UNIVERSITY OF CALIFORNIA
RIVERSIDE

AND

SAN DIEGO STATE UNIVERSITY

Evolutionary Past, Present, and Future of the Yosemite Toad (*Anaxyrus canorus*): A Total Evidence Approach to Delineating Conservation Units

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Evolutionary Biology

by

Paul Andrew Maier

September 2018

Dissertation Committee:
Dr. Andrew J. Bohonak, Co-Chairperson
Dr. Leonard Nunney, Co-Chairperson
Dr. Amy G. Vandergast
Dr. David A. Reznick

The Dissertation of Paul Andrew Maier is approved:

_____

_____

_____

Committee Co-Chairperson

_____

Committee Co-Chairperson

University of California, Riverside
San Diego State University

ACKNOWLEDGEMENTS

I am eternally grateful to so many people, without whom this dissertation would have been just another idea in the wind.

First and foremost, I must acknowledge the extraordinary help I received from the U.S. Geological Survey (USGS), both financially and institutionally. My employment at USGS lasted from before my Ph.D. work until nearly the end (2010–2017), and was the platform for conceiving Chapters 1 and 3. The vast majority of financial support for those projects derived from USGS Ecosystems Mission Area, Natural Resource Preservation Program Research (NRPP) funding.

I would also like to acknowledge my other funding sources: the Harold & June Memorial, Jordan D. Covin, and ARCS scholarships. These foundations generously provided support for much of Chapters 2 and 4.

I was lucky to spend five summers (2010–2014) at the Sierra Nevada Research Institute (SNRI) in Yosemite National Park, the launchpad for fieldwork involving thousands of miles hiked. Similarly, the Sierra Nevada Aquatic Research Laboratory (SNARL) generously provided facilities for my field team in 2016. I am very appreciative to both institutions for helping to manage the chaos of fieldwork, and providing the opportunity for clean socks every so often.

My research on Yosemite toads has now spanned 12 years (2006–2018), and was sparked by a phone interview with a young U.S. Forest Service researcher, who invited me to assist a study that involved meadows, cows, and making toads glow in the dark. I

can finally admit now that I was way out of my depth. I owe the beginning of my career to Robert Grasso, who gave the opportunity of a lifetime to a naïve sophomore.

In a similar vein, I owe so much of my career to Steven Ostoja, Ph.D., and professional facilitator. In the days of my early adulthood, there was a serious dilemma about whether to pursue a biology career, or become a rock-and-roll drummer. Steve simplified the choice vastly with a tongue-and-cheek ultimatum involving a "rusty Buick." He has been a long-time mentor to me, and opened up innumerable professional doors, for which I will always be thankful.

A veritable army of field biologists literally hiked thousands of miles to help me collect my field specimens, and morphological data. I would like to express my warm gratitude to: Michael Hernandez, Johanne Boulat, Alexa Killion, Mara MacKinnon, Alexa Lindauer, Ross Maynard, Nathalie Aall, Corrina Kamoroff, Steven Lee, Amy Patton, Heidi Rodgers, Cameron DeMaranville, Keevan Harding, Brenna Blessing, Megan Sidran, Julia Meyer, and Willa Gruver. Very few other people on the planet would count 11,000 tadpoles in exchange for freeze-dried spaghetti.

All animal handling was performed in accordance with SDSU animal care and use protocols #13-03-001B and #16-04-004B. All data and sample collection was only completed after obtaining and adhering to the regulations of NPS research permits.

A special thank you to Jared Grummer, who provided technical advice that improved library preparation for sequencing, and to Andrew Bohonak, Amy Vandergast, Jonathan Q. Richmond, Brian J. Halstead, Leonard Nunney, Andres Aguilar, Steven Ostoja, David Weisrock, Peter Tiffin, and two anonymous reviewers, who provided

comments and suggestions that significantly improved the writing of various chapters. Amy in particular has gone above and beyond to give critical feedback and mentorship for all aspects of science whenever possible.

It is impossible to quantify the guidance and mentorship that my faculty adviser Andrew Bohonak has provided me; however if I were to quantify them, he would request not only that I calculate their significant difference from 0, but also their 95% CIs, and broader biological meaning. Andy always taught me to manage expectations; I hope that his have been met. The back-and-forth discussions about theory and applications of population genetics have been an absolute pleasure for me over the years.

I would like to express my deep gratitude to my parents, whose love and support have always been a constant in my life. My dad turned me into a professional mountaineer before I turned six, and taught me the value of science and the natural world. My mom instilled in me a sense of hard-working stubbornness and belief in myself that has always endured. Thank you to you both, and I love you.

Lastly, to my partner in life, who has been everything from emotional confidante, to professional toad-wrangler, to pro-bono figure designer: I love you Rachel, and cannot wait to turn the page to our next chapter in life.

DEDICATION

I would like to dedicate this work to my grandfather, Howard Theodore Maier. May we all strive to touch gently a butterfly, and renew our connection with nature.

ABSTRACT OF THE DISSERTATION

Evolutionary Past, Present, and Future of the Yosemite Toad (*Anaxyrus canorus*):
A Total Evidence Approach to Delineating Conservation Units

by

Paul Andrew Maier

Doctor of Philosophy, Graduate Program in Evolutionary Biology
University of California, Riverside and San Diego State University, September 2018
Dr. Andrew J. Bohonak and Dr. Leonard Nunney, Co-Chairpersons

Climate change is ostensibly one of the greatest modern selective pressures, and species with sensitive life histories or physiologies must adapt, migrate, or buffer its effects to persist. Some 15–37% of species are expected to be endangered or extinct by 2050. The most vulnerable include habitat specialists, local endemics, and species with low intrinsic growth rates. Yosemite toads (*Anaxyrus canorus*) are one such alpine endemic, having been extirpated from up to 69% of their historical range. Several features of their natural history make them vulnerable: small population sizes, high larval mortality, infrequent breeding, and specialized, patch-limited habitat prone to premature desiccation. In addition to their role as ecosystem flagships, Yosemite toads provide a model system for the many other specialists with similar life histories that are challenged by environmental change. The goal of this dissertation is to understand how historical evolutionary processes such as lineage divergence and secondary admixture, along with

current levels of genetic connectivity, are expected to shape the future of Yosemite toad persistence in the face of climate change. The first chapter reconstructs phylogeographic patterns of lineage formation and fusion during repeated bouts of Pleistocene glaciation, and showcases a role for refugia in ecological divergence. The second chapter examines three contact zones as replicate tests of the hypothesis that loci associated with incipient speciation are distinct from those that readily cross ancient lineage boundaries. The third chapter models modern genetic connectivity as a network of environmental and climatic interactions, using a novel approach that incorporates phylogeographic structure. The fourth chapter forecasts the future selective pressure of climate change, and predicts where connectivity may be a mitigating force to restore genetic diversity. My dissertation provides an example of how conservation strategies can incorporate the many temporal processes (ancient, recent, and current) that have shaped current genetic diversity patterns, and use a "total evidence" approach to predict future adaptive potential.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

LIST OF SUPPLEMENTARY TABLES

LIST OF SUPPLEMENTARY FIGURES

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

PREFACE

Shortly before I completed this work, we lost the late great Ernest Karlstrom (1928–2018). Ernie was an early student of Robert Stebbins at U.C. Berkeley, whose indelible impact on California herpetology is still felt today. Ernie's work on Yosemite toads embodied the passion of his lab for natural history and evolution. For those unfamiliar, the Yosemite toad is a shining beacon for curious minds, a squishy denizen of a rugged montane wilderness, and an evolutionary relict of an earlier time when glaciers scoured California. Ernie was arguably the first to shed substantial light on this beautiful, sonorous oddity, although his work would not have been possible without that of Joseph Grinnell, and his two students Charles Camp and Tracy Storer, between 1916 and 1925. After finishing his dissertation in 1956, Ernie published his research in his 1962 book entitled "The toad genus *Bufo* in the Sierra Nevada of California." This deceptively simple title belies some of the best and earliest natural history observations of the Yosemite toad's growth, development, physiological relations with the environment, and phenology. However he went much further, and delved into its bizarre adaptations for montane life. He carefully observed its evolved mating trill, size, and pigmentation (Yosemite toads are by far the most sexually dichromatic anurans in North America). He quantified how its close relative the boreal toad exhibits convergent adaptive traits in higher latitude regions such as Montana, Canada, and Alaska. He catalogued areas of sympatry between the two species, but noticed how they rarely co-occur. And he came up

with some tantalizing hypotheses for how speciation in Sierra Nevada toads has progressed, ascribing similar processes in the local frogs and salamanders.

Perhaps most importantly, Ernie opened up a window for numerous other graduate students, allowing them to journey into this amphibian's world, and quench their own scientific curiosity. By my count, 14 students have completed their theses or dissertations on topics of Yosemite toad ecology or evolution: Wood 1977 (M.S.), Sherman 1980 (Ph.D.), Hollis 1997 (M.A.), Stephens 2001 (M.S.), Vredenburg 2002 (Ph.D.), Grasso 2005 (M.S.), Martin 2008 (Ph.D.), Pauly 2008 (Ph.D.), Nelson 2008 (M.S.), Liang 2010 (Ph.D.), Roche 2011 (Ph.D.), Dodge 2012 (M.S.), Lindauer 2018 (M.S.), as well as this author. Although I was very saddened to hear of his passing, Ernie had such a positive influence on so many people that his life should be celebrated. He spent much of the rest of his life teaching (37 years, mostly at University of Puget Sound), studying recolonization of Mt. Saint Helens, being a loving husband to his wife Marilyn (also a teacher and naturalist), and staying engaged with natural science organizations until the end. Ernie had an immeasurable impact on many people, including a strong influence on my own dissertation; he will be greatly missed.

INTRODUCTION OF THE DISSERTATION

Anthropogenic climate change is rapidly reorganizing our natural world. Some 39% of species studied have shifted their distributions away from new physiological stressors, at a rate of approximately 6 km per decade (Parmesan and Yohe 2003; Root et al. 2003). As bicarbonate is forced into the oceans, innumerable mollusks, corals, and coccolithophores are literally dissolving, with almost immeasurable impact on fisheries, food web interactions, and productivity (Hays et al. 2005; Hoegh-Guldberg 2010; Doney et al. 2012). Plant flowering dates are shifting in unpredictable directions (CaraDonna and Inouye 2015), plant-pollinator interactions are becoming decoupled (Forrest 2015), and frogs are breeding earlier, usually to the detriment of population viability (Blaustein et al. 2001; Todd et al. 2011; Sheridan et al. 2018). Approximately 15–37% of species are expected to be on a path toward extinction by 2050, with large and unpredictable consequences for community composition and genetic structure (Thomas et al. 2004; Thuiller 2004; Jay et al. 2012). These most vulnerable of species are not a random assortment—patch-limited, endemic, or otherwise specialized species are more vulnerable, due to their narrow tolerance for change (Warren et al. 2001; Thuiller et al. 2005; IUCN 2018). Taxonomically, amphibians appear to be particularly vulnerable to phenological and range shifts, because their physiology is intimately tied to climate (Alexander and Eischeid 2001; Pounds 2001; Pounds et al. 2006). For example, a staggering 67% of species in the montane-adapted toad genus *Atelopus* are critically endangered, and have virtually disappeared from climate-induced disease outbreak (La

Marca et al. 2005; Pounds et al. 2006; IUCN 2018). Therefore, it is critical that we understand why, how, and where such specialists are vanishing before it is too late. Species may still persist if their natural adaptive response is left unencumbered.

Yosemite toads of the Sierra Nevada Mountains (California) are flagships for other vulnerable specialists. They are federally threatened, because they are believed to be declining in both distribution and abundance (Sherman and Morton 1993; Drost and Fellers 1994; Jennings and Hayes 1994; Brown and Olsen 2013; US Fish & Wildlife Service 2014; Brown et al. 2015; Thompson et al. 2016; but see Ostoja et al., in prep.). Despite numerous efforts to find causative agents for the decline of Yosemite toads, such as UV radiation (Sadinski 2004), exotic predators (Grasso et al. 2010), meadow grazing (Roche et al. 2012a,b; Matchett et al. 2015), chemical deposition (Bradford and Gordon 1994; Davidson 2004; Sadinski 2004), and chytridiomycosis (Dodge and Vredenberg 2012; Lindauer 2018), no single cause has emerged (Brown et al. 2015). Despite this uncertainty about the past, climate change has been identified as the leading future threat to Yosemite toad persistence (US Fish & Wildlife Service 2014). This is because adult toads specifically choose shallow ($\bar{x}$ <5 cm) meadow pools for breeding, and tadpoles behaviorally seek out the shallowest margins of these transient waterbodies (Mullally 1953; Karlstrom 1962; Liang et al. 2017). Significant larval mortality is often observed even during years with above-average snowpack (Sherman 1980; Sherman and Morton 1993; Brown et al. 2015). Adults and post-metamorphic juveniles are also sensitive to climate, because higher winter temperature and lower snowpack incur greater energetic costs of overwintering (Morton 1981). In the future, climate change is expected to

dramatically change the hydrology of Sierra Nevada meadows. If carbon emissions continue unabated, by the end of the 21st century there will be a 7°C rise in average springtime temperature, 64% drop in springtime snowpack volume, and 50 day phenological shift to earlier runoff of snowmelt (Reich et al. 2018). This effect will not be uniform however, and some meadows may become climate refugia for toads (Smith and Tirpak 1988; Viers et al. 2013; Reich et al. 2018). Therefore, there is a need to understand where Yosemite toads are most vulnerable, and whether they have the potential to respond to this selective pressure.

Conservation genetics is in the enterprise of answering such questions about future population persistence, by using indirect (genetic) methods (Frankham 1995). Many species are refractory to direct ecological study because they are difficult to find, capture, or track over time. For example, Yosemite toad adults are rarely found outside their short breeding season, and the majority of meadows they inhabit are >10 miles from any road. Thus, conservation genetics can fill the void that population viability assessments might otherwise fill (Boyce 1992), by estimating average evolutionary and demographic parameters (DeSalle and Amato 2004). However, there are some fundamental flaws to the way conservation genetic assessments are often carried out. One common mistake practitioners make is to conflate genetic patterns—such as the diagnosability of two sister lineages—with the importance of the process they represent, for instance divergent adaptation (Paetkau 1999). This is easy to remedy, by simply distinguishing the conservation outcome (e.g. protecting adaptive diversity) from the operational criterion used to find significant patterns in the data (de Queiroz 1998). A

much more serious error is the conflation of past adaptive processes (e.g. adaptation to Pleistocene climatic niches) with anticipated relevance to the future (e.g. future adaptation to climate change). As outlined above, modern climate change is starkly unpredictable, and it will require unprecedented (10,000× greater than usual) rates of niche evolution (Quintero and Wiens 2013). For conservation units to be meaningful, they should explicitly predict future evolutionary potential (Bowen 1998, 1999).

There are at least four distinct evolutionary processes likely to affect future adaptation and persistence in a vulnerable species, such as the Yosemite toad. (1) A species may consist of multiple intraspecific lineages that became isolated in the ancient past, perhaps allowing them to accumulate divergent adaptations. For example, Rissler and Apodaca (2007) combined traditional phylogeographic approaches with ecological niche modeling, and found that numerous cryptic lineages of black salamander (*Aneides flavipunctatus*) had diverged by climatic niche in California. (2) Two lineages within a species may have secondarily come back into contact, possibly promoting hybrid vigor or adaptive introgression. For example, Fitzpatrick and Shaffer (2007) found that hybrids between native and introduced tiger salamanders in California had higher larval survival than either species due to both overdominant and recombinant hybrid vigor. (3) Present-day populations are structured based on the unit of demographic independence, and are important sources of genetic variation. A sufficiently large number of migrants among populations ($N_e m$) can have the beneficial effect of replenishing any lost variation over time. For example, Blouin et al. (2010) estimated low levels of gene flow between hierarchically structured populations of Oregon spotted frogs (*Rana pretiosa*), and

attributed low N_e to poor dispersal habitat. (4) Future lineage evolution—particularly for specialist amphibians such as Yosemite toads—will depend on which populations can adapt and persist. Adaptation to simple directional selection depends on the strength of selection (i.e. how much will climate shift), and levels of additive genetic variation for the trait(s) being selected. Thus, all of the above processes (ancient adaptation, recent admixture, and current neutral diversity) may be informative about future outcomes. Future adaptation to climate change may then be forecasted using adaptive genetic variation, and anticipated selection pressure. For example, Fitzpatrick and Keller (2015) forecasted where in the U.S. climate change would most strongly shift frequencies of a circadian clock gene, which is under selection by temperature in balsam poplars (*Populus balsamifera*). Unlike trees, Yosemite toads and other specialist amphibians may have weak response to selection due to smaller population sizes (Willi et al. 2006). Therefore, it is critical for climate adaptation models to integrate the total evidence for both selection and heritability of climate traits (Hoffmann and Sgrò 2011; Franks and Hoffmann 2012; Merilä and Hendry 2014).

The overarching goal of my dissertation is to study these four temporal processes in Yosemite toad evolution. In Chapter 1, I test a "glacial pulse" model of ancient lineage diversification in which cycles of allopatry and ecologically divergent glacial refugia play a role in generating biodiversity. In Chapter 2, I capitalize on three contact zones of admixture found in Chapter 1, and test the hypothesis that adaptive divergence and adaptive introgression act on fundamentally different genetic loci. In Chapter 3, I delineate population boundaries, characterize source-sink structure, and introduce a novel

network method to predict the environments that facilitate toad meadow connectivity. The final chapter attempts to integrate the total evidence for Yosemite toad evolution into a predictive map of future evolutionary potential. In this dissertation, I introduce the first two genome-wide datasets (ddRADseq and RNAseq) for the species, for which very limited genetic resources previously existed. This progression of studies is a model for thinking about lineage evolution across the temporal spectrum, and for anticipating conservation outcomes in vulnerable species that are poorly understood.

REFERENCES

Alexander, M. A., and J. K. Eischeid. 2001. Climate variability in regions of amphibian declines. Conserv. Biol. 15:930–942.

Blaustein, A. R., L. K. Belden, D. H. Olson, D. M. Green, T. L. Root, and J. M. Kiesecker. 2001. Amphibian breeding and climate change. Conserv. Biol. 15:1804–1809.

Blouin, M. S., I. C. Phillipsen, and K. J. Monsen. 2010. Population structure and conservation genetics of the Oregon spotted frog, *Rana pretiosa*. Conserv. Genet. 11:2179–2194.

Bowen, B. W. 1999. Preserving genes, species, or ecosystems? healing the fractured foundations of conservation policy. Mol. Ecol. 8:S5–S10.

Bowen, B. W. 1998. What is wrong with ESUs?: the gap between evolutionary theory and conservation principles. J. Shellfish Res. 17:1355–1358.

Boyce, M. S. 1992. Population viability analysis. Annu. Rev. Ecol. Syst. 23:481–506.

Bradford, D., and M. Gordon. 1994. Acidic deposition as an unlikely cause for amphibian population declines in the Sierra Nevada, California. Biol. Conserv. 69:155–161.

Brown, C., M. P. Hayes, G. A. Green, D. C. Macfarlane, and A. J. Lind. 2015. Yosemite toad conservation assessment. USDA Forest Service report. Sonora, CA.

Brown, C., and A. R. Olsen. 2013. Bioregional monitoring design and occupancy estimation for two Sierra Nevadan amphibian taxa. Freshw. Sci. 32:675–691.

CaraDonna, P. J., and D. W. Inouye. 2015. Phenological responses to climate change do not exhibit phylogenetic signal in a subalpine plant community. Ecology 96:355–361.

Davidson, C. 2004. Declining downwind: amphibian population declines in California and historical pesticide use. Ecol. Appl. 14:1892–1902.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation. Pp. 57–75 *in* D. J. Howard and S. H. Berlocher, eds. Endless Forms: Species and Speciation. Oxford University Press, Oxford, UK.

DeSalle, R., and G. Amato. 2004. The expansion of conservation genetics. Nat. Rev. Genet. 5:702–712.

Dodge, C. M., and V. T. Vredenburg. 2012. The sad song of the Yosemite toad: the role of the amphibian chytrid fungus in an enigmatic decline. Paper presented at the annual convention of the Ecological Society of America, Portland, OR.

Doney, S. C., M. Ruckelshaus, J. Emmett Duffy, J. P. Barry, F. Chan, C. A. English, H. M. Galindo, J. M. Grebmeier, A. B. Hollowed, N. Knowlton, J. Polovina, N. N. Rabalais, W. J. Sydeman, and L. D. Talley. 2012. Climate change impacts on marine ecosystems. Ann. Rev. Mar. Sci. 4:11–37.

Drost, C., and G. Fellers. 1994. Decline of frog species in the Yosemite section of the Sierra Nevada. National Park Service report. Davis, CA.

Fitzpatrick, B. M., and H. B. Shaffer. 2007. Hybrid vigor between native and introduced salamanders raises new challenges for conservation. Proc. Natl. Acad. Sci. 104:15793–15798.

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

Forrest, J. R. K. 2015. Plant-pollinator interactions and phenological change: What can we learn about climate impacts from experiments and observations? Oikos 124:4–13.

Frankham, R. 1995. Conservation genetics. Annu. Rev. Genet. 29:305–327.

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

Grasso, R. L., R. M. Coleman, and C. Davidson. 2010. Palatability and antipredator response of Yosemite toads (*Anaxyrus canorus*) to nonnative brook trout (*Salvelinus fontinalis*) in the Sierra Nevada Mountains of California. Copeia 2010:457–462.

Hays, G. C., A. J. Richardson, and C. Robinson. 2005. Climate change and marine plankton. Trends Ecol. Evol. 20:337–344.

Hoegh-Guldberg, O. 2010. The impact of climate change on the world's marine ecosystems. Ecol. Res. 328:1523–1529.

Hoffmann, A. A., and C. M. Sgrò. 2011. Climate change and evolutionary adaptation. Nature 470:479–485.

IUCN. 2018. The IUCN Red List of Threatened Species. Version 2018-1. Available online at: http://www.iucnredlist.org/

Jay, F., S. Manel, N. Alvarez, E. Y. Durand, W. Thuiller, R. Holderegger, P. Taberlet, and O. François. 2012. Forecasting changes in population genetic structure of alpine plants in response to global warming. Mol. Ecol. 21:2354–2368.

Jennings, M., and M. Hayes. 1994. Amphibian and reptile species of special concern in California. California Department of Fish & Game report. Rancho Cordova, CA.

Karlstrom, E. L. 1962. The toad genus *Bufo* in the Sierra Nevada of California. Unviersity Calif. Publ. Zool. 62:1–104.

La Marca, E., K. R. Lips, S. Lötters, R. Puschendorf, R. Ibáñez, J. V. Rueda-Almonacid, R. Schulte, C. Marty, F. Castro, J. Manzanilla-Puppo, J. E. Garcia-Pérez, F. Bolaños, G. Chaves, J. A. Pounds, E. Toral, and B. E. Young. 2005. Catastrophic population declines and extinctions in neotropical harlequin frogs (Bufonidae: *Atelopus*). Biotropica 37:190–201.

Liang, C. T., R. L. Grasso, J. J. Nelson-Paul, K. E. Vincent, and A. J. Lind. 2017. Fine-Scale habitat characteristics related to occupancy of the Yosemite toad, *Anaxyrus canorus*. Copeia 105:120–127.

Lindauer, A. 2018. The role of hydroperiod and fluctuating temperature on disease dynamics: a disease ecology approach to understanding Yosemite toad declines. Reno, NV: University of Nevada Reno. M.S. thesis.

Matchett, J. R., P. B. Stark, S. M. Ostoja, R. A. Knapp, H. C. McKenny, M. L. Brooks, W. T. Langford, L. N. Joppa, and E. L. Berlow. 2015. Detecting the influence of rare stressors on rare species in Yosemite National Park using a novel stratified permutation test. Sci. Rep. 5:1–12.

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

Morton, M. 1981. Seasonal changes in total body lipid and liver weight in the Yosemite toad. Copeia 1981:234–238.

Mullally, D. 1953. Observations on the ecology of the toad *Bufo canorus*. Copeia 1953:182–183.

Ostoja, S. M., S. R. Lee, J. R. Matchett, H. McKenny, M. L. Brooks, R. A. Knapp, P. A. Maier, and N. Danielle. 2018. Yosemite toad (*Anaxyrus* [*Bufo*] *canorus*) breeding occupancy of meadows in Yosemite and Kings Canyon National Parks, California. Unpublished manuscript.

Paetkau, D. 1999. Using genetics to identify intraspecific conservation units: a critique of current methods. Conserv. Biol. 13:1507–1509.

Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42.

Pounds, J. 2001. Climate and amphibian declines. Nature 410:639–640.

Pounds, J. A., M. R. Bustamante, L. A. Coloma, J. A. Consuegra, M. P. L. Fogden, P. N. Foster, E. La Marca, K. L. Masters, A. Merino-Viteri, R. Puschendorf, S. R. Ron, G. A. Sánchez-Azofeifa, C. J. Still, and B. E. Young. 2006. Widespread amphibian extinctions from epidemic disease driven by global warming. Nature 439:161–167.

Quintero, I., and J. J. Wiens. 2013. Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species. Ecol. Lett. 16:1095–1103.

Reich, K. D., N. Berg, D. B. Walton, M. Schwartz, F. Sun, X. Huang, and A. Hall. 2018. Climate change in the Sierra Nevada: California's water future. UCLA Center for Climate Science report. Los Angeles, CA.

Rissler, L., and J. Apodaca. 2007. Adding More Ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). Syst. Biol. 56:924–942.

Roche, L. M., B. Allen-Diaz, D. J. Eastburn, and K. W. Tate. 2012a. Cattle grazing and Yosemite toad (*Bufo canorus*, Camp) breeding habitat in Sierra Nevada meadows. Rangel. Ecol. Manag. 65:56–65.

Roche, L. M., A. M. Latimer, D. J. Eastburn, and K. W. Tate. 2012b. Cattle grazing and conservation of a meadow-dependent amphibian species in the Sierra Nevada. PLoS One 7:e35734.

Root, T., J. Price, and K. Hall. 2003. Fingerprints of global warming on wild animals and plants. Nature 421:57–60.

Sadinski, W. 2004. Amphibian declines: causes. U.S. Geological Survey report. La Crosse, WI.

Sheridan, J. A., N. M. Caruso, J. J. Apodaca, and L. J. Rissler. 2018. Shifts in frog size and phenology: testing predictions of climate change on a widespread anuran using data from prior to rapid climate warming. Ecol. Evol. 8:1316–1327.

Sherman, C. K. 1980. A comparison of the natural history and mating system of two anurans: Yosemite toads (*Bufo canorus*) and black toads (*Bufo exsul*). Ann Arbor, MI: University of Michigan. Ph.D. dissertation.

Sherman, C., and M. Morton. 1993. Population declines of Yosemite toads in the eastern Sierra Nevada of California. J. Herpetol. 27:186–198.

Smith, J. B., and D. A. Tirpak. 1988. The potential effects of global climate change on the United States. U.S. Environmental Protection Agency, draft report to Congress.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk from climate change. Nature 427:145–148.

Thompson, B., A. Wright, and H. B. Shaffer. 2016. Yosemite toad (*Bufo canorus*). Pp. 112–123 *in* California amphibian and reptile species of special concern.

Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. Glob. Chang. Biol. 10:2020–2027.

Thuiller, W., S. Lavorel, and M. B. Araújo. 2005. Niche properties and geographical extent as predictors of species sensitivity to climate change. Glob. Ecol. Biogeogr. 14:347–357.

Todd, B. D., D. E. Scott, J. H. K. Pechmann, and J. Whitfield Gibbons. 2011. Climate change correlates with rapid delays and advancements in reproductive timing in an amphibian community. Proc. R. Soc. B Biol. Sci. 278:2191–2197.

US Fish & Wildlife Service. 2014. Endangered and threatened wildlife and plants; endangered status for the Sierra Nevada yellow-legged frog and the northern distinct population segment of the mountain yellow-legged frog, and threatened status for the Yosemite toad: final rule. Fed. Regist. 1–56.

Viers, J. H., S. E. Purdy, R. A. Peek, A. Fryjoff-Hung, N. R. Santos, J. V. Katz, J. D. Emmons, D. V Dolan, and S. M. Yarnell. 2013. Montane meadows in the Sierra Nevada: changing hydroclimatic conditions and concepts for vulnerability assessment. Centre for Watershed Sciences report. Davis, CA.

Warren, M. S., J. K. Hill, J. A. Thomas, J. Asher, R. Fox, B. Huntley, D. Roy, M. Telfer, S. Jeffcoate, P. Harding, G. Jeffcoate, S. Willis, J. Greatorex-Savies, D. Moss, and C. Thomas. 2001. Rapid responses of British butterflies to opposing forces of climate and habitat change. Nature 414:65–69.

Willi, Y., J. Van Buskirk, and A. A. Hoffmann. 2006. Limits to the adaptive potential of small populations. Annu. Rev. Ecol. Evol. Syst. 37:433–458.

CHAPTER 1: PLEISTOCENE GLACIAL CYCLES DROVE LINEAGE DIVERSIFICATION AND FUSION

ABSTRACT

Species endemic to alpine environments often evolve via steep ecological selection gradients between lowland and upland environments. However, many alpine environments have faced repeated glacial episodes over the past 2 million years, fracturing these endemics into isolated populations. I introduce a "glacial pulse" model of alpine speciation, in which cycles of allopatry and ecologically divergent glacial refugia play a role in generating biodiversity, including novel admixed ("fused") lineages. I tested for patterns of glacial pulse lineage diversification in the Yosemite toad (*Anaxyrus* [*Bufo*] *canorus*), an alpine endemic tied to glacially influenced meadow environments. Using double-digest RADseq on populations densely sampled from a large portion of the species range, I identified nine distinct lineages with divergence times ranging from 214–732 ka, coinciding with multiple Sierra Nevada glacial events. Three lineages have admixed origins, and demographic models suggest these admixed lineages have persisted throughout past glacial cycles. Multiple estimates of genealogy shape supported the hypothesis that some lineages recolonized Yosemite from east of the ice sheet, whereas other lineages remained in western refugia. Finally, I found evidence that low- and high-elevation lineages have convergently adapted to similar climatic niches. My results suggest glacial cycles and secondary contact zones may be important crucibles of adaptive diversity across deep evolutionary time.

INTRODUCTION

A complex history of geological and climatic changes has dynamically altered California's landscape for millions of years, and driven sharp genetic discontinuities in plants and animals. Vicariance by mountain uplift and climatic gradients has long been thought to drive speciation (Fisher 1930; Mayr 1942), and California's juxtaposition of rising mountains, shifting plate tectonics, active volcanism, and resulting sharp ecotones has produced unparalleled levels of biodiversity and endemism (Myers et al. 2000; Lapointe and Rissler 2005). The Sierra Nevada mountains as a whole are a particularly imposing barrier to dispersal, and have repeatedly driven vicariant evolution during the late tertiary and early quaternary periods (Calsbeek et al. 2003; Swenson and Howard 2005). Over 73% of amphibian and squamate species that transect the Sierra Nevada exhibit lineage divergence in this ecoregion, although the biogeographic boundaries often do not coincide (Rissler et al. 2006). Endemism of lineages in California is especially high in patches of the western Sierra Nevada foothills and along the eastern crest, likely a consequence of glacial advance and retreat (Swenson and Howard 2005; Rissler et al. 2006). Although major uplift of the Sierra Nevada by westward tilting probably began in the early Pliocene ca. 5 Ma (Unruh 1991), the earliest known glaciation (McGee) is thought to have occurred only 2.5–1.5 Ma during the early to mid Pleistocene, after the Sierras had attained sufficient height (Huber 1981; Gillespie and Clark 2011). The onset of Sierra Nevada glaciation matches well with major divergence times for many resident species (Calsbeek et al. 2003). Subsequent glacial cycles during the Pleistocene have

played an important role in shaping biodiversity patterns in the region, by repeatedly isolating and reuniting populations (Avise et al. 1998; Hewitt 2004).

Alpine endemics are prone to this evolutionary process because their narrow distributions and specialized habitat requirements are closely linked to the process of glaciation. For example, endemic Sierra Nevada species of butterfly (Schoville and Roderick 2009), chipmunk (Rubidge et al. 2014), and amphibian (Rovito 2010; Schoville et al. 2011) have recently recolonized glacially-deposited lake, meadow, or talus habitats, thus minimizing their niche overlap with closely related lowland species. Many of these alpine endemics originally arose via allopatric speciation during tectonic uplift or ecological speciation along an elevational cline (Chabot and Billings 1972; Billings 1974). However, repeated glacial episodes can fracture these endemics further by restricting them to different isolated refugia, and the resulting lineages may be adapted to non-overlapping climatic regimes. The linear nature of glacial barriers and rain shadow effect provide ample potential for climatic differentiation, since refugia can occur on both sides of mountain ranges (Mulch et al. 2008). Thus, alpine endemics may be prone to repeated glacial "pulses" of isolation, differential adaptation, and secondary contact (Hewitt 1996). I suggest that alpine speciation and lineage divergence is best understood under this model, whereby pulses of glacial action bisect lineages into divergent refugial ecologies, which can then recursively subdivide further during subsequent glacial cycles (Fig. 1.1). This general framework explains ubiquitous patterns of species and lineage endemism in alpine zones, and provides specific testable hypotheses where secondary contact zones emerge.

Secondary contact zones in alpine systems across the world tend to share concordant patterns but different locations among species, suggesting common glacial barriers may affect entire ecosystems, but species-specific traits such as dispersal ability may play a role (Wallis et al. 2016). What is less clear is whether these zones typically maintain lineage integrity, either by lowered or augmented hybrid fitness, or merge lineages back together by unimpeded gene flow (Fig. 1.1). The outcome depends on the extent of reproductive isolation, ecological divergence, and whether admixed populations survive into the next cycle. Patchworks of multiple contact zones that reticulate cyclically across time are not uncommon, further confusing inference about population history (Hewitt 1996, 2011). Following secondary contact, one possible outcome is that divergent adaptation among lineages will produce hybrids with intrinsic genetic incompatibilities via epistasis, regardless of hybrid environment. Such a "tension zone" tends to be restricted in space, owing to a balance between migration and selection (Harrison 1993; Harrison and Larson 2014). For example, admixture between two closely related species of newt in the Sierra Nevada produced a tension zone, as evidenced by high inbreeding, elevated linkage disequilibrium, and heterozygote deficit (Kuchta 2007). Another outcome is for admixture to occur along an ecotone where hybrid fitness surpasses parental fitness, but only in that specific environmental context. This is known as a "bounded hybrid superiority zone," and is especially common along alpine gradients, although is probably widespread (Abbott and Brennan 2014). For example, two closely related species of *Lycaeides* butterfly reach secondary contact at lower montane elevations in the Sierra Nevada, where a reproductively isolated hybrid lineage has

developed alpine adaptations, such as high host specificity for alpine-endemic plants, a loss of egg adhesion, and novel color patterns (Gompert et al. 2006).

Even when early-generation hybrids have reduced fertility or reproductive success, subsequent generations can still form a stable lineage (Arnold et al. 1999). Bounded hybrid superiority zones might be common for alpine endemics, because the temporal scale of glacial cycles is short enough that lineages may have accumulated some adaptive genetic differences without having become intrinsically incompatible. This process of "lineage fusion" essentially amounts to homoploid hybrid speciation (except that lineages are fused instead of species), if the hybrid population maintains its distinctness due to novel combinations of alleles (Abbott et al. 2013). Only scattered instances in animals have been described (e.g. Mavárez and Linares 2008; reviewed in Schumer et al. 2014), possibly because contact zones are usually examined between more distantly related lineages, and the burden of proof is high. Hence, lineage fusion is one hypothesized source of lineage diversity under the glacial pulse model, in contrast to simple divergence, and unimpeded gene flow (Fig. 1.1). Regardless of whether fused lineages persist, introgression of alleles between recently diverged lineages can be an important source of adaptive novelty (Hedrick 2013).

In this study, I tested for patterns of glacial pulse lineage formation in the Yosemite toad (*Anaxyrus canorus*), a Sierra Nevada endemic that is closely associated with Pleistocene glacial cycles. It breeds almost exclusively in the shallow, transient, and highly productive water bodies of mountain meadows (Ratliff 1985), which make up less than 3% of the landscape. However, currently existing meadows only formed ca. 10 ka,

as receding glaciers and rising snowpack levels helped retain more alluvium in stream valleys (Wood 1975). Hence, location of suitable habitat has likely shifted during glacial maxima. In addition, the species utilizes hydrologically distinct meadow types along an elevational gradient, from montane to subalpine (Ratliff 1982). Karlstrom (1962) hypothesized that the most recent common ancestor of Yosemite toads and Western toads (*A. boreas*) colonized the young Sierra Nevada during mountain uplift in the Pliocene, where these progenitors of *A. canorus* became adapted to boreal conditions prior to the major upheaval of these mountains. He speculated that repeated glacial pulses of isolation during the Pleistocene promoted alpine specialization and eventually speciation for *A. canorus*, which could then competitively exclude *A. boreas* at higher elevations (Fig. 1.1). If glacial pulses explain the origin of the species and continued generation of novel lineages, (1) species and lineage divergence times should correspond to dates of Pleistocene glacial oscillations, (2) most sister lineages should occupy divergent ecological niches that relate to hypothesized refugial locations, and (3) size of secondary contact zones should depend on niche distinctness of the two lineages; narrow contact zones should form between ecologically distinct lineages, and wider contact zones should form between lineages with more similar ecologies. Lineage fusion may occur anywhere that secondary admixture occurs, if hybrid genotypes have a local advantage and remain distinct from parental lineages.

Major phylogeographic discontinuities have been suggested within and between Yosemite National Park (YOSE) and Kings Canyon National Park (KICA), but previously sparse sampling makes it difficult to infer the locations, divergence depths,

and presence of contact zones for any discontinuities (Shaffer and Fellers 2000; Stephens 2001; Goebel and Ranker 2009). In addition, the maternal inheritance of mitochondrial DNA can misrepresent lineage history, especially because it is effectively one non-recombining locus that could be subject to selection (Galtier et al. 2009). This may explain previous patterns of mitochondrial paraphyly or polyphyly of *A. canorus* with respect to *A. boreas* (Graybeal 1993; Shaffer and Fellers 2000; Stephens 2001; Pauly 2008; Goebel and Ranker 2009). Alternatively, *A. canorus* may represent multiple lineages that arose in parallel via an ecological-elevational gradient (Stephens 2001). Although both scenarios could be attributed to glacial pulses, the timing of species origin is important to accurately reconstruct, since a pre-Pleistocene origin would not be consistent with Karlstrom's (1962) hypothesis of glacial origins. Inference based on multiple nuclear markers is needed to distinguish these competing hypotheses.

My goals were to: (1) delineate the phylogeographic boundaries of past lineage isolation within YOSE and KICA, (2) detect locations of contact zones between lineages to test hypotheses of age, symmetry, and viability of inter-lineage admixture, (3) reconstruct historical demography and niche occupancy of lineages in YOSE, to test the hypothesis that refugial niches reinforced lineage divergence across this heterogeneous landscape. I used a highly robust spatial sampling scheme to reduce bias associated with unsampled locations, by utilizing a systematic six-year census of Yosemite toad occupancy at the meadow scale (Ostoja et al., in prep.). I also employed a double-digest RADseq genomic sampling scheme to provide thousands of genome-wide nuclear markers that can accurately reconstruct species and population history. The unique

analytical approach in this study leveraged both SNP and haplotype information from RAD loci, and made inference based on coalescent, site frequency spectrum, and inter-lineage heterozygosity information, as well as ecological models. Given that Yosemite toads are federally threatened and particularly susceptible to climatic fluctuations (US Fish & Wildlife Service 2014), it is critical to understand the history of climatic evolution in this species in order to anticipate how different lineages might respond to future climate change. Finally, the glacial pulse model provides a novel framework for shedding light on patterns of alpine speciation and endemism worldwide, by showing that endemic species and their young lineages may have arisen by a common, cyclical process.

MATERIALS AND METHODS

Study region

Yosemite toads (*Anaxryus canorus*) are meadow-breeding specialists restricted to the central Sierra Nevada of California, and found between 1,950–3,444 m (Mullally and Cunningham 1956; Karlstrom 1962). The species breeds almost exclusively in shallow, snowmelt ponds within meadows, which seasonally dry up. Lower elevation sites are typically spring-fed mesic or hydric meadows characterized by adjacent stands of montane red fir (*Abies magnifica*) and lodgepole pine (*Pinus contorta var. murrayana*), while subalpine and alpine meadows are typically larger, snow-fed, more xeric, and surrounded by whitebark pine (*Pinus albicaulis*) or boulders (Keeler-Wolf et al. 2012; Viers et al. 2013) (Fig. 1.2). Yosemite National Park (YOSE; 3,027 km^2) and northern Kings Canyon National Park (KICA; 467 km^2) were chosen as the primary study areas

because they overlap with previous studies (Shaffer and Fellers 2000; Wang 2012; Berlow et al. 2013), likely contain genetic discontinuities, and offer a representative snapshot of ecological conditions experienced by the species (dense sampling across 37% of species-wide localities).

Tissue sampling and site selection

Tail tissue was collected from larval toads using a sterilized razor blade during summers of 2011–2013. All samples were preserved in 95% EtOH and stored at -20°C within one week. An existing USGS meadow layer based on a park-wide vegetation map with 0.5 ha resolution was used to delimit sampling sites (Keeler-Wolf et al. 2012; Berlow et al. 2013). Sites were chosen to maximize representation across all known breeding locations from a recent six-year survey effort (Ostoja et al., in prep) and overlap with previous studies. Tadpoles were sampled across multiple clutches, pools, and years to maximize inclusion of available genetic diversity. Toe clips from eight *Anaxyrus boreas* samples spanning California (including samples from YOSE) and a single *A. punctatus* sample (next closest relative) were obtained as outgroups for phylogenetic analysis from the Museum of Vertebrate Zoology (University of California, Berkeley) and USGS (Western Ecological Research Station).

Molecular methods

A total of 653 samples were chosen for sequencing (535 samples from YOSE, 109 from KICA, and nine outgroup samples) (Appendix A). A minimum of five samples

was used per meadow if additional meadows were sequenced within 1 km, and 10 per meadow for remaining meadows, unless insufficient samples were available. This scheme maximized intra- and inter-meadow sampling representation across the study area. Genomic DNA was extracted using a combination of 96-well glass fiber plate (Ivanova et al. 2006) and DNeasy blood and tissue spin column (Qiagen) protocols. I constructed double-digest RADseq libraries following the protocol of Peterson et al. (2012; S1 Protocol). Details are described in Appendix A.

Bioinformatic data processing

Raw data were filtered and processed using Stacks v1.19 (Catchen et al. 2011, 2013). A detailed bioinformatic pipeline is described in Appendix A. Briefly, two datasets were generated from a total of 3261 loci: (1) SNPs (one SNP kept per locus), and (2) full-locus haplotypes (concatenated paired-end reads were used where possible, otherwise Read 1 was used). Some analyses required SNP data, whereas others could accommodate the full diversity present among multiple SNPs at a locus due to recombination. Therefore, I built and implemented a custom python script (fasta2genotype.py, github.com/paulmaier/fasta2genotype) to output RAD haplotypes in addition to SNPs. After applying a locus genotype coverage threshold of 10, loci were removed from meadows if absent from >25% individuals, and removed from the dataset if absent from >25% individuals overall. Only alleles with a minor allele frequency (MAF) of 0.05 or greater were kept. All processing was performed on a high-

performance biocluster at the Institute for Integrative Genome Biology, UC Riverside, CA.

Population genetic analyses at the meadow scale

All markers in the haplotype dataset were initially tested for random mating genotype (Hardy-Weinberg) frequencies at the meadow scale using the adegenet package (Jombart 2008) in R v3.3.3 (R Core Team 2018). NeEstimator v2.01 was used to estimate N_e values using the single-sample linkage disequilibrium method (Do et al. 2014). Population genetic summary statistics were computed using populations.pl in Stacks, along with Weir and Cockerham's pairwise F_{ST} estimates (Weir 1996), which correct for bias associated with different sample sizes. This estimate was iterated across loci and then averaged for each pair of populations.

Phylogenetic analysis and divergence dating

A spatial principal components analysis (sPCA; Jombart et al. 2008) was performed first, to reveal the extent and location of cryptic phylogeographic discontinuities in YOSE. This method works by creating orthogonal synthetic variables that optimize the product of genetic variance and spatial autocorrelation (measured by Moran's I). I used the haplotype YOSE dataset (2318 loci; 6473 haplotypes), and a spatial connection network of the k=50 nearest neighbors for each sample location. Phylogenetic structure among meadows was estimated by creating a concatenated alignment of loci, and using the GTR + Γ nucleotide model in RAxML v8.2.9

(Stamatakis 2014) and BEAST v2.4.6 (Drummond and Rambaut 2007). This evolutionary model was chosen using jModelTest v2.1.4, (Posada 2008) by selecting the model with lowest AIC score and significant likelihood ratio test. I chose to use meadows (not representative individuals) as operational taxon units (OTUs), because treating polymorphic SNPs as fixed can bias branch lengths and divergence time estimates (Lischer et al. 2014). Therefore, any SNPs that were polymorphic within OTUs were summarized by IUPAC ambiguity codes.

For the RAxML analysis, I concatenated one SNP per locus and applied the Stamatakis method of ascertainment bias correction to account for the known number of invariable sites, and base frequencies. This ensured that bootstrap replicates sampled variable sites and provided reasonable estimates of node confidence. *A. punctatus* was used as the formal outgroup, and all eight samples of *A. boreas* were included. A heuristic search was performed with RAxML for 10 independent reps using TBR branch swapping and 1000 bootstrap replicates for each. The tree with highest likelihood was compared with the combined bootstrap tree to check for differences in topology. In order to further assess the reciprocal monophyly of *A. boreas* and *A. canorus*, I repeated this analysis with a data matrix that balanced representation of *A. boreas* (n=8) and *A. canorus* (n=12; two individuals per major lineage).

Divergence dates were estimated using BEAST v2.4.6 on an alignment of full sequences from all *A. canorus* individuals in the initial RAxML analysis. I used a coalescent Bayesian skyline tree model, because that model had a higher harmonic mean marginal likelihood than either Yule or constant population models during preliminary

runs. I used a strict clock, and calibrated the clockRate parameter with a normal prior bounded by the range of amphibian nuclear DNA rates ($9.24 \times 10^{-10} – 1.53 \times 10^{-9}$ substitutions site^{-1} year^{-1}) described by Crawford (2003). Preliminary runs using a relaxed lognormal clock failed to converge, and ucld.stdev values were close to zero, consistent with a clock-like pattern of evolution across lineages. Divergence times with and without a strict clock were nearly identical. I chose the option to use IUPAC ambiguity code information in tree likelihood calculations. The MCMC chain was run for 5×10^7 generations with 10^7 burnin, logging every 10^3, and checked for convergence across three independent runs using Tracer 1.6.

Several geographic regions appeared to have recent inter-lineage admixture (Miller Lake in northern YOSE, Isberg Pass in southern YOSE) or admixture from outside the study area (Lyell Canyon in southern YOSE), based on sPCA scores that were intermediate between lineages, and conflicting signals of ancestry. Conflicting ancestry was detected during preliminary BEAST runs by visualizing the posterior sample of trees with DensiTree v2.4.6. OTUs were considered to have conflicting ancestry if they had phylogenetic signal from two distinct lineages, and if the OTUs spatially resided at the boundary between those lineages. Such OTUs were removed from all phylogenetic analyses because they violate the assumption of dichotomous branching.

Detecting secondary admixture between lineages

In YOSE, an eastern lineage appeared to have secondary contact with three other lineages (north, west, and south; see Results). However, as described above, recent inter-

lineage admixture was only detected in northern and southern YOSE. Therefore, I chose to focus on the pattern and process of secondary admixture in these two contact zones. The largest inter-lineage contact zone was found in northern YOSE, in contrast to a much smaller contact zone in southern YOSE (see Results). For each contact zone, ancestry from parental lineages was estimated using STRUCTURE v2.3.4 (Pritchard et al. 2000). Using the haplotype dataset (n=2318 loci), I ran STRUCTURE for 10 replicates each at K=2 for individuals within the two lineages, using 100,000 generations and 10,000 burn-in. Results were combined using CLUMPP v1.1.2 and visualized with distruct v1.1. NewHybrids v1.0 (Anderson and Thompson 2002) was run for 500,000 reps and 100,00 burn-in to estimate the posterior probability of individuals falling into six discrete genotype classes of hybrid origin (P1, P2, F1, F2, P1-F1 backcross, and P2-F1 backcross). I then estimated admixture index (S; the proportion of ancestry from one of the two parental populations) and inter-lineage heterozygosity (H_I; the observed heterozygosity for markers with one allele from each parental population). H_I can either be estimated as particular genotype class (e.g. F1), or as a continuous value from 0–1, where H_I=1 is the maximal inter-lineage heterozygosity (i.e. F1 class). This latter type of analysis is more realistic in situations where hybrids are more than two generations old. Both types of analyses require approximately diagnostic markers, i.e. those that approach fixation within the two parental lineages. I chose diagnostic markers by subsampling the genomic data to those within the top 1% of markers contributing to spatial sPC1 (larger contact zone) and sPC2 (smaller contact zone). I opted to choose diagnostic markers using sPCA loadings and not F_{ST}; the former approach is model-free, and the first two

sPCs corresponded perfectly to the north and south contact zones respectively, whereas F_{ST} can fail to detect diagnostic markers if they have high intra-lineage diversity.

I used the R package HIest (Fitzpatrick 2012) to estimate admixture index (S) and inter-lineage heterozygosity (H_I), i.e. whether admixed individuals were admixed beyond two filial generations (i.e. F3, F4, F5, etc.). I ran each individual for 500 iterations, with a starting grid of 20. I contrasted these empirical results with simulated values of S and H_I, assuming a population of size 50 and no immigration (assuming a closed system for hybridization), with 25 diagnostic markers. I used H_I as one possible indication of admixed population fitness, since any reproductive incompatibilities would likely curtail the extent of admixture to one or a few generations. The function dapc in the adegenet package was used to perform a discriminant analysis of principal components (DAPC), to determine whether the admixed meadows formed distinct genetic clusters from either parental lineage. I compared DAPC analyses with meadows vs. lineages as the grouping factor, and retained the first 100/535 principal components.

Demographic reconstruction of admixture zones

Results from the foregoing analyses suggested that three distinct admixed lineages exist within the two contact zones examined (see Results). Such admixed lineages may originate by several processes. For example, lineage fusion may result if hybridization occurred briefly, and produced a genetically distinct lineage that persists through time. Another possibility is that hybrids have lowered fitness, but ongoing introgression toward the contact zone replenishes a distinct admixed lineage. These possibilities assume that

admixture follows secondary contact; if lineages have never been fully isolated, then primary contact (i.e. isolation with migration) may explain the pattern. Finally, the null hypothesis is that admixture has not happened, and the lineage in question may reflect simple divergence. I tested demographic models of lineage origins using fastsimcoal v2.6 (Excoffier et al. 2013). Fastsimcoal can leverage the observed minor allele site frequency spectrum (SFS) to calculate composite likelihoods of complex models by simulating an expected SFS with coalescent simulations. This method is particularly suitable for large SNP datasets. For each contact zone (northern and southern), I constructed plausible models of simple isolation, isolation with migration (primary contact), hybridization (secondary contact, instantaneous origin), and introgression (secondary contact, ongoing directional gene flow) (see Results for model details).

Multidimensional folded SFSs were generated for fastsimcoal analysis using a custom python script (github.com/isaacovercast/easySFS). For each contact zone, SFSs were downsampled to equal sample sizes among lineages, while also maximizing the number of segregating sites to increase statistical power. All parameters were given uniform priors except for time parameters, which were given log-uniform priors. Each model was run for 10^5 coalescent simulations with 40 expectation-maximization (ECM) cycles, and replicated for 100 independent runs, to determine parameter values leading to the maximum likelihood. Likelihood ratios or AIC statistics are imprecisely estimated using composite likelihoods; hence I used the qpcR package (Ritz and Spiess 2008) in R to calculate Akaike's weight of evidence, based on the highest overall likelihoods across 100 independent estimations of each model.

Confidence intervals for parameters were estimated by generating 100 non-parametric bootstrap datasets. Pseudoreplicate datasets were made by sampling SNPs with replacement using the vcfR (Knaus and Grünwald 2017) package in R. Then, each model was rerun 100 times, using the maximum likelihood model parameters as starting values. Parameter estimates of highest likelihood from each pseudoreplicate run were used to construct 95% confidence intervals.

Tests for demographic declines or range expansions following glaciation

Genealogy "shape" was estimated in several ways to determine whether a simple Wright-Fisher model with constant population size could be rejected in favor of a model with recent population size changes. First, Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) were calculated using the SNP dataset in Arlequin v3.5.2 (Excoffier and Lischer 2010). I also tested the null hypothesis of recent spatial expansion under an infinite island model, where T generations ago a small population of size N began to expand and send migrants to neighboring demes (Excoffier 2004). Sum of squared deviations (SSD) between observed and expected distributions, as well as Harpending's raggedness index, were tested for significance using 10^3 bootstrap replicates.

Extended Bayesian skyline plots (EBSPs) were constructed separately for each lineage using BEAST v2.4.6, to test the hypothesis of population expansion in a coalescent framework. I created Nexus files for 50 randomly selected RAD loci, to increase computational efficiency, using a custom R script (github.com/rcristofari/RAD-Scripts/blob/master/RAD_Haplotypes.R). Low coverage reads can underrepresent

heterozygotes, so I randomly selected one haplotype per individual. I chose a "scattered" sampling strategy of one individual per meadow in each region (N=8–17 individuals), to reduce bias associated with population structure (Heller et al. 2013). Concatenated (197 bp) loci were used, and locus selection was restricted to those with 5–6 SNPs. I found this range of SNPs to be a good balance between: (1) containing sufficient coalescent signal, (2) being prevalent enough in the dataset for my sample size (N=50 loci), and (3) only requiring two site models to be specified. To reduce the number of parameters, I used an HKY + Γ model with empirical base frequencies, and a single strict clock (calibrated as before). The MCMC chain was run for 10^8 generations with 2×10^7 burnin, logging every 10^4, and checked for convergence across three independent runs using Tracer v1.6.

I further tested the hypothesis that a spatial range expansion had taken place in YOSE following glacial recession, using the method of Peter and Slatkin (2013) on the SNP dataset. This method infers a directionality index (ψ), based on pairwise allele frequency asymmetries that result from drift fixing lower frequency alleles during an expansion. Based on observed ψ values, the origin is inferred using time difference of arrival methods (TDOA). A key assumption of the method is even sample sizes, so five individuals were randomly chosen from every meadow. A block-jackknife approach was used to generate a null distribution (isolation by distance) to test significance of the model. Only pure lineages were considered because admixed sample sizes were insufficient.

Phylogeographic reconstruction of glacial refugia and climatic niche divergence in Yosemite NP

Hypothesized Pleistocene refugia were reconstructed using Maxent v3.4 (Phillips et al. 2004). Environmental niche models can provide a more spatially explicit and less subjective hypothesis of refugial distribution than coalescent approaches (Waltari et al. 2007). The 19 layers of bioclimatic variables were obtained from WorldClim v1.4 corresponding to present day (1960–1990) and the last glacial maximum (LGM; ~22 ka) (Hijmans et al. 2005). I used the Model for Interdisciplinary Research on Climate Earth System Model (MIROC-ESM), at a resolution of 2.5 arc-minutes. Sampling points for all non-admixed lineages were derived from occurrence data from 1905–present. The number of points used was 222–1,046, depending on lineage, with 25% set aside as test points. Present-day environmental niche models were constructed for each lineage separately using all 19 variables, removing those with permutation importance values of zero, and then rerunning the model. The final number of variables was 5–7 (listed by lineage in Table S1.6), reducing multicollinearity. Each model was projected using the corresponding LGM dataset.

Climatic niche overlap of present-day lineages within YOSE was assessed using two methods. First, a linear discriminant analysis (LDA) on lineages was performed for all lineages (pure/admixed) using the six bioclimatic variables with highest loadings, with the MASS package (Venables and Ripley 2002) in R. Second, Schoener's D measuring niche overlap was calculated for each pair of non-admixed lineages, and a niche

equivalency test was performed (Broennimann et al. 2012). All GIS extraction and manipulation was automated in R.

RESULTS

Phylogenetic analysis and divergence dating

Phylogenetic analyses using RAxML and BEAST recovered a topology supporting four major lineages within YOSE ("Y-North", "Y-East", "Y-South", and "Y-West"), as well as one admixed lineage ("East-North-A1", discussed below). The other two admixed lineages ("East-North-A2" and "East-South-A") were excluded from phylogenetic analysis due to recent admixed ancestry. As mentioned earlier (see Methods), recent admixed ancestry was detected by intermediate sPCA scores, and conflicting signals of ancestry shown by DensiTree. Y-East was paraphyletic with respect to Y-West in the RAxML (but not the BEAST) tree (Figs. 1.3, S1.1, S1.2, S1.3), which likely reflects the recentness of that divergence. Sister to YOSE, the KICA clade contained two lineages ("Goddard" and "Evolution"). A general pattern of increasing genetic diversity from low to high elevation was found (see Appendix A). The oldest divergence was between YOSE and KICA, and was estimated to be 1.99±0.12 Ma (Fig. 1.3, Table S1.3). The second oldest divergence between Y-North and the remainder of YOSE was estimated at 732±45 ka. Only the youngest lineage (Evolution) had a crown date (18±9 ka) corresponding to the most recent glaciation (Tioga; advance 21–20 ka, retreat 15–14 ka). Most divergences predated the Tioga. Additionally, all eight *Anaxyrus boreas* samples from both northern and southern California (including from YOSE)

formed one monophyletic lineage, sister to all *A. canorus* samples. *A. canorus* and *A. boreas* were reciprocally monophyletic regardless of taxon sampling and ascertainment used (Fig. S1.3). The estimated clockRate was 1.507×10^{-9} substitutions site^{-1} year^{-1}, which I used in subsequent fastsimcoal and EBSP analyses. Highest ranked sPCA scores reflected three major barriers: Tuolumne and Merced river gorges, and an east-west elevational cline (white areas of rapid sPC score turnover; Fig. 1.4). Phylogenetic and ordination results revealed identical genetic discontinuities (Figs. 1.3, 1.4), suggesting river gorges (previously filled by glaciers) and elevation have structured phylogeny.

Detecting secondary admixture between lineages

The deep divergence between Y-North and the rest of YOSE showed an allelic, genotypic, and diversity gradient suggestive of isolation followed by recent contact, and genetic admixture between ancient lineages (Fig. 1.5). Several lines of evidence suggested this. First, the estimated divergence time of these lineages (732 ka, 37% the estimated age of the YOSE-KICA divergence, which is 1.99 Ma; Figs. 1.3, S1.2) is very old. Second, although northern YOSE is filled with many potentially isolating river canyons (Fig. 1.2), inter-lineage pairwise F_{ST} values were significantly higher than intra-lineage values when crossing between Y-East and Y-North, where no such topographic barriers seem to exist ($F=207.6$, $df=3,816$, $p<0.001$) (Fig. S1.4). Third, allele frequency showed a sharp cline along the zone, while observed heterozygosity H_O showed a spike, which would be expected from secondary inter-lineage admixture. Finally, STRUCTURE

results from a K=2 run that included meadows from Y-East and Y-North showed a similar cline in genotypic ancestry between those two Hardy-Weinberg gene pools.

NewHybrids found that the two admixture zones have three distinct lineages ("East-North-A1", "East-North-A2", and "East-South-A"), two of which have approximately symmetrical ancestry from parental lineages (designated "F2"). The other has asymmetrical admixture in the northern contact zone (designated "backcrossed" with Y-North). Although the southern zone also appeared to have some backcrossing (Fig. S1.5), the northern backcrossed individuals formed their own clade (Fig. 1.3), hence I treated them separately. Although NewHybrids binned admixed individuals into three general categories, consisting of F2s and each type of F1-parental backcross, the HIest approach suggested these individuals have been admixing for much more than two generations. Based on simulations with N=50 and no gene flow, the observed patterns are consistent with origins between five and 500 generations, although much older hybridization is possible if these parameters are underestimated (Fig. 1.6, Table S1.5). The DAPC analysis confirmed that three admixed regions in YOSE (East-North-A1, East-North-A2, and East-South-A) spanning these two contact zones formed distinct genetic clusters (Fig. S1.6). Using the first 100 PCs and first two discriminant functions, individuals were assigned to the correct lineage 99.4% of the time. When meadow was the grouping factor, correct assignment to lineage was still 98.7%.

Demographic reconstruction of admixture zones

I tested a total of 17 hypothesized models for admixture history in contact zones (10 northern, seven southern). Specifically, I wanted to know whether patterns of admixture reflect true inter-lineage fusion following cessation of gene flow ("hybrid" and "introgression" models), or if gene flow never ceased ("isolation with migration" models, i.e. primary contact) (Fig. 1.7; see Fig. S1.7 for all models). I contrasted all models with the null model of simple isolation. The models with highest likelihood were "Hybrid 3" (northern), and "Hybrid" (southern), both with AIC weights of ~1.0 (Table 1.3). These models inferred an instantaneous origin of admixed lineages T generations ago, after a period of time with no gene flow. The Hybrid 3 model contains two admixture events: East-North-A1 was created from Y-East and Y-North, and then East-North-A2 was created from East-North-A1 and Y-East. I used the divergence date estimates to calibrate those times, and assuming a generation time of five years (estimated mean age of breeding) I estimated admixture dates (in ka) of 473 [203–610] for East-North-A1, 254 [0.055–322] for East-North-A2, and 366 [327–377] for East-South-A (Table 1.4). With the exception of East-North-A2, these events predate the last three glacial episodes. The East-North-A1 lineage had higher genetic input (α) from Y-North (0.44 [0.22–0.82]), whereas East-North-A2 was much closer to Y-East (0.14 [0.14–0.86]). The East-South-A lineage had higher admixture from Y-South (0.61 [0.17–0.77]). These findings were consistent with the HIest results, but large 95% confidence intervals preclude extensive interpretation (e.g. whether backcrossing occurred).

Tests for demographic declines or range expansions following glaciation

Mismatch distributions and EBSPs were consistent with Pleistocene-long population declines across the species, and recent (47–128 ka) demographic expansions. Using the SSD statistic, deviations from expected mismatch distributions under the spatial expansion model were insignificant in all but Y-East and Y-West (Table 1.5, S1.5; Fig. S1.8). All lineages also had insignificant deviations using the raggedness index, except for Y-East. This suggests that most populations have been recolonizing the parks with high levels of population growth and gene flow. This pattern was corroborated by Fu's F_S values, which were all significantly negative ($p<0.02$) except for the two young admixed lineages (East-North-A2, East-South-A) and the most genetically depauperate lineage (Evolution). All values of Tajima's D were not significant which is not unexpected given the lower power of that test. EBSPs showed a strong signal of expansion in Y-North, East-North-A1, Y-West, Goddard, and moderate signal in Y-East, Y-South, and East-South-A (Fig. 1.8). The remaining two originated most recently, and have not have had sufficient time for population sizes to expand since the LGM.

Estimates of pairwise ψ also supported spatial expansions in the pure lineages of YOSE (Table 1.6; Figs. 1.9, S1.9). Interestingly, TDOA-estimated origin points were east of the ice sheet for high-elevation Y-North and Y-East (whose current distributions are underneath the former ice extent), but low-elevation Y-South and Y-West had origin points inside their current distributions, which makes sense because they exist mostly outside the glacial extent.

Phylogeographic reconstruction of glacial refugia and climatic niche divergence in Yosemite NP

The variables with highest permutation importance across all four Maxent models were mean temperature of: wettest quarter, driest quarter, warmest quarter, and coldest quarter (Table S1.6). Area under the curve (AUC) for test data was >0.996 for all models. Climatic predictions of refugial distribution were concordant with the spatial range expansion results, showing largely eastern distributions for Y-North and Y-East, and western distributions for Y-South and Y-West (Fig. 1.9). The LDA on six climatic variables showed that the two low-elevation lineages (Y-South/Y-West) are completely distinct from the others across a complex gradient of temperature and precipitation variables (ignoring elevation) (Fig. S1.10). Y-North and Y-West could be discriminated to a lesser degree from Y-East and Y-South, respectively, based on higher summer precipitations levels. I quantified total pairwise niche overlap while accounting for available niche space, and found Y-South/Y-West to have 46.7% overlap, and Y-North/Y-East to have 10.6% overlap (Table S1.7). A test for niche equivalency found all climatic niches to be significantly divergent except for Y-South/Y-West, which were nearly divergent ($p=0.051$).

DISCUSSION

My results provide abundant evidence that Yosemite toad lineages have evolved across steep climatic selection gradients, and their divergence has been strengthened during bouts of glacially induced vicariance (Table S1.3; Figs 1.3, 1.4, S1.10). Species

and lineage divergence times generally correspond to known dates of glacial maxima (Table S1.3). Lower-elevation toads have repeatedly been isolated in-situ west of Pleistocene ice sheets, while their higher-elevation brethren have been forced eastward into Owens Valley, only to recolonize later (Tables 1.5, 1.6, S1.5; Figs. 1.8, 1.9, S1.8, S1.9). This has resulted in a phylogenetic pattern of sister lineages residing in divergent climatic niches. For example, Y-West and Y-East have been isolated into western and eastern refugia, and now occupy distinct low- and high-elevation niches (Fig. 1.3). Likewise, the ancestor of these two lineages (Y-West + Y-East) likely resided in a high-elevation niche, and diverged from the low-elevation Y-South. This can be inferred from the much higher genetic diversity of Y-East compared to either Y-West or Y-South. This phylogenetic topology means that low-elevation lineages (Y-South and Y-West) have converged on nearly identical low-elevation climatic niches (Table S1.7, Fig. S1.10). Given the configuration of these lineages and major geographic barriers, only three possible secondary contact zones exist within YOSE, and the widest zone is found between the two lineages with most similar niches (Y-East and Y-North). Furthermore, at least two contact zones have apparently generated novel lineages through the process of lineage fusion (Tables 1.3, 1.4; Figs. 1.5, 1.6). These results are consistent with the model of glacial pulses (Fig. 1.1), which may apply to other alpine endemics. I now discuss these results in turn.

Phylogeographic boundaries and dates of past lineage isolation

I detected substantial phylogeographic structure within YOSE and between the two parks (Figs. 1.3, 1.4, S1.6). Previous studies have found two major mitochondrial clades of *A. canorus*, often polyphyletic with respect to *A. boreas*, with alternate samples from Ireland Lake (in Y-East) falling into either clade (Shaffer and Fellers 2000; Stephens 2001; Goebel and Ranker 2009). Although their spatial sampling was limited, my northern contact zone is likely the boundary of those previously described clades.

Previous patterns of paraphyly or polyphyly with respect to *A. boreas* based solely on mitochondrial data (Graybeal 1993; Shaffer and Fellers 2000; Stephens 2001; Pauly 2008; Goebel and Ranker 2009) are unsupported based on my ddRAD dataset. My data strongly supports a monophyletic *A. canorus,* with all eight *A. boreas* specimens from throughout California (including YOSE) forming a monophyletic sister group to the species. Previous unpublished work based on three nuclear genes has shown a similar pattern (G. Pauly, pers. comm.). Therefore, mitochondrial introgression by female-biased dispersal, asymmetrical population sizes, or environmental selection on sex-linked traits during rare hybridization events probably caused the previously observed topologies (McGuire et al. 2007; Fontenot et al. 2011; Pavlova et al. 2013). However, the placement of *A. exsul* (another Pleistocene relict species related to *A. boreas* and *A. canorus*) is still unknown and could still render *A. canorus* paraphyletic. Future studies with the intent of species delimitation in the *A. boreas* group should use high-coverage nuclear markers and full spatial coverage of all four species (*A. boreas, A. canorus, A. exsul,* and *A. nelsoni*), as their phylogenetic relationships are still uncertain.

Using a published molecular clock for anuran non-synonymous nuclear DNA, I estimated that the two parks (possibly representative of the entire species) diverged 1.99±0.12 Ma, corresponding to the approximate onset of Sierra Nevada glaciation (Calsbeek et al. 2003). The median estimate for all subsequent divergences occurred between 732 and 214 ka (Fig. 1.3, Table S1.3). Beginning in the late Pleistocene (ca. 900 ka), glacial periodicity switched from 41 to 100 ka cycles, with longer duration, shorter inter-glacials, and colder average temperature (Mudelsee and Schulz 1997; Tziperman and Gildor 2003). Hence, intra-park divergences could reflect stronger isolating events during this period. Numerous other arctic and alpine temperate species exhibit deep divergences dating back to the mid-Pleistocene, indicating they retain distinctiveness over several ice ages, and lineage diversification proceeds by repeated allopatry ("glacial pulse" mechanism) (Hewitt 1996, 2004; Avise et al. 1998). All of my divergence estimates (1.99 Ma – 214 ka) are much older than the five most recent glaciation events (Gillespie and Clark 2011), which is consistent with remaining distinct during interglacial time. This suggests these lineages have either survived through glacial maxima in high-elevation nunatak or peripheral refugia (Holderegger and Thiel-Egenter 2009), or remained reproductively isolated while in lowland refugia (discussed below).

Although theory predicts that cold-adapted species should expand their ranges during ice ages (Haffer 1969), montane and alpine amphibians may counter this pattern by remaining in refugia within their current distributions. For example, phylogeographic and isolation by distance analyses showed that Pygmy salamanders (*Desmognathus wrighti*) remained fragmented at high elevation during the Pleistocene, and this restriction

was likely driven by ecological interactions (Crespi et al. 2003). Interestingly, another Sierra Nevada endemic amphibian (the sympatric *Rana sierrae*) has three gene pools with almost identical spatial extent to *A. canorus* lineages in YOSE (Poorten et al. 2017). If divergence times turn out to be similar, this would be strong support that common glacial barriers and refugia have structured each species. For my forgoing conclusions, I note the caveat that using a rate calibration from other species assumes similar rates of evolution. This assumption may be violated; for example, a relatively low historical N_e could accelerate fixation, biasing divergence estimates upward. With increasing genome-wide studies of phylogeography, genomic rates of evolution will be better assessed and calibrated across diverse taxa.

Genetic consequences of inter-lineage admixture

My analyses suggest that secondary contact following isolation has fused together new and persistent lineages of admixed origin (Tables 1.3, 1.4; Figs. 1.5, 1.6). Admixture tended to be asymmetrical, suggesting that backcrossing may be important during this process, although the confidence intervals of fastsimcoal estimates were too broad to corroborate this pattern. Lineage fusion presents an intriguing research opportunity for admixed fitness and conservation. Traditionally, hybridization is discussed in the context of interspecific gene flow or introgression between native and exotic species. In these cases of anthropogenic hybridization, two negative consequences are likely: outbreeding depression follows from maladaptive allele combinations and results in low hybrid vigor or fertility, or genetic assimilation erodes locally adapted populations and species

boundaries (Ellstrand and Elam 1993; Weber and D'Antonio 2000). For example, introduction of an exotic ambystomatid salamander into southern California has sent a rapid influx of morphological and genetic changes across populations of *Ambystoma californiense*, where hybrids have outcompeted natives of that and other species (Ryan et al. 2009; Fitzpatrick et al. 2010). However a different view of natural hybridization posits it is common (Mallet 2005), and this new source of genetic variation might be important irrespective of fitness in F1 and other early filial generations (Arnold et al. 1999).

Adaptive introgression could be an essential source of adaptive genetic novelty in species such as the Yosemite toad, with an average N_e of 29.13 (Table S1.2). Preliminary analysis of Yosemite toad evolutionary rates have shown a strong negative relationship with estimated census population size, indicating genetic drift might overpower adaptation from standing variation and new mutations (P. Maier, unpublished data). Although recombination in admixed individuals can disrupt coadapted genotypes underlying polygenic traits, a selective advantage is the shorter waiting time to generation of adaptive recombinant alleles. Given a mutation rate μ for beneficial mutations, and n potentially adaptive loci underlying a trait, adaptive recombinant alleles are generated every n/μ generations (compared with every $1/\mu$ generations for point mutations) (Hedrick 2013). Since admixture is a recurrent process, the continual flow of adaptive alleles can raise the initial frequency of them considerably, and lead to much faster fixation (Hedrick 2013). This process, in concert with possible heterosis experienced by F1s and other early filial individuals, can dramatically expedite adaptive evolution. For example, hybridization between several closely related species of Darwin's finches on

Daphne Island has contributed at least 10% additive genetic variance in beak shape, a trait under intense and episodic selection (Grant and Grant 2008, 2010, 2014; Grant 2015). Subsequent fluctuations in food availability during an El Niño event allowed hybrids to survive longer, and backcrossed individuals eventually formed a morphologically, ecologically, and reproductively distinct lineage.

In the present study, I detected three lineage fusions of moderate age (Table 1.4), whose descendants appear to maintain independent evolutionary trajectories. With ongoing admixture zones involving independently evolving lineages, more extreme traits are possible, for example very high or low tadpole growth rates (Parris 1999; Reyer 2008). Under conditions where ponds are shallow and tend to desiccate, admixture in pond-breeding anurans is more widespread (Kingsolver et al. 2002; Pfennig 2007), and natural selection can more efficiently replace old traits with new, successful ones because they are already at high frequencies (Hedrick 2013). There are many examples of intraspecific admixture increasing the viability of translocated larvae and adults (e.g. Japanese common toads: Hase et al. 2013; European toads: Zeisset and Beebee 2013). Given the extremely short and variable hydroperiods Yosemite toads experience during larval development, it seems likely that tadpole physiology would be under strong selection. Thus, any heritable Yosemite toad fitness traits that have diverged between lineages may be studied for conservation value, to ascertain whether (and in what environmental context) pure versus admixed individuals would make superior stock for translocation efforts (Hamilton and Miller 2016). Although the longevity and number of

lineage fusions hints at their population fitness, my evidence does not directly suggest hybrid advantage, a subject which future studies should examine at the phenotypic level.

Evidence and implications of the glacial pulse model

Recently, there has been a debate about the role of niche evolution in speciation. Proponents of niche conservatism argue that species innately have low tolerance to adapt outside their fundamental niche, meaning that climate shifts can fragment the distribution of that niche, leading to allopatric speciation (Wiens 2004; Wiens and Graham 2005; Wiens et al. 2010). Under this model, ecological divergence is the result (not the cause) of lineage formation. However, this process does not convincingly explain niche divergence for adjacent species pairs occupying steep mountainous gradients. These are commonly observed in animals (e.g. Graham et al. 2004) and plants (e.g. Chabot and Billings 1972). In contrast, proponents of ecological speciation argue that natural selection can overpower gene flow to produce local adaptation, and hence niche divergence directly spawns new lineages (Dobzhansky 1951; Schluter 2001; Rundle and Nosil 2005). Although there are numerous examples of divergent selection driving incipient ecological speciation, advocates readily admit that the process is often incomplete (Nosil et al. 2009).

My glacial pulse model combines elements of niche divergence with the classical and widely accepted model of allopatric speciation and associated reproductive incompatibilities (Mayr 1942; Coyne and Orr 1989; Barraclough and Vogler 2000). It unifies the two principal explanations for speciation in Sierra Nevada alpine endemics:

(1) colonization of mountains followed by physiological or ecological divergence (Went 1948; Chabot and Billings 1972; Billings 1974), and (2) isolation of formerly continuous species in glacial or interglacial refugia (Rovito 2010; Rubidge et al. 2014). My model suggests these both happen, cyclically, and they are mutually reinforcing (Fig. 1.1). In this study, I found evidence that Yosemite toads have repeatedly diverged in climatic niche while isolated into distinct (west vs. east) refugia. The fact that high- and low-elevation niches have evolved more than once (i.e. convergence between Y-South and Y-West) suggests a strong influence of ecological selection, and the contact zones are evidence of associated genetic distinctness via allopatry. Although the Sierra Nevada are unique among mountain ecosystems in having a Mediterranean climate (Rundel 2011; Rundel and Millar 2016), the patterns found here may be widespread. For example, similar patterns (i.e. repeated ecological selection, glacial allopatry, and recolonization) have been observed in European grasshoppers (Hewitt 1996).

One assumption of the model is heterogeneity of refugial climates, because post-glacial recolonization must proceed into (partially) non-overlapping niches. In the case of the Sierra Nevada, it has long been acknowledged that high-Sierra and eastern desert (Owen's Valley) ecosystems have similar growing seasons (Went 1948), explaining why the latter was a suitable refugium for high-elevation and not low-elevation Yosemite toad lineages. Climatic conditions in Owen's Valley over the past 155 ka were also moister and cooler than at present (Menking et al. 1997; Koehler et al. 2005), which might explain the much higher genetic diversity of high-elevation lineages. Another Sierra Nevada endemic amphibian (*Hydromantes platycephalus*) reaches its highest diversity in

its eastern lineage, near Owen's Valley, which seems to corroborate this pattern (Rovito 2010). Diversity patterns may also reflect current habitat: although high-elevation lineages must periodically retreat into refugia, their meadows are rejuvenated with new alluvial deposits that may positively influence hydrology for tadpoles (Wood 1975). Conversely, low-elevation lineages that remain in situ may experience little benefit from glacial action, and higher rates of conifer encroachment into meadows as trees recolonize middle elevations (Woolfenden 1996; Lubetkin et al. 2017).

CONCLUSIONS

In this study, I have provided evidence that pulses of glacial action bisect lineages into divergent climatic niches, which then can form further lineages by fusion. This mechanism is unique from other described mechanisms for lineage diversification, and should be considered in phylogeographic hypothesis testing for other systems. The process of lineage fusion in secondary contact zones is particularly intriguing; given that the species is federally threatened, any adaptive or deleterious effects of inter-lineage admixture should be fully elucidated before attempting conservation measures, such as translocations. I suggest the possibility that such a zone is a crucible for valuable adaptive diversity to combat stressors such as disease and periodic climatic extremes, such as intense drought, but this must be investigated. All nine lineages in my study area should be considered potential evolutionarily significant units (ESUs), since they appear to be evolving independently, and occupy divergent climatic niches in most cases. The future for alpine endemics is grim; one study estimated that up to 66% of California endemic

plants will experience large (>80%) range reductions by 2100 (Loarie et al. 2008). However, my results are promising in that lineages have evolved differing climate optima, and hence should respond differently to climate change. Future studies may capitalize from my results by examining the fitness differences between pure/admixed, and low-/high-elevation lineages, to anticipate future changes.

REFERENCES

* References marked with asterisk have been cited in Appendix A only.

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Abbott, R. J., and A. C. Brennan. 2014. Altitudinal gradients, plant hybrid zones and evolutionary novelty. Philos. Trans. R. Soc. B Biol. Sci. 369:1–12.

Anderson, E. C., and E. A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160:1217–1229.

Arnold, M. L., M. R. Bulger, J. M. Burke, and A. L. Hempel. 1999. Natural hybridization: how low can you go and still be important? Ecology 80:371–281.

Avise, J. C., D. Walker, and G. C. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. R. Soc. B Biol. Sci. 265:1707–1712.

Barraclough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. Am. Nat. 155:419–434.

Berlow, E. L., R. Knapp, S. M. Ostoja, R. J. Williams, H. McKenny, J. R. Matchett, Q. Guo, G. M. Fellers, P. Kleeman, M. L. Brooks, and L. Joppa. 2013. A network extension of species occupancy models in a patchy environment applied to the Yosemite Toad (*Anaxyrus canorus*). PLoS One 8:e72200.

Billings, W. D. 1974. Adaptations and origins of alpine plants. Arct. Alp. Res. 6:129–142.

Broennimann, O., M. C. Fitzpatrick, P. B. Pearman, B. Petitpierre, L. Pellissier, N. G. Yoccoz, W. Thuiller, M. J. Fortin, C. Randin, N. E. Zimmermann, C. H. Graham, and A. Guisan. 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. Glob. Ecol. Biogeogr. 21:481–497.

Calsbeek, R., J. N. Thompson, and J. E. Richardson. 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: The California Floristic Province. Mol. Ecol. 12:1021–1029.

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

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes, Genomes, Genet. 1:171–182.

Chabot, B. F., and W. D. Billings. 1972. Origins and ecology of the Sierran alpine flora and vegetation. Ecol. Monogr. 42:163–199.

Coyne, J., and H. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362–381.

Crawford, A. J. 2003. Relative rates of nucleotide substitution in frogs. J. Mol. Evol. 57:636–641.

Crespi, E. J., L. J. Rissler, and R. A. Browne. 2003. Testing Pleistocene refugia theory: phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. Mol. Ecol. 12:969–984.

Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Resour. 14:209–214.

Dobzhansky, T. 1951. Genetics and the origin of species. 3rd ed. Columbia University Press, New York, NY.

Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:1–8.

Ellstrand, N. C., and D. R. Elam. 1993. Population genetic consequences of small population size: implications for plant conservation. Annu. Rev. Ecol. Evol. Syst. 24:217–242.

Excoffier, L. 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Mol. Ecol. 13:853–864.

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

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University Press, Oxford, UK.

Fitzpatrick, B. M. 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evol. Biol. 12:131.

Fitzpatrick, B. M., J. R. Johnson, D. K. Kump, J. J. Smith, S. R. Voss, and H. B. Shaffer. 2010. Rapid spread of invasive genes into a threatened native species. Proc. Natl. Acad. Sci. USA 107:3606–3610.

Fontenot, B. E., R. Makowsky, and P. T. Chippindale. 2011. Nuclear-mitochondrial discordance and gene flow in a recent radiation of toads. Mol. Phylogenet. Evol. 59:66–80.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925.

Galtier, N., B. Habholz, S. Glemin, and G. D. D. Hurst. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol. Ecol. 18:4541–4550.

Gillespie, A. R., and D. H. Clark. 2011. Glaciations of the Sierra Nevada, California, USA. Pp. 447–462 *in* Developments in Quaternary Science.

Goebel, A., and T. Ranker. 2009. Mitochondrial DNA evolution in the *Anaxyrus boreas* species group. Mol. Phylogenet. Evol. 50:209–225.

Gompert, Z., A. M. Shapiro, and C. C. Nice. 2006. Homoploid hybrid speciation in an extreme habitat. Science 314:1923–1925.

Graham, C. H., S. R. Ron, J. C. Santos, C. J. Schneider, and C. Moritz. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. Evolution 58:1781–1793.

Grant, B. R. 2015. Introgressive hybridization and natural selection in Darwin's finches. Biol. J. Linn. Soc. 117:812–822.

Grant, P. R., and B. R. Grant. 2010. Conspecific versus heterospecific gene exchange between populations of Darwin's finches. Philos. Trans. Biol. Sci. 365:1065–1076.

Grant, P. R., and B. R. Grant. 2014. Synergism of natural selection and introgression in the origin of a new species. Am. Nat. 183:671–681.

Grant, P. R., and R. Grant. 2008. How and why species multiply: the radiation of Darwin's finches. Princeton University Press, Princeton, NJ.

Graybeal, A. 1993. The phylogenetic utility of cytochrome b: lessons from bufonid frogs. Mol. Phylogenet. Evol. 2:256–269.

Haffer, J. 1969. Speciation in Amazonian forest birds. Science 165:131–137.

Hamilton, J. A., and J. M. Miller. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv. Biol. 30:33–41.

Harrison, R. G. (ed). 1993. Hybrid zones and the evolutionary process. Oxford University Press, Ithaca, NY.

Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. J. Hered. 105:795–809.

Hase, K., N. Nikoh, and M. Shimada. 2013. Population admixture and high larval viability among urban toads. Ecol. Evol. 3:1677–1691.

Hedrick, P. W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Heller, R., L. Chikhi, and H. R. Siegismund. 2013. The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. PLoS One 8:e62992.

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

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

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

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25:1965–1978.

* Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6:e1000862.

Holderegger, R., and C. Thiel-Egenter. 2009. A discussion of different types of glacial refugia used in mountain biogeography and phylogeography. J. Biogeogr. 36:476–480.

Huber, N. K. 1981. Amount and timing of late Cenozoic uplift and tilt of the central Sierra Nevada, California: Evidence from the upper San Joaquin River basin. U.S. Geological Survey professional paper. Washington D.C., USA.

Ivanova, N. V, J. R. Dewaard, and P. Hebert. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. Mol. Ecol. Notes 6:998–1002.

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

Jombart, T., S. Devillard, A. Dufour, and D. Pontier. 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:92–103.

Karlstrom, E. L. 1962. The toad genus *Bufo* in the Sierra Nevada of California. Unviersity Calif. Publ. Zool. 62:1–104.

Keeler-Wolf, T., E. T. Reyes, J. M. Menke, D. N. Johnson, and D. L. Karavidas. 2012. Yosemite National Park vegetation classification and mapping project report. National Park Service report. Fort Collins, CO.

Kingsolver, J. G., D. W. Pfennig, and M. R. Servedio. 2002. Migration, local adaptation and the evolution of plasticity. Trends Ecol. Evol. 17:540–541.

Knaus, B. J., and N. J. Grünwald. 2017. vcfr: a package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour. 17:44–53.

Koehler, P. A., R. S. Anderson, and W. G. Spaulding. 2005. Development of vegetation in the Central Mojave Desert of California during the late Quaternary. Palaeogeogr. Palaeoclimatol. Palaeoecol. 215:297–311.

Kuchta, S. R. 2007. Contact zones and species limits: hybridization between lineages of the California newt, *Taricha torosa*, in the southern Sierra Nevada. Herpetologica 63:332–350.

Lapointe, F., and L. J. Rissler. 2005. Congruence, consensus, and the comparative phylogeography of codistributed species in California. Am. Nat. 166:290–299.

* Lemmon, A. R., and E. M. Lemmon. 2012. High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. Syst. Biol. 61:745–761.

Lischer, H. E. L., L. Excoffier, and G. Heckel. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. Mol. Biol. Evol. 31:817–831.

Loarie, S. R., B. E. Carter, K. Hayhoe, S. McMahon, R. Moe, C. A. Knight, and D. D. Ackerly. 2008. Climate change and future of California's endemic flora. PLoS One 3:e2502.

Lubetkin, K. C., A. L. Westerling, and L. M. Kueppers. 2017. Climate and landscape drive the pace and pattern of conifer encroachment into subalpine meadows. Ecol. Appl. 27:1876–1887.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229–237.

Mavárez, J., and M. Linares. 2008. Homoploid hybrid speciation in animals. Mol. Ecol. 17:4181–4185.

Mayr, E. 1942. Systematics and the origin of species: from the viewpoint of a zoologist. Columbia University Press, New York, NY.

McGuire, J. A., C. W. Linkem, M. S. Koo, D. W. Hutchison, A. K. Lappin, D. I. Orange, J. Lemos-Espinal, B. R. Riddle, and J. R. Jaeger. 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. Evolution 61:2879–2897.

Menking, K., J. Bischoff, J. Fitzpatrick, J. W. Burdette, and R. O. Rye. 1997. Climatic/hydrologic oscillations since 155,000 yr B.P. at Owens Lake, California, reflected in abundance and stable isotope composition of sediment carbonate. Quat. Res. 68:58–68.

Mudelsee, M., and M. Schulz. 1997. The mid-Pleistocene climate transition: onset of 100 ka cycle lags ice volume build-up by 280 ka. Earth Planet. Sci. Lett. 151:117–123.

Mulch, A., A. M. Sarna-Wojcicki, M. E. Perkins, and C. P. Chamberlain. 2008. A Miocene to Pleistocene climate and elevation record of the Sierra Nevada (California). Proc. Natl. Acad. Sci. U. S. A. 105:6819–6824.

Mullally, D., and J. Cunningham. 1956. Aspects of the thermal ecology of the Yosemite toad. Herpetologica 12:57–67.

Myers, N., R. Mittermeier, C. Mittermeier, G. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature 403:853–858.

Nosil, P., L. J. Harmon, and O. Seehausen. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.

Ostoja, S. M., S. R. Lee, J. R. Matchett, H. McKenny, M. L. Brooks, R. A. Knapp, P. A. Maier, and N. Danielle. 2018. Yosemite toad (*Anaxyrus* [*Bufo*] *canorus*) breeding occupancy of meadows in Yosemite and Kings Canyon National Parks, California. Unpublished manuscript.

Parris, M. J. 1999. Hybridization in leopard frogs (*Rana pipiens* complex): larval fitness components in single-genotype populations and mixtures. Evolution 53:1872–1883.

Pauly, G. 2008. Phylogenetic systematics, historical biogeography, and the evolution of vocalizations in nearctic toads (*Bufo*). Austin, TX: University of Texas at Austin. Ph.D. dissertation.

Pavlova, A., J. N. Amos, L. Joseph, K. Loynes, J. J. Austin, J. S. Keogh, G. N. Stone, J. A. Nicholls, and P. Sunnucks. 2013. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. Evolution 67:3412–3428.

Peter, B. M., and M. Slatkin. 2013. Detecting range expansions from genetic data. Evolution 67:3274–3289.

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

Pfennig, K. S. 2007. Facultative mate choice drives hybridization. Science 318:965–967.

Phillips, S. J., M. Dudík, and R. E. Schapire. 2004. A maximum entropy approach to species distribution modeling. 21st Int. Conf. Mach. Learn. Banff, Canada 655–662.

Poorten, T. J., R. A. Knapp, and E. B. Rosenblum. 2017. Population genetic structure of the endangered Sierra Nevada yellow-legged frog (*Rana sierrae*) in Yosemite National Park based on multi-locus nuclear data from swab samples. Conserv. Genet. 18:731–744.

Posada, D. 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

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

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ratliff, R. D. 1982. A meadow site classification for the Sierra Nevada, California. U.S. Forest Service report. Berkeley, CA.

Ratliff, R. D. 1985. Meadows in the Sierra Nevada of California: state of knowledge. U.S. Forest Service report. Berkeley, CA.

Reyer, H. U. 2008. Mating with the wrong species can be right. Trends Ecol. Evol. 23:289–292.

Rissler, L. J., R. J. Hijmans, C. H. Graham, C. Moritz, and D. B. Wake. 2006. Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. Am. Nat. 167:655–666.

Ritz, C., and A. N. Spiess. 2008. qpcR: An R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. Bioinformatics 24:1549–1551.

Rovito, S. M. 2010. Lineage divergence and speciation in the web-toed salamanders (Plethodontidae: *Hydromantes*) of the Sierra Nevada, California. Mol. Ecol. 19:4554–4571.

Rubidge, E. M., J. L. Patton, and C. Moritz. 2014. Diversification of the alpine chipmunk, *Tamias alpinus*, an alpine endemic of the Sierra Nevada, California. BMC Evol. Biol. 14:34.

Rundel, P. W. 2011. The diversity and biogeography of the alpine flora of the Sierra Nevada, California. Madroño 58:153–184.

Rundel, P. W., and C. I. Millar. 2016. Alpine ecosystems. Pp. 613–634 *in* H. Mooney and E. Zavaleta, eds. Ecosystems of California. University of California Press, Berkeley, CA.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecol. Lett. 8:336–352.

Ryan, M. E., J. R. Johnson, and B. M. Fitzpatrick. 2009. Invasive hybrid tiger salamander genotypes impact native amphibians. Proc. Natl. Acad. Sci. USA 106:11166–11171.

Schluter, D. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372–380.

Schoville, S. D., and G. K. Roderick. 2009. Alpine biogeography of Parnassian butterflies during Quaternary climate cycles in North America. Mol. Ecol. 18:3471–3485.

Schoville, S. D., T. S. Tustall, V. T. Vredenburg, A. R. Backlin, E. Gallegos, D. A. Wood, and R. N. Fisher. 2011. Conservation genetics of evolutionary lineages of the endangered mountain yellow-legged frog, *Rana muscosa* (Amphibia: Ranidae), in southern California. Biol. Conserv. 144:2031–2040.

Schumer, M., G. G. Rosenthal, and P. Andolfatto. 2014. How common is homoploid hybrid speciation? Evolution 68:1553–1560.

Shaffer, H., and G. Fellers. 2000. The genetics of amphibian declines: population substructure and molecular differentiation in the Yosemite toad, *Bufo canorus* (Anura, Bufonidae) based on single-strand conformation polymorphism analysis (SSCP) and mitochondrial DNA sequence data. Mol. Ecol. 9:245–257.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stephens, M. 2001. Phylogeography of the *Bufo boreas* (Anura, Bufonidae) species complex and the biogeography of California. Rohnert Park, CA: Sonoma State University. M.S. thesis.

Swenson, N. G., and D. J. Howard. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. Am. Nat. 166:581–591.

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

Tziperman, E., and H. Gildor. 2003. On the mid-Pleistocene transition to 100-kyr glacial cycles and the asymmetry between glaciation and deglaciation times. Paleoceanography 18:1–8.

Unruh, J. R. 1991. The uplift of the Sierra Nevada and implications for late Cenozoic epeirogeny in the western Cordillera. Geol. Soc. Am. Bull. 103:1395–1404.

US Fish & Wildlife Service. 2014. Endangered and threatened wildlife and plants; endangered status for the Sierra Nevada yellow-legged frog and the northern distinct population segment of the mountain yellow-legged frog, and threatened status for the Yosemite toad: final rule. Fed. Regist. 1–56.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Springer, New York.

Viers, J. H., S. E. Purdy, R. A. Peek, A. Fryjoff-Hung, N. R. Santos, J. V. Katz, J. D. Emmons, D. V Dolan, and S. M. Yarnell. 2013. Montane meadows in the Sierra Nevada: changing hydroclimatic conditions and concepts for vulnerability assessment. Centre for Watershed Sciences report. Davis, CA.

Wallis, G. P., J. M. Waters, P. Upton, and D. Craw. 2016. Transverse alpine speciation driven by glaciation. Trends Ecol. Evol. 31:916–926.

Waltari, E., R. J. Hijmans, A. T. Peterson, Á. S. Nyári, S. L. Perkins, and R. P. Guralnick. 2007. Locating pleistocene refugia: comparing phylogeographic and ecological niche model predictions. PLoS One 2:e563.

Wang, I. J. 2012. Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. Divers. Distrib. 18:1033–1041.

Weber, E., and C. M. D'Antonio. 2000. Conservation implications of invasion by plant hybridization. Biol. Invasions 2:207–217.

Weir, B. S. 1996. Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, Inc., Sunderland, MA.

Went, F. W. 1948. Parallels between desert and alpine flora in California. Madroño 9:241–249.

Wiens, J. J. 2004. Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. Evolution 58:193–197.

Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, B. A. Hawkins, R. D. Holt, C. M. McCain, and P. R. Stephens. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecol. Lett. 13:1310–1324.

Wiens, J. J., and C. H. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Annu. Rev. Ecol. Evol. Syst. 36:519–539.

Wood, S. H. 1975. Holocene stratigraphy and chronology of mountain meadows, Sierra Nevada, California. Pasadena, CA: California Institute of Technology. Ph.D. dissertation.

Woolfenden, W. B. 1996. Quaternary vegetation history. Pp. 47–70 *in* Sierra Nevada Ecosystem Project: Final Report to Congress.

Zeisset, I., and T. J. C. Beebee. 2013. Donor population size rather than local adaptation can be a key determinant of amphibian translocation success. Anim. Conserv. 16:359–366.

TABLES AND FIGURES

Table 1.1. Summary of double-digest RAD loci. Average number of loci, SNPs, haplotypes, and locus coverage for each park and dataset. "R1" loci contain only the first read (first 96 nt) of each locus, "concatenated" loci contain merged reads where both passed quality filtering (197 nt). "Total" summarizes genomic diversity using the concatenated loci where possible, and R1 otherwise. "Total SNPs" indicates the number present across the given locus, however only the first SNP was retained for SNP marker datasets. "No. Haplotypes" is the number of unique phased SNP haplotypes observed (maximum of two if only one SNP is present at that locus, maximum four for two SNPs, etc.).

	Yosemite NP			Kings Canyon NP			Both Parks		
	R1	Concat.	Total	R1	Concat.	Total	R1	Concat.	Total
No. Loci	2318	1044	2318	1914	1341	1914	3261	1347	3211
Total SNPs	4424	2945	5310	2993	2861	3642	9146	6468	11635
No. Haplotypes	5994	3212	6473	4375	3320	4607	11176	6308	12721
SNPs/Locus	1.91 (1-8)	2.82 (1-14)	2.29 (1-14)	1.56 (1-7)	2.13 (1-9)	1.90 (1-9)	2.80 (1-12)	4.80 (1-17)	3.62 (1-17)
Haps/Locus	2.58 (2-7)	3.07 (2-11)	2.78 (2-11)	2.28 (2-5)	2.48 (2-8)	2.41 (2-8)	3.42 (2-12)	4.68 (2-20)	3.95 (2-20)
Avg. Coverage	95.58	55.15	83.07	79.80	49.69	67.40	92.21	54.38	81.22

Table 1.2. **Population genetic summary statistics.** Summary of parameters for each meadow, averaged at the lineage level. Results summarized at the meadow level are available in Table S1.2.

Lineage	Park	Unit Type	PA	N	P	H_o	H_e	π	F_{IS}	% Pol Loci	Lineage N_e
North	YOSE	Pure	22.900	8.173	0.961	0.061	0.052	0.056	-0.008	0.344	14.300
East-North-A1	YOSE	Admixed	11.857	6.538	0.959	0.063	0.055	0.061	-0.003	0.354	28.400
East-North-A2	YOSE	Admixed	7.333	4.485	0.961	0.061	0.052	0.058	-0.005	0.316	9.800
East	YOSE	Pure	6.811	5.053	0.962	0.060	0.049	0.056	-0.007	0.302	74.900
East-South-A	YOSE	Admixed	6.500	10.136	0.957	0.063	0.058	0.061	-0.001	0.398	17.300
South	YOSE	Pure	5.400	5.702	0.969	0.051	0.040	0.045	-0.012	0.243	25.900
West	YOSE	Pure	5.938	5.154	0.967	0.052	0.042	0.047	-0.009	0.256	36.300
Evolution	KICA	Pure	28.000	6.631	0.966	0.055	0.045	0.050	-0.011	0.291	3.300
Goddard	KICA	Pure	23.700	9.112	0.953	0.068	0.063	0.067	-0.001	0.438	24.300
	All		**13.160**	**6.776**	**0.962**	**0.059**	**0.051**	**0.056**	**-0.006**	**0.327**	**26.056**
	YOSE		**9.534**	**6.463**	**0.962**	**0.059**	**0.050**	**0.055**	**-0.006**	**0.316**	**29.557**
	KICA		**25.850**	**7.872**	**0.960**	**0.062**	**0.054**	**0.058**	**-0.006**	**0.365**	**13.800**
		Pure	**15.458**	**6.638**	**0.963**	**0.058**	**0.049**	**0.053**	**-0.008**	**0.312**	**29.833**
		Admixed	**8.563**	**7.053**	**0.959**	**0.062**	**0.055**	**0.060**	**-0.003**	**0.356**	**18.500**

* Abbreviations: YOSE = Yosemite NP, KICA = Kings Canyon NP, PA = private alleles, N = effective sample size, P = frequency of most frequent allele, H_o = observed heterozygosity, H_e = expected heterozygosity, π = average gene diversity, F_{IS} = fixation index, % Pol Loci = % sites that are locally polymorphic, Lineage N_e = effective population size estimated for entire lineage (estimated using linkage disequilibrium method).

Table 1.3. **Results of model testing using fastsimcoal.** Two admixture zones were tested: East-North and East-South. A graphical depiction of models and model parameters is shown in Figs. 1.7 and S1.7. Each model was run independently 100 times using 100,000 coalescent simulations with 40 ECM cycles each time. The most likely model is shown in bold.

Zone	Model	Max. AIC	No. Param.	ΔAIC	AIC Weight
East-North Admix.	Isolation	-3771.12	7	-51.744	2.49×10^{-22}
	Iso./Migration	-3904.74	15	-185.37	7.75×10^{-84}
	Hybrid 1	-3729.53	9	-10.157	3.88×10^{-5}
	Hybrid 2	-3748.28	9	-28.906	2.79×10^{-13}
	Hybrid 3	**-3719.37**	**9**	**0**	**~1.00**
	Hybrid 4	-3733.56	8	-14.185	1.88×10^{-6}
	Intro. 1	-3854.50	9	-135.131	2.06×10^{-59}
	Intro. 2	-3792.58	9	-73.206	1.61×10^{-32}
	Intro. 3	-3846.50	9	-127.124	6.18×10^{-56}
	Intro. 4	-3767.68	8	-48.302	2.86×10^{-21}
East-South Admix	Isolation 1	-4073.765	5	-23.481	1.72×10^{-10}
	Isolation 2	-4064.046	5	-13.762	2.87×10^{-6}
	Iso./Migration 1	-5784.45	11	-1734.16	~0.00
	Iso./Migration 2	-5700.31	11	-1650.03	~0.00
	Hybrid	**-4050.284**	**6**	**0**	**~1.00**
	Intro. 1	-4182.295	6	-132.011	4.66×10^{-58}
	Intro. 2	-4704.026	6	-653.742	1.21×10^{-284}

Table 1.4. **Demographic parameter estimates from fastsimcoal.** Estimates are obtained from models with highest weighted AIC scores from the East-North and East-South admixture zones. Population sizes are in haploid units. Admixture times are in units of years, with an assumed generation time of five years. See Figs. 1.7 and S1.7 for graphical depictions of models and parameters. Estimates are the highest maximum likelihood scores from 100 replicate fastsimcoal runs. The 95% confidence intervals were estimated using 100 pseudoreplicate SFS datasets and rerunning each model 100 times, using the best model parameters as starting values.

Model		Parameter								
		N_{Anc}	N_N	N_{A1}	N_{A2}	N_E	α_{A1}	α_{A2}	T_{A1}	T_{A2}
East-North Admix.	estimate	1.17×10^7	9.65×10^5	7.11×10^5	4.15×10^5	7.17×10^5	0.44	0.14	4.73×10^5	2.54×10^5
	95% CI	[1.17×10^7 – 1.77×10^7]	[4.55×10^5 – 9.75×10^5]	[8.29×10^5 – 7.11×10^5]	[0.95×10^5 – 4.29×10^5]	[6.34×10^5 – 8.10×10^5]	[0.22 – 0.82]	[0.14 – 0.86]	[2.03×10^5 – 6.10×10^5]	[55 – 3.22×10^5]
		N_{Anc}	N_S	N_A	N_E	α_A	T_A			
East-South Admix.	estimate	1.02×10^7	3.75×10^5	3.34×10^5	6.12×10^5	0.61	3.66×10^5			
	95% CI	[9.53×10^6 – 1.21×10^7]	[1.97×10^5 – 4.05×10^5]	[2.74×10^5 – 5.50×10^5]	[3.99×10^5 – 6.92×10^5]	[0.17 – 0.77]	[3.27×10^5 – 3.77×10^5]			

***** Abbreviations: population sizes for North (N_N), East-North-A1 (N_{A1}), East-North-A2 (N_{A2}), East (N_E), South (N_S), East-South-A (N_A), and ancestral (N_{Anc}) lineages; admixture proportions from East → East-North-A1 (α_{A1}), East-North-A1 → East-North-A2 (α_{A2}), and South → East-South-A (α_A); and admixture times for East-North-A1 (T_{A1}), East-North-A2 (T_{A2}), and East-South-A (T_A) lineages.

Table 1.5. **Results of spatial expansion, Tajima's D, and Fu's F_S tests.** The null hypothesis for the expansion test is actually the hypothesis of interest, namely spatial population expansion, so significant ($p<0.05$) values reject the expansion model. Significance levels for rejecting lack of population expansion by Tajima's D and Fu's F_S are 0.05 and 0.02, respectively. Bold p-values are consistent with population expansion. See Table S1.5 for a complete list of statistics.

Lineage	Spatial Mismatch SSD	SSD p-value	Tajima's D	Tajima's D p-value	Fu's F_S	Fu's F_S p-value
North	8.20×10^{-4}	**0.23**	1.36	0.94	-24.85	**0.00**
East-North-A1	1.56×10^{-3}	**0.37**	1.19	0.90	-25.12	**0.00**
East-North-A2	4.12×10^{-3}	**0.51**	0.84	0.85	-3.28	0.05
East	2.01×10^{-3}	0.00	0.72	0.83	-23.97	**0.00**
East-South-A	5.01×10^{-3}	**0.38**	1.30	0.94	-5.35	0.03
South	1.73×10^{-3}	**0.15**	1.41	0.93	-24.50	**0.00**
West	1.62×10^{-2}	0.00	1.82	0.97	-24.82	**0.00**
Evolution	4.55×10^{-3}	**0.58**	1.32	0.93	-3.93	0.04
Goddard	1.16×10^{-3}	**0.37**	1.69	0.96	-24.78	**0.00**

* SSD = sum of squared deviations between observed and expected mismatch distributions.

Table 1.6. **Inferred origins of post-Pleistocene colonization events.** Only the four pure lineages of Yosemite NP were considered; admixed lineages were excluded due to low sample sizes. Longitude and latitude are given for inferred location.

Lineage	Longitude	Latitude	q	r_1	r_{10}	r_{100}	d	R^2	p-value
North	-119.5173	38.11973	0.015	0.971	0.772	0.253	0.343	0.600	2.62×10^{-6}
East	-119.2167	37.96289	0.008	0.985	0.869	0.398	0.669	0.273	4.87×10^{-24}
South	-119.5223	37.66362	0.012	0.976	0.804	0.291	0.414	0.237	5.82×10^{-3}
West	-119.6657	37.8162	0.020	0.961	0.710	0.196	0.247	0.540	7.76×10^{-18}

***** Abbreviations: q (strength of colonization, given as regression slope in km^{-1}), r_i (decrease in diversity over 1, 10, 100 km given as $N_e^{founder}/N_e$), d (distance in km over which $N_e^{founder}/N_e$ is 0.99), R^2 (adjusted coefficient of determination for the non-linear model), and p-value for significance of spatial range expansion.

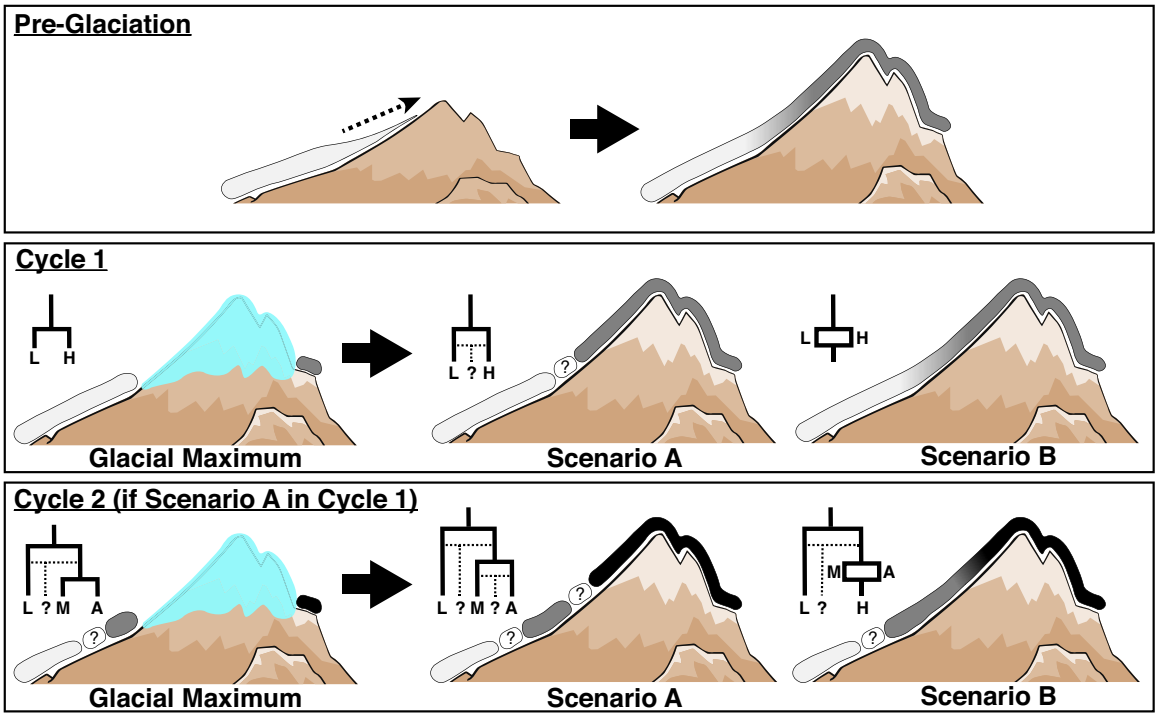


Figure 1.1. The glacial pulse model of alpine speciation
A hypothesized "glacial pulse" mechanism for generating new alpine species and lineages. The profile of a mountain range (e.g., west to east) over time is shown. Prior to Pleistocene glacial cycles, mountains may not have reached modern stature. A hypothetical ancestral species (white) colonizes the young mountain range and eventually shows signs of ecological divergence over an elevational gradient. During the first glacial maximum ("Cycle 1"), vicariant populations experience divergent climates in isolation, due to physiographic differences caused by the mountains (e.g. rain shadow effect). Refugia on one side may resemble higher elevation habitat. Secondary contact during the interglacial either results in reproductive isolation, with or without fusion into a third hybrid lineage (Scenario A), or strong gene flow merges the two populations back together (Scenario B). If Scenario A occurs, low (L) and high (H) species repeat the process during Cycle 2, potentially bisecting the H lineage into montane (M), alpine (A), and fused hybrid lineages. This process may continue recursively, subdividing species into ecologically distinct lineages until ecological or genetic differences are too minor to form permanent barriers (Scenario B).

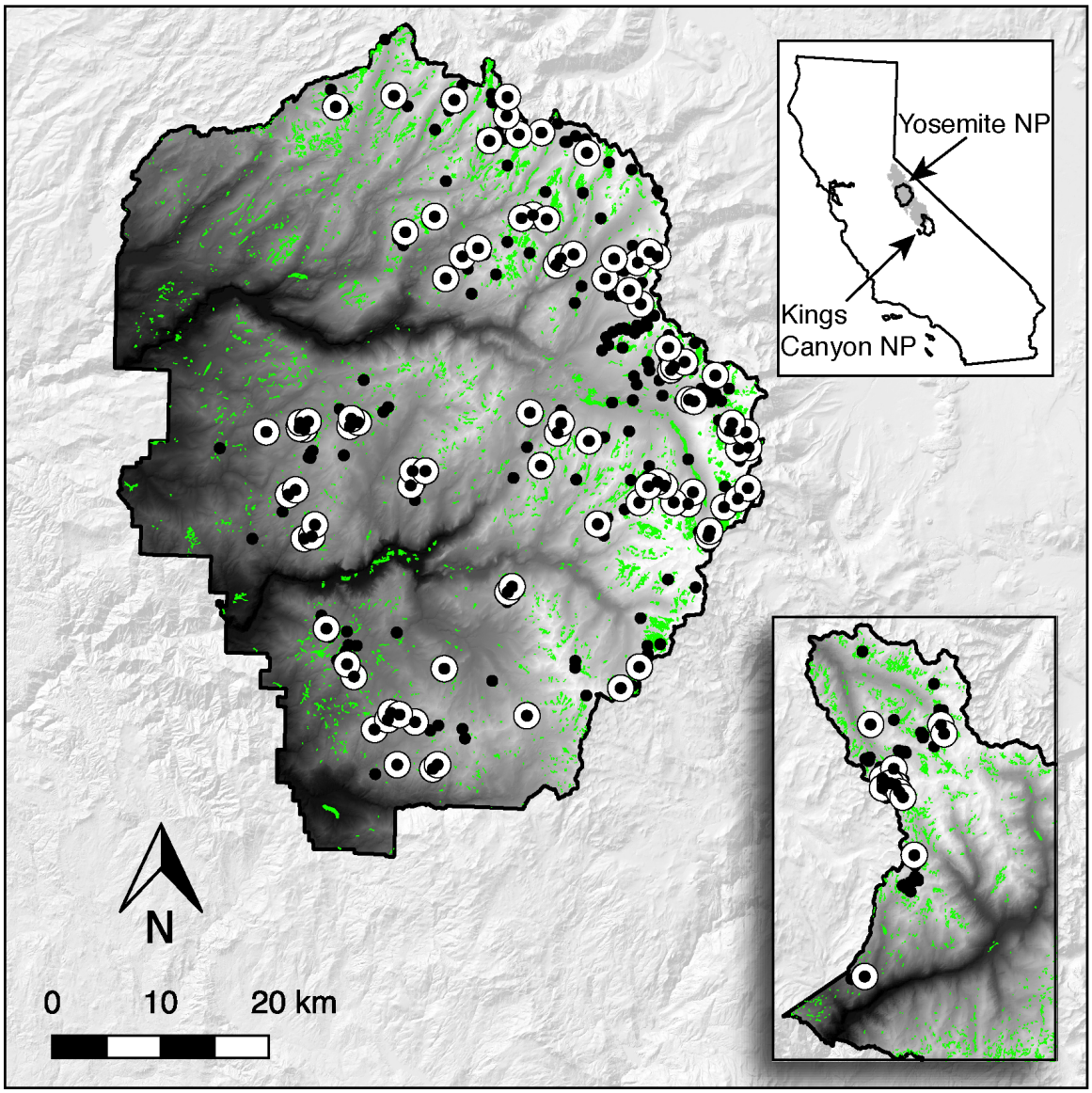


Figure 1.2. **Study region**
Study area in Yosemite National Park (YOSE), CA includes approximately 33% of known Yosemite toad sites, whereas the study area in Kings Canyon National Park (KICA), CA (bottom right inset) includes the southernmost 4% of known Yosemite toad sites. Top right inset shows the range of Yosemite toads in gray, and the boundaries of YOSE and KICA in black. Green polygons are all meadows within the parks. Solid black circles indicate all known Yosemite toad meadows identified between 1905 and the present. White circles indicate the meadows sampled and sequenced in the present study, in YOSE (n=90) and KICA (n=12). Eight samples of *Anaxyrus boreas* and one sample of *A. punctatus* from throughout California are not shown. Random noise is added to locations in order to protect the locations of this threatened species.

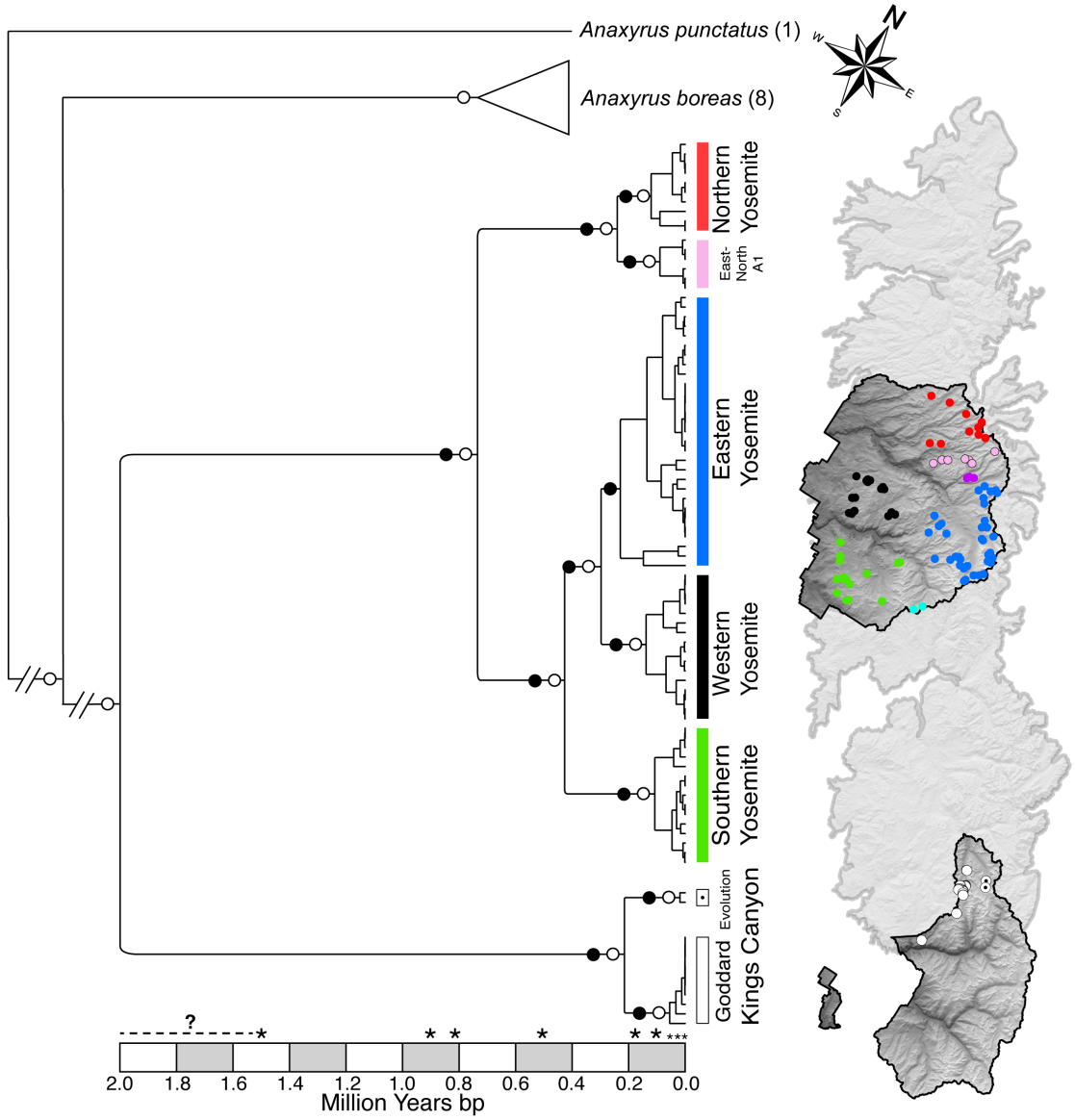

Anaxyrus punctatus (1)
Anaxyrus boreas (8)
Northern Yosemite
East-North A1
Eastern Yosemite
Western Yosemite
Southern Yosemite
Evolution
Goddard Kings Canyon
N
W
E
S
?
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Million Years bp

Figure 1.3. **Phylogenetic reconstruction of Yosemite toad ancestry**
Maximum clade credibility chronogram produced from BEAST, run for 50,000,000 generations using a GTR + Γ model and strict clock. A molecular clock of 9.24×10^{-10} – 1.53×10^{-9} substitutions site^{-1} year^{-1} was used to calibrate node divergence times, following Crawford (2003). A maximum likelihood phylogram from RAxML using 1000 bootstrap replicates and 10 heuristic searches under the same model produced the same ingroup topology for major lineages, with the exception that Y-East was paraphyletic with respect to Y-West (Fig. S1.1). Outgroups were only included for RAxML analysis. Black circles indicate posterior probabilities of 1.0, and open circles indicate bootstrap support values >95 for major clades. Outgroup branches for *Anaxyrus boreas* and *A. punctatus* individuals are broken for clarity. Two of three admixed lineages (purple and cyan circles on map) were excluded from phylogenetic analysis due to recent admixture, although they are still depicted on the map for completeness. Maps of both national parks are overlaid on the distribution of *A. canorus*. Asterisks on the timeline denote ages of known glacial maxima based on Gillespie and Clark (2011).

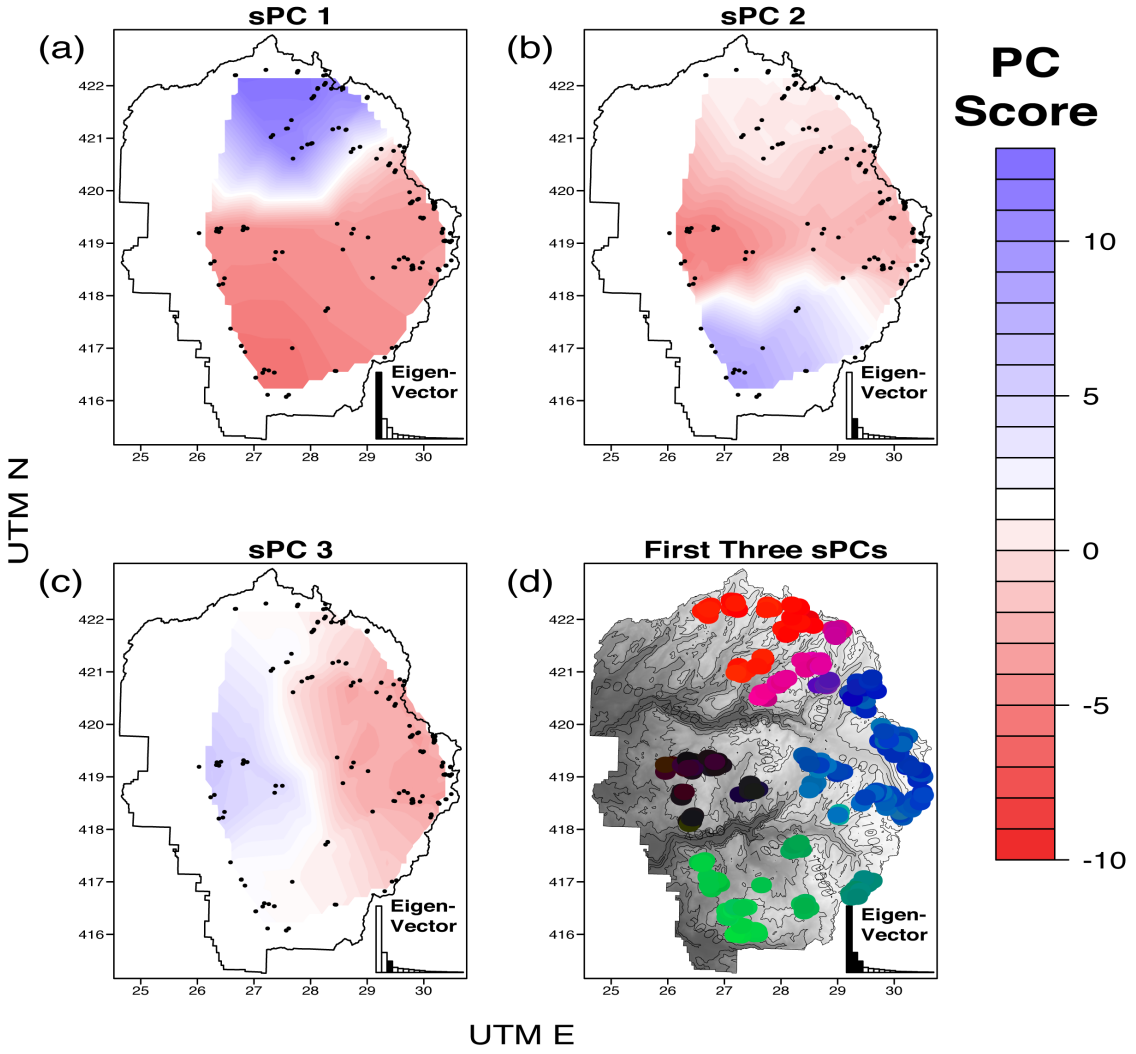


Figure 1.4. **Spatial PCA of Yosemite toad genetic structure in Yosemite NP**
First three panels show the first (a), second (b), and third (c) spatial components contributing to global structure (positive Moran's I values). sPC scores are the product of principal components and Moran's I coefficients, thus they resemble gradients of genetic variance that follow strong spatial patterns. Panel (d) shows all three axes of multivariate spatial genetic structure, with color denoting 3-dimensional scores (red: Y-North, blue: Y-East, black: Y-West, green: Y-South).

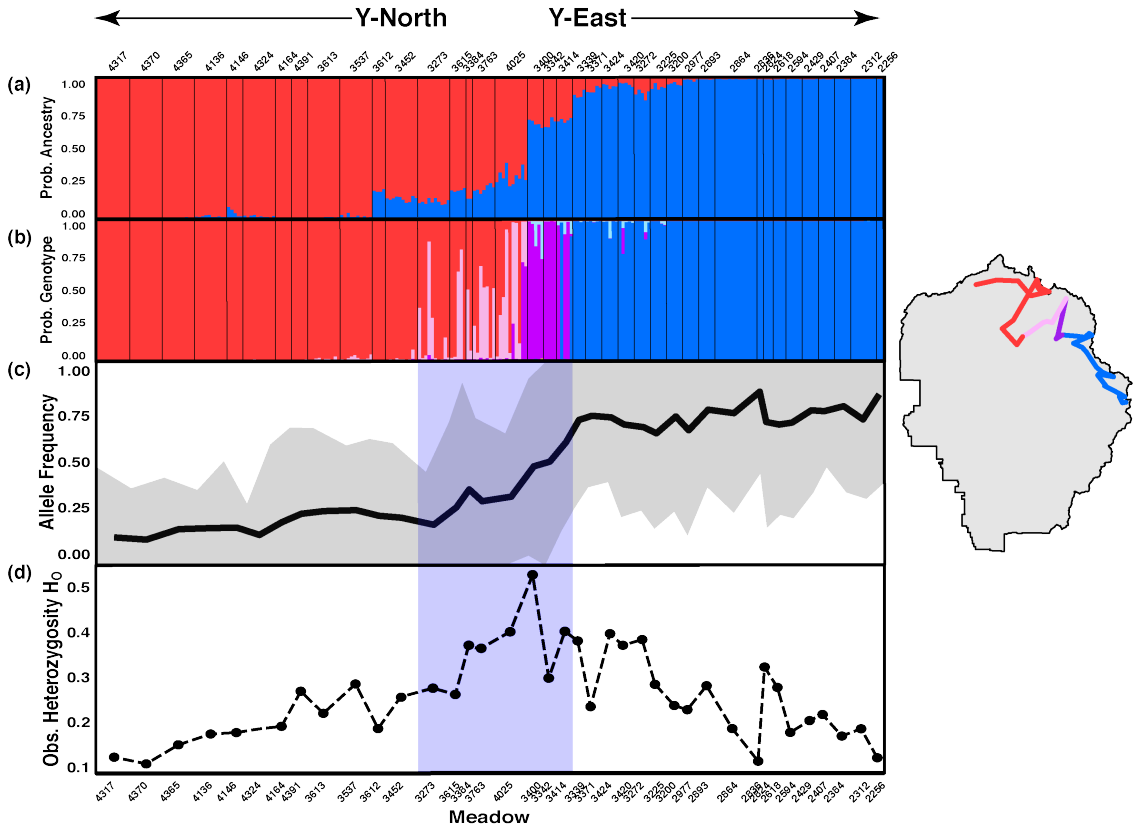


Figure 1.5. **Contact zone dynamics in northern Yosemite NP**
Gradient in ancestry, hybrid genotype classes, allele frequency, and observed heterozygosity following a linear transect between two lineages: Y-North (red) and Y-East (blue), shown on the map. All haplotypic markers are included in (a), whereas only lineage-diagnostic markers (with highest 1% contribution to sPC1) are used in (b-d). (a) STRUCTURE barplot showing probability of ancestry from Y-East and Y-North. K=2 was tested 10 times for 100,000 generations and 10,000 burn-in each, then combined using CLUMPP. (b) NewHybrids barplot showing probability of pure or hybrid genotypes, from the last two generations. Each color represents a hybrid genotype class (P1: red, P2: blue, F2: purple, F1-P1 backcross: pink, F1-P2 backcross: light-blue). (c) Allele frequency gradient across the contact zone. Solid line indicates mean, gray shading indicates 95% confidence interval among loci. (d) Mean observed heterozygosity. Purple box in (c-d) highlights meadows with identified admixture from (b).

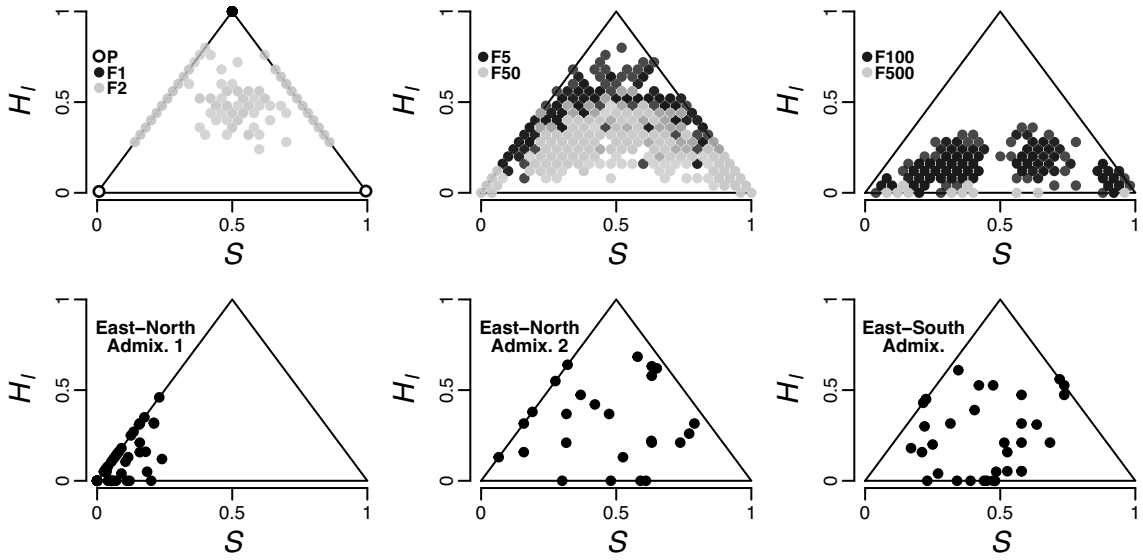


Figure 1.6. **Admixed ancestry and heterozygosity in Yosemite NP**
Estimation of admixture index (S; proportion of ancestry from each lineage) and inter-lineage heterozygosity (H_I; heterozygosity at lineage-diagnostic markers) for three admixed lineages in YOSE, following Fitzpatrick (2012). In each panel, the horizontal axis represents S, and the vertical axis represents H_I. Circles represent individual genotypes. The top row of panels shows simulated individuals with population size of 50, at 25 markers, with no gene flow, and various initial admixture proportions (10% – 90%). The bottom row of panels represents observed admixture scores at East-North-A1, East-North-A2, and East-South-A lineages. On the horizontal axes, from left to right indicates increasing ancestry from Y-East. Pure types (P) are located in the bottom corners, F1 genotypes are in the top corner. Heterozygosity declines from 100% (F1s, black dot in top left panel) as admixed individuals breed in successive generations after initial admixture. Advanced-stage hybrid individuals approach the base (e.g. F500, top right panel), while advanced backcrossed genotypes would be on the sides. Unlike NewHybrids (Figs. 1.5, S1.5), this method can estimate H_I exceeding the F2 generation.

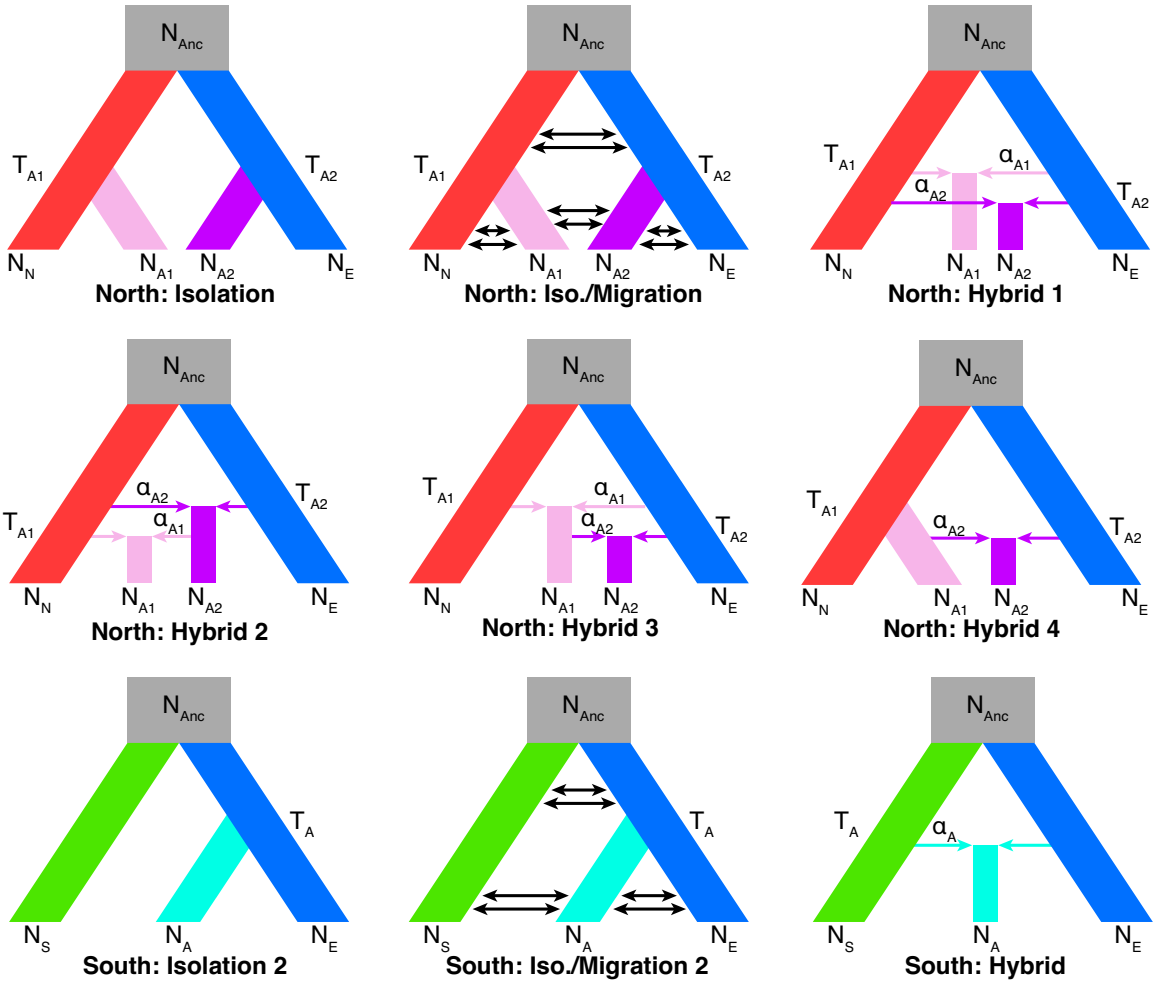


Figure 1.7. **Hypothesized models of admixed lineage origin in Yosemite NP**
A subset of the models tested using fastsimcoal for two admixture zones in Yosemite: East-North and East-South. See Fig. S1.7 for a graphical representation of all 17 models. Parameters include: population sizes for North (N_N), East-North-A1 (N_{A1}), East-North-A2 (N_{A2}), East (N_E), South (N_S), East-South-A (N_A), and ancestral (N_{Anc}) lineages; admixture proportions into East-North-A1 (α_{A1}), East-North-A2 (α_{A2}), and East-South-A (α_A); divergence or admixture times for East-North-A1 (T_{A1}), East-North-A2 (T_{A2}), and East-South-A (T_A) lineages; and asymmetrical migration rates. Migration parameters do not have abbreviations, to save space. Double arrows represent ongoing gene flow, whereas single arrows represent single admixture events.

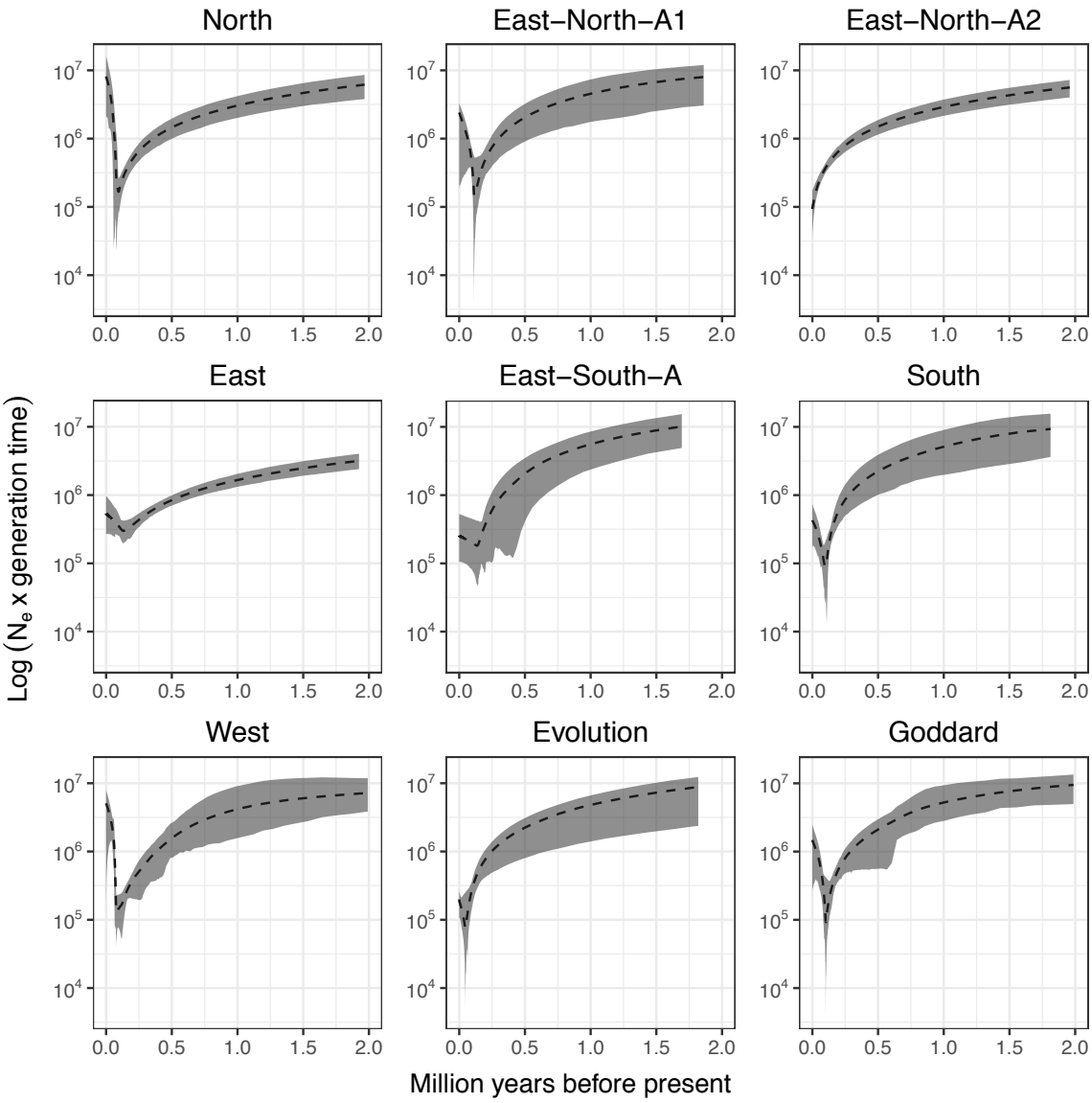


Figure 1.8. **Extended Bayesian skyline plots**
EBSPs estimated using BEAST v2.4.6, using 50 randomly selected RAD loci with 5–6 SNPs. Dashed line shows mean, and gray ribbon contains 95% HPD. Timeline is calibrated using a molecular clock of $9.24 \times 10^{-10} – 1.53 \times 10^{-9}$ substitutions site^{-1} year^{-1}.

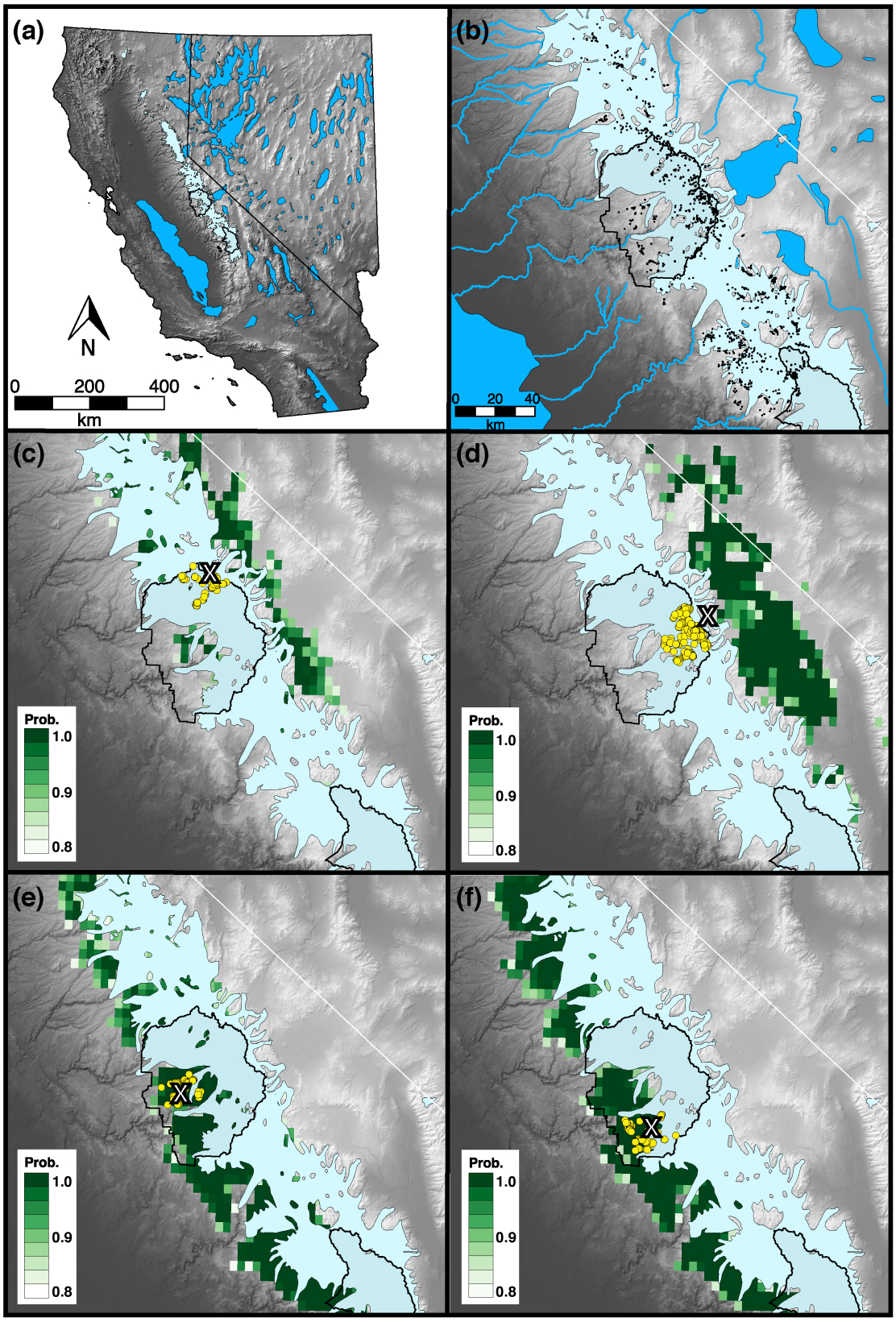

(a)
N
0 200 400
km
(b)
0 20 40
km
(c)
Prob.
1.0
0.9
0.8
(d)
Prob.
1.0
0.9
0.8
(e)
Prob.
1.0
0.9
0.8
(f)
Prob.
1.0
0.9
0.8

Figure 1.9. **Spatial demographic expansion from glacial refugia in Yosemite NP** Location of Pleistocene refugia inferred from climatic niche modeling, and spatial localization of range expansion origins. Panels a–b show the maximum extent of glaciation in the Sierra Nevada, and pluvial lakes from across the period; (b) all known occurrences of *Anaxyrus canorus* are represented by black points. Panels c–f show occurrence data for each of the four pure lineages in Yosemite with yellow circles, niche suitability scores 0.8–1.0 with green pixels, and the inferred origin of spatial expansion with a black 'X'. These panels are as follows: (c) Y-North; (d) Y-East; (e) Y-West; (f) Y-South. Note that origin locations can only be inferred inside each study area, but (c) and (d) suggest locations outside of Yosemite. Interpolations of origins are in Fig. S1.9.

CHAPTER 2: GENOMIC GEOGRAPHY: ISLANDS AND RIVERS OF DIVERGENCE AT SECONDARY CONTACT ZONES

ABSTRACT

Genomes are highly heterogeneous during the early stages of speciation, with small "islands" of DNA appearing to reflect strong adaptive differences, surrounded by vast seas of relative homogeneity. As species begin to diverge, secondary contact zones between them can act as the interface between their diverging genomes, and selectively filter through advantageous hybrid alleles. Such introgression is another important adaptive process, one that allows adaptive mosaics of recombinant DNA to flow like "rivers" from one species into the other. Although genomic islands appear to be associated with reproductive isolation, and genomic rivers form by a different process (adaptive introgression), it is unknown whether islands and rivers are actually different loci. One possibility is that islands and rivers are predictably different loci. This may happen if islands typically include "speciation genes" that evolve by the same mutations, and recombining these mutations lowers hybrid fitness, leading to reproductive isolation. Another possibility is that islands are simply small-effect, polygenic adaptations. In this scenario, numerous genes have the potential to become islands, making the genomic location of islands unpredictable. However, the extreme allele frequency differences at these island loci could fuel more extreme recombinant traits, thus turning small-effect island adaptations into more potent river adaptations. Thus, hybrid advantage from adaptive recombination may supersede any hybrid incompatibility, and predictably convert some islands into rivers. I used three replicate secondary contact zones for the Yosemite toad (*Anaxyrus canorus*), two genomic datasets, and a morphometric dataset to answer the questions: (1) How predictably different are islands and rivers, both in terms

of genomic location and gene function; (2) Are the adaptive genetic trait loci underlying tadpole growth and development reliably islands, rivers, or neither? I found that island and river loci have significant overlap within a contact zone, suggesting that some islands are predictably converted into rivers. However, my gene ontology enrichment analysis showed strong overlap in gene function unique to all island loci, suggesting predictability in overall gene pathways for islands. GWAS outliers for tadpole development included LPIN3, a lipid metabolism gene that is nearly an island for all three contact zones, but also appears to be introgressing (as a river) across one zone. Taken together, my results suggest that a more nuanced view of islands and rivers is needed, where some islands may predictably contribute to reproductive isolation, but others play a vital role in the generation of recombinant adaptation.

INTRODUCTION

Darwin viewed speciation as a process involving differential adaptation to contrasting environments, by either natural or sexual selection (Darwin 1859). However, the neo-Darwinian synthesis placed extreme emphasis on the totality of reproductive isolation between populations in allopatry, regardless of the role selection played (Dobzhansky 1937; Mayr 1942). One solution to these conflicting views came via the "genic species" concept, which imagines incipient species as semipermeable barriers to gene flow, with some adaptively divergent genes playing larger roles in reproductive isolation (Wu 2001; Wu and Ting 2004). Under this view, blocks of coadapted gene complexes become larger as advantageous mutations are clustered together by

hitchhiking selection, and these genomic "islands" of divergence are closely involved in the formation of new species (Rundle and Nosil 2005; Nosil et al. 2009; Nosil and Feder 2012). This more nuanced perspective acknowledges the ongoing hybridization observed at multiple levels of organization (i.e. from lineages to distant species), and shifts the focus away from entire genomes toward genetic loci as the units of speciation (Abbott et al. 2013). Since the paradigm of genic speciation, numerous studies have reported genomic islands of highly differentiated loci surrounded by relatively homogenous genetic backgrounds (e.g. Turner et al. 2005; Ellegren et al. 2012; Malinsky et al. 2015). Although genomic islands were initially conceived to explain primary divergence under parapatric speciation with gene flow, the idea of adaptively diverging species with porous genomes applies equally well to secondary contact between allopatric young species.

If genes (and not populations of organisms) are the units of speciation, a question naturally arises about the numerous genes that continue to recombine in secondary contact zones: what is the consequence of this recombination on adaptive divergence, and the fate of species? One increasingly common pattern found in young species is that hybrid zones generate recombinant diversity that increases hybrid fitness (Arnold and Hodges 1995; Arnold and Martin 2010; Arnold et al. 2012). Although such heterosis wanes in subsequent hybrid generations, what remains are the extreme hybrid phenotypes that result from transgressive segregation (Rieseberg et al. 1999). Such extreme traits likely result from recombination between species, when multiple complementary genes underlie those traits. Normally, a trait may be generated by multiple alleles that oppose each other, but recombination can shuffle those alleles into an order that amplifies or

diminishes the trait value beyond what either species had previously displayed. Transgressive segregation is most common when the two species are more inbred, and more genetically divergent (Rieseberg et al. 1999). If it is adaptive, a transgressive hybrid trait can then flow freely into one or the other species by so-called adaptive introgression, to an extent that is limited by a balance between selection and migration (Arnold and Martin 2009; Rieseberg 2011; reviewed in Hedrick 2013). In contrast to genomic islands, these adaptive recombinant alleles can be thought of as genomic "rivers," because they are new adaptations that flow directionally from hybrid zones into one species. Generally, genomic rivers will flow into whichever genomic and/or environmental background where it confers the highest fitness. Hybrid zones are thus excellent places to study the outcome of ongoing speciation: they are barriers to island loci with species-specific adaptations that may lower hybrid fitness, but they filter through river loci according to levels of recombinant adaptation (Barton and Hewitt 1985; Barton and Bengtsson 1986; Hewitt 1988; Martinsen et al. 2001). Many examples of adaptive introgression are known in plants (Whitney et al. 2006; Castric et al. 2008), and animals (Song et al. 2011; Fraïsse et al. 2014; Norris et al. 2015), including in humans (Hawks and Cochran 2006; Racimo et al. 2015).

Although studies of genomic islands and rivers are widespread, it remains unclear whether islands must be different loci than rivers. In the most extreme scenario (S1), islands cannot become rivers, because islands are composed of "speciation genes" with species-specific adaptations, and recombining them would lead to negative epistasis (Coyne 1992; Orr et al. 2004). The idea is that certain loci, whether related to

reproduction or not, are predisposed to accumulating Bateson–Dobzhansky–Muller (BDM) incompatibilities (Orr and Turelli 2001; Cutter 2012) or chromosomal rearrangements (Kirkpatrick and Barton 2006) that predictably lower hybrid fitness. If genomic islands are predisposed to containing particular genes and mutations, then the same island loci should predictably be found in multiple examples of species divergence. This view of speciation genes has been popular up until very recently (Nosil and Schluter 2011). For example, mutations and chromosomal inversions near loci controlling skeletal armor in three-spine sticklebacks have repeatedly evolved between pairs of freshwater and marine populations (Barrett et al. 2008; Jones et al. 2012). "Magic traits" that are both the subject of divergent selection and cause of non-random mating may add to this predictability (Servedio et al. 2011). For example, the two stickleback phenotypes are associated with body size, which directly leads to assortative mating (Snyder and Dingle 1989; McKinnon et al. 2004). Under this view, island loci should repeatedly become island loci across various contact zones, because they are genomic regions biased toward predictable reproductive incompatibilities. River loci, under this view, should form a distinct class of genes that is non-overlapping with islands.

There are two primary alternatives to this scenario. A second scenario (S2) is that islands emerge in unpredictable genomic locations, because they are typically composed of small-effect trait loci that are involved in polygenic adaptations. There is some evidence that speciation occurs by predictable phenotypic evolution, but unpredictable molecular evolution when the phenotype is polygenic. For example, different genes that all play roles in regulating plumage pigmentation have independently differentiated at

multiple crow contact zones (Poelstra et al. 2014; Vijay et al. 2016). If this pattern is widespread, it could suggest that island loci evolve by parallel evolution. Another explanation is that islands are stochastically influenced by genetic draft, background selection, or drift (Noor and Bennett 2009; Cruickshank and Hahn 2014; Wolf and Ellegren 2017). Regardless of the exact cause, there is an important consequence of island loci containing polygenic adaptations. In this scenario, recombining islands is unlikely to substantially lower hybrid fitness, because each island has a small effect on fitness. However, recombination is more likely to promote extreme hybrid traits for these island loci, because by definition they are found at extreme allele frequencies (Rieseberg et al. 1999). Therefore, in this scenario, adaptive introgression predictably co-opts some islands into rivers, if the recombinant fitness of these rivers surpasses the fitness of islands they are replacing (Rieseberg et al. 1999; Martinsen et al. 2001; Jiggins et al. 2008). Thus, weakly selected island loci may act as fuel for adaptive introgression, which may then erode and homogenize those islands (Clarkson et al. 2014; Bay and Ruegg 2017). For example, an insecticide-resistance mutation that was originally a genomic island later introgressed between *Anopheles* mosquito species, with no perceptible effect on reproductive isolation or hybrid fitness (Clarkson et al. 2014). The expectation under this scenario would be that some loci are both islands and rivers in a given contact zone, but few loci are islands (or rivers) across different contact zones.

In a third possible scenario (S3), neither islands nor rivers predictably emerge in the same genetic loci, simply because every locus reflects unique histories of interacting evolutionary forces such as mutation, recombination, and selection, the so-called "$n = 1$

constraint" (Beaumont and Balding 2004; Buerkle et al. 2011). This situation could also arise if adaptation—both divergent and introgressive adaptation—is primarily through small-effect polygenes (Rockman 2012). This scenario would predict loci to sort into islands and rivers with no discernible pattern. Of course, S1–S3 are not mutually exclusive or exhaustive of all possible scenarios, but they are meant to explain the three most salient patterns found in nature.

I used a series of recently described contact zones between four lineages of the Yosemite toad (*Anaxyrus canorus*; Chapter 1) to address which of these scenarios best describes incipient speciation. In this system, three replicate zones of secondary contact exist for lineages that diverged 214–732 ka. Previous work on this system also found that inter-lineage admixture was extensive enough to promote lineage "fusion" for two of these contact zones, raising the possibility that adaptive introgression is ongoing. Using evolutionary replicates can be a powerful way to address alternative hypotheses, particularly when the focal question involves how predictable a process is in nature (Losos 1992; Hoekstra 2006; Hohenlohe et al. 2010). I tested these hypotheses by first identifying markers as putative genomic islands, then identifying putative genomic rivers, and finally assessing whether there is any predictability to genic patterns across three replicate contact zones. Since genetic predictability can manifest in DNA loci, or in broader protein function (Reid et al. 2000; Stern and Orgogozo 2008, 2009), I also assembled and annotated a de novo larval transcriptome to address whether genetic patterns are mirrored by gene functional patterns.

Additionally, I asked whether genome-wide patterns of island and river predictability applied to genetic trait loci of adaptive importance for Yosemite toad tadpoles. Specifically, I asked: are loci underlying growth and development reliably islands, rivers, or neither? The Yosemite toad is ostensibly under greatest future threat from a warming climate, because its habit of laying eggs in shallow ephemeral ponds predisposes the species to mass mortality when ponds desiccate (US Fish & Wildlife Service 2014). However, different lineages of Yosemite toads may have adapted to different desiccation levels. Tadpole growth and development are generally seen as a life history tradeoff, where faster development comes at the expense of smaller size at metamorphosis (for a comparative analysis, see Richter-Boix et al. 2011).
Species that specialize on shorter hydroperiod ponds do not consistently have different levels of plasticity for growth or development rates (Brady and Griffiths 2000; Leips et al. 2000; Richter-Boix et al. 2006, 2011; Lind and Johansson 2007). However, these species do generally have a heritable basis for developing faster, growing more slowly, and metamorphosing at a smaller size (Morey and Reznick 2004; Richter-Boix et al. 2011). Thus, it is reasonable to expect adaptive genetic differences between Yosemite toad lineages experiencing different levels of desiccation.

I tested scenarios S1–S3 using a ddRADseq dataset, a reference transcriptome, and a tadpole morphometric dataset collected from Yosemite National Park (YOSE). My analytical workflow used several divergence estimators to ensure that putative islands were not artifacts of any measure used (Cruickshank and Hahn 2014), and used Bayesian genomic cline analysis (Gompert and Alex 2011) to identify likely rivers. Understanding

how adaptation and speciation proceed for this federally threatened species is essential, because long-term population declines and susceptibility to climate change make its future uncertain (US Fish & Wildlife Service 2014; Brown et al. 2015). My results will also bear significantly on the nature of speciation generally, by directly testing alternative scenarios of genic speciation in a comparative framework.

MATERIALS AND METHODS

Molecular methods, ddRAD sequencing, and bioinformatics

I used the previously described ddRADseq haplotype dataset (Chapter 1) for all analyses of genetic differentiation and introgression. Details about the library preparation, sequencing, and bioinformatic parameters used to identify variable loci are described therein. Briefly, a total of 535 individual Yosemite toad (*Anaxyrus canorus*) tadpoles sampled from 90 meadows across Yosemite National Park (YOSE) were sequenced using 2×100 bp sequencing on an Illumina HiSeq 2500. Stringent quality thresholds were used, and several measures were used in the Stacks v1.19 (Catchen et al. 2011, 2013) pipeline to avoid recovering markers from paralogous sequences. Wherever quality and coverage thresholds permitted, both reads were concatenated into a ~200 bp sequence, otherwise the first read (~100 bp) was used. Instead of calling one SNP per locus, I used the sequence as the allele for that locus. This gave the dataset an enhanced amount of intra-locus information, and potentially made it more sensitive to intra-locus recombination that can occur within admixture zones. I used a coverage threshold of 10 reads per genotyped individual. SNPs with minor allele frequency (MAF) <0.05 were not called,

and loci were filtered from the dataset if absent from >25% of individuals. Although tadpoles were sampled from different pools wherever possible, Colony v2.0.6.4 (Jones and Wang 2010) was used to remove siblings from the dataset.

RNA sequencing, transcriptome assembly, and annotation

A full de novo transcriptome was constructed to match RAD loci to genes and gene functional information (gene ontologies) where available. For full details of methods, see Appendix C. Briefly, three tadpoles were collected from throughout YOSE, libraries were prepared, and sequenced using 100 bp paired-end reads on an Illumina HiSeq 2500. A complete de novo transcriptome was assembled using standard methods in Trinity r2014-02-04 (Grabherr et al. 2011; Haas et al. 2013). I annotated the de novo transcriptome using the Trinotate suite included in Trinity, using standard methods (https://trinotate.github.io/). Bi-allelic SNPs were called across the de novo transcriptome using the GATK v4.0.1.2 best practices for RNAseq (McKenna et al. 2010). SNPs were then annotated with synonymous/non-synonymous effects and predicted protein changes in two steps: (1) likely ORFs ≥ 100 amino acids were reconstructed using TransDecoder v5.0.2 (Haas et al. 2013); (2) SNP effects were annotated using Ensembl VEP v92.1 (McLaren et al. 2016). I identified ddRADseq markers that fall within coding genes by performing a blast-n nucleotide search on full RAD sequences, with the transcriptome as the database. To minimize the possibility of matching paralogs, I only allowed ungapped alignments with no more than three mismatches beyond the known RAD SNPs, with an e-value cutoff of 1×10^{-6}.

Defining lineages and admixture zones

I used lineage boundaries from previous phylogenetic analyses of the same individuals (i.e. RAxML, BEAST) to identify putatively "pure" and "admixed" individuals. Genomic cline methods are able to identify introgressed loci based on their departure from genome-wise patterns of admixture, but they require a definition of pure (i.e. parental) allele frequencies. A total of four pure lineages (North, East, South, West) exist in YOSE, with admixture definitely found at two contact zones: East-North (hereafter "EN"), and East-South ("ES"). Given the spatial arrangement of these populations, and their phylogenetic relationships, another contact zone (East-West; "EW") was also plausible. To screen individuals as putatively admixed, I used a four-step process: (1) I ran STRUCTURE v2.3.4 (Pritchard et al. 2000) five times for 5.0×10^5 steps and 1.0×10^5 burn-in with K=2 ancestral populations to find individuals with dual ancestry; (2) I used NewHybrids v1.1 (Anderson and Thompson 2002) to screen for signal of recent admixture; (3) I used the HIest package (Fitzpatrick 2012) in R (R Core Team 2018) to screen for more advanced (>2 generations) admixture; (4) I performed a principal components analysis on genotypes to assess whether putatively admixed individuals identified in steps 1–3 clustered separately and intermediate from pure individuals. Any meadows lacking individuals with evidence of admixture were thereafter considered "pure".

Finding genomic islands of divergence

Previous studies (e.g. Hofer et al. 2012; Marques et al. 2016) that identified genomic islands of incipient speciation have often done so in a two-step process: (1) markers with extreme patterns of divergence are flagged; (2) their proximity along chromosomes can add evidence to the contention that individual SNPs are involved in speciation. I implemented a modified form of this workflow due to recent criticisms of F_{ST}-based outlier designation (described below), and because I used anonymous markers without a reference genome.

Typically, divergence metrics (usually F_{ST}) and associated p-values are calculated separately for each marker, and then compared to a null distribution, e.g. expected values based on a hierarchical island model. This step identifies extremely differentiated SNPs, which may have been driven by directional selection. However, if outliers are designated solely using F_{ST}, they will include markers with lower genomic diversity that may have nothing to do with directional selection (Noor and Bennett 2009; Cruickshank and Hahn 2014; Wolf and Ellegren 2017). For example, processes such as background selection or genetic draft can reduce intra-lineage diversity by reducing recombination, which can elevate F_{ST} values. This is because F_{ST} estimates inter-lineage divergence relative to intra-lineage diversity, and hence is a "relative" measure. True genomic islands should have high relative, absolute, and cladistic divergence.

I addressed this potential bias by defining genomic islands with three metrics. First, I performed a hierarchical molecular analysis of variance (AMOVA; Excoffier et al. 1992) on each marker at level of lineages, while accounting for variance at the level of

meadows and individuals. I used the amova function in the R package pegas (Paradis 2010), and extracted Φ_{ST} representing between-lineage variance using the getPhi function. Second, I calculated D_{XY} (Nei 1987), an absolute measure of sequence divergence that quantifies the average number of nucleotide differences among lineages. D_{XY} only reflects differences that accumulated since the lineages split (as well as ancestral polymorphism), and thus is unaffected by levels of intra-lineage diversity. I calculated D_{XY} using custom R scripts. Third, I calculated Slatkin and Maddison's 's' statistic (Slatkin and Maddison 1989), the number of parsimony steps required for each gene tree to be congruent with the "true" lineage tree using DendroPy v4.4.0 (Sukumaran and Holder 2010). This cladistic measure is sensitive to lowered gene flow or increased selection, which should both cause gene trees to be more congruent with the lineage tree. Input gene trees were generated using the BIONJ neighbor joining algorithm in the ape package (Gascuel 1997; Paradis et al. 2004), with the Kimura 2-parameter model of evolution (Kimura 1980). For all three metrics, markers were considered statistical outliers if they: (a) exceeded the 95th quantile of marker-wide values, and (b) rejected the null hypothesis of no differentiation, based on 1000 permutations at the α=0.05 level. I refer to markers that passed two out of three tests as outliers, and those passing all three tests as "stringent" outliers. Outliers (those passing at least two tests) are hereafter referred to as genomic islands. All computations were performed using custom R scripts.

If a reference genome were available, the next step would typically use the physical arrangement of outlier SNPs along chromosomes (e.g. clustered or random) to add evidence that directional selection is a primary force in creating them. Although a

reference genome was unavailable, evidence of selection was inferred by the repeatability of patterns across replicate contact zones, and by matching RAD loci to genes and gene functional information from the de novo transcriptome (discussed below).

Finding genomic rivers using genomic cline analysis

I estimated locus-specific genomic clines for each contact zone using bgc v1.03 (Gompert and Alex 2011; Gompert and Buerkle 2012). A genomic cline is a function of how, for each locus, the proportion of alleles shifts along an average genomic hybridization gradient (between pure lineages 'A' and 'B'). Locus-specific patterns of ancestry (ϕ_{ih}) can differ from the genome-wide average (h, i.e. hybrid index), if a locus introgresses directionally into one lineage more or less than average. Two parameters are used to describe this "excess ancestry": cline center (α_i) and cline steepness (β_i):

$$\phi_{ih} = h + 2(h + h^2)(\alpha_i + \beta_i(2h - 1)) \quad (2.1)$$

The cline center parameter α_i indicates increases (positive values) or decreases (negative values) in the "extent" of ancestry from one particular lineage, for one marker relative to genome-wide expectations. The cline steepness parameter β_i indicates increases (positive values) or decreases (negative values) in the "rate" of introgression from one particular lineage, representing levels of pairwise linkage disequilibrium for that marker compared to all others. Simulations have shown that selection against hybrids that leads to reproductive isolation, such as underdominance or pairwise epistasis, can impact either parameter (α_i or β_i) (Gompert and Alex 2011; Gompert and Buerkle 2012; Gompert et al. 2012). However, adaptive introgression that favors homozygous genotypes by

directional selection should only impact α_i (Gompert and Alex 2011). Therefore, I chose to interpret extreme α_i values as evidence for adaptive introgression.

The bgc method uses a Bayesian framework to estimate the probability of an individual with hybrid index h inheriting a gene copy at locus i from the East lineage (defined here as ϕ_{ih}). The East lineage is chosen for convenience, since all contact zones contain it; the probability for a contrasting lineage is defined by $1 - \phi_{ih}$. I ran models 3× each for 5.0×10^5 steps and 1.0×10^5 burn-in to check for convergence, sampling every 20th step with default settings, except that MCMC tuning parameters were increased 2× to increase mixing of the chain. If a marker had 95% equal-tailed posterior probability credible intervals of α_i that did not include the genome-wide average value (zero), it was designated a genomic river.

Estimating genome-wide and outlier-specific migration

Given the recent controversy over whether genomic islands truly represent regions that are barriers to migration, I estimated rates and symmetries of migration. Once island and river loci were identified, I used migrate-n (Beerli and Felsenstein 2001) to compare rates and direction of gene flow for each marker class (all markers, islands, rivers with positive α, rivers with negative α), and for each contact zone. To reduce the number of parameters, each contact zone was classified into three populations: P1, P2, and admixed, and modeled under a stepping-stone scenario (migration of P1 and P2 through the admixed population). Each model was run 10× each for 2×10^6 steps with 2×10^4 burn-in, using four MCMC chains with static heating.

Comparative analyses

I tested the null hypothesis of no association for genetic loci among the different categories of outlier locus (six total—islands vs. rivers, across three contact zones) in several ways. First, I used pairwise Fisher's exact tests to test for locus overlap between each pair of groups, given the number of loci examined and a p-value correction for multiple tests. I implemented this using the GeneOverlap package in R (https://github.com/shenlab-sinai).

Second, I tabulated gene ontology terms for each outlier locus where that information was available, and performed a Cochran-Mantel-Haenszel (CMH) chi-squared test for count data (Agresti 2002) to determine whether gene functional categories are conditionally independent in each stratum. Gene ontologies (GO) categorize gene products based on biological processes, cellular components and molecular functions. I used Level 2 of the GO term hierarchy, which gives highly specific categories. First I tested for independence of GO terms by islands and rivers while accounting for contact zone, and then I tested for independence of GO terms by contact zone, while accounting for marker type. In order to understand which specific GO terms are sorting by contact zone and marker type, I also performed a hierarchical clustering analysis on the level of GO enrichment or depletion, using the hclust package in R. Enrichment was calculated using two-tailed Fisher's exact tests on counts of each GO term compared to expected values by group. I multiplied the sign of the odds ratio by the $-log_{10}(p-value)$ to get level of enrichment (positive) or depletion (negative) (Sánchez

et al. 2007; Janoušek et al. 2015). I removed GO terms that were only enriched by a single group (singletons; e.g. only EN Islands) before clustering them.

Finally, I performed another pair of CMH tests (as described above) to test whether severity of SNP effects—e.g. missense mutations—differs by either contact zone or marker type. Given the low proportion of RAD markers that successfully matched to gene annotations, I found it necessary to slightly lower the threshold for islands and rivers for tests that relied on gene or SNP annotations. To accomplish this, I used one-tailed and two-tailed 95% confidence intervals for islands and rivers, respectively.

Loci related to tadpole growth and development: islands or rivers?

I opportunistically recorded the total length and larval stage—field categories of Gosner (1960) stages—for 1725 tadpoles from 97 meadows throughout YOSE during the years 2012–2013. I used these data to conduct a genome-wide association study (GWAS), to identify SNPs likely involved in larval growth and development. I first modeled overall growth-by-development using polynomial regression, where growth was an nth degree polynomial of developmental stage. I tested between one and six polynomial terms, and compared model fit using Bayesian Information Criterion (BIC) with the compareLM function in the rcompanion package, as well as with an ANOVA. Tadpoles were only measured once, so I treated deviation from the overall growth-developmental curve as evidence for a tadpole's phenotypic deviation; residuals for 535 genotyped individuals were modeled against 1302 SNP loci, and a generalized linear model (GLM) with Gaussian link function was fit to each SNP following Reed et al.

(2015). Candidate loci were designated with the p-value cutoff of $-log_{10}(5\times10^{-6})$ that is normally used in GWAS studies. Spatial genetic patterns for candidate loci were visualized using the adegenet package (Jombart 2008).

RESULTS

Transcriptome assembly quality

The transcriptome assembly contained 285,751 components (putative "genes"), with 50% of assembled nucleotides residing in transcripts 1,571 bp or longer (defined as the N50). Since the raw N50 value can be biased by an overabundance of longer isoforms, I also calculated the N50 for the single longest isoform per component, and found that to be 657 bp. Overall, 91.3% of RNAseq reads were represented by the transcriptome assembly, which surpassed the baseline expectation of 80% (http://trinityrnaseq.github.io/). This indicated a low proportion of low quality or coverage reads in the dataset. BUSCO recovered 96.2% of the single-copy tetrapod orthologs expected for a tetrapod transcriptome (defined as 3,950 genes present in one copy for 90% of 55 chosen tetrapod species). Only 4.3% of these were fragmented, meaning very few incompletely assembled transcripts. A total of 538,806 biallelic SNPs were found across all three samples in the transcriptome, and this number was reduced to 517,237 after applying quality filters during the joint genotyping process. Out of 1485 unique RAD sequences that passed quality thresholds in the three contact zones, 653 successfully matched transcripts. This relatively high success (44%) is expected due to the high GC content of the RAD 8-cutter used; SbfI (5' CCTGCA*GG / GG*ACGTCC

3') has been found to have bias toward gene-rich regions in other vertebrate studies (Amores et al. 2011; Everett et al. 2012; Gonen et al. 2014).

Lineage boundaries and extent of admixture zones

STRUCTURE and NewHybrids analyses revealed differing patterns for each of the three secondary contact zones, and HIest/PCA suggested differences in overall admixture levels (Figs. S2.1–S2.3). The EN contact zone—also the oldest contact zone—was the most geographically expansive of all three contact zones, and showed an even distribution of hybrid index (h) values (Fig S2.1). Although PCA showed hybrid genotypes intermediate between pure genotypes in all contact zones, EN hybrids were particularly intermixed with pure genotypes. HIest and PCA showed a mix of putatively recent and advanced hybrids clustering together, supporting an overall picture of recent and widespread admixture in the EN contact zone. More individuals (n=105) and meadows (n=23) were defined as admixed for the EN contact zone than for the ES (n=74 individuals, n=14 meadows) and EW (n=58 individuals, n=13 meadows) contact zones. These two contact zones showed the opposite pattern, namely narrower zones, with more distinct clustering of hybrid genotypes (Figs S2.2–S2.3). One possible reason for this discrepancy is the difference in contact zone type: these two younger contact zones are low-high-elevation contrasts, whereas the EN contact zone is a high-high contrast (see Chapter 1). However, with a paucity of contact zones to study, it cannot be known for certain whether climatic divergence is the primary force in shaping the dynamics of the EN, ES, and EW contact zones.

Genomic islands: divergence outliers

I initially removed loci from each contact zone that were invariant in the pure meadows, fell below a MAF of 0.05, or had zone-specific missing data levels >0.5. After applying these thresholds, I retained 1126 EN loci, 1138 ES loci, and 925 EW loci (the total number of unique loci across all zones was 1485). The same locus sets were used in all subsequent analyses. Genomic islands were attained in numbers roughly proportional to these overall locus counts: EN (n=47), ES (n=49), and EW (n=35). Values of hierarchical Φ_{ST} and D_{XY} were highly correlated, especially closer to the upper tails of the distributions (Fig. 2.1). The cladistic measure of divergence $'s'$ had a more nuanced relationship with hierarchical Φ_{ST} and D_{XY} (Figs S2.4–S2.6). However, neighbor joining trees based on concatenated genomic islands were strikingly effective at generating the "correct" lineage phylogeny; in comparison, an equivalent number of loci chosen randomly from the interquartile range of D_{XY} values did not recover the correct topology (Fig. 2.1).

Genomic rivers: introgression outliers

Log likelihood values and parameters converged across replicate bgc runs, and replicate α and β estimates were highly correlated (ρ of 0.76–0.93), so all chains were combined into a single estimate (Figs S2.7–S2.8). Estimates of α showed an abundance of outliers in the EN contact zone (n=193; Fig. 2.2), supporting the pattern of higher admixture found earlier for that contact zone (Fig S2.1). The other two contact zones had

comparatively few α outliers (n=47 for ES; n=15 for EW). The majority of these α outliers were positive, meaning clines are shifted toward East, and more introgression from the East compared to the reverse direction. Overall, the range of α values estimated was larger and more even-tailed for EN (min = -0.62, max = 0.71) compared with the ES (min = -0.37, max = 0.65) and EW (min = -0.38, max = 0.54) ranges.

The parameter β ranged from -0.5 to 0.52, but the 95% CI overlapped zero for all loci. As noted earlier, β is more associated with forms of selection such as underdominance or pairwise epistasis that may confer reproductive isolation, and hence I solely used α to identify introgression outliers. However, it is interesting to note that α and β had a strong positive correlation in the two younger contact zones (ρ of 0.25–0.30), and a small but significantly negative correlation for the EN contact zone (ρ of -0.11), indicating the possibility that divergence and introgression have a different dynamic in that contact zone (Figs. S2.9–S2.11). D_{XY} and β were all positively correlated, as expected (ρ of 0.15–0.26). The parameter estimates of α and β are likely accurate because they were repeatable across replicate bgc runs, and because all genomic cline estimates were based on sufficient coverage of hybrid index values (Fig 2.2).

Direction of migration for islands versus rivers

Migrate-n results generally supported the overall expectations (Fig. 2.3). Genomic islands as a group were migrational sinks, as would be expected if those markers are somewhat deleterious for hybrids possessing them. In contrast, genomic rivers (divided into negative and positive α values) had asymmetric migration rates into one lineage. For

example, positive α rivers (which contain excess East lineage ancestry) showed highest migration between the East lineage and the admixture zone, suggesting introgression from the East lineage. In the negative rivers, where introgression is expected from the other (non-East) lineage, patterns were slightly more nuanced. One river had the same pattern (i.e. EN had elevated North-admixed migration), and the other two had net migration toward East. Collectively, rivers in each admixture zone appeared to act as migration sources, with net migration radiating away from the admixture zone.

Results of comparative analyses

Within each contact zone, I found significant overlap between the composition of loci for islands and rivers (Fig. 2.4). Pairwise Fisher's exact tests found odds ratios of 6.51 for the EN contact zone ($p = 6.0 \times 10^{8}$), 16.90 for the ES contact zone ($p = 6.0 \times 10^{10}$), and 23.77 for the EW contact zone ($p = 7.0 \times 10^{5}$), after correcting p-values with the Benjamin-Hochberg method. No such pattern of significant locus overlap was found for islands as a group, or rivers as a group. No other group combinations had significant overlap, given their respective locus counts, and no locus was unique to islands or rivers only. This suggests that within a particular contact zone, highly introgressed loci may derive from highly diverged loci.

In contrast, when considering only gene functional categories, a high specificity was found for islands compared with rivers. The CMH chi-squared test for independence of marker class (i.e. islands vs. rivers) across GO categories was significant (M^2=360.31, df=260, $p=3.71\times10^{-5}$), when accounting for differences across contact zones. I used a

Woolf test to check that all odds ratios were homogenous across strata, to make sure the CMH test was appropriate; I found no 3-way association of odds ratios across strata (χ^2=0.007, df=2, p=0.9965), indicating the CMH assumption was not violated. However, post-hoc tests showed that no island-river pair was significant within each contact zone (EN: p=0.93, ES: p=1.0, EW: 0.06). This is likely due to low counts in pairwise comparisons (mean count: 0.56 per cell). The opposite pattern was not found; namely, a CMH test for independence of contact zones while accounting for marker class was insignificant (M^2=389.93, df=520, p=1.0).

This pattern of functional difference between islands and rivers was strongly supported by hierarchical clustering of GO terms (Fig. 2.5). The majority of GO terms that were enriched in more than one group were enriched for all three sets of island loci. Although many GO terms were enriched for all islands and depleted from rivers, low sample sizes prevented the odds ratios for these depleted river terms from being significantly different than zero. There was less evidence of functional (GO term) overlap between islands and rivers in a given contact zone, however some GO terms did conform to this pattern (Fig. 2.5). In total, 61 specific GO terms differentiated islands from rivers, whereas 36 specific GO terms were unique to islands and rivers of a specific contact zone. Of this latter category, 10, 23, and three terms were unique to the EN, ES, and EW contact zones, respectively (Table S2.1). These patterns were based entirely on highly specific (Level 2) GO term categories. Sorted into broader GO categories ("biological process", "cellular component", "molecular function"), the functional profiles of "island-only," "EN-only," "ES-only," and "EW-only" were distinct (CMH test; M^2= 25.063,

df=6, p=0.0003; Table 2.1). However, when these latter three groups were lumped into a "zone-only" grouping, there was no significant difference in functional profile between "island-only" and "zone-only" (CMH test; M^2=4.069, df=2, p=0.1308). Overall, this suggests that island loci are functionally the most distinct grouping, yet each contact zone has island/river loci with unique functional patterns.

I also tested whether type of SNP (i.e. the effect of that SNP on gene function) is characteristic of marker class, contact zone, or neither. I used the SNP effect categories defined by VEP: low (no protein change), moderate (e.g. missense), high (e.g. nonsense), and modifier (non-coding, e.g. 3' UTR variant). Using a CMH chi-squared test for independence, I found a significant difference in counts of SNP type by marker class (M^2=10.363, df=3, p=0.0157), and also by contact zone (M^2= 19.702, df=6, p= 0.0031). Woolf tests showed that odds ratios were homogenous across strata for both marker class (χ^2=0.243, df=2, p=0.8857) and contact zone (χ^2=0.024, df=1, p=0.8763), in line with CMH assumptions. To pinpoint which SNP types were most important in differentiating the three contact zones, I pooled the data across both marker types and performed a standard Pearson's chi-squared test. The residuals from the chi-square test show that a different SNP type predominates in each contact zone: moderate (EN), non-coding (ES), and low (EW). To pinpoint which SNP types were most important in differentiating the two marker types, I pooled the data across all three contact zones and performed a standard Pearson's chi-squared test. The residuals from the chi-square test show that differences between islands and rivers were driven almost entirely by non-coding (possibly regulatory) and low-effect SNPs.

Relation of islands and rivers to tadpole phenotype

The best model of tadpole length by stage was a second-order polynomial curve, as determined by lowest BIC score (R^2_{adj} = 0.481, F(2,1702)=791.2, p<0.001; Table S2.2; Fig. 2.7a). Both stage (p<0.001) and stage^2 (p<0.001) were significant variables in the model, and a Q-Q plot showed the residuals to be normally distributed. The GWAS between SNP markers and growth-development residuals found six RAD loci out of 1302 with $-log_{10}(p)$ greater than the candidate threshold of $-log_{10}(5\times10^{-6})$ (Fig. 2.7b). A spatial PCA based upon these six loci found 96.7% of the spatial genetic variance in the first (71.4%) and second (25.3%) spatial principal components (sPCs). The first sPC is divided sharply at the EN contact zone, while the second sPC shows a pattern of East ancestry that has been spread to several meadows in other lineages, particularly the West but also the North lineage (Fig. 2.7c-d).

None of the six loci were islands or rivers strictly speaking, based on my originally defined thresholds. However, I ranked them by quantiles of D_{XY} and α to assess whether they had potential to follow one pattern or the other (Table 2.2). One locus had been removed from island/river analysis due to missing data in key meadows. The other five were above the 90th quantile for islands, above the 90th quantile for rivers, or below the 10th quantile for rivers in at least one contact zone. Two loci were nearly rivers in a single contact zone, one locus was nearly an island in a single contact zone, another locus was nearly an island for two contact zones, and the last locus (C2952) was nearly an island for all three contact zones, and a river for the EW contact zone (Table

2.2). The pattern of introgression from East to West that I found for the six loci (Fig. 2.7d) fits with the extreme negative α value for C2952 in the EW contact zone. C2952 was the only locus that successfully received gene annotation information, and was identified as the phosphatidate phosphatase (LPIN3) gene. This gene produces an enzyme that plays a crucial role in multiple lipid metabolic pathways, by catalyzing (regulating) the conversion of triacylglycerols to phospholipids, and vice versa.

DISCUSSION

In this study, I discovered genomic islands of extreme divergence for Yosemite toads, i.e. loci that segregate by lineage history much more than expected given the genome-wide average (Fig. 2.1). Concomitantly, I found genomic rivers that reflect a history of backcrossing directionally from admixture zones into one of the pure lineages, i.e. introgression outliers (Fig. 2.2). I used replicated contact zones in YOSE to ask whether these islands and rivers were consistent with patterns of parallel evolution for genes or gene functions. Specifically, I was interested in understanding which of three scenarios best explained the data:

S1. Islands and rivers represent generally non-overlapping adaptive processes. Divergent adaptation predictably leads to reproductive isolation for "speciation" loci, whereas adaptive introgression predictably leads to hybrid advantage for "hybridization" loci.

S2. Islands reflect divergent adaptation that is not predictable, and hence they are not biased toward reproductive incompatibilities. Selection on hybrids is more likely to co-opt some locally adapted island alleles into recombinant alleles that fuel adaptive introgression.

S3. No discernible pattern exists for island and river loci in different contact zones, either because of the stochasticity of evolutionary forces at each locus, or because adaptation primarily proceeds through small-effect polygenes with rampant epistasis.

Although these are not mutually exclusive hypotheses, they summarize some of the commonly found patterns of speciation genetics and hybridization in the literature (Colosimo et al. 2005; Reed et al. 2011; Jones et al. 2012; Martin et al. 2012; Rockman 2012; Clarkson et al. 2014; Vijay et al. 2016; Bay and Ruegg 2017). I found Scenario S2 best explains the pattern of islands/rivers significantly overlapping within each contact zone, without any significant overlap of either islands or rivers collectively (Fig. 2.4). However, at the level of gene ontologies, islands cluster together overwhelmingly (Fig. 2.5), suggesting some predictability to functional gene networks involved in divergent adaptation. Altogether, this means adaptation within lineages and adaptation between lineages are compatible processes, and likely synergistic in some cases. This is best exemplified by my GWAS results, showing metabolic adaptation for tadpoles appearing to involve a gene that is highly fixed, and yet introgressing from the East lineage to the West (Fig. 2.7). At the same time, some islands probably bestow hybrids with

incompatibilities, given that islands are migrational sinks overall, within admixture zones (Fig. 2.3). I now discuss each of these results separately.

Compatibility of adaptive introgression with the speciation process

Although there are still detractors (Hedges et al. 2015), the field of evolutionary biology has come full circle, and accepted Darwin's (1859) view that natural selection plays a central role in speciation (Schluter 2001; Via 2002, 2009; Coyne and Orr 2004; Rundle and Nosil 2005; Schluter and Conte 2009; Arnegard et al. 2014). Previous work on the Yosemite toad (Chapter 1) suggests adaptation to different refugial climates has driven lineage formation, and resulted in persistent admixture at secondary contact zones. Ecological speciation-with-gene-flow models such as "Divergence Hitchhiking" and "Genome Hitchhiking" describe this pattern of adaptive divergence, and predict the formation of genomic islands through linked selection of adaptive mutations (Nosil et al. 2009; Feder et al. 2012a,b; Nosil and Feder 2012; Via 2012). In these models, genomic islands are areas of low recombination that support coadapted gene complexes, and may erode if recombination overcomes selection pressure (Barton and Bengtsson 1986; Yeaman 2013; Samuk et al. 2017). Thus, if incipient speciation proceeds with hybrid-mediated gene flow that is also adaptive, then there are two adaptive processes: adaptive divergence, and adaptive introgression. Do these two processes occur independent of one another, perhaps even antagonistically, or are they mutually reinforcing somehow?

My results show that significantly more river loci are co-opted from genomic islands than expected by chance alone (Fig. 2.4), and this pattern is consistently found in

all three contact zones. There are several reasons that recombinant adaptation may be fueled by divergent adaptation. Transgressive segregation, or phenotypes more extreme in hybrids than either parent, should be maximal for additive alleles that are extremely divergent (Rieseberg et al. 1999, 2003). Extreme recombinant variation can be repeatedly and selectively filtered through the F1 generation, expediting selection at loci that were already of adaptive importance (Hedrick 2013). This process could be particularly potent for the Yosemite toad, because its extremely small population sizes (mean $N_e \approx 30$; Chapter 1) make typical selection on novel mutations or standing variation very slow.

As genome-wide rates of divergence and introgression are increasingly measured and compared, other researchers are finding positive relationships between locus-specific levels of divergence and introgression for many hybrid zones, for example in butterflies (Gompert et al. 2012), house mice (Janoušek et al. 2015), budworms (Blackburn et al. 2017), thrushes (Bay and Ruegg 2017), and rattlesnakes (Schield et al. 2017). Analyses of whole-genome datasets have found specific adaptively divergent trait loci that subsequently introgressed across species boundaries. For example, *Heliconius* butterflies among different hybrid zones are highly divergent at large effect loci that control mimicry coloration, and yet some of these loci have introgressed adaptively between species pairs (Joron et al. 2006; Baxter et al. 2008; Reed et al. 2011; Martin et al. 2012; Pardo-diaz et al. 2012; Nadeau et al. 2014). A similar pattern has been found for flower coloration genes in monkeyflowers (Stankowski and Streisfeld 2015), insecticide resistance in mosquitos (Clarkson et al. 2014), and beak morphology in Darwin's finches

(Grant and Grant 2014; Grant 2015). Clearly, a consilience of examples in nature support the idea of genomic rivers co-opting existing genomic islands.

Some aspects of theory may predict the opposite pattern, however. There are constraints on the genomic composition of hybrids, and thus on genomic islands amenable to adaptive introgression. For example, it is well known introgression is limited on sex chromosomes, particularly in species with extreme heterogamety (Coyne and Orr 1989, 1997; Turelli and Hoffmann 1995; Swanson et al. 2001; Presgraves et al. 2003). This is presumably because BDM incompatibilities accumulate on X or W chromosomes, where disruption of coadapted genes is severely limited by negative epistasis (Turelli and Orr 2000; Presgraves 2008). Moreover, when hybrid genomes initially form, portions of the genome are constrained to come from one specific parent (Runemark et al. 2018). However these patterns may only reflect constraints on the most reproductively isolated portions of the genome, and not the vast majority of genomic islands (e.g. Taylor et al. 2014). Ultimately, it may be that only strongly selected islands with epistatic effects on reproduction are refractory to adaptive introgression (Abbott et al. 2013).

It is difficult to estimate what proportion of islands in the Yosemite toad genome may actually confer reproductive isolation; the ddRADseq dataset in this study only represents a small proportion of sites with limited gene annotation available. However, two results may shed some light: (1) collectively, islands in every contact zone are migrational sinks (Fig. 2.3), suggesting some islands may contain speciation genes; and (2) there is a paucity of fully fixed loci, so it would seem that introgression overcomes any reproductive barriers throughout most of the genome. It is also important to note that

"adaptive" introgression is predicated on the assumption that most α outliers are not affected by drift, which is supported by simulations (Gompert and Alex 2011; Gompert and Buerkle 2012). Although it remains possible some rivers are flowing stochastically, this does not explain island/river overlap consistently found across all contact zones.

Predictability of adaptation

Intriguingly, I found that island loci are united by many common gene ontologies, despite being mostly distinct loci (Fig. 2.5). If there is stochasticity at the level of loci, but predictability at the level of gene networks, this favors the view that islands represent parallel divergent adaptation (Stern and Orgogozo 2008, 2009; Stern 2013). Other studies have found similar results. For example, house mouse divergence outliers also cluster by gene ontology (Janoušek et al. 2015), and crow divergence genes affecting plumage coloration show "parallelism by pathway," not identical speciation genes (Vijay et al. 2016). These patterns would imply that although distinct loci are involved in genetic divergence, they might underlie related traits with polygenic inheritance. This raises the question: how predictable is adaptation in speciation, and generally?

One survey of the published literature estimates the mean probability for a gene to be reused in parallel or convergent adaptation is 0.32–0.55, and declines with age of divergence (Conte et al. 2012). However, this high estimate may partly be biased from genes of large effect preferentially studied, when polygenic traits of smaller effect may dominate most of adaptation (Rockman 2012; Berg and Coop 2014; Yeaman 2015). Even if this estimate were off by an order of magnitude, my results do not match any such

pattern of parallel island evolution, at the level of individual RAD loci (Fig. 2.4). It may be that my short-read loci obscure parallelism at the level of genes, but this is unlikely given that loci are enzymatically cut from disparate locations in the genome (Davey et al. 2011; Peterson et al. 2012). Another possibility is that young lineages are more prone to parallelism by pathway than by gene. Rewiring of protein interaction networks is known to be an important driver of phenotypic change, by shifting the importance of individual genes without disrupting pleiotropic interactions (Kim et al. 2012; Stern 2013). This pattern was found in a series of diverging stick insect populations: only a proportion of divergent SNPs were shared among all population pairs and reflected host-plant ecological divergence, whereas the remainder were idiosyncratic SNPs, but shared gene ontologies (Soria-Carrasco et al. 2012). As mentioned above, given this element of stochasticity to island formation, there may be less bias toward developing strong reproductive isolation, thus increasing chances of adaptive introgression at a locus.

Thus, overall predictability in phenotypic evolution might not have a 1:1 relationship with genomic changes, which can manifest as simple mutations, protein interaction shifts (Kim et al. 2012), expression differences (Brawand et al. 2011), or even changes in developmental systems (Verster et al. 2014). This could explain why predictability for single-gene adaptation in bacteria (Weinreich et al. 2006; Roy 2009; Gaut 2015) does not scale up to predictability at the level of speciation genes (Cutter 2015; Gaut 2015; Mandeville et al. 2015). My results are consistent with a growing number of examples that predictability in evolution can manifest at various scales in the genes-to-phenotype pathway, but with stochasticity at other scales. I found predictability

for gene ontologies of divergent adaptation, some predictability for a process of adaptive introgression co-opting island loci, and stochasticity at the level of individual RAD loci and genic SNP effects (Figs. 2.4–2.6). One criticism of genomic island studies generally is that islands may simply be regions of low recombination that are unrelated to levels of adaptation and introgression (Cruickshank and Hahn 2014). However, using semi-relative (Φ_{ST}) as well as absolute (D_{XY}, Maddison and Slatkin's $'s'$) measures of divergence is the prescribed method of minimizing that possibility (Noor and Bennett 2009; Cruickshank and Hahn 2014; Wolf and Ellegren 2017). These measures as well as the use of replicated contact zones to test for adaptation gives increased confidence of the results attained.

Consequences for Yosemite toad tadpole growth and development

Apparent introgression of the highly differentiated LPIN3 candidate gene (Fig. 2.7) is very interesting, because it may indicate the importance of adaptive introgression for the evolution of desiccation resistance. The patterns found for RAD locus C2952—the locus that matched to LPIN3—should be followed up with full gene sequencing to confirm the pattern found. Previous studies have found that expression levels of thyroid receptor genes (TRα and TRβ), as well as thyroid hormone (TH), and corticosterone (CORT) explain the faster development and metamorphosis of certain spadefoot species (Hollar et al. 2011; Kulkarni and Buchholz 2012; Gomez-Mestre et al. 2013). The LPIN3 gene plays a role in regulating lipid metabolism, and so one possibility is it regulates the tradeoff between growth and development in response to a drying environment. This would be consistent with data showing that faster-developing desert spadefoots reduce

their developmental plasticity, and dramatically reduce storage of fat bodies (Kulkarni et al. 2011). Another possibility is that for fast-developing Yosemite toad tadpoles, LPIN3 variants keep growth rate at uniformly low levels. Fast-developing species are known to sacrifice variable growth rates in order to accelerate development rate, at the expense of metamorphosing much smaller (Richter-Boix et al. 2011). Regardless of the exact metabolic role LPIN3 plays, it is part of a larger pattern of likely adaptive introgression for the Yosemite toad. LPIN3 is one of six loci that putatively underlie extreme patterns of growth/development (Fig. 2.7a–b), which are strongly differentiated by lineage (Table 2.2; Fig. 2.7c), and that collectively show a pattern of introgression from East to West lineages (Fig. 2.7d). This is particularly interesting because tadpoles in the East lineage have been shown to possess faster development times than any other lineage (P. Maier, unpublished data), and so this trait may be spreading adaptively.

CONCLUSIONS

Adaptive divergence and introgression may be more complementary forces than currently appreciated. Clearly they are separate forces that can act antagonistically, but my work supports the view that introgression between incipient species may build off the accumulated differences in a constructive way. I showed evidence that admixture in secondary contact zones is a probable source for new adaptive variation in the Yosemite toad. More specifically, I found evidence that genes underlying tadpole growth and development such as LPIN3—a metabolic regulatory gene—may be adaptively introgressing between lineages today. The Yosemite toad is under great future threat from

a warming climate, because it is a species that specializes on breeding in shallow, ephemeral snowmelt ponds; shifting snowmelt phenology and more unpredictable precipitation are expected to directly impact its breeding success (US Fish & Wildlife Service 2014; Brown et al. 2015). It has already been suggested that conservation managers should utilize hybrids—specifically advanced hybrids with well-understood recombinant adaptations to climate—as a resource for conservation genetics (Hamilton and Miller 2016). The series of contact zones in Yosemite may be a perfect opportunity to use this natural process for conservation of the species, given additional experimental research into the fitness of advanced hybrids in nature.

REFERENCES

* References marked with asterisk have been cited in Appendix C only.

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Agresti, A. 2002. Categorical data analysis. 2nd ed. Wiley-Interscience, Hoboken, NJ.

* Altschul, S. F., W. Gish, W. Miller, E. E. W. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Amores, A., J. Catchen, A. Ferrara, Q. Fontenot, and J. H. Postlethwait. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics 188:799–808.

Anderson, E. C., and E. A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160:1217–1229.

* Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arnegard, M. E., M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, D. M. Kingsley, C. L. Peichel, and D. Schluter. 2014. Genetics of ecological divergence during speciation. Nature 511:307–311.

Arnold, M. L., E. S. Ballerini, and A. N. Brothers. 2012. Hybrid fitness, adaptation and evolutionary diversification: lessons learned from Louisiana irises. Heredity 108:159–166.

Arnold, M. L., and S. A. Hodges. 1995. Are natural hybrids fit or unfit relative to their parents? Trends Ecol. Evol. 10:67–71.

Arnold, M. L., and N. H. Martin. 2009. Adaptation by introgression. J. Biol. 8:9–11.

Arnold, M. L., and N. H. Martin. 2010. Hybrid fitness across time and habitats. Trends Ecol. Evol. 25:530–536.

* Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the identification of biology. Nat. Genet. 25:25–29.

Barrett, R. D., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. Science 322:2006–2008.

Barton, N., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57:357–376.

Barton, N., and G. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

* Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy. 2004. The Pfam protein families database. Nucleic Acids Res. 32:D138-D141.

Baxter, S. W., R. Papa, N. Chamberlain, S. J. Humphray, M. Joron, C. Morrison, R. H. Ffrench-Constant, W. O. McMillan, and C. D. Jiggins. 2008. Convergent evolution in the genetic basis of Müllerian mimicry in *Heliconius* butterflies. Genetics 180:1567–1577.

Bay, R. A., and K. Ruegg. 2017. Genomic islands of divergence or opportunities for introgression? Proc. R. Soc. Biol. 284:2229–2243.

Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. 13:969–980.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. Proc. Natl. Acad. Sci. 98:4563–4568.

Berg, J. J., and G. Coop. 2014. A Population genetic signal of polygenic adaptation. PLoS Genet. 10:e1004412.

Blackburn, G. S., B. M. T. Brunet, K. Muirhead, M. Cusson, C. Béliveau, R. C. Levesque, L. M. Lumley, and F. A. H. Sperling. 2017. Distinct sources of gene flow produce contrasting population genetic dynamics at different range boundaries of a *Choristoneura* budworm. Mol. Ecol. 26:6666–6684.

* Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Brady, L. D., and R. A. Griffiths. 2000. Developmental responses to pond desiccation in tadpoles of the British anuran amphibians (*Bufo bufo, B. calamita* and *Rana temporaria*). J. Zool. 252:61–69.

Brawand, D., M. Soumillon, A. Necsulea, P. Julien, G. Csárdi, P. Harrigan, M. Weier, A. Liechti, A. Aximu-Petri, M. Kircher, F. W. Albert, U. Zeller, P. Khaitovich, F. Grützner, S. Bergmann, R. Nielsen, S. Pääbo, and H. Kaessmann. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

Brown, C., M. P. Hayes, G. A. Green, D. C. Macfarlane, and A. J. Lind. 2015. Yosemite toad conservation assessment. USDA Forest Service report. Sonora, CA.

Buerkle, C. A., Z. Gompert, and T. L. Parchman. 2011. The n = 1 constraint in population genomics. Mol. Ecol. 20:1575–1581.

Castric, V., J. Bechsgaard, M. H. Schierup, and X. Vekemans. 2008. Repeated adaptive introgression at a gene under multiallelic balancing selection. PLoS Genet. 4:e1000168.

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

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes, Genomes, Genet. 1:171–182.

Clarkson, C. S., D. Weetman, J. Essandoh, A. E. Yawson, G. Maslen, M. Manske, S. G. Field, M. Webster, T. Antão, B. MacInnis, D. Kwiatkowski, and M. J. Donnelly. 2014. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. Nat. Commun. 5:4248.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter. 2012. The probability of genetic parallelism and convergence in natural populations. Proc. R. Soc. B Biol. Sci. 279:5039–5047.

Coyne, J. A. 1992. Genetics and speciation. Nature 355:511–515.

Coyne, J. A., and A. Orr. 1997. Patterns of speciation in *Drosophila* revisited. Evolution 51:295–303.

Coyne, J., and H. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362–381.

Coyne, J., and H. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

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

Cutter, A. D. 2012. The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. Trends Ecol. Evol. 27:209–218.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 1st ed. John Murray, London, England.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510.

* Davidson, N. M., A. D. Hawkins, and A. Oshlack. 2017. SuperTranscript: a data driven reference for analysis and visualisation of transcriptomes. Genome Biol. 18:1–10.

* Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

Dobzhansky, T. 1937. Genetics and the origin of species. 3rd ed. Columbia University Press, New York, NY.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H. Mäkinen, K. Nadachowska-Brzyska, A. Qvarnström, S. Uebbing, and J. B. W. Wolf. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756–760.

* Engström, P. G., T. Steijger, B. Sipos, G. R. Grant, A. Kahles, G. Rätsch, N. Goldman, T. J. Hubbard, J. Harrow, R. Guigó, P. Bertone, T. Alioto, J. Behr, R. Bohnert, D. Campagna, C. A. Davis, A. Dobin, T. R. Gingeras, G. Jean, P. Kosarev, S. Li, J. Liu, C. E. Mason, V. Molodtsov, Z. Ning, H. Ponstingl, J. F. Prins, P. Ribeca, I. Seledtsov, V. Solovyev, G. Valle, N. Vitulo, K. Wang, T. D. Wu, and G. Zeller. 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 10:1185–1191.

Everett, M. V., M. R. Miller, and J. E. Seeb. 2012. Meiotic maps of sockeye salmon derived from massively parallel DNA sequencing. BMC Genomics 13:1–16.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Feder, J. L., S. P. Egan, and P. Nosil. 2012a. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Feder, J. L., R. Gejji, S. Yeaman, and P. Nosil. 2012b. Establishment of new mutations under divergence and genome hitchhiking. Philos. Trans. R. Soc. B Biol. Sci. 367:461–474.

* Finn, R. D., J. Clements, and S. R. Eddy. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:W29–W37.

Fitzpatrick, B. M. 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evol. Biol. 12:131.

Fraïsse, C., C. Roux, J. J. Welch, and N. Bierne. 2014. Gene-flow in a mosaic hybrid zone: is local introgression adaptive? Genetics 197:939–951.

Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14:685–695.

Gaut, B. S. 2015. Evolution is an experiment: Assessing parallelism in crop domestication and experimental evolution. Mol. Biol. Evol. 32:1661–1671.

Gomez-Mestre, I., S. Kulkarni, and D. R. Buchholz. 2013. Mechanisms and consequences of developmental acceleration in tadpoles responding to pond drying. PLoS One 8:1–12.

Gompert, Z., and B. C. Alex. 2011. Bayesian estimation of genomic clines. Mol. Ecol. 20:2111–2127.

Gompert, Z., and C. A. Buerkle. 2012. bgc: software for bayesian estimation of genomic clines. Mol. Ecol. Resour. 12:1168–1176.

Gompert, Z., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and C. A. Buerkle. 2012. Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. Evolution 66:2167–2181.

Gonen, S., N. R. Lowe, T. Cezard, K. Gharbi, S. C. Bishop, and R. D. Houston. 2014. Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. BMC Genomics 15:1–17.

Gosner, K. 1960. A simplified table for staging anuran embryos and larvae with notes on identification. Herpetologica 16:183–190.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29:644–652.

Grant, B. R. 2015. Introgressive hybridization and natural selection in Darwin's finches. Biol. J. Linn. Soc. 117:812–822.

Grant, P. R., and B. R. Grant. 2014. Synergism of natural selection and introgression in the origin of a new species. Am. Nat. 183:671–681.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, D. Philip, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. D. Macmanes, M. Ott, J. Orvis, and N. Pochet. 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat. Protoc. 8:1494.

Hamilton, J. A., and J. M. Miller. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv. Biol. 30:33–41.

Hawks, J., and G. Cochran. 2006. Dynamics of adaptive introgression from archaic to modern humans. PaleoAnthropology 2006:101−115.

Hedges, S. B., J. Marin, M. Suleski, M. Paymer, and S. Kumar. 2015. Tree of life reveals clock-like speciation and diversification. Mol. Biol. Evol. 32:835–845.

Hedrick, P. W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Hewitt, G. M. 1988. Hybrid zones–natural laboratories for evolutionary studies. Trends Ecol. Evol. 3:158–167.

Hoekstra, H. E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97:222–234.

Hofer, T., M. Foll, and L. Excoffier. 2012. Evolutionary forces shaping genomic islands of population differentiation in humans. BMC Genomics 13:1–13.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6:e1000862.

Hollar, A. R., J. Choi, A. T. Grimm, and D. R. Buchholz. 2011. Higher thyroid hormone receptor expression correlates with short larval periods in spadefoot toads and increases metamorphic rate. Gen. Comp. Endocrinol. 173:190–198.

Janoušek, V., P. Munclinger, L. Wang, K. C. Teeter, and P. K. Tucker. 2015. Functional organization of the genome may shape the species boundary in the house mouse. Mol. Biol. Evol. 32:1208–1220.

Jiggins, C. D., C. Salazar, M. Linares, and J. Mavarez. 2008. Review. Hybrid trait speciation and *Heliconius* butterflies. Philos. Trans. R. Soc. B Biol. Sci. 363:3047–3054.

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

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. a Pollen, T. Howes, C. Amemiya, J. Baldwin, T. Bloom, D. B. Jaffe, R. Nicol, J. Wilkinson, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Jones, O. R., and J. Wang. 2010. COLONY: A program for parentage and sibship inference from multilocus genotype data. Mol. Ecol. Resour. 10:551–555.

Joron, M., R. Papa, M. Beltrán, N. Chamberlain, J. Mavárez, S. Baxter, M. Abanto, E. Bermingham, S. J. Humphray, J. Rogers, H. Beasley, K. Barlow, R. H. Ffrench-Constant, J. Mallet, W. O. McMillan, and C. D. Jiggins. 2006. A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. PLoS Biol. 4:1831–1840.

* Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28:27–30.

Kim, J., I. Kim, S. K. Han, J. U. Bowie, and S. Kim. 2012. Network rewiring is an important mechanism of gene essentiality change. Sci. Rep. 2:1–7.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111–120.

Kirkpatrick, M., and N. Barton. 2006. Chromosome inversions, local adaptation and speciation. Genetics 173:419–434.

* Krogh, A., B. Larsson, G. Von Heijne, and E. L. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–580.

Kulkarni, S. S., and D. R. Buchholz. 2012. Beyond synergy: corticosterone and thyroid hormone have numerous interaction effects on gene regulation in *Xenopus tropicalis* tadpoles. Endocrinology 153:5309–5324.

Kulkarni, S. S., I. Gomez-Mestre, C. L. Moskalik, B. L. Storz, and D. R. Buchholz. 2011. Evolutionary reduction of developmental plasticity in desert spadefoot toads. J. Evol. Biol. 24:2445–2455.

Leips, J., M. G. Mcmanus, and J. Travis. 2000. Response of treefrog larvae to drying ponds: comparing temporary and permanent pond breeders. Ecology 81:2997–3008.

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

Lind, M. I., and F. Johansson. 2007. The degree of adaptive phenotypic plasticity is correlated with the spatial environmental heterogeneity experienced by island populations of *Rana temporaria*. J. Evol. Biol. 20:1288–1297.

Losos, J. B. 1992. The evolution of convergent structure in Caribbean *Anolis* communities. Syst. Biol. 41:403–420.

Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R. Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350:1493–1498.

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

Marques, D. A., K. Lucek, J. I. Meier, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. PLoS Genet. 12:1–34.

Martin, A., R. Papa, N. J. Nadeau, R. I. Hill, B. A. Counterman, G. Halder, C. D. Jiggins, M. R. Kronforst, A. D. Long, W. O. McMillan, and R. D. Reed. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc. Natl. Acad. Sci. 109:12632–12637.

Martinsen, G. D., T. G. Whitham, R. J. Turek, and P. Keim. 2001. Hybrid populations selectively filter gene introgression between species. Evolution 55:1325–1335.

Mayr, E. 1942. Systematics and the origin of species: from the viewpoint of a zoologist. Columbia University Press, New York, NY.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

McKinnon, J. S., S. Mori, B. K. Blackman, L. David, D. M. Kingsley, L. Jamieson, J. Chou, and D. Schluter. 2004. Evidence for ecology's role in speciation. Nature 429:294–298.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17:1–14.

Morey, S. R., and D. N. Reznick. 2004. The relationship between habitat permanence and larval development in California spadefoot toads: field and laboratory comparisons of developmental plasticity. Oikos 104:172–190.

Nadeau, N. J., M. Ruiz, P. Salazar, B. Counterman, J. A. Medina, H. Ortiz-Zuazaga, A. Morrison, W. O. McMillan, C. D. Jiggins, and R. Papa. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. Genome Res. 24:1316–1333.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY.

Noor, M. A. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert examining the role of restricted recombination in maintaining species. Heredity 103:439–444.

Norris, L. C., B. J. Main, Y. Lee, T. C. Collier, A. Fofana, A. J. Cornel, and G. C. Lanzaro. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc. Natl. Acad. Sci. 112:815–820.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. B Biol. Sci. 367:332–342.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Nosil, P., and D. Schluter. 2011. The genes underlying the process of speciation. Trends Ecol. Evol. 26:160–167.

Orr, H. A., J. P. Masly, and D. C. Presgraves. 2004. Speciation genes. Curr. Opin. Genet. Dev. 14:675–679.

Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Evolution 55:1085–1094.

Paradis, E. 2010. Pegas: An R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20:289–290.

Pardo-diaz, C., C. Salazar, S. W. Baxter, C. Merot, W. Figueiredo-ready, M. Joron, W. O. Mcmillan, and C. D. Jiggins. 2012. Adaptive Introgression across species boundaries in *Heliconius* butterflies. PLoS Genet. 8:1–13.

* Petersen, T. N., S. Brunak, G. Von Heijne, and H. Nielsen. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8:785–786.

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

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414.

* Powell, S., K. Forslund, D. Szklarczyk, K. Trachana, A. Roth, J. Huerta-Cepas, T. Gabaldón, T. Rattei, C. Creevey, M. Kuhn, L. J. Jensen, C. Von Mering, and P. Bork. 2014. EggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res. 42:1–9.

Presgraves, D. C. 2008. Sex chromosomes and speciation in Drosophila. Trends Genet. 24:336–343.

Presgraves, D. C., L. Balagopalan, S. M. Abmayr, and H. A. Orr. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature 423:715–719.

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

Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez. 2015. Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. 16:359–371.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reed, E., S. Nunez, D. Kulp, J. Qian, M. P. Reilly, and A. S. Foulkes. 2015. A guide to genome-wide association analysis and post-analytic interrogation. Stat. Med. 34:3769–3792.

Reed, R. D., R. Papa, A. Martin, H. M. Hines, M. R. Kronforst, R. Chen, G. Halder, H. F. Nijhout, and W. O. Mcmillan. 2011. optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333:1137–1141.

Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam. 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature 406:64–67.

Richter-Boix, A., G. A. Llorente, and A. Montori. 2006. A comparative analysis of the adaptive developmental plasticity hypothesis in six Mediterranean anuran species along a pond permanency gradient. Evol. Ecol. Res. 8:1139–1154.

Richter-Boix, A., M. Tejedo, and E. L. Rezende. 2011. Evolution and plasticity of anuran larval development in response to desiccation. A comparative analysis. Ecol. Evol. 1:15–25.

Rieseberg, L. 2011. Adaptive introgression: the seeds of resistance. Curr. Biol. 21:R581–R583.

Rieseberg, L. H., M. A. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaption and speciation. Heredity 83:363–372.

Rieseberg, L. H., A. Widmer, A. M. Arntz, J. M. Burke, D. E. Carr, R. J. Abbott, and T. R. Meagher. 2003. The genetic architecture necessary for transgressive segregation is common in both natural and domesticated populations. Philos. Trans. R. Soc. B Biol. Sci. 358:1141–1147.

Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution: All that's gold does not glitter. Evolution 66:1–17.

Roy, S. W. 2009. Probing evolutionary repeatability: Neutral and double changes and the predictability of evolutionary adaptation. PLoS One 4:e4500.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecol. Lett. 8:336–352.

Runemark, A., C. N. Trier, F. Eroukhmanoff, J. S. Hermansen, M. Matschiner, M. Ravinet, T. O. Elgvin, and G. P. Sætre. 2018. Variation and constraints in hybrid genome formation. Nat. Ecol. Evol. 2:549–556.

Samuk, K., G. L. Owens, K. E. Delmore, S. E. Miller, D. J. Rennison, and D. Schluter. 2017. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. Mol. Ecol. 26:4378–4390.

Sánchez, A., M. Salicrú, and J. Ocaña. 2007. Statistical methods for the analysis of high-throughput data based on functional profiles derived from the Gene Ontology. J. Stat. Plan. Inference 137:3975–3989.

Schield, D. R., R. H. Adams, D. C. Card, B. W. Perry, G. M. Pasquesi, T. Jezkova, D. M. Portik, A. L. Andrew, C. L. Spencer, E. E. Sanchez, M. K. Fujita, S. P. Mackessy, and T. A. Castoe. 2017. Insight into the roles of selection in speciation from genomic patterns of divergence and introgression in secondary contact in venomous rattlesnakes. Ecol. Evol. 7:3951–3966.

Schluter, D. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372–380.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proc. Natl. Acad. Sci. 106:9955–9962.

Servedio, M. R., G. S. Van Doorn, M. Kopp, A. M. Frame, and P. Nosil. 2011. Magic traits in speciation: "magic" but not rare? Trends Ecol. Evol. 26:389–397.

* Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123:603–613.

Snyder, R. J., and H. Dingle. 1989. Adaptive, genetically based differences in life history between estuary and freshwater threespine sticklebacks (*Gasterosteus aculeatus* L.). Can. J. Zool. J. Can. Zool. 67:2448–2454.

Song, Y., S. Endepols, N. Klemann, D. Richter, F. R. Matuschka, C. H. Shih, M. W. Nachman, and M. H. Kohn. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. Curr. Biol. 21:1296–1301.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil. 2012. Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738–742.

Stankowski, S., and M. A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. Proc. R. Soc. B Biol. Sci. 282:1–9.

Stern, D. L. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–764.

Stern, D. L., and V. Orgogozo. 2009. Is genetic evolution predictable? Science 323:746–752.

Stern, D. L., and V. Orgogozo. 2008. The loci of evolution: How predictable is genetic evolution? Evolution 62:2155–2177.

Sukumaran, J., and M. T. Holder. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26:1569–1571.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. 98:7375–7379.

Taylor, S. A., R. L. Curry, T. A. White, V. Ferretti, and I. Lovette. 2014. Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. Evolution 68:3066–3081.

Turelli, M., and A. A. Hoffmann. 1995. Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. Genetics 140:1319–1338.

Turelli, M., and H. A. Orr. 2000. Dominance, epistasis and the genetics of postzygotic isolation. Genetics 154:1663–1679.

Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3:1572–1578.

US Fish & Wildlife Service. 2014. Endangered and threatened wildlife and plants; endangered status for the Sierra Nevada yellow-legged frog and the northern distinct population segment of the mountain yellow-legged frog, and threatened status for the Yosemite toad: final rule. Fed. Regist. 1–56.

Verster, A. J., A. K. Ramani, S. J. McKay, and A. G. Fraser. 2014. Comparative RNAi screens in *C. elegans* and *C. briggsae* reveal the impact of developmental system drift on gene function. PLoS Genet. 10:e1004077.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. R. Soc. B Biol. Sci. 367:451–460.

Via, S. 2009. Natural selection in action during speciation. Proc. Natl. Acad. Sci. 106:9939–9946.

Via, S. 2002. The ecological genetics of speciation. Am. Nat. 159 Suppl:S1-S7.

Vijay, N., C. M. Bossu, J. W. Poelstra, M. H. Weissensteiner, A. Suh, A. P. Kryukov, and J. B. W. Wolf. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat. Commun. 7:1–10.

Weinreich, D. M., N. F. Delaney, M. A. Depristo, and D. L. Hartl. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111–114.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. Am. Nat. 167:794–807.

Wolf, J. B. W., and H. Ellegren. 2017. Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. 18:87–100.

Wu, C. I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Wu, C. I., and C. T. Ting. 2004. Genes and speciation. Nat. Rev. Genet. 5:114–122.

Yeaman, S. 2013. Hybridization and the porous genome: patterns of isolation and introgression in manakins. Mol. Ecol. 22:3195–3197.

Yeaman, S. 2015. Local adaptation by alleles of small effect. Am. Nat. 186:S74–S89.

TABLES AND FIGURES

Table 2.1. **Summary of gene ontology (GO) terms.** GO terms that are unique to a category of locus (e.g. all islands only) are listed. Specific terms are not shown, but rather summarized at the level of broader GO term for brevity. For a complete listing of specific GO terms, see Table S2.1.

Broader (Level 1 and 2) GO Category		**Specific** GO Terms **Only** Enriched In:			
		All Islands	EN Zone	ES Zone	EW Zone
BP	biological regulation	4		1	
BP	cellular component organization or biogenesis	11	2		
BP	cellular process	24	3	5	1
BP	developmental process	17		2	
BP	growth	2			
BP	immune system process	1			
BP	localization		1	2	
BP	metabolic process	9		2	1
BP	multi-organism process	1			
BP	multicellular organismal process	15		1	
BP	negative regulation of biological process	6		2	
BP	positive regulation of biological process	4			
BP	regulation of biological process	11		2	1
BP	reproduction			1	
BP	reproductive process			1	
BP	response to stimulus	1		3	
BP	signaling	1			
BP	single-organism process	25	3	7	
CC	cell part	8	3	2	
CC	extracellular region		1		
CC	extracellular region part	1	1		
CC	macromolecular complex	8			
CC	membrane part				1
CC	membrane-enclosed lumen	3			
CC	organelle	8	4	1	
CC	organelle part	7	3		
MF	binding	8	1	6	1
MF	catalytic activity	3		2	
MF	transcription factor activity, protein binding	3			

* Specific and broad GO terms do not form a simple hierarchy; a specific GO term (e.g. aortic smooth muscle cell differentiation) may fall into multiple categories at Level 2 (i.e. cellular process, developmental process, and multicellular organismal process, and single-organism process).

† Lev. 1 categories: BP = biological process, CC = cellular component, MF = molecular function

‡ GO terms enriched in both islands/rivers of that zone, but no other zones; e.g. EN islands + EN rivers only

Table 2.2. **Genetic markers underlying tadpole growth and development.** Profile of the six markers that were identified by the GWAS. For each RAD locus, quantiles of D_{XY} and α are provided in each contact zone, to illustrate whether that marker is highly fixed (island) and/or highly introgressed (river). Numbers in bold highlight quantiles in the 10% tails of the distribution.

RAD Locus	Island (D_{XY})			River (α)			Transcript*	Mutation†	Gene‡
	EN	ES	EW	EN	ES	EW			
C2952	**0.93**	**0.94**	**0.98**	0.28	0.66	**0.02**	Yes	missense (K/I)	Phosphatidate phosphatase (LPIN3)
C18725	0.48	0.17	-	0.13	**0.09**	-	Yes	-	-
C9564	-	-	-	-	-	-	Yes	-	-
C13533	-	0.38	0.11	-	0.62	**0.02**	No	-	-
I34049	**0.92**	**0.9**	0.77	0.25	0.14	0.47	No	-	-
I49070	-	**0.95**	0.65	-	0.8	0.78	No	-	-

(cont'd)

Protein Domains§	GO Terms¶
Lipin_N (conserved N-terminus) LNS2 (plasmid maintenance, respiration)	fatty acid metabolism glycerophospholipid biosynthesis phosphatidylcholine biosynthesis phosphatidylethanolamine biosynthesis phospholipid metabolism small molecule metabolism
-	-
-	-
-	-
-	-
-	-

* Whether the RAD sequence matched to a transcript in the transcriptome
† Type of mutation, with alternate amino acid states
‡ From SwissProt database
§ From Pfam database
¶ Specific gene ontology (GO) terms describing gene function and localization

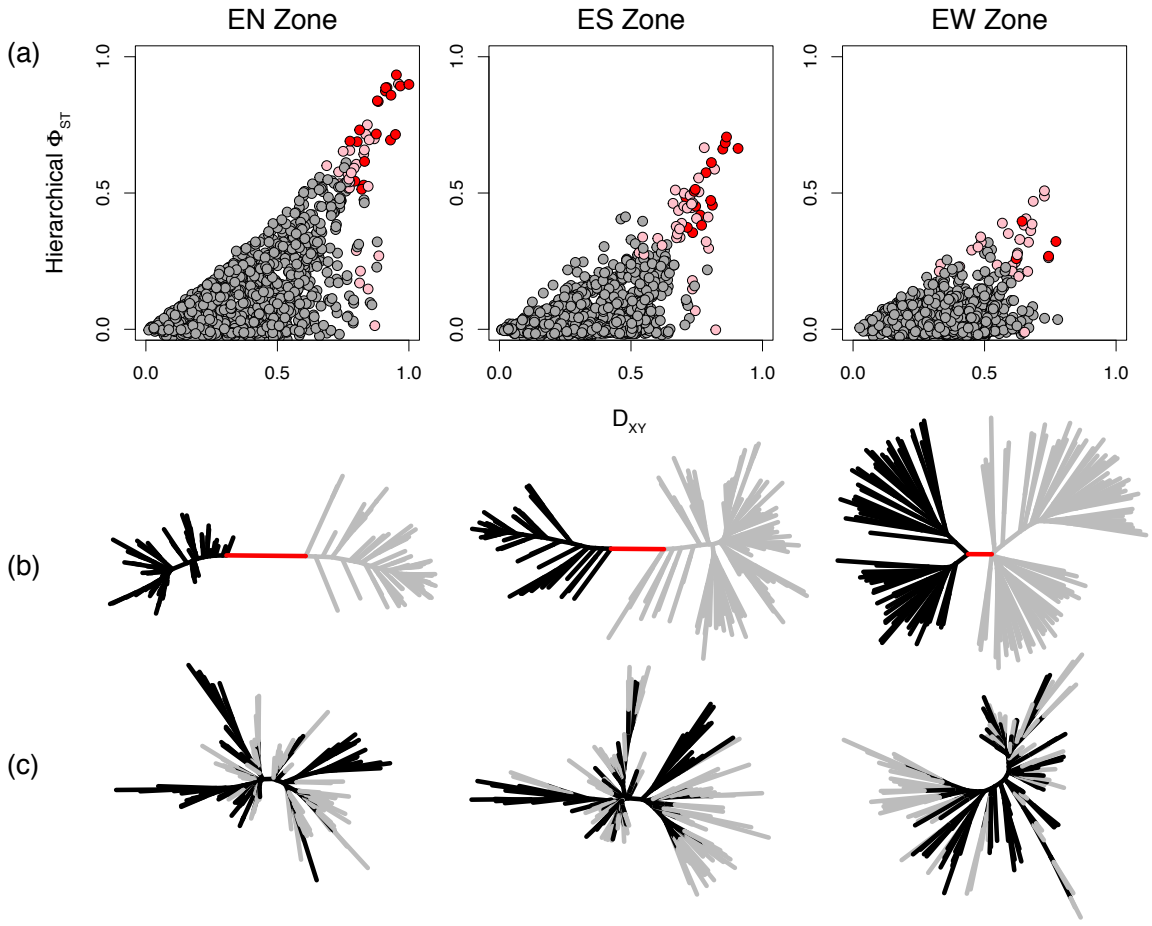


Figure 2.1. **Genomic islands of divergence**
Loci with extreme differentiation between pure individuals, at three contact zones. (a) Scatterplots of hierarchical Φ_{ST} by D_{XY}, showing loci in the 95th quantile that are significantly different than zero; red = outlier loci for Φ_{ST}, D_{XY}, and Maddison and Slatkin's $'s'$, pink = outliers in two out of three tests. (b-c) Neighbor joining trees from RAD sequences at the level of haplotypes (i.e. 2× number of individuals). Loci used are either islands only (b) or an equivalent number of loci chosen randomly from the interquartile range of all D_{XY} values (c), for comparison. Black and gray branches represent ancestry from the pure lineages in that contact zone. The red branch in (b) highlights the fact that island loci produce reciprocal monophyly of lineages.

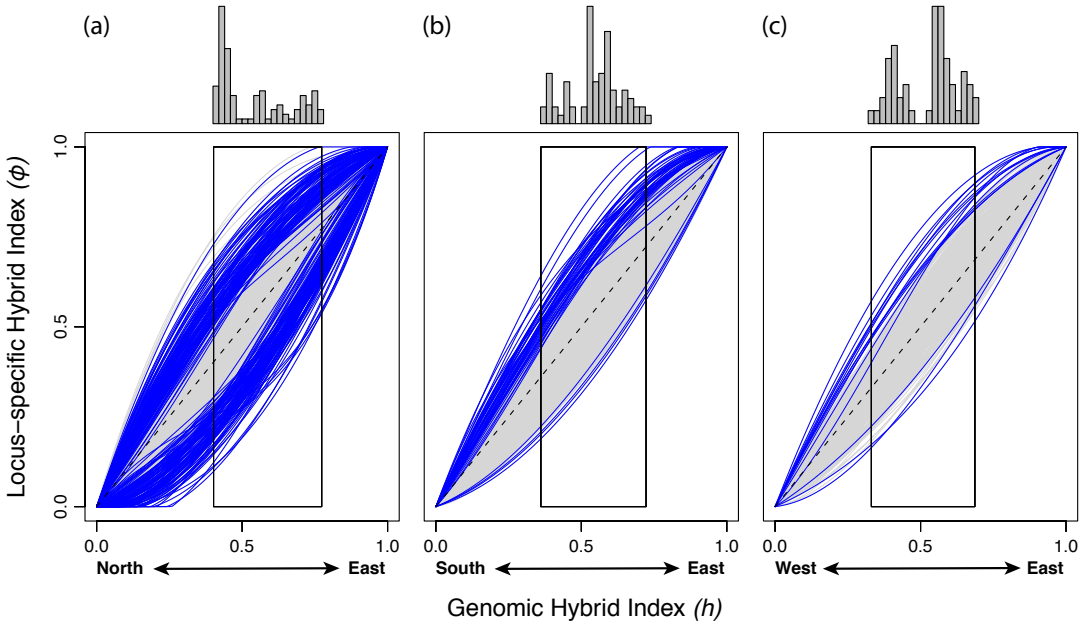


Figure 2.2. **Genomic rivers of introgression**
Bayesian genomic clines for each contact zone: (a) EN, (b) ES, and (c) EW. All loci passing quality and frequency thresholds are shown, with genomic rivers highlighted in blue. Rivers are loci for which 95% credible intervals of its α estimate exclude zero. Hence, rivers can be extreme positive (above stippled line) or negative (below stippled line) values, for loci containing excess East (positive) or North/South/West (negative) ancestry. A histogram of observed genomic hybrid indices (h) is displayed above each plot, with a box highlighting the region of the plot for which h was observed.

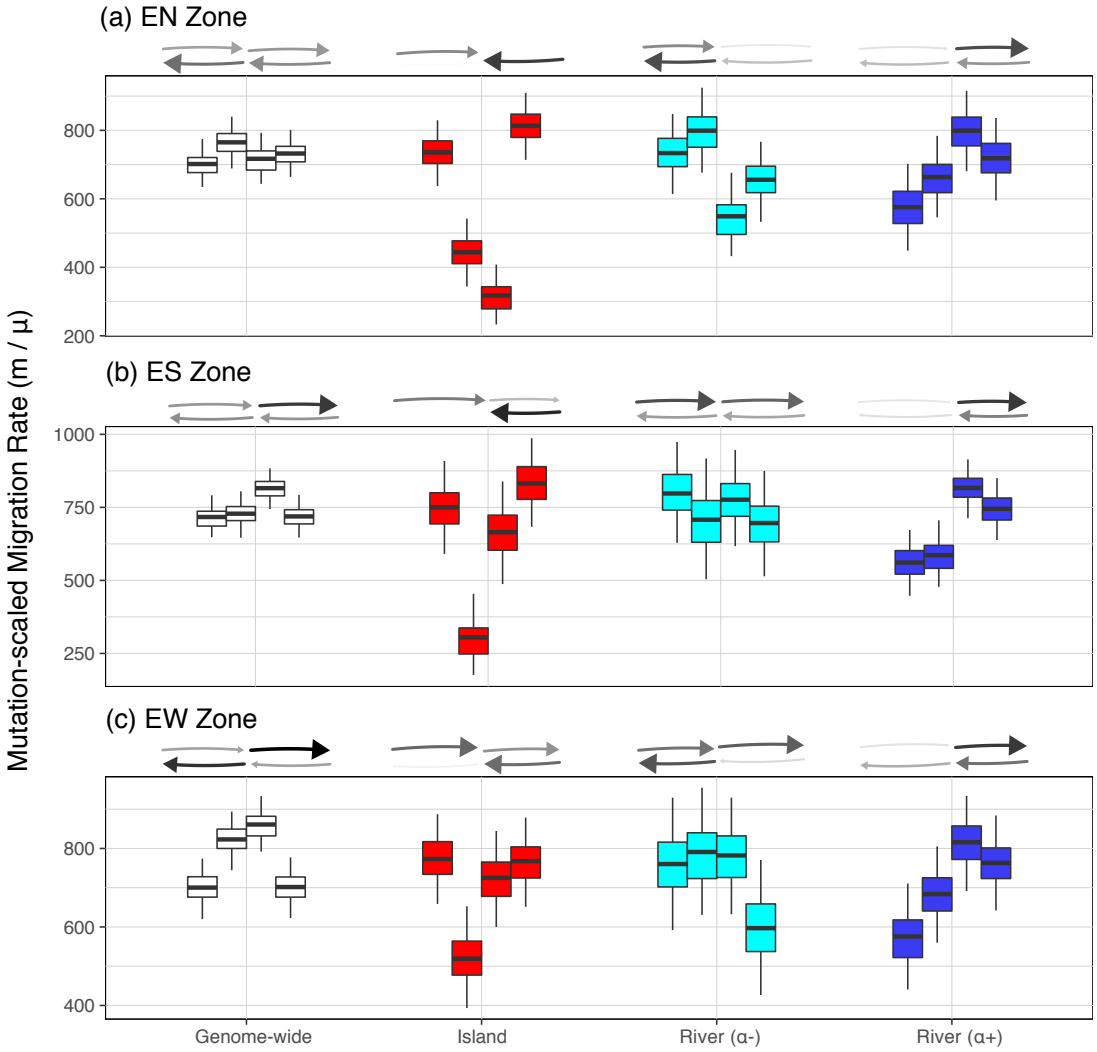


Figure 2.3. **Direction of migration for islands versus rivers**
Results of migrate-n analyses for each contact zone and marker class separately. Relative to genome-wide patterns, island loci tend to act as migrational sinks, and river loci tend to show directional introgression from one lineage or the other. From left to right, each group of four boxplots summarizes the following migration rates: O→Ad, Ad →O, Ad→E, E→Ad, where E = East, Ad = Admixed, and O = Other (North/South/West, depending on contact zone). Each mutation-scaled migration rate (m/μ) is summarized by median and 95% credible intervals. Arrows indicate strength of migration from O (left) to E (right). Shading of arrows is proportional to corresponding migration rates, and scaled by minimum and maximum rates. "Genome-wide" rates are based on 100 randomly sampled loci. Rivers are separated into two classes: $\alpha -$ and $\alpha +$, based on the direction of introgression (see Fig. 2.2).

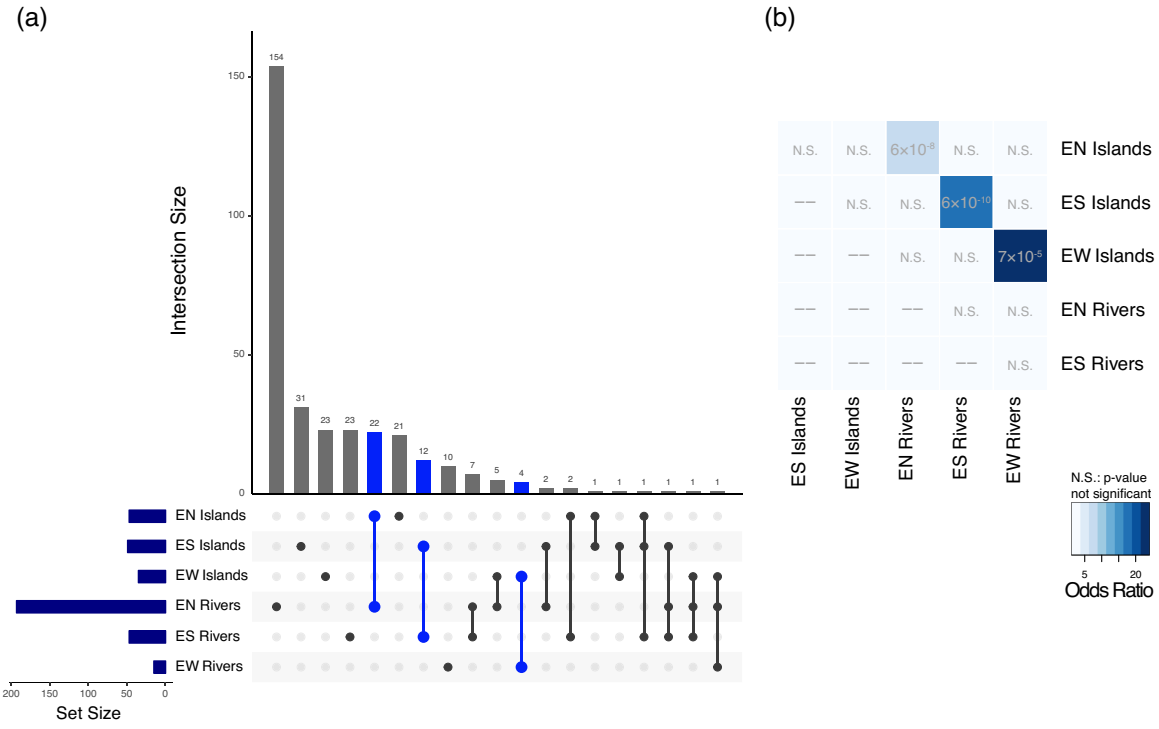


Figure 2.4. **Outlier loci overlap primarily by contact zone**
Islands and rivers share a significant number of loci within the same contact zone only. (a) Upset plot showing the intersection of six categories of loci (islands vs. rivers, across three contact zones). For each intersection shown in a column, the number of shared loci (intersection size) is shown, based on the total number of loci in each category (set size, shown by histogram on rows). Intersections that are specific to a contact zone (e.g. EN Islands + EN Rivers) are highlighted in blue; no intersections involving all islands or all rivers are observed. (b) Results of pairwise Fisher's exact tests of the null hypothesis that categories do not significantly overlap. Odds ratios are represented by the blue heatmap, and significant p-values are shown on their respective box.

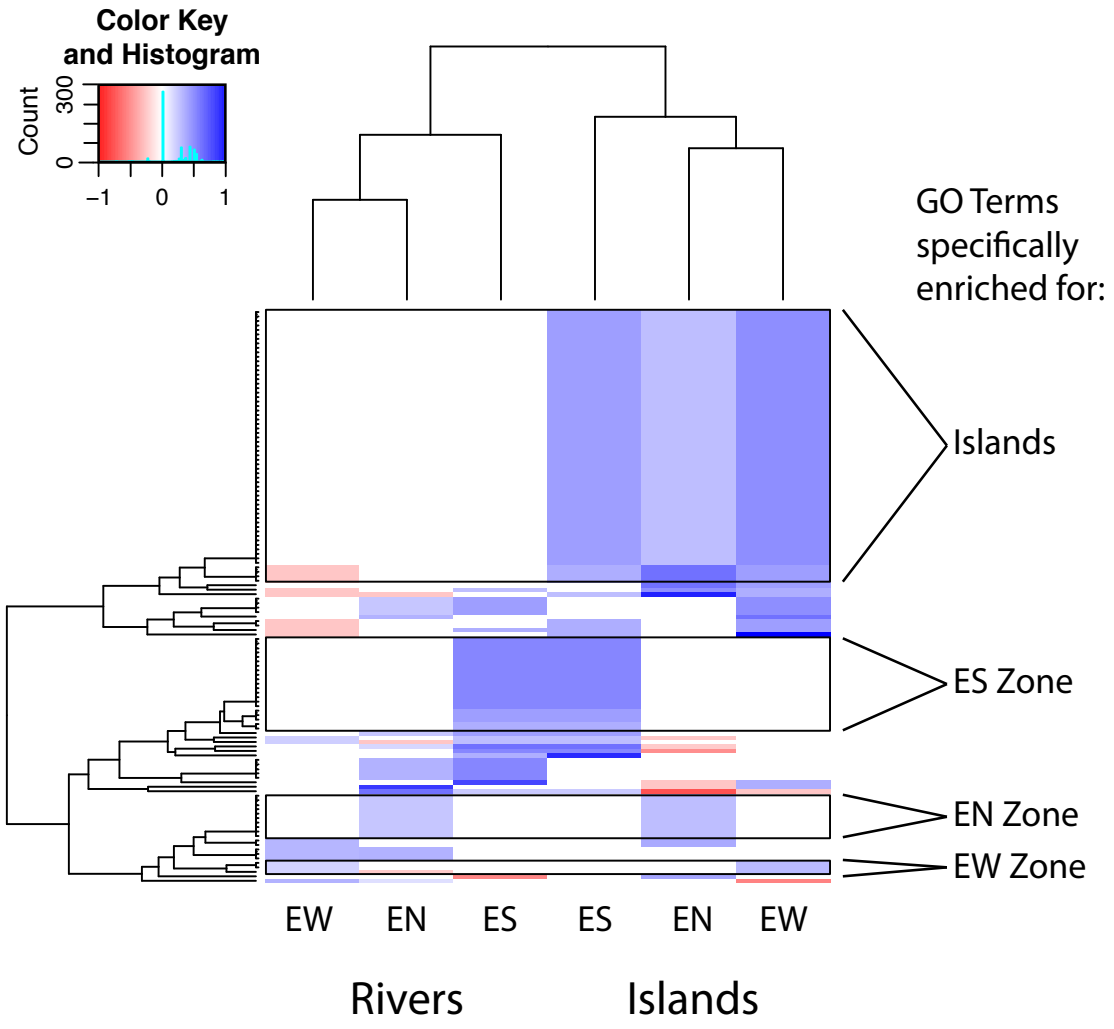


Figure 2.5. **Gene ontologies cluster by marker type and contact zone**
Most shared gene ontologies (GO) are unique to island loci. Hierarchical clustering analysis based on log-transformed p-values of GO term enrichment or depletion in each category of loci. Values representing depleted GO terms are given a (-) sign to polarize enriched/depleted values. Increasing shades of blue show enriched GO terms, and increasing shades of red show depleted GO terms. Marker categories distinctly cluster into islands and rivers (top), whereas GO terms form two primary clusters that include (1) the majority of GO terms specifically enriched for islands, and (2) the majority of GO terms specifically enriched for each of the three contact zones (e.g. EN Islands + EN Rivers only). These GO terms are highlighted with four boxes.

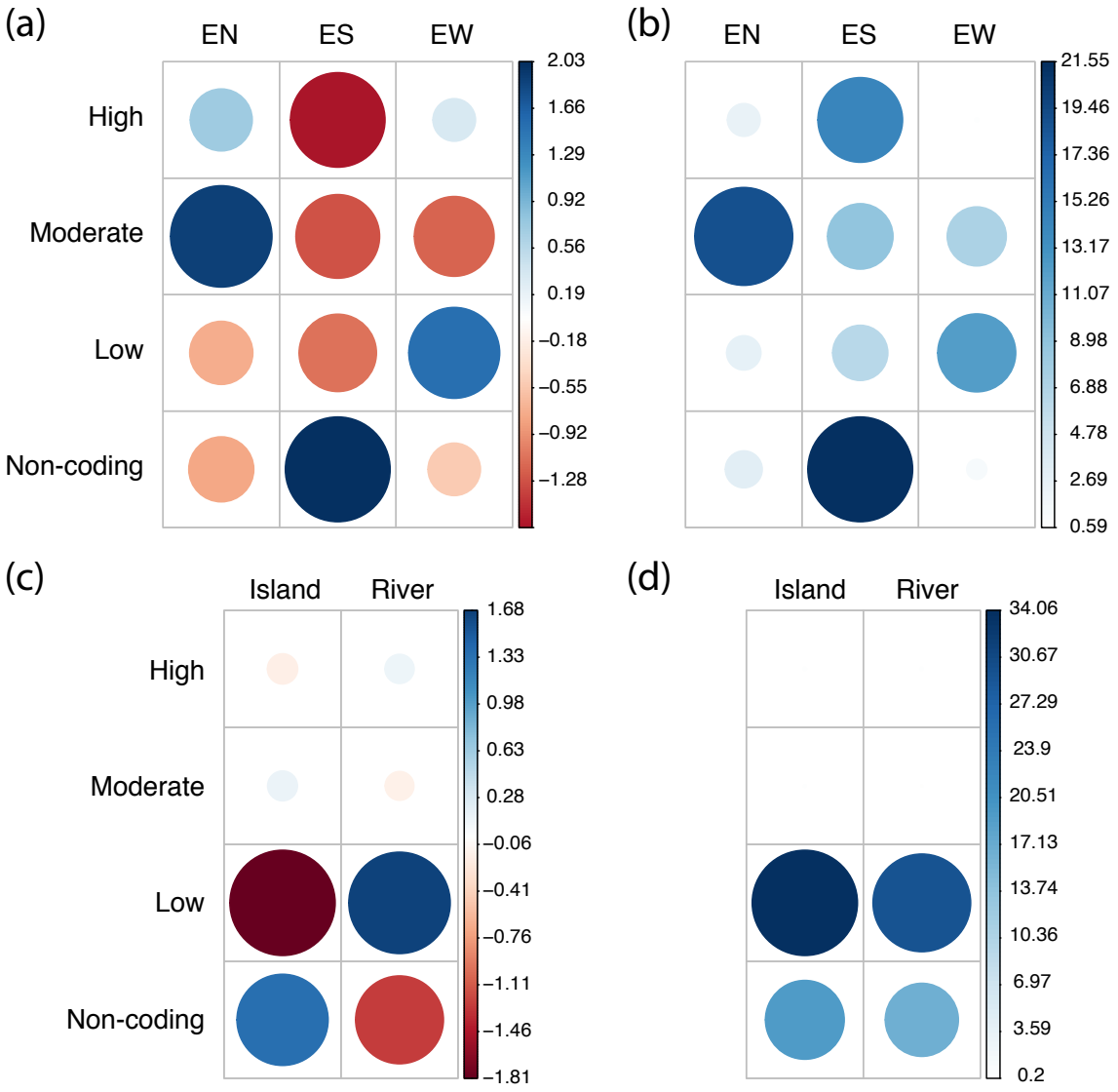


Figure 2.6. **SNP effect categories differ by marker type and contact zone**
Balloon plots summarize the results of a Cochran-Mantel-Haenszel (CMH) chi-squared test, and a subsequent standard chi-squared test, on the independence of SNP effect category across contact zones. (a–b) and marker classes (c–d). Panels (a) and (c) show Pearson residuals of the standard chi-squared test, with increasingly positive values (larger, more blue) suggesting a positive association, and increasingly negative values (larger, more red) suggesting a negative association. Panels (b) and (d) show contribution (%) of each cell to the overall χ^2 score.

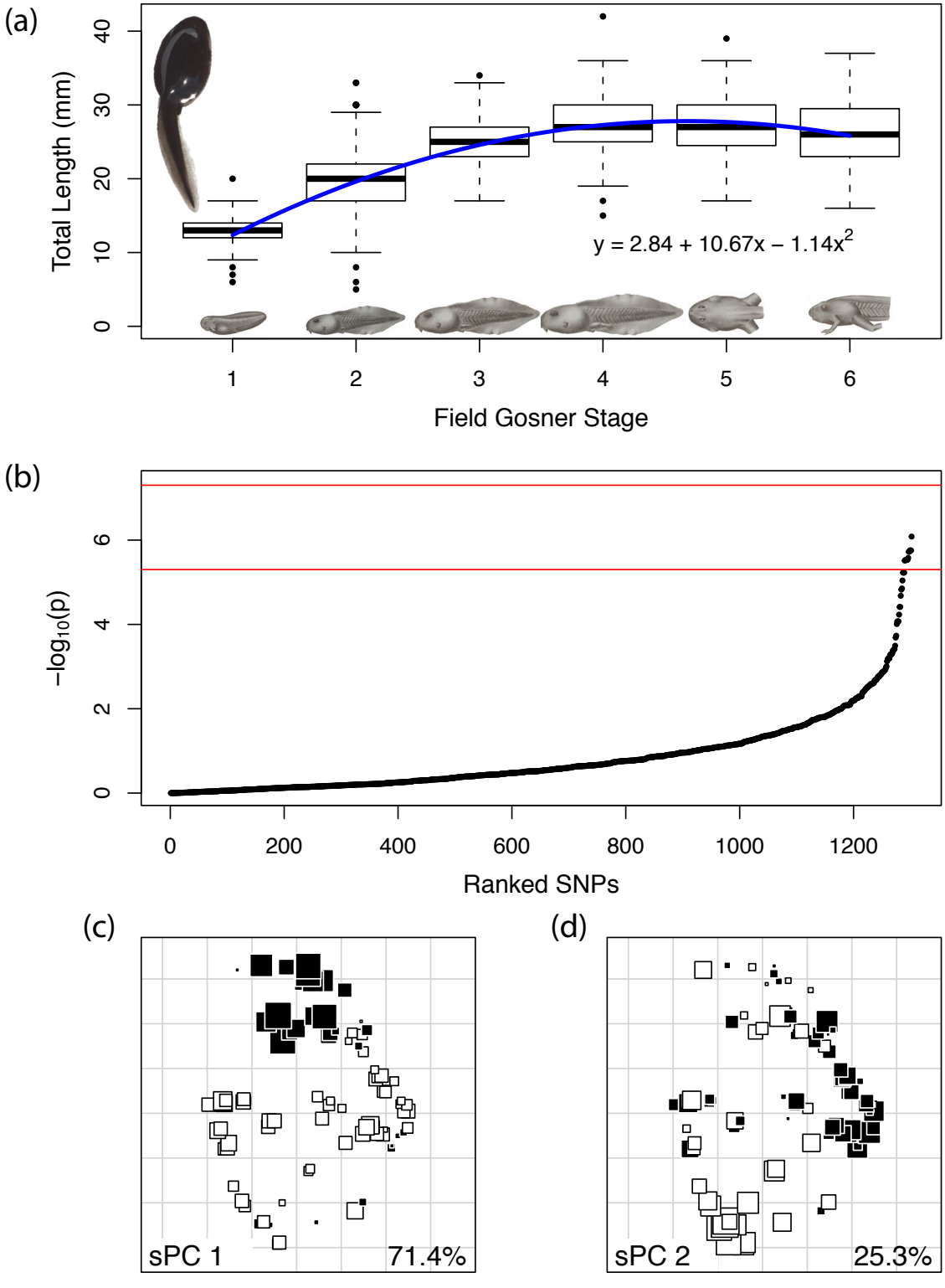

(a)
Total Length (mm)
0
10
20
30
40
y = 2.84 + 10.67x – 1.14x2
1
2
3
4
5
6
Field Gosner Stage
(b)
–log10(p)
0
2
4
6
0
200
400
600
800
1000
1200
Ranked SNPs
(c)
sPC 1
71.4%
(d)
sPC 2
25.3%

Figure 2.7. **Loci associated with tadpole growth and development**
Loci underlying tadpole growth and development show spatial patterns that are both lineage-specific (island-like) and linage-introgressive (river-like). (a) Best polynomial regression model of total tadpole length based on Gosner developmental stage (summarized field version). (b) Results of a GWAS using SNPs as predictors, and residuals from the phenotypic model as the response variable. The two red lines represent "candidate" and "Bonferroni" thresholds. (c-d) Spatial genetic patterns for the six candidate loci, as calculated by a spatial PCA (sPCA), are shown for the first two sPCs. Percent of total variance is shown in the bottom right for each sPC. Squares throughout YOSE represent sPC scores: more positive (larger/black), more negative (larger/white), or closer to zero (smaller).

CHAPTER 3: A NOVEL GENETIC NETWORK MODEL ESTIMATES ENVIRONMENTALLY MEDIATED MIGRATION

ABSTRACT

Amphibian persistence in the 21st century is beset by many anthropogenic impacts including habitat alteration, chemical toxicants, pandemic disease, and climate shifts. The Yosemite toad (*Anaxyrus canorus*) is a species of meadow-specializing amphibian endemic to the high-elevation and federally protected Sierra Nevada of California, yet up to 69% of its populations have recently been extirpated. Although limited evidence exists for species-wide declines due to pesticide use or disease, climate change is projected to have a major hydrological impact on its core meadow habitat. Further, little is known about the relative importance of local habitat stability for larval toads compared to adult dispersal corridors in promoting population persistence. In this study, I used a previously developed double-digest RADseq dataset along with numerous remotely-sensed habitat attributes in a landscape genetics framework to answer three primary questions: (1) Which better approximates the boundaries of current populations, meadows or clusters of meadows? (2) Which environmental, topographic, and climatic attributes most facilitate meadow connectivity? (3) What are the environmental predictors of source-sink structure? I used a novel gravity modeling approach that has wide application to modeling microevolutionary dynamics in patch-separated amphibian populations. Multiple analyses of genetic structure suggested that populations are typically circumscribed by a single meadow, however clusters of meadows display source-sink structure, with large and flat "hub" meadows surrounded by smaller and rugged "satellite" meadows. My results revealed distinct environmental drivers of connectivity among four phylogeographic lineages of toads. I also found that meadows with higher

moisture regimes produce larger tadpoles and toads with higher site fidelity. A better understanding of Yosemite toad natural history applied in this fine-scale ecological genetic framework is a crucial first step toward mitigating further declines, and my results will have significant utility for prioritizing the future management and conservation of this iconic California native.

INTRODUCTION

Genetic connectivity is an essential force for fostering long-term persistence of a species with discrete populations, by mitigating inbreeding depression, replenishing adaptive genetic variation, and promoting heterosis (Gotelli 1991; Hansson 1991; Whitlock et al. 2000; Gaggiotti 2003; Lowe and Allendorf 2010). Pond-breeding amphibians with complex biphasic life histories present a special challenge to the study of landscape genetics, particularly when larvae are restricted to isolated patches. The process of emigration is at least partly influenced by habitat quality, larval recruitment, and density-dependence (Pulliam 1988; Matthysen 2005; Mathieu et al. 2010), while transient migration and immigration depend on adult behavior and limits to dispersal (Baguette and Van Dyck 2007; Clobert et al. 2009). Typically, metapopulation genetic studies focus on local environmental effects (Geffen et al. 2004; Busch et al. 2009; Roy et al. 2012), whereas traditional landscape genetic studies focus on effective dispersal environment (Manel et al. 2003; Storfer et al. 2010; Manel and Holderegger 2013). For biphasic species, estimates of connectivity would benefit from modeling between-site and at-site effects separately as spatial “links” and “nodes” (Pflüger and Balkenhol 2014).

Studies that have successfully done so have found both source and dispersal habitat to strongly influence overall connectivity (Murphy et al. 2010b; Dileo et al. 2014). An added benefit of this approach is that graph theory can be applied to entire networks of species connectivity, to elucidate emergent properties of indirect gene flow such as "centrality" or "modularity" (Dyer and Nason 2004; Garroway et al. 2008; Dyer et al. 2010). It is essential to consider both types of environment to properly understand not only how organisms are genetically structured, but also how and where this network structure is expected to respond to future changes in habitat and climate.

The Yosemite toad (*Anaxyrus canorus*) exemplifies such biphasic complexity; hence, network methods are ideal for answering conservation questions in this species. The Yosemite toad almost exclusively breeds in the transient and highly productive pools of mountain meadows (Ratliff 1985), which make up <3% of the landscape. While tadpoles develop and metamorphose in meadow pools and flooded areas, adult toads forage, hibernate, and can disperse into intervening habitat over 1 km per season (Martin 2008; Liang 2010). Such extreme vagility for a small anuran has the capacity to completely obscure inferences of long-term population stability. For example, Yosemite toads are a federally threatened species that is reputed to be declining in both distribution and abundance (Sherman and Morton 1993; Drost and Fellers 1994; Jennings and Hayes 1994; Brown and Olsen 2013; US Fish & Wildlife Service 2014; Brown et al. 2015; Thompson et al. 2016). However, recent survey efforts have suggested that there is no change or even a modest increase in regional meadow occupancy, and that previous results may reflect bias toward studying historically occupied sites (Ostoja et al., in

prep.). Extreme dispersal ability may be the cause of such discrepancies: migrant adult toads may live over 15 years, only breed every two to four years (Sherman 1980; Sherman and Morton 1984), and thus populations may occasionally undergo spatial shifts in meadow occupancy. Ecological studies conducted in finite areas are unlikely to detect such long-range migration, because the probability of detecting such movement is inversely proportional to the distance traveled (Barrowclough 1978). Thus, ecological methods alone may be inadequate to disentangle patterns of migration from patterns of meadow occupancy, or identify the environmental drivers of either process. Genetic network methods could help address this problem, because they may simultaneously identify environmental drivers of meadow emigration and inter-meadow migration, and test hypotheses for causes of species-wide fragmentation.

Despite numerous efforts to find causative agents for the decline of Yosemite toads, such as UV radiation (Sadinski 2004), exotic predators (Grasso et al. 2010), meadow grazing (Roche et al. 2012a,b; Matchett et al. 2015), chemical deposition (Bradford and Gordon 1994; Davidson 2004; Sadinski 2004), and chytridiomycosis (Dodge and Vredenberg 2012; Lindauer 2018), no clear patterns have emerged (Brown et al. 2015). This is in stark contrast to the "smoking-gun" inferences about causes for other amphibian declines and recoveries in California (Knapp and Matthews 2000; Vredenburg et al. 2010; Knapp et al. 2016). If Yosemite toads are declining in distribution and/or abundance, climate change is a top candidate, because it is predicted to have an inordinately large impact on meadow hydrology, and projected to dramatically reduce the geographic range for Yosemite toads by 2100 (Viers et al. 2013; US Fish & Wildlife

Service 2014). Climate change is known to cause phenological and range shifts as well as genetic modulation in many organisms (reviewed in: Parmesan 2006), and specifically in amphibians (Alexander and Eischeid 2001; Pounds 2001; Pounds et al. 2006). Yosemite toad breeding ecology makes them particularly vulnerable, as they are obligate shallow pond breeders ($\bar{x}$ = 4.35 cm; Liang et al. 2017), thus tadpole survival depends upon snowpack runoff and spring recharge. Significant larval mortality is often observed even during years with above-average snowpack (Sherman 1980; Sherman and Morton 1984, 1993; Brown et al. 2015). Adults and post-metamorphic juveniles are particularly sensitive to the energetic costs of overwintering as it relates to temperature (Morton 1981). Over the next century, California is generally expected to experience an increase in precipitation, decrease in Sierra snowpack, and decrease in spring and summer runoff, although with much spatial heterogeneity in the mountains (Smith and Tirpak 1988). Therefore, a model of fine-resolution climatic impacts on Yosemite toad network connectivity will help address how climate change may fragment meadows in the future.

My first goal in this study was to identify whether meadows or clusters of meadows best approximate population boundaries, which are defined by ecological and genetic independence. A previous study (Shaffer and Fellers 2000) suggested genetic neighborhoods are less than a few kilometers across, however that study had limited sampling coverage and evenness. Since that study, the US Geological Survey used a newly delimited GIS layer of meadow polygons (Keeler-Wolf et al. 2012) to perform a systematic six-year census of Yosemite toad occupancy at the meadow scale (Ostoja et al., in prep.), which revealed that occupied toad meadows form naturally distinct and

hierarchical clusters across the landscape. In their conservation assessment, Brown et al. (2015) identified an important conservation priority for Yosemite toads: quantifying "the spatial scales over which populations function…and [their] variation." This scale of this unit is still unknown, but expected to be at least the size of a meadow, and smaller than previously described lineages (genetically/ecologically distinct units that have evolved in semi-isolation through deep time) (Graybeal 1993, 1997; Shaffer and Fellers 2000; Stephens 2001; Goebel 2005; Goebel and Ranker 2009; Chapter 1). I used a previously described double-digest RADseq (ddRAD) dataset (Chapter 1), with a highly robust sampling scheme (for reducing bias associated with unsampled locations), to examine the scale of population genetic boundaries.

Second, I wanted to identify which environmental, topographic, and climatic attributes most facilitate meadow connectivity. I used a novel combination of elements to answer this question, and to address three ecological and evolutionary challenges with modeling in landscape genetics (Table 3.1; Fig. 3.1): (1) Dispersal routes are complex. Transect methods in landscape genetics seek to extract raw environmental variables from dispersal transects (rectangles) between sites (van Strien et al. 2012; van Strien 2013), but these shapes may not capture the biologically complex corridors used by migrants. To address this challenge, I tested alternative dispersal route models using least cost path (LCP) analysis (Douglas 1994), chose optimal routes using causal modeling (Cushman et al. 2006, 2013), and then extracted variables from buffered routes. (2) Dispersal habitat may be environmentally distinct from at-site habitat, and may have unique contributions to connectivity. I separately modeled node-based (meadow) versus link-based (dispersal

corridor) habitat attributes using gravity modeling (Anderson 1979; Murphy et al. 2010b). Although habitat suitability models are often assumed to explain both at-site and between-site environmental preferences (Epps et al. 2007; Storfer et al. 2007), including for the Yosemite toad (Wang 2012), organisms with distinct breeding and dispersal environments should be modeled as such, because (a) this will improve accuracy of predictions, and (b) manipulating site-specific influence on connectivity may be more tractable for land managers. (3) Individuals in different lineages may possess distinct adaptations, which may change their movement and breeding patterns. Researchers seldom account for phylogeographic signal in genetic differentiation estimates (Dyer et al. 2010), and this unaccounted source of variance may have unpredictable results in landscape genetic models. I corrected for lineage adaptations by incorporating lineage-specific slopes to environmental variables, if doing so significantly improved the model.

In addressing this second goal, I hypothesized that several types of environmental features and stressors may impact Yosemite toad connectivity: topography, vegetation, climate, disturbance, snowmelt timing, and meadow network attributes (the configuration of nearby meadows). I integrated climate and climate change variables by using the 2014 California Basin Characterization Model (BCM; Flint and Flint 2012; Flint et al. 2013), which consists of downscaled 270 m rasters derived from 800 m PRISM data (temperature and precipitation) as well as elevation, geology, and soils data. Given any influence by climate, I asked whether 60-year directional change in those variables had already impacted genetic patterns. Occupancy and breeding of Yosemite toads is known to partly depend on the quality of other meadows in a local neighborhood (Berlow et al.

2013), and Wang (2012) found that 42.6% of genetic structure between meadows was attributable to topography and climate. However virtually nothing is known about how meadows and the intervening landscape specifically influence overall genetic network structure, and the explicit role of climate and climate change remains unknown.

My last goal was to characterize patterns of source-sink dynamics among neighboring meadows, including any environmental predictors of directional migration. I used a frequency-based measure of pairwise genetic structure that assumes the number of migrants is inversely related to the frequency of private alleles (Slatkin 1985; Sundqvist et al. 2016). Theory predicts source-sink differentiation to be large if migration from sources is stochastic or episodic, as occurs with infrequent recolonization events by small founder populations (Gaggiotti 1996). Metapopulations are traditionally defined by a stringent definition involving extinction and recolonization of patches with limited dispersal (Levins 1969; Hanski and Gilpin 1991; Hanski 1998). This model may vastly oversimplify reasons for patch isolation in amphibians, and is rarely supported by sufficient evidence (Marsh and Trenham 2001; Smith and Green 2005). Although dispersal ability may not be limiting for Yosemite toads among most meadows, toads are thought to show strong site fidelity over short time-scales, and this probably has environmental underpinnings (Liang 2010; Berlow et al. 2013). I hypothesized that clusters of nearby occupied meadows (a pattern discovered through recent surveys by Ostoja et al., in prep.) may be recolonized over longer time-scales by “hub” meadows with high quality breeding area, and minimal topographic complexity.

My landscape genetic study will help determine environmental mechanisms driving Yosemite toad population structure and connectivity. My results will be useful for highlighting the most effective mitigation actions against poorly understood threats. Identifying source-sink dynamics will help identify meadows that are more robust to ongoing declines, and which may be utilized for specific conservation actions, such as relocations. In addition, my novel workflow provides a useful framework for characterizing genetic network structure in patch-limited species, and may help managers to make informed judgments about future conservation actions.

MATERIALS AND METHODS

Spatial extent of sample collection

Sites (=meadows) were chosen to maximize representation across all known breeding locations from a recent six-year survey effort (Ostoja et al., in prep) and overlap with previous studies. Tadpoles were sampled across multiple clutches, pools, and years to maximize inclusion of available genetic diversity. A minimum of five samples was used per meadow if additional meadows were sequenced within 1 km, and 10 per meadow for remaining meadows, unless insufficient samples were available. This scheme maximized intra- and inter-meadow sampling representation across the study area.

Molecular methods, ddRAD sequencing, and bioinformatics

I used the previously described ddRADseq haplotype dataset (Chapter 1) for all analyses of genetic differentiation and structure. Details about the library preparation,

sequencing, and bioinformatic parameters used to identify variable loci are described therein. Briefly, a total of 535 individual Yosemite toad (*Anaxyrus canorus*) tadpoles sampled from 90 meadows across Yosemite National Park (YOSE), and 109 individual tadpoles sampled from 12 meadows in Kings Canyon National Park (KICA), were sequenced using 2×100 bp sequencing on an Illumina HiSeq 2500. Stringent quality thresholds were used, and several measures were used in the Stacks v1.19 (Catchen et al. 2011, 2013) pipeline to avoid recovering markers from paralogous sequences. Weir and Cockerham's pairwise F_{ST} estimates (Weir 1996) were computed between meadows using populations.pl in Stacks, which corrects for bias associated with different sample sizes. This estimate is iterated across loci and then averaged for each pair of populations.

Delineating population boundaries

The two previous studies that have measured population genetic patterns of Yosemite toads have found that genetic structure among sampling sites (also sampled at the meadow scale) is strongly affected by distance between sites (Shaffer and Fellers 2000; Wang 2012), a pattern known as isolation by distance (IBD). However, both studies had a limited number of sites sampled in YOSE (n=23; n=24), which was their primary study area, and most sites were strongly biased toward more accessible sites within the East lineage. Therefore, I first wanted to reaffirm IBD patterns using my dataset that consists of 90 meadows sampled evenly from across all known Yosemite toad meadows (Ostoja et al., in prep.), and all major lineages discovered in Chapter 1.

I initially checked all loci for deviations from Hardy-Weinberg expectations at the meadow scale, using the adegenet package (Jombart 2008) in R v3.3.3 (R Core Team 2018). I then used the ade4 package (Dray and Dufour 2007) to test for isolation by distance (IBD) between meadows with a mantel test, and noted that meadows are strongly differentiated from each other as a function of their geographic separation (see Results). A pattern of IBD or hierarchical structure can confound population delineation algorithms that attempt to assign individuals into a set of clearly differentiated Hardy-Weinberg populations (Meirmans 2012). Therefore, I delineated current population boundaries in three distinct ways.

First, I asked which level of population structure is generally the smallest detectable level: meadows or clusters of nearby meadows. I used a hierarchical STRUCTURE analysis (Vähä et al. 2007), which accounts for IBD by re-running each clustering analysis recursively until no further sub-structure is found. This procedure is specifically recommend in later versions of STRUCTURE (≥2.1), because IBD or hierarchical structure can seriously violate the assumption that discrete populations exist, and STRUCTURE will generally find the highest level of hierarchical structure (Evanno et al. 2005; Pritchard et al. 2010). I ran STRUCTURE v2.3.4 (Pritchard et al. 2000) three times for 4.0×10^5 steps and 1.0×10^5 burn-in with K (number of ancestral populations) ranging from 1–10. The optimal K was chosen using the ΔK method of Evanno (2005), and individual runs were combined using CLUMPP (Jakobsson and Rosenberg 2007). This method can only test optimal K values greater than one, so the hypothesis that K=1 was tested using a paired Wilcoxon test (Rosenberg et al. 2001; Emanuelli et al. 2013).

Each cluster of meadows was rerun until reaching the meadow level, or finding no further sub-structure. This workflow addresses the "K=2 conundrum," or inherent bias of ΔK toward finding K=2 when no further substructure exists in reality (Janes et al. 2017).

Second, I asked which level of population structure (meadows or meadow clusters) is found most predictably in the data, without considering population hierarchy. This question is best suited for the program ADMIXTURE, which is efficient at applying the algorithm used in STRUCTURE for many multi-locus SNP genotypes at large values of K (Alexander et al. 2009). The cross-validation procedure employed in ADMIXTURE determines which K has the most predictive accuracy, while withholding data points (Alexander and Lange 2011). I randomly selected one SNP per locus and ran ADMIXTURE v1.3.0 with 10-fold cross-validation at K values ranging from one to the number of meadows sampled (90), then chose the K with lowest cross-validation error.

Finally, I took a hypothesis-driven approach to population delineation. In species with IBD or hierarchical structure such as the Yosemite toad, there is no "true" level of population structure—multiple meaningful levels exist. Therefore, a better question is: which population level has the most biological significance? Instead of simply relying upon genetic clustering algorithms or tests of population differentiation to produce the "correct" K, I decided to test the relative evidence for two biologically meaningful hypotheses. (1) Previous survey and census efforts have found Yosemite toad meadows to be spatially aggregated in space (Brown et al. 2012; Berlow et al. 2013; Matchett et al. 2015; Ostoja et al., in prep.), lending the possibility that a cluster of meadows coincides with the unit of one population. (2) Alternatively, a single meadow may better reflect that

population unit. Toads only travel an average linear distance of 275–279 m each year, with average home range diameter of 92–166 m (Martin 2008; Liang 2010), whereas the average meadow diameter in YOSE is 266 m. Meadows may be the best population unit, given the high rate of philopatry observed in these toads (Sherman 1980; Liang 2010).

I used a hierarchical analysis of molecular variance (AMOVA) to assess which of these two levels is a better reflection of current gene pool (i.e. population) boundaries. For the level of "meadow cluster," I used the optimal K from ADMIXTURE, which coincided well with this hypothesized biological unit (see Results). I used the ade4 function "amova" to perform a hierarchical AMOVA with 100 permutations for significance. In order to correctly partition the variance components of interest, I included all known levels: individuals, meadows, meadow clusters, lineages, and parks. Due to the hierarchical nature of the test, I needed to assign all meadows to one meadow cluster, and all clusters to one lineage. Each meadow was assigned to the cluster that contained the largest proportion of its ancestry. Lineage assignments were designated using previous phylogenetic results (Chapter 1). NeEstimator v2.01 was also used to estimate N_e values for both meadows and meadow clusters using the single-sample linkage disequilibrium method (Do et al. 2014). I compared the ϕ-statistics for meadows and meadow clusters to values expected under Wright's island model (Wright 1931), i.e. $F_{ST} \approx 1/(1 + 4N_e m)$, using mean values of N_e calculated for the two groupings, and assessed population independence with the 10% migration rate (m) cutoff recommended by Hastings (1993) and Palsbøll et al. (2007).

Landscape genetic workflow

I implemented a three-step landscape genetic procedure to determine the most important environmental features impacting toad gene flow between meadows in YOSE (Fig. 3.1). I chose meadows as units, since I found meadows were the best population unit in most cases (see Results). (1) I constructed plausible dispersal paths between each pair of meadows based on hypothesized environmental resistance. (2) I constructed 100 m buffers around these paths as well as the breeding locations themselves, and extracted the average or standard deviation of remotely sensed and USGS survey data. (3) I used linear mixed modeling (a version of gravity modeling) to find likely predictors of meadow connectivity. Each of these steps is described in detail below.

Least cost path transects

Landscape genetic studies are designed to model environmentally mediated gene flow. Most use a statistical framework that falls into one of two categories: (1) resistance surfaces are parameterized with one or more environmental variables so that each pixel represents a hypothesized cost to dispersal (e.g. Peterman et al. 2014), and subsequently least cost paths (LCPs; Douglas 1994), least cost corridors (LCCs; Epps et al. 2007; Pinto and Keitt 2009), or circuit-theoretical currents (McRae 2006; McRae and Beier 2007; McRae et al. 2008) are calculated. The total cost or length of the result is subsequently regressed against some measure of genetic structure, and hypotheses can be evaluated by ranking model fit (e.g. causal modeling). (2) Rectilinear transects based on Euclidean lines between sites are used to collect the raw values of environmental variables (e.g. van

Strien et al. 2012; van Strien 2013). The influence of raw variables can then be individually modeled, e.g. using a multiple regression on distance matrices, random forests, or linear mixed modeling (e.g. gravity modeling). Both approaches have merits and limitations: Cost surfaces directly model the spatial flow of gene copies and presumably model an organism's dispersal corridor better than rectilinear transects, however surface values must be subjectively defined based on the perceived impact of each environmental variable. For hypotheses involving multiple variables, these subjective values must also be combined and transformed (Peterman 2014), which may distort the original variables and make interpreting results difficult. Rectilinear transect methods avoid the subjectivity of cost surfaces, and permit models with a larger number of variables. However, this approach might fail to capture relevant landscape attributes if dispersal corridors deviate drastically from straight-line transects. For example, YOSE includes extremely rugged terrain and amphibians are unlikely to cross major ridgelines (Funk et al. 1999, 2005; Lougheed 1999; Murphy et al. 2010a,b). I implemented a novel approach that combines the benefits of these two approaches (Table 3.1): first, simple LCP models were constructed and buffered similar to the LCTA method described by van Strien et al. (2012; 2013); then environmental variables were individually extracted from LCP transects for downstream variable selection and modeling (Fig. 3.1).

Previous research and expert opinion suggests that higher slope (including ridgelines) and drier vegetation may limit dispersal (Martin 2008; Liang 2010; Wang 2012; Brown et al. 2015). Hence, I constructed simple LCP models to reflect hypothesized dispersal routes between meadows. I built and compared 10 resistance

rasters: slope, vegetation, 1:1 slope/vegetation, 3:1 slope/vegetation, 1:3 slope/vegetation, ridges + 1:1 slope/vegetation, ridges + 3:1 slope/vegetation, ridges + 1:3 slope/vegetation, ridges + slope, and ridges + vegetation. Slope was modeled using a flat linear function where each 10% increment corresponded to values 1–10. Vegetation was modeled using a categorical function based on the Yosemite Vegetation Map (Keeler-Wolf et al. 2012), where each vegetation class was assigned a moisture index from 1–10 (Table S3.1). Ridges (slope >80%) were given pixel values of 1×10^6 to encode them as impenetrable. Cost rasters were resampled to 30 m for computational efficiency. LCP models and topographically corrected distances were calculated using the gdistance package (van Etten 2017) and customized functions for multicore processing. I applied an arcsine transformation for proportional data to normalize the dependent variable ($\sin^{-1}\sqrt{F_{ST}}$), and log-transformed the dispersal path distances. Only comparisons within each lineage were considered. The LCP model with highest significant partial mantel correlation between LCP distance and F_{ST} after accounting for Euclidean distance was buffered and used to extract all environmental data between all sampled meadows. Mantel tests were performed using 1000 random matrix permutations to assess significance using the vegan package (Dixon 2003).

Environmental data

Remotely sensed and USGS survey data were used to parameterize landscape genetic models. A total of 265 between-site and 319 at-site variables were initially collected from 10 sources (Tables 3.2, S3.2): California Basin Characterization Model

2014 (climate and climate change; Flint and Flint 2012; Flint et al. 2013), Daymet (climate; Thornton et al. 2018), Cal Fire (fire frequency; Cal Fire 2015), Yosemite NP (trails and roads), US Geological Survey (meadow network attributes), LANDSAT 5 (moisture and vegetation; US Geological Survey), MODIS (snowmelt timing; NASA), Shuttle Radar Topography Mission (topography; US Geological Survey), Yosemite LiDAR Vegetation Map (vegetation and hydrology; Keeler-Wolf et al. 2012), and Cal Water (watershed attributes; CDWR 2004). Climate change variables from the BCM 2014 data were calculated as a 60-year difference between two 30-year periods: 1980–2010 and 1920–1950. Variables included changes in mean values, as well as changes in the coefficient of variation (CV) of values. Daily MODIS images were derived from previous studies, (Dozier et al. 2008; Dozier and Frew 2009; Painter et al. 2009) and summarized in Octave v3.2.3 and ArcGIS v10.1. LANDSAT surface reflectance images were corrected using the LANDSAT Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm, which corrects raw images for water vapor, ozone, shadows, and other atmospheric aberrations (Masek et al. 2006; Schmidt et al. 2013). Only images taken between the months of May and September were used from LANDSAT, to better reflect the active season of Yosemite toads. Variables meant to represent linear barriers to dispersal (i.e. streams, roads, and trails) were rasterized with each pixel representing presence or absence of the barrier. Additionally, I tried weighting barriers (i.e. roads were weighted by level of traffic, trails by amount of use, and streams by order). All data were sampled in their original resolutions. Resampling is necessary

for creating combined resistance surfaces from multiple variables, however transect sampling extracts each variable independently without the need to combine rasters.

Between-site data were extracted from a LCP buffer distance of 100 m to minimize the influence of terrain outlying the hypothesized dispersal corridors, except for BCM climate data, which were extracted from a 500 m buffer. BCM climate data were only available in >100 m (270 m) resolution and expected to have broader spatial effect. In order to assess the improvement of LCP transects, all between-site data were also collected from Euclidean buffers of identical bandwidth. In previous transect-based landscape genetic studies, bandwidth choice had minimal impact on overall model fit (Murphy et al. 2010b; van Strien et al. 2012), and therefore I chose intermediate bandwidths most likely to encapsulate dispersal without testing multiple bandwidths directly. At-site environmental data were extracted from a 100 m (most data) or 500 m (BCM climate data) buffer around the mean sampling location for each USGS meadow sampled, except MODIS snowmelt and Daymet precipitation data which were extracted directly from meadow polygons. I hypothesized that certain at-site features would impact larval growth, which can impact population growth rate, and have downstream effects on rates of migration. All spatial data extractions were performed using the GDAL module (GDAL Contributors 2018) in a custom Python script (ZonalStatistics.py, Supporting Information).

I chose two variables from each of 11 categories for consideration in gravity models, for a total of 22 variables (Table 3.3). Variables were chosen based on highest variable importance in a random forests analysis. Random forests analysis is a

classification and regression tree method that estimates the importance and accuracy of an ensemble of decision trees (Breiman 2001). It can handle datasets containing many variables, their higher-order interactions, and multicollinearity while modeling a response variable. Variable importance was assessed using the "importance" parameter of the "randomForest" function from the randomForest package (Liaw and Wiener 2002), which evaluates the mean decrease in accuracy and mean squared error (MSE) when permuting each variable. Phylogenetic lineage (discussed below) was found to be the most important variable overall, because between-lineage F_{ST} values are higher than within-lineage F_{ST} values (see Results). To avoid phylogenetic bias in selecting variables, I calculated importance values without transects that cross lineage boundaries. I also excluded any variables that were collinear with other variables. Multicollinearity was assessed using recursive backwards selection until variance inflation factor (VIF) was less than 10 for each variable, using the fmsb package (Nakazawa 2018). Since elevation is often collinear with many other variables, I included elevation during VIF variable selection in all cases, to reduce this potential bias.

Landscape genetic analysis with gravity modeling

Gravity models are made to predict attraction (i.e. gene flow) between locations, while accounting for the non-independence of at- and between-site variables. For example, gene flow between sites A and B is not independent from gene flow between sites B and C. The conditions between each pair of sites (e.g. mean temperature) might partly explain their connectivity, but the conditions at site B (e.g. wetness) may influence

migrants traveling to both A and C. Gravity models were initially constructed following the general form of Anderson (1979) and implemented following Murphy et al. (2010b) and Dileo et al. (2014). Gravity models are a series of related models, most generically:

$$F_{ij} = m_i D_{ij}^{\alpha} R_{ij}^{\beta} S_i^{\gamma} \tag{3.1}$$

where F_{ij} represents the flow of gene copies between sites and m_i is a site-specific random intercept. D_{ij}, R_{ij}, and S_i represent the effects of distance, environmental resistance between sites, and production effects of sites respectively, and α, β, and γ are parameters to be estimated. However, I chose not to log-transform the data for three reasons: (1) models fit the data poorly using this transformation, and caused the residuals to strongly deviate from normality, (2) climate change variables included negative values, and (3) I wanted to fit random (lineage-specific) slopes as well as intercepts. Instead I fitted linear mixed models with the general form:

$$F_{ij} = m_i + \alpha D_{ij} + \beta R_{ij} + \gamma S_i \tag{3.2}$$

An arcsine transformation for proportional data was applied to normalize the dependent variable, and the resultant value was subtracted from 1.0 to represent connectivity instead of differentiation ($1 - \sin^{-1}\sqrt{F_{ST}}$). Sites with <5 observations were removed to decrease the standard deviations associated with each random grouping level. All variables were centered and scaled by subtracting the mean and dividing by the standard deviation.

Linear mixed effects models were constructed with the lme4 package (Bates et al. 2015) using three transect distance cutoffs for each: 10 km, 20 km, and 30 km. These three cutoffs were chosen in order to infer whether different environmental processes influence short, medium, and long-term dispersal, given that these take increasing

amounts of time (Anderson et al. 2010; Galpern et al. 2012). In addition, both Euclidean and LCP transects were modeled to assess whether LCP transect modeling substantially improves or changes models. Stepwise addition of each variable was performed using the variable that would produce the lowest model AICc, and a variable was retained only if a likelihood ratio test between it and its nested model was significant (Fig. 3.1).

In addition, I suspected that toads in separate lineages might have evolved different sensitivities to environmental effects, so I included random slopes by lineage for any environmental variable, if doing so significantly improved the model without over-fitting it:

$$F_{ij} = m_{ik} + \alpha D_{ij} + (\beta + \beta_k)R_{ijk} + (\gamma + \gamma_k)S_{ik} \quad (3.3)$$

Variable random slopes by lineage were added using the stepwise process described above. Marginal (fixed effects) and conditional (random + fixed effects) R^2 values were calculated using the MuMIn package. I simplified lineage assignment to the four main (i.e. pure, non-admixed) lineages (Y-North, Y-East, Y-South, Y-West; see Chapter 1) in YOSE to reduce the complexity of models.

Source-sink dynamics

At the local scale of meadow clusters (see above definition), there may be asymmetries in gene flow indicative of source-sink dynamics, possibly as a function of meadow environment. Long-term survey data suggests that toads breed at certain meadows much more reliably than in neighboring meadows, underscoring variable detection probabilities for the species (Brown et al. 2012, 2015; Fellers et al. 2015;

Ostoja et al., in prep.). I hypothesized that clusters of nearby meadows included "hub" meadows that receive and retain the majority of migrants, and lower quality "satellite" meadows, where toads are migrating over to hubs. I tested this hypothesis by estimating directional migration rates between meadows using the method of Sundqvist et al. (2016). In essence, this method calculates geometric means of allele frequencies for a theoretical pool of migrants between each pair of meadows using alleles present in both meadows. Then asymmetrical G_{ST} values are calculated by comparing each meadow to the migrant pool. Finally, migration rates are estimated from G_{ST} values: $N_e m \approx ((1/G_{ST}) - 1)/4$ (Wright 1931); values are divided by the maximum, and pairwise estimates are normalized to get a unitless measure of net relative migration δM. Relative values are less sensitive to assumptions of the island model, such as drift-migration equilibrium, because they only infer differential migration without explicitly modeling any parameter. For any comparison between two meadows, the meadow with $\delta M < 0$ has net immigration, and meadows with $\delta M > 0$ have net emigration. For each meadow cluster I examined, δM estimates were mapped onto igraph (Csárdi and Nepusz 2006) objects to display the extent of asymmetrical migration in each cluster of meadows.

I evaluated 10 meadow-scale environmental attributes as potential correlates of hubs and satellites (Table 3.4). Six variables were meadow network attributes: log meadow area, meadow perimeter, meadow area within 1 km, wet area of toad breeding meadows within 1 km, wet area of closest meadow, and the network-boosted probability of toad breeding (Keeler-Wolf et al. 2012; Berlow et al. 2013). The other four variables were topographic: mean slope, mean roughness index, standard deviation (SD) of

elevation, and mean slope position index (digital elevation model from SRTM, analyses performed in ArcGIS v10.1). Any variables that did not approximate normality were transformed using either logarithmic, exponential, or root transformations. I identified environmental correlates of net migration (δM) individually, using mantel correlations as implemented by the ncf package (Bjørnstad 2018), with 1000 permutations for significance. I then identified correlations collectively for all variables using a multiple regression on distance matrices (MRDM; Legendre et al. 1994; Lichstein 2007), using the method of Wang (2013), and 1000 permutations.

I also gave "hubs" and "satellites" formal definitions based on extreme values of net migration. Hubs were defined as meadows with average δM that is negative, and at least one SD less than zero. Satellites were defined as meadows with average δM that is positive, and at least one SD greater than zero. I performed a principal components analysis (PCA) and a linear discriminant analysis (LDA) using the MASS package (Venables and Ripley 2002) to identify environmental predictors of the two types. In order to determine the most important discriminatory variables, I performed stepwise variable selection to optimize the accuracy of predictors using the stepclass function of the package klaR (Weihs et al. 2005). The process uses 10-fold cross-validated accuracy of the LDA to add or remove variables until improvement is less than 5%.

RESULTS

Which are populations: meadows or meadow clusters?

I found a strong pattern of isolation by distance up to 25 km distances (Fig. 3.2), suggesting that population structure might be heavily influenced by stepping-stone migration at multiple spatial scales. However, the hierarchical structure analysis revealed significant structure below the level of lineages, with most meadows being genetically distinct populations (Fig. 3.3). Meadows in fused lineages (i.e. with inter-lineage admixture, see Chapter 1) were distinct at high levels in the hierarchy: e.g. Isberg Pass (East-South-A zone), the northern admixture area near Miller Lake and Mule Pass (East-North-A2), and near Slide and Rodgers Canyons (East-North-A1). The results of ADMIXTURE analysis delineated units of genetic cohesion intermediate between meadow and lineage (Fig. 3.4). The lowest ADMIXTURE cross-validation error for Yosemite was found at K=28, and for Kings Canyon was K=6, although subsequent K values elicited only gradual increases in cross-validation error. Most meadows (>50%) were contained in clusters with two to six meadows, but some (e.g. Tioga) had as many as 12. Out of 34 clusters, 16 (47%) contained only one meadow, and the mean cluster size was three meadows. Many singleton clusters were in terrain separated by canyons, such as Rancheria, Thompson, Wells, Tilden, and Twin Lakes in Y-North. In other cases, including Wawona and Summit in Y-South, and Bald and Ribbon 2 in Y-West, and to an extent Cockscomb, Polly, and Tressider in Y-East, topography appeared to be a minor barrier.

The hierarchical AMOVA found a large proportion of total variance between parks (30.45%; ϕ_{PT} = 0.31) as expected (Table 3.5). Meadows within clusters showed the next highest variance (23.43%) and had the highest ϕ-statistic (ϕ_{MC} = 0.41), supporting the idea that meadows structure current populations more than ADMIXTURE meadow clusters (3.82%; ϕ_{CL} = 0.062). Further, the estimated threshold for demographic independence among meadows (assuming 10% dispersal rates, and using the observed meadow N_e of 29.13) was $F_{ST} \geq 0.079$, a value far below the observed statistic ϕ_{MC} = 0.41. In contrast, the estimated threshold for demographic independence among clusters was $F_{ST} \geq 0.097$ (using average cluster N_e of 23.33), a value greater than the observed statistic ϕ_{CL} = 0.062. Patterns of conformity to Hardy Weinberg expectations at the meadow and cluster scales further supported meadow distinctiveness, with meadows deviating at 2.48% of genotypes, compared to 4.86% for meadow clusters, a nearly two-fold difference (Fig. S3.1).

Environmental drivers of source-sink dynamics

I first examined source-sink dynamics in meadow clusters containing at least five meadows, to get an appreciable sample size. Out of these nine meadow clusters, source-sink dynamics were strong in six (67%) based on the presence of at least one hub or satellite meadow (Fig. 3.5). Most meadow clusters contained a few highly diverse hub meadows, with high H_O and much stronger genetic immigration than emigration. Out of nine clusters examined, three had a single hub meadow, and the other three had two to three hubs. Distinct satellite meadows were only identified in three out of nine clusters.

There were environmental patterns characteristic of hubs and satellites, regardless of meadow cluster size (i.e. including <5 meadows). Compared with satellite meadows, hub meadows (defined as having δM values ≤ 1 SD below zero) had significantly larger mean areas (t=-3.34, df=20, p=0.002), perimeters (t=-2.93, df=21, p=0.004), and breeding wet area within 1 km (t=-1.94, df=18, p=0.03) (Fig. 3.6). Hubs also had lower mean slope within 500 m of the sampled location (t=1.77, df=15, p=0.048). Interestingly, although hubs had larger areas and larger nearby wet areas, their nearest neighboring meadow had significantly less wet area (t=3.52, df=22, p<0.001). All tests remained significant after correcting for multiple comparisons with the FDR method (maximum p-value: 0.048).

I also found relationships between δM and each of the 10 environmental variables (Fig. 3.6). For all 10 variables examined (Table 3.4), there was a significant mantel correlation between δM and δ(environment). This result remained after correcting for multiple comparisons (maximum p-value: 0.007). The general result can be summarized as follows: for any two meadows in a cluster, the meadow with larger area and less topographic ruggedness tends to have net immigration (influx of gene copies). Interestingly, this included a small but significant positive correlation between δM and δ(slope position index) meaning that meadows closer to hilltops have net genetic influx (ρ =0.17, df=198, p<0.001). I also found a small but significant positive correlation between δM and δ(breeding probability), indicating that meadows with higher probability of breeding have net genetic influx (ρ=0.17, df=198, p<0.001). The MRDM results supported the above pattern of meadow area strongly influencing source-sink dynamics, and meadow ruggedness having a small to moderate influence. I found that

meadow area explained the largest percentage of the variance (29.8%), followed by wet area of the closest meadow (7.0%), while the remaining variables collectively only accounted for 6.6% of the variance (R^2_{adj}=0.41, F(10,189)=14.52, p<0.001). Five variables had significant coefficients in the model: meadow area (t=4.21, p<0.001), wet area of the closest meadow (t=-4.12, p<0.001), breeding wet area within 1 km (t=3.06, p=0.003), meadow slope (t=2.25, p=0.031), and meadow roughness index (t=2.12, p=0.046). Although roughness index had a positive coefficient in the MRDM model, both δ(roughness index) and δ(slope) had negative mantel correlations with δM.

The LDA function based on all 10 variables had a 90.0% predictive accuracy rate, including 91.7% for hubs, and 87.5% for satellites (Fig. 3.6). The stepwise variable selection for LDA identified two variables that significantly improved discrimination between hubs and satellites: meadow area, and wet area of the closest meadow. The maximum cross-validated rate of accuracy was 62.9%, and the predictive accuracy rate for just those two variables was 90.0%. The first two principal components of the PCA accounted for 74.7% of the variance, and this variance was orthogonally split into the two categories of variables: meadow network attributes, and topography (Fig. 3.6). Hubs appeared to be more variable than satellites for meadow network attributes, whereas satellites appeared to be more variable than hubs for topography. Using F-tests for homogeneity of variances, I confirmed that hubs have significantly greater variance for breeding wet area within 1 km (F=0.35, df=11, p=0.048); satellites had significantly greater variance for slope (F=4.90, df=11, p=0.006), roughness index (F=3.77, df=11, p=0.02), and standard deviation of elevation (F=5.27, df=11, p=0.005).

Meadow, landscape, and climate effects on connectivity

In this study, I combined elements of LCP model selection, least cost transect analysis (LCTA), and gravity modeling into a single workflow in order to build environmental models of genetic structure that are more predictive and biologically plausible than other modeling techniques (Fig. 3.1). In the first step of my three-step process, I found that the LCP model with 3:1 weight of slope/vegetation and ridgeline barriers had the highest significant partial mantel correlation with F_{ST} after accounting for Euclidean distance (ρ=0.176, df=4005, p=0.040; Table 3.6). Although one other model had a significant partial mantel test (slope), this model accounted for less of the matrix covariance than the best model (ρ=0.164, df=4005, p=0.021). I extracted all between-site environmental variables from a buffer around the best model of adult toad dispersal, which assumes that slope and ridgelines impact dispersal primarily, and wet vegetation plays a smaller role.

I initially used random forests on all variables to ascertain how much of the variance could be explained, and which variables were globally most important for Yosemite toad connectivity. The random forests analysis initially found phylogenetic lineage as the most important overall variable, likely because between-lineage F_{ST} values are higher than within-lineage F_{ST} values (Fig. 3.2b). Therefore I repeated the analysis without any between-lineage pairwise comparisons to avoid biasing the variable selection process. Using all variables (265 between-site, 319 at-site, 324 unique), and 500 classification trees, the global model explained 83.3% of the total variance. I assessed

variable importance as the total decrease in residual sum of squares from splitting on the variable, averaged over all trees. The best five global variables were (in order, with importance metric): mean elevation between sites (0.12), SD of May burn ratio between sites (0.09), proportion of coniferous vegetation between sites (0.09), 60-year change in CV of potential evapotranspiration between sites (0.06), and the 60-year change in mean precipitation at sites (0.06); see Tables 3.3 and S3.2 for complete variable definitions. Interestingly, these last two out of the top five most important variables were measures of climate change. After identifying the best two variables in each of the 11 categories (Table 3.3), the maximum VIF for any variable included in a model was 5.96.

Finally, I constructed landscape genetic models using 30 km, 20 km, and 10 km cutoffs for between-site corridors, in order to represent processes occurring over different spatial scales (and temporal scales, given that toad dispersal over space is limited by time). Out of 4005 possible pairwise corridors (90 meadows), the 30 km model included 1298 corridors, the 20 km included 870 corridors, and the 10 km included 320 corridors. In all cases, the best LCP model outperformed the best Euclidean model, based on AICc scores and marginal (fixed-effects) R^2 (Table 3.7). Connectivity was best explained by environment at 30 km cutoffs ($R^2_m = 0.58$), followed by 10 km cutoffs ($R^2_m = 0.41$), and then 20 km cutoffs ($R^2_m = 0.37$). Fixed and random residuals from all models appeared to be independent, homoscedastic, and identically distributed (Figs. S3.2–4).

In the best environmental models, slope between sites was the most important predictor of connectivity at all spatial scales, with lower slope causing higher connectivity (Table 3.7; for values of variable coefficients, see Table S3.3). The exact

relationship had a random effect that depended on lineage at 30 km and 10 km scales. Between-site elevation was the second-most important predictor at 30 km and 20 km, with higher elevation facilitating connectivity at these scales. However at 10 km, elevation inhibited connectivity for all lineages except Y-North. Both roads and trails were found to be barriers at all scales and across all lineages, a result with important implications discussed below. Roads had lineage-specific effects at 20 km, where connectivity of the two lineages bisected by CA-120 was strongly impacted (Y-West), or weakly impacted (Y-East). Roads had the strongest impact on Y-South, a genetically depauperate lineage that is bisected by Glacier Point Road, and virtually no impact on Y-North, which is not adjacent to any roads. Finally, the proportion of conifer cover between sites generally inhibited connectivity, but actually facilitated it for the two lower-elevation lineages (Y-South, Y-West) at 20km, and for Y-West at 30 km.

I found that variables relating to wetter conditions had a mostly inhibitory effect on connectivity, a counterintuitive result that is discussed below (Table 3.7). For example, dispersal paths with either a 60-year increase in mean runoff, or larger proportion of water between sites, inhibited connectivity at all scales. At-site effects were all related to moisture in the best models, and moisture had the effect of reducing gene flow away from meadows (genetic emigration). At 10 km, higher inter-annual meadow precipitation (Daymet) was the second largest inhibitor of connectivity, after slope between sites. At 20 km and 30 km scales, meadows with a higher proportion of intermittently to permanently flooded area also reduced connectivity. These two variables were not independent. For example, meadows with > median precipitation had an average

of 15.0% flooded area, significantly more than those < median precipitation, which had an average of 8.7% flooded area (t=-2.34, df=86, p=0.02). These higher-precipitation meadows were found to be 5.7 times smaller (t=3.84, df=45, p<0.001), with 3.8 times smaller N_e (t=4.14, df=43, p<0.001), surrounded by 4.1 times more coniferous vegetation (t=-8.55, df=87, p<0.001), and 468.1 m lower in elevation (t=10.04, df=73, p<0.001). Additionally, meadows with a 60-year increase in average inter-annual precipitation (BCM) decreased connectivity at 20 km. (The sign of this negative relationship is obscured by multicollinearity in the full model (Table 3.7), however using a univariate regression the coefficient was negative.) Although this general pattern suggested that more hydric conditions inhibit connectivity, I note the seemingly contravening result that more variable hydric conditions also may inhibit connectivity. Specifically, higher spatial variability of two variables within dispersal corridors decreased connectivity: summer moisture index, and May burn ratio (an index of soil moisture). However, this result may not contravene the general pattern of wetter conditions inhibiting connectivity if it reflects a separate process (see Discussion).

Based on these results, I hypothesized that superior (wetter) breeding conditions may improve larval viability, increasing likelihood of metamorphosis, and ultimately causing higher site fidelity. Using the polynomial regression model of tadpole growth and development from Chapter 2, I tested for a difference in mean tadpole size among low rainfall (< median) and high rainfall (> median) meadows. An ANCOVA showed that tadpoles from meadows with higher-than-median precipitation (Daymet) were larger compared to meadows with lower-than-median precipitation, (R^2_{adj}=0.484,

$F(3,1701)=534.7$, $p<0.001$; Fig. 3.7). A Tukey Honest Significant Differences test showed that the difference was significant, and higher-precipitation tadpoles are 0.66 ± 0.38 mm longer for a given developmental stage than lower-precipitation tadpoles. Presumably tadpoles at these meadows are less threatened by early desiccation, and have longer growing periods, better nutrition, or higher insolation, leading to larger body size at metamorphosis. This hypothesis is further discussed below.

DISCUSSION

Are meadows an effective unit for population management?

Identifying the unit at which a population is ecologically and evolutionarily independent is crucial to properly allocating conservation resources. Management Units (MUs) are usually defined as demographically independent populations with sufficiently high growth rates and low connectivity to neighboring units (Palsbøll et al. 2007). In practice, MUs are delineated based on genetic discontinuity or local adaptation to unique environmental conditions (Dizon et al. 1992). Moritz (1994) defined MUs as having significant nuclear or mitochondrial allele frequency divergence, representing present-day genetic or demographic cohesion. The category of MUs is used to reflect present-day population boundaries, and should be contrasted with Evolutionarily Significant Units (ESUs), which are lineages reflecting long-term or ancient isolation sufficient to create fixed genetic or morphological differences, and perhaps even reproductive isolation.

There are many technological challenges involved in accurately delineating MU boundaries. For example, it has been acknowledged that different processes underlie

demographic and genetic connectivity (Slatkin 1987; Bohonak 1999; Lowe and Allendorf 2010). The number of migrants ($N_e m$) is solely responsible for genetic connectivity, while demographic connectivity is influenced by m relative to birth and death rates. Additionally, the flow of gene copies and individuals can occur at dramatically different rates, a phenomenon known as Slatkin's Paradox (Bohonak 1999; Yu et al. 2010). Finally, many tests for population structure or differentiation implicitly assume equilibrium between the forces of genetic drift and gene flow. Many conditions likely violate this assumption in Yosemite toad populations, such as temporally fluctuating gene flow in metapopulations, recent colonization from a source population, or stepping stone migration preventing the formation of Hardy-Weinberg gene pools. For these reasons, and due to individual limitations of programs, I used multiple methods to assess MU boundaries. I encourage future studies to complement my results with demographic (i.e. dispersal and growth rates) and adaptive (i.e. experimental translocations) evidence of MU boundaries.

A priori, I assumed that the unit of a Yosemite toad population is no smaller than one meadow, because the species demonstrates high vagility at that scale. Adult toads are known to disperse an average of 275–279 m (maximum 1136 m) Martin 2008; Liang 2010) outside breeding meadows in only weeks, and live 12–15 years (Sherman and Morton 1984). In contrast, the mean diameter of meadows occupied by toads in YOSE is 266 m. Shaffer and Fellers (2000) found significant among-"pond" (their definition implying "meadow") variation using an AMOVA on a SSCP marker and suggested that "ponds" (meadows) be the unit of management. In this study, I first used a hierarchical

STRUCTURE analysis to ascertain that meadows are typically distinct, with notable exceptions such as the Tioga cluster (Fig. 3.3). Next I used ADMIXTURE to show that clusters of meadows also show population cohesion (Fig. 3.4). In order to assess which scale best reflects MU boundaries I performed a hierarchical AMOVA, which showed that meadows globally cause the largest reduction in heterozygosity (Table 3.5), and assuming equilibrium conditions they have a mean dispersal rate of far less than 10%. Tests for Hardy-Weinberg genotype frequencies also corroborated this pattern, showing that meadow clusters violated expected frequencies at a much higher rate than individual meadows. Collectively, these results suggest that meadows are the most appropriate unit for MUs. However this does not negate the importance of source-sink structure (discussed below).

Alternative methods for delineating population boundaries have advantages, but would have substantial drawbacks given my sampling scheme. For example, one alternative approach to genotype clustering methods is testing for pairwise differentiation between units, such as Fisher's exact tests on contingency tables of allele counts. That method is particularly robust to high migration, isolation by distance, and unbalanced sampling, however small sample sizes can have high Type II error rates (Goudet et al. 1996; Waples and Gaggiotti 2006). This would likely be the case in this study ($\bar{n}$ = 6.1 diploid samples per meadow). Genetic clustering algorithms can be sensitive to fine-scale divergence because they optimize Hardy-Weinberg simultaneously across many loci; however, small deviations from linkage or migration-drift equilibrium (as are expected in Yosemite toads due to isolation by distance) can result in substructure not being detected

(Schwartz and McKelvey 2009; Lowe and Allendorf 2010. My hierarchical clustering analysis has two major advantages to a more general Bayesian clustering method: (1) it overcomes the technical issue mentioned above, whereby stepping stone migration between meadows can prevent typical clustering analyses from detecting substructure, and (2) the results display the order of meadow distinctiveness, and thus allow unique meadows to be identified (Fig. 3.3). Interestingly, many of the most distinct groups of meadows have mixed ancestry (i.e. from two lineages), including meadows from the two major contact zones: seven meadows near Slide Canyon, Rodgers Lake, and Mule Pass (i.e. East-North-A1), three meadows near Miller Lake (i.e. East-North-A2), and two meadows near Isberg Pass (i.e. East-South-A). Previous research (Chapter 1) suggests these admixed areas have been admixing for multiple generations and are therefore stable, producing viable toads.

In contrast to these distinct meadows, recent colonization or high gene flow is likely overpowering meadow differentiation near Tioga, Conness, Ireland, and White Wolf (Fig. 3.3). Two out of four of these undifferentiated clusters (Tioga and White Wolf) are on both sides of a major highway (CA-120), which has been identified as a barrier to connectivity (discussed below), and hence it is important for future work to estimate individual migration rates among these MUs. This could be approached in three possible ways: (1) these analyses could be repeated with larger sample sizes using markers with higher mutation rates (e.g. microsatellites), (2) simulation, Bayesian, or approximate Bayesian methods could be used to estimate migration rates, or (3) a tagged genetic recapture study could be performed to estimate both dispersal/migration rates.

Natural and anthropogenic limitations on Yosemite toad connectivity

Predictive spatial models in landscape genetics and genomics are an important, burgeoning class of tools for summarizing evolutionary processes, and streamlining decision-making (Vandergast et al. 2008; Sork et al. 2010; Fitzpatrick and Keller 2015). In this study, I provided a novel network method for individually modeling variables of interest (Fig. 3.1), which fills an important void in the landscape genetic toolbox. Many studies generate hypothesized "resistance rasters" of environmental variables (McRae 2006; Zeller et al. 2012), but these are: (1) not well-suited to multivariate hypotheses, (2) biased by the subjectivity of translating raw values into resistance values (Shirk et al. 2010; Spear et al. 2010). Moreover, this approach often does a poor job of exploring total parameter space, and may converge on the wrong environmental model (Graves et al. 2013). The alternate approach, transect analysis, does extract raw values of variables from rectilinear corridors, but at the cost of losing biologically meaningful dispersal corridor locations (van Strien et al. 2012; van Strien 2013). My method combines the biological plausibility of LCP-optimized dispersal corridors with a multivariate modeling process of raw environmental variables, making inferences potentially more realistic.

My models revealed that both climate and climate change (60-year shifts in parameter values) have contributed to Yosemite toad genetic structure. First, during the variable selection process, random forests identified two variables within the top five most important (out of 500+): 60-year change in CV of potential evapotranspiration between sites, and the 60-year change in mean precipitation at sites. Second, a host of

variables relating to climate variability (e.g. variability in summer moisture index) and climate change (e.g. 60-year change in mean runoff) negatively impacted connectivity at multiple spatial scales (Table 3.7). Recently, network methods have shown that the number of Sierra Nevada meadows that offer refugia from climate (i.e. less climate variability) is projected to decline, and this will reduce connectivity across multiple meadow-dwelling species (Maher et al. 2017). Furthermore, this concept was applied to the meadow-dwelling Belding's ground squirrel (*Urocitellus beldingi*), which indeed showed a relationship between climate refugia and network genetic connectivity among meadows (Morelli et al. 2017). My results may provide the first detailed look at how climate change is specifically fragmenting Yosemite toad meadows, but may also offer an opportunity to pinpoint refugial meadows for the species.

Slope and elevation were ostensibly the most important landscape predictors of genetic connectivity regardless of spatial scale (Table 3.7). Wang (2012) found a similar pattern using variance partitioning on microsatellite markers. At all spatial scales, lower slope enhanced connectivity, and at larger spatial scales (20 and 30 km cutoffs), dispersal via higher elevation enhanced connectivity. Higher elevation actually inhibited connectivity at shorter (10 km) transect lengths, for all lineages except Y-North (Table S3.3), the lineage with the most spatially isolated meadows. This region of YOSE is defined by a series of glacially carved canyons with steep headwalls (see Figs. 1.2 or 3.4) that toads likely traverse only after extensive horizontal and vertical dispersal. Possibly, Yosemite toads in Y-North are behaviorally biased toward upland dispersal that maximizes the likelihood of encountering suitably wet, subalpine breeding meadows.

Alternatively, during longer excursions between canyons that require crossing lower elevations, toads might be more likely to encounter and breed in marginal non-meadow habitat (e.g. springs, stream channels, lakes, or bogs). Breeding in such inter-meadow habitat would result in lower meadow-to-meadow migration rates, and decrease the probability of toads reaching any particular meadow destination. This hypothesis is supported by the fact that wetter conditions between meadows (60-year increase in mean runoff, and proportion of water between sites) inhibited connectivity across all models, which means that dispersal routes containing non-meadow breeding habitat might be important for toad "layovers." Several such observations of Yosemite toad larvae or subadults in streams and lakes have previously been made (P. Maier, R. Knapp, S. Lee, pers. comm.). I also found that higher variability in summer moisture index, and May burn ratio (an index of soil moisture), decreased connectivity. Dispersal routes with unpredictable desiccation or fire regimes might increase the mortality of migrants.

Conifers increased connectivity among the low-elevation lineages (Y-West and Y-South), which are dominated by California red fir and Sierra lodgepole pine dispersal habitat. Conifers had the opposite effect for both high-elevation lineages (Y-East and Y-North) where meadows are primarily bordered by talus or sparsely vegetated whitebark pine forest. This pattern is consistent with divergent environmental selection causing neutral genetic divergence, or "isolation by adaptation" (IBA; Nosil et al. 2008). In this case, the specific type of IBA is unique in that low- and high-elevation adult toads are presumably isolated not by mating preferences or phenology, but rather dispersal preferences. Theoretical models predict that divergent or heterogeneous environments

should select for higher dispersal rates within the patch type that locally has a higher carrying capacity (McPeek and Holt 1992). If local adaptation of dispersal strategies occurs across a large enough region (e.g. Y-East), selection should overpower gene flow and prevent dispersal across contrasting environmental gradients such as presence or absence of thick conifer forest (Gros et al. 2006). Another possible explanation for opposite signs in conifer random coefficients is that each lineage resides almost entirely in conifer or subalpine habitat, and inter-lineage phylogenetic signal is erroneously measured as conifer-mediated isolation by adaptation. This latter possibility seems unlikely given the patchwork of environmental heterogeneity within and among regions. For all of my non-stationary results that imply possible adaptations in dispersal by lineage (i.e. IBA for slope, elevation, conifers, etc.; Tables 3.7, S3.3), other explanations are possible. The most effective way to disentangle scenarios of isolation by distance, isolation by adaptation, and stochastic divergence such as "isolation by colonization" (IBC) is to compare genomic signals of divergence at neutral and non-neutral markers with morphological or behavioral biases in dispersal capacities (Orsini et al. 2013).

Roads and trails were significant predictors of reduced Yosemite toad connectivity in all models. Tioga Road (CA-120) bisects two lineages, Y-West and Y-East, and both regions include meadow clusters that cross this highway ("Tioga" and "White Wolf"; Figs. 3.4, 3.5). Interestingly, the road apparently has a very large negative influence on Y-West connectivity, and a comparatively minor impact in Y-East. However, larger intrinsic growth rates, population sizes, and $N_e m$ in Y-East might mask considerable road mortality rates. Indeed, many Yosemite toad museum specimens were

Y-East individuals collected on or near Tioga Road (VertNet 2018). The impact of roads and trails is unsurprising given the magnitude of recreation Yosemite experiences each year. In 2017, 4,336,890 people drove into Yosemite National Park, 167,507 of whom hiked on trails and camped in the backcountry (National Park Service 2018). This user load has gradually accumulated to 125-fold the number of visitors one century ago. Automobiles began driving over Tioga Road in 1900, and the route became popular for businessmen beginning in the 1920s (Trexler 1975). The trail system was slowly built up from existing Native American trails during the 1860s by the U.S. Geological Survey of California, and later in the 1920s and 1930s by the Sierra Club and National Park Service (Bingaman 1968). Future mark-recapture studies might reveal that these linear barriers also impact demographic connectivity.

Meadow attributes impact network connectivity

This landscape genetics study is among the first to isolate source-specific environmental effects on genetic connectivity (Murphy et al. 2010b; Dileo et al. 2014), and the first to accomplish this goal for Yosemite toads. I found two clear environmental patterns: one based on gravity modeling results, and the second based on source-sink (hub-satellite) modeling.

First, my gravity modeling analysis found that wetter (i.e. more precipitation, flooded area) meadows at lower elevation have reduced genetic emigration (Table 3.7) and, perhaps as a consequence, these meadows have lower genetic diversity. This effect was present at all scales, but was especially strong for local (10 km) transects in gravity

models. Geographically, these low-elevation meadows are primarily in the Y-West and Y-South lineages, where breeding is most inconsistent (Berlow et al. 2013), and both heterozygosity and N_e are lowest (Chapter 1). However, tadpoles in these montane meadows grow more slowly, allowing them to metamorphose at greater size (Fig. 3.7). Possibly, these tadpoles are less threatened by early desiccation, and have longer growing periods, better nutrition, or higher insolation, leading to larger body size at metamorphosis. Additionally, more resources for fewer tadpoles could reduce larval resource competition, effectively increasing larval carrying capacity (K). Populations under carrying capacity generally produce fewer migrants (McPeek and Holt 1992), and possibly foster higher site-fidelity. The rate of emigration is primarily determined by the ratio of population size (N) to K (Dasmann 1981; Pflüger and Balkenhol 2014). Body size at metamorphosis is an important predictor of successful recruitment (Smith 1987; Altwegg and Reyer 2003), and more consistently high quality meadows might have selected for higher site-fidelity. The precise relationship between hydrology, elevation, and lineage remains to be described. Regardless, it is clear that Yosemite toads are much more isolated in montane meadows than in subalpine ones, and ultimately have lower genetic diversity, and increased genetic structuring (Figs. 3.3, 3.4; Chapter 1). This is in spite of any putative hydrological benefit to larval development at lower elevation.

Second, my source-sink analysis found that larger and topographically flatter meadows predictably (with 90% accuracy) act as regional genetic hubs (Fig. 3.5, 3.6), and often these meadows are centrally positioned in the spatial arrangement of these "metapopulations." This means that one or several meadows within each cluster tend to

receive a disproportionately large amount of genetic input, and send a disproportionately small amount of genetic output. Hubs tended to be not only bigger and flatter, but have more wet area nearby, and be positioned closer to hilltops than satellites. Aside from large meadow area, the second best discriminator of hubs was that they neighbored a meadow with minimal wet area. This may underscore the importance of juxtaposing breeding habitat with more xeric upland meadow habitat, which plays a vital role in Yosemite toad foraging and hibernation (Sherman 1980; Morton 1981; Martin 2008). Contrasting environmental correlates have previously been found for persistent vs. intermittent meadows; proximity to fire perimeters and timber harvest were predictors of intermittently occupied meadows in one study (Liang and Stohlgren 2011). However the present study offered the first insight into genetic source-sink structure, with 10 m resolution of meadow network attributes that predict hubs and satellites. High hub genetic diversity appears to reflect genetic admixture; multiple satellite meadows surround them, and promote the mixing of genetically diverse toads in a central area. The ultimate cause for this is unclear: perhaps temporally fluctuating habitat quality or low carrying capacity in satellite meadows periodically sends migrants seeking refuge in hub meadows (Holt et al. 2004). By the same token, large hub meadows may act as genetic sinks because breeding toads show higher rates of philopatry within superior breeding conditions. I also cannot rule out the possibility that hubs are actually source populations, but with excess private alleles. The premise of the private allele method used in this study is that populations with more unique alleles are more isolated. However, hubs may simply be larger and more stable populations that have accumulated more diversity, and the

homogeneity of satellite populations may be due to recent founder effects (Sundqvist et al. 2016). This alternative is not easily disentangled using a single genetic dataset, and is better tested by tracking studies. Regardless, hub admixture is a common pattern in my study. Since I have previously identified the pervasiveness of admixture on a larger scale (i.e. between lineages; Chapters 1 and 2), it is unsurprising to find admixture plays an important role at a smaller scale. This result comes with the caveat that many clusters have few meadows ($\bar{n}$ = 3), and I focused my analysis on those with >5 meadows. Therefore, there are also many examples of isolated meadows existing without a large, admixed meadow nearby.

CONCLUSIONS AND RECOMMENDATIONS TO MANAGEMENT

In this study, I introduced a new landscape genetic workflow that benefits analysis of patch-limited amphibians with biphasic life histories. My genome-wide marker set of thousands of haplotypic loci sequenced for 644 individuals in YOSE and KICA showed that most meadows are highly genetically distinct in accord with previous studies, but approximately half (47%) of sites contained multiple meadows with hub-satellite structure. Hub meadows were distinguishable 90% of the time, based primarily on meadow size and proximity to upland habitat. There is a strong indication that climate change is already influencing meadow connectivity, and is a central threat to the future persistence of the species. Future studies should attempt to pinpoint meadows that are either susceptible or resistant to the erosion of genetic connectivity by climate instability, using graph theoretical techniques for example (see Chapter 4). I found some evidence

that wetter, low-elevation meadows are less connected to each other in spite of larger tadpole growth, which could mean that lower-elevation toads have higher meadow fidelity at the expense of genetic diversity. My gravity modeling analysis affirmed the impact of slope on Yosemite toad connectivity, and found that other topographic (elevation), vegetation (conifers), hydrological (precipitation and meadow wetness), and disturbance (roads and trails) variables are important too. Road mortality should be studied using genetic tagging and mark-recapture work given that Tioga and Glacier Point Roads intersect three lineages with highly interconnected meadows. Crossing hiking trails between meadows also seems to increase migrant mortality in Yosemite. The most genetically depauperate lineage of toads studied to date is the "Evolution" lineage in KICA, which is adjacent to the PCT. This area deserves future attention to mitigate recreational impacts on toad breeding and dispersal habitat. In the future, my results may be effectively combined with corridor design methods (Chetkiewicz et al. 2006; Beier et al. 2008; Sawyer et al. 2011) to help maintain species persistence.

REFERENCES

Alexander, D. H., and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:1–6.

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Alexander, M. A., and J. K. Eischeid. 2001. Climate variability in regions of amphibian declines. Conserv. Biol. 15:930–942.

Altwegg, R., and H.-U. Reyer. 2003. Patterns of natural selection on size at metamorphosis in water frogs. Evolution 57:872–882.

Anderson, C. D., B. K. Epperson, M. J. Fortin, R. Holderegger, P. M. A. James, M. S. Rosenberg, K. T. Scribner, and S. Spear. 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. Mol. Ecol. 19:3565–3575.

Anderson, J. 1979. A theoretical foundation for the gravity equation. Am. Econ. Rev. 69:106–116.

Baguette, M., and H. Van Dyck. 2007. Landscape connectivity and animal behavior: functional grain as a key determinant for dispersal. Landsc. Ecol. 22:1117–1129.

Barrowclough, G. F. 1978. Sampling bias in dispersal studies based on finite area. Bird-Banding 49:333–341.

Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67:1–48.

Beier, P., D. R. Majka, and W. D. Spencer. 2008. Forks in the road: choices in procedures for designing wildland linkages. Conserv. Biol. 22:836–851.

Berlow, E. L., R. Knapp, S. M. Ostoja, R. J. Williams, H. McKenny, J. R. Matchett, Q. Guo, G. M. Fellers, P. Kleeman, M. L. Brooks, and L. Joppa. 2013. A network extension of species occupancy models in a patchy environment applied to the Yosemite toad (*Anaxyrus canorus*). PLoS One 8:e72200.

Bingaman, J. W. 1968. Pathways: a story of trails and men. End-Kian Publishing Company, Lodi, CA.

Bjørnstad, O. N. 2018. ncf: spatial covariance functions. R package version 1.1-7.

Bohonak, A. 1999. Dispersal, gene flow, and population structure. Q. Rev. Biol. 74:21–45.

Bradford, D., and M. Gordon. 1994. Acidic deposition as an unlikely cause for amphibian population declines in the Sierra Nevada, California. Biol. Conserv. 69:155–161.

Breiman, L. 2001. Random forests. Mach. Learn. 45:5–32.

Brown, C., M. P. Hayes, G. A. Green, D. C. Macfarlane, and A. J. Lind. 2015. Yosemite toad conservation assessment. USDA Forest Service report. Sonora, CA.

Brown, C., K. Kiehl, and L. Wilkinson. 2012. Advantages of long-term, multi-scale monitoring: assessing the current status of the Yosemite toad (*Anaxyrus canorus*) in the Sierra Nevada, California. Herpetol. Conserv. Biol. 7:115–131.

Brown, C., and A. R. Olsen. 2013. Bioregional monitoring design and occupancy estimation for two Sierra Nevadan amphibian taxa. Freshw. Sci. 32:675–691.

Busch, J. D., P. M. Waser, and J. A. Dewoody. 2009. The influence of density and sex on patterns of fine-scale genetic structure. Evolution 63:2302–2314.

Cal Fire. 2015. Fire perimeters. Department of Forestry and Fire Protection, Fire and Resource Assessment Program. Available online at: http://frap.fire.ca.gov/data/frapgisdata-sw-fireperimeters_download

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

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes, Genomes, Genet. 1:171–182.

CDWR. 2004. CalWater 2.2.1 Watershed Boundaries. Department of Forestry and Fire Protection. Available online at: http://frap.fire.ca.gov/data/frapgisdata-sw-calwater_download

Chetkiewicz, C.-L. B., C. C. St. Clair, and M. S. Boyce. 2006. Corridors for conservation: integrating pattern and process. Annu. Rev. Ecol. Evol. Syst. 37:317–342.

Clobert, J., J. F. Le Galliard, J. Cote, S. Meylan, and M. Massot. 2009. Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecol. Lett. 12:197–209.

Csárdi, G., and T. Nepusz. 2006. The igraph software package for complex network research. InterJournal Complex Syst. 1695:1–9.

Cushman, S. A., K. S. McKelvey, J. Hayden, and M. K. Schwartz. 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am. Nat. 168:486–499.

Cushman, S., T. Wasserman, E. Landguth, and A. Shirk. 2013. Re-evaluating causal modeling with Mantel tests in landscape genetics. Diversity 5:51–72.

Dasmann, R. F. 1981. Wildlife biology. 2nd ed. John Wiley and Sons, New York, NY.

Davidson, C. 2004. Declining downwind: amphibian population declines in California and historical pesticide use. Ecol. Appl. 14:1892–1902.

Dileo, M. F., J. C. Siu, M. K. Rhodes, A. Lõpez-Villalobos, A. Redwine, K. Ksiazek, and R. J. Dyer. 2014. The gravity of pollination: integrating at-site features into spatial analysis of contemporary pollen movement. Mol. Ecol. 23:3973–3982.

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

Dizon, A. E., C. Lockyer, W. F. Perrin, D. P. Demaster, and J. Sisson. 1992. Rethinking the stock concept: a phylogeographic approach. Conserv. Biol. 6:24–36.

Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Resour. 14:209–214.

Dodge, C. M., and V. T. Vredenburg. 2012. The sad song of the Yosemite toad: the role of the amphibian chytrid fungus in an enigmatic decline. Paper presented at the annual convention of the Ecological Society of America, Portland, OR.

Douglas, D. H. 1994. Least-cost path in GIS using an accumulated cost surface and slopelines. Cartographica 31:37–51.

Dozier, J., and J. Frew. 2009. Computational provenance in hydrologic science: a snow mapping example. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 367:1021–1033.

Dozier, J., T. H. Painter, K. Rittger, and J. E. Frew. 2008. Time-space continuity of daily maps of fractional snow cover and albedo from MODIS. Adv. Water Resour. 31:1515–1526.

Dray, S., and A. B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22:1–20.

Drost, C., and G. Fellers. 1994. Decline of frog species in the Yosemite section of the Sierra Nevada. National Park Service report. Davis, CA.

Dyer, R. J., and J. D. Nason. 2004. Population graphs: the graph theoretic shape of genetic structure. Mol. Ecol. 13:1713–1727.

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

Emanuelli, F., S. Lorenzi, L. Grzeskowiak, V. Catalano, M. Stefanini, M. Troggio, S. Myles, J. M. Martinez-Zapater, E. Zyprian, F. M. Moreira, and M. S. Grando. 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol. 13:1–17.

Epps, C. W., J. D. Wehausen, V. C. Bleich, S. G. Torres, and J. S. Brashares. 2007. Optimizing dispersal and corridor models using landscape genetics. J. Appl. Ecol. 44:714–724.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Fellers, G. M., P. M. Kleeman, and D. A. W. Miller. 2015. Wetland occupancy of pond-breeding amphibians in Yosemite National Park, USA. J. North Am. Herpetol. 2015:22–33.

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

Flint, L. E., and A. L. Flint. 2012. Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis. Ecol. Process. 1:1–15.

Flint, L. E., A. L. Flint, J. H. Thorne, and R. Boynton. 2013. Fine-scale hydrologic modeling for regional landscape applications: the California basin characterization model development and performance. Ecol. Process. 2:1–21.

Funk, W. C., M. S. Blouin, P. S. Corn, B. a Maxell, D. S. Pilliod, S. Amish, and F. W. Allendorf. 2005. Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. Mol. Ecol. 14:483–496.

Funk, W., D. Tallmon, and F. Allendorf. 1999. Small effective population size in the long-toed salamander. Mol. Ecol. 8:1633–1640.

Gaggiotti, O. E. 2003. Genetic threats to population persistence. Ann. Zool. Fennici 40:155–168.

Gaggiotti, O. E. 1996. Population genetic models of source sink metapopulations. Theor. Popul. Biol. 50:178–208.

Galpern, P., M. Manseau, and P. Wilson. 2012. Grains of connectivity: analysis at multiple spatial scales in landscape genetics. Mol. Ecol. 21:3996–4009.

Garroway, C. J., J. Bowman, D. Carr, and P. J. Wilson. 2008. Applications of graph theory to landscape genetics. Evol. Appl. 1:620–630.

GDAL Contributors. 2018. GDAL geospatial data abstraction software library. Open Source Geospatial Foundation. Available online at: http://gdal.org

Geffen, E., M. J. Anderson, and R. K. Wayne. 2004. Climate and habitat barriers to dispersal in the highly mobile grey wolf. Mol. Ecol. 13:2481–2490.

Goebel, A. 2005. Conservation systematics: the *Bufo boreas species* group. Pp. 210–221 *in* Amphibian Declines: The Conservation Status of United States species. Univ of California Press.

Goebel, A., and T. Ranker. 2009. Mitochondrial DNA evolution in the *Anaxyrus boreas* species group. Mol. Phylogenet. Evol. 50:209–225.

Gosner, K. 1960. A simplified table for staging anuran embryos and larvae with notes on identification. Herpetologica 16:183–190.

Gotelli, N. J. 1991. Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis. Am. Nat. 138:768–776.

Goudet, J., M. Raymond, T. De Meeüs, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144:1933–1940.

Grasso, R. L., R. M. Coleman, and C. Davidson. 2010. Palatability and antipredator response of Yosemite toads (*Anaxyrus canorus*) to nonnative brook trout (*Salvelinus fontinalis*) in the Sierra Nevada Mountains of California. Copeia 2010:457–462.

Graves, T. A., P. Beier, and J. A. Royle. 2013. Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. Mol. Ecol. 22:3888–3903.

Graybeal, A. 1997. Phylogenetic relationships of bufonid frogs and tests of alternate macroevolutionary hypotheses characterizing their radiation. Zool. J. Linn. Soc. 119:297–338.

Graybeal, A. 1993. The phylogenetic utility of cytochrome b: lessons from bufonid frogs. Mol. Phylogenet. Evol. 2:256–269.

Gros, A., H. Joachim Poethke, and T. Hovestadt. 2006. Evolution of local adaptations in dispersal strategies. Oikos 114:544–552.

Hanski, I. 1998. Metapopulation dynamics. Nature 396:41–49.

Hanski, I., and M. Gilpin. 1991. Metapopulation dynamics: brief history and conceptual domain. Biol. J. Linn. Soc. 42:3–16.

Hansson, L. 1991. Dispersal and connectivity in metapopulations. Biol. J. Linn. Soc. 42:89–103.

Hastings, A. 1993. Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. Ecology 74:1362–1372.

Holt, R. D., M. Barfield, and R. Gomulkiewicz. 2004. Temporal variation can facilitate niche evoluiton in harsh sink environments. Am. Nat. 164:187–200.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Janes, J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, C. I. Cullingham, and R. L. Andrew. 2017. The K = 2 conundrum. Mol. Ecol. 26:3594–3602.

Jennings, M., and M. Hayes. 1994. Amphibian and reptile species of special concern in California. California Department of Fish & Game report. Rancho Cordova, CA.

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

Keeler-Wolf, T., E. T. Reyes, J. M. Menke, D. N. Johnson, and D. L. Karavidas. 2012. Yosemite National Park vegetation classification and mapping project report. National Park Service report. Fort Collins, CO.

Knapp, R. A., G. M. Fellers, P. M. Kleeman, D. A. W. Miller, V. T. Vredenburg, E. B. Rosenblum, and C. J. Briggs. 2016. Large-scale recovery of an endangered amphibian despite ongoing exposure to multiple stressors. Proc. Natl. Acad. Sci. 113:11889–11894.

Knapp, R. A., and K. R. Matthews. 2000. Non-native fish introductions and the decline of the mountain yellow-legged frog from within protected areas. Conserv. Biol. 14:428–438.

Legendre, P., F. Lapointe, and P. Casgrain. 1994. Modeling brain evolution from behavior: A permutational regression approach. Evolution 48:1487–1499.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Am. Entomol. 15:237–240.

Liang, C., and T. Stohlgren. 2011. Habitat suitability of patch types: A case study of the Yosemite toad. Front. Earth Sci. 5:217–228.

Liang, C. T. 2010. Habitat modeling and movements of the Yosemite toad (*Anaxyrus* (=*Bufo*) *canorus*) in the Sierra Nevada, California. Davis, CA: University of California, Davis. Ph.D. dissertation.

Liang, C. T., R. L. Grasso, J. J. Nelson-Paul, K. E. Vincent, and A. J. Lind. 2017. Fine-Scale habitat characteristics related to occupancy of the Yosemite toad, *Anaxyrus canorus*. Copeia 105:120–127.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2:18–22.

Lichstein, J. W. 2007. Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecol. 188:117–131.

Lindauer, A. 2018. The role of hydroperiod and fluctuating temperature on disease dynamics: a disease ecology approach to understanding Yosemite toad declines. Reno, NV: University of Nevada Reno. M.S. thesis.

Lougheed. 1999. Ridges and rivers. a test of competing hypotheses of Amazonian. P.B. Sci. 266 266:1829–1835.

Lowe, W. H., and F. W. Allendorf. 2010. What can genetics tell us about population connectivity? Mol. Ecol. 19:3038–3051.

Maher, S. P., T. L. Morelli, M. Hershey, A. L. Flint, L. E. Flint, C. Moritz, and S. R. Beissinger. 2017. Erosion of refugia in the Sierra Nevada meadows network with climate change. Ecosphere 8:1–17.

Manel, S., and R. Holderegger. 2013. Ten years of landscape genetics. Trends Ecol. Evol. 28:614–621.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18:189–197.

Marsh, D. M., and P. C. Trenham. 2001. Metapopulation dynamics and amphibian conservation. Conserv. Biol. 15:40–49.

Martin, D. L. 2008. Decline, movement and habitat utilization of the Yosemite toad (*Bufo canorus*): an endangered anuran endemic to the Sierra Nevada of California. Santa Barbara, CA: University of California, Santa Barbara. Ph.D. dissertation.

Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, F. Gao, J. Kutler, and T. Lim. 2006. A landsat surface reflectance dataset, 1990-2000. IEEE Geosci. Remote Sens. Lett. 3:68–72.

Matchett, J. R., P. B. Stark, S. M. Ostoja, R. A. Knapp, H. C. McKenny, M. L. Brooks, W. T. Langford, L. N. Joppa, and E. L. Berlow. 2015. Detecting the influence of rare stressors on rare species in Yosemite National Park using a novel stratified permutation test. Sci. Rep. 5:1–12.

Mathieu, J., S. Barot, M. Blouin, G. Caro, T. Decaëns, F. Dubs, L. Dupont, P. Jouquet, and P. Nai. 2010. Habitat quality, conspecific density, and habitat pre-use affect the dispersal behaviour of two earthworm species, *Aporrectodea icterica* and *Dendrobaena veneta*, in a mesocosm experiment. Soil Biol. Biochem. 42:203–209.

Matthysen, E. 2005. Density-dependent dispersal in birds and mammals. Ecography 28:403–416.

McPeek, M. A., and R. D. Holt. 1992. The evolution of dispersal in spatially and temporally varying environments. Am. Nat. 140:1010–1027.

McRae, B. 2006. Isolation by resistance. Evolution 60:1551–1561.

McRae, B., B. Dickson, T. Keitt, and V. Shah. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. Ecology 89:2712–2724.

McRae, B. H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proc. Natl. Acad. Sci. USA 104:19885–19890.

Meirmans, P. G. 2012. The trouble with isolation by distance. Mol. Ecol. 21:2839–2846.

Morelli, T. L., S. P. Maher, M. C. W. Lim, C. Kastely, L. M. Eastman, L. E. Flint, A. L. Flint, S. R. Beissinger, and C. Moritz. 2017. Climate change refugia and habitat connectivity promote species persistence. Clim. Chang. Responses 4:8.

Moritz, C. 1994. Defining "evolutionarily significant units" for conservation. Trends Ecol. Evol. 9:373–375.

Morton, M. 1981. Seasonal changes in total body lipid and liver weight in the Yosemite toad. Copeia 1981:234–238.

Murphy, M. A., J. S. Evans, and A. Storfer. 2010a. Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics. Ecology 91:252–261.

Murphy, M., R. Dezzani, D. Pilliod, and A. Storfer. 2010b. Landscape genetics of high mountain frog metapopulations. Mol. Ecol. 19:3634–3649.

Nakazawa, M. 2018. fmsb: functions for medical statistics book with some demographic data. R package version 0.5.2.

National Park Service. 2018. National park service visitor use statistics. Available online at: https://irma.nps.gov/Stats/SSRSReports/Park Specific Reports/Summary of Visitor Use By Month and Year (1979 - Last Calendar Year)?Park=YOSE

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316–336.

Orsini, L., J. Vanoverbeke, I. Swillen, J. Mergeay, and L. De Meester. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol. Ecol. 22:5983–5999.

Ostoja, S. M., S. R. Lee, J. R. Matchett, H. McKenny, M. L. Brooks, R. A. Knapp, P. A. Maier, and N. Danielle. 2018. Yosemite toad (*Anaxyrus* [*Bufo*] *canorus*) breeding occupancy of meadows in Yosemite and Kings Canyon National Parks, California. Unpublished manuscript.

Painter, T. H., K. Rittger, C. McKenzie, P. Slaughter, R. E. Davis, and J. Dozier. 2009. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. Remote Sens. Environ. 113:868–879.

Palsbøll, P., M. Berube, and F. Allendorf. 2007. Identification of management units using population genetic data. Trends Ecol. Evol. 22:11–16.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. Annu. Rev. Ecol. Evol. Syst. 37:637–669.

Peterman, W. E. 2014. Surfaces using genetic algorithms ResistanceGA: an R package for the optimization of resistance. bioRxiv 7575.

Peterman, W. E., G. M. Connette, R. D. Semlitsch, and L. S. Eggert. 2014. Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. Mol. Ecol. 23:2402–2413.

Pflüger, F. J., and N. Balkenhol. 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. Mol. Ecol. 23:2146–2156.

Pinto, N., and T. H. Keitt. 2009. Beyond the least-cost path: evaluating corridor redundancy using a graph-theoretic approach. Landsc. Ecol. 24:253–266.

Pounds, J. 2001. Climate and amphibian declines. Nature 410:639–640.

Pounds, J. A., M. R. Bustamante, L. A. Coloma, J. A. Consuegra, M. P. L. Fogden, P. N. Foster, E. La Marca, K. L. Masters, A. Merino-Viteri, R. Puschendorf, S. R. Ron, G. A. Sánchez-Azofeifa, C. J. Still, and B. E. Young. 2006. Widespread amphibian extinctions from epidemic disease driven by global warming. Nature 439:161–167.

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

Pritchard, J. K., X. Wen, and D. Falush. 2010. Documentation for structure software: version 2.3. Department of Human Genetics, The University of Chicago, Chicago, IL.

Pulliam, H. R. 1988. Sources, sinks, and population regulation. Am. Nat. 132:652–661.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ratliff, R. D. 1985. Meadows in the Sierra Nevada of California: state of knowledge. U.S. Forest Service report. Berkeley, CA.

Roche, L. M., B. Allen-Diaz, D. J. Eastburn, and K. W. Tate. 2012a. Cattle grazing and Yosemite toad (*Bufo canorus*, Camp) breeding habitat in Sierra Nevada meadows. Rangel. Ecol. Manag. 65:56–65.

Roche, L. M., A. M. Latimer, D. J. Eastburn, and K. W. Tate. 2012b. Cattle grazing and conservation of a meadow-dependent amphibian species in the Sierra Nevada. PLoS One 7:e35734.

Rosenberg, N. A., T. Burke, K. Elo, M. W. Feldman, P. J. Freidlin, M. A. M. Groenen, J. Hillel, A. Mäki-Tanila, M. Tixier-Boichard, A. Vignal, K. Wimmers, and S. Weigend. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159:699–713.

Roy, J., G. Yannic, S. D. Côté, and L. Bernatchez. 2012. Negative density-dependent dispersal in the American black bear (*Ursus americanus*) revealed by noninvasive sampling and genotyping. Ecol. Evol. 2:525–537.

Sadinski, W. 2004. Amphibian declines: causes. U.S. Geological Survey report. La Crosse, WI.

Sawyer, S. C., C. W. Epps, and J. S. Brashares. 2011. Placing linkages among fragmented habitats: do least-cost models reflect how animals use landscapes? J. Appl. Ecol. 48:668–678.

Schmidt, G., C. Jenkerson, J. Masek, E. Vermote, and F. Gao. 2013. Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. U.S. Geological Survey report. Reston, VA.

Schwartz, M. K., and K. S. McKelvey. 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conserv. Genet. 10:441–452.

Shaffer, H., and G. Fellers. 2000. The genetics of amphibian declines: population substructure and molecular differentiation in the Yosemite toad, *Bufo canorus* (Anura, Bufonidae) based on single-strand conformation polymorphism analysis (SSCP) and mitochondrial DNA sequence data. Mol. Ecol. 9:245–257.

Sherman, C. K. 1980. A comparison of the natural history and mating system of two anurans: Yosemite toads (*Bufo canorus*) and black toads (*Bufo exsul*). Ann Arbor, MI: University of Michigan. Ph.D. dissertation.

Sherman, C. K., and M. L. Morton. 1984. The toad that stays on its toes. Nat. Hist. 93:72–78.

Sherman, C., and M. Morton. 1993. Population declines of Yosemite toads in the eastern Sierra Nevada of California. J. Herpetol. 27:186–198.

Shirk, a J., D. O. Wallin, S. a Cushman, C. G. Rice, and K. I. Warheit. 2010. Inferring landscape effects on gene flow: a new model selection framework. Mol. Ecol. 19:3603–3619.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science 236:787–792.

Slatkin, M. 1985. Rare alleles as indicators of gene flow. Evolution 39:53–65.

Smith, D. C. 1987. Adult recruitment in chorus frogs: effects of size and date at metamorphosis. Ecology 68:344–350.

Smith, J. B., and D. A. Tirpak. 1988. The potential effects of global climate change on the United States. U.S. Environmental Protection Agency, draft report to Congress.

Smith, M., and D. Green. 2005. Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? Ecography 28:110–128.

Sork, V. L., F. W. Davis, R. Westfall, A. Flint, M. Ikegami, H. Wang, and D. Grivet. 2010. Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. Mol. Ecol. 19:3806–3823.

Spear, S. F., N. Balkenhol, M.-J. Fortin, B. H. McRae, and K. Scribner. 2010. Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. Mol. Ecol. 19:3576–3591.

Stephens, M. 2001. Phylogeography of the *Bufo boreas* (Anura, Bufonidae) species complex and the biogeography of California. Rohnert Park, CA: Sonoma State University. M.S. thesis.

Storfer, A., M. a Murphy, S. F. Spear, R. Holderegger, and L. P. Waits. 2010. Landscape genetics: where are we now? Mol. Ecol. 19:3496–3514.

Storfer, A., M. Murphy, J. Evans, C. Goldberg, S. Robinson, S. Spear, R. Dezzani, E. Delmelle, L. Vierling, and L. Waits. 2007. Putting the "landscape" in landscape genetics. Heredity 98:128–142.

Sundqvist, L., K. Keenan, M. Zackrisson, P. Prodohl, and D. Kleinhans. 2016. Directional genetic differentiation and asymmetric migration. Ecol. Evol. 1–15.

Thompson, B., A. Wright, and H. B. Shaffer. 2016. Yosemite toad (*Bufo canorus*). Pp. 112–123 *in* California amphibian and reptile species of special concern.

Thornton, P. E., M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook. 2018. Daymet: daily surface weather data on a 1-km grid for North America. ORNL DAAC. Oak Ridge, TN.

Trexler, K. A. 1975. The Tioga Road: A History, 1883-1961. Yosemite Natural History Association, El Portal, CA.

US Fish & Wildlife Service. 2014. Endangered and threatened wildlife and plants; endangered status for the Sierra Nevada yellow-legged frog and the northern distinct population segment of the mountain yellow-legged frog, and threatened status for the Yosemite toad: final rule. Fed. Regist. 1–56.

Vähä, J. P., J. Erkinaro, E. Niemelä, and C. R. Primmer. 2007. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Mol. Ecol. 16:2638–2654.

van Etten, J. 2017. R Package gdistance: distances and routes on geographical grids. J. Stat. Softw. 76:1–21.

van Strien, M. 2013. Advances in landscape genetic methods and theory: lessons leart from insects in agricultural landscapes. Zürich, Switzerland: ETH Zürich. Ph.D. dissertation.

van Strien, M. J., D. Keller, and R. Holderegger. 2012. A new analytical approach to landscape genetic modelling: least-cost transect analysis and linear mixed models. Mol. Ecol. 21:4010–23.

Vandergast, A. G., A. J. Bohonak, S. A. Hathaway, J. Boys, and R. N. Fisher. 2008. Are hotspots of evolutionary potential adequately protected in southern California? Biol. Conserv. 141:1648–1664.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Springer, New York.

VertNet. 2018. Distributed databases with backbone. Available online at: http://portal.vertnet.org/search?q=specificepithet:canorus+genus:(anaxyrus+OR+bufo)+mappable:1

Viers, J. H., S. E. Purdy, R. A. Peek, A. Fryjoff-Hung, N. R. Santos, J. V. Katz, J. D. Emmons, D. V Dolan, and S. M. Yarnell. 2013. Montane meadows in the Sierra Nevada: changing hydroclimatic conditions and concepts for vulnerability assessment. Centre for Watershed Sciences report. Davis, CA.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs. 2010. Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proc. Natl. Acad. Sci. 107:9689–9694.

Wang, I. J. 2012. Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. Divers. Distrib. 18:1033–1041.

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

Waples, R. S., and O. Gaggiotti. 2006. What is a population? an empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol. Ecol. 15:1419–1439.

Weihs, C., U. Ligges, K. Luebke, and N. Raabe. 2005. klaR analyzing German business cycles. Pp. 335–343 *in* D. Baier, R. Decker, and L. Schmidt-Thieme, eds. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Heidelberg, Germany.

Weir, B. S. 1996. Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, Inc., Sunderland, MA.

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

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Yu, H., J. D. Nason, X. Ge, and J. Zeng. 2010. Slatkin's paradox: when direct observation and realized gene flow disagree. a case study in *Ficus*. Mol. Ecol. 19:4441–4453.

Zeller, K. A., K. McGarigal, and A. R. Whiteley. 2012. Estimating landscape resistance to movement: A review. Landsc. Ecol. 27:777–797.

TABLES AND FIGURES

Table 3.1. Landscape genetic modeling challenges. Landscape genetics typically aims to model genetic structure as a function of ecological environment, to glean information about environmental pressures on dispersal. Several complexities of ecology and evolution make successful modeling of genetic structure difficult. Most methods in landscape genetics ignore these challenges. Proposed solutions and added benefits are listed for each challenge.

	Ecological/Evolutionary Modeling Challenge		
	Dispersal routes are complex	Different site vs. dispersal habitat	Lineages Adapted Differently
Modeling Solution	Construct realistic transects (LCPs) Extract each variable from transects	Gravity models: both site-specific and dispersal influence on connectivity	Random slopes by lineage
Additional Benefit	Variables are modeled separately	Can determine which is more influential: site or dispersal environment	Can extrapolate connectivity of a lineage outside of current distribution (predict translocation effects)

Table 3.2. Environmental data sources.

Type	Source	Year(s)	Resolution	Details
Climate	CA Basin Characterization Model (BCM) 2014	1920–2010	270 m	Downscaled from 800m PRISM
Climate	Daymet	1980–1997	1000 m	Alternative to PRISM-derived BCM
Disturbance	Cal Fire	1900–2014	10 m	Rasterized from known perimeters
Disturbance	Yosemite National Park (YNP)	2010	10 m	Trails/roads rasterized and weighted by traffic
Meadow Network	USGS survey data	2009–2014	-	See Berlow et al. (2013) for details
Moisture/Vegetation	LANDSAT 5	1984–2011	30 m	Climate Data Record (CDR) processed
Snowmelt	MODIS	2002–2007	500 m	Daily albedo from spectroradiometer
Topography	Shuttle Radar Topography Mission (SRTM)	2000	10 m	National elevation dataset
Vegetation	Yosemite National Park (YNP) LiDAR	2012	10 m	Aerial imagery
Watershed Attributes	Cal Water	2004	10 m	CalWater 2.2.1 watershed boundaries

Table 3.3. Gravity modeling variables. The two most important variables in each of 11 categories were previously selected using random forests. A complete listing of variables considered is in Table S3.2.

Location	Category	Abbreviation	Source	Definition	GIS Summary*
At	Climate†	AT(pptch)	BCM 2014	60-year change in mean precipitation	Mean at sites
		AT(precip)	Daymet	Inter-annual mean of precipitation	Mean at meadows
	Disturbance	AT(roadd)	YNP	Distance to nearest road	Value at sites
		AT(fired)	Cal Fire	Distance to nearest fire perimeter	Value at sites
	Meadow Network	AT(BP)	Berlow et al. 2013	Network-boosted breeding probability	Value at meadows
		AT(breed)	YNP LiDAR	Distance to nearest breeding meadow	Value at meadows
	Snowmelt	AT(p50sd)	MODIS	Proportion days with >50% snow cover	SD at meadows
		AT(meltsd)	MODIS	Meadow meltoff date	SD at meadows
	Topography	AT(spi)	SRTM	Slope position index	SD at sites
		AT(slope2)	SRTM	Second derivative of slope	SD at sites
	Vegetation	AT(pcon)	YNP LiDAR	Proportion coniferous vegetation	Value at sites
		AT(pflood)	YNP LiDAR	Proportion semi- to permanently flooded meadow vegetation	Value at sites
Between	Climate†	BT(petch)	BCM 2014	60-year change in CV of potential evapotranspiration	Mean between sites
		BT(runch)	BCM 2014	60-year change in mean of runoff	Mean between sites
	Disturbance	BT(trail)	YNP	Trail crossings (per pixel)	Sum between sites
		BT(roadw)	YNP	Road crossings (per pixel, weighted by traffic level)	Sum between sites
	Moisture/Vegetation	BT(nbrm)	LANDSAT 5	Burn Ratio in May: interannual mean	SD between Sites
		BT(ndmi)	LANDSAT 5	Moisture Index in May–Sept: interannual mean	SD between sites
	Topography	BT(dem)	SRTM	Elevation	Mean between sites
		BT(slope)	SRTM	Slope	Sum between sites
	Vegetation	BT(pcon)	YNP LiDAR	Proportion coniferous vegetation	Value between sites
		BT(pwat)	YNP LiDAR	Proportion standing water, permanent pool, or stream	Value between sites

* "at sites" = circular buffer around sampling sites; "at meadows" = polygon of meadow boundary; "between sites" = buffer around LCP paths.
† Climate 60-year change = difference between 1980–2010 mean and 1920–1950 mean

Table 3.4. Source-sink predictive variables.

Category	Abbreviation	Source	Definition	GIS Summary
Meadow Network	BP	Berlow et al. 2013	Network-boosted breeding probability	Value
Meadow Network	A	YNP LiDAR	Area of meadow	Value
Meadow Network	BWA	YNP LiDAR	Breeding wet areas	Value in 1 km radius
Meadow Network	MA	YNP LiDAR	Area of meadows	Value in 1 km radius
Meadow Network	P	YNP LiDAR	Length of meadow perimeter	Value
Meadow Network	WAC	YNP LiDAR	Wet area of closest meadow	Value
Topography	ESD	SRTM	Elevation	SD in 500 m radius
Topography	R	SRTM	Roughness index	Mean in 500 m radius
Topography	S	SRTM	Slope	Mean in 500 m radius
Topography	SPI	SRTM	Slope position index	Mean in 500 m radius

Table 3.5. Analysis of Molecular Variance (AMOVA) results. Based on five hierarchical levels: park, lineage, ADMIXTURE clusters, meadows, and individuals. Significance is based on 100 random permutations with one-tailed test.

Variance Component	df[a]	SS[b]	MS[c]	σ[d]	% Total[e]	ϕ-Statistic[f]	p-value[g]
Between parks (P)	1	50745.20	50745.20	116.90	30.45	ϕ_{PT} = 0.305	**
Between lineages (L) within parks (P)	7	41768.02	5966.86	30.15	7.85	ϕ_{LP} = 0.113	**
Between clusters (C) within lineages (L)	27	55756.95	2065.07	14.67	3.82	ϕ_{CL} = 0.062	**
Between meadows (M) within clusters (C)	66	76499.53	1159.08	89.96	23.43	ϕ_{MC} = 0.405	**
Between individuals (I) within meadows (M)	542	95376.63	175.97	43.78	11.41	ϕ_{IM} = 0.331	**
Within individuals (I)	644	56931.50	88.40	88.40	23.03	ϕ_{IT} = 0.77	**
Total	1287	377077.82	292.99	383.87	100.00		

[a] df = degrees of freedom
[b] SS = sums of squares
[c] MS = mean squares
[d] σ = variance
[e] % = proportion of total of variance
[f] ϕ-statistic = ratio of variances
[g] p-value = permuted probability of the observed σ by chance; ** signifies $p < 0.01$

Table 3.6. Model comparison for LCP transects between meadows. Causal modeling considered correlation between LCP distance and F_{ST} for 10 hypothesized resistance rasters. Significant p-values from partial mantel tests after accounting for Euclidean distance are in bold. Slope and vegetation cost values were added together with three alternative weights: 1:1, 3:1, or 1:3. Mantel and partial mantel tests were performed with 1000 random permutations.

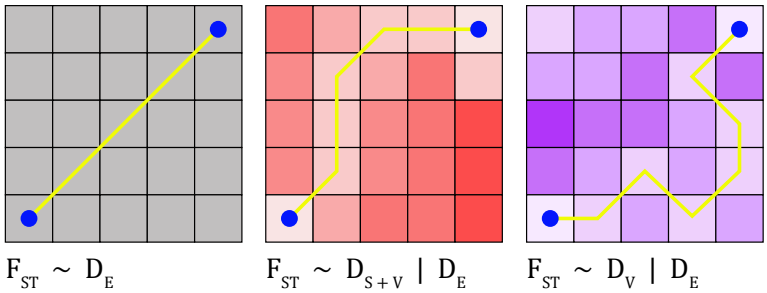
Model*	Mantel ρ	p-value	Partial Mantel ρ	p-value
E	0.311	***	-	-
S	0.324	***	0.164	**0.021**
V	-0.070	0.723	-0.081	0.753
S+V (1:1)	0.314	***	0.042	0.306
S+V (3:1)	0.326	***	0.134	0.057
S+V (1:3)	0.317	***	0.061	0.230
R, S	0.332	***	0.138	0.081
R, V	0.309	***	0.041	0.310
R, S+V (1:1)	0.322	***	0.086	0.195
R, S+V (3:1)	0.341	***	0.176	**0.040**
R, S+V (1:3)	0.313	***	0.051	0.303

* R=Ridges, S=Slope, V=Vegetation, E=Euclidean Distance

Table 3.7. Gravity modeling results. Each of 12 models has the lowest AICc score for its category, where adding each effect gave a significant likelihood ratio test score. Model categories include transect cutoff distance (10, 20, 30 km), transect type (LCP, Euclidean), and lineage-specific random slopes (yes or no). The best model for each transect cutoff distance is highlighted in bold. All variables fitted with lineage-specific slopes are in bold. All variables are listed in order of their inclusion during the forward-selection, and signs represent coefficient directions with respect to meadow connectivity ($1 - \sin^{-1}\sqrt{F_{ST}}$). "AT" and "BT" refer to at-site and between-site measurements, respectively. Only fixed effect signs are shown; for values of fixed and lineage-specific variable coefficients, see Table S3.3.

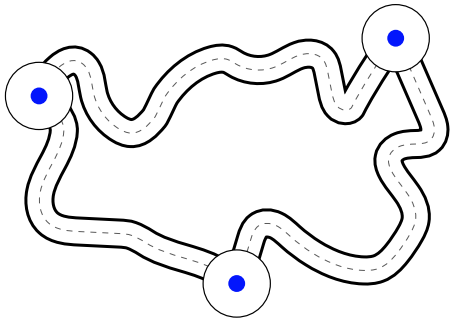
Dist. (km)	Transect	Formula	Rand. Slopes	-lnL	R^2_c	R^2_m	AICc	BIC
30	LCP	D - **BT(slope)** + **BT(dem)** - BT(runch) - **BT(ndmi)** - BT(trail) - BT(roadw) - **BT(pcon)** - BT(pwat) - BT(nbrm) - AT(pflood)	-	-2166.73	0.75	0.59	4361.62	4444.14
			✔	**-2096.76**	**0.81**	**0.58**	**4252.18**	**4422.79**
	Euclid.	D - **BT(slope)** + **BT(dem)** - **BT(runch)** - BT(nbrm) + BT(petch) - BT(trail) - BT(pcon) - BT(ndmi) - BT(roadw) - AT(pflood) - BT(pwat)	-	-2334.41	0.72	0.55	4699.00	4787.40
			✔	-2198.96	0.86	0.47	4448.41	4595.56
20	LCP	D - BT(slope) + **BT(dem)** - BT(runch) - **BT(roadw)** - **BT(ndmi)** - BT(trail) - **BT(pwat)** - **BT(pcon)** - BT(nbrm) - AT(pflood) + AT(pptch)	-	-1627.01	0.67	0.43	3284.31	3365.64
			✔	**-1571.47**	**0.76**	**0.37**	**3216.53**	**3410.82**
	Euclid.	D - **BT(runch)** - BT(ndmi) - BT(nbrm) + **BT(dem)** - **BT(slope)** - **BT(trail)** - **BT(pcon)** + **BT(petch)** - BT(roadw) - AT(pflood) + AT(pptch)	-	-1733.51	0.62	0.37	3497.30	3578.63
			✔	-1635.87	0.87	0.18	3360.03	3591.73
10	LCP	D - **BT(slope)** - AT(precip) - BT(runch) - BT(pwat) - BT(trail) - BT(roadw) - **BT(dem)** - AT(breed)	-	-567.11	0.63	0.41	1158.80	1209.94
			✔	**-549.52**	**0.78**	**0.41**	**1136.34**	**1212.63**
	Euclid.	D - **BT(ndmi)** - BT(runch) - BT(slope) - AT(pflood) - AT(breed) - AT(slope2)	-	-598.44	0.58	0.31	1217.30	1259.99
			✔	-577.44	0.73	0.33	1181.56	1236.91

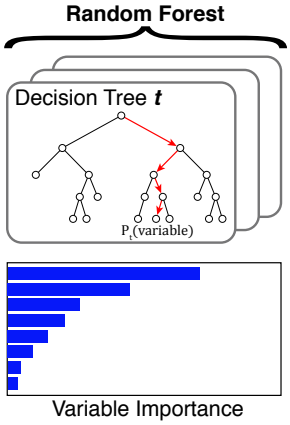
(a) Determine dispersal paths



✔

(b) Extract/choose variables

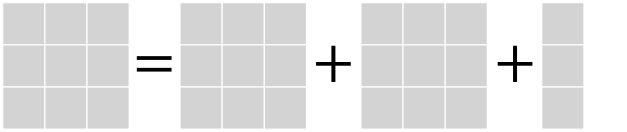




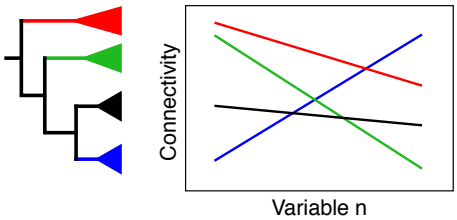
Choose 2 variables for each environmental category
Exclude variables with high multicollinearity (VIF ≥ 10)

(c) Gravity model

C1. Select variables that improve model fit *



C2. Select lineage-specific slopes that improve model fit **



Selection Process (* / **)

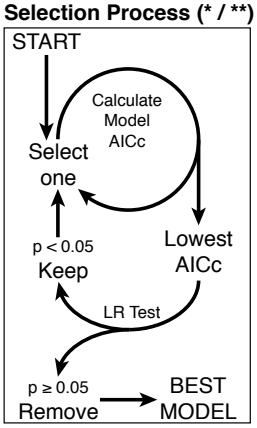


Figure 3.1. Least cost gravity modeling workflow
Workflow for modeling genetic connectivity based on both dispersal and site environments. (a) LCP models are constructed from simple environmental variables known to influence connectivity, such as slope (S) and vegetation (V). A partial mantel test between F_{ST} and LCP distance is performed to choose the best model, while accounting for Euclidean distance (E). (b) Raw environmental data are extracted from dispersal corridors (buffered LCPs) and sites (buffered locations). Variables are categorized, and N (e.g. two) variables are chosen from each category, based on variable importance scores from a random forests analysis. Collinear variables (e.g. variance inflation factor ≥ 10) are excluded. (c) Gravity models of genetic connectivity are constructed using both dispersal and site environmental variables, as well as LCP distance. Lineage-specific environmental effects are accommodated using random slopes. Variables and random slopes are added using a forward-selection process.

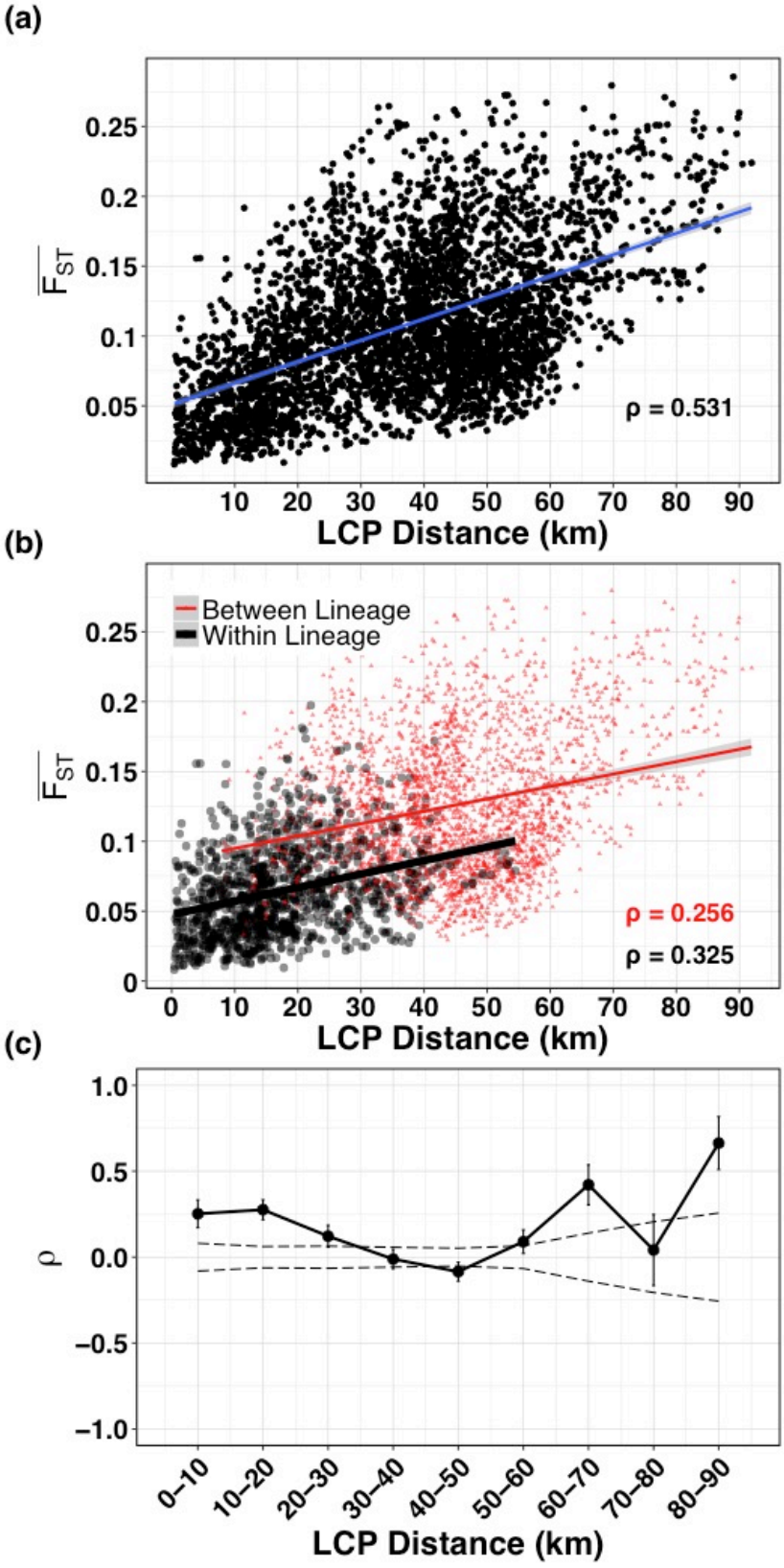

(a)
0.25
0.2
0.15
0.1
0.05
$\overline{F_{ST}}$
ρ = 0.531
10 20 30 40 50 60 70 80 90
LCP Distance (km)
(b)
Between Lineage
Within Lineage
0.25
0.2
0.15
0.1
0.05
0
$\overline{F_{ST}}$
ρ = 0.256
ρ = 0.325
0 10 20 30 40 50 60 70 80 90
LCP Distance (km)
(c)
1.0
0.5
0.0
−0.5
−1.0
ρ
0–10
10–20
20–30
30–40
40–50
50–60
60–70
70–80
80–90
LCP Distance (km)

Figure 3.2. Isolation by distance
Isolation by distance (IBD) based on $\bar{F}_{ST}$ and LCP-derived, topographically corrected distances between meadows. (a) Scatterplot of all pairwise comparisons with mantel correlation of $\rho = 0.531$ with $p < 0.01$. (b) The same scatterplot, separated into between-lineage ($\rho = 0.256$) and within-lineage ($\rho = 0.325$) comparisons, both with $p < 0.01$. Linear regression trend lines are shown for both (a) and (b). (c) Mantel correlogram showing mantel correlation coefficients for successive 10 km distance categories. Solid lines are actual correlations, stippled lines represent confidence intervals for null distribution, and error bars represent bootstrapped confidence intervals for correlation.

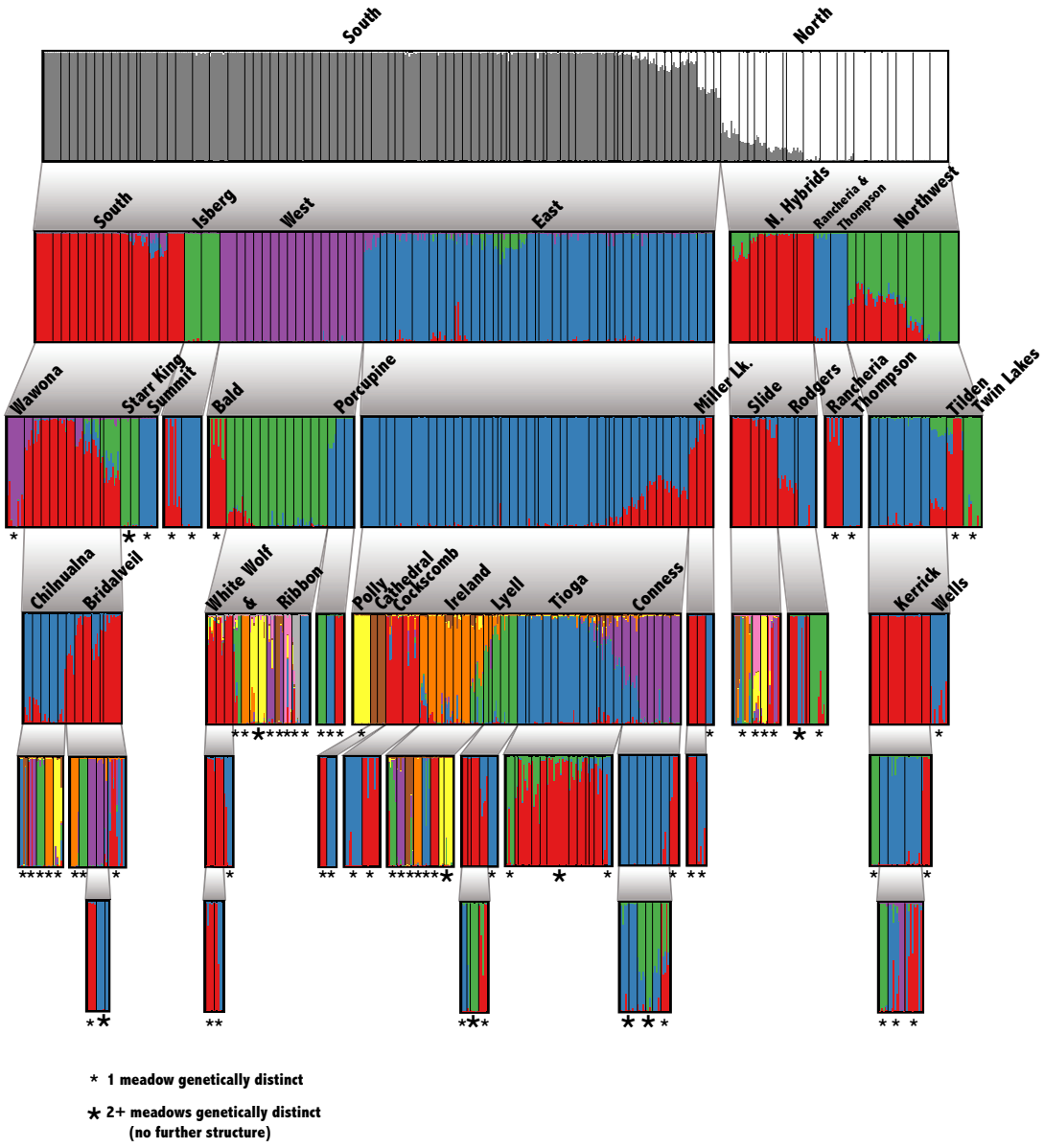


Figure 3.3. Hierarchical genetic structure in Yosemite NP
Barplots of hierarchical genetic structure based on recursive STRUCTURE analyses down to the smallest detectable unit. Thin, black, vertical bars indicate meadow boundaries. All 90 meadows from YOSE are included, and those found to be distinct are labeled with a small (*), while groups of meadows with no further sub-structure are labeled with a large (*). Each analysis considered a range of K from 1–10, and the optimal K was chosen following Evanno (2005). The hypothesis that K=1 was tested using a paired Wilcoxon test. Each cluster was rerun until reaching the meadow level, or finding no further structure.

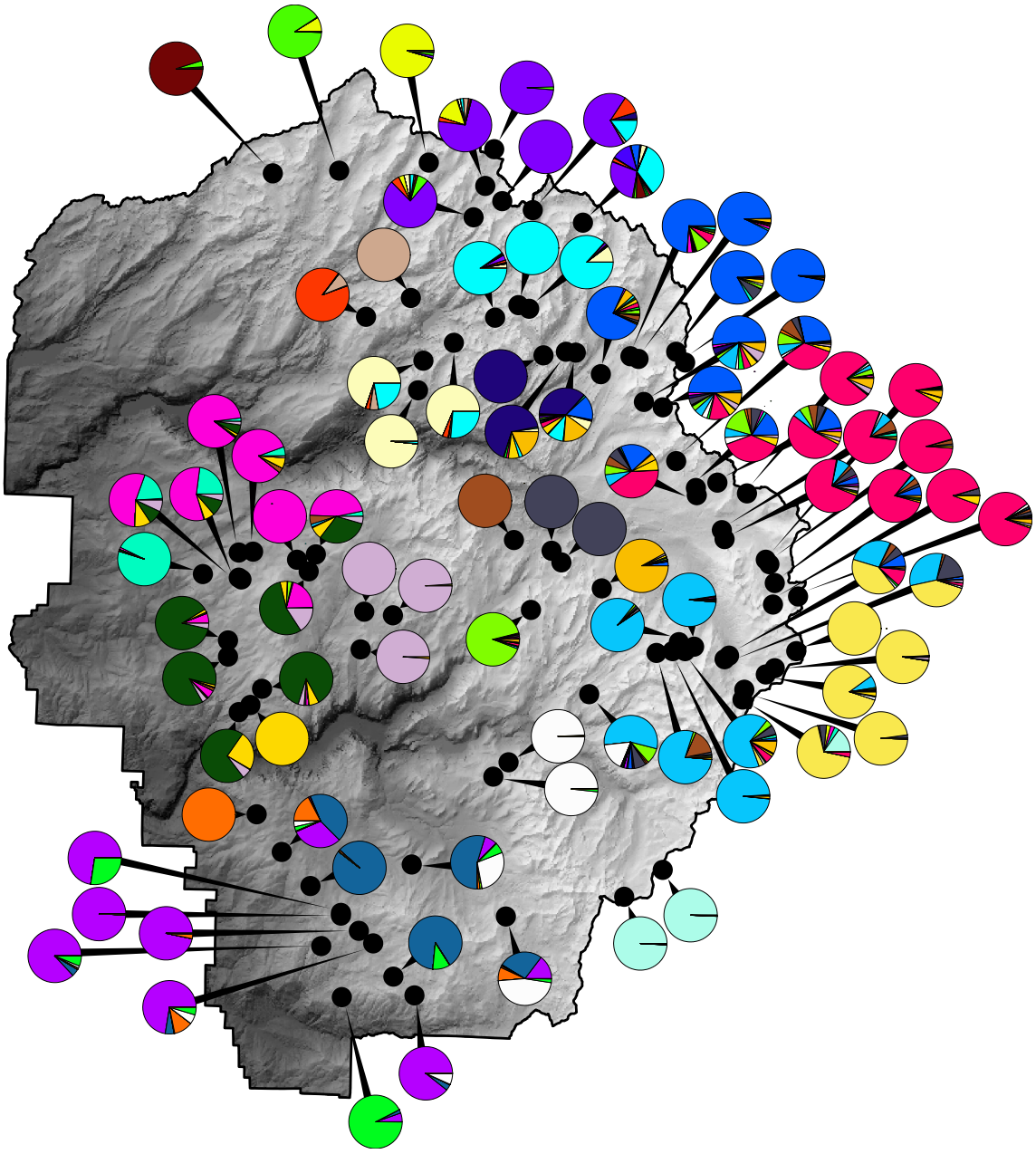


Figure 3.4. Meadow clusters in Yosemite NP
Proportion of individual ancestry from an optimal K=28 gene pools delineated by ADMIXTURE. Each pie chart represents individual ancestry summarized at the meadow scale. Names and summary statistics for each cluster are in Table S1.2. The optimal K was chosen by minimizing cross-validation error, defined as the proportion of mis-assignment when holding out each 10% of the genotypes. The 28 colors are arbitrarily chosen for display purposes.

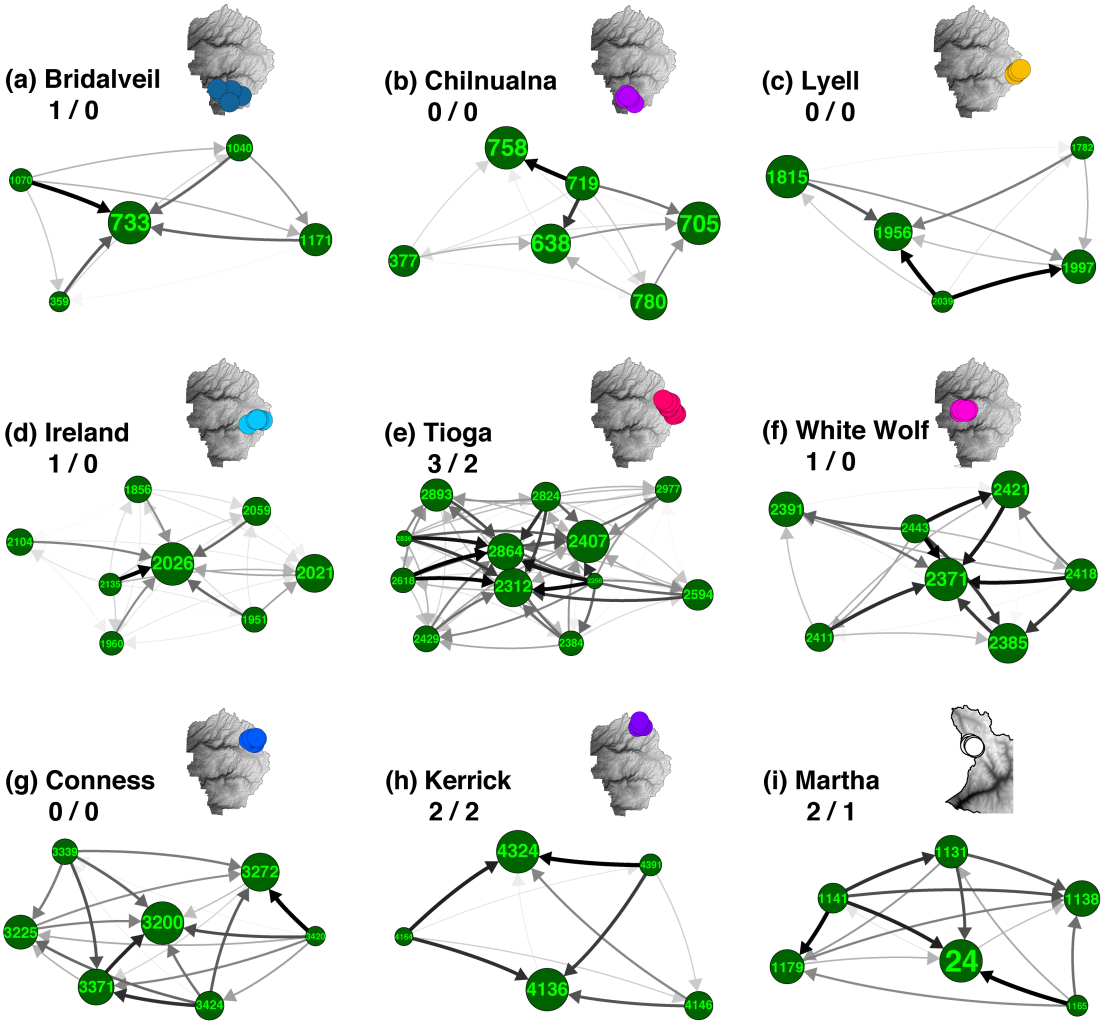


Figure 3.5. Source-sink dynamics in meadow clusters
Most of the ADMIXTURE clusters with at least five meadows are groups of closely related meadows with one or a few highly admixed "hubs" surrounded by more genetically distinct "satellites". For each cluster, net asymmetry in migration rates (difference between emigration and immigration) is summarized by direction and shade of arrow. Meadow IDs and associated cluster names are labeled for all nine clusters in YOSE and KICA with at least five meadows. Meadow node sizes are weighted by relative observed heterozygosity. Inset maps denote location of each cluster in YOSE (a–h have identical colors to Fig. 3.4) and KICA (i). Numbers in parentheses indicate number of identified hubs / satellites.

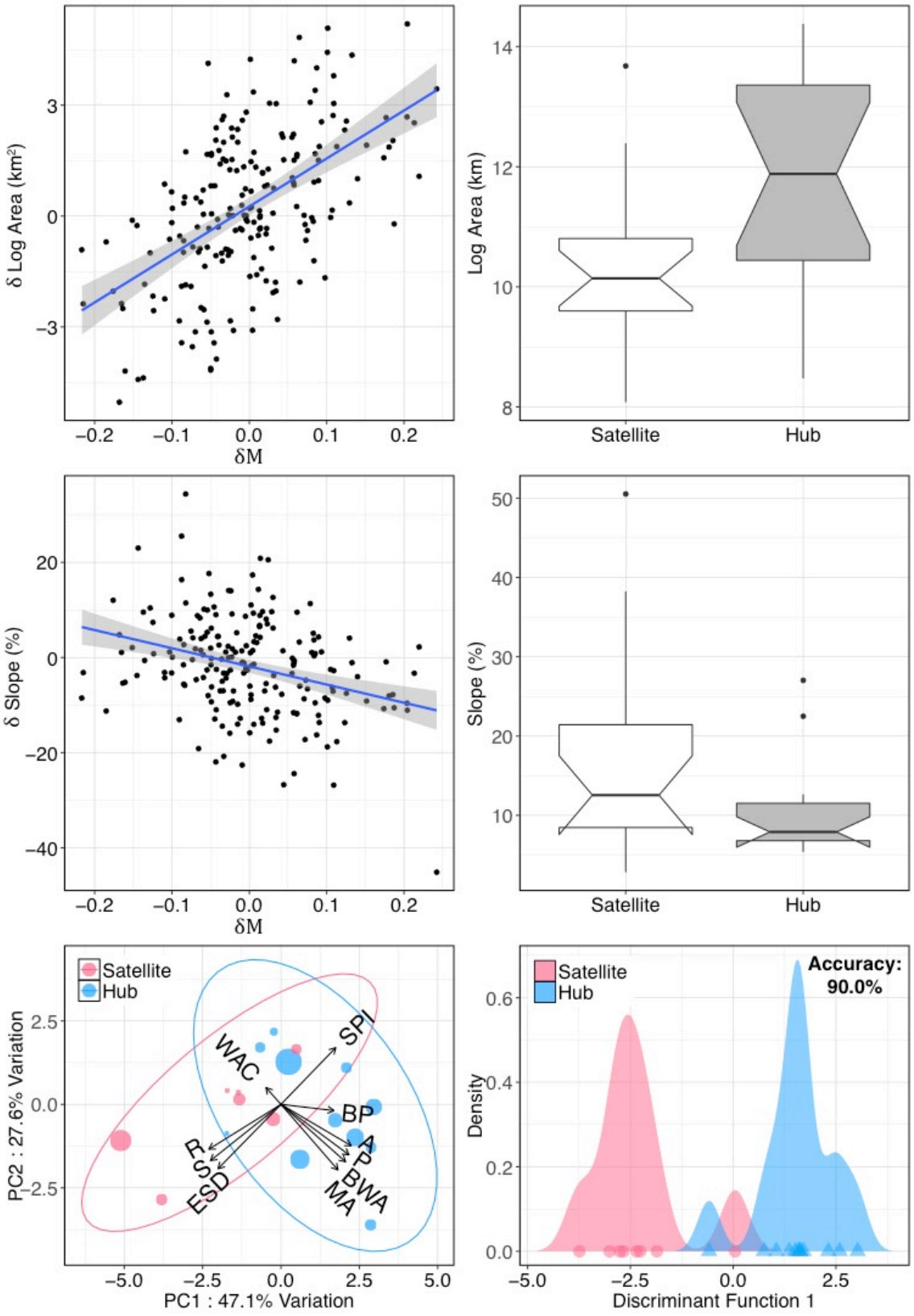

δ Log Area (km²)
δM
Log Area (km)
Satellite
Hub
δ Slope (%)
δM
Slope (%)
Satellite
Hub
Satellite
Hub
WAC
SPT
BP
P
BWA
MA
SSD
ESD
PC2 : 27.6% Variation
PC1 : 47.1% Variation
Satellite
Hub
Accuracy:
90.0%
Density
Discriminant Function 1

Figure 3.6. Attributes of hub and satellite meadows
Hub (exceptional immigration rate) and satellite (exceptional emigration rate) meadows from Fig. 3.5 are related to several environmental variables. Relationship between net relative migration (δM) and relative difference in pairwise meadow area (a) and slope (c) are shown. In (b) and (d), these pairwise differences are averaged by meadow to show mean meadow area and slope for hubs ($\delta M < 0 - 1SD$) and satellites ($\delta M > 0 + 1SD$). In (e), a principal components analysis (PCA) shows the first two PCs for 10 variables that broadly fall into one of two categories: "meadow area" and "topographic roughness." In (f), a discriminant analysis of principal components (DAPC) shows the predictive accuracy of these 10 variables to discriminate between hubs and satellites.

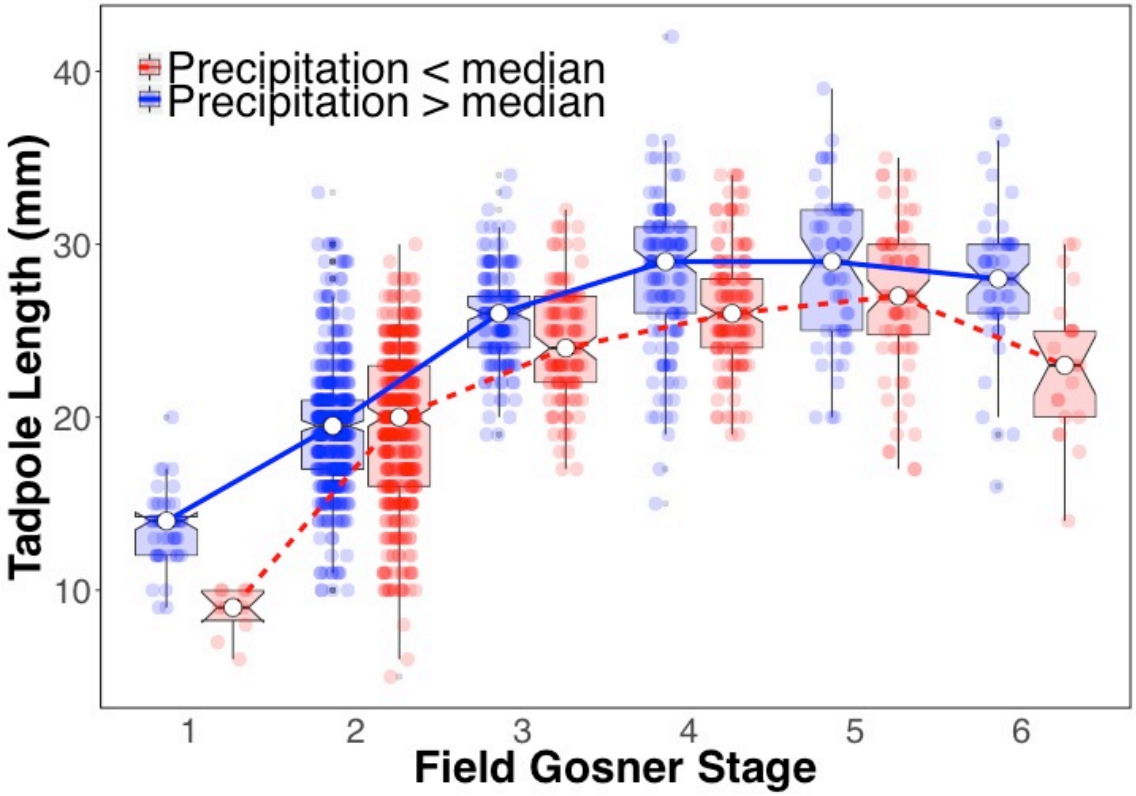


Figure 3.7. Tadpole length and stage differ by precipitation level
Comparison of growth curves among meadow-scaled Daymet precipitation levels, > and < median values. Tadpole field stages correspond to Gosner (1960) stages as follows: 1 (Gosner 1–25), 2 (Gosner 26–30), 3 (Gosner 31–35), 4 (Gosner 36–39), 5 (Gosner 41), 6 (Gosner 42–44). Tadpoles spend the majority of time as Field Stage 2.

CHAPTER 4: PREDICTING THE UNKNOWN: CONSERVATION UNITS FOR EVOLUTIONARY POTENTIAL

ABSTRACT

An essential goal in conservation biology is delineating population units that maximize the probability of species persisting into the future, and adapting to future environmental change. However, the particular discipline and philosophy of the practitioner (i.e. systematics, population genetics, ecology, or evolutionary biology) can largely determine which tools are used for delineation, and what evolutionary processes are protected. I echo the sentiments of Bowen (1998, 1999), that the underlying process of lineage evolution should be the target of conservation, and it must be measured using a temporal framework. Ancient lineages represent past adaptive or reproductive divergence that may or may not be relevant to future selective pressures. Recent inter-lineage contact zones can harbor novel recombinant diversity through transgression, resulting in extreme trait values that may increase the pace of future adaptation. Modern gene pools are defined by demographic independence, and they maintain the bulk of additive genetic variation that may set the pace of future response to selection. All of these processes are informative to conservation insofar as they may influence future population persistence, adaptation to new environments, and continued lineage diversification. I recommend delineating future "Geminate Evolutionary Units" (GEUs), by modeling the most relevant selective pressure to a focal species, and forecasting adaptive shifts into the future. I applied this framework to delineate conservation units for the Yosemite toad (*Anaxyrus* [*Bufo*] *canorus*), a recently federally listed species that is highly vulnerable to climate change. I used a genome-wide dataset and genetic association methods to identify candidate loci responding to climatic selection, and then forecasted future response to selection with the

machine learning algorithm Gradient Forests. I also estimated where additive genetic variation is most likely to be replenished, by forecasting patterns of meadow connectivity into the future. Based on all the available evidence, I found three GEUs in Yosemite National Park, based on differing patterns of snowpack, runoff, and summer precipitation. I project that the highly diverse Y-East lineage will be a future refuge from climate change, as it will contain the most additive genetic variation, with the smallest requirement for adaptive genetic change. My approach provides a comprehensive strategy for protecting an imperiled non-model species with genomic data alone, and has wide applicability to other declining species that are difficult to study directly.

INTRODUCTION

Conservation biologists agree upon the goal of preserving evolutionary processes to help species persist across a diversity of habitats (Frankel and Soulé 1981; Allendorf et al. 2013). However, even-handed delineation of conservation units is often challenging due to disagreement regarding the appropriate processes to preserve, and ambiguity regarding which ecological and genetic patterns are the best surrogates for them. The long-standing debate about how to operationally define Evolutionarily Significant Units (ESUs) exemplifies this: researchers disagree whether ESUs should delineate historically isolated lineages (Avise 1994; Moritz 1994, 2002; Bowen 1998), or circumscribe populations that have historically adapted to similar phenotypes, without specific regard for their shared ancestry (Ryder 1986; Waples 1991; Vogler and Desalle 1994; Crandall et al. 2000). Conceptually, both approaches are an attempt to preserve historical

trajectories that may represent early stages toward formation of new species (de Queiroz 1998), and hence both should be maintained. However, the relative order of events (e.g. lineage vicariance, adaptive evolution, reproductive isolation), and the type of dataset bearing upon them (e.g. molecular, ecological, morphological) might suggest primacy of one process over another. If evolutionary legacy were indeed the ideal process to conserve, then ESUs would ideally be delimited by giving proportional weight to both neutral and adaptive genetic divergence, based on all empirical evidence for each (Fraser and Bernatchez 2001).

This entire ESU approach may be inherently misguided, however, because it implicitly assumes that ancient divergence and/or adaptation are good indicators of future persistence and/or adaptation. This assumption is ubiquitous: most conservation genetic studies delineate units of future evolutionary potential based on phylogenetic or population genetic patterns. This means that "future" units of potential are informed by temporal periods ranging from the ancient past to the present (Fraser and Bernatchez 2001; Allendorf et al. 2013; Casacci et al. 2014). However, adaptive genetic diversity preserved in ancient lineages will not necessarily be equally relevant to future selective environments (Crandall et al. 2000). ESUs defined by supposedly adaptive characters (such as distinct morphologies, behaviors, or a nuclear gene sequence) might present an incomplete picture of future fitness, or be incongruent with future selection, assuming the traits in question are even heritable (Fraser and Bernatchez 2001; Moritz 2002). In addition to (or in lieu of) delineating ESUs, many conservation genetic studies attempt to delineate Management Units (MUs), which are generally defined as demographically

independent populations connected by dispersal rates below a threshold of 10% (Hastings 1993), and diagnosed by significantly distinct allele frequencies (Dizon et al. 1992; Moritz 1994; Palsbøll et al. 2007). MUs are meant to represent the scale at which contemporary genome-wide processes of migration and drift maintain genetic diversity, however they are often delineated with similar operational criteria ("significant divergence of allele frequencies") to those of supposedly historical ESUs. Further complicating MU delineation, genetic structure is known to represent an average snapshot of past dispersal (Mallet 2001), and exact boundaries are thwarted by hierarchical population structure and isolation by distance (Frantz et al. 2009).

Although both ancient lineage distinctiveness and current genetic diversity are important components of evolutionary potential, the persistence of a population into future environmental conditions is ultimately decided by its future adaptive optimum, and whether it can garner sufficient genetic novelty to adapt and grow. Bowen (1998, 1999) channeled the idea of future evolutionary legacy (Waples 1995) into a future-directed conservation unit that he termed the Geminate Evolutionary Unit (GEU). The term "geminate species" was originally coined to refer to young sister species on either side of a geographic barrier, whose minor apomorphic differences would allow their continued interbreeding, but may lead to future adaptive divergence (Jordan 1908). Conceptually, GEUs offer a better framework than ESUs or MUs for thinking about conservation concerns, which are axiomatically forward-looking. I adapt and codify the GEU concept here, and prescribe two patterns useful for identifying GEUs (populations likely to form lineages that persist in future environments): (1) populations characterized by

inordinately high diversity, possibly caused by inter-lineage admixture, and (2) populations expected to adapt similarly to models of future environmental change. I suggest that this temporal framework should be used to distinguish conservation units of interest (Table 4.1): past and present evolutionary processes should be studied, and then used to model the rate of adaptive change necessary to keep apace with future natural selection.

One of the simplest and most widespread models for predicting future evolutionary change by natural selection is the breeder's equation (Lush 1937). In its single-locus form $R = h^2S$, it demonstrates that heritable response to selection across one generation depends on the selection differential S (mean trait difference after phenotypic selection on the parents), and the breeding value h^2 (proportion of offspring trait value that can be predicted from the mean parental value). This last term h^2, also known as the additive genetic variance or heritability, is the crucial link between phenotypic selection and genotypic inheritance, at the population level (Visscher et al. 2008). Hence, the required level of response to future selection is a product of both selection and heritability of the trait involved (Falconer 1960). Of course, the breeder's equation makes a number of simplifying assumptions, chiefly: the focal trait(s) under selection are all measured, and causality between the trait(s) and fitness is known. Even if these are met, long-term prediction is limited because the model assumes constant h^2, which is of course violated as selection depletes adaptive genetic variation (Pigliucci and Schlichting 1997; Roff 2000; Houchmandzadeh 2014). For this reason, some authors have criticized the breeder's equation as being too narrow of a model, in the same vein as Fisher's (1930)

fundamental theorem (Heywood 2005; Pigliucci 2006; Gienapp et al. 2008; Pemberton 2010). Alternative equations have been proposed to ameliorate some of these oft-violated model assumptions, although some of these solutions work by ignoring h^2 altogether (Morrissey et al. 2010, 2012; Houchmandzadeh 2014).

Notwithstanding some of the objections to its use as a predictive model, the breeder's equation has been successfully used to model rates of adaptation to climate change and ocean acidification in dozens of examples (Møller and Szép 2005; Gienapp et al. 2008; Pemberton 2010; Sunday et al. 2011; Charmantier and Gienapp 2014). Moreover, the concept has been used to extend the predictive power of species distribution and community composition models, by integrating ecological niche, physiological tolerance, and rates of adaptive evolution (Catullo et al. 2015; Bush et al. 2016). Thus, for the conservation goal of predicting broad landscape patterns of lineage evolution, the breeder's equation may offer a conceptual model for understanding future adaptive change. Namely, there are two aspects to delineating GEUs: (1) h^2, adaptive genetic variation (heritability), and (2) S, selective pressure. In a conservation genomic context these can be though of as genetic supply and demand for adaptation to a stressor, such as climate change. Without any phenotypic or experimental data, the breeder's equation cannot be explicitly used. However, information about where selection is expected to be highest, in concert with which populations are adapting in similar ways, is valuable information for delineating future conservation units (GEUs; Bowen 1998, 1999). Depending upon how "hard" or "soft" selection is, h^2 relative to S is also informative about whether the number of selective deaths required for adaptation is

compatible with population persistence (Haldane 1957; Wallace 1975; Nunney 2003; Reznick 2016). This can be estimated by (1) modeling associations between adaptive genetic variation and the current environment, (2) projecting those associations into the future (e.g. by forecasting climate), and (3) using the anticipated genetic difference as a relative measure of R, the required amount of genetic response to stay in the adaptive optimum. Relative h^2 can be directly estimated using standing levels of adaptive genetic diversity (at loci under selection), and also indirectly from N_e and genetic connectivity.

I applied this GEU approach to the Yosemite toad (*Anaxyrus canorus*), a species of high-elevation anuran endemic to the central Sierra Nevada of California. This species exclusively breeds in the transient and extremely shallow ponds of mountain meadows (Ratliff 1985) which make up less than 3% of the landscape. Climate change is ostensibly one of the greatest threats to Yosemite toad breeding ecology (Viers et al. 2013; US Fish & Wildlife Service 2014; Brown et al. 2015; Chapter 3), but little is known about how toads will need to adapt to remain viable into the future. Yosemite toads are reputed to have declined in both distribution and abundance from across their range, despite dwelling primarily within wilderness areas (Sherman and Morton 1993; Drost and Fellers 1994; Jennings and Hayes 1994; Green and Sherman 2001; Thompson et al. 2016; but see Ostoja et al., in prep.). Although little is known about where toads are most vulnerable to climate change, much has recently been gleaned about patterns of lineage diversity (Chapter 1), adaptive introgression through secondary admixture (Chapter 2), and current gene pool boundaries (Chapter 3). For example, there is some evidence that lower-elevation lineages have repeatedly undergone adaptive divergence from higher-

elevation lineages, as they inhabited refugia in different climates (Chapter 1). As stated above, these patterns of past and current genetic structure can be used to guide GEU delimitation (Table 4.1).

I leveraged a genome-wide dataset in this study, so that I could model putatively adaptive genetic processes separately from full-genome processes. The use of outlier SNPs for conservation unit delineation is not new (Bonin et al. 2007; Allendorf et al. 2010; Funk et al. 2012), however I apply a newly developed community ecology approach for modeling outlier SNP variation into future conditions using the non-linear regression tree approach, Gradient Forests (Ellis et al. 2012; Fitzpatrick and Keller 2015). I also modeled genetic connectivity of toad-occupied meadows as a function of climate, and used graph theory to forecast future shifts in overall network connectivity as a function of climate change. Very few studies have modeled climate change impacts to gene flow (however, see Razgour 2015), and to my knowledge no studies have explicitly predicted future climatic effects on graph-theoretical connectivity. For all future genetic models, I assume that climatic selection is directional, and hence will deplete h^2 over time. The potential for h^2 to be replenished fast enough to respond to additional climate change will almost certainly depend upon exogenous migrants ($N_e m$) rather than new mutations ($N_e \mu$), given the rapid pace of climate change.

In this study, my goals were to (1) identify populations of Yosemite toads that are adapting to climate in similar ways in Yosemite National Park, (2) quantify the relative extent to which future climate change is anticipated to impact those populations, and (3) forecast where adaptive genetic variation (h^2 for loci under climatic selection) is most

likely to be replenished by gene flow. I used a highly robust spatial sampling scheme to reduce bias associated with unsampled locations, and a double-digest RADseq dataset that was previously described (Chapter 1). My GEU approach to conservation genomics is a novel and informative way for anticipating future patterns of evolutionary potential as they relate to population persistence.

MATERIALS AND METHODS

Molecular methods, ddRAD sequencing, and bioinformatics

I used the previously described ddRADseq dataset (Chapter 1) for all analyses of genetic differentiation and structure. Details about the library preparation, sequencing, and bioinformatic parameters used to identify variable loci are described therein. Briefly, a total of 535 individual Yosemite toad (*Anaxyrus canorus*) tadpoles sampled from 90 meadows across Yosemite National Park (YOSE), and 109 individual tadpoles sampled from 12 meadows in Kings Canyon National Park (KICA), were sequenced using 2×100 bp sequencing on an Illumina HiSeq 2500. Stringent quality thresholds were used, and several measures were used in the Stacks v1.19 (Catchen et al. 2011, 2013) pipeline to avoid recovering markers from paralogous sequences. The particular analyses used in this study were designed for SNP (not haplotype) loci, so I randomly sampled one SNP for each "Read 1" of the dataset. I used a coverage threshold of 10 reads per genotyped individual. SNPs with minor allele frequency (MAF) <0.05 were not called, and loci were filtered from the dataset if absent from >25% of individuals.

Predictions of future selective pressure by climate change

I modeled future selective pressure in three general steps: (1) identification of loci that presently show patterns of climatic selection, (2) statistically modeling the potentially non-linear relationships between climatic selection and allele frequencies, and (3) forecasting those predictions into a plausible scenario of future climate change. Although I possessed data for both YOSE and KICA, I chose to focus on predictions for YOSE, to prevent any bias associated with modeling a disjunct distribution. In addition, YOSE is a useful sample of the species: it contains approximately 30% of the species across nearly its entire elevational distribution, a wide range of climatic conditions, and has seven different historical lineages. However, I did include KICA in step (1) to increase the confidence that outlier loci represent climatic (and not population) patterns.

I extracted mean climatic data values using a 500 m buffer around the coordinates of my sample locations. A total of 30 variables were used (Table 4.2): these included 19 bioclimatic variables from WorldClim 1.4 at 1 km resolution (Hijmans et al. 2005), and 11 hydrologic variables from the 2014 California BCM at 270 m resolution (Flint et al. 2013). The WorldClim (http://www.worldclim.org/) and BCM (http://climate.calcommons.org/) variables were taken from 30-year averages available on their respective hosting websites: 1960–1990 (WorldClim), and 1981–2010 (BCM). In order to account for the differing spatial resolutions, I only averaged the portions of each pixel overlapping the 500 m buffer (i.e. a weighted mean). Future projections of these 30 variables were generated using the Community Climate System Model (CCSM), under representation concentration pathway (RCP) 8.5 (Riahi et al. 2011). The RCP 8.5

scenario models a future where emissions continue to rise throughout the 21st century, leading to increase in global mean surface temperature of 2.6°C – 4.8°C over the century (IPCC 2014). Future climatic variables were taken from the available 20-to-30-year averages: 2061–2080 (WorldClim), and 2070–2099 (BCM).

A principal components analysis (PCA) was used to reduce the climatic data (in addition to elevation, latitude, and longitude) into a set of orthogonal predictors. I also included Moran's Eigenvector Maps (MEMs), orthogonal spatial predictors that describe spatial autocorrelation based on the arrangement of sample locations (Griffith 1996; Dray et al. 2006). These are important to include in landscape genomic models, because they can account for otherwise unmodeled variables, particularly from either isolation by distance or phylogeographic signal in separate lineages. There is some chance these MEMs may represent unsampled environmental variation of interest. However, for the purposes of identifying loci that are putatively responding to climatic selection, MEMs are best used as a nuisance variable that accounts for demographic structure. Failure to properly disentangle demographic and climatic patterns can result in many false positive "outlier" loci that reflect non-adaptive processes. I used the package adespatial (Dray et al. 2016) to build a spatial neighborhood with k=10 neighbors, inversely weight by the distance between points, then test for significance with 100 random permutations. I used the first four of these MEMs because they aligned with phylogeographic breaks. All variables were centered and scaled by standard deviations.

Regions of the genome that are adapting in response to natural selection can be identified in several different ways. Traditionally, the genome scan approach has the most

popular for anonymous loci, because it works from a very simple assumption: selection will produce loci with unusual values of F_{ST} (or other measure of population differentiation) relative to genetic diversity (such as H_E). This began with the Lewontin and Krakauer test (Lewontin and Krakauer 1973), and has since expanded to include many differentiation metrics with increasingly sophisticated null expectations (Beaumont and Nichols 1996; Beaumont and Balding 2004; Foll and Gaggiotti 2008; Excoffier et al. 2009; Bonhomme et al. 2010; Günther and Coop 2013; Duforet-Frebourg et al. 2014). However, with a few recent exceptions (Duforet-Frebourg et al. 2014; Luu et al. 2017), many of these methods do not explicitly control for the effect that population or demographic structure has in generating unusual F_{ST} distributions (Lotterhos and Whitlock 2014, 2015). Additionally, even if false positive or negative rates are adequately controlled, these methods do not explicitly model the cause of natural selection, so that outlier loci may have limited biological meaning. The major alternative to genome scans are so-called tests of genetic-environment association (GEA), which take advantage of the long-standing observation that phenotypic patterns often trace environmental gradients (Huxley 1939; Endler 1986). For anonymous genetic loci, this idea was first implemented using logistic regressions between SNP alleles and environment (Joost et al. 2007) without accounting for population structure (Meirmans 2012; De Mita et al. 2013). This shortcoming has since been remedied in methods that take the conditional likelihood of GEA, either using neutral genetic covariance matrices (Coop et al. 2010; Günther and Coop 2013; Guillot et al. 2014) or latent factors (Frichot

et al. 2013). Overall, the GEA approach has high power to detect true outliers when the spatial sampling of populations is large and evenly distributed.

I identified loci putatively responding to divergent selection (either directly or by physical linkage) using the program bayenv v2.0 (Hancock et al. 2008; Coop et al. 2010; Günther and Coop 2013). This Bayesian method explicitly tests for correlations between the allele frequencies of SNPs and environmental variables putatively selecting them, while accounting for overall genetic covariance by population structure ($X^T X$). Unlike other correlative methods, bayenv accounts for the effects of shared population history and uneven sampling noise, and then tests for linear environmental associations in a Bayesian framework. I used the first five climatic PCs representing >95% variance (along with the four MEMs), for all YOSE and KICA coordinates. I considered as outliers those loci having a significant association with at least one of the five principal components, determined by a Bayes Factor >20, and correlation coefficient (either Pearson or Spearman) >|0.1|. Additionally, I removed any outliers that contained significant correlations with any of the MEMs, as a second pass for removing the effects of population structure. I considered all remaining outlier loci as representatives of adaptive genetic diversity for climatic selection.

Finally, I used Gradient Forests (GF) to model the potentially non-linear relationships between outlier loci and climate, and to forecast the shift in adaptive optimum by future climate change. GF is an extension of random forests (Breiman 2001) developed for community ecology, which can model species abundances over non-linear environmental gradients, where cross-validated R^2 is aggregated across each species to

give an estimated community "turnover" for each predictor (Ellis et al. 2012). This idea has recently been applied to predict adaptive genomic "turnover," which in essence is a summary form of adaptive genetic variation as a function of the landscape (Fitzpatrick and Keller 2015; Bay et al. 2018). The GF model can then be used to forecast the amount of net allelic change required to match the future climate. I built GF models using the gradientForest package (Ellis et al. 2012), predicting the previously identified outlier loci with the 30 climatic variables described above. I used the raw variables instead of principal components because GF (unlike the linear models of bayenv) deals with multicollinearity by using a conditional permutation scheme to assess the importance of each variable, if it is collinear above a certain level (Strobl et al. 2008). I generated 500 bootstrapped trees, and calculated variable importance from conditional permutation distributions. Out-of-bag (OOB) permuting was conditioned on partitions of variables having absolute correlation >0.5.

I then predicted both current and future allelic state for each pixel on the map. For this purpose, variables were rescaled to the lowest resolution of the data (1 km) using bilinear interpolation, to prevent statistical bias that can result from combining different resolutions. However, once multi-variable predictions were made for each pixel, the combined prediction map was again resampled to a finer resolution for display purposes. A PCA was used to visualize the predicted adaptive genetic gradients as a map, with areas of similar adaptive genetic composition clustering together in multidimensional space. Since GF predicts allele frequency for each variable separately (weighted by

cumulative variable importance), I calculated future change as the Euclidean distance between current and future allele frequencies.

Estimated levels of neutral and adaptive genetic diversity

Neutral genetic diversity is indirectly related to heritable variation underlying traits under selection, because (1) populations with larger N_e contain a larger absolute number of favorable gene copies, and (2) larger N_e increases the efficacy of selection, by lowering the effective threshold for selective coefficients. However, adaptive genetic diversity at the specific traits under selection is directly responsible for genetic response to phenotypic selection, and hence is more important in the short term. Neutral and adaptive genetic variation may follow opposite patterns if, for example, strong selection has recently depleted adaptive genetic variation from a large population. Therefore, a comparison of both types of diversity can provide meaningful insight about adaptive potential.

I quantified overall neutral genetic diversity in two ways, based on a previous study (Chapter 1). First, I used explicit estimates of N_e at the meadow cluster scale, based on the single-sample linkage disequilibrium method (Do et al. 2014). I chose to generalize from this scale because individual meadow sample sizes were often too small for meaningful confidence intervals, leading to many missing estimates. Second, I quantified average gene diversity (π) across all loci, using STACKS (Catchen et al. 2011, 2013). I then calculated average gene diversity (π) across the outlier loci identified above (n=35), as an estimate of adaptive genetic variation. All patterns in relative park-wide

diversity were visualized using inverse distance weighted interpolation, implemented in the gstat package (Pebesma 2004; Gräler et al. 2016).

Predictions of future graph theoretical meadow connectivity

Previously (Chapter 3), I determined that isolation by distance (IBD) has a significant effect on genetic connectivity for distances up to 30 km along least cost paths (LCPs). I used that rationale to justify building gravity models for three distinct cutoffs of LCP transect length: 30 km, 20 km, and 10 km. Since adult toads move longer distances given more time, these separate cutoffs implicitly modeled environmental influence on dispersal at different spatiotemporal scales. However, in the present study, I had the goal of forecasting changes in genetic connectivity due to the next century of climate change. Therefore, I chose to only model connectivity with a 30 km LCP cutoff distance, because (1) toads can easily disperse up to 30 km in 100 years given a yearly maximum dispersal distance of >1 km (Liang 2010), and (2) transects spanning the range of plausible dispersal distances (i.e. 0–30 km) are the best indication of overall park-wide connectivity. In addition, I only considered LCPs that fall entirely within the known distribution of the species. All variables (described next) were extracted from a 500 m buffer of LCPs, as this broad bandwidth size should correspond to the broad influence of climate.

I used previously described gravity modeling techniques (see Chapter 3) to predict estimated F_{ST} between 90 Yosemite toad meadows in Yosemite National Park (YOSE). Here, I only mention the deviations from my previous workflow. Only climatic

predictors were included, from two sources (Table 4.2): 19 bioclimatic variables from WorldClim 1.4 at 1 km resolution (Hijmans et al. 2005), and 11 hydrologic variables from the 2014 California BCM at 270 m resolution (Flint et al. 2013). Models were constructed with the sole purpose of predicting present and future connectivity as a function of climate, without any goal of interpreting variable coefficients. Multicollinearity can inflate the variance of predictor coefficients and thus make them unusable for biological interpretation (Farrar and Glauber 1967; Graham 2003; Woolridge 2013; Vatcheva et al. 2016), however it has no effect on overall model fit or prediction accuracy (Kutner et al. 2005). Therefore, I considered all variables regardless of variance inflation factor (VIF), with the exception that any variable that was collinear with elevation (VIF >10) was removed. Elevation can have a strong influence on many other variables, and hence I isolated just the variables that could be distinguished from elevation. VIF between elevation and each other variable was calculated using the fmsb package (Nakazawa 2018) in R v3.3.3 (R Core Team 2018). All other variables were entered into forward stepwise selection, in an order determined by their effect on model AICc, and retained if a likelihood ratio test with the nested model (not containing that variable) was significant (Chapter 3). Lineage-specific slopes were fitted using the same process.

Predictive power of the final model was assessed in several ways, using conditional and marginal coefficients of determination (${R^2}_c$ for all effects, and ${R^2}_m$ for fixed effects; Nakagawa and Schielzeth 2013), and prediction coefficient (P^2). P^2 is defined as $1 - MSPE/{s^2}_y$, where $MSPE$ (mean square prediction error) is the sum of the

squared differences of y_i from $\hat{y}_i$, divided by sample size (Guénard et al. 2013). Predicted values ($\hat{y}_i$) were calculated by k-fold cross validation; I created training and validation datasets k=10 times by setting aside a random 10% of observations, and then predicted each fold of the data.

I then predicted both present and future connectivity across the entire network of Yosemite toad meadows in YOSE. First, connectivity was extrapolated to all 283 meadows known to have toad occupancy at some point during the last century (Berlow et al. 2013; VertNet 2018; Ostoja et al., in prep.; U.S. Geological Survey, unpublished data; R. Knapp, unpublished data), to estimate total connectivity for all potential sites. Second, future connectivity was forecast using projected values of each variable in the years 2061–2080 (WorldClim), or 2070–2099 (BCM). Future projections of these 30 variables were extracted using the Community Climate System Model (CCSM), under representation concentration pathway (RCP) 8.5 (Riahi et al. 2011). The RCP 8.5 scenario models a future where emissions continue to rise throughout the 21st century, leading to increase in global mean surface temperature of 2.6°C – 4.8°C over the century (IPCC 2014). All predictions were mapped onto igraph (Csárdi and Nepusz 2006) objects to measure changes in network attributes. Two graph theoretical attributes were measured for each meadow: average edge weight (mean of $1 - F_{ST}$), and Kleinberg's hub centrality score (Kleinberg 1999). Average edge weight has been shown to most closely reflect simulated and empirical gene flow (Koen et al. 2016). Hub centrality scores are the principal eigenvector of the weighted adjacency matrix $A^T A$, where the weights are

$1 - F_{ST}$. Hub centrality represents park-wide connectedness, with the highest values receiving the most overall (direct and indirect) gene flow.

RESULTS

Predictions of future selective pressure by climate change

The first five components of the PCA on climatic variables cumulatively contained 96.6% of the variance. PC1 (64.7%) had high loadings for all WorldClim variables overall, whereas PC2 (15.7%) and PC3 (6.4%) had highest loadings for differing aspects of runoff, recharge, and snowpack variables of the BCM dataset (Table S4.1; Fig. S4.1). The last two components corresponded most with temperature variability (PC4; 5.7%) and climatic water deficit variability caused by patterns of annual precipitation (PC5; 4.1%). Out of 16 MEMs found to have significantly positive spatial autocorrelation based on 100 random permutations, the first four matched known patterns of phylogeographic structure (Fig. S4.2): MEM1 distinguished the two parks (YOSE and KICA), MEM2 distinguished the most distant lineage in YOSE (Y-North), MEM3 separated Y-East from the remainder of the park, and MEM4 resolved the two lower-elevation lineages (Y-South and Y-West). Hence, these variables were included in the bayenv analysis to flag and remove possible instances where environmental correlations were false positives due to phylogeographic structure. A total of 35 bayenv outliers were identified out of 1672 SNPs examined, after removing 37 loci with MEM associations (Table S4.2). The most common bayenv predictors were PC1 (14 hits), elevation (10

hits), and PC2 (9 hits), whereas PC4 (6 hits) and PC5 (5 hits) were less common; the remaining predictors contained either one hit or no hits.

For the GF model of climatic-genetic association, 32 out of 35 SNPs had some predictive power (R^2 >0), with a mean R^2 0.29 (range: 0.04–0.65). I tried re-running the model with only the SNPs having R^2 >0.25 (16 SNPs, mean R^2=0.43), and the final results described below were not perceptibly changed; hence I kept all SNPs in the model. I found that adaptive genetic variation tracks two primary clines: a cline in summer precipitation (bio17) that roughly increases with elevation, and a cline in snowpack and summer runoff that generally tracks latitude, and increases sharply in the north (Fig. 4.1a). The most important predictors as determined by R^2 values from conditional permutations were: mean April 1 snowpack water content (aprpck_ave), mean summer runoff (runsum_ave), and precipitation of the driest quarter, i.e. summer (bio17), and the eight most important variables were all related to moisture (Fig. 4.1f). Snowpack and runoff variables were found to produce abrupt (non-linear) genetic turnover at the higher values of those predictors (i.e. in the north), whereas summer precipitation was found to have the most genetic turnover at the lower values of those predictors, at lower elevation (Fig. 4.1c–e). Genetic turnover was a more linear function for most other predictors (Fig. S4.3, S4.4).

These three climatic drivers separate the park (YOSE) into three major adaptive zones that roughly correspond to known lineages boundaries (see Chapter 1). Based on adaptive genetic clustering (Fig. 4.1), the northern region appears particularly adapted to a robust snowpack and associated runoff that may last longer into the season, but with

moderate summer precipitation (Y-North, East-North-A1). The eastern region appears adapted to a small to moderate snowpack with foreshortened runoff, but with the highest levels of summer precipitation at higher elevations (Y-East, East-South-A). The lower elevations (Y-South, Y-West) receive low to moderate April 1 snowpack, and the lowest amount of summer precipitation (despite having the highest winter precipitation). I found that $\delta(f)$, the predicted amount of allelic change required to remain in the adaptive optimum, was highest for Y-South, moderately high for Y-West, and variable for Y-North and its associated admixture zone (Fig. 4.1b). Y-East (and the southern admixture zone) had by far the lowest $\delta(f)$, for two reasons. First, snowpack and runoff levels are expected to decrease the least amount in Y-East (Fig. 4.1c–d). Second, although summer precipitation is forecasted to increase everywhere, and increase maximally in the high elevation localities occupied by Y-East, the genetic turnover function is much shallower at these high values (Fig. 4.1e). This means that less genetic change is expected to be necessary for the larger climatic changes in precipitation for Y-East. It may seem counterintuitive that increased summer precipitation could be maladaptive for amphibians, however the suite of changes that includes a shrinking snowpack and altered phenology of rainfall may indeed require important adaptations in behavior and physiology (see Discussion).

Predicted levels of neutral and adaptive genetic variation

Neutral genetic diversity as measured by N_e was found to be vastly higher in Y-East compared with other lineages, particularly in the far eastern region of Tioga Pass

(Fig. 4.2). Neutral diversity more broadly defined by meadow π was also high in Y-East, but high diversity was found across the high elevation meadows of Y-East, Y-North, and was maximal for the admixed meadows in the contact zone adjoining them (East-North-A1). In both cases, diversity at the lower elevations (Y-South and Y-West) was minimal. However, adaptive genetic variation as defined by meadow π for outlier loci showed a somewhat different pattern. Highest levels of diversity were still observed in Y-East, however a sharp drop-off was observed for Y-North. Instead, Y-West (which has previously been found to receive adaptive genetic introgression from Y-East, see Chapter 2) had moderate levels of diversity. Y-South remained the most genetically depauperate for adaptive variation.

My gravity model of current and anticipated future genetic connectivity echoed some of the patterns found in the results described above, but also shed some interesting new light. The best model with random coefficient slopes by lineage was significantly preferred over the best model without random slopes, based on AICc values and a likelihood ratio test (Table 4.3). The final model contained 15 variables, explained 95.1% of the total variance, and explained 35.4% of the climatic (fixed effect) variance. The prediction coefficient (P^2) from k-fold cross validation was 0.67, indicating accurate prediction at a rate of approximately two-thirds. In addition, all model residuals conformed to normality, appearing to be independent and identically distributed around the mean for both fixed and random effects (Fig. S4.5). Across all 283 known toad meadows, I predicted a current maximum value of meadow $\overline{F_{ST}}$ = 0.13, compared with a future maximum value of $\overline{F_{ST}}$ = 0.39, giving an average increase of 0.096 over this this

100-year time period (Fig. 4.3). This sharp decrease in predicted connectivity is not stationary; the most dramatic shifts are high-elevation. Y-North ($\bar{x}$ = 0.25 increase) and the northern admixture zone ($\bar{x}$ = 0.19 and $\bar{x}$ = 0.14 increase, respectively, for East-North-A1 and East-North-A2) are projected to be most impacted, followed by the southern admixture zone ($\bar{x}$ = 0.12 for East-South-A). Y-East, currently has the highest connectivity ($\overline{F_{ST}}$ of $\bar{x}$ = 0.053), and yet is also projected to have a sizeable increase of 0.10. In contrast to this seismic decrease in high-elevation connectivity, the low-elevation lineages (Y-South and Y-West) are actually projected to have slightly increased connectivity ($\bar{x}$ = 0.011 decrease in $\overline{F_{ST}}$).

These anticipated shifts in direct (meadow-meadow) connectivity were mirrored by patterns of graph theoretical connectivity (i.e. gene flow integrated over all indirect connections). The median value of Kleinberg's hub centrality score decreased from 0.68 to 0.65, and eigenvalue of the weighted adjacency matrix $A^T A$ decreased from 18,429.43 to 14,544.52. However, the major reversal in patterns of connectivity as measured by edge weight were not observed for hub centrality score (Fig. S4.6). Instead, I predicted Y-East to dominate the number of meadow connections both before and after 100 years of climate change, with mean hub centrality score shifting from 0.8975 to 0.8950. The largest hub centrality score shift was for Y-North meadows (from $\bar{x}$ = 0.18 to $\bar{x}$ = 0.13), and this actually increased the centrality of many Y-East meadows, by moving the center of the entire network eastward. Graph theoretical measures such as hub centrality score are partly influenced by edge weight ($1 - F_{ST}$ in this case), but more so by the number of connections (direct or indirect) a node has with the entire network. Therefore, if it is

assumed that raised values of F_{ST} do not imply meadow extirpations or zero migrants, hub centrality score should be more robust to climate change than edge weight. Hub centrality score may therefore be a good indicator of how gene flow and N_e interact. For example, I found that a meadow's probability of toad breeding ("Prob100GLM" from Berlow et al. 2013) correlates with hub centrality score, (ρ=0.343, df=279, p<0.001) much more than edge weight (ρ=0.160, df=279, p<0.01).

Although my use of multicollinear predictors in the gravity model precludes my direct interpretation of their coefficients, it is still possible to decipher the environmental cause of these results to some degree. The order of variables entered into the model is still meaningful, because it indicates which ones improved the model most at each step (Table 4.3). The most important variable was thus annual standard deviation of climatic water deficit (CWD), for dispersal environment between sites. CWD (potential minus actual evapotranspiration) is expected to undergo a large clinal shift by elevation. Higher-elevation sites are expected to have much higher standard deviations, the opposite pattern compared to that of low-elevation sites. I found that there is a significantly negative Mantel correlation between annual standard deviation of climatic water deficit and connectivity (ρ=0.304, p<0.001), suggesting that increased variability in CWD at higher elevations may play a role in the anticipated decrease of higher-elevation connectivity.

DISCUSSION

In the beginning, I presented a novel conceptual workflow for imagining conservation units, starting with the unifying process of lineage formation (Table 4.1).

First, all past lineages should be delineated and measured for adaptive divergence; previous work (Chapter 1) suggested that contrasting glacial refugia during the Pleistocene forced low-elevation Yosemite toad lineages (Y-South and Y-West) to favor different environmental conditions than high-elevation lineages (Y-East and Y-North). Therefore, I modeled these potential differences in my gravity models of meadow connectivity, by taking into account potentially different slopes of climatic variable coefficients by lineage. Second, inter-lineage contact zones should be identified, because they may harbor more extreme trait values through transgressive segregation, and possibly spread some of this novel diversity by adaptive introgression. In Chapter 2, six loci were found that putatively influence tadpole development and growth patterns, including one locus (LPIN3, a lipid metabolism gene) that was found introgressing from Y-East to Y-West. Third, the scale of current population boundaries and their main ecological pressures should be studied, because (1) demographic units are bulwarks maintaining genetic diversity for entire lineages, and (2) with appropriate selection pressures, these units could become incipient lineages. In Chapter 3, I found that most lineages are distinct (with only one large exception, a cluster of meadows in Y-East), and inferred that moist meadow environments lower connectivity by increasing site fidelity. In short, I recommend first studying the factors that go into the delimitation of traditional conservation units (ESUs and MUs), because they will likely have some influence on the future. However, my workflow prescribes a future-predictive approach to actually anticipating GEUs, which are the true units of interest for conservation. The conservation practitioner should build an understanding of how lineages of the focal species have

diversified, will diversify, and how best to channel that process into the future based on management goals.

Bowen (1998) partitioned out the concept of GEUs from Waples' (1991, 1995) definition of ESUs, calling them "progenitors of future evolutionary diversity," based on the idea that long-term isolation represented by ESUs is not the only road to future diversification. The problem that many researchers encounter is a lack of availability for morphological, behavioral, or other ecological data that would support the GEU designation, leading to partial ESU designations that only incorporate neutral genetic evidence (Fraser and Bernatchez 2001; de Guia and Saitoh 2007). Although other studies have prescribed the use of adaptive loci to define conservation unit boundaries (Bonin et al. 2007; Funk et al. 2012), or predicted future adaptation using experimental data (Hoffmann and Sgrò 2011), ancient DNA (Orsini et al. 2013; Fordham et al. 2014), or outlier loci (Fitzpatrick and Keller 2015) my study is the first to use both neutral and putatively adaptive genomic variation to build explicitly predictive conservation units.

The geographic distribution of Yosemite toad adaptations to climate

The GF model predicted that April 1 snowpack, summer runoff, and summer precipitation are driving the most climatic adaptation for Yosemite toads (Fig. 4.1). In addition, GF models any non-linear or threshold effects of climate, making it possible to pinpoint where these variables are most important. Snowpack and runoff had the most adaptive impact for Y-North and the northern admixture zone, whereas summer precipitation has an inordinate impact on low elevation lineages (Y-South and Y-West;

Figs. 4.1c–e, S4.3, S4.4). Collectively, there results suggest three adaptive genetic clusters (Fig. 4.1a): the northern region (high snowpack/runoff, moderate precipitation), eastern region (small–moderate snowpack, high precipitation), and low-elevation region (low–moderate snowpack, low precipitation). It should be clarified that summer and winter precipitation patterns show opposite patterns: the former increases with elevation, and the latter decreases with elevation.

There are several possible reasons for an adaptive tradeoff between higher snowpack versus higher summer rainfall. The first concerns the general tendency of Yosemite toads to behaviorally select shallow pools ($\bar{x}$ = 4.35 cm), both as breeding adults and as swimming tadpoles (Karlstrom 1962; Liang et al. 2017). Meadow pools continually fed by runoff from a robust snowpack (Y-North) are likely to retain water longer into the season, allowing tadpoles to invest metabolic resources more into growth than rapid development. Tadpoles in fast-drying pools that sporadically get replenished by summer rainfall (Y-East) may be adapted to the opposite strategy, with the added benefit of escaping predators sooner, and having a longer snow-free season with which to forage or disperse. Tadpole growth and development are generally seen as a life history tradeoff, where faster development comes at the expense of smaller size at metamorphosis (for a comparative analysis, see Richter-Boix et al. 2011). Yosemite toad tadpoles are known to aggregate together in the shallowest available habitat, in order to absorb maximum thermal energy and metamorphose more quickly (Brattstrom 1962), and this explains their abundance of melanism compared to the closely related *Anaxyrus boreas* (Karlstrom 1962). Tadpoles in Y-East also have the fastest observed development

rates, and the smallest metamorphic size of any lineage in YOSE (Chapter 3; P. Maier, unpublished data). In contrast to these two strategies, the low-elevation adaptive advantage is less clear. Tadpoles in these lineages (Y-South, Y-West) receive low to moderate snowmelt, and the lowest summer precipitation, although they experience the most winter precipitation. Thus, their early phenology may benefit from high winter and spring moisture, and afford them a much longer growing season. These low-elevation tadpoles are among the largest recorded in YOSE, supporting this hypothesis (Chapter 3). Amphibians at lower elevation are also adapted to higher critical thermal maxima than high-elevation conspecifics, and this has been shown in *A. boreas* (Brattstrom 1968; Snyder and Weathers 1975). Given their higher tolerance for much higher temperatures, they may be much more efficient at growing large (i.e. converting phytoplankton into biomass).

Relative impact of future climate change to different regions

I estimated values of $\delta(f)$, or the predicted amount of allelic change required to remain in the adaptive optimum. This is analogous to the response to selection S in the breeder's equation, or the amount of genetic change needed to maintain present associations with alleles and climate. I use the notation $\delta(f)$, because my method does not actually measure a shift in phenotype with experimentally validated levels of heritability, and should be taken as a relative rather than absolute measure. I found the largest $\delta(f)$ for Y-South and Y-West, moderate or mixed $\delta(f)$ for Y-North, and relatively low values of $\delta(f)$ for Y-East (Fig. 4.1b). This pattern of climate refugia being

at higher elevation, and especially concentrated in the Tioga Pass region of eastern YOSE, has also been observed for Belding's ground squirrels (*Urocitellus beldingi*; Maher et al. 2017; Morelli et al. 2017). Those authors observed that meadows with relatively low 20th century change in temperature and precipitation also have highest allelic richness and meadow connectivity for squirrels. Yosemite toad climate refugia have been forecast into the future by a previous study, although only based on species distribution modeling, and with now-outdated models of future climate (Viers et al. 2013; IPCC 2014). Species distribution modeling of that sort assumes that occupancy is binary, and predictions can be significantly improved by incorporating the heterogeneity of adaptation (Bush et al. 2016). Nevertheless, that study found an 89% range contraction by 2050–2070, with most refugia distributed at lower latitudes and higher elevations in the Sierra Nevada. Interestingly, the five most important variables in their boosted regression tree model were in the same categories as my three most important variables: snowpack, runoff, and precipitation (Viers et al. 2013).

My results show that Y-East and the southern admixture zone are protected primarily because summer snowpack and runoff have the smallest future decline (Fig. 4.1). Although change in summer precipitation was largest at the higher elevations (particularly for Y-North), the GF models showed much shallower genetic turnover at those particular values. This means that lower elevation sites are more sensitive to changes in precipitation. In addition, the lower elevation sites are faced with the dual challenge of reduced snowpack and relatively little increase of summer rainfall, which may have drastic phenological consequences for breeding. The relative contributions of

these two hydrological sources to breeding pools are almost certainly non-linear, and may have complex consequences, not only for tadpole survival, but also lipid and liver stores of hibernating adults (Licht 1975; Morton 1981). Thus, overall phenological shifts are predicted to inordinately impact the lower elevations, and make Y-East a climate refugium.

For the Sierra Nevada in general, climate change is anticipated to inordinately impact the mid-elevations. Under the same "business-as-usual" RCP 8.5 scenario that I used, Reich et al. (2018) predicted that by the end of the 21st century, there will be a 7°C rise in average springtime temperature, 64% drop in springtime snowpack volume, and 50 day phenological shift to earlier runoff of snowmelt. They also found that the most vulnerable elevations were 5,000–8,000 ft. due to high snow albedo feedback (warming, snowmelt, less reflectivity). This range overlaps with much of the elevational range of Y-South ($\bar{x}$ = 7,962±1,560 ft.) and Y-West ($\bar{x}$ = 7,986±963 ft.). The higher elevation lineages Y-East ($\bar{x}$ = 9,991 ±1,109 ft.) and Y-North ($\bar{x}$ = 9,682 ±739 ft.) are comparatively protected. For example, many of these meadows have a projected 70–90 day earlier mean runoff, compared to only 30–40 days for many high elevation toad meadows (Reich et al. 2018). Thus, Yosemite toad patterns of future adaptive change are expected to parallel patterns in climate change in the Sierra Nevada.

Factors influencing adaptive genetic variation

The breeder's equation (or "Lande's equation," in its multivariate form for polygenic traits), shows that response to selection depends on both selection pressure and

additive genetic variation for the trait(s) involved (Lush 1937; Lande 1976). Thus, it offers a useful conceptual tool for predicting future outcomes, by unifying the demand of environmental change (S) with the supply of adaptive genetic variation (h^2). In this study, I turned this equation upside down, by estimating the product of these two terms (R), i.e. the required genetic response to selection pressure to keep apace with adaptation to climate. This allowed me to ask how much regional genetic variation exists relative to the required response, or R/h^2.

Of course, without observational data on the true values of h^2 or S, my result is a relative one, highlighting which regions deserve the most attention. I assessed relative h^2 in two different ways: (1) by directly measuring neutral and adaptive genetic diversity, and (2) by predicting current and future levels of connectivity. Over the time scale of ~100 years, and for a species with very low $\overline{N_e}$ of ~20, migrants ($N_e m$) are likely to replenish h^2 much more quickly than new mutations ($N_e \mu$). It should be noted that gene flow can swamp out local adaptation (Lenormand 2002), but since adaptation likely occurs at broader scales than one meadow, this effect may be negligible compared to the effect of genetic rescue (Orr and Unckless 2014).

I found that N_e and adaptive genetic diversity (π) were higher in Y-East than anywhere else, although neutral π was highest within the northern admixture zone between Y-East and Y-North (Fig. 4.2). Other Sierra Nevada amphibians also reach their apex of diversity in this region (Rovito 2010), and possible explanations for this have been discussed elsewhere (Chapter 1). This suggests that Y-East currently has the largest h^2 for climate-related traits, but highlights the potential for admixture zones to produce

novel genetic diversity that may become useful in the future. The low-elevation lineages had the lowest neutral genetic diversity (N_e and π), however Y-West had moderate levels of adaptive genetic diversity. This may be related to the adaptive introgression that was previously observed between Y-East and Y-West (Chapter 2). Finally, Y-North had a conspicuous dearth of adaptive genetic diversity at the loci examined, compared to relatively high neutral genetic diversity. One possibility is that climatic selection has recently depleted this variation, which can happen after just a few generations of strong selection (Barton 1989). Y-North is the least connected lineage in YOSE, because it contains meadows that are highly fragmented by rugged terrain. Therefore, recently depleted h^2 would not be replenished by gene flow very quickly.

My gravity modeling prediction of network connectivity is largely concordant with patterns of meadow diversity, with some interesting exceptions. For example, the least diverse lineages (Y-South and Y-West) have the lowest hub centrality scores in the park (Fig. S4.6), but are actually expected to become slightly more connected ($\overline{F_{ST}}$ reduction of 0.011). The opposite pattern is true for the high-elevation lineages. Y-North is expected to undergo the largest decline in connectivity, with an increase in meadow $\overline{F_{ST}}$ of 0.25 compared with a park-wide value of 0.096 (Fig. 4.3). This is consistent with the above contention that climatic selection has recently depleted adaptive genetic variation in Y-North. Interestingly, I projected that the highly diverse Y-East will become slightly less connected too (mean shift in meadow $\overline{F_{ST}}$ from 0.053 to 0.1), however taking all indirect network connections into account, hub centrality score for Y-East remains much higher than any other lineage (Fig. S4.6). Compared to $\overline{F_{ST}}$, hub centrality score is

more correlated with meadow N_e and breeding probability (Berlow et al. 2013), and thus may reflect the combined effects of population connectivity and growth rate better. Networks make good biological models whenever a collection of units has emergent properties that depend on all of their interactions, such as in gene regulation, food web dynamics, and epidemiology (Proulx et al. 2005). Graph theoretical metrics such as hub centrality score may be superior to traditional population genetic statistics, because they integrate the indirect flow of gene copies (Dyer et al. 2010). Additionally, my inclusion of virtually all toad meadows in YOSE avoids the erroneous effect of "ghost populations" (Slatkin 2005), and hence my network predictions should be relatively unbiased. It is interesting to note that variability in climatic water deficit (CWD) seems to partly account for these predictions in altered network structure. This underscores the observation that organism dispersal may be under a distinct set of selection pressures (Lowe and McPeek 2014; Pflüger and Balkenhol 2014), and Yosemite toads may disperse less when hydric conditions are more variable (Chapter 3).

One more complication: selection for phenotypic plasticity

Weak selection due to low levels of h^2 at low elevation may be reinforced by the tendency for individual phenotypic plasticity to become more prevalent there. For example, if tadpole body size at lower elevation is poorly matched to the climate, this could be partly due to low genetic diversity, but also because plasticity is being selected. Larval body size at metamorphosis is an important predictor of successful recruitment, which may be an important determinant of population growth rate (Smith 1987; Berven

1990; Hughes 1990; Grosberg and Levitan 1992; Altwegg and Reyer 2003), although variation in adult survival is probably an equally large or larger predictor in low population sizes (Schmidt et al. 2005). For the wood frog (*Rana sylvatica*), body size at metamorphosis is strongly heritable at high elevation (0.66 ± SE 0.31), yet is not significant at lower elevation (Berven and Berven 1987). Yet there is also a separate pattern of life history traits being more plastic at lower elevation, within a species (Gugger et al. 2015). For example, the relative contributions of genetics versus environment in determining sex have been found to follow elevational and latitudinal gradients in fish (e.g. Lagomarsino and Conover 1993) and squamates (e.g. Pen et al. 2010). This means that phenotypic plasticity may inherently be more advantageous at lower elevations, because a generalist strategy is more successful than at montane or alpine elevations, where phenotypes must be fine-tuned to extreme environments (Huey and Hertz 1984; Gilchrist 1995).

Therefore, in addition to the low levels of h^2 I anticipate for Yosemite toads at low elevation based on lower adaptive genetic diversity and meadow connectivity, these populations may additionally be selecting for higher phenotypic plasticity, weakening the effect of natural selection. This strategy may have been adequate in the past, but will likely be maladaptive in the future. Nunney (2015) showed that for models of directional selection by climate change, plasticity was a double-edged sword: environmental tolerance went up because it was favored by selection in the short-term, but led to extirpations in the long-run, if the trait under selection was determined by multiple loci. This is supported by field studies: plasticity has generated a mixture of adaptive and

maladaptive responses to climate change, and is insufficient to keep up to speed with phenology changes (Phillimore et al. 2010; Urban et al. 2014). This also explains many cases where predictive accuracy of the breeder's equation is low, because plasticity can be conflated with or mask h^2 (Gienapp et al. 2008; Pemberton 2010; Merilä and Hendry 2014). With this caveat in mind, my workflow still provides a reasonable expectation for the future, and a null model for comparing future outcomes.

Are other selective pressures relevant to the future of Yosemite toads?

An important question for future population ecological studies to address is the relative extent to which climate change and disease are impacting local growth rates and demographic connectivity. While Y-East is the most genetically robust and ecologically vibrant region of YOSE for the Yosemite toad, the only recorded demographic bottleneck for the species occurred in meadows surrounding Tioga Pass (Sherman and Morton 1993). Massive extirpations were witnessed, and although only a handful of specimens showed histological evidence of *Bd* infections, Dodge and Vredenberg (2012) later found a correlation between specimen infection intensity and phase of bottleneck. However, they also found that prevalence (0–25%) and infection intensity are currently low, more so in adults than subadults. It has later been shown that *Bd* is deadly to Yosemite toad metamorphs if zoospore equivalents are large enough, especially if those metamorphs were subjected to desiccating pond conditions as tadpoles (Lindauer 2018). Certainly, chytridiomycosis is driving declines in a nearby species of alpine toad (Muths et al. 2003), and although disease refugia seem to exist at extreme temperatures, population

persistence seems to depend on the interaction between disease, climate, and demography (Lambert et al. 2016; Mosher et al. 2018). Given that there is likely a synergistic effect between *Bd* and climate change for Yosemite toads (Lips et al. 2008; Rohr and Raffel 2010), future work should extend my method with a plausible model of future disease dynamics and demography.

CONCLUSIONS AND RECOMMENDATIONS TO MANAGEMENT

Altogether, I summarize the future of adaptation and persistence for Yosemite toad lineages as follows. Three adaptive genetic regions exist in YOSE currently (north, east, and low elevation). I circumscribe these regions as the three GEUs, and rank them in the following order of vulnerability: low elevation, north, and east. Since the northern admixture zone contains large amounts of neutral genetic variation, and no evidence of outbreeding depression, this region should be protected as a potentially critical interface for replenishing lost variation in the "north" GEU. Admixed toads might serve as important stock for reintroduction efforts, but further experimental work is needed to verify their viability under realistic ecological conditions. The "low elevation" GEU consists of two past lineages that are not each other's closest relatives. They are exchanging very few migrants, yet are adapted to similar climates, and experiencing similar selective pressure. Therefore, if reintroductions are deemed necessary, these lineages may be considered interchangeable. Admixing these genetically depauperate toads may have beneficial effects on their response to selection. My novel workflow in this study gives a forward-looking perspective on adaptive genetic cohesion, and provides

an example of modeling tools (i.e. Gradient Forests and gravity modeling) for grouping populations into conservation units. Although my approach was entirely based on molecular data, the Yosemite toad does not lend itself to ecological or experimental evolutionary study due to its secrecy and difficulty accessing its remote, rugged habitat. I suggest my approach as a useful toolkit for planning conservation efforts on poorly studied and non-model species of concern.

REFERENCES

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11:697–709.

Allendorf, F. W., G. Luikart, and S. N. Aitken. 2013. Conservation and the genetics of populations. 2nd ed. Wiley Blackwell, Oxford, UK.

Altwegg, R., and H.-U. Reyer. 2003. Patterns of natural selection on size at metamorphosis in water frogs. Evolution 57:872–882.

Avise, J. C. 1994. Molecular markers, natural history, and evolution. Chapman & Hall, New York, NY.

Barton, N. 1989. Evolutionary quantitative genetics: how little do we know? Annu. Rev. Genet. 23:337–370.

Bay, R. A., R. J. Harrigan, V. L. Underwood, H. L. Gibbs, T. B. Smith, and K. Ruegg. 2018. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science 359:83–86.

Beaumont, M. A., and D. J. Balding. 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. 13:969–980.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. London B 263:1619–1626.

Berlow, E. L., R. Knapp, S. M. Ostoja, R. J. Williams, H. McKenny, J. R. Matchett, Q. Guo, G. M. Fellers, P. Kleeman, M. L. Brooks, and L. Joppa. 2013. A network extension of species occupancy models in a patchy environment applied to the Yosemite toad (*Anaxyrus canorus*). PLoS One 8:e72200.

Berven, K. A. 1990. Factors affecting population fluctuations in larval and adult stages of the wood frog (*Rana sylvatica*). Ecology 71:1599–1608.

Berven, K. A., and K. A. Berven. 1987. The heritable basis of variation in larval developmental patterns in the wood frog (*Rana sylvatica*). Evolution 41:1088–1097.

Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah, S. Blott, and M. SanCristobal. 2010. Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics 186:241–262.

Bonin, A., F. Nicole, F. Pompanon, C. Miaud, and P. Taberlet. 2007. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. Conserv. Biol. 21:697–708.

Bowen, B. W. 1999. Preserving genes, species, or ecosystems? healing the fractured foundations of conservation policy. Mol. Ecol. 8:S5–S10.

Bowen, B. W. 1998. What is wrong with ESUs?: the gap between evolutionary theory and conservation principles. J. Shellfish Res. 17:1355–1358.

Brattstrom, B. H. 1968. Thermal acclimation in anuran amphibians as a function of latitude and altitude. Comp. Biochem. Physiol. 24:93–111.

Brattstrom, B. H. 1962. Thermal control of aggregation behavior in tadpoles. Herpetologica 18:38–46.

Breiman, L. 2001. Random forests. Mach. Learn. 45:5–32.

Brown, C., M. P. Hayes, G. A. Green, D. C. Macfarlane, and A. J. Lind. 2015. Yosemite toad conservation assessment. USDA Forest Service report. Sonora, CA.

Bush, A., K. Mokany, R. Catullo, A. Hoffmann, V. Kellermann, C. Sgrò, S. McEvey, and S. Ferrier. 2016. Incorporating evolutionary adaptation in species distribution modelling reduces projected vulnerability to climate change. Ecol. Lett. 19:1468–1478.

Casacci, L. P., F. Barbero, and E. Balletto. 2014. The "evolutionarily significant unit" concept and its applicability in biological conservation. Ital. J. Zool. 81:182–193.

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

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes, Genomes, Genet. 1:171–182.

Catullo, R. A., S. Ferrier, and A. A. Hoffmann. 2015. Extending spatial modelling of climate change responses beyond the realized niche: estimating, and accommodating, physiological limits and adaptive evolution. Glob. Ecol. Biogeogr. 24:1192–1202.

Charmantier, A., and P. Gienapp. 2014. Climate change and timing of avian breeding and migration: Evolutionary versus plastic changes. Evol. Appl. 7:15–28.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411–1423.

Crandall, K. A., O. R. P. Bininda-Emonds, G. M. Mace, and R. K. Wayne. 2000. Considering evolutionary processes in conservation biology. Trends Ecol. Evol. 15:290–295.

Csárdi, G., and T. Nepusz. 2006. The igraph software package for complex network research. InterJournal Complex Syst. 1695:1–9.

de Guia, A. P. O., and T. Saitoh. 2007. The gap between the concept and definitions in the evolutionarily significant unit: the need to integrate neutral genetic variation and adaptive variation. Ecol. Res. 22:604–612.

De Mita, S., A.-C. Thuillet, L. Gay, N. Ahmadi, S. Manel, J. Ronfort, and Y. Vigouroux. 2013. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Mol. Ecol. 22:1383–1399.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation. Pp. 57–75 *in* D. J. Howard and S. H. Berlocher, eds. Endless Forms: Species and Speciation. Oxford University Press, Oxford, UK.

Dizon, A. E., C. Lockyer, W. F. Perrin, D. P. Demaster, and J. Sisson. 1992. Rethinking the stock concept: a phylogeographic approach. Conserv. Biol. 6:24–36.

Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Resour. 14:209–214.

Dodge, C. M., and V. T. Vredenburg. 2012. The sad song of the Yosemite toad: the role of the amphibian chytrid fungus in an enigmatic decline. Paper presented at the annual convention of the Ecological Society of America, Portland, OR.

Dray, S., G. Blanchet, D. Borcard, G. Guenard, T. Jombart, G. Larocque, P. Legendre, and H. H. Wagner. 2016. adespatial: multivariate multiscale spatial analysis. R package version 0.0-4.

Dray, S., P. Legendre, and P. R. Peres-Neto. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Modell. 196:483–493.

Drost, C., and G. Fellers. 1994. Decline of frog species in the Yosemite section of the Sierra Nevada. National Park Service report. Davis, CA.

Duforet-Frebourg, N., E. Bazin, and M. G. B. Blum. 2014. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. Mol. Biol. Evol. 31:2483–2495.

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

Ellis, N., S. J. Smith, and C. R. Pitcher. 2012. Gradient forests: calculating importance gradients on physical predictors. Ecology 93:156–168.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press, Princeton, N.J.

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

Falconer, D. S. 1960. Introduction to quantitative genetics. Oliver and Boyd Limited, Edinburgh, UK.

Farrar, D. E., and R. R. Glauber. 1967. Multicollinearity in regression analysis: the problem revisited. Rev. Econ. Stat. 49:92–107.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University Press, Oxford, UK.

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

Flint, L. E., A. L. Flint, J. H. Thorne, and R. Boynton. 2013. Fine-scale hydrologic modeling for regional landscape applications: the California basin characterization model development and performance. Ecol. Process. 2:1–21.

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

Fordham, D. A., B. W. Brook, C. Moritz, and D. Nogués-Bravo. 2014. Better forecasts of range dynamics using genetic data. Trends Ecol. Evol. 29:436–443.

Frankel, O. H., and M. E. Soulé. 1981. Conservation and evolution. Cambridge University Press, Cambridge, UK.

Frantz, A. C., S. Cellina, A. Krier, L. Schley, and T. Burke. 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? J. Appl. Ecol. 46:493–505.

Fraser, D. J., and L. Bernatchez. 2001. Adaptive evolutionary conservation: towards a unified concept for delineating conservation units. Mol. Ecol. 10:2741–2752.

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

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

Gienapp, P., C. Teplitsky, J. S. Alho, J. A. Mills, and J. Merilä. 2008. Climate change and evolution: disentangling environmental and genetic responses. Mol. Ecol. 17:167–178.

Gilchrist, G. W. 1995. Specialists and generalists in changing environments I: fitness landscapes of thermal sensitivity. Am. Nat. 146:252–270.

Graham, M. H. 2003. Confronting multicollinearity in cological multiple regression. Ecology 84:2809–2815.

Gräler, B., E. Pebesma, and G. Heuvelink. 2016. Spatio-temporal interpolation using gstat. R J. 8:204–218.

Green, D. E., and C. K. Sherman. 2001. Diagnostic histological findings in Yosemite toads (*Bufo canorus*) from a die-off in the 1970s. J. Herpetol. 35:92–103.

Griffith, D. A. 1996. Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. Can. Geogr. 40:351–367.

Grosberg, R. K., and D. R. Levitan. 1992. For adults only? supply-side ecology and the history of larval biology. Trends Ecol. Evol. 7:130–133.

Guénard, G., P. Legendre, and P. Peres-Neto. 2013. Phylogenetic eigenvector maps: a framework to model and predict species traits. Methods Ecol. Evol. 4:1120–1131.

Gugger, S., H. Kesselring, J. Stöcklin, and E. Hamann. 2015. Lower plasticity exhibited by high- versus mid-elevation species in their phenological responses to manipulated temperature and drought. Ann. Bot. 116:953–962.

Guillot, G., R. Vitalis, A. le Rouzic, and M. Gautier. 2014. Detecting correlation between allele frequencies and environmental variables as a signature of selection: a fast computational approach for genome-wide studies. Spat. Stat. 8:145–155.

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

Haldane, J. 1957. The cost of natural selection. J. Genet. 55:511–524.

Hancock, A. M., D. B. Witonsky, A. S. Gordon, G. Eshel, J. K. Pritchard, G. Coop, and A. Di Rienzo. 2008. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet. 4:e32.

Hastings, A. 1993. Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. Ecology 74:1362–1372.

Heywood, J. S. 2005. An exact form of the breeder's equation for the evolution of a quantitative trait under natural selection. Evolution 59:2287–2298.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25:1965–1978.

Hoffmann, A. A., and C. M. Sgrò. 2011. Climate change and evolutionary adaptation. Nature 470:479–485.

Houchmandzadeh, B. 2014. An alternative to the breeder's and Lande's equations. G3 (Bethesda). 4:97–108.

Huey, R. B., and P. E. Hertz. 1984. Is a jack-of-all-temperatures a master of none? Evolution 38:441–444.

Hughes, T. P. 1990. Recruitment limitation, mortality, and population regulation in open systems: a case study. Ecology 71:12–20.

Huxley, J. S. 1939. Clines: an auxiliary method in taxonomy. Bijdr. tot Dierkd. 27:491–520.

IPCC. 2014. Climate change 2014: synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland.

Jennings, M., and M. Hayes. 1994. Amphibian and reptile species of special concern in California. California Department of Fish & Game report. Rancho Cordova, CA.

Joost, S., a Bonin, M. W. Bruford, L. Després, C. Conord, G. Erhardt, and P. Taberlet. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol. Ecol. 16:3955–3969.

Jordan, D. S. 1908. The law of geminate species. Am. Nat. 42:73–80.

Karlstrom, E. L. 1962. The toad genus *Bufo* in the Sierra Nevada of California. Unviersity Calif. Publ. Zool. 62:1–104.

Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment. J. ACM 46:604–632.

Koen, E. L., J. Bowman, and P. J. Wilson. 2016. Node-based measures of connectivity in genetic networks. Mol. Ecol. Resour. 16:69–79.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li. 2005. Applied linear statistical models. 5th ed. McGraw-Hill, New York, NY.

Lagomarsino, I. V, and D. O. Conover. 1993. Variation in environmental and genotypic sex-determining mechanisms across a latitudinal gradient in the fish, *Menidia menidia*. Evolution 47:487–494.

Lambert, B. A., R. A. Schorr, S. C. Schneider, and E. Muths. 2016. Influence of demography and environment on persistence in toad populations. J. Wildl. Manage. 80:1256–1266.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends Ecol. Evol. 17:183–189.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175–195.

Liang, C. T. 2010. Habitat modeling and movements of the Yosemite toad (*Anaxyrus* (=*Bufo*) *canorus*) in the Sierra Nevada, California. Davis, CA: University of California, Davis. Ph.D. dissertation.

Liang, C. T., R. L. Grasso, J. J. Nelson-Paul, K. E. Vincent, and A. J. Lind. 2017. Fine-Scale habitat characteristics related to occupancy of the Yosemite toad, *Anaxyrus canorus*. Copeia 105:120–127.

Licht, E. 1975. Comparative life history features of the western spotted frog, *Rana pretiosa*, from low- and high-elevation populations. Can. J. Zool. 53:1254–1257.

Lindauer, A. 2018. The role of hydroperiod and fluctuating temperature on disease dynamics: a disease ecology approach to understanding Yosemite toad declines. Reno, NV: University of Nevada Reno. M.S. thesis.

Lips, K. R., J. Diffendorfer, J. R. Mendelson, and M. W. Sears. 2008. Riding the wave: reconciling the roles of disease and climate change in amphibian declines. PLoS Biol. 6:e72.

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

Lotterhos, K. E., and M. C. Whitlock. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol. Ecol. 24:1031–1046.

Lowe, W. H., and M. A. McPeek. 2014. Is dispersal neutral? Trends Ecol. Evol. 29:444–450.

Lush, J. L. 1937. Animal breeding plans. Iowa State College Press, Ames, IA.

Luu, K., E. Bazin, and M. G. B. Blum. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. 17:67–77.

Maher, S. P., T. L. Morelli, M. Hershey, A. L. Flint, L. E. Flint, C. Moritz, and S. R. Beissinger. 2017. Erosion of refugia in the Sierra Nevada meadows network with climate change. Ecosphere 8:1–17.

Mallet, J. 2001. Gene flow. Pp. 337–360 *in* W. I.P., R. D.R., and C. D. Thomas, eds. Insect Movement: Mechanisms and Consequences. CAB International, London, UK.

Meirmans, P. G. 2012. The trouble with isolation by distance. Mol. Ecol. 21:2839–2846.

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

Møller, A. P., and T. Szép. 2005. Rapid evolutionary change in a secondary sexual character linked to climatic change. J. Evol. Biol. 18:481–495.

Morelli, T. L., S. P. Maher, M. C. W. Lim, C. Kastely, L. M. Eastman, L. E. Flint, A. L. Flint, S. R. Beissinger, and C. Moritz. 2017. Climate change refugia and habitat connectivity promote species persistence. Clim. Chang. Responses 4:8.

Moritz, C. 1994. Defining “evolutionarily significant units” for conservation. Trends Ecol. Evol. 9:373–375.

Moritz, C. 2002. Strategies to protect biological diversity and the evolutionary process that sustain it. Syst. Biol. 51:238–254.

Morrissey, M. B., L. E. B. Kruuk, and A. J. Wilson. 2010. The danger of applying the breeder’s equation in observational studies of natural populations. J. Evol. Biol. 23:2277–2288.

Morrissey, M. B., D. J. Parker, P. Korsten, J. M. Pemberton, L. E. B. Kruuk, and A. J. Wilson. 2012. The prediction of adaptive evolution: Empirical application of the secondary theorem of selection and comparison to the breeder's equation. Evolution 66:2399–2410.

Morton, M. 1981. Seasonal changes in total body lipid and liver weight in the Yosemite toad. Copeia 1981:234–238.

Mosher, B. A., L. L. Bailey, E. Muths, and K. P. Huyvaert. 2018. Host–pathogen metapopulation dynamics suggest high elevation refugia for boreal toads. Ecol. Appl. 28:926–937.

Muths, E., P. Stephen Corn, A. P. Pessier, and D. Earl Green. 2003. Evidence for disease-related amphibian decline in Colorado. Biol. Conserv. 110:357–365.

Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol. Evol. 4:133–142.

Nakazawa, M. 2018. fmsb: functions for medical statistics book with some demographic data. R package version 0.5.2.

Nunney, L. 2015. Adapting to a changing environment: modeling the interaction of directional selection and plasticity. J. Hered. 107:15–24.

Nunney, L. 2003. The cost of natural selection revisited. Ann. Zool. Fennici 185–194.

Orr, H. A., and R. L. Unckless. 2014. The population genetics of evolutionary rescue. PLoS Genet. 10:e1004551.

Orsini, L., K. Schwenk, L. De Meester, J. K. Colbourne, M. E. Pfrender, and L. J. Weider. 2013. The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments. Trends Ecol. Evol. 28:274–282.

Ostoja, S. M., S. R. Lee, J. R. Matchett, H. McKenny, M. L. Brooks, R. A. Knapp, P. A. Maier, and N. Danielle. 2018. Yosemite toad (*Anaxyrus* [*Bufo*] *canorus*) breeding occupancy of meadows in Yosemite and Kings Canyon National Parks, California. Unpublished manuscript.

Palsbøll, P., M. Berube, and F. Allendorf. 2007. Identification of management units using population genetic data. Trends Ecol. Evol. 22:11–16.

Pebesma, E. J. 2004. Multivariable geostatistics in S: The gstat package. Comput. Geosci. 30:683–691.

Pemberton, J. M. 2010. Evolution of quantitative traits in the wild: mind the ecology. Philos. Trans. R. Soc. B Biol. Sci. 365:2431–2438.

Pen, I., T. Uller, B. Feldmeyer, A. Harts, G. M. While, and E. Wapstra. 2010. Climate-driven population divergence in sex-determining systems. Nature 468:436–438.

Pflüger, F. J., and N. Balkenhol. 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. Mol. Ecol. 23:2146–2156.

Phillimore, A. B., J. D. Hadfield, O. R. Jones, and R. J. Smithers. 2010. Differences in spawning date between populations of common frog reveal local adaptation. Proc. Natl. Acad. Sci. 107:8292–8297.

Pigliucci, M. 2006. Genetic variance-covariance matrices: a critique of the evolutionary quantitative genetics research program. Biol. Philos. 21:1–23.

Pigliucci, M., and C. D. Schlichting. 1997. On the limits of quantitative genetics for study of phenotypic evolution. Acta Biotheor. 45:143–160.

Proulx, S. R., D. E. L. Promislow, and P. C. Phillips. 2005. Network thinking in ecology and evolution. Trends Ecol. Evol. 20:345–353.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ratliff, R. D. 1985. Meadows in the Sierra Nevada of California: state of knowledge. U.S. Forest Service report. Berkeley, CA.

Razgour, O. 2015. Beyond species distribution modeling: a landscape genetics approach to investigating range shifts under future climate change. Ecol. Inform. 30:250–256.

Reich, K. D., N. Berg, D. B. Walton, M. Schwartz, F. Sun, X. Huang, and A. Hall. 2018. Climate change in the Sierra Nevada: California's water future. UCLA Center for Climate Science report. Los Angeles, CA.

Reznick, D. 2016. Hard and soft selection revisited: how evolution by natural selection works in the real world. J. Hered. 107:3–14.

Riahi, K., S. Rao, V. Krey, C. Cho, V. Chirkov, G. Fischer, G. Kindermann, N. Nakicenovic, and P. Rafaj. 2011. RCP 8.5–a scenario of comparatively high greenhouse gas emissions. Clim. Change 109:33–57.

Richter-Boix, A., M. Tejedo, and E. L. Rezende. 2011. Evolution and plasticity of anuran larval development in response to desiccation: a comparative analysis. Ecol. Evol. 1:15–25.

Roff, D. 2000. The evolution of the G matrix: Selection or drift? Heredity 84:135–142.

Rohr, J. R., and T. R. Raffel. 2010. Linking global climate and temperature variability to widespread amphibian declines putatively caused by disease. Proc. Natl. Acad. Sci. 107:8269–8274.

Rovito, S. M. 2010. Lineage divergence and speciation in the web-toed salamanders (Plethodontidae: *Hydromantes*) of the Sierra Nevada, California. Mol. Ecol. 19:4554–4571.

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

Schmidt, B. R., R. Feldmann, and M. Schaub. 2005. Demographic processes underlying population growth and decline in Salamandra salamandra. Conserv. Biol. 19:1149–1156.

Sherman, C., and M. Morton. 1993. Population declines of Yosemite toads in the eastern Sierra Nevada of California. J. Herpetol. 27:186–198.

Slatkin, M. 2005. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Mol. Ecol. 14:67–73.

Smith, D. C. 1987. Adult recruitment in chorus frogs: effects of size and date at metamorphosis. Ecology 68:344–350.

Snyder, G. K., and W. W. Weathers. 1975. Temperature adaptations in amphibians. Am. Nat. 109:93–101.

Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. BMC Bioinformatics 9:1–11.

Sunday, J. M., R. N. Crim, C. D. G. Harley, and M. W. Hart. 2011. Quantifying rates of evolutionary adaptation in response to ocean acidification. PLoS One 6:1–8.

Thompson, B., A. Wright, and H. B. Shaffer. 2016. Yosemite toad (*Bufo canorus*). Pp. 112–123 *in* California amphibian and reptile species of special concern.

Urban, M. C., J. L. Richardson, and N. A. Freidenfelds. 2014. Plasticity and genetic adaptation mediate amphibian and reptile responses to climate change. Evol. Appl. 7:88–103.

US Fish & Wildlife Service. 2014. Endangered and threatened wildlife and plants; endangered status for the Sierra Nevada yellow-legged frog and the northern distinct population segment of the mountain yellow-legged frog, and threatened status for the Yosemite toad: final rule. Fed. Regist. 1–56.

Vatcheva, K. P., M. Lee, J. B. McCormick, and M. H. Rahbar. 2016. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiol. 6:1–20.

VertNet. 2018. Distributed databases with backbone. Available online at: http://portal.vertnet.org/search?q=specificepithet:canorus+genus:(anaxyrus+OR+bufo)+mappable:1

Viers, J. H., S. E. Purdy, R. A. Peek, A. Fryjoff-Hung, N. R. Santos, J. V. Katz, J. D. Emmons, D. V Dolan, and S. M. Yarnell. 2013. Montane meadows in the Sierra Nevada: changing hydroclimatic conditions and concepts for vulnerability assessment. Centre for Watershed Sciences report. Davis, CA.

Visscher, P. M., W. G. Hill, and N. R. Wray. 2008. Heritability in the genomics era – concepts and misconceptions. Nat. Rev. Genet. 9:255–266.

Vogler, A. P., and R. Desalle. 1994. Diagnosing units of conservation management. Conserv. Biol. 8:354–363.

Wallace, B. 1975. Hard and soft selection revisited. Evolution 29:465–473.

Waples, R. S. 1995. Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. Am. Fish. Soc. Symp. 17:8–27.

Waples, R. S. 1991. Pacific salmon, *Oncorynchus spp.*, and the definition of "species" under the Endangered Species Act. Mar. Fish. Rev. 53:11–22.

Woolridge, J. M. 2013. Introductory econometrics: a modern approach. 5th ed. South-Western CENGAGE Learning, Mason, OH.

TABLES AND FIGURES

Table 4.1 Conservation units reflecting evolutionary pattern and process. From the perspective of present day, population genetic processes from each time period uniquely shape evolutionary potential within a species. The formation of lineages that will persist and adapt through time depends on current adaptive or reproductive differences (ESUs), recent adaptive admixture (GEUs in contact zones), the current maintenance of genetic diversity (MUs), and sufficient population fitness to survive future environmental change (GEUs). Consequently, four classes of genetic patterns reflecting these four processes can and should be separately modeled to accurately delineate intraspecific conservation units.

	Time Frame			
	Ancient Past	**Recent Past**	**Present**	**Near Future**
Process	I. Lineage Isolation II. Lineage Adaptation	Secondary contact with admixture, introgression	Demographically independent populations	Future adaptation, population persistence
Lineage Formation				
Evo. Forces	I. Drift-mutation balance II. Divergent selection	Outbreeding, hybrid vigor, recombination	Drift-migration balance	Hard and soft selection
Pattern	I. Fixation of alleles, or reciprocal monophyly II. Non-exchangeability	High genome-wide π, and inter-population H_O, unique alleles	Distinct gene pools with dispersal rates < 10%	Correlation of outlier alleles with future environmental gradients
Units	ESUs	GEUs (in contact zones)	MUs (potentially hierarchical)	GEUs
References	Ryder (1986), Waples (1991), Moritz (1994), Crandall et al. (2000)	Allendorf et al. (2001), Hedrick et al. (2013), Klütsch et al. (2016)	Dizon et al. 1992, Hastings (1993), Waples and Gaggiotti (2006), Palsbøll et al. (2007)	Bowen (1998, 1999), Fitzpatrick and Keller (2015)

Table 4.2 Climatic data definitions and sources. All precipitation variables are in units of mm, and all temperature variables are in units of °C. The timespans were matched as nearly as possible between datasets, based on the available summaries.

Variable	Definition
Source: WorldClim (Bioclim), current timespan 1960–1990, future timespan 2061–2080	
bio1	Annual Mean Temperature
bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
bio3	Isothermality (bio2/bio7) (*100)
bio4	Temperature Seasonality (standard deviation *100)
bio5	Max Temperature of Warmest Month
bio6	Min Temperature of Coldest Month
bio7	Temperature Annual Range (bio5-bio6)
bio8	Mean Temperature of Wettest Quarter
bio9	Mean Temperature of Driest Quarter
bio10	Mean Temperature of Warmest Quarter
bio11	Mean Temperature of Coldest Quarter
bio12	Annual Precipitation
bio13	Precipitation of Wettest Month
bio14	Precipitation of Driest Month
bio15	Precipitation Seasonality (Coefficient of Variation)
bio16	Precipitation of Wettest Quarter
bio17	Precipitation of Driest Quarter
bio18	Precipitation of Warmest Quarter
bio19	Precipitation of Coldest Quarter
Source: California BCM 2014, current timespan 1981–2010, future timespan 2070–2099	
aprpck_ave	Mean April 1 snowpack snow water equivalent
aprpck_std	Std. dev. April 1 snowpack snow water equivalent
cwd_ave	Annual mean climatic water deficit; potential minus actual evapotranspiration
cwd_std	Annual std. dev. climatic water deficit; potential minus actual evapotranspiration
cwd_sum_ave	Summer mean climatic water deficit; potential minus actual evapotranspiration
rch_ave	Annual mean recharge; amount of water that penetrates below the root zone
rch_std	Annual std. dev. recharge; amount of water that penetrates below the root zone
rch_sum_ave	Annual summer recharge; amount of water that penetrates below the root zone
run_ave	Annual mean runoff; amount of water that becomes stream flow
run_std	Annual std. dev. runoff; amount of water that becomes stream flow
run_sum_ave	Annual summer runoff; amount of water that becomes stream flow

Table 4.3 Climatic gravity modeling results. Models with the maximum number of variables and random slopes that significantly improved model fit are shown. The best model with (bold) and without random slopes is shown for contrast. The overall best model was defined by having the lowest AICc score for each variable added, where adding that variable gave a significant likelihood ratio test score. All variables are listed in order of their inclusion during the forward-selection, and signs represent coefficient directions with respect to meadow connectivity ($1 - \sin^{-1}\sqrt{F_{ST}}$). "AT" and "BT" refer to at-site and between-site measurements, respectively. Only fixed effect signs are shown. Predictive power is represented by conditional and marginal coefficients of determination ($R^2{}_c$ for all effects, and $R^2{}_m$ for fixed effects) and prediction coefficient (P^2).

Formula	Rand. Slopes	-lnL	AICc	BIC	$R^2{}_c$	$R^2{}_m$	P^2
D - BT(cwd_std) - BT(rchsum_ave) - AT(runsum_ave) - BT(bio15) + AT(cwd_ave) + AT(aprpck_std) - BT(aprpck_std) + BT(aprpck_ave) - BT(rch_ave) + BT(run_std) - AT(cwdsum_ave) + BT(rch_std) + BT(cwd_ave) - BT(cwdsum_ave) - AT(aprpck_ave)	-	1715.11	-3391.42	-3299.45	0.72	0.39	0.65
	✔	**1737.11**	**-3422.84**	**-3302.15**	**0.95**	**0.35**	**0.67**

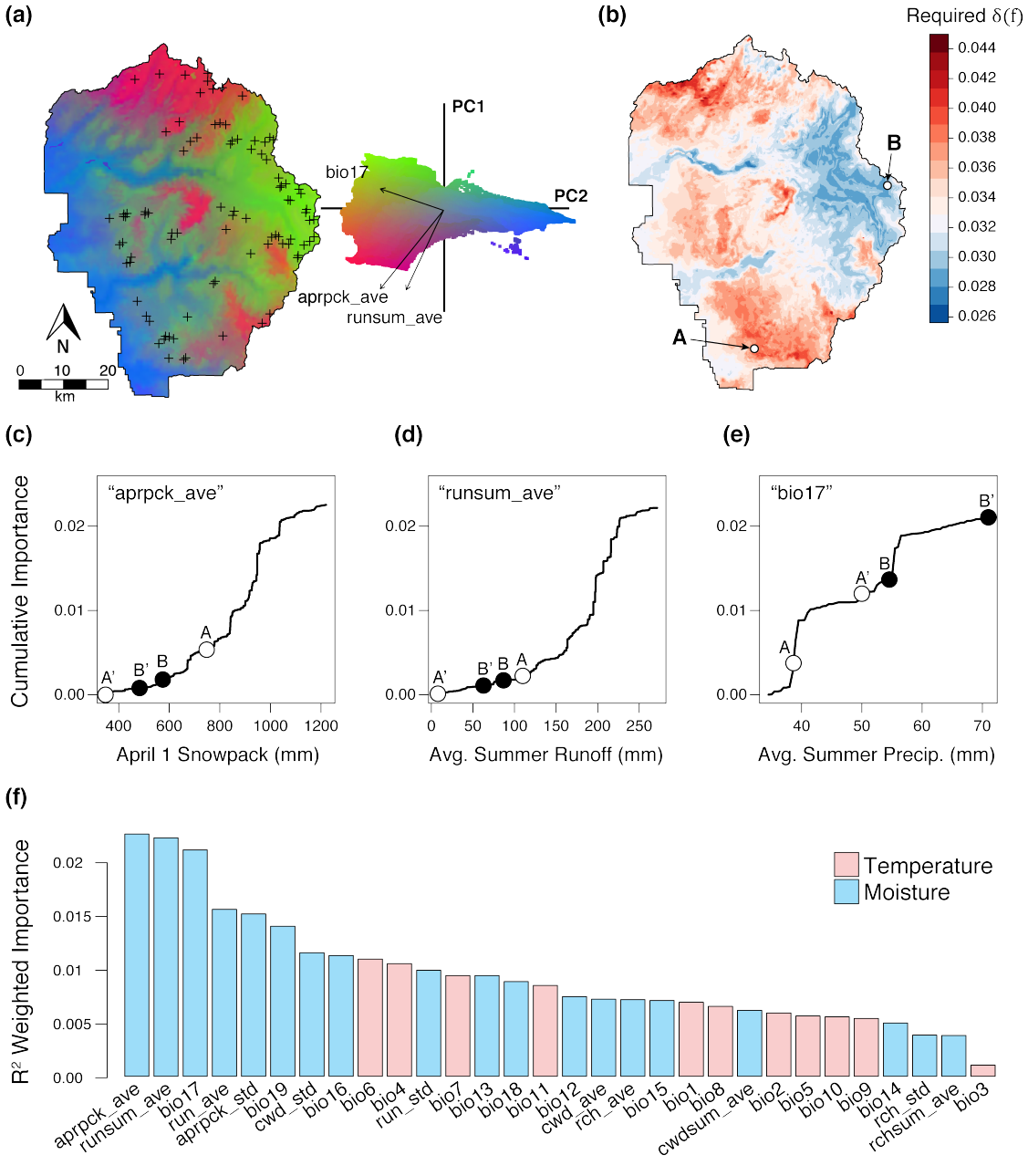

(a)
PC1
PC2
bio17
aprpck_ave
runsum_ave
N
0 10 20
km
(b)
Required δ(f)
0.044
0.042
0.040
0.038
0.036
0.034
0.032
0.030
0.028
0.026
B
A
(c)
"aprpck_ave"
Cumulative Importance
0.02
0.01
0.00
A'
B'
B
A
400 600 800 1000 1200
April 1 Snowpack (mm)
(d)
"runsum_ave"
0.02
0.01
0.00
A'
B'
B
A
0 50 100 150 200 250
Avg. Summer Runoff (mm)
(e)
"bio17"
0.02
0.01
0.00
A
A'
B
B'
40 50 60 70
Avg. Summer Precip. (mm)
(f)
R² Weighted Importance
0.02
0.015
0.01
0.005
0.00
Temperature
Moisture
aprpck_ave
runsum_ave
bio17
run_ave
aprpck_std
bio19
cwd_std
bio16
bio6
bio4
run_std
bio7
bio13
bio18
bio11
bio12
cwd_ave
rch_ave
bio15
bio1
bio8
cwdsum_ave
bio2
bio5
bio10
bio9
bio14
rch_std
rchsum_ave
bio3

Figure 4.1 Projected future selection pressure by climate change
GF model of current adaptive genetic variation among 35 outlier loci identified by bayenv, and forecasted into the years 2061–2099. (a) Regions in YOSE with similar colors represent populations expected to have similar adaptations given the underlying (non-linear) climatic gradient. Sampling locations are indicated by plus symbols. Specifically, colors represent scores along the first three PCs of multi-dimensional allele frequency. The biplot shows their locations along the first two axes, along with vectors of the three most important variables. (b) Change in optimal allele frequency $\delta(f)$ as a consequence of climate is shown: red hues = values > mean, blue hues = values < mean. Sites representative of high and low change are shown by A and B, respectively. (c–e) Cumulative importance curves (effect on allele frequency) for the three most important variables. Current climatic values for sites A and B are indicated, as well as future values (A' and B'). (f) All climatic variables, ranked by R^2 weighted importance.

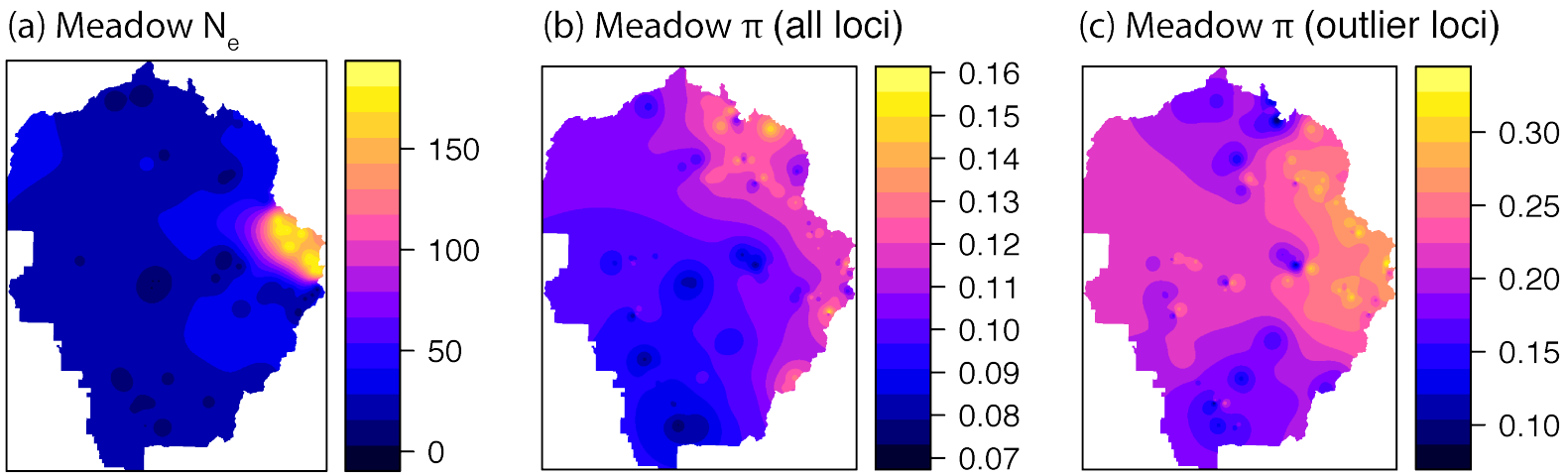


Figure 4.2 Current levels of overall and adaptive genetic diversity
Inverse-distance weighted interpolations of standing genetic diversity for Yosemite toads in YOSE. (a) N_e calculated at the level of meadow clusters (due to sample size limitations), and assigned to each meadow. (b) Meadow π at the level of SNPs, across all loci. (c) Adaptive diversity estimated as meadow π at the level of SNPs, for only the 35 outlier loci identified by bayenv. See main text and Fig. 4.1 for details.

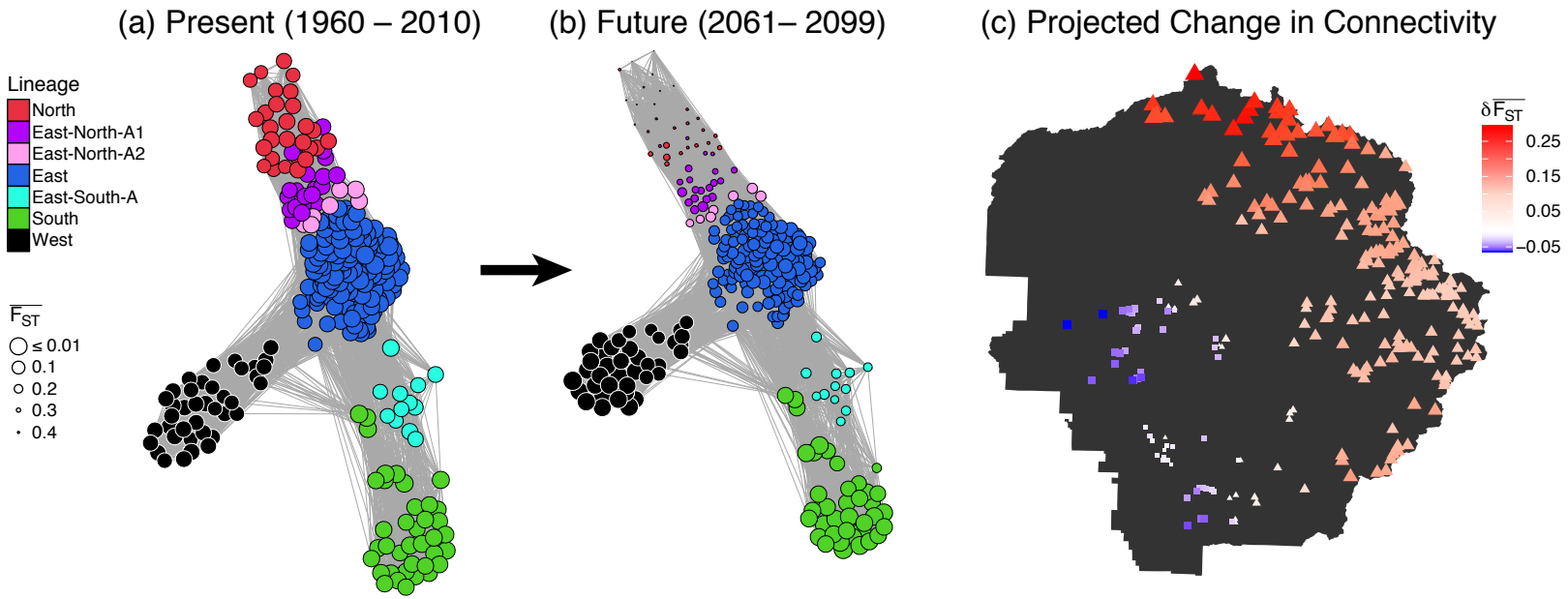


Figure 4.3 Projected future shift in meadow connectivity by climate change
A graph theoretical representation of meadow connectivity, and its forecasted change over time. Connectivity is defined here as the average "edge weight" of each node in the graph, equivalent to mean value of $1 - F_{ST}$ for each meadow. The adjacency matrix of meadows includes all links <30 km in least cost path distance, which do not veer outside the species distribution. Models based on 90 sampled meadows were forecasted to all 283 meadows in YOSE known to historically contain toads. Connectivity is shown (a) before, and (b) ~100 years after climate change, with node size indicating level of connectivity. Lineage membership is indicated by node colors; see Chapter 1 for details. (c) The net change in connectivity is shown, with a decrease indicated by larger triangles of increasingly red hue, and increase indicated by larger squares of increasingly blue hue.

CONCLUSION OF THE DISSERTATION

This dissertation has provided evidence of lineage formation in the past, present and future for Yosemite toads. In Chapter 1, I have provided evidence that pulses of glacial action bisect lineages into divergent climatic niches, which then can form further lineages by fusion. During the course of studying patterns of divergence, I discovered the intriguing process of lineage fusion in secondary contact zones. In Chapter 2, I studied contact zones between lineages further, and found that adaptive divergence and adaptive introgression may be more complementary forces than currently appreciated. In fact, my results support the view that adaptations that rapidly spread between incipient species (such as Y-East and Y-North lineages of Yosemite toads) may erode divergent adaptive variation into recombinant adaptation in a constructive way. I showed evidence that admixture in secondary contact zones is a probable source for new recombinant adaptive variation in the Yosemite toad. More specifically, I found evidence that genes underlying tadpole growth and development such as LPIN3—a metabolic regulatory gene—may be adaptively introgressing between lineages today. These two chapters (1 and 2) together provided novel insight about the past two million years of adaptive evolution in this species. Moving forward, any conservation strategy must not treat populations of Yosemite toads as adaptively homogenous, because this would gravely miscalculate the fit of these animals to their local landscape.

In Chapter 3, I introduced a novel landscape genetic workflow that improves upon existing methods for modeling connectivity in patch-limited amphibians, such as the

Yosemite toad. My method envisions two environment types being important to gene flow: breeding habitat quality, and dispersal corridors. I found that most meadows are highly genetically distinct in accord with previous studies, but approximately half (47%) of sites contained multiple meadows with source-sink structure. Hub meadows were distinguishable 90% of the time, based primarily on larger meadow size and proximity to upland habitat. There is a strong indication that climate change is already influencing meadow connectivity, because multiple climate change variables were important to the overall model. I also found evidence that wetter conditions (both in meadows, and between meadows) generally inhibit genetic connectivity. For example, low-elevation meadows have higher annual precipitation (mostly falling during the winter), and yet are less connected to each other, with lower effective population sizes. In spite of this, low-elevation meadows had larger tadpole sizes, which would indicate tadpoles are in favorable conditions. One possible interpretation is that lower-elevation toads have higher meadow fidelity at the expense of genetic diversity. My gravity analysis affirmed the impact of slope on Yosemite toad connectivity, and found that other topographic (elevation), vegetation (conifers), hydrological (precipitation and meadow wetness), and disturbance (roads and trails) variables are important too.

My last chapter identified three regions of Yosemite National Park (GEUs or conservation units) that are expected to adapt in three distinct ways to climate change: "north," "east," and "low elevation." Their vulnerability to climate change is in the following order: low elevation, north, and east. Since the northern inter-lineage admixture zone contains large amounts of neutral genetic variation, and no sign of outbreeding

depression, this region should be protected as a potentially critical interface for replenishing lost variation in the "north" conservation unit. Admixed toads might serve as important stock for reintroduction efforts to the north, but further experimental work is needed to verify their viability under realistic ecological conditions. The "low elevation" conservation unit consists of two phylogenetic lineages that are not each other's closest relatives. They are exchanging very few migrants, yet are adapted to similar climates, and experiencing similar selective pressure. Therefore, if reintroductions are deemed necessary, these lineages may be considered interchangeable. Admixing these genetically depauperate toads may have beneficial effects on their response to selection.

Overall, more is understood about the past, present, and future of this species than was previously known. Previous models of niche occupancy or breeding probability have implicitly assumed adaptive homogeneity, i.e. that these animals interact with their environment identically throughout the mountains. My dissertation overturns this idea; I have highlighted several ways natural history may be locally adapted, or have higher potential to adapt in the future. We now know that the one location that has definitely experienced long-term declines (Tioga Pass) is also the most viable and climatically stable lineage in Yosemite, perhaps in the entire species. Focusing more of our attention on this "super-toad" core area may be critical to ensure the future survival of the species. Throughout these four studies, I have tried to integrate evolution, ecology, and genetics, and to suggest the utility of emergent natural processes to conservation-minded people. There is grandeur in this heterogeneity of toad adaptation, and it will hopefully guide research and conservation for the many other specialized amphibians worldwide.

APPENDICES

APPENDIX A: SUPPLEMENTARY METHODS/RESULTS FOR CHAPTER 1

Methods: Study region

Yosemite toads are meadow-breeding specialists restricted to the central Sierra Nevada of California, and found from 1,950–3,444 m (Mullally and Cunningham 1956; Karlstrom 1962). The species breeds almost exclusively in shallow, snowmelt ponds within meadows, which seasonally dry up. Lower elevation sites are typically spring-fed mesic or hydric meadows characterized by adjacent stands of montane red fir (*Abies magnifica*) and lodgepole pine (*Pinus contorta var. murrayana*), while subalpine and alpine meadows are typically larger, snow-fed, more xeric, and surrounded by whitebark pine (*Pinus albicaulis*) or boulders (Keeler-Wolf et al. 2012; Viers et al. 2013) (Fig. 1.2). YOSE (3,027 km^2) and northern KICA (467 km^2) were chosen as the primary study areas because they overlap with previous studies (Shaffer and Fellers 2000; Wang 2012; Berlow et al. 2013), likely contain genetic discontinuities, and offer a representative (37% of all localities) snapshot of ecological conditions experienced by the species. YOSE is heavily forested in the southwest and pocked with glacially carved cirques, hanging valleys, moraines, and canyon systems at higher elevations. The Merced and Tuolumne River canyons present major biogeographical barriers along east-west axes, and numerous canyons feed into the Tuolumne from the north. KICA is separated from YOSE by approximately 70 km, and represents both the southern and elevational limit to the species distribution. The snow-free season ranges from 3–6 months depending on elevation, latitude, and year (P. Maier, pers. obs.).

Methods: Sampling Sites and Molecular methods

A total of 653 samples were chosen for sequencing (535 samples from YOSE, 109 from KICA, and nine outgroup samples). Sites were chosen to maximize representation across all known breeding locations from a recent six-year survey effort (Ostoja et al., in prep) and overlap with previous studies. Tadpoles were sampled across multiple clutches, pools, and years to maximize inclusion of available genetic diversity. A minimum of five samples was used per meadow if additional meadows were sequenced within 1 km, and 10 per meadow for remaining meadows, unless insufficient samples were available. This scheme maximized intra- and inter-meadow sampling representation across the study area. Genomic DNA was extracted using a combination of 96-well glass fiber plate (Ivanova et al. 2006) and DNeasy blood and tissue spin column (Qiagen) protocols. Extracted DNA was eluted into 10 mM Tris-Cl, pH 8.0.

Double-digest RADseq libraries were constructed following the protocol of Peterson et al. (2012; S1 Protocol). Briefly, 200–500 ng DNA was digested with 5U SbfI-HF and MspI (New England Biolabs), for 3 hr at 37°C and cleaned with 1.5x Agencourt Ampure XP beads (Beckman Coulter). Digested DNA was quantified by Qubit 2.0 Fluorometer (Life Technologies) and ligated to oligo-nucleotide adapters with one of eight unique 5 bp MID barcode sequences (Table S1.1) at 25°C for 30 min, followed by a 10 min heat kill at 65°C. Ligated DNA was cleaned with the AMPure XP and pooled by adapter, then size selected between 424 and 525 bp using a 1.5% gel cassette (Pippin Prep; Sage Science). This optimal fragment size range was chosen to balance number of loci with projected coverage per locus, by performing an in silico digestion of the

Xenopus tropicalis v4.1 genome following the method of Lemmon and Lemmon (2012). Size-selected DNA was amplified with Illumina primers containing one of 12 unique indices (Table S1.1) using a Phusion PCR kit (New England Biolabs). The following cycle profile was used: 98°C for 30 s, [98°C for 10 s, 72°C for 20 s @ 16% ramp], 72°C for 10 min, 4°C hold. Finally, amplicons were bead-cleaned, quantified by BioAnalyzer (Agilent Technologies), and pooled in equimolar amounts for sequencing. This combinatorial approach allowed 8×12 unique samples to be sequenced in parallel on a single Illumina flowcell, and use of double-restricted fragments dramatically increased locus recovery across samples. All ddRAD libraries were 2×100bp sequenced on seven lanes of an Illumina HiSeq 2500 at the Institute for Integrative Genome Biology, UC Riverside, CA.

Methods: Bioinformatic data processing

Raw data were filtered and processed using Stacks v1.19 (Catchen et al. 2011, 2013). Sequences were demultiplexed using process_radtags with a threshold of 1 nt error in barcodes. Reads with an average phred score of <10 across a sliding window of 20% sequence length were discarded. Paired-end (PE) reads were non-overlapping by design, meaning that both ends of each DNA fragment were sequenced, but the intervening DNA (of variable length) was not sequenced. Therefore, two datasets were subsequently possible: (1) "R1" (only Read 1 was retained) and (2) "concatenated" (PE sequence reads were concatenated). Since concatenated loci are longer sequences, they contain more SNPs on average, and hence more haplotype alleles per RAD locus. By

processing these separately, I was able to build a dataset of RAD haplotypes with maximum length per locus (concatenated loci wherever available, R1 loci if the R2s were lost due to quality or coverage thresholds). Duplicate loci (R1s that are already represented by concatenated R1 + R2) were identified and removed using a standard blast-n search. In this way I built a combined R1/concatenated haplotype dataset to use with analyses that can leverage haplotypic information. "Concatenated" reads are less ideal than contigs because the reads are non-overlapping, and hence the distance between SNPs on different reads is unknown. However they still include information about allelic phase, and thus can increase the number of unique haplotypes per locus. For all other analyses, one SNP per locus was randomly picked. Stacks only considers biallelic SNPs, so only these were included in the output.

ustacks was then used to identify alleles ("stacks") and subsequently call SNPs using a multinomial likelihood algorithm (Hohenlohe et al. 2010). R1 loci (96 nt) were given a maximum stack distance of 3 nt both within and between individuals, and this value was doubled for concatenated (197 nt) loci. I assumed that sequence divergence higher than this threshold was indicative of alleles from different loci, or from paralogs. Each stack had a minimum coverage of three reads. Secondary stacks with 1–2 sequencing errors were retained to increase power for SNP likelihood ratio tests. Loci with more than three stacks were either removed or deleveraged into multiple loci using a minimum spanning tree algorithm; this allowed paralogs and highly repetitive sequences to be either separated or discarded. Catalogs of consensus loci were constructed by

cstacks using a subset (270/653 for R1, 126/653 for concatenated) of individuals due to limited computing power.

The program sstacks was used to match ustacks loci against the cstacks catalog to call genotypes. After applying a locus genotype coverage threshold of 10, loci were removed from meadows if absent from >25% individuals, and removed from the dataset if absent from >25% individuals overall. Only alleles with a minor allele frequency (MAF) of 0.05 or greater were kept. As described above, two final datasets were generated: (1) SNPs (one SNP kept per locus), and (2) haplotypes (concatenated paired-end reads were used where possible, otherwise R1 was used). I built and implemented a custom python script (fasta2genotype.py, github.com/paulmaier/fasta2genotype) to output a dataset containing unique haplotype integers for each locus. All processing was performed on a high-performance biocluster at the Institute for Integrative Genome Biology, UC Riverside, CA.

Results: Illumina data quality and quantity

In total, 1.88 billion reads were returned from seven lanes of an Illumina HiSeq 2500 run, of which 1.60 billion were identifiably barcoded. This included 161.68 gigabases (Gb) of useable sequence data with a mean quality score of 35.57. A total of 535 individuals from 90 meadows were sequenced at 2318 loci for YOSE, and 109 individuals from 12 meadows were sequenced at 1914 loci for KICA (Table 1.1). Concatenating the paired-end reads greatly reduced the number of loci passing coverage and quality thresholds, as only 1044 and 1341 concatenated loci were retained for YOSE

and KICA, respectively. Similarly, coverage was lower for concatenated data (average of 55.15 and 49.69) versus R1 data (average of 95.58 and 79.80). Despite this reduction in data quantity, concatenating loci increased the average number of haplotypes that each contained, thus potentially increasing the genetic diversity contained within any particular locus. However, SNPs across approximately 200 bp of sequence are in tight linkage, and hence only one SNP should be used per locus. Therefore, using haplotypic data effectively increased the number of alleles per locus in all cases from two SNP alleles to an average of 2.41–3.95 haplotypic alleles (depending on dataset), and a maximum of 20 haplotype alleles (Table 1.1).

Results: Patterns of genetic diversity

Genetic diversity showed a dramatic increase from lower (western/southern YOSE) to higher elevation (eastern/northern YOSE, KICA) meadows (Tables 1.2, S1.2). Meadow elevation and observed heterozygosity showed a strong positive correlation, (ρ=0.68, df=100, p<0.001), as did meadow elevation and mean gene diversity (π) (ρ=0.61, df=100, p<0.001). Elevation is a covariate of many ecological patterns and processes, including a transition from montane red fir meadow habitats to subalpine meadow habitats, so this observation by itself is not explanatory. The number of private alleles was also positively correlated with elevation, (ρ=0.27, df=100 p<0.01), although this pattern was not significant when excluding KICA, (p=0.28). KICA meadows not only had more private alleles than YOSE (t=-3.50, df=12.11, p<0.01), but also higher observed heterozygosity (t=-3.62, df=12.67, p<0.01), π (t=-3.48, df=12.72, p<0.01), and

average number of polymorphic sites, (t=-2.85, df=12.48, p<0.05), in spite of having far fewer SNPs (t=6.75, df=34.51, p<0.001). Overall, KICA showed a trend of higher genetic diversity than YOSE. This is possibly because the sites in KICA have a higher mean (t=-5.64, df=16.57, p<0.001), and lower (albeit insignificant) variance in elevation (F=1.73, df=11, p=0.16).

The Goddard lineage along the S. San Joaquin River in KICA had the highest genetic diversity of any lineage in either park (Table 1.2). Compared to the second-most diverse lineage, Goddard has an average meadow H_O of 0.068 (8% > East-North-A1, lineage defined below), π of 0.067 (10% > East-North-A1), and 0.44 of SNPs are polymorphic (10% > East-South-A). This was expected given more adult toads and tadpoles are seen in Goddard canyon than anywhere else (P. Maier, pers. obs.), but surprising given this is the latitudinal and elevational range limit for the species. Its sister lineage Evolution, in contrast, has third lowest H_E, H_O, π, % polymorphic SNPs, and the lowest N_e (3.3). In every metric, admixed lineages (discussed below) were found to have higher genetic diversity than pure lineages (e.g. H_E, H_O, π, P, etc.).

APPENDIX B: SUPPLEMENTARY TABLES/FIGURES FOR CHAPTER 1

Table S1.1. Oligonucleotides used for ddRAD library prep. Libraries were produced following the protocol of Peterson et al. (2012).

Name	Sequence
GCATG_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATGTGCA
AACCA_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCATGCA
CGATC_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCTGCA
TCGAT_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGATTGCA
TGCAT_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCATTGCA
CAACC_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACCTGCA
GGTTG_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTGTGCA
AAGGA_SbfI_A1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGATGCA
GCATG_SbfI_A1.2	/5Phos/CATGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AACCA_SbfI_A1.2	/5Phos/TGGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CGATC_SbfI_A1.2	/5Phos/GATCGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TCGAT_SbfI_A1.2	/5Phos/ATCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
TGCAT_SbfI_A1.2	/5Phos/ATGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
CAACC_SbfI_A1.2	/5Phos/GGTTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
GGTTG_SbfI_A1.2	/5Phos/CAACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
AAGGA_SbfI_A1.2	/5Phos/TCCTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
MspI_A2.1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MspI_A2.2	/5Phos/CGAGATCGGAAGAGCGAGAACAA
PCR1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG
PCR2_Idx_1_ATCACG	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_2_CGATGT	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_3_TTAGGC	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_4_TGACCA	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_5_ACAGTG	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_6_GCCAAT	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_7_CAGATC	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_8_ACTTGA	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_9_GATCAG	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_10_TAGCTT	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_11_GGCTAC	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC
PCR2_Idx_12_CTTGTA	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC

Table S1.2. Summary of population genetic parameters for each meadow. YOSE = Yosemite NP, KICA = Kings Canyon NP, PA = private alleles, N = effective sample size, P = frequency of most frequent allele, H_o = observed heterozygosity, H_e = expected heterozygosity, π = average gene diversity, F_{IS} = fixation index, % Pol Loci = % of sites that are locally polymorphic, N_e = effective population size (estimated at each scale using the linkage disequilibrium method, dashes indicate insufficient power for an estimate). See Chapter 3 for definition of meadow "clusters."

Lineage	Park	Type	Cluster	Meadow	PA	N	P	H_O	H_E	π	F_{IS}	% Pol Loci	Meadow N_e	Cluster N_e	Lineage N_e
North	YOSE	Pure	Rancheria	3537	10	10	0.965	0.051	0.047	0.049	-0.003	0.318	31.7	31.7	14.3
			Thompson	3613	13	10	0.966	0.053	0.045	0.048	-0.011	0.291	9.3	9.3	
			Kerrick	4136	50	9	0.954	0.066	0.063	0.067	0.005	0.452	22.4	22	
				4146	14	5	0.958	0.065	0.055	0.062	-0.004	0.340	-		
				4164	2	5	0.964	0.065	0.046	0.052	-0.026	0.267	-		
				4324	45	10	0.954	0.066	0.063	0.067	0.005	0.454	24.4		
				4391	10	5	0.961	0.067	0.050	0.057	-0.019	0.302	-		
			Wells	4365	23	10	0.958	0.061	0.056	0.059	0.001	0.379	17.3	17.3	
			Tilden	4370	26	10	0.960	0.058	0.053	0.055	-0.005	0.347	8.2	8.2	
			Twin Lakes	4317	36	10	0.967	0.056	0.044	0.047	-0.018	0.288	11	11	
East-North-A1	YOSE	Admixed	Slide	3612	2	4	0.965	0.058	0.046	0.053	-0.009	0.276	-	27.7	28.4
				3615	19	5	0.956	0.064	0.058	0.065	0.004	0.375	-		
				3763	7	7	0.957	0.065	0.058	0.063	-0.003	0.385	11.7		
				4025	30	10	0.950	0.070	0.067	0.071	0.004	0.475	22.1		
			Rodgers	3384	1	2	0.963	0.063	0.045	0.061	-0.003	0.243	-	19	
				3273	3	9	0.963	0.058	0.050	0.053	-0.011	0.326	18.2		
				3452	21	9	0.958	0.062	0.057	0.060	-0.002	0.399	40.4		
East-North-A2	YOSE	Admixed	Miller	3342	3	4	0.966	0.057	0.043	0.050	-0.014	0.235	-	9.2	9.8
				3400	7	5	0.959	0.065	0.054	0.060	-0.009	0.336	-		
				3414	12	5	0.957	0.062	0.058	0.065	0.007	0.377	-		
East	YOSE	Pure	Ireland	1856	15	5	0.964	0.059	0.047	0.052	-0.013	0.289	-	19.3	74.9
				1951	2	3	0.964	0.065	0.046	0.055	-0.017	0.257	-		
				1960	5	5	0.966	0.056	0.045	0.050	-0.011	0.278	-		
				2021	4	5	0.960	0.061	0.053	0.059	-0.002	0.335	-		
				2026	9	5	0.959	0.064	0.055	0.062	-0.004	0.357	-		

Lineage	Park	Type	Cluster	Meadow	PA	N	P	H_O	H_E	π	F_{IS}	% Pol Loci	Meadow N_e	Cluster N_e	Lineage N_e
East	YOSE	Pure	Ireland	2059	0	5	0.963	0.065	0.048	0.053	-0.022	0.288	-	19.3	74.9
				2104	0	5	0.964	0.063	0.047	0.053	-0.021	0.283	-		
				2135	2	5	0.965	0.062	0.044	0.049	-0.025	0.252	-		
			Lyell	1782	2	3	0.963	0.068	0.045	0.054	-0.024	0.246	-	11.9	
				1815	7	2	0.955	0.070	0.056	0.076	0.008	0.306	-		
				1956	13	5	0.959	0.062	0.054	0.061	-0.001	0.343	-		
				1997	8	5	0.961	0.059	0.052	0.058	-0.001	0.326	-		
				2039	2	5	0.965	0.062	0.044	0.049	-0.025	0.248	-		
			Tresidder	2147	2	10	0.962	0.061	0.051	0.054	-0.015	0.338	11.4	11.4	
			Cockscomb	2324	6	10	0.963	0.055	0.049	0.051	-0.007	0.317	12	12	
			Cathedral	2351	2	4	0.972	0.049	0.034	0.040	-0.017	0.188	-	3	
				2410	1	5	0.969	0.055	0.039	0.043	-0.021	0.219	-		
			Polly	2498	3	9	0.970	0.053	0.039	0.042	-0.024	0.224	-	-	
			Tioga	2256	4	2	0.967	0.051	0.040	0.054	0.004	0.213	-	181.2	
				2312	19	7	0.958	0.059	0.057	0.061	0.006	0.385	-		
				2384	6	5	0.963	0.058	0.049	0.055	-0.006	0.307	-		
				2407	6	6	0.957	0.060	0.059	0.065	0.014	0.403	-		
				2429	10	5	0.962	0.057	0.050	0.056	-0.001	0.317	-		
				2594	3	5	0.960	0.063	0.053	0.060	-0.007	0.331	-		
				2618	5	4	0.962	0.055	0.049	0.056	0.004	0.297	-		
				2824	10	3	0.960	0.064	0.052	0.062	-0.003	0.303	-		
				2836	2	2	0.967	0.056	0.040	0.054	-0.004	0.215	-		
				2864	30	12	0.958	0.058	0.056	0.059	0.006	0.412	-		
				2893	13	5	0.959	0.063	0.054	0.061	-0.003	0.349	-		
				2977	3	5	0.962	0.060	0.050	0.056	-0.007	0.316	-		
			Conness	3200	17	5	0.957	0.062	0.057	0.063	0.004	0.363	-	29.8	
				3225	10	5	0.960	0.061	0.053	0.059	-0.002	0.339	-		
				3272	7	5	0.958	0.066	0.055	0.062	-0.007	0.339	-		
				3339	0	4	0.963	0.065	0.048	0.055	-0.018	0.285	-		
				3371	12	5	0.959	0.060	0.054	0.060	0.002	0.332	82.4		

Lineage	Park	Type	Cluster	Meadow	PA	N	P	H_O	H_E	π	F_{IS}	% Pol Loci	Meadow N_e	Cluster N_e	Lineage N_e
East	YOSE	Pure	Conness	3420	6	5	0.967	0.054	0.043	0.049	-0.010	0.267	-	29.8	74.9
				3424	6	5	0.962	0.062	0.050	0.056	-0.012	0.303	17.8		
East-South-A	YOSE	Admixed	Isberg	1097	6	11	0.956	0.062	0.059	0.062	0.002	0.401	18.2	17.3	17.3
				942	7	10	0.958	0.063	0.057	0.060	-0.004	0.394	14.4		
South	YOSE	Pure	Bridalveil	1040	6	5	0.969	0.050	0.040	0.045	-0.009	0.244	-	13.2	25.9
				1070	3	2	0.968	0.054	0.038	0.052	-0.003	0.203	-		
				1171	16	5	0.967	0.050	0.043	0.048	-0.004	0.270	-		
				359	8	5	0.973	0.047	0.036	0.040	-0.014	0.217	-		
				733	11	10	0.965	0.053	0.047	0.050	-0.007	0.316	5.9		
			Chilnualna	377	3	5	0.971	0.051	0.037	0.042	-0.018	0.225	-	15.2	
				638	1	5	0.969	0.048	0.040	0.045	-0.006	0.251	-		
				705	5	5	0.968	0.051	0.041	0.046	-0.011	0.244	27.3		
				719	2	5	0.970	0.054	0.038	0.042	-0.023	0.211	-		
				758	1	5	0.969	0.052	0.041	0.046	-0.012	0.247	-		
				780	4	5	0.970	0.053	0.039	0.044	-0.017	0.240	-		
			Wawona	387	5	10	0.973	0.045	0.036	0.038	-0.014	0.232	15.7	15.7	
			Summit	1369	5	10	0.971	0.053	0.037	0.039	-0.028	0.217	-	-	
			Starr King	1543	10	6	0.968	0.052	0.042	0.046	-0.011	0.268	-	29.7	
				1569	1	5	0.967	0.051	0.042	0.048	-0.006	0.256	-		
West	YOSE	Pure	Bald	2369	4	10	0.968	0.048	0.043	0.045	-0.006	0.272	12	12	36.3
			Ribbon 1	1779	6	5	0.965	0.049	0.046	0.052	0.006	0.284	-	19.3	
				1841	3	5	0.964	0.056	0.047	0.053	-0.004	0.294	-		
				2012	11	5	0.965	0.053	0.046	0.051	-0.004	0.289	-		
				2030	11	5	0.967	0.051	0.043	0.048	-0.005	0.257	8.2		
			Ribbon 2	1756	1	5	0.972	0.049	0.035	0.040	-0.019	0.194	-	-	
			Porcupine	2073	5	5	0.971	0.051	0.038	0.042	-0.016	0.227	-	2.8	
				2118	3	5	0.971	0.050	0.036	0.040	-0.020	0.192	-		
				2132	1	5	0.971	0.053	0.036	0.041	-0.022	0.205	-		
			White Wolf	2371	9	5	0.965	0.054	0.047	0.053	-0.002	0.297	-	22.8	
				2385	7	5	0.965	0.053	0.046	0.051	-0.003	0.286	-		

Lineage	Park	Type	Cluster	Meadow	PA	N	P	H_O	H_E	π	F_{IS}	% Pol Loci	Meadow N_e	Cluster N_e	Lineage N_e
West	YOSE	Pure	White Wolf	2391	8	5	0.966	0.054	0.044	0.049	-0.008	0.270	-	22.8	36.3
				2411	19	5	0.969	0.049	0.041	0.046	-0.005	0.254	7		
				2418	1	5	0.967	0.054	0.043	0.048	-0.012	0.260	-		
				2421	2	5	0.966	0.055	0.045	0.050	-0.008	0.277	-		
				2443	4	5	0.968	0.058	0.041	0.046	-0.022	0.237	-		
Evolution	KICA	Pure	Sapphire	958	24	10	0.959	0.065	0.055	0.058	-0.014	0.375	8.2	3.3	3.3
				902	32	4	0.972	0.046	0.035	0.041	-0.009	0.207	-		
Goddard	KICA	Pure	Tehipite	101	16	6	0.961	0.061	0.052	0.057	-0.006	0.342	20.6	20.6	24.3
			Emerald	916	47	21	0.957	0.062	0.059	0.060	-0.002	0.432	4.2	4.2	
			Finger	1381	34	6	0.962	0.061	0.049	0.054	-0.015	0.285	60.2	60.2	
			Martha	1131	10	7	0.950	0.071	0.067	0.073	0.006	0.470	-	53.7	
				1138	26	9	0.949	0.072	0.069	0.073	0.005	0.505	160.7		
				1141	15	7	0.952	0.070	0.065	0.070	0.002	0.444	47.9		
				1165	12	6	0.957	0.068	0.057	0.062	-0.010	0.386	-		
				1179	12	10	0.950	0.071	0.068	0.072	0.004	0.495	19.9		
				24	56	10	0.947	0.076	0.073	0.077	0.006	0.541	152		
			Peak 12,434	1071	9	9	0.950	0.072	0.068	0.072	0.003	0.485	9.3	9.3	
	All				**10.471**	**6.08**	**0.963**	**0.059**	**0.049**	**0.054**	**-0.007**	**0.309**	**29.125**	**23.332**	**25.767**
	YOSE				**8.611**	**5.74**	**0.964**	**0.058**	**0.047**	**0.053**	**-0.008**	**0.295**	**19.522**	**22.880**	**31.750**
	KICA				**24.417**	**8.70**	**0.955**	**0.066**	**0.060**	**0.064**	**-0.002**	**0.414**	**53.667**	**25.217**	**13.800**
		Pure			**10.556**	**6.01**	**0.963**	**0.058**	**0.048**	**0.054**	**-0.008**	**0.303**	**31.038**	**24.078**	**29.833**
		Admixed			**9.833**	**6.62**	**0.959**	**0.063**	**0.054**	**0.060**	**-0.003**	**0.352**	**20.833**	**18.300**	**18.500**

Table S1.3. Divergence date estimates for major lineages and corresponding ice ages. Dates were estimated by a BEAST ultrametric tree, with clock rate prior set to confidence intervals of best available amphibian nuclear genetic clock. Clock prior was set to a normal distribution with a range of $9.24 \times 10^{-10} – 1.53 \times 10^{-9}$ substitutions/site/year, following Crawford (2003). Ages are in thousands of years. Asterisks (*) indicate estimates of a crown group. Estimated age uses median node height, min and max are based on 95% HPD of node height. Names and dates of closest ice ages are listed, with uncertainty values shown (based on Gillespie and Clark 2011). Where multiple glacial events are plausible, they are all listed.

Lineage 1	Lineage 2	Max. Age	Min. Age	Med. Age	Prop. Oldest	Nearest Ice Age	Date
Kings Canyon NP	Yosemite NP	2118	1881	1993	1.00	McGee	2700–1500
North	(East + West + South)	777	687	732	0.37	Sherwin	~820
South	(East + West)	455	397	425	0.21	Walker Creek	~550
West	East	319	274	297	0.15	Bloody Canyon	220–140
North	*	263	215	239	0.12	Bloody Canyon	220–140
East	*	250	207	229	0.11	Bloody Canyon	220–140
West	*	155	122	138	0.07	Bloody Canyon Tahoe I Casa Diablo	220–140 170–130 126–62
South	*	122	93	107	0.05	Casa Diablo	126–62
Evolution	Goddard	239	191	214	0.11	Bloody Canyon	220–140
Goddard	*	69	42	55	0.03	Mono Basin Tahoe II	80–60 50–42
Evolution	*	27	10	18	0.01	Tioga	21–20

Table S1.4. Comparison of NewHybrids and HIest results. Assignment of admixed individuals to specific hybrid genotype classes for the two contact zones examined. In these two contact zones (East-North, East-South), only individuals from the three admixed lineages (East-North-A1, East-North-A2, East-South-A) are shown. NH.class = hybrid genotype class estimated by NewHybrids, HI.class = hybrid genotype class estimated by HIest, Advanced? = true/false, true if HIest was able to reject the discrete 2-generation classification scheme in favor of the continuous hybrid scheme (although advanced genotype classes such as "F3" cannot be predicted), dAIC = the difference in AIC score between the continuous model log-likelihood and the discrete model log-likelihood (a positive score means the discrete model was rejected in favor of the continuous/advanced classification), S = admixture index (0–1, larger meaning more East ancestry), H = inter-lineage heterozygosity.

Cont.Zone	Lineage	Meadow	Individual	NH.class	HI.class	Advanced?	dAIC	S	H
East-North	East-North-A1	3273	YOSE12-0655	North	North-BC	Advanced	21.834	0.210	0.100
East-North	East-North-A1	3273	YOSE12-0657	North	F2	Advanced	56.829	0.210	0.000
East-North	East-North-A1	3273	YOSE12-0658	North	F2	-	-882.282	0.222	0.111
East-North	East-North-A1	3273	YOSE12-0659	North	North-BC	-	-901.337	0.185	0.070
East-North	East-North-A1	3273	YOSE12-0660	North	F2	Advanced	45.345	0.170	0.220
East-North	East-North-A1	3273	YOSE12-0661	North	North-BC	Advanced	9.997	0.260	0.140
East-North	East-North-A1	3273	YOSE12-0663	North	North-BC	-	-123441.605	1.000	0.111
East-North	East-North-A1	3273	YOSE12-0671	North	North-BC	Advanced	26.638	0.140	0.000
East-North	East-North-A1	3273	YOSE12-0679	North	North-BC	-	-123430.105	1.000	0.111
East-North	East-North-A1	3273	YOSE12-0680	North	North-BC	-	-907.490	0.222	0.111
East-North	East-North-A1	3384	YOSE13-0529	North	North-BC	-	-921.213	0.160	0.320
East-North	East-North-A1	3384	YOSE13-0533	North	North-BC	Advanced	5.928	0.180	0.160
East-North	East-North-A1	3452	YOSE12-0549	North	F2	-	-886.028	0.220	0.100
East-North	East-North-A1	3452	YOSE12-0558	North	North-BC	Advanced	5.084	0.165	0.130
East-North	East-North-A1	3452	YOSE12-0565	North	North-BC	Advanced	6.390	0.150	0.120
East-North	East-North-A1	3452	YOSE12-0567	North	F2	-	-877.379	0.175	0.110
East-North	East-North-A1	3452	YOSE12-0572	North	F2	-	-894.896	0.260	0.120
East-North	East-North-A1	3452	YOSE12-0579	North	F2	-	-880.649	0.160	0.000
East-North	East-North-A1	3452	YOSE13-0531	North	F2	-	-888.438	0.215	0.170
East-North	East-North-A1	3452	YOSE13-0535	North	North-BC	-	-908.570	0.235	0.110

Cont.Zone	Lineage	Meadow	Individual	NH.class	HI.class	Advanced?	dAIC	S	H
East-North	East-North-A1	3452	YOSE13-0539	North	F2	Advanced	26.120	0.222	0.222
East-North	East-North-A1	3452	YOSE13-0558	North	North-BC	Advanced	10.409	0.165	0.170
East-North	East-North-A1	3612	YOSE12-0496	North	F2	-	-881.946	0.180	0.020
East-North	East-North-A1	3612	YOSE12-0517	North	F2	-	-878.965	0.185	0.050
East-North	East-North-A1	3612	YOSE12-0533	North	F2	-	-881.446	0.230	0.000
East-North	East-North-A1	3612	YOSE12-0564	North	F2	-	-889.407	0.220	0.140
East-North	East-North-A1	3615	YOSE13-0742	North	North-BC	Advanced	13.967	0.245	0.190
East-North	East-North-A1	3615	YOSE13-0745	North	F2	Advanced	26.136	0.215	0.270
East-North	East-North-A1	3615	YOSE13-0756	North-BC	North-BC	-	-922.714	0.222	0.333
East-North	East-North-A1	3615	YOSE13-0761	North-BC	F2	-	-1801.876	0.165	0.190
East-North	East-North-A1	3615	YOSE13-0764	North	North-BC	-	-2740.551	0.240	0.000
East-North	East-North-A1	3763	YOSE13-0320	North-BC	F2	-	-902.176	0.225	0.230
East-North	East-North-A1	3763	YOSE13-0321	North	North-BC	-	-924.635	0.310	0.380
East-North	East-North-A1	3763	YOSE13-0736	North	North-BC	-	-1.882	0.230	0.460
East-North	East-North-A1	3763	YOSE13-0737	North	North-BC	Advanced	7.791	0.222	0.222
East-North	East-North-A1	3763	YOSE13-0739	North	F2	Advanced	38.772	0.185	0.110
East-North	East-North-A1	3763	YOSE13-0740	North	F2	Advanced	20.131	0.260	0.200
East-North	East-North-A1	3763	YOSE13-0791	North	North-BC	-	-919.147	0.220	0.320
East-North	East-North-A1	4025	YOSE12-1151	North-BC	F2	-	-903.378	0.250	0.280
East-North	East-North-A1	4025	YOSE12-1170	F2	F2	-	-902.527	0.275	0.150
East-North	East-North-A1	4025	YOSE12-1184	F2	F2	-	-905.235	0.275	0.270
East-North	East-North-A1	4025	YOSE12-1205	North-BC	F2	-	-909.189	0.333	0.222
East-North	East-North-A1	4025	YOSE12-1214	North	North-BC	Advanced	3.306	0.285	0.390
East-North	East-North-A1	4025	YOSE13-0174	North-BC	F2	-	-924.389	0.445	0.530
East-North	East-North-A1	4025	YOSE13-0309	North-BC	North-BC	-	-0.910	0.310	0.520
East-North	East-North-A1	4025	YOSE13-0322	North	North-BC	Advanced	1.106	0.222	0.333
East-North	East-North-A1	4025	YOSE13-0704	North	F2	-	-900.826	0.222	0.222

Cont.Zone	Lineage	Meadow	Individual	NH.class	HI.class	Advanced?	dAIC	S	H
East-North	East-North-A1	4025	YOSE13-0728	North-BC	F2	-	-919.213	0.370	0.560
East-North	East-North-A1	4025	YOSE13-0752	North-BC	F2	Advanced	5.965	0.365	0.490
East-North	East-North-A2	3342	YOSE12-0480	F2	F2	Advanced	13.674	0.444	0.111
East-North	East-North-A2	3342	YOSE12-0508	F2	F2	Advanced	14.886	0.444	0.111
East-North	East-North-A2	3342	YOSE12-0509	F2	F2	Advanced	11.217	0.495	0.170
East-North	East-North-A2	3342	YOSE12-0539	F2	F2	Advanced	17.233	0.425	0.070
East-North	East-North-A2	3400	YOSE12-0059	F2	F2	Advanced	5.598	0.665	0.410
East-North	East-North-A2	3400	YOSE12-0107	F2	F2	Advanced	10.028	0.667	0.333
East-North	East-North-A2	3400	YOSE12-0229	F2	F2	Advanced	4.946	0.615	0.330
East-North	East-North-A2	3400	YOSE12-0339	F2	F2	Advanced	4.831	0.630	0.240
East-North	East-North-A2	3400	YOSE12-0351	F2	F2	Advanced	6.028	0.650	0.420
East-North	East-North-A2	3414	YOSE12-0478	East-BC	F2	Advanced	31.380	0.760	0.160
East-North	East-North-A2	3414	YOSE13-0758	F2	F2	-	-918.918	0.640	0.380
East-North	East-North-A2	3414	YOSE13-0766	East-BC	F2	Advanced	12.125	0.667	0.222
East-North	East-North-A2	3414	YOSE13-0774	F2	East-BC	Advanced	4.764	0.705	0.430
East-North	East-North-A2	3414	YOSE13-0779	F2	F2	Advanced	20.534	0.655	0.130
East-South	East-South-A	942	YOSE12-0163	F2	F2	Advanced	9.944	0.550	0.000
East-South	East-South-A	942	YOSE12-0617	F2	F2	-	-44224.538	1.000	0.053
East-South	East-South-A	942	YOSE12-0788	F2	F2	-	-914.373	0.520	0.000
East-South	East-South-A	942	YOSE12-0825	F2	F2	-	-1.930	0.505	0.430
East-South	East-South-A	942	YOSE12-0831	F2	F2	Advanced	4.621	0.579	0.053
East-South	East-South-A	942	YOSE12-0836	F2	F2	-	-918.166	0.670	0.140
East-South	East-South-A	942	YOSE12-0945	F2	East-BC	-	-923.647	0.632	0.421
East-South	East-South-A	942	YOSE12-0975	F2	F2	Advanced	2.886	0.565	0.170
East-South	East-South-A	942	YOSE12-0985	F2	F2	-	-922.080	0.510	0.160
East-South	East-South-A	942	YOSE12-0996	F2	F2	Advanced	3.962	0.632	0.158
East-South	East-South-A	1097	YOSE12-0736	F2	F2	-	-911.601	0.620	0.000

Cont.Zone	Lineage	Meadow	Individual	NH.class	HI.class	Advanced?	dAIC	S	H
East-South	East-South-A	1097	YOSE12-0837	F2	F2	-	-919.897	0.421	0.053
East-South	East-South-A	1097	YOSE12-0859	F2	F2	-	-1845.556	0.474	0.158
East-South	East-South-A	1097	YOSE12-0995	F2	F2	-	-1834.862	0.570	0.000
East-South	East-South-A	1097	YOSE12-0998	F2	East-BC	-	-44235.003	1.000	0.053
East-South	East-South-A	1097	YOSE13-0363	F2	F2	-	-43744.106	1.000	0.053
East-South	East-South-A	1097	YOSE13-0374	F2	F2	-	-43741.470	1.000	0.053
East-South	East-South-A	1097	YOSE13-0403	F2	F2	-	-925.378	0.535	0.390
East-South	East-South-A	1097	YOSE13-0474	F2	F2	-	-43754.186	1.000	0.053
East-South	East-South-A	1097	YOSE13-0480	F2	F2	-	-44231.480	1.000	0.053
East-South	East-South-A	1097	YOSE13-0697	F2	F2	Advanced	1.007	0.474	0.211

Table S1.5. Spatial expansion, Tajima's D, and Fu's F_S results. The null hypothesis for the expansion test is actually the hypothesis of interest, namely spatial population expansion, so significant ($p<0.05$) values reject the expansion model. Tajima's D and Fu's F_S compare two estimators of θ (segregating sites or alleles, vs. pairwise differences) to infer population expansions (negative values) or contractions (positive values). Significance levels for Tajima's D and Fu's F_S are 0.05 and 0.02, respectively, indicating that stable population size over time can be rejected.

Statistic	North	East-North-A1	East-North-A2	East	East-South-A	South	West	Evolution	Goddard
τ	38.76 (30.15-44.29)	30.81 (23.89-59.36)	82.71 (66.62-128.20)	5.39 (4.57-7.21)	78.25 (57.54-167.92)	29.16 (22.94-35.26)	58.67 (49.47-56.56)	71.50 (56.22-121.27)	21.23 (16.41-27.37)
θ_0	7.85 (0.01-21.49)	11.46 (0.00-19.33)	38.57 (0.01-64.75)	0.59 (0.00-1.40)	40.05 (0.01-51.28)	8.26 (0.01-16.94)	0.76 (0.00-3.23)	38.59 (0.02-57.31)	7.38 (0.67-20.43)
M	8,829.31 (187.06-9,849.74)	4,356.07 (119.96-10,624.43)	9,049.69 (182.63-6,045.91)	8,220.10 (198.30-6,603.85)	8,884.29 (148.00-8,978.08)	4,428.43 (162.29-9,212.64)	2,877.80 (901.21-7,263.95)	8,814.07 (128.43-4,626.06)	8,722.01 (106.23-10,177.01)
SSD	8.20E-04	1.56E-03	4.12E-03	2.01E-03	5.01E-03	1.73E-03	1.62E-02	4.55E-03	1.16E-03
SSD p-value	0.23	0.37	0.51	0.00	0.38	0.15	0.00	0.58	0.37
Raggedness index	4.00E-04	6.40E-04	4.87E-03	1.41E-02	2.35E-03	4.00E-04	3.10E-04	4.31E-03	5.70E-04
Raggedness p-value	0.97	0.79	0.70	0.00	0.67	0.98	1.00	0.86	0.98
Sample size	150.00	116.00	28.00	388.00	42.00	176.00	170.00	28.00	190.00
S	415.00	419.00	337.00	421.00	470.00	349.00	392.00	253.00	417.00
π	105.26	106.76	105.02	79.53	147.45	87.34	107.34	86.92	109.22
Tajima's D	1.36	1.19	0.84	0.72	1.30	1.41	1.82	1.32	1.69
Tajima's D p-value	0.94	0.90	0.85	0.83	0.94	0.93	0.97	0.93	0.96
No. of alleles (unchecked)	150.00	116.00	28.00	388.00	42.00	176.00	170.00	28.00	190.00
θ_π	105.26	106.76	105.02	79.53	147.45	87.34	107.34	86.92	109.22
Exp. no. of alleles	93.54	78.78	24.93	141.29	37.07	96.73	102.20	24.39	110.39
FS	-24.85	-25.12	-3.28	-23.97	-5.35	-24.50	-24.82	-3.93	-24.78
FS p-value	0.00	0.00	0.05	0.00	0.03	0.00	0.00	0.04	0.00

* Parameter abbreviations are as follows: $\tau = 2T\mu$ (T = no. of generations since expansion, μ = mutation rate), $\theta_0 = 2N_0\mu$ (N_0 = initial population size), $M = 2N_0m$ (m = emigration rate during expansion), SSD (sum of squared deviations between observed and expected mismatch distributions), sample size = no. of haplotypes, S = no. of segregating sites, π = no. of pairwise differences, Tajima's D = test statistic, $\theta_\pi = 4N\mu$ estimated from π.

Table S1.6. Environmental variables used for phylogeographic analyses. All variables are from the Community Climate System Model (CCSM), from the WorldClim dataset (BioClim variables). For each pure lineage in Yosemite NP (Y-North, Y-East, Y-South, Y-West) variables were only retained if their permutation importance was >0. Permutation importance is the percentage drop in training AUC when the variable is randomly permuted. Those variables were kept, and the model was rerun. Values shown below are the permutation importance values for the final model. Variables used for the linear discriminant analysis (LDA) of all seven lineages in Yosemite are denoted with an 'x'. All variables were used in a PCA for analysis of niche overlap and divergence in Yosemite.

Variable	Description	North	East	South	West	LDA
bio1	Annual Mean Temperature	78.6	27.3	54.5		x
bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))					
bio3	Isothermality (bio2/bio7) (* 100)	1.2	0		1.6	x
bio4	Temperature Seasonality (standard deviation * 100)	1.7			0.8	
bio5	Max Temperature of Warmest Month			0		
bio6	Min Temperature of Coldest Month					
bio7	Temperature Annual Range (bio5 – bio6)				0.1	
bio8	Mean Temperature of Wettest Quarter		0	11.6	69.2	x
bio9	Mean Temperature of Driest Quarter					
bio10	Mean Temperature of Warmest Quarter					
bio11	Mean Temperature of Coldest Quarter			0		
bio12	Annual Precipitation		0			
bio13	Precipitation of Wettest Month					
bio14	Precipitation of Driest Month	0.7	31.4			x
bio15	Precipitation Seasonality (Coefficient of Variation)	17.7	39.2	33.8	27.5	x
bio16	Precipitation of Wettest Quarter					
bio17	Precipitation of Driest Quarter				0.9	x
bio18	Precipitation of Warmest Quarter					
bio19	Precipitation of Coldest Quarter	0.1	2			

Table S1.7. Climatic niche overlap and equivalency results. Schoener's niche overlap metric D denoting pairwise overlap in climatic niche based on ordinated BioClim values for Yosemite NP. All four pure lineages were considered; admixed lineages were excluded because their sample sizes are low. Parenthetical numbers are p-values for niche equivalency tests, where the alternative hypothesis is that niches are less equivalent than randomly chosen distributions in Yosemite. Significance was assessed with 1000 replicate simulations. All metrics were calculated using the ecospat package in R. See Fig. S1.10 for linear discriminant analysis results.

	North	East	South	West
North		0.106 (0.003)	0.000 (<0.001)	0.000 (<0.001)
East			0.000 (<0.000)	0.000 (<0.000)
South				0.467 (0.051)
West				


A. punctatus
A.boreas (Lassen National Park, CA)
A. boreas (Wawona, CA)
A. boreas (San Diego, CA)
A. boreas (Bakersfield, CA)
A. boreas (Sageland, CA)
A. boreas (Yosemite Valley, CA)
A. boreas (Little Lake, CA)
A. boreas (Westguard Pass, CA)
Y3613
Y3537
Y4164
Y4146
Y4324
Y4391
Y4136
Y4365
Y4370
Y4317
Y3763
Y3452
Y3273
Y3384
Y3615
Y3612
Y3225
Y3200
Y2498
Y2824
Y2618
Y2977
Y2407
Y3420
Y3424
Y2594
Y2147
Y2256
Y3371
Y3339
Y2312
Y2135
Y1960
Y2026
Y2059
Y2104
Y2893
Y3272
Y2429
Y2384
Y2836
Y2351
Y2410
Y2324
Y2864
Y2012
Y1756
Y2443
Y2391
Y2369
Y2411
Y2421
Y1841
Y2030
Y1779
Y2132
Y2118
Y2073
Y2418
Y2371
Y2385
Y1070
Y1569
Y1543
Y1171
Y377
Y1040
Y733
Y719
Y758
Y1369
Y780
Y705
Y359
Y387
Y638
K902
K958
K1131
K1141
K1179
K1165
K24
K1071
K101
K916
K1381
K1138
0.0020

Figure S1.1. Maximum likelihood phylogram for Yosemite toads
A maximum likelihood phylogram from RAxML using 1000 bootstrap replicates and 10 heuristic searches under a GTR + Γ model. Node values denote bootstrap support. A total of 19,132 variable sites were used in the analysis, and the Stamatakis method of ascertainment bias correction was applied to account for the known number of invariable sites, and base frequencies. Meadows with suspected recent admixture among lineages based on preliminary results were excluded from phylogenetic analysis. These included meadows from Y-East-North-A2, Y-East-South-A, and Lyell Canyon, which is suspected to contain admixture with Sierra National Forest. One sample of *Anaxyrus punctatus* was used as outgroup, and eight samples of *A. boreas* were included to assess their monophyly, and to help polarize samples of *A. canorus*.

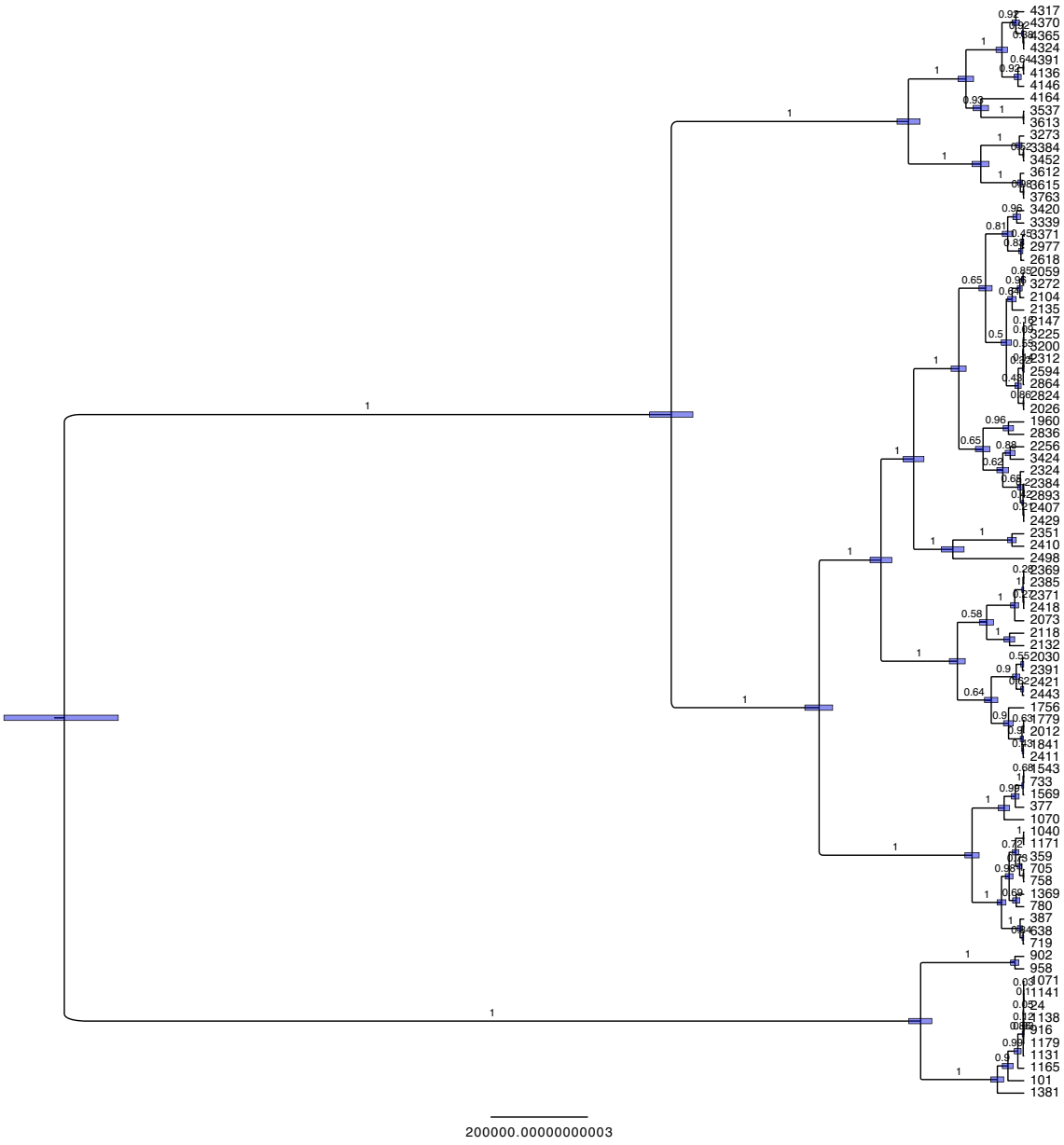


Figure S1.2. Bayesian chronogram for Yosemite toads
A BEAST phylogram produced using a GTR + Γ model and strict clock, and run for 50,000,000 generations. Node values denote posterior probabilities. Node bars denote 95% HPD of node heights, based on a molecular clock of $9.24 \times 10^{-10} – 1.53 \times 10^{-9}$ substitutions site^{-1} year^{-1}, following Crawford (2003). A total of 726,027 sites were used in the analysis. Meadows with suspected recent admixture among lineages based on preliminary results were excluded from phylogenetic analysis. These included meadows from Y-East-North-A2, Y-East-South-A, and Lyell Canyon, which is suspected to contain admixture with Sierra National Forest.

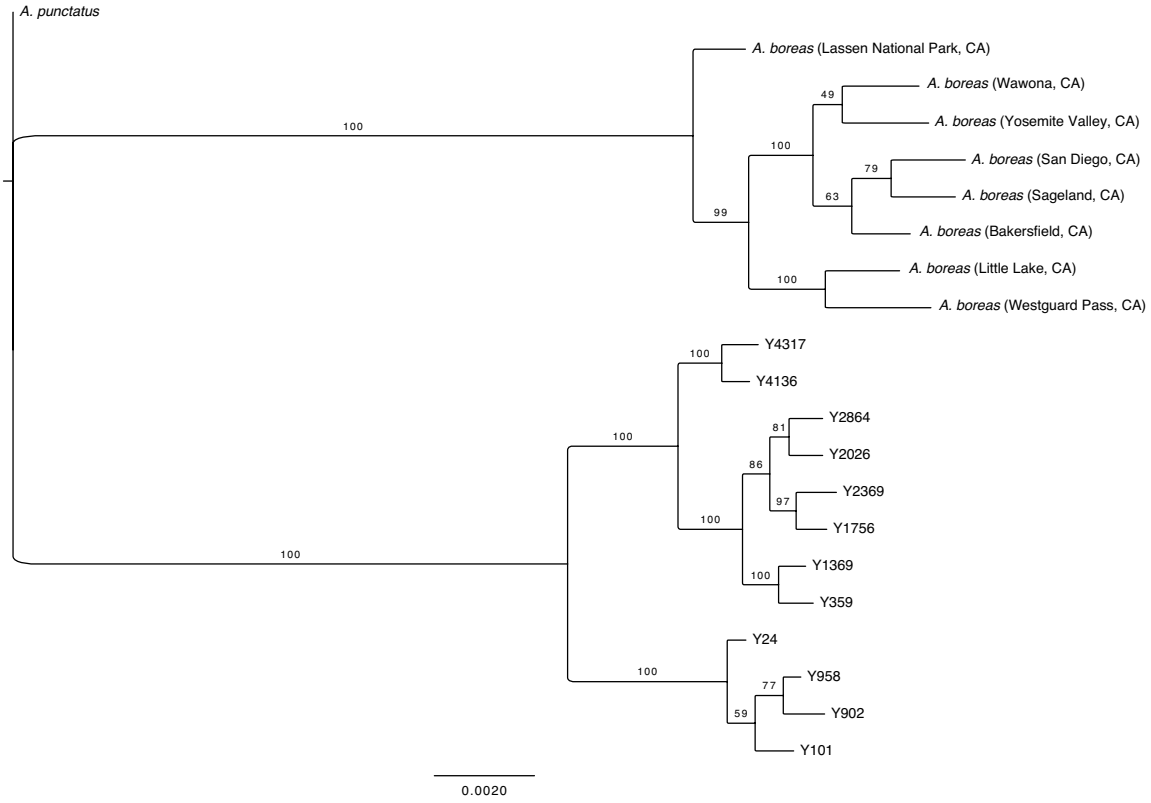


Figure S1.3. Maximum likelihood phylogram balancing Yosemite/western toads
A maximum likelihood phylogram from RAxML that balanced sampling from both species, and minimized missing data to 75% across individuals of both species. All eight samples of *Anaxyrus boreas* were used, along with 12 samples of *A. canorus* (two from each of six pure lineages). Settings used were identical to those from Fig. S1.1: 1000 bootstrap replicates and 10 heuristic searches under a GTR + Γ model. Node values denote bootstrap support. A total of 6,162 variable sites were used in the analysis, and the Stamatakis method of ascertainment bias correction was applied to account for the known number of invariable sites, and base frequencies. One sample of *Anaxyrus punctatus* was used as outgroup.

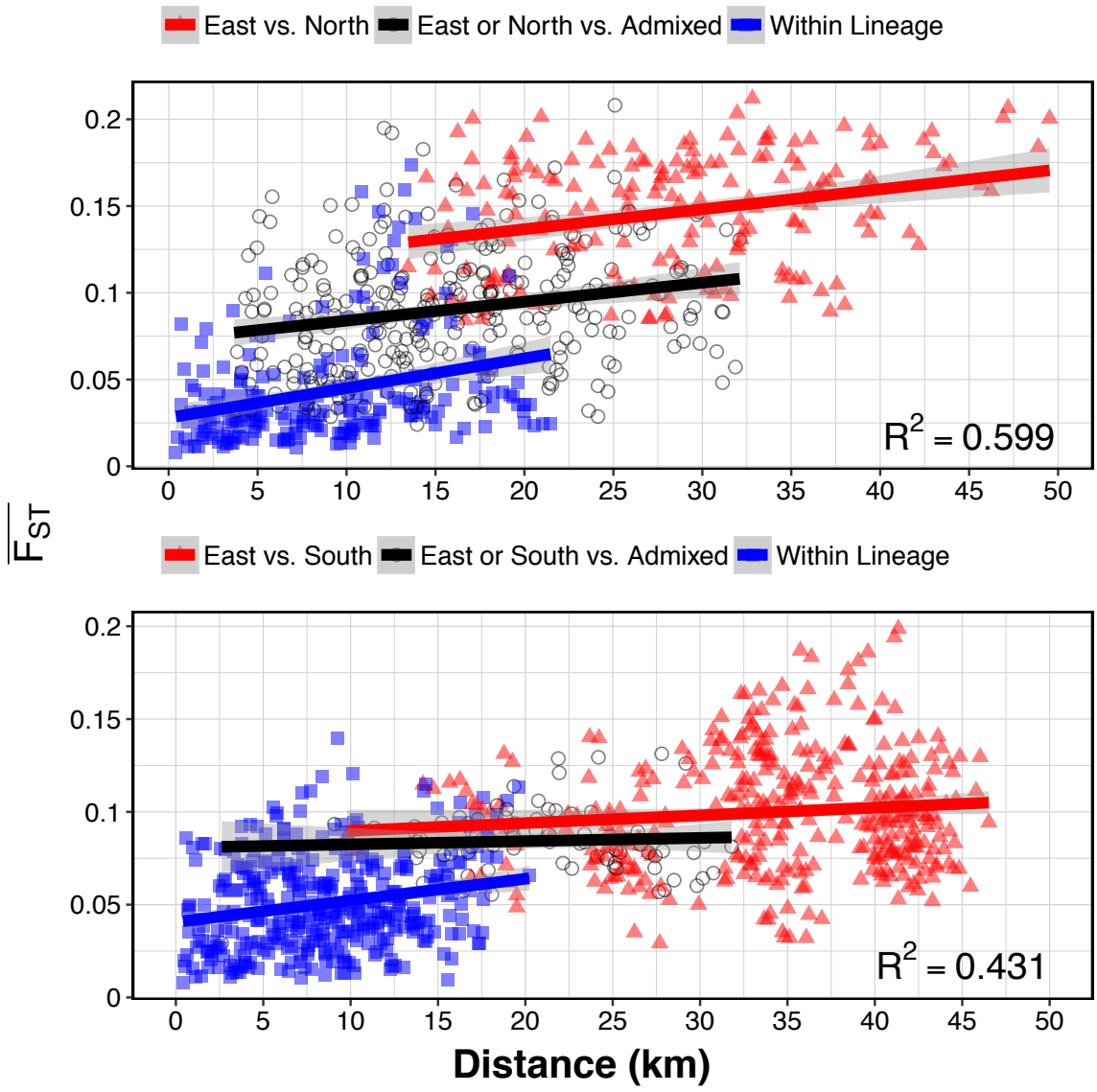


Figure S1.4. Isolation by distance across contact zones in Yosemite NP
Average marker F_{ST} vs. topographically corrected distance, while transecting across each contact zone (northern and southern, in Yosemite NP). R^2 is the overall adjusted value for an ANCOVA model. A Compute Tukey Honest Significant Differences test showed the means of all factor levels to be significantly different, except for the "East vs. South" and "East or South vs. Admixed" contrast.

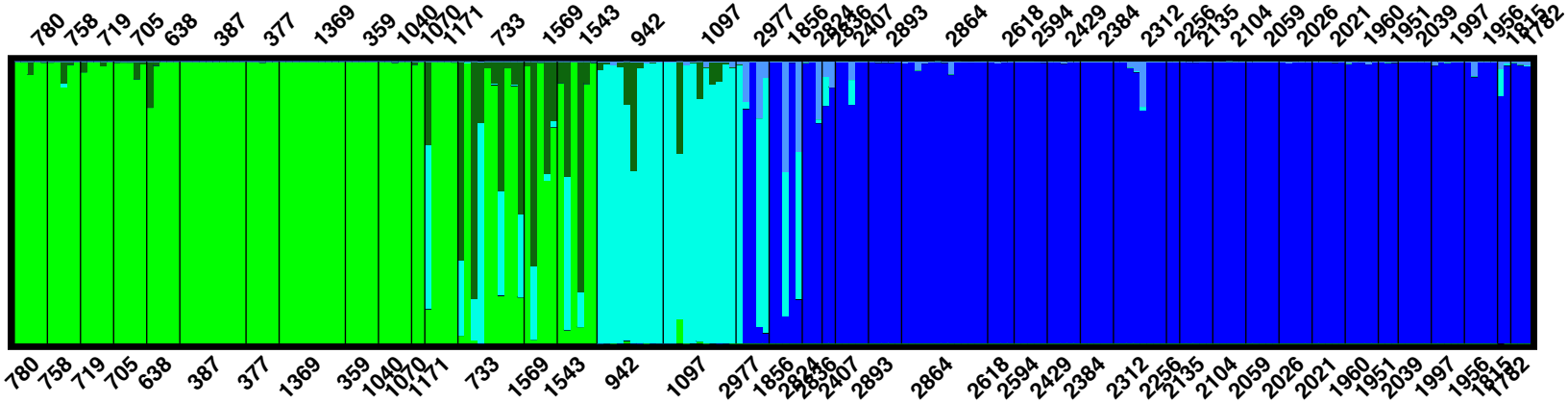


Figure S1.5. Contact zone dynamics in southern Yosemite NP
Gradient of admixture classes following a linear transect across two lineages: Y-East (dark blue) and Y-South (light green). Only lineage-diagnostic markers with highest 1% contribution to sPC2—which corresponds to the East-South zone— are used. NewHybrids barplot shows the probability of recent admixture across a putative contact zone. Each color represents an admixed genotype class (P1: dark blue, P2: light green, F2: cyan, F1-P1 backcross: dull blue, F1-P2 backcross: dark green).

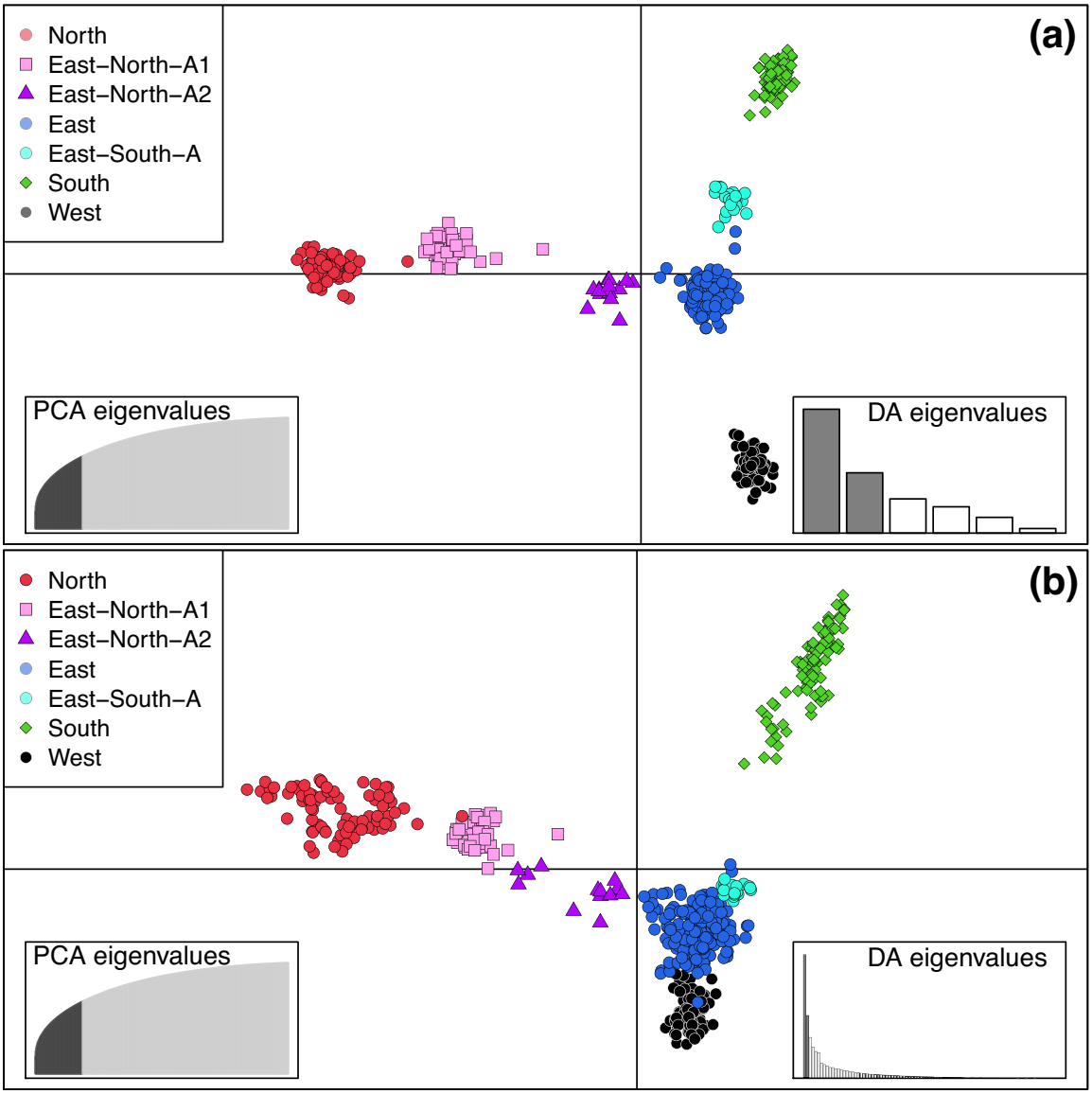


Figure S1.6. Discriminant analysis of principal components in Yosemite NP
DAPC analysis using (a) lineages and (b) meadows as the grouping factor. Meadows cluster into seven pure and admixed lineages along the first two DA axes of each analysis. The first 100/535 PCs were retained. Lineage membership was determined by k-means clustering with 10^9 iterations using the find.clusters function in the adegenet package. Y-North, Y-East, Y-West, and Y-South lineages based on colors in previous figures are shown as circles, with admixed lineages shown as intermediate colors: Y3273, Y3384, Y3452, Y3612, Y3615, Y3763, Y4025 (East-North-A1 shown as pink squares); Y3342, Y3400, Y3414 (East-North-A2 shown as purple triangles); Y942, Y1097 (East-South-A shown as cyan diamonds).

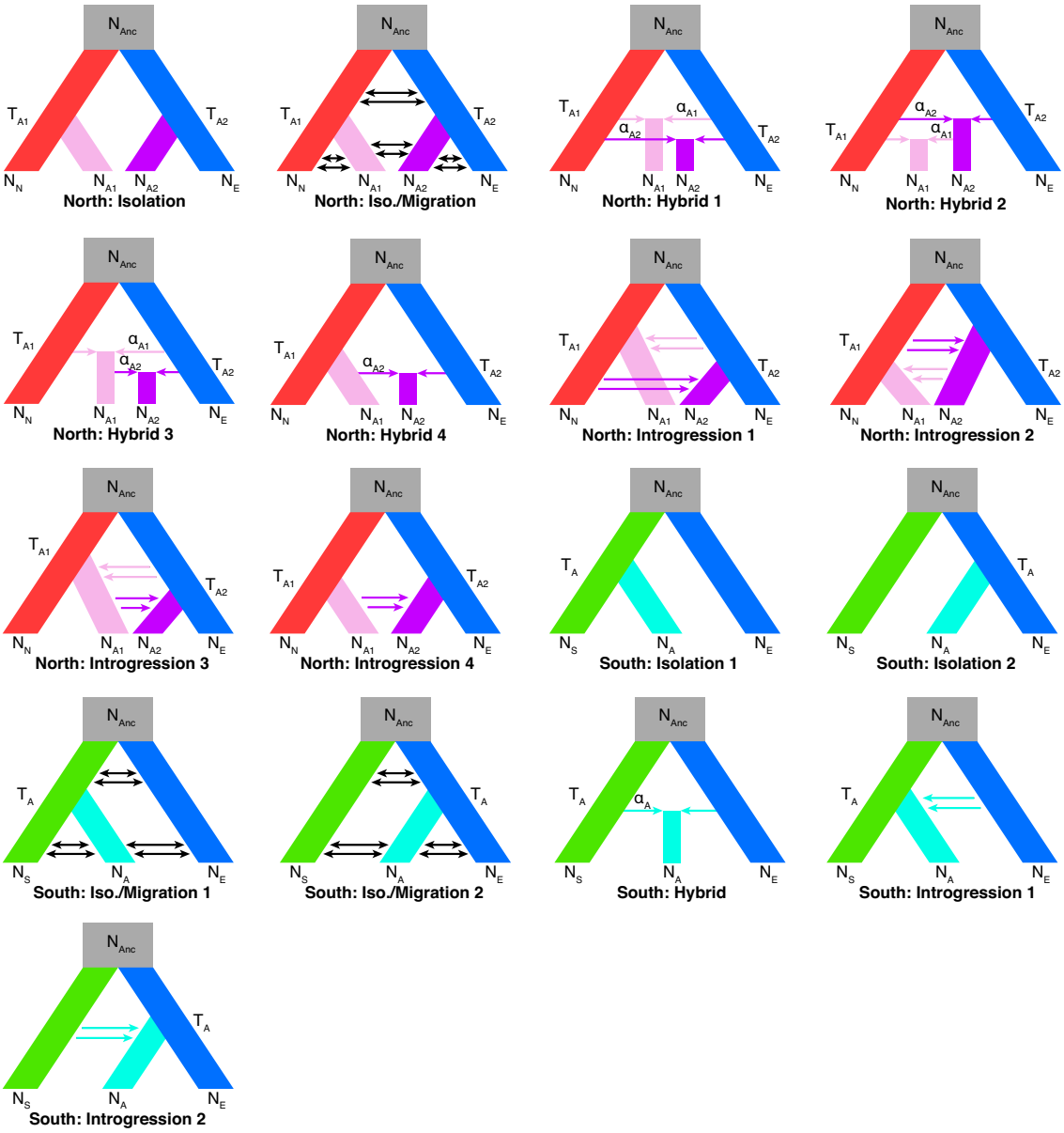


Figure S1.7. Hypothesized models of admixed lineage origin in Yosemite NP
All models tested using fastsimcoal2 for two admixture zones in Yosemite NP: East-North and East-South. Parameters include: population sizes for North (N_N), East-North-A1 (N_{A1}), East-North-A2 (N_{A2}), East (N_E), South (N_S), East-South-A (N_A), and ancestral (N_{Anc}) lineages; admixture proportions into East-North-A1 (α_{A1}), East-North-A2 (α_{A2}), and East-South-A (α_A); divergence or admixture times for East-North-A1 (T_{A1}), East-North-A2 (T_{A2}), and East-South-A (T_A) lineages; and asymmetrical, or unidirectional (i.e. introgression) migration rates. Migration parameters do not have abbreviations, to save space. Double arrows represent ongoing gene flow, whereas single arrows represent single admixture events.

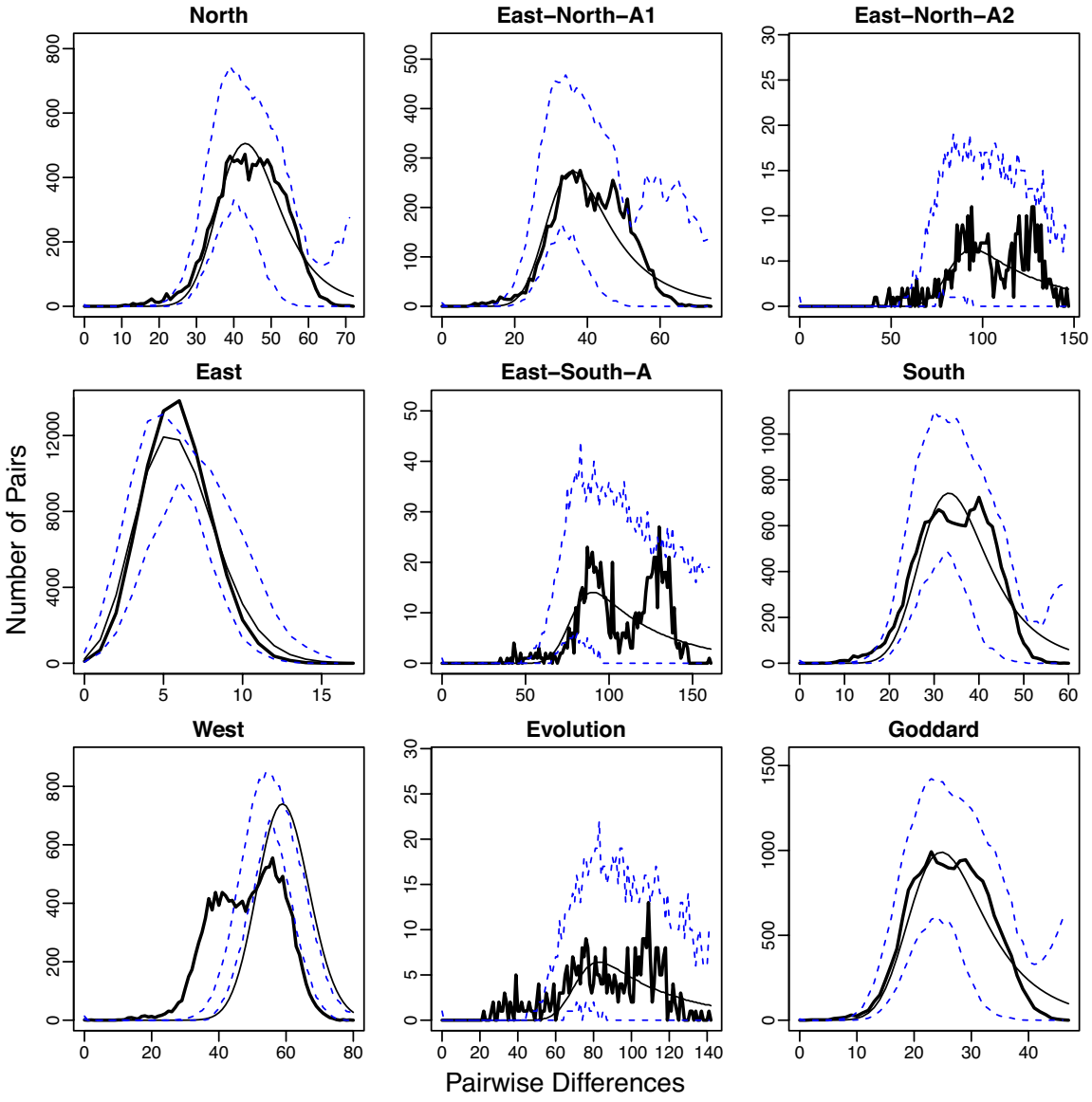


Figure S1.8. Mismatch distributions for all lineages
Mismatch distributions and expected mismatch distribution 95% confidence intervals under the spatial expansion model. Unimodal distributions tend to occur after recent and/or large expansions, whereas multimodal distributions reflect the randomness of coalescence in a stable population.

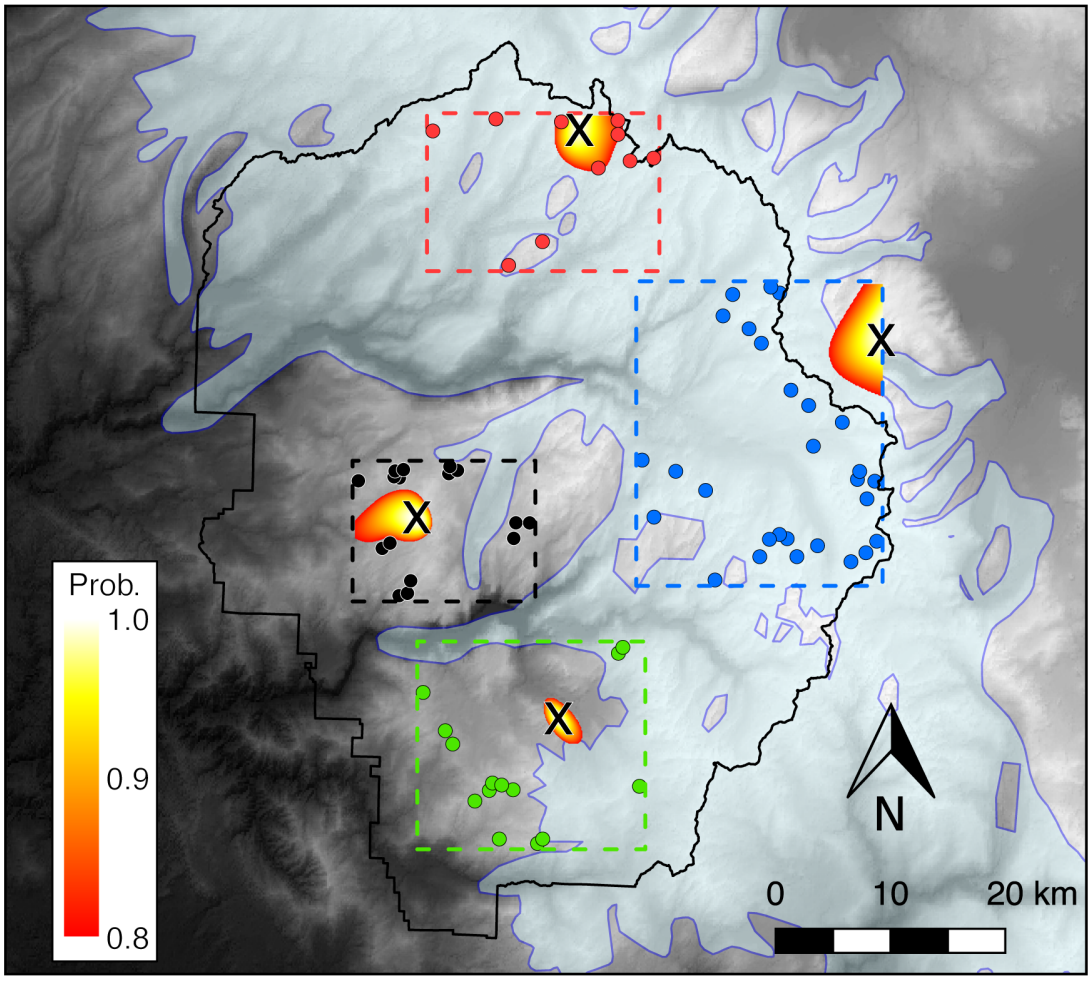


Figure S1.9. Origins of post-glacial range expansions in Yosemite NP
Origins of post-Pleistocene colonization events in the four pure lineages of Yosemite NP; admixed lineages were excluded because their sample sizes are low. The method of Peter and Slatkin (2013) was used to infer the directionality index (ψ), based on pairwise allele frequency asymmetries that result from drift fixing lower frequency alleles during an expansion. Based on observed ψ values, the origin was inferred using time difference of arrival methods. A key assumption of the method is even sample sizes, so five individuals were randomly chosen from every meadow. The inferred origin cannot be extrapolated outside the study areas (denoted with dashed boxes), which makes an external origin more likely for Y-North (red) and Y-East (lineages), compared with Y-West (black) and Y-South (green) lineages. The light blue polygon denotes the maximum extent of glaciation in the Sierra Nevada, which was very similar among different glacial events.

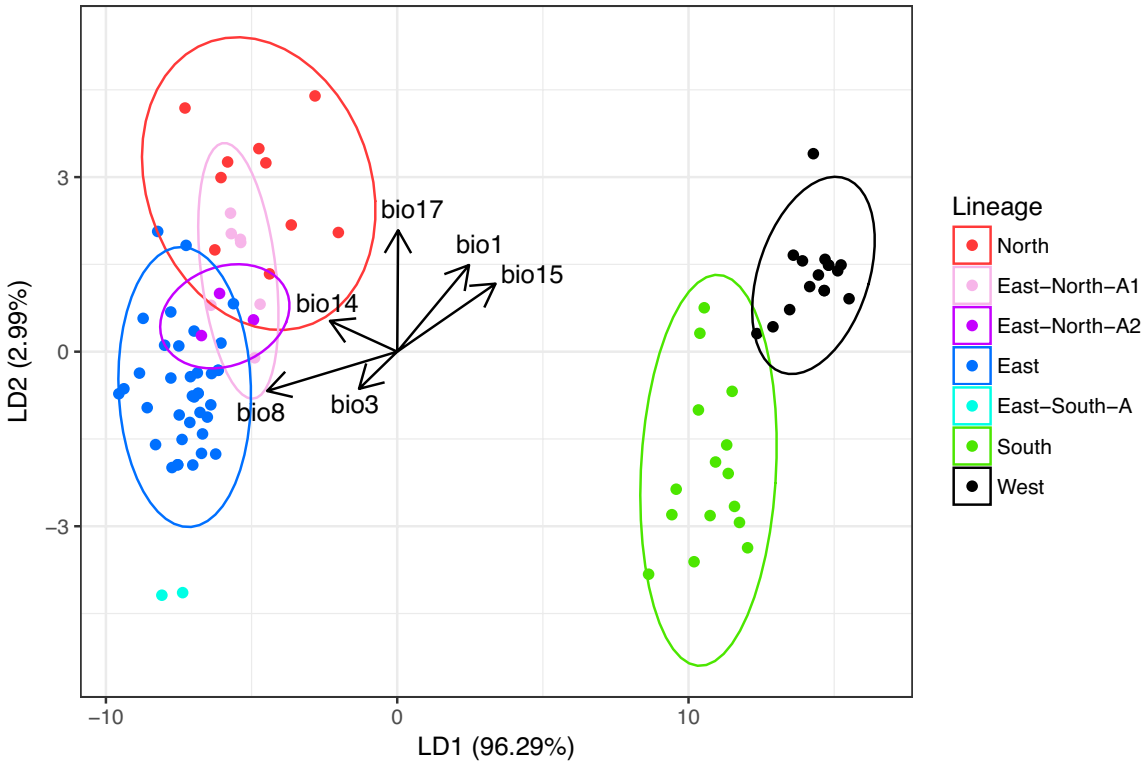


Figure S1.10. Climatic niche overlap and differentiation in Yosemite NP
A linear discriminant biplot of six BioClim variables shows the climatic differentiation between the lineages in Yosemite NP. See Table S1.6 for variable definitions, and Table S1.7 for results of formal niche overlap and equivalency tests.

APPENDIX C: SUPPLEMENTARY METHODS FOR CHAPTER 2

RNA sequencing, transcriptome assembly, and annotation

Three tadpoles were collected for RNAseq and subsequent transcriptomic analysis between June 20 and July 3, 2013. Individuals were chosen to span the lineage diversity in Yosemite: Y-North (Twin Lakes; 38.130080, -119.666431), Y-South (3.8 km SE of Turner Meadow; 37.569148, -119.576916), and Y-East (Summit Lake; 38.051194, -119.323416). All tadpoles sampled were between Gosner stages 26 and 36. Whole tadpoles were euthanized following IACUC protocol, immersed in RNA*later,* and stored at -20 °C within three days of collection. A Bullet Blender was used to homogenize tadpole tissues, and then whole RNA was extracted with a Qiagen RNeasy Mini Kit, and eluted into RNAse free ddH2O. RNAseq libraries were prepared by Hudson Alpha, which used the NEBNext Poly(A) kit to isolate mRNA (transcribed RNAs of interest) from rRNA. Libraries were sequenced using 100 bp paired-end reads on an Illumina HiSeq 2500.

Quality control of raw reads was performed with Trimmomatic v0.32 (Bolger et al. 2014). Illumina adapter sequences were clipped, and reads were trimmed using a sliding window of 4 bp, wherever average quality (Phred) scores were <20. Leading edges were also trimmed until the first position of Phred <20, and trailing edges were trimmed after Phred <10. Resulting reads shorter than 20 bp were removed. Paired end reads were then sorted and isolated from orphan reads. Quality stats and nucleotide distribution per bp position were then visualized using FastQC v0.10.1 (Andrews 2010) to affirm the absence of contamination.

Paired and singleton reads passing quality threshold were assembled into a de novo larval transcriptome using Trinity r2014-02-04 (Grabherr et al. 2011; Haas et al. 2013). Trinity addresses the complexity of alternative isoforms in RNAseq with three successive scripts: (1) Inchworm assembles contigs from the longest isoform, retaining only the unique portions of other isoforms; (2) Chrysalis clusters all isoforms for a given gene together with de Bruijn graphs of overlapping k-mers; and (3) Butterfly traverses those graphs, and reconstructs all likely isoforms while discarding likely paralogs. BUSCO v3 (Simão et al. 2015) was used to check transcriptome quality and completeness, based on a near-universal database of tetrapod single-copy orthologs (Tetrapoda OrthoDB v9).

I annotated the de novo transcriptome using the Trinotate suite included in Trinity, using standard methods (https://trinotate.github.io/). Briefly, Transdecoder v5.0.2 (Haas et al. 2013) identified open reading frames (ORFs) of at least 300 bp. I used the Basic Local Alignment Search Tool (BLAST) to align translated ORFs with the SwissProt database (blastp search), and untranslated transcripts with the non-redundant nucleotide database (blastx search), to assign gene identity (Blast+ v 2.2.28; Altschul et al. 1990). I used a maximum e-value of 1×10^{-10}. Gene ontology (GO) terms were ascribed using three databases: Gene Ontology (Ashburner et al. 2000), KEGG (Kanehisa and Goto 2000), and eggNOG (Powell et al. 2014). Predicted protein domains were found using Pfam (Bateman et al. 2004) and HMMER (Finn et al. 2011), and protein signal peptide and transmembrane domain predictions were made using signalP (Petersen et al. 2011) and tmHMM (Krogh et al. 2001).

Identifying and annotating transcriptomic SNPs

Variants (SNPs and indels) were called across the de novo transcriptome using the GATK v4.0.1.2 best practices for RNAseq (McKenna et al. 2010). First, a reference transcriptome was formed as a set of supertranscripts using Trinity_gene_splice_modeler.py in Trinity v2.5.1. A supertranscript is the concatenation of alternative splicing isoforms produced by a gene, in order to make a single reference sequence (Davidson et al. 2017). Annotated supertranscripts were the indexed with Picard Tools v2.14.1 (broadinstitute.github.io/picard/). Raw reads were aligned with STAR v2.5.3a (Dobin et al. 2013), which GATK developers have found to have the highest true positive and lowest false negative rates of SNP and indel discovery for RNAseq data. I used the STAR 2-pass method, whereby splice junctions detected initially are used to guide a secondary alignment (Engström et al. 2013). Picard Tools was then used to group and sort reads, and mark any PCR duplicates.

A variant-calling pipeline specific to RNA-seq was conducted in GATK v4.0.1.2. First, the tool SplitNCigarReads was used to hard-clip any read fragments spanning exon-intron boundaries. Next, potential variants were called using the tool HaplotypeCaller, which works by assembling reads into potential haplotypes, and compares these back to the reference to calculate confidence in alternate alleles (SNP or indels), and likelihood of genotypes. This tool deals with further challenges of RNAseq data by rescuing variants near splice junctions, while taking into account the graph complexity of alternate splicing. I used a minimum phred-scaled confidence threshold of 20.0, the AF QUAL

score metric (-newQual flag, to minimize loss of singletons), and a ploidy of six to accommodate the pooling of three diploid samples. VariantFiltration was next used to filter clusters of three SNPs in a 35 bp window size with Fisher Strand values > 30.0 and Quality By Depth values (QD) < 2.0. Finally, SelectVariants was used to isolate a dataset with only biallelic SNPs (excluding indels, and multi-allelic SNPs).

I individually genotyped each of the three transcriptomes by re-running the above pipeline, with the following changes: (1) only SNP loci identified above (with 100 bp margin for proper genomic context) were considered; (2) variant files were combined with CombineGVCFs, so that the more powerful joint-genotyping algorithm could be applied in GenotypeGVCFs. All VCF files were combined using the merge function in bcftools v1.6 (Li et al. 2009). SNPs were then annotated with synonymous/non-synonymous effects and predicted protein changes in two steps: (1) I reconstructed likely ORFs ≥ 100 amino acids using TransDecoder v5.0.2; (2) SNP effects were annotated using Ensembl VEP v92.1 (McLaren et al. 2016).

I identified ddRADseq markers that fall within coding genes by performing a blast-n nucleotide search on full RAD sequences, with the transcriptome as the database. To minimize the possibility of matching paralogs, I only allowed ungapped alignments with no more than three mismatches beyond the known RAD SNPs, with an e-value cutoff of 1×10^{-6}.

APPENDIX D: SUPPLEMENTARY TABLES/FIGURES FOR CHAPTER 2

Table S2.1. List of specific gene ontology (GO) terms unique to a category of locus. All specific terms are shown, rather than a summary across broad (Level 2) categories as shown in Table 2.1.

Group*	Ontology†	GOID	GO Term
All Islands	BP	GO:0035887	aortic smooth muscle cell differentiation
All Islands	BP	GO:0043044	ATP-dependent chromatin remodeling
All Islands	BP	GO:0001832	blastocyst growth
All Islands	BP	GO:0001835	blastocyst hatching
All Islands	BP	GO:0000902	cell morphogenesis
All Islands	BP	GO:0006325	chromatin organization
All Islands	BP	GO:0006338	chromatin remodeling
All Islands	BP	GO:0060318	definitive erythrocyte differentiation
All Islands	BP	GO:0010424	DNA methylation on cytosine within a CG sequence
All Islands	BP	GO:0035116	embryonic hindlimb morphogenesis
All Islands	BP	GO:0048562	embryonic organ morphogenesis
All Islands	BP	GO:0048730	epidermis morphogenesis
All Islands	BP	GO:0030198	extracellular matrix organization
All Islands	BP	GO:0030900	forebrain development
All Islands	BP	GO:0007403	glial cell fate determination
All Islands	BP	GO:0060347	heart trabecula formation
All Islands	BP	GO:0030902	hindbrain development
All Islands	BP	GO:0043966	histone H3 acetylation
All Islands	BP	GO:0030216	keratinocyte differentiation
All Islands	BP	GO:0001889	liver development
All Islands	BP	GO:0006346	methylation-dependent chromatin silencing
All Islands	BP	GO:0060766	negative regulation of androgen receptor signaling pathway
All Islands	BP	GO:0030308	negative regulation of cell growth
All Islands	BP	GO:0008285	negative regulation of cell proliferation
All Islands	BP	GO:2000134	negative regulation of G1/S transition of mitotic cell cycle

Group*	Ontology†	GOID	GO Term
All Islands	BP	GO:0007070	negative regulation of transcription from RNA polymerase II promoter
All Islands	BP	GO:0007399	nervous system development
All Islands	BP	GO:0003407	neural retina development
All Islands	BP	GO:0006334	nucleosome assembly
All Islands	BP	GO:0006337	nucleosome disassembly
All Islands	BP	GO:0043923	positive regulation by host of viral transcription
All Islands	BP	GO:0043388	positive regulation of DNA binding
All Islands	BP	GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity
All Islands	BP	GO:0045944	positive regulation of transcription from RNA polymerase II promoter
All Islands	BP	GO:0045893	positive regulation of transcription, DNA-templated
All Islands	BP	GO:0006357	regulation of transcription from RNA polymerase II promoter
All Islands	BP	GO:0019827	stem cell population maintenance
All Islands	CC	GO:0005615	extracellular space
All Islands	CC	GO:0000792	heterochromatin
All Islands	CC	GO:0045111	intermediate filament cytoskeleton
All Islands	CC	GO:0071565	nBAF complex
All Islands	CC	GO:0071564	npBAF complex
All Islands	CC	GO:0000790	nuclear chromatin
All Islands	CC	GO:0005719	nuclear euchromatin
All Islands	CC	GO:0005726	perichromatin fibrils
All Islands	CC	GO:0043234	protein complex
All Islands	CC	GO:0016514	SWI/SNF complex
All Islands	MF	GO:0050681	androgen receptor binding
All Islands	MF	GO:0016887	ATPase activity
All Islands	MF	GO:0003682	chromatin binding
All Islands	MF	GO:0008094	DNA-dependent ATPase activity
All Islands	MF	GO:0004386	helicase activity

Group*	Ontology†	GOID	GO Term
All Islands	MF	GO:0070577	lysine-acetylated histone binding
All Islands	MF	GO:0002039	p53 binding
All Islands	MF	GO:0047485	protein N-terminus binding
All Islands	MF	GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding
All Islands	MF	GO:0001105	RNA polymerase II transcription coactivator activity
All Islands	MF	GO:0030957	Tat protein binding
All Islands	MF	GO:0003713	transcription coactivator activity
All Islands	MF	GO:0003714	transcription corepressor activity
All Islands	MF	GO:0044212	transcription regulatory region DNA binding
EN Zone	BP	GO:0007049	cell cycle
EN Zone	BP	GO:0030030	cell projection organization
EN Zone	BP	GO:0051297	centrosome organization
EN Zone	BP	GO:0008104	protein localization
EN Zone	CC	GO:0005814	centriole
EN Zone	CC	GO:0005813	centrosome
EN Zone	CC	GO:0035253	ciliary rootlet
EN Zone	CC	GO:0070062	extracellular exosome
EN Zone	MF	GO:0019894	kinesin binding
EN Zone	MF	GO:0005198	structural molecule activity
ES Zone	BP	GO:0032502	developmental process
ES Zone	BP	GO:0015074	DNA integration
ES Zone	BP	GO:0001892	embryonic placenta development
ES Zone	BP	GO:0046909	intermembrane transport
ES Zone	BP	GO:0051452	intracellular pH reduction
ES Zone	BP	GO:0060548	negative regulation of cell death
ES Zone	BP	GO:0090212	negative regulation of establishment of blood-brain barrier
ES Zone	BP	GO:0007584	response to nutrient

Group*	Ontology†	GOID	GO Term
ES Zone	BP	GO:0009268	response to pH
ES Zone	BP	GO:0009636	response to toxic substance
ES Zone	BP	GO:0042360	vitamin E metabolic process
ES Zone	BP	GO:0051180	vitamin transport
ES Zone	CC	GO:0005829	cytosol
ES Zone	CC	GO:0005770	late endosome
ES Zone	MF	GO:0004190	aspartic-type endopeptidase activity
ES Zone	MF	GO:0004519	endonuclease activity
ES Zone	MF	GO:0003676	nucleic acid binding
ES Zone	MF	GO:0043325	phosphatidylinositol-3,4-bisphosphate binding
ES Zone	MF	GO:0005546	phosphatidylinositol-4,5-bisphosphate binding
ES Zone	MF	GO:0003723	RNA binding
ES Zone	MF	GO:0005215	transporter activity
ES Zone	MF	GO:0019842	vitamin binding
ES Zone	MF	GO:0008431	vitamin E binding
EW Zone	BP	GO:0006355	regulation of transcription, DNA-templated
EW Zone	CC	GO:0016021	integral component of membrane
EW Zone	MF	GO:0005524	ATP binding

* GO terms enriched in both islands and rivers of that zone, but no other zones; e.g. EN islands + EN rivers only

† Ontology is the Level 1 GO category, of which there are just three: biological process (BP), molecular function (MF), and cellular component (CC)

Table S2.2. Model comparison for tadpole growth-development relationship. Polynomial regressions of tadpole length by developmental stage are shown with increasing polynomial degrees, and respective likelihood-based rankings.

Parameters	AICc	BIC	R^2_{adj}	p-value	Shapiro-Wilk	SW p-value
1	9895	9911	0.382	4.16×10^{-180}	0.996	8.29×10^{-5}
2	9596	9618	0.481	1.10×10^{-243}	0.996	1.25×10^{-4}
3	9598	9625	0.481	3.27×10^{-242}	0.996	1.27×10^{-4}
4	9595	9627	0.482	5.63×10^{-242}	0.997	2.44×10^{-3}
5	9594	9632	0.483	2.75×10^{-241}	0.997	1.53×10^{-3}
6	9596	9639	0.483	2.75×10^{-241}	0.997	1.53×10^{-3}

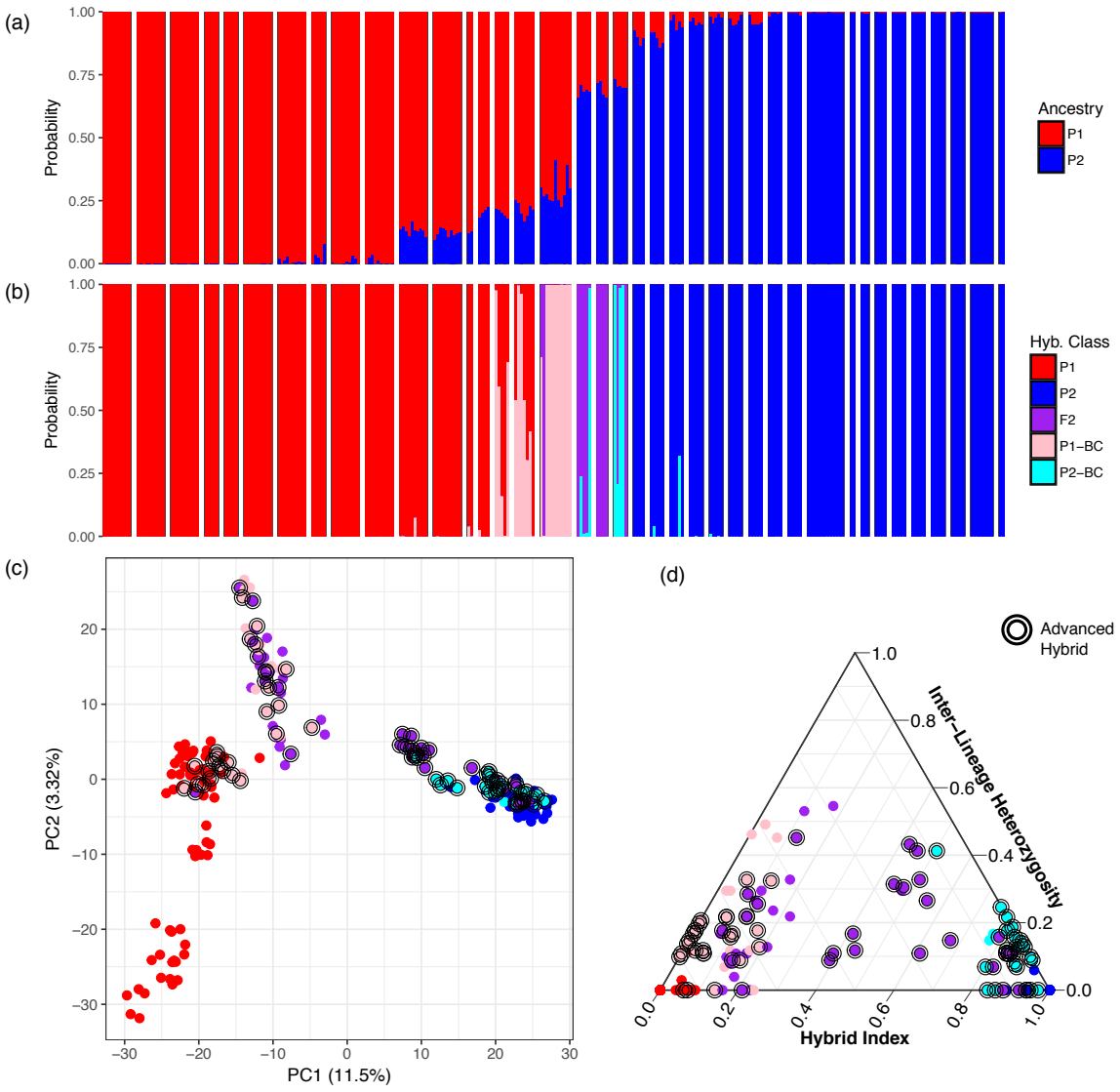


Figure S2.1. Lineage boundaries and admixed individuals for EN contact zone
A four-step procedure was used for determining boundaries of pure lineages, and identifying meadows with putative inter-lineage admixture. (a) STRUCTURE was run 5× with K=2, and runs were combined, to find individuals with dual ancestry. (b) NewHybrids was run to find individuals with evidence of recent admixture. (c-d) HIest was used to find individuals with advanced (>2 generations) admixture, and a PCA was used to assess whether putatively admixed individuals clustered separately and intermediate from pure individuals. Circle colors represent HIest estimates of hybrid category based on the two-generation classification. Double rings indicate individuals identified as "advanced hybrids," i.e. where this two-generation classification can be rejected.

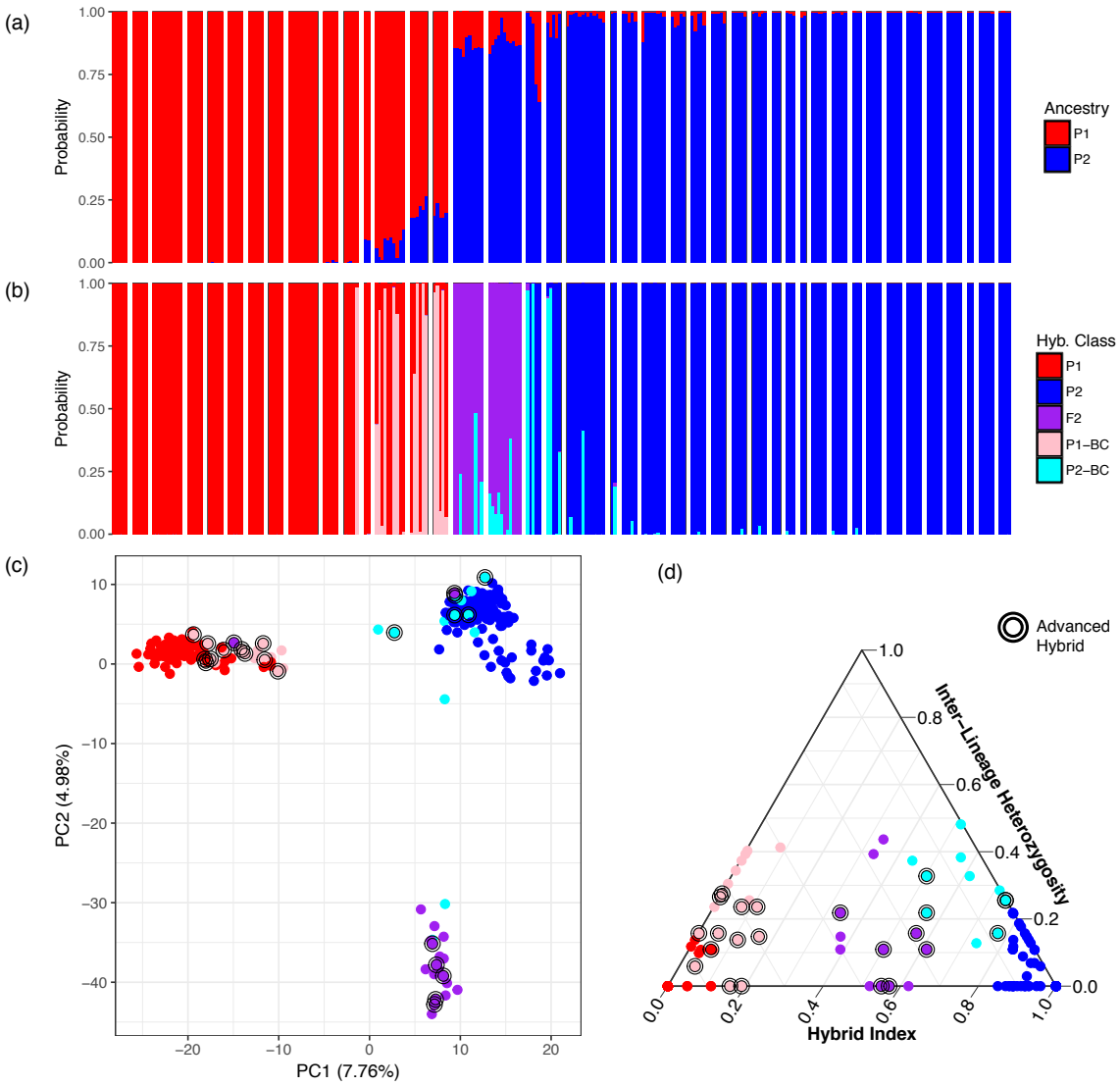


Figure S2.2. Lineage boundaries and admixed individuals for ES contact zone
A four-step procedure was used for determining boundaries of pure lineages, and identifying meadows with putative inter-lineage admixture. (a) STRUCTURE was run 5× with K=2, and runs were combined, to find individuals with dual ancestry. (b) NewHybrids was run to find individuals with evidence of recent admixture. (c-d) HIest was used to find individuals with advanced (>2 generations) admixture, and a PCA was used to assess whether putatively admixed individuals clustered separately and intermediate from pure individuals. Circle colors represent HIest estimates of hybrid category based on the two-generation classification. Double rings indicate individuals identified as "advanced hybrids," i.e. where this two-generation classification can be rejected.

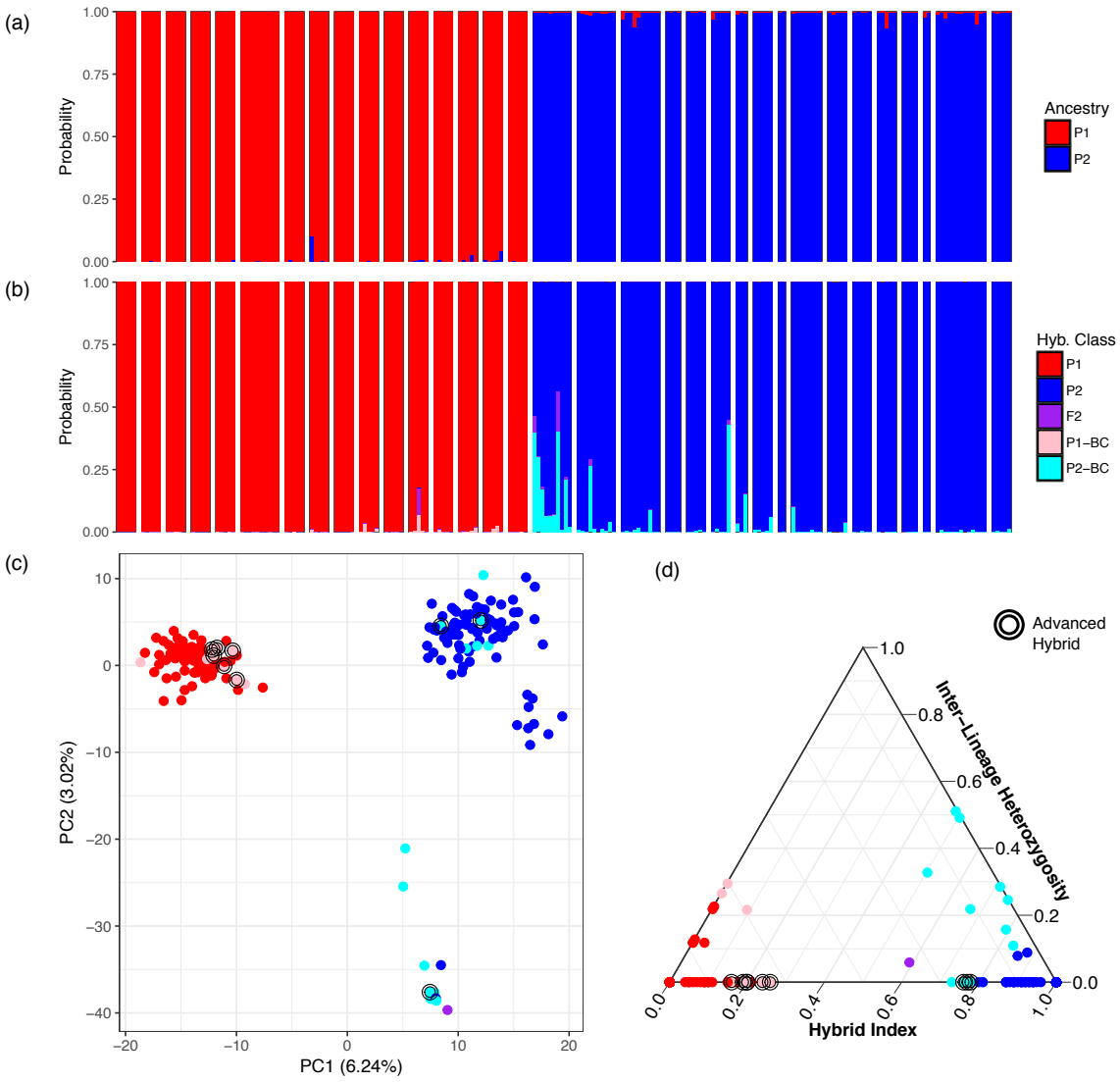


Figure S2.3. Lineage boundaries and admixed individuals for EW contact zone
A four-step procedure was used for determining boundaries of pure lineages, and identifying meadows with putative inter-lineage admixture. (a) STRUCTURE was run 5× with K=2, and runs were combined, to find individuals with dual ancestry. (b) NewHybrids was run to find individuals with evidence of recent admixture. (c-d) HIest was used to find individuals with advanced (>2 generations) admixture, and a PCA was used to assess whether putatively admixed individuals clustered separately and intermediate from pure individuals. Circle colors represent HIest estimates of hybrid category based on the two-generation classification. Double rings indicate individuals identified as "advanced hybrids," i.e. where this two-generation classification can be rejected.

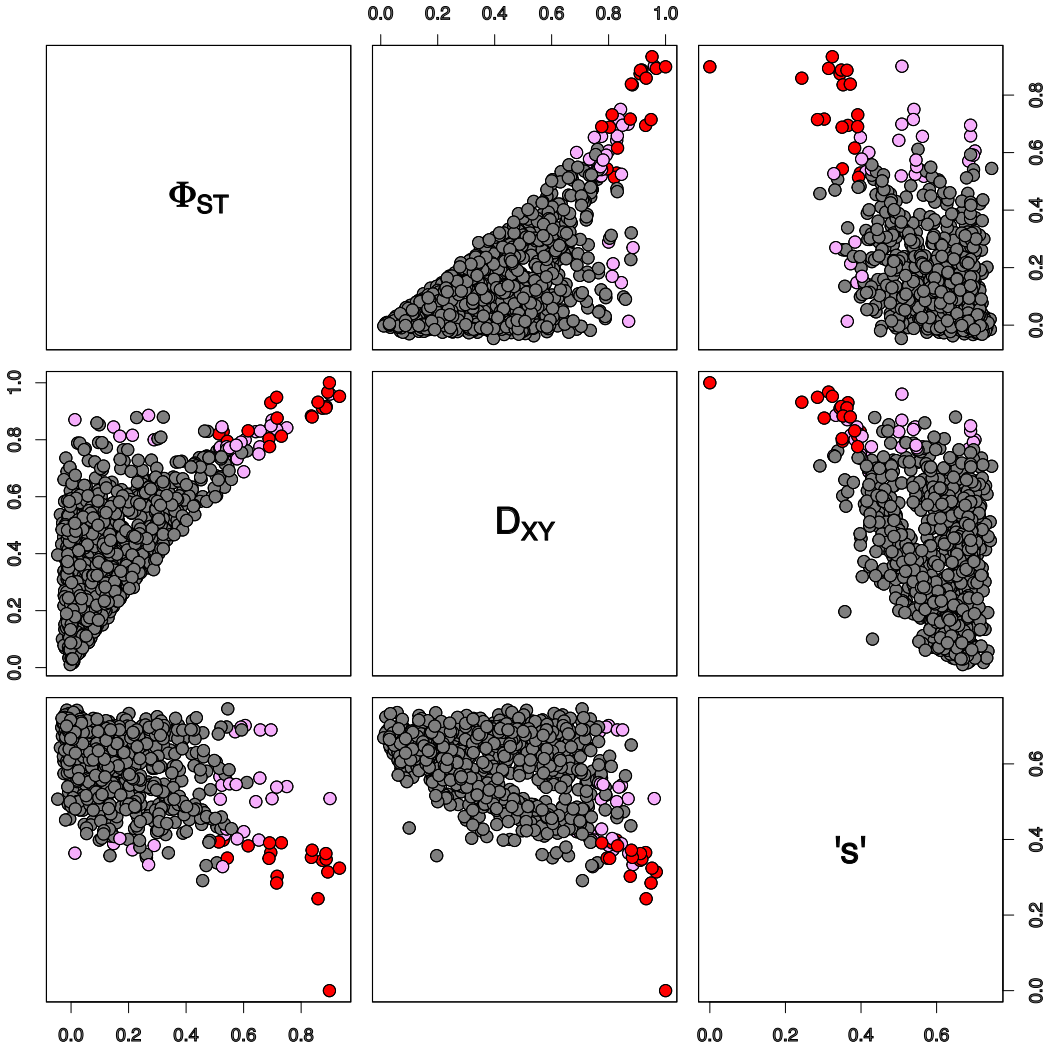


Figure S2.4. Scatterplots of Φ_{ST}, D_{XY}, and Slatkin/Maddison's 's': EN contact zone
Pairwise scatterplots of all three statistics, where 's' is the ratio between the 's' statistic and the number of individuals present in a given gene tree (accounting for missing data).

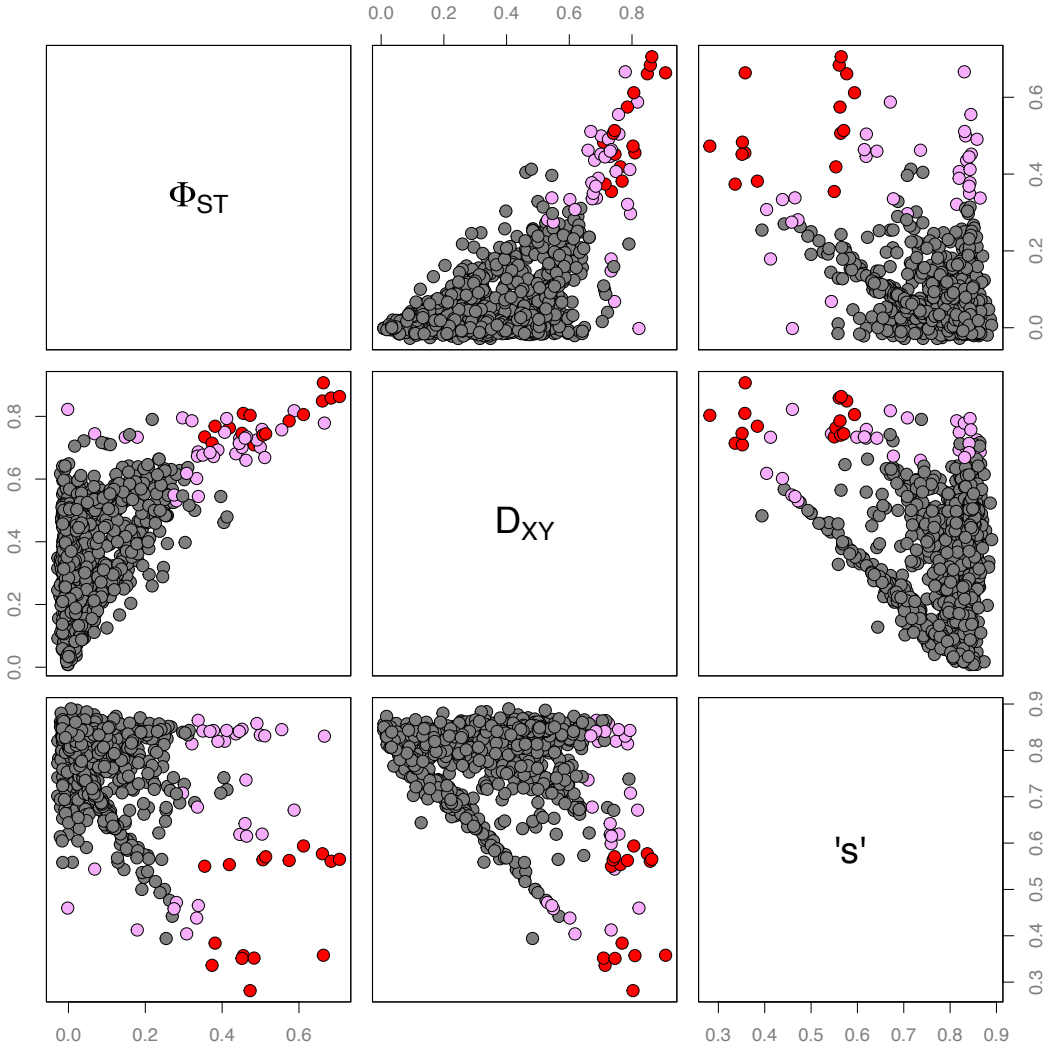


Figure S2.5. Scatterplots of Φ_{ST}, D_{XY}, and Slatkin/Maddison's 's': ES contact zone
Pairwise scatterplots of all three statistics, where 's' is the ratio between the 's' statistic and the number of individuals present in a given gene tree (accounting for missing data).

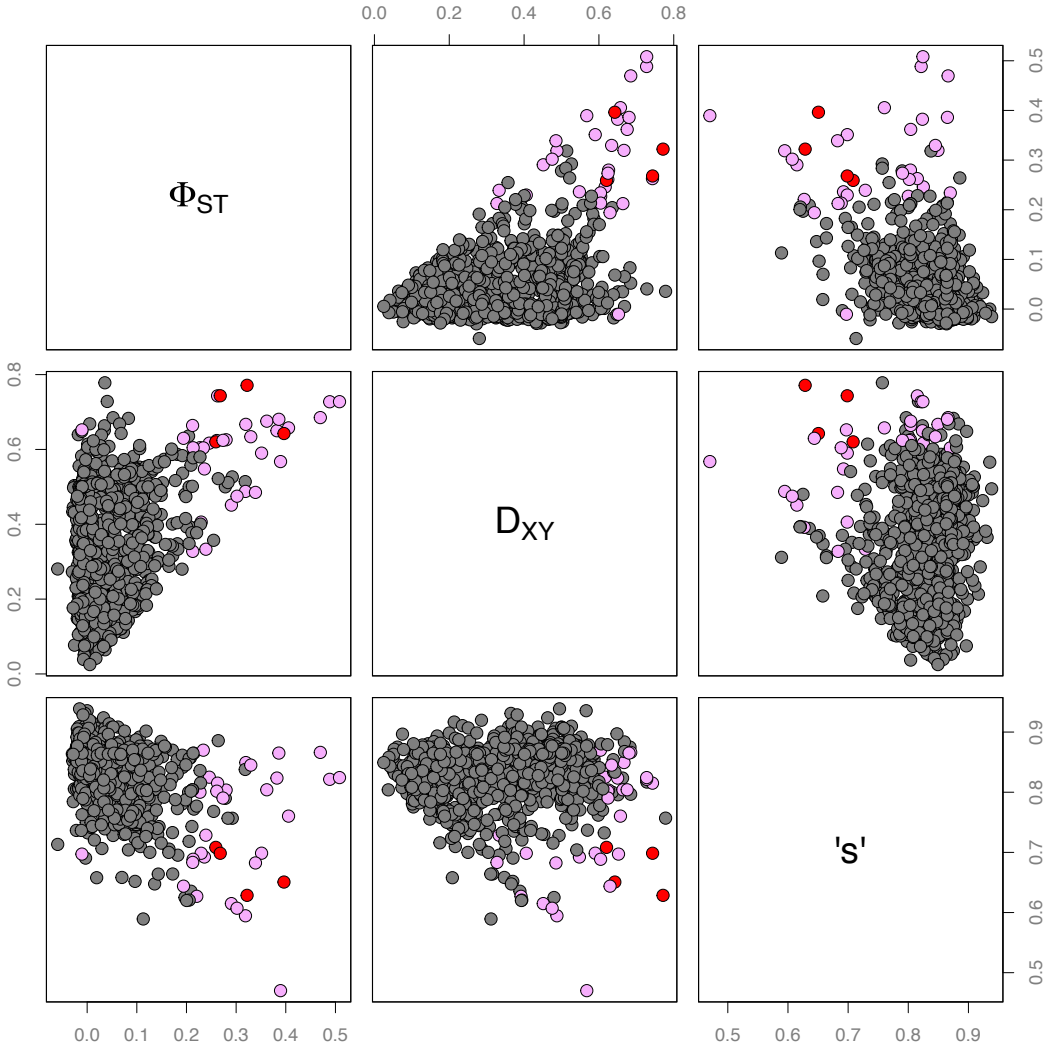


Figure S2.6. Scatterplots of Φ_{ST}, D_{XY}, and Slatkin/Maddison's 's': EW contact zone
Pairwise scatterplots of all three statistics, where 's' is the ratio between the 's' statistic and the number of individuals present in a given gene tree (accounting for missing data).

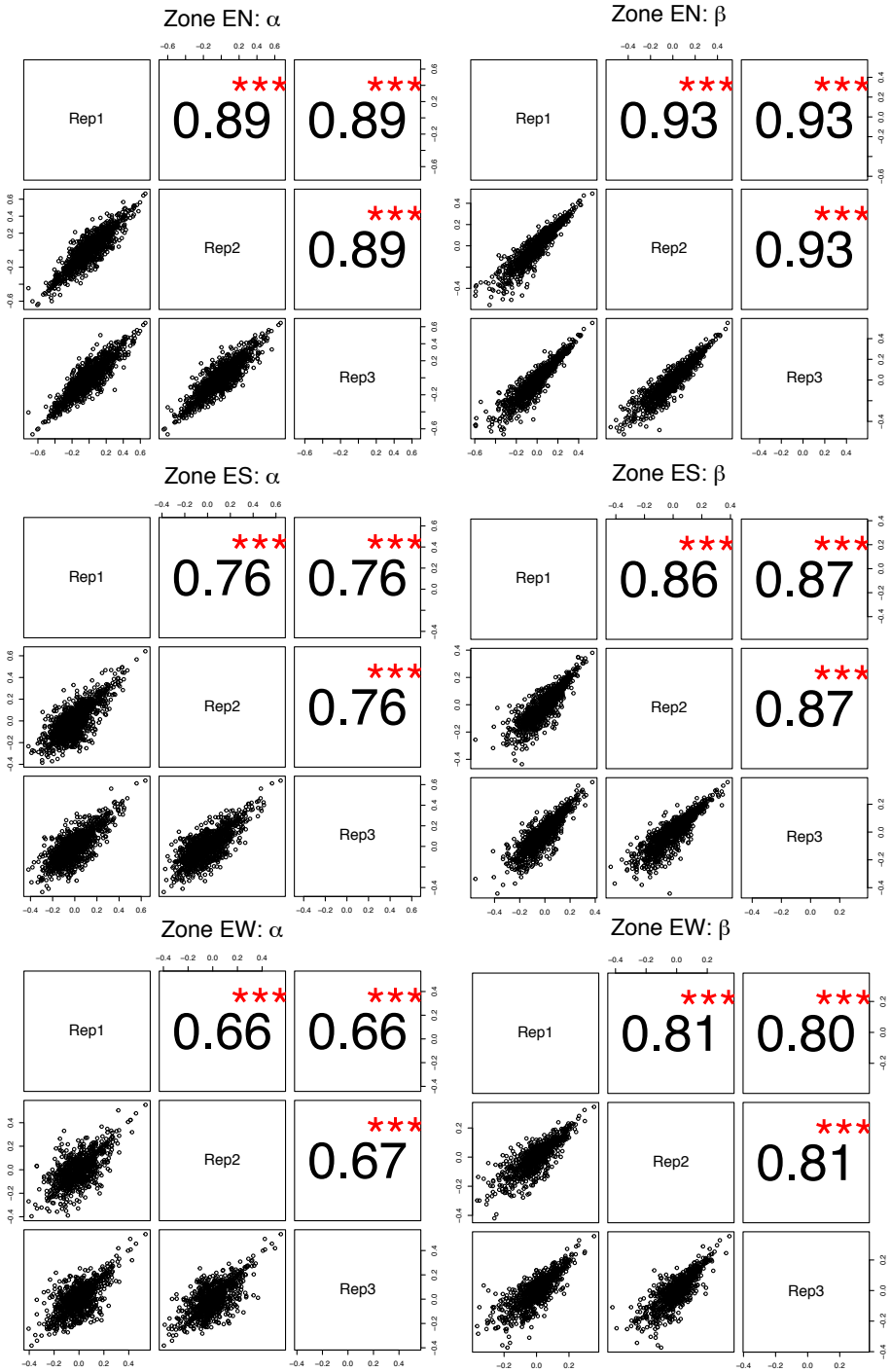


Figure S2.7. Correlation between α and β estimates for three replicate BGC runs
Pairwise scatterplots of all three runs for each zone, showing correlation coefficients and levels of significance (*** = $p < 0.001$).

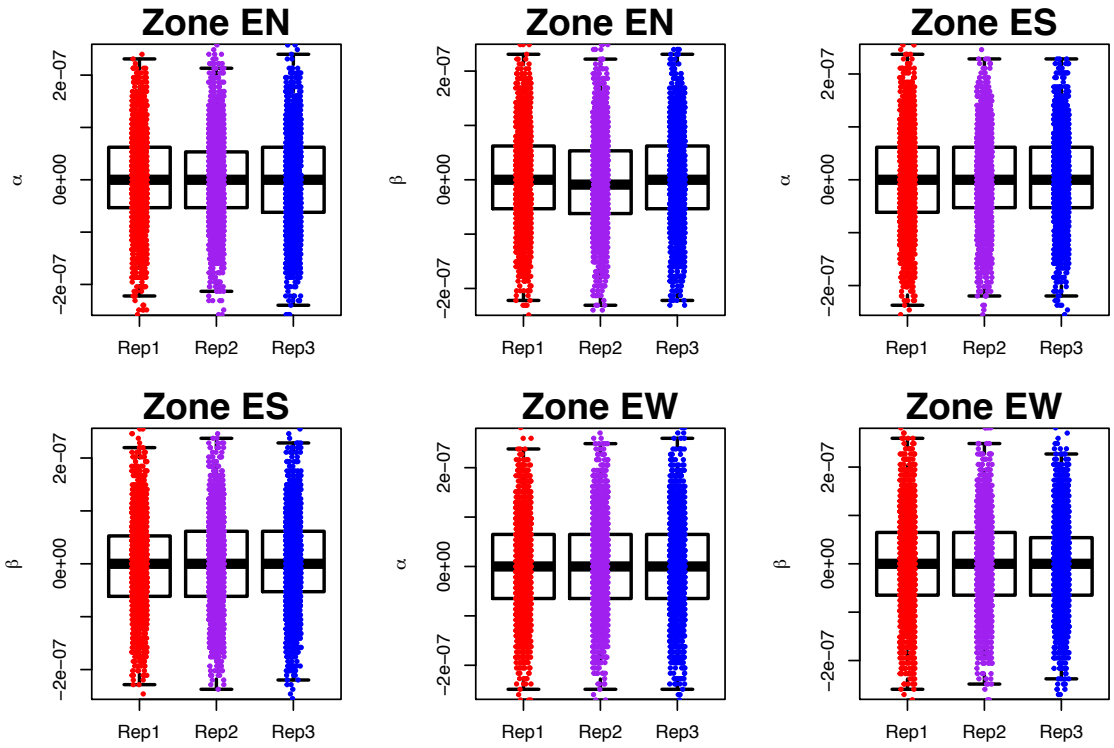


Figure S2.8. Boxplots of MCMC steps for α/β estimates for replicate BGC runs
All three runs converged on the same relatively narrow likelihood space, demonstrating reproducibility of the estimates.

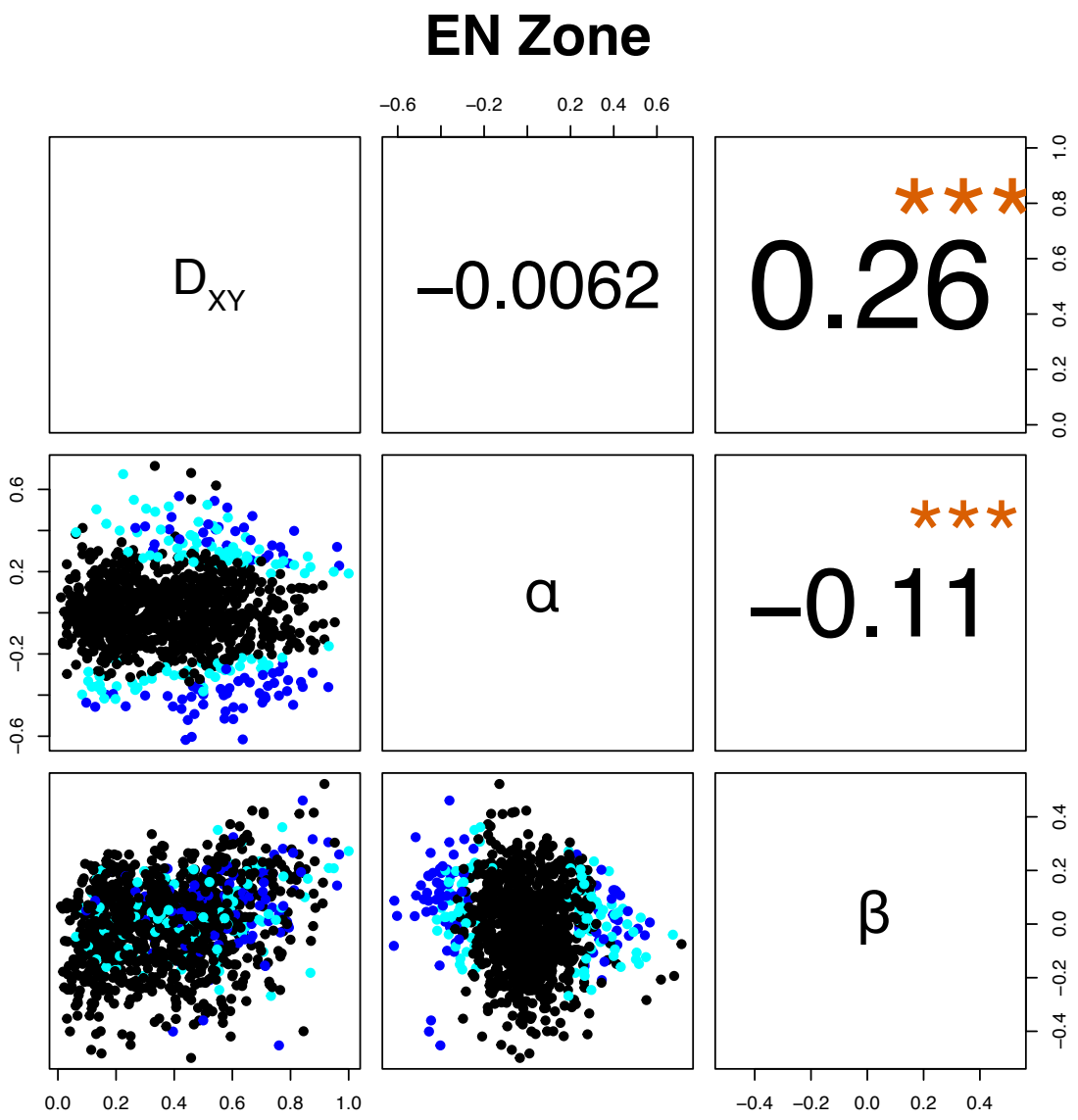


Figure S2.9. Scatterplots of D_{XY}, α, and β: EN contact zone
Pairwise scatterplots of D_{XY}, α, and β for the EN zone, showing correlation coefficients and levels of significance (*** = p < 0.001). Blue circles represent stringent α outliers (99% credible intervals exclude zero), while cyan circles represent regular outliers (95% credible intervals exclude zero).

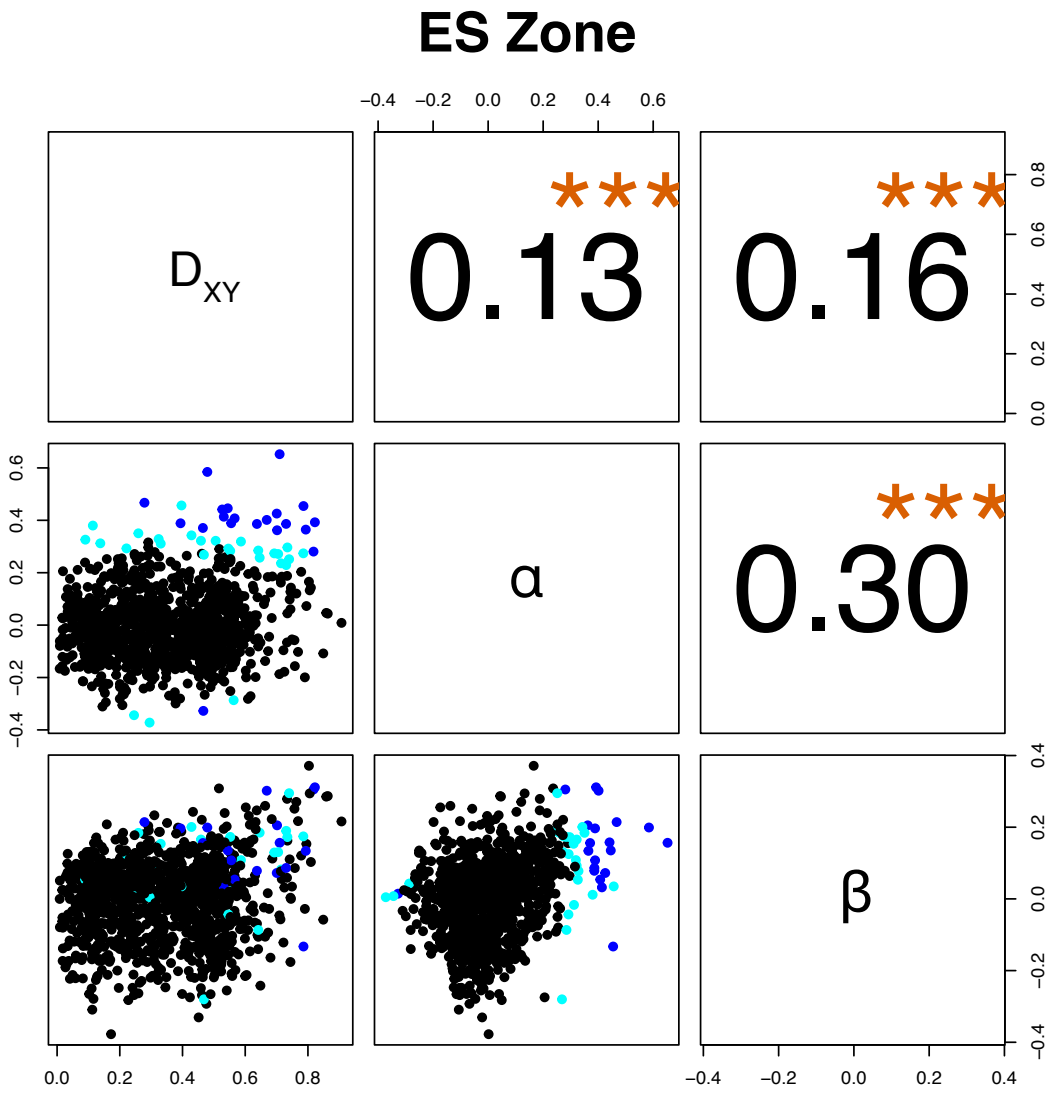


Figure S2.10. Scatterplots of D_{XY}, α, and β: ES contact zone
Pairwise scatterplots of D_{XY}, α, and β for the ES zone, showing correlation coefficients and levels of significance (*** = p < 0.001). Blue circles represent stringent α outliers (99% credible intervals exclude zero), while cyan circles represent regular outliers (95% credible intervals exclude zero).

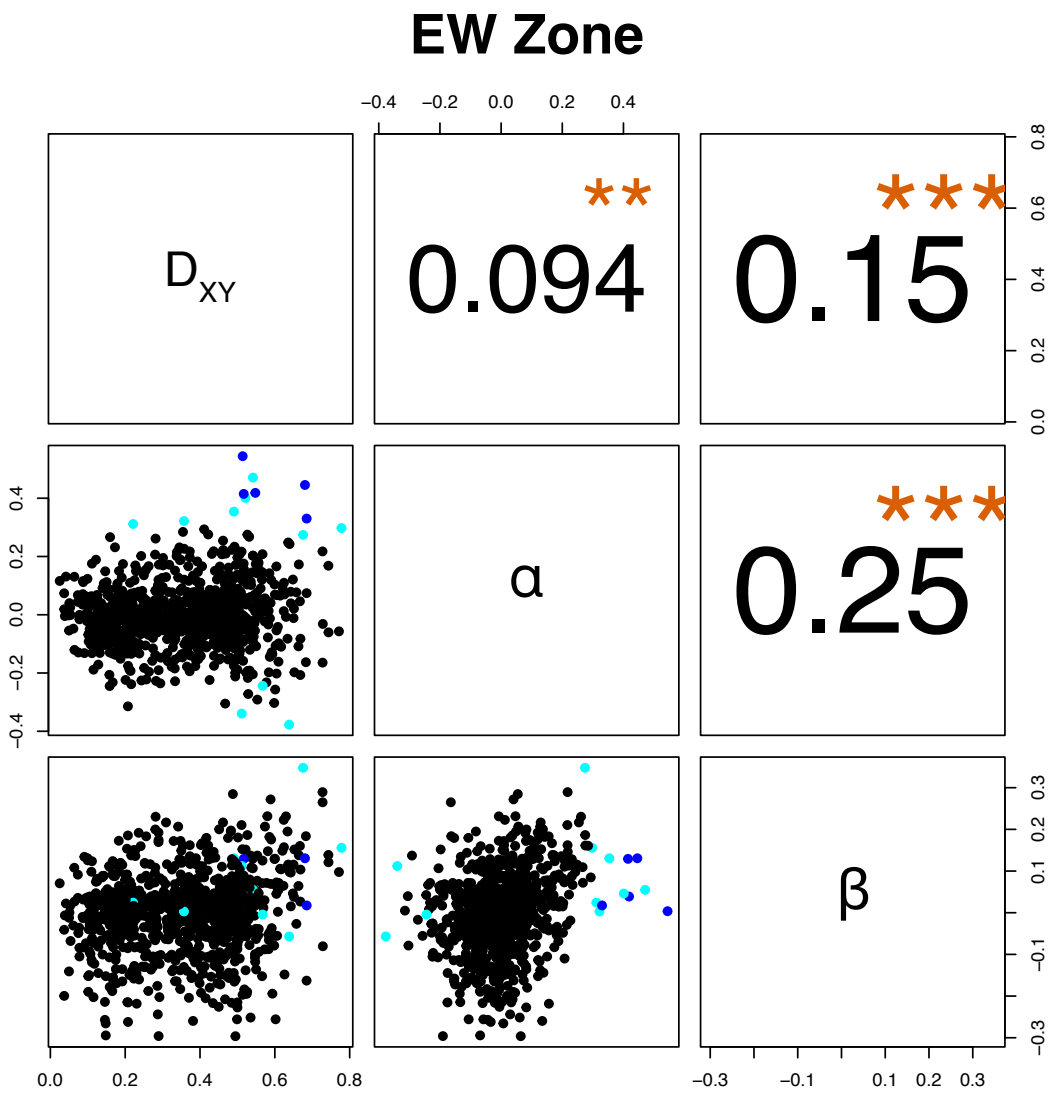


Figure S2.11. Scatterplots of D_{XY}, α, and β: EW contact zone
Pairwise scatterplots of D_{XY}, α, and β for the EW zone, showing correlation coefficients and levels of significance (** = p < 0.01, *** = p < 0.001). Blue circles represent stringent α outliers (99% credible intervals exclude zero), while cyan circles represent regular outliers (95% credible intervals exclude zero).

APPENDIX E: SUPPLEMENTARY TABLES/FIGURES FOR CHAPTER 3

Table S3.1. Vegetation resistance values used for building LCP models.

Vegetation Class	Resistance Value
Wet meadow (intermittently to seasonally flooded)	1
Wet meadow (semi-permanently to permanently flooded)	1
Permanent water	1
Willow	1
Dry meadow (shorthair sedge, annual grassland)	2
Quaking aspen	2
Permanent snowfield or glacier	3
Upland herbaceous	3
Conifers found with mesic understory (e.g. Sierra Lodgepole Pine, Whitebark Pine)	4
Conifers associated with Yosemite toad habitats (California Red Fir, Sierra Lodgepole Pine, Whitebark Pine)	5
Somewhat mesic conifers, no Yosemite toad association (Sugar Pine, White Fir, Western White Pine, Mountain Hemlock, Incense-cedar, Giant Sequoia)	5
Somewhat xeric conifers, no Yosemite toad association (Douglas-fir, Jeffrey Pine, Ponderosa Pine, Sierra Juniper, Limber Pine, Knobcone Pine, Singleleaf Pinyon Pine)	7
Hardwood (e.g. Oak, Mahogany, Cottonwood, Alder, Maple)	7
Shrubland or chaparral	7
Sparesely to non-vegetated	7
Urban	7
Rock outcrop (boulder, scree, talus, dome)	10

Table S3.2. List of All environmental variables used. Abbreviations are defined in the footnotes.

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Climate	At	PrecipDMmean	precip	Daymet	Inter-annual mean of precipitation	Mean at meadows
Climate	At	PrecipDMstd	-	Daymet	Inter-annual standard deviation of precipitation	Mean at meadows
Climate	Both	aet_av	aet_av	BCM 2014	Mean actual evapotranspiration	Mean at/bt sites
Climate	Both	aet_avch	-	BCM 2014	60-year change in mean actual evapotranspiration	Mean at/bt sites
Climate	Both	aet_cv	aet_cv	BCM 2014	CV of actual evapotranspiration	Mean at/bt sites
Climate	Both	aet_cvch	-	BCM 2014	60-year change in CV of actual evapotranspiration	Mean at/bt sites
Climate	Both	cwd_av	-	BCM 2014	Mean climatic water deficit	Mean at/bt sites
Climate	Both	cwd_avch	-	BCM 2014	60-year change in mean climatic water deficit	Mean at/bt sites
Climate	Both	pet_av	pet_av	BCM 2014	Mean potential evapotranspiration	Mean at/bt sites
Climate	Both	pet_avch	-	BCM 2014	60-year change in mean potential evapotranspiration	Mean at/bt sites
Climate	Both	pet_cv	pet_cv	BCM 2014	CV of potential evapotranspiration	Mean at/bt sites
Climate	Both	pet_cvch	petch	BCM 2014	60-year change in CV of potential evapotranspiration	Mean at/bt sites
Climate	Both	ppt_av	ppt_av	BCM 2014	Mean precipitation	Mean at/bt sites
Climate	Both	ppt_avch	pptch	BCM 2014	60-year change in mean precipitation	Mean at/bt sites
Climate	Both	ppt_cv	ppt_cv	BCM 2014	CV of precipitation	Mean at/bt sites
Climate	Both	ppt_cvch	-	BCM 2014	60-year change in CV of precipitation	Mean at/bt sites
Climate	Both	rch_av	rch_av	BCM 2014	Mean recharge	Mean at/bt sites
Climate	Both	rch_avch	-	BCM 2014	60-year change in mean recharge	Mean at/bt sites
Climate	Both	rch_cv	rch_cv	BCM 2014	CV of recharge	Mean at/bt sites
Climate	Both	rch_cvch	-	BCM 2014	60-year change in CV of recharge	Mean at/bt sites
Climate	Both	run_av	-	BCM 2014	Mean runoff	Mean at/bt sites
Climate	Both	run_avch	runch	BCM 2014	60-year change in mean runoff	Mean at/bt sites
Climate	Both	run_cv	run_cv	BCM 2014	CV of runoff	Mean at/bt sites
Climate	Both	run_cvch	-	BCM 2014	60-year change in CV of runoff	Mean at/bt sites
Climate	Both	snp_av	-	BCM 2014	Mean snowpack	Mean at/bt sites
Climate	Both	snp_avch	-	BCM 2014	60-year change in mean snowpack	Mean at/bt sites
Climate	Both	snp_cv	-	BCM 2014	CV of snowpack	Mean at/bt sites
Climate	Both	snp_cvch	-	BCM 2014	60-year change in CV of snowpack	Mean at/bt sites
Climate	Both	tmn_av	-	BCM 2014	Mean minimum temperature	Mean at/bt sites
Climate	Both	tmn_avch	-	BCM 2014	60-year change in mean minimum temperature	Mean at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Climate	Both	tmn_cv	tmn_cv	BCM 2014	CV of minimum temperature	Mean at/bt sites
Climate	Both	tmn_cvch	-	BCM 2014	60-year change in CV of minimum temperature	Mean at/bt sites
Climate	Both	tmx_av	-	BCM 2014	Mean maximum temperature	Mean at/bt sites
Climate	Both	tmx_avch	-	BCM 2014	60-year change in mean maximum temperature	Mean at/bt sites
Climate	Both	tmx_cv	-	BCM 2014	CV of maximum temperature	Mean at/bt sites
Climate	Both	tmx_cvch	-	BCM 2014	60-year change in CV of maximum temperature	Mean at/bt sites
Disturbance	At	FireDist	fired	Cal Fire	Distance to nearest fire perimeter	Value at sites
Disturbance	At	RoadDist	roadd	YNP	Distance to nearest road	Value at sites
Disturbance	At	TrailDist	-	YNP	Distance to trail	Value at sites
Disturbance	Between	FIRE_sum	-	Cal Fire	Number of recorded fire occurrences (per pixel)	Sum bt sites
Disturbance	Between	ROAD_sum	-	YNP	Road crossings (per pixel)	Sum bt sites
Disturbance	Between	ROADW_sum	roadw	YNP	Road crossings (per pixel, weighted by traffic level)	Sum bt sites
Disturbance	Between	TRAIL_sum	trail	YNP	Trail crossings (per pixel)	Sum bt sites
Disturbance	Between	TRAILW_sum	-	YNP	Trail crossings (per pixel, weighted by traffic level)	Sum bt sites
Meadow Network	At	ClustCoeff	-	Berlow et al. 2013	Cluster coefficient	Value at meadows
Meadow Network	At	NetworkSim	BP	Berlow et al. 2013	Network-boosted breeding probability	Value at meadows
Meadow Network	At	NormImpr	-	Berlow et al. 2013	NormImprov breeding probability	Value at meadows
Meadow Network	At	Prob100GLM	-	Berlow et al. 2013	100GLM breeding probability	Value at meadows
Meadow Network	At	HydroName	-	Cal Water	Second largest watershed unit	Value at meadows
Meadow Network	At	HydroUnit	-	Cal Water	Largest watershed unit	Value at meadows
Meadow Network	At	PlanWS	-	Cal Water	Fourth largest watershed unit	Value at meadows
Meadow Network	At	PlanWSacres	-	Cal Water	Area of of smallest watershed unit	Value at meadows
Meadow Network	At	SupPlanWS	-	Cal Water	Third largest watershed unit	Value at meadows
Meadow Network	At	Area	A	YNP LiDAR	Area of meadow	Value at meadows
Meadow Network	At	AreaBrdMdws1 km	-	YNP LiDAR	Area of breeding meadows within 1 km	Value at meadows
Meadow Network	At	ClsArea	-	YNP LiDAR	Area of closest meadow	Value at meadows
Meadow Network	At	ClsAvgSlp	-	YNP LiDAR	Average slope traveling to closest meadow	Value at meadows
Meadow Network	At	ClsBrdArea	-	YNP LiDAR	Area of closest breeding meadow	Value at meadows
Meadow Network	At	ClsBrdAvgSlp	-	YNP LiDAR	Average slope traveling to closest breeding meadow	Value at meadows
Meadow Network	At	ClsBrdDist	breed	YNP LiDAR	Distance to nearest breeding meadow	Value at meadows
Meadow Network	At	ClsBrdElevMn	-	YNP LiDAR	Mean elevation traveling to closest breeding meadow	Value at meadows
Meadow Network	At	ClsBrdElevRange	-	YNP LiDAR	Range in elevation traveling to closest breeding meadow	Value at meadows

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Meadow Network	At	ClsBrdMaxSlp	-	YNP LiDAR	Maximum slope traveling to closest breeding meadow	Value at meadows
Meadow Network	At	ClsBrdPerimeter	-	YNP LiDAR	Perimeter of closest breeding meadow	Value at meadows
Meadow Network	At	ClsBrdWetArea	-	YNP LiDAR	Wet area of closest breeding meadow	Value at meadows
Meadow Network	At	ClsBrdWetPrct	-	YNP LiDAR	Proportion wet area of closest breeding meadow	Value at meadows
Meadow Network	At	ClsDist	-	YNP LiDAR	Distance to closest meadow	Value at meadows
Meadow Network	At	ClsElevMn	-	YNP LiDAR	Mean elevation traveling to closest meadow	Value at meadows
Meadow Network	At	ClsElevRange	-	YNP LiDAR	Range in elevation traveling to closest meadow	Value at meadows
Meadow Network	At	ClsMaxSlp	-	YNP LiDAR	Maximum slope traveling to closest meadow	Value at meadows
Meadow Network	At	ClsMdwArea1 km	MA	YNP LiDAR	Area of meadows within 1 km	Value at meadows
Meadow Network	At	ClsNumMdws1 km	-	YNP LiDAR	Number of meadows within 1 km	Value at meadows
Meadow Network	At	ClsNumWetMdws1 km	-	YNP LiDAR	Number of wet meadows within 1 km	Value at meadows
Meadow Network	At	ClsPerimeter	-	YNP LiDAR	Perimeter of closest meadow	Value at meadows
Meadow Network	At	ClsPrcWet	-	YNP LiDAR	Proportion wet area of closest meadow	Value at meadows
Meadow Network	At	ClsWetArea	WAC	YNP LiDAR	Wet area of closest meadow	Value at meadows
Meadow Network	At	ClsWetArea1 km	-	YNP LiDAR	Area of meadow wet areas within 1 km	Value at meadows
Meadow Network	At	MinTravelTimeMinutes	-	YNP LiDAR	Estimated travel time to meadow from trailhead	Value at meadows
Meadow Network	At	NumBrdMdws1 km	-	YNP LiDAR	Number of breeding meadows within 1 km	Value at meadows
Meadow Network	At	Perimeter	P	YNP LiDAR	Length of meadow perimeter	Value at meadows
Meadow Network	At	WetAreaBrdMdws1 km	BWA	YNP LiDAR	Area of breeding meadow wet areas within 1 km	Value at meadows
Moisture/Vegetation	Both	MSAVI_AUG_MEAN_mean	-	LANDSAT 5	MSA Vegetation Index: August mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_AUG_MEAN_std	-	LANDSAT 5	MSA Vegetation Index: August mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_AUG_STD_mean	-	LANDSAT 5	MSA Vegetation Index: August SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_AUG_STD_std	-	LANDSAT 5	MSA Vegetation Index: August SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_JUL_MEAN_mean	-	LANDSAT 5	MSA Vegetation Index: July mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_JUL_MEAN_std	-	LANDSAT 5	MSA Vegetation Index: July mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_JUL_STD_mean	-	LANDSAT 5	MSA Vegetation Index: July SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_JUL_STD_std	-	LANDSAT 5	MSA Vegetation Index: July SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_JUN_MEAN_mean	-	LANDSAT 5	MSA Vegetation Index: June mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_JUN_MEAN_std	-	LANDSAT 5	MSA Vegetation Index: June mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_JUN_STD_mean	-	LANDSAT 5	MSA Vegetation Index: June SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_JUN_STD_std	-	LANDSAT 5	MSA Vegetation Index: June SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_MAY_MEAN_mean	-	LANDSAT 5	MSA Vegetation Index: May mean, interannual mean	Mean at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Moisture/Vegetation	Both	MSAVI_MAY_MEAN_std	-	LANDSAT 5	MSA Vegetation Index: May mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_MAY_STD_mean	-	LANDSAT 5	MSA Vegetation Index: May SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_MAY_STD_std	-	LANDSAT 5	MSA Vegetation Index: May SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_MEAN_MEAN_mean	-	LANDSAT 5	MSA Vegetation Index: May-Sept mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_MEAN_MEAN_std	-	LANDSAT 5	MSA Vegetation Index: May-Sept mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_MEAN_STD_mean	-	LANDSAT 5	MSA Vegetation Index: May-Sept mean, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_MEAN_STD_std	-	LANDSAT 5	MSA Vegetation Index: May-Sept mean, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_SEP_MEAN_mean	-	LANDSAT 5	MSA Vegetation Index: September mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_SEP_MEAN_std	-	LANDSAT 5	MSA Vegetation Index: September mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_SEP_STD_mean	-	LANDSAT 5	MSA Vegetation Index: September SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_SEP_STD_std	-	LANDSAT 5	MSA Vegetation Index: September SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_STD_MEAN_mean	-	LANDSAT 5	MSA Vegetation Index: May-Sept SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_STD_MEAN_std	-	LANDSAT 5	MSA Vegetation Index: May-Sept SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	MSAVI_STD_STD_mean	-	LANDSAT 5	MSA Vegetation Index: May-Sept SD, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	MSAVI_STD_STD_std	-	LANDSAT 5	MSA Vegetation Index: May-Sept SD, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NBR_AUG_MEAN_mean	-	LANDSAT 5	N Burn Ratio: August mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_AUG_MEAN_std	-	LANDSAT 5	N Burn Ratio: August mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_AUG_STD_mean	-	LANDSAT 5	N Burn Ratio: August SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_AUG_STD_std	-	LANDSAT 5	N Burn Ratio: August SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_JUL_MEAN_mean	-	LANDSAT 5	N Burn Ratio: July mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_JUL_MEAN_std	-	LANDSAT 5	N Burn Ratio: July mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_JUL_STD_mean	-	LANDSAT 5	N Burn Ratio: July SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_JUL_STD_std	-	LANDSAT 5	N Burn Ratio: July SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_JUN_MEAN_mean	-	LANDSAT 5	N Burn Ratio: June mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_JUN_MEAN_std	-	LANDSAT 5	N Burn Ratio: June mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_JUN_STD_mean	-	LANDSAT 5	N Burn Ratio: June SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_JUN_STD_std	-	LANDSAT 5	N Burn Ratio: June SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_MAY_MEAN_mean	-	LANDSAT 5	N Burn Ratio: May mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_MAY_MEAN_std	-	LANDSAT 5	N Burn Ratio: May mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_MAY_STD_mean	-	LANDSAT 5	N Burn Ratio: May SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_MAY_STD_std	-	LANDSAT 5	N Burn Ratio: May SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_MEAN_MEAN_mean	-	LANDSAT 5	N Burn Ratio: May-Sept mean, interannual mean	Mean at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Moisture/Vegetation	Both	NBR_MEAN_MEAN_std	-	LANDSAT 5	N Burn Ratio: May-Sept mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_MEAN_STD_mean	-	LANDSAT 5	N Burn Ratio: May-Sept mean, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NBR_MEAN_STD_std	-	LANDSAT 5	N Burn Ratio: May-Sept mean, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NBR_SEP_MEAN_mean	-	LANDSAT 5	N Burn Ratio: September mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_SEP_MEAN_std	-	LANDSAT 5	N Burn Ratio: September mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_SEP_STD_mean	-	LANDSAT 5	N Burn Ratio: September SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_SEP_STD_std	-	LANDSAT 5	N Burn Ratio: September SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_STD_MEAN_mean	-	LANDSAT 5	N Burn Ratio: May-Sept SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR_STD_MEAN_std	-	LANDSAT 5	N Burn Ratio: May-Sept SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR_STDSTD_mean	-	LANDSAT 5	N Burn Ratio: May-Sept SD, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NBR_STDSTD_std	-	LANDSAT 5	N Burn Ratio: May-Sept SD, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NBR2_AUG_MEAN_mean	-	LANDSAT 5	N Burn Ratio 2: August mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_AUG_MEAN_std	-	LANDSAT 5	N Burn Ratio 2: August mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_AUG_STD_mean	-	LANDSAT 5	N Burn Ratio 2: August SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_AUG_STD_std	-	LANDSAT 5	N Burn Ratio 2: August SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_JUL_MEAN_mean	-	LANDSAT 5	N Burn Ratio 2: July mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_JUL_MEAN_std	-	LANDSAT 5	N Burn Ratio 2: July mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_JUL_STD_mean	-	LANDSAT 5	N Burn Ratio 2: July SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_JUL_STD_std	-	LANDSAT 5	N Burn Ratio 2: July SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_JUN_MEAN_mean	-	LANDSAT 5	N Burn Ratio 2: June mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_JUN_MEAN_std	-	LANDSAT 5	N Burn Ratio 2: June mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_JUN_STD_mean	-	LANDSAT 5	N Burn Ratio 2: June SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_JUN_STD_std	-	LANDSAT 5	N Burn Ratio 2: June SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_MAY_MEAN_mean	-	LANDSAT 5	N Burn Ratio 2: May mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_MAY_MEAN_std	nbrm	LANDSAT 5	N Burn Ratio 2: May mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_MAY_STD_mean	-	LANDSAT 5	N Burn Ratio 2: May SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_MAY_STD_std	-	LANDSAT 5	N Burn Ratio 2: May SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_MEAN_MEAN_mean	-	LANDSAT 5	N Burn Ratio 2: May-Sept mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_MEAN_MEAN_std	-	LANDSAT 5	N Burn Ratio 2: May-Sept mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_MEAN_STD_mean	-	LANDSAT 5	N Burn Ratio 2: May-Sept mean, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_MEAN_STD_std	-	LANDSAT 5	N Burn Ratio 2: May-Sept mean, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NBR2_SEP_MEAN_mean	-	LANDSAT 5	N Burn Ratio 2: September mean, interannual mean	Mean at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Moisture/Vegetation	Both	NBR2_SEP_MEAN_std	-	LANDSAT 5	N Burn Ratio 2: September mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_SEP_STD_mean	-	LANDSAT 5	N Burn Ratio 2: September SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_SEP_STD_std	-	LANDSAT 5	N Burn Ratio 2: September SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_STD_MEAN_mean	-	LANDSAT 5	N Burn Ratio 2: May-Sept SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_STD_MEAN_std	-	LANDSAT 5	N Burn Ratio 2: May-Sept SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NBR2_STD_STD_mean	-	LANDSAT 5	N Burn Ratio 2: May-Sept SD, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NBR2_STD_STD_std	-	LANDSAT 5	N Burn Ratio 2: May-Sept SD, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NDMI_AUG_MEAN_mean	-	LANDSAT 5	ND Moisture Index: August mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_AUG_MEAN_std	-	LANDSAT 5	ND Moisture Index: August mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_AUG_STD_mean	-	LANDSAT 5	ND Moisture Index: August SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_AUG_STD_std	-	LANDSAT 5	ND Moisture Index: August SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_JUL_MEAN_mean	-	LANDSAT 5	ND Moisture Index: July mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_JUL_MEAN_std	-	LANDSAT 5	ND Moisture Index: July mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_JUL_STD_mean	-	LANDSAT 5	ND Moisture Index: July SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_JUL_STD_std	-	LANDSAT 5	ND Moisture Index: July SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_JUN_MEAN_mean	-	LANDSAT 5	ND Moisture Index: June mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_JUN_MEAN_std	-	LANDSAT 5	ND Moisture Index: June mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_JUN_STD_mean	-	LANDSAT 5	ND Moisture Index: June SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_JUN_STD_std	-	LANDSAT 5	ND Moisture Index: June SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_MAY_MEAN_mean	-	LANDSAT 5	ND Moisture Index: May mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_MAY_MEAN_std	-	LANDSAT 5	ND Moisture Index: May mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_MAY_STD_mean	-	LANDSAT 5	ND Moisture Index: May SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_MAY_STD_std	-	LANDSAT 5	ND Moisture Index: May SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_MEAN_MEAN_mean	-	LANDSAT 5	ND Moisture Index: May-Sept mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_MEAN_MEAN_std	ndmi	LANDSAT 5	ND Moisture Index: May-Sept mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_MEAN_STD_mean	-	LANDSAT 5	ND Moisture Index: May-Sept mean, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_MEAN_STD_std	-	LANDSAT 5	ND Moisture Index: May-Sept mean, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NDMI_SEP_MEAN_mean	-	LANDSAT 5	ND Moisture Index: September mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_SEP_MEAN_std	-	LANDSAT 5	ND Moisture Index: September mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_SEP_STD_mean	-	LANDSAT 5	ND Moisture Index: September SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_SEP_STD_std	-	LANDSAT 5	ND Moisture Index: September SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_STD_MEAN_mean	-	LANDSAT 5	ND Moisture Index: May-Sept SD, interannual mean	Mean at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Moisture/Vegetation	Both	NDMI_STD_MEAN_std	-	LANDSAT 5	ND Moisture Index: May-Sept SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDMI_STD_STD_mean	-	LANDSAT 5	ND Moisture Index: May-Sept SD, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NDMI_STD_STD_std	-	LANDSAT 5	ND Moisture Index: May-Sept SD, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NDVI_AUG_MEAN_mean	-	LANDSAT 5	ND Vegetation Index: August mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_AUG_MEAN_std	-	LANDSAT 5	ND Vegetation Index: August mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_AUG_STD_mean	-	LANDSAT 5	ND Vegetation Index: August SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_AUG_STD_std	-	LANDSAT 5	ND Vegetation Index: August SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_JUL_MEAN_mean	-	LANDSAT 5	ND Vegetation Index: July mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_JUL_MEAN_std	-	LANDSAT 5	ND Vegetation Index: July mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_JUL_STD_mean	-	LANDSAT 5	ND Vegetation Index: July SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_JUL_STD_std	-	LANDSAT 5	ND Vegetation Index: July SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_JUN_MEAN_mean	-	LANDSAT 5	ND Vegetation Index: June mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_JUN_MEAN_std	-	LANDSAT 5	ND Vegetation Index: June mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_JUN_STD_mean	-	LANDSAT 5	ND Vegetation Index: June SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_JUN_STD_std	-	LANDSAT 5	ND Vegetation Index: June SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_MAY_MEAN_mean	-	LANDSAT 5	ND Vegetation Index: May mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_MAY_MEAN_std	-	LANDSAT 5	ND Vegetation Index: May mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_MAY_STD_mean	-	LANDSAT 5	ND Vegetation Index: May SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_MAY_STD_std	-	LANDSAT 5	ND Vegetation Index: May SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_MEAN_MEAN_mean	-	LANDSAT 5	ND Vegetation Index: May-Sept mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_MEAN_MEAN_std	-	LANDSAT 5	ND Vegetation Index: May-Sept mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_MEAN_STD_mean	-	LANDSAT 5	ND Vegetation Index: May-Sept mean, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_MEAN_STD_std	-	LANDSAT 5	ND Vegetation Index: May-Sept mean, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	NDVI_SEP_MEAN_mean	-	LANDSAT 5	ND Vegetation Index: September mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_SEP_MEAN_std	-	LANDSAT 5	ND Vegetation Index: September mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_SEP_STD_mean	-	LANDSAT 5	ND Vegetation Index: September SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_SEP_STD_std	-	LANDSAT 5	ND Vegetation Index: September SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_STD_MEAN_mean	-	LANDSAT 5	ND Vegetation Index: May-Sept SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_STD_MEAN_std	-	LANDSAT 5	ND Vegetation Index: May-Sept SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	NDVI_STD_STD_mean	-	LANDSAT 5	ND Vegetation Index: May-Sept SD, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	NDVI_STD_STD_std	-	LANDSAT 5	ND Vegetation Index: May-Sept SD, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	SAVI_AUG_MEAN_mean	-	LANDSAT 5	SA Vegetation Index: August mean, interannual mean	Mean at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Moisture/Vegetation	Both	SAVI_AUG_MEAN_std	-	LANDSAT 5	SA Vegetation Index: August mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_AUG_STD_mean	-	LANDSAT 5	SA Vegetation Index: August SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_AUG_STD_std	-	LANDSAT 5	SA Vegetation Index: August SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_JUL_MEAN_mean	-	LANDSAT 5	SA Vegetation Index: July mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_JUL_MEAN_std	-	LANDSAT 5	SA Vegetation Index: July mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_JUL_STD_mean	-	LANDSAT 5	SA Vegetation Index: July SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_JUL_STD_std	-	LANDSAT 5	SA Vegetation Index: July SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_JUN_MEAN_mean	-	LANDSAT 5	SA Vegetation Index: June mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_JUN_MEAN_std	-	LANDSAT 5	SA Vegetation Index: June mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_JUN_STD_mean	-	LANDSAT 5	SA Vegetation Index: June SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_JUN_STD_std	-	LANDSAT 5	SA Vegetation Index: June SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_MAY_MEAN_mean	-	LANDSAT 5	SA Vegetation Index: May mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_MAY_MEAN_std	-	LANDSAT 5	SA Vegetation Index: May mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_MAY_STD_mean	-	LANDSAT 5	SA Vegetation Index: May SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_MAY_STD_std	-	LANDSAT 5	SA Vegetation Index: May SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_MEAN_MEAN_mean	-	LANDSAT 5	SA Vegetation Index: May-Sept mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_MEAN_MEAN_std	-	LANDSAT 5	SA Vegetation Index: May-Sept mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_MEAN_STD_mean	-	LANDSAT 5	SA Vegetation Index: May-Sept mean, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_MEAN_STD_std	-	LANDSAT 5	SA Vegetation Index: May-Sept mean, interannual SD	SD at/bt sites
Moisture/Vegetation	Both	SAVI_SEP_MEAN_mean	-	LANDSAT 5	SA Vegetation Index: September mean, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_SEP_MEAN_std	-	LANDSAT 5	SA Vegetation Index: September mean, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_SEP_STD_mean	-	LANDSAT 5	SA Vegetation Index: September SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_SEP_STD_std	-	LANDSAT 5	SA Vegetation Index: September SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_STD_MEAN_mean	-	LANDSAT 5	SA Vegetation Index: May-Sept SD, interannual mean	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_STD_MEAN_std	-	LANDSAT 5	SA Vegetation Index: May-Sept SD, interannual mean	SD at/bt sites
Moisture/Vegetation	Both	SAVI_STD_STD_mean	-	LANDSAT 5	SA Vegetation Index: May-Sept SD, interannual SD	Mean at/bt sites
Moisture/Vegetation	Both	SAVI_STD_STD_std	-	LANDSAT 5	SA Vegetation Index: May-Sept SD, interannual SD	SD at/bt sites
Snowmelt	At	MdwMltDateMean	-	MODIS	Meadow meltoff date	Mean at meadows
Snowmelt	At	MdwMltDateSD	meltsd	MODIS	Meadow meltoff date	SD at meadows
Snowmelt	At	MdwP50Mean	-	MODIS	Proportion days with >50% snow cover	Mean at meadows
Snowmelt	At	MdwP50SD	p50sd	MODIS	Proportion days with >50% snow cover	SD at meadows
Topography	At	pctMdwEAST	-	SRTM	Proportion meadow area facing east	Value at meadows

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Topography	At	pctMdwNOREA	-	SRTM	Proportion meadow area facing northeast	Value at meadows
Topography	At	pctMdwNORTH	-	SRTM	Proportion meadow area facing north	Value at meadows
Topography	At	pctMdwNORWE	-	SRTM	Proportion meadow area facing northwest	Value at meadows
Topography	At	pctMdwSOUEA	-	SRTM	Proportion meadow area facing southeast	Value at meadows
Topography	At	pctMdwSOUTH	-	SRTM	Proportion meadow area facing south	Value at meadows
Topography	At	pctMdwSOUWE	-	SRTM	Proportion meadow area facing southwest	Value at meadows
Topography	At	pctMdwWEST	-	SRTM	Proportion meadow area facing west	Value at meadows
Topography	Both	CTI_mean	-	SRTM	Compound topographic index	Mean at/bt sites
Topography	Both	CTI_std	-	SRTM	Compound topographic index	SD at/bt sites
Topography	Both	CTI_sum	-	SRTM	Compound topographic index	Sum at/bt sites
Topography	Both	CURVE_mean	-	SRTM	Curvature	Mean at/bt sites
Topography	Both	CURVE_std	-	SRTM	Curvature	SD at/bt sites
Topography	Both	CURVE_sum	-	SRTM	Curvature	Sum at/bt sites
Topography	Both	DEM_mean	dem	SRTM	Elevation	Mean at/bt sites
Topography	Both	DEM_std	ESD	SRTM	Elevation	SD at/bt sites
Topography	Both	DEM_sum	-	SRTM	Elevation	Sum at/bt sites
Topography	Both	HLOAD_mean	-	SRTM	Heat load index	Mean at/bt sites
Topography	Both	HLOAD_std	-	SRTM	Heat load index	SD at/bt sites
Topography	Both	HLOAD_sum	-	SRTM	Heat load index	Sum at/bt sites
Topography	Both	IMPEDANCE_sum	-	SRTM	Slope impedance	Sum at/bt sites
Topography	Both	pctEAST	-	SRTM	Proportion facing east	Value at/bt sites
Topography	Both	pctNEAST	-	SRTM	Proportion facing northeast	Value at/bt sites
Topography	Both	pctNORTH	-	SRTM	Proportion facing north	Value at/bt sites
Topography	Both	pctNWEST	-	SRTM	Proportion facing northwest	Value at/bt sites
Topography	Both	pctSEAST	-	SRTM	Proportion facing southeast	Value at/bt sites
Topography	Both	pctSOUTH	-	SRTM	Proportion facing south	Value at/bt sites
Topography	Both	pctSWEST	-	SRTM	Proportion facing southwest	Value at/bt sites
Topography	Both	pctWEST	-	SRTM	Proportion facing west	Value at/bt sites
Topography	Both	RIDGE_sum	-	SRTM	Ridgeline (per pixel)	Sum at/bt sites
Topography	Both	ROUGH_mean	R	SRTM	Roughness index	Mean at/bt sites
Topography	Both	ROUGH_std	-	SRTM	Roughness index	SD at/bt sites
Topography	Both	ROUGH_sum	-	SRTM	Roughness index	Sum at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Topography	Both	SAR_mean	-	SRTM	Surface to area ratio	Mean at/bt sites
Topography	Both	SAR_std	-	SRTM	Surface to area ratio	SD at/bt sites
Topography	Both	SAR_sum	-	SRTM	Surface to area ratio	Sum at/bt sites
Topography	Both	SLOPE_mean	S	SRTM	Slope	Mean at/bt sites
Topography	Both	SLOPE_std	-	SRTM	Slope	SD at/bt sites
Topography	Both	SLOPE_sum	slope	SRTM	Slope	Sum at/bt sites
Topography	Both	SLOPE2ND_mean	-	SRTM	Slope second derivative	Mean at/bt sites
Topography	Both	SLOPE2ND_std	slope2	SRTM	Slope second derivative	SD at/bt sites
Topography	Both	SLOPE2ND_sum	-	SRTM	Slope second derivative	Sum at/bt sites
Topography	Both	SPI_mean	SPI	SRTM	Slope position index	Mean at/bt sites
Topography	Both	SPI_std	spi	SRTM	Slope position index	SD at/bt sites
Topography	Both	SPI_sum	-	SRTM	Slope position index	Sum at/bt sites
Topography	Both	SRR_mean	-	SRTM	Surface relief ratio	Mean at/bt sites
Topography	Both	SRR_std	-	SRTM	Surface relief ratio	SD at/bt sites
Topography	Both	SRR_sum	-	SRTM	Surface relief ratio	Sum at/bt sites
Topography	Both	TRASP_mean	-	SRTM	Radiation index	Mean at/bt sites
Topography	Both	TRASP_std	-	SRTM	Radiation index	SD at/bt sites
Topography	Both	TRASP_sum	-	SRTM	Radiation index	Sum at/bt sites
Vegetation	At	LakeDist	-	YNP LiDAR	Distance to lake	Value at sites
Vegetation	At	mdwpctIntFlood	-	YNP LiDAR	Proportion transiently flooded meadow	Value at sites
Vegetation	At	mdwpctSemiPermFlood	pflood	YNP LiDAR	Proportion semi- to permanently flooded meadow vegetation	Value at sites
Vegetation	At	mdwpctShSedge	-	YNP LiDAR	Proportion short-hair sedge in meadow	Value at sites
Vegetation	At	mdwpctWillow	-	YNP LiDAR	Proportion willow in meadow	Value at sites
Vegetation	At	StreamDist	-	YNP LiDAR	Distance to stream	Value at sites
Vegetation	Between	STREAMW_sum	-	YNP LiDAR	Stream crossings (per pixel, weighted by magnitude)	Value bt sites
Vegetation	Both	pctCONHERB	-	YNP LiDAR	Proportion conifers and herbaceous	Value at/bt sites
Vegetation	Both	pctCONIFER	pcon	YNP LiDAR	Proportion coniferous vegetation	Value at/bt sites
Vegetation	Both	pctCONSHRUB	-	YNP LiDAR	Proportion conifers and shrub	Value at/bt sites
Vegetation	Both	pctCONWOOD	-	YNP LiDAR	Proportion conifers and hardwood	Value at/bt sites
Vegetation	Both	pctDMEADOW	-	YNP LiDAR	Proportion dry meadow	Value at/bt sites
Vegetation	Both	pctHARDWOOD	-	YNP LiDAR	Proportion hardwood	Value at/bt sites
Vegetation	Both	pctRIPARIAN	-	YNP LiDAR	Proportion riparian	Value at/bt sites

Category	At or Between Sites	Name	Table Abbreviation	Source	Definition	GIS Summary
Vegetation	Both	pctROCK	-	YNP LiDAR	Proportion rock	Value at/bt sites
Vegetation	Both	pctSHRUB	-	YNP LiDAR	Proportion chaparral	Value at/bt sites
Vegetation	Both	pctSNOW	-	YNP LiDAR	Proportion permanent snow	Value at/bt sites
Vegetation	Both	pctSTREAM	-	YNP LiDAR	Proportion stream	Value at/bt sites
Vegetation	Both	pctWATER	pwat	YNP LiDAR	Proportion standing water, permanent pool, or stream	Value at/bt sites
Vegetation	Both	pctWMEADOW	-	YNP LiDAR	Proportion wet meadow	Value at/bt sites

* bt = between, CV = coefficient of variation, SD = standard deviation, SA = soil adjusted, MSA = modified soil adjusted, N = normalized, ND = normalized difference, 100GLM = 100-subsample general linear model (see Berlow et al. 2013), BCM = CA Basin Characterization Model, YNP = Yosemite National Park

Table S3.3. Lineage-specific variable coefficients from gravity models. Terms in bold have random effects by lineage.

30 km LCP Model

Lineage	**Intercept**	D	**BT(slope)**	**BT(dem)**	BT(runch)	**BT(ndmi)**	BT(trail)	BT(roadw)	**BT(pcon)**	BT(pwat)	BT(nbrm)	AT(pflood)
North	**0.25**	0.16	**-0.56**	**0.22**	-0.27	**-0.06**	-0.12	-0.07	**-0.33**	-0.04	-0.08	-0.12
East	**-0.07**	0.16	**-0.34**	**0.05**	-0.27	**-0.11**	-0.12	-0.07	**-0.15**	-0.04	-0.08	-0.12
South	**0.37**	0.16	**-0.61**	**0.37**	-0.27	**0.05**	-0.12	-0.07	**-0.15**	-0.04	-0.08	-0.12
West	**-0.20**	0.16	**-0.51**	**0.36**	-0.27	**0.10**	-0.12	-0.07	**0.19**	-0.04	-0.08	-0.12

20 km LCP Model

Lineage	**Intercept**	D	BT(slope)	**BT(dem)**	BT(runch)	**BT(roadw)**	**BT(ndmi)**	BT(trail)	**BT(pwat)**	**BT(pcon)**	BT(nbrm)	AT(pflood)	AT(pptch)
North	**0.12**	0.22	-0.45	**0.42**	-0.27	**-0.06**	**-0.27**	-0.09	**0.05**	**-0.19**	-3.4×10^{-4}	-0.15	0.04
East	**-0.19**	0.22	-0.45	**0.08**	-0.27	**-0.07**	**-0.06**	-0.09	**-0.12**	**-0.36**	-3.4×10^{-4}	-0.15	0.04
South	**-0.06**	0.22	-0.45	**0.35**	-0.27	**-0.21**	**-0.04**	-0.09	**-0.19**	**0.18**	-3.4×10^{-4}	-0.15	0.04
West	**-0.51**	0.22	-0.45	**0.52**	-0.27	**-0.24**	**0.03**	-0.09	**-0.30**	**0.53**	-3.4×10^{-4}	-0.15	0.04

10 km LCP Model

Lineage	**Intercept**	D	**BT(slope)**	AT(precip)	BT(runch)	BT(pwat)	BT(trail)	BT(roadw)	**BT(dem)**	AT(breed)
North	**0.54**	0.12	**-0.73**	-0.60	-0.31	-0.08	-0.09	-0.07	**0.55**	-0.13
East	**-0.06**	0.12	**-0.37**	-0.60	-0.31	-0.08	-0.09	-0.07	**-0.36**	-0.13
South	**-0.17**	0.12	**-0.31**	-0.60	-0.31	-0.08	-0.09	-0.07	**-0.54**	-0.13
West	**0.13**	0.12	**-0.48**	-0.60	-0.31	-0.08	-0.09	-0.07	**-0.09**	-0.13

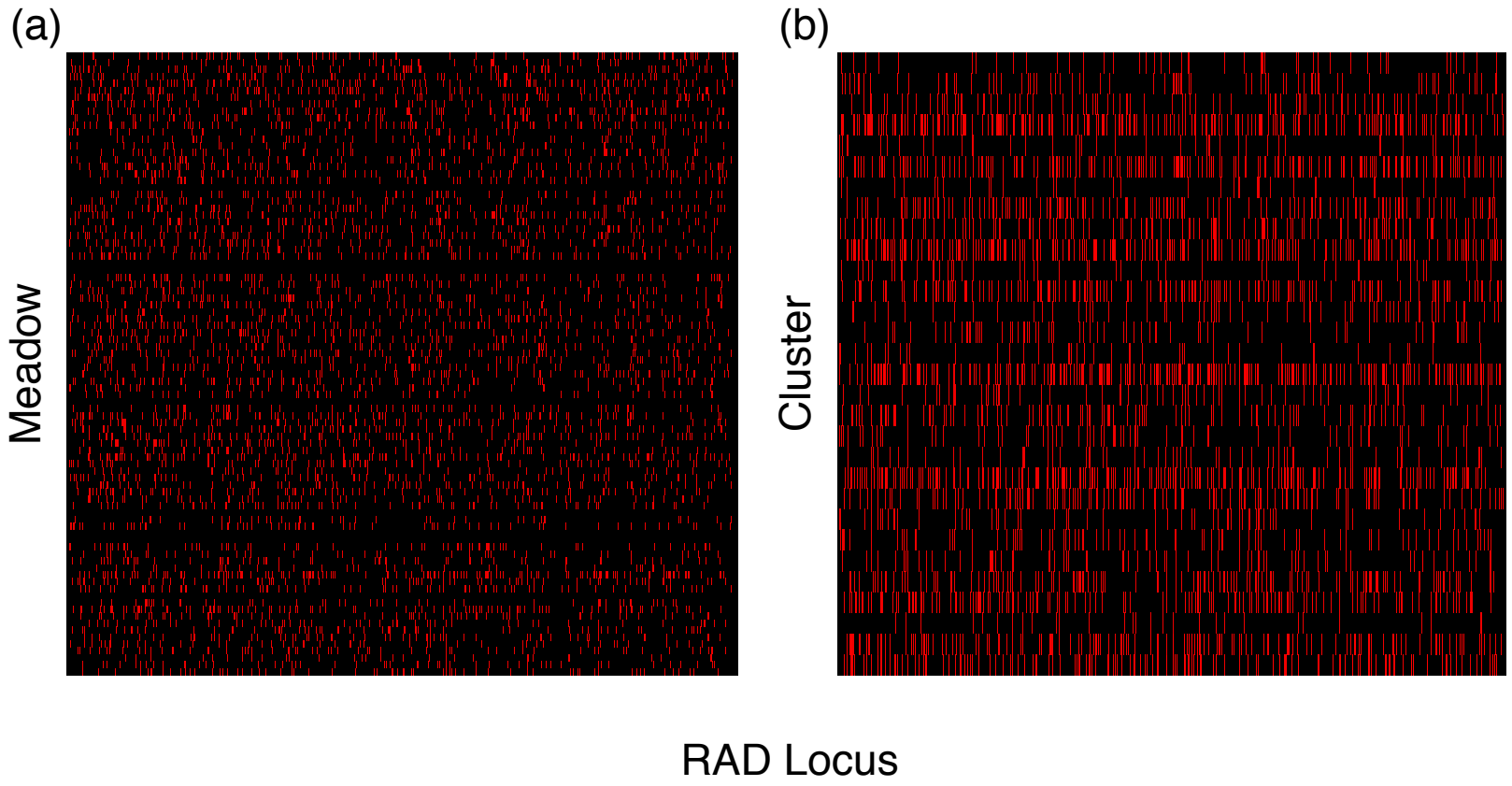


Figure S3.1. Hardy-Weinberg deviations
Rectangles in red represent loci that significantly deviate from Hardy-Weinberg genotype frequencies, when structured by (a) meadows (2.48% non-conformity), and (b) clusters (4.86% non-conformity).

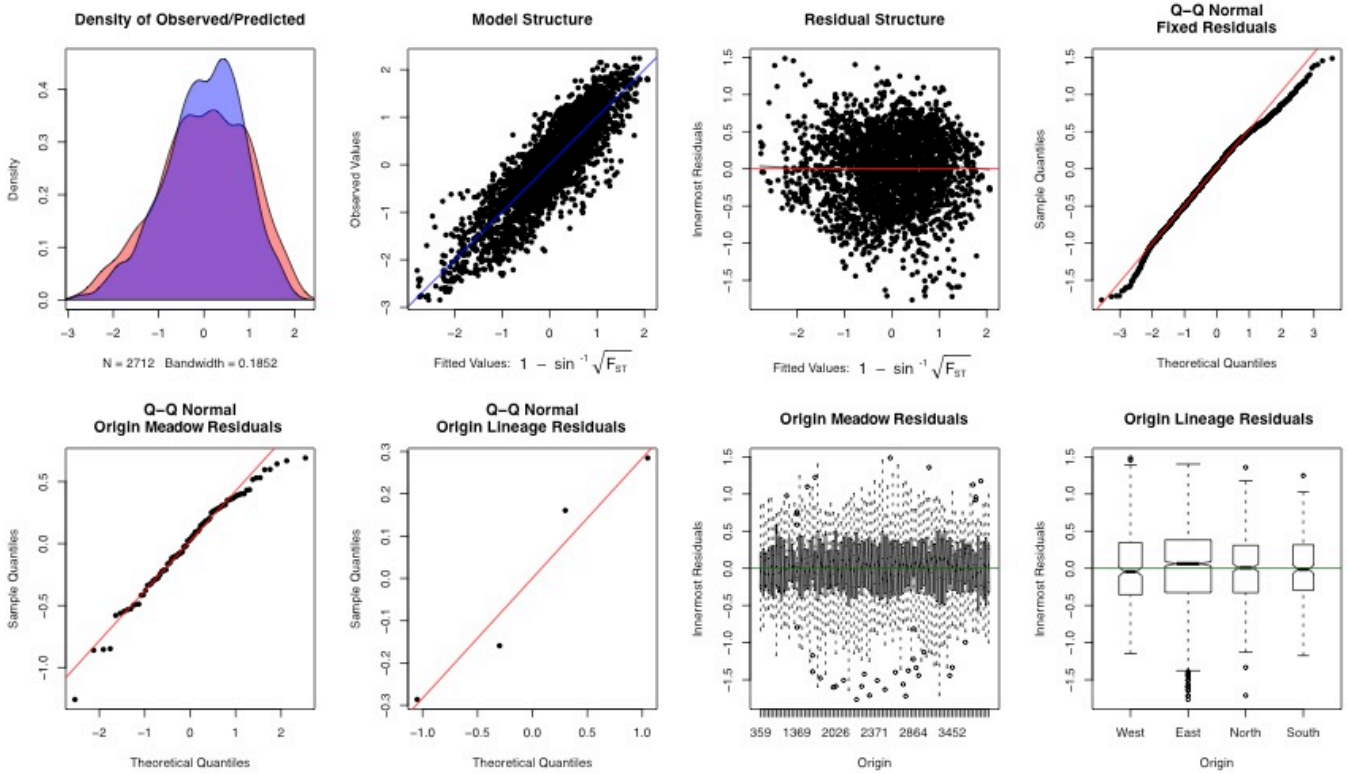


Figure S3.2. Gravity model fit statistics, 30 km cutoff
Example model performance (30 km model). (a) Distribution of observed (red) and predicted (blue) values. (b) Fit of observed values to predicted values on regression line. (c) Deviation of residuals from central predicted values. (d) Q-Q plot of fixed residual quantiles against normally distributed quantiles. (e) Q-Q plot of random intercept residual quantiles against normally distributed quantiles. (f) Q-Q plot of random slope residual quantiles against normally distributed quantiles. (g) Deviation of population residuals from central predicted values, by population. (h) Deviation of lineage residuals from central predicted values, by lineage.

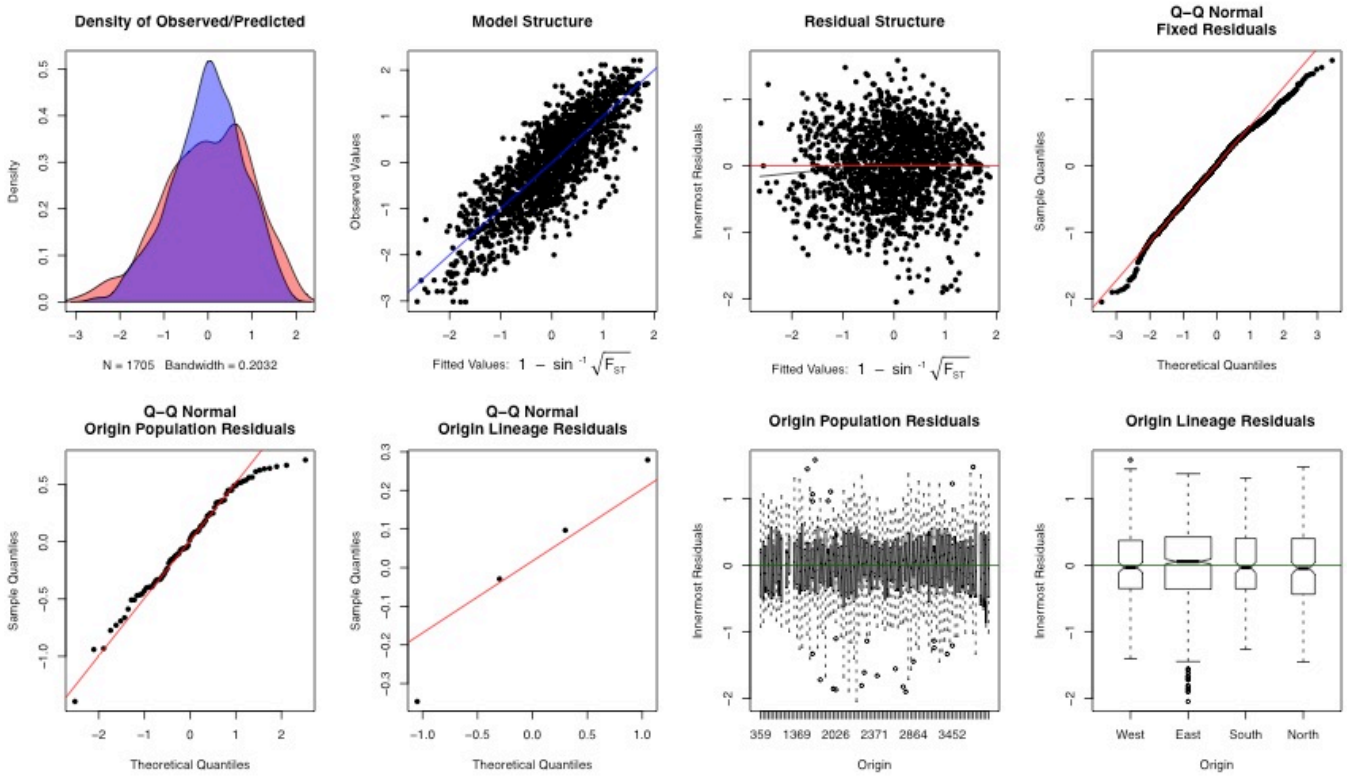


Figure S3.3. Gravity model fit statistics, 20 km cutoff
Example model performance (20 km model). (a) Distribution of observed (red) and predicted (blue) values. (b) Fit of observed values to predicted values on regression line. (c) Deviation of residuals from central predicted values. (d) Q-Q plot of fixed residual quantiles against normally distributed quantiles. (e) Q-Q plot of random intercept residual quantiles against normally distributed quantiles. (f) Q-Q plot of random slope residual quantiles against normally distributed quantiles. (g) Deviation of population residuals from central predicted values, by population. (h) Deviation of lineage residuals from central predicted values, by lineage.

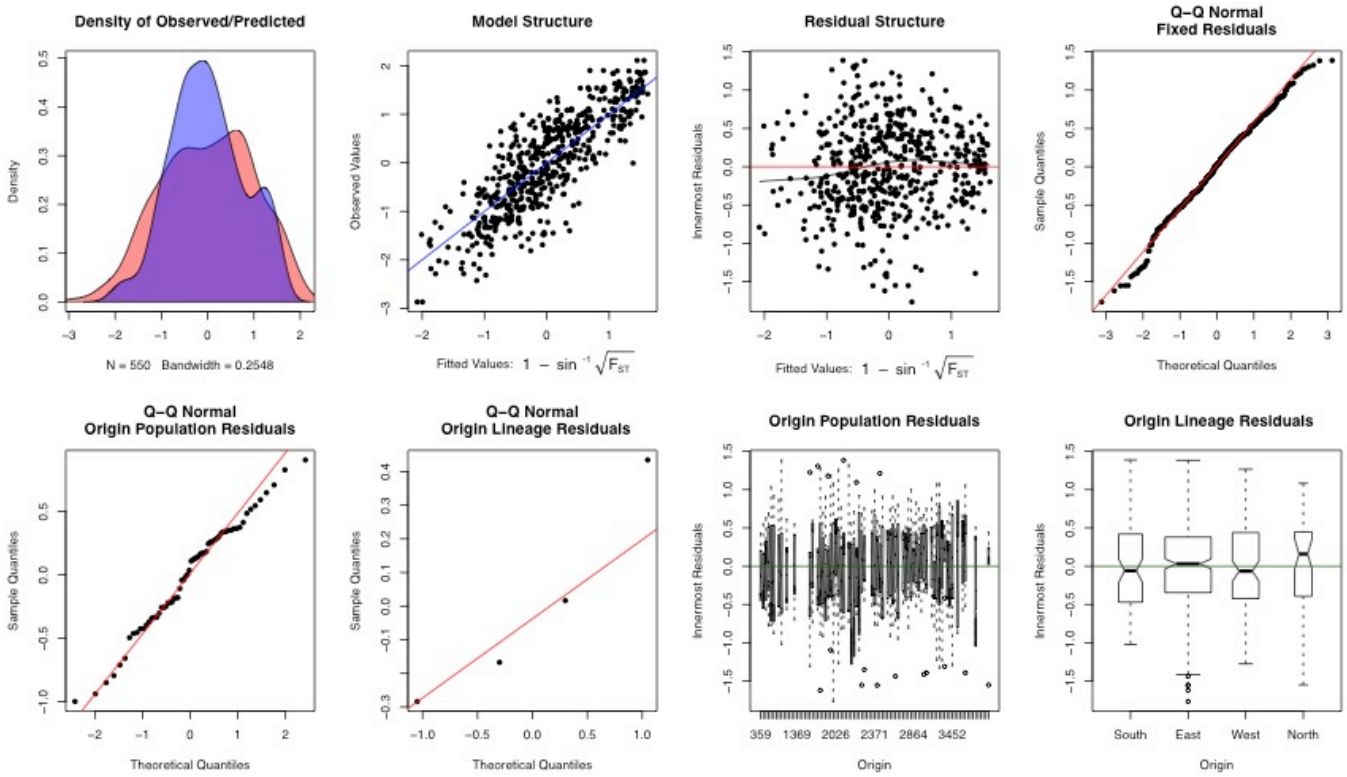


Figure S3.4. Gravity model fit statistics, 10 km cutoff
Example model performance (10 km model). (a) Distribution of observed (red) and predicted (blue) values. (b) Fit of observed values to predicted values on regression line. (c) Deviation of residuals from central predicted values. (d) Q-Q plot of fixed residual quantiles against normally distributed quantiles. (e) Q-Q plot of random intercept residual quantiles against normally distributed quantiles. (f) Q-Q plot of random slope residual quantiles against normally distributed quantiles. (g) Deviation of population residuals from central predicted values, by population. (h) Deviation of lineage residuals from central predicted values, by lineage.

APPENDIX F: SUPPLEMENTARY TABLES/FIGURES FOR CHAPTER 4

Table S4.1. Loadings for principal components analysis of climatic data. Darker cells have higher absolute variable loadings.

Variable	PC1	PC2	PC3	PC4	PC5
bio1	0.22	0.04	-0.06	0.02	0.10
bio2	0.19	0.00	-0.15	-0.35	-0.03
bio3	0.20	-0.02	-0.12	-0.23	-0.07
bio4	0.16	0.01	-0.20	-0.48	0.05
bio5	0.22	0.03	-0.08	-0.07	0.08
bio6	0.22	0.04	-0.03	0.11	0.11
bio7	0.18	0.00	-0.17	-0.44	-0.02
bio8	0.22	0.03	-0.09	-0.07	0.03
bio9	0.22	0.03	-0.07	-0.03	0.10
bio10	0.22	0.03	-0.06	-0.02	0.10
bio11	0.22	0.04	-0.04	0.07	0.10
bio12	0.16	0.19	0.23	-0.02	-0.42
bio13	0.18	0.17	0.20	0.03	-0.33
bio14	-0.20	-0.03	-0.02	-0.27	-0.17
bio15	0.22	0.09	0.06	0.09	-0.01
bio16	0.18	0.18	0.19	-0.01	-0.35
bio17	-0.21	-0.02	0.03	-0.19	-0.20
bio18	-0.22	-0.02	0.07	-0.06	-0.17
bio19	0.19	0.16	0.19	0.07	-0.29
aprpck_ave	-0.15	0.26	0.24	-0.20	0.24
cwd_ave	0.20	0.02	-0.12	0.20	-0.02
cwdsum_ave	0.20	0.02	-0.14	0.20	-0.04
rch_ave	0.16	-0.18	0.37	-0.09	0.07
rchsum_ave	0.06	-0.32	0.36	-0.24	0.16
run_ave	-0.10	0.40	-0.09	-0.05	0.07
runsum_ave	-0.16	0.30	0.05	-0.11	0.11
aprpck_std	-0.02	0.36	0.36	-0.13	0.27
cwd_std	0.16	0.14	0.14	0.17	0.39
rch_std	0.14	-0.23	0.40	-0.08	0.08
run_std	-0.05	0.44	-0.10	-0.05	0.05

Table S4.2. **Climatic correlates of outlier loci.** Variables indicated by 'x' had either Pearson or Spearman correlation coefficients of >|0.1|, Bayes Factor of >0.2 compared to the model based on population structure ($X^T X$) alone. All outliers with any significant correlation with a MEM were removed.

Locus	long	lat	elev	PC1	PC2	PC3	PC4	PC5	MEM1	MEM2	MEM3	MEM4
l196					x							
l2794							x					
l18477					x							
l20128				x								
l20277					x							
l20524				x								
l22381								x				
l27196							x					
l28646					x							
l28769				x								
l30836							x					
l34048							x					
l38163			x	x								
l38427			x	x								
l43123			x		x							
l43270			x	x								
l44224			x	x								
l46715			x	x								
l48794								x				
l49074				x								
l49136				x								
l54027	x				x							
l54854								x				
l55572								x				
l58960					x							
l61882			x	x								
l63530			x	x								
l64503							x					
l64511			x	x								
l64697			x	x								
l66687					x							
l67875					x							
l68654						x						
l70715							x					
l71969								x				

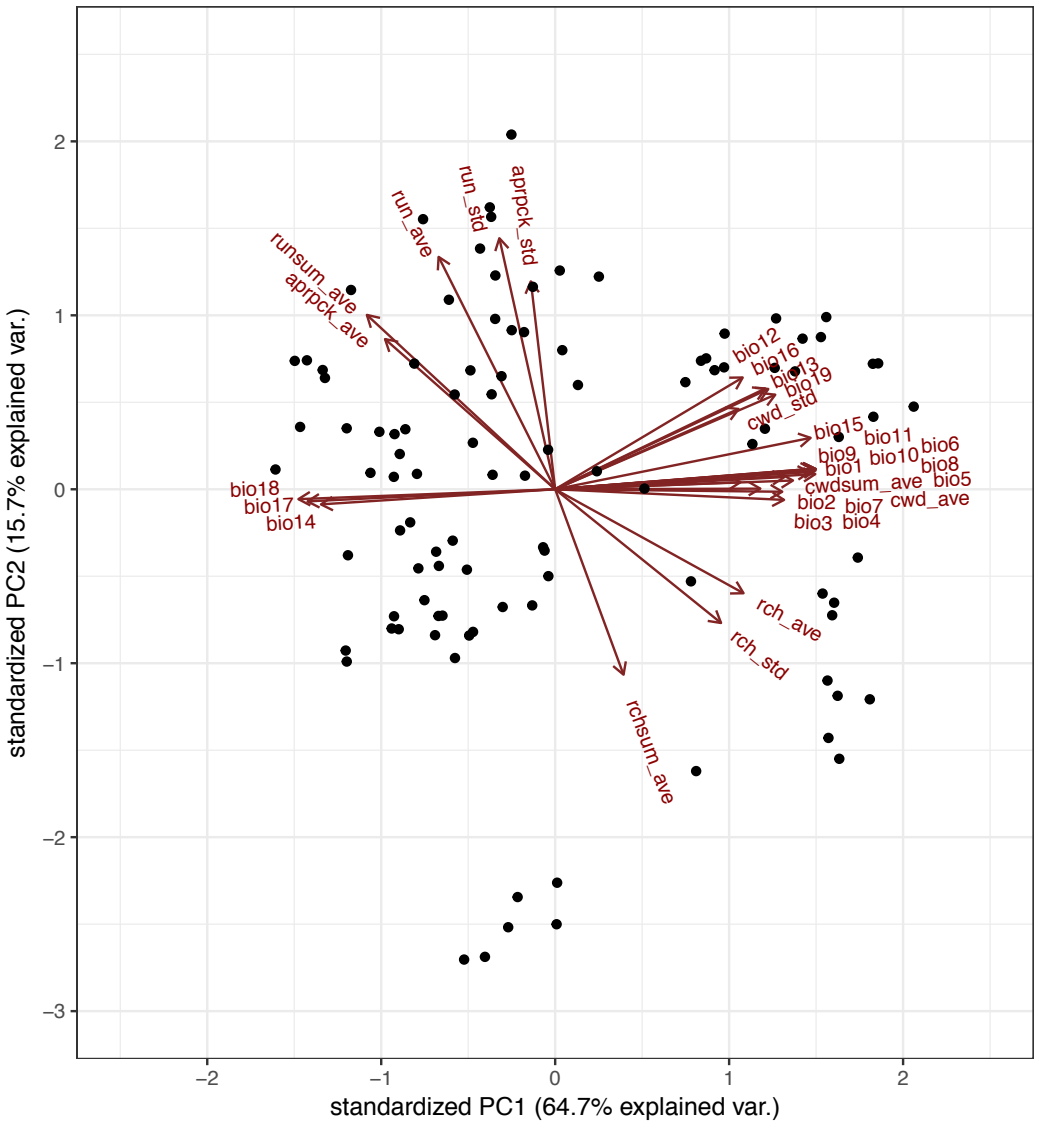


Figure S4.1. Biplot of principal components analysis of climatic data
A PCA biplot shows the first two principal components summarizing the 30 climatic variables (19 WorldClim and 11 BCM variables, see Table 2). The first two axes comprise 80.4% of the environmental variance (PC1 = 64.7%, PC2 = 15.7%). Loadings for the first five PCs are shown in Table S1.

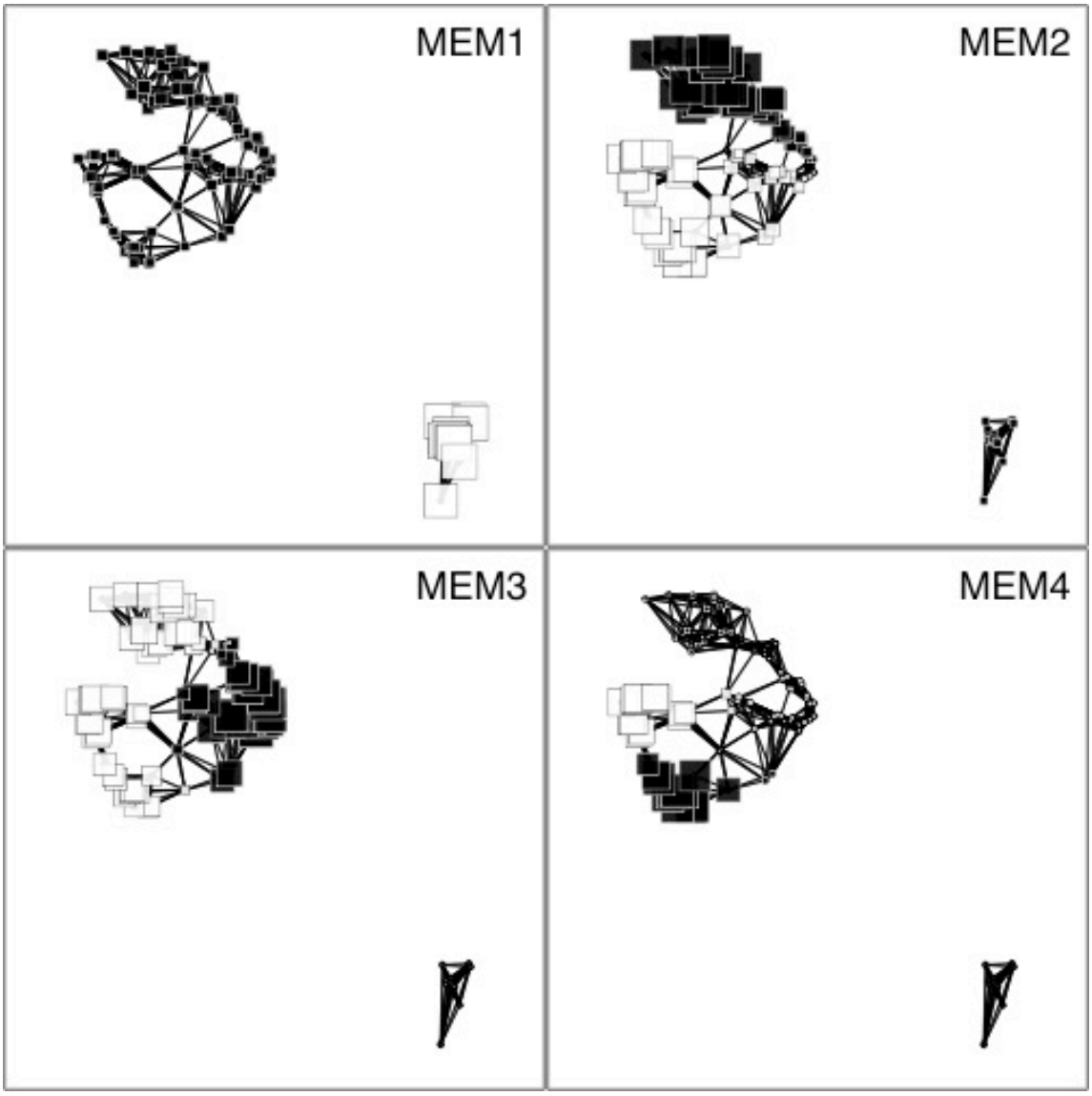


Figure S4.2. Moran's eigenvector maps of Yosemite toad sample locations
A total of 16 MEMs were found to have significant positive spatial autocorrelation (Moran's I), and were ranked by level of spatial autocorrelation. The first four (shown here) were retained to use as nuisance variables in the bayenv analysis because they corresponded closely with phylogeographic structure.

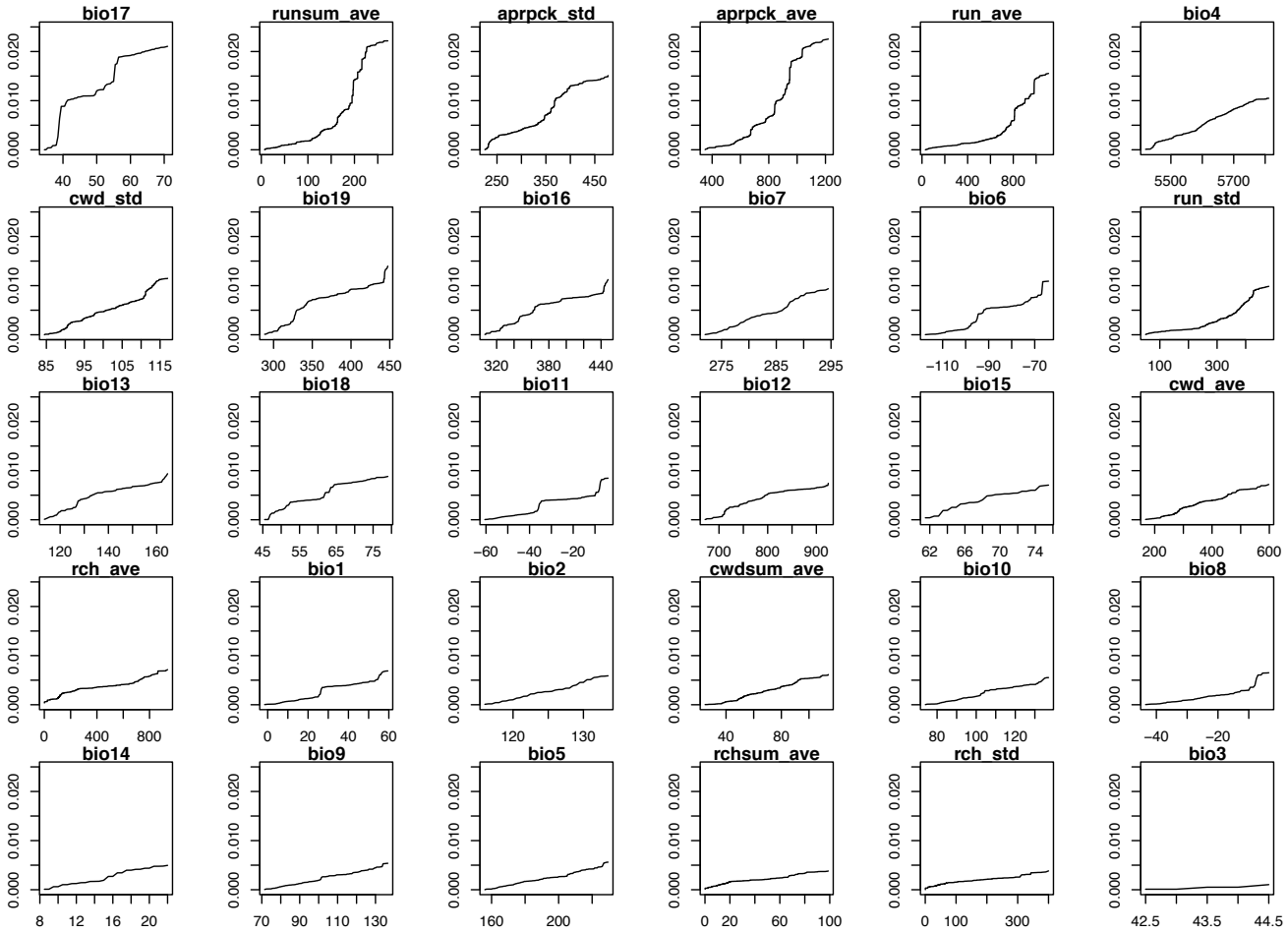


Figure S4.3. Cumulative importance curves of allelic turnover by climate
Cumulative importance distributions showing the amount of allelic change (Y axis) associated with change in each climatic variable (X axis), for all 30 predictors. Maximum height of each distribution indicates the maximum allelic change influenced by that predictor.

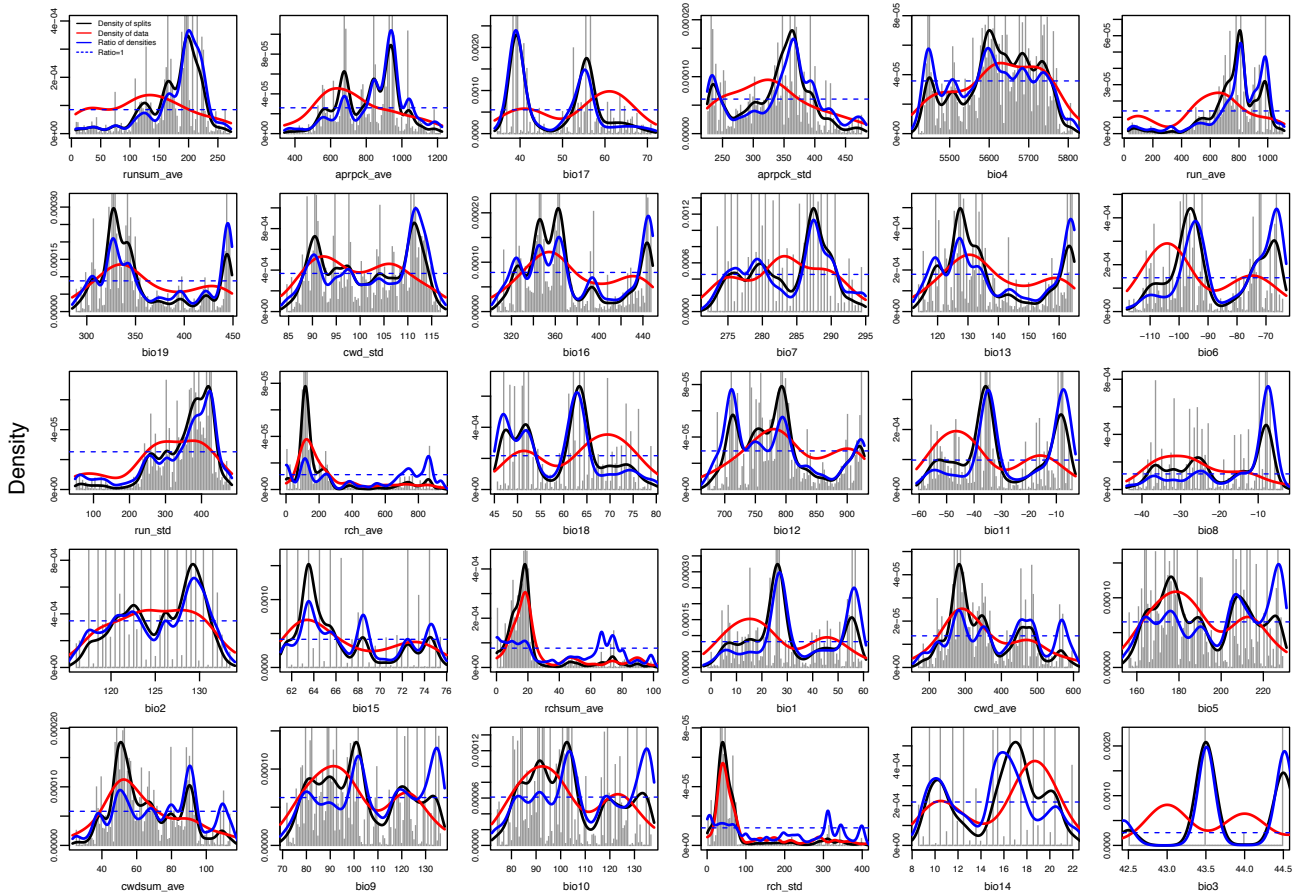


Figure S4.4. Splits importance of allelic turnover by climate

Distributions indicate where the rate of allelic change (Y axis) is highest, compared to values of each climatic variable (X axis), for all 30 predictors. Gray histogram = splits importance, black lines = kernel density of splits, red lines = kernel density of observations, blue lines = kernel density of splits standardized by observations (where the blue dotted line indicates a ratio of 1.0). All distributions integrate to variable importance values.

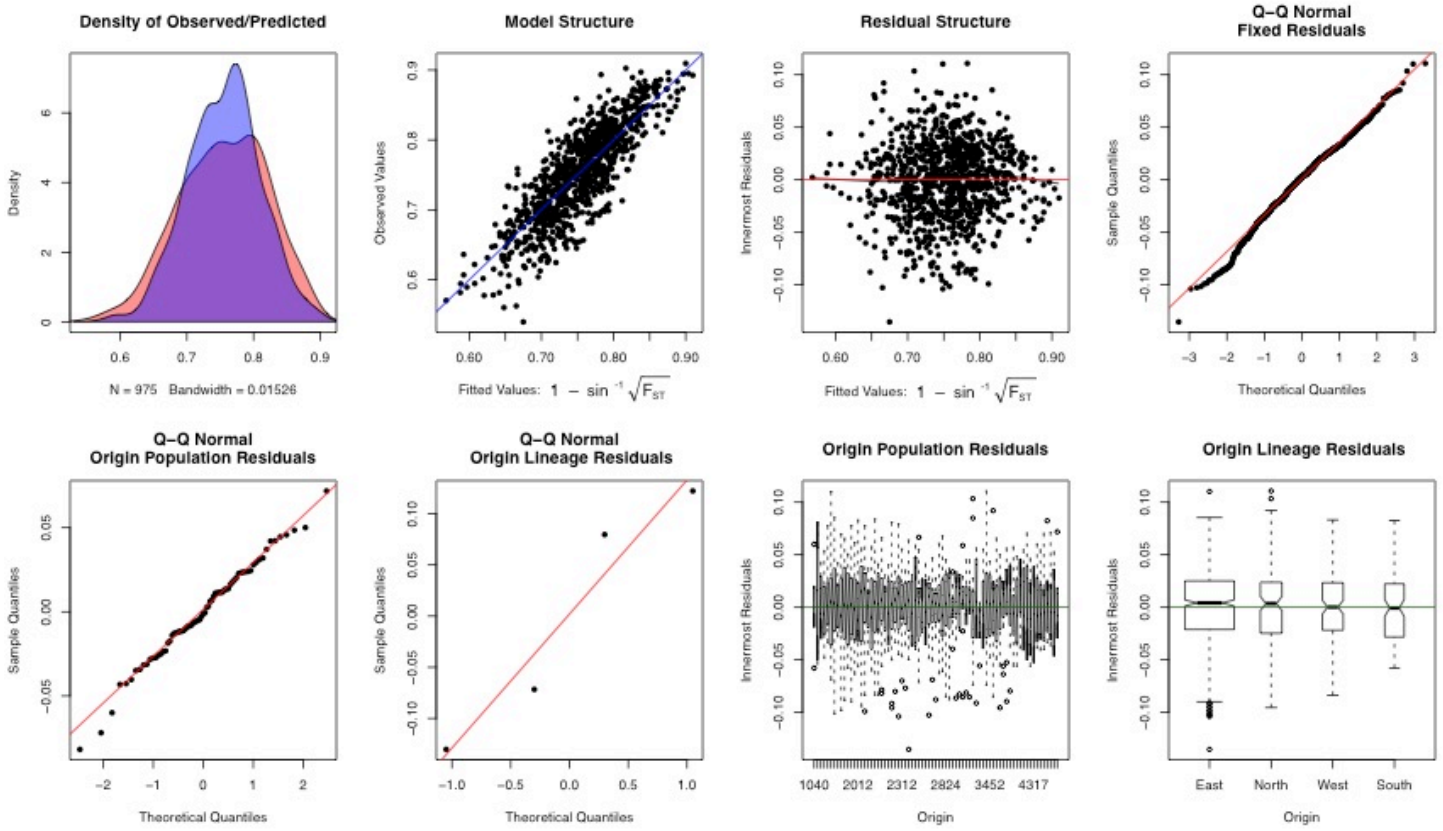


Figure S4.5. Gravity model fit statistics: climate-mediated connectivity
Performance of 30 km climate model using LCP dispersal corridors. (a) Distribution of observed (red) and predicted (blue) values. (b) Fit of observed values to predicted values on regression line. (c) Deviation of residuals from central predicted values. (d) Q-Q plot of fixed residual quantiles against normally distributed quantiles. (e) Q-Q plot of random intercept residual quantiles against normally distributed quantiles. (f) Q-Q plot of random slope residual quantiles against normally distributed quantiles. (g) Deviation of population residuals from central predicted values, by population. (h) Deviation of lineage residuals from central predicted values, by lineage.

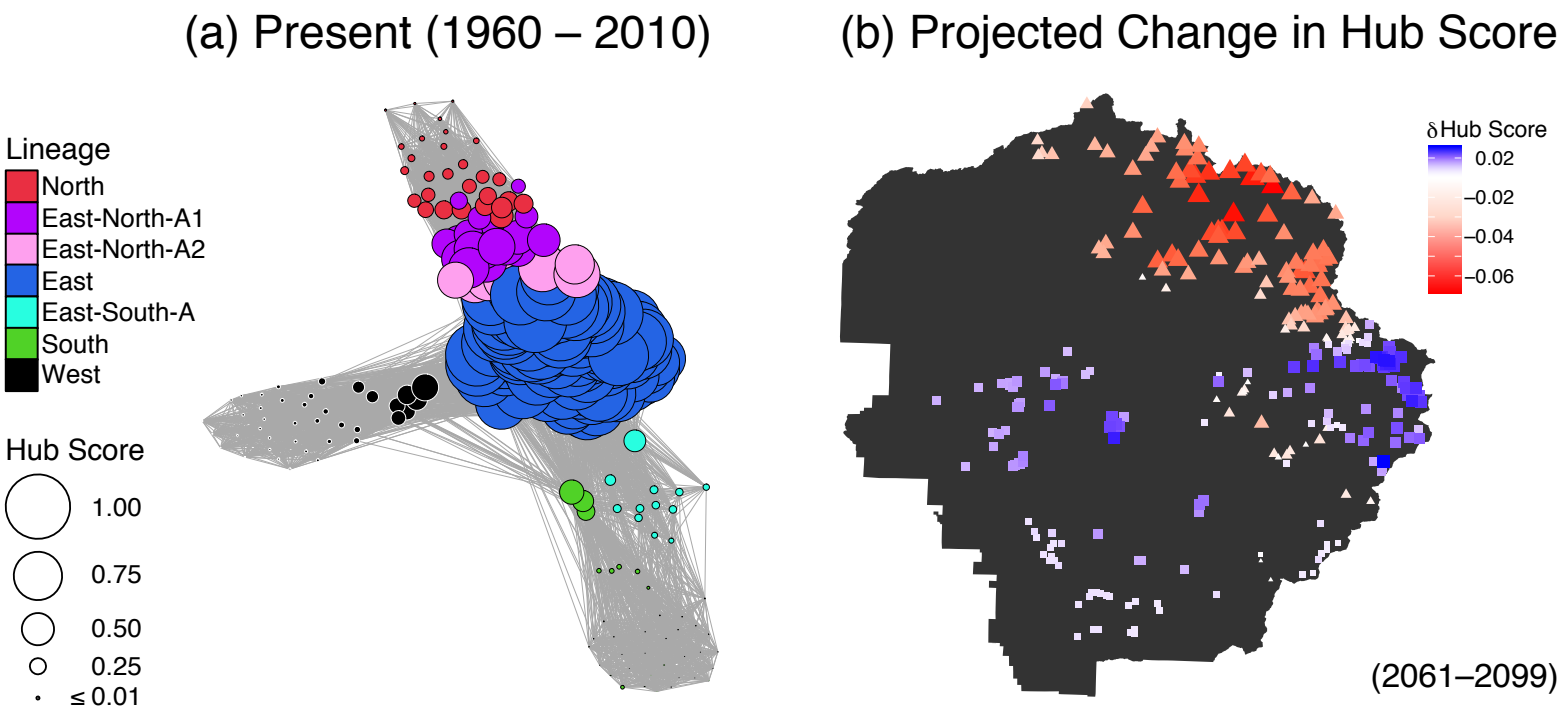


Figure S4.6. Projected future shift in meadow hub centrality score by climate change
A graph theoretical representation of meadow connectivity, and its forecasted change over time. Connectivity is defined here as Kleinberg's hub centrality score of each node in the graph, equivalent to the principal eigenvector of the weighted adjacency matrix $A^T A$, and weighted by $1 - F_{ST}$. Hub centrality represents park-wide connectedness, with the highest values receiving the most overall (direct and indirect) gene flow. The adjacency matrix of meadows includes all links <30 km in least cost path distance, which do not veer outside the species distribution. Models based on 90 sampled meadows were forecasted to all 283 meadows in YOSE known to historically contain toads. Connectivity is shown (a) before climate change, with node size indicating level of connectivity. Lineage membership is indicated by node colors; see Chapter 1 for details. (b) The net change in connectivity after ~100 years of climate change is shown, with a decrease indicated by larger triangles of increasingly red hue, and increase indicated by larger squares of increasingly blue hue.