2020

Landscape Genetics of Salamander Populations at Mammoth Cave National Park

Jackson Resser
Western Kentucky University, john.resser412@topper.wku.edu

LANDSCAPE GENETICS OF SALAMANDER POPULATIONS AT MAMMOTH CAVE NATIONAL PARK

A Capstone Experience/Thesis Project Presented in Partial Fulfillment

of the Requirements for the Degree Bachelor of Science

with Mahurin Honors College Graduate Distinction

at Western Kentucky University

By

John J. Resser

May 2020

*****

CE/T Committee:

Dr. Jarrett Johnson, Chair

Dr. Scott Grubbs

Dr. Doug McElroy

ABSTRACT

Habitat connectivity affects the distribution of genetic diversity among populations by influencing the movements of individuals and the resulting pattern of gene flow across landscapes. It has become evident that amphibians are experiencing a period of worldwide population declines brought about by environmental change. An understanding of the effects of habitat structure on landscape connectivity is important for developing effective amphibian conservation strategies. The purpose of this study is to investigate the effect of landscape characteristics on gene flow and population structure of the marbled salamander (*Ambystoma opacum*) in Mammoth Cave National Park, Kentucky, USA. Salamander larvae were sampled from 50 ponds and screened at eight microsatellite loci to estimate genetic population structure. We used the R package ResistanceGA to build and evaluate models of landscape resistance using five different habitat categories: coniferous forest, dry deciduous forest, wet deciduous forest, human influence, and surface water. Our data reveal strong support for an 'isolation by distance' model in which interpond distances are a reliable predictor of the pattern of gene flow observed.

ACKNOWLEDGEMENTS

I would like to thank Dr. Jarrett Johnson for his consistent support and guidance throughout the duration of our project. He has taught me a great deal, and the knowledge I have gained under his tutelage has been and will continue to be extremely valuable. Thank you also to Dr. Scott Grubbs and Dr. Doug McElroy for being a part of my CE/T committee.

I would also like to thank James Kevin Martin for collecting the genetic data and the WKU Biodiversity Center for providing the computational workstation that were both integral to the success of this project.

Thank you to all of my college friends, including the men of Phi Delta Theta, as well as my roommate, Matthew, and my girlfriend, Shannon. Thank you for pushing me to be the best that I can be personally and academically. I will treasure the memories that I have made in my four years at Western Kentucky University for the rest of my life.

To my family, thank you for your investment in my life at an early age. Thank you for teaching me, encouraging me to follow my dreams, and for your unwavering support. I hope to continue making you proud in the years to come.

Thank you to Vanderbilt University, where I will be continuing my education in the Fall 2020 semester by pursuing a Master of Science in Biostatistics. Thank you for believing in me and inviting me into your program. I am excited for what the future holds! Lastly, and most importantly, I would like to thank my Lord and Savior Jesus

Christ. “Whatever you do, do your work heartily, as for the Lord rather than for men” - Colossians 3:23.

VITA

Education

Western Kentucky University, Bowling Green, KY — May 2020

Bachelor of Science - Mahurin Honors College Graduate

McCracken County High School, Paducah, KY — May 2016

Professional Experience

Primary Data Analyst-Bingocize — Summer 2019-Present

Awards & Honors

Greek Academic Hall of Fame

President's List — Fall 2016, Spring 2019

Dean's List — Spring 2017, Fall 2018, Fall 2019

WKU Academic Merit Scholarship — Fall 2016

Honors Travel Abroad Grant — Fall 2017

World Topper Scholarship — Fall 2017

International Experience

Studied Abroad at Harlaxton College in Grantham, England — Fall 2017

CONTENTS

LIST OF FIGURES & TABLES

INTRODUCTION

Globally, amphibians are experiencing a period of population decline as a result of environmental change (Collins & Storfer 2003; Storfer 2003; Stuart et al. 2004; Cole & North 2014; Smith et al. 2018). Human alteration of natural landscapes results in habitat fragmentation, which can negatively affect both patterns of gene flow between populations as well as genetic diversity within populations (Noel et al. 2007; Storfer et al. 2009; Sunny et al. 2014; Cayuela et al. 2020). Quantifying the relationship between habitat fragmentation and gene flow will aid in the development of conservation strategies that enhance landscape connectivity and encourage gene flow between populations (Manel & Holderegger 2013; Mims et al. 2018; Hebbar et al. 2019).

Habitat fragmentation reduces landscape connectivity and gene flow, resulting in the increased isolation of populations. Isolation can occur from natural landscape features (e.g., rivers or mountains) or from anthropogenic landscape features (e.g., roads, buildings, or agriculture; Garcia-Gonzalez et al. 2011). In the absence of gene flow, genetic diversity is decreased and inbreeding activity is increased due to lack of genetic inputs from immigrating individuals (Sunny et al. 2014; Arntzen et al. 2017). Inbreeding is commonly associated with isolated populations and can lead to the increased expression of deleterious mutations that are harmful to amphibian fitness (Emaresi et al. 2011). In contrast, groups of populations that frequently exchange individuals are more genetically diverse and less likely to suffer from the effects of genetic drift and inbreeding. The implementation of management practices that will stimulate gene flow

are increasingly vital to the persistence of amphibian species inhabiting landscapes affected by human influences.

The principles of landscape ecology and population genetics can be incorporated into a single approach known as landscape genetics (Manel et al. 2003). The landscape genetics discipline involves the modeling of differences in allele frequencies for adjacent populations while also accounting for the various intervening landscape features (Storfer et al. 2007). The field of landscape genetics relates spatial patterns of geographic variation in habitat (e.g., forests, grasslands, urban areas) and barriers to movement (e.g., roads, rivers) to patterns of genetic diversity using procedures for optimizing statistical models (Manel et al. 2003).

All landscapes are heterogeneous in some way, and the resources contained within landscapes are patchy. In a landscape ecology context, a basic binary description of landscape heterogeneity is between "matrix" and "habitat". Habitat comprises the areas of the landscape that are suitable for individuals to carry out life history processes and for populations to persist. Matrix comprises the areas where life history processes cannot be suitably completed in such a way for populations persist. Even landscapes that are entirely comprised of habitat can be considered heterogeneous based the distribution of different types of plant communities. The matrix and different vegetation types serve to affect an individual's ability to traverse the landscape during dispersal movements between populations (i.e., impeding or facilitating gene flow), and so the effect of landscape heterogeneity on population dynamics is measured in terms of landscape resistance, analogous to the movement of electricity moving through a circuit.

It is important to note, however, that not every barrier or vegetation type that impedes movement is equally resistant. For example, while multi-lane interstate highways with concrete barriers likely prevent all salamanders from completely traversing them, other features, like unpaved country roads are much less formidable and may have a negligible effect on movement. In addition, qualities of the environment can affect different taxa in different ways. For example, birds would be much less affected by an interstate highway due to their ability to fly over the barrier.

The problem is that it is unclear what resistance values to attribute to each landscape feature. Early landscape genetics studies relied heavily on "expert opinion" to model the resistance of landscape features. For example, salamander experts might agree that forests have low resistance to movements and parking lots have high resistance to movements because the former provides a better physiological environment than the latter; amphibian skin must stay moist. Such logic may be true, but what happens on rainy nights? A parking lot devoid of cars and saturated with water may provide an excellent surface through which to move unimpeded by downed logs and dense brush.

A better strategy than relying on expert opinion to parameterize resistance surfaces is to determine the resistance values through an iterative optimization procedure that derives the values from the data. ResistanceGA is an R package that optimizes resistance surfaces by making use of genetic algorithms that simulate hundreds of generations of gene flow for the study landscape (Peterman 2014). ResistanceGA uses maximum-likelihood population-effects (MLPE) mixed models to describe the relationship between pairwise genetic distances and pairwise geographic distance in the case of isolation by distance (IBD) models and pairwise resistance distances in the case

of isolation by resistance (IBR) models. The MLPE models use each pairwise population combination as a data point, but because each population is included in multiple pairwise combinations, each data point is not independent, and thus violates an assumption of linear models. The lack of independence is accommodated through the incorporation of a population-level factor (i.e., a random factor in mixed effect model terminology) that distinguishes between data points that are independent from those that are not (Clarke et al. 2002).

The objective of this study is to examine landscape characteristics and its effect on gene flow and population genetic structure of the marbled salamander (*Ambystoma opacum*) at Mammoth Cave National Park (MCNP). Studying the marbled salamander populations within MCNP will serve as a model system to reveal patterns of gene flow within a large-continuous tract of habitat for a species that is not experiencing declines, to help in developing strategies that will both encourage genetic diversity and impede population decline elsewhere. While most conservation approaches tend to be species- or landscape-specific, the results of this study have the potential to provide broad strategies that can be implemented when specialized local data are not available.

Both IBD and IBR models will be evaluated to explain the observed pattern of gene flow on the MCNP landscape. The IBD model explains the genetic variation found among populations on a landscape solely based on the Euclidean distance between each pair of sampled populations. The isolation-by resistance model (IBR) explains the genetic variation by incorporating the measures of heterogeneity associated with the landscape. For the IBR model, the “resistance” of the landscape is determined by the pattern of heterogeneity in landscape features. It is expected that an IBR model will best explain the

genetic distribution present in these salamander populations because the MCNP landscape comprises both heterogeneity in vegetation communities and potential barriers to salamander movement. Specifically, it is expected that forest vegetation type, the Green River, and human developments within the park have influence on salamander movements, with an *a priori* expectation that the river will be found to provide the most resistance to gene flow and wet deciduous forest will be found to provide the least resistance.

METHODS

Study Species

The marbled salamander (*Ambystoma opacum*) is a species in the "mole" salamander family (Ambystomatidae) inhabiting the state of Kentucky as well as a significant portion of the Eastern United States (Barbour 1971). During the non-breeding season, marbled salamander adults can typically be found in damp underground burrows near breeding sites (Horton & Kemp 2013). Most species of *Ambystoma* breed in the spring, but marbled salamanders breed in the fall following migration back from their summer habitats to natal ponds to undertake courtship and mating (Taylor & Scott 1997). Eggs are laid on land in shallow, self-excavated burrows within dry pond basins. After a period of overwintering in the egg state, nests are inundated during spring rains. The larvae are aquatic, and owing to their early hatching date, gain a significant size advantage over the larvae of spring-breeding salamander species. Metamorphosis takes place in the summer, at which time the metamorphosed juvenile salamanders emigrate from the natal pond to a suitable upland burrow within which to develop to maturity over the period of several years (Barbour 1971).

Study Landscape

Mammoth Cave National Park consists of more than 50,000 acres of forest along with the longest known cave system in the world that spans more than 400 miles (National Park

Service 2018). The cave exists due to erosional forces acting on a sandstone layer, exposing the limestone underneath. Surface runoff moves underground through sinkholes, forming the caves present today (National Park Service 2018). Surface ponds then form in these sinkholes during periods of heavy rain and the ephemeral nature of the ponds keeps them fishless, making them an attractive habitat for breeding amphibians (Figure 1). Martin (2013) sampled larvae from 50 different ponds in Mammoth Cave National Park: (MCNP; Figure 2). The park is home to a plethora of different amphibian species, including the marbled salamander (*Ambystoma opacum*). The park is bisected by the Green River and comprises a variety of vegetation types, including forest/savanna, oak-hickory/savanna, karst valley forest/prairie, mesic slope and floodplain forests, cedar-oak forest glades, ridgetop pine-oak stands, and prairie. Mammoth Cave National Park earned the title of International Biosphere Reserve in 1990 (National Park Service 2018) and offers a largely undisturbed model system that can be used to analyze the genetic structure and gene flow present in the native populations.

Sample Collection

Sample sites were selected based on topographic map data, GIS wetlands layers, Google Earth imagery, park ranger knowledge, and random encounters (Martin 2013). Larvae were captured using dip nets and a 1 cm tail clipping was collected from 12-30 larvae within each pond. Larval tail-clipping has proven to be an efficient method for collecting genetic data while also exhibiting little effect on the survival of the individual (Wilbur and Semlitsch 1990). After the tail clippings were collected, the individual larvae were promptly released and the tissues were placed in 95% ethanol for storage. All tissue

collection was supported by Kentucky Department of Fish and Wildlife permit #SC1211057 and Mammoth Cave Scientific Research and Collecting permit #MACA-2012-SCI-0001.

Genetic Data Collection

DNA extraction was performed on the collected tissues using standard phenol-chloroform, DNEasy Blood and Tissue Kits® (Qiagen Inc.) or protein precipitation procedures (Martin 2013). The extracted DNA was screened for amplification and polymorphism of microsatellite markers at 10 loci designed for the marbled salamander (Nunziata et al. 2011). Using the universal fluorescence labeled primer method (Nunziata et al. 2011), polymerase chain reaction (PCR) products at all genetic markers were generated. For each individual PCR, fluorescently-labeled PCR products were multiplexed. The multiplexed samples were then scored either at the Western Kentucky University Biotechnology Center using a 3130 Genetic Analyzer or at the University of Georgia Genomic Facility with a 3730xl 96-capillary DNA Analyzer. Resulting genotypic data were analyzed with GeneMapper® v3.7 software (Applied Biosystems, Inc.).

Geographic Data Collection

Pairwise geographic distances between the ponds were calculated using the GPS coordinates at each sample site as well as ArcMap (Esri, Inc.). A raster surface was generated using data from the United States Geological Survey (USGS 2011). Each different type of habitat represented in the layer was assigned a resistance category

depending on the primary vegetation type as well as expert opinion concerning the wetness or dryness of the habitat (Martin 2013; Table 1). The raster surface consisted of five landscape categories: coniferous forest, wet deciduous forest, dry deciduous forest, human influence, and water. The category termed human influence comprised areas affected by anthropogenic activity, including water-willow rock bar and shore, highland rim limestone cliff and talus seep, rock, soil, agriculture, lawn, power line easement, building, commercial, parking lot, road, residential, successional broomsedge vegetation, cultivated meadow, and blackberry-greenbrier successional shrubland thicket. The water category referenced the Green River, which divides the park.

Relationship between Geographic and Genetic Data

The data used by ResistanceGA to run the computations consisted of pairwise geographic data and pairwise genetic data. The pairwise geographic data contained x and y values denoting the coordinates belonging to each of the sample ponds. The pairwise genetic data consisted of a matrix denoting the F_{ST} values between one pond and another. An F_{ST} value is a measure of the genetic differentiation between two populations, with values close to zero indicating low differentiation (i.e., high gene flow) and values close to one representing high differentiation (i.e., low gene flow). Using these data, a scatterplot was constructed in order to visualize the statistical relationship between the Euclidean distance and genetic diversity of the samples. The code used to construct the scatterplot is included in Appendix 1. Additionally, because the river was expected to be a significant barrier to movement based on prior research, a matrix was constructed that describes the

comparisons between genetic distance and geographic distance for each pair of sample sites.

Partial Mantel Test

A partial Mantel Test was conducted as a basic comparison of the pairwise genetic, geographic, and resistance distance matrices. A Mantel test investigates the correlation between two N x N matrices, and a partial Mantel test allows for the comparison of two matrices while controlling for a third. In this case we wish to test for a relationship between pairwise geographic distance (matrix 1), pairwise genetic distance (matrix 2), and the pairwise "resistance distance" optimized in ResistanceGA (matrix 3; see below). Strong correlation between pairwise geographic and genetic distance matrices supports an IBD model of genetic structure, and if the correlation is improved with the inclusion of the pairwise resistance data, an IBR model of genetic structure is favored. The code used to run the partial Mantel Test is included in Appendix 2.

Analysis with ResistanceGA

The R package ResistanceGA makes use of a genetic algorithm to model gene flow among breeding ponds on a landscape (Peterman 2014). As described above, ResistanceGA optimizes models of habitat resistance based on a landscape-specific GIS habitat layer through the iterative estimation of model likelihood values across a range of parameter estimates for each habitat type. In other words, ResistanceGA assigns a resistance value to each of the different habitat types comprising the landscape by evaluating the likelihood that the resulting landscape resistance pattern explains the

genetic structure on the landscape. Genetic structure is described as pairwise genetic distances between the sample ponds along with the coordinates of the ponds. As a result, we can test between IBD and IBR models of gene flow for the study landscape.

Three models were considered (Table 2). The first model was a null model with only a constant explanatory term and the pairwise pond ID. The second model was an IBD model, which considered the Euclidean distance between the ponds as an additional explanatory variable along with the pairwise pond ID. The third model was an IBR model, which accounts for the variation in habitat resistance between the ponds along with the pairwise pond ID. The code used to run the Resistance GA package in R is included in (Appendix 3). For this study, the landscape data used was the categorical raster surface derived from the US Geological Survey data from Mammoth Cave National Park, while the pairwise genetic distances were calculated by Martin (2013) as normalized F_{ST} values. The coordinate locations for each of the 50 ponds were used as sample locations (Table 3).

The quantification of resistance values for particular landscape features in IBR models is accomplished through the construction of a resistance surface, which is a spatial GIS layer that describes the locations of the various features on the landscape. For each category of landscape feature, a resistance value is attributed that represents the degree to which the feature either inhibits or enhances individual movements and thus gene flow and landscape connectivity (Spear et al. 2010). Resistance surfaces represent hypothesized relationships between landscape variables and movement or gene flow that can be tested for statistical significance (O'Brien et al. 2006; Wang et al. 2008).

Several representations of the study landscape were generated because the ResistanceGA algorithm is sensitive to the scale of the landscape under analysis, and to generate a computationally reasonable time frame for completing the analytical iterations. Using ArcGIS, the vegetation raster layer was converted to an ASCII file, which is a text file detailing the landscape numerically by assigning a value to each pixel. The ASCII file was resized to three different resolutions: 50, 75, and 100 meters per pixel. For each iteration of ResistanceGA, a linear mixed-effects model with maximum likelihood population effects parametrization (MLPE) was created. For each landscape feature, resistance values ranging from 1 to 2500 were assigned by first scaling one landscape feature to 1. From there, the other landscape features were assigned values based on how conducive they were to movement across the landscape: 1 denoting a landscape feature least resistant to movement and 2500 denoting a landscape feature most resistant to movement.

Analyses were performed on a custom-built 40-core Linux computational workstation provided by the WKU Biodiversity Center. Using multiple computer cores as opposed to one improves the processing efficiency of the analyses as multiple iterations of ResistanceGA can be running simultaneously. ResistanceGA selected models based on the best log-likelihood values and created new models in an attempt to further improve the values.

RESULTS

Relationship Between Geographic and Genetic Data

The scatterplot depicting the relationship between the distance of the sample pond locations and the F_{ST} values is shown in Figure 3. The x-axis represents the natural logarithm of the distance between each pond pair, while the y-axis represents the F_{ST} value associated with the two ponds. The slope of the line of best fit associated with the natural logarithm of the distance was found to be significant at $\alpha=0.05$ with a p-value of <0.001, but the adjusted R^2 value was 0.044 which indicates that only 4.4% of the variance in F_{ST} is explained by geographic distance.

The matrix describing the genetic distance and geographic distance between each pair of sample sites is included in Figure 4. The upper triangular portion represents the genetic distance and the lower triangular portion represents the geographic distance. The top-most rows and right-most columns represent the eight ponds located north of the river. Lighter colors represent values closer to zero for both pairwise genetic distance (F_{ST}, shown in blue) and pairwise geographic distance (shown in red). A strong signal of IBD would be suggested if the pattern of variation in color intensity was similar for both the upper and lower halves of the matrix. If the Green River consists of a strong barrier to marbled salamander gene flow at MCNP, F_{ST} are expected to be higher for pond pairings that span the river, irrespective of geographic distance. The preponderance of dark blue values in the top rows of the matrix and lack of corresponding dark red values in the

right-most columns suggest that the Green River is a landscape feature that reduces gene flow.

The results of the Mantel Tests are included in Table 4. The Mantel R for the comparison of the pairwise geographic and genetic distance matrices (R=0.1962, P=0.0073) indicates support for the IBD model of population structure at MCNP. Further, the inclusion of the resistance data matrix does not appreciably improve the correlation between genetic and geographic distances (R=0.1979, P=0.0069), indicating little support for the IBR model. These results suggest that the modulation of straight-line geographic distance to represent heterogeneity in landscape resistance does not improve our ability to describe the pattern of genetic structure on the landscape.

Landscape Model Comparisons

Despite the poor proportion of variation explained by a simple relationship between F_{ST} and geographic distance, ResistanceGA converged on an isolation-by-distance model as the best explanation of the genetic distribution between the sampled ponds (Table 5). Based on the AICc values, the IBD model performed better than the null model at the 50-m resistance surface resolution, at the 75-m resolution, and at the 100-m resolution. The resistance model term did not significantly improve the algorithm's ability to successfully predict the pairwise genetic distances between the ponds as shown by the lack of improvement in model performance for the IBR model versus the IBD model. The isolation-by-distance model is a more parsimonious explanation for the data. Models with ΔAICc values >2.0 are generally viewed as significantly poorer performing than the model with the lowest AICc.

Although the landscape resistance term was not significant in our model, Table 6 shows the coefficients used for the resistance surface of the best IBR model. Table 6 shows that the wet deciduous forest vegetation category was assigned the smallest resistance value (1), followed by the coniferous forest (4.0), then the dry deciduous forest (4.4), then human influence (10.2), and water (880.7) after model optimization.

DISCUSSION

Based on the results of the partial Mantel Test and the model comparison analysis, the IBD model is a better model to describe the gene flow among the sampled populations of *Ambystoma opacum* at Mammoth Cave National Park than the IBR model, although both the IBD and the IBR models outperformed the null model. The vegetation type as an explanatory variable did not result in a significant improvement to the model. Thus, vegetation type in Mammoth Cave National Park does not have a significant effect on gene flow between adjacent populations of marbled salamanders.

These results do not align with what was expected, as the initial hypothesis was that the IBR model would do the best job of explaining the gene flow within the sampled populations. According to McRae (2007), the resistance distance is more theoretically justified and more robust to spatial heterogeneity as a predictor of genetic differentiation than Euclidean or least cost path-based distance measures. One potential explanation for why the initial expectations were incorrect is that landscape heterogeneity within MCNP is not sufficient to affect the movements of individuals and influence patterns of gene flow. Similarly, Hagerty et al. (2011) and Latch et al. (2011) pointed to the IBD model as

the best model for summarizing the genetic connectivity of the desert tortoise *Gopherus agassizii* in a landscape without high landscape heterogeneity. While MCNP appears to display heterogeneity of forest types within the landscape (Figure 2), the degree to which the forest types vary does not seem to have a strong influence on gene flow.

Although the IBR model was not the "best" model, the resistance values assigned to landcover types matched our *a priori* predictions. For example, the wet deciduous forest vegetation type was determined to be the least resistant to gene flow. The coniferous forest and dry deciduous forest vegetation types were the second and third least resistant vegetation types, respectively. These results are consistent with the findings of Martin (2013) and Burgess & Garrick (2020), and it makes sense for wetter microclimates to offer the least resistance to movement because of the physiology of amphibians. Maintaining proper water balance is difficult in dry environments, and wet forests would provide a more suitable microhabitat throughout the year. Dryer forests might not restrict movements during wet periods but could restrict movements during certain times of year. Thus, there will be fewer individuals in these areas, which will decrease the potential for gene flow to occur.

The most resistant vegetation type according to the model was the water, followed by the human influence. We can infer based on the significant disparity between the water resistance value (880.7) and the human influence resistance value (10.2) that the river presents a considerable obstacle to gene flow. The Green River acting as a barrier is consistent with what was expected, as it would be difficult for the terrestrial individuals to traverse this aquatic habitat. The effect of the river barrier was also supported by the pattern observed in the heatmap in Figure 4, as the level of genetic variation was greater

for pond pairs separated by the river than for ponds on the same side of the river. The matrix displays a discernable pattern of higher F_{ST} values between ponds north of the river and ponds south of the river. This pattern of F_{ST} variation suggests that sites separated by the river potentially poses a significant barrier to gene flow. Similarly, Yamane & Nishida (2010) found that rivers are significant barriers to salamander gene flow and concluded that even a small lowland river could present a landscape barrier to gene flow for the clouded salamander (*Hynobius nebulosus*). The Green River likely does not present an absolute barrier to movements because the current is low in the summer when juvenile marbled salamanders are dispersing, but predation pressure is likely higher in the water than on land. Further analysis should focus on the effect of the river specifically, rather than in the context of all habitat heterogeneity categories (i.e., forest).

It was also expected that the human influence vegetation type would be more resistant to gene flow than the forest types, and the results reflected this to be so. Aside from the bordering areas that likely did not factor into the analysis, the parts of the park considered to be "human influence" include paved two-lane roads, and the visitor center (including parking lots). It is likely that the visitor center and parking lots affect dispersal because an asphalt substrate is a less than exemplary migration medium for marbled salamanders, or that chemical contamination present in the parking lots affect individuals in a negative way. In addition, human-built curbs and fences could also potentially impede salamanders from dispersing from one pond to another, as well as direct mortality from, motor vehicle traffic.

Strong effects of human-modified landscapes have been found previously. Apodaca et al. (2012) found that patterns of gene flow for the red-hills salamander

(*Phaeognathus hubrichti*) have been altered substantially as a result of habitat modification by humans. Reduced migration, increased population bottlenecks, and high levels of inbreeding are each cited as directly caused by human interference. Bartoszek & Greenwald (2009) cited railroad tracks as an agent of habitat fragmentation for two populations of marbled salamanders inhabiting southwestern Ohio. Titus et al. (2014) asserted that further fragmentation of the remaining habitat for their study species, the Eastern tiger salamander, will potentially restrict dispersal among breeding ponds, cause the erosion of genetic diversity, and exacerbate already high levels of inbreeding. Overall, though, the human influence category does not represent a large proportion of the MCNP landscape.

Conclusion

Populations of marbled salamanders at MCNP are genetically structured based on the distance between breeding ponds. The composition of landscape features intervening the breeding sites was not shown to be strongly influencing patterns of gene flow on this landscape. While the IBD model was sufficient to explain the genetic variation on this landscape the IBR model suggests that wet deciduous forest is the least resistant to dispersal and the Green River provides the most resistance to movement. These findings will hopefully aid in continuing to shed light on and developing effective conservation strategies for amphibians.

REFERENCES

Anderson SJ, Kierepka EM, Swihart RK, Latch EK, Rhodes Jr. OE. 2015. Assessing the permeability of landscape features to animal movement: using genetic structure to infer functional connectivity. PLoS ONE 10:e011750.

Apodaca JJ, Rissler LJ, Godwin JC. 2012. Population structure and gene flow in a heavily disturbed habitat: implications for the management of the imperiled Red Hills salamander (*Phaeognathus hubrichti*). Conservation Genetics 13:913-923.

Arntzen JW, Abrahams C, Meilink WRM, Iosif R, Zuiderwijk A. 2017. Amphibian decline, pond loss and reduced population connectivity under agricultural intensification over a 38 year period. Biodiversity and Conservation 26:1411-1430.

Austin, H. 2018. Landscape genetics of *Ambystoma opacum* in Mammoth Cave National Park. Western Kentucky University.

Barbour RW. 1971. Amphibians and Reptiles of Kentucky. Lexington (KY): University Press of Kentucky.

Bartoszek J & Greenwald KR. 2009. A population divided: railroad tracks as barriers to gene flow in an isolated population of marbled salamanders (*Ambystoma opacum*). Herpetological Conservation and Biology 4:191-197.

Burgess SM & Garrick RC. 2020. Regional replication of landscape genetics analyses of the Mississippi slimy salamander, *Plethodon Mississippi*. Landscape Ecology 35:337-351.

Cayuela H, Lemaitre JF, Bonnaire E, Pichenot J, Schmidt BR. 2020. Population position along the fast–slow life history continuum predicts intraspecific variation in actuarial senescence. Journal of Animal Ecology. 89:1069-1079.

Clarke RT, Rothery P, Raybould AF. 2002. Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. Journal of Agricultural, Biological, and Environmental Statistics 7:361.

Cole EM, North MP. 2014. Environmental influences on amphibian assemblages across subalpine wet meadows in the Klamath Mountains, California. Herpetologica 70:135- 148.

Collins J, Storfer A. 2003. Global amphibian declines: sorting the hypotheses. Diversity and Distributions 9:89-98.

Emaresi G, Pellet J, Dubey S, Hirzel AH, Fumagalli L. 2011. Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. Conservation Genetics 12:41-50.

Galpern P, Manseau M, Wilson P. 2012. Grains of connectivity: analysis at multiple spatial scales in landscape genetics. Molecular Ecology 21:3996-4009.

Hagerty BE, Nussear KE, Esque TC, Tracy CR. 2011. Making molehills out of mountains: landscape genetics of the Mojave desert tortoise. Landscape Ecology 26:267-280.

Hebbar P, Ravikanth G, Aravind NA. 2019. A review on the conservation genetic studies of Indian amphibians and their implications on developing strategies for conservation. Journal of Genetics 98:114.

Horton E, Kemp A. 2017. State salamander of North Carolina: marbled salamander. NC Government & Heritage Library. URL https://ncpedia.org/symbols/salamander.

Latch EK, Boarman WI, Walde A, Fleischer RC. 2011. Fine-scale analysis reveals cryptic landscape genetic structure in desert tortoises. PLoS ONE 6:e27794.

Manel S, Holderegger R. 2013. Ten years of landscape genetics. Trends in Ecology & Evolution 28:614-621.

Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology and Evolution 18:157-206.

Martin, JC. 2013. Landscape genetics of the marbled salamander (*Ambystoma opacum*) at Mammoth Cave National Park. Western Kentucky University.

McRae BH. 2006. Isolation by resistance. Evolution 60:1551-1561.

Mims MC, Hartfield Kirk EE, Lytle DA, Olden JD. 2018. Traits-based approaches support the conservation relevance of landscape genetics. Conservation Genetics 19:17-26.

National Park Service. 2018. URL https://www.nps.gov/maca/learn/nature/forests.htm

Noel S, Ouellet M, Galois P, Lapointe FJ. 2007. Impact of urban fragmentation on the genetic structure of the eastern red-backed salamander. Conservation Genetics 8:599–606.

Nunziata SO, Scott DE, Jones KL, Hagen C, Lance SL. 2011. Twelve novel microsatellite markers for the marbled salamander, *Ambystoma opacum*. Conservation Genetics Resources 3:773-775.

O'Brien D, Manseau M, Fall A, Fortin M-J. 2006. Testing the importance of spatial configuration of winter habitat for woodland caribou: An application of graph theory. Biological Conservation 130:70-83.

Peterman WE. 2014. ResistanceGA: an R package for the optimization of resistance surfaces using genetic algorithms. Methods in Ecology and Evolution 9:1638-1647.

Smith, RK, Meredith H, and Sutherland WJ. 2018. Amphibian conservation. Pp 9-66 in What Works in Conservation, edited by Smith RK, Sutherland WJ, Dicks LV, Ockendon N, and Petrovan SO. Cambridge, UK: Open Book Publishers.

Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K. 2010. Use of resistance surfaces for landscape genetic studies: considerations for parametrization and analysis. Molecular Ecology 19:3576-3591.

Storfer A, Eastman JM, Spear SF. 2009. Modern Molecular Methods for Amphibian Conservation. BioScience 59:559-571.

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP. 2007. Putting the 'landscape' in landscape genetics. Heredity 98:128-142.

Storfer A. 2003. Amphibian declines: future directions. Diversity and Distributions 9:151-163.

Stuart SN, Chanson JS, Cox NA, Young BE, Rodrigues ASL, Fischman DL, Waller RW. 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306:1783-1786.

Sunny A, Monroy-Vilchis O, Reyna-Valencia C, Zarco-González MM. 2014. Microhabitat types promote the genetic structure of a micro-endemic and critically endangered mole salamander (*Ambystoma leorae*) of Central Mexico. PLoS ONE 9:e103595.

Taylor BE, Scott DE. 1997. Effects of Larval Density Dependence on Population Dynamics of *Ambystoma opacum*. Herpetologica 53:132-145.

Titus VR, Bell RC, Becker CG, Zamudio KR. 2014. Connectivity and gene flow among Eastern Salamander (*Ambystoma tigrinum*) populations in highly modified anthropogenic landscapes. Conservation Genetics 15:1447-1462.

United States Geological Survey. 2011. Vegetation Map for Mammoth Cave National Park. URL: http://www.usgs.gov/core_science_systems/csas/vip/parks/maca.html.

Wang IJ, Savage WK, Shaffer HB. 2009. Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (*Ambystoma californiense*). Molecular Ecology 18:1365-1374.

Wilbur HM, Semlitsch RD. 1990. Ecological consequences of tail injury in *Rana* tadpoles. Copeia 1:18-24.

Yamane A & Nishida S. 2010. Fine-scale spatial genetic structure and genetic diversity among clouded salamander (*Hynobius nebulosus*) populations. Current Herpetology 29:79-90.

Figure 1. Example of marbled salamander breeding habitat.

Figure 2. Plot of the GIS input data as seen in Austin (2018). Orange represents areas of human influence, blue represents water, light green represents the dry deciduous forest, medium green represents the wet deciduous forest, and dark green represents the coniferous forest. Each of the white points represents a pond that was a sample site for *Ambystoma opacum*.

Figure 3. Scatterplot using the natural logarithm transformation of the distance as the explanatory variable and F_{ST} value as the response variable. The effect was found to be significant at $\alpha=0.05$ with a p-value of <0.001.

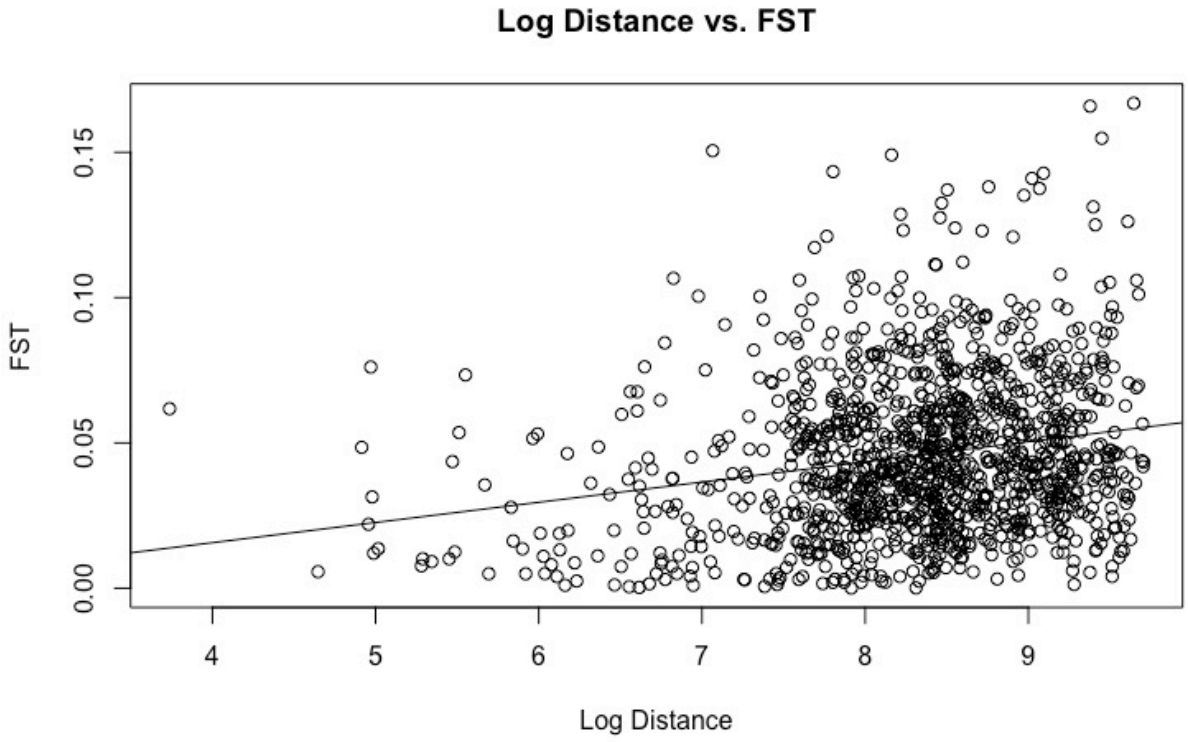

Figure 4. Pairwise geographic and genetic distance heatmap. Rows and columns represent each sampled pond. Pairwise genetic distance (F_{ST} values) are colored blue and located in the upper triangular of the matrix and pairwise Euclidean distances (m) are colored red and located in the lower triangular of the matrix. The top-most and right-most rows and columns correspond to sample sites north of the river.

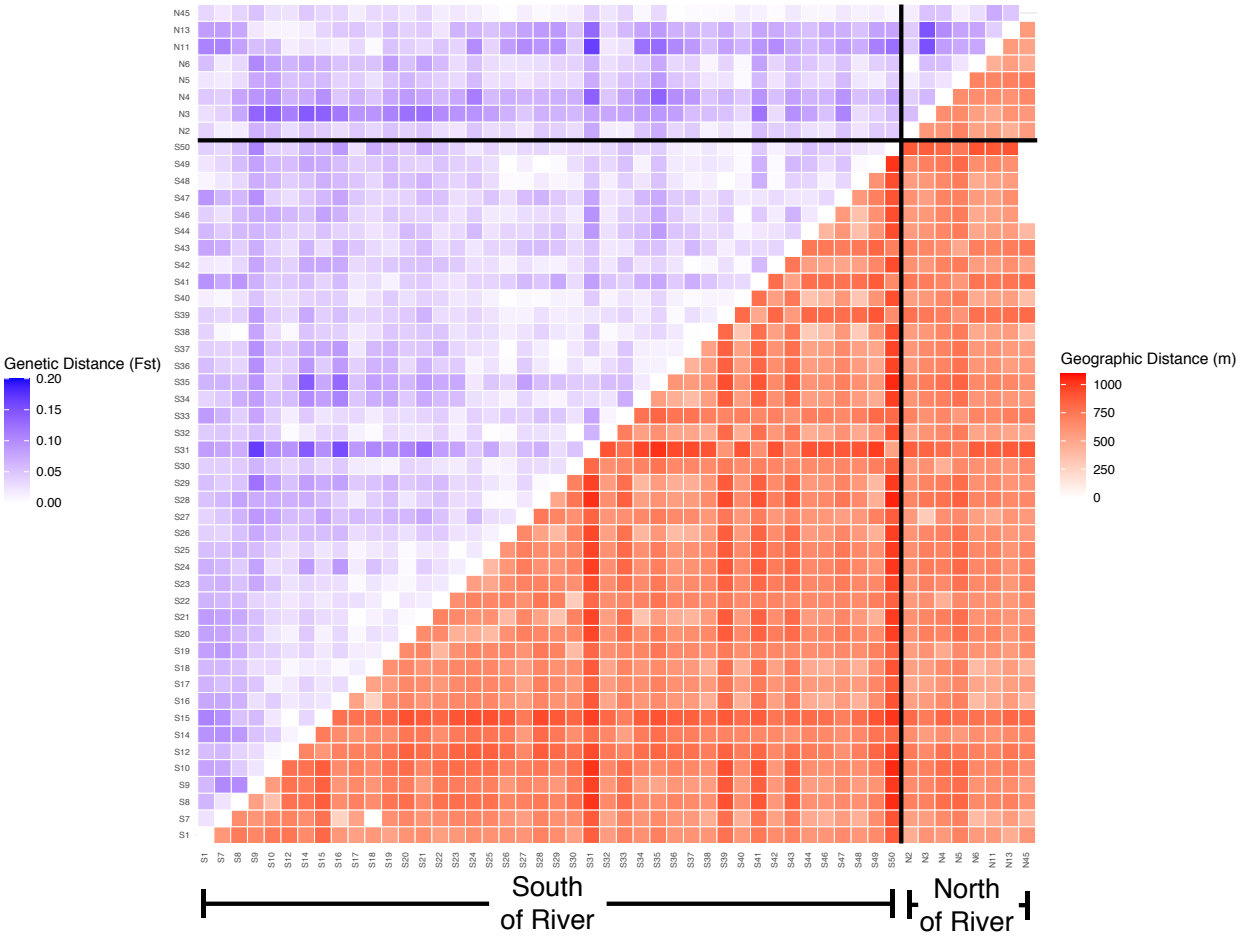

Table 1. Descriptions of each habitat type as seen in Martin (2013).

Habitat Category	*List of Included Habitats*
Wet Deciduous Forest	Successional Tuliptree Forest(Acidic Type)
	Successional Black Walnut Forest
	Beech - Maple Unglaciated Forest
	Interior Low Plateau Mesic Sugar Maple - Hickory Forest
	Successional Sweetgum Floodplain Forest
	Sycamore - Silver Maple Calcareous Floodplain Forest
	Rich Levee Mixed Hardwood Bottomland Forest
	Southeastern Successional Black Cherry Forest
	Successional Tuliptree Forest(Circumneutral Type)
	Rich Appalachian Red Oak - Sugar Maple Forest
	CentralInterior Beech - White Oak Forest
	Shumard Oak - Chinquapin Oak Mesic Limestone Forest
	Pin Oak Mixed Hardwood Depression Forest
	Sinkhole Pond Marsh
	Southern Cattail Marsh
	Buttonbush Sinkhole Pond Swamp
Dry Deciduous Forest	Interior Low Plateau Chestnut Oak - Mixed Oak Forest
	Interior Dry-Mesic White Oak - Hickory Forest
	Chinquapin Oak Unglaciated Bluff Woodland
	Western Highland Rim Post Oak Barrens
	White Oak – Mixed Oak Dry-Mesic Alkaline Forest
	Nashville Basin Shingle Oak - Shumard Oak - Chinquapin Oak Forest
	Southern Red Oak Flatwoods Forest
	Southern Red Oak - Mixed Oak Forest
	Interior Low Plateau Chestnut Oak Forest
Coniferous Forest	Eastern Red-cedar Successional Forest
	Early-Successional Shortleaf Pine Forest
	Appalachian Low-Elevation Mixed Pine/Hillside Blueberry Forest
	East Central Hemlock Hardwood Forest
	Virginia Pine Successional Forest
	Virginia Pine - Red-cedar Successional Forest
Human Influence	Water-Willow Rock Bar and Shore
	Highland Rim Limestone Cliff/Talus Seep
	Rock
	Soil
	Agriculture
	Lawn
	Power Line Easement
	Building
	Commercial
	Human Influence
	Parking Lot
	Road
	Residential

	Successional Broomsedge Vegetation
	Cultivated Meadow
	Blackberry - Greenbrier Successional Shrubland Thicket
Water	Water

Table 2. Description of the three models considered. FST represents the pairwise genetic distances between the ponds; PPID is a random effect term that identifies each unique pairwise pond combination; GEO is a fixed-effect term representing the pairwise geographic distance between the ponds; LAND is a fixed-effect term representing the land cover categories.

Name	**Model**
Null Model	FST ~ 1 + (PPID)
Distance Model (Isolation-by-Distance)	FST ~ GEO + (PPID)
Vegetation Model (Isolation-by-Resistance)	FST ~ LAND + (PPID)

Table 3. Latitude and longitude coordinates for each sample site.

Latitude (X)	Longitude (Y)
37.15324	-86.108440
37.15800	-86.100000
37.18200	-86.092000
37.16362	-86.134970
37.20574	-86.139200
37.16560	-86.084560
37.16600	-86.080000
37.16300	-86.042000
37.15222	-86.053360
37.16489	-86.040740
37.17234	-86.086690
37.20800	-86.052000
37.15093	-86.099041
37.19800	-86.111000
37.23225	-86.057710
37.16471	-86.080190
37.16861	-86.098640
37.16565	-86.080180
37.15972	-86.123340
37.12800	-86.105000
37.13867	-86.070580
37.15900	-86.129000
37.13040	-86.111900
37.12600	-86.096000
37.12700	-86.101000
37.13808	-86.075940
37.18083	-86.092468
37.12580	-86.071570
37.13649	-86.071170
37.16000	-86.128000
37.21474	-86.208940
37.15300	-86.083000
37.21500	-86.113000
37.14099	-86.070530
37.14104	-86.050760
37.14000	-86.080000
37.14345	-86.073260
37.15910	-86.074400
37.21480	-86.174800
37.15880	-86.076600
37.20440	-86.170194
37.17266	-86.062460
37.20150	-86.151800
37.15780	-86.074600
37.16170	-86.075200
37.16196	-86.073010
37.18326	-86.069610
37.16040	-86.074900
37.13251	-86.072190
37.20247	-86.219000

Table 4. Results of Mantel and Partial Mantel Tests. The 'Mantel R' value depicts the correlation between the compared matrices. 'FST' and 'GEO' represent the pairwise genetic and geographic distance matrices, respectively. 'RES' is the third matrix of resistance values used for the partial mantel test. The 'P' value for the correlation was determined by 10,000 resampling iterations of the data. The alpha value is 0.05.

	Mantel R	P
FST x GEO	0.1962	0.0073
GEN x GEO | RES	0.1979	0.0069

Table 5. Model comparisons across three resistance surface resolutions (50m, 75m, & 100m)

	Model	df	AIC	AICc	ΔAICc
A. 50m	Null	1	-5957.1	-5961.1	36.4
	Distance (IBD)	2	-5993.8	-5997.5	0
	Vegetation (IBR)	6	-5991.8	-5985.8	11.7
B. 75m	Null	1	-5957.1	-5961.1	39.5
	Distance (IBD)	2	-5996.8	-6000.6	0
	Vegetation (IBR)	6	-5995.8	-5989.8	10.8
C. 100m	Null	1	-5957.1	-5961.1	36.7
	Distance (IBD)	2	-5994.1	-5997.8	0
	Vegetation (IBR)	6	-5978.7	-5972.8	25.0

Table 6. Resistance values assigned to each habitat type (50m resolution grain).

Vegetation Category	Associated Resistance Value
Wet Deciduous Forest	1 (least resistant)
Dry Deciduous Forest	4.4
Coniferous Forest	4.0
Human Influence	10.2
Water	880.7 (most resistant)

Appendix 1: LogDistance vs. FST scatterplot.

```
fst<-read.csv(".../fstnormalizedmatrix.csv",header=FALSE)
pondlocs<-read.csv(".../UTMpondlocs.csv")

#FST: Transforming FST values into a dataframe. Will use “value”
column.

fst[upper.tri(fst,diag=TRUE)]<-NA
library(reshape2)
fstdf<-melt(fst)

#Distance: Using pointDistance function to convert x and y
coordinates into a vector. Will use “value” column in dataframe
called distance.

install.packages("raster")
pondlocsxy<-pondlocs[,2:3]
library(raster)
distance<-pointDistance(pondlocsxy,lonlat=FALSE)
distance[upper.tri(distance,diag=TRUE)]<-NA
distance<-melt(distance)

#Plot: Generating a linear model with the FST values as the
response and the natural logarithm of the distance as the
predictor. Plotting the values in a scatterplot and then
embedding a line of best fit for the linear model.

lm(fstdf$value~log(distance$value))
plot(log(distance$value),fstdf$value,
      main="Log Distance vs. FST",
      ylab="FST",xlab="Log Distance")
abline(lm(fstdf$value~log(distance$value)))
summary(lm(fstdf$value~log(distance$value)))
```

Appendix 2: Partial Mantel Test.

```
fst<-read.csv(".../fstnormalizedmatrix.csv",header=FALSE)
dist<-read.csv(".../Distance_commuteDistance_distMat.csv",
                      header=FALSE)
resist<-read.csv(".../veii50_commuteDistance_distMat.csv",
                      header=FALSE)

resist1<-resist/dist

fst<-as.matrix(fst)
is.matrix(fst)
dist<-as.matrix(dist)
is.matrix(dist)
resist1<-as.matrix(resist1)
is.matrix(resist1)

resist1[is.nan(resist1)]=0

library(ncf)

partial.mantel.test(fst, dist, resist1, resamp=10000,
method="pearson", quiet=FALSE)
```

Appendix 3: ResistanceGA.

```
library(ResistanceGA)

#Import data (FST values and pond coordinates)
fstnorm<read.csv("/.../fstnormalizedmatrix.csv",
                 header=FALSE)
pondlocs<-read.csv("/.../UTMpondlocs.csv")
ponds<-SpatialPoints(pondlocs)

#Plot map
library(raster)
#Importing raster file (50 meter resolution grain)
veg<- raster("/.../vegascii50.asc")
plot(veg)
points(x=pondlocs$X, y=pondlocs$Y,
       bg="blue", pch=21, cex=.8)

#Run Categorical analysis
# Defining GA.inputs: tells where ASCII file is saved, where to
 save the results, what random number seed to use, and how many
 computer cores to use during the computational process
GA.inputs<-GA.prep(ASCII.dir="/.../",
                   Results.dir="/.../",
                   select.trans = NA,
                   seed = 111,
                   parallel = 15)

#Defining gdist.inputs: indicates where samples and response are
 located and the method to be used

gdist.inputs<-gdist.prep(n.Pops = length(ponds),
                         samples = ponds,
                         response = fstnorm,
                         method ='commuteDistance')

#Defining SS_RESULTS.gdist: uses GA.inputs and gdist.inputs
 objects

SS_RESULTS.gdist<-SS_optim(gdist.inputs = gdist.inputs,
                           GA.inputs = GA.inputs)
```