Effects of the Dynamic Pleistocene Climate on Plant Evolution: Integrative Tests Across Multiple Spatial and Temporal Scales

by

Robert T. Massatti

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Ecology and Evolutionary Biology)
in the University of Michigan
2015

Doctoral Committee:

Professor L. Lacey Knowles, Co-Chair
Research Scientist Anton A. Reznicek, Co-Chair
Assistant Professor Jeremy N. Bassis
Assistant Professor Stephen A. Smith

Acknowledgements

I significantly benefited from the support, advice, and friendship of many people during my tenure at the University of Michigan. I'd like to thank Lacey Knowles for generously opening her lab to me, providing mentorship, and for always finding time to listen to my ideas and provide feedback. Her insights shaped my dissertation into an interesting and meaningful body of research. I enjoyed many stimulating conversations with Tony Reznicek, who was always willing to speculate on the evolutionary history of sedges and montane plants. I am also grateful to Tony for providing valuable assistance with matters related to the herbarium, for example helping me obtain loans pertinent to my phylogenetic research. Stephen Smith and Jeremy Bassis provided valuable suggestions that helped this dissertation be thoughtful and well rounded. I am grateful to all of my committee members for helping me develop into a better scientist.

I'd like to thank all of the members of the Knowles lab, from whom I've learned so much. Jen-Pan Huang, Andréa Thomaz, and Carlos Muñoz provided companionship, ears to bounce ideas off of, and valuable distractions when I couldn't stand my research any longer. Anna Papadopoulou, Mark Christie, Jess Pierson, Hayley Lanier, and Dan Edwards generously read and edited many of my grant proposal, as well as helped me learn analytical and/or laboratory techniques. Other lab members who significantly benefited my ideas and research include: Qixin He, Tristan McKnight, Lucy Tran, Diego Alvarado-Serrano, and Huateng Huang.

Finally, I'd like to acknowledge my family for their support during this process. Sirène Lipschutz edited much of my writing, provided respite from science, and took on many tasks around the house to help free up time for me to work. Her support over the past five years allowed me to succeed in my studies and start a wonderful little family. My parents and brother offered constant encouragement, and their interest in my research challenged me to learn how to talk about science in an easily understandable way.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ABC approximate Bayesian computation

ENM ecological niche modeling

PCA principle component analyses

SNP single nucleotide polymorphism

ABSTRACT

New analytical techniques and abundant genomic data provide an unparalleled opportunity to illuminate evolutionary questions. This dissertation utilizes these resources to investigate plant evolution in ecosystems that were highly impacted by Pleistocene climatic oscillations. At the broadest spatial and temporal scales, I resolve relationships among species in *Carex* section *Racemosae*, which are primarily distributed at high elevations and high latitudes. Analyses utilizing the resulting phylogeny suggest that much of the extant diversity within sect. *Racemosae* was generated during the Pleistocene, and that montane clades diversified regionally, but failed to disperse between continents. This latter result, in light of floristic similarities of Asian and North American mountains, suggests that shared evolutionary lineages were generated when widespread species adapted to high latitudes became isolated in mountains at southern latitudes during glacial periods. Refocusing on a regional scale, I investigate whether genomic patterns of variation in two closely related and co-distributed species differ in a manner consistent with expectations of how species' respective microhabitats were influenced by Pleistocene glaciations. Results support plant populations of wet-adapted species being more isolated through time compared to populations of a dry-adapted species. To test whether interactions with glaciers caused these genomic patterns, I use demographic modeling that incorporates spatial and temporal heterogeneity. Coalescent simulations generate expectations for geographic patterns of molecular variation under different combinations of parameters for two models, and I subsequently use approximate Bayesian computation to select the most likely model for each species. Glaciers impacted wet and dry microhabitats differently, forcing wet-adapted species to establish isolated populations at lower elevations around the margins of glaciers. Alternatively, populations of dry-adapted species remained more connected by persisting within glaciated regions as well as establishing populations at lower elevations. These results suggest that deterministic processes are important to consider in comparative phylogeographic studies, and may partially explain discordant patterns commonly resolved among species. Many factors influence the evolution of organisms adapted to ecosystems highly

impacted by Pleistocene glacial cycles, and here I demonstrate how genomic data and new analytical techniques can be utilized to shed light on processes across spatial and temporal scales.

CHAPTER 1

Introduction

Climatic oscillations during the Pleistocene substantially impacted ecosystems, and in this dissertation I use genomic data and recently developed analytical techniques to illuminate the interaction between organisms' shifting habitats and the generation of interspecific and intraspecific diversity. Plants adapted to the cold, harsh climates of high latitudes and high elevations represent a prime system for demonstrating the utility of these data and techniques because cyclical glaciations forced species' populations to fragment, reestablish, and/or reconnect over relatively short periods of time. By inferring relationships in a large clade of predominantly montane and arctic species and reconstructing ancestral distributions, I investigate dispersal and diversification over a large spatial scale. Subsequently, I elucidate closely related species' regional patterns of molecular variation to determine if their traits deterministically affected their interactions with glaciations. By investigating processes over multiple spatial and temporal scales, this dissertation exemplifies an integrative approach for studying patterns of biodiversity and provides a framework for future research into factors affecting plant evolution in dynamic ecosystems.

Applying abundant genomic data and appropriate analytical techniques to resolve biological questions within dynamic systems is important because evolutionary processes may mislead inference on traditional molecular datasets. For example, if clades radiated rapidly or recently, conflict among gene trees may deceive phylogenetic estimations using too few loci or individuals (Maddison and Knowles, 2006). Recently, genomic data have been applied to difficult phylogenetic problems (Heliconius Genome Consortium, 2012; Eaton and Ree, 2013, Wagner et al., 2013; Takahashi et al., 2014). While these researchers were able to resolve relationships among species with high support, their concatenation of genomic loci ignores fundamental biological processes; for example, concatenation does not allow loci to have

independent histories (Maddison, 1997). In Chapter 2, I combine genomic datasets with analytical techniques that treat loci both as a unit and independently to develop a phylogenetic hypothesis for a clade of sedges (*Carex* L. section *Racemosae* G. Don, Cyperaceae) adapted predominantly to montane and tundra habitats. Rapid and recent radiations are especially apparent in clades adapted to high elevation and arctic ecosystems (e.g., Carstens and Knowles, 2007) due to their recent development (Matthews, 1979; Matthews and Ovenden, 1990; Moran et al., 2006) and the amplitude and frequency of disturbances (e.g., Hewitt, 2000). Resolved and accurate phylogenetic hypotheses are critical for downstream evolutionary analyses that attempt to attribute historical significance to specific topologies (e.g., Ree and Smith, 2008). In Chapter 3, I utilize the sect. *Racemosae* phylogeny to reconstruct the ranges of these plants' ancestors and shed light on how Pleistocene climatic oscillations influenced clade diversification, especially in reference to the Bering Land Bridge.

Genomic data can also be used in combination with demographic modeling techniques that integrate spatial and temporal heterogeneity to elucidate processes that affected species' patterns of molecular variation. The importance of such integrative research is demonstrated by He et al. (2013), who discovered that distinguishing among factors structuring an Australian lizard's phylogeographic pattern required incorporating dynamic historical events over multiple spatial and temporal scales; furthermore, they determined that results inferred using traditional analytical techniques were erroneous. However, new analytical techniques have been infrequently applied, arising in part from the difficulties in designing appropriate biological models to test among alternative evolutionary scenarios (i.e., Knowles, 2009). Within habitat heavily influenced by Pleistocene climatic oscillations, many opportunities exist for species' traits to facilitate idiosyncratic responses to glaciations. Therefore, in Chapter 4 I investigate the genetic patterns of two closely related and codistirbuted species (*C. chalciolepis* and *C. nova*) to determine whether the species' microhabitat preferences caused them to interact differently with glaciations, thereby generating discordant phylogeographic patterns. Subsequently, in Chapter 5 I test whether *in situ* survival within glaciated terrain was the mechanism by which discordant genetic patterns were formed. The importance of applying appropriate models to investigate evolutionary questions, especially now that sufficient data can be generated to resolve among alternative hypotheses, cannot be overstated.

1.1 Estimating an accurate phylogeny for a recently-radiated clade

Phylogenetic hypotheses are critical for many types of analyses throughout the ecological and evolutionary sciences. For example, shared traits among closely related taxa may influence patterns of community assembly, and therefore resolving relationships may help predict patterns of biodiversity (Cavender-Bares et al., 2004; Wiens et al., 2010). In addition, the dynamics of biogeographic barriers may facilitate different patterns of dispersal versus diversification, and these patterns should be imprinted on a phylogeny (Donoghue, 2008; Tkach et al., 2008). However, different processes during clade diversification may obfuscate phylogenetic reconstruction. For example, gene incongruence due to horizontal gene transfer, gene duplication and loss, hybridization, and/or incomplete lineage sorting may lead to difficulties reconstructing phylogenetic relationships (Maddison, 1997). In addition, the time between speciation events may be relatively short, causing a high mutational variance among loci to contribute insufficient data for estimating relationships (Huang and Knowles, 2014). Now that genomic data are fairly easy to generate, we have the opportunity to create well-resolved phylogenetic hypotheses while potentially alleviating or circumventing altogether some of the problems involved with traditional loci (e.g., lack of power to resolve nodes at various ages). In particular, restriction-associated DNA sequencing (RADseq; Miller et al., 2007; Rowe et al., 2011) data may be useful, as it produces abundant, anonymous data from throughout the genome that can be used for phylogenetic inference (Eaton and Ree, 2013; Hipp et al., 2014).

By generating phylogenetic hypotheses using both traditional Sanger sequencing and RADseq datasets, in Chapter 2 I investigate the utility of different types of data to resolve relationships within *Carex* sect. *Racemosae* across multiple phylogenetic depths (i.e., among and within species groups). I also analyze the next-generation sequencing data as both a concatenated dataset and as independent loci to determine the robustness of the inferred topologies. Based on these analyses, I evaluate the hypothesis that sect. *Racemosae sensu lato* is a monophyletic clade, as well as its purported distinctiveness from other species and sections, including some postulated close relatives based on morphological similarities. My analyses provide a highly refined estimate of phylogenetic relationships among *Carex* species in one of the largest sections in the genus, and they also provide general insights about the utility of RADseq data and inference methods that are applicable to non-model taxa.

1.2 The influence of Pleistocene climatic oscillations on plant diversification

Plants distributed throughout high elevations and northerly latitudes are adapted to harsh climates and habitats that have been repeatedly disturbed by Pleistocene climatic oscillations. For example, during the majority of the Pleistocene, cooler climates facilitated widespread continental glaciers at high latitudes (Dyke and Prest, 1987), a land bridge between Asia and North America (Hopkins, 1967), and valley glaciers in many mountain ranges unaffected by continental glaciers (Richmond, 1986; Hewitt, 1996), which caused local and regional plant extirpations and the reestablishment of populations in newly available habitat. These periods were interrupted by warmer, interglacial periods (Siegenthaler et al., 2005), during which distributions were generally restricted to higher elevations for montane plants and higher latitudes for arctic plants (Prentice et al., 2000). Increased levels of intraspecific molecular variation, as indicated by organisms' phylogeographic patterns (Avise, 2000; Hewitt, 2000; Soltis et al., 2006), and greater local species richness, as witnessed by regional patterns of arctic and montane biodiversity (Hultén, 1937; Weber, 1965; Murray, 1995), illustrate the consequences of shifting climates on these habitats.

Pleistocene climatic oscillations were especially important in facilitating plant dispersal across arctic tundra habitat, for example between Asia and North America via the Bering Land Bridge (Simpson, 1953; Waltari et al., 2007; Tkach et al., 2008). A connection between Asia and North America was also important in structuring montane floras at lower latitudes, as evidenced by the floristic similarity of the southern Rocky Mountains and the Altai Mountains, which share evolutionary lineages in more than 140 genera (Weber, 2003). However, the dynamics of plants' movements through Beringia may have been biased due to the glacial dynamics on the Asian and North American continents (Simpson, 1953; Rausch, 1994). Whether dispersal events between continents were frequent or rare is not well investigated, especially for montane plants, due to the lack of phylogenetic hypotheses of diverse groups distributed throughout these regions. Dispersal events followed by vicariance and limited to no diversification have occurred many times within genera adapted to lower elevation and steppe habitats (Tkach et al., 2008; Lu et al., 2011).

In Chapter 3 I investigate how the dynamic nature of the Pleistocene climate facilitated the creation of biodiversity using *Carex* section *Racemosae,* which contains about 70 species primarily restricted to habitats at northerly latitudes and within the montane regions of eastern Asia and western North America. Species' ranges vary from narrowly endemic to widespread,

and recently derived clades generally contain species with similar habitat preferences (Murray, 2002). In particular, I focus on identifying the timing of diversification events within sect. *Racemosae*, followed by elucidating the distributions of sect. *Racemosae* ancestors in order to infer how clades dispersed and diversified. Illuminating the impact of the dynamic climate during the Pleistocene will be generally informative for the study of biodiversity patterns within montane and arctic ecosystems.

1.3 Phylogeographic concordance in co-distributed species with different microhabitat preferences

Concordance in phylogeographic structure across species provides a comparative context critical for testing hypotheses regarding the temporal influence of environmental factors on communities (Avise, 1992; Soltis et al., 2006; Shafer et al., 2010; Ribas et al., 2012). If membership in a specific biological community is the primary factor that determines population connectedness through time, it follows that species with similar life histories and morphological traits should show concordant phylogeographic patterns. Although the impact of historical and regional processes is often emphasized in comparative phylogeographic studies by identifying and quantifying the extent of concordance in patterns of genetic variation among species (Hickerson et al., 2010), there could also be deterministic processes that would produce discord. For example, the interaction between species' microhabitat affinities and environmental perturbations may be key contributors to phylogeographic discord. Microhabitat preferences are widely recognized as important to a variety of ecological and evolutionary phenomena. They influence the richness of biological communities and partitioning of resources therein (Hixon and Beets, 1993; Ackerly, 2003; Cavender-Bares et al., 2004), lead to diversification through character displacement dictated by local environmental variables (Schluter and Grant, 1983; Losos et al., 1997; Rosenblum, 2006), and even facilitate speciation (Feder et al., 1988; Chunco et al., 2007). It follows that microhabitat may also affect species' phylogeographic structure (Hugall et al., 2002; Whiteley et al., 2004; Hodges et al., 2007).

In Chapter 4, I investigate whether two closely-related and co-distributed montane sedge species, *Carex nova* and *C. chalciolepis*, share similar phylogeographic patterns. Their many shared traits suggest that they should have concordant phylogeographic patterns. However, there is one primary difference between these species – their microhabitat affinities. Within alpine

tundra and unforested subalpine habitats, *C. nova* is restricted to wetlands and mesic meadows, whereas *C. chalciolepis* occurs on drier montane slopes, meadows, and ridges. Therefore, in Chapter 4 I ask: Are microhabitat differences of significant evolutionary consequence? That is, instead of phylogeographic concordance, do species-specific differences have the potential to affect phylogeographic structure?

1.4 The role of glaciers in impacting co-distributed species' patterns of molecular variation

Understanding the contribution of species-specific attributes to observed patterns of genetic variation is critical for determining why taxa responded similarly (or dissimilarly) to historical climate changes. Such inferences are often out of reach because many studies attempting to illuminate processes structuring patterns of genetic variation rely upon correlative analyses, in which it is not possible to rule out alternative causes for the observed patterns. Fortunately, newly developed methodologies for modeling molecular expectations in a spatially explicit framework have the potential to capture species-specific attributes that may structure population genetic variation. In Chapter 5, I use integrative distributional, demographic, and coalescent (iDDC) modeling to evaluate the relative support for a set of biologically informed models. In particular, I test whether microhabitat differences between *C. nova* and *C. chalciolepis* caused populations to be impacted differently by the Pleistocene glaciations, as was suggested by analyses in Chapter 4.

The iDDC methodology utilizes a population demographic model to make explicit predictions for patterns of genetic variation (see also Currat and Excoffier, 2004; Wegmann et al., 2006). Both spatial and temporal heterogeneity in climatically suitable areas are integrated into the modeling using information derived from ecological niche models (ENMs) for the present and the Last Glacial Maximum (see details in Knowles and Alvarado-Serrano, 2010; Brown and Knowles, 2012). To test whether current genetic structure results from a species' ability/inability to persist within mountains during glacial periods, I constructed two alternative demographic models. Species were either allowed to persist within glaciated areas or were altogether excluded. Coalescent simulations were used to predict patterns of genetic variation under these models, taking into account how shifts in climatic conditions over time would have impacted local population demography (i.e., population size and dispersal probabilities). As a result, predicted patterns of genetic variation are species specific, reflecting the interaction

between the physical environment and biological parameters that determine the level and pattern of gene flow across the landscape (see Knowles and Alvarado-Serrano, 2010; Morgan et al., 2011; Brown and Knowles, 2012). I then use these simulations in conjunction with approximate Bayesian computation to identify which model is most probable for each species, as well as to assess the quality of parameter estimates. Integrating species specific information into modeling represents a fundamental advance in testing among alternative historical scenarios, and this research represents one of the earliest demonstrations of its utility.

1.5 Literature Cited

Ackerly, D.D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Science* 164: S165–S184.

Avise, J.C. 1992. Molecular Population Structure and the Biogeographic History of a Regional Fauna: A Case History with Lessons for Conservation Biology. *Oikos* 63: 62–76.

Avise, J.C. 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts, USA.

Brown J.L., and L.L. Knowles 2012. Spatially explicit models of dynamic histories: examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Molecular Ecology* 21: 3757–3775.

Carstens, B.C., and L.L. Knowles. 2007. Estimating Species Phylogeny from Gene-Tree Probabilities Despite Incomplete Lineage Sorting: An Example from *Melanoplus* Grasshoppers. *Systematic Biology* 56: 400-411.

Cavender-Bares, J., et al. 2004. Phylogenetic Overdispersion in Floridian Oak Communities. *The American Naturalist* 163: 823–843.

Chunco, A.J., et al. 2007. Microhabitat variation and sexual selection maintain male color polymorphisms. *Evolution* 61: 2504–2515.

Currat M., and L. Excoffier. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biology* 2: 2264–2274.

Donoghue, M. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences, USA* 105: 11549-11555.

Dyke, A.S., and V.K. Prest. 1987. Late Wisconsinan and Holocene History of the Laurentide Ice Sheet. *Geographie physique et Quaternaire.* 41(2): 237-263.

Eaton, D.A.R., and R.H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62: 689–706.

Feder, J.L., et al. 1988. Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* 336: 61–64.

He, Q., et al. 2013. Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution* 67: 3386-3402.

Heliconius Genome Consortium. 2013. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–98.

Hewitt, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907-913.
Hickerson, M.J., et al. 2010. Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution* 54: 291–301.
Hipp, A.L., et al. 2014. A Framework Phylogeny of the American Oak Clade Based on Sequenced RAD Data. *PLoS ONE* 9: e93975.
Hixon, M.A., and J.P. Beets. 1993. Predation, Prey Refuges, and the Structure of Coral-Reef Fish Assemblages. *Ecological Monographs* 63: 77–101.
Hodges, K.M., et al. 2007. Remarkably different phylogeographic structure in two closely related lizard species in a zone of sympatry in southeastern Australia. *Journal of Zoology* 272: 64-72.
Hopkins, D.M., ed. 1967. *The Bering Land Bridge*. Stanford University Press, Stanford, California, USA.
Huang, H., and L.L. Knowles. 2014. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology,* In Press.
Hugall A., et al. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences, USA* 99: 6112–6117.
Hultén, E. 1937. *Outline of the history of arctic and boreal biota during the Quaternary period: their evolution during and after the glacial period as indicated by the equiformal progressive areas of present plant species*. Bokfoerlags Aktiebolaget Thule.
Knowles, L.L. 2009. Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics* 40: 593–612.
Knowles, L.L., and D.F. Alvarado-Serrano. 2010. Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology* 19: 3727–3745.
Losos, J.B., et al. 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. *Nature* 387: 70–73.
Lu, J., et al. 2011. Biogeographic disjunction between eastern Asia and North America in the *Adiantum pedatum* complex (Pteridaceae). *American Journal of Botany* 98(10): 1680-1693.
Maddison, W.P., and L.L. Knowles. 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic Biology* 55: 21-30.
Maddison, W.P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.
Matthews, J.V. 1979. Tertiary and Quaternary environments: Historical background for an analysis of the Canadian insect fauna. *In* H.V. Danks [ed.], *Canada and its Insect fauna*, 31-86. Entomological Society of Canada, Ottawa, Canada.
Matthews, J.V., and L.E. Ovenden. 1990. Late Tertiary plant macrofossils from localities in Arctic/Subarctic North America: a review of the data. *Arctic* 43: 364–392.
Miller, M.R., et al. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240–248.
Moran, K., et al. 2006. The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature* 441: 601-605.

Morgan K., et al. 2011. Comparative phylogeography reveals a shared impact of Pleistocene environmental change in shaping genetic diversity within nine *Anopheles* mosquito species across the Indo-Burma biodiversity hotspot. *Molecular Ecology* 20: 4533–4549.
Murray, D.F. 1995. Causes of arctic plant diversity: origin and evolution. *In* F. S. Chapin and C. Körner [eds.], *Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences,* 21-32. Springer, Heidelberg, Germany.
Murray, D.F. 2002. *Carex* L. sect. *Racemosae* G. Don. *In* Flora of North America Editorial Committee [eds.], *Flora of North America, north of Mexico*, vol. 23, 401-414. Oxford University Press, New York, New York, USA.
Prentice, I.C., et al. 2000. Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography* 27(3): 507-519.
Ree, R.H., and S.A. Smith. 2008. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. *Systematic Biology* 57: 4-14.
Richmond, G.M. 1986. Stratigraphy and correlation of glacial deposits of the Rocky Mountains, the Colorado Plateau and the ranges of the Great Basin. *Quaternary Science Reviews* 5: 99-127.
Rausch, R. 1994. Transberingian dispersal of cestodes in mammals. *International Journal of Parasitology* 24: 1203-1212.
Ribas, C.C., et al. 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Philosophical Transactions of the Royal Society B: Biological Sciences* 22: 681–689.
Rosenblum, E.B. 2006. Convergent evolution and divergent selection: lizards at the White Sands ecotone. *The American Naturalist* 167: 1–15.
Rowe, H.C., et al. 2011. RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* 20: 3499–3502.
Schluter, D., and P.R. Grant. 1983. Determinants of Morphological Patterns in Communities of Darwin's Finches. *The American Naturalist* 123: 175–196.
Siegenthaler, U., et al. 2005. Stable Carbon Cycle-Climate Relationship During the Late Pleistocene. *Science* 310(5752): 1313-1317.
Simpson, G.L. 1953. Evolution and Geography: An Essay on Historical Biogeography with Special Reference to Mammals. Condon Lectures, Oregon State System of Higher Education. Eugene, Oregon, USA.
Shafer, A.B.A., et al. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* 19: 4589–4621.
Soltis, D.E., et al. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15: 4261–4293.
Takahashi, T., et al. 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution* 80: 137–144.
Tkach, N.V., et al. 2008. Parallel evolutionary patterns in multiple lineages of arctic *Artemisia* L. (Asteraceae). *Evolution* 62: 184–198.
Wagner, C.E., et al. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787–798.
Waltari, E., et al. 2007. Eastward Ho: phylogeographical perspectives on colonization of hosts and parasites across the Beringian nexus. *Journal of Biogeography* 34: 561-574.

Weber, W.A. 1965. Plant geography in the southern Rocky Mountains. *The Quaternary of the United States*. 453–468.

Weber, W.A. 2003. The middle Asian element in the southern Rocky Mountain flora of the western United States: a critical biogeographical review. *Journal of Biogeography* 30(5): 649–685.

Wegmann D., et al. 2006. Molecular diversity after a range expansion in heterogeneous environments. *Genetics* 174: 2009–2020.

Whiteley, A.R., et al. 2004. Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. *Molecular Ecology* 13: 3675-3688.

Wiens, J.J., et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13: 1310–1324.

CHAPTER 2

Utilizing RADseq Data for Phylogenetic Analysis of Challenging Taxonomic Groups: A Case-study in *Carex* Sect. *Racemosae* (Cyperaceae)

2.1 Abstract

Premise of study: Relationships among closely related and recently diverged taxa can be especially difficult to resolve. Here we use both traditional Sanger datasets and next-generation RADseq datasets to estimate phylogenetic relationships among species of *Carex* section *Racemosae* (Cyperaceae), a clade largely restricted to high latitudes and elevations. Interest in the relationships among these taxa derives from questions about their biogeographic history and possible links of diversification to Pleistocene glaciations.

Methods: A combination of approaches and different molecular markers were used to estimate relationships among *Carex* species within sect. *Racemosae* and taxa from closely related sections. Traditional molecular markers generated by Sanger sequencing were analyzed with *BEAST, and SNP data from RADseq loci were analyzed as a concatenated dataset using maximum likelihood and as independent loci using SVDquartets.

Key Results: Traditional molecular markers resolved relationships among taxa at intermediate phylogenetic depths (albeit with low levels of support). Only the RADseq data resolved relationships with strong support at all phylogenetic depths. Moreover, different methods and data-partitions of the RADseq data resulted in nearly identical topologies. *Carex* sect. *Racemosae* is a strongly supported clade, although a handful of species were found to group with closely related sections.

Conclusions: Despite the short read lengths of RADseq data, they nevertheless resolved relationships that commonly-used molecular loci did not. Resolution of the phylogenetic relationships among recently and rapidly diversifying taxa within sect. *Racemosae* clades suggest a role for the Pleistocene glaciations in clade diversification.

2.2 Introduction

Phylogenetic hypotheses provide a key framework for testing mechanisms underlying macroevolutionary patterns, including processes important in structuring communities (Cavender-Bares et al., 2004; Donoghue, 2008; Wiens et al., 2010), the distribution of morphological traits within and among communities (Kraft et al., 2007), and the dispersal and diversification of clades across geographic barriers (Tkach et al., 2008). Today, phylogenies commonly incorporate nuclear and chloroplast (and/or mitochondrial) loci. Although the inclusion of such sequence data has generally led to a vast improvement over exclusively using morphological data (e.g., see classifications by Cronquist, 1981 or Takhtajan, 1987 vs. APG III, 2009), phylogenetic inference with molecular datasets is not free of challenges. For example, gene incongruence due to horizontal gene transfer, gene duplication and loss, hybridization, and/or incomplete lineage sorting may lead to difficulties reconstructing phylogenetic relationships (Maddison, 1997). In addition, phylogenetic analyses that rely on limited numbers of loci, even with methods that accommodate processes such as gene-lineage coalescence (Knowles, 2009), may leave a proportion of relationships unresolved. Obtaining resolved phylogenetic relationships becomes especially challenging when the times between speciation events are relatively short, such that a high mutational variance among loci may contribute to insufficient data for estimating relationships (Huang and Knowles, 2014).

Now that genomic data are fairly easy to generate, we have the opportunity to create well-resolved phylogenetic hypotheses while potentially alleviating or circumventing altogether some of the problems involved with traditional loci (e.g., lack of power to resolve nodes at various ages). In particular, restriction-associated DNA sequencing (RADseq; Miller et al., 2007; Baird et al., 2008; Rowe et al., 2011) data may be useful, as it produces abundant, anonymous data from throughout the genome that can be used for phylogenetic inference (Eaton and Ree, 2013; Hipp et al., 2014). Because of the importance of identifying a sufficient number of orthologous restriction sites among species (e.g., Rubin et al., 2012), RADseq data has typically

been used in phylogenetic analysis of recently derived clades (Heliconius Genome Consortium, 2012; Eaton and Ree, 2013, Wagner et al., 2013; Takahashi et al., 2014). However, temporal boundaries are being continually expanded. For example, using a RADseq dataset, Hipp et al. (2014) infer relationships within a clade of oaks estimated to be 23-33 million years old. However, the limit of their utility at greater phylogenetic depths, where homoplasy might erode the phylogenetic signal contained across the single nucleotide polymorphisms (SNPs) of independent loci, is not yet known.

Carex L. is a large and diverse genus for which traditional sequence data have failed to resolve well-supported relationships across multiple phylogenetic levels (Roalson et al., 2001; Hipp et al., 2006; Starr and Ford, 2009). The taxa are ecologically important and distributed throughout many habitat types worldwide (Reznicek, 1990). *Carex* section *Racemosae* G. Don, one of the largest sections within the genus, contains roughly 70 species mostly distributed at higher latitudes and elevations in the northern hemisphere (Egorova, 1999; Murray, 2002; Lunkai et al., 2010). Section *Racemosae* species are predominantly restricted to cold, harsh habitats that initially became widespread and interconnected during the Pliocene (Matthews, 1979; Matthews and Ovenden, 1990), suggesting that the crown group diversified relatively recently. In addition, montane clades impacted by glaciations may represent more recent, Pleistocene diversifications (e.g., Massatti and Knowles, 2014). As such, sect. *Racemosae* offers an opportunity to test the utility of RADseq data for resolving relationships over recent to intermediate phylogenetic depths. Few of these taxa have been included in previous phylogenetic analyses of the genus, and while they are generally united by shared morphological traits including trigonous achenes and multiple spikes, where the terminal spike is gynecandrous or unisexual, multiple studies have indicated that the section is non-monophyletic (Hendrichs et al., 2004; Waterway et al., 2009). However, these studies suffer from a paucity of molecular data that leaves many relationships unsupported, as well as from a dependency on high copy nuclear loci (e.g., ITS) that may mislead phylogenetic inference due to intra-individual polymorphisms and incomplete concerted evolution (see King and Roalson, 2008).

By generating phylogenetic hypotheses from both traditional and RADseq datasets, here we investigate the utility of different data types to resolve relationships within sect. *Racemosae* across multiple phylogenetic depths (i.e., among and within species groups). We also analyze the RADseq dataset with multiple methodologies to determine the robustness of the inferred

topologies. Based on these analyses, we evaluate the hypothesis that sect. *Racemosae* sensu lato is a monophyletic clade, and its purported distinctiveness from other species and sections, including some postulated close relatives based on morphological similarities. Our analyses provide a highly refined estimate of phylogenetic relationships among *Carex* species in one of the largest sections in the genus, and they also provide general insights about the utility of RADseq data and inference methods that are applicable to non-model taxa.

2.3 Materials and Methods

Taxon sampling—We sampled 111 plants specimens representing 48 species within sect. *Racemosae*, as well as 21 species from eight closely related sections (Table A.1). Sampling represents nearly all of the sect. *Racemosae* taxa from North America (excluding a couple of recent segregates, e.g., *C. orestera* Zika) and roughly 75% of taxa found worldwide. Unsampled taxa occur mainly as narrow endemics in Asian mountain ranges and are not well represented within accessible herbarium collections. Multiple individuals were sampled where possible to facilitate species tree reconstructions (see below). About 0.30 grams of leaf tissue was sampled from herbarium specimens housed at A, COLO, DAO, GH, MICH, MO, NAVA, RM, and SI or from live plants; the leaf material was stored in silica gel until DNA was extracted using DNeasy Plant Mini Kits (QIAGEN, Valencia, California, USA) following the manufacturer's protocol.

Sanger sequencing dataset—We amplified two nuclear and one chloroplast marker using polymerase chain reaction (PCR) on an Eppendorf Thermal Cycler (Hauppauge, New York, USA); these loci were shown to have phylogenetic utility in previous *Carex* studies (Starr et al., 2003; Starr et al., 2009). All loci were successfully amplified in the following 20 μL reaction mixture: 10x reaction buffer, 10 mM dNTPs, 1 μM of both primers, 1 U TaKaRa Ex Taq (Clontech, Mountain View, California, USA), 25 mM $MgCl_2$, and 1 M Betaine. The primers ITS4 and ITS1 (White et al., 1990) were used to amplify the internal transcribed spacer region (ITS) and the included 5.8S rDNA. DNA amplification followed a thermal cycler protocol that included: 95°C initial denaturing for 4 min; followed by 35 cycles of 95°C denaturing for 1 min, 50°C annealing for 1 min, and 72°C extension for 1 min; followed by a final extension at 72°C for 15 min. PCR of the external transcribed fragment (ETS) employed the primers ETS-1f and 18S-R (Starr et al., 2003); amplification followed the thermal cycler protocol of Starr et al.

(2003). PCR of the chloroplast *matK* locus was completed using the primers MOmatK480F and trnK2R (Hilu et al., 2003). While most accessions were successfully amplified using the two external primers, some were amplified in two parts using two internal primers: matK-2.1f (5' CCT ATC CAT CTG GAA ATC TTA G 3') and matK-5r (5' GTT CTA GCA CAA GAA AGT CG 3'). Amplification of the chloroplast DNA employed the following thermal cycler program: 94°C initial denaturing for 2 min; followed by 35 cycles of 94°C denaturing for 1 min, 48°C annealing for 30 sec, and 72°C extension for 1 min; followed by a final extension at 72°C for 10 min.

The PCR products were electrophoresed on 2% agarose gels in 1x Tris-borate-EDTA (TBE) buffer (pH 8.0) and stained with ethidium bromide. The amplified templates were processed at the University of Michigan Sequencing Core using the same primers used in the PCR reactions. Chromatograms were proofed by visual inspection and edited using the program Sequencher v.4.2 (Gene Codes Corporation, Ann Arbor, MI, USA). Sequences were initially aligned in MEGA V.6.06 (Tamura et al., 2013) with MUSCLE (Edgar, 2004) using a 'Gap Open Penalty' of -400 and the UPGMB clustering method. Subsequently, the initial alignment was visually inspected for accuracy. Indels and the poly A tail at the 3' end of ITS1 were excluded from phylogenetic analyses. Sequences for all specimens are available on GenBank (Table A.1).

RADseq dataset—We selected 58 specimens representing 48 species within sect. *Racemosae* and 7 species from closely related sections for our RADseq data collection efforts (bolded taxa, Table A.1). For these specimens, extracted genomic DNA was individually barcoded and processed into a reduced complexity library using a restriction fragment-based procedure (for details see Peterson et al., 2012). Briefly, DNA was doubly digested with *Eco*RI and *Mse*I restriction enzymes, followed by the ligation of Illumina adaptor sequences and unique barcodes. Ligation products were pooled among samples and the fragments were amplified by 12 cycles of PCR. A Pippin Prep (Sage Science, Beverly, MA, USA) was used to size select fragments between 400 and 500 base pairs. The library was sequenced at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, Canada) in one lane on the Illumina HiSeq2500 platform to generate 150 base pair, single-end reads. Sequences were demultiplexed using process_radtags.pl, which is distributed as part of the Stacks pipeline (Catchen et al., 2013); only

reads with Phred scores ≥ 32, no adaptor contamination, and unambiguous barcode and restriction cut sites were retained.

Data were analyzed in the pyRAD pipeline, which accounts for indels that may be present among species' homologous loci (for details see Eaton, 2014). Briefly, sequences of each individual are clustered into highly similar stacks using USEARCH (Edgar, 2010), followed by the estimation of rates of heterozygosity and sequencing error (Lynch, 2008). Heterozygotes are inferred by a binomial probability based on these parameters. Each resulting stack is hereafter referred to as a RADseq locus. Each individual's RADseq loci are independently summarized into consensus sequences, which are subsequently clustered among individuals to generate a data matrix. Because not every individual has a sequence for every RADseq locus, due to both variation in sequencing coverage and mutations in the restriction site defining the RAD loci, the resulting data matrix is expected to be incomplete. Parameter values used for clustering were chosen based on suggestions from the literature (Eaton and Ree, 2013; Eaton, 2014; Takahashi et al., 2014). The percent similarity required to cluster sequences into a stack and individuals into a locus was 0.88, the minimum stack depth for each individual was 6, and the minimum number of individuals per locus cluster was 4; full parameter details are archived with associated datasets on Dryad (doi:xxxx). The amount of shared loci among taxa was visualized using the corrplot function in the 'corrplot' package (Wei, 2015) in R (R Core Team, 2014).

Phylogenetic inference: Sanger dataset—In order to allow idiosyncratic patterns of descent from ancestral alleles (e.g., Degnan and Rosenberg, 2009), we estimated a species tree using the *BEAST template in BEAST v2.1.3 (Bouckaert et al., 2014). We utilized two datasets in species tree reconstructions. The first included 76 accessions representing 47 taxa for which we acquired RADseq data, and is hereafter referred to as the 'reduced Sanger dataset'. In contrast, the 'full Sanger dataset' was composed of 107 accessions representing 70 taxa; this analysis included all Sanger data that we generated. We linked the ITS and ETS tree models because of their physical linkage, while the *matK* tree model remained unlinked. Substitution (GTR + Γ) and lognormal relaxed clock models were unlinked across loci, and we used a Yule process species tree prior. Three independent runs per dataset (six runs total) were conducted to compare the similarity of the maximum clade credibility trees and nodal posterior probabilities. Each replicate ran for 1E9 MCMC iterations with trees sampled every 10 000 generations (for a

total of 100 000 sampled trees). Visual inspection of the marginal densities of the estimated parameters and their associated Effective Sample Size (ESS) scores in Tracer v1.6 (Rambaut et al., 2014) was used to assess convergence, which occurred within the first 20 000 MCMC generations in the six independent *BEAST runs. Maximum clade credibility trees were generated using TreeAnnotator v.2.1.2 (distributed with BEAST v2.1.3) after discarding the first 20% of sampled trees (resulting in 80 000 trees used per run).

Phylogenetic inference: SNP dataset—A dataset consisting of unlinked SNPs created by pyRAD from the RADseq loci was used in multiple phylogenetic analyses. First, a phylogeny was estimated from the concatenated SNP dataset using maximum likelihood (ML) in RAxML v.8.1.16 (Stamatakis, 2014), with a GTR + Γ model of nucleotide evolution. Likelihood calculations were corrected for ascertainment bias (-m ASC_GTRGAMMA); this option is useful for SNP datasets that contain no invariable sites, which will cause the algorithm to exit with errors. Therefore, prior to analysis, the dataset was filtered with a custom script (doi:xxxx) to exclude sites that were potentially invariable due to ambiguous characters. Support was assessed by 500 nonparametric bootstrap replicates, followed by a search for the best-scoring maximum likelihood tree.

The relationships among and within clades suggested by the ML analysis of the concatenated SNP dataset were further investigated by analyses that account for differences in the genealogical histories of individual loci. Specifically, we used the program SVDquartets (Chifman and Kubatko, 2014), which is a quartet sampling method that accounts for sequence variability due to both mutational and coalescent variance. Because SVDquartets uses the sequence data directly, and it does not require long sequence reads to estimate a gene tree, it provides an advantage over summary-statistics methods for estimating species trees such as STEM (Kubatko et al., 2009), STAR (Liu et al., 2009), and MP-EST (Liu et al., 2010). In addition, the method is rapid and results are straightforward to interpret, in contrast to other SNP-based approaches that use MCMC frameworks (e.g., SNAPP, Bryant et al., 2012), which can be slow for large datasets and difficult to assess convergence. Note, however, that a large number of quartets must be sampled to estimate phylogenetic relationships. To maximize the coverage of shared data across taxa (i.e., the overlap in shared loci decreases as a function of the time since common ancestry; Huang and Knowles, 2014), only one outgroup and sect.

Racemosae sensu stricto (s.s.; 39 species total - see Results) were included in the analyses. Three independent runs of SVDquartets were conducted to assess topological convergence, for each of which 1E9 quartets were generated.

2.4 Results

Sequence analysis: Sanger dataset—The ITS and ETS nuclear regions and the *matK* chloroplast region were successfully amplified in 106, 105, and 95 specimens, respectively (Table A.1). Alignment of the full nuclear dataset (ITS + ETS) yielded 1274 characters, of which 474 were variable and 199 (15.6%) were phylogenetically informative. Alignment of all chloroplast sequences (*matK*) yielded 1209 characters, of which 192 were variable and 88 (7.3%) were phylogenetically informative. All phylogenetically informative sites were retained after pruning out taxa to create the reduced Sanger dataset.

Sequence analysis: SNP dataset—Of the 58 specimens included in the Illumina library, nine were discarded because they had low numbers of raw reads and few loci (Table A.2). While four of the excluded taxa were sampled between 1984 and 1997, the other five represent 5 out of 6 taxa sampled from vouchers collected prior to 1980, suggesting a limit to the usefulness of herbarium collections more than roughly 35 years old. In addition, the *C. atrosquama* Mack. accession was found to be misidentified (likely a specimen of *C. scopulorum* Holm) and was excluded from analyses.

After data processing, taxa had on average 12 598 unlinked SNPs in the final dataset, based on an average of 1 674 947 reads per specimen (Table 2.1 and Table A.2). However, the proportion of shared loci varied among taxa. Specifically, the initial concatenated data matrix contained 41 615 loci, with as few as 1.2% shared between more distantly related taxa (e.g., *C. adelostoma* V.I. Krecz. and *C. vaginata* Tausch) and as many as 37.2% shared between closely related taxa (e.g., *C. nova* L.H. Bailey and *C. nelsonii* Mack.; Fig. A.1). We used 12 689 SNPs (doi:xxxx) for the maximum likelihood analysis of the concatenated data with RAxML, and 9317 unlinked SNPs for the SVDquartets analyses.

Phylogenetic inference: Sanger dataset—The 3 loci comprising the full and reduced Sanger datasets consistently resolved recently derived clades (numbered circles in Fig. 2.1 and

Fig A.2), except for the placements of *C. obscura* Nees and *C. melanocephala* Turcz., which were nested within Clade 5 in analyses on the full Sanger dataset, but independent of defined clades in analyses on the reduced Sanger dataset. However, relationships among these clades (i.e., the backbone topology) and those within many clades were unresolved and differed between the two datasets. In addition, relationships within and among recently derived clades were variable across independent *BEAST runs, even though convergence appeared to be reached in every instance (see details in Methods). We only present the maximum clade credibility trees from the first of the three *BEAST runs, and except where noted, we only refer to the tree generated with the reduced Sanger dataset (Fig. 2.1). All posterior probabilities (PPs) were low, except on some branches subtending aforementioned recently derived clades and a couple of branches within these clades (Fig. 2.1 and Fig. A.2). Nevertheless, moderate support (82% PP) was found for the monophyly of sect. *Racemosae* s.s., which excludes five taxa typically treated as members of this clade: *C. augustinowiczii* Meinsh. ex Korsh., *C. curvicollis* Franch. & Sav., *C. mertensii* J.D. Prescott ex Bong., *C. meyeriana* Kunth, and *C. stylosa* C.A. Mey.

Phylogenetic inference: SNP dataset—The ML and SVDquartets analyses of the respective SNP datasets produced very consistent results (Figs. 2.2 and 2.3). The monophyly of sect. *Racemosae* s.s. was highly supported, as was the sister relationship between sect. *Racemosae* and *C. bicolor* Bellardi ex All. (and additionally *C. hassei* L.H. Bailey, *C. vaginata,* and *C. williamsii* Britton in the ML analyses that included more outgroup taxa). These analyses also corroborated phylogenetic estimates of the recently derived clades that were consistently resolved using the Sanger datasets (as noted by the same circled numbers in Figs. 2.2 and 2.3 as in Fig. 2.1), including the placement of *C. obscura* and *C. melanocephala* as independent of defined clades, similar to the reduced Sanger dataset (see above). The best-scoring ML tree contained many well-supported nodes, most with 100% bootstrap support, although multiple nodes had lower bootstrap values (Fig. 2.2). The only difference between the ML and SVDquartets topologies was the placement of the clade containing *C. helleri* Mack. and *C. raynoldsii* Dewey (see Figs. 2.2 and 2.3); while these taxa were placed as sister to Clade 2 in the ML phylogeny, the quartet phylogeny placed them sister to the clade containing Clade 2, Clade 3, *C. melanocephala*, *C. gmelinii* Hook. & Arn., and *C. hancockiana* Maxim.

Table 2.1 Data processing summary statistics detailing next-generation sequencing and the single nucleotide polymorphism (SNP) dataset. Raw reads refers to the total reads produced by Illumina sequencing, while post-processing reads are those that remained after filtering for adaptor contamination, quality, and ambiguous barcodes. The post-processing reads were utilized by pyRAD to create clusters of homologous sequences per species (Total clusters, Mean depth of clusters). After the heterozygosity (H) and error-rate (E) were estimated across clusters, consensus sequences were created for each cluster; those that passed the pyRAD filtering parameters were retained (Loci). Variable and invariable DNA sites were summed across all loci (Total sites), and the percentage of polymorphic sites (Percent poly) is reported. Consensus sequences were clustered across species, and loci that passed filtering parameters were included in the final data matrix (Final loci). Refer to Table A.2 for a breakdown by species, as well as for the taxa not included in the final SNP dataset.

Summary statistic	Raw reads	Post-processing reads	Percent retained	Total clusters	Mean depth of clusters	H	E	Loci	Total sites	Percent poly	Final loci
Average	1 882 068	1 674 947	88.0	23 741	53.4	0.0104	0.0009	20 864	2 815 076	0.18	12 598
Min	138 966	107 088	77.1	4214	13.0	0.0068	0.0004	3234	436 159	0.06	1564
Max	6 307 952	5 746 439	93.1	49 313	114.8	0.0228	0.0032	42 722	5 766 499	0.47	27 988
St. Dev.	1 321 887	1 195 946	3.0	10 853	21.3	0.0036	0.0005	9933	1 340 434	0.12	6693

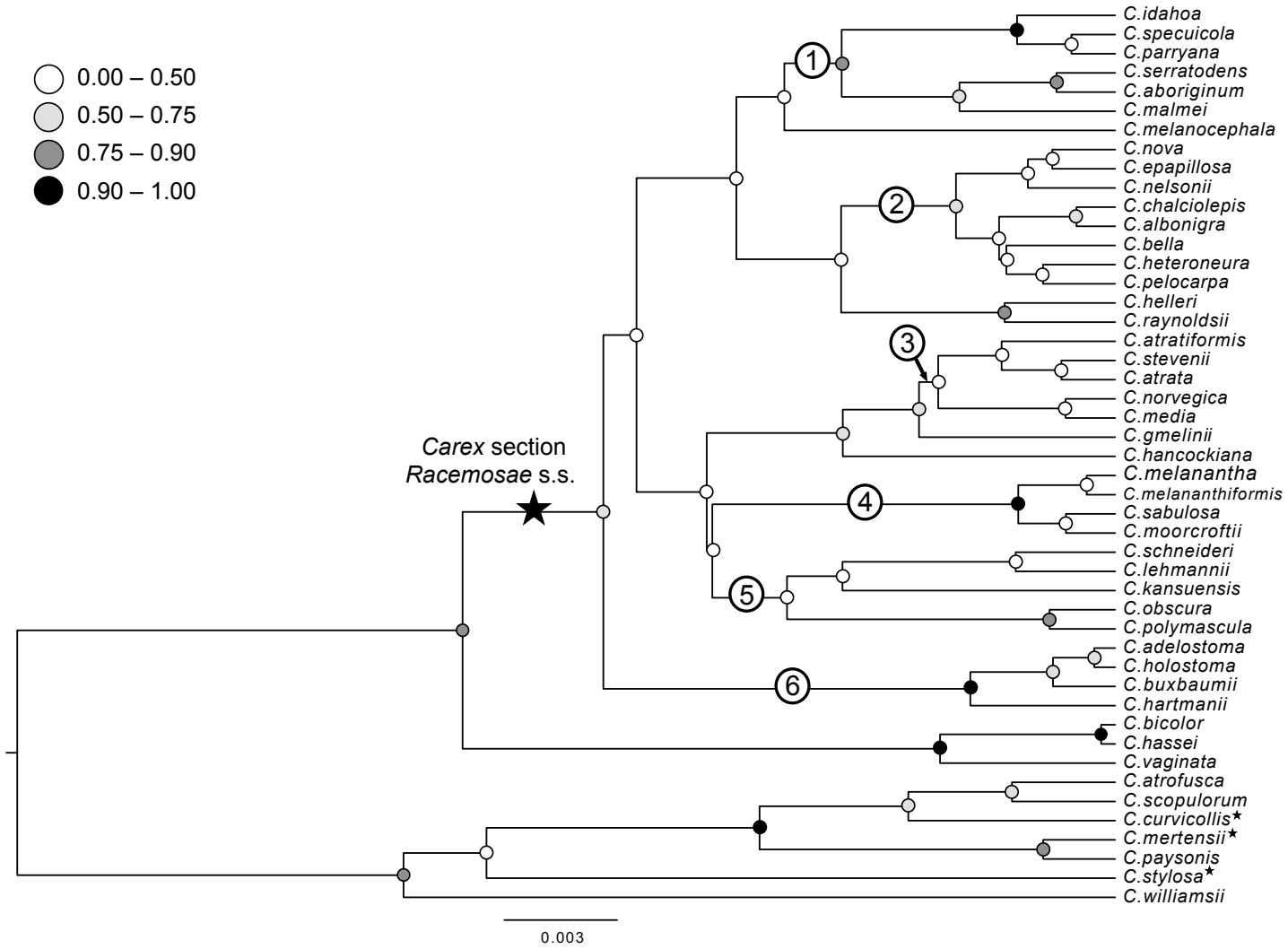

Figure 2.1 Maximum clade credibility tree detailing the relationships of sect. *Racemosae* and closely related taxa, estimated using *BEAST and the reduced Sanger dataset. Shaded circles on nodes represent posterior probabilities estimated from 80,000 post burn-in trees. Numbered circles representing clades discussed in the text are used throughout the figures. The branch subtending sect. *Racemosae* s.s. is starred, as are the sect. *Racemosae* taxa that cluster with taxa representing other sections. See Fig. A.2 for the phylogeny estimated using all taxa for which Sanger data was collected.

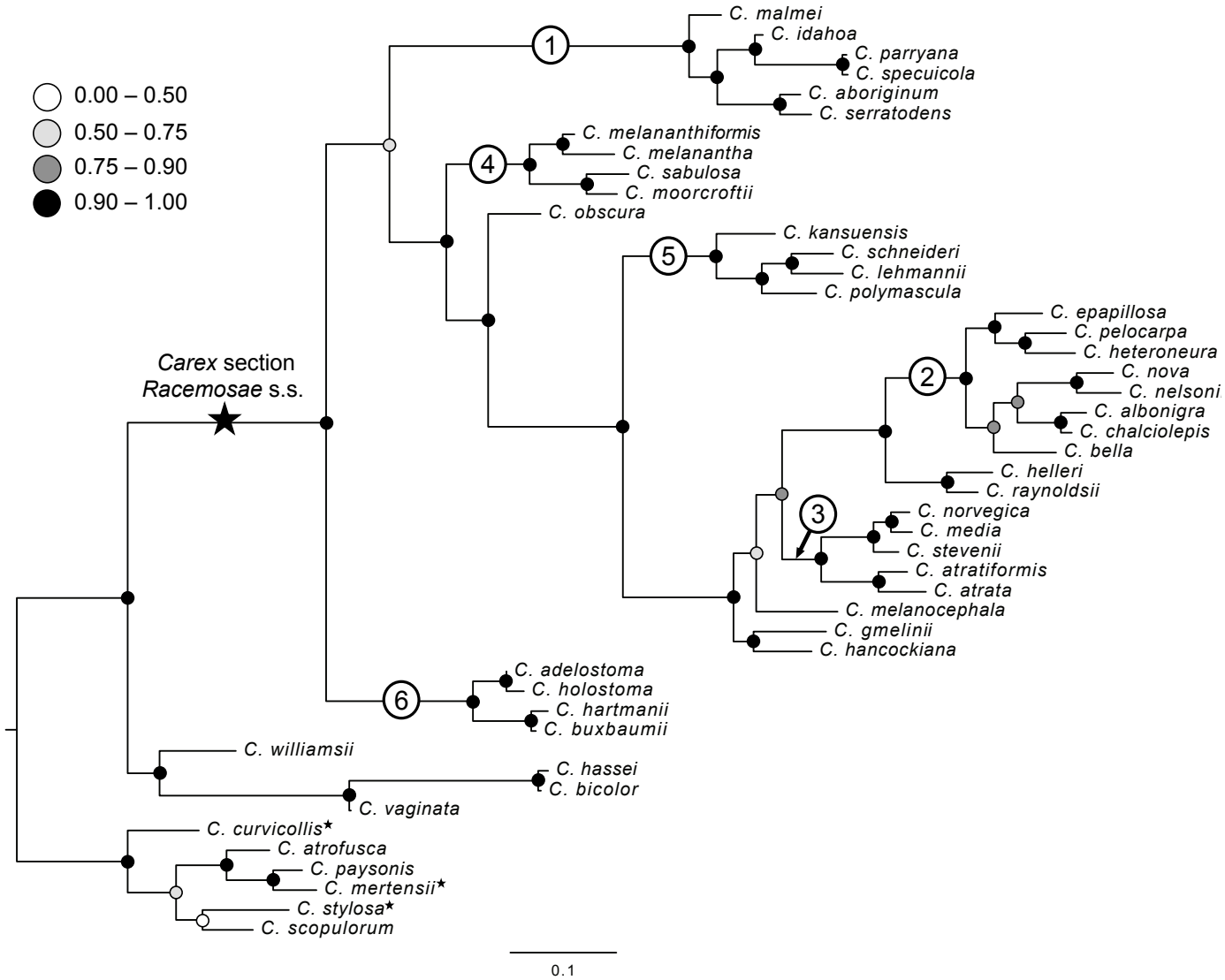

Figure 2.2 The best-scoring maximum likelihood tree estimated by RAxML. Shaded circles on the nodes illustrate nonparametric bootstrap values. Numbered circles correspond to clades depicted in other figures and are discussed in the text. The branch subtending sect. *Racemosae* s.s. is starred, as are the sect. *Racemosae* taxa that cluster with taxa representing other sections.

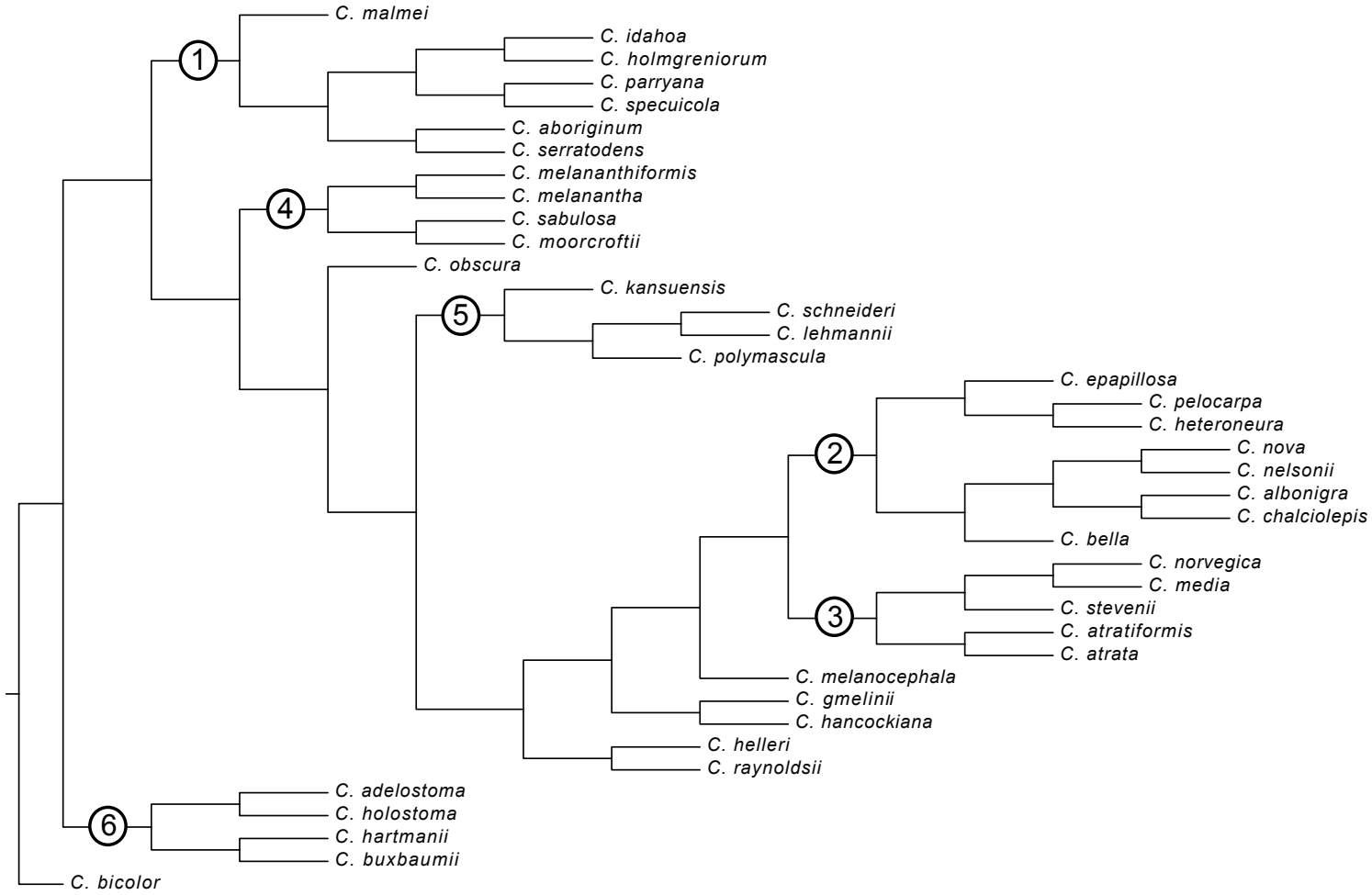

Figure 2.3 Relationships among sect. *Racemosae* species as resolved by the generation of 1E9 quartets in SVDquartets. Numbered circles correspond to clades depicted in other figures and are discussed in the text. All species belong to sect. *Racemosae* s.s. except for one outgroup taxon (*C. bicolor*). Branch lengths are not meaningful.

2.5 Discussion

Our study highlights how next-generation sequencing, and in particular RADseq data, can provide phylogenetic information for resolving relationships among taxa over a variety of temporal depths, including for relationships that lack resolution because traditional Sanger loci are uninformative or morphological characters are ambiguous. Moreover, recently derived clades suggested by Sanger loci, and in some instances morphology, are not only corroborated by SNP data, but relationships within and among these clades are consistent across multiple inference methods. Our analyses suggest that RADseq data have high potential for future phylogenetic research, including instances where conflict among subsets of Sanger loci and/or methods of analyses is commonplace. This is especially heartening in groups where particular biological phenomena that directly contribute to difficulties with phylogenetic analysis may be of interest (e.g., rapid and recent species diversification). In contrast, when using only traditional Sanger loci, it is unclear whether low support for relationships among taxa should be attributed to biological phenomena (e.g., short internodes associated with rapid diversification) versus a paucity of data (e.g., relying on traditional markers that amplify well, but have low mutations rate; Lanier et al., 2014). Finally, our data suggest that herbarium specimens up to 35 years old may by utilized within library construction protocols with little to no modification for specimen age.

Sanger vs. RADseq data for phylogenetic inference—Several patterns stand out when comparing the phylogenetic hypotheses inferred utilizing Sanger loci versus the SNP datasets. First, both types of data and all methods of inference resolved the same recently derived clades (albeit with low support in the Sanger datasets; see numbered clades throughout all figures), supporting the robustness of these relationships and suggesting that both types of data are informative at this phylogenetic depth (see Escudero et al., 2014 for a similar result in *Carex* section *Ovales* Kunth). However, whereas ML and quartet analyses of the SNP datasets resolved largely the same, highly supported topology (Figs. 2.2 and 2.3; see below for differences), no consistent relationships among recently derived clades are apparent among the Sanger datasets (either the full dataset or the reduced dataset that matches the SNP dataset; Fig. 2.1 and Fig. A.2). Similarly, relationships within recently derived clades are labile among analyses on the Sanger datasets, with the exception of a few highly supported nodes that are consistent across

analyses (e.g., *C. serratodens* W. Boott—*C. aboriginum* M.E. Jones and *C. moorcroftii* Falc. ex Boott—*C. sabulosa* Turcz. ex Kunth; Figs. 2.1-2.3).

While phylogenetic inference of the SNP dataset produced a strongly supported and robust hypothesis for sect. *Racemosae,* there are several issues that need special attention before this methodology is broadly applied across clades. Where divergence times between taxa or clades are known to be old, care should be taken to identify and exclude loci that have high mutation rates because they may display extensive homoplasy. Our investigation of mutation rates across SNP loci reinforced the phylogenetic utility of the SNPs with the fastest rates within our dataset, and therefore they were not excluded (data not shown). As mentioned previously, because the taxa in sect. *Racemosae* are almost exclusively alpine and artic, and this flora has only been well established and interconnected since roughly the Pliocene, it is possible that the genetic variation within these taxa has not reached the threshold beyond which homoplasy degrades phylogenetic signal.

Library construction is another important issue to consider when using RADseq data for phylogenetic inference. The number of samples to include in highly multiplexed libraries must be carefully calibrated to ensure sufficient depth is achieved across all taxa. Even with the most careful preparation, significant variability in coverage among taxa may be present, and may necessitate excluding individuals from downstream analyses. In our analyses, nine individuals were excluded; five of these may have suffered from DNA degradation that inhibited sequencing due to their ages (i.e., the specimens were older than 35 years). While our data suggest that herbarium specimens yield high-quality sequences, care must be taken to select specimens that have been appropriately collected and stored. Herbarium specimens older than 35 years may be amenable to RADseq data generation, although more work may be required during specimen selection, DNA extraction, and quality control prior to library construction. While the variability of coverage among taxa and missing data may affect the quality of the phylogenetic inference, it may not be a problem in many cases (see Hovmöller et al., 2013) and excluding data can create a host of potential problems (Huang and Knowles, 2014). Within our SNP datasets there is not a correlation between the amount of loci shared among taxa and the level of support between pairs of taxa or clades, despite high total levels of missing data. For example, the position within the phylogeny of *C. melanocephala* is not well supported (Fig. 2.2), even though it shares relatively many loci with other species (Fig. A.1). Alternatively, the relationships between *C. adelostoma*

and *C. stevenii* (Holm) Kalela and their respective sister taxa are strongly supported, even though they share relatively few loci (Fig. 2.2 and Fig. A.1). In addition, some relationships within Clade 2 are not well supported, even though these taxa are among the best represented within the RADseq dataset (Fig. A.1). That said, there are examples of taxa that share relatively few loci and that also have low support values in their respective clades, for example *C. scopulorum* and *C. stylosa,* which are outgroup species (Fig. 2.2). Overall, better coverage and a more complete dataset should have benefits for creating a well-supported topology, reinforcing the necessity of careful library planning and preparation.

Finally, another complication with using RADseq data for phylogenetic estimation regards the variability of topologies inferred from different subsets of a RADseq dataset. For example, Takahashi et al. (2014) used a range of criteria to filter a RADseq dataset and identify orthologous loci, which resulted in different topologies among 15 species in a ground beetle species flock (see also Wagner et al., 2013; Cruaud et al., 2014). In contrast, our analyses of different datasets identified the same recently derived clades, some of which are also highly supported by the morphological similarity among species (discussed in detail below). In addition, multiple analyses of the SNP dataset largely resolved the same backbone topology and relationships within recently derived clades, suggesting that phylogenetics utilizing SNP data within sect. *Racemosae* may be less prone to these complications. However, while we used parameter values informed by the literature, we acknowledge that a more thorough exploration of how parameter space affects our RADseq dataset may benefit our final phylogenetic hypothesis. Overall, concerns about the information contained within a SNP dataset highlight the importance of investigating data using multiple methodologies and an appropriate dataset, and reinforce the potential utility of resolving discordance within a SNP dataset as a means for generating biologically interesting hypotheses.

General biological phenomena that impact phylogenetic resolution—Backbone nodes with lower support in the ML analysis of the SNP dataset may reflect evolutionary scenarios that facilitated rapid diversification within sect. *Racemosae*. For example, Clade 6 is strongly supported as sister to the remaining sect. *Racemosae* s.s., and it contains species that are predominantly widespread at higher latitudes. Furthermore, all of the clades within the group sister to Clade 6, except for Clade 3, are endemic to single continents (or North America + South

America for Clade 1). If the ancestor of Clade 6 was distributed across high latitudes similar to the extant species, multiple instances of peripatric speciation over a short period of time may have occurred, which is biologically reasonable in habitats heavily disturbed by Pleistocene climatic oscillations (e.g., Hewitt, 1996; Weir and Schluter, 2004; Carstens and Knowles, 2007). This process, in turn, may have created significant conflict and unresolved gene trees, which could manifest as low nodal support among these clades (Huang et al., 2010; Huang et al., 2014). A similar process may be reiterated at a smaller scale within Clade 2, which also contains nodes that have lower support values within the ML analysis of the SNP dataset (Fig. 2.2). In fact, preliminary analyses on a dataset with more individuals from just Clade 2 species suggest these taxa diversified rapidly during the Pleistocene in conjunction with glacial cycles and may be subject to incomplete lineage sorting, in addition to present-day introgression (e.g., Massatti and Knowles, 2014; Massatti, unpublished data).

Phylogenetic relationships of sect. **Racemosae**—The data presented herein suggest that sect. *Racemosae* is a strongly supported, monophyletic clade. However, the boundary of sect. *Racemosae* needs to be modified to exclude *C. augustinowiczii, C. curvicollis, C. mertensii, C. meyeriana,* and *C. stylosa*; these species have been suggested by previous studies (Hendrichs et al., 2004; Waterway et al., 2009) as not belonging with other members of this section. Elucidating the true sectional affiliations of the excluded taxa these will require an analysis of all sections closely related to sect. *Racemosae.* Within sect. *Racemosae* s.s., six clades were consistently resolved throughout our phylogenetic inferences. While relationships among these clades differed across analyses based on the Sanger datasets, a backbone was firmly established and supported in both gene tree and species tree inferences of the SNP datasets. Differences among these analyses reflect only the placement of individuals or pairs of taxa (discussed in detail below).

Clades 1 and 2 are exclusive to the New World. Clade 1 includes multiple species endemic to western North America and one endemic to South America. The North American group (seven species included here, *C. utahensis* Reznicek & D.F. Murray was not sampled) has not been previously included in a subsection within sect. *Racemosae,* though they were thought to be closely related based on morphology (e.g., '*Carex parryana* complex' of Mackenzie, 1935 and Reznicek and Murray, 2013). The South American taxon sampled here was included in

subsect. *Atropictae* G. Wheeler, which also includes the only two other South American sect. *Racemosae* species (*C. atropicta* Steud. and *C. monodynama* (Griseb.) G. Wheeler). These three South American species are very similar morphologically. This study is the first to suggest that subsect. *Atropictae* and the North American '*Carex parryana* complex' should be treated as one clade. Clade 2 contains eight species that are endemic to montane habitats in western North America. Several species, including *C. nelsonii*, *C. nova*, and *C. pelocarpa* F.J. Herm., have previously been included in subsect. *Alpinae* Kalela (Murray, 1969; Egorova, 1999), along with species that, in this study, are either members of Clade 3 or unresolved (e.g., *C. melanocephala*); the remaining Clade 2 species have not been previously included within a subsection. This clade also contains *C. orestera* (not included in the present study), which is morphologically similar to *C. albonigra* Mack. (Zika, 2012), and may include *C. atrosquama*, which groups with Clade 2 in the full *BEAST cladogram (Fig. A.2).

Clade 3 contains a mixture of species that are geographically widespread (*C. atrata* L.) to narrowly endemic (*C. stevenii*). The five species included in this study have not previously been treated as a clade, most commonly being separated into subsect. *Alpinae*, subsect. *Aterrimae* T.V. Egorova, or remaining unplaced (e.g., Egorova, 1999). Unlike Clades 1 and 2, it is unclear which additional taxa belong in this group, but some of the unsampled Asian endemics may cluster with these species (see below for details). The full *BEAST analysis suggests that *C. aterrima* Hoppe (alternatively treated as *C. atrata* ssp. *aterrima* (Hoppe) S.Yun Liang) and *C. parviflora* Host may belong in this group, along with *C. gmelinii* and *C. hancockiana* (see below for details on the alternative placement of the two latter taxa). The ML and quartet analyses of the SNP dataset were in agreement regarding the relationships within this clade.

Clades 4 and 5 are almost exclusively confined to the Asian continent, parallel to the pattern of Clades 1 and 2 being restricted to the New World. Species composing Clade 4 have been previously considered closely related based on morphology (subsect. *Sabulosae* T.V. Egorova; Kreczetovicz, 1935), and the clade likely does not contain additional species. *Carex sabulosa* is the only Clade 4 or 5 species distributed outside of Asia, with a few occurrences in Alaska and Yukon Territory in North America. Clade 5 species have previously been considered members of multiple subsections, including subsect. *Longibracteatae* T.V. Egorova and subsect. *Aterrimae* (Egorova, 1999). Much like Clade 3, it is unclear which additional taxa may be

closely related to these species. The full *BEAST analysis suggests that *C. infuscata* Nees and *C. caucasica* Steven may belong in this group, along with *C. obscura* and *C. melanocephala*.

Finally, Clade 6 is well supported in analyses of both the traditional and SNP datasets (Figs. 2.1-2.3). This clade consistently occupies a basal position within sect. *Racemosae* s.s., except in the *BEAST analysis on the full traditional dataset (Fig. A.2). The close relationship among these species has been previously established (e.g., subsect. *Papilliferae* T.V. Egorova), except for *C. holostoma,* which has been treated as a monospecific subsection (i.e., subsect. *Holostomae* T.V. Egorova). The only additional species that Clade 6 likely includes is *C. tarumensis* Franch., which is scarcely different from *C. buxbaumii* Wahlenb. Both the ML and quartet analyses support the same relationships among Clade 6 species.

Differences between the ML and quartet phylogenies are apparent in the placement of the clade containing *C. helleri* and *C. raynoldsii* as either sister to Group 2 or sister to the combination of Group 2, Group 3, *C. melanocephala*, *C. gmelinii* (previously treated as a monospecific subsect. *Longiaristatae* T.V. Egorova by Egorova, 1999), and *C. hancockiana* (previously treated as subsect. *Longibracteatae* by Egorova, 1999; Figs. 2.2 and 2.3). Both of these taxa are endemic to montane habitat in western North America, although *C. raynoldsii* has a much broader distribution than *C. helleri*. In addition to morphological similarities shared with species included in Clade 2, their distributions suggest that a close relationship with Clade 2 may be more appropriate than a more basal position (i.e., see Fig. 2.3). The placements of *C. obscura, C. melanocephala*, and the clade containing *C. gmelinii* and *C. hancockiana* are also questionable, based on their lability within the phylogeny at low and intermediate levels of quartet sampling by SVDquartets (data not shown), as well as lower bootstrap support for relevant nodes in the ML analysis. The most common alternative topology nested *C. melanocephala* and the *C. gmelinii* - *C. hancockiana* clade within Group 3. *Carex obscura* may have a close relationship with the Asian endemic Clade 5, based on its well supported sister relationship to *C. polymascula* P.C. Li in the analyses on the Sanger datasets (Fig. 2.1 and Fig. A.2).

Relationships among outgroup taxa—Taxa from two sections (*Bicolores* (Tuck. ex L.H. Bailey) Rouy and *Paniceae* G. Don) were resolved as sister to sect. *Racemosae* in all analyses, including *C. bicolor, C. hassei*, and *C. vaginata*; in addition, *C. williamsii* (sect. *Chlorostachyae* Tuck. ex Meinsh.) nested with these taxa only in the ML analysis of the SNP dataset. While sect.

Bicolores is monophyletic within this study, sect. *Paniceae* is not, as *C. livida* (Wahlenb.) Willd. does not cluster with the immediate sister group to sect. *Racemosae* in the full *BEAST analysis (Fig. A.2). This pattern is reiterated among the outgroup taxa, in that no traditionally circumscribed sections appear to be monophyletic, either because the taxa are split among multiple clades, or because of the inclusion of taxa formerly considered to be part of sect. *Racemosae* (current sectional assignments of the outgroup taxa are given in Table A.1). Furthermore, while many authors have considered taxa from sect. *Scitae* Kük. to be very closely related to sect. *Racemosae* (some authors even treat them together; Kreczetovicz, 1935; Mackenzie, 1935), we note here that taxa from sect. *Scitae* are more distantly related than the immediate sister clade to sect. *Racemosae*. Given the lack of support among outgroup taxa in the phylogenies inferred with the Sanger datasets, it may be wise to utilize next-generation sequencing to develop robust hypotheses for these taxa.

Implications of findings for morphological evolution and taxonomy—Our data suggest that *Carex* morphology, both within and among sections, is extremely labile over evolutionary time. For example, despite previous suggestions that *C. nova* and *C. melanocephala* might better be treated as the same species based on morphological similarities (see Murray, 1969), no evidence for a close relationship is suggested by molecular data. Alternatively, this study is the first to treat the South American taxa in Clade 1 as sister to the remaining North American '*Carex parryana'* complex, because morphological traits linking these clades have not previously been identified. However, some recently derived clades (e.g., the North American species in Clade 1, Clade 4, and the majority of Clade 6) are strongly united by morphology, which may suggest that revisiting newly resolved clades and searching for uniting morphological characters will be fruitful.

Given the difficulty in creating monophyletic groups of species based on morphology (e.g., Clades 1-6 often contain species formerly thought to belong in various subsections, or at least not previously grouped together), we do not recommend classifying all but the most morphologically similar taxa (e.g., *C. orestera* in Clade 2*, C. utahensis* in Clade 1*,* and *C. tarumensis* in Clade 6) into one of the clades recovered herein by morphological features alone. That said, the remaining sect. *Racemosae* species not hitherto discussed likely belong in Clades 3, 4, or 5, or will group independently, with *C. obscura, C. melanocephala,* or outside of sect.

Racemosae. These species are mostly Asian endemics and include: *C. alsia* Raymond*, C. aristulifera* P.C. Li*, C. bijianensis* S.Yun Liang & S.R. Zhang*, C. decaulescens* V.I. Krecz., *C. hongyuanensis* Y.C. Tang & S.Yun Liang*, C. macrostigmatica* Kük.*, C. minxianensis* S.Yun Liang*, C. montis-wutaii* T. Koyama*, C. nigerrima* Nelmes*, C. obliquitruncata* Y.C. Tang & S.Yun Liang*, C. oligantha* Steud.*, C. peiktusanii* Kom.*, C. pirinensis* Acht.*, C. praeclara* Nelmes*, C. pseudobicolor* Boeckeler*, C. serreana* Hand.-Mazz.*,* and *C. tatjanae* Malyschev. The last sect. *Racemosae* taxon, *C. urostachys,* will likely not fall within sect. *Racemosae* based on its morphological similarity to *C. mertensii*, which is not included in sect. *Racemosae* s.s.

Conclusion—Next-generation sequencing, in particular RADseq data, provides resolution of the phylogeny of sect. *Racemosae* across multiple taxonomic levels; no previous study within *Carex* utilizing morphological traits or Sanger loci has been able to achieve comparable resolution. In fact, our data suggest that, while morphology may accurately group related species, at least across deeper phylogenetic depths, it can be especially misleading when used to infer more recent relationships. While support values for species' relationships can be low in inferences utilizing the SNP dataset, lower values may also suggest biologically interesting phenomena that affected the taxa over evolutionary time. As next-generation sequencing technologies continue to develop, it will become easier and cheaper to generate large datasets of homologous loci distributed throughout organisms' genomes. Furthermore, refined library construction protocols may alleviate technical problems associated with the variability of coverage and missing data among the specimens included in the library, and may allow researchers to easily utilize older museum specimens. Nevertheless, our study supports the utility of RADseq data for phylogenetic estimation, including for taxa where the lack of well resolved and accurate phylogenies have hampered downstream evolutionary inferences.

2.6 Literature Cited

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

BAIRD, N.A., P.D. ETTER, T.S. ATWOOD, M.C. CURREY, A.L. SHIVER, Z.A. LEWIS, E.U. SELKER, ET AL. 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* 3: e3376.

BOUCKAERT, R., J. HELED, D. KÜHNERT, T. VAUGHAN, C.-H. WU, D. XIE, M.A. SUCHARD, ET AL. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology* 10: e1003537.

BRYANT, D., R. BOUCKAERT, J. FELSENSTEIN, N.A. ROSENBERG, AND A. ROYCHOUDHURY. 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution* 29: 1917–1932.

CARSTENS, B.C., AND L.L. KNOWLES. 2007. Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology* 16: 619–627.

CATCHEN, J., P.A. HOHENLOHE, S. BASSHAM, A. AMORES, AND W.A. CRESKO. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22: 3124–3140.

CAVENDER-BARES, J., D.D. ACKERLY, D.A. BAUM, AND F.A. BAZZAZ. 2004. Phylogenetic Overdispersion in Floridian Oak Communities. *The American Naturalist* 163: 823–843.

CHIFMAN, J., AND L. KUBATKO. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317–3324.

CRONQUIST, A. 1981. An integrated system of classification of flowering plants. Columbia University Press, New York, New York, USA.

CRUAUD, A., M. GAUTIER, M. GALAN, J. FOUCAUD, J. SAUNE, G. GENSON, E. DUBOIS, ET AL. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. *Molecular and Biological Evolution* 32: 1272–1274.

DEGNAN, J.H., AND N.A. ROSENBERG. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

DONOGHUE, M. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences, USA* 105: 11549-11555.

EATON, D.A.R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30: 1844–1849.

EATON, D.A.R., AND R.H. REE. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology* 62:

689–706.

EDGAR, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.

EDGAR, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26: 2460–2461.

EGOROVA, T. V. 1999. The sedges (*Carex* L.) of Russia and adjacent states (within the limits of the former USSR). St. Petersburg State Chemical-Pharmaceutical Academy, St. Petersburg, Russia, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.

ESCUDERO, M., D.A.R. EATON, M. HAHN, AND A.L. HIPP. 2014. Genotype-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution* 79: 359-367.

HELICONIUS GENOME CONSORTIUM. 2013. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–98.

HENDRICHS, M., F. OBERWINKLER, D. BEGEROW, AND R. BAUER. 2004. *Carex*, subgenus *Carex* (Cyperaceae) - A phylogenetic approach using ITS sequences. *Plant Systematics and Evolution* 246: 89–107.

HEWITT, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.

HILU, K., T. BORSCH, K. MULLER, D. SOLTIS, P. SOLTIS, V. SAVOLAINEN, M. CHASE, ET AL. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758.

HIPP, A.L., A.A. REZNICEK, P.E. ROTHROCK, AND J.A. WEBER. 2006. Phylogeny and classification of *Carex* section *Ovales* (Cyperaceae). *International Journal of Plant Sciences* 167: 1029–1048.

HIPP, A.L., D.A.R. EATON, J. CAVENDER-BARES, E. FITZEK, R. NIPPER, AND P.S. MANOS. 2014. A Framework Phylogeny of the American Oak Clade Based on Sequenced RAD Data. *PLoS ONE* 9: e93975.

HOVMOELLER, R., L.L. KNOWLES, AND L.S. KUBATKO. 2013. Effects of missing data on species tree estimation under the coalescent. *Molecular Phylogenetics and Evolution* 69: 1057–1062.

HUANG, H., AND L.L. KNOWLES. 2014. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology*. In Press.

HUANG, H., L.A.P. TRAN, AND L.L. KNOWLES. 2014. Do estimated and actual species phylogenies match? Evaluation of East African cichlid radiations. *Molecular Phylogenetics and Evolution* 78: 56–65.

HUANG, H., Q. HE, L.S. KUBATKO, AND L.L. KNOWLES. 2010. Sources of Error Inherent in Species-Tree Estimation: Impact of Mutational and Coalescent Effects on Accuracy and Implications for Choosing among Different Methods. *Systematic Biology* 59: 573–583.

KING, M.G., AND E.H. ROALSON. 2008. Exploring Evolutionary Dynamics of nrDNA in *Carex* Subgenus *Vignea* (Cyperaceae). *Systematic Botany* 33: 514–524.

KNOWLES, L.L. 2009. Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics* 40: 593–612.

KRAFT, N.J.B., W.K. CORNWELL, C.O. WEBB, AND D.D. ACKERLY. 2007. Trait Evolution, Community Assembly, and the Phylogenetic Structure of Ecological Communities. *The American Naturalist* 170: 271–283.

KRECZETOVICZ, V. I. 1935. *Carex* L. *In* V. L. Komarov [ed.], Flora URSS, vol. 3, 114-464. Academy of Science, Leningrad, URSS.

KUBATKO, L.S., B.C. CARSTENS, AND L.L. KNOWLES. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25: 971–973.

LANIER, H.C., H. HUANG, AND L.L. KNOWLES. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. *Molecular Phylogenetics and Evolution* 70: 112–119.

LIU, L., L. YU, D.K. PEARL, AND S.V. EDWARDS. 2009. Estimating Species Phylogenies Using Coalescence Times among Sequences. *Systematic Biology* 58: 468–477.

LIU, L., L. YU, AND S.V. EDWARDS. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* 10: 302.

LUNKAI, D., L. SONGYUN, Z. SHUREN, T. YANCHENG, T. KOYAMA, AND C.G. TUCKER. 2010. Flora of China, vol. 23 (Acoraceae through Cyperaceae). Missouri Botanical Garden Press, St. Louis, Missouri, USA.

LYNCH, M. 2008 Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Molecular Biology and Evolution* 25: 2409–2419.

MACKENZIE, K.K. 1931-1935. Cyperaceae. North American Flora, vol. 18, 1-478. New York Botanical Garden, New York, New York, USA.

MADDISON, W.P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

MASSATTI, R., AND L.L. KNOWLES. 2014. Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution* 68: 2833–2846.

MATTHEWS, J.V. 1979. Tertiary and Quaternary environments: Historical background for an

analysis of the Canadian insect fauna. *In* H.V. Danks [ed.], Canada and its Insect fauna, 31-86. Entomological Society of Canada, Ottawa, Canada.

MATTHEWS, J.V., AND L.E. OVENDEN. 1990. Late Tertiary plant macrofossils from localities in Arctic/Subarctic North America: a review of the data. *Arctic* 43: 364–392.

MILLER, M.R., J.P. DUNHAM, A. AMORES, W.A. CRESKO, AND E.A. JOHNSON. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240–248.

MURRAY, D.F. 1969. Taxonomy of *Carex* sect. *Atratae* (Cyperaceae) in the southern Rocky Mountains. *Brittonia* 21: 55–76.

MURRAY, D.F. 1995. Causes of arctic plant diversity: origin and evolution. *In* F. S. Chapin and C. Koerner [eds.], Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences*,* 21-32. Springer, Heidelberg, Germany.

MURRAY, D.F. 2002. *Carex* L. sect. *Racemosae* G. Don. *In* Flora of North America Editorial Committee [eds.], Flora of North America, north of Mexico, vol. 23, 401-414. Oxford University Press, New York, New York, USA.

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.

R CORE TEAM. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. website: http://www.R-project.org/.

RAMBAUT, A., M.A. SUCHARD, D. XIE, AND A.J. DRUMMOND. 2014. Tracer v1.6. Computer program and documentation distributed by the author, website: http://beast.bio.ed.ac.uk/Tracer [accessed 1 July 2014].

REZNICEK, A.A. 1990. Evolution in sedges (*Carex*, Cyperaceae). *Canadian Journal of Botany* 68: 1409–1432.

REZNICEK, A.A. AND D.F. MURRAY. 2013. A Re-evaluation of *Carex specuicola* and the *Carex parryana* complex (Cyperaceae). *Journal of the Botanical Research Institute of Texas 7*: 37-51.

ROALSON, E., J. COLUMBUS, AND E. FRIAR. 2001. Phylogenetic relationships in Cariceae (Cyperaceae) based on ITS (nrDNA) and trnT-LF (cpDNA) region sequences: Assessment of subgeneric and sectional relationships in *Carex* with emphasis on section *Acrocystis*. *Systematic Botany* 26: 318–341.

ROWE, H.C., S. RENAUT, AND A. GUGGISBERG. 2011. RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* 20: 3499–3502.

RUBIN, B.E.R., R.H. REE, AND C.S. MOREAU. 2012. Inferring Phylogenies from RAD Sequence Data. *PLoS ONE* 7: e33394.

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

STARR, J., AND B. FORD. 2009. Phylogeny and Evolution in Cariceae (Cyperaceae): Current Knowledge and Future Directions. *The Botanical Review* 1–28.

STARR, J., S. HARRIS, AND D. SIMPSON. 2003. Potential of the 5' and 3' Ends of the Intergenic Spacer (IGS) of rDNA in the Cyperaceae: New Sequences for Lower-Level Phylogenies in Sedges with an Example from *Uncinia* Pers. *International Journal of Plant Sciences* 164: 213–227.

STARR, J.R., R.F.C. NACZI, AND B.N. CHOUINARD. 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). *Molecular Ecology Resources* 9: 151–163.

TAKAHASHI, T., N. NAGATA, AND T. SOTA. 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution* 80: 137–144.

TAKHTAJAN, A. 1987. Systema magnoliophytorum. Editoria Nauka, Leningrad (St. Petersburg), Russia.

TAMURA, K., G. STECHER, D. PETERSON, A. FILIPSKI, AND S. KUMAR. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

TKACH, N.V., M.H. HOFFMANN, M. RÖSER, A.A. KOROBKOV, AND K.B. VON HAGEN. 2008. Parallel evolutionary patterns in multiple lineages of arctic *Artemisia* L. (Asteraceae). *Evolution* 62: 184–198.

WAGNER, C.E., I. KELLER, S. WITTWER, O.M. SELZ, S. MWAIKO, L. GREUTER, A. SIVASUNDAR, AND O. SEEHAUSEN. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787–798.

WATERWAY, M.J., T. HOSHINO, AND T. MASAKI. 2009. Phylogeny, Species Richness, and Ecological Specialization in Cyperaceae Tribe Cariceae. *The Botanical Review* 75: 138–159.

WEI, T. 2015. Package 'corrplot'. *Statistician* 56: 316-324.

WEIR, J.T., AND D. SCHLUTER. 2004. Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society B: Biological Sciences* 271: 1881–1887.

WHITE, T.J., T.D. BRUNS, S.B. LEE, AND J.W. TAYLOR. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* M.A. Innis, D.H. Gelfand,

J.J. Sninsky, and T.J. White [eds.], PCR protocols: A guide to methods and applications, 315–322. Academic Press, San Diego, California, USA.

WIENS, J.J., D.D. ACKERLY, A.P. ALLEN, B.L. ANACKER, L.B. BUCKLEY, H.V. CORNELL, E.I. DAMSCHEN, ET AL. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13: 1310–1324.

ZIKA, P.F. 2012. *Carex orestera* (Cyperaceae), a New Sedge from the Mountains of California. *Novon: A Journal for Botanical Nomenclature* 22: 118–124.

CHAPTER 3

The influence of Pleistocene climatic oscillations on plant diversification

3.1 Introduction

Plants distributed throughout high elevations and northerly latitudes are adapted to harsh climates and habitats that have been repeatedly disturbed by Pleistocene climatic oscillations. For example, during the majority of the Pleistocene, cooler climates facilitated widespread continental glaciers at high latitudes (Dyke and Prest, 1987), a land bridge between Asia and North America (Hopkins, 1967), and valley glaciers in many mountain ranges unaffected by continental glaciers (Richmond, 1986; Hewitt, 1996), which caused local and regional plant extirpations and the reestablishment of populations in newly available habitat. These periods were interrupted by warmer, interglacial periods (Siegenthaler et al., 2005), during which distributions were generally restricted to higher elevations for montane plants and higher latitudes for arctic plants (Prentice et al., 2000). Increased levels of intraspecific molecular variation, as indicated by organisms' phylogeographic patterns (Avise, 2000; Hewitt, 2000; Soltis et al., 2006), and greater local species richness, as witnessed by regional patterns of arctic and montane biodiversity (Hultén, 1937; Weber, 1965; Murray, 1995), illustrate the consequences of shifting climates on these habitats.

Based upon paleoclimate reconstructions and geologic evidence, the montane and arctic floras developed recently. The ancestral montane flora likely originated when locally distributed plants adapted to harsher climates at higher elevations as global temperatures dropped from the mid-Eocene onwards (Moran et al., 2006). The arctic flora, which likely appeared in the late Pliocene contemporaneously with the expansion of the Greenland ice shield (about 3.2 million years ago) (Matthews and Ovenden, 1990), was not continuously distributed in a circumarctic belt until about 3 million years ago (Matthews, 1979). It was likely derived from the ancestral stocks of montane plants occupying the high mountains in Asia and North America as cooler

climates brought these two habitats in closer proximity (Hultén, 1937; Weber, 1965; Hedberg, 1992; Murray, 1995). An additional source may have been elements of the Arcto-Tertiary flora that occupied bog, riparian, and upland habitats (Murray, 1995). The origin of these floras and the history of paleoclimatic events suggest that widespread species and clades must have attained their distributions after arctic habitat became continuous in the Pliocene. While high elevation and high latitude habitats require similar adaptations to cold, harsh climates, their floras are not directly interchangeable; for example, plants within these habitats require distinctive metabolic tolerances to cope with differences in daily temperature fluctuations during the growing season (Körner, 2003). Some widespread species at high latitudes have locally adapted populations in montane environments at more southerly latitudes (e.g., Mooney and Billings, 1961), while other clades have species that are adapted to one habitat or the other (Conti et al., 1999).

Here I investigate how the dynamic nature of the Pleistocene climate facilitated the creation of biodiversity using *Carex* section *Racemosae,* which contains about 70 species primarily restricted to habitats at northerly latitudes and within the montane regions of eastern Asia and western North America. Species' ranges vary from narrowly endemic to widespread, and recently derived clades generally contain species with similar habitat preferences (see Chapter 2 and Murray, 2002). Illuminating the impact of the dynamic climate during the Pleistocene on dispersal and diversification within this clade of sedges will be generally informative for the study of biodiversity patterns within montane and arctic ecosystems.

3.2 Materials and Methods

I utilized two of the phylogenetic hypotheses developed in Chapter 2 to investigate the influence of Pleistocene climatic oscillations on patterns of diversification. Because a general substitution rate for the internal transcribed spacer (ITS) nuclear region has been estimated across many plant clades (Kay et al., 2006), I used the species tree estimated by *BEAST to determine the general timing of diversification events in sect. *Racemosae*. While this phylogeny is not well resolved compared to the phylogeny created using SNP data, it groups taxa into the same intermediate-level clades as the SNP dataset, and thus provides a means to generally estimate the timing of sect. *Racemosae* diversification events. Branch lengths in the *BEAST species tree were scaled proportionally to an average substitution rate for herbaceous plants (4E-9 substitutions/site/year), as well as to an exceptionally slow substitution rate comparable to what

is generally associated with woody plants (2E-9 substitutions/site/year), to determine how rate variation may affect the general inference.

To determine the distributions of ancestral taxa within sect. *Racemosae*, I used the fully resolved and well supported phylogeny inferred with RAxML and SNP data (see Chapter 2) in conjunction with LAGRANGE (Ree and Smith, 2008), which implements dispersal-extinction-cladogenesis models of geographic range evolution. The phylogeny inferred with RAxML was nearly identical to the phylogeny created using SVDquartets (a SNP-based species tree methodology that should generate a more trustworthy result compared to RAxML), but it included branch lengths necessary for the maximum likelihood algorithm implemented in LAGRANGE. The RAxML phylogeny was transformed into a chronogram using the 'chronos' function in the 'ape' package (Paradis et al., 2004) in R (R Core Team, 2012). Changing the value of lambda between 0 and 1 in the chronos function did not impact the ancestral state reconstruction (data shown only for lambda = 1). LAGRANGE CONFIGURATOR v20130526 (http://www.reelab.net/lagrange/configurator) was used to configure the settings for the ancestral state reconstruction. Species' present distributions were either constrained to North America or Asia, or they occupied multiple continents (any combination of North America, Asia, and Europe); present distributions were determined using the sect. *Racemosae* treatments in Flora of North America (Murray, 2002) and Flora of China (Lunkai et al., 2010), and The Sedges (*Carex* L.) of Russia and Adjacent States (Egorova, 1999). No range constraints were imposed, and dispersal events between Europe and North America were assumed to be slightly less frequent (0.75) compared to dispersal between any other pairs of continents (1.0) because discrete land bridges were not formed in the Pleistocene (Hewitt, 1996).

3.3 Results

Using the average ITS substitution rate for herbaceous plants, the median values of the posterior distributions for the ancestral diversification events of recently derived clades (numbered clades in Figs. 3.1 and 3.2) occurred during the Pleistocene. Furthermore, only Clade 5 had a 95% highest posterior density whose range included times substantially older than the Pleistocene (Fig. 3.1). When using the substantially slower substitution rate for woody plants, the median values of the posterior distributions for three diversification events were pulled into the Pliocene (two in Clade 1 and one in Clade 5). However, the remaining diversification events

were still nested within the Pleistocene. These data suggest that much of the diversity within sect. *Racemosae* was generated during the Pleistocene.

Ancestral range reconstructions are unambiguous for many of the nodes in the SNP-based phylogeny (Fig. 3.2). Ancestral ranges associated with Clades 1, 2, 4, and 5, whose taxa are almost exclusively endemic to either Asia or North America, are restricted to either Asia or North America. Clade 3 contains a mixture of North American endemics and widespread species (species distributed on two or more continents), and Clade 6 contains exclusively widespread taxa, resulting in more uncertainty within ancestral ranges. For some of these nodes, a widespread ancestral range is equally or more probable than an ancestor distributed across only one continent. While the uncertainty affects nodes within these clades, it remains locally isolated and only affects the most closely related ancestors, if at all (i.e., see relationships below Clade 3 in Fig. 3.2). However, uncertainty is also evident at the most basal ancestral node, in which an Asian or North American distribution is almost equally probable.

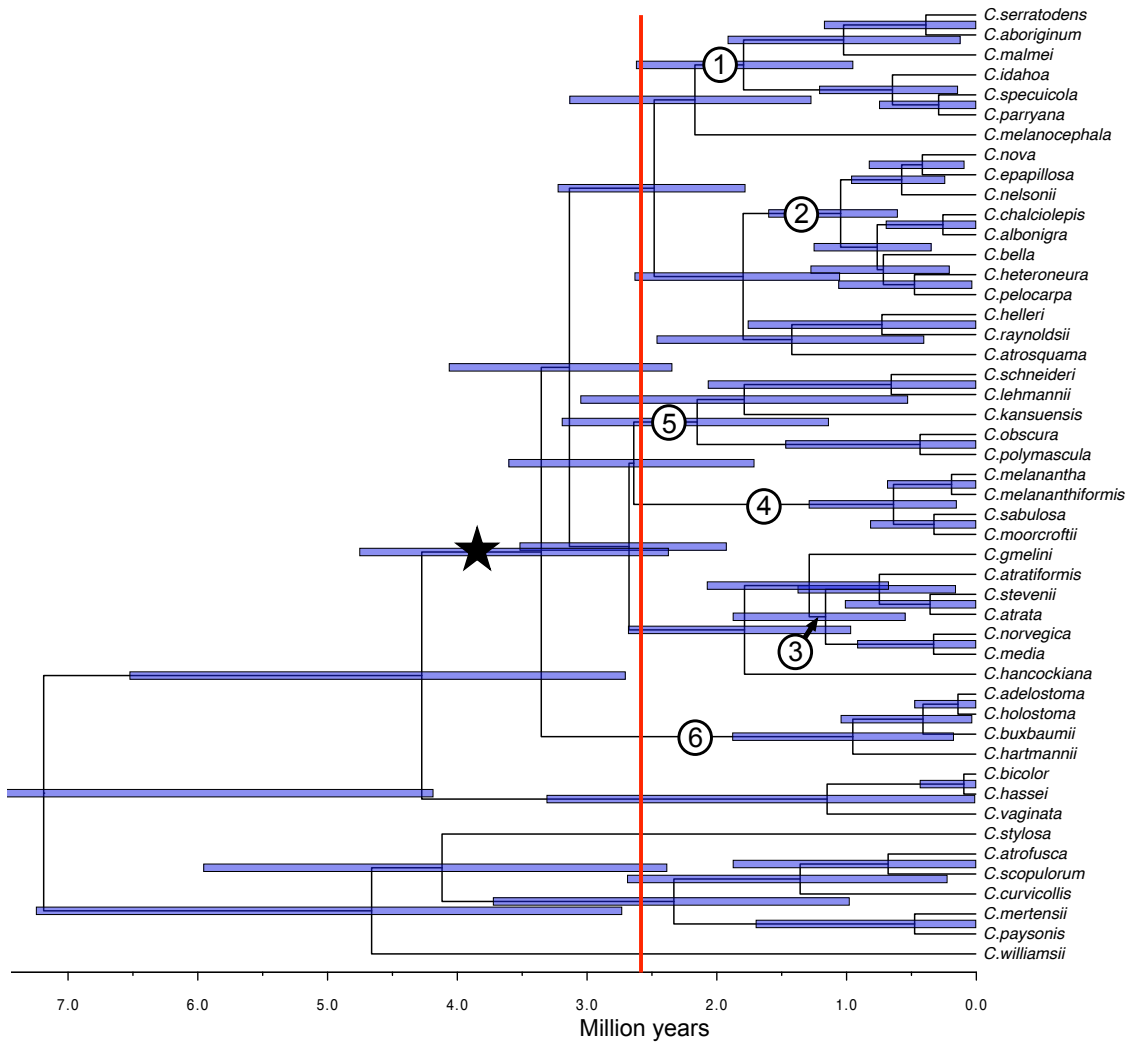

Figure 3.1 Time-calibrated phylogeny for *Carex* sect. *Racemosae* created using *BEAST (see Chapter 2) and an average ITS substitution rate for herbaceous plants (4E-9). The red line indicates the beginning of the Pleistocene and the star denotes sect. *Racemosae*. Bars on the nodes are the interval of 95% parameter regions with the highest posterior density. Numbered clades are discussed in the text and demarcate the same clades as in Fig. 3.2 and Chapter 2.

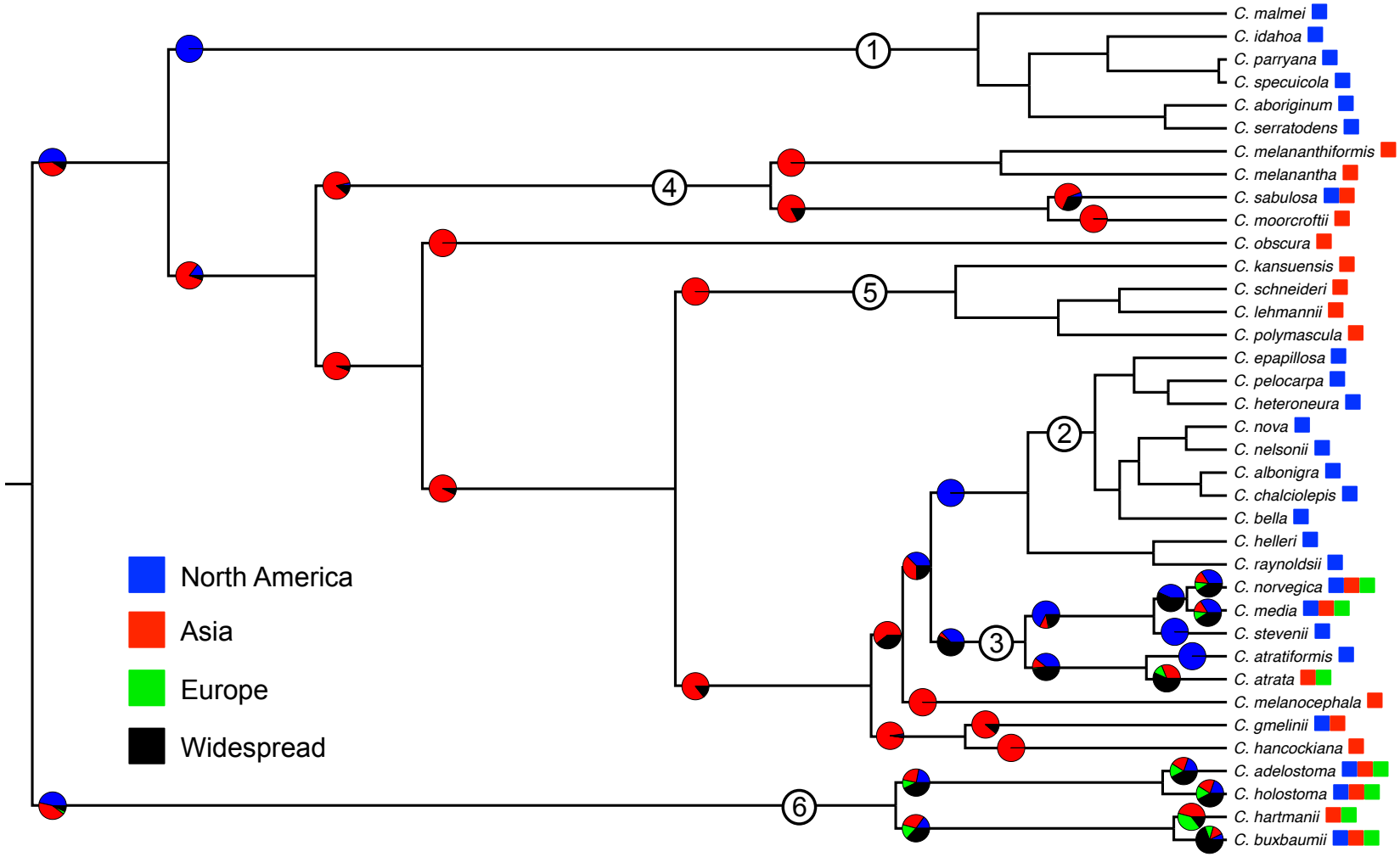

Figure 3.2 Ancestral range reconstruction for *Carex* sect. *Racemosae,* based on a phylogeny estimated with RAxML and a concatenated SNP dataset (see Chapter 2). Pie graphs show the relative likelihoods for alternative ancestral distributions. The 'Widespread' distribution includes any combination of two or more continents. Present distributions for the taxa are shown by colored squares next to the tip labels. Pie graphs are not shown on branches when all descendants and present-day taxa have the same distribution as their ancestor. Numbered clades are discussed in the text and demarcate the same clades as in Fig. 3.1 and Chapter 2.

3.4 Discussion

Shifting climates had a profound influence on landscapes at high elevations and northerly latitudes during the Pleistocene, and here I utilize a clade of sedges predominantly restricted to these habitats to explore the interplay between climatic oscillations and plant dispersal and diversification. This investigation relies upon having a well-sampled, diverse clade for which relationships among taxa are resolved, and it is useful for generating hypotheses about the specific factors influencing diversification in cold-adapted clades.

Clade diversification and dispersal dynamics during the Pleistocene – Many of the diversification events within *Carex* section *Racemosae* occurred contemporaneously with Pleistocene climatic oscillations, including all of the events within the recently derived clades (Fig. 3.1). Significantly slower rates of molecular evolution do not change the high correlation between diversification and the Pleistocene. While I do not test whether climatic oscillations and the shifting of suitable habitat throughout landscapes caused these diversification events, this mechanism has been supported in plants as well as in other organisms. For example, Levsen et al. (2012) utilized coalescent simulations and measures of niche overlap to provide evidence that two keystone tree species important in North American ecosystems, *Populus balsamifera* and *P. trichocarpa,* diverged approximately 75,000 years ago, corresponding to a period of widespread glaciation. Knowles (2000) determined that ten species of grasshoppers adapted to sky islands in the northern Rocky Mountains originated within the last 1.7 million years, coinciding with multiple rounds of regional glaciation. Finally, Weir and Schluter (2004) determined that divergence between multiple pairs of boreal bird species date to the Pleistocene, linking fragmentation of the boreal forest by continental ice sheets to the generation of regional endemism.

This dataset may be especially relevant for deciphering the influence of Pleistocene climatic oscillations on plant dispersal across northerly latitudes, for example between Asia and North America via the Bering Land Bridge (Simpson, 1953; Waltari et al., 2007; Tkach et al., 2008). A connection between Asia and North America was important in structuring montane floras at lower latitudes, as evidenced by the floristic similarity of the southern Rocky Mountains in North America and the Altai Mountains in Asia, which share evolutionary lineages in more than 140 genera (Weber, 2003). However, the dynamics of plants' movements through Beringia

may have been biased due to the glacial dynamics on the Asian and North American continents (Simpson, 1953; Rausch, 1994). Briefly, the localized continental glaciers in Asia, unlike the widespread continental glaciers at high latitudes across North America, may have facilitated an Asian bias in the identity of the Beringian flora due to longer periods of connectivity between high latitude habitats. Subsequently, the Asian-biased Beringian flora would have been optimally situated to colonize the abundant, open habitat available as North American continental glaciers melted during interglacial periods. In addition, whether dispersal events were frequent or rare between continents is not well investigated, especially for montane plants, due to the lack of phylogenetic hypotheses of diverse groups distributed throughout these regions. Dispersal events followed by vicariance and limited to no diversification have occurred many times within genera adapted to lower elevation and steppe habitats (Tkach et al., 2008; Lu et al., 2011).

Within sect. *Racemosae*, dispersal from Asia to North America, likely facilitated by the Bering Land Bridge, is evident in *C. gmelinii* and *C. sabulosa*. Both of these taxa originated in the Pleistocene (Fig. 3.1), have ancestors with Asian distributions (Fig. 3.2), and presently have populations at higher latitudes in North America. A similar scenario may have occurred in Clade 2, but the timing of the ancestral dispersal event into North America is unclear based on the time calibration presented here (i.e., it is very close to the Pliocene-Pleistocene transition). In addition, our data suggest several instances of species restricted to one continent becoming widespread during the Pleistocene. For example, the ancestor of Clade 6 is inferred to be either Asian or North American, while all of the extant taxa are distributed on two or more continents (Fig. 3.2), and the diversification events occurred during the Pleistocene (Fig. 3.1). Clade 3 also contains ancestral taxa predicted to have North American distributions, while extant taxa are more widely distributed (e.g., see *C. stevenii, C. norvegica,* and *C. media,* and their respective ancestors, Fig. 3.2). Widely distributed species may have dispersed across the Bering Land Bridge, though evidence also indicates that dispersal between North America and Europe may have been a viable route (Abbott et al., 2000).

While dispersal between continents occurred during the Pleistocene, it seems to have been more common for species adapted to high latitudes, compared to those adapted to montane environments (i.e., high elevations). Mountainous regions have never been continuous for the dispersal of montane organisms between Asia and North America, although habitat in which montane species may have survived was brought into closer proximity during glacial periods

(Simpson, 1953; Murray, 1995). However, habitat for arctic plants has been continuously present at similar elevations and latitudes since the inception of this flora, although connectivity among tundra habitats varied through time (e.g., the availability of the Bering Land Bridge). Of the five clades containing species adapted to montane environments, four clades (1, 2, 4, and 5) diversified in and are restricted to their respective continents (one exception is *C. sabulosa* in Clade 4, which is primarily Asian, but with populations in Alaska and Yukon Territory). The fifth clade (Clade 3) is composed of a mixture of a North American montane species (*C. stevenii*) and species generally restricted to high latitudes (e.g., *C. norvegica* and *C. media*), further supporting the limited dispersal ability of montane taxa. Together, these data suggest that montane diversification within sect. *Racemosae* relied upon ancestors that were adapted to high latitude habitats.

An alternative evolutionary hypothesis – The dataset presented herein suggests that diversity within sect. *Racemosae* was generated as an ancestral taxon endemic to a single continent diversified *in situ,* with infrequent dispersal events to and subsequent diversification in an adjacent continent (whether Asia or North America is ancestral is uncertain) (Fig. 3.2). More recently, taxa within certain clades have attained more widespread distributions and/or further diversified. However, the ancestral range reconstruction employed here is insufficient for distinguishing between this scenario and one in which a widespread ancestor, potentially closely related to the extant taxa in Clade 3, gave rise to the recently derived clades. In this scenario, populations of the widespread ancestral taxon may have been repeatedly isolated on the Asian and North American continents due to Pleistocene glaciations, facilitating adaptation to montane or closely related habitats, and in some cases generating reproductive isolation. These newly derived taxa may then have gone on to diversify within montane habitats to create the diversity we see in Clades 1, 2, 4, and 5. Phylogeographic studies have revealed Pleistocene survival of arctic-adapted species in southern refugia and subsequent migration of populations to higher latitudes after the retreat of the ice sheets (Hewitt, 1996; Abbott et al., 2000; Hewitt, 2000), as well as instances where southern populations gained reproductive isolation (e.g., Conti et al., 1999). The ancestral range reconstruction presented here may fail to identify this scenario because Clade 3 is nested so deeply within the phylogeny that the widespread distribution of the extant taxa (i.e., the taxa displaying the ancestral range) is overwhelmed by the influence of the

clades restricted to single continents. Clades' climatic tolerances may lend support to this alternative hypothesis if, for example, taxa within Clade 3 encompass the entire range of climatic variation present within the clades adapted to single continents (i.e., Clades 1, 2, 4, and 5).

Conclusion – *Carex* sect. *Racemosae* clades accrued diversity during the Pleistocene in both high elevation montane habitat and habitat at high latitudes. These habitats were highly affected by Pleistocene climatic fluctuations, which have been shown to facilitate diversification in other groups of organisms. The ancestral range reconstruction presented here suggests that species dispersed from Asia to North America, likely via the Bering Land Bridge, as well as from single to multiple continents, likely by various dispersal routes. There is no evidence that montane species dispersed between continents; instead, ancestral montane taxa diversified *in situ*. Future investigations should attempt to determine whether Clade 3, which contains both widespread species and continental endemics, gave rise to the other recently derived clades, potentially as widespread ancestors became isolated at southern latitudes on continents, and subsequently diversified locally.

3.5 Literature Cited

Abbott, R.J., et al. 2000. Molecular analysis of plant migration and refugia in the Arctic. *Science*. 289(5483): 1343-1346.

Avise, J. C. 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, Massachusetts, USA.

Conti, E., et al. Phylogenetic Relationships of the Silver Saxifrages (*Saxifraga* Sect. *Ligulatae* Haworth): Implications for the Evolution of Substrate Specificity, Life Histories, and Biogeography. *Molecular Phylogenetics and Evolution*. 13(3): 536-555.

Dyke, A.S., and V.K. Prest, 1987. Late Wisconsinan and Holocene History of the Laurentide Ice Sheet. *Geographie physique et Quaternaire.* 41(2): 237-263.

Egorova, T.V. 1999. The sedges (*Carex* L.) of Russia and adjacent states (within the limits of the former USSR). Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Hedberg, K.O. 1992. Taxonomic differentiation in *Saxifraga hirculus* L. (Saxifragaceae) – a circumpolar Arctic-Boreal species of Central Asiatic origin. *Botanical Journal of the Linnean Society*. 58: 247-276.

Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*. 58: 247-276.

Hewitt, G.M. 2000. The genetic legacy of the Quaternary ice ages. *Nature*. 405(6789): 907–913.

Hopkins, D.M., ed. 1967. The Bering Land Bridge. Stanford University Press, Stanford, California, USA.

Hultén, E. 1937. Outline of the history of arctic and boreal biota during the Quaternary period: their evolution during and after the glacial period as indicated by the equiformal progressive areas of present plant species, Bokfoerlags Aktiebolaget Thule.

Kay, K.M., et al. 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology*. 6:36.

Knowles, L.L. 2000. Tests of Pleistocene Speciation in Montane Grasshoppers (Genus *Melanoplus*) from the Sky Islands of Western North America. *Evolution*. 54(4): 1337-1348.

Körner, C. 2003. Alpine plant life: functional plant ecology of high mountain ecosystems. Springer, Berlin, Germany.

Levsen, N.D., et al. 2012. Pleistocene Speciation in the Genus *Populus* (Salicaceae). *Systematic Biology*. 61(3): 401-412.

Lu, J., et al. 2011. Biogeographic disjunction between eastern Asia and North America in the *Adiantum pedatum* complex (Pteridaceae). *American Journal of Botany.* 98(10): 1680-1693.

Lunkai, D., et al. 2010. Flora of China, vol. 23 (Acoraceae through Cyperaceae). Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Matthews, J.V. 1979. Tertiary and Quaternary environments: Historical background for an analysis of the Canadian insect fauna. In: *Canada and its Insect Fauna* (ed. Danks, H.V.). Pages 31-86. Entomological Society of Canada, Ottawa, Canada.

Matthews, J.V., and L.E. Ovenden. 1990. Late Tertiary plant macrofossils from localities in arctic, sub-arctic North America – a review of the data. *Arctic*. 43: 364-392.

Mooney, H.A., and W.D. Billings. 1961. Comparative Physiological Ecology of Arctic and Alpine Populations of *Oxyria digyna*. *Ecological Monographs.* 31(1): 1-29.
Moran, K., et al. 2006. The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature*. 441: 601-605.
Murray, D.F. 1995. Causes of arctic plant diversity: origin and evolution. In: *Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences* (eds. Chapin, F.S., and Körner, C.). Pages 21-32. Springer, Heidelberg, Germany.
Murray, D.F. 2002. *Carex* sect. *Racemosae*. Pages 401-414 in Flora of North America Editorial Committee, eds. 1993+ *Flora of North America North of Mexico*. 16+ vols. New York and Oxford. Vol. 23.
Paradis, E., et al. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 20(2): 289-290.
Prentice, I.C., et al. 2000. Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography*. 27(3): 507-519.
R Core Team. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rausch, R. 1994. Transberingian dispersal of cestodes in mammals. *International Journal of Parasitology*. 24: 1203-1212.
Ree, R.H., and S.A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*. 57: 4-14.
Richmond, G.M. 1986. Stratigraphy and correlation of glacial deposits of the Rocky Mountains, the Colorado Plateau and the ranges of the Great Basin. *Quaternary Science Reviews*. 5: 99-127.
Siegenthaler, U., et al. 2005. Stable Carbon Cycle-Climate Relationship During the Late Pleistocene. *Science*. 310(5752): 1313-1317.
Simpson, G.L. 1953. Evolution and Geography: An Essay on Historical Biogeography with Special Reference to Mammals. Condon Lectures, Oregon State System of Higher Education. Eugene, Oregon, USA.
Soltis, D.E., et al. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*. 15: 4261–4293.
Tkach, N.V., et al., 2008. Parallel Evolutionary Patterns in Multiple Lineages of Arctic *Artemisia* L. (Asteraceae). *Evolution*. 62(1): 184-198.
Waltari, E., et al. 2007. Eastward Ho: phylogeographical perspectives on colonization of hosts and parasites across the Beringian nexus. *Journal of Biogeography*. 34: 561-574.
Weber, W.A. 1965. Plant geography in the southern Rocky Mountains. *The Quaternary of the United States*. 453–468.
Weber, W.A. 2003. The middle Asian element in the southern Rocky Mountain flora of the western United States: a critical biogeographical review. *Journal of Biogeography*. 30(5): 649–685.
Weir, J.T., and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society B: Biological Sciences* 271: 1881–1887.

CHAPTER 4

Microhabitat differences impact phylogeographic concordance of co-distributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains

4.1 Abstract

By selecting co-distributed, closely related montane sedges from the Rocky Mountains that are similar in virtually all respects but one – their microhabitat affinities – we test predictions about how patterns of genetic variation are expected to differ between *Carex nova*, an inhabitant of wetlands, and *Carex chalciolepis*, an inhabitant of drier meadows, slopes, and ridges. Although contemporary populations of the taxa are similarly isolated, the distribution of glacial moraines suggests that their past population connectedness would have differed. Sampling of co-distributed population pairs from different mountain ranges combined with the resolution provided by over 24,000 SNP loci supports microhabitat-mediated differences in the sedges' patterns of genetic variation that are consistent with their predicted differences in the degree of isolation of ancestral source populations. Our results highlight how microhabitat preferences may interact with glaciations to produce fundamental differences in the past distributions of presently co-distributed species. We discuss the implications of these findings for generalizing the impacts of climate-induced distributional shifts for communities, as well as for the prospects of gaining insights about species-specific deterministic processes, not just deterministic community level responses, from comparative phylogeographic study.

4.2 Introduction

Concordance in phylogeographic structure across species provides a comparative context critical for testing hypotheses regarding the temporal influence of environmental factors on communities (Avise 1992; Soltis et al. 2006; Shafer et al. 2010; Ribas et al. 2012). If membership in a specific biological community is the primary factor that determines population connectedness through time, it follows that species with similar life histories and morphological traits should show concordant phylogeographic patterns. Although the impact of historical and regional processes is often emphasized in comparative phylogeographic studies by identifying and quantifying the extent of concordance in patterns of genetic variation among species (Hickerson et al. 2010), there could also be deterministic processes that would produce discord.

Deterministic processes associated with the interaction between species' microhabitat affinities and environmental perturbations may be key contributors to phylogeographic discord. Microhabitat preferences are widely recognized as important to a variety of ecological and evolutionary phenomena. They influence the richness of biological communities and partitioning of resources therein (Hixon and Beets 1993; Ackerly 2003; Cavender-Bares et al. 2004), lead to diversification through character displacement dictated by local environmental variables (Schluter and Grant 1983; Losos et al. 1997; Rosenblum 2006), and even facilitate speciation (Feder et al. 1988; Chunco et al. 2007). It follows that microhabitat may also affect species' phylogeographic structure (Hugall et al. 2002; Whiteley et al. 2004; Hodges et al. 2007). For example, an interaction between Australian funnel web spiders' preferences for wood or ground-dwelling microhabitats and Pleistocene climatic cycles has been proposed as an explanation for phylogeographic disparities (Beavis et al. 2011). Even at large geographic scales, patterns of genetic variation may differ across taxa in a manner consistent with species-specific habitat affinities, as suggested by correlations between species' preferences for particular forest canopy strata and genetic differentiation among co-distributed South American birds (Burney and Brumfield 2009). Such studies highlight how the deterministic effects of species-specific traits on patterns of genetic variation may be dissected from a comparative phylogeographic perspective.

Here we test for phylogeographic concordance in two montane sedge species, *Carex nova* and *C. chalciolepis* (*Carex* section *Racemosae,* Cyperaceae; Fig. 4.1). There are many reasons to expect concordance between these taxa. They are endemic to high-elevation montane habitat in

western North America and are common and co-distributed throughout the southern Rocky Mountains (from the Sangre de Cristo Mountains, NM, to the Medicine Bow Mountains, WY). They are not distantly related, instead belonging to a clade of exclusively montane taxa that share a common ancestor during the Pleistocene (Massatti, In preparation). With many shared life history characteristics, the inherent dispersal capabilities do not differ between the species (e.g., they are long-lived perennials with locally dispersed seeds). Both *C. nova* and *C. chalciolepis* flower as soon as environmental conditions are appropriate for plant growth, reflecting the seasonal constraints on the time to seed maturity imposed by high-elevation environments. In these respects, both taxa are representative of many montane plants not only in terms of shared life history characteristics, but also as members of montane plant communities that respond to shifts in climatic conditions. Yet, there is one primary difference between these species – their microhabitat affinities. Within alpine tundra and unforested subalpine habitats, *C. nova* is restricted to wetlands and mesic meadows, whereas *C. chalciolepis* occurs on drier montane slopes, meadows, and ridges (Fig. 4.1). The question is: are microhabitat differences of significant evolutionary consequence? That is, instead of phylogeographic concordance, do species-specific differences have the potential to affect phylogeographic structure?

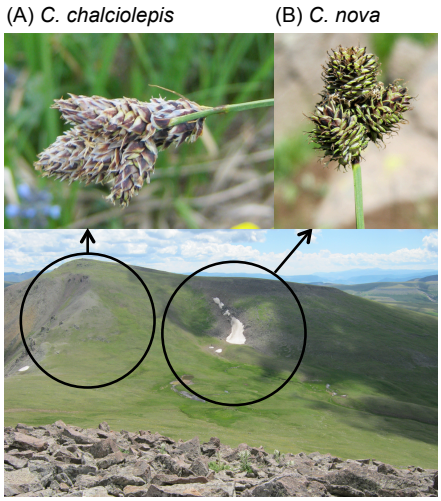

Figure 4.1 *Carex chalciolepis* (A) and *C. nova* (B) inflorescence morphologies and representative habitats.

Montane plants are adapted to the strenuous physiological demands of high-elevation habitats. In addition, as inhabitants of areas directly impacted by glacial cycles, they experienced repeated displacements from their current distributions (Pierce 2003). This includes tracking suitable habitat to lower elevations during glacial periods (in addition to potentially surviving *in situ*, see below), followed by re-colonization from ancestral populations to distributions established during interglacials (as established by palynological records and macrofossils; e.g., Betancourt et al. 1990; Thompson et al. 1993; Thompson and Anderson 2000). Despite the similar adaptations that all montane plants share, their microhabitat affinities may have

facilitated disparate responses to glaciations. Plants inhabiting montane microhabitats that dry out relatively quickly (e.g., ridges and south-facing slopes) are among the earliest flowering plants in montane ecosystems because wind redistributes snow to places like drainages and leeward slopes (Holway and Ward 1965). Although the cooler climates of glacial periods likely shortened the growing season throughout montane habitats (Pierce 2003), plants on slopes and ridges, like *C. chalciolepis*, may have persisted *in situ* because they are adapted to initiate growth as soon as environmental conditions permit (e.g., Stehlik et al. 2002).

In contrast to ridge and slope microhabitats, drainages and landscape depressions preferentially accumulate snow and ice. Consequently, the majority of plants in wetter microhabitats, such as *C. nova*, flower later compared to plants in other montane microhabitats (Holway and Ward 1965) when meltwater subsides and the substrate warms (Körner 2003). Glacial periods impacted the majority of wetland microhabitats in the southern Rocky Mountains by facilitating the growth of valley glaciers down drainages, as evidenced by the distribution of moraines and glacial till (Ehlers and Gibbard 2004). Furthermore, wetland microhabitats adjacent to glaciers would have had a shortened growing season due to an increased volume of meltwater resulting from the increased accumulation of snow and ice during glacial periods. Therefore, it is unlikely that wetland plants persisted locally, instead being disproportionately displaced to lower elevations in drainages.

Based on the general assumption that co-distributed montane species with similar inherent dispersal capabilities will respond similarly to historical events, we test the null hypothesis of concordant phylogeographic structure between *C. nova* and *C. chalciolepis*. Alternatively, if microhabitat differences are of significant consequence, we predict the species will show discordant patterns of genetic variation. Moreover, we predict that *C. nova* populations will be more differentiated from one another compared to *C. chalciolepis* populations, in line with expectations based on the interactions between species' microhabitat affinities and climatic oscillations. Specifically, populations of *C. nova* from unconnected drainages should have experienced prolonged disjunctions during the glaciations, especially given the relative brevity of interglacial compared to glacial periods (Clark et al. 1999), when they were displaced to lower elevations in drainages around the margins of mountain ranges. Prolonged disjunction*s* are not expected to be characteristic of *C. chalciolepis* populations, where population differentiation is expected to reflect colonization from within-glacier refugial ridge and slope microhabitats, as

well as from lower elevation populations (which were not restricted to drainages). With fewer expected differences between the current and past distributions of *C. chalciolepis*, the genetic signature of distinct ancestral refugia is not predicted to be as strong as in *C. chalciolepis* as in *C. nova.* Note that unlike other comparative phylogeographic studies where inferences about causality rely on repeated patterns observed across taxa (e.g., Hugall et al. 2002; Burney and Brumfield 2009), our sampling design is structured to assess repeated genetic patterns at the population level (see also Grahame et al. 2006; Rosenblum 2006; Ryan et al. 2007). Specifically, by sampling two populations per mountain range across multiple mountain ranges in each taxon, we can evaluate whether the patterns of genetic variation differ consistently between the species.

Phylogeographic concordance (or lack thereof) was assessed in *C. nova* and *C. chalciolepis* using a genomic dataset of more than 24,000 anonymous SNP loci sequenced in seven pairs of sympatric populations across five mountain ranges. We used environmental niche modeling (ENM) to confirm similar projected distributions of the species during the Last Glacial Maximum (LGM). We then investigated the impact of glaciations on genomic differentiation using STRUCTURE and principal component analysis (PCA). Our results show that the sedges have similar magnitudes and ranges of genetic differentiation; however, the gene pool of *C. nova* displays hierarchical structure while that of *C. chalciolepis* does not, as predicted under our hypothesis that microhabitat does influence a species' response to climatic cycling. These results highlight how difficult it might be to predict the consequences of future climate change on communities as a whole. We discuss the implications of our results for interpreting previously documented, but difficult to interpret, patterns of community assemblages from regions impacted by glacial cycles (e.g., Soltis et al. 2006), as well as generalizations that have been made from comparative phylogeographic study.

4.3 Methods

Sample collection and DNA extraction

To minimize the collection of related individuals, tissue samples (i.e., 2 cm of leaf material) were collected from well-dispersed *C. chalciolepis* and *C. nova* individuals (average distance between samples of 300 m, and a minimum distance of 35 m) at each sampling locality during the summer of 2011 (Fig. 4.2). A total of 40 individuals were sampled per species; five individuals were sampled per population except for Blanca, where ten individuals were sampled (see Table

4.1 for details). Leaf material was stored in silica gel until DNA was extracted with DNeasy Plant Mini Kits (QIAGEN) following the manufacturer's protocol.

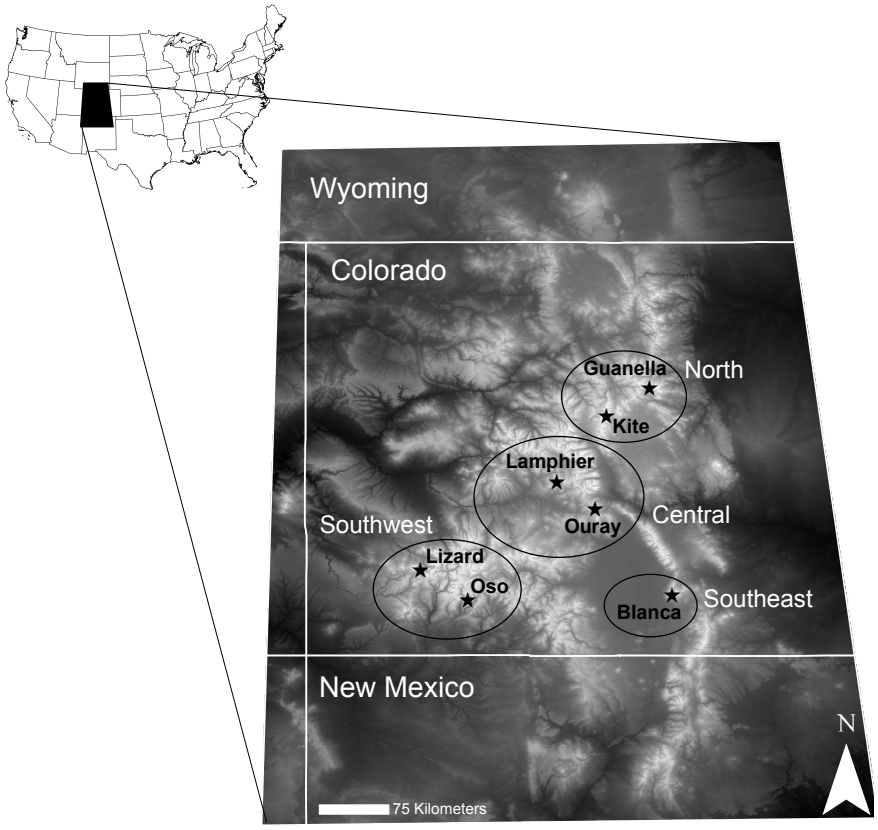

Figure 4.2 Map of the southern Rocky Mountains showing the distribution of sampled populations (marked by stars) of *C. chalciolepis* and *C. nova* in each of the four geographic regions referred to in the text; whiter colors indicate higher elevations (elevations range from 1200 to 4350 m). Populations are located in the San Juan Mountains (Oso and Lizard), the Sangre de Cristo Mountains (Blanca), the Sawatch Mountains (Ouray and Lamphier), the Mosquito Mountains (Kite), and the Front Range (Guanella).

Environmental niche modeling

Environmental niche models (ENMs) were used to confirm the similarity of the sedges' responses to glaciations, given that there are areas within their respective ranges where they do not co-occur. ENMs were generated from bioclimatic variables for the present and the LGM for each species with MAXENT v3.3.3e (Phillips et al. 2006) using the following parameters: regularization multiplier = 1, max number of background points = 10,000, replicates = 50, replicated run type = cross-validate. Georeferenced distribution points used in the modeling were representative of species' entire ranges throughout western North America and were collected

from personal fieldwork and validated voucher specimens housed at the Rocky Mountain Herbarium (species distribution points are available at http://doi.org/10.5061/dryad.4158c). We utilized 19 bioclimatically informative variables to model present-day distributions (WorldClim v1.4; Hijmans et al. 2005) and LGM distributions (PMIP2-CCSM; Braconnot et al. 2007). Full details on ENM modeling procedures are available in the Supplementary Material.

Table 4.1 Details for the sampled populations in both *C. chalciolepis* and *C. nova*, including population name, geographic region within the Southern Rocky Mountains (see Fig. 4.2), the number of individuals collected per species, collection site coordinates (latitude, longitude), and the range of elevations (m) over which plants were sampled.

Population	Geographic region	Number of individuals per species	GIS coordinates	Elevation (m)
Oso	Southwest (SW)	5	37.5357, -107.4817	3500 - 3900
Lizard	Southwest (SW)	5	37.8256, -107.9391	3200 - 3700
Blanca	Southeast (SE)	10	37.5991, -105.4782	3250 - 3800
Ouray	Central (C)	5	38.4329, -106.2407	3400 - 3800
Lamphier	Central (C)	5	38.6772, -106.6069	3600 - 3900
Kite	North (N)	5	39.3279, -106.1360	3700 - 4000
Guanella	North (N)	5	39.5941, -105.7117	3600 - 3900

RAD library preparation and sequence analysis

Extracted genomic DNA was individually barcoded and processed into a reduced complexity library using a restriction fragment-based procedure (for details see Gompert et al. 2010). Briefly, DNA was doubly digested with *Eco*RI and *Mse*I restriction enzymes, followed by the ligation of Illumina adaptor sequences and unique barcodes. Ligation products were pooled among samples and the fragments were amplified by PCR. Gel purification was used to size select fragments between 300 and 400 base pairs. The library was sequenced in one lane on the Illumina HiSeq2000 platform according to manufacturer's instructions to generate 50 base pair, single-end reads. Sequences were demultiplexed using custom scripts and only reads with Phred scores ≥ 32 and that had an unambiguous barcode and restriction cut site were retained (scripts are available at http://doi.org/10.5061/dryad.4158c). Potential chloroplast and mitochondrial sequences were filtered from the processed dataset using Bowtie 0.12.8 (see Supplementary Information for more details; Langmead et al. 2009).

SNPs were determined from loci formed from overlapping sets of homologous fragments and genotypes were assigned using a maximum-likelihood statistical model (Catchen et al. 2011; Hohenlohe et al. 2012) with the Stacks v1.03 pipeline (Catchen et al. 2013); default settings were used except where noted below. Specifically, loci and polymorphic nucleotide sites were identified in each individual using the USTACKS program, which groups reads with a minimum coverage depth (m) into a "stack". The data were processed with m = 3; increasing the minimum depth helps to avoid erroneously calling convergent sequencing errors as stacks. A catalog of consensus loci among individuals was constructed with the CSTACKS program from the USTACKS output files for each species, where loci were merged together across individuals if the distance between them (n) was ≤ 2. This catalog was used to determine the allele(s) present in each individual at each homologous locus using the SSTACKS program. Our choice of parameters was determined with consideration of avoiding both over- and under-merging of homologous loci in the focal taxa, as well as with reference to other studies (e.g., Renaut et al. 2011). Similarity of the number of loci identified in the species for different parameter values used in USTACKS (m) and CSTACKS (n) suggests that the properties of the genomic libraries were similar (i.e., that the potential errors associated with over- or under-merged homologous loci did not differ substantially between the species). The close relatedness of the taxa and the short reads makes an n ≤ 2 reasonable (also see Renaut et al. 2011), although we acknowledge this could be conservative if catalogs were assembled for taxa that were more distantly related, and/or for read lengths larger than 50 base pairs.

Population genetics statistics, including major allele frequency, nucleotide diversity (π), and Wright's *F*-statistics F_{IS} and F_{ST} were calculated for each SNP using the POPULATIONS program in the Stacks pipeline (Catchen et al. 2013). For bi-allelic SNP markers, π is a measure of expected heterozygosity, and is therefore a useful measure of genetic diversity in populations; furthermore, F_{IS} measures the reduction in observed heterozygosity as compared to expected heterozygosity for an allele in a population, and positive values indicate nonrandom mating or cryptic population structure (Nei and Kumar 2000; Hartl and Clark 2006; Holsinger and Weir 2009). One SNP with two alleles was identified in every available homologous locus for each sedge species; this resulted in 24,574 and 25,670 SNPs available for analyses in *C. nova* and *C. chalciolepis*, respectively. While we could have included SNPs with more than 2 segregating alleles or additional SNPs segregating at a locus, we opted for the conservative approach of

excluding these SNPs since they may disproportionately represent sequencing errors, given the close relatedness of the taxa and short read lengths of 50 base pairs (see also Renaut et al. 2011). Note that SNPs with more than 2 segregating alleles were uncommon, representing less than 1% of the loci.

Only loci present in at least two populations (p = 2) and genotyped in at least 50% or 75% of the individuals of each population (r = 0.50 or r = 0.75) were used to create molecular summary statistics for each species; in instances where 50% or 75% did not result in a round number of individuals in a population, the number of individuals required before the locus was processed was rounded up (e.g., a locus would need to be genotyped in 3 out of 5 individuals with r = 0.50, and 4 out of 5 individuals with r = 0.75). In addition, per locus F_{ST}-values were calculated only for loci with minimum minor allele frequencies of 5% or higher (a = 0.05), the lowest possible minor allele frequency given our sampling effort within populations. Isolation by distance was assessed within each species by graphing populations' pairwise F_{ST} against Euclidean distance (km) in the Isolation By Distance Web Service (Jensen et al. 2005); one-sided *p*-values were computed by randomizing the data 30,000 times and both geographic distance and genetic distance were log-transformed to determine their effect on results.

Characterization of population genetic structure

Population genetic structure was characterized within *C. chalciolepis* and *C. nova* using STRUCTURE 2.3.4 (Pritchard et al. 2000). SNP data were exported from the POPULATIONS program in Stacks into a STRUCTURE file using the parameter settings r = 0, p = 2, and a = 0 (see above for parameter descriptions). We created different data sets using more inclusive parameters in order to take advantage of variation in individuals within populations that was filtered out when calculating population-level summary statistics; this variation may be important for assessing genetic similarity between individuals from different populations when population membership isn't assumed *a priori*. *K*-values ranging from 1 to 9 (two more than the total number of populations in each species) were analyzed in STRUCTURE. Ten independent runs per *K* were conducted, each with 100,000 burn-in and 150,000 MCMC iterations, using the 'Admixture Model' and the 'Correlated Allele Frequency Model' with default settings. Results were not different using more burn-in or MCMC iterations. STRUCTURE HARVESTER (Earl and

vonHoldt 2012) and DISTRUCT (Rosenberg 2004) were used to visualize results, and the most probable *K* was chosen based on Δ*K* (Evanno et al. 2005).

To explore genetic structure that might be present within initial clusters identified by STRUCTURE (i.e., hierarchical structuring of genetic variation), an iterative approach was used (e.g., Ryan et al. 2007). This involved the analysis of subsets of the data that corresponded to the respective genetic clusters identified in previous runs. For the analyses of data subsets, values of *K* ranging from 1 to 5 were explored using the same parameter settings as those used in the initial STRUCTURE analyses. As with the analyses of the entire dataset across all populations, only loci with SNPs were included in the analyses.

In addition to the analyses described above, all the analyses were repeated with a dataset in which three putative hybrid individuals were removed to confirm the robustness of the results. These individuals were identified from a STRUCTURE and principal components analysis (PCA) of the pooled datasets of *C. chalciolepis* and *C. nova*, along with a third closely related sedge species (*C. nelsonii*), which was not included in this study because of limited sampling. The genomic composition of these three individuals contained significant heterospecific contributions; one *C. nova* individual was clearly admixed with *C. chalciolepis* and two *C. chalciolepis* individuals were hybridized with *C. nelsonii* (Fig. B.1).

To visualize the major axes of population genetic variation, a PCA was performed using the 'adegenet' package (Jombart 2008) in R (R Core Team 2012). PCA is a natural companion to STRUCTURE because it is free from many of the population genetics assumptions underlying STRUCTURE (Gao et al. 2007; Jombart et al. 2009), and it can be more useful with continuous patterns of differentiation (e.g., isolation by distance; Engelhardt and Stephens 2010). The datasets requiring either 50% or 75% completeness within populations' loci (see above) were utilized for PCA; in addition, analyses were repeated excluding the putative hybrids, similar to the method described for STRUCTURE analyses (above).

4.4 Results

Ecological niche modeling

Present-day and LGM ENMs were generated using 4 bioclimatic variables: temperature seasonality (Bio4), maximum temperature of the warmest quarter (Bio5), precipitation of the driest month (Bio14), and precipitation of the warmest quarter (Bio18). Not only are the present-

day ENMs of *C. chalciolepis* and *C. nova* nearly identical within the southern Rocky Mountains (Fig. B.2), but their distributions are also predicted to be largely congruent during the LGM (Fig. 4.3). The similarity of the ENMs between the species confirms the predicted phylogeographic concordance of the two sedge taxa under the assumption that montane taxa respond similarly to past climatic cycles. The maximum training sensitivity plus specificity threshold (MTSS) was among the most over-predictive of the thresholds calculated by MAXENT, deeming habitat with ≥ 18% and ≥ 11% probability of being suitable as suitable habitat for *C. nova* and *C. chalciolepis*, respectively. Models created using more over-fitted thresholds only decreased the area surrounding predicted regions using MTSS and never identified regions as no longer habitable, suggesting the approach used was appropriate given our intentions to assess the overall similarity of *C. chalciolepis'* and *C. nova's* LGM distributions.

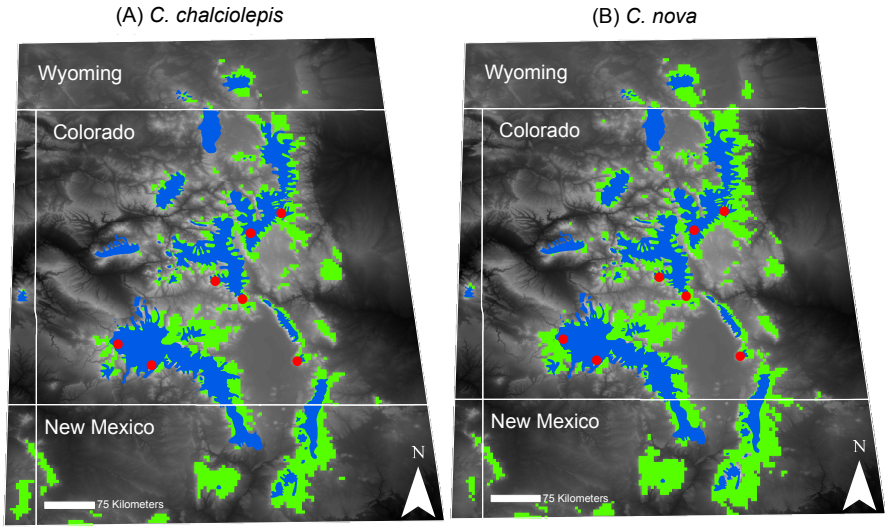

Figure 4.3 The predicted LGM distributions of *C. chalciolepis* (A) and *C. nova* (B) are shown in green using the maximum training sensitivity plus specificity threshold. Glaciers (i.e., non-suitable habitat) are shown in blue (data extracted from Glaciers of the American West, http://glaciers.us), and red dots depict the sampling locations.

Sequence data quality and processing

One lane of Illumina sequencing produced more than 132 million reads derived from 80 individuals (average of 1,659,473 ± 481,727), of which more than 127 million reads had Phred scores ≥32 and an unambiguous barcode (Table 4.2 and Fig. B.3). On average, an additional 6% of reads per individual were discarded because they aligned with either chloroplast or mitochondrial genomes. Based on low coverage or low quality of reads, 5 individuals were

excluded from further analyses, including four from *C. nova* and one from *C. chalciolepis* (Table 4.2 and Fig. B.3). Filtering loci containing one SNP and two alleles from the species' MySQL databases resulted in 24,574 loci for *C. nova* and 25,670 loci for *C. chalciolepis* (data are available at http://doi.org/10.5061/dryad.4158c), with all populations well represented (Fig. B.4).

Table 4.2 Summary of genomic data collected in each population, presented as averages across individuals for a given population in each species. Shown are the raw count of reads from the Illumina run and the number of reads after processing for quality control (i.e., after excluding reads with low quality scores, ambiguous barcodes, and that aligned with a haploid genome), as well as the number of reads analyzed with Stacks to identify homologous loci.

Populations	*n**	Raw read count	Post quality control	Analyzed reads	% Overall of raw reads used
C. chalciolepis					
Oso	4	1,631,473	1,485,349	1,411,438	86.51
Lizard	5	1,378,608	1,281,107	1,225,386	88.89
Blanca	10	1,492,198	1,416,754	1,359,250	91.09
Ouray	5	1,898,460	1,747,426	1,691,626	89.11
Lamphier	5	1,904,632	1,791,184	1,735,826	91.14
Kite	5	1,606,399	1,504,082	1,444,351	89.91
Guanella	5	1,728,853	1,628,912	1,569,360	90.77
C. nova					
Oso	5	2,301,651	2,126,856	2,054,335	89.25
Lizard	5	2,355,473	2,075,661	2,021,638	85.83
Blanca	9	1,602,688	1,484,440	1,411,294	88.06
Ouray	4	1,779,119	1,656,305	1,580,144	88.82
Lamphier	5	1,466,146	1,359,203	1,282,636	87.48
Kite	5	1,922,637	1,774,684	1,707,109	88.79
Guanella	3	1,447,511	1,038,007	971,188	67.09

Genetic diversity within and among southern Rocky Mountain Carex *populations*

Characterization of genetic diversity within and among populations was conducted on the two datasets that differed in the tolerance levels for missing data, requiring either 50% or 75% of individuals within a population to have a locus to be included in the dataset. The results were qualitatively similar and therefore we only present data for the dataset restricted to loci present in 50% or more of the individuals within a population. *Carex nova* and *C. chalciolepis* are quite comparable with respect to common measures of genetic diversity (e.g., average major allele frequencies and observed heterozygosities; see Tables 4.3 and B.1 for details). Likewise, their populations are characterized by similar ranges of genetic diversity, π (0.390 to 0.581 and 0.359

to 0.504 in *C. nova* and *C. chalciolepis,* respectively), with the lowest diversity in the most isolated population (Blanca) among the sites collected. Observed heterozygosities are lower than π in each population for both species, resulting in positive F_{IS}-values within all populations (Table 4.3).

Genetic relatedness between population pairs within *C. chalciolepis* and *C. nova* was measured using F_{ST}. The range of F_{ST}-values was similar between the two species (0.053 to 0.088 in *C. chalciolepis,* and 0.041 to 0.071 in *C. nova*). However, a significant pattern of isolation by distance was only apparent in *C. chalciolepis* (*p*-value = 0.0419; Table B.2). Log transformations of the geographic distance or genetic distance had no impact on the significance of the results.

Table 4.3 Summary statistics for the sampled populations of *C. chalciolepis* and *C. nova*. Results are presented for only polymorphic nucleotide positions. Shown are the number of loci, the average frequency of the major allele (*P*), the average observed heterozygosity per locus (H_{obs}), the average nucleotide diversity (π), and the Wright's inbreeding coefficient (F_{IS}). See Table B.1 for summary statistics including polymorphic + fixed nucleotide positions.

	Population	Loci	*P*	H_{obs}	π	F_{IS}
C. chalciolepis	Oso	8129	0.662	0.262	0.504	0.4470
	Lizard Head	4471	0.717	0.238	0.434	0.4251
	Blanca	5493	0.744	0.271	0.359	0.2645
	Ouray	3547	0.674	0.305	0.474	0.3773
	Lamphier	3560	0.679	0.327	0.455	0.3198
	Kite	7298	0.728	0.353	0.428	0.1484
	Guanella	4694	0.699	0.262	0.440	0.4062
C. nova	Oso	5041	0.694	0.225	0.448	0.5269
	Lizard Head	2753	0.667	0.289	0.503	0.4227
	Blanca	5265	0.719	0.204	0.390	0.4910
	Ouray	7160	0.743	0.336	0.411	0.1493
	Lamphier	3949	0.689	0.274	0.454	0.4125
	Kite	4895	0.688	0.232	0.456	0.5060
	Guanella	1947	0.614	0.431	0.581	0.2517

Population structure of southern Rocky Mountain Carex

Analyses with STRUCTURE across different values of *K* identified *K* = 2 as the most probable number of genetic groups (the *K* with the highest Δ*K*, hereafter referred to as the most probable *K*) in both *C. chalciolepis* and *C. nova* (Table 4.4). Groups identified at *K*=2 showed a strong correspondence with geography in *C. nova* (separating the North and Central populations from

the Southeast and Southwest populations), but were not interpretable geographically in *C. chalciolepis* (Fig. 4.4). Moreover, in contrast to the large difference in Δ*K* between the first and second most probable *K*-values in *C. nova* (*K* = 2 and *K* = 3), the small difference in Δ*K* between the most probable *K*-values in *C. chalciolepis* (*K* = 2 and *K* = 4), indicates the lack of equivalent distinctiveness of genetic clusters in *C. chalciolepis* compared to *C. nova* (Table 4.4). Considering the second most probable *K*-value in *C. chalciolepis* (*K* = 4), there is more of a geographical correspondence than at *K* = 2 (Fig. 4.4), suggesting that genetic variation in *C. chalciolepis* is structured. However, this structuring is less pronounced than the hierarchical nature of *C. nova's* genetic variation (see below).

Table 4.4 Summary of results for STRUCTURE analyses for each species. Each row represents a separate analysis with the first and second most probable *K*-value and their associated Δ*K* shown. The number of individuals, the geographic regions represented by those individuals, and the number of SNPs used in the respective analyses is also given. Note the putative hybrid individuals were not included in these analyses (see Methods and Fig. B.1). Inclusion of the two putative hybrid individuals did not affect the most probable *K*-value in *C. chalciolepis* (i.e., the analyses were robust). However, inclusion of the putative hybrid in *C. nova* from the Ouray population (Central region) affected the most probable *K*-value inferred by STRUCTURE in the two analyses of regional subsets of the data (i.e., those containing individuals from the Central region). Specifically, the hybrid individual was distinguished from the other individuals from the Ouray population, thereby making the most probable *K*-value = 3.

Species	# loci	# individuals	Geographic Region(s)	1st *K*	Δ*K*	2nd *K*	Δ*K*
C. chalciolepis	24,566	37	SW, SE, C, N	2	14.6	4	6.4
C. nova	23,405	35	SW, SE, C, N	2	675.1	3	2.4
	22,950	19	SW, SE	2	503.9	3	71.6
	22,223	10	SW	2	68.3	3	2.5
	22,565	16	C, N	2	67.3	4	2.7
	21,579	8	C	2	10.9	4	0.6
	21,365	8	N	2	8.6	4	2.2

In contrast with *C. chalciolepis*, not only was there structuring of genetic variation at the regional scale within *C. nova*, but the series of STRUCTURE analyses on regional subsets of the data identified significant structuring of genetic variation within mountain ranges (Fig. 4.4). Moreover, *C. nova* populations were repeatedly delimited across multiple mountain ranges. The pronounced difference in Δ*K* separating the most probable *K* (*K* =2) and the second most

probable *K* (either *K* = 3 or 4; Table 4.4) in most analyses confirms that the major axes of genetic variation within subsequent subsets of data are strictly concordant with geography in *C. nova*.

The genetic composition of *C. chalciolepis* individuals generally differed from that of *C. nova* individuals. For example, the probability of mixed ancestry was very low in *C. nova* individuals at both the regional scale of mountain ranges and the local scale of populations within mountain ranges (Fig. 4.4B). The only exception was the mixed ancestry of some individuals from Ouray within the Central region (i.e., the Sawatch Mountains; Fig. 4.2). In contrast, *C. chalciolepis* not only displays a consistent lack of clear genetic differences between individuals from different populations within each of the mountain ranges, but it also shows evidence of mixed ancestry that traces to populations from different mountain ranges (i.e., regions; Fig. 4.4A).

Results from the PCA analyses parallel results from STRUCTURE. Individuals from *C. chalciolepis* exhibited some geographic structuring of genetic variation (Fig. 4.5A). However, with the exception of individuals from Blanca, the populations were not separated along the PCA axes to the same degree as *C. nova* populations (Fig. 4.5B). *Carex nova* populations were separated on both PC1 and PC2 (Fig. 4.5B) into clusters at the regional level (i.e., between mountain ranges from the Southwest, Southeast, and Central + North regions), as well as at the population level for most, but not all, populations. Specifically, populations from the Central region (i.e., the Ouray and Lamphier populations) are overlapping with those from the North region (i.e., the Kite and Guanella populations). The two individuals from Oso that are separated from the other *C. chalciolepis* individuals along PC1 (i.e., the 2 left-most red points in Fig. 4.5A) correspond to those that are suspected to be putative hybrids based on species-level analyses with the third closely related species, *C. nelsonii* (as discussed in the methods, *C. nelsonii* was not included in this study because of limited geographic sampling). Analyses without these putative admixed individuals, or allowing different amounts of missing data, did not change the overall patterns of genetic variation among individuals or populations of *C. chalciolepis* or *C. nova* (results not shown).

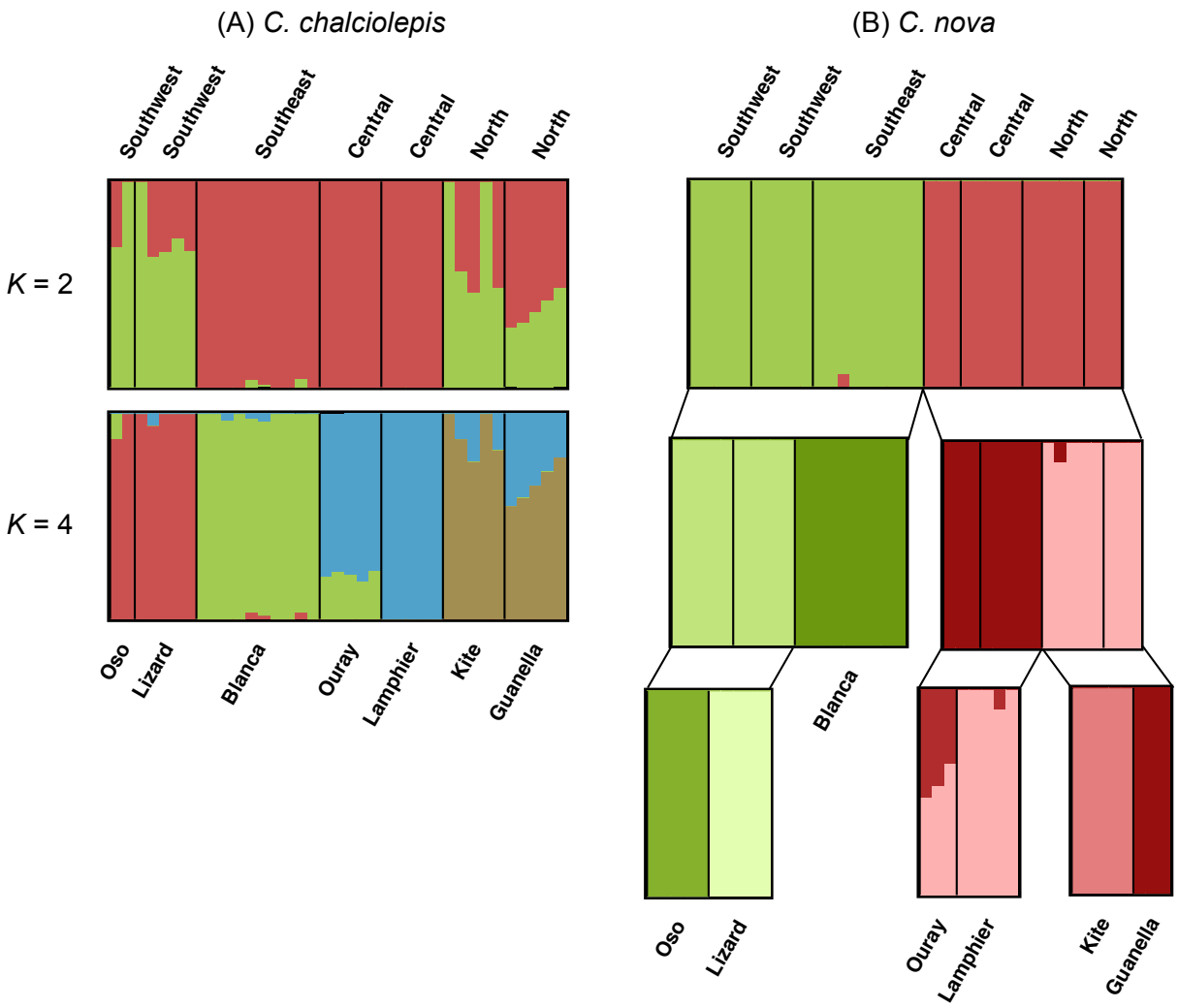

Figure 4.4 Plots of posterior probabilities for individuals assigned to *K* groups from STRUCTURE analyses (each separate block corresponds to one analysis). Each of the *K* groups within an analysis is shown as a different color for *C. chalciolepis* (A) and *C. nova* (B), and thin black lines delimit populations, whose names are listed along the bottom. The regional membership of populations (see Fig. 4.2) is listed along the top-most analysis. The two most probable values of *K* are shown for analyses of all the individuals of *C. chalciolepis*, whereas the results from a series of hierarchical analyses of subsets of data with $K = 2$ are shown for *C. nova* (slanted lines show the subsets of data analyzed, starting from the entire dataset depicted at the top, down to individuals within regions, shown at the bottom, excluding the Southeast region because of the lack of multiple sampled populations). Note that no hierarchical geographic structuring was detected in *C. chalciolepis* (see text and Table 4.4 for details).

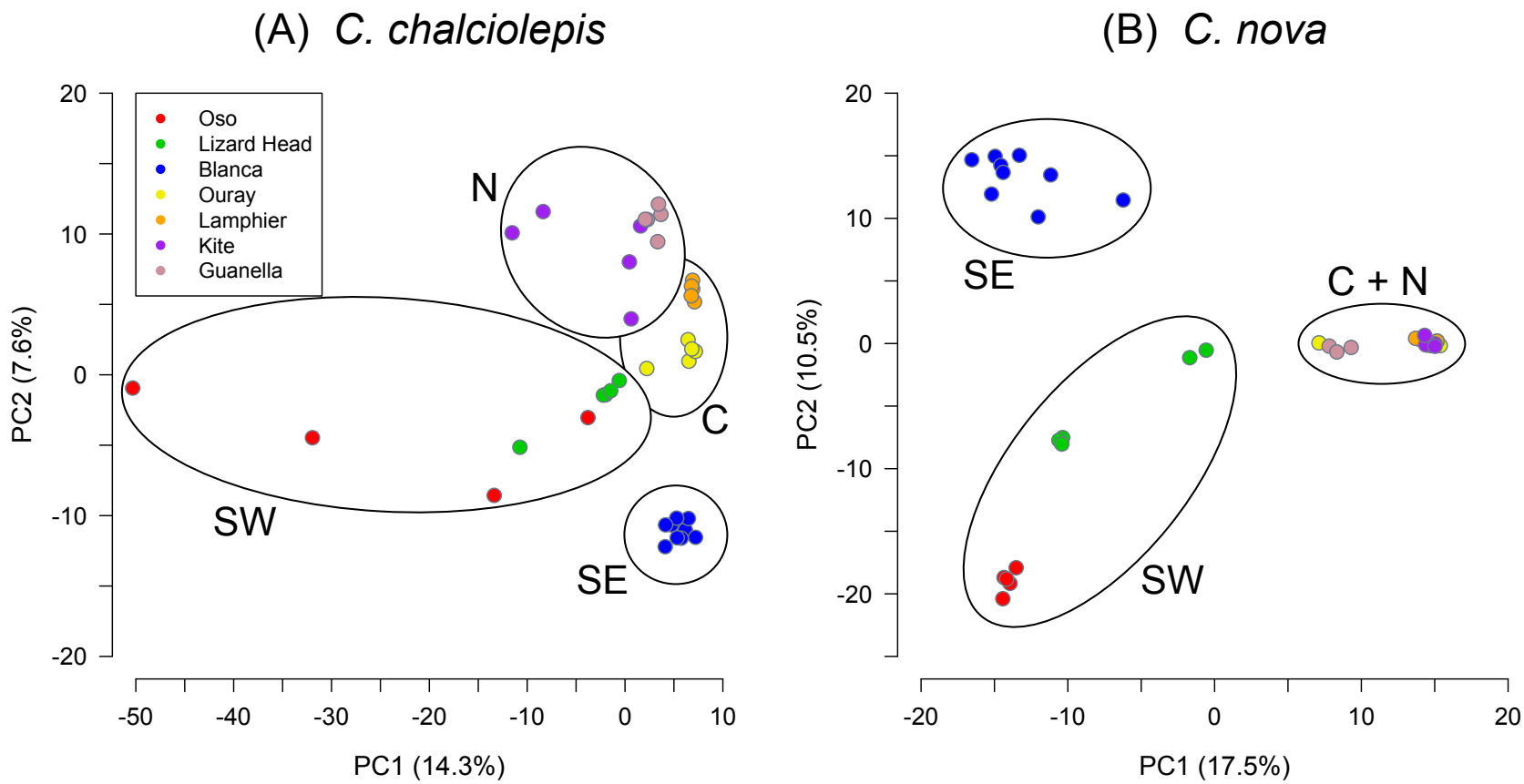

Figure 4.5 Distribution of individuals along PC1 and PC2 axes of genetic variation for *C. chalciolepis* (A) and *C. nova* (B), with the amount of variation explained by each axis given in parentheses. The pattern among *C. nova's* Central and North individuals is shown in detail in Fig. B.5. Colors indicate population identity in each species, and ellipses demarcate regions (SW = Southwest, SE = Southeast, C = Central, and N = North; see Fig. 4.2).

4.5 Discussion

This study highlights how comparative phylogeography may be used to provide insights about the interaction between species' traits (e.g., morphological traits or life history characteristics) and historical environmental perturbations. Our approach – that is, selecting closely related species with many similarities, thereby controlling for the phylogeographic disparity that may be due to suites of trait differences, and sampling co-distributed populations from multiple mountain ranges – provides an opportunity to explore whether patterns of genetic variation differ between taxa in a predictable fashion. Below we discuss the relative strengths and weaknesses of this approach, and argue that the manner in which phylogeographic study is pursued can contribute to a bias in both our perceptions about the factors structuring patterns of genetic variation and the tendency to interpret discordant patterns of genetic variation as reflecting the vagaries of history rather than deterministic processes.

Species-specific traits structuring patterns of genetic variation

Phylogeographic concordance of co-distributed species can provide important details about the impact of environmental changes on communities over time. For example, investigation of disparate taxa distributed throughout Europe has elucidated refugia, migration routes, and suture zones resulting from Pleistocene glacial cycles (Hewitt 2004). Similar scenarios have been reconstructed for the Pacific Northwest in North America (reviewed in Shafer et al. 2010), and to a limited extent for the Rocky Mountains (Brunsfeld and Sullivan 2005; DeChaine and Martin 2005a; Spellman et al. 2007). In addition, concordance among species may help elucidate the factors that shaped past refugia, as for example in the Brazilian Atlantic Coastal Forest (Carnaval et al. 2009).

As may be expected based on the many studies of montane taxa (Galbreath et al. 2010; Knowles and Alvarado-Serrano 2010), including tests of phylogeographic concordance within montane regions of North America (Soltis et al. 1997; Carstens et al. 2005), there are similarities in the geographic patterns of genetic variation of *C. chalciolepis* and *C. nova*. For example, within both species, populations are regionally differentiated, and they exhibit similar ranges of genetic diversity (Fig. 4.5 and Table 4.3). In additional, relatedness is generally correlated with geographic proximity (Fig. 4.5), with the geographically isolated Blanca population from the Southeast region (Fig. 4.2) well differentiated from other populations in both species. These

similarities are evidence of common processes that structure patterns of genetic variation. However, they are not unexpected given that the sedges have similar dispersal capabilities and are co-distributed across the southern Rocky Mountains. For example, similar results were found for co-distributed butterfly species sampled across the Rocky Mountains (DeChaine and Martin 2005b). The similarities in patterns of genetic variation not only reinforce the commonalities between the species as montane taxa with limited dispersal, but they also provide a compelling framework for interpreting the potential genetic consequences of microhabitat differences.

The proposed differences in past population connectedness based on their microhabitats are supported by differences in the extent of geographic structuring of molecular variation in *C. nova* compared to *C. chalciolepis* (Fig. 4.4). For example, in contrast to the regional correspondence of a Central and North genetic cluster and a Southeast and Southwest genetic cluster at $K = 2$ in *C. nova*, there are no regional groups at $K = 2$ in *C. chalciolepis* (in fact, the delimited clusters display no obvious geographic pattern). Only at $K = 4$ does the signature of regional structuring become apparent in *C. chalciolepis*. However, in contrast to the distinctiveness of regional groups in *C. nova* (i.e., there is little evidence of admixture between individuals from different mountain ranges), there is clear evidence of admixture among regional groups in *C. chalciolepis* (Fig. 4.4). This is also reflected in the PCA, where *C. chalciolepis* populations are continuously distributed along PC2 from south to north, while *C. nova* populations are regionally distinct (Fig. 4.5). Only among populations within the Central region (i.e., Ouray and Lamphier populations) is there a lack of genetic distinctiveness in *C. nova* (Fig. 4.4). These are the only two *C. nova* populations sampled in this study that may have been recolonized by ancestral populations that shared a common drainage during Pleistocene glaciations, consequently experiencing less isolation compared to other regional population pairs.

Paleoclimatic modeling provides additional evidence that *C. chalciolepis* potentially persisted within suitable habitat exposed in glaciated areas (i.e., ridges and slopes). Four environmental variables had comparable ranges during the LGM compared to the present for the southern Rocky Mountains: temperature seasonality (Bio4), maximum temperature of the warmest quarter (Bio5), precipitation of the driest month (Bio14), and precipitation of the warmest quarter (Bio18). Relative stability of these climatic variables suggests that a montane growing season may have persisted within and adjacent to glaciers. This supports our hypothesis that montane plants adapted to ridges and slopes may have persisted not only at lower elevations,

but also within the margins of glaciers (for a detailed perspective of the interaction of ridges, slopes, and valley glaciers, see Fig. B.6). Although survival within high elevation refugia (i.e., nunataks) has been supported in other montane systems (e.g., Schönswetter et al. 2005), the additional example represented by *C. chalciolepis* (but not *C. nova*) is informative to ongoing debates about the generality of nunataks for *in situ* survival within mountains and their contribution to geographic patterns of genetic variation (i.e., Brochmann et al. 2003). In particular, the contrast between the likelihood of survival in the mountains during glacial periods between the sedge species highlights the importance of considering microhabitat when evaluating the proposed role for nunataks.

In contrast to many phylogeographic studies that are motivated by seeking concordance among taxa to elucidate historical processes, our work highlights how considering species-specific traits may facilitate investigations focusing on phylogeographic disparities. Our work complements past studies that have shown how genetic variation can differ in a fashion consistent with differences in the species' ecological traits. For example, in darters (Turner and Trexler 1998), differences in life history strategies among taxa from headwater habitats are associated with differing levels of gene flow and subsequently disparate phylogeographic patterns (i.e., species with small clutches and large eggs are characterized by low gene flow compared to species with high fecundity and small eggs), whereas phylogeographic discord has been linked to differences in spawning locations in Bull trout and mountain whitefish from the northern Rocky Mountains (Whiteley et al. 2004). Instead of reflecting intrinsic characteristics of the taxa (e.g., differences in inherent dispersal capabilities or other life history traits; see Knowles and Alvarado-Serrano 2010), aspects of the habitats themselves might also contribute to predictable discordant patterns of genetic variation across taxa. For example, the degree of similarity in the structuring of genetic variation among co-distributed darkling beetle taxa is associated with whether the taxa inhabit ephemeral versus stable habitats (Papadopoulou et al. 2009). As with the work presented herein, such studies suggest that deterministic processes may indeed underlie some of the differences observed in patterns of genetic variation across taxa (but see below).

Although the results we present are consistent with the predicted patterns of genetic variation, there are certainly limitations to our analyses. The evaluation of the data with respect to the predictions based on the effects of microhabitat differences is correlative, and in this

respect, our study is not unlike many phylogeographic studies (see references in Shafer et al. 2010). Specifically, without an explicit model capable of generating predictions that reflect species-specific microhabitat differences, we cannot distinguish the statistical support for alternative hypotheses. At this point, no such model exists (see review in Alvarado-Serrano and Knowles 2013). There have been promising advances for testing hypotheses that capture biological phenomena that cannot be accommodated in more generic modeling approaches (see Knowles 2009). Fortunately, with the potential power provided by the genomic dataset presented here, we will be poised to take advantage of new methodologies for generating species-specific predictions as they develop (e.g., Neuenschwander et al. 2008; He et al. 2013; Martinkova et al. 2013). Moreover, by identifying the potential importance of microhabitat in structuring patterns of genetic variation, our study can serve to motivate and guide future methodological developments. Such work could also be extended to additional co-distributed *Carex* species within the same clade as *C. nova* and *C. chalciolepis* (i.e., there are four other taxa that overlap geographically; www.rmh.uwyo.edu/data/search.php), which would both complement the aforementioned modeling approaches and provide a means for evaluating the impact of species-specific attributes through tests of phylogeographic concordance among taxa with similar microhabitats (e.g., see Whiteley et al. 2004; Burney and Brumfield 2009).

Deterministic processes reflected in discordant phylogeographic patterns

When both deterministic and stochastic events contribute to observed patterns of genetic variation, how do we recognize their relative contributions? Phylogeographic concordance has been the primary method to identify deterministic processes, whether to identify biogeographic processes structuring genetic variation of whole communities (Avise 1992; Soltis et al. 2006), to identify how past climate changes have impacted co-distributed species (Hewitt 2004), or to prioritize areas for biological conservation (Moritz 1994). Notwithstanding the merit of such study, focusing on concordance obviously limits the types of deterministic processes that can be investigated. For example, tests of concordance across disparate organisms (e.g., Carstens and Richards 2007) does not allow for tests about the genetic consequences of a trait that varies in a species-specific fashion (like microhabitat affinity) because of the lack of a control (i.e., more than the particular trait of interest differs across taxa). Moreover, focusing on concordant patterns of genetic variation may also introduce a bias in our general perception of the relative

predominance of factors structuring genetic variation. For example, biogeographic barriers may shape gene flow patterns, as a vast number of phylogeographic studies attest (Avise 2000). The importance of factors that act in a species-specific manner, such as the impact of habitat specialization on patterns of genetic variation (Neuenschwander et al. 2008; Knowles and Alvarado-Serrano 2010), may go undetected because of a study's design (e.g., when taxa differ in many characteristics, as discussed above).

Even if there is a signal of species-specific traits on the patterns of genetic variation, it may go overlooked, mistakenly assigned to chance because of the tendency to interpret the lack of concordance as being symptomatic of the vagaries of history (e.g., Taberlet et al. 1998; DeChaine and Martin 2005b; Kropf et al. 2012). This is in some ways not surprising given the historical focus of the field on geologic and environmental influences that affect taxa in a similar, deterministic fashion, when they predominate (Avise 2000). Alternatively, and as our results support, the lack of concordance may also reflect deterministic processes associated with species-specific traits. Elucidating how phylogeographic processes are influenced by these traits, whether the traits encompass microhabitat requirements, species interactions, or differing degrees of habitat specialization, will lead to a better resolution of the interactions of taxa with their environments and the resulting consequences for divergence and diversification processes. The only difference is the practical challenges associated with their study.

Research into the effects of climate change on communities is one area where phylogeographic biases may have a profound impact. Our study suggests that it may be quite difficult to make generalizations about the effects of climate-induced distributional shifts on whole communities. In particular, despite taxa sharing the strenuous physiological demands of montane environments and the disturbance regime imposed by glaciations on suitable habitat, microhabitat preferences may interact with glaciations to result in fundamental differences in the past distributions of presently co-distributed species. This isn't to deny that observed concordance in past studies (Carstens and Richards 2007) reflects the common effects of climate change. However, given the broad geographic scales of such phylogeographic studies (e.g., montane taxa and the disjunction between populations in the Rocky Mountains and the Cascade Range in the Pacific Northwest), the conclusions may not be applicable at local geographic scales, which are arguably the most relevant to the persistence of populations across a taxon's range (Hanski 1998; Saccheri et al. 1998).

4.6 Literature Cited

Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. Int. J. Plant Sci. 164:S165–S184.

Alvarado-Serrano, D. F., and L. L. Knowles. 2013. Ecological niche models in phylogeographic studies: applications, advances, and precautions. Mol. Ecol. Resour.: In press.

Avise, J. C. 1992. Molecular Population Structure and the Biogeographic History of a Regional Fauna: A Case History with Lessons for Conservation Biology. Oikos 63:62–76.

Avise, J. C. 2000. Phylogeography: The History and Formation of Species. Havard Univ. Press, Cambridge, MA.

Beavis, A. S., P. Sunnucks, and D. M. Rowell. 2011. Microhabitat preferences drive phylogeographic disparities in two Australian funnel web spiders. Biol. J. Linn. Soc. 104:805–819.

Betancourt, J. L., Van Devender, T. R., and P. S. Martin, eds. 1990. Packrat Middens: the Last 40,000 Years of Biotic Change. Univ. Ariz. Press, Tucson, AZ.

Braconnot, P., B. Otto-Bliesner, S. Harrison, S. Joussaume, J.-Y. Peterchmitt, A. Abe-Ouchi, M. Crucifix, E. Driesschaert, T. Fichefet, C. D. Hewitt, M. Kageyama, A. Kitoh, A. Laine, M.-F. Loutre, O. Marti, U. Merkel, G. Ramstein, P. Valdes, S. L. Weber, Y. Yu, and Y. Zhao. 2007. Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum–Part 1: experiments and large-scale features. Clim. Past 3:261–277.

Brochmann C., T. M. Gabrielsen, I. Nordal, J. Y. Landvik, and R. Elven. 2003. Glacial survival or *tabula rasa*? The history of North Atlantic biota revisited. Taxon 52:417–450.

Brunsfeld, S. J., and J. Sullivan. 2005. A multi-compartmented glacial refugium in the northern Rocky Mountains: Evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). Conserv. Genet. 6:895–904.

Burney, C. W., and R. T. Brumfield. 2009. Ecology Predicts Levels of Genetic Differentiation in Neotropical Birds. Am. Nat. 174:358-368.

Carnaval, A. C., M. J. Hickerson, C. F. B. Haddad, M. T. Rodrigues, and C. Moritz. 2009. Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. Science 323:785–789.

Carstens, B. C., S. J. Brunsfeld, J. R. Demboski, J. M. Good, and J. Sullivan. 2005. Investigating the evolutionary history of the Pacific Northwest Mesic Forest Ecosystem: Hypothesis testing within a comparative phylogeographic framework. Evolution 59:1639–1652.

Carstens B. C., and C. L. Richards. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. Evolution 61:1439–1454.

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.

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

Cavender-Bares J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz. 2004. Phylogenetic Overdispersion in Floridian Oak Communities. Am. Nat. 163:823–843.

Chunco, A. J., J. S. McKinnon, and M. R. Servedio. 2007. Microhabitat variation and sexual selection maintain male color polymorphisms. Evolution 61:2504–2515.

DeChaine, E. G., and A. P. Martin. 2005a. Marked genetic divergence among sky island populations of *Sedum lanceolatum* (Crassulaceae) in the Rocky Mountains. Am. J. Bot. 92:477–486.

DeChaine, E. G., and A. P. Martin. 2005b. Historical biogeography of two alpine butterflies in the Rocky Mountains: broad-scale concordance and local-scale discordance. J. Biogeogr. 32:1943–1956.

Earl, D. A., and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Ehlers, J., and P. L. Gibbard, eds. 2004. Quaternary Glaciations - Extent and Chronology II: North America. Elsevier, London, U.K.

Engelhardt, B. E., and M. Stephens. 2010. Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis. PLoS Genet. 6:e1001117.

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.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1988. Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. Nature 336:61–64.

Gao, H., S. Williamson, and C. D. Bustamante. 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635–1651.

Galbreath, K. E., D. J. Hafner, and K. R. Zamudio. 2010. Isolation and introgression in the Intermountain West: contrasting gene genealogies reveal the complex biogeographic history of the American pika (*Ochotona princeps*). J. Biogeogr. 37:344–362.

Gompert, Z., M. L. Forister, J. A. Fordyce, C. C. Nice, R. J. Williamson, and C. A. Buerkle. 2010. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of Lycaeides butterflies. Mol. Ecol. 19:2455–2473.

Grahame, J. W., C. S. Wilding, and R. K. Butlin. 2006. Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. Evolution 60:268-278.

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

Hartl, D. L., and A. G. Clark. 2006. Principles of Population Genetics. 4th ed. Sinauer Associates, Sunderland, MA.

He, Q., D. L. Edwards, and L. L. Knowles. 2013. Integrative testing of how environments from the past to the present shape genetic structure across landscapes. Evolution 67:3386-3402.

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

Hickerson, M. J., B. C. Carstens, J. Cavender-Bares, K. A. Crandall, C. H. Graham, J. B. Johnson, L. Rissler, P. F. Victoriano, and A. D. Yoder. 2010. Phylogeography's past, present, and future: 10 years after. Mol. Phylogenet. Evol. 54:291–301.

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.

Hixon, M. A., and J. P. Beets. 1993. Predation, Prey Refuges, and the Structure of Coral-Reef Fish Assemblages. Ecol. Monogr. 63:77–101.

Hodges, K. M., D. M. Rowell, and J. S. Keogh. 2007. Remarkably different phylogeographic structure in two closely related lizard species in a zone of sympatry in south-eastern Australia. J. Zool. 272:64-72.
Hohenlohe, P. A., J. Catchen, and W. A. Cresko. 2012. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. Methods Mol. Biol. 888:235–260.
Holsinger, K. E., and B. S. Weir. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. Nat. Rev. Genet. 10:639–650.
Holway, J. G., and R. T. Ward. 1965. Phenology of Alpine Plants in Northern Colorado. Ecology 46:73–83.
Hugall A., C. Moritz, A. Moussalli, and J. Stanisic. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proc. Natl. Acad. of Sci. USA 99:6112–6117.
Jensen, J. L., A. J. Bohonak, and S. T. Kelley. 2005. Isolation by distance, web service (v.3.23 http://ibdws.sdsu.edu/). BMC Genet. 6:13.
Jombart, T. 2008. *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.
Jombart, T., D. Pontier, and A. B. Dufour. 2009. Genetic markers in the playground of multivariate analysis. Heredity 102:330–341.
Knowles, L. L. 2009. Statistical Phylogeography. Annu. Rev. Ecol. Evol. Syst. 40:593–612.
Knowles, L. L., and D. F. Alvarado-Serrano. 2010. Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. Mol. Ecol. 19:3727–3745.
Körner, C. 2003. Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems. 2nd ed. Springer-Verlag, Berlin.
Kropf, M., J. W. Kadereit, and H. P. Comes. 2003. Differential cycles of range contraction and expansion in European high mountain plants during the Late Quaternary: insights from *Pritzelago alpina* (L.) O. Kuntze (Brassicaceae). Mol. Ecol. 12:931–949.
Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.
Losos, J. B., K. I. Warheit, and T. W. Schoener. 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. Nature 387:70–73.
Martinkova, N., R. Barnett, T. Cucchi, R. Struchen, M. Pascal, M. Pascal, M. C. Fischer, T. Higham, S. Brace, S. Y. W. Ho, J. Quere, P. O'Higgins, L. Excoffier, G. Heckel, A. R. Hoelzel, K. M. Dobney, and J. B. Searle. 2013. Divergent evolutionary processes associated with colonization of offshore islands. Mol. Ecol. 22:5205–5220.
Moritz, C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. Mol. Ecol. 3:401–411.
Nei, M., and S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford Univ. Press, NY.
Neuenschwander, S., C. R. Largiader, N. Ray, M. Currat, P. Vonlanthen, and L. Excoffier. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. Mol. Ecol. 17:757–772.
Papadopoulou, A., I. Anastasiou, B. Keskin, and A. P. Vogler. 2009. Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: the effect of dispersal ability and habitat preference. Mol. Ecol. 18:2503-2517.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190:231–259.
Pierce, K. L. 2003. Pleistocene glaciations of the Rocky Mountains. Devel. Quat. Sci. 1:63–76.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
R Core Team. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Renaut, S., G. L. Owens, and L. H. Rieseberg. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Mol. Ecol. 23:311-324.
Ribas, C. C., A. Aleixo, A. C. R. Nogueira, C. Y. Miyaki, and J. Cracraft. 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. Philos. Trans. R. Soc., B 22:681–689.
Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138.
Rosenblum, E. B. 2006. Convergent evolution and divergent selection: lizards at the White Sands ecotone. Am. Nat. 167:1–15.
Ryan, P. G., P. Bloomer, C. L. Moloney, T. J. Grant, and W. Delport. 2007. Ecological Speciation in South Atlantic Island Finches. Science 315:1420–1423.
Saccheri, I., M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius, and I. Hanski. 1998. Inbreeding and extinction in a butterfly metapopulation. Nature 392:491–494.
Schluter, D., and P. R. Grant. 1983. Determinants of Morphological Patterns in Communities of Darwin's Finches. Am. Nat. 123:175–196.
Schönswetter, P., I. Stehlik, R. Holderegger, and A. Tribsch. 2005. Molecular evidence from glacial refugia of mountain plants in the European Alps. Mol. Ecol. 14:3547–3555.
Shafer, A. B. A., C. I. Cullingham, S. D. Côté, and D. W. Coltman. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Mol. Ecol. 19:4589–4621.
Soltis, D. E., M. A. Gitzendanner, D. D. Strenge, and P. S. Soltis. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Syst. Evol. 206:353–373.
Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. Mol. Ecol. 15:4261–4293.
Spellman, G. M., B. Riddle, and J. Klicka. 2007. Phylogeography of the mountain chickadee (*Poecile gambeli*): diversification, introgression, and expansion in response to Quaternary climate change. Mol. Ecol. 16:1055–1068.
Stehlik, I., F. R. Blattner, R. Holderegger, and K. Bachmann. 2002. Nunatak survival of the high Alpine plant *Eritrichum nanum* (L.) Gaudin in the central Alps during the ice ages. Mol. Ecol. 11:2027–2036.
Taberlet, P., L. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7:453–464.
Thompson, R. S., C. Whitlock, P. J. Bartelin, S. P. Harrison, and W. G. Spaulding. 1993. Climatic changes in the western United States since 18,000 yr B.P. Pp 468-513 *in* H. E. Wright Jr., J. E. Kutzbach, W. F. Ruddiman, F. A. Street-Perrott, and T. Webb III, eds. Global climates since the last glacial maximum. Univ. of Minnesota Press, Minneapolis, MN.

Thompson, R. S., and K. H. Anderson. 2000. Biomes of western North America at 18,000, 6000, and 0 ^{14}C yr_{BP} reconstructed from pollen and packrat midden data. J. Biogeogr. 27:555-584.

Turner, T. F., and J. C. Trexler. 1998. Ecological and historical associations of gene flow in darters (Teleostei: Percidae). Evolution 52:1781-1801.

Whiteley, A. R., P. Spruell, and F. W. Allendorf. 2004. Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. Mol. Ecol. 13:3675-3688.

CHAPTER 5

Integrative tests support the role of microhabitat preference in shaping genetic structure

5.1 Abstract

Deterministic processes may affect co-distributed species' phylogeographic patterns in unique ways, yet incorporating them into a statistical framework where they can be explicitly tested remains challenging. Here, we construct spatially and temporally dynamic models to investigate whether two closely related montane species' microhabitat preferences caused them to interact differently with Pleistocene glaciations. In particular, we use environmental niche models from the Last Glacial Maximum and the present to perform demographic simulations through time. This information is integrated into tests of alternative models – in which glaciated regions are either a permeable barrier or a complete barrier to gene flow – using spatially explicit coalescent simulations and approximate Bayesian computation. The results indicate that genomic data are consistent with predictions about the role of microhabitat in shaping patterns of genetic variation. Specifically, genomic variation in *Carex nova*, a wet-adapted species, supports a model of isolated populations around the margins of glaciated areas that were impermeable during glacial periods, which results in a higher level of partitioning of genomic variation compared to *C. chalciolepis*, a dry-adapted species. In contrast, *C. chalciolepis'* patterns of genomic variation support a model in which populations traversed glaciated areas, with individuals dispersing from locally persistent populations, as well as establishing populations at lower elevations. Our results imply that differences in the species' phylogeographic patterns have a biological component. That is, discordant genomic patterns between species can be explained by models formulated to capture species-specific responses to climate change. As such, our work illustrates how such inferences can be used to distinguish whether deterministic processes contribute to observed discord in patterns of genomic variation, rather than assuming such discord reflects the idiosyncrasies of historical processes.

5.2 Introduction

Understanding the contribution of species-specific attributes to observed patterns of genetic variation is critical for determining why taxa responded similarly (or dissimilarly) to historical climate changes. However, such inferences often are out of reach. For example, it is not possible to rule out alternative causes for observed patterns with correlative analyses between genetic variation and geography. Newly developed methodologies that model molecular expectations in a spatially explicit framework can overcome such limitations, and even have the potential to capture how species-specific attributes may structure population genetic variation. For example, they can capture how aspects of habitat requirements or differences in habitat stability across a landscape might influence genetic divergence (e.g., He *et al.* 2013). However, given that the number of biologically informed models that could be tested for any given study system is limitless, the challenge is how to decide what models to test (Knowles 2009).

Deciding which hypotheses to test necessarily limits the scope of analyses (even when testing multiple hypotheses in a Bayesian framework; see Knowles & Sukumaran 2015). Specifically, even though it might be possible to compare over a hundred different models (e.g., Pelletier and Carstens 2014), the insights provided by selecting one model over another will be limited if the differences among the models are trivial (e.g., the models differ in parameters that are more or less equivalent to nuisance parameters in that such parameters do not impact their biological interpretation). Such inherent constraints of model-based approaches (but see O'Meara et al. 2015 for a dissenting point of view) reinforce the importance of developing models that capture processes of biological interest (Papadopoulou & Knowles 2015).

Given that the number of hypotheses that can be tested is constrained, choosing among models to test will determine the potential insights that statistical evaluations might provide. Therefore, it is not the statistical approach per se, but the creativity and intimate knowledge of a study system that a researcher brings to such tests that ultimately determines how much insight will be extracted from patterns of genetic variation. In other words, rather than building a narrative centered upon how well the data should fit a predicted evolutionary scenario (e.g., tests for a correlation under an isolation by distance model), we can use narratives derived from the natural history, as well as the ecological and climatic history of a region, to generate alternative hypotheses to be tested (Knowles & Sukumaran 2015). In conjunction with data simulated under models of such hypotheses, the empirical genetic data then forms the basis for rigorous statistical

tests to select among models (see also Bruggeman *et al*. 2010; Epperson *et al*. 2010; Landguth *et al*. 2010; Morgan *et al*. 2011; Shirk *et al*. 2012).

Here we develop models to test the role of habitat specificity, and its potential interactions with shifting distributions associated with climate change, in structuring the geographic distribution of genetic variation in two species of montane sedges (*Carex* L., Cyperaceae). As such, the study is itself more than a detailed analysis of how species-specific properties might determine whether taxa respond similarly (dissimilarly) to climate change. The work illustrates how general narratives about the processes structuring genetic variation can be formalized into models for hypothesis testing (i.e., narratives about the role of microhabitat in structuring genetic variation; see Massatti & Knowles 2014). Specifically, we use integrative distributional, demographic, and coalescent (iDDC) modeling (see He et al. 2013 and references therein for more information about the iDDC procedure and philosophy) to evaluate the relative support for a set of biologically informed models. In particular, we test whether microhabitat differences between two closely related sedge species (*C. nova* and *C. chalciolepis*) caused the impact of the Pleistocene glaciations to differ between species. Briefly, *C. nova*, an inhabitant of wet microhabitats within high elevation montane habitat in the southern Rocky Mountains (i.e., occurring within drainages), may have been disproportionately displaced to lower elevations (relative to dry-adapted species) because of the greater accumulation of snow and ice in this microhabitat during glacial periods. As a consequence of population persistence being limited to lower elevations around the margins of glaciated habitat, populations of *C. nova* may have been relatively isolated during the glacial periods that were dominant throughout the Pleistocene (i.e., most of the time during the Pleistocene is represented by glacial, not interglacial, periods). This narrative contrasts with plant species that inhabit meadows, ridges, and slopes, such as *C. chalciolepis*. Inhabitants of drier microhabitats may have persisted in high elevation areas that remained relatively undisturbed (i.e., free of glaciers and persistent snow) throughout glacial cycles, in addition to populations that may have been established at lower elevations. As a result, *C. chalciolepis* populations may have remained more interconnected throughout glacial periods, and as a result, might show correspondingly less geographic partitioning of their molecular variation today. The geographic patterning of population structure observed in these two species is consistent with such a narrative (as described in Massatti & Knowles, 2014). However, evaluating the extent to which differences in the patterns of genetic variation can be ascribed to

microhabitat differences requires formalization of these narratives into models for statistical testing.

To test whether current genetic structure results from a species' ability/inability to persist within mountains during glacial periods, we constructed two alternative models to test under the iDDC framework. The iDDC methodology utilizes a population demographic model to make explicit predictions for patterns of genetic variation (see also Currat & Excoffier 2004; Wegmann *et al*. 2006) accounting for both spatial and temporal heterogeneity in climatically suitable areas (e.g., based on information derived from ecological niche models, ENMs, for the present and during the Last Glacial Maximum, LGM; see details in He et al. 2013). For the models tested here, species either persisted within glaciated areas or were excluded, and coalescent simulations were used to predict the impact of these differences in population persistence on patterns of genetic variation. As a result, predicted patterns of genetic variation are species specific. Using these simulations and approximate Bayesian computation (ABC; see Beaumont *et al.* 2002 for an overview of ABC), we test whether the most probable model for each species supports the hypothesis that microhabitat differences determine how species respond to shifts in climate, as well as conduct tests of model validation.

5.3 Methods

Genomic data generation and processing

Carex chalciolepis ($N = 110$) and *C. nova* ($N = 109$) leaf material was field collected from 11 sampling localities across the full extent of their geographical ranges in the southern Rocky Mountains (Fig. 5.1; Table 5.1). Within populations, the collecting distance among individuals was

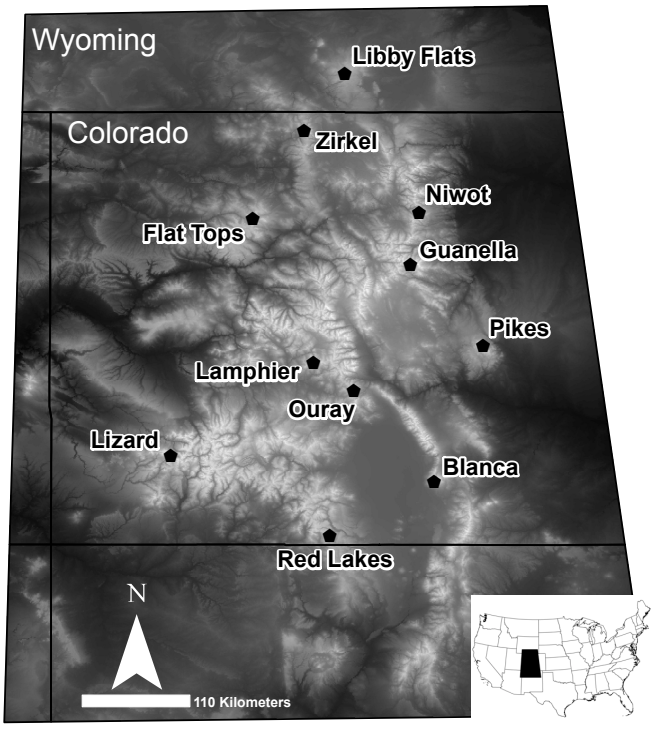

Figure 5.1 Collecting localities throughout the southern Rocky Mountains (for population details, see Table 5.1). Whiter colors represent higher elevations.

maximized to decrease the probability of sampling related individuals (average distance between samples of 300 m, and a minimum distance of 35 m). Leaf material was stored in silica gel until DNA was extracted with DNeasy Plant Mini Kits (Qiagen, Hilden, Germay) following the manufacturer's protocol. As with previous libraries (see Massatti & Knowles 2014), anonymous genomic loci were developed using a restriction associated DNA sequencing (RADseq) approach (for details see Peterson *et al.* 2012); library construction and data processing is described in full detail in the Supporting Information. Briefly, fragments ranging in size from 400 to 500 base pairs were sequenced at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, Canada) to generate 50 base pair, single-end reads. Reads were demultiplexed and single nucleotide polymorphisms (SNPs) were identified using a multinomial-based likelihood model that accounts for sequencing error implemented in Stacks v1.25 (Hohenlohe *et al*. 2010; Catchen *et al*. 2011; Catchen *et al*. 2013).

Five Illumina 2500 sequencing runs were used to generate data for this project. In order to i) maximize the number of unlinked loci for SPLATCHE2 modeling, ii) reduce missing data to the fullest extent, iii) maximize the number of individuals per population, and iv) ensure that the subsampled SNP datasets displayed the same genetic patterns among populations as the larger, unfiltered datasets, we employed the following post-processing procedures. Only RADseq loci containing up to three SNPs and four alleles were used to construct the species' datasets. For each RADseq locus, one randomly selected SNP was exported into a STRUCTURE-formatted file if the locus contained less than 50% missing data. Patterns of genome-wide SNP variation among each species' individuals and populations were then visualized with PCAs in R (R Core Team 2014) using the 'adegenet' package (Jombart 2008) and the 'dudi.pca' function; missing data were replaced by the mean frequency of the corresponding allele. Subsequently, we minimized missing data by iteratively removing SNPs and individuals containing an excess of missing data and rechecking PCAs to ensure that the subsampling procedure did not alter the major axes of genetic variation among populations.

Custom scripts were used to convert the STRUCTURE-formatted files into ARLEQUIN-formatted files, which were input into ARLSUMSTAT to create the empirical summary statistics used in modeling and model selection (see below). We also used the empirical ARLEQUIN-formatted files to create masks that were applied to the simulated data sets so the amount and pattern of missing data in the simulated data would match that of the empirical datasets.

iDDC procedure

We modeled expected patterns of genetic variation for each species under two historical scenarios: i) glaciers as barriers, and ii) persistence within glaciated habitat (Fig. 5.2). Specifically, simulated datasets were generated for these scenarios using spatially and temporally explicit demographic models that were informed by habitat suitabilities extracted from past and present ENMs. Subsequent coalescent simulations were informed by parameter combinations used in the demographic modeling. These steps are detailed below, along with the approximate Bayesian computation (ABC) procedure used for model selection and validation of the parameter estimations. In sum, we test whether, based on empirical data from *C. chalciolepis* and *C. nova*, differences in species' microhabitat affinities translate into species-specific patterns of genetic variation that match predictions based on an interaction between biological adaptations and the physical environment. ENMs and the settings for demographic modeling in SPLATCHE2 are deposited in Dryad (doi:xxx).

Modeling Habitat Suitability across the Landscape. Habitat suitability across the southern Rocky Mountains during the present and LGM was estimated for *C. chalciolepis* and *C. nova* with MAXENT v3.3.3e (Phillips *et al*. 2006). Nineteen bioclimatically informative variables for the present (WorldClim v1.4; Hijmans *et al*. 2005) and the LGM (PMIP2-CCSM; Braconnot *et al.* 2007) were used to generate the environmental niche models (ENMs); full procedural details are presented in the Supporting Information. Georeferenced distribution points used in the modeling were representative of species' entire ranges throughout western North America and were collected from personal fieldwork and validated voucher specimens housed at the Rocky Mountain Herbarium (species distribution points are available at dryad:xxx). To have a tractable number of cells for demographic simulations (detailed below), we statistically downscaled the cell sizes of the ENMs to 0.42 decimal degrees (~16.5 km^2 per cell)(e.g., Ray *et al*. 2010; He et al. 2014). Subsequently, the values of the cells in the LGM and present ENMs denoting the logistic habitat suitability scores (continuously ranging from 0 to 1) determined by MAXENT were reassigned. Specifically, the logistic values were grouped into ten categories using the 'equal interval' clustering method in ArcMap 10.0 (ESRI, Redlands, California, USA) and assigned values ranging from 1-10.

Four landscapes were generated for use in iDDC modeling in which different arrangements of habitat suitabilities represented different time periods and/or ecological conditions (Fig. 5.2). These included a landscape configuration used to initiate the simulations (generations 1-750; Fig 5.2a), two alternative landscapes in which glaciated areas were either an inhospitable barrier to dispersal or were permeable due to persistent populations (generations 751-1750; Fig. 5.2b or 5.2c), and a landscape with habitat suitabilities informed from contemporary climatic conditions (generations 1751-2083; Fig. 5.2d). Note that with a generation time of 3 years for these high elevation species (Körner 2003), and the scaling of the generations by a factor of 15 (to make simulations computationally tractable), the combined generations of the three stages span a time period from the present to the LGM and beyond. Specifically, the first 750 generations under the initial landscape (Fig. 5.2a) represent the period from ~93,750 – 60,000 years before present (YBP). This length of time was chosen to provide a sufficient number of generations for all suitable habitat to be occupied over all possible combinations of population demographic parameters. The next 1000 generations were modeled under one of the two alternative landscapes where glaciated areas were either barriers or not (Fig. 5.2b or 5.2c), and represent the height of the last Pleistocene glaciation from 60,000 – 15,000 YBP (which has left an indelible signature on patterns of genetic variation in montane organisms; e.g., Avise & Walker 1998; Hewitt 2000; Carstens & Knowles 2007). The final 333 generations represent the time since the LGM (15,000 YBP to present), as modeled by the landscape with habitat suitabilities informed by contemporary climatic conditions (Fig. 5.2d). Because of this scaling, any biological interpretation of absolute values of population genetic parameters would need to be adjusted accordingly.

To generate expectations for patterns of genetic variation when a species did (or did not) persist within glaciated terrain, the habitat suitability scores from ENMs for *C. nova* and *C. chalciolepis* were averaged (i.e., an average ENM was calculated for the LGM and an average ENM was calculated for the present). Note that the ENMs for both periods were nearly identical in the two species across their distributions in the southern Rocky Mountains (Fig. C.1). By using estimates of habitat suitabilities for the present and past based on averages for the two species, we provide a standardized model that avoids confounding influences of subtle differences in the unique ENMs of the taxa; even subtle differences in species' models might affect their posterior probabilities and deceive model selection. To capture the different

hypothesized effects of glaciers on the taxa, habitat suitabilities for cells corresponding to previously glaciated regions were either rescaled to values of 0, as would be expected if glaciers were a complete barrier to gene flow (Fig. 5.2a), or they were rescaled to values of 15% of the carrying capacity selected for a given simulation (see details below), which corresponds to an impeded, rather than unfettered, dispersal (Fig. 5.2c). Preliminary analyses confirmed that, even at these reduced carrying capacities, gene flow was able to occur among populations separated by glaciers. The specific cells corresponding to glaciated areas were determined by reconstructions of glacial moraines and glacial till (see Ehlers & Gibbard 2004; Colorado Geological Survey http://coloradogeologicalsurvey.org/).

Table 5.1 Sampling localities for *C. chalciolepis* and *C. nova* individuals.

Population	Geographic region	Mountain Range	GIS coordinates	Elevation (m)
Libby Flats	North	Medicine Bow	41.3499, -106.3202	2970 - 3550
Zirkel	North	Park	40.8184, -106.7015	3000 - 3430
Niwot	Central-North	Front	40.0694, -105.6216	3240 - 3630
Flat Tops	Central-North	Flat Tops	39.8649, -107.2383	3170 - 3570
Guanella	Central-North	Front	39.5941, -105.7117	3600 - 3900
Pikes	Central-South	Front	38.8621, -105.0668	3010 - 4200
Lamphier	Central-South	Sawatch	38.6772, -106.6069	3600 - 3900
Ouray	Central-South	Sawatch	38.4329, -106.2407	3400 - 3800
Lizard	South	San Juan	37.8256, -107.9391	3200 - 3700
Blanca	South	Sangre de Cristo	37.5991, -105.4782	3250 - 3800
Red Lakes	South	San Juan	37.1002, -106.4636	3430 - 3540

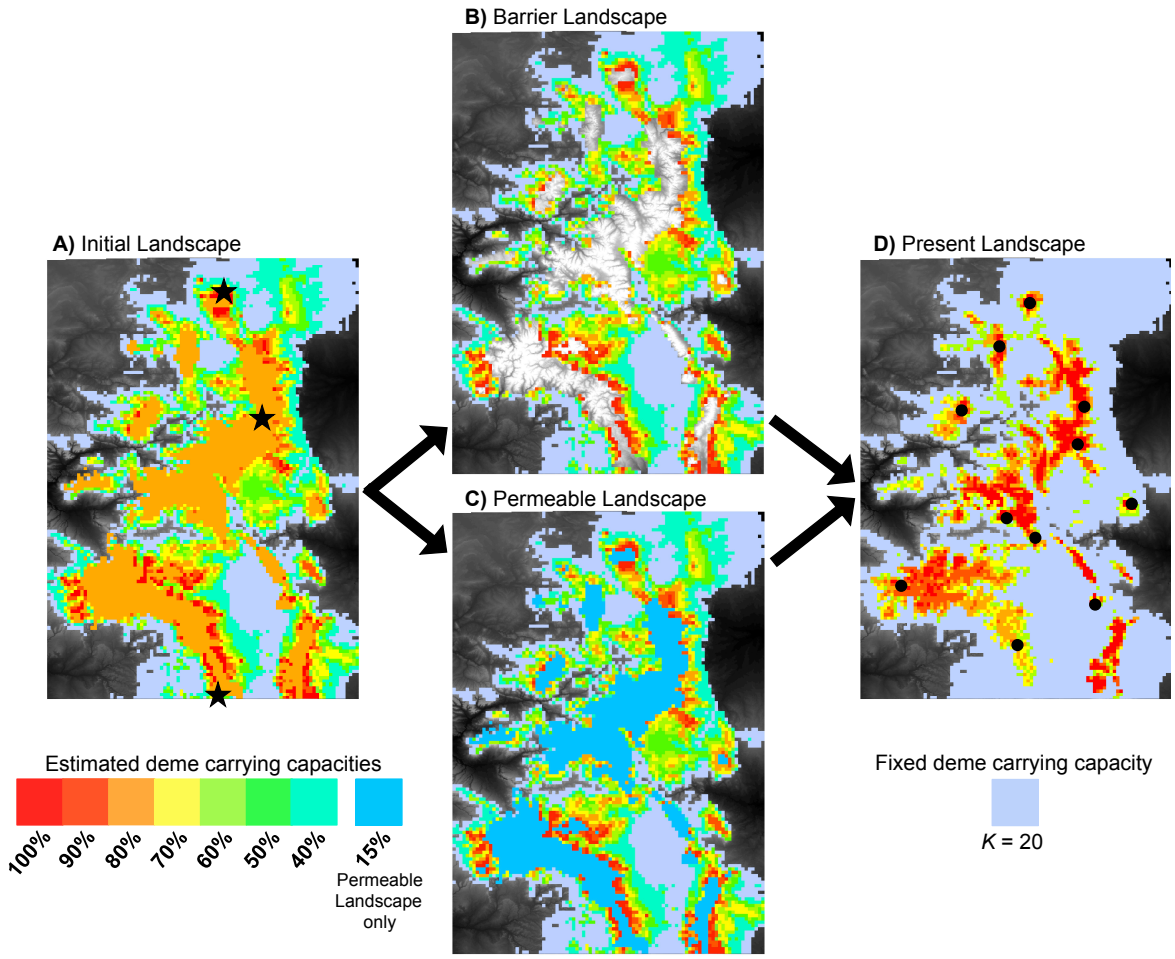

Figure 5.2 Alternative scenarios used in iDDC modeling. Both scenarios start with the same Initial Landscape (A) and end with the same Present Landscape (D). The intermediate landscape varies; the Barrier Landscape (B) represents glaciers completely displacing populations to lower elevations, while the Permeable Landscape (C) represents suitable habitat remaining in glaciated area. Colored demes use either a percentage of the carrying capacity chosen for the current simulation (Estimated deme carrying capacity) or a fixed value (Fixed deme carrying capacity). Colored demes overlay a digital elevation model for the modeling region; wherever the elevation model shows (greyscale cells), demes have a $K = 0$. Black stars in the Initial Landscape indicate where populations were initialized, while black circles in the Present Landscape represent the locations where simulated data was sampled (the same locations in which empirical data was sampled, see Fig. 5.1).

Simulated Datasets - Demographic and coalescent simulations were performed in SPLATCHE2 (Ray *et al.* 2010), where the amount and directionality of gene flow was determined by the habitat suitabilities of a particular landscape (Fig. 5.2). For each of the two models tested (i.e., glaciers as barriers vs. persistence within glaciated habitats), 1,000,000 simulations were generated for each species (4,000,000 total). Uniform priors were used for each population demographic parameter, which included: between-deme migration rate (m), maximum carrying capacity of a deme (K), and the population sizes of the initial populations (N_{Anc}).

All demes' carrying capacities were scaled proportionally to their habitat suitability values (see Fig. 5.2). In other words, the highest quality habitat (demes with values of 10) was allowed to have the full carrying capacity (100%), while the carrying capacities of lower quality habitat were decreased proportionally (i.e., demes with values of 9 had 90% of K, demes with values of 8 had 80% of K, etc.). As noted above, demes directly affected by glaciers were either assigned a K-value of 0 (in the Barrier Landscape) or a K-value reduced to 15% of K (in the Permeable Landscape)(Fig. 5.2b or Fig. 5.2c). A lower bound of $K = 20$ was used for demes whose habitat suitability scores were below the maximum training sensitivity plus specificity threshold identified by MAXENT because of the uncertainty surrounding habitat quality estimates under such conditions (the spatial distribution of such areas is shown in Fig. 5.2 for each of the landscapes). Varying the K-values for demes with uncertain habitat suitability could introduce demographic consequences that would have undue influence on the resulting patterns of genetic variation. Finally, because *C. chalciolepis* and *C. nova* were predicted to occur in and adjacent to montane habitat during the LGM, but not in the lower elevation basins and plains further away from montane areas (which are predominantly represented by demes with values of 1), all demes with habitat suitability values of 1 were assumed to be uninhabitable (see Fig 5.2).

In each demographic simulation, populations were initialized in southern, central, and northern locations (identified in Fig. 5.2a), each with a population size of N_{Anc}. Note that population trees constructed using empirical SNP data for *C. chalciolepis* and *C. nova* clustered populations sampled from these regions (see supplementary material), validating our use of three initial populations. Each generation, m proportion of the population migrates out of the local deme; migration occurs to the adjacent four cells (north, south, west, east). After the exchange of individuals, local demes grow logistically at the rate of 1, regulated by the carrying capacity inferred from habitat suitability.

Spatially explicit coalescent simulations followed every demographic simulation to generate patterns of genetic variation (i.e., genetic variation differed across the landscape depending on the specific combination of *m*, *K,* and N_{Anc}; Excoffier *et al*. 2000; Currat *et al.* 2004). We ran 1142 and 1010 coalescent simulations for *C. chalciolepis* and *C. nova*, respectively, corresponding to each locus for which a SNP was scored in the empirical datasets. Also note that the same number of individuals were sampled in the simulated datasets as were sampled in the empirical data (e.g., same population locations and number of individuals per population), with the same amount of missing data in the simulated data as in the empirical data.

As in the empirical data, nine summary statistics were calculated using ARLSUMSTAT v.3.5.2 (Excoffier & Lischer 2010) for the simulated data. These include the number of segregating sites (*S*) for each population and across populations, mean heterozygosity across loci for each population and across populations (*H*), and pairwise population F_{ST} (Weir & Cockerham 1984), which results in 83 values calculated per demographic simulation.

Model Selection and Validation - We implemented approximate Bayesian computation (ABC) with ABCestimator in ABCtoolbox (Wegmann *et al*. 2010) to select between alternative models (for additional details, see He *et al*. 2013). Specifically, partial least squares (PLS) components (Boulesteix & Strimmer 2007) were extracted from the three estimated parameters using the "PLS" package (Mevik & Wehrens 2007) with boxcox treatment (Box & Cox 1964) in R for the first 20,000 runs for each model; this method removes the effects of interactions between summary statistics and reduces "the curse of dimensionality" associated with using a large number of summary statistics (Boulesteix & Strimmer 2007). Subsequently, we examined the root mean squared error (RMSE) prediction for each parameter to decide how many PLS components should be utilized (Fig. C.2). Five thousand simulations (0.5%) whose summary statistics were closest to those calculated from each species' empirical genomic data were retained from each model and used for parameter estimation and model selection. Postsampling regression adjustment was applied using the ABC-GLM (general linear model) function (Leuenberger & Wegmann 2010) to obtain posterior distributions of the parameters. Bayes factors, which are the ratios between marginal densities of two models, were used for model selection; a higher ratio indicates more support for the first model (Jeffreys 1961). Under the estimated GLM model, the likelihood of the empirical data was evaluated and compared with the

likelihoods of the retained simulations. The fraction of simulations that had a smaller likelihood than the empirical data was shown as *P*-value to check if each model is capable of generating the observed data, with small *P*-values indicating that a model is highly unlikely (Wegmann *et al*. 2010). Finally, a coefficient of variation (R^2) of each parameter explained by the five PLS components was computed and used as an indicator for the power of estimation (Neuenschwander *et al.* 2008).

After selecting the model with the highest support for each species, we validated the accuracy of parameter estimations. One thousand pseudo-observations were generated from prior distributions of the parameters; if the estimation of the parameters is unbiased, posterior quantiles of the parameters from pseudo runs should be uniformly distributed (Cook *et al*. 2006; Wegmann *et al.* 2010). The posterior quantiles of true parameters for each pseudo run were also calculated based on the posterior distribution of the regression adjusted 5000 simulations closest to the pseudo-observation.

5.4 Results

Empirical genomic dataset

Almost 180,000,000 reads were generated for the 110 *C. chalciolepis* individuals (average 1,630,280 ±719,977 per individual); similarly, about 172,000,000 reads were generated for the 109 *C. nova* individuals (average 1,579,380 ±733,329 per individual)(Table C.1). The retention of reads after data processing and assembly with *Stacks* was on average 86% per individual for both species. The initial *C. chalciolepis* dataset contained 24,497 unlinked SNPs (i.e., one SNP from each RADseq locus), with individuals missing 29-94% of the total SNPs (average missing data per individual/SNP was 67%). The final postprocessed genomic dataset contained 101 individuals and 1142 SNPs, with individuals averaging 4.9% missing data. The initial *C. nova* dataset contained 21,706 unlinked SNPs, with individuals missing 29-93% of the total SNPs (average missing data per individual/SNP was 62%). The final postprocessed genomic dataset contained 99 individuals and 1010 SNPs, with individuals averaging 5.3% missing data. Both genomic datasets used in analyses contained less than 15% missing data. When individuals were grouped by sampling locality, all populations for both species were well represented (Table 5.2).

Test of hypotheses and model validation

For the ABC analyses, the first six PLSs contained the majority of information pertinent to the three estimated parameters (Fig. C.2) and were selected to calculate the distance between simulations and the empirical observation. Based on the marginal density calculated from the 5000 closest simulations for each model, the model in which the glaciated areas remained permeable to dispersal and permitted local persistence (Fig. 5.2c) best explains the genetic patterns observed in *C. chalciolepis*, while the model in which the glaciated areas represented a barrier (Fig. 5.2b) best explains the genetic patterns observed in *C. nova* (Table 5.3). In both taxa, the best model had a higher *P*-value than the alternative model, suggesting a better correspondence between the observed empirical data and the simulated data under the more probable model (Table 5.3).

Accuracy estimations of the three demographic modeling parameters differ significantly in *C. chalciolepis* and *C. nova* (Table 5.3 and Fig. 5.3). The posterior probability of the ancestral population size (N_{Anc}) is much flatter than the other two parameters (notice the density of the highest peak) (Fig. 5.3), and there is also the lowest power to estimate N_{Anc}, as indicated by the lowest R^2 values across models (0.379 – 0.497) (Table 5.3). Tests of estimation bias of the parameters show that posterior distributions of N_{Anc} is uniformly distributed in both species, whereas the histograms of the posterior quantiles of *m* and *K* deviate significantly from a uniform distribution for both species (*P*-values <0.05; Fig. 5.4).

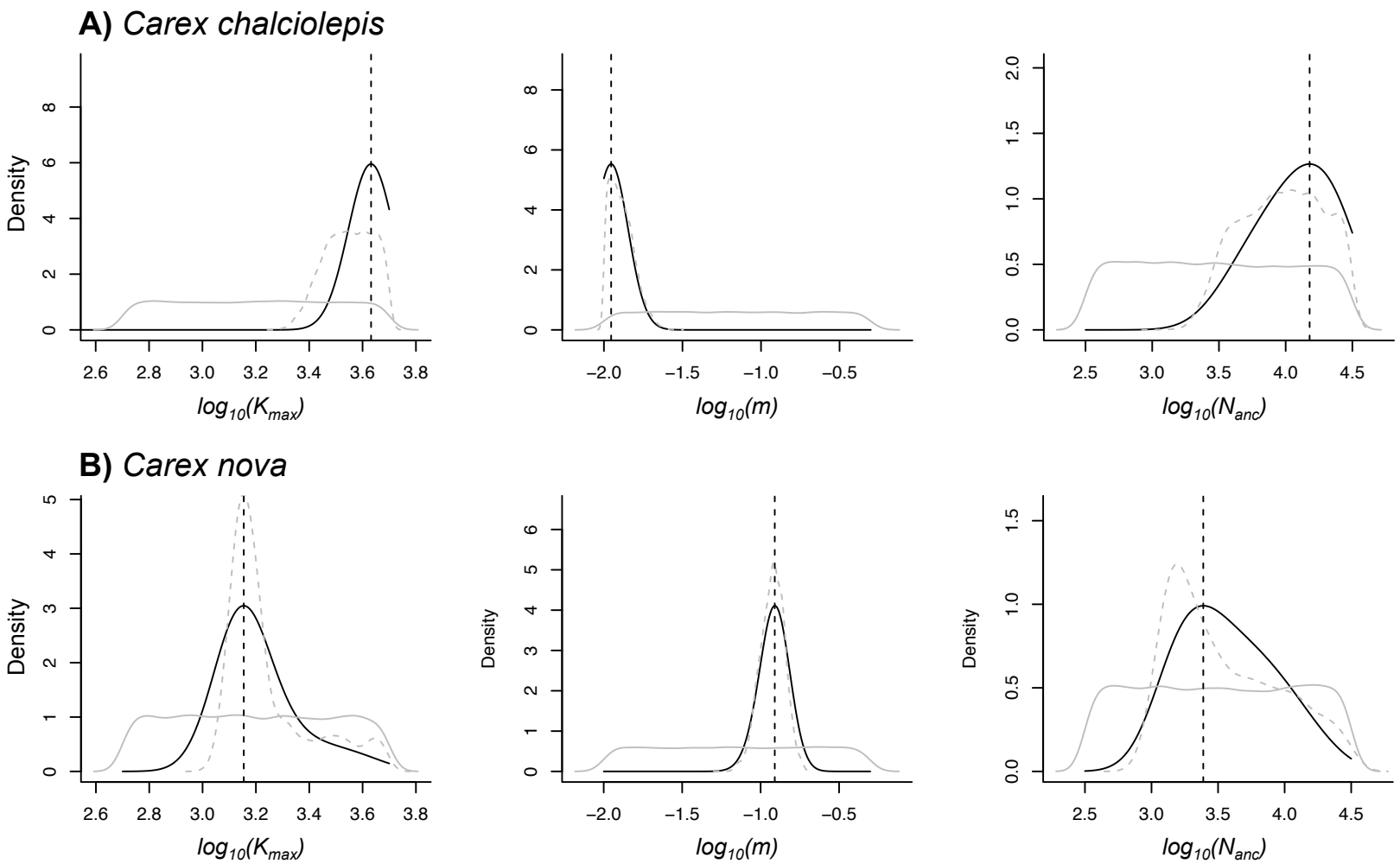

Figure 5.3 Posterior distribution (black line) and mode (vertical dotted line) of parameter estimates for the most probable model for *C. chalciolepis* (Permeable model) and *C. nova* (Barrier model). Results are based on a GLM regression adjustment of the 5000 closet simulations. The distribution of the retained simulations (dashed line) and the prior (gray line) demonstrate the improvement that the GLM procedure had on parameter estimates and that the data contained information relevant to estimating the parameters.

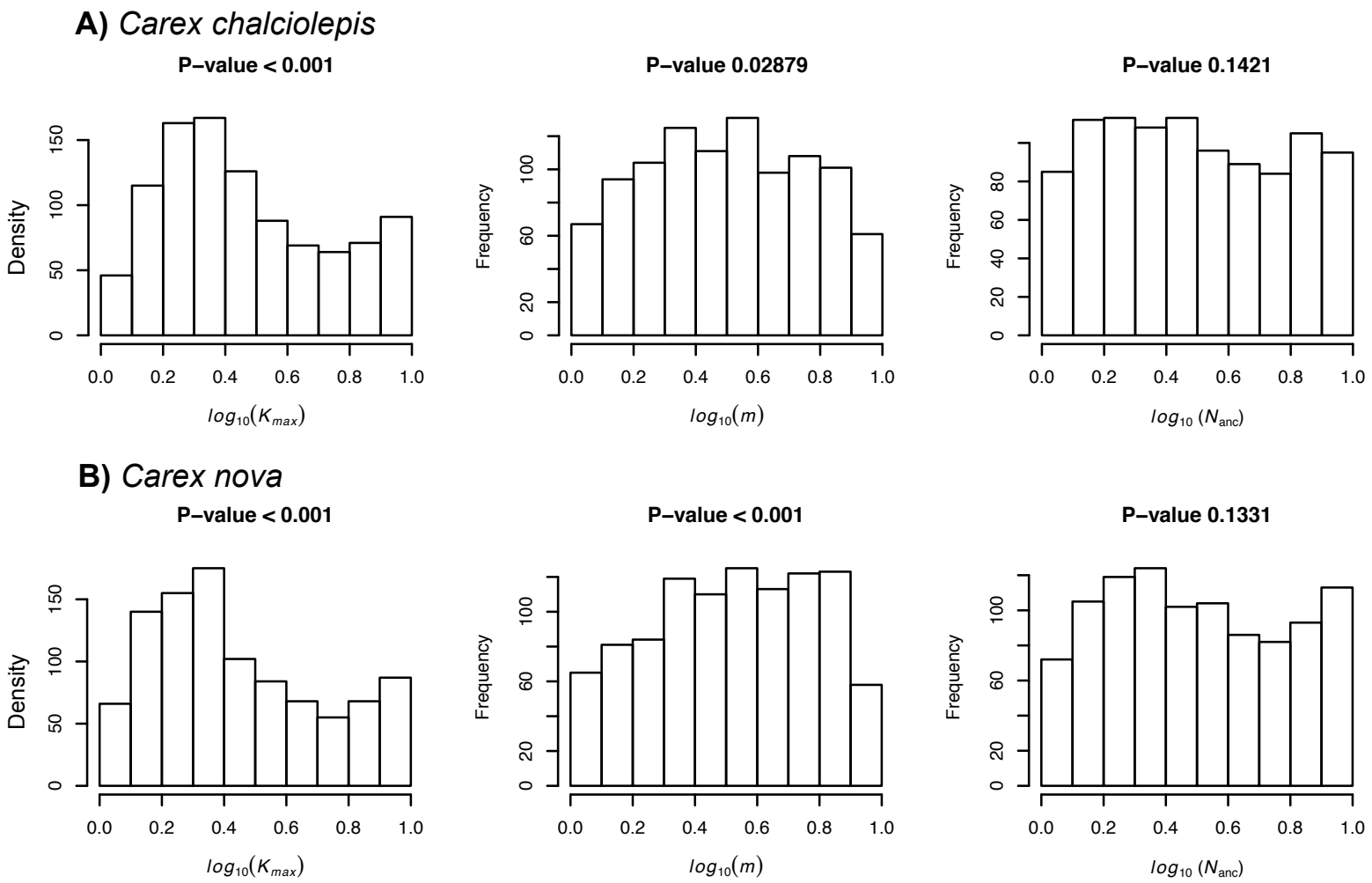

Figure 5.4 Distribution of posterior quantiles of parameters for A) *C. chalciolepis'* Permeable model and B) *C. nova's* Barrier model. The distributions help evaluate potential bias in the parameter estimates, as measured by a departure from a uniform distribution, using a Kolmogorov-Smirnov test. Analyses are based on 1000 pseudo-observations. Estimation of N_{Anc} is unbiased, while the distributions for K and m are too wide for both species.

Table 5.2 Processing data averaged across individuals (± 1 SD) within populations for A) *C. chalciolepis* and B) *C. nova*. The raw reads (Total) were filtered to exclude low quality reads and those without barcodes and a restriction cutsite (No RadTag). The demultiplexed reads (Retained) were further filtered to exclude haploid data (post-bowtie). These data were the input for Stacks, which has additional internal filters to exclude potential paralogs and over-merged loci. The total number of reads used in construction of the final dataset (# retained by Stacks) in also reported as a proportion of the total number of raw reads (total % retained).

A) *Carex chalciolepis*

Population	Total	No RadTag	Low Quality	Retained	post-bowtie	# retained by stacks	total % retained
Libby Flats	1409665 (648596)	2130 (1639)	6225 (2781)	1392172 (640041)	1288722 (594364)	1237680 (587291)	0.87 (0.02)
Zirkel	1317118 (732607)	1985 (1005)	5949 (2874)	1300039 (723972)	1204260 (676465)	1155094 (666738)	0.87 (0.02)
Niwot	1183643 (497878)	2224 (868)	5435 (2311)	1169109 (492840)	1072434 (448979)	1018465 (442073)	0.85 (0.02)
Flat Tops	1447565 (302467)	3903 (2721)	9088 (1126)	1427897 (299391)	1321389 (279198)	1273646 (276458)	0.88 (0.01)
Guanella	2376443 (926967)	6260 (3544)	11307 (4671)	2342005 (913325)	2125198 (820562)	2076221 (821515)	0.87 (0.01)
Pikes	1618699 (211976)	3155 (1065)	10643 (2261)	1597353 (208583)	1493806 (197116)	1446672 (194732)	0.89 (0.01)
Lamphier	1833680 (768785)	3712 (1512)	9862 (4360)	1805993 (758212)	1629347 (671004)	1582182 (668465)	0.86 (0.02)
Ouray	2011868 (721359)	5334 (3244)	10971 (4245)	1867398 (533658)	1698314 (474795)	1644646 (479188)	0.84 (0.08)
Lizard	1906259 (658434)	2803 (1415)	9647 (4856)	1879958 (646192)	1692056 (575632)	1639293 (580534)	0.86 (0.01)
Blanca	1750788 (512394)	7028 (6296)	10238 (3237)	1717764 (502546)	1562420 (458337)	1501078 (448923)	0.85 (0.02)
Red Lakes	1077351 (416764)	1665 (897)	5145 (2298)	1063633 (411406)	974799 (379863)	927497 (372987)	0.85 (0.02)

B) *Carex nova*

Population	Total	No RadTag	Low Quality	Retained	post-bowtie	utilized by stacks	total % retained
Libby Flats	1926102 (639294)	4785 (2616)	10994 (4194)	1892947 (626813)	1727116 (566969)	1685275 (571077)	0.87 (0.01)
Zirkel	1872212 (655942)	5732 (2853)	11141 (3238)	1838018 (646500)	1709975 (606864)	1661876 (608927)	0.88 (0.02)
Niwot	1790854 (645816)	3952 (2152)	10224 (3731)	1760409 (635233)	1641491 (589545)	1596402 (593673)	0.89 (0.02)
Flat Tops	791741 (321133)	2176 (1898)	5053 (2631)	779351 (315807)	724842 (290576)	647542 (290579)	0.76 (0.26)
Guanella	1969498 (961299)	5358 (2858)	9772 (4538)	1940127 (949380)	1763711 (849065)	1712345 (847969)	0.86 (0.01)
Pikes	1037894 (341618)	3548 (1780)	6435 (2325)	1020302 (336852)	967444 (316093)	923417 (316767)	0.88 (0.03)
Lamphier	1543261 (826571)	2696 (1899)	8921 (5006)	1529473 (820181)	1424171 (764452)	1377393 (766317)	0.88 (0.02)
Ouray	1720823 (806145)	4694 (3714)	9548 (4491)	1676352 (796020)	1541581 (735238)	1491146 (735221)	0.86 (0.04)
Lizard	1748043 (560715)	2437 (1471)	8912 (2763)	1722763 (555205)	1558465 (492788)	1509244 (488461)	0.86 (0.02)
Blanca	1241575 (332677)	4358 (3791)	5285 (2015)	1231933 (329821)	1150089 (300909)	1096118 (295411)	0.88 (0.01)
Red Lakes	1765853 (491296)	5783 (6052)	9899 (3136)	1733579 (481989)	1597540 (448828)	1550596 (448903)	0.88 (0.01)

Table 5.3 Model statistics and prior and posterior distributions of estimated parameters. The Bayes factor represents the ratio between the model with the highest marginal density and the alternative. Parameters include: K_{max}, the carrying capacity of the deme with the highest suitability; m, the migration rate per deme per generation; and N_{Anc}, the ancestral population sizes of initial populations before expansion from refugia. Logarithms of all priors are uniformly distributed and have the same ranges across models. R^2 is the coefficient of determination between a parameter and the six PLSs used herein. HPD 50 and HPD 90 represent the 50% and 90% parameter regions with the highest posterior density.

							Posterior		
Species	Model	Marginal density (P-value)	Bayes factor	Parameters	Prior [min, max]	R^2	Mode	HPD 50	HPD 90
C. chalciolepis	Barrier	4.87×10^{-5} (0.650)	2.84	$\log_{10}(K_{\max})$	[2.7, 3.7]	0.642	3.7	[3.652, 3.700]	[3.579, 3.700]
				$\log_{10}(m)$	[-2, -0.3]	0.966	-1.906	[-1.959, -1.855]	[-2.000, -1.776]
				$\log_{10}(N_{Anc})$	[2.5, 4.5]	0.404	4.296	[4.116, 4.452]	[3.745, 4.500]
	Permeable	1.38×10^{-4} (0.970)	-	$\log_{10}(K_{\max})$	[2.7, 3.7]	0.698	3.632	[3.588, 3.674]	[3.514, 3.700]
				$\log_{10}(m)$	[-2, -0.3]	0.965	-1.954	[-2.000, -1.908]	[-2.000, -1.794]
				$\log_{10}(N_{Anc})$	[2.5, 4.5]	0.379	4.18	[3.953, 4.366]	[3.607, 4.500]
C. nova	Barrier	1.29×10^{-4} (0.844)	-	$\log_{10}(K_{\max})$	[2.7, 3.7]	0.548	3.154	[3.067, 3.248]	[2.951, 3.476]
				$\log_{10}(m)$	[-2, -0.3]	0.961	-0.91	[-0.976, -0.845]	[-1.075, -0.753]
				$\log_{10}(N_{Anc})$	[2.5, 4.5]	0.497	3.389	[3.177, 3.717]	[2.962, 4.196]
	Permeable	5.68×10^{-6} (0.078)	22.69	$\log_{10}(K_{\max})$	[2.7, 3.7]	0.585	3.132	[3.059, 3.208]	[2.948, 3.327]
				$\log_{10}(m)$	[-2, -0.3]	0.962	-0.985	[-1.042, -0.927]	[-1.129, -0.842]
				$\log_{10}(N_{Anc})$	[2.5, 4.5]	0.479	4.084	[3.855, 4.310]	[3.457, 4.500]

5.5 Discussion

Our study highlights how knowledge of species' traits and ecological requirements can be formalized into a statistical framework to test among potential evolutionary scenarios. In doing so, we were able to show that species' responses to glaciations depend upon their microhabitat preferences. Specifically, with patterns of genomic variation in the two species corresponding to either a model in which the glaciers were a permeable, or impermeable, barrier, the results suggest that species adapted to wetter microhabitats were more isolated around the margins of glaciers, while species adapted to drier microhabitats persisted within glaciated regions and remained relatively connected by gene flow. These analyses allowed us to determine *how* historical environmental events shaped present patterns of molecular variation across the landscape; these ideas are easily transferable to other sets of species or different biological systems. Below we discuss what role iDDC modeling may play in evolutionary studies, how modeling helped us resolve between alternative historical scenarios for montane plants, and what implications our results may generally have for montane ecosystems.

The role demographic modeling in resolving historical processes - Using intimate knowledge of the interactions between climate and topography within montane ecosystems, as well as utilizing data from other disciplines (e.g., maps of glacial till and glacial moraines), we constructed models that integrated spatial and temporal variation in order to test whether species differ in their response to glaciated areas during the Pleistocene. Such models go beyond traditional analyses that test, for example, whether species' molecular patterns fit standard null models, such as isolation by distance (Slatkin 1993). For example, He *et al.* (2013) contrast correlative and explicit modeling approaches to investigate different candidate factors structuring the genetic variation in *Lerista lineopunctulata*, an Australian lizard adapted to sandy habitat that was presumptively highly impacted by Pleistocene climatic fluctuations. While the correlative analyses identified isolation by distance as the most likely model explaining the variation among this lizard's populations (i.e., strictly geography with no contribution from the past or present climate), iDDC modeling highly supported the role of environmental variables as well as shifting habitat suitability over time. As this example demonstrates, correlative approaches utilizing general models may mislead research testing factors influencing genetic patterns. Furthermore, integrating species-specific information into carefully designed models may help illuminate the

complexity underlying species histories and the forces that shaped them. Because this or similar modeling approaches have only been applied in a limited number of instances (e.g., Neuenschwander *et al*. 2008; Martinkova *et al*. 2013), we can expect to learn much about any organisms for which sophisticated models are tested. However, applying such modeling to systems in which many generalizations have been made across organisms (e.g., from comparative phylogeographic studies) (e.g., Soltis *et al*. 2006; Shafer *et al.* 2010) may lead to reassessing the status quo. In fact, with genomic data and refined modeling procedures (Papadopoulou and Knowles 2015), we may discover that generalizations are hard to make, because all species react to environmental changes and biotic pressures individualistically (sensu Gleason 1926).

Validation and interpretation of models - We acknowledge that the modeling we apply to investigate historical scenarios is not without difficulties. For example, post-sampling adjustment, such as regression (Beaumont *et al*. 2002) or GLM (Leuenberger & Wegmann 2010), can pose problems when the relationship between parameters and summary statistics is extrapolated beyond the region of the observed data set. Moreover, ABC will always produce a posterior distribution, even if the model is wrong (Bertorelle *et al*. 2010). Accordingly, model validation is pertinent, especially for ABC, which approximates the likelihood of models with summary statistics (Pritchard *et al*. 1999; Beaumont *et al.* 2002) instead of using all of the data like full likelihood-based models (Hey & Nielsen 2004, 2007; Kuhner 2006; Nielsen & Beaumont 2009; Hey 2010). Given the complexity of the models used herein, one concern was whether the data would be sufficient to discriminate between alternative models. We validated the models using several approaches, which suggested that our results are generally robust.

The primary goal of our research was to choose among alternative models depicting species' interactions with historical events. Therefore, we applied a standard rejection sampling scheme (Beaumont *et al.* 2002), which should not create bias among models (Wegmann *et al*. 2010). We not only assessed model support by comparing the marginal densities of each model with Bayes factors, but also by examining the percentage of the simulated data whose likelihood matched that of the empirical data to check if the model is capable of generating the observed data (i.e., a small proportion of simulated datasets with smaller or equal likelihood values as the empirical dataset – a small *P*-value – would suggest the model is not capable of generating the

data). The Barrier model (Fig. 5.2c) is strongly supported for *C. nova* based on the Bayes factor (Table 5.3). In addition, this model has a much higher probability of generating simulations with smaller or equal likelihoods than the empirical observation compared to the Permeable model (Table 5.3). In *C. chalciolepis,* the Barrier and Permeable models have much closer marginal densities, resulting in a Bayes factor that shows only marginally positive support for the Permeable model. However, the Permeable model generates most simulations with equal or smaller likelihoods than the empirical data, while the Barrier model generates far fewer ($P = 0.65$ versus $P = 0.97$; Table 5.3). In other words, even though the some combinations of parameters produce data sets that match the empirical data under the Barrier model for *C. chalciolepis*, the Permeable model has much wider parameter region that generate data close to the observation.

In addition to choosing between models for each species, we estimated parameter values that may illustrate differential population dynamics. The posterior distributions of the parameters are similar between models for both species, only differing significantly in the estimation of N_{Anc} for *C. nova* (Table 5.3). All parameters for both species are informed to different degrees by the PLSs in the ABC analyses; while N_{Anc} has moderate R^2 values (0.38-0.48), the R^2 for the migration rate (m) is very high across models (~0.96). The estimation of maximum carrying capacity (K_{max}) and m both show a biased estimation (i.e., the posterior distribution in too wide) based their non-uniform posterior distributions from pseudo-observations (Fig. 5.4). These parameters affect the change of a series of local parameters, potentially contributing to the difficulty of their estimation (Wegmann *et al.* 2010). Because our primary concern was model selection and not parameter estimation, it is possible to ignore biased parameters, but interpretation of them should also be avoided (Wegmann *et al.* 2010). With this caveat in mind, we only note that *C. chalciolepis* tended to higher values of K_{max} and lower values of m compared to *C. nova*, while the difference in N_{Anc} was more ambiguous. These parameter estimates may inform future investigations about the dynamics of microhabitat affinity and glaciations, for example, whether habitat stability for dry microhabitats facilitated larger population effective sizes and lower rates of migration, while continual disturbance within wet microhabitats fostered relatively lower population sizes and higher rates of migration (i.e., the reestablishment of disturbed populations).

Implications for montane ecosystems - Our results suggest that plant species had different responses to glacial periods depending on their microhabitat preferences. In particular, *C. chalciolepis* was able to persist in ridge and slope microhabitats as well as establish populations in newly available habitat at lower elevations during glacial periods, while *C. nova* was predominantly displaced to suitable habitat at lower elevations around the margins of the glaciers. In turn, these responses facilitated population connectedness within *C. chalciolepis*, while populations of *C. nova* remained in isolation for the temporally dominant glacial periods. While these responses are completely in line with expectations derived from knowledge of the interactions of persistent snow and ice and habitat within montane ecosystems (Ehlers & Gibbard 2004), this is the first study to directly link genetic patterns with this causal mechanism.

Our data are highly suggestive of *in situ* survival for organisms adapted to drier ridge and slope microhabitats in mountain ranges. Survival within glaciated regions has been suggested for other organisms based on the geographic distribution of their molecular variation. For example, Westergaard *et al*. (2011) determine that two arctic-alpine pioneer plant species, *Sagina caespitosa* and *Arenaria humifusa*, survived in available habitat protruding from North European ice sheets based on distinct genetic groups resolved from amplified fragment length polymorphisms. Similarly, Lohse *et al.* (2011) investigate survival within glaciers versus peripheral to glaciers using a group of alpine carabid beetles in the Orobian Alps. They determine that both processes likely affect phylogeographic patterns based on the presence of populations with either paraphyletic or reciprocally monophyletic gene trees. However, these studies do not test whether survival within glaciated regions caused their present molecular patterns. Linking genetic patterns to specific mechanisms is critical because similar genetic patterns may result form different demographic processes (Knowles 2009; Csillery *et al.* 2010).

By combining spatial and temporal heterogeneity into demographic modeling, the iDDC methodology has the capacity to test species-specific historical scenarios in a rigorous and statistical framework. Such modeling, if implemented across a broad range of species representing different traits and adaptations within an ecosystem, may reveal unparalleled insights regarding the response of species to environmental perturbations over time. Montane ecosystems may be especially amenable to these types of studies, because the climatic dynamics over the Pleistocene and their effects on the landscape are relatively well known. Moreover, many species show relatively shallow genetic histories due to their interactions with climatic

oscillations (Avise & Walker 1998; Hewitt 2000; Carstens & Knowles 2007), and species are co-distributed locally and regionally, facilitating a range of comparisons.

5.6 Literature Cited

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society B,* **265**, 1707–1712.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics,* **162**, 2025–2035.

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

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology,* **19**, 2609–2625.

Boulesteix A-L, Strimmer K (2007) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics,* **8**, 32–44.

Box GEP, Cox DR (1964) An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **26**, 211–252.

Braconnot P, Otto-Bliesner B, Harrison S *et al.* (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum–Part 1: experiments and large-scale features. *Climate of the Past,* **3**, 261–277.

Brown JL, Knowles LL (2012) Spatially explicit models of dynamic histories: examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Molecular Ecology,* **21**, 3757–3775.

Bruggeman DJ, Wiegand T, Fernandez N (2010) The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). *Molecular Ecology,* **19**, 3679–3691.

Carstens BC, Knowles LL (2007) Estimating Species Phylogeny from Gene-Tree Probabilities Despite Incomplete Lineage Sorting: An Example from *Melanoplus* Grasshoppers. *Systematic Biology*, **56**, 400–411.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) *Stacks*: Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3 Genes, Genomes, Genetics*, **1**, 171–182.

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

Cook SR, Gelman A, Rubin DB (2006) Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics,* **15**, 675–692.

Csillery K, Blum MGB, Gaggiotti OE, Francois O (2010) Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution,* **25**, 410–418.

Currat M, Ray N, Excoffier L (2004) SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes,* **4**, 139–142.

Currat M, Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biology,* **2**, 2264–2274.

Ehlers J, Gibbard PL eds. (2004) *Quaternary Glaciations - Extent and Chronology II: North America*. Elsevier, London, United Kingdom.

Epperson BK, McRae BH, Scribner K *et al.* (2010) Utility of computer simulations in landscape genetics. *Molecular Ecology,* **19**, 3549–3564.
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity,* **91**, 506–509.
Gleason HA (1926) The Individualistic Concept of the Plant Association. *Bulletin of the Torrey Botanical Club*, **53**, 7-26.
He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution,* **67**, 3386–3402.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.
Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution,* **27**, 905–920.
Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics,* **167**, 747–760.
Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.
Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* **25**, 1965–1978.
Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. PLoS *Genetics*, **6**, e1000862.
Jeffreys H (1961) *Theory of probability*, 3rd ed. Clarendon Press, Oxford, United Kingdom.
Jombart T (2008) *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
Knowles LL (2009) Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 593–612.
Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology,* **19**, 3727–3745.
Körner C (2003) *Alpine plant life: functional plant ecology of high mountain ecosystems.* Springer, Berlin, Germany.
Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics,* **149**, 429–434.
Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics,* **22**, 768–770.
Landguth EL, Cushman SA, Murphy MA, Luikart G (2010) Relationships between migration rates and landscape resistance assessed using individual-based simulations. *Molecular Ecology Resources,* **10**, 854–862.
Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology,* **10**, R25.

Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics,* **184**, 243–252.
Lohse K, Nicholls JA, Stone GN *et al.* (2011) Inferring the colonization of a mountain range-refugia vs. nunatak survival in high alpine ground beetles. *Molecular Ecology*, **20**, 394-408.
Martinkova N, Barnett R, Cucchi T *et al*. (2013) Divergent evolutionary processes associated with colonization of offshore islands. *Molecular Ecology,* **22**, 5205–5220.
Massatti R, Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, **68**, 2833–2846.
Mevik B-H, Wehrens R (2007) The PLS package: principal component and partial least squares regression in R. *Journal of Statistical Software*, **18**, 1–24.
Morgan K, O'Loughlin SM, Chen B *et al.* (2011) Comparative phylogeography reveals a shared impact of Pleistocene environmental change in shaping genetic diversity within nine *Anopheles* mosquito species across the Indo-Burma biodiversity hotspot. *Molecular Ecology,* **20**, 4533–4549.
Neuenschwander S, Largiader CR, Ray N *et al.* (2008) Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Molecular Ecology,* **17**, 757–772.
Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology,* **18**, 1034–1047.
Papadopoulou A, Knowles LL (2015) Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution*, **69**, 1501-1517.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, **7**, e37135.
Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling,* **190**, 231– 259.
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution,* **16**, 1791–1798.
Ray N, Currat M, Foll M, Excoffier L (2010) SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, **26**, 2993–2994.
Renaut S, Owens GL, and Rieseberg LH (2014) Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology,* **23**, 311–324.
Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences, USA,* **108**, 15112–15117.
Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589–4621.
Shirk AJ, Cushman SA, Landguth EL (2012) Simulating pattern-process relationships to validate landscape genetic models. *International Journal of Ecology,* **2012**, 539109.

Slatkin M (1993) Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution*, **47**, 264–279.
Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.
Wegmann D, Currat M, Excoffier L (2006) Molecular diversity after a range expansion in heterogeneous environments. *Genetics,* **174**, 2009–2020.
Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics,* **11**, 7.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution,* **38**, 1358–1370.
Westergaard KB, Alsos IG, Popp M *et al*. (2011) Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. *Molecular Ecology*, **20**, 376-393.

CHAPTER 6

Conclusion

Evolutionary phenomena occurred at multiple spatial and temporal scales for high latitude and high elevation plant species. In this dissertation, I used novel analytical methods and abundant genomic data to investigate patterns of biodiversity within montane and arctic ecosystems. In Chapter 2, I demonstrated how genomic data can be used to construct a resolved and well-supported phylogeny for a clade that was highly affected by Pleistocene climatic oscillations. Subsequently, I used this phylogeny to infer ancestral distributions within *Carex* section *Racemosae* and investigate interactions between Pleistocene climatic fluctuations and clade dispersal and diversification (Chapter 3). Refocusing on a regional scale, I examined closely related and co-distributed species in order to test whether species' traits facilitated different interactions with Pleistocene glaciations (Chapters 4 and 5). Below, I discuss the continental and regional processes that influenced the patterns and accumulation of biodiversity in habitats highly influenced by Pleistocene climatic oscillations, and I suggest future research directions that may further clarify these processes.

The dynamic Pleistocene climate had a profound impact on the habitat available for montane and arctic plant species. This is illustrated well by the shared ancestry of the montane floras of the southern Rocky Mountains in North America and the Altai Mountains in Asia (Weber, 2003). Without a connection between habitat in Asia and North America during glaciations (i.e., the Bering Land Bridge), such floristic similarity would not exist. My analyses in Chapters 2 and 3 support the influence of a bridge between Asia and North America by suggesting that widespread species attained their distributions from more narrowly distributed ancestors during the Pleistocene and by reiterating the similarity of the Asian and North American montane floras (i.e., each of these floras contains multiple clades of *Carex* sect. *Racemosae*). However, my data offer conflicting evidence on the process by which floristic

similarity between Asian and North American mountains accrued. The ancestral range reconstruction that I present in Chapter 3 suggests that ancestral taxa confined to a single continent infrequently dispersed to the adjacent continent (both from Asia to North America and vice versa). While I do not explicitly reconstruct ancestral habitat preferences, the ancestors that dispersed between continents would likely have been adapted to montane habitats, given the habitat preferences of extant species. However, the clades of montane taxa within sect. *Racemosae* are almost always exclusively endemic to either Asia or North America; if montane taxa dispersed between continents and subsequently diversified, we may expect to see montane clades that include species on both continents.

An alternative hypothesis that may explain the endemic montane clades within sect. *Racemosae* requires that species adapted to high latitudes be the progenitors of montane clades. Widely distributed species across high latitudes are known to disperse into montane habitat at more southerly latitudes during glacial periods (Hultén, 1937; Abbott et al., 2000), only to become isolated and subjected to extreme environmental conditions during subsequent interglacial periods. While most of these populations vanish as their habitat disappears, some isolated populations may become reproductively isolated (e.g., Conti et al., 1999). If these new species persist and subsequently diversify regionally, a similar topology would be expected as was resolved for sect. *Racemosae*. This scenario may also be more likely than what is suggested by analyses in Chapter 3 because montane habitat was never continuous between continents, even if it was more widespread across lower elevations during glacial periods. In order to test whether adaptation to montane habitats facilitates diversification, future research could take advantage of Bayesian analysis of macroevolutionary mixtures (Rabosky et al., 2013), which identifies shifts in diversification rates within a phylogeny. Furthermore, using an analytical method that allows a broader spectrum of possible ancestral ranges (e.g., BayArea, Landis et al., 2013) may resolve a different ancestral range reconstruction compared to the method used here. Finally, investigating the breadth of the climatic space occupied by the recently derived clades in sect. *Racemosae* may provide evidence for the distribution of the ancestral sect. *Racemosae* species. For example, determining that the clade including both widespread and endemic species (Clade 3 in Fig. 3.1) represents the entire scope of climatic variability present in Clades 1, 2, 4, and 5 would suggest that these clades are derived from species in Clade 3.

Even though I demonstrate in Chapters 2 and 3 that genomic data provide significant power to resolve relationships among species such that the resulting phylogeny can be confidently used in downstream evolutionary analyses, future phylogenetic studies utilizing these data will benefit from investigations into the ramifications of data processing on the loci available for analysis. For example, Takahashi et al. (2014) discovered that varying the amount of missing data within data matrices affected the topologies of trees with the highest support, and that the best-supported tree did not necessarily match up with expected relationships based on morphology. Their study reinforces the importance of knowing how data processing affects datasets and constructing datasets that are optimized for the inference methods utilized. In addition, because significant rate heterogeneity may exist among genomic loci, future research into the temporal limits of SNP data for resolving phylogenetic relationships, in other words the point at which homoplasy destroys phylogenetic signal, will be critical. This could be easily implemented within *Carex* if, for example, clades more distantly related to sect. *Racemosae* were included in phylogenetic analyses. Finally, while my research demonstrates that herbarium specimens are amenable to next-generation sequencing library construction protocols, the ability to use older herbarium specimens within this framework would represent a significant advance due to the taxonomic diversity of older specimens available from larger institutions.

While Pleistocene climatic oscillations facilitated the repeated isolation of widespread, high latitude taxa in montane habitat and their subsequent diversification, they also drove montane clade diversification at a regional scale. Within mountain ranges impacted by valley glaciers, such as those composing the southern and central Rocky Mountains in North America, cooler climates of the temporally dominant glacial periods presumably necessitated that species' populations be distributed at lower elevations surrounding glaciated terrain. Alternatively, interglacial periods forced species to find refuge from warmer climatic conditions at higher elevations. The cyclical nature of glacial and interglacial periods must have facilitated species' populations repeatedly joining together and fragmenting. While the short durations of glacial and interglacial periods likely preclude strict genetic drift as the main driver of species accumulation in montane clades (TMCS Network, 2012), it may be that isolation in conjunction with strong selective forces promoted diversification (Nosil, 2012). For example, populations at the periphery of a species' range may have experienced environmental conditions outside their normal tolerances (e.g., extreme temperatures or amount of precipitation), and while many of

these populations were likely extirpated, some may have persisted by adapting to these disparate environmental conditions. If such adaptation facilitated change in genomic regions controlling important aspects of the species' life history, for example by shifting flowering time, reproductive barriers may have formed. The plausibility of such a scenario would be higher if adaptation in a hypothetical population occurred during a longer glacial period compared to the relatively short interglacial periods.

In support of the hypothetical scenario developed above, my research in Chapters 4 and 5 suggests that certain species' populations had the propensity to be isolated on the margins of mountain ranges during glacial periods, in particular species adapted to wet microhabitats. Isolation not only occurred, but drift-induced differentiation was greater among wet-adapted populations versus the populations of a dry-adapted species. While my research in Chapters 4 and 5 does not link processes promoting differentiation with those promoting diversification, it does imply a specific macroevolutionary pattern, namely that if microhabitat affinity were to regularly facilitate speciation, many endemic clades within montane ecosystems should contain at least one species adapted to wet microhabitats. Future investigations may benefit from environmental niche models of species' LGM distributions to identify geographic areas that may have been subjected to extreme environmental conditions. If such areas are identified, genomic data similar to what was generated herein may help determine if present-day populations from these areas are more differentiated than expected by chance (i.e., as is seen among populations not expected to have been subjected to extreme environmental conditions during the LGM). Overall, new analytical techniques, such as the iDDC modeling demonstrated in Chapter 5, in combination with carefully designed sampling informed by extensive knowledge of organisms' biological requirements and of the environments in which they live, as demonstrated in Chapter 4, will help researchers decipher the roles of species-specific traits in influencing evolutionary processes.

Many factors influence the evolution of organisms adapted to ecosystems highly impacted by Pleistocene glacial cycles, and here I demonstrate how genomic data and new analytical techniques can be utilized to shed light on processes across spatial and temporal scales. While many aspects of evolutionary science are becoming more computational as the sizes of datasets grow and as analytical methods become more complex, it is heartening that the capacity of evolutionary methods to capture the influence of species' traits and specific

environmental influences is also expanding. In order to decipher the impact of historical processes on organisms, it is as important as ever that a researcher knows the biology of her or his organism and the environmental conditions affecting the organism's habitat, and that the researcher be creative in designing tests to distinguish among alternative hypotheses.

Literature Cited

Abbott, R.J., et al. 2000. Molecular analysis of plant migration and refugia in the Arctic. *Science*. 289(5483): 1343-1346.

Conti, E., et al. Phylogenetic Relationships of the Silver Saxifrages (*Saxifraga* Sect. *Ligulatae* Haworth): Implications for the Evolution of Substrate Specificity, Life Histories, and Biogeography. *Molecular Phylogenetics and Evolution*. 13(3): 536-555.

Hultén, E. 1937. *Outline of the history of arctic and boreal biota during the Quaternary period: their evolution during and after the glacial period as indicated by the equiformal progressive areas of present plant species*. Bokfoerlags Aktiebolaget Thule.

Landis, M.J., et al. 2013. Bayesian Analysis of Biogeography when the Number of Areas is Large. *Systematic Biology.* Accepted.

Nosil, P. 2012. *Ecological Speciation.* Oxford University Press, Oxford, United Kingdom.

Rabosky, D.L., et al. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*. 4: 1958.

Takahashi, T., et al. 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution.* 80: 137–144.

TMCS Network. 2012. What do we need to know about speciation? *Trends in Ecology & Evolution*. 27(1): 27–39.

Weber, W.A. 2003. The middle Asian element in the southern Rocky Mountain flora of the western United States: a critical biogeographical review. *Journal of Biogeography*. 30(5): 649–685.

APPENDIX A

Chapter 2 Supplementary Material

Table A.1 Sampled *Carex* taxa, with voucher information, country of origin, and GenBank accession numbers for nuclear and chloroplast loci. Taxon names for vouchers used in next-generation sequencing are shown in bold. Five specimens were obtained after Sanger data collection efforts and were included only in the RADseq dataset. Herbarium acronyms follow Index Herbariorum (Thiers, B. [continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. http://sweetgum.nybg.org/ih/).

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
Carex section *Racemosae* G. Don								
***Carex aboriginum* M.E. Jones**	MICH	1359664	5/28/99	United States (ID)	R. Bjork (4263)	KT021126	KT021021	KT021448
***Carex adelostoma* V.I. Krecz.**	DAO	761392	8/4/92	Canada (Quebec)	M. Blondeau (TQ92265)	KT021124	KT021019	KT021447
Carex albonigra Mack.	MICH	NA	8/3/11	United States (CO)	R. Massatti (9105)	KT021127	KT021022	KT021446
***Carex albonigra* Mack.**	MICH	NA	7/31/11	United States (CO)	R. Massatti (9086)	KT021128	KT021023	KT021445
***Carex aterrima* Hoppe**	MICH	NA	7/7/88	Russia	D. F. Murray (9950)	KT021129	KT021024	KT021444
Carex aterrima Hoppe	MO	4937738	8/10/97	Georgia	R. Gagnidze (2237)	KT021130	KT021025	KT021443
***Carex atrata* L.**	MICH	NA	8/11/01	Norway	P. Zika (16447)	KT021131	KT021026	KT021442
Carex atrata L.	MO	5112118	7/19/74	Sweden	E. Evers (s.n.)	KT021132	KT021027	
***Carex atrata* L.**	A	248871	7/11/84	China (Yunnan)	Sino-Amer. Bot. Exped. (1046)			
Carex atratiformis Britton	MICH	1362493	7/17/96	Canada (New Found.)	M. J. Oldham (18982)	KT021133	KT021029	KT021441
***Carex atratiformis* Britton**	MICH	1362457	8/12/98	United States (AK)	M. Duffy (98-401)	KT021134	KT021030	KT021440
Carex atrosquama Mack.	MICH	NA	8/6/11	United States (CO)	R. Massatti (9113)	KT021136	KT021032	

Table A.1 Continued

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
Carex atrosquama Mack.	MICH	1398996	6/25/70	Canada (Yukon)	M. G. Duman (70-145)	KT021137	KT021033	KT021438
***Carex atrosquama* Mack.**	MICH	1399013	8/18/00	United States (MT)	P. Lesica (8191)			
***Carex augustinowiczii* Meinsh. ex Korsh.**	MICH	NA	6/26/54	Russia	Herb. Fl. URSS (7166)	KT021138	KT021035	KT021437
Carex bella L.H. Bailey	MICH	NA	7/17/11	United States (CO)	R. Massatti (9055)	KT021140	KT021037	KT021436
Carex bella L.H. Bailey	MICH	NA	7/13/11	United States (CO)	R. Massatti (9034)	KT021141	KT021038	KT021435
***Carex bella* L.H. Bailey**	MICH	1399064	7/24/86	United States (AZ)	R. Tallent (402)	KT021142	KT021039	KT021434
Carex buxbaumii Wahlenb.	MICH	1364940	7/11/93	United States (ME)	A. A. Reznicek (9651)	KT021145	KT021041	KT021431
***Carex buxbaumii* Wahlenb.**	MICH	NA	7/2/98	Finland	H. Kamarainen (1998-49)	KT021146	KT021042	KT021430
***Carex caucasica* Steven**	MO	4937735	8/10/97	Georgia	R. Gagnidze (2241)	KT021147	KT021043	KT021429
Carex caucasica Steven	MO	4056429	7/17/87	Russia	A. K. Skvortsov (s.n.)	KT021148	KT021044	
Carex chalciolepis Holm	MICH	NA	8/14/11	United States (WY)	R. Massatti (9119)	KT021149	KT021045	KT021428
Carex chalciolepis Holm	MICH	NA	7/13/11	United States (CO)	R. Massatti (9038)	KT021150	KT021046	KT021427
***Carex chalciolepis* Holm**	MICH	1398953	8/8/86	United States (CO)	R. Tallent (490)	KT021151	KT021047	KT021426
***Carex curvicollis* Franch. & Sav.**	MICH	NA	6/26/02	Japan	K. Yonekura (8721)	KT021153	KT021049	KT021424
Carex epapillosa Mack.	MICH	1399090	8/21/04	United States (CA)	L. P. Janeway (8294)	KT021154	KT021050	KT021423
***Carex epapillosa* Mack.**	MICH	NA	8/14/11	United States (WY)	R. Massatti (9123)	KT021155	KT021051	KT021422
Carex epapillosa Mack.	MICH	1369737	8/2/91	United States (MT)	P. Lesica (5566)	KT021156	KT021052	KT021421

Table A.1 Continued

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
***Carex gmelinii* Hook. & Arn.**	MICH	1371412	6/3/97	United States (AK)	P. Zika (13115)	KT021157	KT021053	KT021420
***Carex hallii* Olney**	MICH	1372735	6/6/64	United States (WY)	Johnson (412)			
Carex hallii Olney	RM	758356	7/10/00	United States (CO)	R. L. Hartman (67856)	KT021158	KT021054	KT021419
***Carex hancockiana* Maxim.**	MICH	NA	6/26/88	Russia	D. F. Murray (9806b)	KT021159	KT021055	KT021418
***Carex hancockiana* Maxim.**	MO	4496057	7/1/93	China	Q. R. Wu (835)			
***Carex hartmannii* Cajander**	COLO	249147	7/18/66	Sweden	C. M. Norrman (s.n.)	KT021123	KT021018	KT021417
***Carex helleri* Mack.**	MICH	1372895	8/2/87	United States (CA)	R. Tallent (849)	KT021161	KT021057	KT021415
Carex heteroneura W. Boott	MICH	1399086	7/30/04	United States (CA)	L. P. Janeway (8211)	KT021162	KT021058	KT021414
***Carex heteroneura* W. Boott**	MICH	1398989	7/31/04	United States (NV)	R. F. C. Naczi (10651)	KT021163	KT021059	KT021413
***Carex holmgreniorum* Reznicek & D.F. Murray**	MICH	NA	7/4/08	United States (UT)	A. A. Reznicek (11937)	KT021198	KT021034	KT021449
***Carex holostoma* Drejer**	MICH	1373187	8/10/87	Greenland	B. Predskild (87-1242)	KT021164	KT021060	KT021412
***Carex idahoa* L.H. Bailey**	MICH	1373774	7/12/93	United States (MT)	P. Lesica (6088)	KT021165	KT021061	KT021411
Carex infuscata Nees	A	309410	7/9/05	China (Sichuan)	D. E. Boufford (32614)	KT021166	KT021062	KT021410
***Carex infuscata* Nees**	MO	5108536	6/21/62	Afghanistan	I. Hedge (4372)	KT021167	KT021063	
***Carex kansuensis* Nelmes**	GH	249417	8/6/93	China (Qinghai)	T. N. Ho (858)	KT021168	KT021064	KT021409
Carex kansuensis Nelmes	MO	4647489	7/30/93	China (Qinghai)	T. N. Ho (551)	KT021169	KT021065	KT021408
***Carex lehmannii* Drejer**	MO	6296541	8/10/07	China (Sichuan)	D. E. Boufford (39446)	KT021170	KT021066	KT021407
***Carex malmei* Kalela**	SI	NA	2/11/11	Argentina	Zuloaga (12789)	KT021125	KT021020	KT021400

Table A.1 Continued

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
***Carex media* R. Br. ex Richardson**	MICH	1398895	6/22/99	United States (AK)	P. Zika (13827)	KT021177	KT021073	KT021399
Carex media R. Br. ex Richardson	MO	4622484	7/2/89	Russia	Unknown (s.n.)	KT021178	KT021074	KT021398
***Carex melanantha* C.A. Mey.**	MICH	NA	6/28/99	Krygystan	L. R. Phillippe (30798)	KT021179	KT021075	KT021397
***Carex melananthiformis* Litv.**	COLO	444483	7/2/82	Russia	Unknown (s.n.)	KT021122	KT021017	KT021396
***Carex melanocephala* Turcz.**	MICH	NA	6/20/99	Krygystan	L. R. Phillippe (30664)	KT021180	KT021076	KT021395
***Carex mertensii* J.D. Prescott ex Bong.**	MICH	1378626	7/29/05	Canada (British Columbia)	P. M. Peterson (18752)	KT021181	KT021077	KT021394
***Carex meyeriana* Kunth**	MICH	NA	7/14/79	Russia (Siberia)	H. H. Iltis (859)	KT021182	KT021078	
Carex meyeriana Kunth	MO	4512400	6/16/90	China	X. Lin (61604)	KT021183	KT021079	
***Carex moorcroftii* Falc. ex Boott**	MICH	NA	8/1/99	Tibet	D. F. Murray (13027)	KT021186	KT021082	KT021391
Carex nelsonii Mack.	MICH	NA	8/14/11	United States (WY)	R. Massatti (9124)	KT021187	KT021083	KT021390
***Carex nelsonii* Mack.**	MICH	NA	8/7/11	United States (CO)	R. Massatti (9117)	KT021188	KT021084	KT021389
Carex nelsonii Mack.	MICH	NA	7/13/11	United States (CO)	R. Massatti (9037)	KT021189	KT021085	KT021388
Carex nelsonii Mack.	MICH	1398801	8/8/86	United States (CO)	R. Tallent (498)	KT021190	KT021086	KT021387
***Carex norvegica* Retz.**	MICH	1380837	8/26/00	Canada (Ontario)	M. J. Oldham (24532)	KT021191	KT021087	KT021386
Carex nova L.H. Bailey	MICH	NA	7/20/11	United States (CO)	R. Massatti (9065)	KT021192	KT021088	KT021385
***Carex nova* L.H. Bailey**	MICH	NA	7/13/11	United States (CO)	R. Massatti (9036)	KT021193	KT021089	KT021384
Carex nova L.H. Bailey	MICH	1398827	8/10/86	United States (CO)	R. Tallent (548)	KT021194	KT021090	KT021383
***Carex obscura* Nees**	A	307351	8/15/07	China (Sichuan)	D. E. Boufford (39774)	KT021195	KT021091	KT021382

Table A.1 Continued

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
Carex obscura Nees	MO	6296543	7/11/04	China (Sichuan)	D. E. Boufford (30513)	KT021196	KT021092	KT021381
***Carex parryana* Dewey**	MICH	1398914	7/19/90	United States (MT)	P. Lesica (5184)	KT021199	KT021094	KT021379
Carex parryana Dewey	MICH	1398913	7/18/97	Canada (Ontario)	M. J. Oldham (20222)	KT021200	KT021095	KT021378
***Carex parviflora* Host**	DAO	25637	7/2/51	Russia	Unknown (s.n.)	KT021121	KT021016	KT021377
Carex parviflora Host	MICH	NA	8/28/62	Switzerland	A. Broecker (s.n.)	KT021201	KT021096	
***Carex pelocarpa* F.J. Herm.**	MICH	NA	8/5/11	United States (CO)	R. Massatti (9111)	KT021203	KT021098	KT021375
Carex pelocarpa F.J. Herm.	MICH	NA	7/31/11	United States (CO)	R. Massatti (9089)	KT021204	KT021099	KT021374
Carex pelocarpa F.J. Herm.	MICH	1398817	7/31/04	United States (NV)	R. F. C. Naczi (10660)	KT021205	KT021100	KT021373
***Carex polymascula* P.C. Li**	A	309944	7/18/05	China (Sichuan)	D. E. Boufford (33235)	KT021208	KT021103	KT021371
***Carex raynoldsii* Dewey**	MICH	1384614	7/22/04	United States (CA)	R. F. C. Naczi (10484)	KT021210	KT021105	KT021369
***Carex sabulosa* Turcz. ex Kunth**	MICH	1204205	6/13/80	Canada (Yukon)	W. J. Cody (25606)	KT021211	KT021106	KT021368
Carex sabulosa Turcz. ex Kunth	MO	4984386	7/24/84	Russia	Grubov & Ivanina, Herb. Fl. URSS (6557)	KT021212	KT021107	
Carex schneideri Nelmes	A	249833	6/24/84	China (Yunnan)	Sino-Amer. Bot. Exped. (430)	KT021213	KT021108	KT021367
***Carex schneideri* Nelmes**	A	307329	8/7/07	China (Sichuan)	D. E. Boufford (39239)	KT021214	KT021109	KT021366
***Carex serratodens* W. Boott**	MICH	1386711	5/9/02	United States (CA)	D. W. Taylor (18037)	KT021217	KT021111	KT021364
***Carex specuicola* J.T. Howell**	NAVA	NA	7/12/06	United States (UT)	Roth (1869)	KT021219	KT021113	KT021362

Table A.1 Continued

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
Carex specuicola J.T. Howell	MICH	NA	8/19/11	United States (AZ)	Licher (3210)	KT021220	KT021114	KT021361
***Carex stevenii* (Holm) Kalela**	MICH	1398853	7/14/84	United States (WY)	R. S. Kirkpatrick (3923)	KT021221	KT021115	KT021360
***Carex stylosa* C.A. Mey.**	MICH	1388536	8/12/06	United States (AK)	S. S. Talbot (LKI011-15)	KT021222	KT021116	KT021359
Carex stylosa C.A. Mey.	MICH	1388535	7/17/96	Canada (New Found.)	M. J. Oldham (18963)	KT021223	KT021117	KT021358
Carex stylosa C.A. Mey.	MICH	NA	7/28/93	Russia	C. L. Parker (4610)	KT021224	KT021118	KT021357
Carex section *Aulocystis* Dumort.								
***Carex atrofusca* Schkuhr**	MICH	1362501	7/14/05	Canada (NWT)	P. M. Peterson (18581)	KT021135	KT021031	KT021439
Carex luzulina Olney	MICH	1377890	8/9/03	United States (CA)	L. P. Janeway (7956)	KT021173	KT021069	KT021404
Carex section *Bicolores* (Tuck. ex L.H. Bailey) Rouy								
Carex aurea Nutt.	MICH	1362563	6/5/91	Canada (Ontario)	M. J. Oldham (12663)	KT021139	KT021036	
***Carex bicolor* Bellardi ex All.**	MICH	1363370	8/4/06	Canada (Yukon)	B. A. Bennett (06-516)	KT021143	KT021040	KT021433
Carex hassei L.H. Bailey	MICH	1372767	7/10/01	United States (CA)	B. Wilson (10542)	KT021160	KT021056	KT021416
***Carex hassei* L.H. Bailey**	MICH	NA	7/2/08	United States (UT)	A. A. Reznicek (11931)			
Carex section *Chlorostachyae* Tuck. ex Meinsh.								
***Carex williamsii* Britton**	MICH	1201565	7/7/01	Canada (Ontario)	M. J. Oldham (26147)	KT021226	KT021120	KT021355
Carex section *Griseae* (L.H. Bailey) Kuk.								
Carex ouachitana Kral, Manhart & Bryson	MICH	1381553	4/27/02	United States (KY)	R. F. C. Naczi (9241)	KT021197	KT021093	KT021380

Table A.1 Continued

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
Carex section *Limosae* (Heuff.) Meinsh.								
Carex limosa L.	MICH	1376679	6/13/96	United States (WI)	A. A. Reznicek (10235)	KT021171	KT021067	KT021406
Carex macrochaeta C.A. Mey.	MICH	1378081	8/13/06	United States (AK)	S. S. Talbot (BUL023-02)	KT021174	KT021070	KT021403
Carex macrochaeta C.A. Mey.	RM	318882	8/20/79	United States (AK)	E. Neese (8394)	KT021175	KT021071	KT021402
Carex magellanica Lam.	MICH	1201305	7/1/00	United States (NY)	A. A. Reznicek (11158)	KT021176	KT021072	KT021401
Carex section *Paniceae* G. Don								
Carex livida (Wahlenb.) Willd.	MICH	1282624	7/11/03	Canada (Ontario)	M. J. Oldham (28968)	KT021172	KT021068	KT021405
***Carex vaginata* Tausch**	MICH	1391220	6/29/96	Canada (Ontario)	M. J. Oldham (18782)	KT021225	KT021119	KT021356
Carex section *Phacocystis* Dumort.								
Carex lugens Holm	MICH	1201267	8/4/03	United States (AK)	A. A. Reznicek (11536)	KT021144	KT021028	KT021432
***Carex scopulorum* Holm**	MICH	1386578	9/6/03	United States (CO)	C. Lea (3439)	KT021216		KT021365
Carex section *Scitae* Kuk.								
Carex coriophora Fisch. & C.A. Mey. ex Kunth	MICH	NA	6/29/64	Russia	Bobrov, Herb. Fl. URSS (5004)	KT021152	KT021048	KT021425
Carex microchaeta Holm	MICH	1378756	7/19/94	United States (AK)	R. Lipkin (893)	KT021184	KT021080	KT021393
Carex microchaeta Holm	MICH	1378755	7/28/05	Canada (Yukon)	B. A. Bennett (05-0795)	KT021185	KT021081	KT021392
***Carex paysonis* Clokey**	MICH	1398939	8/12/03	United States (MT)	P. Lesica (8767)	KT021202	KT021097	KT021376

Table A.1 Continued

Taxon	Herbarium	Herbarium Accession #	Date collected	Country (state/ province)	Collector (Collection #)	ITS	ETS	matK
Carex podocarpa R. Br. ex Richardson	MICH	1383472	7/10/05	Canada (Yukon)	B. A. Bennett (05-0391)	KT021206	KT021101	KT021372
Carex podocarpa R. Br. ex Richardson	MO	4984483	7/27/64	Russia	Grubov & Ivanina, Herb. Fl. URSS (6608)	KT021207	KT021102	
Carex rariflora (Wahlenb.) Sm.	MICH	NA	8/5/79	Russia	Grubov & Ivanina, Herb. Fl. URSS (6310)	KT021209	KT021104	KT021370
Carex scita Maxim.	MO	4021644	4/19/90	Japan	S. Tsugaru (13097)	KT021215	KT021110	
Carex spectabilis Dewey	MICH	1387201	7/7/03	United States (CA)	L. P. Janeway (7890)	KT021218	KT021112	KT021363

Table A.2 Processing information and pyRAD summary statistics for species sequenced on the Illumina platform. Raw reads refers to the total reads produced during Illumina sequencing, while post-processing reads are those that remained after filtering for adaptor contamination, quality, and ambiguous barcodes. The post-processing reads were utilized by pyRAD to create clusters of homologous sequences (Total clusters, Mean depth of clusters). After the heterozygosity (H) and error-rate (E) were estimated across clusters, consensus sequences were created for each cluster; those that passed the pyRAD filtering parameters were retained (Loci). Variable and invariable DNA sites were summed across all loci (Total sites), and the percentage of polymorphic sites (Percent poly) is reported. Consensus sequences were clustered across species, and loci that passed filtering parameters were included in the final data matrix (Final loci). Species that were excluded from analyses are starred (see text).

Species	Raw reads	Post-processing reads	Percent retained	Total clusters	Mean depth of clusters	H	E	Loci	Total sites	Percent poly	Final loci
C. aboriginum	1742636	1543026	88.5	27338	47.6	0.0077	0.0008	24720	3335201	0.08	13725
C. adelostoma	138966	107088	77.1	4214	13.0	0.0155	0.0025	3234	436159	0.46	1564
C. albonigra	2246117	1999389	89.0	25167	63.2	0.0142	0.0006	21384	2884925	0.40	14784
C. atrata	877334	763700	87.0	16455	37.1	0.0095	0.0013	14403	1942980	0.12	9175
*C. aterrima**	28699	21151	73.7	333	18.0	0.0256	0.0041	212	28593	0.12	183
*C. atrata**	16284	13565	83.3	203	17.8	0.0130	0.0014	10731	1447660	0.10	125
C. atratiformis	3000145	2701088	90.0	39270	60.9	0.0080	0.0006	35562	4798366	0.09	23541
C. atrofusca	2072743	1855382	89.5	19210	71.0	0.0158	0.0010	15429	2081546	0.30	9167
*C. atrosquama**	1762146	1577753	89.5	20851	58.2	0.0165	0.0010	16614	2241353	0.34	9809
*C. augustinowiczii**	33307	23220	69.7	227	55.6	0.0131	0.0011	184	24824	0.05	149
C. bella	2564877	2221773	86.6	37181	51.9	0.0072	0.0006	33908	4575031	0.10	22703
C. bicolor	5525190	4926895	89.2	36058	93.3	0.0119	0.0007	30955	4176655	0.10	8943
C. buxbaumii	501690	426935	85.1	10712	29.7	0.0154	0.0013	8332	1124099	0.37	3893
*C. caucasica**	92527	61763	66.8	1385	15.3	0.0196	0.0032	995	134143	0.12	824
C. chalciolepis	2512641	2194906	87.4	33660	54.3	0.0080	0.0007	30355	4095368	0.14	21466
C. curvicollis	2282178	2089820	91.6	29423	57.0	0.0091	0.0006	25923	3497772	0.06	13601
C. epapillosa	2864948	2593687	90.5	34934	64.9	0.0080	0.0006	31365	4232010	0.09	22001

Table A.2 Continued

Species	Raw reads	Post-processing reads	Percent retained	Total clusters	Mean depth of clusters	H	E	Loci	Total sites	Percent poly	Final loci
C. gmelinii	2083327	1851484	88.9	23672	66.7	0.0094	0.0007	20965	2828296	0.11	13153
*C. hallii**	7764	3660	47.1	42	10.6	0.0163	0.0048	29	3909	0.26	22
C. hancockiana	844833	733533	86.8	18892	31.8	0.0089	0.0013	16721	2255620	0.13	10815
*C. hancockiana**	11496	4477	38.9	73	13.0	0.0053	0.0014	66	8905	0.02	59
C. hartmanii	697694	575397	82.5	13923	31.2	0.0150	0.0012	11463	1546172	0.47	5709
C. hassei	1856304	1670514	90.0	15844	68.8	0.0178	0.0011	12548	1692874	0.11	5815
C. helleri	1034681	920288	88.9	17210	42.8	0.0085	0.0010	15256	2058198	0.11	9765
C. heteroneura	2325647	2052299	88.2	32130	53.9	0.0079	0.0008	28798	3885202	0.10	19718
C. holmgreniorum	3266924	2901112	88.8	36718	69.0	0.0077	0.0005	33037	4458045	0.10	18256
C. holostoma	1033051	869602	84.2	20799	35.7	0.0071	0.0011	18894	2548813	0.10	9359
C. idahoa	2778168	2518694	90.7	38135	54.4	0.0100	0.0006	33902	4574603	0.27	18253
*C. infuscata**	33095	19876	60.1	222	11.7	0.0208	0.0059	147	19797	0.17	109
C. kansuensis	1421914	1245589	87.6	29105	36.4	0.0100	0.0008	25786	3478921	0.25	16223
C. lehmannii	1692298	1460379	86.3	33414	38.2	0.0068	0.0006	30538	4120032	0.10	19867
C. malmei	1301754	1180094	90.7	14470	68.2	0.0073	0.0005	13188	1779175	0.10	7505
C. media	841376	746638	88.7	14294	43.8	0.0081	0.0009	12792	1725904	0.10	8627
C. melanantha	635725	545124	85.7	11533	38.0	0.0103	0.0008	10102	1362981	0.25	5747
C. melananthiformis	1020891	895654	87.7	22572	31.8	0.0159	0.0014	17429	2350676	0.47	9834
C. melanocephala	1814678	1623727	89.5	28129	50.8	0.0070	0.0006	25718	3469882	0.09	16768
C. mertensii	3513366	3124699	88.9	35903	70.1	0.0093	0.0006	31540	4255496	0.07	17092
*C. meyeriana**	102225	67108	65.6	1574	15.1	0.0210	0.0033	1130	152331	0.28	839
C. moorcroftii	1774701	1503973	84.7	27844	47.5	0.0106	0.0006	24806	3346887	0.34	13816
C. nelsonii	4980914	4468185	89.7	37619	103.3	0.0073	0.0004	34210	4616020	0.09	22882
C. norvegica	2023702	1779662	87.9	29191	53.8	0.0074	0.0004	26583	3586358	0.09	17785
C. nova	6307952	5746439	91.1	49313	102.8	0.0097	0.0005	42722	5766499	0.09	27988
C. obscura	494064	420812	85.2	12312	25.8	0.0111	0.0010	10679	1440643	0.31	6132

Table A.2 Continued

Species	Raw reads	Post-processing reads	Percent retained	Total clusters	Mean depth of clusters	H	E	Loci	Total sites	Percent poly	Final loci
C. parryana	387530	333514	86.1	7886	32.8	0.0081	0.0009	7016	946572	0.14	4056
*C. parviflora**	56160	44396	79.1	1352	14.3	0.0141	0.0028	1063	143350	0.11	889
C. paysonis	3816086	3436303	90.0	22446	114.8	0.0173	0.0005	17947	2421371	0.34	10656
C. pelocarpa	2343242	2069481	88.3	29757	57.5	0.0076	0.0005	26909	3630674	0.09	18965
C. polymascula	3070534	2735307	89.1	42463	57.8	0.0082	0.0007	37517	5063020	0.09	23837
C. raynoldsii	831714	755980	90.9	13183	46.9	0.0092	0.0010	11545	1557629	0.11	7436
C. sabulosa	808652	652260	80.7	24606	21.3	0.0096	0.0009	21991	2966853	0.30	11960
C. schneideri	1749399	1616882	92.4	34509	41.9	0.0069	0.0007	31110	4197762	0.09	20356
C. scopulorum	1077146	969696	90.0	10616	69.7	0.0163	0.0012	8472	1143218	0.30	5141
C. serratodens	1253384	1122834	89.6	21530	45.1	0.0072	0.0006	19630	2648258	0.08	11309
C. specuicola	588283	525254	89.3	10079	42.8	0.0081	0.0008	9024	1217571	0.13	5244
C. stevenii	339728	282316	83.1	7612	28.0	0.0092	0.0014	6636	895353	0.11	4261
C. stylosa	746217	660991	88.6	10350	46.8	0.0118	0.0011	8706	1174431	0.10	5484
C. vaginata	1328591	1187011	89.3	12549	64.8	0.0228	0.0032	8729	1177953	0.31	3738
C. williamsii	1925348	1792026	93.1	15358	83.1	0.0128	0.0008	13012	1755578	0.08	6582

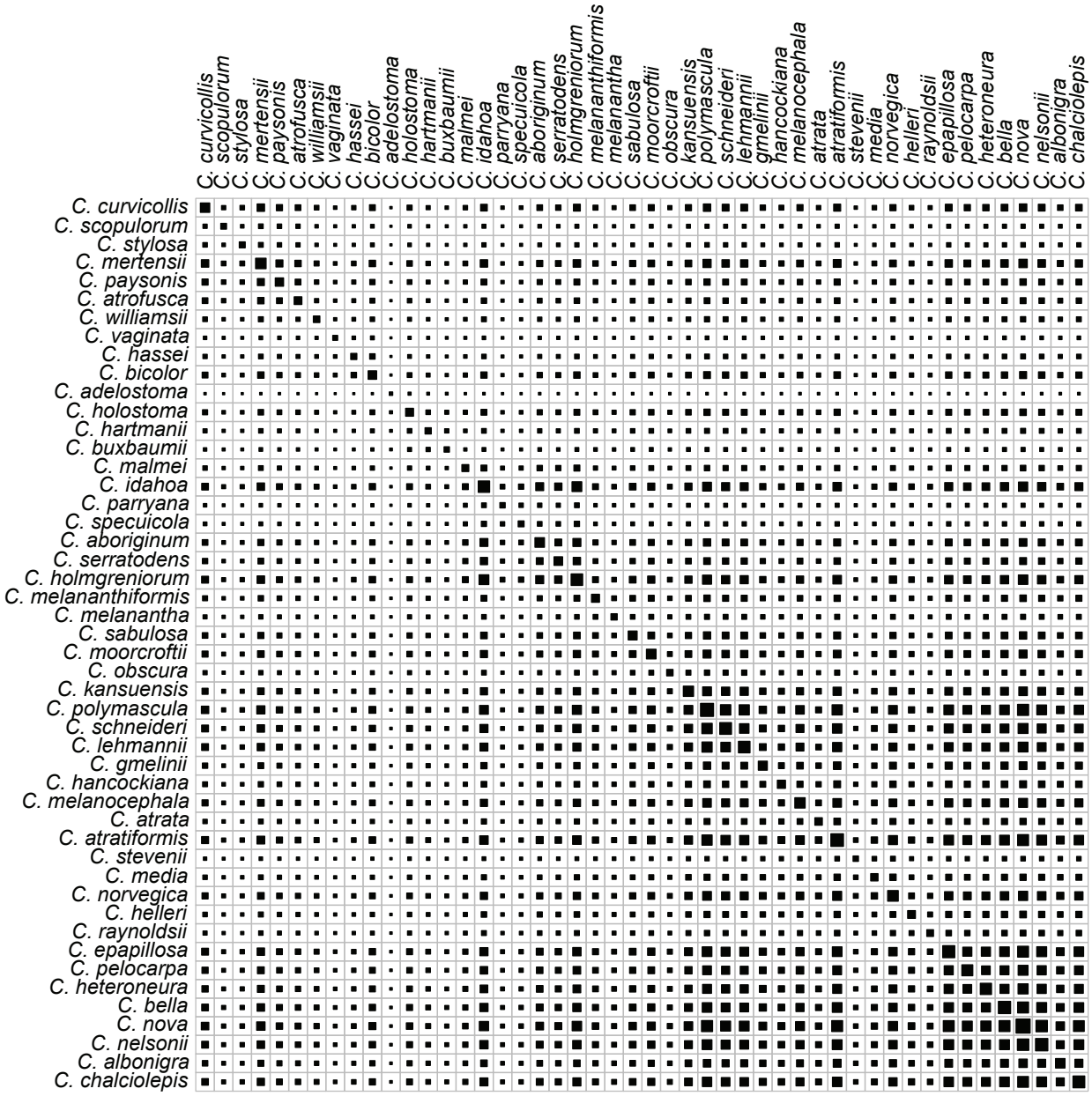

Figure A.1 Loci shared among individuals, where the size of the square correlates to the proportion shared (from 0 to 1) either between individuals (off-diagonal cells) or successfully amplified within an individual (diagonal cells). Values range from 1.2% loci shared between *C. adelostoma* and *C. vaginata* to 37.2% loci shared between *C. nova* and *C. nelsonii*.

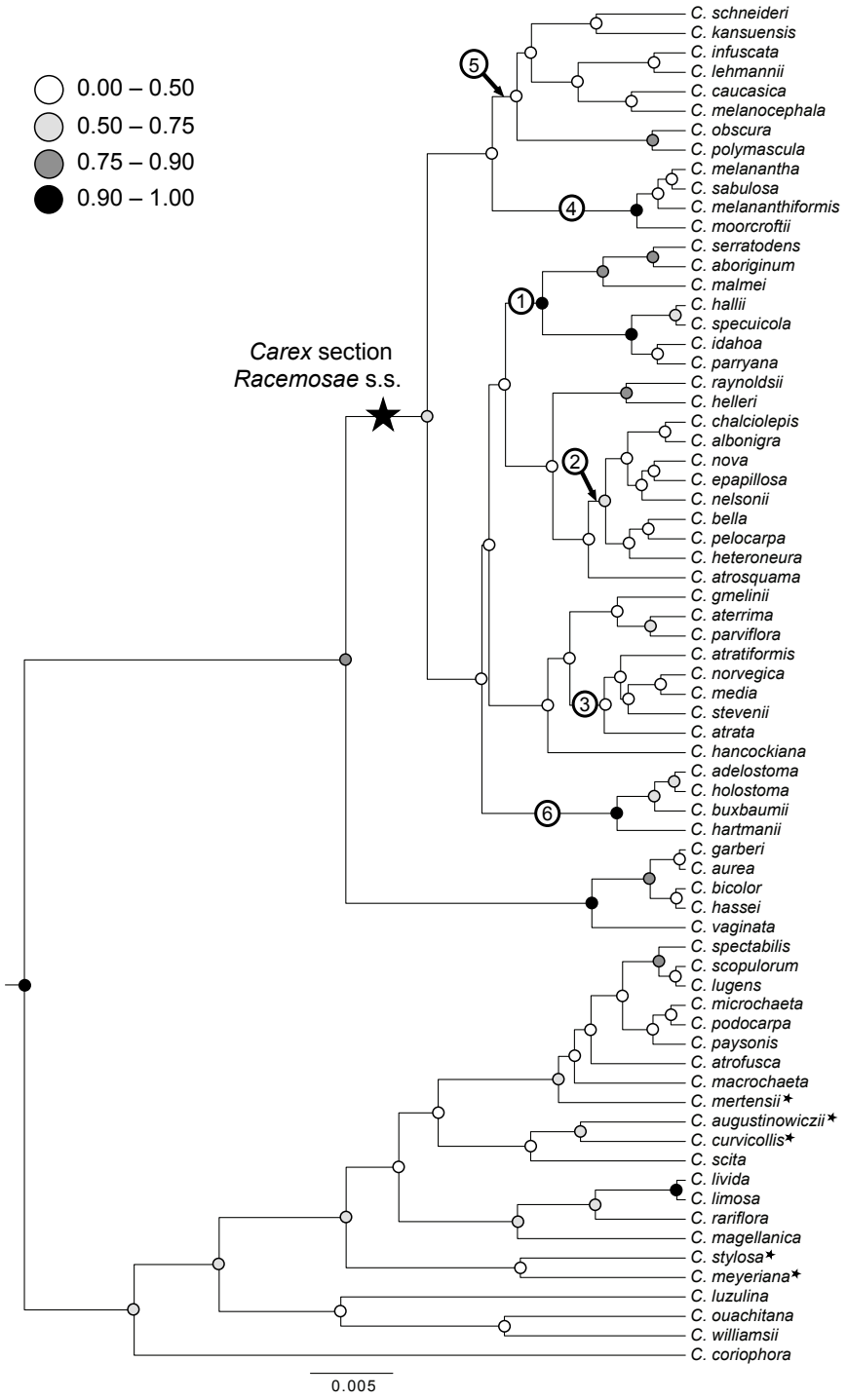

Figure A.2 Maximum clade credibility tree detailing the relationships of *Carex* section *Racemosae* and closely related taxa, estimated using *BEAST and all of the traditional loci and taxa detailed in Table A.1. Shaded circles on nodes represent posterior probabilities estimated from 80,000 post burn-in trees. Numbered circles representing clades discussed in the text are used throughout the figures, and are the same except that here and in Fig. 2.1 Clade 5 contains *C. obscura*. The branch subtending *Carex* section *Racemosae* s.s. is starred, as are the *Carex* section *Racemosae* taxa that cluster with taxa representing other sections.

APPENDIX B

Chapter 4 Supplementary Material

Environmental niche modeling methodology

To avoid overfitting of the distribution models, the geographic extent of the environmental layers was reduced to an area approximately 20% larger than the known distribution of the species (Anderson and Raza 2010). To guard against the inherent difficulties involved in extrapolating distributions into novel climates (reviewed in Alvarado-Serrano and Knowles 2013), an iterative approach was used to generate ENMs for the LGM. Multivariate environmental similarity surfaces (MESS maps) were used to identify which of the 19 bioclimatic variables resulted in areas of low reliability predictions due to the variables being outside of the range present in the present-day environmental data (Elith et al. 2010). MAXENT was rerun excluding these out-of-range variables, and this process of analysis with MESS maps was repeated until no LGM variables were out-of-range compared to present-day bioclimatic variables. Because MESS maps do not indicate changes in the correlations among the environmental variables used for LGM reconstructions (Elith et al. 2010), we checked our LGM ENM using only the most informative variable (Bio5) to ensure we were not reporting errant distributional patterns. Additionally, a present-day ENM was generated for the subset of variables that were not out-of-range and compared to a ENM constructed using all climatic variables with greater than 5% importance (determined by jackknifing) to assess their similarity.

Because our goal was to assess the overall similarity of *C. chalciolepis* and *C. nova's* LGM distributions, and given the difficulties with reconstructing past distributions (Elith and Leathwick 2009), we visualized ENMs by utilizing threshold values based on maximum training sensitivity plus specificity (MTSS). As such, models of past distributions are less likely to be severely constrained, and hence, it is more likely that the habitable area common to *C. chalciolepis* and *C. nova* will be represented in the projected LGM distributions.

RAD library data processing methodology

Potential chloroplast and mitochondrial sequences were filtered from the processed dataset using Bowtie 0.12.8. Because of the lack of sequenced chloroplast and mitochondrial genomes in *Carex*, chloroplast and mitochondrial genomes within Poaceae were used. Specifically, the data were compared to genomes downloaded from GenBank, including *Agrostis*

stolonifera (NC_008591.1), *Oryza nivara* (AP006728.1), and *Zea mays* (X86563.2) for identifying potential chloroplast genes, and *Zea perennis* (DQ645538.1) and *Triticum aestivum* (EU534409.1) for identifying potential mitochondrial genes. Given the relative slow rates of molecular evolution characterizing chloroplast and plant mitochondrial genomes (Wolfe et al. 1987), a tolerance of 2 mismatches (-v 2) between *Carex* sequences and these genomes was used to identify chloroplast and mitochondrial sequences that were removed from the dataset.

Literature Cited

Alvarado-Serrano, D. F., and L. L. Knowles. 2013. Ecological niche models in phylogeographic studies: applications, advances, and precautions. Mol. Ecol. Resour.: In press.

Anderson, R. P., and A. Raza. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. J. Biogeogr. 37:1378–1393.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40:677–697.

Elith, J., M. Kearney, and S. Phillips. 2010. The art of modeling range-shifting species. Method. Ecol. Evol. 1:330–342.

Wolfe, K. H., W. H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84:9054–9058.

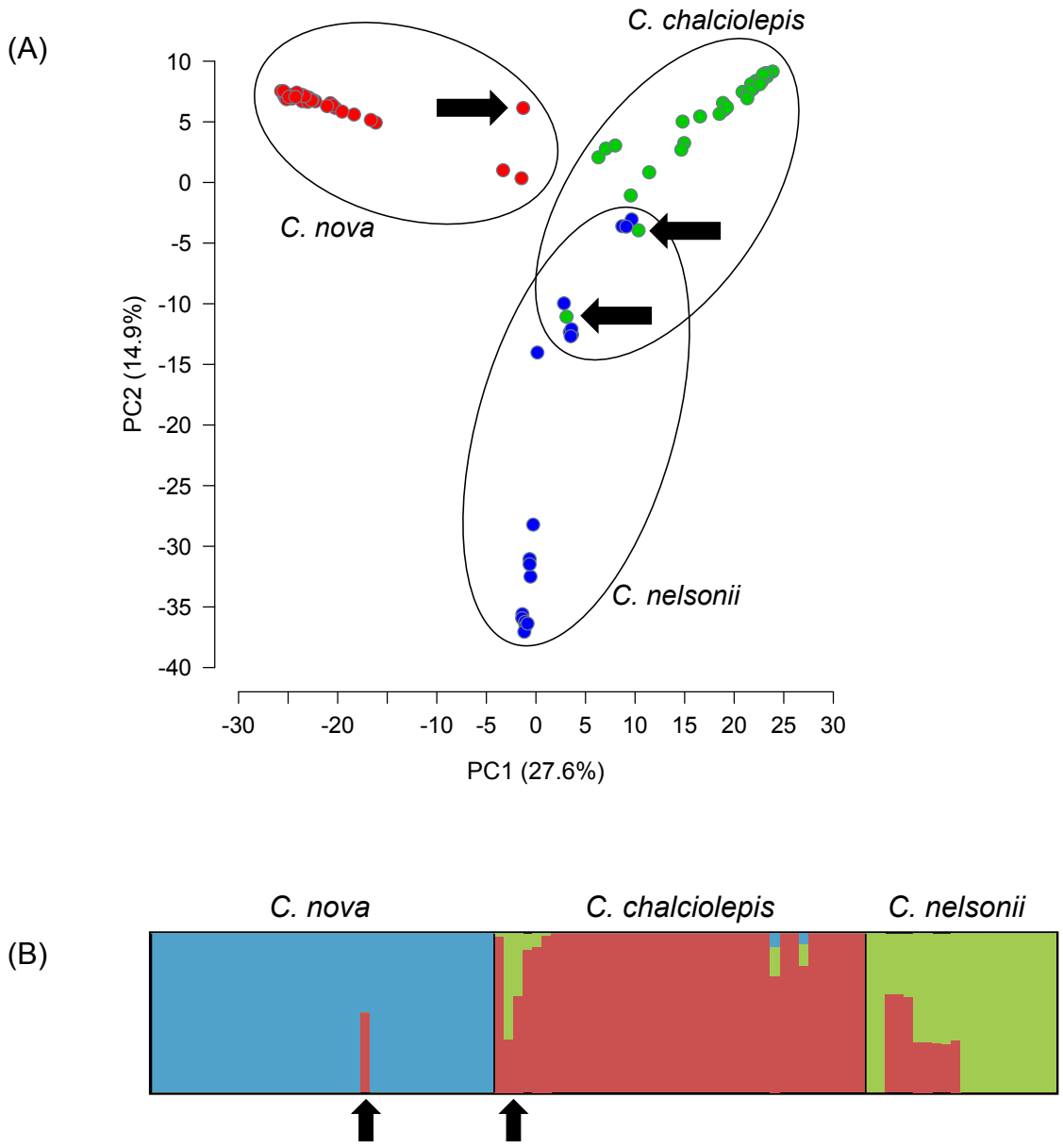

Figure B.1 PCA results (A) and STRUCTURE results (B) identifying three *Carex* individuals with significant contributions of heterospecific genomic material. All three individuals are denoted with black arrows in (A), while the arrow pointing out *C. chalciolepis* individuals in (B) refers to two individuals that are side-by-side.

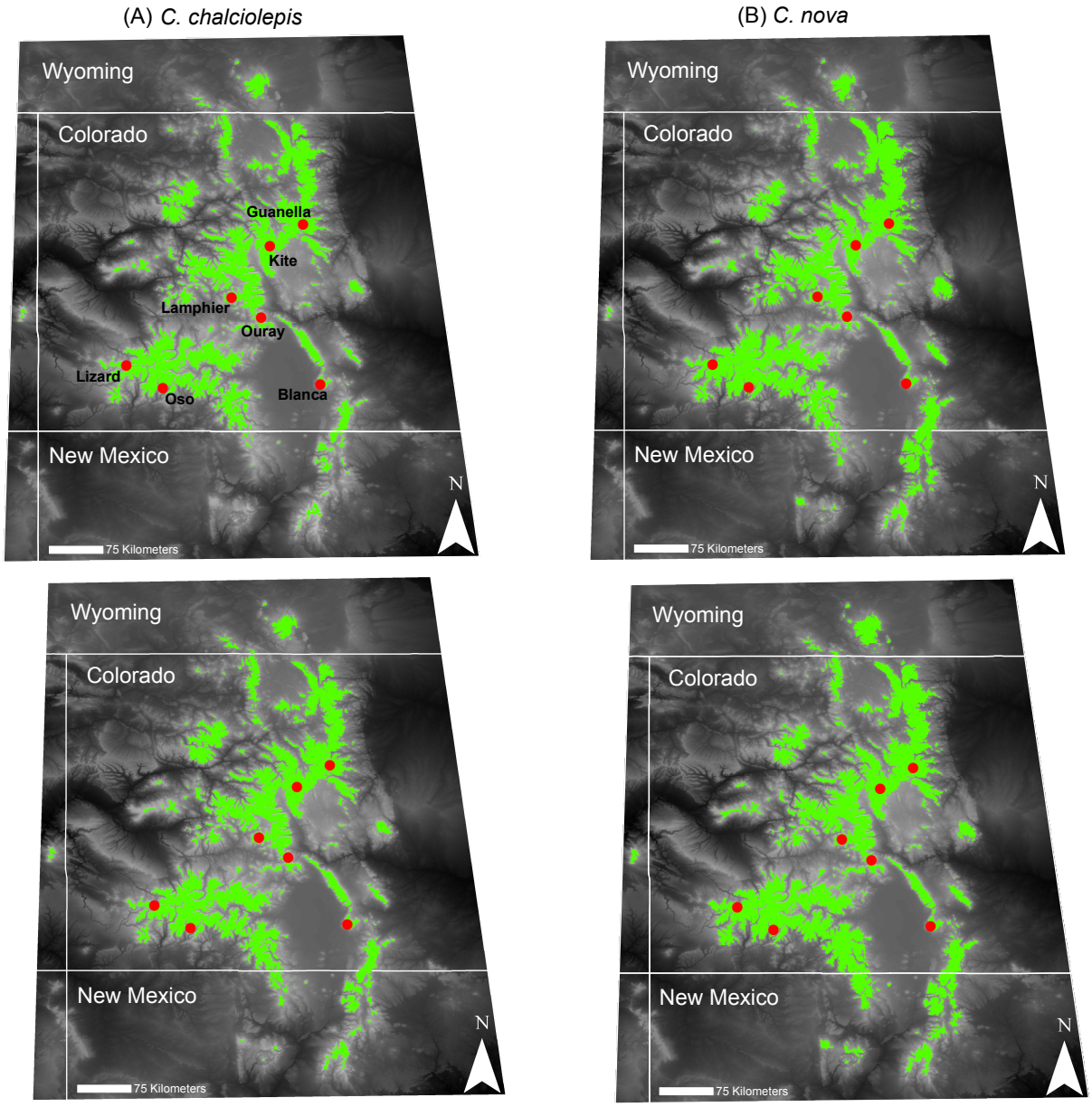

Figure B.2 Present-day ENMs for *C. chalciolepis* (A) and *C. nova* (B). The top row contains ENMs created using all environmental variables with greater than 5% importance, while the bottom row contains models created using only environmental variables that have similar present and LGM ranges (Bio4, Bio5, Bio14, and Bio18). Sampling sites are indicated with red dots.

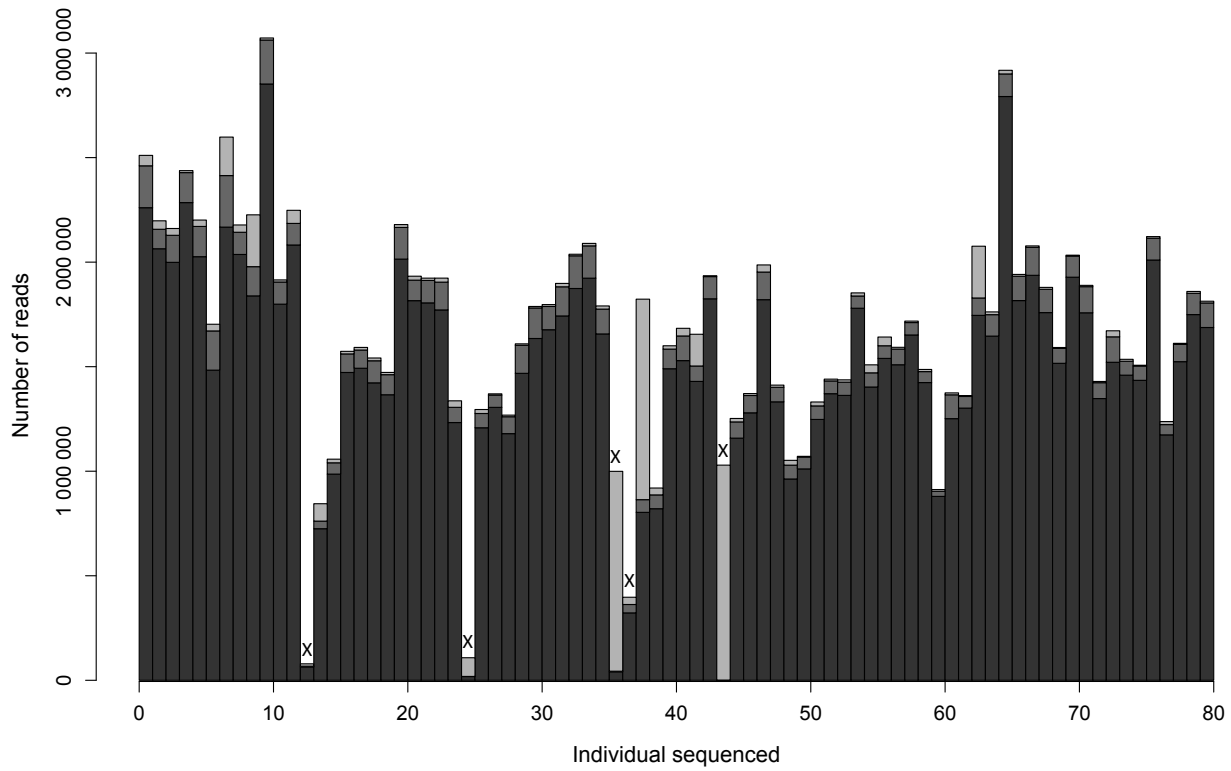

Figure B.3 The number of reads per individual, where individuals 1 through 40 are *C. nova* and individuals 41 through 80 are *C. chalciolepis*. The cumulative stacked bars represent the number of raw reads per individual. Within each individual, the light gray color represents reads discarded due to low quality scores or ambiguous barcodes, the medium gray color represents reads discarded because they aligned with chloroplast or mitochondrial genomes, and the dark gray color represents reads that were available for analyses. Individuals that were removed from all analyses because of insufficient high quality reads are marked with Xs.

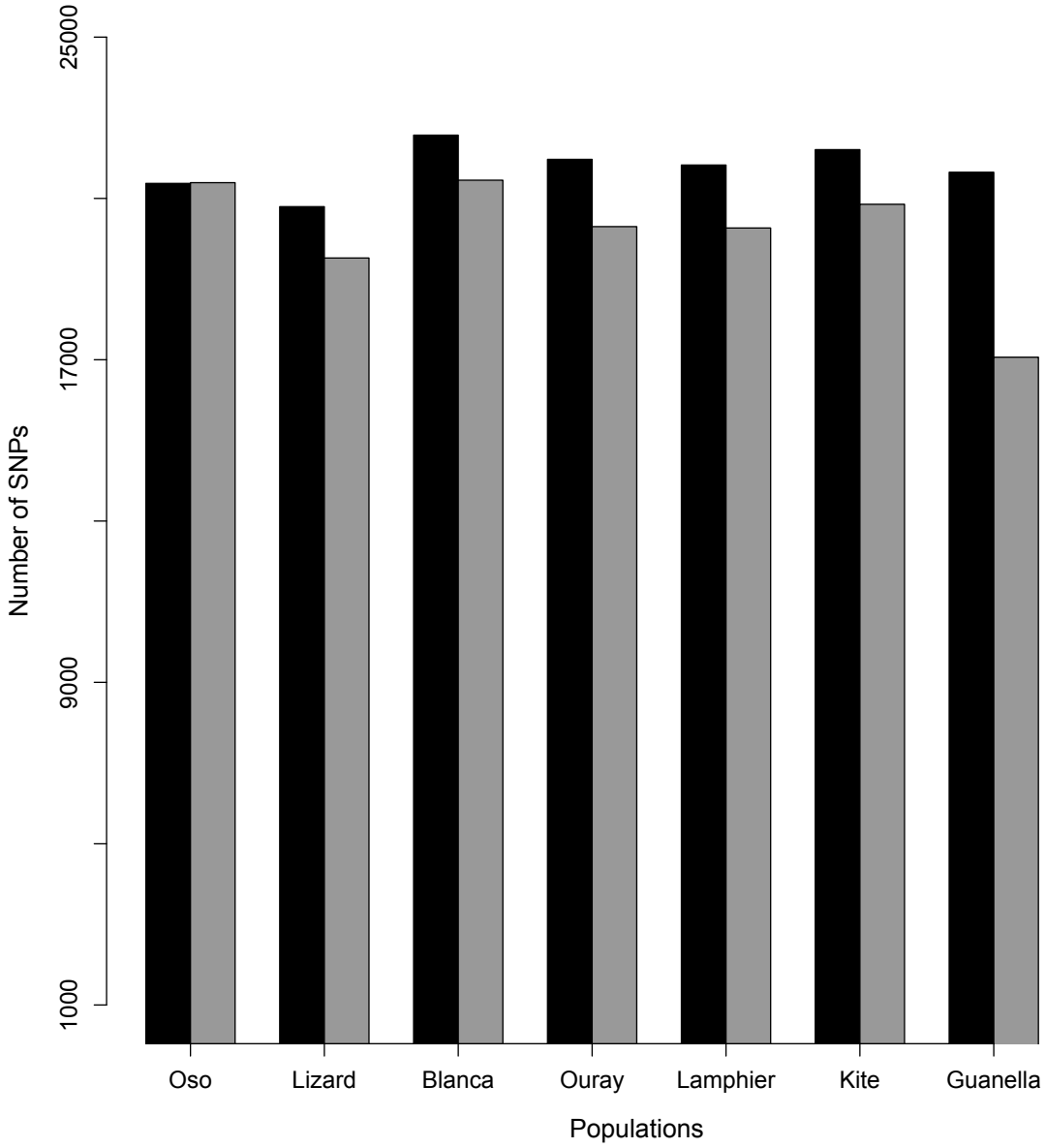

Figure B.4 The number of SNPs present in each population of *C. chalciolepis* (in black) and *C. nova* (in grey). See Table 4.2 for a summary of genomic data collected in each population, including the average number of reads per population and the number of individuals analyzed per population.

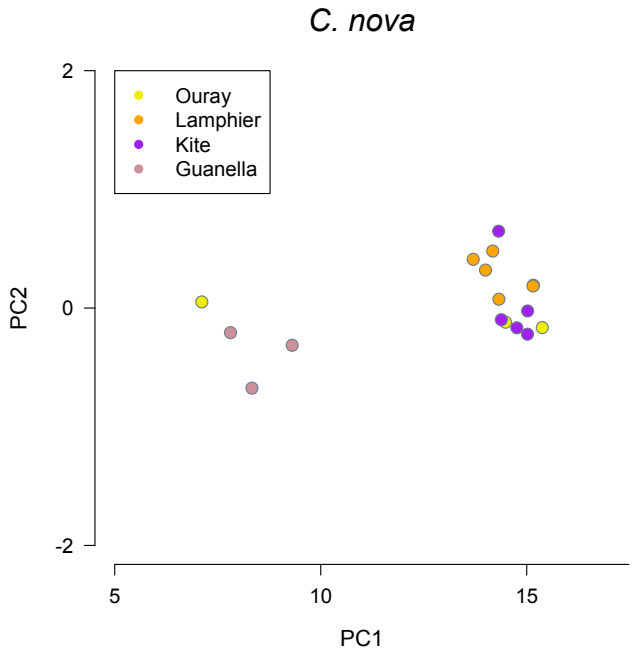

Figure B.5 PCA detail of *C. nova's* Central and North populations (see Fig. 4.5B). The Ouray individual that groups with the Guanella population was identified as having a heterospecific genomic contribution (see Fig. B.1 and Methods).

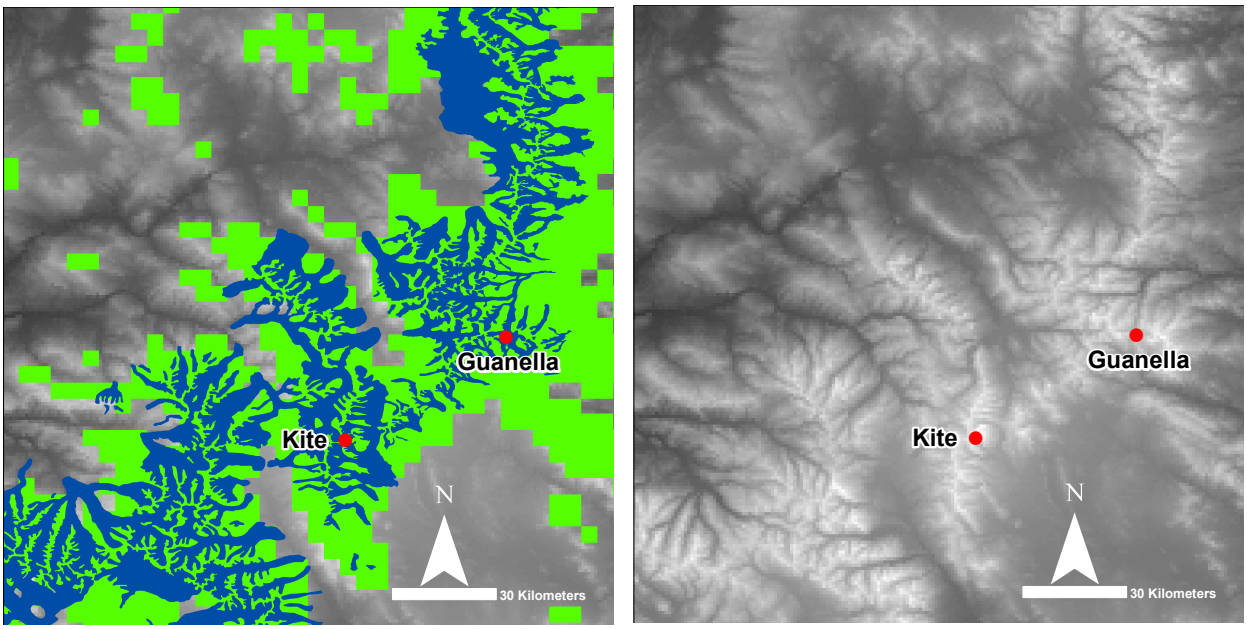

Figure B.6 Detail of the North region (see Fig. 4.2) to illustrate the interaction of topography, glaciers, and predicted habitat. The left illustration shows a detailed reconstruction of Pleistocene glaciers (blue) based on glacial moraines and glacial till (data extracted from http://geosurvey.state.co.us/geology/Pages/GlacialGeology.aspx). The green polygons represent *C. chalciolepis'* LGM ENM (see above). Potentially suitable habitat within the glacier polygons represents ridges and slopes, whereas glaciers disproportionately affect drainages (where the majority of wetland habitat is located). The right illustration shows the montane landscape, with whiter colors representing higher elevations. For another perspective on Pleistocene glaciers within the southern Rocky Mountains, watch the 'Late Pleistocene glaciers of Colorado' video created by the Integrative Geology Project at the University of Colorado at Boulder (http://igp.colorado.edu/animations.html).

Table B.1 Summary statistics for the sampled populations of *C. chalciolepis* and *C. nova*. Results are presented for all nucleotide positions (polymorphic + fixed). Shown are the number of loci, the percentage of loci that are polymorphic, the average frequency of the major allele (*P*), the average observed heterozygosity per locus (H_{obs}), the average nucleotide diversity (π), and the Wright's inbreeding coefficient (F_{IS}).

	Population	Loci	% polymorphic loci	*P*	H_{obs}	π	F_{IS}
C. chalciolepis	Oso	593,856	1.37	0.995	0.004	0.007	0.0061
	Lizard Head	481,613	0.93	0.997	0.002	0.004	0.0039
	Blanca	616,201	0.89	0.998	0.002	0.003	0.0024
	Ouray	586,951	0.60	0.998	0.002	0.003	0.0023
	Lamphier	614,807	0.58	0.998	0.002	0.003	0.0019
	Kite	536,372	1.36	0.996	0.005	0.006	0.0020
	Guanella	590,817	0.79	0.998	0.002	0.003	0.0032
C. nova	Oso	589,187	0.86	0.997	0.002	0.004	0.0045
	Lizard Head	412,113	0.67	0.998	0.002	0.003	0.0028
	Blanca	532,231	0.99	0.997	0.002	0.004	0.0049
	Ouray	589,264	1.22	0.997	0.004	0.005	0.0018
	Lamphier	511,657	0.77	0.998	0.002	0.004	0.0032
	Kite	563,135	0.87	0.997	0.002	0.004	0.0044
	Guanella	399,860	0.49	0.998	0.002	0.003	0.0012

Table B.2 Population pairwise F_{ST} values (below diagonal) and Euclidean distances (above diagonal) for *C. chalciolepis* (A) and *C. nova* (B).

(A) *Carex chalciolepis*

	Oso	Lizard	Blanca	Ouray	Lamphier	Kite	Guanella
Oso		51.5	177	148	148	231	276
Lizard	0.083		218	163	150	229	276
Blanca	0.088	0.080		115	155	201	222
Ouray	0.060	0.061	0.061		42	100	137
Lamphier	0.063	0.061	0.064	0.053		83	128
Kite	0.078	0.072	0.072	0.058	0.064		47
Guanella	0.069	0.070	0.076	0.058	0.061	0.062	

(B) *Carex nova*

	Oso	Lizard	Blanca	Ouray	Lamphier	Kite	Guanella
Oso		51.5	177	148	148	231	276
Lizard	0.064		218	163	150	229	276
Blanca	0.062	0.065		115	155	201	222
Ouray	0.061	0.057	0.071		42	100	137
Lamphier	0.052	0.056	0.066	0.056		83	128
Kite	0.058	0.059	0.064	0.064	0.059		47
Guanella	0.041	0.043	0.043	0.051	0.045	0.048	

APPENDIX C

Chapter 5 Supplementary Material

Environmental niche modeling methodology

ENMs were generated from bioclimatic variables for the present and the LGM for each species with MAXENT v3.3.3e (Phillips et al. 2006) using the following parameters: regularization multiplier = 1, max number of background points = 10,000, replicates = 50, replicated run type = cross-validate. To avoid overfitting of the distribution models, the geographic extent of the environmental layers was reduced to an area approximately 20% larger than the known distribution of the species (Anderson and Raza 2010). To guard against the inherent difficulties involved in extrapolating distributions into novel climates (reviewed in Alvarado-Serrano and Knowles 2013), an iterative approach was used to generate ENMs for the LGM. Multivariate environmental similarity surfaces (MESS maps) were used to identify which of the 19 bioclimatic variables resulted in areas of low reliability predictions due to the variables being outside of the range present in the present-day environmental data (Elith et al. 2010). MAXENT was rerun excluding these out-of-range variables, and this process of analysis with MESS maps was repeated until no LGM variables were out-of-range compared to present-day bioclimatic variables. Because MESS maps do not indicate changes in the correlations among the environmental variables used for LGM reconstructions (Elith et al. 2010), we checked our LGM ENM using only the most informative variable (Bio5) to ensure we were not reporting errant distributional patterns. In addition, a present-day ENM was generated for the subset of variables that were not out-of-range and compared to a ENM constructed using all climatic variables with greater than 5% importance (determined by jackknifing) to assess their similarity.

Library construction and data processing

Anonymous genomic loci were developed from five Illumina 2500 sequencing runs using a restriction associated DNA sequencing (RADseq) approach (for details see Peterson et al. 2012). Briefly, DNA was doubly digested with *Eco*RI and *Mse*I restriction enzymes, followed by the ligation of Illumina adaptor sequences and unique 10 base pair barcodes. Ligation products were pooled among samples and the fragments were amplified by 12 cycles of PCR. A Pippin Prep (Sage Science) was used to size select fragments between 400 and 500 base pairs. The library

was sequenced at The Centre for Applied Genomics (Hospital for Sick Children, Toronto, Canada) to generate 50 base pair, single-end reads. Sequences were demultiplexed using process_radtags.pl, which is distributed as part of the Stacks pipeline (Catchen et al. 2013); only reads with Phred scores ≥ 32, no adaptor contamination, and that had an unambiguous barcode and restriction cut site were retained. Potential chloroplast and mitochondrial sequences were filtered from the processed dataset using Bowtie 0.12.8 (Langmead et al. 2009) (see Massatti and Knowles 2014). Reads showed consistently high sequence quality and remained untrimmed, except for the barcode and restriction enzyme cut site.

Single nucleotide polymorphisms (SNPs) were identified at each RADseq locus and genotypes were called using a multinomial-based likelihood model that accounts for sequencing error as implemented in Stacks v1.25 (Catchen *et al*. 2011; Hohenlohe *et al*. 2012; Catchen *et al*. 2013). A conservative upper bound of the error rate (ε) was set to 0.1 to avoid underestimating heterozygotes (Catchen *et al*. 2013). In the first step of the Stacks pipeline, loci and polymorphic nucleotide sites were identified in each individual using the USTACKS program, which groups reads with a minimum coverage depth (m) into a "stack". The data were processed with m = 3; increasing the minimum depth helps to avoid erroneously calling convergent sequencing errors as stacks. Reads were filtered using a removal algorithm that eliminated highly repetitive stacks (i.e., stacks that exceed the expected number of reads for a single locus given the average depth of coverage, for example, when loci are members of multi-gene families) and a 'deleveraging algorithm' to resolve over-merged loci (i.e., non-homologous loci misidentified as a single locus). A catalog of consensus loci among individuals was constructed with the CSTACKS program from the USTACKS output files for each species, where loci were merged together across individuals if the distance between them (n) was ≤ 2. This catalog was used to determine the allele(s) present in each individual at each homologous locus using the SSTACKS program. Our choice of parameters was determined with consideration of avoiding both over- and under-merging of homologous loci in the focal taxa, as well as with reference to other studies (e.g., Renaut et al. 2014). Similarity of the number of loci identified in the species for different parameter values used in USTACKS (m) and CSTACKS (n) suggests that the properties of the genomic libraries were similar (i.e., that the potential errors associated with over- or under-merged homologous loci did not differ substantially between the species). The close relatedness of the taxa and the short reads makes an n ≤ 2 reasonable (also see Renaut et al. 2014), although

we acknowledge this could be conservative if catalogs were assembled for taxa that were more distantly related, and/or for read lengths larger than 50 base pairs.

Literature Cited

Alvarado-Serrano, D. F., and L. L. Knowles. 2013. Ecological niche models in phylogeographic studies: applications, advances, and precautions. Mol. Ecol. Resour. 14:233-248.

Anderson, R. P., and A. Raza. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. J. Biogeogr. 37:1378–1393.

Catchen, J., P. 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.

Elith, J., M. Kearney, and S. Phillips. 2010. The art of modeling range-shifting species. Method. Ecol. Evol. 1:330–342.

Hohenlohe, P. A., J. Catchen, and W. A. Cresko. 2012. Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. Methods Mol. Biol. 888:235–260.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Massatti, R. and L. L. Knowles. 2014. Microhabitat differences impact phylogeographic concordance of codistributed species: Genomic evidence in montane sedges (Carex L.) from the Rocky Mountains. Evolution 68:2833-2846.

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.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190:231–259.

Renaut, S., G. L. Owens, and L. H. Rieseberg. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Mol. Ecol. 23:311-324.

Table C.1 Summary of genomic data collected for each individual for A) *Carex chalciolepis* and B) *C. nova*. Shown are the raw counts of reads from the Illumina run ('Total') and the number of 'Retained' reads after processing for quality control (i.e., after excluding 'Low Quality' reads and 'No RadTag' reads). We also report the number of remaining reads after filtering out potential chloroplast and mitochondrial DNA ('post-bowtie', see additional Methods above) and the number of reads retained by Stacks after filtering out potential paralogous loci and over-merged loci; this latter number (also represented by 'total % retained') is the data which was used to identify homologous loci. Starred individuals were excluded from analyses because they had too few reads.

A) *Carex chalciolepis*

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	# retained by stacks	total % retained
Libby Flats 1	2269864	6142	10146	2240426	2048756	1985763	0.87
Libby Flats 2	2573898	3335	10804	2541563	2379309	2320139	0.90
Libby Flats 3	1504829	980	7860	1485750	1375851	1319219	0.88
Libby Flats 4	1755431	2211	7437	1732805	1612477	1554853	0.89
Libby Flats 5	970330	1102	4127	958742	900554	855331	0.88
Libby Flats 6	742915	818	3583	734324	681104	633809	0.85
Libby Flats 7	835137	1030	3431	825805	759477	712137	0.85
Libby Flats 8	437541	344	2044	432501	400529	365003	0.83
Libby Flats 9	1341269	3005	6190	1324363	1201995	1157381	0.86
Libby Flats 10	1665439	2337	6625	1645436	1527170	1473164	0.88
Zirkel 1	1063134	1777	4531	1047625	992120	942644	0.89
Zirkel 2	1346474	2608	6102	1328265	1219778	1166800	0.87
Zirkel 3	576525	1267	2762	568487	525998	491704	0.85
Zirkel 4	607590	706	3442	599091	547372	513109	0.84
Zirkel 5	872180	798	5114	861119	793244	749972	0.86
Zirkel 6	647245	3598	3081	635940	594267	552664	0.85
Zirkel 7	2139688	2319	9483	2114055	1970048	1907541	0.89
Zirkel 8	2944771	3545	12273	2907035	2710240	2645211	0.90
Zirkel 9	1818223	2204	7449	1796689	1664415	1604871	0.88
Zirkel 10	1155351	1025	5252	1142082	1025115	976428	0.85
Niwot 1	1550383	1115	7668	1532100	1423214	1365514	0.88
Niwot 2	790991	1313	3446	781804	713742	665585	0.84
Niwot 3	405252	1647	2353	398680	364674	331555	0.82
Niwot 4	715457	2334	3606	704831	647777	601083	0.84
Niwot 5	1745519	3812	7043	1724517	1567758	1509504	0.86
Niwot 6	801720	1477	3199	791859	732237	681640	0.85
Niwot 7	913065	2052	3840	901643	832593	778875	0.85
Niwot 8	1563337	2572	9074	1543008	1407370	1345624	0.86
Niwot 9*	1357108	2323	5841	1339197	1238262	1163446	0.86
Niwot 10	1993594	3599	8279	1973450	1796709	1741820	0.87

Table C.1 Continued

A) *Carex chalciolepis* (continued)

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	# retained by stacks	total % retained
Flat Tops 1	1281735	2146	9117	1264583	1170587	1124335	0.88
Flat Tops 2	949146	1502	8949	932787	878658	843867	0.89
Flat Tops 3	1222176	766	10798	1205174	1101017	1057760	0.87
Flat Tops 4*	1159794	4026	7675	1142138	1057284	999909	0.86
Flat Tops 5	1639642	5258	9267	1617543	1507478	1456427	0.89
Flat Tops 6	1509101	3735	7108	1492288	1371029	1324377	0.88
Flat Tops 7	1665466	2654	9352	1645704	1535509	1483874	0.89
Flat Tops 8*	1274850	3290	8434	1257126	1143598	1096990	0.86
Flat Tops 9	1860772	11046	10889	1831375	1684843	1635382	0.88
Flat Tops 10	1912967	4609	9293	1890250	1763882	1713542	0.90
Guanella 1	1458271	3534	7190	1434169	1308310	1261141	0.86
Guanella 2	2971729	5286	16340	2926574	2584773	2541363	0.86
Guanella 3	1492005	2609	7948	1469964	1323578	1273845	0.85
Guanella 4	3127383	8849	14782	3080278	2759085	2719368	0.87
Guanella 5	2715139	6064	14184	2672038	2462410	2420913	0.89
Guanella 6	1537937	6191	5841	1515598	1378722	1323276	0.86
Guanella 7	1669606	3243	6980	1648436	1518827	1472562	0.88
Guanella 8	4211731	15359	19153	4148991	3755221	3705054	0.88
Guanella 9	1469160	4198	6315	1449804	1335305	1280162	0.87
Guanella 10*	3111472	7269	14339	3074200	2825746	2764526	0.89
Pikes 1	1619770	2917	13632	1595195	1510606	1462900	0.90
Pikes 2	1222818	2240	6408	1208780	1117193	1079582	0.88
Pikes 3	1638230	2437	8842	1619832	1480713	1435332	0.88
Pikes 4	1775795	3662	13461	1751523	1615896	1568548	0.88
Pikes 5	1833297	3720	11525	1809224	1693295	1648734	0.90
Pikes 6	1268042	1322	8963	1251667	1180047	1130373	0.89
Pikes 7	1862884	5477	13155	1834941	1723379	1666946	0.89
Pikes 8	1807463	3582	9117	1786450	1687921	1638245	0.91
Pikes 9*	1569717	2645	10025	1550728	1442070	1390945	0.89
Pikes 10	1588975	3551	11303	1565188	1486937	1445114	0.91

Table C.1 Continued

A) *Carex chalciolepis* (continued)

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	# retained by stacks	total % retained
Lamphier 1	1636336	2996	9463	1609267	1445797	1400435	0.86
Lamphier 2	1841268	4584	13063	1807495	1663778	1617433	0.88
Lamphier 3	2873259	5229	16737	2823532	2561462	2518649	0.88
Lamphier 4	3004800	3924	14867	2961664	2589076	2545177	0.85
Lamphier 5	1905873	5031	10807	1875059	1660991	1611309	0.85
Lamphier 6	2545975	2975	9972	2519010	2292639	2238556	0.88
Lamphier 7	394867	6493	1749	383721	354203	328423	0.83
Lamphier 8	1687309	1819	10905	1663247	1501340	1445282	0.86
Lamphier 9	1035984	2084	4542	1023101	940008	887141	0.86
Lamphier 10	1411129	1982	6511	1393838	1284176	1229416	0.87
Ouray 1	3506390	9107	19502	2310542	2138571	2094172	0.60
Ouray 2	1500608	2701	13081	1471015	1354440	1306657	0.87
Ouray 3	2834687	11417	14591	2787490	2427877	2385520	0.84
Ouray 4*	1118144	2585	7568	1098638	1006616	941060	0.84
Ouray 5*	2389863	7812	14712	2345401	2176674	2123558	0.89
Ouray 6	2045589	4848	10946	2018812	1826892	1775259	0.87
Ouray 7	2027806	7582	9901	1998708	1841996	1785629	0.88
Ouray 8	1994387	3058	7691	1972088	1806177	1750776	0.88
Ouray 9*	1677496	3152	6086	1660990	1485781	1418445	0.85
Ouray 10	1023713	1078	5628	1010293	918117	865384	0.85
Lizard 1	1675863	1920	7351	1654055	1471929	1417190	0.85
Lizard 2	2148107	3077	11186	2119317	1896099	1842942	0.86
Lizard 3	3131109	5919	20730	3075716	2836319	2790949	0.89
Lizard 4	2207051	2811	11723	2176388	1955333	1908209	0.86
Lizard 5	2994538	3486	15385	2953216	2568118	2524555	0.84
Lizard 6	1177007	1173	4976	1164175	1063852	1003116	0.85
Lizard 7	1508159	2379	6375	1488943	1342463	1284054	0.85
Lizard 8	1472946	1284	6976	1454974	1318444	1265080	0.86
Lizard 9	1417728	4366	6103	1399090	1258805	1203953	0.85
Lizard 10	1330078	1613	5665	1313702	1209197	1152877	0.87

Table C.1 Continued

A) *Carex chalciolepis* (continued)

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	# retained by stacks	total % retained
Blanca 1	1562719	4225	10372	1532369	1386816	1329958	0.85
Blanca 2	1753655	5743	10744	1723013	1576806	1525892	0.87
Blanca 3	2234245	25494	11511	2177979	2012159	1830924	0.82
Blanca 4	1979090	7448	11922	1944761	1797427	1755035	0.89
Blanca 5	835078	2307	4543	819391	740622	690421	0.83
Blanca 6	2116307	5558	12194	2078169	1886437	1842756	0.87
Blanca 7	1007990	5456	5325	986646	889344	844002	0.84
Blanca 8	1502721	3533	7672	1478947	1344292	1291318	0.86
Blanca 9	2577594	4851	15463	2533168	2275093	2232178	0.87
Blanca 10	1938477	5664	12638	1903194	1715201	1668293	0.86
Red Lakes 1	927162	861	5189	915076	841728	791382	0.85
Red Lakes 2	1040497	1014	5199	1027506	927637	878116	0.84
Red Lakes 3	949432	2037	4225	936043	865723	814952	0.86
Red Lakes 4	1585695	2465	7265	1566591	1436470	1380761	0.87
Red Lakes 5	666296	850	2604	658327	603939	562694	0.84
Red Lakes 6*	376839	398	1624	372826	332856	306750	0.81
Red Lakes 7	732250	947	2566	724151	663214	617900	0.84
Red Lakes 8	1780382	2317	7837	1759078	1613762	1558417	0.88
Red Lakes 9	1293224	3017	8868	1272830	1155884	1110401	0.86
Red Lakes 10	1421735	2741	6072	1403902	1306778	1253594	0.88

Table C.1 Continued

B) *Carex nova*

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	utilized by stacks	total % retained
Libby Flats 1	2839544	8949	15530	2784008	2533428	2500386	0.88
Libby Flats 2	1424530	4104	7377	1398837	1294738	1251058	0.88
Libby Flats 3	2786782	8502	18805	2731054	2534674	2495825	0.90
Libby Flats 4	1140254	2343	6497	1118951	1046745	991595	0.87
Libby Flats 5	1594281	2699	10072	1565307	1467794	1423378	0.89
Libby Flats 6	1539315	3731	8628	1515512	1362497	1318823	0.86
Libby Flats 7	1331829	2004	6243	1312948	1180772	1140793	0.86
Libby Flats 8	1965130	3114	11005	1937037	1731268	1693540	0.86
Libby Flats 9	2713249	7620	14792	2672865	2392125	2352075	0.87
Zirkel 1	1073382	1911	8342	1051600	978356	919552	0.86
Zirkel 2	1289822	1206	9941	1267232	1157776	1104180	0.86
Zirkel 3	1552260	7397	9337	1521159	1418164	1372391	0.88
Zirkel 4	1668094	3858	9057	1640267	1521502	1476176	0.88
Zirkel 5	3281087	9389	16949	3225692	3003577	2962763	0.90
Zirkel 6	2498606	5670	13717	2457125	2279927	2215288	0.89
Zirkel 7	1904644	4513	11585	1869362	1766131	1724315	0.91
Zirkel 8	1101950	10667	6303	1073662	978695	932277	0.85
Zirkel 9	2064636	6312	10224	2030063	1897568	1856236	0.90
Zirkel 10	2287641	6399	15951	2244013	2098058	2055577	0.90
Niwot 1	2030466	3500	14676	1995845	1829983	1785708	0.88
Niwot 2	2001624	7733	12220	1962323	1829839	1788093	0.89
Niwot 3	2710605	6361	13526	2666675	2505161	2464430	0.91
Niwot 4	1291454	2088	5799	1271894	1184785	1139019	0.88
Niwot 5	1291026	2446	6703	1268767	1183554	1137097	0.88
Niwot 6	1190172	2026	9293	1166854	1107520	1056340	0.89
Niwot 7	2343084	4146	12808	2302162	2147806	2107614	0.90
Niwot 8	1809134	3966	9589	1779032	1666026	1622921	0.90
Niwot 9	2620391	6479	14318	2579303	2387382	2345214	0.89
Niwot 10	620585	772	3307	611239	572858	517588	0.83

Table C.1 Continued

B) *Carex nova* (continued)

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	utilized by stacks	total % retained
Flat Tops 1	810170	949	5067	798597	756411	697079	0.86
Flat Tops 2	1023357	7056	8562	1000654	938142	893188	0.87
Flat Tops 3	1164665	3420	7727	1146008	1058107	1019706	0.88
Flat Tops 4	526658	1044	2463	519550	480866	454152	0.86
Flat Tops 5	1102549	1543	9717	1085986	1012971		0.00
Flat Tops 6	1321773	3391	5767	1303501	1193859	1153086	0.87
Flat Tops 7	507149	786	3326	499780	460283	430725	0.85
Flat Tops 8	343938	421	2591	338730	317708	265775	0.77
Flat Tops 9	567203	1942	2997	557796	530467	470203	0.83
Flat Tops 10	549947	1210	2316	542903	499604	443961	0.81
Guanella 1	1211961	7079	5487	1189739	1072559	1024498	0.85
Guanella 2	1231002	2506	5772	1214365	1120575	1066358	0.87
Guanella 3	3584067	11845	14019	3536390	3182524	3128547	0.87
Guanella 4	3201045	3856	14415	3162440	2838739	2780679	0.87
Guanella 5	2491327	4947	10452	2460699	2200591	2140934	0.86
Guanella 6	1544413	5246	8453	1517411	1406985	1364925	0.88
Guanella 7	819044	2015	4184	806862	745568	698961	0.85
Guanella 8	857027	2261	4801	843634	770759	709405	0.83
Guanella 9	2948088	7397	18044	2897628	2654812	2608655	0.88
Guanella 10	1807001	6426	12091	1772102	1644002	1600492	0.89
Pikes 1	1602846	5186	8286	1579714	1454106	1406076	0.88
Pikes 2	414410	2696	3090	405745	391030	334458	0.81
Pikes 3	1008799	3460	7084	990296	945571	901448	0.89
Pikes 4	884506	7799	4566	865769	824611	779829	0.88
Pikes 5	1564878	3466	11058	1538011	1492117	1444438	0.92
Pikes 6	710198	2505	4052	698471	666069	628063	0.88
Pikes 7	1151385	2118	7379	1133267	1072035	1030474	0.89
Pikes 8	984305	4585	5024	966830	912318	874828	0.89
Pikes 9	898930	1622	5300	886068	838042	795041	0.88
Pikes 10	1158679	2038	8509	1138853	1078539	1039516	0.90

Table C.1 Continued

B) *Carex nova* (continued)

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	utilized by stacks	total % retained
Lamphier 1	1207998	2612	8513	1205269	1112923	1055301	0.87
Lamphier 2	584168	830	4448	583221	544214	506544	0.87
Lamphier 3	598713	593	1433	598003	559209	518713	0.87
Lamphier 4	1058901	1699	5229	1057085	978864	925553	0.87
Lamphier 5	1340400	2932	5540	1337351	1248669	1190282	0.89
Lamphier 6	1768369	2249	16084	1734711	1580734	1536471	0.87
Lamphier 7	1378070	2089	7599	1355526	1256279	1202878	0.87
Lamphier 8	1497783	1915	9865	1471572	1403911	1361772	0.91
Lamphier 9	2603490	4631	12481	2564801	2408237	2367964	0.91
Lamphier 10	3394716	7405	18013	3387194	3148673	3108451	0.92
Ouray 1	925230	2458	4293	912831	809603	761216	0.82
Ouray 2	3932347	9885	21090	3875347	3581075	3531749	0.90
Ouray 3	1736257	8139	7764	1541492	1395841	1339145	0.77
Ouray 4	1977064	7707	12509	1943969	1731037	1671777	0.85
Ouray 5	1802514	10704	7747	1770767	1641652	1588909	0.88
Ouray 6	1395209	1631	7004	1374300	1265140	1221092	0.88
Ouray 7	1568824	2354	7592	1543603	1416011	1369568	0.87
Ouray 8	1116198	1387	7429	1096480	1038920	989745	0.89
Ouray 9	1670200	1868	12303	1638869	1541219	1498637	0.90
Ouray 10	1084383	803	7744	1065857	995310	939625	0.87
Lizard 1	1536985	2310	5476	1517004	1378427	1321729	0.86
Lizard 2	2487989	5924	9890	2458239	2180260	2125758	0.85
Lizard 3	2715118	3354	9835	2687524	2406839	2343368	0.86
Lizard 4	1331422	2163	5861	1314424	1173924	1121514	0.84
Lizard 5	682648	772	3637	673691	602966	563331	0.83
Lizard 6	1562742	1120	10000	1534944	1387372	1340917	0.86
Lizard 7	2137479	2462	12451	2102749	1912627	1867963	0.87
Lizard 8	1937613	3616	10180	1906210	1726094	1681383	0.87
Lizard 9	1583255	1612	11994	1553392	1427103	1379960	0.87
Lizard 10	1505177	1037	9793	1479450	1389034	1346518	0.89

Table C.1 Continued

B) *Carex nova* (continued)

Individual	Total	No RadTag	Low Quality	Retained	post-bowtie	utilized by stacks	total % retained
Blanca 1	927985	1007	3217	923761	881899	832796	0.90
Blanca 2	1305721	1703	4760	1299258	1183868	1126511	0.86
Blanca 3	935678	2365	2537	930776	885593	836250	0.89
Blanca 4	968218	5351	6616	956251	901180	851092	0.88
Blanca 5	1103123	2611	2190	1098322	1055760	998406	0.91
Blanca 6	1472012	14483	5556	1451973	1328549	1271697	0.86
Blanca 7	1098224	3993	7485	1086746	998237	954405	0.87
Blanca 8	1755689	2960	8116	1744613	1628886	1570506	0.89
Blanca 9	970736	2027	5045	963664	899690	850013	0.88
Blanca 10	1878365	7078	7326	1863961	1737225	1669506	0.89
Red Lakes 1	1391009	1480	7197	1370117	1225492	1183143	0.85
Red Lakes 2	2616139	4289	15347	2572327	2334633	2283918	0.87
Red Lakes 3	1635262	2222	7779	1610001	1487183	1446554	0.88
Red Lakes 4	1123129	1115	6570	1104941	1020171	967227	0.86
Red Lakes 5	1551242	2554	6827	1528216	1399386	1349936	0.87
Red Lakes 6	1673811	2329	9655	1648000	1499799	1448806	0.87
Red Lakes 7	1318874	5296	13129	1284025	1200077	1153884	0.87
Red Lakes 8	2318899	3787	12724	2281156	2108525	2062472	0.89
Red Lakes 9	1548360	14584	6888	1512846	1407475	1362440	0.88
Red Lakes 10	2481807	20173	12870	2424165	2292654	2247581	0.91

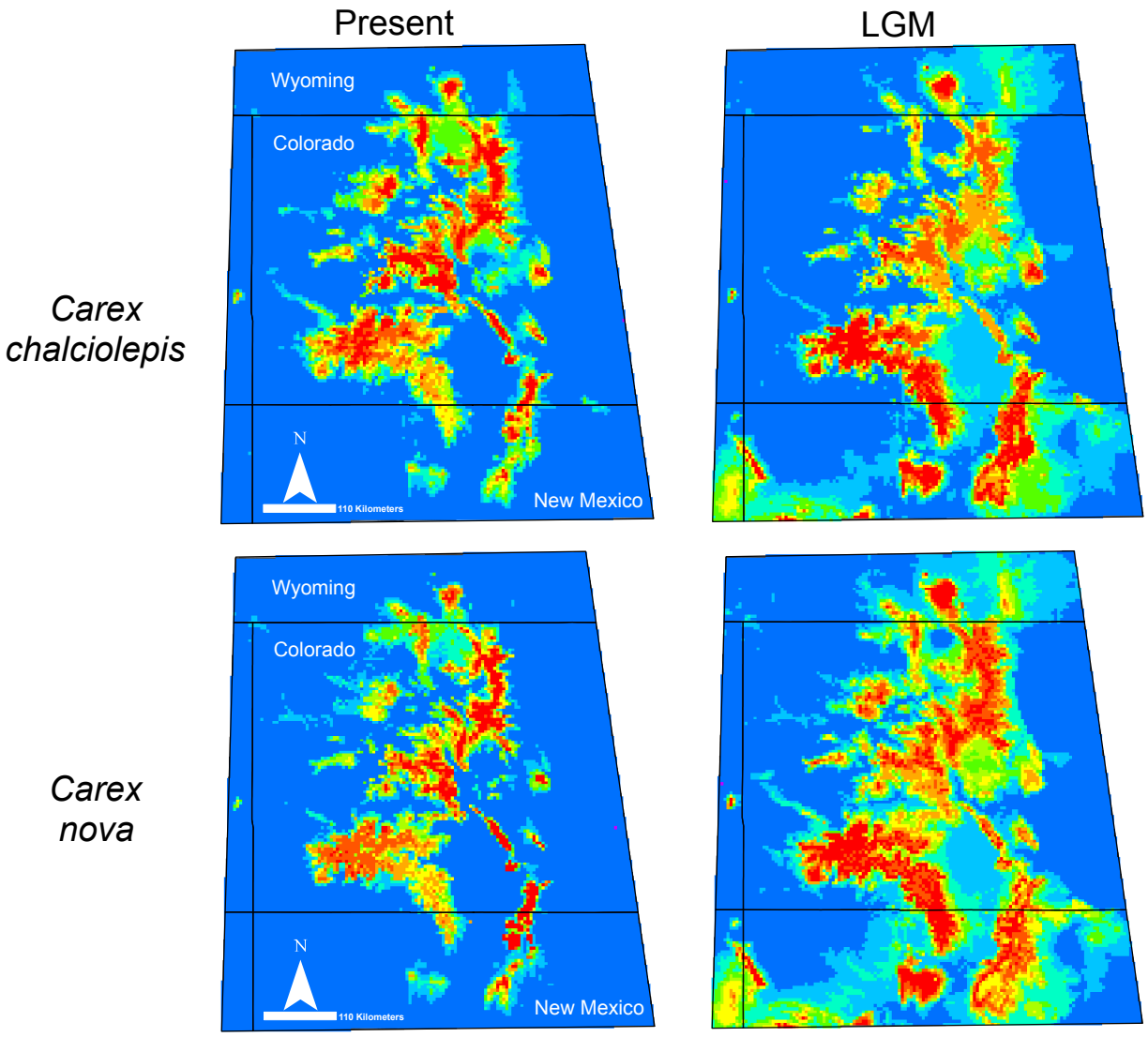

Figure C.1 Present (left) and LGM (right) ENMs for *Carex chalciolepis* and *C. nova,* as estimated with MAXENT. Due to their similarity within the time periods, the ENMs were averaged before they were modified into the landscapes used in the modeling process (see Methods).

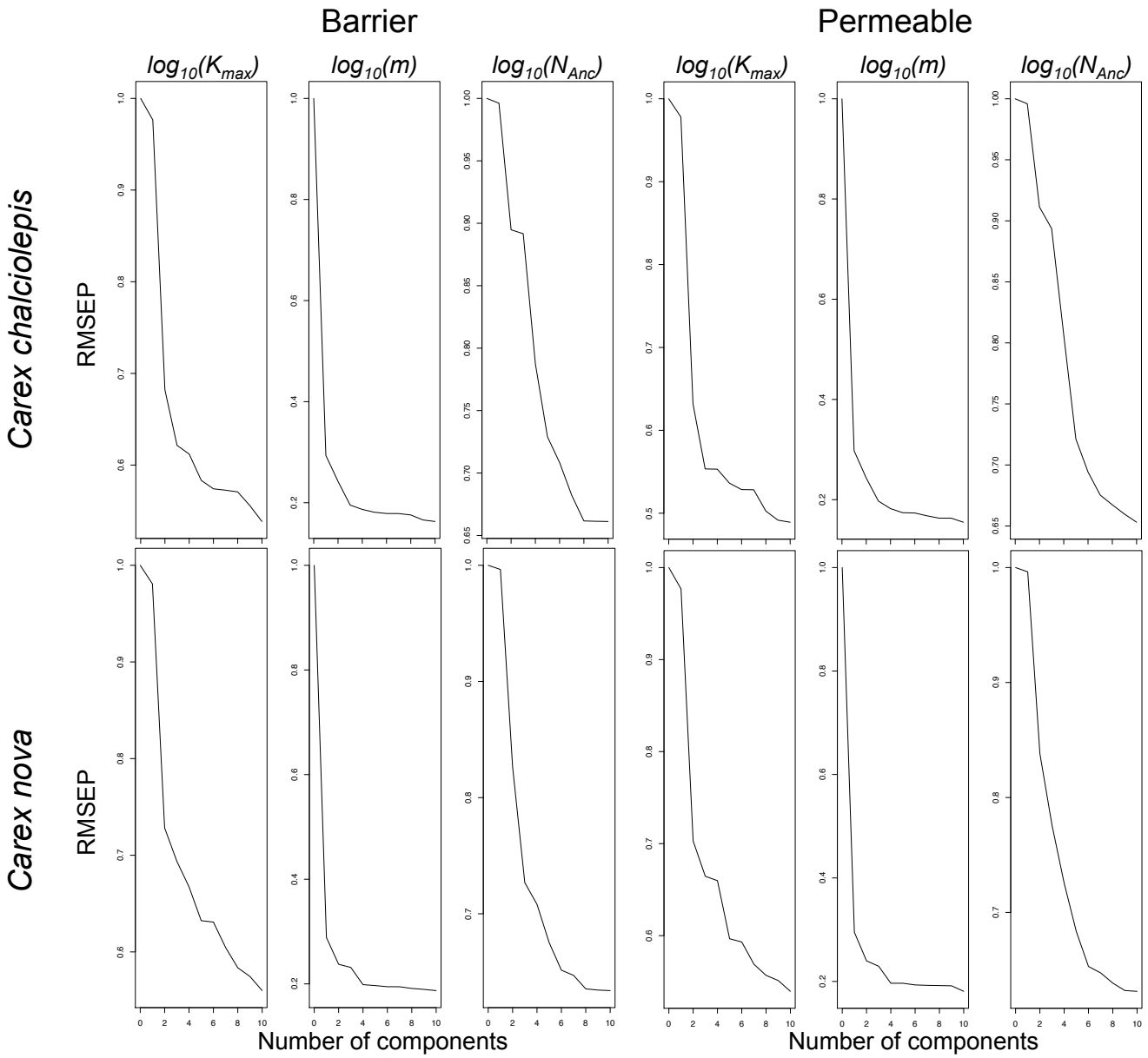

Figure C.2 Root Mean Square Error (RMSE) of parameter estimation against the number of PLSs included under two demographic models: Barrier (left column) or Permeable (right column) for *C. chalciolepis* (top) and *C. nova* (bottom).