PHYLOGEOGRAPHY AND POPULATION GENOMICS OF THE AMERICAN BLACK BEAR (URSUS AMERICANUS)

EMILY ELIZABETH PUCKETT

A dissertation submitted in partial fulfillment of the requirements for the degree
Doctor of Philosophy

Presented to the Faculty of the Graduate School
University of Missouri
Columbia, Missouri

MAY 2015

The undersigned, appointed by the Dean of the Graduate School,
have examined the dissertation entitled:

PHYLOGEOGRAPHY AND POPULATION GENOMICS OF THE
AMERICAN BLACK BEAR (URSUS AMERICANUS)

Presented by Emily E Puckett, a candidate for the degree of Doctor of Philosophy,
and hereby certify that it is worthy of acceptance.

Lori S. Eggert	J. Chris Pires
Robert D. Schnabel	Raymond D. Semlitsch

ACKNOWLEDGEMENTS

I sincerely appreciated the support of my family: Jay Puckett, Tatum, Oslo, and Cinco de Gato Pulley-Puckett, and Lily Puckett. Thank you for moving to Missouri so I could fulfill this life goal.

I appreciated the opportunity to train provided by my advisor Lori Eggert. Sincere thanks to my committee Chris Pires, Bob Schnabel, and Ray Semlitsch for challenging me both methodologically and in data interpretation, both of which proved invaluable to my scientific growth. I also appreciated each of your open door policies.

Sincere thanks to Dylan Kesler for working on a side project with me that provide an often needed break from my dissertation and a unique point of view. I appreciated the bioinformatics support and training provided by Bill Spollen and Scott Given of the Informatics Research Core Facility at the University of Missouri. Thanks to Mari Ruiz-Lopez for helpful conversations regarding RADseq bioinformatics and molecular ecology.

I appreciated my Eggert Lab colleagues Stephanie Schuttler, Tabitha Finch, Sheena Feist, Bill Peterman, Karen DeMatteo, Jake Burkhart, and Martha Fischer for the friendly and helpful atmosphere. Thanks to my undergraduate assistances Michelle Anderson and Jessica Wayhart for a whole lot of PCRs, and the honor to watch you each grow as scientists. Thanks also to the Pires-Conant Lab members for their support and feedback,

specifically Gavin Conant, Kate Hertweck, Pat Edger, Dustin Mayfield-Jones, Jacob Washburn, Sarah Unruh, Andrea Ravelo, Wade Dismukes, and Kevin Bird.

I appreciated the University of Missouri Life Sciences Fellowship which allowed me to focus on research for four years. This work was partially funded by the Missouri Department of Conservation.

Finally, this project would not have been possible without the samples generously provided by Rod Flynn and Gino DelFrate (Alaska Department of Fish and Game), Brian Wakeling (Arizona Game and Fish Department), Bradley Bowden (Cariboo Mountain Outfitters), Kevin Bracewell (BC Trophy Mountain Outfitters) and Wilderness Stewardship Foundation, Doug McMann (Skinner Creek Hunts), Mat Alldredge, Chad Bishop, and Jerry Apker (Colorado Department of Wildlife), Jason Hawley (Connecticut Department of Energy), Mike Orlando and Mike Connelly (Florida Fish and Wildlife Conservation Commission), Jim Hayden and Jeff Rohlman (Idaho Fish and Game), Randy Cross (Maine Department of Inland Fisheries and Wildlife), Laura Hajduk-Conlee (Massachusetts Department of Fish and Game), Hank Hristienko (Manitoba Wildlife Branch), Dwane Etter (Michigan Department of Natural Resources), Karen Noyce (Minnesota Department of Natural Resources), Mississippi Museum of Natural Science, Jeff Beringer (Missouri Department of Conservation), Neil Anderson and Justin Gude (Montana Fish, Wildlife, and Parks), Rick Winslow (New Mexico Department of Game and Fish), Matthew Walter, Jeremy Hurst, and Ed Reed (New York State Department of Environmental Conservation), Colleen Olfenbuttel and Mike Carraway (North Carolina

Wildlife Resources Commission), Marty Obbard (Ontario Ministry of Natural Resources and Forestry), Agnes Pelletier and Chris Kyle (Trent University), Tim Hiller (Oregon Department of Fish and Wildlife), Mark Ternent (Pennsylvania Game Commission), Colin Carpenter (West Virginia Division of Natural Resources), Dave MacFarland (Wisconsin Department of Natural Resources), and Ramona Marja (Environment Yukon). Additionally, I thank any field crew members that participated in sample collection even if not listed above.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABSTRACT

The American black bear (*Ursus americanus*) is one of the eight living species of Ursidae, and the only one to have speciated in North America (1.8 – 1.2 Mya). This dissertation investigates the contemporary population structure of American black bears across their range; and specifically asks how a translocation of bears affected the population genetics of individuals in the Central Interior Highlands. Black bear mitochondrial lineages began forming within the last 170 kya, whereas the eastern and western nuclear genomes diverged 67 kya. A third nuclear lineage was discovered in contemporary Alaska, which diverged from the eastern lineage 31 kya. These three lineages harbor nine genetic clusters, and potentially more in unsampled portions of the range. These nine clusters may represent evolutionary significant units for the species; however, more work would be needed before proposing taxonomic revisions. The regional population genetics of the Central Interior Highlands (Arkansas, Oklahoma, and Missouri, USA) showed that the majority of genetic diversity in contemporary populations of the Ozark and Ouachita Mountains was introduced from Minnesota, USA and Manitoba, Canada during a translocation of bears from 1958 - 1968. Analyses also indicated that the contemporary Ozark and Ouachita populations were genetically differentiated. Additionally, bears that form a low diversity genetic cluster in Missouri were highly similar to bears from the Ozarks in genomic analyses, indicating a small founding population dispersed northwards following the reintroduction. Finally, I analyzed the accuracy and precision with which the natal location of a black bear may be identified using different inference methods and dataset compositions. While samples

were estimated within 201 km of their sample site and with high precision, there was a low correlation between the state or province of sampling and that estimated. These results suggest caution when using genetic data for natal inference problems in cases for trade of wildlife products.

CHAPTER 1

Biogeography, evolution, and contemporary conservation issues within the genus *Ursus*

Emily E. Puckett

Division of Biological Sciences, 226 Tucker Hall, University of Missouri, Columbia, MO 65211

The Ursidae are a family of eight extant bear species distributed across a range of forested ecosystems. Based on the fossil record, bears most likely diverged from canids approximately 20 - 25 Mya from a Holarctic range in contemporary Eurasia (McLellan & Reiner 1992). Two bear subfamilies (Hemicyoninae and Agriotherinnae) went extinct in the Pliocene (McLellan & Reiner 1992). The earliest diverging lineage with an extant species, the giant panda (*Ailuropoda melanoleuca*), is the Ailuropodinae. Species of *Aluriopoda* spread throughout eastern Asia; however, giant panda is now confined to isolated bamboo forests in China. The second subfamily with a single extant species is the Tremarctinae. Fossils of the species in this lineage are found throughout North and South America suggesting divergence following dispersal from an Asian-North American landbridge. Only the spectacled bear (*Tremarctos ornatus*) remains, living in the forests from Bolivia to Venezuela in the Andes Mountain chain (Figure 1.1). The Ursinae has one extant genus (*Ursus*) with six species: *U. ursinus* (sloth), *U. malayanus* (sun), *U. thibetanus* (Asian black), *U. americanus* (American black), *U. arctos* (brown), and *U. maritimus* (polar). The sloth bear is primarily distributed in India, the sun bear has a southeast Asian distribution, and the Asiatic black bear is broadly distributed throughout Asia (Figure 1.1). The American black bear is distributed across North America and the

polar bear has an artic distribution. The brown bear is widely distributed from Europe, across Eurasia as far south as Tibet, and into western North America.

Evolutionarily *Ursus* is a young genus diversifying 3 - 7 Mya (Krause *et al.* 2008). This Pliocene radiation resulted in numerous species; however, many went extinct during the Pleistocene (Krause *et al.* 2008). Two of the species that went extinct were *U. minimus* and *U. abstrusus* which are the respective Old and New World species from which contemporary species shared common ancestors (McLellan & Reiner 1992). The rapid radiation is thought to have contributed to the mitochondrial-nuclear discordance observed between phylogenies using genes from each genome (Kutschera *et al.* 2014; Nakagome *et al.* 2008; Yu *et al.* 2004). The primary discordance is between the placement of the Asiatic and American black bears. In the mitochondrial phylogeny, the American black is sister to the Asiatic black, where the sun bear is sister to the black bear clade (Figure 1.2A). However, in the nuclear genome, the American black bear is sister to the brown-polar clade, and Asiatic black, sun, and sloth form a second clade (Figure 1.2B). The nuclear phylogeny has greater concordance with the biogeography of the bears as it groups Asian and distantly Eurasian species together.

Below I summarize the evolutionary history and conservation status for each of the six *Ursus* species with particular attention on the American black bear, the subject of this dissertation.

Sloth bear

Evolution- The sloth bear ranges across the central and southern extents of the Indian subcontinent and Sri Lanka. This tropical distribution is poorly suited to the preservation of fossils, hence the evolutionary history of the sloth bear has been under studied. The species most likely diverged from *U. minimus* following its expansion into Asia from Eurasia (McLellan & Reiner 1992). The mainland (*U. ursinus ursinus*) and Sri Lankan (*U. u. inornatus*) populations have been split into two subspecies; however, no genetic studies have estimated divergence age, genetic diversity, or adaptation.

Conservation- Sloth bears share all of the conservation challenges of Asian bear species, particularly increasing habitat destruction, habitat fragmentation, killing to prevent depredation or agricultural damage, and poaching. Dwindling population size due to decreasing habitat led researchers to conclude that sloth bears have been extirpated in Bangladesh (Islam *et al.* 2013). While all four species of Asian bears are poached for viscera, bile, and meat, an analysis of trade from 2000 to 2011 observed that sloth bears represented the smallest fraction of this trade (Burgress *et al.* 2014). However, sloth bear cubs are also poached for a trade of live individuals trained to perform with the nomadic Kalandar ethnic group in India, although surveys since the mid-1990s indicate a decline in the number of dancing bears (D'Cruze *et al.* 2011). The IUCN Red List classifies sloth bears as vulnerable to extinction, with estimated population sizes less than 20,000 individuals (Garshelis *et al.* 2008b). Sloth bears are a CITES (Convention on International Trade of Endangered Species) Appendix I species meaning that international trade is prohibited.

Sun bear

Evolution- The sun bear ranges from southeast Asia and along a portion of the Indonesian archipelago (Figure 1.1). Similar to the sloth bear, divergence of the sun bear occurred following movement of *U. minimus* into southern Asia. Based on cranial morphometrics and body size, two subspecies have been proposed: *U. malayanus malayanus* on the mainland and Sumatra, and *U. m. euryspilus* on Borneo (Meijaard 2004). Few genetic studies have been conducted with sun bears to understand population, and potentially subspecies level, differentiation or genetic diversity. One study observed two mitochondrial clades; however, samples were from confiscated illegal harvests on Borneo and natal locations were unknown (Onuma *et al.* 2006). This resulted in two hypotheses; first, that sun bears on Borneo currently represent two clades, and second, that some of the confiscated animals were brought to Borneo from the mainland which is the natal location of the second mitochondrial clade.

Conservation- The IUCN Red List classifies sun bears as vulnerable to extinction; while a population size estimate is not available, the species has been estimated to have declined 30% since the 1980s (Fredriksson *et al.* 2008). Sun bears are threatened by both deforestation and poaching (Foley *et al.* 2011). Sun bear viscera has been confiscated in China and Thailand, which border range countries, but also internationally in North America, Europe, and Australia indicating broad international trade networks for this species (Burgress *et al.* 2014; Shepherd & Nijman 2008). Sun bears are a CITES Appendix I species.

Asiatic black bear

Evolution- Similar to divergence of the sloth and sun bears, the Asiatic black bear was also thought to have diverged from *U. minimus* in Asia (McLellan & Reiner 1992). In genetic studies with nuclear intron sequences and Y-chromosome haplotypes, extensive incomplete lineage sorting (ILS) was observed in the three Asian species (Kutschera *et al.* 2014). The genetic data supports the biogeography from the fossil record; specifically, ILS may be prevalent due to the three extant bears sharing a recent common ancestor. Mitochondrial genomes place the Asiatic and American black bears as sister taxa (Figure 1.2A) indicating mito-nuclear discordance within *Ursus*. Kutschera *et al.* (2014) proposed two hypotheses for mito-nuclear discordance in black bears; first, an Asiatic mitochondrial genome introgressed into the ancestors that expanded into North America so that American black bears have mitochondrial genomes more similar to Asiatic black bears. Second, ancient hybridization of brown bears into American black bears results in the nuclear genome being placed as sister to the brown-polar clade instead of a closer relationship with the other Asian species. The addition of more within species lineages may elucidate these relationships further. There are seven named subspecies of Asiatic black bears, defined by country of origin, body size, coat color, and mitochondrial sequence, including: *U. thibetanus formosanus* in Taiwan, *U. t. gedrosianus* in Pakistan, *U. t. japonicus* in Japan, *U. t. laniger* in the western Himalyan Mountains, *U. t. mupinensis* in the eastern Himalayan Mountains, *U. t. thibetanus* in southeastern Asia, and *U. t. ussuricus* in southern Russia, China, and Korea. Choi *et al.* (2010) presented a phylogeny with four of the seven subspecies, I extended this to hypothesize that *U. t. japonicus* was the earliest diverging lineage (Chapter 3) although full mitogenomes are

needed for several subspecies to fully understand phylogenetic relationships and mitochondrial diversity (Yasukochi *et al.* 2009).

Conservation- The IUCN Red List classifies Asiatic black bears as vulnerable to the threat of extinction, noting declining population sizes throughout the range (Garshelis & Steinmetz 2008). Asiatic black bears have the same conservation challenges as the sun and sloth bears. Populations have declined due to habitat reduction and fragmentation (Islam *et al.* 2013). Additionally, Asiatic black bears are killed because of their status as an agricultural pest (Ngoprasert *et al.* 2011; Sakurai *et al.* 2013), for trade in bear viscera, or for bear farming (Burgress *et al.* 2014; Peppin *et al.* 2008). Asiatic black bears are a CITES Appendix I species.

Brown bear

Evolution- The brown bear shared a most recent common ancestor with the cave bear about 1.2 - 1.6 Mya (Bon *et al.* 2008; Knapp *et al.* 2009). The cave bear may be several different species (*U. spelaeus*, *U. ingressus*, and *U. deningeri*) or a single species with spatially distributed mitochondrial clades as observed in other bears (Stiller *et al.* 2014). Cave bear populations began declining approximately 50 kya (Stiller *et al.* 2010) where the youngest fossils were found in the eastern (e.g. central Europe) portion of the range and suggest western populations (which extended throughout contemporary Russia) were extirpated first (Stiller *et al.* 2014). In contrast, European brown bear population size was stable throughout the Pleistocene (Stiller *et al.* 2010) and predicted to have begun

expanding its range throughout Europe and Eurasia before the extinction of cave bears (Saarma *et al.* 2007).

Brown bears occupy the largest geographic area (Figure 1.1) and have spatially distributed genetic diversity throughout. There are six brown bear mitochondrial clades that began diversifying 263 kya out of Europe (Davison *et al.* 2011). Two clades dispersed towards the south where clade 6 or *U. arctos syriacus* is now found in the Caucasus Mountains and towards the western Himalayan Mountains, and clade 5 includes *U. a. pruinosus* and *U. a. gobiensis*/*U. a. isabellinus* along the Tibetan plateau and Gobi desert (Galbreath *et al.* 2007). Clades 5 and 6 are sister to clades 3 and 4 which spread northward across Eurasia (Hirata *et al.* 2013; Saarma *et al.* 2007). Clade 4 is found in northern Japan (*U. a. lasiotus*) and in North America (*U. a. horribillis*). In North America, clade 4 was south of the glacial ice sheets during the last glacial maximum (Barnes *et al.* 2002). Clade 3 is distributed throughout Siberia (*U. a. collaris*), the Kamchatka peninsula of Russia and Japan (*U. a. beringianus*), and into Alaska (*U. a. middendorffi*). Sister to this large and geographically distributed clade, is the branch with clades 1 and 2. Clade 1 (*U. a. arctos*) is found in Europe. Clade 2 is found on the ABC Islands of Alaska in North America and in polar bears (discussed below).

Conservation- Although the IUCN Red List classifies brown bears as least concern (McLellan *et al.* 2008), their conservation status may be the most dynamic across its range. In Europe and North America, historic persecution reduced population size numbers so that animals contracted into isolated habitat patches (Chapron *et al.* 2014;

Frosch *et al.* 2014). In North America, the southern Rocky Mountain populations of brown bears were extirpated in the mid-1900s (Miller *et al.* 2006) and the species was listed as threatened under the precursor to the US Endangered Species Act in 1967. Brown bears in the Himalayas, Tibetan plateau, and Gobi desert are all isolated and have small population sizes (McCarthy *et al.* 2009). Urbanization across the brown bear range has contributed to limited recovery in some areas (Mowat *et al.* 2013; Schwartz *et al.* 2010). Climate change has been implicated as a factor limiting population recovery of brown bears in North America. As an omnivorous species, acorn mast may comprise a large portion of brown bear diets (Bojarska & Selva 2012); however, warmer winters prevent freezing of mountain pine beetles on white pine (Raffa *et al.* 2013). The beetles, in conjunction with rusts and abiotic changes, limit mast production, thereby reducing food for brown bears (Costello *et al.* 2014; Tomback & Achuff 2010). Finally, as with all Asian species, brown bears are frequently harvested for the trade in bear viscera particularly along the Russia-China border (Burgress *et al.* 2014). Through most of their range brown bears are a CITES Appendix II species meaning that commercial trade is regulated by CITES; however, in brown bears are Appendix I in Bhutan, China, Mongolia, and Mexico, although the Mexican population was extirpated (Miller *et al.* 2006).

Polar bear

Evolution- The evolution of polar bears is a particularly contentious research topic, specifically dating of the divergence from brown bears. The crux of the debate centers on the placement of the mitochondrial haplotype, clade 2, which is nested within clade 1 of

the brown bear mitochondrial phylogeny (Edwards *et al.* 2011). Mitochondrial clade 2 is also found in a brown bear population on the ABC Islands of Alaska but not in any other brown bear population (Barnes *et al.* 2002; Waits *et al.* 1998). Several research groups posited that clade 2 was associated with brown bears and introgressed into polar bears near the ABC Islands before fixing within *U. maritimus* (Hailer *et al.* 2012; Lindqvist *et al.* 2010; Miller *et al.* 2012). However, Cahill *et al.* (2013) proposed an alternative hypothesis: a population of polar bears became isolated on the ABC Islands and male mediated brown bear introgression into this population explains both polar bears and ABC Island brown bears sharing clade 2 mitochondrial haplotypes. This hypothesis is more parsimonious with the biogeographic distribution of clade 2 allowing circumpolar movement of the clade, followed by the formation of a single North American population instead of extirpation of clade 2 throughout a Eurasian range expansion (Davison *et al.* 2011).

The divergence date between polar and brown bears is important for understanding the rate of adaptation to different climatic environments. Polar bears have several adaptations including white fur, lower center of gravity, larger paws, rougher paw pads for gripping ice, and the most carnivorous diet within the bears (Liu *et al.* 2014; Miller *et al.* 2012; Slater *et al.* 2010; Stirling 2011). A recent divergence date would suggest that mammals will rapidly adapt to climate change, whereas an older divergence date would suggest that adaptation is slower. Estimates for divergence between mitochondrial clades 1 and 2 vary from 88 to 152 kya (Krause *et al.* 2008; Lindqvist *et al.* 2010); while estimates in the nuclear genome vary: 343-479 kya (Liu *et al.* 2014), 603 kya (Hailer *et*

al. 2012), and 4-5 Mya (Miller *et al.* 2012). Polar bear divergence timing remains an open question as these estimates contradict one another.

Conservation- The IUCN Red List classifies polar bears as vulnerable to extinction with decreasing populations across the range (Schliebe *et al.* 2008). The unique biology of polar bears contributes to the challenge of their conservation. Polar bears are the most carnivorous Ursidae species where seal, whale, and walrus blubber comprise the largest portion of their diet followed by fish, ungulates, berries, and birds when available (Gormezano & Rockwell 2013; Rode *et al.* 2010; Thiemann *et al.* 2008). Polar bears hunt for these high fat species when sea ice is at its greatest extent during North American winters, as prey species gather for breeding. Unlike the hibernating American black and brown bears, polar bears hunt in the winter and have a low energetic state in the summer, living off their fat stores (Rode *et al.* 2010). This biology partially informs how the species has come to symbolize a broader political debate regarding climate change. Weak and disjunct sea ice limits polar bears' access to prey which for some individuals means they do not store enough fat to survive summer, or drown while hunting off of floes.

The other significant conservation challenge for polar bears is their distribution across the Arctic where five countries (the United States, Canada, Denmark, Norway, and Russia) all have territory. While polar bears do have population structure around Hudson Bay, movement of individuals between the political borders is not impeded (Peacock *et al.* 2015). Additionally, several indigenous tribes traditionally hunted polar bears for meat

and skins; thus legal harvesting agreements with these tribes has been negotiated. The five nations within the polar bear range signed the Agreement on the Conservation of Polar Bears in 1973 to limit unregulated trophy hunting and other human activities which had reduced population sizes. Polar bears were listed as threatened under the US Endangered Species Act in 2008 and are a CITES Appendix II species.

American black bear

Evolution- The American black bear evolved in North America following expansion of *U. abstrusus* onto the continent across an Asian land bridge (Kurten & Anderson 1980). Previous work based on morphology and geography designated 16 subspecies including: *U. americanus altifrontalis* in British Columbia west of the Rocky Mountains, *U. a. amblyceps* in the southern Rocky Mountains, *U. a. americanus* found throughout the east, central Canada, and Alaska, *U. a. californiensis* in California, *U. a. carlottae* on the near shore islands of British Columbia, *U. a. cinnamomum* in the northern Rocky Mountains, *U. a. emmonsii* in Yakutat Bay, Alaska, *U. a. eremicus* in Sierra Madre Oriental Mountains, *U. a. floridanus* in Florida, *U. a. hamiltoni* on Newfoundland Island, *U. a. kermodei* along a coastal and near shore island patch of British Columbia, *U. a. luteolus* in Louisiana, *U. a. machetes* in Sierra Madre Occidental Mountains, *U. a. perniger* on the Kenai peninsula of Alaska, *U. a. pugnax* on the southern islands of Alaska, and *U. a. vancouveri* on Vancouver Island (Lariviere 2001). Three mitochondrial lineages have been described, A-east in the eastern portion of the range, A-west in the western range, and B along the Pacific coast from California to northern British Columbia (Wooding &

Ward 1997). In Chapter 3, I provide the most extensive analysis of nuclear and mitochondrial genetic diversity of the species to date.

Conservation- The IUCN Red List classifies American black bears as least concern with a population size greater than 1 million animals and either stable or increasing across most of the range (Garshelis *et al.* 2008a). Despite the large population size, the species is still listed as CITES Appendix II due to the similarity of appearance with other bear species and high levels of demand for bear viscera in those species. Due to local small population sizes, *U. a. eremicus* and *U. a. machetes* were listed as endangered in Mexico in 1985, *U. a. luteolus* was placed on the threatened list under the US Endangered Species Act in 1992, and *U. a. floridanus* was placed on the Florida threatened list in 1974 but delisted in 2012 (FFWCC 2012). Finally, *U. a. kermodei* individuals with white pelage have been protected from hunting since the early 1900s (Marshall & Ritland 2002). Much of the current conservation concern with American black bears focuses on human-bear conflict management particularly as development encroaches on habitat. In the same vein, as carnivore persecution has decreased, local populations have grown and expanded their range (Scheick & McCown 2014), again bringing bears and humans into broader contact.

Questions answered in this dissertation

This dissertation focuses on the evolution and conservation of the American black bear. In Chapter 2, I focus on the regional population structure of black bears in the Central Interior Highlands of North America which include Oklahoma, Arkansas, and Missouri.

Following purported extirpation in early 1900s, bears were translocated to the Ozark and Ouachita Mountains of Arkansas from Minnesota, USA and Manitoba, Canada (Clark & Smith 1994). I investigated the effect of the translocation on the contemporary genotypes and observed first, a signature of admixture between reintroduced bears and genotypes on the landscape at the time of translocation, and secondly, strong genetic drift between the Ozark and Ouachita populations.

In Chapter 3, I analyze the first range wide phylogeographic dataset using nuclear genomic markers for American black bears. I first describe patterns of nuclear and mitochondrial divergence in time and space, then relate the current distribution of genetic lineages and nested clusters to range expansion patterns out of glacial refugia. This provides the most comprehensive evaluation of the evolutionary history within the species and opens the door for new analyses into adaptive variation across the landscape accounting for underlying population structure. The conservation implications of this work include a call to reevaluate subspecies designations.

In Chapter 4, I utilize the range wide data from Chapter 3 for an analysis of spatial assignment of samples. Spatial assignment methods are of great interest to the conservation community for the role they may play for understanding the sources of illegal trade and distance to market, which may elucidate trade routes (Banks *et al.* 2015). I compare how varying marker type, loci number, and training sample number affect the accuracy and precision of natal area estimates. I observed that dense SNP marker sets may overcome small sample sizes; however, microsatellite marker sets uniformly

performed poorly which is concerning given that they are the genetic marker of choice for most applied conservation studies (Shafer *et al.* 2015). This research implies that extending these methods to the heavily traded Asian bear species should begin by building detailed range wide SNP genotyping panels with greater spatial resolution. This data may also contribute to a better understanding of the evolution of those species, as the work in this dissertation draws on the same datasets for insight into both the evolution and conservation of the species.

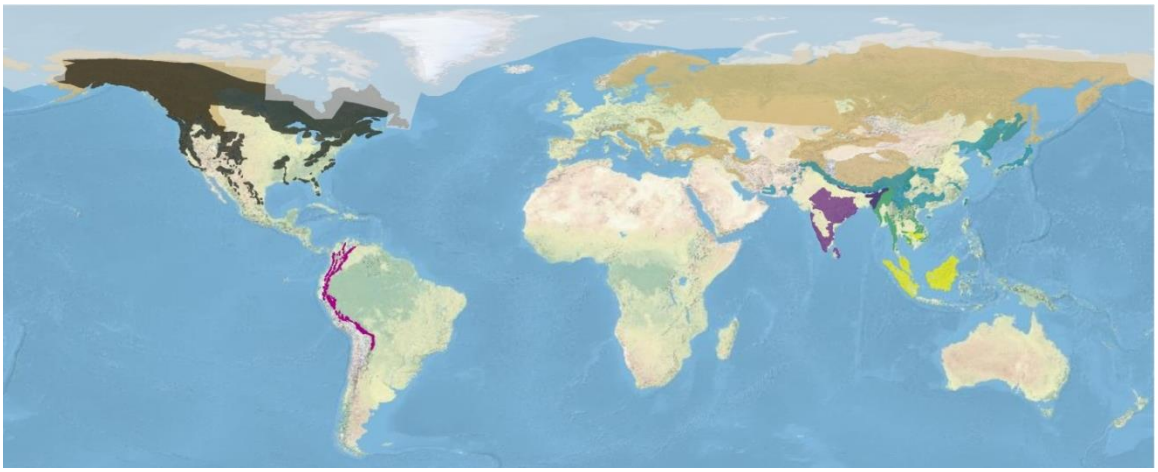

Figure 1.1- Geographic distribution of the eight extant Ursidae species, including: the giant panda (*Ailuropoda melanoleuca*, orange and not visible due to small area in China), spectacled bear (*Tremarctos ornatus*, magenta), sloth bear (*Ursus ursinus*, purple), sun bear (*U. malayanus*, yellow), Asiatic black bear (*U. thibetanus*, blue), American black bear (*U. americanus*, black), brown bear (*U. arctos*, brown), and polar bear (*U. maritimus*, white).

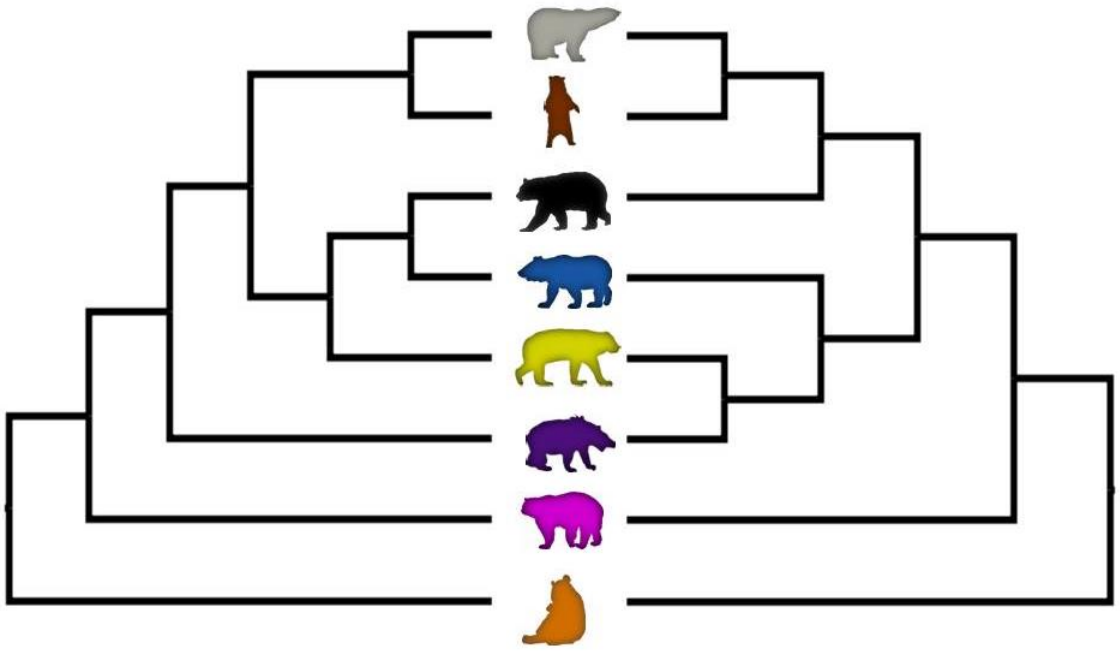

Figure 1.2- The mitochondrial (left) and nuclear (right) phylogenies for the Ursidae. The bear pictograms were colored similarly to the range map in Figure 1.1, including: the giant panda (*Ailuropoda melanoleuca*, orange), spectacled bear (*Tremarctos ornatus*, magenta), sloth bear (*Ursus ursinus*, purple), sun bear (*U. malayanus*, yellow), Asiatic black bear (*U. thibetanus*, blue), American black bear (*U. americanus*, black), brown bear (*U. arctos*, brown), and polar bear (*U. maritimus*, grey).

LITERATURE CITED

Banks NC, Paini DR, Bayliss KL, Hodda M (2015) The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecology Letters* **18**, 188-199.

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* **295**, 2267-2270.

Bojarska K, Selva N (2012) Spatial patterns in brown bear Ursus arctos diet: the role of geographical and environmental factors. *Mammal Review* **42**, 120-143.

Bon C, Caudy N, de Dieuleveult M, *et al.* (2008) Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17447-17452.

Burgress EA, Stoner SS, Foley K-E (2014) Brought to bear: An analysis of seizures across Asia (2000-2011), p. 46. TRAFFIC, Petaling Jaya, Selangor, Malaysia.

Cahill JA, Green RE, Fulton TL, *et al.* (2013) Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLoS Genet* **9**, e1003345.

Chapron G, Kaczensky P, Linnell JDC, *et al.* (2014) Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science* **346**, 1517-1519.

Choi EH, Kim SK, Ryu SH, Jang KH, Hwang UW (2010) Mitochondrial genome phylogeny among Asiatic black bear *Ursus thibetanus* subspecies and comprehensive analysis of their control regions. *Mitochondrial DNA* **21**, 105-114.

Clark JD, Smith KG (1994) A demographic comparison of two black bear populations in the Interior Highlands of Arkansas *Wildlife Society Bulletin* **22**, 593-603.

Costello CM, van Manen FT, Haroldson MA, *et al.* (2014) Influence of whitebark pine decline on fall habitat use and movements of grizzly bears in the Greater Yellowstone Ecosystem. *Ecology and Evolution* **4**, 2004-2018.

D'Cruze N, Sarma UK, Mookerjee A, *et al.* (2011) Dancing bears in India: A sloth bear status report. *Ursus* **22**, 99-105.

Davison J, Ho SYW, Bray SC, *et al.* (2011) Late-Quaternary biogeographic scenarios for the brown bear (*Ursus arctos*), a wild mammal model species. *Quaternary Science Reviews* **30**, 418-430.

Edwards Ceiridwen J, Suchard Marc A, Lemey P, *et al.* (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. *Current Biology* **21**, 1251-1258.

FFWCC (2012) Florida Black Bear Management Plan: *Urus americanus floridanus*, p. 215. Florida Fish and Wildlife Conservation Commission, Tallahassee, Florida.

Foley K-E, Stengel CJ, Shepherd CR (2011) Pills, powders, vials, and flakes: The bear bile trade in Asia. TRAFFIC Southeast Asia, Petaling Jaya, Selangor, Malaysia.

Fredriksson G, Steinmetz R, Wong S, Garshelis DL (2008) *Helarctos malayanus*. In: *The IUCN Red List of Threatened Species. Version 2014.3*.

Frosch C, Dutsov A, Zlatanova D, *et al.* (2014) Noninvasive genetic assessment of brown bear population structure in Bulgarian mountain regions. *Mammalian Biology - Zeitschrift für Säugetierkunde* **79**, 268-276.

Galbreath GJ, Groves CP, Waits LP (2007) Genetic resolution of composition and phylogenetic placement of the isabelline bear. *Ursus* **18**, 129-131.

Garshelis DL, Crider D, Van Manen FT (2008a) *Ursus americanus*. In: *The IUCN Red List of Threatened Species. Version 2014.3*.

Garshelis DL, Ratnayeke S, Chauhan NPS (2008b) *Melursus ursinus*. In: *The IUCN Red List of Threatened Species. Version 2014.3*.

Garshelis DL, Steinmetz R (2008) *Ursus thibetanus*. In: *The IUCN Red List of Threatened Species. Version 2014.3*.

Gormezano LJ, Rockwell RF (2013) What to eat now? Shifts in polar bear diet during the ice-free season in western Hudson Bay. *Ecology and Evolution* **3**, 3509-3523.

Hailer F, Kutschera VE, Hallström BM, *et al.* (2012) Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science* **336**, 344-347.

Hirata D, Mano T, Abramov AV, *et al.* (2013) Molecular phylogeography of the brown bear (*Ursus arctos*) in northeastern Asia based on analyses of complete mitochondrial DNA sequences. *Molecular Biology and Evolution* **30**, 1644-1652.

Islam MA, Uddin M, Aziz MA, *et al.* (2013) Status of bears in Bangladesh: going, going, gone? *Ursus* **24**, 83-90.

Knapp M, Rohland N, Weinstock J, *et al.* (2009) First DNA sequences from Asian cave bear fossils reveal deep divergences and complex phylogeographic patterns. *Molecular Ecology* **18**, 1225-1238.

Krause J, Unger T, Nocon A, *et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *Bmc Evolutionary Biology* **8**.

Kurten B, Anderson E (1980) *Pleistocene Mammals of North America* Columbia University Press, New York.

Kutschera VE, Bidon T, Hailer F, *et al.* (2014) Bears in a forest of gene trees: Phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Molecular Biology and Evolution* **31**, 2004-2017.

Lariviere S (2001) *Ursus americanus*. *Mammalian Species* **647**, 1-11.

Lindqvist C, Schuster SC, Sun Y, *et al.* (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences*.

Liu S, Lorenzen Eline D, Fumagalli M, *et al.* (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* **157**, 785-794.

Marshall HD, Ritland K (2002) Genetic diversity and differentiation of Kermode bear populations. *Molecular Ecology* **11**, 685-697.

McCarthy TM, Waits LP, Mijiddorj B (2009) Status of the Gobi bear in Mongolia as determined by noninvasive genetic methods. *Ursus* **20**, 30-38.

McLellan B, Reiner DC (1992) A review of bear evolution **9**, 85-96.

McLellan BN, Servheen C, Huber D (2008) *Ursus arctos*. In: *The IUCN Red List of Threatened Species. Version 2014.3*.

Meijaard E (2004) Craniometric differences among Malayan sun bears (*Ursus malayanus*); Evolutionary and taxonomic implications. *Raffles Bulletin of Zoology* **52**, 665-672.

Miller CR, Waits LP, Joyce P (2006) Phylogeography and mitochondrial diversity of extirpated brown bear (*Ursus arctos*) populations in the contiguous United States and Mexico. *Molecular Ecology* **15**, 4477-4485.

Miller W, Schuster SC, Welch AJ, *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences* **109**, E2382–E2390.

Mowat G, Heard DC, Schwarz CJ (2013) Predicting grizzly bear density in western North America. *Plos One* **8**, e82757.

Nakagome S, Pecon-Slattery J, Masuda R (2008) Unequal rates of Y chromosome gene divergence during speciation of the family Ursidae. *Molecular Biology and Evolution* **25**, 1344-1356.

Ngoprasert D, Steinmetz R, Reed DH, *et al.* (2011) Influence of fruit on habitat selection of Asian bears in a tropical forest. *The Journal of Wildlife Management* **75**, 588-595.

Onuma M, Suzuki M, Ohtaishi N (2006) Possible conservation units of the sun bear (*Helarctos malayanus*) in Sarawak based on variation of mtDNA control region. *Japanese Journal of Veterinary Research* **54**, 135-139.

Peacock E, Sonsthagen SA, Obbard ME, *et al.* (2015) Implications of the circumpolar genetic structure of polar bears for their conservation in a rapidly warming Arctic. *Plos One* **10**, e112021.

Peppin L, McEwing R, Carvalho GR, Ogden R (2008) A DNA-based approach for the forensic identification of Asiatic black bear (*Ursus thibetanus*) in a traditional Asian medicine. *Journal of Forensic Sciences* **53**, 1358-1362.

Raffa KF, Powell EN, Townsend PA (2013) Temperature-driven range expansion of an irruptive insect heightened by weakly coevolved plant defenses. *Proceedings of the National Academy of Sciences* **110**, 2193-2198.

Rode KD, Reist JD, Peacock E, Stirling I (2010) *Comments in response to "Estimating the energetic contribution of polar bear (Ursus maritimus) summer diets to the total energy budget" by Dyck and Kebreab (2009)*.

Saarma U, Ho SYW, Pybus OG, *et al.* (2007) Mitogenetic structure of brown bears (*Ursus arctos* L.) in northeastern Europe and a new time frame for the formation of European brown bear lineages. *Molecular Ecology* **16**, 401-413.

Sakurai R, Jacobson SK, Carlton JS (2013) Media coverage of management of the black bear Ursus thibetanus in Japan. *Oryx* **47**, 519-525.

Scheick BK, McCown W (2014) Geographic distribution of American black bears in North America. *Ursus*, 24-33.

Schliebe S, Wiig O, Derocher AE, Lunn N (2008) Ursus maritimus. In: *The IUCN Red List of Threatened Species. Version 2014.3*.

Schwartz CC, Haroldson MA, White GC (2010) Hazards affecting grizzly bear survival in the Greater Yellowstone Ecosystem. *Journal of Wildlife Management* **74**, 654-667.

Shafer ABA, Wolf JBW, Alves PC, *et al.* (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**, 78-87.

Shepherd C, Nijman V (2008) The trade in bear parts from Myanmar: an illustration of the ineffectiveness of enforcement of international wildlife trade regulations. *Biodiversity and Conservation* **17**, 35-42.

Slater GJ, Figueirido B, Louis L, Yang P, Van Valkenburgh B (2010) Biomechanical consequences of rapid evolution in the polar bear lineage. *Plos One* **5**, e13870.

Stiller M, Baryshnikov G, Bocherens H, *et al.* (2010) Withering away-25,000 years of genetic decline preceded cave bear extinction. *Molecular Biology and Evolution* **27**, 975-978.

Stiller M, Molak M, Prost S, *et al.* (2014) Mitochondrial DNA diversity and evolution of the Pleistocene cave bear complex. *Quaternary International* **339–340**, 224-231.

Stirling I (2011) *Polar Bears: The Natural History of a Threatened Species* Fitzhenry and Whiteside, Brighton, MA.

Thiemann GW, Iverson SJ, Stirling I (2008) Polar bear diets and arctic marine food webs: Insights from fatty acid analysis. *Ecological Monographs* **78**, 591-613.

Tomback DF, Achuff P (2010) Blister rust and western forest biodiversity: ecology, values and outlook for white pines. *Forest Pathology* **40**, 186-225.

Waits LP, Talbot SL, Ward RH, Shields GF (1998) Mitochondrial DNA phylogeography of the North American brown bear and implications for conservation. *Conservation Biology* **12**, 408-417.

Wooding S, Ward R (1997) Phylogeography and Pleistocene evolution in the North American black bear. *Molecular Biology and Evolution* **14**, 1096-1105.

Yasukochi Y, Nishida S, Han S-H, *et al.* (2009) Genetic structure of the Asiatic black bear in Japan using mitochondrial DNA analysis. *Journal of Heredity* **100**, 297-308.

Yu L, Li QW, Ryder OA, Zhang YP (2004) Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. *Molecular Phylogenetics and Evolution* **32**, 480-494.

CHAPTER 2

Influence of drift and admixture on population structure of American black bears (*Ursus americanus*) in the Central Interior Highlands, USA 50 years after translocation

Emily E. Puckett[1], Thea V. Kristensen[2], Clay M.Wilton[3], Sara B. Lyda[4], Karen V. Noyce[5], Paula M. Holahan[6], David M. Leslie, Jr.[7], Jeff Beringer[8], Jerrold L. Belant[3], Don White, Jr.[9], and Lori S. Eggert[1]

[1]- Division of Biological Sciences, 226 Tucker Hall, University of Missouri, Columbia, MO 65211
[2]- Department of Biological Sciences, Science and Engineering 601, University of Arkansas, Fayetteville, AR 72701
[3]- Carnivore Ecology Laboratory, Forest and Wildlife Research Center, Mississippi State University, Mississippi State, MS 39762
[4]- Oklahoma Cooperative Fish and Wildlife Research Unit, Department of Natural Resource Ecology and Management, 007 Agriculture Hall, Oklahoma State University, Stillwater, OK 74078
[5]- Forest Wildlife Populations and Research Group, Minnesota Department of Natural Resources, 1201 East Highway 2, Grand Rapids, MN 55744
[6]- University of Wisconsin Zoological Museum, 250 North Mills Street, Madison, WI 53706
[7]– U.S. Geological Survey, Oklahoma Cooperative Fish and Wildlife Research Unit, Department of Natural Resource Ecology and Management, 007 Agriculture Hall, Oklahoma State University, Stillwater, OK 74078
[8]- Missouri Department of Conservation, Resource Science Center, 3500 Gans Creed Road, Columbia, MO 65201
[9]- University of Arkansas Agricultural Experiment Station, Arkansas Forest Resources Center,110 University Court, University of Arkansas-Monticello, Monticello, AR 71656

ABSTRACT

Bottlenecks, founder events, and genetic drift often result in decreased genetic diversity and increased population differentiation. These events may follow abundance declines due to natural or anthropogenic perturbations, where translocations may be an effective conservation strategy to increase population size. American black bears (*Ursus americanus*) were nearly extirpated from the Central Interior Highlands, USA by 1920. In an effort to restore bears, 254 individuals were translocated from Minnesota, USA and

Manitoba, Canada, into the Ouachita and Ozark Mountains from 1958 to 1968. Using 15 microsatellites and mitochondrial haplotypes, we observed contemporary genetic diversity and differentiation between the source and supplemented populations. We inferred four genetic clusters: Source, Ouachitas, Ozarks, and a cluster in Missouri where no individuals were translocated. Coalescent models using approximate Bayesian computation identified an admixture model as having the highest posterior probability (0.942) over models where the translocation was unsuccessful or acted as a founder event. Nuclear genetic diversity was highest in the source (A_R = 9.11) and significantly lower in the translocated populations (A_R = 7.07 - 7.34; P = 0.004). The Missouri cluster had the lowest genetic diversity (A_R = 5.48) and served as a natural experiment showing the utility of translocations to increase genetic diversity following demographic bottlenecks. Differentiation was greater between the two admixed populations than either compared to the source, suggesting that genetic drift acted strongly over the eight generations since the translocation. The Ouachitas and Missouri were previously hypothesized to be remnant lineages. We observed a pre-translocation remnant signature in Missouri but not in the Ouachitas.

INTRODUCTION

In North America during the 1800s, human population expansion resulted in local extirpation of many native vertebrates (Laliberte & Ripple 2004). Declining populations resulted from direct effects such as harvest for meat, fat, skins, and predator control (Larson *et al.* 2002; Sobey 2007), and indirect effects including habitat destruction and fragmentation, changes in food resources, and introduction of diseases (Baker *et al.* 2008;

Miller & Cully 2001). Translocation is a conservation management tool that may be used to supplement (Grenier *et al.* 2007; Larson *et al.* 2002) or reestablish (Van Houtan *et al.* 2009; Williams & Scribner 2010) populations following declines and have also been used to increase genetic diversity of isolated populations (Johnson *et al.* 2010). Translocations increase populations by increasing the number of reproductively mature adults, especially females that may increase the number of births in a population.

Genetic drift may rapidly decrease genetic diversity (Nei & Tajima 1981) following abundance declines or founder events (Nei *et al.* 1975). Translocation events may act as bottlenecks, capturing a subset of the allelic diversity of the source population resulting in the loss of low frequency alleles (Garza & Williamson 2001). Decreased allelic richness has been observed in numerous studies of translocated birds and mammals (Groombridge *et al.* 2012; O'Ryan *et al.* 1998; Pruett & Winker 2005), although researchers have also observed no decrease in richness when the source population has few alleles (Hicks *et al.* 2007; Larson *et al.* 2002; Taylor & Jamieson 2008) or the number of founding individuals was large (Miller *et al.* 2009). Slow population growth following a translocation may also contribute to loss of genetic diversity, although supplementing a population with individuals over multiple years may decrease loss of genetic diversity compared with single translocation events (Wisely *et al.* 2008).

During the nineteenth and early twentieth centuries, the American black bear (*Ursus americanus*) experienced both overall range contraction and local extirpation (Lackey *et al.* 2013; Laliberte & Ripple 2004; Smith & Clark 1994; Sobey 2007) following human

westward expansion. Currently in the Central Interior Highlands, USA (hereafter CIH), black bears occur in most forested habitats including the Ouachita (OU) and Ozark (OZ) Mountains (Figure 2.1). The Ouachitas span Arkansas (AR) and Oklahoma (OK), while the Ozarks cross AR, OK, and Missouri (MO). Both OU and OZ populations were considered extirpated by 1920 (Smith & Clark 1994). The CIH is isolated from other suitable bear habitat which may have increased extirpation risk (Yackulic *et al.* 2011). However, reports after 1920 document a limited number of bear sightings in OZ including a bear shot in 1931 and a cub shot in 1950 suggesting both low density and breeding during this time and not extirpation (Beringer *et al.* 2008).

In an effort to increase Arkansas' bear population, the Arkansas Game and Fish Commission translocated 254 bears from Minnesota, USA (MN) and Manitoba, Canada (MB) from 1958 to 1968 and released them within AR in both OU and OZ (Smith & Clark 1994). Detailed records of the translocation were not recorded including information on the number of individuals translocated each year, the source and drop-off locations, and the sex and age of translocated individuals. The AR translocation was successful as the population size increased (up to 3,000 bears in 2013; M Means, personal communication) in part due to the three factors associated with successful reintroductions: use of a large number of animals, use of wild caught individuals, and reduction in human persecution as a factor in the original decline (Fischer & Lindenmayer 2000). Additionally, as population size increased in core areas of the translocation, bears dispersed and recolonized former habitat throughout the CIH, a trend observed throughout the range (Pelton & Van Manen 1994). Although objective

population estimates of bears from OK are not available, their distribution and presumably numbers are expanding within the state as of 2013 (Bales *et al.* 2005). In Missouri, public sightings of bears have increased annually since 1987 (Beringer *et al.* 2008) and a recent genetic mark-recapture survey estimated population size between 250 to 300 individuals within the known breeding range (CM Wilton, personal communication).

Due to the history of extirpation and reintroduction, bears have been well studied in the CIH as managers measured population density, abundance, and growth rate. Abundance estimates using genetic mark-recapture were completed in OK-Ou, AR-Ou, AR-Oz, and MO (Brown 2007; Gardner-Santana 2007; Kristensen 2013). Additional studies focused on population differentiation to determine the effect of the translocation on CIH population genetics. Of particular interest was if CIH subpopulations were significantly differentiated from source populations and may represent remnant lineages suggesting that the demographic decline resulted in a bottleneck and not extirpation. Microsatellite studies arrived at different interpretations where Csiki *et al.* (2003) did not observe differentiation between OU, OZ, and MN and Warrillow *et al.* (2001) observed differentiation between OZ and MN, each suggesting a strong effect of the translocation on the CIH. Faries *et al.* (2013) added samples from MO and observed significant differentiation between all four areas and suggested the presence of remnant lineages in OU and MO. Van Den Bussche and colleagues (2009) also suggested the presence of a remnant lineage in OK-Ou due to exclusive geographic location of two mitochondrial haplotypes. Our study improves on previous work by increasing both sample size and

marker number thereby providing more power to detect differences between populations; additionally, we used a hypothesis testing framework which had not been applied to this system.

Bears in each of the five CIH areas investigated in this study could have been extirpated or experienced a bottleneck leaving remnant genotypes in the region. If bears were extirpated, the translocation would have been a founder event in both AR-Ou and AR-Oz. We tested three hypotheses about contemporary genetic diversity in the CIH. First, translocated bears did not survive to reproduce and contemporary genetic diversity is a product of CIH lineages. Second, the translocation resulted in admixture between CIH and source lineages. Third, before translocation, bears were extirpated so that the reintroduction was a founder event. Our objectives were to identify the population structure of bears in the CIH 50 years following translocation and test for signatures of remnant genetic lineages.

MATERIALS and METHODS

Sample collection

We obtained black bear samples from seven geographic locations (Figure 2.1, Table 2.1). In Minnesota (MN), state agents collected blood samples from denning bears. Twenty-nine samples collected from 2006 to 2013 and distributed throughout the range within MN were provided for analysis. Additionally, we obtained small amounts of tissue from eight black bear skulls harvested in MN in 1899 and 1900 from the University of Wisconsin Zoological Museum. Provincial wildlife agents collected tissue from hunter

harvested individuals in Manitoba (MB) in 2006 and provided nine samples. Hair samples from unique individuals identified during three genetic mark-recapture studies in the Arkansas Ouachita Mountains (AR-Ou; 2006-2008), Arkansas Ozark Mountains (AR-Oz; 2009-2011), and Missouri (MO; 2011-2012) were included in the analysis. Hair samples in Oklahoma from the Ouachita (OK-Ou; 2010-2011) and Ozark Mountains (OK-Oz; 2012) were collected by state personnel from hunter harvested individuals.

DNA extraction

DNA from blood samples was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) following the manufacturer's protocol. DNA from hair samples was extracted using InstaGene (BioRad, Laboratories, Hercules, CA) following the protocol of Eggert *et al.* (2005). DNA from tissue samples including the historic samples was extracted using a standard phenol-chloroform protocol (Sambrook *et al.* 1989). The historic samples were extracted in a separate lab, free of PCR products; all reactions using DNA from these samples employed filter tips and were not performed alongside contemporary samples.

Microsatellite genotyping

For genotyping modern samples, we optimized 15 microsatellite loci in two multiplex panels; the first panel contained markers G1A, G10B, G10C, G1D, G10L, G10M, and G10P (Paetkau *et al.* 1998) with alternative primer sets detailed in Kristensen *et al.* (2011). The second panel contained markers G10J, G10O, G10U (Kristensen *et al.* 2011; Paetkau *et al.* 1998); UarMU05, UarMU10, UarMU23, UarMU59 (Taberlet *et al.* 1997);

and UA-P2H03 (Sanderlin *et al.* 2009). Similar to Kristensen *et al*. (2011), we redesigned primer pairs to shorten the microsatellite to increase genotyping efficiency from potentially fragmented DNA obtained from hair samples (Table A1.1).

All PCR reactions were performed in a UV-sterilized hood. Multiplex genotyping reactions were performed in 8µL volumes with final concentrations of 1X Multiplex PCR Master Mix (Qiagen, Valencia, CA), 0.1 µM of each primer, and 15 ng DNA. The thermocycler settings were: 95°C for 15 min; 40 cycles of 94°C for 30 s, 57°C for 90 s, 72°C for 60 s; and 60°C for 30 min. Forward primers were fluorescently labeled with 6FAM, VIC, NED, or PET (Table A1.1). Each 96 well plate of samples also contained the same positive control (a bear tagged by state agents in MO) and a negative control, which were scored alongside samples for each of the three replicate PCR reactions per sample. Products were separated in an ABI 3730 DNA Analyzer at the University of Missouri DNA Core Facility (Columbia, MO) and scored against internal lane standards (Genescan LIZ 600) in GeneMarker v1.97 (SoftGenetics, State College, PA). Wildlife Genetics International (Nelson, British Columbia, CA) genotyped samples from AR-Ou at loci G1A, G10B, G10C, G1D, G10J, G10L, G10M, and G10P. Four samples were genotyped at the same loci at the University of Missouri to calibrate datasets for comparison.

Population genetic analyses

We removed samples from analyses if they genotyped at fewer than 12 loci. As samples from AR-Ou, AR-Oz, and MO were obtained from gridded hair snares which may

sample related individuals, we removed one individual from putative parent-offspring pairs identified in ML-Relate (Kalinowski *et al.* 2006). We used ARLEQUIN v3.5 (Excoffier & Lischer 2010) to test for deviations from expected heterozygosity values under Hardy-Weinberg Equilibrium (HWE; 10^6 steps in Markov chain and 10^5 dememorization steps) and linkage disequilibrium (LD; with 10^4 permutations) in each area. We applied a Bonferroni correction per locus ($\alpha < 0.003$) for HWE and over all 120 pairwise comparisons ($\alpha < 0.0004$) for LD. We used Micro-Checker (Van Oosterhout *et al.* 2004) to test for null alleles in each area.

We estimated expected (H_E) and observed (H_O) heterozygosity, mean number of alleles (A), inbreeding coefficients (F_{IS}) at the population and study area levels (20,000 permutations with R_{ST} distance matrix), and pairwise F_{ST} in ARLEQUIN v3.5 (Excoffier & Lischer 2010). Mean number of alleles adjusted for sample size (allelic richness, n = 38) was calculated using FSTAT v2.9.3.2 (Goudet 1995) for the four genetic clusters. We tested for differences in allelic richness using a Kruskal-Wallis rank test in R (R Core Team 2013) both globally and between all pairwise populations.

We analyzed population structure using the Bayesian assignment software STRUCTURE v2.3 (Pritchard *et al.* 2000). STRUCTURE assigns proportions of within sample genetic diversity to a priori selected genetic clusters (K). We applied the admixture model and correlated allele frequencies options. We performed 10 repetitions of each independent K value between one and eight; burn-in was set to 10^6 iterations, followed by sampling of 10^6 iterations. Models for the 10 technical replicates were averaged in CLUMPP v1.2

(Jakobsson & Rosenberg 2007) and displayed using DISTRUCT v1.1 (Rosenberg 2004). We interpreted successive levels of K within the biological and geographic context of our system. Additionally, we calculated the log probability of the data, LnP(K), and rate of change in the second order derivatives of the log probability between successive K values (ΔK) (Evanno *et al.* 2005) with the aid of STRUCTURE HARVESTER (Earl & vonHoldt 2012). To test for hierarchical substructure within the data, we analyzed each of the four clusters identified in separate STRUCTURE analyses again applying the admixture model and correlated allele frequencies options. We performed five repetitions of each K between one and three except for OZ where we also ran at K = 4; burn-in was set to $5x10^5$, followed by sampling of $1x10^6$ iterations. Run averaging, display, and summary statistics for each K value were determined as for the full analysis.

We estimated migration between OU and OZ using BayesAss v3.0 (Wilson & Rannala 2003). We averaged migration estimates from three repetitions each with a different starting seed, run for 10^7 iterations with the first 10^6 discarded as burn-in. We adjusted the migration, allele frequency, and inbreeding coefficient mixing parameters to 0.20; this resulted in acceptance rates of 0.21, 0.30, and 0.23 for those parameters respectively. Convergence was assessed using Tracer (Rambaut & Drummond 2009).

Mitochondrial control region sequencing and analysis

We redesigned primers (F: 5'-ATGAATCGGAGGRCAACCAG-3'; R: 5'-CATAGAAACCCCCACAYTTCA-3') to amplify 566bp of the 3' end of cytochrome *b* and the control region of the mitochondrion in black bears then trimmed the sequence to

the 556bp used in previous studies (Onorato *et al.* 2004; Van Den Bussche *et al.* 2009). We randomly chose a subset of samples from OU, OZ, and MO for haplotype sequencing while sequencing all samples from MB, MN, and historic MN; however, three samples from MN amplified poorly and were not sequenced. Polymerase chain reactions were performed in 25µL volumes with final concentrations of 1X AmpliTaq Gold PCR Buffer, 200 µM dNTPs, 2 mM $MgCl_2$, 0.8 mM BSA, 0.4 µM of each primer, 1U Taq-Gold, and 15 ng DNA (Life Technologies, Grand Island, NY). Thermocycler settings were: 95°C for 15 min; 35 cycles of 95°C for 60 s, 55.4°C for 60 s, 72°C for 60 s; and 72°C for 10 min. Products were sequenced on an ABI 3730 DNA Analyzer at the University of Missouri DNA Core Facility.

We aligned our data along with Haps A-M (Van Den Bussche *et al.* 2009) and Hap4 (Pelletier *et al.* 2011) using Geneious v5.5.3 (Drummond *et al.* 2011). To assign the new haplotypes to clades A-east, A-west, or B (Wooding & Ward 1997), we input sequences into MrBayes v3.1 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003). First, we determined that the General-Time Reversible (GTR) plus gamma distributed substitution rates was the best-fit model of evolution following comparison to other models in FindModel (Tao *et al.* 2005). We ran two chains for 10^7 generations with trees sampled every 100 generations. We assessed convergence of the runs in Tracer, then discarded the first 25% of the data as burnin. We visualized the consensus tree as a median-joining network using SplitsTree4 (Huson & Bryant 2006) to assign new haplotypes to previously described clades. We calculated pairwise F_{ST} for each area,

then population level summary statistics of haplotype (h) and nucleotide (π) frequencies in ARLEQUIN.

Population structure analysis with approximate Bayesian computation

We inferred the demographic history of black bears in the CIH via approximate Bayesian computation (ABC) implemented in DIYABC v2 (Cornuet *et al.* In Press). Using the microsatellite data we tested three models (*Split, Admix,* and *Founder*; Figure 2.2) with three populations: MB and MN combined into the Source, OU, and OZ. The *Split* model represents the distribution of genetic diversity had the translocation failed and no Source genotypes contributed to contemporary CIH individuals. In the model, the OU and OZ populations merge t_1 generations before present and the combined lineage merges with the Source t_2 generations after t_1 (Figure 2.2A). The population splitting remains in the *Admix* model with the addition of the translocation of individuals from the Source into OU and OZ (Figure 2.2B). Using a generation time of 6.3 years (Onorato *et al.* 2004), we set the admixture event to occur eight bear generations ago or approximately half-way through the decade long translocation program. We allowed the proportion of admixed alleles from the Source to vary between OU and OZ (λ_{OU} and λ_{OZ}) where the proportion remaining from CIH lineages was 1-λ. The *Founder* model differed from the *Admix* in that we assumed extirpation instead of a bottleneck; therefore, λ equals 1 for both OU and OZ (Figure 2.2C). Alternatively, the *Founder* model could represent a scenario in which bears in OU and OZ before the reintroduction did not contribute to reproduction.

We allowed the effective population sizes (N_e) of OU and OZ to vary and assumed they have been stable for the past four generations. Though we believe CIH bear populations continue to increase, DIYABC only allows discrete changes in N_e and not population growth or decline models. We modeled a bottleneck 16 generations ago (or approximately 1912) in the *Split* and *Admix* models. Further, an additional parameter allowed N_e to vary between four and eight generations (post-translocation) in the *Admix* and *Founder* models.

For each model we simulated $1.5x10^6$ datasets drawing from uniform prior distributions of N_e, divergence time, and admixture proportion (Table A1.2). For each simulation a set of summary statistics was calculated to compare to the observed dataset in the model selection procedure. The summary statistics recorded were: mean allelic richness (A_R), observed heterozygosity (H_O), mean allele size variation, pairwise F_{ST} (Weir & Cockerham 1984), and the genetic distance between two populations $(\delta\mu)^2$ (Goldstein *et al.* 1995).

We assumed that microsatellites followed a generalized stepwise mutation model (Estoup *et al.* 2002) with the prior of the mean mutation rate set as a uniform distribution from 10^{-5} to 10^{-3} (Ellegren 2000). We allowed microsatellites to vary up to 40 alleles as suggested in the DIYABC manual. All other microsatellite mutation parameters were at default settings including: individual locus mutation rate with a gamma distribution from 10^{-5} to 10^{-2}, mean coefficient of the geometric distribution (P) with a uniform distribution from 10^{-5} to 30^{-1}, individual locus coefficient P with a gamma distribution from 10^{-2} to

90^{-1}, a mean single nucleotide insertion or deletion (SNI) rate with a uniform log distribution from 10^{-8} to 10^{-5}, and an individual locus SNI rate with a gamma distribution from 10^{-9} to 10^{-4}.

To select between the three models, we used DIYABC's logistic regression. The posterior probability of each model was determined with the 1% (n = 15,000) of simulated datasets with the shortest Euclidean distance to the observed dataset in multivariate space (Cornuet *et al.* 2008; Storz & Beaumont 2002). Since the *Admix* model was highly supported (posterior probability 0.942, Table A1.2) over the *Split* (posterior probability 0.044) and *Founder* (posterior probability 0.014) models we moved forward with parameter estimation of *Admix*. To estimate our confidence for support of *Admix* we used DIYABC's built-in function for estimating type-I and type-II error across models (Cornuet *et al.* 2008). We simulated 500 pseudo-observed datasets for each model where model parameters were drawn from the posterior distributions instead of the priors (Cornuet *et al.* 2010).

We estimated the posterior distributions of θ (= $N_e\mu$) and admixture rates under the *Admix* model using a local linear regression of the closest 1% simulated datasets to the observed data following logistic transformation of the parameter values (Cornuet *et al.* In Press; Cornuet *et al.* 2008).

To estimate bias in the parameter estimates we used the built-in function in DIYABC and simulated 500 pseudo-observed datasets from the 90% highest posterior density

distributions for each parameter under the *Admix* model. DIYABC computes the true value then compares the estimated values to the variance in the simulated datasets. We report root of the relative mean integrated square error (RRMISE), and average relative bias and factor 2 score (which are the proportion of simulated datasets for which the estimate is at least half and at most twice the true value where values closer to 1 are desirable) of the mode.

RESULTS

Heterozygosity was similar across all seven study areas although mean number of alleles (A) was not (Table 2.1). Mean number of alleles was 9.33 in the Source (MN and MB samples combined) and 8.33 and 8.00 in OU and OZ, respectively. Missouri had the lowest A at 5.00. We estimated allelic richness for the four clusters and observed diversity patterns similar to that of the mean number of alleles (Source = 9.11; OU = 7.34; OZ = 7.07; MO = 5.48). The global Kruskal-Wallis test for allelic richness was significant (H = 23.5, df = 3, $P < 0.0001$) and only the pairwise comparison between OU and OZ was not significantly different. Estimates of F_{IS} were 0.055, 0.021, 0.016, and -0.144 respectively for the Source, OU, OZ, and MO, although no population was significantly different ($P = 0.263 - 0.930$) from the null hypothesis of random mating.

We assessed HWE in each area, including separate analyses for MO and MO-Oz (see below), and observed three loci deviated from the null hypothesis at the Bonferroni corrected $\alpha < 0.003$. In the MO-Oz group, loci G1D and G10U did not conform to expectations under HWE, nor did locus UarMU23 in MB. We observed varying amounts

of non-random assortment of alleles between areas (Table A1.3); as there were no consistent patterns between geographic areas, we did not exclude any loci from the analyses. The elevated non-random assortment may be due to bottlenecks or admixture, both of which were hypothesized for the CIH. Further, LD decays more slowly when admixture events occurred over an extended period (Chakraborty & Weiss 1988), as in the 10-year AR-Ou and AR-Oz translocations. Although Microchecker suggested the presence of null alleles (UA-P2H03 and G10O in AR-Ou, G10B in AR-Oz, and MU23 in MB), they were not at the same loci between sampling areas; therefore, no loci were removed due to null alleles.

Population structure

Three groups of geographically connected study areas were not significantly differentiated from each other at a Bonferonni corrected $\alpha < 0.006$ (Table 2.2). We reran F_{ST} tests following STRUCTURE results, which identified two clusters in MO (detailed below) where individual admixture coefficients at K = 3 were used to discriminate samples between MO and MO-Oz. The MB and MN samples (F_{ST} 0.015) were combined into the "Source," the AR-Ou and OK-Ou (F_{ST} 0.015) were combined into "OU," and the AR-Oz, OK-Oz, and MO-Oz (F_{ST} 0.005-0.013) were not differentiated (AR and OK samples were combined into "OZ").

Results from STRUCTURE at successive numbers of genetic clusters reflect the varying processes effecting differentiation among populations (Figure 2.3). Individuals from the Ouachitas (OK-Ou and AR-Ou) grouped as one genetic cluster while individuals from the

Ozarks (OK-Oz and AR-Oz) grouped as a second cluster; both results support observed non-significant F_{ST} values (Table 2.2). The OU and OZ clusters appeared at K = 2 and represent the largest separation of the data (Table A1.4). Recent migration from OU to OZ was 0.025 (95% CI = 0.007 – 0.043), and from OZ to OU was 0.014 (95% CI = 0.0003 – 0.028).

Individuals captured in MO grouped with the Ozarks at K = 2 as expected based on geography; however, half of the individuals in MO formed a unique cluster at K = 3 (Figure 2.3). The high proportion of OZ genotypes identified in MO would be consistent with a population expansion from AR-Oz; therefore, we analyzed the unique MO genotypes separately from the Ozark genotypes identified in MO (MO-Oz). The unique MO cluster was significantly different from all other areas in pairwise F_{ST} comparisons (Table 2.2); therefore, future reference to MO only includes the unique genetic cluster.

The Source formed a unique cluster at K = 4 (Figure 2.3). At both K = 2 and 3, the Source appeared to be admixed between OU and OZ (Figure 2.3). Given the known direction of the translocation, the unique clustering at K = 4 appears more relevant given the discontinuous geographic distribution of the study areas (Figure 2.1A). Additionally, unique clustering of the Source at K = 4 reflected both the non-significant F_{ST} values observed between MN and MB, and the significant differentiation observed with CIH populations. Thus four genetic clusters best represent the population structure within this dataset. We did not observe hierarchical substructure within any of the four clusters

(Figure 2.3); additionally, ΔK ranged from 0-95, which in combination with the lack of geographic pattern, we considered too low for further clustering.

Mitochondrial haplotype frequencies

From our sampling of MB, MN (contemporary and historic), OU, and OZ, we identified 22 haplotypes (Table A1.5), of which 13 were previously undescribed (GenBank accessions KJ406324-KJ406336). We assigned each of the 13 haplotypes to clades based on their position within the haplotype network (Appendix 2) that coincided with fixed differences between the clades. A single historic sample from MN assigned to clade B is the most eastern longitude ever reported for that clade which Byun *et al.* (1997) hypothesized to have originated on the coastal islands of contemporary British Columbia. Sampling areas differed in proportions of clade A-east and A-west haplotypes [so named due to their hypothesized location in Pleistocene refugia (Wooding & Ward 1997); Table 2.3] with the average proportion of A-east haplotypes equal to 0.40, 0.77, 0.73, and 0.03 in contemporary samples of the Source, OU, OZ, and MO areas, respectively. The haplotype network was distinguished by the location of high frequency haplotypes (HapB and E, both in CIH and Source) occupying internal nodes in clades A-east and A-west. Lower frequency haplotypes appeared at the tips of the network and suggest newer mutations derived from haplotypes occupying internal nodes (Appendix 2). The highest frequency haplotypes in OU and OZ were HapB, F, H, and E, while the highest frequencies in the Source were HapB, T, and Y (additionally HapF, X, and E were high frequency in geographically adjacent Ontario, Table A1.5).

Haplotypic diversity mirrored nuclear genetic diversity. Specifically, the source areas (contemporary and historic) had the highest haplotypic diversity (0.837-0.944; Table 2.3) while CIH subpopulations had moderate diversity (OU = 0.660; OZ = 0.513). Individuals captured in Missouri that clustered with OZ genotypes had diversity (0.695) similar to OZ. Haplotypic diversity of the individuals assigned to MO was 0.054 where most individuals shared HapE, a single individual whose nuclear genome assigned half to MO and half to OZ had HapB.

Population modeling

The demographic model accounting for admixture (*Admix*) was the best supported model in a logistic regression of posterior probabilities (0.942) comparing the 1% of datasets closest to the observed data for each model. None of the simulated summary statistics were significantly different (P = 0.140-0.629) from the observed data. We calculated the empirical type I error of the *Admix* model at 3.8%. Combining all simulations under the two alternative models (n = 1,000), we calculated the type II error at 1.3%.

We estimated the parameters θ and λ of the model using the 1% of simulated datasets closest to the observed data in multivariate space (Table 2.4). We estimated contemporary N_e of OU, OZ, and Source to be 382, 193, and 19,600 individuals respectively. The model estimated N_e historically (OU = 2,320, OZ = 7,580), following the bottleneck (OU = 29, OZ = 111), and following the translocation (OU = 27, OZ = 21). The bias estimates suggested that N_e was underestimated (except historic N_e of OU which was overestimated; Table 2.4). This was not unexpected because effective population

size is the geometric mean of a population over time. As DIYABC does not model continuous population growth, we expected point estimates of N_e over multiple generations to be underestimated. The admixture rates for OU and OZ were 0.863 and 0.638, respectively, and were not biased (Table 2.4).

DISCUSSION

Decreased genetic diversity in reintroduced populations

Our results support the hypothesis that black bears in the CIH were not extirpated but underwent a bottleneck in the early 1900s before occurring at low densities. We also observed lower genetic diversity at nuclear and mitochondrial markers in OU and OZ compared to the Source. Population genetic theory predicts a loss of genetic diversity following bottleneck and founder events (Nei *et al.* 1975). Reintroduced populations often experience both events, as a population goes through a demographic bottleneck resulting in reduced N_e of the remnant population before low population size prompts management action for a translocation, which may also function as a bottleneck moving a subset of the source population's alleles. In a survey of genetic diversity in translocated populations, approximately half of the reviewed studies reported decreased diversity in the translocated compared to the source population (Groombridge *et al.* 2012). Rapid population expansion and a large number of translocated individuals were associated with no difference in diversity between translocated and source populations (Groombridge *et al.* 2012). While these latter attributes describe the CIH translocation, this system was also characterized by high diversity in the Source, an attribute associated with the loss of genetic diversity following translocation bottlenecks. These results highlight the

interplay of variables in translocation design, some of which managers may control (i.e.- using a genetically diverse source, removal of pressure for population decline) while others may be beyond direct management control (i.e.- rate of population growth, immigration). We expected genetic diversity in OU and OZ to be higher because contemporary diversity resulted from admixture, thereby combining remnant and translocated alleles. Low frequency alleles may have been lost from remnant and source lineages due to their respective demographic and translocation bottlenecks.

The MO genetic cluster offers a natural experimental comparison of expected genetic diversity had the translocation not taken place. We hypothesize that bears in the Ozark Mountains were separated into at least two habitat patches during the demographic bottleneck of the 1900s, where the MO patch was isolated from other Ozark patches (Appendix 1 and Figure A1.1). The unique MO cluster had an allelic richness with 3.6 fewer alleles per locus than the Source, while average allelic richness in the admixed (OU and OZ) populations was 1.9 fewer alleles per locus than the Source. We do not know the number of alleles in the CIH before the demographic bottleneck nor what that parameter would be if the translocation were a founder event rather than an admixture event. However, the presence of more alleles in OU and OZ than in MO, where management action was not taken, illustrates the utility of translocations for increasing genetic diversity despite the loss of some alleles due to the translocation bottleneck.

Genetic drift in reintroduced populations

Genetic drift was the strongest evolutionary force in the CIH following the translocation. Two lines of evidence support this; first, F_{ST} values between the Source and either OU or OZ (0.046-0.056; Table 2.2) were approximately half those between OU and OZ (0.080-0.093). Thus, differentiation was greater between the geographically adjacent OU and OZ subpopulations than between those subpopulations and the Source, despite the Source contributing the majority of genetic diversity during the admixture event ($\lambda_{OU} = 0.863$, $\lambda_{OZ} = 0.638$). Other mammalian studies where multiple reintroductions used the same source population also observed drift and significant F_{ST} differentiation between reintroduced populations (Biebach & Keller 2009; Hicks *et al.* 2007; Wisely *et al.* 2008). We observed drift eight generations after the translocation, a similar time frame to the 10, 6.5, and 5 to 18 generations (depending on when reintroduction occurred) generations observed in Alpine ibex (Biebach & Keller 2009), black footed ferrets (Wisely *et al.* 2008), and elk (Hicks *et al.* 2007), respectively. We believe this effect was due to the stochastic distribution of alleles between AR-Ou and AR-Oz so that following the translocation allele frequency distributions were already divergent. Drift exacerbated these differences so that OU and OZ were less similar to each other than each subpopulation was to the Source. The magnitude of θ for OU and OZ suggest similar levels of drift since the translocation in contrast to θ_{Source} which was two orders of magnitude greater than in the CIH suggesting limited drift in the Source and allele frequencies similar to those when the translocation occurred. Migration between populations can decrease the effects of drift due to gene flow (Le Corre & Kremer 1998); however, the CIH contains three potential partial barriers: the Arkansas River,

discontinuous forest habitat in the river valley, and US Interstate-40. We hypothesize that these partial barriers kept gene flow low between OU and OZ (1.4 to 2.5% per generation), and that gene flow may be lower than historic levels following construction of the AR portion of US Interstate-40, which was constructed between the early 1960s and 1973 (AHTD 2004).

Second, we observed a false signature of admixture within the Source at two and three a priori genetic clusters (Figure 2.3). Biebach and Keller (2009) also observed the source population clustering as admixed when it produced the founding individuals for three captive breeding programs which subsequently reintroduced Alpine ibex throughout Switzerland. The north central bear range where MB and MN are located contains continuous habitat and our estimate of the genetic distance between them suggests they represent a single bear population (Table 2.2). At four a priori clusters, the Source formed a unique cluster where individual-level genetic diversity was not partitioned between OU and OZ. We believe this pattern represents the history of these populations, specifically that the Source represents a single cluster of genetic diversity that was partitioned during the translocation (observed at K = 3, Figure 2.3). The subsequent influence of genetic drift in OU and OZ resulted in clusters distinct from the Source. We hypothesize that identity by descent and few mutations since the translocation created both the admixture pattern in the Source at K = 3 as well as the assignment of individuals to the Source despite collection in the OU and OZ subpopulations at K = 4.

Remnant genetic lineages in the Central Interior Highlands

We observed that genetic differentiation in both OU and OZ was strongly influenced by the translocation. Csiki *et al*. (2003) did not observe differentiation between OU, OZ, and Source and interpreted that as a signature of the translocation, while Warrillow *et al.* (2001) only observed differentiation between OZ and the Source. Additionally, we did not observe that OU contained remnant lineages as hypothesized based upon mitochondrial (Van Den Bussche *et al.* 2009) and nuclear (Faries *et al.* 2013) markers. Van Den Bussche *et al*. (2009) used the exclusive presence of HapF and H in OU and OZ to propose remnant bear lineages persisted through the bottleneck; however, Pelletier *et al.* (2011) observed a high frequency of HapF in Ontario, a province adjacent to MB and MN. We would expect high frequency haplotypes in the source to also be high frequency in the translocated individuals. We propose that contemporary high frequencies of HapF in the CIH may be a result of the translocation and not a unique lineage. Similarly to Van Den Bussche *et al.* (2009), we also observed CIH specific haplotypes (H, V, and W; Table A1.5). Each CIH specific haplotype was at the tip of either the A-east or A-west clade within the haplotype network (Appendix 2). While these CIH specific haplotypes have not been observed in northern localities, we cannot determine whether these haplotypes are remnant within CIH, recent mutations following translocation of common haplotypes to CIH, or introduced during the translocation. Extensive geospatial sampling may identify CIH specific haplotypes across a broader scale. Faries *et al.* (2013) observed microsatellite differentiation between OU and Source and also interpreted it as a remnant signal. We suggest that the increased number of loci and samples in this study increased the power to differentiate between subpopulations and that genetic diversity in

OU is a product of the translocation and genetic drift. Even if unique CIH alleles persist on the landscape today, the strong influence of the admixture event ($\lambda_{OU} = 0.863$) suggests that remnant and Source lineages have been thoroughly mixed.

We were unable to differentiate between alternative hypotheses regarding when the unique MO cluster arose in CIH (Appendix 1). The remnant status of MO is dependent on whether "remnant" is defined as before the translocation of bears from the Source or before European settlement. The distinctiveness of the MO cluster was driven by low genetic diversity and not private alleles; further the MO cluster was nested within the OZ cluster (Figure 2.3) suggesting that MO represents the diversity of the Ozarks before the translocation albeit having gone through a severe demographic bottleneck and sustained low N_e. Our data were unable to determine if MO and OZ were separate subpopulations before human settlement. There are no obvious physical barriers between the OZ and MO clusters that suggest a geographic hypothesis for two clusters within the Ozark Mountains. Additionally, the recolonization of southern MO by bears with OZ genotypes suggests that the species is capable of dispersal within the mountains. If we assume that habitat fragmentation from the late 1800s to the present is greater than pre-settlement fragmentation, it would suggest a single subpopulation existed in the Ozarks before European settlement.

Conclusions

Our results suggest that three management actions could benefit black bears in the Central Interior Highlands. First, it may be beneficial for state agencies in Arkansas,

Oklahoma, and Missouri to work together to unify bear management decisions. While habitat availability and bear densities may vary between study areas, it is clear that subpopulations in OU and OZ cross state boundaries. Collaborative management among state agencies could help ensure that gene flow continues within and among subpopulations and future barriers (e.g. roads) are avoided or mitigated. Second, our data do not support designation of OU or OZ genetic clusters as remnant lineages of the pre-translocation population; however, MO may represent a remnant lineage albeit with reduced genetic diversity compared to the population before the demographic bottleneck. Protection afforded to subpopulations may be made based on population size or resource availability (e.g. differential management in low mast or drought years), but Source lineages have been thoroughly mixed in OU and OZ and genetic diversity is low in MO, thereby removing truly remnant lineages from the landscape. Third, breeding between MO and MO-Oz could improve genetic diversity in the current MO cluster over time. Potential barriers that reinforce separation between MO and MO-Oz could be identified and alleviated to accelerate gene flow.

ACKNOWLEDGEMENTS

We thank Ricky Eastridge and Myron Means (AGFC) for assistance with sample collection in AR; Craig Endicott, Joe Hemphill, and Jeff Ford (ODWC) for collection in OK; and Jessica A. Philbrick and Michelle R. Anderson for lab assistance. We thank David Onorato and three anonymous reviewers for comments that improved the manuscript. Sample collection in Oklahoma was supported in part by Federal Aid Project F10AF00137 of the Oklahoma Department of Wildlife Conservation and Oklahoma State

University, administered through the Oklahoma Cooperative Fish and Wildlife Research Unit (Oklahoma State University, Oklahoma Department of Wildlife Conservation, U.S. Geological Survey, Wildlife Management Institute, and U.S. Fish and Wildlife Service cooperating). Funding for this project was provided by the Missouri Department of Conservation, Arkansas Game and Fish Commission, Federal Aid in Wildlife Restoration program, and Safari Club International Foundation. EEP was supported by a University of Missouri Life Sciences Fellowship. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

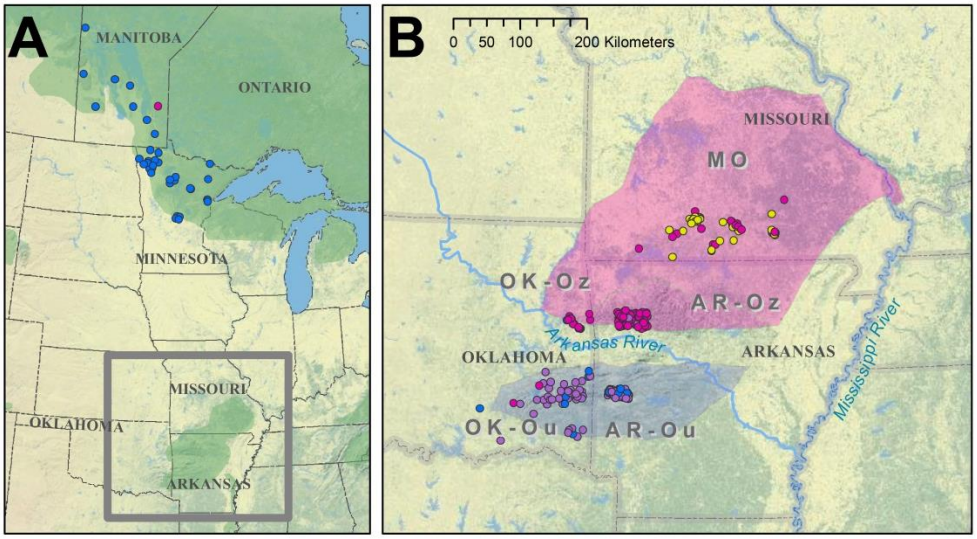

Figure 2.1- (A) Map of American black bear distribution (green) in central North America with names of the states (Oklahoma, OK; Arkansas, AR; Missouri, MO; Minnesota, MN) and provinces (Manitoba, MB; Ontario, ON) discussed in this paper. The grey box delineates the inset map (B) with the Ouachita (OU, purple) and Ozark (OZ, pink) land features and the five study areas of the CIH: OK-Ou, AR-Ou, OK-Oz, AR-Oz, and MO. Sampling locations were colored according to assignment of the four genetic clusters identified in Figure 2.3.

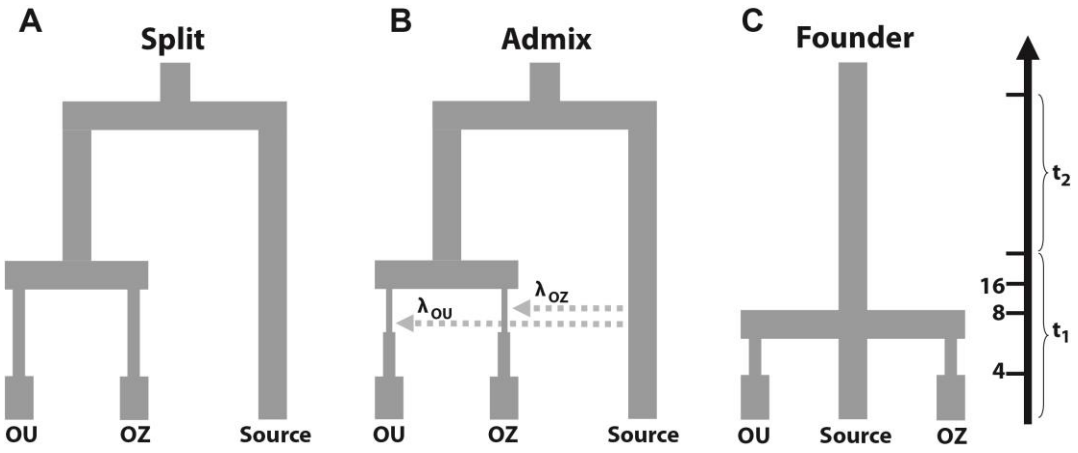

Figure 2.2- Three models of population divergence (A), divergence with admixture (B), or founding events (C) tested in DIYABC with variable effective population (N_e) sizes. The black bar indicates the number of generations from the present (t = 0) where demographic events were modeled, including: t = 4 current N_e; t = 8 admixture or founding events of Source into OU and OZ; t = 16 population bottlenecks in OU and OZ; t_1 the estimated time since divergence of the Central Interior Highlands (OU and OZ) populations; and t_2 the time between divergence of CIH and Source. The admixture parameters (λ_{OU} and λ_{OZ}) relate to the proportion of admixture coming from the Source.

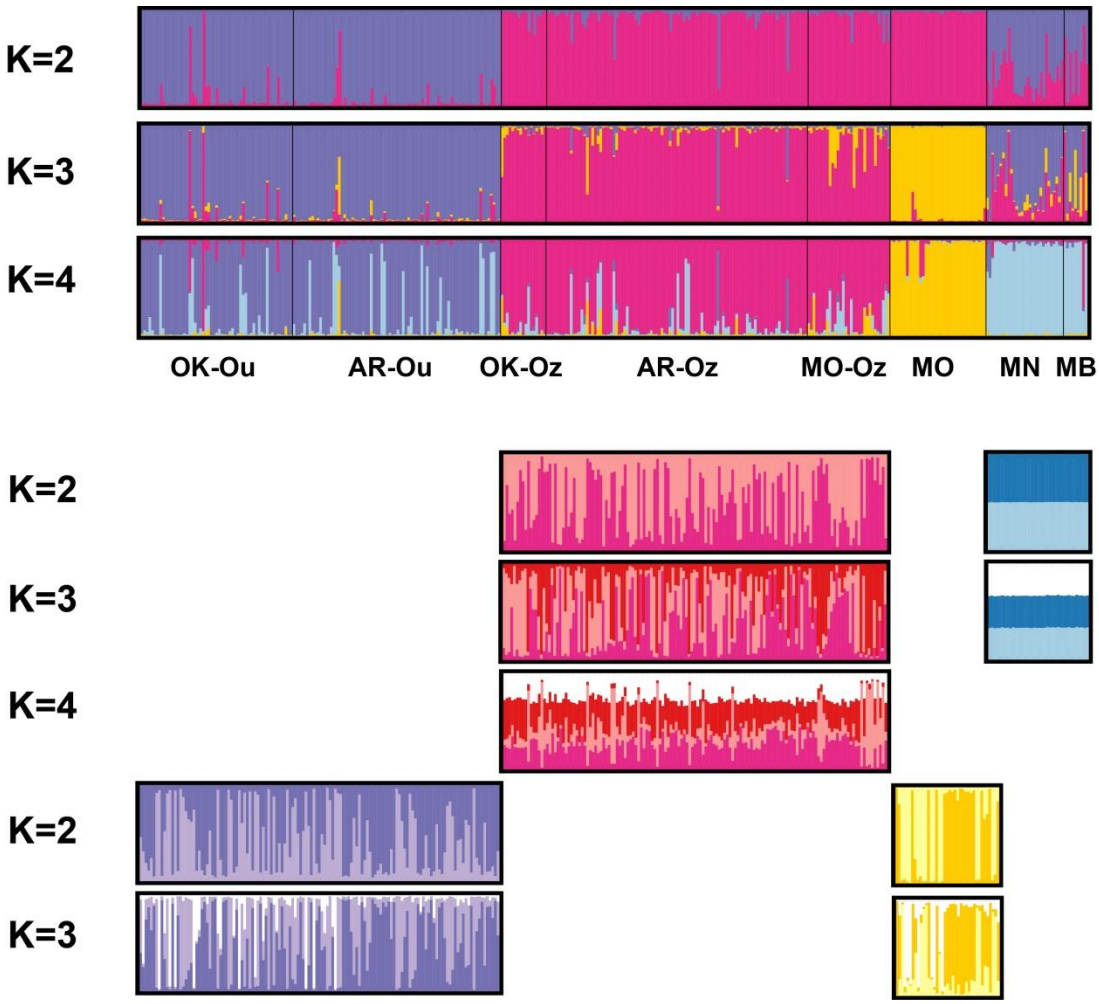

Figure 2.3- Admixture coefficients of individual black bears from eight study areas (as abbreviated in Table 2.1) assigned between an a priori number of genetic clusters (K). The top three rows reflect the full analysis, while the bottom five rows reflect independent analyses within each of the four clusters identified in the full analysis.

Table 2.1- Sample size (n), mean number of alleles (A), expected (H_E) and observed (H_O) heterozygosity, and inbreeding coefficient (F_{IS}). Genetic data were calculated with 15 microsatellite loci in seven study areas.

Population	Abbreviation	n	A	H_E	H_O	F_{IS}
Source	Source	38	9.33	0.7980	0.7929	0.055
Manitoba	MB	9	6.80	0.8196	0.8148	0.025
Minnesota	MN	29	8.40	0.7857	0.7862	0.073
Ouachitas	OU	135	8.33	0.7380	0.7331	0.021
Oklahoma	OK-Ou	57	7.33	0.7387	0.7350	-0.101
Arkansas	AR-Ou	78	7.73	0.7277	0.7317	0.095
Ozarks[1]	OZ	115	8.00	0.7764	0.7790	0.016
Oklahoma	OK-Oz	17	6.13	0.7748	0.8000	-0.109
Arkansas	AR-Oz	98	7.87	0.7756	0.7754	0.062
Missouri[2]	MO-Oz	28	7.40	0.7819	0.7689	-0.076
Missouri[3]	MO	39	5.00	0.6752	0.7767	-0.144

[1]- Combined Ozark cluster only includes samples from AR and OK.
[2]- Samples collected in Missouri and clustered with the Ozark cluster at K = 3 (see Figure 2.3).
[3]- Samples collected in Missouri and clustered as their own genetic cluster at K = 3 (see Figure 2.3).

Table 2.2- Pairwise F_{ST} based on genotypes of 15 microsatellite loci (below diagonal) and mitochondrial haplotypes (above diagonal) of black bears from historic (MN-H) and contemporary study areas (abbreviations as in Table 2.1). Significant values following Bonferroni correction ($P < 0.006$) are in bold.

	MB	**MN**	**MN-H**	**OK-Ou**	**AR-Ou**	**OK-Oz**	**AR-Oz**	**MO-Oz**	**MO**
MB	-	0.032	0.155	**0.692**	0.130	**0.680**	**0.436**	0.063	**0.288**
MN	0.015	-	0.155	**0.496**	0.033	**0.510**	**0.302**	0.022	**0.401**
MN-H	n/a	n/a	-	**0.441**	**0.149**	**0.377**	**0.269**	0.135	**0.601**
OK-Ou	**0.046**	**0.047**	n/a	-	**0.414**	**0.473**	**0.216**	**0.423**	**0.931**
AR-Ou	**0.055**	**0.048**	n/a	0.015	-	**0.463**	**0.206**	-0.023	**0.530**
OK-Oz	**0.047**	**0.056**	n/a	**0.086**	**0.093**	-	0.072	**0.436**	**0.953**
AR-Oz	**0.047**	**0.056**	n/a	**0.080**	**0.093**	0.005	-	0.187	**0.786**
MO-Oz	**0.050**	**0.053**	n/a	**0.084**	**0.094**	0.013	0.006	-	**0.471**
MO	**0.104**	**0.117**	n/a	**0.139**	**0.134**	**0.077**	**0.077**	**0.069**	-

Table 2.3- Summary statistics of mitochondrial haplotypes in Source (Manitoba and Minnesota), Central Interior Highlands (Ouachitas, Ozarks, and Missouri), and historic (Minnesota) areas with sample size (n), proportion of clade assignment, haplotype diversity and standard deviation (*h*), and nucleotide diversity and standard deviation (π).

		Proportion						
	n	**A-east**	**A-west**	**B**	***h***	h_{sd}	π	π_{sd}
Ontario[1]	660	0.39	0.61	0.00	0.691	0.071	0.015	0.008
Manitoba	9	0.22	0.78	0.00	0.944	0.070	0.007	0.005
Manitoba[2]	20	0.15	0.85	0.00	0.858	0.042	0.006	0.003
Minnesota	26	0.46	0.54	0.00	0.837	0.039	0.008	0.004
Minnesota[2]	6	0.83	0.17	0.00	0.333	0.215	0.001	0.001
MN-Historic	8	0.63	0.25	0.13	0.929	0.084	0.018	0.011
Ouachitas	52	0.77	0.23	0.00	0.660	0.033	0.006	0.004
Ouachitas[2]	114	0.84	0.16	0.00	0.438	0.050	0.004	0.003
Ozarks	27	0.89	0.11	0.00	0.513	0.090	0.005	0.003
Ozarks[2]	11	0.64	0.36	0.00	0.618	0.104	0.008	0.005
Missouri-OZ	21	0.52	0.48	0.00	0.695	0.070	0.008	0.004
Missouri	37	0.03	0.97	0.00	0.054	0.051	0.001	0.001

[1]- Data from Pelletier *et al.* (2011).
[2]- Data from Van Den Bussche *et al.* (2009) with MN summary statistics adjusted for six individuals reported in original table.

Table 2.4- Demographic parameter estimates (mode and 90% highest posterior density, HPD) from posterior distributions modeled under *Admix*. Estimates of bias include root of the relative mean integrated square error (RRMISE), average relative bias and factor 2 score of the mode of the posterior distribution. Parameters include theta ($\theta = N_e\mu$) for varying segments of the population history, divergence time ($\tau = t\mu$), admixture rate from the Source (λ), and microsatellite mutation rate (μ).

Parameter	Mode	90% HPD	RRMISE	Average Relative Bias	Factor 2
θ_{OU}	5.70×10^{-2}	2.17×10^{-2}-2.23×10^{-1}	1.558	-0.114	0.632
θ_{OZ}	2.88×10^{-2}	1.71×10^{-2}-2.21×10^{-1}	1.674	-0.189	0.550
θ_{Source}	2.92×10^{0}	1.73×10^{0}-8.50×10^{0}	0.792	-0.008	0.900
$\theta_{OU\text{-}Post\text{-}translocation}$	4.02×10^{-3}	4.22×10^{-3}-3.59×10^{-2}	2.419	-0.112	0.696
$\theta_{OZ\text{-}Post\text{-}translocation}$	3.17×10^{-3}	2.15×10^{-3}-1.91×10^{-2}	3.299	-0.039	0.730
$\theta_{OU\text{-}Bottleneck}$	4.32×10^{-3}	4.95×10^{-3}-2.26×10^{-1}	3.798	-0.306	0.276
$\theta_{OZ\text{-}Bottleneck}$	1.65×10^{-2}	7.75×10^{-3}-2.38×10^{-1}	3.025	-0.422	0.266
$\theta_{OU\text{-}Historic}$	3.46×10^{-1}	1.00×10^{-1}-2.81×10^{0}	3.845	0.302	0.508
$\theta_{OZ\text{-}Historic}$	1.13×10^{0}	3.41×10^{-1}-4.32×10^{0}	1.048	-0.101	0.672
λ_{OU}	0.863	0.587-0.948	0.173	0.027	1.000
λ_{OZ}	0.638	0.361-0.838	0.283	0.064	0.994
τ_1	5.79×10^{-1}	2.19×10^{-1}-2.82×10^{0}	2.325	-0.354	0.398
τ_2	1.82×10^{0}	1.67×10^{0}-5.64×10^{0}	0.877	-0.064	0.772
μ	1.49×10^{-4}	8.87×10^{-5}-3.97×10^{-4}	0.651	0.017	0.936

LITERATURE CITED

AHTD (2004) Historical Review, p. 264. Arkansas State Highway and Transportation Department, Little Rock, AR.

Baker PJ, Boitani L, Harris S, Saunders G, White PCL (2008) Terrestrial carnivores and human food production: impact and management. *Mammal Review* **38**, 123-166.

Bales SL, Hellgren EC, Leslie DM, Jr., Hemphill J, Jr. (2005) Dynamics of a recolonizing population of black bears in the Ouachita Mountains of Oklahoma. *Wildlife Society Bulletin* **33**, 1342-1351.

Beringer J, Pondrom A, Mays A, *et al.* (2008) Management plan for the black bear in Missouri. Missouri Department of Conservation.

Biebach I, Keller LF (2009) A strong genetic footprint of the re-introduction history of Alpine ibex (Capra ibex ibex). *Molecular Ecology* **18**, 5046-5058.

Brown AG (2007) *Demographic characteristics and habitat associations of an expanding black bear (Ursus americanus) population in Oklahoma*, Oklahoma State University.

Byun SA, Koop BF, Reimchen TE (1997) North American black bear mtDNA phylogeography: Implications for morphology and the Haida Gwaii glacial refugium controversy. *Evolution* **51**, 1647-1653.

Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proceedings of the National Academy of Sciences* **85**, 9119-9123.

Cornuet J-M, P P, J V, *et al.* (In Press) DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using single nucleotide polymorphism, DNA sequence, and microsatellite data. *Bioinformatics*.

Cornuet J-M, Ravigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11**, 1-11.

Cornuet JM, Santos F, Beaumont MA, *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719.

Csiki I, Lam C, Key A, *et al.* (2003) Genetic variation in black bears in Arkansas and Louisiana using microsatellite DNA markers. *Journal of Mammalogy* **84**, 691-701.

Drummond A, Ashton B, Buxton S, *et al.* (2011) Geneious v5.4. http://www.geneious.com.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* **4**, 359-361.

Eggert LS, Maldonado JE, Fleischer RC (2005) Nucleic acid isolation from ecological samples: Animal scat and other associated materials. In: *Methods in Enzymology* (eds. Elizabeth AZ, Eric HR), pp. 73-82. Academic Press.

Ellegren H (2000) Microsatellite mutations in the germline: Implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.

Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**, 1591-1604.

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

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.

Faries KM, Kristensen TV, Beringer J, *et al.* (2013) Origins and genetic structure of black bears in the Interior Highlands of North America. *Journal of Mammalogy* **94**, 369-377.

Fischer J, Lindenmayer DB (2000) An assessment of the published results of animal relocations. *Biological Conservation* **96**, 1-11.

Gardner-Santana LC (2007) *Patterns of genetic diversity in black bear (Ursus americanus) during a range expansion into Oklahoma*, Oklahoma State University.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Grenier MB, McDonald DB, Buskirk SW (2007) Rapid population growth of a critically endangered carnivore. *Science* **317**, 779.

Groombridge JJ, Raisin C, Bristol R, Richardson DS (2012) Genetic consequences of reintroductions and insights from population history. In: *Reintroduction biology:*

integrating science and management (eds. Ewen JG, Armstrong DP, Parker KA, Seddon PJ). Wiley-Blackwell, Oxford, UK.

Hicks JF, Rachlow JL, Rhodes OE, Williams CL, Waits LP (2007) Reintroduction and genetic structure: Rocky Mountain elk in Yellowstone and the western states. *Journal of Mammalogy* **88**, 129-138.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.

Johnson WE, Onorato DP, Roelke ME, *et al.* (2010) Genetic restoration of the Florida panther. *Science* **329**, 1641-1645.

Kalinowski ST, Wagner AP, Taper ML (2006) ML Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* **6**, 576-579.

Kristensen TV (2013) *Ecology and structure of black bear (Ursus americanus) populations in the interior highlands of Arkansas*, University of Arkansas.

Kristensen TV, Faries KM, White Jr. D, Eggert LS (2011) Optimized methods for multiplex genotyping analysis of hair samples for American black bears (*Ursus americanus*). *Wildlife Biology in Practice* **7**, 123-128.

Lackey CW, Beckmann JP, Sedinger J (2013) Bear historical ranges revisited: Documenting the increase of a once-extirpated population in Nevada. *The Journal of Wildlife Management* **77**, 812-820.

Laliberte AS, Ripple WJ (2004) Range contractions of North American carnivores and ungulates. *Bioscience* **54**, 123-138.

Larson S, Jameson R, Bodkin J, Staedler M, Bentzen P (2002) Microsatellite DNA and mitochondrial DNA variation in remnant and translocated sea otter (*Enhydra lutris*) populations. *Journal of Mammalogy* **83**, 893-906.

Le Corre V, Kremer A (1998) Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology* **11**, 495-512.

Miller KA, Nelson NJ, Smith HG, Moore JA (2009) How do reproductive skew and founder group size affect genetic diversity in reintroduced populations? *Molecular Ecology* **18**, 3792-3802.

Miller SD, Cully JF (2001) Conservation of black-tailed prairie dogs (*Cynomys ludovicianus*). *Journal of Mammalogy* **82**, 889-893.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.

Nei M, Tajima F (1981) Genetic drift and estimation of effective population size. *Genetics* **98**, 625-640.

O'Ryan C, Harley EH, Bruford MW, *et al.* (1998) Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Animal Conservation* **1**, 85-94.

Onorato DP, Hellgren EC, Van Den Bussche RA, Doan-Crider DL (2004) Phylogeographic patterns within a metapopulation of black bears (*Ursus americanus*) in the American Southwest. *Journal of Mammalogy* **85**, 140-147.

Paetkau D, Shields GF, Strobeck C (1998) Gene flow between insular, coastal, and interior populations of brown bears in Alaska. *Molecular Ecology* **7**, 1283-1292.

Pelletier A, Obbard ME, White BN, Doyle C, Kyle CJ (2011) Small-scale genetic structure of American black bears illustrates potential postglacial recolonization routes. *Journal of Mammalogy* **92**, 629-644.

Pelton MR, Van Manen FT (1994) Distribution of black bears in North America **12**, 133-138.

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

Pruett CL, Winker K (2005) Northwestern song sparrow populations show genetic effects of sequential colonization. *Molecular Ecology* **14**, 1421-1434.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A, Drummond AJ (2009) Tracer v1.5, http://tree.bio.ed.ac.uk/software/tracer/.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Rosenberg NA (2004) distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: A laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory Press.

Sanderlin JS, Faircloth BC, Shamblin B, Conroy MJ (2009) Tetranucleotide microsatellite loci from the black bear (*Ursus americanus*). *Molecular Ecology Resources* **9**, 288-291.

Smith KG, Clark JD (1994) Black bears in Arkansas: Characteristics of a successful translocation. *Journal of Mammalogy* **75**, 309-320.

Sobey D (2007) An analysis of the historical records for the native mammalian fauna of Prince Edward Island. *Canadian Field-Naturalist* **121**, 384-396.

Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* **56**, 154-166.

Taberlet P, Camarra JJ, Griffin S, *et al.* (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology* **6**, 869-876.

Tao N, Richardson R, Bruno W, Kruiken C (2005) FindModel, http://hcv.lanl.gov/content/hcv–db/findmodel/findmodel.html.

Taylor SS, Jamieson IG (2008) No evidence for loss of genetic variation following sequential translocations in extant populations of a genetically depauperate species. *Molecular Ecology* **17**, 545-556.

Van Den Bussche RA, Lack JB, Onorato DP, *et al.* (2009) Mitochondrial DNA phylogeography of black bears (*Ursus americanus*) in central and sourthern North America: Conservatoin implications *Journal of Mammalogy* **90**, 1075-1082.

Van Houtan KS, Halley JM, Van Aarde R, Pimm SL (2009) Achieving success with small, translocated mammal populations. *Conservation Letters* **2**, 254-262.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Warrillow J, Culver M, Hallerman E, Vaughan M (2001) Subspecific affinity of black bears in the White River National Wildlife Refuge. *Journal of Heredity* **92**, 226-233.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Williams BW, Scribner KT (2010) Effects of multiple founder populations on spatial genetic structure of reintroduced American martens. *Molecular Ecology* **19**, 227-240.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Wisely S, Santymire R, Livieri T, Mueting S, Howard J (2008) Genotypic and phenotypic consequences of reintroduction history in the black-footed ferret (*Mustela nigripes*). *Conservation Genetics* **9**, 389-399.

Wooding S, Ward R (1997) Phylogeography and Pleistocene evolution in the North American black bear. *Molecular Biology and Evolution* **14**, 1096-1105.

Yackulic CB, Sanderson EW, Uriarte M (2011) Anthropogenic and environmental drivers of modern range loss in large mammals. *Proceedings of the National Academy of Sciences* **108**, 4024-4029.

CHAPTER 3

Phylogeographic analyses of American black bears (*Ursus americanus*) suggest four glacial refugia and complex patterns of post-glacial admixture

Emily E. Puckett[1], Paul D. Etter[2], Eric A. Johnson[2], and Lori S. Eggert[1]

[1]- Division of Biological Sciences, 226 Tucker Hall, University of Missouri, Columbia, MO 65211
[2]- Institute of Molecular Biology, 315 Streisinger Hall, University of Oregon, Eugene, OR 97403

ABSTRACT

Studies of species with continental distributions continue to identify intraspecific lineages despite continuous habitat. Lineages may form due to isolation by distance, adaptation, divergence across barriers, or genetic drift following range expansion. We investigated lineage diversification and admixture within American black bears (*Ursus americanus*) across their range using 22k SNPs and mitochondrial DNA sequences. We identified three subcontinental nuclear clusters which we further divided into nine geographic regions: Alaskan (Alaska-East), eastern (Central Interior Highlands, Great Lakes, Northeast, Southeast), and western (Alaska-West, West, Pacific Coast, Southwest). We estimated that the western cluster diverged 67 kya, before eastern and Alaskan divergence 31 kya; these divergence dates contrasted with those from the mitochondrial genome where clades A and B diverged 1.07 Mya, and clades A-east and A-west diverged 169 kya. We combined estimates of divergence timing with hindcast species distribution models to infer glacial refugia for the species in Beringia, Pacific Northwest, Southwest, and Southeast. Our results show a complex arrangement of admixture due to expansion out of multiple refugia. The delineation of the genomic population clusters was inconsistent with the ranges for 16 previously described subspecies. Ranges for *U.*

a. pugnax and *U. a. cinnamomum* were concordant with admixed clusters, calling into question how to order taxa below the species level. Additionally, our finding that *U. a. floridanus* has not diverged from *U. a. americanus* also suggests that morphology and genetics should be reanalyzed to assess taxonomic designations relevant to the conservation management of the species.

INTRODUCTION

Generalist species may occupy continental distributions as they can utilize a variety of resources.

Studies of genetic lineages within continentally distributed generalists have identified multiple broadly geographically distributed lineages even when ranges were continuous (Burbrink, et al. 2008; Reding, et al. 2012). This raises questions about how lineage diversification, or population structure at a finer geographic scale, arises in continuous habitat. Hypotheses that explain lineage divergence for continuously distributed species include dispersal limitation and isolation by distance (Platt, et al. 2010), local adaptation (Richardson, et al. 2014), divergence across barriers (Sacks, et al. 2005), and genetic drift following range expansion (Hallatschek, et al. 2007; Hofer, et al. 2009).

Given the strong influence of Pleistocene (2.6-0.012 Mya) glacial cycles on species distributions, many continentally distributed species have endured one or more range contraction and expansion events. Ice sheets covered northern North America during the Last Glacial Maximum (LGM; 22-18 kya), meaning that species with present day distributions across modern Canada necessarily had to contract into refugia before

recolonizing. Generally, refugial areas have higher genetic diversity than expansion areas due to the compression of diversity into a limited portion of the range while expansion areas only receive a subset of the refugial diversity reaching the range front (Austerlitz, et al. 1997; Hewitt 1999); however, admixture following lineage diversification also produces a signature of high genetic diversity (Petit, et al. 2003). Additionally, range contraction and expansion should, respectively, decrease and increase effective population size (N_e) thereby influencing genetic drift and differentiation (Li and Durbin 2011).

We examined lineage diversification in the American black bear (*Ursus americanus*), a forest generalist with a continental distribution (Figure 3.1). Black bears are endemic to North America and estimated to have speciated 1.8 – 1.2 Mya (Kurten and Anderson 1980). Three hypotheses regarding black bear phylogeography have been proposed based on mitochondrial data. Wooding and Ward (1997) and Byun *et al.* (1997) each identified two clades: A and B (where Byun *et al.* named the clades continental and coastal, respectively). Within Wooding and Ward's data, clade A further separated in geographic space into eastern and western clades. The authors agreed that the refugium for clade A was located in the eastern broadleaf forests; however, they disagreed in the placement of the refugium for clade B. Wooding and Ward (1997) hypothesized the location in the coniferous forests along the Pacific coast while Byun *et al.* (1997) hypothesized it was located on the near shore islands or continental shelf off modern day British Columbia. Neither hypothesized specific refugial locations for clades A-east and A-west, in part because the clades were not recognized as being distinct. The third

hypothesis proposed that the Sierra Madre Mountains served as a glacial refugium from which bears dispersed north following deglaciation (Varas-Nelson 2010). This hypothesis was based upon a high frequency of A-west haplotypes in Arizona and northern Mexico and contemporary distribution of those haplotypes north of the author's southwestern study area. In this study, we employ both nuclear and mitochondrial data to resolve both the number of lineages and refugial locations. Nuclear markers may delimit intraspecific lineages at a finer resolution than mitochondrial haplotypes and provide multiple realizations of the coalescent process to estimate variation in population genetic parameters (Edwards and Bensch 2009; Wu and Drummond 2011). Thus, identification of intraspecific lineages may lead to a more detailed understanding of population genetic and phylogeographic processes.

Sixteen black bear subspecies were named between 1790 and 1957 (Lariviere 2001). Many were based on small differences in cranial morphology that may be the result of natural variation. Incongruence between subspecies designations and genetics has been observed in multiple taxa, specifically where there were more described subspecies than genetic lineages (Zink 2004; Torstrom, et al. 2014). We sampled individuals across eight of the subspecies ranges to identify intraspecific lineages in the nuclear genome and compare their spatial distribution to lineages detected in the mitochondrial genome. We combined species distribution modeling and estimates of divergence times in a phylogeographic analysis to infer locations of glacial refugia to compare to previous hypotheses.

MATERIALS and METHODS

Sample collection

Samples were collected throughout the range by US state and Canadian provincial wildlife managers, university researchers, or private guides. Most samples were from hunter harvested individuals with a limited number from road kills or nuisance bears euthanized by wildlife agents. Blood samples from live individuals were collected in Minnesota and Alaska-Kenai; hair samples were collected from live individuals in Ontario. We combined these samples (n = 439) with those collected in Missouri (n = 67), which formed two clusters Ozark Mountains (OZ) and remnant Missouri (MO) as described in Puckett *et al*. (2014). We designate the combined OZ and MO samples as the Central Interior Highlands (CIH) samples, as the CIH represents the mid-continental Ozark and Ouachita Mountains. Samples were collected between 2010 and 2013 (Table A3.1).

A geographic location was reported with each sample; however, locations varied in accuracy from GPS coordinates of harvest or den location to the county or state delineated management unit from which the bear was sampled. When a GPS location was not provided, we created a coordinate by placing a point in a random direction of the nearest town or creek provided as the harvest location. When only a county or management unit was provided, the center point of the polygon was used as the sample location.

DNA extraction, RAD sequencing, and SNP calling

Tissue specific DNA extraction and microsatellite genotyping (Appendix 1) followed the protocols of Puckett *et al*. (2014). We selected 95 tissue or blood samples from across the geographic range for restriction-site associated DNA (RAD) sequencing. We prepared RAD-Seq libraries with up to 100ng of genomic DNA from each sample and one negative control made up of water following the protocol described in Ruiz-Lopez *et al*. (Accepted). Briefly, samples were digested with SbfI before ligation of the P1 adapter (Etter, et al. 2011; Ruiz-Lopez, et al. Accepted). We pooled 24 samples each in 4 sub-libraries, inputting proportionally more of the reaction if less than 100ng of DNA was used in the digestion, before shearing in a Bioruptor for 6 minutes to obtain fragments with an average size of 500bp. Following end repair we ligated the P2 adapter before 14 cycles of PCR amplification (Etter, et al. 2011). We normalized the size of the library fragments using AMPure XP beads (Agencourt), pooled the sub-libraries based on yield and sequenced the library on two lanes of an Illumina HiSeq at the University of Oregon Core facility.

We processed the raw reads in STACKS v1.09 (Catchen, et al. 2013). To assess the number of mismatches allowed both within an individual and between loci when building RADtags (-M and -n flags respectively), we processed two samples under a number of scenarios and compared the number of RADtags that formed as de novo loci versus those that mapped to the polar bear (*U. maritimus*) genome (Liu, et al. 2014) as detailed in Catchen *et al*. (2013). We first assessed the M parameter by holding n constant at two while varying M between two and five. We observed a decrease in the overmerged

RADs at M = 3 and no improvement using higher M values. Therefore, we selected M = 3 for both the final RAD processing and as the constant level when we allowed n to vary between zero and two. At n = 2 there were the fewest overmerged RADs; therefore, we retained that parameter for the full processing. We used the BioCluster maintained by the Informatics Research Core Facility at the University of Missouri to process RADtags first by mapping reads to the polar bear genome using Bowtie v0.12.7 (Langmead, et al. 2009) before running the ref_map.pl script within the STACKS software. Following processing, we filtered tags to the set with one SNP and two alleles, where at least 85 bears were genotyped for each SNP. This resulted in 31,447 SNPs that mapped to 505 scaffolds of the polar bear genome. We selected scaffolds with at least 125 SNPs (81 scaffolds with 21,995 SNPs) for population genomic analyses.

Population genomic analyses

We used CHROMOPAINTER v0.0.4 and FINESTRUCTURE v0.0.2 to estimate the number of populations across the geographic range (Lawson, et al. 2012). Data was first phased and imputed using fastPHASE v1.2 (Scheet and Stephens 2006) for each scaffold. We estimated the recombination rate between each SNP on a chromosome using the makeuniformrecfile.pl script accompanying the CHROMOPAINTER software. We ran a preliminary run of CHROMOPAINTER on each scaffold with nine samples and 10 E-M iterations; estimates of N_e and mutation rate were averaged using the neaverage.pl script. CHROMOPAINTER was run independently on each scaffold using the estimated parameters for N_e (1,913) and mutation rate (0.000416) with no donor samples identified and 10 iterations maximized over both N_e and mutation rate. Output between scaffolds

was combined with ChromoCombine before analysis in FINESTRUCTURE where following $5x10^5$ burnin iterations, $5x10^5$ iterations were run sampling every 10^3 iteration; we observed stability for each parameter for which a trace file was produced. The FINESTRUCTURE tree was produced using the full hill climb setting with $5x10^4$ iterations. We observed the first two axes of variation from a principal component analysis (PCA) calculated within FINESTRUCTURE. We identified populations as the branches from the FINESTRUCTURE tree without polytomies; these appear in the heat map as the samples that lie along the diagonal with equal coancestry values as adjacent samples. We collapsed the populations into nine geographic regions using both deeper branches in the FINESTRUCTURE tree and our knowledge of habitat connectivity within the species. Differentiation in the nine regions was quantified in ARLEQUIN v3.5.1.3 (Excoffier and Lischer 2010). Finally, three subcontinental clusters were inferred both from the FINESTRUCTURE tree and heat map, where clusters encompassed one or more geographic regions.

We also estimated mitochondrial diversity by sequencing 556bp (3' Cytochrome b and control region) on a subset of samples (n = 402) from across the range similarly to Puckett *et al.* (2014). We aligned sequences with *U. americanus* haplotypes A-Z (Onorato, et al. 2004; Van Den Bussche, et al. 2009; Puckett, et al. 2014) before identifying haplotypes using COLLAPSE v1.2 (Posada 2004). We visualized haplotypes with SPLITSTREE v4 (Huson and Bryant 2006), then assigned to clades based on their position in the network. We estimated haplotype (h) and nucleotide (π) diversity, and

differentiation where samples were assigned to the approximate geographic group as identified from the nuclear data in ARLEQUIN.

Nuclear divergence modeling

Using the regional clusters inferred above (excluding AK-West and Pacific Coast due to small sample sizes), we inferred admixture using MIXMAPPER v2 (Lipson, et al. 2013; Lipson, et al. 2014) with the following parameters: to compute moment stats we used 1,000 bootstrap replicates over 50 blocks, to compute most additive trees we used 10,000 data subsets, and for each admixture analysis we ran 500 replicate trees. The three population test (Patterson, et al. 2012) identified three regions with significant admixture. MIXMAPPER then inferred drift distances of the four source regions, the sources and admixture proportions for each admixed region, and the drift distance for the admixed regions.

We estimated the timing of divergence from the inferred drift units and N_e using the equation from Lipson *et al.* (2013):

$$D \approx 1 - exp^{-\frac{t}{2Ne}}$$

As there was not a linkage map available for bears, the LD based methods to estimate admixture timing were inappropriate for this dataset. Additionally, the small sample size for each group in the SNP dataset appeared to bias downwards estimates of N_e; therefore, we utilized the microsatellite dataset with much larger sample sizes to estimate N_e. We estimated N_e for each group independently by simulating 10^6 datasets in DIYABC v2.0.3 (Cornuet, et al. 2014). Since structured populations will bias N_e estimates upwards, we

allowed population structure, inferred by state or province of the sampling location, within each model. A uniform prior of 10^1 to 10^4 was used for N_e in each population, and divergence time was set to 10^1 to $3x10^3$ (see Appendix for individual model parameters). The summary statistics recorded were mean number of alleles, mean genic diversity, mean size variance, and $(\delta\mu)^2$ (Goldstein, et al. 1995). We estimated the posterior distribution using a local linear regression of the closest 1% of simulated datasets to the observed data. We calculated N_e from the modal estimate of both θ and μ, and also report the 90% highest density probability (HPD) of N_e. We used the mode and HPD of N_e in the divergence time calculation above, and converted from generations to years using the black bear specific generation time of 6.3 years (Onorato, et al. 2004).

Mitochondrial clade divergence

To estimate lineage divergence of the three mitochondrial clades, we first sequenced and assembled genomes from one bear each from clades A-west and A-east (accession numbers KM257059 and KM257060, respectively). We followed the Illumina library preparation methods of Steele *et al.* (2012). Briefly, we repaired ends of 2-5µg of sonicator sheared genomic DNA with the NEB Prep kit E600L (New England Biolabs, Ipswich, Massachusetts, USA). We ligated barcoded adaptors before size selecting fragments between 300-350bp. We PCR amplified barcoded products before 100bp single-end sequencing on the Illumina GAIIx Genome Analyzer at the University of Missouri DNA Core facility. We assembled reads into contigs in YASRA (Ratan 2009), a reference based assembly program, using the clade B *U. americanus* mitogenome (AF303109) as the reference. Following an initial assembly of contigs, we replaced the

mitogenome reference with the newly created contigs and mapped previously unassembled reads to the contigs to create a single scaffold, again using YASRA.

We estimated divergence dates between *U. americanus* clades using a time calibrated tree across *Ursus* species and within species lineages. We aligned 16 *Ursus* mitochondrial genomes using GENEIOUS v5.4 (Drummond *et al.* 2011) before manually removing the variable repeat region. We included three aDNA samples with associated tip dates; these included the extinct bears *U. deningeri* (GenBank Accession: KF437625; 409 kya) and *U. spelaeus* (EU327344; 32 kya), and one aDNA sample from *U. maritimus* (GU573488; 120 kya) (Bon *et al.* 2008; Dabney *et al.* 2013; Lindqvist *et al.* 2010). For other bear species, we selected genomes representative of within species lineages when available, including: *U. arctos* (GU573489, AF303110, EU497665, and JX196367), *U. maritimus* (JX196390), *U. thibetanus* (DQ402478, EF076773, and EF587265), *U. americanus* included the two genomes sequenced above and a clade B representative (AF303109), *U. malayanus* (NC_009968), and *U. ursinus* (NC_009970) (Bon *et al.* 2008; Delisle & Strobeck 2002; Hou *et al.* 2007; Lindqvist *et al.* 2010; Miller *et al.* 2012; Tsai *et al.* 2009; Yu *et al.* 2007). Using the program BEAUTI, we set up a BEAST v1.8 input file with the following parameters: no partitioning of the data, a lognormal relaxed substitution model (Drummond *et al.* 2006), and a constant coalescent (Kingman 1982) process on the tree model. Following the method of Heath (2012), we added three fossil calibrations to the tree. The first fossil was placed at the crown of *U. americanus* (node N) with an exponential prior (mean = 0.45, offset = 0.90) (Kurten & Anderson 1980). The second calibration for *U. etruscus* was placed at the crown of the brown and cave bear clade

(node C) with a lognormal distribution (mean = 0.20, std = 1.00, offset = 2.60) (Krause *et al.* 2008). The last fossil was placed on the root (node A) of the tree with a lognormal distribution (mean = 0.50, std =1.00, offset = 3.50) (Hunt Jr 1998). We also included the three tip dates in the analysis. We ran 10^8 MCMC steps sampling every 10^4 with 10% of the run discarded as burnin following observation of the posterior trace in TRACER v1.6. Finally, we output the tree with the highest median log credibility score using TREE ANNOTATOR v1.8 and report the node age and 95% HPD from this tree.

Ecological niche modeling

We used MAXENT v3.3.3k (Phillips *et al.* 2006) to model the current distribution of American black bears across North America, then hindcast the distribution to climates 21 kya and 120 kya. We combined presence localities from this study (n = 453) with samples from Missouri, Arkansas, and Oklahoma (n = 284) sampled in Puckett *et al.* (2014), and 597 locations from the Global Biodiversity Information Facility (GBIF 2013). When multiple samples occur in one grid cell MAXENT randomly removes samples so that only one sample per cell remains. We used all 19 variables in the WorldClim dataset for both current and historical climates (Hijmans *et al.* 2005). All data was interpolated on 2.5 arc-minute grids. In MAXENT we ran five replicates (5,000 iterations per replicate) with subsampling where 40% of the data was used to train the model and the remaining 60% for the test. We enabled all auto features. Model outputs were imported into ArcView v10.1 (ESRI, Redlands, CA) where the probability of occurrence scale was standardized so that visual representation between the three models was congruent.

RESULTS

Population structure, divergence, and admixture

Our analysis of population structure using 22k SNPs identified 26 populations nested into nine geographic regions within three subcontinental clusters (western, eastern, and Alaskan; Figures 3.1 and 3.2, Table A3.1). Within the western cluster, samples collected in the southern Rocky Mountains (Southwest: New Mexico, Arizona, and Colorado) had a higher drift estimate than samples from the northern Rocky Mountains (West: Idaho, Oregon, and Montana) with the exception of two samples from the Coast Range in Oregon (Pacific Coast; Figures 3.1A and 3.2). While the eastern cluster had a lower overall drift parameter (Figure 3.2), geographic regions were still identified. We grouped samples from the Atlantic Coast (Florida and North Carolina-coast) and Appalachian Mountains (North Carolina-mountains and West Virginia) together as the Southeast; and samples to the east of the Great Lakes (Pennsylvania, Michigan-Lower Peninsula, New York, and Maine) as the Northeast. Samples west of the Great Lakes (Wisconsin, Michigan-Upper Peninsula, and Minnesota) grouped broadly with those in Missouri which we separated into the Great Lakes and Central Interior Highlands (CIH; Figure 3.1A). Notably, the Great Lakes was paraphyletic in the tree produced by FINESTRUCTURE; the decision to group was made based on geography and known translocation of bears from Minnesota (MN) to the CIH (Puckett, et al. 2014) which may have produced the relationship of MN samples grouping with CIH instead of Wisconsin and Michigan-UP. We observed a signature of range expansion in the eastern cluster based on PCA analysis (Figure A3.1). Finally, we identified the Alaskan cluster from samples collected from the mainland near Alexander Archipelago and Kenai Peninsula

(hereafter referred to as AK-East). However, three samples from the Alexander Archipelago grouped with the broader western cluster (AK-West). FINESTRUCTURE identified multiple populations within the Alexander Archipelago samples, including three in AK-East and two in AK-West; mapping these populations revealed they were isolated on mainland peninsulas and Kuiu Island. While we inferred the AK-East group as a separate Alaskan cluster, we note coancestry from both the eastern and western clusters (Figure 3.2).

We observed significant differentiation between most pairwise comparisons of the nine geographic regions using F_{ST} for both SNP (0.025-0.429) and mitochondrial (0.081-0.917) datasets (Table 3.1). We caution that interpretation of F_{ST} values for the Pacific Coast and AK-West SNP geographic regions could be problematic due to low sample sizes; thus, we removed these two groups from further analyses. MIXMAPPER identified the Great Lakes, Northeast, and West as admixed (Figure 3.3). The best supported source populations for both the Great Lakes and Northeast were the Southeast and AK-East (Table 3.2). The Southwest and CIH were the best supported sources for the West.

Using the drift distances inferred from MIXMAPPER (Table 3.2), and N_e estimates (Table A3.3) from coalescent simulations of each region, we inferred the time since either admixture or divergence. We estimated divergence of the Southwest from the remaining eastern cluster at 66.6 kya (HPD: 44.7 – 144 kya). In the eastern cluster, the divergence time between the Southeast and CIH was 10.5 kya (HPD: 7.2 – 17.9 kya)

when estimated using values from the Southeast and 13.6 kya (HPD: 11.5 – 28.2 kya) with values from the CIH. Divergence of AK-East from the ancestral eastern cluster was estimated at 1.6 kya (HPD: 1.3 – 3.6 kya); however, this estimate seems low and may be a result of low diversity in the AK-Kenai Peninsula samples (Table A3.2) which may have reduced the drift estimate. As an alternative measure of divergence between AK-East and the eastern cluster, we used the N_e for the Southeast and the full drift distance, as it is additive (Lipson, et al. 2013), along the Southeast branch (0.346) to estimate a divergence time of 30.6 kya (HPD: 21.0 – 51.9 kya). Using the same method we estimated the time since admixture in the Northeast (0.78 kya; HPD: 0.66 – 3.6 kya), Great Lakes (1.5 kya; HPD: 0.65 – 4.8 kya), and West (9.4 kya; HPD: 8.0 – 33.2 kya).

Mitochondrial haplotype network

We identified 81 CytB-control region haplotypes (Table A3.4) across the range, of which 63 were previously undescribed (GenBank accessions KM230048-KM230113). We assigned haplotypes to clades based on their position within the haplotype network (Figure A2.1) that coincided with fixed differences between the three clades. The proportion of clades A-east, A-west, and B differed across the range where 74% and 26% of samples were in clades A-west and A-east respectively in the Southwest (n = 34); however, in the West (n = 64) 73% of samples were in A-west and 27% assigned to B (Table A3.4). In the Southeast (n = 60), 85% and 15% assigned to clades A-east and A-west. All haplotypes in the Northeast assigned to A-east (n = 63); in the Great Lakes (n = 52), 56% and 44% assigned to clades A-east and A-west, respectively. Alaskan and

western Canadian samples (n= 71) contained all three clades, where clades A-east, A-west, and B had respective percentages of 2%, 70%, and 28%.

Divergence of mitochondrial clades

We estimated the median substitution rate for the *Ursus* mitochondrial DNA to be 0.0145 substitutions per site per Mya (95% HDP: 0.0123 – 0.0176). This result was similar to the 0.0120 substitutions per site per Mya (95% HPD: 0.0090 – 0.0150) estimated across the Ursidae (Krause *et al.* 2008). The crown of the American black bear clade was 1.068 Mya (HPD: 0.900 – 1.281; Figure 3.3). Divergence between the A-west and A-east clades of American black bears was 169 kya (HPD: 115 – 230).

Species distribution models

We observed a marked difference in the probability of occurrence between the interglacial (current and 120 kya) and glacial (21 kya) palaeodistribution models (Figure 3.4). Specifically, the interglacial models were distinguished by a wide latitudinal range and distinct areas of high occurrence probability in the west and east. In both models, the areas of high occurrence were connected by a corridor across central Canada, although the probability of the corridor was higher in the current climate than 120 kya. In contrast, the glacial model had a distinct high probability across southern North America (Figure 3.4B) with decreased connectivity in modern Louisiana and across the Rocky Mountains. Northern coastlines, particularly bordering the Pacific Ocean and into Beringia (contemporary Alaska), were identified as moderate probability refugial habitat. We inferred four refugial areas based on the LGM palaeodistribution model (Figure 3.4B),

mitochondrial distributions (Figure A3.1B), and timing of nuclear lineage divergence during the LGM. Specifically, we infer that refugia were broadly located in Beringia, the Pacific Northwest, Southwest, and Southeast. These areas were broadly distributed when accounting for cells with both high and medium occurrence probabilities as may be expected for a generalist species.

DISCUSSION

Phylogeography

Our data suggest that geographic clustering within black bears arose both from divergence processes and differential admixture proportions following multiple range expansions. While the divergence between the eastern and western ranges was due to isolation and lack of gene flow along the northern corridor, the admixture within the geographic regions appears to be due to differences between short and long range expansions. Below we detail regional and cluster divergence in relation to the four inferred refugia.

Mid-Pleistocene Divergence Patterns- Lineage divergence in both the mitochondrial and nuclear genomes began before isolation in LGM refugia. We estimated divergence between clades A and B at 1.07 Mya, suggesting an early population divergence following speciation. Given the contemporary limited spatial distribution of clade B and wide distribution of clade A, clade B may have been geographically isolated along the Pacific coast for much of black bear evolution. Within clade A, we estimated divergence between A-east and A-west haplotypes at 169 kya; divergence of the nuclear eastern and

western clusters occurred 67 kya. These results, in conjunction with the 120 kya species distribution model (Figure 3.5C), suggest that black bears occupied a northern range before expansion and divergence around the North American grasslands. The grasslands emerged 7-5 Mya (Axelrod 1985) and appear to be a primary barrier to dispersal. Divergence data for both the mitochondrial clades and nuclear clusters suggest the establishment of structure before the LGM; although finer scale patterns (such as some of the geographic regions and admixture) occurred following the LGM.

We propose that the southern range had increased connectivity before the LGM. Mitochondrial haplotypes from clades A-east and A-west were respectively observed in Arizona (Southwest) and the North Carolina-mountains (Southeast; Figure 3.1B, Table A3.4). We propose this gene flow occurred before the LGM given the lack of nuclear admixture in the sampling areas with the discordant mitochondrial haplotypes, a signature we would expect if gene flow were more recent.

Western Refugia- Based on our hindcast distribution models, we propose a refugium along the Pacific coast and into the Sierra Nevada Mountains. Our probability of occurrence data suggest that the Pacific Northwest refugium was primarily continentally located south of the Cordilleran ice sheet (Demboski, et al. 1999) and not along the near shore islands from Alaska to British Columbia. The Pacific Coast and AK-West groups may represent descendants from this refugium albeit with varying proportions of admixture from the Alaskan cluster (Figure A3.1). We hypothesize that clade B haplotypes were isolated in the Pacific Northwest refugia and expanded northwards

following the LGM as these haplotypes were found from California (Wooding and Ward 1997) to southeastern Alaska (Table A3.4). However, we note this hypothesis does not preclude the presence of A-west haplotypes in the refugia which likely supported the ancestral populations for the West geographic region. Our results do not identify micro-refugia which could lead to substructuring within the refugia as previously observed in the Pacific Northwest (Swenson and Howard 2005; Galbreath, et al. 2010; Shafer, et al. 2010). There is little reason to think the northern Rocky Mountains were devoid of bears during this period, as other species with northern and southern Rocky Mountain populations such as tree squirrels (Chavez, et al. 2014), pikas (Galbreath, et al. 2010), and Douglas-fir (Gugger, et al. 2010) inhabited both ranges continuously through time. Given our data that the West was admixed, the genomic makeup of the West may have been similar to the Southwest before admixture 9.7 kya. In regards to the admixture, the Great Lakes may be a more reasonable second source population than the CIH. It is likely that the Great Plains acted as a barrier between the forests of the CIH and northern Rocky Mountains. In addition, individuals from Minnesota and Manitoba (Great Lakes genotypes) were translocated to the CIH eight generations before collection of samples for this study. While this hypothesis was not supported by a goodness of fit test in MIXMAPPER, we present it as an alternative explanation of the data.

We estimated the Southwest diverged from the eastern and Alaskan clusters 67 kya, although this time may be underestimated given the mitochondrial divergence. Although we were unable to sample bears in Mexico, we hypothesize that they would cluster with the Southwest. Our LGM hindcast (Figure 3.5B) shows the Colorado Plateau and Sierra

Madres as connected during this period and suggests that the Southwest refugium was broadly distributed. If additional data supports this hypothesis, a revision of the southwestern (*U. a. amblyceps*) and Mexican (*U. a. eremicus* and *U. a. machetes*) subspecies designations may be warranted.

Beringia Refugium- We inferred divergence of the Alaskan cluster before the LGM (31 kya). This suggested that bears living in the northern range contracted into both the Beringia and Southeast refugia. This conclusion was supported by data from the Y-chromosome (Bidon, et al. 2014), in which a haplotype observed in Alaska was closely related to one from Vermont, while two western haplotypes from Montana and Oregon formed a second Y chromosome clade. We propose that following contraction into the Beringia refugium bears diverged to form the Alaskan cluster, while at the same time A-west mitochondrial haplotypes introgressed into the population setting up the observed mito-nuclear discordance. The eastward expansion of AK-East out of the refugium and across modern Canada explains both the moderate and low frequencies of A-west haplotypes in the Great Lakes and the Northeast, respectively (Table A3.4); both regions were hypothesized to be admixed with AK-East as one of the sources. This long distance range expansion was congruent with our admixture time analysis where admixture of AK-East into the Great Lakes occurred earlier (1.5kya) than into the Northeast (0.78 kya). STRUCTURE analyses using microsatellites also supported this hypothesis showing decreasing proportions of the Alaskan cluster from Yukon to Maine (unpublished data).

Southeast Refugium- Divergence between the Southeast and CIH did not occur until after the LGM (10.5 – 13.6 kya). This divergence date was supported by our LGM hindcast model (Figure 3.5B) which indicated the Ozark and Ouachita Mountains, which comprise the CIH, were poor habitat for black bears 21 kya. We hypothesize that bears used the Mississippi River valley to colonize from southern Louisiana northwards into the CIH and that gene flow with the Southeast was limited due to low forest connectivity.

We observed a signature of range expansion from the Southeast to Northeast then Great Lakes in the PCA data (Figure A3.1); however, isolation by distance, the low Alaskan admixture proportion, or uneven sampling may also have produced this signature (McVean 2009; Meirmans 2012). This data suggests that Lake Michigan, and its predecessor lakes, were a barrier to westward expansion, and that bears expanded counter clockwise around the Great Lakes landform following deglaciation. The assignment of Michigan-Lower Peninsula to the Northeast and Michigan-Upper Peninsula to the Great Lakes supports this hypothesis which was also observed in bobcats (Reding, et al. 2012) and chipmunks (Rowe, et al. 2004).

Implications for bear research and conservation

Both the three nuclear clusters and three mitochondrial clades were wider than the ranges of the 16 named black bear subspecies (Lariviere 2001), where the nine regional clusters may be better units to discuss subspecies designations. Our sampling covered eight subspecies ranges: *U. a. altifrontalis* (Pacific coast from British Columbia to California), *U. a. amblyceps* (Colorado Plateau), *U. a. americanus* (majority of range covering

Alaska, northern and eastern Canada, Great Lakes region, and the Atlantic coast), *U. a. cinnamomum* (northern Rocky Mountains), *U. a. floridanus* (Florida), *U. a. luteolus* (Louisiana and Mississippi), *U. a. perniger* (Alaska-Kenai Peninsula), and *U. a. pugnax* (Alaska-Alexander Archipelago). The Southwest had strong concordance with the defined range for *U. a. amblyceps*. The West and Pacific Coast respectively aligned with the ranges of *U. a. cinnamomum* and *U. a. altifrontalis*. While subspecies delimitation criteria are not agreed upon (Braby, et al. 2012), the association of the West with the *U. a. cinnamomum* subspecies raises the question of whether or not geographic ranges with recently admixed genomes qualify for subspecies status. We observed two areas of incongruence between the geographic clusters and subspecies ranges. First, bears sampled along the Alexander Archipelago (designated *U. a. pugnax*) split into two regional clusters: AK-East, which was closely associated with bears from Alaska-Kenai Peninsula (*U. a. perniger*), and AK-West. We propose that the broad geographic designation of the Alexander Archipelago as *U. a. pugnax* is reductive as the area supports two genetic clusters (Alaskan and western). We hypothesize that range expansion of *U. a. perniger* out of the Beringia refugium resulted in substructure within AK-East, and that *U. a. pugnax* may be related to northward range expansion from bears in the Pacific Northwest refugium. As suggested by the coancestry data (Figure 3.2), bears across the Alexander Archipelago and mainland were admixed, albeit in different proportions given higher Alaskan or western ancestry. Second, the Southeast spans the historically designated ranges of *U. a. americanus* and *U. a. floridanus*; additionally, our microsatellite data support grouping samples from Mississippi (*U. a. luteolus*) within the Southeast. We interpret the increased coancestry within FL as a result of recently

reduced population size and genetic diversity (Dixon, et al. 2007) resulting in higher drift. This suggests that these three subspecies represent a single genetic cluster. In combination with our hypothesis of Southeast and AK-East range expansion to produce the majority of genetic diversity in the Northeast and Great Lakes, we suggest that *U. a. americanus* may be the most accurate subspecies designation for bears across the eastern range. We are not arguing for the removal of protected status from populations in Florida (delisted from the state threatened list in 2012) or Louisiana (listed as threatened under the US Endangered Species Act since 1992) as demography and resource availability should weigh heavily in management decisions; however, bears in these geographic regions may warrant protection as distinct population segments rather than as subspecies.

Our results also have broader implications related to evolutionary inference in *Ursus*. Divergence within *U. thibetanus* (Asiatic black bear) was similar in age to divergence within both *U. americanus* and *U. arctos* (Figure 3.4). The topology we recovered was consistent with the topology of Choi *et al.* (2010); however, we only included three of the seven described subspecies of *U. thibetanus* due to the availability of sequenced mitogenomes. Therefore it is possible that the 720 kya divergence date within *U. thibetanus* was underestimated, although this could be resolved by fully understanding diversity across the species' range.

American black bears are frequently used as outgroups in Ursid phylogenomic studies, particularly those that elucidate divergence between brown and polar bears (Cahill *et al.* 2013; Hailer *et al.* 2012; Lindqvist *et al.* 2010; Miller *et al.* 2012). However, most

studies use one black bear sample (Cahill *et al.* 2013; Miller *et al.* 2012), or when using multiple samples skew the geographic distribution to the west (Kutschera *et al.* 2014). We believe these studies underestimate black bear diversity as the three lineages harbor private alleles. By underestimating genetic diversity, the mutation rate is also likely to be underestimated which may in turn underestimate divergence times and theta (Leaché *et al.* 2014). For example, Kutschera *et al.* (2014) estimated N_e of bear species with intron sequence data, and estimated that *U. americanus* N_e was half that of *U. thibetanus* and *U. arctos*. While this may reflect the truth, particularly if *U. americanus* ancestors went through a bottleneck during expansion into North America from east Asia, an alternative hypothesis suggests that greater sampling diversity across the black bear range may result in different parameter estimates. Our results may aid bear evolutionary research by identifying diverse geographic areas to sample from considering the known genetic structuring across multiple bear species.

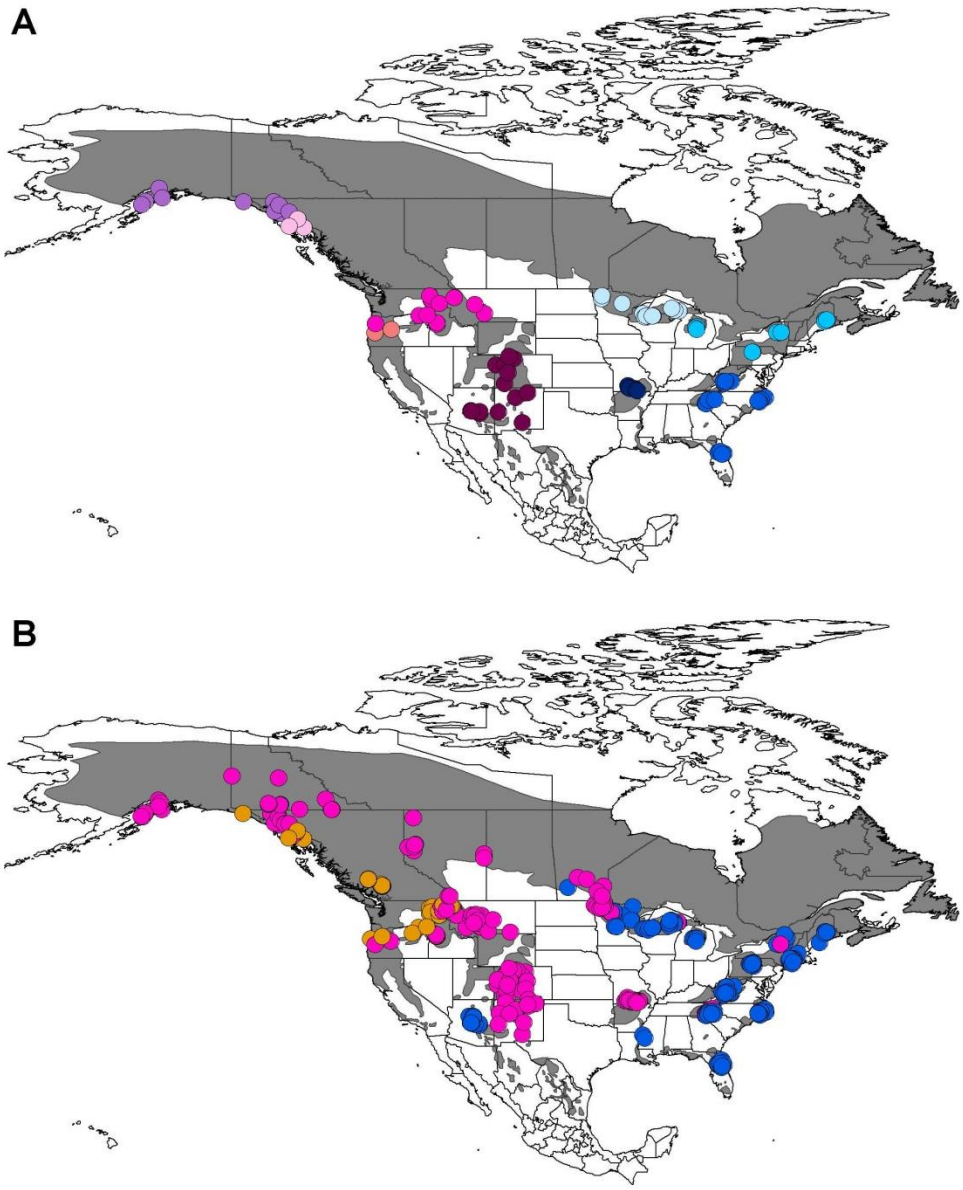

Figure 3.1- (A) Map of sample locations colored by inferred nuclear genomic cluster (Alaska-East: purple; Central Interior Highlands: navy; Great Lakes: light blue; Northeast: bright blue; Southeast: medium blue; Alaska-West: light pink; West: pink; Pacific Coast: coral; Southwest: raspberry) with American black bear distribution (grey). (B) Map of sample locations colored by mitochondrial haplotype clade (A-east: blue; A-west: pink; B: orange).

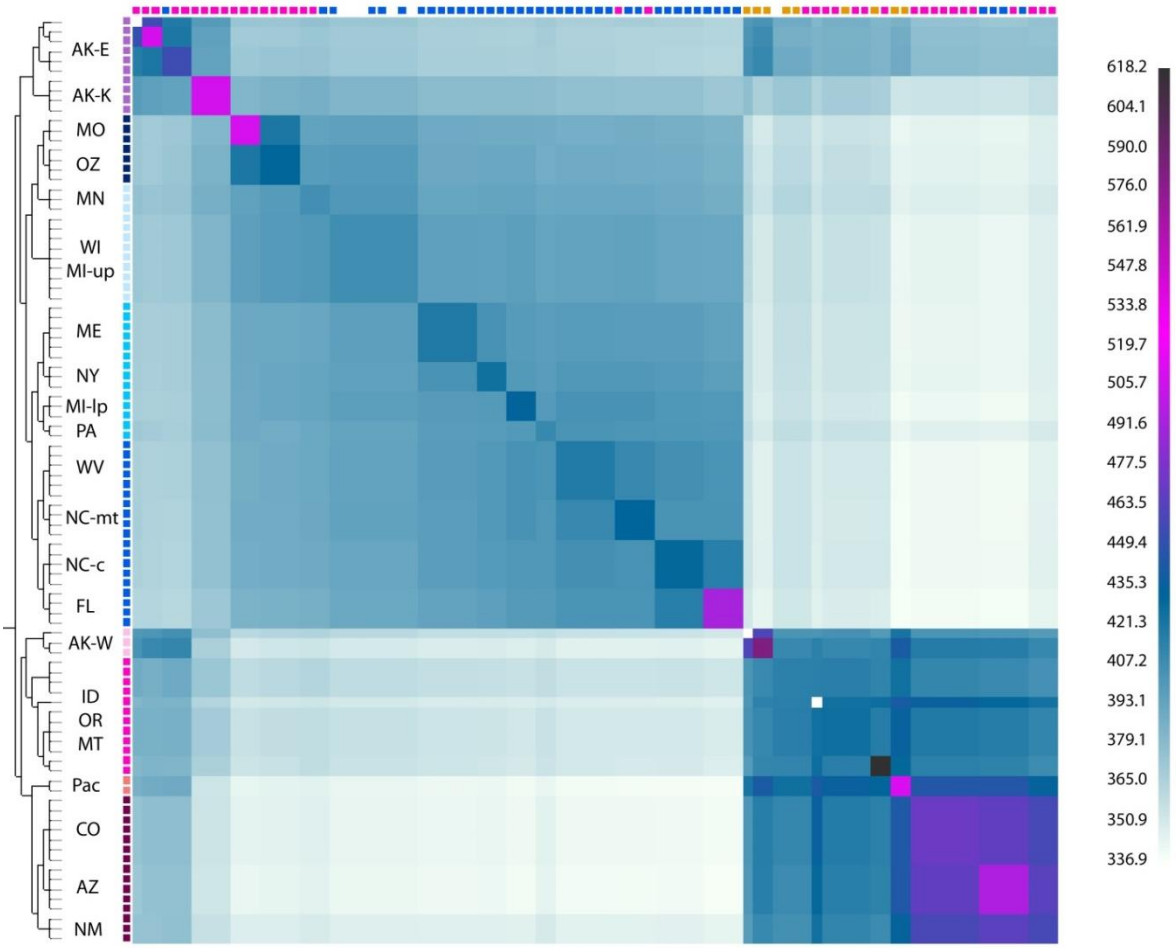

Figure 3.2- Coancestry heat map of the American black bear, where light blue and black respectively denote lower and higher shared ancestry. Samples were ordered based on the 26 populations identified within FINESTRUCTURE and appear within the population tree and as adjacently colored boxes along the diagonal. Vertical boxes were colored to represent the nine clusters used in subsequent analyses; from the top moving downwards the clusters were AK-East (Alaska-ABC Islands and Alaska-Kenai Peninsula), CIH (Missouri and Ozarks), Great Lakes (Minnesota, Wisconsin, and Michigan-Upper Peninsula), Northeast (Maine, New York, Michigan-Lower Peninsula, and Pennsylvania), Southeast (West Virginia, North Carolina mountains and coast, and Florida), AK-West (Alaska-ABC Islands), West (Idaho, Montana, eastern Oregon), Pacific Coast (Oregon-Coast Range), and Southwest (Colorado, Arizona, and New Mexico). Horizontal boxes denote the mitochondrial clade (A-east: blue; A-west: pink; B: orange; and no box: not sequenced) of each sample. See Figure A3.1 for the spatial distribution of the samples. The adjacently colored boxes along the diagonal indicate the 26 populations inferred.

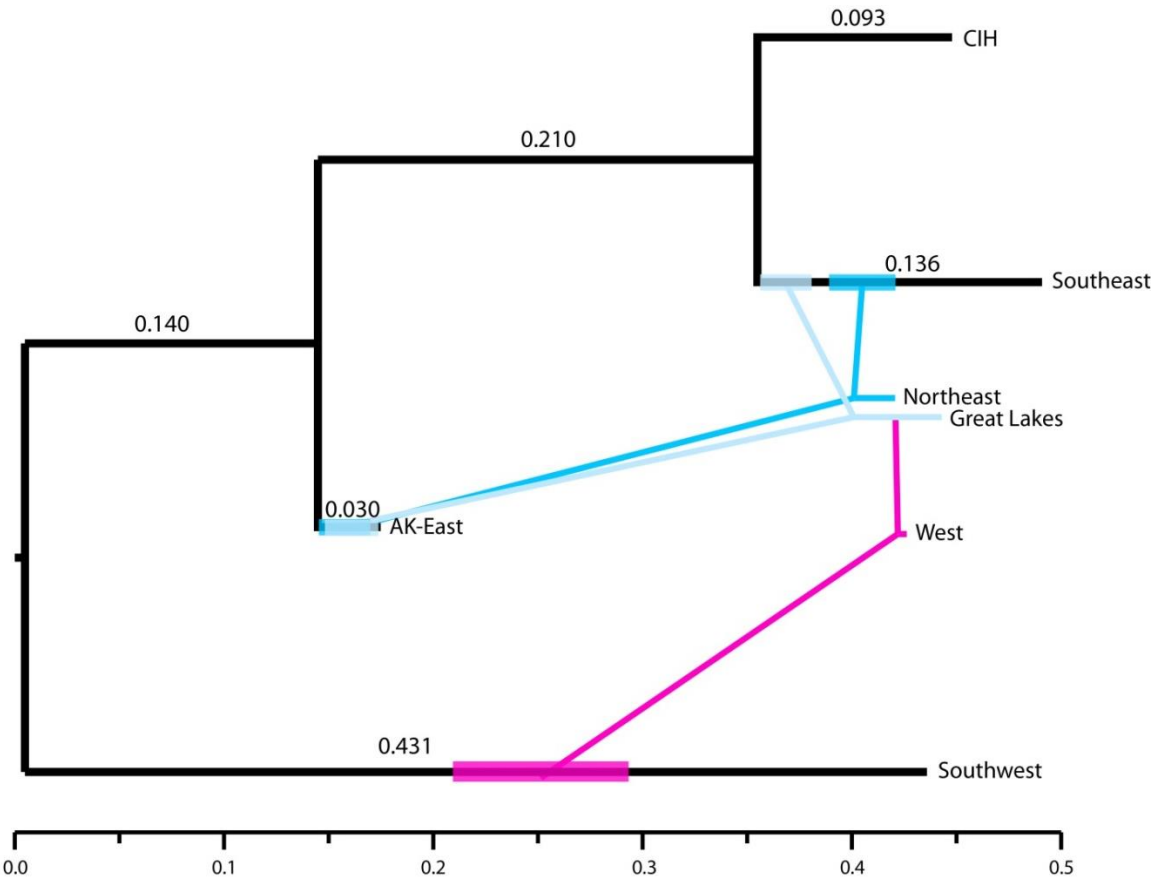

Figure 3.3- Drift tree of unadmixed and admixed (Northeast: bright blue; Great Lakes: light blue; West: pink) black bear populations. Tree branches, branches for drift following admixture, and pre-split range (see Table 3.2) for admixed lineages all to scale in drift units estimated by MIXMAPPER.

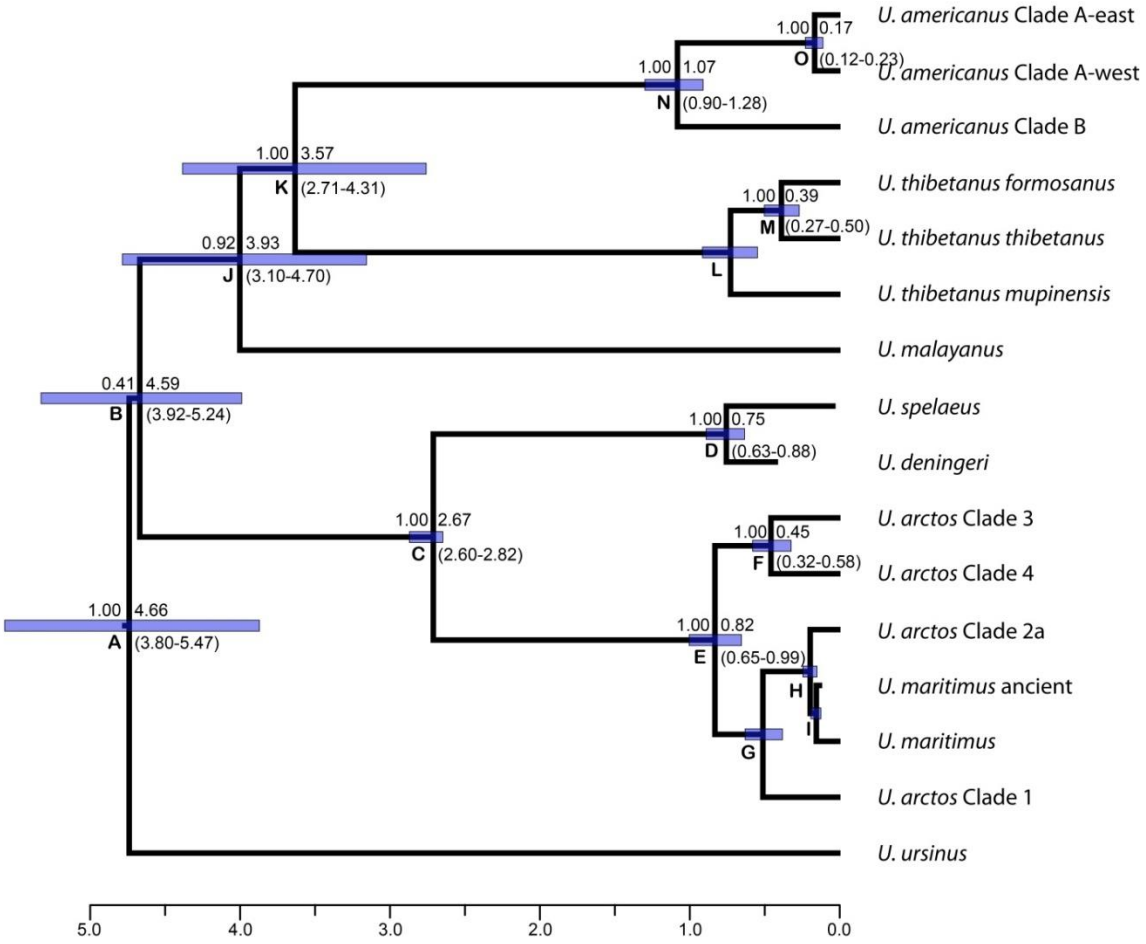

Figure 3.4- Phylogenetic tree of *Ursus* mitochondrial genomes including within species lineages. Each node was labeled (bottom left) for reference, and includes Bayesian posterior support (upper left), and node age (upper right) with 95% highest posterior density (HPD, bottom right) in millions of years (Mya). Values for nodes G, H, I, and L were excluded for readability where each node had a Bayesian posterior probability of 1.00. Nodes G, H, and I had respective divergence times of 0.51 (HPD: 0.38-0.62), 0.20 (0.15-0.25), and 0.16 (0.13-0.20) Mya. Node L diverged 0.72 (0.54-0.90) Mya.

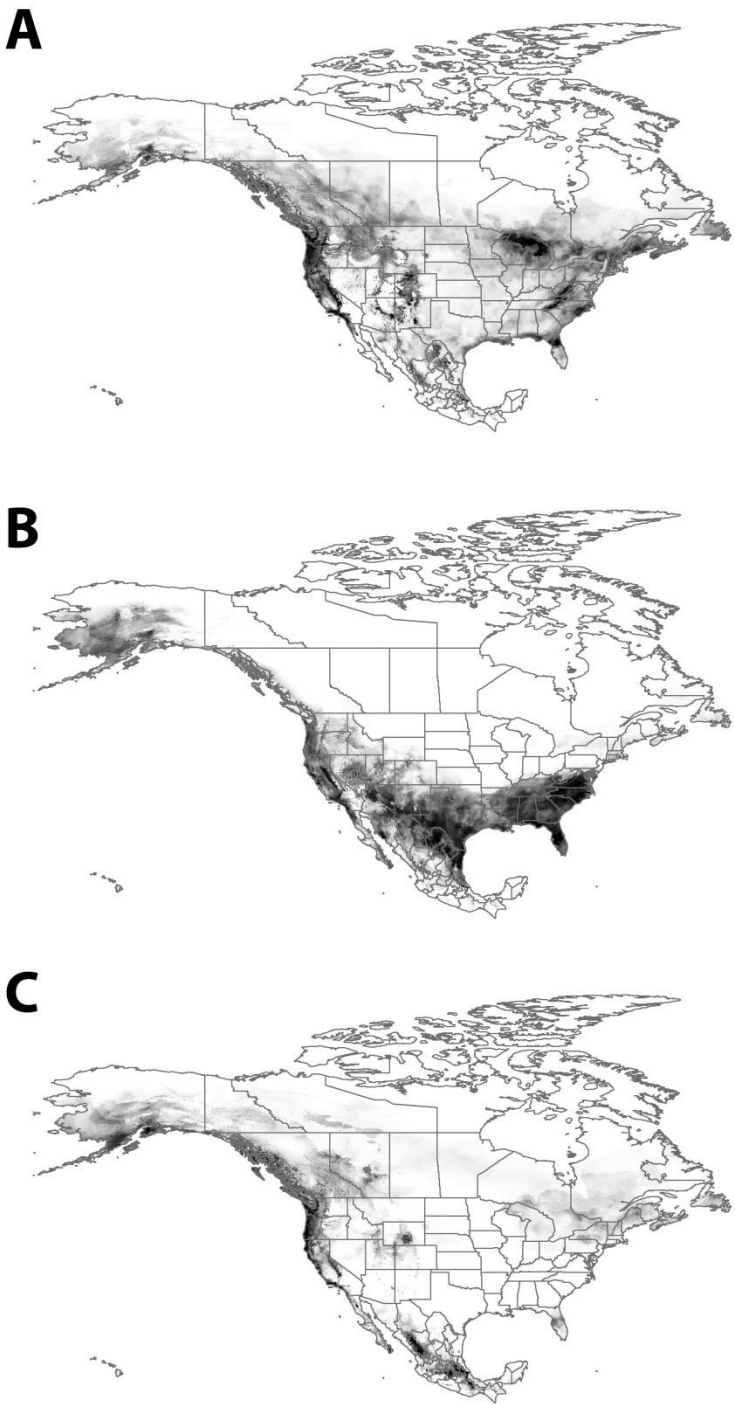

Figure 3.5- American black bear species distribution model for current (A) climate, and hindcasts for 21 kya (B) and 120 kya (C). Probability of occurrence was on a continuous color scale where white indicates low probability and black indicates high probability.

Table 3.1- Pairwise F_{ST} based on SNP genotypes (below diagonal) and mitochondrial haplotypes (above diagonal) of black bears grouped into nine clusters across the range. Comparisons in bold were significant following Bonferroni correction ($P < 0.005$).

	AK-East[1]	CIH	Great Lakes	Northeast	Southeast	AK-West[2]	West	Pacific Coast	Southwest
AK-East[1]	*	**0.591**	**0.237**	**0.494**	**0.573**	**0.127**	**0.174**	**0.912**	**0.119**
CIH	**0.161**	*	**0.534**	**0.588**	**0.619**	**0.525**	**0.424**	**0.861**	**0.545**
Great Lakes	**0.122**	**0.062**	*	**0.148**	**0.143**	**0.199**	**0.190**	**0.848**	0.081
Northeast	**0.143**	**0.080**	**0.025**	*	**0.039**	**0.421**	**0.319**	**0.862**	**0.322**
Southeast	**0.184**	**0.118**	**0.064**	**0.037**	*	**0.466**	**0.318**	**0.917**	**0.360**
AK-West[2]	**0.219**	**0.352**	**0.316**	**0.329**	**0.366**	*	0.025	**0.687**	**0.107**
West	**0.164**	**0.264**	**0.229**	**0.243**	**0.280**	**0.156**	*	**0.484**	**0.154**
Pacific Coast	0.276	0.393	0.360	0.373	0.408	0.218	0.098	*	**0.860**
Southwest	**0.322**	**0.429**	**0.393**	**0.386**	**0.414**	**0.299**	**0.138**	0.197	*

[1]- For mitochondrial haplotypes, this group only comprises samples from AK-Kenai Peninsula and Yukon.

[2]- For mitochondrial haplotypes, this group comprises samples both from both the eastern and western associated AK-ABC samples as they could not be separated without sequencing all of the samples.

Table 3.2- Admixed and inferred source population parameters for two-way admixtures inferred from 500 replicates in MIXMAPPER. Parameters include the 95% bootstrap confidence interval for the proportion of ancestry from Source 1 (α), the drift distance for both source branches (including drift before admixture / total along the source branch), and drift following admixture.

Admixed	Source 1	Source 2	Replicates	α	Branch 1	Branch 2	Mixed Drift
West	Southwest	CIH	442	0.599-0.822	0.260-0.421 / 0.431	0.002-0.106 / 0.215	0.005-0.093
	Great Lakes[1]	Southwest	307	0.092-0.203	0.013-0.037 / 0.136	0.214-0.296 / 0.431	0.000-0.000
Great Lakes	Southeast	AK-East	307	0.915-0.997	0.013-0.037 / 0.136	0.004-0.029 / 0.029	0.002-0.041
	Southeast	AK-East	75	0.896-0.991	0.013-0.038 / 0.136	0.013-0.138 / 0.139	0.005-0.051
	Southeast	AK-East	55	0.926-0.996	0.017-0.036 / 0.135	0.000-0.000 / 0.000	0.006-0.038
Northeast	Southeast	AK-East	201	0.935-0.995	0.049-0.080 / 0.136	0.002-0.026 / 0.026	0.000-0.019
	Southeast	AK-East	119	0.930-0.989	0.049-0.081 / 0.133	0.020-0.132 / 0.136	0.000-0.022
	Southeast	CIH	79	0.380-0.764	0.078-0.133 / 0.138	0.186-0.213 / 0.215	0.008-0.031
	Southeast	Southwest	81	0.954-0.998	0.045-0.075 / 0.133	0.006-0.435 / 0.435	0.000-0.025

[1]- Denotes the three population test for admixture where following admixture creating the Great Lakes, that population serves as the source for a second admixture event into the West.

LITERATURE CITED

Austerlitz F, JungMuller B, Godelle B, Gouyon PH (1997) Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology* **51**, 148-164.

Axelrod DI (1985) Rise of the grassland biome, central North America. *The Botanical Review* **51**, 163-201.

Bidon T, Janke A, Fain SR, *et al.* (2014) Brown and polar bear Y chromosomes reveal extensive male-biased gene flow within brother lineages. *Molecular Biology and Evolution*.

Bon C, Caudy N, de Dieuleveult M, *et al.* (2008) Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17447-17452.

Braby MF, Eastwood R, Murray N (2012) The subspecies concept in butterflies: has its application in taxonomy and conservation biology outlived its usefulness? *Biological Journal of the Linnean Society* **106**, 699-716.

Burbrink FT, Fontanella F, Alexander Pyron R, Guiher TJ, Jimenez C (2008) Phylogeography across a continent: The evolutionary and demographic history of the North American racer (Serpentes: Colubridae: *Coluber constrictor*). *Molecular Phylogenetics and Evolution* **47**, 274-288.

Byun SA, Koop BF, Reimchen TE (1997) North American black bear mtDNA phylogeography: Implications for morphology and the Haida Gwaii glacial refugium controversy. *Evolution* **51**, 1647-1653.

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

Chavez AS, Maher SP, Arbogast BS, Kenagy GJ (2014) Diversification and gene flow in nascent lineages of island and mainland North American tree squirrels (*Tamiasciurus*). *Evolution* **68**, 1094-1109.

Choi EH, Kim SK, Ryu SH, Jang KH, Hwang UW (2010) Mitochondrial genome phylogeny among Asiatic black bear *Ursus thibetanus* subspecies and comprehensive analysis of their control regions. *Mitochondrial DNA* **21**, 105-114.

Cornuet J-M, Pudlo P, Veyssier J, *et al.* (2014) DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using single nucleotide polymorphism, DNA sequence, and microsatellite data. *Bioinformatics* **30**, 1187-1189.

Dabney J, Knapp M, Glocke I, *et al.* (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*.

Delisle I, Strobeck C (2002) Conserved primers for rapid Sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears. *Molecular Biology and Evolution* **19**, 357-361.

Demboski JR, Stone KD, Cook JA (1999) Further perspectives on the Haida Gwaii Glacial refugium. *Evolution* **53**, 2008-2012.

Dixon JD, Oli MK, Wooten MC, *et al.* (2007) Genetic consequences of habitat fragmentation and loss: the case of the Florida black bear (*Ursus americanus floridanus*). *Conservation Genetics* **8**, 455-464.

Drummond A, Ashton B, Buxton S, *et al.* (2011) Geneious v5.4. http://www.geneious.com.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *Plos Biology* **4**, 699-710.

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology* **18**, 2930-2933.

Etter PD, Bassham S, Hohenlohe PA, Johnson E, Cresko WA (2011) SNP discovery and genotyping for evolutinary genetics using RAD sequencing. In: *Molecular Methods for Evolutionary Genetics, Methods in Molecular Biology* (eds. Orgogozo V, Rockman MV). Springer Science.

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.

Galbreath KE, Hafner DJ, Zamudio KR, Agnew K (2010) Isolation and introgression in the Intermountain West: contrasting gene genealogies reveal the complex biogeographic history of the American pika (*Ochotona princeps*). *Journal of Biogeography* **37**, 344-362.

GBIF (2013) *GBIF Backbone Taxonomy.* Global Biodiversity Information Facility. http://www.gbif.org/species/2433407

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record. *Molecular Ecology* **19**, 1877-1897.

Hallatschek O, Hersen P, Ramanathan S, Nelson DR (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19926-19930.

Heath TA (2012) A hierarchical Bayesian model for calibrating estimates of species divergence times. *Systematic Biology*.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.

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.

Hofer T, Ray N, Wegmann D, Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have cccurred by drift during range expansions than by selection. *Annals of Human Genetics* **73**, 95-108.

Hou WR, Chen Y, Wu X, *et al.* (2007) A complete mitochondrial genome sequence of Asian black bear Sichuan subspecies (*Ursus thibetanus mupinensis*). *International Journal of Biological Sciences* **3**, 85-90.

Hunt Jr RM (1998) Ursidae. In: *Evolution of Tertiary Mammals of North America* (eds. Janis CM, Scott KM, Jacobs LL), pp. 174-189. Cambridge University Press, Cambridge, UK.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267.

Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications* **13**, 235-248.

Krause J, Unger T, Nocon A, *et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *Bmc Evolutionary Biology* **8**.

Kurten B, Anderson E (1980) *Pleistocene Mammals of North America* Columbia University Press, New York.

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.

Lariviere S (2001) *Ursus americanus*. *Mammalian Species* **647**, 1-11.

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *Plos Genetics* **8**, e1002453.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496.

Lindqvist C, Schuster SC, Sun Y, *et al.* (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences*.

Lipson M, Loh P-R, Levin A, *et al.* (2013) Efficient moment-based inference of admixture parameters and sources of gene flow. *Molecular Biology and Evolution* **30**, 1788-1802.

Lipson M, Loh P-R, Patterson N, *et al.* (2014) Reconstructing Austronesian population history in island Southeast Asia. *Nat Commun* **5**.

Liu S, Lorenzen Eline D, Fumagalli M, *et al.* (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* **157**, 785-794.

McVean G (2009) A genealogical interpretation of principal components analysis. *Plos Genetics* **5**, e1000686.

Meirmans PG (2012) The trouble with isolation by distance. *Molecular Ecology* **21**, 2839-2846.

Miller W, Schuster SC, Welch AJ, *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences* **109**, E2382–E2390.

Onorato DP, Hellgren EC, Van Den Bussche RA, Doan-Crider DL (2004) Phylogeographic patterns within a metapopulation of black bears (*Ursus americanus*) in the American Southwest. *Journal of Mammalogy* **85**, 140-147.

Patterson N, Moorjani P, Luo Y, *et al.* (2012) Ancient admixture in human history. *Genetics* **192**, 1065-1093.

Petit RJ, Aguinagalde I, de Beaulieu J-L, *et al.* (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**, 1563-1565.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.

Platt A, Horton M, Huang YS, *et al.* (2010) The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet* **6**, e1000843.

Posada D (2004) Collapse v1.2.

Puckett EE, Kristensen TV, Wilton CM, *et al.* (2014) Influence of drift and admixture on population structure of American black bears (*Ursus americanus*) in the Central Interior Highlands, USA 50 years after translocation. *Molecular Ecology* **23**, 2414-2427.

Ratan A (2009) *Assembly algorithms for next-generation sequence data*, Pennsylvania State University.

Reding DM, Bronikowski AM, Johnson WE, Clark WR (2012) Pleistocene and ecological effects on continental-scale genetic differentiation in the bobcat (*Lynx rufus*). *Molecular Ecology* **21**, 3078-3093.

Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution* **29**, 165-176.

Rowe KC, Heske EJ, Brown PW, Paige KN (2004) Surviving the ice: Northern refugia and postglacial colonization. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10355-10359.

Ruiz-Lopez MJ, Goldberg TL, Chapman CA, *et al.* (Accepted) Identification of SNP markers for the endangered Ugandan red colobus (*Procolobus rufomitratus tephrosceles*) using RAD sequencing. *Molecular Ecology Resources*.

Sacks BN, Mitchell BR, Williams CL, Ernest HB (2005) Coyote movements and social structure along a cryptic population genetic subdivision. *Molecular Ecology* **14**, 1241-1249.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* **78**, 629-644.

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.

Steele PR, Hertweck KL, Mayfield D, *et al.* (2012) Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* **99**, 330-348.

Swenson Nathan G, Howard Daniel J (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist* **166**, 581-591.

Torstrom SM, Pangle KL, Swanson BJ (2014) Shedding subspecies: The influence of genetics on reptile subspecies taxonomy. *Molecular Phylogenetics and Evolution* **76**, 134-143.

Tsai C-L, Chou Y-C, Shih C-C, *et al.* (2009) The complete mitochondrial genome of the Formosan black bear (*Ursus thibetanus formosanus*). *Zootaxa* **1971**, 50-58.

Van Den Bussche RA, Lack JB, Onorato DP, *et al.* (2009) Mitochondrial DNA phylogeography of black bears (*Ursus americanus*) in central and sourthern North America: Conservatoin implications *Journal of Mammalogy* **90**, 1075-1082.

Varas-Nelson AC (2010) *Conservation genetics of black bears in Arizona and Northern Mexico* PhD, University of Arizona.

Wooding S, Ward R (1997) Phylogeography and Pleistocene evolution in the North American black bear. *Molecular Biology and Evolution* **14**, 1096-1105.

Wu C-H, Drummond AJ (2011) Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. *Genetics* **188**, 151-164.

Yu L, Li YW, Ryder OA, Zhang YP (2007) Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. *Bmc Evolutionary Biology* **7**.

Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 561-564.

CHAPTER 4

Comparison of SNP and microsatellite genotyping panels for spatial assignment of individuals to natal range: A case study using the American black bear (*Ursus americanus*)

Emily E. Puckett and Lori S. Eggert

Division of Biological Sciences, 226 Tucker Hall, University of Missouri, Columbia, MO 65211

ABSTRACT

The use of statistical spatial smoothing methods to infer the natal population for a sample of unknown origin from its genotype was an exciting development for wildlife forensics. These methods trace illegally harvested samples to their natal origin so that either prosecution or additional enforcement resources may be allocated to the correct jurisdiction. Spatial smoothing studies have thus far used microsatellite genotyping panels; however, many researchers have shifted towards collecting SNP datasets. We investigated the accuracy and precision of two methods (spatial smoothing and principal components regression) with datasets that varied in marker type (microsatellites or SNPs), loci number, and number of training samples. Accuracy varied between datasets where the median difference between true and estimated geographic locations ranged from 174 – 1,212 km. A dataset using 1,000 SNP loci and the spatial smoothing method was the most precise and had a median accuracy of 201 km. We observed that natal inference from SNPs was more accurate and precise than when estimated using microsatellites, and that large numbers of SNP loci could overcome having few samples in the training dataset. Our results suggest cautious interpretation of natal assignments, as only one dataset assigned 50% of test samples to the correct management jurisdiction

and this dataset was neither the most accurate nor precise. The use of natal inference as a tool for law enforcement may work well at the regional level, given sufficient input data; however, we clearly observe limits on the spatial scale of inference.

INTRODUCTION

International wildlife trade has been estimated at $300 billion annually for legal products (Roe 2008) with an additional $8-20 billion in illegal products (Ferrier 2009). Exact estimates of the volume and value of illegal trade are unknown due to the illicit nature of the industry; however, as legal trade volume has increased in the past 20 years analysts believe there has been a parallel increase in illegal trade (Wyler and Sheikh 2008). Rising illegal trade volume has been associated with both rising wealth and global immigration, as immigrants import familiar products from their home countries to their residence (Wyler and Sheikh 2008). Improving infrastructure, particularly access to roads that transport wildlife products to market, may also contribute to increasing trade (Corlett 2007). Finally, human population growth may increase local demand for wildlife products, straining biodiversity without affecting international markets (Corlett 2007).

In wildlife forensics cases, inference of the natal population of an individual can be achieved through a variety of assignment methods (Manel, Gaggiotti, and Waples 2005). Both the accuracy and precision of the inferred natal location are important as they determine the jurisdiction from which wildlife was harvested and who may bring suit against poachers or which jurisdictions need additional enforcement resources. Assignment methods rely on sampling all of the potential natal populations which an

individual may belong. A useful extension of discrete population assignment employed spatial smoothing of allele frequencies between populations (Wasser et al. 2004; Wasser et al. 2007). This method was a significant development for the application of assignment methods to wildlife trade product tracking because it did not require all populations to be sampled. However, one assumption of spatially smoothed allele frequencies is that populations are related to each other by simple isolation-by-distance processes (Meirmans 2012).

Trade of bear species is concentrated around their value as trophies and for traditional Asian medicine (Burgress, Stoner, and Foley 2014); however, bears face additional population pressure from killing for predator defense (Fredriksson 2005; Can et al. 2014; Qashqaei, Karami, and Etemad 2014) and from habitat degradation which exacerbates declining population sizes. Asiatic black (*Ursus thibetanus*), sun (*U. malayanus*), and brown (*U. arctos*) bears are the primary species involved in trade of bile, gallbladders, skin, skulls, teeth, claws, and paws for traditional medicine (Foley, Stengel, and Shepherd 2011). These species are native to Asia where the majority of the production and consumption of bear parts occurs (Foley, Stengel, and Shepherd 2011). Trade of bears and their viscera is illegal under CITES as all eight species are listed under either Appendix I (threatened with extinction) or II (listed due to similarity of appearance). From 1991 to 2013, there were 304 illegal American black bear (*U. americanus*) confiscations comprised of over 14,800 products and 687 kg of meat (UNEP-WCMC 2013). This data show that while the market for American black bear products is smaller

than that for Asian species (Burgress, Stoner, and Foley 2014), illegal trade does exist for the species.

We leveraged our extensive collection of American black bear samples to test the population allele frequency based method SCAT (Wasser et al. 2004; Wasser et al. 2007) and an individual based principal components (PC) regression method (Jolliffe 2002). Studies that employ spatially smoothed allele frequencies often validate the accuracy of the method for their system using leave-one-out cross validation (Ghobrial et al. 2010; Dellicour et al. 2011; Mondol et al. 2014). This paper extends previous work by comparing accuracy of spatial assignment methods under varying scenarios of genetic marker type (microsatellites or single nucleotide polymorphisms, SNPs), marker number, and training sample number. Marker selection is critical for the generation of an effective assignment panel. Despite individual microsatellite loci having higher informativeness than individual SNP loci (Liu et al. 2005), when multilocus genotypes were used, SNPs consistently outperform microsatellites for population structure and assignment analyses due to strong segregation between populations (Helyar et al. 2011). The primary advantage of microsatellites include higher mutation rate leading to more alleles, thus information content, per locus; where the primary disadvantage of microsatellites is that alleles may differ by research group based on sequencing platform and scoring protocols. The advantages of SNPs include genome-wide distribution and standard scoring following identification. The generation of an efficient genotyping panel for natal population inference problems will maximize information content while minimizing marker number both due to genotyping costs and analysis time.

MATERIALS AND METHODS

Datasets

We used microsatellite and SNP genotypes previously collected for a phylogeographic study of the American black bear (Puckett et al. In Review). Briefly, 506 samples were genotyped at 15 microsatellite loci and 94 of those samples were genotyped at 29,166 SNP loci identified using RAD-Seq (Table A4.1). We built eight datasets to test the accuracy and precision of natal location inference of samples based on varying genetic markers and numbers of loci. Each sample was assigned to a sampling area based on habitat patches within states or provinces, and sampling areas were assigned to regional clusters as identified in Puckett *et al.* (In Review) (Table A4.1).

SNPs- We first tested three sampling strategies (*Equal*, *Proportional*, and *Top*) to select a tractable number (n = 200) of SNP loci for assignment from the total available pool of 29k markers. We ran a principal component analysis (PCA) on the 29k data with the smartpca program within EIGENSOFT (Patterson, Price, and Reich 2006; Price et al. 2006). We calculated the number of significant eigenvectors using Tracy-Widom statistics with the twstats program within EIGENSOFT. Our three sampling strategies selected variable numbers of SNPs from the significant principal components (PC) (Monzón, Kays, and Dykhuizen 2014). In the *Equal* dataset, an equal number of SNPs were selected from each significant PC. In the *Proportional* dataset, SNPs were selected proportionally to the eigenvalue associated with each significant PC. In the *Top* dataset, an equal number of SNPs were selected from the top two PCs. Loci with the highest absolute PC loading values were selected from PCs in descending order where SNPs

from PCs that explained more variance were selected before those that explained less variance. If a SNP was already selected for inclusion in the study, then an additional SNP was selected for that PC so that the number of unique SNPs per PC would equal the desired number (e.g. n = 200 or 1000). Subsequently, we ran PCA on the SNPs selected for inclusion to observe the decomposition patterns for individual datasets. In all PCAs, test samples were excluded from the initial analysis and projected into PCA space based on the training sample data.

Following observation of accuracy between the three sampling strategies, we selected the *Equal* strategy for further analysis as it had the lowest median difference between true and estimated locations (Table A4.2). We then constructed datasets with 40 and 1000 SNPs to understand how marker number affected accuracy. All SNP datasets had 87 training samples and seven test samples (Table A4.1).

Microsatellites- The *Micro* dataset included all 506 samples genotyped at all microsatellite loci. Twenty samples were selected as test samples and included the seven samples used as test samples in the SNP datasets (Table A4.1). The *PartialMicro* dataset used the same 20 test samples as the *Micro* dataset and the same 87 training samples as the SNP dataset. In this way we were able to compare the decrease in sample number from the *Micro* to *PartialMicro* datasets, and alternately the variability between marker types. We ran a PCA on both datasets to observe the decomposition patterns and compare to the SNP datasets.

Natal range assignment

We estimated natal location using both the population based SCAT v1.0.2 (Wasser et al. 2004) and an individual based PC regression (Jolliffe 2002). SCAT assumes that allele frequencies vary continuously across a landscape and may be estimated by smoothing frequencies between sampled populations. Following creation of the smoothed allele frequency surface, the program then estimates the spatial location for samples of unknown origin. Due to the smoothing in continuous space, we created a boundary file of the contemporary black bear range that linked discontinuous habitat but did not include all of North America, particularly the Great Plains and southwestern deserts. We ran three repetitions of SCAT, each starting from a different seed for $5x10^4$ MCMC steps, retaining every 500^{th} iteration for a total of 200 iterations from which we discarded the first 100 retained iterations as burnin. In SCAT, we grouped the 33 sampling areas into six regional clusters (Alaska, West, Southwest, Southeast, Northeast and Great Lakes combined, and Central Interior Highlands) based on our previous results (Puckett et al. In Review).

For the individual based PC regression, latitude and longitude of the training samples were independently regressed against the significant PC loading scores obtained for each dataset from PCA using a linear model in program R (R Core Team 2013). The coefficients of the linear model and individual PC loadings for each test sample were used to estimate latitude and longitude using the formula:

$$L = x + \sum_{i=1}^{K} b_j * PC_j$$

Where L was latitude or longitude, K was the number of significant PCs estimated using Tracy-Widom statistics, j was the individual sample, PC was the eigenvector, b was the slope, and x was the x-intercept.

To estimate accuracy, we calculated the Euclidian distance between the estimated and true sample locations with the raster package (Hijmans et al. 2014) in R, then calculated summary statistics. For data from SCAT, the estimated location was the median center from the retained locations (n = 300) which was calculated using the spatial statistics tools within ArcMAP v10.0 (ESRI, Redlands, CA). To estimate precision of SCAT, the retained locations for each test sample were converted to a kernel density (kde) using the GEOSPATIAL MODELING ENVIRONMENT v0.7.2 (GME) (Beyer 2012), with the plugin bandwidth which was independently estimated for each test sample and sampling scheme, cell size of five, and Gaussian kernel. The output was piped to the isopleth function where the 50% and 90% isopleths were calculated. Since the PC regression method only output a single point estimate, precision could not be estimated for that method. ANOVAs were applied to comparable datasets (*Micro* and *PartialMicro*; and *PartialMicro* [using the same seven test samples as the SNP datasets], *Equal200*, and *Equal1000*; additionally 29k was included for the PC regression method) to test for significant differences in accuracy and precision.

RESULTS

Principal component analyses

Using the 29k SNP dataset, the first 10 PCA axes were statistically significant (Table A4.3), where the first PC described 10.9% of the variation in the data and distinguished between the eastern and western genomic clusters (Figure 4.1A). The second PC described 3.3% of the variation and was oriented along a north-south geographic axis.

We ran the *Equal* datasets with 40, 200, and 1000 loci through PCA to confirm that the shape of the decomposition was equivalent between datasets; however, it was not (Figure A4.1). The PCA with 40 loci did not resemble the shape of the other datasets, and no PCs were significant indicating lack of resolution in the data; therefore, this dataset was not used in further analyses. The 29k and *Equal200* SNP loci datasets had broadly the same shape, where the eastern and western samples separated along PC1, and the Alaskan samples separated along PC2. The *Equal200* PCA appeared to lose resolution by broadly clumping samples in the east and west. While the eastern, western, and Alaskan clusters continued to separate in the *Equal1000* PCA, PC2 was characterized by separation between two samples from Alaska-ABC Islands west cluster and a sample from both Oregon and Montana.

For the *Micro* and *PartialMicro* datasets, the first seven and three PCA axes were significant, respectively (Table A4.3). The shape of the PCA for the *Micro* dataset was markedly different from that of the 29k SNP dataset (Figure 4.1). While the PCA for *Micro* separated eastern and western samples along PC1, the pattern of separation

differed where samples from Missouri and the Ozark Mountains appeared as the most differentiated eastern samples; whereas samples from the southeast were the most differentiated in the 29k SNP dataset. The *Micro* PCA did not differentiate samples from the southeast and northeast. Additionally, the Alaskan lineage was shown as an admixed cluster between eastern and western genotypes versus its own cluster (Figure 4.1). The *PartialMicro* PCA had little geographic resolution (Figure A4.1C).

Accuracy and Precision

Spatial Smoothing- Accuracy varied between the datasets but was highest using the *Equal1000* dataset and spatial smoothing which gave a median distance between true and estimated locations of 201 km (Table 4.1, Figures 4.2, 4.3, and A4.3). For the spatial smoothing method, there were not significant differences ($P = 0.116$) in distance between *Micro* and *PartialMicro* datasets (with 20 test samples), while there were differences between the *PartialMicro*, *Equal200*, and *Equal1000* datasets ($P = 0.030$). Estimates of precision were significantly different for the 50% ($P = 0.011$) but not the 90% ($P = 0.053$) isopleths when comparing the *Micro* and *PartialMicro* datasets. However, when comparing the *PartialMicro* to the *Equal200* and *Equal1000* datasets, the area of both the 50% ($P = 0.004$) and 90% ($P = 0.0004$) isopleths were significantly different.

PC Regression- We compared accuracy for the individual sample based PC regression with the 29k, *Equal* 200 and 1000, and *Micro* datasets (Figures A4.2 and A4.3). Similar to the spatial smoothing results, there was no significant difference between the *Micro*

and *PartialMicro* datasets ($P = 0.161$), but were differences for the *PartialMicro*, *Equal200*, *Equal1000*, and 29k datasets ($P = 0.036$).

DISCUSSION

We aimed to understand how sampling density and marker selection affected accuracy and precision when estimating the location of samples of unknown origin. In the two microsatellite datasets used to compare sampling density, neither accuracy nor precision were significantly different; however, this appears to be an effect of several poorly estimated samples as the medians were lower when training sample size increased. Intuitively, this suggests that larger sample sizes will better estimate natal locations. That said, marker type was able to overcome deficiencies in sample size to produce both the most accurate and precise results when using the spatial smoothing method. Finally, we observed that very high numbers of SNPs (at least at the low sampling density) produced the most accurate and precise estimates of natal location. We initially selected 200 SNPs for our analyses based on estimates that microsatellites have approximately 10-fold greater informativness per marker than SNPs (Liu et al. 2005). However, several studies observed higher resolution of population structure (Liu et al. 2005; Hess, Matala, and Narum 2011; Rasic et al. 2014) and pedigree assignment (Schopen et al. 2008; Santure et al. 2010) when using SNP panels only two to three times larger than a microsatellite panel. Our results suggest that for natal inference problems, large SNP genotyping panels are necessary. We discuss accuracy and precision, and marker selection below from a conservation management perspective.

Accuracy and Precision- While we observed individual samples which were well estimated, median estimates were hundreds of kilometers from sampling sites. Joshi *et al.* (2013) interpreted the difference between sampling site and estimated natal location to represent dispersal; similarly, Jay *et al*. (2013) interpreted differences between known and estimated locations as historic signatures of population movement. While both are plausible hypotheses, we interpreted our differences as a representation of uncertainty. Our sampled distances were equivalent to similar research using SCAT with elephants where accuracy varied with population structure from 135 – 697 km (Wasser et al. 2004), and with PC regression in humans which ranged from 250 – 510 km (Jay et al. 2013). In each of these studies, larger continuous tracks of habitat had lower accuracy; conversely isolated areas had higher accuracy. Our test sample from Ocala, Florida was consistently well estimated (Figure 4.2), which we believe was due to the isolation of that habitat patch (Dixon et al. 2007). Jay *et al*. (2013) also observed this north of the Sahara desert, a strong barrier to dispersal. Conversely, bear test samples from the southeast and northeast genetic clusters were poorly estimated and tended to converge on the same incorrect region (southeastern Ontario; Figure 4.2). We previously hypothesized that the northeast represents a post-Pleistocene range expansion from a Southeastern refugium (Puckett et al. In Review). This large and continuous habitat means that estimated allele frequencies within individual populations may look more similar than in isolated areas, although see below for a discussion of marker based population structure differences.

We tested a special case of the isolated habitat where no training samples were used, as incomplete population sampling may be a realistic scenario for applying spatial

smoothing to natal inference problems. We estimated the location of two samples collected in Mississippi as being from southeastern Ontario (Figure 4.2). Individuals in Mississippi most likely colonized from neighboring Louisiana or Florida (Simek et al. 2012) which would have been reasonable, although still inaccurate, assignment locations. Much of the appeal of the spatial estimation methods comes from the ability to assign samples to unsampled locations, and our data suggest limits to this application, at least for isolated habitat patches.

We were also interested in the accuracy of spatial assignment methods for regions with known translocations as samples may be expected to assign to source populations. From 1958 to 1968, 254 black bears were translocated from Minnesota, USA and Manitoba, Canada to Arkansas, USA (Smith and Clark 1994). Puckett *et al*. (2014) observed admixture between the population on the landscape at the time of the reintroduction and the translocated individuals. This was also apparent within our PCA data (Figure 4.1) especially in the microsatellite PCA where differences between Great Lakes and Central Interior Highlands genotypes account for significant separation along the top two PCs. Despite these differences in underlying PCA structure, source and reintroduced areas were both accurately estimated. However, there were differences between the spatial smoothing and PC regression methods, where the PC regression estimated a more northernly location for the reintroduced sample as if it were a source sample (Figures 4.2 and A4.1). This result highlights the sensitivity of the PC regression method to the underlying genomic results, where the spatial smoothing method accounts for geography in translocation scenarios. We note that this translocation occurred 16 generations ago

and was characterized by strong genetic drift (Puckett et al. 2014), assignment in species with more recently translocated individuals may not produce the robust results observed in this study.

Despite our results being within the range of accuracy for the methods employed (SCAT and PC regression), the question of usefulness for law enforcement of confiscated viscera is pertinent. Specifically, the area and shape of states and provinces are highly variable across North America; for instance, a circle with a 200 km radius in larger western states and smaller northeastern states may not have equivalent proportions of the management jurisdiction within that defined area. From the perspective of assignments being in the "correct" state or province, using SCAT with the *Micro* dataset resulted in 50% accurate assignment to jurisdiction; all other methods and datasets performed less well although results were within the range of other spatial assignment studies (Yang et al. 2012). This has implications for interpreting results for unknown samples as large management areas may have more confidence in estimates that a sample came from their jurisdiction simply due to the fact that they have more area and lower need for precision. Notably, this disregards the accuracy and precision of any dataset and is a sole property of political boarders. Results may be less useful for understanding which of several habitat patches an animal came from within a state or province. Larger jurisdictions may manage animals in multiple habitat patches (which may or may not be genetically distinct populations) requiring a division of enforcement resources between patches. The ability to map samples to a habitat patch and not just a state or province would be useful,

although the literature strongly suggests limits on accuracy suggesting that often state or province may be the smallest unit of inference.

Marker Selection- Our initial aim was to compare microsatellite and SNP datasets; however, this aim was predicated on the assumption that the markers inferred the same underlying population structure. Our PCA results show that, for the American black bear, microsatellites and SNPs infer substantially different population structure (Figure 4.1). Specifically, the microsatellites showed less clustering in PCA space than the SNPs; this was also observed for domesticated dogs (Karlsson et al. 2007) and mosquitos (Rasic et al. 2014). Hess and colleagues (2011) proposed that population structure differed between microsatellites and SNPs because the former delineates recent structure and the later deep time events. However, we argue that the multi-allelic states of microsatellites, higher mutation rate, and the possibility of identity-by-state smooth over underlying population differentiation. Vali *et al*. (2008) observed that the magnitude of heterozygosity differences between populations was significantly less for microsatellites than for SNPs, meaning that PCA will not tightly cluster the data when using microsatellites. Additionally, McVean (2009) showed that coalescent time between two samples explained the distance apart in PC space. From this perspective our microsatellite data show individuals from the Southwest as equally related as the entire eastern seaboard (Figure 4.1B), which would represent an unrealistically high level of gene flow.

Our results demonstrate that it is critical to understand the underlying population structure of the training samples before use in wildlife forensic applications. We suggest that PCA is an appropriate analysis for underlying structure where lack of clustering in PCA space suggests that the data lacks resolution for wildlife forensic inference. However, without comparative data from our 29k PCA, we would not have been able to detect the lack of clustering in the microsatellites or could not have inferred that the clustering patterns from the *Equal200* and *Equal1000* SNP sets were skewed in PC space. Notably the SNP subsets were chosen from the PC loadings that best discriminated PCs within the 29k dataset, indicating that those loadings did not fully capture the variation in the data. Spatial smoothing was able to overcome some of these deficiencies as it produced the most accurate and precise natal estimates with the *Equal1000* dataset. For this to be a useful method to track trade of Asian bear species, we recommend large genome-wide genotyping panels like that obtained from RADseq, to ascertain structure and highly informative markers before selection of smaller SNP genotyping panels.

Conservation implications

We are cautious regarding the application of these methods to real world natal mapping problems. In a perfect analysis the estimated (natal) and true (sampled) locations would not be the same due to natal dispersal alone. For American black bears, average male and female dispersal varies from 40-60 and 2-7 km, respectively, although dispersal events up to 250 km have been reported (Costello 2010; Moore et al. 2014). This means that natal dispersal may account for a proportion, but not all, of the difference between true and estimated locations. An exception to this concern for spatial smoothing as an application

was shown by Mondol *et al*. (2014) where the difference between true and estimated locations from the training samples was itself a fraction of the sampled and estimated differences for test samples, thereby suggesting seizure markets far from poaching sites. However, even in this instance, the precision from the training samples was not accounted for in the final analysis. Buffering either point or kernel density estimates using either dispersal or the median accuracy of the training datasets, may provide a way to spatially account for dispersal or uncertainty (estimated by the training samples) in final analyses. While buffering would make the natal inference area larger, it may also represent a more realistic range from which the individual originated, and may implicate one or more management jurisdictions that could reallocate enforcement resources.

Above we noted the difficulty of assigning to state or province which is understandable as genetically distinct populations do not lie solely within political borders. This suggests that regional or international management agreements and law enforcement structures should be considered for illegal trade cases. We also show that the applicability of natal inference methods will rely on geographically and genomically dense genotyping efforts, where concerted international efforts to build species specific databases will aid enforcement efforts. Research efforts for many species of bears, as well as other elusive animals, rely on the use of hair snares for genetic mark-recapture estimates of population size monitoring. The hairs collected from these efforts may be used to genotype at SNP loci given recent technological developments for SNP genotyping from hair (Kraus et al. 2015). This suggests that initiating range wide genotyping projects for species of trade

concern may be fruitful if geographic coverage is large and habitat connectivity is understood across a species' range.

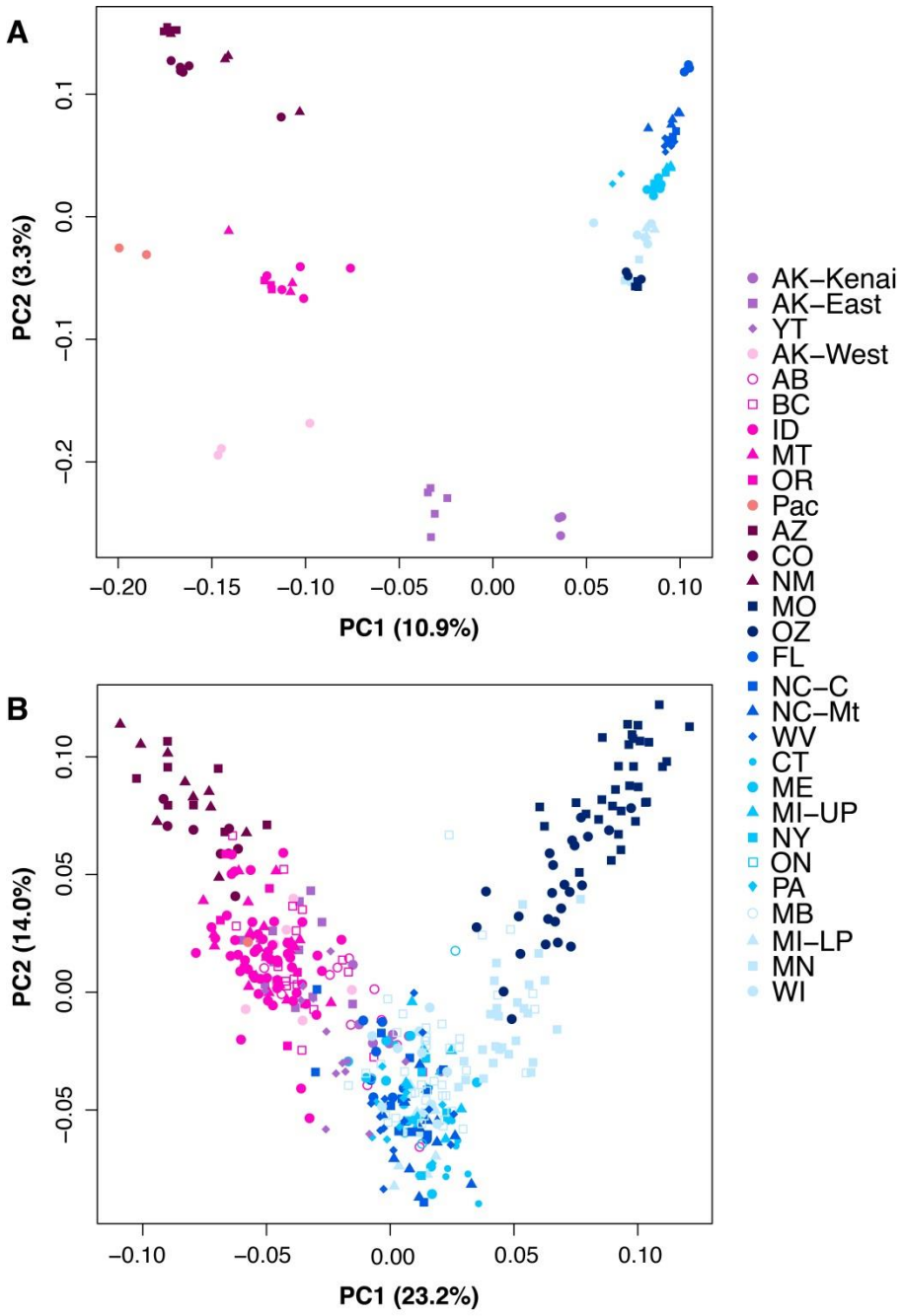

Figure 4.1- PCA of (A) 29k SNPs (n = 94) and (B) 15 microsatellites (n = 506) for black bears across the range. Sample locations were organized by symbology (abbreviations described in Table A4.1) where colors identified genomic cluster (raspberry, Southwest; pink, West; coral, Pacific Coast; light pink, Alaska- ABC Islands west cluster; purple, Alaska- ABC Islands east cluster; navy, Central Interior Highlands; light blue, Great Lakes; bright blue, Northeast; medium blue, Southeast).

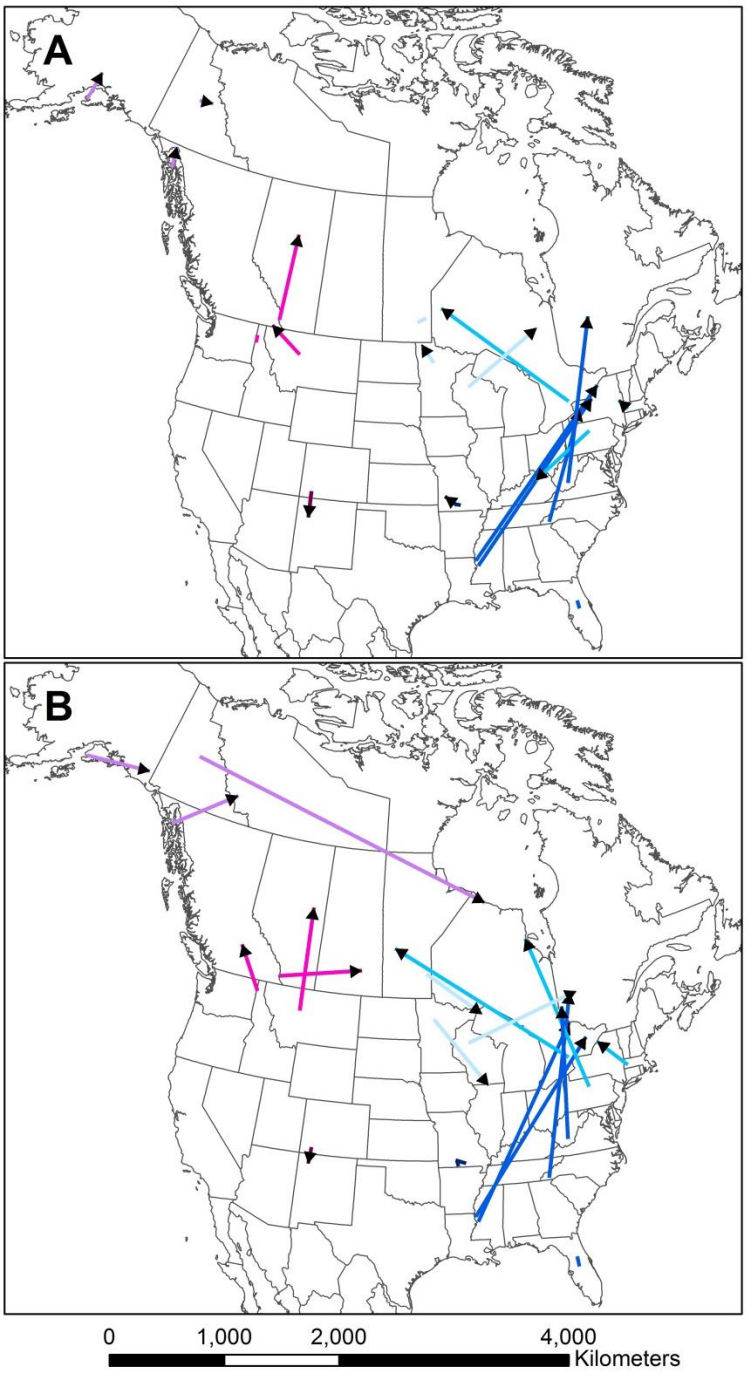

Figure 4.2- Map of true and estimated locations for test samples with the spatial smoothing method with the *Micro* (A) and *PartialMicro* (B) datasets. Arrows point towards estimated location.

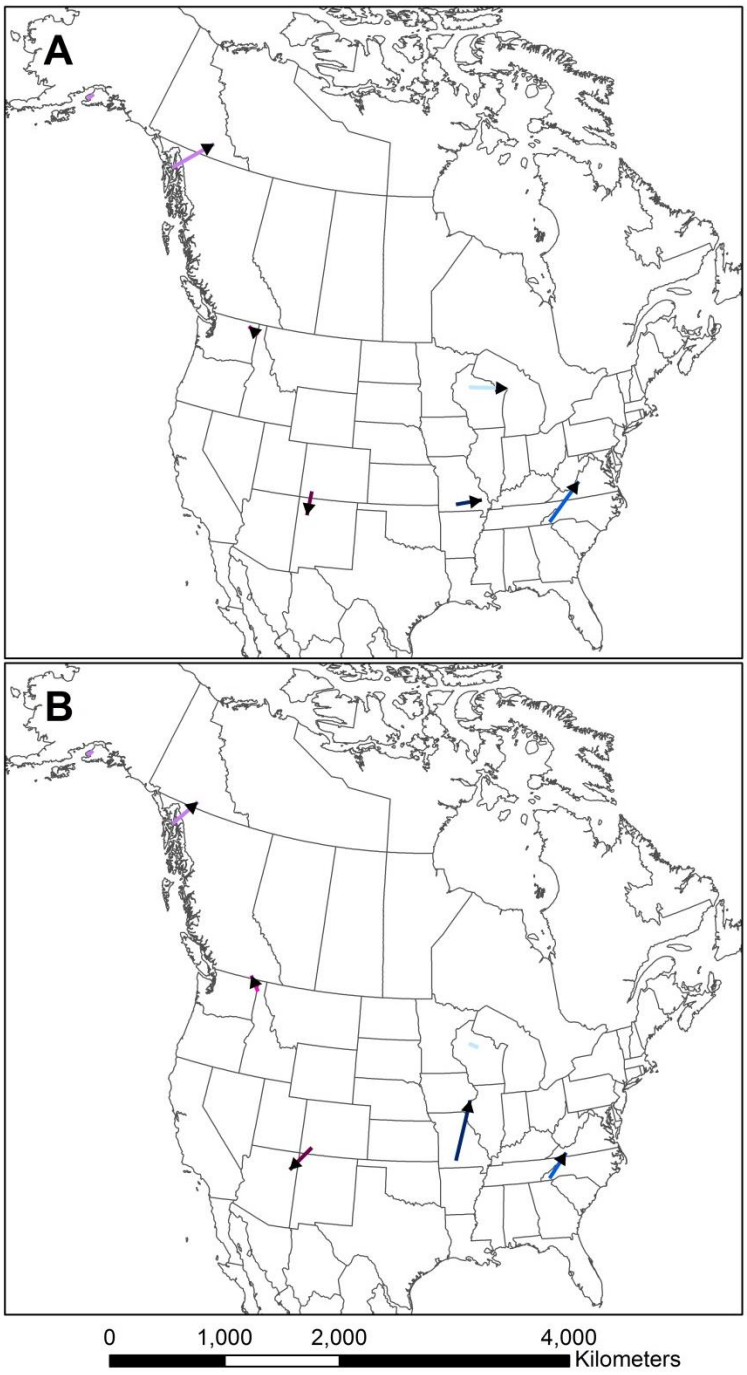

Figure 4.3- Map of true and estimated locations for test samples with the spatial smoothing method with the *Equal200* (A) and *Equal1000* (B) datasets. Arrows point towards estimated location.

Table 4.1- The median distance (km) and range between true and estimated locations using both the spatial smoothing and principal component regression methods across marker sets. Datasets varied in the type of genetic marker (microsatellite or SNP), the number of loci, and the number of training and test samples.

					Spatial Smoothing			PC Regression		
Dataset	**Marker**	**Loci**	**Training**	**Test**	**Median**	**Min**	**Max**	**Median**	**Min**	**Max**
Micro	Micro	15	486	20	249	13	1695	1212	22	4668
PartialMicro	Micro	15	87	20	669	12	5334	1809	108	6832
PartialMicro	Micro	15	87	7	662	45	1537	2675	829	6832
Equal	SNP	200	87	7	302	66	555	174	80	1903
Equal	SNP	1000	87	7	201	64	310	763	130	7769
29k	SNP	29,166	87	7	-	-	-	291	55	1133

Table 4.2- The median and range of the area (km^2) for the 50% and 90% isopleths estimated for test samples in each dataset from 300 point locations using the spatial smoothing method.

					50% Isopleth			90% Isopleth		
Dataset	**Marker**	**Loci**	**Training**	**Test**	**Median**	**Min**	**Max**	**Median**	**Min**	**Max**
Micro	Micro	15	486	20	9.60×10^{5}	0	2.55×10^{6}	1.61×10^{6}	1.00×10^{-6}	1.45×10^{7}
PartialMicro	Micro	15	87	20	9.18×10^{5}	1.00×10^{-6}	3.10×10^{6}	4.44×10^{6}	2.53×10^{5}	1.53×10^{7}
PartialMicro	Micro	15	87	7	8.59×10^{5}	0	2.55×10^{6}	3.87×10^{6}	8.87×10^{5}	7.30×10^{6}
Equal	SNP	200	87	7	1.00×10^{-6}	1.00×10^{-6}	1.48×10^{5}	5.42×10^{5}	1.00×10^{-6}	8.55×10^{5}
Equal	SNP	1000	87	7	1.00×10^{-6}	0	3.00×10^{-6}	2.00×10^{-6}	1.00×10^{-6}	1.41×10^{5}

LITERATURE CITED

Beyer, Hawthorne L. 2012. Geospatial Modelling Environment: http://www.spatialecology.com/gme/.

Burgress, Elizabeth A, Sarah S Stoner, and Kaitlyn-Elizabeth Foley. 2014. Brought to bear: An analysis of seizures across Asia (2000-2011). Petaling Jaya, Selangor, Malaysia: TRAFFIC.

Can, Özgün Emre, Neil D'Cruze, David L. Garshelis, John Beecham, and David W. Macdonald. 2014. Resolving human-bear conflict: A global survey of countries, experts, and key factors. *Conservation Letters* 7 (6):501-513.

Corlett, Richard T. 2007. The impact of hunting on the mammalian fauna of tropical Asian forests. *Biotropica* 39 (3):292-303.

Costello, C. M. 2010. Estimates of dispersal and home-range fidelity in American black bears. *Journal of Mammalogy* 91 (1):116-121.

Dellicour, S., A. C. Frantz, M. Colyn, S. Bertouille, F. Chaumont, and M. C. Flamand. 2011. Population structure and genetic diversity of red deer (*Cervus elaphus*) in forest fragments in north-western France. *Conservation Genetics* 12 (5):1287-1297.

Dixon, J. D., M. K. Oli, M. C. Wooten, T. H. Eason, J. W. McCown, and M. W. Cunningham. 2007. Genetic consequences of habitat fragmentation and loss: the case of the Florida black bear (*Ursus americanus floridanus*). *Conservation Genetics* 8 (2):455-464.

Ferrier, Peyton. 2009. The economics of agricultural and wildlife smuggling, edited by U. D. o. Agriculture. Washington, DC: Economics Research Service.

Foley, Kaitlyn-Elizabeth, Carrie J. Stengel, and Chris R. Shepherd. 2011. Pills, powders, vials, and flakes: The bear bile trade in Asia. Petaling Jaya, Selangor, Malaysia: TRAFFIC Southeast Asia.

Fredriksson, Gabriella. 2005. Human–sun bear conflicts in East Kalimantan, Indonesian Borneo. *Ursus* 16 (1):130-137.

Ghobrial, L., F. Lankester, J. A. Kiyang, A. E. Akih, S. de Vries, R. Fotso, E. L. Gadsby, P. D. Jenkins, Jr., and M. K. Gonder. 2010. Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. *BMC Ecol* 10:2.

Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. I. Taylor, R. Ogden, M. T. Limborg, A. Cariani, G. E. Maes, E. Diopere, G. R. Carvalho, and E. E. Nielsen. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* 11:123-136.

Hess, J. E., A. P. Matala, and S. R. Narum. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources* 11:137-149.

raster: Geographic data analysis and modeling 2.3-12. CRAN.

Jay, Flora, Per Sjödin, Mattias Jakobsson, and Michael G.B. Blum. 2013. Anisotropic isolation by distance: The main orientations of human genetic differentiation. *Molecular Biology and Evolution* 30 (3):513-525.

Jolliffe, Ian T. 2002. *Principal Component Analysis*. 2nd ed, *Springer Series in Statistics*: Springer.

Joshi, Aditya, Srinivas Vaidyanathan, Samrat Mondol, Advait Edgaonkar, and Uma Ramakrishnan. 2013. Connectivity of tiger (*Panthera tigris*) populations in the human-influenced forest mosaic of central India. *Plos One* 8 (11):e77980.

Karlsson, Elinor K., Izabella Baranowska, Claire M. Wade, Nicolette H. C. Salmon Hillbertz, Michael C. Zody, Nathan Anderson, Tara M. Biagi, Nick Patterson, Gerli Rosengren Pielberg, Edward J. Kulbokas, Kenine E. Comstock, Evan T. Keller, Jill P. Mesirov, Henrik von Euler, Olle Kampe, Ake Hedhammar, Eric S. Lander, Goran Andersson, Leif Andersson, and Kerstin Lindblad-Toh. 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* 39 (11):1321-1328.

Kraus, Robert H. S., Bridgett vonHoldt, Berardino Cocchiararo, Verena Harms, Helmut Bayerl, Ralph Kühn, Daniel W. Förster, Jörns Fickel, Christian Roos, and Carsten Nowak. 2015. A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources* 15 (2):295-305.

Liu, N., L. Chen, S. Wang, C. Oh, and H. Zhao. 2005. Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *Bmc Genetics* 6 (SUPPL.1).

Manel, S., O. E. Gaggiotti, and R. S. Waples. 2005. Assignment methods: Matching biological questions techniques with appropriate. *Trends in Ecology & Evolution* 20 (3):136-142.

McVean, Gil. 2009. A genealogical interpretation of principal components analysis. *Plos Genetics* 5 (10):e1000686.

Meirmans, Patrick G. 2012. The trouble with isolation by distance. *Molecular Ecology* 21 (12):2839-2846.

Mondol, S., V. Sridhar, P. Yadav, S. Gubbi, and U. Ramakrishnan. 2014. Tracing the geographic origin of traded leopard body parts in the Indian subcontinent with DNA-based assignment tests. *Conserv Biol*.

Monzón, J., R. Kays, and D. E. Dykhuizen. 2014. Assessment of coyote–wolf–dog admixture using ancestry-informative diagnostic SNPs. *Molecular Ecology* 23 (1):182-197.

Moore, Jennifer A., Hope M. Draheim, Dwayne Etter, Scott Winterstein, and Kim T. Scribner. 2014. Application of large-scale parentage analysis for investigating natal dispersal in highly vagile vertebrates: A case study of American black bears (*Ursus americanus*). *Plos One* 9 (3):e91168.

Patterson, Nick, Alkes L. Price, and David Reich. 2006. Population structure and eigenanalysis. *PLoS Genet* 2 (12):e190.

Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38 (8):904-909.

Puckett, Emily E., Paul D. Etter, Eric A. Johnson, and Lori S. Eggert. In Review. Phylogeographic analyses of American black bears (*Ursus americanus*) suggest four glacial refugia and complex patterns of post-glacial admixture.

Puckett, Emily E., Thea V. Kristensen, Clay M. Wilton, Sara B. Lyda, Karen V. Noyce, Paula M. Holahan, D. M. Leslie, J Beringer, Jerrold L. Belant, D. White Jr., and L. S. Eggert. 2014. Influence of drift and admixture on population structure of American black bears (*Ursus americanus*) in the Central Interior Highlands, USA 50 years after translocation. *Molecular Ecology* 23 (10):2414-2427.

Qashqaei, Ali Turk, Mahmoud Karami, and Vahid Etemad. 2014. Wildlife conflicts between humans and Brown Bears, *Ursus arctos*, in the Central Zagros, Iran. *Zoology in the Middle East* 60 (2):107-110.

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

Rasic, G., I. Filipovic, A. R. Weeks, and A. A. Hoffmann. 2014. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *Bmc Genomics* 15:275.

Roe, Dilys. 2008. Trading nature: A report, with case studies, on the contribution of wildlife trade management to sustainable livelihoods and the Millennium Development Goals. Cambridge, UK: TRAFFIC International and WWF International.

Santure, Anna W., Jessica Stapley, Alexander D. Ball, Tim R. Birkhead, Terry Burke, and J. O. N. Slate. 2010. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology* 19 (7):1439-1451.

Schopen, G. C. B., H. Bovenhuis, M. H. P. W. Visker, and J. A. M. Van Arendonk. 2008. Comparison of information content for microsatellites and SNPs in poultry and cattle. *Animal Genetics* 39 (4):451-453.

Simek, Stephanie L., Jerrold L. Belant, Brad W. Young, Catherine Shropshire, and Bruce D. Leopold. 2012. History and status of the American black bear in Mississippi. *Ursus* 23 (2):159-167.

Smith, Kimberly G. and Joseph D. Clark. 1994. Black bears in Arkansas: Characteristics of a successful translocation. *Journal of Mammalogy* 75 (2):309-320.

UNEP-WCMC. 2014. *CITES Trade Database*. CITES 2013 [cited July 23 2014]. Available from http://trade.cites.org/.

Vali, U., A. Einarsson, L. Waits, and H. Ellegren. 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* 17 (17):3808-3817.

Wasser, S. K., A. M. Shedlock, K. Comstock, E. A. Ostrander, B. Mutayoba, and M. Stephens. 2004. Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade. *Proceedings of the National Academy of Sciences of the United States of America* 101 (41):14847-14852.

Wasser, Samuel K., Celia Mailand, Rebecca Booth, Benezeth Mutayoba, Emily Kisamo, Bill Clark, and Matthew Stephens. 2007. Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proceedings of the National Academy of Sciences* 104 (10):4228-4233.

Wyler, Liana Sun and Pervaze A Sheikh. 2008. International illegal trade in wildlife: Treats and US policy, edited by C. R. Service. Washington DC.

Yang, Wen-Yun, John Novembre, Eleazar Eskin, and Eran Halperin. 2012. A model-based approach for analysis of spatial structure in genetic data. *Nat Genet* 44 (6):725-731.

CHAPTER 5

Population structure and genetic diversity across the range of the American black bear (*Ursus americanus*)

Emily E. Puckett

Division of Biological Sciences, 226 Tucker Hall, University of Missouri, Columbia, MO 65211

ABSTRACT

Species level genetic diversity is hierarchically organized where all of the diversity is encompassed at the species level and each nested division (lineages, clusters, populations, extended family groups, and families) have less of the total diversity. This hierarchical ordering also relates to population structure across the range, where larger spatial extents are associated with higher levels of organization and genetic diversity. These factors, in combination with the type of genetic marker, number of loci, and number of samples affect the level of inference that may be made in studies of population structure and genetic diversity. While studies may be combined across the range to understand genetic diversity, inference on population structure must come from studies that use same markers across the range so that the data is directly comparable. I synthesize knowledge of the hierarchical population structure and distribution of genetic diversity within American black bears. I describe the distribution of three nuclear genetic lineages and their subsequent divergence into clusters and populations across the range. I show that genetic diversity for mean number of alleles and observed heterozygosity was higher across the northern range than the southern range. I hypothesized that high habitat fragmentation in the south contributed to this pattern despite the location of glacial

refugia in both the southeast and southwest which should have higher genetic diversity. Finally, I identify understudied geographic areas which may provide additional insight into the evolutionary history and conservation management of black bear populations.

INTRODUCTION

The distribution of genetic variation within species lies along a continuum from individuals to within species lineages (Lawson 2013). Understanding how genetic diversity is distributed across a species' range informs not only the evolutionary history but also conservation and management. At the scale of the individual, genetic diversity is comprised of the specific alleles at each locus. This information has been widely used to identify individual animals based on their genotypes allowing estimates of population census size using genetic mark-recapture methods (Pearse *et al.* 2001). At the next hierarchical level, individuals are organized into families then extended family. Pedigrees (Blouin 2003) are a useful tool for understanding how individuals are related to each other within families that form an individual population. Further, pedigrees describe how traits, including heritable diseases, are distributed and transmitted (i.e. autosomal or sex-linked), and the level of inbreeding. The next hierarchical level of genetic diversity is that of the population, where multiple family groups form a single breeding unit. Populations may take several forms including being arranged sympatrically, as a metapopulation, or as a large panmictic population (Ciannelli *et al.* 2013), where population type may depend on both the geography of available habitat and the life history of the species. A population may be distributed over one or more habitat patches. Individual habitat patches do not necessarily comprise unique genetic populations as the

presence of barriers and levels of gene flow between patches influence whether individuals are randomly mating (Cushman *et al.* 2006; Dixon *et al.* 2007). I argue that there can be two additional levels of genetic structure above the population level, which I refer to as clusters and lineages. Cluster is an admittedly poor word choice for describing genetic variation due to its common use to describe genetic structure ranging from extended family groups to different species. This emanates out of the prevalent use of the Bayesian software STRUCTURE (Evanno *et al.* 2005; Pritchard *et al.* 2000) which identifies clusters (K) in the data by minimizing Hardy-Weinberg equilibrium covariance matrixes. The definition of cluster used in this paper includes the populations that comprise the next hierarchical level of population structure often representative of populations within a geographic region; in limited cases this may include a single isolated population. Finally, clusters may be organized into lineages which represent the broadest within species division of genetic diversity; together all of the lineages represent the total genetic variation for the species.

The levels of hierarchical genetic structure are related to both geographic area and time. In relation to area, as structure increases from individuals to species, area also increases from the home range of single individuals to the full species distribution. The relationship between population structure and time is similar as individuals exist on the shortest time scales (years) while within species lineages and/or species, through divergence processes, may be tens of thousands to millions of years old.

The continuous nature of genetic diversity interacts with the type of genetic marker, number of loci, number of samples, and spatial extent of sampling to determine which hierarchical level of structure at which inference may be made. These factors, in combination with the spatial extent of sampling, may confuse which level of genetic structure is being measured despite the objectives of the study, where the most commonly confused levels appear to be between extended family groups and populations. However, understanding the distribution of genetic diversity across the entire species range may aid in putting local studies within a range wide context, especially as researchers begin to estimate the average geographic area of populations in continuous habitat, which will be a function of life history specific dispersal processes.

The terminology outlined above ignores the conservation relevant terms subspecies, evolutionary significant unit, and management unit (Moritz 1994). As subspecies delimitation is poorly defined (Braby *et al.* 2012; Haig *et al.* 2006; Hawlitschek *et al.* 2012; Zink 2004) there is vigorous debate on if subspecies are real and, if so, how to define them. My definition of lineage as the first level of within species variation lends itself well to a view of subspecies as incipient species that could, through genetic drift and selection, become reproductively isolated at some point in the future. Despite this evolutionary debate, conservation organizations and governments continue to use subspecies to define populations to protect; thus, better defining how to delineate subspecies should aid in their conservation. Evolutionary significant units may be well aligned with genetic clusters which represent unique evolutionary trajectories within the lineages. Management units are best defined as the habitat patches within a management

agency's jurisdiction (Green 2005). As stated above habitat patches may or may not comprise a single genetic population.

I conducted a meta-analysis of genetic diversity across the range of the American black bear (*Ursus americanus*). Specifically, I aimed to describe how mean number of alleles and heterozygosity change over the range. Low heterozygosity and mean number of alleles may be a result of low effective population size. Following demographic bottlenecks, populations may lose alleles due to genetic drift; however, heterozygosity lags as an indicator of a decline in population size (Allendorf 1986; Balloux *et al.* 2004; Nei *et al.* 1975). In this way, genetic diversity can identify small populations which may represent isolated populations. Additionally, I synthesized population structure and genetic diversity results from local studies within a broader range wide perspective. From a management perspective, local estimates of genetic diversity contain little information without understanding the variation across populations. I end by highlighting unanswered questions related to the population genetics of black bears.

MATERIALS and METHODS

I surveyed 30 studies that genotyped bears using microsatellites and reported mean number of alleles (A) and/or observed heterozygosity (H_O) across the range of the American black bear (Atwood *et al.* 2011; Belant *et al.* 2005; Breck *et al.* 2008; Brown *et al.* 2009; Costello *et al.* 2008; Coster *et al.* 2011; Csiki *et al.* 2003; Dixon *et al.* 2007; Drewry *et al.* 2013; Frary *et al.* 2011; Gardner *et al.* 2009; Laufenberg 2014; Marshall & Ritland 2002; Marshall *et al.* 2011; Onorato *et al.* 2007; Paetkau *et al.* 1997; Peacock *et*

al. 2007; Pelletier *et al.* 2012; Puckett *et al.* Accepted; Puckett *et al.* 2014; Robinson *et al.* 2007; Roy *et al.* 2012; Sanderlin *et al.* 2009; Sawaya *et al.* 2012; Schwartz *et al.* 2006; Sinclair *et al.* 2003; Triant *et al.* 2004; Warrillow *et al.* 2001; Wilder 2003; Winslow 2012). I recorded sample size, number of microsatellite loci genotyped, A, H_O, and a geographic coordinate for each population. When two studies sampled the same population, I selected the study with larger sample size unless both studies had greater than 30 individuals (Hale *et al.* 2012), in which I selected the study with more loci for the analysis. All recorded data were reported in Table 1. I interpolated a surface for both A and H_O by ordinary kriging with a Gaussian semicovariogram, a cell size of 0.181 decimal degrees, and used the six closest data points in ArcGIS v10 (ESRI, Redlands, CA). Kriging was done within the mask of the black bear range using an IUCN range map (Garshelis *et al.* 2008).

RESULTS and DISCUSSION

Population Structure at Multiple Scales

Within species lineages

My colleagues and I previously conducted the largest study of black bear genomics (Puckett *et al.* Accepted) and I base much of the discussion of hierarchical genetic organization on that study. We observed three genetic lineages: western, eastern, and Alaskan. These results highlight that the spatial scale of evolution in black bears is subcontinental similar to brown bears (Davison *et al.* 2011). The geographic distribution of the three nuclear and three mitochondrial lineages was not congruent nor was the divergence timing. While the observed ranges of the nuclear eastern and mitochondrial

A-east lineages overlap, the Alaskan lineage primarily has A-west (with a low proportion of B) mitochondrial haplotypes. The western nuclear lineage is primarily associated with A-west haplotypes except along the Pacific coast where the frequency of B haplotypes increases (Puckett *et al.* Accepted). For the mitochondrial lineages, clade B diverged from clade A 1.07 Mya, where clades A-east and A-west diverged 169 kya; whereas in the nuclear genome the western lineage diverged 67 kya, then the Alaskan and eastern lineages diverged 31 kya (Puckett *et al.* Accepted). Thus, the mito-nuclear discordance in black bears is both spatial and temporal. This suggests introgression of the mitochondrial genome in the western range, and that mitochondrial haplotypes alone are not sufficient for evolutionary inference in black bears.

Black bears contracted into glacial refugia during the Last Glacial Maximum (LGM, 21 kya), then expanded and recolonized previously glaciated habitat in North America. Expansion out of refugia resulted in differential amounts of admixture, or the mixing of clusters or lineages, across the range. Admixture was most apparent in the northwest. I hypothesize that the western lineage which formed following divergence spread across the northern and southern Rocky Mountains and westward towards the Pacific coast before a southward range contraction as the Cordilleran ice sheet covered modern Canada. In the same way, bears that formed the Alaskan lineage contracted into Beringia and were separated by the ice. As the glaciers receded, the western lineage expanded northwards and the Alaskan lineage expanded southward. This resulted in bears in the northwest of the continent (Montana, Idaho, Wyoming, Oregon, Washington, British Columbia, Alberta) having an admixed signature between the western and Alaskan

lineages (Puckett *et al.* Accepted). The second admixture event was formed from the long distance eastward range expansion of the Alaskan lineage. The proportion of the genome for which the Alaskan lineage contributes steadily decreases towards the east across Canada.

Regional and local structure

Within the three lineages, Puckett *et al.* (Accepted) identified nine genetic clusters (Table 1). Based upon the data from this meta-analysis, 52 populations and an additional 12 management units have been identified to date. Increasing the geographic range of sampling would invariably identify more populations and additional clusters, as well as better define the spatial extent of known populations. I discuss these levels of genetic structuring inferred from local studies that used microsatellite markers.

Southeast- The Southeast cluster represents descendants from the ancestral cluster that contracted into the Southeast refugium during the LGM (Puckett *et al.* Accepted). Since that time the landscape has become fragmented, leaving several isolated populations. These include seven populations in Florida (Dixon *et al.* 2007), at least three management units in Louisiana (Laufenberg 2014), two populations (mountains and coastal) in North Carolina where bears in Tennessee may belong to the mountain population, and one population each in West Virginia, Kentucky (Frary *et al.* 2011), coastal South Carolina (Drewry *et al.* 2013), and Georgia (Sanderlin *et al.* 2009). I used management unit and not population to describe areas in Louisiana as limited studies of gene flow have been conducted, whereas the extreme isolation of other southeastern areas

would lend itself to hypotheses that each area makes up a single breeding unit. Local range expansion appears to be ongoing in eastern Kentucky, with bears from the Appalachian Mountains serving as a likely source (Frary *et al.* 2011). Finally, while it is unknown which genetic cluster the Arkansas-White River population belongs to, Laufenberg's (2014) data suggest it is the Southeast cluster.

Northeast- Following the LGM, bears expanded out of the Southeast refugium and recolonized the northeast. As bears are found in Quebec and Newfoundland, it is likely that the Northeast genetic cluster extends far into Canada or that a secondary range expansion resulted in another genetic cluster further in the northeast or on Newfoundland. Habitat is continuous in the northeast and few studies have investigated population structure within the region. There are a minimum of six populations in the Northeast cluster; one each in Newfoundland, Maine, Pennsylvania, the Lower Peninsula of Michigan, the Bruce Peninsula of Ontario, and southeastern Ontario which may extend into the Adirondacks of New York (Marshall *et al.* 2011; Pelletier *et al.* 2012; Puckett *et al.* Accepted). The extent of the southeastern Ontario population is broadly distributed in eastern and central Ontario; however, connectivity with bears in western Quebec is unknown. There is evidence of local expansion into previously extirpated parts of the range including Ohio, New Jersey, Maryland, and Connecticut (Scheick & McCown 2014). These areas should be considered management units pending further research into their source populations and gene flow with other populations to determine population status. The source population for the Connecticut expansion appears to be from southeastern Ontario/Adirondack Mountains of New York (Puckett, unpublished data).

Great Lakes- Following the range expansion out of the southeast refugium that produced the Northeast genetic cluster, the expansion continued counterclockwise around the Great Lakes and across central Canada (Puckett *et al.* Accepted). Similarly to the long distance range expansion and clinal structure of Great Lakes genotypes across Canada, Alaskan lineage genotypes made an eastward expansion, with decreasing frequencies the further east across the Great Lakes and Northeast genetic clusters. These opposing range expansion clines clarify the clinal structure observed across Ontario (Pelletier *et al.* 2012) as well as the higher prevalence of clade A-west mitochondrial haplotypes in Manitoba (Puckett *et al.* Accepted) and western Ontario (Pelletier *et al.* 2011). There are at least three populations within the Great Lakes cluster, one each across: Wisconsin and the Upper Peninsula of Michigan, Minnesota, and western Ontario (Pelletier *et al.* 2012; Puckett *et al.* Accepted). Data suggests that bears in Manitoba are more closely related to those in Minnesota than other populations; however, it is unclear if Manitoba and Minnesota comprise a single or multiple populations. The Great Lakes cluster is data deficient in the northern ranges of Canadian provinces and the western extent of the cluster is unknown, specifically if it extends into Saskatchewan and eastern Alberta.

Central Interior Highlands- The population genetics of the Central Interior Highlands (CIH) have received substantial attention (Csiki *et al.* 2003; Puckett *et al.* 2014; Van Den Bussche *et al.* 2009; Warrillow *et al.* 2001). The crux of the debate has been the extent to which the translocation of bears from the Great Lakes cluster influenced the genetics of bears in Arkansas, Oklahoma, and Missouri. Two analyses appear to resolve these

questions. Based on population divergence analyses, the CIH genetic cluster diverged from the Southeast approximately 11 kya (Puckett *et al.* Accepted). Some of these bears remained on the landscape following what was long thought to be extirpation throughout the CIH (Clark & Smith 1994); however, admixture between CIH and Great Lakes genotypes was observed in bears in both the Ozark and Ouachita Mountains, which are distinct populations (Puckett *et al.* 2014). Finally, genomic data support the hypothesis that bears in Missouri were not a remnant population representing the CIH cluster but an isolated population that founded Missouri following the Arkansas reintroduction (Puckett *et al.* Accepted). Contemporary northward range expansion of bears from the Arkansas Ozarks into Missouri may connect these populations in the future (Puckett *et al.* 2014).

Southwest- The Southwest genetic cluster is found throughout the southern Rocky Mountains and Mogollon Rim, and may extend into western Mexico. The data suggest there are at least four populations including: the Mongollon Rim of Arizona and New Mexico, the Sky Islands of southern Arizona and northern Mexico, the Sangre de Cristo Mountains in New Mexico, and Colorado (Atwood *et al.* 2011; Costello *et al.* 2008; Puckett *et al.* Accepted). Onorato *et al.* (2007) also identified the Mongollon Rim population, as well as two management units in Texas and northern Mexico. However, the bears sampled in Texas and nearby Mexican locations (Carmens and Burros) are probably related to the eastern lineage and not the western lineage which the Southwest cluster belongs, despite geographic proximity. This hypothesis is based upon the Texas and eastern Mexican bears having A-east mitochondrial haplotypes (Onorato *et al.* 2004) and nuclear (Onorato *et al.* 2007) genetic signatures unique from that of the Mongollon

Rim population which is part of the Southwest cluster. I further hypothesize that the Arizona sky islands and Sierra Nevadas are part of the Southwest genetic cluster based upon the high proportion of A-west haplotypes and the inferred glacial history as the entire Southwest served as an LGM refugium (Puckett *et al.* Accepted). Finally, I hypothesize that eastern Utah and southern Wyoming are both management units and potentially populations within the Southwest cluster due to continuous habitat across the Colorado plateau.

Pacific Coast- The Coastal Range between Oregon and California formed a distinct genetic cluster (Puckett *et al.* Accepted); I hypothesize that this cluster extends from Washington state to both the coastal and Sierra Nevada Mountains in California. California has three populations: North Coast/Klamath, central Sierra Nevadas, and southern Sierra Nevadas (Brown *et al.* 2009). Brown and colleagues (2009) identified substructure within the central Sierra Nevadas; however, these may represent management units instead of populations as related individuals were not removed from the analysis which may result in overestimation of structure.

West- Following divergence of the western lineage, bears spread throughout the northern and southern Rocky Mountains and along the Pacific Coast. The LGM pushed bears into two glacial refugia, Southwest and Pacific Northwest, thereby beginning divergence between the geographic regions, although both have high frequencies of A-west mitochondrial haplotypes (Puckett *et al.* Accepted). Following expansion out of glacial refugia, bears from the Alaskan and eastern lineages (from the Great Lakes cluster)

admixed in to the West (Puckett *et al.* Accepted). Sampling along a north-south axis, shows higher proportions of Alaskan lineage admixture in more northern samples across the western range (Pacific Coast, West, and Southwest clusters). Previous studies identified very fine scale population structure of black bears within the Alexander Archipelago corresponding to isolation on islands or mainland peninsulas (Peacock *et al.* 2007; Puckett *et al.* Accepted); given that British Columbia has a similar coastline, I hypothesize fine scale population structure but not necessarily unique clusters, lineages, or subspecies along the coast. Given the extent of the range studied to date, I inferred 12 populations including: the Idaho panhandle and northwestern Montana which likely extends into southern British Columbia, central Idaho and eastern Oregon, central Montana, southern Montana which may extend into northwestern Wyoming (Puckett *et al.* Accepted), Banff National Park in Alberta (Sawaya *et al.* 2012), the six coastal and island sampling sites of Marshall and Ritland (2002) including Vancouver Island, and Haida Gwaii due to isolation. Given both the coastline geography and sparse sampling of the interiors of British Columbia, Alberta, and Saskatchewan, there are likely more populations in the West cluster.

Alaska-East and Alaska-West- As described above, the coastal geography along the Pacific coast isolates bears on islands and peninsulas that are geographically close but which limit gene flow. The Alaskan lineage which comprises the Alaska-East genetic cluster has at least five populations including: Kenai Peninsula, Yakutat, Chilkat Peninsula, Juneau, and the central mainland of the Alexander Archipelago (Peacock *et al.* 2007; Puckett *et al.* Accepted). I hypothesize that the Alaskan lineage extends into

mainland Alaska and eastward into Yukon, evidenced by the declining proportion of Alaskan genotypes observed in the Great Lakes and Northeast clusters which I hypothesize expanded north around the Yukon Range.

Despite the adjacent geographic proximity of Alaska-East and Alaska-West clusters, the Alaska-West cluster is a western lineage with substantial Alaskan lineage admixture (Puckett *et al.* Accepted). I hypothesize that this cluster established following deglaciation from northward range expansion out of the Pacific Northwest refugium. The Alaska-West cluster includes four populations within the southern extent of the Alexander Archipeligo on Kuiu, Kupreanof, and Prince of Wales Islands, and the southeastern mainland (Peacock *et al.* 2007).

Genetic Diversity

Species that contracted into glacial refugia then expanded their ranges in interglacial climates are expected to show a pattern of higher genetic diversity in the refugial area and lower diversity in the expanded range (Excoffier & Ray 2008). This pattern is expected during range expansions because only a subset of alleles expand with the expansion front, thereby reducing genetic diversity before a reduction of heterozygosity (Allendorf 1986). The American black bear does not exhibit this pattern of diversity for either mean number of alleles (A) nor observed heterozygosity (H_O; Figure 5.1) despite contraction into multiple refugia. Specifically, the refugial areas in the Southeast, Southwest, and Pacific Northwest have lower genetic diversity than the continuous, and previously glaciated, range in the north (Figure 5.1).

One explanation for the patterns of genetic diversity in black bears is the variability of continuous habitat across the range. Fragmented habitat would be expected to support fewer bears due to limited resources including food and territory. In these smaller patches, genetic drift will have a larger effect and alleles will be lost from the population; the converse is expected for continuous habitat that may support large populations, higher levels of gene flow, and weak genetic drift. This is congruent with data for black bears, as the isolated mountain ranges of the southwest and the Sierra Nevadas of California (both surrounded by deserts), islands (Newfoundland and Vancouver), and peninsulas (Kenai, Alaska and Bruce, Ontario) have the lowest genetic diversity (Table 5.1). The highest levels of genetic diversity were observed across Canada and in the northern Rocky Mountains (Figure 5.1).

Further, anthropogenic processes such as abundance declines due to overharvest or habitat fragmentation may also affect genetic diversity similarly to areas of natural geographic isolation. As the pressures from both overharvest and habitat fragmentation occurred within the past 200 years, populations may not have had time to recover lost genetic diversity via gene flow or mutation. For example, despite long-term habitat stability on the Florida peninsula, recent habitat fragmentation and lack of dispersal corridors have been implicated in both reduced population size and genetic diversity in the seven habitat patches within the state (Dixon *et al.* 2007). Similarly, historic overharvest decreased bear abundance; currently, these populations are expanding locally (in population size and geographic range) as hostility towards carnivores has waned in

recent decades. Local bear population expansions into previously extirpated regions have been documented in Ohio and Connecticut (Scheick & McCown 2014). The low genetic diversity observed in Connecticut may be due to a recent founder event (Table 5.1).

The high genetic diversity across the northern range may have several explanations. First, as mentioned above the larger population sizes and weaker genetic drift result in more alleles being maintained in the population. Second, long distance dispersal events, as opposed to a purely stepping-stone expansion, may increase genetic diversity and heterozygosity and decrease differentiation along the expansion axis by continually bringing alleles to the expanding edge (Ray & Excoffier 2010). Black bear natal dispersal distances up to 250 km have been reported for both sexes (Costello 2010; Moore *et al.* 2014), indicating that while rare, this may be a viable hypothesis for long distance gene flow. Third, admixed regions have higher genetic diversity than un-admixed populations due to unique sets of alleles, including private alleles, combining, thereby increasing genetic diversity. As admixture was observed across the Alaska-East, Alaska-West, West, Great Lakes, and Northeast clusters, admixture may in part explain the higher genetic diversity across the northern range. Of note, Frary *et al.* (2011) hypothesized that allelic richness and heterozygosity were high in the recently expanded population of eastern Kentucky due to admixture. However, genetic diversity in Kentucky was within the range of nearby bears sampled in West Virginia and North Carolina (Table 5.1, Figure 5.1A). While admixture between the Southeast and Northeast clusters is a testable hypothesis, current evidence indicates that the Northeast expansion into eastern Ohio has yet to expand as far south as Kentucky (Scheick & McCown 2014).

Both this example from Kentucky, as well as the reintroduction of bears into Arkansas (Clark & Smith 1994), show that genetic diversity will not necessarily be reduced during founding events, although large numbers of founding individuals will aid in maintaining high genetic diversity (Puckett *et al.* 2014).

Open Questions in American Black Bear Population Genetics

While black bears are intensely studied due to their game status, there remains much to learn about the species. Given that the range map (Scheick & McCown 2014) for the species has a more northern distribution than mapped presence locations from the Global Biodiversity Information Facility database, one open question is: how far north does the species extend? Specifically, does this mismatch represent a lack of information in the database or an overestimation of the range? Either may be possible as many studies are biased to either the continental United States or populations along the Pacific Coast. This bias extends to studies of genetic structure; I will outline open questions on the genetics of the species and present my hypotheses below.

First, to which genetic cluster(s) do bears in Mexico belong? Two subspecies have been described in Mexico, *U. a. machetes* in the west and *U. a. eremicus* in the east (Lariviere 2001). Considering that Varas-Nelson (2010) observed clade A-west mitochondrial haplotypes in her study of Mexican populations semi-continuous with Arizona, and Onorato *et al.* (2007) observed A-east haplotypes in his study of Mexican populations continuous with Texas (*U. a. machetes* and *U. a. eremicus*, respectively), I hypothesize that the western populations are an extension of the Southwest genetic cluster and the

eastern populations share a most recent population level ancestor with the CIH cluster of the eastern lineage.

Second, to which genetic cluster do bears in California belong? Brown *et al.* (2009) identified four populations within the state of California. As stated above, I hypothesize that the North Coast/Klamath population is contiguous with bears in coastal Oregon (Puckett *et al.* Accepted) based upon work in other species which observed a phylogeographic suture zone between coastal Oregon and California (Soltis *et al.* 1997; Swenson & Howard 2005). While F_{ST} statistics were significantly different between all pairwise comparisons of the North Coast, and northern, central, and southern Sierra Nevada Mountains (Brown *et al.* 2009), the significance may be at the level of populations and not genetic clusters, thereby suggesting that all of California belongs to the Pacific Coast cluster.

Third, when did bears colonize islands? Black bears live on Newfoundland, Vancouver, Haida Gawaii, and the Princess Charlotte Islands, where each of these island populations has been described as a subspecies: *U. a. hamiltoni*, *U. a. vancouveri*, *U. a. carlottae*, and *U. a. kermodei* respectively. The primary question relates to if bears colonized islands before or after the LGM. While bear fossils on some of these islands have been identified and dated to before the LGM (Wooding & Ward 1997), that does not mean that those islands were unglaciated and supported bears during the LGM. Thus genetic analyses that estimated population divergence times between mainland and island populations would provide a clearer understanding of colonization history, including

patterns of range expansion out of glacial refugia. Population genetics could also be used to infer the mainland population which founded each island population. Understanding the level of divergence and any adaptations between the mainland and island populations would also help to clarify the level of differentiation needed to warrant subspecies status. Answering these questions would also answer broader evolutionary questions about the extent of genetic drift at neutral and potentially adaptive loci, as the colonization of islands represents a natural experiment within the species.

Fourth, what is the spatial extent of the Alaskan lineage? The samples from Puckett *et al.* (Accepted) observed the Alaskan lineage along the Pacific coast (Kenai Peninsula and the Alexander Archipelago) but did not sample inland. The subspecies designations for Alaska include *U. a. perniger* on Kenai Peninsula, *U. a. emmonsii* along the southern coast of Alaska, *U. a. pugnax* on the Alexander Archipelago, and *U. a. americanus* in central Alaska and Yukon (Lariviere 2001). I hypothesize that the Alaskan lineage is found throughout central Alaska and adjacent Yukon and Northwest Territories. There may be admixed signatures from long distance range expansion of the Great Lakes or a western lineage signature; however, I hypothesize that the Alaskan lineage is not confined to coastal areas.

Answering these questions regarding population structure will contribute to both the conservation and evolutionary understanding of black bears. Identifying geographic regions with unique diversity will aid future translocations of bears by identifying source populations for translocations that are genetically similar to recipient populations.

Further, in the process of identifying genetic structure across the range, we subsequently estimate genetic diversity and inbreeding levels. This information may be used along with census size estimates to target populations where supplementation would be an effective conservation management strategy. Understanding population structure will also aid evolutionary inference by accounting for neutral differentiation due to population structure in statistical estimates of adaptation and selection.

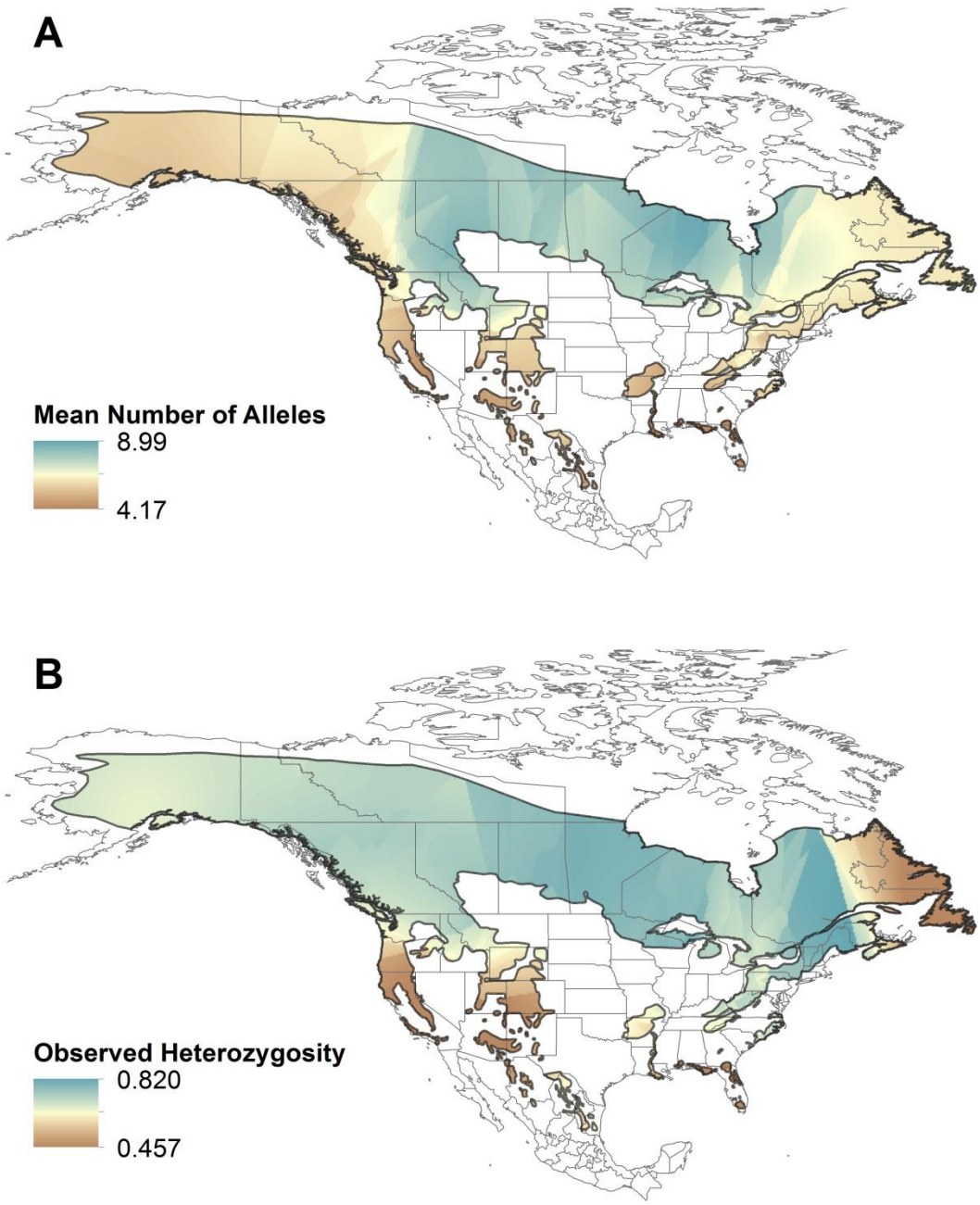

Figure 5.1- Interpolated surfaces for (A) mean number of alleles and (B) observed heterozygosity across the range of the American black bear. In each map, darker brown and darker green regions respectively represent areas of low and high genetic diversity.

Table 5.1- Population genetic estimates of American black bears across studies showing estimates of mean number of alleles (A) and observed heterozygosity (H_O) for varying sample sizes (n) and number of microsatellite loci.

Location	n	Loci	A	H_O	Reference
Southwest					
Arizona- North	102	11	5.8	0.538	Atwood et al 2011
Arizona- Sky Islands	33	11	4.3	0.422	Atwood et al 2011
Arizona	8	15	4.1	0.515	Puckett et al In Review
Colorado	20	15	5.3	0.615	Puckett et al In Review
New Mexico- Sangre de Cristo Mountains	211	11	6.5	0.626	Costello et al 2008
New Mexico- Mogollon Mountains	195	11	5.9	0.549	Costello et al 2008
New Mexico- Mogollon Mountains	29	7		0.493	Onorato et al 2007
New Mexico	11	15	4.8	0.551	Puckett et al In Review
New Mexico	143	12	4.8	0.559	Winslow 2012
Utah- East Tavaputs Plateau	71	7		0.529	Sinclair et al 2003
West					
Alberta- Banff National Park	85	13	10.3	0.850	Sawaya et al 2010
Alberta	13	15	7.4	0.739	Puckett et al In Review
British Columbia- Mainland East of Princess Royal	21	8	6.6		Marshall and Ritland 2002
British Columbia- Princess Royal Island	50	8	6.5		Marshall and Ritland 2002
British Columbia- Vancouver	19	8	4.4		Marshall and Ritland 2002
British Columbia- Terrace	17	8	7.5		Marshall and Ritland 2002
British Columbia- Don Peninsula	23	8	5.9		Marshall and Ritland 2002
British Columbia- Hawkesbury	20	8	5.7		Marshall and Ritland 2002
British Columbia- Golden (West Slopes)	116	8	9.5		Paetkau et al 1997
British Columbia	19	15	8.1	0.799	Puckett et al In Review
Nevada- Lake Tahoe	50	9	4.6	0.465	Breck et al 2008
Idaho- Purcell Mountains	72	9	9.0	0.760	Schwartz et al 2006
Idaho- Selkirk Mountains	73	9	9.3	0.800	Schwartz et al 2006
Idaho	57	15	9.3	0.679	Puckett et al In Review
Montana	28	15	7.8	0.784	Puckett et al In Review
Pacific Coast					
California- North Coast	167	12	9.2	0.610	Brown et al 2009

Location	n	Loci	A	H_O	Reference
California- Cascades	93	12	5.5	0.520	Brown et al 2009
California- Central Sierra Nevada	161	12	4.9	0.510	Brown et al 2009
California- Southern Sierra Nevada	117	12	4.5	0.440	Brown et al 2009
California- Yosemite	46	9	3.8	0.450	Breck et al 2008
Oregon	9	15	5.6	0.783	Puckett et al In Review
Alaska					
Alaska- Kenai Peninsula	79	13		0.671	Robinson et al 2007
Alaska- Kenai Peninsula	10	15	5.4	0.725	Puckett et al In Review
Alaska- Mainland	20	13		0.704	Robinson et al 2007
Alaska- ABC Islands	254	7		0.690	Peacock et al 2007
Alaska-ABC Islands East	14	15	5.8	0.759	Puckett et al In Review
Alaska-ABC Islands West	6	15	5.2	0.77	Puckett et al In Review
Alaska- Kennicott	84	6		0.751	Wilder 2003
Yukon	15	15	7.5	0.714	Puckett et al In Review
Central Interior Highlands					
Ozarks- Arkansas and Oklahoma	114	15	8.0	0.779	Puckett et al 2014
Ouachitas- Arkansas and Oklahoma	135	15	8.3	0.733	Puckett et al 2014
Missouri	36	15	5.0	0.777	Puckett et al 2014
Great Lakes					
Manitoba	9	15	6.9	0.822	Puckett et al In Review
Michigan- Upper Peninsula	10	15	6.2	0.822	Puckett et al In Review
Minnesota	28	15	8.5	0.791	Puckett et al In Review
Wisconsin- Apostle Islands	32	6		0.840	Belant et al 2005
Wisconsin	10	15	6.1	0.713	Puckett et al In Review
Northeast					
Connecticut	12	15	4.7	0.694	Puckett et al In Review
Massachusetts	5	15	4.5	0.923	Puckett et al In Review
Maine	10	15	6.7	0.853	Puckett et al In Review
Michigan- Lower Peninsula	10	15	6.1	0.808	Puckett et al In Review
New York	47	6		0.800	Gardner et al 2009
New York	7	15	5.5	0.719	Puckett et al In Review

Location	n	Loci	A	H_O	Reference
Ontario	2839	15	8.8	0.750	Pelletier et al 2012
Ontario-Red Lake	44	15	9.4	0.769	Pelletier et al 2012
Ontario-Nipigon	49	15	9.8	0.813	Pelletier et al 2012
Ontario- Wawa	37	15	9.3	0.761	Pelletier et al 2012
Ontario- Sault Ste. Marie	52	15	9.8	0.765	Pelletier et al 2012
Ontario- Bruce	135	15	4.6	0.549	Pelletier et al 2012
Ontario- Peterborough	10	15	6.1	0.783	Pelletier et al 2012
Ontario[1]	51	15	9.5	0.800	Puckett et al In Review
Pennsylvania	15	15	6.8	0.702	Puckett et al In Review
Quebec	141	10	11.5	0.829	Roy et al 2012
Quebec- La Mauricie National Park	32	8	8.75		Paetkau et al 1997
Southeast					
Florida- Apalachicola	40	12	5.9	0.690	Dixon et al 2007
Florida- Aucilla	40	12	5.0	0.566	Dixon et al 2007
Florida- Big Cypress	41	12	5.5	0.642	Dixon et al 2007
Florida- Chassahowitzka	29	12	2.3	0.287	Dixon et al 2007
Florida- Eglin	40	12	4.1	0.613	Dixon et al 2007
Florida- Highlands/Glades	28	12	2.8	0.327	Dixon et al 2007
Florida- Ocala	40	12	4.8	0.579	Dixon et al 2007
Florida- Ocala	9	15	4.3	0.601	Puckett et al In Review
Florida- Osceola	41	12	6.7	0.705	Dixon et al 2007
Florida- St. Johns	40	12	5.8	0.650	Dixon et al 2007
Georgia	23	21	2.9	0.500	Sanderlin et al 2009
Georgia	23	8	2.6		Sanderlin et al 2009
Louisiana- Coastal	20	5	4.2		Csiki et al 2003
Louisiana- Coastal	57	8		0.506	Triant et al 2004
Louisiana- Inland	32	8		0.704	Triant et al 2004
Louisiana- Lower Atchafalaya	26	8	6.1	0.420	Warrillow et al 2001
Louisiana- Three Rivers Complex	39	23	4.4		Laufenberg 2014
Louisiana- Upper Atchafalaya	60	23	4.1		Laufenberg 2014
Mississippi	59	23	4.8		Laufenberg 2014

Location	n	Loci	A	H_O	Reference
Kentucky	54	6	7.5	0.81	Frary et al 2011
North Carolina- Mountains	17	15	6.8	0.801	Puckett et al In Review
North Carolina- Coast	14	15	6.5	0.765	Puckett et al In Review
South Carolina- Lewis Bay	16	21	5.1	0.630	Drewry et al 2013
West Virginia	20	15	7.2	0.743	Puckett et al In Review
Eastern Lineage- Unknown Cluster					
Arkansas- White River	6	5	1.8		Csiki et al 2003
Arkansas- White River	22	8	2.4	0.380	Warrillow et al 2001
Arkansas- White River	105	23	2.2		Laufenberg 2014
Newfoundland	32	8	3		Paetkau et al 1997
Newfoundland[1]	93	12	3.9		Marshall et al 2011
Newfoundland- Northern Peninsula	25	12		0.443	Marshall et al 2011
Newfoundland- Baie Verte Peninsula	47	12		0.42	Marshall et al 2011
Newfoundland- Bonavista Peninsula	21	12		0.373	Marshall et al 2011
Texas- Big Bend	32	7		0.722	Onorato et al 2007
Mexico- Burros	58	7		0.732	Onorato et al 2007

LITERATURE CITED

Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* **5**, 181-190.

Atwood TC, Young JK, Beckmann JP, *et al.* (2011) Modeling connectivity of black bears in a desert sky island archipelago. *Biological Conservation* **144**, 2851-2862.

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* **13**, 3021-3031.

Belant JL, Van Stappen JF, Paetkau D (2005) American black bear population size and genetic diversity at Apostle Islands National Lakeshore. *Ursus* **16**, 85-92.

Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution* **18**, 503-511.

Braby MF, Eastwood R, Murray N (2012) The subspecies concept in butterflies: has its application in taxonomy and conservation biology outlived its usefulness? *Biological Journal of the Linnean Society* **106**, 699-716.

Breck SW, Wllliams CL, Beckmann JP, *et al.* (2008) Using genetic relatedness to investigate the development of conflict behavior in black bears. *Journal of Mammalogy* **89**, 428-434.

Brown SK, Hull JM, Updike DR, Fain SR, Ernest HB (2009) Black bear population genetics in California: Signatures of population structure, competitive release, and historical translocation *Journal of Mammalogy* **90**, 1066-1074.

Ciannelli L, Fisher JA, Skern-Mauritzen M, *et al.* (2013) Theory, consequences and evidence of eroding population spatial structure in harvested marine fishes: a review. *Marine Ecology Progress Series* **480**, 227-243.

Clark JD, Smith KG (1994) A demographic comparison of two black bear populations in the Interior Highlands of Arkansas *Wildlife Society Bulletin* **22**, 593-603.

Costello CM (2010) Estimates of dispersal and home-range fidelity in American black bears. *Journal of Mammalogy* **91**, 116-121.

Costello CM, Creel SR, Kalinowski ST, Vu NV, Quigley HB (2008) Sex-biased natal dispersal and inbreeding avoidance in American black bears as revealed by spatial genetic analyses. *Molecular Ecology* **17**, 4713-4723.

Coster SS, Kovach AI, Pekins PJ, Cooper AB, Timmins A (2011) Genetic mark-recapture population estimation in black bears and issues of scale. *Journal of Wildlife Management* **75**, 1128-1136.

Csiki I, Lam C, Key A, *et al.* (2003) Genetic variation in black bears in Arkansas and Louisiana using microsatellite DNA markers. *Journal of Mammalogy* **84**, 691-701.

Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: Testing multiple hypotheses with causal modeling. *American Naturalist* **168**, 486-499.

Davison J, Ho SYW, Bray SC, *et al.* (2011) Late-Quaternary biogeographic scenarios for the brown bear (*Ursus arctos*), a wild mammal model species. *Quaternary Science Reviews* **30**, 418-430.

Dixon JD, Oli MK, Wooten MC, *et al.* (2007) Genetic consequences of habitat fragmentation and loss: the case of the Florida black bear (*Ursus americanus floridanus*). *Conservation Genetics* **8**, 455-464.

Drewry JM, Van Manen FT, Ruth DM (2013) Density and genetic structure of black bears in coastal South Carolina. *The Journal of Wildlife Management* **77**, 153-164.

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

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution* **23**, 347-351.

Frary VJ, Duchamp J, Maehr DS, Larkin JL (2011) Density and distribution of a colonizing front of the American black bear *Ursus americanus*. *Wildlife Biology* **17**, 404-416.

Gardner B, Royle JA, Wegan MT (2009) Hierarchical models for estimating density from DNA mark-recapture studies. *Ecology* **90**, 1106-1115.

Garshelis DL, Crider D, Van Manen FT (2008) *Ursus americanus*. In: *The IUCN Red List of Threatened Species. Version 2014.3*.

Green DM (2005) Designatable units for status assessment of endangered species. *Conservation Biology* **19**, 1813-1820.

Haig SM, Beever EA, Chambers SM, *et al.* (2006) Taxonomic considerations in listing subspecies under the U.S. Endangered Species Act. *Conservation Biology* **20**, 1584-1594.

Hale ML, Burg TM, Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *Plos One* **7**, e45170.

Hawlitschek O, Nagy ZT, Glaw F (2012) Island Evolution and Systematic Revision of Comoran Snakes: Why and When Subspecies Still Make Sense. *Plos One* **7**, e42970.

Lariviere S (2001) *Ursus americanus*. *Mammalian Species* **647**, 1-11.

Laufenberg JS (2014) *Population dynamics and genetic structure of Louisiana black bears in the Lower Mississippi Alluvial Valley of Louisiana*, University of Tennessee.

Lawson DJ (2013) Populations in statistical genetic modelling and inference. *arXiv*, 1306.0701.

Marshall HD, Ritland K (2002) Genetic diversity and differentiation of Kermode bear populations. *Molecular Ecology* **11**, 685-697.

Marshall HD, Yaskowiak ES, Dyke C, Perry EA (2011) Microsatellite population structure of Newfoundland black bears (*Ursus americanus hamiltoni*). *Canadian Journal of Zoology* **89**, 831-839.

Moore JA, Draheim HM, Etter D, Winterstein S, Scribner KT (2014) Application of large-scale parentage analysis for investigating natal dispersal in highly vagile vertebrates: A case study of American black bears (*Ursus americanus*). *Plos One* **9**, e91168.

Moritz C (1994) Defining evolutionary significant units for conservation. *Trends in Ecology & Evolution* **9**, 373-375.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.

Onorato DP, Hellgren EC, Van Den Bussche RA, Doan-Crider DL (2004) Phylogeographic patterns within a metapopulation of black bears (*Ursus americanus*) in the American Southwest. *Journal of Mammalogy* **85**, 140-147.

Onorato DP, Hellgren EC, Van Den Bussche RA, Doan-Crider DL, Skiles JR (2007) Genetic structure of American black bears in the desert southwest of North America: conservation implications for recolonization. *Conservation Genetics* **8**, 565-576.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics* **147**, 1943-1957.

Peacock E, Peacock MM, Titus K (2007) Black bears in Southeast Alaska: the fate of two ancient lineages in the face of contemporary movement. *Journal of Zoology* **271**, 445-454.

Pearse DE, Eckerman CM, Janzen FJ, Avise JC (2001) A genetic analogue of 'mark–recapture' methods for estimating population size: an approach based on molecular parentage assessments. *Molecular Ecology* **10**, 2711-2718.

Pelletier A, Obbard ME, Mills K, *et al.* (2012) Delineating genetic groupings in continuously distributed species across largely homogeneous landscapes: a study of American black bears (*Ursus americanus*) in Ontario, Canada. *Canadian Journal of Zoology* **90**, 999-1014.

Pelletier A, Obbard ME, White BN, Doyle C, Kyle CJ (2011) Small-scale genetic structure of American black bears illustrates potential postglacial recolonization routes. *Journal of Mammalogy* **92**, 629-644.

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

Puckett EE, Etter PD, Johnson EA, Eggert LS (Accepted) Phylogeographic analyses of American black bears (*Ursus americanus*) suggest four glacial refugia and complex patterns of post-glacial admixture. *Molecular Biology and Evolution*.

Puckett EE, Kristensen TV, Wilton CM, *et al.* (2014) Influence of drift and admixture on population structure of American black bears (*Ursus americanus*) in the Central Interior Highlands, USA 50 years after translocation. *Molecular Ecology* **23**, 2414-2427.

Ray N, Excoffier L (2010) A first step towards inferring levels of long-distance dispersal during past expansions. *Molecular Ecology Resources* **10**, 902-914.

Robinson SJ, Waits LP, Martin ID (2007) Evaluating population structure of black bears on the Kenai Peninsula using mitochondrial and nuclear DNA analyses. *Journal of Mammalogy* **88**, 1288-1299.

Roy J, Yannic G, Cote SD, Bernatchez L (2012) Negative density-dependent dispersal in the American black bear (*Ursus americanus*) revealed by noninvasive sampling and genotyping. *Ecology and Evolution* **2**, 525-537.

Sanderlin JS, Faircloth BC, Shamblin B, Conroy MJ (2009) Tetranucleotide microsatellite loci from the black bear (*Ursus americanus*). *Molecular Ecology Resources* **9**, 288-291.

Sawaya MA, Stetz JB, Clevenger AP, Gibeau ML, Kalinowski ST (2012) Estimating grizzly and black bear population abundance and trend in Banff National Park using noninvasive genetic sampling. *Plos One* **7**, e34777.

Scheick BK, McCown W (2014) Geographic distribution of American black bears in North America. *Ursus*, 24-33.

Schwartz MK, Cushman SA, McKelvey KS, Hayden J, Engkjer C (2006) Detecting genotyping errors and describing American black bear movement in northern Idaho. *Ursus* **17**, 138-148.

Sinclair EA, Black HL, Crandall KA (2003) Population structure and paternity in an American Black Bear (*Ursus americanus*) population using microsatellite DNA. *Western North American Naturalist* **63**, 489-497.

Soltis D, Gitzendanner M, Strenge D, Soltis P (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution* **206**, 353-373.

Swenson Nathan G, Howard Daniel J (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist* **166**, 581-591.

Triant DA, Pace RM, Stine M (2004) Abundance, genetic diversity and conservation of Louisiana black bears (*Ursus americanus luteolus*) as detected through noninvasive sampling. *Conservation Genetics* **5**, 647-659.

Van Den Bussche RA, Lack JB, Onorato DP*, et al.* (2009) Mitochondrial DNA phylogeography of black bears (*Ursus americanus*) in central and sourthern North America: Conservatoin implications *Journal of Mammalogy* **90**, 1075-1082.

Varas-Nelson AC (2010) *Conservation genetics of black bears in Arizona and Northern Mexico* PhD, University of Arizona.

Warrillow J, Culver M, Hallerman E, Vaughan M (2001) Subspecific affinity of black bears in the White River National Wildlife Refuge. *Journal of Heredity* **92**, 226-233.

Wilder JM (2003) Quantifying bear populations and bear-human conflicts using non-invasive genetic sampling in the Kennicott Valley of Wrangell-St. Elias National Park and Preserve, Alaska. University of Idaho.

Winslow FS (2012) *Effects of translocation and climatic events on the population genetic structure of black bears in New Mexico,* University of New Mexico.

Wooding S, Ward R (1997) Phylogeography and Pleistocene evolution in the North American black bear. *Molecular Biology and Evolution* **14**, 1096-1105.

Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 561-564.

APPENDIX 1- Approximate Bayesian computation of five models including four populations to understand the relationship between the unique MO cluster and the OU and OZ genetic clusters.

MATERIALS and METHODS

Model selection for demographic history in Missouri

Following identification of the demographic model for the three population scenarios, we wanted to determine how the hierarchically substructured MO population was related to OU and OZ using *Admix* as the base model. In $Split_{MO}$ (Figure A1.1A), MO diverges from CIH lineages before OU and OZ diverge; while in $Split_{OU}$, OU diverges from CIH before OZ and MO (Figure A1.1B). In $Sequential_{MO}$ (Figure A1.1C), one generation after the Source-OZ admixture event, MO was founded from the OZ lineage. In the $Remnant_{MO}$ (Figure A1.1D) and $Admix_{MO}$ (Figure A1.1E) models all three CIH areas diverge at the same time, where the OZ and MO lineages admix seven generations from the present in the $Admix_{MO}$ model. Each model was simulated $1.5x10^6$ with priors equal to those used for the *Admix* model. For parameters not present in the *Admix* model, priors included: N_{MO} [1-1,000], λ_{MO} [0.001-0.999], and t_{MO} and t_{OU} [1-20,000]. Again we used DIYABC's logistic regression on the 1% of data with the shortest Euclidian distance in multivariate space to calculate the posterior probabilities of the models. We calculated type I and II error rate as described in the main text.

RESULTS

The logistic regression of the five models including MO supported $Remnant_{MO}$ as the most likely model with posterior probability 0.581 where $Admix_{MO}$ and $Split_{MO}$ had

posterior probabilities of 0.228 and 0.164, respectively (Figure A1.1). The type I and II errors of the $Remnant_{MO}$ model were 4.0% and 88.0% respectively.

DISCUSSION

The coalescent modeling could not distinguish between the five models with four populations. The $Remnant_{MO}$ model was the simplest of the models and only distinguishes that the four contemporary populations have divergent levels of allele frequencies and genetic diversity. In the calculation of Type II error, this model was repeatedly preferred over the model of interest indicating that the data have low power to distinguish between more complex alternative scenarios. This may be due to the short time periods in which differentiation occurred, shared ancestry, high genetic drift, and/or loss of alleles in MO.

We hypothesize that the $Split_{OU}$ model is the most congruent with knowledge about habitat fragmentation in the Ozark Mountains due to clear cutting. In this model, the OZ and MO genetic clusters would separate concurrent with or prior to the demographic bottleneck due to fragmentation of the landscape. The fragmented landscape in conjunction with low population size may have prevented gene flow within the Ozark Mountains thereby allowing drift to act independently in OZ and MO before the translocation. Missouri and Arkansas were heavily logged between 1880 to 1920 (Giessman *et al.* 1986; Stausberg & Hough 1997). Pre-settlement forest cover in Missouri was estimated at 70% with approximately half removed through intensive logging (Giessman *et al.* 1986). While the distribution of clear cutting and the resultant

forest patches are unknown, it is possible that connectivity between patches suitable for bears was removed during intensive logging. Forest protection began in Arkansas between 1907 to 1909 as the Ouachita and Ozark National Forests were established (Stausberg & Hough 1997). In Missouri, the Mark Twain and Clark National Forests were established in 1934 and 1935 (Guldin *et al.* 2008). These federal land buys along with new management practices allowed forests to regrow and may have connected disjunct patches over time.

LITERATURE CITED

Giessman NF, Barney TW, Haithcoat TL, Myers JW, Massengale R (1986) Distribution of forestland in Missouri. *Transactions of the Missouri Academy of Science* **20**, 5-14.

Guldin JM, Iffrig GF, Flader SL (2008) Pioneer Forest: A half century of sustainable uneven-aged forest management in the Missouri Ozarks, p. 123.Gen. Tech. Report SRS-108. U.S. Department of Agriculture, Forest Service, Southern Research Station, Ashville, NC.

Stausberg S, Hough WA (1997) The Ouachita and Ozark-St Francis National Forests: A history of the lands and USDA Forest Service tenure, p. 45.Gen. Tech. Report SO-I 21. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Figure A1.1- Five models used to identify the most likely demographic scenario for the unique MO genetic cluster in DIYABC. Patterns for OU, OZ, and Source were fixed as in the *Admix* model. In the $Split_{MO}$ (A) and $Split_{OU}$ (B) models, MO or OU diverge (t) from the ancestral Central Interior Highlands (CIH) lineage first. In the $Seqential_{MO}$ (C) model, MO was founded from OZ following the admixture. In the $Remnant_{MO}$ (D) and $Admix_{MO}$ (E) models, all CIH subpopulations diverged at the same time; where the models differ in that in $Admix_{MO}$ MO genotypes admix with OZ following the admixture from the Source.

Table A1.1- Primer sequences used in this study to shorten PCR amplified products for more efficient amplification from hair samples.

Locus Name	Florescent Dye	Orientation	Sequence (5' to 3')
G10O[a]	NED	Forward	CTTTGGCTACCTCAGATGG
		Reverse	TGCCTACTGCACCAACAG
G10U[a]	VIC	Forward	CAGTGTCAGTTGTTAGGAAGCAG
		Reverse	TATTTCCAATGCCCTAAGTGAT
UarMU05[b]	PET	Forward	AATCTTTTCACTTATGCCCA
		Reverse	GAAACTTGTTATGGGAACCA
UarMU10[b]	NED	Forward	TTCAGATTTCATCAGTTTGAC
		Reverse	CAGCATAGTTACACAAATCTCC
UarMU23[b]	PET	Forward	GCCTGTGTGCTATTTTATCC
		Reverse	TTTAGGATTCGGCTTGCTTG
UarMU59[b]	NED	Forward	GCTCCTTTGGGACATTGTAA
		Reverse	AGGCATGGGGAAGAAATTG
UA-P2H03[c]	6FAM	Forward	CCACCATTTTGGCATTACTT
		Reverse	TCCCTCTGTGCACTCATCTG

[a]- Paetkau, D. and C. Strobeck. 1996. Mitochondrial DNA and the phylogeography of Newfoundland black bears. *Canadian Journal of Zoology* **74**, 192-196.

[b]- Taberlet, P., J.J. Camarra, S. Griffin, E. Uhres, O. Hanotte, L.P. Waits, C. Dubois Paganon, T. Burke, and J. Bouvet. 1997. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology* **6**, 869-876.

[c]- Sanderlin, J.S., B.C. Faircloth, B. Shamblin, and M.J. Conroy. 2009. Tetranucleotide microsatellite loci from the black bear (*Ursus americanus*). *Molecular Ecology Resources* **9**, 288-291.

Table A1.2- Range of prior distributions for each parameter used in three population models between Ouachita (OU), Ozark (OZ), and Source black bear populations. Parameters included in each model are noted. See Figure 2.2 for a pictorial representation.

	Models			
Parameter	***Split***	***Admix***	***Founder***	**Priors**
Posterior Probability	0.044	0.942	0.014	
N_{OU}	✓	✓	✓	1-10,000
N_{OZ}	✓	✓	✓	1-10,000
N_{Source}	✓	✓	✓	10,000-40,000
$N_{OU\text{-}Post\text{-}translocation}$		✓	✓	1-1,000
$N_{OZ\text{-}Post\text{-}translocation}$		✓	✓	1-1,000
$N_{OU\text{-}Bottleneck}$	✓	✓		1-1,000
$N_{OZ\text{-}Bottleneck}$	✓	✓		1-1,000
$N_{OU\text{-}Historical}$	✓	✓		100-10,000
$N_{OZ\text{-}Historical}$	✓	✓		100-10,000
λ_{OU}		✓		0.100-0.999
λ_{OZ}		✓		0.100-0.999
t_1	✓	✓		14-20,000
t_2	✓	✓		1-10,000

Table A1.3- Microsatellite loci that display non-random assortment (+) or random assortment (-) of alleles within each study area. First panel: AR-Ou (above diagonal), OK-Ou (below diagonal); second panel: AR-Oz (above diagonal), OK-Oz (below diagonal); third panel: Manitoba (above diagonal), Minnesota (below diagonal); fourth panel: MO-Oz (above diagonal), MO (below diagonal).

	G10L	G10P	G1D	G10C	G10M	G1A	G10B	G10J	P2H03	G10U	MU10	G10O	MU59	MU05	MU23
G10L	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G10P	-	*	+	-	-	-	-	-	-	-	+	-	-	-	-
G1D	-	+	*	-	+	-	-	-	+	-	+	+	+	-	-
G10C	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-
G10M	-	-	-	-	*	-	+	-	-	-	-	-	-	-	+
G1A	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-
G10B	-	-	-	-	-	+	*	-	-	-	-	-	-	-	-
G10J	+	-	-	-	-	-	-	*	-	-	-	-	-	-	-
P2H03	-	-	-	-	-	-	-	-	*	-	-	-	+	-	-
G10U	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-
MU10	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-
G10O	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-
MU59	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-
MU05	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-
MU23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*

	G10L	G10P	G1D	G10C	G10M	G1A	G10B	G10J	P2H03	G10U	MU10	G10O	MU59	MU05	MU23
G10L	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G10P	-	*	+	-	-	-	-	-	-	-	-	-	-	-	-
G1D	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-
G10C	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-
G10M	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-
G1A	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-
G10B	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-
G10J	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-
P2H03	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-
G10U	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-
MU10	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-
G10O	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-
MU59	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-
MU05	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-
MU23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*

	G10L	G10P	G1D	G10C	G10M	G1A	G10B	G10J	P2H03	G10U	MU10	G10O	MU59	MU05	MU23
G10L	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-
G10P	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-
G1D	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-
G10C	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-
G10M	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-
G1A	-	-	-	-	-	*	-	-	-	-	-	-	-	-	-
G10B	-	-	-	-	-	-	*	-	-	-	-	-	-	-	-
G10J	-	-	-	-	-	-	-	*	-	-	-	-	-	-	-
P2H03	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-
G10U	-	-	-	-	-	-	-	-	-	*	-	-	-	-	-
MU10	-	-	-	-	-	-	-	-	-	-	*	-	-	-	-
G10O	-	-	-	-	-	-	-	-	-	-	-	*	-	-	-
MU59	-	-	-	-	-	-	-	-	-	-	-	-	*	-	-
MU05	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-
MU23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*

	G10L	G10P	G1D	G10C	G10M	G1A	G10B	G10J	P2H03	G10U	MU10	G10O	MU59	MU05	MU23
G10L	*	+	-	-	-	-	-	-	-	-	-	-	-	-	-
G10P	-	*	-	-	-	-	-	-	-	-	-	-	-	-	-
G1D	-	-	*	-	-	-	-	-	-	-	-	-	-	-	-
G10C	-	-	-	*	-	-	-	-	-	-	-	-	-	-	-
G10M	-	-	-	-	*	-	-	-	-	-	-	-	-	-	-
G1A	+	-	-	+	-	*	-	-	-	-	-	-	-	-	-
G10B	-	-	-	-	-	-	*	-	+	-	-	-	-	-	-
G10J	-	-	-	+	-	-	-	*	-	-	-	-	-	-	-
P2H03	-	-	-	-	-	-	-	-	*	-	-	-	-	-	-
G10U	-	-	-	-	+	-	-	-	-	*	-	-	-	-	-
MU10	-	-	-	-	-	-	-	-	-	-	*	-	-	+	-
G10O	-	-	+	-	+	+	-	-	-	-	-	*	-	-	-
MU59	-	-	-	+	-	-	-	-	-	-	-	+	*	-	-
MU05	-	-	-	-	-	-	-	-	-	-	-	-	-	*	-
MU23	-	-	-	-	-	+	-	-	-	-	-	+	-	-	*

Table A1.4- Mean and standard deviation of 10 replicate STRUCTURE runs at each number of a priori clusters (K), and change in the second order derivatives (ΔK) between successive K values.

	LnP(K)		
K	**Mean**	**Std**	**ΔK**
1	-19228.15	0.242	-
2	-18056.19	0.396	1,823.813
3	-17605.86	0.775	327.340
4	-17409.18	2.132	33.077
5	-17283.01	19.125	0.131
6	-17159.35	23.521	1.827
7	-17078.67	33.371	0.157
8	-16992.74	11.039	-

Table A1.5- Frequency of mitochondrial haplotypes in Source (Manitoba and Minnesota), Central Interior Highlands (Ouachitas, Ozarks, and Missouri), and historic (Minnesota) areas.

	Clade A- East														
	A	**B**	**C**	**F**	**G**	**H**	**K**	**O**	**P**	**R**	**S**	**V**	**X[3]**	**Z**	**Hap4**
Ontario[1]	16	111	68	103						1			116		17
Manitoba	1												1		
Manitoba[2]					2		1								
Minnesota		7					2						2	1	
Minnesota[2]		5													
MN-Historic								1	1	2	1				
Ouachitas		14		24								2			
Ouachitas[2]		12		83		1									
Ozarks		6				18									
Ozarks[2]		1				6									
Missouri-OZ		6		2		3									
Missouri		1													

	Clade A- West											Clade B
	D	**E**	**I**	**J**	**L**	**M**	**N**	**T**	**U**	**W**	**Y**	**Q**
Ontario[1]	19	83	2									
Manitoba	2	2		1	1		1					
Manitoba[2]	5	5	3	3		1						
Minnesota		1						6	1		6	
Minnesota[2]					1							
MN-Historic	2											1
Ouachitas		11								1		
Ouachitas[2]		18										
Ozarks		3										
Ozarks[2]		4										
Missouri-OZ		10										
Missouri		36										

[1]- Data from Pelletier *et al.* (2011) where summary statistics (n, *h*, and π) represent all samples (n=660) but only high frequency haplotypes from each clade were reproduced (n=419) for this table.
[2]- Data from Van Den Bussche *et al.* (2009) with MN summary statistics adjusted for six individuals reported in original table.
[3]- Identical at all 312 bp of Hap5 from Pelletier *et al.* (2011).

APPENDIX 2

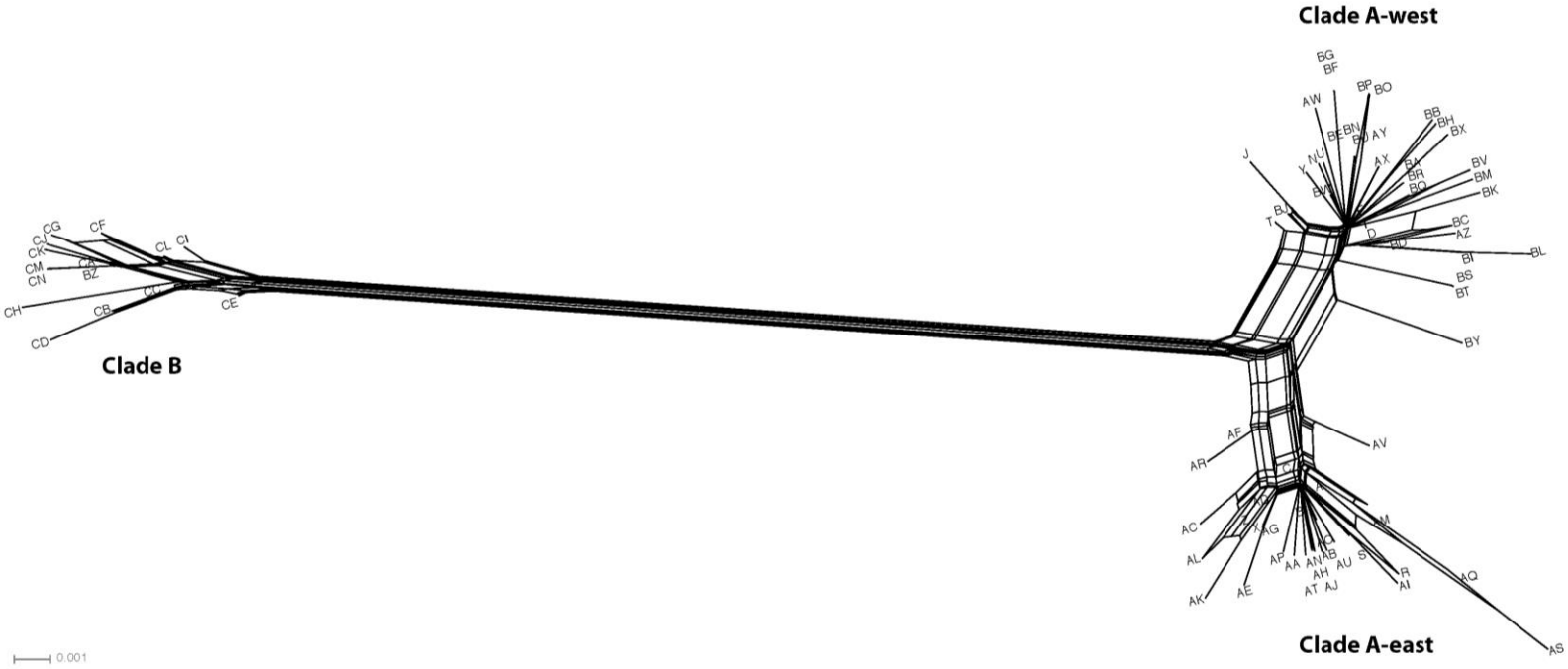

Figure A2.1- Network of mitochondrial haplotypes identified in Chapters 2 and 3. Clade B haplotypes are oriented towards left of page, clade A-east haplotypes oriented towards bottom of page, and clade A-west haplotypes oriented towards top of page. Individual haplotype names are at the tips of the network and their frequencies were described in Tables A1.5 and A3.4.

APPENDIX 3- Rangewide population genetic analyses of American black bears at 15 microsatellite loci.

Microsatellite genotyping and genetic diversity analysis

We genotyped all samples using the same loci and procedures as in Puckett *et al*. (2014). We removed samples that genotyped at fewer than 10 markers. As samples may have come from related individuals, we identified parent-offspring pairs within each sampling area in MLRelate (Kalinowski, et al. 2006) and removed one individual from each pair for a final sample size of 506 bears. We estimated expected (H_E) and observed (H_O) heterozygosity, and mean number of alleles (A) in each sampling area and genetic cluster (where samples were grouped to the closest cluster inferred from the SNP analysis) in ARLEQUIN v3.5 (Excoffier and Lischer 2010). See Table A3.1 for results.

Effective population size analysis

We used the microsatellite data to estimate effective population size (N_e) in the following groups identified in the SNP analysis: Southeast, Northeast, Great Lakes, AK-East, CIH, Southwest and West. We previously attempted to do this analysis with the SNP data; however, the models would not coalesce, hence the decision to use the microsatellite data. Additionally, the larger sample sizes per group in the microsatellite data also suggested that bias would be less than if using the SNP datasets.

Since population structure upwardly biases estimates of θ, samples were structured by state within each group as informed by the FINESTRUCTURE analysis (Figure 3.1). The structure for each group was as such:

Southeast: ((Florida, North Carolina-coast), (North Carolina-mountains, West Virginia))

Northeast: ((Maine/Massachusetts, New York), (Michigan-Lower Peninsula, Pennsylvania))

Great Lakes: ((Michigan-Upper Peninsula, Wisconsin), (Manitoba, Minnesota))

AK-East: ((Alaska-ABC, Yukon), Alaska-Kenai Peninsula)

CIH: (Missouri, Arkansas-Ozarks)

Southwest: ((Arizona, Colorado), New Mexico)

West: ((Oregon, Idaho), Montana)

Simulation models were run in DIYABC v2 (Cornuet, et al. 2014) with 10^6 simulations per group. We estimated a single N_e over the full model with a uniform prior of 10-10,000. Each divergence event was given the uniform prior of 10-3,000 generations. All models were checked using DIYABC's internal PCA program to ensure that the observed data fell within the simulation state space.

We used the 1% of the data closest to the observed data to estimate θ and μ. We report the mode and 90% highest density probability (HPD) for both metrics; however, the calculations for N_e used only the modal value of μ for both the point estimate and 90% HPD (Table A3.3).

LITERATURE CITED

Cornuet J-M, Pudlo P, Veyssier J, *et al.* (2014) DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using single nucleotide polymorphism, DNA sequence, and microsatellite data. *Bioinformatics* **30**, 1187-1189.

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.

Kalinowski ST, Wagner AP, Taper ML (2006) ML Relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* **6**, 576-579.

Puckett EE, Kristensen TV, Wilton CM, *et al.* (2014) Influence of drift and admixture on population structure of American black bears (*Ursus americanus*) in the Central Interior Highlands, USA 50 years after translocation. *Molecular Ecology* **23**, 2414-2427.

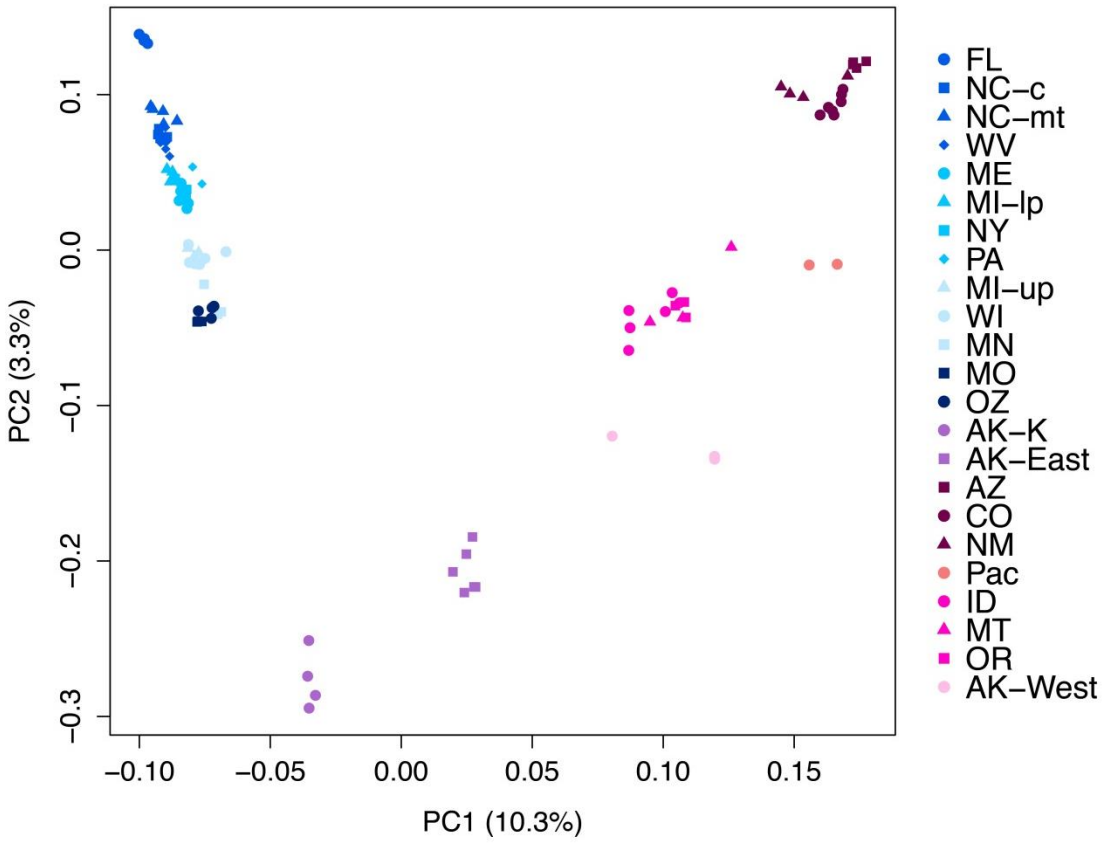

Figure A3.1- The first two principal components from a PCA using 22k SNPs in 94 American black bear samples distributed across the range. Samples were colored according to their nuclear genomic cluster (Alaska-East: purple; Central Interior Highlands: navy; Great Lakes: light blue; Northeast: bright blue; Southeast: medium blue; Alaska-West: light pink; West: pink; Pacific Coast: coral; Southwest: raspberry).

Table A3.1- Sample size (n), mean number of alleles (A), observed (H_O) and expected (H_E) heterozygosity, haplotype diversity (*h*) and standard deviation, and nucleotide diversity (π) and standard deviation. Microsatellite data were calculated with 15 loci.

		Microsatellites				Mitochondrion				
Study Area	**Abbreviation**	**n**	**A**	**H_O**	**H_E**	**n**	***h***	**h_{sd}**	**π**	**π_{sd}**
Southwest		**39**	**6.8**	**0.576**	**0.624**	**34**	**0.8699**	**0.0367**	**0.0064**	**0.0037**
Arizona	AZ	8	4.1	0.515	0.577	10	0.8444	0.1029	0.0051	0.0033
Colorado	CO	20	5.3	0.615	0.649	15	0.8286	0.0643	0.0039	0.0026
New Mexico	NM	11	4.8	0.551	0.600	9	0.6389	0.1258	0.0013	0.0012
West		**94**	**10.5**	**0.700**	**0.770**	**64**	**0.9484**	**0.0121**	**0.0203**	**0.0104**
Idaho	ID	57	9.3	0.679	0.760	25	0.9067	0.0350	0.0213	0.0111
Montana	MT	28	7.8	0.748	0.786	29	0.9286	0.0251	0.0128	0.0069
Oregon	OR	9	5.6	0.783	0.775	10	0.7778	0.0907	0.0198	0.0111
Alaska and Canada		**77**	**10.2**	**0.719**	**0.833**	**71**	**0.9131**	**0.0177**	**0.0160**	**0.0082**
Alaska- Kenai Peninsula	AK-K	10	5.4	0.725	0.713	10	0.3778	0.1813	0.0007	0.0008
Alaska- ABC Islands	AK-ABC	20	7.1	0.627	0.795	19	0.9591	0.0277	0.0144	0.0078
Alberta	AB	13	7.4	0.739	0.811	10	0.7111	0.1175	0.0027	0.0020
British Columbia	BC	19	8.1	0.799	0.814	18	0.7516	0.1031	0.0194	0.0104
Yukon	YT	15	7.5	0.714	0.795	14	0.7033	0.1088	0.0043	0.0028
Great Lakes		**57**	**10.9**	**0.789**	**0.814**	**52**	**0.8824**	**0.0298**	**0.0069**	**0.0039**
Manitoba	MB	9	6.9	0.822	0.821	9	0.9167	0.0725	0.0068	0.0043
Michigan- Upper Peninsula	MI-up	10	6.2	0.822	0.808	10	0.7778	0.1374	0.0070	0.0043
Minnesota	MN	28	8.5	0.791	0.787	26	0.8492	0.0382	0.0069	0.0040
Wisconsin	WI	10	6.1	0.713	0.766	7	0.8571	0.1371	0.0059	0.0040
Northeast		**110**	**10.7**	**0.778**	**0.812**	**63**	**0.8715**	**0.0214**	**0.0052**	**0.0031**
Connecticut	CT	12	4.7	0.694	0.667	14	0.2747	0.1484	0.0010	0.0009
Massachusetts[1]	MA	5	4.5	0.923	0.793	14	0.5714	0.1322	0.0029	0.0021
Maine[1]	ME	10	6.7	0.853	0.830					

		Microsatellites				Mitochondrion				
Study Area	**Abbreviation**	**n**	**A**	**H_O**	**H_E**	**n**	***h***	**h_{sd}**	**π**	**π_{sd}**
Michigan- Lower Peninsula	MI-lp	10	6.1	0.808	0.790	10	0.2000	0.1541	0.0022	0.0017
New York	NY	7	5.5	0.719	0.796	6	0.8667	0.1291	0.0078	0.0052
Ontario[2]	ON	51	9.5	0.800	0.805					
Pennsylvania	PA	15	6.8	0.702	0.760	19	0.5789	0.1138	0.0053	0.0032
East		**62**	**8.6**	**0.747**	**0.793**	**60**	**0.7192**	**0.0379**	**0.0034**	**0.0022**
Florida[1]	FL	9	4.3	0.601	0.636	11	0.5091	0.1008	0.0009	0.0009
Mississippi[1]	MS	2	2.3	0.682	0.636					
North Carolina- Coast	NC-c	14	6.5	0.765	0.781	14	0.6264	0.1098	0.0016	0.0013
North Carolina- Mountains	NC-mt	17	6.8	0.801	0.786	17	0.8015	0.0761	0.0055	0.0034
West Virginia	WV	20	7.2	0.743	0.779	18	0.3072	0.1316	0.0013	0.0011
CIH[3]		**67**	**7.6**	**0.774**	**0.748**	**58**	**0.3587**	**0.0753**	**0.0038**	**0.0024**
Missouri	MO	39	5.6	0.775	0.687	37	0.0541	0.0505	0.0006	0.0007
Missouri	OZ	28	6.9	0.773	0.772	21	0.6952	0.0697	0.0066	0.0039

[1]- Adjacent sampling areas were combined for haplotype analysis when one study area had a small sample size.

[2]- See Pelletier *et al*. (2011) for an extensive survey of haplotype diversity in Ontario.

[3]- All samples collected in Missouri. MO refers to unique cluster of samples identified in Puckett *et al.* (2014). OZ refers to samples from Missouri that clustered with the larger Ozark Mountains cluster in Puckett *et al.* (2014).

Table A3.2- Sample sizes, average minor allele frequency (MAF), and average inbreeding coefficient (F_{IS}) of the SNP dataset reported by inferred clusters.

Study Area	**n**	**MAF**	**F_{IS}**
AK-East	10	0.097	0.358
CIH	7	0.093	0.273
Great Lakes	12	0.092	0.221
Northeast	14	0.091	0.248
Southeast	19	0.089	0.361
AK-West	3	0.110	0.436
West	12	0.105	0.300
Pacific Coast	2	0.118	0.397
Southwest	15	0.111	0.509

Table A3.3- Estimates of the mode and 90% highest density probabilities for θ (=$N_e\mu$) and μ for each of the seven unadmixed (Southeast, AK-East, CIH, and Southwest) and admixed (Northeast, Great Lakes, and West) groups inferred from coalescent models in DIYABC using 15 microsatellites. Calculations for N_e are also presented where the point estimate was used for final calculations of divergence or admixture time.

	θ	**μ**	**N_e**
Southeast	1.19	2.08×10^{-4}	5720
	0.82 - 2.02	1.30×10^{-4} - 7.58×10^{-4}	3940 - 9710
Northeast	1.05	3.24×10^{-4}	3240
	0.88 -4.83	2.09×10^{-4} - 8.56×10^{-4}	2720 - 14900
Great Lakes	2.50	8.84×10^{-4}	2830
	1.10 - 8.10	3.01×10^{-4} - 9.44×10^{-4}	1240 - 9160
AK-East	1.33	3.24×10^{-4}	4100
	1.11 - 3.00	2.04×10^{-4} - 8.66×10^{-4}	3430 - 9260
CIH	0.82	1.85×10^{-4}	4450
	0.70 - 1.71	1.23×10^{-4} - 8.19×10^{-4}	3780 - 9240
Southwest	0.91	9.70×10^{-5}	9380
	0.61 - 1.97	6.17×10^{-5} - 5.29×10^{-4}	6290 - 20300
West	1.21	1.59×10^{-4}	7610
	1.03 - 4.30	1.06×10^{-5} - 6.87×10^{-4}	6480 - 27000

Table A3.4- Counts of mitochondrial haplotypes in each sampling area with sums across each genetic cluster and the range (Total). Proportion of haplotypes belonging to clades A-east, A-west, and B for each sampling area, cluster, and range are also presented.

	Clade A-East										
	A	**B**	**C**	**F**	**H**	**R**	**S**	**X**	**Z**	**AA**	**AB**
Southwest	0	2	0	0	0	0	0	0	0	1	4
AZ		2								1	4
NM											
CO											
West	0	0	0	0	0	0	0	0	0	0	0
ID											
MT											
OR											
Alaska & Canada	0	1	0	0	0	0	0	0	0	0	0
AK-ABC											
AK-Kenai											
AB											
BC											
YT		1									
Great Lakes	2	15	0	0	0	0	0	4	3	0	0
MB	1							2			
MN		7						2	1		
MI-up		5							1		
WI	1	3							1		
Northeast	0	1	12	0	0	2	1	9	0	0	0
CT											
MA											
ME		1									
MI-lp								9			
NY						2	1				
PA			12								
Southeast	0	25	19	0	0	0	0	0	0	1	0
FL		2	7								
MS		2									
NC-mt		2	3								
NC-c		4	8							1	
WV		15	1								
CIH[1]	0	7	0	2	3	0	0	0	0	0	0
MO		1									
OZ		6		2	3						
Total	**2**	**51**	**31**	**2**	**3**	**2**	**1**	**13**	**3**	**2**	**4**

	Clade A-East									
	AC	**AD**	**AE**	**AF**	**AG**	**AH**	**AI**	**AJ**	**AK**	**AL**
Southwest	1	1	0	0	0	0	0	0	0	0
AZ	1	1								
NM										
CO										
West	0	0	0	0	0	0	0	0	0	0
ID										
MT										
OR										
Alaska & Canada	0	0	0	0	0	0	0	0	0	0
AK-ABC										
AK-Kenai										
AB										
BC										
YT										
Great Lakes	0	0	0	0	0	0	0	1	1	1
MB										
MN										
MI-up								1	1	1
WI										
Northeast	0	0	1	1	16	9	1	0	0	0
CT			1	1	12					
MA					3	1				
ME						8	1			
MI-lp										
NY										
PA					1					
Southeast	0	0	0	0	0	0	0	0	0	0
FL										
MS										
NC-mt										
NC-c										
WV										
CIH[1]	0	0	0	0	0	0	0	0	0	0
MO										
OZ										
Total	**1**	**1**	**1**	**1**	**16**	**9**	**1**	**1**	**1**	**1**

	Clade A-Easat										Proportion
	AM	**AN**	**AO**	**AP**	**AQ**	**AR**	**AS**	**AT**	**AU**	**AV**	
Southwest	0	0	0	0	0	0	0	0	0	0	0.26
AZ											0.90
NM											0.00
CO											0.00
West	0	0	0	0	0	0	0	0	0	0	0.00
ID											0.00
MT											0.00
OR											0.00
Alaska & Canada	0	0	0	0	0	0	0	0	0	0	0.02
AK-ABC											0.00
AK-Kenai											0.00
AB											0.00
BC											0.00
YT											0.07
Great Lakes	0	2	0	0	0	0	0	0	0	0	0.56
MB											0.33
MN		2									0.46
MI-up											0.90
WI											0.71
Northeast	1	0	0	0	1	2	4	1	1	0	1.00
CT											1.00
MA											1.00
ME											1.00
MI-lp	1										1.00
NY					1	2					1.00
PA							4	1	1		1.00
Southeast	0	0	3	1	0	0	0	0	0	2	0.85
FL											1.00
MS											1.00
NC-mt			2	1							0.47
NC-c			1								1.00
WV										2	1.00
CIH[1]	0	0	0	0	0	0	0	0	0	0	0.21
MO											0.03
OZ											0.52
Total	**1**	**2**	**3**	**1**	**1**	**2**	**4**	**1**	**1**	**2**	**0.45**

	Clade A-West											
	D	**E**	**I**	**J**	**N**	**T**	**U**	**Y**	**AW**	**AX**	**AY**	**AZ**
Southwest	10	5	1	0	0	0	0	0	0	0	0	0
AZ	1											
NM	5	3	1									
CO	4	2										
West	13	8	2	0	0	0	0	0	0	0	0	0
ID	4	6	1									
MT	5	2	1									
OR	4											
Alaska & Canada	8	15	5	0	1	0	0	0	1	8	1	1
AK-ABC	3	2										1
AK-Kenai	1									8	1	
AB	1	5	3						1			
BC	2	1	2		1							
YT	1	7										
Great Lakes	2	6	0	1	1	6	1	5	0	0	0	0
MB	2	2		1	1							
MN		2				6	1	5				
MI-up		1										
WI		1										
Northeast	0	0	0	0	0	0	0	0	0	0	0	0
CT												
MA												
ME												
MI-lp												
NY												
PA												
Southeast	7	2	0	0	0	0	0	0	0	0	0	0
FL												
MS												
NC-mt	7	2										
NC-c												
WV												
CIH[1]	0	46	0	0	0	0	0	0	0	0	0	0
MO		36										
OZ		10										
Total	**40**	**82**	**8**	**1**	**2**	**6**	**1**	**5**	**1**	**8**	**1**	**1**

	Clade A-West											
	BA	**BB**	**BC**	**BD**	**BE**	**BF**	**BG**	**BH**	**BI**	**BJ**	**BK**	**BL**
Southwest	0	0	0	0	0	0	0	0	5	2	1	1
AZ												
NM												
CO									5	2	1	1
West	0	0	0	0	0	0	0	0	0	0	0	0
ID												
MT												
OR												
Alaska & Canada	2	1	1	2	1	1	1	1	0	0	0	0
AK-ABC	2	1	1	2	1	1	1	1				
AK-Kenai												
AB												
BC												
YT												
Great Lakes	0	0	0	0	0	0	0	0	0	0	0	0
MB												
MN												
MI-up												
WI												
Northeast	0	0	0	0	0	0	0	0	0	0	0	0
CT												
MA												
ME												
MI-lp												
NY												
PA												
Southeast	0	0	0	0	0	0	0	0	0	0	0	0
FL												
MS												
NC-mt												
NC-c												
WV												
CIH[1]	0	0	0	0	0	0	0	0	0	0	0	0
MO												
OZ												
Total	**2**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**5**	**2**	**1**	**1**

	Clade A-West													Proportion
	BM	**BN**	**BO**	**BP**	**BQ**	**BR**	**BS**	**BT**	**BU**	**BV**	**BW**	**BX**	**BY**	
Southwest	0	0	0	0	0	0	0	0	0	0	0	0	0	0.74
AZ														0.10
NM														1.00
CO														1.00
West	1	3	2	3	5	3	1	1	1	1	0	0	0	0.73
ID	1	3												0.60
MT			2	3	5	3	1	1	1	1				0.86
OR														0.40
Alaska & Canada	0	0	0	0	0	0	0	0	0	0	0	4	1	0.70
AK-ABC														0.84
AK-Kenai														1.00
AB														1.00
BC														0.33
YT												4	1	0.93
Great Lakes	0	0	0	0	0	0	0	0	0	0	1	0	0	0.44
MB														0.67
MN														0.54
MI-up														0.10
WI											1			0.29
Northeast	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00
CT														0.00
MA														0.00
ME														0.00
MI-lp														0.00
NY														0.00
PA														0.00
Southeast	0	0	0	0	0	0	0	0	0	0	0	0	0	0.15
FL														0.00
MS														0.00
NC-mt														0.53
NC-c														0.00
WV														0.00
CIH[1]	0	0	0	0	0	0	0	0	0	0	0	0	0	0.79
MO														0.97
OZ														0.48
Total	**1**	**3**	**2**	**3**	**5**	**3**	**1**	**1**	**1**	**1**	**1**	**4**	**1**	**0.47**

	Clade B								
	BZ	**CA**	**CB**	**CC**	**CD**	**CE**	**CF**	**CG**	**CH**
Southwest	0	0	0	0	0	0	0	0	0
AZ									
NM									
CO									
West	3	3	0	0	0	0	3	1	1
ID	2						3	1	1
MT	1								
OR		3							
Alaska & Canada	1	2	9	1	1	1	0	0	0
AK-ABC	1	2							
AK-Kenai									
AB									
BC			9	1	1	1			
YT									
Great Lakes	0	0	0	0	0	0	0	0	0
MB									
MN									
MI-up									
WI									
Northeast	0	0	0	0	0	0	0	0	0
CT									
MA									
ME									
MI-lp									
NY									
PA									
Southeast	0	0	0	0	0	0	0	0	0
FL									
MS									
NC-mt									
NC-c									
WV									
CIH[1]	0	0	0	0	0	0	0	0	0
MO									
OZ									
Total	**4**	**5**	**9**	**1**	**1**	**1**	**3**	**1**	**1**

	Clade B						Proportion
	CI	**CJ**	**CK**	**CL**	**CM**	**CN**	
Southwest	0	0	0	0	0	0	0.00
AZ							0.00
NM							0.00
CO							0.00
West	3	1	1	1	2	1	0.27
ID	1	1	1				0.40
MT	2			1			0.14
OR					2	1	0.60
Alaska & Canada	0	0	0	0	0	0	0.28
AK-ABC							0.16
AK-Kenai							0.00
AB							0.00
BC							0.67
YT							0.00
Great Lakes	0	0	0	0	0	0	0.00
MB							0.00
MN							0.00
MI-up							0.00
WI							0.00
Northeast	0	0	0	0	0	0	0.00
CT							0.00
MA							0.00
ME							0.00
MI-lp							0.00
NY							0.00
PA							0.00
Southeast	0	0	0	0	0	0	0.00
FL							0.00
MS							0.00
NC-mt							0.00
NC-c							0.00
WV							0.00
CIH[1]	0	0	0	0	0	0	0.00
MO							0.00
OZ							0.00
Total	**3**	**1**	**1**	**1**	**2**	**1**	**0.08**

[1]- Samples for the CIH were from Puckett *et al.* (2014) where MO was from the unique Missouri cluster, and OZ were samples collected in Missouri that grouped with the broader Ozarks cluster.

APPENDIX 4- Supplemental figures and tables for Chapter 4

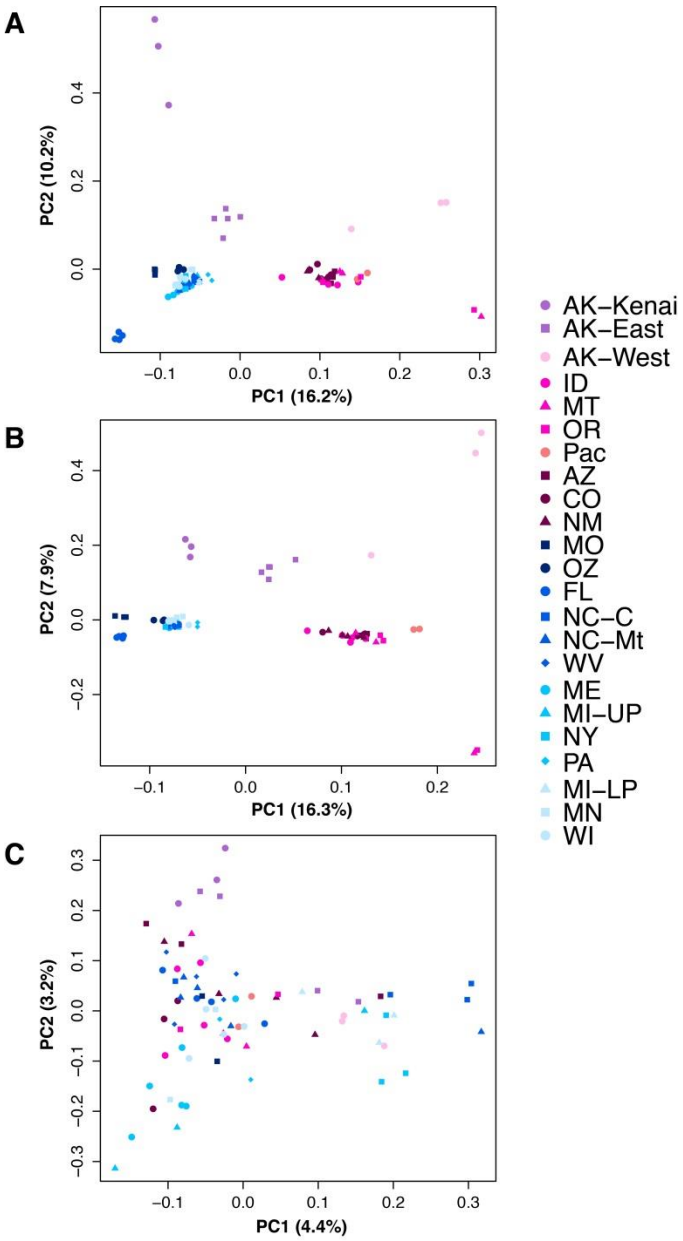

Figure A4.1- PCA of the (A) *Equal200* and (B) *Equal1000* SNP datasets, and (C) *PartialMicro* microsatellite dataset for black bears across the range. Sample locations were organized by symbology (abbreviations described in Table A4.1) where colors identified genomic cluster (raspberry, Southwest; pink, West; coral, Pacific Coast; light pink, Alaska- ABC Islands west cluster; purple, Alaska- ABC Islands east cluster; navy, Central Interior Highlands; light blue, Great Lakes; bright blue, Northeast; medium blue, Southeast).

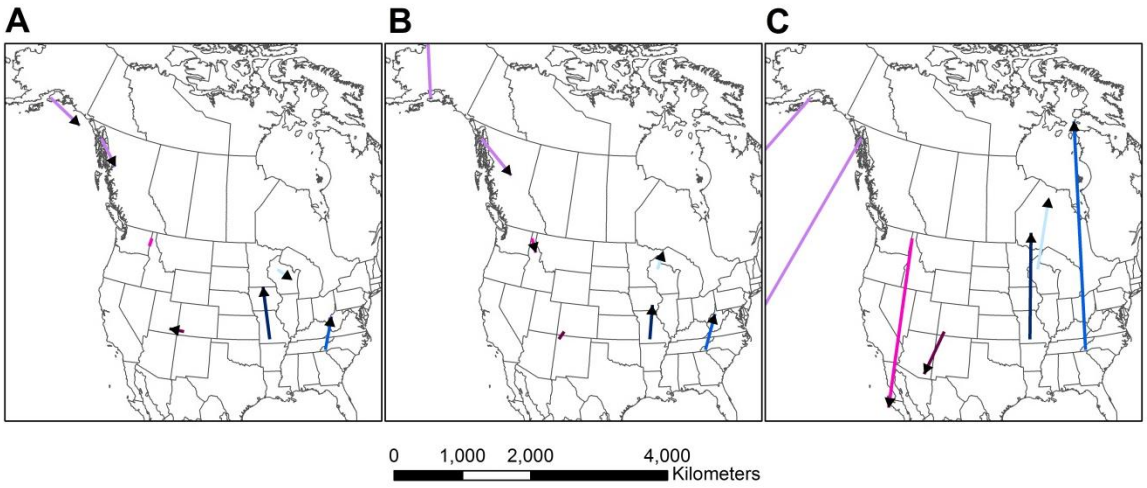

Figure A4.2- Map of true and estimated locations for test samples with the spatial smoothing method with the *Equal200* (A) and *Equal1000* (B) datasets. Arrows point towards estimated locations.

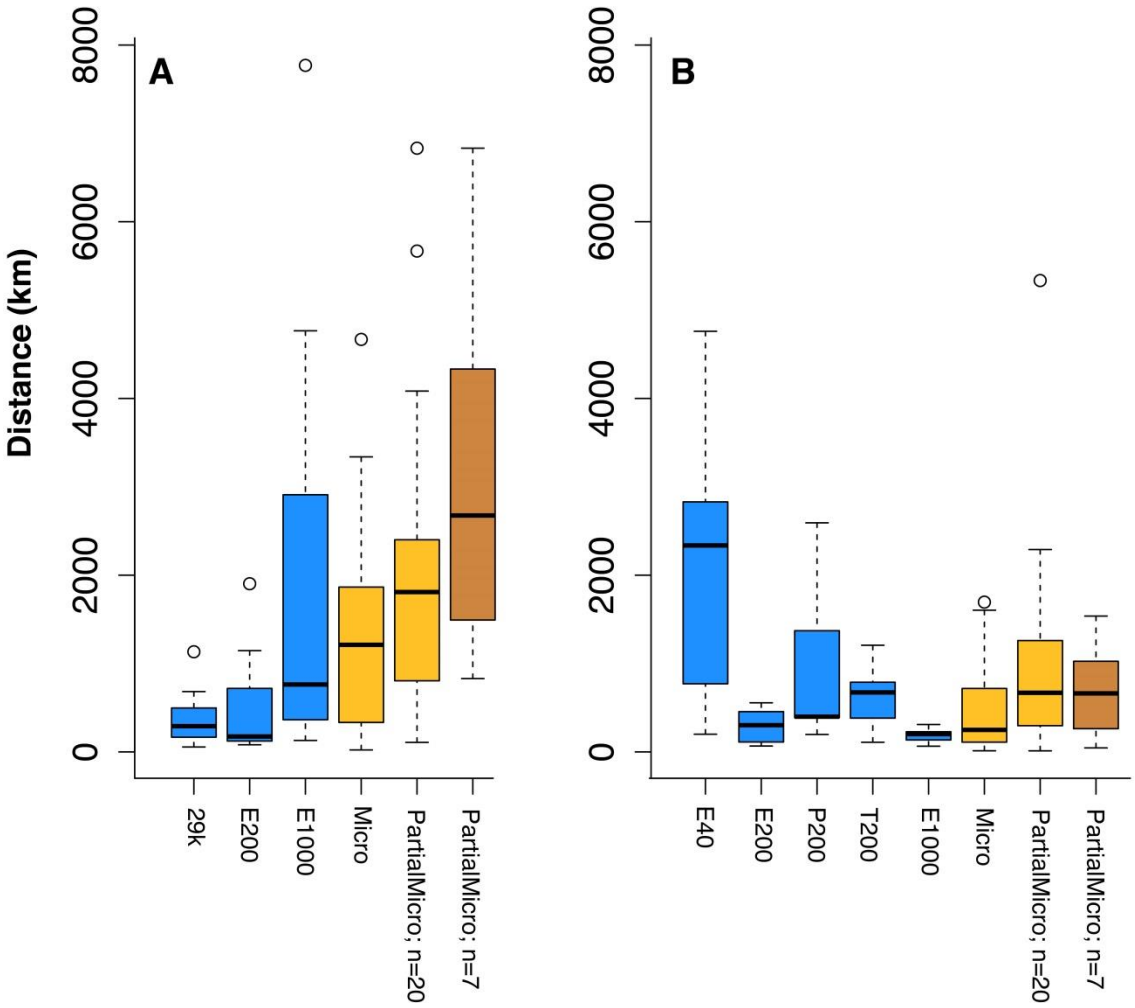

Figure A4.3- Box plots of distance between true and estimated locations using both the principal component regression (A) and spatial smoothing (B) methods with SNP (blue) and microsatellite (gold and brown for the microsatellite dataset comparable to the SNPs) datasets. Properties of datasets were detailed in Table 4.1.

Table A4.1- Counts of training and test samples used for spatial smoothing and PC regression analyses by marker type. Study area corresponds to the regional clusters employed within SCAT.

		Microsatellites		SNPs	
Study Area	**Abbreviation**	**Training**	**Test**	**Training**	**Test**
Southwest					
Arizona	AZ	8	0	4	0
Colorado	CO	19	1	6	1
New Mexico	NM	11	0	4	0
West					
Idaho	ID	56	1	5	1
Montana	MT	28	0	3	0
Oregon	OR	7	0	3	0
Oregon- Pacific Coast cluster	Pac	2	0	2	0
Alaska and Canada					
Alaska- Kenai Peninsula	AK-Kenai	9	1	3	1
Alaska- ABC Islands (East cluster)	AK-East	9	1	5	1
Alaska- ABC Islands (West cluster)	AK-West	10	0	3	0
Alberta	AB	12	1	0	0
British Columbia	BC	19	0	0	0
Yukon	YT	14	1	0	0
Great Lakes					
Manitoba	MB	8	1	0	0
Michigan- Upper Peninsula	MI-up	10	0	3	0
Minnesota	MN	27	1	3	0
Wisconsin	WI	9	1	5	1
Northeast					
Connecticut	CT	12	0	0	0
Massachusetts	MA	4	1	0	0
Maine	ME	10	0	6	0
Michigan- Lower Peninsula	MI-lp	10	0	3	0
New York	NY	7	0	3	0
Ontario	ON	50	1	0	0

Study Area	Abbreviation	Microsatellites		SNPs	
		Training	Test	Training	Test
Pennsylvania	PA	14	1	2	0
Southeast					
Florida	FL	8	1	4	0
Mississippi	MS	0	2	0	0
North Carolina- Coast	NC-c	14	0	4	0
North Carolina- Mountains	NC-mt	16	1	5	1
West Virginia	WV	19	1	5	0
Central Interior Highlands					
Missouri	MO	39	0	3	0
Ozark Mountains	OZ	26	2	3	1

Table A4.2- The median distance (km) and range between true and estimated locations using spatial smoothing with the *Equal200*, *Proportional200*, and *Top200* SNP datasets. All datasets had the same 87 training samples and 7 test samples. The datasets were not significantly different ($P = 0.15$) in an ANOVA.

	Spatial Smoothing		
Dataset	**Median**	**Min**	**Max**
Equal	302	66	555
Top	673	109	1207
Proportional	398	197	2592

Table A4.3- Eigenvalues for significant PCA axes using the SNP (29k, *Equal200*, and *Equal1000*) and microsatellite (*Micro* and *PartialMicro*) datasets, where ns indicates not significant.

PC	29k	Equal200	Equal1000	Micro	PartialMicro
1	10.9	16.2	16.3	23.2	4.4
2	3.3	10.2	7.9	14.0	3.2
3	2.4	9.4	7.0	9.0	2.8
4	2.2	7.5	5.8	8.5	ns
5	1.9	6.1	5.6	7.1	ns
6	1.7	5.0	4.4	6.5	ns
7	1.5	3.0	2.9	6.5	ns
8	1.4	2.5	2.2	ns	ns
9	1.4	2.0	1.9	ns	ns
10	1.3	1.9	1.1	ns	ns
11	ns	1.2	1.0	ns	ns
12	ns	1.2	ns	ns	ns
13	ns	1.0	ns	ns	ns

VITA

Emily E. Puckett was born May 24, 1982 and grew up in Winston-Salem, NC. Since a young age she has had an innate appreciation for nature despite not spending much time outside. This greatly informs her saying, "biology is an indoor activity." She graduated with a BS in Botany from North Carolina State University in 2004, then worked as a research assistant at the University of North Carolina until 2006. She earned her MS in Plant Physiology at the State University of New York- College of Environmental Science and Forestry in Syracuse, NY, studying the competitive uptake of phosphate and arsenate with Dr. Larry Smart. While at SUNY-ESF she learned of conservation biology and thought conservation genetics would well combine her desire to protect species with her strong molecular biology bench skills. Following graduation in 2008 she took a research technician position with Edenspace Systems Corporation in their Chantilly, VA office. She worked in plant biotechnology until moving to Columbia, MO to pursue a PhD with Dr. Lori Eggert in conservation genetics at the University of Missouri. Emily graduated in 2015 and looks forward to continuing her research in evolutionary biology.

Emily married Jay Puckett on February 4, 2010, the couple's 12^{th} dating anniversary. Emily and Jay have four cats, Tatum, Oslo, Cinco de Gato, and Lily. Their family turned 10 years old in 2014. Emily and Jay enjoy cooking, traveling, and spending time with each other.