University of Nevada, Reno

The roles of history, geography, and environment in shaping landscape genetic variation and its applied significance

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Ecology, Evolution, and Conservation Biology

by

Lanie M. Galland

Dr. Thomas L. Parchman/Dissertation Advisor

May, 2022

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

LANIE M. GALLAND

entitled

The roles of history, geography, and environment in shaping landscape genetic variation and its applied significance

be accepted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

Thomas L. Parchman, Ph.D.
Advisor

Mary M. Peacock, Ph.D.
Committee Member

Matthew L. Forister, Ph.D.
Committee Member

Chris R. Feldman, Ph.D.
Committee Member

Marjorie D. Matocq, Ph.D.
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

May, 2022

Abstract

The decline and loss of species and genetic diversity as a result of anthropogenic change is occurring at an unprecedented rate, reshaping biodiversity and restructuring ecosystems. Population genetic variation is shaped by evolutionary processes and in turn determines the evolutionary potential of natural populations. Facilitated by recent improvements in DNA sequencing technologies, population genomic analyses can resolve patterns of genetic differentiation and evolutionary history, characterize the effects of evolutionary processes on genome variation, and facilitate an understanding of how environmental variation may underlie local adaptation. Such analyses can inform conservation and restoration by establishing baseline patterns of genetic variation across the landscape, recognizing evolutionary significant units, sourcing propagules for restoration, and predicting species response to changing environmental conditions. Here, I applied high throughput DNA sequencing approaches to characterize the historical, spatial, and environmental factors shaping genetic variation in several systems of conservation and restoration significance. First, I investigated hierarchical genetic structure and evolutionary history of *Hucho taimen* (taimen, the world's largest salmonid), listed as vulnerable by the International Union for Conservation of Nature (IUCN), across multiple river basins in Russia and Mongolia. Second, I characterized patterns of emergent population genetic structure of nonnative *Oncorhynchus mykiss* (rainbow trout) in the Lake Tahoe basin to inform reintroduction of the U.S. Endangered Species Act listed native cutthroat trout *Oncorhynchus clarkii henshawi* (Lahontan cutthroat trout). Rainbow trout have been widely introduced across the globe, stocked for

>50 years into Lake Tahoe, and an understanding of population genetic structure may help inform strategies for successful native species reintroduction. Finally, I quantified spatial genetic structure, identified environmental variables potentially involved in local adaptation, and predicted variation in maladaptation under projected climate change across the range of *Pinus muricata*, a closed-cone pine occurring in a small number of isolated and disjunct stands along the coast of California, and also listed as vulnerable by the IUCN. Collectively, my research highlights the wide utility of population genomic analyses for taxa of conservation and restoration significance.

For my best girl.

Acknowledgements

This research would not have been possible without the immense guidance and support from friends, family, and colleagues. I am grateful for the support from the Ecology, Evolution, and Conservation Biology program, the Biology Department, the College of Science, and the University of Nevada, Reno, as a whole. I would first like to thank my advisor, Tom Parchman, for his support and mentorship, without which this research and my journey would not have been possible. I would like to thank my coadvisor Mary Peacock for her continued mentorship and friendship, and for giving me my first chance in this graduate program. I would also like to thank committee members Matt Forister, Marjorie Matocq, and Chris Feldman, each of whom have uniquely contributed to my development as an accomplished and well-rounded scientist.

I am extremely grateful for the numerous undergraduates who helped collect rainbow trout samples during my first summer, without whom I could not have completed this research. I am also immensely grateful to the diverse science team consisting of faculty, university donors, and volunteers who introduced me to taimen in Mongolia and helped in sample collection a half a world away. I would like to extend special thanks to Jeff Thompson for his encouragement and friendship, and for showing me how to dance around to land a true monster fish without falling out of the boat.

Thank you to past and present members of the Parchman laboratory. Josh Jahner, Lana Sheta, Josh Hallas, Trevor Faske, and Katie Uckele, your collaboration, guidance, and friendship will always be special to me. I would like to especially thank Agusto Luzuriaga-Neira and Trevor Faske for going beyond normal collaborative mentorship for their patience in leading me through complex bioinformatic analyses.

I am forever indebted to my wide net of family and friends. You all have cheered me on since the beginning, and you have been an endless sounding board for all my complaints, triumphs, failures, and accomplishments. And to my pup, Bailey, I am forever grateful for your endless love and positivity. You will always be my best girl.

Table of Contents

List of tables

List of figures

Introduction

Globally, species have experienced sharp declines as a result of anthropogenically induced environmental change. Ecosystem degradation and habitat destruction (Tilman *et al.*, 1994; Travis, 2003, Watson *et al.*, 2018; Bergstrom *et al*., 2021), overharvest (Russ & Alcala, 2011), population fragmentation (Lamont *et al.*, 1993; Morita & Yamamoto, 2002; Wu *et al.*, 2003; Jaeger & Fahrig, 2004), and species introductions (Bax *et al.*, 2003; Davis, 2003; Pearson *et al*., 2022) represent some of the most significant threats to biodiversity. Perhaps the greatest threat, the rapidly changing climate, has resulted in changes in phenology (Edwards & Richardson, 2004; Visser & Both, 2005; Inouye, 2008; Piao *et al.*, 2019), shifts in species' range and distribution (Colwell *et al.*, 2008; Chen *et al.*, 2011; Wallingford *et al*., 2020), and disruption of community organization (Suttle *et al.*, 2007; Walther, 2010; Glassman *et al*., 2018). Together, these disruptions to ecosystems threaten biodiversity at unprecedented scales, warranting an increase in studies to better understand the evolutionary processes underlying the production, organization, and maintenance of diversity.

Evolutionary biologists have long recognized how population genetic variation shapes the form and outcome of evolutionary processes. Rapid recent growth in our ability to generate and analyze genomic data has transformed our understanding of diverse patterns and processes in evolutionary biology and ecology (Andrews *et al.*, 2016; Porter & Hajibabaei, 2018; Breed *et al.*, 2019; Hohenlohe *et al.*, 2021). Population genetic variation provides a critical axis of our understanding of evolutionary and ecological processes in both natural and managed systems as genetic diversity underlies

evolutionary potential (Reed & Frankham, 2001; Hughes *et al.*, 2008) and has been empirically associated with population persistence (Palstra & Ruzzante, 2008) and even ecosystem function (Reynolds *et al.*, 2012; Wernberg *et al.*, 2018). The production and analysis of genetic data for rare, declining, or managed populations or ecosystems can be used to: (1) identify unique populations or lineages, (2) preserve genetic diversity within populations, (3) understand and preserve adaptive potential in the context of environmental change, and (4) guide genetically appropriate source material for translocation or restoration. Historically, conservation geneticists explored variation in threatened or endangered species with small extant populations, characterizing patterns of standing variation, identifying bottlenecks, and estimating evolutionary potential (e.g., Florida panthers, Roelke *et al.*, 1993; Isle Royale wolves, Wayne *et al.*, 1991; northern elephant seals, Bonnell & Selander, 1974). Modern applied population genetic analyses focus on understanding how genetic variation is partitioned across the landscape and the underlying causal mechanisms, identifying unique lineages, uncovering important reductions in genetic diversity, and understanding the genetic basis of adaptation. As a result, population genomic analyses could provide a critical perspective for understanding how populations may respond to global change, and thus can inform recovery and restoration strategies.

Recent DNA sequencing innovations have dramatically increased our ability to quantify genome-wide variation spanning large numbers of loci, individuals, and populations, and have transformed basic and applied research in genetics (Andrews *et al.*, 2016; Porter & Hajibabaei, 2018; Breed *et al.*, 2019; Hohenlohe *et al.*, 2021). Genome-level perspectives on population genetic variation have resolved the genetic structure of

populations and species at increasingly fine spatial and temporal scales (e.g., Novembre & Stephens, 2008; Kashtan *et al.*, 2014; Larson *et al.*, 2014; Leslie *et al.*, 2015; Shannon *et al.*, 2015) and have dramatically increased our ability to detect and characterize the influence of evolutionary processes on genome variation. For example, we are now capable of quantifying the genetic basis of phenotypes with fitness consequences in non-model organisms (Ellegren *et al.*, 2012; Jones *et al.*, 2012; Riedelsheimer *et al.*, 2012; Shapiro *et al.*, 2013; Poelstra *et al.*, 2014; Lamichhaney *et al.*, 2015), quantifying fine-scale variation in admixture and introgression across hybridizing lineages (Gompert *et al,.* 2017), anticipating species' adaptive responses to global climate change (Garrett *et al.*, 2006; Sork *et al.*, 2013; Exposito-Alonso *et al.*, 2018), and inferring how time, migration, and changes in effective population size shape population divergence (Edwards & Beerli, 2000; Hey & Nielsen, 2004; Excoffier *et al.*, 2013). Further, landscape genomic analyses can facilitate an understanding of the genetic signatures of local adaptation and its environmental causes (Rellstab *et al.*, 2015; Forester *et al.*, 2016, 2018). Collectively, these advances provide a wide variety of applied ecological and evolutionary genetic tools to guide the conservation and restoration of ecosystems.

By addressing basic evolutionary questions in species of applied significance, my dissertation research has utilized population and landscape genomic analyses to resolve complex evolutionary histories, to understand the consequences of a long history of under-informed management practices, to detect the genetic signature of local adaptation and its environmental drivers, and to assess potential maladaptation under projected climate change. I investigate these questions across two very different groups of taxa: two species of fish in Salmonidae and a rare *Pinus* species in the *Pinus* subsection Australes.

For one long-lived, native salmonid species, *Hucho taimen*, I investigated hierarchical genetic structure and the evolutionary history underlying it. For one short-lived salmonid species, *Oncorhynchus mykiss*, I characterized the population and landscape genetic structure of naturalized nonnative populations for this repeatedly and deliberately introduced species, representing one of the most successful invasives in the world. Finally, for one species of *Pinus*, I investigated the extent to which history and environment predict current patterns of standing variation while considering the consequences of these patterns for long-lived and fragmented forest tree populations threatened by climate change. Naturally, none of these taxa can escape the effects of environmental change; one is long-lived and immobile, while the others are restricted to riverscapes and the downstream flow effects of within-river anthropogenic modifications and disturbances. An understanding of the processes shaping and maintaining genetic diversity among and within populations of these taxa will be critical for understanding their potential future response to human mediated environmental change.

In Chapter One, I quantified patterns of genetic variation across basins and drainages for *Hucho taimen*. The taimen is the world's largest and one of the most ancient extant members of Salmoninae (salmon, trout, and char; Crête-Lafrenière *et al.*, 2012), historically occurring from Eastern Europe across much of Asia to the Pacific Ocean (Holčík *et al.*, 1988). The species has suffered dramatic population declines across much of its range, including extirpations in three large river basins (Rand, 2013) and is now extirpated, endangered, or threatened across much of its historic range (Ocock *et al.*, 2006). Here, I used high throughput sequencing of reduced representation libraries (ddRADseq, ~6,000 single nucleotide polymorphisms) to characterize levels of genetic

differentiation of taimen among several river drainages spanning Mongolia and Russia. I identified substantial genetic differentiation consistent with historical isolation between taimen from the Arctic (Selenge) and Pacific (Amur and Tugur) drainages. I did not detect genetic differentiation among sites in the Selenge Basin of Mongolia and Russia, which is consistent with large-scale movements of taimen and with past studies which used smaller sets of genetic markers. Among the Pacific drainages, however, I found clear differentiation between the Amur and Tugur basins. Most importantly, these analyses revealed surprising evidence for genetic differentiation between two groups of taimen sampled in the tributaries and the mainstem of the Tugur basin, despite a lack of movement barriers. Coalescent modeling suggested contemporary Tugur tributary taimen diverged in isolation, likely in the eastern Amur Basin, before paleohydrological changes allowed them to colonize the Tugur. My results are consistent with reproductive isolation between two groups of taimen, despite a lack of current geographic isolation after a period of allopatric divergence. The hierarchical population structure recovered in this study suggests that ecologically relevant genetic variation might be partitioned at smaller spatial scales than previously considered, warranting additional study and perhaps identification of distinct evolutionarily significant units with unique conservation considerations.

In Chapter Two, I characterized population genetic structure of introduced, naturalized rainbow trout (*Oncorhynchus mykiss*, RBT) to understand and inform removal strategies as ecosystem restoration proceeds for the native but locally extirpated cutthroat trout subspecies within the Lake Tahoe basin. The Lahontan cutthroat trout (*Oncorhynchus clarkii henshawi*, LCT), endemic to the hydrographic Lahontan basin of

northeastern California, northern Nevada, and southeastern Oregon, has been listed as threatened under the ESA since 1975 (Coffin & Cowan, 1995). Considered the largest inland trout in North America (Peacock *et al.*, 2018), LCT were extirpated in the Lake Tahoe basin due to overharvest, lack of access to suitable spawning habitat, water diversions, and introduction of nonnative salmonids (Gerstung, 1988; Peacock *et al.*, 2018). As with the other CT subspecies, perhaps the most formidable threat to survival and successful reintroduction of LCT into its historical habitat are naturalized populations of nonnative rainbow trout, a close congener that threatens the integrity of the native LCT genome through hybridization and introgression (Leary *et al.*, 1987; Allendorf *et al.*, 2001; Campbell *et al.*, 2002). Large, naturalized populations of RBT have been established in the Tahoe basin for >50 years (Cordone & Frantz, 1968), presenting a formidable challenge to LCT reintroduction in an ecosystem where nonnative RBT introductions were consistent, deliberate, and widespread. Here, I used high throughput sequencing of reduced representation libraries (ddRADseq, ~13,000 single nucleotide polymorphisms) to characterize population genetic differentiation and diversity within and among RBT sampled from different tributary streams. Despite dispersal from stocking locations across all regions, these analyses revealed some genetic differentiation among tributaries, with individuals from spatially proximate streams clustering across multiple population genetic analyses. I detected evidence for genetic differentiation among tributaries from the southern, western, and northern regions, including surprising structure involving a single tributary (McKinney Creek). These results illustrate the extent of differentiation within and among streams and could inform possibilities for and implications of RBT removal and LCT reintroduction. Importantly, this research

represents a unique perspective on the establishment of population genetic structure in a system where intentionally introduced but successfully naturalized populations not only represent immensely successful invasions but have also prevented successful restoration of an iconic salmonid to its native habitat.

In Chapter Three, I characterized the distribution of genetic diversity and differentiation across the entire range of *Pinus muricata* (Bishop pine) to understand how history and environment have shaped genetic variation and to anticipate variation in the degree of local maladaptation under several future climate projection models. *P. muricata* occurs across environmental conditions occupying a narrow coastal band that encompass the foggy, maritime conditions of the Pacific coast (Millar, 1986, 1988). Populations are highly isolated and disjunct (Little, 1975; Millar, 1986, 1988), unlike other widely distributed pine species. I sampled nearly all extant populations of *P. muricata* and used high throughput sequencing of reduced representation libraries (ddRADseq, ~8,000 single nucleotide polymorphisms) to quantify the distribution of genetic diversity and differentiation across the landscape. I found pronounced spatial genetic structure following a latitudinal gradient, where populations showed no evidence of reduced genetic diversity despite high levels of isolation. Phylogenetic analyses suggested older populations in the south than in the north, coinciding with the origin of the Attenuatae clade from mid-latitudes in the late Miocene between 5-10 mya followed by subsequent northward expansion (Eckert & Hall, 2006; Hernandez-Leon *et al.*, 2013; Saladín *et al.*, 2017; Gernandt *et al.*, 2018). Notably, populations from Santa Cruz Island showed fine scale genetic structure despite occurring in highly proximate stands. I then used genome-environment association (GEA) analyses to identify environmental variables (both

climate and soil) that may contribute to variation in local adaptation across the range. Using the resultant subset of climate variables, I used genetic offset approaches to quantify the relative degree of maladaptation of each population under several different climate projection models at time intervals 2041–2060 and 2081–2100. Genomic offset analyses revealed variation in the relative degree of maladaptation of populations, with southern populations generally experiencing lower levels of offset than northern populations. These results suggest that isolation and local adaptation have shaped genetic variation among disjunct populations and illustrate the consequences of this variation for *P. muricata* under projected climate change.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK. **2001**. The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution* **16**: 613–622.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. **2016**. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**: 81–92.

Bax N, Williamson A, Aguero M, Gonzalez E, Geeves W. **2003**. Marine invasive alien species: a threat to global biodiversity. *Marine Policy* **27**: 313–323.

Bergstrom DM, Wienecke BC, van den Hoff J, Hughes L, Lindenmayer DB, Ainsworth TD, Baker CM, Bland L, Bowman DMJS, Brooks ST, *et al.* 2021. Combating ecosystem collapse from the tropics to the Antarctic. *Global Change Biology* **27**: 1692–1703.

Bonnell ML, Selander RK. **1974**. Elephant seals: genetic variation and near extinction. *Science* **184**: 908–909.

Breed MF, Harrison PA, Blyth C, Byrne M, Gaget V, Gellie NJC, Groom SVC, Hodgson R, Mills JG, Prowse TAA, *et al.* 2019. The potential of genomics for restoring ecosystems and biodiversity. *Nature Reviews Genetics* **20**: 615–628.

Campbell MR, Dillon J, Powell MS. **2002**. Hybridization and introgression in a managed, native population of Yellowstone cutthroat trout: genetic detection and management implications. *Transactions of the American Fisheries Society* **131**: 364–375.

Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD. **2011**. Rapid range shifts of species associated with high levels of climate warming. *Science* **333**: 1024–1026.

Coffin PD, Cowan WF. **1995**. *Lahontan cutthroat trout (Oncorhynchus clarki henshawi) recovery plan*. US Fish and Wildlife Service, Region 1.

Colwell RK, Brehm G, Cardelús CL, Gilman AC, Longino JT. **2008**. Global warming, elevational range shifts, and lowland biotic attrition in the wet tropics. *Science* **322**: 258–261.

Cordone AJ, Frantz TC. **1968**. An evaluation of trout planting in Lake Tahoe. *California Fish and Game* **54**: 68–89.

Crête-Lafrenière A, Weir LK, Bernatchez L. **2012**. Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PloS one* **7**: e46662.

Davis MA. **2003**. Biotic globalization: does competition from introduced species threaten biodiversity? *Bioscience* **53**: 481–489.

Eckert AJ, Hall BD. **2006**. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): Phylogenetic tests of fossil-based hypotheses. *Molecular Phylogenetics and Evolution* **40**: 166–182.

Edwards SV, Beerli P. **2000**. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**: 1839–1854.

Edwards M, Richardson AJ. **2004**. Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* **430**: 881–884.

Ellegren H, Smeds L, Burri R, Olason PI, Backstrom N, Takeshi K, Kunster A, Makinen H, Nadachowska-Brzyska K, Qvarnstrom A, *et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**: 756–760.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. **2013**. Robust demographic inference from genomic and SNP data. *PLoS genetics* **9**: e1003905.

Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D. **2018**. Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nature Ecology & Evolution* **2**: 352–358.

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

Forester BR, Lasky JR, Wagner HH, Urban DL. **2018**. Comparing methods for detecting multilocus adaptation with multivariate genotype--environment associations. *Molecular Ecology* **27**: 2215–2233.

Garrett KA, Dendy SP, Frank EE, Rouse MN, Travers SE. **2006**. Climate change effects on plant disease: genomes to ecosystems. *Annu. Rev. Phytopathol.* **44**: 489–509.

Gernandt DS, Aguirre Dugua X, Vázquez-Lobo A, Willyard A, Moreno Letelier A, de la Rosa JA, Piñero D, Liston A. **2018**. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection Australes. *American Journal of Botany* **105**: 711–725.

Gerstung ER. **1988**. Status, life history, and management of the Lahontan cutthroat trout. In: American Fisheries Society Symposium. 93–106.

Glassman SI, Weihe C, Li J, Albright MBN, Looby CI, Martiny AC, Treseder KK, Allison SD, Martiny JBH. **2018**. Decomposition responses to climate depend on microbial community composition. *Proceedings of the National Academy of Sciences* **115**: 11994–11999.

Gompert Z, Mandeville EG, Buerkle CA. **2017**. Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics* **48**: 207–229.

Hernandez-Leon S, Gernandt DS, de la Rosa JA, Jardon-Barbolla L. **2013**. Phylogenetic relationships and species delimitation in *Pinus* section Trifoliae inferrred from plastid DNA. *PloS one* **8**: e70501.

Hey J, Nielsen R. **2004**. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**: 747–760.

Hohenlohe PA, Funk WC, Rajora OP. **2021**. Population genomics for wildlife conservation and management. *Molecular Ecology* **30**: 62–82.

Holčík J, Hensel K, Nieslanik J, Skacel L. **1988**. *The Eurasian huchen, Hucho hucho: largest salmon of the world*. Dr. W. Junk Publishers.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M. **2008**. Ecological consequences of genetic diversity. *Ecology Letters* **11**: 609–623.

Inouye DW. **2008**. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology* **89**: 353–362.

Jaeger JAG, Fahrig L. **2004**. Effects of road fencing on population persistence. *Conservation Biology* **18**: 1651–1657.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, *et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.

Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding H, Marttinen P, Malmstrom RR, Stocker R, *et al.* 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344**: 416–420.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, *et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**: 371–375.

Lamont BB, Klinkhamer PGL, Witkowski ETF. **1993**. Population fragmentation may reduce fertility to zero in *Banksia goodii*---a demonstration of the Allee effect. *Oecologia* **94**: 446–450.

Larson WA, Seeb LW, Everett M V, Waples RK, Templin WD, Seeb JE. **2014**. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications* **7**: 355–369.

Leary RF, Allendorf FW, Phelps SR, Knudsen KL. **1987**. Genetic divergence and identification of seven cutthroat trout subspecies and rainbow trout. *Transactions of the American Fisheries Society* **116**: 580–587.

Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe

B, Lawson DJ, *et al.* 2015. The fine-scale genetic structure of the British population. *Nature* **519**: 309.

Little EL. **1975**. *Rare and local conifers in the United States*. Washington, D.C.: US Department of Agriculture, Forest Service.

Millar CI. **1986**. The Californian closed cone pines (subsection Oocarpae Little and Critchfield): a taxonomic history and review. *Taxon* **35**: 657–670.

Millar CI. **1988**. Allozyme differentiation and biosystematics of the Californian closed-cone pines (*Pinus* subsection Oocarpae). *American Society of Plant Taxonomists* **13**: 351–370.

Morita K, Yamamoto S. **2002**. Effects of habitat fragmentation by damming on the persistence of stream-dwelling charr populations. *Conservation Biology* **16**: 1318–1323.

Novembre J, Stephens M. **2008**. Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics* **40**: 646–649.

Ocock J, Baasanjav G., Baillie J.E.M., Erdenebat M., Kottelat M., Mendsaikhan B., Smith K. **2006**. Mongolian Red List of Fishes. *Regional Red List Series* **3**.

Palstra FP, Ruzzante DE. **2008**. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology* **17**: 3428–3447.

Peacock MM, Neville H, Finger AJ. **2018**. The Lahontan Basin evolutionary lineage of cutthroat trout. *Cutthroat Trout Evolutionary Biology and Taxonomy*: 231–260.

Pearson DE, Clark TJ, Hahn PG. **2022**. Evaluating unintended consequences of intentional species introductions and eradications for improved conservation management. *Conservation Biology* **36**: e13734.

Piao S, Liu Q, Chen A, Janssens IA, Fu Y, Dai J, Liu L, Lian XU, Shen M, Zhu X. **2019**. Plant phenology and global climate change: current progresses and challenges. *Global Change Biology* **25**: 1922–1940.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, *et al.* 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**: 1410–1414.

Porter TM, Hajibabaei M. **2018**. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology* **27**: 313–338.

Rand PS. **2013**. Current global status of taimen and the need to implement aggressive conservation measures to avoid population and species-level extinction. *Archives of Polish Fisheries* **21**: 119–128.

Reed DH, Frankham R. **2001**. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* **55**: 1095–1103.

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

Reynolds LK, McGlathery KJ, Waycott M. **2012**. Genetic diversity enhances restoration success by

augmenting ecosystem services. *PloS one* **7**: e38397.

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE. **2012**. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proceedings of the National Academy of Sciences* **109**: 8872–8877.

Roelke ME, Martenson JS, O'Brien SJ. **1993**. The consequences of demographic reduction and genetic depletion in the endangered Florida panther. *Current Biology* **3**: 340–350.

Russ GR, Alcala AC. **2011**. Enhanced biodiversity beyond marine reserve boundaries: the cup spillith over. *Ecological Applications* **21**: 241–250.

Saladín B, Leslie AB, Wüest RO, Litsios G, Conti E, Salamin N, Zimmermann NE. **2017**. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC evolutionary biology* **17**: 1–15.

Shannon LM, Boyko RH, Castelhano M, Corey E, Hayward JJ, McLean C, White ME, Said MA, Anita BA, Bondjengo NI, *et al.* 2015. Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences* **112**: 13639–13644.

Shapiro MD, Kronenberg Z, Li C, Domyan ET, Pan H, Campbell M, Tan H, Huff CD, Hu H, Vickrey AI, *et al.* 2013. Genomic diversity and evolution of the head crest in the rock pigeon. *Science* **339**: 1063–1067.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. **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.

Suttle KB, Thomsen MA, Power ME. **2007**. Species interactions reverse grassland responses to changing climate. *Science* **315**: 640–642.

Tilman D, May RM, Lehman CL, Nowak MA. **1994**. Habitat destruction and the extinction debt. *Nature* **371**: 65–66.

Travis JMJ. **2003**. Climate change and habitat destruction: a deadly anthropogenic cocktail. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**: 467–473.

Visser ME, Both C. **2005**. Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences* **272**: 2561–2569.

Wallingford PD, Morelli TL, Allen JM, Beaury EM, Blumenthal DM, Bradley BA, Dukes JS, Early R, Fusco EJ, Goldberg DE, *et al.* 2020. Adjusting the lens of invasion biology to focus on the impacts of climate-driven range shifts. *Nature Climate Change* **10**: 398–405.

Walther G-R. **2010**. Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 2019–2024.

Watson JEM, Venter O, Lee J, Jones KR, Robinson JG, Possingham HP, Allan JR. **2018**. Protect the last of the wild. *Nature* **563**: 27–30.

Wayne RK, Lehman N, Girman D, Gogan PJP, Gilbert DA, Hansen K, Peterson RO, Seal US, Eisenhawer A, Mech LD, *et al.* 1991. Conservation genetics of the endangered Isle Royale gray wolf. *Conservation Biology* **5**: 41–51.

Wernberg T, Coleman MA, Bennett S, Thomsen MS, Tuya F, Kelaher BP. **2018**. Genetic diversity and kelp forest vulnerability to climatic stress. *Scientific Reports* **8**: 1–8.

Wu J, Huang J, Han X, Xie Z, Gao X. **2003**. Three-Gorges dam--experiment in habitat fragmentation? *Science* **300**: 1239–1240.

Chapter 1: Hierarchical genetic structure and implications for conservation of the world's largest salmonid, *Hucho taimen*

*Lanie M. Galland[1,2], James B. Simmons[1,2], Joshua P. Jahner[2], Agusto R. Luzuriaga-Neira[2], Matthew R. Sloat[3], Sudeep Chandra[1,2,4], Zeb Hogan[2,4], Olaf P. Jensen[5], & Thomas L. Parchman[1,2]

[1]Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, Reno, NV, USA

[2]Department of Biology, University of Nevada, Reno, Reno, NV, USA

[3]Wild Salmon Center, Portland, OR, USA

[4]Global Water Center, University of Nevada, Reno, Reno, NV, USA

[5]Center for Limnology, University of Wisconsin - Madison, Madison, WI, USA

***Author for correspondence:** Lanie M. Galland, lgalland@unr.edu

Abstract

Population genetic analyses can evaluate how evolutionary processes shape diversity and inform conservation and management of imperiled species. Taimen (*Hucho taimen*), the world's largest freshwater salmonid, is threatened, endangered, or extirpated across much of its range due to anthropogenic activity including overfishing and habitat degradation. We generated genetic data using high throughput sequencing of reduced representation libraries for taimen from multiple drainages in Mongolia and Russia. Nucleotide diversity estimates were within the range documented in other salmonids, suggesting moderate diversity despite widespread population declines. Similar to other recent studies, our analyses revealed pronounced differentiation among the Arctic (Selenge) and Pacific (Amur and Tugur) drainages, suggesting historical isolation among these systems. However, we found evidence for finer-scale structure within the Pacific drainages, including unexpected differentiation between tributaries and the mainstem of the Tugur River. Differentiation across the Amur and Tugur basins together with coalescent-based demographic modeling suggests the ancestors of Tugur tributary taimen likely diverged in the eastern Amur basin, prior to eventual colonization of the Tugur basin. Our results suggest the potential for differentiation of taimen at different geographic scales, and suggest more thorough geographic and genomic sampling may be needed to inform conservation and management of this iconic salmonid.

Key words: Amur, demographic inference, genetic diversity, *Hucho taimen*, RADseq, Salmoninae, Selenge, Tugur.

Introduction

Population genetic data is relevant for shaping conservation, restoration, and management activities, and for understanding the response of populations to environmental change. Modern high-throughput sequencing technologies have enabled genome-wide perspectives and improved our ability to quantify genetic variation across populations, including those of conservation concern[1, 2]. Reduced representation approaches such as restriction site-associated DNA sequencing (RADseq) and genotyping-by-sequencing (GBS)[3] have facilitated genome-wide population genetic analyses in organisms without genomic resources, and have often recovered patterns of fine-scale genetic structure and resolved patterns of recent diversification that were not evident with traditional molecular marker systems[4, 5]. Such approaches have improved our understanding of patterns of population structure and connectivity[6, 7], the frequency and dynamics of hybridization[8, 9], and the potential for species response to environmental change[10, 11], and have guided the identification of conservation or management units[12, 13].

Globally, many salmonid fishes have experienced sharp declines, especially in recent decades[14], due to anthropogenic factors including aquaculture[15], introduced species, habitat degradation, overfishing, and climate change[16, 17, 18]. Population genetic data have been central to understanding evolutionary history of sensitive salmonid populations[19, 20] and for guiding their conservation and management[19, 21]. High throughput sequencing in salmonids has improved the delineation of evolutionarily significant units[12] and the detection of introgression between introduced and native populations[22, 23], as well as identifying the genetic basis of adaptive phenotypes[24, 25, 26].

These perspectives, however, have mostly been limited to certain regions where substantial funding has been allocated to bolster fisheries conservation and management, such as the North American Pacific Northwest (e.g., see Ref.[27])

In contrast, salmonids occurring in sparsely populated regions of northern Asia have received comparatively little research attention. Despite hosting some of the world's most isolated aquatic ecosystems, and being among the least densely populated regions in the world, salmonid populations in this region are rapidly declining as a result of anthropogenic influences with cascading effects on riverine ecosystems[28]. Still, headwater regions across Siberia and northern Mongolia host some of the world's most pristine rivers and wetlands, receiving some of the highest ecological and chemical status ratings from the European Water Framework Directive standards (e.g., regions including the Selenge River basin headwaters in northern Mongolia[29]). Here, in the remote headwaters of river basins, species including Siberian taimen (*Hucho taimen*), lenok (*Brachymystax lenok* and *B. tumensis*), grayling (*Thymallus spp.*), and pike (*Esox lucius* and *E. reichertii*) are thought to be more abundant compared to other parts of northern Asia[30,31], as there are fewer disturbances from hatcheries, large habitat modifications to the landscape (e.g., deforestation), and dam development.

The taimen is the world's largest salmonid, and one of the most ancient extant members of subfamily Salmoninae, living 80 or more years[32, 33]. Reaching up to 2 m in length and 100 kg in weight[34], taimen reside in home ranges of large and variable size (mean 23 km, maximum up to nearly 100 km[35]) with movements over 200 km recorded for tagged individuals in northern Mongolia (O.P. Jensen, *unpublished data*), a characteristic that could limit the potential for spatial genetic differentiation. However,

taimen form pair bonds from weeks to months before moving to spawning areas (unlike other salmonids[36]), and upon reaching an acceptable spawning area, a male will aggressively attack other males within a 20 m radius of the female[37]. Thus, the apparent monogamous nature of taimen spawning behavior, as well as the timing of pair bond formation, could potentially promote spatial genetic differentiation at some scales despite large movements.

Taimen were historically found from the west slope of the Ural Mountains in Eastern Europe to the Pacific Ocean in the east, to the Arctic Circle in the north and the Gobi Desert in southern Mongolia[34]. Similar to most of the world's largest freshwater fish species[38], taimen have been negatively affected by human activity and are listed as "Vulnerable" on the IUCN Red List[39]. Anthropogenic disturbances (e.g., overfishing, pollution, mining contamination, energy development) have substantially decreased its native range, including putative extirpations of populations in the Volga, Ural, and Pechora River basins[40]. Further, large stretches of rivers in northern Mongolia have seen local extirpations associated with rapid increases in industrialization[28, 40]. Consequently, taimen are listed as threatened or endangered in Mongolia, portions of Russia, Kazakhstan, and China[28]. Though experiencing population declines across much of its historic range, several of the remaining population strongholds exist in the rivers of northern Mongolia and Siberia[31].

Understanding patterns of genetic diversity and differentiation among river systems and basins at different geographic scales will be critical for informing taimen conservation strategies. Previous research utilized mtDNA and microsatellite markers to illustrate broad phylogeographic relationships in taimen across its native range[41, 42, 43, 44,

[45]. Variation across three mtDNA regions has shown substantial divergence between populations in different basins, with one distinct clade consisting of the Amur (Pacific) and Lena (Arctic) drainages and the other consisting of the Yenesei (Arctic, specifically referring to the drainage downstream of Lake Baikal) and Khatanga (Arctic) drainages[41]. More recently, Marić et al.[45] found two distinct haplogroups representing individuals from the (1) Lena and Amur basins and the (2) Volga (Caspian Sea outflow), Ob (Arctic), Yenesei (Arctic), and Khatanga (Arctic) basins. Kaus et al.[46] used both mitochondrial and nuclear markers and found pronounced population differentiation between populations from the Amur basin (Pacific) and the Upper Yenesei (Arctic) and Selenge (Arctic, specifically referring to the drainage upstream of Lake Baikal) basins with analyses identifying two ancestral genetic clusters that the authors suggested should be considered separate evolutionary significant units (ESUs). Importantly, none of the aforementioned studies detected any evidence for genetic differentiation or isolation by distance within these large basins, perhaps suggesting that management plans should be implemented at the basin scale. While the lack of evidence for genetic structure across finer geographic scales within basins is consistent with the potential for large movement and population connectivity among river systems, it could alternatively reflect gaps in our understanding and sampling of taimen genetic variation.

Here, we used high throughput sequencing of reduced representation libraries (ddRADseq[47]) to characterize population genetic structure and diversity of taimen within and among several major river drainages in Mongolia and Russia. Specifically, we used single nucleotide polymorphism (SNP) data from multiple sampling sites within one Arctic and two Pacific drainages to (1) characterize levels of genetic

differentiation among drainages, (2) evaluate the potential for fine-scale differentiation within drainages, and (3) quantify levels of genetic diversity across sampling sites. We additionally used coalescent-based demographic modeling to infer the demographic and historical context of divergence between two groups of taimen within one Pacific drainage for which we detected unanticipated genetic differentiation at smaller spatial scales. Our results shed further light on the evolutionary history of taimen and suggest the need for more thorough geographic and genomic sampling to facilitate the development of effective strategies for conservation and management.

Methods

Sample collection, DNA sequencing, and quality filtering

We used catch and release fly fishing to obtain samples from 174 taimen using single, barbless hooks, from five sites across northern Mongolia and four sites in southeastern Russia (Fig. 1, Table 1). These sites are distributed across river systems that drain to the Arctic (Eg, Uur, Delgermörön) and those that drain to the Pacific (Amur basin: Upper and Lower Onon; Tugur basin: Konin, Munikan, Konin/Assyni Junction, and Tugur mainstem). Upon capture, small pelvic fin clip samples were taken and the fish were released. Fin clips were stored either in ethanol or dried in coin envelopes for transport. All methods were carried out in accordance with local and national regulations and guidelines (including fishing permits obtained from the local governments), and all experimental protocols were approved by the University of Nevada, Reno, Institutional Animal Care and Use Committee (IACUC protocol ID 20-10-1098). Genomic DNA was

isolated from fin clips using Qiagen DNeasy Blood and Tissue kits (Qiagen Inc., Valencia CA), and DNA concentrations were quantified using a Qiagen QIAxpert microfluidics analyzer. Due to variability in DNA yield from fin clip extractions, DNA samples were standardized to within the range of 20-40 ng/uL to ensure similar template concentrations for sequencing library preparation.

We used a reduced representation approach based on two restriction endonucleases to generate ddRADseq libraries. DNA was digested with restriction endonucleases *Mse*I (4-base recognition site) and *Eco*RI (6-base recognition site). To the *Eco*RI cut-sites, we ligated Illumina adaptors embedded with unique 8-10 bp barcode sequences which were used to tag DNA from each individual. An Illumina sequencing adaptor was ligated to the *Mse*I cut-sites. We then pooled the uniquely barcoded samples and amplified fragments using Illumina PCR primers. To reduce the portion of the genome sampled for sequencing, libraries were size selected for fragments ranging from 350-450 bp using a Pippin Prep unit (Sage Science, Beverly, MA) at the University of Texas Genome and Sequencing Analysis Center (Austin, TX). Full details on the laboratory methods used for library preparation can be found in Ref.[48] Size selected libraries were then sequenced on two lanes of an Illumina HiSeq 2500 at the University of Wisconsin-Madison Biotechnology Center's Genome Center (Madison, WI).

Raw sequencing data was filtered for contaminant sequences including *E. coli* and *PhiX*, and for Illumina sequencing adaptors, using `bowtiedb2`[49] and a pipeline of Perl and bash scripts (http://github.com/ncgr/tapioca). Importantly, barcode sequences all differed by a minimum of three bases, allowing us to detect one or two base sequencing errors within them. A custom Perl script (available at DRYAD doi: 10.5061/dryad.wstqjq2kd)

was used to correct one or two bp sequencing errors in barcode sequences, remove barcodes and cut-site associated bases, and match sequences with individual sample names. Reads were then split into fastq files specific to each individual. We then removed individuals represented by volumes of sequencing data lower than the 1st quartile of the distribution spanning all samples to reduce the fraction of missing data and to increase the number of loci retained for analyses of a sufficient number of samples per sampled locality.

Alignment, variant calling, and filtering

Since we quantified substantial genetic differentiation among samples from the two major outflows (Pacific and Arctic, see Results below), we conducted separate analyses based on: 1) all individuals (full dataset); and 2) a subset of individuals sampled from populations in the Pacific drainages (Pacific subset). We did not analyze data separately for the Arctic subset, as analysis of the full dataset (in Results) suggested limited genetic structure among populations within this drainage. All methods for alignment, variant calling, and filtering were identical for both datasets, except for the *de novo* reference generation, which was implemented separately with the different sets of samples. As there were no reference genomes available for any closely related taxa, we used a *de novo* clustering approach to build a reference of genomic regions sampled with our sequencing approach as a basis for aligning all of the reads. We used `CD-HIT`[50] to generate contig consensus sequences (partial reference hereafter) built from clustering the unique sequences in our entire dataset with a minimum match percentage of 90%. This *de novo* clustering algorithm, also utilized by the commonly used

RADseq pipeline `dDocent`[51] was used to generate a partial reference to serve as a target for subsequent read mapping. We used `bwa v0.7.5`[52] to map sequences generated for each individual to the partial reference based on an edit distance of three. Sequence alignment map (.sam) files were converted to binary alignment map (.bam) files with `samtools v1.3`[53], before `samtools v1.3` and `bcftools v1.3`[53] were used to identify and call variants across the alignments of all individuals. We calculated genotype likelihoods for SNPs at sites with a minimum base quality of 20, maximum coverage depth of 100, minimum map quality of 20, minimum site quality of 20, and minimum genotype quality of 10. We used `vcftools v0.1.14`[54] to further filter called variants. We retained only bi-allelic SNPs with minor allele frequencies (MAF) greater than 0.04, and those where at least 60% of individuals had at least one read. We randomly sampled one SNP per 100 bp contig and discarded individuals with missing data at more than 30% of loci.

As mis-assembly of reads representing paralogous regions can lead to genotyping error in high throughput sequencing data[55], we took several steps to mitigate the potential influence of such loci. First, we used `vcftools v0.1.14` to remove loci with exceptionally high coverage depth per individual, greater than or equal to 50. We then additionally identified and removed potentially paralogous loci using the `HDplot` approach described in Ref.[55]. This method identifies potential paralogs in sequence data from populations based on deviations from the expected frequency of heterozygotes and from the expected 1:1 ratio of read counts for alternate alleles in heterozygotes. Here, we retained loci with heterozygosity (*H*) levels between 0 and 0.6, and read ratio

deviance (*D*) between -18 and 18. We took these steps to exclude potentially misassembled genomic regions representing duplicate or diverged duplicate loci.

Population genetic analyses

We quantified population structure without *a priori* sample information using the Bayesian ancestry-based model `entropy v1.2`[56, 57] which is based on the correlated allele frequency model of `structure`[58]. We used `entropy` to infer the number of *k* ancestral populations represented by the data and to estimate admixture proportions (*q*) for individuals. Importantly, this model accounts for statistical uncertainty arising from sequencing and alignment error and stochastic variation in coverage depth inherent in low to medium coverage sequencing data[59, 60]. Because `entropy` provides posterior estimates of genotype probabilities at each locus for each individual, it allows for the incorporation of genotype uncertainty into downstream population genetic analyses. To seed and speed the convergence of Markov chain Monte Carlo runs, we first generated starting values for the *q* parameter. We conducted a principal component analysis (PCA) on the covariance matrix of genotype likelihoods calculated above using the `prcomp` function in `R` version 3.4[61] and then used *k*-means clustering and linear discriminant analyses (LDA) to estimate ancestry proportions for each individual for models representing $k = 2$ through $k = 9$ (or through $k = 6$ for the Pacific dataset). We ran `entropy` models for $k = 2$ through $k = 9$ (or $k = 6$) ancestral groups, with 5 replicate chains per *k*. We ran 100,000 MCMC iterations, retaining every tenth step after a burn-in of 30,000 steps. Model fit was assessed using the deviance information criterion (DIC), where lower DIC values represent better model fit. We conducted `entropy` runs

separately for (1) the set of SNP genotype likelihoods for all individuals and (2) the subset of samples representing locations within the Pacific drainages. We used genotype probabilities from entropy for the majority of the population genetic analyses described below.

We additionally characterized genetic variation with PCA using the `prcomp` function in `R`. As metrics of pairwise genetic differentiation among populations, we calculated Hudson's F_{ST}[62] and Nei's D[63] based on population allele frequencies. As metrics of genetic diversity for each sampling location, we calculated nucleotide diversity (θ_π, or the average number of pairwise differences between sequences[64]), Watterson's theta (θ_W, or the number of segregating sites[65]), and the scaled difference between the two (Tajima's *D*[66]) using methods that incorporate genotype uncertainty implemented in `ANGSD`[67, 68]. We used the *de novo* artificial reference genome and individual .bam files to estimate site allele frequency likelihoods using the "doSaf 1" tool. We then used site allele frequency likelihoods as input for REALSFS[68] to generate folded site frequency spectrum likelihoods. Using "doSaf 1," we calculated posterior allele frequency probabilities. Lastly, we used the `thetastat` utility from `ANGSD` to estimate per-locus measures of each diversity metric and generated the per-population average of these values over all contigs and nucleotides. For comparison with other studies, we also calculated expected heterozygosity based on allele frequencies.

As we found unexpected evidence for divergence between taimen in the Tugur mainstem and its tributaries, we used a coalescent-based demographic modeling approach to explore parameters characterizing the divergence and demographic history of these two populations. The site frequency spectrum (SFS) contains the signatures of

divergence and demographic processes (e.g., time, migration, changes in effective population size), and high throughput sequencing data has substantially improved our ability to estimate such parameters from population genetic data[69, 70]. Before generating the SFS, we further filtered the vcf file generated above. First, we removed variants with MAF < 0.1 to guard against rare variants that could represent sequencing errors and trimmed outlier loci with $F_{ST} > 0.15$ (the 0.95 quantile of the F_{ST} distribution) between the two populations. We generated the unfolded SFS for each population using `easySFS` (https://github.com/isaacovercast/easySFS#easysfs) on the stringently filtered .vcf file, down sampling populations to sizes of 10 and 10 (--proj 10, 10).

Using the unfolded SFS, we estimated demographic parameters for eight different models using coalescent simulation and a maximum likelihood framework in `fastsimcoal2`[69]. These models represented two-population divergence (Tugur mainstem and Tugur tributaries), with and without migration, and with and without population expansion or contraction. We ran 50,000 coalescent simulations per replicate and a total of 50 replicates, with minimum (-n) of 100,000 simulations for a total (-L) of 40 cycles. We used a mutation rate for salmonids (Salmoninae) of 8e-9 bp per generation for model estimation (coho salmon[71]), and to estimate coalescent effective population size. For each model, the replicate with the smallest difference between the maximum expected likelihood (MEL) and the maximum observed likelihood (MOL) represented the best-fit run[67]. To account for differences in the number of parameters included in each model, we calculated AIC scores for each model's best-fit run. We then calculated ΔAIC to compare models.

Following parameter estimation for the best fit model, we calculated 95% confidence intervals for each parameter. Using the maximum likelihood parameters from the *_maxL.par file, we generated 100 bootstrap replicates of the SFS for each model. Next, we estimated parameters of these 100 SFS using the same 50-replicate analyses described above. The best-fit model parameter estimates for each of the new 100 simulated SFS were used to calculate mean parameter estimates and subsequently to infer 95% confidence intervals.

Results

Full dataset

After filtering for contaminants and removing individuals lacking sufficient sequencing data, we retained 174 individuals with a mean number of 1,898,867 reads per individual. `bwa` mapped reads from all individuals onto the *de novo* partial reference consisting of 221,178 genomic regions. After variant calling and filtering based on sequencing coverage and quality, we retained 7,597 loci with MAF > 0.04. We additionally discarded 1,551 SNPs that potentially represented paralogous regions, leaving a final set of 6,046 SNPs from 174 individuals (mean coverage = 10.1X per locus per individual). Sequence data is available on NCBI's Short Read Archive (accession PRJNA745962; https://dataview.ncbi.nlm.nih.gov/object/ PRJNA745962). Both the sequence data and the vcf file are available at DRYAD (doi: 10.5061/dryad.wstqjq2kd).

Pronounced genetic differentiation was evident between taimen sampled from the Arctic and Pacific drainages across all analyses (Figs. 2A, 3). DIC values from

`entropy` indicated that the $k = 2$ model best fit the data, though the $k = 3$ model had similar support and illustrated additional population differentiation (Fig. 2B, Supplementary Table S1). Individuals were assigned nearly 100% to one of two ancestries (Fig. 2A): the Arctic with the single Selenge drainage (Eg, Uur, and Delgermörön Rivers), and the Pacific with the Amur (Upper and Lower Onon River sites) and the Tugur (Konin, Munikan, Konin/Assyni Junction, and Tugur Rivers) drainages (Fig. 1). The $k = 3$ model reflected the same pattern for the Arctic versus Pacific drainages, but individuals were assigned with variable ancestry across several Pacific sites exhibiting differentiation (Fig. 2B).

PCA of the genotype probabilities revealed genetic structure similar to ancestry estimates from `entropy` (Fig. 3A). PC 1 explained 59.99% of variation in the data, separating individuals and populations from the Arctic and Pacific drainages. PC 2 explained only 0.98% of variation in the data, but suggested finer-scale structure within the Pacific drainages. Pairwise measures of genetic divergence among sites from the two outflows (Arctic and Pacific) were high (F_{ST} range = 0.209 - 0.258, Nei's *D* range = 0.1405– 0.1823; Supplementary Table S2) consistent with independent evolutionary histories and substantial isolation of drainages (Table 2). A neighbor joining tree based on pairwise estimates of Nei's D among sampling localities provided and alternative visualization of hierarchical genetic structure consistent with all other population genetic analyses (Fig. 3D).

Genetic diversity varied across the sampled geographic regions. No populations from the Arctic drainage had confidence intervals that overlapped with those in the Pacific drainages. Mean nucleotide diversity across all nine sampling sites was 0.00164

for θ_π (range 0.00119 - 0.00209) and 0.00132 for θ_W (range 0.00077 - 0.00178). Nucleotide diversity was consistently higher in Pacific drainages (mean θ_π = 0.00181, range 0.00165 - 0.00209; mean θ_W = 0.00152, range 0.00119 - 0.00178) than in the Arctic drainage (mean θ_π = 0.00130, range = 0.00119 - 0.00138; mean θ_W = 0.00093, range = 0.00077 - 0.00101), as was H_E (Arctic mean = 0.1195, range = 0.118 – 0.121; Pacific mean = 0.2203, range = 0.213 – 0.230). Importantly, for taimen localities sampled in the current work, sample size was unrelated to both genetic diversity metrics (sample size vs. heterozygosity: r = -0.327, P – 0.323; sample size vs. nucleotide diversity: r = -0.424, P = 0.253; Supplementary Fig. S1).

Analyses of the Pacific drainages

Given pronounced genetic differentiation between the Arctic and Pacific drainages and evidence for finer-scale differentiation among river systems within the Pacific drainages, we conducted assembly and variant calling separately for Pacific populations. Unique reads were assembled into a partial reference consisting of 160,858 contigs. Following reference mapping, variant calling, subsequent bioinformatic processing, and paralog filtering, we retained 3,961 SNPs from 83 individuals (mean coverage = 10.4X per locus per individual) for the Pacific subset. The vcf file associated with the Pacific subset is available at DRYAD (doi: 10.5061/dryad.wstqjq2kd).

Although pairwise measures of genetic differentiation among sampling sites across this drainage were relatively low (Supplementary Table S2), analyses illustrated finer-scale patterns of differentiation that were less evident in analyses of the full data set.

The $k = 2$ entropy model fit the data best, although the $k = 3$ model had similar DIC support (Supplementary Table S1), and further illustrated population structure. Across analyses, clear genetic differentiation was evident between the Amur (Upper and Lower Onon sites) and Tugur drainages (Figs. 2C, D, 3B). Both the $k = 2$ and $k = 3$ entropy models assigned individuals from the Amur and Tugur drainages to different ancestral clusters (Fig. 2C, D), and PCA clearly separated individuals from the two drainages (Fig. 3C).

There was unexpected genetic differentiation among sampling sites within the Tugur basin; taimen from the headwater tributaries (KO, MO, AJ; n = 37) were differentiated from samples taken in the mainstem Tugur (TU; n = 43; Fig. 1B). Ancestry was assigned differentially to separate clusters in both the $k = 2$ and $k = 3$ entropy models (Fig. 2C, D), and the tributary populations formed a non-overlapping cluster in PC space intermediate between samples from the Tugur and the Amur drainages (Fig. 3C). Although overall genetic differentiation was subtle (F_{ST} range = 0.012 - 0.029; Supplementary Table S2), these analyses nonetheless indicate the presence of finer-scale differentiation among sampling sites in the Tugur than in the other drainages we examined.

After the more stringent filtering with `vcftools v0.1.14`, we retained 1,971 variants from which we constructed the unfolded SFS. Coalescent simulations using `fastsimcoal2` were run for eight models spanning variation in divergence and demography of the Tugur mainstem and Tugur tributary populations (see Table 3 for parameter estimates and model comparison metrics, and Fig. 4 for model schematics). Importantly, all models including migration had substantially better fits than the model

without migration. Model likelihoods were similar across models including migration, but the best fit model included initial divergence, after which coalescent N_e remained constant in the tributary population but contracted in the mainstem population (Table 3; Fig. 5). Parameter estimates indicated divergence at ≈ 28k generations ago based on a mutation rate of 8e-9 (mean ≈ 393 kya based on mean generation time of 13.8 years [range ≈ 195 - 594 kya based on generation time range 6.9 - 20.8 years (mean ± 2 standard deviations)][34, 72]; Fig. 5). Generation time was calculated from age-frequency data from the Tugur River (M.R. Sloat, *unpublished data*) and fecundity-at-age data presented by Ref.[34]. Following the ancestral divergence event, asymmetrical gene flow was inferred between mainstem and tributary clusters, with greater gene flow from the mainstem to tributaries. Coalescent N_e for tributaries was substantially lower than that of the mainstem (≈ 10,000 versus ≈ 41,000), and the model indicated population contraction in the latter.

Discussion

We documented hierarchical patterns of genetic differentiation among taimen sampled across multiple drainages. We found pronounced divergence among taimen from the Arctic and Pacific drainages, but also recovered more subtle patterns of differentiation within and among river systems that drain to the Pacific. Although our spatial sampling of the distribution was limited, our analyses indicate the potential for genetic differentiation at finer scales within basins and even within specific river systems. This aspect of our results differs from several past studies and was potentially influenced by

different spatial and more thorough genomic sampling. Below, we discuss our results in the context of past genetic work on taimen, with consideration of how current and future population genetic analyses could inform their management and conservation.

Deep genetic divergence between Arctic and Pacific drainages

Our population genetic analyses illustrate substantial genetic differentiation consistent with significant historical isolation between taimen from the Arctic and Pacific drainages (mean F_{ST} = 0.24; Fig. 1, Supplementary Table S2), similar to past studies based on smaller sets of genetic markers[41, 45, 46]. Additionally, maximum likelihood phylogenetic analyses (RAxML[73]) based on 1,229 SNPs in a concatenated alignment similarly illustrated deep divergence among taimen sampled from the Pacific and Arctic drainages (Supplementary Methods, Supplementary Fig. S2). Genetic differentiation across systems would be expected given the geologic barriers separating these drainages. Though the Amur and Selenge basins are separated by only a few kilometers (Fig. 1), the Khentii Mountains arising from the Baikal Rift system (last active during the Pliocene[74]) form a continental divide that likely severed any recent connectivity between the Arctic and Pacific river systems. This geologic feature is associated with deep phylogeographic breaks for other taxa occurring in this region, including other salmonids (*Thymallus*[75, 76], *Brachymystax*[77, 46]), as well as other groups of freshwater fishes (*Cottus*[78], *Esox*[79]). Divergence time estimates from mtDNA analyses indicate divergence across this divide in the range of 1 - 2.5 mya for *Cottus*[78] and 0.5 - 2 mya for taimen[41, 45].

Taimen from the Arctic and Pacific outflows do differ in a few key morphological traits, including mean relative mass, and density and size of spot patterns (Mikhail Skopets and M.R. Sloat, *unpublished data.*). This variation is consistent with long-term isolation and independent evolution. Variation could possibly be due in part to ecological differences among drainages on either side of this divide, such as riverscape and landscape topography[35, 72], constituent riparian species composition[35], and marine food web subsidies from seasonal spawning runs of chum salmon in the Tugur basin, but not the Selenge basin[72]. Unfortunately, the geographically sparse and clumped nature of our sampling led to a strong correlation between geographic and environmental distances among sampling localities, which precluded formal tests of environmental influences on spatial genetic structure.

We did not detect genetic differentiation among sites within the Arctic drainage (Figs. 2, 3), though only three locations, all within the Selenge basin, were sampled in this drainage. Future studies could benefit from additional samplings across the Arctic drainage, particularly upstream and downstream of Lake Baikal, which may have presented a migration barrier to the predominantly riverine taimen. This apparent absence of genetic structure is consistent with past studies based on microsatellite and mtDNA data[45, 46], and with the possibility of population connectivity among sites separated by hundreds of kilometers. Limited genetic differentiation across broad spatial scales would not be surprising given the large size, long lifespan, and substantial opportunity for movement in taimen[35]. Moreover, the region's glacial history suggests that populations in the Arctic are likely younger than those in the Pacific drainages. Northwestern Siberia experienced repeated Pleistocene glaciations which blocked

north-flowing rivers and formed ice dam lakes[80, 81, 82], whereas the more southern Pacific drainages are thought to have been less influenced by glaciation[83] (but see Ref.[80]). A younger taimen lineage in the Arctic drainage would be consistent with limited differentiation among populations often seen in Arctic fishes[84, 85], and also with variation in genetic diversity among the regions we sampled.

For all metrics, levels of genetic diversity were consistently lower for the three Arctic drainage sites than those from Pacific drainage sites (Table 2). Measures of nucleotide diversity were relatively low, but well within the range of published nucleotide diversity estimates across salmonids[86, 87, 88, 89]. For example, our values for taimen were similar to those reported for Atlantic salmon (*Salmo salar*, overall nucleotide diversity of 3.99e-4; Ref.[90]), a species that has also been strongly influenced by recent glacial periods[90]. Lower diversity in the Arctic drainage is consistent with the hypothesis that Arctic populations are younger, and/or suffered recent bottlenecks during Pleistocene glacial activity in this region[82].

Fine-scale genetic structure within the Pacific drainages

Our analyses revealed clear, though less pronounced, genetic differentiation between the two Pacific drainages (Amur and Tugur). Taimen from the Onon River (Amur basin) were differentiated from those occurring 1,800-2,000 km to the east in the Tugur basin (mean F_{ST} = 0.033). This pattern could be consistent with the Yablonovy and Stanovoi Mountain ranges acting as a divide for aquatic fauna separating the headwaters of the Lena (just north of the Tugur) and Amur basins[91], although a lack of sampling over a large area between the Onon and Tugur sites limits our understanding

of geographic features associated with spatial genetic structure. As in the Arctic drainage, we did not detect evidence for differentiation among sampling sites within the Amur basin. However, given the size and complexity of the Amur basin combined with the population structure documented across the basin in other salmonids (e.g., *Thymallus*[76]), it is possible that more thorough sampling could reveal additional structure.

Our analyses revealed unexpected evidence for fine-scale genetic differentiation between two groups of taimen sampled in the Tugur basin. Though sampling locations within this basin are highly proximate (less than 10 km, Fig. 1) and well within average home ranges of taimen[35, 92], we observed distinct genetic differentiation between the Tugur River mainstem and its tributaries (Figs. 1B, 2C, D, and 3C). Differentiation among these groups was subtle but clear (Supplementary Table S2), as individuals from the mainstem and tributary groups were completely identifiable and formed non-overlapping groups in PCA and ancestry-based analyses (Figs. 2D, and 3C). No evident barriers to movement (and thus spawning) have been observed in this region, and taimen from these sites are not known to differ morphologically. Nonetheless, the pattern of consistent differentiation among these sites indicates that some barrier to gene flow likely exists. Ecological variation among tributaries and the mainstem has been noted, including differences in water flow, water level, and chemical composition, as well as the timing of food availability (e.g., variation in chum salmon spawning[93]), suggesting ecological factors may underlie isolation. Reproductive isolation has evolved within a number of salmonid species due to spatiotemporal differences in spawning (spring versus fall Chinook salmon[26]), or morphological specialization (dwarf

and normal *Coregonus clupeaformis*[94]; benthic and limnetic *Salvinus alpinus*[95]; pelagic and littoral feeding *Thymallus nigrescens*[96]). Additionally, although taimen populations typically do not differentiate at smaller spatial scales, anadromous Sakhalin taimen (*Parahucho perryi*) demonstrate population genetic structure arising from differences in spawning grounds and homing behavior[97]. While we are unaware of such mechanisms underlying divergence in taimen, our results indicate the potential for differentiation at smaller geographic scales than previously detected and the need for further study to understand its evolutionary causes and consequences for management.

Given the unexpected differentiation within the Tugur basin, we compared empirical and simulated SFS for models representing different divergence scenarios to consider the timing and demographic context of this divergence. All seven of the models incorporating migration had strongly improved fit compared to the model without migration (Table 3), consistent with a history of allopatric divergence followed by secondary contact and gene flow. The best fit model included bidirectional gene flow, with population size constant in the Tugur tributaries and contracting in the mainstem. The divergence time estimate for this model of ~28k generations would correspond to divergence at 195 - 594 kya, depending on generation time estimates. The estimated coalescent N_e was substantially larger in the Tugur mainstem than the tributaries, while migration probabilities were higher from the tributaries into the mainstem. It is worth noting that parameter estimates for these models can be affected by mutation rate, changes in migration and population size over time, and bioinformatic processing of sequence data[98, 99]. Furthermore, the additional models including migration had similar likelihoods to the best fit model yet very different parameter

estimates (Table 3). Importantly, coalescent N_e estimates can be heavily affected by changes in these modeling input values, as seen in our substantially higher coalescent N_e for the best fit model than empirical N_c estimates for taimen[72]. Denser data and sampling would improve our ability to evaluate such models and characterize this history.

A possible, if not likely, scenario underlying divergence in the Tugur basin is that ancestral tributary taimen diverged in isolation in the eastern Amur, before paleohydrological connections allowed colonization of the Tugur basin. Consistent with this hypothesis, taimen in the Tugur tributaries are intermediate with respect to those from the Tugur mainstem and the western Amur basin samples in both PCA and ancestry-based analyses (Figs. 2 and 3). Although the Tugur and Amur basins have no contemporary hydrologic connectivity and the hydrologic history of this region is not well understood, faunal similarities are consistent with paleohydrological connections among these drainages during the Pleistocene. For example, mtDNA haplotype sharing among the Amur and Tugur basins has been documented in blunt-nosed lenok (*Brachymystax lenok*), suggesting a history of such connectivity[77]. The location of a potential paleo connection between the Tugur and Nimelen, a lower Amur tributary, is apparent in a low-relief area where active channels in tributaries of the two rivers are separated by <1 km and by a drainage divide of <10 m. One hypothesized scenario is that the Tugur was forced south and joined the lower Amur during Pleistocene "back-arc glaciation" of a Sea of Okhotsk marine ice sheet and the Stanovoi glacier complex to the west before drainages reorganized and became independent during glacier retreat[80, 100]. A similar pattern of glaciation has been hypothesized to generate

paleohydrological connections between the proto Tugur drainage and mid-Amur drainages of the Bureya and Zeya Rivers (see Fig. 6 of Ref.[100]).

Understanding the historical and geographic context of differentiation in the Tugur basin as well as the mechanisms potentially maintaining this genetic variation will require further study. Our results are based on relatively sparse spatial sampling only in the western Amur basin. More thorough sampling of the eastern Amur basin, especially where it nears the Tugur, will be necessary to more thoroughly characterize the geographical variation and the origin of divergence among taimen in the Tugur basin. Similarly, additional sampling within the Tugur basin could improve understanding of the spatial distribution of the genetically differentiated groups detected here. As importantly, an understanding of ecological, morphological, and life history variation among taimen in this system will be critical for understanding potential mechanisms underlying and maintaining differentiation. Further work here is warranted as the Tugur River continues to support the largest individuals of the world's largest salmonid[101], and the majority of the watershed is protected within the Tugursky Nature Reserve, a regional zakaznik (equivalent to an IUCN Category IV protected area).

Implications for conservation and management

Given substantial and widespread declines of taimen populations, an understanding of genetic diversity and spatial genetic structure could be critical for informing taimen conservation. The demarcation of taimen management units could become important as anthropogenic influences known to impact fish populations (e.g., see Ref.[102, 103]) increase

in frequency and intensity, and as translocation efforts are considered for regions where populations have declined or have been extirpated. Genetic diversity is often considered a key parameter for conservation and restoration, as it is commonly viewed as a proxy for population resilience and evolutionary potential[104, 105]. Nucleotide diversity estimates for all of our sampling sites were in the range of estimates published for other salmonids[86, 87, 88, 89], and do not reflect severely reduced diversity. Although, Tajima's *D* estimates were slightly positive for the Amur basin, which could be consistent with population declines here. However, our sampling was limited to regions with healthy river systems and those that have yet to exhibit strong population declines, and may not represent standing variation in other areas of the distribution.

Evolutionarily significant units (ESUs[106, 107]) are often used to designate lineages of conservation importance. Kaus et al.[46] proposed two taimen ESUs across northern Mongolia that largely correspond to our Selenge (Arctic) and Amur (Pacific) basin sites. Our results indicating the presence of substantial genetic divergence between these basins lend support for two separate units that might require separate conservation and management strategies. The hierarchical population structure recovered in this study suggests that ecologically relevant genetic variation might be partitioned at smaller spatial scales than previously considered, and we thus caution against the translocation of taimen among geographically and ecologically distinct populations before more thorough genetic sampling can be completed. The differentiated groups we detected in the Tugur basin could reflect ecological and historical variation warranting unique conservation consideration, though further study is clearly needed. Indeed, the additional evidence for population structure within the

Pacific drainages suggests that our understanding of taimen genetic structure across different riverscapes is likely limited by both the geographic and genomic extent of sampling. Due to difficulty in sampling via fly fishing, our sampling was opportunistic, geographically sparse, and less than ideal for quantifying how environmental and hydrological variation may influence spatial genetic structure across the distribution. Future studies with more comprehensive sampling within and across basins could be essential for developing a finer scale understanding of the factors influencing evolutionary history of taimen across Siberia and the rest of its range.

Data availability

The datasets generated for this study are available at the Dryad Digital Repository (doi: 10.5061/dryad.wstqjq2kd; https://doi.org/10.5061/dryad.wstqjq2kd) and NCBI's Short Read Archive (accession PRJNA745962; https://dataview.ncbi.nlm.nih.gov/object/PRJNA745962).

References

1. Avise, J. C. Perspective: conservation genetics enters the genomics era. *Conserv. Genet*. **11,** 665–669 (2010).

2. Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K. & Allendorf, F. W. Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conserv. Genet*. **11,** 355–373 (2010).

3. Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G. & Hohenlohe, P. A. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet*. **17,** 81–92 (2016).

4. Wagner, C. E. et al. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Mol. Ecol*. **22,** 787–798 (2013).

5. Vendrami, D. L. J. et al. RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. *Roy. Soc. Open Sci.* **4,** 160548 (2017).

6. Xuereb, A. et al. Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Mol. Ecol*. **27,** 2347–2364 (2018).

7. DeSaix, M. G. et al. Population assignment reveals low migratory connectivity in a weakly structured songbird. *Mol. Ecol.* **28,** 2122–2135 (2019).

8. Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W. & Luikart, G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Mol. Ecol. Resour*. **11,** 117–122 (2011).

9. Mandeville, E. G. et al. Inconsistent reproductive isolation revealed by interactions between *Catostomus* fish species. *Evol. Lett.* **1,** 255–268 (2017).

10. Bay, R. A. et al. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359,** 83–86 (2018).

11. Razgour, O. et al. Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. *Proc. Natl. Acad. Sci*. **116,** 10418–10423 (2019).

12. Larson, W. A. et al. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evol. Appl*. **7,** 355–369 (2014).

13. Bernos, T. A., Jeffries, K. M. & Mandrak, N. E. Linking genomics and fish

conservation decision making: a review. *Rev. Fish Biol. Fish.* **30,** 587–604 (2020).

14. Montgomery, D. R. *King of fish: the thousand-year run of salmon*. (Westview Press, 2003).

15. Costello, M. J. How sea lice from salmon farms may cause wild salmonid declines in Europe and North America and be a threat to fishes elsewhere. *Proc. R. Soc. B.* **276,** 3385–3394 (2009).

16. Sepulveda, A. J., Rutz, D. S., Dupuis, A. W., Shields, P. A. & Dunker, K. J. Introduced northern pike consumption of salmonids in Southcentral Alaska. *Ecol. Freshw. Fish* **24,** 519–531 (2015).

17. Otero, J. et al. Basin-scale phenology and effects of climate variability on global timing of initial seaward migration of Atlantic salmon (*Salmo salar*). *Glob. Chang. Biol*. **20,** 61–75 (2014).

18. Clavero, M. et al. Historical citizen science to understand and predict climate-driven trout decline. *Proc. R. Soc. B.* **284,** 20161979 (2017).

19. Utter, F. Population genetics, conservation and evolution in salmonids and other widely cultured fishes: some perspectives over six decades. *J. Fish Biol*. **65,** 323–324 (2004).

20. Bernatchez, L. et al. Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol*. **32,** 665–680 (2017).

21. Ryman, N. Conservation genetics considerations in fishery management. *J. Fish Biol*. **39,** 211–224 (1991).

22. Hohenlohe, P. A. et al. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Mol. Ecol*. **22,** 3002–3013 (2013).

23. Bay, R. A., Taylor, E. B. & Schluter, D. Parallel introgression and selection on introduced alleles in a native species. *Mol. Ecol*. **28,** 2802–2813 (2019).

24. Brieuc, M. S. O., Ono, K., Drinan, D. P. & Naish, K. A. Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol*. **24,** 2729–2746 (2015).

25. Hecht, B. C., Matala, A. P., Hess, J. E. & Narum, S. R. Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Mol. Ecol*. **24,** 5573–5595 (2015).

26. Prince, D. J. et al. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci. Adv*. **3,** e1603198 (2017).

27. Habicht, C. et al. *Harvest and harvest rates of sockeye salmon stocks in fisheries of the Western Alaska Salmon Stock Identification Program (WASSIP), 2006-2008.* Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Services, 2012.

28. Ocock, J. et al. Mongolian Red List of Fishes. Reg. Red List Ser. 3, (2006).

29. Hofmann, J. et al. Initial characterization and water quality assessment of stream landscapes in northern Mongolia. *Water* **7,** 3166–3205 (2015).

30. Mercado-Silva, N. et al. Fish community composition and habitat use in the Eg-Uur River system, Mongolia. *Mong. J. Biol. Sci*. **6,** 21–30 (2008).

31. Jensen, O. P. et al. Evaluating recreational fisheries for an endangered species: a case study of taimen, *Hucho taimen*, in Mongolia. *Can. J. Fish. Aquat. Sci*. **66,** 1707–1718 (2009).

32. Crête-Lafrenière, A., Weir, L. K. & Bernatchez, L. Framing the Salmonidae family phylogenetic portrait: a more complete picture from increased taxon sampling. *PLoS One* **7,** e46662 (2012).

33. Shedko, S. V, Miroshnichenko, I. L. & Nemkova, G. A. Phylogeny of salmonids (Salmoniformes: Salmonidae) and its molecular dating: Analysis of nuclear RAG1 gene. *Russ. J. Genet*. **48,** 575–579 (2012).

34. Holčík, J., Hensel, K., Nieslanik, J. & Skacel, L. *The Eurasian huchen,* Hucho hucho*: Largest Salmon of the World*. (vol. 5) (Dr. W. Junk Publishers, 1988).

35. Gilroy, D. J. et al. Home range and seasonal movement of taimen, *Hucho taimen*, in Mongolia. *Ecol. Freshw. Fish* **19,** 545–554 (2010).

36. Esteve, M. Observations of spawning behaviour in Salmoninae: *Salmo*, *Oncorhynchus* and *Salvelinus*. *Rev. Fish Biol. Fish*. **15,** 1–21 (2005).

37. Esteve, M., Gilroy, D. & McLennan, D. A. Spawning behaviour of taimen (*Hucho taimen*) from the Uur River, Northern Mongolia. *Environ. Biol. fishes* **84,** 185–189 (2009).

38. He, F. et al. The global decline of freshwater megafauna. *Glob. Chang. Biol.* **25,** 3883–3892 (2019).

39. Hogan Z. & Jensen, O. *Hucho taimen. The IUCN Red List of Threatened Species* https://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T188631A22605180.en (2013).

40. Rand, P. S. Current global status of taimen and the need to implement aggressive conservation measures to avoid population and species-level extinction. *Arch. Polish Fish*. **21,** 119–128 (2013).

41. Froufe, E., Alekseyev, S., Knizhin, I. & Weiss, S. Comparative mtDNA sequence

(control region, ATPase 6 and NADH-1) divergence in *Hucho taimen* (Pallas) across four Siberian river basins. *J. Fish Biol*. **67,** 1040–1053 (2005).

42. Tong, G., Kuang, Y., Yin, J., Liang, L. & Sun, X. Isolation of microsatellite DNA and analysis on genetic diversity of endangered fish, *Hucho taimen* (Pallas). *Mol. Ecol*. *Notes* **6,** 1099–1101 (2006).

43. Kuang, Y., Tong, G., Xu, W., Sun, X. & Yin, J. Analysis on population genetic structure of taimen (*Hucho taimen*) in the Heilongjiang River [J]. *J. Fish. Sci*. China **6,** (2010).

44. Balakirev, E. S., Romanov, N. S., Mikheev, P. B. & Ayala, F. J. Mitochondrial DNA variation and introgression in Siberian taimen *Hucho taimen*. *PLoS One* **8,** e71147 (2013).

45. Marić, S. et al. First mtDNA sequencing of Volga and Ob basin taimen *Hucho taimen*: European populations stem from a late Pleistocene expansion of *H. taimen* out of western Siberia and are not intermediate to *Hucho hucho*. *J. Fish Biol*. **85,** 530–539 (2014).

46. Kaus, A. et al. Fish conservation in the land of steppe and sky: evolutionarily significant units of threatened salmonid species in Mongolia mirror major river basins. *Ecol. Evol*. **9,** 3415–3433 (2019).

47. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7,** e37135 (2012).

48. Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F., Benkman, C.W. & Buerkle, C.A. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol*. **21**, 2991–3005 (2012).

49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357 (2012).

50. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28,** 3150–3152 (2012).

51. Puritz, J.B., Hollenbeck, C.M. & Gold, J.R. *dDocent*; a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431 (2014).

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

53. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

54. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27,** 2156–

2158 (2011).

55. McKinney, G. J., Waples, R. K., Seeb, L. W. & Seeb, J. E. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Mol. Ecol.* Resour. **17,** 656–669 (2017).

56. Gompert, Z. et al. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Mol. Ecol*. **23,** 4555–4573 (2014).

57. Shastry, V. et al. Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. *bioRxiv* (2020).

58. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164,** 1567–1587 (2003).

59. Buerkle, C. A. & Gompert, Z. Population genomics based on low coverage sequencing: how low should we go? *Mol. Ecol*. **22,** 3028–3035 (2013).

60. Fumagalli, M. et al. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* **195,** 979–992 (2013).

61. R Core Team. R: A Language and Environment for Statistical Computing. (2017).

62. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132,** 583–589 (1992).

63. Nei, M. Genetic distance between populations. *Am. Nat*. **106,** 192–283 (1972).

64. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105,** 437–460 (1983).

65. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol*. **7,** 256–276 (1975).

66. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595 (1989).

67. Korneliussen, T. S., Moltke, I., Albrechtsen, A. & Nielsen, R. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* **14,** 289 (2013).

68. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15,** 356 (2014).

69. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet*. **9,** e1003905 (2013).

70. Beichman, A. C., Huerta-Sanchez, E. & Lohmueller, K. E. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu. Rev. Ecol. Evol. Syst*. **49,** 433–456 (2018).

71. Rougemont, Q. et al. Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. *PLoS Genet*. **16,** e1008348 (2020).

72. Kul'bachnyi, S. E. & Kul'bachnaya, A. V. Some Features of Biology of the Siberian Taimen *Hucho taimen* (Salmonidae) from the Tugur River basin. *J. Ichthyol*. **58,** 765–768 (2018).

73. Stamatakis, A. The RAxML v8. 2. X manual. Heidleberg Institute for Theoretical Studies. (2016).

74. Petit, C. & Deverchere, J. Structure and evolution of the Baikal rift: a synthesis. *Geochemistry, Geophys. Geosystems* **7**, (2006).

75. Koskinen, M. T., Knizhin, I., Primmer, C. R., Schlötterer, C. & Weiss, S. Mitochondrial and nuclear DNA phylogeography of *Thymallus spp*. (grayling) provides evidence of ice-age mediated environmental perturbations in the world's oldest body of fresh water, Lake Baikal. *Mol. Ecol*. **11,** 2599–2611 (2002).

76. Froufe, E., Knizhin, I. & Weiss, S. Phylogenetic analysis of the genus *Thymallus* (grayling) based on mtDNA control region and ATPase 6 genes, with inferences on control region constraints and broad-scale Eurasian phylogeography. *Mol. Phylogenet*. Evol. **34,** 106–117 (2005).

77. Froufe, E., Alekseyev, S., Alexandrino, P. & Weiss, S. The evolutionary history of sharp-and blunt-snouted lenok (*Brachymystax lenok* (Pallas, 1773)) and its implications for the paleo-hydrological history of Siberia. *BMC Evol. Biol*. **8,** 1–18 (2008).

78. Yokoyama, R., Sideleva, V. G., Shedko, S. V & Goto, A. Broad-scale phylogeography of the Palearctic freshwater fish *Cottus poecilopus* complex (Pisces: Cottidae). *Mol. Phylogenet. Evol.* **48,** 1244–1251 (2008).

79. Skog, A., Vøllestad, L. A., Stenseth, N. C., Kasumyan, A. & Jakobsen, K. S. Circumpolar phylogeography of the northern pike (*Esox lucius*) and its relationship to the Amur pike (*E. reichertii*). *Front. Zool*. **11,** 67 (2014).

80. Grosswald, M. G. Late-Weichselian ice sheets in arctic and Pacific Siberia. *Quat. Int*. **45,** 3–18 (1998).

81. Karabanov, E. B., Prokopenko, A. A., Williams, D. F. & Colman, S. M. Evidence from Lake Baikal for Siberian glaciation during oxygen-isotope substage 5d. *Quat. Res*. **50,** 46–55 (1998).

82. Mangerud, J. et al. Ice-dammed lakes and rerouting of the drainage of northern Eurasia during the Last Glaciation. *Quat. Sci. Rev*. **23,** 1313–1332 (2004).

83. Simonov, E. A. & Dahmer, T. D. Amur-Heilong river basin reader. (*Ecosystems* Hongkong, 2008).

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

85. Bernatchez, L. & Wilson, C. C. Comparative phylogeography of nearctic and palearctic fishes. *Mol. Ecol*. **7,** 431–452 (1998).

86. Ferchaud, A.-L., Laporte, M., Perrier, C. & Bernatchez, L. Impact of supplementation on deleterious mutation distribution in an exploited salmonid. *Evol. Appl*. **11,** 1053–1065 (2018).

87. Kijas, J. et al. Evolution of sex determination loci in Atlantic salmon. *Sci. Rep*. **8,** 1–11 (2018).

88. Leitwein, M. et al. Genome-wide nucleotide diversity of hatchery-reared Atlantic and Mediterranean strains of brown trout *Salmo trutta* compared to wild Mediterranean populations. *J. Fish Biol*. **89,** 2717–2734 (2016).

89. Matthaeus, W. J. *Investigation of fine spatial scale population genetic structure in two Alaskan salmonids*. (Doctoral dissertation, 2016).

90. Ryynänen, H. J. & Primmer, C. R. Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. *BMC Genomics* **7,** 1–11 (2006).

91. Banarescu, P. *Zoogeography of Fresh Waters. Volume 2: Distribution and Dispersal of Freshwater Animals in North America and Eurasia*. 519–1091 (Aula-Verlag, 1991).

92. Kaus, A. et al. Seasonal home range shifts of the Siberian taimen (*Hucho taimen* Pallas 1773): evidence from passive acoustic telemetry in the Onon River and Balj tributary (Amur River basin, Mongolia). *Int. Rev. Hydrobiol*. **101,** 147–159 (2016).

93. Kul'bachnyi, S. E. & Ivankov, V. N. Temporal differentiation and conditions of reproduction of chum salmon *Oncorhynchus keta* (Salmoniformes, Salmonidae) from the Tugur River basin (Khabarovsk Krai). *J. Ichthyol*. **51,** 63–72 (2011).

94. Bernatchez, L. et al. On the origin of species: insights from the ecological genomics of lake whitefish. *Philos. Trans. R. Soc. B Biol. Sci*. **365,** 1783–1800 (2010).

95. Gordeeva, N. V, Alekseyev, S. S., Matveev, A. N. & Samusenok, V. P. Parallel evolutionary divergence in Arctic char *Salvelinus alpinus* complex from Transbaikalia: variation in differentiation degree and segregation of genetic diversity

among sympatric forms. *Can. J. Fish. Aquat. Sci*. **72,** 96–115 (2015).

96. Olson, K. W., Krabbenhoft, T. J., Hrabik, T. R., Mendsaikhan, B. & Jensen, O. P. Pelagic--littoral resource polymorphism in Hovsgol grayling *Thymallus nigrescens* from Lake Hovsgol, Mongolia. *Ecol. Freshw. Fish* **28,** 411–423 (2019).

97. Zhivotovsky, L. A. et al. Eco-geographic units, population hierarchy, and a two-level conservation strategy with reference to a critically endangered salmonid, Sakhalin taimen *Parahucho perryi. Conserv. Genet*. **16,** 431–441 (2015).

98. Han, E., Sinsheimer, J. S. & Novembre, J. Characterizing bias in population genetic inferences from low-coverage sequencing data. *Mol. Biol. Evol*. **31,** 723–735 (2014).

99. Shafer, A. B. A. et al. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol*. **8,** 907–917 (2017).

100. Grosswald, M. G. & Hughes, T. J. 'Back-arc' marine ice sheet in the Sea of Okhotsk. *Russ. J. Earth Sci.* **7,** (2005).

101. Zolotukhin, S. F. & Shcherbovich, I. V. Maximum weight of the Siberian *Hucho taimen* (Pallas) in the area. *Fisheries Russia* **1,** 47–51 (2021).

102. Reid, S. M., Wilson, C. C., Mandrak, N. E. & Carl, L. M. Population structure and genetic diversity of black redhorse (*Moxostoma duquesnei*) in a highly fragmented watershed. *Conserv. Genet*. **9,** 531 (2008).

103. Wu, H. et al. Effects of dam construction on biodiversity: a review. *J. Clean. Prod*. **221,** 480–489 (2019).

104. Allendorf, F. W. & Luikart, G. *Conservation and the genetics of populations*. (John Wiley & Sons, 2009).

105. Weeks, A. R. et al. Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evol. Appl*. **4,** 709–725 (2011).

106. Ryder, O. A. Species conservation and systematics: the dilemma of subspecies. *Trends Ecol. Evol.* **1,** 9–10 (1986).

107. Moritz, C. Defining `evolutionarily significant units' for conservation. *Trends Ecol. Evol*. **9,** 373–375 (1994).

Acknowledgements

This research was funded by The Taimen Fund (previously The Tributary Fund), the Wild Salmon Center, and the University of Nevada, Reno. We thank The Taimen Fund and the Taimen Conservation Fund for accessibility and facilitation of cooperative sampling across fly fishing outfitters. We thank the staff and fishing guides of Konin Lodge, Sweetwater Travel, Hovsgol Travel, and Mongolia River Outfitters for substantial aid in logistics of lodging, sampling, and local permitting. We thank Iain McClaren for his aid in acquiring samples from the Tugur River basin. We thank the Global Water Center Science Team and the University of Nevada, Reno, for support in sampling and outreach. Finally, we thank Michael R. Miller from the University of California, Davis, for helpful discussion and guidance on demographic inference.

Author contributions

S.C., Z.H., and J.B.S. identified the need for this research and secured funding. L.M.G., J.B.S., S.C., and Z.H. collected specimens. L.M.G. extracted DNA, generated ddRADseq libraries, and conducted population genetic analyses. A.R.L.N. executed demographic inference modeling. M.R.S. contributed natural history of taimen and geological history of the Tugur River basin and helped interpret demographic modeling results. L.M.G. and T.L.P. wrote the manuscript, with substantial input from J.P.J. All authors contributed to the final revisions of the manuscript.

Competing interests

The authors declare no competing interests.

Table 1. Geographic locations, sample sizes, and capture information for taimen analyzed in this study. Site abbreviations correspond to those in Figs. 1 and 2.

Outflow	Basin	Location Name	Latitude, Longitude	*N*
Arctic	Selenge	Eg (EG)	49.84, 102.57	13
		Uur (UR)	50.34, 101.87	62
		Delgermörön (DL)	50.11, 98.85	18
Pacific	Amur	Upper Onon (UO)	49.24, 112.02	6
		Lower Onon (LO)	49.12, 111.96	5
	Tugur	Konin (KO)	53.28, 136.26	5
		Munikan (MO)	53.27, 136.12	3
		Konin / Assyni Junction (AJ)	53.07, 136.03	19
		Tugur (TU)	53.11, 136.25	43

Table 2. Population genetic diversity for the full set of 174 individuals, including estimates based on population allele frequencies (H_E) and those based directly on DNA sequence variation across the sampled genomic regions (the average number of pairwise differences between sequences, or θ_π, the number of segregating sites, or θ_W, and the scaled difference between these two, or Tajima's D). High and low columns represent upper and lower 95% confidence intervals for each metric.

Outflow	Basin	Site	H_E	θ_π	θ_π low	θ_π high	θ_W	θ_W low	θ_W high	D	D low	D high
Arctic	Selenge	DL	0.118	0.001187	0.001177	0.001197	0.000774	0.000768	0.000779	-0.00665	-0.00668	-0.00662
		EG	0.121	0.001348	0.001338	0.001358	0.001006	0.000100	0.001013	-0.01024	-0.01026	-0.01021
		UR	0.120	0.001377	0.001367	0.001387	0.000999	0.000993	0.001004	-0.01266	-0.01270	-0.01263
Pacific	Amur	UO	0.227	0.002090	0.002077	0.002103	0.001777	0.001767	0.001787	0.00210	0.00207	0.00213
		LO	0.217	0.001757	0.001744	0.001770	0.001506	0.001496	0.001516	0.00475	0.00472	0.00478
	Tugur	KO	0.214	0.001646	0.001635	0.001658	0.001391	0.001382	0.001400	-0.00025	-0.00028	-0.00022
		MO	0.213	0.001875	0.001861	0.001889	0.001672	0.001661	0.001684	0.00835	0.00832	0.00838
		AJ	0.221	0.001775	0.001765	0.001786	0.001585	0.001578	0.001591	-0.02457	-0.02460	-0.02454
		TU	0.230	0.001694	0.001682	0.001705	0.001186	0.001179	0.001192	-0.00942	-0.00946	-0.00938

Table 3. Demographic modeling parameter estimates for each of eight 2-population models. TU refers to Tugur mainstem, while TT refers to Tugur tributaries. Growth parameters for each population represent expansion (+), contraction (-), and no change (0) following the ancestral divergence. Model numbers correspond to those in Fig. 4. For the best fit model (model 1), 95% confidence intervals are given in parentheses.

Model description	Model #	Growth Parameters		N_e		ancestral N_e	$T_{divergence}$ (generations)	Migration		ΔModel likelihood	ΔAIC
		TU	TT	TU	TT			TU ➔ TT	TT ➔ TU		
2 population with gene flow	1	-	0	40,999 (38,052 - 43,946)	10,058 (7,998 - 12,119)	446,164 (398,862 - 493,466)	28,480 (23,685 - 33,276)	0.00356 (0.00279 - 0.00434)	0.00435 (0.00386 - 0.00485)	31.13	-
	2	0	-	3,302	83,980	734,783	5,046	0.00251	0.59000	39.68	39.36
	3	-	-	47,370	81,411	253,462	53,643	0.00377	0.00027	36.32	23.88
	4	+	+	21,677	72,582	380,356	2,587	0.00702	0.00574	35.73	21.17
	5	0	+	505	54,433	743,812	2,274	0.00489	0.00605	38.05	31.84
	6	+	0	18,046	7,769	816,466	5,739	0.00170	0.00401	36.72	25.73
	7	0	0	687	1,724	523,521	1,247	0.00596	0.00464	35.67	20.88
2 population	8	0	0	435	791	110,201	38	-	-	136.83	484.7

Figure legends

Figure 1. Map illustrating the main drainages from which taimen were sampled from Mongolia and Russia. One Arctic (purple) and two Pacific (green) drainage sampling sites (a). The red rectangle denotes the Tugur basin sampling area, shown in greater detail (b). Sampling sites correspond to those in Table 1 and include the Delgermörön (DL), Eg (EG), Uur (UR), Lower Onon (LO), Upper Onon (UO), Tugur (TU), Konin/Assyni Junction (AJ), Konin (KO), and Munikan (MU) sampling sites. A recently released taimen rests in the flooded grassland (c).

Figure 2. Ancestry coefficient estimates (q) generated with `entropy` for analyses based on all sampled individuals (a, b) and for separate analyses based on the subset of individuals from the two Pacific drainages (c, d). Vertical bars represent individuals, and colors correspond to the admixture proportions for each of k clusters. For both sets of analyses, the $k = 2$ models (a, c) fit the data best. Models for $k = 3$ (b, d) are additionally shown for each set of analyses as the revealed patterns of clustering further illustrate population genetic structure within river systems. As in Fig. 1, sampling sites correspond to those in Table 1 and include the Delgermörön (DL), Eg (EG), Uur (UR), Lower Onon (LO), Upper Onon (UO), Tugur (TU), Konin/Assyni Junction (AJ), Konin (KO), and Munikan (MO) sampling sites.

Figure 3. Genetic variation among taimen illustrated with PCA (calculated using `R` software) of 6,046 SNPs called in all sampled individuals (a), Arctic individuals (b),

and in 3,961 SNPs called separately for individuals from the Pacific drainages (c). In panel (a), samples from the Pacific drainages are represented by circles, and those from the Arctic drainage by squares. In panel (c), samples from the Amur basin are represented by triangles while those from the Tugur basin are represented by circles. The neighbor joining tree in panel (d) supports the deep divergence between Arctic and Pacific drainages, where symbols represent populations shown in panel (a).

Figure 4. Representations of each two-population model tested with `fastsimcoal2`, where TU represents the mainstem population, TT represents the tributaries population, and arrows represent gene flow. Model numbers and parameters correspond to those listed in Table 3.

Figure 5. The expected versus observed SFS for the best fit model shows substantial overlap (a). Panel B shows the schematic representing the best fit model (model 1) from demographic inference using `fastsimcoal2`. Coalescent *Ne* is given for population, with branch and arrow width corresponding to population size and level of gene flow, respectively (but not drawn to scale). Numbers in parentheses represent 95% confidence intervals.

Figure 1

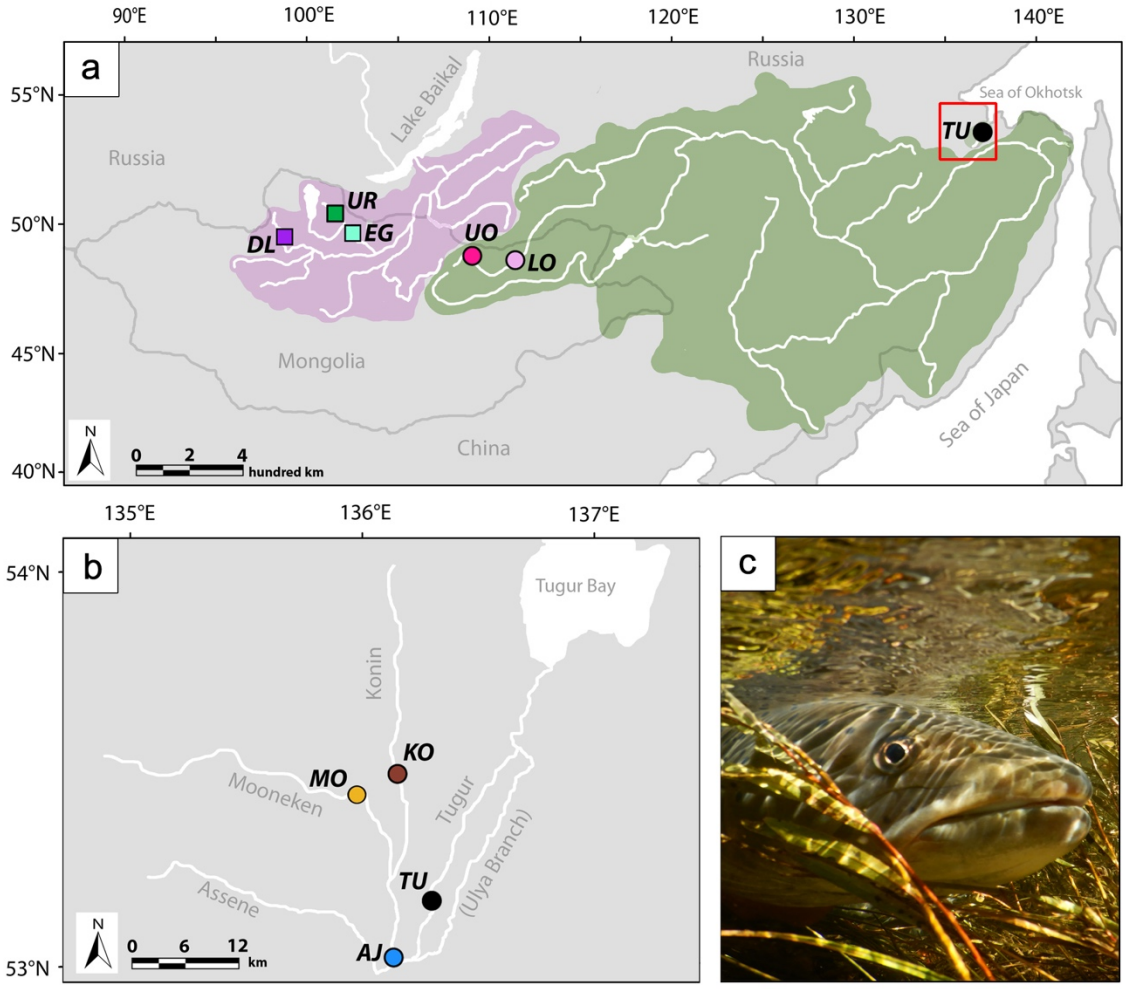

a
90°E
100°E
110°E
120°E
130°E
140°E
55°N
50°N
45°N
40°N
Russia
Lake Baikal
Russia
Sea of Okhotsk
TU
UR
EG
DL
UO
LO
Mongolia
China
Sea of Japan
N
0 2 4
hundred km
b
135°E
136°E
137°E
54°N
53°N
Tugur Bay
Konin
KO
MO
Mooneken
Tugur
(Ulya Branch)
TU
Assene
AJ
N
0 6 12
km
c

Figure 2

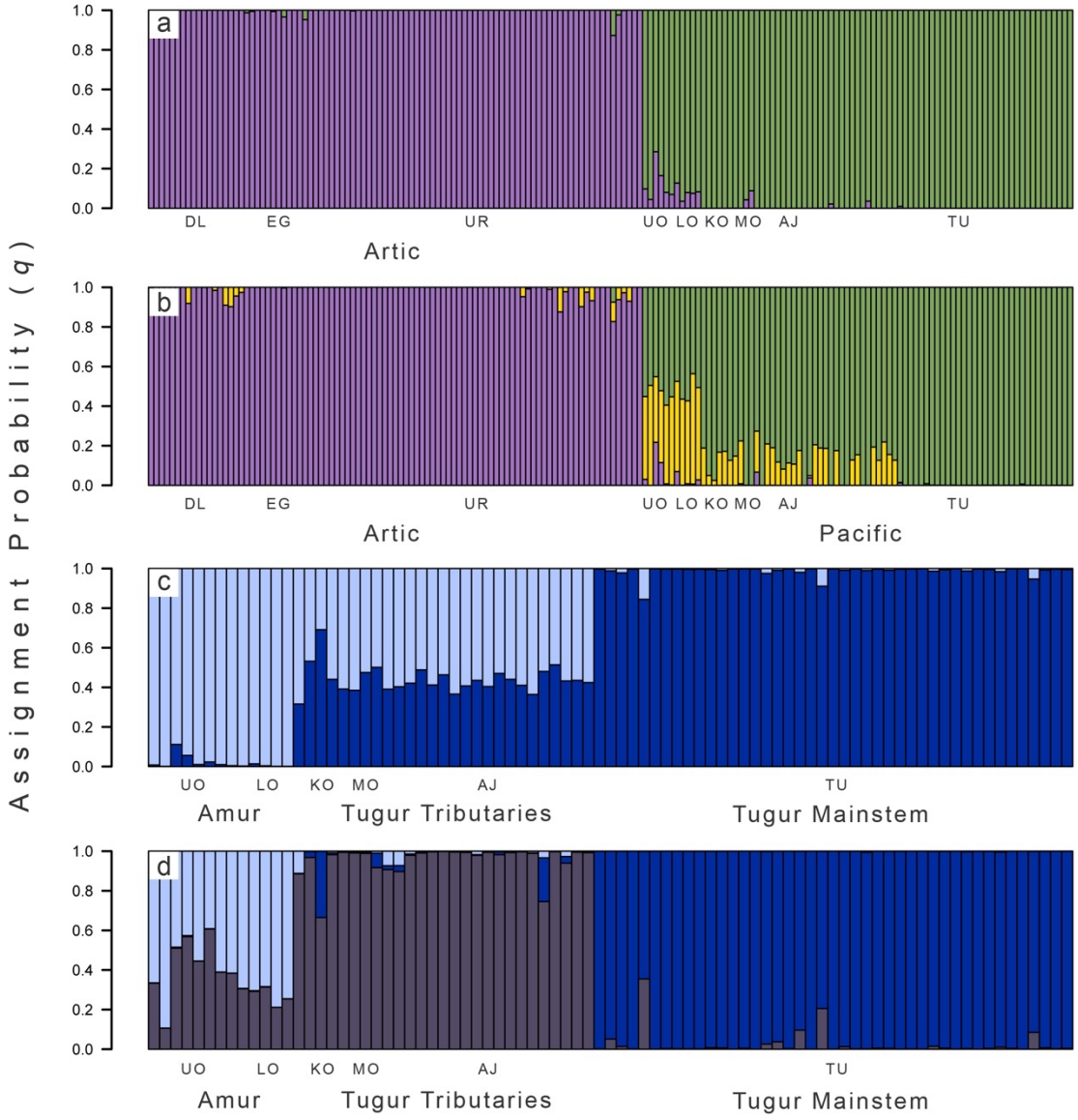

a
b
c
d
Assignment Probability (q)
1.0
0.8
0.6
0.4
0.2
0.0
DL
EG
UR
UO LO KO MO
AJ
TU
Artic
Pacific
Amur
Tugur Tributaries
Tugur Mainstem

Figure 3

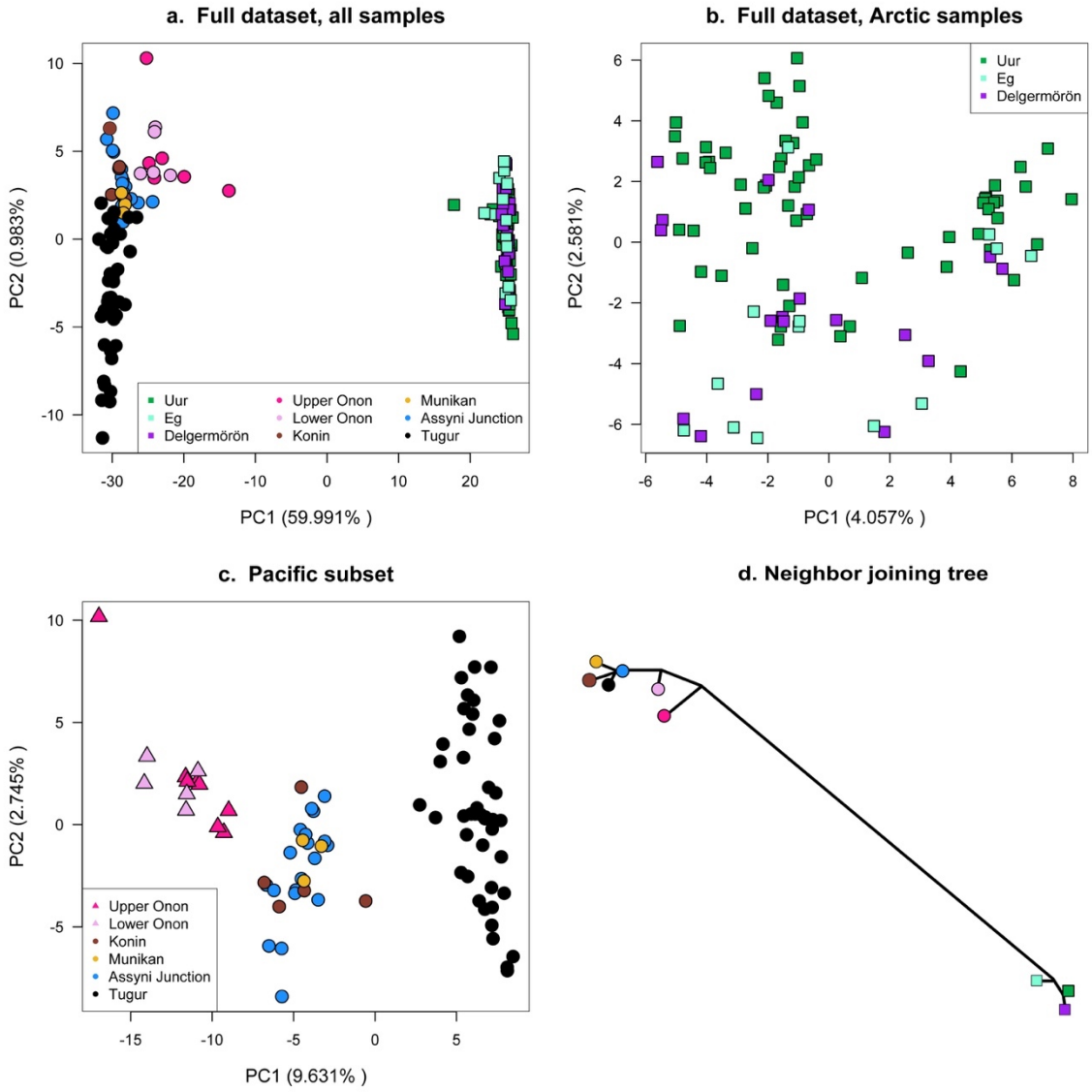

a. Full dataset, all samples
PC2 (0.983%)
PC1 (59.991%)
Uur
Eg
Delgermörön
Upper Onon
Lower Onon
Konin
Munikan
Assyni Junction
Tugur
b. Full dataset, Arctic samples
PC2 (2.581%)
PC1 (4.057%)
Uur
Eg
Delgermörön
c. Pacific subset
PC2 (2.745%)
PC1 (9.631%)
Upper Onon
Lower Onon
Konin
Munikan
Assyni Junction
Tugur
d. Neighbor joining tree

Figure 4

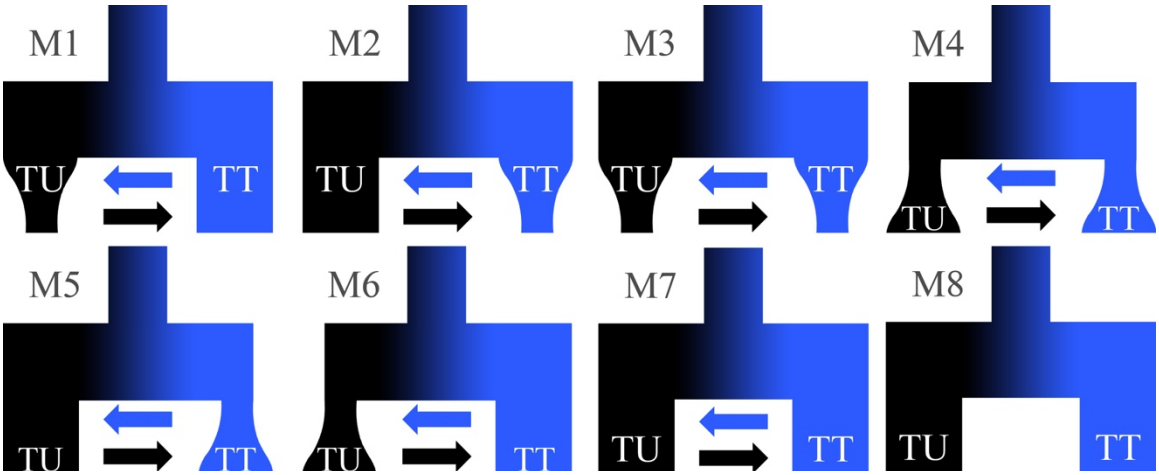
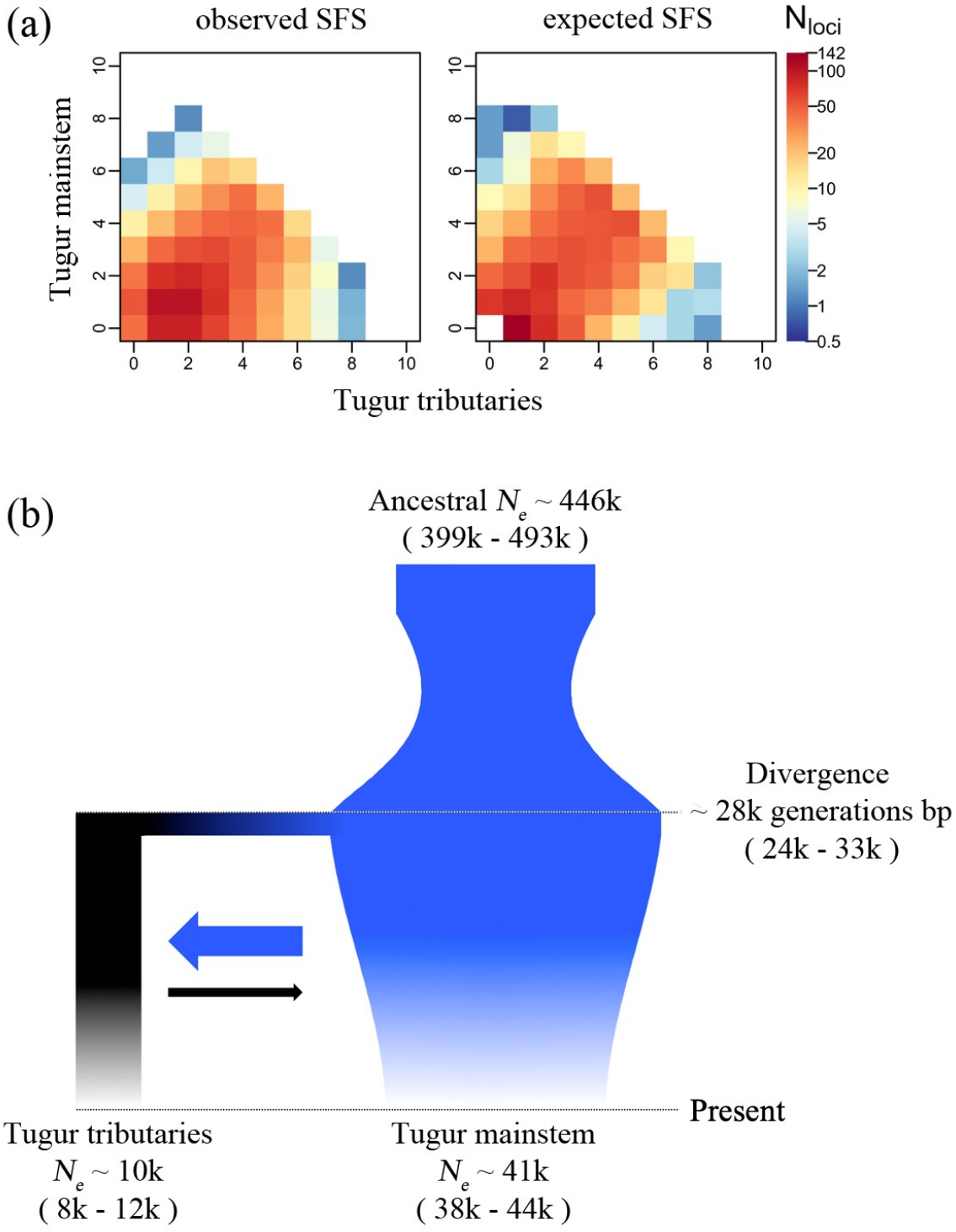


Figure 5


(a)
observed SFS
expected SFS
N_loci
142
100
50
20
10
5
2
1
0.5
Tugur mainstem
Tugur tributaries
0
2
4
6
8
10
(b)
Ancestral N_e ~ 446k
(399k - 493k)
Divergence
~ 28k generations bp
(24k - 33k)
Present
Tugur tributaries
N_e ~ 10k
(8k - 12k)
Tugur mainstem
N_e ~ 41k
(38k - 44k)

Supplementary material

Chapter 1: Hierarchical genetic structure and implications for conservation of the world's largest salmonid, *Hucho taimen*

*Lanie M. Galland[1,2], James B. Simmons[1,2], Joshua P. Jahner[2], Agusto R. Luzuriaga-Neira[2], Matthew R. Sloat[3], Sudeep Chandra[1,2,4], Zeb Hogan[2,4], Olaf P. Jensen[5], & Thomas L. Parchman[1,2]

[1]Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, Reno, NV, USA

[2]Department of Biology, University of Nevada, Reno, Reno, NV, USA

[3]Wild Salmon Center, Portland, OR, USA

[4]Global Water Center, University of Nevada, Reno, Reno, NV, USA

[5]Center for Limnology, University of Wisconsin - Madison, Madison, WI, USA

***Author for correspondence:** Lanie M. Galland, lgalland@unr.edu

Supplementary Methods

We conducted a phylogenetic analysis of all sampling localities based on a concatenated alignment of ddRADseq loci using a maximum likelihood approach based. Given patterns of hierarchical genetic structure in previous analyses, and to produce an illustrative tree without an unnecessarily large number of tips, we subsampled 3 individuals from each sampling locality for this analysis. We additionally included three individuals of *Brachymystax lenok* (a closely related salmonid, also sampled in Mongolia) as an outgroup. We generated a multiple alignment of sequenced ddRADseq loci using `ipyrad v. 0.9.15`[1]. Default values were chosen for most parameters, unless otherwise stated below. Nucleotide sites with phred scores less than 33 were considered missing and were replaced with "N," representative of an ambiguous nucleotide base. We began with *de novo* assembly using `vsearch`[2] with a clustering similarity threshold (*clust_threshold*) of 0.85. To account for uneven sequence depth, statistical (*mindepth_statistical*) and majority-rule (*mindepth_majrule*) base calling were set to 5 and 4, respectively. Contigs were reduced to consensus sequences within each individual at each site, and sequences with more than 5% heterozygous bases (*max_Ns_consens*) or more than 8% heterozygous sites (*max_Hs_consens*) were discarded. Next, the clustering step was repeated using identical parameters, but across, rather than within, all individuals. Clusters were then filtered and discarded if they contained more than eight indels (*max_Indels_locus*) or if more than 20% of sites were variable (*max_SNPs_locus*). Finally, we retained all loci found in more than 21 individuals (*min_samples_locus*).

We used the resultant `phylip` output file from `ipyrad` as input for phylogenetic analyses using a maximum likelihood approach. We inferred a maximum likelihood phylogeny used `RAxML v. 8.2.12`[3] using the "-f a" option, which searches for the best--scoring tree and performs a bootstrap analysis. We conducted searches using the GTR + GAMMA evolutionary model of sequence evolution, and the number of bootstrap replicates was assessed using the `autoMRE` option, resulting in 1,000 replicates. Although bootstrap support was low for many nodes at more recent scales, we rendered the tree in Fig. S2 to show all nodes in order to illustrate relevant patterns of divergence and support across the sampled localities.

Table S1. DIC values for 5 replicate `entropy` runs of each *k* ancestral model for both the full dataset and the Pacific subset. Mean and standard deviation (SD) are included for each model. Lower DIC scores represent better model fit.

	k	1	2	3	4	5	mean	SD
Full dataset	2	2,563,950	2,574,427	2,556,251	2,564,459	2,542,892	2,560,396	11,721
	3	3,397,146	3,389,373	3,417,252	3,453,044	3,465,691	3,424,501	33,712
	4	3,363,991	3,257,494	3,420,572	3,348,324	3,379,798	3,354,036	60,297
	5	3,361,903	3,786,156	3,844,955	3,863,475	3,827,988	3,736,895	211,567
	6	3,751,559	3,697,491	3,835,392	3,427,559	3,998,918	3,742,184	209,568
	7	3,664,111	6,597,391	4,099,092	5,211,434	3,300,373	4,574,480	1,339,402
	8	3,531,952	4,009,169	3,711,057	3,885,740	3,542,185	3,736,020	210,277
	9	4,397,436	3,400,362	3,503,357	12,730,622	3,806,935	5,567,742	4,022,938
Pacific subset	2	1,409,122	1,362,080	1,388,500	1,371,819	1,393,363	1,384,977	18,460
	3	1,641,797	2,423,495	1,847,685	1,874,910	2,097,337	1,977,045	297,214
	4	1,623,020	1,567,422	1,487,859	1,661,693	1,557,531	1,579,505	66,476
	5	2,152,078	2,219,632	2,222,038	1,965,042	1,639,002	2,039,558	247,176
	6	2,126,383	2,191,929	1,656,468	2,765,256	1,593,665	2,066,740	474,151

Table S2. Pairwise estimates of mean F_{ST} (upper diagonal) and Nei's *D* (lower diagonal) among all sampling sites. Site abbreviations correspond to those in Table 1.

	DL	EG	UR	UO	LO	KO	MO	AJ	TU
DL	-	0.009	0.006	0.213	0.224	0.255	0.250	0.247	0.258
EG	0.0014	-	0.008	0.209	0.221	0.252	0.247	0.244	0.256
UR	0.0004	0.0008	-	0.212	0.223	0.255	0.250	0.248	0.256
UO	0.1441	0.1405	0.1438	-	0.017	0.039	0.038	0.026	0.036
LO	0.1492	0.1474	0.1485	0.0035	-	0.034	0.034	0.023	0.031
KO	0.1797	0.1767	0.1794	0.0155	0.0109	-	0.029	0.017	0.023
MO	0.1702	0.1698	0.1698	0.0128	0.0085	0.0039	-	0.018	0.022
AJ	0.1736	0.1709	0.1734	0.0093	0.0060	0.0041	0.0027	-	0.012
TU	0.1823	0.1804	0.1797	0.0153	0.0107	0.0042	0.0031	0.0033	-

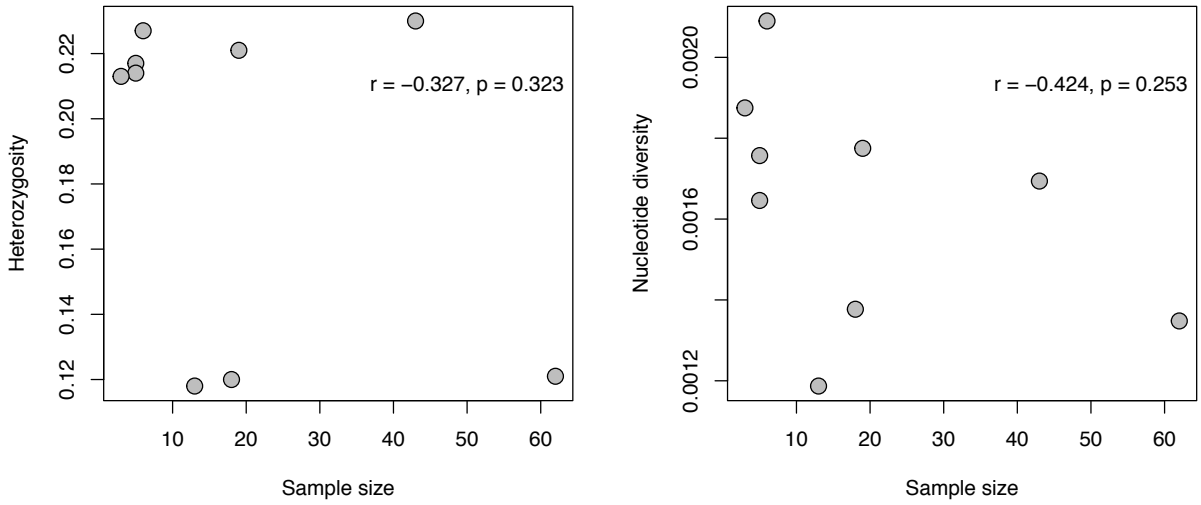


Figure S1. Plots of genetic diversity metrics (heterozygosity and nucleotide diversity) by sample size for each locality. Sample size does not influence estimates of genetic diversity.

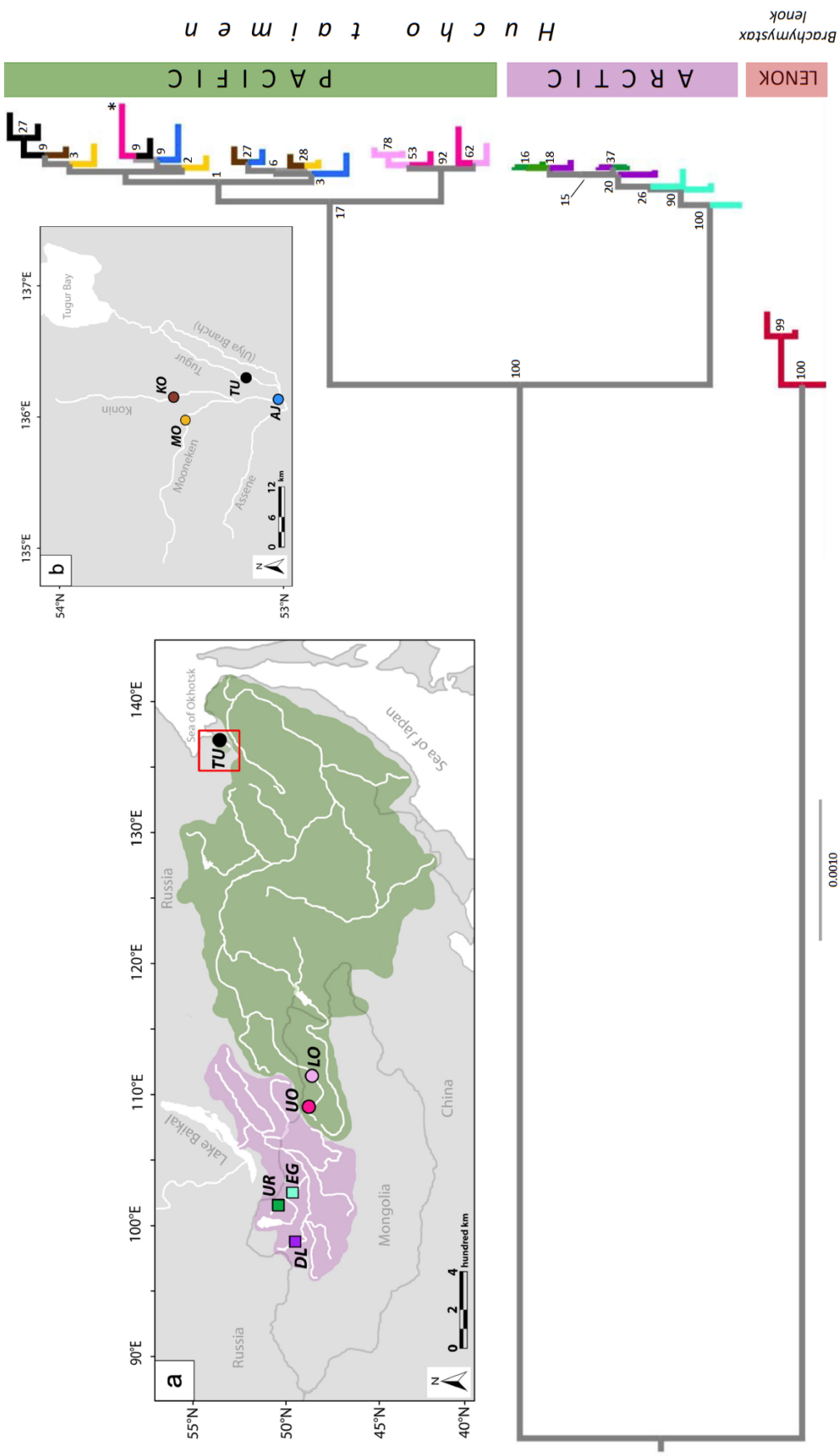


Figure S2. Phylogenetic tree inferred from a concatenated alignment (1,229 RADseq loci, 1,642 phylogenetically informative sites) using RAxML. Bootstrap support (out of 100) is illustrated at nodes. *Brachymystax lenok* was used as an outgroup.

Supplementary References

1. Eaton, D. A. R. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* btu121 (2014).

2. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

3. Stamatakis, A. The RAxML v8. 2. X manual. Heidleberg Institute for Theoretical Studies. (2016).

Chapter 2: Assessing the population genetic structure of introduced rainbow trout (*Oncorhynchus mykiss*) in the Lake Tahoe basin: a case for understanding hybridization potential during the reintroduction of the native, threatened Lahontan cutthroat trout (*O. clarkii henshawi*)

L.M. Galland[12], T.L. Parchman[12], and M.M. Peacock[12]

[1]Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, NV, USA

[2]Department of Biology, University of Nevada, Reno, NV, 89557, USA

Author for correspondence:

Email: lgalland@unr.edu

Acknowledgements This research was funded by the University of Nevada, Reno. We thank the undergraduate research assistants for their extensive aid in sample collection. We thank Travis Hawks from the Nevada Department of Wildlife and Mitch Lockhart from the California Department of Fish and Wildlife for their information on rainbow trout stocking in the Lake Tahoe basin. We further thank Travis Hawks for providing the catch record of tagged fish stocked into Lake Tahoe. We also thank Tyler Hern, Supervisory Fish Biologist, United States Fish and Wildlife Service (USFWS) Erwin National Fish Hatchery; Robert Null, Deputy Project Leader

USFWS Coleman National Fish Hatchery Complex; and April Gregory and Carlos Martinez of D.C. Booth Historic National Fish Hatchery and Fish and Aquatic Conservation Archives for information on rainbow trout strain origins and stocking history. All samples were collected in compliance with the following: Nevada Department of Wildlife scientific collecting permit #386188, California Department of Fish and Wildlife scientific collecting permit SC-13232, the Lake Tahoe Basin Management Unit (written approval), and the Institutional Animal Care and Use Committee protocol 2015-00641.

Abstract Hybridization with introduced or invasive species is a driver of population declines in native salmonids and a major threat to their persistence in the wild. The rainbow trout (*Oncorhynchus mykiss*, RBT) has been widely introduced globally and represents an important invasive species, often establishing naturalized populations. The cutthroat trout (*Oncorhynchus clarkii*, CT), a close congener, is particularly susceptible to competition and hybridization from RBT introductions which has led to range-wide population declines and loss of CT genetic variation. The Lahontan cutthroat trout (*O. c. henshawi*, LCT) whose historic distribution included the Lake Tahoe basin, was extirpated by the 1940s due to overfishing and introduction of nonnative salmonids, including now naturalized RBT. Diploid reproductively viable RBT were stocked annually into Lake Tahoe from the late 1800s until the mid-2000s by California and Nevada fish and wildlife agencies, planting the same commonly raised hatchery strains over time. Since 2007, triploid (sterile) RBT comprise the bulk of RBT planted. To assess potential homing of RBT to streams for spawning, thereby informing LCT reintroduction, we characterize genetic variation of RBT in a subset of Lake Tahoe tributaries. Despite

extensive dispersal from stocking locations, our analyses revealed variation in population differentiation among tributaries, with individuals from spatially proximate streams clustering. Although subtle, we also detected evidence for genetic differentiation among tributaries from the southern, western, and northern regions, including surprising structure involving a single tributary. These results illustrate the extent of differentiation within and among streams, and could inform possibilities for and implications of RBT removal and LCT reintroduction.

Key words: RADseq, genetic diversity, genetic structure, *Oncorhynchus clarkii*, *Oncorhynchus mykiss*, introduced species

Introduction

In recent decades, populations of many North American native freshwater fishes have dramatically declined and are under increasing risk of extinction (Abell et al. 2000; Moyle et al. 2011). Over the 20th and 21st centuries, the cold-water salmonids have suffered significant range contractions globally due to anthropogenic disturbances. This is especially pronounced among the native inland species, where local population extirpations have been widespread, resulting in threatened or endangered listing under the United States Endangered Species Act (ESA) or as vulnerable on the International Union for Conservation of Nature (IUCN) red list (Dauwalter et al. 2020; Fagan et al. 2005; Maxwell and Jennings 2005; Muhlfeld et al. 2019). Major threats to native salmon and trout include habitat loss, fragmentation, and degradation; nonnative salmonid introductions and invasions; introductions of hatchery raised stocks; and climate change (Bonar et al. 2005; Clavero et al. 2017; Costello 2009; Ford and Myers 2008; Kareiva et al. 2000; Wenger et al. 2011; Otero et al. 2014; Sepulveda et al. 2015). Specifically, the inland forms of cutthroat trout (*Oncorhynchus clarkii*, CT), consisting of multiple subspecies native to the intermountain western United States, are at increased risk of extirpation from habitat alterations, nonnative salmonid introductions, and impacts of climate change (Dunham et al. 1997, 2003; Kruse et al. 2000; Peterson et al. 2004; Sanderson et al. 2009; Wenger et al. 2011, 2017). Nonnative salmonid introductions, in particular, represent a triumvirate of threats to native species from competition, predation, and hybridization (Dunham et al. 1997, 2003; Koel et al. 2005).

The Lahontan cutthroat trout (*O. clarkii henshawi*, LCT), endemic to the hydrographic Lahontan basin of northeastern California, northern Nevada, and

southeastern Oregon, has been listed as threatened under the ESA since 1975 (Coffin and Cowan 1995). This subspecies has been extirpated from >90% of its historic stream habitat and 99% of its historic lake habitat (Coffin and Cowan 1995; Dunham et al. 1997). The only native lacustrine population remaining in the western Lahontan basin (Truckee, Walker, and Carson river watersheds) exists in Independence Lake in the upper Truckee River watershed (where the majority of historically occupied lacustrine habitat is found: Independence, Tahoe, Donner, Cascade, Fallen Leaf, Walker, and Pyramid lakes; Peacock et al. 2018). A relatively small geographic area of the larger Lahontan basin, the Truckee River watershed spans the California-Nevada border and is home to the two largest lacustrine habitats occupied historically by LCT: Lake Tahoe, a large oligotrophic subalpine lake representing the tenth deepest lake in the world, and the endorheic and mesotrophic Pyramid Lake, a remnant of the large pluvial Lake Lahontan, which covered most of northwestern Nevada during the Pleistocene (Thompson et al. 1986). The two large lakes are connected by the 195 km Truckee River corridor. LCT in these large lacustrine habitats grew to exceptionally large sizes and were considered the largest inland trout in North America (Peacock et al. 2018). In addition to being the apex predator in these ecosystems, these fish were an important part of the diet and cultural heritage of the native Paiute and Washoe tribes. In the 1940s, LCT were extirpated from the Lake Tahoe and Truckee River basins due to overharvest, lack of access to suitable spawning habitat, water diversions, and introduction of nonnative salmonids including lake trout (*Salvelinus namaycush*), brook trout (*S. fontinalis*), kokanee salmon (*O. nerka*), brown trout (*Salmo trutta*), and rainbow trout (*O. mykiss*) (Gerstung 1988; Peacock et al. 2018).

As with the other CT subspecies, perhaps the most formidable threat to survival and successful reintroduction of LCT into its historical habitat are naturalized populations of nonnative rainbow trout (RBT, *O. mykiss*), a close congener with which CT subspecies readily interbreed often forming hybrid swarms (Allendorf et al. 2001; Campbell et al. 2002; Leary et al. 1987). RBT, native to the west coast of North America and the northeastern coast of Russia south through Kamchatka, exhibit well documented homing behavior, returning to natal streams to spawn and thus giving rise to population genetic structure specific to populations by basin, stream, and even location within stream (Altukhov et al. 2000; Heath et al. 2001; Weigel et al. 2013). RBT have been artificially propagated in the United States for >150 years, with the first RBT hatcheries in the United States dating back to the early 1870s, originally established on San Leandro Creek (tributary to San Francisco Bay) and Campbell Creek (tributary to the McCloud River in northern California) (Schley 1971). Shortly thereafter, RBT were shipped across the United States, starting with New York and Michigan (Schley 1971). Since that time, RBT have been widely introduced across the globe, and naturalized populations (those that were originally stocked but achieved natural reproduction, spawning in the ecosystems where they were introduced) have flourished in a significant number of ecosystems, earning the species a place on the world's worst 100 invasive species list (Lowe et al. 2000). The history of rainbow trout hatchery stocks involves multiple source populations of inland and anadromous forms as well as intermixing of these source populations in hatchery stocks over time. Although hatcheries have not focused on maintaining genetic variation in their stocks, the mixing of genomes across these different source populations could facilitate survival and naturalization potential of stocked RBT.

Introduction and success of RBT in nonnative ecosystems has led to the loss of biodiversity and has dramatically impacted the fates of native CT due to hybridization. The relationship between RBT and CT is one of the most well- known examples of introgressive hybridization in salmonids (Mandeville et al. 2019; Ostberg and Rodriguez 2004). Past work in the Truckee River suggests an unequal mating dynamic between LCT and RBT, as female LCT mtDNA haplotypes accounted for 14% of the F_1 hybrids, despite the fact that LCT make up less than 5% of the salmonids in the river. This suggests that female LCT mate with male RBT in higher proportion than would be expected by chance (Kirchoff 2016). Depending on rates of backcrossing, such a dynamic could lead to the potential loss of native LCT genetic variation through genetic assimilation (Allendorf et al. 2001). Competitive exclusion and reduced fitness of hybrids can contribute to greater introgression with RBT, resulting in genetic erosion and eventual loss of pure CT populations across their range (Allendorf and Leary 1988; Muhlfeld et al. 2009; Rhymer and Simberloff 1996; Todesco et al. 2016).

In the Lake Tahoe basin, where LCT were historically the only native salmonid, naturalized RBT populations have been established for >50 years (Cordone and Frantz 1968). In recent years, both California and Nevada fish and wildlife agencies have either stopped planting RBT (California 2007) or have switched to planting primarily infertile triploid RBT along with a small number of hatchery raised diploid RBT from naturalized stream populations (Incline and Third Creeks) that drain into Lake Tahoe (Nevada 2011). Thus, reestablishing naturally reproducing populations of LCT in Lake Tahoe basin is constrained by multiple threats including hybridization with these robust, naturalized populations of RBT that now occupy most tributary streams. Both RBT and CT are

spring stream spawners, and both have stream resident and migratory forms, where homing to natal streams is common (Dunham et al. 2002; Keefer and Caudill 2014; Labar 1971; Muhlfeld et al. 2012; Neville et al. 2006; Wenburg et al. 1998). We currently lack knowledge of population genetic variation in naturalized RBT of the Lake Tahoe basin, including the extent to which genetic structure could have arisen as a result of homing to specific streams. Such perspectives could be an important step towards assessing the threat of hybridization among naturalized RBT and reintroduced LCT, identifying streams best suited for RBT removal and LCT reintroduction.

Here, we establish a baseline understanding of population genetic structure of nonnative, naturalized populations of RBT found in tributaries draining into Lake Tahoe. Characterizing genetic structure may aid in understanding and anticipating challenges associated with reintroduction of native LCT, where introduced nonnative salmonid populations have been regularly bolstered by stocking for recreational angling purposes. Specifically, we sought to understand the occurrence and extent of genetic differentiation of RBT among different regions of the basin, including the potential for naturalized populations to be associated with spawning in specific tributaries. Genetic differentiation among RBT from specific streams or groups of streams may indicate homing behavior and decreased migration among tributaries and could thus aid in the identification of streams where mitigation strategies could facilitate LCT reintroduction into streams with suitable spawning and rearing habitat. We used single nucleotide polymorphism (SNP) data generated with a reduced-representation sequencing approach for RBT sampled from a subset of streams that drain into Lake Tahoe to better understand the genetic consequences of deliberate species introduction by: (1) characterizing basin-wide

population genetic structure, and (2) quantifying genetic diversity within and among RBT sampled from different streams. An understanding of population genetic substructure of naturalized nonnative salmonids will be critical to developing effective conservation and management strategies in systems where restoration of native ecosystems is a priority.

Methods

Stocking and catch records

Rainbow trout have been planted into waters of the Lake Tahoe basin since the late 1800s; several local within-basin hatcheries were established early on and supplied RBT for the lakes and tributaries of the basin (Online Resource 1; Leitritz 1970; Sigler and Sigler 1986). The specific strains of trout raised in these hatcheries are typically not identified in any available records. However, these strains likely originate from the McCloud River (redband trout) and other tributaries in the upper Sacramento River watershed, as there were multiple egg collection stations and several hatcheries on these rivers in the early 20th century (Leitritz 1970). Trout from these within-basin hatcheries were widely planted around Lake Tahoe and in its tributaries from the early 1900s to mid-century (Kelley 1957). Further, rainbow trout appear to have already been naturalized in the Truckee River by the early 1900s, as eggs were collected and housed at the Verdi hatchery (1902-1905) on the Truckee River and then shipped to the Tahoe and Mt. Shasta (Siskiyou County, California) hatcheries to be reared for later out-planting (Leitritz 1970). Still, despite decades of stocking, by the end of the 1950s, Lake Tahoe was considered a “poor fishing lake” as few of the planted rainbow were actually caught by anglers (Kelley 1957). All within-basin hatcheries were decommissioned by the 1950s

(Online Resource 1), and California had closed the majority of its over 120 hatcheries and egg collection stations by 1960, consolidating its hatchery program of all fish species to 11 modernized hatcheries (Leitritz 1970). In the Lake Tahoe basin, California and Nevada state fish and wildlife agencies began to focus on habitat improvement and use of specific RBT strains to establish and increase the size of naturalized RBT populations around the lake.

Records indicate the early efforts to establish robust naturalized populations of RBT within the Lake Tahoe basin were limited at best. Here, we focus on stocking records from the past ~50 years, after the change in stocking strategy by California and Nevada game and fish agencies, to assess whether stocking patterns are associated with any observed genetic structure of naturalized RBT populations. We obtained records of the strains, number of individuals, and stocking locations around the lake from both California Department of Fish and Wildlife (CDFW) and Nevada Department of Wildlife (NDOW). We also obtained information on strain origin and history from archives of the United States Fish and Wildlife Service (USFWS) and individual hatcheries. Further, we evaluated mark-recapture data from recreational catch records of tagged, stocked RBT to investigate movement patterns and dispersal of newly stocked individuals along inshore, lacustrine habitat.

Sample collection

To determine which tributaries to Lake Tahoe were occupied by RBT, we consulted with state and federal fish and wildlife agencies regarding recent electrofishing survey data. We attempted to sample fish from the majority of streams where RBT were previously

observed. Several tributaries suffered detrimental effects of severe multi-year drought, resulting in the absence of RBT or streambeds that were completely dry. Despite sampling efforts spanning 28 streams, we were only able to collect data from 8 tributaries for this study (Fig. 1a, Table 1). These tributaries likely represent stable habitat that would be continuously occupied by RBT. We used electrofishing methodology for individual capture and collected small fin clip samples from the left pelvic fin from all individuals sampled (following state permit regulations and IACUC protocols; see Acknowledgments).

DNA sequencing

Genomic DNA was isolated from fin clips using Qiagen DNeasy Blood and Tissue kits (Qiagen Inc., Valencia, CA). We used a reduced-representation technique (ddRADseq; Peterson et al. 2012) to create DNA sequencing libraries using the protocol described in Parchman et al. (2012). Briefly, we digested DNA with restriction endonucleases *Mse*I and *Eco*RI. We ligated Illumina adaptors containing unique 8-10 bp DNA barcodes to each *Eco*RI cutsite, and Illumina DNA sequencing adaptors to each *Mse*I cutsite. Barcoded samples were pooled and amplified using PCR primers and a proofreading polymerase (Iproof, BioRad, Hercules, CA). See the protocol document at the DRYAD repository (doi: 10.5061/dryad.15dv41nzk) for more details on sequencing library preparation. The reduced-representation library was sent to the University of Texas Genome and Sequencing Analysis Center, where DNA fragments 350-450 bp in length were selected using the Pippin Prep quantitative electrophoresis unit (Sage Science, Beverly, MA) and subsequently sequenced across two lanes on the Illumina HiSeq 4000.

Alignment and variant calling

We used `bowtie2_db` (Langmead and Salzberg 2012) and Perl and bash scripts (http://github.com/ngcr/tapioca) to filter raw sequencing data for contaminant DNA including *PhiX* and *E. coli*, and for PCR primers and sequencing adaptors. We used a custom Perl script to correct 1-2 bp sequencing errors in unique barcode sequences as all barcodes differed by ≥3 bp. We then matched barcoded sequences with unique sample IDs, removing barcode associated bases from reads, and we subsequently split fastq files into individual fastqs for each sample. Samples yielding sequence data <100 megabytes were removed from further analyses.

We mapped reads to the *O. mykiss* reference genome (GenBank assembly accession GCA_900005705.1) using `bwa v0.7.5` (Li and Durbin 2009). Next, we converted sequence alignment map files to binary alignment map files using `samtools v1.3` before calling variants using `samtools v1.3` and `bcftools v1.3` (Li et al. 2009). To calculate genotype likelihoods, we used minimum site quality of 20, minimum genotype quality of 10, minimum base quality of 20, minimum map quality of 20, and maximum coverage depth of 100. Using `vcftools v0.1.14` (Danecek et al. 2011), variants were filtered to include only bi-allelic SNPs with minor allele frequencies >0.05 and where at least 60% of individuals had at least one read. Next, we removed samples if an individual was missing data at more than 30% of its loci. To avoid over-assembled loci that could represent paralogous regions, we filtered out loci with coverage depth >12x using `vcftools v0.1.14`. As a further step against genotyping errors representing potentially misassembled paralogous regions, we used the *HDplot* approach (McKinney

et al. 2017), retaining loci with heterozygosity (*H*) between 0 – 0.55 and read ratio deviance (*D*) from -4 through 4 to ensure we only retained singleton loci.

Population genetic analyses

To infer genotypes probabilistically and to infer population structure via ancestry, we used `entropy` (Gompert et al. 2014). Similar to the admixture model in `structure` (Falush et al. 2003; Pritchard et al. 2000), `entropy` is a Bayesian model that estimates the number of *k* ancestral groups and estimates admixture proportions (*q*) within and among populations without *a priori* sample information. Stochastic differences in coverage depth and sequencing and alignment errors are characteristic of low- to medium-coverage sequencing data (Buerkle and Gompert 2013; Fumagalli et al. 2013; Nielsen et al. 2011); `entropy` accounts for this uncertainty by producing posterior probabilities for each individual genotype at each locus. To begin, we used the `prcomp` function in `R version 3.4` (R Core Team, 2017) to run PCA on the covariance matrix of genotype likelihoods. We then used linear discriminant analysis (LDA) and *k*-means clustering to assign initial ancestry proportion estimates to each individual for $k = 2$ through $k = 8$ groups. We ran `entropy` models for each of the *k* ancestral groups using 5 chains per *k*, running 100,000 MCMC iterations and retaining the 10^{th} step after a burn-in of 30,000 steps. To assess model fit, we compared the deviance information criterion (DIC) for each model, where the lowest DIC represented the best-fit model.

As a model free method for examining population genetic variation, we used the `prcomp` function in `R` to run PCA on genotype probabilities generated with `entropy`. We additionally conducted discriminant analysis of principal components (DAPC; Jombart et

al. 2010) to analyze differences between groups of individuals. Where PCA maximizes total variance between individuals, DAPC minimizes variation within groups and maximizes variation between groups (Jombart et al. 2010). First, we ran DAPC without *a priori* group assignment by using `find.clusters` in the `adegenet R` package to infer the number of *k* genetic clusters (Jombart 2008; Jombart and Ahmed 2011), where Bayesian Information Criterion (BIC) scores were used to assess model fit. Next, considering patterns of subtle differentiation among groups of streams in prior analyses in addition to ecological and biological relevance, we conducted DAPC using *a priori* group assignments based on region ($k = 3$: North, South, West), and again with the same three regions but with McKinney Creek individuals grouped separately based on clustering in PCA and `entropy` analyses above ($k = 4$: North, South, West, McKinney). To additionally analyze population genetic differentiation, we ran three separate analyses of molecular variance (AMOVAs). Using the `poppr` (Kamvar et al. 2014) package in `R`, we analyzed variance explained by among versus (a) within stream, (b) within region ($k = 3$; North, South, West), and (c) within region, with McKinney grouped separately ($k = 4$; North, South, West, McKinney).

To summarize genome-wide levels of differentiation among RBT sampled from each stream, we calculated Nei's *D* (Nei 1972) and Hudson's F_{ST} (Hudson et al. 1992) on allele frequencies. To assess levels of genetic variation for RBT sampled from each stream, we calculated Θ_W (the number of segregating sites; Watterson 1975) and Θ_π (the average number of pairwise differences between sequences; Tajima 1983) using `ANGSD`, where genotype uncertainty is incorporated (Korneliussen et al. 2013, 2014). We used the "doSaf 1" tool within `ANGSD` to estimate the site allele frequency likelihoods from the

reference genome and BAM files. Site allele frequency likelihoods were then used to generate folded site frequency spectrum likelihoods using `REALSFS` (Fumagalli et al. 2013). We calculated posterior allele frequency probabilities, and then used the `thetastat` tool in `ANGSD` to estimate per-locus diversity metrics. We averaged these values across all contigs and nucleotides to generate population-level measures of diversity.

Results

Review of stocking and catch records

Multiple RBT hatchery strains have been stocked into Lake Tahoe since 1960, and often multiple stocks were planted in a single year (Online Resource 2). The same stocking stations around the lake have been used annually by California (8 sites on the West and South shores) and Nevada (4 or 5 sites on the East, North, and South shores) fish and wildlife agencies. The preponderance of diploid RBT stocked into Lake Tahoe from 1980-2010 by both state fish and wildlife agencies were the Eagle Lake strain (*O. mykiss aquilarum*, ~500,000 adult individuals), representing 50% of all RBT planted among the 16 strains stocked into the lake over this time period (Online Resource 2). Eagle Lake is an endorheic lake found in Lassen County in northern California, and this strain is raised in multiple hatcheries across the country. Additionally, from 1991-present, NDOW biologists also recovered sperm and eggs from naturalized RBT males and gravid females spawning in North Lake tributaries (Incline and Third creeks) and fertilized eggs were transferred to and raised in the NDOW Mason Valley hatchery with the aim of planting these offspring from successful spawn runs back into the lake. A total of 691,688

hatchery raised RBT from these creeks have been stocked back into the lake since 1991, representing ~43% of all individuals from the 8 diploid strains stocked 1991-present. In recent years, however, the number of hatchery raised RBT stocked from these naturalized populations in Incline and Third creeks has been greatly reduced (~2000 per year 2016-2020). In 2006, the only trout planted by CDFW were 2000 Golden trout (*O. mykiss aguabonita)* fingerling, stocked near Taylor Creek, which drains into Lake Tahoe in South Lake. Golden trout are distinctive in coloration and readily distinguishable from other hatchery RBT strains. From 2007-present, CDFW has stocked only kokanee salmon into Lake Tahoe (total 605,508 individuals). Beginning in 2011, NDOW switched to primarily planting infertile triploid RBT (191,265), which now account for 91% of all planted RBT. All samples genotyped in this study are from naturalized RBT caught in tributaries around the lake. Analysis of recent recapture records by NDOW biologists illustrated wide-ranging dispersal of RBT from two stocking locations in East Lake (Cave Rock and Sand Harbor) and mixing across all regions of the lake (Online Resource 3). These data, in addition to the preponderance of stocking Eagle Lake and naturalized RBT from the Lake Tahoe basin, strongly suggest that any observed genetic structure will not be an artefact of stocking history.

Alignment and variant calling

Following contaminant filtering and removal of individuals with insufficient sequencing data, we retained 150 individuals with a mean of 1,416,116 reads per individual. `bwa` mapped reads from all individuals to the *O. mykiss* reference genome (GenBank assembly accession GCA_900005705.1). After variant calling and filtering based on

sequencing coverage and quality, we retained 13,162 loci with MAF > 0.05. Additional filtering for potentially paralogous regions resulted in the removal of an additional 355 SNPs, leaving a final dataset of 12,807 SNPs. The final dataset consisted of 150 individuals with a mean coverage of 4.6X per locus per individual. Sequence data (fastq files) for each individual are deposited at the SRA of NCBI (accession PRJNA7497193), and the vcf file, sequence data, and other supplementary files are available at the DRYAD repository (doi: 10.5061/dryad.15dv41nzk).

Population genetic analyses

DIC values from `entropy` indicated that the $k = 2$ model best fit the data (Fig. 2, Online Resource 4), although the $k = 3$ and 4 models had similar support and illustrated additional aspects of spatial genetic structure (described below). PCA based on genotype probabilities revealed subtle genetic structure, where PC 1 explained 3.06% of the variation in the data while PC 2 explained only 1.60% (Fig. 1b, c). When plotted by region of the lake (North, West, South, McKinney), PCA revealed subtle but apparent differentiation of RBT from different sets of streams across the basin. Specifically, RBT in the three regions (North, West, South) exhibited differentiation, with McKinney Creek as a uniquely differentiated, fourth group (Figs. 1b, c, and Fig. 2). Finally, using AMOVA to analyze variance explained by among versus (a) within stream, (b) within region ($k = 3$; North, South, West), and (c) within region, with McKinney grouped separately ($k = 4$; North, South, West, McKinney), we found that 95.47% of variance explained was attributed to variation within individuals (Table 2).

Results from DAPC without *a priori* clustering are largely concordant with those from PCA and `entropy`, with the three regions (North, West, South) illustrating divergence, and with McKinney forming a uniquely differentiated cluster (Fig. 3a, b). The $k = 4$ analysis clustered individuals differently than analyses using PCA or `entropy`, with main clusters corresponding to three main clusters representing North Lake, McKinney Creek, and South and West Lake, and a fourth cluster representing few anomalous individuals (Fig. 3b, e, Online Resource 5). Analyses using *a priori* group assignments grouped populations more distinctly than in PCA (Fig. 3c, d).

Pairwise measures of genetic divergence indicated low levels of differentiation among sampled streams but were also consistent with evidence for differentiation among groups of streams from different regions of the basin (F_{ST} range = 0.0182 – 0.0487, Nei's *D* range = 0.0090 – 0.0271; Table 3). Genetic diversity for 150 individuals across 8 sampling sites was 0.00270 for Θ_π (range 0.00242 – 0.00287) and 0.00263 for Θ_W (range 0.00219 – 0.00311) (see Table 4). Genetic diversity was relatively consistent across regions, with North Lake mean $\Theta_\pi = 0.00275$ (range = 0.00271 – 0.00279) and mean $\Theta_W = 0.00269$ (range = 0.00262 – 0.00276), South Lake mean $\Theta_\pi = 0.00265$ (range = 0.00242 – 0.00287) and mean $\Theta_W = 0.00265$ (range = 0.00219 - .00311), West Lake (not including McKinney) mean $\Theta_\pi = 0.00270$ (range = 0.00261 – 0.0281) and mean $\Theta_W = 0.00263$ (range = 0.00248 – 0.00277), and McKinney $\Theta_\pi = 0.00269$ and $\Theta_W = 0.00245$. Interestingly, Tajima's *D* varied among all populations, including variation within regions of the lake (range = -0.3151 – 0.2683).

Discussion

Nonnative trout species pose multiple threats and challenges to conserving and restoring native trout fisheries. Predation, competition, and hybridization all represent obstacles to native CT populations and to the potential success of reintroductions (Dunham et al. 1997, 2003; Koel et al. 2005). These issues are especially acute in the Lake Tahoe basin, where native LCT populations were extirpated and replaced with multiple nonnative salmonid species. Both CT and RBT spawn in streams or near stream outflows (Arostegui and Quinn 2019), making both hybridization and competition for rearing and suitable spawning habitat problematic. For the inland CT subspecies and RBT, both stream resident and migratory life histories have been identified (Arostegui and Quinn 2019; Neville et al. 2006, 2016). Individuals with the migratory life history spend one or more years in the natal stream before migrating to lake or main stem river habitat for the remainder of maturity (Arostegui and Quinn 2019). Greater differences in habitat use within lakes arise as adults, where large-bodied, adult LCT inhabit pelagic zones (Al-Chokhachy et al. 2009; Meeuwig and Peacock 2017), while adult RBT primarily inhabit inshore habitats (Swales 2006), suggesting less interspecific competition in later adult stages. Thus, we aimed to gain a better understanding of where (or if) lake dwelling, naturalized RBT home for spawning and the extent of differentiation within and among streams to understand the possibilities for and implications of RBT removal and LCT reintroduction.

Here, we quantified fine scale population genetic structure and genetic diversity of naturalized RBT sampled from different tributary streams, establishing baseline population genetic data for naturalized RBT of the Lake Tahoe basin. Our analyses

revealed subtle but consistent population differentiation among tributaries, with individuals from spatially proximate populations primarily grouping closer together than others across multiple population genetic analyses (Figs. 1, 2, 3). Differentiation became more evident as populations were grouped by geographic region (Figs. 1, 2, 3). Many stream beds were entirely dry during the sampling period, meaning RBT that may have attempted spawning in these tributaries had to move to the lake for the remainder of the dry season. Additionally, the few streams of East Lake where sampling was possible were predominantly or entirely inhabited by other nonnative, competitor species (e.g., brook trout, *Salvelinus fontinalis*).

Identifying plausible explanations for this population genetic structure presents a challenge in this system, where nonnative RBT introductions of fish from mixed sources were consistent, deliberate, and widespread, . Possibilities include long-term homing behavior to certain tributaries, which has been observed in some (but not all) stocked populations of RBT in other systems (Biette et al. 1981; Quinn 1993; Schroeder et al. 2001). Homing behavior by planted nonnative trout could arise from overwintering adults seeking spawning habitat in the spring. The larger tributaries with more consistent water flow would be natural targets for spawning adults and eventually act as natal streams for homing behavior of surviving, reproducing offspring.

We originally hypothesized that genetic structure may be related to the historical stocking of genetically differentiated strains into specific tributaries or regions. Earlier work, however, has shown that in streams where multiple hatchery raised RBT strains were planted over multiple years, the resulting naturalized populations were thoroughly admixed between otherwise genetically distinct strains (Kirchoff 2016; Peacock and

Kirchoff 2004). Additionally, a study on naturalized RBT in the mainstem Truckee River, which flows from Lake Tahoe to Pyramid Lake, showed that admixture among hatchery strains was evident along the entire length of the river (though subtle genetic population structure corresponding to instream barriers was also observed; Kirchoff 2016). Review of the Tahoe stocking records also suggests that the strains and stocking events are unlikely to have caused genetic differentiation we observed as stocking events regularly included individuals of mixed ancestry, with multiple stocking events within the same year and recaptured fish routinely found throughout the basin. We found substantial overlap in strains planted by both CDFW and NDOW, with no convergence of patterns between stocked strains and stocking locations with population genetic differentiation. Further, there have been no RBT stocking events on the California side of the lake since 2007, and primarily sterile, triploid RBT have been stocked on the Nevada side since mid-2011 (with few exceptions of naturalized RBT from North Lake; see Results). Combined with the average life expectancy of 3-5 years and wild RBT reproductive maturity at approximately three years of age (McAfee 1966), we can infer that RBT currently occurring in the Lake Tahoe basin are naturalized populations that are reproducing across the landscape and are now assorted in the patterns we observed.

As stocking events of reproductively viable RBT have almost entirely ceased in the past decade, it is possible that we observed a snapshot of the ongoing, increasing admixture of populations from different genetic sources. Alternatively, we may have identified nascent population structure developing as a result of what may be reduced survival of stocked individuals that were maladapted to local environments, combined with drift arising from naturalized RBT that are philopatric to individual streams. Some

combination of these explanations is likely, as most stocked individuals are unlikely to survive and reproduce, something that has been documented across multiple stocked salmonid species (Brunner et al. 1998; Heggenes et al. 2002; Benavente et al. 2015). However, given that the Eagle Lake strain of RBT together with hatchery raised RBT from naturalized populations within the Lake Tahoe basin represent the majority of stocked trout since 1980, the population structure we observed strongly suggests an emergent homing behavior – if not to particular streams – to particular regions of the lake, where habitat attributes may influence the spatial dispersion of RBT.

Such genetic differentiation due to habitat variation has been observed in a large number of other studies. In a study on population genetic structure among coastal cutthroat trout populations, Wenburg et al. (1998) showed that genetic clustering was explained by physiogeographic region, and although individual stream homing was important, the dynamic between gene flow among streams and genetic drift accounted for the regional pattern. Further, Gresswell et al. (1994) showed that life history traits including size, age, migration strategy, and migration timing among spawning populations of Yellowstone cutthroat trout (*O. clarkii bouvieri*) in tributary streams of Yellowstone Lake vary considerably across populations, suggesting genetic differentiation likely driven by habitat variability. Landscape features also influence population genetic structure among life history types of rainbow trout (anadromous and resident) in the Pacific Northwest (Narum et al. 2008). Instream landscape characteristics explained patterns of genetic variation among resident populations, while isolation by distance explained patterns in anadromous forms, which in many instances were confined

to the lower reaches of these watersheds by high gradient streams and barriers to upstream movement (Narum et al. 2008).

Results from previous studies are consistent with recent comparison of two different systems in Patagonia where naturalized RBT populations flourish. The first lake represents the largest RBT smolt producer in Chile, where aquaculture escapees regularly establish and integrate with naturalized populations of RBT (Canales-Aguirre et al. 2018). The other, contrasting lake is protected via national park and reserve status, where aquaculture is prohibited and only two stocking events since 1900 occurred, leading to the development of a fully naturalized population (Canales-Aguirre et al. 2018). In this case, the lake permitting aquaculture yielded reduced genetic structure and higher genetic diversity stemming from multiple genetic sources of farmed individuals, while the protected lake showed marked genetic structure and reduced genetic diversity from lower propagule pressure (Canales-Aguirre et al. 2018). Our results likely fall somewhere between these two extremes as we found subtle but consistent population structure potentially influenced by intermittent stocking events in the last 50 years along with variation in genetic diversity estimates (Table 4). The comparison between results of the current and future studies could provide a perspective vital for LCT reintroduction success.

Our results carry substantial implications for conservation and management of LCT in this system, where both hybridization and ecological challenges that have arisen from an altered food web constrain recovery of the native apex predator to this ecosystem. Amelioration of the effects of nonnative salmonids in the western United States has led to substantial research into both the impacts of these nonnative fishes and

ways to remove them (Al-Chokhachy et al. 2009; Meeuwig and Peacock 2017; Ruzycki et al. 2003; Syslo et al. 2011). A case study in lake trout removal for Yellowstone CT conservation illustrates the challenges of removing entrenched nonnative salmonids: despite decades of gillnet removal efforts and over 450,000 pounds of lake trout removed from Yellowstone Lake, the lake trout population continues to increase (Syslo et al. 2011). Similar gillnetting efforts in Fallen Leaf Lake, a small subalpine lake which drains into Lake Tahoe, where LCT have been reestablished, have reduced the average size of nonnative lake trout and therefore some of the predation threat for LCT, but have not eliminated them (Al-Chokhachy et al. 2020). Instream electrofishing removal of nonnative salmonid populations across ecosystems also yields mixed results, with success somewhat dependent upon habitat size (Brunson 2020; Pacas and Taylor 2015; Rytwinski et al. 2019). Such mechanical removal efforts are unlikely to be wholly successful when used in isolation. Still, the ecological similarities with CT including stream spawning life history offer opportunities to target resident RBT stream populations as well as lake dwelling RBT entering the streams for spawning. For example, based upon pilot studies conducted in the smaller Fallen Leaf Lake ecosystem (Al-Chokhachy et al. 2009; Meeuwig and Peacock 2017), current USFWS management activities include: (1) the use of weirs to reduce RBT access to the single available spawning stream, (2) instream electrofishing removal of adults, and (3) redd disruption to reduce RBT recruitment into the Fallen Leaf Lake population (Al-Chokhachy et al. 2020). The efficacy of such an approach as a long-term solution to the hybridization threat posed by RBT remains unknown, even in this smaller ecosystem, and will require ongoing genetic monitoring. The results of the study in Fallen Leaf Lake may help refine the effectiveness of such an

approach and allow expansion to the much larger Lake Tahoe and similar sized landscapes.

Moving forward, tracking changes to RBT genetic structure on a temporal scale spanning time periods long enough to observe any potential changes in genetic structure (i.e., across multiple generations) would increase our understanding of the population dynamics and habitat utilization in this complex, altered system and could help fine tune reintroduction strategies for LCT in Lake Tahoe. Should the observed patterns represent emergent RBT population genetic structure based upon homing behavior to natal streams, targeted RBT removal (mechanical and/or chemical) would open habitat for LCT reintroduction and potentially facilitate similar homing behavior to tributaries for spawning LCT. From there, active management of trout entering the streams to spawn using weirs together with ongoing genetic monitoring of reestablished LCT stream populations may allow this iconic trout to gain a foot hold in its historic habitat.

Data availability

The datasets generated for this study are available at the Dryad Digital Repository (doi: 10.5061/dryad.15dv41nzk; https://doi.org/10.5061/dryad.15dv41nzk) and NCBI's Short Read Archive (accession PRJNA7497193; https://dataview.ncbi.nlm.nih.gov/object/PRJNA7497193).

References

Abell RA, Olson DM, Dinerstein E, Hurley PT, Eichbaum W, Diggs JT, Walters S, Wettengel W, Allnutt T, Loucks CJ, Hedao P (2000) Freshwater ecoregions of North America: a conservation assessment, Vol. 2. Island Press.

Al-Chokhachy R, Peacock M, Heki LG, Thiede G (2009) Evaluating the reintroduction potential of Lahontan cutthroat trout in Fallen Leaf Lake, California. North Am J Fish Manag 29(5):1296–1313.

Al-Chokhachy R, Heki L, Loux T, Peka R (2020) Return of a giant: coordinated conservation leads to the first wild reproduction of Lahontan Cutthroat Trout in the Truckee River in nearly a century. Fisheries 45(2):63–73.

Allendorf FW, Leary RF (1988) Conservation and distribution of genetic variation in a polytypic species, the cutthroat trout. Conserv Biol 2(2):170–184.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16(11):613–622.

Altukhov YP, Salmenkova EA, Omelchenko VT (2008). Salmonid fishes: population biology, genetics, and management. John Wiley and Sons.

Arostegui MC, Quinn TP (2019) Reliance on lakes by salmon, trout and charr (*Oncorhynchus*, *Salmo* and *Salvelinus*): an evaluation of spawning habitats, rearing strategies and trophic polymorphisms. Fish Fish 20(4):775–794.

Benavente JN, Seeb LW, Seeb JE, Arismendi I, Hernandez CE, Gajardo G, Galleguillos R, Cadiz MI, Musleh SS, Gomez-Uchida D (2015) Temporal genetic variance and propagule-driven genetic structure characterize naturalized rainbow trout (*Oncorhynchus mykiss*) from a Patagonian lake impacted by trout farming. PloS One 10(11): e0142040.

Biette RM, Dodge DP, Hassinger RL, Stauffer TM (1981) Life history and timing of migrations and spawning behavior of rainbow trout (*Salmo gairdneri*) populations of the Great Lakes. Can J Fish Aquat Sci 38(12):1759–1771.

Blankenship SM, Campbell MR, Hess JE, Hess MA, Kassler TW, Kozfkay CC, Matala AP, Narum SR, Paquin MM, Small MP, Stephenson JJ, Warheit KI, Moran P (2011) Major lineages and metapopulations in Columbia River Oncorhynchus mykiss are structured by dynamic landscape features and environments. Trans Am Fish Soc 140(3):665–684.

Bonar SA, Bolding BD, Divens M, Meyer W (2005) Effects of introduced fishes on wild juvenile coho salmon in three shallow Pacific Northwest lakes. Trans Am Fish Soc 134(3):641–652.

Brunner PC, Douglas MR, Bernatchez L (1998) Microsatellite and mitochondrial DNA assessment of population structure and stocking effects in Arctic charr *Salvelinus alpinus* (Teleostei: Salmonidae) from central Alpine lakes. Mol Ecol 7(2):209–223.

Brunson C (2020) Evaluating the Effect of the Removal of Non-Native Trout in Two High Elevation Tributary Streams in the Intermountain West. Thesis, Utah State University.

Buerkle CA, Gompert Z (2013) Population genomics based on low coverage sequencing: how low should we go? Mol Ecol 22(11):3028–3035.

Campbell MR, Dillon J, Powell MS (2002) Hybridization and introgression in a managed, native population of Yellowstone cutthroat trout: genetic detection and management implications. Trans Am Fish Soc 131(3):364–375.

Canales-Aguirre CB, Seeb LW, Seeb JE, Cádiz MI, Musleh SS, Arismendi I, Gajardo G, Galleguillos R, Gomez-Uchida D (2018). Contrasting genetic metrics and patterns among naturalized rainbow trout (*Oncorhynchus mykiss*) in two Patagonian lakes differentially impacted by trout aquaculture. Ecol Evol 8(1):273–285.

Clavero M, Ninyerola M, Hermoso V, Filipe AF, Pla M, Villero D, Brotons L, Delibes M (2017) Historical citizen science to understand and predict climate-driven trout decline. Proc R Soc B Bio. Sci 284(1846):20161979.

Coffin PD, Cowan WF (1995) Lahontan cutthroat trout (*Oncorhynchus clarki henshawi*) recovery plan. US Fish and Wildlife Service, Region 1. Portland, Oregon, USA.

Cordone AJ, Frantz TC (1968) An evaluation of trout planting in Lake Tahoe. Calif Fish Game 54:68–89.

Costello MJ (2009) How sea lice from salmon farms may cause wild salmonid declines in Europe and North America and be a threat to fishes elsewhere. Proc R Soc B Biol Sci 276(1672):3385–3394.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158.

Dauwalter DC, Duchi A, Epifanio J, Gandolfi A, Gresswell R, Juanes F, Kershner J, Lobón-Cerviá J, McGinnity P, Meraner A, Mikheev P, Morita K, Muhlfeld CC, Pinter K, Post JR, Unfer G, Vøllestad LA, William JE (2020) A call for global action to conserve native trout in the 21st century and beyond. Ecol Freshw Fish 29(3):429–432.

Dunham JB, Vinyard GL, Rieman BE (1997) Habitat fragmentation and extinction risk of Lahontan cutthroat trout. North Am J Fish Manag 17(4):1126–1133.

Dunham JB, Rieman BE, Peterson JT (2002) Patch-based models to predict species occurrence: lessons from salmonid fishes in streams. Predicting Species Occurrences: Issues of Scale and Accuracy. Island Press, Covelo, California, pp 327–334.

Dunham J, Schroeter R, Rieman B (2003) Influence of maximum water temperature on occurrence of Lahontan cutthroat trout within streams. North Am. J. Fish. Manag 23(3):1042–1049.

Fagan WF, Aumann C, Kennedy CM, Unmack PJ (2005) Rarity, fragmentation, and the scale dependence of extinction risk in desert fishes. Ecology 86(1):34–41.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4):1567–1587.

Ford JS, Myers, RA (2008) A global assessment of salmon aquaculture impacts on wild salmonids. PLoS Biol 6(2):0411–0417.

Fumagalli M, Vieira FG, Korneliussen TS, Linderoth T, Huerta-Sánchez E, Albrechtsen A, Nielsen R (2013) Quantifying population genetic differentiation from next-generation sequencing data. Genetics 195(3):979–992.

Gerstung ER (1988) Status, life history, and management of the Lahontan cutthroat trout. American

Fisheries Society Symposium 4:93–106.

Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, Nice CC (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Mol Ecol 23(18):4555–4573.

Gresswell RE, Liss WJ, Larson GL (1994) Life-history organization of the Yellowstone cutthroat trout (*Oncorhynchus clarki bouvieri*) in Yellowstone Lake. Can J Fish Aquat Sci 51(S1)298–309.

Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM (2013) Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. G3 Genes, Genomes, Genet 3(8):1273–1285.

Hand BK, Muhlfeld CC, Wade AA, Kovach RP, Whited DC, Narum SR, Matala AP, Ackerman MW, Garner BA, Kimball JS, Stanford JA, Luikhart G (2015). Mol Ecol 25(3):689–705.

Heath DD, Pollard S, Herbinger C (2001) Genetic structure and relationships among steelhead trout (*Oncorhynchus mykiss*) populations in British Columbia. Heredity 86(5):618–627.

Heggenes J, Røed KH, Høyheim B, Rosef L (2002) Microsatellite diversity assessment of brown trout (*Salmo trutta*) population structure indicate limited genetic impact of stocking in a Norwegian alpine lake. Ecol Freshw Fish 11(2):93–100.

Hoyer AB, Schladow SG, Rueda FJ (2015) Local dispersion of nonmotile invasive bivalve species by wind-driven lake currents. Limnol Oceanogr 60(2):446–462.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132(2):583–589.

Jeton AE (1999) Precipitation-Runoff Simulations for the Lake Tahoe Basin, California and Nevada. Vol. 99, No. 4110. US Department of the Interior, US Geological Survey.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11(1):94.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21):3070–3071.

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

Kareiva P, Marvier M, McClure M (2000) Recovery and management options for spring/summer chinook salmon in the Columbia River Basin. Science 290(5493):977–979.

Keefer ML, Caudill CC (2014) Homing and straying by anadromous salmonids: a review of mechanisms and rates. Rev fish Biol Fish 24(1):333–368.

Kelley DW (1957) A report on Lake Tahoe and its tributaries: fisheries management vs. trial and error, California Department of Fish and Game.

Kirchoff VS (2016) Impact of triploid Rainbow Trout and naturalized Rainbow Trout (*Oncorhynchus*

mykiss) on recovery of Lahontan Cutthroat Trout (*Oncorhynchus clarkii henshawi*) in the Truckee River watershed. Dissertation, University of Nevada, Reno.

Koel TM, Bigelow PE, Doepke PD, Ertel BD, Mahony DL (2005) Nonnative lake trout result in Yellowstone cutthroat trout decline and impacts to bears and anglers. Fisheries 30(11):10–19.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's *D* and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics 14(1):289.

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15(1):356.

Kruse CG, Hubert WA, Rahel FJ (2000) Status of Yellowstone cutthroat trout in Wyoming waters. North Am J Fish Manag 20(3):693–705.

Labar GW (1971) Movement and homing of cutthroat trout (*Salmo clarki*) in Clear and Bridge creeks, Yellowstone National Park. Trans Am Fish Soc 100(1):41–49.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357.

Leary RF, Allendorf FW, Phelps SR, Knudsen KL (1987) Genetic divergence and identification of seven cutthroat trout subspecies and rainbow trout. Trans Am Fish Soc 116(4):580–587.

Leitritz E (1970) A history of California's fish hatcheries, 1870-1960, Vol. 150. California Department of Fish and Game.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Proc (2009) The sequence alignment map format and SAMtools. Bioinformatics 25(16):2078–2079.

Lowe S, Browne M, Boudjelas S, De Poorter M (2000) 100 of the world's worst invasive alien species: a selection from the global invasive species database, Vol. 12. Auckland: Invasive Species Specialist Group.

Mandeville EG, Walters AW, Nordberg BJ, Higgins KH, Burckhardt JC, Wagner CE (2019) Variable hybridization outcomes in trout are predicted by historical fish stocking and environmental context. Mol Ecol 28(16): 3738–3755.

Maxwell D, Jennings S (2005) Power of monitoring programmes to detect decline and recovery of rare and vulnerable fish. J Appl Ecol 42(1):25–37.

McAfee WR (1966) Rainbow trout. Inl Fish Manag, CDFG:192–215.

McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017) Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. Mol Ecol Resour 17(4):656–669.

Meeuwig MH, Peacock MM (2017) Food web interactions associated with a Lahontan cutthroat trout reintroduction effort in an alpine lake. J Fish Wildl Manag 8(2):449–464.

Moyle PB, Katz JVE, Quiñones RM (2011) Rapid decline of California's native inland fishes: a status

assessment. Biol Conserv 144(10):2414–2423.

Muhlfeld CC, Kalinowski ST, McMahon TE, Taper ML, Painter S, Leary RF, Allendorf FW (2009) Hybridization rapidly reduces fitness of a native trout in the wild. Biol Lett 5(3):328–331.

Muhlfeld CC, Thorrold SR, McMahon TE, Marotz B (2012) Estimating westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) movements in a river network using strontium isoscapes. Can J Fish Aquat Sci 69(5):906–915.

Muhlfeld CC, Dauwalter DC, D'Angelo VS, Ferguson A, Giersch JJ, Impson D, Koizumi I, Kovach R, McGinnity P, Schöffmann J, Vøllestad LA, Epifanio J (2019) Global Status of Trout and Char: Conservation Challenges in the Twenty-First Century. In Trout and char of the world. American Fisheries Society. Bethesda, Maryland.

Narum SR, Zendt JS, Graves D, Sharp WR (2008) Influence of landscape on resident and anadromous life history types of *Oncorhynchus mykiss*. Can J Fish Aquat Sci 65(6):1013–1023.

Nei M (1972) Genetic distance between populations. Am Nat 106(949):192–283.

Neville HM, Dunham JB, Peacock MM (2006) Landscape attributes and life history variability shape genetic structure of trout populations in a stream network. Landsc Ecol 21(6):901–916.

Neville HM, Dauwalter D, Peacock M (2016) Monitoring demographic and genetic responses of a threatened inland trout to habitat reconnection. Trans Am Fish Soc 145(3):610–626.

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12(6):443–451.

Ostberg CO, Rodriguez RJ (2004) Bi-parentally inherited species-specific markers identify hybridization between rainbow trout and cutthroat trout subspecies. Mol Ecol Notes 4(1):26–29.

Otero J, L'Abée-Lund JH, Castro-Santos T, Leonardsson K, Storvik GO, Jonsson B, Dempson B, Russell IC, Jensen AJ, Baglinière J-L, Dionne M, Armstrong JD, Romakkaniemi A, Letcher BH, Kocik JF, Erkinaro J, Poole R, Rogan G, Lundqvist H, MacLean JC, Jokikokko E, Arnekleiv JV, Kennedy RJ, Niemelä E, Caballero P, Music PA, Antonsson T, Gudjonsson S, Veselov AE, Lamberg A, Groom S, Taylor BH, Taberner M, Dillane M, Arnason F, Horton G, Hvidsten NA, Jonsson IR, Jonsson N, McKelvey S, Næsje TF, Skaala O, Smith GW, Sægrov H, Stenseth NC, Vøllestad LA (2014) Basin-scale phenology and effects of climate variability on global timing of initial seaward migration of Atlantic salmon (*Salmo salar*). Glob Chang Biol 20(1):61–75.

Pacas C, Taylor MK (2015) Nonchemical eradication of an introduced trout from a headwater complex in Banff National Park, Canada. North Am J Fish Manag 35(4):748–754.

Parchman TL, Gompert Z, Mudge J, Schilkey F, Benkman CW, Buerkle CA (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol Ecol 21(12):2991–3005.

Peacock MM, Kirchoff V (2004) Assessing the conservation value of hybridized cutthroat trout populations in the Quinn River drainage, Nevada. Trans Am Fish Soc 133(2):309–325.

Peacock MM, Neville H, Finger AJ (2018) The Lahontan Basin evolutionary lineage of cutthroat trout. Cutthroat trout Evol Biol Taxon 231–260.

Peterson DP, Fausch KD, White GC (2004) Population ecology of an invasion: effects of brook trout on native cutthroat trout. Ecol Appl 14(3):754–772.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PloS One 7(5):e37135.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Quinn TP (1993) A review of homing and straying of wild and hatchery-produced salmon. Fish Res 18(1–2): 29–44.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. Annu Rev Ecol Syst 27(1):83–109.

Ruzycki JR, Beauchamp DA, Yule DL (2003) Effects of introduced lake trout on native cutthroat trout in Yellowstone Lake. Ecol Appl 13(1):23–37.

Rytwinski T, Taylor JJ, Donaldson LA, Britton JR, Browne DR, Gresswell RE, Lintermans M, Prior KA, Pellatt MG, Vis C, Cooke SJ (2019) The effectiveness of non-native fish removal techniques in freshwater ecosystems: a systematic review. Environ Rev 27(1):71–94.

Sanderson BL, Barnas KA, Rub AMW (2009) Nonindigenous species of the Pacific Northwest: an overlooked risk to endangered salmon? BioScience 59(3):245–256.

Schley B (1971) A century of fish conservation (1871-1971). US Fish and Wildlife Service. https://nctc.fws.gov/History/Articles/FisheriesHistory.html. Accessed 1 January 2022.

Schroeder RK, Lindsay RB, Kenaston KR (2001) Origin and straying of hatchery winter steelhead in Oregon coastal rivers. Trans Am Fish Soc 130(3):431–441.

Sepulveda AJ, Rutz DS, Dupuis AW, Shields PA, Dunker KJ (2015) Introduced northern pike consumption of salmonids in Southcentral Alaska. Ecol Freshw Fish 24(4):519–531.

Sigler JW, Sigler WF (1986) History of fish hatchery development in the Great Basin states of Utah and Nevada. Gt Basin Nat 583–594.

Steissberg TE, Hook SJ, Schladow SG (2005) Measuring surface currents in lakes with high spatial resolution thermal infrared imagery. Geophys Res Lett 32(11):L11402.

Swales S (2006) A review of factors affecting the distribution and abundance of rainbow trout (*Oncorhynchus mykiss* Walbaum) in lake and reservoir systems. Lake Reserv Manag 22(2):167–178.

Syslo JM, Guy CS, Bigelow PE, Doepke PD, Ertel BD, Koel TM (2011) Response of non-native lake trout (*Salvelinus namaycush*) to 15 years of harvest in Yellowstone Lake, Yellowstone National Park. Can J Fish Aquat Sci 68(12):2132–2145.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105(2):437–460.

Thompson RS, Benson L, Hattori EM (1986) A revised chronology for the last Pleistocene lake cycle in the central Lahontan Basin. Quat Res 25(1):1–9.

Todesco M, Pascual MA, Owens GL, Ostevik KL, Moyers BT, Hübner S, Heredia SM, Hahn MA, Caseys

C, Bock DG, Rieseburg LH (2016) Hybridization and extinction. Evol Appl 9(7):892–908.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7(2):256–276.

Weigel DE, Connolly PJ, Martens KD, Powell MS (2013) Colonization of steelhead in a natal stream after barrier removal. Trans Am Fish Soc 142(4):920–930.

Wenburg JK, Bentzen P, Foote CJ (1998) Microsatellite analysis of genetic population structure in an endangered salmonid: the coastal cutthroat trout (*Oncorhynchus clarkii clarkii*). Mol Ecol 7(6):733–749.

Wenger SJ, Isaak DJ, Luce CH, Neville HM, Fausch KD, Dunham JB, Dauwalter DC, Young MK, Elsner MM, Rieman BE, Hamlet AF, William JE (2011) Flow regime, temperature, and biotic interactions drive differential declines of trout species under climate change. Proc Natl Acad Sci 108(34):14175–14180.

Wenger SJ, Leasure DR, Dauwalter DC, Peacock MM, Dunham JB, Chelgren ND, Neville HM (2017) Viability analysis for multiple populations. Biol Conserv 216:69–77

Table 1. Sampling information for RBT from streams in the Lake Tahoe basin. Here, streams are grouped according to geographic region.

Region	Stream	Latitude, Longitude	*N*
North	Incline (IN)	39.24, -119.94	13
	Third (TH)	39.24, -119.95	25
West	Ward (WR)	39.14, -120.20	16
	Blackwood (BK)	39.11, -120.18	16
	Madden (MD)	39.09, -120.18	23
	McKinney (MC)	39.06, -120.15	23
South	Taylor (TA)	38.93, -120.05	7
	Upper Truckee (UT)	38.86, -120.03	27

Table 2. AMOVA results of both $k = 3$ (North, South, West) and $k = 4$ (North, South, West, McKinney), where df represents degrees of freedom and PVE represents percent of variance explained.

		df	sum of squares	variance (σ)	PVE	*p*-value
$k = 3$	between populations	2	5425	13.1	1.057	0.02
	between samples within populations	3	9626	43.08	3.475	0.01
	within samples	142	168051	1183.46	95.468	0.01
	total	149	183102	1239.64	100	NA
$k = 4$	between populations	3	8254	24.11	1.945	0.02
	between samples within populations	4	6796	32.08	2.588	0.01
	within samples	142	168051	1183.46	95.468	0.01
	total	149	183102	1239.64	100	NA

Table 3. Pairwise estimates of Nei's *D* (Nei, 1972) and mean F_{ST} (Hudson et al., 1992) among all sampling sites. F_{ST} estimates are given on the upper diagonal, and Nei's *D* on the lower diagonal. Population abbreviations correspond to those in Table 1.

	BK	IN	MC	MD	TA	TH	UT	WR
BK	-	0.0375	0.0314	0.0198	0.0459	0.0276	0.0217	0.0252
IN	0.0199	-	0.0294	0.0302	0.0487	0.0224	0.0326	0.0365
MC	0.0158	0.0148	-	0.0248	0.0483	0.0229	0.0302	0.0334
MD	0.0099	0.0161	0.0120	-	0.0414	0.0217	0.0182	0.0208
TA	0.0254	0.0271	0.0268	0.0234	-	0.0405	0.0401	0.0452
TH	0.0142	0.0112	0.0109	0.0104	0.0229	-	0.0231	0.0283
UT	0.0109	0.0161	0.0154	0.0090	0.0222	0.0111	-	0.0225
WR	0.0122	0.0192	0.0173	0.0102	0.0241	0.0144	0.0107	-

Table 4. Population genetic diversity of 150 individuals from 8 populations based directly on DNA sequence variation across sampled genomic regions. Below are estimates of θ_π, or the average number of pairwise differences between sequences, θ_W, or the number of segregating sites, and Tajima's *D*, or the scaled difference between these two. Confidence intervals (95%) are included in high and low columns.

Region	Pop	*N*	H_E	θ_π	θ_π low	θ_π high	θ_W	θ_W low	θ_W high	*D*	*D* low	*D* high
North	IN	13	0.221	0.00271	0.00270	0.00272	0.00262	0.00261	0.00263	0.0224	0.0132	0.0315
	TH	25	0.226	0.00279	0.00278	0.00280	0.00276	0.00275	0.00277	-0.0436	-0.0523	-0.0348
West	WR	16	0.214	0.00261	0.00260	0.00263	0.00248	0.00247	0.00249	0.0754	0.0661	0.0846
	BK	16	0.220	0.00269	0.00268	0.00270	0.00265	0.00264	0.00266	-0.0379	-0.0475	-0.0284
	MD	23	0.222	0.00281	0.00280	0.00282	0.00277	0.00277	0.00278	-0.0497	-0.0588	-0.0406
	MC	23	0.222	0.00269	0.00268	0.00270	0.00245	0.00244	0.00245	0.2107	0.2015	0.2199
South	TA	7	0.206	0.00242	0.00241	0.00243	0.00219	0.00218	0.00220	0.2683	0.2595	0.2770
	UT	27	0.216	0.00287	0.00286	0.00288	0.00311	0.00310	0.00311	-0.3151	-0.3240	-0.3062

Figure legends

Fig. 1 Lake Tahoe basin rainbow trout sampling sites (a). PCA of genotype probabilities from `entropy`, based on 12,807 SNPs from 150 samples across 8 sites, grouped and colored by region (b) and stream (c).

Fig. 2 Estimates of ancestry coefficients (*q*) generated with `entropy` are scaled along the y-axis for each of *k* = 2 (a), *k* = 3 (b), and *k* = 4 (c) models. Vertical bars represent individuals, and colors represent ancestry proportion for *k* clusters. Population abbreviations correspond to those in Table 1. Here, the *k* = 2 model best fit the data, supported by the lowest deviance information criterion (DIC; Online Resource 4).

Fig. 3 Discriminant analyses of principal components without *a priori* clustering based on the `find.clusters` function (a, b), and with *a priori* cluster assignment (c, d), where plots a and c represent *k* =3, and b and d represent *k* = 4 models. Panel e represents the *k* = 4 model without *a priori* clustering, where vertical bars represent individuals, and colors represent proportion of ancestry from each ancestral group.

Figure 1

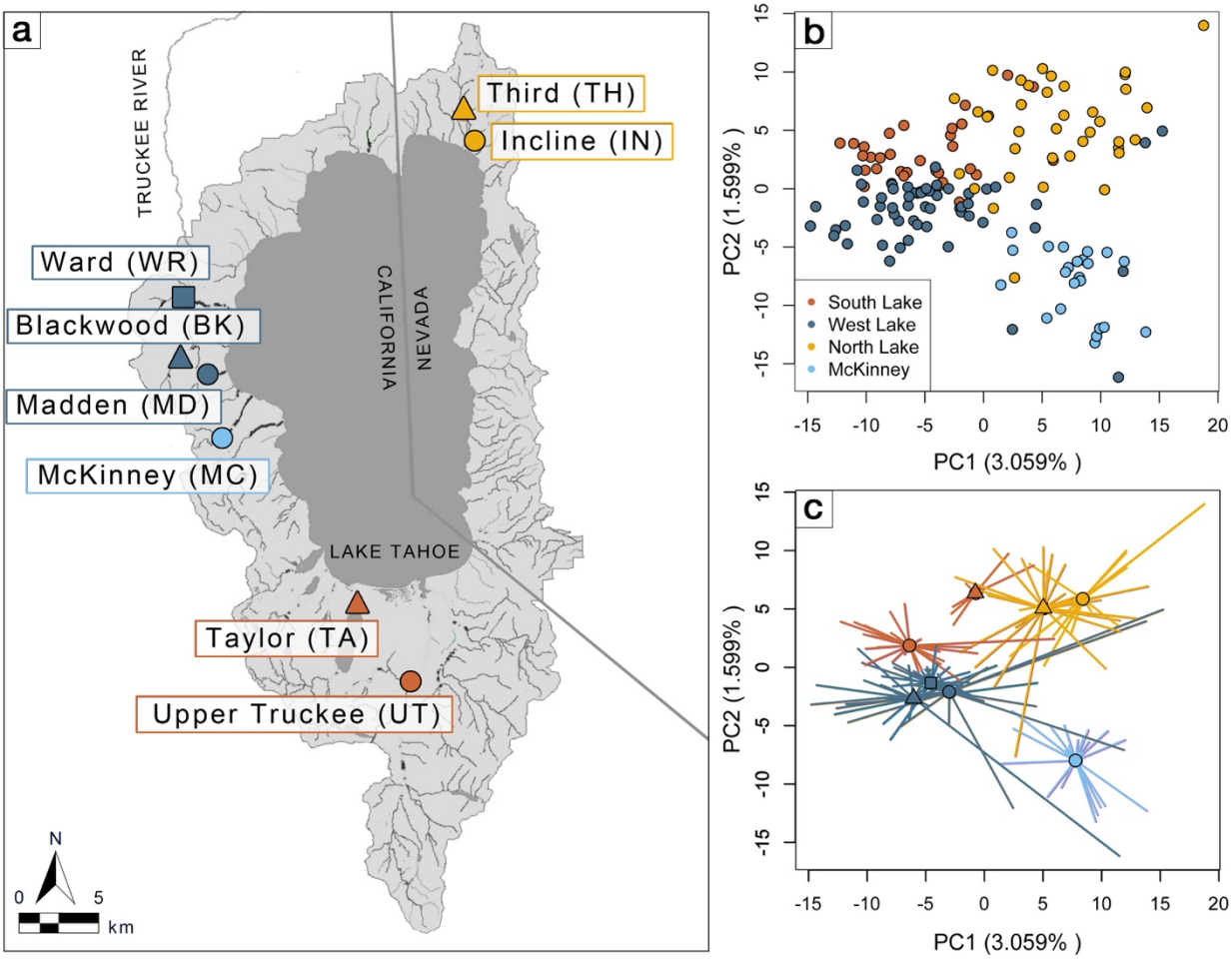

a
TRUCKEE RIVER
Third (TH)
Incline (IN)
Ward (WR)
Blackwood (BK)
Madden (MD)
McKinney (MC)
CALIFORNIA
NEVADA
LAKE TAHOE
Taylor (TA)
Upper Truckee (UT)
N
0
5
km
b
South Lake
West Lake
North Lake
McKinney
PC2 (1.599%)
PC1 (3.059%)
c
PC2 (1.599%)
PC1 (3.059%)

Figure 2

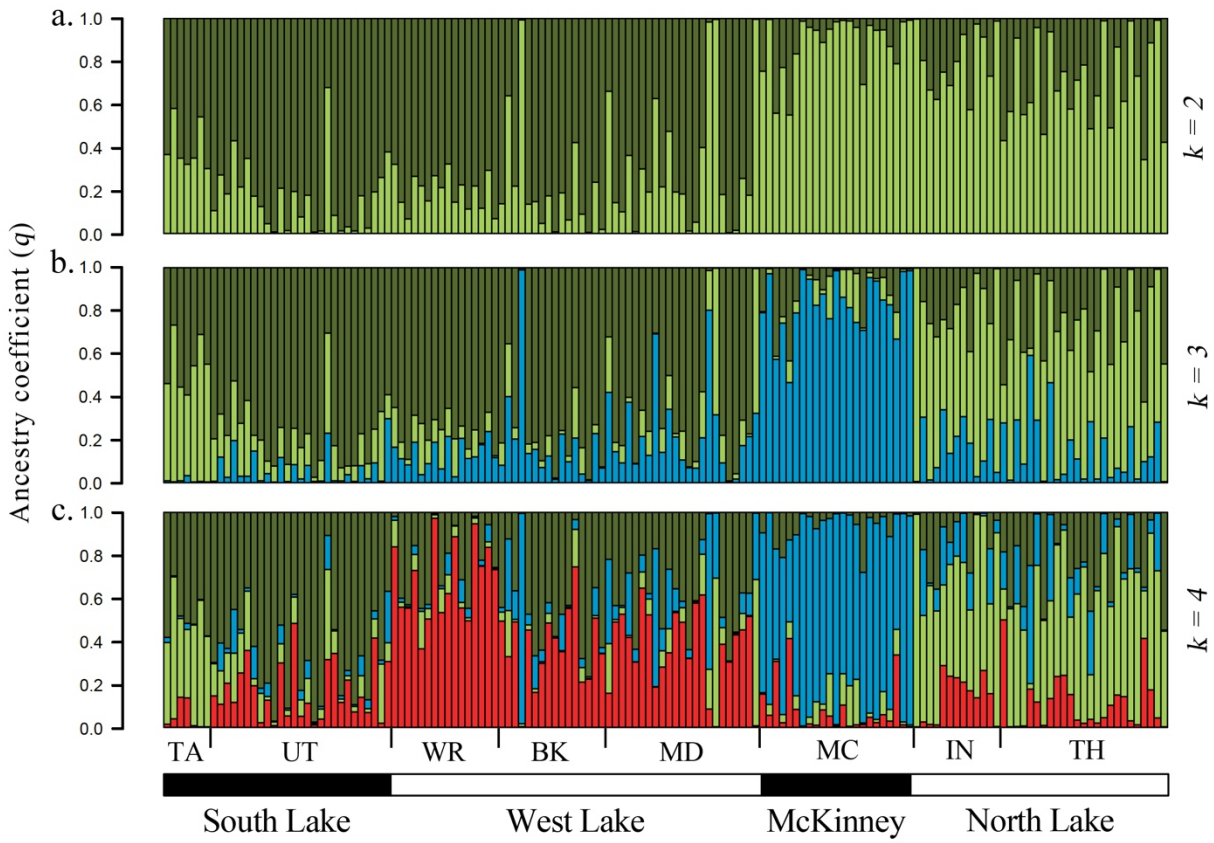
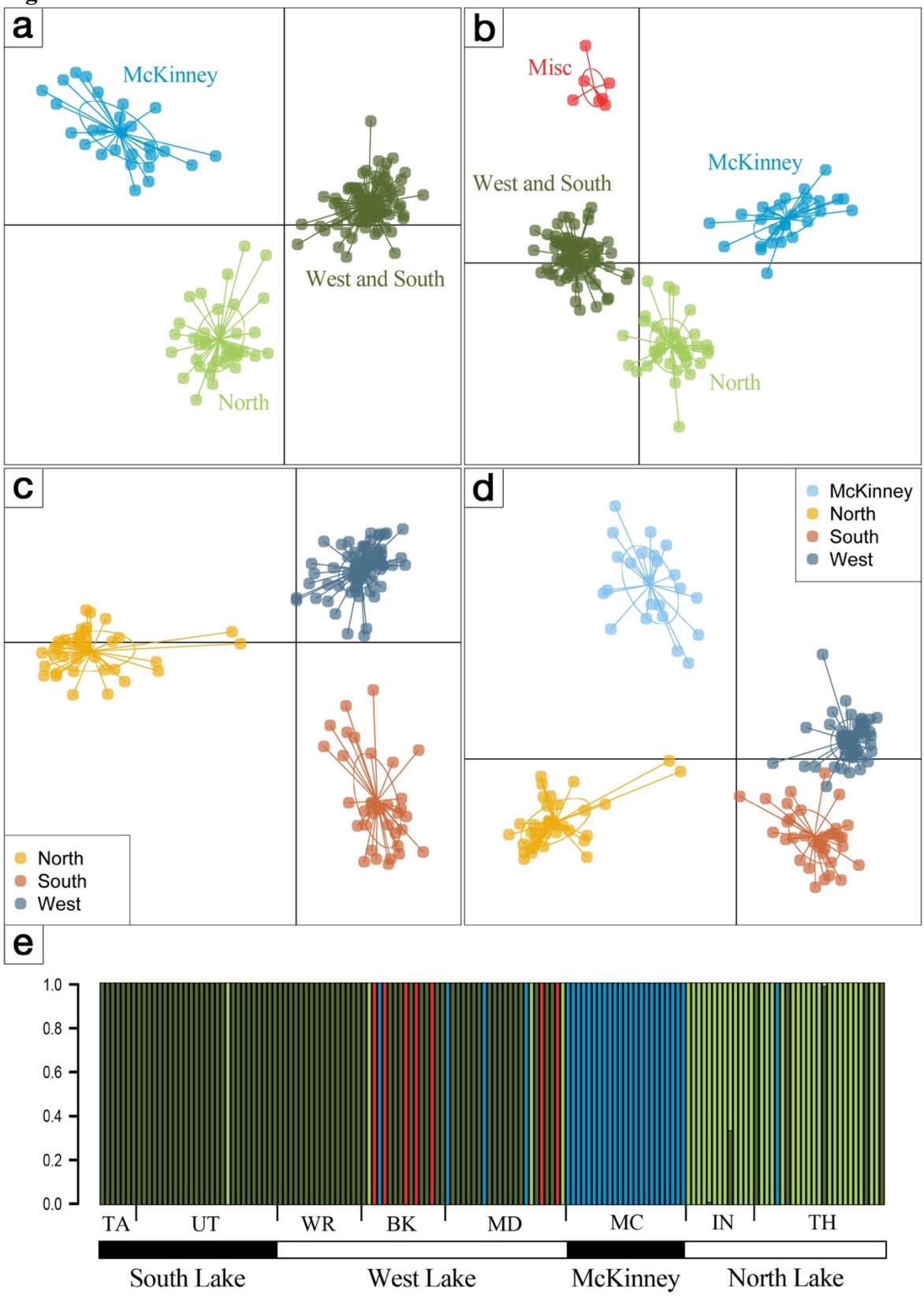

a.
b.
c.
Ancestry coefficient (q)
1.0
0.8
0.6
0.4
0.2
0.0
k = 2
k = 3
k = 4
TA
UT
WR
BK
MD
MC
IN
TH
South Lake
West Lake
McKinney
North Lake

Figure 3


a
McKinney
West and South
North
b
Misc
West and South
McKinney
North
c
North
South
West
d
McKinney
North
South
West
e
1.0
0.8
0.6
0.4
0.2
0.0
TA
UT
WR
BK
MD
MC
IN
TH
South Lake
West Lake
McKinney
North Lake

Supplementary material

Chapter 2: Assessing the population genetic structure of introduced rainbow trout (*Oncorhynchus mykiss*) in the Lake Tahoe basin: a case for understanding hybridization potential during the reintroduction of the native Endangered Species Act listed Lahontan cutthroat trout (*O. clarkii henshawi*)

L.M. Galland[12], T.L. Parchman[12], and M.M. Peacock[12]

[1]Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, NV, USA

[2]Department of Biology, University of Nevada, Reno, NV, 89557, USA

Author for correspondence:

Email: lgalland@unr.edu

Online Resource 1 Historical rainbow trout hatcheries in the Tahoe basin.

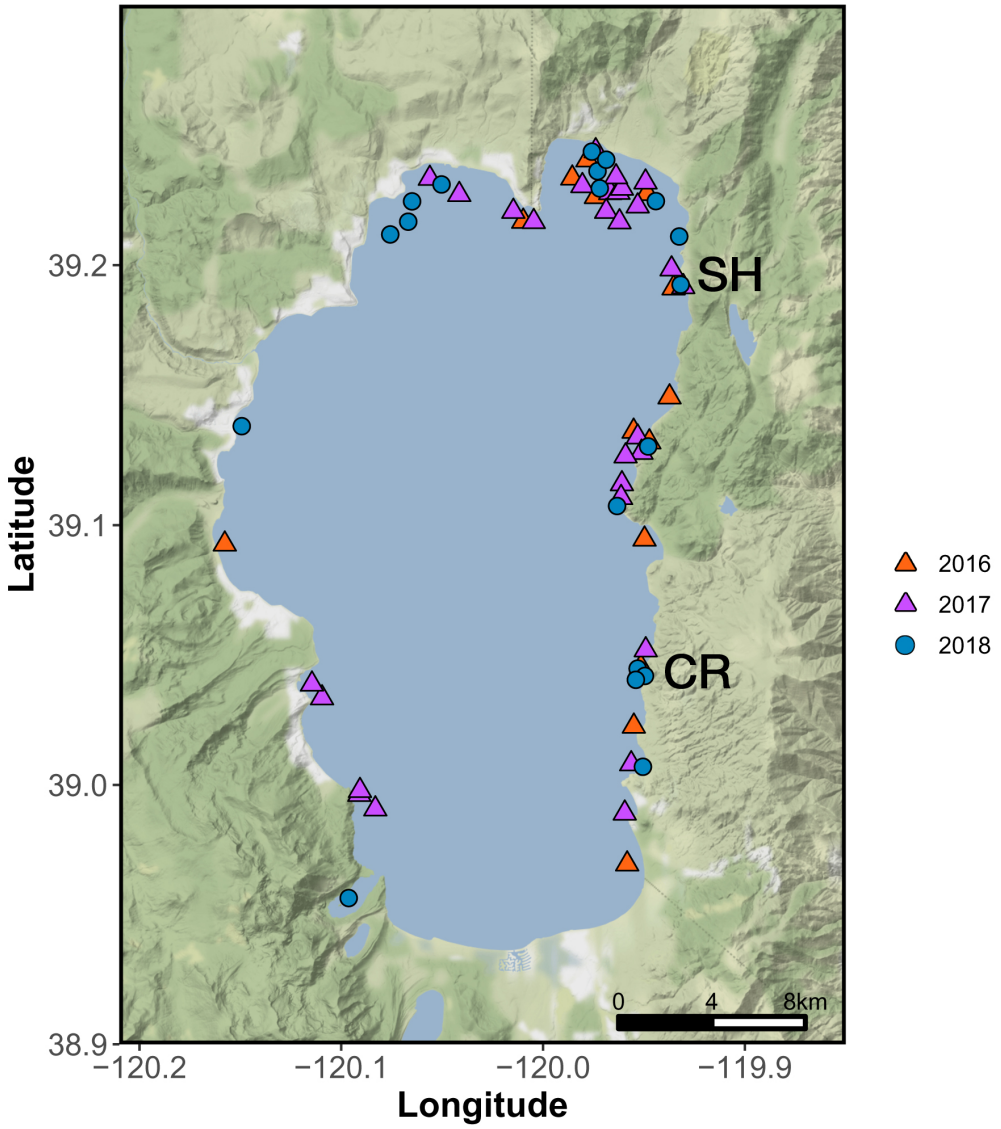
Hatchery	Location	County	State	Years
Frasier	Squaw Creek	Placer	CA	1875-1880
Hurley	Tahoe City	Placer	CA	1880-1888
Phipps	Lake Tahoe	El Dorado	CA	1884-1888
Mt. Tallac	Taylor Creek	El Dorado	CA	1895-1909
Marlette-Carson	Carson City	none[a]	NV	1916-1917
Blackwood Creek	Blackwood Creek	Placer	CA	1921-1932
Tahoe	Tahoe City	Placer	CA	1889-1956
Verdi	Verdi	Washoe	NV	1902-1905, 1909-1986

[a]Carson City is an independent city

Online Resource 2 Rainbow trout stocking records (1960-2020) from the California Department of Fish and Wildlife (CDFW) and Nevada Department of Wildlife (NDOW).

STRAIN	YEAR(S)	No. stocked per year
Fall spawning strain	1960	5,941
Kamloops (British Columbia)	1962	127,000
Trophy	1975	12,261
Trophy/Shasta cross	1975	2,396
No strain ID	1975-1983	198,748
Whitney	1978-1979, 1980, 1984, 1986	33,441
Wigwam	1982-1983, 1991	242,404
Wigwam/Shasta cross	1982	2,200
Blue Mountain	1983	44,280
Nashua	1983	11,898
Junction	1984	9,895
Coleman (anadromous steelhead)	1985	13,125
Kamloops/Junction cross	1986	15,200
Sand Creek	1983-1984, 1987-1988, 1991	66,200
Shasta (mixed ancestry including Eagle Lake)	**1977-1988**	**180,791**
Eagle Lake	**1971, 1980, 1984-2010**	**422,172**
Hot Creek (Eagle Lake origin)	**1980-1985, 1997**	**69,437**
Wytheville	1988, 1990	27,803
Marlette Lake Tahoe Basin[a]	1988	19,600
Marlette Lake Tahoe Basin[b]	1990	13,790
Erwin (Wytheville origin)	2000	13,941
Kamloops	2006-2008	37,065
Tasmanian (anadromous steelhead, Sonoma Creek)	1986, 1995-1999, 2004, 2008-2010	124,130
Jumper	2012	8,126
No strain ID	2011, 2014	13,576
Tahoe Basin[c] (Incline and Third Creeks)	**1991-2010, 2015-2018, 2020**	**691,688**
Kamloops (triploid)	2007, 2015	10,812
Trout Lodge (triploid)	2018	13,602
Triploid RBT	2011-2017, 2019, 2020	177,663
TOTAL		2,609,185

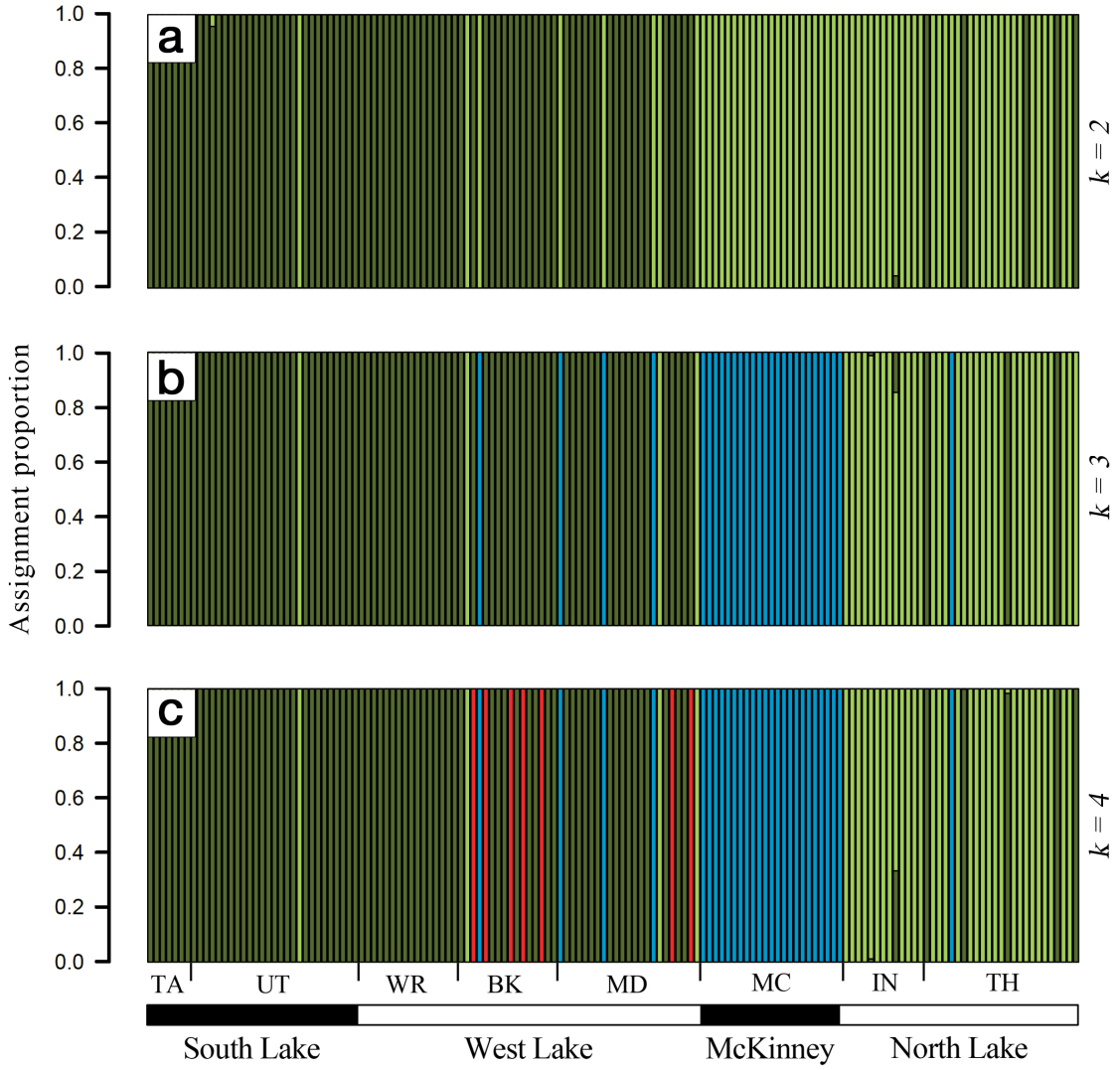
[a] naturalized RBT from Lake Tahoe raised in Marlette Lake, Lake Tahoe basin (fingerlings)
[b] naturalized RBT from Lake Tahoe raised in Marlette Lake, Lake Tahoe basin (adults)
[c] naturalized RBT from Incline and Third creeks, which drain into North Lake Tahoe, raised in NDOW Mason Valley hatchery



Online Resource 3 Individuals stocked at Cave Rock (CR, circles) and Sand Harbor (SH, triangles) across three years were recovered and reported to the Nevada Department of Wildlife. Results indicate mixing of stocked individuals across all regions of the lake.

Online Resource 4 Deviance information criteria for each of 5 replicate `entropy` runs, including mean and standard deviation, where a lower score indicates better model fit.

k	**1**	**2**	**3**	**4**	**5**	**mean**	**SD**
2	6,849,821	7,056,445	6,695,732	7,112,544	7,058,046	6,954,518	176,076
3	12,076,597	10,519,135	9,712,600	9,747,084	9,930,560	10,397,195	993,016
4	11,888,946	11,154,106	13,788,505	11,506,529	10,873,331	11,842,283	1,152,730
5	16,200,905	10,521,209	11,029,413	15,003,581	14,091,069	13,369,236	2,489,804
6	15,672,915	15,905,383	17,747,334	20,482,260	16,204,417	17,202,462	2,004,494
7	15,336,136	51,082,844	15,853,549	75,907,098	26,233,082	36,882,542	26,193,710
8	19,268,427	82,991,043	17,512,223	16,024,636	129,551,613	53,069,588	51,293,193



Online Resource 5 Discriminant analyses of principal components without *a priori* clustering based on the `find.clusters` function where vertical bars represent individuals, and colors represent proportion of ancestry from each ancestral group.

Chapter 3: History and environment shape spatial genetic variation and predict climate maladaptation in a narrowly distributed serotinous pine, *Pinus muricata*.

Lanie M. Galland[12], Trevor M. Faske[12], Carolina Osuna-Mascaró[2], Thomas E. Dilts[13], Sarah Bisbing[13], & Thomas L. Parchman[12]

[1]Graduate Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, NV, USA

[2]Department of Biology, University of Nevada, Reno, NV, 89557, USA

[3]Natural Resources and Environmental Science, University of Nevada, Reno, NV, 89557, USA

Author for correspondence:

Email: lgalland@unr.edu

Abstract

Understanding the distribution of genetic diversity and differentiation in species with disjunct populations is critical for assessing local adaptive variation, as well as evolutionary potential, especially in the context of climate change. In contrast to the large distributions and population sizes of most pine species, *Pinus muricata* (Bishop pine) occurs in a small number of isolated populations occupying a narrow band of environmental conditions along the coast of western North America. To describe the spatial arrangement of genetic differentiation and diversity in this species, we used genotyping by sequencing to generate population genomic data for trees sampled from nearly all existing populations of *P. muricata* (12 populations, 213 individuals, 7,828 loci). We used genetic-environment association (GEA) analyses to quantify the contribution of environmental variables to local adaptation and spatial genetic structure. Based on these results, we quantified genomic offset as a relative estimate of potential maladaptation given future climate projections at 2041–2060 and 2081–2100. Our analyses reveal pronounced spatial genetic structure across the distribution, with most populations forming genetically identifiable groups across a latitudinal gradient, and differentiation at a remarkably fine scale among three stands on Santa Cruz Island. Despite occurring in small, isolated stands, *P. muricata* populations do not exhibit reduced diversity relative to many other more widespread pines. GEA analyses suggested that specific soil and climate variables have contributed to local patterns of genetic differentiation. Genomic offset analyses suggest geographic variation in potential maladaptation, with northern populations generally experiencing higher levels of offset under projected climate change models. Overall, our results suggest that isolation and

local adaptation have shaped current patterns of genetic variation among disjunct populations and illustrate the evolutionary consequences of this variation for *P. muricata*.

Key words: climate maladaptation, genetic diversity, genetic offset, genetic-environment association (GEA), GEA offset, *Pinus muricata* (Bishop pine), RADseq

Introduction

Variation in gene flow, drift, and local adaptation are all expected to shape spatial genetic structure across species with fragmented distributions. Fragmented populations of rare plant species often suffer decreased levels of genetic diversity, whether a cause or consequence of rarity (Gitzendanner & Soltis, 2000; Nybom, 2004), that can decrease evolutionary potential and increase extinction risk at geographic, temporal, and climatic scales. Disjunct distributions, coupled with processes that reduce diversity (e.g., genetic drift), often result in pronounced levels of spatial genetic structure (Young *et al.*, 1996). While such patterns are common across plant systems, numerous studies have identified inconsistencies, where broad distributions and large population sizes do not always protect a species from genetic erosion, and where genetic variation of rare species is not necessarily reduced but is instead overgeneralized as low (Gitzendanner & Soltis, 2000; Ellis *et al.*, 2006; Honnay & Jacquemyn, 2007; Kramer *et al.*, 2008; Ægisdóttir *et al.*, 2009). Rather, populations with high genetic diversity despite fragmentation may be resistant to genetic erosion due to high levels of outcrossing, long lifespans (i.e., older individuals with increased diversity from a previously expanded range), and overlapping generations (Nybom, 2004; Ægisdóttir *et al.*, 2009; Escaravage *et al.*, 2011; Walisch *et al.*, 2015), all of which are characteristic of several conifer species.

Pines are among the most ecologically and economically important plants on Earth, occurring across diverse landscapes and playing an essential role in structuring forest ecosystems (Farjon, 2008). Pines typically occur in large, contiguous populations that exhibit little genetic structure (González-Martínez *et al.*, 2006; Petit & Hampe, 2006) due to high levels of gene flow from wind dispersal (Slavov & Zhelev, 2004; González-

Martínez *et al.*, 2006). Despite the prevalence of gene flow, a long history of common garden studies has demonstrated that pine populations are locally adapted to diverse environmental conditions (Langlet, 1971; Matyas, 1996; Morgenstern, 1996; Sork *et al.*, 2013; Lind *et al.*, 2018), and more recent studies have quantified adaptation across the landscape with both phenotypic and genomic approaches (Eckert *et al.*, 2015; Yeaman *et al.*, 2016; De La Torre *et al.*, 2019; Mahony *et al.*, 2020; Hall *et al.*, 2021). Pines generally have exceptionally long lifespans, but the seedling phase is typically when selection acts strongest (Savolainen *et al.*, 2007; Alberto *et al.*, 2013; Lind *et al.*, 2018). As a result, adults can occur in environments that differ from what they experienced as seedlings, and climate change is creating an increasing mismatch between the genetics of pine populations and their environments (Aitken *et al.*, 2008; O'Connor *et al.*, 2014; Tíscar *et al.*, 2018; De La Torre *et al.*, 2019). Moreover, recent models predict dramatically high mortality rates as a result of warming temperatures and associated physiological stresses for pines distributed in western North America (e.g., Alberto *et al.*, 2013; Mcdowell *et al.*, 2016). Understanding the distribution of phenotypic and genetic variation across populations of pines will be critical for understanding environmental variation driving local adaptation and the potential response of forests to environmental change (Savolainen *et al.*, 2013; Ellegren, 2014; Prunier *et al.*, 2016).

Pinus muricata (Bishop pine), *P. radiata* (Monterey pine), and *P. attenuata* (knobcone pine) are emblematic California closed-cone pines, and form a monophyletic clade (Attenuatae, *sensu*; Gernandt *et al.*, 2018) within the Australes subsection of the North American hard pines (Gernandt *et al.*, 2005, 2018). Phylogenetic analyses based on a variety of data indicate that the Attenuatae clade originated in the late Miocene between

5-10 mya (Eckert & Hall, 2006; Hernandez-Leon *et al.*, 2013; Saladín *et al.*, 2017). In contrast to the wide distributions and large population sizes of many pine species, each of these species is distributed in restricted ranges and smaller populations (Little, 1975). *P. muricata* occurs in few disjunct coastal populations in California, USA, and Baja California, Mexico (Fig. 1a,b). These populations are narrowly distributed within a couple kilometers of the coastline where they occupy slopes and terraces and experience a Mediterranean climate with moisture coming from fall and winter precipitation and spring and summer coastal fog (Millar, 1986, 1988). While populations in the north occupy a more continuous stretch of coastline, others are highly isolated, including stands on two of the Channel Islands (Critchfield & Little, 1966; Little, 1975). As with the other Attenuatae, *P. muricata* populations exhibit varying degrees of serotiny (the retention of seeds in closed cones), a fire adaptation that has evolved repeatedly in *Pinus* (Lamont & Enright, 2000; He *et al.*, 2012).

While early taxonomic treatments of the California closed-cone pines were varied, subsequent phylogenetic and population genetic analyses illustrated *P. muricata* as a monophyletic lineage characterized by some level of geographic structure (Millar, 1988; Strauss *et al.*, 1993; Gernandt *et al.*, 2018). *P. muricata* exhibit considerable phenotypic variability, including variation in stomatal form (Millar, 1983), stem form (Duffield, 1951), cone morphology (Duffield, 1951; Linhart *et al.*, 1965), and xylem monoterpene characteristics (Mirov *et al.*, 1966) that vary among geographic regions. Based on early analysis of phenotype, Duffield (1951) recognized distinct northern, central, and southern races, and crossing trials by Millar and Critchfield (1988) indicated potential reproductive barriers among these three geographic groups of populations.

Analyses of genetic variation across these populations have been limited to allozymes and small segments of chloroplast and mitochondrial DNA, but indicated significant differentiation among the different regions of the distribution (Millar, 1988; Hong *et al.*, 1993; Strauss *et al.*, 1993). These analyses also suggested that despite current small population sizes, *P. muricata* populations do not lack genetic diversity. Fossil cones dating to 26 kya or older from many currently unoccupied regions in coastal and central California illustrate that *P. muricata* had much more continuous distributions at times in the past (Axelrod, 1980, 1981). Cyclical climate fluctuations during the Pleistocene (Heusser *et al.*, 1985; Mann & Hamilton, 1995; Hewitt, 2004) as well as tectonic events reshaping the California coast (i.e., the uplifting of the Cascade Range and its division moving south from Mt. Shasta into the Sierra Nevada and Coastal Ranges; Hewitt, 2004) could have thus driven repeated cycles of contraction and expansion in *P. muricata*.

Models based on climate change projections predict extraordinarily high mortality for conifers in western North America (Lenihan *et al.*, 2008; Davis *et al.*, 2019; Halofsky *et al.*, 2020), even for broadly distributed pines (McDowell *et al.*, 2013; Mcdowell *et al.*, 2016). Bishop pine represents one of the most range-restricted pines on Earth, and is listed as threatened by the IUCN (Farjon, 2013). The severe degree of isolation could make *P. muricata* particularly susceptible to the effects of warming temperatures as its disjunct, restricted populations could further constrict with insufficient northward range expansion in response to environmental change, making it a potential future candidate for assisted migration. As past analyses of genetic variation in *P. muricata* were based on small numbers of traditional molecular markers, current high throughput sequencing data stand to improve our understanding of genetic variation in *P. muricata* along several

fronts. Genome-wide data should generate a more comprehensive understanding of patterns and levels of genetic differentiation among populations, as well as a more thorough understanding of variation in genetic diversity. Importantly, recent methods for genetic environment association (GEA) analyses applied to genome-wide data can facilitate an understanding of how environmental variation contributes to the spatial genetic structure of populations and local adaptation (Forester *et al.*, 2018; Capblancq *et al.*, 2020a). An understanding of associations among genetic and environmental variation can then predict the relative extent of maladaptation populations might experience under projected future climate variation (Fitzpatrick *et al.*, 2018; Capblancq *et al.*, 2020a). Such analyses may be especially relevant given that *P. muricata* occurs in small and disjunct populations, as well as the outsized importance of conifer species that are often foundational, dominant plant species with extended ecological significance.

Here we used high-throughput sequencing of reduced representation libraries (ddRADseq) to quantify spatial variation in genetic differentiation and diversity to further understand the geographic and historical context of divergence across the range of *P. muricata*. Next, we used two GEA approaches to evaluate the extent to which climate and soil variation have shaped spatial genetic structure and potentially contributed to local adaptation. Finally, given the occurrence of *P. muricata* across few small and disjunct populations and its status as threatened under the IUCN (Farjon, 2013), we estimated the extent of maladaptation under multiple future climate projections using genomic offset techniques (designated as "GEA offset," detailed below). Our results provide a perspective on the association of spatial and environmental variation with population

genetic variation in a rare, range-restricted conifer, and provide an applied perspective on the potential of these populations to respond to environmental change.

Materials and Methods

Sample collection, sequencing, alignment, and variant calling. Needle samples were collected from up to 30 trees per population as DNA sources from mature trees at 12 site across the range of *P. muricata*, including from three stands of Santa Cruz Island (Table 1, Fig. 1a,b). Genomic DNA was isolated at Ag-Biotech (Monterey, CA) and subsequently used in a reduced-representation workflow to create DNA sequencing libraries following the protocol described in Parchman *et al*. (2012). For details on DNA sequencing library preparation, see the Supporting Information and the protocol at the DRYAD repository (DOI: 10.5061/dryad.sqv9s4n61). Fragments 350-450 bp in length were size-selected using the PippinPrep quantitative gel electrophoresis unit (Sage Science, Beverley, MA) at the University of Texas Genome and Sequencing Analysis Center (Austin, TX) and subsequently sequenced with S2 chemistry on one lane of the Illumina NovaSeq platform.

We filtered sequencing data for contaminants (e.g., *PhiX*, *E. coli*) and for Illumina sequencing oligos using `bowtie_db2` (Langmead & Salzberg, 2012) and Perl and bash scripts. We then used a custom Perl script to demultiplex reads by individual by first correcting 1 – 2 bp errors in barcode sequences and removing restriction site-associated bases, and subsequently to match each individual to its unique DNA barcode sequence, splitting sequencing data into individual fastq files. Next, we used the `CD-HIT` clustering algorithm (Fu *et al.*, 2012) to align contig consensus sequences using a minimum match

percentage of 90%, thereby generating a *de novo* assembly for mapping reads. Reads for each individual were mapped to the partial reference using `bwa v0.7.5a` (Li & Durbin, 2009) with a minimum edit distance of 3. We called single nucleotide variants using `samtools v1.3` (Li *et al.*, 2009) and `bcftools v1.3` (Li *et al.*, 2009) across the alignments. Here, we generated genotype likelihoods from `samtools v1.3` (Li *et al.*, 2009) using `--min-MQ 20`, `--min-BQ 20`, `--max-depth 100`, and minimum genotype quality (`GQ`) 10. We then used `vcftools v0.1.14` (Danecek *et al.*, 2011) to ensure we only retained biallelic variants that were found in a minimum of 70% of individuals with minor allele frequencies greater than 0.05, and removed individuals that were missing data at >20% of loci. Finally, to avoid genotyping error due to paralogous regions, we used two methods. First, we filtered over-assembled loci with coverage depth >40x. We then removed potentially duplicate or diverged duplicate loci using the `HDplot` approach to ensure we retained only singletons (McKinney *et al.*, 2017). Here, we retained loci with heterozygosity between 0 – 0.5 and read ratio deviance from -20 – 30. For phylogenomic analyses, we performed variant calling with the same parameters but included *P. radiata* individuals (see below).

Spatial genetic structure. To assess spatial genetic structure, we used both model-based (`entropy`; Gompert *et al.*, 2014) and model-free (principal components analysis, PCA) approaches. We began by using `entropy` (Gompert *et al.*, 2014), a hierarchical Bayesian model, to estimate genotype probabilities and admixture proportions *q* from *k* ancestral demes without the use of *a priori* sample origin information. `entropy` employs a model similar to that of `structure` (Pritchard *et al.*, 2000), but accounts for genotype

uncertainty stemming from variation in sequencing coverage depth across individuals and loci while estimating genotype and ancestry coefficient posterior probabilities for each individual at each locus. First, we ran PCA on genotype likelihoods using the `prcomp` function in `R version 3.4` (R Core Team, 2017). We then used the `lda` function from the `MASS` package (Venables & Ripley, 2002) in `R version 3.4` (R Core Team, 2017) to run linear discriminant analysis. We ran `kmeans` clustering for $k = 2$ through $k = 12$ demes to find initial starting values for the model. The model was run for 100,000 MCMC iterations (5 chains per k) after an initial burn-in of 30,000, saving every 10^{th} iteration. We used deviance information criterion (DIC) to assess model fit, where lower values represent better model fit.

Using genotype probabilities generated with `entropy`, we examined genetic variation among individuals and populations using the `prcomp` function in `R`, maximizing total variance between individuals. As pairwise metrics of genetic differentiation among populations, we used allele frequencies to estimate Hudson's F_{ST} (Hudson *et al.*, 1992) and Nei's D (Nei, 1972). We also calculated expected population-level heterozygosity (H_E) and individual inbreeding coefficients (F), which were averaged across individuals within each population. Individual inbreeding coefficients (F) were calculated while incorporating genotype uncertainty using `ngsF` (Vieira *et al.*, 2013), implemented through `angsd-wrapper` (Durvasula *et al.*, 2016). We then calculated population levels of genetic diversity while incorporating genotype uncertainty using the `doSaf 1`, `REALSFS`, and `thetastat` calls in `ANGSD` (Korneliussen *et al.*, 2013, 2014) We retained estimates for both Θ_π (the average number of pairwise differences between sequences; Tajima, 1983) and Θ_W (the number of segregating sites; Watterson, 1975), and calculated the scaled

difference between the two, or Tajima's *D* (Tajima, 1989) for each locus. To assess how geography and environment influence genetic variation, we used Mantel tests (N_{perm} = 999; Mantel, 1967) to calculate isolation-by-distance (IBD; Wright, 1943) and isolation-by-environment (IBE; Wang & Bradburd, 2014), comparing geographic (Haversine) and environmental (Euclidean, described in the next section) distance to genetic distance (Nei's *D*).

Phylogenetic analyses. As an alternative assessment of genetic differentiation across the range of *P. muricata*, we conducted Maximum Likelihood based phylogenetic inference. We converted the SNP matrix generated above into PHYLIP format using `vcf2phylip v 2.0` (Ortiz, 2019) and inferred a phylogeny with `IQ-TREE` (Nguyen *et al.*, 2015) using *P. radiata* as an outgroup. We ran `IQ-TREE` using ModelFinder (-m MFP; Kalyaanamoorthy *et al.*, 2017) to infer the best substitution model that minimized the BIC (Bayesian Information Criterion) score for building the Maximum Likelihood phylogenetic tree. A total of 50 models were tested and evaluated with 1,000 ultrafast bootstrap (UFBoot) replicates to assess statistical support (-bb 1000; hoang18). We also used the ASC option to correct for ascertainment bias (Lewis, 2001). We trimmed the final tree for easy visualization using the `drop.tip` function from the `R` package `phytools v 0.7-80` (Revell, 2012; Revell & Revell, 2014), keeping a maximum of four individuals per *P. muricata* population and one individual of *P. radiata*.

We used the `drop.tip` function to simplify the tree to retain one sample tip per population. Then, we used the `phylo.to.map` function in the `R` package `phytools v`

`0.7-80` (Revell & Revell, 2014) to plot a map of California and link tree tips with each population's geographic coordinates.

Genetic-environment association analyses. Detecting correlated shifts in allele frequencies and environmental variation facilitates an understanding of the genomic signatures of local adaptation to specific environmental variables (Savolainen *et al.*, 2013; Rellstab *et al.*, 2015; Forester *et al.*, 2018) and can be used to anticipate species response to changing environmental conditions (Fitzpatrick & Keller, 2015; Brauer *et al.*, 2016; Bay *et al.*, 2018). Multiple model-based GEA methods that account for population genetic structure have been utilized, including latent factor mixed models (Frichot *et al.*, 2013; Frichot & François, 2015; Caye *et al.*, 2019), `BAYENV` (Günther & Coop, 2013), `BAYESCENV` (De Villemereuil & Gaggiotti, 2015). Here, we used two methods to identify environmental drivers of local adaptation: (1) redundancy analysis (RDA; Legendre & Legendre, 2012), an ordination approach valuable for its low false-positive and high true-positive rates while accounting for complex demographic histories (Forester *et al.*, 2018), and (2) Gradient Forests (GF; Ellis *et al.*, 2012; Fitzpatrick & Keller, 2015), a machine-learning algorithm that fits an ensemble of decision trees using Random Forests and constructs cumulative importance turnover functions from these models, associating sets of loci with a multivariate assortment of predictors.

We pulled environmental data for each sampling location, incorporating elevation, 19 bioclimatic variables from the WorldClim v2.1 database at 30-arcsecond resolution (~1 x 1 km; Fick & Hijmans, 2017), and 14 soil variables from SoilGrids 2.0 (see Supporting Information for details on soil data generation; Poggio *et al.*, 2021). We

also considered fire as a potential variable, but there have been only two fires across all sampling sites since fire history recording began, so we did not include this factor. To avoid multicollinearity, soil variables were collapsed into the first five PC axes (cuml. *PVE* = 93.89%; Supporting Information Table S3), and we removed all bioclim variables with $|r| > 0.75$, retaining elevation, mean annual temperature (bio1), isothermality (bio3), temperature seasonality (bio4), and temperature annual range (bio7). Soil PC2 was removed due to strong correlation with bio1 ($|r| = 0.85$). The environmental (Euclidean) distance matrix was used to calculate IBE as described in the previous section. As we were not concerned with identifying specific outlier loci, population structure was not partialled out before conducting our GEA analyses, as we aimed to assess only the magnitude and directionality of specific environmental variables' influence on spatial genetic variation. We did, however, include latitude as a covariate in our GEA analyses as it is strongly associated with PC1 (or population structure). The `rda` function in the `vegan` package (Oksanen *et al.*, 2013, 2019) was used for RDA and the `gradientforest` package was used for GF in `R`. Both analyses used all genomic loci (i.e., genotype probabilities) as the response variables and the center and standardized environmental variables as predictors. Total importance for each environmental variable was calculated as the total weighted loadings (sum of *eigenvalue*loading* for each axis) in the RDA and the weighted importance (R^2) in GF.

GEA offset analyses. Predicting changes to the spatial distribution of local adaptation will be essential for both anticipating species response and effectively conserving and managing populations as environmental conditions rapidly change. Using future climate

models in combination with results from GEA, researchers have been able to identify the rate of allele frequency change across an environmental gradient at a single locus and have coined this "genetic offset" (Fitzpatrick & Keller, 2015) or "genomic vulnerability" (Bay *et al.*, 2018; Ruegg *et al.*, 2018). We use caution with these terms, as suggested by Láruson et al (2021), and refer to the predicted mismatch of genetic-environment associations across the landscape under different climate scenarios as "GEA offset." These changes to the strength and direction of association between present and future may be especially threatening to longer-lived, sessile species that will experience environmental change over the course of individual lifespans (e.g., coniferous trees). Here, we quantified GEA offset using the set of climatic variables from the previous section to compare with future climate change scenarios using both GF and RDA (Referred to as GF offset and RDA offset, respectively).

Future climate data was extracted from WorldClim v2.1 CMIP6 (Eyring *et al.*, 2016) at 30-arcsecond resolution (~1 x 1 km) for time intervals 2041–2060 and 2081–2100 (approximately one and two generations from present, respectively) under two climate change scenarios: Shared Socioeconomic Pathway (SSP) 1-2.6 (mild change) and SSP5-8.5 (severe change). These models utilize adjusted greenhouse gas scenarios and are widely accepted as the most likely climatic shifts (see Van Vuuren *et al.*, 2011 for details). All available global climate models (GCMs) for each time interval/SSP combination at the sampling locations were extracted, resulting in the mean estimate for each bioclimatic variable across eight GCMs that were subsequently used as projected climate data (see Supporting Information Materials and Methods for full details on data extraction and summary statistics). The same variables used in the GEA analyses were

used for offset analyses, only allowing the bioclimatic variables to change between current and future climate scenarios. We, again, avoided basing inference on any specific outlier loci and instead used all available genomic loci, as suggested by Láruson et al. 2021. We calculated the offset between current and future climate scenarios using the Euclidean distance of predicted environmental importance for each individual with both RDA and GF, following Capblancq & Forester (2021) and Fitzpatrick & Keller (2015), respectively. For the RDA offset, we chose to use all RDA axes to calculate the total weighted loading of each variable as the associative measure. We additionally calculated the GEA offset between the different generational time intervals (2041–2060 versus 2081–2100) and climate change scenarios (SSP1-2.6 versus SSP5-8.5).

Results

Sample collection, sequencing, alignment, and variant calling. After removing contaminant DNA and eliminating individuals with insufficient sequencing data, we retained a mean of 3,309,158 reads per individual. Subsequent filtering steps resulted in retention of 9,260 loci in 213 individuals. Finally, we removed 1,432 loci representing potentially paralogous regions, resulting in retention of 7,828 SNPs with mean coverage of 10.5X per locus per individual.

Spatial genetic structure. We found pronounced genetic differentiation across the landscape (Figs 1c, S1) across a hierarchy of scales. DIC values from `entropy` indicated the $k = 3$ model as best fit number of demes (Supporting Information Table S1), corresponding to individuals from populations from the north coast (*N*pop = 6, north of

38°N; PP, NR, PA, SP, FR, and PR populations), the central coast (*N*pop = 3, from 34.5°N to 38°N; DM, DC, and LO populations), and SCI (*N*pop = 3, south of 34.5°N, CP, CH, and PB populations) (Supporting Information Fig. S1). Demes showed mixed but consistent ancestry within populations, indicating marked differentiation following a latitudinal gradient (Figs 1c,d, S1). PCA based on genotype probabilities from `entropy` provided additional support for landscape genetic differentiation, with PC axes 1 and 2 explaining 14.4% and 3.1% of the variation in the data, respectively. All populations from the north coast grouped together within a tight cluster, overlapping in PC space, while individuals from the central coast and Santa Cruz Island (SCI) populations grouped tightly in non-overlapping clusters (with the exception of overlap of the Christy Pine and Pelican Bay populations from SCI; Fig. 1c). Patterns of differentiation along the first principal component axis were almost perfectly concordant with the latitudinal gradient (latitude-PC1: $r = 0.93$, $p < 0.001$) and greater population structure was highlighted in the south than in the north.

To identify additional population structure within the regional demes, we separately analyzed 89 individuals from six sampling sites along the north coast, and then 74 individuals from three sampling sites on SCI. The north coast populations showed some genetic differentiation, with differentiated populations following a latitudinal gradient, although levels of differentiation among sampling sites were subtle (overall mean $F_{ST} = 0.067$, overall F_{ST} range = 0.011 – 0.117; Fig. 2a,b). SCI populations, on the other hand, showed marked fine scale population differentiation despite highly geographically proximate stands (<10 km), grouping in nonoverlapping PC space (Fig. 2c).

Population differentiation increased in magnitude as latitude decreased, with northern populations yielding lower values of pairwise F_{ST} and Nei's D than in southern populations (Supporting Information Fig. S2). Individual inbreeding estimates (F) varied across all sampling sites (Table 1, Supporting Information Fig. S3c), but were minimal overall (F mean = 0.0499, F range = 0 – 0.3135), showing no observable latitudinal pattern. Mean nucleotide diversity varied across the landscape but was relatively high compared to similar taxa, with mean Θ_{π} = 0.00695 (range = 0.00616 – 0.00831) and mean Θ_{W} = 0.00659 (range = 0.00539 – 0.00838) (Supporting Information Table S2, Supporting Information Fig. S3a,b). All measures of Tajima's D were negative (Supporting Information Fig. S3d), consistent with population expansion. (See Table 1 and Supporting Information Table S2 for additional diversity metrics, including 95% confidence intervals in the latter.) Individually, geographic distance and environmental distance were each strong predictors of genetic distance, indicating both IBD (Mantel r = 0.8081, p = 0.001) and IBE (Mantel r = 0.4177, p = 0.008). Further, the association between geographic and environmental distance was strong (Mantel r = 0.5608, p = 0.001).

Phylogenetic analyses. The phylogenetic results based on the same filtering parameters (resulting in 5,427 SNPs from 12 *P. muricata* populations and the single *P. radiata* population) were congruent with a latitudinal gradient of differentiation, where population differentiation was more remarkable as latitude decreased. The Maximum Likelihood phylogeny based on the optimal substitution model (PMB+F+ASC+R6) showed three well-supported clades (Fig. 3). The first clade (UFBoot = 97) grouped the

northern populations (PR, FR, SP, PA, NR, and PP; north of 38°N) and one central coast population (DM; 36.59°N). The second clade (UFBoot = 100) grouped the southern mainland population (LO; 34.73°N) and SCI populations (CH, PB, CP, south of 34.5°N). Finally, the DC population on the central coast (35.24°N) appeared in a single clade (UFBoot = 75). Moreover, individuals from southern and central coast populations were grouped by population, whereas individuals from northern populations were not.

Genetic-environment association analyses. Both of the GEA analyses provided evidence of specific environmental variables influencing spatial genetic variation and potentially local adaptation. The most influential variables in the RDA and GF showed some degree of concordance ($r = 0.5767$, $p = 0.0801$), where the top variables were latitude, elevation, and mean annual temperature (bio1) for RDA, and mean annual temperature (bio1), latitude, and soilPC3 for GF (Fig. 4; soil PC loadings in Supporting Information Table S3). Within the RDA, the environmental variables explained a large portion of the spatial genetic structuring with an adjusted $r^2 = 0.2083$, where the first axis alone explained the majority of that variance (*PVE* = 57.96%). When assessing the total weighted loadings of each variable, latitude contributed 3.54 and 5.43 times more than the next two variables (elevation and bio1, respectively; Fig. 4b). Alternatively, when looking only at the RDA1 loadings, latitude and bio1 were roughly equivalent and had much higher loadings than the remaining variables, similar to the results of GF. The cumulative weighted R^2 of the GF model was 0.2160, where bio1 and latitude were roughly equivalent and yielded the greatest importance (Fig. 4b). Soil PCs ranked 3rd - 6th, and bio3, elevation, bio7, and bio4 followed in rank in order of importance. The top ranked variable (mean annual

temperature, or bio1) had 4.58 times greater R^2 than the lowest ranked variable (temperature seasonality, or bio4).

GEA offset analyses. The GEA offset analyses indicated a large amount of variation in offset among populations under future climate change scenarios. Differences based on the severity and time interval of climate projections were variable. There was a discernible difference in population-specific results from RDA and GF offset across all four of the possible time interval/SSP combinations (Fig. 5, Supplemental Information Fig. S4), likely due to the variation in association of environmental variables found in the GEA analyses above (Fig. 4). Still, GEA offset across all climate scenario combinations was associated with changes in mean annual temperature (bio1), which showed the strongest association with sampled loci (Fig. 4b, Supplemental Information Fig. S5). Despite differences arising from variation in environmental variables, results of each method were consistent when assessing relative offset (Fig. 5). For both GEA offset methods, both SSP1-2.6 and SSP5-8.5 scenarios projected to 2041–2060 (roughly one generation from present) indicated minimal offset. Projected to 2081–2100 (roughly two generations from present), however, measures of offset under the SSP5-8.5 scenario were drastically increased, with the SSP5-8.5 scenario yielding offset 2.35 and 3.41 times more severe than the offset under the SSP1-2.6 scenario for GF and RDA, respectively. While the offset between 2041–2060 and 2081–2100 for SSP1-2.6 was only a 1.04- and 1.07-fold increase, the same comparison yielded a 1.86- and 2.66-fold increase for SSP5-8.5.

Discussion

Spatial variation in differentiation and diversity. Our results illustrate genetic differentiation at a hierarchy of scales across the range of *P. muricata*. At the broadest spatial scale, we detected clear evidence for differentiation among groups of populations in northern, central, and southern California (Figs 1c,d, S1). Genetic differentiation was moderate among these groups (mean F_{ST} = 0.067, F_{ST} range = 0.011 – 0.117; Supporting Information Fig. S2), but trees from each group formed strongly identifiable units in ordination, ancestry, and phylogenetic analyses. These results are not surprising in light of past studies illustrating differentiation within and among some of these groups based on allozymes (Millar, 1983, 1988), morphology (Duffield, 1951), turpentine composition (Forde & Blight, 1964; Mirov *et al.*, 1966), phenology (Millar, 1983), and tests of reproductive compatibility (Critchfield, 1967; Millar & Critchfield, 1988). Cone serotiny levels, which are likely shaped by variation in fire, are also higher in the south than in the north (Duffield, 1951). Clear patterns of genetic and phenotypic differentiation in this and past studies are consistent with disjunct *P. muricata* populations evolving independently in different geographic regions of western North America.

Past phylogenetic analyses of *Pinus* subsection Australes based on a variety of data and methods inferred crown ages ranging from 6 – 43 mya (Eckert & Hall, 2006; Saladín *et al.*, 2017; Gernandt *et al.*, 2018). While subsection Australes has a present day distribution in southeastern North America, Mexico, and the coastal and Pacific slope regions of western North America (Gernandt *et al.*, 2018), fossil data suggests the group was previously more widespread in North America (Axelrod, 1986). The monophyletic group of *P. muricata*, *P. radiata*, and *P. attenuata* (Attenuatae clade) has been previously

suggested, based on paleontological data, to have originated in Mexico from an oocarpae-like ancestor (Axelrod, 1980). Our phylogenetic analyses suggest southern *P. muricata* populations are older than those in the north, consistent with a southern origin followed by a northward expansion (Fig. 3). Also consistent with such a history, genetic differentiation is stronger among southern populations than among those in the northern part of the distribution. More recent shifts of *Pinus* along coastal California have been illustrated by pollen records from the Santa Barbara basin, where pines were replaced by woodland and chaparral vegetation as wetter, cooler climates were replaced by warmer, drier climates ~5.7 kya (Heusser, 1978; Axelrod, 1981). Further, cone fossils likely representing *P. muricata* and sister species *P. radiata* have been recovered throughout coastal regions of California (Axelrod, 1981) illustrating more continuous distribution during the cooler, wetter Late Wisconsin climate before drought and heat during the Xerotherm (~4 – 8 kya) led to range fragmentation (Axelrod, 1967a,b, 1980). Indeed, fossil pollen illustrates repeated cycles of *Pinus* expansion and contraction coinciding with climate fluctuations over the last 160k years (Heusser & Sirocko, 1997; Millar, 1999) consistent with the hypothesis (Millar, 1999) that these populations may have persisted as a dynamic metapopulation.

While the broad patterns of differentiation above were expected based on past work, our analyses revealed genetic differentiation among sampling sites within these groups at much finer geographic scales than anticipated. Most individuals were fully identifiable to sampling site based on genotypic data (Figs 1c, 2a,c), and we observed a distinct latitudinal gradient of population differentiation (Figs 1c inset, 3b). Individuals from populations in southern and central regions formed largely non overlapping and

clearly differentiated clusters in PC space, while those from northern regions showed differentiation only at much finer scales (Figs 1c, 2).

Previous work on continuously distributed *P. muricata* along the northern coast near Sea Ranch between Point Arena (PA) and Salt Point (SP) sampling sites (Fig. 1a) described a notable break in phenotype and phenology. A steep clinal transition occurs across this region as illustrated by differences in monoterpene composition, leaf color, and stomatal form (Duffield, 1951; Millar, 1983). Alpha-pinene composes the majority of monoterpene fraction in xylem resin north of the transition zone, while delta-3-carene prevails to the south (Mirov *et al.*, 1966), and populations north of the transition have blue needles while green needles predominate south of the transition (Duffield, 1951) due to the waxiness of stomatal chambers (Millar, 1983). Further, Millar (1983) found significant differences in allozyme GOT-1 between trees north and south of the transition zone. Importantly, however, no differentiation was evident along the north coast for the other 19 allozymes queried by Millar (1983), nor was it evident in other studies. Rather, the phenotypic differentiation among *P. muricata* north and south of Sea Ranch appears to persist despite a lack of stark differences in genetic differentiation among the southern and northern groups of the north coast.

While we did not find genome-wide structure mirroring the distinct phenotypic break within the north coast region described by Millar (1983) and others, we did recover a subtle but clear pattern of differentiation among sampling sites across latitude here. This, and evidence for population differentiation across small distances (described below), was likely evident due to the much higher marker density applied here compared to past studies based on fewer molecular markers. One mechanism likely underlying this

pattern is the north to south distributional band of trees with prevailing winds blowing pollen inland to closely neighboring trees, and then into habitat unsuitable for *P. muricata*. This hypothesis was proposed by Millar (1983), who suggested that this pattern may be due to prevailing wind directions in combination with changes in flowering time and thus reproductive receptivity. When trees release pollen along the coast from April through May, winds primarily originate from the NNW, indicating dispersal that would inhibit pollen movement in opposing directions (Millar, 1983), concordant with our results. Here, we further observed differentiation among sites in this north coast region at a finer scale than was evident in past studies (Fig. 2a; Millar, 1988, 1989). This scale of spatial genetic structure in a conifer is notable, even if characterized by subtle overall levels of differentiation. While our results are supported by previous work, future study of this cline with denser individual and genomic sampling is warranted.

We also documented evidence of differentiation among three distinct *P. muricata* stands on SCI (Fig. 2c). Although overall levels of differentiation were subtle (F_{ST} mean = 0.0387), that we detected any differentiation among stands separated by 10 km or less was surprising. Most conifers exhibit little genetic structure even over great distances (González-Martínez *et al.*, 2006; Petit & Hampe, 2006), with few exhibiting population structure over small spatial scales. A number of factors may have influenced genetic differentiation on the island, but most notable are stand-replacing drought in combination with foraging and trampling by feral sheep, both of which could reduce genetic diversity and promote genetic drift. Bishop pine on SCI has an extensive history of known mortality events occurring in the late 1940s, mid 1970s, late 1980s, and mid 2010s (Fischer *et al.*, 2009; Robeson, 2015; Taylor *et al.*, 2020). Feral sheep foraged on and

trampled native vegetation, contributing to forest die-off before being removed by the mid 1980s (Wehtje, 1994; Junak, 1995). As the forests began rapidly recovering from sheep removal, two severe droughts caused extensive die-offs in the following 40 years (Taylor *et al.*, 2020). A severe drought from 1987-1991 caused extensive mortality (Wehtje, 1994; Walter & Taha, 1999) with the loss of 70-90% of island pines (Walter & Taha, 1999). A second event occurred from 2012-2016 (extended through 2017 in the south) when California experienced the most severe drought since record keeping began in the mid 1800s (Robeson, 2015; Taylor *et al.*, 2020). The severity of these mortality events combined with genetic drift in these small stands likely influenced the genetic differentiation we observed on SCI, where at least partial stand replacing events occurred (Baguskas *et al.*, 2014; Taylor *et al.*, 2020). In addition, the CP stand was distinguished from the other SCI stands by climate and soil variation and was differentiated in RDA-based GEA analyses (Fig. 4; see below). This suggests that local adaptation may be another mechanism contributing to genetic differentiation across fine spatial scales.

Given the narrow coastal and disjunct distribution of *P. muricata*, there has been concern about its persistence and evolutionary potential, as reflected by its listing as threatened by the IUCN (Farjon, 2013). Past studies indicated that these populations were not genetically depauperate (Millar, 1983, 1999; Hong *et al.*, 1993),and we also found that populations did not have reduced diversity, despite small population size and isolation. In fact, genetic diversity estimates were high within the range of comparable estimates from other recent DNA sequencing studies in conifers (*Picea glauca*, *Pic. sitchensis*, *Pic. mariana*, *Pic. pungens*, and *Pic. breweriana*, Haselhorst *et al.*, 2019; *Cryptomeria japonica*, Uchiyama *et al.*, 2012; *Pin. elliotii* and *Pin. taeda*, Acosta *et al.*,

2019; *Pic. abies*, Wang *et al.*, 2020). Additionally, consistently negative estimates for Tajima's *D* suggest population expansion (Tajima, 1989), inconsistent with the idea that these populations are experiencing ongoing contraction. While negative Tajima's *D* estimates can also arise from selective sweeps, it is unlikely to find this pattern across the whole genome (Stajich and Hahn, 2005); thus, genome-wide patterns of Tajima's *D* are likely due to such demographic changes. Our results are consistent with the suggestion of Millar (1999), that populations may have persisted as disjunct entities, experiencing periodic episodes of contraction, expansion, and recolonization in a metapopulation-like dynamic during shifting climates of the late Quaternary.

GEA analyses. Despite its limited range, both demographic (e.g., genetic drift) and adaptive (i.e., local adaptation to environmental variation) processes appear to have shaped spatial genetic structure in *P. muricata*, as indicated by IBD and IBE, respectively. We did, however, find a strong association between geographic and environmental distance (Mantel's $r = 0.5608$, $p = 0.001$), which could confound the inference of genetic-environment associations (Wang & Bradburd, 2014; Rellstab *et al.*, 2015; Forester *et al.*, 2016). Additionally, as the range of *P. muricata* is limited narrowly to the western coast of North America (Fig. 1a,b), genetic differentiation was strongly predicted by latitude (latitude-PC1: $r = 0.93$, $p < 0.001$). Studies using landscape genetic approaches have commonly corrected for population structure using partial Mantel tests or RDAs that partial out PC axes or Moran Eigenvalue Maps (MEMs; Manel *et al.*, 2010; Rellstab *et al.*, 2015). Here, we instead included latitude as a covariate in our models (statistically equivalent to partialling out latitude or PC 1) to reduce environmental

associations arising purely from demographic processes, allowing for a more conservative interpretation of both GEA and GEA offset analyses.

Both GEA methods indicated that allele frequency shifts across the disjunct coastal distribution of *P. muricata* were strongly associated with latitude, elevation, mean annual temperature (bio1), and composite soil variables loaded by soil organic carbon, sand, and clay (Fig. 4). Sampling sites in the northern part of the range were associated with lower temperature (bio1), lower elevation, and lower soil PC1 compared to those in the central and southern regions (Fig. 4). As might be expected due to different approaches and assumptions underlying the models in the two different analyses, there was some discordance between the RDA and GF associations (Fig. 4b). One factor likely contributing to discordance among the methods is that latitude and mean annual temperature (bio1) were strongly negatively correlated ($r = -0.9657$), skewing the linear regression in the RDA model. GF, on the other hand, uses regression tree approaches that are better equipped for variable selection with highly collinear data (Genuer & Tuleau-Malot, 2010). Still, both analyses found strong associations between allele frequencies and both elevation and mean annual temperature (bio1), indicating local adaptation to those variables, or to unmeasured variables correlated with them. Elevation is commonly associated with temperature, precipitation, and soil type, and can be used as a proxy for variables likely unmeasured within the data. Additionally, mean annual temperature (bio1) was implicated in both analyses. While mean annual temperature (bio1) has been shown to be important for local adaptation in pines, it is strongly correlated with other bioclimatic variables we removed from our model, each of which may represent more causal association (i.e., bio4, bio5, and bio8–19 range $|r|$: 0.844 – 0.967). Landscape

genetic studies of other *Pinus* species have identified similar genetic-environment associations, most consistently citing variation associated with geography (e.g., longitude, elevation), precipitation, and temperature (e.g., *P. taeda*, Eckert *et al.*, 2010a,b; *P. albicaulis*, Lind *et al.*, 2017; *P. pinaster*, Jaramillo-Correa *et al.*, 2015, *P. lambertiana*, Eckert *et al.*, 2015).

GEA offset. Genetic-environment association studies reveal how spatial genetic variation and its environmental correlates are distributed across landscapes, after attempting to control for neutral processes (Rellstab *et al.*, 2015; Hoban *et al.*, 2016; Forester *et al.*, 2018). Results of GEA analyses can then form the basis for genetic offset analyses that quantify how the identified genetic-environment associations may shift with changing climatic conditions, and thus how populations may become maladapted to novel environments (Fitzpatrick & Keller, 2015; Capblancq *et al.*, 2020a). Presumably due to their ecological importance and long life spans, forest trees have been the focus of numerous such studies (*Populus balsamifera*, Fitzpatrick & Keller, 2015; *Po. tremula*, Ingvarsson & Bernhardsson, 2020; *Quercus rugosa*, Martins *et al.*, 2018; *Q. suber*, Pina-Martins *et al.*, 2019; *Eucalyptus melliodora*, Supple *et al.*, 2018; *Pin. taeda*, Lu *et al.*, 2019; *Fagus sylvatica*, Capblancq *et al.*, 2020b; *Pin. contorta*, Mahony *et al.*, 2020; *Pin. cembra*, Dauphin *et al.*, 2021; *Po. balsamifera*, Gougherty *et al.*, 2021; reviewed thoroughly in Capblancq *et al.*, 2020a). Most genetic offset studies to date have focused on loci associated with adaptation by either conducting GEA outlier analyses or by using previously identified causal loci (e.g., Gougherty *et al*., 2021). Outlier-based offset analyses may have issues, however, as population structure and clinal associations can

result in false-positives (Lotterhos & Whitlock, 2015; Rellstab *et al.*, 2015), and local adaptation to similar environments can occur via different subsets of loci in separate populations (Crow *et al.*, 2021; Lopez-Arboleda *et al.*, 2021). GEA approaches (especially RDA), however, appear robust in identifying multilocus selection given complex demographic and evolutionary histories (Lotterhos & Whitlock, 2015; Forester *et al.*, 2016, 2018; Capblancq *et al.*, 2018). Here, we applied offset analyses based on all available loci, rather than merely outlier or causal loci, as this approach is known to perform well across diverse demographic scenarios in assessing how environmental associations will shift in response to predicted future climate (see Láruson *et al.*, 2021). We use the suggestion from Láruson *et al*. (2021) of referring to analysis-specific offset and thus use the term "GEA offset" (described in detail in Materials and Methods) to describe our results.

The degree of GEA offset varied based on analysis (RDA versus GF), climate projection severity (low change SSP1-2.6 versus extreme change SSP5-8.5), and time interval (2041–2060 versus 2081–2100). As expected, the relative degree of GEA offset increased with both model severity and time from present. GEA offset was higher for SSP5-8.5 than for SSP1-2.6, and increased across time within each model (Fig. 5). Mean annual temperature (bio1) was most strongly correlated with spatial allele frequency shifts, and also strongly predicted GEA offset (Supplemental Information Fig. S5). For example, the Lompoc (LO) population had the greatest predicted change in mean annual temperature and subsequently the most severe GEA offset across all climate projections and time intervals. We identified considerable geographic variation in the degree of maladaptation predicted across the sampled range with few inconsistencies between GF

offset and RDA offset (sup Fig. S5). While geographic variation in offset was not strongly consistent across latitude, models, or time frames, populations in northern regions were generally predicted to experience greater GEA offset, indicating increased levels of future maladaptation, while those in the central and southern regions tended to have lower levels of offset. A possible explanation for the lack of clear, north to south increasing maladaptation pattern is that *P. muricata* may be environmentally buffered along the coast. This could also be influenced by independent evolutionary histories in southern and northern populations.

Forest tree populations can respond to climate change via local adaptation, migration, or phenotypic plasticity (Aitken *et al.*, 2008). Genomic perspectives on maladaptation to future climates can be used to predict fitness outcomes across suitable environments and can inform conservation and restoration by identifying habitats and geographic regions of increased vulnerability or suitability (Bay *et al.*, 2018; Supple *et al.*, 2018; Capblancq *et al.*, 2020a; Capblancq & Forester, 2021; Gougherty *et al.*, 2021; Rellstab, 2021). Despite variation across space, time, and methods, all *P. muricata* populations are predicted to experience some GEA offset, suggesting persistence will rely to some degree on adaptation to changing climate. Evidence here for genetic differentiation across both broad and fine spatial scales, and the fact that *P. muricata* is distributed in a small number of highly disjunct populations, means that contemporary gene flow among disjunct populations is an unlikely source of novel allelic variation to aid adaptation to shifting climate. Further, the long life span of *P. muricata* will limit the rate of local adaptation to rapid climate change in isolated populations (as recognized in other long-lived conifers; e.g., Jia *et al.*, 2020). Given these constraints, assisted

migration (i.e., assisted gene flow; Aitken & Whitlock, 2013; Aitken & Bemmels, 2016; Grummer *et al.*, 2022) may be an approach worth future consideration to ameliorate threats to persistence by introducing and/or increasing adaptive variation in populations particularly at risk.

Conclusions. Consistent with past studies of morphology, allozymes, and reproductive compatibility, our results indicate defined spatial genetic structure across disjunct populations occurring in different geographic regions, as well as differentiation within these regions at much finer scales than is typical of pines. Although standing genetic variation does not appear reduced within populations, evidence for genetic differentiation among isolated, disjunct populations of *P. muricata* is consistent with a lack of contemporary gene flow among populations evolving independently in different locales and environmental conditions. Indeed, we identified specific soil and climate variables that contribute to allele frequency shifts across the distribution. Evolution to changing climate in *P. muricata* may thus be constrained by a lack of gene flow among different geographic regions and even populations, in addition to the environmental mismatch experienced by seedlings and much older adult trees. The large and genome-wide set of polymorphisms analyzed here provided fine-scale resolution of spatial genetic structure, but also facilitated inference of the environmental variables shaping spatial genetic variation. Our analyses suggest that the consequences of a changing climate may include local maladaptation. Although the extent of genetic offset varied across climate model projections and time intervals, geographic variation in the extent of potential maladaptation was a consistent feature of our results. Both *ex situ* and *in situ*

conservation considerations for isolated populations of this relatively rare pine should recognize this endemic genetic variation and its potential causes, in addition to the impact that climate change may have on the fitness of local populations.

Acknowledgements

This research was funded by startup funds to TLP from the University of Nevada Reno, in addition to funds from the BLM. We thank John Knapp from The Nature Conservancy and Mark DiMaggio from Paso Robles High School for collecting samples from Santa Cruz Island. We thank Julia Harenčar from the University of California-Santa Cruz for collecting samples from Monterey.

Data availability

The data that support the findings of this study are openly available at the Dryad Digital Repository (doi: 10.5061/dryad.sqv9s4n61; https://doi.org/10.5061/dryad.sqv9s4n61) and NCBI's Short Read Archive (accession PRJNA824947; https://dataview.ncbi.nlm.nih.gov/object/PRJNA824947).

References

Acosta JJ, Fahrenkrog AM, Neves LG, Resende MFR, Dervinis C, Davis JM, Holliday JA, Kirst M. **2019**. Exome resequencing reveals evolutionary history, genomic diversity, and targets of selection in the conifers *Pinus taeda* and *Pinus elliottii*. *Genome Biology and Evolution* **11**: 508–520.

Ægisdóttir HH, Kuss P, Stöcklin J. **2009**. Isolated populations of a rare alpine plant show high genetic diversity and considerable population differentiation. *Annals of Botany* **104**: 1313–1322.

Aitken SN, Bemmels JB. **2016**. Time to get moving: assisted gene flow of forest trees. *Evolutionary Applications* **9**: 271–290.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. **2008**. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**: 95–111.

Aitken SN, Whitlock MC. **2013**. Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics* **44**: 367–388.

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, *et al.* 2013. Potential for evolutionary responses to climate change--evidence from tree populations. *Global Change Biology* **19**: 1645–1661.

Axelrod DI. **1967a**. Proceedings of the Symposium on the Biology of the California Islands: Evolution of the Californian closed-cone pine forest. In: Santa Barbara, California: Santa Barbara Botanic Garden in collaboration with the California Native Plant Society, 91–149.

Axelrod DI. **1967b**. Proceedings of the Symposium on the Biology of the California Islands: Geologic history of the Californian insular flor. In: Santa Barbara, California: Santa Barbara Botanic Garden in collaboration with the California Native Plant Society, 267–315.

Axelrod DI. **1980**. *History of the maritime closed-cone pines, Alta and Baja California*. Berkeley, CA: University of California Press.

Axelrod DI. **1981**. Holocene climatic changes in relation to vegetation disjunction and speciation. *The American Naturalist* **117**: 847–870.

Axelrod DI. **1986**. Cenozoic history of some western American pines. *Annals of the Missouri Botanical Garden* **73**: 565–641.

Baguskas SA, Peterson SH, Bookhagen B, Still CJ. **2014**. Evaluating spatial patterns of drought-induced tree mortality in a coastal California pine forest. *Forest Ecology and Management* **315**: 43–53.

Bay RA, Harrigan RJ, Le Underwood V, Gibbs HL, Smith TB, Ruegg K. **2018**. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**: 83–86.

Brauer CJ, Hammer MP, Beheregaray LB. **2016**. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology* **25**: 5093–5113.

Capblancq T, Fitzpatrick MC, Bay RA, Exposito-Alonso M, Keller SR. **2020a**. Genomic prediction of (mal)adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics* **51**: 245–269.

Capblancq T, Forester BR. **2021**. Redundancy analysis: A Swiss Army Knife for landscape genomics.

Methods in Ecology and Evolution **12**: 2298–2309.

Capblancq T, Luu K, Blum MGB, Bazin E. **2018**. Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources* **18**: 1223–1233.

Capblancq T, Morin X, Gueguen M, Renaud J, Lobreaux S, Bazin E. **2020b**. Climate-associated genetic variation in *Fagus sylvatica* and potential responses to climate change in the French Alps. *Journal of Evolutionary Biology* **33**: 783–796.

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

Critchfield WB. **1967**. Crossability and relationships of the closed-cone pines. *Silvae Genetica* **16**: 89–97.

Critchfield WB, Little EL. **1966**. *Geographic Distribution of the Pines of the World*. US Department of Agriculture Forest Service, Washington, DC.

Crow TM, Buerkle CA, Runcie DE, Hufford KM. **2021**. Implications of genetic heterogeneity for plant translocation during ecological restoration. *Ecology and Evolution* **11**: 1100–1110.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Dauphin B, Rellstab C, Schmid M, Zoller S, Karger DN, Brodbeck S, Guillaume F, Gugerli F. **2021**. Genomic vulnerability to rapid climate warming in a tree species with a long generation time. *Global Change Biology* **27**: 1181–1195.

Davis KT, Dobrowski SZ, Higuera PE, Holden ZA, Veblen TT, Rother MT, Parks SA, Sala A, Maneta MP. **2019**. Wildfires and climate change push low-elevation forests across a critical climate threshold for tree regeneration. *Proceedings of the National Academy of Sciences* **116**: 6193–6198.

De La Torre AR, Puiu D, Crepeau MW, Stevens K, Salzberg SL, Langley CH, Neale DB. **2019**. Genomic architecture of complex traits in loblolly pine. *New Phytologist* **221**: 1789–1801.

De Villemereuil P, Gaggiotti OE. **2015**. A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution* **6**: 1248–1258.

Duffield JW. **1951**. *Interrelationships of the California closed-cone pines: with special reference to Pinus muricata D. Don*. PhD thesis, University of California, Berkeley, Berkeley, CA, USA.

Durvasula A, Hoffman PJ, Kent TV, Liu C, Kono TJY, Morrell PL, Ross-Ibarra J. **2016**. ANGSD-wrapper: utilities for analysing next-generation sequencing data. *Molecular Ecology Resources* **16**: 1449–1454.

Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB. **2010a**. Back to nature: ecological genomics of loblolly pine (*Pinus taeda,* Pinaceae). *Molecular Ecology* **19**: 3789–3805.

Eckert AJ, Hall BD. **2006**. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): Phylogenetic tests of fossil-based hypotheses. *Molecular Phylogenetics and Evolution* **40**: 166–182.

Eckert AJ, Maloney PE, Vogler DR, Jensen CE, Mix AD, Neale DB. **2015**. Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genetics & Genomes* **11**: 42.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC, Neale DB. **2010b**. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**: 969–982.

Ellegren H. **2014**. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution* **29**: 51–63.

Ellis JR, Pashley CH, Burke JM, Mccauley DE. **2006**. High genetic diversity in a rare and endangered sunflower as compared to a common congener. *Molecular Ecology* **15**: 2345–2355.

Ellis N, Smith SJ, Pitcher CR. **2012**. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**: 156–168.

Escaravage N, Cambecedes J, Largier G, Pornon A. **2011**. Conservation genetics of the rare Pyreneo-Cantabrian endemic *Aster pyrenaeus* (Asteraceae). *AoB Plants*.

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE. **2016**. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9**: 1937–1958.

Farjon A. **2008**. *A natural history of conifers*. Portland, OR: Timber Press.

Farjon A. **2013**. *Pinus muricata*. *The IUCN Red List of Threatened Species*. [WWW document] URL https://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T34058A2841776.en. Accessed on 4 January 2022.

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

Fischer DT, Still CJ, Williams AP. **2009**. Significance of summer fog and overcast for drought stress and ecological functioning of coastal California endemic plant species. *Journal of Biogeography* **36**: 783–799.

Fitzpatrick MC, Blois JL, Williams JW, Nieto-Lugilde D, Maguire KC, Lorenz DJ. **2018**. How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. *Global Change Biology* **24**: 3575–3586.

Fitzpatrick MC, Keller SR. **2015**. Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology letters* **18**: 1–16.

Forde MB, Blight MM. **1964**. Geographical variation in the turpentine of Bishop pine. *New Zealand Journal of Botany* **2**: 44–52.

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

Forester BR, Lasky JR, Wagner HH, Urban DL. **2018**. Comparing methods for detecting multilocus adaptation with multivariate genotype--environment associations. *Molecular Ecology* **27**: 2215–2233.

Frichot E, François O. **2015**. *A short manual for LFMM version 1.4*. [WWW document] URL http://membres-timc.imag.fr/Olivier.Francois/lfmm/files/note.pdf [accessed 2 March 2022].

Frichot E, Schoville SD, 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.

Fu L, Niu B, Zhu Z, Wu S, Li W. **2012**. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Genuer R, Poggi J-M, Tuleau-Malot C. **2010**. Variable selection using random forests. *Pattern Recognition Letters* **31**: 2225–2236.

Gernandt DS, Aguirre Dugua X, Vázquez-Lobo A, Willyard A, Moreno Letelier A, de la Rosa JA, Piñero D, Liston A. **2018**. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection Australes. *American Journal of Botany* **105**: 711–725.

Gernandt DS, López GG, García SO, Liston A. **2005**. Phylogeny and classification of *Pinus*. *Taxon* **54**: 29–42.

Gitzendanner MA, Soltis PS. **2000**. Patterns of genetic variation in rare and widespread plant congeners. *American Journal of Botany* **87**: 783–792.

Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, Nice CC. **2014**. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology* **23**: 4555–4573.

González-Martínez SC, Krutovsky KV, Neale DB. **2006**. Forest-tree population genomics and adaptive evolution. *New Phytologist* **170**: 227–238.

Gougherty AV, Keller SR, Fitzpatrick MC. **2021**. Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change* **11**: 166–171.

Grummer JA, Booker TR, Matthey-Doret R, Nietlisbach P, Thomaz AT, Whitlock MC. **2022**. The immediate costs and long-term benefits of assisted gene flow in large populations. *Conservation Biology*: e13911.

Günther T, Coop G. **2013**. Robust identification of local adaptation from allele frequencies. *Genetics* **195**: 205–220.

Hall D, Olsson J, Zhao W, Kroon J, Wennström U, Wang X-R. **2021**. Divergent patterns between phenotypic and genetic variation in Scots pine. *Plant Communications* **2**: 100139.

Halofsky JE, Peterson DL, Harvey BJ. **2020**. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA. *Fire Ecology* **16**: 1–26.

Haselhorst MSH, Parchman TL, Buerkle CA. **2019**. Genetic evidence for species cohesion, substructure and hybrids in spruce. *Molecular Ecology* **28**: 2029–2045.

He T, Pausas JG, Belcher CM, Schwilk DW, Lamont BB. **2012**. Fire-adapted traits of *Pinus* arose in the fiery Cretaceous. *New Phytologist* **194**: 751–759.

Hernandez-Leon S, Gernandt DS, de la Rosa JA, Jardon-Barbolla L. **2013**. Phylogenetic relationships

and species delimitation in *Pinus* section Trifoliae inferrred from plastid DNA. *PloS one* **8**: e70501.

Heusser L. **1978**. Pollen in Santa Barbara Basin, California: a 12,000-yr record. *Geological Society of America Bulletin* **89**: 673–678.

Heusser CJ, Heusser LE, Peteet DM. **1985**. Late-Quaternary climatic change on the American North Pacific coast. *Nature* **315**: 485–487.

Heusser LE, Sirocko F. **1997**. Millennial pulsing of environmental change in southern California from the past 24 ky: A record of Indo-Pacific ENSO events? *Geology* **25**: 243–246.

Hewitt GM. **2004**. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**: 183–195.

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

Hong Y-P, Hipkins VD, Strauss SH. **1993**. Chloroplast DNA diversity among trees, populations and species in the California closed-cone pines (*Pinus radiata, Pinus muricata and Pinus attenuata*). *Genetics* **135**: 1187–1196.

Honnay O, Jacquemyn H. **2007**. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conservation Biology* **21**: 823–831.

Hudson RR, Slatkin M, Maddison WP. **1992**. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Ingvarsson PK, Bernhardsson C. **2020**. Genome-wide signatures of environmental adaptation in European aspen (*Populus tremula)* under current and future climate conditions. *Evolutionary Applications* **13**: 132–142.

Jaramillo-Correa J-P, Rodríguez-Quilón I, Grivet D, Lepoittevin C, Sebastiani F, Heuertz M, Garnier-Géré PH, Alía R, Plomion C, Vendramin GG, *et al.* 2015. Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster* Aiton, Pinaceae). *Genetics* **199**: 793–807.

Jia K-H, Zhao W, Maier PA, Hu X-G, Jin Y, Zhou S-S, Jiao S-Q, El-Kassaby YA, Wang T, Wang X-R, *et al.* 2020. Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Evolutionary Applications* **13**: 665–676.

Junak S. **1995**. *Flora of Santa Cruz Island*. Santa Barbara Botanic Garden in collaboration with the California Native Plant Society.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. **2017**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Korneliussen TS, Albrechtsen A, Nielsen R. **2014**. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**: 356.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. **2013**. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* **14**: 289.

Kramer AT, Ison JL, Ashley MV, Howe HF. **2008**. The paradox of forest fragmentation genetics. *Conservation Biology* **22**: 878–885.

Lamont BB, Enright NJ. **2000**. Adaptive advantages of aerial seed banks. *Plant Species Biology* **15**: 157–166.

Langlet O. **1971**. Two hundred years genecology. *Taxon* **21**: 653–721.

Langmead B, Salzberg SL. **2012**. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357.

Láruson ÁJ, Fitzpatrick MC, Keller SR, Haller BC, Lotterhos KE. **2021**. Seeing the forest for the trees: Assessing genetic offset predictions with Gradient Forest. *bioRxiv*. doi: 10.1101/2021.09.20.461151

Legendre P, Legendre LFJ. **2012**. *Numerical Ecology*. Elsevier.

Lenihan JM, Bachelet D, Neilson RP, Drapek R. **2008**. Response of vegetation distribution, ecosystem productivity, and fire to climate change scenarios for California. *Climatic Change* **87**: 215–230.

Lewis PO. **2001**. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**: 913–925.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Proc. **2009**. The sequence slignment/map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lind BM, Friedline CJ, Wegrzyn JL, Maloney PE, Vogler DR, Neale DB, Eckert AJ. **2017**. Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology* **26**: 3168–3185.

Lind BM, Menon M, Bolte CE, Faske TM, Eckert AJ. **2018**. The genomics of local adaptation in trees: Are we out of the woods yet? *Tree Genetics & Genomes* **14**: 29.

Linhart YB, Burr B, Conkle MT. **1965**. The closed-cone pines of the northern Channel Islands. *Proceedings of the Symposium on the biology of the California islands. Santa Barbara Botanic Garden*: 151–177.

Little EL. **1975**. *Rare and local conifers in the United States*. Washington, D.C.: US Department of Agriculture, Forest Service.

Lopez-Arboleda WA, Reinert S, Nordborg M, Korte A. **2021**. Global genetic heterogeneity in adaptive traits. *Molecular Biology and Evolution* **38**: 4822–4831.

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

Lu M, Krutovsky KV, Loopstra CA. **2019**. Predicting adaptive genetic variation of loblolly pine (*Pinus taeda* L.) populations under projected future climates based on multivariate models. *Journal of Heredity* **110**: 857–865.

Mahony CR, MacLachlan IR, Lind BM, Yoder JB, Wang T, Aitken SN. **2020**. Evaluating genomic data for management of local adaptation in a changing climate: a lodgepole pine case study. *Evolutionary Applications* **13**: 116–131.

Manel S, Poncet BN, Legendre P, Guerli F, Holderegger R. **2010**. Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology* **19**: 3824–3835.

Mann DH, Hamilton TD. **1995**. Late Pleistocene and Holocene paleoenvironments of the North Pacific coast. *Quaternary Science Reviews* **14**: 449–471.

Mantel N. **1967**. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.

Martins K, Gugger PF, Llanderal-Mendoza J, González-Rodríguez A, Fitz-Gibbon ST, Zhao J-L, Rodríguez-Correa H, Oyama K, Sork VL. **2018**. Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evolutionary Applications* **11**: 1842–1858.

Matyas C. **1996**. Climatic adaptation of trees: rediscovering provenance tests. *Euphytica* **92**: 45–54.

McDowell NG, Fisher RA, Xu C, Domec JC, Hölttä T, Mackay DS, Sperry JS, Boutz A, Dickman L, Gehres N, *et al.* 2013. Evaluating theories of drought-induced vegetation mortality using a multimodel--experiment framework. *New Phytologist* **200**: 304–321.

Mcdowell NG, Williams AP, Xu C, Pockman WT, Dickman LT, Sevanto S, Pangle R, Limousin J, Plaut J, Mackay DS, *et al.* 2016. Multi-scale predictions of massive conifer mortality due to chronic temperature rise. *Nature Climate Change* **6**: 295–300.

McKinney GJ, Waples RK, Seeb LW, Seeb JE. **2017**. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources* **17**: 656–669.

Millar CI. **1983**. A steep cline in *Pinus muricata*. *Evolution* **37**: 311–319.

Millar CI. **1986**. The Californian closed cone pines (subsection Oocarpae Little and Critchfield): a taxonomic history and review. *Taxon* **35**: 657–670.

Millar CI. **1988**. Allozyme differentiation and biosystematics of the Californian closed-cone pines (*Pinus* subsection Oocarpae). *American Society of Plant Taxonomists* **13**: 351–370.

Millar CI. **1989**. Allozyme variation of bishop pine associated with pygmy-forest soils in northern California. *Canadian Journal of Forest Research* **19**: 870–879.

Millar CI. **1999**. Evolution and biogeography of *Pinus radiata*, with a proposed revision of its Quaternary history. *New Zealand Journal of Forestry Science* **29**: 335–365.

Millar CI, Critchfield WB. **1988**. Crossability and relationships of *Pinus muricata* (Pinaceae). *Madrono* **35**: 39–53.

Mirov NT, Zavarin E, Snajberk K, Costello K. **1966**. Further studies of turpentine composition of *Pinus muricata* in relation to its taxonomy. *Phytochemistry* **5**: 343–355.

Morgenstern EK. **1996**. *Geographic Variation in Forest Trees: genetic basis and application of knowledge in silviculture*. Vancouver, Canada: UBC Press.

Nei M. **1972**. Genetic distance between populations. *American Naturalist* **106**: 192–283.

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. **2015**. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.

Nybom H. **2004**. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* **13**: 1143–1155.

O'Connor CD, Falk DA, Lynch AM, Swetnam TW. **2014**. Fire severity, size, and climate associations diverge from historical precedent along an ecological gradient in the Pinaleño Mountains, Arizona, USA. *Forest Ecology and Management* **329**: 264–278.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, *et al.* 2019. vegan: Community Ecology Package. 2019. *R package version 2.5-6*.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. **2013**. Vegan: community ecology package. *R package version 2.0-9*.

Ortiz EM. **2019**. vcf2phylip v2. 0: convert a VCF matrix into several matrix formats for phylogenetic analysis. doi: 10.5281/zenodo.1257057.

Parchman TL, Gompert Z, Mudge J, Schilkey F, Benkman CW, Buerkle CA. **2012**. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**: 2991–3005.

Petit RJ, Hampe A. **2006**. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* **37**: 187–214.

Pina-Martins F, Baptista J, Pappas Jr G, Paulo OS. **2019**. New insights into adaptation and population structure of cork oak using genotyping by sequencing. *Global Change Biology* **25**: 337–350.

Poggio L, de Sousa LM, Batjes NH, Heuvelink G, Kempen B, Ribeiro E, Rossiter D. **2021**. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**: 217–240.

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

Prunier J, Verta J-P, MacKay JJ. **2016**. Conifer genomics and adaptation: at the crossroads of genetic diversity and genome function. *New Phytologist* **209**: 44–62.

Rellstab C. **2021**. Genomics helps to predict maladaptation to climate change. *Nature Climate Change* **11**: 85–86.

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

Revell LJ. **2012**. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*: 217–223.

Revell LJ, Revell MLJ. **2014**. *Package `phytools'*. [WWW document] URL https://cran.r-project.org/web/packa ges/phytools [accessed 2 February 2022].

Robeson SM. **2015**. Revisiting the recent California drought as an extreme value. *Geophysical Research Letters* **42**: 6771–6779.

Ruegg K, Bay RA, Anderson EC, Saracco JF, Harrigan RJ, Whitfield M, Paxton EH, Smith TB. **2018**. Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters* **21**: 1085–1096.

Saladín B, Leslie AB, Wüest RO, Litsios G, Conti E, Salamin N, Zimmermann NE. **2017**. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology* **17**: 1–15.

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

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

Slavov GT, Zhelev P. **2004**. Allozyme variation, differentiation, and inbreeding in populations of *Pinus mugo* in Bulgaria. *Canadian Journal of Forest Research* **34**: 2611–2617.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. **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.

Stajich JE, Hahn MH. **2005**. Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution* **22**: 63–73.

Strauss SH, Hong Y-P, Hipkins VD. **1993**. High levels of population differentiation for mitochondrial DNA haplotypes in *Pinus radiata, muricata,* and *attenuata. Theoretical and Applied Genetics* **86**: 605–611.

Supple MA, Bragg JG, Broadhurst LM, Nicotra AB, Byrne M, Andrew RL, Widdup A, Aitken NC, Borevitz JO. **2018**. Landscape genomic prediction for restoration of a Eucalyptus foundation species under climate change. *Elife* **7**: e31835.

Tajima F. **1983**. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.

Tajima F. **1989**. Statistical-method for testing the neutral mutation hypothesis by {DNA} polymorphism. *Genetics* **123**: 585–595.

Taylor A, Biswas T, Randall JM, Klausmeyer K, Cohen B. **2020**. Parched pines: a quantitative comparison of two multi-year droughts and associated mass mortalities of bishop pine (*Pinus muricata*) on Santa Cruz Island, California. *Remote Sensing in Ecology and Conservation* **6**: 20–34.

Tíscar PA, Lucas-Borja ME, Candel-Pérez D. **2018**. Lack of local adaptation to the establishment conditions limits assisted migration to adapt drought-prone *Pinus nigra* populations to climate change. *Forest Ecology and Management* **409**: 719–728.

Uchiyama K, Ujino-Ihara T, Ueno S, Taguchi Y, Futamura N, Shinohara K, Tsumura Y. **2012**. Single nucleotide polymorphisms in *Cryptomeria japonica*: their discovery and validation for genome mapping and diversity studies. *Tree Genetics & Genomes* **8**: 1213–1222.

Vieira FG, Fumagalli M, Albrechtsen A, Nielsen R. **2013**. Estimating inbreeding coefficients from NGS data: impact on genotype calling and allele frequency estimation. *Genome Research* **23**: 1852–1861.

Van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, *et al.* 2011. The representative concentration pathways: an overview. *Climatic Change* **109**: 5–31.

Venables WN, Ripley BD. **2002**. MASS (R package). In: Modern Applied Statistics with S. New York: Springer Verlag.

Walisch TJ, Matthies D, Hermant S, Colling G. **2015**. Genetic structure of *Saxifraga rosacea* subsp. *sponhemica*, a rare endemic rock plant of Central Europe. *Plant Systematics and Evolution* **301**: 251–263.

Walter HS, Taha LA. **1999**. Regeneration of bishop pine (*Pinus muricata*) in the absence and presence of fire: a case study from Santa Cruz Island, California. In: Proceedings of the fifth California Islands symposium. 38–99.

Wang X, Bernhardsson C, Ingvarsson PK. **2020**. Demography and natural selection have shaped genetic variation in the widely distributed conifer Norway spruce (*Picea abies*). *Genome Biology and Evolution* **12**: 3803–3817.

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

Watterson GA. **1975**. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**: 256–276.

Wehtje W. **1994**. *Response of a Bishop pine (Pinus muricata) population to removal of feral sheep on Santa Cruz Island, California*.

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

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, *et al.* 2016. Convergent local adaptation to climate in distantly related conifers. *Science* **353**: 1431–1433.

Young A, Boyle T, Brown T. **1996**. The population genetic consequences of habitat fragmentation for plants. *Trends in ecology & evolution* **11**: 413–418.

Tables

Table 1 We collected 213 individuals from 12 locations spanning the range of *P. muricata*. Population abbreviations (in parentheses) correspond to those in Fig. 1. Population genetic estimates based on sequence variation included nucleotide diversity Θ_π, or the average number of pairwise differences between sequences, and Tajima's *D*, or the scaled difference between Θ_π and the number of segregating sites (Θ_W; see Supporting Information Table S2 for an expanded list of diversity metrics with 95% confidence intervals). Finally, we calculated the mean individual inbreeding coefficient (*F)* at each sampling location.

Site	**Lat, Long**	***N***	**Elev. (m)**	Θ_π	***D***	***F***
Patrick's Point (PP)	41.140, -124.154	9	58	0.00616	-0.0251	0.0741
Navarro River (NR)	39.195, -123.765	4	75	0.00727	-0.0168	0.0362
Point Arena (PA)	38.876, -123.663	14	27	0.00719	-0.0567	0.0261
Salt Point (SP)	38.577, -123.334	20	51	0.00729	-0.0612	0.0081
Fort Ross (FR)	38.519, -123.247	19	59	0.00698	-0.0536	0.0312
Point Reyes (PR)	38.063, -122.849	23	241	0.00689	-0.0582	0.0062
Del Monte (DM)	36.594, -121.926	19	142	0.00674	-0.0470	0.1430
Diablo Canyon (DC)	35.245, -120.879	3	379	0.00831	-0.0715	0.0220
Lompoc (LO)	34.734, -120.440	28	292	0.00788	-0.0152	0.0990
Christy Pines (CP)	34.014, -119.797	37	421	0.00629	-0.0496	0.0350
Pelican Bay (PB)	34.024, -119.692	18	87	0.00617	-0.0370	0.0899
China Pines (CH)	34.003, -119.614	19	363	0.00625	-0.0447	0.1058

Figure legends

Fig. 1 Genetic structuring among populations was pronounced and followed a latitudinal gradient. Sampling sites along the west coast of North America (a), where sites correspond to those in Table 1 and include Patrick's Point (PP), Navarro River (NR), Point Arena (PA), Salt Point (SP), Fort Ross (FR), Point Reyes (PR), Del Monte (DM), Diablo Canyon (DC), and Lompoc (LO). Dark green shading represents the limited distribution of *P. muricata* along coastal California (distribution map from the Conservation Biology Institute, Data Basin, databasin.org) and provides support for the sampling regime at nearly all locations, where lines indicate sampling sites. The red box in panel (a) represents Santa Cruz Island, illustrated in greater detail in panel (b), with sampling sites including Christy Pines (CP), Pelican Bay (PB), and China Pines (CH). Genetic variation was visualized using principal components analysis (PCA) based on genotype probabilities from `entropy` for all individuals sampled (c), where a clear latitudinal gradient is evident (PC 1 versus latitude; inset, panel (c)). Colors in panel (c) correspond to colors and sampling locations in panels (a) and (b). Finally, ancestry estimates (q) from `entropy` for $k = 3$ (best fit model, top) and $k = 4$ (bottom) models are shown in panel (d), where vertical bars represent individuals and colors correspond to admixture proportions. See Supporting Information Fig. S1 for additional plots of the $k = 2$ and $k = 5$ models.

Fig. 2 Stronger regional genetic differentiation was evident among southern populations. We ran PCA using genotype probabilities from `entropy` of only the north

coast populations (a), and subsequently plotted PC 1 against latitude (b). Additionally, we ran PCA of only the Santa Cruz Island populations (c). Colors correspond to those in Fig. 1a,b,c.

Fig. 3 Environmental variation was associated with spatial genetic structure. RDA associated genotype probabilities from `entropy` with the set of predictor variables (a). Arrow length and direction correspond to loadings of each variable onto the RDA axes. The inset plot shows the percent of variance explained (PVE) for each of the first 6 RDA axes (cuml. *PVE* = 93.62%), with RDA axis 1 explaining the majority of variance (*PVE* = 57.96%). Colors match with the map and PCA in Fig. 1, and GEA variables bio1, bio3, bio4, and bio7 refer to mean annual temperature, isothermality, temperature seasonality, and temperature annual range, respectively. Relative importance (%) is shown for each variable for both RDA (dark grey; weighted axis loadings) and GF (light grey; R^2) along with the correlation between the two (b). Numbers accompanying bars correspond to the rank of each variable within each method, where a rank of 1 indicates the variable explaining the greatest portion of spatial genetic structuring within the particular method.

Fig. 4 Phylogenetic inference supported population genetic structure across the landscape, showing evidence of southern origin populations and subsequent northward expansion. Maximum likelihood phylogeny of *P. muricata* with *P. radiata* as an outgroup, inferred with `IQ-TREE` (a). The scale bar represents the expected number of nucleotide substitutions per site. Ultrafast bootstrap support values are indicated in the branches. Projected phylogeny onto the California geographic map showing each

population's location (b). One individual is represented for each population to avoid overlapping.

Fig. 5 Geographic variation in GEA offset across different time intervals, climate change scenarios, and modeling approaches. The degree of GEA offset for each population is represented as the difference between present conditions and four projected time interval/SSP combinations using both RDA (top row) and GF (bottom row). While there was some discordance across methods, sites suffering the most severe GEA offset remained consistent. Additionally, dramatic shifts in associations were predicted under both offset modeling approaches (RDA and GF) at time interval 2081-2100 for the SSP1-2.6 climate scenario, and at both time intervals for the SSP5-8.5 climate scenario.

Figures

Fig. 1

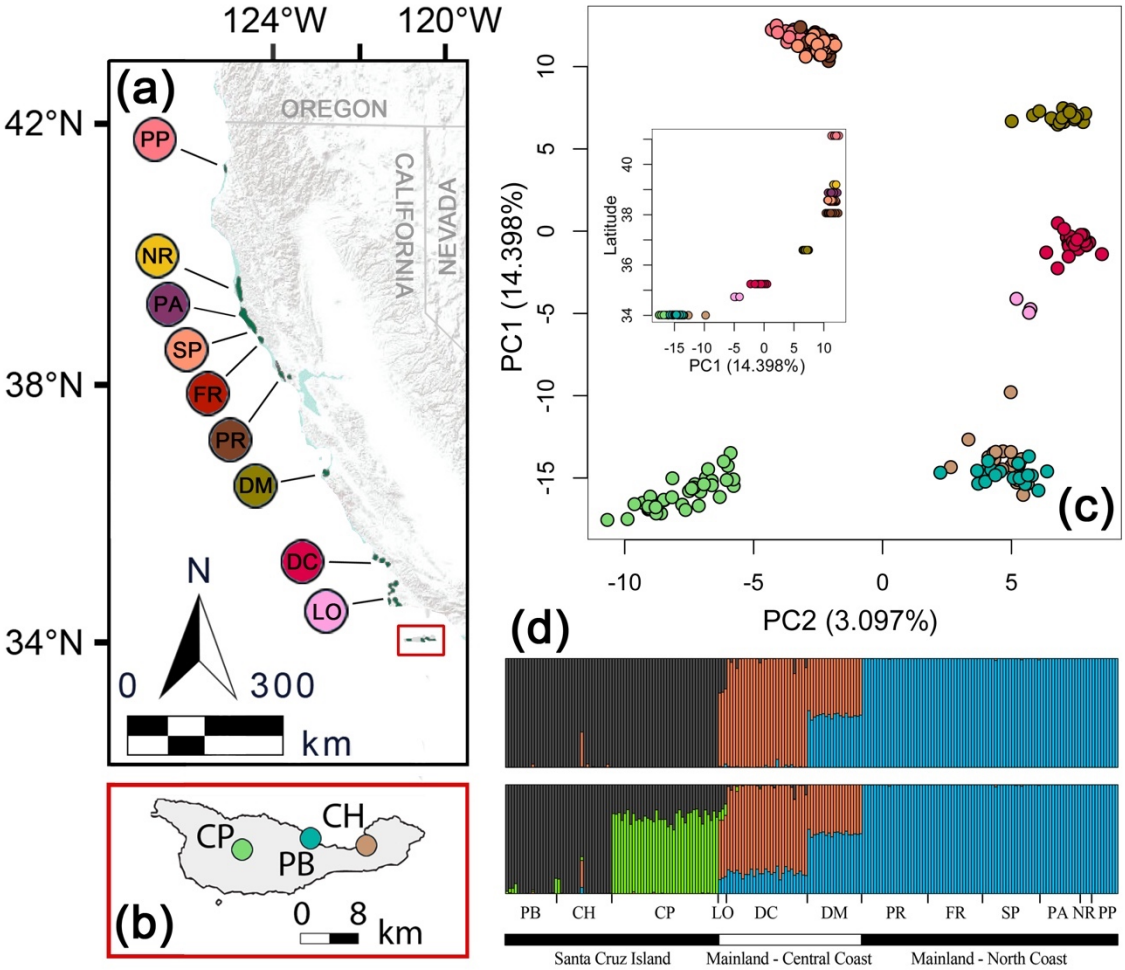


Fig. 2

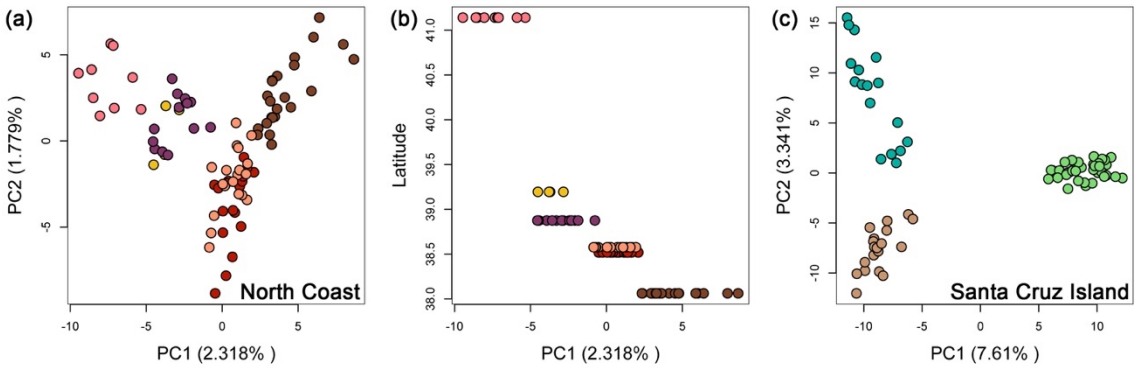

(a)
PC2 (1.779%)
PC1 (2.318%)
North Coast
(b)
Latitude
PC1 (2.318%)
(c)
PC2 (3.341%)
PC1 (7.61%)
Santa Cruz Island

Fig. 3

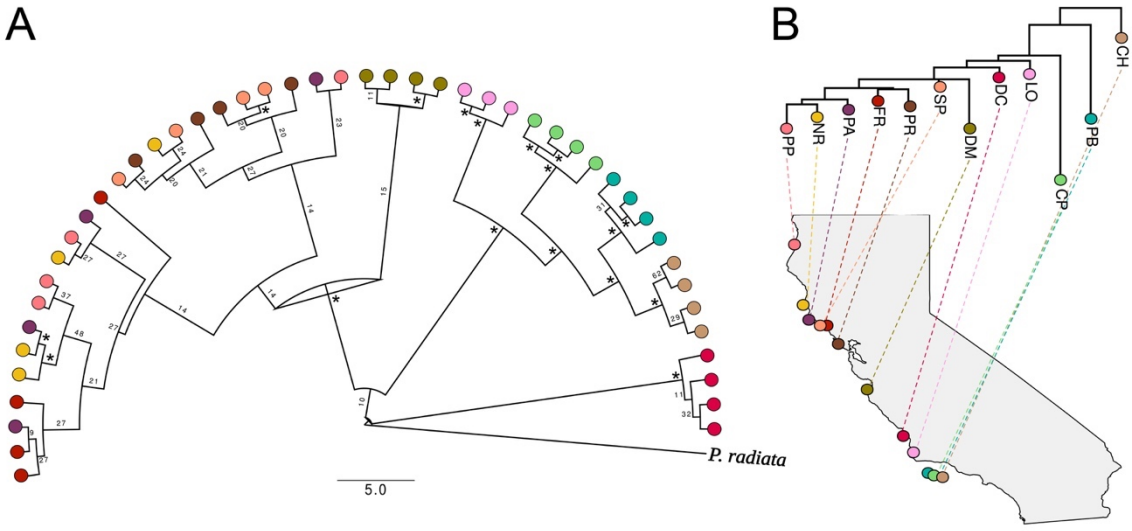


Fig. 4

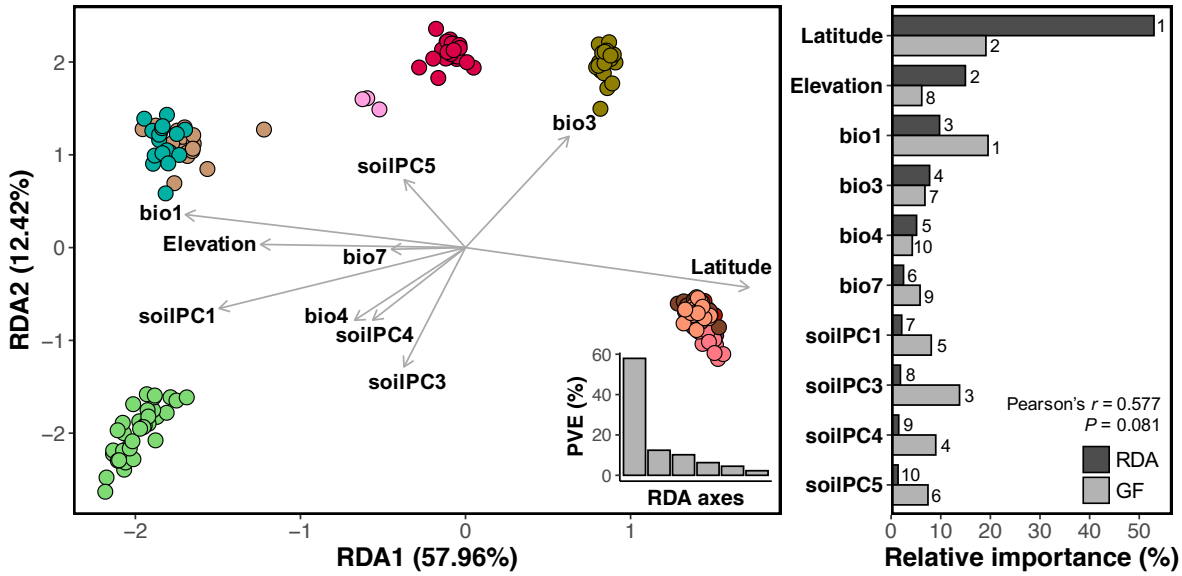
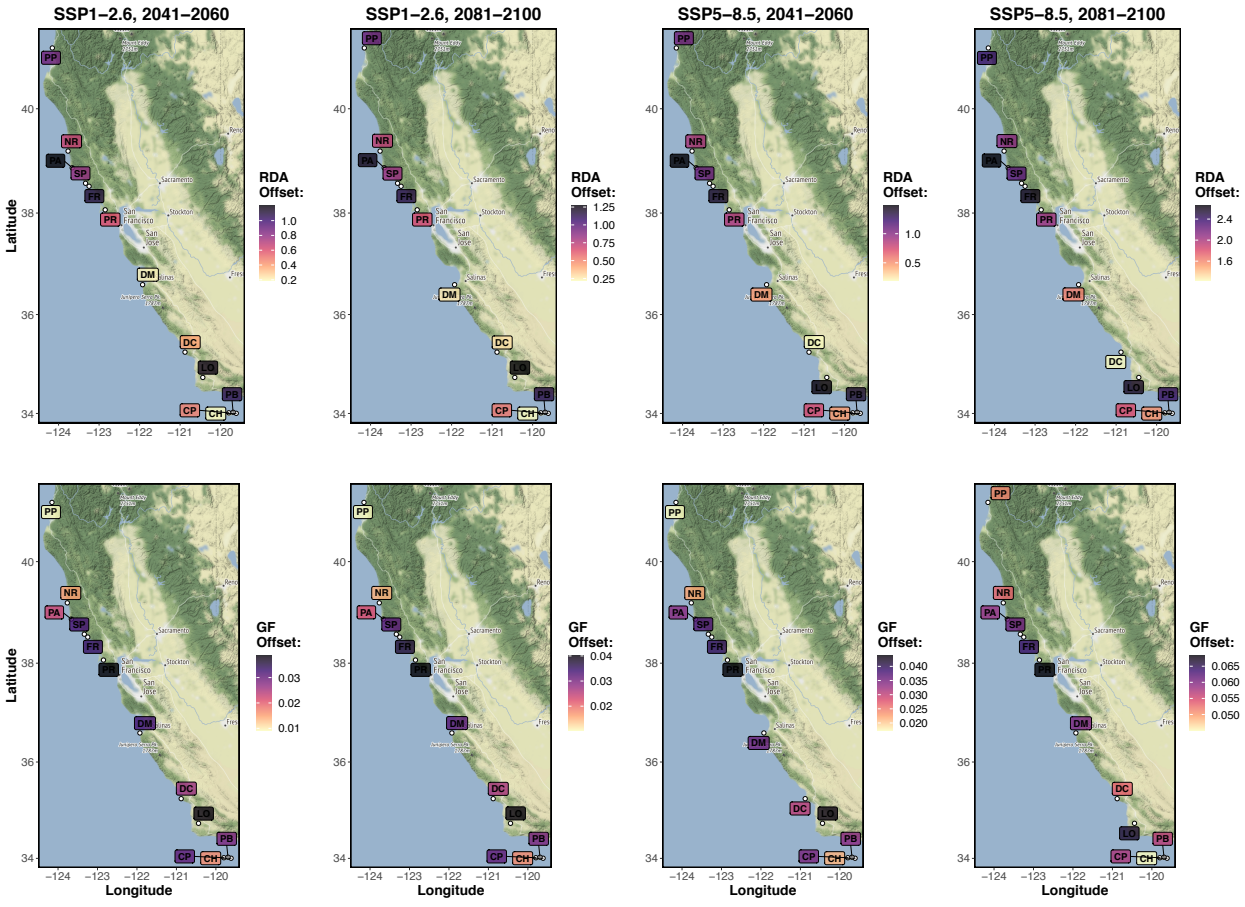

RDA2 (12.42%)
RDA1 (57.96%)
bio1
Elevation
bio7
soilPC1
bio4
soilPC4
soilPC3
soilPC5
bio3
Latitude
PVE (%)
RDA axes
Latitude
Elevation
bio1
bio3
bio4
bio7
soilPC1
soilPC3
soilPC4
soilPC5
Pearson's r = 0.577
P = 0.081
RDA
GF
Relative importance (%)

Fig. 5


SSP1−2.6, 2041−2060
SSP1−2.6, 2081−2100
SSP5−8.5, 2041−2060
SSP5−8.5, 2081−2100
RDA Offset:
GF Offset:
Latitude
Longitude
PP
NR
PA
SP
FR
PR
DM
DC
LG
PB
CP
CH

Supporting Information

1. Soil data extraction and summarization

To include aspects of soil variation, we downloaded data from SoilGrids 2.0 (www.soilgrids.org; Poggio *et al*., 2021). Across sampling sites, we extracted data for ten soil variables at six soil depths (0-5, 5-15, 15-30, 30-60, 60-100, 100-200 cm). Because sampling sites occurred on along the coast and SoilGrids resolution is ~250 x 250m, four locations were assigned no data as the average of the grid overlapped with the Pacific Ocean. For these sites, we averaged the three nearest grids with positive estimates as the point estimate of missing values for each soil variable. Next, we assessed correlations and standard deviations across soil depths, within each variable at each locality, to decide between two options: average across all depths or split into a shallow and deep depth variable. The final dataset included 14 soil variables. Exact descriptions of variables, depth decisions, and units are available online. For the final analyses, the 14 soil variables were highly correlated and collapsed into composite variables using principal components analyses (PCA) for downstream analyses (Supporting Information Table S3). See Materials and Methods in main text for additional information.

2. Future climatic variable extraction and summarization

Future climate data was extracted from WorldClim v2.1 CMIP6 (Eyring *et al.*, 2016) at 30-arcsecond resolution (~1 x 1 km) for time intervals 2041–2060 and 2081–2100 (approximately one and two generations from present, respectively) under two

climate change scenarios: Shared Socioeconomic Pathway (SSP) 1-2.6 and SSP5-8.5 (www.worldclim.org/data/cmip6/cmip6_clim30s). All available global climate models (GCMs) for each time interval/SSP combination at the sampling locations were extracted (25 total available). Not every GCM was available at each of our sampling locations and desired time interval / SSP combinations. As such, we used GCMs that matched across SSPs. Within SSP1-2.6, a total of eight GCMs were used: ACCESS.ESM1.5, CNRM.CM6.1, CNRM.CM6.1.HR, FIO.ESM.2.0, GFDL.ESM4, INM.CM4.8, INM.CM5.0, MIROC.ES2L. Within SSP1-2.6, a total of six GCMs were used: ACCESS.ESM1.5, FIO.ESM.2.0, GISS.E2.1.G, GISS.E2.1.H, MIROC6, MPI.ESM1.2.HR. Within each year interval, SSP, and chosen bioclimatic variable (bio1, bio3, bio4, bio7; see Materials and Methods of main text for selection process), we assessed correlations and percent change across GCMs and averaged across GCMs for the point estimate for each sampling location and future climate projection. The correlations across GCMs combinations were high (mean $r = 0.976$, range = 0.907 – 0.999) and percent change was low (mean = 2.782%, range = 1.381 – 5.175). Averages and descriptions for all future bioclimatic variables at each site is available online.

Supporting Information Table S1 DIC values for each of 5 replicate `entropy` runs for *k* demes. Genotype probabilities from the best fit model (the fifth replicate of the *k* = 3 model, in **bold**) were used for subsequent analyses.

k	1	2	3	4	5	mean	SD
2	4,694,069	4,730,125	4,691,057	4,612,379	4,704,762	4,686,478	44,183
3	4,573,061	4,569,639	4,525,149	4,577,898	**4,521,519**	4,553,453	27,681
4	5,202,771	5,307,073	5,359,470	5,316,896	5,270,698	5,291,382	58,764
5	5,378,464	5,395,386	5,023,277	5,226,875	5,206,145	5,246,029	151,181
6	11,050,473	15,687,610	14,395,517	15,386,044	11,128,048	13,529,539	2,278,526
7	6,865,803	6,922,160	6,839,397	7,327,504	7,303,602	7,051,693	242,865
8	23,748,442	18,486,485	67,706,204	12,030,852	19,406,859	28,275,768	22,436,805
9	2,719,179	45,738,006	11,256,110	11,256,110	13,778,710	16,949,623	16,629,517
10	68,820,992	152,029,331	13,460,452	21,236,365	21,942,261	55,497,880	58,229,723
11	11,000,307	11,000,307	16,821,156	12,876,326	11,909,958	12,721,610	2,419,440
12	12,873,377	12,873,377	15,950,014	10,585,768	12,022,045	12,860,916	1,963,596

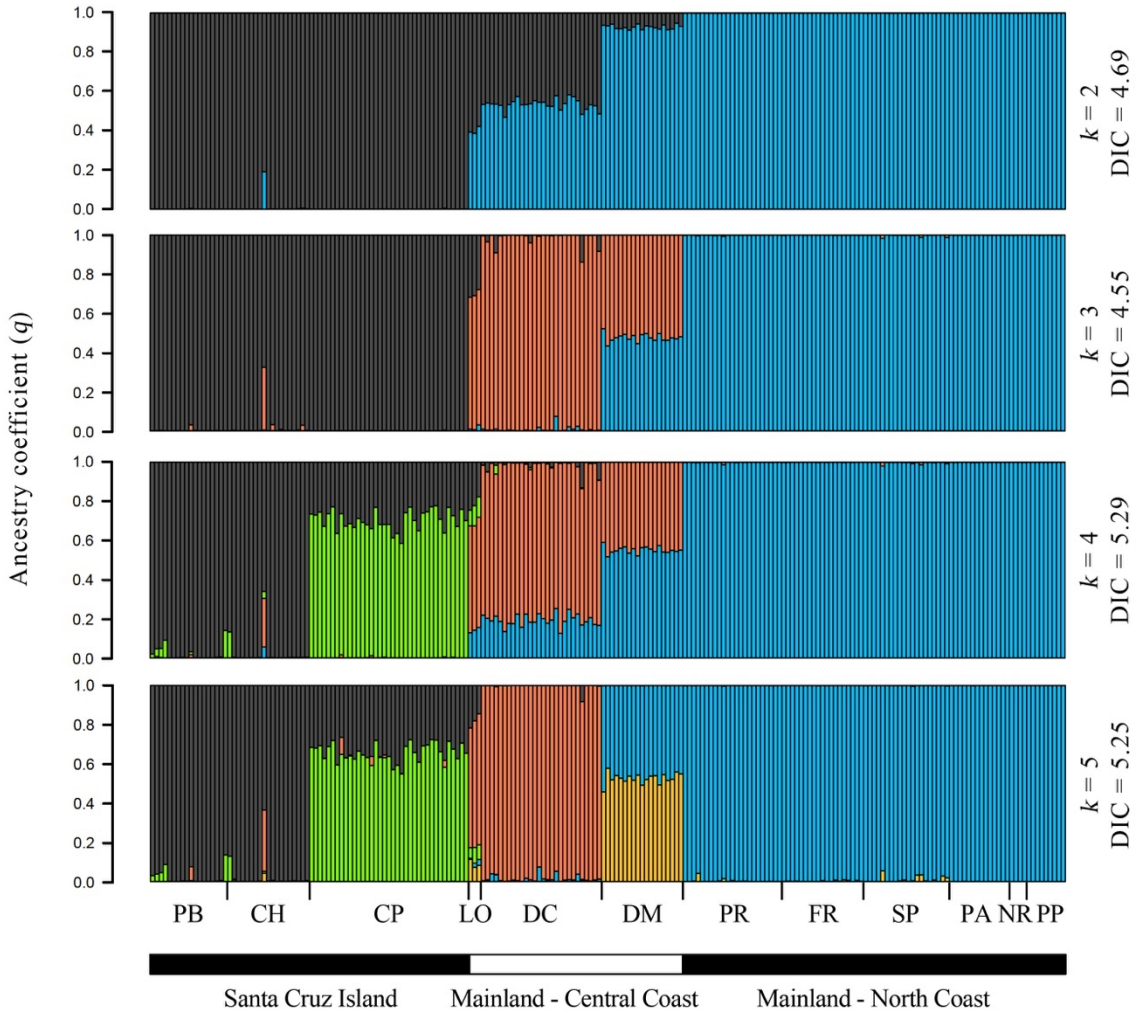
Supporting Information Table S2. Genetic diversity based on 213 individuals from 12 sampling sites, where site abbreviations correspond to those in Table 1 and Figure 1A. Estimates based on allele frequencies (H_E) and those based on DNA sequence variation (θ_π, θ_W, and the scaled difference between these two, or Tajima's D) are included. High and low columns represent 95% confidence intervals.

Site	H_E	θ_π	θ_π low	θ_π high	θ_W	θ_W low	θ_W high	D	D low	D high
PP	0.144	0.00616	0.00615	0.00618	0.00556	0.00555	0.00557	-0.02510	-0.02512	-0.02505
NR	0.141	0.00727	0.00726	0.00729	0.00684	0.00682	0.00685	-0.01680	-0.01685	-0.01679
PA	0.145	0.00719	0.00718	0.00721	0.00723	0.00722	0.00724	-0.05670	-0.05678	-0.05670
SP	0.150	0.00729	0.00728	0.00731	0.00727	0.00726	0.00728	-0.06120	-0.06123	-0.06115
FR	0.149	0.00698	0.00697	0.00700	0.00672	0.00671	0.00673	-0.05360	-0.05368	-0.05360
PR	0.148	0.00689	0.00687	0.00690	0.00675	0.00674	0.00675	-0.05820	-0.05826	-0.05817
DM	0.171	0.00674	0.00673	0.00676	0.00623	0.00622	0.00624	-0.04700	-0.04704	-0.04695
DC	0.167	0.00831	0.00829	0.00832	0.00838	0.00837	0.00839	-0.07150	-0.07153	-0.07144
LO	0.164	0.00788	0.00786	0.00790	0.00744	0.00742	0.00745	-0.01520	-0.01523	-0.01516
CP	0.210	0.00629	0.00628	0.00630	0.00563	0.00563	0.00564	-0.04960	-0.04961	-0.04951
PB	0.203	0.00617	0.00615	0.00618	0.00539	0.00538	0.00540	-0.03700	-0.03700	-0.03692
CH	0.200	0.00625	0.00624	0.00627	0.00565	0.00564	0.00566	-0.04470	-0.04477	-0.04468

Supporting Information Table S3. Principal component axis (PCA) loadings of soil variables for axes 1-5 with parenthetic *PVE* for each axis (cuml. *PVE* = 93.89%). Full description of variables, units, and raw values available online. Extracted from www.soilgrids.org.

Soil variable	Description (depth in cm)	PC1 (36.1%)	PC2 (33.9%)	PC3 (11.8%)	PC4 (6.2%)	PC5 (5.9%)
ocs	Organic carbon stock (0-30)	-0.0186	-0.4241	-0.2577	-0.1335	0.0116
ocd shallow	Organic carbon density (0-30)	0.1794	-0.3715	0.2586	-0.0860	0.2175
ocd deep	Organic carbon density (60-200)	-0.0064	-0.3915	-0.1864	0.3568	0.1509
bdod	Bulk density (0-200)	-0.1667	0.3462	0.1320	0.5355	0.0538
clay	Proportion of clay (0-200)	0.0735	0.2395	-0.5709	-0.2339	-0.0188
cfvo	% coarse fragments (0-200)	0.3482	0.2239	-0.0682	0.0905	-0.1553
sand	Proportion of sand (0-200)	-0.4029	-0.1102	0.0206	-0.1213	0.1492
silt	Proportion of silt (0-200)	0.3881	-0.0046	0.2652	0.2452	-0.1506
cec	Cation exchange capacity (0-200)	0.3771	0.1685	-0.2040	-0.1675	0.1040
nitrogen shallow	Soil nitrogen (0-30)	0.4020	0.0398	-0.1453	-0.1830	0.0483
nitrogen deep	Soil nitrogen (60-200)	0.0956	-0.2493	-0.3870	0.5039	-0.4519
soc shallow	Soil organic carbon (0-15)	0.3718	0.0083	0.3493	0.0534	0.1559
soc deep	Soil organic carbon (15-200)	0.2169	-0.3215	-0.1276	0.1278	0.4584
phh2o	Soil pH (0-200)	-0.0435	0.3063	-0.2501	0.2973	0.6369

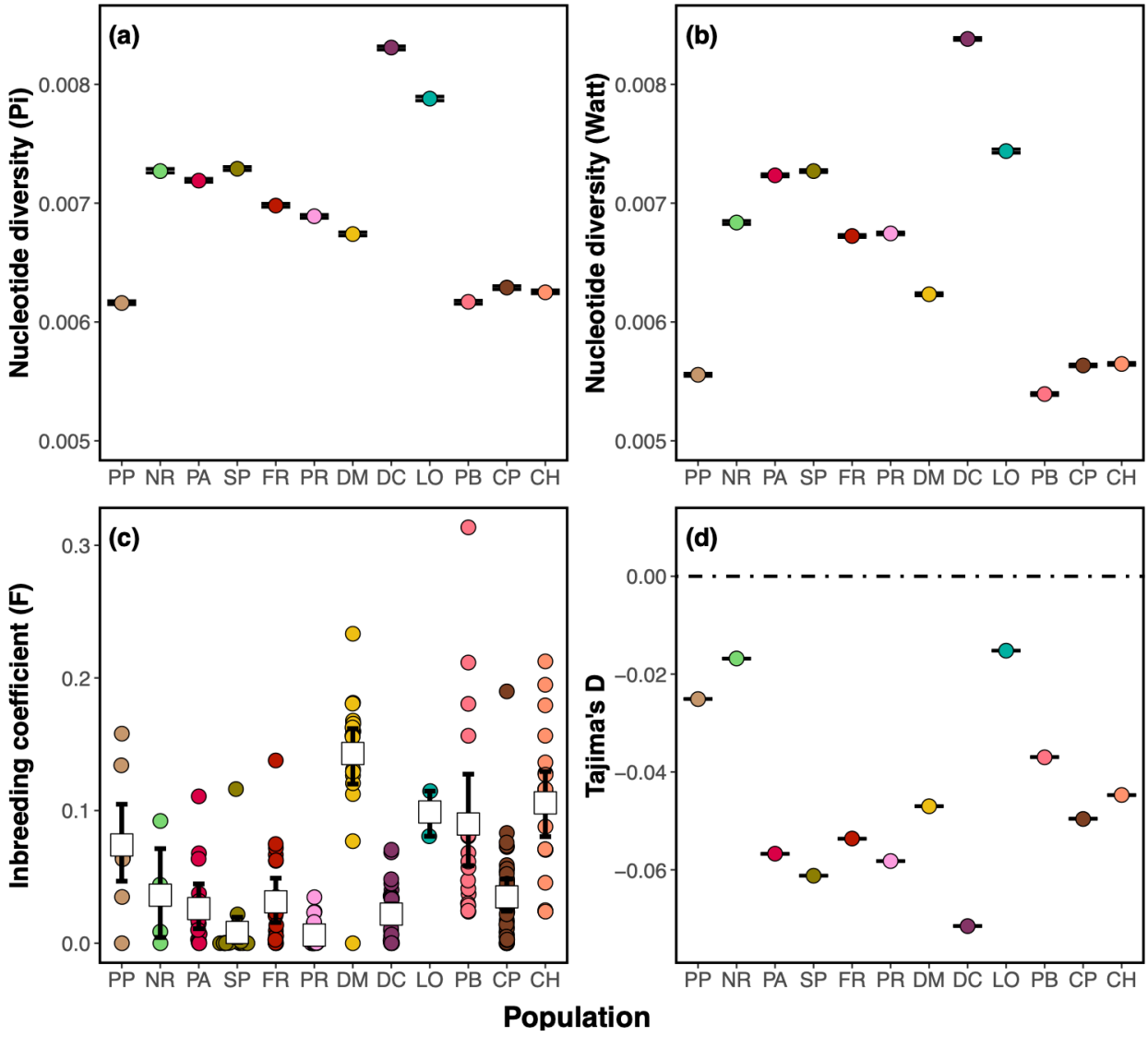
Supporting Information Fig. S1. Estimates of ancestry coefficients (q) for $k = 2 – 5$ models, where each vertical bar represents an individual and colors correspond to the proportion of ancestry for k demes. The $k = 3$ model fit the data best, yielding the lowest deviance information criterion (DIC). Full DIC scores for all replicate models are supplied in Supplementary Table 1. Here, the lowest DIC for each model is provided to the right of each plot, in millions. Population abbreviations correspond to those in Table 1.



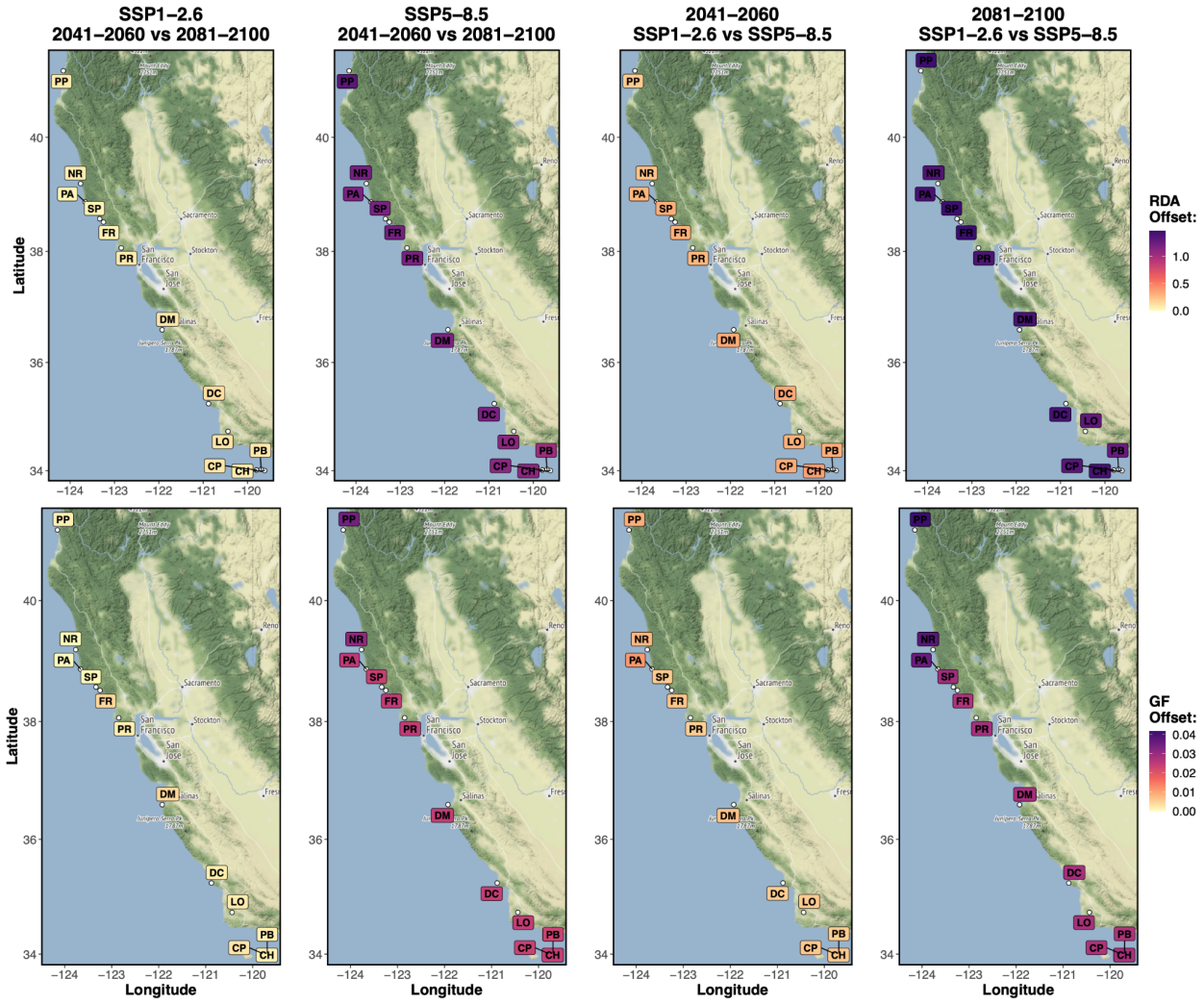
Supporting Information Fig. S2 Pairwise estimates of F_{ST} (upper diagonal; Hudson *et al.*, 1992) and Nei's *D* (lower diagonal; Nei, 1972), where populations are listed latitudinally and population abbreviations correspond to those in Table 1.

	PP	NR	PA	SP	FR	PR	DM	DC	LO	PB	CP	CH
PP	0	0.033	0.02	0.021	0.022	0.023	0.053	0.066	0.097	0.111	0.11	0.108
NR	0.01	0	0.027	0.029	0.029	0.03	0.059	0.071	0.1	0.117	0.116	0.113
PA	0.006	0.008	0	0.014	0.015	0.016	0.046	0.058	0.091	0.105	0.104	0.1
SP	0.006	0.009	0.004	0	0.011	0.012	0.04	0.054	0.089	0.102	0.101	0.098
FR	0.007	0.009	0.004	0.003	0	0.013	0.042	0.056	0.09	0.104	0.103	0.099
PR	0.007	0.009	0.004	0.003	0.004	0	0.041	0.055	0.089	0.103	0.101	0.098
DM	0.019	0.021	0.016	0.014	0.014	0.014	0	0.044	0.081	0.093	0.094	0.089
DC	0.025	0.027	0.021	0.02	0.021	0.02	0.015	0	0.056	0.064	0.063	0.06
LO	0.04	0.041	0.037	0.036	0.036	0.036	0.032	0.019	0	0.084	0.081	0.079
PB	0.048	0.05	0.044	0.043	0.044	0.043	0.039	0.023	0.032	0	0.043	0.031
CP	0.047	0.049	0.044	0.042	0.043	0.043	0.04	0.023	0.031	0.014	0	0.042
CH	0.046	0.048	0.042	0.041	0.042	0.041	0.037	0.022	0.03	0.009	0.014	0

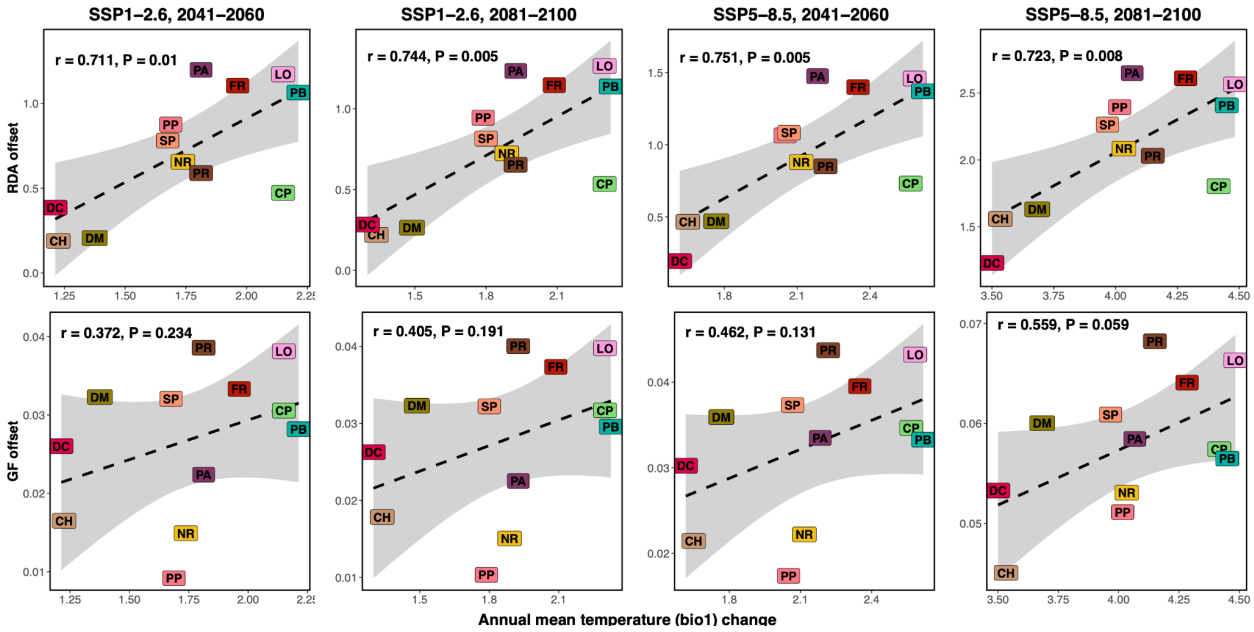
Supporting Information Fig. S3 Estimates of nucleotide diversity, inbreeding (*F*), and Tajima's *D* for each population, ordered latitudinally. (a) and (b) population mean and standard deviation of nucleotide diversity, Θ_π and Θ_W, respectively. (c) Individual inbreeding coefficients (*F*) at each sampling site, showing relatively low inbreeding overall (mean = 0.0499) but also highlighting variation across the landscape. Colored points represent the raw data shaped by the distribution, and squares and error bars represent the population mean and a 95% bootstrapped confidence intervals. (d) Population mean and standard deviation of Tajima's *D* were consistently negative across all sampling sites, suggesting population expansion. Colors correspond to those in Fig. 1a,b,c.



Supporting Information Fig. S4. GEA offset was calculated for each population as the difference between two future scenarios using both RDA (top row) and GF (bottom row). The changes in environmental associations between SSP1-2.6 and SSP5-8.5 at year interval 2041–2060 and between year intervals 2081–2100 and 2041–2060 at SSP1-2.6 were minimal. Dramatic shifts in associations were predicted under the SSP5-8.5 scenario from year intervals 2041–2060 and 2081–2100 and between SSP1-2.6 and SSP5-8.5 at year interval 2081–2100.



Supporting Information Fig. S5 Associations between the GEA offsets using both RDA (top row) and GF (bottom row) and the estimated increase from present mean annual temperature (bio1, °C) across the four time interval / SSP combinations. A generalizable pattern emerged, illustrating that GEA offset increases with projected increases in mean annual temperature and is consistent between the two different approaches.



References

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE. **2016**. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9**: 1937–1958.

Hudson RR, Slatkin M, Maddison WP. **1992**. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Nei M. **1972**. Genetic distance between populations. *American Naturalist* **106**: 192–283.

Poggio L, de Sousa LM, Batjes NH, Heuvelink G, Kempen B, Ribeiro E, Rossiter D. **2021**. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**: 217–240.