POPULATION GENETICS OF THE GIANT RED CENTIPEDE ALONG THE APPALACHIAN MOUNTAINS

by
Kayla Elayne Ladner

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
December 2017

Approved By

Advisor: Dr. Ryan Garrick

Reader: Dr. Rebecca Symula

Reader: Dr. Sarah Liljegren

© 2017
Kayla Elayne Ladner
ALL RIGHTS RESERVED

ACKNOWLEDGEMENTS

I would like to thank Dr. Ryan Garrick for his guidance in the laboratory and in the classroom. Because of him I have a greater appreciation and understanding for the impacts of biological research in our society. I would also like to thank Dr. Rebecca Symula for encouraging me to participate in research within my major field of study, without whose initial support, I would have missed out on a big part of my undergraduate career. Finally, I would like to thank everyone in the Garrick Lab for being patient with me as I learned how to do independent research and for always celebrating my successes and encouraging me when I was defeated.

ABSTRACT

KAYLA ELAYNE LADNER: Population Genetics of the Giant Red Centipede along the Appalachian Mountains
(Under the direction of Dr. Ryan Garrick)

This study focused on population genetics of the giant red centipede, *Scolopocryptops sexpinousus*, in the Southern Appalachian Mountains. Two genetic markers were used—the mitochondrial COI gene, and the nuclear RNA Polymerase II gene. The goal of this study was to determine how many genetically distinct populations of this invertebrate species exist within the study area, and to understand the spatial distribution of genetic diversity within and among populations. I hypothesized several genetically distinct populations of *S. sexpinosus* would be detected because the complex topographic characteristics and historical climate cycles of the region are thought to limit movement of many forest invertebrates. Methods used in this study included sampling of centipedes within the known range of *S. sexpinosus*, extracting DNA, PCR amplification of target genes, DNA sequence editing and alignment, estimation of phylogenetic tree based on data from the COI gene, and mapping of the spatial distributions of well-supported clades. The results from this study revealed high levels of genetic diversity among populations for the mitochondrial COI genetic marker, and natural genetic clusters based on an estimated phylogenetic tree were found to comprise individuals that occurred in close geographic proximity to one another. The RNA Polymerase II gene was successfully amplified using primers created in this study, and these can be used in future studies of *S. sexpinosus*. The results from this study fit within a larger-scaled project

focused on population genetics of several different species found within the Southern Appalachians region.

TABLE OF CONTENTS

ABSTRACT……………………………………………………………………………….iv

INTRODUCTION……………………………………………………………………….1

METHODS……………………………………………………………………………5

RESULTS……………………………………………………………………………12

DISCUSSION………………………………………………………………………..14

LIST OF REFERENCES……………………………………………………………18

INTRODUCTION

Species of Interest

The giant red centipede, *Scolopocryptops sexpinosus*, is commonly found throughout the Southern Appalachian Mountain ranges (Figure 1). This species is a brightly colored red-orange centipede that is usually 5–7.5 cm long when fully grown (Newton 2011). It can be found under rocks, inside logs, and on loose tree bark, and the moist climate of the Southern Appalachian region provides large areas of suitable habitat for *S. sexpinosus* (Highlands Biological Station 2017). The genus name *Scolopocryptops* means "dead eyes", which indicates that *S. sexpinosus* is blind. Because these centipedes spend the majority of their time on the forest floor, there is no need for them to spend energy on the sense of sight which allows their blindness to carry little or no cost. This species of centipede has 23 pairs of legs, having one pair of legs on each body segment, and the first pair of legs evolved to be venomous fangs (Newton 2011). The basic biology of the red centipede is poorly known. Therefore, the present research could provide useful data to forest conservationists for identifying the habitat needs of this species, and the new information may also help make plans for sustaining Southern Appalachian forest invertebrate communities as a whole.

Geographic Setting

The Appalachian Mountains are a series of mountain ranges in the Eastern United States, and they lie within the known geographic range of *S. sexpinosus* (Figure 1, left). The Appalachian Mountains are some of the oldest mountain ranges in the world, with

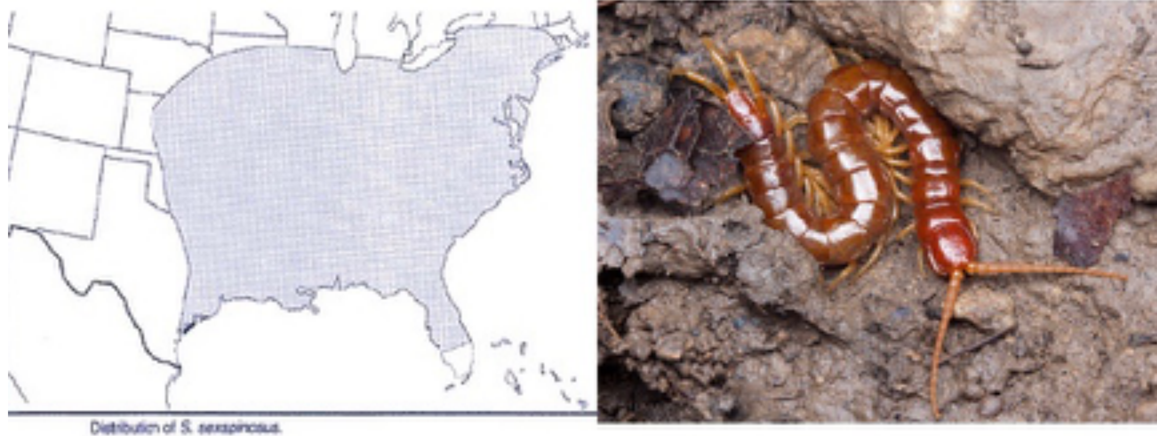

Figure 1: Map of the Eastern United States with state borders shown. Grey shading indicates the known geographic distribution (left) of *Scolopocryptops sexpinosus* (right).

their origin of formation dating back to the middle of the Ordovician Period (about 495-440 million years ago; New World Encyclopedia 2015). The Appalachian Mountain ranges extend for 2,400 km along an approximately north-south axis, spanning from central Alabama into Canada. For this study, the portion of the geographic range of *S. sexpinosus* that lies within the Southern Appalachian Mountains was the focus. These mountains contain the highest point in the United States east of the Mississippi River (i.e., Mt. Mitchell in North Carolina), and although they have an average elevation of 900 meters, there is considerable variation in peak heights over relatively short distances. In addition, the geographical barriers created by high elevation expansive plateaus, and rivers, allow for increased biological diversity. These geographic characteristics may limit gene flow among populations in low mobility taxa such as *S. sexpinosus* (Garrick 2011). The Appalachian Mountains are well known for their beauty, and they are studied for their ecological diversity by many scientists (New World Encyclopedia 2015). In particular, Southern Appalachia is recognized for having the most extensive broad-leaved

deciduous forests in the world and is home to over 140 species of trees (Dykeman 2017a). The diverse climate and precipitation that occurs along these ranges fosters high biodiversity in flora and fauna, and this also allows different species to find suitable habitats along the Appalachian Mountain range (Dykeman 2017b). The Southern Appalachians in particular are also home to different suitable habitats, so this area is able to sustain a larger variety of species (Thomas & Hedin 2008). Nearly 10,000 species are known to exist in Southern Appalachia, and many of these species are endemic to the region (Highlands Biological Station 2017). Increased biodiversity can be seen at every level of organization including genetic diversity within populations, which makes these mountain ranges an important region for biological study.

Glacial Maxima

Unlike the Northern Appalachians and other places in the United States, the Southern Appalachians were never covered by an ice sheet during the last glacial maximum or earlier (Pickering et al. 2002). This means that animals and plants found in Southern Appalachia were not completely wiped out, but instead, many of these species retreated into areas of refuge allowing historic genetic diversity to be maintained. Consequently, many species in this region have long histories of occupancy, so as the ice sheets retreated, recolonization was able to occur rapidly. The glacial cycles of the mid-to-late Pleistocene caused continual range contractions and expansions of species within the region, and these paleoclimatic events are reflected in their population genetic structures (Soltis et al. 2006; Walker et al. 2009). The effects of the last glacial period on

temperature and rainfall contributed greatly to the changes in range of habitats found in the region, and patterns of diversity along these mountains can be attributed to the heterogeneous assemblage of these habitats.

Genetic Markers

In this study, the mitochondrial cytochrome oxidase I (COI) gene was targeted. The decision to choose a mitochondrial gene was made based on the advantages of mitochondrial sequences over nuclear sequences such as lack of introns, limited exposure to recombination and haploid mode of inheritance (Herbert et al. 2003). The universality of previously developed Polymerase Chain Reaction (PCR) primers for this gene is another great advantage. The COI gene is also known to be useful when building phylogenies because its evolution is rapid enough to allow the discrimination of closely related species, as well as phylogeographic groups within a single species (Herbert et al. 2003). In addition to the COI gene, in this study primers were developed for the nuclear RNA Polymerase II gene. These primers were tested in hope of successful gene amplification, and for use in future studies.

METHODS

Sample Collection

Individuals of *S. sexpinosus* were collected from various sites along the Southern Appalachian Mountain range (Table 1; Figure 2), and then stored in 95% ethanol.

Table 1: Sample collection sites including the GPS coordinates, site name, and number of individuals collected from each site. The abbreviations NF, NP, and SP denote National Forest, National Park, and State Park, respectively.

Site Name	GPS Coordinates		Site Description	# of Individuals
A20	34.87367	-84.56690	Chattahoochee NF	2
A22	34.68311	-84.25093	Chattahoochee NF	1
A28	35.60975	-83.44732	Smoky Mtn NP	1
A30	35.65682	-83.51849	Smoky Mtn NP	1
A37	35.77140	-83.21343	Smoky Mtn NP	1
A38	35.61933	-83.66993	Smoky Mtn NP	1
A40	34.75931	-84.69140	Chattahoochee NF	1
A51	36.73672	-81.43256	Jefferson NF	1
A60	38.29123	-78.64308	Shenandoah NP	1
A63	38.76103	-78.28269	Shenandoah NP	4
A64	38.62592	-78.34060	Shenandoah NP	3
A65	38.55696	-78.37899	Shenandoah NP	1
A90	35.27122	-83.68566	Nantahala NF	2
A91	35.25083	-83.64008	Nantahala NF	1
A95	35.37037	-82.79151	Pisgah NF	1
A98	35.74876	-82.33347	Pisgah NF	7
A101	38.56375	-79.69815	Monongahela NF	6
A102	38.88536	-79.46547	Monongahela NF	4
A103	39.09540	-79.43826	Monongahela NF	3
A106	38.82374	-79.38618	Monongahela NF	1
A108	38.72694	-79.48494	Monongahela NF	5
A109	38.71836	-79.51450	Monongahela NF	4
A110	38.71320	-79.55144	Monongahela NF	3
A113	38.74870	-79.69157	Monongahela NF	4

Table 1 Continued

Site Name	GPS Coordinates		Site Description	# of Individuals
A114	38.79810	-79.54447	Monongahela NF	4
A115	38.47525	-79.69947	Monongahela NF	1
A117	33.47105	-85.80658	Talladega NF	1
A127	33.33344	-86.02572	Talladega NF	1
A128	34.06701	-87.32520	Bankhead NF	1
A141	34.86200	-83.10755	Oconee SP	4
A154	34.75933	-84.69117	Chattahoochee NF	1
A155	34.77750	-84.32607	Chattahoochee NF	1
A156	35.52144	-83.30981	Smoky Mtn NP	1
O03	35.87976	-84.88071	Black Mtn Rd.	1
O14	37.28419	-80.46925	Pandapas Pond	1

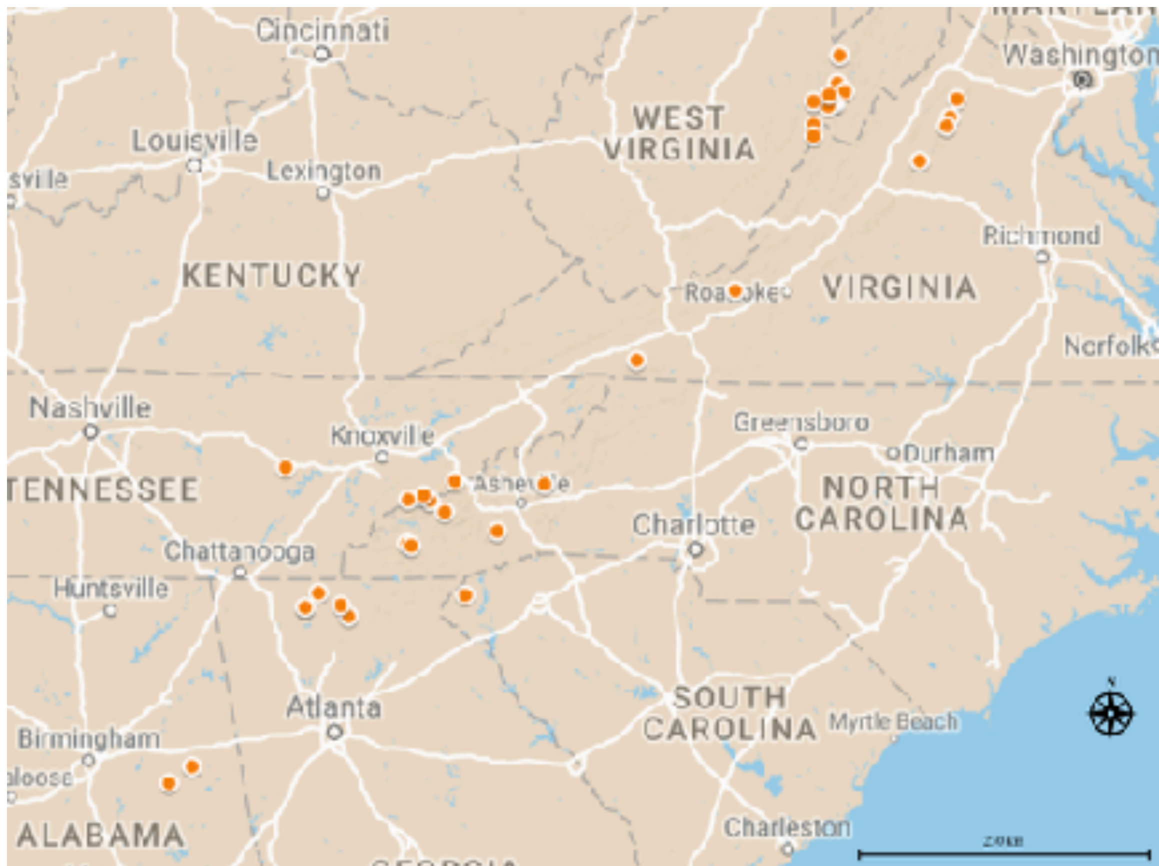

Figure 2: Map of the Eastern United States showing sample sites, as represented by orange dots.

DNA Extraction

Several legs from each *S. sexpinosus* individual were removed and crushed to extract DNA using a Qiagen DNeasy Blood and Tissue Kit, following the manufacturer's protocol. Modifications to this protocol included the following: (1) before starting, legs that had been removed from each centipede were air dried for at least 45 minutes on a kimwipe, under a petri dish to allow the ethanol that had been used for specimen storage to completely evaporate; (2) dried legs were manually crushed with a pestle to break up as much tissue as possible; (3) incubation in lysis buffer was always allowed to continue overnight; (4) the optional step of adding 4 μL RNase A (100 mg/mL) was always performed; (5) an extra 1 minute centrifugation step at 14,000 rpm was conducted immediately prior to elution of DNA in order to further prevent carry-over of ethanol on the spin column membrane; and (6) two elutions of the DNA were performed, each time using 65 μL of AE that had been warmed to 56° C.

Polymer Chain Reaction

PCR was used to amplify portions of the mitochondrial COI and nuclear RNA Polymerase II genes. A master mix was created using the components shown in Table 2. With the exception of PCR primers, the same master mix was used for amplification of both gene regions. The COI forward and reverse primes were LCO and HCO respectively (Folmer et al. 1994; Table 3), and the RNA Polymerase II forward and reverse primers were SsRNP2-F2 and SsRNP-R2 respectively (Garrick & Ladner unpublished; Table 3).

Amplifications of the two gene regions were performed in a BioRad T100 Thermal Cycler, using the conditions shown in Table 4.

Table 2: PCR master mix components with their respective concentrations and volumes that were used in each amplification.

PCR Master Mix Components	Concentration	Volume (µL)
5x PCR Buffer	n/a	3.00
$MgCl_2$	25 mM	1.20
dNTPs	1.25 mM	2.40
Bovine Serum Albumin	10 mg/mL	0.75
Distilled H_2O	n/a	4.50
Forward Primer	10 µM	0.75
Reverse Primer	10 µM	0.75
Go*Taq*. DNA Polymerase	5 µM	0.15

Table 3: Forward and reverse primers for the mitochondrial COI gene, LCO-1490 and HCO-2198, along with the forward and reverse primers for the nuclear RNA Polymerase II gene, SsRNP2-F2 and SsRNP2-R2.

Primer Name	Primer Sequence	Length (bp)
LCO - 1490	5'-GGTCAACAAATCATAAAGATATTGG-3'	25
HCO - 2198	5'-TAAACTTCAGGGTGACCAAAAAATCA-3'	26
SsRNP2-F2	5'-CACATCAACACTCGAATCC-3'	19
SsRNP2-R2	5'-TTGCGGTCAAGTTCGATACGC-3'	21

Table 4: Thermal Cycler conditions when amplifying both the mitochondrial COI and the nuclear (nDNA) RNA Polymerase II genetic markers including temperature in °C, time in minutes, and number of cycles.

Temperature (°C)	Time (min.)	# of Cycles
95	2.0	1
95	0.5	35
50	0.5	
72	1.0	

Table 4 Continued

Temperature (°C)	Time (min.)	# of Cycles
72	2.0	1
12	∞	1

Gel Electrophoresis

PCR products were separated through the process of gel electrophoresis using 1% agarose gel and 1X TBE buffer. A 100 base pair (bp) ladder (~5 µL) was added in the left-most well of each gel for comparison with amplified products from the target gene regions. Each subsequent well was filled with 5 µL of PCR products and 3 µL of loading dye. The last well of each gel only contained the negative control. Each gel was run at 80 to 100 volts for 60 to 90 minutes. The gels were cast and allowed to run using a Thermo Scientific Owl Easycast B2, and the voltage supplier was a Fisher Scientific FB300. A BioRad Gel Doc XR gel documentation system was used to photograph each gel and review amplification results. PCR amplification was used for amplifying both the COI gene and the RNA Polymerase II gene. Successful PCR amplification products were sent away for sequencing at Yale University's DNA Analysis Facility on Science Hill, in New Haven, Connecticut.

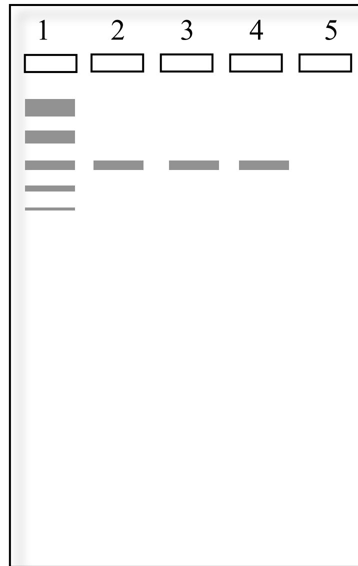

Figure 3: An example gel. Lane 1 represents the 100 base pair ladder, lanes 2-4 represent PCR products, and lane 5 represents the negative control.

DNA Sequence Analysis

DNA sequence analysis was performed only on the mitochondrial COI gene, given that a sufficiently large number of *S. sexpinosus* individuals had been screened using the marker. The software MEGA V6 (Tamura et al. 2013) to build and edit DNA sequence alignments as follows. First, an AB1 file was opened and the overall quality of a sequence was gauged by adjusting the width and height of the peaks displayed. Then the low-quality ends were trimmed, and for the remaining high-quality sequence, four steps were involved in sequence editing, as follows: (1) agree with the automated software's base call, (2) replace a base, (3) mark a base as ambiguous, or (4) insert or delete bases. Once the appropriate edits were made the sequence was saved as an SCF file. Once all sequenced had been edited, the SCF files were imported into MEGA, the

sequences were aligned to one another, and a second round of editing was performed to confirm any rare mutational differences within the alignment.

Phylogenetic Tree Building and Analysis

After editing and aligning mitochondrial COI gene sequences from the sampled individuals, a phylogenetic tree was estimated using MEGA. The tree was built using the maximum likelihood statistical method. The General Time Reversible model was used to account for unequal nucleotide frequencies, and for substitutions that occurred twice in the same location, potentially causing reversals back to the original state. Support for relationships in the estimated phylogeny was tested using the bootstrap method with 1000 replications. Based on the estimated phylogenetic tree, natural groupings of sequences were identified by eye, based on the criterion of identifying well-supported monophyletic groups that included multiple individuals.

Geographic Analysis

After identifying natural genetic clusters from the estimated phylogenetic tree, the geographic distribution of these groups was assessed by plotting GPS coordinates of sequenced individuals using Google Maps.

RESTULTS

Based on the reconstructed phylogenetic tree, there were five distinct, well supported groups, each comprised of multiple individuals (Figure 4). Well supported groups had a node support value of ≥ 75%. Upon finding natural clusters of individuals based on phylogenetic analysis of DNA sequence data, the geographic distribution of these groups was assessed (Figure 5). This spatial distribution was interesting, particularly around the Great Smoky Mountains National Park, which is located on the North Carolina and Tennessee border. There was a noteworthy spatial pattern between two of the groups (Figure 5) where an abrupt east vs. west divide occurs between them. Within the two distinct groups, there are 9 individuals in the light blue group and 12 individuals in the lime green group (Figure 5). These groupings occur in very close spatial proximity to one another, yet there appears to be an abrupt transition from one to the other, with no indication of intermixing. The primers for the RNA Polymerase II gene successfully amplified the target sequence, and these primers will be used for future studies involving *S. sexpinosus* to amplify the gene region for the same individuals that have already sequenced for the mitochondrial COI gene.

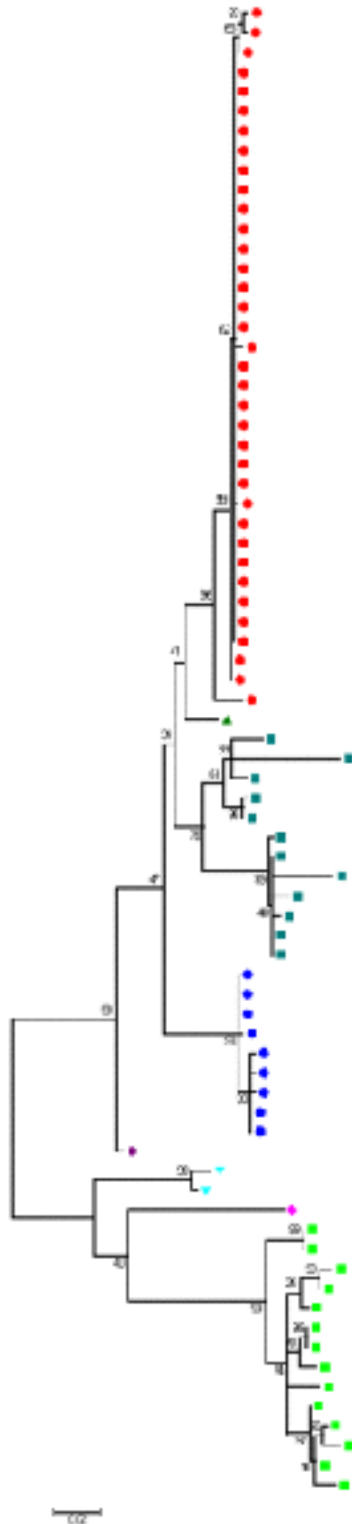

Figure 4: Estimated phylogenetic tree using MEGA V6 (Tamura et al. 2013) with 1000 bootstrap replications using the maximum likelihood statistical method.

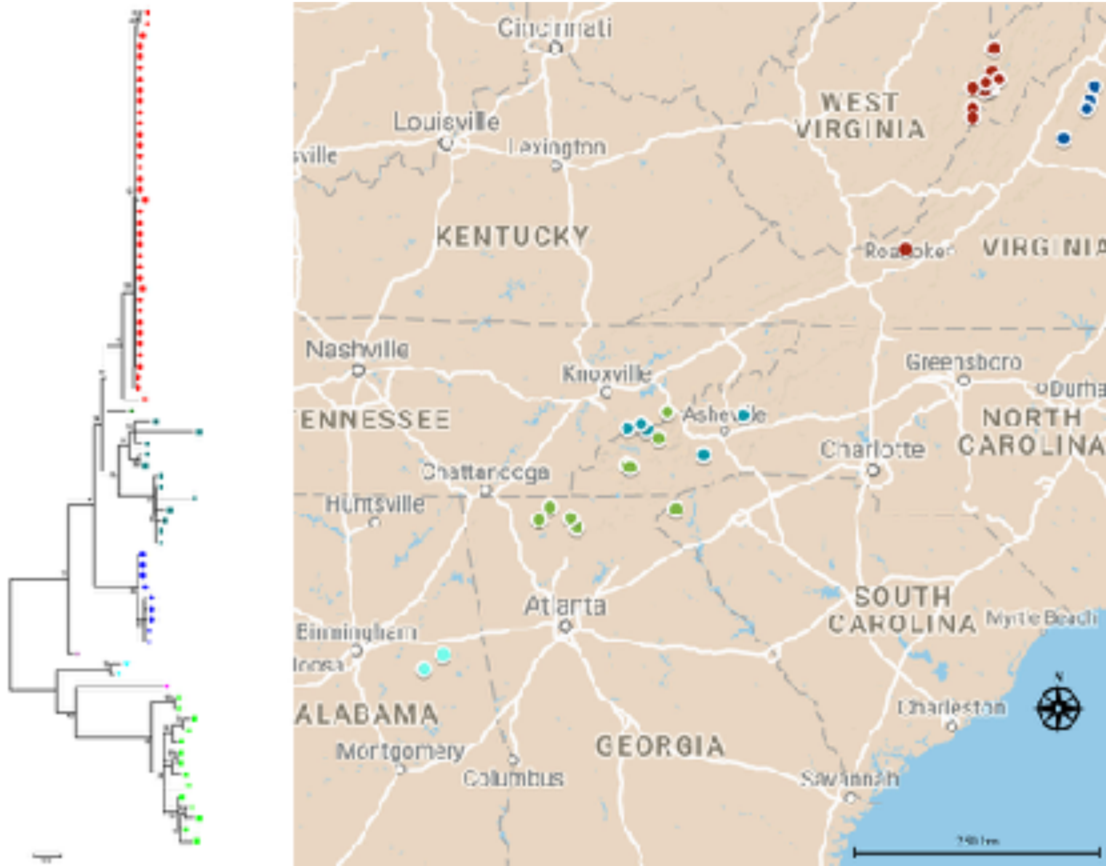

Figure 5: Estimated phylogenetic tree from Figure 4 (left) along with a map of the Eastern United States (right). The dots represent individuals sampled, and the 5 different colors represent the 5 natural groupings from the estimated phylogenetic tree which monophyletic group they belong to from the tree. Nodes on the estimated phylogenetic tree with only one individual were excluded from the map.

DISCUSSION

Importance for Other Studies

The results from this study have many parallels with some other inveterate species found within this region, such as a wood-feeding cockroach, *Cryptocercus punctulatus* (Garrick et al. 2017), and members of the millipede genus *Narceus* (Walker et al. 2009). The study on *C. punctualatus* reported that a region within the central area of study, in the

area of the Great Smoky Mountains, where two genetically distinct clades almost meet (Garrick et al. 2017). This zone shows an abrupt transition from one genetic population to another, and was also seen in *S. sexpinosus*. This particular finding is also reflected by millipedes in the genus *Narceus*. In that study by Walker et al. (2009), there were three distinct clades in close proximity but with minimal overlap, and as with the centipede and cockroach, location of this was along the east-west divide created by the Great Smoky Mountains. The similarities between these three distantly related groups of forest invertebrates should be useful for other scientists studying population genetics in the Southern Appalachian region, as they may may find this geographic distribution trend in their their own studies. The RNA Polymerase II primers developed in the present study did successfully amplify the target gene, and was it was also variable among the small number centipede examined. These primers could be used to determine if this same east-west division in the Smoky Mountains is seen using a bi-parentally inherited nuclear genetic marker.

Limitations

For this study, the genetic data set that was analyzed to determine the number and locations of distinct populations was limited because it includes information from only one gene. If the patterns of variation at this gene are not representative of other parts of the genome, then the genetic groups identified here may not be very genetically distinct. Also, the mitochondrial COI gene is maternally inherited, which gives a female-biased view of population biology for this species. Furthermore, inferences made from the

phylogenetic tree are limited because there is not strong support across all nodes in the tree. This is due to a lack of information in the data for estimating some relationships, and so these tentative inferences may either be further supported or refuted in the future with additional sequence data. There was also a challenge in identifying the target species because of the morphological characteristics. These characteristics of size, shape, and color make it difficult to differentiate *S. sexpinosus* from other *Scolopcyptops* species. A lot of sequences were excluded because they were thought to belong to other species of *Scolopocryptops* centipedes. In the future, centipede samples could be looked at more closely, and expert advice could be sought out to help distinguish *S. sexpinousus* from other species prior to sequencing.

Glacial Impacts on Population Genetics

It is important to remember that current distribution of genetic diversity within *S. sexpinousus* has at least partly resulted from historical climate change, and its associated effects on distributions of forest habitats. Populations in areas of refuge during the last glacial maximum likely served as source populations for recolonization of neighboring areas. The continual change in habitat range due to varying climate characteristics gave rise to the patterns of genetic diversity among populations seen in this study. Identification of the number and locations of genetically distinct populations is the first step in reconstructing the history of refugial isolation and recolonization events, and so this project is an important springboard for follow-up work.

Future Research

Moving forward, the nuclear marker for RNA Polymerase II should be applied to all of the individuals that the mitochondrial marker was used on, and a comparison of the diversity of the genetic markers could be done. Also, more geographic sampling can be done to improve the density of individuals for the study along the Appalachian Mountains specifically in Virginia, North Carolina, and Alabama. Focus can also be directed to the area of the Great Smoky Mountains between North Carolina and Tennessee where two distinct groups are spatially close, and there is potential for gene flow between them.

BIBLIOGRAPHY

Dykeman, W. (2017a) Appalachian Mountains. In: *Encyclopedia Britannica*. Available at: https://www.britannica.com/place/Appalachian-Mountains/Plant-and-animal-life (accessed November 17, 2017).

Dykeman, W. (2017b) Appalachian Mountains. In: *Encyclopedia Britannica*. Available at: https://www.britannica.com/place/Appalachian-Mountains/Geology (accessed November 17, 2017).

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, 294–299.

Garrick, R. (2011) Montane refuges and topographic complexity generate and maintain invertebrate biodiversity: Recurring themes across space and time. *Journal of Insect Conservation*, 15, 469–478.

Garrick, R., Sabree, Z., Jahnes, B., Oliver, J. (2017) Strong spatial-genetic congruence between a wood-feeding cockroach and its bacterial endosymbiont, across a topographically complex landscape. *Journal of Biogeography*, 44, 1500–1511.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, 270, 313–321.

Highlands Biological Station contributors (2017) Biodiversity of the Southern Appalachians. Available at: http://highlandsbiological.org/nature-center/biodiversity-of-the-southern-appalachians/ (accessed November 17, 2017).

New World Encyclopedia contributors (2015) Appalachian Mountains. In: *New World Encyclopedia.* Available at: http://www.newworldencyclopedia.org/p/index.php?title=Appalachian_Mountains&oldid=989897 (accessed November 20, 2017).

Newton, B. (2011) Kentucky Centipedes. In: Kentucky Critter Files. Available at: http://www.uky.edu/Ag/CritterFiles/casefile/relatives/centipedes/centipede.htm (accessed November 17, 2017).

Pickering, J., Kays R., Meier, A., Andrew, S., Yatskievych, K. (2002) The Appalachians. In: Wilderness -- Earth's Last Wild Places (eds. P. R. Gil, R. A. Mittermeier, C. G. Mittermeier, J. Pilgrim, G. Fonseca, W. R. Konstant and T. Brooks). Pages 458–467, Conservational International. Available at: www.discoverlife.org

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S., Soltis, P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15, 4261–4293.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729.

Thomas, S.M., Hedin, M. (2008) Multigenic phylogeographic divergence in the paleoendemic southern Appalachian opilionid *Fumontana deprehendor* Shear (Opiliones, Laniatores, Triaenonychidae). *Molecular Phylogenetics and Evolution*, 46, 645–658.

Walker, M.J., Stockman, A.K., Marek, P.E., Bond, J.E. (2009) Pleistocene glacial refugia across the Appalachian Mountains and coastal plain in the millipede genus *Narceus*: Evidence from population genetic, phylogeographic, and paleoclimatic data. *BMC Evolutionary Biology*, 9, 25.