Title

Pattern and process: An examination of how evolutionary forces shape patterns of genetic diversity and adaptive potential in the long-lived tree species, giant sequoia

Author

DeSilva, Rainbow

Publication Date

2020

Peer reviewed|Thesis/dissertation

Pattern and process: An examination of how evolutionary forces shape patterns of genetic diversity and adaptive potential in the long-lived tree species, giant sequoia

By

Rainbow DeSilva

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge

Professor Richard S. Dodd, Chair
Professor David D. Ackerly
Professor Erica B. Rosenblum

Fall 2020

Abstract

Pattern and process: An examination of how evolutionary forces shape patterns of genetic diversity and adaptive potential in the long-lived tree species, giant sequoia

By

Rainbow DeSilva

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Richard S. Dodd, Chair

During this century, climate warming and altered precipitation patterns will lead to habitat changes that may be detrimental to long-lived tree species. Giant sequoia, *Sequoiadendron giganteum*, is an iconic Sierra Nevada tree species with populations that tend to be small and highly fragmented, making them especially vulnerable to rapid environmental change. For tree species like giant sequoia, long generation times can limit migration outside of current range boundaries to track climate change. Thus, attention needs to be paid to the risks of adaptative mismatches between a population and its environment. In the face of climate change, genetic diversity is the ultimate source of variation upon which selection can act to allow adaptive responses that mitigate this phenomenon. This dissertation is an investigation of the patterns of extant genetic diversity in giant sequoia at multiple spatial scales, and the processes that shape and move diversity, with a focus on the adaptive potential of populations. Specifically, the chapters of this dissertation will address the following topics in detail: 1) Patterns of range-wide population connectivity and estimation of recent and historic gene-flow; 2) Mating parameters and the dispersal dynamics of pollen and seed within giant sequoia populations; 3) Pattern and climatic drivers of local adaptation and identification of potential genomic regions of adaptive significance. The knowledge gleaned from this research will provide comprehensive background information for giant sequoia regarding the extant genetic diversity, the potential for diversity to spread through gene flow, and the risks of future loss of genetic diversity for a locally restricted species under a changing climate. By highlighting both populations of conservation need, (i.e. isolated or genetically depauperate populations) and those that may contain diversity that can serve as pre-adapted variation for future conditions, this research will be an invaluable asset to forest managers seeking to maintain viable giant sequoia populations into the future.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am truly grateful to all the people who helped make this dissertation possible. I would like to start by recognizing my superb dissertation committee. First and foremost, I want to thank my advisor, Richard S. Dodd, for his patience and guidance throughout this entire experience. Richard gave me the space to figure out my true passion and determine the bounds of my dissertation research, yet he was always there to provide ideas and constructive feedback when I needed guidance. I appreciate the support and patience he demonstrated throughout this extended process, which was pleasantly interrupted by the birth of my two amazing children. Richard inspired and fostered my love for teaching. As a GSI for the classes he led, I always felt that I was part of a teaching team and that my ideas were appreciated. Richard, thank you for sharing your immense knowledge of trees with me. It has been an honor learning from you. Second, a heartfelt thank you to Erica (Bree) Rosenblum for being my champion during this experience. Through the years, Bree provided moral support and helped me navigate the challenging endeavor of raising children while conducting scientific research. Bree has challenged me to be a better scientist and fostered confidence in my own abilities. Third, a sincere thank you to David Ackerly, who somehow always knew the perfect questions to ask which helped improve my work. Over the years, your insightful feedback has been invaluable. When you were part of my qualifying exam committee, you continually inspired me to learn more. I greatly enjoyed the challenge of trying to keep up with your quick thinking. David is an inspiration for me to be a better, and more well-rounded, scientist.

I also want to acknowledge the pivotal role of a few graduate students. First, to Matthew Hughes, thank you for being a great lab-mate. You provided invaluable lab, field, and moral support. Plus, I enjoyed all of our great conversations about trees. Also, to Prahlada Papper, thank you for walking me though lab work and the initial steps of data analysis. You made bioinformatics look easy. Thank you for your patience and ongoing support with R. Your help has been invaluable. With how demanding our graduate research can be, it is humbling that you took so much time to help me.

Most of all, I would like to thank my amazing family. To my husband, Jason, you are my everything. It is because of you that I have gotten to this point in my education. At every critical juncture, you believed in me, encouraged me, and been unwavering in your support for my aspirations. Words cannot express how lucky I am to have you as my partner in this life. To my children, Autumn and Jake, you are my inspiration. The sincerity, purity, and joy that you bring to everything you do, inspires me to work harder towards creating a world worthy of the potential you both have.

Introduction

SPECIES BACKGROUND

Giant sequoia, *Sequoiadendron giganteum* [Cupressaceae] is the only extant member of its genus. It is a paleoendemic, long-lived tree species whose current range stretches approximately 400km North-to-South from Placer to Tulare counties and consists of ~70 groves, scattered across the mid-elevation Sierra Nevada mountains. Its elevational range predominantly occurs within 1800-2100m, yet a few groves extend into elevations between 1500-2200m and scattered individual trees have been noted as low as 900m along drainages (Hartesveldt et al. 1975). The entire range of giant sequoia is fragmented. However, populations are highly disjunct from Placer county to the Kings River, and somewhat more contiguous from the Kings river south to the southernmost grove, Deer Creek, in Tulare county. Giant sequoia range is situated within the mixed-conifer forest. Although giant sequoia is often dominant within groves, it co-occurs with many tree species including, *Abies concolor, Calocedrus decurrens*, *Pinus ponderosa*, and *Pinus lambertiana.*

The timeline for the establishment of the extant range of giant sequoia on the western slopes of the Sierra Nevada remains somewhat obscure due to limited fossil and molecular evidence. Some fossil evidence suggests changes in local abundance, or short distance range shifts at various time periods in the late Pleistocene and Holocene (Anderson and Smith 1994, Cole 1983, Anderson 1994, Koehler and Anderson 1994). For instance, in one record from a full glacial period in Sequoia and Kings Canyon National Park, (SKCNP), Cole (1983) notes the presence of giant sequoia remains in packrat middens from 40-18 kya (thousands of years before present), which subsequently drops out of the more modern record. Currently, giant sequoia grows approximately 2 km from the midden site, which suggests that a short distance shift in grove boundaries may have occurred in the recent past. Moreover, another pollen core from SKCNP, within present day Giant Forest grove, shows low levels of sequoia pollen from approximately 10.5 kya, that slowly increases until 5kya, indicating an increase in abundance at this site (Anderson 1994). In the central and northern portion of the giant sequoia range, fossil records are extremely sparse. Two studies of note in this region include Davis and Moratto (1988) who reported giant sequoia pollen at Exchequer Meadow at approximately 11 kya and Koehler and Anderson (1994) who found macrofossils present at Nelder grove from 11-9.5 kya. Currently, giant sequoia is absent from Exchequer Meadow, but it is present 5 km south at McKinley Grove, again suggesting a shift in grove location. Unfortunately, the record for Nelder grove lacks resolution and thus the changes in distribution or abundance of giant sequoia at this site remains unclear. Taken together this limited pollen and fossil evidence suggests that shifts in giant sequoia range have been ongoing until recently (5kya). Yet, molecular data do not show imprints of recent colonization for present-day groves (Dodd and DeSilva 2016). Instead, extant patterns of genetic diversity suggest a long-history for many groves (Dodd and DeSilva 2016, DeSilva and Dodd 2020). Thus, the most parsimonious explanation that reconciles fossil and molecular evidence is that modern-day groves have only shifted short distances by way of local elevational or microclimatic changes during recent glacial-interglacial phases (Dodd and DeSilva 2016). Of course, more extensive pollen sites or detailed molecular work would be very helpful to better understand the evolutionary history of giant sequoia range.

Giant sequoia exhibits a unique regeneration strategy that is tied to fire and an environmentally sensitive seedling stage. The reproductive biology of giant sequoia is commonly characterized by episodic bursts of regeneration when conditions are optimal. Mature giant sequoia trees produce an abundance of wind-pollinated seed-cones each year. These cones remain closed and attached to the tree for many years, and thus provide a large aerial seed bank (Hartesveldt et al. 1975). Although Douglas squirrels contribute annually to seed dispersal, successful regeneration often occurs after fire, which not only triggers extensive seed dispersal, but also creates canopy gaps and a mineral substrate that increases successful seed germination and growth (Rundel 1972; Hartesveldt et al. 1975, Harvey et al. 1980; York et al 2003, Shellhammer and Shellhammer 2006, York et al. 2009). Many ecological studies have noted that giant sequoia seedlings are sensitive to water and light availability during their establishment phase and grow best under conditions of high resource availability (Rundel 1972, Hartesveldt et al. 1975, York et al. 2003, Shellhammer and Shellhammer 2006). Giant sequoia can grow very fast under prime conditions of high moisture and light (York et al. 2003). It is commonly noted that the lower elevational range limit of giant sequoia is controlled by a lack of sufficient soil moisture to sustain sequoia seedlings and the upper elevational limit is controlled by cold temperatures (Rundel 1972, Hartesveldt et al. 1975). In support of this assertion, researchers have found that giant sequoia seedlings experience high mortality due to desiccation (Rundel 1972, Weatherspoon 1990). In contrast, mature giant sequoias can be fairly robust to short periods of unfavorable climate (Stephenson et al. 2018).

ENVIRONMENTAL CONTEXT

Climate change poses great threats to biodiversity. In California, end of century climate predictions include temperature increases ranging from 1.5-4.5ºC (Cayan et al. 2008) and altered precipitation patterns. The Sierra Nevada is a high mountain range that collects precipitation from the Pacific Ocean mostly in the form of winter rain and snowfall. The slow release of water from snowmelt in the spring is an important source of moisture for seedling growth and establishment. Sierra snowpack has declined in recent years (Fyfe et al. 2017). With warmer temperatures, regional climate models suggest that spring snow-water equivalent will decline by 73% by the end of the century, with mid-elevations (1500-2500m) experiencing the greatest declines (Sun et al. 2018). Taken together, increases in temperature and earlier snowmelt will cause tree seedlings to experience additional heat stress coupled with an accentuation of the summer drought that is typical of Mediterranean climates. A recent drought in California resulted in massive tree mortality (USDA 2016), exposing the risks of catastrophic disease and mortality for some tree species from climate change.

During this century, climate warming and altered precipitation patterns will lead to habitat changes that may be especially detrimental to long-lived tree species. To persist through these changes, tree species can shift their range or adapt to new conditions (Aitken et al. 2008). Although species distribution models predict shifts of a species suitable habitat poleward and upslope (Iverson et al. 2008), for many tree species, long generation-times limit the potential for a rapid migratory response. As a long-lived tree species, giant sequoia exhibits generation-times of ~305 years (Dodd and DeSilva 2016). Thus, given both the pace of climate change and the specific suite of conditions needed for giant sequoia recruitment it is unlikely that range shifts can keep pace with climate changes. Under migratory limitation, populations of giant sequoia are likely to experience conditions to which they are maladapted. Thus, the adaptive potential of

giant sequoia populations will be a major determinant of persistence. For giant sequoia, the current pool of genetic diversity will provide the fuel for adaptive evolution in the face of climate change (Aitken et al. 2008; Barrett and Schluter 2008). Thus, understanding patterns of extant diversity and the factors that create, erode, or move diversity is of primary concern.

The overarching goal of this dissertation is to understand whether viable populations of giant sequoia are likely to exist in the future and highlight at-risk populations, as well as populations that may contain important functional variation. The results of this research can provide invaluable information to forest managers who wish to take steps to maintain genetic diversity and facilitate adaptive responses across the range of giant sequoia.

CHAPTER ONE

Considering the fragmented nature of the giant sequoia range, connectivity among spatially separated populations through gene flow has numerous implications for the maintenance and distribution of genetic variation. Gene-flow is an important mechanism by which variation is moved within and maintained across a species' range. Fragmented populations, typical of giant sequoia, tend to have reduced gene flow (Aguilar et al. 2008, Dubreuil et al. 2010). Yet, connectivity can be maintained across fragmented landscapes via long distance gene flow (Bacles et al. 2006, O'Connell et al. 2007, Colabella et al. 2014), commonly by pollen. In addition, species that exhibit small population sizes, a typical feature of giant sequoia, can be vulnerable to losing genetic diversity over time through genetic drift. Gene flow can counteract this trend and reduce diversity loss (Slatkin 1987), as well as to rapidly introduce new variation into populations, which can be a source of adaptive potential (Young et al. 1996, Kremer et al. 2012). Under climate change, gene flow has the added potential to spread pre-adapted variation (Kremer et al. 2012). For example, locally adapted populations within the warm-dry climatic range extremes could be pre-adapted to the future conditions expected in other areas of giant sequoia distribution.

The dynamics of gene flow and range-wide population connectivity are examined in the first chapter of this dissertation. Using patterns in range-wide genetic structure among 568 individuals, from 19 groves, I report long-term genetic connectivity with estimates of recent and historic gene-flow across spatially adjacent populations. Gene flow among populations may depend on the success of immigrant genes in their novel environment, so I assessed the role of landscape dissimilarity, including climatic variation, on restricting gene flow and ultimately shaping genetic differentiation among populations.

CHAPTER TWO

In Chapter One, I reported surprisingly low gene flow between spatially close groves in the northern range of giant sequoia that was not evident in southern groves. In this chapter, I addressed the fine scale processes of gene flow at the scale of individual groves. This included possible evidence for local demic structure that can arise from short distance pollen and seed dispersal and subsequent mating among close relatives (bi-parental inbreeding). Dispersal shapes the clustering of related individuals on the landscape and thus contributes to changes in genetic diversity over time. Two aspects of dispersal are of major importance, mean dispersal distance and the shape of the dispersal kernel. Mean dispersal provides important insights on the extent to which seed and pollen cluster near the parent tree. Subsequently, the spatial clustering of related

individuals can affect mating parameters such as biparental inbreeding (Krakosksi et al. 2003, Lloyd et al. 2018) and also increase the potential for genetic diversity loss through genetic drift or from a mortality disturbance (Willi and Määttänen 2011). Whereas, the shape of the dispersal kernel (kurtosis) can be an important indicator of the potential for long-distance gene flow between fragmented populations (Nathan et al. 2008, discussed in more detail below). The dispersal distance of pollen is generally greater than that of seed in wind-pollinated trees such as giant sequoia (Dow and Ashley 1996, Bittencourt and Sebbenn 2007). Thus, understanding the pollen dispersal kernel is a major focus of this work, for which I used progeny arrays, in the form of laboratory germinated seed.

The information presented in this chapter synthesizes my estimates of gene dispersal among adult trees and pollen dispersal with estimates of local mating and fine-scale grove genetic structure. It will help determine how diversity may change through time within populations, (i.e. how populations may be impacted by genetic drift or disturbance), and the potential for long-distance gene flow among fragmented giant sequoia populations.

CHAPTER THREE

Natural selection is of key importance in shaping patterns of genetic diversity across a species range. When populations occur across a heterogenous landscape, divergent selective pressures can result in local adaptation of populations to their home environment. Patterns of local adaptation can contain information about species tolerances and sensitivity to different ecological factors. In the context of climate change, identifying the major climatic factors that influence the selective landscape will inform how the species may respond to altered conditions. Moreover, understanding the distribution of adaptive diversity in relation to climatic gradients is a critical first step in promoting the adaptive potential of populations (Holderegger et al. 2006; Aitken and Whitlock 2013). Furthermore, highlighting populations or regions that potentially contain adaptively important variation will inform future management efforts that seek to maximize adaptive capacity in the species.

In the third chapter of this dissertation, I utilize a genome-wide data set of 1364 bi-allelic single nucleotide polymorphisms, to uncover signatures of local adaptation by identifying highly differentiated loci and those with a strong association to climate factors. This analysis provides evidence of the specific climate factors that are of major selective importance for giant sequoia and can highlight loci of potential adaptive significance. Considering the anticipated climatic changes for sierra Nevada ecosystems, we also identify populations that could contain genetic variation that may offer ‘pre-adaptations’ to future conditions.

OVERARCHING THEMES

The persistence of giant sequoia under climate change will be aided by adaptation of populations to new environmental conditions. The capacity for this response (adaptive potential) will depend on the extent and distribution of genetic diversity, whether it can move among populations, and whether variation exists that is pre-adapted to future conditions. Thus, assessing extant patterns of diversity and the factors that shape changes in diversity over time is of utmost importance to predicting species response. In this dissertation, I address the major determinants of genetic diversity in giant sequoia at multiple spatial scales and in relation to various evolutionary forces. During the course of this study, I identified populations of giant sequoia that

are at high risk (i.e. those that are genetically depauperate), those of high conservation priority (i.e. those that are genetically distinct), and populations that may harbor vital diversity to aide in adaptive responses across the range (i.e. those that are likely adapted to arid/warm conditions).

REFERENCES

Aguilar, R., M. Quesada, L. Ashworth, Y. Herrerias-Diego, and J. Lobo. 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology* 17: 5177-5188.

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. L., and Curtis-McLane, S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications. 1: 95–111.

Aitken, S. N., and Whitlock, M. C., 2013. Assisted gene flow to facilitate local adaptation to climate change. Annual Review of Ecology, Evolution, and Systematics 44: 367–388.

Anderson, R. S. 1994. Paleohistory of a Giant Sequoia Grove: The Record from Log Meadow, Sequoia National Park. Pp. 49–55 in P. S. Aune, ed. Proceedings of symposium on giant sequoias: their place in the ecosystem and society. USDA Forest Service Gen. Tech. Rep.PSW-151.

Anderson, R. S., and S. J. Smith. 1994. Paleoclimatic interpretations of meadow sediment and pollen stratigraphies from California. Geology 22:723–726.

Bacles, C. F. E., A. Lowe, and R. Ennos. 2006. Effective Seed Dispersal across a Fragmented Landscape. *Science,* 311(5761): 628-628.

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

Bittencourt, J., Sebbenn, A. Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. *Heredity* 99, 580–591 (2007). https://doi.org/10.1038/sj.hdy.6801019

Cayan D., Maurer E. P., Dettinger M., Tyree M., and K. Hayhoe. 2008. Climate change scenarios for the California region. Climatic Change. 87 (Suppl 1):S21-S42.

Colabella, F., L.A. Gallo, A. C. Moreno, and P. Marchelli. 2014. Extensive pollen flow in a natural fragmented population of patagonian cypress austrocedrus chilensis. *Tree Genetics and Genomes* 10(6): 1519-1529.

Cole, K. L. 1983. Late Pleistocene vegetation of Kings Canyon, Sierra Nevada, California. Quaternary Research. 19:117–129.

Davis, O.K., and M.J. Moratto. 1988. Evidence for a warm dry early Holocene in the western Sierra Nevada of California: pollen and plant macrofossil analysis of Dinkey and Exchequer Meadows. Madroño 35:132-149.

DeSilva R, and Dodd R. S. 2020. Fragmented and isolated: limited gene flow coupled with weak isolation by environment in the paleoendemic giant sequoia (Sequoiadendron giganteum). American Journal of Botany. 107 (1): 45-55

Dodd, R. A., and DeSilva, R. 2016. Long-Term Demographic Decline and Late Glacial Divergence in a Californian Paleoendemic: Sequoiadendron giganteum, giant sequoia. Ecology and Evolution 6: 3342-3355.

Dow, B.D. and Ashley, M.V. (1996), Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Molecular Ecology, 5: 615-627. doi:10.1111/j.1365-294X.1996.tb00357.x

Dubreuil, M., M. Riba, S. C. González-Martínez, G. G. Vendramin, F. Sebastiani, and M. Mayol. 2010. Genetic effects of chronic habitat fragmentation revisited: Strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability. *American Journal of Botany* 97: 303-310.

Fyfe, J., Derksen, C., Mudryk, L. Flato, G. M., Santer B. D., Swart, N. C., Molotch, N. P., Zhang. X., Wan, H., Arora. V. K., Scinocca, J., and Jiao, Y. 2017. Large near-term projected snowpack loss over the western United States. Nature Communications 8: 14996. doi:10.1038/ncomms14996

Hartesveldt, R. J., H. T. Harvey, H. S. Shellhammer, and R. E. Stecker. 1975. The giant sequoia of the Sierra Nevada. U. S. Department of the Interior, National Park Service, Washington DC, USA.

Harvey, T. H., Shellhammer, H. S. and Stecker, R.E. 1980. Giant Sequoia Ecology: Fire and Reproduction. Scientific Monograph Series No. 12. U.S. Department of the Interior, National Park Service. Washington, D.C.

Iverson, L.R., Prasad, A.M., Matthews, S.N. & Peters, M. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. Forest Ecology and Management, 254: 390–406.

Holderegger, R., Kamm, U., and Gugerli, F. 2006. Adaptive vs. neutral genetic diversity: Implications for landscape genetics. Landscape Ecology. 21(6): 797-807.

Koehler and Anderson 1994. The Paleoecology and stratigraphy of Nichols Meadow, Sierra National Forest, California, USA. Palaeogeography, Palaeoclimatology, Palaeoecology. 112: 1-17

Kremer, A., O. Ronce, J. J. Robledo-Arnuncio, F. Guillaume, G. Bohrer, R. Nathan, J. R. Bridle, et al. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* 15: 378-392.

Krakowski, J., Aitken, S. & El-Kassaby, Y. Inbreeding and conservation genetics in whitebark pine. *Conservation Genetics* 4, 581–593 (2003). https://doi.org/10.1023/A:1025667700479

Maurer E. P. 2007. Uncertainty in hydrologic impacts of climate change in the Sierra Nevada, California, under two emissions scenarios RID C-7190-2009. Climatic Change 82:309-325.

Lloyd, M. W., Tumas, H. R., and Neel, M. C.. 2018. Limited pollen dispersal, small genetic neighborhoods, and biparental inbreeding in *Vallisneria americana*. *American Journal of Botany* 105(2): 227– 240.

Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. and Tsoar, A. 2008. Mechanisms of long-distance seed dispersal. *Trends Ecol. Evol.*, 23, 638– 647.

O'Connell, L. M., A. Mosseler, and O. P. Rajora. 2007. Extensive Long-Distance Pollen Dispersal in a Fragmented Landscape Maintains Genetic Diversity in White Spruce. *Journal of Heredity* 98(7): 640–645.

Petit and Hampe. 2006. Some Evolutionary Consequences of Being a Tree.Annual Review of Ecology, Evolution, and Systematics.37: 187-214

Rehfeldt GE, Ying CC, Spittlehouse DL, Hamilton DA. 1999. Genetic responses to climate change in Pinus contorta: niche breadth, climate change, and reforestation. Ecol. Monogr. 69:375–407

Rundel P. 1972.Habitat Restriction in Giant Sequoia - Environmental Control of Grove Boundaries. American Midland Naturalist 87:81-99

Savolainen, O., T. Pyhajarvi, and T. Knurr. 2007. Gene Flow and Local Adaptation in Trees. Rev. Ecol. Evol. Syst. 2007. 38:595–619

Shellhammer, H. S., and T. H. Shellhammer. 2006. Giant sequoia (Sequoiadendron giganteum Taxodiaceae) seedling survival and growth in the first four decades following managed fires. Madrono 53:342-350.

Slatkin M. 1987. Gene flow and the Geographic Structure of Natural Populations. *Science* 236: 787-792

Stephenson, N. L., Das, A. J. Ampersee, N. J. Cahill, K. G., Caprio, A. C., Sanders., J. E., and Williams, A. P. 2018. Patterns and correlates of giant sequoia foliage dieback during California's 2012–2016 hotter drought. Forest Ecology and Management, Volumes 419–420, 268-278

Stewart I., D. Cayan, and M. Dettinger. 2004. Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. Climatic Change 62:217-232.

Sun, F., Berg, N., Hall. A., Schwartz, M., and Walton, D. 2018. Understanding end-of-century snowpack changes over California's Sierra Nevada. Geophysical Research Letters. 46: 933-943

[USDA–FS] US Department of Agriculture–Forest Service. 2016. New Aerial Survey Identifies More than 100 Million Dead Trees in California. Accessed January 1st 2020. https://www.fs.usda.gov/news/releases/new-aerial-survey-identifies-more-100-million-dead-trees-california

Weatherspoon, C. P., 1990. Sequoiadendron giganteum (Lindl.) Buchholz Giant Sequoia. In: Burns, R.M., Honkala, B.H. (Eds.), Silvics of North America, USDA Forest Service Agricultural Handbook 654, Washington, DC, pp. 552–562.

Willi, Y. and Määttänen, K. 2011. The relative importance of factors determining genetic drift: mating system, spatial genetic structure, habitat and census size in *Arabidopsis lyrata*. New Phytologist, 189: 1200-1209. doi:10.1111/j.1469-8137.2010.03569.x

York R., J. Battles, and R. Heald. 2003. Edge effects in mixed conifer group selection openings: tree height response to resource gradients. Forest Ecology and Management 179:107-121

York, R. A., Z. Thomas, and J. Restaino. 2009. Influence of Ash Substrate Proximity on Growth and Survival of Planted Mixed-Conifer Seedlings. Western Journal of Applied Forestry 24:117-123.

Young, A., T. Boyle, and T. Brown. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends Ecology and Evolution* 11: 413–418

Chapter 1

Fragmented and isolated: limited gene flow coupled with weak isolation by environment in the paleoendemic giant sequoia (*Sequoiadendron giganteum*)

Rainbow DeSilva, Richard S. Dodd

ABSTRACT

Patterns of genetic structure across a species' range reflect the long-term interplay between genetic drift, gene flow, and selection. Given the importance of gene flow in preventing the loss of diversity through genetic drift among spatially isolated populations, understanding the dynamics of gene flow and the factors that influence connectivity across a species' range is a major goal for conservation of genetic diversity. Here we present a detailed look at gene flow dynamics of *Sequoiadendron giganteum*, a paleoendemic tree species that will likely face numerous threats due to climate change. We used microsatellite markers to examine nineteen populations of *S. giganteum* for patterns of genetic structure and to estimate admixture and rates of gene flow between eight population pairs. Also, we used Generalized Dissimilarity Models to elucidate landscape factors that shape genetic differentiation among populations. We found minimal gene flow between adjacent groves in the northern disjunct range. In most of the southern portion of the range, groves showed a signal of connectivity which degrades to isolation in the extreme south. Geographic distance was the most important predictor of genetic dissimilarity across the range, with environmental conditions related to precipitation and temperature explaining a small, but significant, portion of the genetic variance. Due to their isolation and unique genetic composition, northern populations of *S. giganteum* should be considered a high conservation priority. In this region, we suggest germplasm conservation as well as restoration planting to enhance genetic diversity.

INTRODUCTION

Understanding the patterns of genetic diversity and population differentiation is a central goal of evolutionary biology. These patterns are a result of long-term trends in species distribution and evolutionary forces (e.g., gene flow, selection, and genetic drift) occurring over space and time (Loveless and Hamrick, 1984; Slatkin, 1985, 1987). Many demographic changes, such as population size reductions or range contractions with subsequent population isolation, create fragmented populations that tend to experience reduced gene flow, which in turn results in

strong population structure and the loss of genetic diversity over time (Young et al., 1996; Segelbacher et al., 2003; Aguilar et al., 2008; Dubreuil et al., 2010; Arenas et al., 2012). Yet, this is not always the case. In forest trees, connectivity can be maintained across fragmented landscapes via long distance dispersal, as in *Austrocedrus chilensis* (D.Don) Pic. Serm. & Bizzarri [Cupressaceae], *Picea glauca* (Moench) Voss [Pinaceae], and *Fraxinus excelsior* L. [Oleaceae] (Bacles et al., 2006; O'Connell et al., 2007; Colabella et al., 2014). Even at low levels, gene flow has the potential to reduce diversity lost via drift (Slatkin, 1987) and quickly introduce new variation into populations, which can be a source of adaptive potential (Young et al., 1996; Kremer et al., 2012).

A major goal of landscape genetics is to disentangle the relative contribution of landscape factors in influencing the spatial variation in gene flow (Holderegger and Wagner, 2008). One important factor inhibiting gene flow is dispersal limitation, which results in patterns of isolation by distance (IBD), meaning proximate populations are more likely to share migrants than distant populations (Wright, 1943). However, many other factors, including population size, landscape barriers, and ecological gradients, play an important role in shaping gene flow and differentiation among populations (Ellstrand and Elam, 1993; Lee and Mitchell-Olds, 2011; Shafer and Wolf, 2013). Of particular concern is elucidating how local adaptation can negatively affect immigrant success and thereby reduce gene flow and result in a pattern of increased genetic divergence with environmental distance, commonly referred to as isolation by environment (IBE) (Nosil et al., 2008; Wang and Summers, 2010).

Sequoiadendron giganteum (Lindl.) J.Buchh., giant sequoia, [Cupressaceae] is an iconic tree species, renowned for its massive size and long lifespan, often exceeding 2000-3000 years in age (Stephenson, 2000; Sillett et al., 2015). This species occurs in ~70 groves in a narrow elevation band (mostly within 1400–2200 m) along the western slope of the Sierra Nevada Mountains, California, USA, that extends approximately 420 km north-to-south (Figure 1). *Sequoiadendron giganteum* groves vary in size from 1 ha to 1624 ha and occupy a total area of ~14,600 ha (Harvey et al., 1980; York et al., 2013; Nydick et al., 2018). The vast majority of *S. giganteum* grove area is concentrated in the southern portion of the range (York et al., 2013). Here, populations are fairly continuous, yet often separated from each other by approximately 2–10 km, whereas in the northern two thirds of the latitudinal range, the remaining eight populations are highly disjunct. In view of the potential threats posed by climate change and by fire for this restricted paleoendemic, we have been studying genetic structure of populations to better understand their recent evolution, genetic potential to respond to environmental challenges, and distribution of genetic resources that can provide material for future restoration and reforestation.

It is surprising that so little is known of the population genetics of such an iconic species. In our previous work (Dodd and DeSilva, 2016), we showed that the northern isolated populations were genetically distinct from those in the southern range and concluded from demographic simulations that *Sequoiadendron giganteum* had likely experienced a long-term decline in population size over the last ~2 My, with a more severe contraction prior to the last glacial maximum. We inferred that the fragmented distribution today was a result of this demographic decline, rather than a result of colonization by long-distance dispersal. Indeed, we found some evidence that would suggest that gene flow is quite restricted in the northern range

of the species (Dodd and DeSilva, 2016). The pollen record suggests subsequent local expansion of *S. giganteum* in a few regions during the Holocene (Anderson, 1994; Anderson and Smith, 1994).

Currently, range boundaries of *Sequoiadendron giganteum* remain stable, as the vast majority of groves are located in protected areas, but in the future, climate change may result in a mismatch between protected areas and suitable habitat. This raises two questions as to the maintenance of genetic diversity through gene flow amongst extant populations and the degree to which environment may affect migrant success: (1) Do fragmented populations of *S. giganteum* have the potential to maintain genetic connectivity through gene flow? and (2) Is there evidence that isolation by environment has played a role in population genetic structure? Gene flow dynamics will shape how *S. giganteum* populations respond to rapidly changing environmental conditions and will help in designing strategies for the protection of this remarkable natural resource.

In this study, we tested three hypotheses: (1) gene flow is restricted in the northern disjunct range, even between locally close populations. This would support our previous findings that the northern groves are likely to have had a longer history than previously thought, or that they arose through relatively short distance spatial shifts of paleo-groves; (2) southern populations should show a signal of greater gene flow that would be consistent with connectivity among the network of groves in this region; and (3) isolation by environment, potentially resulting from constraints on the success of migrants, has been important in shaping genetic dissimilarity. We tested these hypotheses by analyzing allelic variation in 19 natural populations for a suite of microsatellite markers designed for giant sequoia (DeSilva and Dodd, 2014).

MATERIALS AND METHODS

Sampling and SSR genotyping

Between 2015 and 2018, foliage was collected from 562 individuals representing eighteen groves distributed across the range of *Sequoiadendron giganteum* (Figure 1). Isolated DNA from an additional grove (PLAC) was obtained from Valerie Hipkins (USDA Forest Service). Geographic coordinates for all samples were obtained at the time of collection using a Garmin eTrex Global Positioning System (Garmin, Olathe, Kansas, USA). All samples were collected with permission.

The resulting data set consists of 568 individuals from 19 populations (Table 1) spanning the entire latitudinal range of the species, including all eight populations in the northern two thirds of the range (Figure 1, 'northern group') and eleven populations spread across the southern portion of the range, (Figure 1, 'southern group'). DNA was extracted from 100mg of leaf tissue using Norgen Biotek Plant/Fungi DNA Isolation kits (Norgen Biotek, Thorold, ON, Canada). All individuals were genotyped at eleven microsatellite loci described in DeSilva and Dodd (2014). DNA amplifications were achieved using a M13 tail attached to the forward primer and a FAM-labeled universal M13 fluorescent primer, following the technique outlined by Schuelke (2000). Each PCR reaction consisted of a 10µl reaction volume containing 1 µl 10x PCR buffer, 0.2µl $MgCl_2$, 200µmol/L of each dNTP (Roche, Santa Clara, Calif., USA), 0.16µmol/L reverse and universal M13 primer, 0.04µmol/L forward primer, 0.5 units of HotStarTaq polymerase (Qiagen, Valencia, Calif., USA), and 20-40ng of extracted DNA. The resulting solution was subjected to the following temperature treatments: initial denaturation at 95°C for 15min; 30 cycles of 30s at

95°C, 45s at 56°C, and 45s at 72°C; 8 cycles of 30s at 95°C, 45s at 53°C, and 45s at 72°C; finishing with a final extension of 30min at 72°C. The PCR product was then mixed with a solution containing 0.5 µl Genescan LIZ500 size standard (Applied Biosystems, Waltham, Mass., USA) and 8.0 µl formamide and processed on an ABI 3730 automated sequencer. SSR fragment lengths were visualized and called using GeneMapper 5 software (Applied Biosystems).

Genetic diversity, structure, and differentiation

All SSR data were checked for null alleles and evidence of scoring errors using Microchecker (Oosterhout et al., 2004). Deviations from Hardy-Weinberg equilibrium and Linkage Disequilibrium were tested using Genepop through their web interface (Raymond and Rousset, 1995; Rousset, 2008) and significance for both analyses was assessed after Bonferroni corrections for multiple tests (Rice, 1989).

We used FSTAT version 2.9.3 (Goudet, 2001) to calculate gene diversity (H_E) allelic richness (corrected for sample size) (A_R), and pairwise F_{ST}, and we performed permutation tests to obtain one-tailed significance tests in FSTAT with 1000 iterations to compare H_E, A_R, and F_{ST} between the northern and southern population groups, assuming less connected populations in the north would on average have lower genetic diversity and greater divergence. PLAC, an outlier with low A_R, and high F_{ST} (Table 1) was excluded from this test. We estimated conditional genetic distance (cGD), a measure of genetic dissimilarity between populations, using the GSTUDIO package implemented in R (Dyer et al., 2010; Dyer, 2014; R Core Team, 2018). This metric simultaneously takes into account the genetic covariance among all populations and thereby measures the genetic similarity that is created by both direct and indirect gene flow (Dyer and Nason, 2004; Dyer et al., 2010).

To assess the evolution of population structure we tested for a signal of mutations contributing to the observed population structure, which would be a signal of ancient divergence between northern and southern groups (Figure 1). These tests compare the observed R_{ST} with a value of R_{ST} obtained after permuting allele sizes within loci (R_{STperm}). Permuting allele sizes removes the effects of stepwise mutation among microsatellite alleles but retains the divergence among populations due to allele frequency differences and thereby R_{STperm} approximates F_{ST} (Hardy et al. 2003). R_{STperm} is then compared to R_{ST} to estimate the effects of mutation on population divergence. We performed analyses in SPAGeDi V 1.5 (Hardy and Vekemans 2002), using 20,000 permutations.

To estimate patterns of genetic structure across the species range, we utilized two complementary approaches, STRUCTURE (Pritchard et al., 2000) and TESS (Chen et al., 2007; Durand et al., 2009b). The program STRUCTURE version 2.3.3, utilizes a Bayesian clustering algorithm to simultaneously estimate the number of genetic clusters (K) and assigns individuals to a genetic cluster based on their multi-locus genotypes, without a priori information about geographic origin (Pritchard et al., 2000). To visualize partitions in genetic clusters and find likely K values, we utilized CLUMPAK (Kopelman et al., 2015). The best K values were determined using the K value that maximized Pr(Data| K) (Pritchard et al., 2000), and the ΔK statistic as suggested by Evanno et al. (2005). To reduce computing time, we narrowed the range of potential K values by completing exploratory tests using 10 repetitions of K = 1–19. After

initial runs, we used 20 replications and a burnin of 300,000 and 1,000,000 Markov Chain Monte Carlo (MCMC) repeats after burnin with K = 2–13 (a reduced range of reasonable K-values obtained from exploratory runs).

We also utilized TESS, a spatially explicit method for genetic clustering (Chen et al. 2007). TESS utilizes the geographic coordinates of each individual to inform the estimation of spatial autocorrelation in allele frequencies (Durand et al. 2009b). In this algorithm, the probability that two individuals belong to the same cluster decreases with geographic distance (Durand et al. 2009b). We used the admixture model (CAR) as it is robust to various levels of admixture in the data set (Durand et al. 2009a, Francois and Durand 2010). To determine the most likely number of clusters (K_{max}) we plotted the deviance information criterion (DIC) against the K_{max} and visually chose values of K_{max} where DIC reached a plateau, as suggested in the program documentation (Durand et al. 2009b). Using the same range of likely K_{max} values as in STRUCTURE, TESS was run using 120000 sweeps and a burnin of 20000 with 20 replicates for each K_{max}= 2 - 13.

Climatic effects on genetic divergence

Due to the long lifespan of *Sequoiadendron giganteum* (~3000 years), we chose to run our analyses using climate data for two time periods, Mid-Holocene (6kya) and current (1970–2000). For the current time period, nineteen bioclimatic variables were obtained from the WorldClim database at a spatial resolution of 30 arc-seconds (Fick and Hijmans, 2017). For the analysis of current climate only, we obtained Climate Water Deficit (CWD) from the California Basin Characterization Model (Flint et al., 2013). CWD is the difference between potential and actual evapotranspiration and provides an indication of aridity that is important for Mediterranean climate systems, such as in California (Stephenson, 1998). From the resulting set of 20 environmental variables, we used the sampling centroid of each of the populations to extract average site conditions for each population. To improve accuracy in differentiating between IBD and IBE, we removed variables that were highly correlated with geographic distance using Spearman's $\rho \geq 0.7$. We then further reduced redundancy among our environmental predictors by removing highly correlated variables ($\rho \geq 0.7$). When choosing between two highly correlated variables, we kept the one that was less correlated to the other remaining predictors. Our remaining set of predictors representing the current climate included eight variables; five associated with temperature, two with precipitation, and CWD (an index of aridity) (Appendices 1.1 and 1.2). Nineteen downscaled bioclimatic variables were obtained from WorldClim for the mid-Holocene time period under the CCSM4 model (Fick and Hijmans, 2017). After removal of correlated variables, the mid-Holocene climate was represented by nine climate variables (Appendix 1.1).

We utilized Generalized Dissimilarity Modeling (GDM) (Ferrier et al., 2007) to assess the correlations between environmental variation and geographic distance with genetic dissimilarity among populations. This method uses matrix regression to look for relationships between dissimilarities in predictor and response variables and has been increasingly applied to landscape genetic studies (Freedman et al., 2010; Thomassen et al., 2010; Geue et al., 2016). We used both environmental variation and geographic distance as predictor variables, and we used genetic dissimilarity, as measured by cGD, as the response variable. Conditional genetic distance was the chosen metric of dissimilarity due to its increased potential to detect landscape genetic patterns under various migration models and its reduced sensitivity to confounding effects of

phylogenetic history (Dyer et al., 2010). In addition, cGD captures dimensional migrant exchange among multiple populations, making it potentially superior to traditional metrics (e.g., pairwise F_{ST}) for assessing connectivity in *Sequoiadendron giganteum* (Dyer and Nason, 2004).

For these analyses, we utilized the 'gdm' package in R (Manion et al., 2015). The northernmost population, PLAC, was a strong geographic and cGD outlier and was therefore excluded from all analyses. Due to evidence of a strong genetic divide at the Kings River drainage (Dodd and DeSilva, 2016) separating the northern and southern groups (Figure 1), we ran our analyses separately for both the range-wide data set and the eleven southernmost populations. We conducted model selection among un-correlated environmental variables with the varImp function within the 'gdm' package using 1000 permutations. This function uses matrix permutation and backward variable removal to aid in model selection (Manion et al., 2015). The 'best' model was chosen by removing all predictor variables that did not contribute to the explanatory power of the model while considering the point at which removing additional variables increased the P-value of the model.

Population connectivity

To capture genetic connectivity at short distances (<10 km), we employed STRUCTURE version 2.3.3 (Pritchard et al., 2000) to infer the degree of admixture for eight population pairs, three in the northern group and five in the southern group (Table 2). The Bayesian clustering framework implemented in STRUCTURE estimates a proportional assignment of each individual to a genetic cluster, and thus also provides a measure of joint ancestry to multiple clusters. We ran STRUCTURE using 10 replications with a burnin of 300,000 and 1,000,000 MCMC repeats after burnin with K varying from 1 to 4. We determined the level of admixture within populations as the average proportion of individuals per population pair that show between 30 and 70% assignment to each genetic cluster at K = 2. Using CLUMPAK (Kopelman et al., 2015), we determined the best K using the ΔK statistic as suggested by (Evanno et al., 2005) and visualized the degree of admixture.

For these same population pairs, we also examined the rates of gene flow through time using two complementary approaches, MIGRATE-n and BAYESASS. BAYESASS 3.0.4 employs Markov Chain Monte Carlo (MCMC) analyses to estimate recent (within the last ~3–5 generations) levels of interpopulation migration based on the linkage disequilibrium imprint it creates (Wilson and Rannala, 2003; Rannala, 2007). MIGRATE-n estimates long-term (approximately $4N_e$ generations) migration rates using a coalescent based approach (Beerli and Felsenstein, 2001; Beerli, 2006). For both analyses, we removed one locus (31670) that exhibited strong deviations from Hardy-Weinberg equilibrium as we were concerned that this locus may violate the requirement of MIGRATE-n that populations be in mutation-drift equilibrium (Beerli, 2012). In addition, for population pairs with unequal sample sizes, we randomly removed individuals from the larger sample of the pairwise grove comparisons to equal the smaller sample i.e., we re-sampled RMNT and GFOR, and NELD. For NELD, our migrate runs after re-sampling failed to converge, which appeared to be due to population structure within the two regions within NELD grove (see Appendix 1.3), Therefore, we randomly removed samples from one area within NELD.

During initial runs of BAYESASS we altered the mixing parameters Δa, Δm, Δf between 0.2 and 0.6 to achieve the optimal acceptance values that facilitate parameter mixing, as suggested by Rannala (2007). After finalizing mixing parameters, we completed 10 runs for each population pair, each with a different seed to check run consistency. Each run consisted of 10,000,000 iterations, a burnin of 1,000,000, and parameter values were sampled every 5000 iterations. As recommended by Meirmans et al. (2014), we report the parameter estimates from the 'best fit model' as indicated by the lowest Bayesian deviance (Faubet et al., 2007).

Using MIGRATE-n, we estimated mutation scaled migration rates with the requirement that loci are in migration-drift equilibrium and the assumption that migration rates are constant through time (Beerli, 2012). We used the Bayesian framework and a Brownian motion mutation model as it approximates the step-wise mutation rate for SSR markers and is computationally less intensive (Beerli, 2012). We also employed a constant mutation rate and used F_{ST} calculations as starting parameter values. We conducted exploratory runs to determine the range of values appropriate for uniform priors for θ ($4N_e\mu$) and M (m/μ), where N_e is the effective population size and m and μ are the per generation migration and mutation rates respectively. After uniform priors were determined, we ran one long chain that visited 100,000,000 parameter values and had a burnin of 50,000, and sampled parameter values every 100 steps. To compare long-term and recent migration rates, we converted the mutation-scaled migration rate M (m/μ), using an estimated mutation rate of 5×10^{-4} per generation as outlined in Chiucchi and Gibbs (2010). Using estimates of mutation scaled effective population size, θ, we also converted M to number of migrants per generation (Nm) using the formula $\theta \times M/4$.

RESULTS

Genetic diversity, structure, and differentiation

MicroChecker found significant evidence for null alleles at two loci (31670 and 31267), in seven and eight populations, respectively. The presence of null alleles was not consistent across populations for the two loci, so we chose to retain all loci. We detected significant linkage disequilibrium for 85 of 1045 locus pairwise tests among populations, but these were spread across populations and pairs of loci.

Genetic diversity, as measured by allelic richness (A_R) and gene diversity (H_E), was highest in GFOR, with $A_R = 4.08$, and $H_E = 0.64$, and lowest in PLAC, with $A_R = 1.73$, and $H_E = 0.20$ (Table 1; Appendix 1.4). FSTAT permutation tests show no significant difference in mean A_R (3.6 > 3.2 P = 0.11) or mean H_E (0.58 > 0.59 P = 0.59) for the northern region as compared to the southern region. Average pairwise F_{ST} was highest in PLAC (0.39) and lowest in GFOR and RMNT (0.08) (Table 1). PLAC was a clear F_{ST} outlier, as other Northern populations exhibited F_{ST} ranging from 0.11-0.24 (Table 1). F_{ST} was not significantly higher in the northern region when compared to the southern ($F_{STnorth} = 0.13$, $F_{STsouth} = 0.06$, $P_{north>south} = 0.07$), whereas Global F_{ST} was 0.143. R_{ST} permutation tests suggested an effect of stepwise mutations on divergence between the northern and southern groups of populations: $R_{ST} = 0.04$ and the $R_{STperm} = 0.02$, P ($R_{STobs} > R_{STperm}$) = 0.01.

Population genetic structure

In the STRUCTURE analyses, we found an optimum number of clusters at K = 5 (Figure 2). Among the key trends at K=5, southern groves (LMDW to DCRK) appeared quite homogeneous, but northern groves (PLAC to MKLY) tended to be more distinct one from

another (except NCAL and SCAL). Of the northern groves, PLAC formed a distinct cluster and MERC appeared highly distinct from its neighboring grove TUOL, which is only 3 km to the north. Six of the seven central groves were highly admixed, but LOST was relatively uniform and showed affinities to MERC and MKLY. The TESS analysis recovered nine clusters (Figure 2). It provided improved resolution over STRUCTURE by separating LOST, MERC and MKLY and by showing increased differentiation between PKSD and DCRK and between NCAL and SCAL, with SCAL exhibiting considerable admixture (Figure 2).

Climatic and geographic effects on genetic dissimilarity

Using GDM on the range-wide dataset with current climate conditions, our best model explained 22.4% of the variation in the observed genetic distance among populations ($P = 0.001$; Table 3). The predictor variables in this model included geographic distance, as well as three climatic variables, Precipitation of Warmest Quarter (PWQ), Temperature Seasonality (TS), and Climate Water Deficit (CWD). Geographic distance was important in explaining genetic dissimilarity in the range-wide analysis, explaining 12.6% of the variation when used as the only predictor ($P < 0.001$) (Table 3). The combination of climatic predictors was significant in explaining a small portion of the observed variation (1.8%, $P = 0.003$) (Table 3). Range-wide GDM models with Mid-Holocene climate predictors were consistent with contemporary models (Table 3). In addition to the three overlapping predictors (Table 3), Precipitation of Driest Quarter (PDQ) was also included in the model. GDM models on the southern group only were non-significant.

Population connectivity

Rates of STRUCTURE admixture at similar spatial scales were highly variable for the eight paired populations from across the range of *S. giganteum.* For the three northern population pairs, individual assignments were strongly associated with groves from which samples were obtained, and rates of admixture were only 5-7% (Table 2; Appendix 1.3). By contrast, admixture levels in the southern group were 70-100% for all but the two southernmost pairs, LMDW and CNHM, PKSD and DCRK, which exhibited only 12% and 4% admixture respectively and strong association of individuals with their grove of origin (Table 2; Appendix 1.3).

We found significant recent gene-flow (BAYESASS) only in population pairs from the mid-southern range; GRNT/RMNT, GFOR/ATWL and FMAN/MCTR (Figure 3). For each of these three population pairs, our inferred migration rates (fraction of individuals of migrant ancestry) showed evidence of asymmetry. For example, significant migration was detected from GRNT to RMNT (M=0.18, 95% confidence interval [CI] 0.05-0.31), ATWL to GFOR (M=0.26, CI 0.10 - 0.41), and FMAN to MCTR (M=0.22, CI 0.05-0.31) (Figure 3). For these population pairs none of the reverse estimates of migrant ancestry were significant. Across the three northern and two southernmost population pairs M was low (from 0.02- 0.07) and non-significant (Figure 3).

Mutation-scaled migration rates estimated by MIGRATE-n were only significant for RMNT to GRNT (Figure 3). Although no other long-term rates of gene flow obtained by MIGRATE-N were significant, we believe that estimates of Nm remain somewhat informative in providing an assessment of the relative connectivity across the range (Figure 3, Appendix 1.5).

Mean long-term rates of migration showed strong variation across the range that was largely concordant with contemporary gene flow patterns. We found low levels of long-term migration between each of the three northern population pairs (Nm = 0.01-0.73), higher rates of gene flow for the three out of five southern population pairs (Nm = 1.77-19.14), and limited migration among the two southernmost population pairs (Nm = 0.04-0.54). Spearman's correlation shows that estimates of long-term and recent gene flow show consistent patterns across the range (ρ = 0.22, P = 0.04) (Appendix 1.5).

DISCUSSION

Range-wide patterns of population structure and differentiation

Sequoiadendron giganteum and its close relatives once existed across a larger portion of North America (Axelrod, 1964; Harvey, 1985). Currently, this paleoendemic exhibits a fragmented range that includes a series of highly disjunct northern groves and a southern network of groves that are more spatially connected. Previously, we detected a strong genetic divergence among plants in a common garden originating from northern and southern groves (Dodd and DeSilva, 2016). Here, samples collected directly from nineteen groves covering the range of the species confirmed our previous results of a strong north-south divergence. Among the nineteen groves of *S. giganteum*, we found between five and nine genetic clusters depending on the clustering algorithm used. TESS, which uses spatial data, returned a greater optimal number of clusters than STRUCTURE, providing fewer anomalous cluster assignments and should probably be considered as being more biologically meaningful. These clusters indicated strong genetic differentiation among populations in the northern range, a distinct cluster in the extreme southern range, and central southern populations that appear more genetically cohesive. Increased genetic differentiation at the range periphery has been reported for many species, such as *Thuja occidentalis* L. [Cupressaceae] and *Pinus sylvestris* L. [Pinaceae], and has been attributed to isolation and lower rates of gene flow (Pandey and Rajora, 2012; Wójkiewicz et al., 2016). The results of our R_{ST} permutations indicate a potential impact of mutation on the population structure separating the northern and southern groups (Hardy et al., 2003). This finding is consistent with a long-term divergence of populations in the northern portion of the range as reported by Dodd and DeSilva (2016). Lower population sizes at the range periphery may also be contributing to genetic differentiation in this region by increasing the effects of genetic drift (Eckert, 2008 and references therein).

Connectivity across adjacent populations

We assessed connectivity among approximately equidistant population pairs in two ways: (1) from evidence of mixed ancestry, as inferred from admixture proportions estimated by STRUCTURE; and (2) from estimates of long-term (MIGRATE-n) and recent (BAYESASS) rates of migration. The variation in admixture proportions among population pairs was broadly supported by estimates of recent and long-term migration rates from BAYESASS and MIGRATE-n respectively. All methods indicated connectivity between three population pairs from the southern range, but no significant migration among population pairs in the northern range, or between the southernmost population pairs.

BAYESASS estimates showed asymmetric migration for two of the southern grove pairs, resulting in net immigration to RMNT from GRNT and to GFOR from ATWL, whereas migration between FMAN and MCTR was similar in both directions. Grove size is larger at

RMNT than at GRNT and at GFOR than at ATWL and approximately equal at FMAN and MCTR. The possibility that grove size can play a role in source-sink relationships is interesting and worth further investigation.

Estimates of long-term migration from MIGRATE-n were more equivocal due to lack of significance of most estimates. However, mean estimates of number of migrants per generation showed a pattern concordant with recent migration rates. Therefore, it seems likely that, although the rates of migration may have varied over time, the pattern among population pairs has likely remained more or less constant; less gene flow over approximately equal distances in the northern paired groves compared with those in the south of the range. This pattern suggests greater genetic connectivity in the south, even over similar spatial distance between groves. This could be explained by: (1) interactions among the network of groves in the south; and (2) by greater grove size in the south. Over 90% of giant sequoia range area is clustered in the southern portion of the distribution, where populations tend to be larger and more abundant over the landscape (York et al., 2013). This increased density of populations would lead to higher levels of genetic connectivity through stepwise gene flow across the region. Larger average grove size (numbers of individuals) should result in greater numbers of pollen and seed that might enhance opportunities for long-distance dispersal events (Nathan et al., 2008; Purves, 2009).

An important conclusion that can be drawn from our connectivity data is that direct gene flow in the northern range of *Sequoiadendron giganteum* appears limited across distances of 3–6 km, the maximum distance between adjacent groves in this region. Interestingly, this is not true of the southern range where significant migration was detected over a distance of ~10 km from GFOR to ATWL. We assume this to be due to indirect gene flow among the network of groves in the southern range. Consistency in range-wide patterns of current and historic gene flow suggests biological factors are limiting dispersal. This finding is supported by research showing poor pollen dispersal in *S. giganteum* (Anderson, 1990a). On an evolutionary scale, the virtual lack of gene flow between northern paired groves is surprising and seems to suggest that the groves do not have a very long shared spatial proximity. In other words, at least some of these northern pairs of groves might have originated from divergent source populations. Could these have colonized from ghost populations that have now gone extinct, from long-distance dispersal, or even from ancient human-assisted dispersal? Previously, we suggested that fragmentation of the range of giant sequoia probably dated to around the last glacial maximum and that subsequent spatial shifts in populations may have occurred over relatively short distances (Dodd and DeSilva, 2016). This was based on palynological studies that showed the presence of giant sequoia close to contemporary groves during the Holocene (Cole, 1983; Davis and Moratto, 1988). It is difficult to envisage highly divergent ghost populations giving rise to such genetically divergent, but spatially close groves, such as TUOL and MERC. To our knowledge, no other studies have attempted to determine the source populations for the northern portion of *S. giganteum* range. Moreover, the fossil pollen record for *S. giganteum* in this region is extremely limited and provides little clarity on the availability of source populations to colonize the northern range (Davis and Moratto, 1988; Anderson, 1990b; Koehler and Anderson, 1994). More extensive pollen sites would be very helpful in understanding the evolutionary history of giant sequoia and, in particular, the perplexing distribution of northern groves. The possibility of ancient human-assisted plantings is intriguing and worthy of further study. We hope to provide

some added clarity on these questions through direct studies of pollen and seed dispersal between the grove pairs and more detailed range-wide genetic data that we are currently analyzing.

Limitations to estimates of migration

Some surprising results raise questions concerning the reliability of our estimates of migration. Indeed, BAYESASS has been found to provide reliable estimates of migration only when differentiation among populations is high ($F_{ST} \geq 0.05$) (Faubet et al., 2007). Whereas our northern population pairs satisfy this criterion, some southern population pairs do not (Appendix 1.6). Most concerning was the lack of significance for most of the MIGRATE-n estimates. Due to the assumptions of a coalescent-based approach, divergence time (in generations) among populations should ideally be larger than N_e (Beerli, 2010). Our prior work (Dodd and DeSilva, 2016) suggested divergence ~68 generations ago and Ne for the northern population pairs, as a combined estimate per pair, ranging from 329 to 592 (~165 to 296 per population). Therefore, divergence time in *Sequoiadendron giganteum* is likely to be too recent for us to reject zero migration rates with confidence. Another caveat to our analysis is that considering that *S. giganteum* has generation times of more than 300 years (Dodd and DeSilva, 2016) and that individuals can live upwards of 3000 years, any distinctions between contemporary and long-term gene flow may be artificial. Despite these caveats, all the methods converge on similar patterns of gene flow: low in the northern pairs, higher in the mid-southern pairs, and low again in the southernmost populations.

Climatic and geographic effects on genetic dissimilarity

Our GDM models for the range-wide population set explained a moderate proportion of the observed genetic dissimilarity among populations and indicated a prevailing pattern of isolation by distance (IBD) with a more limited importance of isolation by environment (IBE). Although both IBD and IBE have been shown to be integral factors shaping genetic composition across many plant taxa including *Fraxinus angustifolia* Vahl [Oleaceae]*, Quercus engelmannii* Greene [Fagaceae]*, Caragana microphylla* Lam. [Fabaceae]*,* and *Helianthus petiolaris* Nutt. [Asteraceae] (Andrew et al., 2012; Ortego et al., 2012; Temunović et al., 2012; Xu et al., 2017), the results presented here should be treated with caution. The signal of IBE detected in our study is weak and explains a small portion of the variance. Our models indicated that both temperature and moisture related variables played a role in the genetic structure that we detected, which is consistent with ecological studies reporting that *Sequoiadendron giganteum* is highly sensitive to water availability, light, and temperature during its establishment phase (Rundel, 1972; Hartesveldt et al., 1975; York et al., 2003; Shellhammer and Shellhammer, 2006). However, GDM models were non-significant in the portion of the range that contains the vast majority of *S. giganteum* populations.

Although significant, a historical component of divergence confounds our IBD estimates. The IBD that the GDM model detects is likely driven by the strong divergence between the northern and southern groups as we reported previously (Dodd and DeSilva, 2016), and is also supported by the R_{ST} permutation tests presented here. Our earlier reports of a lack of IBD across *Sequoiadendron giganteum* range (Dodd and DeSilva, 2016) separated a historical component of divergence that is included in our IBD estimates here. Thus, the detected signal of IBD here might be better thought of as an unknown combination of IBD and isolation by time. More

comprehensive data is needed to provide clarity on IBE, IBD, and historic divergence across the range of *S. giganteum*.

CONCLUSIONS

Isolated and genetically distinct northern and extreme southern populations of *Sequoiadendron giganteum* should be a high conservation priority. In these regions, suggested conservation actions include building an extensive ex-situ germplasm reserve and restoration planting to promote the maintenance of genetic diversity. In an era of climate change, *S. giganteum* populations are facing novel environmental conditions to which they may be maladapted. Additional research aimed at understanding the adaptive potential of *S. giganteum* populations will be of paramount value in informing our response to this threat.

ACKNOWLEDGMENTS

The authors thank Matthew Hughes for providing vital assistance in foliage sampling efforts and Valerie Hipkins for providing leaf tissue from Placer grove. We also thank reviewers whose feedback helped improve the manuscript. This research was partially supported by a Save the Redwoods League-USA (Grant/Award number) '070' awarded to R.S.D. This dissertation chapter was published in American Journal of Botany in January 2020.

REFERENCES

Aguilar, R., M. Quesada, L. Ashworth, Y. Herrerias-Diego, and J. Lobo. 2008. Genetic consequences of habitat fragmentation in plant populations: Susceptible signals in plant traits and methodological approaches. *Molecular Ecology* 17: 5177–5188.

Anderson, R. S. 1990a. Modern pollen rain within and adjacent to two giant sequoia *(Sequoiadendron giganteum)* groves, Yosemite and Sequoia National Parks, California. *Canadian Journal of Forest Research* 20: 1289–1305.

Anderson, R. S. 1990b. Holocene forest development and paleoclimates within the central Sierra Nevada, California. *Journal of Ecology* 78: 470–489.

Anderson, R. S. 1994. Paleohistory of a Giant Sequoia grove: The record from Log Meadow, Sequoia National Park. *In* Proceedings of the symposium on Giant Sequoias: Their place in the ecosystem and society, General Technical Report PSW-151, U.S. Department of Agriculture, Forest Service Pacific Southwest Research Station, Albany, California, USA.

Anderson, R. S., and S. J. Smith. 1994. Paleoclimatic interpretations of meadow sediment and pollen stratigraphies from California. *Geology* 22: 723–726.

Andrew, R. L., K. L. Ostevik, D. P. Ebert, and L. H. Rieseberg. 2012. Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology* 21: 2078–2091.

Arenas, M., N. Ray, M. Currat, and L. Excoffier. 2012. Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution* 29: 207–218.

Axelrod, D. I. 1964. The Miocene Trapper Creek flora of southern Idaho. University of California Press, Berkeley, California, USA.

Bacles, C. F. E., A. Lowe, and R. Ennos. 2006. Effective dispersal across a fragmented landscape. *Science* 311: 628.

Beerli, P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22: 341–345.

Beerli, P. 2010. Violation of assumptions, or are your migration estimates wrong when the populations split in the recent past? Website http://popgen.sc.fsu.edu/Migrate/Blog /Entries/2010/8/15_Violation_of_assumptions%2C_or_are_your_migration_estimates_wro ng_when_the_populations_split_in_ the_recent_past.html [accessed 01 August 2019].

Beerli, P. 2012. MIGRATE, version 3.2.1 documentation. Website http://popgen. sc.fsu.edu/migratedoc.pdf [accessed 01 July 2018].

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migra- tion matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences, USA* 98: 4563–4568.

Chen, C., E. Durand, F. Forbes, and O. François. 2007. Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Notes* 7: 747–756.

Chiucchi, J. E., and H. L. Gibbs. 2010. Similarity of contemporary and historical gene flow among highly-fragmented populations of an endangered rattlesnake. *Molecular Ecology* 19: 5345–5348.

Colabella, F., L. A. Gallo, A. C. Moreno, and P. Marchelli. 2014. Extensive pollen flow in a natural fragmented population of patagonian cypress *Austrocedrus chilensis*. *Tree Genetics and Genomes* 10: 1519–1529.

Cole, K. L. 1983. Late Pleistocene vegetation of Kings Canyon, Sierra Nevada, California. *Quaternary Research* 19: 117–129.

Davis, O. K., and M. J. Moratto. 1988. Evidence for a warm dry early Holocene in the western Sierra Nevada of California: Pollen and plant macrofossil analysis of Dinkey and Exchequer Meadows. *Madroño* 35: 132–149.

DeSilva, R., and R. S. Dodd. 2014. Development and characterization of microsatellite markers for giant sequoia, *Sequoiadendron giganteum* (Cupressaceae). *Conservation Genetic Resources* 6: 173–174.

DeSilva, R., and R. S. Dodd. 2019. Data from: Fragmented and isolated: limited gene flow coupled with weak isolation by environment in the paleoendemic giant sequoia (*Sequoiadendron giganteum*). *Dryad Digital Repository*. https:// doi.org/10.6078/D10Q3N.

Dodd, R. S., and R. DeSilva. 2016. Long-term demographic decline and late glacial divergence in a Californian paleoendemic: *Sequoiadendron giganteum* (giant sequoia). *Ecology and Evolution* 6: 3342–3355.

Dubreuil, M., M. Riba, S. C. González-Martínez, G. G. Vendramin, F. Sebastiani, and M. Mayol. 2010. Genetic effects of chronic habitat fragmentation revisited: Strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability. *American Journal of Botany* 97: 303–310.

Durand, E., C. Chen, and O. François 2009a. TESS version 2.3, reference manual. Website http://membrestimc.imag.fr/Olivier.Francois/manual.pdf [accessed 01 May 2018].

Durand, E., F. Jay, O. E. Gaggiotti, and O. François. 2009b. Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution* 26: 1963–1973.

Dyer, R. J. 2014. An introduction to the gstudio package. Website http://dyerlab.github.io /gstudio [accessed 01 March 2019].

Dyer, R. J., and J. D. Nason. 2004. Population graphs: The graph theoretic shape of genetic structure. *Molecular Ecology* 13: 1713–1727.

Dyer, R. J., J. D. Nason, and R. C. Garrick. 2010. Landscape modelling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* 19: 3746–3759.

Eckert, C. G., K. E. Samis, and S. C. Lougheed. 2008. Genetic variation across species' geographical ranges: The central–marginal hypothesis and beyond. *Molecular Ecology* 17: 1170–1188.

Ellstrand, N. C., and D. R. Elam. 1993. Population genetic consequences of small population size: Implications for plant conservation. *Annual Review of Ecology and Systematics* 24: 217–242.

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

Faubet, P., R. S. Waples, and O. E. Gaggiotti. 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology* 16: 1149–1166.

Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions* 13: 252–264.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.

Flint, L. E., A. L. Flint, J. H. Thorne, and R. Boynton. 2013. Fine-scale hydrologic modeling for regional landscape applications: The California Basin Characterization Model development and performance. *Ecological Processes* 2: 25.

François, O., and E. Durand. 2010. Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources* 10: 773–784.
Freedman, A. H., H. A. Thomassen, W. Buermann, and T. B. Smith. 2010. Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology* 19: 3773–3788.
Geue, J. C., C. I. Vágási, M. Schweizer, P. L. Pap, and H. A. Thomassen. 2016. Environmental selection is a main driver of divergence in house sparrows (*Passer domesticus*) in Romania and Bulgaria. *Ecology and Evolution* 6: 7954–7964.
Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices, version 2.9.3. Website http://www2.unil.ch/popgen/softwares/ fstat.htm [accessed 01 May 2018].
Hardy, O. J., and X. Vekemans. 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.
Hardy, O. J., N. Charbonnel, H. Freville, and M. Heuertz. 2003. Microsatellite allele size: A simple test to assess their significance on genetic differentiation. *Genetics* 163: 1467–1482.
Hartesveldt, R. J., H. T. Harvey, H. S. Shellhammer, and R. E. Stecker. 1975. The giant sequoia of the Sierra Nevada. U.S. Department of the Interior, National Park Service, Washington District of Columbia, USA.
Harvey, H. T. 1985. Evolution and history of Giant Sequoia. *In* C. P. Weatherspoon, Y. R. Iwamoto, and D. D. Piirto [Tech. Coord.], Proceedings of the workshop on the management of Giant Sequoia, General Technical Report PSW- GTR-95. U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, California, USA.
Harvey, H. T., H. S. Shellhammer, and R. E. Stecker. 1980. Giant sequoia ecology: Fire and reproduction. Scientific Monograph Series 12. U.S. Department of the Interior, National Park Service. Washington, District of Columbia, USA.
Holderegger, R., and H. H. Wagner. 2008. Landscape genetics. *BioScience* 58: 199–207.
Koehler, P. A., and R. S. Anderson. 1994. The paleoecology and stratigraphy of Nichols Meadow, Sierra National Forest, California, USA. *Palaeogeography, Palaeoclimatology, Palaeoecology* 112: 1–17.
Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. CLUMPAK: A program for identifying clustering modes and packag- ing population structure inferences across K. *Molecular Ecology Resources* 15: 1179–1191.
Kremer, A., O. Ronce, J. J. Robledo-Arnuncio, F. Guillaume, G. Bohrer, R. Nathan, J. R. Bridle, et al. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* 15: 378–392.
Lee, C., and T. Mitchell-Olds. 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* 20: 4631–4642.
Loveless, M. D., and J. L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics* 15: 65–95.
Manion, G., M. Lisk, S. Ferrier, D. Neito-Lugilde, and M. C. Fitzpatrick. 2015. GDM: Functions for Generalized Dissimilarity Modeling. Website https:// cran.r-project.org/web/packages/ gdm/gdm.pdf [accessed 01 February 2018].
Meirmans, P. G. 2014. Nonconvergence in Bayesian estimation of migration rates. *Molecular Ecology Resourses* 14: 726–733.

Nathan, R., F. M. Schurr, O. Spiegel, O. Steinitz, A. Trakhtenbrot, and A. Tsoar. 2008. Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution* 23: 638–647.
Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* 62: 316–336.
Nydick, K. R., N. L. Stephenson, A. R. Ambrose, G. P. Asner, W. L. Baxter, A. J. Das, T. Dawson, et al. 2018. Leaf to landscape responses of giant sequoia to hotter drought: An introduction and synthesis for the special section. *Forest Ecology and Management* 419–420: 249–256.
O'Connell, L. M., A. Mosseler, and O. P. Rajora. 2007. Extensive long-distance pollen dispersal in a fragmented landscape maintains genetic diversity in white spruce. *Journal of Heredity* 98: 640–645.
Oosterhout, C. V., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. Micro- checker: Software for identifying and correcting genotyping errors in micro- satellite data. *Molecular Ecology Notes* 4: 535–538.
Ortego, J., E. C. Riordan, P. F. Gugger, and V. L. Sork. 2012. Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Molecular Ecology* 21: 3210–3223.
Pandey, M., and O. P. Rajora. 2012. Genetic diversity and differentiation of core vs. peripheral populations of eastern white cedar, *Thuja occidentalis* (Cupressaceae). *American Journal of Botany* 99: 690–699.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
Purves, D.W. 2009. The demography of range boundaries versus range cores in eastern US tree species. *Proceedings of the Royal Society, B, Biological Sciences* 276: 1477–1484.
R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https:// www.R-project.org/ [accessed 01 January 2018].
Rannala, B. 2007. BayesAss edition 3.0, user's manual. Website http://www. rannala.org/?page_xml:id=245 [accessed 01 April 2018].
Raymond, M., and F. Rousset. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.
Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* 43: 223–225. Rousset, F. 2008. Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.
Rundel, P. 1972. Habitat restriction in Giant Sequoia - Environmental control of grove boundaries. *American Midland Naturalist* 87: 81–99.
Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.
Segelbacher, G., J. Höglund, and I. Storch. 2003. From connectivity to isolation: Genetic consequences of population fragmentation in capercaillie across Europe. *Molecular Ecology* 12: 1773–1780.
Shafer, A. B. A., and J. B. W. Wolf. 2013. Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters* 16: 940–950.
Shellhammer, H., and T. Shellhammer. 2006. Giant Sequoia (*Sequoiadentron giganteum*

[Taxodiaceae]) seedling survival and growth in the first four decades following managed fires. *Madroño* 53: 342–350.
Sillett, S. C., R. Van Pelt, A. L. Carroll, R. D. Kramer, A. R. Ambrose, and D. A. Trask. 2015. How do tree structure and old age affect growth potential of California redwoods? *Ecological Monographs* 85: 181–212.
Slatkin, M. 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16: 393–430.
Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236: 787–792.
Stephenson, N. L. 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. *Journal of Biogeography* 25: 855–870.
Stephenson, N. L. 2000. Estimated ages of some large Giant Sequoias: General Sherman keeps getting younger. *Madroño* 47: 61–67.
Temunović, M., J. Franjić, Z. Satovic, M. Grgurev, N. Frascaria-Lacoste, and J. F. Fernández-Manjarrés. 2012. Environmental heterogeneity explains the ge- netic structure of continental and mediterranean populations of *Fraxinus angustifolia* Vahl. *PLoS ONE* 7: e42764.
Thomassen, H. A., W. Buermann, B. Milá, C. H. Graham, S. E. Cameron, C. J. Schneider, J. P. Pollinger, et al. 2010. Modeling environmentally associated morphological and genetic variation in a rainforest bird, and its application to conservation prioritization. *Evolutionary Applications* 3: 1–16.
Wang, I. J., and K. Summers. 2010. Genetic structure is correlated with pheno- typic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology* 19: 447–458.
Wilson, G. A., and B. Rannala. 2003. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163: 1177–1191.
Wójkiewicz, B., M. Litkowiec, and W. Wachowiak. 2016. Contrasting pat- terns of genetic variation in core and peripheral populations of highly outcrossing and wind pollinated forest tree species. *AoB Plants* 8: plw054.
Wright, S. 1943. Isolation by distance. *Genetics* 28: 114–138.
Xu, B., G. Sun, X. Wang, J. Lu, I. J. Wang, and Z. Wang. 2017. Population genetic structure is shaped by historical, geographic, and environmental factors in the leguminous shrub *Caragana microphylla* on the Inner Mongolia Plateau of China. *BMC Plant Biology* 17: 200.
York, R. A., J. J. Battles, and R. C. Heald. 2003. Edge effects in mixed conifer group selection openings: tree height response to resource gradients. *Forest Ecology and Management* 179: 107–121.
York, R. A., N. L. Stephenson, M. Meyer, S. Hanna, T. Moody, A. C. Caprio, and J. J. Battles. 2013. Appendix 11a—Giant sequoias. *In* Sydoriak, C., J. A. Panek, J. J. Battles, and K. R. Nydick [Eds.], A natural resource condition assessment for Sequoia and Kings Canyon National Parks, Natural Resource Report NPS/SEKI/NRR—2013/665.11a. Published Report 2195490. U.S. Department of Interior, National Park Service, Fort Collins, Colorado, USA.
Young, A., T. Boyle, and T. Brown. 1996. The population genetic consequences of habitat fragmentation for plants. *Trends Ecology and Evolution* 11: 413–418.

TABLES AND FIGURES

TABLE 1. General population information and summary statistics.

Grove Name	Code	GPS location	Sample size	Gene diversity 'H_E'	Allelic richness 'A_R'	Average pairwise F_{ST}	Average cGD
Placer	PLAC	39.06, −120.57	6	0.20 (0.09)	1.73 (0.30)	0.39 (0.04)	–
North Calaveras	NCAL	38.28, −120.30	41	0.56 (0.07)	3.32 (0.36)	0.14 (0.02)	12.50 (1.02)
South Calaveras	SCAL	38.24, −120.25	46	0.61 (0.06)	3.74 (0.43)	0.12 (0.02)	13.35 (1.07)
Tuolumne	TUOL	37.77, −119.81	8	0.41 (0.06)	2.38 (0.24)	0.24 (0.02)	13.79 (1.07)
Merced	MERC	37.75, −119.84	12	0.54 (0.06)	2.89 (0.27)	0.17 (0.02)	15.25 (0.85)
Mariposa	MPSA	37.51, −119.60	23	0.59 (0.05)	3.29 (0.33)	0.14 (0.02)	12.49 (0.84)
Nelder	NELD	37.43, −119.59	48	0.64 (0.04)	3.29 (0.39)	0.11 (0.02)	11.34 (1.05)
McKinley	MKLY	37.03, −119.11	26	0.55 (0.07)	3.28 (0.48)	0.15 (0.02)	15.56 (1.15)
Grant	GRNT	36.75, −118.97	42	0.60 (0.07)	3.78 (0.55)	0.10 (0.02)	11.63 (1.08)
Redwood Mountain	RMNT	36.60, −118.92	51	0.59 (0.08)	3.90 (0.35)	0.08 (0.02)	10.66 (0.94)
Lost	LOST	36.65, −118.83	26	0.53 (0.06)	3.20 (0.49)	0.15 (0.02)	11.99 (0.97)
Giant Forest	GFOR	36.57, −118.76	62	0.64 (0.05)	4.08 (0.44)	0.08 (0.02)	11.09 (1.14)
Atwell	ATWL	36.47, −118.67	23	0.59 (0.06)	3.72 (0.51)	0.10 (0.02)	12.39 (1.07)
McIntyre	MCTR	36.13, −118.58	25	0.61 (0.06)	3.94 (0.50)	0.09 (0.02)	11.81 (1.27)
Freeman Creek	FMAN	36.14, −118.52	26	0.61 (0.06)	4.01 (0.44)	0.09 (0.02)	12.41 (1.26)
Long Meadow	LMDW	35.96, −118.60	29	0.56 (0.07)	3.63 (0.30)	0.11 (0.02)	11.70 (1.13)
Cunningham	CNHM	35.92, −118.57	22	0.54 (0.06)	3.22 (0.33)	0.13 (0.02)	13.14 (1.21)
Packsaddle	PKSD	35.93, −118.59	28	0.51 (0.06)	3.10 (0.39)	0.12 (0.02)	10.39 (1.08)
Deer Creek	DCRK	35.88, −118.61	24	0.50 (0.07)	3.00 (0.53)	0.15 (0.02)	11.28 (0.95)

TABLE 2. Distance between grove perimeters for paired groves and the observed level of admixture based on STRUCTURE (Pritchard et al., 2000) between pairs.

Paired groves	Approximate intergrove distance	Proportion of admixed individuals
Northern group		
NCAL – SCAL	5.3 km	0.07
TUOL – MERC	3.0 km	0.05
MPSA – NELD	6.4 km	0.05
Southern group		
GRNT – RMNT	4.4 km	0.70
GFOR – ATWL	9.8 km	0.89
MCTR – FMAN	2.0 km	1.00
LMDW – CNHM	2.4 km	0.12
PKSD – DCRK	5.2 km	0.04

TABLE 3. Results of generalized dissimilarity modeling showing percent of genetic dissimilarity explained by model using various combinations of geographic and climate predictors. Rows 2–5 indicate models fit using current climate data and rows 6–9 represent mid-Holocene climate.

	Range-wide	Model *P*-value	Southern populations	Model *P*-value
Current Climate				
Full model	22.4	0.001	12.9	0.15
Environmental variables included in full model	Geographic distance, PWQ, TS, CWD	–	Geographic distance, CWD, TS, PS,	–
Geographic distance only	12.6	< 0.001	2.9	0.29
Environmental distance only	1.8	0.003	11.9	0.14
Mid-Holocene Climate				
Full model	20.9	0.002	8.7	0.20
Environmental variables included in full model	Geographic distance, TS, PDQ, PWQ	–	Geographic distance, PDQ, TS	–
Geographic distance only	12.6	< 0.001	2.9	0.29
Environmental distance only	16.08	0.02	6.1	0.17

Notes: PWQ = Precipitation of Warmest Quarter; TS = Temperature Seasonality; CWD = Climate Water Deficit; PDQ = Precipitation of Driest Quarter; PS = Precipitation Seasonality.

Figure 1: *Sequoiadendron giganteum* distribution. A map of California, USA with the distribution of *Sequoiadendron giganteum* shown in black. Each sampled population is labeled with a location code (Table 1), and the northern and southern groups are outlined.

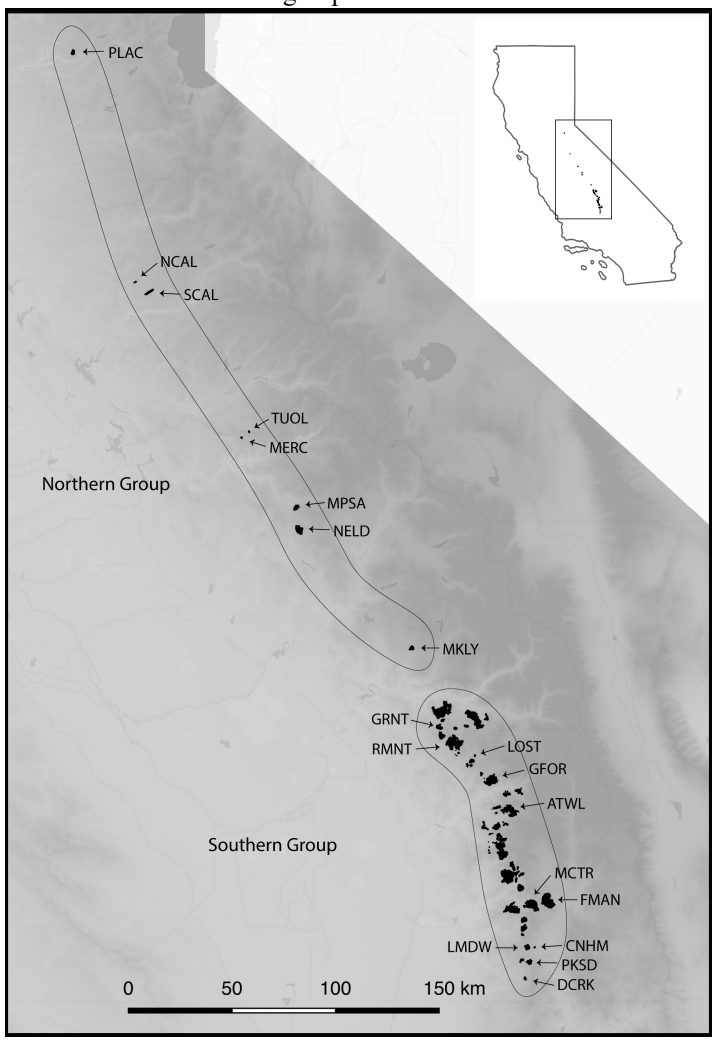


Figure 2: Results of population structure analyses from STRUCTURE (Pritchard et al., 2000) and TESS (Chen et al., 2007; Durand et al., 2009b). (A) STRUCTURE clusters for all populations sampled: Vertical bars represent a sampled individual, color-coded for assigned cluster at K = 5. Populations are arranged from north to south. (B) TESS clusters for all populations sampled: Vertical bars represent a sampled individual, color-coded for assigned cluster at K = 9. Populations are arranged from north to south.

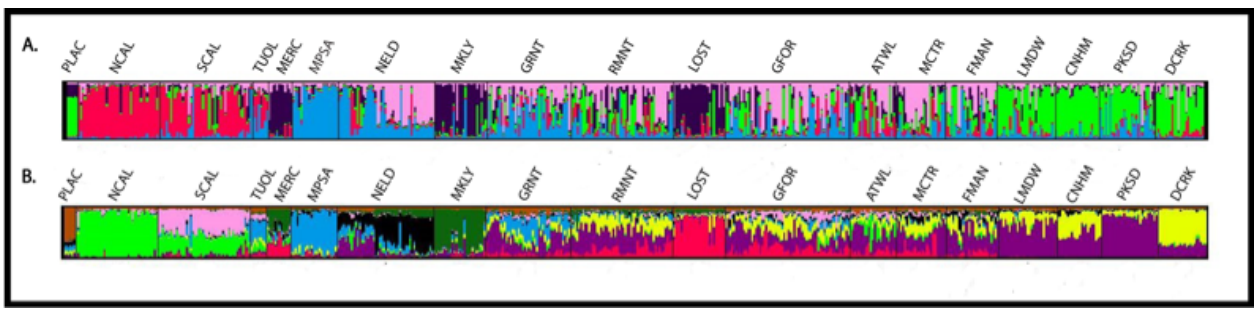
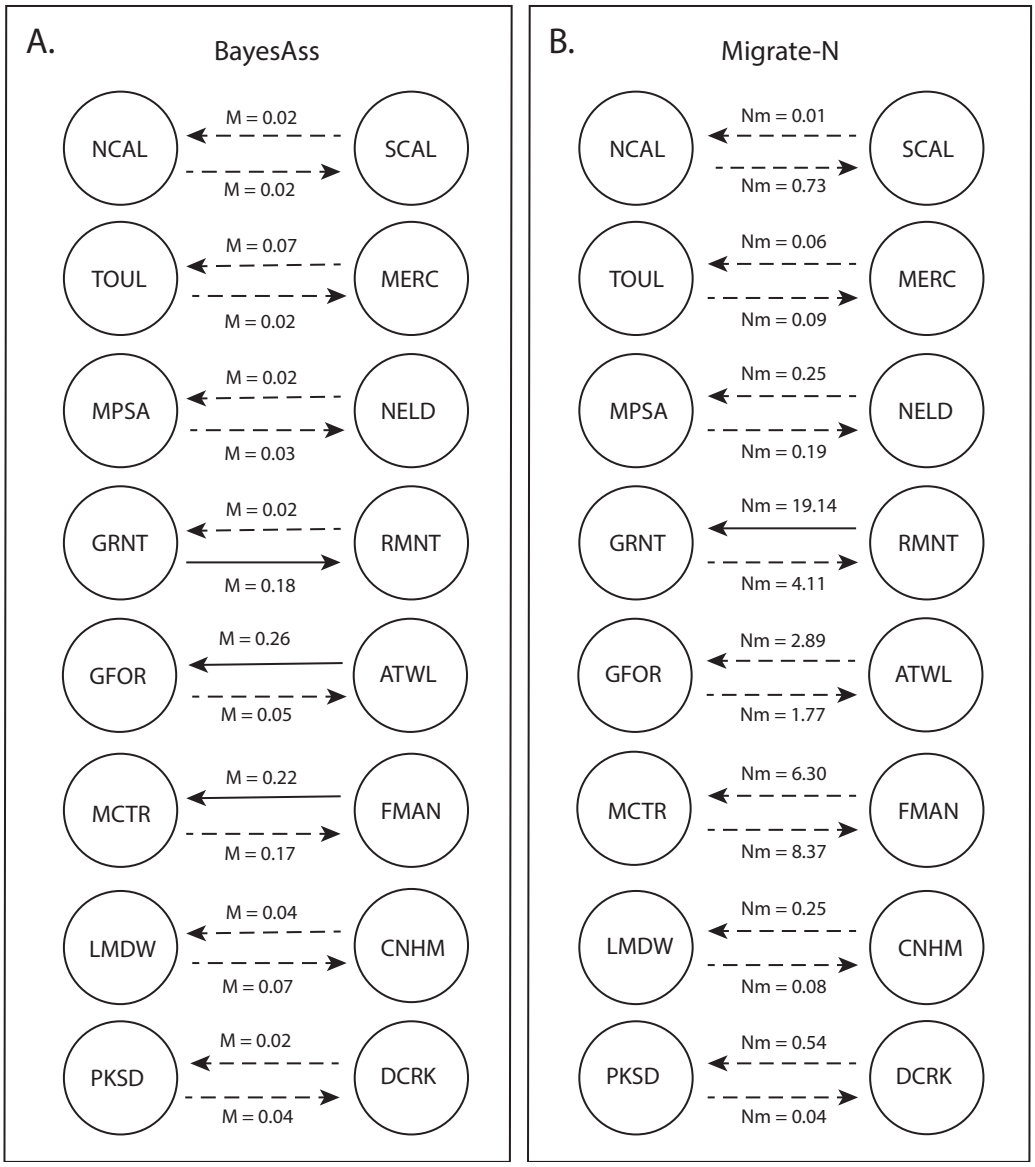


Figure 3: Recent and historic gene flow across paired populations. Migration across selected population pairs. (A) Recent migration estimated as the fraction of individuals of migrant ancestry (M), (B) Historic migration rates estimated as the number of migrants per generation (Nm). Solid lines indicate significant values and dotted arrows are non-significant. Population pairs are arranged from north to south, where the first three pairs are in the northern group of populations, and the remaining pairs are in the southern group of populations



APPENDIX

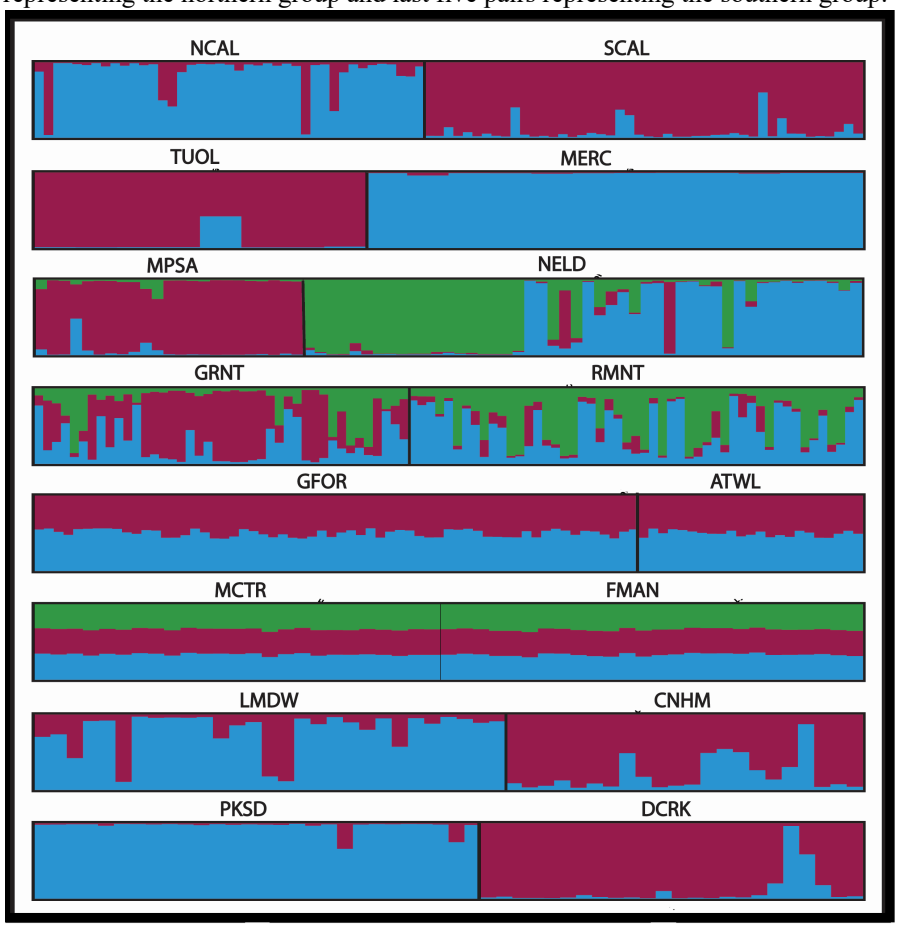
Appendix 1.1: Environmental variables included in GDM model selection

Current Conditions	**Mid-Holocene Conditions**
Bio 2: Mean Diurnal Range	Bio 2: Mean Diurnal Range
Bio 3: Isothermality	Bio 3: Isothermality
Bio 4: Temperature Seasonality	Bio 4: Temperature Seasonality
Bio 6: Min Temperature of the Coldest Month	Bio 6: Min Temperature of the Coldest Month
Bio 10: Mean Temperature of the Warmest Quarter	Bio 10: Mean Temperature of the Warmest Quarter
Bio 15: Precipitation Seasonality	Bio 14: Precipitation of Driest Month
Bio 18: Precipitation of Warmest Quarter	Bio 15: Precipitation Seasonality
Climate Water Deficit	Bio 17: Precipitation of Driest Quarter
	Bio 18: Precipitation of Warmest Quarter

Appendix 1.2: The full set of climate data used in the GDM analyses

	Longitude	Latitude	bio2	bio3	bio4	bio6	bio10	bio14	bio15	bio17	bio18	cwd
NCAL	-120.303	38.280	136	44	6255	-27	183	6	80	30	42	568.79
SCAL	-120.250	38.244	138	44	6247	-29	182	6	79	32	43	557.92
TUOL	-119.808	37.770	142	45	6108	-44	167	6	78	31	40	487.88
MERC	-119.840	37.749	142	45	6145	-37	175	5	78	29	37	564.47
MPSA	-119.601	37.509	137	45	5831	-48	155	6	78	28	39	558.3
NELD	-119.591	37.429	142	46	5987	-34	174	4	80	24	33	767.12
MKLY	-119.108	37.029	124	44	5643	-42	151	5	79	22	38	634.56
GRNT	-118.972	36.747	118	42	5849	-35	161	4	81	19	36	625.43
RMNT	-118.916	36.704	127	42	6053	-32	175	4	81	17	32	536.71
LOST	-118.828	36.651	119	42	5843	-42	155	5	80	21	37	467.18
GFOR	-118.760	36.567	123	42	5975	-39	163	5	80	20	35	581.12
ATWL	-118.675	36.467	120	42	5872	-48	150	6	77	25	39	651.59
MCTR	-118.576	36.132	129	43	6061	-34	173	4	78	20	32	670.34
FMAN	-118.519	36.144	130	42	6136	-36	174	5	78	21	33	709.77
LMDW	-118.598	35.964	127	43	5929	-33	168	4	80	20	32	711.46
CNHM	-118.569	35.917	123	43	5834	-39	157	5	78	22	36	653.06
PKSD	-118.592	35.926	125	43	5846	-35	163	5	79	22	35	704.53
DCRK	-118.611	35.876	139	44	6069	-19	192	2	81	16	25	724.52

Appendix 1.3: Observed admixture among paired populations. STRUCTURE plots for paired populations: Vertical bars represent a sampled individual, color-coded for assigned cluster at best K, with a black line indicating the separation of each population. Paired populations are arranged from North to South, with the first three pairs representing the northern group and last five pairs representing the southern group.



Appendix 1.4: Gene Diversity (H_E), Allelic Richness (A_R), Inbreeding Coefficient (F_{IS}), and Estimated Null Alleles by locus for the entire sample

Locus ▵	Gene diversity (He)	Allelic Richness (Ar)	Inbreeding Coeficient (Fis)	Estimated Null Alleles*
31267	0.82	16.9	0.32	0.05
31670	0.71	13.0	0.59	0.22
39473	0.92	24.0	0.14	0.00
30133	0.71	9.0	0.18	0.00
40527	0.58	6.0	0.08	0.00
29596	0.75	14.9	0.09	0.00
30510	0.30	6.0	0.02	0.00
36493	0.72	16.9	0.25	0.01
7365	0.63	6.0	0.13	0.00
17786	0.67	4.0	0.19	0.00

Appendix 1.5: Migration coefficient estimates from MIGRATE-n and BAYESASS

Direction of Gene Flow	BayesAss: (M) Fraction of individuals of migrant ancestry (95% credible interval)	Migrate-n: (M) Mutation scaled Migration Mean (95% confidence interval)	Theta (Migrate)	Historical migration rate $m_h=M\mu$ *($\mu=5x10^{-4}$)*	Number of migrants per generation (Nm) historic ($\theta\times M/4$)
NCAL -> SCAL	M= 0.02 (0.00-0.04)	3.26 (0-28)	NCAL 0.02	0.0016	0.73
SCAL -> NCAL	M=0.02 (0.00 – 0.05)	12.43 (0-38)	SCAL 0.24	0.0062	0.01
TUOL -> MERC	M=0.02 (0.00 – 0.07)	5.58 (0-22.7)	TUOL 0.06	0.0028	0.09
MERC -> TUOL	M=0.07 (0.00 – 0.15)	3.55 (0-20)	MERC 0.06	0.0018	0.06
MPSA -> NELD	M=0.03 (0.00 – 0.08)	5.01 (0-21.3)	MPSA 0.13	0.0013	0.19
NELD -> MPSA	M=0.02 (0.00 – 0.05)	5.69 (0-22.6)	NELD 0.09	0.0014	0.25
GRNT -> RMNT	M=0.18 (0.05 – 0.31)	6.26 (0-37)	GRNT 2.49	0.0031	4.11
RMNT -> GRNT	M=0.02 (0.00 – 0.06)	30.81 (3-57)	RMNT 2.63	0.0154	19.14
GFOR -> ATWL	M=0.05 (0.00 – 0.12)	15.00 (0-41)	GFOR 1.56	0.0075	1.77
ATWL -> GFOR	M=0.26 (0.10 – 0.41)	7.43 (0-33)	ATWL 0.47	0.0031	2.89
MCTR -> FMAN	M=0.17 (0.00 – 0.37)	22.85 (0-56)	MCTR 1.45	0.0114	8.37
FMAN -> MCTR	M=0.22 (0.05 – 0.38)	17.43 (0-50.7)	FMAN 1.47	0.0087	6.30
LMDW -> CNHM	M=0.07 (0.00 – 0.17)	6.42 (0-14)	LMWD 0.03	0.0032	0.08
CNHM -> LMDW	M=0.04 (0.00 – 0.11)	9.90 (0-18)	CNHM 0.15	0.0050	0.25
PKSD -> DCRK	M=0.04 (0.00 – 0.09)	8.95 (0-34)	PKSD 0.24	0.0045	0.04
DCRK -> PKSD	M=0.02 (0.00 – 0.06)	6.61 (0-32)	DCRK 0.03	0.0033	0.54

Appendix 1.6: Pairwise F_{ST} values for all pairwise population combinations.

	PLAC	NCAL	SCAL	TUOL	MERC	MPSA	NELD	MKLY	GRNT	RMNT	LOST	GFOR	ATWL	MCTR	FMAN	LMDW	CNHM	PKSD	DCRK
PLAC	0.00	0.40	0.39	0.58	0.43	0.39	0.33	0.42	0.36	0.35	0.44	0.32	0.38	0.34	0.35	0.36	0.38	0.41	0.46
NCAL	0.40	0.00	0.09	0.18	0.19	0.17	0.13	0.15	0.09	0.11	0.15	0.09	0.08	0.11	0.10	0.13	0.16	0.13	0.13
SCAL	0.39	0.09	0.00	0.18	0.16	0.15	0.09	0.12	0.09	0.06	0.11	0.06	0.07	0.07	0.09	0.10	0.14	0.10	0.10
TUOL	0.58	0.18	0.18	0.00	0.31	0.21	0.19	0.25	0.15	0.20	0.25	0.17	0.18	0.17	0.20	0.23	0.27	0.25	0.26
MERC	0.43	0.19	0.16	0.31	0.00	0.16	0.14	0.13	0.15	0.12	0.11	0.10	0.14	0.11	0.13	0.17	0.17	0.22	0.20
MPSA	0.39	0.17	0.15	0.21	0.16	0.00	0.09	0.17	0.07	0.08	0.16	0.08	0.11	0.11	0.08	0.12	0.14	0.14	0.18
NELD	0.33	0.13	0.09	0.19	0.14	0.09	0.00	0.11	0.05	0.05	0.14	0.06	0.08	0.05	0.04	0.09	0.08	0.07	0.14
MKLY	0.42	0.15	0.12	0.25	0.13	0.17	0.11	0.00	0.11	0.07	0.15	0.09	0.12	0.09	0.09	0.11	0.15	0.14	0.15
GRNT	0.36	0.09	0.09	0.15	0.15	0.07	0.05	0.11	0.00	0.04	0.12	0.03	0.06	0.05	0.05	0.07	0.10	0.06	0.12
RMNT	0.35	0.11	0.06	0.20	0.12	0.08	0.05	0.07	0.04	0.00	0.09	0.02	0.04	0.03	0.02	0.04	0.06	0.04	0.07
LOST	0.44	0.15	0.11	0.25	0.11	0.16	0.14	0.15	0.12	0.09	0.00	0.09	0.10	0.09	0.12	0.15	0.17	0.16	0.17
GFOR	0.32	0.09	0.06	0.17	0.10	0.08	0.06	0.09	0.03	0.02	0.09	0.00	0.04	0.03	0.02	0.04	0.07	0.06	0.08
ATWL	0.38	0.08	0.07	0.18	0.14	0.11	0.08	0.12	0.06	0.04	0.10	0.04	0.00	0.06	0.04	0.07	0.07	0.04	0.11
MCTR	0.34	0.11	0.07	0.17	0.11	0.11	0.05	0.09	0.05	0.03	0.09	0.03	0.06	0.00	0.04	0.06	0.09	0.09	0.10
FMAN	0.35	0.10	0.09	0.20	0.13	0.08	0.04	0.09	0.05	0.02	0.12	0.02	0.04	0.04	0.00	0.02	0.06	0.06	0.08
LMDW	0.36	0.13	0.10	0.23	0.17	0.12	0.09	0.11	0.07	0.04	0.15	0.04	0.07	0.06	0.02	0.00	0.08	0.06	0.09
CNHM	0.38	0.16	0.14	0.27	0.17	0.14	0.08	0.15	0.10	0.06	0.17	0.07	0.07	0.09	0.06	0.08	0.00	0.08	0.11
PKSD	0.41	0.13	0.10	0.25	0.22	0.14	0.07	0.14	0.06	0.04	0.16	0.06	0.04	0.09	0.06	0.06	0.08	0.00	0.12
DCRK	0.46	0.13	0.10	0.26	0.20	0.18	0.14	0.15	0.12	0.07	0.17	0.08	0.11	0.10	0.08	0.09	0.11	0.12	0.00

TRANSITION TO CHAPTER TWO

Considering the fragmented nature of the giant sequoia range and the prevalence of small populations, the first chapter of this dissertation focused on the broad factors that affect connectivity among spatially separated populations and the extant patterns of genetic diversity at neutral molecular markers. Our research estimating rates of gene flow among spatially proximal groves indicated minimal gene exchange between adjacent groves (3.0 - 6.5 km apart) in the northern range of giant sequoia, which is consistent with limited dispersal in the species. In contrast, southern populations appeared to maintain genetic connectivity at similar spatial scales. Because the biological explanations for this dichotomy were speculative, we decided to follow up this work (Chapter Two) with a fine-scale study on dispersal dynamics within groves that could add clarity to the potential for long-distance dispersal in the species. Moreover, by using progeny arrays at a local scale, we are able to assess pollen dispersal characteristics, the quality (diversity) of the pollen pool and rates of inbreeding.

An overarching goal of this dissertation is to understand patterns of extant genetic diversity across giant sequoia range. The initial assessment presented in chapter one, indicates low levels of genetic diversity, with the smallest and most isolated populations typically being the most genetically depauperate. For sessile organisms such as forest trees, how far propagules disperse away from the parent tree has many consequences for the distribution of genetic diversity at fine-spatial scales. Thus, to further understand patterns of diversity within groves and how diversity might change over time, spatial genetic structure within groves is examined in Chapter Two.

Overall, Chapter Two addressed the fine scale processes of dispersal at the scale of individual groves and builds upon the knowledge gained in Chapter One by adding clarity to aspects of gene flow and genetic diversity.

Chapter 2

Patterns of fine-scale spatial genetic structure and pollen dispersal in giant sequoia (*Sequoiadendron giganteum*)

Rainbow DeSilva, Richard S. Dodd

ABSTRACT

Patterns of dispersal shape the distribution and temporal development of genetic diversity both within and among populations. In an era of unprecedented environmental change, maintenance of extant genetic diversity is crucial to population persistence. We investigate patterns of pollen dispersal and spatial genetic structure within populations of giant sequoia *(Sequoiadendron giganteum*), using both progeny arrays and leaf collections. Our results indicate that giant sequoia is predominantly outcrossing but exhibits moderate levels of bi-parental inbreeding (0.155). The diversity of the pollen pool is low with an average of 7.5 pollen donors per mother tree. As revealed by the Sp-statistic, we find significant genetic structure in ten of twelve populations examined, which indicates clustering of related individuals at fine spatial scales. Estimates of pollen and gene dispersal indicate predominantly local dispersal, with the majority of pollen dispersal < 253 m, and some populations showing fat-tailed dispersal curves, suggesting potential for long-distance dispersal. The results presented here represent the first detailed examination of the reproductive ecology of giant sequoia, which will provide necessary background information for conservation of genetic resources in this species.

INTRODUCTION

Dispersal is a key ecological process that influences the evolution of genetic diversity both within and among populations (Nathan et al. 2008, Kremer et al. 2012, Ellstrand, 2014, Robledo-Arnuncio et al. 2014). For sessile organisms such as forest trees, how far propagules disperse away from the parent tree has many consequences for the distribution of genetic diversity at fine-spatial scales (Vekemans and Hardy 2004). Moreover, characteristics of the dispersal kernel play a large role in determining the extent of long-distance dispersal (gene flow) and thus modulate large scale patterns of genetic diversity and structure across a species range (Slatkin 1987, Cain et al. 2000, Riesenberg and Burke 2001). Since an adequate pool of genetic diversity, on which selection can act, is critical for the success of populations under changing environments, understanding dispersal dynamics is important for successful management of species.

At the scale of a population, dispersal dynamics shape the clustering of related individuals on the landscape. Fine-scale spatial genetic structure (FSGS), can be defined as the non-random arrangement of genotypes on a landscape (Heywood 1991, Vekemans and Hardy 2004). In plants, FSGS is caused by the interplay of many evolutionary forces, but of key importance is dispersal limitation, which creates patterns of isolation by distance between parents and offspring (Wright 1943, Vekemans and Hardy 2004). When related genotypes aggregate together in space, this can increase rates of bi-parental inbreeding, eventually leading to a reduction in genetic diversity (i.e. loss of heterozygosity, Ellstrand and Elam 1993, Krakowski et al 2003). Moreover, populations with strong genetic structure can be more vulnerable to genetic diversity loss from stochastic events, such as disturbance or genetic drift (Willi and Määttänen 2011).

In plants, gene dispersal occurs at two distinct phases, pollen and subsequent seed dispersal. For forest trees, pollen and seed differ in abundance and dispersal dynamics. Empirical studies have often shown that for wind-dispersed species, pollen can have a particularly large dispersal potential (Dow and Ashley 1996, Bacles et al. 2005, Robledo-Arnuncio and Gil 2005, deLucas et al. 2008, Chybicki and Dzialuk 2014). Moreover, the shape of the dispersal kernel (kurtosis) is an important indicator of the potential for long-distance dispersal (LDD) as fat-tailed curves have increased likelihood of LDD (Nathan et al. 2008). In an era of ever-increasing habitat fragmentation, long distance pollen dispersal can be especially important for maintaining connectivity among habitat patches (O'Connell et al. 2007, Colabella et al. 2014) and mitigating risks of inbreeding. Thus, determining the characteristics of the pollen dispersal curve is critical to understanding how genetic diversity will change within populations and across fragmented landscapes.

Giant sequoia, *Sequoiadendron giganteum,* is a paleoendemic long-lived tree species occupying ~70 groves, scattered across mid-elevations in the Sierra Nevada mountains of California. Its range stretches approximately 400 km from Placer county in the north to Tulare county in the south. The entire range of giant sequoia is fragmented. However, populations tend to become smaller and more disjunct in the northern ~2/3rds of the latitudinal range (Figure 1). Giant sequoia is wind pollinated, and mature trees produce an abundance of pollen and seed cones each year (Hartesveldt et al. 1975). Mature seed cones remain closed and attached to the tree for many years, and thus provide a large aerial seed bank (Hartesveldt et al. 1975). Successful regeneration for this species can be episodic and often occurs after a fire, which triggers seed dispersal and creates canopy gaps (Rundel 1972; Harvey et al. 1980; York et al 2003, Shellhammer and Shellhammer 2006, York et al. 2009). This species' reliance on fire, highlights a potential for environmental mismatch as forest management policies and shifting climate alter fire regimes across California (Goss et al. 2020).

Given the cultural and ecological value of this species, it is surprising how little is known regarding the fine-scale patterns of genetic diversity and dispersal within the extant groves. To date, no studies have addressed the extent of FSGS within giant sequoia populations, and only a single study (Anderson 1990), which investigated pollen rain, demonstrated patterns consistent with short-distance pollen dispersal. Yet, the scope of this work was extremely limited as it only covered two extremely small populations of giant sequoia < 0.21 km^2 (Anderson 1990). Our

previous work indicated minimal gene exchange between adjacent groves (3.0 - 6.5 km apart) in the northern range of giant sequoia (DeSilva and Dodd 2020), consistent with limited dispersal in the species. However, no studies have used progeny arrays at a local scale to understand the dynamics of pollen dispersal distance, the quality (diversity) of the pollen pool and rates of inbreeding in determining FSGS. Here we attempt to fill this gap in giant sequoia reproductive ecology by: 1) determining the degree of fine-scale spatial genetic structure within twelve populations of giant sequoia; 2) estimating mating parameters (i.e. number of pollen donors, rates of outcrossing and bi-parental inbreeding) using progeny collected from five populations; and 3) investigating characteristics of pollen dispersal including mean dispersal distance and kurtosis of the dispersal curve.

METHODS

Sampling and data preparation

Fine-scale spatial genetic structure (FSGS)

FSGS within groves was estimated using leaf tissue collected from twelve giant sequoia groves. Leaf collections, DNA extraction, and DNA preparation are described in detail in DeSilva and Dodd (2020). In brief, for this study we utilized leaf material collected in twelve populations with high sampling density. All individuals were genotyped at ten microsatellite loci described in DeSilva and Dodd (2014).

Mating parameters and pollen dispersal

We obtained fallen cones that were in clusters close to a putative parent tree, within five *S. giganteum* groves. Due to the height of reproductive branches in mature *S. giganteum* trees, often >10 m, cones were collected from the ground beneath potential maternal trees in 7-18 locations per population (Table 1). Geographic coordinates of all collection sites were taken at the time of collection. From each sampling location, seeds were extracted from multiple cones and subjected to 30 days of moist cold-stratification and 30 days of dry cold-stratification at 1-2° Celsius. Seeds were then germinated on filter-paper lined Petri dishes. Subsequently, DNA was isolated from the seedling radicle using the CTAB method (Doyle and Doyle 1990). A total of 1070 seeds were genotyped using ten microsatellite markers outlined in DeSilva and Dodd (2014).

Census density of mature trees was estimated from population surveys for all southern groves, using the density of all trees >75cm d.b.h. (York et al. 2013). For the northern groves, (NELD, CALN, CALS), census densities of mature trees were estimated from Willard (2000).

Assigning maternity

Because each cone collection locality potentially included progeny from more than a single mother tree, maternal families were identified using the likelihood method implemented in ML-RELATE (Kalinowski et al. 2006). Subsequently, the largest maternal families were retained from each cone collection locality. We then removed individuals with potential null alleles or genotyping errors (those that were incompatible with a single mother tree), while maximizing the number of individuals retaining the most common alleles within the maternal family. On occasion, this process resulted in more than one potential maternal family. In this case, all potential families were retained for subsequent analyses and average statistics for the population are reported. We then utilized MLTR to determine the most likely maternal genotype

for each maternal family (Ritland 1996, Ritland 2002), as potential maternal genotypes were not sampled. After filtering, DNA from 629 seeds were used in pollen dispersal analyses and assessing mating parameters.

Data Analysis

Mating parameters

Using the seed genotype data, we first estimated the mating parameters (selfing, bi-parental inbreeding and outcrossing) from single locus and multilocus estimates of outcrossing using MLTR (Ritland 2002). Multilocus outcrossing rate (t_m) provides an estimate of the true level of selfing, whereas single locus outcrossing rate (t_s) accounts for all inbreeding (selfing and mating with close relatives). Thus, we estimated bi-parental inbreeding as (t_m - t_s) following Ritland (2002).

Then, we used TwoGener (Smouse et al. 2001) to estimate the effective number of pollen donors per mother tree (N_{ep}) as $N_{ep}=1/2\Phi_{FT}$ (Austerlitz and Smouse 2001), where ΦFT is the differentiation between the pollen clouds sampled by pairs of maternal trees within a population.

Spatial Genetic Structure and Spatial genetic structure and gene dispersal

Based on leaf genotypes of established trees we estimate the extent of spatial genetic structure within populations using the Sp statistic (Vekemens and Hardy 2004), calculated as $-b_F / [1 - F_{(1)}]$, where b_F represents the linear decay in the pairwise kinship coefficient (F_{ij}) for all pairs of individuals with the logarithm of geographic distance and $F_{(1)}$ is the mean kinship coefficient between individuals within the first distance class. The Sp statistic is useful in our case because it provides a meaningful way to compare spatial genetic structure across populations despite variation in population sizes and sampling scheme (Vekemans and Hardy 2004). For this calculation, we used the pairwise kinship coefficients (F_{ij}) of Loiselle et al. (1995), as this measure of relatedness shows less statistical bias than many others (Hardy and Vekemans 1999; Vekemans and Hardy 2004). We assigned the number of distance classes based on the geographic area of each grove as follows: number of distance classes 0-125 ha = 2, 125 - 275 ha = 3, 275-500 ha = 4, 500-850 ha = 5, and >900 ha = 6 (Appendix 2.1). We determined these grove area cut-offs, after completing exploratory analyses, to inform the spatial resolution that allowed for a large number of pairwise comparisons per distance class and an adequate spatial resolution within the first distance class. Significance of 'b_F' was obtained using 20,000 permutations of sampling locations within populations. All calculations were completed using SPAGeDi 1.5 software (Hardy and Vekemans 2002).

Also, we utilized SPAGeDi 1.5 (Hardy and Vekemans 2002) to obtain an indirect estimate of evolutionary-scale gene dispersal, an effective pollen and seed average, from patterns of FSGS. Under equilibrium isolation-by-distance conditions, the scale of effective gene dispersal can be estimated from Wright's neighborhood size equation ($N_b \equiv 4\pi D_e\sigma^2$), where D_e is the effective density and σ^2 is half the mean-squared parent-offspring distance (i.e. gene dispersal; Hardy and Vekemans 2002, Fenster 2003). For the estimation of gene dispersal (σ^2) we set the effective density to 1/2 the adult census density. We recognize this represents a high estimate of effective density as evidence suggests that for adult populations the ratio of Ne/N often ranges between 0.1 and 0.5 (Frankham 1995).

Pollen dispersal parameters

We estimated the characteristics of pollen dispersal using both the TwoGener and KinDist approaches, as implemented in POLDISP 1.0 (Robledo-Arnuncio et al., 2007). Here, maternal trees are considered to serve as pollen traps, and their progeny represent a sample of the available pollen pool. The TwoGener method uses the differential structure of the pollen sampled by each mother tree across the landscape (Smouse et al. 2001), whereas the related KinDist method uses the relationship between correlated paternity and the pollen dispersal kernel (Robledo-Arnuncio et al. 2006). Both methods can have some drawbacks. To fit the pollen dispersal curve, TwoGener requires an independent estimate of effective density which can be hard to obtain for some species. Yet, accuracy can be increased with a reliable external estimate of effective density (Robledo-Arnuncio et al. 2006). On the other hand, KinDist requires a threshold distance for unrelated pollen pools, which can be difficult to determine. For both KinDist and TwoGener, we applied the two parameter (a and b) exponential power distribution to fit the pollen distribution curve as recommended by Austerlitz et al. (2004). This distribution allows for the leptokurtic pattern of pollen dispersal that is commonly observed in wind-pollinated trees (Austerlitz et al. 2004, de-Lucas et al 2008, Marchelli et al. 2012, Chybicki and Diazuk 2014, Moracho et al. 2016, Burczyk et al. 2019). We estimated the scale (a) and shape (b) parameters of the dispersal kernel and calculated mean pollen dispersal distances (d) according to Austerlitz et al. (2004). The shape parameter (b) provides an indication of the potential for LDD, through determining the degree of kurtosis (i.e. how fat-tailed the dispersal curve is). For TwoGener calculations we set the Ne/N ratio as 0.1, and 0.5. Because effective density often ranges between 0.1 and 0.5 of census density (Frankham 1995), we chose these values as they potentially represent a high and low estimate of the effective density within giant sequoia populations.

RESULTS

Mating parameters

Average multi-locus outcrossing rate (t_m) across five groves was 91 percent, with the highest outcrossing observed for GFOR (0.95) and the lowest in GRNT (0.84) (Table 1), which indicates that low levels (9 percent) of selfing is also occurring. The average rate of bi-parental inbreeding was moderate (t_m-t_s) = 0.155. Bi-parental inbreeding was lowest in LOST, (t_m-t_s) = 0.112 (0.037), and highest in GFOR, (t_m-t_s) = 0.189 (0.107) (Table 1). The number of effective pollen donors per mother tree (N_{ep}), ranged from 4.7 - 8.5, with an average of 7.5 (Table 1). We observed the fewest pollen donors in GRNT (4.7) and the most in CALN (8.5) (Table 1).

Spatial genetic structure

Significant genetic structure was found in ten of the twelve populations assessed (Table 2). For these ten populations, Sp ranged from 0.0235 to 0.0444 and was lowest in RMNT and highest in CALS. When estimating Wright's gene dispersal from SPAGeDi, which represents an effective pollen and seed average, we found mean gene dispersal distances between 120.4 - 374.0 m (Table 2). Our iterative procedure to estimate gene dispersal failed to converge for three populations, likely due to insufficient sampling density within these populations.

Dispersal dynamics

The TwoGener method indicated average pollen dispersal distance (d) ranging from 64.6 – 252.1 m with a trend of increased dispersal distance when the effective density was reduced

from 50 to 10 percent census density (Table 3). Evidence for fat-tailed dispersal kernels ($b < 1.0$) was consistently found for GRNT, LOST, and GFOR but was absent in CALS and detected in CALN only when effective density was set to 50 percent census density (Table 3). When using the KinDist approach, the correlated paternity among maternal families did not show a significant decrease with distance for CALN, CALS, or LOST populations (Pearson's product-moment correlation p-value = 0.21, 0.47, and 0.69, respectively). Thus, further analysis using the KinDist method is not recommended for these populations (Robledo-Arnuncio et al. 2007). Estimated dispersal parameters for GRNT and GFOR showed average pollen dispersal distance ranging from 572.8 - 2133.8 m respectively and leptokurtic dispersal kernels (i.e. $b < 1.0$). Since our goal here is to uncover general pollen dispersal characteristics across the giant sequoia range, we focus the discussion on the TwoGener results as they provide evidence for more general patterns in the species.

DISCUSSION

Our prior work on genetic structure within the range of giant sequoia indicated relatively strong divergence among groves in the northern range (north of the Kings River watershed) despite the close proximity of some of the groves e.g. the two Calaveras groves (CALN and CALS), Tuolumne and Merced, Nelder and Mariposa (DeSilva and Dodd 2020). We proposed low rates of seed and pollen dispersal to account for divergence over such short distances. In the southern range, groves appeared to be more admixed, which raised questions as to whether dispersal distances might be greater in the south, or whether the increased admixture was a result of a lack of lineage sorting in this more contiguous range.

In the present work, we have addressed gene dispersal by estimating the distance of total gene flow and of pollen dispersal inferred from progeny arrays sampled within groves from the northern and southern range of the species. In addition, we assessed mating system parameters in the same groves to determine whether their size, or isolation contributed to any differences in levels of inbreeding through selfing and bi-parental inbreeding. Overall, we found that although giant sequoia is generally an outbreeding species, it exhibits a low degree of selfing and moderate rates of bi-parental inbreeding. The scales of pollen and gene dispersal were consistent across groves, suggesting dispersal distances for the species are predominantly short, which indicates that most pollination is localized to within groves. Dispersal curves showed evidence of fat-tails, which could indicate potential for some long-distance dispersal events.

Mating system and pollen pool diversity

Consistent with our observations for giant sequoia, outcrossing rates in many wind-pollinated trees typically range from 90-100% (Burczyk et al. 1996, Wasieliwska et al. 2005, Burczyk et al. 1996, Sork et al. 2002, O'Connell et al. 2006, Bower and Aitken 2007, de-Lucas et al. 2008). In slight contrast, we found higher levels of bi-parental inbreeding in giant sequoia than for many other conifer species (Ledig et al. 2005, O'Connell et al. 2006, Mantovani et al. 2006, de-Lucas et al. 2008). Links between inbreeding and population size have been established for some tree species, which can exhibit higher inbreeding (bi-parental or selfing) in smaller populations (Rajora et al 2002, O'Connell et al. 2006, Chybicki and Dzialuk 2014). From the five groves for which we obtained estimates of inbreeding, we found no evidence for a relationship between grove size and level of inbreeding, despite having data from two groves of less than 25 ha (CALN and LOST). Interestingly, the average bi-parental inbreeding estimated

from progeny arrays exceeded the degree of inbreeding (F_{IS}-statistic) estimated from adult trees in these five groves using the same microsatellite loci (DeSilva, unpublished data). We believe this difference could be attributable to a post germination selective filter acting against inbred progeny in natural populations.

We observe low pollen pool diversity in comparison to many other conifers, such as *Larix occidentalis* ($N_{ep} \cong 35$), *Picea glauca* ($N_{ep} = 62–143$), *Pinus pinaster* ($N_{ep} = 21–56$), and *Austrocedrus chilensis* (average $N_{ep} = 13.9$) (El-Kassaby and Jaquish 1996, O'Connell et al. 2006, de-Lucas et al. 2008, Colabella et al. 2014), which supports the idea that pollen dispersal in these giant sequoia groves is spatially restricted. However, it should be noted that our N_{ep} estimates are similar to those reported for the wind pollinated angiosperm trees *Quercus lobata* and *Nothofagus nervosa* (Sork et al. 2002, Marchelli et al. 2012) and our estimate of N_{ep} is likely somewhat reduced due to bi-parental inbreeding (Austerlitz and Smouse 2001b).

Evidence for limited dispersal

The high levels of fine-scale spatial genetic structure (FSGS) across the giant sequoia range suggest limited dispersal capacity in this species. Although forest tree species commonly show significant genetic structure, the Sp-statistic values are typically lower than found here (de-Lucas et al. 2009, Aleksic et al. 2017, Eliades et al 2018, Mosca et al. 2018). For instance, in 25 populations of four conifer species, Sp values varied from 0.0018 to 0.0035 (Mosca et al. 2018). Although peripheral populations (due to small size or lower density) often demonstrate stronger FSGS as compared to core populations (Gapare and Aiken 2005, de-Lucas et al. 2009), for giant sequoia we find no relationship between the degree of FSGS and population size, isolation, or density. In contrast, the degree of FSGS was fairly consistent across populations (average Sp = 0.0349, SD 0.0076), which points to underlying biological constraints on gene dispersal.

Our results indicate that the majority of pollination in giant sequoia occurs over short distances < 253 m, which is typical for many tree species including *Pinus pinaster*, *Quercus lobata*, *Nothofagus nervosa*, and *Larix decidua* (de-Lucas et al. 2008, Pluess et al. 2009, Marchelli et al. 2012, Burczyk et al. 2019). Moreover, gene dispersal (a measure of the effective pollen and seed average) also appears to occur over short distances <370 m and we observe no significant differences across the range due to grove area or density. Although we are unable to make direct comparisons between pollen and gene dispersal for GRNT and LOST, estimated dispersal parameters show general correspondence. High congruence in dispersal distance despite differences among groves indicate biological controls on dispersal capacity. Interestingly, this finding also suggests that seed and pollen may disperse at similar scales in giant sequoia, which is in contrast to many wind-pollinated trees where pollen travels farther than seed (Latta et al.1998, Schuster and Mitton 2000, Heuertz et al. 2003, Chybicki and Burczyk 2010, Sork et al. 2015, Browne et al, 2018). More limited pollen dispersal in giant sequoia may be a result of less buoyant pollen, as giant sequoia pollen lack the air-filled sacs typical of the members of the *Pinaceae* (e.g. *Pinus spp. Abies spp. Picea spp.*) (Ting 1965, Bortenschlager 1990, Schwendemann et al. 2007). It is important to note that estimates of dispersal using the Twogener method are highly tied to effective density (Smouse and Sork 2004). Here, we assume that effective density is between 10-50 percent of census density. Yet, we recognize that no formal studies have examined the effective density of any giant sequoia populations. Our results suggest a general tendency for increased average dispersal when effective density is reduced.

Thus, our estimates of average dispersal are likely an underestimate if effective density is lower than 10 percent of census density.

Evidence for long-distance pollen dispersal

In some conifer species, wind dispersed pollen can account for extremely long-distance dispersal (O'connell et al. 2007, Robledo-Arnuncio 2011, Chybicki and Dzialuk 2014, Colabella et al. 2014). Here, we find leptokurtic pollen dispersal for four out of five populations examined, highlighting the potential for long-distance dispersal in giant sequoia. This finding is consistent with our previous research that indicates connectivity in the southern range of giant sequoia across populations between 2-10km apart (DeSilva and Dodd 2020). Yet, it is at odds with the apparent lack of gene flow between adjacent giant sequoia populations in the north (DeSilva and Dodd 2020). In the southern section of giant sequoia range, multiple populations are often within 10km of each other, potentially allowing for more regional LDD opportunities than in the north where very few populations exist. Our progeny data from the northern section of giant sequoia range has important limitations. The indirect TwoGener method works best when few offspring are sampled from many mother trees that are sampled both near and far (Smouse and Sork 2004). Our data has relatively few maternal trees representing CALN and CALS populations. Thus, a more extensive sampling of mother trees, or a detailed parentage analysis, is likely needed to determine the existence of gene-flow across these populations. Empirical studies consistently suggest that 'fat-tailed' pollen dispersal curves are typical for many wind-dispersed tree species (Robledo-Arnuncio and Gill 2005, de-Lucas et al 2008, Marchelli et al. 2012, Chybicki and Diazuk 2014, Colabella et al. 2014, Moracho et al. 2016, Burczyk et al. 2019). Although we find evidence for leptokurtic dispersal kernels, due to the lack of air sacs in giant sequoia pollen, its capacity for LDD may be more limited than many conifers with which it shares habitat. Fat-tailed dispersal curves allow more opportunities for gene flow among fragmented populations (Nathan et al. 2008). Gene flow can be a crucial factor facilitating population persistence as it can replenish diversity lost through genetic drift and introduce new variation into populations, which can be a source of adaptive potential (Young et al. 1996, Kremer et al. 2012).

Evidence for demic structure in giant sequoia groves

The reproductive dynamics of giant sequoia suggest that demes within populations are an important factor influencing changes in genetic diversity over time. Dispersal limitation can result in demic structure within non-selfing species, as mating among close relatives becomes more important. Our data indicate relatively high rates of bi-parental inbreeding that were more or less consistent across groves and highest in GRNT and GFOR groves. We observe predominantly local dispersal, which, coupled with low diversity in the effective pollen pool sampled by mother trees, indicates the importance of reproductive groupings within populations. Moreover, the coupling of high levels of FSGS with predominantly local dispersal can beget further clustering of related individuals over time. Demic patterns within populations are important because small reproductive neighborhoods can reduce the effective population size and increase the risk of diversity erosion due to genetic drift (Whitlock and Barton 1997, Willi and Määttänen 2011). Moreover, reduced genetic diversity can eventually lead to inbreeding depression, which is a potential threat to population survival (Frankham 1995b).

CONCLUSIONS

We present evidence for predominantly local pollen dispersal in giant sequoia and potential for a limited degree of long-distance dispersal. The spatial restriction of the majority of pollen dispersal has likely influenced the observed strong spatial genetic structure and created demic structure within giant sequoia groves. We warn of potential genetic diversity loss in many giant sequoia populations that may be effectively operating as smaller reproductive units. Thus, we suggest that small, isolated, and highly structured giant sequoia populations are at highest risk for erosion of genetic diversity. These populations include, but are not limited to, CALN, NELD, GRNT, and DCRK. We end on the hopeful note that evidence of fat-tailed dispersal suggests some of this diversity loss may be mitigated by long distance gene-flow. We suggest that assisting in the movement of genetic resources by planting seedlings from both local and non-local sources in these high-risk populations can be an effective means to enhance genetic diversity. As climate changes, extant genetic diversity will be important for the long-term persistence of giant sequoia populations.

ACKNOWLEDGMENTS

We sincerely thank Glenn Lunak and Sierra Pacifica Industries for providing seed collections for our study. We also thank Zoe Caron for laboratory assistance with seed germination and DNA extraction.

REFERENCES

Anderson R.S. 1990. Modern pollen rain within and adjacent to two giant sequoia (Sequoiadendron giganteum) groves, Yosemite and Sequoia national parks, California. *Canadian Journal of Forest Research*, 20 (9): 1289-1305.

Austerlitz F, Smouse P.E. 2001. Two-generation analysis of pollen flow across a landscape; II. Relation between Φ(ft), pollen dispersal and interfemale distance. Genetics 157: 851–857

Austerlitz F, Smouse P.E. 2001b. Two-generation analysis of pollen flow across a landscape. III. Impact of within-population structure. Genetical Research, 78: 271–278

Austerlitz F, Smouse P.E. 2002. Two-generation analysis of pollen flow across a landscape IV. Estimating the dispersal parameter. Genetics 161: 355–363

Bacles C.F.E, Burczyk J, Lowe A.J, Ennos R.A. 2005. Historical and contemporary mating patterns in remnant populations of the forest tree *Fraxinus excelsior* L. *Evolution* 59 :979–90

Bortenschlager, S. 1990. Aspects of pollen morphology in the cupressaceae. Grana. 29(2): 129-138

Bower A.D., Aitken S.N. (2007) Mating system and inbreeding depression in whitebark pine (Pinus albicaulis Engelm.). Tree Genet Genome 3: 379–388

Burczyk J., Adams W.T., Shimizu J.Y. 1996. Mating patterns and pollen dispersal in a natural knobcone pine (Pinus attenuata Lemmon.) stand. Heredity 77: 251–260.

Burczyk, J., Sandurska, E., Lewandowski, A. 2019. Patterns of Effective Pollen Dispersal in Larch: Linking Levels of Background Pollination with Pollen Dispersal Kernels. Forests 10: 1139. doi:10.3390/f10121139

Cain M.L., Brook G. Milligan, and A.E. Strand. 2000. Long-distance seed dispersal in plant populations. American Journal of Botany. 87(9): 1217–1227.

Chybicki IJ, Burczyk J. 2010. Realized gene flow within mixed stands of Quercus robur L. and Q. petraea (Matt.) L. revealed at the stage of naturally established seedling. Molecular Ecology. 19(10): 2137–2151. doi:10. 1111/j.1365-294X.2010.04632.x

Chybicki and Dzialuk 2014. Bayesian approach reveals confounding effects of population size and seasonality on outcrossing rates in a fragmented subalpine conifer. *Tree Genetics & Genomes* 10: 1723–1737. https://doi.org/10.1007/s11295-014-0792-3

Colabella, F., L.A. Gallo, A. C. Moreno, and P. Marchelli. 2014. Extensive pollen flow in a natural fragmented population of patagonian cypress austrocedrus chilensis. *Tree Genetics and Genomes* 10(6): 1519-1529.

de-Lucas, A., Robledo-Arnuncio, J., Hidalgo, E. and Gonzalez-Martinez, S.C. 2008. Mating system and pollen gene flow in Mediterranean maritime pine. *Heredity* 100, 390–399.

de-Lucas, A.I., Gonzalez-Martinez, S.C., Vendramin, G.G., Hidalgo, E. and Heuertz, M. 2009. Spatial genetic structure in continuous and fragmented populations of *Pinus pinaster* Aiton. Molecular Ecology, 18: 4564-4576. doi:10.1111/j.1365-294X.2009.04372.x

DeSilva, R., and R. S. Dodd. 2014. Development and characterization of microsatellite markers for giant sequoia, Sequoiadendron giganteum (Cupressaceae). *Conservation Genetic Resources* 6: 173–174.

DeSilva, R., and Dodd, R. S. (2020). Fragmented and isolated: Limited gene flow coupled with weak isolation by environment in the paleoen- demic giant sequoia (*Sequoiadendron giganteum*). *American Journal of Botany*, *107*(1), 45–55

Dow, B. D., and M. V. Ashley, 1996 Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa.* Molecular Ecology. 5**:** 615–627.
Doyle, J.J.; Doyle J.L. Isolation of plant DNA from fresh tissue. Focus, v.12, p.13-15, 1990.
El-Kassaby, Y. A.; Jaquish, B. 1996. Population density and mating pattern in western larch. Journal of Heredity 87(6): 438-443.
Eliades, N.H., Fady, B., Gailing, O. Leinemann, L. and R. Finkeldey. 2018. Significant patterns of fine-scale spatial genetic structure in a narrow endemic wind-dispersed tree species, *Cedrus brevifolia* Henry. *Tree Genetics & Genomes* 14: 15. https://doi.org/10.1007/s11295-018-1231-7
Ellstrand, N. C. and Elam, D. R. 1993. Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. Annual Review of Ecology and Systematics. 24: 217-242. doi: 10.1146/annurev.es.24.110193.001245
Ellstrand, N. C. 2014. Is gene flow the most important evolutionary force in plants? *American Journal of Botany* 101: 737–753.
Ennos RA. 1994. Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 72:250–59
Fenster CB, Vekemans X, Hardy OJ (2003) Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (leguminosae). *Evolution*, **57**, 995–1007.
Frankham R,. 1995. Effective population-size adult-population size ratios in wildlife—a review. Genet Res 66(02):95–107. https://doi.org/10. 1017/S0016672300034455
Frankham R,. 1995b. Inbreeding and extinction; a threshold effect. Conservation Biology. 9: 792-799.
Gapare, W.J. and Aitken, S.N. 2005. Strong spatial genetic structure in peripheral but not core populations of Sitka spruce [*Picea sitchensis* (Bong.) Carr.]. Molecular Ecology, 14: 2659-2667. doi:10.1111/j.1365-294X.2005.02633.x
Goss M., D.L Swain, J.T Abatzoglou, A. Sarhadi, C. A. Kolden, A.P. Williams and N.S. Diffenbaugh. 2020. Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *Environmental Research Letters.* 15: 094016.
Hardy OJ, and X. Vekemans. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation and population genetics models. Heredity 83(2):145–154. https://doi.org/10. 1046/j.1365-2540.19 99.00558.x
Hardy, O. J. and X. Vekemans. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618-620.
Hartesveldt, R. J., H. T. Harvey, H. S. Shellhammer, and R. E. Stecker. 1975. The giant sequoia of the Sierra Nevada. U. S. Department of the Interior, National Park Service, Washington DC, USA.
Harvey, T. H., Shellhammer, H. S. and Stecker, R.E. 1980. Giant Sequoia Ecology: Fire and Reproduction. Scientific Monograph Series No. 12. U.S. Department of the Interior, National Park Service. Washington, D.C.
Heywood J.S. 1991. Spatial analysis of genetic variation in plant populations. Annual Review Ecology Systematics 22(1):335–355. https://doi.org/10. 1146/annurev.es.22.110191.002003
Heuertz M, Vekemans X, Hausman JF, Palada M, Hardy OJ. 2003. Estimating seed versus pollen dispersal from spatial genetic structure in the common ash. Mol Ecol 12(9): 2483–2495. doi:10.1046/j.135-294X. 2003.01923.x

Kalinowski, S.T., Wagner, A.P. and Taper, M.L. 2006. ml-relate : a computer program for maximum likelihood estimation of relatedness and relationship. Molecular Ecology Notes, 6: 576-579.

Krakowski, J., Aitken, S. & El-Kassaby, Y. 2003. Inbreeding and conservation genetics in whitebark pine. *Conservation Genetics* 4, 581–593. https://doi.org/10.1023/A:1025667700479

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J. R., et al. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* 15: 378-392.

Latta, R.G., Linhart, Y.B., Fleck, D. and Elliot, M. 1998. Direct and indirect estimates of seed versus pollen movement within a population of Ponderosa pine. Evolution, 52: 61-67. doi:10.1111/j.1558-5646.1998.tb05138.x

Ledig, F.T., Hodgskiss, P.D. and Johnson, D.R. 2005. Genic diversity, genetic structure, and mating system of Brewer spruce (Pinaceae), a relict of the Arcto-Tertiary forest. American Journal of Botany. 92: 1975-1986. doi:10.3732/ajb.92.12.1975

Mantovani, A.L. Patrícia C. Morellato, Maurício S. dos Reis, 2006. Internal Genetic Structure and Outcrossing Rate in a Natural Population of *Araucaria angustifolia* (Bert.) O. Kuntze, *Journal of Heredity*, 97 (5): 466–472. https://doi.org/10.1093/jhered/esl031

Mosca, E, Di Pierro, EA, Budde, KB, Neale, DB, González-Martínez, SC. 2018. Environmental effects on fine-scale spatial genetic structure in four Alpine keystone forest tree species. *Molecular Ecology*. 27: 647– 658. https://doi.org/10.1111/mec.14469

Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. and Tsoar, A. 2008. Mechanisms of long-distance seed dispersal. *Trends Ecology and Evolution.* **23**: 638– 647.

O'Connell LM, Mosseler A, Rajora OP. 2006. Impacts of forest fragmentation on the mating system and genetic diversity of white spruce (Picea glauca) at the landscape level. Heredity, 97, 418–426. https://doi.org/10.1038/sj.hdy.6800886

O'Connell, L. M., Mosseler, A., and O. P. Rajora. 2007. Extensive Long-Distance Pollen Dispersal in a Fragmented Landscape Maintains Genetic Diversity in White Spruce. *Journal of Heredity* 98(7): 640–645.

Marchelli, P., Smouse, P.E. & Gallo, L.A. 2012. Short-distance pollen dispersal for an outcrossed, wind-pollinated southern beech (*Nothofagus nervosa* (Phil.) Dim. et Mil.). *Tree Genetics & Genomes.* 8: 1123–1134 https://doi.org/10.1007/s11295-012-0500-0

Petit and Hampe. 2006. Some Evolutionary Consequences of Being a Tree.Annual Review of Ecology, Evolution, and Systematics. 37: 187-214

Pluess, Andrea R., Victoria L. Sork, Brian Dolan, Frank W. Davis, Delphine Grivet, Kurt Merg, Jeanette Papp, Peter E. Smouse. 2009. Short distance pollen movement in a wind-pollinated tree, Quercus lobata (Fagaceae). Forest Ecology and Management. 258:(5). 735-744.

Rajora OP, Mosseler A, Major JE (2002) Mating system and reproductive fitness traits of eastern white pine (Pinus strobus) in large, central versus small, isolated, marginal populations. Canadian Journal of Botany 80: 1173-1184.

Ritland, K. 1989. Correlated matings in the partial selfer, *Mimulus guttatus*. *Evolution* 43: 848–859.

Rieseberg LH, Burke JM (2001) The biological reality of species: gene flow, selection, and collective evolution. Taxon 50:47–67

Ritland, K. 1996. MLTR: Multilocus mating system program, version 1.1. Available from the author. University of British Columbia, Vancouver, British Columbia, Canada.

Ritland K. 2002. Extensions of models for the estimation of mating systems using n independent loci. Heredity 88:221–228
Robledo-Arnuncio JJ, Gil L. 2005. Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity* 94:13– 22
Robledo-Arnuncio JJ, Austerlitz F, Smouse P.E. 2006. A new method of estimating the pollen dispersal curve independently of effective density. Genetics 173:1033–1045
Robledo-Arnuncio JJ, Austerlitz F, Smouse P.E. 2007. POLDISP: a software package for indirect estimation of contemporary pollen dispersal. *Molecular Ecology Notes* 7: 763-766.
Robledo-Arnuncio, J.J., 2011. Wind pollination over mesoscale distances: an investigation with Scots pine. New Phytologist. 190: 222–233.
Robledo-Arnuncio, J. J., E. K. Klein, H. C. Muller-Landau, and L. Santamaria. 2014. Space, time and complexity in plant dispersal ecology. *Movement Ecology* 2: 16.
Rundel P. 1972. Habitat Restriction in Giant Sequoia - Environmental Control of Grove Boundaries. American Midland Naturalist 87: 81-99
Schuster, W.S.F. and Mitton, J.B. 2000. Paternity and gene dispersal in limber pine (*Pinus flexilis* James). Heredity, 84: 348-361. doi:10.1046/j.1365-2540.2000.00684.x
Schwendemann, A.B., Wang, G., Mertz, M.L., McWilliams, R.T., Thatcher, S.L. and Osborn, J.M. 2007. Aerodynamics of saccate pollen and its implications for wind pollination. American Journal of Botany. 94: 1371-1381. doi:10.3732/ajb.94.8.1371
Shellhammer, H. S., and T. H. Shellhammer. 2006. Giant sequoia (Sequoiadendron giganteum Taxodiaceae) seedling survival and growth in the first four decades following managed fires. Madrono. 53: 342-350.
Smouse, P. E., R. J. Dyer, R. D. Westfall, and V. L. Sork. 2001. Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution.* 55: 260–271.
Smouse, P. E., and V. L. Sork 2004. Measuring pollen flow in forest trees: an exposition of alternative approaches. Forest Ecology and Management. 197: 21–38.
Sork VL, Dyer RJ, Davis FW, Smouse PE. 2002. Mating patterns in a savanna population of valley oak (*Quercus lobata* Neé). In: *Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in Cali- fornia's Changing Landscape, October 22–25, 2001* (eds Standiford R, McCreary D, Purcell KL), pp. 427–439. Pacific SW Research Station, US Forest Service, USDA, San Diego, CA.
Sork, V., Smouse, P., Grivet, D., and D. Scofield. 2015. Impact of asymmetric male and female gamete dispersal on allelic diversity and spatial genetic structure in valley oak (Quercus lobata Nee). Evolutionary Ecology. 29: 927-945
Slatkin M. 1987. Gene flow and the Geographic Structure of Natural Populations. *Science* 236: 787-792
Ting W.S. 1965. The Saccate Pollen Grains of Pinaceae Mainly of California, Grana, 6:2, 270-289, DOI: 10.1080/00173136509429148
Vekemans X, Hardy OJ. 2004. New insights from fine-scale spatial genetic structure analyses in plants populations. Mol Ecol 13(4):921–935. https://doi.org/10.1046/j.1365-294X.2004.02076.x
Wasieliwska M, LKlemm M, Burczyk J. 2005. Genetic diversity and mating system of Scots pine plus trees. Dendrobiology 53: 57–62.
Whitlock M.C. and N.H. Barton 1997. The Effective Size of a Subdivided Population. Genetics 146: 427-441

Willi, Y. and Määttänen, K. 2011. The relative importance of factors determining genetic drift: mating system, spatial genetic structure, habitat and census size in *Arabidopsis lyrata*. New Phytologist, 189: 1200-1209. doi:10.1111/j.1469-8137.2010.03569.x

Willard, Dwight. 2000. A guide to the Sierra Groves of California. Yosemite Conservancy.

Wright S 1943. Isolation by distance. Genetics 28: 114–138.

York R., J. Battles, and R. Heald. 2003. Edge effects in mixed conifer group selection openings: tree height response to resource gradients. Forest Ecology and Management **179**:107-121

York, R. A., Z. Thomas, and J. Restaino. 2009. Influence of Ash Substrate Proximity on Growth and Survival of Planted Mixed-Conifer Seedlings. Western Journal of Applied Forestry 24:117-123.

York, R. A., N. L. Stephenson, M. Meyer, S. Hanna, T. Moody, A. C. Caprio, and J. J. Battles. 2013. A natural resource condition assessment for Sequoia and Kings Canyon National Parks: Appendix 11a – giant sequoias. Natural Resource Report NPS/SEKI/NRR—2013/665.11a. National Park Service, Fort Collins, Colorado.

TABLES AND FIGURES

Table 1: Seed collection and mating system parameters for five populations.

Population (ha)	Seed collection sites	Average Maternal Family Size	Census Density (m^2)	Multi-locus outcrossing rate t_m (SD)	Sinlge-locus outcrossing rate t_s (SD)	Bi-parental inbreeding t_m-t_s	Nep (TwoGener)
CALN (25)	7	10	0.00052	0.92 (0.11)	0.81 (0.09)	0.117 (0.09)	8.5
CALS (182)	9	11	0.00055	0.94 (0.12)	0.76 (0.04)	0.178 (0.11)	8.4
GRNT (163)	11	14	0.00099	0.84 (0.03)	0.66 (0.03)	0.180 (0.03)	4.7
LOST (21)	8	13	0.00127	0.88 (0.04)	0.77 (0.04)	0.112 (0.04)	7.8
GFOR (935)	18	12	0.00104	0.95 (0.12)	0.76 (0.02)	0.189 (0.11)	7.0
AVERAGE	-	-	-	0.91	0.75	0.155	7.5

Table 2: Sp-statistic and estimated gene dispersal for twelve giant sequoia populations distributed across the range, based on analyses using SPAGeDi.

Population (ha)	Sp-statistic	Sigma*2 (mean gene dispersal distance, m)
CALN (25)	0.0354*	252.6
CALS (182)	0.0444**	374.0
NELD (195)	0.0417**	153.6
GRNT (163)	0.0388**	_
RMNT (1466)	0.0235*	120.4
LOST (21)	0.0339^{NS}	_
GFOR (935)	0.0238**	185.6
ATWL (542)	0.0292^{NS}	_
MCTR (700)	0.0414**	264.2
FMAN (580)	0.0397**	159.0
LMDW (138)	0.0308*	260.2
DCRK (21)	0.0303*	236.0

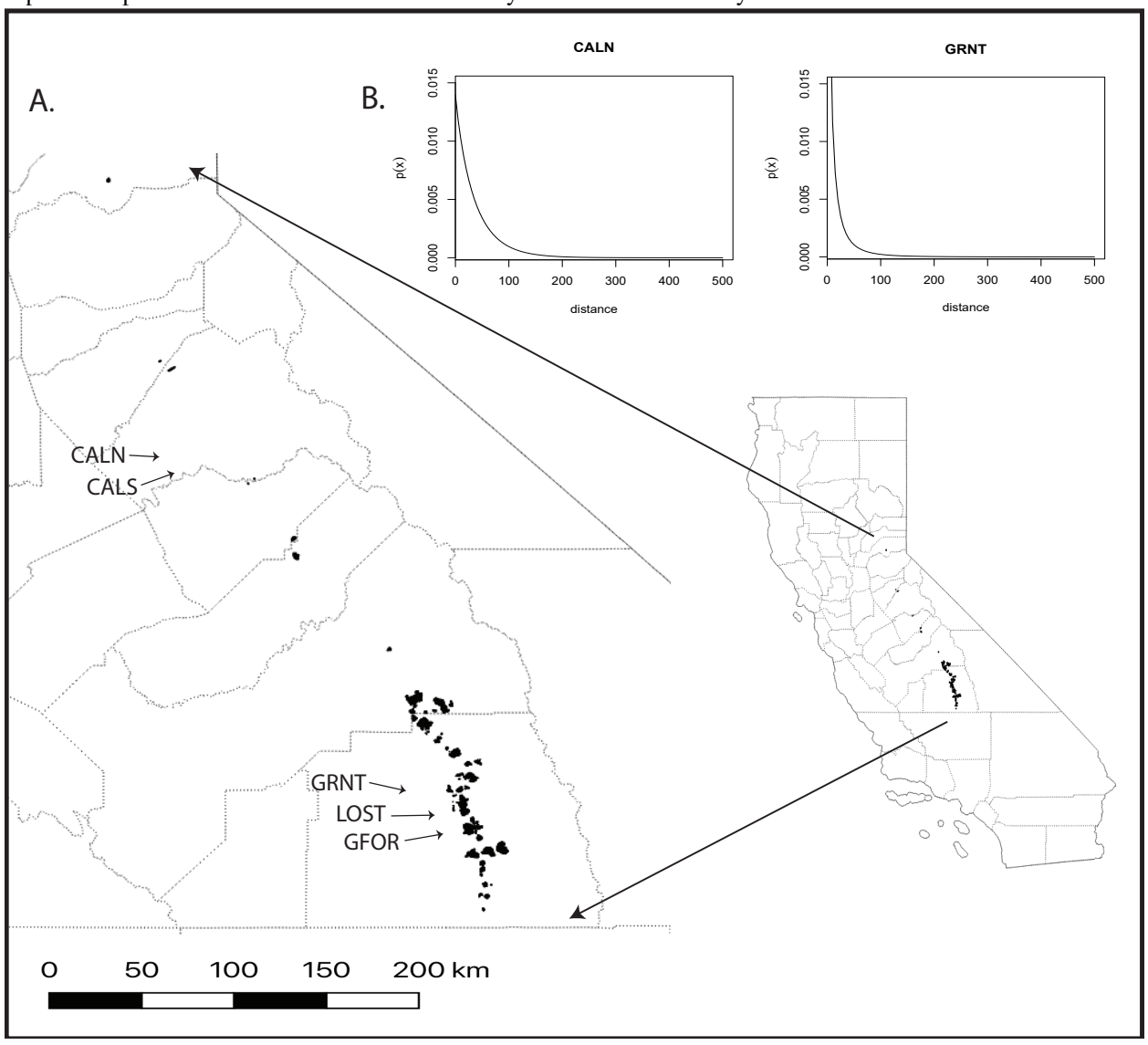
*indicates a significant result (p 0.05), **(p 0.01)

^{NS} = result is not significant

Table 3: Pollen dispersal parameters obtained using the TwoGener approach.

	(Ne/N=0.5)			**(Ne/N=0.1)**		
Population	Scale (a)	Shape (b)	Average distance, d (m)	Scale (a)	Shape (b)	Average distance, d (m)
CALN	34.62	0.93	80.7	215.96	6.84	141.9
CALS	98.54	25.94	64.6	82.85	1.05	152.5
GRNT	0.55	0.37	74.2	4.75	0.47	129.0
LOST	34.62	0.93	80.7	1.05	0.35	213.5
GFOR	0.004	0.22	203.0	0.37	0.31	252.1

Figure 1: A) Range map of giant sequoia. Populations where seeds were collected indicated are noted by a population code. B) Pollen dispersal kernels for CALN (left) and GRNT (right) estimated in TwoGener by fitting an exponential power distribution with effective density set to ½ census density.



APPENDIX

Appendix 2.1: Grove size, number of spatial groups, and distance within the first distance class for populations analyzed with SPAGeDi

Population Code	Grove Size (ha)	Number of spatial groups	Distance range within the first distance class (m)
CALN	25	2	0 – 428
CALS	182	3	0 – 403
NELD	195	3	0 – 558
GRNT	163	3	0 – 293
RMNT	1466	6	0 – 415
LOST	21	2	0 – 138
GFOR	935	6	0 – 767
ATWL	542	5	0 – 236
MCTR	700	5	0 – 419
FMAN	580	5	0 – 456
LMDW	138	3	0 – 347
DCRK	21	2	0 – 269

TRANSITION TO CHAPTER THREE

The first two chapters of this dissertation have provided a comprehensive picture of patterns in extant genetic diversity across giant sequoia range at neutral markers and how they are influenced by dispersal and gene-flow. Chapter Two adds detail to patterns of genetic diversity in the species, showing that within groves diversity is highly structured, meaning related genotypes tend to cluster in space. Thus far, we have addressed how non-selective factors (dispersal and gene flow) shape genetic diversity of at multiple spatial scales. Yet, natural selection is of key importance in shaping patterns of genetic diversity across a species range. Therefore, in Chapter Three we complete our genetic diversity assessment by seeking to better understand adaptively important genetic variation and its distribution across giant sequoia range.

Building on the knowledge regarding gene flow gained from Chapter One, Chapter Two provided evidence that the majority of pollen and gene dispersal occurs over short distances, while also highlighting a non-trivial potential for long distance pollen dispersal. Moreover, in our study of the factors shaping gene-flow (Chapter One) we found some evidence for isolation by environment (IBE), linking genetic divergence at putatively neutral loci to dissimilarity in precipitation and temperature related variables. Although IBE is consistent with local adaptation of populations, many other factors can create or contribute to patterns of IBE. Thus, in Chapter Three we build upon this earlier work by assessing the prevalence of local adaptation to climate among giant sequoia populations.

Adding to the knowledge contained in the earlier chapters of this dissertation, Chapter Three provides a more complete understanding of how giant sequoia may respond to global change. Specifically, in the next chapter we assess the distribution of adaptive variation across giant sequoia range, which is a prerequisite to predicting this species' response to climate change. In addition, we uncover signatures of local adaptation through association of genomic and climatic variation. These associations provide a glimpse into the climatic factors that have a major influence on selection, which in turn inform us about the environmental tolerances of giant sequoia. Lastly, we highlight specific populations that may offer genetic variants that represent 'pre-adaptations' to future conditions.

Chapter 3

Association of genetic and climatic variability in giant sequoia, *Sequoiadendron giganteum,* reveals signatures of local adaptation along moisture-related gradients.

Rainbow DeSilva, Richard S. Dodd

ABSTRACT

Uncovering the genetic basis of local adaptation is a major goal of evolutionary biology and conservation science alike. In an era of climate change, an understanding of how environmental factors shape adaptive diversity is crucial to predicting species response and directing management. Here, we investigate patterns of genomic variation in giant sequoia, an iconic and ecologically important tree species, using 1364 bi-allelic single nucleotide polymorphisms (SNPs). We use an F_{ST} outlier test and two genotype-environment association methods, latent factor mixed models (LFMM) and redundancy analysis (RDA), to detect complex signatures of local adaptation. Results indicate 79 genomic regions of potential adaptive importance, with limited overlap between the detection methods. Of the 58 loci detected by LFMM, 51 showed strong correlations to a precipitation driven composite variable and seven to a temperature-related variable. RDA revealed 24 outlier loci with association to climate variables, all of which showed strongest relationship to summer precipitation. Nine candidate loci were indicated by two methods. After correcting for geographic distance, RDA models using climate predictors accounted for 49% of the explained variance and showed significant correlations between SNPs and climatic factors. Here, we present evidence of local adaptation in giant sequoia along gradients of precipitation and provide a first step towards identifying genomic regions of adaptive significance. The results of this study will provide information to guide management strategies that seek to maximize adaptive potential in the face of climate change.

INTRODUCTION

In an era of unprecedented climate change, the adaptive potential of populations has become an increasingly important topic to conservation biologists, raising questions of landscape partitioning of adaptive variation and management strategies to maintain population viability. Given the rapid rate of climate change, new beneficial mutations are

expected to play a limited role for species with low mutation rates and long generation times. Therefore, adaptive evolution under climate change for many species will depend on standing genetic variation (Aitken et al. 2008; Barrett and Schluter 2008) that may vary across the landscape and include alleles that gain adaptive value as selection pressures change (Olson-Manning et al 2012). Reliance on standing genetic variation is likely to be particularly true for long-lived sedentary species, such as forest trees that are characterized by adaptive constraints that can limit their evolutionary response to rapid environmental change: extended generation times that result in local persistence, increased rates of genetic drift associated with overlapping generations (Rogers and Prügel-Bennett 2000) and limited rates of migration due to long generation times. For these species, understanding the distribution of adaptive diversity in relation to recent, or past climatic gradients is a critical first step in promoting the adaptive potential of populations in hopes of maintaining future viability (Holderegger et al. 2006; Aitken and Whitlock 2013).

The rich history of field research on phenotypic traits in plants (common gardens and reciprocal transplant studies) provides evidence for abundant heritable variation for quantitative traits that is organized along environmental clines (Morgenstern 1996; Savolainen et al. 2007). Until recently, determining the molecular basis of this variation has been less tractable. However, the rapid advances in genome sequencing, including methods that use reduced genomic complexity (e.g. Genotyping-By-Sequencing (GBS), Restriction-site Associated DNA sequencing (RADseq)), has opened the door to more comprehensive assessments of population level diversity and allowed for the detection of regions under selection. Although some instances of strong selection on few genes of major effect have been noted (Akey 2009; Linnen et al. 2009; Sella et al. 2009), many traits of adaptive importance in plants are believed to be polygenic in nature (Holland 2007; Pritchard and Di Rienzo 2010; Le Corre and Kremer 2012; Yeaman et al. 2016). Under selection these traits can exhibit subtle changes in frequency across many loci of small effect. Further, demographic processes can shape genetic diversity in ways that mimic selective gradients, as geographic distance and climatic gradients are often autocorrelated. As a result, imprints of selection within the genome can be difficult to detect (Yeaman 2015), and it is necessary to parcel out the contribution of geographic space in order to successfully identify regions of functional importance (Excoffier et al. 2009; Rellstab et al. 2015).

By coupling genome-wide markers with landscape genomics analyses, many researchers have successfully uncovered patterns of adaptive variation and identified potential genomic regions under selection across a wide variety of species (De Kort et al. 2014; Benestan et al. 2016; Pais et al. 2016; Harrisson et al. 2017; Lind et al. 2017; Dudaniec et al. 2018). F_{ST} outlier tests, that scan for highly differentiated loci as candidates for divergent selection, have proven useful in detecting regions under selection but often cannot detect weak or polygenic selection (Pritchard and Di Rienzo 2010; Narum and Hess 2011; Lotterhos and Whitlock 2015). Genotype-environment association (GEA) tests have demonstrated high power to detect signals of adaptive evolution under varying demographic scenarios (de Villemereuil et al. 2014; Lotterhos and Whitlock 2015; Forester et al. 2018). Univariate association methods that test for single-locus-single-predictor correlation after accounting for population structure are powerful tools to accurately detect even weak signatures of adaptation (Frichot et al. 2013; Gunther and Coop 2013; de Villemereuil et al. 2014; Lotterhos and Whitlock 2015; Rellstab et al. 2015). However, a short-coming of assessing each locus independently is a potential failure to detect

signals of polygenic selection (Forester et al. 2018). Multivariate approaches can fill this gap by assessing the combined effects of multiple loci and predictors (Rellstab et al. 2015; Capblancq et al. 2018; Forester et al. 2018), which is perhaps more reflective of real-life evolutionary pressures. Given the advantages of each method, combining outlier tests with GEA can increase the likelihood of detecting complex patterns of selection (Rellstab et al. 2015).

Determining the presence of adaptively important genetic variation and its distribution across a species range is crucial to predicting species' responses to global climate change and directing biodiversity conservation and management efforts (Sgrò et al. 2011; Funk et al. 2012; Aitken and Whitlock 2013; Alberto et al. 2013; Sork et al. 2013). This has become an urgent challenge in California, where a protracted drought has resulted in massive tree mortality (USDA 2016). The Sierra Nevada of California is a high mountain range that collects precipitation from the Pacific Ocean mostly in the form of winter rain and snowfall. The slow release of water from snowmelt in the spring is an important source of moisture for seedling growth and establishment. Sierra snowpack has declined in recent years (Fyfe et al. 2017) and high-resolution regional climate models suggest that spring snow water equivalent will decline by 73% by the end of the century, with mid-elevations (1500-2500m) experiencing the greatest declines (Sun et al. 2018).

This elevational range includes the extant groves of the iconic, long-lived conifer, giant sequoia (*Sequoiadendon giganteum* [Lindl.] Buchholz) that occur in a highly disjunct range consisting of ~70 groves spanning approximately 400km north-to-south (Figure 1). Currently, most giant sequoia populations are in protected areas as this species is valued both culturally and for ecotourism. However, despite this protected status, a key question is whether populations of giant sequoia will remain viable under changing climate. Our previous work has shown very restricted gene flow (DeSilva and Dodd 2020), suggesting that natural dispersal outside of existing groves will be unlikely. Long generation times, (~305 years; Dodd and DeSilva 2016) will slow the expansion of new variants that may arise through mutation, which underscores the role of standing genetic variation in determining the future viability of giant sequoia populations. In our landscape genetics study of microsatellite variation, we found some evidence for isolation by environment (IBE), linking genetic divergence at putatively neutral loci to dissimilarity in precipitation and temperature related variables (DeSilva and Dodd 2020). Although IBE is consistent with local adaptation, it is dependent on a reduction of gene flow from divergent habitats due to selection against non-adapted immigrants, and therefore, patterns of IBE may not be reflective of local adaptation when gene flow is low or absent (Nosil et al. 2005; Wang and Bradburd 2014), as is likely the case in sections of giant sequoia range (DeSilva and Dodd 2020). A recent common garden study reported provenance variation in growth performance, providing support for the existence of adaptive genetic variation across the species' range (Valness 2016). Yet, to date, no studies have investigated local adaptation in giant sequoia using genomic data. Our ultimate goal is to detect populations that may be genetically responsive to anticipated climate change, so here, we build upon this earlier work by reporting on genomic signatures of selection using a range-wide Genotyping-By-Sequencing dataset. Specifically, we utilize an F_{ST} outlier test and gene-environment association methods (LFMM, and RDA), to find signatures of local adaptation among giant sequoia populations and locate potential genomic regions under selection.

METHODS

DNA extraction, GBS library preparation, and data processing

Foliage was collected from 6-9 trees within each of 18 populations of giant sequoia distributed throughout the range (Figure 1). To reduce the potential of sampling related individuals, we aimed to sample individual trees >40 meters apart. However, this was not possible in some small and highly clustered populations. In this latter case, we attempted to maximize the distance between sampled individuals, with the exception of the PLAC population, where all individuals were sampled. Our goal is to maximize the capture of variation across the range of our study species. Thus, we prioritize increased sampling of populations across the *S. giganteum* range, with the tradeoff of limited sampling within each population. Appropriate permits were obtained for all sampling.

High purity genomic DNA from 143 individuals was isolated from leaf tissue using Plant/Fungi DNA Isolation kits (Norgen Biotek, Thorold, ON, Canada). We constructed three sequencing libraries using a double-digest restriction-enzyme associated genotyping-by-sequencing (GBS) protocol outlined in Peterson et al. (2014). Genomic DNA was digested using SbfI and EcoRI restriction enzymes (New England Biolabs, Ipswich, MA, USA). The resulting product was ligated to barcoded adapters and purified, 46-48 individuals per-library were then pooled and subjected to PCR amplification using Phusion High Fidelity PCR Kit (New England Biolabs) and an automated size selection for fragments between 430-570bp using Pippen Prep. The resulting three libraries were sequenced on an Illumina HiSeq 4000 platform using 150bp pair-end reads. Sequence data were then demultiplexed using the process_radtags module within the STACKS pipeline (Catchen et al. 2013), during which reads with a phred quality score < 10 were removed. Sequences were then aligned to the giant sequoia reference genome v1.0 (Redwood Genome Project 2019), using the software Bowtie-2 and SAMtools (Li et al. 2009; Langmead and Salzberg 2012). Variable sites were called using FreeBayes (Garrison and Marth 2012) and filtered to remove low quality reads, potential sequencing errors, and paralogs. Data filtering steps included removing loci with uneven mapping quality and those with average read depth > 200, requiring a minimum read depth of 5x and a minor allele count > 3, removal of loci with more than 80% missing data, and a thinning step that retains one SNP per DNA fragment to remove potentially linked loci (Appendix 3.1). This filtering protocol resulted in a final data set of 1364 bi-allelic SNPs used for outlier tests and environmental association analyses and to obtain genetic diversity statistics.

Environmental data

To characterize the climatic conditions for each population, we used the spatial centroid of each population to extract and compile twenty-one environmental variables at a spatial resolution of approximately 1km^2. Nineteen climate variables were obtained from the WorldClim database (Fick and Hijmans, 2017), and elevation and Climate Water Deficit (CWD) were obtained from the California Basin Characterization Model (Flint et al. 2013). CWD provides an indication of aridity that is important for Mediterranean climate systems, such as in California (Stephenson, 1998). We conducted a principal component analysis (PCA) on the full environmental data set (twenty-one variables) after standardization, to reduce dimensionality in the climate data. We retained the first two axes (hereafter PC1 and PC2), which together explained 82% of the climate variation (Appendix 3.2, and 3.3). PC1 was driven predominantly by temperature and elevation variables, with a small contribution from annual and winter precipitation, whereas PC2 was determined mostly by precipitation related variables and CWD

with a minor contribution from variables related to temperature seasonality (Appendix 3.2, and 3.3).

Genetic diversity

Genetic diversity and differentiation statistics were calculated using both the 'diveRsity' package in R and GenoDive (Meirmans and Van Tienderen 2004, Keenan et al. 2013). Calculated statistics included observed and unbiased expected heterozygosity (H_o and uHe respectively), the inbreeding coefficient (F_{IS}) and the pairwise fixation index (G'_{ST}). Since the removal of rare alleles (minor allele count filtering) can bias genetic diversity estimation, we also calculated genetic diversity statistics without this filtering step for comparison. To further investigate the partitioning of genetic variation, we used AMOVA with 10000 permutations to estimate F_{ST} across all populations as well as between northern and southern regions which previous evidence suggested were divergent (Dodd and DeSilva 2016; DeSilva and Dodd 2020). For regional diversity comparisons, groves north of GRNT were grouped as northern populations and groves from GRNT to the south as southern populations (Figure 1; Dodd and DeSilva 2020).

Genomic signatures of selection: F_{ST} outliers and gene-environment association tests

To detect F_{ST} outliers that are candidates for selection we utilized the Bayesian likelihood approach implemented in BayeScan v.2.1 (Foll and Gaggiotti 2008). This method scans the genome for highly differentiated SNPs that potentially have been subjected to divergent selection while accounting for neutral genetic structure (Narum and Hess 2011). BayeScan was run using the false discovery rate (FDR) set to 0.05 under the following parameters: 20 pilot runs of 5000 with an additional burn in of 50000 iterations and a subsequent run with 5000 iterations and a thinning interval of 10. The prior odds for the neutral model was increased to 100 (default is 10) as raising this value has been shown to reduce false positives with little effect on false negatives (Lotterhos and Whitlock 2014). Loci with Log_{10} values of the posterior odds > 1.0 were retained, as the program documentation suggests these loci show 'strong' evidence for selection (Foll 2010).

To test for associations between genomic variation and environmental factors, we utilized latent factor mixed models (LFMM; Frichot et al. 2013), as implemented in the LEA package in R (Frichot and Francois 2015a). LFMM is a univariate approach that treats each individual locus as a response variable with climate data (PC1 and PC2 separately) as the explanatory variable, while incorporating neutral structure using latent factors (Frichot et al. 2013). In simulation studies, LFMM has demonstrated a good balance between high power and low false-positive rate (Frichot et al. 2013; de Villemereuil et al. 2014). As suggested by Frichot et al. (2013) and Frichot and Francois (2015a), we used two methods to determine the optimal number of latent factors (K-value) that correct for the neutral genetic structure of our data. First, we ran a principal components analysis on the individual allele frequencies. We then determined the number of components that explain the genetic variance, based on the Tracy-Wisdom test on the eigenvalues, as an estimate of K (Frichot and Francois 2015a). Second, we utilized the Bayesian clustering algorithm STRUCTURE that estimates the number of genetic clusters (K) without prior information about geographic origin (Pritchard et al. 2000). The best K value was determined using the ΔK statistic as

suggested by Evanno et al. (2005). We used four replicates and a burnin of 300000 and 1000000 MCMC repeats after burnin for K = 2 - 12.

We ran LFMM to test for associations between SNP's and two composite climate variables (PC1 and PC2) using ten independent replications at 50000 iterations after a burnin period of 25000 with the number of latent factors (K) ranging from 8 to 12, as the methods outlined above suggested K equal to 10 and 9 respectively. We chose high run length parameters because of the relatively small number of individuals and loci. LFMM uses the z-scores to indicate the strength of the gene-environment association (Frichot and Francois 2015b). As suggested by the authors, we calculated the median z-score from ten replicate runs, re-adjusted the p-values, controlled for FDR using the q-value of 0.05, and determined candidate SNPs based on the Benjamini-Hochberg procedure (Frichot and Francois 2015b).

We also utilized RDA, a multivariate GEA method, to test for more subtle polygenic signatures of adaptation and detect outlier loci as candidates of functional importance. Redundancy Analysis (RDA) is an extension of multiple regression to multivariate response variables (Legendre and Legendre, 2012). In finding the ideal combination of predictor and explanatory variables RDA has shown high power to detect potential signals of polygenic adaptation (Harrison et al. 2017; Forester et al. 2018). For these analyses, Hellinger transformed allele frequencies (Legendre and Gallagher 2001) were treated as response variables. Because RDA models do not allow missing data, we imputed allele frequency data using probabilistic principal components analysis (ppca) as implemented in the 'pcaMethods' package in R (Stacklies et al. 2007). Ppca uses a decomposition of SNP frequencies to create principal components, the components with the largest eigenvalues are then used to impute the missing data. We evaluated space and climate as explanatory variables. Space was defined by distance-based Moran's eigenvector maps (dbMEMs; Borcard and Legendre 2002, Dray et al. 2006) based on Euclidean distances between all giant sequoia groves (sampled and unsampled) and extracting the values that correspond to our sample sites. Then we conducted backward model selection, using the 'ordistep' function within the vegan package for R (Oksanen et al. 2013), to reduce the number of dbMEM vectors. For climate, we reduced the twenty-one untransformed environmental variables described above, first by removing highly correlated environmental variables, ($|r| < 0.7$), and subsequently by using the 'ordistep' function for backwards model selection to remove variables lacking explanatory power. The above process resulted in climate being represented by 'Isothermality' (ISO), a measure diurnal and annual temperature fluctuation, 'Precipitation of Driest Quarter' (PDQ), a measure of summer precipitation in Mediterranean climates, and 'Climate Water Deficit' (CWD) , a measure of aridity, in all RDA models, and space represented by two dbMEM vectors, MEM3 and MEM5. All variables were centered and standardized before use in each model.

We set up multiple RDA models to determine the relative amount of variation in allele frequency explained by climate after correcting for geographic space as a signature of local adaptation (Lasky et al. 2012; Sork et al. 2016; Harrison et al 2017). First, to elucidate the major factors shaping genetic variation and to detect potential signals of local adaptation, we set up three models for comparison: a full RDA model where allele frequencies were associated with both climate and spatial explanatory variables, a partial RDA in which the effects of climate were conditioned on geography (dbMEMs) and a second partial RDA, where the effects of

geography were conditioned on climate. Next, to detect outlier loci, allele frequencies were associated with climate predictors after removing the effects of spatial predictors (Lasky et al. 2012; Harrison et al. 2017; Forester et al. 2018). Using the first constrained axis, we identified candidate SNPs of potential adaptive importance as those with loadings in the tails of a 95% confidence interval from the mean, or (2.0SD from the mean loadings). One risk of using such a low cutoff is an elevated rate of false positives. However, we chose this to maximize the number of SNPs detected, as we did not expect to find single loci that would be under very strong selection for climate variation in the range of giant sequoia. Moreover, we also identified the climate predictor with the highest correlation to each indicated SNP. In all RDA models we assessed model and constrained-axis significance using 999 permutations.

Genomic context of outlier loci

To gain insights into the potential adaptive significance of outlier loci, we obtained the flanking sequence of each outlier SNP locus from the giant sequoia reference sequence (Redwood Genome Project 2019). Since the giant sequoia reference genome is not annotated, functional annotation was performed using the online BLAST (Basic Local Alignment Search Tool) database. Using a 601bp sequence (300bp up and downstream of the SNP site) we searched the NCBI database using BLASTn with an e-value cut-off set to 1×10^5 and the requirement of > 70% sequence similarity.

RESULTS

Genetic diversity and differentiation

Genetic diversity and differentiation differed substantially across the eighteen sampled populations (Table 1). Observed heterozygosity (Ho) ranged from 0.09 to 0.17 and was lowest in PLAC and highest in ATWL and FMAN (Table 1). Unbiased expected heterozygosity (uHe) was also lowest in PLAC (0.07) and highest in ATWL (0.21) (Table 1). Average pairwise population differentiation (G'_{ST}) varied from 0.09-0.32 and was lowest for GFOR and highest for PLAC (Table 1). Average G'_{ST} was significantly higher in the northern populations than in the southern populations ($G'_{ST}N = 0.235$ $G'_{ST}S = 0.109$, Prob $G'_{ST}N \neq G'_{ST}S = 0.017$). Diversity analysis of SNP's without minor allele count (MAC) filtering yielded significantly different results: Ho and uHe were lower and ranged from 0.08 to 0.15 and 0.06 to 0.18 respectively ($P < 0.001$, $P = 0.01$ respectively, Table 1). Whereas, G'_{ST} was slightly higher in the dataset without MAC filtering (G'_{ST} 0.09-0.34, $P = 0.003$, Table 1).

Hierarchical AMOVA found a small, but significant variance due to regions ($F_{CT} = 0.02$, $P = 0.000$) and a larger portion of genomic variation distributed among populations ($F_{ST} = 0.15$, $P = 0.000$; Appendix 3.4). Population clustering (STRUCTURE) at $K = 9$, indicated strong differentiation among many of the northern populations (north of GRNT) with little admixture (Figure 2). In addition, populations NELD, GFOR, ATWL, GRNT, MCTR, and FMAN were assigned to the same cluster and the four southernmost populations, LMDW, CNHM, PKSD, and DCRK, consisted of two clusters with PKSD as a transitional population exhibiting admixture from both clusters (Figure 2). Finally, RMNT and GRNT show admixture between the neighboring GFOR, ATWL, GRNT, MCTR, FMAN, cluster, and the geographically separate LMDW, CNHM, cluster (Figure 2).

F_{ST} outliers

BayeScan indicated seven F_{ST} outliers, six demonstrating evidence for divergent selection with F_{ST} values ranging from 0.55-0.72 (Table 2, Appendix 3.5) and one showing signs of balancing selection (locus 1114, F_{ST} = 0.04). Since our focus here is on patterns of spatially varying selection, no further discussion is presented for locus 1114. A BLAST search found one of these loci exhibited functional significance (Locus 828; Table 2).

Candidate genomic regions associated with climate variables

Univariate environmental association analyses (LFMM with K = 9) indicated a total of 58 loci with strong correlations to composite environmental variables (Figure 3; Appendix 3.5). Of these, 51 were correlated with PC2 that was predominantly driven by precipitation, and seven were correlated with the temperature-driven PC1 (Figure 3; Table 2; Appendix 3.5). An examination of the adjusted P-values from all runs (K set from 8-12) provided additional support for K = 9 (Frichot and Francois 2015b). Our BLAST analysis was successful in finding functional annotation for three loci that were strongly associated with PC1 and 11 loci that were associated with PC2 (Table 2).

We used the partial RDA model to detect outlier loci as candidates of importance in selection in a multivariate context, where allele frequencies were associated with climate after removing the effect of spatial predictors. Using the SNP loadings on the first RDA axis, we identified 24 outlier loci beyond the 95% confidence that demonstrated strong correlations to environmental variation, all of which were most correlated with Precipitation of the Driest Quarter (PDQ) (Figure 4; Appendix 3.5 and 3.6). Our annotation procedure supported functional importance for seven of the 24 loci (Table 2).

Concordance among tests for signatures of selection

Overall, eight loci were detected as outliers by both RDA and LFMM. All of these loci were most associated with precipitation related variables (PC2 in LFMM, and PDQ in RDA). Annotation through BLAST identified two of these loci as having a putative function (Table 2). Overlap was found between a BayeScan (F_{ST} outlier) and LFMM at one locus (1123; Table 2).

Partitioning Variation Between Climate and Geographic Space

The full RDA model explained 45 percent of the total variation in allele frequency and supported an influence of climate and/or space in shaping allelic variation (P = 0.001; adjusted R^2 = 0.22). The first two canonical axes from the full RDA model were significant (P = 0.002, and 0.013 respectively) and together accounted for 77 percent of the explained variation (Figure 4). The partial RDA model, with climate conditioned on space, was significant (P = 0.019; adjusted R^2 = 0.09) and constrained 49 percent of the variance explained by the full model. The first partial RDA axis was significant at the 0.1 level (P = 0.098) and accounted for 45 percent of the variation. The partial RDA model with space conditioned on climate accounted for 24 percent of the explained variation and was non-significant (P = 0.207). The remaining 27 percent of the explained variation was confounded between climate and geography.

DISCUSSION

Giant sequoia is a paleoendemic of California that has likely suffered from a long-term demographic decline (Dodd and DeSilva 2016). Today, it is limited to a number of restricted groves in the Sierra Nevada mountain chain. Small grove sizes and limited gene exchange

among populations (DeSilva and Dodd 2020) might be expected to limit its adaptive potential through inbreeding effects and genetic drift. However, through different approaches we have found evidence for a signal of spatially varying local adaptation associated with climate variables and, in particular, along gradients dominated by precipitation. We report here that population genetic structure in giant sequoia has been shaped by local adaptation overlain on historical population processes. From our study of genomic variation, we detected 79 loci as either F_{ST} outliers, or loci with strong associations to climate as candidate regions of adaptive importance. Of these, we highlight 26 SNPs, found from multiple methods, or that correspond with functional annotation, as prime candidates for additional research. We emphasize that these outlier loci may include false positives and that experimental studies are needed to demonstrate functional significance of putative adaptive genomic regions (Kawecki and Ebert 2004; Barrett and Hoekstra 2011). Here, we present a first step towards understanding local adaptation in an iconic forest tree.

Population divergence and structure

We found evidence for strong population differentiation in our genomic data ($F_{ST} = 0.15$), which was close to our earlier estimate of $F_{ST} = 0.14$ from microsatellite variation (DeSilva and Dodd 2020). Such high levels of differentiation are unusual in wind-pollinated tree species, for which population differentiation is typically low and suggests that, at least some populations have been isolated for a considerable time (McKay and Latta 2002; Petit and Hampe 2006). The 18 giant sequoia populations that we sampled, covering the range of the species, could be partitioned into nine clusters. Six of these clusters were restricted to northern isolated groves and the remaining three clusters include all southern populations that are somewhat more contiguous. Our results confirm the previously report of strong population structure among populations north of GRNT based on microsatellite data (Dodd and DeSilva 2016; DeSilva and Dodd 2020). Genetic diversity (unbiased heterozygosity) was no lower in the northern fragmented groves than in most populations within the range, supporting our earlier inference that northern groves have a long evolutionary history (Dodd and DeSilva 2016). Moreover, estimates of uHe (calculated after removing putative adaptive sites; see Appendix 3.7), show consistent patterns across populations as previous estimates of He from microsatellite markers (Dodd and DeSilva 2020), although uHe is lower in the SNP dataset. This pattern of population and genetic structure is unusual for north temperate conifers, for which higher latitude populations are commonly thought of as "leading edge" colonization following glacial retreat. Current groves extend above the lower extent of late Pleistocene glaciers (Moore and Moring 2013), so either some short distance upward colonization must have occurred, or pockets of unglaciated terrain may have served as very local refugia. Given the pattern of genomic diversity that we have detected and the long generation time (~300 years) of giant sequoia, it seems most likely that extant groves have either persisted through many generations or were colonized by short distance migrations.

Evidence for local adaptation

Despite the strong structure among populations, analysis of our genomic data revealed a signal of divergent selection associated with climatic variables. BayeScan detected few F_{ST} outliers due, in part, to the high population structure and increasing the prior odds for the neutral model to 100. However, each of these outliers exhibited high levels of differentiation (F_{ST} 0.55 - 0.72; Appendix 3.5). Univariate (LFMM) and multivariate (RDA) environmental association approaches identified more loci indicative of local adaptation. Although F_{ST} outlier approaches

have been found to be more robust with respect to false positives than other methods (Lotterhos and Whitlock 2014), environmental association studies are more successful in detecting loci under selection and can provide context for selective forces as well (De Mita et al. 2013). Our environmental association studies found climate to be an important predictor of allele frequency and accounted for the largest portion of explained variation after correcting for geographic space; a pattern consistent with local adaptation. Signatures of local adaptation to climate are prevalent in many tree species, including *Picea mariana*, *Alnus glutinosa*, *Populus trichocarpa*, *Cornus florida*, and *Quercus lobata* (Prunier et al. 2011; De Kort et al. 2014; Geraldes et al. 2014; Pais et al. 2016; Sork et al. 2016). Moreover, the association of genomic variation with climate in giant sequoia is consistent with our previous work that found precipitation-related variables play a role in patterns of isolation by environment at neutral genetic markers (DeSilva and Dodd 2020). Here, a genome-wide data set that includes putative functional regions showed a signal of climatic factors shaping genomic variation, which suggests that local adaptation in-situ, likely under conditions of limited gene-flow, is important in this species.

Local adaptation is further supported by nine loci detected by multiple methods and 19 candidate loci with functional annotation (Table 2). Overlap in detection of outlier loci has been reported in numerous field studies (Hess et al. 2016; Sork et al. 2016; Harrison et al. 2017). A carefully designed simulation study demonstrated that overlap between GEA methods was found more often for actual targets of selection rather than false positives (Forester et al. 2018). In addition, two of the eight loci detected by both RDA and LFMM have relevant functional annotation, (loci 827 and 1229; Table 2). A BLAST search suggests that Locus 827 is a kinesin-like protein, KIN-13A, which has been found to be involved in trichome morphogenesis (Lu et al. 2005). In plants, trichome occurrence and density is associated with increased drought resistance (Sletvold and Agren 2012; Galdon-Almero et al. 2018). Locus 1229 represents a potential pollen allergen gene, which is thought to be involved in plant responses to stress (Chen et al. 2016). Although our BLAST search suggested functional importance for 20 loci, many of the annotated regions are characterized only as mRNA with further functional roles yet to be determined (Table 2). The non-annotated outlier loci are promising candidates for future research as they may be of unknown importance, linked to adaptive genes, in regulatory regions, or represent false positives. Any future annotation of the giant sequoia genome will provide valuable clarity as to the specific role of all outlier loci. Yet, we emphasize the candidate loci noted in this study demonstrate only strong associations with climate and identifying the exact targets of selection involves rigorous experimental research. Taken together, outlier loci with functional annotation or those detected by multiple methods provide strong support for adaptive variation across the range of giant sequoia.

Outlier SNPs driven by precipitation

Interestingly, variables associated with precipitation appeared to be the major drivers of local adaptation, which perhaps reflects the strong gradients of water relations on the western slope of the Sierra Nevada. Using gene-environment association methods (LFMM and RDA) we found evidence for adaptive differentiation across giant sequoia populations in response to gradients in precipitation and a more limited signal of local adaptation to temperature. LFMM analyses demonstrated seven times as many outliers correlated with precipitation-related PC2 than to temperature-related PC1 (Figure 3, Appendix 3.5). Although, all RDA outliers were correlated with three environmental factors, PDQ, a measure of summer precipitation, CWD, a

measure of aridity, and ISO, a measure diurnal and annual temperature fluctuation, outlier loci showed the strongest relationship to PDQ (Appendix 3.6). Thus, both LFMM and RDA indicate a subtle signal of adaptation to temperature and a stronger signature of divergent selection in response to gradients in water-related variables. Gradients of water availability are important selective agents for many tree species including *Picea mariana*, *Cornus florida, Fagus sylvatica, Quercus spp.*, and *Pinus albicaulis* (Prunier et al. 2011; Pais et al. 2016; Pluess et al. 2016; Sork et al. 2016, Lind et al. 2017; Martins et al. 2018). In addition, many ecological studies have noted that giant sequoia is sensitive to water availability during its establishment phase (Rundel 1972; Hartesveldt et al. 1975; York et al. 2003; Shellhammer and Shellhammer 2006). Giant sequoia is known to have bursts of reproduction after fire and subsequently experience high seedling mortality due to desiccation (Weatherspoon 1990). Considering the reproductive biology of this species, water availability is a highly plausible selective agent. For Mediterranean type climates PDQ is of particular relevance for desiccation sensitive species as it equates to summer precipitation, which may represent a vital water source for giant sequoia seedlings during a vulnerable establishment phase.

Limitations, opportunities, and future implications

It is important to note that GEA methods can suffer from low power or high false positive rate under some demographic scenarios. Although LFMM has been shown to be robust to various demographic scenarios, including those that create high levels of population structure, this method can have elevated false discovery rates (FDR) under scenarios that create IBD (de Villemereuil et al. 2014; Lotterhos and Whitlock 2015; Forester et al. 2016). Here, we do not find a significant signal of IBD in our data. Yet, the role of IBD or other neutral factors affecting population structure in giant sequoia has not been fully elucidated. Previous research has indicated isolation by distance (IBD) and/or ancient divergence separating the northern and southern populations of giant sequoia (Dodd and DeSilva 2016; DeSilva and Dodd 2020). Here, a large portion of the explained variance in our data (27%) was confounded between climatic variation and geographic space, which is perhaps due, to the strictly north-south range of giant sequoia, making it inherently difficult to decouple distance from environmental gradients that vary latitudinally. Therefore, LFMM results should be treated with some caution due to the potential contribution of IBD to population structure. In contrast to LFMM, RDA models show high power and low false positive rates under IBD (Forester et al. 2016). Simulation studies indicate that the performance of RDA also remains high when population structure and selective gradients are explicitly correlated (Capblancq et al. 2018). Yet, RDA is not without limitations, as it can have low power under island demographic models (Forester et al. 2018) or when selective pressures are highly clustered (Capblancq et al. 2018). Given that each GEA method has particular limitations, we believe the outlier loci detected by both LFMM and RDA, as well as outliers with functional annotation remain strong candidate regions of adaptive importance.

There is an ongoing need for future studies to provide additional clarity on the distribution of adaptive variation and genetic architecture of local adaptation in giant sequoia. To our knowledge, this study represents the first investigation of adaptive variation using genome-wide data. Yet, the results presented here are based on a small subset of the genomic variation within the species, as the giant sequoia genome is very large (8.5Gb, Redwood Genome Project). More comprehensive sampling of the genome as well as an incorporation of phenotypic information will greatly improve our understanding of local adaptation in this species. In

addition, future annotation of the giant sequoia genome will provide opportunities to better understand the genetic underpinnings of many phenotypic traits.

A trend towards increased aridity along mid-elevation Sierra Nevada forests could undermine the long-term persistence of giant sequoia. With end-of-century predictions for this region that include decreasing snowfall and earlier snowmelt, forests of the Sierra Nevada mountains will likely experience an accentuation of the summer drought that is typical of Mediterranean climates (Stewart et al. 2004; Fyfe et al 2017; Sun et al. 2018). Considering the evidence presented here, we highlight the potential that increased water stress may create mal-adaptation of giant sequoia populations to their environment. It has been suggested that a long-term decline (over the last ~2My) of giant sequoia is tied to increasing aridity during the development of current climate regimes (Dodd and DeSilva 2016). Moreover, some giant sequoia populations suffered extensive foliage die back during the drought period from 2012-2016 (Stephenson et al. 2018), providing further indication of sensitivity of giant sequoia to arid conditions.

CONCLUSIONS

We provide evidence of local adaptation along gradients of precipitation and highlight genomic regions of potential adaptive importance for additional research. This information can aid in determining the best course of action to preserve giant sequoia into the future. Locally adapted populations of giant sequoia are facing an accentuation of summer drought to which they may be mal-adapted. Genomic variation currently present in more arid regions of the giant sequoia range could include 'pre-adapted' variants that might enhance the adaptive response of nearby populations (Kremer et al. 2012; Aitken and Bemmels 2016). Currently, DCRK, GRNT, MCTR, and RMNT inhabit areas experiencing the lowest levels of summer precipitation (PDQ), and thus may provide potential sources of adaptive alleles. Given the limited gene flow in much of giant sequoia range, it is unlikely this variation will spread quickly by natural means (DeSilva and Dodd 2020). Thus, forest managers may consider assisting in the movement of genetic resources in order to enhance the adaptive potential of giant sequoia populations.

ACKNOWLEDGEMENTS

We are grateful to Matthew Hughes for assistance with leaf collections and to Prahlada Papper for vital help with data processing. This research was partially funded by Mcintire Stennis grant CA-B-ECO-0165-MS awarded to R.S.D. We also thank reviewers whose feedback helped improve the manuscript. In August 2020, this chapter was published in Ecology and Evolution.

REFERENCES

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. L., and Curtis-McLane, S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications. 1: 95–111.

Aitken, S. N., and Bemmels, J. B. 2016. Time to get moving: assisted gene flow of forest trees. Evolutionary Applications, 9: 271–290.

Aitken, S. N., and Whitlock, M. C.. 2013. Assisted gene flow to facilitate local adaptation to climate change. Annual Review of Ecology, Evolution, and Systematics 44: 367–388.

Akey, J. M. 2009. Constructing genomic maps of positive selection in humans: where do we go from here? Genome Research, 19: 711–722.

Alberto, F. J. , Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., Lefèvre, F., Lenormand, T., Yeaman, S., Whetten, R., and Savolainen, O. 2013. Potential for evolutionary responses to climate change—evidence from tree populations. Global Change Biology 19: 1645–1661

Barrett, R. D. H., and Hoekstra, H. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nature Reviews Genetics. 12: 767–780.

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

Borcard, D., and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling. 153: 51–68.

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., Rochette, R. and Bernatchez, L. 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). Molecular Ecology, 25: 5073-5092. doi:10.1111/mec.13811

Capblancq, T., Luu, K., Blum, M. G. B., and Bazin, E. Evaluation of redundancy analysis to identify signatures of local adaptation. Molecular Ecology Resoures. 18: 1223– 1233. https://doi.org/10.1111/1755-0998.12906

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. and Cresko, W.A. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology. 22: 3124–3140.

Chen, M., Xu, J., Devix, D., Shi, J., Ren, K., Searle, I., and Zhang, D., 2016. Origin and Functional Prediction of Pollen Allergens in Plants. Plant Physiology 172: 341-357

De Kort, H., Vandepitte, K., Bruun, H. H. Closset-Kopp, D., Honnay, O., and Mergeay, J. 2014. Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species Alnus glutinosa. Molecular Ecology. 23: 4709–4721

De Mita S., Thuillet A-C, Gay L. et al. (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing popula- tions. Molecular Ecology, 22, 1383–1399.

DeSilva R, and Dodd R. S. 2020. Fragmented and isolated: limited gene flow coupled with weak isolation by environment in the paleoendemic giant sequoia (Sequoiadendron giganteum). American Journal of Botany. 107 (1): 45-55

Dodd, R. A., and DeSilva, R. 2016. Long-Term Demographic Decline and Late Glacial Divergence in a Californian Paleoendemic: Sequoiadendron giganteum, giant sequoia. Ecology and Evolution 6: 3342-3355.

Dray, S., Legendre, P., and Peres-Neto, P. R. 2006. Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling. 196: 483–493. https:// doi.org/10.1016/j.ecolmodel.2006.02.015

Dudaniec, R. Y., Yong, C. J., Lancaster, L. T., Svensson, E. I., and Hansson, B. Signatures of local adaptation along environmental gradients in a range-expanding damselfly (Ischnura elegans). Molecular Ecology. 2018: 27:2576–2593. https://doi.org/10.1111/mec.14709

Excoffier, L., Hofer, T., and Foll, M. 2009 Detecting loci under selection in a hierarchically structured population. Heredity. 103: 285–298.

Emerson, K. J., Merz, C. R., Catchen, J. M. Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., and Holzapfel, C. M. 2010. Resolving post-glacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences of the United States of America. 107: 16196–16200.

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

Fick, S. E. and Hijmans, R. J. 2017. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology. doi: 10.1002/joc.5086.

Flint, L. E., Flint, A. L., Thorne, J. H., and Boynton, R. 2013. Fine-scale hydrologic modeling for regional landscape applications: the California Basin Characterization Model development and performance. Ecological Processes. 2:25. doi: https://doi.org/10.1186/2192-1709-2-25

Foll M, 2010. BayeScan v2.0 User Manual. Accessed November 1st 2019. http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.0_manual.pdf

Foll, M., and Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics. 180: 977–993.

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L. and Lasky, J. R. 2016. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular Ecology, 25: 104-120. doi:10.1111/mec.13476

Forester, B. R., Lasky, J. R., Wagner, H. H., and Urban, D. L. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Molecular Ecology. 27: 2215– 2233. https://doi.org/10.1111/mec.14584

Frichot E., Schoville S. D., Bouchard G., François O., 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Molecular Biology and Evolution 30: 1687–1699

Frichot and François, O. 2015a. LEA: An R package for landscape and ecological association studies. Methods Ecology and Evolution, 6: 925-929. doi:10.1111/2041-210X.12382

Frichot and François, O. 2015b A short manual for LFMM version 1.4. Accessed November 15th 2019. http://membres-timc.imag.fr/Olivier.Francois/lfmm/files/note.pdf

Funk, W. C., McKay, J. K., Hohenlohe, P. A., and Allendorf, F. W. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution. 27: 489–496.

Fyfe, J., Derksen, C., Mudryk, L. Flato, G. M., Santer B. D., Swart, N. C., Molotch, N. P., Zhang. X., Wan, H., Arora. V. K., Scinocca, J., and Jiao, Y. 2017. Large near-term projected snowpack loss over the western United States. Nature Communications 8: 14996. doi:10.1038/ncomms14996

Galdon-Armero, J., Fullana-Pericas, M., Mulet, P. A., Conesa, M. A., Martin, C. and Galmes, J. 2018. The ratio of trichomes to stomata is associated with water use efficiency in Solanum lycopersicum (tomato). Plant J, 96: 607-619. doi:10.1111/tpj.14055

Garrison, E, Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN]

Geraldes, A., Farzaneh, N., Grassa, C.J., McKown, A. D., Guy, R. D., Mansfield, S. D., Douglas, C. J., and Cronk, Q. C. B. 2014. Landscape genomics of Populus Trichocarpa: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. Evolution, 68(11): 3260–3280. https://doi.org/10.1111/evo.12497

Gunther, T., Coop, G. 2013. Robust identification of local adaptation from allele frequencies. Genetics. 195; 205–220

Harrisson, K. A., Amish, S. J., Pavlova, A., Narum, S. R., Telonis-Scott, M., Rourke, M. L., Lyon, J., Tonkin, Z., Gilligan, D. M., Ingram, B. A., and Lintermans, M. 2017. Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. Molecular Ecology. 26: 6253– 6269. https://doi.org/10.1111/mec.14368

Hartesveldt, R. J., Harvey, H. T., Shellhammer, H. S., and Stecker, R. E. 1975. The giant sequoia of the Sierra Nevada. U. S. Department of the Interior, National Park Service, Washington, DC; 1975.

Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. American Naturalist 173: 579–588.

Hess, J. E., Zendt, J. S., Matala, A. R., and Narum, S. R. 2016. Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. Proceedings of the Royal Society. B, Biological Sciences, 283: 20153064.

Holderegger, R., Kamm, U., and Gugerli, F. 2006. Adaptive vs. neutral genetic diversity: Implications for landscape genetics. Landscape Ecology. 21(6): 797-807. doi:http://dx.doi.org.libproxy.berkeley.edu/10.1007/s10980-005-5245-9

Holland, J. 2007. Genetic architecture of complex traits in plants. Current Opinion in Plant Biology. 10: 156–161.

Holliday, J.A., Zhou, L., Bawa, R., Zhang, M., and Oubida, R. W. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in Populus trichocarpa. New Phytologist. 209: 1240–1251.

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

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., and Prodöhl, P. A. 2013. diveRsity: An R package for the estimation of population genetics parameters and their associated errors, Methods in Ecology and Evolution. 4: 782-788.

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J. R., Gomulkiewicz, R., Klein, E. K., Ritland, K., Kuparinen, A., Gerber, S. and Schueler, S. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters. 15: 378-392

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with bowtie 2. Nature Methods. 9: 357–360.

Lasky, J. R., Des Marais, D. L., McKay, J. K., Richards, J. H., Juenger, T. E. and Keitt, T. H. 2012. Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. Molecular Ecology. 21: 5512-5529. doi:10.1111/j.1365-294X.2012.05709.x

Le Corre, V. and Kremer, A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Molecular Ecology. 21: 1548-1566. doi:10.1111/j.1365-294X.2012.05479.x

Legendre, P. & Gallagher, E.D. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia**.** 129: 271–280. https://doi.org/10.1007/s004420100716

Legendre, P., and Legendre, L. 2012. Numerical ecology, 3rd English ed., vol. 24. Elsevier, Boston, Massachusetts, USA. 990.

Leimu, R., and Fischer, M. 2008. A meta-analysis of local adaptation in plants. PLoS One, 3: e4010.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., 2009. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Linnen, C. R., Kingsley, E. P., Jensen, J. D., Hoekstra, H. E. 2009. On the origin and spread of an adaptive allele in deer mice. Science. 325: 1095–98

Lind, B. M., Friedline, C. J., Wegrzyn, J. L., Maloney, P. E., Vogler, D. R., Neale, D. B. and

Lotterhos, K. E., and Whitlock, M. C. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Molecular Ecology. 23: 2178–2192.

Lotterhos, K. E. and Whitlock, M. C. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology. 24: 1031-1046. doi:10.1111/mec.13100

Lu, L., Lee, Y. J., Pan, R., Maloof, J. N., and Liu B. 2005. An internal Motor Kinesin Is Associated with the Golgi Apparatus and Plats a Role in Trichome Morphogenesis in Arabidopsis. Molecular Biology of the Cell. 16: 811-823

McKay, J. K., and Latta, R. G. 2002. Adaptive population divergence: Markers, QTL and traits. Trends in Ecology & Evolution. 17: 285–291. https://doi.org/10.1016/S0169-5347(02)02478-3

Mitton, J. B., Stutz, H. P., and Schuster, W. S. 1989. Genotypic differentiation at PGM in Engelmann spruce from wet and dry sites. Silvae Genetica. 38: 217–221.

Martins, K., Gugger, P. F., Llanderal-Mendoza, J., Gonzalez-Rodrigruez, A., Fitz-Gibbon, S. T., Zhao, J., Rodriguez-Correa H., Oyama, K., and Sork V. L. 2018. Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of Quercus rugosa. Evolutionary Applications. 11: 1842– 1858. https://doi.org/10.1111/eva.12684

Meirmans, P.G., and Van Tienderen, P.H. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms, Molecular Ecology Notes. 4:792-794.

Moore, J. G. and Morning, B. C. 2013. Rangewide glaciation in the Sierra Nevada, California. Geosphere. 9(6): 1804–1818. https://doi.org/10.1130/GES00891.1

Morgenstern, E. K. 1996. Geographic variation in forest trees: genetic basis and application of knowledge in silviculture. UBC Press, Vancouver.

Narum, S. R., and Hess, J. E. 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. Molecular Ecology Resources. 11: 184-194. doi:10.1111/j.1755-0998.2011.02987.x

Nosil, P., Vines T., and Funk, D. 2005. Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution; international journal of organic evolution. 59: 705-719.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. 2013. vegan: Community Ecology Package. Assessed December 1st 2019. https://cran.r-project.org/web/packages/vegan/vegan.pdf

Olson-Manning, C. F., Wagner, M. R., Mitchell-Olds, T. 2012. Adaptive evolution: evaluating empirical support for theoretical predictions. Nature Review Genetics. 13(12): 867–877. doi:10.1038/nrg3322

Pais, A. L., Whetten, R. W., and Xiang, Q. J. 2016 Ecological genomics of local adaptation in Cornus florida L. by genotyping by sequencing. Ecology and Evolution. 7(1): 441-465. doi:10.1002/ece3.2623

Peterson, G. W, Dong, Y. Horback, C. and Yong-Bi, F. 2104. Genotyping-By-Sequencing for Plant Genetic Diversity Analysis: A Lab Guide for SNP Genotyping. Diversity. 6(4): 665-680

Petit, R. J., and Hampe, A. 2006. Some evolutionary consequences of being a tree. Annual Review of Ecolology, Evolution, and Systematics. 37: 187–214

Pluess, A. R., Frank, A., Heiri, C., Lalagüe, H., Vendramin, G. G. and Oddou-Muratorio, S. 2016. Genome–environment association study suggests local adaptation to climate at the regional scale in Fagus sylvatica. New Phytologist. 210: 589-601. doi:10.1111/nph.13809

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

Pritchard, J. K., and Di Rienzo, A. 2010. Adaptation–not by sweeps alone. Nature Reviews Genetics. 11: 665–667.

Prunier, J., Laroche, J., Beaulieu, J., and Bousquet, J. 2011. Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. Molecular Ecology. 20: 1702–16

Redwood Genome Project. 2019. Sequoiadendron giganteum reference genome 1.0. Accessed December 1st 2018. https://nealelab.ucdavis.edu/redwood-genome-project-rgp/

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., and Holderegger, R. 2015. A practical guide to environmental association analysis in landscape genomics. Molecular Ecology. 24: 4348–4370.

Rogers, A., and Prügel-Bennett, A. 2000. Evolving populations with overlapping generations. Theoretical Population Biology. 57: 121-129

Rundel, P. 1972. Habitat Restriction in Giant Sequoia - Environmental Control of Grove Boundaries. American Midland Naturalist. 87: 81-99

Savolainen O, Pyhäjärvi T, and Knurr T (2007) Gene flow and local adaptation in trees. Annual Review of Ecology, Evolution and Systematics, 38, 595–619.

Sella, G., Petrov, D. A., Przeworski, M., and Andolfatto, P. 2009. Pervasive natural selection in the Drosophila genome? PLoS Genetics. 5:e1000495

Shellhammer, H., and Shellhammer, T. 2006. Giant Sequoia (Sequoiadentron Giganteum [Taxodiaceae]) Seedling Survival and Growth in the First Four Decades Following Managed Fires. Madroño, 53(4): 342-350.

Sgrò, C. M., Lowe, A. J., and Hoffmann, A. A. 2011. Building evolutionary resilience for conserving biodiversity under climate change. Evolutionary Applications. 4: 326–337.

Sletvold, N. and Ågren, J. 2012. Variation in tolerance to drought among Scandinavian populations of Arabidopsis lyrata. Evolutionary Ecology. 26: 559-577. 10.1007/s10682-011-9502-x.

Sork, V., Aitken, S., Dyer, R., Eckert, A., Legendre, P., and Neale, D. 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genetics & Genomes 9: 901–911

Sork, V. L., Squire, K., Gugger, P. F., Steele, S. E., Levy, E. D., and Eckert, A. J. 2016. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, Quercus lobata. American Journal of Botany. 103: 33–46

Stacklies, W., Redestig, H., Scholz, M., Walther, D., Selbig, J. 2007. pcaMethods—a bioconductor package providing PCA methods for incomplete data, Bioinformatics. 23: 1164–1167.

Stewart, I., Cayan, D., and Dettinger, M. 2004. Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. Climatic Change. 62: 217-232.

Stephenson, N. L. 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. Journal of Biogeography. 25: 855–870.

Stephenson, N. L., Das, A. J. Ampersee, N. J. Cahill, K. G., Caprio, A. C., Sanders., J. E., and Williams, A. P. 2018. Patterns and correlates of giant sequoia foliage dieback during California's 2012–2016 hotter drought. Forest Ecology and Management, Volumes 419–420, 268-278

Sun, F., Berg, N., Hall. A., Schwartz, M., and Walton, D. 2018. Understanding end-of-century snowpack changes over California's Sierra Nevada. Geophysical Research Letters. 46: 933-943

[USDA–FS] US Department of Agriculture–Forest Service. 2016. New Aerial Survey Identifies More than 100 Million Dead Trees in California. Accessed January 1st 2020. https://www.fs.usda.gov/news/releases/new-aerial-survey-identifies-more-100-million-dead-trees-california

Valness, C. M. 2016. Performance and Morphology in Sequoiadendron Genotypes Outside of Their Range. Master's Thesis. Humboldt State University, Arcata, CA, USA.

de Villemereuil, P., Frichot, É., Bazin, É., François, O., and Gaggiotti, O. E. 2014. Genome scan methods against more complex models: When and how much should we trust them? Molecular Ecology. 23: 2006–2019.

Wang, I. J. and Bradburd, G. S. 2014. Isolation by environment. Molecular Ecology. 23: 5649-5662.

Weatherspoon, C. P., 1990. Sequoiadendron giganteum (Lindl.) Buchholz Giant Sequoia. In: Burns, R.M., Honkala, B.H. (Eds.), Silvics of North America, USDA Forest Service Agricultural Handbook 654, Washington, DC, pp. 552–562.

Wójkiewicz, B., Litkowiec, M., and Wachowiak, W. 2016. Contrasting patterns of genetic variation in core and peripheral populations of highly outcrossing and wind pollinated forest tree species. AoB PLANTS. 8: plw054.

Yeaman, S., 2015. Local Adaptation by Alleles of Small Effect. The American Naturalist. 186:S1 S74-S89

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., Nurkowski, K. A., Smets, P., Wang, T., Gray, L. K., Liepe, K. J., Hamann, A., Holliday, J. A., Whitlock, M. C., Rieseberg, L. H., and Aitken, S. N. 2016. Convergent local adaptation to climate in distantly related conifers. Science. 353: 1431–1433.

York, R., Battles, J., and Heald, R. 2003. Edge effects in mixed conifer group selection openings: tree height response to resource gradients. Forest Ecology and Management. 179: 107-121

TABLES AND FIGURES

TABLE 1 Population information and genetic diversity summary statistics calculated for each population. Diversity statistics calculated without minor allele filtering are noted within parentheses.

Grove name	Population code	GPS location	Sample size	Ho	uHe	F_{IS}	*Mean G′st*
Placer	PLAC	39.06, −120.57	6	0.09 (0.08)	0.07 (0.06)	0.00 (0.01)	0.32 (0.34)
North Calaveras	CALN	38.28, −120.30	8	0.14 (0.13)	0.19 (0.16)	0.17 (0.17)	0.17 (0.17)
South Calaveras	CALS	38.24, −120.25	8	0.16 (0.14)	0.18 (0.16)	0.13 (0.13)	0.15 (0.15)
Tuolumne	TUOL	37.77, −119.81	8	0.16 (0.14)	0.17 (0.14)	−0.04 (−0.04)	0.24 (0.25)
Merced	MERC	37.75, −119.84	8	0.16 (0.14)	0.17 (0.15)	0.00 (−0.01)	0.21 (0.21)
Mariposa	MPSA	37.51, −119.60	8	0.13 (0.11)	0.14 (0.12)	0.11 (0.11)	0.20 (0.20)
Nelder	NELD	37.43, −119.59	8	0.16 (0.15)	0.20 (0.17)	0.13 (0.11)	0.13 (0.14)
McKinley	MKLY	37.03, −119.11	8	0.15 (0.13)	0.17 (0.15)	0.07 (0.06)	0.17 (0.17)
Grant	GRNT	36.75, −118.97	8	0.12 (0.10)	0.13 (0.11)	0.17 (0.18)	0.14 (0.15)
Redwood Mountain	RMNT	36.60, −118.92	8	0.12 (0.10)	0.11 (0.10)	0.17 (0.18)	0.13 (0.13)
Giant Forest	GFOR	36.57, −118.76	8	0.15 (0.14)	0.20 (0.18)	0.18 (0.17)	0.09 (0.09)
Atwell	ATWL	36.47, −118.67	8	0.17 (0.15)	0.21 (0.18)	0.08 (0.08)	0.11 (0.12)
Mcintyre	MCTR	36.13, −118.58	8	0.13 (0.11)	0.11 (0.09)	0.10 (0.11)	0.12 (0.13)
Freeman Creek	FMAN	36.14, −118.52	8	0.17 (0.12)	0.19 (0.16)	0.08 (0.07)	0.10 (0.11)
Long Meadow	LMDW	35.96, −118.60	8	0.12 (0.11)	0.11 (0.10)	0.16 (0.15)	0.15 (0.16)
Cunningham	CNHM	35.92, −118.57	9	0.16 (0.15)	0.17 (0.16)	0.00 (0.00)	0.16 (0.16)
Packsaddle	PKSD	35.93, −118.59	8	0.15 (0.13)	0.15 (0.13)	0.08 (0.08)	0.16 (0.16)
Deer Creek	DCRK	35.88, −118.61	8	0.16 (0.14)	0.18 (0.16)	0.05 (0.04)	0.16 (0.16)

Table 2: Functional significance, detection method, and associated variable for highly supported outlier loci

Locus ID	**Detection Method**	**Associated Variable* (F_{ST})**	**Adjusted P-value**	**| Axis 1 Loading |**	**Annotation**
828	BayeScan	(0.58)			Unknown mRNA
1123	BS, LFMM	PC1(0.66)	1.72E-04		--
186	LFMM	PC1	7.39E-06	--	Unknown mRNA
218	LFMM	PC1	5.50E-05	--	Unknown mRNA
452	LFMM	PC1	5.30E-05	--	Unknown mRNA
251	LFMM	PC2	1.91E-05	--	Unknown mRNA
722	LFMM	PC2	1.52E-03	--	Unknown mRNA
870	LFMM	PC2	9.48E-04	--	Unknown mRNA
1029	LFMM	PC2	6.56E-06	--	Unknown mRNA
1062	LFMM	PC2	1.66E-04	--	Unknown mRNA
1214	LFMM	PC2	1.52E-03	--	magnesium transporter MRS2-4-like
1253	LFMM	PC2	2.91E-04	--	Unknown mRNA
1313	LFMM	PC2	5.87E-06	--	Pleiotropic drug resistance protein 1-like
368	LFMM, RDA	PC2, PDQ	2.35E-05	-0.21132	--
421	LFMM, RDA	PC2, PDQ	9.56E-05	-0.19423	--
471	LFMM, RDA	PC2, PDQ	2.03E-05	-0.19892	--
515	LFMM, RDA	PC2, PDQ	1.46E-03	0.21181	--
679	LFMM, RDA	PC2, PDQ	3.51E-04	0.22348	--
827	LFMM, RDA	PC2, PDQ	9.12E-04	0.20579	kinesin-like protein KIN-13A mRNA
1229	LFMM, RDA	PC2, PDQ	1.65E-03	0.17822	pollen allergen gene
1286	LFMM, RDA	PC2, PDQ	4.41E-07	0.26369	--
338	RDA	PDQ	--	0.18769	Wall-associated receptor kinase-like 1
612	RDA	PDQ	--	-0.24267	arogenate dehydratase gene
617	RDA	PDQ	--	0.17057	Unknown mRNA
940	RDA	PDQ	--	0.23536	signal peptidase I AT2G30440 mRNA
1066	RDA	PDQ	--	0.19520	Unknown mRNA

Figure 1: Range map of giant sequoia (black) showing the gradient of Precipitation of Driest Quarter (mm) across a section of California. Sampled populations indicated by a population code.

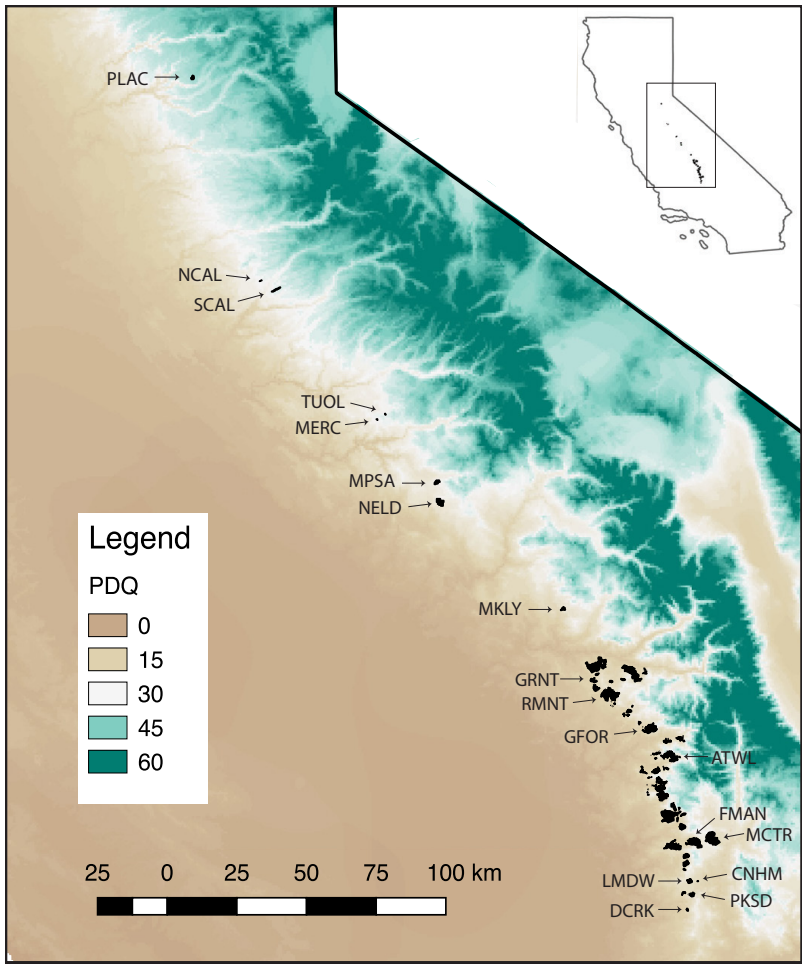


Figure 2: Results of population structure analyses from STRUCTURE. Vertical bars represent a sampled individual, color-coded for assigned cluster at K=9.

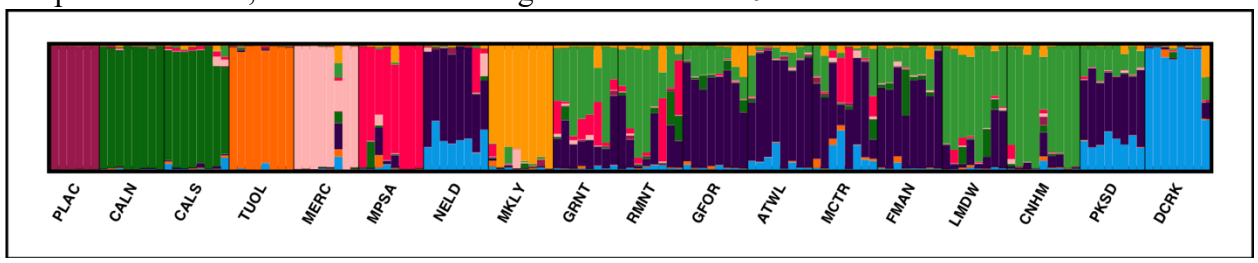


Figure 3: Adjusted P-values from LFMM for association with PC1 and PC2. Outliers are outlined in blue.

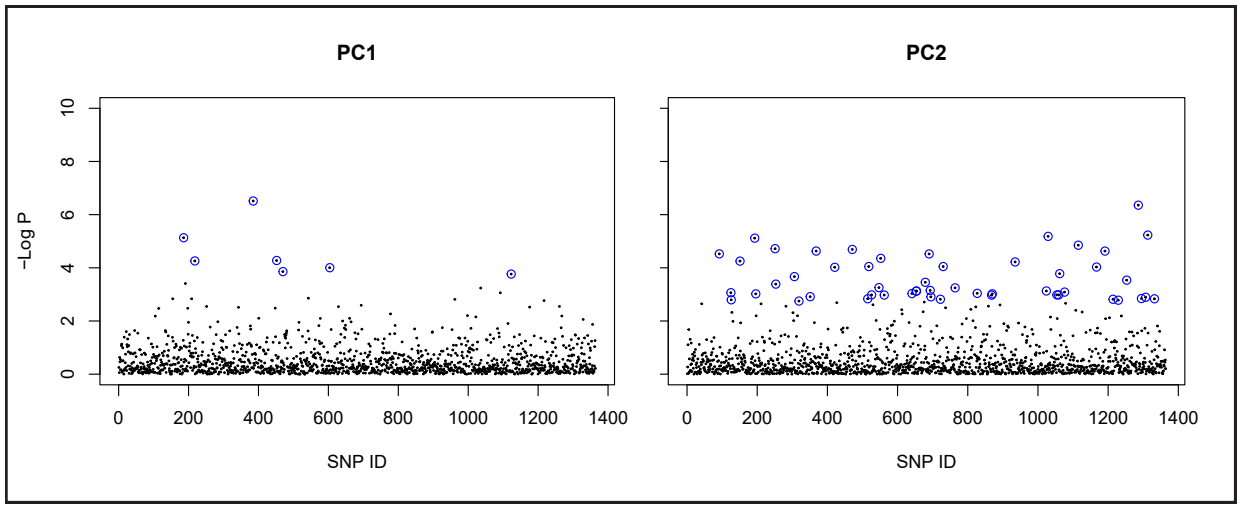
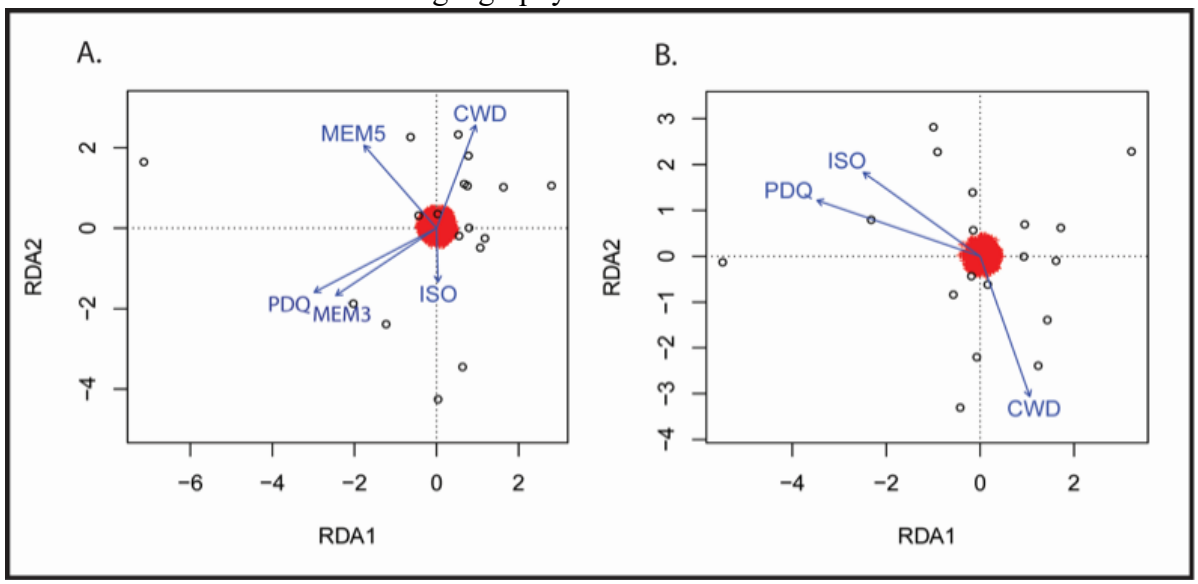


Figure 4: Triplot from Redundancy analysis showing how each explanatory variable affects the RDA axis with (A) representing the full RDA model and (B) a partial RDA model with the effects of climate conditioned on geography.



APPENDIX

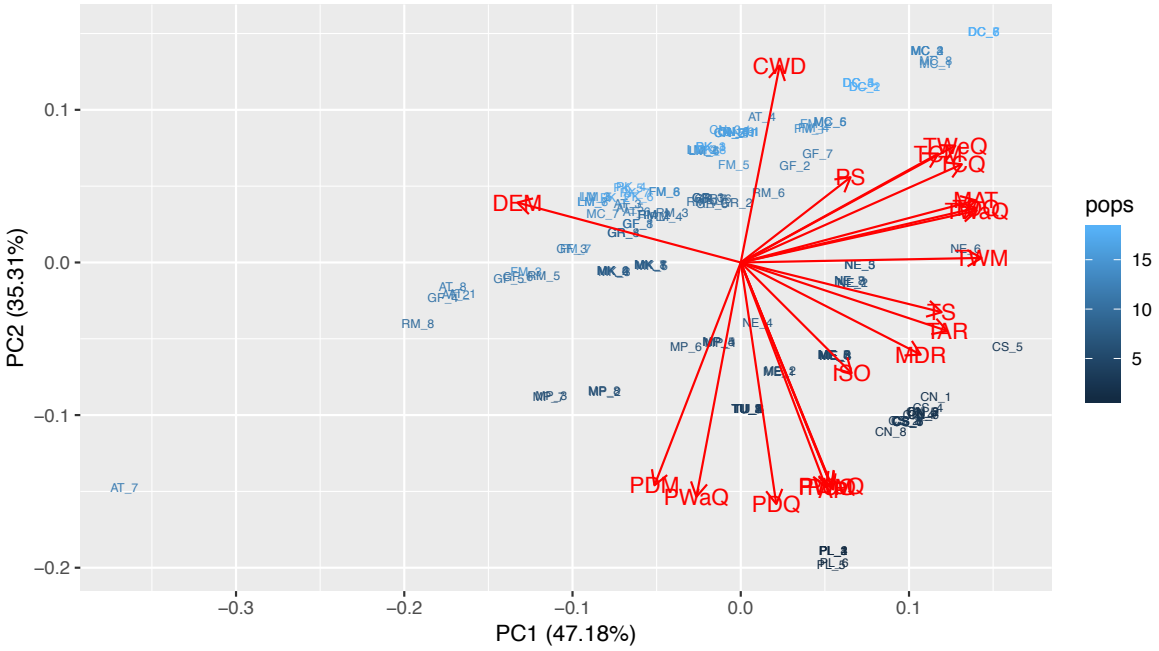
Appendix 3.1: Additional data filtering details: Number of Contigs and SNP's at each filtering step.

Filtering Step	**Number of Contigs**	**Number of SNPs**
Total Variable sites	8701	3,915,935
Removal of loci with uneven mapping quality	8333	1,344,117
Removal of unbalanced reads in heterozygotes	8329	1,300,206
Remove sites with average read depth >200	8329	1,299,461
Remove samples with read depth <5	8329	1,299,461
Remove sites with MAC <3	7557	244,489
Remove sites with >80% missing data	1133	4,276
Remove Indels	1120	4,235
Remove multiallelic sites	1118	4,208
Thinning to 1-SNP per fragment	1118	1,364

Appendix 3.2: Variable loadings and proportion of variance explained by each axis for Principal Components Analysis

Environmental Variable	Loadings on PC1 (47% of variance)	Loadings on PC2 (35% of variance)
Annual Mean Temperature	0.30533764	0.08895816
Mean Diurnal Range	0.23382154	-0.1319765
Isothermality	0.14402502	-0.1584969
Temperature Seasonality	0.2610051	-0.0711812
May Temperature of Warmest Month	0.3128064	0.00651353
Min Temperature of Coldest Month	0.25664444	0.15588655
Temperature Annual Range	0.26828944	-0.0980023
Mean Temperature of Wettest Quarter	0.27880069	0.16738419
Mean Temperature of Driest Quarter	0.30597282	0.07812717
Mean Temperature of Warmest Quarter	0.30670671	0.07428924
Mean Temperature of Coldest Quarter	0.28724379	0.1406514
Annual Precipitation	0.12018293	-0.3250514
Precipitation of Wettest Month	0.11053911	-0.3215777
Precipitation of Driest Month	-0.1120967	-0.3191191
Precipitation Seasonality	0.14263436	0.12207502
Precipitation of Wettest Quarter	0.11777354	-0.3202812
Precipitation of Driest Quarter	0.04615045	-0.3459344
Precipitation of Warmest Quarter	-0.0576154	-0.3355527
Precipitation of Coldest Quarter	0.11802076	-0.3218542
Elevation	-0.2902339	0.08573527
Climate Water Deficit	0.05050955	0.28198092

Appendix 3.3: PCA plot showing the behavior of the climate variables across PC1 and PC2. Genomic data is represented by individual ID's and blue shading represents populations, with color becoming lighter from North-to-South.



Appendix 3.4: Full AMOVA results

Source of Variation	F-stat	F-value	Std.Dev.	c.i.2.5%	c.i.97.5%	P-value
Among population	F_{ST}	0.151	0.005	0.141	0.161	0.000
Among Group	F_ct	0.023	0.002	0.018	0.027	0.000

Appendix 3.5: All outlier loci detected by BayeScan, LFMM, and RDA

Locus ID	**Detection Method**	**Associated Variable**	**F_{ST}**	**Adjusted P-value**	**Axis 1 Loading**	**Annotation**
522	BS	--	0.59	--	--	--
701	BS	--	0.72	--	--	--
801	BS	--	0.55	--	--	--
828	BS	--	0.58	--	--	Unknown mRNA
1324	BS	--	0.55	--	--	--
1123	BS, LFMM	PC1	0.66	1.72E-04	--	--
186	LFMM	PC1		7.39E-06	--	Unknown mRNA
218	LFMM	PC1		5.50E-05	--	Unknown mRNA
385	LFMM	PC1		3.07E-07	--	--
452	LFMM	PC1		5.30E-05	--	Unknown mRNA
470	LFMM	PC1		1.39E-04	--	--
604	LFMM	PC1		9.89E-05	--	NA

92	LFMM	PC2		2.99E-05	--	--
125	LFMM	PC2		8.62E-04	--	--
126	LFMM	PC2		1.59E-03	--	--
151	LFMM	PC2		5.59E-05	--	--
193	LFMM	PC2		7.66E-06	--	--
196	LFMM	PC2		9.55E-04	--	--
251	LFMM	PC2		1.91E-05	--	Unknown mRNA
253	LFMM	PC2		4.10E-04	--	--
306	LFMM	PC2		2.14E-04	--	--
319	LFMM	PC2		1.78E-03	--	--
351	LFMM	PC2		1.22E-03	--	--
518	LFMM	PC2		8.94E-05	--	--
526	LFMM	PC2		1.05E-03	--	--
547	LFMM	PC2		5.54E-04	--	--
552	LFMM	PC2		4.40E-05	--	--
562	LFMM	PC2		1.06E-03	--	--
641	LFMM	PC2		9.30E-04	--	--
653	LFMM	PC2		7.57E-04	--	--
654	LFMM	PC2		7.42E-04	--	--
690	LFMM	PC2		3.01E-05	--	--
693	LFMM	PC2		7.09E-04	--	--
695	LFMM	PC2		1.24E-03	--	--
722	LFMM	PC2		1.52E-03	--	Unknown mRNA
730	LFMM	PC2		8.95E-05	--	--
764	LFMM	PC2		5.68E-04	--	--
868	LFMM	PC2		1.07E-03	--	--
870	LFMM	PC2		9.48E-04	--	Unknown mRNA
935	LFMM	PC2		6.01E-05	--	--
1024	LFMM	PC2		7.43E-04	--	Retrotransposon
1029	LFMM	PC2		6.56E-06	--	Unknown mRNA
1054	LFMM	PC2		1.04E-03	--	--
1059	LFMM	PC2		1.04E-03	--	--
1062	LFMM	PC2		1.66E-04	--	Unknown mRNA
1076	LFMM	PC2		8.18E-04	--	--
1115	LFMM	PC2		1.41E-05	--	--
1167	LFMM	PC2		9.34E-05	--	--
1191	LFMM	PC2		2.34E-05	--	--
1214	LFMM	PC2		1.52E-03	--	magnesium transporter MRS2-4-like
1253	LFMM	PC2		2.91E-04	--	Unknown mRNA
1295	LFMM	PC2		1.43E-03	--	--
1307	LFMM	PC2		1.28E-03	--	--
1313	LFMM	PC2		5.87E-06	--	pleiotropic drug resistance protein 1-like
1332	LFMM	PC2		1.46E-03	--	--
368	LFMM, RDA	PC2, PDQ		2.35E-05	-0.211321	--
421	LFMM, RDA	PC2, PDQ		9.56E-05	-0.194232	--
471	LFMM, RDA	PC2, PDQ		2.03E-05	-0.198918	--
515	LFMM, RDA	PC2, PDQ		1.46E-03	-0.211814	--
679	LFMM, RDA	PC2, PDQ		3.51E-04	-0.223479	--
827	LFMM, RDA	PC2, PDQ		9.12E-04	-0.205785	kinesin-like protein KIN-13A mRNA
1229	LFMM, RDA	PC2, PDQ		1.65E-03	-0.178222	pollen allergen gene
1286	LFMM, RDA	PC2, PDQ		4.41E-07	0.263690	--
5	RDA	PDQ		--	-0.171565	--

90	RDA	PDQ		--	-0.167273	--
122	RDA	PDQ		--	-0.167507	-
166	RDA	PDQ		--	-0.202460	--
273	RDA	PDQ		--	-0.166244	--
338	RDA	PDQ		--	-0.187693	Wall-associated receptor kinase-like 1
494	RDA	PDQ		--	-0.163779	--
572	RDA	PDQ		--	-0.171818	--
612	RDA	PDQ		--	-0.242672	arogenate dehydratase gene
617	RDA	PDQ		--	-0.223479	Unknown mRNA
673	RDA	PDQ		--	-0.175162	--
709	RDA	PDQ		--	-0.175226	--
940	RDA	PDQ		--	0.235355	signal peptidase I AT2G30440 mRNA
1066	RDA	PDQ		--	-0.195195	Unknown mRNA
1116	RDA	PDQ		--	-0.163279	--
1305	RDA	PDQ		--	0.235376	--

Appendix 3.6: Results from partial redundancy analysis, with climate conditioned on geographic space, showing loadings of climate variables on RDA axis 1, and the relative contribution of each climate factor.

SNP ID	Loading	PDQ	ISO	CWD
5	-0.172	0.585	0.237	-0.079
90	-0.167	0.529	0.028	-0.024
122	-0.168	0.533	0.072	-0.169
166	-0.202	0.748	0.165	-0.313
273	-0.166	0.294	0.046	0.266
338	-0.188	0.584	0.258	-0.189
368	-0.211	0.607	0.321	-0.084
421	-0.194	0.822	0.257	-0.599
471	-0.199	0.633	0.163	-0.144
494	-0.164	0.768	0.294	-0.584
515	-0.212	0.696	0.265	-0.229
572	-0.172	0.470	-0.089	0.040
612	-0.243	0.690	0.252	-0.208
617	-0.171	0.713	0.278	-0.351
673	-0.175	0.603	0.158	-0.161
679	-0.223	0.710	0.635	-0.397
709	-0.175	0.619	-0.069	-0.194
827	-0.206	0.682	0.012	-0.148
940	0.235	-0.755	-0.048	0.300
1066	-0.195	0.632	0.258	-0.257
1116	-0.163	0.590	0.212	-0.205
1229	-0.178	0.429	0.219	-0.115
1286	0.264	-0.812	-0.294	0.389
1305	0.235	-0.710	-0.364	0.183

Appendix 3.7: Population information and Genetic diversity summary statistics calculated for each population after removal of putative adaptive loci.

Grove Name	Population Code	GPS Location	Sample Size	Ho	uHe	F_{IS}	G'st (average)
Placer	PLAC	39.06, -120.57	6	0.09	0.07	0.02	0.34
North Calaveras	CALN	38.28, -120.30	8	0.14	0.18	0.16	0.17
South Calaveras	CALS	38.24, -120.25	8	0.15	0.18	0.13	0.15
Tuolumne	TUOL	37.77, -119.81	8	0.16	0.16	-0.03	0.25
Merced	MERC	37.75, -119.84	8	0.16	0.17	0.00	0.21
Mariposa	MPSA	37.51, -119.60	8	0.13	0.14	0.12	0.20
Nelder	NELD	37.43, -119.59	8	0.16	0.19	0.12	0.14
McKinley	MKLY	37.03, -119.11	8	0.15	0.17	0.07	0.17
Grant	GRNT	36.75, -118.97	8	0.12	0.13	0.17	0.15
Redwood Mountain	RMNT	36.60, -118.92	8	0.12	0.11	0.16	0.13
Giant Forest	GFOR	36.57, -118.76	8	0.15	0.20	0.18	0.09
Atwell	ATWL	36.47, -118.67	8	0.17	0.21	0.08	0.11
Mcintyre	MCTR	36.13, -118.58	8	0.13	0.11	0.09	0.13
Freeman Creek	FMAN	36.14, -118.52	8	0.17	0.18	0.08	0.11
Long Meadow	LMDW	35.96, -118.60	8	0.12	0.11	0.16	0.16
Cunningham	CNHM	35.92, -118.57	9	0.16	0.17	0.00	0.16
Packsaddle	PKSD	35.93, -118.59	8	0.15	0.15	0.08	0.16
Deer Creek	DCRK	35.88, -118.61	8	0.16	0.18	0.04	0.16

CONCLUSIONS

Trends in extant genetic diversity

Within populations, spatial genetic structure is often high with significant clustering of related individuals on the landscape. Although giant sequoia is predominately an outcrossing species, both selfing and bi-parental inbreeding occur at low to moderate rates. Yet, perhaps due to selection against inbred progeny, established populations do not show strong inbreeding. Across the range, giant sequoia populations exhibit relatively low levels of genetic diversity with the smallest and most isolated populations often being the most depauperate. Groves with the lowest levels of genetic diversity include Placer, Tuolumne, Merced, Deer Creek, and Packsaddle.

We found a high degree of divergence among populations across many parts of the range. Northern and extreme southern populations are divergent from each other and the remaining groves and, thereby, represent important facets of the diversity within the species. We highlight the distinct genetic composition of the following groves, Placer, the two Calaveras groves, Tuolumne, Merced, Nelder, Mariposa, McKinley, and Deer Creek.

Populations appear to be locally adapted to gradients in water availability. Thus, the distribution of functional genetic variation will likely show correspondence with climatic gradients. Since Deer Creek, Grant, McIntyre, and Redwood Mountain are in areas with increased summer aridity, these groves represent important potential sources of drought adaptive variation.

Dispersal and gene flow

Likely due to biological constraints on dispersal capability (wingless seed and pollen lacking sacci), giant sequoia populations show a predominance of short distance dispersal. This has resulted in groves that exhibit a high degree of genetic structure at fine spatial scales and varying degrees of isolation from neighboring groves. Although many pollen dispersal kernels show evidence of 'fat-tails', indicating potential for long-distance dispersal, much of the range appears to lack direct gene flow between adjacent populations. Gene flow is low or absent in the northern and southernmost populations although some are only 2-7km away from neighboring groves. In contrast, the less fragmented central-southern range appears to maintain genetic connectivity. Yet, our estimates of average dispersal distances do not support high rates of gene flow across large spatial distances. Detailed future research can provide more clarity as to the extent to which the observed connectivity in the central-southern range is due to long-distance gene flow or a lack of time that would allow populations to diverge.

Adaptive capacity

Giant sequoia is a moisture-sensitive species whose natural distribution is becoming increasingly arid. Populations that are locally adapted to historic climate regimes will likely become mal-adapted to future conditions. Thus, adaptation to increased water stress is vital to the future viability of giant sequoia populations. Genetic diversity is the fuel for adaptive evolution.

This pool of diversity will be predominantly determined by extant diversity within groves and the potential for an influx of diversity from neighboring populations via gene flow.

Currently, genetic diversity is low in many groves and small population sizes, fine-scale spatial genetic structure, and lack of gene-flow among many groves all point to potential erosion of genetic diversity over time. Although some populations may be more drought tolerant than others, due to constraints on dispersal, the genetic variation underlying these traits is unlikely to move rapidly via gene-flow or spread to distant populations unaided. For these reasons, the adaptive capacity of giant sequoia appears to be quite limited. This void can be filled by carefully employed management that seeks to enhance adaptive capacity. An effective strategy to facilitate adaptation would involve collecting genetic resources (seed stock) from around the range, with special attention paid to capturing variation from potentially drought adapted populations, and planting seeds within current range boundaries in a way that maximizes the genetic diversity on the landscape.