Identification of evolutionary hotspots based on genetic data from multiple terrestrial and aquatic taxa and gap analysis of hotspots in protected lands encompassed by the South Atlantic Landscape Conservation Cooperative.

J. Robinson[1,3], M. Snider[2], J. Duke[2], and G.R. Moyer[1]

[1] U.S. Fish and Wildlife Service, Warm Springs Fish Technology Center, Conservation Genetics Lab, 5308 Spring Street, Warm Springs, GA 31830

[2] U.S. Fish and Wildlife Service, Regional GIS Center, Cookeville Ecological Field Office, 446 Neal Street Cookeville, TN 38501

[3] Current address: Department of Biology, City College of New York, 160 Convent Ave., New York, NY 10031

Corresponding author: Gregory R. Moyer, U.S. Fish and Wildlife Service, Warm Springs Fish Technology Center, Conservation Genetics Lab, 5308 Spring Street, Warm Springs, GA 31830

Email: Greg_Moyer@fws.gov, Phone: 706.655.3382, Fax: 706.655.9034

Abstract. The southeastern United States is a recognized hotspot of biodiversity for a variety of aquatic taxa, including fish, amphibians, and mollusks. Unfortunately, the great diversity of the area is accompanied by a large proportion of species at risk of extinction. Gap analysis was employed to assess the representation of evolutionary hotspots in protected lands where an evolutionary hotspot was defined as an area with high evolutionary potential and measured by atypical patterns of genetic divergence, genetic diversity, and to a lesser extent genetic similarity across multiple terrestrial or aquatic taxa. A survey of the primary literature produced 16 terrestrial and 14 aquatic genetic datasets for estimation of genetic divergence and diversity. Relative genetic diversity and divergence values for each terrestrial and aquatic dataset were used for interpolation of multispecies genetic surfaces and subsequent visualization using ArcGIS. The multispecies surfaces interpolated from relative divergences and diversity data identified numerous evolutionary hotspots for both terrestrial and aquatic taxa, many of which were afforded some current protection. For instance, 14% of the cells identified as hotspots of aquatic diversity were encompassed by currently protected areas. Additionally, 25% of the highest 1% of terrestrial diversity cells were afforded some level of protection. In contrast, areas of high and low divergence among species, and areas of high variance in diversity were poorly represented in the protected lands. Of particular interest were two areas that were consistently identified by several different measures as important from a conservation perspective. These included an area encompassing the panhandle of Florida and southern Georgia near the Apalachicola National Forest (displaying varying levels of genetic divergence and greater than average levels of genetic diversity) and a large portion of the coastal regions of North and South Carolina (displaying low genetic divergence and greater than average levels of genetic diversity). Our results show the utility of genetic datasets for identifying cross-species patterns of genetic

diversity and divergence (i.e., evolutionary hotspots) in aquatic and terrestrial environments for use in conservation design and delivery across the southeastern United States.

Introduction

The importance of genetic diversity from a conservation standpoint is recognized by both the International Union for Conservation and Nature (McNeely et al. 1990) and the United Nations Convention on Biological Diversity. The loss of genetic diversity is expected to limit the ability of a species to adapt to changing environmental conditions (Frankham 2005), and recent empirical work has shown that intraspecific genetic variation can have important effects on a variety of population and ecosystem-level processes (Hughes et al. 2008). For instance, diverse populations appear to be better able to adapt to novel environments (Frankham et al. 1999; Agashe 2009), more resistant to local disturbance (Hughes and Stachowicz 2004), and more productive (Crutsinger et al. 2006). Unfortunately, assessment and protection of genetic diversity has lagged behind that of other, more recognizable biodiversity components (Laikre 2010).

A growing number of studies have shown that diversity at the genetic level may correlate with species-level diversity for a variety of systems (e.g., Vellend and Geber 2005; Robinson et al. 2010; Blum et al. 2012). Additionally, some previous work has documented congruence between priority areas identified for different suites of species (Myers et al. 2000). These observations raise the possibility that information for one level of biodiversity (or taxonomic group) could be useful in conserving diversity across levels of organization, although in practice this may not always be the case (Moritz 2002; Forest et al. 2007). In addition to local genetic diversity, the level of genetic divergence among populations is also important to consider in conservation planning (Petit et al. 1998; Moritz 2002). Analyses of population structure can identify areas that are important for their unique genetic composition, as well as migration corridors that help to maintain demographic and genetic connectivity among disjunct populations across a landscape.

Priorities must be set for conservation, because the available funds are insufficient to address the global need (Myers et al. 2000). This necessitates the identification of biodiversity hotspots (i.e., biogeographic regions with a significant reservoir of biodiversity that is under threat from humans) where conservation funds are most efficiently deployed (Myers et al. 2000). Additionally, gap analysis provides an explicitly spatial approach to prioritization where information on, for example, vegetation type and species ranges are used to assess the representation of species diversity in protected lands (Scott et al. 1993; Kiester et al. 1996). This methodology allows the identification of diversity hotspots and the simultaneous assessment of the degree to which diversity is currently protected. Priority areas are then identified as areas of interest that fall outside of presently defined nature reserves.

The focus on species diversity in conservation planning persists today, despite the large number of population genetic datasets published over the last half century. Because of this body of work, existing data can be used to provide information on patterns of genetic diversity with little direct cost. Recently, gap analysis (i.e., a specific, stepwise method of assessing and mapping the level of biodiversity protection for a given area; Scott et al. 1993) has been combined with published genetic data in conservation assessments (Ji and Leberg 2002; Vandergast et al. 2008). Vandergast et al. (2008) combined genetic landscapes across twenty-one codistributed species in southern California to identify areas of high conservation value. In cases where phylogeographic patterns are concordant, priority areas may cluster across multiple taxa (both sampled and unsampled). The comparative approach adopted by Vandergast et al. (2008) can also help to highlight the influences of shared historical processes on multiple species in an ecosystem (Avise 2000).

The southeastern United States is a recognized hotspot of biodiversity for a variety of aquatic taxa, including fish (Warren et al. 2000), amphibians (Rissler and Smith 2010), and mollusks (Lydeard and Mayden 1995). Unfortunately, the great diversity of the area is accompanied by a large proportion of species at risk of extinction (Lydeard and Mayden 1995; Warren et al. 2000). Both species distributional data and intraspecific genetic data show consistent patterns of differentiation between eastern and western areas of the region in terrestrial (Hayes and Harrison 1992; Ellsworth et al. 1994), as well as aquatic and semi-aquatic species (Avise 2000). More recent work has identified shared phylogeographic breaks among species along the Appalachian Mountains and the Apalachicola-Chattahoochee-Flint River basin (Soltis et al. 2006; Rissler and Smith 2010). The high diversity, threatened status of the regional biota, and wealth of previous research (*e.g.,* Walker and Avise 1998; Soltis et al. 2006; Rissler and Smith 2010) make the southeastern United States an ideal area in which to assess shared patterns of genetic diversity and divergence among species.

Our study sought to use a combination of published genetic datasets and geographic information systems (GIS) to determine how well priority areas identified on the basis of genetic data across multiple species are protected by existing conservation plans that were developed without considering patterns of intraspecific genetic diversity. We focused on the region defined by the South Atlantic Landscape Conservation Cooperative (SALCC; an applied conservation science partnership among federal agencies, regional organizations, states, tribes, NGOs, universities and other entities), which spans the area from eastern Alabama and northern Florida to southern Virginia, from the Atlantic Ocean to the lower slopes of the Appalachian Mountains (Fig. 1). Our interest lies in areas of high genetic diversity and areas of atypical genetic divergence. Areas that show unexpectedly high divergence will help to identify barriers to

migration across the landscape and suture zones, while those that show high connectivity (or low divergence) may correspond to migration corridors (Vandergast et al. 2008). Given previous information on biogeographic boundaries in the southern U.S. (Avise 2000; Soltis et al. 2006), we hypothesized that, for terrestrial taxa, high divergence areas would be clustered along the southern and western boundaries of the study region. In contrast, we expected that the central domain of the SALCC would be characterized generally by high connectivity.

Methods

Genetic diversity and divergence estimation

We surveyed the primary literature for population genetic studies with sample sites located within the bounds of the SALCC study area (Fig. 1). We included studies that sampled at least four unique locations within the area and sequenced or genotyped at least three individuals per collection locality. The final dataset consisted of a total of thirty species (Table 1). For genetic diversity, we included datasets from sixteen terrestrial and fourteen aquatic species (Table 1). For genetic divergence, fifteen species were included in the terrestrial dataset and nine species in the aquatic dataset (Table 1). The datasets used in our analysis varied in terms of the genetic markers employed. We included studies that reported DNA sequences, microsatellite or allozyme allele frequencies, and RFLP genotypes (Table 1).

The statistics used to estimate genetic diversity and divergence varied by marker type (Table 1). For DNA sequence datasets available from GenBank (http://www.ncbi.nlm.nih.gov/genbank/), we calculated the average number of pairwise nucleotide differences (π; Nei and Li 1979) as our measure of genetic diversity and the absolute average divergence (D_{xy}; Nei 1987) as our measure of genetic distance, using DnaSP v5 (Librado and Rozas 2009). In cases where diversity and divergence estimates were taken directly from the

cited manuscript (i.e., Philipp et al. 1983, Donovan et al. 2000, Scott et al. 2009, Hasselman 2010) other statistics were employed (*e.g.*, net divergence, D_a, Nei and Li 1979, see Table 1). In other cases, haplotype diversity (H; Nei and Tajima 1981) was calculated from published data, given the number of unique haplotypes recovered and the number of sequences obtained. For microsatellite datasets, expected heterozygosity (H_e; gene diversity) and F_{ST} (or analogue R_{ST}) were used to measure diversity and divergence, respectively. For allozyme datasets, when allele frequencies were given in the text of the manuscript, H_e at each locus was calculated in an Excel (Microsoft Corp.) spreadsheet as

$$H_e = 1 - \sum p_i^2$$

where p_i was the frequency of the ith allele, and averaged across loci. Nei's standard genetic distances (Nei's D; Nei 1972) among population pairs were calculated using GeneDist (http://www.biology.ualberta.ca/jbrzusto/GeneDist.php). Finally, for restriction fragment length polymorphism data, we estimated genetic diversity as

(note that IBD analyses were not attempted for aquatic datasets because of the difficulty in defining the distance, in river km, between collection sites in different drainage basins). Isolation-by-distance analysis was implemented using the residuals from a reduced major axis (RMA) regression of geographic distance and relative genetic distance. All RMA regressions were conducted in the R statistical computing environment (R Development Core Team 2012) using scripts written by the authors. Residuals from these regressions along with relative genetic diversity and divergence values were collected for each dataset and passed to the Genetic Landscapes GIS toolbox (Vandergast et al. 2011) for interpolation of genetic surfaces and subsequent visualization using ArcGIS v10.0 (ESRI, Redlands, CA).

GIS interpolation and visualization - genetic divergence

Associated with the Genetic Landscapes GIS toolbox (Vandergast et al. 2011) are the following four modules: Single Species Genetic Divergence, Single Species Genetic Diversity, Multiple Species Genetic Landscape, and Create Feature Class from Table. For terrestrial and aquatic datasets, we first created single species genetic divergence surfaces from the relative genetic divergence estimates (as well as RMA residuals for terrestrial datasets) using the Single Species Genetic Divergence module. Data input for the Single Species module was in the form of an Excel spreadsheet containing X and Y coordinates for each collection location as well as relative genetic divergence values. We used a point feature class of collection locations to create a triangular irregular network (TIN) such that all collection locations were connected to their nearest neighbors with non-overlapping edges; thus, forming irregularly distributed triangles. Each relative genetic divergence value (or RMA residual divergence value depending on the dataset) was mapped to the geographic midpoint between locations along each edge of the TIN. Next, we employed inverse distance weighted interpolation to generate single species surfaces

clipping individual species surfaces to the extent of the collection locations to prevent extrapolation beyond the original collection locations.

Species-specific genetic divergence surfaces were then averaged into a multiple species genetic landscape using the Multiple Species Genetic Landscape module. The module created an average surface, a variance surface, and a count surface, where the average surface was the mean raster surface for all single species divergence surfaces, the sample variance surface was a measure of the dispersion of the individual species genetic landscape values, and the count surface was the number of input surfaces that overlap in each cell of the multiple species genetic landscape. All surfaces were clipped to the boundary of the SALCC via the Extract by Mask tool in the Spatial Analyst Toolbox of ArcGIS and the mean, range, and standard deviation for composite genetic divergence surfaces (both relative divergence and IBD residuals) calculated. Areas of exceptionally high (and low) genetic divergence were defined as those more than 1.5 standard deviations (SD) above (or below) the mean value for the genetic landscape (Vandergast et al. 2008).

GIS interpolation and visualization - genetic diversity

For terrestrial and aquatic datasets, we first created single species genetic diversity surfaces from the relative genetic diversity estimates using the Single Species Genetic Diversity module. As above, data input for the Single Species module was in the form of an Excel spreadsheet containing genetic diversity values for each collection location and corresponding X and Y coordinates. Using inverse distance weighted interpolation as described above, we interpolated the genetic diversity surface for each species then calculated the average diversity multi-species genetic landscape along with corresponding variance and count surfaces. Note that surface interpolation of the multi-species genetic diversity landscape was conducted around

actual collection locations, rather than the midpoints between them (as done for the multi-species genetic divergence landscape).

Gap analyses

We used data from the protected areas database (http://gapanalysis.usgs.gov/padus/) for the SALCC to measure the proportion of the cells in the multi-species diversity and divergence raster layers (for terrestrial and aquatic species) that were both identified as hotspots of diversity or divergence and presently protected in conservation easements. Additionally, we calculated similar measures for the upper percentiles of diversity and divergence surfaces (highest 10%, 5%, 1% diversity, highest/lowest 10%, 5%, 1% divergence). These measures provide an indication of the extent to which areas currently set aside for conservation in the SALCC include areas of high value from a population genetic perspective.

Confidence in hotspot designations

We assessed the confidence in our hotspot designations for diversity and divergence surfaces first by considering the variance plots associated with the identified diversity or divergence (IBD multispecies surface only) hotspots. High variance is an indication that, for example, some taxa are highly divergent while others are less divergent across the landscape. Low variance is suggestive of consistent patterns across multiple taxa; however, low variance could also be a result of limited overlap of sampling sites. Thus, multispecies count surfaces were viewed to see if the low variance was associated with limited taxon coverage or truly represented a pattern consistent with multiple taxa. Identified hotspots with taxon coverage values greater than three were considered hotspots of greater corroboration and were the focus of further discussion.

Results

Study Area

In total, the SALCC encompasses an area of 359,345 km^2, covering a latitudinal range from southern Virginia to northern Florida and a longitudinal range from the Atlantic Ocean to eastern Alabama and the foothills of the Appalachian Mountains. Based on data provided by the National Gap Analysis Program (http://gapanalysis.usgs.gov/padus/), only about 7.6% of this area is currently composed of protected lands (Fig. 1). This percentage includes public lands such as county and state parks, wildlife management areas, game lands, national forests, and military bases, as well as private land set aside for conservation (*e.g.*, Nature Conservancy Preserves).

Terrestrial Diversity

The terrestrial diversity dataset included genetic data for a total of 16 species in the SALCC (Table 1). The relative diversity (scale 0 to 1) across species ranged from 0.186 to 0.747. The mean diversity value across species was 0.429 (SD = 0.086). Our analysis identified hotspots as areas where the relative diversity across species was greater than 0.559 (i.e., 0.429 × 1.5 × SD; Table 2). Six different hotspots of diversity were defined in this fashion (Fig. 2A). These areas were clustered in the southern and northern portions of the SALCC, with the largest in coastal North Carolina, running from the border with South Carolina north to the Pasquatank River and inland to Greenville and Goldsboro, NC (Fig. 2A). The second largest hotspot was located in southern Georgia and northern Florida, near the Apalachicola National Forest. This hotspot stretched north and east to interstate 75 north of Valdosta, GA. Three additional hotspots were located in Georgia, one north of Columbus and south of La Grange, GA, another north of Atlanta, GA, and a third between Atlanta and Athens, GA (Fig. 2A). Another small area of high diversity was found in the northern portion of the SALCC, near Danville, VA. In total, these

hotspots comprised approximately 8.2% of the area of the multispecies genetic landscape, and 9.5% of the cells defined as hotspots of diversity were currently covered by protected areas (Table 3). Areas of higher genetic diversity were better represented in protected areas. The protected fractions of cells with the highest 10%, 5%, and 1% diversity values were 9.7%, 11.2% and 25.1%, respectively (Table 3).

The mean variance in relative genetic diversity across species was 0.07 (SD = 0.045; Table 2). Hotspots identified on the basis of high variance in genetic diversity were generally closely associated with areas of high diversity (Fig. 2B). Hotspots of variance were located near the diversity hotspots north of Atlanta, north of Columbus, near the Apalachicola National Forest, and near Danville, VA (Fig. 2B). Another area of high variance in diversity across species was located inland from the large diversity hotspot along the coast of North Carolina, indicating that for some species, this hotspot may extend inland towards Raleigh, NC (Fig. 2B). A final small area of high variance was located between Asheville and Hickory, NC along interstate 40. These areas comprised approximately 4.8% of the landscape. Of these cells, only about 3.9% fell within currently protected areas in the SALCC. This fraction was higher for the highest 10% (6.4% protected), lower for the highest 5% (4.1% protected), and highest for the most variable 1% of cells (9.3% protected; Table 3).

Terrestrial Divergence

The terrestrial divergence dataset considered data from a total of 15 species (Table 1). We assessed divergence in the terrestrial landscape using both the relative divergence values and the residuals from IBD analysis. Across species, relative genetic divergence averaged 0.447 (SD = 0.093) and ranged from 0.190 to 0.790 (Table 2). The surfaces interpolated from relative divergences identified a number of areas of interest (Fig. 3A). Both regions of high and low

genetic divergence are of conservation importance, as high divergence may indicate barriers to the movement of individuals, while low divergence areas help to identify migration corridors. Several areas of high genetic divergence were located in southern Georgia and northern Florida, including areas around Lake City, FL; Tallahassee, FL; and Valdosta, GA (Fig. 3A). Additionally, a very large hotspot of divergence was identified running from Columbus and La Grange, GA to areas north of Atlanta, paralleling interstate 85 (Fig. 3A). Several smaller areas in the northern portion of the SALCC, near Raleigh, NC and the Pasquatank River, and a small area along the western border of the region in the general vicinity of Hickory, NC, were also identified as hotspots of genetic divergence.

In contrast to the pattern for areas of high genetic divergence, areas of low divergence (coolspots) were found along the coasts of North and South Carolina and along the northern portion of the SALCC, particularly along interstate 40 running from Winston-Salem, NC to Raleigh, Goldsboro, and Rocky Mount, NC (Fig. 3A). Low divergence areas were clustered in the latter area, suggesting that, while connectivity within each area appeared to be high, there was also evidence for some divergence among populations in these separate regions (Fig. 3A). The largest area of high connectivity ran along the Atlantic coast, from the Pamlico River near Greenville, NC south to Beaufort, SC. This area of low divergence extended further inland in North Carolina than in South Carolina (Fig. 3A). One additional area of low divergence was located in eastern Alabama, along the border of the SALCC north of Auburn, AL and west of La Grange, GA. A total of 8.8% of hotspot cells (which totaled 5.4% of the landscape) and 7% of coolspot cells (10% of the landscape) were represented in current protected areas (Table 3). The fraction of hotspots of divergence protected increased to 13.8% when only the highest areas of

divergence were considered (Table 3). In contrast, protection of areas of low divergence did not increase substantially in the lower quantiles of the distribution (6.4% to 6.6%; Table 3).

Residuals of the IBD analysis may be a more appropriate measure for identifying areas of high genetic divergence in our study area. This analysis accounted for the influence of geographic distance among sample sites, and therefore revealed areas where divergence was unexpectedly high (or low), given the spatial separation of populations. Very similar patterns resulted when considering the IBD residuals instead of relative divergence values (Fig. 3A vs. Fig. 4A). The mean residual value across the multispecies surface was 0.466 (SD = 0.083) and averaged residuals ranged from 0.203 to 0.781 (Table 2). Many of the same areas in southern Georgia and northern Florida were identified as displaying high genetic divergence (compare Figs. 3A and 4A). There were large areas of low divergence along the Atlantic coast of North and South Carolina, although the width along the coast of North Carolina more closely matched that in South Carolina when the surfaces were based on IBD residuals (Fig. 3A vs. Fig. 4A). Generally speaking, the northern boundary of the study area was again characterized by a lack of genetic divergence (Fig. 4A). Hotspots of divergence encompassed a relatively small fraction of the overall landscape (6.8%), and a minority of these cells were included in current protected areas (6.1%; Table 3). In contrast to the increasing protection in the upper quantiles of relative genetic divergence, only 1.6% of the cells with the highest average IBD residuals were currently protected (Table 3). Areas of low divergence occupied a similar fraction of the landscape (8.2%), but were somewhat better represented in conservation areas (10.9%). For the areas of low divergence revealed by the IBD residual analysis, roughly 6.5% were protected for all of the lower quantiles of the distribution (Table 3).

Patterns for the variance in genetic divergence across terrestrial species were also highly consistent between the two measures of divergence (Fig. 3B vs. Fig. 4B) The mean interpolated variance in relative genetic divergence across species was 0.042 (SD = 0.027), while the mean interpolated variance in IBD residuals was 0.044 (SD = 0.030; Table 2). Variance in genetic divergence across species identified a number of additional areas of biological interest (Figs. 3B and 4B). The high variance areas showed a great deal of overlap between the two measures of genetic divergence and were generally found in association with hotspots or coolspots of divergence. This pattern may indicate that the hot and coolspots would extend further for some species than for others. Areas of consistently high variance in divergence included three regions in the north of the study area, an area near Greenville, NC, one near Raleigh, NC, and another near Salisbury, NC, south of Winston-Salem, NC. These three areas were located adjacent to, or in close proximity to previously described areas of low divergence. Similarly, a large area of high variance was found near the hotspots of divergence in southern Georgia and northern Florida (see Figs. 3B and 4B). For relative divergence and the IBD residuals, 7.5% and 6.6% of the landscape consisted of areas defined by our cutoff as highly variable across species (Table 3). Of these hotspots, 4.3% and 5.3% were covered by presently defined protected areas, respectively (Table 3). Considering the upper quantiles of the variance in relative genetic divergence across species, 4.7% to 6.2% (90th percentile and 99th percentile) were protected (Table 3). These fractions were slightly higher considering the areas of highest variance in IBD residuals (5.1% to 9.7%, respectively; Table 3).

Aquatic Diversity

Our aquatic dataset included measures of genetic diversity for 14 total species (Table 1). The cross-taxon average relative diversity was 0.476 (SD = 0.130), and average values in the

landscape ranged from 0.039 to 0.853 (Table 2). We identified five hotspots of genetic diversity where relative diversity was higher than 0.671. These included one large area running from Fayetteville, NC south to Florence, SC, and two much smaller regions in the same general vicinity, between Florence and Charleston, SC (Fig. 5A). Additionally there was a diversity hotspot that encompassed a large area along the southern border of the SALCC, including Lake City, FL and Perry, FL (Fig. 5A). Finally, a small area near the Roanoke and Pasquatank Rivers of North Carolina was also identified as a hotspot in our analysis. These areas encompassed a total of 5.1% of the interpolated surface and were generally better protected than areas of high terrestrial diversity (14.3% protected; Table 3). The upper quantiles of the distribution were also better protected than areas of interest identified based on genetic divergence or variance in diversity or divergence across species (Table 3).

Areas of high variance in diversity were generally not associated with areas of high diversity, contrasting the pattern observed for the terrestrial dataset; however, there were two small areas of high variance associated with the large hotspot identified in northern Florida (Fig. 5B). The mean variance in relative aquatic diversity was 0.066 (SD = 0.046; Table 2). In addition to the two areas mentioned above, a third hotspot of variance was found near the Apalachicola National Forest (Fig. 5B). The largest contiguous hotspot identified by high variance in genetic diversity across species (Fig. 5B) was located to the north of the largest hotspot of aquatic genetic diversity (Fig. 5A vs. 5B). This area of high variance ran south from Winston-Salem, NC to just north of Columbia, SC, and east to an area near Spartanburg, SC (Fig. 5B). Several other areas of interest were located along the western border of the SALCC, near the Appalachian Mountains (north of Marietta, north of Athens, GA, between Asheville and Hickory, NC; Fig. 5B). Finally, a sizeable area to the north and west of Aiken, SC, including

Augusta, GA and running along interstate 20 towards Atlanta (Fig. 5B), also showed high variance in diversity across aquatic taxa. These hotspots made up a total of 8.4% of the area of the plotted surface, and 6.6% of the hotspots were protected (Table 3). This fraction increased in the higher variance cells (6.6% in the 90th percentile, 7.4% in the 95th percentile, and 8.5% in the 99th percentile; Table 3).

Aquatic Divergence

The aquatic divergence dataset was substantially smaller, considering data from only nine species (Table 1). The small number of species in our aquatic divergence dataset resulted in a much smaller portion of the SALCC covered in our divergence surfaces (Fig. 5 vs. 6). Mean relative divergence across species was 0.176 (SD = 0.128) and ranged from 0 to 0.718 (Table 2). Several areas of high divergence were detected between populations (Fig. 6A). The largest area of high divergence partially overlapped the area of high diversity found in central North Carolina (Fig. 6A). This area ran from Rocky Mount, NC to Florence, SC and included areas around Lumberton, Goldsboro, and Fayetteville, NC. This hotspot runs primarily North-South, and spans several river basins in the area (*e.g.*, Great Pee Dee, Lumber, Haw, and Neuse Rivers). The second largest hotspot of divergence was found in the north of the SALCC, in the area around Danville, VA. Two adjacent areas near Macon, GA were also identified as displaying elevated genetic divergence for aquatic species. Additionally, several small areas along the western boundary of the SALCC were identified as hotspots of genetic divergence (Fig. 6A). In total the high divergence areas comprised 8.7% of the total area covered by the surface, but of these cells, only 2.9% were protected by presently defined conservation areas (Table 3). This fraction increased somewhat in the upper quantiles (2.7%, 3.8%, and 7.3% for the upper 10%, 5%, and 1% of divergence values, respectively; Table 3). No areas of exceptionally low genetic

divergence were identified within the bounds of the SALCC. Nonetheless, a small fraction of cells in the lower percentiles of genetic divergence were protected (~4.7%, Table 3).

Across the nine aquatic species considered, variance in divergence was greatest in the northern portion of the study area (Fig. 6B). The mean variance in relative divergence across species was 0.062 (SD = 0.077) with a maximum of 0.5 (Table 2). Approximately 10.3% of the surface's area was classified as having high variance in divergence across species, and 2.4% of these cells were protected by currently defined protected areas (Table 3). Hotspots identified based on these data were situated near the hotspots of divergence in the northern portion of the SALCC (near Greensboro and Raleigh, NC; Fig. 6B). The fraction of protected cells increased for the 99th percentile of variance, with 6.5% of these cells currently protected (Table 3).

Confidence in hotspot designations

For the terrestrial divergence and diversity multispecies surfaces, 14 and 6 hotspots were identified, respectively (Figs 2 and 4; Table 4). Of these hotspots, only 35% (five) of the divergence hotspots and 50% (three) of the diversity hotspots were represented by more than three taxa (Table 4). Similarly, only one aquatic divergence and one aquatic diversity hotspot had a taxonomic coverage value greater than three (Table 4). Note that we chose to report and include all identified hotspots for gap analyses as a first approximation of the percent to which hotspots of genetic diversity are protected by existing conservation measures. As more data become available, these hotspots will either be verified or refuted and the estimate of percent protected upheld or refined.

Discussion.

Following Vandergast et al (2008) we defined evolutionary hotspots as areas with high evolutionary potential as measured by atypical patterns of genetic divergence, genetic diversity

and to a lesser extent genetic similarity across multiple taxa. Geographic areas displaying high genetic divergence among populations for multiple taxa may be of great evolutionary potential because they typically reflect abiotic drivers of adaptive variation (Avise 2000, Riddle and Hafner 2006). High levels of genetic diversity provide populations with a source for evolutionary change. Finally, areas of low genetic divergence may be poised for rapid evolutionary change because they may reflect relatively recent and rapid range expansion where genetic differences among populations have yet to accumulate (Lee 2002); alternatively, areas of high connectivity may reflect high ongoing rates of gene flow due to few ecological or topographic barriers to gene flow.

While our results are preliminary, they imply that, despite identifying protected areas on the basis of other factors, substantial portions of the highest areas of genetic diversity were currently protected for both aquatic and terrestrial landscapes. For instance, 14% of the cells identified as hotspots of aquatic diversity were found in currently protected areas. Additionally, 25% of the highest 1% of terrestrial diversity cells were afforded some level of protection. In contrast, areas of high and low divergence among species, and areas of high variance in diversity were poorly represented in the protected lands of the SALCC. Barriers to migration are of conservation importance as they constitute local areas where genetic differences accumulate. Occasional migration across these barriers leads to the influx of divergent alleles, which may have important implications for the potential for adaptation in response to environmental changes and for species-level retention of genetic variation. Additionally, migration corridors are of interest in that they identify areas where human modifications of the environment have not yet isolated geographically separated populations and they may allow for population connectivity on a level of demographic and ecological, rather than purely genetic, importance. Our study

provides a first assessment of the patterns of diversity and divergence across a wide variety of species in the terrestrial and aquatic habitats of the SALCC.

The degree of protection afforded to hotspots of diversity in both aquatic and terrestrial environments was substantially higher than for hotspots identified on the basis of genetic divergence or variance across species in diversity or divergence. Using the protected areas data (http://gapanalysis.usgs.gov/padus/), it was apparent that the elevated protection was due to several large protected areas that overlapped with hotspots of genetic diversity. For the terrestrial dataset, the Apalachicola National Forest, along with smaller parks like Wakulla State Forest and Lake Talquin State Forest, protected a large portion of the southern half of the northern Florida / southern Georgia hotspot identified by our study. Additionally, the hotspot in coastal North Carolina included a number of game lands, a military installation (MCB Camp Lejeune), the Croatan National Forest, the Roanoke River National Wildlife Refuge, and several other private and public conservation areas. In total, the protected areas within the SALCC encompassed an area that included 9.5% of the hotspot cells, and an impressive 25.1% of the highest 1% of diversity values in the interpolated multispecies terrestrial genetic diversity surface.

Similar patterns were evident for aquatic genetic diversity. The largest diversity hotspot included portions of Sand Hills State Forest in South Carolina and two military installations (Fort Bragg and Camp Mackall) in North Carolina. Additionally, the large hotspot of aquatic diversity in northern Florida encompassed portions of Osceola National Forest, Big Gum Swamp Wilderness, Twin Rivers State Forest, Big Bend Wildlife Management Area, and a number of smaller public and private conservation and restoration areas. It is also notable that this hotspot of diversity was adjacent to, and slightly overlapping, the Okefenokee Wilderness which covers

an area of more than 1400 km^2 in southern Georgia. In total, more than 14% of the area covered by the aquatic genetic diversity hotspots identified in this study was protected. In contrast to the relatively high degree of overall protection for genetic diversity hotspots, areas of interest identified on the basis of genetic divergence or variance across species in genetic diversity and divergence were generally less well protected. Additionally, the large geographic area covered by some of the protected lands mentioned above (Fort Bragg for instance), means that certain diversity hotspots were overrepresented when compared to others.

Given the sparse sampling of populations in many of the datasets included in our metaanalysis (especially the aquatic divergence data) and the relatively low count surfaces for many of our peripherally identified hotspots, our results must be interpreted with caution as future collection efforts could lead to significant changes in the patterns documented here and subsequently the degree of protection afforded to hotspots. Additionally, our surfaces, particularly the genetic divergence surfaces, should not be interpreted as identifying particular areas in need of conservation. The areas identified as exhibiting high genetic divergence were based on the interpolation of divergence values at the midpoints between sampled populations. For instance, if two populations displayed high genetic divergence across a geographic area, this divergence would contribute evidence that a barrier existed at the geographical midpoint between the two populations. In fact, this barrier could have been located anywhere between the two populations. With more thorough sampling, it might be possible to narrow the region of interest and identify the true barrier to migration. However, our results are unlikely to identify the exact location of barriers, and instead highlight general areas where they may exist.

Despite these caveats, there were two hotspots identified by our cross-species analysis showing low surface variance and high taxonomic representation (i.e., represented by > four

species). These areas, which may be of general biological interest and warrant further study, were as follows: an area encompassing the panhandle of Florida and southern Georgia near the Apalachicola National Forest and a large portion of the coastal regions of North and South Carolina. We discuss the significance of these areas below.

A large area of conservation interest was identified in southern Georgia and northern Florida. The area encompassing much of the Florida panhandle is recognized as major suture zone (contact zone) in North America for terrestrial and aquatic biota (Remmington 1968, Avise 2000, Rissler and Smith 2010). We would expect that a geographic area such as a suture zone would display varying levels of genetic divergence (due to species specific isolating mechanisms or lack thereof) and greater than average levels of genetic diversity (due to hybridization and backcrossing) depending on the degree of isolation between species and populations and the level of secondary contact in and around a suture zone (Vandergast et al 2008). Our terrestrial multispecies data fit this pattern, as the area included terrestrial hotspots of genetic divergence (IBD divergence hotspot no. 9), diversity (diversity hotspot no. 3), and variance in genetic diversity (Fig. 4B). The area was also an aquatic hotspot of genetic diversity (diversity hotspot no. 4). The congruence of our observed data to that of expected is testament that the multi-species genetic landscape approach may be a valid tool for identifying general areas of high evolutionary potential for conservation design and delivery. For example, the area is currently afforded much protection primarily due to the presence of the Apalachicola National Forest, Okefenokee Wilderness Area, Osceola National Forest, and several other management areas and conservation easements in the region (see protected areas in Fig. 1)

The other large region of biological interest identified by our multispecies genetic landscape was an area primarily running along the coast of North Carolina, but also into South

Carolina. The region was defined by a large area of genetic diversity for terrestrial species that overlapped with several smaller areas of low terrestrial genetic divergence suggesting that this area may be a genetic corridor of high connectivity maintaining significant levels of genetic variation. The geographic area highlighted by the hotspots is associated with the Mid-Atlantic Coastal Plain ecoregion (ranked in the top 10 in the continent in number of reptile, bird, and tree species; Ricketts et al 1999), which is typified by flat land and encompasses much of the coastal beach and dune systems of North and South Carolina. Overall, this region also has a substantial number of protected areas, as might be expected given its coastal orientation. Conservation areas in the region included (from south to north), portions of the Ashepoo, Combahee, and Edisto basin National Estuarine Research Reserve, coastal portions of Francis Marion National Forest, Marine Corps Base Camp Lejeune, the Croatan National Forest, Pocosin Lakes National Wildlife Refuge, the Roanoke River National Wildlife Refuge, and a large number of smaller game lands and conservation easements (Fig. 1).

Finally, we hypothesized that areas of high genetic divergence in terrestrial taxa would be clustered along the southern and western boundaries of the SALCC study area with areas of high connectivity in the central portion of the SALCC. This expectation was based on observations of boundaries along the Appalachian Mountains and the Apalachicola-Chattahoochee-Flint River basin previously seen in other studies (e.g., Soltis et al. 2006). While there were numerous putative terrestrial divergence hotspots identified along the western and southern boundaries of the SALCC in support of our hypothesis, the count surfaces revealed poor taxonomic coverage along the western SALCC border. Our results highlight the importance of broader geographic sampling of populations and species outside the predetermined SALCC boundaries in order to accurately determine hotspots associated along the periphery of the SALCC. In contrast we

observed lower levels of terrestrial genetic divergence coupled with low variance and high count surfaces for much of the SALCC interior suggesting that, indeed, the central portion of this region was characterized generally by high connectivity especially in areas of central Georgia and South Carolina, as well as, portions of eastern North and South Carolina.

Our study provides information about the cross-species patterns of genetic diversity and divergence in the aquatic and terrestrial environments of the SALCC. Our approach closely followed that of Vandergast et al. (2008); although, it considered a geographic area that was much larger than this previous study. Given the large amount of population genetic data that has been (and continues to be) generated and deposited in online data repositories, similar analyses in other areas can provide low-cost information with the potential to complement conservation assessments focusing on habitat types, species diversity, and patterns of endemicity. Additionally, these efforts can be updated in the future, by including samples of additional species or populations. We attempted to streamline and automate the meta-analysis of genetic datasets for use with GIS. Given the easy access to genetic datasets provided by online data repositories (*e.g.*, GenBank, DRYAD), it should be possible to substantially reduce the effort required to perform a meta-analysis such as that presented herein, or to update the results of previous assessments of diversity patterns. However, the full automation of the meta-analysis was not possible. Briefly, there were several problems with complete automation of the process. These included inconsistencies with the datasets uploaded to online repositories (lack of haplotype frequency information, lack of GPS data), separation of datasets across multiple repositories (GenBank has only DNA sequence data), and the need for extensive quality control in the data analysis phase. We discuss these limitations and provide R scripts for two functions in the Appendix. These functions should provide a starting point for future attempts to automate

the comparative analysis of genetic diversity and divergence patterns across codistributed species in a given geographic area.

Our results clearly show the need for additional population genetic studies in the southeastern United States, particularly with a focus on genetic divergence among populations of aquatic organisms. Incorporation of these additional datasets would strengthen (or perhaps change) the patterns depicted in our genetic diversity and divergence surfaces. Nonetheless, this work represents an initial assessment of cross-species genetic patterns in the SALCC study area and identifies several regions of interest from a population genetic perspective. Overall, our results show the promise of genetic datasets for identifying patterns across species. This work should stimulate future genetic monitoring and assessment in the SALCC, with a particular focus on species that are widespread and common (see the *Gambusia* dataset of Hernandez-Martich and Smith or the studies of *Pleurocera* from Dillon and colleagues). This focus would allow for adequate population-level sampling across the large geographic area considered here.

Acknowledgements

Funding for this research was provided by the South Atlantic Landscape Conservation Cooperative and the United States Fish and Wildlife Service. The authors would like to thank A. Vandergast for help with the genetic landscapes GIS toolbox. Additionally, we gratefully acknowledge the assistance with datasets we received from corresponding authors from many of the studies included in our meta-analysis.

Literature Cited

Agashe D (2009) The stabilizing effect of intraspecific genetic variation on population dynamics in novel and ancestral habitats. *The American Naturalist* **174**: 255-267.

Austin JD, Lougheed SC, Boag PT (2004) Discordant temporal and geographic patters in maternal lineages of eastern north American frogs, *Rana catesbeiana* (Ranidae) and *Pseudacris crucifer* (Hylidae). *Molecular Phylogenetics and Evolution* **32**: 799-816.

Avise JC (2000) *Phylogeography: The History and Formation of Species.* Harvard University Press: Cambridge, MA. 447 pp.

Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* **113**: 939-965.

Blum MJ, Bagley MJ, Walters DM, Jackson SA, Daniel FB, Chaloud DJ, Cade BS (2012) Genetic diversity and species diversity of stream fishes covary across a land-use gradient. *Oecologia* **168**: 83-95.

Bozarth CA, Lance SL, Civitello DJ, Glenn JL, Maldonado JE (2011) Phylogeography of the gray fox (*Urocyon cinereoargenteus*) in the eastern United States. *Journal of Mammalogy* **9**: 283-294.

Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR (2003) Evidence for multiple Pleistocene refugia in the postglacial expansion of the eastern tiger salamander, *Ambystoma tigrinum tigrinum. Evolution* **57**: 372-383.

Crutsinger GM, Collins MD, Fordyce JA, Gompert Z, Nice CC, Sanders NJ (2006) Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science* **313**: 966-968.

Degner JF, Silva DM, Hether TD, Daza JM, Hoffman EA (2010) Fat frogs, mobile genes: unexpected phylogeographic patterns for the ornate chorus frog (*Pseudacris ornate*). *Molecular Ecology* **19**: 2501-2515.

Dillon RT (1984) Geographic distance, environmental difference, and divergence between isolated populations. *Systematic Zoology* **33**: 69-82.

Dillon RT, Reed AJ (2002) A survey of genetic variation at allozyme loci among *Goniobasis* populations inhabiting Atlantic drainages of the Carolinas. *Malacologia* **44**: 23-31.

Dillon RT, Robinson JD (2011) The opposite of speciation: genetic relationships among the populations of *Pleurocera* (Gastropoda: Pleuroceridae) in central Georgia. *American Malacological Bulletin* **29**: 159-168.

Donovan MF, Semlitsch RD, Routman EJ (2000) Biogeography of the southeastern United States: a comparison of salamander phylogeographic studies. *Evolution* **54**: 1449-1456.

Ellsworth DL, Honeycutt RL, Silvy NJ, Bickham JW, Klimstra WD (1994) Historical biogeography and contemporary patterns of mitochondrial DNA variation in white-tailed deer from the southeastern United States. *Evolution* **48**: 122-136.

Forest F, Grenyer R, Rouget M, Davies TJ, Cowling RM, Faith DP, Balmford A, Manning JC, Proches S, van der Bank M, Reeves G, Hedderson TAJ, Savolainen V (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**: 757-760.

Frankham R, Lees K, Montgomery ME, England PR, Lowe EH, Briscoe DA (1999) Do population size bottlenecks reduce evolutionary potential? *Animal Conservation* **2**: 255-260.

Frankham R (2005) Genetics and extinction. *Biological Conservation* **126**: 131-140.

Grunwald C, Maceda L, Waldman J, Stabile J, Wirgin I (2008) Conservation of Atlantic sturgeon *Acipenser oxyrinchus oxyrinchus*: delineation of stock structure and distinct population segments. *Conservation Genetics* **9**: 1111-1124.

Hasselman DJ (2010) Spatial distribution of neutral genetic variation in a wide ranging anadromous clupeid, the American shad (*Alosa sapidissima*). Dissertation. Dalhousie University: Halifax, Nova Scotia. 244 pp.

Hayes JP, Harrison RG (1992) Variation in mitochondrial DNA and the biogeographic history of woodrats (*Neotoma*) of the eastern United States. *Systematic Biology* **41**: 331-344.

Herman TA (2009) Range-wide phylogeography of the four-toed salamander (*Hemidactylium scutatum*): out of Appalachia and into the glacial aftermath. Thesis. Bowling Green State University: Bowling Green, Ohio. 57 pp.

Hernandez-Martich JD (1988) Genetic variation of eastern mosquitofish (*Gambusia holbrooki* Girard) from the piedmont and coastal plain of the Altamaha, Broad-Santee and Pee Dee drainages. Thesis. University of Georgia: Athens, Georgia. 141 pp.

Hernandez-Martich JD, Smith MH (1990) Patterns of genetic variation in eastern mosquitofish (*Gambusia holbrooki* Girard) from the piedmont and coastal plain of three drainages. *Copeia* **1990**: 619-630.

Howes BJ, Lindsay B, Lougheed SC (2006) Range-wide phylogeography of a temperate lizard, the five-lined skink (*Eumeces fasciatus*). *Molecular Phylogenetics and Evolution* **40**: 183-194.

Howes BJ, Lougheed SC (2008) Genetic diversity across the range of a temperate lizard. *Journal of Biogeography* **35**: 1269-1278.

Hughes AR, Stachowicz JJ (2004) Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. *Proceedings of the National Academy of Sciences USA* **101**: 8998-9002.

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

Jackson ND, Austin CC (2010) The combined effects of rivers and refugia generate extreme cryptic fragmentation within the common ground skink (*Scincella lateralis*). *Evolution* **64**: 409-428.

Ji W, Leberg P (2002) A GIS-based approach for assessing the regional conservation status of genetic diversity: an example from the southern Appalachians. *Environmental Management* **29**: 531-544.

Kiester AR, Scott JM, Csuti B, Noss RF, Butterfield B, Sahr K, White D (1996) Conservation prioritization using GAP data. *Conservation Biology* **10**: 1332-1342.

Kozak KH, Blaine RA, Larson A (2006) Gene lineages and eastern North American paleodrainage basins: phylogeography and speciation in salamanders of the *Eurycea bislineata* species complex. *Molecular Ecology* **15**: 191-207.

Laikre L (2010) Genetic diversity is overlooked in international conservation policy implementation. *Conservation Genetics* **11**: 349-354.

Lee, CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology and Evolution* **17**: 386-391.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.

Liu FR, Moler PE, Miyamoto MM (2006) Phylogeography of the salamander genus *Pseudobranchus* in the southeastern United States. *Molecular Phylogenetics and Evolution* **39**: 149-159.

Lydeard C, Mayden RL (1995) A diverse and endangered aquatic ecosystem of the southeast United States. *Conservation Biology* **9**: 800-805.

McNeely J, Miller K, Reid W, Mittermeier R, Werner T (1990) *Conserving the World's Biological Diversity*. IUCN, World Resources Institute, Conservation International, WWF-US and the World Bank: Washington, DC. 193 pp.

Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology* **51**: 238-254.

Morton PK, Foley CJ, Schemerhorn BJ (2011) Population structure and spatial influence of agricultural variables on Hessian fly populations in the southeastern United States. *Environmental Entomology* **40**: 1303-1316.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**: 853-858.

Nei M (1972) Genetic distances between populations. *American Naturalist* **106**: 283-292.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press: New York, NY. 512 pp.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences: U.S.A* **76**: 5269-5273.

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases *Genetics* **97**:145-163

Newman CE, Rissler LJ (2011) Phylogeographic analyses of the southern leopard frog: the impact of geography and climate on the distribution of genetic lineages vs. subspecies. *Molecular Ecology* **20**: 5295-5312.

Osentoski MF, Lamb T (1995) Intraspecific phylogeography of the gopher tortoise, *Gopherus polyphemus*: RFLP analysis of amplified mtDNA segments. *Molecular Ecology* **4**: 709-718.

Peterson MA, Denno RF (1997) The influence of intraspecific variation in dispersal strategies on the genetic structure of planthopper populations. *Evolution* **51**: 1189-1206.

Petit RJ, el Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**: 844-855.

Philipp DP, Childers WF, Whitt GS (1983) A biochemical genetical evaluation of northern and Florida subspecies of largemouth bass. *Transactions of the American Fisheries Society* **112**: 1-20.

Quattro JM, Greig TW, Coykendall DK, Bowen BW, Baldwin JD (2002) Genetic issues in aquatic species management: the shortnose sturgeon (*Acipenser brevirostrum*) in the southeastern United States. *Conservation Genetics* **3**: 155-166.

R Development Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Remington, CL (1968) Suture-zones of hybrid interaction between recently joined biotas. Pages 321–428 in T. Dobzhansky, M. K. Hecht, and W. C. Steere, eds. Evolutionary biology. Plenum, NewYork.

Ricketts, T. H., E. Dinerstein, D. M. Olson, C. J. Loucks, W. Eichbaum, D. DellaSala, K. Kavanagh, P. Hedao, P. T. Hurley, K. M. Carney, R. Abell, and S. Walters. (1999) Terrestrial ecoregions of North America: a conservation assessment. Island Press, Washington, DC.

Riddle, BR, Hafner DJ (2006). A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *Journal of Arid Environments* **66:** 435-461.

Rissler LJ, Smith WH (2010) Mapping amphibian contact zones and phylogeographical break hotspots across the United States. *Molecular Ecology* **19**: 5404-5416.

Robinson JD, Diaz-Ferguson E, Poelchau MF, Pennings S, Bishop TD, Wares J (2010) Multiscale diversity in the marshes of the Georgia Coastal Ecosystems LTER. *Estuaries and Coasts* **33**: 865-877.

Schrey NM, Schrey AW, Heist EJ, Reeve JD (2011) Genetic heterogeneity in a cyclical forest pest, the southern pine beetle, *Dendroctonus frontalis*, is differentiated into east and west groups in the southeastern United States. *Journal of Insect Science* **11**: 1-10.

Scott JM, Davis F, Csuti B, Noss R, Butterfield B, Groves C, Anderson H, Caicco S, D'Erchia F, Edwards TC, Ulliman J, Wright RG (1993) Gap analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs* **123**: 3-41.

Scott CH, Cashner M, Grossman GD, Wares JP (2009) An awkward introduction: phylogeography of *Notropis lutipinnis* in its 'native' range and the Little Tennessee River. *Ecology of Freshwater Fish* **18**: 538-549.

Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15**: 4261-4293.

Stephens JD, Santos SR, Folkerts DR (2011) Genetic differentiation, structure, and a transition zone among populations of the pitcher plant moth *Exyra semicrocea*: implications for conservation. *PLoS ONE* **6**: e22658.

Vandergast AG, Bohonak AJ, Hathaway SA, Boys J, Fisher RN (2008) Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation* **141**: 1648-1664.

Vandergast AG, Perry WM, Lugo RV, Hathaway S (2011) Genetic landscapes GIS toolbox: tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources* **11**: 158-161.

Vellend M, Geber MA (2005) Connections between species diversity and genetic diversity. *Ecology Letters* **8**: 767-781.

Walker D, Avise JC (1998) Principles of phylogeography as illustrated by freshwaterand terrestrial turtles in the southeastern United States. *Annual Review of Ecology and Systematics* **29**: 23-58.

Warren ML, Burr BM, Walsh SJ, Bart HL, Cashner RC, Etnier DA, Freeman BJ, Kuhajda BR, Mayden RL, Robison HW, Ross ST, Starnes WC (2000) Diversity, distribution, and conservation status of the native freshwater fishes of the southern United States. *Fisheries* **25**: 7-31.

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

Wirgin I, Grunwald C, Stabile J, Waldman JR (2010) Delineation of discrete population segments of shortnose sturgeon *Acipenser brevistrostrum* based on mitochondrial DNA control region sequence analysis. *Conservation Genetics* **11**: 689-708.

Table 1. Taxonomic datasets, molecular markers, and genetic diversity/divergence estimators used to infer patterns of genetic diversity and divergence across species in the SALCC. The abbreviation *N* designates the number of populations surveyed for each species. Molecular marker (Marker) abbreviations are sequence (S), microsatellite (M), restriction fragment length polymorphism (R), and allozymes (A). Genetic diversity/divergence estimator (diversity; divergence) abbreviations are haploype diversity (*H*), pairwise nucleotide differences (π); expected heterozygosity (H_e), Watterson estimator (θ_w), absolute average divergence (D_{xy}), net divergence (D_a), F-statistics and analogues (F_{ST}, R_{ST}, φ_{ST}), and Nei's standard genetic distances (Nei's D). An asterisk denotes estimate taken directly from the reference.

Dataset	Species	Common name	*N*	Marker	Estimator	Reference
Terrestrial						
	Ambystoma talpoideum	Mole salamander	4	S	H; φ_{ST}*	Donovan et al. (2000)
	Ambystoma tigrinum	Tiger salamander	7	S	π; D_{xy}	Church et al. (2003)
	Dendroctonus frontalis	Southern pine beetle	9	M	H_e; F_{ST}	Schrey et al. (2011)
	Eumeces fasciatus	Five-lined skink	7	S, M	π/H_e; F_{ST}	Howes et al. (2006); Howes & Lougheed (2008)
	Exyra semicrocea	Pitcher plant moth	4	S	π; D_{xy}	Stephens et al. (2011)
	Gopherus polyphemus	Gopher tortoise	11	R	θ_W; --	Osentoski & Lamb (1995)
	Hemidactylium scutatum	Four-toed salamander	16	S	π; R_{ST}	Herman (2009)
	Mayetiola destructor	Hessian fly	7	M	H_e; R_{ST}	Morton et al. (2011)
	Prokelisia dolus	Planthopper	10	A	H_e; Nei's D	Peterson & Denno (1997)
	Prokelisia marginata	Planthopper	10	A	H_e; Nei's D	Peterson & Denno (1997)
	Pseudacris crucifer	Spring peeper	4	S	π; D_{xy}	Austin et al. (2004)
	Pseudacris ornata	Ornate chorus frog	9	M	H_e; F_{ST}	Degner et al. (2010)
	Rana catesbeiana	American bullfrog	6	S	π; D_{xy}	Austin et al. (2004)
	Rana sphenocephala	Southern leopard frog	4	S	π; D_{xy}	Newman & Rissler (2011)
	Scincella lateralis	Common Ground Skink	11	S	π; D_{xy}	Jackson & Austin (2010)
	Urocyon cinereoargenteus	Gray fox	7	S	π; D_{xy}	Bozarth et al. (2011)
Aquatic						
	Acipenser brevirostris	Shortnose sturgeon	10	S	π; D_{xy}	Quattro et al. (2002); Wirgin et al. (2010)
	Acipenser oxyrinchus	Atlantic sturgeon	7	S	π; D_{xy}	Grunwald et al. (2008)
	Alosa sapidissima	American Shad	13	M	H_e*; F_{ST}*	Hasselman (2010)
	Amia calva	Bowfin	8	R	θ_W; --	Bermingham & Avise (1986)
	Eurycea cirrigera	Southern two-lined salamander	28	S	π; D_{xy}	Kozak et al. (2006)
	Gambusia holbrooki	Eastern mosquitofish	88	A	H_e; Nei's D	Hernandez-Martich (1988); Hernandez-Martich & Smith (1990)

Lepomis gulosus	Warmouth	7	R	θ_W; --	Bermingham & Avise (1986)
Lepomis microlophus	Redear Sunfish	7	R	θ_W; --	Bermingham & Avise (1986)
Lepomis punctatus	Spotted Sunfish	8	R	θ_W; --	Bermingham & Avise (1986)
Micropterus salmoides	Largemouth bass	20	A	H_e*; --	Philipp et al. (1983)
Notropis lutipinnis	Yellowfin shiner	14	S	π*; D_a*	Scott et al. (2009)
Pleurocera catenaria	Gravel elimia	12	A	H_e; Nei's D	Dillon & Reed (2002); Dillon & Robinson (2011)
Pleurocera proxima	Sprite elimia	29	A	H_e; Nei's D	Dillon (1984); Dillon & Reed (2002); Dillon & Robinson (2011)
Pseudobranchus striatus	Northern dwarf siren	15	S	π; D_{xy}	Liu et al. (2006)

Table 2. Statistics calculated from the ten interpolated surfaces produced in this study. Hotspot (coolspot) cutoffs are 1.5*SD above (below) mean values for the surface.

		Relative Diversity		Relative Divergence		IBD Residuals	
Dataset	Statistics	Mean surface	Variance surface	Mean surface	Variance surface	Mean surface	Variance surface
Terrestrial							
	Mean	0.429	0.070	0.447	0.040	0.466	0.044
	SD	0.086	0.045	0.093	0.030	0.083	0.030
	Min	0.186	0.000	0.190	0.000	0.203	0.000
	Max	0.747	0.496	0.790	0.170	0.781	0.242
	Hotspot Cutoff	0.559	0.138	0.587	0.080	0.590	0.089
	Coolspot Cutoff	N/A	N/A	0.307	N/A	0.341	N/A
Aquatic							
	Mean	0.476	0.066	0.176	0.060	N/A	N/A
	SD	0.130	0.046	0.128	0.080	N/A	N/A
	Min	0.039	0.00	0.00	0.000	N/A	N/A
	Max	0.853	0.394	0.718	0.500	N/A	N/A
	Hotspot Cutoff	0.671	0.135	0.367	0.180	N/A	N/A
	Coolspot Cutoff	N/A	N/A	0.000	N/A	N/A	N/A

Table 3. Percent Protected

Dataset	Statistics	Relative Diversity		Relative Divergence		IBD Residuals	
		Mean surface	Variance surface	Mean surface	Variance surface	Mean surface	Variance surface
Terrestrial							
	Landscape Hotspot	8.20%	4.80%	5.40%	7.50%	6.80%	6.60%
	Hotspot Protected	9.50%	3.90%	8.80%	4.30%	6.10%	5.30%
	q90 protected	9.70%	6.40%	6.10%	4.70%	5.10%	5.10%
	q95 protected	11.20%	4.10%	9.20%	4.50%	6.80%	6.30%
	q99 protected	25.10%	9.30%	13.80%	6.20%	1.60%	9.70%
	Landscape Coolspot	N/A	N/A	10.00%	N/A	8.20%	N/A
	Coolspot Protected	N/A	N/A	7.00%	N/A	10.90%	N/A
	q10 protected	N/A	N/A	6.40%	N/A	6.50%	N/A
	q05 protected	N/A	N/A	6.40%	N/A	6.50%	N/A
	q01 protected	N/A	N/A	6.60%	N/A	6.40%	N/A
Aquatic							
	Landscape Hotspot	5.10%	8.40%	8.70%	10.30%	N/A	N/A
	Hotspot Protected	14.30%	6.60%	2.90%	2.40%	N/A	N/A
	q90 protected	12.60%	6.60%	2.70%	2.40%	N/A	N/A
	q95 protected	14.50%	7.40%	3.80%	2.20%	N/A	N/A
	q99 protected	9.80%	8.50%	7.30%	6.50%	N/A	N/A
	Landscape Coolspot	N/A	N/A	N/A	N/A	N/A	N/A
	Coolspot Protected	N/A	N/A	N/A	N/A	N/A	N/A
	q10 protected	N/A	N/A	4.70%	N/A	N/A	N/A
	q05 protected	N/A	N/A	4.70%	N/A	N/A	N/A
	q01 protected	N/A	N/A	4.70%	N/A	N/A	N/A

Table 4. Confidence in hotspot designations. Bold rows indicate hotspots represented by greater than three species. See figures 2-6 for specific hotspot designations, variance interpretation, and taxon number. Note that terrestrial hotspots were only identified for the isolation by distance (IBD) multispecies surfaces. Hotspots (i.e., areas of exceptionally high (and low) genetic divergence/diversity) were defined as those more than 1.5 standard deviations (SD) above (or below) the mean value for the genetic landscape.

Database	Surface	Hotspot	Type of hotspot	Variance	Taxon number
Terrestrial	IBD Divergence	1	Low	High	1-2
		2	Low	High	2-4
		3	Low	Low	1
		4	**Low**	**Low**	**4-5**
		5	Low	Low	1-4
		6	**Low**	**Low**	**4-5**
		7	**Low**	**Low**	**4-5**
		8	**Low**	**High**	**1-7**
		9	**High**	**Low**	**4-5**
		10	High	High	2-3
		11	High	High	2-3
		12	High	Low	1-2
		13	High	High	1-3
		14	High	High	3-4
	Diversity	1	High	Low	1-2
		2	**High**	**Low**	**3-6**
		3	**High**	**Low**	**2-6**
		4	High	High	2-3
		5	**High**	**Low**	**7**
		6	High	Low	3-4
Aquatic	Divergence	1	High	High	2-3
		2	High	High	1-2
		3	**High**	**Low**	**3-4**
		4	High	Low	2
		5	High	Low	1
		6	High	Low	1
		7	High	low	1
	Diversity	1	High	Low	2
		2	High	Low	2
		3	High	Low	3
		4	**High**	**Low**	**6**

Figure 1. Map of the South Atlantic Landscape Conservation Cooperative study area including protected areas from the National Gap Analysis Project database (http://gapanalysis.usgs.gov/padus/).

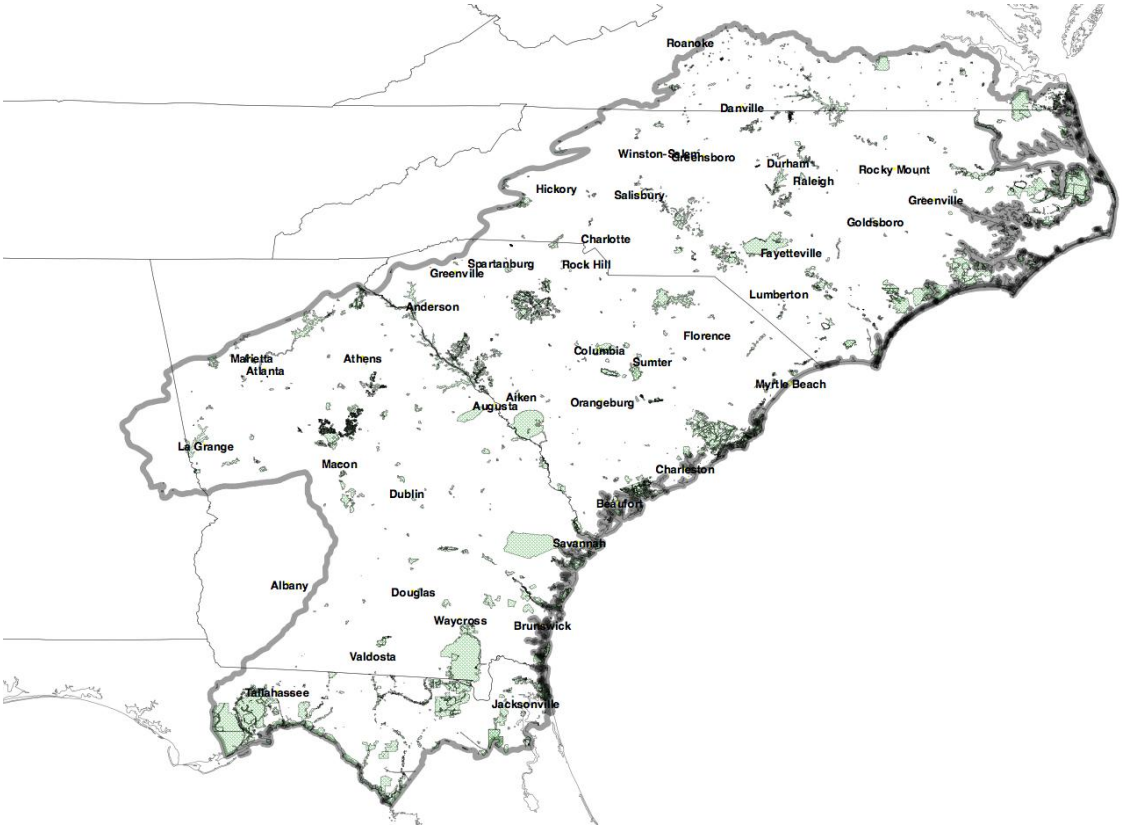
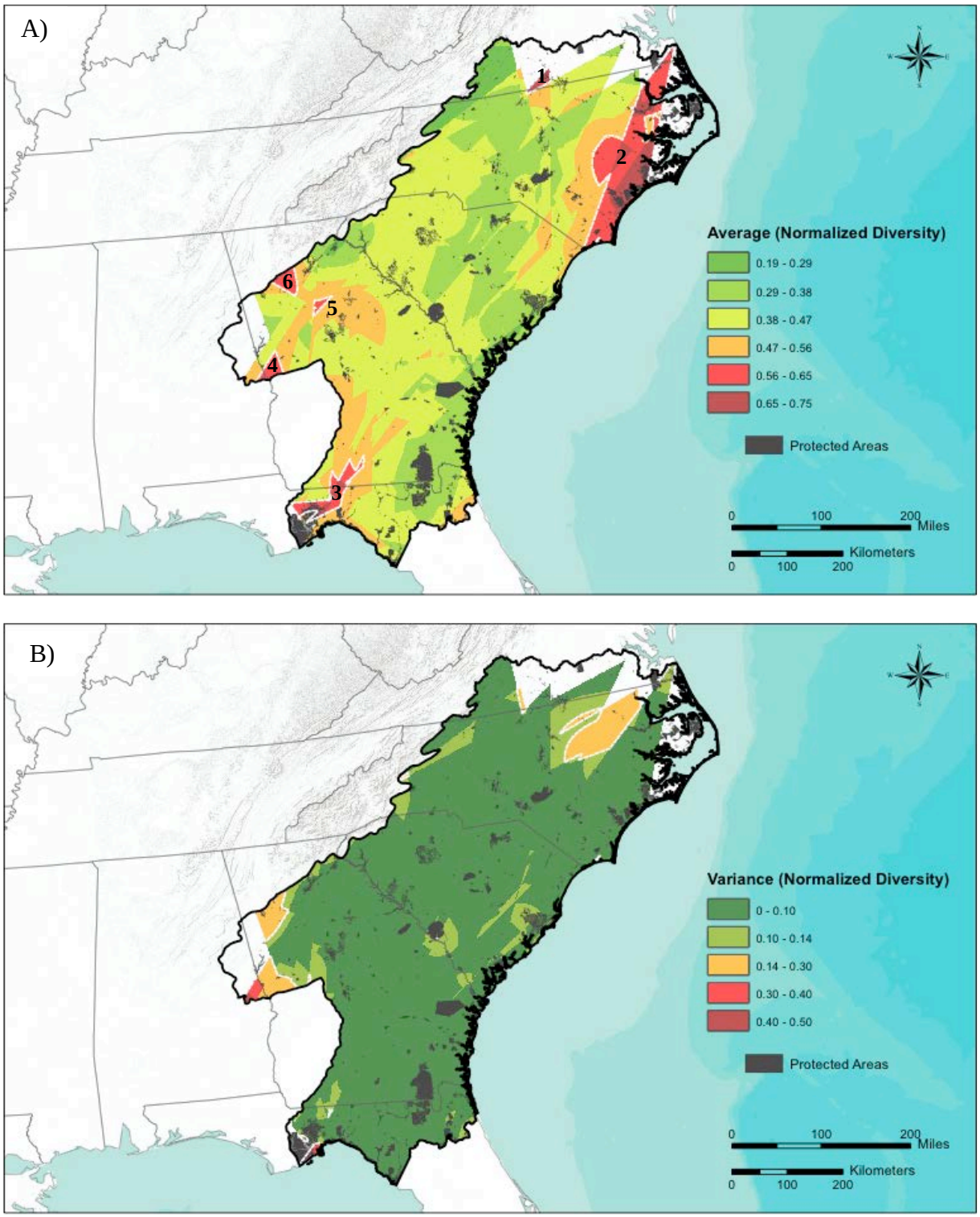


Figure 2. Multispecies terrestrial genetic diversity A) average surface, B) variance surface, and C) count surface. A total of 16 species were used to generate surfaces (Table 1). Areas outlined in white and identified with Arabic numbers are designated as hotspots (i.e., those more than 1.5 standard deviations above the mean value for the genetic landscape; Table 2).



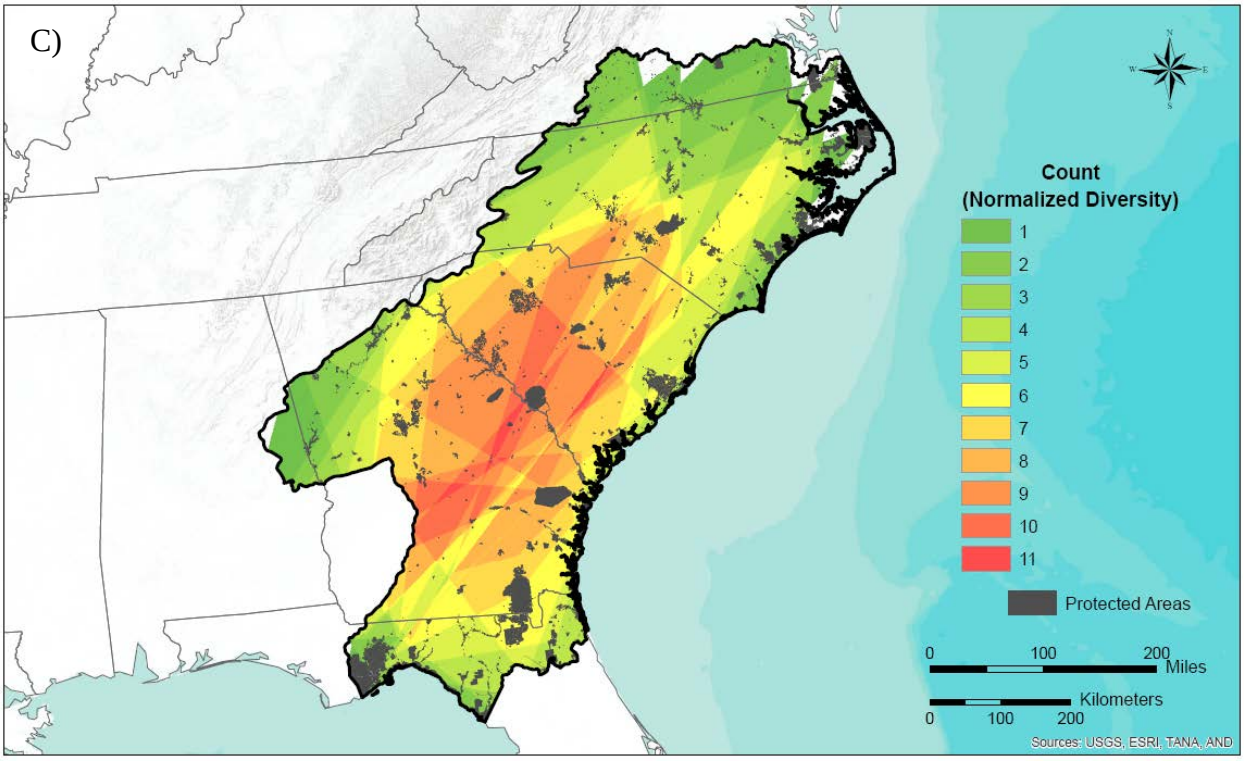
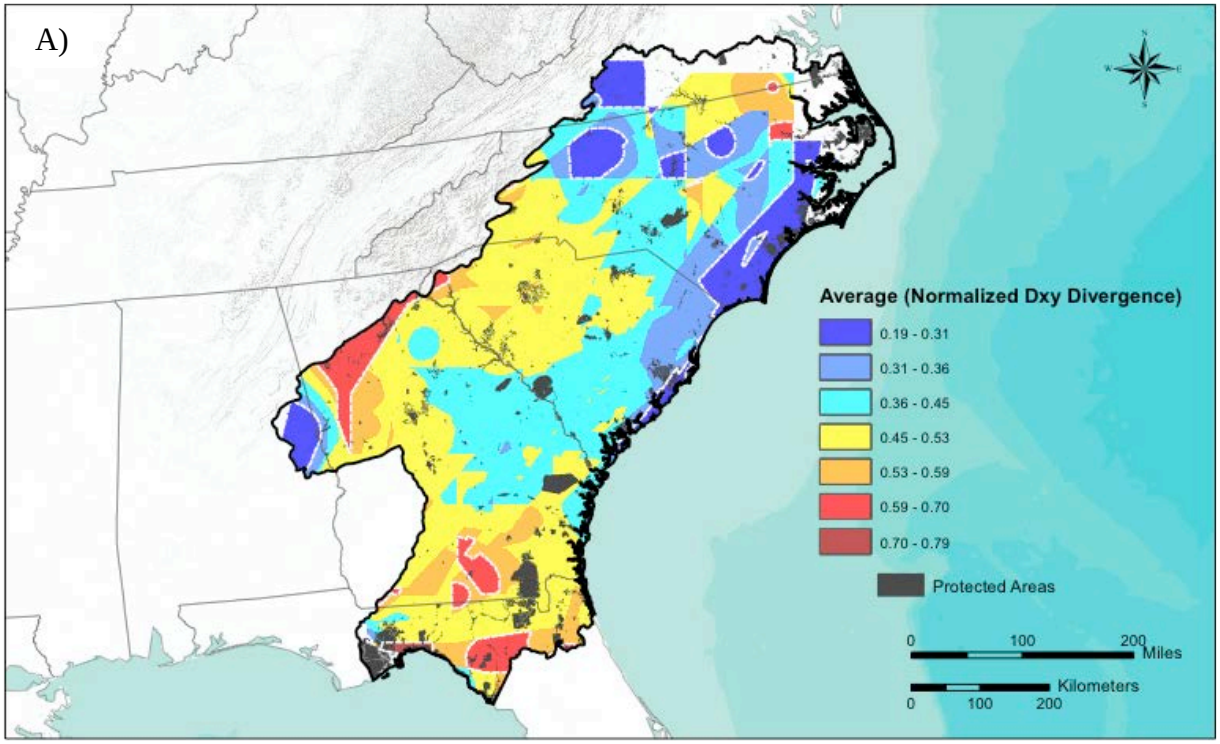
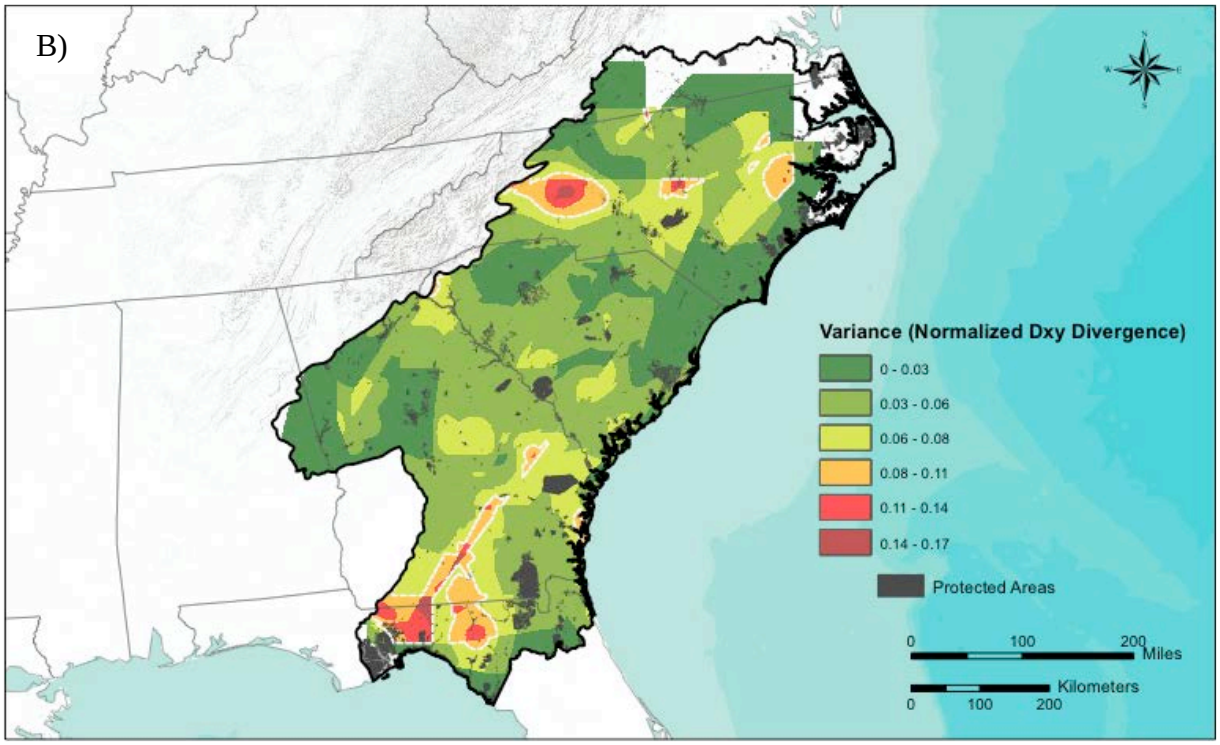

C)
Count
(Normalized Diversity)
1
2
3
4
5
6
7
8
9
10
11
Protected Areas
0
100
200
Miles
Kilometers
0
100
200
Sources: USGS, ESRI, TANA, AND

Figure 3. Multispecies terrestrial genetic divergence A) average surface, B) variance surface, and C) count surface. A total of 15 species were used to generate surfaces (Table 1). Areas outlined in white and identified with Arabic numbers are designated as hotspots (i.e., those more than 1.5 standard deviations above the mean value for the genetic landscape; Table 2).





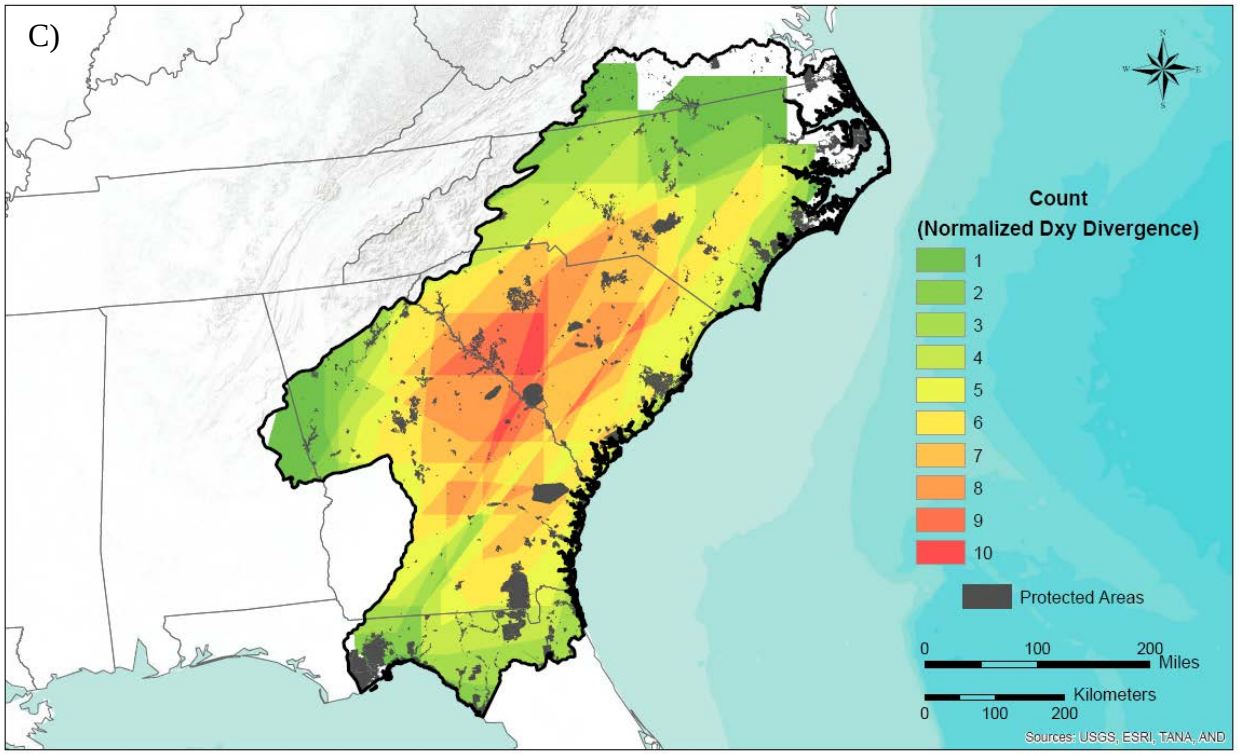
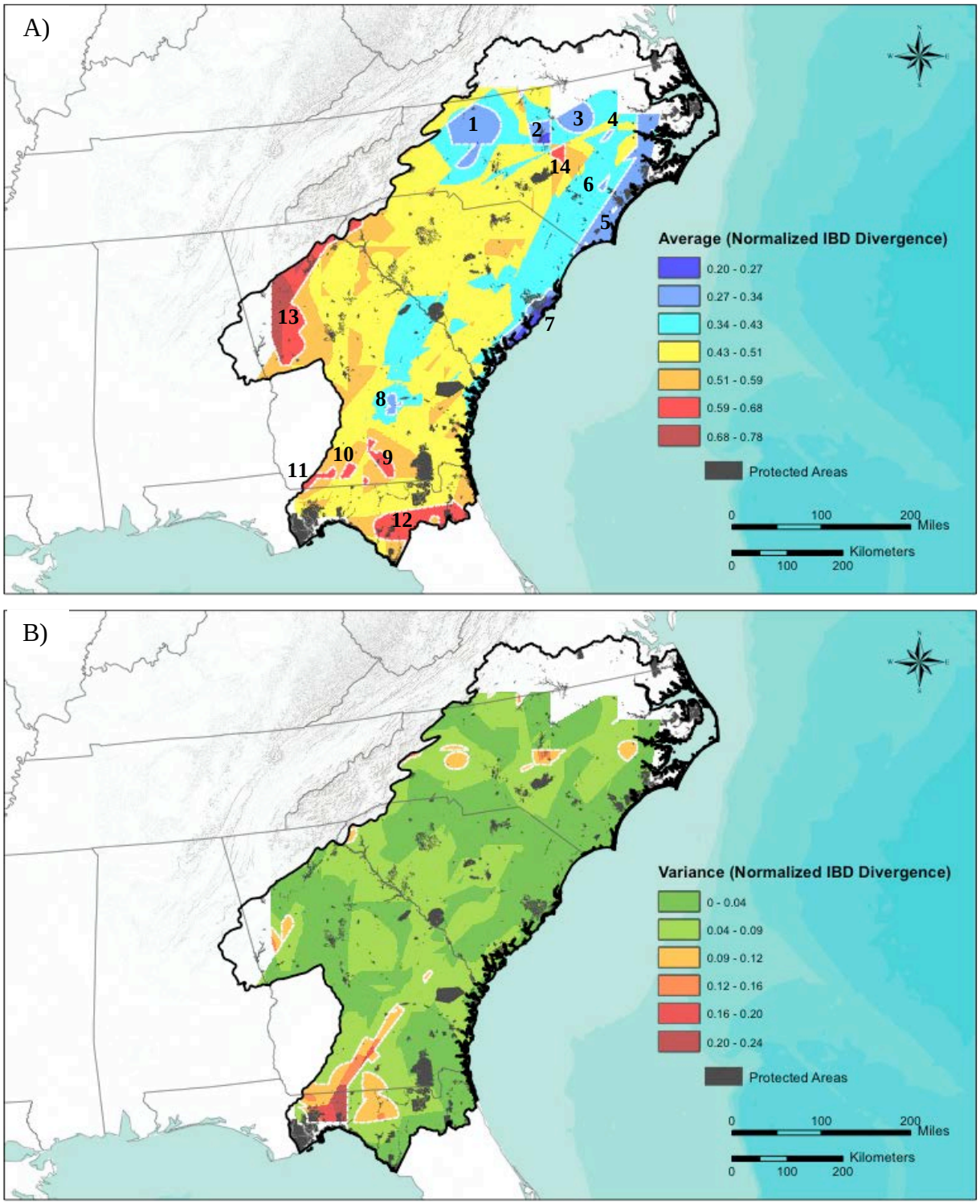

C)
Count
(Normalized Dxy Divergence)
1
2
3
4
5
6
7
8
9
10
Protected Areas
0
100
200
Miles
Kilometers
0
100
200
Sources: USGS, ESRI, TANA, AND

Figure 4. Multispecies terrestrial genetic divergence A) average surface, B) variance surface, and C) count surface based on residuals from isolation-by distance analyses. A total of 15 species were used to generate surfaces (Table 1). Areas outlined in white white and identified with Arabic numbers are identified as hotspots or coolspots (i.e., those more than 1.5 standard deviations greater or less than the mean value for the genetic landscape; Table 2).



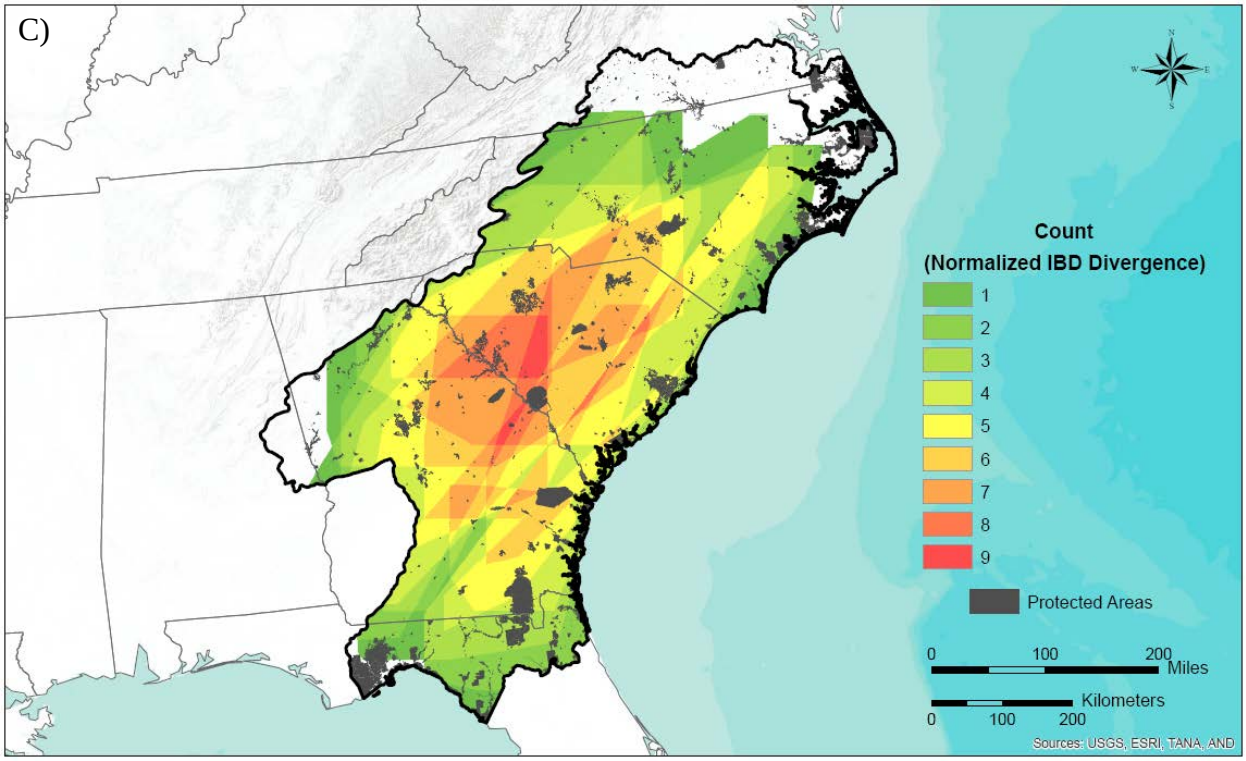
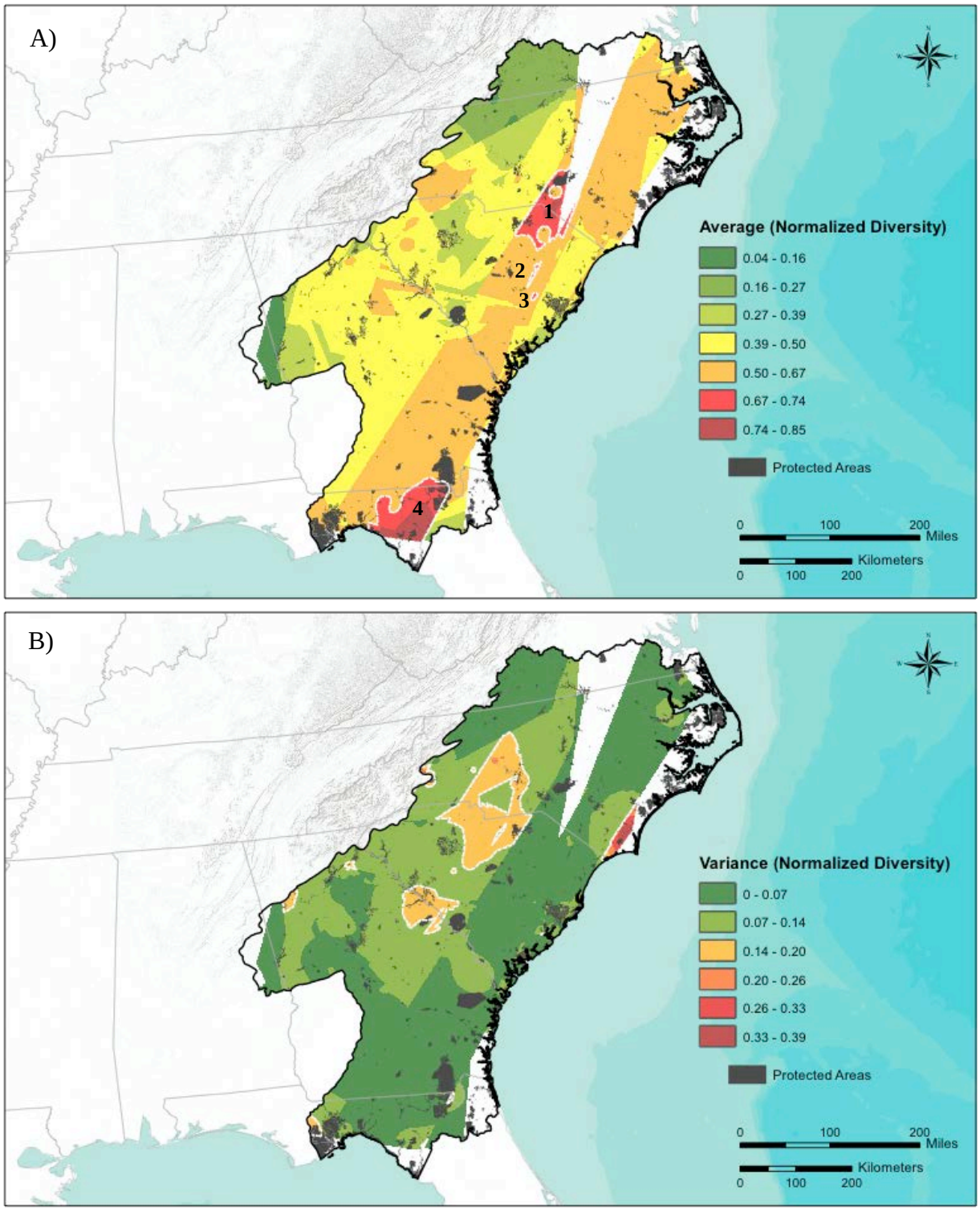

C)
Count
(Normalized IBD Divergence)
1
2
3
4
5
6
7
8
9
Protected Areas
0
100
200
Miles
Kilometers
0
100
200
Sources: USGS, ESRI, TANA, AND

Figure 5. Multispecies aquatic genetic diversity A) average surface, B) variance surface, and C) count surface. A total of 14 species were used to generate surfaces (Table 1). Areas outlined in white and identified with Arabic numbers are designated as hotspots (i.e., those more than 1.5 standard deviations above the mean value for the genetic landscape; Table 2).



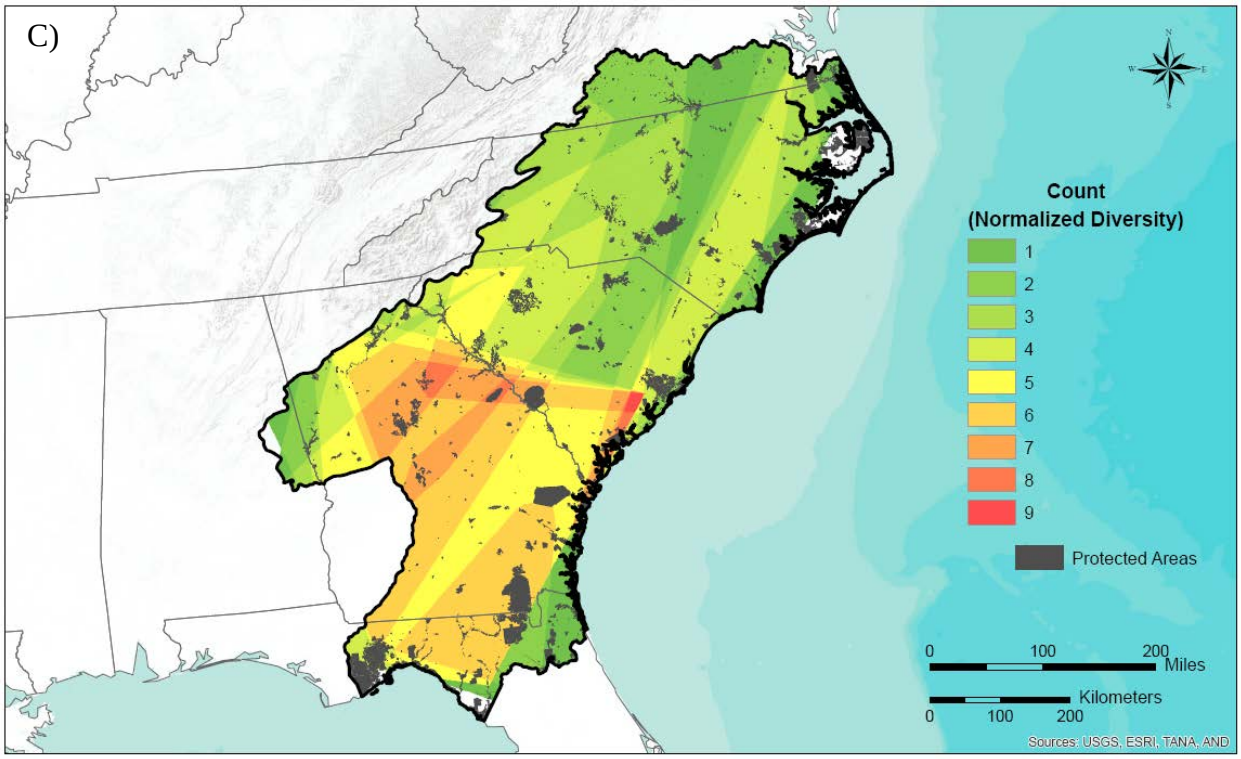
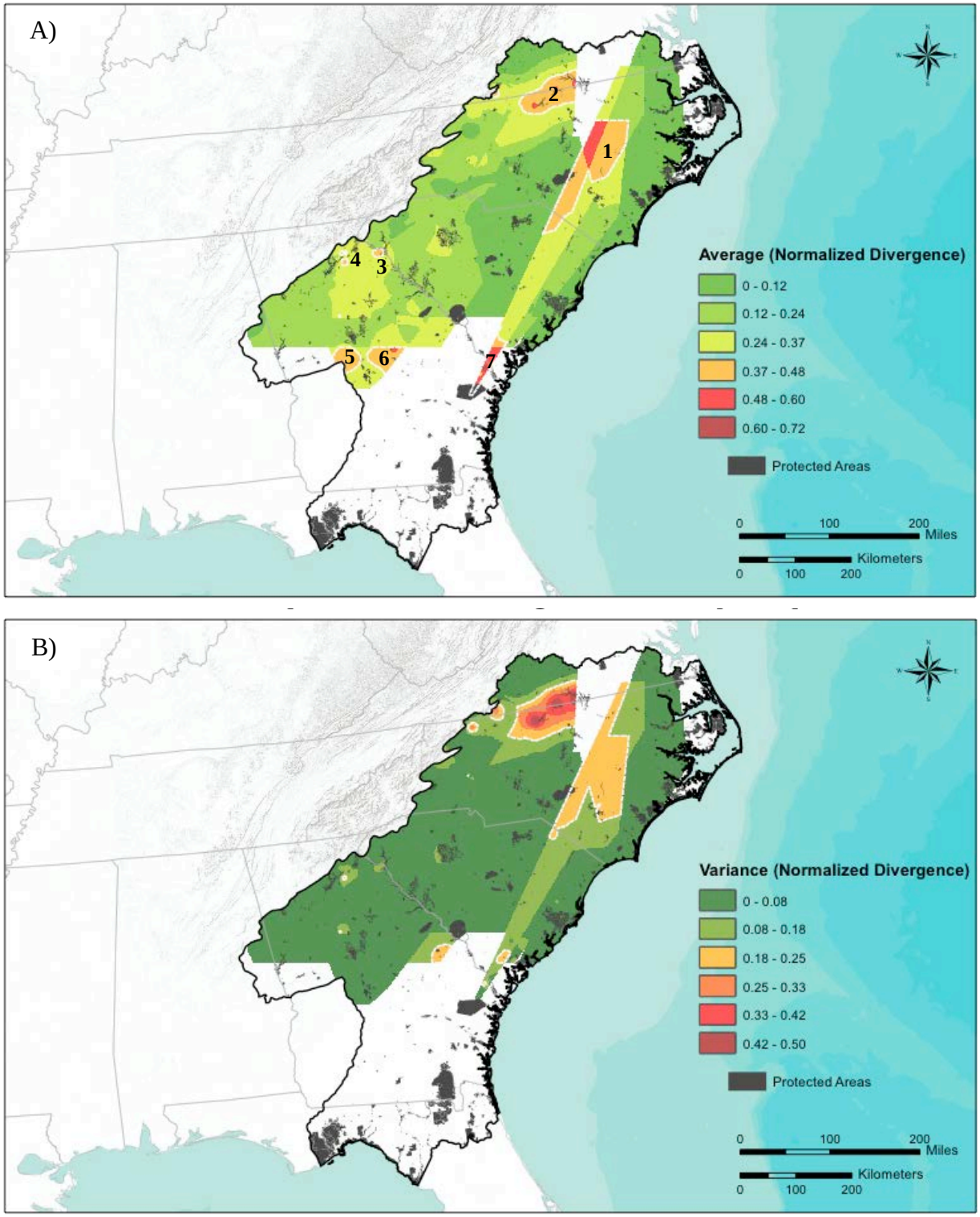
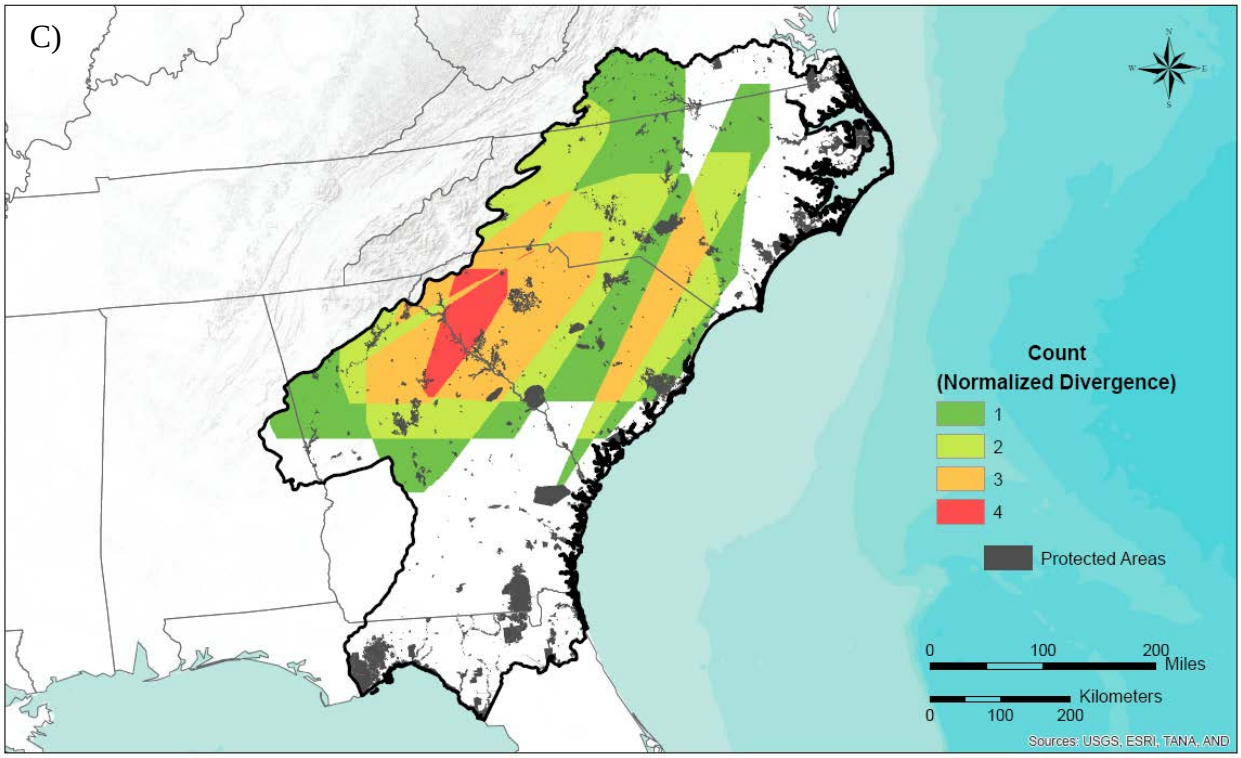

C)
Count
(Normalized Diversity)
1
2
3
4
5
6
7
8
9
Protected Areas
0
100
200
Miles
Kilometers
0
100
200
Sources: USGS, ESRI, TANA, AND

Figure 6. Multispecies aquatic genetic divergence A) average surface, B) variance surface, and C) count surface. A total of nine species were used to generate surfaces (Table 1). Areas outlined in white and identified with Arabic numbers are designated as hotspots (i.e., those more than 1.5 standard deviations above the mean value for the genetic landscape; Table 2).
No coolspots of divergence were found.




C)
Count
(Normalized Divergence)
1
2
3
4
Protected Areas
0
100
200
Miles
Kilometers
0
100
200
Sources: USGS, ESRI, TANA, AND

Appendix. Scripts for Genetic Diversity Surface Interpolation and Visualization in R

Motivation

We describe two functions written in the R statistical computing environment (R development core team 2012) to 1) use inverse distance-weighted (IDW) interpolation to project a genetic diversity surface and 2) visualize surfaces for individual species. Much of the methodology behind this work follows that of Vandergast et al. (2008, 2011). Our original goal was to develop a streamlined analytical framework for depicting genetic patterns across a geographic area using GIS tools in the R environment. This toolset would allow the user to quickly visualize genetic patterns for multiple species, aiding in the prioritization of areas for landscape conservation. To meet our goal, we needed a pipeline that included identifying and downloading appropriate data sets from online data repositories, aligning and analyzing DNA sequence data, and interpolating surfaces for each species. Several issues, which are discussed below, prevented the attainment of our original goal.

Considerations and Problematic Issues

In order to meet our goals, our tools needed to first identify genetic datasets with appropriate population-level sampling within a geographic area of interest. In general, one of the best online data repositories for population genetic studies is operated by the National Center for Biotechnology Information (NCBI – GenBank). GenBank includes data for a wide variety of taxa, but it is somewhat uncommon for users to specify the geographic location where individuals were collected. Furthermore, it is not currently possible to return a list of sequences collected within a bounding pair of coordinates. An additional complication is that users often upload a subset of the sequences they obtain (*e.g.*, only unique haplotypes). This practice would cause problems for our analyses, as the frequencies of individual haplotypes are important in calculating measures of population genetic diversity. For this reason, the first step of our analytical framework – identifying useful DNA sequence datasets – must be performed beforehand.

Because datasets from GenBank consist exclusively of DNA sequences, the framework presented in this Appendix ignores datasets that employ other molecular markers (*e.g.*, microsatellites, allozymes, RFLPs). This is problematic, as it reduces the number of species with appropriate samples in any given geographic area. Future work could attempt to develop similar tools for online data repositories that consider other data types (*e.g.*, DRYAD – microsatellites), but a large portion of the data will still be missed without a survey of the primary literature. Our pipeline would allow for other datasets to be included along with the DNA sequence data downloaded by our scripts, but the process of data compilation and analysis could be time-consuming.

Vandergast et al. (2008) used inverse distance weighted (IDW) interpolation to depict patterns of genetic diversity across the landscape and triangular irregular network (TIN) interpolation (on the midpoint between two sample sites) for genetic divergence. Our scripts include the former, but not the latter method. The “akima” R package performs TIN interpolation, but not at the midpoint between two observations. Incorporation of genetic divergence surfaces into our framework will require that a suitable R package is written, more closely following the approach of Vandergast et al. (2008).

It is important to note that the surfaces interpolated using the defaults provided below do not match those produced by the Vandergast et al. (2011) Genetic Landscapes GIS toolbox. For

instance, the surfaces interpolated by the following R scripts are bounded by the maximum and minimum latitude and longitude (*i.e.*, they are square). A more sensible method might be to trim this surface to the geographic polygon formed by the outermost sample sites. The latter approach was used in the production of our multispecies diversity and divergence figures.

Our Analytical Pipeline

The first step in order to depict genetic diversity across a geographic area is to identify appropriate population-level studies within the region. This requires the user to decide the minimum number of populations to include, and the required minimum per-population sample size. In our analysis of the SALCC, we set the minimum number of populations to four, requiring that at least three individuals were sampled per site. Following this decision, the primary literature is surveyed and GenBank Accession numbers, population assignments, and sample locations (GPS coordinates) are recorded (*n.b.*, in cases where only unique haplotypes are deposited in GenBank, accession numbers should be repeated in the list to match their frequency in the sampled population). These three pieces of data are required for the function proj.diversity() that uses IDW interpolation to project the diversity surface for a given dataset.

proj.diversity(accnos, popassign, lats, longs, model = "JC69", idpower = 3, cellsize = 0.009)

This function requires the 'ape' (Paradis *et al.* 2004) and 'gstat' (Pebesma 2004) R packages, along with an installed version of command-line clustal alignment software ('clustalw2' on Mac OS X – see ?clustal for the 'ape' package documentation; Larkin *et al.* 2007). The function proceeds to download the list of sequences whose accession numbers are given in the vector 'accnos'. Sequences are then aligned using clustal(), and genetic distances among the sequences are calculated with dist.dna() from the ape package, assuming the model of molecular evolution given by the 'model' argument. The default distance model is the simple Jukes-Cantor model ("JC69"). Next, genetic diversity is calculated as the average genetic distance between individuals within a population. Population specific values are divided by the maximum value within the dataset to give the relative diversity of a given population. These relative values are then used for surface interpolation. The 'idpower' argument sets the IDW power (larger values weight nearby points more heavily with less influence of surrounding points, producing smoother surfaces). The 'cellsize' argument determines the number of cells in the output raster layer. This value is in units of decimal degrees. The default (0.009) produces cells that are approximately 1 km^2.

```
proj.diversity = function(accnos,popassign,lats,longs, model = "JC69",idpower = 3, cellsize =
0.009) {

        require(ape);
        require(gstat);

        #Make sure all the vectors supplied have equal lengths
        test1 = (length(accnos) == length(popassign) & length(popassign) == length(lats) &
length(lats) == length(longs));
        test2 = is.function(clustal);
        if(test1 == FALSE) {
                        print(paste("ERROR: Vectors supplied are not of the same length.  accnos
= ", length(accnos),", popassign = ", length(popassign), ", lats = ",length(lats),", longs =
",length(longs),".", sep = ""));
        } else if(test2 == FALSE){
                        print("ERROR: No version of clustal found.  Please install clustal to
continue (see ?clustal).")
```

```
        } else {


                #Make a matrix of GPS coordinates

                coords = cbind(lats,longs);

                gendata=read.GenBank(accnos);
                print("Downloaded sequences from GenBank");

                #ape has a clustal() command, need to install clustalw in the /usr/local/bin
folder (Mac OSX)

        alignment = clustal(gendata);
                print("Sequences aligned with clustal");


                neworder = c();
                for(j in 1:length(popassign)) {
                        neworder[j] = popassign[dimnames(alignment)[[1]][j] == accnos];
                }

                #Use dist.dna to calculate distances (pi) between pairs of sequences

                distmat = dist.dna(alignment, model = model);
                print(paste("Distance matrix calculated based on model = ",model));

                #Calculate the average distance within and between populations here
                #Use the framework developed by sppDist in the spider package to do this... will
end up with a vector if intrapopulation diversities
                #and a matrix of divergences


        obj = as.matrix(distmat);
                attr(obj, "dimnames")[[1]] = neworder;
                pops = unique(neworder);
                piwithin = matrix(data = NA, nrow = length(pops), ncol = 2);
                piwithin[,1] = unique(pops);
                dxymat = matrix(data = NA, nrow = length(pops), ncol = length(pops));
                attr(dxymat, "dimnames") = list(unique(pops), unique(pops));

                for(i in 1:length(pops)) {
                        for(j in 1:length(pops)) {
                                newobj = obj[which(dimnames(obj)[[1]] ==
pops[i]),which(dimnames(obj)[[1]] == pops[j])];
                                if(pops[i] == pops[j]) {
                                        piwithin[i,2] = mean(newobj[lower.tri(newobj)]);
                                } else {
                                        dxymat[i,j] = mean(newobj);
                                }
                        }
                }

                dimnames(piwithin)[[2]] = c("pop","pi");
                dimnames(dxymat) = list(pops,pops);

                #Packaging all relevant information from sequences into an output for use in
mapping

                gpsdata = matrix(data = NA, nrow = length(pops), ncol = 3);

                for(x in 1:length(pops)){
                        gpsdata[x,1] = pops[x];
                        gpsdata[x,2] = unique(lats[popassign == pops[x]]);
                        gpsdata[x,3] = -1*unique(longs[popassign == pops[x]]);
                }
                out = list(coords = gpsdata, diversity = piwithin, divergence = dxymat);

                reldiv = out$diversity[,2]/max(out$diversity[,2]);
```

```
                #The "gstat" library also has functions for idw

                plot.data = data.frame(x = out$coords[,3], y = out$coords[,2], z = reldiv);
                coordinates(plot.data) = ~x+y;

                #Making a grid with cells 1km^2

                gridx = seq(from = min(out$coords[,3]), to = max(out$coords[,3]), by = cellsize);
                gridy = seq(from = min(out$coords[,2]), to = max(out$coords[,2]), by = cellsize);
                griddy = expand.grid(x = gridx, y = gridy);
                coordinates(griddy) = ~x+y;
                gridded(griddy) = TRUE;

                idwout = idw(z~1,plot.data,griddy,idp = idpower);
        }
}
```

The output from the proj.diversity() function can then be given as an argument to the vis.divsurf() function.

vis.divsurf(projdiv, bounds, alpha = 0.85)

The vis.divsurf() function is a simple function that helps to visualize the individual genetic diversity surfaces generated by proj.diversity(). It requires the R packages ‘maptools’ (Lewin-Koh *et al.* 2012), ‘maps’ (Becker *et al.* 2012), and ‘raster’ (Hijmans & van Etten 2012). Current versions of the script are specific to geographic areas in the continental United States, but could easily be modified by the user. The function begins by converting the object (class = SpatialPixelDataFrame) produced by proj.diversity() to an object compatible with the ‘raster’ package (class = “RasterLayer”). It then plots the interpolated surface in the ‘projdiv’ object, masked by the geographic area defined in the ‘bounds’ object (class = SpatialPolygon). Alpha sets the transparency of the plot, with the default allowing borders of states to be seen through the interpolated surface. The color palette ranges from purple (low diversity) to red (high diversity). The function defines the scale based on the observed values. Since relative diversities (ranging from 0 to 1) are output by proj.diversity(), values near 1 are always red, while those near 0 are always blue or purple. The scales do not exactly match between interpolated surfaces, but they are close enough to allow for quick comparison. To generate composite multispecies surfaces like those presented in our report, these individual surfaces can be averaged, using the mosaic() function in the ‘raster’ package [*e.g.*, using the following line - `mosaic(as(surf1,”RasterLayer”), as(surf2, “RasterLayer”), tolerance = 1000, fun = “mean”)`].

```
vis.divsurf = function(projdiv, bounds, alpha = 0.85) {

        require(raster);
        require(maps);
        require(maptools);

        #'bounds' will be the object (of class SatialPolygon) for defining the boundary for the
projection
        #'projdiv' is the rasterized version of output from proj.diversity()
        #'alpha' is the transparency, set default to 0.85

        projdiv2 = as(projdiv,"RasterLayer");

        x = raster();
        x = resample(x,projdiv2);
```

```
        x = rasterize(bounds,x);

        #Using the minimum and maximum diversity to define the color palette
        #This practice should make the surfaces more comparable across species, when plotted
individually
        mindiv=min(getValues(projdiv2));
        maxdiv=max(getValues(projdiv2));
        if(mindiv < 0.8) {
                blueend = 0.8-mindiv;
        } else {
                blueend = 1-mindiv;
        }
        redend = 1-maxdiv;


        #Plotting the bounding area first
        plot(bounds, lwd = 2.5);
        map(database = "state", add = TRUE);
        plot(mask(projdiv2, x), add = TRUE, alpha = alpha, col = rev(rainbow(round(50*(maxdiv-
mindiv)), end = blueend, start = redend)));


}
```

Future Directions

The next step to more fully depict genetic patterns across a landscape within our framework would be the incorporation of genetic divergence data into these plots. Vandergast et al. (2008) used TIN interpolation with divergences mapped to the midpoints between two sites along edges of the network to do this, but a similar implementation is not currently available in R. The values used could either be the relative divergence between two sites, or more appropriately, the residuals from a reduced major-axis regression of genetic distance versus geographic distance. These residuals depict the deviation from the expectation that genetic distance increases with geographic distance, and therefore are largest in absolute value when divergence is either higher, or lower, than expected given the geographic distance between two sampling sites.

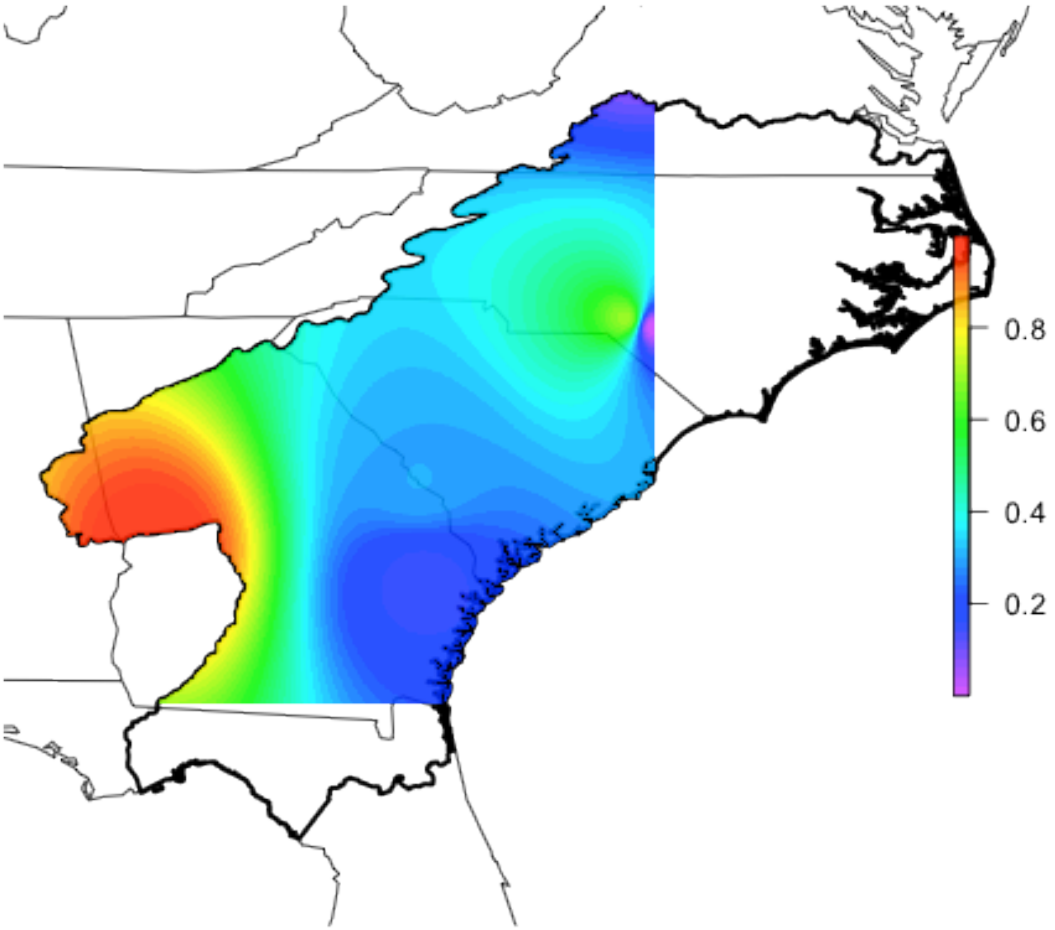
This implementation would be much less time consuming if the initial literature review step were eliminated. This would require either using a different data repository that allowed queries based on geographic coordinates and included better reporting of GPS data, or the use of bioinformatics tools to search GenBank based on the geographic position of the sample sites. It is important to note that other considerations will also apply even if this goal becomes feasible. For instance, the taxonomic status of the populations sampled may need to be verified and the frequencies of haplotypes will often need to be recorded from associated manuscripts. At this point, the datasets available in online repositories like GenBank are not complete enough to eliminate the literature survey portion of a meta-analysis across taxa. Until better information on geographic position, taxonomic status, and haplotype frequencies are regularly incorporated in data online, meta-analyses integrating population genetic data across species with GIS tools will likely be time consuming and work-intensive.

The R scripts given in this appendix should provide an avenue of exploration for researchers interested in automating this process. We have attempted to identify the major limitations to streamlined metaanalyses of population genetic datasets. At this point, these include incomplete reporting of GPS data in online data repositories, separation of datasets based on different marker types (*e.g.*, microsatellite genotypes and DNA sequences), misleading practices in data reporting (*i.e.*, only uploading unique haplotypes), and current limitations of spatial R packages. Given the abilities of the Vandergast et al. (2011) toolbox, we separated the

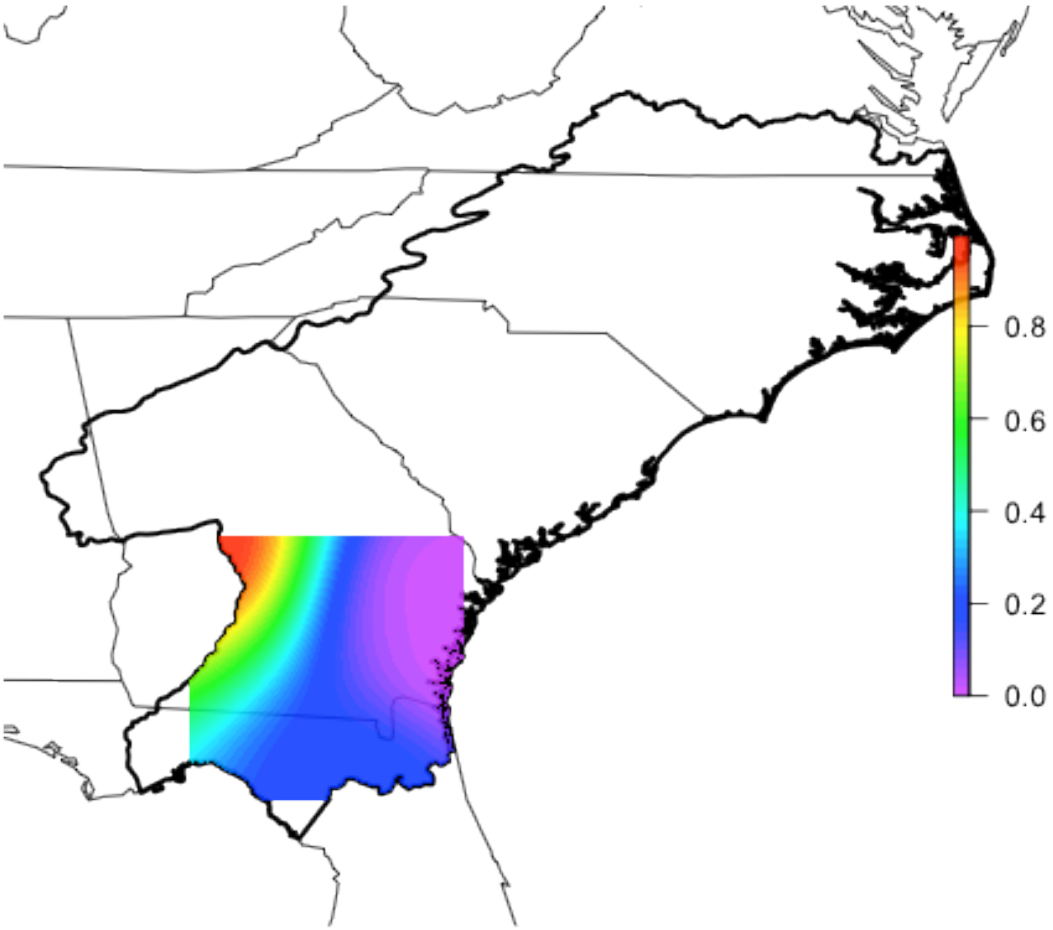
literature review, genetic analysis, and surface visualization steps in this project. At this point, this appears to be the best approach for the integration of comparative phylogeographic data with GIS.

Example Output

1. Diversity surface for *Ambystoma tigrinum* dataset (Church *et al.* 2003)



2. Diversity surface for *Rana sphenocephala* dataset (Newman & Rissler 2011)



Literature Cited

Becker RA, Wilks AR, Brownrigg R, Minka TP (2012) maps: draw geographical maps. R package ver. 2.2-6. http://CRAN.R-project.org/package=maps.

Church SA, Kraus JM, Mitchell JC, Church DR, Taylor DR (2003) Evidence for multiple Pleistocene refugia in the postglacial expansion of the eastern tiger salamander, *Ambystoma tigrinum tigrinum. Evolution* **57**: 372-383.

Hijmans RJ, van Etten J (2012) raster: geographic analysis and modeling with raster data. R package ver. 2.0-12. http://CRAN.R-project.org/package=raster.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Lewin-Koh NJ, Bivand R, Pebesma EJ, Archer E, Baddeley A, Bibiko H-J, Callahan J, Carrillo G, Dray S, Forrest D, Friendly M, Giraudoux P, Hufthammer KO, Jagger T, Luque SP, MacQueen D, Niccolai A, Short T, Snow G, Stabler B, Turner R (2012) maptools: tools for reading and handling spatial objects. R package ver. 0.8-16. http://CRAN.R-project.org/package=maptools.

Newman CE, Rissler LJ (2011) Phylogeographic analyses of the southern leopard frog: the impact of geography and climate on the distribution of genetic lineages vs. subspecies. *Molecular Ecology* **20**: 5295-5312.

Paradis E, Claude J, and Strimmer K (2004) APE: analysis of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289-290. R package ver. 3.0-5.

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**: 683-691. R package ver. 1.0-14.

R Development Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Vandergast AG, Bohonak AJ, Hathaway SA, Boys J, Fisher RN (2008) Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation* **141**: 1648-1664.

Vandergast AG, Perry WM, Lugo RV, Hathaway S (2011) Genetic landscapes GIS toolbox: tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources* **11**: 158-161.