ABSTRACT

HESS, BENJAMIN MATTHEW. Distribution and Taxonomic Status of the Short-tailed Shrew (Genus *Blarina*) in North Carolina. (Under the direction of Roger A. Powell).

Identifying and describing species is a key component in the study of biology. Species concepts are used to define species based on a specific set of criteria. The phenetic or morphological species concept uses measured characters to create clusters that represent the separation of species or subspecies. The genetic species concept investigates differences in genetic material to separate species or subspecies with a common gene pool. I examined the taxonomic status of the short-tailed shrew (Genus *Blarina*) in North Carolina using both the morphological and genetic species concepts.

Short-tailed shrews have gone through several taxonomic revisions in North Carolina and throughout their ranges in North America. In North Carolina, the northern short-tailed shrew (*Blarina brevicauda*) with multiple subspecies and the Dismal Swamp short-tailed shrew (*B. telmalestes*) were recognized in the state. Currently, *B. brevicauda knoxjonesi* and *B. b. talpoides* are recognized as the northern short-tailed shrew subspecics, and *B. carolinensis carolinensis* is recognized as the northern-most subspecies of the southern short-tailed shrew in North Carolina. *B. b. talpoides* populations are found in the mountains and the northeast part of North Carolina, where these populations are connected with *B. b. talpoides* populations in Virginia. *B. b. knoxjonesi* was identified as a disjunct population in the southeast part of North Carolina.

Morphological analyses done with the measurements on skulls and mandibles from museum specimens in North Carolina confirmed that *B. brevicauda* and *B. carolinensis* are distinct species. *B. b. knoxjonesi* and *B. b. talpoides* have substantial overlap in multivariate analyses (principal components analysis, linear discriminant function analysis and classification tree

analysis), and exhibit core-edge population size variation in eastern North Carolina based on the classification tree analysis with localities. Genetic analyses done with the cytochrome-b gene extracted from liver tissue from museum specimens also confirm *B. brevicauda* and *B. carolinensis* as distinct species. A small samples size (n=1) for *B. b. knoxjonesi* does not provide substantial evidence for subspecies status, but this sample grouped with all samples of *B. b. talpoides* in the phylogram created with the cytochrome-b sequence data in Bayesian and maximum likelihood analyses. Both morphological and genetic analyses showed little variation within *B. b. carolinensis* confirming one subspecies is in North Carolina.

My results confirm that two species of the short-tailed shrew (genus *Blarina*) exist in North Carolina. The subspecies status for the northern short-tailed shrew is questionable based on morphological analyses, but requires more research with genetic analyses. More *B. b. knoxjonesi* specimens should be analyzed with additional mitochondrial DNA genes sequenced to verify if the status based on genetic analyses is in congruence with the morphological analyses.

Distribution and Taxonomic Status of the Short-tailed Shrew (Genus *Blarina*) in North Carolina

by

Benjamin Matthew Hess

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
Requirements for the Degree of
Master of Science

ZOOLOGY

Raleigh, North Carolina

2016

APPROVED BY:

Dr. Jaime A. Collazo	Dr. Kevin Gross Minor Representative
Dr. Bryan L. Stuart	Dr. W. David Webster
Dr. Roger A. Powell Chair of Advisory Committee	

DEDICATION

To my family: Jonathan, Georgia, Missy and Tim. This dissertation would not have been possible without all the early times shared in the woods to drive my curiosity. To the late Dr. Gordon L. Kirkland Jr., who took a chance with me and introduced me to museums after my initial field work with Jim Hart – thank you Shippensburg University Vertebrate Museum ("There's no business like shrew business!"). And to my wife, Shelora, for providing continual love and support during this journey.

BIOGRAPHY

I, Benjamin Matthew Hess, was born and grew up in York, Pennsylvania. I began camping and hiking when I was young, which fostered a great interest in the natural world. While attending Shippensburg University in Pennsylvania, I started my education as a math major with the prospects of becoming a teacher. Knowing that I also wanted to do field work, I ended with a Bachelor of Science (with what is now a designated concentration in Ecology and Environmental Biology) as a major and a minor in mathematics. During this undergraduate work, a summer position as a temporary game farm worker with the Pennsylvania Game Commission led to a position with the Shippensburg University Vertebrate Museum that allowed me the opportunity for extensive field and museum work. I maintained this position until graduation, which sparked a profound interest in the ecology, differences and similarities, and the relationships of mammals.

After completion of my undergraduate degree from Shippensburg in 1999, I took non-degree graduate courses and worked various jobs that made me realize that I truly missed biology. I began taking graduate class again in 2004 with the thought of attaining a master's degree from Shippensburg University. After one year, however, I took on a new job opportunity as the Mammal Collections Manager at the North Carolina State Museum of Natural Sciences in 2005. I started this position with the prospects of continuing my education and after three years, I began my master's degree at North Carolina State University.

Upon graduation with a Master of Science in Zoology, I hope to continue my education with new endeavors. My work and education in North Carolina forged a desire to answer new questions and rekindled my love of teaching. I look forward to my next challenge.

ACKNOWLEDGEMENTS

I would like to first thank my wife Shelora, whose patience and support make this degree more meaningful. Thank you for persevering with me.

I thank my committee chair, Dr. Roger Powell, for taking me on as a prospective student and giving me guidance during this degree. I also thank the remainder of my committee, Dr. Jaime Collazo, Dr. Kevin Gross, Dr. Bryan Stuart and Dr. W. David Webster. A special thank you goes to Dr. W. David Webster for the early conversations about shrews and access to the University of North Carolina at Wilmington (UNCW) to begin my work at the mammal collection there. Additional access to the UNCW Mammal Collection was provided by Dr. Brian Chapman, Dr. Sentiel "Butch" Rommel, Amy Cherry Millis and Sarah Hutchinson-Karfas including many late evenings and weekends.

Thank you to the "unofficial Powell" lab of Roger Powell, Alana Sullivan, Aimee Rockhill, Scott Robertson, Aaron Facka, Rob Sweirs and others for great conversations during our meetings.

Additional work done for specimens and the data associated with the specimens was a group effort at the North Carolina State Museum of Natural Sciences (NCSM – now called the North Carolina Museum of Natural Sciences). Thank you to Lisa Gatens, Bryan McLean, Matt Owen, Bronson Curry, Lydia Fraser, Alexandra Mash, Adrianna Cardinal De Casas, Dawn Keyser, Sam Freeze and others within the Mammal Unit. Early thanks to Jonathan Raine and recent help from M. Ben Norton for GIS mapping of localities at the NCSM. Skull illustrations and layout would not have been possible without the assistance of Liz Bradford and Brenda Wynne.

Funding for genomics research was made possible by the NCSM Research and Collections Grant 2013-2014. Thank you to Dr. Bryan Stuart for my initial genetics training, and Dr. Julie Horvath and Dr. Julie Urban for allowing me access and use of the NCSM Genomics Lab. I would also like to thank Dr. Heather Farrington, Dr. Olivia Evangelista and Russell Wisser for additional assistance and guidance during my lab work and sequence analysis. Lastly, a special thank you to Dr. Julie Urban for the extra help and trouble-shooting to increase my sample size and final analysis help.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

Chapter 1

Introduction: Defining and Designating Species, and the Taxonomic History of the Short-tailed Shrew (Genus *Blarina*) with an Emphasis on North Carolina

Defining and designating species and subspecies should be done with care. Designating a new taxonomic name with few samples can give the impression of differences where they may not exist, and can give implications of a threatened population without seeing the whole picture of the species or subspecies. A larger sample size can more clearly document the variation in morphology, genetics, ecology, development, behavior and other traits. This leads to a better understanding of the variation that informs a taxonomist before categorical naming or separation is determined.

One of the key components of species is barriers to gene flow. If a barrier exists, reproductive isolation combined with differences in habitat can promote variation in morphology, genetics, ecology and behavior, which can send different groups of organisms on their own evolutionary trajectories (Mayr, 1942; 1996). When identifying species and subspecies, a large sample size can distinguish the difference between a gradual change of characteristics suggesting environmental variation versus an abrupt change, suggesting a barrier to gene flow. The splitting of species into subspecies is equally difficult, requiring the identification of localized populations of a species that have distinct habitats, and ecologies and that are reproductively isolated from the main species distribution. To improve this distinction, large sample sizes from a large spatial scale can aid in examining the variation. Identifying species and subspecies is a crucial step for

improving or maintaining the protection of one or more species or the land that the species inhabit.

Biology is the study of life and organisms, and identifying species is a way to group or categorize organisms. Species concepts use different sets of characteristics of organisms to designate species. The biological species concept (Dobzhansky, 1935; Mayr, 1942) identifies interbreeding populations that are reproductively isolated as species. The phenetic or morphological species concept (Beckner 1959; Cronquist, 1978; Smith, 1994; Sokal and Sneath, 1963) groups individuals that share a set of unique physical characteristics. The genetic species concept (Dobzansky, 1950; Mayr, 1969; Simpson, 1943) identifies a species as those individuals with a common gene pool. Contraditions among these and other species concepts have led to discussion about how we define and designate species (De Querioz, 1998; Mayden, 1997; Mishler, 2010).

De Quieroz (2005, 2007) developed a unified concept of species as a lineage of metapopulations, where the traits of divergent lineages vary over time. The different species concepts take different positions along the lineage. In other words, a lineage only has to be separately evolving from other lineages (De Quieroz, 2007). All species have characteristics that describe them, but taxonomists must determine what characteristics delimit species and systematists must determine the evolutionary relationships between species.

As for all organisms, early descriptions of species of mammals were based on variations in external morphology. For mammals, species descriptions often use skull and external morphological characters (often called traits or characteristics). Characters on skulls provide information about a mammal's diet (i.e. tooth type correlated with diet), whether it is a predator or prey (i.e. eye orbit positions differ between predators and prey) and what senses are dominant

(i.e. sizes of nasal openings, auditory bullae and orbits correlated with use of smell, hearing and vision). Characters of the skulls of mammals can exhibit convergent evolution. Therefore, skull characters do not always indicate phylogenetic relatedness.

Closely related cryptic species can not always be identified using skull characters alone. Often, cryptic species can be identified by quantifying genetic differences among individuals. Currently, most mammalian taxonomic revisions must use both morphological and genetic analyses. My research used both morphological and genetic analyses to evaluate the status of short-tailed shrews of the genus *Blarina* in North Carolina. Historically, shrews of the genus *Blarina* were divided into two species (Hall, 1981), but are now classified as four species (Wilson and Reeder, 2005). Many subspecies have been classified and reclassified as well.

Species of shrews and moles are grouped within the order Soricimorpha (Wilson and Reeder, 2005). The systematic relationships within the family Soricidae, the red-toothed shrews, have been relatively well studied, especially within the genera *Sorex* (Diersing, 1980; Foresman and Jensen, 1992; Huggins and Kennedy, 1989; Van Zyll De Jong, 1980; Van Zyll De Jong and Kirkland, 1989) and *Blarina* (Genoways and Choate, 1998).

Shrews in the genus *Blarina* are short-tailed shrews, which have relatively short tails compared to their body lengths. These short-tailed shrews are easily distinguished from the genus *Cryptotis*, which also have a short tail relative to their body length, by the presence of 5 unicuspids and 32 total teeth as compared to *Cryptotis*, which have 4 unicuspids with only 3 visible from a lateral view and 30 total teeth (George et al., 1986; Whitaker, 1974). Shrews of the genus *Blarina* are large shrews living only in North America and have dense short grayish-brown to dark gray dorsal pelage with lighter ventral pelage. Their small ears are concealed by their pelage; they have tiny eyes and pointed snouts. Short-tailed shrews live throughout most of the

central and eastern United States from eastern Texas to Florida and north to southern Saskatchewan and Nova Scotia (George et al., 1986; Hall, 1981; Kurta, 1995; McCay, 2001; Thompson et al., 2011). They are mostly nocturnal and are considered habitat generalists, living in coastal salt marshes, old fields, deciduous and mixed forests and other diverse habitats (Trani et al., 2007; Whitaker and Hamilton, 1998). Short-tailed shrews may be soil specialists because they are often associated with loose damp soil, allowing semi-fossorial movements within habitats.

The original use of *Blarina* is credited to Gray (1838), and *Blarina* was elevated to the genus level by Lesson (1842). Merriam (1895) considered *Blarina* a subgenus of *Sorex* and designated three species: *Sorex* [*B.*] *brevicauda* (Baird, 1858; Say, 1823), *Sorex* [*B.*] *carolinensis* (Bachman, 1837) and *Sorex* [*B.*] *telmalestes*. Hall (1981) designated two species in the genus *Blarina*: *B. brevicauda*, with 15 subspecies, and *B. telmalestes* (swamp or Dismal Swamp short-tailed shrew), without any subspecies. More recent analyses have revealed 3 species: *B. brevicauda* (northern short-tailed shrew – Baird, 1858; Say, 1823), *B. carolinensis* (southern short-tailed shrew – Bachman, 1837), and *B. hylophaga* (Elliot's short-tailed shrew – Elliot, 1899) based on multiple studies (Baumgardner et al., 1992; Garland and Heidt, 1989; Genoways and Choate, 1998; George et al., 1981; George et al., 1986; McCay, 2001; Schmidly and Brown, 1979; Thompson et al., 2011). Wilson and Reeder (2005) designated a fourth species, *B. peninsulae* (Everglades short-tailed shrew – Merriam, 1895) based on the taxonomic work on the short-tailed shrews in Florida (Benedict et al., 2006; Genoways and Choate, 1998; George et al., 1981; Wilson and Ruff, 1999).

Molecular phylogenetic research on *Blarina* supports the monophyly of the genus and suggests that speciation occurred 3.7 to 4.6 mya, before the onset of Pleistocene glaciations

(Brant and Ortí, 2002). Fossil records suggest that the origins of *B. brevicauda* date to the Pliocene (~2 mya), *B. carolinensis* dates to the late Pliocene to early Pleistocene (~1.7 mya) and *B. hylophaga* dates to the late Pleistocene (Brant and Ortí, 2002; Harris, 1998). The earliest specimens of *B. brevicauda* were recovered from fauna in Kansas, and specimens of *B. carolinensis* were first known from fauna in western Florida in the early Pleistocene (Jones et al., 1984). Handley (1971) indicated that the current range of *B. carolinensis* emanated from western Florida. Cave deposits near Savanna, Georgia, contained *Blarina* of two sizes suggesting that ancestral species of *Blarina* could have exhibited sympatry in this region (Hulbert and Pratt, 1998). Recent distributions, however, of *B. brevicauda* and *B. carolinensis* appear parapatric, segregated by both temperature and moisture extremes (Graham and Semken, 1976; Jones et al., 1984).

Lee et al. (1982) created an early distribution map (Figure 1.1) of the genus *Blarina* in North Carolina, showing *B. brevicauda telmalestes* (Merriam, 1895) in the Dismal Swamp region, *B. brevicauda kirtlandi* (Bole and Moulthrop, 1942) at low elevations in the mountains and in the upper piedmont, *B. b. churchi* (Bole and Moulthrop, 1942) at the high mountain elevations, and *B. carolinensis carolinensis* (Bachman, 1837) in the Coastal Plain region. *B. brevicauda telmalestes* was considered smaller than the other northern short-tailed shrew subspecies, and was thought to be isolated within the Dismal Swamp in North Carolina and Virginia (Handley, 1979; Lee et al., 1982; Paul, 1965; Rhoads and Young, 1897). Nonetheless, short-tailed shrew specimens collected from the southeast part of North Carolina were comparable in size to *B. b. telmalestes*, where only *B. carolinensis* was assumed to be found (French, 1981; Webster et al., 1984). Clark et al. (1985) identified specimens from upland areas near the Carolina Bays of

Bladen County as *B. carolinensis*. They presumed that large specimens in areas with wet forest floors to be *B. b. telmalestes*. Thus, the taxonomy of *Blarina* in North Carolina was confused.

Using more *Blarina* specimens from North Carolina, Webster et al. (1985) described the distribution of *B. carolinensis* to be throughout the piedmont and the coastal plain, and the distribution of *B. brevicauda* to be split between the foothills and mountain regions in the west and 2 disjunct populations in the northeastern and southeastern coastal plain (Figure 1.2). In a study of *B. brevicauda* in eastern North Carolina, Webster (1996) described a new subspecies, *B. b. knoxjonesi*, in the southeastern coastal plain. In this research, he defined a separation of the species based on habitat where "*B. carolinensis* occupies relatively dry well-drained uplands", while *B. brevicauda* inhabits areas that retain more moisture. The most recent update on the taxonomy of *B. brevicauda* combined the subspecies *B. b. churchi*, *B. b. kirtlandi* and *B. b. telmalestes* into *B. b. talpoides* (Gapper, 1830), but retained the subspecies *B. b. knoxjonesi* (Figure 1.3; Webster, 1996; Webster et al., 2011).

The goal of my research was to examine the current taxonomy of the short-tailed shrews of the genus *Blarina* in North Carolina. I used phenotypic and genetic data from *Blarina* specimens collected in North Carolina. From museum specimens, I analyzed skull and mandible measurements and analyzed mtDNA sequences extracted from liver tissues. If multiple species and subspecies exist in North Carolina, the analyses done with skull and mandible measurements and mtDNA should separate specimens into groups. Using a historical perspective of the classification for the short-tailed shrew in North Carolina, I compared past and present species and subspecies designations in the state, mapped the species and subspecies ranges using locality information available with specimens, and investigated areas where species or subspecies may overlap to determine if any barriers may exist.

LITERATURE CITED

Bachman, J. 1837. Some remarks on the genus *Sorex*, with a monograph of the North American species. Journal of the Academy of Natural Sciences of Philadelphia 7(2): 262-402.

Baird, S. F. 1858. Mammals. In Reports of explorations and surveys, to ascertain the most practicable and economic route for a railroad from the Mississippi River to the Pacific Ocean, 8(part 1): l-757.

Baumgardner, G. O., N. O. Dronen and D. J. Schmidly. 1992. Distributional status of short-tailed shrews (genus *Blarina*) in Texas. The Southwestern Naturalist 37(3): 326-328.

Beckner, M. 1959. The biological way of thought. Columbia University Press, New York.

Benedict, R. A., H. H. Genoways, and J. R. Choate. 2006. Taxonomy of short-tailed shrews (genus *Blarina*) in Florida. Occasional Papers, Museum of Texas Tech University 251: 1-19.

Bole, B. P., Jr., and P. N. Moulthrop. 1942. The Ohio Recent mammal collection in the Cleveland Museum of Natural History. Scientific Publication, Cleveland Museum of Natural History, 5:83-181.

Brant, S. V. and G. Ortí. 2002. Molecular phylogeny of short-tailed shrews, *Blarina* (Insectivora: Soricidae). Molecular Phylogenetics and Evolution 22(2): 163-173.

Clark, M. K., D. S. Lee and J. B. Funderburg, Jr. 1985. The mammal fauna of Carolina bays, pocosins, and associated communities in North Carolina: an overview. Brimleyana 11: 1-38.

Cronquist, A. 1978. Once again, what is a species? Pp. 3–20 in J. Romberger, ed. Biosystematics in Agriculture. Allanheld and Osmun, Montclair, New Jersey.

De Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. Pp. 57-75, in Endless Forms: Species and Speciation (D. J. Howard and S. H. Berlocher, eds.). Oxford University Press, New York.

De Queiroz, K. 2005. Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences of the United States of America 102 Suppl: 6600-6607.

De Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology 56(6): 879-886.

Diersing, V. E. 1980. Systematics and evolution of the pygmy shrews (subgenus *Microsorex*) of North America. Journal of Mammalogy 61(1): 76-101.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York.

Dobzhansky, T. 1950. Mendelian populations and their evolution. The American Naturalist 74: 312-321.

Elliot, D. G. 1899. Descriptions of apparently new species and subspecies of mammals from the Indian Territory. Field Columbian Museum Publications. Zoology Series 1: 285-288.

Foresman, K. R. and R. E. Jensen, Jr. 1992. A morphometric comparison of the masked shrew (*Sorex cinereus*) and the Hayden's shrew (*Sorex haydeni*) in Montana and bordering states. Northwestern Naturalist 73: 15-21.

French, T. W. 1981. Notes on the distribution and taxonomy of short-tailed shrews (genus *Blarina*) in the southeast. Brimleyana 6: 101-110.

Gapper, A. 1830. Observations on the quadrupeds found in the district of Upper Canada extending between York and Lake Simcoe, with the view of illustrating their geographical distribution, as well as of describing some species hitherto unnoticed. Zoological Journal 5: 201-207.

Garland, D. A. and G. A. Heidt. 1989. Distribution and status of shrews in Arkansas. Proceedings of the Arkansas Academy of Science 43: 35-38.

Genoways, H. H. and J. R. Choate. 1998. Natural history of the southern short-tailed shrew, *Blarina carolinensis*. Occasional Papers, The Museum of Southwestern Biology 8: 1-43.

Genoways, H. H. and J. R. Choate. 1998. Natural history of the southern short-tailed shrew, *Blarina carolinensis*. Occasional Papers, The Museum of Southwestern Biology 8: 1-43.

George, S. B., J. R. Choate and H. H. Genoways. 1981. Distribution and taxonomic status of *Blarina hylophaga* Elliot (Insectivora: Soricidae). Annals of Carnegie Museum, 50(21): 493-513.

George, S. B., J. R. Choate and H. H. Genoways. 1986. *Blarina brevicauda*. Mammalian Species 261: 1-9.

Graham, R. W. and H. A. Semken. 1976. Paleoecological significance of the short-tailed shrew (*Blarina*), with a systematic discussion of *Blarina ozarkensis*. Journal of Mammalogy 57(3): 433-449.

Gray, J. E. 1838. [Untitled paper read to the Zoological Society of London]. Proceedings of the Zoologicial Society of London 1837: 123-126.

Hall, E. R. 1981. The Mammals of North America. 2nd edition. John Wiley and Sons, New York. 1181 pp.

Handley, C. O., Jr. 1971. Appalachian mammalian geography – recent epoch. Pp. 263-303. The distributional history of the biota of the Southern Appalachians. Part III: vertebrates (P. C. Holt, R. A. Paterson, J. P. Hubbard, eds.). Virginia Polytechnic Institute Research Division Monograph 4, Virginia Polytechnic and State University Blacksburg, Virginia.

Handley, C. O., Jr. 1979. Mammals of the Dismal Swamp: a historic account. Pp. 297 – 357, in The Great Dismal Swamp (P. W. Kirk, Jr., ed.). University Press of Virginia, Charlottesville, Virginia.

Harris, A. H. 1998. Fossil history of shrews in North America. Pp 133-156, in Evolution of Shrews (J. M. Wójcik and M. Wolsan, eds.) Mammal Research Institute, Polish Academy of Sciences, Bialowieża.

Huggins, J. A. and M. L. Kennedy. 1989. Morphological variation in the masked shrew (*Sorex cinereus*) and the smoky shrew (*S. fumeus*). American Midland Naturalist 122: 11-25.

Hulbert, R. C. Jr. and A. E. Pratt. 1998. Pleistocene (Rancholabrean) vertebrate faunas from coastal Georgia. Journal of Vertebrate Paleontology 18(2): 412-429.

Jones, C. A., J. R. Choate, and H. H. Genoways. 1984. Phylogeny and paleobiogeography of short-tailed shrews (genus *Blarina*). Pp. 56-148, in Genoways, H. H., and M. R. Dawson, eds. Contributions in Quaternary vertebrate paleontology: a volume in memorial to John E. Guilday. Carnegie Museum of Natural History, Special Publication 8: 1-538.

Kurta, A. 1995. Mammals of the Great Lakes Region. The University of Michigan Press, Michigan. 376 pp.

Lee, D. S., J. B. Funderburg, Jr. and M. K. Clark. 1982. A distributional survey of North Carolina mammals. Occasional Papers of the North Carolina Biological Survey. 72 pp.

Lesson, R. P. 1842. Nouveau tableau du regne animal. Mammiferes. Arthus Bertrand, Paris, 204 pp.

Mayden, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. Pp. 381-423, in Species: The Units of Diversity (M. F. Claridge, H. A. Dawah & M. R. Wilson, eds.). Chapman and Hall, London.

Mayr, E. 1942. Systematics and the origin of species. Columbia University Press, New York. 334 pp.

Mayr, E. 1969. Principles of systematic zoology. McGraw Hill, New York.

Mayr, E. 1996. What is a species, and what is not? Philosophy of Science: 262-277.

McCay, T. S. 2001. *Blarina carolinensis*. Mammalian Species 673: 1-7.

Merriam, C. H. 1895. Revision of the shrews of the American genera *Blarina* and *Notiosorex*. North American Fauna 10: 5-34, 102-107.

Mishler, B. D. 2010. Species are not uniquely real biological entities. Pp. 110-122, in Contemporary Debates in Philosophy of Biology (F. J. Ayala and R. Arp, eds.). Blackwell Publishing Ltd., Oxford.

Paul, J. R. 1965. *Blarina telmalestes* in North Carolina. Journal of Mammalogy 46(3): 496.

Rhoads, S. N. and R. T. Young. 1897. Notes on a collection of small mammals from northeastern North Carolina. Proceedings of the Academy of Natural Sciences of Philadelphia 49: 303-312.

Say, T. 1823. Account of an expedition from Pittsburgh to the Rock Mountains performed in the years 1819 and 20, under the command of Major Stephen H. Long. Compiled by Edwin James, 2 vols., Philadelphia, Pennsylvania.

Schmidly, D. J. and W. A. Brown. 1979. Systematics of short-tailed shrews (genus *Blarina*) in Texas. The Southwestern Naturalist 24(1): 39-48.

Simpson, G. G. 1943. Criteria for genera, species and subspecies in zoology and paleontology. Annals of the New York Academy of Sciences 44: 145-178.

Smith, A. B. 1994. Systematics and the fossil record: documenting evolutionary patterns. Blackwell Science, Cambridge, Massachusetts.

Sokal, R. R. and P. H. A. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman, San Francisco, California.

Thompson, C. W., J. R. Choate, H. H. Genoways and E. J. Finck. 2011. *Blarina hylophaga* (Soricomorpha: Soricidae). Mammalian Species 43(878): 94-103.

Trani, M. K., W. M. Ford and B. R. Chapman (editors). 2007. The Land Manager's Guide to Mammals of the South. The Nature Conservancy, Southeastern Region, Durham, North Carolina. 546 pp.

Van Zyll De Jong, C.G. 1980. Systematic relationships of woodland and prairie forms of the common shrew, *Sorex cinereus cinereus* Kerr and *S. c. haydeni* Baird, in the Canadian prairie provinces. Journal of Mammalogy 61(1): 66-75.

Van Zyll De Jong, C.G. and G. L. Kirkland, Jr. 1989. A morphometric analysis of the *Sorex cinereus* group in central and eastern North America. Journal of Mammalogy 70: 110-122.

Webster, W. D., P. B. Colwell and M. A. Shields. 1984. Noteworthy records of mammals from North Carolina. The Journal of the Elisha Mitchell Scientific Society 100(3): 112-115.

Webster, W. D., J. F. Parnell, and W. C. Biggs, Jr. 1985. The mammals of the Carolinas, Virginia, and Maryland. University of North Carolina Press, Chapel Hill, 255 pp.

Webster, W. D. 1996. Geographic variation in *Blarina brevicauda* (Insectivora: Soricidae) from eastern North Carolina, with description of a new subspecies. Pp. 47-56, in Contributions in mammalogy: a memorial volume honoring Dr. J. Knox Jones, Jr. (H. H. Genoways and R. J. Baker, eds.). Museum of Texas Tech University, Lubbock, 315 pp.

Webster, W. D., N. D. Moncrief, J. R. Choate and H. H. Genoways. 2011. Systematic revisions of the northern short-tailed shrew, *Blarina brevicauda* (Say). Virginia Museum of Natural History. Memoir 10. 77 pp.

Whitaker, J. O., Jr. and W. J. Hamilton, Jr. 1998. Mammals of the eastern United States. 3rd edition. Cornell University Press, Ithaca, New York. 583 pp.

Wilson, D. E. and D. M. Reeder (editors). 2005. Mammal species of the world: a taxonomic and geographic reference. 3rd edition. The Johns Hopkins University Press, Baltimore, Maryland. 2142 pp.

Wilson, D. E. and S. Ruff (editors). 1999. North American I Mammals. Smithsonian Institute, Washington, DC.

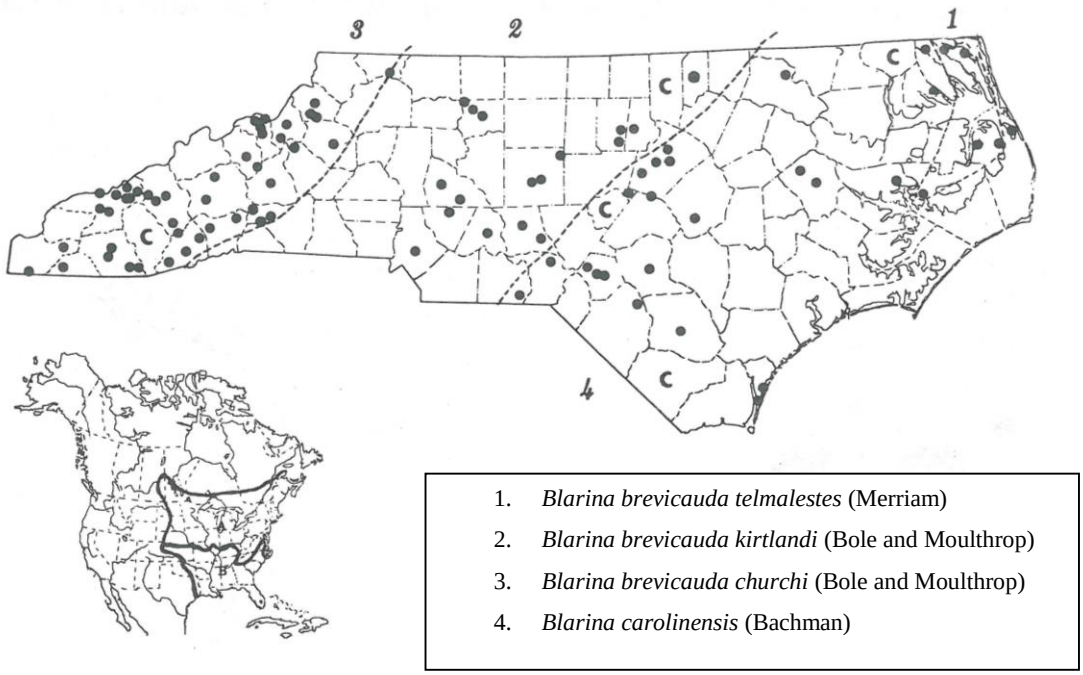
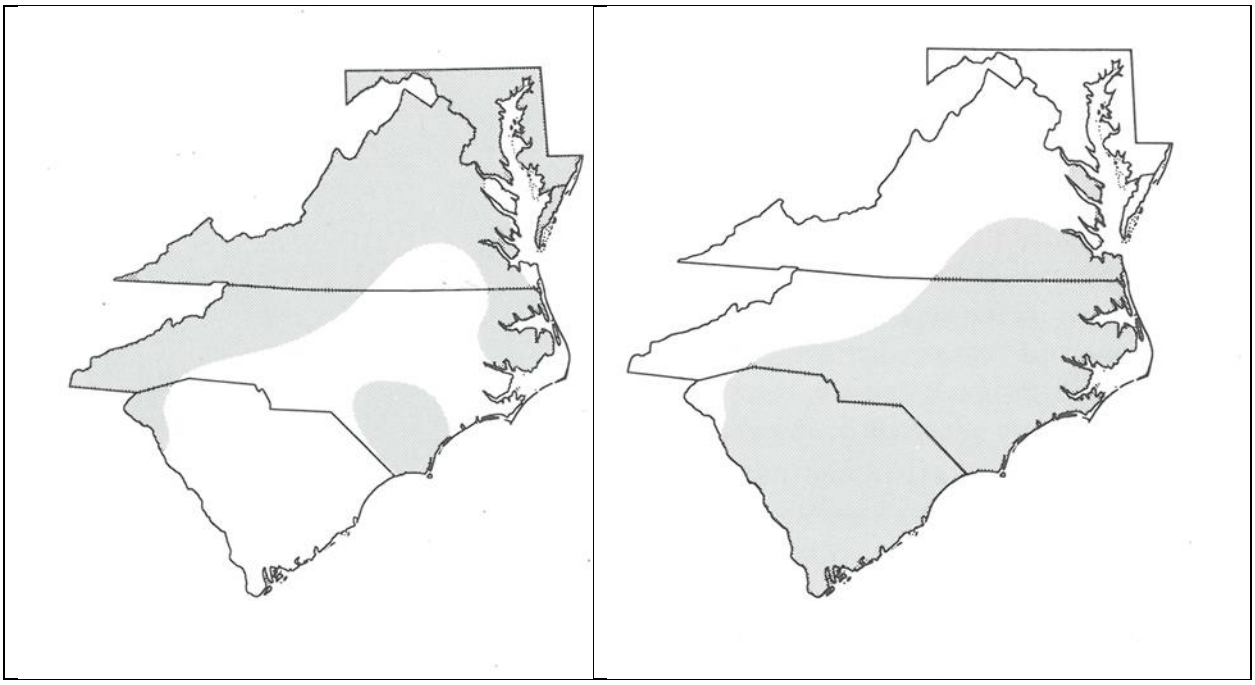


Figure 1.1: Proposed distributions of species and subspecies designations for the short-tailed shrews in North Carolina by Lee et al. (1982). Three subspecies of *Blarina brevicauda* were proposed to occur in North Carolina: *B. b. telmalestes* (1) was thought to be confined to the Dismal Swamp region only, *B. b. kirtlandi* (2) was thought to occupy most of the upper piedmont and lower elevations of the mountains, and *B. b. churchi* (3) was documented as the largest subspecies in North Carolina and was thought to only be found at the highest elevations of the mountains. The short-tailed shrews in southeast North Carolina were classified as *B. carolinensis* (4) and thought to be found in the coastal plain and lower piedmont region. Two or more museum records in close proximity are indicated with a dot, while a "c" reflects a record indicating only the county level locality. The North American distribution map was split as *Blarina brevicauda* (A) to the north and *B. carolinensis* (B) to the south. This figure was taken from A Distributional Survey of North Carolina Mammals by Lee et al. (1982).



Northern short-tailed shrew | Southern short-tailed shrew

Figure 1.2: Proposed distributions of the short-tailed shrews in North Carolina by Webster et al. (1985). The map on the left shows the distribution of the northern short-tailed shrew (*Blarina brevicauda*), while the map on the right shows the distribution of the southern short-tailed shrew (*B. carolinensis*). A disjunct population of the northern short-tailed shrew is shown in southeastern North Carolina. These figures were taken from Mammals of the Carolinas, Virginia, and Maryland by Webster et al. (1985).

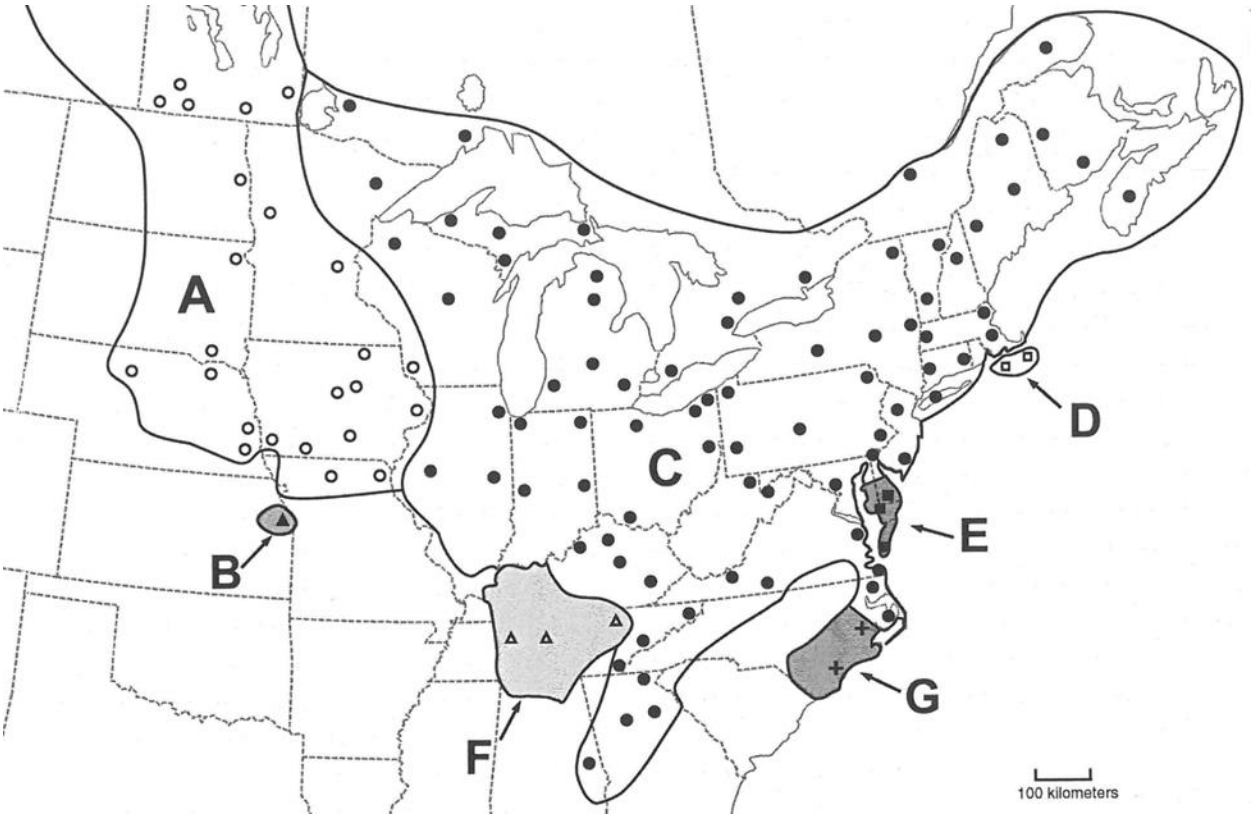


Figure 1.3: Geographic distribution of northern short-tailed shrew (*Blarina brevicauda*) in the United States and southern Canada based on museum voucher specimens by Webster et al. (2011). Each symbol represents an operational taxonomic unit (OTU) that is combination of multiple samples in a given geographic region. The subspecies designations are: A (○) = *B. b. brevicauda*; B (▲) = *B. b. jerrychoatei*; C (●) = *B. b. talpoides*; D (□) = *B. b. aloga*; E (■) = *B. b. delmarvensis*; F (Δ) = *B. b. cumberlandensis*; G (+) = *B. b. knoxjonesi*. Shaded areas indicate peripheral subspecies distribution with *B. b. aloga* (not shaded) occurring on Martha's Vineyard and Nantucket Island, Massachusetts. This figure and text with minor changes were taken from the Systematic Revision of the Northern Short-tailed Shrew, *Blarina brevicauda* (Say) by Webster et al. (2011).

Chapter 2

Morphological Variation in Short-tailed Shrews (Genus *Blarina*) in North Carolina

ABSTRACT

How we define species depends on the questions being asked and the methodology used. The phenetic (or morphological) species concept is used in systematics of mammals to group organisms based on measureable differences. I examined the current taxonomy of the short-tailed shrews in the genus *Blarina* in North Carolina by comparing the morphological variation in skulls and mandibles from museum specimens of *Blarina brevicauda knoxjonesi, B. b. talpoides,* and *B. carolinensis carolinensis*. I aged individuals with a four-part classification criterion, and compared multivariate analyses conducted on 9 skull and 6 mandible variables to separate species and to look for subspecies. Each specimen was georeferenced with a point radius uncertainty to map the distribution of *Blarina brevicauda* and *B. carolinensis* in North Carolina. Throughout the Blue Ridge and Piedmont ecoregions, the species are parapatric in distribution. In the Middle Atlantic Coastal Plain and the southern part of the Southeastern Plains, the two species appear more sympatric. In southeastern North Carolina, specimens designated as *B. b. knoxjonesi* are not as disjunct as previously thought and taxonomic revisions are warranted. My analyses suggest that reclassifying specimens identified as *B. b. knoxjonesi* as *B. b. talpoides* should be considered.

Keywords: *Blarina*; classification tree analysis; cranial morphometrics; discriminant function analysis; morphology; multi-variate analysis; principle components analysis; short-tailed shrews; taxonomy

INTRODUCTION

The morphological (or phenetic or phenotypic) species concept is used in systematics to group organisms based on measureable (or quantitative) physical differences (Ridley, 1989; Simpson, 1961; Smith, 1994; Sneath and Sokol, 1973). These measureable characters are used to cluster individual characters statistically to distinguish groups that separate visually when plotted (Michener, 1970; Sokol and Crovello, 1970; Sneath and Sokol, 1973). Very occasionally, one or two quantitative characters are able to split groups (Handley and Varn, 1994; Kirkland et al., 1987; Schmidly et al., 1988) but, almost universally, using all the measured characters in a multivariate approach is required (Fisher, 1936; Johnson and Wichern, 2007).

Multivariate analysis can be used to combine observed or measured variables into a reduced set of independent variables that separate the groups the best. In mammals, skulls exhibit many critical, diagnostic characters that have been used to separate species. A skull is involved with environmental interactions, like eating, movement and senses, that can influence skull size and shape (Holmes et al., 2016). The most common multivariate analyses used for group separation are principal component analysis (Jackson, 1993), discriminant function analysis (Mitteroecker and Bookstein, 2011) and canonical variate analysis (Albrecht, 1980). Combinations of these analyses are often used when separating mammal species by measurements done on skulls (Bronner et al., 2007; French et al., 1988; Lydeard and Kennedy, 1988; Patton, 1973). Despite

their wide use, multivariate approaches must be applied and interpreted with care as derived factors or clusters do not necessarily reflect those in nature (James and McCulloch, 1990).

Principal components analysis (PCA) is exploratory in nature, and describes the variance-covariance structure in a dataset (Johnson and Wichern, 2007). PCA is often visualized as a scatter plot to see if the data splits into clusters and reduces the number of observed variables that can describe the variance-covariance structure. Linear discriminant function analysis (LDFA) examines which variables discriminate between defined groups (Johnson and Wichern, 2007). LDFA is often visualized graphically as a frequency histogram where the combination of the variables that are most common cause separation between the known groups. Canonical variate analysis – also called canonical correlation analysis – is equivalent to discriminant function analysis in some instances, especially in how canonical variates and discriminant functions are extracted in linear combinations (Glahn, 1968; Hastie et al., 1995; Hotelling, 1936). Visually, canonical analysis is often presented with either a scatter plot, a histogram, or both.

In addition to the traditional multivariate analyses used to separate or group individuals, decision trees can split data based on classes or values. Classification and regression tree (CART) analysis is an umbrella term where classification trees split data into a finite number of classes, and regression trees split data with continuous real numbers (Brieman et al., 1984; Loh, 2011). In classification tree analysis, recursive partitioning creates a decision tree that splits the data into a finite number of classes, and continues to split each new sub-group of data until an end criterion is reached (Therneau and Atkinson, 1997). Classification trees have an advantage of examining individual variable importance that have the greatest impact on splitting the finite number of classes (Strobl et al., 2009).

Short-tailed shrews of the genus *Blarina* were once considered two species (Hall, 1981). Additional samples examined and advances in systematics has now revealed four species within this genus (Wilson and Reeder, 2005). Variation and systematics in shrews of the genus *Blarina* were studied in Pennsylvania (Guilday, 1957), Kentucky (Rippy, 1967), Nebraska (Genoways and Choate, 1972), Connecticut (Choate, 1972), Illinois (Ellis et al., 1978), south-central Virginia (Tate et al., 1980), Kansas-Iowa-Missouri (Moncrief et al., 1982), Tennessee (Braun and Kennedy, 1983), Arkansas (Garland and Heidt, 1989), Texas (Schmidly and Brown, 1979; Baumgardner et al., 1992), eastern North Carolina (Webster, 1996) and Florida (Benedict et al., 2006). In the studies that examined what are now considered *Blarina brevicauda* and *B. carolinensis*, the distributions showed a narrow zone of overlap. In southeastern Virginia and eastern North Carolina, however, the distributions seem to have broad overlap.

If the distributions of two species abut, but do not overlap, the species are considered parpapatric. If the distributions overlap, the two species are sympatric. No two species occupy the exact same places at the same time (i.e. syntopic distribution) as some resources differ between them. Benedict (1999) suggested that the potential movement of a parapatric boundary could give the impression of broad sympatry between species if specimens were collected over a long period of time. Whether the distributions of *Blarina brevicauda* and *B. carolinensis* are actually sympatric in eastern North Carolina or have a moving line of parapatry is not known.

Historically, only the northern short-tailed shrew (*Blarina brevicauda carolinensis*, *B. b. churchi*, and *B. b. kirtlandi*) and the Dismal Swamp short-tailed shrew (*B. telmalestes*) were recognized in North Carolina (Hall, 1981; Paul, 1965). George et al. (1986) continued to acknowledge *B. b. churchi* and *B. b. kirtlandi*, but changed *B. telmalestes* to *B. b. telmalestes* in North Carolina and southeastern Virginia. Genoways and Choate (1998) examined the natural

history of the southern short-tailed shrew (*Blarina carolinensis*), and determined correct species designations for specimens that had originally been designated as *B. brevicauda*. They summarized potential areas of overlap with the northern short-tailed shrew in North Carolina.

This taxonomic confusion meant that many museum specimens collected prior to 1998 were potentially identified as *Blarina brevicauda*, as *B. telmalestes* or as old *B. brevicauda* subspecies, and may warrant updates within museum collections.

Lee et al. (1982) proposed distributions of *Blarina* species and subspecies (*B. brevicauda churchi, B. b. kirtlandi, B. b. telmalestes* and *B. carolinensis*) in North Carolina. In their state distribution map for *Blarina*, *B. carolinensis* was mapped as confined to the eastern third of the state. The distributions of the two species of *Blarina* in North Carolina was updated by Webster et al. (1985), and Webster (1996) designated a new subspecies (*B. brevicauda knoxjonesi*) in the southeastern portion of the state. Webster et al. (2011) conducted a systematic revision of the northern short-tailed shrew in North America. In North Carolina, they combined *Blarina brevicauda churchi, B. b. kirtlandi* and *B. b. telmalestes* into *B. brevicauda talpoides*, but maintained *B. brevicauda knoxjonesi* as a current subspecies in the state (Webster et al., 2011). Although many changes in the distributions and species designations of *Blarina* have occurred in North Carolina just since 1981, additional samples and analyses may show that further revisions are still needed.

The goal of my research was to examine the current taxonomy of the short-tailed shrews of the genus *Blarina* in North Carolina. To do this, I used the phenetic species concept to examine the morphological variation in *Blarina* specimens in the collections of the North Carolina Museum of Natural Sciences (NCSM – formerly the North Carolina State Museum of Natural Sciences) and the University of North Carolina Wilmington (UNCW) mammal collections.

These museum collections offer the best array of specimens over space and time in North Carolina. I used measurements from museum specimens to confirm or correct species identification and investigated subspecies designations. In my research, I determined species where analyses on measured characters completely separate groups with no overlap. I determined subspecies where analyses on measured characters show some overlap within a species, but show an abrupt change that occurs at some geographic region separated from different subspecies. I also used the locality information from the specimens to create an updated distribution map for *Blarina* in North Carolina. With the map showing the presence of *Blarina* in North Carolina, I examined areas where *B. brevicauda* and *B. carolinensis* are in contact to determine the extent of overlap.

MATERIALS AND METHODS

I collected data and measured museum specimens from the mammal collections at the University of North Carolina at Wilmington (UNCW) and the North Carolina Museum of Natural Sciences (NCSM) (Table 2.1). For each individual specimen from UNCW, I took 15 cranial and mandibular measurements (Figure 2.1) with a Mitutoyo Absolute Digimatic caliper (Model No. NTD12-6" CX – level of accuracy, 0.01 mm) and Input Tool (IT-012U) for the transfer into an Excel spread sheet. Specimens from the NCSM were corrected for species identification based on UNCW specimen data analysis. See Appendix I for information about museum specimens used in this study.

Choate (1972) established a basic list of skull measurements that have been used, with some variation, by other researchers (Braun and Kennedy, 1983; Ellis et al., 1978; Genoways and

Choate, 1972; George et al., 1981; Handley and Varn, 1994; Moncrief et al., 1982; Tate et al., 1980). I looked for a consensus on skull and mandible characters that provided a good overall comparison and were good predictors for splitting taxa.

Descriptive statistics of all external measurements from the museum specimens, skull and mandible measurements, and all analyses were done in RStudio statistical software (RStudio and Inc., 2014; R package Version 0.98.1102).

I assessed first how well the currently designated taxa match the variation of critical morphological characters of *Blarina* skulls from North Carolina. To gain a rough overview of the variation found in the skulls and mandibles, I used bivariate scatter plots. I then used multivariate analyses to examine what contributed to the variation, and if the variation corresponded to current taxonomy. For multivariate analyses, I used only data from specimens with complete data. I created separate data sets for the skulls proper and the mandibles.

To test if any variation was due to sex or age, I ran a one-way multivariate analysis of variance (MANOVA—Johnson and Wichern, 2007) on a dataset with complete sex, age and measurement data. I used sex and age as the independent variables, and the 9 skull characters or the 6 mandibular characters as response or dependent variables to determine if there were differences between means of each group. The sex was used only when indicated from specimen data (Table 2.1), and I designated the age (i.e. sub-adult, adult, or old adult) by a 4-part classification criterion looking at degree of tooth wear (Pearson, 1945), upper incisor root exposure (Diersing, 1980; Guilday, 1957), cranial suture closure (Guilday, 1957) and the degree to which the cranial suture was pronounced (Table 2.2). Shrews generally devote much of their time foraging, and consequently, tooth wear occurs early in development due to the nature of their diet (Churchfield, 1990). Therefore, I weighted tooth wear and upper root exposure more

heavily than the cranial suture closure and crest pronouncing (Table 2.2). If either sex or age was significant, I ran an analysis of variance on each dependent skull and mandible character to determine exactly what characters were variable with the independent class selection.

I performed a principal components analysis to see specifically what skull and mandible characters combined best to explain the observed variation. A scree plot determines the minimum number of principal components needed to explain the variation, but I examined the first three principal components. Since all the measured characters are in the same units, I used a covariance matrix as no standardization of measure was needed.

I performed a linear discriminant function analysis to separate group membership based on a linear combination of the skull and mandible characters. I ran one discriminant analysis with species as the known groups, and ran one with subspecies as the known groups for both skull and mandible characters.

My final analyses generated taxonomic relationships based on the skull and mandible data. I performed a classification tree analysis to generate taxonomic trees based on measured characters using the RPART (Recursive PARTintioning; Therneau and Atkinson, 1997) package in R.

The RPART package allows the option of two splitting indices. A Gini splitting index looks at the probability of a class being chosen times the probability of a mistake in categorizing (Strobl et al., 2007), and splits data based on the largest improvement (i.e. gain or impurity reduction) that also has the lowest complexity parameter (i.e. α = cost of a tree penalized by the number of terminal nodes) and the lowest expected loss. An information splitting index uses information gain, which is the probability of a randomly chosen example belonging to a class and a classification entropy that involves a logarithmic transformation (Berzal et al., 2003; Raileanu

and Stoffel, 2004). The information gain is also called the Kullback-Leibler divergence (Kullback, 1959; Kullback and Leibler, 1951), and is most commonly used in the iterative dichotomiser 3 (ID3 – Quinlan, 1986) and the C4.5 (Quinlan, 1993) algorithms. The Gini splitting index is recommended for classification tree analysis, so I have used this for the analysis.

I calculated two measures of misclassification rates. The re-substitution error rate is the error rate computed on the training data in the decision tree (Therneau and Atkinson, 1997). This misclassification is calculated by multiplying the root node error (i.e. total number of specimens not of the predicted class divided by the total number of total specimens) and the relative error (i.e. error on the observations used to estimate the regression model; one minus the root mean squared error; R^2 is similar to linear regression). The cross-validated error rate uses a 10-fold cross validation that is an objective indicator of predictive accuracy (Therneau and Atkinson, 1997). This misclassification is calculated by multiplying the root node error and the apparent error (i.e. error on the observations from the cross validation data; predicted residual error sum of squares).

Georeferencing museum voucher specimens

I obtained latitude and longitude coordinates based on the locality information provided for each specimen using the mapping software Topo USA version 6.0 (DeLorme, 2006) and Topo North America version 10.0 (DeLorme, 2013). For specimens located only to the county level, I used the geographic center of the county as the best coordinates for the true location of a specimen. If a specific town was used, I used the population center or specific locale as designated by the DeLorme software. I determined the coordinates for distance and directional localities (e.g. Raleigh, 5 mi E; Stumpy Point, on US 264, 1 mi S from Navy Shell Rd) with the

'Draw' function or contouring the roads and determining the road length. For locations, historic landmarks or old road names that were unable to be found, I used the larger general location that encompassed the detailed locality or used the geographic center of the county.

The confidence for each point is distance away from the given latitude/longitude point where it is assumed that the true point exists. I used a diameter based confidence (Table 2.3: Metzler, 1992; Metzler, 1994) with early specimens, but was converted to a point-radius confidence interval in meters representing the measurement error. Specimens that were georeferenced with the Topo North America software used a point-radius uncertainty specific to the variations in the written locality (Wieczorek et al., 2004). Specimens with localities and uncertainties beyond the county level were not used in the final distribution map. Figure 2.2 gives examples for the latitude and longitude coordinates and the point radius uncertainty measurement errors.

RESULTS

Taxa selection

I examined and measured the skulls of 491 total specimens from UNCW of short-tailed shrews (131 *Blarina brevicauda knoxjonesi*, 197 *B. b. talpoides*, and 163 *B. carolinensis carolinensis*) because initially, only these specimens were identified to the subspecies level. All the UNCW specimen measurements were used in the descriptive statistics and bivariate analysis. For the multivariate analysis, I reduced this original data set to 448 total specimens with complete skull measurements and 461 total specimens with complete mandibular measurements. I also reduced this for a complete sex and age data set of 304 total specimens. I examined 584 short-tailed shrews (36 *B. b. knoxjonesi*, 207 *B. b. talpoides*, and 341 *B. c. carolinensis*) from

NCSM. I proofed and corrected species identifications on all NCSM *Blarina* specimens in North Carolina based on the UNCW findings. NCSM *B. b. knoxjonesi* designations were made based on subspecies distribution range (Webster, 2011).

Descriptive statistics and bivariate scatter plots

Table 2.4 shows the descriptive statistics of external measurements from specimen data, and all skull and mandible measurements for *Blarina brevicauda* and *B. carolinensis*. All mean comparisons were significant, while the variance comparison showed all measurements were significant except tail vertebrae and ear lengths. Table 2.5 shows the descriptive statistics between *B. b. knoxjonesi* and *B. b. talpoides,* where the mean comparisons were all significant ($P < 0.01$) except the tail vertebrae, hind foot and ear. The subspecies variance comparison showed only UCL, MNH, ARB and CorHt as significant ($P < 0.01$) measurements. Figure 2.1 defines the abbreviations used for all the skull and mandible variables measured in this analysis.

Bivariate scatter plots showed some poor (i.e. with overlap) and some good (i.e. little to no overlap) separation for *B. brevicauda* and *B. carolinensis* with combinations of measured skull or mandible characters. *B. b. knoxjonesi* and *B. b. talpoides* showed consistent overlap when using both the skull and mandible characters. Figure 2.3 shows skull character combinations, and Figure 2.4 shows mandible character combinations illustrating this. The R-squared or coefficient of determination value ranged from 0.6231 to 0.9474 for the skull and mandible bivariate plots, where the higher the value, the better one value can predict another. Consequently, bivariate plots were adequate in separating species with characters of low or high measured values that did not overlap, but did a poor job in separating species and subspecies with characters of intermediate values that overlapped.

Multivariate analysis of variance

I ran a one-way multivariate analysis of variance (MANOVA) with sex and age as the independent variables and the cranial measurements as the dependent variables. 304 samples (75 *Blarina brevicauda knoxjonesi*, 146 *B. b. talpoides*, and 83 *B. carolinensis carolinensis*) of known sex and age with complete measurements were used in this analysis. Sex (163 females and 141 males) was not significant ($p = 0.2609$) as a single dependent variable, so both sexes were pooled for the future analyses.

Combining the 4-part classification criteria (Table 2.2) resulted in 117 sub-adults, 151 adults and 36 old-adults in the analysis. Age was significant ($p < 0.001$) as a single dependent variable, but an analysis of variance done on each skull and mandible character only showed UTR ($p = 0.0113$) and UCL ($p < 0.001$) as significant characters. Boxplots of age versus these toothrow variables and MdbTR (Figure 2.5) show a slightly lower median and mean for the old adults, but are very similar for the values in the sub-adult and adult age classes. These boxplots also show overlap between the measured characters across each age classes. As *Blarina* shrews age, the angle of their upper incisor changes, causing the UTR to decrease, and wear and loss of the unicuspids can decrease the UCL significantly. As toothrow characters decrease in *Blarina*, other characters, like MAB, CRB and CorHt, increase (Figure 2.6). Since multivariate analyses combine multiple measured skull and mandible characters, I combined all age classes for future analyses, but checked if the UTR and UCL were significant in multivariate analyses. Figure 2.1 defines the abbreviations used for all the skull and mandible variables measured in this analysis.

Principal components analysis

I performed a principal components analysis (PCA) on the *Blarina* skull data (n = 448) with a covariance matrix, which explains the variance of each measured variable. The scree plot (Figure 2.7) showed that 2 principal components were appropriate to compare the total sample variance, and is a plot of the magnitude of an eigenvalue (i.e. measure of the amount of variation explained by the principal component) versus the principal component number (i.e. PC1, PC2, PC3, etc.). The scatter plot of the first two principal components separated *Blarina brevicauda* and *B. carolinensis*, but showed some overlap in *B. b. knoxjonesi* and *B. b. talpoides* (Figure 2.8). The first principal component (PC1) had an eigenvalue of 7.192. Since this value was greater than 1, it indicated that the principal component accounts for more variance than is explained by one of the original variables alone in standardized data. The second principal component (PC2) is calculated in the same way as the first, with the condition that it is uncorrelated with (i.e., perpendicular to) the first principal component and that it accounts for the next highest variance. PC2 has an eigenvalue of 0.182, and the third principal component (PC3) has an eigenvalue of 0.059. PC1 explains 94.9% of the total sample variance, while PC2 explains 2.4%, and the first two principal components collectively explain 97.3% of the total sample variance. PC3 only explains 0.8% of the total sample variance, so adding this to the first two principal components results in 98.1% of the total sample variance explained by the collective sum. The eigenvalues and the percent of sample variance support the scree plot to show that only two principal components are needed to sufficiently explain the total sample variance.

PC1 is a weighted sum of GRL, OPML, CRB and UTR as determined by their respective weighted eigenvector scores (i.e. weights of the original variables used to calculate the principal

component) of -0.608, -0.571, -0.313 and -0.305 respectively with the other variables being less of an influence. PC2 is a weighted difference primarily between the variables CRB and MAB, and the variables GRL and UTR with respective weights of -0.785, -0.382, 0.281 and 0.249. The correlation coefficients, which are the correlations of the original variables with the principal components, show a slightly different order of the variables. PC1 shows the most correlated variables are a weighted sum of GRL, OPML, UTR and MTL with correlation coefficients of -0.995, -0.993, -0.972 and -0.950. PC2 has a weighted difference primarily between the variables CRB and MAB, and the variables UCL and UTR with the correlation coefficients as -0.369, -0.301, 0.186 and 0.126. Table 2.6 summarizes the skull covariance matrix PCA and Figure 2.1 defines the abbreviations used for all the skull and mandible variables measured in this analysis.

I also examined the *Blarina* mandible data (n = 461) PCA with the covariance matrix. The scatter plot of the first two principal components from the mandible data (Figure 2.9) separated *Blarina brevicauda* and *B. carolinensis* with some overlap between the species, but showed more overlap in the subspecies than with the skull data. PC1, PC2 and PC3 had eigenvalues of 3.235, 0.130 and 0.085 with a cumulative proportion of the sample variance explained as 0.9300, 0.9674 and 0.9917 respectively. PC1was a weighted sum of GrLgt, MNL, MNH and CorHt (-0.634, -0.515, -0.378 and -0.354), with the correlation coefficients weighted by GrLgt, MNH, CorHt and MNL (-0.984, -0.959, -0.956 and -0.952). PC2 was a weighted difference primarily between the MNL and the variables MNH and CorHt (0.776, -0.430 and -0.422). The correlation coefficients also showed a weighted difference between MNL and the variables CorHt, ARB and MNH (0.287, -0.228, -0.220 and -0.219). Table 2.7 summarizes the mandible covariance matrix PCA.

In addition, I performed a PCA with the covariance matrix with only *Blarina brevicauda* samples on both the skull (n = 297) and mandible (n = 306) data to investigate subspecies separation and measured characters used. There was no improvement in the separation of the subspecies in the scatter plot from the first two principal components. Furthermore, the weighted characters in both PC1 and PC2 were nearly identical to the PCA including both *B. brevicauda* and *B. carolinensis*. There were some differences, however, in the order of characters weighted to explain the variation in each principal component.

Discriminant function analysis

The linear discriminant function analysis histogram created from the skull characters separate *Blarina brevicauda* and *B. carolinensis* (Figure 2.10) with approximate group centroids of -3.0 and 2.5 respectively. The histogram for *B. b. knoxjonesi* and *B. b. talpoides* (Figure 2.11) do not show a clear group subspecies separation with their overlapping distributions, but their group centroids of -1.1 and 1.1 suggest a slight separation. The skull species misclassification error rate of 0.0045 indicates that 99.55% of the species were classified correctly, while the skull subspecies had a higher misclassification error rate of 0.1852 indicating that only 81.48% of the subspecies were classified correctly. The error rate may also suggest that some specimens could be mislabeled. The ratio of between- and within-group standard deviations for the linear discriminant variables is 3251.456 for the species group, and 241.3292 for the subspecies group.

Beyond the differences between the classification error and group comparison, the first discriminant function (i.e. LD1 – the factor that separates the species or subspecies by contrasts of the measured variables) also has its differences (Table 2.8). The skull species comparison shows LD1 to be a contrast between the highest factor loadings of UTR (-2.492), MTL (1.782),

OPML (-1.417), IOB (-1.244) and MAB (1.206) with less emphasis on the UCL, GRL, CRH and CRB (-0.906, 0.621, -0.269 and -0.042 respectively). The skull subspecies comparison shows a factor loading contrast between the UTR (1.283), UCL (1.264), CRB (1.189), GRL (-0.797) and MAB (0.737) with less emphasis on MTL, OPML, IOB and CRH (0.511, 0.452, 0.192 and 0.060). Figure 2.1 defines the abbreviations used for all the skull and mandible variables measured in this analysis.

The mandible measurements showed a slightly higher percent for the species (99.57%) and the subspecies (83.33%) classified correctly when compared to the skull measurements (Table 2.9). The mandible species comparison had group centroids of -2.5 and 2.2 for *B. brevicauda* and *B. carolinensis* (Figure 2.12) with the ratio of between- and within-group standard deviations of 2434.129, and the subspecies comparison had group centroids of -0.8 and 0.8 for *B. b. knoxjonesi* and *B. b. talpoides* (Figure 2.13) with a ratio of 259.4218. The mandible species factor loadings of LD1 contrasted GrLgt (-1.907), MdbTR (-1.742) and ARB (0.938) with less weight on CorHt, MNH and MNL (0.377, -0.106 and 0.101). The factor loadings for the mandible subspecies contrasted ARB (1.265), MNH (1.142) and MNL (0.741) with little emphasis on GrLgt, MdbTR and CorHt (0.379, -0.302 and 0.265).

Classification tree analysis

I performed a classification tree analysis on the *Blarina* skull (n = 448) and mandible (n = 461) data using a Gini splitting index. Figure 2.14 shows all of the skulls identified as *Blarina carolinensis carolinensis* had a smaller occipital-premaxilla skull length (OPML < 19.735, as the first split or node, where all the skulls identified as *B. brevicauda* had a larger skull length (OPML ≥ 19.735). The second split at node 3 used a cranial breadth measure (CRB < 11.805).

Node 5 split 65.1% of all the skulls identified as *B. b. talpoides* with a larger cranial breadth. This node also split 8 skulls (UNCW 483, 4401, 3682, 2610, 3595, 792, 11938 and 4975) identified as *B. b. knoxjonesi* with a larger cranial breadth. 55.2% of all remaining skulls identified as *B. b. knoxjonesi* have a more narrow maxillary breadth (MAB < 7.165), but there are also 11 skulls identified as *B. b. talpoides* with a narrow maxillary breadth at node 6.

For the remainder of the skulls with a larger maxillary breadth, the main tree splits are OPML, MAB, OPML, UTR (upper tooth row) and CRH (cranial height). This resulted in 11 *B. b. knoxjonesi* and 39 *B. b. talpoides* specimens being grouped as the *talpoides* group to the right side of the tree, while 37 *B. b. knoxjonesi* and 10 *B. b. talpoides* specimens were grouped to the left as the *knoxjonesi* group (Figure 2.14). The skull classification analysis had a re-substitution error rate of 0.0893 or 91.07% classified correctly and a cross-validated error rate of 0.1674 or 83.53% classified correctly. The error rate may also suggest that some specimens could be mislabeled. Table 2.10 summarizes each of the tree nodes in the skull classification analysis.

The classification analysis performed on the mandible data showed nearly a perfect split of *Blarina brevicauda* and *B. carolinensis* (Figure 2.15). All of the mandibles identified as *B. c. carolinensis* had a shorter greatest length of mandible including the incisor (GrLgt < 13.18), but 2 mandibles (UNCW 3596 and 0477) identified as *B. b. knoxjonesi* also had a short mandible length. The next split at node 3 showed 66.1% of all *B. b. talpoides* specimens with a taller coronoid process height (CorHt >6.095), but 3 specimens (UNCW 11938, 1480 and 11554) identified as *B. b. knoxjonesi* also grouped here. The last split at node 4 separated the remaining specimens with a length of mandible without the incisor (MNL < 12.94), where 88.7% of all *B. b. knoxjonesi* had a shorter mandible were classified as the *knoxjonesi* group. However, this small mandible size also grouped 33 specimens identified as *B. b. talpoides*. For the remaining

mandibles with a longer length (MNL ≥ 12.94) at node 7, 11 specimens identified as *B. b. knoxjonesi* and 28 specimen identified as *B. b. talpoides* were classified as the *talpoides* group. The mandible classification analysis had a re-substitution error rate of 0.1063 or 89.37% classified correctly and a cross-validated error rate of 0.1280 or 87.20% classified correctly. Table 2.11 summarizes each of the tree nodes in the mandible classification analysis.

See Appendix II for information about the specimens split in the classification analysis. Additional analyses that support the split of *Blarina brevicauda* and *B. carolinensis* and acknowledge the difficulty in splitting *B. b. knoxjonesi* from *B. b. talpoides* can be found in Appendix III.

Georeferencing museum voucher specimens

All georeferenced specimens were placed on a geographic information system (GIS) map layer by M. B. Norton – NCSM Database and GIS Manager. I had 1226 *Blarina* specimens georeferenced for the maps, comprising 584 from NCSM and 642 from UNCW. Of the total number examined, 169 (34 NCSM and 135 UNCW) specimens were identified as *B. brevicauda knoxjonesi*, 505 (210 NCSM and 295 UNCW) specimens were identified as *B. b. talpoides*, and 552 (340 NCSM and 212 UNCW) specimens were identified as *B. carolinensis carolinensis*.

Based on the museum records from NCSM and UNCW, short-tailed shrews are present in all the Level III North Carolina ecoregions: Blue Ridge, Piedmont, Southeastern Plains and Middle Atlantic Coastal Plain (Figure 2.16). *Blarina brevicauda talpoides* occupies the Blue Ridge to the western extent of the Piedmont in western NC, and the northeast corner of the state in the Middle Atlantic Coastal Plain. *B. b. knoxjonesi* is found south of the Pamlico River in the Middle Atlantic Coastal Plain and Southeastern Plains within the Sandhills region. *B. carolinensis*

carolinensis does not occur in the Blue Ridge, but occurs through the remaining eastern ecoregions. In the southern portion of the Southeastern Plains and the Middle Atlantic Coastal Plain, there appears to be overlap between the two species. Only *B. carolinensis carolinensis* has been documented on the Outer Banks, but samples are very limited.

The species of *Blarina* were found in 87 of the 100 NC counties, excluding Anson, Bertie, Cherokee, Cleveland, Davidson, Green, Lenoir, Lincoln, Martin, Pamlico, Person, Warren and Yadkin. They undoubtedly exist in these counties but simply were not sampled (Figure 2.16). In the western counties, only Burke and Guilford County have both species present from museum records with no apparent overlap. In the eastern counties, 19 have both species present with most of the counties being found in the Middle Atlantic Coastal Plain. In Currituck, Camden, Chowan and Hertford Counties in the northeast corner of the state, the species show more overlap than appears in the Alligator River National Wildlife Refuge and Lake Mattamuskeet area in Dare and Hyde Counties. The other counties making up this peninsular region west of the Alligator River NWR only show the presence of the southern short-tailed shrew. For the eastern counties south of the Pamlico River, the species of *Blarina* appear to overlap more with some locations showing both species in close proximity.

DISCUSSION

Short-tailed shrews (genus *Blarina*) live in all the major ecoregions and, presumably, all the counties in North Carolina. In the western part of the state, the northern and southern short-tailed shrews (*B. brevicauda* and *B. carolinensis*) are parapatric in their distributions based on museum records. They separate roughly by the delineation between the Blue Ridge and Piedmont

ecoregions. In the Middle Atlantic Coastal Plain and the southern part of the Southeastern Plains, the two species of short-tailed shrews appear to have overlapping distributions in select areas. *B. b. knoxjonesi* was designated as a disjunct population of the northern short-tailed shrew in southeastern North Carolina but, recently collected specimens show that the distribution of the supspecies is not truly disjunct from that of *B. b. talpoides* in northeastern North Carolina.

The goal of my research was to examine the current taxonomy of the short-tailed shrews of the genus *Blarina* in North Carolina with the aid of the phenetic species concept. The results of my morphological analyses on skulls and mandibles from short-tailed shrew specimens clearly separate *Blarina brevicauda* from *B. carolinensis*. The analyses comparing the subspecies *B. b. knoxjonesi* and *B. b. talpoides* show overlap in individual measurements (i.e. principal components analysis – PCA and linear discriminant function analysis – LDFA) that involve a combination of measurements. The branching in the classification tree analysis equally had overlap between the 'knoxjonesi' and 'talpoides' classification names and the variables and measurements used to split them. Although the skull characters of UTR and UCL (upper toothrow and unicuspid length) showed a significant variation with age, UTR was not weighted heavily in the PCA or the classification tree, but UTR was heavily weighted in the LDFA species split and UTR and UCL was heavily weighted with the LDFA subspecies split. The combination of these tooth row characters with other skull characters, however, did not change the overall interpretation of the multivariate analyses. Therefore, I propose that *Blarina brevicauda knoxjonesi* should be synonymized with *B. b. talpoides*, and that *B. b. knoxjonesi* does not represent a distinct morphological subspecies.

In the systematic revision of the northern short-tailed shrew (*Blarina brevicauda*), Webster et al. (2011) used operational taxonomic units (OTUs) that combined a sample of specimens for a

particular geographic region. The OTUs have the advantage of comparing variation across a large geographic spatial area. The disadvantage of OTUs is that if small, continuous variation exists, the average value of a trait or traits may show a difference when one does not exist across a localized spatial area. Webster et al. (2011) also considered samples that showed "abrupt changes in size or morphological features between populations of *B. brevicauda*" to "indicate evolutionary significance and to warrant taxonomic recognition". The classification tree analysis shows a gradual change when you look closer at the locations of the samples split as detailed in Appendix II.

As with most of the multivariate analyses, *Blarina brevicauda* and *B. carolinensis* split almost perfectly. *B. c. carolinensis* continues north from North Carolina to south-central Virginia and south into Florida (Bennedict et al., 2006; Genoways and Choate; 1998; McCay, 2001; Webster, 1985). The largest skulls of *B. carolinensis* in North Carolina are generally in the Coastal Plain ecoregion and near the center of their range in the state, while the smaller skulls of *B. carolinensis* tend to be at edges of their range in the state. NCSM 16800 is from Morganton in Burke County and represents a county record and the western-most specimen in North Carolina. This is an old adult male with extensive tooth wear and dental abnormalities, but is smaller than most specimens toward the center of the state's distribution. The location of this individual was within the Piedmont "peninsula" created by the Blue Ridge ecoregion around the area (Figure 2.16).

The subspecies *B. b. knoxjonesi* and *B. b. talpoides* have a fair amount of overlap in the size of the skulls. With the skull data, the specimens identified as *B. b. talpoides* that split in the classification tree analysis with the 'talpoides' group were mostly from the Blue Ridge ecoregion at the highest elevation for the largest skulls, and in the NE Coastal Plain for the smallest skulls.

Other small skulls that grouped with ‘talpoides’ group were found near the eastern edge of the Blue Ridge and the Alligator River National Wildlife Refuge region north of the Pamlico River. The specimens identified as *B. b. knoxjonesi* that also were in the ‘talpoides’ group were primarily near the center of the Southeastern and Coastal Plain ecoregion north of the Cape Fear River. The largest specimens identified as *B. b. knoxjonesi* that were in the ‘knoxjonesi’ group were also near the center of the known subspecies distribution.

What can account for this difference in size?

Population densities are often greatest at the core of a distribution and, thus, factors like competition, environmental interactions, resource partitioning, life history and demographics influence dispersal (Brown, 1984; Brown et al., 1996). This can cause the larger, and more fit, individuals to be found at the core of a distribution. Inter- and intraspecific competition also affects where shrew species exist. Competition appears to have led to partitioning of habitat and food sources among the shrew species, creating differences in niche size and shape (Churchfield, 1991; Kirkland, 1991; Brannon, 2000). Fox and Kirkland (1992) demonstrated that soricid functional groups are defined by their body size, and that the largest body size yields a competitive advantage for the highest quality food. If more than one species exists in the same time and space, there must be a way to reduce this competition. Kirkland (1991) found that five or more shrew species existed in some areas. These shrew assemblages vary with latitude and environmental moisture (Berman et al., 2007), elevation and forest type (Ford et al., 2005) or diet and vertical foraging mode (McCay et al., 2004). Therefore, these factors allow shrew species to coexist by partitioning habitat and resources.

Generally, body size increases with latitude for most vertebrates (McNab, 1971). This rule is called Bergmann's rule. This rule does not apply to many mammal species due to factors like temperature, range in elevation, precipitation and land cover (Blackburn and Hawkins, 2004; McNab, 1971). At low latitude, sizes of some mammal species can be large where temperature is low, where elevation is high, or where annual precipitation is high. The largest skulls in my analysis were from the highest elevations in North Carolina, where temperatures get the coldest and land cover is continuous forest.

Geology and climate may also have an important impact on the coexistence of *B. brevicauda* and *B. carolinensis* in eastern North Carolina. The dunes and Carolina Bays may have resources that may allow the two species of shrews to coexist (Soller, 1988). These physical features present relatively dry upland areas to be adjacent to wet regions (Webster, 1996). The eastern coast of North Carolina has higher annual precipitation, especially in July through September, than the rest of the state, creating moist soil moisture (http://climate.ncsu.edu/climate/monthlyprecip.html). These two factors can partition both the land and resources to promote coexistence of the two *Blarina* species.

Potential issues with georeferencing

Species distribution modelling is used increasingly in both applied and theoretical research to predict how species are distributed and to understand environmental requirements affecting them. With this modelling, species occurrence data are combined with spatial data to predict suitability of any location for that species. While data sharing initiatives (i.e. VertNet, IDigBio and GBIF) involving species' occurrences have increased over the past few years,

various data quality and methodological concerns, related to using these data for species distribution modelling, have not been addressed adequately (Graham et al., 2008).

Georeferencing museum voucher specimens is only as good as the data provided by the collector. Using locality coordinates whose GPS precision or uncertainty is at the county level may not give enough detail for habitat modeling. Coordinates derived from a cell phone usually have greater error than coordinates derived from most GPS units, but still offer adequate precision for many applications (Zandbergen, 2009, Zandbergen and Barbeau, 2011). Changes in best practices for georeferencing museum bio-collections are continually improving the standard and quantity of localities that can be georeferenced in a batch (Chapman and Wieczorek, 2006; Rios and Bart, 2010; Wieczorek et al., 2004). These improvements will continue to make collection specimens more valuable. The latitude and longitude data for recent specimens in the NCSM and UNCW mammal collections represent the best practices coordinates with the given data. The point radius uncertainty for the coordinates of these samples reflect a measurement error that was the best practice from when it was determined, but does not reflect a maximum error distance that adds additional uncertainties (Wieczorek and Wieczorek, 2015).

LITERATURE CITED

Albrecht, G.H. 1980. Multivariate analysis and the study of form, with special reference to canonical variate analysis. American Zoologist 20(4): 679-693.

Baumgardner, G. D., N. O. Dronen and D. J. Schmidly. 1992. Distributional status of short-tailed shrews (genus *Blarina*) in Texas. The Southwestern Naturalist 37(3): 326-328.

Benedict, R. A. 1999. Morphological and mitochondrial DNA variation in a hybrid zone between short-tailed shrews (*Blarina*) in Nebraska. Journal of Mammalogy 80(1): 112-134.

Benedict, R. A., H. H. Genoways, and J. R. Choate. 2006. Taxonomy of short-tailed shrews (genus *Blarina*) in Florida. Occasional Papers, Museum of Texas Tech University 251: 1-19.

Berman, J., T. McCay and P. Scull. 2007. Spatial analysis of species richness of shrews (Soricomorpha: Soricidae) in North America north of Mexico. Acta Theriologica 52(2): 151-158.

Berzal, F., J. C. Cubero, F. Cuenca and M. J. Martín-Bautista. 2003. On the quest for easy-to-understand splitting rules. Data & Knowledge Engineering 44(1): 31-48.

Blackburn, T. M. and B. A. Hawkins. 2004. Bergmann's rule and the mammal fauna of northern North America. Ecography 27: 715-724.

Brannon, M. P. 2000. Niche relationships of two syntopic species of shrews, *Sorex fumeus* and *S. cinereus*, in the southern Appalachian Mountains. Journal of Mammalogy 81(4): 1053-1061.

Braun, J. K. and M. L. Kennedy. 1983. Systematics of the genus *Blarina* in Tennessee and adjacent areas. Journal of Mammalogy 64(3): 414-425.

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1984. Classification and regression trees. Wadsworth International Group. Belmont, California.

Bronner, G. N., M. V. D. Merwe and K. Njobe. 2007. Nongeographic cranial variation in to medically important rodents from South Africa, *Mastomys natalensis* and *Mastomys coucha*. Journal of Mammalogy 88(5): 1179-1194.

Brown, James H. 1984. On the relationship between abundance and distribution of species. American Naturalist 124(2): 255-279.

Brown, James H., Stevens, George C., & Kaufman, Dawn M. 1996. The geographic range: size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics 27: 597-623.

Chapman, A.D. and J. Wieczorek (eds). 2006. Guide to best practices for georeferencing. Copenhagen: Global Biodiversity Information Facility.

Choate, J. R. 1972. Variation within and among populations of short-tailed shrew in Connecticut. Journal of Mammalogy 53(1): 116-128.

Churchfield, S. 1990. The natural history of shrews. Comstock Publishing Associates, Ithaca, New York. 178 pp.

Churchfield, S. 1991. Niche dynamics, food resources, and feeding strategies in multispecies communities of shrews. Pp. 23-34, in The Biology of the Soricidae (J. S. Findley and T. L. Yates, eds.), Special Publication of the Museum of Southwestern Biology 1: 1-91.

DeLorme. 2006. Topo USA version 6.0. Two DeLorme Drive. P.O. Box 298. Yarmouth, ME, 04096. USA.

DeLorme. 2013. Topo North America version 10.0. Two DeLorme Drive. P.O. Box 298. Yarmouth, ME, 04096. USA.

Diersing, V. E. 1980. Systematics and evolution of the pygmy shrews (subgenus *Microsorex*) of North America. Journal of Mammalogy 61(1): 76-101.

Ellis, L. S., V. E. Diersing and D. F. Hoffmeister. 1978. Taxonomic status of short-tailed shrews (*Blarina*) in Illinois. Journal of Mammalogy 59(2): 305-311.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(2): 179-188.

Ford, W. M., T. S. McCay, M. A. Menzel, W. D. Webster, C. H. Greensberg, J. F. Pagels and J. F. Merritt. 2005. Influence of elevation and forest type on community assemblage and species distribution of shrews in the central and southern Appalachian Mountains. Special Publication of the International Society of Shrew Biologists 1: 303-315.

Fox, B. J. and G. L. Kirkland, Jr. 1992. An assembly rule for functional groups applied to North American soricid communities. Journal of Mammalogy 73(3): 491-503.

French, D. D., L. K. Corbett and N. Easterbee. 1988. Morphological discriminants of Scottish wildcats (*Felis silvestris*), domestic cats (*F. catus*) and their hybrids. Journal of Zoology 214(2): 235-259.

Garland, D. A. and G. A. Heidt. 1989. Distribution and status of shrews in Arkansas. Proceedings of the Arkansas Academy of Science 43: 35-38.

Genoways, H. H. and J. R. Choate. 1972. A multivariate analysis of systematic relationships among populations of the short-tailed shrew (genus *Blarina*) in Nebraska. Systematic Zoology 21(1): 106-116.

Genoways, H. H. and J. R. Choate. 1998. Natural history of the southern short-tailed shrew, *Blarina carolinensis*. Occasional Papers, The Museum of Southwestern Biology 8: 1-43.

George, S. B., J. R. Choate and H. H. Genoways. 1981. Distribution and taxonomic status of *Blarina hylophaga* Elliot (Insectivora: Soricidae). Annals of Carnegie Museum, 50(21): 493-513.

George, S. B., J. R. Choate and H. H. Genoways. 1986. *Blarina brevicauda*. Mammalian Species 261: 1-9.

Glahn, H. R. 1968. Canonical correlation and its relationship to discriminant analysis and multiple regression. Journal of the Atmospheric Sciences 25(1): 23-31.

Graham, C. H., J. Elith, R. J. Hijmans, A. Guisan, A. Townsend Peterson, A. and B. A. Loiselle. 2008. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45(1): 239-247.

Guilday, J. E. 1957. Individual and geographic variation in *Blarina brevicauda* from Pennsylvania. Annals of Carnegie Museum 35: 41-68.

Hafner, M. S., W. L. Gannon, J. Salazar-Bravo and S. T. Alvarez-Castañedo. 1997. Mammal collections in the Western Hemisphere: a survey and directory of existing collections. Allen Press, Lawrence, Kansas. 93 pp.

Hall, E. R. 1981. The Mammals of North America. 2nd edition. John Wiley and Sons, New York. 1181 pp.

Handley, C. O., Jr., and M. Varn. 1994. Identification of the Carolinian shrews of Bachman 1837. Pp. 393-406, in Advances in the biology of shrews (J. F. Merritt, G. L. Kirkland, Jr., and R. K. Rose, eds.). Special Publications, Carnegie Museum of Natural History, 18: 1-458.

Hastie, T., A. Buja and R. Tibshirani. 1995. Penalized discriminant analysis. The Annals of Statistics 23(1): 73-102.

Holmes, M. W., G. K. R. Boykins, R. C. K. Bowie and E. A. Lacey. 2016. Cranial morphological variation in *Peromyscus maniculatus* over nearly a century of environmental change in three areas of California. Journal of Morphology 277(1): 96-106.

Hotelling, H. 1936. Relations between two sets of variates. Biometrika 28(3/4): 321-377.

Jackson, D.A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology 74(8): 2204-2214.

James, F. C. and C. E. McCulloch. 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box. Annual Review of Ecological Systems 21: 129-166.

Johnson, R. A. and D. W. Wichern. 2007. Applied multivariate statistical analysis. 6th edition. Pearson Prentice Hall, Upper Saddle River, New Jersey. 773 pp.

Kirkland, G. L., Jr. 1991. Competition and coexistence in shrews (Insectivora: Soricidae). Pp. 15-22, in The Biology of the Soricidae (J. S. Findley and T. L. Yates, eds.), Special Publication of the Museum of Southwestern Biology 1: 1-91.

Kirkland, G. L. Jr., A. M. Wilkinson, J. V. Planz and J. E. Maldonado. 1987. *Sorex* (*Microsorex*) *hoyi* in Pennsylvania. Journal of Mammalogy 68(2): 384-387.

Kullback, S. and L. A. Leibler. 1951. On information and sufficiency. The annals of mathematical statistics 22(1): 79-86.

Kullback, S. 1959. Information theory and statistics. John Wiley. New York.

Lee, D. S., J. B. Funderburg, Jr. and M. K. Clark. 1982. A distributional survey of North Carolina mammals. Occasional Papers of the North Carolina Biological Survey. 72 pp.

Loh, W-Y. 2011. Classification and regression trees. WIREs Data Mining and Knowledge Discovery 1: 14-23.

Lydeard, C. and M. L. Kennedy. 1988. Mophological assessment of recently founded populations of the coyote, *Canis latrans*, in Tennessee. Journal of Mammalogy 69(4): 773-781.

McCay, T. S. 2001. *Blarina carolinensis*. Mammalian Species 673: 1-7.

McCay, T. S., M. J. Lovallo, W. M. Ford and M. A. Menzel. 2004. Assembly rules for functional groups of North American shrews: effects of geographical range and habitat partitioning. Oikos 107: 141-147.

McNab, B. K. 1971. On the ecological significance of Bergmann's rule. Ecology 52: 845-854.

Metzler, E. H. 1992. The comprehensive survey of Ohio moths and butterflies. The Ohio Lepidopterists Project No. NGSCW-91-11.

Metzler, E. H. 1994. Scale of relative certainty. Association of Systematics Collections Newsletter 22(5).

Michener, C. D. 1970. Diverse approaches to systematics. Journal of Evolutionary Biology 4: 1-38.

Mitteroecker, P. and F. Bookstein. 2011. Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. Evolutionary Biology 38(1): 100-114.

Moncrief, N. D., J. R. Choate and H. H. Genoways. 1982. Morphometric and geographic relationships of short-tailed shrews (genus *Blarina*) in Kansas, Iowa, and Missouri. Annals of Carnegie Museum 51(9): 157-180.

Paul, J. R. 1965. *Blarina telmalestes* in North Carolina. Journal of Mammalogy 46(3): 496.

Patton, J.L. 1973. An analysis of natural hybridization between the pocket gophers, *Thomomys bottae* and *Thomomys umbrinus*, in Arizona. Journal of Mammalogy 54(3): 561-584.

Pearson, O. P. 1945. Longevity of the short-tailed shrew. The American Midland Naturalist 34: 531-546.

Quinlan, J. R. 1986. Induction of decision trees. Machine learning 1(1): 81-106.

Quinlan, J. R. 1993. C4.5: programs for machine learning. Morgan Kaufmann Publishers, Inc. San Francisco, California.

Raileanu, L. E. and K. Stoffel. 2004. Theoretical comparison between the Gini index and information gain criteria. Annals of Mathematics and Artificial Intelligence 41(1): 77-93.

Ridley, M. 1989. The cladistics solution to the species problem. Biology and Philosophy 4: 1-16.

Rios, N. E. and H. L. Bart. 2010. GEOLocate (Version 3.22) [Computer software]. Belle Chasse, LA: Tulane University Museum of Natural History.

Rippy, C. L. 1967. The taxonomy and distribution of the short-tailed shrew, *Blarina brevicauda*, in Kentucky. M. S. thesis, University of Kentucky, Lexington, 83 pp.

RStudio and Inc. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/ and https://www.rstudio.com/.

Schmidly, D. J. and W. A. Brown. 1979. Systematics of short-tailed shrews (genus *Blarina*) in Texas. The Southwestern Naturalist 24(1): 39-48.

Simpson, G. G. 1961. Principles of animal taxonomy. Columbia University Press, New York. 253 pp.

Smith, A. B. 1994. Systematics and the fossil record: documenting evolutionary patterns. Blackwell Science, Cambridge, Massachussetts.

Sneath, P. H. A. and R. R. Sokal. 1973. Numerical taxonomy. 2nd edition. W. H. Freeman, San Francisco, California. 573 pp.

Sokol, R. R. and T. J. Crovello. 1970. The biological species concept: a critical evaluation. The American Naturalist 104: 127-153.

Soller, D. R. 1988. Geology and tectonic history of lower Cape Fear River valley, Southeastern North Carolina. U.S. Geological Survey Professional Paper 1466-A. 60 pp.

Strobl, C., A. L. Boulesteix and T. Augustin. 2007. Unbiased split selection for classification trees based on the Gini index. Computational Statistics and Data Analysis 52(1): 483-501.

Strobl, C. J. Malley and G. Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods 14(4): 323-348.

Tate, C. M., J. F. Pagels and C. O. Handley, Jr. 1980. Distribution and systematic relationship of two kinds of short-tailed shrews (Soricidae: *Blarina*) in south-central Virginia. Proceedings of the Biological Society of Washington 93(1): 50-60.

Therneau, T. M. and E. J. Atkinson. 1997. An introduction to recursive partitioning using the RPART routine. Technical Report 61, Mayo Clinic, Section of Biostatistics, Rochester, Minnesota, USA.

Webster, W. D., J. F. Parnell, and W. C. Biggs, Jr. 1985. The mammals of the Carolinas, Virginia, and Maryland. University of North Carolina Press, Chapel Hill, 255 pp.

Webster, W. D. 1996. Geographic variation in *Blarina brevicauda* (Insectivora: Soricidae) from eastern North Carolina, with description of a new subspecies. Pp. 47-56, in Contributions in mammalogy: a memorial volume honoring Dr. J. Knox Jones, Jr. (H. H. Genoways and R. J. Baker, eds.). Museum of Texas Tech University, Lubbock, 315 pp.

Webster, W. D., N. D. Moncrief, J. R. Choate and H. H. Genoways. 2011. Systematic revisions of the northern short-tailed shrew, *Blarina brevicauda* (Say). Virginia Museum of Natural History. Memoir 10. 77 pp.

Wieczorek, J., Q. Guo and R. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. International Journal of Geographical Information Science 18(8): 745-767.

Wieczorek C, J Wieczorek (2015) Georeferencing Calculator (version 20151221). Museum of Vertebrate Zoology, University of California, Berkeley. Available: http://manisnet.org/gci2.html.

Wilson, D. E. and D. M. Reeder (editors). 2005. Mammal species of the world: a taxonomic and geographic reference. 3rd edition. The Johns Hopkins University Press, Baltimore, Maryland. 2142 pp.

Zandbergen, P. A. 2009. Accuracy of iPhone locations: a comparison of assisted GPS, WiFi and cellular positioning. Transactions in GIS 13(s1): 5-26.

Zandbergen, P. A and S. J. Barbeau. 2011. Positional accuracy of assisted GPS data from high-sensitivity GPS-enabled mobile phones. Journal of Navigation 64(3): 381-399.

Table 2.1: List of data recorded from museum records for each specimen examined from North Carolina.

Data	Description
Museum Acronym	University of North Carolina at Wilmington (UNCW) or North Carolina Museum of Natural Sciences (NCSM)
Museum Number	Museum number associated with respective museum acronym (e.g. NCSM 18)
Genus	*Blarina*
Species	*brevicauda* or *carolinensis*
subspecies	*talpoides, knoxjonesi* or *carolinensis*
Sex	M = male or F = female
Reproduction	Testes (length x width) or female reproductive status
State	NC = North Carolina
County	county within state (e.g. Wake County)
Locality	major or minor town within county (e.g. Raleigh)
Latitude	Decimal degrees from World Geodetic System 1984 (WGS84) derived from verbal locality information in DeLorme topographical software (e.g. 35.780102)
Longitude	Decimal degrees from WGS84 derived from verbal locality information in DeLorme topographical software (e.g. -78.638820)
Elevation	Elevation in meters for given coordinates (e.g. 109 meters)
Confidence	Point radius uncertainty for locality (e.g. 25470 meters)
Date	Date specimen was collected (e.g. 23 June 1975)
Collector	Full name, or portion indicated with voucher specimen data
Field Number	Other number and acronym associated with specimen
Total Length	TL = Total length – distance from tip of nose to the end of the tail
Tail Vertebrae	TV = Tail vertebrae – distance from the sacral/caudal vertebrae junction to the last caudal vertebrae
Hind Foot	HF = Hind foot – greatest distance from calcaneus to the distal phalange excluding the nail
Ear	E = Ear length – greatest distance from ear attachment to tip of pinna
Weight	Wt = Weight – measured in grams for specimen prior to preparation
Nature of Specimen	Nat_of_spec = material collected and retained in collection: SS = skin and skull, SO - skin only, SK = skull only, SN = skull and skeleton

Table 2.2: Age classification criteria for each short-tailed shrew skull examined. Every skull was characterized by the amount of: tooth wear evident, upper incisor root exposure, cranial suture closure, and crest pronounced. Qualitative characteristics were recorded at the time of skull examination, and converted to a quantitative scale with a detailed description indicated below. Each of the 4 observations were combined to determine the age class where: 4-9 = subadult, 10-19 = adult, 20-23 = old adult.

Qualitative characteristic	Quantitative conversion	Detailed description of aging criteria based on each respective qualitative characteristic
Tooth wear evident – especially in skull, unicuspids are first 5 teeth after incisor, followed by molariforms		
Little to no	1	perfect points evident on all teeth especially to unicuspids
Little	2	almost all teeth have perfect points, little flattening to first 2 unicuspids
Little to some	3	some wear evident to first 2 unicuspids, some wear noted on 3rd and 4th
Some	4	some rounding to tips of all unicuspids, very little wear noted on incisor and molariform teeth
Some to moderate	5	some to a medium level of roundness to unicuspid tips, little wear noted on incisor and molariform teeth
Moderate	6	medium wear to tips of incisors with evidence to incisors and molariform
Moderate to excessive	7	relative flattening to cusps of all teeth
Excessive	8	excessive flattening and wear evident on all teeth especially molariform
Upper incisor root exposure – upper incisor will change angle with wear resulting in tooth root being visible		
Little to no	1	no root evident
Little	2	little root of upper incisor visible
Little to some	3	little to some root evident – no change in angle of incisor
Some	4	some root evident – little change in angle of incisor
Some to moderate	5	some root evident – some change in angle of incisor
Moderate	6	noticeable root evident – change in angle of incisor
Moderate to excessive	7	noticeable to large amount of root evident – close to vertical angle of incisor
Excessive	8	large amount of root evident – almost complete vertical angle of incisor
Cranial suture closure – looking at suture between parietal bones and parietal-occipital suture		
open	1	cranial sutures almost completely open, little cartilage evident and parietal/occipital bones very thin/translucent
< 1/2 closed	2	cranial sutures mostly open with excessive cartilage evident
>1/2 closed	3	cranial sutures mostly closed with some cartilage evident
closed with cartilage evident	4	cranial sutures completely closed with very little cartilage evident
fused (no cartilage evident)	5	cranial sutures completely fused with no signs of cartilage, parietal/occipital bones opaque
Crest pronounced - sagittal crest (parietal-parietal suture) and nuchal crest (parietal-occipital suture)		
not pronounced	1	no rise in sagittal crest, or nuchal crest
lightly pronounced	2	slight rise to sagittal crest, no rise in nuchal crest
pronounced	3	rise to sagittal crest, and slight rise to nuchal crest
excessively pronounced	4	rise to sagittal crest, and rise to nuchal crest

Table 2.3: List of confidence intervals used for georeferenced specimens, and point radius conversion. The latitude and longitude are in the center of the interval, and the confidence shape (circular or linear) is determined by the written locality associated with the museum voucher specimens. The contents were taken from the Scale of Relative Certainty (with slight modification) which was written by Dr. Eric H. Metzler and published in the Association of Systematics Collections October 1994/ vol. 22 no. 5 newsletter.

Confidence Code	Description	Point radius (meters)
S0	Location known to be within a circle 200 foot in diameter	30.48
S1	Location known to be within a circle 0.25 mile in diameter	201.168
S2	Location known to be within a circle 0.5 mile in diameter	402.336
S3	Location known to be within a circle 1 mile in diameter	804.672
S4	Location known to be within a circle 2 miles in diameter	1609.34
S5	Location known to be within a circle 5 miles in diameter	4023.36
S6	Location known to be within a circle 15 miles in diameter	12070.1
S7	Location known to be within county or marine area 50 miles in diameter	40233.6
S8	Location known to be within ½ of the state or marine area 100 miles in diameter	80467.2
S9	Location known to be within the state or marine are 500 miles in diameter	402336
S10	Location unknown	
L0	Precise location known	
L1	The linear site is known to be no more than 0.5 mile long	402.336
L2	The linear site is known to be no more than 1 mile long	804.672
L3	The linear site is known to be no more than 2 miles long	1609.34
L4	The linear site is known to be no more than 6 miles long	4828.03
L5	The linear site is known to be no more than 25 miles long	20116.8
L6	The linear site does not exceed a distance equal to ½ the state	40233.6
L7	The linear site does not exceed a distance equal to diameter of the state	

Table 2.4: Descriptive statistics of measurements from museum specimens between species of short-tailed shrews (*Blarina brevicauda* and *B. carolinensis*) in North Carolina. Description of external measurement data provided in Table 1.1, and description of cranial measurement data provided in Figure 1.4. For each character measurement, the mean, standard deviation, and the minimum and maximum range in parentheses are provided. A Welch two sample t-test compared the means and a F-test compared the variances for the measured characters with the p-values provided.

	Blarina brevicauda **n = 328**	***Blarina carolinensis*** **n = 163**	**Comparison of Means**		**Comparison of Variance**	
			t statistic	**p-value**	**F statistic**	**p-value**
TL	110.8 ± 8.58 (83 – 145)	92.6 ± 6.25 (78 – 112)	23.6654	< 2.2e-16	1.8831	9.13e-05
TV	21.61 ± 3.09 (10 – 30)	16.76 ± 2.67 (10 – 23)	15.8955	< 2.2e-16	1.3351	0.06831
HF	14.04 ± 1.70 (7 – 20)	11.16 ± 0.81 (10 – 13)	22.7617	< 2.2e-16	4.4361	< 2.2e-16
E	3.758 ± 1.84 (1 – 8)	1.846 ± 1.41 (1 – 6)	3.7911	0.0007112	1.7104	0.3238
Wt	13.88 ± 4.04 (4.8 – 27)	8.417 ± 1.74 (4.5 – 14)	14.9646	< 2.2e-16	5.3926	2.22e-15
GRL	22.69 ± 0.73 (20.63 – 24.51)	19.54 ± 0.45 (18.31 – 20.73)	57.5649	< 2.2e-16	2.5618	2.351e-10
OPML	21.7 ± 0.69 (19.77 – 23.55)	18.71 ± 0.45 (17.67 – 19.7)	56.2325	< 2.2e-16	2.3144	1.129e-08
MAB	7.417 ± 0.39 (6.46 – 8.49)	6.528 ± 0.25 (5.84 – 7.24)	30.2713	< 2.2e-16	2.3886	2.375e-09
IOB	5.827 ± 0.24 (5.18 – 6.47)	5.112 ± 0.17 (4.76 – 5.53)	37.8022	< 2.2e-16	2.0125	1.288e-06
CRB	11.67 ± 0.61 (10.06 – 13.17)	10.13 ± 0.35 (9.1 – 10.95)	35.0516	< 2.2e-16	3.066	9.637e-14
CRH	6.94 ± 0.33 (5.94 – 7.83)	6.09 ± 0.22 (5.46 – 6.64)	33.4434	< 2.2e-16	2.2889	1.811e-08
UTR	10.33 ± 0.37 (9.16 – 11.25)	8.707 ± 0.25 (7.98 – 9.28)	56.4573	< 2.2e-16	2.0977	2.927e-07
MTL	5.944 ± 0.22 (5.12 – 6.5)	5.203 ± 0.16 (4.78 – 5.58)	42.2733	< 2.2e-16	1.9744	2.239e-06
UCL	2.924 ± 0.18 (2.5 – 3.47)	2.309 ± 0.15 (1.89 – 2.66)	39.76	< 2.2e-16	1.4482	0.008652
GrLgt	14.55 ± 0.54 (12.85 – 15.97)	12.32 ± 0.34 (11.33 – 13.04)	54.4597	< 2.2e-16	2.564	3.091e-10
MNL	12.66 ± 0.63 (11.24 – 14.75)	10.97 ± 0.39 (9.97 – 11.77)	35.5515	< 2.2e-16	2.6457	8.685e-11
MNH	6.49 ± 0.48 (5.52 – 7.72)	5.267 ± 0.19 (4.8 – 5.67)	39.1241	< 2.2e-16	6.7508	< 2.2e-16
ARB	2.352 ± 0.19 (1.9 – 2.91)	1.988 ± 0.11 (1.72 – 2.22)	26.2167	< 2.2e-16	3.2804	6.661e-15
CorHt	6.027 ± 0.46 (5.17 – 7.45)	4.882 ± 0.17 (4.41 – 5.24)	38.7004	< 2.2e-16	7.6057	< 2.2e-16
MdbTR	6.441 ± 0.21 (5.7 – 7.06)	5.623 ± 0.15 (5.28 – 6.05)	47.4088	< 2.2e-16	1.9403	6.134e-06

Table 2.5: Descriptive statistics of measurements from museum specimens between subspecies of northern short-tailed shrews (*Blarina brevicauda knoxjonesi* and *B. b. talpoides*) in North Carolina. Description of external measurement data provided in Table 1.1, and description of cranial measurement data provided in Figure 1.4. For each character measurement, the mean, standard deviation, and the minimum and maximum range in parentheses are provided. A Welch two sample t-test compared the means and a F-test compared the variances for the measured characters with the p-values provided.

	Blarina brevicauda knoxjonesi **n = 131**	***Blarina brevicauda talpoides*** **n = 197**	**Comparison of Means**		**Comparison of Variance**	
			t statistic	**p-value**	**F statistic**	**p-value**
TL	107.1 ± 8.22 (83 – 145)	113.1 ± 7.97 (89 – 131)	-5.9243	1.233e-08	1.0631	0.7217
TV	21.13 ± 2.77 (15 – 29)	21.91 ± 3.25 (10 – 30)	-2.0997	0.03679	0.7288	0.08361
HF	13.78 ± 1.49 (9 – 19)	14.2 ± 1.81 (7 – 20)	-2.0879	0.03784	0.6764	0.03308
E	3.5 ± 2 (2 – 8)	3.84 ± 1.82 (1 – 7)	-0.4276	0.6772	1.2097	0.6709
Wt	12.29 ± 3.69 (4.8 – 20.1)	15.2 ± 3.85 (6.4 – 27.0)	-4.958	1.803e-06	0.9199	0.7138
GRL	22.26 ± 0.60 (20.63 – 23.64)	22.99 ± 0.65 (20.93 – 24.51)	-10.2719	< 2.2e-16	0.8371	0.2832
OPML	21.28 ± 0.62 (19.77 – 22.59)	22.0 ± 0.57 (20.14 – 23.55)	-10.5136	< 2.2e-16	1.2008	0.2559
MAB	7.144 ± 0.28 (6.49 – 7.93)	7.604 ± 0.34 (6.46 – 8.49)	-13.1961	< 2.2e-16	0.722	0.04784
IOB	5.693 ± 0.21 (5.18 – 6.09)	5.92 ± 0.22 (5.37 – 6.47)	-9.4113	< 2.2e-16	0.93	0.6621
CRB	11.24 ± 0.46 (10.06 – 12.27)	11.97 ± 0.51 (10.13 – 13.17)	-13.2923	< 2.2e-16	0.8247	0.2438
CRH	6.775 ± 0.29 (5.94 – 7.50)	7.055 ± 0.30 (6.11 – 7.83)	-8.217	7.595e-15	0.9213	0.6227
UTR	10.11 ± 0.30 (9.16 – 10.99)	10.48 ± 0.33 (9.5 – 11.25)	-10.4621	< 2.2e-16	0.7913	0.1543
MTL	5.809 ± 0.20 (5.12 – 6.30)	6.036 ± 0.18 (5.51 – 6.5)	-10.2548	< 2.2e-16	1.1885	0.2783
UCL	2.857 ± 0.14 (2.52 – 3.21)	2.969 ± 0.19 (2.5 – 3.47)	-6.104	3.018e-09	0.5415	0.0002268
GrLgt	14.17 ± 0.42 (12.85 – 15.01)	14.81 ± 0.45 (13.45 – 15.97)	-12.8722	< 2.2e-16	0.857	0.3574
MNL	12.24 ± 0.52 (11.24 – 13.50)	12.95 ± 0.53 (11.59 – 14.75)	-11.7229	< 2.2e-16	0.9713	0.8676
MNH	6.122 ± 0.26 (5.52 – 6.71)	6.748 ± 0.43 (5.64 – 7.72)	-15.7994	< 2.2e-16	0.3719	1.136e-08
ARB	2.22 ± 0.13 (1.9 – 2.51)	2.445 ± 0.17 (1.99 – 2.91)	-12.993	< 2.2e-16	0.5698	0.0008975
CorHt	5.678 ± 0.23 (5.17 – 6.30)	6.271 ± 0.43 (5.37 – 7.45)	-15.7743	< 2.2e-16	0.2852	1.065e-12
MdbTR	6.333 ± 0.20 (5.70 – 7.06)	6.516 ± 0.19 (6.09 – 6.98)	-8.0689	2.784e-14	1.172	0.3295

Table 2.6: Principal components analysis table using the covariance matrix of the *Blarina* skull data (n = 448) showing what factors contribute to each principal component explaining the sample variation. The variables are the skull characters (Figure 1.4) measured, while the eigenvectors (e^_1) and correlation coefficients (ry^1) give the relative weight each variable contributes. The variance (lambda – λ_i) is an eigenvalue that is the measure of the amount of the variation explained by the principal component. Each principal component accounts for a smaller proportion of the total sample variation, but when combined, the first three principal components account for most of the total sample variation.

	Principal Component 1		**Principal Component 2**		**Principal Component 3**	
Variable	**e^_1**	**ry^1**	**e^_2**	**ry^2**	**e^_3**	**ry^3**
GRL	-0.6081996	-0.9954447	0.28116321	0.07315567	-0.012914768	-0.001912321
OPML	-0.5709861	-0.9933669	0.16148701	0.04466222	0.565968491	0.089079896
MAB	-0.1815011	-0.9000182	-0.38221556	-0.30129985	-0.332204653	-0.149032506
IOB	-0.1375705	-0.9137035	-0.13792909	-0.14563116	-0.099930963	-0.060045886
CRB	-0.3125942	-0.9232915	-0.78497008	-0.36857808	-0.003747593	-0.001001414
CRH	-0.1664465	-0.8926916	-0.16801498	-0.14324958	0.12250895	0.059442615
UTR	-0.3052743	-0.9719565	0.2493131	0.12618864	-0.625687557	-0.180225987
MTL	-0.143558	-0.9498013	0.03901157	0.04103148	-0.238136027	-0.142538979
UCL	-0.1139252	-0.9058453	0.14720746	0.18607265	-0.309758896	-0.222823636
Variance (λ_i)	7.191833631		0.181750951		0.058863514	
↘ % of Total Variance	0.949		0.024		0.0078	
∑ % of Total Variance	0.949		0.973		0.9808	

Table 2.7: Principal components analysis table using the covariance matrix of the *Blarina* mandible data (n = 461) showing what factors contribute to each principal component explaining the sample variation. The variables are the mandible characters (Figure 1.4) measured, while the eigenvectors (e^_1) and correlation coefficients (ry^1) give the relative weight each variable contributes. The variance (lambda – λ_i) is an eigenvalue that is the measure of the amount of the variation explained by the principal component. Each principal component accounts for a smaller proportion of the total sample variation, but when combined, the first three principal components account for most of the total sample variation.

	Principal Component 1		**Principal Component 2**		**Principal Component 3**	
Variable	**e^_1**	**ry^1**	**e^_2**	**ry^2**	**e^_3**	**ry^3**
GrLgt	-0.63385	-0.98445	-0.114244388	-0.03558	0.666212	0.16734
MNL	-0.51539	-0.95165	0.776334136	0.287464	-0.36248	-0.10825
MNH	-0.37717	-0.95927	-0.430006963	-0.21931	-0.39281	-0.16157
ARB	-0.11632	-0.87157	-0.14620436	-0.21969	-0.17897	-0.21687
CorHt	-0.35404	-0.95593	-0.421850626	-0.22842	-0.38239	-0.16698
MdbTR	-0.22688	-0.94307	0.003724787	0.003105	0.30367	0.20414
Variance (λ_i)	3.235202		0.130101955		0.084618	
↘ % of Total Variance	0.9300		0.0374		0.0243	
∑ % of Total Variance	0.9300		0.9674		0.9917	

Table 2.8: Character loadings for the skull measurements on the first coefficient of linear discriminants for short-tailed shrews in North Carolina. The variables in the left hand column are the skull characters measured (Figure 1.4), which is followed by a species comparison of *Blarina brevicauda* (Blbr) and *B. carolinensis* (Blca), and a subspecies comparison of *B. b. knoxjonesi* and *B. b. talpoides*. A comparison of the misclassification error rates, percent of specimens classified correctly and the ratio of between- and within-group standard deviations for the linear discriminant variables are also provided.

	Between species Blbr (n=297) and Blca (n=151)		**Between Blbr subspecies *knoxjonesi* (n=125) and *talpoides* (n=172)**	
Variable	**LD1**	**Factor importance**	**LD1**	**Factor importance**
GRL	0.62117339	7	-0.7969931	4
OPML	-1.41746699	3	0.45245974	7
MAB	1.20602585	5	0.73710139	5
IOB	-1.24350704	4	0.19236811	8
CRB	-0.04199313	9	1.18936612	3
CRH	-0.26935501	8	0.05970473	9
UTR	-2.49172784	1	1.2827919	1
MTL	1.7819057	2	0.51086301	6
UCL	-0.90606576	6	1.26408903	2
Misclassification Error Rate	0.004464286		0.1851852	
% Classified Correctly	99.55357		81.48148	
Between/Within Group	3251.456		241.3292	

Table 2.9: Character loadings for the mandible measurements on the first coefficient of linear discriminants for short-tailed shrews in North Carolina. The variables in the left hand column are the mandibular characters measured (Figure 1.4), which is followed by a species comparison of *Blarina brevicauda* (Blbr) and *B. carolinensis* (Blca), and a subspecies comparison of *B. b. knoxjonesi* and *B. b. talpoides*. A comparison of the misclassification error rates, percent of specimens classified correctly and the ratio of between- and within-group standard deviations for the linear discriminant variables are also provided.

	Between species Blbr (n=306) and Blca (n=155)		**Between Blbr subspecies *knoxjonesi* (n=126) and *talpoides* (n=180)**	
Variable	**LD1**	**Factor importance**	**LD1**	**Factor importance**
GrLgt	-1.9073451	1	0.3789902	4
MNL	0.1005763	6	0.7408481	3
MNH	-0.1062194	5	1.1420423	2
ARB	0.9377071	3	1.2651959	1
CorHt	0.377029	4	0.2647015	6
MdbTR	-1.7423685	2	-0.30235	5
Misclassification Error Rate	0.004338395		0.1666667	
% Classified Correctly	99.56616		83.33333	
Between/Within Group	2434.129		259.4218	

Table 2.10: Classification tree table with Gini splitting index from *Blarina* skull data (n = 448) showing what factors best split the data into subpecies classes. The subspecies of the southern short-tailed shrew (*Blarina carolinensis carolinensis* – Blcaca) and the northern short-tailed shrew (*B. brevicauda knoxjonesi* – Blbrkn and *B. b. talpoides* – Blbrta) represent the classes used while measured cranial characters (Figure 1.4) determine the primary split in the tree. Relative error (rel error), apparent error (x error) and apparent standard deviation (x std) are given for each complexity parameter to compute two measures of predictive performance using the root node error.

Tree Node #	**Blcaca**	**Blbrkn**	**Blbrta**	**Total at Node**	**Proportion at Node (Total / 448)**	**Complexity Parameter CP = (α)**	**Expected Loss = # of other classes / total at node**	**Predicted Class**	**Primary Split**	**Improvements = Gain = impurity reduction**
1	151	125	172	448	1.0000	0.5471	0.6161	talpoides	OPML < 19.735	151.4108
2	151	0	0	151	0.3371		0.0000	carolinensis		
3	0	125	172	297	0.6629	0.2065	0.4209	talpoides	CRB < 11.805	50.5258
4	0	117	60	177	0.3951	0.0199	0.3390	knoxjonesi	MAB < 7.165	11.8522
5	0	8	112	120	0.2679		0.0667	talpoides		
6	0	69	11	80	0.1786		0.1375	knoxjonesi		
7	0	48	49	97	0.2165	0.0199	0.4948	talpoides	OPML < 21.98	3.8201
8	0	44	34	78	0.1741	0.0199	0.4359	knoxjonesi	MAB < 7.565	3.4375
9	0	4	15	19	0.0424		0.2105	talpoides		
10	0	43	27	70	0.1563	0.0133	0.3857	knoxjonesi	OPML < 21.885	3.0446
11	0	1	7	8	0.0179		0.1250	talpoides		
12	0	12	1	13	0.0290		0.0769	knoxjonesi		
13	0	31	26	57	0.1272	0.0133	0.4561	knoxjonesi	UTR < 10.035	2.5474
14	0	10	2	12	0.0268		0.1667	knoxjonesi		
15	0	21	24	45	0.1004	0.0133	0.4667	talpoides	CRH < 6.835	3.9850
16	0	15	7	22	0.0491		0.3182	knoxjonesi		
17	0	6	17	23	0.0513		0.2609	talpoides		
						CP	**rel error**	**x error**	**x std**	
						0.5417	1	1	0.0373	
						0.2065	0.4529	0.4565	0.0345	
						0.0199	0.2464	0.2935	0.0295	
						0.0133	0.1848	0.2971	0.0297	
						0.0100	0.1449	0.2717	0.0286	

Root Node Error = 276/448 = 0.6161

Re-substitution Error Rate = 0.6161 * 0. 1449 = 0.0893 or 8.9% misclassified

Cross-validated Error Rate = 0.6161 * 0.2717 = 0.1674 or 16.7% misclassified

Table 2.11: Classification tree table with Gini splitting index from *Blarina* mandible data (n = 461) showing what factors best split the data into subpecies classes. The subspecies of the southern short-tailed shrew (*Blarina carolinensis carolinensis* – Blcaca) and the northern short-tailed shrew (*B. brevicauda knoxjonesi* – Blbrkn and *B. b. talpoides* – Blbrta) represent the classes used while measured cranial characters (Figure 1.4) determine the primary split in the tree. Relative error (rel error), apparent error (x error) and apparent standard deviation (x std) are given for each complexity parameter to compute two measures of predictive performance using the root node error.

Tree Node #	**Blcaca**	**Blbrkn**	**Blbrta**	**Total at Node**	**Proportion at Node (Total / 461)**	**Complexity Parameter (CP = α)**	**Expected Loss = # of other classes / total at node**	**Predicted Class**	**Primary Split**	**Improvements = Gain = impurity reduction**
1	155	126	180	461	1.0000	0.5516	0.6095	talpoides	GrLgt < 13.185	153.3737
2	155	2	0	157	0.3406		0.0127	carolinensis		
3	0	124	180	304	0.6594	0.2135	0.4079	talpoides	CorHt < 6.095	59.8798
4	0	121	61	182	0.3948	0.0605	0.3352	knoxjonesi	MNL < 12.94	14.5458
5	0	3	119	122	0.2646		0.0246	talpoides		
6	0	110	33	143	0.3102		0.2308	knoxjonesi		
7	0	11	28	39	0.0846		0.2821	talpoides		
						CP	**rel error**	**x error**	**x std**	
						0.5516	1.0000	1.0000	0.0373	
						0.2135	0.4484	0.4520	0.0341	
						0.0605	0.2349	0.2598	0.0279	
						0.0100	0.1744	0.2100	0.0255	

Root Node Error = 281/461 = 0.6095

Re-substitution Error Rate = 0.6095 * 0. 1744 = 0.1063 or 10.6% misclassified

Cross-validated Error Rate = 0.6095 * 0.2100 = 0.1280 or 12.8% misclassified

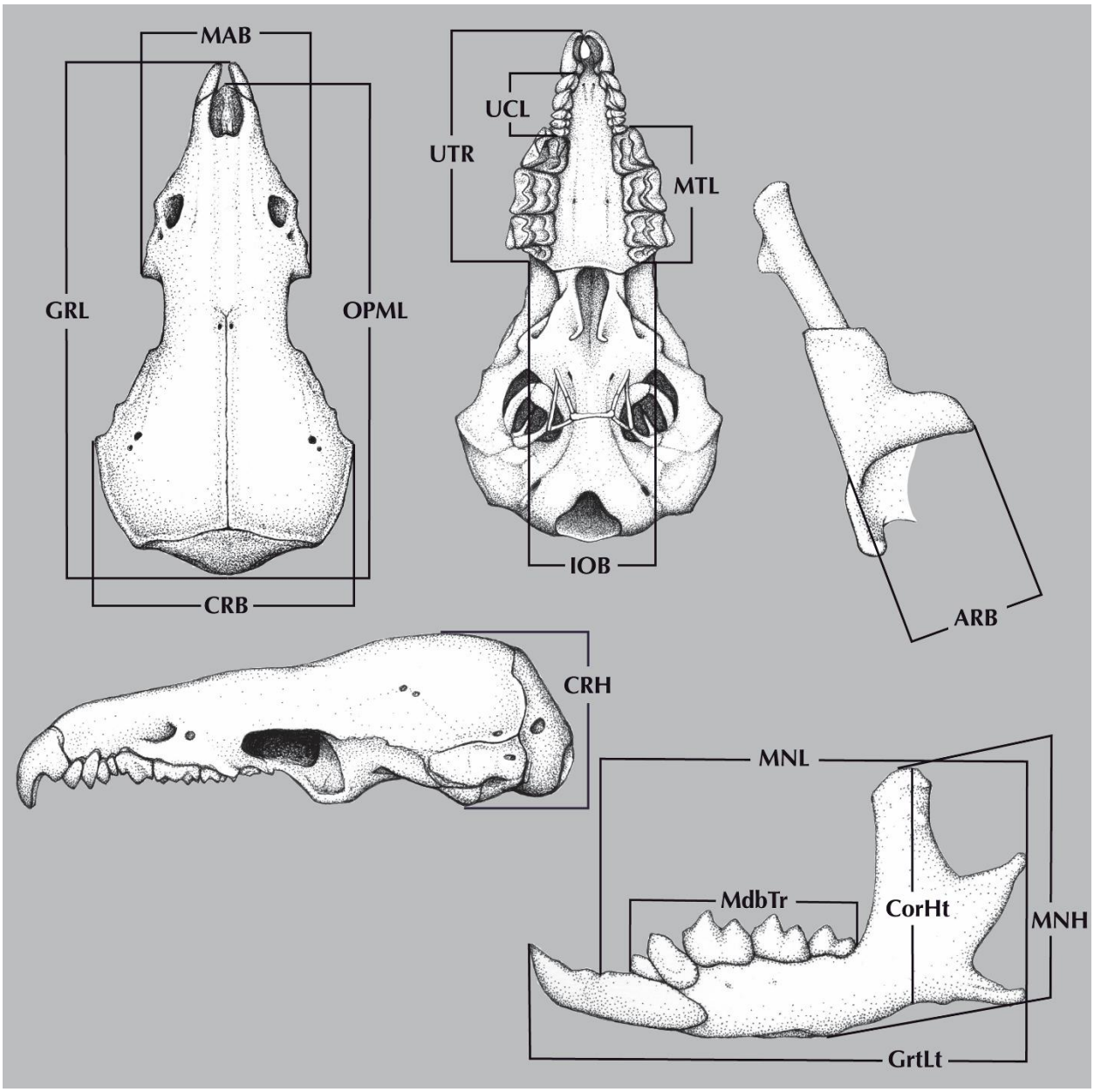
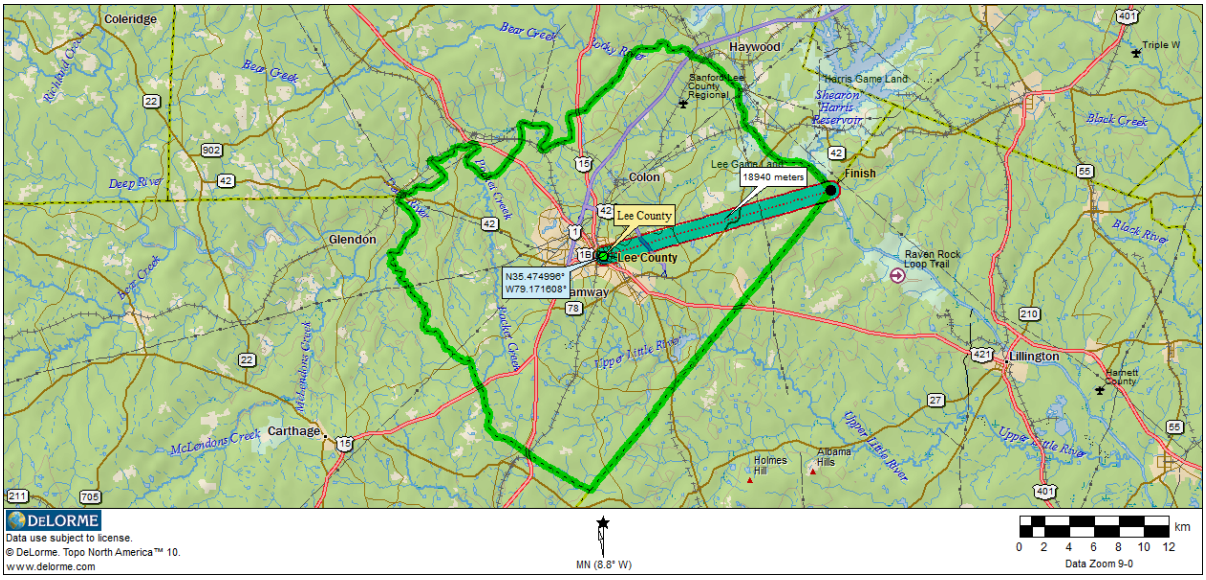
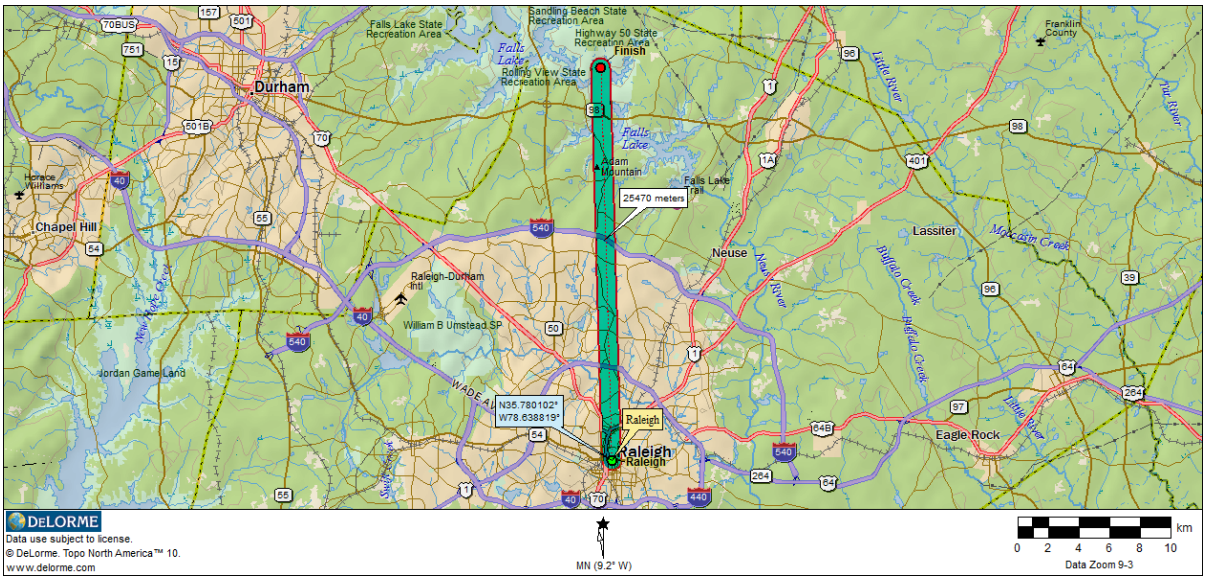


Figure 2.1: Description of the skull and mandible characters measured for each *Blarina* museum specimen sampled. Character descriptions: GRL = greatest skull length (includes incisor); OPML = occipital-premaxilla length (length minus incisor); MAB = maxillary breadth; IOB = interorbital breadth; CRB = cranial breadth; CRH = cranial height; UTR = upper toothrow (includes upper incisor); MTL = length of molariform toothrow (P4-M3); UCL = unicuspid toothrow length; GrtLt = greatest length of mandible with incisor; MNL = length of mandible without incisor; MNH = height of mandible; CorHt = coronoid process height; ARB = articular breadth; MdbTr = mandibular toothrow. Dorsal, ventral and lateral view of cranium and lateral view of lower jaw of *Blarina brevicauda* (NCSM 14408 male and NCSM 14409 female; NC, Polk County, Saluda, Green River Game Land). Drawn by L. Bradford. Layout by B. W. Wynne.

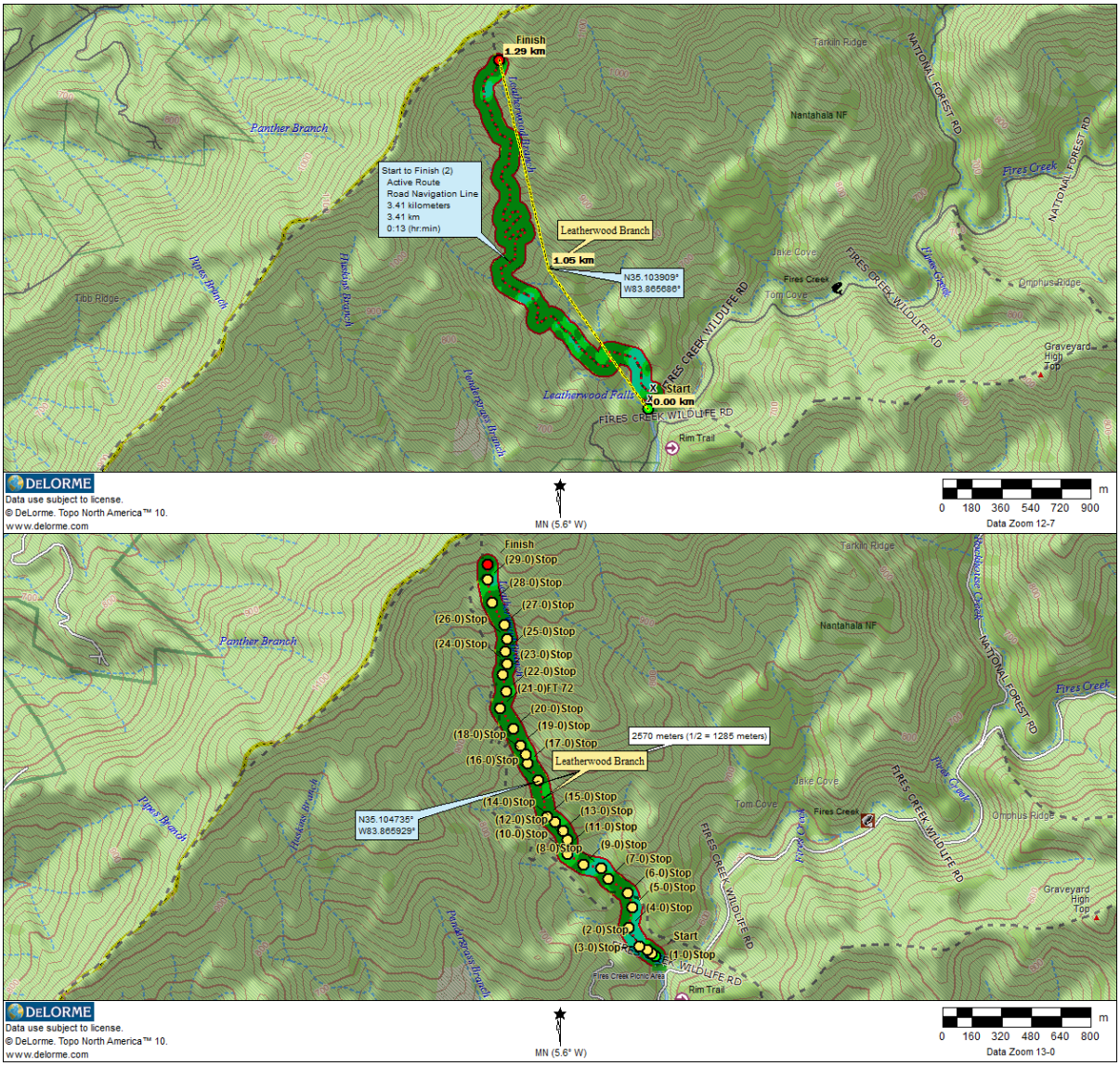


a. County only: Lat/Long taken at geographic center of Lee County NC. Point radius uncertainty is greatest extent of county.



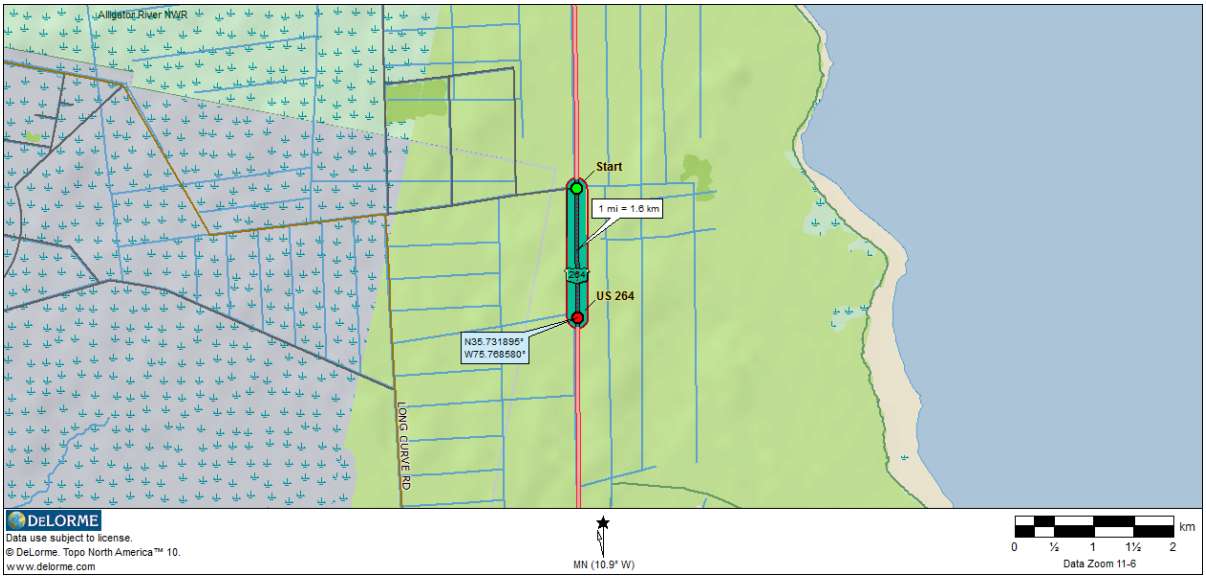
b. City, town or locale: Lat/Long taken at population center of Raleigh. Point radius uncertainty is greatest extent of locale.

Figure 2.2: Examples of point used for latitude and longitude coordinates, and confidence of localities using the mapping software Topo North America version 10.0 (DeLorme, 2013). Examples are: a. county only; b. city, town or locale; c. stream, creek or branch; d. road distance; and e. GPS points.

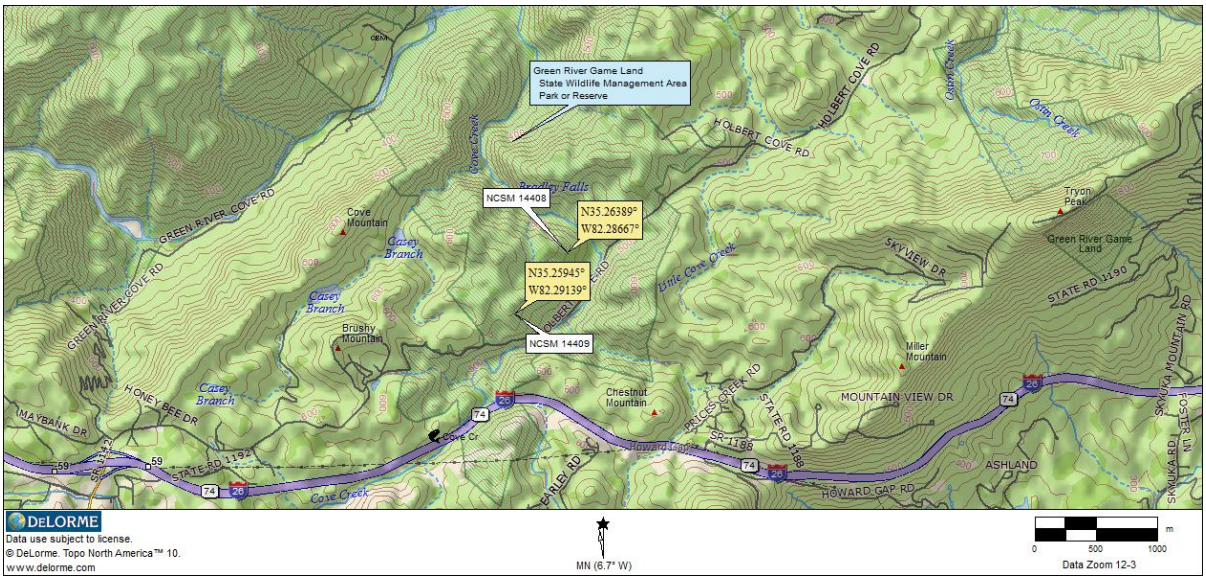


c. Stream, creek or branch: Lat/Long taken at approximate center of Leatherwood Branch. Point radius uncertainty is half the stream length.

Figure 2.2 (continued): Examples of point used for latitude and longitude coordinates, and confidence of localities using the mapping software Topo North America version 10.0 (DeLorme, 2013). Examples are: a. county only; b. city, town or locale; c. stream, creek or branch; d. road distance; and e. GPS points.

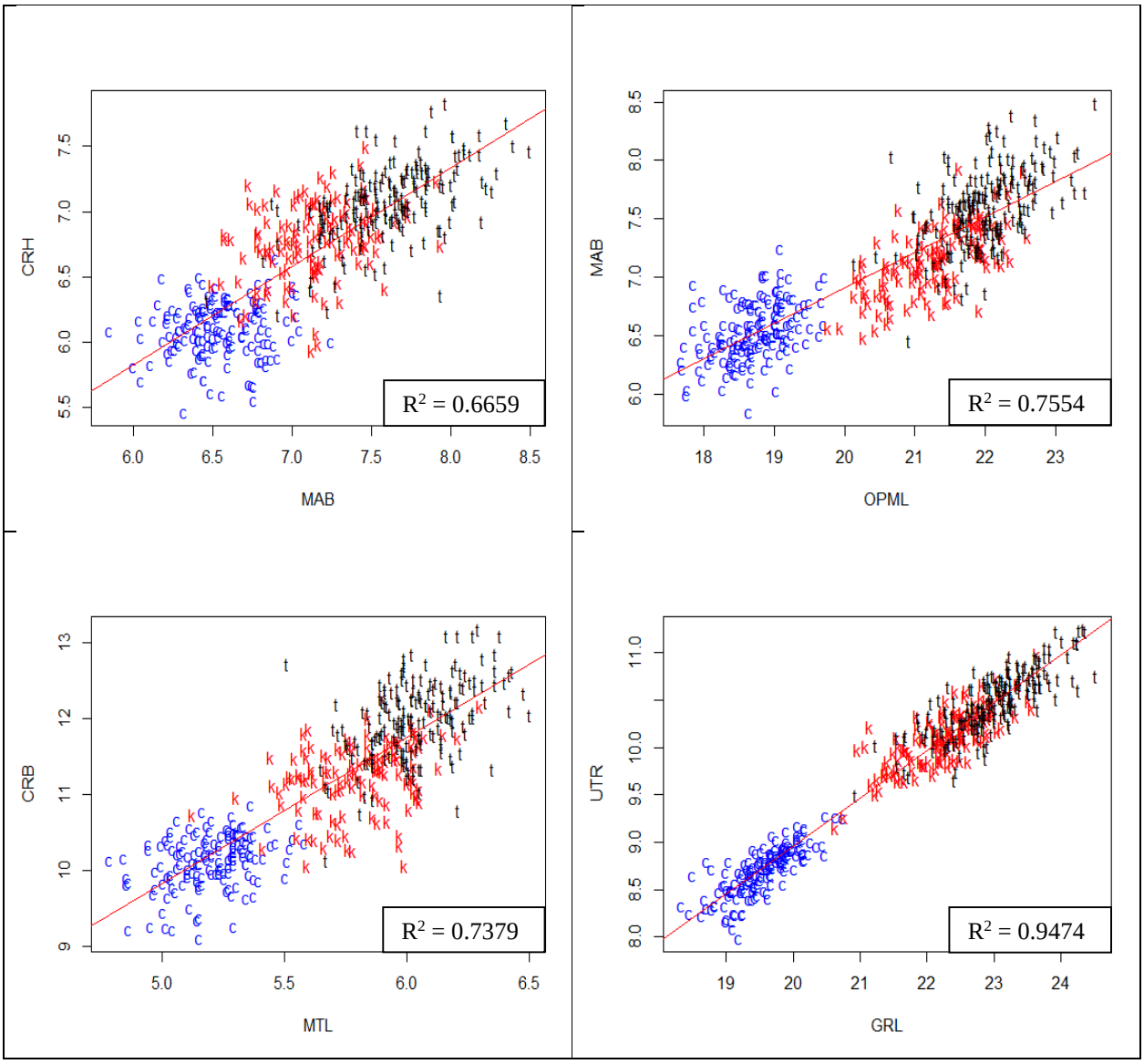


d. Road distance: Lat/Long taken on US 264, 1.0 road mile S from junction with Navy Shell Road. Point radius uncertainty is 1.0 mile (1609.34 meters).



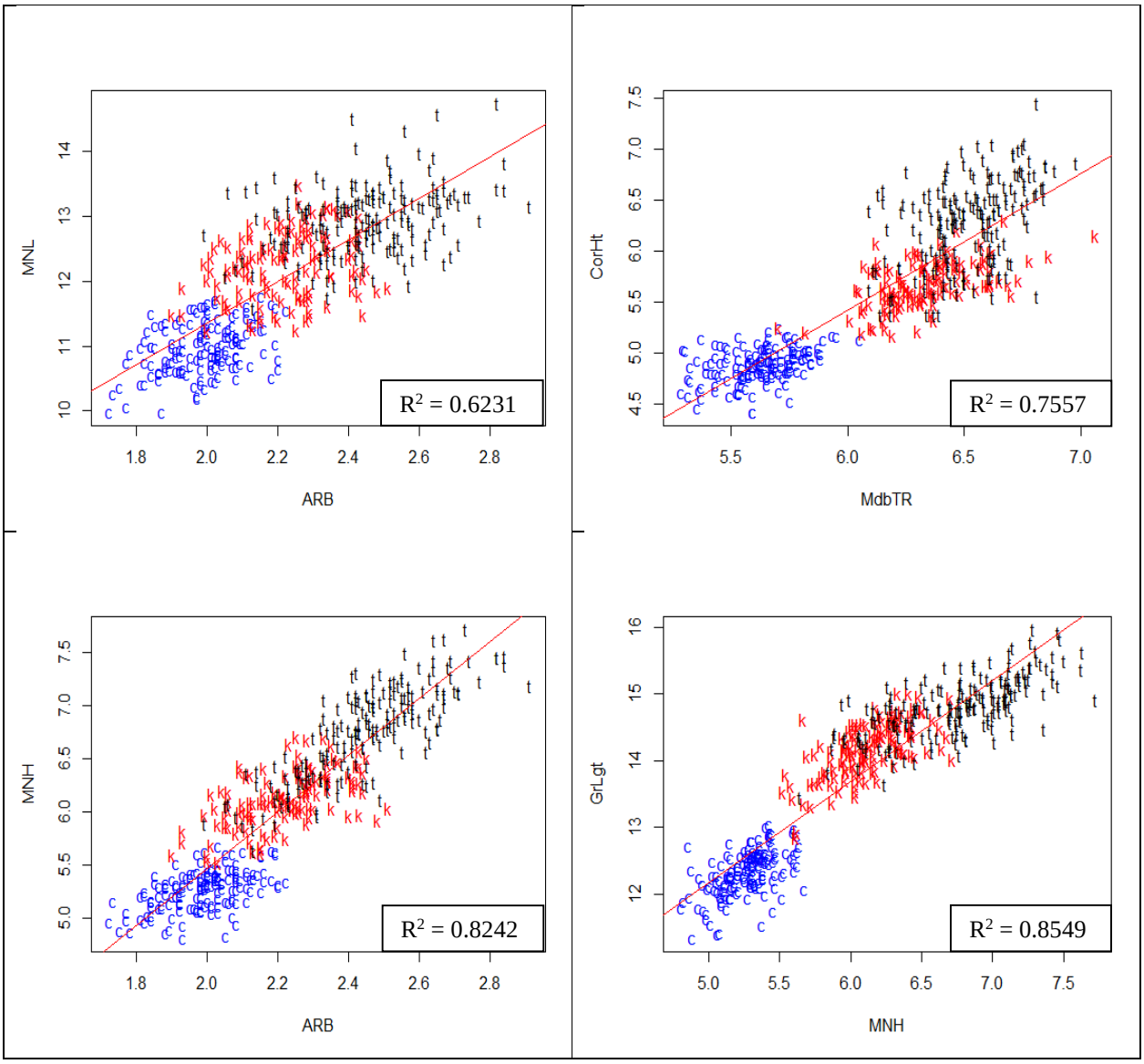
e. GPS points: Lat/Long taken with GPS – converted in DeLorme 10.0. Point radius uncertainty is average GPS precision = 10 meters.

Figure 2.2 (continued): Examples of point used for latitude and longitude coordinates, and confidence of localities using the mapping software Topo North America version 10.0 (DeLorme, 2013). Examples are: a. county only; b. city, town or locale; c. stream, creek or branch; d. road distance; and e. GPS points.



POOR GOOD

Figure 2.3: Bivariate plots of measurements taken from the skulls of short-tailed shrew specimens (genus *Blarina*) showing a poor and good split between species. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. The plots given as x-axis vs y-axis showing poor species splitting are: TOP LEFT = maxillary breadth (MAB) vs. cranial height (CRH), BOTTOM RIGHT = length of molariform toothrow (MTL) vs. cranial breadth (CRB); and the plots showing good species splitting are TOP RIGHT = occipital-premaxilla length without incisor (OPML) vs. maxillary breadth (MAB), BOTTOM LEFT = length of molariform toothrow (MTL) vs. OPML, BOTTOM RIGHT = greatest length of skull including incisor (GRL) vs. UTR. The regression line and multiple R-squared (R2) or coefficient of determination value gives information about how well the measure of one character can predict the measure of another character.



POOR GOOD

Figure 2.4: Bivariate plots of measurements taken from the mandibles of short-tailed shrew specimens (genus *Blarina*) showing a poor and good split between species. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. The plots given as x-axis vs y-axis are: TOP LEFT = articular breadth (ARB) vs. greatest length of mandible with incisor (GrLgt), TOP RIGHT = mandibular tooth row without incisor (MdbTR) vs. coronoid process height (CorHt), BOTTOM LEFT = CorHt vs GrLgt, BOTTOM RIGHT = mandible height (MNH) vs. GrLgt. The regression line and multiple R-squared (R^2) or coefficient of determination value gives information about how well the measure of one character can predict the measure of another character.

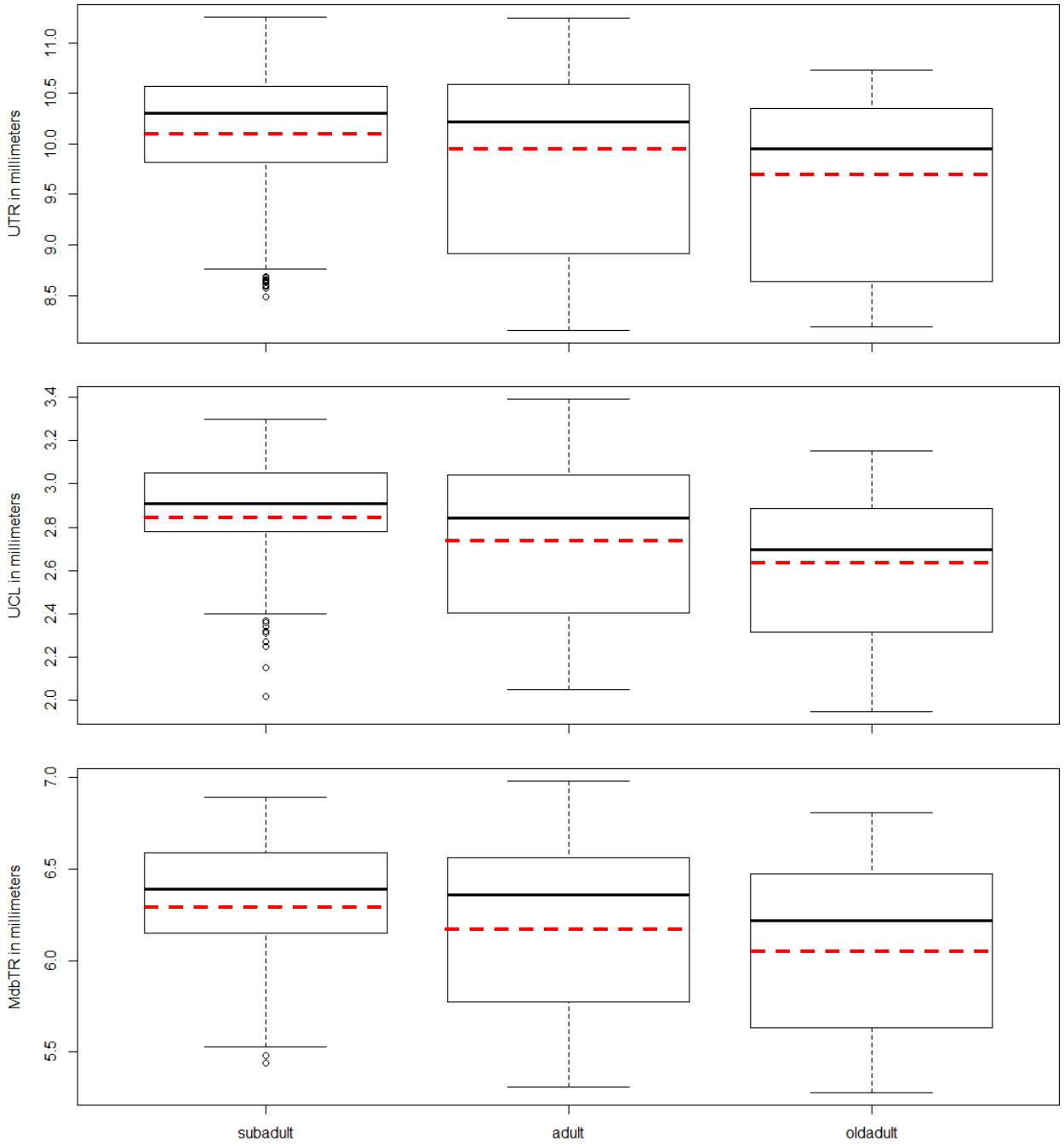


Figure 2.5: Boxplots of age (subadult: n=117, adult: n=151, oldadult: n=36) versus toothrow variables from skull and mandible characters. Each box splits the data into the first quartile (Q1), median (Q2) and third quartile (Q3) for the measured variables. The red dashed horizontal line represents the mean for each measurement within each category. The dashed vertical line extending beyond the box is the whisker that shows the smallest to largest non-outliers in the data to the horizontal solid line. In the subadult age category, the additional open dots are outliers.

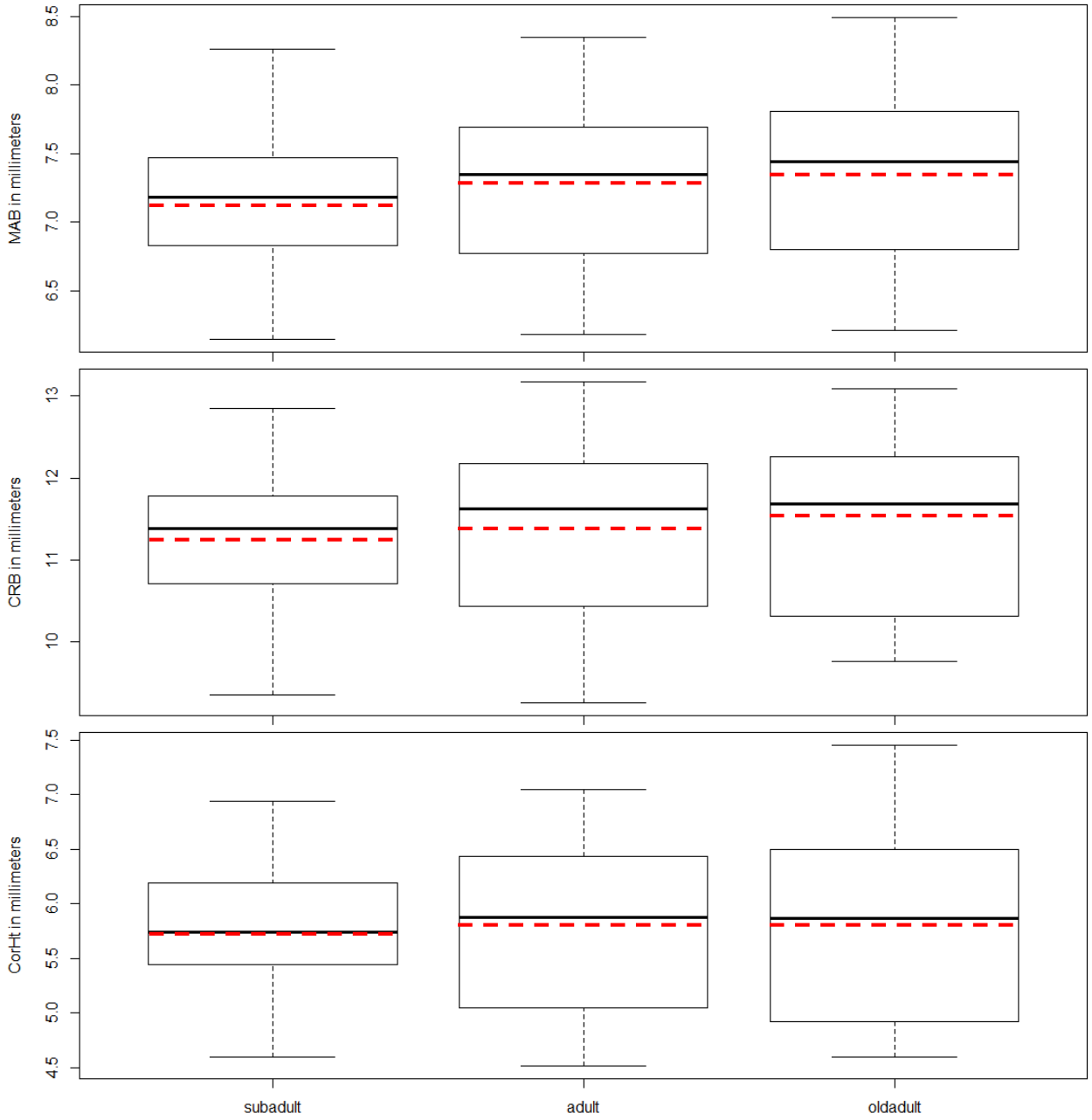


Figure 2.6: Boxplots of age (subadult: n=117, adult: n=151, oldadult: n=36) versus width and height variables from skull and mandible characters. Each box splits the data into the first quartile (Q1), median (Q2) and third quartile (Q3) for the measured variables. The red dashed horizontal line represents the mean for each measurement within each category. The dashed vertical line extending beyond the box is the whisker that shows the smallest to largest non-outliers in the data to the horizontal solid line. In the subadult age category, the additional open dots are outliers.

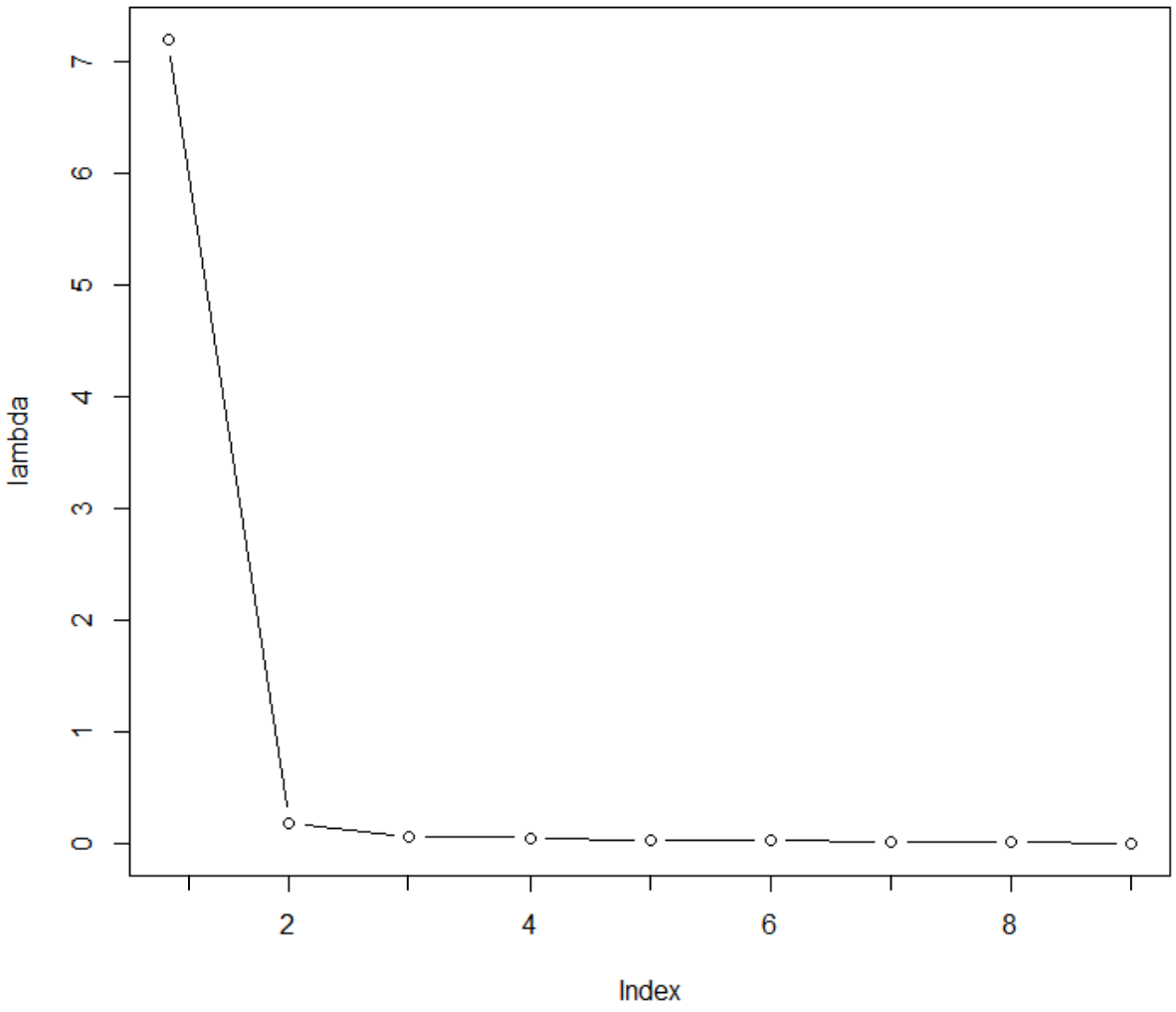


Figure 2.7: Scree plot from the covariance matrix calculations showing how many principal components are needed to explain the total sample variance of the *Blarina* skull data. This is a plot of the magnitude of an eigenvalue (i.e. lambda = measure of the amount of variation explained by the principal component) versus the principal component number (i.e. Index = PC1, PC2, PC3, etc.). An elbow occurs in the plot at the index number 2, where after the second eigenvalue, the remaining eigenvalues are small and about the same size. In this case, two principal components effectively summarize the total sample variance.

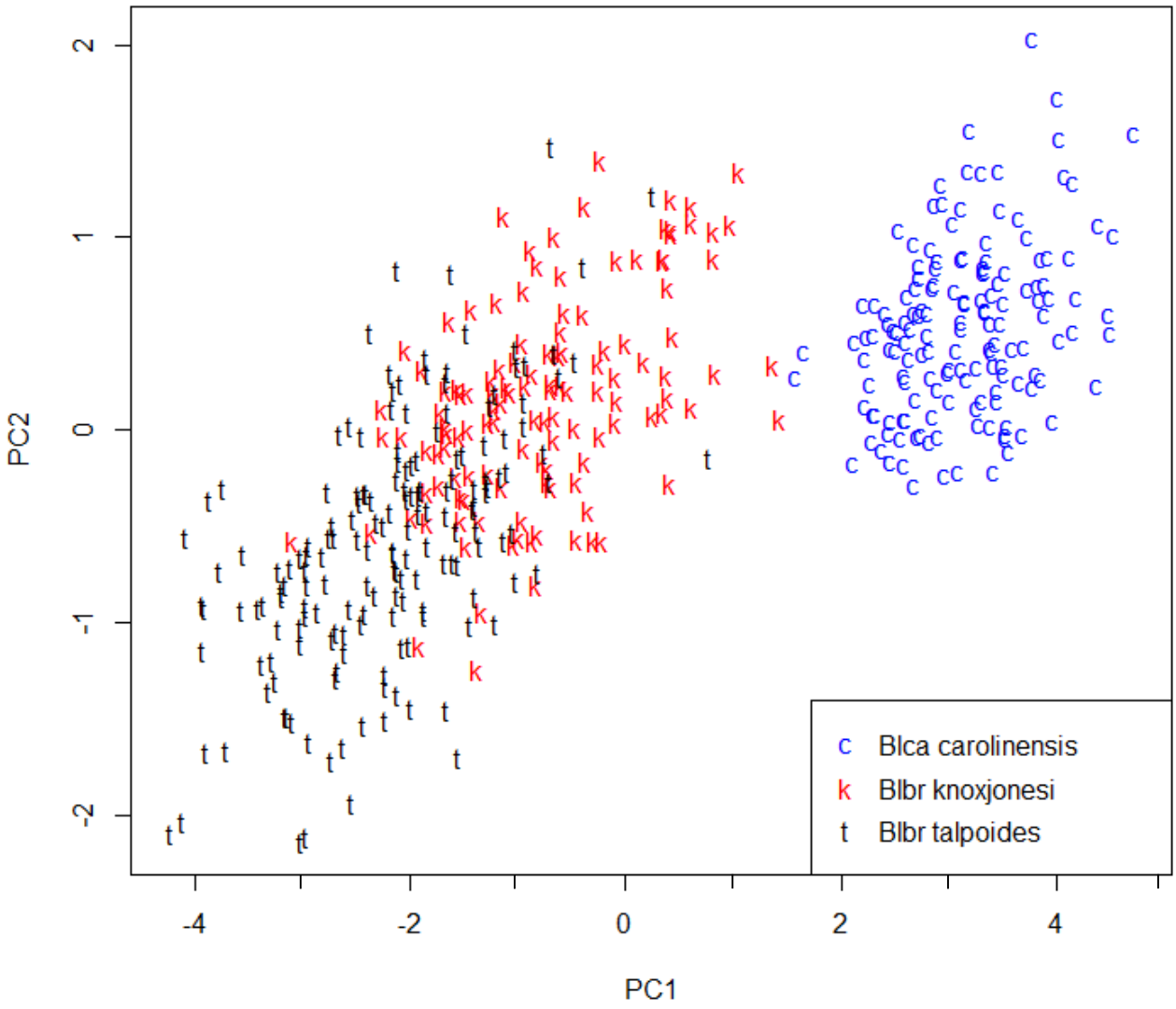


Figure 2.8: Scatter plot of the first two principal components derived from the covariance matrix for the *Blarina* skull data. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. The first principal component (PC1) accounts for 94.9% of the total sample variance, while the second principal component (PC2) accounts for 2.4%, for a combination of 97.3% of the total sample variance explained by PC1 and PC2.

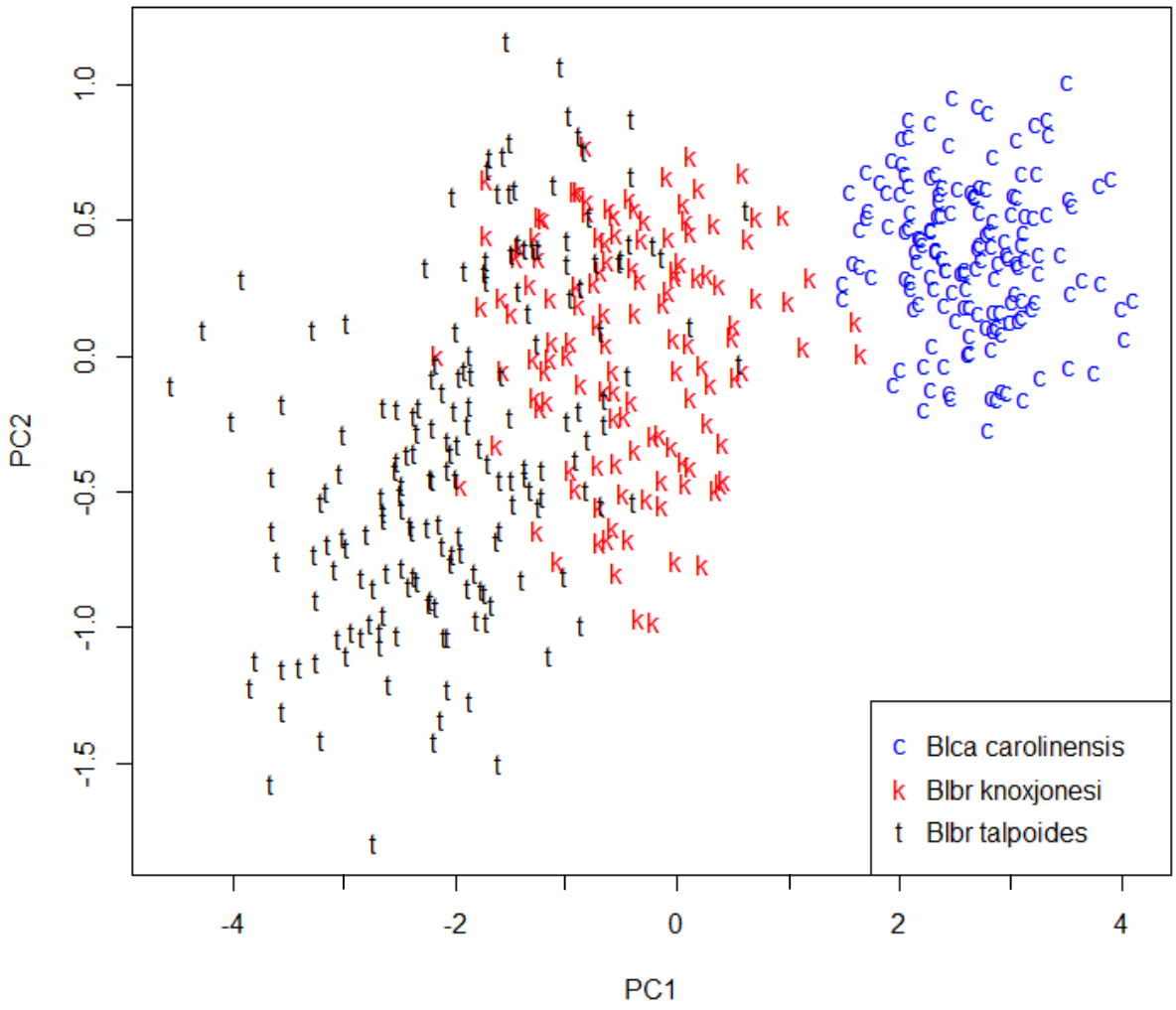


Figure 2.9: Scatter plot of the first two principal components derived from the covariance matrix for the *Blarina* mandible data. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. The first principal component (PC1) accounts for 93.0% of the total sample variance, while the second principal component (PC2) accounts for 3.7%, for a combination of 96.7% of the total sample variance explained by PC1 and PC2.

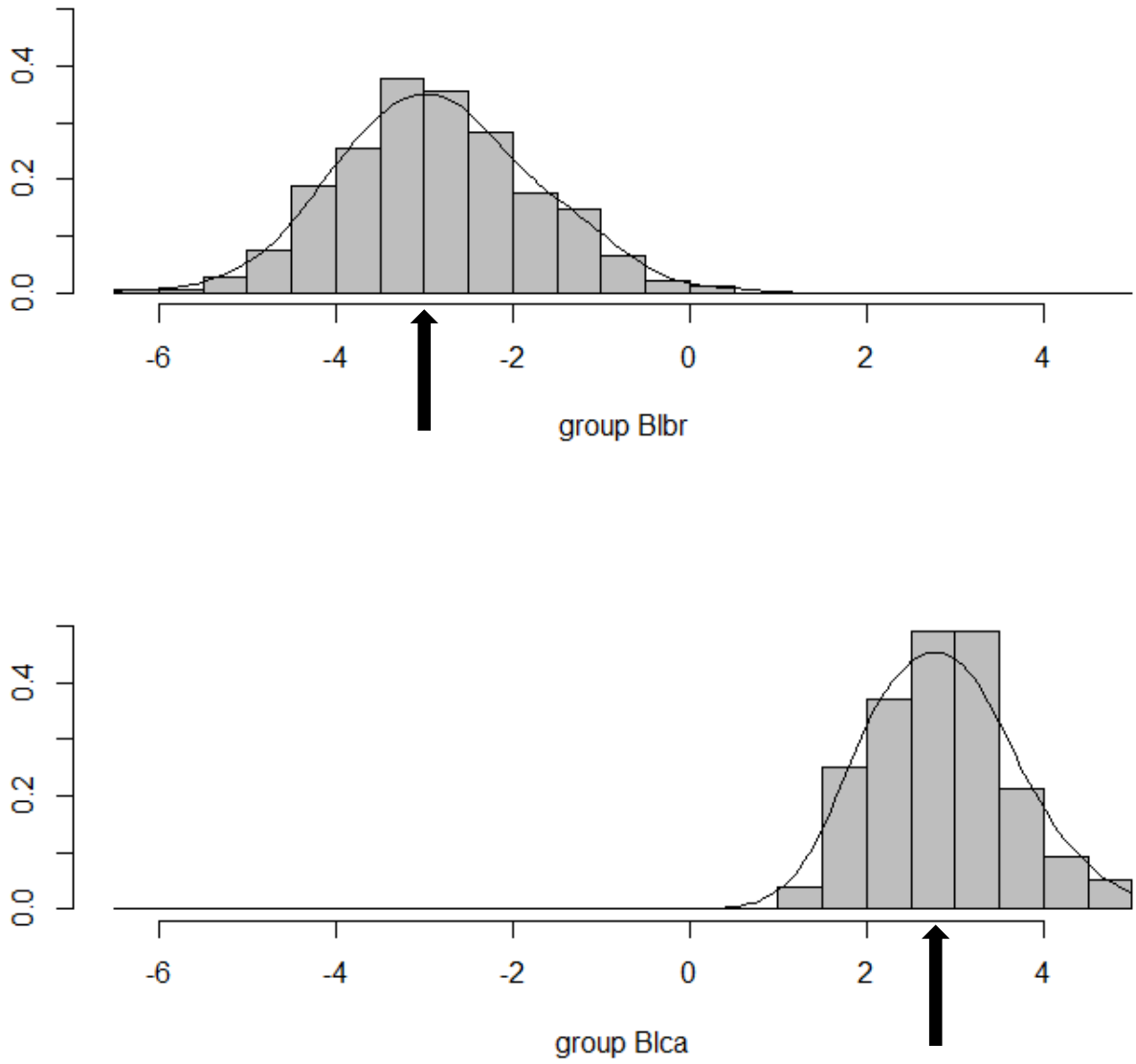


Figure 2.10: Linear discriminant function analysis histogram comparing the first discriminant function versus the frequency for the skull measurements of short-tailed shrew species. The histogram on top represents *Blarina brevicauda* (group Blbr) and the histogram on bottom represents *B. carolinensis* (group Blca). The arrows represent the approximate group centroid placement.

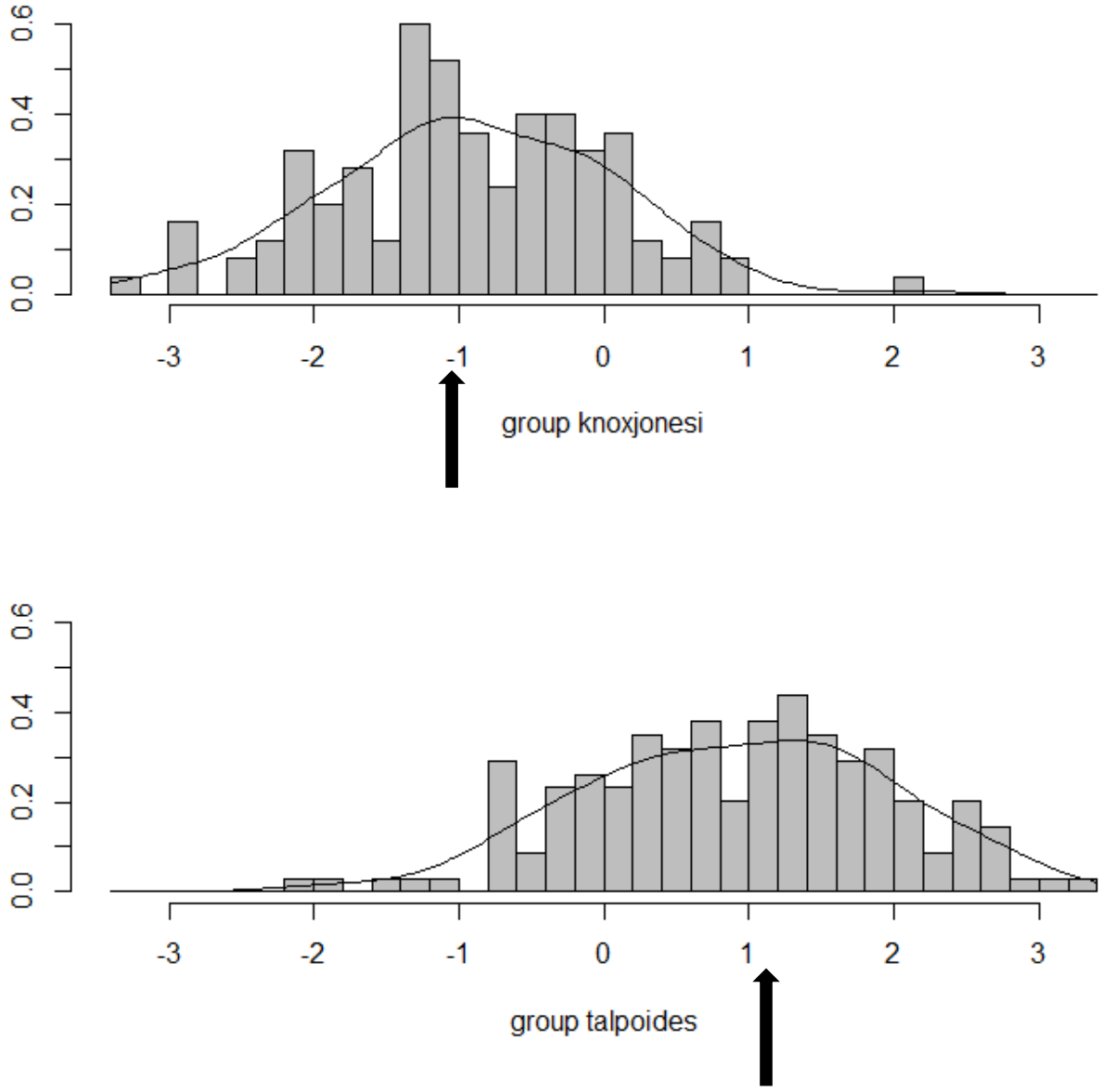


Figure 2.11: Linear discriminant function analysis histogram comparing the first discriminant function versus the frequency for the skull of the northern short-tailed shrew subspecies. The histogram on top represents *Blarina brevicauda knoxjonesi* (group knoxjonesi) and the histogram on bottom represents *B. b. talpoides* (group talpoides). The arrows represent the approximate group centroid placement.

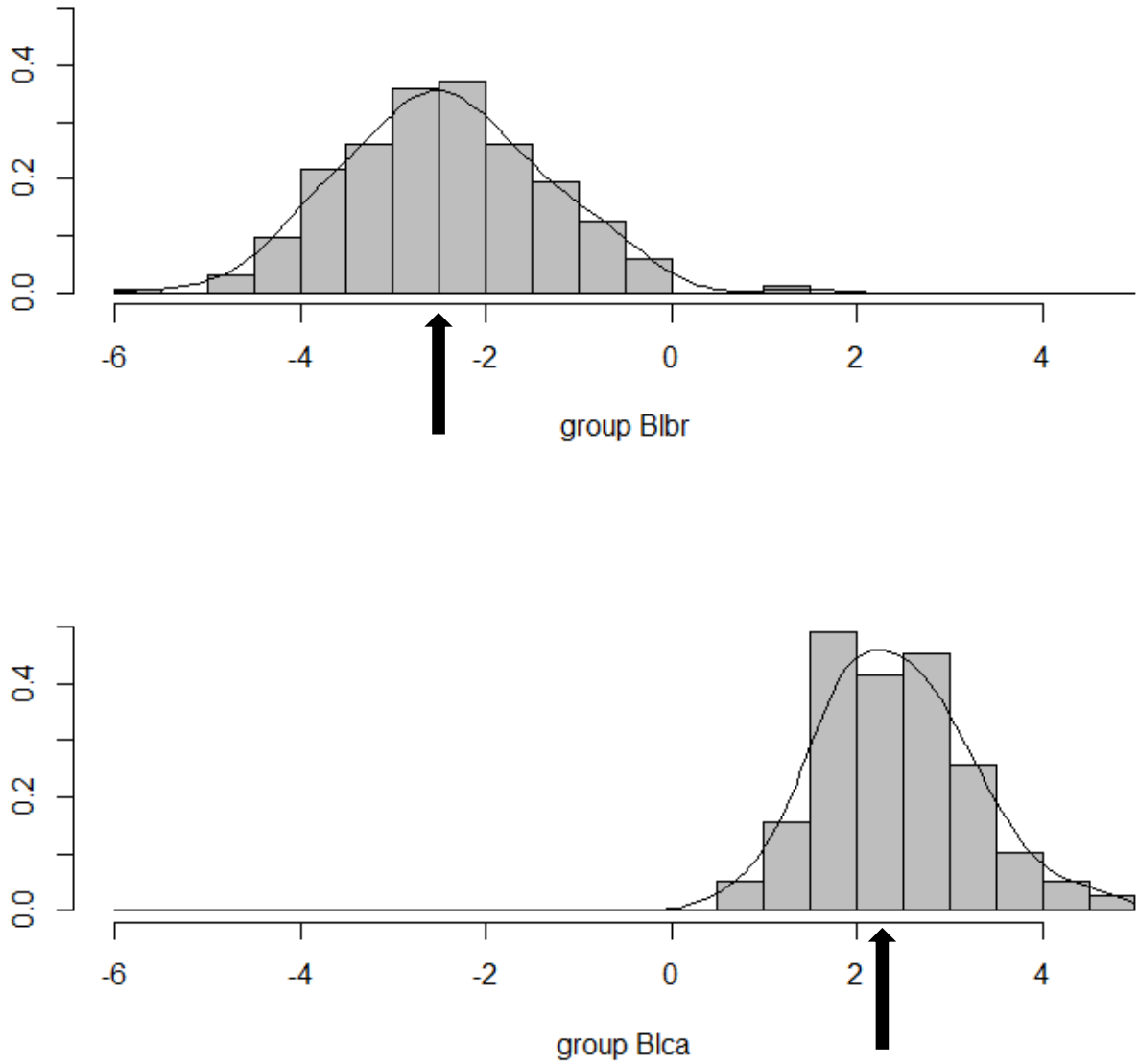


Figure 2.12: Linear discriminant function analysis histogram comparing the first discriminant function versus the frequency for the mandible measurements of short-tailed shrew species. The histogram on top represents *Blarina brevicauda* (group Blbr) and the histogram on bottom represents *B. carolinensis* (group Blca). The arrows represent the approximate group centroid placement.

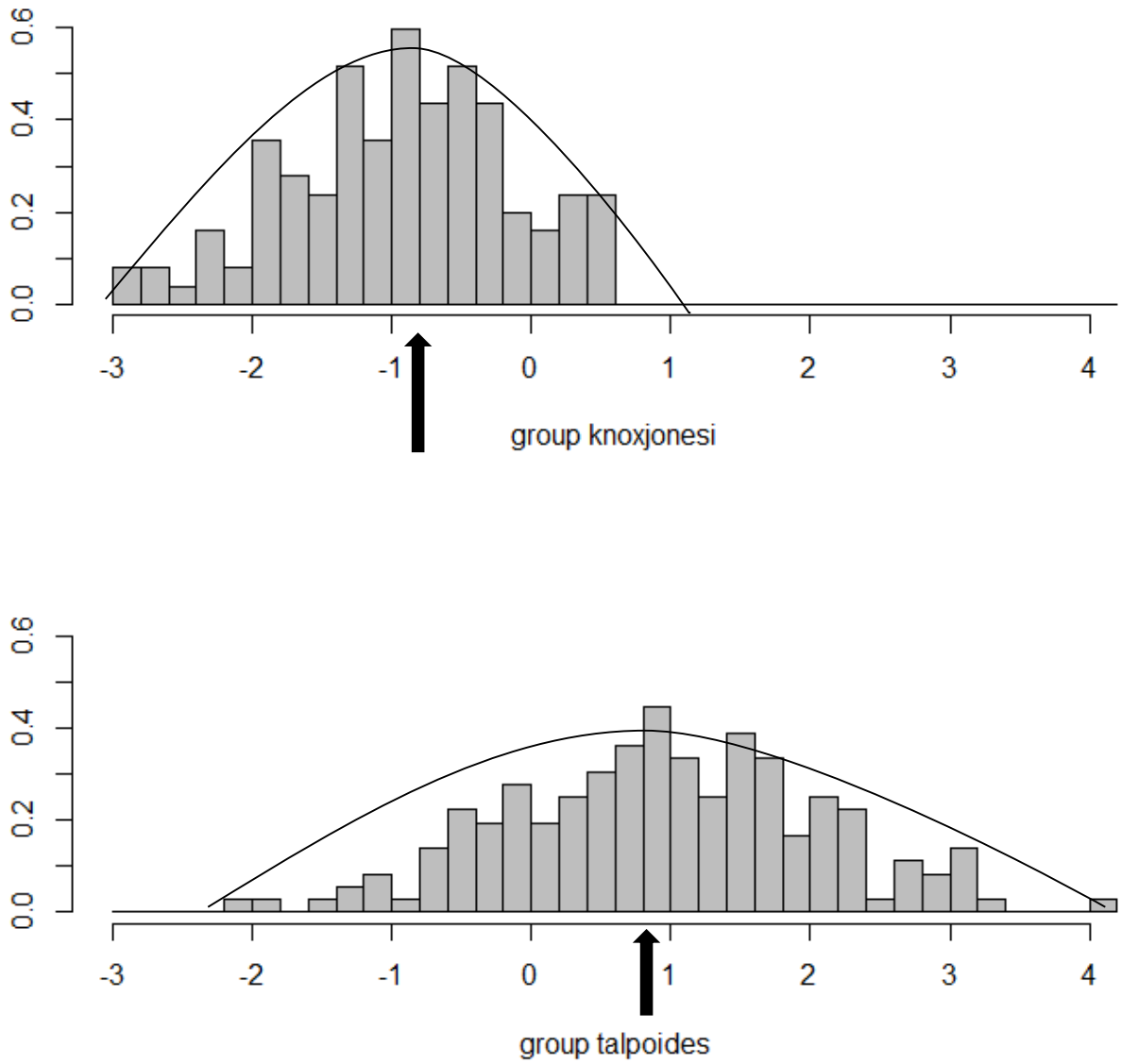


Figure 2.13: Linear discriminant function analysis histogram comparing the first discriminant function versus the frequency for the mandible of the northern short-tailed shrew subspecies. The histogram on top represents *Blarina brevicauda knoxjonesi* (group knoxjonesi) and the histogram on bottom represents *B. b. talpoides* (group talpoides). The arrows represent the approximate group centroid placement.

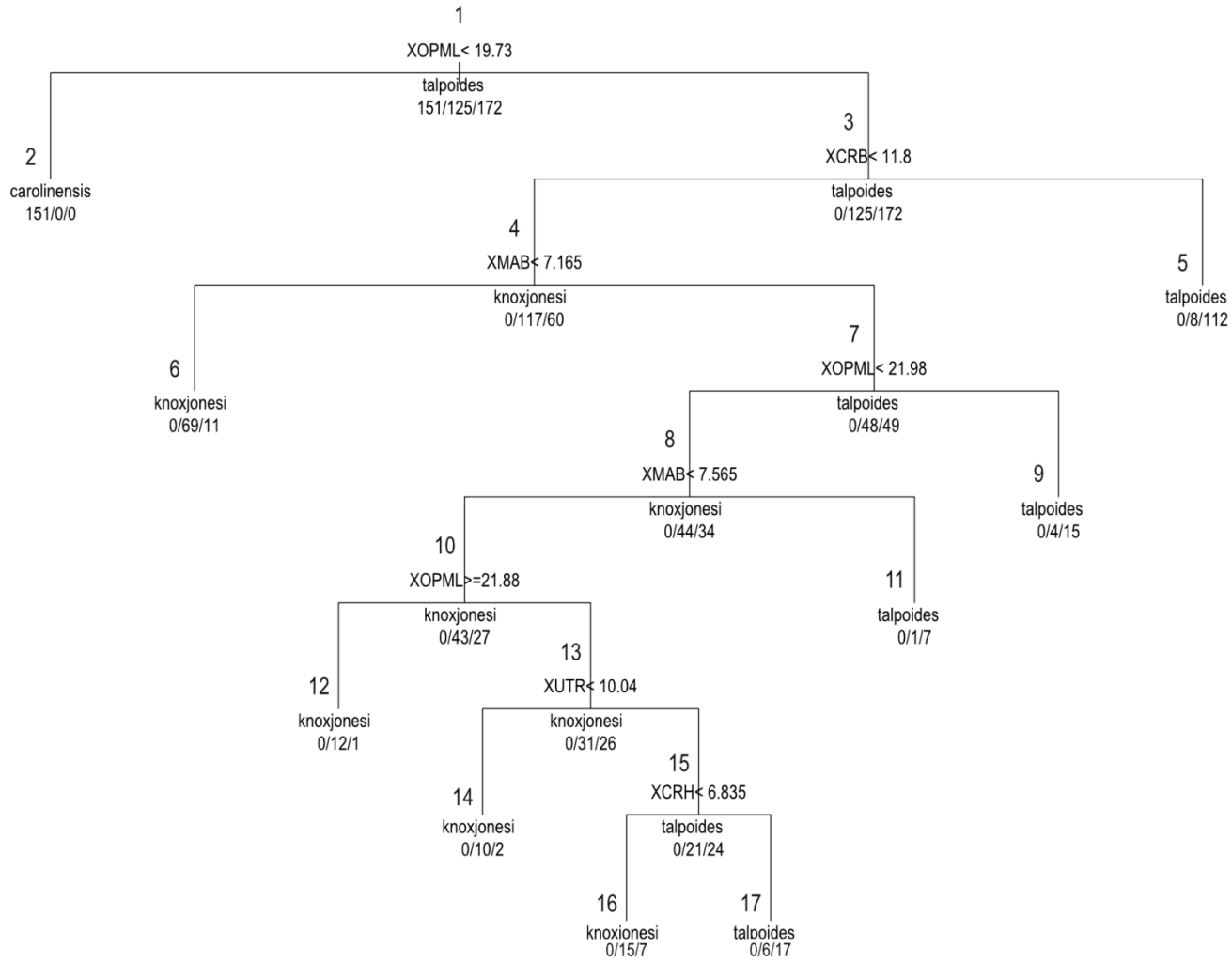


Figure 2.14: Classification tree with Gini splitting index for *Blarina* skull data. Node numbers 1 – 17 correspond to details in Table 2.10.

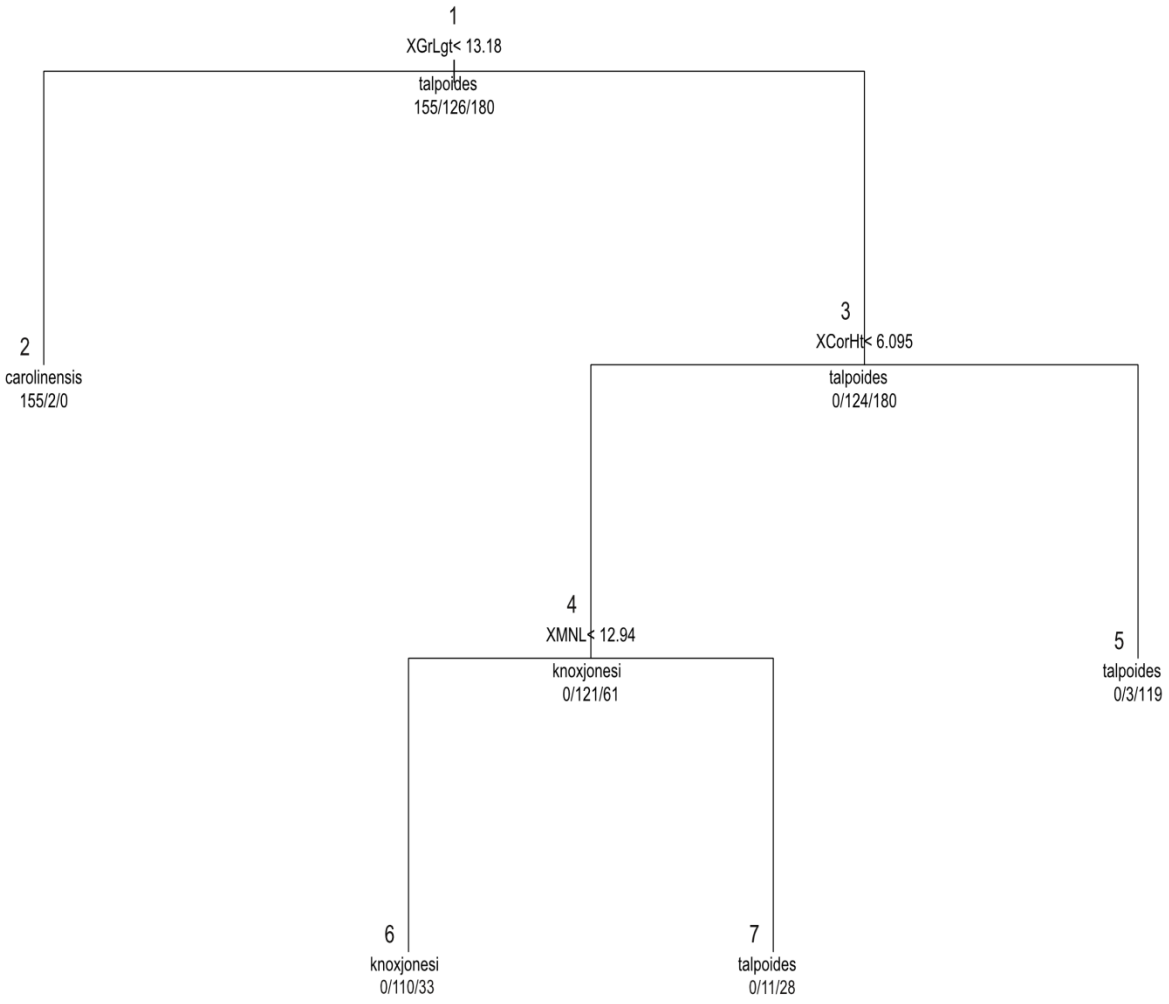


Figure 2.15: Classification tree with Gini splitting index for *Blarina* mandible data. Node numbers 1 – 7 correspond to details in Table 2.11.

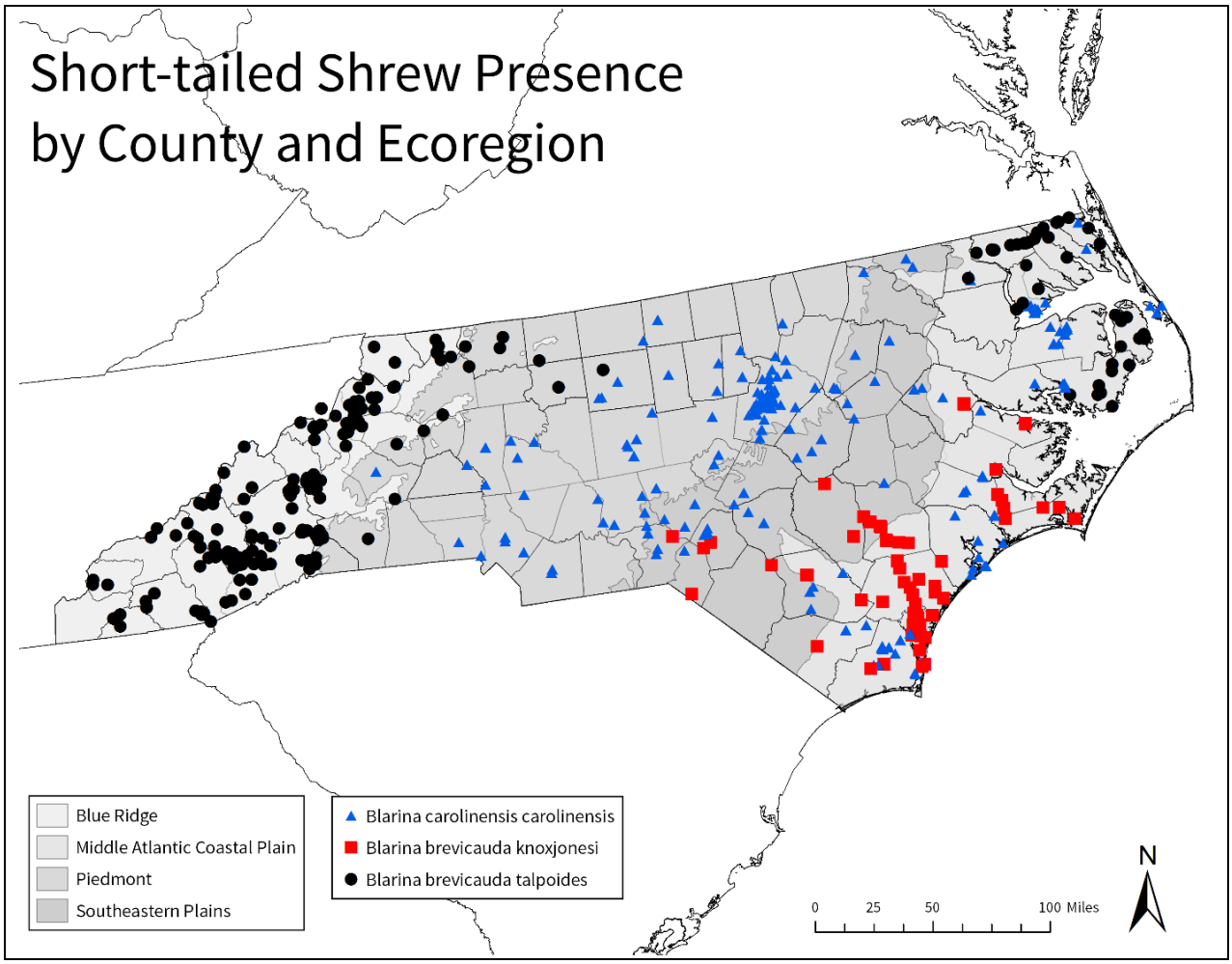


Figure 2.16: Distribution of short-tailed shrew voucher specimens in North Carolina counties and ecoregions. These voucher specimens represent holdings of the genus *Blarina* held at the North Carolina Museum of Natural Sciences and the University of North Carolina Wilmington. Points represent the center of a confidence region from the locations indicated with specimen data. The Pamlico River to the north and the Cape Fear River to the south are indicated as potential barriers for *Blarina brevicauda knoxjonesi* in eastern North Carolina. Map by M. B. Norton.

Chapter 3

Genetic Analyses of Short-tailed Shrews (Genus *Blarina*) in North Carolina

ABSTRACT

The genetic species concept is used in systematics for mammals, and the cytochrome-*b* gene has been used extensively for understanding phylogenetic relationships. I examined the sequence variation in 1080 bp of the mitochondrial cytochrome-b gene from 100 samples of the short-tailed shrew in the genus *Blarina* and select outgroups. Genetic distances show a separation between *Blarina brevicauda, B. carolinensis, B. hylophaga, Cryptotis parva, Sorex cinereus, S. longirostris* and *Condylura cristata* with a sequence divergence between 2.6% and 19.3%. Separation between all subspecies ranged from 1.2% to 3.6%. *B. brevicauda* and *B. carolinensis* occur in North Carolina with a mean genetic distance difference of 0.6% and 0.4% respectively. Phylogenetic analysis under Bayesian and maximum likelihood methods produced trees with identical topology where *Blarina brevicauda*, *B. carolinensis* and *B. hylophaga* represent three clear clades. All samples of *B. brevicauda* from North Carolina group with the *B. b. talpoides* subspecies and included a single sample indicated as *B. b. knoxjonesi*. Samples of *B. brevicauda* from Alabama clustered into a separate clade and may be the first confirmation for *B. b. cumberlandensis* in the state.

Keywords: *Blarina*; cytochrome-b; genetic species concept; mitochondrial DNA; molecular systematics; phylogenetics; short-tailed shrews

INTRODUCTION

The genetic species concept (Dobzansky, 1950; Mayr, 1969; Simpson, 1943) identifies a species as those individuals with a common gene pool. Using this concept, genetic material from different individual animals like RNA, DNA and other coding regions of an organism can be measured for its similarity. Carson (1957) defined a genetic species as "fields for gene recombination" where genetic material is able to pass throughout individuals unobstructed. Masters and Spencer (1989) argued that the genetic species concept has a number of flaws. They compared past literature to convey how the genetic species concept applies only to sexually reproducing individuals and to describe the inconsistencies in describing and delineating barriers to gene flow that isolate species. In addition to the difficulty in determining if barriers to gene flow exist, one of the most debatable components of the genetic species concept is how much genetic variation is actually needed to split species and subspecies.

DNA can be altered and changed in many ways, resulting in different mutations. A common mutation is a point mutation that occurs at a specific point along a DNA strand. Phylogenetic substitution models, or models of DNA sequence evolution, help to explain the observed proportions of bases, variations in point mutations, and the different rates of change present in transition and transversion substitutions (Table 3.1). One can select the best-fit phylogenetic model based on a particular set of DNA sequence data using various approaches (Posada and Crandall, 2001; Sullivan and Joyce, 2005) and computer software allows fast comparisons of models using different selection criteria (Lanfear et al., 2012; Posada, 2008; Posada and Crandall, 1998). These models use information about the sequence data to create phylogenetic trees that group organisms by the similarities of their sequences. Early phylogenetic tree

construction used genetic distances between sequences where the tree topology (i.e. the phylogenetic tree branching configuration) depended upon the order by which the data was entered into the computer software but did not use a substitution model (Farris, 1970; Fitch, 1971; Murtagh, 1984; Saitou and Nei, 1987; Sokal and Michener, 1958). More recently, maximum likelihood methods (Guindon and Gascuel, 2003; Tamura et al., 2011; Yang, 2007; Zwickl, 2006; Zwickl, 2008) and Bayesian analysis methods (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Ronquist et al., 2011; Ronquist et al., 2012) incorporate a specific substitution model for an improved understanding about how sample organisms are related base on the DNA sequences.

Designations of mammal species were initially based solely on phenotypic variation. This has traditionally been done by looking at the morphology of skull characteristics. Species that are closely related and share many similarities in morphology are considered cryptic species. In cryptic species, traditional morphological analyses may not recognize the separation of species or subspecies, and the addition of genetic analyses can reveal new species or species boundaries (Baker and Bradley, 2006; Bradley and Baker, 2001). Baker and Bradley (2006) defined a genetic species for mammals as a "group of genetically compatible, interbreeding natural populations that is genetically isolated from other such groups". With this definition, they considered speciation to occur if a divergence in genetic material caused genetic isolation of the two gene pools of the respective lineages, which was due to an accumulation of genetic change between two lineages.

Molecular markers (e.g. allozymes, ribosomal DNA and mitochondrial DNA) have a long history of use in species differentiation but mitochondrial DNA is especially useful for within species differentiation due to its maternal inheritance and high rate of evolutionary substitutions

(Avise, 2004; Kocher et al., 1989). Within sequences of mitochondrial DNA, the rate of change differs between the respective bases that code for amino acids, where position-one is slightly variable, position-two is the most stable, and position-three is the most variable (i.e. the wobble hypothesis; Crick, 1966). Sequences of the mitochondrial cytochrome-b gene can be used to split species and subspecies by examining their genetic distances, which can infer isolation of interbreeding natural populations. Bradley and Baker (2001) suggested that the average genetic distance calculated using the Kimura 2-parameter model of nucleotide substitution (Kimura, 1980) between sampled mammal species was 2.49, and that the average genetic distance between sampled subspecies was 1.04. They also argued that small sample sizes could yield genetic distances lower or higher than the average genetic distances listed above, which warrant additional studies to confirm if genetic distance actually reflect biological differentiation. The values can be considered as a percent difference between the sequences by multiplying the genetic distance by 100 where in a similar assessment, Larsen et al. (2012) indicated that values greater than 0.02 (2.00) indicate different species. Analyses of mitochondrial cytochrome-b genes have clarified relationships among rodents (Bellinger et al., 2005; Blois and Arbogast 2006; Bradley et al., 2007; Harris et al., 2000; Sudman et al., 2006), among carnivores (Helgen et al., 2013; Masuda and Yoshida, 1994) and among long-tailed shrews that included short-tailed shrews (genus *Blarina*) as an outgroup (Fumagalli et al., 1999; Maldonado et al., 2001; Ohdachi et al., 2006; Willows-Munro and Matthee, 2011).

The short-tailed shrews in the genus *Blarina* were initially designated using morphological analyses and historically were split into two species (Hall, 1981). This genus of shrews has overlap in many morphological characters but genetic analyses continue to provided clarity within this group. Early genetic differentiation of the species of *Blarina* was done by

karyotyping. The karyotype describes the appearance and number of chromosomes found in the nucleus, and often compares the diploid number (i.e. the number of chromosomes in a somatic cell), the fundamental number (i.e. the number of chromosomal arms in a somatic cell) or the complete set of chromosomes (Stebbins, 1950; White, 1973). The northern short-tailed shrew (*B. brevicauda*) has a diploid number (2N) = 48, 49 or 50, and a fundamental number (FN) = 48 (Genoways et al., 1977; George, et al., 1986; George et al., 1982). The southern short-tailed shrew (*B. carolinensis*) has a lower karyotype with 2N = 35-41, 46 and FN = 41-45 (Beck et al., 1991; Elrod et al., 1996; George et al., 1982; Qumsiyeh et al., 1997). Elliot's short-tailed shrew (*B. hylophaga*) is characterized with a higher karyotype of 2N = 52 and FN = 60, 61 or 62 (Genoways et al., 1977; George et al., 1982; Thompson et al., 2011). The Everglades short-tailed shrew (*B. peninsulae* – formerly *B. c. peninsulae*) has a 2N = 50-52 and FN = 52, and was only recently elevated to the species level (Genoways and Choate, 1998; George et al., 1982; Wilson and Reeder, 2005). North Carolina specimens of *B. brevicauda* (George et al., 1982; Webster 1996) and *B. carolinensis* (Qumsiyeh et al., 1997) were karyotyped with 2N = 49 or 50 and FN = 48, and 2N = 46 and FN = 44 respectively for the two species. The differences in the diploid and fundamental numbers separate the species of *Blarina*, but do not clarify the subspecies.

Recent studies examining genetic variation beyond chromosomes has clarified within species variation of *Blarina*. The cytochrome-b gene has been studied exclusively in *Blarina* to show that species select against interbreeding and have very narrow zones of overlap (Bennedict, 1999a; Bennedict, 1999b), and to identify species correctly in new areas or in old areas where measurement data yielded wrong identifications (Pfau and Braun, 2013; Pfau et al., 2011; Reilly et al., 2005). In addition, the cytochrome-b gene has shown well-differentiated genetic groups

within species of *Blarina* suggesting barriers to gene flow and giving evidence for subspecies designations (Brant and Ortí, 2002; Brant and Ortí, 2003; Reilly et al., 2005).

Webster (1996) designated a new subspecies (*Blarina brevicauda knoxjonesi*) in the southeast part of North Carolina, where other northern short-tailed shrews in the state are considered to be *B. b. talpoides* (Webster, 2011). This designation was done with karyotype species confirmation along with morphological and ecological differences to separate subspecies (Webster, 1996). *B. carolinensis carolinensis* has been the only subspecies designation for the southern short-tailed shrew in North Carolina since being designated (Lee et al., 1982; McCay, 2001; Webster et al., 1985). No genetic research has been done on *Blarina* in North Carolina beyond karyotyping, and DNA sequences may provide clarity if one or multiple subspecies exist within *B. brevicauda* or *B. carolinensis*.

The goal of my research was to examine the current taxonomy of the short-tailed shrews of the genus *Blarina* in North Carolina (McCay, 2001; Webster, 2011). To do this, I used the genetic species concept and examined the genetic variation in *Blarina* specimens from the North Carolina Museum of Natural Sciences (NCSM – formerly the North Carolina State Museum of Natural Sciences). I examined sequence diversity in the cytochrome-b gene in tissue samples from *Blarina brevicauda* and *B. carolinensis* to determine if the current classification within *B. brevicauda* in the state is valid or that further taxonomic revision is needed. I hypothesize that *B. brevicauda* and *B. carolinensis* should have sequence divergences consistent with their species designations, should show genetic distances greater than 2.49 and should separate into distinct clades on a phylogram. I also hypothesize that if the systematic status of the two subspecies within *B. brevicauda* is correct in North Carolina, sequence variation should show a separation in samples from the respective geographic regions of the subspecies. Subspecies should show

genetic distances greater than 1.04, should separate into distinct branches within their respective species clade, and should be separated geographically from other subspecies.

MATERIALS AND METHODS

Taxa and tissue selection

I collected tissue samples from 118 specimens of *Blarina brevicauda* and *B. carolinensis* plus specimens of outgroup species from the collection at the North Carolina Museum of Natural Sciences (NCSM). 50 specimens of *B. brevicauda* and 50 specimens of *B. carolinensis* came from North Carolina (Figure 4 and 5), with an additional 10 specimens of *B. brevicauda* from Alabama, and 1 *B. brevicauda* specimen each from both New York and Virginia. I selected 2 samples for each outgroup species from North Carolina represented by the least shrew (*Cryptotis parva*), southeastern shrew (*Sorex longirostris*), masked shrew (*Sorex cinereus*) and star-nosed mole (*Condylura cristata*).

Liver and muscle tissues from these voucher specimens were extracted, diced, placed in a 2.0 mL Nalgene cryo-tube with a 95% ethyl alcohol (EtOH) buffer solution and stored in a refrigerator at 4° Celsius (C) for a minimum of 24 hours. After refrigeration, the ethanol extracted excess water from the tissue samples and was poured out for the addition of new 95% EtOH in each cryo-tube for the long-term storage in a -80°C ultra-cold freezer (Camacho-Sanchez, 2013; Corriveau et al., 2009; Shokralla et al., 2010). I took 2-3 rice sized pieces of tissue from each tissue voucher in cold storage as a permanent loan to myself for the study of genetic variation in short-tailed shrews in North Carolina from the NCSM Mammal Unit. See Appendix IV for information about museum specimens used in this study.

DNA isolation, amplification, and sequencing

I measured the available DNA concentration in each sample using a NanoDrop ND-1000 V3.7.0 Spectrophotometer (Thermo Fisher Scientific - NanoDrop products, Wilmington, Delaware). For any sample with a concentration greater than 600 nannograms per microliter (ηg/µL), I diluted the respective sample with a 1X TE buffer.

Polymerase chain reaction (PCR; Mullis et al., 1986; Saiki et al., 1986; Saiki et al., 1988) was performed on each sample to amplify the entire cytochrome-b (*cyt b*) gene. Initially, I created an amplification cocktail containing 12.5 µL of Thermo Scientific Dream Taq Master Mix, 1.0 µL of each primer and 9.5 µL of PCR dH_2O for all samples. The primers used in the reactions were forward LGL 765 (5'-GAA AAA CCA YCG TTG TWA TTC AAC T-3') and reverse LGL 766 (5'GTT TAA TTA GAA TYT YAG CTT TGG G-3') (Bickham et al., 1995; Bickham et al., 2004; Pfau and Braun 2013). I added 1.0 µL of each sample for a 25 µL total PCR reaction and ran the following sequencing parameters: 1 cycle of 94°C (5 minutes); 35 cycles of 94°C (45 seconds denaturing), 50°C (30 seconds annealing), and 72°C (1 minute); followed by 1 cycle of 72°C (7 minutes); and held at 4°C (until removed).

After PCR, I made a 1% agarose gel to view if the amplification process was successful. SYBR Safe DNA stain (Life Technologies, Carlsbad, California) was added to the melted agarose and TAE buffer to aid the visualization of amplified PCR reactions under ultraviolet (UV) light. On each gel, I added 5 µL of a 100 base pair DNA ladder (Ausubel et al., 1992) to the left chamber for comparison. I placed 2 µL of purple loading dye on a piece of Parafilm with a multi-channel pipette for each sample. To this, I added 5 µL of each sample, changed the pipet volume to 9 µL and mixed the sample before placing the sample in the respective agarose well.

This was repeated until all samples were loaded on the agarose gel. After the DNA moves from the negative to positive direction in the electrophoresis set-up, each positive sample will show a bright band when the UV transilluminator is turned on in the BioDoc-It Imaging System (UVP, LLC, Upland, California). I printed each respective gel image on the attached thermal printer to record all positive reactions (Figure 3.2).

After one round of sequencing, some samples did not work. All DNA samples that did not yield a positive gel band were re-extracted from the sample tissue, and run through the PCR with different combinations of polymerases and reaction cocktails. I ran the following varying cocktails: the original Dream Taq Master Mix, Dream Taq Master Mix under variable annealing temperatures, 5 PRIME Perfect Taq under variable annealing temperatures, Phusion polymerase and 5 PRIME Hot Master Mix.

The resulting double-stranded amplified product (i.e. the samples with a bright band under UV) was purified with ExoSAP-IT (USB Products, Cleveland, Ohio) to eliminate unincorporated primers and dNTPs. I obtained sequences for both direction primers (LGL 765 and LGL 766) by adding a Big Dye Terminator and buffer (Life Technologies, Carlsbad, California) to 2.0 µL of each amplified and cleaned product. Sequencing parameters used were: 1 cycle of 96°C (1 minute); 25 cycles of 96°C (10 seconds), 50°C (5 seconds), and 68°C (4 minutes); and held at 4°C (until removed). An ethanol-salt-acid solution precipitated the reactions no more than 24 hours after sequencing, with the resulting reaction held at 4°C until sequenced. Within four hours of sequencing, 10µL of Hi-Di was added into each well to denature all samples. All samples were run on an Applied Biosystems/Hitachi 3130xl Genetic Analyzer (Applied Biosystems, Foster City, California) resulting in a forward and reverse single-stranded sequence. Details of all DNA protocols are available upon request.

Sequence alignment

I aligned and proofed sequences via the chromatogram visualization in Geneious version 7 (http://www.geneious.com; Kearse et al., 2012). The forward and reverse single stranded sequences formed a set of overlapping DNA segments that together represent a consensus region of DNA called an assembly contig, which created the preliminary alignments. I visually inspected and adjusted the base calls of the overlap regions for each sample capable of creating a contig (Figure 3.3). Once all the samples were individually aligned, each contig became a consensus sequence in Geneious, and combined into one file.

In addition to the consensus sequences obtained from tissue samples at the NCSM, I downloaded and added 20 cytochrome-b gene DNA sequences from GenBank (Benson, et al., 2008; Sayers et al., 2009). Ten sequences from *Blarina brevicauda* with a length of 1131 base pairs (bp) from GenBank (accession number: AF533609, AF533619, AF533641, AF534115, AF534117, AF534116, AF534123, AF533632, AF534121 and AF533642) were obtained from Brant and Ortí (2003), five sequences from *B. carolinensis* with a length of 1140 bp (GenBank AF395449, AF395456, AF395458, AF395460 and AF395454) were obtained from Brant and Ortí (2002), and five sequences from *B. hylophaga* with a length of 1140 bp (GenBank AY546659, AY546662, AY546669, AY546681 and AY546677) were obtained from Reilly et al. (2005). See Appendix IV for information about GenBank specimens used in this study.

To these sequences, I manually added my forward and reverse primer sequences along with transfer RNA (tRNA: TRNT – tRNA threonine, TRNP – tRNA proline and TRNE – tRNA glutamate) sequences found in GenBank. I exported all sequences out of Genious as a FASTA (Lipman and Pearson, 1985) file, and imported this file of consensus sequences into multiple

sequences alignment software (MAFFT version 7; Katoh and Standley, 2013). When this new multiple alignment was complete, I imported this file back into Geneious to continue revising the multiple sequences. Once aligned, I trimmed off the primer sequences and adjacent tRNA sequences to ensure that I was only analyzing sequences of cytochrome-b.

MAFFT aligned and organized the samples by sequence similarity. I looked at each gap and fixed discrepancies within the original contig file. When all gaps were adjusted, the translation function within Geneious changes the individual base pairs into codons. I selected the vertebrate mitochondrial genetic code and kept the translation relative to the alignment. Different frame views yield more or less stop codons in the translation, with the frame 1 option giving the best translation. With the codons present, I made more adjustments to each of the sequences in the contig files and repeated the Geneious-MAFFT process until no additional changes were needed. I trimmed each end of the sequences in the Geneious multiple alignment sequence view to equal length comparing the codons names from the downloaded GenBank sequences.

Genetic distances

Once all the sequences were aligned and corrected, I exported a NEXUS (Maddison et al., 1997) file out of Geneious and imported this file into the Phylogenetic Analysis Using Parsimony software (PAUP* 4.0a147; Swofford, 2002) to calculate the uncorrected genetic distances on a small sample to confirm the codon position. With the codon position reading at the first position and the sequences trimmed, I calculated the uncorrected genetic distances on the entire cytochrome-b sequence data set.

Model selection

In PAUP, I separated the sequences by codon position and exported a separate NEXUS file for each position. Each codon position (i.e. CytB_1_only, CytB_2_only and CytB_3_only) sequence was imported into the online portal Cyberinfrastructure for Phylogenetic Research (CIPRES Science Gateway, Miller et al., 2010). I used the NEXUS files to run jModelTest2 (Darriba et al., 2012; Guindon and Gascuel, 2003) to assess the best fit model independently on each codon position by comparing common phylogenetic models in a likelihood analysis (Table 3.1). JModelTest2 compared 22 unique models, where each model is compared with its base properties, with invariable sites (+I), with rate variation among sites (+G), or with both invariable sites and rate variation among sites (+I+G). This ultimately compared 88 candidate models, where the program ran the models from most complex (+I+G) to simple (+G and +I independently), and then to the base model where each model received a calculated a likelihood score (-lnL). Models were eliminated based on hypothesis testing (and accepted or rejected) of my sequence data separated by codon position for base frequencies, substitution types, proportion of invariable sites and substitution rate variation (Posada and Crandall, 2001). The final substitution models of sequence evolution were determined using the corrected Akaike Information Criteria (AICc – Hurvich and Tsai, 1989; Sugiura, 1978) for small sample size. The small sample size is defined as $n / k_i \leq 40$, where n is the number of sites and k_i is the number of model parameters (Sullivan and Joyce, 2005).

Bayesian analysis

I ran a mixed-model Bayesian analysis in MrBayes 3.2 (Ronquist et al., 2011 and 2012) on the CIPRES portal with an imported NEXUS file. I ran the sequences partitioned by first-codon,

second-codon and third-codon position. The substitution models of sequence evolution selected by jModelTest2 with AICc were K80 + G (Kimura, 1980) for the first-codon position, F81 (Felsenstein, 1981) for the second-codon position, and TIM2 + G (Posada, 2003) for the third-codon position. The Bayesian analysis was run with different substitution rates (nst = 2, nst = 1 and nst = 6; Table 3.1) for the first, second and third-codon position respectively for 10 million generations. Four independent runs were done, each with one cold chain and 3 heated chains. Every 1000 generations, trees were sampled with a Marcov Chain Monte Carlo (MCMC – Huelsenbeck and Ronquist, 2001) method. The first 25% were discarded as a burn-in (i.e. a fraction of samples discarded before lack of improvement in likelihood score), leaving the remaining trees sampled for the construction of a 50% majority rule consensus tree. The four runs were combined to construct the final consensus tree, and the likelihood scores and posterior probabilities were determined for the tree summary by using the *sumt* command. I analyzed the results in the Java program Tracer (Rambaut et al, 2014) to test for normal or skewed data distribution, and to test for convergence of data.

Maximum likelihood analysis

I performed a maximum likelihood (ML) analysis on the cytochrome-b sequence data using the Genetic Algorithm for Rapid Likelihood Inference software (GARLI; Zwickl, 2006) and run on a computer desktop application with an imported Nexus file. I used the model of evolution for each codon position to set the prior rate matrix code, the state frequencies, rate-heterogeneity distribution model, and the number of rate categories as detailed in the GARLI manual (Zwickl, 2008). The first-codon position with a K80 + G model of evolution gave the parameters: prior rate matrix code = (0 1 0 0 1 0), state frequencies = estimate, rate-heterogeneity distribution

model = gamma, and the number of rate categories = 4; the second-codon position with a F81 model of evolution gave the parameters: prior rate matrix code = (0 0 0 0 0 0), state frequencies = estimate, rate-heterogeneity distribution model = none, and the number of rate categories = 1; and the third-codon position with a GTR + I + G model of evolution gave the parameters: prior rate matrix code = (0 1 0 2 3 2), state frequencies = estimate, rate-heterogeneity distribution model = gamma, and the number of rate categories = 4. I ran 1000 bootstrap replicates for 100,000 generations, with trees stored every 100 generations and a significant topology change criterion set at 0.01 (this is the first condition for termination and is the likelihood increase value needed for a new topology to be considered significant) and a score threshold set at 0.05 (this is the second condition for termination that stops when the total improvement in the likelihood score is < 0.05 when compared to the last solution) as default values (Zwickl, 2008). Bootstrap support values were calculated in PAUP*. I imported the bootstrap tree file from GARLI into PAUP* and used the majority rule command to give all bootstrap values over 50% and show the consensus tree.

Phylogenetic topology construction

I viewed all tree files in FigTree v1.4.2 (Rambaut, 2014). With each respective tree viewed in FigTree, I altered the tree with increasing node order, rerooted with outgroups, and rotated branches for consistency of location of major clades. The final tree topology for the Bayesian analysis and ML bootstrap were imported into Scribus v1.4.5 (Schmid et al., 2015) where I redrew each trees for clarity and additions.

RESULTS

Out of the 118 original samples, 80 individuals yielded a genetic sequence for the cytochrome-b gene. The final sample set resulted in 36 *Blarina brevicauda* and 39 *B. carolinensis* for the ingroup, and the outgroup was 2 *Cryptotis parva*, 1 *Sorex cinereus*, 1 *Sorex longirostris* and 1 *Condylura cristata*. After alignment with the 20 GenBank cytochrome-b sequences (10 *B. brevicauda*, 5 *B. carolinensis* and 5 *B. hylophaga*), I trimmed the ends for final sequence length was 1080 base pairs and 100 samples.

Genetic distances

Uncorrected pairwise genetic distances (Table 3.2) indicate intraspecific or interspecific cytochrome-b sequence divergence on the samples in the genus *Blarina* and select outgroups. The genetic distances for the sample sequences ranged from minimum value of 0 to a maximum of 19.7%. The mean value between all outgroups and all *Blarina* was 16.7%. The greatest difference from *Blarina* occurred with *Condylura cristata* (19.3%) and the smallest difference occurred with *Cryptotis parva* (15.2%). The lowest mean distance within the outgroup was the comparison of *Sorex cinereus* and *S. longirostris* with a 2.7% difference.

The mean genetic distance between all samples of *Blarina brevicauda* (Blbr), *Blarina carolinensis* (Blca) and *Blarina hylophaga* (Blhy) was 7.6%, where Blbr vs Blca = 7.6%, Blbr vs Blhy = 8.7% and Blca vs Blhy = 6.5%. Within Blbr, the distance varied on average by 0.9% with the largest differences among different localities. There was little difference of Blbr within North Carolina (NC - 0.6%), but comparisons with other states ranged from 0.4% (West Virginia Blbr vs NC Blbr) to 2.4% (Iowa Blbr vs NC Blbr). NCSM 16801 represented a sample from the

locality for *B. b. knoxjonesi* (Webster, 1996; Webster et al., 2011), but it only had one comparison that had a distance above 1.0% (NCSM 16801 vs NCSM 17616 – NC, Haywood Co, Clyde at Harmon Den WMA). Blca also showed little difference (0.4%) in North Carolina, but comparisons with Arkansas, Illinois and Florida were all greater than 3.1% difference in sequences.

Model selection

I selected the top model based on the corrected Akaike information criterion (AICc). I used the small sample size definition $n / k_i \leq 40$ (Sullivan and Joyce, 2005), where n is the number of sites (n = 360 bases or 1080/3) and k_i is the number of model parameters (k_i = K = 200, 201 and 205 for codon position 1, 2 and 3 respectively). The top models selected with AICc in jModelTest2 were K80+G (Kimura, 1980) for the first-codon, F81 (Felsentein, 1981) for the second-codon, and TIM2+G (Posada, 2003) for the third-codon position. K80 + G is the Kimura 2-parameter model that has equal base frequencies, one transition and one transversion rate (nst = 2), and a gamma distributed rate variation among sites. F81, also known as the TN84 model, has variable base frequencies and all substitutions are equally likely (nst = 1). Lastly, TIM2 + G, or transitional model, has variable base frequencies, variable transition rates and two transversion rates (nst = 6).

Bayesian analysis

The Bayesian analysis produced a bootstrap consensus tree from four independent runs (Figure 3.4) with one cold and three hot chains. The –ln likelihood score of the best state for the cold chain of run 1 was 5105.20, for run 2 was 5133.20, for run 3 was 5162.79, and for run 4 was

5288.23. The average standard deviation of split frequencies, which was computed across the four runs was 0.004751. The program Tracer showed no convergence and did show stabilization in the trace file. This indicated that the sequence data was sampled randomly in each of the four runs across the 10 million generations. Also, the mean estimates parameters showed that all sequence data was normally distributed across the four runs.

All shrews of the genus *Blarina* split from the outgroups at node 1 with a posterior probability of 100%. Figure 3.4 showed *B. hylophaga* split from *B. brevicauda* and *B. carolinensis* at node 3 (100%), while *B. brevicauda* split from *B. carolinensis* at node 5 (95%). Within the *B. brevicauda* clade (Figure 3.5) at node 9, the GenBank sequence sample form Wisconsin split in a basal position at node 9. At node 10, the GenBank samples from Iowa and Kentucky split with samples from Alabama. The remainder of the *B. brevicauda* samples, including all North Carolina samples split at node 11. The North Carolina samples grouped with the GenBank sequence samples from New York, Pennsylvania, Tennessee, Virginia and West Virginia. NCSM 16801, which represented a sample from the locality for *B. b. knoxjonesi*, grouped within the group at node 11, and within a group including New York, Pennsylvania, Virginia and West Virginia as well as NCSM 17595 (NC, Watauga Co., Banner Elk at Bear Paw SNA). The *B. carolinensis* clade (Figure 3.6) at node 6, separated the GenBank sequence samples from Arkansas, Florida and Illinois in separate groups from the North Carolina specimens and the one GenBank sample from Georgia.

Maximum likelihood analysis

The ML consensus tree showed a nearly identical topology to the Bayesian analysis (Figure 3.4) with the exception of an extra branch for the outgroup *Cryptotis parva*. All branches within

each species of *Blarina* were identical. The ML consensus tree had a –ln likelihood score of 4724.0013 after twenty partitioned search replicates., and a final –ln likelihood score after 1000 bootstrap replications of 4477.1039.

DISCUSSION

The analyses presented in this study identified two species of shrews in the genus *Blarina*, *B. brevicauda* and *B. carolinensis*, in North Carolina. The two species occur in North Carolina with little sequence variation. Within *B. brevicauda*, all samples group with the *B. b. talpoides* clade (Webster et al, 2011). The one sample of *B. b. knoxjonesi* (NCSM 16801) has been described as a distinct subspecies, but the cytochrome-b sequence from this sample did not separate it into a separate branch, suggesting that this subspecies should possibly be grouped with *B. b. talpoides* (Webster, 1996; Webster, et al., 2011). All samples of *B. carolinensis* in North Carolina showed less variation than *B. brevicauda* and represent the *B. c. carolinensis* subspecies (McCay, 2001; Merriam, 1895a). The use of *Cryptotis parva* as an outgroup species revealed it is more closely related to *Blarina* than to the other outgroup species. In addition, these analyses identify three species of shrews in the genus *Blarina*: *B. brevicauda*, *B. carolinensis* and *B. hylophaga*. This result is consistent with other studies that used the mitochondrial DNA cytochrome-b gene (Brant and Orti, 2002; Pfau and Braun, 2013; Pfau et al., 2011; Reilly et al., 2005).

Larsen et al. (2012) suggested that closely related North American related mammal species have a low level of interspecific sequence divergence (i.e. >2.0% or 0.02 is found between species). The sample sequences used to determine my distances held true to their findings (Table 3.2). The lowest interspecific sequence divergence was between *Sorex cinereus* and *S.*

longirostris (2.7%), but this only compared 2 samples and may possibly be higher with a larger sample size. This was followed by *B. carolinensis* and *B. hylophaga* with a mean distance of 6.5%. Comparisons between *Blarina* and outgroup species showed the closest relative is *Cryptotis parva*, which has been shown in other studies (Brant and Ortí, 2002; Choate, 1970; Driskell and Feldhamer, 2003, Whitaker, 1974). *C. parva* is followed by *Sorex longirostris* and *S. cinereus* in decreasing relatedness as determined by the mean genetic distances (15.2%, 16.9% and 17.1% respectively) to *Blarina*. *Condylura cristata* had the greatest mean genetic distance (19.3%), which should occur as this is the only species in the Talpidae Family and not the Soricidae Family (Petersen and Yates, 1980).

Samples with similar sequences have small genetic distances and are more related. The genetic species concept looks at the similarity of DNA to determine species or populations. Bradley and Baker (2001) examined variation in the mitochondrial cytochrome-b gene in mammals and determined that sequence divergences can split species, but values between 2 and 11% needed more study than the sequences alone in some species. They also noted that an average sequence divergence greater than 1.04 may indicate subspecies status. Genetic isolation is also a stipulation for interbreeding natural populations of mammals in which zones of contact can cause genetic hybrids (Baker and Bradley, 2006). Hybridization depends on the ecological competitiveness of hybrid genotypes, where the recombination of genetic material among lineages can give rise to novel phenotypes (Stelkens and Seehausen, 2009). In this study, both *Blarina* and *Sorex* had genetic distance values between 2 and 11%, which may account for the vast number of species and subspecies designations over time (Hall, 1981; Merriam, 1895a; Merriam 1895b; Miller, 1895; Wilson and Reeder, 2005).

Bayesian and ML bootstrap analyses for constructing tree topologies use methods that rely on an explicit model of evolution. Separating the sequences by codon position allowed the best analysis of this set of sequence data by using a separate model of evolution for each codon position. The variation within each codon position, especially the third position (Crick, 1966), would have been missed if the sequences were analyzed in an un-partitioned manner.

The Bayesian posterior probabilities values were higher than the maximum likelihood bootstrap support values for the samples used in the phylogenetic analysis (Figure 3.4 – 3.6). Bayesian posterior probabilities are derived from a Markov Chain Monte Carlo (MCMC) method that generates the set of trees with the highest posterior probabilities and are less conservative. (Huelsenbeck and Rannala, 2004; Rannala and Yang, 1996; Yang and Rannala, 1997). Alfaro et al. (2003) suggested that bootstrap support values are more conservative than Bayesian posterior probabilities and tend to be less prone to error. The bootstrap support values give the likelihood of the best tree in support of the given sequence data (Felsenstein, 1981; Kishino and Hasegawa, 1989). Both Bayesian posterior probabilities and bootstrap support values give a notion of confidence, but the posterior probabilities derived from the MCMC sampling method tend to sample very similar tree topologies based on their higher clade support than the bootstrap support values (Garcia-Sandoval, 2014). Despite the different approaches in finding the "best" consensus tree, both analyses gave proportional confidence values.

The phylogram from this analysis separated all the species examined in this study. Within the *Blarina* group, *B. brevicauda*, *B. carolinensis* and *B. hylophaga* split into three clades (Figure 3.4). In this phylogram, *B. hylophaga* is in a basal position to *B. brevicauda* and *B. carolinensis* due to rooting with outgroups and rotating branches, which showed a similar topology to that in Brant and Ortí (2002). Before this was change was done, the phylogram showed *B. brevicauda* in

a basal position from *B. carolinensis* and *B. hylophaga*, but had the outgroups split off from *B. hylophaga*. Jones et al. (1984) concluded that an ancestral species similar to *B. brevicauda* divided to become *B. brevicauda* and *B. carolinensis*, and *B. hylophaga* split from *B. carolinensis*. The genetic distances in this study (Table 3.2) indicate that *B. brevicauda* is more similar to *B. carolinensis* than to *B. hylophaga*, and *B. carolinensis* is more similar *B. hylophaga* than to *B. brevicauda*. Therefore, the arrangement of *B. carolinensis* between the other two *Blarina* species is correct, but *B. brevicauda* should be in a basal position from *B. carolinensis* and *B. hylophaga*.

The *B. brevicauda* clade (Figure 3.5) shows a split of specimens at node 9 and 10, which splits North Carolina and other eastern states from Alabama, Kentucky, Iowa and Wisconsin specimens. I did not have tissue samples from northeast North Carolina, which was the location for the former subspecies *B. b. telmalestes* (Merriam, 1895a; Paul, 1965) in the Dismal Swamp. Samples from high elevations in western North Carolina represent the former subspecies *B. b. churchi*, while the samples from Polk County would have been *B. b. kirtlandi* (Bole and Moulthrop, 1942; Lee, et al., 1982). Webster et al. (2011) combined these former subspecies into *B. b. talpoides* by comparing cranial characters and the geographic zone of intergradation. My results support the decision of Webster et al. to reduce all previous subspecies in the mountains to *B. b. talpoides*. See Appendix IV for the locations of each specimen used in the genetic analyses.

The *B. brevicauda* sample from eastern North Carolina is from Holly Ridge in Onslow County (NCSM16801), and represents the locality region for *B. b. knoxjonesi* (Webster, 1996; Webster et al., 2011). This is the only specimen of *B. brevicauda* from the Middle Atlantic Coastal Plain, while the remainder of the North Carolina specimens came from the Blue Ridge

ecoregion. If *B. b. knoxjonesi* is a subspecies, it should show genetic divergence and separation from the remainder of the *B. b. talpoides* group. The sequence from NCSM 16801 grouped with the sequence from NCSM 17595 (NC, Watauga County, Banner Elk) suggesting very little genetic variation between the western and eastern part of the state for this species. These two samples also grouped with GenBank samples from New York, Pennsylvania, Virginia and West Virginia suggesting that there is little variation between North Carolina *B. brevicauda* and these eastern states to the north.

Besides the lack of confidence imparted by a sample size of one, the limitations of mitochondrial DNA prevent a conclusive result for the taxonomic status of *Blarina brevicauda knoxjonesi* with this genetic analysis. Mitochondrial DNA (mtDNA) is only inherited from the mother. If the sample (NCSM16801) represented a first filial (F1) generation and the mother was not *B. b. knoxjonesi*, the 'knoxjonesi' group would not be present with the mtDNA sequences even if the father was *B. b. knoxjonesi*. Therefore, additional mtDNA sequences would most likely capture the variation with a larger sample size, but analyzing other molecular markers like nuclear DNA that is inherited by both parents would ensure differences are not missed based on genetic inheritance. It is also possible that *B. b. knoxjonesi* represents a newly forming subspecies that has not been isolated long enough to show sequence divergence. Additional samples of *B. b. knoxjonesi* should also investigate divergence times to examine this possibility. With the added information about the limitations of mtDNA, my results are unable to propose synonymizing *B. b. knoxjonesi* with the *B. b. talpoides* subspecies. Therefore, I suggest that *B. b. knoxjonesi* is retained as a recognized subspecies in southeast North Carolina until further genetic research occurs.

The *B. carolinensis* clade (Figure 3.6) showed early separations at node 6 and node 7 with GenBank samples from Arkansas, Illinois and Florida, but then showed little variation at node 8 with very short branch lengths in the phylogram. All the samples from North Carolina represent the subspecies *B. c. carolinensis*, which continues north to south-central Virginia and south into Florida (Bennedict et al., 2006; Genoways and Choate; 1998; McCay, 2001; Webster, 1985). NCSM 16800 is from Morganton in Burke County and represents a county record and the western-most specimen in North Carolina for *B. carolinensis*. The GenBank sample from Georgia (AF395449) grouped within the North Carolina samples suggesting that there is little genetic difference in *B. carolinensis* between these two states, and presumably in South Carolina as well.

Non-North Carolina *Blarina* samples

The *B. brevicauda* samples from Jackson County in Alabama (NCSM13838, 13839, 14397, 14383 and 13843) separated into a clade different from the North Carolina specimens at node 10 (Figure 3.5). Theses samples are grouped close to GenBank samples AF534121 and AF533632 from Trigg County, Kentucky. Linear geographic distances calculated in the mapping software DeLorme Topo North America version 10.0 (DeLorme, 2013) show the closest distance for an Alabama genetic sample (NCSM 13843) to a North Carolina specimen is a little over 190 km (distance to NCSM 14639 – NC, Graham Co., Sand Creek) and 200 km (distance to NCSM 13770 – NC, Clay Co., Leatherwood Branch). These distances are slightly shorter than the North Carolina mountain samples: NCSM 17595 (NC, Watauga Co., Bear Paw State Natural Area) to NCSM 14639 is a little over 210 km, and NCSM 17595 to NCSM 13770 is almost 220 km in linear distance. In contrast, the *B. b. knoxjonesi* sample (NCSM 16801) was over 410 km from

NCSM 17595. If sequences diverge between the Alabama and North Carolina specimens but not the North Carolina mountain specimens, something beyond distance must be preventing gene flow between these locations in the two states.

The Alabama specimens are located in the southern extent of the Cumberland Plateau, and are divided by the Tennessee River to the east of these specimens' localities. Webster et al. (2011) described a new subspecies *Blarina brevicauda cumberlandensis* that has a distribution of the "Cumberland Plateau region of central Tennessee and adjacent parts of western and south-central Kentucky." They also suggested that *B. b. cumberlandensis* "likely occurs in northern Alabama, north of the Tennessee River." These Alabama specimens match this description, and I propose that the samples from Alabama are *B. b. cumberlandensis*. Additional genetic and morphometric analyses are needed, but genetic sequence divergence and location supports this subspecies designation. Likewise, I assume that the samples from Kentucky also represent this subspecies based on locality and genetic separation with cytochrome-b sequences.

The samples from Iowa and Wisconsin also split from the North Carolina group in this study's phylogram (Figure 3.5). The sample from Iowa (AF533609) is from Muscatine County near the eastern border of Iowa and is within the locality of *Blarina brevicauda brevicauda* (Webster et al., 2011). GenBank sample AF533642 is from Columbia County, Wisconsin and is within the locality of the northwestern extent of *B. b. talpoides* (Webster et al., 2011). The locality for the sample from Iowa is slightly more than 1000 km, and the sample from Wisconsin is almost 1050 km, from the sample in the northwest corner of North Carolina (NCSM 17599 – NC, Ashe Co., Fleetwood) based on measurements in DeLorme 10.0. In contrast, the GenBank sample from Hamilton County, New York (AF534115) had the greatest distance to Ashe County, North Carolina (NCSM 17599) with an approximate geographic distance of 1060 km, and grouped with

all the North Carolina samples. It is difficult to separate out the difference between sequence divergence based on subspecies identification versus the separation based geographical distances, but the distance from North Carolina to New York exceeded the distance to both Iowa and Wisconsin. Therefore, something else in the landscape is causing this difference in genetic sequences to diverge between these locations. Additional sampling is needed in these areas to differentiate whether the sequence variation observed is a gradual change across the landscape, or if it represents an abrupt change, and subspecies status is warranted.

All the GenBank *Blarina carolinensis* cytochrome-b sequence samples, except that from Georgia, separated from the North Carolina sequence samples. The GenBank sample from Arkansas (AF395456) and samples from Illinois (AF395458, AF395460) grouped together, while the sample from Florida (AF395454) grouped as an independent branch in the phlyogram (Figure 3.6). The linear geographic distances calculated in DeLorme 10.0 from the Polk County in Arkansas to the geographic county center of Jackson County, Illinois is roughly 560 km, but compared to the closest North Carolina specimen (NCSM 16800 – NC, Burke County, Morganton), the distance is greater than 600 km to Arkansas and 1260 km to Illinois. These two localities either represent the western *B. c. carolinensis* or *B. c. minima* subspecies, but the western part of the range of the southern-short tailed shrew is uncertain (Genoways and Choate; 1998; McCay, 2001).

The sequence sample from Florida is from Highlands County, and it is located in the southern half of the state. The closest North Carolina specimen to this is NCSM 17592 (NC, Gaston Co., Belmont) and is just over 880 km away from the Florida sample. This sample from Florida is in the locality designated as *Blarina carolinensis peninsulae* (Bennedict et al., 2006; Genoways and Choate; 1998; McCay, 2001). Wilson and Reeder (2005) named the Everglades short-tailed

shrew (*Blarina peninsulae*) as a distinct species based on differences in the karyotype and morphology (Genoways and Choate, 1998; George et al., 1982), but my analysis questions the species designation. The *Blarina carolinensis* clade (Figure 3.6) shows the Florida sample from the cytochrome-b sequence within this group, and not in its own separate clade, suggesting that it should remain a subspecies status only.

The *Blarina hylophaga* samples (AY546659, AY546662, AY546669, AY546681 and AY546677) were all from Texas. The first four were from Bastrop County, Texas and represent *B. h. hylophaga*, while the last sample (AY546677) was from Aransas County, Texas and represents *B. h. plumbea* (George et al., 1981; Thompson et al., 2011). All samples separated into a separate clade at node 4, with AY546677 having a longer branch length showing a slight divergence in the sequence from the other samples of *B. hylophaga* (Figure 3.4).

LITERATURE CITED

Alfaro M. E., S. Zoller and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Molecular Biology and Evolution 20(2): 255-266.

Ausubel F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, K. Struhl. 1992. Short Protocols in Molecular Biology 2nd edition. Greene Publishing Associates and John Wiley & Sons, New York.

Avise, J. C. 2004. Molecular markers, natural history, and evolution. 2nd edition. Sinauer Associates, Inc., Sunderland, Massachusetts. 684 pp.

Baker, R. J. and R. D. Bradley. 2006. Speciation in mammals and the genetic species concept. Journal of Mammalogy 87(4): 643-662.

Beck, M. L., C. J. Biggers and J. A. Huggins. 1991. Variation in chromosome number in the southern short-tailed shrew *Blarina carolinensis*. Mammalia 55(4): 623-625.

Bellinger, M. R., S. M. Haig, E. D. Forsman and T. D. Mullins. 2005. Taxonomic relationships among *Phenacomys* voles as inferred by cytochrome b. Journal of Mammalogy 86(1): 201-210.

Benedict, R. A. 1999a. Morphological and mitochondrial DNA variation in a hybrid zone between short-tailed shrews (*Blarina*) in Nebraska. Journal of Mammalogy 80(1): 112-134.

Benedict, R. A. 1999b. Characteristics of a hybrid zone between two species of short-tailed shrews (*Blarina*). Journal of Mammalogy 80(1): 135-141.

Benedict, R. A., H. H. Genoways, and J. R. Choate. 2006. Taxonomy of short-tailed shrews (genus *Blarina*) in Florida. Occasional Papers, Museum of Texas Tech University 251: 1-19.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers. 2009. GenBank. Nucleic Acids Research 37(Database issue): D26-31.

Bickham, J. W., C. C. Wood and J. C. Patton. 1995. Biogeographic implications of cytochrome b sequences and allozymes in sockeye (*Oncorhynchus nerka*). Journal of Heredity 86: 140-144.

Bickham, J. W., J. C. Patton, D. A. Schlitter, I. L. Rautenbach and R. L. Honeycutt. 2004. Molecular phylogenetics, karyotypic diversity, and partition of the genus *Myotis* (Chiroptera: Vespertilionidae). Molecular Phylogenetics and Evolution 33: 333-338.

Blois, J. L. and B. S. Arbogast. 2006. Conservation genetics of the Sonoma tree vole (*Arborimus pomo*) based on mitochondrial and amplified fragment length polymorphism markers. Journal of Mammalogy 87(5): 950-960.

Bole, B. P., Jr., and P. N. Moulthrop. 1942. The Ohio Recent mammal collection in the Cleveland Museum of Natural History. Scientific Publication, Cleveland Museum of Natural History, 5:83-181.

Bradley, R. D. and R. J. Baker. 2001. A test of the genetic species concept: cytochrome-b sequences and mammals. Journal of Mammalogy 82(4): 960-973.

Bradley, R. D., N. D. Durish, D. S. Rogers, J. R. Miller, M. D. Engstrom and C. W. Kilpatrick. 2007. Toward a molecular phylogeny form *Peromyscus*: evidence from mitochondrial cytochrome-b sequences. Journal of Mammalogy 88(5): 1146-1159.

Brant, S. V. and G. Ortí. 2002. Molecular phylogeny of short-tailed shrews, *Blarina* (Insectivora: Soricidae). Molecular Phylogenetics and Evolution 22(2): 163-173.

Brant, S. V. and G. Ortí. 2003. Phylogeography of the northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): past fragmentation and postglacial recolonization. Molecular Ecology 12(6): 1435-1449.

Burnham, K. P. and D. R. Anderson. 2002. Model selectin and multimodel inference: a practical information theoretic approach. Ecological Modelling Vol. 2. Springer, New York. 488pp.

Burnham, K. P. and D. R. Anderson. 2004. Multimodel inference understanding AIC and BIC in model selection. Sociological methods & research 33(2): 261-304.

Camacho-Sanchez, M., P. Burraco, I. Gomez-Mestre and J. A. Leonard. 2013. Preservation of RNA and DNA from mammal samples under field conditions. Molecular Ecology Resources 13(4): 663-673.

Carr, S. P. 2014. Transition versus transversion mutations. Accessed at: http://www.mun.ca/biology/scarr/Transitions_vs_Transversions.html.

Carson, H. L. 157. The species as a field for gene recombination. Pp 23-38 in The Species Problem (E. Mayer, ed.). American Association for the Advancement of Science, Washington, D. C.

Choate, J. R. 1970. Systematics and zoogeography of Middle American shrews of the genus *Cryptotis*. University of Kansas. Museum of Natural History.

Corriveau, M., K. Uppstrom and H. Klompen. 2009. Effect of eight storage modes on DNA preservation. Trends in Acarology: 553-556.

Crick, F. H. C. 1966. Codon-anticodon pairing: the wobble hypothesis. Journal of Molecular Biology 19(2): 548-555.

Darriba, D., G. L. Taboada, R. Doallo and D. Posada. 2012. JModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9(8): 772-772.

DeLorme. 2013. Topo North America version 10.0. Two DeLorme Drive. P.O. Box 298. Yarmouth, ME, 04096. USA.

Dobzhansky, T. 1950 Mendelian populations and their evolution. The American Naturalist. 84: 401-418.

Driskell, A. C and G. A. Feldhamer. 2003. Allozyme differentiation in five species of North American shrews (Soricidae: Mammalia). Biochemical Systematics and Ecology 31: 739-750.

Elrod, D. A., M. L. Beck and M. L. Kennedy. 1996. Chromosomal variation in the southern short-tailed shrew (*Blarina carolinensis*). Geneica 98(2): 199-203.

Farris, J. S. 1970. Methods for computing Wagner trees. Systematic Zoology 19: 83-92.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17(6): 368-376.

Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. Systematic Zoology 20(4): 406-416.

Fumagalli, L., P. Taberlet, D. T. Stewart, L. Gielly, J. Hausser and P. Vogel. 1999. Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. Molecular Phylogenetics and Evolution 11: 222-235.

Garcia-Sandoval, R. 2014. Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. Israel Journal of Ecology & Evolution 60(1): 41-44.

Genoways, H. H. and J. R. Choate. 1998. Natural history of the southern short-tailed shrew, *Blarina carolinensis*. Occasional Papers, The Museum of Southwestern Biology 8: 1-43.

Genoways, H. H., J. C. Patton, III and J. R. Choate. 1977. Karyotypes of shrews of the genera *Cryptotis* and *Blarina* (Mammalia: Soricidae). Experientia 33:1294-1295.

George, S. B., J. R. Choate and H. H. Genoways. 1981. Distribution and taxonomic status of *Blarina hylophaga* Elliot (Insectivora: Soricidae). Annals of Carnegie Museum 50(21): 493-513.

George, S. B., J. R. Choate and H. H. Genoways. 1986. *Blarina brevicauda*. Mammalian Species 261: 1-9.

George, S. B., H. H. Genoways, J. R. Choate and R. J. Baker. 1982. Karyotypic relationships within the short-tailed shrews, genus *Blarina*. Journal of Mammalogy 63(4): 639-645.

Guindon, S. and O. Gascuel. 2003. A simple, fast, and accurate method to estimate large phylogenies by maximum likelihood. Systematic Biology 52(5): 696-704.

Hafner, M. S., W. L. Gannon, J. Salazar-Bravo and S. T. Alvarez-Castañedo. 1997. Mammal collections in the Western Hemisphere: a survey and directory of existing collections. Allen Press, Lawrence, Kansas. 93 pp.

Hall, E. R. 1981. The Mammals of North America. 2nd edition. John Wiley and Sons, New York. 1181 pp.

Harris, D., D. S. Rogers and J. Sullivan. 2000. Phylogeography of *Peromyscus furvus* (Rodentia; Muridae) based on cytochrome b sequence data. Molecular Ecology 9(12): 2129-2135.

Helgen, K. M., C. M. Pinto, R. Kays, L. E. Helgen, M. T. N. Tsuchiya, A. Quinn, D. E. Wilson and J. E. Maldonado. 2013. Taxonomic revision of the olingos (*Bassaricyon*), with description of a new species, the olinguito. ZooKeys 324: 1-83.

Hasegawa, M., H. Kishino and T. Yano. 1985. Dating of human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22(2): 160-174.

Huelsenbeck, J. P. and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Systematic Biology 53(6): 904-913.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17(8): 754-755.

Hurvich, C. M. and C. L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76(2): 297-307.

Jones, C. A., J. R. Choate and H. H. Genoways. 1984. Phylogeny and paleobiogeography of short-tailed shrews (genus *Blarina*). Pp. 56-148, in Contributions in Quaternary Vertebrate Paleontology: a volume in memorial to John E. Guilday (Hugh H. Genoways and Mary R. Dawson, editors). Special Publication of the Carnegie Museum of Natural History 58(8): 1-538.

Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-123 in Mammalian Protein Metabolism (H. N. Munro, ed.). Academic Press, New York, New York.

Katoh, K. and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30(4): 772-780.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies and A. Drummond. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647-1649.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16(2): 111-120.

Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proceedings of the National Academy of Sciences 78(1): 454-458.

Kishino, H. and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution 29(2): 170-179.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, F. X. and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences 86(16): 6196-6200.

Lanave, C., G. Preparata, C. Saccone and G. Serio. 1984. A new method for calculating evolutionary substitution rates. Journal of Molecular Evolution 20(1): 86-93.

Lanfear, R., B. Calcott, S. Y. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29(6): 1695-1701.

Larsen, R. J., M. C. Knapp, H. H. Genoways, F. A. A. Kahn, P. A. Larson, D. E. Wilson, R. J. Baker. 2012. Genetic Diversity of Neotropical *Myotis* (Chiroptera: Vespertilionidae) with an emphasis on South America species. PLoS ONE 7(10): e46578.

Lee, D. S., J. B. Funderburg, Jr. and M. K. Clark. 1982. A distributional survey of North Carolina mammals. Occasional Papers of the North Carolina Biological Survey. 72 pp.

Lipman, D. J. and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. Sciences 227(4693): 1435-1441.

Maddison D. R., D. L. Swofford and W. P. Maddison. 1997. NEXUS: an extensible file format for systematic information. Systematic Biology 46(4): 590-621.

Maldonado, J. E., C. Vilà and R. K. Wayne. 2001. Tripartite genetic subdivisions in the ornate shrew (*Sorex ornatus*). Molecular Ecology 10(1): 127-147.

Masters, J. C. and H. G. Spencer. 1989. Why we need a new genetic species concept. Systematic Zoology 38(3): 270-279.

Masuda, R. and M. C. Yoshida. 1994. A molecular phylogeny of the family Mustelidae (Mammalia, Carnivora), based on comparison of mitochondrial cytochrome b nucleotide sequences. Zoological Science 11(4): 605-612.

Mayr, E. 1969. Principles of systematic zoology. McGraw-Hill, New York.

McCay, T. S. 2001. *Blarina carolinensis*. Mammalian Species 673: 1-7.

Merriam, C. H. 1895a. Revision of the shrews of the American genera *Blarina* and *Notiosorex*. North American Fauna 10: 5-34, 102-107.

Merriam, C. H. 1895b. Synopsis of the American shrews of the genus *Sorex*. North American Fauna 10: 57-97.

Miller, M. A., W. Pfeiffer and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010. Pp. 1-8.

Murtagh F. 1984. Complexities of Hierarchic Clustering Algorithms: the state of the art. Computational Statistics Quarterly 1: 101–113.

Ohdachi, S. D., M. Hasegawa, M. A. Iwasa, P. Vogel, T. Oshida, L-K. Lin and H. Abe. 2006. Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome b gene sequences: with special reference to the Soricinae. Journal of Zoology 270(1): 177-191.

Paul, J. R. 1965. *Blarina telmalestes* in North Carolina. Journal of Mammalogy 46(3): 496.

Petersen, K. and T. Yates. 1980. *Condylura cristata*. Mammalian Species 129: 1-4.

Pfau, R. S. and J. K. Braun. 2013. Occurrence of the northern short-tailed shrew (Mammalia: Soricomorpha: Soricidae: *Blarina brevicauda*) in Oklahoma. Proceedings of the Oklahoma Academy of Science 93: 1-6.

Pfau, R. S., D. B. Sasse, M. B. Connior and I. F. Guenther. 2011. Occurrence of *Blarina brevicauda* in Arkansas and notes on the distribution of *Blarina carolinensis* and *Cryptotis parva*. Journal of the Arkansas Academy of Science 65: 61-66.

Posada, D. 2003. Using ModelTest and PAUP to select a model of nucleotide substitution. Current Protocols in Bioinformatics 6.5: 6.5.1-6.5.14.

Posada, D. 2008. JModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25(7): 1253-1256.

Posada, D. and Crandall, K. A. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14(9): 817-818.

Posada, D. and K. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. Systematic Biology 50(4): 580-601.

Qumsiyeh, M.B., J. L Coate, J. A. Peppers, P. K. Kennedy and M. L. Kennedy. 1997. Robertsonian chromosomal rearrangements in the short-tailed shrew, *Blarina carolinensis*, in western Tennessee. Cytogenetics and Cell Genetics 76: 153-158.

Rambaut, A. 2012. FigTree: tree figure drawing tool version 1.4.2. Institute of Evolutionary Biology, University of Edinburgh, Scotland. Accessed at: http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut, A., M. A. Suchard, D. Xie and A. J. Drummond. 2014. Tracer version 1.6. Institute of Evolutionary Biology, University of Edinburgh, Scotland. Accessed at: http://beast.bio.ed.ac.uk/Tracer.

Rannala, B. and Z. Yang. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. Journal of Molecular Evolution 43(3): 304-311.

Reilly, S. M., R. W. Manning, C. C. Nice and M. R. J. Forstner. 2005. Systematics of isolated populations of short-tailed shrews (Soricidae: *Blarina*) in Texas. Journal of Mammalogy 86(5): 887-894.

Rodriguez, F., J. L. Oliver, A. Marín and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142(4): 485-501.

Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19(12): 1572-1574.

Ronquist, F., J., Huelsenbeck and M. Teslenko. 2011. MrBayes version 3.2 manual: tutorials and model summaries. 183 pp.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539-542.

Rosalind. 2012. On-line learning bioinformatics platform. Transisitions-transversions image. Accessed at: http://rosalind.info/media/problems/tran/transitions-transversions.png.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4(4): 406-425.

Sayers, E. W., T. Barret, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D.

Landsman, D. J. Lipman, T. L. Madden, D. R. Maglott, V. Miller, I. Mizrachi, J. Ostell, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, E. Yaschenko and J. Ye. 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 37(database issue): D5-15.

Schmid, F., P. Linnell, C. Bradney, P. Vaněk. R. Leino, A. Vox, J. Ghali, O. Moskalenko, C. Schäfer, H. Kraus, P. Marchand, J. B. Aliyas, P. Smedley. 2015. Scribus: open source desktop publishing version 1.4.5. Accessed at: http://sourceforge.net/projects/scribus/files/scribus/1.4.5/.

Shokralla, S., G. A. C. Singer and M. Hajibabaei. 2010. Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. BioTechniques 48(3): 233-234.

Simpson, G. G. 1943. Criteria for genera, species, and subspecies in zoology and paleontology. Annals of the New York Academy of Sciences 44:145–178.

Sokal, R. and C. Michener. 1958. A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38: 1409-1438.

Stebbins, G.L. 1950. Chapter XII: the karyotype in variation and evolution in plants. Columbia University Press, New York.

Stelkens, R. and O. Seehausen. 2009. Genetic distances between species predict novel trait expression in their hybrids. Evolution 63(4): 884-897.

Sudman, P. D., J. K. Wickliffe, P. Horner, M. J. Smolen, J. W. Bickham, and R. D. Bradley. 2006. Molecular systematics of pocket gophers of the genus *Geomys*. Journal of Mammalogy 87(4): 668-676.

Sugiura, N. 1978. Further analysts of the data by Akaike's information criterion and the finite corrections. Communication in Statistics-Theory and Methods 7(1): 297-307.

Sullivan, J. and P. Joyce. 2005. Model selection in phylogenetics. Annual Review of Ecology, Evolution, and Systematics 36: 445-466.

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tamura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10(3): 512-526.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance,

and maximum parsimony methods. Molecular Biology and Evolution 28(10): 2731-2739.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Pp. 57-86 in American Mathematical Society: Lectures on Mathematics in the Life Sciences.

Thompson, C. W., J. R. Choate, H. H. Genoways and E. J. Finck. 2011. *Blarina hylophaga* (Soricomorpha: Soricidae). Mammalian Species 43(878): 94-103.

Webster, W. D., J. F. Parnell, and W. C. Biggs, Jr. 1985. The mammals of the Carolinas, Virginia, and Maryland. University of North Carolina Press, Chapel Hill, 255 pp.

Webster, W. D. 1996. Geographic variation in *Blarina brevicauda* (Insectivora: Soricidae) from eastern North Carolina, with description of a new subspecies. Pp. 47-56, *in* Contributions in mammalogy: a memorial volume honoring Dr. J. Knox Jones, Jr. (H. H. Genoways and R. J. Baker, eds.). Museum of Texas Tech University, Lubbock, 315 pp.

Webster, W. D., N. D. Moncrief, J. R. Choate and H. H. Genoways. 2011. Systematic revisions of the northern short-tailed shrew, *Blarina brevicauda* (Say). Virginia Museum of Natural History. Memoir 10. 77 pp.

Whitaker, J. 1974. *Cryptotis parva*. Mammalian Species 43: 1-8.

White, M. J. D. 1973. Animal cytology and evolution. 3rd edition. Cambridge University Press, England.

Willows-Munro, S. and C. A. Matthee. 2011. Exploring the diversity and molecular evolution of shrews (family Soricidae) using mtDNA cytochrome b data. African Zoology 46(2): 246-262.

Wilson, D. E. and D. M. Reeder (editors). 2005. Mammal species of the world: a taxonomic and geographic reference. 3rd edition. The Johns Hopkins University Press, Baltimore, Maryland. 2142 pp.

Yang, Z. and B. Rannala. 1997. Bayesian phylogenetic inference using DNA sequences: a Marcov Chain Monte Carlo method. Molecular Biology and Evolution 14(7): 717-724.

Zharkikh, A. 1994. Estimation of evolutionary distances between nucleotide sequences. Journal of Molecular Evolution 39(3): 315-329.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin. Accessed at: http://www.bio.utexas.edu/faculty/antisense/garli/garli.html.

Zwickl, D. J. 2008. GARLI manual, program version 0.96. Accessed at: http://www.phylo.org/tools/Garli0.96Manual.pdf.

Table 3.1: Named substitution models in jModelTest2 compared for model selection. These 22 models (a few of the 1624 possible) can include invariable sites (+I), rate variation among sites (+G), or both (+I+G) for a combination of 88 models tested. This table includes: model abbreviation and reference for that model, nst is the total number of substitution rates used in the model prior parameters, base frequency is the proportion of each nitrogenous base or nucleobase (A, C, G, T) in the model and whether it is equal or unequal (estimated), the free parameters are the number of branch lengths needed to be estimated for every model, the detailed substitution rates in the model, and the substitution code used for GARLI (Zwickl, 2006). Details for this table were taken from jModelTest2 manual (Darriba et al., 2012) and a summary about these substitution models (http://www.molecularevolution.org/resources/models/nucleotide).

Model	Reference	nst	Base Frequency	Free parameters	Substitution Rates	Substitution Code
JC	Jukes and Cantor, 1969	1	equal	0	AC=AG=AT=CG=CT=GT	000000
F81	Felsenstein, 1981	1	estimated	3	AC=AG=AT=CG=CT=GT	000000
K80	Kimura, 1980	2	equal	1	AC=AT=CG=GT;AG=CT	010010
HKY	Hasegawa et al., 1985	2	estimated	4	AC=AT=CG=GT;AG=CT	010010
TrNef	Tamura and Nei, 1993	6	equal	2	AC=AT=CG=GT;AG;CT	010020
TrN	Tamura and Nei, 1993	6	estimated	5	AC=AT=CG=GT;AG;CT	010020
TPM1 = K81	Kimura, 1981	6	equal	2	AC=GT;AG=CT;AT=CG	
TPM1uf = K81uf	Kimura, 1981	6	estimated	5	AC=GT;AG=CT;AT=CG	012210
TPM2	See JModelTest Manual	6	equal	2	AC=AT;CG=GT;AG=CT	010212
TPM2uf	See JModelTest Manual	6	estimated	5	AC=AT;CG=GT;AG=CT	010212
TPM3	See JModelTest Manual	6	equal	2	AC=CG;AT=GT;AG=CT	012012
TPM3uf	See JModelTest Manual	6	estimated	5	AC=CG;AT=GT;AG=CT	012012
TIM1ef	Posada, 2003	6	equal	3	AC=GT;AT=CG;AG;CT	012230
TIM1	Posada, 2003	6	estimated	6	AC=GT;AT=CG;AG;CT	012230
TIM2ef	See JModelTest Manual	6	equal	3	AC=AT;CG=GT;AG;CT	010232
TIM2	See JModelTest Manual	6	estimated	6	AC=AT;CG=GT;AG;CT	010232
TIM3ef	See JModelTest Manual	6	equal	3	AC=CG;AT=GT;AG;CT	012032
TIM3	See JModelTest Manual	6	estimated	6	AC=CG;AT=GT;AG;CT	012032
TVMef	Posada, 2003	6	equal	7	AC;CG;AT;GT;AG=CT	012314
TVM	Posada, 2003	6	estimated	4	AC;CG;AT;GT;AG=CT	012314
SYM	Zharikikh, 1994	6	equal	5	AC;CG;AT;GT;AG;CT	012345
GTR	Lanave et al., 1994; Rodriguez et al., 1990; Tavare, 1986	6	estimated	8	AC;CG;AT;GT;AG;CT	012345

Table 3.2: Uncorrected pairwise genetic distances for cytochrome-b sequences from samples in the genus *Blarina* and selected outgroups calculated in PAUP* 4.0a147 software (Swofford, 2002). Distance values greater than 0.02 (>2.0%) indicate interspecific sequence divergence, while values less than 0.02 indicate intraspecific sequence divergence (Larson et al., 2012). Legend: Blbr = *Blarina brevicauda*, Blca = *Blarina carolinensis*, Blhy = *Blarina hylophaga*, NC = North Carolina, AL = Alabama, IA = Iowa, WI = Wisconsin, KY = Kentucky, TN = Tennessee, VA = Virginia, WV = West Virginia, PA = Pennsylvania, NY = New York, GA = Georgia, AR = Arkansas, IL = Illinois, FL = Florida, Outgroups = (Crpa = *Cryptotis parva*, Soci = *Sorex cinereus*, Solo = *Sorex longirostris*, Cocr = *Condylura cristata*), Blsp = all *Blarina* species. IA, WI, KY, TN, VA, WV, PA, NY, GA, AR, IL and FL samples downloaded from GenBank.

Taxa	n = Total Samples	Distances Compared	Mean	Minimum	Maximum	Variance	Standard Deviation
Blbr vs Blbr	45	1035	0.00902	0.00000	0.02685	2.57E-05	0.00507
Blca vs Blca	43	946	0.00921	0.00000	0.03981	0.000134	0.01159
Blhy vs Blhy	4	10	0.00296	0.00000	0.00741	7.78E-06	0.00279
Blbr vs Blca	90	2024	0.07575	0.06389	0.08611	9.86E-06	0.00314
Blbr vs Blhy	51	230	0.08704	0.08148	0.09444	8.51E-06	0.00292
Blca vs Blhy	49	220	0.06461	0.05926	0.07130	5.04E-06	0.00224
NC Blbr vs NC Blbr	31	465	0.00620	0.00000	0.01204	5.30E-06	0.00230
AL Blbr vs AL Blbr	5	10	0.00371	0.00000	0.00833	7.42E-06	0.00272
AL Blbr vs NC Blbr	35	155	0.01159	0.00833	0.01667	3.26E-06	0.00181
GenBank Blbr vs NC Blbr	41	341	0.01087	0.00093	0.02593	4.16E-05	0.00645
IA Blbr vs NC Blbr	32	31	0.02410	0.02222	0.02593	1.86E-06	0.00136
WI Blbr vs NC Blbr	32	31	0.01398	0.01204	0.01759	2.07E-06	0.00144
KY Blbr vs NC Blbr	32	31	0.01674	0.01481	0.01944	1.56E-06	0.00125
AL Blbr vs KY Blbr	7	10	0.01361	0.01204	0.01574	1.53E-06	0.00124
AL Blbr vs TN Blbr	7	10	0.01176	0.00926	0.01481	3.43E-06	0.00185
NC Blbr vs NC Blbr	31	465	0.00620	0.00000	0.01204	5.30E-06	0.00230
AL Blbr vs AL Blbr	5	10	0.00371	0.00000	0.00833	7.42E-06	0.00272
TN Blbr vs NC Blbr	33	62	0.00617	0.00278	0.01019	2.81E-06	0.00168
VA Blbr vs NC Blbr	32	31	0.00687	0.00185	0.01019	3.19E-06	0.00179
WV Blbr vs NC Blbr	32	31	0.00427	0.00093	0.00741	2.67E-06	0.00163
PA Blbr vs NC Blbr	32	31	0.00765	0.00463	0.01111	2.86E-06	0.00169
NY Blbr vs NC Blbr	32	31	0.00600	0.00278	0.00926	2.79E-06	0.00167
Blbr_kn vs All Blbr	46	45	0.00712	0.00093	0.02315	1.95E-05	0.00442
NC Blca vs NC Blca	39	741	0.00401	0.00000	0.01111	4.26E-06	0.00206
GA Blca vs all Blca	43	42	0.00696	0.00185	0.03611	8.19E-05	0.00905
AR Blca vs all Blca	43	42	0.03254	0.00093	0.03981	5.04E-05	0.00710
IL Blca vs all Blca	43	42	0.03163	0.00093	0.03889	4.90E-05	0.00700
IL Blca vs all Blca	43	42	0.03249	0.00093	0.03796	4.98E-05	0.00706
FL Blca vs all Blca	43	42	0.036005	0.03426	0.03981	1.68E-06	0.00130
Outgroups vs Blsp	100	475	0.16725	0.13796	0.19722	0.000271	0.01646
Crpa vs Blsp	97	194	0.15173	0.13796	0.16667	6.99E-05	0.00837
Soci vs Blsp	96	95	0.17057	0.16111	0.17870	3.07E-05	0.00554
Solo vs Blsp	96	95	0.16915	0.16204	0.17778	1.57E-05	0.00396
Cocr vs Blsp	96	95	0.19306	0.18796	0.19722	5.12E-06	0.00226
Crpa vs Soci	3	2	0.18241	0.18056	0.18426	6.85E-06	0.00262
Crpa vs Solo	3	2	0.18056	0.17685	0.18426	2.75E-05	0.00524
Crpa vs Cocr	3	2	0.18241	0.17778	0.18704	4.29E-05	0.00655
Soci vs Cocr	2	1	0.17593	0.17593	0.17593	*******	*******
Solo vs Cocr	2	1	0.17870	0.17870	0.17870	*******	*******
Soci vs Solo	2	1	0.02685	0.02685	0.02685	*******	*******

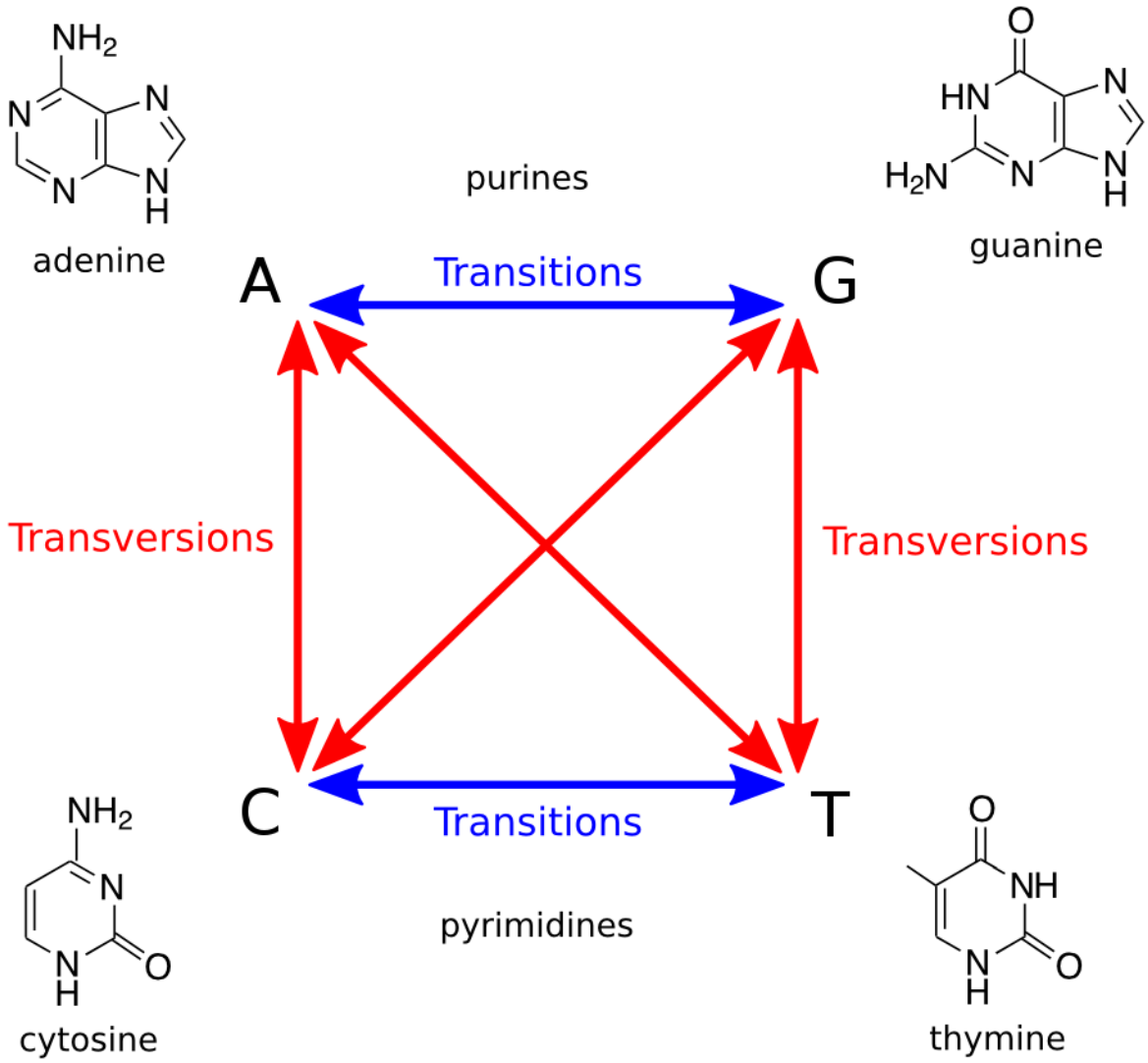


Figure 3.1: DNA base substitution mutations: transitions versus transitions. Transitions are interchanges between the two-ring purines (adenine to guanine, A ←→ G), or between the one-ring pyrimidines (cytocine to thymine, C ←→ T). Transversions are base interchanges between one purine and one pyrimidine (A ←→ C, A ←→ T, C ←→ G, G ←→T). Although there are more possible transversions than transistions, transition base mutations occur at a higher rate and often do not change the amino acids. Image accessed unchanged from Rosalind (2012) and information derived from Carr (2014).

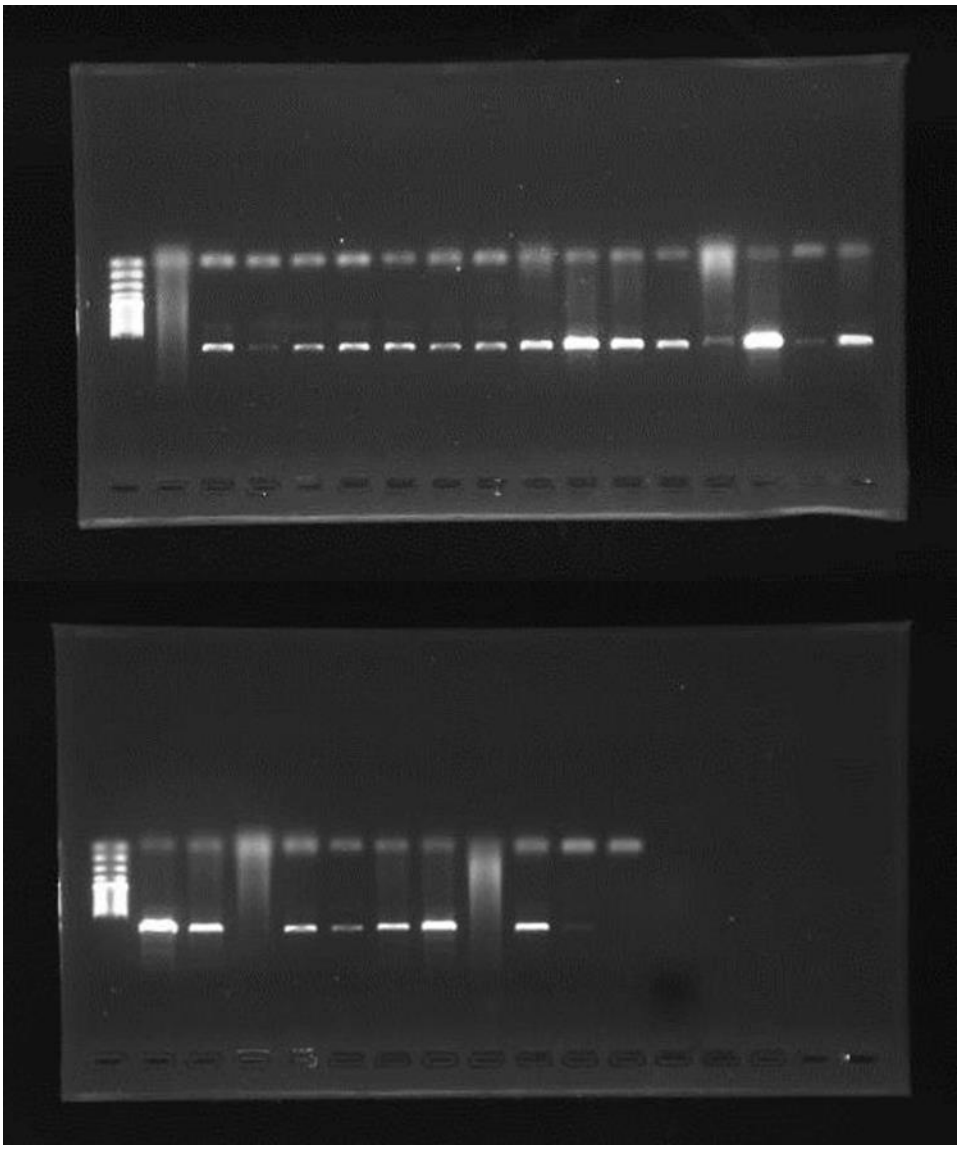


Figure 3.2: Ultraviolet (UV) illuminated gel picture – bright bands represent positive reactions. Both gel pictures are from polymerase chain reaction BMH004 for with a 100 base pair DNA ladder in the left chambers and samples 1-16 (top) and 17-26 (bottom) running left to right respectively. The image was taken with a BioDoc-It FluorCam 220 Camera with a resolution of 1.3MP.

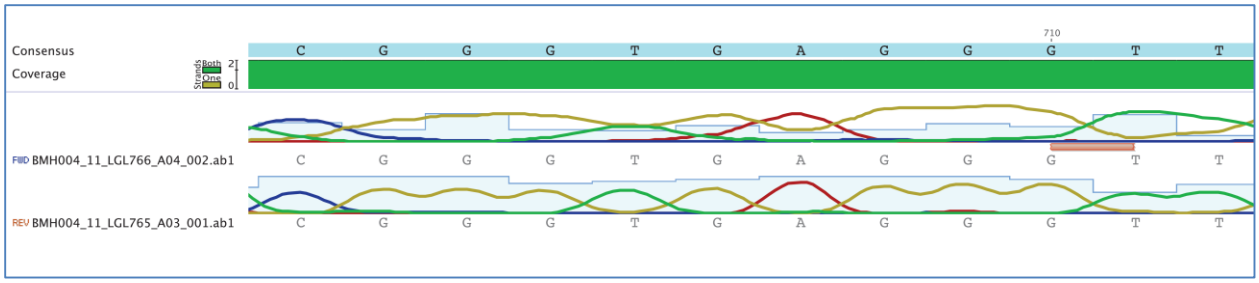


Figure 3.3: DNA sequence chromatogram for visual and qualitative alignment in Geneious version 7 (http://www.geneious.com; Kearse et al., 2012). The overlap region is created by the forward primer LGL 765 and the reverse primer LGL 766 contig assembly until a final consensus sequence is reached. The blue underlying boxes indicate the quality of each respective base call and each base call should yield a definitive peak in the chromatogram. Deletions must be made when the forward and reverse direction do not match indicated by the pink bar at base 710.

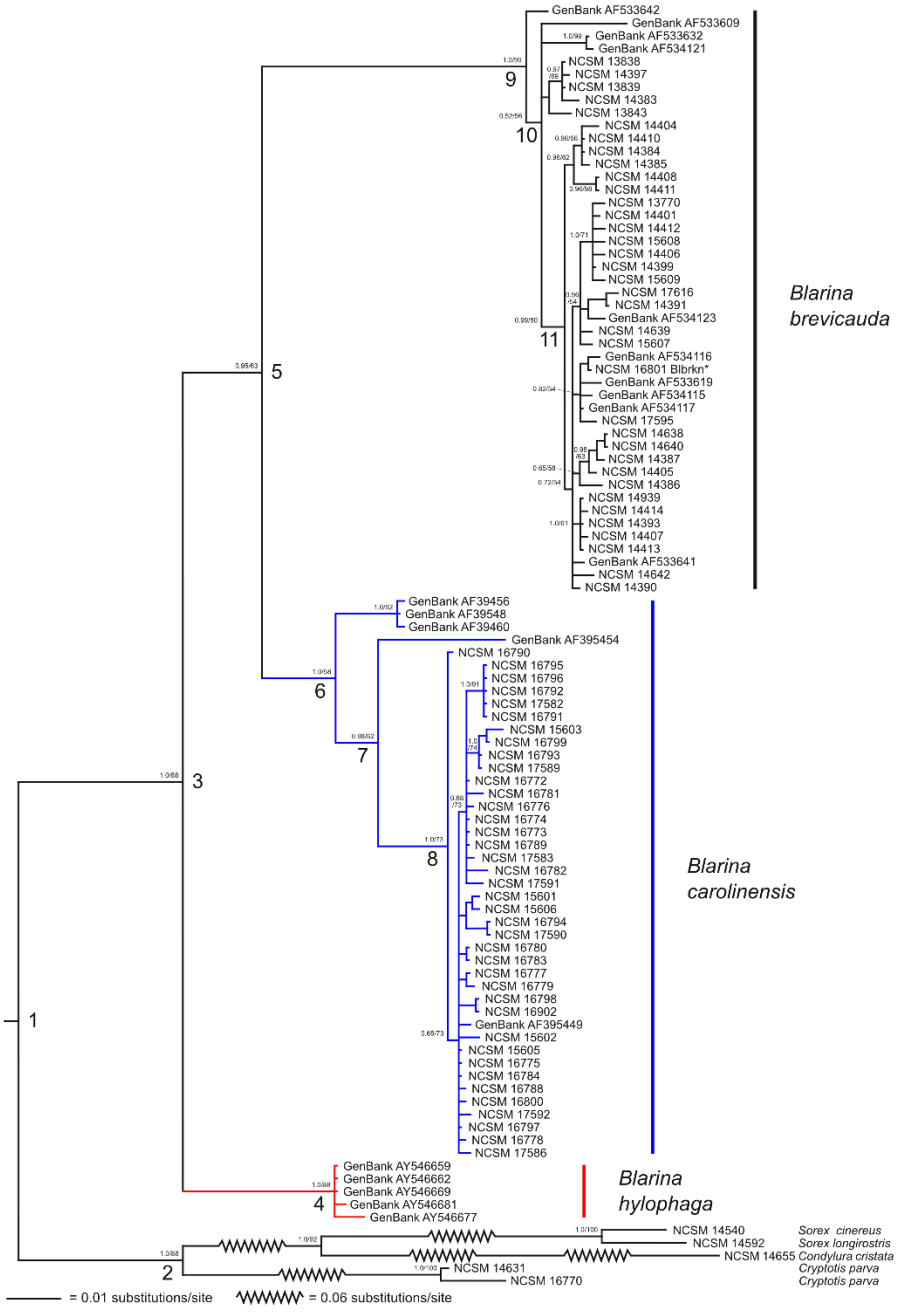


Figure 3.4: Fifty-percent majority-rule consensus phylogram resulting from mixed-model Bayesian analysis of the mitochondrial cytochrome-b gene from short-tailed shrews (genus *Blarina*) and outgroups. Maximum likelihood analysis recovered the same topology. Numbers next to nodes are Bayesian posterior probabilities and bootstrap support values derived from 1000 bootstrap replications. Outgroup lineages at node 2 are depicted as broken due to the extremely long branch length.

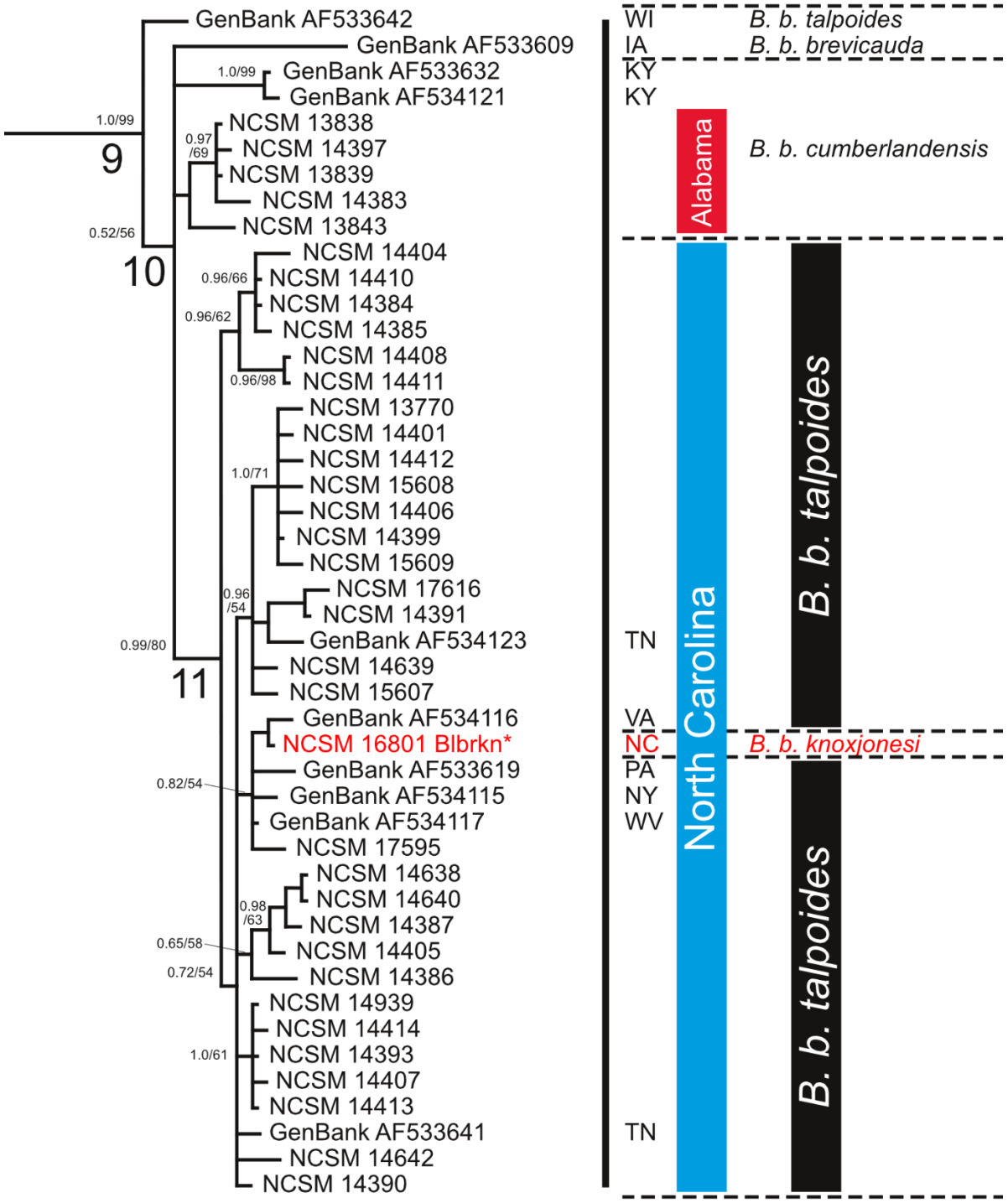


Figure 3.5: Partial fifty-percent majority-rule consensus phylogram showing the *Blarina brevicauda* clade from the Bayesian and maximum likelihood analysis. Subspecies designations are based on the analysis in Webster et al. (2011) and are noted on the right. All samples are from North Carolina except where noted with two-letter state code.

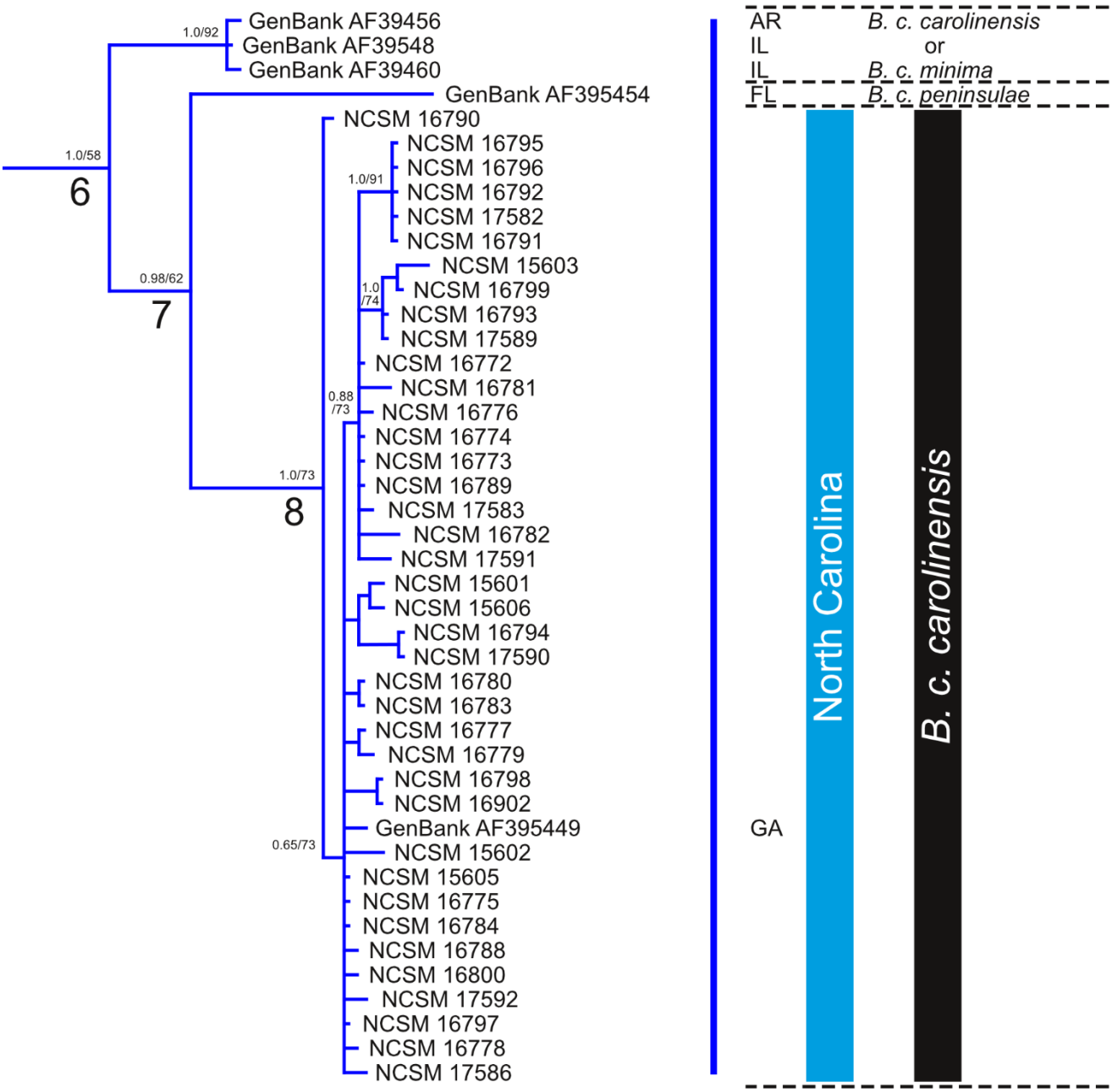


Figure 3.6: Partial fifty-percent majority-rule consensus phylogram showing the *Blarina carolinensis* clade from the Bayesian and maximum likelihood analysis. Subspecies designations noted on the right are based on Benedict et al. (2006) and McCay (2011). All samples are from North Carolina except where noted with two-letter state code.

Chapter 4

Conclusions: Short-tailed Shrews (Genus *Blarina*) in North Carolina and Future Research

Identifying and designating species of mammals have historically used the comparison of external and skull characters, but comparisons of DNA sequences have helped clarify difficult taxonomy. The short-tailed shrews in the genus *Blarina* form a group that has undergone many taxonomic changes, and the taxonomic status of *Blarina* in North Carolina have been no exception.

The goal of my research was to examine current taxonomy of the short-tailed shrews of the genus *Blarina* in North Carolina. To do this, I asked whether enough variation exist to support the current two species: the northern short-tailed shrew (*Blarina brevicauda*) and the southern short-tailed shrew (*B. carolinensis*). Both the morphological and genetic analyses showed a clear separation of the specimens identified as *B. brevicauda* and *B. carolinensis*. In the morphological analyses, all the multivariate analyses (principal components analysis – PCA, linear discriminant function analysis – LDFA, and classification tree analysis - CTA) consistently separated the two species. The genetic phylograms from Bayesian and maximum likelihood analyses showed identical topologies that showed clades for both species in North Carolina, as well as a separate clade for Elliot's short-tailed shrew (*B. hylophaga*).

The current subspecies of the northern short-tailed shrew are *Blarina brevicauda knoxjonesi* in southeastern North Carolina, and *B. b. talpoides* in the mountains and the northeast corner in the state. *B. carolinensis carolinensis* is the only subspecies recognized in North Carolina for the southern short-tailed shrew. I asked whether enough variation exist to support the two current

subspecies of the northern short-tailed shrew, and does the variation support one subspecies of the southern short-tailed shrew in North Carolina. The morpological analyses showed a slight separation with overlap for *B. brevicauda knoxjonesi* and *B. b. talpoides*, while the genetic analyses show no difference between *B. b. knoxjonesi* and *B. b. talpoides*, but only used one sample of *B. b. knoxjonesi*. Both the morphological and genetic analyses for *B. carolinensis carolinensis* showed very little variation suggesting that there is only one subspecies in North Carolina. The one sample of *B. carolinensis carolinensis* (NCSM 16800) from Burke County suggests that the southern short-tailed shrew may be more common at the western extent of the Piedmont ecoregion than evident from museum records. Southern short-tailed shrews may also be more common on the Outer Banks, so additional work should be done there.

When all the specimens were georeferenced and mapped, the two species were parapatric where the Piedmont and the Blue Ridge ecoregions abut, but showed areas of sympatry in the Middle Atlantic Coastal Plain and the Southeastern Plain ecoregions. In the northeast corner of North Carolina, overlap between *B. brevicauda* and *B. carolinensis* exist but, based on the museum specimens, *B. brevicauda* appears to be in Dare and Hyde County with the only overlap existing in western Hyde County in this peninsular region. *B. b. knoxjonesi* is found in the southeastern part of North Carolina, south of the Pamlico River (Webster, 1996; Webster et al., 2011). No specimens of *Blarina* have been collected north of the Pamlico River for the North Carolina Museum of Natural Sciences, for the University of North Carolina Wilmington mammal collections, or for any other institution who have on-line records (i.e. GBIF, IDigBio, and VertNet). Specimens are needed from north and south of the Pamlico River to determine if the river blocks gene flow, as well as from the Pamlico River to the Albemarle Sound and around the sound to the Chowan River area.

If the Pamlico River was supposed to be a barrier to gene flow (Webster, 1996), the gradual change in skull characters determined by the classification tree analysis suggests that the river is not a barrier. Specimens from northeast North Carolina show a trend of larger skull characters in the core of the local range than at the edges of the local range, such as in the Alligator River National Wildlife Refuge and near Lake Mattamuskeet in Dare and Hyde County. The same core-edge size variation exists south of the Pamlico River within the specimens identified as *B. b. knoxjonesi*.

The notion that a size-based variation within skull characters, and presumably body size, occurs on the east coast within a core-edge population dynamic is perhaps only part of the story. Webster (1996) defined a separation of the *B. brevicauda* and *B. carolinensis* based on habitat. He stated that "*B. carolinensis* occupies relatively dry well-drained uplands" while *B. brevicauda* will inhabit areas that retain more moisture where the species overlap on the east coast. Perhaps competition between species is driving the larger body size, and resource partitioning between species is responsible for some of the observable variation in the skull characters.

A sample-size of one is very problematic, and should not be weighted heavily for any generalizations. The phylograms created with the cytochrome-b gene were able to separate the species with about 30 samples of each species from North Carolina. I propose that roughly 30 samples of *B. b. knoxjonesi* throughout the range should be adequate to show if genetic variation and gene flow exists down the east coast of North Carolina for *B. brevicauda*. These genetic samples need to be taken from specimens throughout the Middle Atlantic Coastal Plains and Southeastern Plains ecoregions to represent the range of *B. brevicauda* accurately in eastern North Carolina. Targeted trapping efforts and sampling from museum specimens (Pääbo, 1989; Stuart and Fritz, 2008; Thomas et al., 1990; Wandeler et al., 2007) can help to clear up the

taxonomic uncertainty created by adding critical genetic samples within the range of *B. brevicauda* on the east coast of North Carolina. Therefore, until more genetic research is complete, my research is unable to suggest synonymizing *Blarina brevicaud knoxjonesi* with *B. b. talpoides*, and the *B. b. knoxjonesi* subspecies should remain a valid subspecies in North Carolina.

Non-North Carolina *Blarina* samples

As for the few specimens I sequenced the cytochrome-b gene from Alabama, these correspond to the description of *Blarina brevicauda cumberlandensis* located within the region of the Cumberland Plateau (Webster et al., 2011). These samples represent the southern extent of the range of this subspecies, and the first described in Alabama. It appears that there is no gene flow to the east as evidenced by the phylogenetic relationship with North Carolina samples, and that the Tennessee River is perhaps an isolating mechanism. Future work with this subspecies should address morphology and genetic analysis for potential gradual clines to the north into Tennessee, and to the east of the Tennessee River in Alabama and into Georgia.

LITERATURE CITED

Pääbo, S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences 86(6): 1939-1943.

Stuart, B. L. and U. Fritz. 2008. Historical DNA from museum type specimens clarifies diversity of Asian leaf turtles (*Cyclemys*). Biological Journal of the Linnean Society 94(1): 131-141.

Thomas, W. K., S. Pääbo, F. X. Villablanca and A. C. Wilson. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. Journal of Molecular Evolution 31(2): 101-112.

Wandeler, P., P. E. Hoeck and L. F. Keller. 2007. Back to the future: museum specimens in population genetics. Trends in Ecology & Evolution 22(12): 634-642.

Webster, W. D. 1996. Geographic variation in *Blarina brevicauda* (Insectivora: Soricidae) from eastern North Carolina, with description of a new subspecies. Pp. 47-56, in Contributions in mammalogy: a memorial volume honoring Dr. J. Knox Jones, Jr. (H. H. Genoways and R. J. Baker, eds.). Museum of Texas Tech University, Lubbock, 315 pp.

Webster, W. D., N. D. Moncrief, J. R. Choate and H. H. Genoways. 2011. Systematic revisions of the northern short-tailed shrew, *Blarina brevicauda* (Say). Virginia Museum of Natural History. Memoir 10. 77 pp.

APPENDICES

APPENDIX I – SPECIMENS EXAMINED FOR MORPHOLOGICAL ANALYSES

Specimens examined - The 1216 specimens examined in this study are listed below by museum acronym (Hafner et al., 1997). All samples are from the North Carolina State Museum of Natural Sciences (NCSM - now called the North Carolina Museum of Natural Sciences) and the University of North Carolina Wilmington (UNCW). All localities are in the United States and from North Carolina. Each sample represents a voucher specimen housed at each respective museum and is listed by county within subspecies designation (Webster et al., 2011.)

Blarina brevicauda knoxjonesi – total 169: *Beaufort Co*.: 4 mi N Aurora (UNCW 1476, 1480, 1481, 1483, 1485, 1835, 1836, 2309, 2343, 2356, 2420, 2430, 2476, 728, 2787, 2793, 2887, 2293, 1352). *Bladen Co*.: Colly Creek (UNCW 0789, 0790, 0791, 0792), White Oak (NCSM 4548, 4549, 5379). *Brunswick Co*.: 4 km NWN Bolivia (UNCW 20394, 20395, 20398, 2463), 4 mi NNW Supply (NCSM 8001). *Carteret Co*.: Harlowe (NCSM 6777, 6802), Straits (UNCW 3101). *Columbus Co*.: county only (NCSM 2575). *Craven Co*.: Croatan (NCSM 6809, 6795, 6808, 7040, 7254, 7654), county only (UNCW 8801, 8861). *Duplin Co*.: Warsaw (UNCW 4976, 3393, 3394, 3502, 3504, 5840, 5841, 5917, 11548, 11554, 11818, 11938, 11565), Rose Hill (UNCW 4975, 5021, 11125, 11553,11814), Magnolia (UNCW 5158, 5259, 5277, 5278, 11126, 12300), Wallace (UNCW 2891, 2608, 2111, 2606, 2659, 2610). *Hoke Co*.: Raeford (NCSM 2965; UNCW 784, 785). *Jones Co*.: 11.25 mi WSW Havelock (NCSM 6803, 6855), 7 mi ESE Pollocksville (NCSM 6785, 7236, 7237, 7657, 7658). *New Hanover Co*.: Carolina Beach (NCSM 4031; UNCW 0477, 0535, 0536, 0537, 0538, 0539), Castle Hayne (UNCW 4798, 5082, 5089, 12842), Wilmington (NCSM 6279, 7661, UNCW 0223, 2890, 2130, 3596, 3599, 3682, 4400, 3098, 0499, 0502, 4008, 3595). *Onslow Co*.: Holly Ridge (NCSM 16801). *Pender Co*.: Burgaw (UNCW 3389, 3390, 4994, 4995, 4996, 5165, 5230, 11555, 11564, 11994), Castle Hayne (UNCW 5087, 5191), Hamstead (UNCW 4484, 4495, 4496, 3328), Currie, Moores Creek National Battlefield (UNCW 10091, 10108, 10109, 10111, 10115, 10129), Scotts Hill (UNCW 0483, 0501, 0504, 0527, 0506, 0526, 0528, 0529, 0530, 0532, 0533, 0534, 2164), Wallace (UNCW 4978), 2 km E Watha (UNCW 5084, 5085, 5088, 5161, 11996, 11559, 11939), 3 km E Willard (UNCW 5083, 5916, 5918, 11927), county only (NCSM 17601; UNCW 3500, 3505). *Pitt Co*.: Grimesland (NCSM 4956, 4964, 4970, 4971). *Richmond Co*.: 4.1 mi NNW Hoffman (NCSM 7223). *Robeson Co*.: 5 mi E Saint Pauls (UNCW 0783). *Sampson Co*.: 5.9 mi NNW Delway (NCSM 6256), 2 km SW Newton Grove (UNCW 4401, 4402). *Scotland Co*.: 8 km SW Laurinburg (UNCW 4399)

Blarina brevicauda talpoides – total 504: *Alexander Co*.: 6 km NW Ellendale (UNCW 5096, 5192). *Alleghany Co*.: Air Bellows Gap (UNCW 3117, 3096), Beach Mountain (UNCW 3108), 2 mi SW Glade Valley (UNCW 3254), county only (UNCW 3051). *Ashe Co*.: Fleetwood (NCSM 17599, 17600), Jefferson (NCSM 14379, UNCW4162). *Avery Co*.: Collettsville (UNCW 15845), Cranberry Mine (UNCW 10582), Crossnore (UNCW 14450, 14451, 14452, 14453), Kentucky Creek (NCSM 14638, 14640), Linville (NCSM 4877, 7526, 7201, 8150, 7296, 6896, 6895, 6622, 7244, 4878, 4879, 4880, 6282, 6283, 6284, 6285; UNCW 16767, 17105, 17114, 17115, 15516, 15517, 15518, 15519, 15520, 16777, 2136, 2161, 13802), Pineola (NCSM 14384, 14385, 14387,

14405), Sugar Mountain Bog (NCSM 14386). *Buncombe Co*.: Arden (UNCW 15725), Asheville (NCSM 17605, 17606), Candler (NCSM 3372, 3376, 3377, 3394, 3395), 3 km S Dillingham (UNCW 13391, 13393, 13402, 13409, 13533, 13534, 13535, 13536, 13537, 13538, 13590), 4 km S Dillingham (UNCW 13512, 13513), 5 km S Dillingham (UNCW 13367, 13456, 13457), 6 km S Dillingham (UNCW 13582), Craggy Gardens (UNCW 1680), Jones Mountain (NCSM 13793, 13794, 14399, 14388, 14390, 14378, 14391, 14392, 14393), Swannanoa (NCSM 14400). *Burke Co*.: Linville Falls (NCSM 1504), 5 km W Ramsey (UNCW 5842, 5843). *Caldwell Co*.: Lenoir (NCSM 1949). *Camden Co*.: Dismal Swamp SNA (UNCW 3398), 2.5 mi NE Horseshoe (NCSM 1352), 6 mi W Moyock (UNCW 3303, 3304). *Chowan Co*.: 3 km NNW Edenton (UNCW 4056, 4196), 5 mi W Edenton (UNCW 3295, 3296, 3297). *Clay Co*.: Hayesville (NCSM 933; UNCW 2462), Sweetwater, Leatherwood Branch (NCSM 13771, 13772, 13773, 13774, 13770). *Currituck Co*.: Coinjock (NCSM 2723, 2699), Currituck (NCSM 917), Moyock (NCSM 3218, 467, 471; UNCW 3269, 3278, 3290). *Dare Co*.: Manns Harbor (NCSM 3818, 3916), Buffalo City (UNCW 4743, 4744, 4801, 5086), Stumpy Point (NCSM 2653; UNCW 10581, 3728, 4971, 4972), county only (NCSM 7905). *Forsyth Co*.: Winston Salem (NCSM 932, 3301). *Gates Co*.: Roduco (UNCW 3015, 3053, 3066, 3076, 3175, 10375, 3594), Storys (UNCW 3072), Sunbury (NCSM 1353; UNCW 3615, 5177, 5178, 5179, 5180, 3965, 0484, 3054, 3138, 3148, 3188, 3195, 3396). *Graham Co*.: Robbinsville (NCSM 14402, 14403; UNCW 1688), Sand Creek (NCSM 14639). *Guilford Co*.: Lake Higgin (UNCW 3610, 3611, 3612, 3613, 3614). *Haywood Co*.: Canton, 4 km S Big East Fork (UNCW 13890, 13921, 13922, 13926, 13928), Canton, Shining Rock (NCSM 3099, 3102, 3125, 3123, 3172; UNCW 1146, 1221, 15474, 15483, 15501, 15502, 15530, 15531, 15534), Canton, Black Balsam Knob (UNCW 15807, 1903), Canton, Grassy Cove (UNCW 15709), Balsum Grove (UNCW 2114, 2023, 1895, 1894), Big East Fork (UNCW 13446), Cold Mountain (NCSM 13767), Crusco (UNCW 2460), Maggie Valley (UNCW 1675, 13714, 13715, 13716, 13721), High Top (NCSM 13777), Clyde (NCSM 17610, 17611, 17612, 17613, 17614, 17615, 17616), Cecil, Bubbling Spring Branch (NCSM 14377, 13775), Sunburst (NCSM 1929), Sunburst, Big Beartrail Ridge (NCSM 13795, 13768, 13769, 13792), Dogwood Flats Creek (NCSM 14389, 13796), Black Camp Gap (UNCW 1675, 1676, 2392, 2393, 2796), Tom Branch Tributary (NCSM 15607), 3 mi S Waterville (UNCW 0496), Waynesville, Old Bald Creek (UNCW 13428, 13983, 13982, 13984, 13936, 13937), Waynesville, LTLT boulderfield (UNCW 13782, 13822, 13754, 13952), Waynesville, Steestachee Branch (UNCW 13629, 13630, 13898), county only (UNCW 2038). *Henderson Co*.: Bat Cave (NCSM 8343, 8344, 8345; UNCW 3598), Gerton (NCSM 3373), Gerton, Little Bearwallow Mtn (NCSM 13297, 13306, 17602, 17603, 17604), Humphrey's Bog (UNCW 15451, 15452, 15453, 15455, 15456, 15457, 15458, 15460, 15461, 15462, 15463, 15466, 15467, 15468), 2 km E Upward (UNCW 9542). *Hertford Co*.: 2 km NW Ahoski (UNCW 3730, 3731). *Hyde Co*.: Engelhard (UNCW 4788, 4746, 4800, 4799, 5047, 5098, 5093), Lake Mattamuskeet NWR (UNCW 5239, 5240, 1243, 2118), New Holland (2786, 5094, 5097), Swan Quarter (NCSM 439). *Jackson Co*.: Waterrock Knob (NCSM 13775; UNCW 985, 986), Cullowhee (NCSM 7246). *Jones Co*.: 7 mi ESE Pollocksville (NCSM 7655, 7656). *Macon Co*.: 4 km SSE Aquone (UNCW 13860), 6 km NW Rainbow Springs (UNCW 13826, 13827, 13828, 13829, 13830, 13831, 13845), Burningtown Creek (NCSM 1063), Highlands (NCSM 5607, 5610, 2043, 1174, 1173, 1168, 1172, 5606; UNCW 9815). *Madison Co*.: 3 mi W Hot Springs (UNCW 3600, 3601), 2 km S Marshall (UNCW 9543), Max Patch Mtn (NCSM15608, 17608, 17609), Whiterock Cliffs (NCSM 14380). *McDowell Co*.: Old Fort (UNCW 13819, 13820, 13818, 4973, 4974, 4977), 10 mi NW Marion (NCSM 3963). *Mitchell Co*.: Roan High Bluff (UNCW 14470,

14476, 14508, 17570, 17571, 17602, 17603), Bald Mtn (NCSM 2594), Balsam Road Switchback (UNCW 17564, 17565, 17655), Roan Mtn (NCSM 846, 5618, 5605, 5609, 5608, 7804; UNCW 390, 391, 392, 497, 498, 531). *Pasquotank Co*.: Elizabeth City (NCSM 7879), 9 km WNW Morgan's Corner (UNCW 3357, 3364). *Perquiman Co*.: 5 km SSW Hertford (UNCW 4195, 4525, 4526), 1.3 mi S Hickory Crossroads (UNCW 3298). *Polk Co*.: 1 km NE Saluda (UNCW 4490, 4492, 4497), 10 km NE Saluda (UNCW 4486), 3 km ENE Saluda (UNCW 4487), 6 km NE Saluda (UNCW 4489, 4483, 4493, 4494), 8 km NE Saluda (UNCW 4480, 4481, 4482, 4485, 4488, 4498, 4499, 4491), Saluda, Green River Gamelands (NCSM 7747, 14406-14414, 14641, 14642, 14938, 14939, 15609, 15610, 17596, 17597, 17598). *Rutherford Co*.: Bat Cave (NCSM 357, 358, 359), Chimney Rock (NCSM 6275, 6276), Lake Lure (NCSM 356, 13305, 13307, 8133, 8346, 8347), Mill Spring (NCSM 17607), Rumbling Bald Mtn (NCSM 8180, 8194, 8195). *Stokes Co*.: King (UNCW 1988, 2035). *Surry Co*.: 4.5 mi E Dobson (NCSM 3464), 4.75 mi NNW Mountain Park (NCSM 7633), 3 mi S Mt Airy (NCSM 3462). *Swain Co*.: Clingmans Dome (UNCW 2360, 2365, 2367, 3597), Old Indian Gap Road (UNCW 1825), Straight Fork Creek (UNCW 2408), Smokemont (UNCW 981). *Transylvania Co*.: Wagon Road Gap (UNCW 1933, 2018), Balsam Grove (UNCW 1153), Blantyre (NCSM 1924), 9 mi SW Brevard (UNCW 2024), Sapphire (NCSM 4784, 8792), Highway 280 (NCSM 1888, 1887, 1891, 2473), 7 mi S Blue Ridge Parkway (UNCW 2150), Thompson Ridge (UNCW 1824), county only (NCSM 8794, 8937). *Watauga Co*.: Beech Creek Bog (NCSM 17594, 13765), Blowing Rock (NCSM 1922, 1923), Boone (NCSM 10359, 13784, 13785), Brookhollow, S fork New River (NCSM 2002, 2003, 2004), Elk Knob (NCSM 13764), Foscoe (NCSM 13762, 13763, 13783, 13786, 13787, 13788, 13789, 13790), Hanging Rock Mtn (NCSM 6208), Banner Elk, Bear Paw SNA (NCSM 17595), Valle Crucis (NCSM 3976, 3977). *Wilkes Co*.: Air Bellows Gap (UNCW 3052, 3216, 3102, 3602), 4.25 mi ESE Austin (NCSM 14363), Pores Knob (UNCW 5090, 5091, 5092, 5095), 4.75 mi WNW Traphill (NCSM 8220). *Yancey Co*.: Burnsville, Mt Craig (UNCW 15544), Burnsville, Mt Mitchell (UNCW 2888), Burnsville, Toe River and Hwy 80 (UNCW 393, 2196-2204), Balsam Gap (UNCW 3344), Busick (UNCW 2041), 4 mi SSW Busick (UNCW 3622), 4 mi SW Busick (NCSM 17), 4.8 mi SW Busick (NCSM 397), 5 mi SSW Busick (UNCW 3359), 5 mi W Busick (UNCW 1641), Busick, Black Mountain Campground (UNCW 975, 1375, 1992, 1993, 1994, 1995, 1996, 2892, 2605, 2632, 2633, 1119), Carolina Hemlocks Campground (UNCW 816, 817, 847), Mt Mitchell (NCSM 15, 16; UNCW 1890, 1923, 1948, 2888, 2653), Mt Mitchell SP (NCSM 4986, 7770), Balsam Cone (NCSM 14404), Deep Gap Trail (NCSM 14401).

Blarina carolinensis carolinensis – total 543: *Alamance Co*.: county only (NCSM 7880). *Beaufort Co*.: 1.5 mi S Chocowinity (NCSM 3422), Pantego (NCSM 2422). *Bladen Co*.: Clarkton (UNCW 2631, 2884, 2885), 2.5 mi SW Elizabethtown (NCSM 4832), 3.7 mi NW Elizabethtown (NCSM 8336), Elizabethtown (NCSM 7088), Harrells (NCSM 4687). *Brunswick Co*.: Leland (UNCW 14040), Winnabow, Bell Swamp (UNCW 10986), Rabontown (UNCW 14066, 14067, 14071, 14073, 14074, 14075, 14076, 13706), 0.5 km N Rabontown (UNCW 11625, 11626, 11627, 13312, 13315, 13318, 13687, 13688, 13695, 13696), 1 km N Rabontown (UNCW 11628, 11629, 14031, 14035, 14036, 14037, 14041, 14037), 1 km NW Rabontown (UNCW 14027, 13313, 13315, 13316, 13317, 13319, 13679, 13680, 13681, 13682, 13694), 2 km NW Rabontown (UNCW 14040, 13314, 13683, 13684, 13685, 13686, 14030, 14033), 5.5 mi NNE Southport (NCSM 6280), Southport, MOTSU (NCSM 5209), 5 mi N Supply (NCSM

5338), 8 km NNE Supply (UNCW 20611), 1 mi W Wilmington (UNCW 3514), 5.3 mi W Wilmington (UNCW 494, 500, 507), 3 mi NW Winnabow (UNCW 3966, 3967, 3513). *Burke Co*.: 4.5 km WNW Morganton (NCSM 16800). *Cabarrus Co*.: 1 mi W Kannapolis (NCSM 1064). *Camden Co*.: Shawboro (NCSM 3973). *Caswell Co*.: 4 mi W Yanceyville (NCSM 4479, 4480). *Chatham Co*.: 6.5 km NE Pittsboro (NCSM 13755). *Chowan Co*.: Edenton (NCSM 8137, 8397), 2 mi E Edenton (NCSM 4792, 4793), 4 mi SE Edenton (NCSM 5328), 4.5 mi SE Edenton (NCSM 8337, 5423, 6713, 6609, 6612, 6613, 5348, 6365, 6366, 6367, 6746, 6714, 6816, 6817, 6818, 6819, 6843, 6844, 6845, 6854, 6861, 6862, 7066, 7067, 7304, 7399, 6277, 6712, 7971, 8338, 8339), 5.2 mi SE Edenton (NCSM 7221), 5.5 mi SE Edenton (NCSM 6278), 8.5 mi E Edenton (NCSM 5339). *Columbus Co*.: 2 km NW Bolton (UNCW 4021, 4022, 4023, 4066, 4067, 4068), 3 km NNE Freeman (UNCW 4069). *Craven Co*.: 2 km NNW Rhems (UNCW 5121, 5153, 5169, 5170). *Cumberland Co*.: 7 km W Fayetteville (NCSM 8341, 8342), Fort Bragg (NCSM 8512, 8791). *Currituck Co*.: 9 km ESE Moyock (UNCW 3293, 3294). *Dare Co*.: Bodie Island (UNCW 2017), 1 mi N Wanchese (UNCW 46). *Davie Co*.: 5 km S Davie Crossroads on SR 1882 (UNCW 12872). *Durham Co*.: Bahama (NCSM 17591), 3.5 mi WNW Durham (NCSM 7512, 6700, 6701, 6702, 6703, 6708), Durham (NCSM 7908). *Edgecombe Co*.: 4 mi ESE Battleboro (NCSM 7607-7612, 7625, 8076, 8086, 8340). *Franklin Co*.: 2 km W Youngsville (UNCW 4778, 4779). *Gaston Co*.: Belmont (NCSM 17592), Gastonia (UNCW 2304). *Granville Co*.: 9.2 km NW Youngsville (NCSM 13760). *Guilford Co*.: Greensboro (UNCW 3624, 3964, 4127), 1 mi E Jamestown (UNCW 3340), 2 km W Jamestown (UNCW 4513). *Halifax Co*.: 7 mi W Roanoke Rapids (NCSM 469). *Harnett Co*.: Fuquay Varina (NCSM 2560), 3.5 mi SSW Spout Springs (NCSM 7228), 6 km E Spout Springs (UNCW 4232, 4327). *Hertford Co*.: Ahoskie (NCSM 2431), 2 mi N Ahoskie (NCSM 4637). *Hoke Co*.: McCain (NCSM 3827, 7049, 7511, 7299, 4515, 4155). *Hyde Co*.: 4 km N Scranton (UNCW 4982, 4983, 4985, 4983, 4982, 4780, 4781, 4810, 4811, 4812, 4813), 6 km N Scranton (UNCW 4986, 4986, 4774, 4775, 4776, 4777). *Iredell Co*.: 1 km N Buffalo (UNCW 4984, 5190), 7 mi W Mooresville (NCSM 4805), county only (NCSM 4819, 4820). *Johnson Co*.: 3 km N Benson (UNCW 4165), 10 km WSW Clayton (UNCW 4257, 4315, 4316, 4323, 4325, 4330, 4352), Four Oaks (NCSM 1057), Smithfield (UNCW 540), 5.5 mi SE Maysville (NCSM 6810, 7199), 6 mi SE Maysville (NCSM 6794, 6807), county only (NCSM 8395, 8793, 8802). *Lee Co*.: county only (NCSM 2424). *Mecklenburg Co*.: county only (UNCW 16802, 16809, 16814, 16815, 16825, 16826, 16828, 16829, 16833, 16834), Charlotte (NCSM 926, 927; UNCW 3625), 2.8 km NNE Matthews (UNCW 2650, 2886). *Montgomery Co*.: 2 km E Biscoe (UNCW 4235), Mount Gilead (UNCW 2158, 2159, 641, 2098). *Moore Co*.: 0.9 mi SSE Jackson Springs (NCSM 14365), 1.5 mi NNW Pinebluff (NCSM 17638), 7 km SW Robbins (UNCW 4233, 4234, 4239, 4324, 4328, 4329, 4347), Whispering Pines (NCSM 8795). *Nash Co*.: 3 km WNW Bailey (UNCW 4168, 4169, 4236), 3 km ENE Middlesex (UNCW 4318, 4242, 4317, 4348, 4238, 4240, 4322), county only (UNCW 4241). *New Hanover Co*.: Carolina Beach (NCSM 3816), Carolina Beach SP (NCSM 3056), 1 mi NW Wilmington (UNCW 1174), Eagle Island in Wilmington (UNCW 2010). *Northampton Co*.: Seaboard off Hwy 186 (UNCW 3120), county only (NCSM 8504). *Onslow Co*.: Camp Lejeune (NCSM 7990), Richlands in Hofmann State Forest (NCSM 17593), 0.75 mi SE Piney Green (NCSM 6904, 6905), Sneads Ferry (UNCW 5009), 10 km E Sneads Ferry (UNCW 4128-4134, 4514, 4515), 4.2 km SSW Sneads Ferry (UNCW 4783), 8 km E Sneads Ferry (UNCW 4507, 4508, 4510, 4511, 4512), Swansboro at Hammocks Beach SP (UNCW 2302, 2303, 2295). *Orange Co*.: Chapel Hill (NCSM 2419, 8864), Durham at Eno River SP (NCSM 7542, 7543). *Pitt Co*.: 3 km WSW Dupree Crossroads (UNCW 4326, 4350, 4782),

Falkland (NCSM 1166), Greenville at Tar River Swamp (NCSM 458). *Randolph Co*.: 1.5 mi SSE (NCSM 7130), 5 mi E Asheboro (UNCW 338), 4.24 km SSE Asheboro (UNCW 3519), Liberty (NCSM 1065), 3.8 mi ESE Ulah (NCSM 6614, 7347). *Richmond Co*.: 4 mi N Ellerbe (NCSM 3445), 6 km NNE Rockingham (UNCW 4321), 7.2 km SSE Ellerbe (NCSM 3059), 7.2 mi N Ellerbe (NCSM 3443). *Rockingham Co*.: 3 mi E Williamsburg (NCSM 3438). *Rowan Co*.: 0.5 mi SW Barber (NCSM 6497), 4 km WNW Needmore (UNCW 10574). *Scotland Co*.: 4 mi E Marston (NCSM 16786-16788), 11 mi NW Wagram (NCSM 14421). *Stanly Co*.: 2 mi SE Badin (UNCW 495). *Tyrrell Co*.: Columbia at Scuppernong River SNA (NCSM 16772-16777), 4 km SE Creswell (UNCW 5051, 5122-5126, 5171), Creswell at Scuppernong River SNA (NCSM 16778-16784). *Union Co*.: 2.25 mi NE Monroe (NCSM 7892), county only (NCSM 2981). *Vance Co*.: 2.25 mi NW Kittrell (NCSM 4833). *Wake Co*.: Apex (NCSM 2433), 4.5 mi ENE Auburn (NCSM 6204), Bayleaf (NCSM 13756), Boneyard Lake (NCSM 2517), Cary (NCSM 806), 1.75 mi ESE Cary (NCSM 6744), 5.3 km SSW Cary (NCSM 14415-14420, 15604), 3.4 km SSW Cary (NCSM 16798, 16799), 3.9 mi ESE Cary (NCSM 17648), 2 km SW Cary (NCSM 8396), 4 mi S Cary (NCSM 1359), 3.74 km S Cary (NCSM 5195), 3.5 mi WNW McCullers (NCSM 6417, 6418), 7.75 mi NW Millbrook (NCSM 7224), 5 mi WNW Neuse (NCSM 7411, 7452), Raleigh (NCSM 18-22, 24-31, 819, 820, 928, 616, 739, 741, 744, 877, 914-916, 918-924, 931, 934, 999, 1916, 1917, 1926-1928, 2416-2418, 2420, 2421, 2425, 2427, 2429, 2430, 2432, 2962, 2963, 3209, 3229, 3231, 4690, 6205-6207, 6286, 6600, 7063, 7870, 8021, 8114, 9146, 9147, 16902, 17581; UNCW M48WC, 3709, 4167), 8.66 km NNE Raleigh (NCSM 8398), 17.8 km NNE Raleigh (NCSM 16797), 3.8 km NW Raleigh (NCSM 1119, 1223, 1229), 4.7 km NW Raleigh (NCSM 15603), 8.8 km NW Raleigh (NCSM 2428, 2436), 3.6 NW Raleigh (NCSM 7599), 3 mi W Raleigh (NCSM 6509), 3.8 km WNW Raleigh (NCSM 7529), 4 mi W Raleigh (NCSM 930), 5 mi E Raleigh (NCSM 23), 12.2 km N Raleigh (NCSMNCSM 16791-16796, 17589, 17590), 7.4 km E Raleigh (NCSM 15601), 4.2 mi N Raleigh (NCSM 7243), 9.5 mi N Raleigh (NCSM 6742), 4.3 km SW Raleigh (NCSM 2434), Raleigh at Crabtree Valley (NCSM 3261), 7 km WSW Raleigh (NCSM 17588), 4.2 mi WSW Raleigh (NCSM 848, 8020), 4.8 km NNW Raleigh (UNCW 478), 5.1 km WNW Raleigh (NCSM 2435, 2441, 2576, 2426), 7.9 km NW Raleigh (NCSM 15602, 15177, 17582-17587), Raleigh at NCSU (NCSM 929, 2423, 5996-5998), Raleigh at NCSU Centennial Campus (NCSM 5951-5953, 5977-5995, 5999-6001), 8.9 km NW Raleigh (NCSM 15606), 7.6 km N Raleigh (NCSM 8141, 8152), Raleigh at Schenk Forest (NCSM 2437), west Raleigh (NCSM 935), Raleigh at W. B. Umstead SP (NCMS 3199, 16789, 16790), Stony Hill (NCSM 1925), 7.9 km W Raleigh (NCSM 12910), 3.25 mi NW Willow Spring (NCSM 8991, 8994, 13761), Zebulon (UNCW 3681), county only (NCSM 831, 15178). *Washington Co*.: Creswell (UNCW 4135), Creswell at Scuppernong River SNA (NCSM 16785), Creswell at Pettigrew SP (NCSM 6891, 7308, 7370, 7600-7606, 15605), 1 km NW Scuppernong (UNCW 4170). *Wayne Co*.: 7 km W Kenly (UNCW 4237), 8 mi E Mount Olive (UNCW 781, 782). *Wilson Co*.: 3 km SE Elm City (UNCW 4320, 4349), 16 km W Wilson (UNCW 4313, 4314, 4319, 4351). North Carolina only (NCSM 16903).

LITERATURE CITED

Hafner, M. S., W. L. Gannon, J. Salazar-Bravo and S. T. Alvarez-Castañeda. 1997. Mammal collections in the Western Hemisphere. American Society of Mammalogists. Allen Press, Lawrence, Kansas. 93 pp.

Webster, W. D., N. D. Moncrief, J. R. Choate and H. H. Genoways. 2011. Systematic revisions of the northern short-tailed shrew, *Blarina brevicauda* (Say). Virginia Museum of Natural History. Memoir 10. 77 pp.

APPENDIX II – SPECIMENS SPLIT WITH CLASSIFICATION TREE ANALYSIS WITH GINI SPLITTING INDEX

Specimens examined - The 448 specimens analyzed with the skull data and the 461 specimens analyzed for the mandible data. All samples from this classification tree analysis are from the University of North Carolina Wilmington (UNCW). All localities are in the United States and from North Carolina. Each sample represents a voucher specimen housed at each respective museum and is listed by county within subspecies designation.

SKULL – Blcaca = *Blarina carolinensis carolinensis*, Blbrkn = *Blarina brevicauda knoxjonesi*, Blbrta = *Blarina brevicauda talpoides*

1st split is OPML < 19.73 → (NODE 2) Splits all Blcaca; OPML ≥ 19.73 is all Blbr

30 smallest (small to large) Blcaca by OPML: 1-10 (UNCW 3966, Brunswick Co., Winnabow, SubAd female **SE corner Coastal**; UNCW 4134, Onslow Co., Sneads Ferry, OldAd male **SE Coastal ocean edge**; UNCW 4233, Moore Co., Robbins, OldAd male **S Central Southeastern Plains near Piedmont edge**; UNCW 4320, Wilson Co., Elm City, SubAd no sex indicated **central NC**; UNCW 14037, Brunswick Co., Rabontown, SubAd no sex indicated **SE corner Coastal**; UNCW 5190, Iredell Co., Statesville, SubAd no sex indicated **W Piedmont**; UNCW 4782, Pitt Co., Dupree Crossroads, Ad no sex indicated **SE plain near coastal plain**; UNCW 4514, Onslow Co., Sneads Ferry, SubAd no sex inicated **SE Coastal ocean edge**; UNCW 5009, Onslow Co., Sneads Ferry, Ad female **SE Coastal ocean edge**; UNCW 4326, Pitt Co., Dupree Crossroads, Ad no sex indicated ** SE plain near coastal plain **) 11-20 (UNCW 14033, Brunswick Co., Rabontown, SubAd no sex indicated **SE corner Coastal**; UNCW 5124, Tyrrell Co., Creswell, SubAd no sex indicated **near Alligator River area edge of coastal BLBR**; UNCW 5170, Craven Co., Rhems, Ad male **near middle of Coastal Plain**; UNCW 781, Wayne Co., Mount Olive, Ad no sex indicated **central NC**; UNCW 4325, Johnston Co., Clayton, SubAd no sex indicated **central NC**; UNCW 2158, Montgomery Co., Mount Gilead, Ad no sex indicated **SW NC – S central Piedmont**; UNCW 4232, Harnett Co., Spout Springs, SubAd no sex indicated **central NC**; UNCW 2098, Montgomery Co., Mount Gilead, Ad no sex indicated **SW NC – S central Piedmont**; UNCW 4775, Harnett Co., Spout Springs, OldAd no sex indicated **central NC**; UNCW 4781, Hyde Co., Scranton, Ad no sex indicated **near Alligator River area edge of coastal BLBR**) 21-30 (UNCW 5121, Craven Co., Rhems, SubAd no sex indicated **near middle of Coastal Plain**; UNCW 500, Brunswick Co., Wilmington, SubAd no sex indicated **SE corner Coastal**; UNCW 4813, Hyde Co., Scranton, Ad no sex indicated **near Alligator River area edge of coastal BLBR**; UNCW 2159, Montgomery Co., Mount Gilead, Ad no sex indicated **SW NC – S central Piedmont**; UNCW 14041, Brunswick Co., Rabontown, Ad no sex indicated **SE corner Coastal**; UNCW 4127, Guilford Co., Greensboro, Ad no sex indicated **NW corner of distribution – central piedmont adjacent to E extent of Blue Ridge BLBR**; UNCW 14067, Brunswick Co., Rabontown, SubAd female **SE corner Coastal**; UNCW 11627, Brunswick Co., Rabontown, Ad-OldAd male **SE corner Coastal**; UNCW 3709, Wake Co., Raleigh, Ad male **central NC**; UNCW 495, Stanly Co., Badin, Ad-OldAd male **SW NC – S central Piedmont**)

*** 16/30 Coastal Plain; 14/30 is near center of state distribution; 2/30 [6.7%] female, 6/30 [20.0%] male, 22/30 [73.3%] no sex indicated; 12/30 [40.0%] sub adult, 13/30 [43.3%] Adult, 5/30 [16.7%] old adult

30 largest (small to large) Blcaca by OPML: 1-10 largest (small to large) (UNCW 3120, Northampton Co., Seaboard, SubAd-Ad male **NE NC – W of NE Coastal BLBR**; UNCW 14075, Brunswick Co., Rabontown, Ad no sex indicated **SE corner Coastal**; UNCW 4165, Johnston Co., Benson, Ad-OldAd no sex indicated **central NC**; UNCW 4135, Washington Co., Creswell SubAd male **near Alligator River area edge of coastal BLBR**; UNCW 5123, Tyrrell Co., Creswell, SubAd no sex indicated **near Alligator River area edge of coastal BLBR**; UNCW 5171, Tyrrell Co., Creswell, Ad male **near Alligator River area edge of coastal BLBR**; UNCW 1174, New Hanover Co., Wilmington, Ad female **SE corner Coastal**; UNCW 11629, Brunswick Co., Rabontown, Ad male **SE corner Coastal**; UNCW 2295, Onslow Co., Swansboro, Ad female **SE Coastal ocean edge**; UNCW 3514, Brunswick Co., Wilmington, Ad female lactating **SE corner Coastal**) 12-20 largest (small to large) (UNCW 11625, Brunswick Co., Rabontown, Ad male **SE corner Coastal**; UNCW 10574, Rowan Co., Needmore, Ad male **W Piedmont**; UNCW 3513, Brunswick Co., Winnabow, Ad female lactating **SE corner Coastal**; UNCW 3293, Currituck Co., Tulls Creek, SubAd-Ad no sex indicated **NE corner Coastal**; UNCW 4023, Columbus Co., Bolton, Ad male **SE corner Coastal**; UNCW 4067, Columbus Co., Bolton, SubAd female **SE corner Coastal**; UNCW 4021, Columbus Co., Bolton, SubAd female **SE corner Coastal**; UNCW 2885, Bladen Co., Clarkton, OldAd male **SE Southeastern Plains**; UNCW 4132, Onslow Co., Sneads Ferry, SubAd female **SE Coastal ocean edge**; UNCW 4352, Johnston Co., Benson, SubAd no sex indicated **central NC**) 21-30 largest (small to large) (UNCW 4239, Moore Co., Robbins, Ad male **S Central Southeastern Plains near Piedmont edge**; UNCW 2304, Gaston Co., Gastonia, Ad female **S central to SW Piedmont**; UNCW 4982, Hyde Co., Scranton, Ad female **near Alligator River area edge of coastal BLBR**; UNCW 4066, Columbus Co., Bolton, SubAd male **SE corner Coastal**; UNCW 3964, Guilford Co., Greensboro, Ad no sex indicated **NW corner of distribution – central piedmont adjacent to E extent of Blue Ridge BLBR**; UNCW 4316, Johnston Co., Clayton, SubAd no sex indicated **central NC**; UNCW 4780, Hyde Co., Scranton, SubAd female **near Alligator River area edge of coastal BLBR**; UNCW 4512, Onslow Co., Sneads Ferry, Ad male **SE Coastal ocean edge**; UNCW 4511, Onslow Co., Sneads Ferry, Ad male **SE Coastal ocean edge**; UNCW 4168, Nash Co., Bailey, OldAd male **central NC**)

*** 20/30 Coastal Plain; 10/30 is near center of state distribution; 10/30 [33.3%] female, 13/30 [43.3%] male, 7/30 [23.3%] no sex indicated; 11/30 [36.7%] sub adult, 16/30 [53.3%] Adult, 3/30 [10.0%] old adult

2nd split is CRB < 11.8 → (NODE 5) 112 Blbrta CRB > 11.8

1-10 smallest Blbrta (UNCW 3731, Hertford Co., Ahoski, Ad female **NE Coastal**; UNCW 4487, Polk Co., Saluda, SubAd female; UNCW 5092, Wilkes Co., Pores Knob, SubAd male **NE Blue Ridge**; UNCW 1680, Buncombe Co., Swannanoa, SubAd female; UNCW 1824, Transylvania Co., Thompson Ridge, Ad male; UNCW 4974, McDowell Co., Old Fort, Ad male;

UNCW 2786, Hyde Co., New Holland, SubAd male **NE Coastal**; UNCW 3066, Gates Co., Roduco, SubAd male **NE Coastal**; UNCW 3072, Gates Co., Storys, SubAd female **NE Coastal**; UNCW 9542, Henderson Co., Upward, Ad male) 11-20 smallest Blbrta: (UNCW 4490, Polk Co., Saluda, SubAd male; UNCW 1146, Haywood Co., Shining Rock, Ad female; UNCW 10581, Dare Co., Stumpy Point, SubAd female **edge of Alligator River area**; UNCW 13457, Buncombe Co., Dillingham, Ad female; UNCW 3622, Yancey Co., Busick, SubAd female; UNCW 4483, Polk Co., Saluda, Ad male; UNCW 3278, Currituck Co., Moyock, Ad no sex indicated **edge of NE Coastal**; UNCW 3965, Gates Co., Sunbury, Ad female **NE Coastal**; UNCW 3076, Gates Co., Roduco, SubAd female **NE Coastal**; UNCW 3188, Gates Co., Sunbury, Ad female **NE Coastal**) 21-30 Blbrta: (UNCW 4481, Polk Co., Saluda, SubAd female; UNCW 5095, Wilkes Co., Pores Knob, SubAd male **NE Blue Ridge – E point on northern BR peninsula**; UNCW 392, Mitchell Co., Roan Mountain, Ad female; UNCW 4480, Polk Co., Saluda, Ad male; UNCW 2136, Avery Co, Linville, Ad female; UNCW 3364, Pasquotank Co., Morgans Corner, Ad male ** NE Coastal**; UNCW 4486, Polk Co., Saluda, Ad male; UNCW 1988, Stokes Co., King, Ad male **NW Piedmont**; UNCW 4488, Polk Co., Saluda, Ad female; UNCW 1994, Yancey Co., Busick, Ad-OldAd no sex indicated)

*** 10/30 [33.3%] is NE Coastal Plain; 10/30 is edge of Blue Ridge, 10/30 is Blue Ridge; 15/30 [50%] female, 13/30 [46.7%] male; 12/30 [40.0%] sub adult, 18/30 [60.0%] Adult, 0/30 old adult

1-10 Largest (small to large) Blbrta: (UNCW 1221 Haywood Co., Shining Rock, OldAd male; UNCW 13629, Haywood Co., Waynesville, Ad male; UNCW 2633, Yancey Co., OldAd male; UNCW 13714, Haywood Co., Maggie Valley, SubAd female; UNCW 2393, Haywood Co., Soco Gap, SubAd female; UNCW 1890, Yancey Co., Busick, OldAd female; UNCW 0497, Mitchell Co., Roan Mtn, SubAd male; UNCW 498, Mitchell Co., Roan Mtn, OldAd male; UNCW 2114, Haywood Co., Balsum Grove, OldAd male; UNCW 847, Yancey Co., Carolina Hemlocks Cmpgd, Ad female) 11-20 largest (small to large) Blbrta: (UNCW 13582, Bucombe Co., Dillingham, Ad no sex indicated; UNCW 390, Mitchell Co., Roan Mtn, Ad male; UNCW 4492, Polk Co., Saluda, OldAd male **SE Blue Ridge**; UNCW 4497, Polk Co., Saluda, Ad female **SE Blue Ridge**; UNCW 2041, Yancey Co., Busick, Ad male; UNCW 2392, Haywood Co., Soco Gap, SubAd female; UNCW 13456, Buncombe Co., Dillingham, Ad male; UNCW 2408, Swain Co., GSMNP, Ad female; UNCW 13630, Haywood Co., Waynesville, Ad male; UNCW 13715, Haywood Co., Maggie Valley, OldAd female) 21-30 largest (small to large) Blbrta: (UNCW 13716, Haywood Co., Maggie Valley, SubAd male; UNCW 391, Mitchell Co., Roan Mtn, SubAd male; UNCW 3597, Swain Co., Clingmans Dome, SubAd female; UNCW 13402, Buncombe Co., Dillingham, Ad male; UNCW 3344, Yancey Co., Busick, Ad female; UNCW 3601, Madison Co., Hot Springs, Ad female; UNCW 3096, Alleghany Co., Aire Bellows Gap, SubAd female; UNCW 975, Yancey Co., Busick, Ad female; UNCW 985, Jackson Co., Watterock Knob, Ad female; UNCW 13367, Buncombe Co., Dillingham, Ad male)

***28/30 [93.3%] is high elevations Blue Mtn region; 14/30 [46.7%] female, 15/30 [50%] male; 8/30 [26.7%] sub adult, 15/30 [50.0%] adult, 7/30 [23.3%] old adult

8 Blbrkn CRB > 11.8 (UNCW 0483, Pender Co., Scotts Hill, Ad female; UNCW 4401, Sampson Co., Newton Grove, SubAd-Ad no sex indicated; UNCW 3682, New Hanover Co., Wilmington, Ad-OldAd female; UNCW 2610, Duplin Co., Wallace, Ad-OldAd male; UNCW 3595, New

Hanover Co., Wilmington, Ad female; UNCW 0792, Bladen Co., Ivanhoe, SubAd-Ad no sex indicated; UNCW 11938 Duplin Co, 5 km S Warsaw *almost center of distribution*, Ad no sex indicated; UNCW 4975, Duplin Co., Rose Hill, Ad-OldAd male

*** 6/8 [75.0%] are near ~center of Southeastern and Coastal Plain populations, all from Middle Atlantic Coastal Plain, north of Cape Fear River, 2/8 male 3/8 female 3/8 no sex indicated; 2/8 [25.0%] sub adult, 3/8 [37.5%] adult, 3/8 [37.5%] old adult

3rd split MAB < 7.165 → (NODE 6)

69 Blbrkn (69/125 [55.2%] of total *knoxjonesi* specimens)
10 largest (small to large, 60-69) Blbrkn (UNCW 5278,Duplin Co., Magnolia, SubAd male *almost center of distribution*; UNCW 0528, Pender Co., Scotts Hill, SubAd male *coastal center of distribution*; UNCW 11996, Pender Co., Watha, Ad no sex indicated *center of distribution*; UNCW 4400, New Hanover Co., Wilmington, Ad female *SE corner of distribution* ; UNCW 0223, New Hanover Co., Wilmington, Ad female *SE corner of distribution*; UNCW 2890, New Hanover Co., Wilmington, SubAd female *SE corner of distribution*; UNCW 1476, Beaufort Co., Aurora, Ad no sex indicated *NE corner of distribution*; UNCW 5840, Duplin Co., Warsaw, SubAd female *center of distribution*; UNCW 1483, Beaufort Co., Aurora, SubAd male *NE corner of distribution*; UNCW 5021, Duplin Co., Rose Hill, Ad male *center of distribution*) next 10 largest (small to large, 50-59) Blbrkn (UNCW 11994, Pender Co., Burgaw, Ad no sex indicate *center of distribution*; UNCW 0530, Pender Co., Scotts Hill, SubAd female *center of distribution*; UNCW 11555, Pender Co., Burgaw, SubAd no sex indicated *center of distribution*; UNCW 0527, Pender Co., Scotts Hill, SubAd no sex indicated *center of distribution*; UNCW 0501, Pender Co., Scotts Hill, SubAd female *center of distribution*; UNCW 1836, Beaufort Co., Aurora, SubAd male *NE corner of distribution*; UNCW 3394, Duplin Co., Warsaw, SubAd no sex indicated *center of distribution*; UNCW 3505, Pender Co., no locality, Ad female *center of distribution*; UNCW 2343, Beaufort Co., Aurora, SubAd female *NE corner of distribution*; UNCW 10129, Pender Co., Moores Creek Nat'l Battlefield, SubAd female *NE corner of distribution*) Next 10 largest (small to large, 40-49) (UNCW 5918, Pender Co., Willard, SubAd female *center of distribution, coastal plain close to southeastern plain*; UNCW 11125, Duplin Co., Rose Hill, Ad male *center of distribution*; UNCW 3328, Pender Co., Hamstead Holly Shelter Game Land, SubAd male *coastal center of distribution*; UNCW 3599, New Hanover Co., Wilmington, SubAd female *SE corner of distribution*; UNCW 4994, Pender Co., Burgaw, SubAd female *center of distribution*; UNCW 0536, New Hanover Co., Carolina Beach, SubAd no sex indicated *SE corner of distribution*; UNCW 4798, New Hanover Co., Castle Hayne, Ad female *SE corner of distribution*; UNCW 0506, Pender Co., Scotts Hill, SubAd female*center of distribution*; UNCW 0537, New Hanover Co., Carolina Beach, Ad no sex indicated *SE corner of distribution*; UNCW 3389, Pender Co., Burgaw, SubAd male *center of distribution*)

***18/30 [60.0%] is approximate center of distribution, 12/30 [40.0%] is edges of distribution; 14/30 [46.7%] female, 8/30 [26.7%] male; 20/30 [66.7%] sub adult, 10/30 [33.3%] adult, 0/30 [0.00%] old adult

11 Blbrta (UNCW 5086, Dare Co., Buffalo City, SubAd no sex indicated *Alligator NWR region*; UNCW 1993, Yancey Co., Busick, OldAd male *Blue Ridge*; UNCW 5097, Hyde Co., New Holland, SubAd female *Lake Mattamuskeet / Alligator NWR region*; UNCW 5094, Hyde Co., New Holland, SubAd no sex indicated *Lake Mattamuskeet / Alligator NWR region*; UNCW 3728, Dare Co., Stumpy Point, SubAd female *Alligator NWR region*; UNCW 5098, Hyde Co., Engelhard, SubAd male *Lake Mattamuskeet / Alligator NWR region*; UNCW 4196, Chowan Co., Edenton, Ad male *NE corner of state*; UNCW 5093, Hyde Co., Engelhard, SubAd no sex indicated *Lake Mattamuskeet / Alligator NWR region*; UNCW 3015, Gates Co., Roduco, Ad female *NE corner of NC*; UNCW 4526, Perquimans Co., Hertford, Ad no sex indicated *NE corner of NC*; UNCW 3594, Gates Co., Roduco, Ad female *NE corner of NC*)

*** 10/11 [90.9%] are in NE Middle Atlantic Coastal Plain, north of Pamlico River, with 6/10 [60.0%] of those in Lake Mattamuskeet/Alligator River NWR region; 4/11 [36.4%] female, 4/11 [36.4%] male, 3/11 [27.3%] no sex indicated; 6/11 [54.5%] sub adult, 4/11 [36.4%] adult, 1/11 [9.1%] old adult

4th split OPML >21.98 → (NODE 9)

4 Blbrkn (UNCW 0532, Pender Co., Scotts Hill, Ad no sex indicated *center of distribution*; UNCW 11554, Duplin Co., Warsaw, SubAd male *center of distribution*; UNCW 4399, Scotland Co., Laurinburg, SubAd no sex indicated *SW edge of distribution*; UNCW 4496, Pender Co., Wallace, SubAd no sex indicated *center of distribution*)

15 Blbrta (UNCW 1676, Haywood Co., Soco Gap, Ad male *Blue Ridge*; UNCW 1995, Yancey Co., Busick, SubAd no sex indicated *Blue Ridge*; UNCW 4493, Polk Co., Saluda, SubAd female *SE Blue Ridge*; UNCW 4743, Dare Co., Buffalo City, SubAd male *Alligator River NWR region*; UNCW 4746, Hyde Co., Engelhard, SubAd female *Lake Mattamuskeet / Alligator NWR region*; UNCW 1243, Hyde Co., Swanquarter, SubAd male *Lake Mattamuskeet / Alligator River NWR region*; UNCW 4788, Hyde Co., Engelhard, Ad female *Lake Mattamuskeet / Alligator River NWR region*; UNCW 4801, Dare Co., Buffalo City, SubAd male *Alligator River NWR region*; UNCW 1992, Yancey Co., Busick, Ad female *Blue Ridge*; UNCW 4972, Dare Co., Stumpy Point, SubAd female *Alligator River NWR region*; UNCW 3298, Perquimans Co., Hickory Crossroads, SubAd no sex indicated *NE corner of state*; UNCW 1948, Yancey Co., Busick, Ad female *Blue Ridge*; UNCW 13721, Haywood Co., Maggie Valley, Ad female *Blue Ridge*; UNCW 3195, Gates Co., Sunbury Ad female *NE corner of state*; UNCW 9815, Macon Co., Highlands, SubAd male *Blue Ridge*

*** 7/15 Blue Ridge, 6/15 Lake Mattamuskeet / Alligator River NWR region; 8/15 [53.3%] female, 5/15 [33.3%] male, 2/15 [13.3%] no sex indicated; 9/15 [60.0%] sub adult, 6/15 [40.0%] adult, 0/15 [0.0%] old adult

5th split MAB > 7.565 → (NODE 11)

1 Blbrkn (UNCW 11814) UNCW 11814, Duplin Co., Rose Hill, Ad no sex indicated *center of distribution*;

7 Blbrta (UNCW 3297, Chowan Co., Edenton, SubAd no sex indicated **NE Coastal Plain**; UNCW 3730, Hertford Co., Ahoski, Ad female **NE Coastal Plain**; UNCW 4482, Polk Co., Saluda, SubAd female **SE Blue Ridge**; UNCW 2365, Swain Co., Clingmans Dome, SubAd female **Blue Ridge**; UNCW 2367, Swain Co., Clingmans Dome, Ad female **Blue Ridge**; UNCW 2161, Avery Co., Linville SubAd female **Blue Ridge**; UNCW 2018, Transylvania Co., Wagon Road Gap, SubAd female **Blue Ridge**)

*** 5/7 Blue Ridge, 2/7 NE Coastal Plain; 6/7 female, 1/7 no sex indicated; 5/7 sub adult, 2/7 adult

6^{th} split OPML ≥ 21.88 → (NODE 12)

12 Blbrkn (UNCW 12842, New Hanover Co., Castle Hayne, SubAd no sex indicated *SE corner of distribution*; UNCW 5088, Pender Co., Watha, SubAd female *center of distribution*; UNCW 2891, Duplin Co., Wallace, SubAd female *center of distribution*; UNCW 783, Robeson Co., Saint Pauls, Ad no sex indicated *SE to central distribution*; UNCW 0533, Pender Co., Scotts Hill, SubAd male *coastal center of distribution*; UNCW 784, Hoke Co., Raeford, SubAd no sex indicated *SW corner of distribution*; UNCW 5917, Duplin Co., Warsaw, SubAd female *center of distribution*; UNCW 2356, Beaufort Co., Aurora, SubAd no sex indicated *NE corner of distribution*; UNCW 2293, Beaufort Co., Aurora, SubAd no sex indicated *NE corner of distribution*; UNCW 4495, Pender Co., Hamstead, SubAd no sex indicated *coastal center of distribution*; UNCW 0534, Pender Co., Scotts Hill, SubAd no sex indicated *coastal center of distribution*; UNCW 11565, Duplin Co., Warsaw, SubAd male *center of distribution*)

*** 8/12 near center of distribution, 4/12 near corner of distribution; 3/12 female, 2/12 male, 7/12 no sex indicated; 11/12 [91.7%] sub adult, 1/12 adult

1 Blbrta (UNCW 3357, Pasquotank Co., Morgans Corner, Ad male *NE corner of state*)

7^{th} split UTR < 10.04 → (NODE 14)

10 Blbrkn (UNCW 4996, Pender Co., Burgaw, OldAd female *center of distribution*; UNCW 2606, Duplin Co., Wallace, OldAd male *center of distribution*; UNCW 11548, Duplin Co., Warsaw, OldAd no sex indicated *center of distribution*; UNCW 3098, New Hanover Co., Wilmington, Ad female* SE corner of distribution*; UNCW 4976, Duplin Co., Warsaw, Ad male *center of distribution*; UNCW 11126, Duplin Co., Magnolia, SubAd no sex indicated *center of distribution*; UNCW 10111, Pender Co., Moores Creek, SubAd male *center to SE of distribution*; UNCW 5259, Duplin Co., Magnolia, Ad male *center of distribution*; UNCW 0526, Pender Co., Scotts Hill, SubAd female *coastal center of distribution*; UNCW 5158, Duplin Co., Magnolia, Ad male *center of distribution*)

2 Blbrta (UNCW 13393, Buncombe Co., Dillingham Ad male *Blue Ridge*; UNCW 3303, Camden Co., Moyock, SubAd no sex indicated *NE corner of state*)

8th split CRH < 6.835 → (NODE 16)

15 Blbrkn (UNCW 5084, Pender Co., Watha, SubAd no sex indicated *center of distribution*; UNCW 11818, Duplin Co., Warsaw, SubAd no sex indicated *center of distribution*; UNCW 10109, Pender Co., Moores Creek, SubAd male *center to SE of distribution*; UNCW 11553, Duplin Co., Rose Hill, Ad no sex indicated *center of distribution*; UNCW 11559, Pender Co., Watha, Ad no sex indicated *center to SE of distribution*; UNCW 5191, Pender Co., Castle Hayne, Ad male *center to SE of distribution*; UNCW 2476, Beaufort Co., Aurora, SubAd female *NE corner of distribution*; UNCW 10108, Pender Co., Moores Creek, Ad male *center to SE of distribution*; UNCW 2659, Duplin Co., Wallace, OldAd female *center of distribution*; UNCW 5841, Duplin Co., Warsaw, Ad female *center of distribution*; UNCW 11927, Pender Co., Willard, SubAd no sex indicated *center to SE of distribution*; UNCW 0499, New Hanover Co., Wilmington, Ad male* SE corner of distribution*; UNCW 5916, Pender Co., Willard, SubAd no sex indicated *center to SE of distribution*; UNCW 4978, Pender Co., Wallace, OldAd male *center to SE of distribution*; UNCW 5085, Pender Co., Watha, OldAd female *center to SE of distribution*)

7 Blbrta (UNCW 5047, Hyde Co., Engelhard, SubAd no sex indicated *Lake Mattamuskeet / Alligator NWR region*; UNCW 2118, Hyde Co., Swanquarter, SubAd female *Lake Mattamuskeet / Alligator NWR region*; UNCW 2038, Haywood Co., no specific locality, SubAd no sex indicated *Blue Ridge*; UNCW 4800, Hyde Co., Engelhard, Ad female *Lake Mattamuskeet / Alligator NWR region*; UNCW 4195, Perquimans Co., Hertford SubAd female *NE corner of state*; UNCW 4489, Polk Co., Saluda, SubAd female *SE Blue Ridge*; UNCW 4744, Dare Co., Buffalo City, SubAd female *Alligator NWR region*)

9th split CRH > 6.835 → (NODE 17)

6 Blbrkn (UNCW 2130, New Hanover Co., Wilmington, SubAd female* SE corner of distribution*; UNCW 2608, Duplin Co., Wallace, SubAd female *center of distribution*; UNCW 11939, Pender Co., Watha, SubAd male *center to SE of distribution*; UNCW 4402, Sampson Co., Newton Grove, SubAd no sex indicated *center of distribution*; UNCW 1480, Beaufort Co., Aurora, SubAd female *NE corner of distribution*; UNCW 5083 Pender Co., Willard, SubAd no sex indicated *center to SE of distribution*)

17 Blbrta (UNCW 3290, Currituck Co., Moyock, Ad no sex indicated *NE corner of state coastal*; UNCW 3269, Currituck Co., Moyock, SubAd no sex indicated *NE corner of state coastal*; UNCW 4494, Polk Co., Saluda, SubAd no sex indicated *SE Blue Ridge*; UNCW 3108, Alleghany Co., Beach Mountain, OldAd female *Blue Ridge*; UNCW 4525, Perquimans Co., Hertford SubAd male *NE corner of state*; UNCW 4799, Hyde Co., Engelhard, Ad male *Lake Mattamuskeet / Alligator NWR region*; UNCW 3175, Gates Co., Roduco, Ad male *NE corner of state*; UNCW 3148, Gates Co., Sunbury, Ad female *NE corner of state*; UNCW 1996, Yancey Co., Busick, Ad male *Blue Ridge*; UNCW 4485, Polk Co., Saluda, SubAd female *SE Blue Ridge*; UNCW 13446, Haywood Co., Big East Fork, SubAd female *Blue Ridge*; UNCW 9543, Madison Co., Marshall, Ad no sex indicated *Blue Ridge*; UNCW 2796, Haywood Co., Soco Gap, Ad female *Blue Ridge*; UNCW 4971, Dare Co., Stumpy Point, SubAd no sex indicated *Alligator NWR region coastal*; UNCW 5090, Wilkes Co., Pores

Knob, SubAd no sex indicated *Blue Ridge – NE edge with Piedmont*; UNCW 13409, Buncombe Co., Dillingham, SubAd female *Blue Ridge*; UNCW 4491, Polk Co., Saluda, SubAd female *SE Blue Ridge*)

MANDIBLE

1st split is GrLgt < 13.185 → (NODE 2)

2 Blbrkn with GrLgt < 13.185 (UNCW 3596, GrLgt = 12.85, New Hanover Co., Wilmington; UNCW 0477, GrLgt = 12.93, New Hanover Co., Carolina Beach *both near type locality southern part of distribution* *SE corner of distribution*)
All remaining are Blcaca

2nd split is CorHt > 6.095 → (NODE 5)

3 Blbrkn with CorHt ≥ 6.095 (UNCW 11938, CorHt = 6.16, Duplin Co, 5 km S Warsaw *almost center of distribution*, Ad no sex indicated; UNCW 1480, CorHt = 6.21, Beaufort Co., 4 mi (~6.4 km) N Aurora *northern part of distribution – on peninsula created by Pamilico River to N, Durham Creek to W and South Creek to E – phosphate mine with many pools of water*; UNCW 11554, CorHt = 6.3, Duplin Co, 5 km S Warsaw *almost center of distribution*)

119 Blbrta: 10 smallest Blbrta (UNCW 4499, Polk Co, Saluda, SubAd-Ad male; UNCW 5090, Wilkes Co, Pores Knob, SubAd no sex indicated; UNCW 3600, Madison Co, Hot Springs, Old Ad female, UNCW 3188, Gates Co, Sunbury *Coastal Plain*, Ad female; UNCW 13409, Buncombe Co, Dillingham, SubAd-Ad female; UNCW 4488, Polk Co, Saluda, SubAd female; UNCW 4491, Polk Co, Saluda, SubAd female; UNCW 4493, Polk Co, Saluda, SubAd female; UNCW 3072, Gates Co, Storys *Coastal Plain*, SubAd female; UNCW 4486, Polk Co, Saluda, Ad male); Next 10 smallest Blbrta: UNCW 9542, Henderson Co., 2 km E Upward, Ad male; UNCW 2796, Haywood Co., Soco Gap, Ad female; UNCW 13446, Haywood Co., Big East Fork, SubAd female; UNCW 1996, Yancey Co., 1 mi W Busick, Ad male; UNCW 9815, Macon Co., Highlands, SubAd male; UNCW 13457, Buncombe Co., Dillingham, Ad female; UNCW 5842, Burke Co., Ramsey, Ad – Old Ad no sex indicated; UNCW 3601, Madison Co., Hot Springs, Ad female; UNCW 1988, Stokes Co., King, Ad male; UNCW 392, Mitchell Co., Roan Mountain, Ad female)

*** 7/10 edge (BR & Piedmont) or Coastal Plain; 7/10 female; 6/10 subadult to young adult age

3rd split is MNL < 12.94 → (NODE 6)

33 Blbrta with MNL < 12.94 (UNCW 5086, Dare Co., Buffalo City, SubAd-Ad no sex indicated; UNCW 4489, Polk Co., Saluda, SubAd-Ad female; UNCW 4494, Polk Co., Saluda, SubAd-Ad no sex indicated; UNCW 9543, Madison Co., Marshall, Ad no sex indicated **very small body and skull BLUE RIDGE**; UNCW 5240, Hyde Co., Lake Mattamuskeet, Ad-oldAD male; UNCW 4526, Perquimans Co., Hertford, Ad no sex indicated; UNCW 4525, Perquimans Co., Hertford, SubAd male; UNCW 5098, Hyde Co., Engelhard, SubAd-Ad male; UNCW 4799,

Hyde Co., Englehard, Ad male; UNCW 3357, Pasquotank Co., Morgans Corner, Ad male; UNCW 3175, Gates Co., Roduco, Ad male; UNCW 4788, Hyde Co., Englehard, Ad female; UNCW 3364, Pasquotank Co., Morgans Corner, Ad male; UNCW 5047, Hyde Co., Englehard, SubAd no sex indicated; UNCW 4485, Polk Co, Saluda, SubAd female; UNCW 3731, Hertford Co. Ahoski, Ad female; UNCW 3066, Gates Co., Roduco, Ad male; UNCW 3108, Alleghany Co., Beach Mountain, OldAd female **BLUE RIDGE**; UNCW 4800, Hyde Co., Englehard, Ad female; UNCW 5093, Hyde Co., Englehard, SubAd no sex indicated; UNCW 2118, Hyde Co., Swanquarter, OldAd male; UNCW 1243, Hyde Co., Swanquarter, SubAd male; UNCW 4482, Polk Co., Saluda, SubAd female; UNCW 3290, Currituck Co., Moyock, Ad-OldAd no sex indicated; UNCW 3594, Gates Co., Roduco, Ad female; UNCW 4196, Chowan Co., Edenton, Ad male; UNCW 2786, Hyde Co., New Holland, Ad male; UNCW 5094, Hyde Co., Englehard, SubAd no sex indicated; UNCW 3730, Hertford Co., Ahoski, Ad female; UNCW 4195, Perquimans Co., Hertford, Ad female; UNCW 5843, Burke Co., Ramsey, Ad-OldAd no sex indicated; UNCW 3015, Gates Co., Roduco, Ad female; UNCW 4746, Hyde Co., Englehard, Ad female)

*** 13/33 [39.4%] Alligator River/Lake Mattamuskeet 13/33 [39.4%] northeast coastal plain (8/13 [61.5%] near edge) 7/33 [21.2%] in Blue Ridge (4/7 [57.1%] at SE edge of blue ridge in NC); 12/33 [36.4%] female, 12/33 [36.4%] male, 9/33 [27.3%] no sex indicated; 11/33 [33.3%] sub adult, 17/33 [51.5%] adult, 5/33 [15.2%] old adult

MNL > 12.94 → (NODE 7)

11 Blbrkn >12.94 (UNCW 12842, New Hanover, Castle Hayne, SubAd-Ad no sex indicated; UNCW 0528, Pender Co., Scotts Hill, SubAd-Ad male; UNCW 10108, Pender Co., Moores Creek, Ad-old Ad male; UNCW 2891, Duplin Co., Wallace, Young Ad female; UNCW 5191, Pender Co., Castle Hayne, Ad male; UNCW 11939, Pender Co., Watha, SubAd-Ad male; UNCW 0533, Pender Co., Scotts Hill, SubAd male; UNCW 4496, Pender Co., Hamstead, SubAd no sex indicated; UNCW 0527, Pender Co., Scotts Hill, SubAd no sex indicated; UNCW 5088, Pender Co., Watha, SubAd female; UNCW 0792, Bladen Co., Ivanhoe, SubAd-Ad no sex indicated)

*** 10/11 [90.9%] are near ~center of coastal populations (8/11 [72.7%] Pender county), all from Middle Atlantic Coastal Plain, north of Cape Fear River; 5/11 [45.4%] male, 2/11 [18.2%] female, 4/11 [36.4%] no sex indicated; 4/11 [36.4%] sub adult, 6/11 [54.5%] adult, 1/11 [9.1%] old adult

28 Blbrta with MNL ≥ 12.94 (UNCW 3053, Gates, Roduco, Old Ad female; UNCW 5097, Hyde Co., New Holland, SubAd female; UNCW 3054, Gates Co., Sunbury, Ad female; UNCW 4801, Dare Co., Buffalo City, SubAd male; UNCW 4481, Polk Co., Saluda, SubAd female; UNCW 3965, Gates Co., Sunbury, Ad female; UNCW 4744, Dare Co., Buffalo City, SubAd female; UNCW3297, Chowan Co., Edenton, SubAd-Ad no sex indicated; UNCW 5239, Hyde Co., Lake Mattamuskeet, SubAd-Ad male; UNCW 3278, Currituck Co., Moyock, Ad no sex indicated; UNCW 0484, Gates Co., Sunbury, Ad female; UNCW 10581, Dare Co., Stumpy Point, SubAd-Ad female; UNCW 3298, Perquimans Co., Hickory Crossroads, SubAd-Ad no sex indicated; UNCW 3138, Gates Co., Sunbury, Ad female; UNCW 5180, Gates Co., Sunbury, Ad no sex

indicated; UNCW 4972, Dare Co., Stumpy Point, SubAd female; UNCW 4971, Dare Co., Stumpy Point, SubAd no sex indicated; UNCW 3076, Gates Co., Roduco, SubAd female; UNCW 3398, Camden Co., Dismal Swamp, SubAd no sex indicated; UNCW 3303, Camden Co., Moyock, SubAd no sex indicated; UNCW 10375, Gates Co., Roduco, Ad female; UNCW 3195, Gates Co., Sunbury, Ad female; UNCW 3295, Chowan Co., Edenton, SubAd no sex indicated; UNCW4743, Dare Co., Buffalo City, SubAd-Ad male; UNCW 3396, Gates Co., Sunbury, Ad so sex indicated; UNCW 3148, Gates Co., Sunbury, Ad no sex indicated; UNCW 3269, Currituck Co., Moyock, SubAd no sex indicated; UNCW 3728, Dare Co., Stumpy Point, SubAd-Ad female)

*** 27/28 [96.4%] northeast Middle Atlantic Coastal Plain; 14/28 [50.0%] female, 3/28 [10.7%] male, 11/28 [39.3%] no sex indicated; 17/28 [60.7%] sub adult, 10/28 [35.7%] adult, 1/28 [3.6%] old adult

APPENDIX III – ADDITIONAL MORPHOLOGICAL ANALYSES - ANALYSES NOT USED IN MOPHOLOGICAL CHAPTER SUPPORT AND CONFIRM SPLITTING OF SKULL AND MANDIBLE MEASUREMENTS FROM SHORT-TAILED SHREW SAMPLES.

Specimens examined - The 448 specimens analyzed with the skull data and the 461 specimens analyzed for the mandible data. All samples are from the University of North Carolina Wilmington (UNCW).

Bivariate pairs plots

I examined the pairwise combination of the 5 external measurements, and there was no clear separation in the five unique bivariate scatter plots (Figure AIII-2.1). Pairwise combinations of the cranial measurements resulted in 105 unique combinations of bivariate plots. Since some of the museum samples examined consisted of only the upper portion of the skull or only the mandible, I examined the pairwise combination on skull (Figure AIII-2.2) and the mandible (Figure AIII-2.3) separately. For combinations examined further, I plotted the points as their respective subspecies designation where: 'c' = *Blarina carolinensis carolinensis*, 'k' = *B. brevicauda knoxjonesi*, and 't' = *B. b. talpoides*.

The skull had 29 of the 36 (80.6%) combinations and the mandible had 10 of the 15 (66.7%) combinations that show a separation of the points created by the pairs of measurements. Figure AIII-2.4 compares the bivariate plots for select skull and mandible combinations by x and y axes. The skull comparisons were: top-left – upper tooth row including incisor (UTR) vs. cranial height (CRH), and the bottom-left – length of molariform tootrow (MTL) vs. occipital-premaxilla length without incisor (OPML) with R-squared or coefficient of determination values of 0.7124 and 0.8749 respectively. The mandible comparisons were: top-right – articular breadth (ARB) vs. greatest length of mandible with incisor (GrLgt), and the bottom-right – coronoid

process height (CorHt) vs. GrLgt with R-squared or coefficient of determination values of 0.6898 and 0.8475 respectively.

Principal components analysis

I performed the PCA with a correlation matrix, which shows how correlated the variables are to one another. The correlation matrix is essential for different units of measure and is a standardized conversion between variables, where the within variable correlation is 1 showing that it is perfectly correlated. The correlation matrix was analyzed on the skull data as a comparison to the covariance matrix. A scree plot showed that 2 principal components were sufficient, so I plotted a scatter plot of the first two principal components (Figure AIII-2.5). This plot separated *Blarina brevicauda* and *B. carolinensis* with little overlap, but showed more overlap in *B. b. knoxjonesi* and *B. b. talpoides*.

PC1, PC2 and PC3 had eigenvalues of 8.030, 0.353 and 0.208 with a cumulative proportion of the sample variance explained as 0.8922, 0.9314 and 0.9546 respectively. PC1 was a weighted sum of GRL, OPML, UTR and MTL (-0.347, -0.346, -343 and -0.337), and with no change in the order, the associated correlation coefficients for these variables were -0.984, -0.979, -0.972 and -0.955. PC2 was a weighted difference primarily between the variables UCL and UTR and the variables MAB and CRB (0.569, 0.332, -0.453 and -0.450). The correlation coefficients also showed a weighted difference of the variables UCL and UTR and the variables MAB and CRB (0.338, 0.197, -0.269 and -0.267). Table AIII-2.1 summarizes the skull correlation matrix PCA.

Principal components and maximum likelihood factor analysis

Factor analysis is considered an extension of principal components analysis (Johnson and Wichern, 2007). I performed an un-rotated and rotated principle components and maximum likelihood factor analysis to examine the covariance relationship between the measured variables.

Principal components factor analysis

I performed a principal components factor analysis (PCFA) on the *Blarina* skull data (n = 448) with a covariance matrix to describe the shared variance, or covariance relationship, among the measured variables in terms of factors. A scree plot is used, like in PCA, to show that 2 factors are sufficient to compare the variable covariance. The PCFA scatter plot (Figure AIII-2.6) showed a similar separation for *B. brevicauda* and *B. carolinensis* and the subspecies *B. b. knoxjonesi* and *B. b. talpoides* as the PCA. The PCFA yielded a sum of squares on the variable factor loadings for factor 1 (F1) of 7.19 and 0.18 for factor 2 (F2). Table AIII-2.2 summarizes the output from the PCFA. F1 represents high loadings for GRL, OPML, CRB and UTR (-1.631, -1.531, -0.838 and -0.819), while F2 has highest loading on CRB (-0.335) followed by MAB, GRL and UTR (-0.163, 0.120 and 0.106). F1 explains 79.9% of the sample variance and F2 explains 2.0% for a combined 81.9% of the total sample variance explained.

Factor analysis provides a communality that indicates the portion of the variance of each variable contributed by the factors, and a specific variance that is the portion of the variance due to the specific factors (Johnson and Wichern, 2007). The communality is the sum of the squared factor loadings across all factors for a given variable, and measures the percent of variance in a given variable explained by all the factors jointly and may be interpreted as the reliability of the

indicator. The PCFA shows the most reliable indicators of the variance are GRL, OPML followed by CRB and UTR (2.675, 2.349, 0.815 and 0.682) for the communalities. The specific variance or uniqueness of a variable is 1 minus the communality, where the larger the number, the more unique the variable. The UCL, IOB, MTL CRH and MAB (0.903, 0.860, 0.852, 0.796 and 0.737) are the most unique variables, which are more unrelated to the other variables, to explain the observed variability with two factors.

I then examined a rotated PCFA with a varimax rotation, which maximizes the sum of the variances of the squared loadings and changes the orthogonal basis of the two factors (Kaiser, 1958). The rotated PCFA scatter plot (Figure AIII-2.7) showed the separation of the species, but shows a better grouping for the subspecies. By rotating the factor loadings, the rotated F1 has a slightly lower SS loading of 7.14, while the rotated F2 has a slightly higher SS loading of 0.22 when compared to the un-rotated PCFA. Table AIII-2.2 summarizes the output from the rotated PCFA. The rotated F1 represents high loadings for GRL, OPML, UTR and CRB (-1.631, -1.531, -0.838 and -0.819), while the rotated F2 reduced the variables to only CRB, MAB and CRH (-0.403, 0.203 and -0.108). The rotated F1 explains 79.4% of the sample variance and rotated F2 explains 2.5% for a combined 81.9% of the total sample variance explained. The rotated F1 explains slightly less and the rotate F2 explains slightly more of the total sample variance when compared to the un-rotated PCFA. Rotating the factors did not change the variable communality or the specific variances.

Maximum likelihood factor analysis

I performed a maximum likelihood factor analysis (MLFA) on the *Blarina* skull data (n = 448) with a covariance matrix to describe the shared variance, or covariance relationship, among

the measured variables in terms of factors. The MLFA standardizes the original covariance matrix to a correlation matrix, where the factor scores are optimized by the log likelihood over the uniqueness and estimated by weighted least squares from the maximum likelihood estimates (Bartlett, 1937). The MLFA tests the null hypothesis that two factors are not sufficient to explain the observed variation. A chi square statistic of 551.57 on 19 degrees of freedom yielded a $p < 0.001$, so I rejected the null hypothesis and accepted that two factors are not sufficient. The scatter plot of the first two factor loadings from the skull data (Figure AIII-2.8) separated *Blarina brevicauda* and *B. carolinensis* with some overlap between the species, but created a tighter cluster of points.

Factor 1 (F1) explains 87.5% of the sample variance and factor 2 (F2) explains 3.5% for a combined 91.0% of the total proportion of the sample variance explained. Table AIII-2.3 summarizes the output from the MLFA. F1 has a SS loading of 7.87, and is weighted by the loadings for GRL, OPML, UTR and MTL (0.997, 0.990, 0.975 and 0.948). F2 has a SS loading of 0.30, and is reduced to the variables CRB, MAB, IOB, UCL and CRH (0.346, 0.316, 0.184, -0.168 and 0.143). The MLFA communalities show the most reliable indicators are GRL, OPML, UTR and CRB (0.995, 0.980, 0.959 and 0.945), and the most unique variables are CRH, IOB, UCL, MAB and MTL (0.198, 0.143, 0.138, 0.112 and 0.101) with the specific variance values.

I also examined a rotated MLFA with a varimax rotation, where the scatter plot of the rotated MLFA (Figure AIII-2.9) has a condensed distribution with little scatter of the factor loadings. The rotated factor 1 (F1) explains 47.2% of the sample variance and the rotated factor 2 (F2) explains 43.8% for a combined 91.0% of the total proportion of the sample variance explained. Table AIII-2.3 summarizes the output from the rotated MLFA. The rotated factors lowered the rotated F1 SS loading to 4.25, and increased the rotated F2 SS loading to 3.94 to balance the

weight of the factors. The rotated F1 is weighted by UCL, UTR, GRL, OPML and MTL (0.814, 0.812, 0.796, 0.766 and 0.730), while rotated F2 is weighted on the loadings for CRB, MAB, IOB and CRH (0.842, 0.806, 0.715 and 0.669). Rotating the factors did not change the variable communality or the specific variances as in the principal components factor analysis.

Classification tree analysis

I performed a classification tree analysis on the *Blarina* skull (n = 448) and mandible (n = 461) data with an information splitting index using the RPART (Recursive PARTintioning; Therneau and Atkinson, 1997) package in R. An information splitting index uses information gain, which is the probability of a randomly chosen example belonging to a class and a classification entropy that involves a logarithmic transformation (Berzal et al., 2003; Raileanu and Stoffel, 2004). The information gain is also called the Kullback-Leibler divergence (Kullback, 1959; Kullback and Leibler, 1951), and is most commonly used in the iterative dichotomiser 3 (ID3 – Quinlan, 1986) and the C4.5 (Quinlan, 1993) algorithms. Although used for different decision trees, I compared this with the Gini information index in the classification tree analysis.

The skull data yielded slightly different tree with the information splitting index when comparing it to the Gini splitting index (Figure AIII-2.10). All of the skulls identified as *Blarina carolinensis carolinensis* had a smaller occipital-premaxilla skull length (OPML < 19.735, see Figure 2.1 for cranial descriptions) as the first split or node, where all the skulls identified as *B. brevicauda* had a larger skull length (OPML ≥ 19.735). The second split at node 3 used a cranial breadth measure (CRB < 11.845). Node 5 split 62.8% of all the skulls identified as *B. b. talpoides* with a larger cranial breadth. This node also split 6 skulls (UNCW 3682, 2610, 3595,

792, 11938 and 4975) identified as *B. b. knoxjonesi* with larger cranial breadths. 55.2% of all remaining skulls identified as *B. b. knoxjonesi* have a more narrow maxillary breadth (MAB < 7.165) in both trees, but there are also 11 skulls identified as *B. b. talpoides* with a narrow maxillary breadth at node 6.

The information tree continues with maxillary breadth (MAB), OPML, OPML, length of molariform toothrow (MTL) and cranial height (CRH) as the main splits. The proportion of specimens grouped was close to that in the Gini tree with 10 *B. b. knoxjonesi* and 42 *B. b. talpoides* specimens being grouped as the *talpoides* group and 40 *B. b. knoxjonesi* and 11 *B. b. talpoides* specimens grouped as the *knoxjonesi* group (Figure 1.23). The misclassification error rate for the skull gave a resubstitution error rate of 8.5% or 91.5% classified accurately, and a cross-validated error rate of 16.5% or 83.5% classified correctly (Table AIII-2.4).

The information splitting index had slightly higher improvements (i.e. gain or impurity reduction) at the primary splits when compared to the Gini splitting index due to the differences in scaling (Table AIII-2.4). The first primary split compared multiple measured variables and resulted in OPML < 19.735 that had greatest improvement (286.2997), and had the lowest complexity parameter ($\alpha = 0.5471$). Both the improvements and the complexity parameter decreased toward the final split at node 15. This decrease was the result of a decreasing number of terminal nodes and the number of specimens with the class grouping.

The mandible data resulted in identical trees for the information splitting indices (Figure AIII-2.11) when compared to the Gini splitting index. All of the mandibles identified as *B. c. carolinensis* had a shorter greatest length of mandible including the incisor (GrLgt < 13.18), but 2 mandibles (UNCW 3596 and 0477) identified as *B. b. knoxjonesi* also had a short mandible length. The next split showed 66.1% of all *B. b. talpoides* specimens with a taller coronoid

process height (CorHt >6.095), but 3 specimens (UNCW 11938, 1480 and 11554) identified as *B. b. knoxjonesi* also grouped here. The last split separated the remaining specimens with a length of mandible without the incisor (MNL < 12.94), where 88.7% of all *B. b. knoxjonesi* had a shorter mandible were classified as the 'knoxjonesi' group. However, this small mandible size also grouped 33 specimens identified as *B. b. talpoides*. For the mandibles with a longer length, 11 specimens identified as *B. b. knoxjonesi* and 28 specimen identified as *B. b. talpoides* were classified as the 'talpoides' group.

The misclassification error rate for the mandible data with the information splitting index gave a re-substitution error rate of 10.6% or 89.4% classified accurately, and a cross-validated error rate of 12.8% or 87.2% classified correctly (Table AIII-2.5). The improvements reduced quicker with the mandible data due to only 3 primary splits in the tree, and resulted in higher values than the Gini splitting index due to the scaling. The information splitting index tree with the mandible data showed improvements of 285.4192 at node 1, 73.3750 at node 3 and 15.6257 at node 4.

LITERATURE CITED

Bartlett, M. S. 1937. The statistical conception of mental factors. British Journal of Psychology 28: 97-104.

Berzal, F., J. C. Cubero, F. Cuenca and M. J. Martín-Bautista. 2003. On the quest for easy-to-understand splitting rules. Data & Knowledge Engineering 44(1): 31-48.

Johnson, R. A. and D. W. Wichern. 2007. Applied multivariate statistical analysis. 6th edition. Pearson Prentice Hall, Upper Saddle River, New Jersey. 773 pp.

Kaiser, H. F. 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23: 187-200.

Kullback, S. and L. A. Leibler. 1951. On information and sufficiency. The annals of mathematical statistics 22(1): 79-86.

Kullback, S. 1959. Information theory and statistics. John Wiley. New York.

Quinlan, J. R. 1986. Induction of decision trees. Machine learning 1(1): 81-106.

Quinlan, J. R. 1993. C4.5: programs for machine learning. Morgan Kaufmann Publishers, Inc. San Francisco, California.

Raileanu, L. E. and K. Stoffel. 2004. Theoretical comparison between the gini index and information gain criteria. Annals of Mathematics and Artificial Intelligence 41(1): 77-93.

Therneau, T. M. and E. J. Atkinson. 1997. An introduction to recursive partitioning using the RPART routine. Technical Report 61, Mayo Clinic, Section of Biostatistics, Rochester, Minnesota, USA.

Table AIII-2.1: Principal components analysis table using the correlation matrix of the *Blarina* skull data (n = 448) showing what factors contribute to each principal component explaining the sample variation. The variables are the skull characters (Figure 1.4) measured, while the eigenvectors (e^_1) and correlation coefficients (ry^1) give the relative weight each variable contributes. The variance (lambda – λ_i) is an eigenvalue that is the measure of the amount of the variation explained by the principal component. Each principal component accounts for a smaller proportion of the total sample variation, but when combined, the first three principal components account for most of the total sample variation.

	Principal Component 1		**Principal Component 2**		**Principal Component 3**	
Variable	**e^_1**	**ry^1**	**e^_2**	**ry^2**	**e^_3**	**ry^3**
GRL	-0.34735	-0.98427	0.173082	0.102839	0.005994	0.002737
OPML	-0.34559	-0.97928	0.094983	0.056435	-0.00583	-0.00266
MAB	-0.32583	-0.9233	-0.45314	-0.26924	-0.29791	-0.13603
IOB	-0.32858	-0.93109	-0.25909	-0.15394	-0.27234	-0.12435
CRB	-0.32987	-0.93474	-0.45009	-0.26742	0.03674	0.016776
CRH	-0.31999	-0.90674	-0.17693	-0.10512	0.867322	0.396028
UTR	-0.34305	-0.97209	0.332119	0.197333	-0.08776	-0.04007
MTL	-0.33706	-0.95511	0.142524	0.084682	-0.26953	-0.12307
UCL	-0.32137	-0.91067	0.569263	0.338235	0.055351	0.025274
Variance (λ_i)	8.029607		0.353029		0.208493	
↘ % of Total Variance	0.8922		0.0392		0.0232	
∑ % of Total Variance	0.8922		0.9314		0.9546	

Table AIII-2.2: Principal components factor analysis table using the covariance matrix of the *Blarina* skull data (n = 448) showing what variables (Figure 1.4) contribute to each factor explaining the sample variation. The estimated factor loadings and the rotated factor loadings are the sample principal component coefficients (eigenvector) scaled by the square root of the corrsponding eigenvalue. The communality is the proportion of the variation for the respective measured variable contributed by the two factor loadings and is the sum of the sqauared factor loadings for each given variable. The specific variance is the portion of variance due to the specific variable. The sum of squares loadings (SS loadings) indicates the weight of each factor, and the proportion and cummulative proportion of the total variance expained is indicated.

	Estimated Factor Loadings		**Rotated Factor Loadings**			
Variable	**F1**	**F2**	**F1 rot**	**F2 rot**	**Communalities**	**Specific Variances**
GRL	-1.631	0.1199	-1.635		2.6746759	-1.674676
OPML	-1.5312	0.0688	-1.532		2.34945792	-1.3494579
MAB	-0.4867	-0.1629	-0.472	-0.203	0.26346984	0.7365302
IOB	-0.3689	-0.0588	-0.363		0.13956781	0.8604322
CRB	-0.8383	-0.3347	-0.808	-0.403	0.81474203	0.185258
CRH	-0.4464	-0.0716	-0.439	-0.108	0.20437633	0.7956237
UTR	-0.8187	0.1063	-0.825		0.68152124	0.3184788
MTL	-0.385	0.0166	-0.385		0.1484925	0.8515075
UCL	-0.3055	0.0628	-0.310		0.09728103	0.902719
SS loadings	7.191543	0.181773	7.145337	0.215282		
Proportion of Total Sample Variance Explained	0.79909263	0.02019455	0.79376011	0.02552706	Total = 7.37358459	
Cumulative Proportion of Total Sample Variance Explained	0.79909263	0.8192872	0.79376011	0.8192872	Average = 0.819287177	

Table AIII-2.3: Maximum likelihood factor analysis table using the covariance matrix of the *Blarina* skull data (n = 448) showing what variables (Figure 1.4) contribute to each factor explaining the sample variation. The estimated factor loadings and the rotated factor loadings are the sample principal component coefficients (eigenvector) scaled by the square root of the corrsponding eigenvalue. The communality is the proportion of the variation for the respective measured variable contributed by the two factor loadings and is the sum of the sqauared factor loadings for each given variable. The specific variance is the portion of variance due to the specific variable. The sum of squares loadings (SS loadings) indicates the weight of each factor, and the proportion and cummulative proportion of the total variance expained is indicated.

	Estimated Factor Loadings		**Rotated Factor Loadings**			
Variable	**F1**	**F2**	**F1 rot**	**F2 rot**	**Communalities**	**Specific Variances**
GRL	0.997		0.796	0.602	0.995141	0.004859
OPML	0.990		0.766	0.628	0.9803735	0.01962655
MAB	0.888	0.316	0.489	0.806	0.8882447	0.11175526
IOB	0.907	0.184	0.588	0.715	0.8570694	0.14293065
CRB	0.908	0.346	0.486	0.842	0.9446636	0.0553364
CRH	0.884	0.143	0.595	0.669	0.8016846	0.19831541
UTR	0.975		0.812	0.548	0.9588263	0.0411737
MTL	0.948		0.730	0.605	0.8985502	0.10144978
UCL	0.913	-0.168	0.814	0.447	0.8622256	0.1377744
SS loadings	7.87412	0.302101	4.250298	3.940312		
Proportion of Total Sample Variance Explained	0.87507948	0.03456261	0.4720485	0.4375936	Total = 8.1867789	
Cumulative Proportion of Total Sample Variance Explained	0.8750795	0.9096421	0.4720485	0.9096421	Average = 0.9096421	

Table AIII-2.4: Classification tree table with information splitting index from *Blarina* skull data (n = 448) showing what factors best split the data into subpecies classes. The subspecies of the southern short-tailed shrew (*Blarina carolinensis carolinensis* – Blcaca) and the northern short-tailed shrew (*B. brevicauda knoxjonesi* – Blbrkn and *B. b. talpoides* – Blbrta) represent the classes used while measured cranial characters (Figure 1.4) determine the primary split in the tree. Relative error (rel error), apparent error (x error) and apparent standard deviation (x std) are given for each complexity parameter to compute two measures of predictive performance using the root node error.

Tree Node #	**Blcaca**	**Blbrkn**	**Blbrta**	**Total at Node**	**Proportion at Node (Total / 448)**	**Complexity Parameter (CP = α)**	**Expected Loss = # of other classes / total at node**	**Predicted Class**	**Primary Split**	**Improvements = Gain = impurity reduction**
1	151	125	172	448	1.0000	0.5471	0.6161	talpoides	OPML < 19.735	286.2997
2	151	0	0	151	0.3371		0.0000	carolinensis		
3	0	125	172	297	0.6629	0.1993	0.4209	talpoides	CRB < 11.845	60.1725
4	0	119	64	183	0.4085	0.0229	0.3497	knoxjonesi	MAB < 7.165	15.0694
5	0	6	108	114	0.2545		0.0526	talpoides		
6	0	69	11	80	0.1786		0.1375	knoxjonesi		
7	0	50	53	103	0.2299	0.0229	0.4854	talpoides	MAB < 7.585	6.3862
8	0	50	44	94	0.2098	0.0229	0.4681	knoxjonesi	OPML < 21.98	4.4472
9	0	0	9	9	0.0201		0.0000	talpoides		
10	0	46	30	76	0.1696	0.0157	0.3947	knoxjonesi	OPML < 21.885	4.5338
11	0	4	14	18	0.0402		0.2222	talpoides		
12	0	13	1	14	0.0313		0.0714	knoxjonesi		
13	0	33	29	62	0.1384	0.0157	0.4677	knoxjonesi	MTL < 5.745	3.5561
14	0	11	2	13	0.0290		0.1538	knoxjonesi		
15	0	22	27	49	0.1094	0.0157	0.4490	talpoides	CRH < 6.835	4.6553
16	0	16	8	24	0.0536		0.3333	knoxjonesi		
17	0	6	19	25	0.0558		0.2400	talpoides		
						CP	**rel error**	**x error**	**x std**	
						0.5471	1.0000	1.0000	0.0373	
						0.1993	0.4529	0.4565	0.0345	
						0.0229	0.2536	0.2717	0.0286	
						0.0157	0.1848	0.2754	0.0288	
						0.0100	0.1377	0.2681	0.0285	

Root Node Error = 276/448 = 0.6161

Re-substitution Error Rate = 0.6161 * 0. 1377 = 0.0848 or 8.5% misclassified

Cross-validated Error Rate = 0.6161 * 0.2681 = 0.1652 or 16.5% misclassified

Table AIII-2.5: Classification tree table with information splitting index from *Blarina* mandible data (n = 461) showing what factors best split the data into subpecies classes. The subspecies of the southern short-tailed shrew (*Blarina carolinensis carolinensis* – Blcaca) and the northern short-tailed shrew (*B. brevicauda knoxjonesi* – Blbrkn and *B. b. talpoides* – Blbrta) represent the classes used while measured cranial characters (Figure 1.4) determine the primary split in the tree. Relative error (rel error), apparent error (x error) and apparent standard deviation (x std) are given for each complexity parameter to compute two measures of predictive performance using the root node error.

Tree Node #	**Blcaca**	**Blbrkn**	**Blbrta**	**Total at Node**	**Proportion at Node (Total / 461)**	**Complexity Parameter (CP = α)**	**Expected Loss = # of other classes / total at node**	**Predicted Class**	**Primary Split**	**Improvements = Gain = impurity reduction**
1	155	126	180	461	1.0000	0.5516	0.6095	talpoides	GrLgt < 13.185	285.4192
2	155	2	0	157	0.3406		0.0127	carolinensis		
3	0	124	180	304	0.6594	0.2135	0.4079	talpoides	CorHt < 6.095	75.3750
4	0	121	61	182	0.3948	0.0605	0.3352	knoxjonesi	MNL < 12.94	15.6257
5	0	3	119	122	0.2646		0.0246	talpoides		
6	0	110	33	143	0.3102		0.2308	knoxjonesi		
7	0	11	28	39	0.0846		0.2821	talpoides		
						CP	**rel error**	**x error**	**x std**	
						0.5516	1.0000	1.0000	0.0373	
						0.2135	0.4484	0.4520	0.0341	
						0.0605	0.2349	0.2598	0.0279	
						0.0100	0.1744	0.2100	0.0255	

Root Node Error = 281/461 = 0.6095

Re-substitution Error Rate = 0.6095 * 0. 1744 = 0.1063 or 10.6% misclassified

Cross-validated Error Rate = 0.6095 * 0.2100 = 0.1280 or 12.8% misclassified

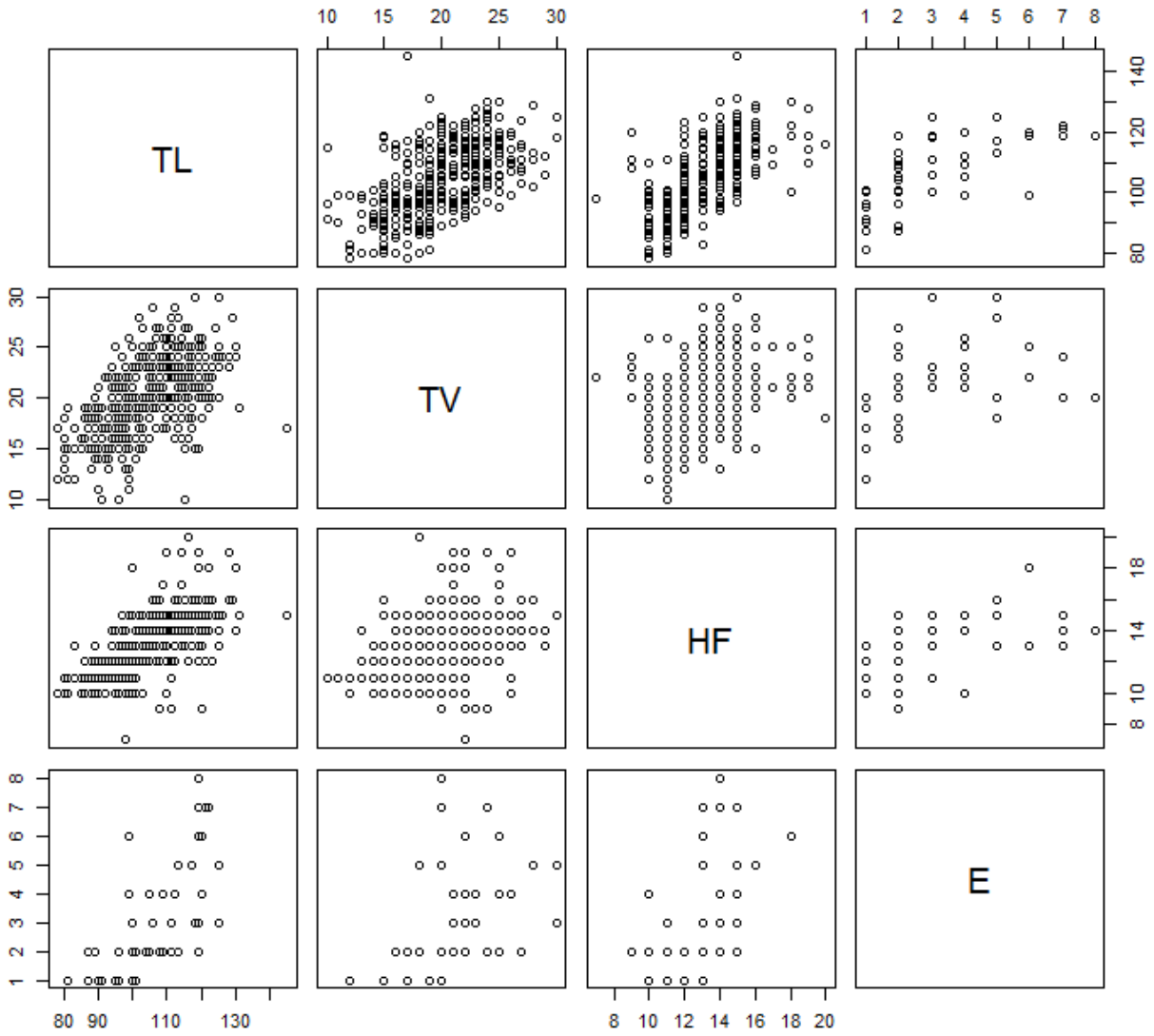


Figure AIII-2.1: Bivariate plot with 'pairs' function in R statistical software showing the combinations of external measurements from short-tailed shrew specimens (*Blarina*). The measurements are: TL = total length of specimen taken from tip of nose to the end of tail, TV = tail vertebrae length taken from the sacral/caudal vertebrae junction to the last caudal vertebrae, HF = hind foot taken as greatest distance from calcaneus to the distal phalange excluding the toe nail, E = ear length taken as greatest distance from the ear attachment to the tip of the pinna. Each point represents a voucher specimen from the University of North Carolina Wilmington where not every specimen had all measurements recorded, thus the difference between the amount of points on plot.

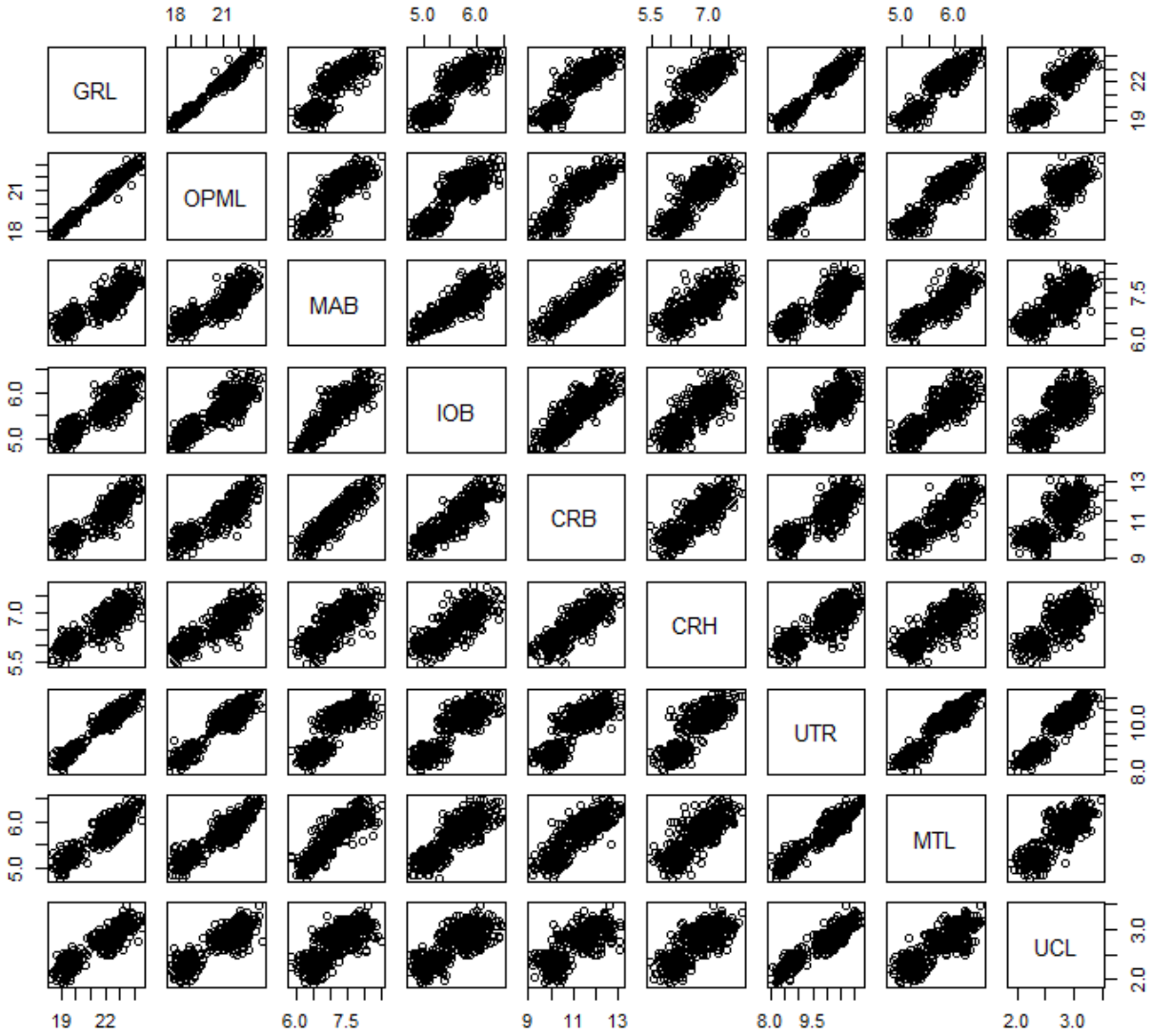


Figure AIII-2.2: Bivariate plot with 'pairs' function in R statistical software showing the combinations of cranial measurements from short-tailed shrew specimens (*Blarina*).

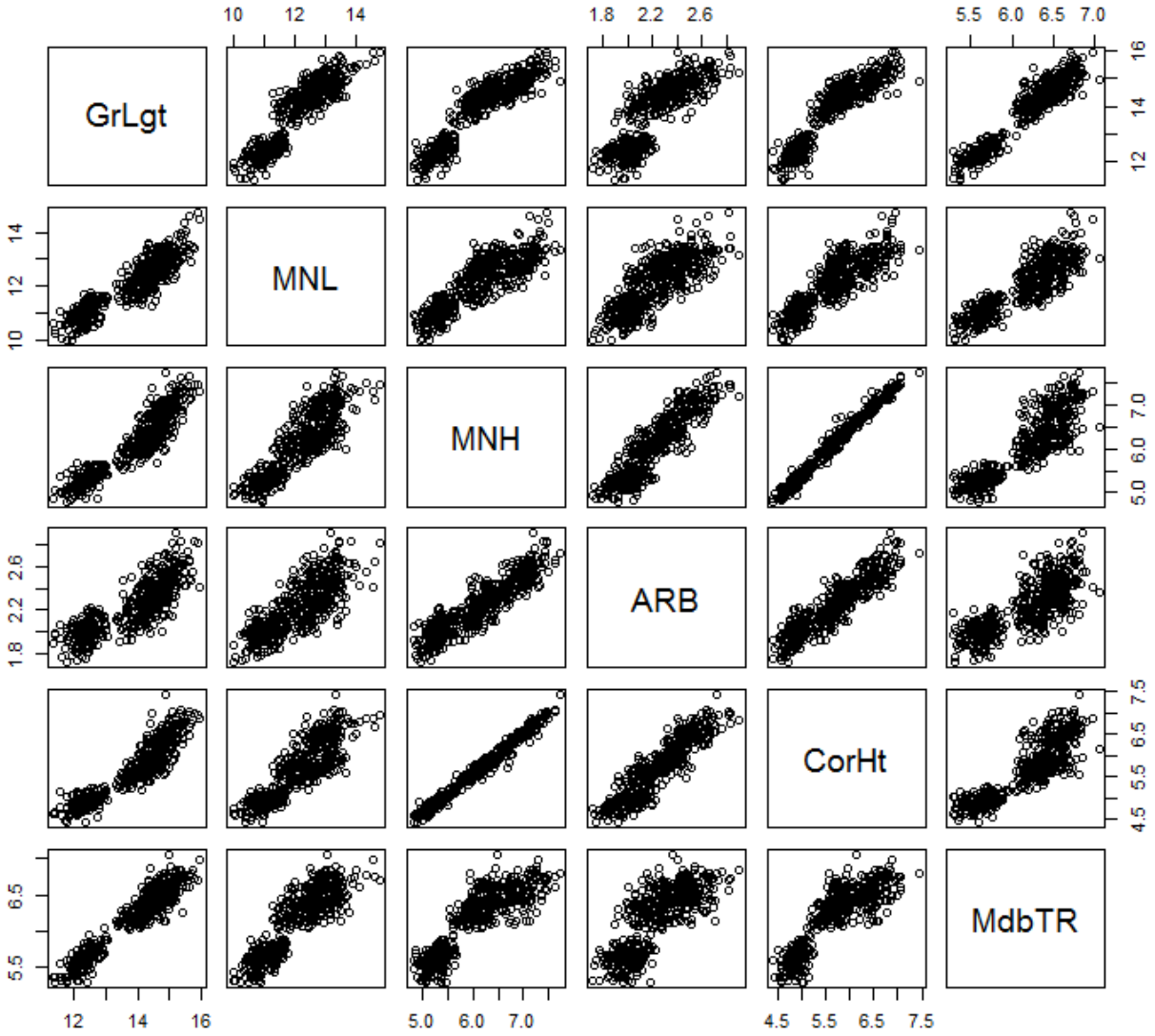
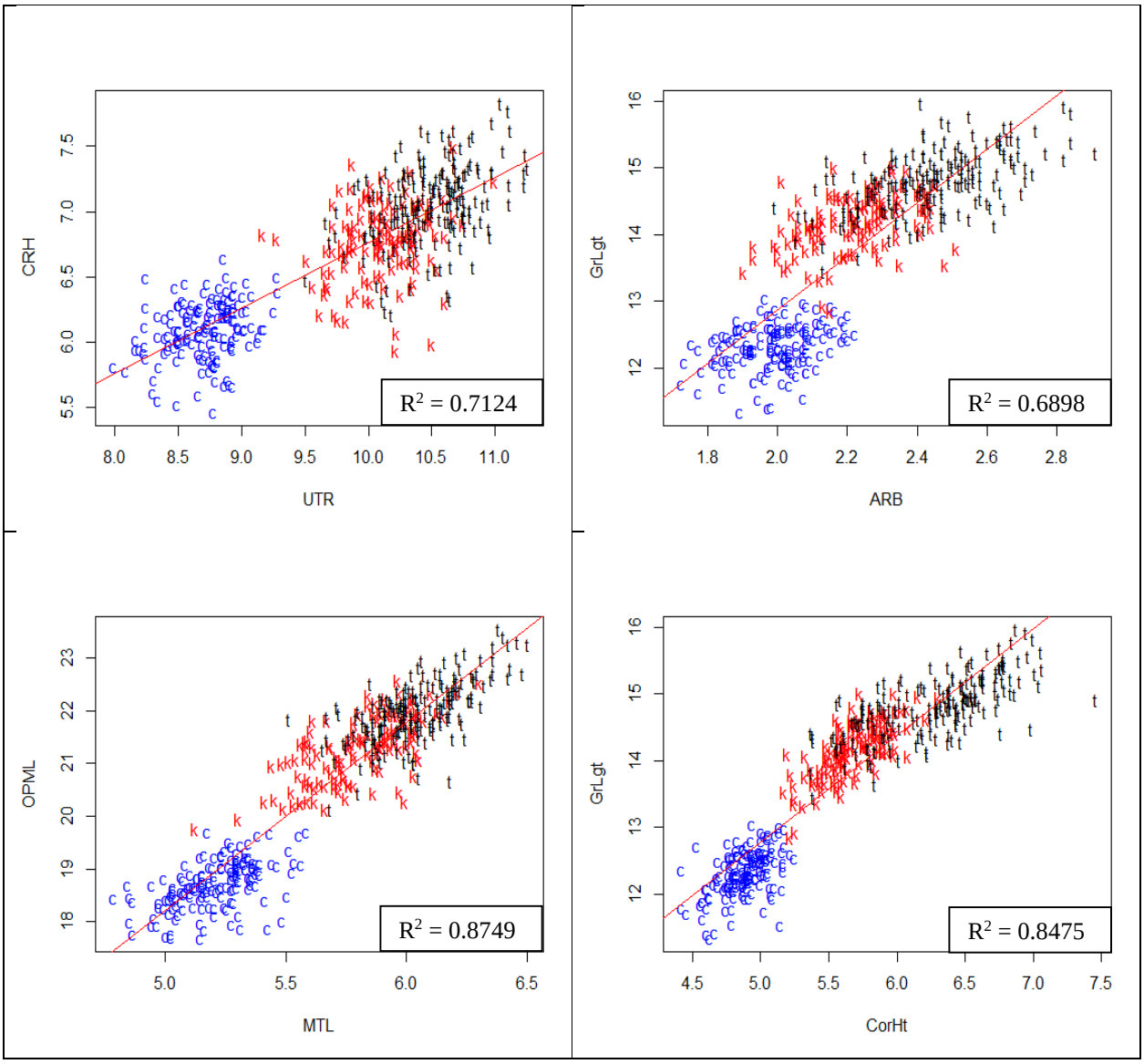


Figure AIII-2.3: Bivariate plot with 'pairs' function in R statistical software showing the combinations of mandibular measurements from short-tailed shrew specimens (*Blarina*).



SKULL MANDIBLE

Figure AIII-2.4: Additional bivariate plots of measurements taken from the skulls and mandibles of short-tailed shrew specimens (genus *Blarina*). The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. The plots given as x-axis vs y-axis for the skull are: TOP LEFT = upper tooth row including incisor (UTR) vs. cranial height (CRH), BOTTOM LEFT = length of molariform toothrow (MTL) vs. occipital-premaxilla length without incisor OPML; and for the mandible are TOP RIGHT = articular breadth (ARB) vs. greatest length of mandible with incisor (GrLgt), BOTTOM RIGHT = coronoid process height (CorHt) vs. GrLgt. The regression line and multiple R-squared (R^2) or coefficient of determination value gives information about how well the measure of one character can predict the measure of another character.

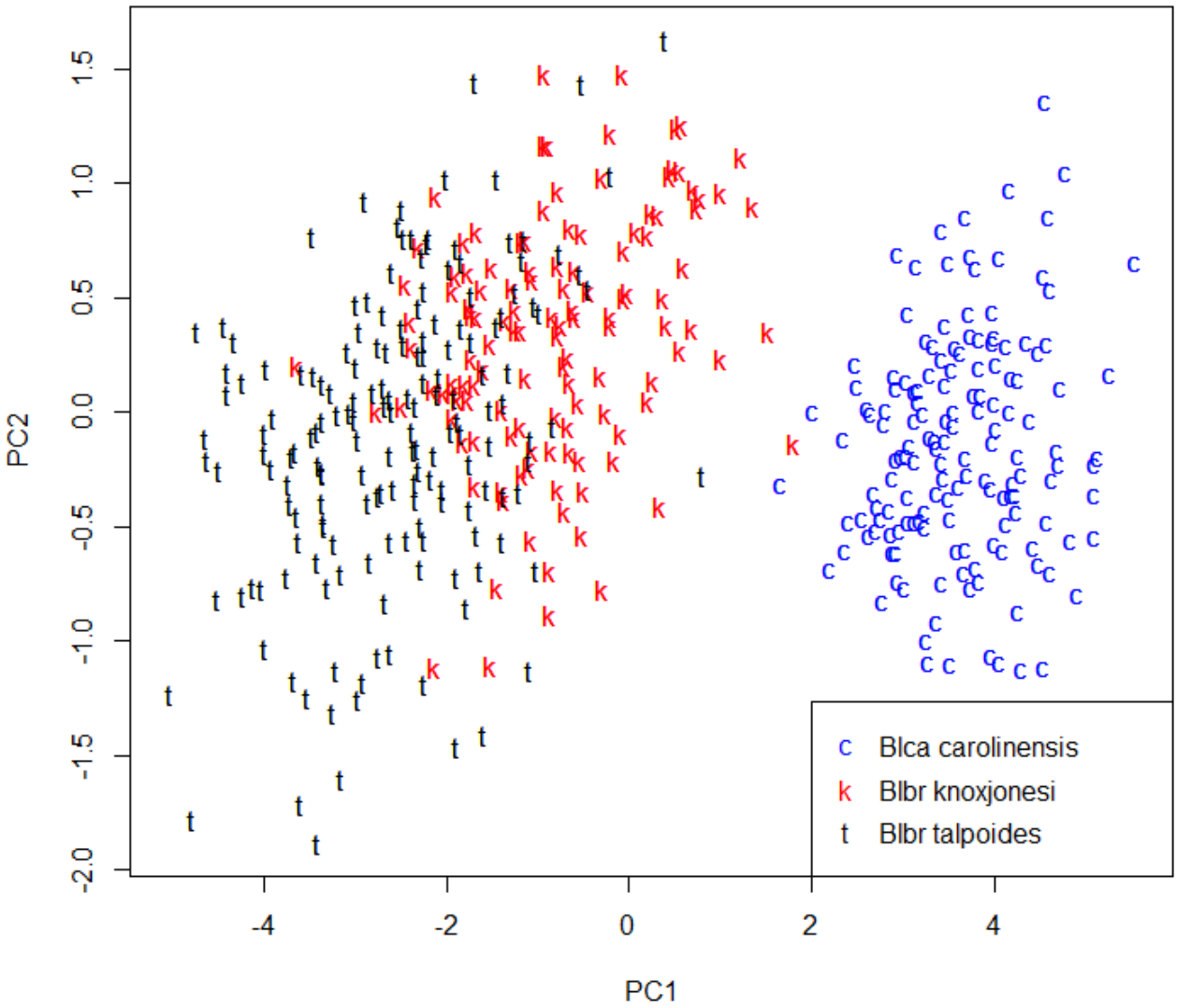


Figure AIII-2.5: Scatter plot of the first two principal components derived from the correlation matrix for the *Blarina* skull data. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. The first principal component (PC1) accounts for 89.2% of the total sample variance, while the second principal component (PC2) accounts for 3.9%, for a combination of 93.1% of the total sample variance explained by PC1 and PC2.

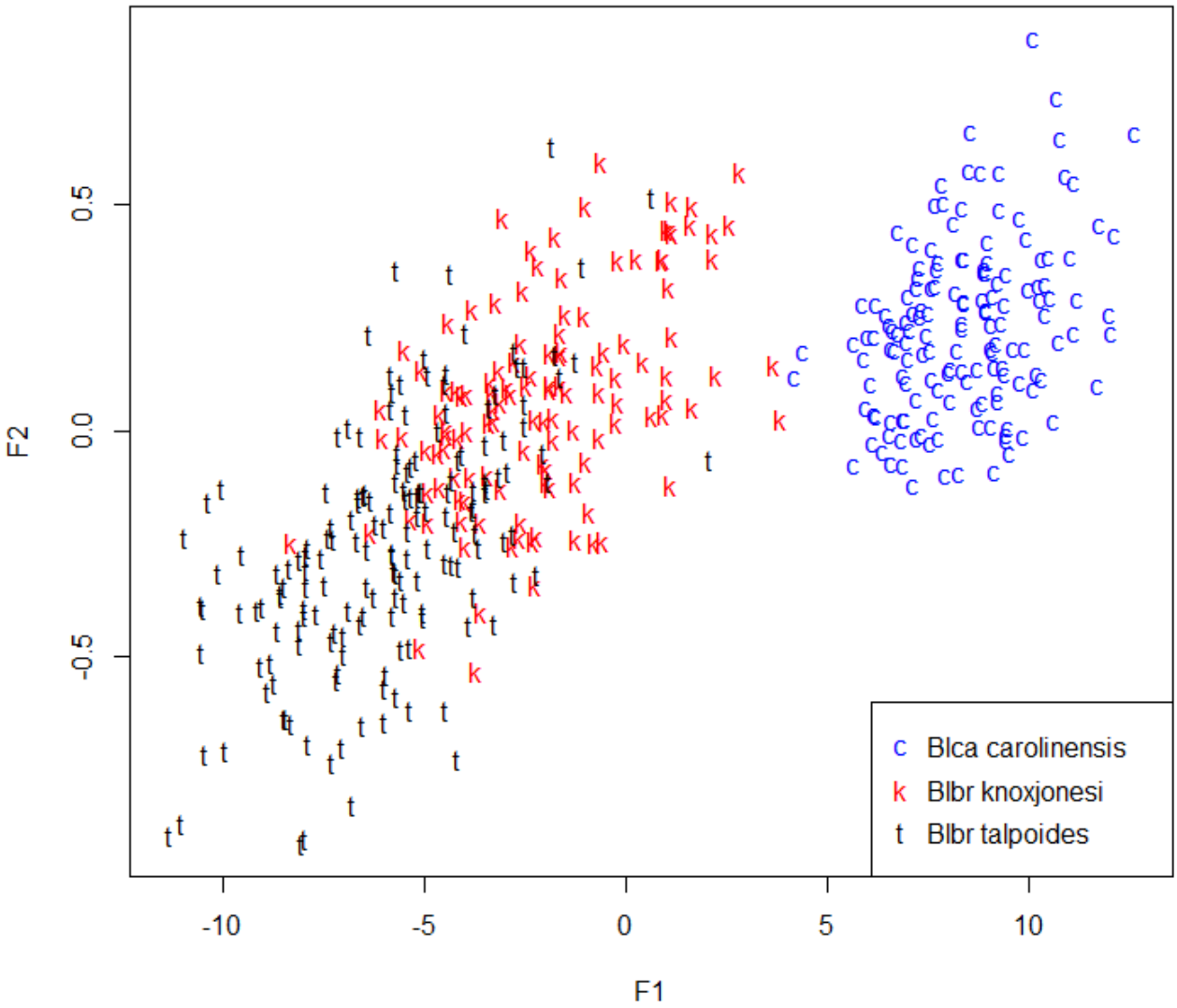


Figure AIII-2.6: Scatter plot of the first two factor loadings of the principal components factor analysis from the covariance *Blarina* skull data. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. Factor 1 (F1) explains 79.9% of the sample variance and Factor 2 (F2) explains 2.0% for a combined 81.9% of the total proportion of the sample variance explained.

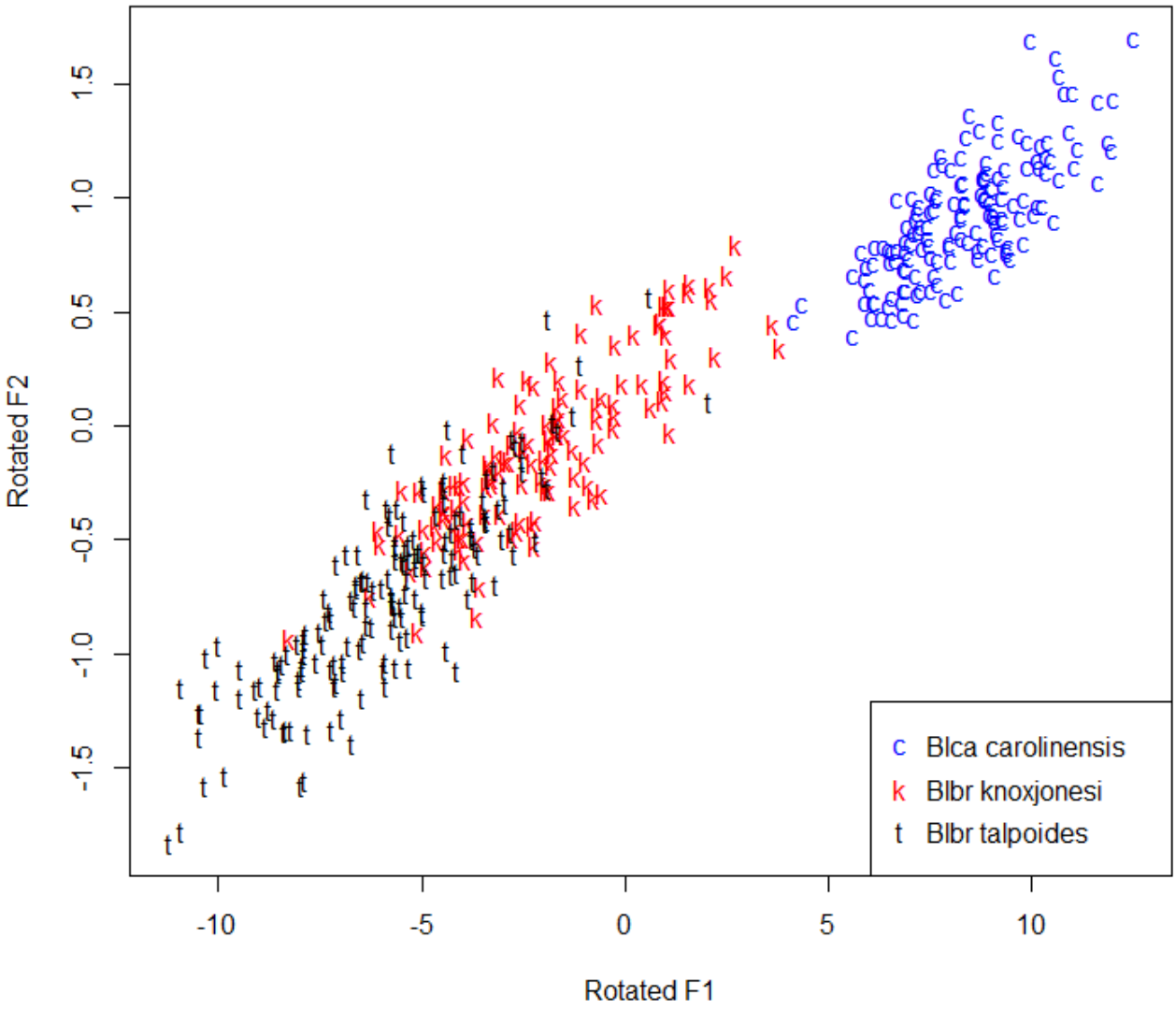


Figure AIII-2.7: Scatter plot of the first two factor loadings of the rotated principal components factor analysis from the covariance *Blarina* skull data. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. Rotated factor 1 (F1) explains 79.4% of the sample variance and the rotated factor 2 (F2) explains 2.5% for a combined 81.9% of the total proportion of the sample variance explained.

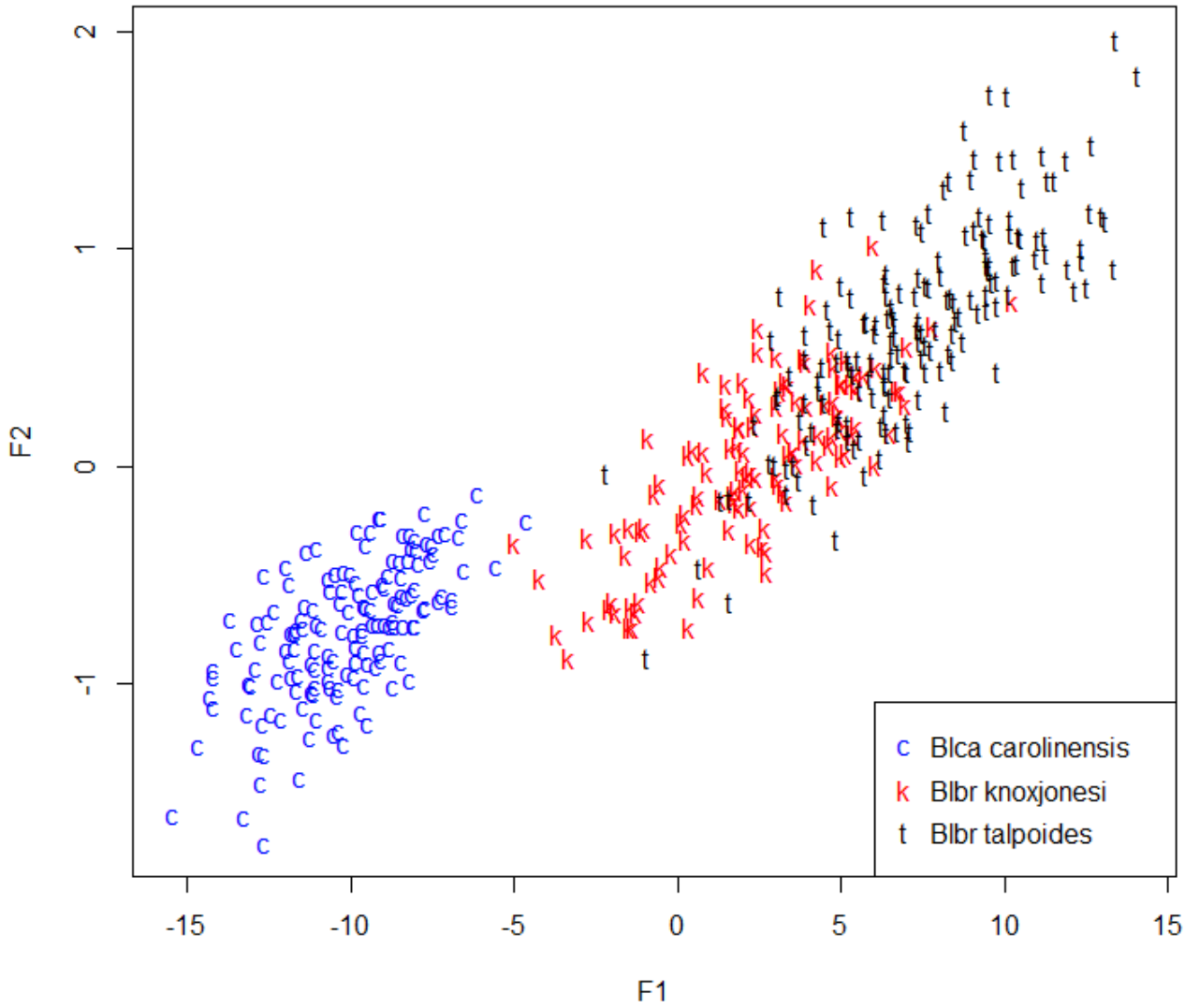


Figure AIII-2.8: Scatter plot of the first two factor loadings of the maximum likelihood factor analysis from the covariance *Blarina* skull data. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. Factor 1 (F1) explains 87.5% of the sample variance and Factor 2 (F2) explains 3.5% for a combined 91.0% of the total proportion of the sample variance explained.

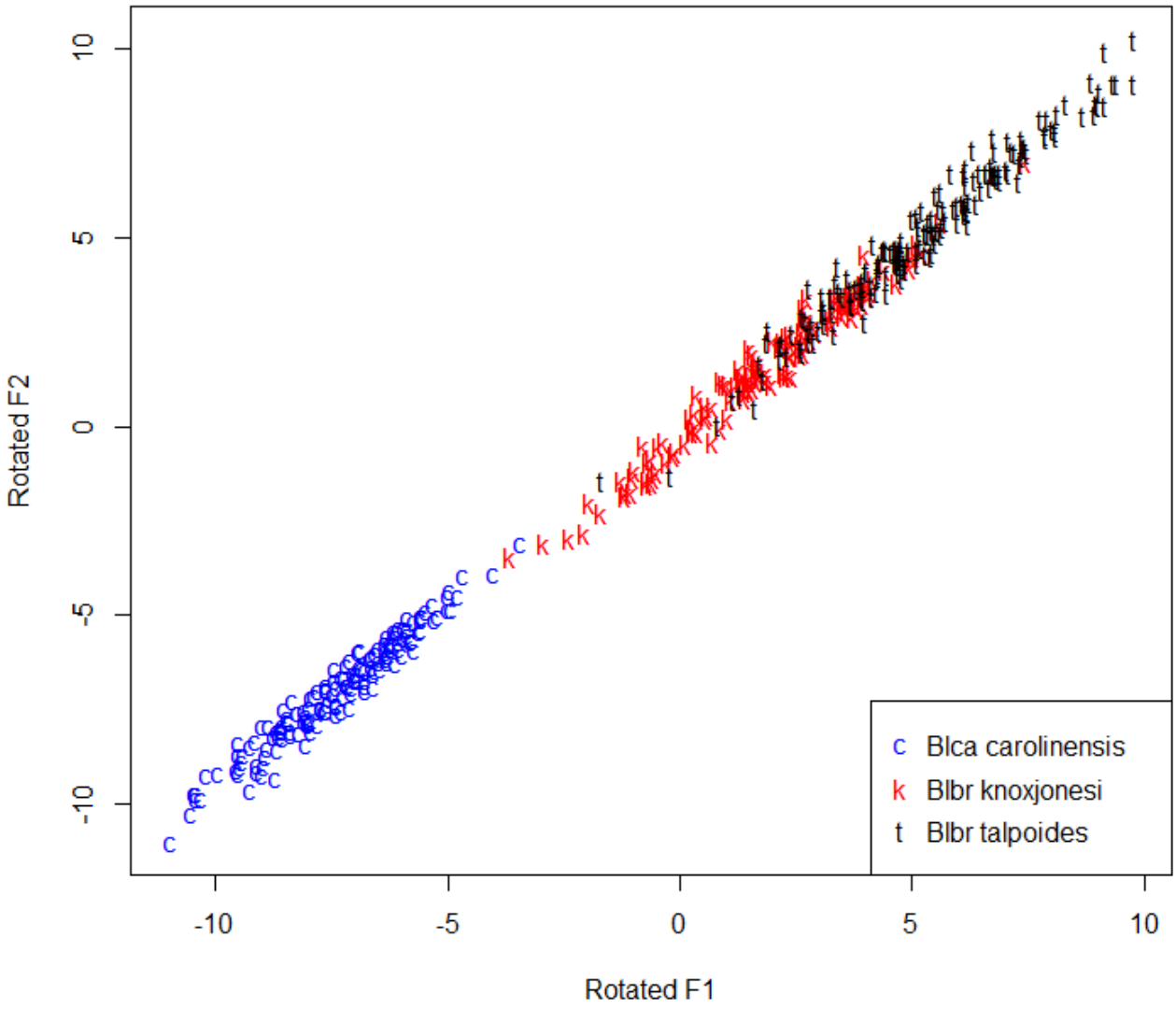


Figure AIII-2.9: Scatter plot of the first two factor loadings of the rotated maximum likelihood factor analysis from the covariance *Blarina* skull data. The points on the plot are: c = *Blarina carolinensis carolinensis*, k = *B. brevicauda knoxjonesi*, t = *B. b. talpoides*, and are colored blue, red and black respectively. Rotated factor 1 (F1) explains 47.2% of the sample variance and the rotated factor 2 (F2) explains 43.8% for a combined 91.0% of the total proportion of the sample variance explained.

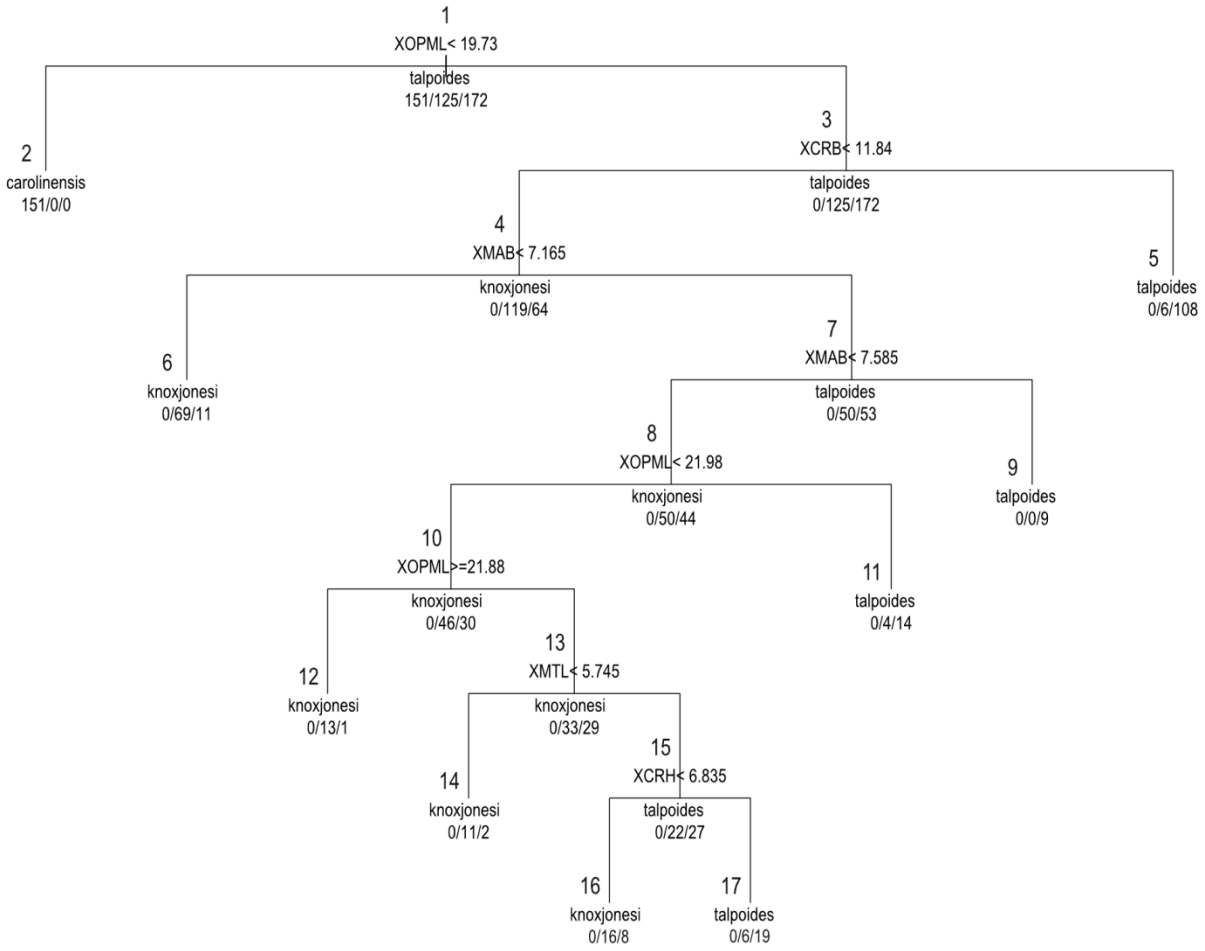


Figure AIII-2.10: Classification tree with information splitting index for *Blarina* skull data. Node numbers 1 – 17 correspond to details in Table AIII-1.4.

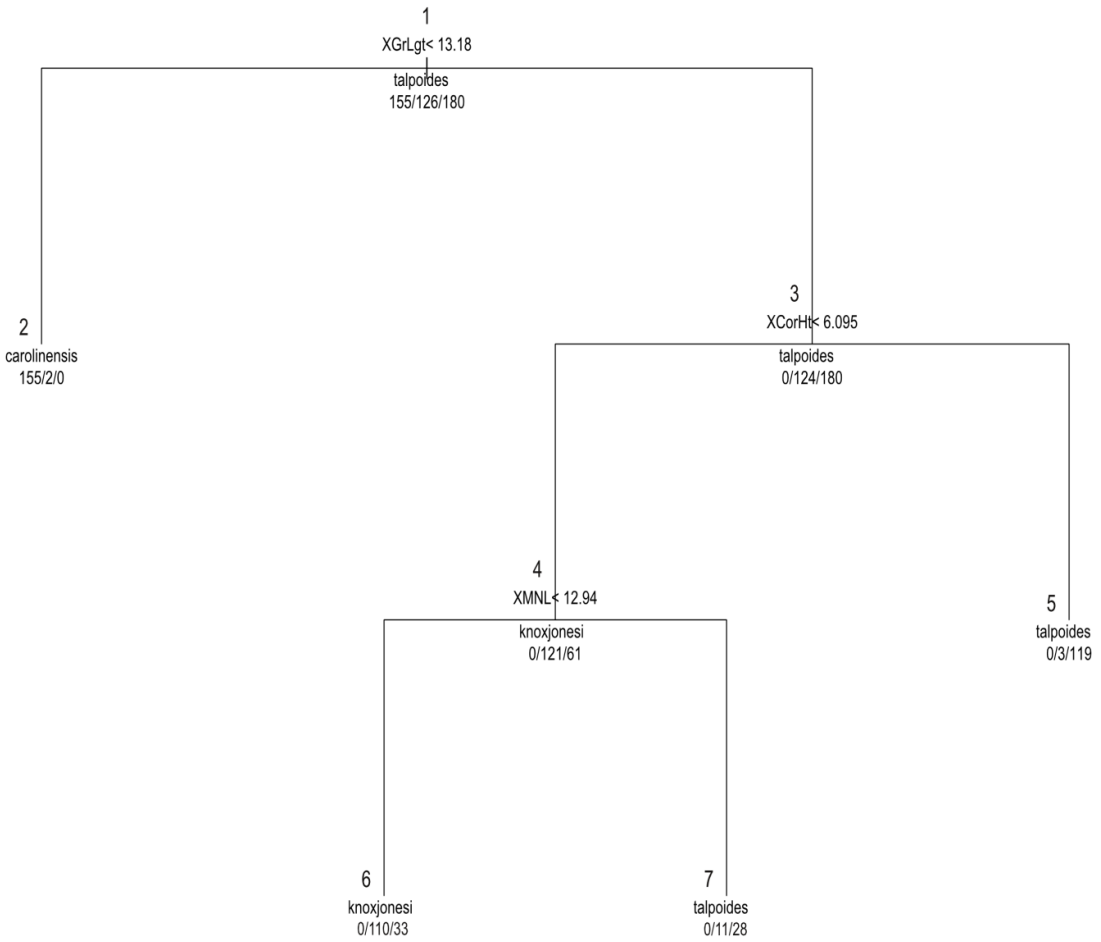


Figure AIII-2.11: Classification tree with information splitting index for *Blarina* mandible data. Node numbers 1 – 7 correspond to details in Table AIII-1.5.

APPENDIX IV - SPECIMENS EXAMINED FOR GENETIC ANALYSES

Specimens examined - The 118 specimens examined in this study are listed below by museum acronym (Hafner et al., 1997). All samples are from the North Carolina State Museum of Natural Sciences (NCSM - now called the North Carolina Museum of Natural Sciences). All localities are in the United States and from Alabama, New York, North Carolina or Virginia. Locality within county and sample number in parentheses is provided for each species indicated. All tissue samples used have associated skin and/or skeletal voucher material. Additional cytochrome-b DNA sequences from GenBank listed below with accession number and locality. Each sample represents a voucher specimen listed by county within subspecies designation (Webster et al., 2011.)

Blarina brevicauda cumberlandensis – total 10: ALABAMA: *Jackson Co*.: 0.5 mi N Money Hollow at Jack Gap (NCSM 13809, 13835, 13839, 13843, 14383), SW slope of Hogyard Ridge and N of Miller Mtn (NCSM 13838, 13851, 13869, 14394, 14397).

Blarina brevicauda knoxjonesi – total 1: NORTH CAROLINA: *Onslow Co*.: Holly Ridge at Sandy Run Swamp (NCSM 16801).

Blarina brevicauda talpoides – total 49: NORTH CAROLINA: *Ashe Co*.: Fleetwood at US 221/West Pine Swamp Rd JCT (NCSM 17599). *Avery Co*.: Newland at Kentucky Creek (NCSM 14638, 14640), Pineola Bog SNA (NCSM 14384, 14385, 14387, 14405), Banner Elk at Sugar Mtn Bog SNA (NCSM 14386). *Buncombe Co*.: Asheville at Turtle Creek Dr (NCSM 17605, 17606), Swannanoa at Four-Brothers Knob (NCSM 14400), Swannanoa at Jones Mtn Boulderfield (NCSM 14390, 14399), Swannanoa at Jones Mtn Spring (NCSM 14391-14393). *Clay Co*.: Hayesville at Leatherwood Branch (NCSM 13770, 13771). *Graham Co*.: Robbinsville at Sand Creek (NCSM 14639). *Haywood Co*.: Clyde at Harmon Den WMA (NCSM 17616), Clyde at Toms Branch Tributary (NCSM 15607). *Henderson Co*.: Gerton at Little Bearwallow Mtn (NCSM 17602-17604). *Madison Co*.: Marshall at Whiterock Cliff (NCSM 14380), Hot Springs at Max Patch Mtn (NCSM 15608, 17608, 17609). *Polk Co*.: Saluda at Green River Gamelands (NCSM 14406-14414, 14642, 14939, 15609, 17596-17598). *Rutherford Co*.: Mill Spring at Cane Creek Mtn (NCSM 17607). *Watauga Co*.: Banner Elk at Bear Paw SNA (NCSM 17595). *Yancey Co*.: Burnsville at Balsam Cone (NCSM 14404), Burnsville at Mt Mitchell SP (NCSM 14401). NEW YORK: Suffolk Co.: East Hampton (NCSM 17617). VIRGINIA: Arlington Co.: Arlington at Bluemont Junction Park (NCSM 17618).

Blarina carolinensis carolinensis – total 50: NORTH CAROLINA: *Burke Co*.: 4.5 km WNW Morganton (NCSM 16800). *Durham Co*.: Bahama (NCSM 17591). *Gaston Co*.: Belmont (NCSM 17592). *Onslow Co*.: Richlands at Hofmann SF (NCSM 17593). *Scotland Co*.: 4 mi E Marston at Sandhills Gameland (NCSM 16786-16788). *Tyrrell Co*.: Columbia at Scuppernong River SNA (NCSM 16772-16777), Creswell at Scuppernong River SNA (NCSM 16778-16784). *Wake Co*.: 3.4 km SSW Cary (NCSM 16798, 16799), Raleigh (NCSM 16902, 17581), 7.4 km E Raleigh (NCSM 15601), 5.3 km SSW Raleigh (NCSM 15604), 12.2 km N Raleigh (NCSM

16791-16796, 17589, 17590), 7 km WSW Raleigh at Lake Johnson Nature Park (NCSM 17588), 7.9 km NW Raleigh at Prairie Ridge Ecostation (NCSM 15606, 17582-17587), 8.9 km NW Raleigh at NCSU Horse Farm (NCSM 15606), 4.7 km NW Raleigh (NCSM 15603), 17.8 km NNE Raleigh (NCSM 16797), Raleigh at W. B. Umstead SP (NCSM 16789, 16790). *Washington Co*.: Creswell at Pettigrew SP (NCSM 15605), Creswell at Scuppernong River SNA (NCSM 16785). State only indicated, but possibly Raleigh (NCSM 16903).

Condylura cristata parva – total 2: NORTH CAROLINA: *Bladen Co*.: White Oak at Jones Lake SP (NCSM 14636). *Haywood Co*.: Clyde at Toms Branch Tributary (NCSM 14655).

Cryptotis parva parva – total 2: NORTH CAROLINA: *Polk Co*.: Saluda at Green River Gamelands (NCSM 14631). *Scotland Co*.: 4 mi E Marston at Sandhills Game Land (NCSM 16770).

Sorex cinereus cinereus – total 2: NORTH CAROLINA: *Buncombe Co*.: Swannanoa at Jones Mtn Boulderfield (NCSM 14538). *Yancey Co*.: Burnsville at Commissary Ridge east (NCSM 14540).

Sorex longirostris longirostris – total 2: NORTH CAROLINA: *Bladen Co*.: White Oak at Jones Lake SP (NCSM 14587). *Polk Co*.: Saluda at Green River Gamelands (NCSM 14592).

Cytochrome-B sequences downloaded from GenBank – Locality given with GenBank Accession number in parentheses.

Blarina brevicauda – total 10: IOWA: *Muscatine Co*.: Durant (AF533609). KENTUCKY: *Trigg Co*., Cadiz (AF533632, AF534121). NEW YORK: *Hamilton Co*., Long Lake (AF534115). PENNSYLVANIA: *Lawrence Co*.: New Galilee (AF533619). TENNESSEE: *Monroe Co*.: Vonore (AF533641, AF534123). VIRGINIA: *James City Co*.: Williamsburg (AF534116). WEST VIRGINIA: *Wayne Co*.: Wayne (AF534117). WISCONSIN: *Columbia Co*.: Portage (AF533642).

Blarina carolinensis – total 5: ARKANSAS: *Polk Co*.: Ouachita NF (AF395456). FLORIDA: *Highlands Co*.: Venus (AF395454). GEORGIA: no county or specific locality given (AF395449). ILLINOIS: *Jackson Co*.: no specific locality given (AF395458, AF395460).

Blarina hylophaga – total 5: TEXAS: *Aransas Co*.: Aransas National Wildlife Refuge (AY546677). *Bastrop Co*.: 5 mi N of Bastrop (AY546659, AY546662, AY546669, AY546681)

LITERATURE CITED

Hafner, M. S., W. L. Gannon, J. Salazar-Bravo and S. T. Alvarez-Castañeda. 1997. Mammal collections in the Western Hemisphere. American Society of Mammalogists. Allen Press, Lawrence, Kansas. 93 pp.

Webster, W. D., N. D. Moncrief, J. R. Choate and H. H. Genoways. 2011. Systematic revisions of the northern short-tailed shrew, *Blarina brevicauda* (Say). Virginia Museum of Natural History. Memoir 10. 77 pp.