American Genetic Association

Journal of Heredity, 2015, 1–11
doi:10.1093/jhered/esv061
Original Article

OXFORD

Original Article

Genetic Structure across Broad Spatial and Temporal Scales: Rocky Mountain Tailed Frogs (*Ascaphus montanus*; Anura: Ascaphidae) in t he Inland Temperate R ainforest

Genevieve Metzger, Anahi Espindola, Lisette P. Waits, and Jack Sullivan

From the Department of Biological Sciences, University of Idaho, Moscow, ID 84844-3051 (Metzger, Espindola, and Sullivan); Center for Research on Invasive Species and Small Populations, University of Idaho, Moscow, ID (Waits and Sullivan); Program in Bioinformatics and Computational Biology, University of Idaho, Moscow, ID (Metzger and Sullivan); and Department of Fish and Wildlife Sciences, University of Idaho, Moscow ID 83844-1136 (Waits).

*Address correspondence to Jack Sullivan at the address above, or e-mail: jacks@uidaho.edu.

Data deposited at Dryad: http://dx.doi.org/doi:10.5061/dryad.2pb57

Received December 15, 2014; First decision March 20, 2015; Accepted July 24, 2015.

Corresponding editor: H. Bradley Shaffer

Abstract

Contemporary and historical processes interact to structure genetic variation, however discerning between these can be difficult. Here, we analyze range-wide variation at 13 microsatellite loci in 2098 Rocky Mountain tailed frogs, *Ascaphus montanus*, collected from 117 streams across the species distribution in the Inland Northwest (INW) and interpret that variation in light of historical phylogeography, contemporary landscape genetics, and the reconstructed paleodistribution of the species. Further, we project species distribution models (SDMs) to predict future changes in the range as a function of changing climate. Genetic structure has a strong spatial signature that is precisely congruent with a deep (~1.8 MY) phylogeographic split in mtDNA when we partition populations into 2 clusters ($K = 2$), and is congruent with refugia areas inferred from our paleorange reconstructions. There is a hierarchical pattern of geographic structure as we permit additional clusters, with populations clustering following mountain ranges. Nevertheless, genetic diversity is the highest in populations at the center of the range and is attenuated in populations closer to the range edges. Similarly, geographic distance is the single best predictor of pairwise genetic differentiation, but connectivity also is an important predictor. At intermediate and local geographic scales, deviations from isolation-by-distance are more apparent, at least in the northern portion of the distribution. These results indicate that both historical and landscape factors are contributing to the genetic structure and diversity of tailed frogs in the Inland Northwest.

Subject areas: Population structure and phylogeography
Key words: landscape genetics, phylogeography, microsatellites, species distribution model.

Understanding the interactions between historical, evolutionary processes and contemporary, landscape processes in structuring patterns of genetic diversity is one of the biggest challenges in ecological and landscape genetics (Balkenhol et al. 2009; Sork and Waits 2010). That is, processes acting across diverse temporal and geographic scales influence genetic diversity in ways that can be difficult to

disentangle (Anderson et al. 2010). This is critical, however, because different questions in evolution and conservation may focus on different temporal or regional scales. For example, conservation biologists may be interested in current processes such as population connectivity and gene flow, whereas phylogeographers may be interested in historical processes, such as vicariance and dispersal from refugia, which often have acted across range-wide scales; this has impeded the integration of the inference across scales.

However, historical and contemporary processes can each be important at both local and range-wide scales (Table 1). Thus, investigations into the genetic diversity and structure of a taxon should be interpreted at multiple time scales, regardless of the geographic scale in which the researcher is primarily interested. This is especially true for species that have a long evolutionary history in a particular ecosystem. Here, we provide a broad, multiscale investigation into ecological and population genetics of *Ascaphus montanus* Mittleman and Myers 1949, the Rocky Mountain tailed frog, a species endemic to inland rainforests of the northwestern US (Figure 1), and elucidate the interaction of these factors in governing the genetic structure of the species across its range.

The cedar-hemlock forests of the Inland Northwest (INW) of North America represent the largest inland temperate rainforest in the world and, along with the coastal rainforest, forms a large disjunction that occurs in >156 species or species complexes (Nielson et al. 2001; Gavin 2009). Inland and coastal rainforest ecosystems are separated by >160 km of the xeric Columbia Basin, and comparative distributional and phylogeographic studies (Carstens et al. 2005; Gavin 2009; Bjork 2009; Carstens et al. 2013) have indicated that the inland rainforest ecosystem is composed of both old endemics (the result of a pre-Pleistocene vicariance) and recent arrivals (i.e., post-Pleistocene dispersers). Furthermore, phylogeographic studies of old endemics such as Constance's bittercress (*Cardamine constancei*; Brunsfeld and Sullivan 2005), Coeur d'Alene salamanders (*Plethodon idahoensis*; Carstens et al. 2004), and Rocky Mountain tailed frogs (see below; Nielson et al. 2006) have indicated that there was more than a single inland Pleistocene refugium for rainforest taxa.

The genus *Ascaphus* Stejneger is part of an ancient lineage that is either sister group to all other frogs (Ford and Canatella 1993) or is sister to the New Zealand genus *Leiopelma*, which then is sister to the remaining frogs (Pyron and Wiens 2011; Zhang et al. 2013). Regardless of which hypothesis represents the actual order of divergences, *Ascaphus* occupies a critical phylogenetic position relative to anuran amphibians. Tailed frogs occur in and around high-gradient mountain streams and suffer from anthropogenic habitat alteration (Spear and Storfer 2008, 2010). In spite of retention of several primitive features as adults (e.g., aspects of the palatoquadrate), they exhibit a number of derived features associated with their unusual habitat (e.g., oral suction discs in larvae and internal fertilization via an intromission organ).

Within *Ascaphus*, coastal and inland entities have been recognized as distinct species (*A. truei* and *A. montanus*, respectively) based on mitochondrial DNA (mtDNA; Nielsen et al. 2001), morphology, and allozymes (Nielson et al. 2006). Using coalescent analyses, Carstens et al. (2005) suggested that *A. montanus* has persisted in the inland region since the Pliocene and throughout Pleistocene climatic fluctuations. Further, Carstens and Richards (2007) suggested, based on projections of ecological niche models (ENMs) onto past climatic reconstructions, that *A. montanus* persisted in 2–3 refugia (a northern refugium in the Clearwater Drainage and possible southern refugia in the Salmon River Drainage and the Blue Mountains). The hypothesis of 2 refugia is supported by reciprocally monophyletic mtDNA clades that differ by as much as 2–3% uncorrected sequence divergence, and the divergence between these clades originated approximately 1 MYA (with a credibility interval of 0.092–5.4 MY; Nielson et al. 2006). The contact zone between northern and southern mtDNA clades is in the East Fork of the South Fork of the Salmon River (Figure 1, lower inset; Nielson et al. 2006). Thus, given this ancient phylogeograhic structure, long-term occupation of the INW, and the habitat specialization described above, current patterns of genetic variation and diversity are likely the result of complex interactions between both ancient and contemporary processes and between landscape and range-wide phenomena.

Here, we address the genetic structure and diversity of this ecologically specialized frog species. To do so, we address 3 hypotheses regarding geographic patterns of genetic diversity in *A. montanus*. First, we hypothesize that refugial structure will be detectable at the range-wide spatial scale (i.e., geographic congruence between microsatellite and mtDNA divergences). Second, habitat variables should influence dispersal in such a way that a correlation between divergence and environmental variation (as represented by a species distribution model, or SDM) will be detectable across the range. Third, habitat variables may have a stronger effect on divergence at local geographic scales, whereas historical factors (i.e., refugial structure) will be less important locally. We address these hypotheses with fine-scale, range-wide sampling using nuclear DNA microsatellite loci and high-resolution SDMs.

Table 1. Temporal and spatial scales of phenomena that structure patterns of genetic diversity

	Local-scale	Rangewide
Contemporary	SDM/connectivity	Isolation by distance
	Gene flow	Patterns of diversity
Historical	Introgression	Refugium/refugia
	Suture zones	Climatic fluctuations

Methods

Sampling

Two hundred twenty-eight streams were sampled for the presence of *Ascaphus* (Figure 1; Supplementary Table S1 online); this entailed kick-sampling for an hour at each site. If no tadpoles were found, frogs were considered to be absent at that site. Sampling typically occurred over a 40-m stretch of a stream and we collected multiple age classes at each site (*A. montanus* has a 4-year larval stage). Site selection was designed to cover the entire distribution of *A. montanus* except the very small portion of the distribution in Canada (i.e., the Yaak River drainage). In addition, several paired headwater sites in different drainages were sampled, and one region in northern Idaho was sampled very intensively to provide for a future analysis of local-scale landscape genetics. All sampling was conducted under University of Idaho ACUC protocol 2007–14 and collecting permits from Idaho, Montana, Oregon, and Washington.

Species Distribution Modeling

We extended the SDM for *Ascaphus* conducted by Carstens and Richards (2007) by refining the spatial resolution of 19 bioclimatic variables to 1 km^2 (WORLDCLIM, Hijmans et al. 2005) and by adding one land-cover (i.e., coniferous cover) variable (USGS). We restricted our SDM to *A. montanus*, used presences previously

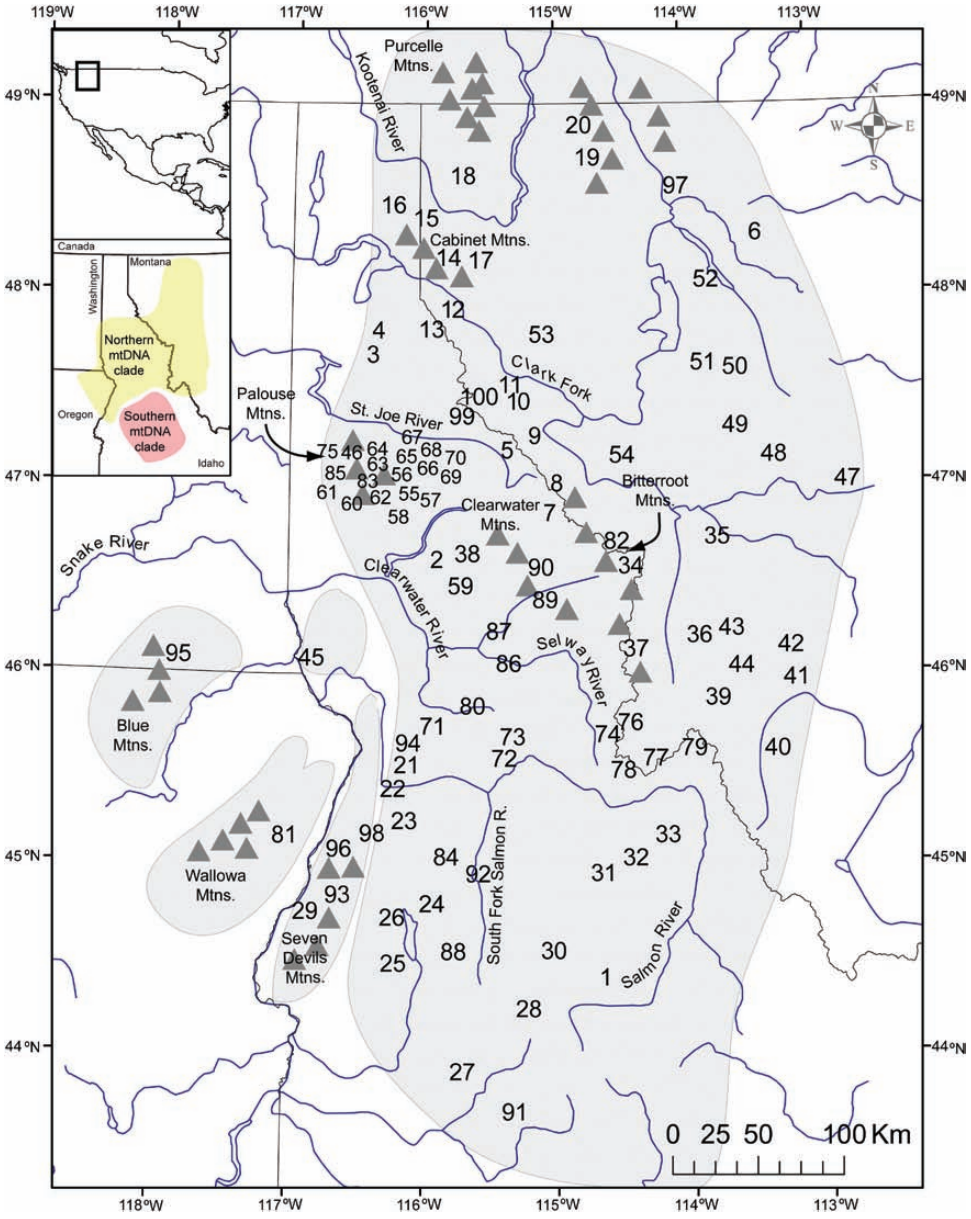

Figure 1. Range of *Ascaphus montanus* (shaded) and collection localities. The upper inset indicates the location of the study region in the continental US; the lower inset shows the study region and illustrates the distributions of northern and southern mtDNA clades reported in Nielson et al. (2006). The entire region is mountainous; names of rivers and subranges discussed in the text are indicated by grey triangles.

known (11) in addition to those obtained from our kick-sampling of 228 sites (176 presences total), and 10 000 background samples. Background samples included 126 locations where we searched but did not observe the species (i.e., true absences), and a set of 9874 background points in the study area that we selected using geospatial manipulations. We used an approach similar to that used by Warren et al. (2014); we first created a continuous density surface that represented the sampling effort, with higher values in better-sampled areas. We then sampled background points using the values of this surface to influence the probability of a point being drawn such that a well-sampled region had a higher probability of contributing to the background set (Searcy and Schaffer 2104).

The SDM was generated using the maximum entropy machine-learning algorithm implemented in Maxent (Phillips et al. 2006), and we randomly selected 30% of the samples for cross-validation. In addition, in order to address the correlation among climatic variables, we conducted the SDM on a reduced set of uncorrelated variables (Peterson et al. 2011; Warren et al 2014). The SDM calculated from the reduced data set was nearly identical to the full data set (see below).

As in Carstens and Richards (2007), we also projected a paleodistribution for *A. montanus* using 19 bioclimatic variables estimated for the last glacial maximum (LGM; 21 KYA) under the community climate system model (CCM3; Collins et al. 2006). Further, to estimate future variation in the range of the species, we used climatic projections of the same 19 bioclimatic variables (WORLDCLIM; Hijmans et al. 2005). These projections correspond to predictions for the year 2070, using a Global Circulation Model (MRI-CGCM3),

a worst-case CO_2-emission scenario (8.5 representative concentration pathways), and a 30-s resolution. Because strong range changes affect the spatial genetic structure of species, and can influence the survival of particular genetic groups in the absence of adaptation (Espíndola et al. 2012), we used the outputs from the current and forecasted projections to quantify the spatial and temporal variation of niche suitabilities for the total range as well as for each genetic cluster at $K = 2$ (see below). Using ArcGIS tools, we extracted all grid values within either the study area or in a polygon containing all sampled locations included in each genetic cluster for $K = 2$. To improve the biological reality of this last analysis, and to implement some measure of dispersal potential, we added a 20-km ("low") and a 100-km ("high") buffer area to the polygons, and we also extracted all suitability values contained under these new polygons. We then calculated the proportion of grids harboring suitability values higher than 0.5 (>0.5 dataset) or 0.7 (>0.7 dataset) in the future, respective to current values. This was done in R using custom scripts (available in the Supplementary Material online).

Genetic Data Collection

Genomic DNA was extracted from approximately 25 mg of tail tissue from 2098 individuals (primarily tadpoles) collected at 117 sites using standard protocols for the DNeasy Tissue kit (Qiagen). Following the procedure of Spear et al. (2008) for *A. truei*, we performed 3 multiplex PCR panels on each sample using the Qiagen Multiplex PCR kit and included a negative control in each run, for a maximum of 13 microsatellite loci amplified for each sample. Products from each PCR panel were run on an ABI 3730 automated sequencer (Applied Biosystems, Inc.) at the University of Idaho College of Natural Resources and genotyped using GeneMapper 3.6 Software (Applied Biosystems, Inc.). To improve the accuracy of allele calls, alleles were binned using Flexibin (Amos et al. 2007) and, genotyping was repeated for 250–600 individuals for each locus. We used FreeNA (Chapuis and Estoup 2007) to calculate the proportion of null alleles present for each microsatellite locus and each locus-population combination. We genotyped a range of 1–40 individuals from each site with an average of 18.4. Seventeen sites with <10 individuals genotyped were excluded from the STRUCTURE and regression analyses described below. In accordance with data archiving policy (Baker 2013), we have deposited collection localities and multilocus genotypes in Dryad.

Genetic Diversity and Structure

Standard descriptive statistics (expected and observed heterozygosity, allelic richness, F_{ST}) were calculated with GenePop 4.0.10 (Raymond and Rousset 1995; Rousset 2008). We also calculated $F_{ST,}$ corrected for the presence of null alleles, using FreeNA (Chapuis and Estoup 2007). We then evaluated population structure with no a priori constraints using the program STRUCTURE (Pritchard et al. 2000) under its admixture model. Because several populations exhibited extremely low allelic diversity (Figure 2A), we conducted separate analyses assuming codominance and recessive expression, but based most of our inferences from the analyses (correctly) assuming codominance. We varied K (the number of clusters of multilocus genotypes) from 1 to 15, and conducted 10 replicate runs at each value of K. Each of the 150 analyses was run for 1 million generations prior to sampling the posterior distributions for another 1 million generations. We assessed the relative importance of each value of K using 2 methods. First, we used the ΔK criterion, which was calculated by dividing the mean difference in likelihood for successive values of K by the standard deviation of $L(K)$ (Evanno et al. 2005). Second, we assessed $P(D|K)$ qualitatively and then compared that to the geographic distribution of the clusters for each value of K. This occurred very close to the value of K that maximized the average $P(D|K)$. This approach is particularly appropriate when there is hierarchical structure in the data and we also used FreeNA (Chapuis and Estoup 2007) to visualize this structure with a NJ-tree calculated from chord distances (Cavalli-Sforza and Edwards 1967).

Geography and Genetics

We performed linear and multivariate regression analyses using *R* to examine correlations between geographic variables and genetic distance (linearized F_{ST}). We performed linear regressions to evaluate the correlations between different distance metrics (see below) and genetic divergence. Inter-population distances were obtained following different approaches: 2 habitat aware [least cost path (LCP) and CircuitScape (CS) distances] and 2 raw geographic distances (Euclidian and aquatic distances).

Results obtained from our SDMs were used to calculate LCP distances between populations; we transformed the suitability values assigned to each raster cell by calculating 1-SDM suitability value for each cell. Thus, values with high suitabilities in the SDM projection convey a minimal cost in the transformed cost raster. We then transformed the resulting raster by multiplying the values of those cells with values higher or equal to 0.85 by 1, 10, 100, or 1000 to increase progressively the cost assigned to low suitability cells. The final LCP for each of these 4 cost layers was calculated using the Cost Path tool in ArcMap 9.3 and the (ESRI, Redlands, CA).

We also used circuit theory implemented the program CS (McRae et al. 2008) to calculate environmental resistance between populations. This approach integrates habitat suitability across a broader swath of area between 2 populations, rather than just focusing on the single path of least resistance. Therefore, we used the SDM as a univariate measure of resistance and calculated costs using 4 values: 1× (values the same as the SDM), 10×, and 100×.

We then used ArcMap 9.3 (ESRI, Redlands, CA) to obtain Euclidean geographic distance between sites ignoring the quality of habitat between populations. Finally, we used ArcMap 9.3, the Pacific North West River Reach database (PNWRF3, StreamNet) and the ArcMap toolbox FLoWS v1 (Theobald et al. 2006) to calculate aquatic distance between sites, as the shortest distance between a pair of sites that involved only travel along water sources (rivers, streams, lakes). Where there was no water-based connection between population pairs, those populations were considered unconnected and N/A values were used for analyses (following Spear and Storfer 2008).

We then performed multiple regression analyses in R between allelic richness, edge distance, elevation, and latitude to accommodate associations among these variables. We also performed multiple regression analyses of the relationships between linearized F_{ST} and the distance metrics (Euclidian by LCP and Euclidian by CS distances). In order to assess the effect of scale on landscape genetic inference, we ran each analysis on 3 data sets. To assess range-wide patterns, we used all the data (called range wide). To assess regional, intermediate-scale patterns, we analyzed northern and southern populations separately, with the criterion of congruent genetic breaks observed in mtDNA (Nielson et al. 2006) and our data at $K = 2$ (see below; called Regional). Finally, to assess local small-scale phenomena, we analyzed just the densely sampled populations from the North Fork of the Clearwater, St. Joe, and St. Maries vicinities (called Local).

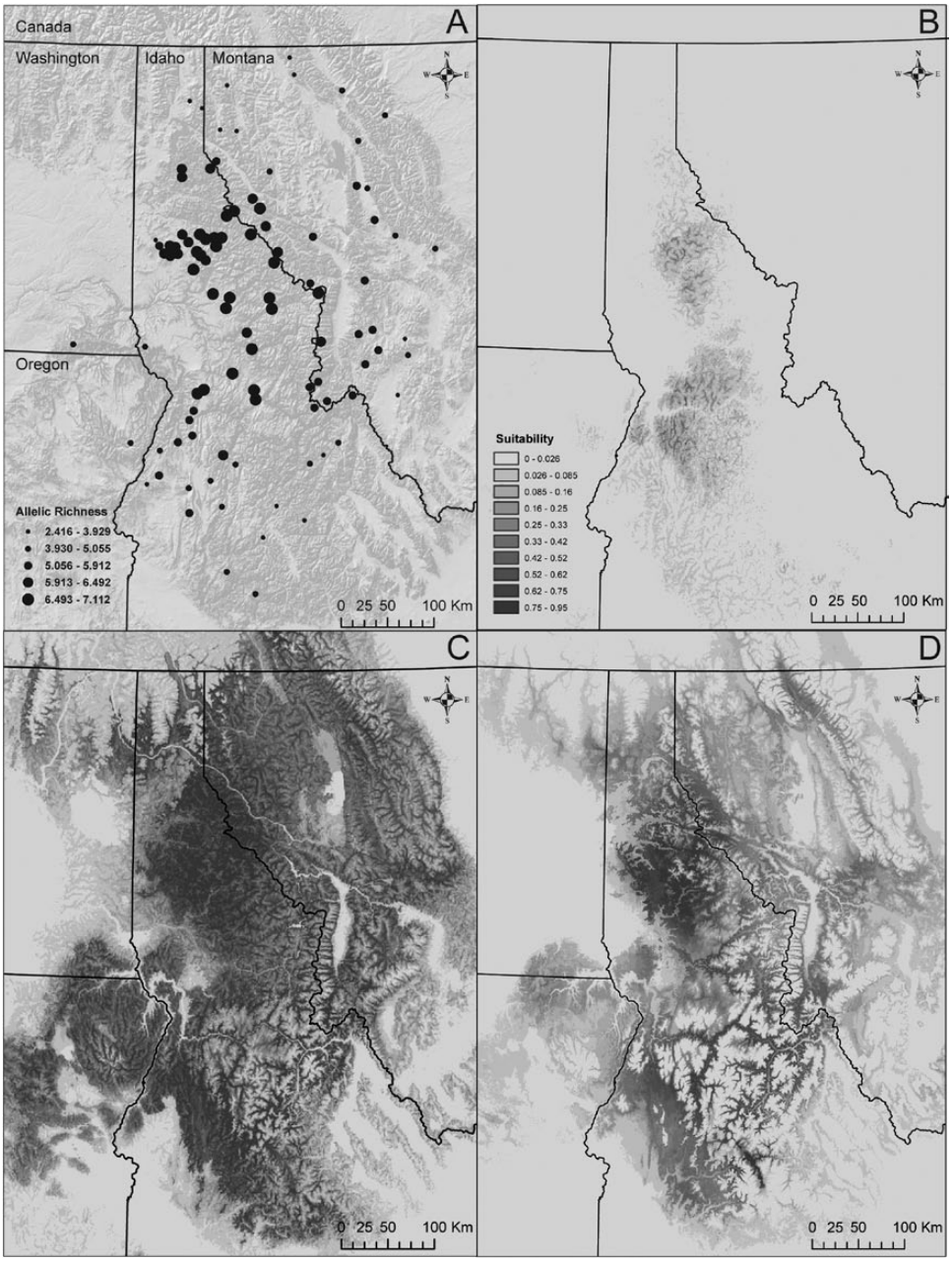

Figure 2. Projected suitable habitat for *A. montanus* derived from our SDM. (**A**) The pattern of allelic richness observed at our 99 sampling locations. (**B**) Suitabilities inferred by projecting our SDM onto climate reconstructions for the last glacial maximum. (**C**) Distribution of suitabilities for current habitat. (**D**) Suitabilities inferred by projecting or SDM onto climate projections for 2070 under a worst-case CO_2-emissions scenario.

Results

SDM

The SDM approach provided a high area under the curve value (0.9668), indicating a very good model fit. The variables most strongly contributing to the model were coniferous cover (a preference for higher covers; 30.41%), flow accumulation (a preference for headwaters; 17.28%), and maximum temperature during the warmest month (12.98%). The SDMs estimated from the reduced data (7 uncorrelated climatic variables) are nearly identical (Supplementary Figure S1 online; correlation coefficients of suitabilities scores per pixel calculated from full versus reduced data is 0.92, 0.96, and 0.80 for current, future, and past SDMs, respectively); therefore we focus on the SDM from the full data. Our SDM for current conditions (Figure 2C) is similar to the *A. montanus* portion of the projections produced by Carstens and Richards (2007), but has higher resolution. It broadly captures the species range of *A. montanus*, however it predicts areas of suitable habitat in some locations where *A. montanus* is known to be absent, notably in central Oregon. On a finer scale, poor habitat is often associated with deep canyons, such as the Hells and Salmon River Canyons, and the high resolution of our SDM renders those canyons visible in our projections (Figure 2C).

Our paleodistribution reconstruction based on climatic variables (Figure 2B) suggests that 2 areas of high climatic suitability (i.e., primary refugia) existed for *A. montanus* during the LGM, both contained

within the current range of the species. However, the location of one of the refugial areas differed from that inferred by Carstens and Richards (2007). We infer the southern refugium to have been higher in the Salmon River drainage than was inferred by Carstens and Richards (2007), which was estimated to be outside the current distribution.

Under current CO_2 emission rates, our model predicts that there will be a severe reduction in suitable habitat for *A. montanus* (Figure 2D). This is particularly true for the southern portion of the current distribution; tailed frog populations in this region will likely be extremely fragmented, and this inference is robust to different dispersal rates and suitability thresholds (not shown). Our predictions are not simply that suitable habitat will be shifted northward; indeed there is very little predicted latitudinal shift. Instead, suitable habitat is predicted to become restricted within the current species range, with a (perhaps) counterintuitive concentration to lower elevations (Figure 2D; Supplementary Figure S2 online).

Genetic Diversity and Structure

Multilocus genotypes for 2098 individuals are available on Dryad and in the supplement (Supplementary Table S2 online). The number of alleles per locus ranged from 1 to 15 across all samples. An edge effect was apparent in both allelic richness (Figure 2A) and, to a lesser extent, heterozygosity (not shown); populations from the center of the distribution had higher diversity than populations from the edges of the range, whereas many edge populations exhibited very low allelic richness (r^2 = 0.75; Table 2, Supplementary Figure S2 online). Overall levels of allelic richness and heterozygosity were moderate-to-high but lower than those observed in studies of *A. truei* (Spear and Storfer 2008, 2010). Pairwise estimates of F_{ST} suggest low to moderate genetic differentiation between sites (deposited in Dryad).

In the STRUCTURE analyses, likelihood scores, $P(D|K)$, increased with increasing K until K = 10, with a plateau across higher values (Figure 3A). The largest increase in $P(D|K)$ was associated with increasing from a single cluster to K = 2 (Figure 3A). At K = 2, individuals were clustered into northern versus southern populations (red vs. yellow in Figure 3B,C) in a manner very similar to the north/south phylogeographic split in mtDNA haplotypes (Figure 1, inset; Nielson et al. 2006). This is particularly true for analyses using a recessive inheritance model (Figure 3C). Analyses run assuming codominant inheritance and K = 2 (Figure 3B) tended to cluster 5 northern samples with the southern cluster; we view this as spurious and the north-south clustering (Figure 3C) is more plausible (see Discussion).

At K = 3 (Supplementary Figure S4A online), the individuals from the northern cluster were split into northeastern (purple) and northwestern (yellow), at K = 4 (Supplementary Figure S4B online), individuals from the purple northeastern cluster were further split into a north-northeastern cluster (light blue) and an east-central cluster (purple). At K = 5 (Supplementary Figure S4C online), the red southern cluster split into southwestern (green) and southeastern clusters (red), and at K = 6 (Figure 3D), the light blue north-northeastern cluster was split into clusters found in the Cabinet and Purcell Mountains (orange) versus those collected east of the Kootenay River (light blue). At K=7 (Supplementary Figure S4D online), individuals from the yellow north-northwestern group were split into clusters from the Palouse Range (dark blue) versus Clearwater/Bitterroot Ranges (yellow). Increasing K further simply resulted in splitting individuals into peripheral populations (not shown).

Table 2. Correlations (above diagonal) and partial correlations (below diagonal) between geographic variables

	Allelic richness	Latitude	Elevation	Edge distance
Allelic richness	0	0.0946	−0.4302	0.7509
Latitude	−0.0936	0	−0.5804	0.0779
Elevation	−0.2254	−0.6044	0	−0.4014
Edge Distance	0.6855	−0.0866	−0.1567	0

Geography and Genetics

At the range-wide scale, Euclidean distance between populations (i.e., isolation-by-distance; IBD) explained much of the variance in pairwise differentiation between populations (as measured by linearized F_{ST}; Figure 4A). Addition of the SDM information to the model through LCP and CS distances (visualized in Supplementary Figure S5 online) increased its explanatory power slightly relative to Euclidean distance (Figure 4B,C). Additionally, for both LCP and CS distances, the distance measures obtained using the lower costs (1× or 10×) explained more of the variation than those obtained using higher (100× or 1000×) costs (Supplementary Table S3 online). Aquatic distance explained very little of the variation in genetic distances between populations (Figure 4D). Furthermore, our multivariate analysis of geographic variables showed a positive correlation between distance from the nearest range edge and allelic richness, as well as a negative correlation between elevation and latitude (Table 2). Multivariate analysis of the range-wide distance metrics illustrated that that CS and LCP distances explain a slightly higher portion of variation in linearized F_{ST} values after the effect of geographic distance are removed (r^2 = 0.345; Table 3).

In analyses at the intermediate scale, in the northern populations, the SDM-based distances explained more of the variation in linearized F_{ST} than did simple geographic distance (r^2 = 0.16 for the IBD, and r^2 = 0.28 for CS distance; Figure 5A). This was not the case for the analyses of the southern populations, where simple geographic distance and CS distance have similar r^2 values (0.24 for geographic distance vs. 0.26 for the CS distance). Thus, at least for the northern populations, patterns of genetic variation are more strongly associated with climatic and environmental variables than at the range-wide scale, and this is particularly true once the effect of geographic distance in removed; partial correlation coefficients for CS distance are 0.231 and 0.422 for southern and northern populations, respectively (Table 3).

This effect is seen more strongly at the local scale for the densely sampled populations in the Clearwater Drainage (the most densely clustered yellow populations in Figure 3D). At this local scale, geographic distances explain little variation (r^2 = 0.115) whereas the CS distance explains substantially more (r^2 = 0.32; Figure 5C). In this analysis, a single population appears to be an outlier with relatively higher F_{ST} values than comparisons involving the other populations in the local sample (probably due to its extremely low allelic diversity). The correlation between geographic distance and differentiation is much higher at the local level once this differentiated population is removed (r^2 = 0.487). Thus, analyses at this local scale reveal patterns that are not detectable in analyses at larger spatial scales, especially regarding identification of populations that deviate from IBD.

Discussion

Genetic structure and patterns of gene flow are often characterized as being either governed by an IBD model or by being mediated by habitat features (i.e., Isolation-by-Environment, or IBE—e.g.,

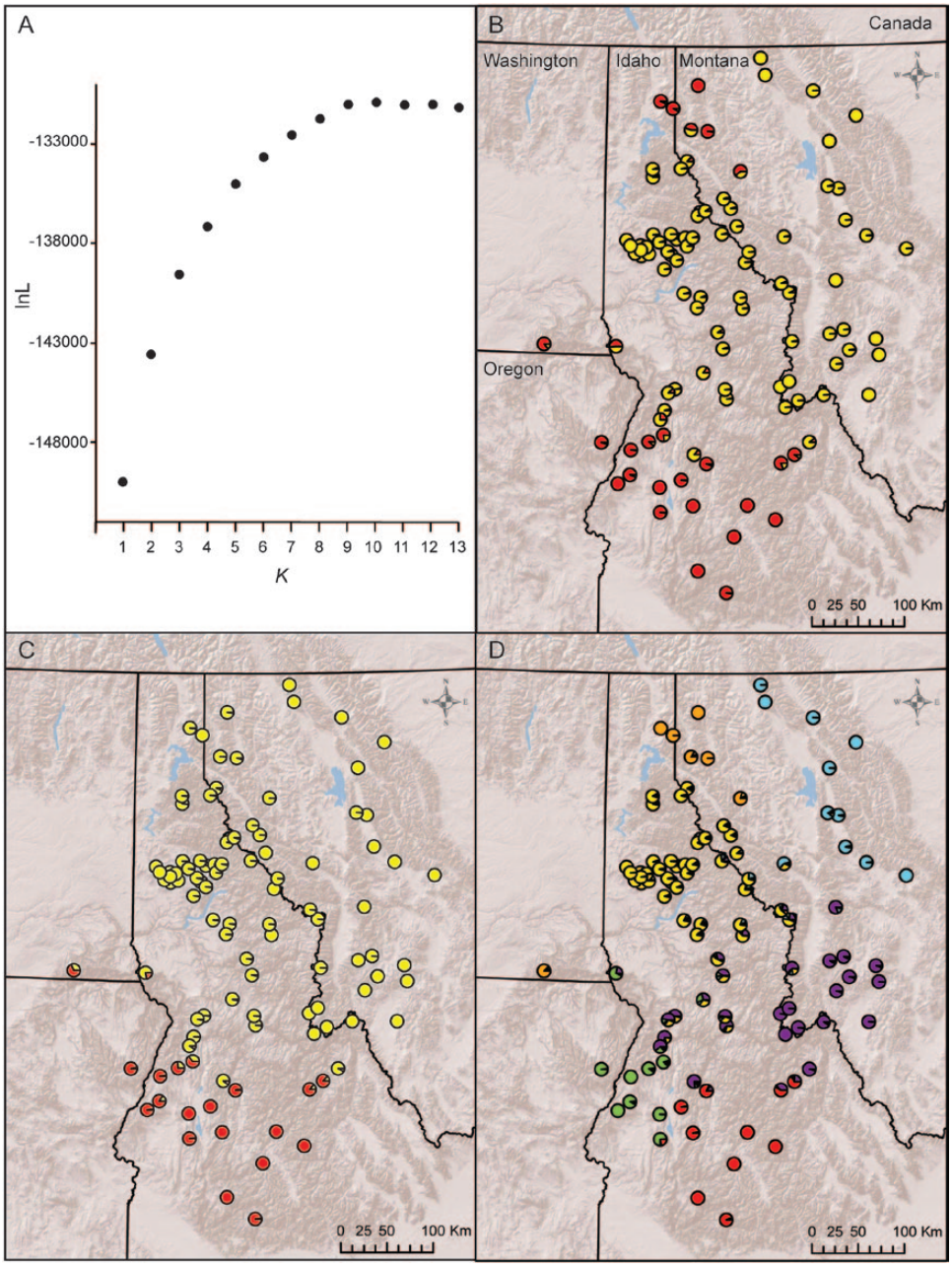

Figure 3. (**A**) Average log likelihood for each value of *K* used for structure. (**B**) Structure results for *K* = 2 based on microsatellite data under a codominant mode of inheritance, or (**C**) assuming recessive inheritance, both of which show Northern (yellow) and Southern (red) clusters. (**D**) Structure results for *K* = 6 based on microsatellite data assuming codominant inheritance. In B, C, and D, pie-chart colors indicate genetic clusters.

Wang and Summers 2010; Sexton et al. 2014; c.f. Shafer and Wolf 2013). Our method of assessing the influence of environment on patterns of genetic variation results in colinearity between habitat-aware distances (LCP and CS distances) and habitat-ignorant distance (geographic, or Euclidian, distances) that can complicate interpretation (Graham 2003). We have addressed this indirectly by using different weightings of intervening habitat suitabilities (1×, 10×, 100×, and 1000×) in calculating the CS and LCP distances. At the high weightings, the correlations between raw distances and the habitat-aware distances become small, diminishing the effect of colinearity for the 100× and 1000× distances (Supplementary Table S3 online).

Nevertheless, even in a homogeneous environment, geographic distances greater than dispersal distances represent barriers to gene flow. Both LCP and CS distances integrate geographic distance with obstacles to dispersal across intervening areas of poor habitat. Thus, such composite measures (that incorporate geographic distance with measures of habitat quality) are directly relevant to population connectivity.

In addition, the geographic structure of genetic variations across a species' range is a multiscale phenomenon, both temporally and spatially. Thus, understanding the geographic structure of genetic variation requires the inference of interactions between historical and contemporary processes and analyses at multiple spatial scales.

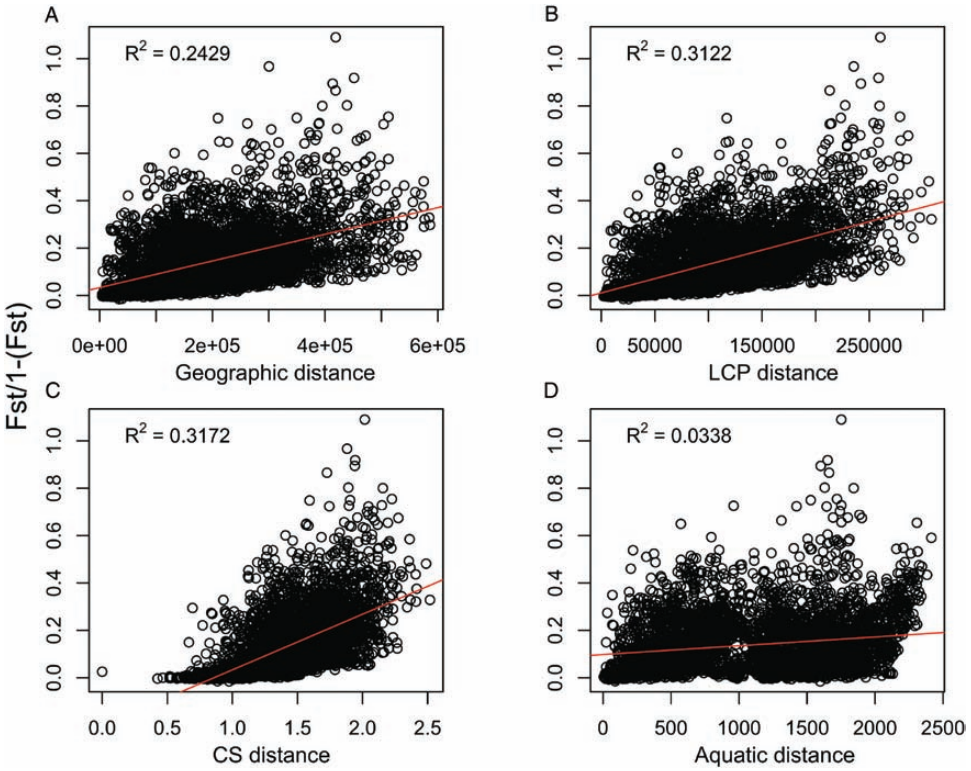

Figure 4. Range-wide linear regression of linearized F_{ST} and (**A**) Euclidean geographic distance calculated in ArcGIS, (**B**) least cost path (LCP) distance calculated from the SDM at cost of 1×, (**C**) lowest resistance (CS) distance calculated from the SDM using CircuitScape at cost of 1×, and (**D**) Aquatic distance calculated in ArcGIS. All correlations are significant ($P < 0.001$).

Table 3. Multiple regressions of Euclidian distances and CS (top) or LCP (bottom) distances on genetic differences for the entire range, only northern, and only southern populations

Rangewide	*Fst*	Euclidean distance	CS1× distance
Fst	0	0.2429	0.3178
Euclidean distance	0.0016	0	0.7530
CS1× distance	0.3479	0.8130	0
North			
Fst	0	0.1604	0.2789
Euclidean distance	−0.1226	0	0.7395
CS1× distance	0.4218	0.8256	0
South			
Fst	0	0.2370	0.2641
Euclidean distance	0.1131	0	0.7149
CS1× distance	0.2312	0.7920	0

Rangewide	*Fst*	Euclidean distance	LCP1× distance
Fst	0	0.2429	0.3122
Euclidean distance	−0.0960	0	0.8714
LCP1× distance	0.3057	0.8881	0
North			
Fst	0	0.1604	0.2660
Euclidean distance	−0.2270	0	0.8577
LCP1× distance	0.4276	0.9124	0
South			
Fst	0	0.2370	0.2738
Euclidean distance	0.0400	0	0.8191
LCP1× distance	0.2264	0.8723	0

Scale-specific patterns can be obscured in analyses of either smaller or larger geographic scope.

Genetic Diversity and Structure: Broadest Scale

At the broadest scale, allelic richness and heterozygosity are highest in populations at the center of the species' range and decline in populations sampled closer to the range edge, a pattern expected under equilibrium conditions. We hypothesized that our combination of molecular and SDM approaches would allow us to identify the location of ancient refugia for the species. Indeed, *A. montanus* exhibits a hierarchical pattern of geographic structure (Figure 3), and at the lowest level ($K = 2$), the genetic clusters correspond almost precisely (Figure 3C) with the deep phylogeographic split in mtDNA identified by Nielson et al. (2006). Furthermore, this partitioning of populations is also the deepest split in our tree-based analysis of our microsatellite data (Supplementary Figure S6 online). We interpret congruence of this result in the microsatellite data (Figure 3C) with mtDNA phylogeography (Nielson et al. 2006; illustrated here in the inset of Figure 1) and the location of refugia paleodistribution reconstructions (Figures 2B; Supplementary Figure S1A online) as strong support for a dual-refugium hypothesis. With $K = 6–9$, clusters of multilocus genotypes roughly correlate with mountain ranges. In addition, as we increased the number of clusters in the STRUCTURE analyses from 1 to 9, new clusters of populations were split from clusters identified by a previous run (i.e., a run with one fewer cluster).

The northern-most populations sampled occur at localities that were glaciated during the LGM; these populations therefore are necessarily the result of expansion from the Clearwater refugium.

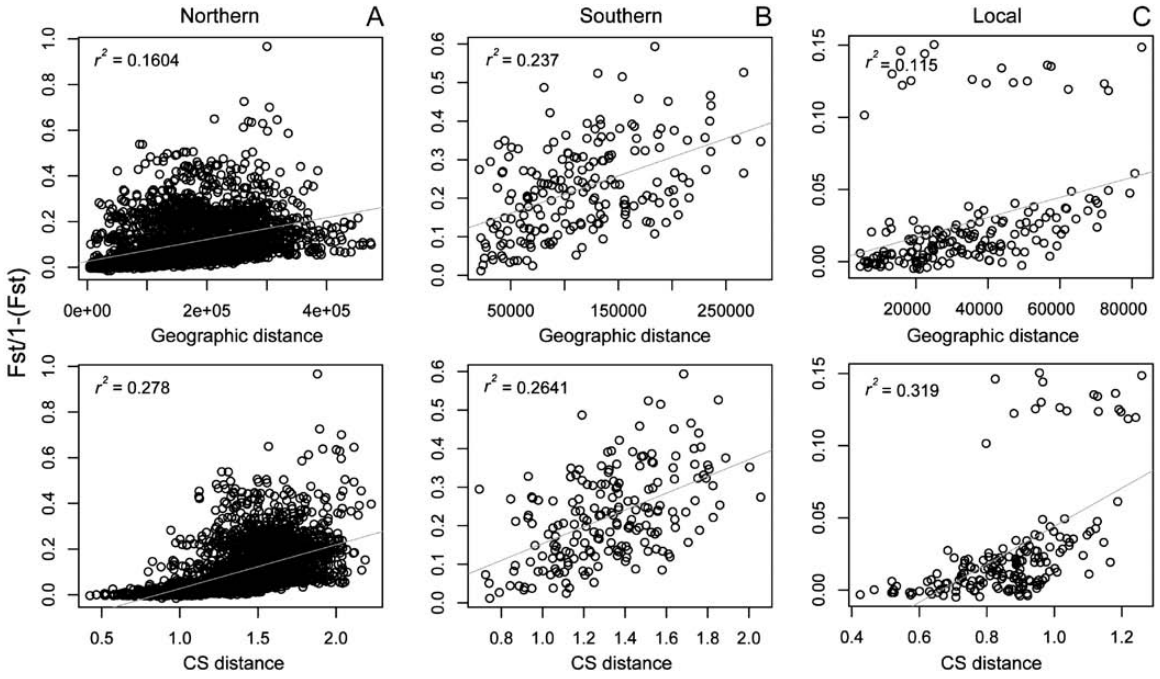

Figure 5. Linear regressions of linearized F_{ST} and geographic or CS distance for (**A**) northern, (**B**) southern, and (**C**) local samples.

Counter-intuitively, however, the route of expansion appears to not have been directly northward from the adjacent Clearwater Drainage, but based on the NJ tree (Supplementary Figure S6 online), appears to have occurred in a counter-clockwise fashion around the Missoula Basin (which is defined by the extent of proglacial Lake Missoula, and from which there are no current records of *A. montanus*). Thus, these populations in the Cabinet and Purcell ranges have become established by relatively long-range dispersal after glacial retreat. This explains the dramatic reduction in allelic diversity exhibited by these populations (Figure 2A), which in turn may be causing them to cluster spuriously with the southern populations in STRUCTURE analyses assuming codominant inheritance (Figure 3B). Populations with low allelic richness due (as we hypothesize) to the serial range expansions will have a stochastically reduced pool of genotypes. Our intuition suggests that applying a recessive alleles model (admittedly incorrectly) may reduce the sampling error in estimating coancestry, and the average coancestries of the Purcell and Cabinet mountains populations estimated this way (Figure 3C) make far more geographic sense than those estimated with the (appropriate) codominance model.

As a result of our extensive sampling within the species range, we were able to detect associations of small effect between genetic distance and climatic variables. We had, however, hypothesized that habitat variables play a role in tail frog dispersal, and that we would be able to identify correlations between habitat variables and genetic diversity. We chose therefore to focus on the magnitude of the association between variables. Our regressions and multivariate analysis indicate that inclusion of SDM-based distance measures (CS and LCP distances) into models explains more variation in the genetic distances between populations than models with geographic distance alone at the broadest spatial scale. The additional variation explained by inclusion of IBE is modest, but increases when the effect of geographic distance alone is removed in our partial regressions (Table 3). Additionally, for both CS and LCP distances, the maximum proportion of variation in genetic distance between populations was explained when both CS and LCP costs were lowest for the environmental variables (1×). Together, these findings all support the conclusion that environmental variables included in the SDM affect gene flow in *Ascaphus,* but that geographic distance between sites is perhaps the most influential determinant of differentiation between populations. One important caveat to our analyses is that we accounted for multiple environmental variables by using a single metric, the SDM, to calculate resistance for our analyses. This prevents us from being able to determine which environmental variables may be most important in mediating gene flow and may disguise the effect of environment if variables have opposing influences.

As was seen by Spear and Storfer (2008) on a more local scale, riparian distances seem to be wholly unrelated to population differentiation (Figure 4D); although the regression is significant (due to our very large sample size), the effect size is negligible ($r^2 < 0.034$). This was true across all 3 scales examined, and indicates that dispersal and gene flow are concentrated to postlarval life stages; dispersal by tadpoles would be expected to lead to an association between riparian and genetic distances. More specifically, given that mark-recapture data suggest strong philopatry of breeding adults (Daugherty and Sheldon 1982), dispersal in tailed frogs appears to be restricted to postmetamorphic subadults and to be occurring over land rather than along water-based corridors (i.e., saddle hopping). Although there is no direct evidence for this, it is more plausible to us that dispersal of small subadults has gone undetected than that dispersal of large adult frogs has.

Accounting for Deep History: Regional Scale

Our third hypothesis predicted that landscape variables may strongly influence gene flow at local scales, and are less important at the total range level. Thus, while geographic distance explains a large portion of the variation in genetic distance between populations at broad spatial scales (i.e., in the range-wide analyses), environmental factors may be more important at smaller spatial scales. To assess this for *Ascaphus*, we conducted separate analyses on northern and

southern groups, as defined by the deep phylogeographic divergences in the mitochondrial data (Nielson et al. 2001, 2006) and the congruent results of the STRUCTURE analyses of the current microsatellite data with $K = 2$ (Figure 3C). The results of our analyses at this intermediate spatial scale, with the deepest historical divergence factored out, are largely consistent with those for the species range as a whole; geographic distance still explains much of the variation in genetic distance between populations. However the SDM-based distance metrics explain more variation in genetic divergence than they do in the range-wide analysis (Figure 5A,B). This is particularly true in the northern cluster, and less so in the southern cluster, perhaps due to the more patchy habitat in the southern portion of the range (Figure 2C). This difference in effect of IBE between northern and southern regions can only be detected at the appropriate (regional) scale; it may be due to the greater current fragmentation of the southern portion of the range (especially evident in Supplementary Figure S1B online), which also likely has led to generally low allelic richness in these populations (Figure 2A).

Landscape Genetics at Local Scale

Further supporting our third hypothesis, even more variation in *F*st is explained in the analyses at the local scale (Figure 5C). First, there is a marked increase in the r^2 value when the SDM-based distance is included, compared to simple geographic distance, suggesting that deviation from IBE is more easily detected at the smallest scale. Second, the impact of a single aberrant population is detectable only at this local scale; all the highest *F*ST values at this scale involve pairwise comparisons of a single population collected at locality 85 (Latah Creek) with the rest in the cluster. There are several unique features of this population; it is the only locality we examined from a drainage that flows directly into the Spokane River, it is separated from the rest of the nearby populations in the cluster by State Highway 6, and it exhibits extremely low genetic diversity (Figure 2A). Application of our SDM-based (CS) distance shifts the comparisons involving this population on the *x*-axis to the right compared to strictly geographic distance (Figure 5D), suggesting that the low allelic diversity in this population (and consequently high *F*STS) may be mediated by ecological differences. Partial correlations between CS distances and genetic divergence are the highest we observed in this study, regardless of weather the Latah Creek population in included or not ($r^2 = 0.642$), suggesting a strong effect of habitat on connectivity at the smallest scale.

Modern and Paleodistribution Models of *A. montanus*

Our SDM for *A. montanus* (Figures 2C, Supplementary Figure S1B online) is broadly consistent with that constructed by Carstens and Richards (2007), although we find a wider range of suitable habitat, especially within the Clearwater River drainage area, and portions of northwestern Montana. This difference is likely the result of our expanded sampling of *A. montanus*, and thus more complete presence/absence data used to construct the SDM.

The hindcast of our SDM using paleoclimatic reconstructions (Figures 2B, Supplementary Figure S1A online) provides strong support that *A. montanus* has had a very long history in the INW; further, these projections support the persistence of 2 distinct refugia (1 in the Clearwater Drainage and 1 in the Salmon River Drainage) during Pleistocene glacial maxima. These refugia are consistent both with previous SDM-based reconstruction (Carstens and Richards 2007) and the deep divergence the mtDNA sequence data identified by Nielson et al. (2006; Figure 1, inset). Furthermore, the signature of these dual refugia is seen in our microsatellite data, both by the results of our STRUCTURE analysis at the level of $K = 2$ (Figure 3C) and by our NJ tree (Supplementary Figure S6 online). In contrast to the analysis by Carstens and Richards (2007), we do not find evidence for refugial habitat present in the Blue Mountains of Oregon, perhaps due to a combination of differences in the spatial scale of the environmental variables used in construction of our SDM (which allowed for a finer scale analysis than was possible for Carstens and Richards 2007), and the updated presence/absence data we used to construct our SDM that arose from our extended sampling efforts. Additionally, whereas Carstens and Richards (2007) found evidence for this third refugium in their paleoclimate model, DNA evidence does not support the hypothesis of 3 refugia; thus the 2-refugia model is consistently supported by available evidence.

Finally, projections into predictions of future climatic conditions (Figure 2D) indicate that the range of the species is expected to reduce drastically and, more importantly, become highly fragmented. Because environmental variables have been shown to influence species connectivity at local scales, this is expected to affect the genetic diversity of the species through reduced gene flow. Although the region with the highest allelic richness (Figure 2A) is expected to harbor high ecological suitability (Figure 2C), the southern range of the species will become extremely fragmented and reduced. This will likely drastically affect the survival and persistence of the southern genetic clusters (Figure 3B–D).

Conclusion

The structure of genetic diversity within *A. montanus* is the result of processes acting across multiple geographic and temporal scales (Table 1). Here, we show a strong geographic structure across the species range, with the deep phylogeographic split between northern and southern range portions detected in mtDNA (Nielson et al. 2006) representing the deepest divergence in microsatellite loci as well. This is likely the result of the species restriction to 2 Pleistocene refugia, the expansion from which has produced a contact zone in the South Fork of the Salmon River detectable in both mtDNA and our data. Analyses at decreasing spatial scales indicate that the smaller the scale, the more additional variance in *F*ST is explained by our SDM-based CS and LCP distances, and only at the smallest scale are outlier populations identified. This suggests that meta-analyses that attempt to identify discern the relative importance of IBD versus IBE (Sexton et al. 2014) should include multiple geographic spatial scales to account for scale-specific phenomena.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

Idaho State Board of Education through a grant to the Center for Research on Invasive Species and Small Populations (CRISSP). Analyses were run on the Computational Resources Core facility supported by the Institute for Bioinformatics and Evolutionary Studies (IBEST) and funded by National Institutes of Health (NCRR 1P20RR016454-01; NCRR 1P20RR016448-01) and National Science Foundation (EPS-809935).

Acknowledgments

We would like to acknowledge the enormous amount of field and lab work done for this project by C. Drummond. Multilocus genotypes and collection localities for each individual are available online supplement and have been deposited in Dryad. Brad Shaffer and 3 anonymous reviewers provided thorough and extremely helpful comments that helped us improve this article tremendously.

References

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS. 2007. Automated binning of microsatellite alleles: problems and solutions. *Mol Ecol Notes*. 7:10–14.

Anderson CD, Epperson BK, Fortin MJ, Holderegger R, James PM, Rosenberg MS, Scribner KT, Spear S. 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol Ecol*. 19:3565–3575.

Baker CS. 2013. Journal of heredity adopts joint data archiving policy. *J Hered*. 104:1.

Balkenhol N, Gugerli F, Cushman SA, Waits LP, Coulon A, Arntzen JW, Wagner HH. 2009. Identifying future research needs in landscape genetics: where to from here? *Landsc Ecol*. 4:455–463.

Bjork C. 2009. Distribution patterns of disjunct and endemic vascular plants in the interior wetbelt of northwestern North America. *Botany*. 88:409–428.

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

Carstens BC, Brennan RS, Chua V, Duffie CV, Harvey MG, Koch RA, McMahan CD, Nelsen BJ, Newman CE, Satler JD, *et al*. 2013. Model selection as a tool for phylohgeographic inference: An example from the willow *Salix melanopsis*. *Mol Ecol*. 22:4014-4028.

Carstens BC, Stevenson AL, Degenhardt JD, Sullivan J. 2004. Testing nested phylogenetic and phylogeographic hypotheses in the *Plethodon vandykei* species group. *Syst Biol*. 53:781–792.

Carstens BC, Brunsfeld SJ, Demboski JR, Good JM, Sullivan J. 2005. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. *Evolution*. 59:1639–1652.

Carstens BC, Richards CL. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*. 61:1439–1454.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol*. 24:621–631.

Cavalli-Sforza LL, Edwards AW. 1967. Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet*. 19(3 Pt 1):233–257.

Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CG, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, *et al*. 2006. The community climate system model version 3 (CCSM3). *J Climate*. 19:2122–2143.

Daugherty CH, Sheldon AL. 1982. Age-specific movement patterns of the frog *Ascaphus truei*. *Herpetologica*. 38:468–474.

Espíndola A, Pellissier L, Maiorano L, Hordijk W, Guisan A, Alvarez N. 2012. Predicting present and future intra-specific genetic structure through niche hindcasting across 24 millennia. *Ecol Letters* 15:649-657.

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

Ford LS, Cannatella DC. 1993. The major clades of frogs. *Herpetol Monogr*. 7:94–117.

Gavin DD. 2009. The coastal-disjunct mesic flora in the inland Pacific Northwest of USA and Canada: refugia, dispersal and disequilibrium. *Divers Distrib*. 15:972–982.

Graham MH. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology*. 84:2809–2815.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 25:1965–1978.

McRae BH, Dickson BG, Keitt TH, Shah VB. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*. 89:2712–2724.

Nielson M, Lohman K, Sullivan J. 2001. Phylogeography of the tailed frog (Ascaphus truei): implications for the biogeography of the Pacific Northwest. *Evolution*. 55:147–160.

Nielson M, Lohman K, Daugherty CH, Allendorf FW, Knudsen KL, Sullivan J. 2006. Allozyme and mitochondrial DNA variation in the tailed frog (Anura: *Ascaphus*): the influence of geography and gene flow. *Herpetologica*. 62:235–258.

Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB. 2011. *Ecological niches and geographic distributions*. Princeton (NJ): Princeton University Press.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecol Modell*. 190:231–259.

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

Pyron RA, Wiens JJ. 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Mol Phylogenet Evol*. 61:543–583.

Raymond M, Rousset F. 1995. genepop (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered*. 86: 248–249.

Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour*. 8:103–106.

Searcy CA, Shaffer HB. 2014. Field validation supports novel niche modeling strategies in a cryptic endangered amphibian. *Ecography*. 37:983–992.

Sexton JP, Hangartner SB, Hoffmann AA. 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*. 68:1–15.

Shafer AB, Wolf JB. 2013. Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecol Lett*. 16:940–950.

Sork VL, Waits L. 2010. Contributions of landscape genetics - approaches, insights, and future potential. *Mol Ecol*. 19:3489–3495.

Spear SF, Storfer A. 2008. Landscape genetic structure of coastal tailed frogs (*Ascaphus truei*) in protected vs. managed forests. *Mol Ecol*. 17:4642–4656.

Spear SF, Storfer A. 2010. Anthropogenic and natural disturbance lead to differing patterns of gene flow in the Rocky Mountain tailed frog, *Ascaphus montanus*. *Biol Conserv*. 14:778–786.

Theobald DM, Norman JB, Peterson EE, Ferraz SB, Wade A, Sherburne MR. 2006. Functional Linkage of Water Basins and Streams (FLoWS) v1 User's Manual: ArcGIS Tools for Network-based Analysis of Freshwater Ecosystems. Natural Resource Ecology Lab, Colorado State University, Fort Collins, CO.

Warren DL, Wright AN, Seifert SN, Shaffer HB. 2014. Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Divers Distrib*. 20:334–343.

Wang IJ, Summers K. 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Mol Ecol*. 19:447–458.

Zhang P, Liang D, Mao RL, Hillis DM, Wake DB, Cannatella DC. 2013. Efficient sequencing of Anuran mtDNAs and a mitogenomic exploration of the phylogeny and evolution of frogs. *Mol Biol Evol*. 30:1899–1915.

Metzger et al. Supplemental Figure S1-S6. Genetic Variation in Tailed Frogs JHered

Figure S1 – Comparison of SDMs calculated from the full set vs. reduced set of climatic variables. Top. Full vs. Reduced data sets for LGS SDM; Middle. Full vs. Reduced data sets for current SDM; Bottom. Full vs. Reduced data sets for 2070 SDM.

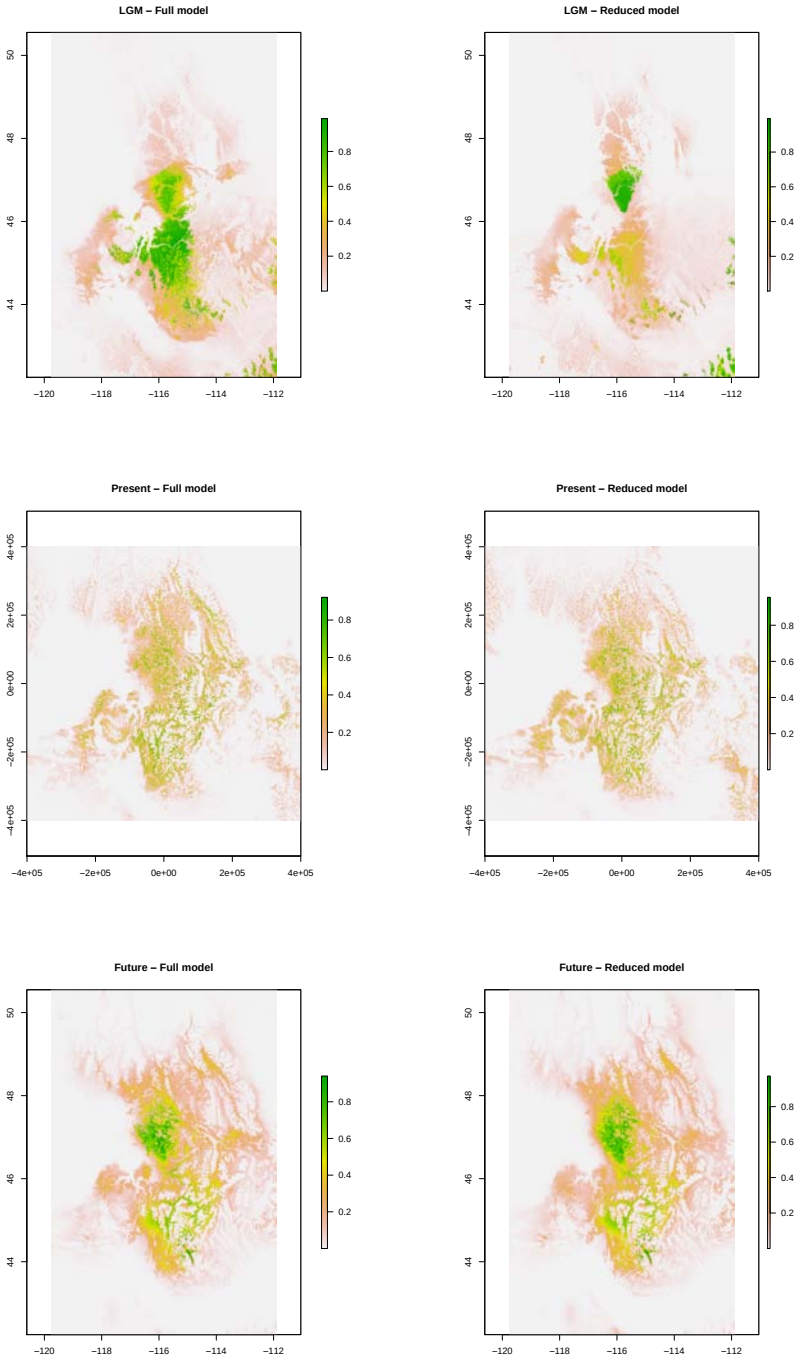

Figure S2 – Top. Distribution of suitabilities by latitude for (left) the Latest Glacial Maximum, (middle) the current conditions, and (right) the 2070 ENM forecast. The horizontal line is at the contact between northern and southern genetic clusters at *k*=2, as well as mtDNA clades.

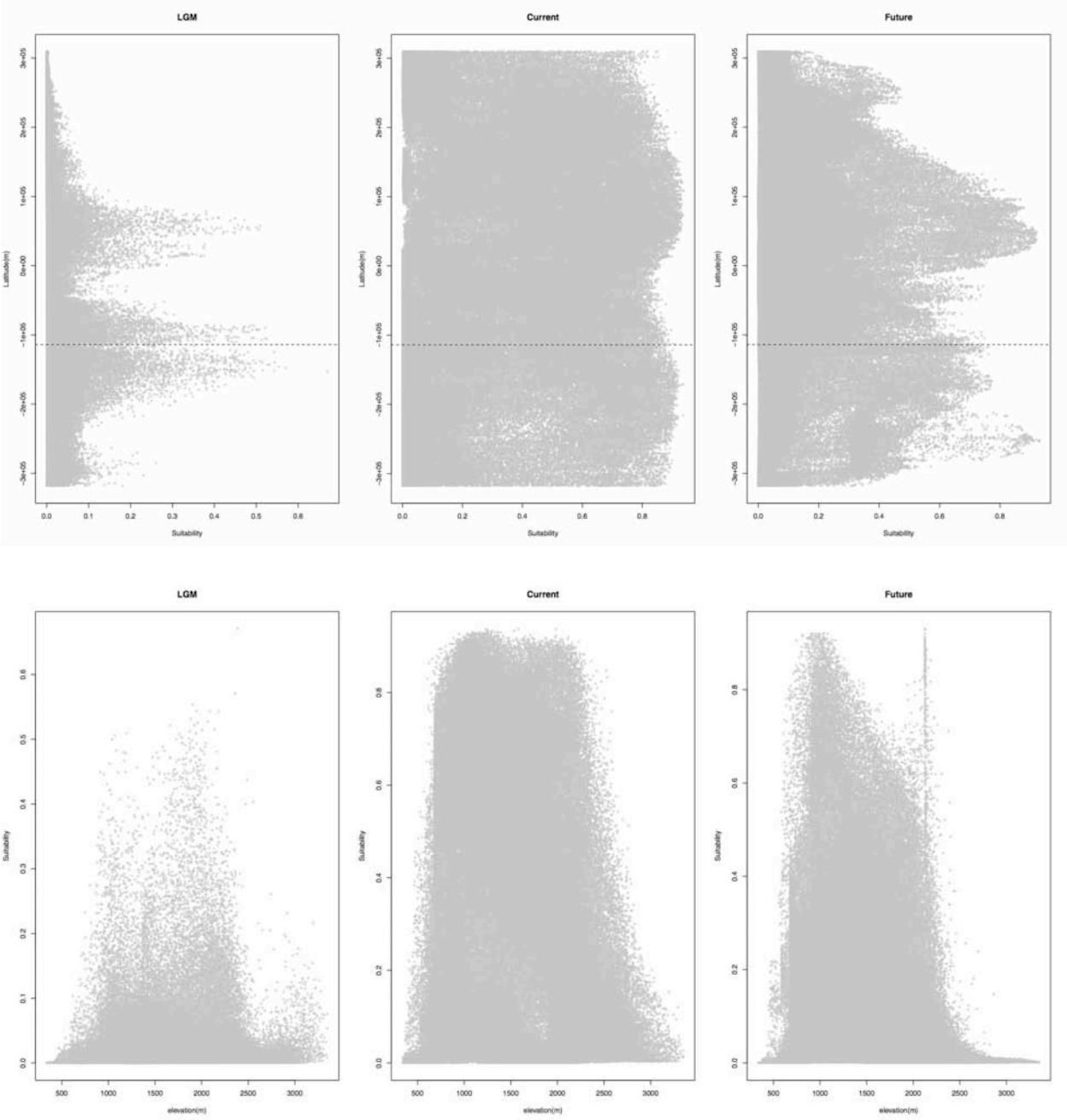

Figure S3 – Allelic richness versus A. Distance from edge of range, and B. Latitude.

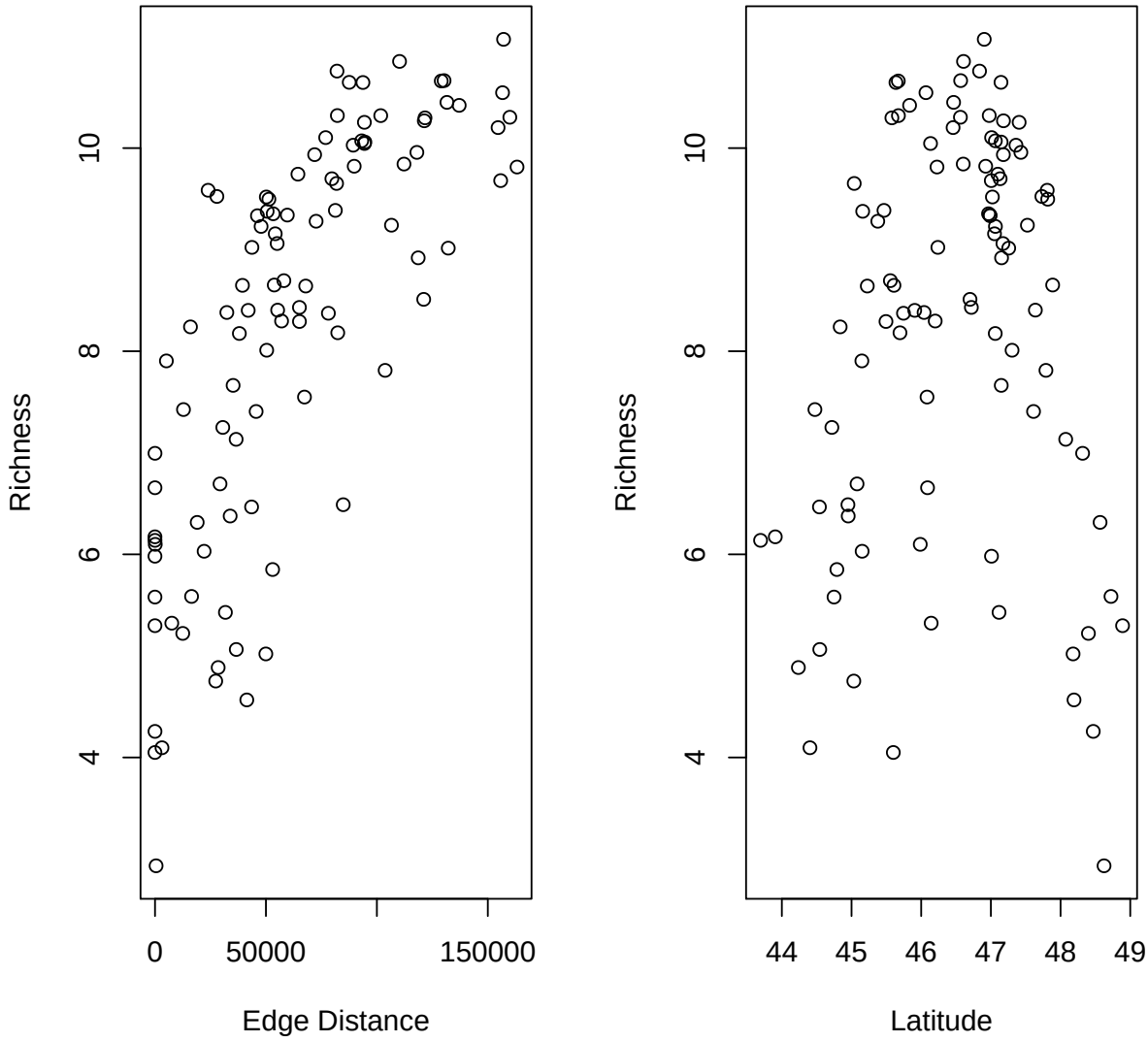

Figure S4 - Hierarchical pattern of clusters from Structure analyses with A. k = 3, B. k = 4, C. k = 5, and D. k = 7 under a co-dominant admixture model.

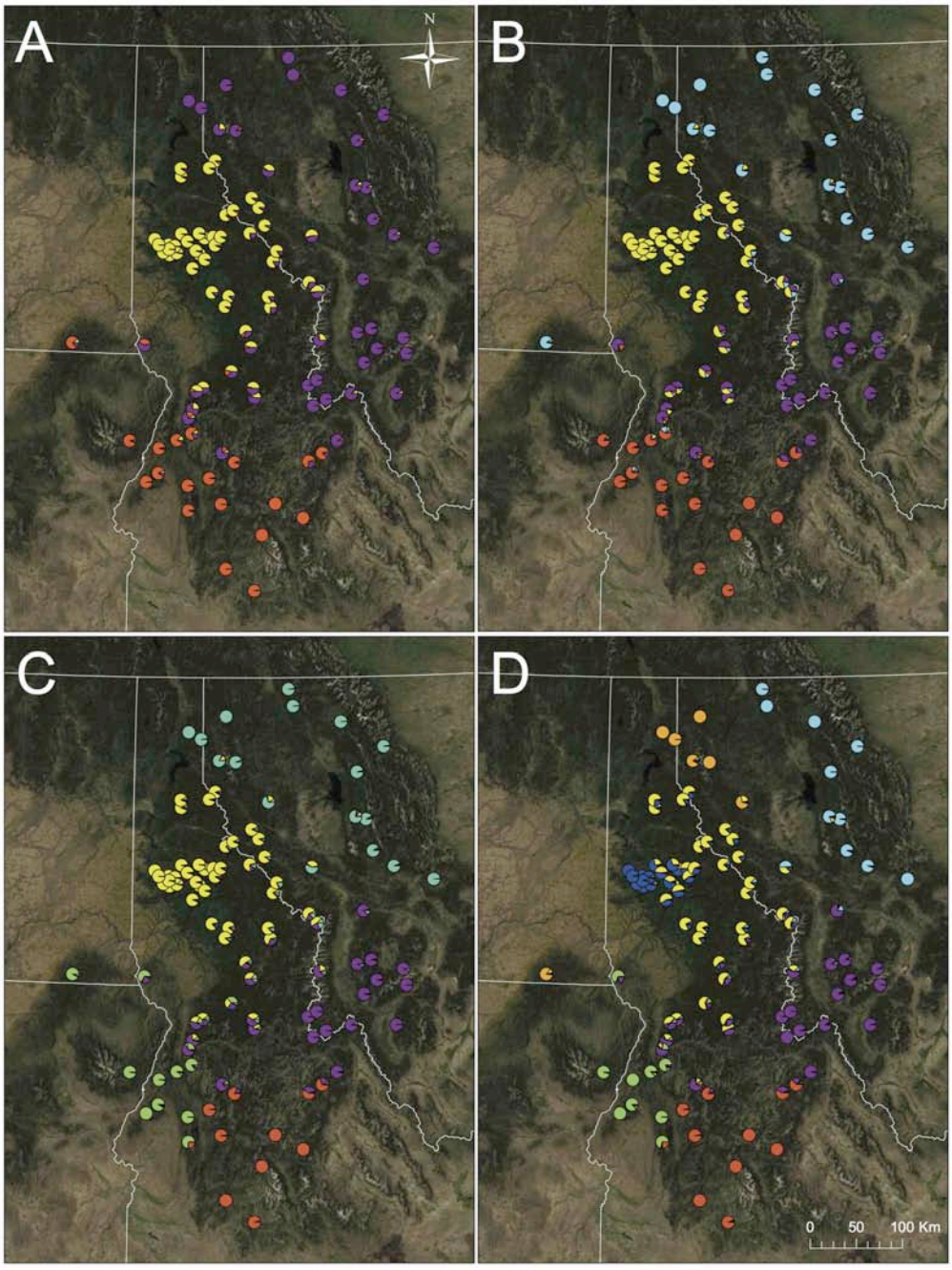

Figure S5 – Visualization of A. Least-Cost Paths, and B. CircuitScape resistance at 10X suitabilities.

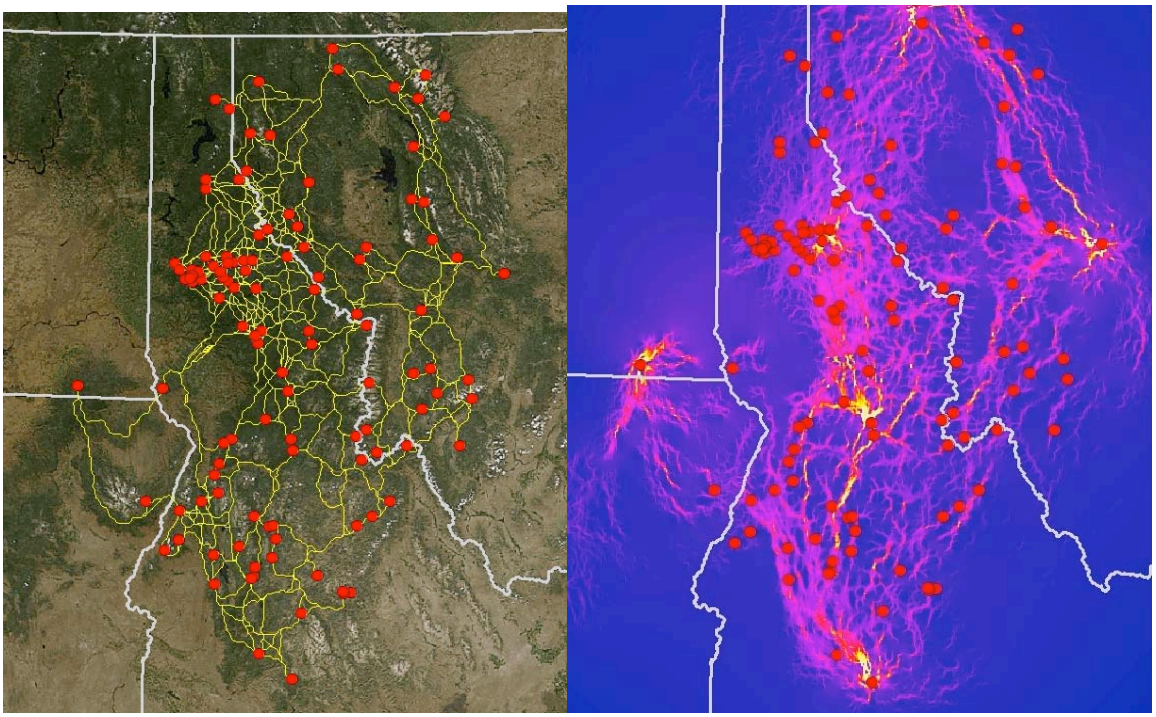

Figure S6. A, neighbor-joining tree calculated from cord distances. Colors correspond to the result of Structure analysis with *K*=10 (B), with major lineages reduced and mapped geographically.