). BayeScEnv requires information on the environmental distance of each population to an environmental centroid for the species. To calculate environmental distance, I computed a PCA using all 19 bioclimatic factors in WorldClim 2.0 for all iNaturalist records of research grade quality uploaded as of 05/13/2022 (n=7,781; Fick and Hijmans 2017; iNaturalist Community n.d.). I extracted bioclimatic factors for each observation at a resolution of 2.5 arcseconds. I retained the smallest set of principal components that explained at least 95% of the variance among bioclimatic factors (n=5). I then computed Euclidean distances between the PCA median centroid and each population. I ran BayeScEnv out-of-the-box with default parameters. I corrected significance values of loci identified by BayeScEnv for false discovery by applying a q-value FDR threshold of 0.05. It is important to note that because RAD-sequencing sub-samples regions of the genome (Baird et al. 2008), loci identified by genome scanning methods likely represent only a portion of loci under selection in the genome (Lowry et al. 2017).

To evaluate environmental factors associated with adaptive differentiation, I computed generalized dissimilarity models (GDMs). Briefly, GDMs associate a metric of population divergence, here F_{ST}, with population-level environmental conditions while controlling for variance associated with neutral divergence, such as isolation-by-distance (Fitzpatrick & Keller,

2015). I calculated genetic divergence (F_{ST}) for pools of climate-adapted loci among populations to identify patterns and signals of genomic adaptation along environmental gradients. First, I extracted loci identified as significant by both PCAdapt and BayeScEnv. Pools of loci identified by both methods did not overlap between lineages or environmental gradients. I calculated F_{ST} between populations using the subset of SNPs under selection for climate adaptation in the R package StAMPP (Pembleton et al. 2013). I then used matrices of genetic distance (F_{ST}) to inform generalized dissimilarity models (GDM; Fitzpatrick et al. 2022). I computed GDMs for each region and lineage subset separately. I extracted I-splines from models, analogous to partial regression coefficients, for geographic and significant bioclimatic factors. I applied a maximum ceiling of one to estimates of adaptive genetic turnover. I extracted predicted adaptive genetic turnover values for each population, then plotted estimates against population range position (core/mid/edge). Finally, I used significant bioclimatic factors identified in GDMs to predict range-scale patterns of genetic turnover by projecting model-predicted allelic turnover using rasterized WorldClim 2.0 bioclimatic data at a resolution of 2.5 arcminutes.

Gene flow into a population from a different climate may influence the population's ability to respond to selection, decreasing adaptive genetic differentiation along the gradient. I regressed genetic offset, or the difference in adaptive turnover between range positions within transects, against population range position (core/mid/edge) using linear mixed effects models, where transect (i.e., spatial replicate) was a random effect (*genetic offset* = *population range position* + *transect*(*population range position*)). Where genetic offsets are greatest, adaptations to environmental conditions are likely more susceptible to contemporary maladaptive gene flow from the range core. Where genetic offsets are small, adaptive differentiation between range positions is minimal.

RESULTS

1. Phenotypes of cold adaptation

Cold tolerance (chlorophyll content and acute cold stress survival)

Environmental variation was strongly correlated with latitude (R = 0.97, SI Fig. 1B) and elevation (R = 0.95, SI Fig. 1C) for their respective gradients (Supplementary Results 1.1). Plants that experienced an acute cold stress had lower chlorophyll content regardless of which environmental gradient plants hailed from (SI Fig. 3). The effect of cold stress on chlorophyll content did not depend on local climate across elevations or latitudes (interactions not significant). Two weeks after exposure to the acute cold event, no plants survived from populations along the elevational gradient. Along the latitudinal gradient, populations from colder climates of origin were more likely to have survived the acute cold stress (p-value=0.01; Fig. 2A).

Frost resistance

Contrary to expectations, populations from colder climates did not have lower leaf freezing temperatures. The climate at the site of a population's origin (i.e., PC1) did not affect the freezing temperature of leaf tissue during vernalization along either environmental gradient (Fig. 2B). After plants had been allowed to acclimate in the greenhouse for two weeks to simulate spring emergence, patterns differed between gradients. For elevational populations, freezing temperature was not related to a population's climate (PC1 p-value=0.81). For latitudinal populations, I found that plants from colder climates froze at warmer temperatures (PC1 p-value=0.003).

2. Fitness across environmental gradients

For local adaptation to be present across a range, local populations must outperform foreign populations across common gardens (Kawecki and Ebert 2004). Along the shallow latitudinal environmental gradient, local populations did not have higher fitness than foreign populations in common gardens. Absolute individual fitness did not vary among populations from different range positions within common gardens (Fig. 3B), and the estimated magnitude of local adaptation was close to zero across the latitudinal gradient (Fig. 3A). In models accounting for relative performance within common gardens, range core populations outperformed range edge populations in range edge common gardens (SI Fig. 4A), though this effect was significant only for models of relative total theoretical reproduction and relative stem diameter (SI Table 4).

Across the steep elevational environmental gradient, range core populations routinely outperformed range edge populations in both range core and range edge common gardens. I found this pattern across models of absolute fitness (Fig. 3B), relative fitness, survival, and binary reproduction (SI Fig. 4). In only one model, reproductive output ≥ 1, did I find local range-edge populations outperforming foreign range core populations in range edge common gardens (SI Fig. 4C; SI Table 4). Similarly, the magnitude of local vs. foreign local adaptation showed range core populations from elevational transects strongly outperformed range edge populations across common garden range positions (i.e., positive effect size; Fig. 3A). I found similar results in survival rates in common gardens planted in 2022 with field-collected seeds (SI Fig. 5), suggesting that maternal environment is not responsible for the reduced performance of range edge populations relative to range core populations. In comparisons between home vs. away performance, range core populations had greater success at home garden sites than away in terms of absolute measurements of fitness (Fig. 3), but range edge populations did not. In terms of

relative measurements of fitness (SI Fig. 4B), range core and range edge populations tended to perform better at home garden sites than away, suggesting some local adaptation exists within range edge populations.

3. Genomic signals of adaptative differentiation

I identified loci associated with signals of environmental selection using multiple genome-scanning approaches. Final pools of putatively adaptive loci that were identified by both BayesScEnv and PCAdapt totaled five loci for Appalachian elevational populations, nine for Western elevational populations, and five for Western latitudinal populations (details in Supplementary Results 1.2). Pools of loci did not overlap between lineages or gradients.

GDMs identified bioclimatic factors important in explaining patterns of adaptive genetic differentiation among populations. For Western-lineage latitudinal populations, precipitation seasonality explained the greatest proportion of genetic differentiation (Table 1, SI Fig. 6, 7), and was significantly associated with range position along the latitudinal gradient. For Appalachian-lineage elevational populations, the mean temperature of the wettest quarter was the most important bioclimatic factor underlying adaptive genetic differentiation, though range position was not associated with this adaptive differentiation (SI Fig. 8, 9). Finally, for Western lineage elevational populations, precipitation of the wettest month was the most important bioclimatic factor underlying adaptive genetic differentiation, but no bioclimatic factors were significantly explained by range position (SI Fig. 9, 10).

Genetic offsets, the difference in adaptive genetic turnover between the range core and other range positions (mid/edge), generally increased with distance from the range core along the latitudinal gradient, though offsets tended to level off between mid-range and range-edge

populations (Fig. 4). The absence of substantive differences in adaptive genetic turnover and genetic offsets between mid-range and range-edge populations suggests that climate adaptation near the latitudinal range edge is limited, and range edge populations resemble mid-latitude populations. For elevational populations, genetic offsets were distinct among transects, bioclimatic variables, and lineages (Fig. 5).

In several cases, range position did not explain variation in adaptive genetic turnover or genetic offset (e.g., mean temperature of the wettest quarter for Appalachian-lineage elevational populations). Such cases reflect bioclimatic factors associated with heterogeneous patterns of selection and local adaptation that are unassociated with postglacial range expansion and the direction of the environmental gradient. Alternatively, several bioclimatic factors differed significantly among range positions but did not describe substantive levels of genetic differentiation in GDMs (e.g., precipitation of the wettest month for Appalachian-lineage elevational populations). Such cases reflect instances of adaptive genetic differentiation along the environmental gradient but are generally less meaningful in describing overarching patterns of selection and climate variation driving genetic differentiation.

DISCUSSION

Theoretical models predict that patterns of adaptation should differ near range limits depending on how rapidly environments change over space (Polechová and Barton 2015; Gilbert et al. 2017; Bridle et al. 2019). Yet, experimental studies have found mixed support for such patterns (Bridle et al. 2009; Halbritter et al. 2015; Sexton et al. 2016; Zhang et al. 2019; Bachmann et al. 2020). Here, I examined patterns of adaptation at the level of phenotype, genotype, and population fitness. Across approaches, local adaptation was stronger in range core populations, regardless of the

environmental gradient, though the magnitude of local adaptation was greatest along the elevational gradient. Local adaptation along the steep environmental gradient is likely constrained by strong genetic drift and maladaptive gene flow. While along the shallow latitudinal gradient, fitness was comparable across range positions and local adaptation was generally absent, suggesting generalist adaptation in lieu of local adaptation. Yet, reductions in the relative fitness of range edge populations near latitudinal range limits suggests some limits to generalist adaptation. Broadly, my results find agreement among tests of local adaptation at multiple levels (phenotype/genotype/population fitness) and suggest that adaptation along steep environmental gradients may be biased toward specialization, while adaptation along shallow environmental gradients may be biased toward generalization.

Phenotypes of cold adaptation

Frost resistance and cold tolerance are important traits underlying survival in temperate overwintering plant species. I found that chlorophyll content was modulated in response to cold stress but did not vary with either environmental gradient (SI Fig. 3). In contrast, frost resistance did not vary along the steeper elevational gradient but was countergradient with it. Lack of a clinal response during vernalization indicates that frost resistance is homogeneous across environmental gradients during winter. Interestingly, survival through an acute cold stress (i.e., cold tolerance) was concurrent with the latitudinal gradient, suggesting that countergradient frost resistance during spring emergence is unrelated to survival through significant cold stress at higher latitudes. That no plants survived through the acute cold stress from the elevational gradient suggests that tests were too extreme for any potential adaptation to be detected.

Near latitudinal range limits, tradeoffs between growth and tolerance, or resistance to climate extremes, may be reinforced by shorter growing seasons. Tradeoffs between climate resilience (i.e., resistance or tolerance) and phenology near range limits have been well described in plant literature (Willi and Van Buskirk 2022). Thus, adaptations that enable cold resistance may be reduced in phenological windows associated with rapid growth, such as spring emergence, if mechanisms enabling climate resilience impede growth. Previous work in northern *Arabidopsis lyrata* populations found countergradient clines of frost tolerance (Wos and Willi 2015). Moreover, at lower latitudes, false spring events are more likely to occur, and the time from spring onset to first bloom is typically longer (Allstadt et al. 2015). Consequently, plants near lower latitudes may be more likely to retain expensive metabolic adaptations to frost resistance during spring bolting. In colder, northern climates, where false spring events are less likely to occur, adaptations to accommodate them may be less important.

Fitness across environmental gradients

Local adaptation, in the form of local advantage among range positions (Kawecki and Ebert 2004), was absent from both environmental gradients (Fig. 3). Along the latitudinal gradient, all populations had near identical relative fitness in the range core and mid-range common gardens (SI Fig. 4A), regardless of where populations came from. Such patterns indicate limited local adaptation along the latitudinal gradient. Yet, it is noteworthy that range edge populations were relatively unfit near the range edge, while being equally as fit as range core populations in common gardens near the range core. Namely, recent range expansion may deplete genetic diversity in range-edge populations through serial founder effects (Slatkin and Excoffier 2012), leaving populations less diverse and potentially less robust to novel or stressful environmental conditions

(Oakley 2013). Previous work in other species, namely *Plantago lanceolata* (Halbritter et al. 2015) and *Arabidopsis lyrata* (Sánchez-Castro et al. 2022), found that absences of local adaptation near latitudinal range limits are associated with signatures of reduced genetic diversity, suggesting that genetic load and reductions in genetic variation associated with serial colonization (i.e., genetic drift) may play an important role in determining where local adaptation occurs. The influence of range expansion on patterns of adaptation may be especially relevant in *C. americana*, which is known to have lower genetic diversity, smaller effective population size (Chapter 4), and higher rates of inbreeding near its latitudinal range limits (Koski et al. 2019b). This idea is explored in greater depth in Chapters 3 and 4.

Local adaptation can also be constrained along shallow environmental gradients if costs of dispersal along the gradient are low and selection against gene flow is relatively weak (Polechová and Barton 2015). In theoretical models and empirical studies, weak selection against introgression among environmentally differentiated populations has been found to generate intermediate, generalist phenotypes that perform well in a wide range of environmental conditions (Spichtig and Kawecki 2004; Tusso et al. 2021). Weak selection against introgression and ecological generalization may in part explain comparable performance of populations across latitudinal common gardens and the plateau of adaptive genetic differentiation noted in genetic offset models.

Along the steep elevational gradient, range core populations tended to outperform range edge populations regardless of the position of the common garden (Fig. 3B, SI Fig. 4). However, weak local adaptation along elevational gradients was present in the form of 'home vs. away' fitness (Kawecki and Ebert 2004), wherein populations have higher relative fitness at garden sites near their origin than away (SI Fig. 4). The presence of home vs. away local adaptation for relative fitness across the elevational gradient suggests a level of ecological specialization among

elevational populations. Yet, absence of local advantage near elevational range limits indicates that range edge populations are generally unfit compared to range core counterparts. Overperformance by a single population or cluster of populations is not unique to *C. americana* (Zovi et al. 2008; Ortegón-Campos et al. 2009; Vesakoski and Jormalainen 2013), and may be a common feature of adaptation in nature (Hereford 2009).

Overperformance of range core populations along the steep elevational gradient may emerge from evolutionary dynamics of range expansion. For example, populations near range limits are more likely to suffer from genetic load as a result of serial founder effects (Slatkin and Excoffier 2012; Peischl et al. 2013). Fitness in range edge populations may also be modulated by interactions of environmental conditions and a history of recent range expansions. For example, inbreeding and drift-associated genetic load may be more strongly expressed in ecologically marginal habitat (Chapter 3; Armbruster and Reed 2005; Galloway and Etterson 2007), and asymmetric gene flow from the range core resulting from recent range expansion can slow or halt local adaptation in marginal populations (Ronce and Kirkpatrick 2001; Bridle et al. 2009; Bachmann et al. 2020). Here, it is plausible that such factors compound to produce fitness losses in populations of *C. americana* near elevational range limits.

Genomic signals of adaptation and sensitivity to gene flow

Adaptive genetic differentiation declined toward range limits across both environmental gradients. Along the shallow latitudinal gradient, adaptive genetic differentiation and genetic offsets plateaued between mid-range and range-edge populations. Where genetic offsets between adjacent populations are high, gene flow among populations may disrupt patterns of local adaptation by introducing maladaptive genetic variance. Weak and plateauing adaptive genetic differentiation

along the latitudinal gradient suggests that populations near the range edge likely resemble mid-range populations. Yet, it is unclear whether this pattern emerges from recent colonization of the range edge, which can limit genetic differentiation among populations (Slatkin and Excoffier 2012), from gene flow impeding adaptive genetic differentiation near range limits, or some combination of these factors.

Along the steep elevational gradient, adaptive genetic differentiation among populations was often absent. Additionally, patterns of adaptive genetic differentiation present among populations were often independent of the elevational gradient, suggesting maladaptive gene flow is inhibiting differentiation. Gene flow among populations can block adaptive genetic differentiation among populations by importing maladaptive alleles that reduce population fitness and break down local adaptation in recipient populations (Bridle et al. 2009, 2019; Bachmann et al. 2020). While such dynamics are more readily applied to steep environmental gradients, adaptive differentiation may also be suppressed by gene flow along environmental gradients of any steepness. For example, weak selection against gene flow can homogenize populations and promote ecological generalization (Tusso et al. 2021). Along shallow environmental gradients, this process can eliminate adaptive differentiation among populations while maintaining fitness across the breadth of the range.

Summary and synthesis

Environmental gradients that differ in steepness drive distinct patterns of adaptation. Along the steep elevational environmental gradient, range core populations universally outperformed range edge populations, though modest specialization was present in the form of local adaptation to home environmental conditions. Additionally, patterns of adaptive genetic differentiation were

often idiosyncratic among transects and independent of the steep environmental gradient. Finally, I found no evidence of physiological trait divergence, measured through frost resistance and cold tolerance along the steep environmental gradient. These results generally accord with findings in other study systems where maladaptive gene flow is present (Bridle et al. 2009; Bachmann et al. 2020), and may indicate the presence of source-sink dynamics near elevational range limits.

Along the shallow latitudinal environmental gradient, adaptive genetic differentiation plateaued toward range limits and local adaptation was largely absent across the range. In phenotypic analyses, I found countergradient patterns of frost resistance and co-gradient patterns of increased survival associated with colder climate of origin. Co-gradient survival accorded with findings of adaptive genetic differentiation associated with the mean temperature of the coldest quarter. These results indicate that adaptation to climate and phenology exists along the latitudinal gradient that was not detected in common garden tests. Further, these results indicate that while local adaptation is limited along the shallow environmental gradient, fitness is broadly maintained, suggesting ecological generalization. Yet, range edge populations performed poorly in common gardens near latitudinal range limits, suggesting that range-edge populations have experienced more genetic drift than populations near the range core. These outcomes hint at conditional fitness of latitudinal populations with high genetic load associated with stressful environmental conditions.

Studies which explore patterns of adaptation across levels of organization (i.e., genotypes, phenotypes, and fitness), are few and far between. As a result, it is often unclear if patterns discovered at one level reflect patterns across others. Here, I explored local adaptation at the level of genes, phenotypes, and *in situ* fitness, and found synergy across tests. Broadly, my findings indicate that local adaptation is limited along both steep and shallow environmental

gradients, while population fitness is largely limited near expanding range edges. These findings are consistent with general patterns of local adaptation identified in a large meta-analysis of transplant experiments, which showed that population fitness in common gardens beyond contemporary range limits is typically lower along steep elevational gradients than shallow latitudinal gradients (Hargreaves et al. 2014). However, my study adds an important caveat to these findings—adaptive differentiation along environmental gradients of any steepness can be limited by postglacial range expansion and maladaptive gene flow. As a result, adaptive potential of range-edge populations is likely poor regardless of the steepness of the environmental gradient. Temperate species, which have likely undergone some form of postglacial range expansion, may be especially likely to exhibit these patterns (Qian and Ricklefs 2007; Hargreaves et al. 2014). As old hypotheses about biogeography are adapted into and replaced by modern phylogeographic frameworks (Pironon et al. 2017), it has become increasingly apparent that gene flow and range expansion must be accounted for (Normand et al. 2011).

Furthermore, my findings have important implications for understanding how genetic drift influences adaptation in range-edge populations, and how future climate change may influence range shifts in other species. Where populations near range edges exhibit limited adaptive potential, work in other systems has demonstrated that the colonization potential of these populations may be reduced during warming-induced range shifts (Hargreaves and Eckert 2019). My results emphasize this outcome and suggest that conditional fitness near range limits associated with climate may play an important role in determining the colonization success of populations displaced in the course of climate change.

LITERATURE CITED

Alleaume-Benharira, M., I. R. Pen, and O. Ronce. 2006. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. J. Evol. Biol. 19:203–215.

Allstadt, A. J., S. J. Vavrus, P. J. Heglund, A. M. Pidgeon, W. E. Thogmartin, and V. C. Radeloff. 2015. Spring plant phenology and false springs in the conterminous US during the 21st century. Environ. Res. Lett. 10:104008.

Armbruster, P., and D. H. Reed. 2005. Inbreeding depression in benign and stressful environments. Heredity 95:235–242.

Bachmann, J. C., A. Jansen Van Rensburg, M. Cortazar-Chinarro, A. Laurila, and J. Van Buskirk. 2020. Gene Flow Limits Adaptation along Steep Environmental Gradients. Am. Nat. 195:E67–E86.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE 3:e3376.

Barnard-Kubow, K. B., C. L. Debban, and L. F. Galloway. 2015. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am. J. Bot. 102:1842–1853.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2014. Fitting Linear Mixed-Effects Models using lme4. arXiv.

Bridle, J. R., S. Gavaz, and W. J. Kennington. 2009. Testing limits to adaptation along altitudinal gradients in rainforest *Drosophila*. Proc. R. Soc. B Biol. Sci. 276:1507–1515.

Bridle, J. R., M. Kawata, and R. K. Butlin. 2019. Local adaptation stops where ecological gradients steepen or are interrupted. Evol. Appl. 12:1449–1462.

Buttery, B. R., and R. I. Buzzell. 1977. The relationship betwen chlorophull content and rate of photosynthesis in soybeans. Can. J. Plant Sci. 57:1–5.

Capblancq, T., M. C. Fitzpatrick, R. A. Bay, M. Exposito-Alonso, and S. R. Keller. 2020. Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. Annu. Rev. Ecol. Evol. Syst. 51:245–269.

Chouard, P. 1960. Vernalization and its Relations to Dormancy. Annu. Rev. Plant Physiol. 11:191–238.

Christiansen, F. B. 1974. Sufficient Conditions for Protected Polymorphism in a Subdivided Population. Am. Nat. 108:157–166.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

De Villemereuil, P., and O. E. Gaggiotti. 2015. A new F_{ST} -based method to uncover local adaptation using environmental variables. Methods Ecol. Evol. 6:1248–1258.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302–4315.

Fitzpatrick, M. C., and S. R. Keller. 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol. Lett. 18:1–16.

Fitzpatrick, M., K. Mokany, D. Nieto-Lugilde, and S. Ferrier. 2022. gdm: Generalized Dissimilarity Modeling.

Fox, J., and S. Weisberg. 2019. car: An R Companion to Applied Regression. Sage, Thousand Oaks, CA.

Furrer, R. D., and G. Pasinelli. 2016. Empirical evidence for source–sink populations: a review on occurrence, assessments and implications. Biol. Rev. 91:782–795.

Galloway, L. F., and J. R. Etterson. 2007. Inbreeding depression in an autotetraploid herb: a three cohort field study. New Phytol. 173:383–392.

Galloway, L. F., J. R. Etterson, and J. L. Hamrick. 2003. Outcrossing rate and inbreeding depression in the herbaceous autotetraploid, Campanula americana. Heredity 90:308–315.

Gilbert, K. J., N. P. Sharp, A. L. Angert, G. L. Conte, J. A. Draghi, F. Guillaume, A. L. Hargreaves, R. Matthey-Doret, and M. C. Whitlock. 2017. Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. Am. Nat. 189:368–380.

Haggerty, B. P., and L. F. Galloway. 2011. Response of individual components of reproductive phenology to growing season length in a monocarpic herb. J. Ecol. 99:242–253.

Halbritter, A. H., R. Billeter, P. J. Edwards, and J. M. Alexander. 2015. Local adaptation at range edges: comparing elevation and latitudinal gradients. J. Evol. Biol. 28:1849–1860.

Hargreaves, A. L., and C. G. Eckert. 2019. Local adaptation primes cold-edge populations for range expansion but not warming-induced range shifts. Ecol. Lett. 22:78–88.

Hargreaves, A. L., K. E. Samis, and C. G. Eckert. 2014. Are Species' Range Limits Simply Niche Limits Writ Large? A Review of Transplant Experiments beyond the Range. Am. Nat. 183:157–173.

Hereford, J. 2009. A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. Am. Nat. 173.

iNaturalist Community. n.d. Observations of Campanulastrum americanum from North America, United States of America observed on/between 08/2011-05/2022.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225–1241.

Koç, E., C. Đşlek, and A. S. Üstün. 2010. Effect of Cold on Protein, Proline, Phenolic Compounds and Chlorophyll Content of Two Pepper (Capsicum annuum L.) Varieties.

Koski, M. H., L. F. Galloway, and J. W. Busch. 2019a. Pollen limitation and autonomous selfing ability interact to shape variation in outcrossing rate across a species range. Am. J. Bot. 106:1240–1247.

Koski, M. H., N. C. Layman, C. J. Prior, J. W. Busch, and L. F. Galloway. 2019b. Selfing ability and drift load evolve with range expansion. Evol. Lett. 3:500–512.

Kottler, E. J., E. E. Dickman, J. P. Sexton, N. C. Emery, and S. J. Franks. 2021. Draining the Swamping Hypothesis: Little Evidence that Gene Flow Reduces Fitness at Range Edges. Trends Ecol. Evol. 36:533–544.

Krause, G. H., and E. Weis. 1991. Chlorophyll Fluorescence and Photosynthesis: The Basics. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42:313–349.

Leimu, R., and M. Fischer. 2008. A Meta-Analysis of Local Adaptation in Plants. PLoS ONE 3:e4010.

Li, H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30:2843–2851.

Liang, Y., H. Chen, M. Tang, P. Yang, and S. Shen. 2007. Responses of *Jatropha curcas* seedlings to cold stress: photosynthesis-related proteins and chlorophyll fluorescence characteristics. Physiol. Plant. 131:508–517.

Lowry, D. B., S. Hoban, J. L. Kelley, K. E. Lotterhos, L. K. Reed, M. F. Antolin, and A. Storfer. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol. Ecol. Resour. 17:142–152.

Naciuk, M. 2015. Latitudinal Gradient Facilitates Adaptive Population Differentiation of Campanulastrum americanum (American Bellflower). Columbus State University.

Normand, S., R. E. Ricklefs, F. Skov, J. Bladt, O. Tackenberg, and J.-C. Svenning. 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. Proc. R. Soc. B Biol. Sci. 278:3644–3653.

Oakley, C. G. 2013. Small effective size limits performance in a novel environment. Evol. Appl. 6:823–831.

Ortegón-Campos, I., V. Parra-Tabla, L. Abdala-Roberts, and C. M. Herrera. 2009. Local adaptation of *Ruellia nudiflora* (Acanthaceae) to biotic counterparts: complex scenarios revealed when two herbivore guilds are considered. J. Evol. Biol. 22:2288–2297.

Peischl, S., I. Dupanloup, M. Kirkpatrick, and L. Excoffier. 2013. On the accumulation of deleterious mutations during range expansions. Mol. Ecol. 22:5972–5982.

Pembleton, L. W., N. O. I. Cogan, and J. W. Forster. 2013. St AMPP : an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. 13:946–952.

Pironon, S., G. Papuga, J. Villellas, A. L. Angert, M. B. García, and J. D. Thompson. 2017. Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. Biol. Rev. 92:1877–1909.

Polechová, J., and N. H. Barton. 2015. Limits to adaptation along environmental gradients. Proc. Natl. Acad. Sci. 112:6401–6406.

Prior, C. J., N. C. Layman, M. H. Koski, L. F. Galloway, and J. W. Busch. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. Mol. Ecol. 29:4473–4486.

PRISM Climate Group, Oregon State University. n.d.

Privé, F., K. Luu, B. J. Vilhjálmsson, and M. G. B. Blum. 2020. Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. Mol. Biol. Evol. 37:2153–2154.

Qian, H., and R. E. Ricklefs. 2007. A latitudinal gradient in large-scale beta diversity for vascular plants in North America. Ecol. Lett. 10:737–744.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reyes-Diaz, M., N. Ulloa, A. Zuniga-Feest, A. Gutierrez, M. Gidekel, M. Alberdi, L. J. Corcuera, and L. A. Bravo. 2006. Arabidopsis thaliana avoids freezing by supercooling. J. Exp. Bot. 57:3687–3696.

Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28:4737–4754.

Ronce, O., and M. Kirkpatrick. 2001. When sources become sinks: migrational meltdown in heterogeneous habitats. Evolution 55:1520–1531.

Sánchez-Castro, D., A. Perrier, and Y. Willi. 2022. Reduced climate adaptation at range edges in North American *Arabidopsis lyrata*. Glob. Ecol. Biogeogr. 31:1066–1077.

Sexton, J. P., M. B. Hufford, A. C.Bateman, D. B. Lowry, H. Meimberg, S. Y. Strauss, and K. J. Rice. 2016. Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. Mol. Ecol. 25:911–928.

Slatkin, M., and L. Excoffier. 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics 191:171–181.

Spichtig, M., and T. J. Kawecki. 2004. The Maintenance (or Not) of Polygenic Variation by Soft Selection in Heterogeneous Environments. Am. Nat. 164:70–84.

Talebzadeh, F., and C. Valeo. 2022. Evaluating the Effects of Environmental Stress on Leaf Chlorophyll Content as an Index for Tree Health. IOP Conf. Ser. Earth Environ. Sci. 1006:012007.

Tusso, S., B. P. S. Nieuwenhuis, B. Weissensteiner, S. Immler, and J. B. W. Wolf. 2021. Experimental evolution of adaptive divergence under varying degrees of gene flow. Nat. Ecol. Evol. 5:338–349.

Van Daele, F., O. Honnay, and H. De Kort. 2022. Genomic analyses point to a low evolutionary potential of prospective source populations for assisted migration in a forest herb. Evol. Appl. 15:1859–1874.

Vesakoski, O., and V. Jormalainen. 2013. Ignored patterns in studies of local adaptations: When the grass is greener on the allopatric site. Ideas Ecol. Evol. 6.

Willi, Y., and J. Van Buskirk. 2022. A review on trade-offs at the warm and cold ends of geographical distributions. Philos. Trans. R. Soc. B Biol. Sci. 377:20210022.

Wos, G., and Y. Willi. 2015. Temperature-Stress Resistance and Tolerance along a Latitudinal Cline in North American Arabidopsis lyrata. PLOS ONE 10:e0131808.

Zhang, M., H. Suren, and J. A. Holliday. 2019. Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa. Genome Biol. Evol. 11:2256–2272.

Zhen, Y., and M. C. Ungerer. 2008. Clinal variation in freezing tolerance among natural accessions of *Arabidopsis thaliana*. New Phytol. 177:419–427.

Zovi, D., M. Stastny, A. Battisti, and S. Larsson. 2008. Ecological Costs on Local Adaptation of an Insect Herbivore Imposed by Host Plants and Enemies. Ecology 89:1388–1398.

AUTHOR CONTRIBUTIONS

Keric Lamb conceived of experiments, constructed common gardens, performed all analyses, and wrote the manuscript. Yimfoong Ho helped conduct frost resistance and cold tolerance experiments. Antoine Perrier assembled raw genetic data. Laura Galloway assisted in the design of experiments, consulted on analyses, and edited the manuscript.

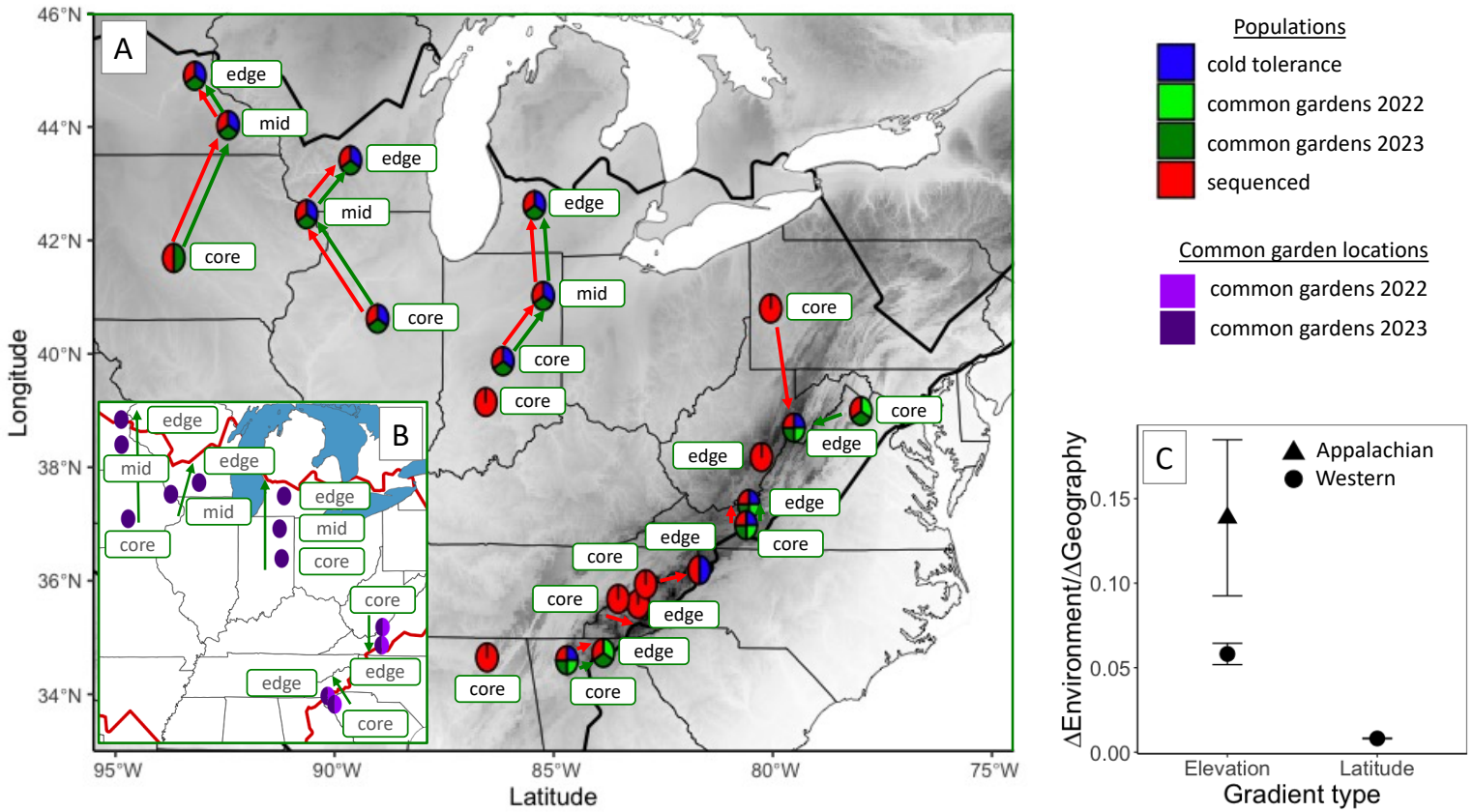


Figure 1) Populations of *C. americana* used in experiments. Black borders denote species range edge; grayscale gradient indicates elevation. (A) Population locations including range position and experiments in which populations were used; arrows indicate population transects from core to edge. Red arrows indicate transects of sequenced populations for GDM, green indicate transects of common garden populations. (B) Common garden locations and direction to range edge; arrows indicate garden transects. Populations used in each experiment are denoted by color: blue indicates populations used in cold tolerance experiments; light green indicates populations used in 2022 common gardens; dark green are populations used in 2023 common gardens; and red populations were sequenced for genomic work. (C) Steepness of environmental gradients. Steepness is calculated as the environmental distance between common garden populations divided by the geographic distance between them. Error bars indicate standard errors of mean gradient steepness estimates.

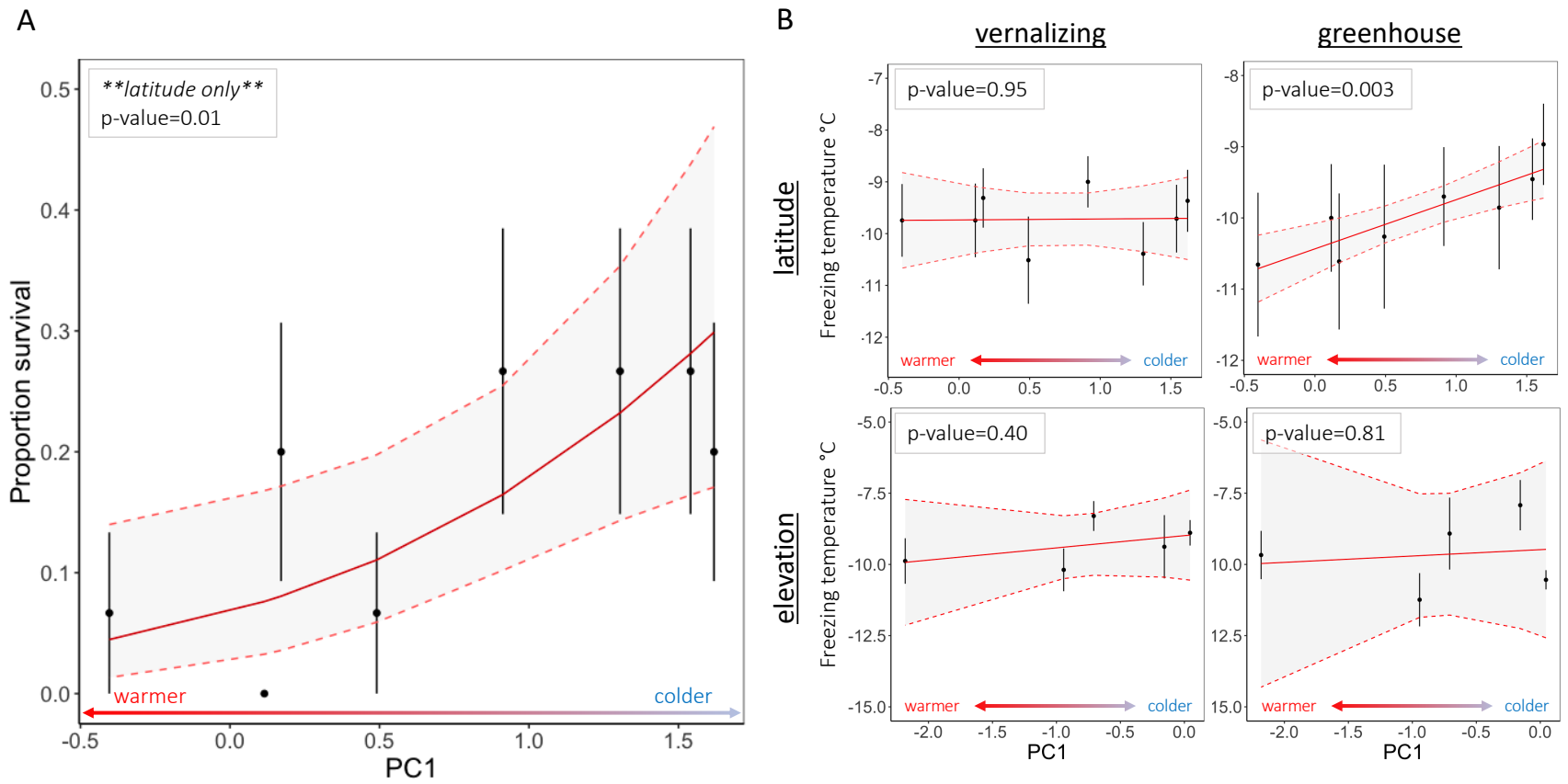


Figure 2) Cold tolerance measured as (A) survival through an acute cold stress (-20°C) during vernalization for populations along a latitudinal gradient where local climates are represented by PC1, and (B) frost resistance, measured by leaf freezing temperature, of leaf tissue during vernalization and greenhouse acclimation for populations along elevation and latitude environmental gradients. Error bars indicate standard errors for population means.

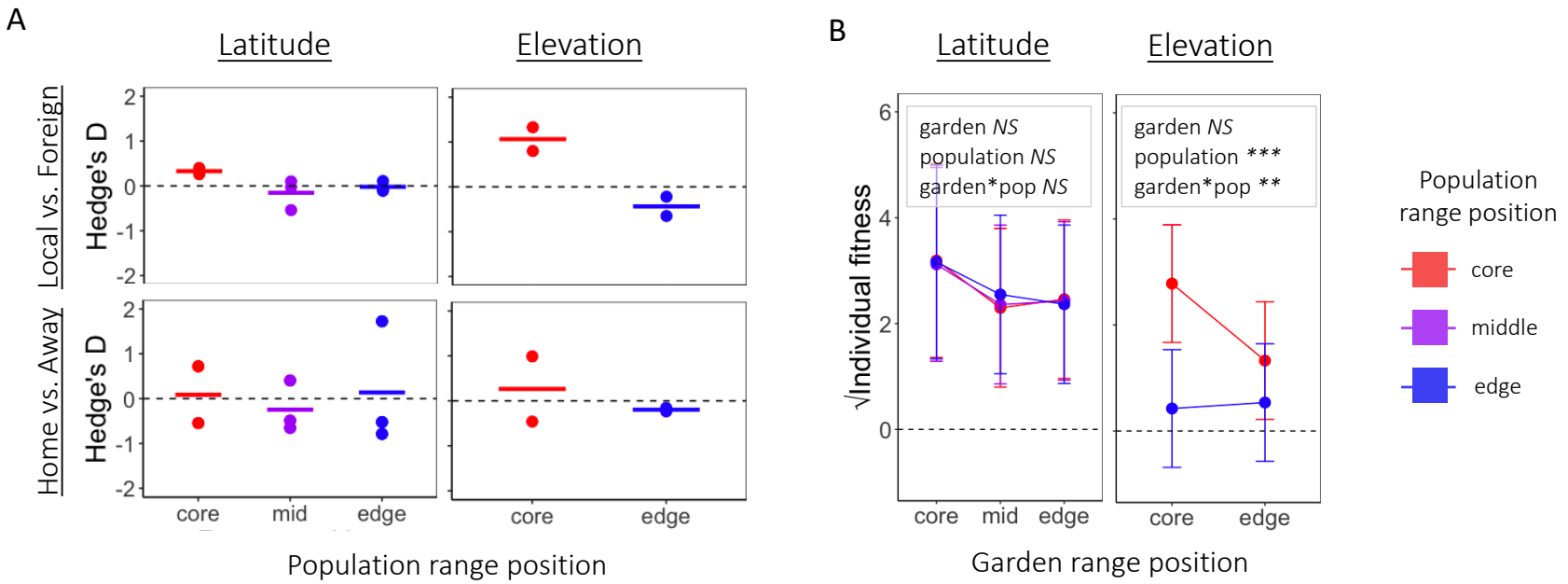


Figure 3) (A) Mean hedge's D estimates of the magnitude of local adaptation across the shallow latitudinal and steep elevational environmental gradients. Local adaptation was estimated as both performance of local vs. foreign populations within common gardens nested within range positions (core/mid/edge) and as performance of populations when home vs. away across common gardens nested within range positions. Positive values of Hedge's D indicate that local populations outperform foreign, or that populations perform better at the home garden than away. (B) Models of absolute fitness across the latitudinal and elevational environmental gradients. Full statistical results in SI Table 4. Error bars are standard errors for model means from the generalized mixed effect ANOVAs. Across all panels, range position along x-axes refers garden range position.

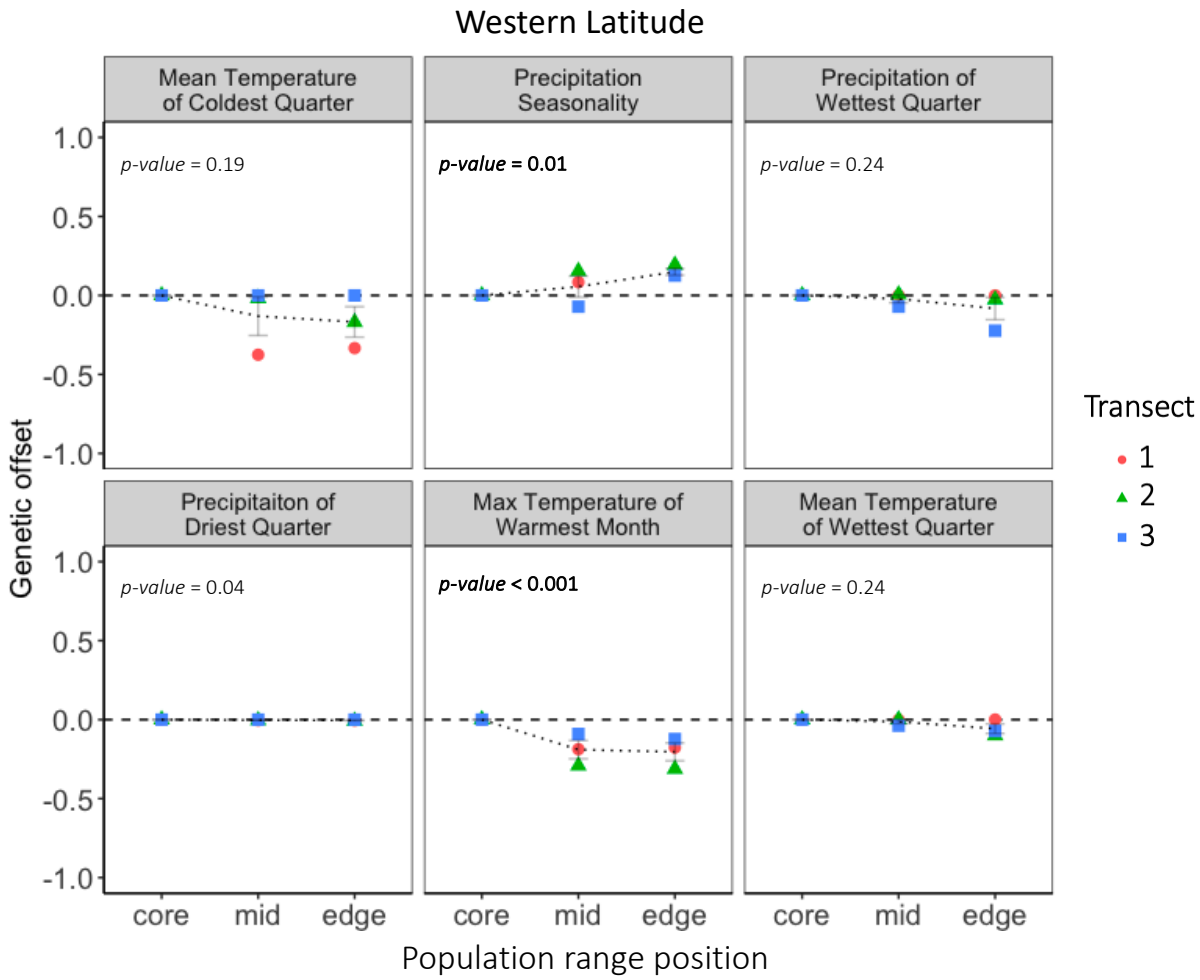


Figure 4) Predicted genetic offset for bioclimatic factors related to local adaptation in latitudinal populations. Genetic offset indicates the difference in adaptive differentiation between expanding range positions (mid/edge) and the range core for each bioclimatic factor. Error bars indicate standard error. Dotted lines indicate range position means, while the dashed line at y-intercept=0 indicates the null expectation of no change in adaptive differentiation (i.e., genetic offset) associated with range expansion and the latitudinal environmental gradient. Statistical test p-values indicate whether differences in genetic differentiation between the range core and other range positions are significant, while the y-axis, genetic offset, indicates the magnitude of genetic differentiation explained by the bioclimatic factor.

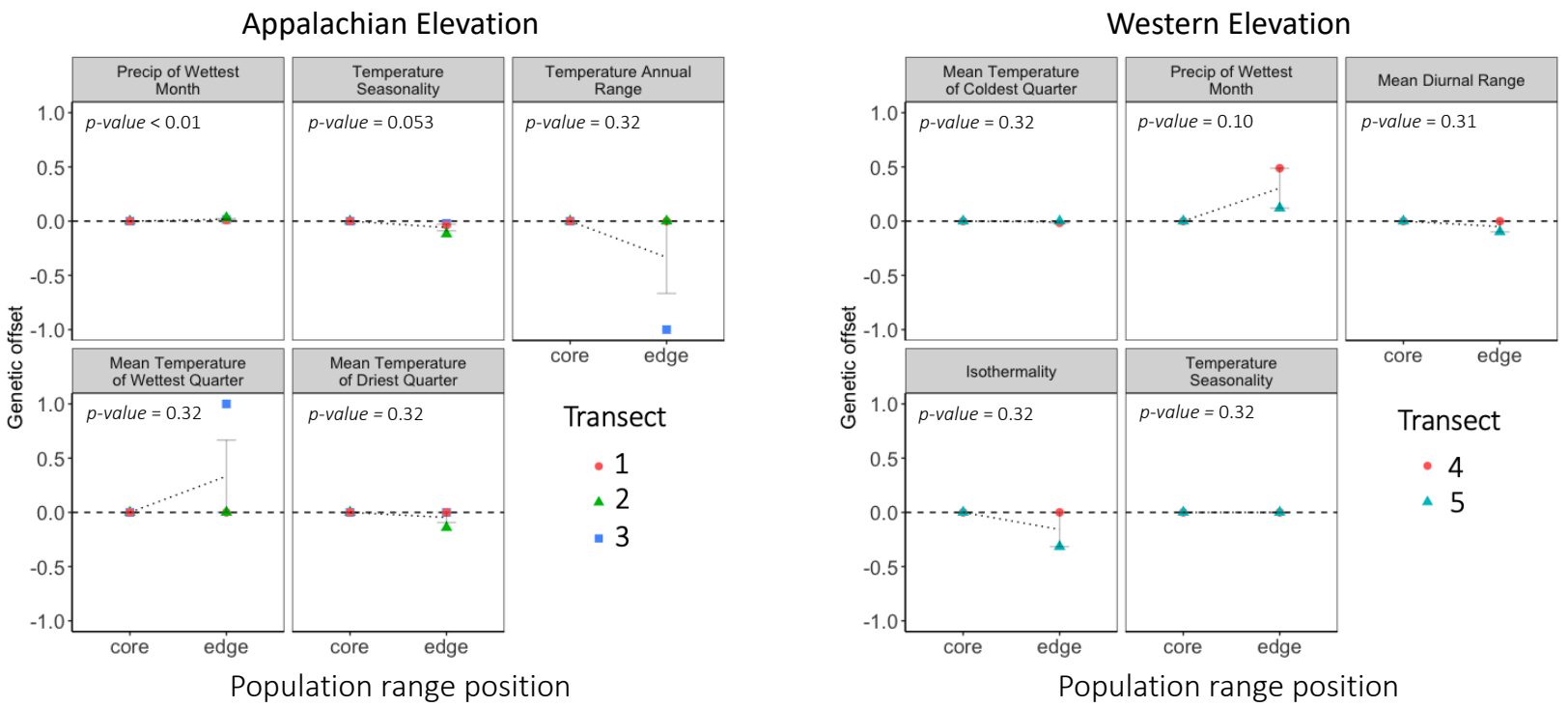


Figure 5) Predicted genetic offset for bioclimatic factors related to local adaptation in elevational populations. Genetic offset indicates the difference in adaptive differentiation between the edge and core of the range for each bioclimatic factor. Error bars indicate standard error. Dashed lines indicate range position means, while the dashed line at y-intercept=0 indicates the null expectation of no change in adaptive differentiation (i.e., genetic offset) associated with range expansion and the elevational environmental gradient. p-values indicate whether differences in genetic differentiation between the range core and other range positions are significant, while the y-axis, genetic offset, indicates the magnitude of genetic differentiation explained by the bioclimatic factor.

GDM summaries				partial coefficients	offset ~ range position	
gradient	lineage	deviance explained	predictor	of genetic turnover	chi^2	p-value
elevation	Appalachian	81.61%	geography	0.13		
			μ temp. wettest quarter	2.07	1.00	0.317
			temp. annual range	1.62	1.00	0.317
			μ temp. driest quarter	0.30	1.00	0.317
			temp. seasonality	0.29	3.75	0.053
			precip. wettest month	0.09	7.30	0.007
	Western	81.04%	geography	1.18		
			precip. wettest month	1.69	2.73	0.098
			isothermality	0.32	1.00	0.317
			μ diurnal range	0.10	1.03	0.310
			μ temp. coldest quarter	0.02	1.00	0.320
			temp. seasonality	0.01	1.00	0.317
latitude	Western	66.22%	geography	0.50		
			precip. seasonality	0.97	9.07	0.011
			μ temp. coldest quarter	0.39	3.37	0.185
			max temp. warmest month	0.32	23.28	< 0.001
			precip. wettest quarter	0.22	2.89	0.236
			precip. driest quarter	0.02	6.57	0.037
			μ temp. wettest quarter	0.10	5.47	0.065

Table 1) Summaries of generalized dissimilarity models (GDMs) exploring patterns of adaptive genetic turnover along environmental gradients. Geographic and bioclimatic variables which were included in final models explained differing amounts of genetic turnover among populations. Higher partial coefficient estimates imply the variable explained more variance and was more important to genetic structure along the gradient and within the lineage. Tests of genetic offset as a function of range position indicate whether differences in adaptive genetic turnover between populations near the range core and other positions (mid/edge) were significant. These tests match results in Figure 4.

SUPPLEMENTARY METHODS

1.1 *Common garden preparation and construction (2022)*

To detect signatures of local adaptation while allowing for the influence of maternal environment, we planted common gardens across Appalachia using field-collected seeds. In 2022, I germinated seeds (Appendix 1) from six elevational populations and then vernalized seedlings under controlled conditions. I planted seedlings in the field at five Appalachian elevational gardens (see SI Table 1; Fig. 1) in mid-March (3/6-3/17). At time of planting, I cleared the ground of herbaceous vegetation and litter, and transplanted individuals 30cm apart in 10 blocks of 12 plants each (2 replicates/population/block; n=120). After planting, I fenced gardens to prevent herbivory. The garden at MLBS, VA, was weeded in late June (6/27). In early August, I visited all gardens and recorded bolting, flowering count, and survival. No plants survived at Clemson or Blandy due to fence damage and herbivory, so I removed the garden from final analyses.

1.2 *Library preparation and sequencing*

Library prep and sequencing were outsourced to the University of Oregon genomics core and Floragenex. Dried leaf samples were processed for sequencing by the genomics core at University of Oregon. UO extracted DNA using kit extractions, then sent to Floragenex for library prep and sequencing. Floragenex digested DNA using the PstI enzyme to generate paired end 150kbp fragments, then multiplexed using RAD and sequence adapters. Fragments underwent size selection, retaining 300-500bp lengths, then PCR amplification. Floragenex performed QA/QC of the library preparation using Qubit and Nanodrop. Finally, Floragenex sequenced libraries using 2 lanes of Illumina NovaSeq6000 Sp300 cycle.

I assembled sequencing reads to filter genetic data to variant loci of interest. First, I quality-checked reads using fastqc (Andrews, 2010), then demultiplexed reads to trim barcodes and assemble reads by individual using process-radtags in Stacks (Catchen et al. 2013). I removed PCR duplicates using clone_filter in Stacks, then merged sequences within individuals across plates. I then aligned sequences within individuals to a reference genome for *C. americana* (unpublished data). Finally, a VCF file was produced using the populations module in Stacks with parameters of -p 1, which requires loci to be in at least one population, -r 0.5, which requires loci to be in at least 50% of individuals within a population, and -maf 0.05, which requires the minimum allele frequency for a given locus to be greater than 5%. Further filtering was performed using VCFTools, the description of which can be found in the main text document.

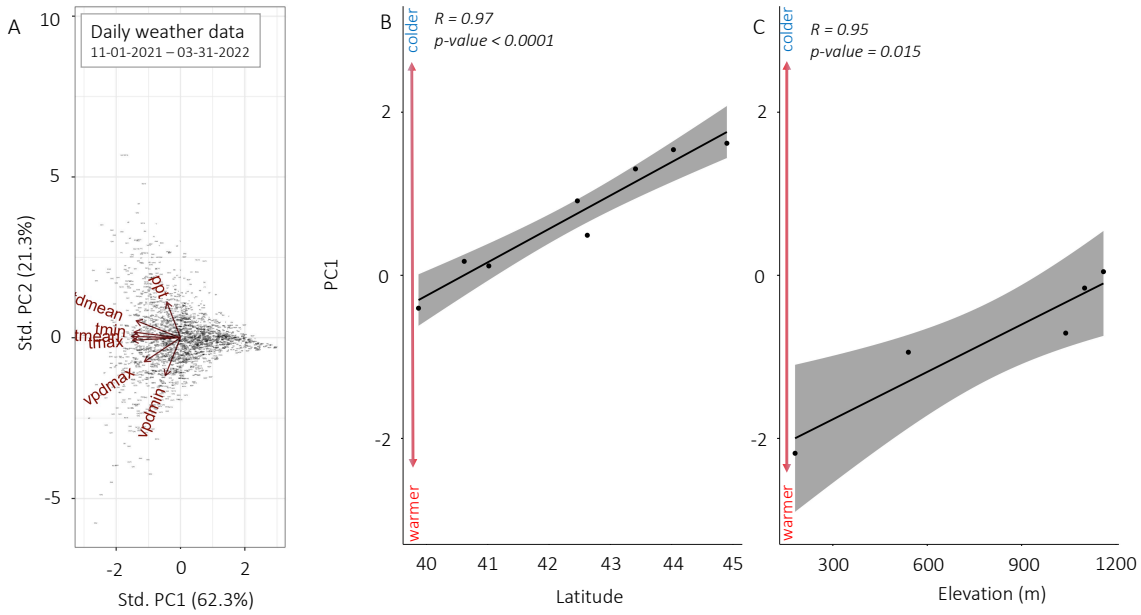
SUPPLEMENTARY RESULTS

1.1 *Environmental variance across gradients (PC1 in photosynthesis, freezing point, and survival analyses)*

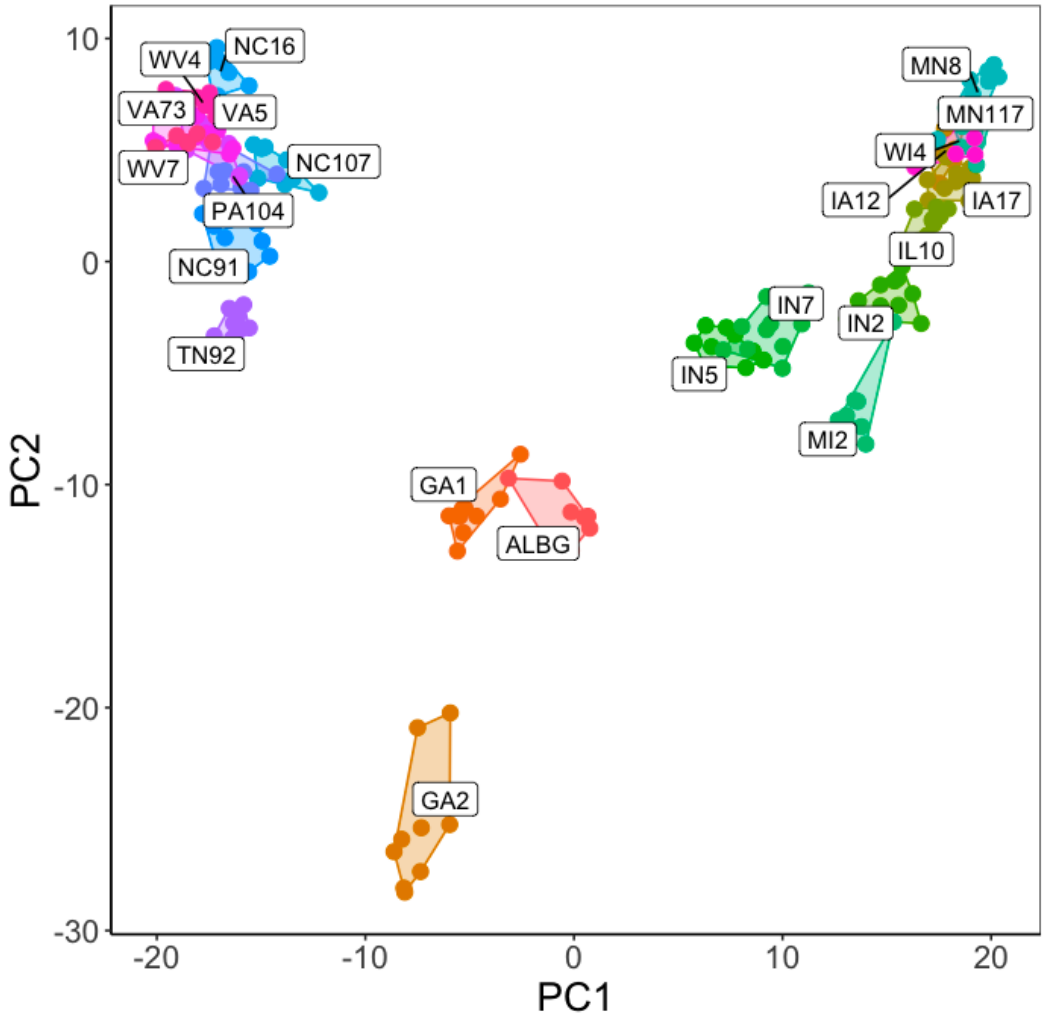
Environmental variation was well explained by latitude and elevation across environmental gradients. The first component from the PCA of daily weather conditions at population origin sites (explaining 62.3% of variance; SI. Fig. 2) was strongly correlated with latitude (R=0.97) for populations hailing from the latitudinal environmental gradient, and elevation (R=0.95) for populations hailing from the elevational environmental gradient. Additionally, while elevational and latitudinal populations did not substantively overlap in values of PC1 (SI Fig. 2B-C), the environmental range, as measured by PC1, was similar, suggesting that the amount of environmental variance experienced is similar across gradients.

1.2 *Genome-scanning methods*

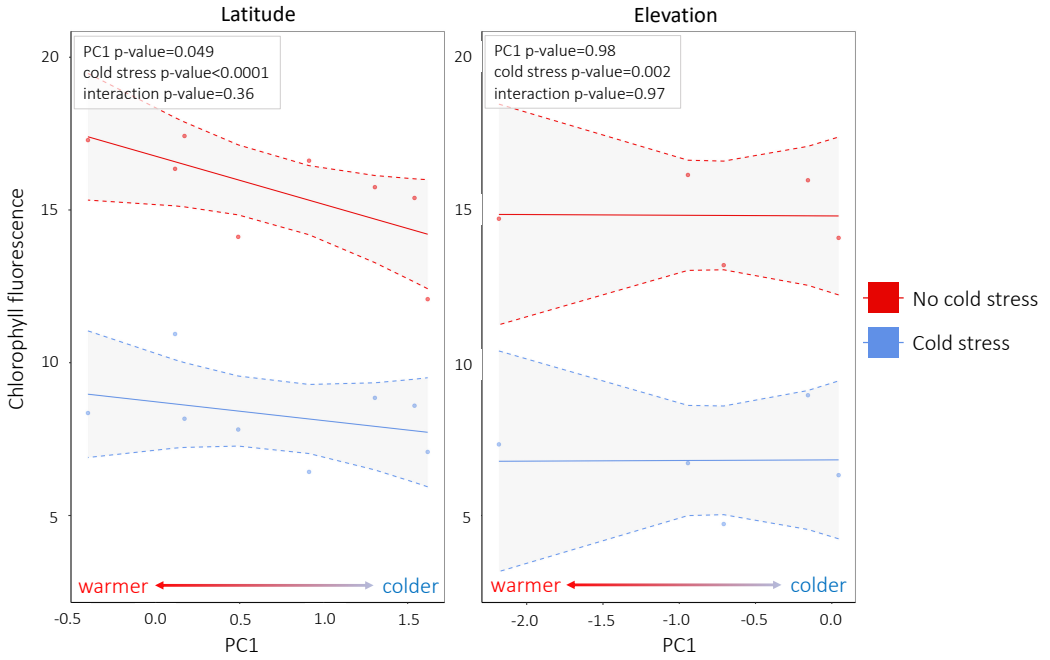
I identified loci underlying environmental selection using PCAdapt and BayeScEnv. We retained loci only if found in common between the two genome-scanning methods. For latitudinal populations, PCAdapt identified 119 loci related to local adaptation with a q-value FDR correction of 0.01 and BayeScEnv identified seven loci with a q-value FDR correction of 0.05, and an overlap of five loci. For the elevational gradient, across Appalachian populations, PCAdapt identified 116 loci and BayeScEnv identified five, with five loci in common. For Western lineage populations in the elevational gradient, PCAdapt identified 264 loci, BayeScEnv identified nine loci, with an intersect of eight loci. Loci which were identified by PCAdapt but not identified by BayeScEnv may in part represent false positives, though others may underlie selection on loci from other environmental sources, such as photoperiod and soil conditions.



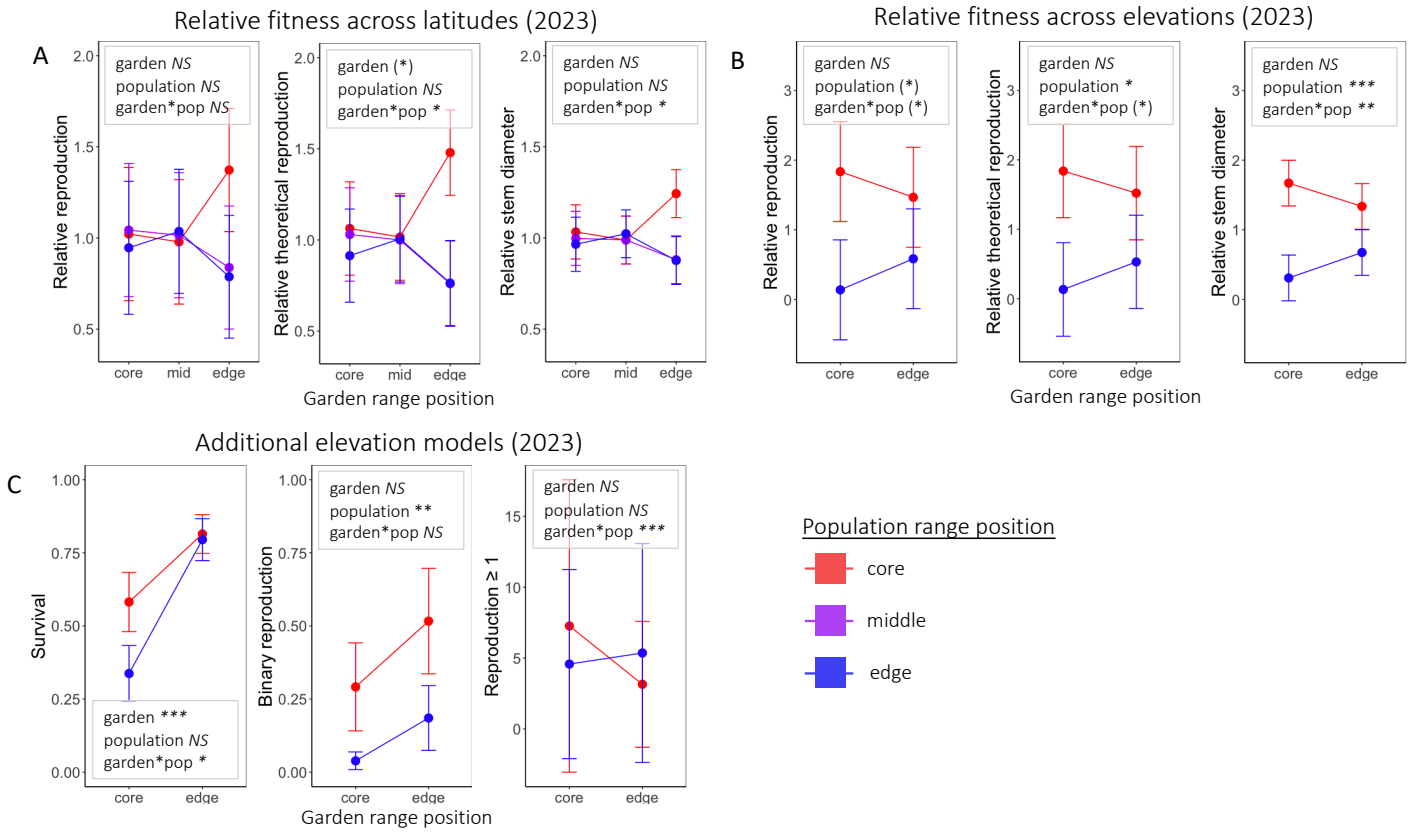
SI Figure 1) Principal component analysis for daily weather data for winter and spring months (November-May) for the years 2017-2022. (A) The principal component analysis; PC1 and PC2 together explained 83.6% of variance in climate. Strong correlations of (B) population latitude of origin and (C) population elevation of origin with PC1; both gradients traversed a similar magnitude of environmental variation. Gray bands denote 95% confidence intervals for regressions.



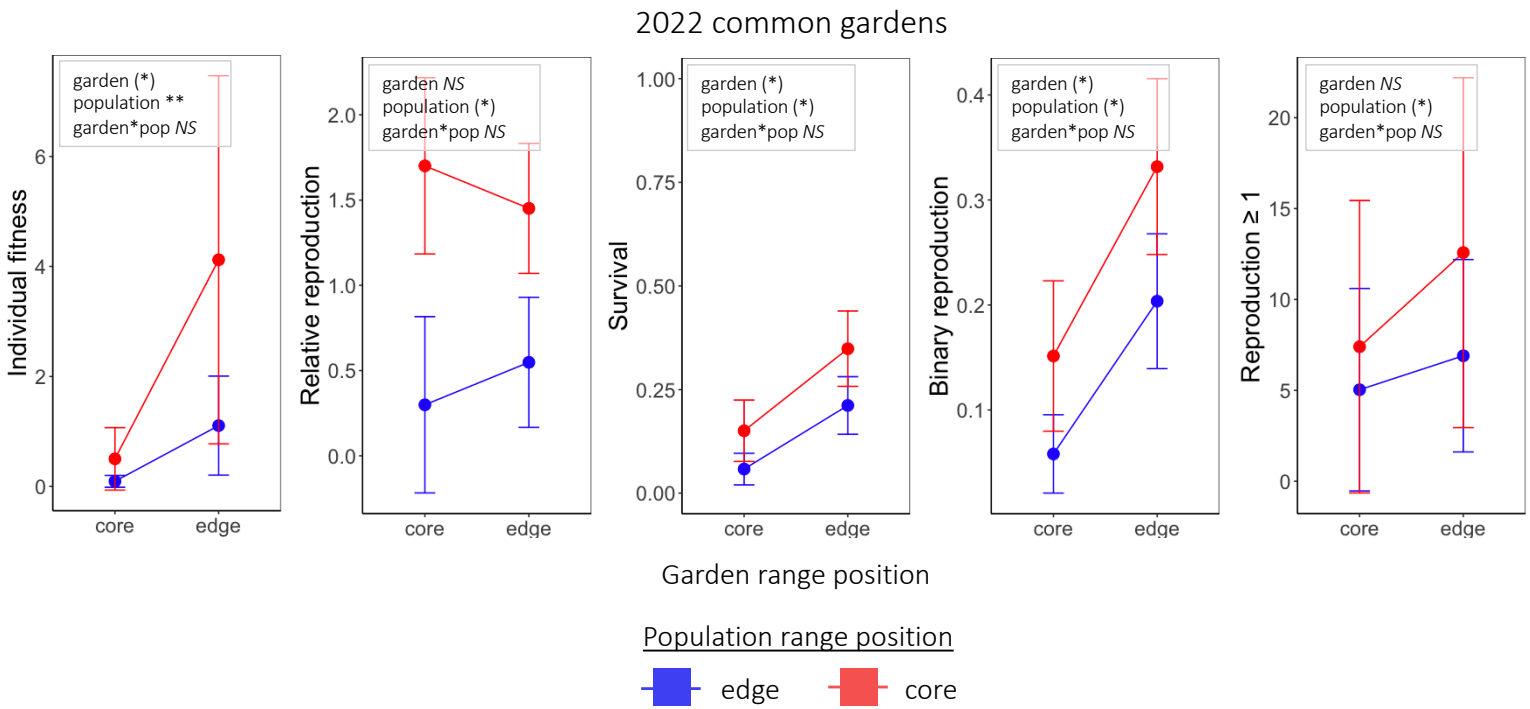
SI Figure 2) Principal component analysis of filtered SNPs which informed gene-by-environment association analyses (PCAdapt and BayeScEnv). Populations formed well-defined clusters.



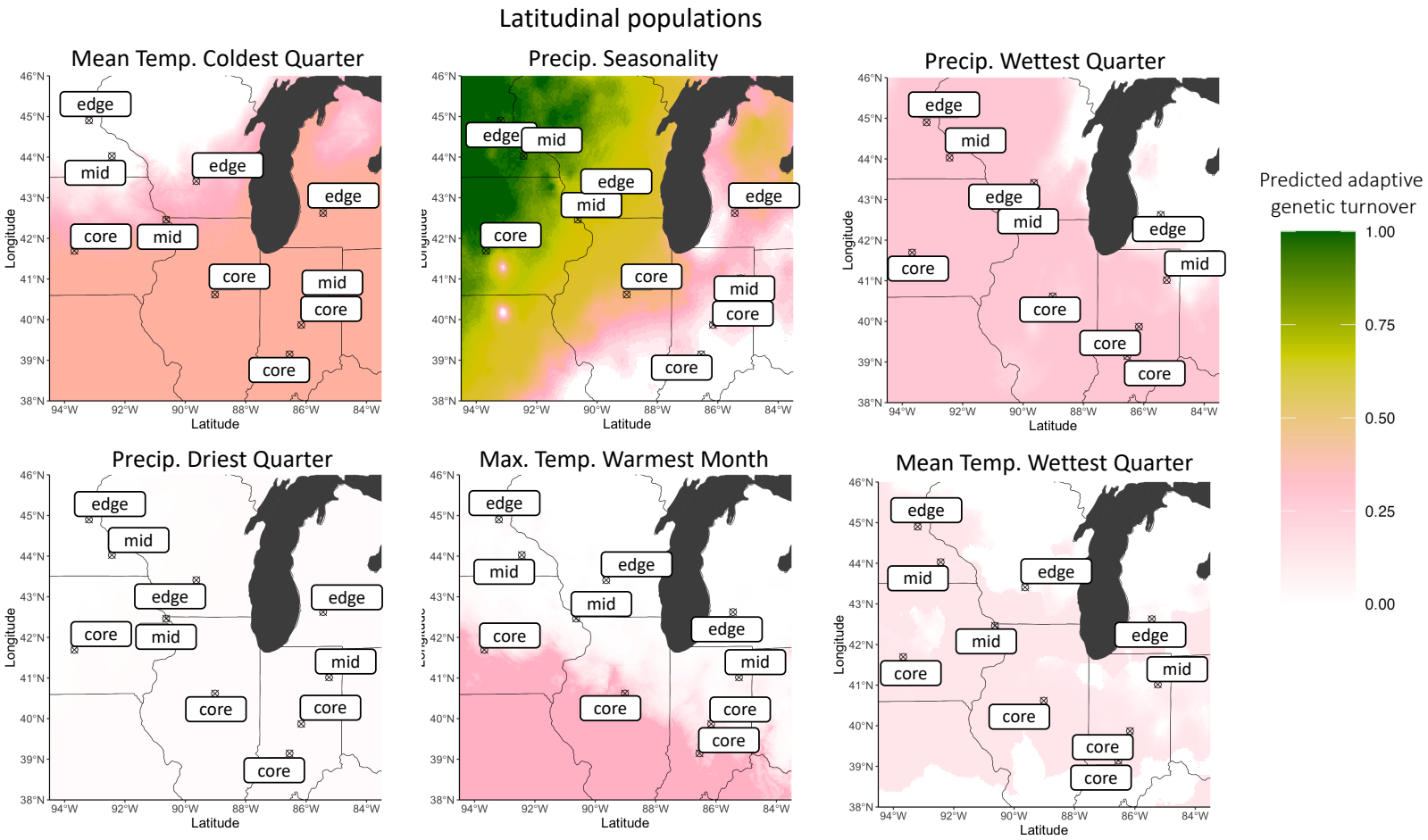
SI Figure 3) The association of chlorophyll content and climate of the site of population origin (PC1) for the latitude and elevation gradients, comparing plants that experienced an acute cold treatment with those that did not. Dashed bands denote 95% confidence intervals for regressions.



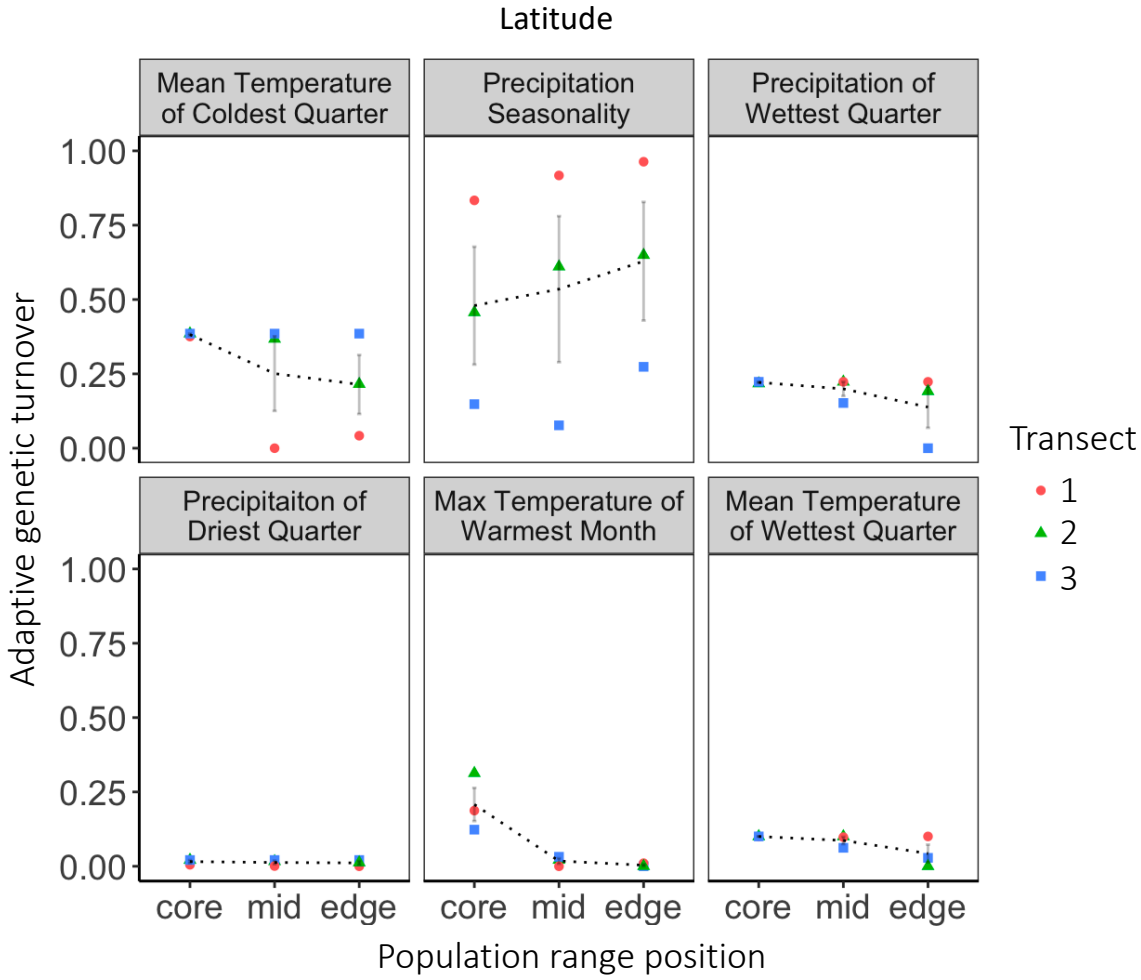
SI Figure 4) Models of local adaptation from common gardens planted in 2023 across latitudinal and elevational environmental gradients. Across all models, x-axis range position refers to garden range positions. (A) Mixed effect models estimated means for latitudinal populations, with each garden nested within a range position (core, middle, edge) for metrics of fruit and flower count relativized within gardens, relativized total theoretical reproduction, and relativized stem diameter. (B) Mixed effect models estimated means for elevational populations with each garden nested within a range position (core, edge) for relativized metrics found in (A). (C-E) Additional models related to elevational populations to account for poor survival and flowering at high elevation sites. (C) Percentage survival modeled for elevational populations across gardens using binomial generalized modeling. (D) Binomial flowering model for elevational populations, wherein plant fitness was simplified as whether surviving plants flowered or failed to. (E) Modeled means of raw fruit and flower counts of plants which survived and flowered or fruited at least once in elevational gardens. (E) demonstrates the sole instance in which local adaptation was detected among high elevation populations at high elevation sites. Across plots, error bars indicate 95% confidence intervals. Full statistical results for each model can be found in SI Table 4. For p-values, (*NS*) indicates non-significance ($p>0.10$), (*) indicates $p\leq0.10$, * indicates $p\leq0.05$, ** indicates $p\leq0.01$, and *** indicates $p\leq0.001$.



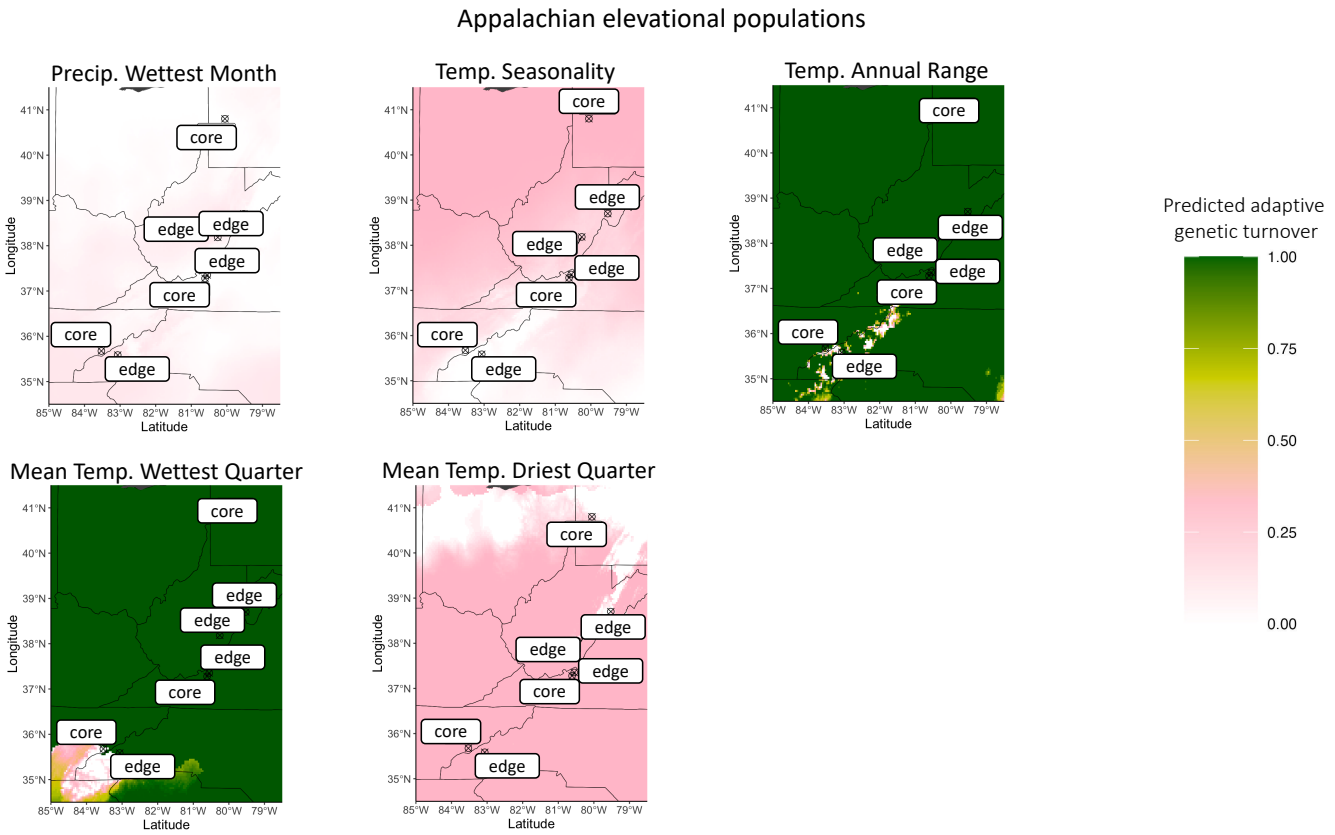
SI Figure. 5) Models of local adaptation from common gardens planted in 2022 across the elevational environmental gradient. Across all models, x-axis range position refers to garden range positions. Full statistical results for each model can be found in SI Table 4. For p-values, (*NS*) indicates non-significance (p>0.10), (*) indicates p≤0.10, * indicates p≤0.05, ** indicates p≤0.01, and *** indicates p≤0.001.



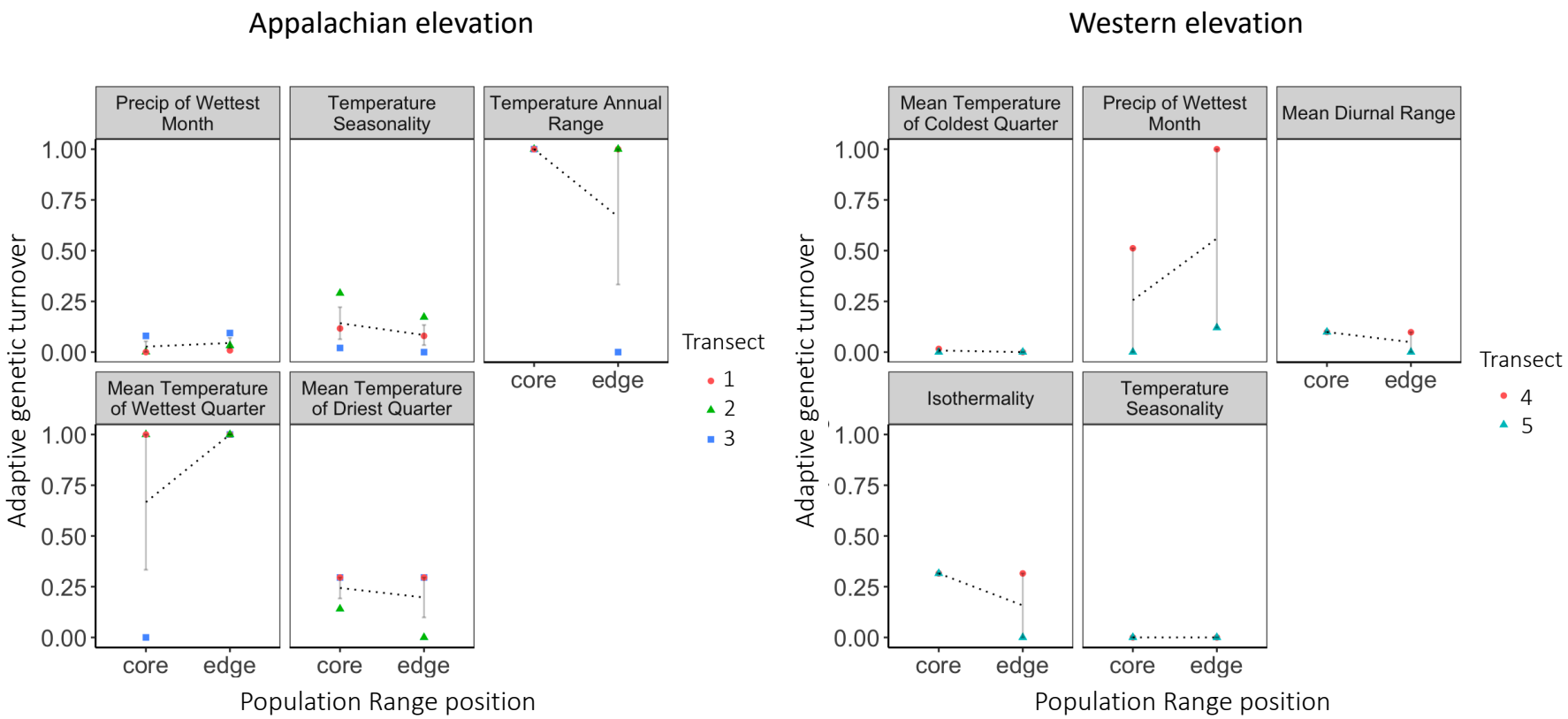
SI Figure 6) Map projections for all variables retained in the Generalized Dissimilarity Models for latitudinal populations. Populations included in the model are shown. Colors indicate amount of adaptive genetic turnover predicted from GDMs. Bioclimatic factors are arranged in order of sum of coefficients size from the GDM, thus predictors which explain less adaptive genetic differentiation produce more blanched map projections.



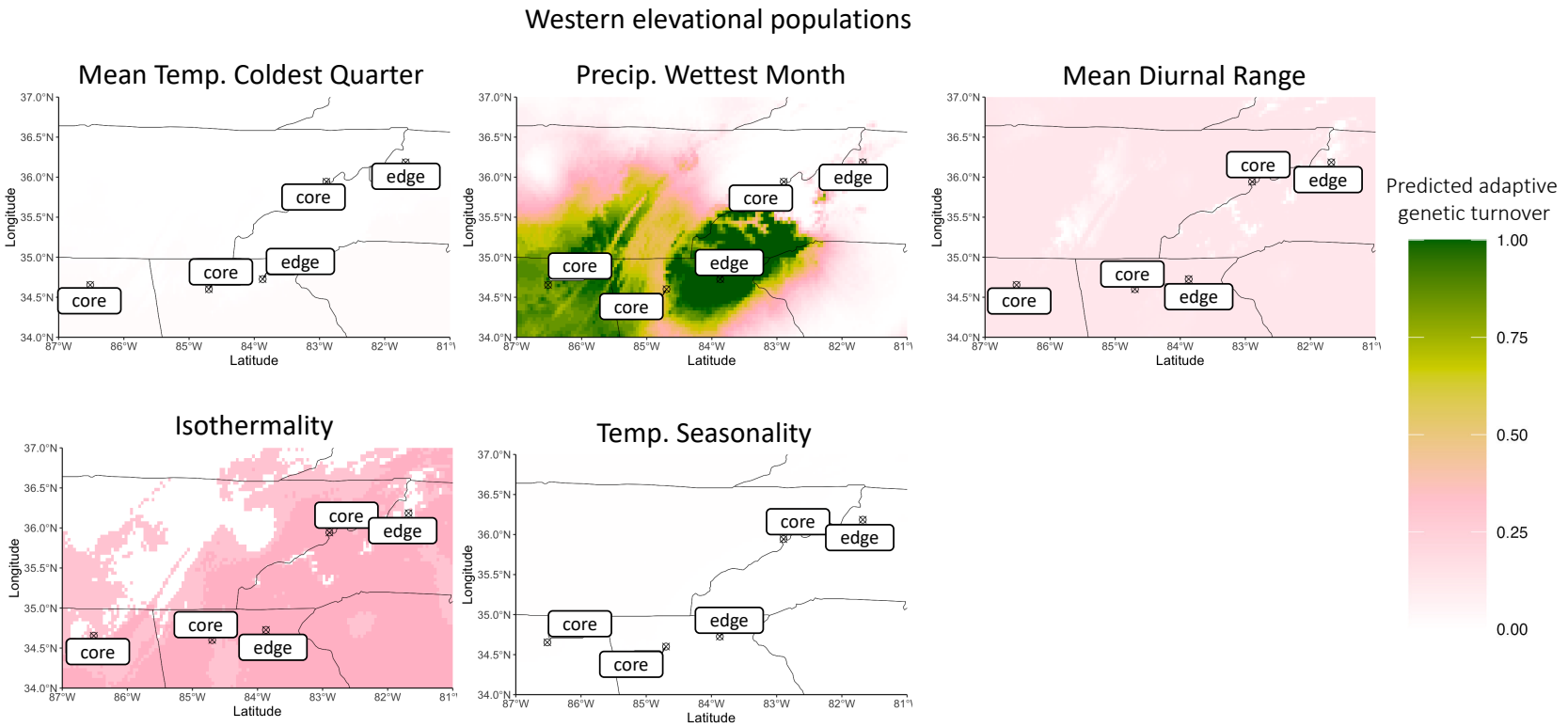
SI Figure 7) Predicted adaptive genetic turnover, which indicates the degree of adaptive genetic differentiation associated with each bioclimatic factor, of Western-lineage latitudinal populations nested within range position for each bioclimatic factor contributing to the GDM. Significance was determined by univariate ANOVA models with a random term for transect. See Figure 1 for transect locations; dotted lines show the mean by range position with standard error bars.



SI Figure 8) Map projections for all variables retained in the GDM for Appalachian-lineage elevational populations. White textboxes and black x-marks indicate the locations of populations informing the model. Geographic distance between populations was significant in the GDM, with a sum of coefficients of 0.155. Colors indicate the adaptive genetic turnover predicted from the GDM. Bioclimatic factors are arranged in order of sum of coefficients size from the GDM.



SI Figure 9) Predicted adaptive genetic turnover of elevational populations, which indicates the degree of adaptive genetic differentiation associated with each bioclimatic factor, nested within range position for each bioclimatic factor contributing to the GDM. Significance was determined by univariate ANOVA models with a random term for transect. Dotted lines show the mean by range position with standard error bars.



SI Figure 10) Map projections for all variables retained in the GDM for Western-lineage elevational populations. White textboxes and black x-marks indicate the locations of populations informing the model. The red border indicates the species range edge, as calculated from iNaturalist observations. Geographic distance between populations was the best predictor of genetic differentiation (sum of coefficients = 1.182). Colors indicate the adaptive genetic turnover predicted from the GDM. Bioclimatic factors are arranged in order of sum of coefficients size from the GDM.

population	transect	gradient	range group	lineage	latitude	longitude	cold tolerance	gardens 2022	gardens 2023	sequencing
TN92	8	elevation	core	Appalachian	35.68	-83.53				1
VA5	7	elevation	core	Appalachian	37.28	-80.61	1	1	1	1
PA104	6	elevation	core	Appalachian	40.80	-80.05				1
NC91	8	elevation	edge	Appalachian	35.59	-83.07				1
VA73	7	elevation	edge	Appalachian	37.35	-80.55	1	1	1	1
WV7	NA	elevation	edge	Appalachian	38.18	-80.26				1
WV4	6	elevation	edge	Appalachian	38.70	-79.52	1	1	1	1
VA3	6	elevation	core	Western	38.99	-77.99		1	1	
GA1	4	elevation	core	Western	34.60	-84.70	1	1	1	1
ALBG	NA	elevation	core	Western	34.65	-86.52				1
NC107	5	elevation	core	Western	35.94	-82.90				1
GA2	4	elevation	edge	Western	34.73	-83.87		1	1	1
NC16	5	elevation	edge	Western	36.19	-81.68	1			1
IN5	3	latitude	core	Western	39.15	-86.55				1
IN7	3	latitude	core	Western	39.87	-86.16	1		1	1
IL10	2	latitude	core	Western	40.62	-89.02	1		1	1
IA12	1	latitude	core	Western	41.69	-93.67			1	1
MI2	3	latitude	edge	Western	42.62	-85.44	1		1	1
WI4	2	latitude	edge	Western	43.41	-89.64	1		1	1
MN117	1	latitude	edge	Western	44.90	-93.19	1		1	1
IN2	3	latitude	mid	Western	41.02	-85.24	1		1	1
IA17	2	latitude	mid	Western	42.46	-90.64	1		1	1
MN8	1	latitude	mid	Western	44.03	-92.43	1		1	1

SI Table 1) Populations used across experiments. Populations are divided by environmental gradient, range position, and lineage. Experiments which populations were used in are denoted with '1'.

	Growth Chambers			Cold Room		
Experiment	**time**	**DL cycle**	**temp.**	**time**	**DL cycle**	**temp.**
photosynthetic function / survival	5 weeks	14h/10h	21°C/14°C	8 weeks	12h/12h	4°C
frost resistance	5 weeks	14h/10h	21°C/14°C	6 weeks	12h/12h	4°C
common gardens 2022 Elevation	6 weeks	14h/10h	21°C/14°C	7 weeks	12h/12h	4°C
common gardens 2023 Elevation	6 weeks	14h/10h	21°C/14°C	7 weeks	12h/12h	4°C
common gardens 2023 Latitude	6 weeks	14h/10h	21°C/14°C	6 weeks	12h/12h	4°C

SI Table 2) We sowed field seeds in 3:1 (LMM1 garden mix:turface) soil mixture and placed them in growth chambers for 5-6 weeks. Unless otherwise specified, we sowed 5 seeds per cell. Growth chambers cycled 12h light (21°C) and 12h dark (14°C) periods. After, we allowed seedlings to vernalize for 6-8 weeks in a cold room held at 4°C and cycled with 12h light and 12h dark periods. We then transplanted plants into cone-tainers in identical soil mix, moved individuals to the greenhouse at the University of Virginia in Charlottesville, VA, and used daylight extenders to ensure plants received at least 16 hours of light per day.

garden	gradient	range group	latitude	longitude	gardens 2022	gardens 2023
CLEMSON E.F.	elevation	core	34.73	-82.85	1	1
KENTLAND FARM	elevation	core	37.20	-80.56	1	1
BLANDY	elevation	core	39.06	-78.07	1	1
HIGHLANDS L.T.	elevation	edge	35.05	-83.19	1	1
M.L.B.S.	elevation	edge	37.37	-80.52	1	1
BALL STATE	latitude	core	40.20	-85.42		1
C.E.R.A.	latitude	core	41.69	-92.86		1
MERRY LEA	latitude	mid	41.31	-85.52		1
WOLTER WOODS	latitude	mid	42.61	-90.79		1
ST. OLAF	latitude	mid	44.46	-93.18		1
P.C.C.I.	latitude	edge	42.54	-85.30		1
U.W. ARBORETUM	latitude	edge	43.04	-89.42		1
CEDAR CREEK	latitude	edge	45.40	-93.20		1

SI Table 3) Garden locations and their respective positions along either gradient used for common gardens during 2022 and 2023.

gradient	year	model type	term	chi^2	p-value
elevation	**2023**	**√individual fitness**	garden.group	0.96	0.328
			pop.group	**13.54**	**0.000**
			garden.group*pop.group	**8.16**	**0.004**
		survival	**garden.group**	**12.41**	**0.000**
			pop.group	2.19	0.139
			garden.group*pop.group	**4.07**	**0.044**
		binomial reproduction	garden.group	2.10	0.147
			pop.group	**7.23**	**0.007**
			garden.group*pop.group	1.99	0.158
		reproduction ≥ 1	garden.group	0.15	0.696
			pop.group	0.15	0.701
			garden.group*pop.group	**16.82**	**0.000**
		relative reproduction	garden.group	0.43	0.510
			pop.group	3.56	0.059
			garden.group*pop.group	3.64	0.056
		relative theoretical reproduction	garden.group	0.38	0.540
			pop.group	**4.05**	**0.044**
			garden.group*pop.group	3.14	0.076
		relative stem diameter	garden.group	1.10	0.294
			pop.group	**13.44**	**0.000**
			garden.group*pop.group	**10.27**	**0.001**
	2022	**individual fitness**	garden.group	3.36	0.067
			pop.group	**10.17**	**0.001**
			garden.group*pop.group	2.09	0.149
		survival	garden.group	0.54	0.462
			pop.group	2.29	0.130
			garden.group*pop.group	0.08	0.772
		binomial reproduction	garden.group	3.63	0.057
			pop.group	2.97	0.085
			garden.group*pop.group	0.32	0.570
		reproduction ≥ 1	garden.group	0.06	0.809
			pop.group	3.23	0.072
			garden.group*pop.group	0.52	0.469
		relative reproduction	garden.group	0.17	0.680
			pop.group	2.97	0.085
			garden.group*pop.group	0.36	0.546
latitude	2023	√individual fitness	garden.group	0.16	0.925
			pop.group	0.04	0.979
			garden.group*pop.group	2.13	0.711
		relative reproduction	garden.group	2.06	0.356
			pop.group	0.05	0.976
			garden.group*pop.group	5.33	0.255
		relative theoretical reproduction	garden.group	5.15	0.076
			pop.group	0.22	0.895
			garden.group*pop.group	**11.31**	**0.023**
		relative stem diameter	garden.group	3.76	0.152
			pop.group	0.16	0.922
			garden.group*pop.group	**12.45**	**0.014**

SI Table 4) Results from statistical models testing local adaptation across elevational and latitudinal environmental gradients. Fitness was relativized within gardens. For elevational populations, high elevation, range edge populations often failed to bolt so we used models of survival (0|1), binary flowering (0|1) and logarithmic reproductive output among plants that flowered (≥1).

CHAPTER 2:

Steep environmental gradients drive divergent patterns of selection on phenology and loci associated with adaptive differentiation.

ABSTRACT

Environmental selection undergirds the process of local adaptation in wild populations. Yet, selection functions by reducing genetic diversity and may degrade the adaptive potential of populations in novel habitat. Here, I explored how the strength of selection shapes patterns of adaptive differentiation along environmental gradients of differing steepness. I found that strong selection along steep environmental gradients, drives polarized allele frequencies and divergent timing of reproduction. Along shallow environmental gradients, I found evidence of weak selection and gradual clines of adaptive allele frequencies, suggesting that adaptive differentiation is weak across the span of the shallow gradient. Together, these patterns suggest that strong and divergent patterns of selection along environmental gradients can maintain adaptive differentiation among interconnected populations. However, selection along the steep environmental gradient was also associated with downstream costs to adaptive potential and weak fitness of high elevation populations (Chapter 1), suggesting that divergent selection along environmental gradients can push populations toward source-sink dynamics.

INTRODUCTION

Survival in novel habitat often requires adaptation to local environmental conditions enabled by selection. Environmental gradients, where environments change linearly in space, represent unique opportunities to understand how patterns of adaptation develop across heterogeneous landscapes under different selection regimes. Where environmental gradients are shallow, differences in the strength and direction of selection along the gradient may be minimal, resulting in weak adaptive differentiation among populations (Kawecki and Ebert 2004; Spichtig and Kawecki 2004). In contrast, where environmental gradients are steep, differences in the strength and direction of selection along the gradient are greater, driving increased adaptive differentiation among populations. Accordingly, adaptive differentiation along environmental gradients is expected to be driven foremost by differences in the strength and direction of selection among populations.

Selection on vegetative traits and reproductive phenology can vary along environmental gradients (Angert et al. 2011; Ochocki and Miller 2017). These gradients can generate stabilizing selection within populations and divergent selection among populations. Divergent selection on vegetative traits and climate adaptation along environmental gradients can produce tradeoffs in fitness associated with dispersal and local adaptation (here, costs of local adaptation; Kassen 2002). As a byproduct of these costs, gene flow among populations along the gradient may decrease (Tigano and Friesen 2016).

Divergent selection among populations may be common along steep environmental gradients (Wong et al. 2020), where environmental conditions change rapidly over space. Consequently, steep environmental gradients may favor the evolution of ecological specialists (Kassen 2002; Kawecki and Ebert 2004). Yet, selection against gene flow along steep environmental gradients may limit the introduction of genetic variation (Tigano and Friesen 2016),

leading to reductions in effective population size (Falk et al. 2012). In turn, these reductions in effective size can limit the adaptive potential of populations to persist through environmental disruption (Oakley 2013) or seed further range expansion. Furthermore, one theoretical model has posited that interactions between selection and gene flow likely exhibit threshold behavior (Yeaman and Otto 2011), where minor shifts in their balance can induce rapid transitions in the maintenance of adaptive polymorphism across range space. Where local adaptation is disrupted along steep environmental gradients due to maladaptive gene flow or reduced effective population size, fitness may be reduced due to conditional (i.e., costs associated with being non-locally adapted) and unconditional costs (i.e., costs associated with genetic drift).

In contrast, differences in the strength and direction of selection are often minimal along shallow environmental gradients (Polechová and Barton 2015), allowing gene flow to perfuse freely among populations. Increased gene flow among populations can increase heterozygosity in adaptive loci, reducing local adaptation. However, weak fitness costs associated with gene flow among populations can promote different patterns of adaptation, including ecological generalism (Spichtig and Kawecki 2004; Tusso et al. 2021). Consequently, local adaptation may be inhibited along shallow environmental gradients, while fitness across a range of environmental conditions is maintained.

Yet, it is unclear what circumstances may lead to the erosion of local adaptation and fitness along environmental gradients of differing steepness. To understand adaptation along different environmental gradients, I analyzed patterns of selection along a steep and shallow environmental gradient. First, I addressed whether and how selection on vegetative traits and reproductive phenology differs along a steep environmental gradient using wild populations of *Campanula americana,* an herbaceous wildflower. Second, I analyzed patterns of genetic differentiation in

populations of *C. americana* along a steep and a shallow environmental gradient to infer patterns of selection on adaptive loci. Finally, I relate patterns of selection in vegetative traits and adaptive loci to *in situ* phenology and discuss how local adaptation can be maintained or disrupted along environmental gradients.

METHODS

Study system

Campanula americana is an herbaceous monocarpic wildflower native to eastern North America. *C. americana* is found throughout the Appalachian Mountains across elevations, though typically in patchy, locally abundant distributions. In general, *C. americana* is divided into three genetic lineages. Here, the Eastern and Western lineages, which are closely related (Barnard-Kubow et al. 2015), are combined. The Appalachian lineage, which resides solely in the Appalachian Mountains, expanded along an elevational environmental gradient after the last glacial maximum (Chapter 5; Appendix 1; Barnard-Kubow et al. 2015). The Western lineage, which resides in the American Southeast and Midwest, experienced two separate range expansions after the last glacial maximum. In the north, the Western lineage expanded along a latitudinal gradient from a mid-latitude refugium after the last glacial maximum (Barnard-Kubow et al. 2015; Koski et al. 2019; Prior et al. 2020). Here, the latitudinal gradient traversed by the Western lineage is treated as equivalent to a shallow environmental gradient (Fig. 1B). In the southern Appalachian Mountains, populations from the Western lineage's relictual southern range expanded independently along a steep elevational gradient (Barnard-Kubow et al. 2015). Here, the elevational gradient is treated as equivalent to a steep environmental gradient (Fig. 1B).

Flowering phenology differs across the Western lineage's latitudinal range (Prendeville et al. 2013; Perrier et al., *in prep*), with more plasticity in onset of flowering toward the northern range limit. Reproductive duration and onset of flowering is also highly plastic across elevational populations in the Appalachian lineage (Haggerty and Galloway 2011), though onset of flowering is at least in part genetically determined. Additionally, selection favors earlier flowering and shorter flowering duration in low elevation sites than high elevation sites (Haggerty and Galloway 2011).

Patterns of local adaptation differ between the latitudinal and elevational environmental gradients. Local adaptation is absent across the Western lineage's shallow latitudinal gradient, though fitness across the range is generally high (Chapter 1). Along the steep elevational gradient in the Appalachian Mountains, populations from the low elevation range core were locally adapted and generally fit across the range, while populations from the high elevation range edge were performed better at home but were generally unfit across the range (Chapter 1). Such patterns suggest ecological generalism along the shallow latitudinal gradient, and ecological specialization coupled with high genetic drift along the steep elevational gradient.

1. Phenotypic selection

Populations and experimental design

To understand how patterns of selection on vegetative and reproductive traits differ across the elevational gradient, I conducted a study of phenotypic selection in seven wild *Campanula americana* populations located throughout the Appalachian Mountains (Fig. 1; SI Table 1). I selected populations to encompass both high and low elevation habitat along the species' elevational gradient (244–1411m), with four populations classified as high elevation (1073–1411m), and three as low elevation (244-503m). Mid-elevation populations are uncommon and

therefore were not included. I tagged all individuals present (n=24–94) in each population with paper jewelry tags in June (6/2/21–6/21/21), prior to flowering, and took measurements of stem diameter and height. I harvested the third leaf below the shoot apical meristem and dried samples from each individual to assess specific leaf area (see Supplementary Methods; SI Fig. 1). After the initial visit, I returned twice (23–41-day intervals) to assess phenology and record survival, with the exception of GH, which I returned to three times due to late-season flowering (SI Table 2). Phenology was assessed by counting the number of immature buds, mature buds (within three days of flowering), flowers, and fruits on each plant at each visit. I removed plants from the study if the plant was missing upon subsequent visits and mortality could not be verified.

Phenotypic differentiation (vegetative and reproductive)

To investigate whether plants found at higher elevations had distinct vegetative phenotypes from lower elevations, I compared vegetative traits measured at the beginning of the selection window. I computed Gaussian ANOVAs for each vegetative trait (*SLA, height, stem diameter*), treating individual phenotypes as a function of population (*individual trait values = population*). I then estimated marginal means for each elevation group (*high/low*) using post-hoc linear contrasts.

To determine if plants from high and low elevation populations flowered at different times in the growing season, I compared reproductive phenology across elevations. To model reproductive phenology, I used nested mixed effect ANCOVAs. I constructed models with fixed and interaction effects of elevation group and the date of each visit to count reproductive traits (*counts of immature buds, mature buds + flowers, fruits*). To account for repeated measures and non-independence, I nested repeated phenology measures within individual ID. Individual ID was nested in population and elevation group (*count = elevation group + date + elevation group * date*

+ *individual*(*population*(*elevation group*))). To meet basic model assumptions of normality, I log-transformed response variables. I validated models with the multcomp package (Hothorn et al., 2008).

Selection on vegetative and reproductive traits

I estimated selection differentials for each vegetative trait to compare variation in the strength and direction of selection across elevations. I first variance-standardized phenotypes of vegetative traits within populations (*standardized phenotype value* = $(y - \mu) \div \sigma$), then calculated relative fitness of individuals in each population using fruit count (*relative fitness* = $y \div \mu$). I constructed mixed effects Gaussian ANCOVAs to quantify selection using the lme4 package (Bates et al. 2014). Models included population nested within elevation group as a random term. I included elevation group and trait values as fixed effect covariates and as an interaction term (*relative reproduction = elevation group + standardized trait + elevation group * standardized trait + population*(*elevation group*)). I calculated selection coefficients from the mixed models using the emtrends function from the emmeans package (Lenth 2023). To determine whether differences in coefficients were significant, I ran generalized mixed effect ANCOVAs (GLMM) with identical model structure but substituted absolute measures of fitness for relative. I ran GLMMs with a Poisson distribution and log-link function.

To estimate the strength and directionality of selection on phenology, I calculated an individual-level index of reproductive phenology. To compute this index, I conducted regressions for each individual in the study, regressing the number of fruits produced by the date of each visit (*fruit count = date*; SI Table 2). I extracted the slope from each regression and used the point estimate as an index of reproductive phenology (here, referred to as the reproductive index). Higher

values of the index indicate that plants produced more fruits toward the end of the growing season, while lower values indicate that individual plants produced more fruits earlier in the growing season. Thus, this index describes the time period of maturity within each phenological phase. For analysis, I variance-standardized the index within populations. A lower value of the standardized reproductive index indicates that the individual produced fruits earlier than other individuals within the population, while a higher value indicates delayed production of fruits compared to other individuals within the population. To calculate selection differentials, I modeled relativized fruit counts as a function of the reproductive index and elevation group using mixed effect Gaussian ANCOVAs and identical model structure as described above. I also calculated reproductive indices for immature bud counts and flower counts. I estimated selection coefficients from the mixed effect ANCOVAs using the emtrends function and calculated the significance of selection coefficients from generalized versions of the mixed effect ANCOVAs. Generalized model counterparts used identical model structure but used absolute of fitness and incorporated a Poisson distribution and log-link function.

2. Genomic selection

Genomic data sequencing and preparation

Divergent selection can produce distinct patterns of allele frequencies for loci associated with local adaptation. Along environmental gradients, divergent selection can produce clines of adaptive allele frequencies, where gradual changes in the strength and direction of selection along the gradient lead to corresponding shifts in allele frequencies (Aguirre-Liguori et al. 2017). To determine if patterns of divergent genomic selection vary between environmental gradients of differing steepness, I investigated clines of adaptive allele frequencies. To generate genomic data,

I grew plants from field-collected seeds and harvested leaf tissue from seedlings (Fig. 1; SI Table 1). Details of DNA extractions from leaf tissue, RAD-Seq library preparation, sequencing, assembly, and SNP filtering are found in Chapter 1.

Identifying putatively adaptive genetic markers

To identify general patterns of divergent selection across the genome, I first identified loci that are associated with population structure and environmental differentiation along each environmental gradient. Loci that exhibit strong signals of both environmental differentiation and population structure are likely to be associated with either local adaptation, driven by divergent selection, or neutral processes confounded with the environmental gradient, like range expansion. To identify putatively adaptive loci, I ran BayeScEnv and PCAdapt (Chapter 1; De Villemereuil and Gaggiotti 2015; Privé et al. 2020). BayeScEnv uses F_{ST}-outlier discovery coupled with information about environmental differentiation among populations to identify putatively adaptive loci. I calculated environmental differentiation among populations by computing a PCA using WorldClim 2.0 bioclimate variables for all research-grade iNaturalist observations of *C. americana* (Fick and Hijmans 2017; iNaturalist Community n.d.). I then calculated PC coordinates for each sequenced population, grouped by lineage and environmental gradient, and found their average PC centroid. Finally, I calculated the distance from each population to its respective PC centroid. I used this index of environmental differentiation among populations distributed along environmental gradients to inform BayeScEnv and identify putatively adaptive loci within each lineage and environmental gradient. I also identified putatively adaptive loci using PCAdapt, which identifies outlier loci driving population structure within metapopulations (Privé et al. 2020). Briefly, I computed a PCA using genetic SNPs in the PCAdapt package framework, then

determined the optimal number of PC axes to retain by visually assessing how many principal components had discrete clustering of populations. I then used this number of principal components for downstream identification of outlier SNPs. To determine the significance of outlier SNPs identified by either PCAdapt or BayeScEnv, I used an alpha level of 0.05, and adjusted p-values using Benjamini-Hochberg correction. I retained the intersection of putatively adaptive loci identified as significant outliers by both PCAdapt and BayeScEnv for use in modeling clines of divergent allele frequencies. BayeScEnv and PCAdapt identified pools of putatively adaptive loci for latitudinal (Western: n=5) and elevational populations (Appalachian: n=5; Western: n=8), though pools did not overlap between gradients or lineages.

RAD-sequencing library preparation randomly samples segments of the genome determined by the location and frequency of restriction sites. Because of this sampling process, loci can be used to gain a coarse understanding of patterns of selection across the genome. However, it is generally not possible to know whether loci are the direct targets of selection or merely linked to other loci under selection. Consequently, exact function of putatively adaptive loci identified by BayeScEnv and PCAdapt are not considered (though genomic positions are reported in SI Table 3).

Genomic selection

To identify whether frequencies of putatively adaptive alleles were associated with environmental differentiation along gradients, I fit models of geographic distance to one-dimensional allele frequency clines (SI Fig. 2). I fit competing clines for each locus identified by genome-scanning methods using the HZAR package (Derryberry, 2019) and compared fits using AICc. In total, I ran two geographic models (latitude/longitude) for elevational populations and

three geographic models for latitudinal populations (latitude/longitude/distance to refugium), where distance to the mid-latitude refugium is the distance from populations to hypothesized origin of northern range expansion. Geographic models were used to ensure that putatively adaptive loci identified by genome-scanning methods were not associated with neutral genetic divergence (i.e., genetic drift) from postglacial range expansion and instead reflect adaptive genetic differentiation among populations. Geographic cline models used free scaling of allele frequencies without exponential tails (i.e., models were not bound to the minimum and maximum allele frequency sampled per cline) to account for incomplete sampling of range space.

I also ran environmental models to determine whether adaptive allele frequencies were associated with environmental gradients. I calculated environmental differentiation across the range by first computing a PCA that incorporated all WorldClim 2.0 bioclimatic factors for *C. americana* observations from iNaturalist (n=7,781). I then predicted PC1 values from the PCA for all populations with sequencing data. Environmental models regressed allele frequency against PC1 as a metric of environmental differentiation among populations. Environmental models were run with both fixed and free allele frequencies without exponential tails, allowing flexible assumptions about whether sampling reflected true environmental range limits. Fixed allele frequencies assume that the full extent of range space has been sampled, while free allele frequencies assume that the geographic distribution of populations in the range may exceed samples included in the model. For environmental models, null models, where slope was equal to zero and intercept was equal to the mean allele frequency across populations, act as the null hypothesis that AICc scores are first compared against. Finally, I compared geographic and environmental models of allele frequencies for each lineage and environmental gradient. I retained

only loci for which the AICc of the environmental model fit was ≤2 AIC points than the AICc any geographic model fit.

RESULTS

1. Phenotypic selection

Phenotypic differentiation

Some vegetative traits were differentiated between high elevation and low elevation populations along the steep environmental gradient. High elevation populations had wider stems (p-value=0.049) and higher SLA (e.g., less expensive leaves; p-value < 0.001) than low elevation populations, though plants did not differ in height across elevations (p-value=0.10, SI Table 4). Likewise, phenology differed by elevation for all phases of flowering and fruiting (*immature budding, flowering, fruiting*; Fig. 2; SI Table 5). In particular, flowering in low elevation populations occurred earlier in the year, whereas high elevation populations located at elevational range limits flowered later (Fig. 2B).

Phenotypic selection

Selection on vegetative phenotypes was largely consistent across elevations. The direction and strength of selection did not differ across elevations for height or SLA (SI Table 5). Generally, selection across elevations favored taller plants and less expensive leaves (SI Fig. 3A,C). However, the strength of selection on stem diameter significantly differed among elevations, with stronger selection for wider stems in low elevation populations than in high elevation populations (SI Fig. 3B). While populations from higher elevations had wider stems and less expensive leaves,

phenotypic differentiation of vegetative traits does not appear to be driven by contemporary divergent selection across elevations.

Selection on budding and flowering phenology generally favored earlier flowering in low elevation populations than in high elevation populations. Individuals had higher fitness within their respective populations if they budded and flowered earlier, though this effect was more dramatic for low elevation populations than high elevation populations, which experienced little selection on flowering phenology (Fig. 3A-B; SI Table 5). In contrast, selection on fruiting phenology favored later fruiting within low elevation populations than within high elevation populations (Fig. 2C; SI Table 5). Thus, net selection on reproduction appears to favor both earlier development of reproduction and longer duration of reproduction at lower elevations when compared to higher elevations.

2. Genomic selection

Patterns of genomic selection

Evidence of adaptive genetic differentiation was present along both environmental gradients, though allele frequency clines were generally steeper for elevational populations than latitudinal (Fig. 4). Geographic models of allele frequency clines, used to exclude loci likely associated with neutral genetic divergence during postglacial range expansion or within-lineage phylogeographic divergence, modestly reduced pools of loci (latitude Western: n=3; elevation Appalachian: n=3; elevation Western: n=2). For the steep elevational gradient, selection was strong, with little to no heterozygosity in adaptive loci (Fig. 4). Such patterns suggest strong selection against introgression among elevational populations for adaptive loci. Conversely, adaptive allele frequency clines were somewhat shallower for latitudinal populations and featured more

intermediate allele frequencies (i.e., heterozygosity) than did elevational populations, suggesting that selection against gene flow may be weaker among latitudinally distributed populations.

DISCUSSION

Divergent patterns of selection among populations are likely ubiquitous features of species distributed in heterogeneous habitat, seeding expectations of abundant local adaptation. Yet, local adaptation may be less common than is often assumed (Leimu and Fischer 2008; Hereford 2009), particularly in small populations. I found strong divergent selection for reproductive phenology among high and low elevation populations, while the strength and direction of selection on vegetative traits did not vary. Selection for earlier flowering and later fruiting in low elevation populations likely indicates that plants that are able utilize more of the local growing season window have higher fitness than plants that do not. In contrast, selection for earlier flowering and later fruiting was weak in high elevation populations, suggesting stabilizing selection on phenology. In early flowering primroses and poppies, pollinator emergence dictates stabilizing selection on phenology at higher elevations (Wu and Li 2017; Kudo and Cooper 2019), while drought strongly influences stabilizing on flowering phenology across elevations in a common Brassica (MacTavish and Anderson 2022).

Additionally, I identified polarized allele frequencies at adaptive loci along a steep environmental gradient, suggesting strong selection for adaptive differentiation among populations at differing elevations. Similar patterns of clinal adaptive allele frequencies have been found along steep elevational gradients in numerous other species, including deer mice (Schweizer et al. 2021), and humans (Scheinfeldt et al. 2012), and are often associated with asymmetric gene flow along the elevational gradient that preserve local adaptation (Chapter 4; Funk et al. 2016; Waterhouse et

al. 2018). Yet, previous work in *C. americana* established that fitness is generally low in populations near elevational range limits (Chapter 1). How then can we accord patterns of strong selection for adaptive differentiation at high elevations with poor fitness of high elevation populations?

Divergent selection along environmental gradients can reduce gene flow among populations, preserving adaptive genetic differentiation. Such patterns have been demonstrated in many species (Tigano and Friesen 2016), and may constitute the primary mechanism by which local adaptation is maintained among interconnected populations. For example, divergent selection maintains phenotypic differentiation of vegetative traits and reproductive phenology among interconnected populations of *Ranunculus bulbosus* found along an elevational gradient (Frei et al. 2014). Similarly, selection against migration has been identified as a key factor contributing to adaptive radiation in monkeyflowers (Stankowski et al. 2019). Yet, selection against migration near range limits may also restrict the introduction of genetic diversity into range-edge populations.

Limited genetic diversity can restrict effective population size, diminishing the opportunity for selection by decreasing heritable genetic variance within populations. Small effective population size has previously been associated with a reduced capacity to respond to environmental disturbance in St. John's wort (Oakley 2013), and increased genetic drift in populations near range limits in *Mercurialis annua* (González-Martínez et al. 2017). This pattern suggests that reductions of effective population size near range limits may strongly constrain the adaptive potential of range edge populations to weather environmental disturbance and seed further range expansion (for more on adaptive potential, see Chapters 1 & 4). Under such conditions, theoretical models predict that range expansion along steep environmental gradients

may be limited (Gilbert et al. 2017), and the evolution of ecological specialist phenotypes favored (Kassen 2002; Kawecki and Ebert 2004). However, such patterns may also lead to the evolution of source-sink dynamics, where high elevation populations near range limits become demographic sinks due to a combination of receiving maladaptive gene flow from the range core and strong genetic drift within populations.

Conversely, differences in the strength and direction of selection among populations may be weak along shallow environmental gradients, resulting in little selection against migration along the gradient. Theoretical models predict that weak differences in selection among populations can increase heterozygosity in adaptive loci (Spichtig and Kawecki 2004), producing gradual clines of allele frequencies with intermediate values (Polechová and Barton 2015). I found more modest clines of adaptive allele frequencies along the shallow latitudinal gradient than I found along the steep elevational gradient (Fig. 4). Concordant with this trend, previous work found limited local adaptation along the shallow latitudinal gradient and high population fitness across latitudes (Chapter 1), except near range limits, where fitness is conditionally reduced by genetic drift (Chapter 3).

Together with previous work, I identified differences in the strength and direction of selection along an elevational gradient, yet weak local adaptation and poor fitness near range limits. While such patterns suggest source-sink dynamics are occurring along the steep elevational gradient, they appear to be the exception across taxa, rather than the rule. Local adaptation along steep environmental gradients is common in many taxa (Halbritter et al. 2015; Sexton et al. 2016), and source-sink dynamics appear to be rare byproducts of gene flow among environmentally differentiated populations (Kottler et al. 2021). Among neighboring populations, adaptive differentiation can be maintained via isolating barriers that reduce the opportunity for maladaptive

gene flow, such as phenological differentiation. Here, stronger selection for early flowering near low elevations may incidentally protect adaptive divergence between populations along the environmental gradient through temporal isolation. Differentiation in phenological windows of reproduction has been shown to increase gene flow among populations of *Nothofagus dombeyi* and *N. pumilio* located in similar environmental conditions, while reducing gene flow between the environmentally-differentiated species (Juri and Premoli 2021). Accordingly, high elevation populations, which lacked a period of temporally isolated flowering (Fig. 2), may be more likely to suffer from gene swamping than low elevation populations. Consequently, temporal separation of reproduction may be an important component of maintaining adaptive differentiation among interconnected populations, particularly when populations are well-differentiated by environmental conditions but not geographic distance.

LITERATURE CITED

Aguirre-Liguori, J. A., M. I. Tenaillon, A. Vázquez-Lobo, B. S. Gaut, J. P. Jaramillo-Correa, S. Montes-Hernandez, V. Souza, and L. E. Eguiarte. 2017. Connecting genomic patterns of local adaptation and niche suitability in teosintes. Mol. Ecol. 26:4226–4240.

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

Barnard-Kubow, K. B., C. L. Debban, and L. F. Galloway. 2015. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am. J. Bot. 102:1842–1853.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2014. Fitting Linear Mixed-Effects Models using lme4. arXiv.

De Villemereuil, P., and O. E. Gaggiotti. 2015. A new F_{ST} -based method to uncover local adaptation using environmental variables. Methods Ecol. Evol. 6:1248–1258.

Falk, J. J., C. E. Parent, D. Agashe, and D. I. Bolnick. 2012. Drift and selection entwined: asymmetric reproductive isolation in an experimental niche shift. Evol. Ecol. Res. 14:403–423.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302–4315.

Frei, E. R., T. Hahn, J. Ghazoul, and A. R. Pluess. 2014. Divergent selection in low and high elevation populations of a perennial herb in the Swiss Alps. Alp. Bot. 124:131–142.

Funk, W. C., M. A. Murphy, K. L. Hoke, E. Muths, S. M. Amburgey, E. M. Lemmon, and A. R. Lemmon. 2016. Elevational speciation in action? Restricted gene flow associated with adaptive divergence across an altitudinal gradient. J. Evol. Biol. 29:241–252.

Gilbert, K. J., N. P. Sharp, A. L. Angert, G. L. Conte, J. A. Draghi, F. Guillaume, A. L. Hargreaves, R. Matthey-Doret, and M. C. Whitlock. 2017. Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. Am. Nat. 189:368–380.

González-Martínez, S. C., K. Ridout, and J. R. Pannell. 2017. Range Expansion Compromises Adaptive Evolution in an Outcrossing Plant. Curr. Biol. 27:2544-2551.e4.

Haggerty, B. P., and L. F. Galloway. 2011. Response of individual components of reproductive phenology to growing season length in a monocarpic herb. J. Ecol. 99:242–253.

Halbritter, A. H., R. Billeter, P. J. Edwards, and J. M. Alexander. 2015. Local adaptation at range edges: comparing elevation and latitudinal gradients. J. Evol. Biol. 28:1849–1860.

Hereford, J. 2009. A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. Am. Nat. 173.

iNaturalist Community. n.d. Observations of Campanulastrum americanum from North America, United States of America observed on/between 08/2011-05/2022.

Juri, G., and A. C. Premoli. 2021. Allochrony of neighbour ecological species: Can isolation by time maintain divergence? The natural experiment of sympatric Nothofagus. For. Ecol. Manag. 497:119466.

Kassen, R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. J. Evol. Biol. 15:173–190.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225–1241.

Koski, M. H., N. C. Layman, C. J. Prior, J. W. Busch, and L. F. Galloway. 2019. Selfing ability and drift load evolve with range expansion. Evol. Lett. 3:500–512.

Kottler, E. J., E. E. Dickman, J. P. Sexton, N. C. Emery, and S. J. Franks. 2021. Draining the Swamping Hypothesis: Little Evidence that Gene Flow Reduces Fitness at Range Edges. Trends Ecol. Evol. 36:533–544.

Kudo, G., and E. J. Cooper. 2019. When spring ephemerals fail to meet pollinators: mechanism of phenological mismatch and its impact on plant reproduction. Proc. R. Soc. B Biol. Sci. 286:20190573.

Leimu, R., and M. Fischer. 2008. A Meta-Analysis of Local Adaptation in Plants. PLoS ONE 3:e4010.

Lenth, R. V. 2023. emmeans: Estimated marginal means, aka least-squares means.

MacTavish, R., and J. T. Anderson. 2022. Water and nutrient availability exert selection on reproductive phenology. Am. J. Bot. 109:1702–1716.

Oakley, C. G. 2013. Small effective size limits performance in a novel environment. Evol. Appl. 6:823–831.

Ochocki, B. M., and T. E. X. Miller. 2017. Rapid evolution of dispersal ability makes biological invasions faster and more variable. Nat. Commun. 8:14315.

Polechová, J., and N. H. Barton. 2015. Limits to adaptation along environmental gradients. Proc. Natl. Acad. Sci. 112:6401–6406.

Prendeville, H. R., K. Barnard-Kubow, C. Dai, B. C. Barringer, and L. F. Galloway. 2013. Clinal variation for only some phenological traits across a species range. Oecologia 173:421–430.

Prior, C. J., N. C. Layman, M. H. Koski, L. F. Galloway, and J. W. Busch. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. Mol. Ecol. 29:4473–4486.

Privé, F., K. Luu, B. J. Vilhjálmsson, and M. G. B. Blum. 2020. Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. Mol. Biol. Evol. 37:2153–2154.

Scheinfeldt, L. B., S. Soi, S. Thompson, A. Ranciaro, D. Woldemeskel, W. Beggs, C. Lambert, J. P. Jarvis, D. Abate, G. Belay, and S. A. Tishkoff. 2012. Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol. 13:R1.

Schweizer, R. M., M. R. Jones, G. S. Bradburd, J. F. Storz, N. R. Senner, C. Wolf, and Z. A. Cheviron. 2021. Broad Concordance in the Spatial Distribution of Adaptive and Neutral Genetic Variation across an Elevational Gradient in Deer Mice. Mol. Biol. Evol. 38:4286–4300.

Sexton, J. P., M. B. Hufford, A. C.Bateman, D. B. Lowry, H. Meimberg, S. Y. Strauss, and K. J. Rice. 2016. Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. Mol. Ecol. 25:911–928.

Spichtig, M., and T. J. Kawecki. 2004. The Maintenance (or Not) of Polygenic Variation by Soft Selection in Heterogeneous Environments. Am. Nat. 164:70–84.

Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph, and M. A. Streisfeld. 2019. Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. PLOS Biol. 17:e3000391.

Tigano, A., and V. L. Friesen. 2016. Genomics of local adaptation with gene flow. Mol. Ecol. 25:2144–2164.

Treurnicht, M., J. Pagel, J. Tonnabel, K. J. Esler, J. A. Slingsby, and F. M. Schurr. 2020. Functional traits explain the Hutchinsonian niches of plant species. Glob. Ecol. Biogeogr. 29:534–545.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic Islands of Speciation in Anopheles gambiae. PLoS Biol. 3:e285.

Tusso, S., B. P. S. Nieuwenhuis, B. Weissensteiner, S. Immler, and J. B. W. Wolf. 2021. Experimental evolution of adaptive divergence under varying degrees of gene flow. Nat. Ecol. Evol. 5:338–349.

Waterhouse, M. D., L. P. Erb, E. A. Beever, and M. A. Russello. 2018. Adaptive population divergence and directional gene flow across steep elevational gradients in a climate-sensitive mammal. Mol. Ecol. 27:2512–2528.

Willi, Y., and J. Van Buskirk. 2019. A Practical Guide to the Study of Distribution Limits. Am. Nat. 193:773–785.

Wong, E. L. Y., B. Nevado, O. G. Osborne, A. S. T. Papadopulos, J. R. Bridle, S. J. Hiscock, and D. A. Filatov. 2020. Strong divergent selection at multiple loci in two closely related species of ragworts adapted to high and low elevations on Mount Etna. Mol. Ecol. 29:394–412.

Wu, Y., and Q. Li. 2017. Phenotypic selection on flowering phenology and pollination efficiency traits between Primula populations with different pollinator assemblages. Ecol. Evol. 7:7599–7608.

Yeaman, S., and S. P. Otto. 2011. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution 65:2123–2129.

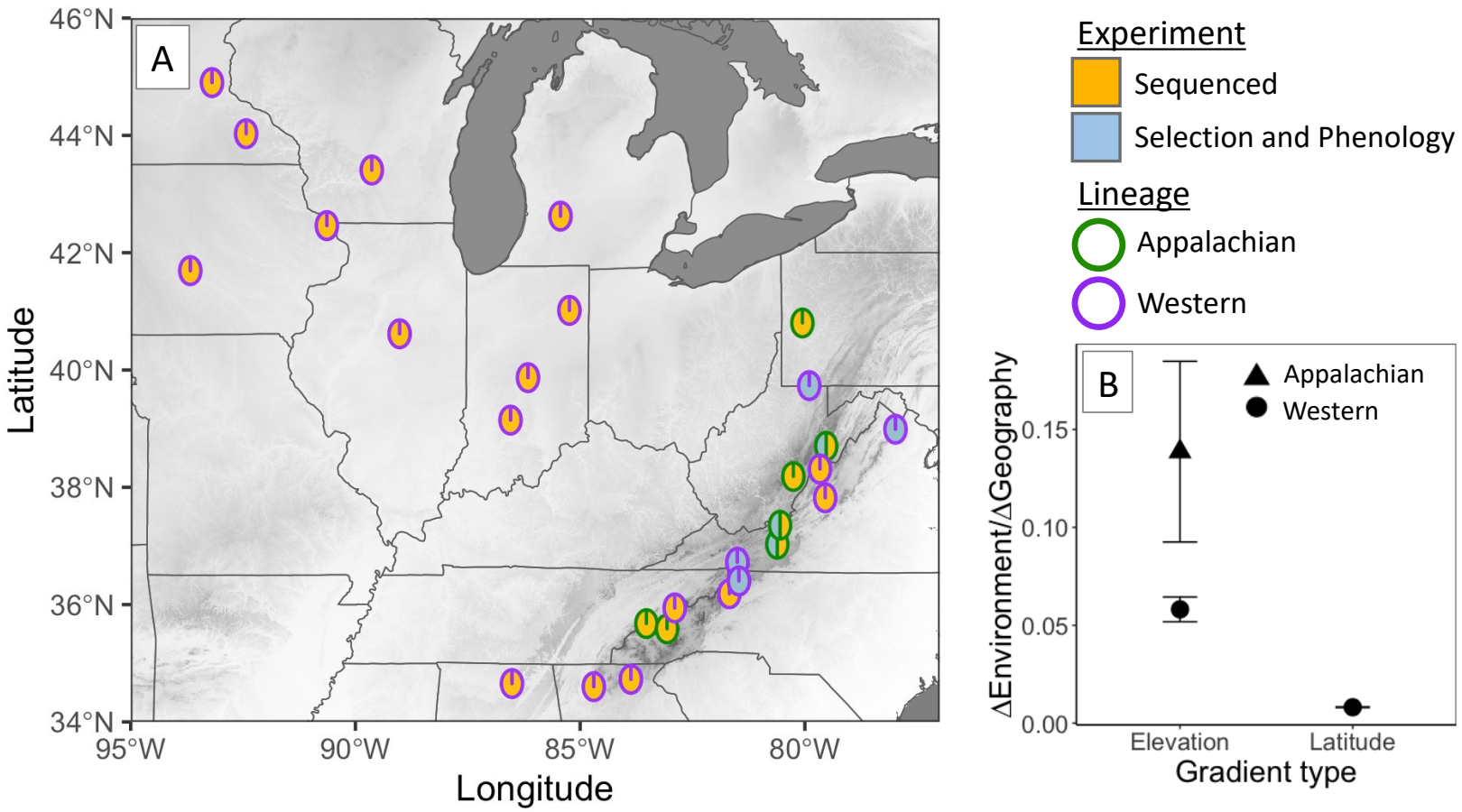


Figure 1) (A) *Campanula americana* populations across an elevational gradient in the Appalachian Mountains. Populations used for the phenotypic selection and phenology study are shown in blue, while sequenced populations are shown in gold. We populations were sampled from three genetic lineages, shown by outline color (green=Appalachian, purple=Western, blue=Eastern). (B) Steepness of environmental gradients along the latitudinal and elevational gradients for genomic selection analyses (*Adapted from Chapter 1*).

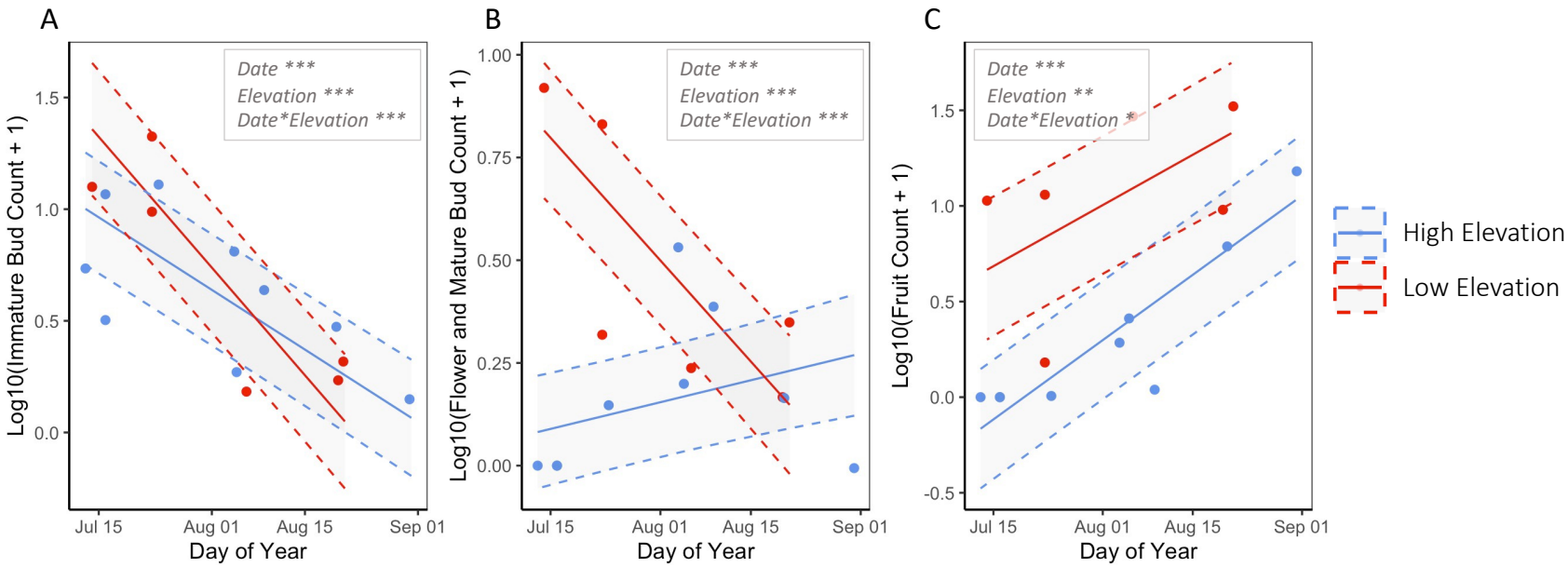


Figure 2) Phenology of populations by range position differed for all reproductive traits analyzed. Low elevation populations produced buds and flowers earlier than high elevation populations. Fruiting was delayed in high elevation populations compared to low elevation populations.

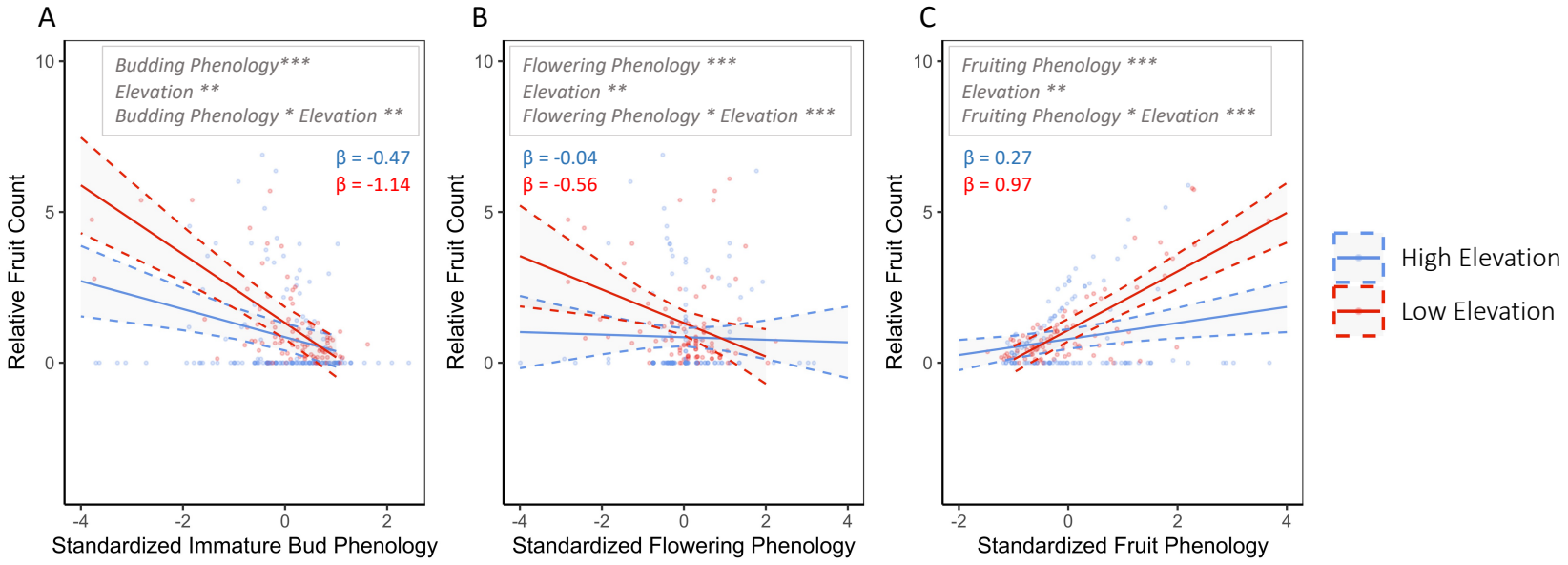


Figure 3) Standardized linear selection gradients for reproductive phenology. Higher values of standardized phenology indicate steeper slopes of individual phenology during the selection window. For budding and flowering phenology, plants which budded and flowered earlier than other individuals in their respective populations tended to have higher fitness. For fruiting phenology, plants which fruited later tended to produce more fruit than other individuals in their respective populations. Generally, selection was weaker in high elevation populations than in low elevation populations.

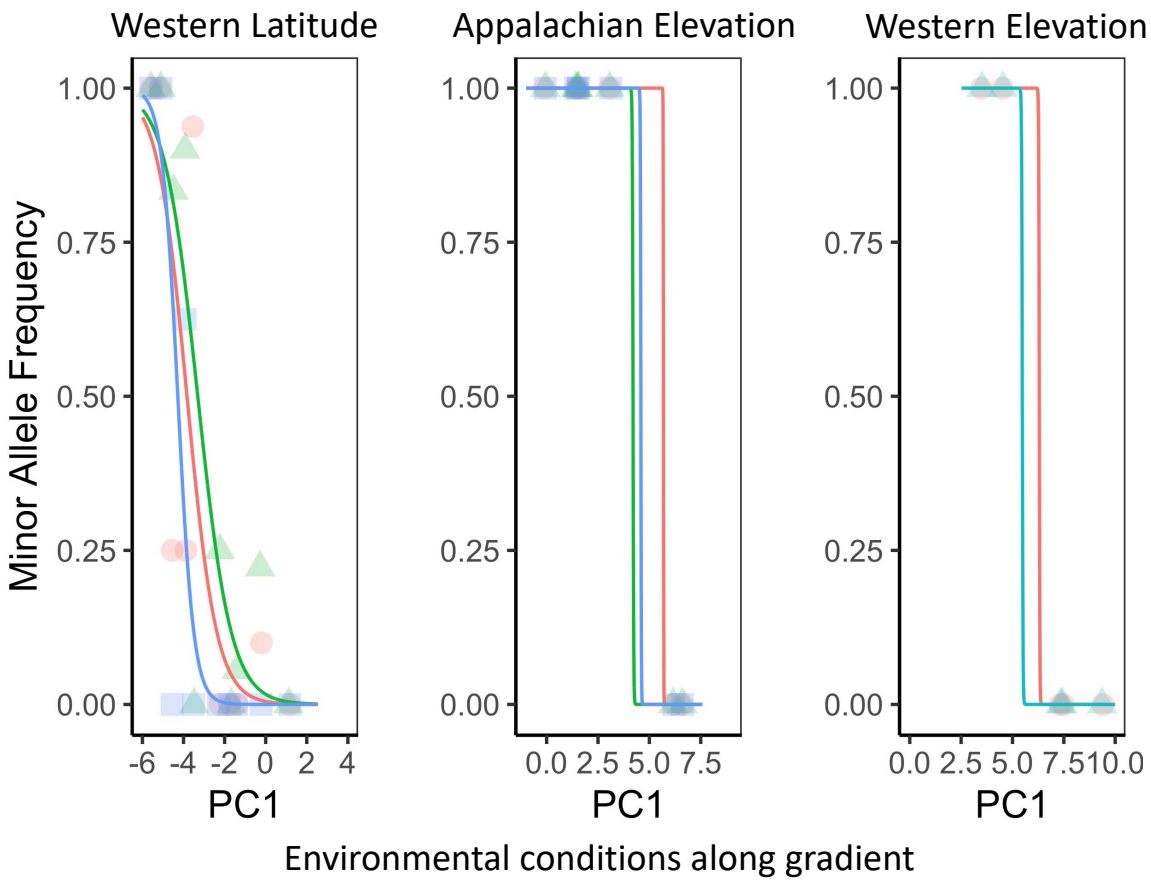
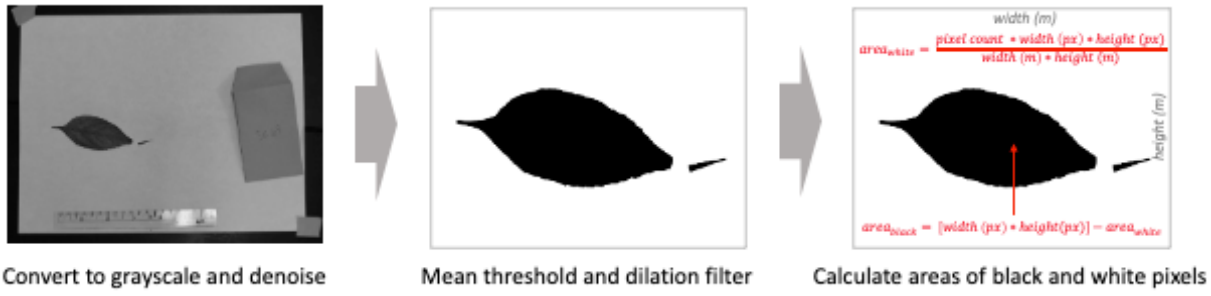


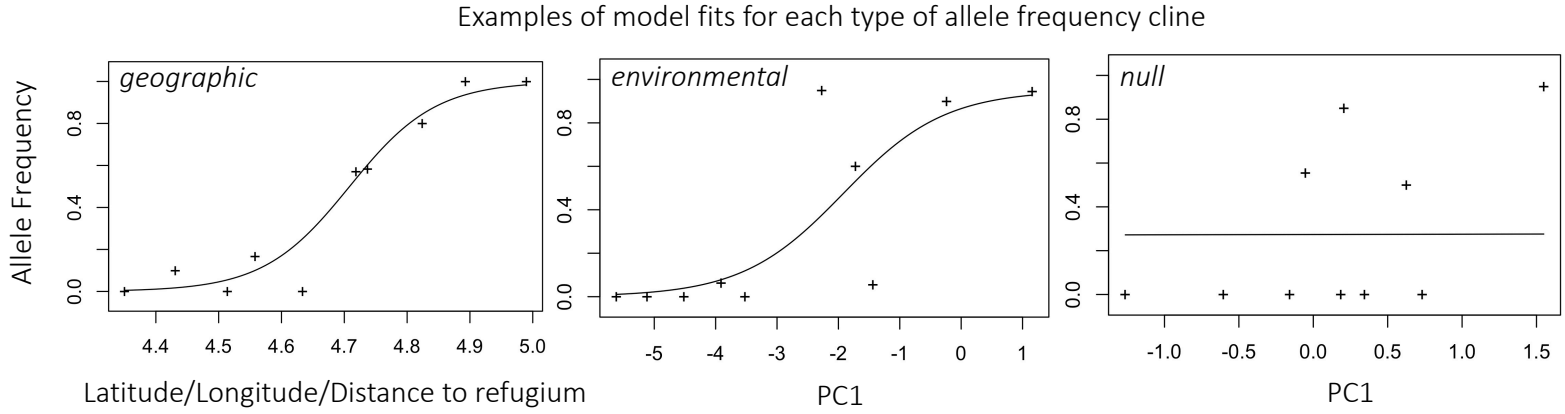
Figure 4) Genomic selection gradients for one-dimensional allele frequency clines. Shapes indicate different adaptive loci. Models shown are for loci where the best fit model was PC1 environmental differentiation and not a null or geographic cline. Loci are shown as separate clines, segregated by lineage. Points, shown by different shapes, are allele frequencies for SNPs associated with adaptive loci in each population. Across lineages and loci, intermediate allele frequencies were notably absent.

SUPPLEMENTARY METHODS

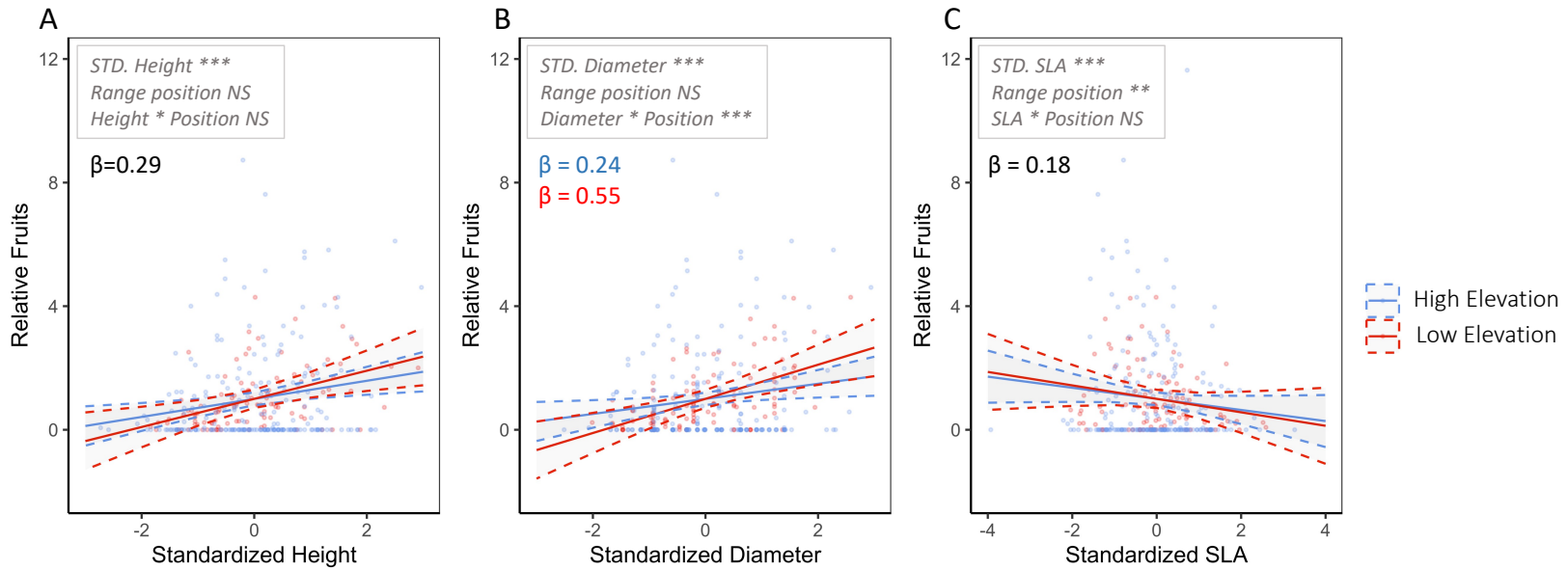
To measure specific leaf area (SLA), we developed a custom python pipeline (SI Fig. 4). In brief, I dried leaf tissue and imaged samples against a white background from a standardized height and with standardized magnification. I then rendered images in grayscale, denoised them with a median filter of 2, and thresholded images via mean values using scikit-image (der Walt 2014). I filled minor holes and pockmarks in images via reconstruction dilation, which recolors pixels based on nearest neighbors, using OpenCV (Bradski & Kaehler, 2000). Finally, I cropped silhouetted images to a known size and tabulated the number of black and white pixels in metric units. For the first five images processed, I verified the accuracy of the pipeline using ImageJ (Schneider et al. 2012). I weighed leaf tissue to nearest 4th digit using a digital scale, and calculated SLA as the area of the leaf in mm^2 per unit of leaf mass in mg^{-1}.



SI Figure 1) Process of calculating specific leaf area (SLA) using custom python script in OpenCV and scikit-image (Bradski & Kaehler, 2000; Van der Walt et al., 2014). I took images with a scale included, then converted images to grayscale and denoised them. I then applied a mean threshold, which rendered images as black and white silhouettes, and applied a dilation filter to fill gaps. Finally, I took the known area of the cropped region and converted the tabulated counts of white and black pixels to leaf area per unit mass.



SI Figure 2) Examples of models where (A) a geographic allele frequency cline is the best fit, (B) environmental allele frequency cline is the best fit, and (C) a null cline is the best fit. We compared models using AICc.



SI Figure 3) Standardized linear selection gradients on vegetative traits across elevational populations. Selection was significant for all traits, favoring taller plants with wider stems and less expensive leaves. The strength of selection (*β*) did not differ between high and low elevation populations for height or SLA. Selection for wider stems was stronger for low elevation populations than high elevation populations.

population	gradient	lineage	elevation group	latitude	longitude	elevation	sequenced	selection
NC91	elevation	Appalachian	high	35.59	-83.07	1456	1	
VA73	elevation	Appalachian	high	37.35	-80.55	1044	1	1
WV7	elevation	Appalachian	high	38.18	-80.26	1085	1	
WV4	elevation	Appalachian	high	38.70	-79.52	1295	1	1
TN92	elevation	Appalachian	low	35.68	-83.53	480	1	
VA5	elevation	Appalachian	low	37.28	-80.61	496	1	1
PA104	elevation	Appalachian	low	40.80	-80.05	289	1	
MJ	elevation	Eastern	high	36.40	-81.47	1275		1
VA3	elevation	Eastern	low	38.99	-77.98	429		1
GA2	elevation	Western	high	34.73	-83.87	1078	1	
NC16	elevation	Western	high	36.19	-81.68	1031	1	
GH	elevation	Western	high	36.62	-81.50	1414		1
GA1	elevation	Western	low	34.60	-84.70	213	1	
ALBG	elevation	Western	low	34.65	-86.52	396	1	
NC107	elevation	Western	low	35.94	-82.90	382	1	
PM	elevation	Western	low	39.73	-79.90	244		1
MI2	latitude	Western	high	42.62	-85.44	291	1	
MN117	latitude	Western	high	44.90	-93.19	213	1	
WI4	latitude	Western	high	43.41	-89.64	260	1	
IN7	latitude	Western	low	39.87	-86.16	228	1	
IA17	latitude	Western	mid	42.46	-90.64	205	1	
IN2	latitude	Western	mid	41.02	-85.24	231	1	
MN8	latitude	Western	mid	44.03	-92.43	308	1	
IA12	latitude	Western	low	41.69	-93.67	246	1	
IL10	latitude	Western	low	40.62	-89.02	229	1	
IN5	latitude	Western	low	39.15	-86.55	234	1	

SI Table 1) Populations used in genomic (sequenced) and field-based phenotypic analyses.

Site name	elevation group	Population size	Initiation day of year	Visit #1 day of year	Visit #2 day of year	Visit #3 day of year
VA73	high	51	152	193	216	
Mount Jefferson (MJ)	high	36	158	196	220	
Grayson Highlands	high	65	156	196	215	242
Spruce Knob (SK)	high	94	171	204	231	
Eggleston (EGG)	low	33	153	194	217	
Point Marion (PM)	low	24	170	203	231	
Sky Meadows (SM)	low	62	168	203	232	

SI Table 2) Selection experiment sites and timing of visits given in number of days since beginning of year. Traits were measured at the of initiation of the selection window and visited upwards of three additional times.

Lineage	Gradient	Scaffold	Contig	Position	SNP ID
Western	latitude	7	9	23496658	393604:23:+
		7	9	41638283	411784:43:-
		12	12	175705	634989:114:-
		14	9	22389472	755658:190:-
		27	9	167699	1395277:53:-
Western	elevation	3	7	39745455	177964:29:+
		3	7	60283746	190899:24:+
		7	9	66425379	436714:20:-
		12	12	14510250	651687:11:+
		13	2	30888183	716925:14:+
		19	4	47347597	1034649:6:-
		20	14	51911654	1075997:47:-
		20	14	51911711	1075996:15:+
		24	3	21392065	1274056:187:-
Appalachian	elevation	2	4	763290	68304:141:+
		7	9	66420029	436704:13:+
		7	9	66425379	436714:20:-
		7	9	66489761	436786:61:-
		12	12	54625834	681469:652:-

SI Table 3) Putatively adaptive loci identified by both BayeScEnv and PCAdapt. Population-level allele frequencies for each SNP informed clinal genomic selection analyses.

Trait	Range edge	Range core	p-value
Height (cm)	57.16	66.25	0.1000
Diameter (mm)	5.17	4.26	<0.0001
SLA (mm^2/mg^-1)	76.64	56.11	0.0002

SI Table 4) Mean values of vegetative traits for high and low elevation populations and significance of post-hoc comparisons of population differences lumped by elevation.

Model	Trait	elevation group	selection coefficient	Term	Chi^2	p-value
vegetative selection	height	high	0.29	Height	356.99	**<0.0001**
		low	0.46	Elevation	1.21	0.2709
				Height*elevation	0.06	0.8013
	stem diameter	high	0.24	Stem diameter	189.58	**<0.0001**
		low	0.55	Elevation	1.10	0.2937
				Diameter*elevation	98.27	**<0.0001**
	SLA	high	-0.18	SLA	61.41	**<0.0001**
		low	-0.22	Elevation	8.54	**0.0035**
				SLA*elevation	3.01	0.0828
phenology selection	immature buds	high	-0.47	Slope	258.57	**<0.0001**
		low	-1.14	Elevation	6.74	**0.0094**
				Date*elevation	13.04	**0.0003**
	flowers and mature buds	high	-0.04	Slope	26.24	**<0.0001**
		low	-0.56	Elevation	7.28	**0.0070**
				Date*elevation	34.53	**<0.0001**
	fruits	high	0.27	Slope	117.22	**<0.0001**
		low	0.97	Elevation	9.71	**0.0018**
				Date*elevation	126.48	**<0.0001**
phenology	immature buds			Date	258.70	**<0.0001**
				Elevation	46.49	**<0.0001**
				Date*elevation	51.12	**<0.0001**
	flowers and mature buds			Date	14.46	**0.0001**
				Elevation	153.83	**<0.0001**
				Date*elevation	141.31	**<0.0001**
	fruits			Date	380.74	**<0.0001**
				Elevation	12.80	**0.0003**
				Date*elevation	6.27	**0.0123**

SI Table 5) Models either evaluate selection on vegetative traits (height, stem diameter, SLA) or reproductive phenology (immature budding, flowering, fruiting). Elevation position specifically refers whether populations are at high elevation or low elevation. Selection differentials are the elevation group slopes (β) for high and low elevation populations. Significant p-values (≤0.05) are bolded. Phenology and phenological selection models correspond to Fig. 2, 3, and SI Fig. 3.

CHAPTER 3:

Fitness consequences of genetic load are modified by environmental conditions along a shallow environmental gradient.

ABSTRACT

Genetic load is expected to accumulate via serial founder effects during range expansion and can reduce population fitness and impede adaptation. Along shallow environmental gradients, where environments change slowly relative to space, genetic load may accumulate rapidly from serial founder events and accelerated range expansion. In contrast, along steep environmental gradients, range expansion may occur more slowly due to high costs of dispersal along the gradient, allowing migration from the range core to reach the range edge, and ameliorating genetic load in range edge populations before the range further expands Furthermore, in ecologically marginal habitat, genetic load may produce greater costs to fitness. Here, I evaluated both the accumulation of load over range space and its influence on fitness in common gardens across separate elevational and latitudinal environmental gradients. I then modeled patterns of population fitness across common gardens as a function of genetic load and environmental conditions at common garden sites. I found genetic load accumulated along the shallower latitudinal environmental gradient, and that load was associated with reduced fitness when populations were under more stressful environmental conditions. In contrast, populations along the steep elevational gradient had more genetic load than latitudinal populations, but I did not find an interaction effect of stressful environmental conditions and the amount of genetic load. My results suggest that environment-dependent expression of genetic load may restrict range expansion, particularly along shallow environmental gradients and when range limits coincide with ecological marginality.

INTRODUCTION

Range expansions occur through successive waves of colonization to habitat beyond current range limits. Colonizations often occur with small numbers of founding individuals. Limited population size carries the potential for deleterious genetic consequences, including the accumulation of genetic load via drift (King 1966; Whitlock et al. 2000; Slatkin and Excoffier 2012; Peischl et al. 2013). Genetic load can accrue during serial founder events, resulting in the accumulation of load in the direction of expansion (Koski et al. 2019) and increased signatures of genetic drift near range edges (Willi et al. 2018; Foutel-Rodier and Etheridge 2020). Genetic load can diminish population fitness (Perrier et al. 2020) and reduce population growth rates (Willi et al. 2018), making it more difficult for populations to adapt and colonize. Theoretical models posit that environmental gradients, the rate that environments change in space, may influence the accumulation and consequences of genetic load (Gilbert et al. 2017). Yet, few empirical studies have attempted to validate these models. Furthermore, while genetic load is known to be consequential for population fitness, it is unclear how environmental conditions, such as the ecological marginality of habitat, influence the fitness consequences of genetic load.

As environmental gradients become steeper, theory suggests that the accumulation of genetic load should slow. Slower accumulation of load results from high fitness costs of dispersal to novel habitat along steeper gradients, which can slow the rate of range expansion and allow more time for gene flow from the range core to percolate toward the range edge (Gilbert et al. 2017). Gene flow toward expanding range edges from populations closer to the range core may help ameliorate the accumulation of genetic load associated with founder effects (Slatkin and Excoffier 2012), improving population fitness, and the capacity to respond to selection (Alleaume-

Benharira et al. 2006). As a result, genetic load is expected to accumulate more slowly over steep environmental gradients than shallow gradients.

While the interplay of adaptation and genetic load during range expansions has been the subject of extensive theoretical modeling efforts (Peischl et al. 2013; Gilbert et al. 2017; Foutel-Rodier and Etheridge 2020), experimental evidence remains lean. Experimental work has predominantly focused on the accumulation of load over shallow environmental gradients, like latitude (Willi et al. 2018; Koski et al. 2019; Perrier et al. 2020), rather than steep environmental gradients, like elevation. Additionally, while deleterious effects of genetic load have been well documented (Willi 2013; Perrier et al. 2020), it is unclear whether and how environmental conditions can modify the influence of genetic load on reproductive output and population fitness (but see: Parsons 1971; Perrier et al. 2022). Whether the fitness consequences of genetic load are modified by the environment is important as it may impose constraints on adaptation that depend on environmental conditions. Inbreeding depression, which results from segregating genetic load expressed via mating with closely related individuals, has been shown to increase when environmental conditions are more stressful (Armbruster and Reed 2005; Galloway and Etterson 2007). Thus, if expression of genetic load accumulated during range expansion is affected by the environment similarly to that of the segregating load underlying inbreeding depression, environmental conditions may be key to understanding adaptation during range expansion. For example, increased expression of genetic load with ecological marginality may depress population fitness near range edges and impede further range expansion.

Here, I used *Campanula americana,* a native North American wildflower, to experimentally assess how environmental gradients influence the accumulation of genetic load during range expansion and how genetic load influences adaptation and fitness in different

environmental conditions. First, I estimated genetic load near range limits along a steep environmental gradient and a shallow latitudinal gradient using heterosis of within- and between-population crosses. I then modeled population fitness in common gardens, estimated in Chapter 1 as a function of genetic load and environmental conditions. I used this framework to address the following questions: (1) How does the steepness of environmental gradients affect the distribution and accumulation of genetic load? (2) How do interactions of genetic load and the environment affect patterns of local adaptation and fitness during range expansion?

METHODS

Study system

Campanula americana is an outcrossing monocarpic herb native to the eastern United States. The species is generally divided into three genetic lineages, with the Western and Appalachian lineages composing the majority of populations and territory (Barnard-Kubow et al. 2015). The Western lineage is largely restricted to the Midwestern and Southern U.S., and the Appalachian lineage to the extent of the Appalachian Mountains (Barnard-Kubow et al. 2015). Since the last glacial maximum (~22kya), the Western lineage has expanded from a mid-latitude refugium located in Kentucky (Koski et al. 2019; Prior et al. 2020; Chapter 5; Appendix 1), while the Appalachian lineage has largely remained confined to the Appalachian Mountains (Barnard-Kubow et al. 2015). Previous work in the species identified a positive association between deleterious genetic load (i.e., drift load) and distance from the mid-latitude refugium (Koski et al. 2019), suggesting that northward range expansion in the Western lineage has progressed through serial colonization events with conspicuous signatures of genetic drift. However, little is known about genetic load in other lineages and regions of the species' range.

Methodological framework

To understand how genetic load evolves across heterogeneous landscapes, I first sampled populations in replicate transects (Fig. 1) across a shallow latitudinal environmental gradient and a steeper elevational environmental gradient. To estimate genetic load, I created between-population F_1 hybrids to determine the gain in fitness from masking deleterious alleles in hybrids relative to within-population crosses (i.e., heterosis). I then determined whether genetic load was associated with position relative to the range edge (core/mid/edge) along latitudinal and elevational gradients. To investigate the relationship between fitness and genetic load, I planted common gardens across the leading elevational and latitudinal range margins and calculated relative fitness of individuals and populations. I then modeled the relative fitness of populations within each common garden as a function of genetic load and the environment. I estimated environmental conditions in two ways. First, I compared the difference in environmental conditions between a population's origin and each respective common garden using PCA. Second, I extracted habitat suitability estimates for each common garden location from a species distribution model.

Population selection and greenhouse crosses

I sampled populations in portions of the range that represent post-glacial expansion in *C. americana* across a steep and shallow environmental gradient (Fig. 1C). I selected range-edge populations ('*edge*') based on proximity to the northern range edge in the Midwest (latitudinal range limit), and to the high elevation range edge in Appalachia (elevational range limit). I also selected populations on the basis of proximity to their respective expansion origins. Populations nearest to the expansion origin ('*core'*), reflect long-established range space, while populations

intermediate (‘*mid’*) to the range core, nearer the range edge, reflect more recent range expansion. I made collections of wild *C. americana* fruits in the fall of 2019 and 2020. In the upper Midwest, I sampled nine populations arranged in three replicate transects (Fig. 1A; SI Table 1; *3 core, 3 mid, 3 edge*). In Appalachia, I sampled 12 populations: six from the Appalachian genetic lineage in three replicate transects (*3 core, 3 edge*), and six from the Western genetic lineage in three replicate transects (Fig. 1A; *3 core, 3 edge*). I germinated field-collected seeds (~13 *plants/population*) for each population according to protocols outlined in Chapter 1.

I created between-population hybrids to estimate heterosis. I crossed individuals within and between populations. All crosses were unidirectional (Fig. 1B), where populations served as either paternal donors or maternal recipients. In total, I used 21 populations, producing 21 sets of within-population crosses and 22 sets of between-population crosses (lines in Fig. 1A; SI Table 2). I performed 15 unique crosses, i.e. genetically distinct, within each population and for each between-population combination, though fewer when field seed collections were limited (SI Table 2). I segregated the elevational crossing design by genetic lineage (*Appalachian/Western*) so as not to confound large-scale genetic differentiation with that accumulated during expansion.

To perform crosses, I emasculated maternal recipient flowers of all pollen when in male phase. The following day, I selected male-phase donor flowers and applied their pollen to the stigmas of recipient female-phase flowers. If the recipient was still in male phase, I delayed the cross until the following day. I used plants only once as paternal donors and once as maternal recipients for each cross type (*within/ between*). I collected fruits when mature, roughly five weeks after performing the cross, then stored fruits at 4°C to maintain seed quality.

Estimating genetic load

I measured fitness components on within- and between-population crosses. I sowed seeds in two cohorts (n=364, n=422) approximately one week apart. All between- and within-population crosses were present in both cohorts. In total, I sowed 786 cells in germination trays representing 42 within- and between-population crosses (~19 *individuals/cross*). I sowed five seeds per cell in each germination tray for cohort one, and three seeds per cell for cohort two. I germinated seeds and vernalized rosettes using identical protocols as the parental generation. Each cohort spent six weeks in growth chambers where I recorded the proportion germination per cell biweekly. I then thinned seedlings to one plant per cell and moved them to a cold room at 4°C to vernalize for seven weeks. I then moved plants to the greenhouse, transplanted individuals to separate cone-tainers, and tracked the number of days to first flower biweekly for seven weeks. Once each plant began flowering, I recorded the number of flowers on the plant weekly for four weeks and summed these to serve as an estimate of reproductive output. To determine if reproductive output differed between cohorts, I tested reproductive output as a function of cohort, cross type (*between/within*) and the environmental gradient (*latitude/elevation*; *reproductive output = cohort + cross type + env. gradient + cohort * cross type + cohort * env. gradient + cross type * env. gradient + cohort * cross type * env. gradient + replicate*(*cross*(*cross type*))). Cohort was not significant as a fixed effect or in any interaction (SI Table 3). Consequently, data from both cohorts was combined in the calculation of heterosis.

I calculated heterosis for each population for germination, reproductive output, and lifetime fitness. For lifetime fitness, I multiplied the germination proportion per cell (0-1) by the total reproductive output (≥0). I first computed averages per cross to calculate population-level estimates of heterosis. Then, for each fitness component (germination/reproductive output/lifetime

fitness) and each cross, I took the average value for the between-population cross ($B_{i,j}$) and subtracted it from the average mid-parent within-population value (W_i, W_j), before dividing it by the mid-parent value (W_i, W_j). I then found heterosis for both between-population crosses that populations were involved in ($B_{i,j}$ and $B_{k,i}$), and calculated the average value of heterosis for each population by averaging the values, as shown in the following equation:

$$heterosis\ population_i = \mu\left(\frac{\mu(B_{i,j}) - \mu(W_i, W_j)}{\mu(W_i, W_j)}, \frac{\mu(B_{k,i}) - \mu(W_k, W_i)}{\mu(W_k, W_i)}\right)$$

Positive values of heterosis indicate that between-population crosses are more fit than expected by mid-parent fitness. Consequently, heterosis results from the masking of deleterious recessive alleles fixed in one parental population that are not fixed in the other. I used heterosis as an index of genetic load for each population.

Distance of populations to the range edge and mid-latitude refugium

I estimated the distance of each population to the range edge and to the mid-latitude refugium. To generate estimates of population distance to the range edge, I fit a spatial hull around the range of *C. americana*. First, I downloaded iNaturalist research-grade observations of *C. americana* up to 2021 (iNaturalist Community n.d.; n=7,781). I then fit a concave polygon using the alphahull R package (Pateiro-Lopez and Rodriguez-Casal 2022). I calculated the distance of populations to the closest range edge as the distance between the polygon edge and population locations using the dist2Line function from the geosphere package (Hijmans 2021). I calculated the distance of populations to the mid-latitude refugium in Kentucky as the distance between populations and the approximate coordinates of the refugium obtained from Koski et al. (2019). I calculated distances using haversine geodesic distances and the geodist function from the geodist package (Padgham 2021). To estimate the influence of range expansion on the accumulation of genetic load, I fit

Gaussian linear models of population genetic load as a function of distance to range edge and of distance to the mid-latitude refugium.

Fitness estimates from common gardens

To understand how genetic load influences the fitness of populations along an expansion front, I used estimates of fitness previously measured in common gardens across latitudinal and elevational environmental gradients (Chapter 1). Briefly, I constructed common gardens near the home sites of populations used in the between- and within-population crosses. These included eight gardens that traversed the latitudinal environmental gradient. For each garden, I planted two individuals from each of the nine populations into each of 10 blocks, totaling 180 plants/garden. Across the elevational environmental gradient, I constructed four gardens. For each garden, I planted two individuals from each of the six populations into each of 10 blocks (2 plants/block/population; n=120). I placed vernalized plants grown from greenhouse produced seeds in each garden. For details on garden construction, see Chapter 1. I assessed the total fruit count per plant as an estimate of reproduction. Plants that died after transplanting into gardens were assigned a reproduction value of zero. Reproduction was highly variable between gardens. To standardize performance across gardens and estimate local adaptation, I relativized reproduction within each garden by dividing individual fitness by the mean fitness of its garden.

Statistical analysis

Genetic load and range position

I analyzed the relationship between population genetic load (i.e., lifetime heterosis) and range position (*core/mid/edge*) for latitudinal and elevational gradients to estimate how genetic load

accumulates along environmental gradients that differ in steepness. First, I tested whether genetic load differed significantly from zero for each lineage (*Appalachian/Western*) and gradient (*latitude/elevation*) separately, ignoring population range position. I then examined whether genetic load increased across range space using Gaussian ANOVAs, treating average genetic load calculated from each life stage (*germination, reproductive output, lifetime*) as a function of categorical range position (*edge/mid/core*; *genetic load = range position + transect*(*range position*)). For the elevational gradient, I included an additional categorical predictor of lineage (*genetic load = range position + lineage + range position * lineage + transect*(*range position*)). I also conducted linear regressions with continuous predictors of latitude and distance to the range edge for populations distributed across latitudinal environmental gradients. I weighted linear regressions as needed to account for the influence of outliers and heteroscedasticity. I calculated regression weights as the inverse of the squared fitted values of a regression between the absolute values of residuals from the naïve linear model (i.e., unweighted) and its fitted values. I did not analyze elevation as a continuous predictor, as elevational range limits vary across the Appalachian Mountains making values of the population's elevation unreliable estimators of distance to range edges.

Genetic load and population fitness

To better understand how environment affects the capacity of genetic load to constrain population fitness near range edges, I assessed fitness in common gardens as a function of genetic load and of environmental conditions. Environmental conditions were assessed in two manners: comparing displacement of environmental conditions between a population's origin and the common garden site, and in the context of range-wide habitat suitability. To determine environmental displacement

I extracted all bioclimatic factors from the WorldClim 2 data set (n=19; Fick and Hijmans 2017) for the locations of all iNaturalist research-grade *C. americana* observations (n=7,781; iNaturalist Community) and performed a PCA. I predicted coordinates for PC1 (62.18% variance; SI Fig. 1) and PC2 (15.44% variance) for all locations of common gardens and population origins. I then estimated environmental displacement by finding the difference between the site of population origin and the common garden site for PC1 and PC2 ($\Delta PC_{origin\text{-}garden}$) for each population in each garden.

I then modeled the effects of genetic load and environmental displacement on population relative fitness within common gardens via mixed effect Gaussian linear regression. Models regressed relative fitness (Chapter 1) of individuals in common garden sites against population-level cumulative genetic load (i.e., lifetime heterosis) and $\Delta PC_{origin\text{-}garden}$. I added block and garden as nested random effect terms ($relative\ fitness = PC1_{origin\text{-}garden} + genetic\ load + genetic\ load * PC1_{origin\text{-}garden} + block(garden) + population(genetic\ load)$) using the lme4 package in R 4.2.0 (Bates et al. 2014; R Core Team 2022). I constructed models for populations distributed across latitude and elevation separately and did not combine lineages within elevational gradients. To determine significance of model terms, I ran type III ANOVAs on all models using the car package (Fox and Weisberg 2019). To visualize model interactions, I used the sjPlot (Lüdecke 2023), ggplot2 (Wickham 2016), and visreg packages (Breheny and Burchett 2017).

An additional set of Gaussian linear mixed effect models assessed whether habitat suitability at each common garden site explained relative fitness and whether an interaction between the site's suitability and genetic load significantly contributed to variance in relative performance. I generated estimates of habitat suitability for common garden sites from a species distribution model (SDM; Appendix 1). I extracted habitat suitability estimates of common garden

sites as a decimal score between 0 and 1 from the SDM. Models followed identical structure to those detailed above (*relative fitness = genetic load + habitat suitability + genetic load * habitat suitability + block*(*garden*(*habitat suitability*)) + *population*(*genetic load*)). I performed type III ANOVAs to determine significance of model terms. Western-lineage elevational populations were rank-deficient for models both of environmental suitability and of displacement and so were dropped from final analysis.

RESULTS

Populations across environmental gradients displayed genetic load, though the degree of accumulation varied by gradient and lineage. Western-lineage populations sampled along the shallow latitudinal environmental gradient displayed modest lifetime heterosis, indicating genetic load (μ=0.14, range=-0.01–0.22). Appalachian-lineage populations sampled along the steep elevational environmental gradient had somewhat greater mean lifetime heterosis (μ=0.26, range=0.17–0.49). In contrast, Western-lineage populations along a steep elevational gradient had mean negative estimates of lifetime heterosis (μ=-0.07, range=-0.21–0.10), suggesting modest outbreeding depression (Fig. 2).

Genetic load and range position

Estimates of genetic load derived from lifetime heterosis did not increase toward the range edge along either environmental gradient (Fig. 2; SI Fig. 2). For populations distributed across elevation, lifetime genetic load differed somewhat between lineages (p-value=0.014) but not by range position (p-value=0.24), though the interaction was significant (p-value=0.027). The significant interaction of lineage and range position likely reflects differential expression of load across life

stages and outbreeding depression among Western-lineage elevation populations (Fig. 2). For estimates of genetic load based on germination rates, edge populations had significantly greater load on average than core populations (p-value=0.001), largely due to greater load in range edge Appalachian populations (interaction p-value=0.01; Fig. 2). Negative estimates of germination genetic load for Western-lineage elevational populations indicate outbreeding depression. For estimates of genetic load based on flowering, elevational populations near the range core carried significantly more genetic load than did high elevation, range edge populations (range position p-value=0.001; Fig. 2). Lineage was not a significant predictor of genetic load associated with flowering. For Western-lineage latitudinal populations, estimates of genetic load did not vary across range positions for either life stage (Fig. 2), but was near significant for the lifetime heterosis metric (p-value=0.057).

Genetic load has been previously shown to correlate with distance from the mid-latitude glacial refugium (Koski et al. 2019). Here, genetic load was near significant when regressed against distance to range edge (SI Fig. 3A; p-value=0.08), though distance to the mid-latitude refugium was not (p-value= 0.52; SI Fig. 3B). Latitude, when weighted and treated continuously, did not explain variance in lifetime genetic load (p-value=0.11; SI Fig. 2). For elevational populations, accumulation of genetic load was unrelated to continuous measures of elevation (p-value=0.18; SI Fig. 2).

Genetic load and population fitness

The effect of genetic load on fitness depended on the environmental conditions that populations experienced within common gardens. Relative fitness within common gardens was explained by interactions of genetic load and climatic differentiation between population origin and common

garden sites for $\Delta PC1_{origin\text{-}garden}$ (Fig. 3; SI Table 4; SI Fig. 4). In particular, where environmental conditions at latitudinal garden sites were more temperate (more seasonal, cooler winters, less precipitation), i.e. potentially more stressful, than locations of population origin, populations with low genetic load outperformed populations with high genetic load. Where garden site conditions were more subtropical (less seasonal, warmer winters, more precipitation), relative fitness did not vary with genetic load, suggesting that load does not produce high fitness costs when plants experience benign conditions than at origins. In concert with this, genetic load more strongly constrained performance in less suitable habitats (Fig. 3B). More specifically, in models incorporating habitat suitability (Fig. 3), the interactions of genetic load and habitat suitability were significant, indicating that ecological marginality of common garden sites influences the fitness consequences of genetic load. Variance in relative fitness was not significantly explained by $\Delta PC2_{origin\text{-}garden}$, nor its interactions with genetic load (SI Table 4) though it is worth noting that PC2 explained substantively less variance in bioclimatic conditions across the species' range (15.44% variance; SI Fig. 1) than PC1 (62.18% variance).

Across elevational common gardens, environmental conditions did not modify the effect of genetic load to constrain local adaptation or fitness. Within common gardens, relative fitness of Appalachian-lineage populations was not explained by the interaction of $\Delta PC1_{origin\text{-}garden}$ and genetic load (SI Table 4). In contrast, $\Delta PC2_{origin\text{-}garden}$ was significant in models, though neither genetic load nor the interaction between ΔPC2 and genetic load were significant (SI Table 4; SI Fig. 4). Finally, habitat suitability, which estimated ecological marginality of common garden sites, was not significant in any models (SI Table 4).

DISCUSSION

Genetic load across environmental gradients

During range expansions, range-edge populations are the product of serial colonizations and may be subject to strong genetic drift (Peischl et al. 2013), resulting in elevated genetic load associated with range limits (Willi et al. 2018) and distance from the origin of range expansion (Koski et al. 2019). Previous work in *C. americana* identified an association between genetic load and distance from the origin of latitudinal range expansion in a near range-wide sample of populations (Koski et al. 2019). Here, lifetime genetic load was marginally explained by range position and distance from the range edge along the latitudinal gradient (Fig. 2; SI Fig. 3), replicating earlier findings. These patterns also indicate that recent range expansion has resulted in modest accumulation of genetic load near range limits, potentially related to more recent founding effects.

Accumulation of genetic load was similar across the steep and the shallow environmental gradients, in contrast to predictions from theoretical models (Gilbert et al. 2017). Sampling of populations near less genetically diverse range limits may have limited detection of genetic load among populations using heterosis. *C. americana* populations near the northern latitudinal range edge are less genetically diverse (Koski et al., 2019, Scott et al., *in prep*) and likely descend from a single mid-latitude refugium (Koski et al. 2019; Prior et al. 2020). Reduced genetic diversity may prevent deleterious loci from being masked in between-population crosses as populations are genetically similar, resulting in downward-bias in estimates of genetic load (-0.01-0.22). In contrast, high elevation Appalachian populations had greater average genetic load (0.17-0.49), concordant with greater genetic variation found in the lineage and a more complex history of postglacial range expansion seeded by multiple microrefugia (Barnard-Kubow et al., 2015;

Chapter 5; Appendix 1). Thus, my ability to detect differences in genetic load using heterosis may be much greater among elevational populations than latitudinal.

Accumulation of genetic load can be moderated during range expansions via episodic purging. While theoretical models have posited that gene flow most effectively purges genetic load along steep environmental gradients (Gilbert et al. 2017), other factors may influence the accrual of load independent of the environment. For example, cycles of inbreeding and outbreeding induced via intermittent gene flow events and post-colonization inbreeding have both been shown to efficiently purge genetic load accumulated during simulated range expansions of the invasive plant *Brachypodium sylvaticum* (Marchini et al. 2016). Selfing ability in the Western lineage is significantly associated with distance from a mid-latitude glacial refugium (Koski et al. 2019). In addition, inbreeding depression declines toward latitudinal range limits (Barringer et al. 2012). Both suggest that genetic load may be actively purged during range expansion. Hence, genetic load may be moderated along the shallow latitudinal environmental gradient by cyclic inbreeding and outbreeding, overriding theoretical expectations of greater load over shallow gradients.

Fitness consequences of genetic load across environmental gradients

Genetic load influenced population fitness only along shallow environmental gradients and where differences in environmental conditions were greatest, suggesting that fitness consequences of load may be exacerbated by ecological marginality. Specifically, there were few fitness costs associated with genetic load in more benign environments (i.e., more suitable habitat, warmer winters, more precipitation; Fig. 3); while more stressful environments yielded significant fitness costs associated with genetic load (i.e., less suitable habitat, colder winters, and less precipitation; Fig. 3). These patterns suggest a synergistic effect of genetic load and environment, where expression

of load is compounded by deleterious environmental conditions to produce stronger constraints of fitness in ecologically marginal habitat than either factor in isolation (i.e., environmental conditions or genetic load). Previous work in other systems ranging from plants (*Arabidopsis lyrata*; Perrier et al. 2022) to insects (*Drosophila melanogaster*; Parsons 1959) have found similar patterns of environment-dependent expression of genetic load. However, trends have generally either been marginal or considered only single elements of environmental variation, like cold temperature resistance (Parsons 1971).

Dynamics of environment-dependent expression of genetic load may be especially relevant to processes of adaptation at or near species range limits. Geographic limits of species' ranges often align with fundamental ecological niche limits (Hargreaves et al. 2014), suggesting that habitat beyond range limits represent worsening environmental conditions for the species. Where habitat degrades in quality beyond range limits, high genetic load in range edge populations may prevent successful colonization beyond the range edge by compounding the fitness consequences of genetic load via environmental stress, potentially inducing 'mutational meltdown' and source-sink dynamics (Gabriel et al. 1993; Willi 2013; Willi and Van Buskirk 2019). Together, fitness may be strongly limited by interactions of environment and genetic load, impeding range expansion, and potentially leading to the formation of stable range limits.

LITERATURE CITED

Alleaume-Benharira, M., I. R. Pen, and O. Ronce. 2006. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. J. Evol. Biol. 19:203–215.

Armbruster, P., and D. H. Reed. 2005. Inbreeding depression in benign and stressful environments. Heredity 95:235–242.

Barnard-Kubow, K. B., C. L. Debban, and L. F. Galloway. 2015. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am. J. Bot. 102:1842–1853.

Barringer, B. C., E. A. Kulka, and L. F. Galloway. 2012. Reduced inbreeding depression in peripheral relative to central populations of a monocarpic herb. J. Evol. Biol. 25:1200–1208.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2014. Fitting Linear Mixed-Effects Models using lme4. arXiv.

Breheny, P., and W. Burchett. 2017. Visualization of Regression Models Using visreg. R J. 9:56–71.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302–4315.

Foutel-Rodier, F., and A. M. Etheridge. 2020. The spatial Muller's ratchet: Surfing of deleterious mutations during range expansion. Theor. Popul. Biol. 135:19–31.

Fox, J., and S. Weisberg. 2019. car: An R Companion to Applied Regression. Sage, Thousand Oaks, CA.

Gabriel, W., M. Lynch, and R. Bürger. 1993. Muller's Ratchet and mutational meltdowns. Evolution 47:1744–1757.

Galloway, L. F., and J. R. Etterson. 2007. Inbreeding depression in an autotetraploid herb: a three cohort field study. New Phytol. 173:383–392.

Gilbert, K. J., N. P. Sharp, A. L. Angert, G. L. Conte, J. A. Draghi, F. Guillaume, A. L. Hargreaves, R. Matthey-Doret, and M. C. Whitlock. 2017. Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. Am. Nat. 189:368–380.

Hargreaves, A. L., K. E. Samis, and C. G. Eckert. 2014. Are Species' Range Limits Simply Niche Limits Writ Large? A Review of Transplant Experiments beyond the Range. Am. Nat. 183:157–173.

Hijmans, R. 2021. geosphere: spherical trigonometry.

iNaturalist Community. n.d. Observations of Campanulastrum americanum from North America, United States of America observed on/between 08/2011-05/2022.

King, J. L. 1966. The gene interaction component of the genetic load. Genetics 53:403–413.

Koski, M. H., N. C. Layman, C. J. Prior, J. W. Busch, and L. F. Galloway. 2019. Selfing ability and drift load evolve with range expansion. Evol. Lett. 3:500–512.

Lüdecke, D. 2023. sjPlot: Data Visualization for Statistics in Social Science.

Marchini, G. L., N. C. Sherlock, A. P. Ramakrishnan, D. M. Rosenthal, and M. B. Cruzan. 2016. Rapid purging of genetic load in a metapopulation and consequences for range expansion in an invasive plant. Biol. Invasions 18:183–196.

Padgham, M. 2021. geodist: Fast, Dependency-Free Geodesic Distance Calculations.

Parsons, P. A. 1971. Extreme-environment heterosis and genetic loads. Heredity 26:479–483.

Parsons, P. A. 1959. Genotypic-environmental interactions for various temperatures in Drosophila melanogaster. Genetics 44:1325–1333.

Pateiro-Lopez, B., and A. Rodriguez-Casal. 2022. alphahull: Generalization of the convex hull of a sample of points in the plane.

Peischl, S., I. Dupanloup, M. Kirkpatrick, and L. Excoffier. 2013. On the accumulation of deleterious mutations during range expansions. Mol. Ecol. 22:5972–5982.

Perrier, A., D. Sánchez-Castro, and Y. Willi. 2022. Environment dependence of the expression of mutational load and species' range limits. J. Evol. Biol. 35:731–741.

Perrier, A., D. Sánchez-Castro, and Y. Willi. 2020. Expressed mutational load increases toward the edge of a species' geographic range. Evolution 74:1711–1723.

Prior, C. J., N. C. Layman, M. H. Koski, L. F. Galloway, and J. W. Busch. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. Mol. Ecol. 29:4473–4486.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Slatkin, M., and L. Excoffier. 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics 191:171–181.

Whitlock, M. C., P. K. Ingvarsson, and T. Hatfield. 2000. Local drift load and the heterosis of interconnected populations. Heredity 84:452–457.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, NY.

Willi, Y. 2013. Mutational meltdown in selfing Arabidopsis lyrata. Evolution 67:806–815.

Willi, Y., M. Fracassetti, S. Zoller, and J. Van Buskirk. 2018. Accumulation of Mutational Load at the Edges of a Species Range. Mol. Biol. Evol. 35:781–791.

Willi, Y., and J. Van Buskirk. 2019. A Practical Guide to the Study of Distribution Limits. Am. Nat. 193:773–785.

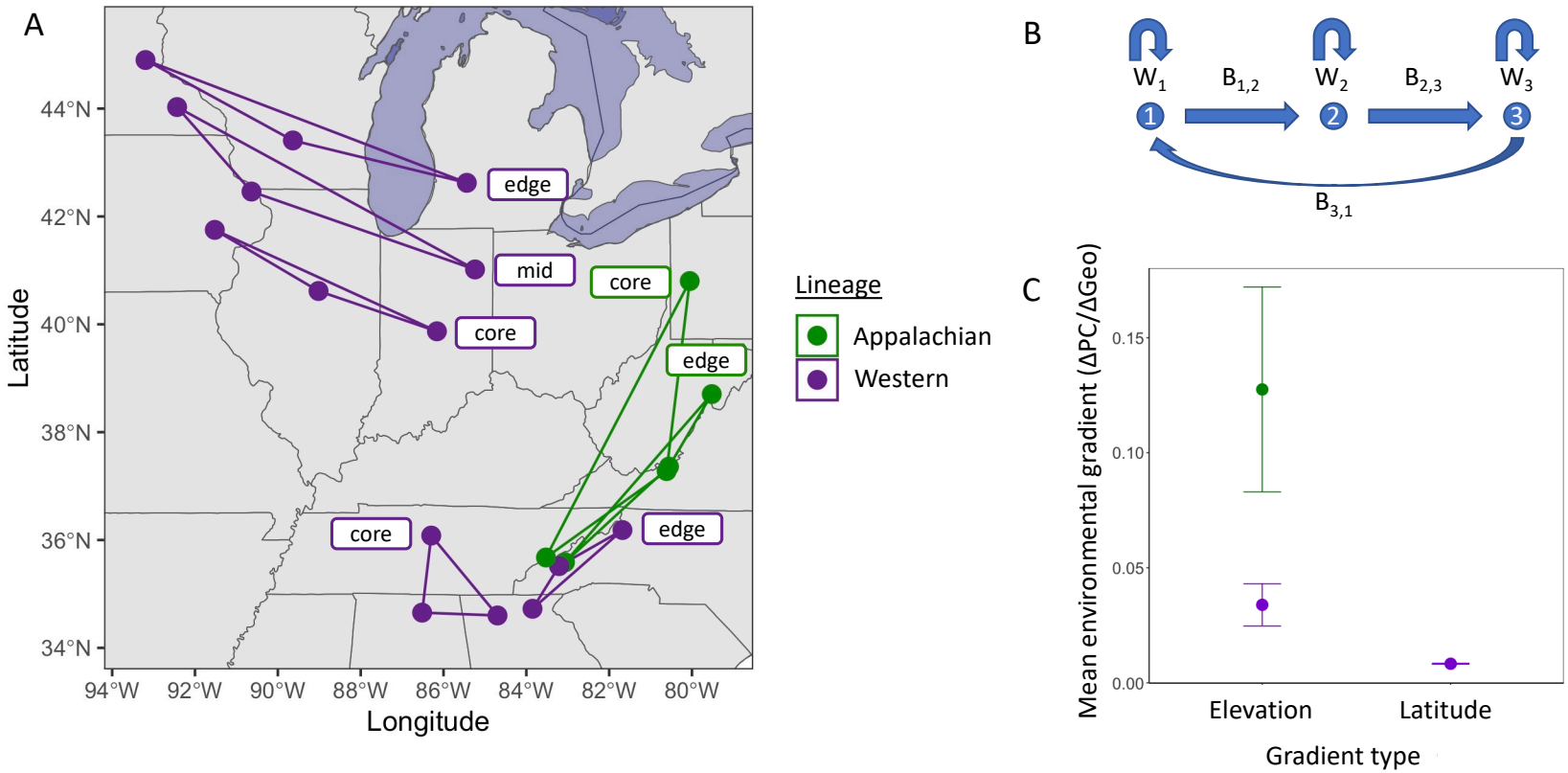


Figure 1) (A) Locations of *Campanula americana* populations sampled across latitudinal and elevational gradients (core, mid, edge). Lines connect populations that were hybridized to estimate genetic load via heterosis. (B) Crossing scheme depicting triads of between (B) and within (W) population crosses. Crosses were performed unidirectionally, such that populations were used only as a maternal recipient or paternal donor once. (C) ΔPC1/ΔGeographic distance (km) between along transects from the range core toward the range edge (core-edge/core-mid/mid-edge). Error bars show standard error of mean gradient estimates.

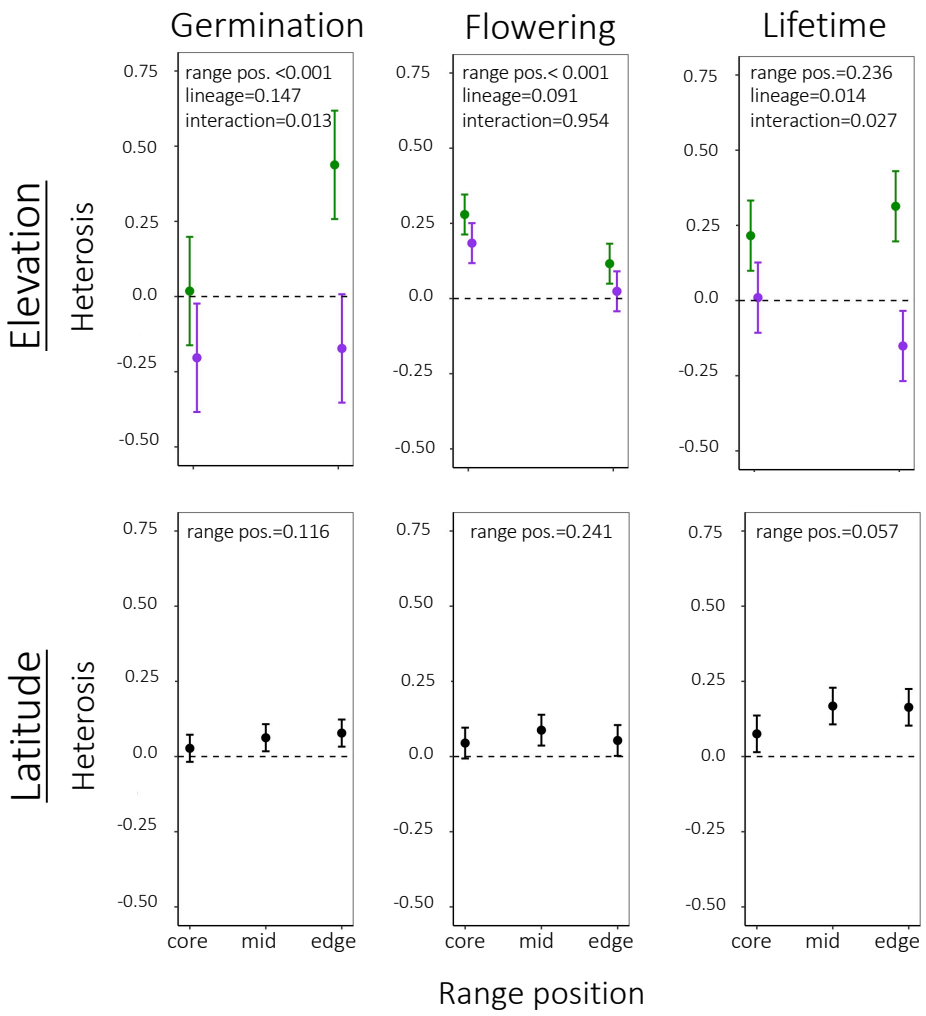


Figure 2) Heterosis estimates, an index of genetic load, expressed in hybrids between populations from the range edge, core and midway between along both latitudinal and elevational gradients for germination, flowering, and lifetime fitness. Elevational populations are separated by lineage, with the Appalachian-lineage populations in green and the Western in purple. Error bars show 95% confidence intervals for estimated marginal group means.

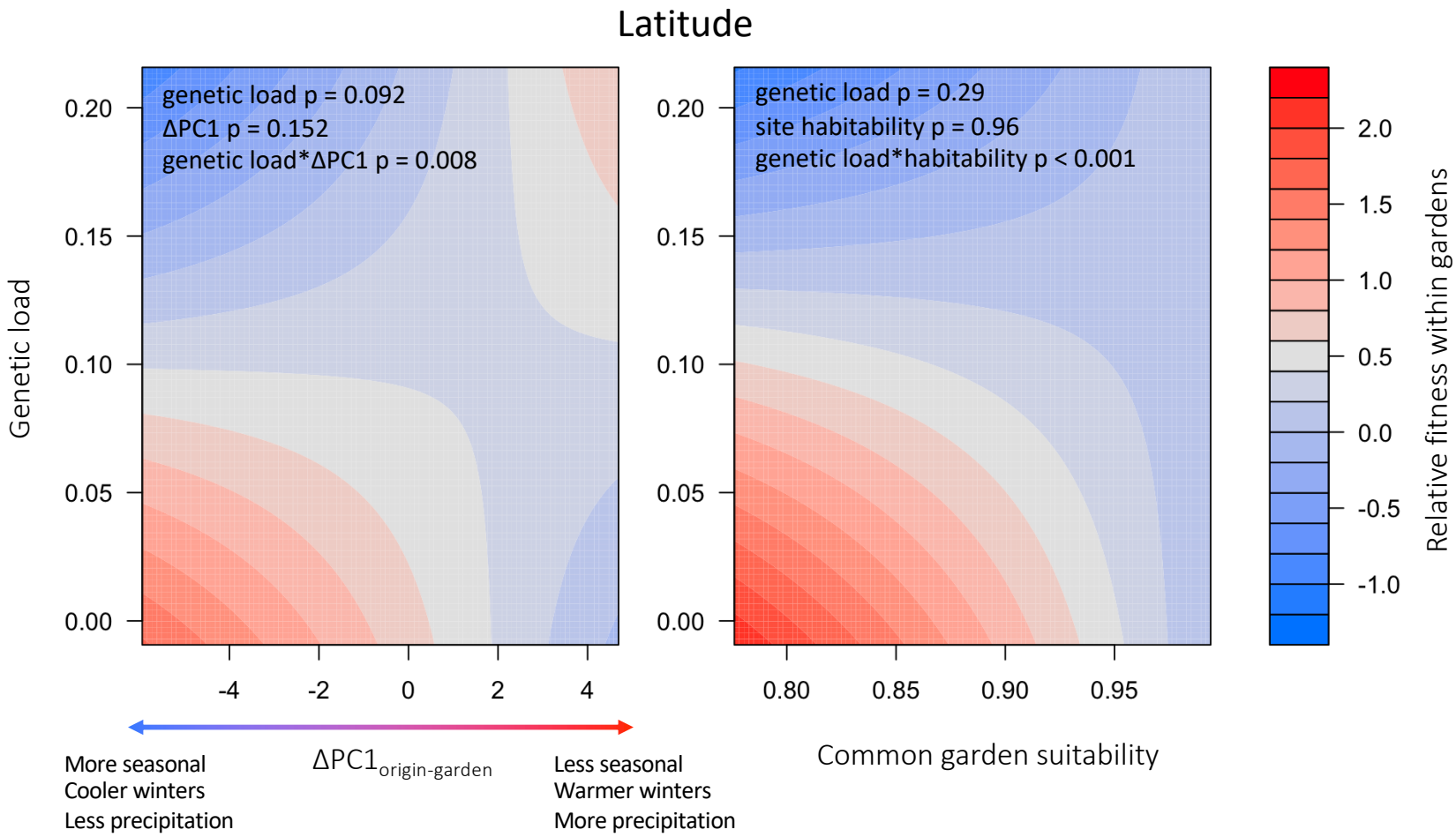
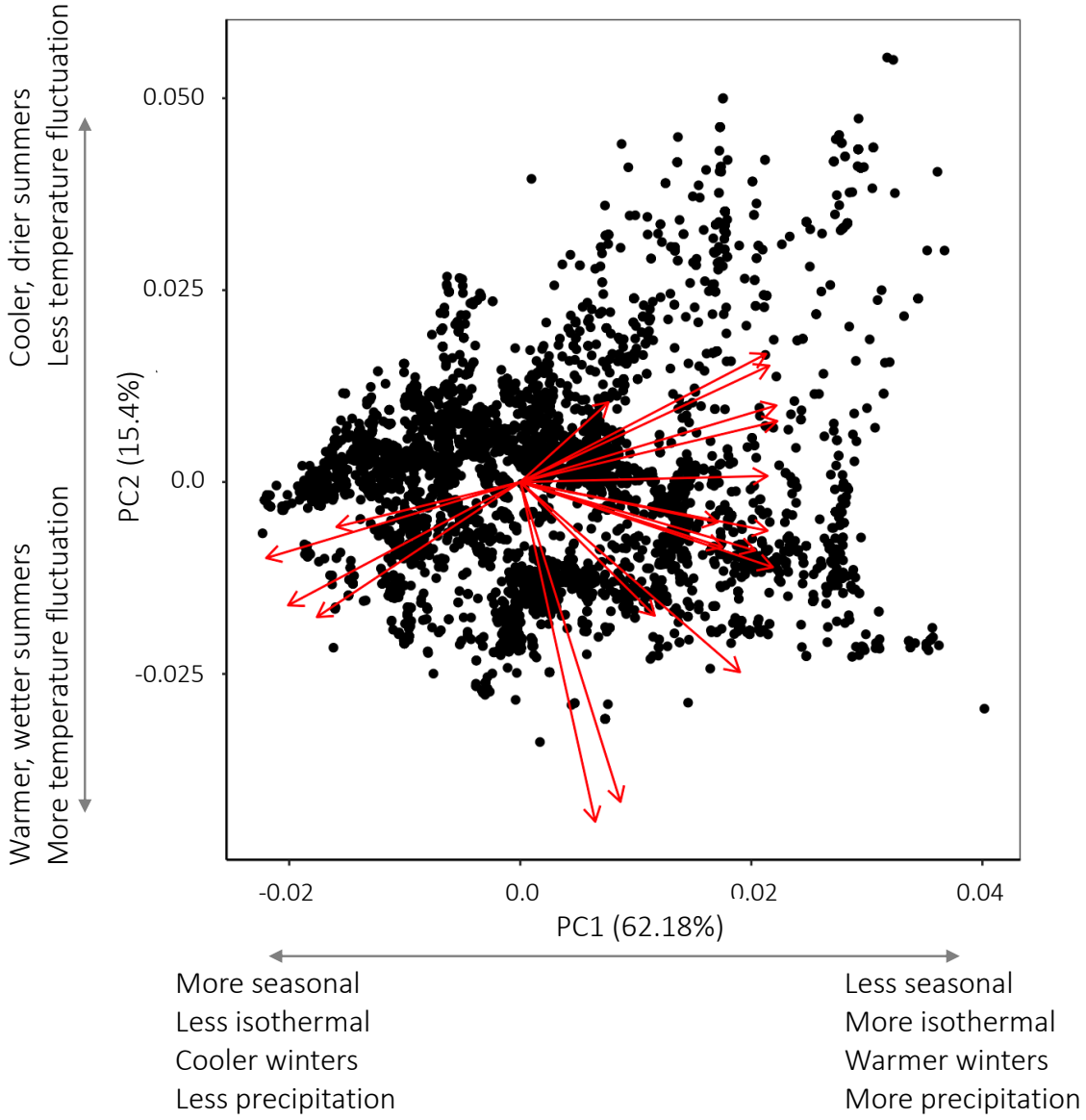
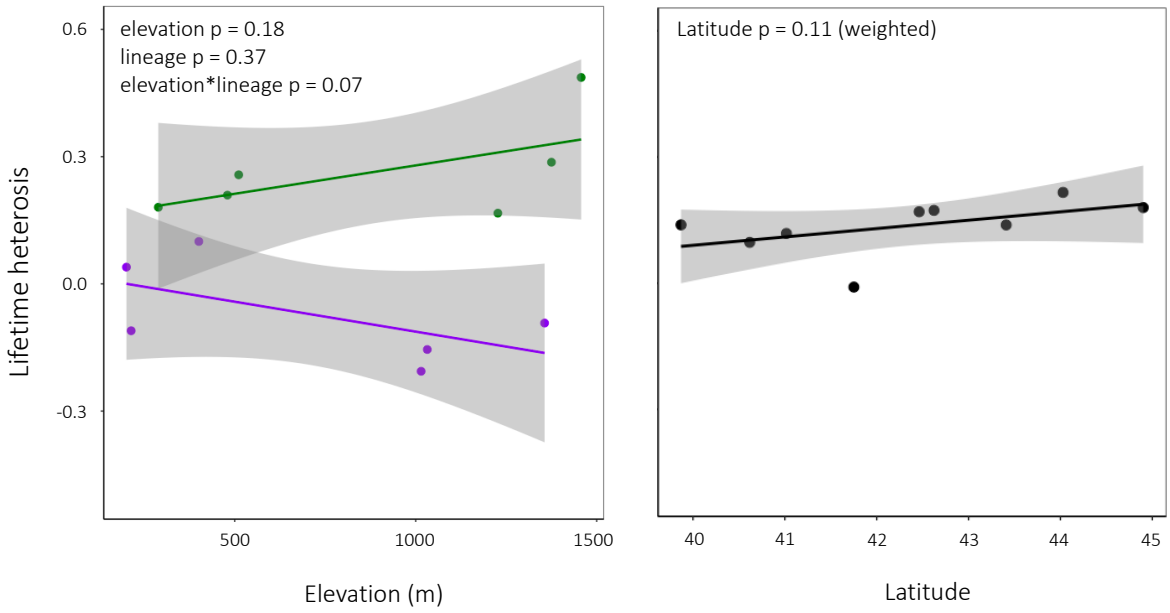


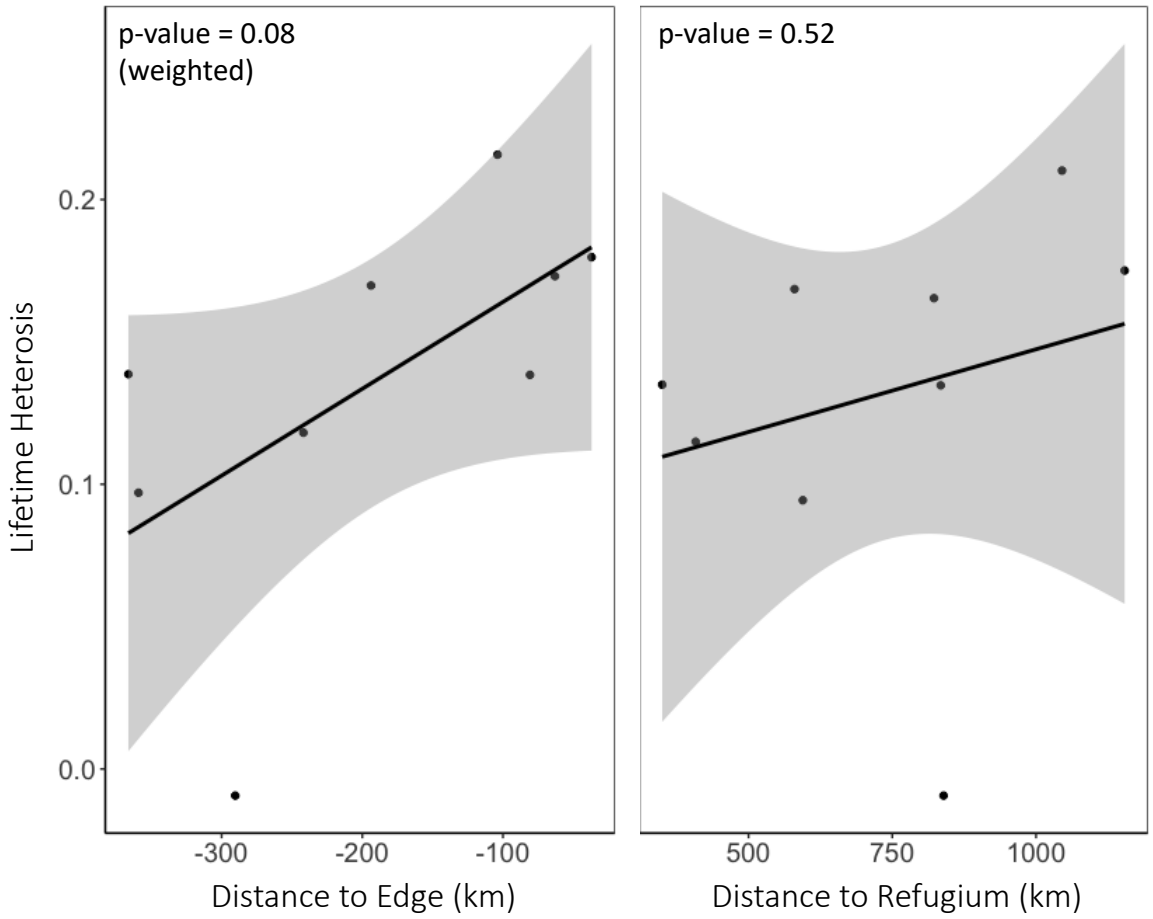
Figure 3) For the latitudinal gradient, multivariate models evaluating the relationship of relative fitness in common gardens with genetic load and environmental displacement (ΔPC1) or common garden site suitability. For both measures of climate, as sites became less suitable and climates harsher, populations with less genetic load tended to outperform populations with high load.



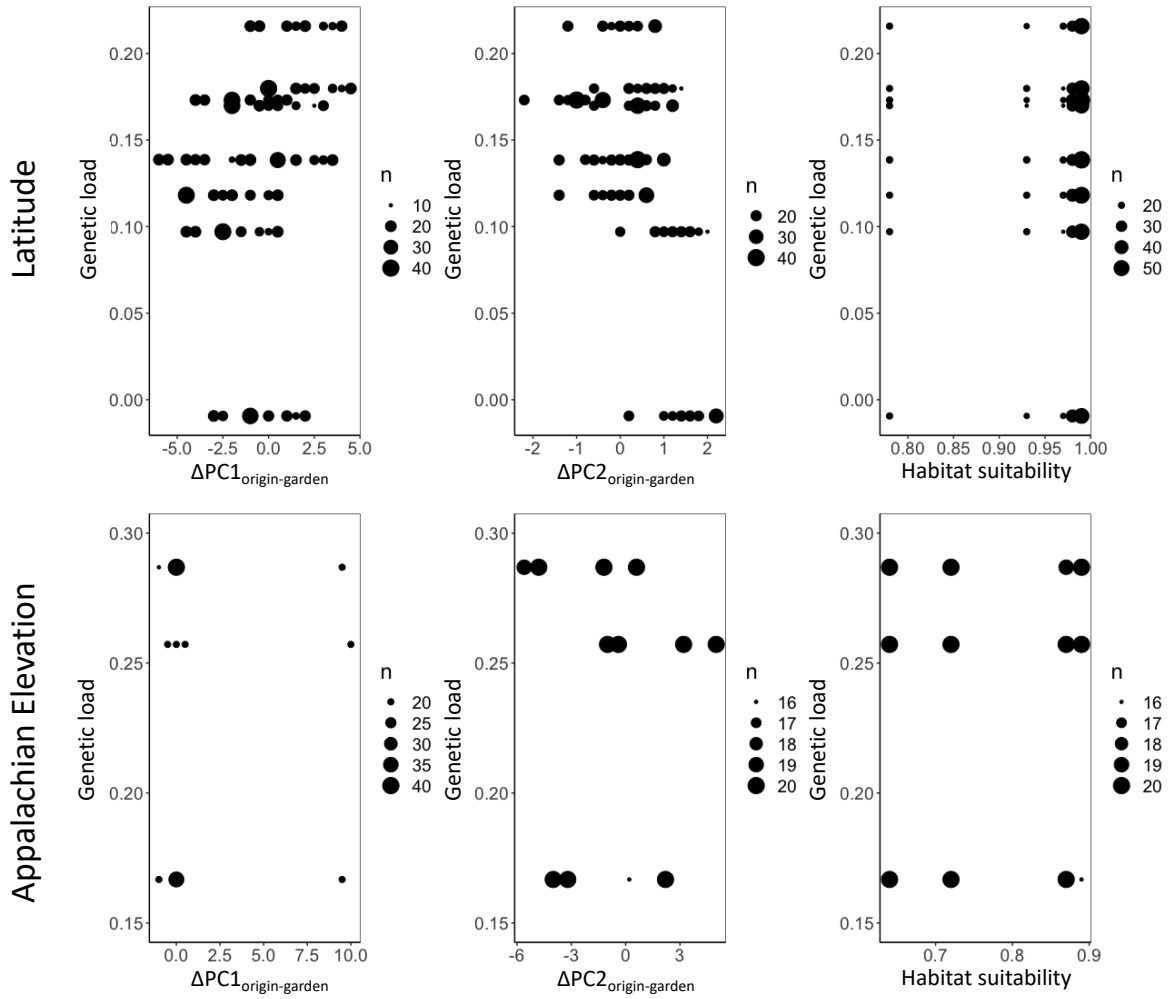
SI Figure 1) Principal component analysis of bioclimatic factors for all iNaturalist research-grade observations, shown as points (n=7,781). Red arrows show eigenvectors for WorldClim 2.0 bioclimatic variables (n=19) in the PC1 and PC2 dimensions. Together, PC1 and PC2 explain 77.62% of variance in bioclimatic factors.



SI Figure 2) Heterosis for lifetime fitness as a function of elevational and latitudinal gradients. Environmental gradient did not explain heterosis. For elevational populations this was also true when lineages were modeled separately, Appalachian p-value=0.21 (green); Western p-value=0.21 (purple). I weighted the latitude regression to account for the influence of an outlier.



SI Figure 3) Modeled relationship of heterosis for lifetime fitness and distance from the range edge and the mid-latitude refugium for populations from the latitudinal gradient.



SI Figure 4) Data distributions underlying models exploring the interactions of genetic load and environmental context on relative fitness within and among common gardens (Figure 3). Symbol sizes are scaled by the number of points in the approximate area of the surface. Distributions are not shown for Western lineage elevational populations, as models were rank deficient.

population	gradient	lineage	range position	transect	latitude	longitude	elevation	germination load	flowering load	lifetime load
EGG	elevation	Appalachian	core	1	37.28	-80.61	511	0.02	0.22	0.26
PA104	elevation	Appalachian	core	2	40.8	-80.05	289	0.01	0.27	0.18
TN92	elevation	Appalachian	core	3	35.68	-83.53	480	0.03	0.34	0.21
VA73	elevation	Appalachian	edge	1	37.36	-80.56	1227	0.42	0.08	0.17
SK	elevation	Appalachian	edge	2	38.71	-79.52	1375	0.17	0.19	0.29
NC91	elevation	Appalachian	edge	3	35.59	-83.07	1457	0.73	0.07	0.49
GA1	elevation	Western	core	4	34.6	-84.7	214	-0.32	0.17	-0.11
ALBG	elevation	Western	core	5	34.65	-86.52	401	-0.07	0.17	0.1
TN34	elevation	Western	core	6	36.08	-86.3	201	-0.22	0.22	0.04
GA2	elevation	Western	edge	4	34.72	-83.85	1015	-0.07	-0.04	-0.21
NC16	elevation	Western	edge	5	36.19	-81.68	1032	-0.22	0.09	-0.15
NC130	elevation	Western	edge	6	35.52	-83.21	1356	-0.23	0.02	-0.09
HP	latitude	Western	core	7	39.87	-86.16	NA	0.09	0.03	0.14
SH	latitude	Western	core	8	40.62	-89.02	NA	0.01	0.05	0.1
IJ	latitude	Western	core	9	41.75	-91.52	NA	-0.02	0.05	-0.01
FC	latitude	Western	mid	7	41.02	-85.24	NA	0.05	0.06	0.12
HB	latitude	Western	mid	8	42.46	-90.64	NA	0.08	0.06	0.17
QH	latitude	Western	mid	9	44.03	-92.43	NA	0.06	0.14	0.22
FT	latitude	Western	edge	7	42.62	-85.44	NA	0.12	-0.01	0.17
V	latitude	Western	edge	8	43.41	-89.64	NA	0.05	0.11	0.14
MDP	latitude	Western	edge	9	44.9	-93.2	NA	0.07	0.07	0.18

SI Table 1) Populations sampled with assignment to respective gradient, lineage, and range position. Coordinates and heterosis/genetic load estimates are included. Negative values for heterosis are suggestive of mild reproductive isolation within crosses, preventing the use of Western lineage elevational populations in models comparing patterns of genetic load.

maternal population	paternal population	type	count
ALBG	ALBG	WI	11
Egg	Egg	WI	11
FC	FC	WI	13
FT	FT	WI	12
GA1	GA1	WI	13
GA2	GA2	WI	3
HB	HB	WI	13
HP	HP	WI	12
IJ	IJ	WI	8
MDP	MDP	WI	14
NC130	NC130	WI	3
NC16	NC16	WI	9
NC91	NC91	WI	5
PA104	PA104	WI	3
QH	QH	WI	15
SH	SH	WI	11
SK	SK	WI	8
TN92	TN92	WI	10
V	V	WI	11
VA73	VA73	WI	12
ALBG	GA1	BW	8
EGG	TN92	BW	9
FC	QH	BW	10
FT	MDP	BW	11
GA1	TN34	BW	8
GA2	NC130	BW	3
HB	FC	BW	10
HP	IJ	BW	4
IJ	SH	BW	2
MDP	V	BW	9
NC130	NC16	BW	2
NC16	GA2	BW	2
NC91	SK	BW	5
PA104	EGG	BW	4
QH	HB	BW	12
SH	HP	BW	13
SK	VA73	BW	10
TN34	ALBG	BW	8
TN92	PA104	BW	4
VA73	NC91	BW	7
V	FT	BW	12

SI Table 2) Cross within and between populations with the number of each cross performed. WI indicates the cross was performed within the population, while BW indicates the cross was performed between the populations. Population 1 indicates the population of the maternal recipient in the cross, while population 2 indicates the population of the paternal donor in the cross.

term	df	F-value	p-value
Cohort	1	1.30	0.254
Cross Type	1	4.09	**0.043**
Environmental Gradient	1	10.19	**0.001**
Cohort:Cross Type	1	0.00	0.988
Cohort:Environmental Gradient	1	0.74	0.389
Cross Type:Environmental Gradient	1	0.50	0.479
Cohort:Cross Type:Environmental Gradient	1	0.24	0.622
Residuals	735		

SI Table 3) Type III ANOVA results testing from significance of cohort effects on reproductive output with factors of the cross type (within/between) and environmental gradient (elevation/latitude; *reproductive output ~ cohort * cross type * gradient type*). Cohort was not significant as a fixed effect in the model, nor in any interaction. Thus, reproductive output per cohort did not vary according to gradient, cross type, or gradient and cross type together.

gradient	lineage	environment term	model terms	Chi^2	p-value
latitude	Western	habitat suitability	genetic load	1.14	0.288
			habitat suitability	< 0.01	0.957
			load * habitat	**27.83**	**< 0.001**
		ΔPC1 origin - garden	genetic load	2.83	0.092
			ΔPC1 origin - garden	2.05	0.152
			load * ΔPC1	**7.11**	**0.008**
		ΔPC2 origin - garden	genetic load	0.64	0.423
			ΔPC2 origin - garden	0.49	0.486
			load * ΔPC2	1.23	0.268
elevation	Appalachian	habitat suitability	genetic load	0.02	0.900
			habitat suitability	0.95	0.330
			load * habitat	0.86	0.355
		ΔPC1 origin - garden	genetic load	0.01	0.906
			ΔPC1 origin - garden	0.43	0.510
			load * ΔPC1	0.13	0.723
		ΔPC2 origin - garden	genetic load	0.13	0.720
			ΔPC2 origin - garden	**17.03**	**< 0.001**
			load * ΔPC2	0.00	0.964

SI Table 4) Type III ANOVA results for mixed effect models testing how relative fitness in common gardens is influenced by genetic load and environmental conditions. Environmental conditions were estimated in two ways: first, by the ΔPC value between population origin and common garden site for PC1 and PC2. Second, by the habitat suitability for each common garden, estimated from the SDM (Appendix 1).

CHAPTER 4:

Consequences of selection and gene flow along environmental gradients.

ABSTRACT

Genetic differentiation among populations can be promoted via genetic drift and selection or constrained by gene flow. Along environmental gradients, where environments change linearly in space, divergent selection among populations can reduce gene flow, enhancing local adaptation but limiting effective population size. Here, I investigated whether patterns of genetic differentiation along gradients were broadly explained by drift, environmental selection, or gene flow using tests of isolation by distance and environment. I then explored asymmetries in migration and effective population size among populations located along both environmental gradients using demographic inference. I found isolation-by-distance along the steep environmental gradient, where migration was most asymmetric, and a decline in effective population size toward elevational range limits. Along the shallow gradient, I found weak signals of isolation-by-distance, symmetric migration, and reductions in effective population size toward latitudinal range limits. I conclude that gene flow reduces local adaptation along both environmental gradients but has led to source-sink dynamics only along the steep gradient. These results suggest that environmental gradients play an important role in mediating patterns of adaptation across range-wide scales by modulating fitness costs associated with gene flow.

INTRODUCTION

Genetic differentiation between populations can be promoted by the diversifying influences of genetic drift and selection, or constrained by gene flow (Wang 2013). The balance of these microevolutionary drivers often depends on their interactions with environment. Along shallow environmental gradients, where environments change slowly in space, weak selection against migration can allow high levels of gene flow along the gradient, reducing signatures of adaptive and neutral genetic differentiation (Polechová and Barton 2015). Along steep environmental gradients, where environments are more locally differentiated, divergent selection along the gradient drives increased costs of dispersal and local adaptation (Hereford 2009; Polechová and Barton 2015).

Where selection against gene flow is weak relative to the amount of migration, gene flow can impede local adaptation in recipient populations (Bridle et al. 2009; Bachmann et al. 2020). Yet, examples of maladaptive gene flow reducing local adaptation are uncommon (Halbritter et al. 2015; Sexton et al. 2016; Kottler et al. 2021). Understanding when and how maladaptive gene flow shapes interpopulation dynamics and local adaptation along environmental gradients can offer a clearer understanding of how microevolutionary dynamics drive range expansions across heterogeneous habitats and how gene flow influences the evolution of populations near range limits. In particular, previous theoretical (Ronce and Kirkpatrick 2001; Alleaume-Benharira et al. 2006; Henry et al. 2015) and empirical work (Strasburg et al. 2011; Fedorka et al. 2012; Brady et al. 2019) both suggest that effective population size influences and is influenced by the relationship between gene flow and local adaptation.

Effective population size (N_e), which in part describes genetic diversity within populations (Frankham 1996), is closely linked with adaptive potential. For example, effective population size

is positively correlated with adaptive divergence across species of sunflower (Strasburg et al. 2011), and small effective population size has been shown to limit population establishment of St John's wort (Oakley 2013). In general, populations near range limits are predisposed to have smaller effective population size (Cisternas-Fuentes and Koski 2023), in part due to lingering demographic effects of range expansions, like serial founder events (Slatkin and Excoffier 2012; Willi 2019). Strong selection against gene flow can reduce effective population size by sweeping adaptations of large effect to fixation (Furrer and Pasinelli 2016). In experimental evolution of adapting to novel resources in *Trilobium* flour beetles, strong environmental selection reduced effective population size within experimental populations, increasing the relative strength of drift over the course of multiple generations, eventually leading to demographic collapse (Falk et al. 2012). Near range limits, the compounding loss of effective population size due to serial founder effects and strong selection may drastically limit the adaptive potential of populations, curtailing range expansion and limiting local adaptation. Yet, few studies have attempted to characterize how effective population size influences microevolutionary dynamics and patterns of local adaptation during range expansions.

Here, I explore the relationship between migration and effective population size on microevolutionary dynamics along elevational and latitudinal environmental gradients in *Campanula americana,* a North American wildflower. Along the elevational gradient*,* populations performed better in common gardens near their home environment than away from it (range edge or range core; Chapter 1). Yet, range-core populations outperformed range-edge populations in both range edge and range core common gardens. The presence of home vs. away local adaptation in range-edge populations indicates some ecological specialization to local conditions, yet the general overperformance of range core populations indicates that populations at elevational range

limits have poor fitness. Sharply divergent allele frequencies at adaptive loci along the elevational gradient (Chapter 2), reveal strong environmental selection and high costs associated with gene flow along the gradient. In contrast, population fitness along the latitudinal environmental gradient is generally high and local adaptation is limited (Chapter 1), suggesting ecological generalization. Here, clines of adaptive allele frequencies are gradual, suggesting weak environmental selection along the latitudinal gradient (Chapter 2) and few costs to gene flow and dispersal.

I compared environmental gradients that differ in steepness in *C. americana* to reveal contributions of drift, gene flow, and selection to geographic patterns of genetic differentiation. I used isolation-by-distance tests to evaluate the balance of gene flow and genetic drift, and isolation-by-environment tests to evaluate the influence of selection relative to genetic drift and gene flow. I then examined whether migration rates among populations and effective population size within populations differ along each environmental gradient with demographic inference models. I used this suite of analyses to address the following specific questions: (1) What microevolutionary pressures are driving genetic differentiation across environmental gradients that differ in steepness? (2) Are migration rates asymmetric among populations along environmental gradients and do they indicate maladaptive gene flow? (3) Does effective population size vary along environmental gradients and does this effect depend on the steepness of the gradient? Finally, I interpret results in a broader framework of selection and adaptation across the native range of *Campanula americana* and discuss outcomes of gene flow along environmental gradients that differ in steepness.

METHODS

Study system

Campanula americana is a monocarpic wildflower distributed across the eastern United States. It is primarily outcrossing and insect pollinated (Galloway et al. 2003; Koski et al. 2019a). Additionally, *C. americana* is frequently locally abundant but geographically patchy in its distribution (field observations). *C. americana* has a complex phylogeographic history of range expansion since the last glacial maximum (Barnard-Kubow et al. 2015; Chapter 5). It is generally divided into three genetic lineages that differ in their contact and phylogeographic history. Here, I focus on the two lineages with the broadest distribution: the Appalachian lineage and the Western lineage.

The Appalachian lineage is confined to the Appalachian Mountains and likely expanded toward its contemporary elevational range limits after the last glacial maximum from multiple microrefugia scattered in and around the Appalachian Mountains (Chapter 5; Appendix 1; Barnard-Kubow et al. 2015). The Western lineage is found in the southern Appalachian Mountains and has a broad latitudinal distribution in the eastern United States. The Western lineage persisted through the last glacial maximum in relictual habitat in the American south near the Gulf of Mexico and in a smaller mid-latitude refugium (Barnard-Kubow et al. 2015). Expansion toward contemporary latitudinal range limits was largely seeded by the mid-latitude refugium and occurred in the Holocene period (Koski et al. 2019b; Prior et al. 2020), while expansion of the Western lineage toward contemporary elevational range limits was likely seeded by relictual southern populations (Barnard-Kubow et al. 2015; Perrier et al., unpublished data). As a result, the Western lineage traverses both a northward shallow latitudinal environmental gradient and a steep elevational

environmental gradient in the southern Appalachian Mountains. Whereas the Appalachian lineage traverses a steep elevational gradient throughout the Appalachian Mountains.

Genomics

To estimate genetic differentiation near the species' elevational and latitudinal range limits, I sampled individuals from populations across lineages and environmental gradients (Fig. 1; SI Table 1). In total, I sampled 19 populations from portions of the species' range nearest to latitudinal and elevational limits. Specifically, I sampled nine populations along the latitudinal gradient from positions near the range core, range edge, and intermediate between them (3 core/3 mid/3 edge); and 10 populations across two genetic lineages from the elevational gradient (4 Western/6 Appalachian) from range positions near the low elevation range core (2 Western/3 Appalachian) and the higher elevation range edge (2 Western/3 Appalachian). I sampled populations to produce three replicate latitudinal transects (Fig. 1), each containing a population from near the range core, the mid-range, and the range edge; and five elevational transects (2 Western, 3 Appalachian), each containing a population from near the low-elevation range core and the high-elevation range edge. Transects contained only populations within lineages (Western or Appalachian) to avoid comparisons confounded by phylogenetic relationships.

I germinated seeds from these 19 wild *C. americana* populations following protocols outlined in Chapter 1. I harvested leaves from young rosettes (~9 ind./pop.), then dried the leaf tissue. I sent tissue samples for RAD-Seq library prep, including DNA extraction, and Illumina sequencing at Floragenex. Details of library prep, sequencing, and assembly can be found in Chapter 1. I aligned and assembled sequences using Stacks 2 (Rochette et al. 2019). I filtered SNPs using a minor allele frequency of 0.05, and a missing threshold of 20%. I further reduced the SNP

dataset by selecting only biallelic SNPs, SNPs with a minimum read depth equal to or greater than 10, and SNPs with a maximum read depth equal to or less than 50 to ensure only high-quality SNPs were retained. In total, I retained 6,378 SNPs for the full dataset.

Structure of isolation (IBD vs. IBE) across environmental clines

I computed genetic, environmental, and geographic distance matrices between populations for the Western elevational gradient, the Appalachian elevational gradient and the Western latitudinal gradient. I determined environmental distance between populations using all bioclimatic factors from WorldClim 2.0 (n=19; Fick and Hijmans 2017). I first extracted bioclimatic factors for the coordinates of all research grade iNaturalist observations of *C. americana* as of 05/2022 (n=7,781; iNaturalist Community n.d.). I conducted a PCA on the extracted climate data, retained the first five PC axes that explained at least 95% of the variance in bioclimatic data among populations, and computed the Euclidean distance between populations in PCA space. I calculated geographic distance between pairs of populations by finding the shortest path through suitable habitat in a species distribution model (Xia et al. 2024), a metric I call 'effective geographic distance'. Effective geographic distance accounts for contiguity of habitat between populations and assesses the distance between populations that migrants would traverse remaining within the species' ecological niche (Supplementary Methods; Appendix 1). Finally, I estimated genetic distance between populations using F_{ST} calculated from the high-quality SNP set (n=6,378) with 100 bootstraps in the StAMPP package using the stamppFst function in R 4.2.0 (Pembleton et al. 2013).

I performed isolation-by-distance (IBD) and isolation-by-environment tests (IBE) analyses for each environmental gradient (elevation/latitude) and lineage (Western/Appalachian) evaluated the relative strength of selection, drift, and gene flow by. I analyzed IBE using two methods. First,

I computed multiple matrix regressions with randomization (MMRR), following Wang (2013). Briefly, MMRR works by regressing matrices of genetic distance (F_{ST}) against matrices of geographic and environmental distance, while performing 1,000 random permutations of the data to account for non-independence of pairwise population comparisons. I performed MMRR tests for each environmental gradient and lineage group using the function lgrMMRR from the PopGenReport package (Adamack and Gruber 2014). Second, I computed partial Mantel tests using identical matrices to those used in MMRR with 1,000 random permutations and Pearson correlations. To assess IBD without accounting for environmental variation across gradients, I performed Mantel tests using the mantel.partial and mantel functions from the vegan package (Dixon 2003).

Migration asymmetry and effective population sizes

Maladaptive gene flow can be inferred by asymmetric rates of migration among populations (Fedorka et al. 2012), though asymmetries do not necessarily indicate a loss of local adaptation (Tigano and Friesen 2016). I used demographic inference to evaluate whether migration is asymmetric along environmental gradients. Specifically, I simulated demographic models of population divergence using the python3 package *Moments* (Jouganous et al. 2017). Because demographic inference models operating with site frequency spectra cannot handle missing data, I maximized the number of loci retained by down-sampling the number of 2N individuals in each population that loci were drawn using *easySFS* (SI Table 2; https://github.com/isaacovercast/easySFS). I fit models to pairs of populations across positions in each transect (i.e., latitude: core/mid, mid/edge, core/edge; elevation: core/edge; Fig. 1).

I fit two demographic models to each pair of populations: split with migration ($IM_{SPLIT;}$ SI Fig. 1) and split with no migration ($NM_{SPLIT;}$ SI Fig. 1). I iterated models 50 times per pair of populations with new parameters drawn at random from uniform prior distributions (SI Table 3; Chapter 5), with the best iteration retained per model by log-likelihood. I retained the best model for each pair of populations by AIC. I translated raw parameters from native units of 4Neμ to migration rates and effective population sizes using a roughly estimated mutation rate (μ=2e-9) and the equations (Gillespie 2004; Gutenkunst et al. 2009):

$$migration\ rate_{i\rightarrow j} = \frac{m_{1\rightarrow 2}}{2*\left(\frac{\theta}{4\mu L}\right)} \qquad N_e\ of\ population_i = nu_1 * \frac{\theta}{4\mu L}$$

I compared migration toward the range edge with that away from the range edge between pairs of populations along each transect. Asymmetry in migration was evaluated using paired t-tests. Finally, I compared relative effective population sizes along the environmental gradient by taking the ratio of the N_e for each pair of populations, using the population closer to the range edge as the numerator and the population closest to the range core as the denominator. As such, values of N_e < 1 indicate a reduction in N_e toward range limits.

RESULTS

Structure of isolation (IBD vs. IBE) across environmental gradients

Isolation-by-distance and isolation-by-environment were largely absent from both environmental gradients and lineages, with the exception of Appalachian-lineage populations along the steep elevational gradient. Across elevations, patterns of IBD and IBE varied by lineage. For Appalachian-lineage populations, genetic differentiation was structured by geographic distance but not environmental distance (Table 1), suggesting that drift is responsible for most genetic divergence between the range core and range edge. For Western-lineage populations along the

steep elevational gradient, environmental distance was near significant in Partial Mantel tests (p-value=0.075) but not MMRR tests (p-value=0.19), indicating a weak contribution of environmental selection to differentiation. In cases where p-values ≤ 0.05, MMRR, partial Mantel, and Mantel tests agreed with one another.

Across the latitudinal environmental gradient, genetic differentiation was not significantly structured by geographic distance among populations (Mantel p-value=0.25; Table 1). Lack of IBD across latitudes suggests that the differentiating effect of genetic drift and homogenizing influence of gene flow are largely in equilibrium along the shallow environmental gradient.

Migration asymmetry and effective population sizes

Across environmental gradients, range positions and lineages, migration models (IM_{SPLIT}) fit genetic data better by AIC than no migration models (NM_{SPLIT}; SI Fig. 2). IM_{SPLIT} models identified migration asymmetry. Across elevational transects, migration was consistently greater from low elevation, range-core populations into high elevation, range edge populations than vice versa (p-value=0.028; Fig. 2). For latitudinal transects, similarly, migration was consistently greater from range core into range edge populations, though the asymmetry was only near significant (p-value=0.070). Migration was not significantly asymmetric in the two components of the latitudinal transect. Specifically, migration was not significantly greater from mid-range populations to range-edge populations than vice versa (p-value=0.121; Fig. 2), nor from range-core populations to mid-range populations than the converse (p-value=0.119). In total, seven of nine comparisons showed greater migration toward range limits than vice versa.

Estimates of effective population size from IM_{SPLIT} models declined toward range limits. For the steeper elevational gradient, populations near the elevational range edge had smaller

effective population size than populations near the range core (Fig. 3), though standard errors of the ratio of effective population sizes somewhat overlapped with 1.0 (i.e., equal N_e). For populations along the shallow latitudinal gradient, effective population sizes were similar between mid-range and range-edge populations but declined between the mid-range and range core as well as between the range edge and range core (Fig. 3).

DISCUSSION

Isolation by distance and isolation by environment

Genetic differentiation among populations results from the sum of diversifying processes and the homogenizing influence of gene flow. Across the shallower latitudinal environmental gradient, I did not detect isolation-by-environment (Table 1), indicating that latitudinal populations of *C. americana*, which rapidly expanded across the shallow environmental gradient since the last glacial maximum (Koski et al. 2019b; Prior et al. 2020), are not well-differentiated by adaptation to their respective environments. Absence of isolation-by-environment along the latitudinal gradient accords with previous work finding an absence of local adaptation among latitudinal populations of *C. americana* (Chapter 1). Such findings could be explained by selection for generalist adaptations under regimes of strong gene flow—in yeast, strong gene flow between distinct experimental environments promoted ecological generalization with limited adaptive differentiation (Tusso et al. 2021).

The presence of isolation-by-distance and weak isolation-by-environment near elevational range limits in *C. americana* suggests that these populations may be more strongly differentiated by drift than selection. Isolation-by-distance but not environment accords with findings of weak local adaptation and poor fitness of range edge populations across common gardens (Chapter 1).

Such patterns are also supported by a meta-analysis of local adaptation along environmental gradients (Hargreaves et al. 2014), which found that range expansions along steep gradients are more likely to be constrained by failure to adapt to environmental conditions near range limits than limitations to dispersal. Yet, it is important to note that RAD-sequencing may underestimate signals of isolation-by-environment. Reduced coverage sequencing, like RAD-seq, only samples a small fraction of the genome (Baird et al. 2008) and may inadequately capture genomic patterns of selection and adaptation (Lowry et al. 2017).

Migration asymmetry and effective population size

Effective population size shrank toward latitudinal range limits (Figs. 2, 3), declining in mid-range populations and stabilizing toward the range edge. Reductions in effective population size toward range limits accords with previous findings of an adaptive plateau between mid-range and range-edge populations (Chapter 1), and suggest lingering founder effects associated with range expansion are influencing the evolution of populations near latitudinal range limits (e.g., Slatkin and Excoffier 2012).

Migration rates between adjacent populations along the latitudinal gradient were not significantly asymmetric (i.e., core/mid and mid/edge), though migration rates tended to be higher toward range limits than vice versa. These patterns accord with previous findings of limited local adaptation in common gardens along the latitudinal gradient (Chapter 1). Significant differences in relative fitness were limited to range edge common gardens, where range edge and mid-range populations underperformed relative to range core populations. These patterns suggest that populations near latitudinal range limits suffer fitness losses in stressful environmental conditions. Thus, while local adaptation is limited along the latitudinal gradient, it appears to be constrained

by lower effective population size in novel and more stressful environments, but not within the established species' range where fitness was high.

Along the steep elevational gradient, migration was highly asymmetric and effective population size declined toward range limits (Fig. 2). While such patterns may indicate lingering demographic effects of post-glacial range expansion, the close proximity of range core and range edge populations suggests that serial founder events are unlikely to be solely responsible. Instead, asymmetry in migration among populations and small effective population size near range limits suggest fitness of range edge populations is constrained by maladaptive gene flow—a common finding in theoretical models examining patterns of local adaptation along steep environmental gradients (Polechová and Barton 2015; Gilbert et al. 2017; Bridle et al. 2019).

Where population size contracts along an environmental gradient, migration toward the smaller population, here the range edge, often represents a greater proportion of total gene exchange in recipient populations (Antonovics 1976). Asymmetric gene flow among populations can promote gene swamping, impeding local adaptation near range limits. However, costs of dispersal along steep environmental gradients can result in strong selection against maladaptive gene flow (Tigano and Friesen 2016). In *C. americana,* elevational populations have divergent allele frequencies in adaptive loci (Chapter 2), suggesting strong selection against gene flow along the steep gradient. Selection against gene flow among populations can protect nascent adaptive differentiation in range edge populations. Yet, selection against gene flow along steep gradients, where migration is common (Fig. 2), can reduce effective population size (Tigano and Friesen 2016), predisposing populations to genetic drift (Willi and Van Buskirk 2019) and limiting the potential for adaptation to novel environments near range limits. In *C. americana,* there is selection against maladaptive gene flow (Chapter 1, 2) and limited home vs. away local adaptation (Chapter

1) along an elevational gradient, yet fitness of range edge populations is generally poor across environments (Chapter 1). hypothesize that strong selection against maladaptive gene flow has reduced fitness in elevational range-edge populations in *C. americana* through reductions in effective population size, increasing the local strength of genetic drift and decreasing the potential of populations to adapt to novel and stressful environments.

Synthesis

The degree of adaptive differentiation and genetic isolation promoted by environmental gradients is often idiosyncratic to the species studied (Halbritter et al. 2015; Sexton et al. 2016; Bachmann et al. 2020; Aguirre-Liguori et al. 2021). An increasing body of evidence suggests that such conclusions are products of latent variables not included in studies (Willi and Van Buskirk 2019), including asymmetry of migration (Tigano and Friesen 2016), which can indicate directional selection dynamics, and effective population size, which limits the genetic substrate that selection can act (Oakley 2013). Additionally, the steepness of environmental gradients may interact with these latent variables to affect the balance of drift and selection on the dynamics of range expansion and adaptation near range limits (Gaggiotti and Smouse 1996).

Near ecologically marginal habitat, like elevational range limits (Hargreaves et al. 2014), gene flow from neighboring populations can introduce maladaptive genetic variance (Bridle et al. 2009, 2019; Gilbert et al. 2017; Bachmann et al. 2020). Previous work in *C. americana* found adaptive loci fixed in populations inhabiting similar range positions that were absent in divergent habitats (Chapter 2), suggesting selection against maladaptive genetic variants. Selection against maladaptive gene flow (Polechová and Barton 2015), particularly when heterozygous underdominance and negative epistasis are at play (Christiansen 1974; Ziehe and Gregorius 1985;

Wilson and Turelli 1986), can protect adaptive polymorphism along elevational gradients where populations are well-connected by migration but inhabit different environmental conditions (Chapter 2). Yet, such processes can also drive effective population size down through strong selective sweeps (Falk et al. 2012), limiting adaptive potential (Oakley 2013; Brady et al. 2019) and ultimately driving ecologically marginal populations toward extinction (Ronce and Kirkpatrick 2001; Willi 2013; Foutel-Rodier and Etheridge 2020). Together, interactions of strong selection and migration may predispose interconnected populations along steep environmental gradients toward narrow ecological specialization.

In contrast, selection against gene flow is weak along shallow latitudinal gradients, where costs of dispersal and local adaptation are low (Chapter 2; Polechová and Barton 2015). Weak selection against gene flow can promote the development of ecological generalization (Spichtig and Kawecki 2004) and minimize adaptive differentiation over the breadth of the shallow latitudinal environmental gradient (Chapter 1). Experimental evolution has found similar patterns of ecological generalization in continuous environments in yeast under strong gene flow with specialization only occurring when gene flow is interrupted (Tusso et al. 2021). Together, interactions of strong migration and weak selection against gene flow may predispose latitudinal populations of *C. americana* toward ecological generalization.

Studying patterns of migration and effective population size near range limits may yield insight into how species' ranges respond to changing climate. As climates warm and species ranges shift (Appendix 1), maladaptive gene flow along environmental gradients may become adaptive, assisting populations near range limits to succeed in warming climates. Alternatively, populations near range limits may be subsumed or displaced by range core populations tracking their respective climate niches. In other plant species, it has been found that local adaptation in populations near

cool, high elevation range limits does not better enable them to track suitable environmental conditions (Hargreaves and Eckert 2019). This pattern suggests that cold-adapted populations near range limits may benefit from assisted migration that introduces warm-adapted alleles. However, empirical tests examining how gene flow along environmental gradients are shaped by changing environmental conditions are limited (but see Hargreaves and Eckert 2019). Initial studies of assisted migration indicate that moderate gene flow between populations along environmental gradients may assist adaptation to warming climates (Aitken and Whitlock 2013). Research examining the relationship between gene flow and local adaptation across environmental gradients, as reported here, can provide important insight into our understanding of whether gene flow will facilitate or constrain climate-induced range shifts.

LITERATURE CITED

Adamack, A. T., and B. Gruber. 2014. POP GEN REPORT: simplifying basic population genetic analyses in R. Methods Ecol. Evol. 5:384–387.

Aguirre-Liguori, J. A., S. Ramírez-Barahona, and B. S. Gaut. 2021. The evolutionary genomics of species' responses to climate change. Nat. Ecol. Evol. 5:1350–1360.

Aitken, S. N., and M. C. Whitlock. 2013. Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu. Rev. Ecol. Evol. Syst. 44:367–388.

Alleaume-Benharira, M., I. R. Pen, and O. Ronce. 2006. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. J. Evol. Biol. 19:203–215.

Antonovics, J. 1976. The Nature of Limits to Natural Selection. Ann. Mo. Bot. Gard. 63:224.

Bachmann, J. C., A. Jansen Van Rensburg, M. Cortazar-Chinarro, A. Laurila, and J. Van Buskirk. 2020. Gene Flow Limits Adaptation along Steep Environmental Gradients. Am. Nat. 195:E67–E86.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE 3:e3376.

Barnard-Kubow, K. B., C. L. Debban, and L. F. Galloway. 2015. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am. J. Bot. 102:1842–1853.

Brady, S. P., D. I. Bolnick, R. D. H. Barrett, L. Chapman, E. Crispo, A. M. Derry, C. G. Eckert, D. J. Fraser, G. F. Fussmann, A. Gonzalez, F. Guichard, T. Lamy, J. Lane, A. G. McAdam, A. E. M. Newman, A. Paccard, B. Robertson, G. Rolshausen, P. M. Schulte, A.

M. Simons, M. Vellend, and A. Hendry. 2019. Understanding Maladaptation by Uniting Ecological and Evolutionary Perspectives. Am. Nat. 194:495–515.

Bridle, J. R., S. Gavaz, and W. J. Kennington. 2009. Testing limits to adaptation along altitudinal gradients in rainforest *Drosophila*. Proc. R. Soc. B Biol. Sci. 276:1507–1515.

Bridle, J. R., M. Kawata, and R. K. Butlin. 2019. Local adaptation stops where ecological gradients steepen or are interrupted. Evol. Appl. 12:1449–1462.

Christiansen, F. B. 1974. Sufficient Conditions for Protected Polymorphism in a Subdivided Population. Am. Nat. 108:157–166.

Cisternas-Fuentes, A., and M. H. Koski. 2023. Drivers of strong isolation and small effective population size at a leading range edge of a widespread plant. Heredity 130:347–357.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14:927–930.

Falk, J. J., C. E. Parent, D. Agashe, and D. I. Bolnick. 2012. Drift and selection entwined: asymmetric reproductive isolation in an experimental niche shift. Evol. Ecol. Res. 14:403–423.

Fedorka, K. M., W. E. Winterhalter, K. L. Shaw, W. R. Brogan, and T. A. Mousseau. 2012. The role of gene flow asymmetry along an environmental gradient in constraining local adaptation and range expansion. J. Evol. Biol. 25:1676–1685.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302–4315.

Foutel-Rodier, F., and A. M. Etheridge. 2020. The spatial Muller's ratchet: Surfing of deleterious mutations during range expansion. Theor. Popul. Biol. 135:19–31.

Frankham, R. 1996. Relationship of Genetic Variation to Population Size in Wildlife. Conserv. Biol. 10:1500–1508.

Furrer, R. D., and G. Pasinelli. 2016. Empirical evidence for source–sink populations: a review on occurrence, assessments and implications. Biol. Rev. 91:782–795.

Gaggiotti, O. E., and P. E. Smouse. 1996. Stochastic Migration and Maintenance of Genetic Variation in Sink Populations. Am. Nat. 147:919–945.

Galloway, L. F., J. R. Etterson, and J. L. Hamrick. 2003. Outcrossing rate and inbreeding depression in the herbaceous autotetraploid, Campanula americana. Heredity 90:308–315.

Gilbert, K. J., N. P. Sharp, A. L. Angert, G. L. Conte, J. A. Draghi, F. Guillaume, A. L. Hargreaves, R. Matthey-Doret, and M. C. Whitlock. 2017. Local Adaptation Interacts with Expansion Load during Range Expansion: Maladaptation Reduces Expansion Load. Am. Nat. 189:368–380.

Gillespie, J. 2004. Population Genetics: A Concise Guide. Second. The John Hopkins University Press.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLoS Genet. 5:e1000695.

Halbritter, A. H., R. Billeter, P. J. Edwards, and J. M. Alexander. 2015. Local adaptation at range edges: comparing elevation and latitudinal gradients. J. Evol. Biol. 28:1849–1860.

Hargreaves, A. L., and C. G. Eckert. 2019. Local adaptation primes cold-edge populations for range expansion but not warming-induced range shifts. Ecol. Lett. 22:78–88.

Hargreaves, A. L., K. E. Samis, and C. G. Eckert. 2014. Are Species' Range Limits Simply Niche Limits Writ Large? A Review of Transplant Experiments beyond the Range. Am. Nat. 183:157–173.

Henry, R. C., A. Coulon, and J. M. J. Travis. 2015. Dispersal asymmetries and deleterious mutations influence metapopulation persistence and range dynamics. Evol. Ecol. 29:833–850.

Hereford, J. 2009. A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. Am. Nat. 173.

iNaturalist Community. n.d. Observations of Campanulastrum americanum from North America, United States of America observed on/between 08/2011-05/2022.

Jouganous, J., W. Long, A. P. Ragsdale, and S. Gravel. 2017. Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. Genetics 206:1549–1567.

Koski, M. H., L. F. Galloway, and J. W. Busch. 2019a. Pollen limitation and autonomous selfing ability interact to shape variation in outcrossing rate across a species range. Am. J. Bot. 106:1240–1247.

Koski, M. H., N. C. Layman, C. J. Prior, J. W. Busch, and L. F. Galloway. 2019b. Selfing ability and drift load evolve with range expansion. Evol. Lett. 3:500–512.

Lowry, D. B., S. Hoban, J. L. Kelley, K. E. Lotterhos, L. K. Reed, M. F. Antolin, and A. Storfer. 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol. Ecol. Resour. 17:142–152.

Oakley, C. G. 2013. Small effective size limits performance in a novel environment. Evol. Appl. 6:823–831.

Pembleton, L. W., N. O. I. Cogan, and J. W. Forster. 2013. St AMPP : an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. 13:946–952.

Polechová, J., and N. H. Barton. 2015. Limits to adaptation along environmental gradients. Proc. Natl. Acad. Sci. 112:6401–6406.

Prior, C. J., N. C. Layman, M. H. Koski, L. F. Galloway, and J. W. Busch. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. Mol. Ecol. 29:4473–4486.

Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28:4737–4754.

Ronce, O., and M. Kirkpatrick. 2001. When sources become sinks: migrational meltdown in heterogeneous habitats. Evolution 55:1520–1531.

Sexton, J. P., M. B. Hufford, A. C.Bateman, D. B. Lowry, H. Meimberg, S. Y. Strauss, and K. J. Rice. 2016. Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. Mol. Ecol. 25:911–928.

Slatkin, M., and L. Excoffier. 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics 191:171–181.

Spichtig, M., and T. J. Kawecki. 2004. The Maintenance (or Not) of Polygenic Variation by Soft Selection in Heterogeneous Environments. Am. Nat. 164:70–84.

Strasburg, J. L., N. C. Kane, A. R. Raduski, A. Bonin, R. Michelmore, and L. H. Rieseberg. 2011. Effective Population Size Is Positively Correlated with Levels of Adaptive Divergence among Annual Sunflowers. Mol. Biol. Evol. 28:1569–1580.

Tigano, A., and V. L. Friesen. 2016. Genomics of local adaptation with gene flow. Mol. Ecol. 25:2144–2164.

Tusso, S., B. P. S. Nieuwenhuis, B. Weissensteiner, S. Immler, and J. B. W. Wolf. 2021. Experimental evolution of adaptive divergence under varying degrees of gene flow. Nat. Ecol. Evol. 5:338–349.

Wang, I. J. 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: a muliple matrix regression appraoch for quantifying geographic and ecological isolation. Evolution 67:3403–3411.

Willi, Y. 2013. Mutational meltdown in selfing Arabidopsis lyrata. Evolution 67:806–815.

Willi, Y. 2019. The relevance of mutation load for species range limits. Am. J. Bot. 106:757–759.

Willi, Y., and J. Van Buskirk. 2019. A Practical Guide to the Study of Distribution Limits. Am. Nat. 193:773–785.

Wilson, D. S., and M. Turelli. 1986. Stable Underdominance and the Evolutionary Invasion of Empty Niches. Am. Nat. 127:835–850.

Xia, M., Y. Luo, Y. Suyama, A. Matsuo, S. Sakaguchi, Y. Wang, and P. Li. 2024. Genetic divergence and ecological adaptation of an eastern North American spring ephemeral *Sanguinaria canadensis*. Divers. Distrib. 30:e13813.

Ziehe, M., and H.-R. Gregorius. 1985. The Significance of Over- and Underdominance for the Maintenance of Genetic Polymorphisms I. Underdominance and Stability. J. Theor. Biol. 117:493–504.

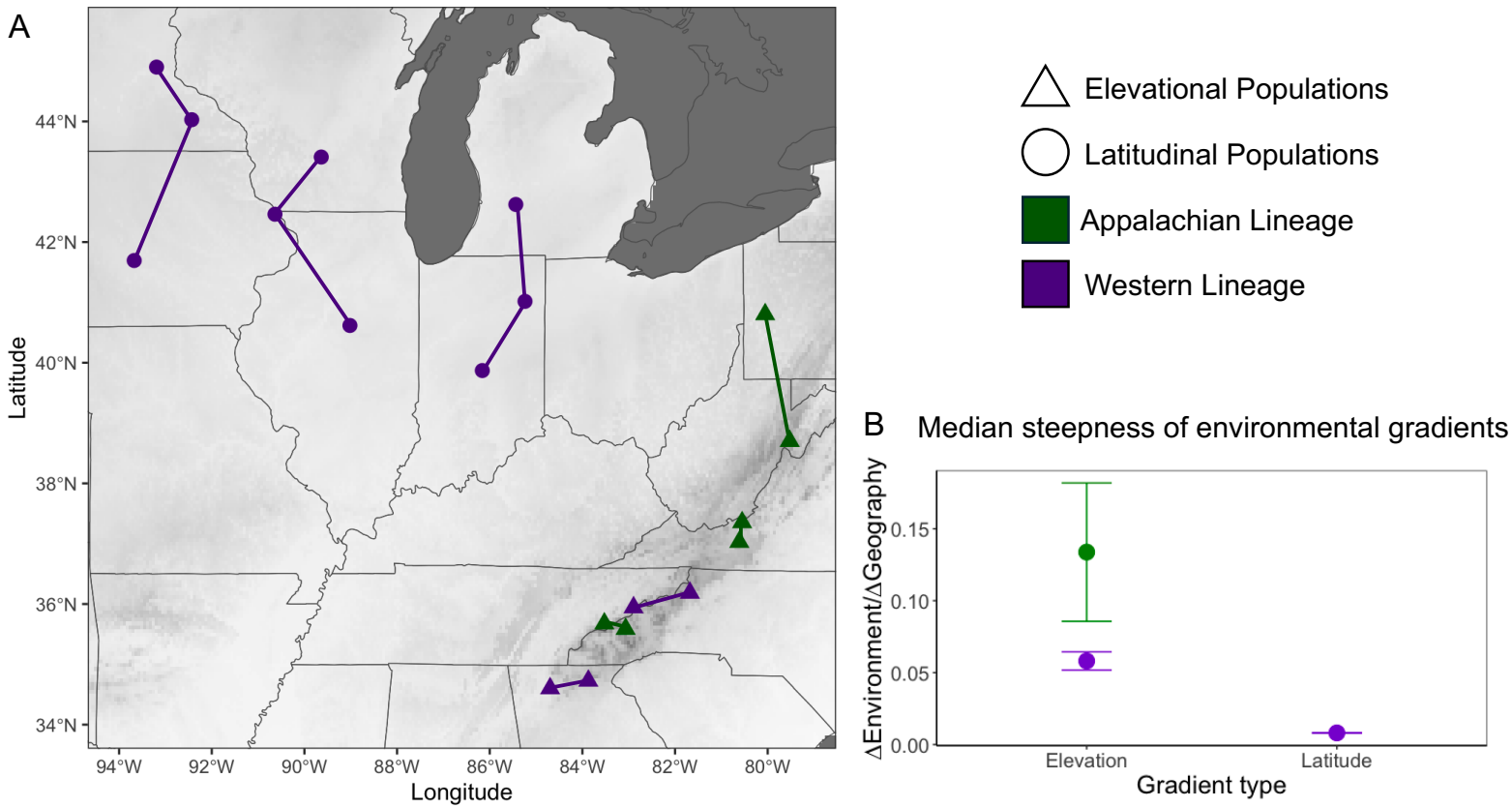


Figure 1) (A) Populations used for genomic analyses. Connections between populations signify transects used for migration analyses. Western lineage populations spanned both a latitudinal gradient near the northern range limits, and an elevational gradient in the southern Appalachian Mountains. Appalachian lineage populations are found only around the Appalachian Mountains, so are confined to the elevational gradient. (B) The median steepness of environmental gradients by lineage and gradient type. Error bars are the standard errors of median gradient estimates per transect. Environmental distance between populations was calculated using a PCA of all WorldClim 2.0 bioclimatic factors (n-=19). Units of environmental gradients are in units of PC distance divided by geographic distance in kilometers.

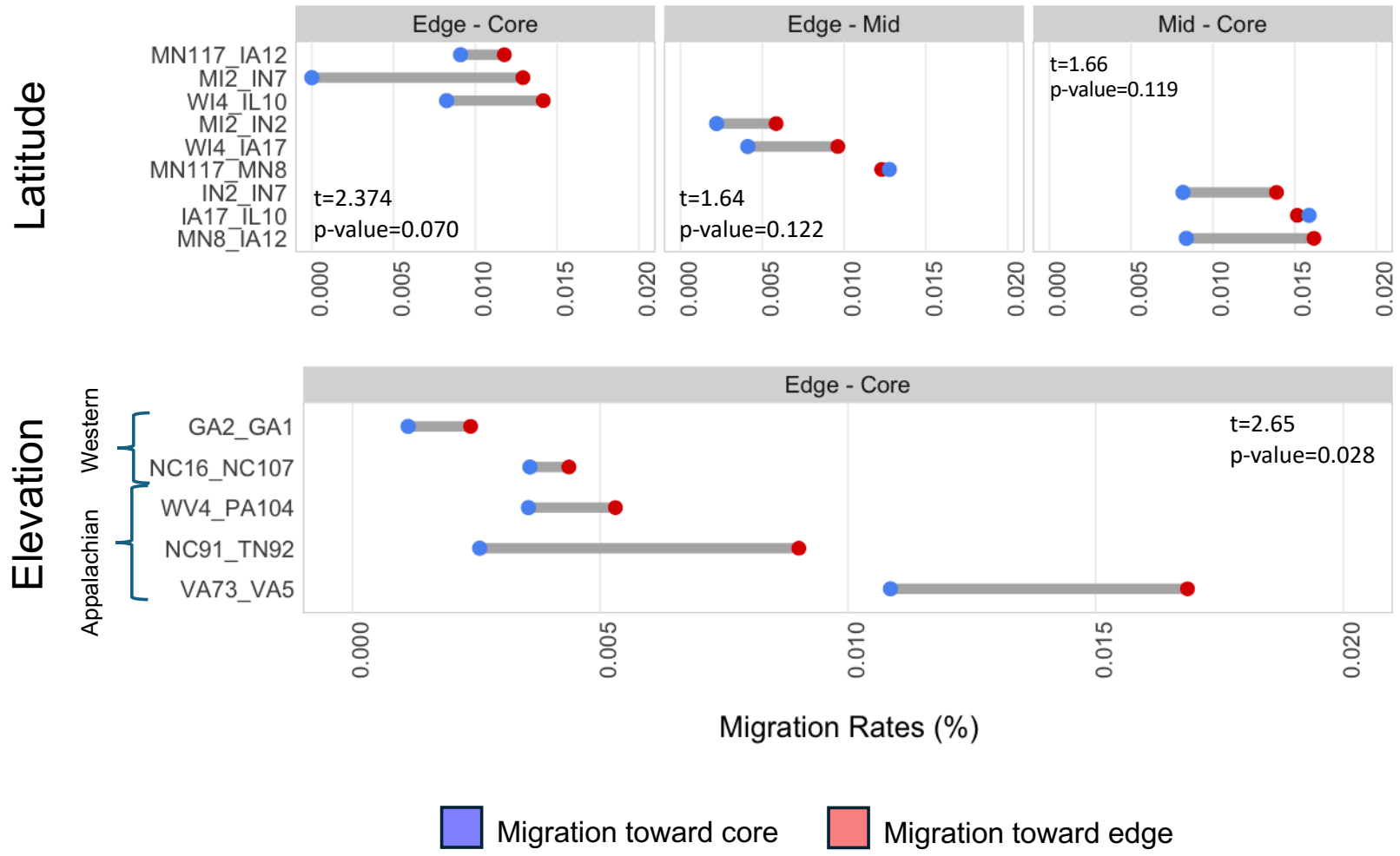


Figure 2) Pairwise migration rates derived from demographic inference models. Pairs of populations are neighboring range positions on transects along latitudinal and elevation gradients (Fig. 1). Migration is shown toward the range edge (blue) and the range core (red). Migration is asymmetric where points do not overlap within each pair of populations.

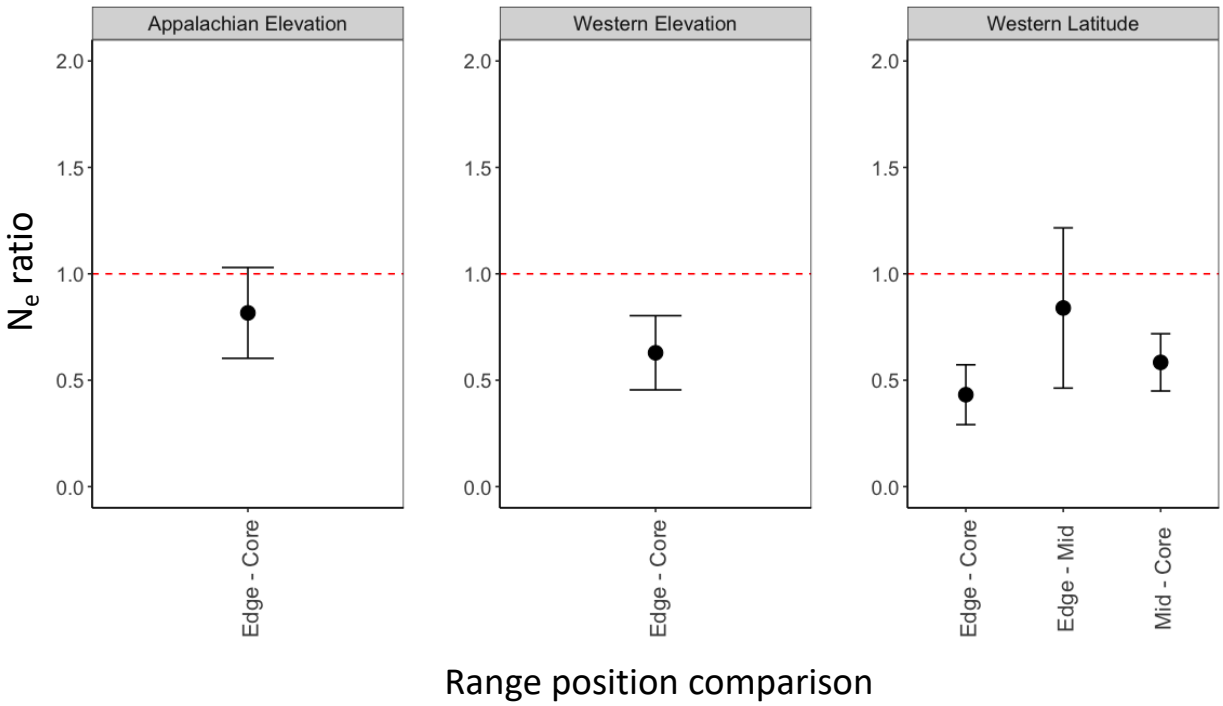


Figure 3) The ratio of effective population sizes (N_e), comparing populations further from the range core to populations nearer to the range core. Ratios <1 suggest that populations farther from the range core have smaller N_e. Ratios equal to zero suggest that populations do not differ in N_e between range positions. Error bars indicate standard errors of mean N_e ratios.

Gradient	Lineage	Test	Comparison	R^2	R	geography coefficient	geography p-value	environment coefficient	environment p-value
Latitude	**Western**	IBE/IBD	MMRR	0.02		0.1081	0.671	-0.0030	0.947
		IBE/IBD	Partial Mantel		-0.02				0.503
		IBD	Mantel		0.11		0.245		
Elevation	**Appalachian**	**IBE/IBD**	**MMRR**	**0.35**		**0.1788**	**0.014**	0.0092	0.627
		IBE/IBD	Partial Mantel		0.14				0.295
		IBD	**Mantel**		**0.58**		**0.002**		
Elevation	**Western**	IBE/IBD	MMRR	0.35		0.3909	0.423	0.0820	0.186
		IBE/IBD	Partial Mantel		0.53				0.075
		IBD	Mantel		0.32		0.158		

Table 1) Results from isolation-by-distance (IBD: Mantel) and isolation-by-environment (IBE: MMRR, Partial Mantel) analyses. Significant terms are in bold. Tests are divided by gradient (latitude/elevation) and population lineage (Western/Appalachian).

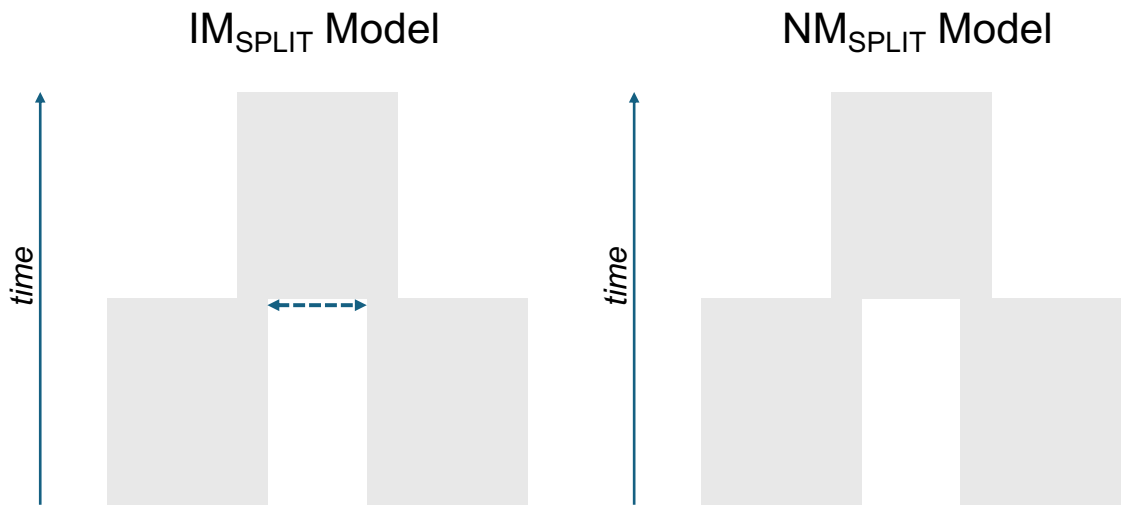
SUPPLEMENTARY METHODS

Effective geographic distance

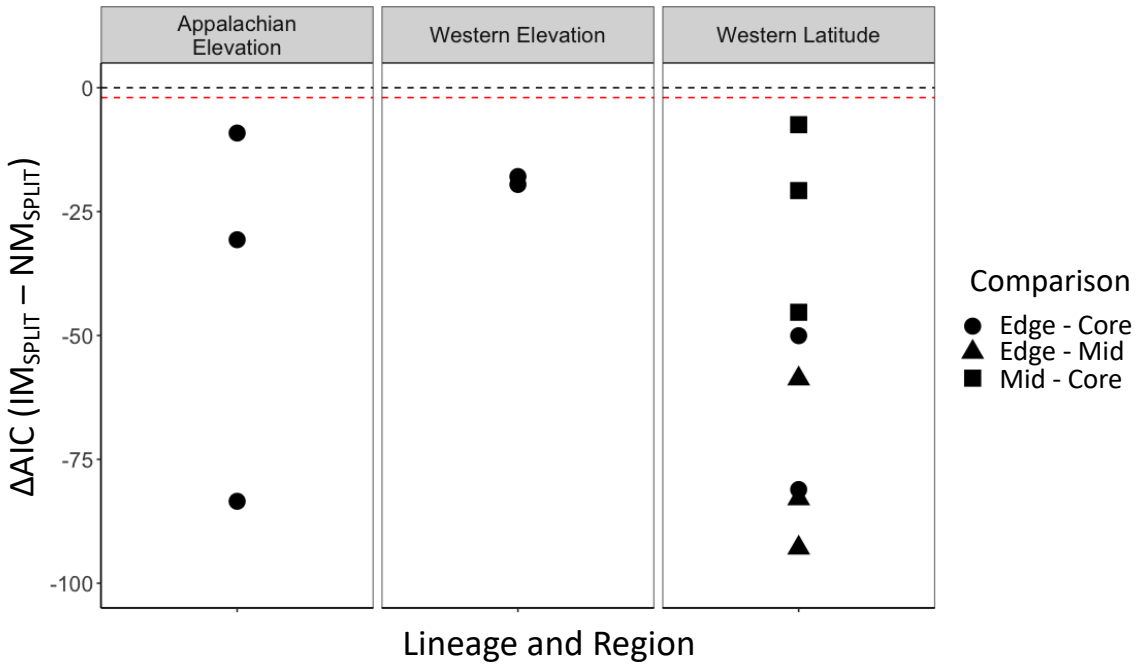
Geographic distance may fail to accurate describe connectivity between populations via migration in heterogeneous habitats. For example, populations separated by a mountain range may have short geographic distances between them even though populations are only connected around the perimeter of the mountains. In such cases, distance through connected, i.e. suitable, habitat may better estimate the influence of neutral processes like genetic drift as it represents the distance of migration that would underlie gene exchange. Here, I developed a procedure to calculate the distance between two points in a species distribution that traverses suitable habitat. I used *Campanula americana* as a model system and determined the shortest path between points through the ecological niche determined with a species distribution model (SDM; Appendix 1). I call the metric 'effective geographic distance'.

To calculate distances, I generated friction rasters using the SDM which penalize movement through less suitable habitat. I use the rasters to measure potential migration distance through contiguous range space. To compute friction rasters, I obtained rasterized predictions of the probability of occurrence from the SDM. I adjusted prediction data to include a minimum prediction of 0.01 to ensure that isolated populations which exist on islands of suitable habitat could still be reached. I translated prediction rasters to transition layers, allowing nonsymmetric 16-directional movement, and geo-corrected the layer to account for bias in map projection using the transition and geoCorrection functions from gdistance (van Etten, 2017). Finally, I calculated the shortest path in the friction raster between populations using the function shortestPath in the gdistance package (van Etten, 2017), then found the length of the path using the SpatialLinesLength function from the sp package (Pebesma and Bivan, 2005). Briefly, shortest

paths assess movement across the friction raster by penalizing movements in vectors with lower predictive scores. Paths circumnavigate obstacles where predictions are lower, thereby maximizing passage through contiguous habitat.



SI Figure 1) Demographic inference models tested to determine evolutionary history and asymmetry of migration rates. For migration (IM$_{SPLIT}$) models, migration was not restrained to a specific time interval and was continuous from the time of population divergence. Dashed line for migration (IM$_{SPLIT}$) model indicates migration. No migration (NM$_{SPLIT}$) models assumed no migration at any point in time.



SI Figure 2) Differences in AIC scores for migration (IM_{SPLIT}) and no migration (NM_{SPLIT}) models (ΔAIC). The red dashed line shows the ΔAIC=-2 line while the black dashed line shows the ΔAIC=0 line. Negative values of ΔAIC indicate better fit of migration models compared to no migration models.

population	gradient	lineage	transect	range group	latitude	longitude	elevation (m)
IA12	latitude	Western	1	core	41.69	-93.67	246
MN8	latitude	Western	1	mid	44.03	-92.43	308
MN117	latitude	Western	1	edge	44.90	-93.19	213
IL10	latitude	Western	2	core	40.62	-89.02	229
IA17	latitude	Western	2	mid	42.46	-90.64	205
WI4	latitude	Western	2	edge	43.41	-89.64	260
IN7	latitude	Western	3	core	39.87	-86.16	228
IN2	latitude	Western	3	mid	41.02	-85.24	231
MI2	latitude	Western	3	edge	42.62	-85.44	291
GA1	elevation	Western	4	core	34.60	-84.70	213
GA2	elevation	Western	4	edge	34.73	-83.87	1078
NC107	elevation	Western	5	core	35.94	-82.90	382
NC16	elevation	Western	5	edge	36.19	-81.68	1031
PA104	elevation	Appalachian	6	core	40.80	-80.05	289
WV4	elevation	Appalachian	6	edge	38.70	-79.52	1295
VA5	elevation	Appalachian	7	core	37.28	-80.61	496
VA73	elevation	Appalachian	7	edge	37.35	-80.55	1044
TN92	elevation	Appalachian	8	core	35.68	-83.53	480
NC91	elevation	Appalachian	8	edge	35.59	-83.07	1456

SI Table 1) Metadata information for locations and elevations of all populations sequenced.

gradient	lineage	population1	population2	projection1	projection2	comparison
latitude	Western	IN2	IN7	16	14	mid - core
latitude	Western	IA17	IL10	18	14	mid - core
latitude	Western	MN8	IA12	18	10	mid - core
latitude	Western	MI2	IN2	16	16	edge - mid
latitude	Western	WI4	IA17	10	10	edge - mid
latitude	Western	MN117	MN8	14	14	edge - mid
latitude	Western	MI2	IN7	16	16	edge - core
latitude	Western	WI4	IL10	10	14	edge - core
latitude	Western	MN117	IA12	14	10	edge - core
elevation	Western	GA2	GA1	12	12	edge - core
elevation	Western	NC16	NC107	12	12	edge - core
elevation	Appalachian	NC91	TN92	14	14	edge - core
elevation	Appalachian	WV4	PA104	14	14	edge - core
elevation	Appalachian	VA73	VA5	16	16	edge - core

SI Table 2) Metadata information for the 2N projection sizes of each population for demographic inference. Pairs of populations for which demographic simulations were run are shown by row (e.g., IN2 – IN7). Analyses included 12 pairs of populations and two demographic models for each pair of populations (IM_{SPLIT} and NM_{SPLIT}).

	nu1	nu2	Ts	m12	m21
upper bound	10	10	5	20	20
lower bound	0.001	0.001	0.001	0.001	0.001

SI Table 3) Ranges used to inform uniform starting parameter distributions for demographic simulations. For each iteration, a new set of parameters was drawn at random from the distributions. M12 and m21, migration parameters, were not used for NM_{SPLIT} models.

CHAPTER 5:

Phylogeography and paleoclimatic range dynamics explain variable outcomes to contact across a species' range

Co-authored with contributions from the work and dissertation of Catherine L. Debban, Ph.D.

Contributions are noted in the text, as well as figures and tables adapted from Catherine Debban's work:

Debban, C. L. 2019. Reproductive isolation and gene flow vary among contact zones between incipient species. University of Virginia.

ABSTRACT

Replicability of speciation and the maintenance of divergence after contact are poorly characterized processes, particularly in context of phylogeography and post-glacial range dynamics. Using contact zones located at the leading- and rear-edges of a species' range, I examined variation in outcomes to contact between divergent lineages of *Campanula americana*. I investigated whether contact zones vary in quantity and directionality of gene flow, how phylogeographic structure differs between contact zones, and how historic range dynamics may affect outcomes to contact. I found that all contact zones formed at similar times via primary contact yet detected significant admixture in only the rear-edge contact zone. In the northerly leading-edge contact zone and the mid-range Virginia contact zone, gene flow was minimal and asymmetric. In the southern rear-edge contact zone, gene flow was strong and symmetric. My results emphasize the dependence of speciation processes on phylogeographic structure, demographic history, and paleoclimatic range dynamics. My results suggest that caution need be taken when treating species as cohesive or uniform evolutionary units.

INTRODUCTION

Species often exhibit complex patterns of phylogeographic structure and intraspecific patterns of reproductive isolation across their respective ranges (Mandeville et al. 2015; Barnard-Kubow et al. 2017; Van Riemsdijk et al. 2023; Sianta et al. 2024). For example, in *Catostomus spp.,* reproductive isolation was found to vary strongly among geographically-isolated rivers (Mandeville et al. 2015). Similarly, other studies examining contact zones between divergent species of Sepsid flies (Giesen et al. 2023), intertidal macroalgae (Hoarau et al., 2015), and toads (Van Riemsdijk et al. 2023) have all independently found evidence of variable introgression and reproductive isolation among closely-related species. Yet, the phylogeographic contexts that presage the development of intraspecific genetic divergence is often ignored in speciation literature (but see: Hewitt, 1996, 2000), leaving broad gaps in our understanding of how range dynamics inform both intraspecific patterns of phylogeographic divergence and broader processes of interspecific speciation (Cutter 2015).

During range expansions, populations near the expanding range front, or leading edge, experience serial founder events as the range boundary is pushed forward (Slatkin and Excoffier 2012). Such expansions can produce strong genomic signatures including reduced genetic diversity and increased genetic load (Petit et al. 2004; Koski et al. 2019b). In contrast, populations near the relictual rear edge (i.e., the rear edge) may serve as genetic reservoirs, maintaining greater genetic diversity (Hampe and Petit 2005). However, as climates shift, suitable habitat near the rear edge may contract through habitat fragmentation and degradation, resulting in population bottlenecks and stronger selection for climate adaptation (Dynesius and Jansson 2000; Hampe and Petit 2005). Across both the leading- and rear-edges of a species' range, range expansions and

contractions can restrict gene flow among populations in different portions of the range, crystallizing nascent genetic differentiation.

When divergent lineages come into contact, latent reproductive isolation is exposed. If hybrids are less fit than parents, selection against introgression may reinforce reproductive barriers and enhance isolation within the contact zone (Hoskin et al. 2005; Liao et al. 2019). Yet, reinforcement is not an inevitable conclusion of contact. Outcomes to contact are influenced by, and sensitive to, initial conditions, including the magnitude of reproductive isolation (Bank et al. 2012), coupling of reproductive barriers (Butlin and Smadja 2018; Kulmuni et al. 2020), nature of the origin of the contact (Pettengill and Moeller 2012; Hudson et al. 2020; Johannesson et al. 2020), geographic structure of contact zones (Cain et al. 1999; Sadedin and Littlejohn 2003; Corbett-Detig et al. 2013), and genomic architecture of reproductive incompatibilities (Gavrilets et al. 2000; Bank et al. 2012; Lindtke and Buerkle 2015).

Importantly, multiple factors influencing outcomes to contact can interact to produce heterogeneous distributions of reproductive isolation across a species' range. For example, where the genomic architecture of reproductive isolation is predicated on Dobzhansky-Muller incompatibilities (DMIs), i.e., negative epistasis of neutral loci, multiple independent incompatible loci can produce multiplicative increases in total reproductive isolation (Moyle and Nakazato 2010). DMIs are thought to arise as a function of neutral genetic divergence (Orr 1995), so likely vary across species' ranges as a product of geographic and genetic isolation. Yet, individual DMIs are often weak barriers to reproduction (Orr 1995). In some special cases, like cytonuclear incompatibilities, barriers to reproduction are also functionally asymmetric (Sloan et al. 2018), producing permeable reproductive barriers that impede gene flow between lineages in only one direction. Consequently, regional differences in the quantity, nature, and accumulation of

reproductive barriers may strongly influence the downstream outcomes of contact between divergent lineages.

Areas of species ranges where divergent lineages are in contact may arise from parapatric speciation and primary contact, or range expansion and secondary contact. In primary contact zones, populations of diverging lineages remain in parapatry, with no period of isolation between divergence and contact. Additionally, primary contact zones typically form when environmental selection favors segregating loci (Momigliano et al. 2017; Johannesson et al. 2020). In contrast, secondary contact zones occur when divergent lineages meet following a period of geographic isolation. In secondary contact zones, divergence can be driven by neutral genetic processes related to genomic architecture (Coyne and Orr 2004; Hanzl et al. 2014), such as negative epistasis (Orr 1995; Moyle and Nakazato 2010; Hill et al. 2019); or by extrinsic environmental factors, such as divergent selection (Berg et al. 2015; Guo et al. 2015; Akopyan et al. 2020). Such differences may influence both the genetic background of populations in contact and the architecture of reproductive isolation.

Together, previous work suggests that outcomes to contact may differ across contact zones due to phylogeographic context and the evolutionary history of populations in contact. Yet, few studies have sought to directly compare how outcomes of contact may vary between lineages of a species, nor whether variation in outcomes is associated with paleoclimatic range dynamics (but see Hewitt, 1996, 2000, 2004). Here, I investigate multiple contact zones between lineages of *Campanula americana*, a North American wildflower, to ask: (1) whether contact zones vary in the quantity and directionality of gene flow, (2) how the phylogeography of populations in contact differs between contact zones, and (3) how paleoclimatic history may affect outcomes to contact.

To answer these questions, I sample populations from three disjunct contact zones found near the leading- and rear-edges of the species' range and evaluate range-wide and contact zone-specific patterns of population structure. I then assess the quantity and direction of gene flow between lineages at each contact zone. Finally, I investigate phylogenetic and demographic history of lineage contact each contact zone to infer their relative age and nature of origin. I predict that (1) populations in contact near the relictual rear edge will harbor greater between-lineage gene flow and admixture; (2) the mode of contact will differ across the species' range, with recent secondary contact in the northern leading-edge contact zone and ancient contact in the rear-edge contact zone; (3) phylogeographic patterns of range expansion and lineage divergence will explain downstream outcomes of contact.

METHODS

Study System

Campanula americana is a monocarpic autotetraploid herb distributed across much of the eastern United States. *C. americana* primarily outcrosses and is insect-pollinated (Galloway et al. 2003; Koski et al. 2019a). Polyploidization in the species is likely ancient, occurring between the divergence with the nearest known diploid species, *Triodanis perfoliata* (11.78 m.y.a.; Gadella 1966; Mansion et al. 2012), and divergence of the most basal intraspecific lineages within *C. americana* (approx. 2.3 m.y.a.; Barnard-Kubow et al. 2015), which are tetraploid. Further, the tetraploid chromosome number (Gadella 1966) is less than expected by the haploid chromosome number (Rogers 1965), suggesting loss of several chromosomes.

C. americana is generally divided into three lineages distinguished by geographic and reproductive isolation (Fig. 1; Barnard-Kubow et al., 2015, 2017). The Appalachian lineage is

largely constrained to the Appalachian Mountains between the leading- and rear-edge contact zones located in Pennsylvania and North Carolina, respectively (Fig. 1B). The Western lineage is cosmopolitan, found across much of the midwestern and southeastern United States. The Eastern lineage is found primarily east of the Appalachian Mountains in Virginia and is likely derived from relictual southeastern populations of the Western lineage (Barnard-Kubow et al. 2015). Hybrids between Appalachian and Western lineages have reduced germination and asymmetric survival, indicating strong nuclear-nuclear and cytonuclear incompatibility (Barnard-Kubow et al. 2016, 2017; Debban 2019). Reduced hybrid survival is primarily driven by chlorosis, the loss of photosynthetic function, of plants with Western chloroplasts on hybrid nuclear backgrounds. Hybrids with Appalachian maternal backgrounds rarely expressed chlorosis, suggesting gene flow may be asymmetric in contact zones where the Appalachian lineage is present (Barnard-Kubow et al. 2016; Debban 2019).

The split between the ancestral *C. americana* lineages occurred approximately 2.3 mya (Barnard-Kubow et al. 2015), producing distinct clades that eventually gave rise to the contemporary Appalachian and Western lineages. The initial split was followed by additional within-lineage cladogenesis and divergence. During the last glacial maximum, Western-lineage populations became geographically isolated in disjunct refugia that seeded post-glacial range expansion in different geographic regions (Appendix 1; Barnard-Kubow et al. 2015). Western-lineage populations isolated in a mid-latitude refugium located in Kentucky expanded north- and westward after the last glacial maximum, invading much of the Western lineage's contemporary range (Prior et al. 2020). In contrast, Western-lineage populations isolated in southeastern refugia near the Gulf of Mexico appear to have undergone limited postglacial range expansion. Similarly, the Eastern and Appalachian lineages have likely undergone limited postglacial range expansion

(Appendix 1). While little is known about postglacial range dynamics of the Eastern lineage, the Appalachian lineage likely persisted through the last glacial maximum in microrefugia scattered throughout and in proximity to the Appalachian Mountains (Barnard-Kubow et al. 2015).

Contact zones between populations from the Appalachian and Western lineages have been found in North Carolina (NC), the rear-edge (RE) contact zone, and Pennsylvania (PA), the leading-edge (LE) contact zone (Fig. 1A). A contact zone between the Eastern and Appalachian lineages has also been found in Virginia (VA). While these contact zones demarcate known regions of contemporary geographic sympatry, additional sampling across the breadth of the Appalachian Mountains suggests that the mountain range may act as a broad mosaic of contact between lineages (Barnard-Kubow, unpublished data). Timing and origin of the rear-edge, Virginia, and leading-edge contact zones is unknown, though reproductive isolation varies among them (Debban 2019). Multiple intrinsic reproductive barriers have been found between the Appalachian and Western lineages at the leading-edge contact zone; yet only an asymmetric cytonuclear incompatibility exists in the rear-edge contact zone (Debban 2019).

Population Sampling & DNA Sequencing

Sampling, sequencing, and data processing were performed as a part of dissertation work by Catherine Debban (2019). In brief, populations were sampled at each contact zone and from allopatric habitat neighboring contact zones for use in genetic analyses. Contact zones were defined as regions where populations of different lineages were found within 80km of each other. In total, 29 populations were sampled (Fig. 1A, SI Table 1): 12 Western (4 Allopatric, 5 RE, 3 LE), 12 Appalachian (3 Allopatric, 4 RE, 2 VA, 3 LE), and 5 Eastern (3 Allopatric, 2 VA). Population lineage was determined via chloroplast genotype (Barnard-Kubow et al. 2015). In total,

326 individuals were sampled, including 19 individuals from *Triodanis perfoliata*, a close relative of *C. americana* (diverged 11.78mya; Mansion et al., 2012), for use as an outgroup.

DNA was harvested from leaf samples. Field-collected seeds were germinated according to protocols outlined in Chapter 1, then leaf tissue was collected from the seedlings. DNA extractions were performed using a modified CTAB chloroform extraction protocol (Barnard-Kubow et al. 2017). Samples were randomized within plates, then DNA was digested with the enzymes ApoI HF and SphI HF simultaneously (NEB, Inc.). After digestion, samples were multiplexed with a unique combination of indices (SI Fig. 1; SI Table 2). After, standard library prep was followed according to Peterson et al. (2012). Libraries were sent to BGI Genomics (Cambridge, MA) for sequencing on two lanes of Illumina HiSeq 4000.

Data processing

Sequencing produced a median of 1.2 million reads per individual. Of these, an average of 3% of reads aligned to chloroplast sequences and 10% aligned to mitochondrial sequences. Samples in the bottom 5th percentile of read number were removed (less than 15,000 reads per individual). Given the ancient origin polyploidization within *C. americana,* paralogous RAD loci are unlikely to bias downstream population genetic analyses if assembled as a putative diploid, while artificial reductions in allelic diversity associated with paralog filtering can inflate estimates of homozygosity (Ilut et al. 2014). Consequently, sequences were assembled under a model of diploidy. To assemble sequences, reads were filtered and assembled *de novo* using the program STACKS/1.48 (Julian M Catchen et al. 2011; Julian Catchen et al. 2013). Optimal STACKS parameters (-m 8 -M 4 -n 4) were determined by running STACKS on a subset of 13 samples, permuting each parameter (-m tested at 4 and 8, -M and –n tested at 0-9 and held equal). Parameters

were selected to maximize polymorphism without combining loci (Paris et al. 2017). Finally, loci were retained if present in at least 25 of the 29 populations. Coverage averaged 27.4 reads per site ($\sigma = 11.3$) across samples.

Patterns of genomic inheritance can vary within individuals. Generally, while nuclear DNA is subject to gene flow and recombination, mitochondrial and chloroplast DNA are uniparentally inherited and not subject to recombination, thereby preserving patterns of neutral genetic divergence among lineages. To analyze parts of the genome with different inheritance patterns, I created alternative sets of reads. These included: a set of reads aligned to either the chloroplast or mitochondrial genomes (cytoplasmic dataset), and a set of reads that did not align to either the chloroplast or mitochondrial genomes (nuclear dataset). These data sets were used to generate consensus sequences for phylogenetic analysis.

I further filtered the nuclear data set to generate a table of high-quality SNPs for demographic inference and analyses of population structure (nuclear demography dataset). Specifically, I removed individuals with more than 50% missing data, then removed sites with a read depth of less than 10 and a minor allele frequency of less than 0.01 using VCFTools (Danecek et al. 2011). I allowed 20% missingness within SNPs to reduce noise while maintaining sample sizes for downstream demographic inference. Finally, I removed populations with $n \leq 3$. The filtering process resulted in a small, high-quality set of 3,568 nuclear SNPs from 26 populations. I conducted all downstream analysis in R 4.2.0 (R Core Team 2022) and python 3.9 (Python Software Foundation n.d.).

Population Structure and Gene Flow

I evaluated population structure across the species' range to assess phylogeographic patterns between and within lineages using the nuclear demography dataset. I analyzed population structure using PCA, and evaluated population-level admixture and ancestry using the snmf function in the R package LEA (Frichot and François 2015; Gain and François 2021). For analyses of admixture and ancestry, I evaluated K=1:50 with 10 repetitions of each K cluster value. To determine the optimal K-value, I first evaluated whether isolation-by-distance (IBD) was present in any lineage, as IBD can bias the optimal number of K clusters upward (Perez et al. 2018). To assess IBD, I used Mantel tests in the vegan package (Dixon 2003). Because IBD was present in multiple lineages, I selected the optimal number of K clusters by finding the K value with the smallest change in slope in the upper quartile of cross-entropy values (SI Fig. 2), indicating where information explained by additional K clusters plateaus. I confirmed this selection by computing a DAPC, then identifying the optimal number of groups to retain using a-scores (Jombart 2008). After selecting the optimal number of clusters to retain, I projected spatial interpolations of major ancestry cluster assignments using the function ggtess3Q in the R package tess3r (Caye et al. 2016).

I evaluated contemporary patterns of gene flow between lineages at contact zones using ABBA-BABA tests. To maximize my ability to detect gene flow, I assessed patterns of gene flow using the nuclear demographic dataset without a missing SNP threshold applied (20%) but all other filters, including a minor allele frequency of 0.01, in place. To conduct ABBA-BABA tests, I computed D-statistics for all between-lineage pairs of populations by contact zone. I used *Triodanis perfoliata* as the outgroup species to establish derived allele states, and allopatric populations as a within-lineage comparison for tests. I performed ABBA-BABA tests using the

Dsuite program (Malinsky et al. 2021). After, I corrected p-values to account for multiple comparisons using Benjamini-Hochberg FDR correction.

Phylogenetics

As a part of dissertation work by Catherine Debban (2019), phylogenetic history was evaluated across components of the genome with different patterns of inheritance by constructing separate phylogenies from cytoplasmic and nuclear population consensus sequences. While the cytoplasmic phylogeny reveals patterns of lineage divergence independent of between-lineage admixture, the nuclear phylogeny demonstrates the influence of admixture on patterns of population and lineage differentiation in each contact zone. As a result, clades that have experienced significant introgression may be intermixed in the nuclear phylogeny while being well-differentiated in the cytoplasmic phylogeny. Phylogenies were constructed using the program BEAST 2 (Bouckaert et al. 2014) and population-level consensus sequences from the nuclear and cytoplasmic datasets. These datasets were used as partitions with independent clock and site models. To determine site models, the function bModelTest was used (Bouckaert and Drummond 2017), and a $1x10^7$ generation MCMC chain with a coalescent constant population prior. Finally, phylogenies were time-calibrated based on an estimated divergence between *T. perfoliata* and *C. americana* 11.78 million years ago from a fossil-calibrated phylogeny of Campanulaceae (Mansion et al. 2012).

Demographic Inference

I performed demographic inference to assess the demographic and evolutionary history of populations at contact zones. Specifically, I modeled potential demographic histories consistent

with the divergence of lineages occurring during primary contact and secondary contact (Fig. 2). Models included: constant migration (IM), where lineages diverge in geographic contact and maintain migration between lineages; ancestral migration (AM), where migration begins at the time of divergence and ceases at some timepoint in the past; secondary migration (SC), where lineage divergence occurs during a phase of geographic isolation, followed by migration; and no migration (NM), where populations do not experience any migration between lineages after divergence.

While the models discussed above assume a singular episode of isolation or contact, it is plausible that contact has occurred many times as a result of glacial cycling. To incorporate this complexity, I also parameterized several cyclical contact models that integrate multiple episodes of contact and ancestral population growth (Fig. 2B). Models included: cyclic secondary migration (SC2C), which institutes two cycles of secondary contact between lineages; cyclic migration (IM2C), which institutes a primary origin with constant migration interrupted by a single intermediate episode of isolation; and ancient secondary migration (A SC), which institutes a single episode of secondary contact that ceased at some time point in the past.

Within each demographic model (e.g., SC2C, IM), I incorporated sub-models with several variant parameters to explore how changes in ancestral and derived population sizes alter model selection (Momigliano et al. 2021). Sub-models parameterized instantaneous ancestral population growth (AE) and growth of descendent populations from an initial bottleneck occurring at the time of lineage divergence (B/B(R); Fig. 2B; *models adapted from* Momigliano et al., 2021). I parameterized all demographic models with migration between lineages as asymmetric migration models, forcing migration parameters between lineages to optimize independently. I computed

demographic models using the python package Moments (Jouganous et al. 2017), which uses allele frequency spectra to evaluate model fit.

I combined populations within lineages at each contact zone to model the demographic and evolutionary history of contact zones. To maximize locus retention while maintaining dimensionality of joint allele frequency spectra, I down-projected 2N population sizes for pools of populations in each contact zone. I used *easySFS* to determine optimal 2N down-projections (https://github.com/isaacovercast/easySFS; SI Table 3). I ran 50 iterations of each demographic model for each contact zone, selecting initial parameters by randomly sampling bounded uniform distributions for each iteration (SI Table 4). I retained the best iteration by log-likelihood for each demographic model in each contact zone. I selected the best model within each contact zone by AIC weights and model scores (Rougeux et al., 2017). I calculated model scores as the ΔAIC between each model and the worst model for a given contact zone, then dividing this by the ΔAIC between the best and worst model for a given contact zone ($\Delta AIC_{model - worst}$ / $\Delta AIC_{best - worst}$). I calculated AIC weights using the MuMIn R package (Barton 2022).

I evaluated asymmetry in migration rates among lineages and the relative time of split using the best fit demographic model for each contact zone. I translated migration parameters to rates from native units of 4Neμ to compare their relative scaling and asymmetry. I translated raw parameter estimates of migration, and relevant time parameters (T_{SPLIT}, $T_{CONTACT}$) using the equations (Gutenkunst et al. 2009):

$$\text{time} = \frac{2\theta}{4\mu L} * T_s \qquad \text{migration rate}_{i \rightarrow j} = \frac{m_{1 \rightarrow 2}}{2 * \frac{\theta}{4\mu L}};$$

For μ, I selected an *ad-hoc* mutation rate of 2.8×10^{-9}. I compared directionality of between-lineage migration rates to evaluate whether gene flow is significantly asymmetric within contact zones in the directionality of known cytonuclear incompatibility, i.e., greater gene flow into

Appalachian-lineage populations. Because the actual mutation rate for *C. americana* is not known, I relativized estimates of the time of divergence and contact (T_{SPLIT}, $T_{CONTACT}$) using the range-wide mean (i.e., across contact zones) for each parameter.

RESULTS

Population Structure and Gene Flow

Populations across the species' range were structured by lineage and geography (Fig. 3, SI Fig. 3, 4). Isolation-by-distance was significant for Appalachian (p-value=0.001, R=0.46) and Western-lineage populations (p-value=0.008, R=0.37) across the species' range, but was not significant for Eastern-lineage populations (p-value=0.17). Because of significant isolation-by-distance identified across the range, I used a conservative approach to selecting the optimal number of K-clusters. For tests of population-level ancestry and admixture, the optimal number of K-clusters, determined by DAPC a-score and cross-entropy quantile, was six.

Populations in the leading-edge and Virginia contact zones clustered within their respective lineages, though Appalachian-lineage populations tended to exhibit more population sub-structuring (Fig. 3, SI Fig. 3). At the rear-edge contact zone, populations are more genetically homogeneous between lineages than elsewhere (Fig. 3, SI Fig. 3). In concert with this finding, ABBA-BABA tests predicted significant gene flow between lineages at the rear-edge contact zone, but little to no gene flow among populations of differing lineages at the leading-edge and Virginia contact zones (Fig. 3B).

Phylogenetics

The phylogeny built from the nuclear consensus sequences (Fig. 4B) provided higher resolution of relationships among populations than the cytoplasmic phylogeny (Fig. 4A) but was discordant with it for populations located in the rear-edge contact zone. For clades containing populations from the leading-edge and Virginia contact zones, the nuclear and cytoplasmic phylogenies were concordant. For rear-edge populations, the nuclear phylogeny placed populations from the Western and Appalachian lineages in the same clade. In contrast, the cytoplasmic phylogeny separated populations into highly divergent Appalachian and Western lineages (Fig. 4A). In the nuclear phylogeny, the Appalachian and Western lineages were further subdivided by geography. The Western lineage was split into two clades: one composed of populations from across the Western range, and one composed solely of populations from the rear-edge and Virginia (Fig. 4A). Similarly, the Appalachian lineage was split into two clades split along latitudes—one clade in the northern Appalachian Mountains near the leading edge and one clade in the southern Appalachian Mountains near the rear edge. Finally, the Eastern lineage was rooted as a basal branch to the Western lineage in the cytoplasmic tree.

Demographic Inference

I evaluated models of demographic history to infer the evolutionary history of lineages in each contact zone. Across contact zones, the best model by AIC weights, raw AIC, and model score was population divergence with constant migration and ancestral population expansion (IM_{AE}; Fig. 5A, SI Fig. 5). The relative time of divergence between lineages differed by contact zone, with the time of divergence between lineages being younger at the leading-edge contact zone than either the Virginia or rear-edge contact zones. Asymmetric migration favored Western and

Eastern introgression into Appalachian-lineage populations at the leading-edge and Virginia contact zones (Fig. 5B). In the rear-edge contact zone, migration rates between lineages were functionally symmetric. In keeping with this finding, estimates of migration rates into Appalachian populations were modest across contact zones, but migration rates from Appalachian populations into Eastern and Western populations were low only in the leading-edge and Virginia contact zones (SI Fig. 6).

DISCUSSION

Outcomes to contact may vary between lineages and across species' ranges depending on the evolutionary and demographic history of the populations in contact. I used multiple contact zones between differentiated lineages as natural replicates to explore whether and how outcomes to contact vary, and whether historical range dynamics explains variation in outcomes. I found that populations at all contact zones have admixed since the time of lineage divergence, though at very different rates. In the leading edge and Virginia contact zones, admixture between lineages is modest and migration over evolutionary timescales has been highly asymmetric, following the direction of a known cytonuclear incompatibility. In the rear-edge contact zone, admixture is substantial and migration over evolutionary timescales has been symmetric. These results indicate that outcomes to contact can be strongly affected by the relative positions of contact zones within species' ranges. Specifically, contact zones near leading range edges may be more likely to produce reinforcement and maintain genetic divergence between lineages than contact zones near relictual or ancestral rear edges. My findings demonstrate intraspecific variation in speciation potential is predicated on intraspecific phylogeography divergence and dynamics of postglacial range expansion.

Outcomes to contact

Contact between divergent lineages within a species can reinforce existing reproductive barriers, driving speciation, or eliminate reproductive barriers through gene flow and admixture, driving lineage homogenization. The probability of either outcome largely depends on the underlying genetics and demographics of populations in contact. In the leading-edge and Virginia contact zones, I found little evidence of contemporary gene flow between lineages. Additionally, historic patterns of migration between lineages are strongly asymmetric in demographic models in these contact zones, favoring Western and Eastern migration into the Appalachian lineage but not vice versa. These results are congruent with previous work identifying a cosmopolitan cytonuclear incompatibility between lineages that impedes gene flow from the Appalachian lineage into the Eastern and Western lineages (Barnard-Kubow et al. 2016; Debban 2019). Conversely, in the rear-edge contact zone, lineage divergence was not maintained. Instead, I found strong evidence of contemporary gene flow and historic patterns of symmetric migration between lineages, suggesting ongoing and extensive admixture between lineages.

Across contact zones, the best model of lineage divergence was constant migration with ancestral population expansion (IM_{AE}), supporting a dynamic of admixture between lineages since the time of divergence. These results are perhaps not surprising, considering the lineage are within differentiated groups in a single species, and reproductive isolation between the lineages is incomplete (Barnard-Kubow et al. 2016; Debban 2019). However, it is not clear whether this dynamic indicates true primary contact, with lineages diverging in parapatry with gene flow, or repeated episodes (>2) of secondary contact. Given the relatively old time of lineage divergence (2.3mya; Barnard-Kubow et al., 2015), weak signatures of secondary contact could be obscured

by repeated episodes of contact induced by glacial cycling over evolutionary timescales.

Spatial heterogeneity in reproductive isolation between and within contact zones

Outcomes to contact can vary between lineages depending on the evolutionary history of populations in contact. Evolutionary history is particularly relevant when reproductive isolation between lineages is the product of cytonuclear conflict and other epistatic (i.e., Dobzhansky-Muller) incompatibilities. Epistatic incompatibilities can arise quickly and are typically neutral on their own respective genetic background (Orr 1995; Burton et al. 2013; Sloan et al. 2017), facilitating heterogeneous fixation and accumulation within lineages. Reproductive isolation may also vary within lineages due to different rates of accumulating reproductive incompatibilities (Orr 1995; Moyle and Nakazato 2010).

Heterogeneous accumulation of reproductive isolation within lineages seems particularly likely in *C. americana*, given geographically-structured genetic divergence within lineages. For example, the cytoplasmic phylogeny reveals deep within-lineage divergence in the Appalachian lineage into separate northern and southern clusters; while more recent divergence is apparent within the Western lineage in a northerly expansion cluster (Prior et al. 2020) and a relictual southern cluster. Trends of geographically heterogeneous reproductive isolation has been reported in other work in *C. americana* (Barnard-Kubow et al. 2016), finding that isolation increased between lineages as a function of chloroplast genetic distance and geographic distance.

Variation in outcomes to contact can also vary within contact zones due to asymmetric reproductive isolation. Asymmetries are frequently due to cytonuclear incompatibilities (Turelli and Moyle 2007; Sloan et al. 2017), which promote unidirectional admixture between lineages (Burton et al. 2013). Across taxa, asymmetric reproductive isolation appears to be a common

phenomenon, having been found in numerous species of flowering plants (Tiffin et al. 2001), freshwater fish (Bolnick et al. 2008), and insects like *Drosophila* (Clancy et al. 2011); though the importance of asymmetric cytonuclear incompatibilities to speciation processes is contested (Burton 2022). I find some signature of incomplete asymmetric reproductive isolation in models of evolutionary history across contact zones. Introgression into Appalachian-lineage populations is similar across contact zones over evolutionary timescales, however, introgression in the opposing direction (i.e., Appalachian into Eastern and Western), was strongly limited in the leading-edge and Virginia contact zones. Differences in historic patterns of migration between lineages across contact zones suggests that permeability of asymmetric reproductive barriers in *C. americana* varies widely across geographic scales.

Synthesis and summary

Taken together, my data provide robust support for variable outcomes to contact among three distinct lineages of *C. americana*. Coupled with previous work demonstrating differing levels of nuclear-nuclear and cytonuclear reproductive isolation across contact zones (Debban 2019), it is apparent that spatial heterogeneity in the accumulation of reproductive isolation influences downstream speciation potential in a phylogeographic context. Variable outcomes to contact between lineages and across species' ranges may have complex ramifications for our understanding of range dynamics as well as micro- and macroevolutionary transitions (Cutter 2015). In particular, differences in outcomes to contact between intraspecific lineages highlights the complexity of how variation within species' ranges informs speciation outcomes. Where lineages come into contact, it is clear that processes that sustain and reinforce divergence are dependent on local makeup of environmental and evolutionary histories of populations in contact.

For example, at ecological spatial scales, introgression between divergent lineages of *Clarkia* is significantly correlated with greater fluctuations of spring precipitation (Sianta et al. 2024). At evolutionary timescales, variance in admixture and reproductive isolation has been found even in different tributaries of the same major river system (Mandeville et al. 2017).

Together, within-lineage divergence over geographic space and evolutionary time opens the door for within-lineage variation in the generation and accumulation of reproductive isolation. Variation in reproductive isolation within lineages presents exciting opportunities for studying how complex evolutionary histories within- and between-lineages influence speciation. However, as noted by Hewitt (1996, 2000), patterns of range dynamics linked with phylogeographic history are difficult to fit into standard models of speciation. Yet, it is increasingly clear that many species are governed by relictual influences of such dynamics. In grasshoppers, geographic subdivision during glacial cycling drove divergence among lineages (Hewitt 1996). In the common frog, *Rana temporaria*, genetic divergence and reproductive isolation among lineages is associated with geographic subdivision of glacial refugia and cytonuclear incompatibility among lineages (Dufresnes et al. 2020). Across species, glacial cycling is hypothesized to be the primary source of patterns of increased intraspecific genetic diversity toward southerly latitudes (Fonseca et al. 2023). Thus, it has become increasingly evident that, to understand speciation outcomes, we must wrangle with complexities of the process introduced by historic and intraspecific range dynamics.

In an era of biological research that increasingly recognizes both the value and degree of intraspecific biodiversity (Dufresnes et al. 2023; Vences et al. 2024), it has become increasingly difficult to shy away from the uncomfortable reality that species are not holistic units with uniform evolutionary trajectories. Instead, resolving standing questions in the field, particularly how phylogeography influences modes of speciation and how the relationship between hybrid zone

outcomes influences downstream speciation potential (e.g., Santini et al., 2012), will demand clear understandings of phylogenetic relationships among populations in contact and robust understandings of the species' natural history. Studies examining the relationships between lineages across species' ranges can provide important insight into how intraspecific phylogeographic variance leads to speciation.

LITERATURE CITED

Akopyan, M., Z. Gompert, K. Klonoski, A. Vega, K. Kaiser, R. Mackelprang, E. B. Rosenblum, and J. M. Robertson. 2020. Genetic and phenotypic evidence of a contact zone between divergent colour morphs of the iconic red-eyed treefrog. Mol. Ecol. 29:4442–4456.

Bank, C., J. Hermisson, and M. Kirkpatrick. 2012. Can reinforcement complete speciation? Evolution 66:229–239.

Barnard-Kubow, K. B., C. L. Debban, and L. F. Galloway. 2015. Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. Am. J. Bot. 102:1842–1853.

Barnard-Kubow, K. B., M. A. McCoy, and L. F. Galloway. 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol. 213:1466–1476.

Barnard-Kubow, K. B., N. So, and L. F. Galloway. 2016. Cytonuclear incompatibility contributes to the early stages of speciation. Evolution 70:2752–2766.

Barton, K. 2022. MuMIn: multi-model inference.

Berg, P. R., S. Jentoft, B. Star, K. H. Ring, H. Knutsen, S. Lien, K. S. Jakobsen, and C. André. 2015. Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (Gadus morhua L.). Genome Biol. Evol. 7:1644–1663.

Bolnick, D. I., M. Turelli, H. López-Fernández, P. C. Wainwright, and T. J. Near. 2008. Accelerated Mitochondrial Evolution and “Darwin’s Corollary”: Asymmetric Viability of Reciprocal F1 Hybrids in Centrarchid Fishes. Genetics 178:1037–1048.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10:e1003537.

Bouckaert, R. R., and A. J. Drummond. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17:42.

Burton, R. S. 2022. The role of mitonuclear incompatibilities in allopatric speciation. Cell. Mol. Life Sci. 79:103.

Burton, R. S., R. J. Pereira, and F. S. Barreto. 2013. Cytonuclear Genomic Interactions and Hybrid Breakdown. Annu. Rev. Ecol. Evol. Syst. 44:281–302.

Butlin, R. K., and C. M. Smadja. 2018. Coupling, Reinforcement, and Speciation. Am. Nat. 191:155–172.

Cain, M. L., V. Andreasen, and D. J. Howard. 1999. Reinforcing selection is effective under a relatively broad set of conditions in a mosaic hybrid zone. Evolution 53:1343–1353.

Caye, K., T. M. Deist, H. Martins, O. Michel, and O. François. 2016. TESS3: fast inference of spatial population structure and genome scans for selection. Mol. Ecol. Resour. 16:540–548.

Clancy, D. J., G. R. Hime, and A. D. Shirras. 2011. Cytoplasmic male sterility in Drosophila melanogaster associated with a mitochondrial CYTB variant. Heredity 107:374–376.

Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Hartl, and J. F. Ayroles. 2013. Genetic incompatibilities are widespread within species. Nature 504:135–137.

Coyne, J., and H. A. Orr. 2004. Speciation. Sinauer Associates.

Cutter, A. D. 2015. Repeatability, ephemerality and inconvenient truths in the speciation process. Mol. Ecol. 24:1643–1644.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Debban, C. L. 2019. Reproductive isolation and gene flow vary among contact zones between incipient species.

Dixon, P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14:927–930.

Dufresnes, C., A. G. Nicieza, S. N. Litvinchuk, N. Rodrigues, D. L. Jeffries, M. Vences, N. Perrin, and Í. Martínez-Solano. 2020. Are glacial refugia hotspots of speciation and cytonuclear discordances? Answers from the genomic phylogeography of Spanish common frogs. Mol. Ecol. 29:986–1000.

Dufresnes, C., N. Poyarkov, and D. Jablonski. 2023. Acknowledging more biodiversity without more species. Proc. Natl. Acad. Sci. 120:e2302424120.

Dynesius, M., and R. Jansson. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. Proc. Natl. Acad. Sci. 97:9115–9120.

Fonseca, E. M., T. A. Pelletier, S. K. Decker, D. J. Parsons, and B. C. Carstens. 2023. Pleistocene glaciations caused the latitudinal gradient of within-species genetic diversity. Evol. Lett. 7:331–338.

Frichot, E., and O. François. 2015. LEA: An R package for landscape and ecological association studies. Methods Ecol. Evol. 6:925–929.

Gadella, T. W. J. 1966. Some notes on the delimitation of genera in the Campanulaceae. II. Meded. Van Het Bot. Mus. En Herb. Van Rijksuniv. Te Utrecht 269:509–521.

Gain, C., and O. François. 2021. LEA 3: Factor models in population genetics and ecological genomics with R. Mol. Ecol. Resour. 21:2738–2748.

Galloway, L. F., J. R. Etterson, and J. L. Hamrick. 2003. Outcrossing rate and inbreeding depression in the herbaceous autotetraploid, Campanula americana. Heredity 90:308–315.

Gavrilets, S., H. Li, and M. D. Vose. 2000. Patterns of parapatric speciation. Evolution 54:1126–1134.

Giesen, A., W. U. Blanckenhorn, M. A. Schäfer, K. K. Shimizu, R. Shimizu-Inatsugi, B. Misof, L. Podsiadlowski, O. Niehuis, H. E. L. Lischer, S. Aeschbacher, and M. Kapun. 2023. Geographic Variation in Genomic Signals of Admixture Between Two Closely Related European Sepsid Fly Species. Evol. Biol. 50:395–412.

Guo, Y., Y. Luo, Z. Liu, and X. Wang. 2015. Reticulate evolution and sea-level fluctuations together drove species diversification of slipper orchids (*P aphiopedilum*) in S outh- E ast A sia. Mol. Ecol. 24:2838–2855.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante. 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLoS Genet. 5:e1000695.

Hampe, A., and R. J. Petit. 2005. Conserving biodiversity under climate change: the rear edge matters. Ecol. Lett. 8:461–467.

Hanzl, M., F. Kolář, D. Nováková, and J. Suda. 2014. Nonadaptive processes governing early stages of polyploid evolution: Insights from a primary contact zone of relict serpentine *Knautia arvensis* (Caprifoliaceae). Am. J. Bot. 101:935–945.

Hewitt, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58:247–276.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907–913.

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359:183–195.

Hill, G. E., J. C. Havird, D. B. Sloan, R. S. Burton, C. Greening, and D. K. Dowling. 2019. Assessing the fitness consequences of mitonuclear interactions in natural populations. Biol. Rev. 94:1089–1104.

Hoarau, G., J. A. Coyer, M. C. W. G. Giesbers, A. Jueterbock, and J. L. Olsen. 2015. Pre-zygotic isolation in the macroalgal genus *Fucus* from four contact zones spanning 100–10 000 years: a tale of reinforcement? R. Soc. Open Sci. 2:140538.

Hoskin, C. J., M. Higgie, K. R. McDonald, and C. Moritz. 2005. Reinforcement drives rapid allopatric speciation. Nature 437:1353–1356.

Hudson, J., K. Johannesson, C. D. McQuaid, and M. Rius. 2020. Secondary contacts and genetic admixture shape colonization by an amphiatlantic epibenthic invertebrate. Evol. Appl. 13:600–612.

Ilut, D. C., M. L. Nydam, and M. P. Hare. 2014. Defining Loci in Restriction-Based Reduced Representation Genomic Data from Nonmodel Species: Sources of Bias and Diagnostics for Optimal Clustering. BioMed Res. Int. 2014:1–9.

Johannesson, K., A. Le Moan, S. Perini, and C. André. 2020. A Darwinian Laboratory of Multiple Contact Zones. Trends Ecol. Evol. 35:1021–1036.

Jombart, T. 2008. *adegenet* : a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jouganous, J., W. Long, A. P. Ragsdale, and S. Gravel. 2017. Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. Genetics 206:1549–1567.

Julian Catchen, Paul A. Hohenlohe, Susan Bassham, Angel Amores, and William A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

Julian M Catchen, Angel Amores, Paul Hohenlohe, William Cresko, and John H Postlethwait. 2011. *Stacks* : Building and Genotyping Loci *De Novo* From Short-Read Sequences. G3 GenesGenomesGenetics 1:171–182.

Koski, M. H., L. F. Galloway, and J. W. Busch. 2019a. Pollen limitation and autonomous selfing ability interact to shape variation in outcrossing rate across a species range. Am. J. Bot. 106:1240–1247.

Koski, M. H., N. C. Layman, C. J. Prior, J. W. Busch, and L. F. Galloway. 2019b. Selfing ability and drift load evolve with range expansion. Evol. Lett. 3:500–512.

Kulmuni, J., R. K. Butlin, K. Lucek, V. Savolainen, and A. M. Westram. 2020. Towards the completion of speciation: the evolution of reproductive isolation beyond the first barriers. Philos. Trans. R. Soc. B Biol. Sci. 375:20190528.

Liao, W.-J., B.-R. Zhu, Y.-F. Li, X.-M. Li, Y.-F. Zeng, and D.-Y. Zhang. 2019. A comparison of reproductive isolation between two closely related oak species in zones of recent and ancient secondary contact. BMC Evol. Biol. 19:70.

Lindtke, D., and C. A. Buerkle. 2015. The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. Evolution 69:1987–2004.

Malinsky, M., M. Matschiner, and H. Svardal. 2021. Dsuite - Fast *D* -statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21:584–595.

Mandeville, E. G., T. L. Parchman, D. B. McDonald, and C. A. Buerkle. 2015. Highly variable reproductive isolation among pairs of *Catostomus* species. Mol. Ecol. 24:1856–1872.

Mandeville, E. G., T. L. Parchman, K. G. Thompson, R. I. Compton, K. R. Gelwicks, S. J. Song, and C. A. Buerkle. 2017. Inconsistent reproductive isolation revealed by interactions between *Catostomus* fish species. Evol. Lett. 1:255–268.

Mansion, G., G. Parolly, A. A. Crowl, E. Mavrodiev, N. Cellinese, M. Oganesian, K. Fraunhofer, G. Kamari, D. Phitos, R. Haberle, G. Akaydin, N. Ikinci, T. Raus, and T. Borsch. 2012. How to Handle Speciose Clades? Mass Taxon-Sampling as a Strategy towards Illuminating the Natural History of Campanula (Campanuloideae). PLoS ONE 7:e50076.

Momigliano, P., A.-B. Florin, and J. Merilä. 2021. Biases in demographic modelling affect our understanding of recent divergence. Mol. Biol. Evol. msab047.

Momigliano, P., H. Jokinen, A. Fraimout, A.-B. Florin, A. Norkko, and J. Merilä. 2017. Extraordinarily rapid speciation in a marine fish. Proc. Natl. Acad. Sci. 114:6074–6079.

Moyle, L. C., and T. Nakazato. 2010. Hybrid Incompatibility "Snowballs" Between *Solanum* Species. Science 329:1521–1523.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139:1805–1813.

Paris, J. R., J. R. Stevens, and J. M. Catchen. 2017. Lost in parameter space: a road map for STACKS. Methods Ecol. Evol. 8:1360–1373.

Perez, M. F., F. F. Franco, J. R. Bombonato, I. A. S. Bonatelli, G. Khan, M. Romeiro-Brito, A. C. Fegies, P. M. Ribeiro, G. A. R. Silva, and E. M. Moraes. 2018. Assessing population structure in the face of isolation by distance: Are we neglecting the problem? Divers. Distrib. 24:1883–1889.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7:e37135.

Petit, R. J., R. Bialozyt, P. Garnier-Géré, and A. Hampe. 2004. Ecology and genetics of tree invasions: from recent introductions to Quaternary migrations. For. Ecol. Manag. 197:117–137.

Pettengill, J. B., and D. A. Moeller. 2012. Phylogeography of speciation: allopatric divergence and secondary contact between outcrossing and selfing *Clarkia*. Mol. Ecol. 21:4578–4592.

Prior, C. J., N. C. Layman, M. H. Koski, L. F. Galloway, and J. W. Busch. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. Mol. Ecol. 29:4473–4486.

Python Software Foundation. n.d. Python Language Reference.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rogers, J. L. 1965. Documented plant chromosome numbers 65:1. SIDA 65:163–165.

Rougeux, C., L. Bernatchez, and P.-A. Gagnaire. 2017. Modeling the Multiple Facets of Speciation-with-Gene-Flow toward Inferring the Divergence History of Lake Whitefish Species Pairs (Coregonus clupeaformis). Genome Biol. Evol. 9:2057–2074.

Sadedin, S., and M. J. Littlejohn. 2003. A spatially explicit individual-based model of reinforcement in hybrid zones. Evolution 57:962–970.

Santini, F., M. P. Miglietta, and A. Faucci. 2012. Speciation: Where Are We Now? An Introduction to a Special Issue on Speciation. Evol. Biol. 39:141–147.

Sianta, S. A., D. A. Moeller, and Y. Brandvain. 2024. The extent of introgression between incipient *Clarkia* species is determined by temporal environmental variation and mating system. Proc. Natl. Acad. Sci. 121:e2316008121.

Slatkin, M., and L. Excoffier. 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics 191:171–181.

Sloan, D. B., J. C. Havird, and J. Sharbrough. 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol. Ecol. 26:2212–2236.

Sloan, D. B., J. M. Warren, A. M. Williams, Z. Wu, S. E. Abdel-Ghany, A. J. Chicco, and J. C. Havird. 2018. Cytonuclear integration and co-evolution. Nat. Rev. Genet. 19:635–648.

Tiffin, P., S. Olson, and L. C. Moyle. 2001. Asymmetrical crossing barriers in angiosperms. Proc. R. Soc. Lond. B Biol. Sci. 268:861–867.

Turelli, M., and L. C. Moyle. 2007. Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule. Genetics 176:1059–1088.

Van Riemsdijk, I., J. W. Arntzen, G. M. Bucciarelli, E. McCartney-Melstad, M. Rafajlović, P. A. Scott, E. Toffelmier, H. B. Shaffer, and B. Wielstra. 2023. Two transects reveal remarkable variation in gene flow on opposite ends of a European toad hybrid zone. Heredity 131:15–24.

Vences, M., A. Miralles, and C. Dufresnes. 2024. Next-generation species delimitation and taxonomy: Implications for biogeography. J. Biogeogr. jbi.14807.

AUTHOR CONTRIBUTIONS

Keric Lamb performed analyses related to population structure, isolation by distance, demographic inference, and gene flow directionality, and wrote the manuscript. Catherine Debban designed the experiments, performed DNA extractions, assembled genetic resources, performed analyses related to phylogenetics and gene flow quantity, and assisted in the writing of the methods and results sections. Laura Galloway assisted in the design of experiments, consulted on analyses, and edited the manuscript.

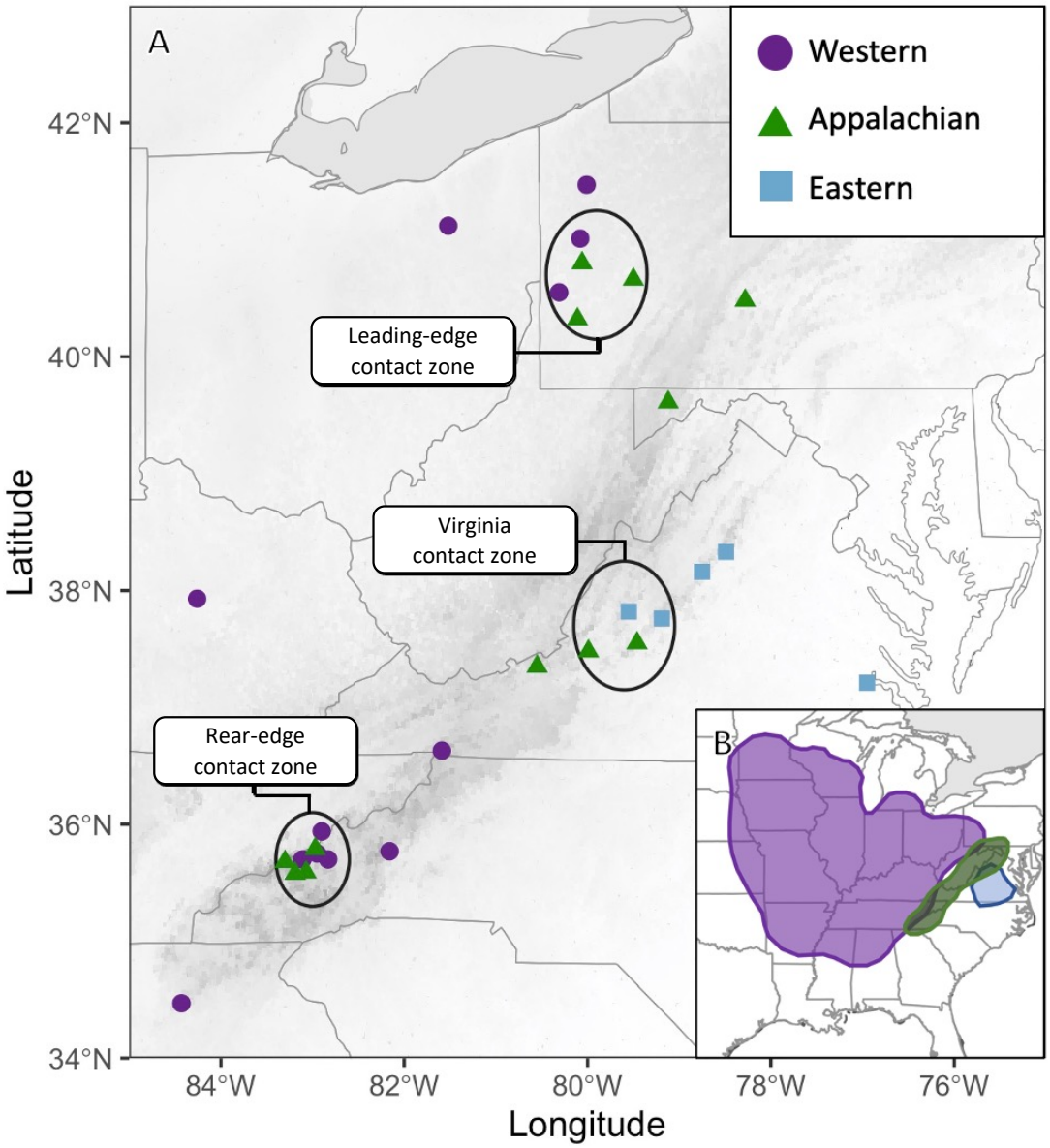


Figure 1) (A) Populations of *Campanula americana* from three contact zones: the leading edge (Pennsylvania), Virginia, and the rear edge (North Carolina). Populations within circles were sampled within 80km of populations of the alternate lineage; populations outside of circles were treated as allopatric. (B) Range map of the Appalachian (green), Eastern (blue), and Western (purple) lineages of *C. americana*.

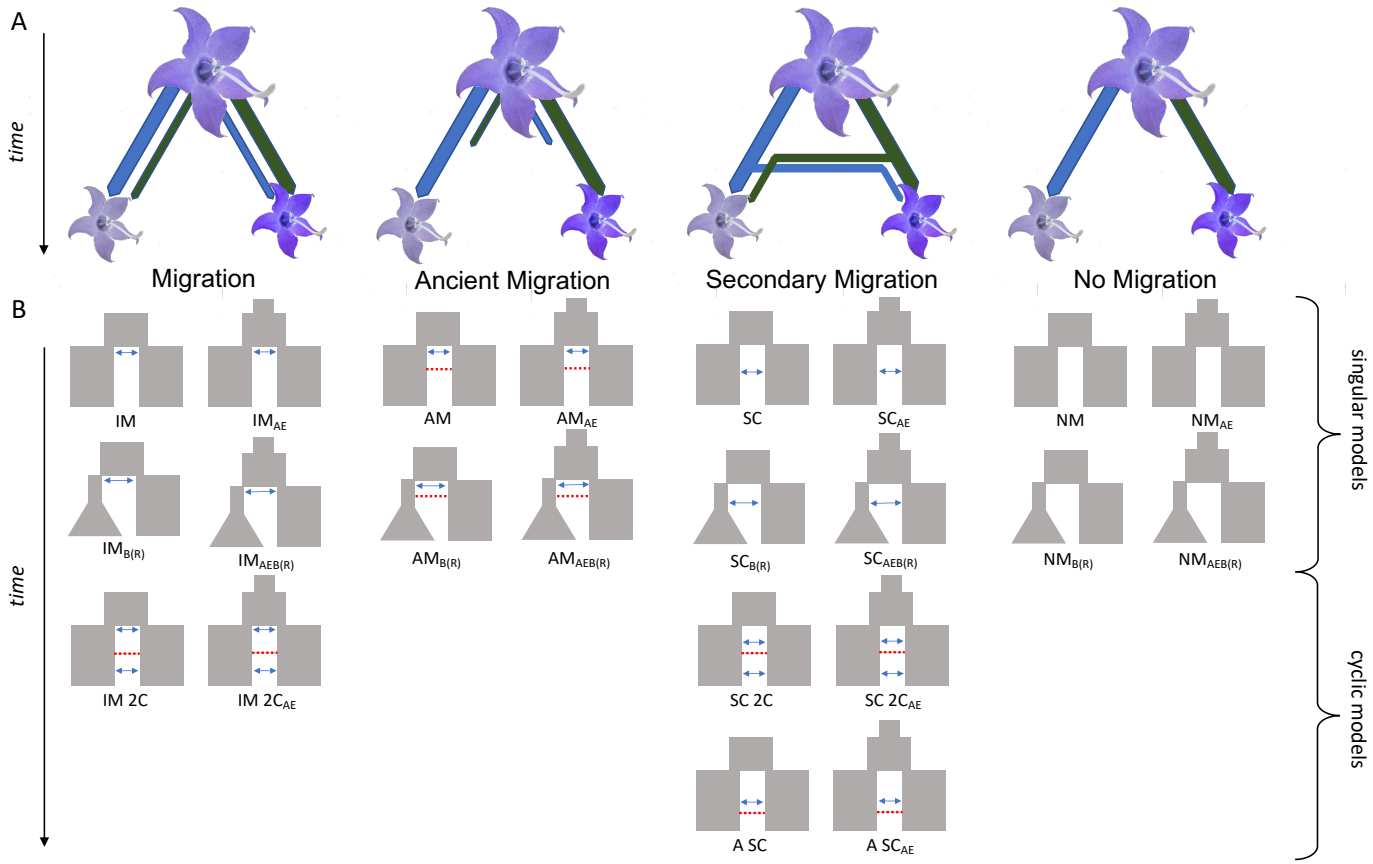


Figure 2) (A) Scenarios describing potential evolutionary histories of lineages in contact. Contact may arise between lineages diverging in parapatry with constant migration (IM); in parapatry with migration ending at some point before the present (AM); in isolation with migration beginning at some point before the present (SC); or in isolation without migration (NM). (B) Divergence between lineages may be accompanied by other demographic shifts including instantaneous expansion of the ancestral population (AE), bottle-growth of either descendent population (B/B(R)), and combinations of the two. There may be cyclic contact where lineages experience multiple episodes of migration including migration after a period of isolation that ceases at some point before the present (A SC); migration after a period of isolation two times (SC 2C); and constant migration with an intermediate break (IM 2C). Cyclic models correspond to evolutionary history with glacial cycling and repeated interglacial contact.

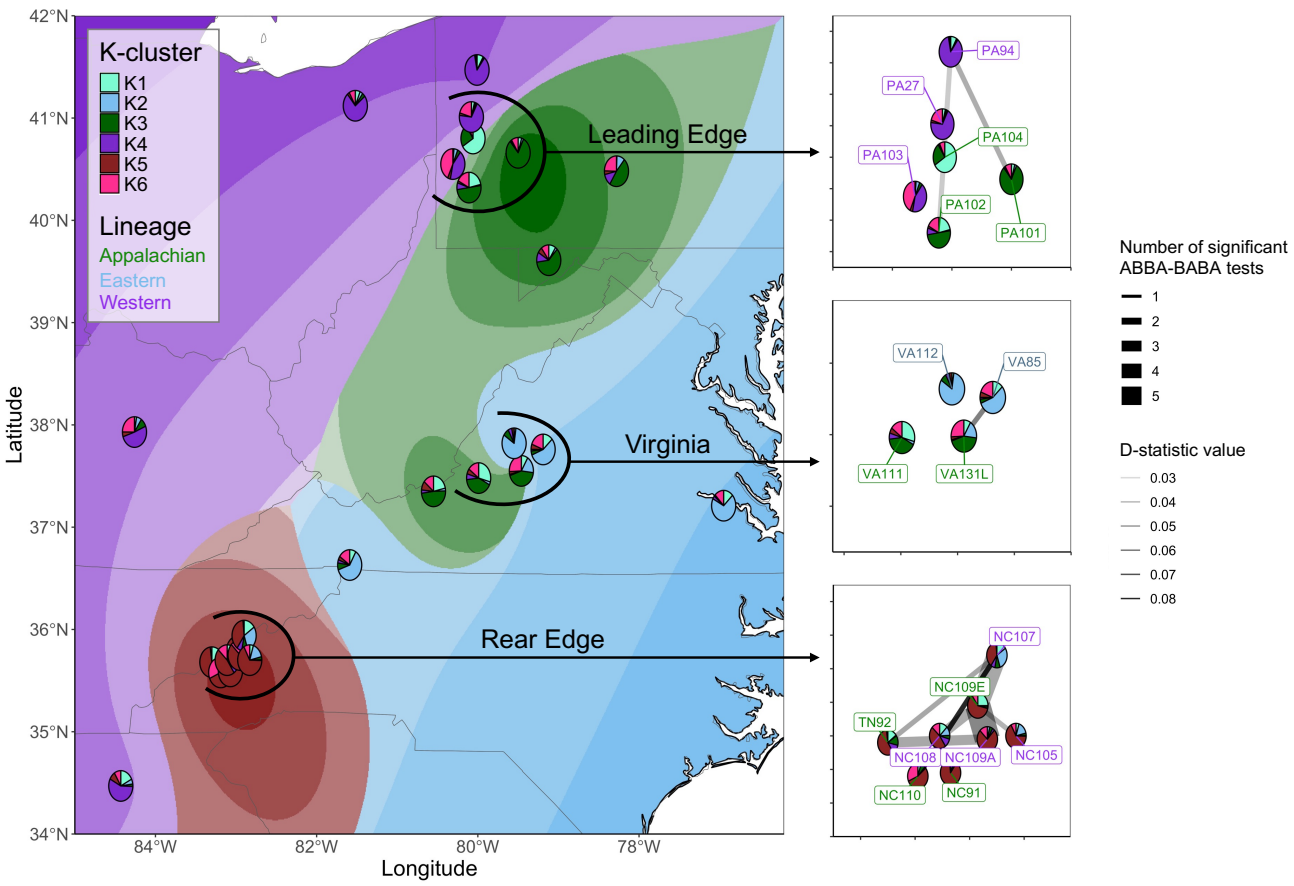


Figure 3) (A) Spatial interpolation of population ancestry and admixture across the species' range. In the leading-edge and Virginia contact zones, lineages are distinct and well-defined, despite contact. In the rear-edge contact zone, populations are not well-differentiated by lineage affiliation. Map colors indicate the primary ancestry assignment in each geographic region, while shading represents the expected proportion of ancestry related to the primary assignment in each region. (B) Regional cutouts of panel A showing estimates of gene flow from ABBA-BABA tests (Table S5). The width of segments representing gene flow indicate the number of significant tests which found gene flow between the pair of populations, indicating the support for the finding of gene flow. Opacity of segments representing gene flow indicates the quantity of gene flow between the populations. In the leading-edge and Virginia contact zones, little gene flow is exchanged among populations of differing lineages. In the rear-edge contact zone, gene flow between lineages is high (i.e., higher values of *D*-statistics) and well-supported (i.e., more significant tests).

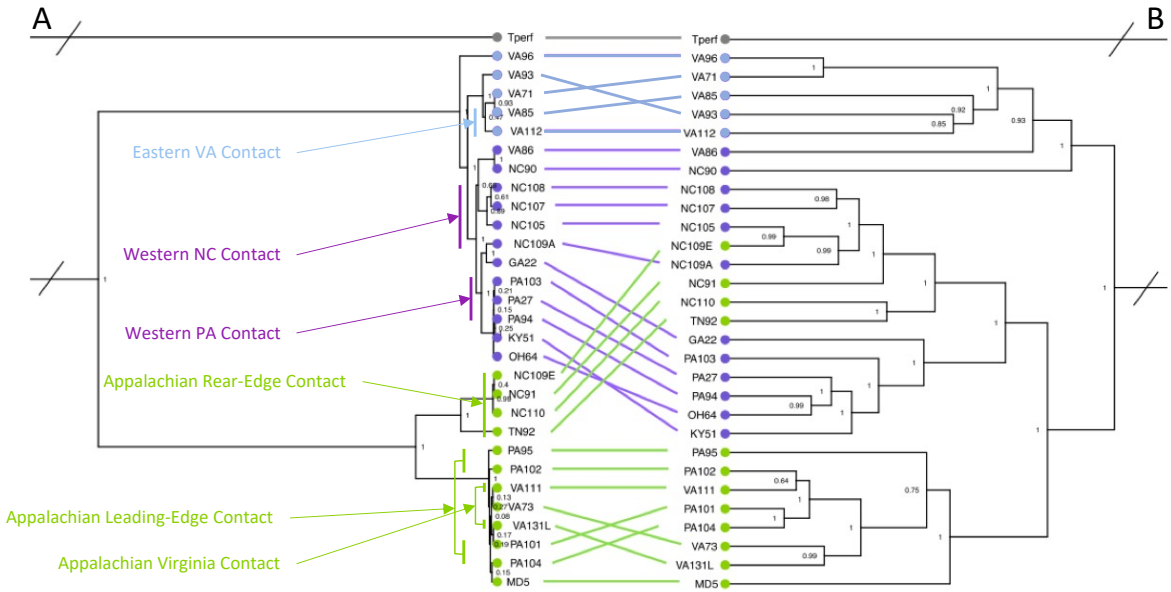


Figure 4) Population phylogenies built with a Bayesian method from (A) the cytoplasmic dataset and (B) the nuclear dataset. Nodes are marked with posterior probabilities. Western lineage populations are marked in purple, Eastern in blue, and Appalachian in green. Phylogeny topology is largely concordant for populations in the leading-edge and Virginia contact zones, suggesting minimal inter-lineage admixture. Phylogenies are discordant for populations in the rear-edge contact zone, with the nuclear phylogeny placing Appalachian-lineage rear-edge populations into a Western-lineage clade. *Adapted from Debban (2019).*

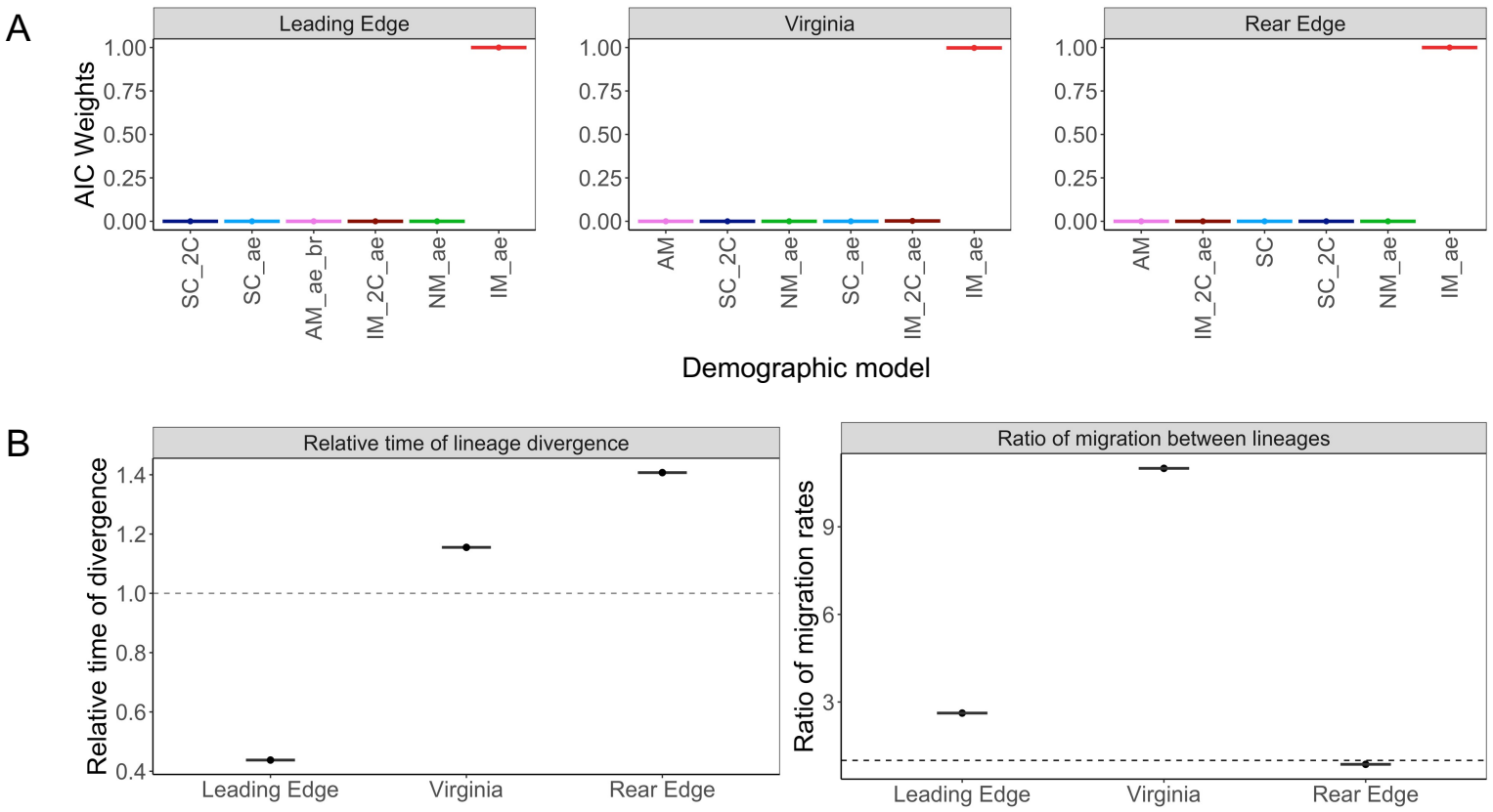
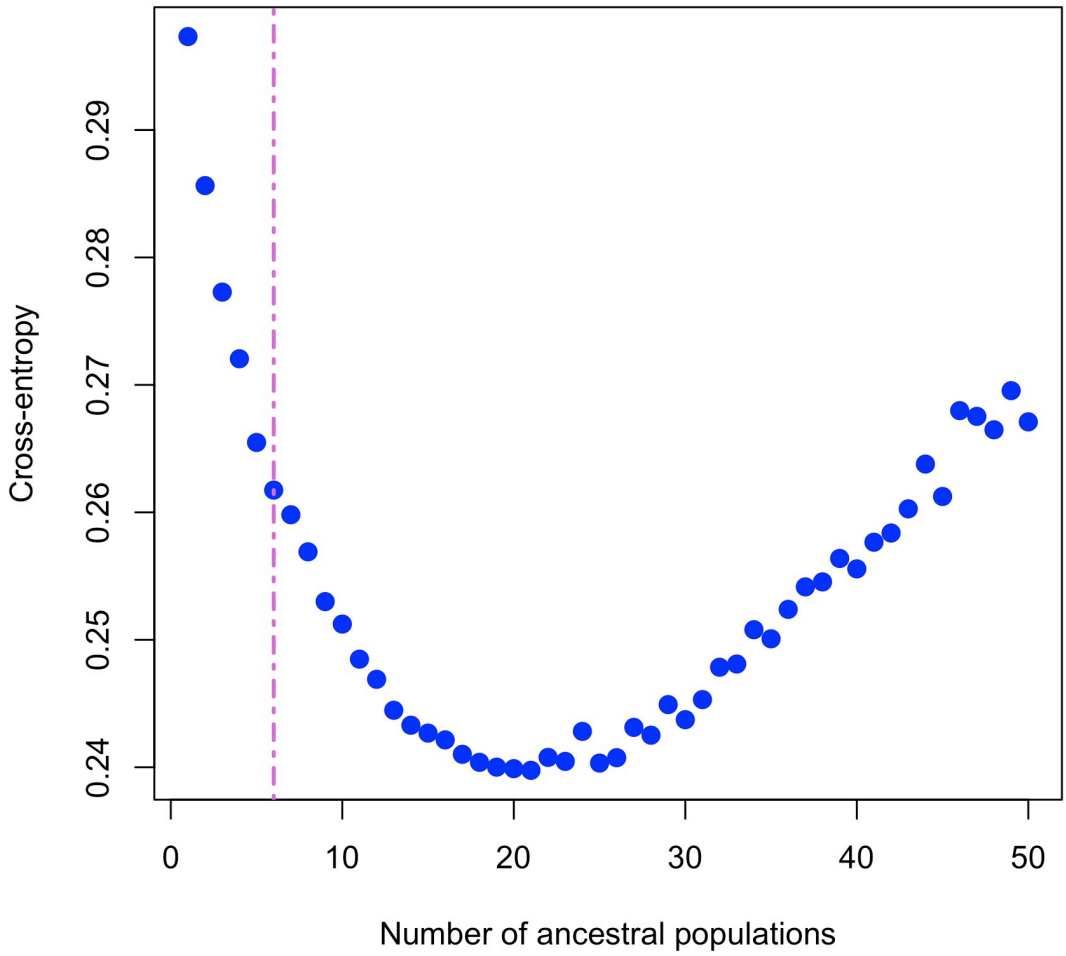
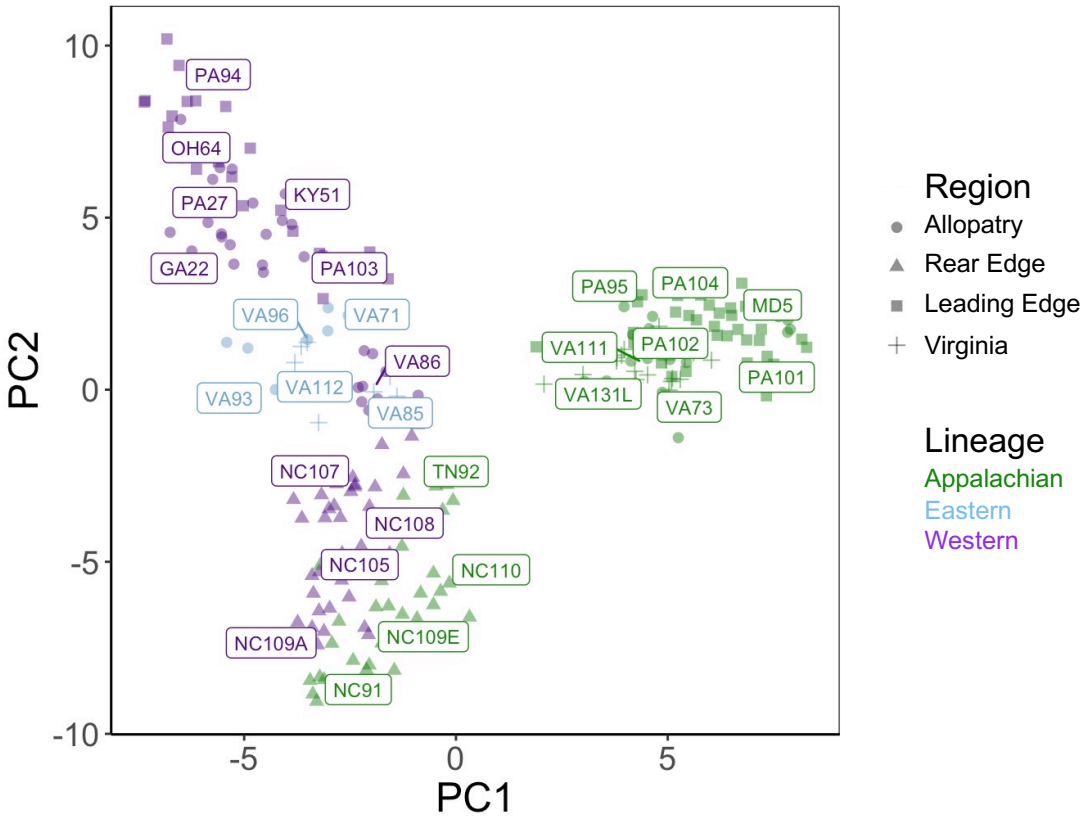


Figure 5) (A) AIC weights of the best model (AE/BR) within each major type of evolutionary history scenario (see Figure S6 for details). Across contact zones, the best model is divergence with constant migration (IM_{AE}). (B) Relative time of lineage divergence, indicating more recent divergence for leading-edge lineages, and the ratio of migration among lineages from IM_{AE} models for each contact zone. Dashed lines indicate 1.0. Asymmetric migration for the leading-edge and Virginia contact zones were concordant with known cytonuclear incompatibility between lineages whereas migration was not asymmetric in the rear-edge contact zone.

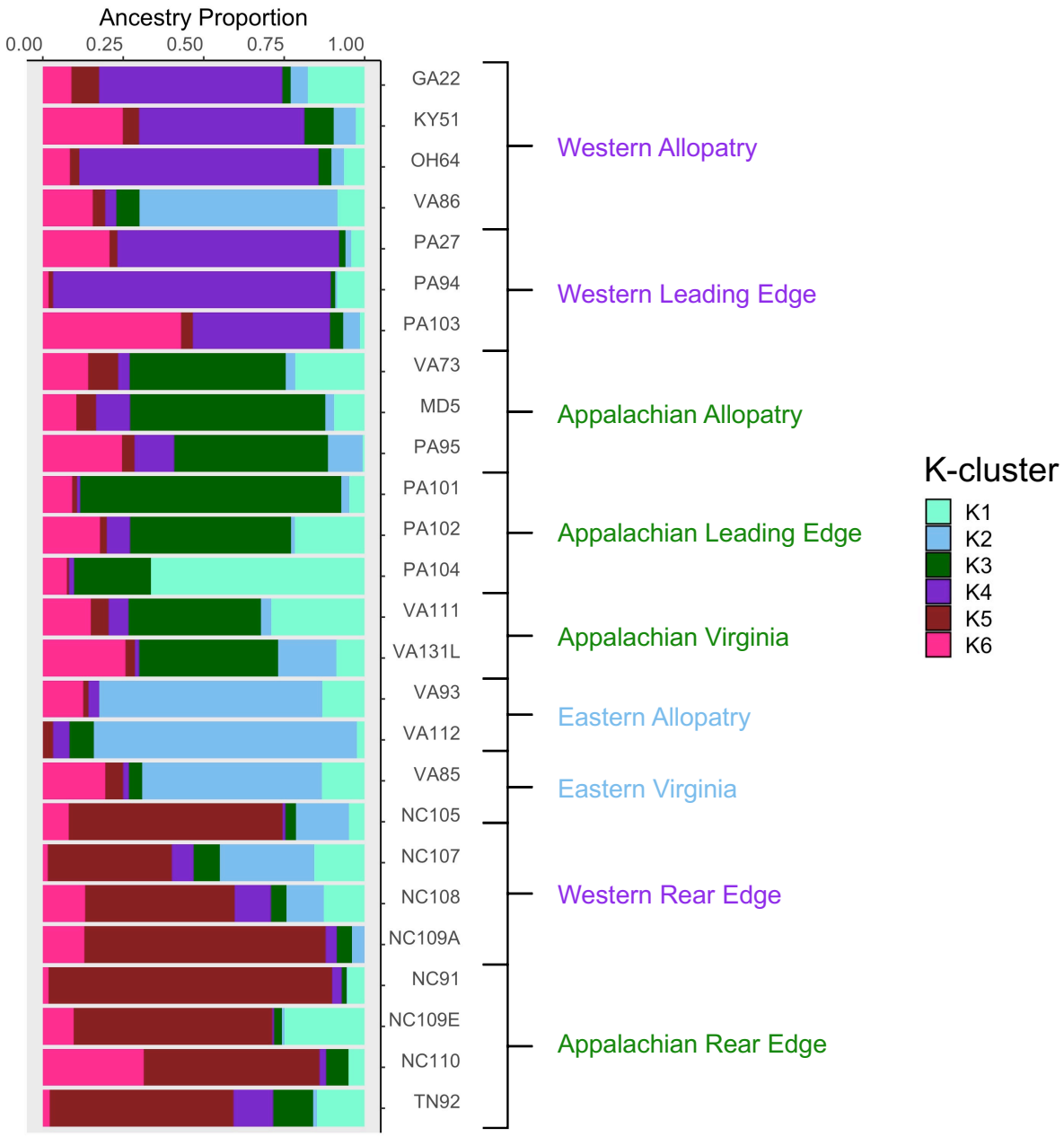
SI Figure 1) Schematic of adapter sequences and PCR indices used in library construction. *Adapted from Debban (2019).*



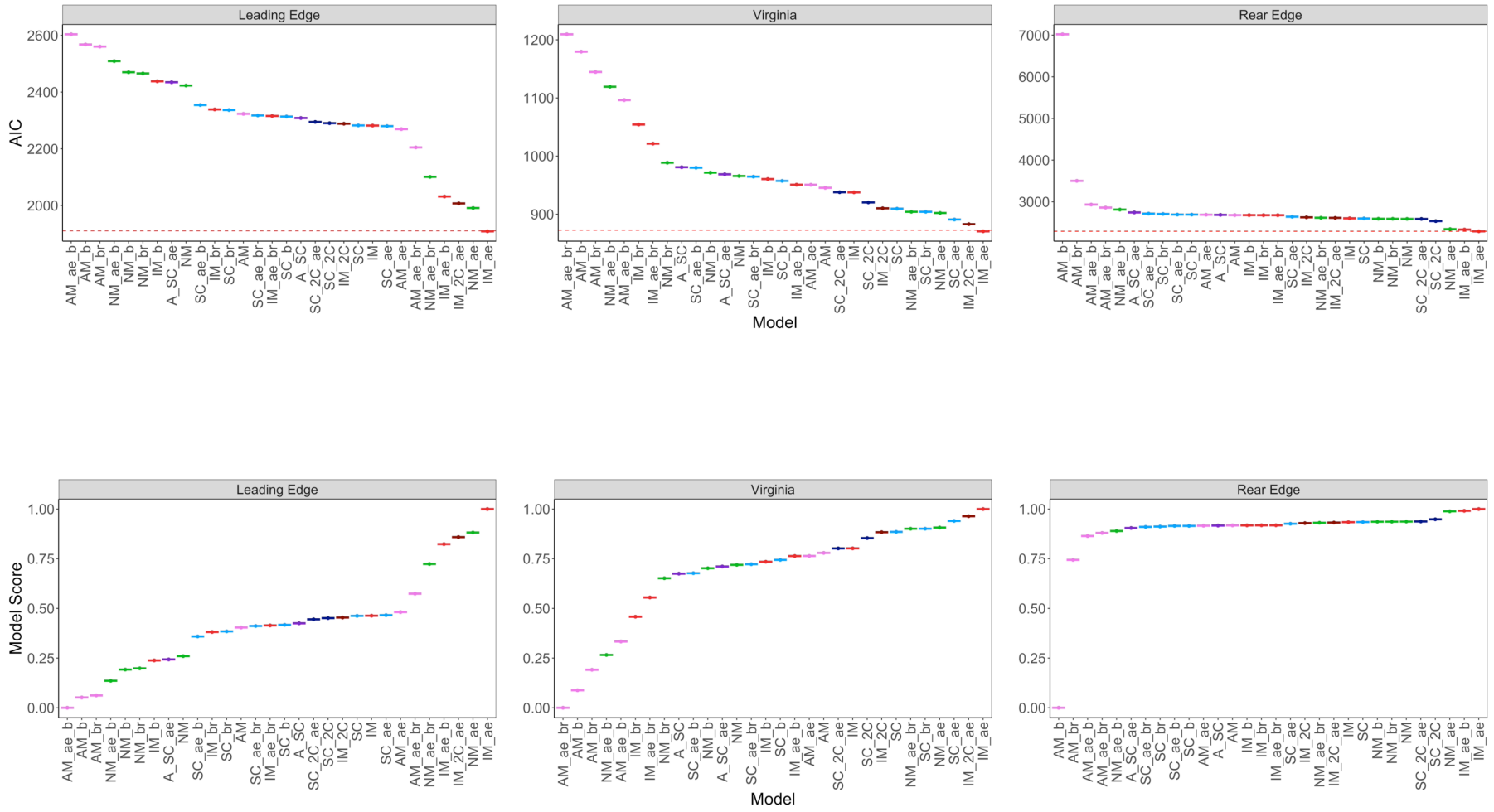
SI Figure 2) Cross-entropy values from ancestry coefficient estimation. The optimal number of clusters (n=6) is shown in pink.



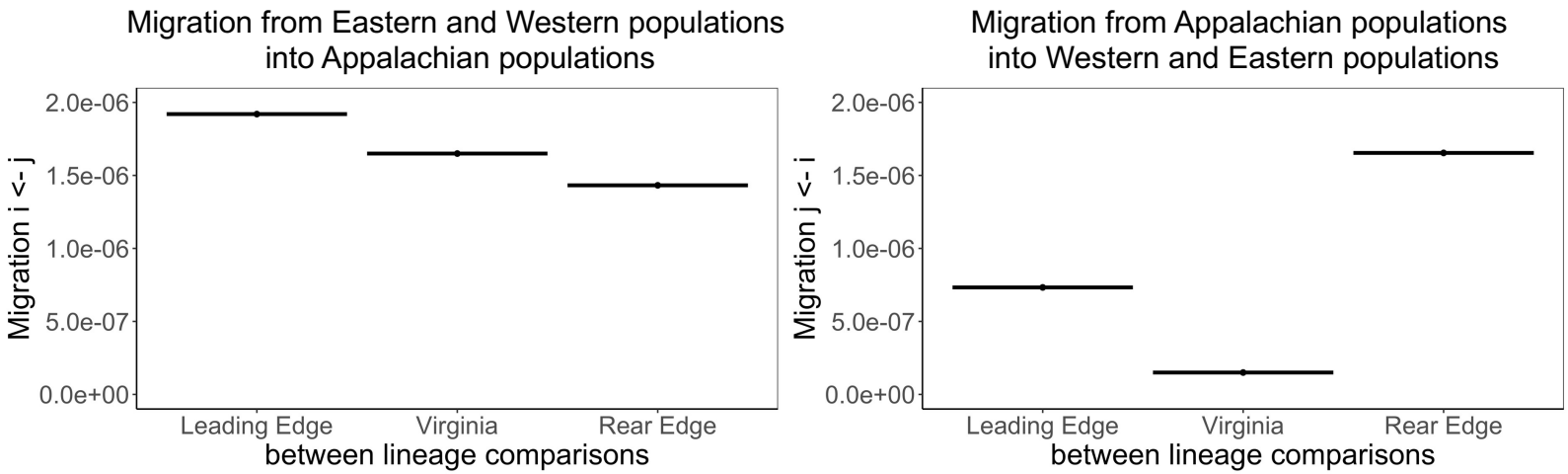
SI Figure 3) Population structure estimated from principal component analysis using the high-quality nuclear SNP dataset. Appalachian-lineage populations from the leading-edge and Virginia contact zones cluster separate from rear-edge Appalachian populations. Western-lineage populations are broken into two primary clusters—a rear-edge cluster near Appalachian-lineage rear-edge populations, and a leading-edge cluster composed of populations hailing from northward postglacial range expansion.



SI Figure 4) Ancestry coefficients estimated from the high-quality nuclear dataset. Estimates are identical to pie charts in Figure 3 but are rendered here in barplot form.



SI Figure 5) Full results of demographic inference including all sub-models of evolutionary history. Models are ranked in order by AIC and Model Score, which is calculated by finding the ΔAIC between the best model for a contact zone and any given model, then dividing it by the ΔAIC between the best and worst model for a contact zone. Across all contact zones and methods, IM_{AE} was selected as the best model. For AIC plots, dashed lines indicate $AIC_{MIN}+2$ to show whether other high-performing models are distinguishable from the best.



SI Figure 6) Migration rates inferred from the optimal demographic model of contact in each contact zone. Migration rates were asymmetric in the direction of known cytonuclear incompatibility (compatible: Western/Eastern --> Appalachian), except in North Carolina, where migration rates were estimated to be approximately symmetric.

Population	Samples	Lineage	Region
MD5	8	Appalachian	Allopatry
PA95	12	Appalachian	Allopatry
VA73	14	Appalachian	Allopatry
GA22	12	Western	Allopatry
KY51	10	Western	Allopatry
OH64	15	Western	Allopatry
VA71	3	Western	Allopatry
VA86	10	Western	Allopatry
VA93	4	Western	Allopatry
VA96	1	Western	Allopatry
NC109E	19	Appalachian	NC Contact
NC110	10	Appalachian	NC Contact
NC91	17	Appalachian	NC Contact
TN92	17	Appalachian	NC Contact
NC105	13	Western	NC Contact
NC107	12	Western	NC Contact
NC108	15	Western	NC Contact
NC109A	6	Western	NC Contact
NC90	1	Western	NC Contact
PA101	17	Appalachian	PA Contact
PA102	17	Appalachian	PA Contact
PA104	12	Appalachian	PA Contact
PA103	8	Western	PA Contact
PA27	15	Western	PA Contact
PA94	10	Western	PA Contact
VA111	13	Appalachian	VA Contact
VA131L	17	Appalachian	VA Contact
VA112	4	Western	VA Contact
VA85	7	Western	VA Contact
Triodanis perfoliata	19	Outgroup	Outgroup

SI Table 1) Population sampling information. Lineage assignments were determined by chloroplast genotyping. *Adapted from Debban (2019).*

NAME	BARCODE	SEQ (5' --> 3')
Index1_EcoRI_P1a	ATCACG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCACG
Index1_EcoRI_P1b		/5Phos/AATTCGTGATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index2_EcoRI_P1a	CGATGT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATGT
Index2_EcoRI_P1b		/5Phos/AATTACATCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index3_EcoRI_P1a	TTAGGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTAGGC
Index3_EcoRI_P1b		/5Phos/AATTGCCTAAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index4_EcoRI_P1a	TGACCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGACCA
Index4_EcoRI_P1b		/5Phos/AATTTGGTCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index5_EcoRI_P1a	ACAGTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACAGTG
Index5_EcoRI_P1b		/5Phos/AATTCACTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index6_EcoRI_P1a	GCCAAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAAT
Index6_EcoRI_P1b		/5Phos/AATTATTGGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index7_EcoRI_P1a	CAGATC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGATC
Index7_EcoRI_P1b		/5Phos/AATTGATCTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index8_EcoRI_P1a	ACTTGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTGA
Index8_EcoRI_P1b		/5Phos/AATTTCAAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index9_EcoRI_P1a	GATCAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCAG
Index9_EcoRI_P1b		/5Phos/AATTCTGATCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index10_EcoRI_P1a	TAGCTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGCTT
Index10_EcoRI_P1b		/5Phos/AATTAAGCTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index11_EcoRI_P1a	GGCTAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCTAC
Index11_EcoRI_P1b		/5Phos/AATTGTAGCCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index12_EcoRI_P1a	CTTGTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGTA
Index12_EcoRI_P1b		/5Phos/AATTTACAAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index13_EcoRI_P1a	AGTCAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCAA
Index13_EcoRI_P1b		/5Phos/AATTTTGACTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index14_EcoRI_P1a	AGTTCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTTCC
Index14_EcoRI_P1b		/5Phos/AATTGGAACTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index15_EcoRI_P1a	ATGTCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGTCA
Index15_EcoRI_P1b		/5Phos/AATTTGACATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index16_EcoRI_P1a	CCGTCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGTCC
Index16_EcoRI_P1b		/5Phos/AATTGGACGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Index17_EcoRI_P1a	GTAGAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGAG
Index17_EcoRI_P1b		/5Phos/AATTCTCTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index18_EcoRI_P1a	GTCCGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCCGC
Index18_EcoRI_P1b		/5Phos/AATTGCGGACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index19_EcoRI_P1a	GTGAAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGAAA
Index19_EcoRI_P1b		/5Phos/AATTTTTCACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index20_EcoRI_P1a	GTGGCC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGGCC
Index20_EcoRI_P1b		/5Phos/AATTGGCCACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index21_EcoRI_P1a	GTTTCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTTCG
Index21_EcoRI_P1b		/5Phos/AATTCGAAACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index22_EcoRI_P1a	CGTACG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACG
Index22_EcoRI_P1b		/5Phos/AATTCGTACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index23_EcoRI_P1a	GAGTGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGTGG
Index23_EcoRI_P1b		/5Phos/AATTCCACTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index24_EcoRI_P1a	GGTAGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTAGC
Index24_EcoRI_P1b		/5Phos/AATTGCTACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index25_EcoRI_P1a	ACTGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTGAT
Index25_EcoRI_P1b		/5Phos/AATTATCAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index26_EcoRI_P1a	ATGAGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGC
Index26_EcoRI_P1b		/5Phos/AATTGCTCATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index27_EcoRI_P1a	ATTCCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTCCT
Index27_EcoRI_P1b		/5Phos/AATTAGGAATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index28_EcoRI_P1a	CAAAAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAAAG
Index28_EcoRI_P1b		/5Phos/AATTCTTTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index29_EcoRI_P1b	CAACTA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACTA
Index29_EcoRI_P1b		/5Phos/AATTTAGTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index30_EcoRI_P1a	CACCGG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACCGG
Index30_EcoRI_P1b		/5Phos/AATTCCGGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index31_EcoRI_P1a	CACGAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACGAT
Index31_EcoRI_P1b		/5Phos/AATTATCGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index32_EcoRI_P1a	CACTCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACTCA
Index32_EcoRI_P1b		/5Phos/AATTTGAGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index33_EcoRI_P1a	CAGGCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGCG

Index33_EcoRI_P1b		/5Phos/AATTCGCCTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index34_EcoRI_P1a	CATGGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGGC
Index34_EcoRI_P1b		/5Phos/AATTGCCATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index35_EcoRI_P1a	CATTTT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATTTT
Index35_EcoRI_P1b		/5Phos/AATTAAAATGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index36_EcoRI_P1a	CCAACA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAACA
Index36_EcoRI_P1b		/5Phos/AATTTGTTGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index37_EcoRI_P1a	CGGAAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGAAT
Index37_EcoRI_P1b		/5Phos/AATTATTCCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index38_EcoRI_P1a	CTAGCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTAGCT
Index38_EcoRI_P1b		/5Phos/AATTAGCTAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index39_EcoRI_P1a	CTATAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTATAC
Index39_EcoRI_P1b		/5Phos/AATTGTATAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index40_EcoRI_P1a	CTCAGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCAGA
Index40_EcoRI_P1b		/5Phos/AATTTCTGAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index41_EcoRI_P1a	GACGAC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACGAC
Index41_EcoRI_P1b		/5Phos/AATTGTCGTCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index42_EcoRI_P1a	TAATCG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAATCG
Index42_EcoRI_P1b		/5Phos/AATTCGATTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index43_EcoRI_P1a	TACAGC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAGC
Index43_EcoRI_P1b		/5Phos/AATTGCTGTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index44_EcoRI_P1a	TATAAT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTATAAT
Index44_EcoRI_P1b		/5Phos/AATTATTATAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index45_EcoRI_P1a	TCATTC	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCATTC
Index45_EcoRI_P1b		/5Phos/AATTGAATGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index46_EcoRI_P1a	TCCCGA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCCGA
Index46_EcoRI_P1b		/5Phos/AATTTCGGGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index47_EcoRI_P1a	TCGAAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGAAG
Index47_EcoRI_P1b		/5Phos/AATTCTTCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index48_EcoRI_P1a	TCGGCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGGCA
Index48_EcoRI_P1b		/5Phos/AATTTGCCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
SphI_P2a		/5Phos/AGATCGGAAGAGCGAGAACAA
SphI_P2b		GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATG

SI Table 2) Sequences of P1 and P2 adapter oligonucleotides and PCR indices used for ddRAD library preparation. *Adapted from Debban (2019).*

Lineage	Region	2N down-sample	maximum 2N value	SNPs retained
Western	Allopatry	48	70	1741
Western	rear edge	64	92	1812
Western	leading edge	24	44	1239
Appalachian	rear edge	48	70	1650
Appalachian	leading edge	68	90	1863
Appalachian	Virginia	28	42	1254
Appalachian	Allopatry	30	46	1421
Eastern	Virginia	12	16	782
Eastern	Allopatry	12	18	721

SI Table 3) 2N down-sampling for demographic inference. We pooled populations by region and lineage to increase dimensionality and resolution of joint site frequency spectra. For each lineage and region, the maximum possible 2N value is given, as well as the 2N value which provided the maximum number of SNPs possible to retain.

	nu1	nu2	nu_ae	S	Tae	Ts	Tsc	Tsc2	m12	m21	m12_1	m21_2
upper bound	20	20	20	0.999	10	10	10	10	20	20	20	20
lower bound	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

SI Table 4) Initial parameter bounds for demographic inference. Not all parameters were included in all models. Nu1 and nu2 parameterize the population size of contemporary population pools, while nu_ae parameterizes the size of the common ancestral population from which lineages are derived. Tae parameterizes the timing of nu_ae expansion; Ts, the time of divergence between lineages; Tsc, the time of secondary contact beginning or ancient migration ceasing; Tsc2, for cyclic models, the time of the most recent period of contact initiating. M12 and m21 parameterize asymmetric migration between populations; m12_2 and m21_2 the asymmetric migration of additional cycles of contact.

P2	P3	significant tests	mean D-statistic	mean Z-score	mean F4	adjusted p-value
NC105	NC109E	1	0.035	3.156	0.291	0.00969
NC107	NC109E	4	0.041	3.052	0.307	0.01515
NC107	NC110	1	0.088	4.890	0.272	0.00005
NC107	TN92	1	0.040	5.282	0.203	0.00001
NC108	NC109E	2	0.039	3.338	0.357	0.00828
NC109A	NC109E	5	0.054	3.492	0.514	0.00692
NC109A	TN92	2	0.040	3.235	0.185	0.00918
PA101	PA94	1	0.038	3.895	0.144	0.00215
PA102	PA94	1	0.025	2.992	0.095	0.01481
VA131L	VA85	1	0.057	3.850	0.165	0.00215

SI Table 5) significant pairwise ABBA-BABA tests after multiple comparison correction. All population comparisons are between lineages. In total, seven pairwise comparisons were significant for rear-edge populations, while only two were significant for leading-edge populations and one for Virginia populations.

APPENDIX 1:

Species Distribution Model

OVERVIEW

Habitat can vary in quality over a species' geographic range, affecting connectivity among populations and species' ecological niche limits (Hargreaves et al. 2014). Incorporating metrics of habitat suitability into model predictions can help to better estimate the environmental contexts that populations are found in (Aguirre-Liguori et al. 2021), and reduce bias in models seeking to correlate genetic distance with geographic distance (i.e., isolation-by-distance). For example, populations found in lowland areas may be substantively less connected than expected by geographic distance if mountains intercede portions of the range. Thus, connectivity and differentiation measured by raw geographic distance would underestimate actual differentiation among populations associated with neutral processes, like genetic drift. To account for such variation, I first generated a species distribution model (SDM) for *Campanula americana* using random forest modeling, data collected from iNaturalist research-grade observations (n=7781; iNaturalist Community, n.d.) and randomly generated pseudo-absence data. I utilize the SDM across chapters to provide environmental context to the relationship between genetic load and population fitness (Chapter 3); to produce less biased estimates of geographic connectivity for models of isolation-by-distance and isolation-by-environment (Chapter 4); and to provide context and insight into historical range dynamics (Chapter 5). Here, I outline methods for generation of the SDM, and in particular methods comparison (generalized linear models/random forest/maximum entropy). I find that random forest models provide higher resolution distribution maps when sample sizes allow. Finally, I briefly discuss conclusions from each model projection.

METHODS

Generating the SDM requires knowing where in a geographic area a species is present and, by extrapolation, where it is absent. For *C. americana,* presence data were derived from observations in iNaturalist. I downloaded all presence records available from iNaturalist as of May 2022 (iNaturalist Community, n.d.). I determined *C. americana*'s absence by assigning areas where there were no observations as "pseudo-absence". To spatially restrict the generation of pseudo-absence data, and prevent its generation far outside the native range of *C. americana*, I first generated a geographic hull for the USA using the rasterize function (Hijmans 2023). I generated pseudo-absence data within this region using the function pseudo.absence in the spatialEco package (Evans 2021). In total, I generated 23,343 (3n presence) random pseudo-absence points, then merged with presence data. The combined dataset was down-sampled to one observation per 10km grid cell by first converting latitudes and longitudes to kilometers, then rounding to the nearest 10km. I then subset presence and pseudo-absence data to only the first per matching latitude and longitude kilometer pair. Where presence and pseudo-absence data overlapped within 10kmx10km grid cells, I retained the presence data point, and discarded the pseudo-absence point. I also removed points west of the hundredth meridian or north of the hundredth parallel, to constrain the dataset to the native range. This final set contained 2,417 presence points and 1,860 pseudo-absence points (SI Fig. 1A).

SDM require climate data to associate with probable occurrence of species. I downloaded bioclimatic variables associated with each location in my final data from WorldClim 2 at a spatial resolution of 2.5 arcseconds (Fick and Hijmans 2017). To account for autocorrelation among bioclimatic variables (SI Fig. 1B), I generated a PCA using all 19 available bioclimatic factors.

Five principal components explained 95% of the variance in bioclimatic factors among data points (SI Fig. 1C). I used these five components for generating the SDM.

I evaluated three different modeling techniques for generating SDMs. These included generalized linear model (GLM), random forest, and maximum entropy modeling approaches. Each approach has different strengths that will affect resolution of the final SDM. Specifically, GLMs account for the interaction among variables but cannot handle non-linear relationships as effectively as explicitly nonlinear models. Random forest models are inherently nonlinear, and may better describe relationships at finer resolution scales without increasing false positive rates compared to generalized regression approaches (Kirasich et al. 2018). Finally, maximum entropy models effectively describe potential climatic niche limits, particularly in habitat where presence points do not exist (Fitzgibbon et al. 2022).

I tested the performance of different modeling techniques by first training models on a subset of 70% of the data (i.e. training data), then comparing their type I and type II error rates using a non-overlapping testing data set (30%). For the GLM, I included all climate principal components as main effects as well as their interactions (*presence ~ PC1*PC2*PC3*PC4*PC5*). I calculated GLMs using the glm function in base R (R Core Team, 2022) and generated spatial predictions using the predict function from base R. I ran random forest models using 5,000 trees and a node size of 10. For random forest models, I predicted spatial data using the predictSDM function from the mecofun package (Zurell 2020). I ran maximum entropy models out-of-the-box, without modifications to parameterization using the maxent function from the dismo package (Hijmans et al. 2021). I generated spatial predictions maximum entropy models using the predict function from the dismo package.

I tested the accuracy of each model type by predicting the testing dataset (30%) and comparing the type I and II errors with confusion matrices. Confusion matrices show the frequency that models correctly and incorrectly predict presence and absence. For random forest models, I selected the random seed that had the highest accuracy (i.e., highest rate of correctly predicting presence and absence) to compute the confusion matrix. I compared models on the basis of type I and II error, indicated by false presence and false absence predictions respectively, and were able to accurately predict known changes in prevalence across the species range (e.g., higher density of *C. americana* along the southern blackbelt prairies). After determining random forest modeling provided the best fit to the data, I then predicted historic and projected future distributions of *C. americana* using the WorldClim projected datasets (http://www.worldclim.com/past). I first generated 100 random forest models using contemporary WorldClim 2.0 data, then predicted presence and absence during last glacial maximum (22kya), mid-Holocene (6kya), and for RCP 4.5 (2070), the most likely climate scenario given current warming trends. For each time point, I then averaged predictions for each grid cell across all 100 random forest models.

RESULTS

Comparison of SDM methods

SDMs generally performed well across methods. Random forest and maximum entropy models weighted the importance of principal component variables approximately in order of components that explained the most bioclimatic variance (SI Fig. 2), suggesting that the models fit the principal component data effectively. The random forest model outperformed the GLM and maximum entropy model for type I error (false presence), though performance of the model was somewhat worse in type II error (false absence) relative to the maximum entropy model (SI Fig. 3). Briefly,

the GLM SDM had false presence rates of 53.93% and false absence rates of 16.9%. The maximum entropy SDM had false presence rates of 37.2% and false absence rates of 7.6%. Finally, the random forest SDM had false presence rates of 28.6% and false absence rates of 9.7%. The difference in type I error between random forest and maximum entropy models results in stronger definition of range boundaries and known zones of low and high prevalence in the maximum forest models. For example, the southern blackbelt prairies, where *C. americana* is frequently found, are well-defined in map projections of the random forest model but poorly defined for projections from the maximum entropy model. Thus, while the maximum entropy model appears to define the potential climate niche more broadly and with less type II error, the random forest defined the contemporary realized extent of the species range more precisely (Fig. 1C).

Historical and future climate projections

Final projections of random forest models using historical and future climate projections reveal changes in distribution over time (Barnard-Kubow et al. 2015, Koski et al. 2019, Prior et al. 2020). The projection to the last glacial maximum found the majority of suitable habitat laid in the southern U.S., near the Gulf of Mexico (Fig. 1A) but allowed for mildly suitable habitat to exist in several mid-latitude regions hypothesized to be glacial refugia, particularly Kentucky and in and around the Appalachian Mountains and eastern coastal areas. In the mid-Holocene, mid-latitude climates in the eastern U.S. were drier and warmer than the current climate (Shin et al. 2006), and projections of the SDM to ~6kya demonstrate expansion of the range relative to the last glacial maximum in response (Fig. 1B). Interestingly, high suitability of habitat beyond the Mississippi and Ohio rivers into southern Illinois aligns with genomic predictions of range expansion beyond the river basins (Prior et al. 2020). During the mid-Holocene (10-6kya), water flow into the

Mississippi river basin was curtailed by drying climates (Knox and Wright 1983), increasing the likelihood of successful dispersal across the river. Finally, the RCP 4.5 (2070) projection demonstrates that range expansions during contemporary climate change do not simply result in northward range shifts or range expansions, but also idiosyncratic range constrictions (Fig. 1D). Importantly, the RCP 4.5 projections suggest that the native range of *C. americana* is likely to experience some contraction toward mid-latitudes in the near future as climate becomes less suitable. Thus, in the absence of adaptation and local response to selection, populations may fail to adjust to shifting climates *in situ*. If regions of the range are subject to constraints of adaptation through microevolutionary or environmental pressures, range constriction, rather than range expansion, is likely.

LITERATURE CITED

Barnard-Kubow, K. B., Debban, C. L., & Galloway, L. F. (2015). Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. *American Journal of Botany*, *102*(11), 1842-1853.

Evans JS (2021). spatialEco: R package version 1.3-6, https://github.com/jeffreyevans/spatialEco

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

Hijmans RJ, Phillips S, Leathwick J, Elith J (2021). dismo: Species Distribution Modeling_. R package version 1.3-5, https://CRAN.R-project.org/package=dismo

Knox, J. C., & Wright, H. E. (1983). Responses of river systems to Holocene climates. *Late quaternary environments of the United States*, *2*, 26-41.

Koski, M. H., Layman, N. C., Prior, C. J., Busch, J. W., & Galloway, L. F. (2019). Selfing ability and drift load evolve with range expansion. *Evolution Letters*, *3*(5), 500-512.

Prior, C. J., Layman, N. C., Koski, M. H., Galloway, L. F., & Busch, J. W. (2020). Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. *Molecular Ecology*, *29*(22), 4473-4486.

Shin, S. I., Sardeshmukh, P. D., Webb, R. S., Oglesby, R. J., & Barsugli, J. J. (2006). Understanding the mid-Holocene climate. *Journal of Climate*, *19*(12), 2801-2817.

Zurell, D. (2020). mecofun: useful functions for macroecology and species distribution modelling version 0.0.0.9. University of Potsdam, Potsdam. https://gitup.uni-potsdam.de/macroecology/mecofun

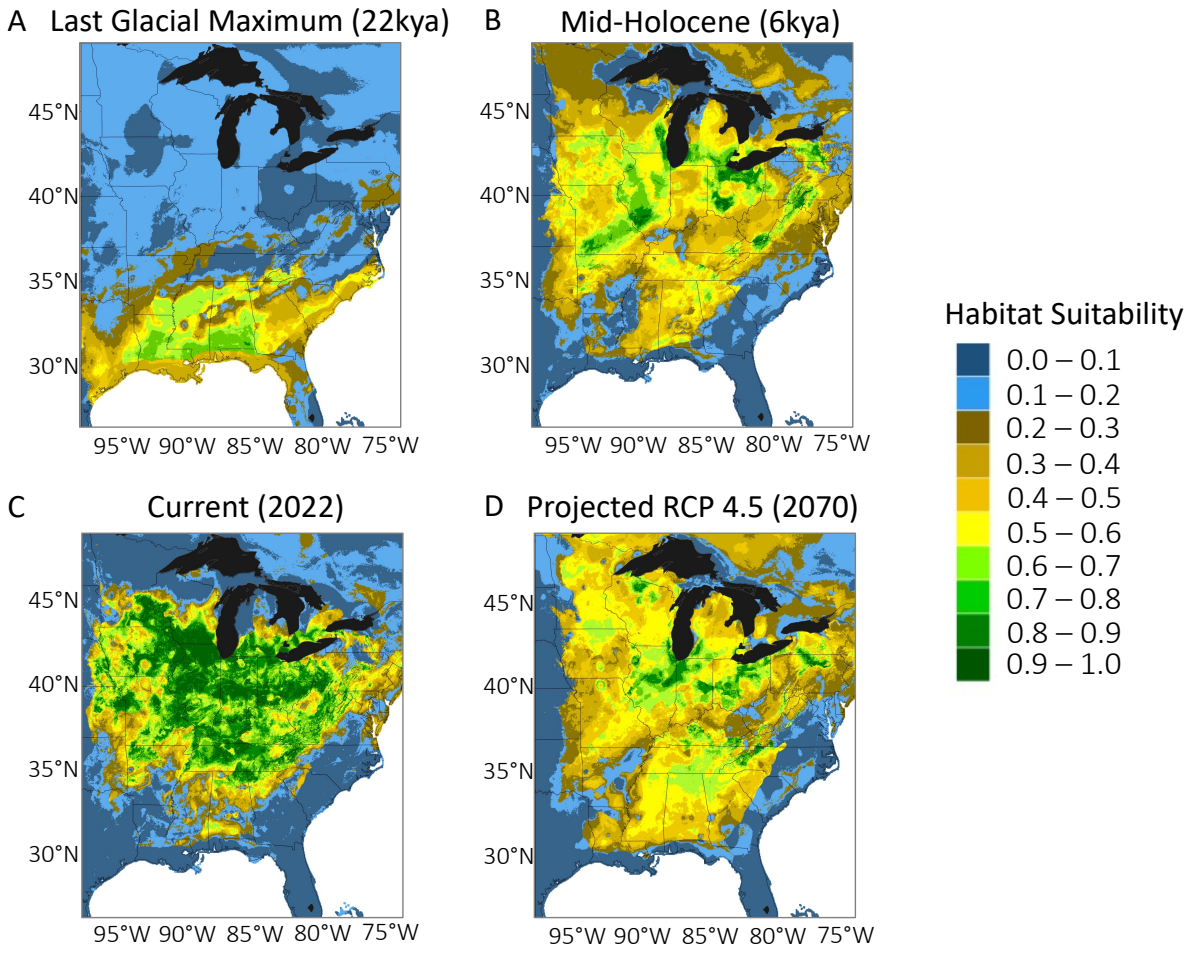
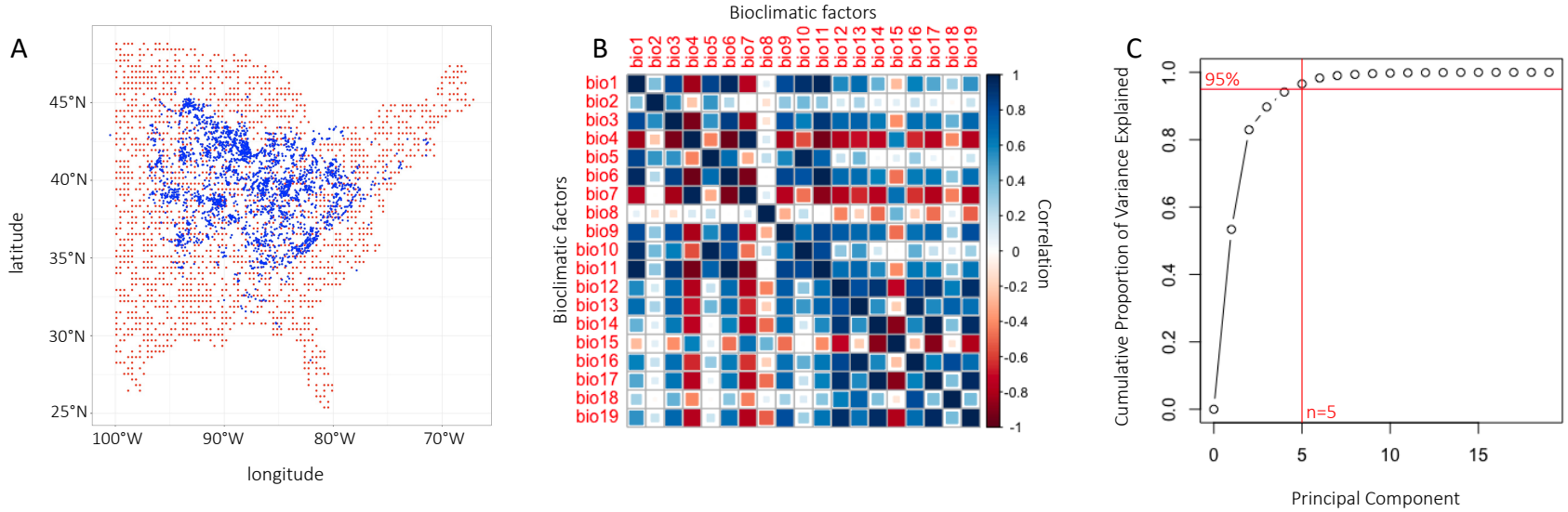
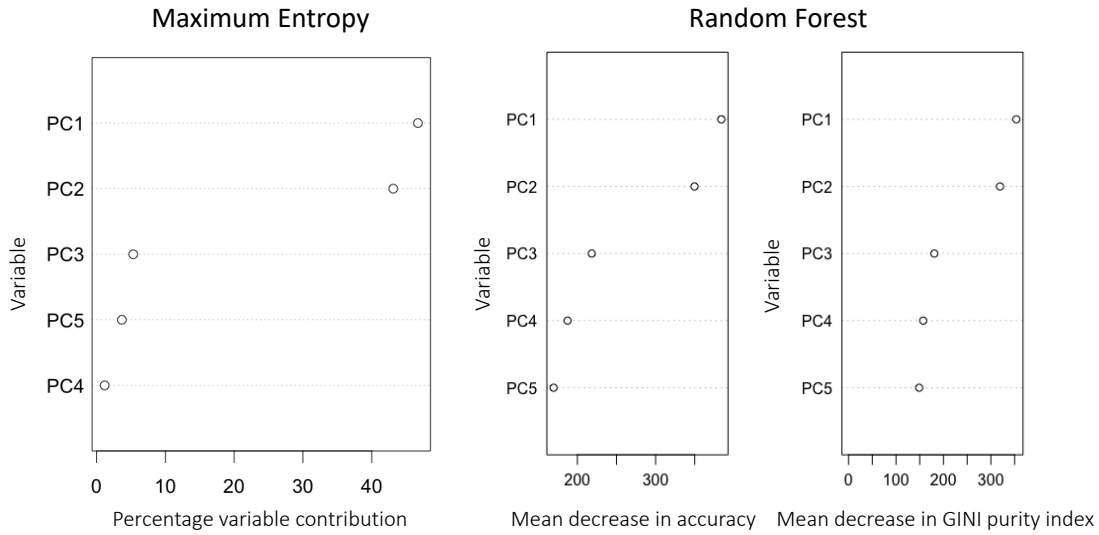


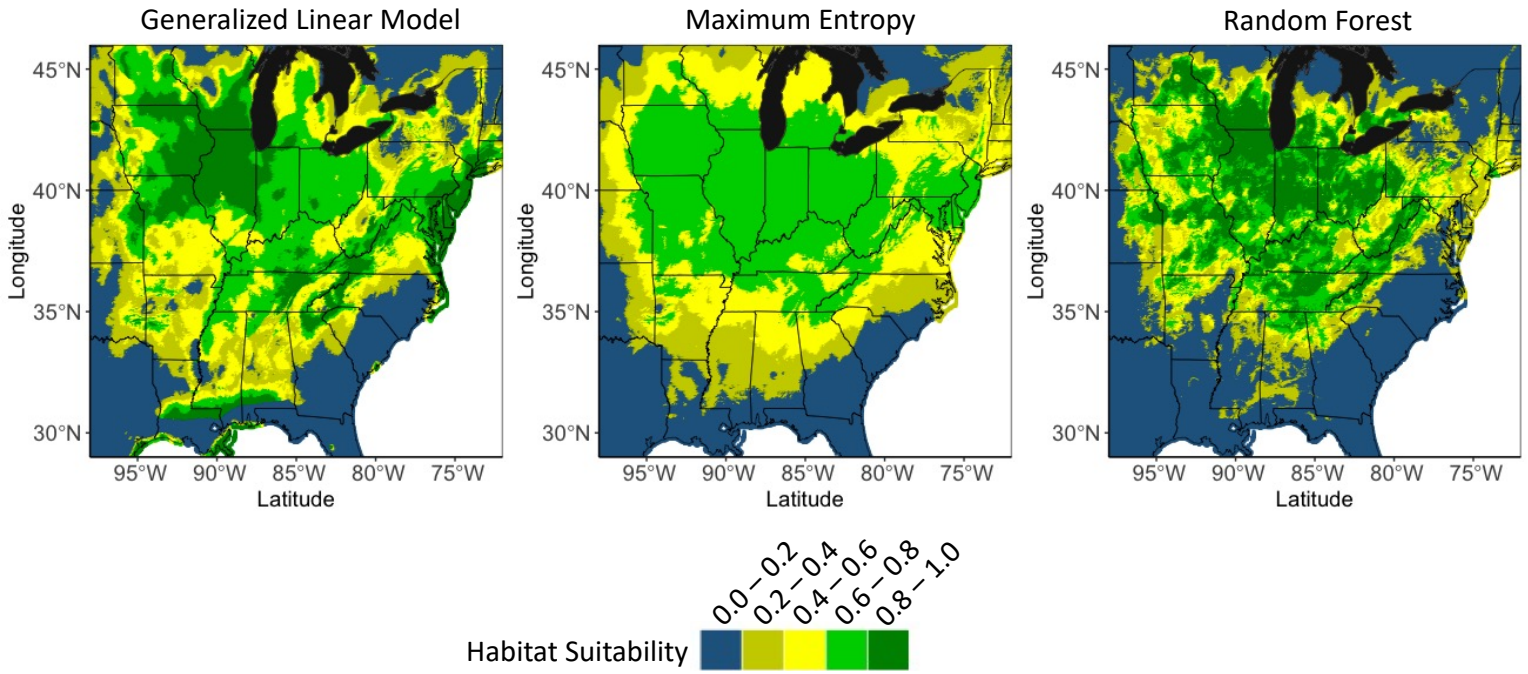
Figure 1) Distribution projections to different time points using the Random Forest model. Distributions reflect those projected for (A) the last glacial maximum ~22kya, (B) the mid-Holocene ~6kya, (C) current climate, and (D) the RCP 4.5 climate scenario (2070). Projections for each time point are binned by decile.



SI Figure 1) (A) Down-sampled presence data from iNaturalist research-grade observations (blue) and pseudo-absence data generated for SDM training (red). Points were constrained to Eastern USA. (B) Correlation of WorldClim bioclimatic factors. (C) To reduce correlation of variables, we performed a PCA and retained five principal components for modeling efforts, as they explained 95% of variance in bioclimatic factors.



SI Figure 2) Climate variable importance for maximum entropy and random forest models. Well-fit models will likely rank principal components variables by the amount of climate variation they explain. For maximum entropy models, the percentage shows the contribution and permutation importance of the variable in the model. For random forest models, mean decrease in accuracy shows the decrease in the model's performance if the variable is excluded. Mean decrease Gini shows the degree of disorganization across trees in the random forest introduced by excluding the factor.



SI Figure 3) Mapped projections from GLM, Maximum Entropy, and Random Forest models, with predictions of habitat suitability binned by quantile.