CURRICULUM VITAE

John David McVay

ACADEMIC DEGREES

2008-	Louisiana State University	PhD, Biology
2007	Texas Tech University	MS, Biology
	Thesis title: Population genetic and phylogeographic analysis of Nerodia erythrogaster *in Texas, using mitochondrial data.*	
2001	The University of Texas at Austin	BS, Zoology

RELEVANT PROFESSIONAL EXPERIENCE

2007-2008	Research Associate, Museum of Natural Science LSU
2002-2007	Graduate Teaching Assistant, Texas Tech University
2006	Laboratory Technician, Texas Tech University
1999	Curatorial Assistant (Herpetology and Ichthyology), Texas Natural History Collection, The University of Texas at Austin

AWARDS

2008	Graduate Student Enhancement, LSU Biological Sciences
2003	Summer Research Award, Department of Biological Sciences
2002	SBC Chancellor's Fellowship, Texas Tech University

PUBLICATIONS

Makowski, R., Marshall, J., McVay, J. D., Rissler, L. J. Phylogeographic analysis, subspecies tests, and ecological niche modeling in the plain-bellied watersnake, *Nerodia erythrogaster*. *Accepted: Molecular Phylogenetics and Evolution.*

Austin. C. C., Spataro, M., Peterson, S. N., Jordan, J. L., and McVay, J. D. 2009. Conservation genetics of Boelen's python (*Morelia boeleni*) from New Guinea: reduced genetic diversity and divergence of captive and wild animals. *Conservation Genetics:* Online.

Siler, C. D., McVay, J. D., Diesmos, A. C., and Brown, R. M. 2009. A New Species of Fanged Frog, Genus *Limnonectes* (Amphibia: Anura: Dicroglossidae) from Southeast Mindanao Island, Philippines. *Herpetologica* 65(1): 105-114.

Williams, D. J., O'Shea, M., Daguerre, R. L., Pook, C. E., W. Wuster, W., Hayden, C. J., McVay, J. D., Paiva, O., Matainaho, T., Winkel, K. D., and Austin, C. C. 2008. Origin of the eastern brown snake, *Pseudonaja textilis* (Duméril, Bibron and Duméril)(Serpentes: Elapidae: Hydrophiinae) in New Guinea: evidence of multiple dispersals from Australia, and comments on the status of *Pseudonaja textilis pughi* Hoser 2003. *Zootaxa* 1703: 47-61.

McVay, J. D., Rodriguez, D., Rainwater, T. R., Ray, D. A., Dever, J. A., Platt, S. G., McMurry, S. T., Forstner, M. R. J., and Densmore, L. D. 2008. Evidence of multiple paternity in Morelet's Crocodile (Crocodylus moreletii) in Belize, CA, inferred from microsatellite markers. Journal of Experimental Zoology Part A: Ecological Genetics and Physiology 309A(10): 643-648.

Using transcriptomics to test hypothesis of selection and dispersal across hybrid zones

John D. McVay, Louisiana State University Department of Biological Sciences

I am requesting money to sequence the transcriptome of three species of water snake (genus *Nerodia*) to study the hybridization that occurs among them.

Background

Recently, it has become increasingly clear that phylogeny is affected by processes occurring at the population level, and incorporation of our understanding of the latter into analysis of the former is crucial [1-3]. Understanding the origin and maintenance of genotypic diversity in the face of gene flow is important to understanding diversification and divergence. Gene flow may occur between two diverging populations, or may occur as a result of secondary contact between two more distantly related species. Hybridization, the latter mechanism, may lead to fixation of alleles (due to drift or selection) that have crossed species boundaries through introgression. Understanding the nature of this introgression and fixation becomes exceedingly important when estimating molecular phylogenies, as topologies of individual gene trees may be influenced by past introgression events, and may bias phylogenetic estimation, especially when genes are concatenated. Past molecular-based investigation of hybrid zone genetics has been conducted using both DNA sequence and fragment data, though both methods have been limited in scale. With the advent of "next-generation" sequencing, there is now opportunity to investigation hybrid zone dynamics by sampling on a genomic scale. I propose to investigate the population dynamics and phenotypic evolution in a hybrid zone where differential selection across an environmental gradient is though to maintain species boundaries; my approach will incorporate massively-parallel sequencing of whole transcriptomes.

Within the North American snake tribe Thamnophiini (9 genera, ~65 species), morphological and molecular evidence has pointed to at least three instances of ongoing hybridization between species [4, 5], some of which occurs between non-sister species [6]. In at least one instance in watersnakes (genus *Nerodia*), hybridization occurs between a freshwater species (*N. fasciata*) and its saltwater-tolerant sister taxon, *N. clarkii* [7]. Further inland, *N. fasciata* is known to hybridize with a more distantly related freshwater congener, the Northern Water snake (*N. sipedon*). Evidence of hybridization (phenotypic intermediates) occurs throughout the contact zones, along the Gulf Coast from Texas to Florida and to the southern edge of the Atlantic coast in Georgia for the former species pair, and within a large contact zone in the Southeastern U.S. for the latter (figure 1). Introgression of the inland species towards is the coast is putatively limited by the inability to tolerate saline environments [8], however the coastal snake appears to be euryhaline, and may even prefer freshwater to seawater [9]. Further, the mechanisms for salt tolerance is this species are poorly understood (no known salt glands, no appreciable difference in kidney function [10]), though it may be based in olfaction or behavior [8]. Better understanding of the molecular underpinnings of hybridization within this species group may lead to a better understanding of speciation and divergence, and with it more robust models for phylogenetic estimation. The study of this system is particularly

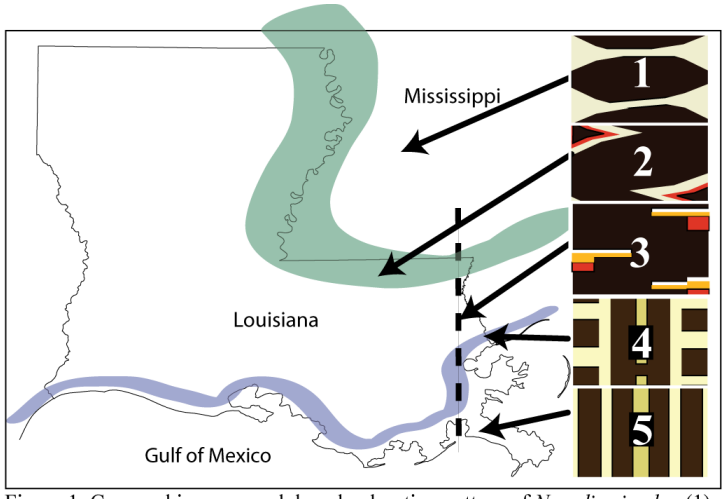

Figure 1. Geographic range and dorsal coloration pattern of *Nerodia sipedon* (1), *N. sipedon* x *fasciata* (2), *N. fasciata* (3), *N. fasciata* x *clarkii* (4), and *N. clarkii* (5). Dashed line represents proposed collection transect.

intriguing, in that it offers an opportunity to study related hybrid zones from two perspectives: one that addresses hypotheses concerning selection across an environmental gradient, and a "control" hybrid zone, where no such drastic gradient is known; and a second approach that compares hybrid zone dynamics between sister and non-sister pairs of species.

Goals

The goals of this project are to better understand the population dynamics in the hybrid zone between these three species of *Nerodia*, and to evaluate the possible mechanisms of salinity tolerance and phenotypic evolution in this group. The ancillary goal of this project is to develop transcriptome libraries from these three species for use in this and other systems.

Methods

Sample collection.—Starting at the coast of the Gulf of Mexico, I will collect individuals of *Nerodia clarkii*, *N. fasciata*, and *N. sipedon* along a longitudinal transect in Louisiana and southern Mississippi. This transect will cross the range boundary of all species, and will include areas where hybrids (based on phenotypic designation) are known to occur. I will initially collect a single female from each region of the transect where only one of the species occurs, for construction of the transcriptome library.

Data collection.— I propose to collect whole transcriptomes from a female (the heterogametic sex) of *N. clarkii*, *N. fasciata*, and *N. sipedon*, by extracting RNA from a multitude of tissues (liver, kidney, integument, brain, salivary gland, Jacobson's organ, heart, muscle, gut) and building a normalized cDNA library. Each species' library will be sequenced individually an in-house Genome Analyzer II_X (Illumina Inc., San Diego, CA), using paired-end reads; the collected data will be ~10Gb per species. Stepwise validation of the library will be conducted with vector cloning. Following data collection, I will align fragments from each species separately, using Velvet [11], then these alignments will be secondarily aligned together. Areas of overlap will be screened for single nucleotide polymorphisms (SNPs) in an automated fashion, and selected SNPs will be used for microarray chip construction. The remaining individuals collected along the transect will be scored for SNPs using the microarray chip.

Data analysis.—I will use NewHybrids [12] to assign individuals to classes of pure or hybrid origin based on SNP signature; these data will be compared qualitatively and quantitatively with body coloration and head shape, the latter has been tied to feeding behavior which differs between species [13, 14]. Hypotheses concerning selection on individual SNPs and fitness of hybrid classes will be tested using simulation approaches.

Advancement in the assembly and annotation of the *Anolis carolinensis* genome may prove to be extremely useful in the coming months, particularly in assessment of sex-biases in dispersal and integration across the hybrid zone.

Significance

It is anticipated that the outcome of this study will help us better understand hybrid zone and dispersal dynamics and the genetic basis of salt tolerance in this group of species, and that this knowledge can lead to insights in mechanisms of diversification and divergence within Thamnophiini. Additionally, we expect that an equally important outcome of this research is the development a vast library of genetic markers that can be used to address a variety of questions in this tribe, from intrapopulation- to intergeneric-level studies. The SNP chips technology may also prove useful in assessment of status and implementation of conservation techniques in populations of *Nerodia clarkii* and other *Nerodia* species considered to need protection [15, 16]. Ultimately, we aim to demonstrate that next-generation sequencing technology will transform research in non-model systems, allowing researchers to take advantage of natural experiments for a better understanding of evolutionary mechanisms.

Timeline

Collection of individuals for library construction will take place in late Spring, 2010, and collection of individuals along the transect throughout the following Summer and Fall. Time will be split in the Summer between field collection and library construction; this can be accomplished as all field sites are no more than three hours drive from the laboratory. Anticipated completion date of data collection will be December 2010, and completion of data analyses June 2011.

Literature Cited

1. Maddison, W. Syst Biol, 1997. **46**: 523-536.
2. Carstens, B.C. and L.L. Knowles. Syst Biol, 2007. **56**: 400-11.
3. Edwards, S.V. Evolution, 2009. **63**: 1-19.
4. Mebert, K. Mol Ecol, 2008. **17**: 1918-1929.
5. Fitzpatrick, B.M., et al. Mol Ecol, 2008. **17**: 4107-17.
6. Rossman, D.A., N.B. Ford, and R.A. Seigel, *The garter snakes: evolution and ecology*. 1996. Norman: University of Oklahoma Press.
7. Lawson, R., et al. Copeia, 1991. **1991**: 638-659.
8. Dunson, W.A. Am J Physiol, 1978. **235**: R151-9.
9. Zug, D.A. and W.A. Dunson. Florida Scientist, 1979. **42**: 1-8.
10. Pettus, D. Copeia, 1958. **1958**: 207-211.
11. Zerbino, D.R. and E. Birney. Genome Res, 2008. **18**: 821-9.
12. Anderson, E.C. and E.A. Thompson. Genetics, 2002. **160**: 1217-29.
13. Vincent, S.E., et al. J Evol Biol, 2009. **22**: 1203-11.
14. Vincent, S.E., et al. J Evol Biol, 2006. **19**: 1545-54.
15. Jansen, K.P., H.R. Mushinsky, and S.A. Karl. Conserv Genet, 2008. **9**: 410-410.
16. Dodd, C.K., Jr., *Status, conservation, and management.*, in *Snakes: ecology and evolutionary biology*, R.A. Seigel, J.T. Collins, and S.S. Novak, Editors. 1987. New York: MacMillan Publ. Co.

Budget – Library construction

Expense	Cost	Amount Requested from SSB
Normalized cDNA library construction -	$950	$950
Paired end Sample preparation -	$980	$980
Cluster generation -	$2875	-
Illumina Sequencing -	$3600	-
Total	$8405	$1930

The budget above constitutes the cost of reagents and kits for cDNA library construction and sequencing from 3/4 of a flow cell (~30Gb). Sequencing will be performed at cost on an in-house Genome Analyzer®. While the total cost is significant, the amount of data collected in a single run is immense, and the relative cost per base pair is very low: ~3-4Mb/dollar (~500b/dollar for conventional Sanger sequencing). I am requesting funding from SSB for this portion of my research, because I feel that the construction of this library will prove to be useful in a variety of future systematic studies. The remainder of the cost will be paid through other currently available funds. Funding for other portions of this project (field work, SNP chip construction) have been requested from other sources.