

# Contents

1

<b>Table of Contents</b> . . . . .	<b>1</b>	2
<b>List of Figures</b> . . . . .	<b>2</b>	3
<b>List of Tables</b> . . . . .	<b>3</b>	4
<b>1 Phylogeography</b> . . . . .	<b>4</b>	5
1.1 Introduction . . . . .	4	6
1.2 Methods . . . . .	6	7
1.2.1 Sampling and DNA Isolation . . . . .	6	8
1.2.2 RADseq Library Preparation . . . . .	6	9
1.2.3 Phylogenetic Data Processing . . . . .	7	10
1.2.4 Maximum Likelihood . . . . .	7	11
1.2.5 Multispecies Coalescent . . . . .	8	12
1.2.6 Test for Historic Admixture . . . . .	8	13
1.2.7 Population Structure Data Processing . . . . .	8	14
1.2.8 Population Structure . . . . .	9	15
1.2.9 Recent <i>A. fowleri</i> x <i>A. woodhousii</i> hybridization . . . . .	9	16
1.3 Results . . . . .	9	17
1.3.1 Assembly and alignment with <i>ipyrad</i> . . . . .	9	18
1.3.2 Maximum Likelihood Phylogeny . . . . .	10	19
1.3.3 Coalescent Phylogeny . . . . .	10	20
1.3.4 Historic Introgression . . . . .	10	21
1.3.5 Population Structure . . . . .	11	22
1.3.6 Hybridization between <i>A. fowleri</i> and <i>A. woodhousii</i> . . . . .	11	23
1.4 Discussion . . . . .	11	24
1.4.1 Relationships . . . . .	11	25
1.4.2 Divergence Time . . . . .	11	26
1.4.3 Population Structure . . . . .	11	27
1.4.4 . . . . .	11	28
1.5 Figures . . . . .	12	29
1.6 Tables . . . . .	23	30

# List of Figures

31

1.1.	Distribution of <i>americanus</i> group samples	13	32
1.2.	Iqtree	14	33
1.3.	Iqtree	15	34
1.4.	Multispecies coalescent phylogeny	16	35
1.5.	<i>f</i> -branch statistics	17	36
1.6.	Americanus Population Structure	18	37
1.7.	Fowleri Population Structure	19	38
1.8.	Terrestris Population Structure	20	39
1.9.	Woodhousii Population Structure	21	40
1.10.	Fowleri Population Structure	22	41

# List of Tables

42

1.1 Samples used in this study . . . . .	24	43
--	----	----

# Chapter 1

44

## Phylogeography

45

### 1.1 Introduction

46

Many factors are understood to be important in driving and shaping the diversification and evolutionary history of organisms. Chief among them is the interplay between climatic conditions and geologic processes. Changes in these environmental variables can alter the distributions of organisms and result in changes in the connectivity of populations. Disconnected populations may undergo genetic divergence from one another due to adaptive evolution in response to changing abiotic or biotic conditions . Or they might simply diverge via neutral evolution driven by the effects of drift. Environmental changes can also reconnect previously isolated populations resulting in hybridization and gene flow, another very important process shaping patterns of diversity. Understanding the interplay of all of these factors is critical for understanding the evolutionary history of organisms.

47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57

The North American toads in the genus *Anaxyrus* are a group of organisms with a poorly understood evolutionary history. Although, not for lack of trying. Multiple studies of the evolutionary relationships among species in the genus have produced conflicting results (Fontenot et al., 2011; Graybeal, 1997; Masta et al., 2002; Portik et al., 2023; Pramuk et al., 2007; Pyron & Wiens, 2011). Particularly within the *A. americanus* group composed of *A. americanus*, *A. baxteri*, *A. fowleri*, *A. hemiophrys*, *A. houstonensis*, *A. microscaphus*, *A. terrestris*, and *A. woodhousii*. Some phylogenies inferred solely from mitochondrial data have even been inconsistent with the current taxonomy, with *A. fowleri* forming a polytomy (Fontenot et al., 2011; Masta et al., 2002). These conflicting results could be due to methodological differences such as the species included, the number of individuals of each species sequenced, inference methods used, or the sequenced loci. But the differences in inferred relationships could also result from real biological processes. Incomplete lineage sorting is one source of possible discordance between datasets in phylogenetic inference (Kubatko & Degnan, 2007). Gene flow via hybridization is another potential source of discordance (Degnan & Rosenberg, 2009).

58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72

While incomplete lineage sorting is very likely to have impacted patterns of genetic variation in *Anaxyrus*, gene flow due to hybridization is a distinct possibility as well. There are numerous reports of natural hybridization between several different species of *Anaxyrus* (Green, 1996). A study of allozyme variation across a hybrid zone between *A. americanus* and *A. hemiophrys* revealed introgression taking place across a more than 50km wide hybrid zone. Hybrid zones are also suspected to exist between *A. americanus* and *A. terrestris* and between *A. woodhousii* and *A. fowleri* (Green, 1996; Weatherby,

1982). Furthermore, numerous laboratory crosses have been performed between pairs of *Anaxyrus* species with currently overlapping distributions (Blair, 1963, 1972). Some of which produce viable and fertile backcross progeny (Blair, 1963, 1972). These studies suggest that gene flow could very easily have played a role in shaping patterns of diversity in *Anaxyrus*. However, they provide only a snapshot in time with no indication of the long term consequences. There are many potential consequences of hybridization such as adaptive introgression, introgression of neutral genetic variation, reinforcement, lineage fusion, polyploidization, hybrid speciation, or transition to unisexual reproduction (Abbott et al., 2013). Inference of past introgression is an important starting point for exploring these outcomes yet it remains a challenging problem. The network structure of phylogenetic networks are far less tractable to infer than the more simple bifurcating phylogenies for which there has been extensive method development. There has been some recent work to overcome this challenge as well as increased feasibility of obtaining appropriate genome wide datasets to investigate past gene flow.

Apart from the significant evolutionary implications of hybridization which need to be understood, it also presents a valuable opportunity for investigating the mechanisms that drive divergence and the evolution of reproductive incompatibility (Rieseberg et al., 1999). Many generations of backcrossing within hybrid zones can produce a large number of highly recombinant genomes that allow for the observation of many possible hybrid genotypes under natural conditions in order to identify advantageous or disadvantageous hybrid genotypes. In most species it is not feasible to produce such a large number of highly recombinant offspring in order to make such observations. The evolutionary history of hybridizing species is important context to have when studying patterns of introgression within hybrid zones. Context such as the phylogenetic relationship of hybridizing species relative to other closely related species, the amount of genetic differentiation, the time since divergence and by extension the biogeographic processes driving initial divergence and subsequent secondary contact in cases of allopatric divergence. This important context is currently missing for *Anaxyrus* which limits the inferences that can be made regarding hybrid zones in this genus.

Ultimately, environmental change is what leads to population divergence and subsequent or concurrent gene flow except in cases of vicariant dispersal. Therefore, environmental variables are also important context for making inferences from hybrid zones in addition to understanding the process of diversification more generally. North America has had a very complex geologic and climatic history (Lyman & Edwards, 2022). The effects of which are often clade specific (Nuñez et al., 2023). But large scale environmental changes can often impact multiple species simultaneously (Oaks, 2019). There has been recent development of methods to infer these events (Oaks, 2019; Oaks et al., 2022). Population structure...

In this study, I investigate the evolutionary history of North American toads in the genus *Anaxyrus* using genome wide sequence data. For this I obtained restriction enzyme-associated DNA sequence (RADseq) data from 12 species of *Anaxyrus* including dense sampling representing a large portion of the ranges of *A. americanus*, *A. fowleri*, *A. terrestris*, and *A. woodhousii*.

With these data I infer evolutionary relationships using different methods.

Evolutionary relationships are needed for inferring historic gene flow.

I test for shared divergence time which would suggest divergence is being driven by the same environmental changes. Or large scale biogeographic processes.

I test for past introgression between lineages.

I determine if there is population structure.

128

## 1.2 Methods

129

### 1.2.1 Sampling and DNA Isolation

130

I obtained tissue samples from museum tissue collections as well as from individuals I collected from 2017 to 2020. I selected samples to represent as much of the range of each species of *Anaxyrrus* as possible. I also included one *Rhinella marina* and one *Incilius nebulifer* for as outgroups for phylogenetic analyses.

131

I isolated DNA from tissues by first lysing a piece of tissue approximately the size of a grain of rice in 300  $\mu\text{L}$  of a solution of 10mM Tris-HCL, 10mM EDTA, 1% SDS (w/v), and nuclease free water along with 6 mg Proteinase K that was incubated for 4-16 hours at 55°C in a 1.5 mL microcentrifuge tube. To purify the DNA and separate it from the lysis product, I mixed the lysis product with a 2X volume of SPRI bead solution containing 1 mM EDTA, 10 mM Tris-HCl, 1 M NaCl, 0.275% Tween-20 (v/v), 18% PEG 8000 (w/v), 2% Sera-Mag SpeedBeads (GE Healthcare PN 65152105050250) (v/v), and nuclease free water. I then incubated the samples at room temperature for 5 minutes, placed the beads on a magnetic rack, and discarded the supernatant once the beads had collected on the side of the tube. I then performed two ethanol washes by adding 1 mL of 70% ETOH to the beads while still placed in the magnet stand and allowing it to stand for 5 minutes before discarding the ethanol. After removing all ethanol from the second wash, I removed the tube from the magnet stand and allowed the sample to dry for 1 minute before mixing the beads with 100  $\mu\text{L}$  of TLE solution containing 10 mM Tris-HCL, 0.1 mM EDTA, and nuclease free water. After allowing the bead mixture to stand at room temperature for 5 minutes I returned the beads to the magnet stand, pipetted all of the TLE solution into another microcentrifuge tube, and discarded the beads. I quantified DNA with a Qubit fluorometer (Life Technologies, USA) and diluted samples with TLE solution to bring the concentration to 20 ng/ $\mu\text{L}$ .

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

### 1.2.2 RADseq Library Preparation

154

I prepared RADseq libraries using the 2RAD approach outlined by Bayona-Vásquez et al., 2019. On 96 well plates, I ligated 100 ng of sample DNA in 15  $\mu\text{L}$  of a solution with 1X CutSmart Buffer (New England Biolabs, USA; NEB), 10 units of XbaI, 10 units of EcoRI, 0.33  $\mu\text{M}$  XbaI compatible adapter, 0.33  $\mu\text{M}$  EcoRI compatible adapter, and nuclease free water with a 1 hour incubation at 37°C. I then immediately added 5  $\mu\text{L}$  of a solution with 1X Ligase Buffer (NEB), 0.75 mM ATP (NEB), 100 units DNA Ligase (NEB), and nuclease free water and incubated at 22°C for 20 min and 37°C for 10 min for two cycles, followed by 80°C for 20 min to stop enzyme activity. For each 96 well plate, I pooled 10  $\mu\text{L}$  of each sample and split this pool equally between two microcentrifuge tubes. I purified each pool of libraries with a 1X volume of SpeedBead solution followed by two ethanol washes as described in the previous section except that the DNA was resuspended in 25  $\mu\text{L}$  of TLE solution.

155

156

157

158

159

160

161

162

163

164

165

166

In order to be able to detect and remove PCR duplicates, I performed a single cycle of PCR with the iTru5-8N primer which adds a random 8 nucleotide barcode to each library construct. For each plate, I prepared four PCR reactions with a total volume of 50  $\mu\text{L}$  containing 1X Kapa Hifi Buffer (Kapa Biosystems, USA; Kapa), 0.3  $\mu\text{M}$  iTru5-8N

167

168

169

170

Primer, 0.3 mM dNTP, 1 unit Kapa HiFi DNA Polymerase, 10  $\mu$ L of purified ligation product, and nuclease free water. I ran reactions through a single cycle of PCR on a thermocycler at 98°C for 2 min, 60°C for 30 s, and 72°C for 5 min. I pooled all of the PCR products for a plate into a single tube and purified the libraries with a 2X volume of SpeedBead solution as described before and resuspended in 25  $\mu$ L TLE. I added the remaining adapter and index sequences unique to each plate with four PCR reactions with a total volume of 50  $\mu$ L containing 1X Kapa Hifi (Kapa), 0.3  $\mu$ M iTru7 Primer, 0.3  $\mu$ M P5 Primer, 0.3 mM dNTP, 1 unit of Kapa Hifi DNA Polymerase (Kapa), 10  $\mu$ L purified iTru5-8N PCR product, and nuclease free water. I ran reactions on a thermocycler with an initial denaturation at 98°C for 2 min, followed by 6 cycles of 98°C for 20 s, 60°C for 15 s, 72°C for 30 s and a final extension of 72°C for 5 min. I pooled all of the PCR products for a plate into a single tube and purified the product with a 2X volume of SpeedBead solution as described before and resuspended in 45  $\mu$ L TLE.

I size selected the library DNA from each plate in the range of 450-650 base pairs using a BluePippin (Sage Science, USA) with a 1.5% dye free gel with internal R2 standards. To increase the final DNA concentrations I prepared four PCR reactions for each plate with 1X Kapa Hifi (Kapa), 0.3  $\mu$ M P5 Primer, 0.3  $\mu$ M P7 Primer, 0.3 mM dNTP, 1 unit of Kapa HiFi DNA Polymerase (Kapa), 10  $\mu$ L size selected DNA, and nuclease free water and used the same thermocycling conditions as the previous (P5-iTru7) amplification. I pooled all of the PCR products for a plate into a single tube and purified the product with a 2X volume of SpeedBead solution as before and resuspended in 20  $\mu$ L TLE. I quantified the DNA concentration for each plate with a Qubit fluorometer (Life Technologies, USA) then pooled each plate in equimolar amounts relative to the number of samples on the plate and diluted the pooled DNA to 5 nM with TLE solution. The pooled libraries were pooled with other projects and sequenced on an Illumina HiSeqX by Novogene (China) to obtain paired end, 150 base pair sequences.

### 1.2.3 Phylogenetic Data Processing

To produce alignments for phylogenetic analysis, I first I demultiplexed the iTru7 indexes using the *process\_radtags* command from *Stacks* v2.6.4 (Rochette et al., 2019) and allowed for two mismatches for rescuing reads. I removed PCR duplicates using the the *clone\_filter* command from *Stacks*. To demultiplex individual samples I used *ipyrad* v0.9.90 and allowed for one mismatch for rescuing reads. I assembled and aligned reads with *ipyrad* using default parameters and a clustering threshold of 0.8. Using *ipyrad*, I filtered loci not present in at least 75% of samples and filtered samples with fewer than 200 loci.

### 1.2.4 Maximum Likelihood

Phylogenetic methods that do not account for incomplete lineage sorting do not perform well with data impacted by this process. However, methods that do account for incomplete lineage sorting are far more computationally demanding. As a result, these methods cannot be performed with a large number of samples. I therefore conducted maximum likelihood phylogenetic inference in order to infer a phylogeny with all of the sequenced samples and to be able to identify samples that may be problematic for other methods due to recent admixture or data quality. I conducted the maximum likelihood phylogenetic inference with *IQ-TREE* v1.6.12 (Nguyen et al., 2015) with the *ipyrad*

alignment as input in order. I ran *IQ-TREE* with 1000 ultrafast bootstrap replicates  
215  
(Hoang et al., 2018) under the GTR substitution model.  
216

### 1.2.5 Multispecies Coalescent

217

In order to account for incomplete lineage sorting in the inference of phylogenetic  
218  
relationships and to infer shared divergence times, I used the program *phycoeval*. I  
219  
selected a subset of samples with up to four from each species due to the infeasible run  
220  
times for *phycoeval* with greater numbers of samples (see table 1). I used *ipyrad* to filter  
221  
loci not present in at least 75% of samples. Using a custom script I filtered the phylip  
222  
alignment file produced by *ipyrad* to exclude sites with more than two characters and  
223  
output the filtered alignment to nexus format with a biallelic character encoding. I ran  
224  
*phycoeval* with state frequencies fixed at 0.5. I set the mutation rate equal to one so that  
225  
divergence times are in units of expected substitutions per site. I set the prior on the  
226  
age of the root as an exponential distribution with a mean of 0.01. I ran *phycoeval* with  
227  
the assumption of a single effective population size shared across all of the branches of  
228  
the tree. The prior on the effective population size was a gamma prior with a shape of  
229  
four and mean of 0.0005 I ran five independent MCMC chains for 10,000 generations,  
230  
sampling every 10 generations. Each chain was started with a comb tree topology with  
231  
all branches sharing the same divergence time. I summarized the posterior sample of  
232  
tree topologies and parameters using *sumphycoeval*. To assess convergence and mixing,  
233  
I used *sumphycoeval* to calculate the potential scale reduction factor (PSRF) and the  
234  
effective sample size (ESS). I discarded the first 100 samples from each chain as burnin.  
235  
I used *sumphycoeval* to rescale the branch lengths of the maximum a posteriori (MAP)  
236  
tree produced by *sumphycoeval* so that the posterior mean root age was 16.5 million years  
237  
ago based on the estimate of Feng et al., 2017.  
238

### 1.2.6 Test for Historic Admixture

239

In order to test for a history of introgression between species of *Anaxyrus* I used the  
240  
program *dsuite* v0.5r50 (Malinsky et al., 2021) to compute the *f*-branch statistic for each  
241  
pair of *Anaxyrus* species for which the statistic can be calculated (Malinsky et al., 2018;  
242  
Reich et al., 2009). I used *ipyrad* to filter all loci that were not found in at least 50% of  
243  
the samples that passed filtering and excluded one *A. fowleri* sample (sample 006 from  
244  
??) which falls outside of the *A. fowleri* clade inferred by *IQ-TREE*. For the input tree  
245  
topology required to run *dsuite*, I used the topology inferred by *phycoeval* and I specified  
246  
*Incilius nebulifer* as the outgroup species. I ran the *dsuite* Dtrios command to compute  
247  
Patterson's the *f4*-ratio statistic for all possible trios with 20 block-jackknife replicates.  
248  
I then ran the Fbranch command from *dsuite* to compute the *f*-branch statistics from  
249  
the computed *f4*-ratio statistics. I plotted the *f*-branch statistics with *dtools* v0.1 which  
250  
is packaged with the *dsuite* program (Malinsky et al., 2021).  
251

Say something about how *f*-branch takes into account correlation among branches  
252

### 1.2.7 Population Structure Data Processing

253

I processed reads differently for the analysis of population structure following PCR  
254  
duplicate filtering. I demultiplexed individual samples, trimmed adapter sequence, and  
255  
filtered reads with low quality scores as well as reads with any uncalled bases using the  
256

*process\_radtags* command and allowed for the rescue of restriction site sequence as well as barcodes with up to two mismatches. I allowed for 14 mismatches between alleles within, as well as between individuals (M and n parameters). This is equivalent to a sequence similarity threshold of 90% for the 140 bp length of reads post trimming. I also allowed for up to 7 gaps between alleles within and between individuals. I used the *populations* command from *Stacks* to filter loci missing in more than 5% of individuals, filter all sites with minor allele counts less than 3, filter any individuals with more than 90% missing loci, and randomly sample a single SNP from each locus.

## 1.2.8 Population Structure

To investigate population structure within *A. americanus*, *A. fowleri*, *A. terrestris*, and *A. woodhousii*, I used the demultiplexed and de-cloned reads used for the phylogenetic analyses for producing alignments. I assembled and aligned these reads using *Stacks* for each species separately. I allowed for 7 mismatches between alleles within, as well as between individuals (M and n parameters). This is equivalent to a sequence similarity threshold of 95% for the 140 bp length of reads post trimming. I also allowed for up to 7 gaps between alleles within and between individuals. I used the *populations* command from *Stacks* to filter loci missing from more than 5% of samples, filter all sites with minor allele counts less than 3, filter any individuals with more than 90% missing loci and to randomly sample a single site per locus.

I ran the program *STRUCTURE* v2.3.4 (Pritchard et al., 2000) for each species separately using the admixture model in order to cluster individuals and estimate ancestry proportions for each individual. I ran *STRUCTURE* under four different models differing in the number of populations assumed (K parameter), with the parameter ranging from 1-4. I ran 10 iterations of *STRUCTURE* for each value of K for a total of 100,000 steps and burnin of 50,000 for each iteration. I used the R package *POPHELPER* v2.3.1 (Francis, 2017) to combine iterations for each value of K and to select the model producing the largest  $\Delta K$  which is the the model that has the greatest increase in likelihood score from the previous model having one fewer populations as described by (Evanno et al., 2005). I also investigated population structure with a non-parametric approach, using principle component analysis (PCA) implemented in the R package *adegenet adegenet* v2.1.10 (Jombart, 2008).

## 1.2.9 Recent *A. fowleri* x *A. woodhousii* hybridization

## 1.3 Results

### 1.3.1 Assembly and alignment with *ipyrad*

A total of 436,265,266 reads were obtained for all samples. After filtering low quality reads and reads without restriction site sequence, 435,650,926 total reads remained for assembly. The number of filtered reads per individual was highly variable with a mean of 4,538,030 (sd=3,619,076). Prior to filtering there were 171,174 loci total loci which was reduced to 659 after filtering loci not present in at least 75% of samples and filtering ?? samples which had fewer than 200 loci (Table 1). Mean sequence read coverage of the loci passing filter was 54x. The final alignment contained a total of 184,453 sites with 20,361 SNPs with 14.96% of sites and 14.71% of SNPs missing.

<b>1.3.2 Maximum Likelihood Phylogeny</b>	299
The full majority rule consensus tree inferred by <i>IQ-TREE</i> is presented in 1.2. All species were inferred as a single monophyletic group with the exception of <i>A. fowleri</i> . A single <i>A. fowleri</i> sample (sample 006) does not form a monophyletic group with other <i>A. fowleri</i> samples but is instead sister to the branch containing <i>A. woodhousii</i> and <i>A. fowleri</i> samples. A representation of the tree inferred by <i>IQ-TREE</i> with the tips within species specific clades collapsed is presented in 1.3. Each species specific clade for which there are at least two representatives samples all have ultrafast bootstrap support values of 100%. All branches below the level of the species specific clades have ultrafast bootstrap support values ranging from 70-100% with the majority being 100%. The most basal internal branch of the tree, marking the split between most of <i>Anaxyrus</i> and <i>A. punctatus</i> along with the outgroup <i>Incilius nebulifer</i> has an ultrafast bootstrap support value of 99%. The sister branch to <i>A. terrestris</i> , which contains the spurious <i>A. fowleri</i> sample (sample 006) and the clade containing <i>A. fowleri</i> and <i>A. woodhousii</i> , has an ultrafast bootstrap support value of 96%. The lowest ultrafast bootstrap support value is found on the branch sister to the <i>A. cognatus/A. speciosus</i> clade with a value of only 70%.	300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315
<b>1.3.3 Coalescent Phylogeny</b>	316
The maximum a posteriori (MAP) tree inferred under the multispecies coalescent model using <i>phycoeval</i> has a topology differs from the maximum likelihood topology inferred by <i>IQ-TREE</i> Fig. 1.4. The MAP tree produced by <i>phycoeval</i> does not have any shared divergence times among any of the 10 internal nodes of the tree. The frequency of topologies in the posterior sample that have 10 independent divergence times is 0.5. The next most frequent topology in the posterior are topologies with a single shared divergence time and nine independent divergences and occur with a frequency of 0.24. One major difference between the maximum likelihood tree inferred by <i>IQ-TREE</i> and the MAP tree inferred by <i>phycoeval</i> is that the MAP tree has one multifurcation. This multifurcation happens at the ancestor of the <i>A. quercicus</i> , <i>A. speciosus/A. cognatus</i> , and <i>A. americanus</i> group lineages. However, this node has a low posterior probability of only 0.51. All other branches in the MAP tree have high posterior probabilities of 0.98 or more. Most divergence events within <i>Anaxyrus</i> have occurred in the past 3.5 million years and most diversification within the <i>A. americanus</i> group is less than 2.5 million years old.	317 318 319 320 321 322 323 324 325 326 327 328 329 330 331
<b>1.3.4 Historic Introgression</b>	332
I used the program <i>dsuite</i> to compute the <i>f</i> -branchstatistic which is an estimate of excess allele sharing between species pairs that is not due to incomplete lineage sorting. I used the species tree topology produced by <i>phycoeval</i> for estimating the <i>f</i> -branchstatistics. The <i>f</i> -branchestimates for each species pair are presented with a heatmap in figure 1.5. Most <i>f</i> -branchestimates produced by <i>dsuite</i> were zero or very near zero. Only 24 out of 112 <i>f</i> -branchestimates were greater than 0 and 11 of those were greater than 0.05 Fig. 1.5. <i>A. americanus</i> and <i>A. woodhousii</i> had the largest number of estimates greater than zero associated with them with nearly every pairwise comparison greater than 0 Fig. 1.5. The highest <i>f</i> -branchstatistic values are between <i>A. americanus</i> and two other species: <i>A. hemiophrys</i> (0.24) and <i>A. baxteri</i> (0.22) Fig. 1.5. The values	333 334 335 336 337 338 339 340 341 342

associated with <i>A. woodhousii</i> are appreciably lower with none exceeding 0.1 ??.	343
The branch preceding <i>A. speciosus</i> and <i>A. cognatus</i> tested against <i>A. punctatus</i> along with	344
the tests of <i>A. quercicus</i> with <i>A. cognatus</i> and <i>A. speciosus</i> all exceeded 0.1.	345
<b>1.3.5 Population Structure</b>	346
<b>1.3.6 Hybridization between <i>A. fowleri</i> and <i>A. woodhousii</i></b>	347
<b>1.4 Discussion</b>	348
<b>1.4.1 Relationships</b>	349
Phylogeny of toads, doesn't place Americanus and Terrestris as sister, finds mitonuclear discordance and Finds two fowleri clades (Fontenot et al., 2011)	350
Inconsistent phylogeny (Masta et al., 2002) Inconsistent phylogeny (Pramuk et al., 2007) Inconsistent phylogeny (Graybeal, 1997) Inconsistent phylogeny (Pyron & Wiens, 2011)	351
	352
	353
	354
<b>1.4.2 Divergence Time</b>	355
<i>A. americanus</i> group diversification largely took place during the Pleistocene. A period marked by repeated glaciations. The <i>phycoeval</i> analysis does not support shared divergence events during this period suggesting that diversification in toads was driven by different	356
	357
	358
	359
<b>1.4.3 Population Structure</b>	360
Given the appearance that there are many secondary contact zones. It seems probable that toad species have undergone range expansions. Following these range expansions, are there any barriers that are now reducing gene flow? We can test that by looking for population structure within species that aligns with possible biogeographic barriers. The maximum likelihood tree along with the <i>STRUCTURE</i> analyses do not support the existence of any unrecognized species diversity or significant population structure as some mitochondrial studies have suggested.	361
	362
	363
	364
	365
	366
	367
<b>1.4.4</b>	368
Population structure in <i>A. woodhousii</i> with two overlapping mtDNA clades with one more associated with the Southwest and one more associated with the great planes (Masta et al., 2003)	369
	370
	371

## 1.5 Figures

372

## Sampling Distribution

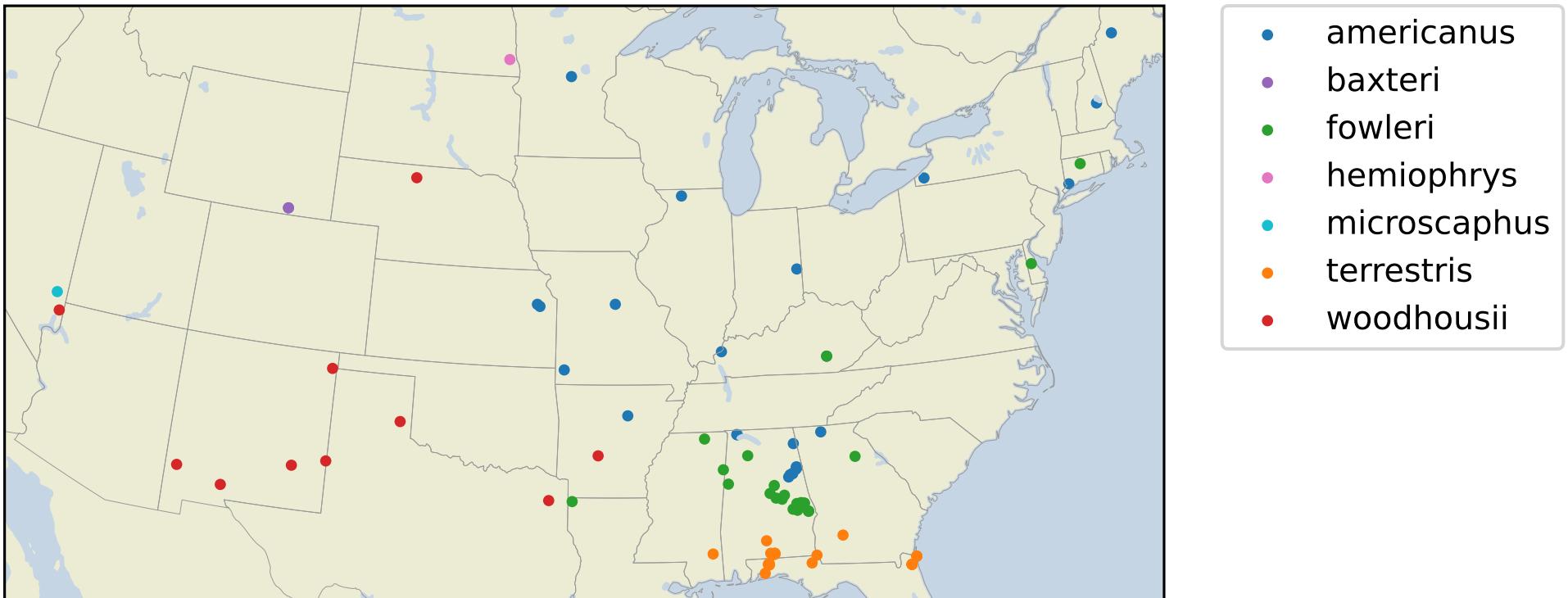


Figure 1.1. Distribution of *americanus* group samples

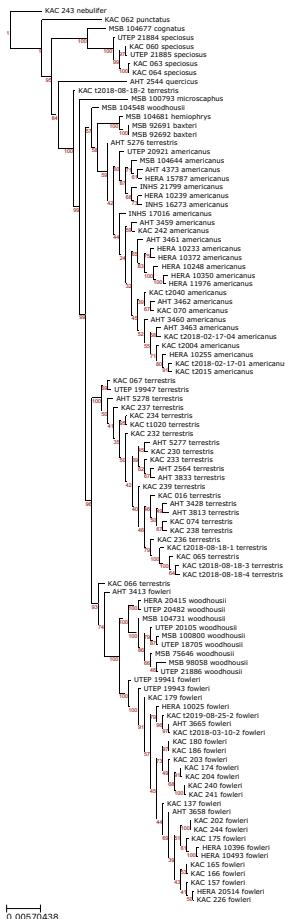


Figure 1.2. Maximum Likelihood Phylogeny

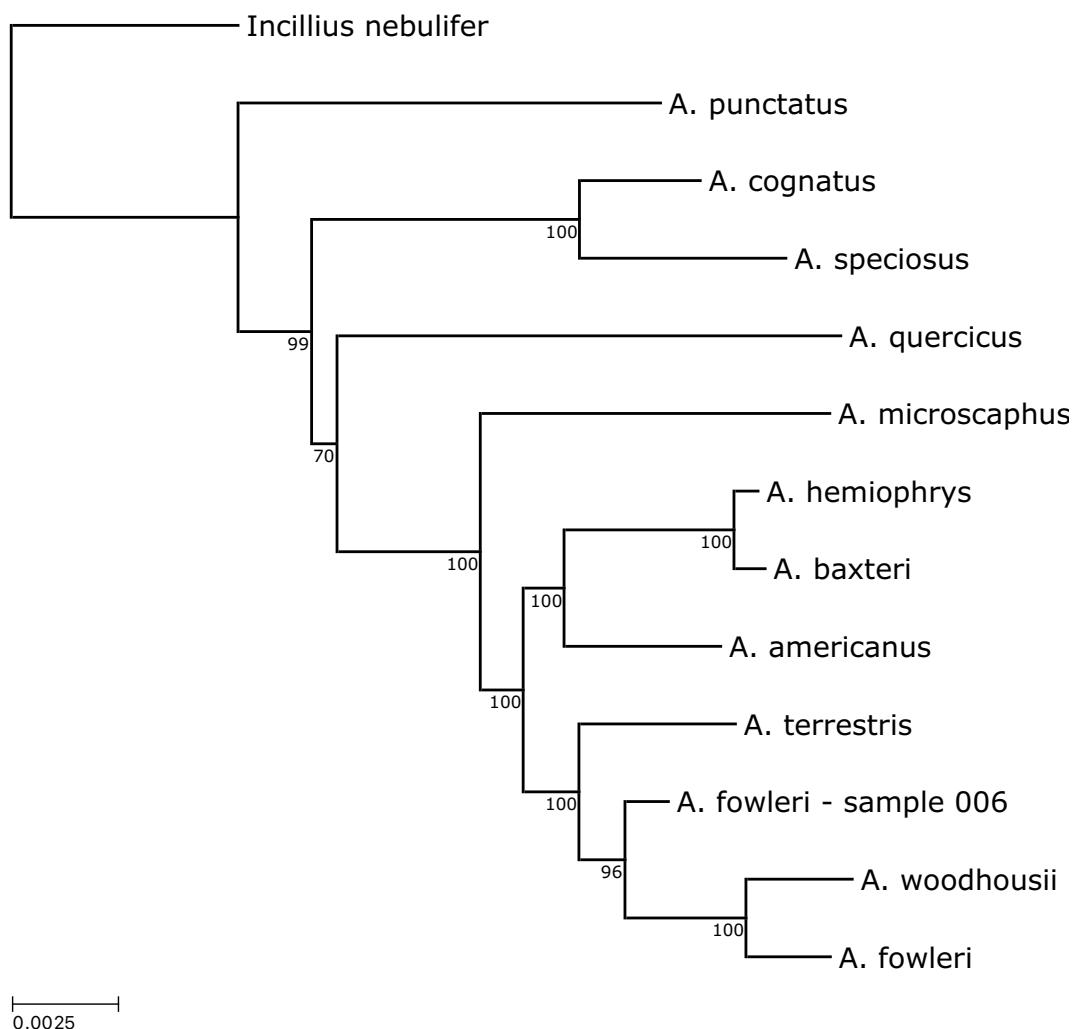
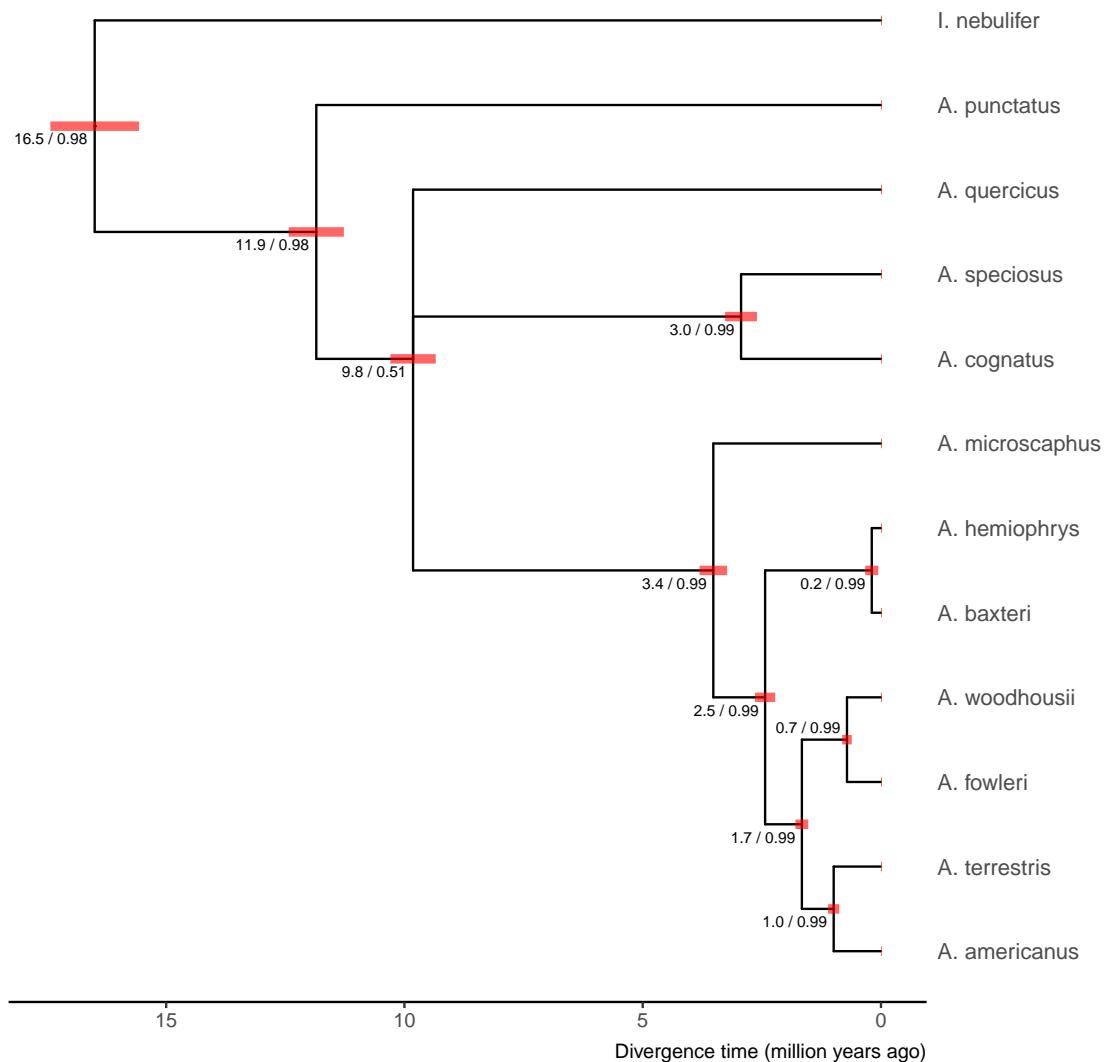
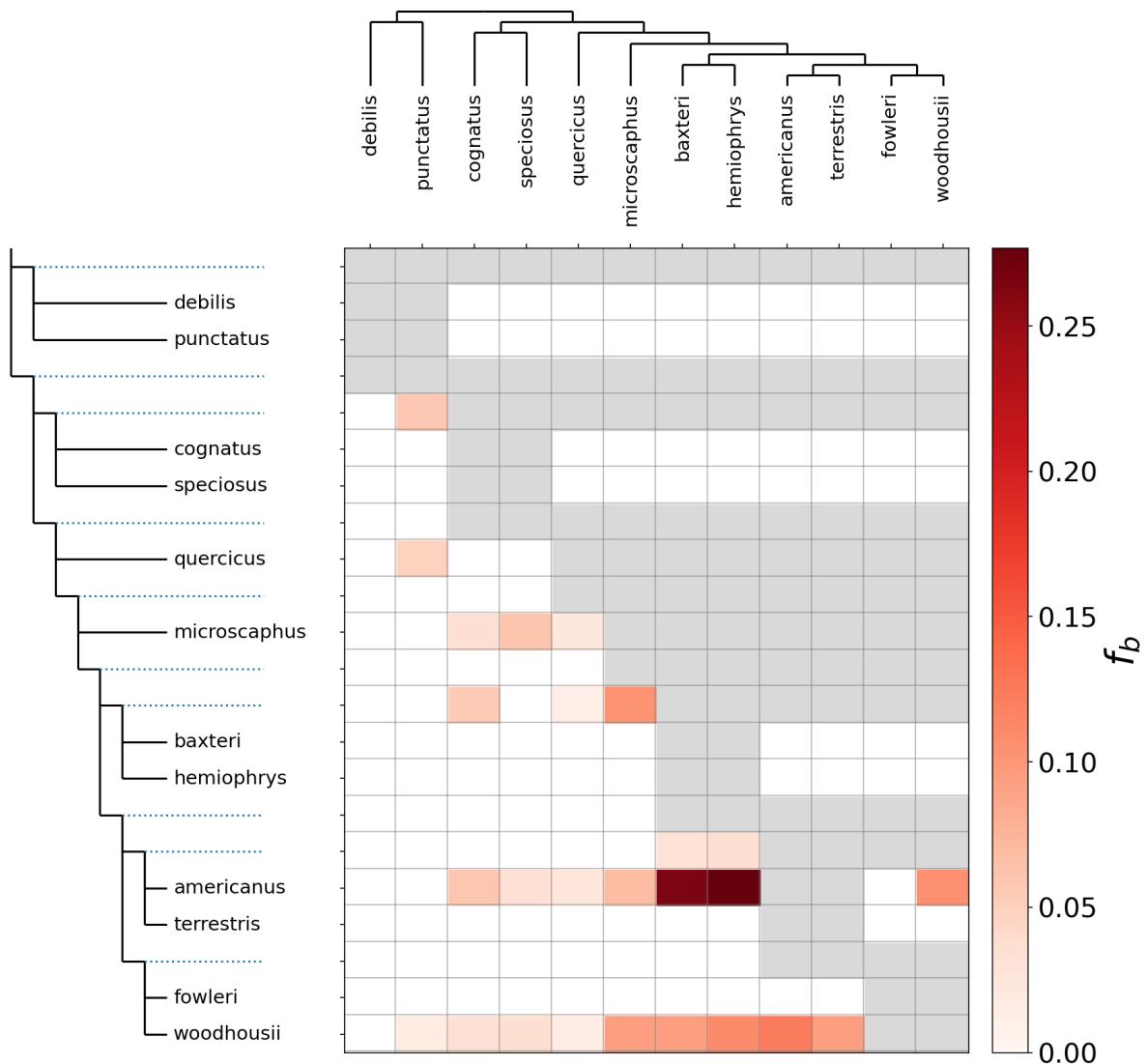


Figure 1.3. Maximum Likelihood Phylogeny with species clades collapsed. Tip lengths are the mean of the height of all collapsed tips to the base of the collapsed clade.



**Figure 1.4.** The maximum a posteriori tree inferred under a multispecies coalescent model by *phycoeval*. Branch lengths are rescaled from expected substitutions per site to millions of years using secondary time calibrations (*Materials and Methods*). Numbers displayed at each node are the mean posterior node age followed by the approximate posterior probability of the node rounded down to the nearest hundredth. Red bars show the 95% HPDI for the scaled node age at each node. Created using ggplot2 (wickham2016), ggtree (Yu et al., 2017), and treeio (Wang et al., 2020)



**Figure 1.5.** Heatmap showing the value of the  $f$ -branch statistic computed for all pairs possible pairs of *Anaxyrus* species. The  $f$ -branch statistic indicates the proportion of excess allele sharing between a species on the x-axis and branch on the y-axis (relative to its sister branch). Excess allele sharing between species identifies possible gene flow between them. Grey boxes indicate that the given tips cannot be tested by Dsuite for the given tree topology.

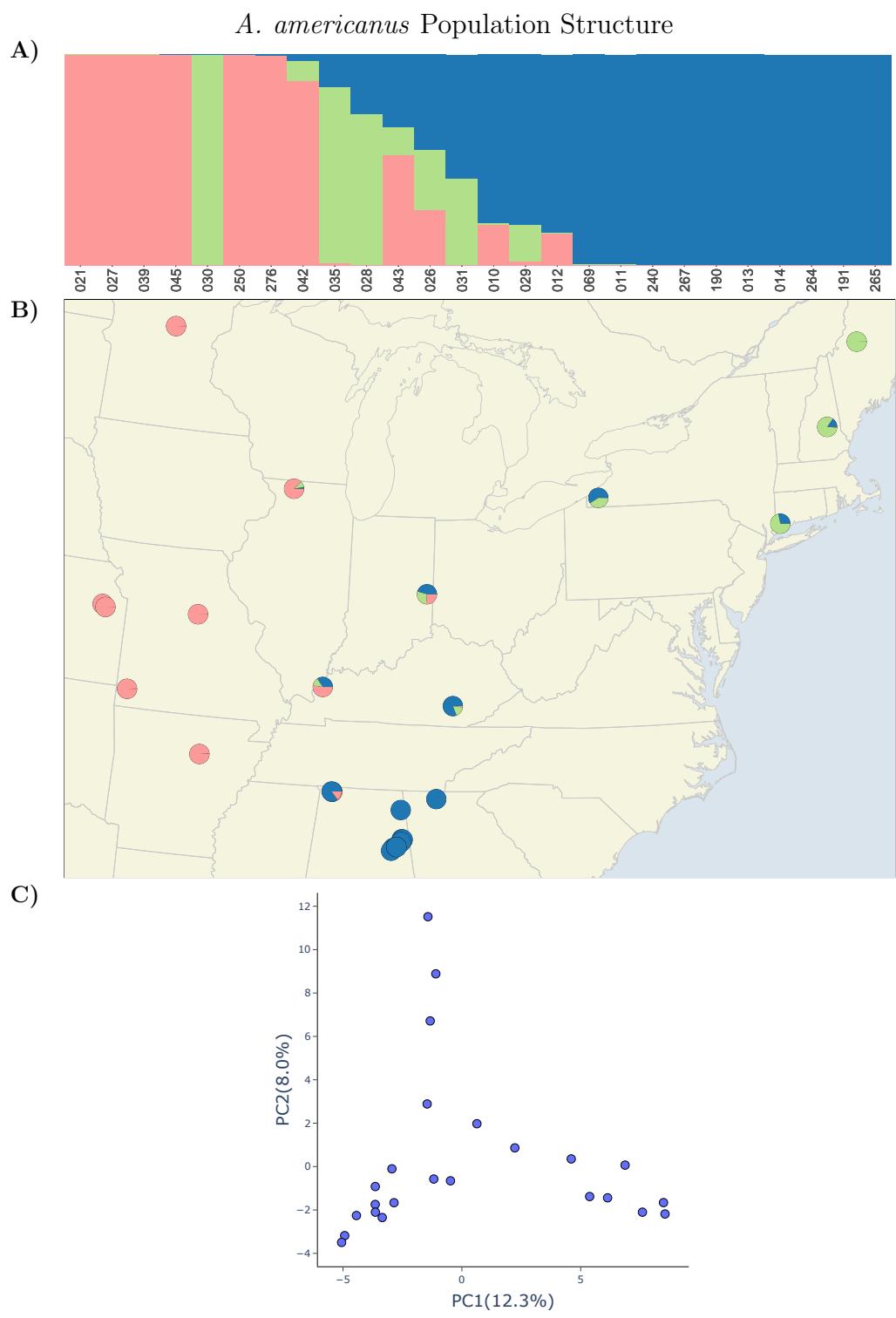


Figure 1.6. Population structure of *A. americanus* with PCA and map

*A. fowleri* Population Structure

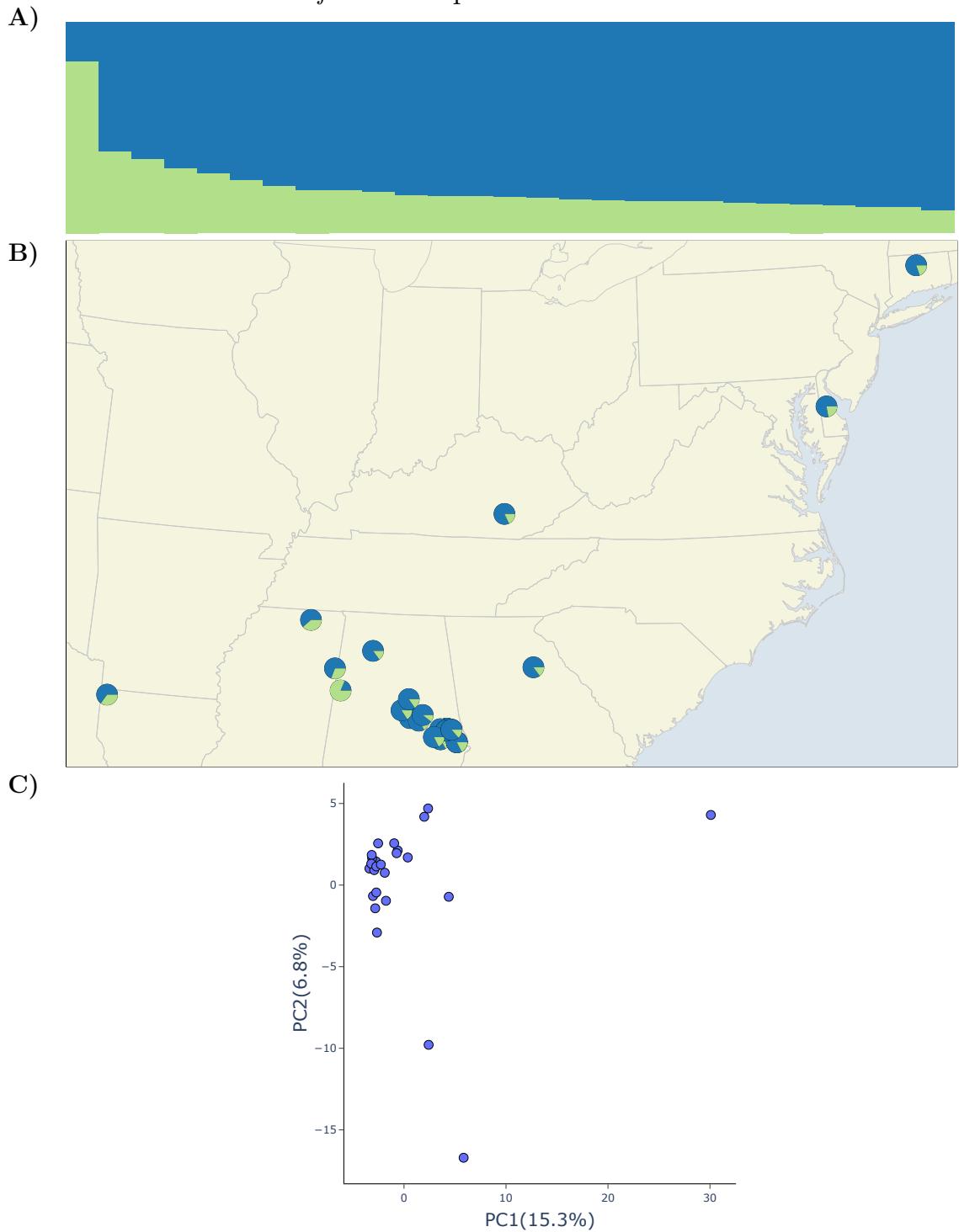


Figure 1.7. Population structure of *A. fowleri* with PCA and map

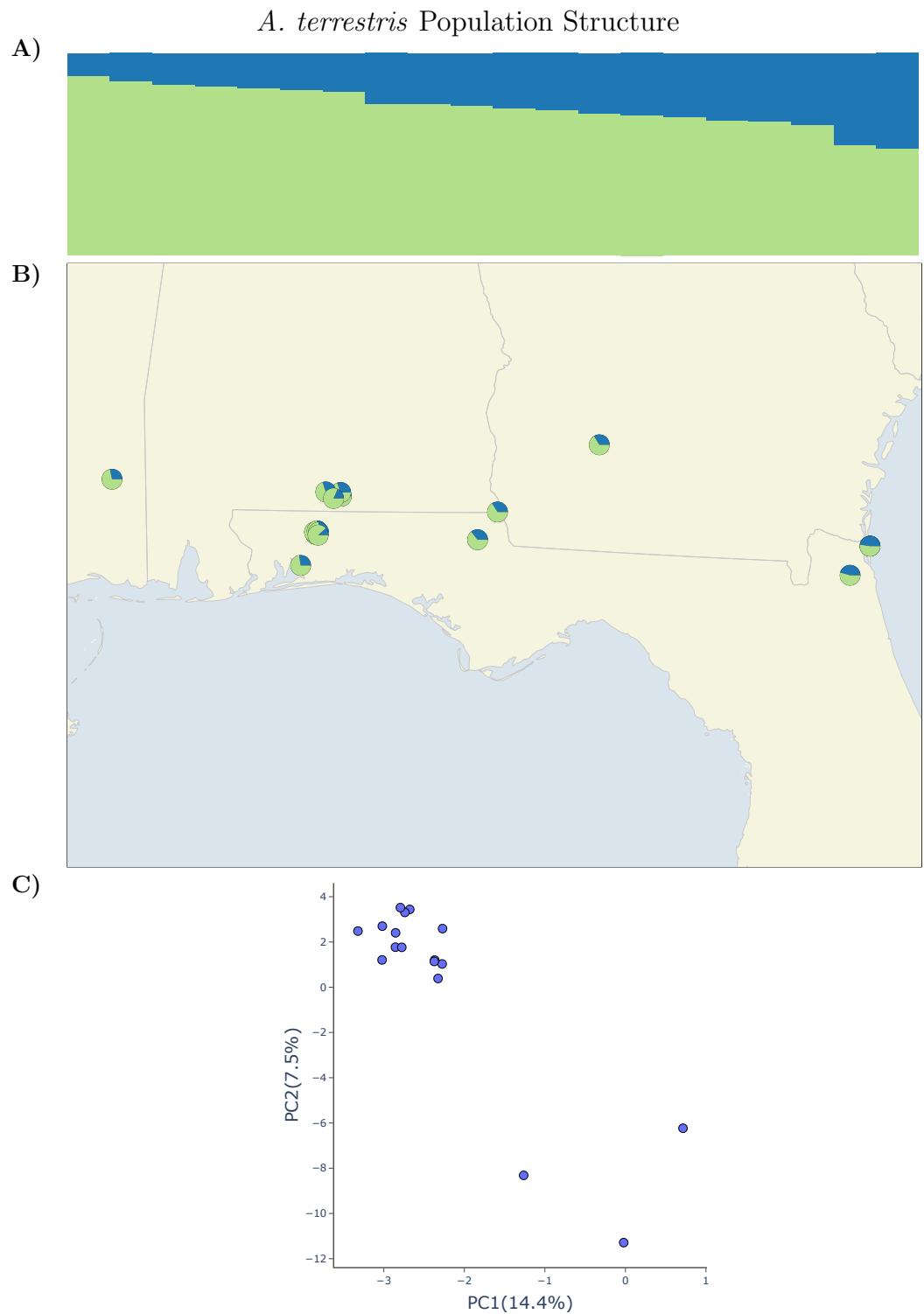


Figure 1.8. Population structure of *A. terrestris* with PCA and map

*A. woodhousii* Population Structure

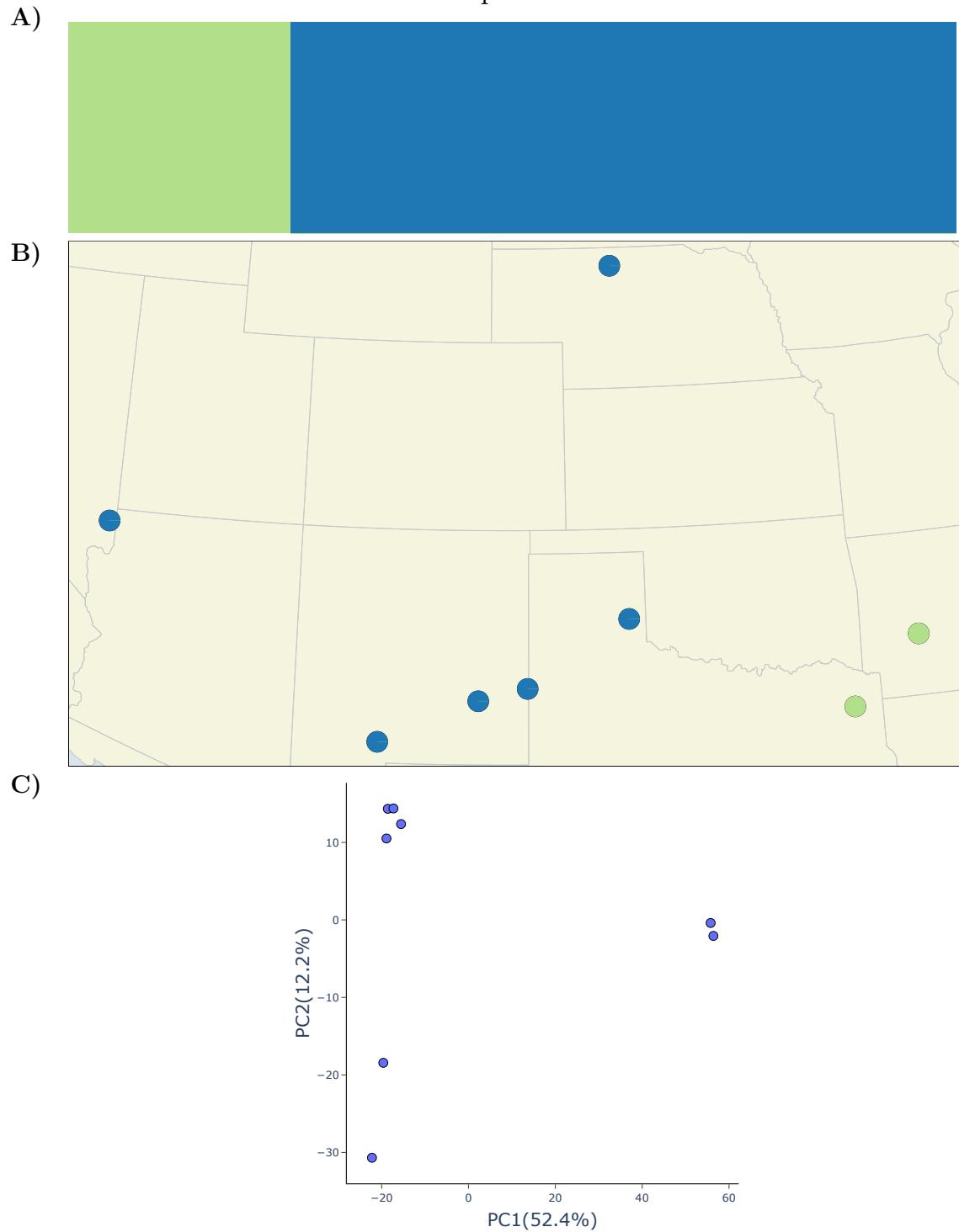


Figure 1.9. Population structure of *A. woodhousii* with PCA and map

*A. fowleri + A. woodhousii* Population Structure

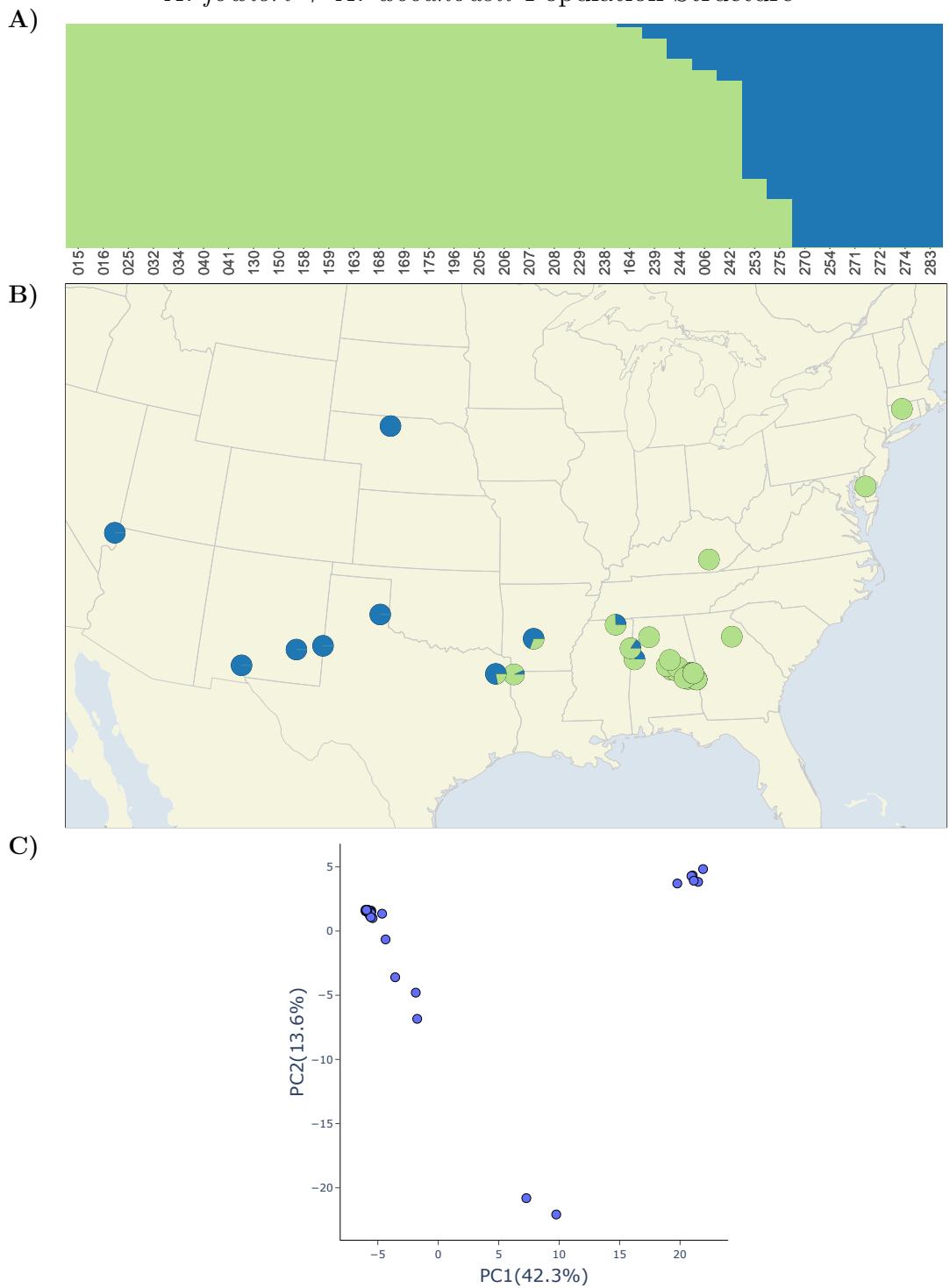


Figure 1.10. Population structure of *A. fowleri* with PCA and map

## 1.6 Tables

373

Table 1.1. Samples used in this study

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
003	AHT 2544	<i>quercicus</i>	30.99523	-86.23332	X	X	
004	AHT 2564	<i>terrestris</i>	31.55752	-84.04267	X	X	X
006	AHT 3413	<i>fowleri</i>	33.36940	-88.12941	X		X
009	AHT 3428	<i>terrestris</i>	31.12679	-86.54755	X		X
010	AHT 3459	<i>americanus</i>	34.88028	-87.71849	X		X
011	AHT 3460	<i>americanus</i>	33.78013	-85.58421	X		X
012	AHT 3461	<i>americanus</i>	34.88779	-87.74103	X		X
013	AHT 3462	<i>americanus</i>	33.77001	-85.55434	X		X
014	AHT 3463	<i>americanus</i>	33.71125	-85.59762	X		X
015	AHT 3658	<i>fowleri</i>	32.85842	-86.39697	X		X
016	AHT 3665	<i>fowleri</i>	32.81220	-86.17698	X		X
017	AHT 3813	<i>terrestris</i>	31.13854	-86.53906	X		
018	AHT 3833	<i>terrestris</i>	31.00422	-85.03427	X		X
021	AHT 4373	<i>americanus</i>	38.94913	-95.39818	X		X
022	AHT 5276	<i>terrestris</i>	31.55613	-86.82514			
023	AHT 5277	<i>terrestris</i>	31.15830	-86.55430	X		X
024	AHT 5278	<i>terrestris</i>	31.16105	-86.69868	X		X
025	HERA 10025	<i>fowleri</i>	37.11151	-84.11812	X	X	X
026	HERA 10233	<i>americanus</i>	39.86453	-85.01037	X	X	X
027	HERA 10239	<i>americanus</i>	38.99151	-92.31078	X		X
028	HERA 10248	<i>americanus</i>	41.27319	-73.38974	X		X
029	HERA 10255	<i>americanus</i>	37.11151	-84.11812	X		X
030	HERA 10350	<i>americanus</i>	45.51396	-69.95928	X	X	X
031	HERA 10372	<i>americanus</i>	42.22795	-79.36759	X		X
032	HERA 10396	<i>fowleri</i>	41.80663	-72.73281	X	X	X
033	HERA 10484	<i>marina</i>	25.61296	-80.56606			
034	HERA 10493	<i>fowleri</i>	39.08588	-75.56844	X	X	X
035	HERA 11976	<i>americanus</i>	43.51819	-71.42336	X		X

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
036	HERA 13722	<i>fowleri</i>	36.55514	-89.18929			
037	HERA 14196	<i>retiformis</i>	33.34906	-112.49010			
038	HERA 14926	<i>microscaphus</i>	33.73033	-113.98078			
039	HERA 15787	<i>americanus</i>	38.88546	-95.29399	X	X	X
040	HERA 20415	<i>woodhousii</i>	34.31743	-92.94602	X	X	X
041	HERA 20514	<i>fowleri</i>	33.95140	-83.36715	X		X
042	INHS 16273	<i>americanus</i>	42.30245	-89.55950	X		X
043	INHS 17016	<i>americanus</i>	37.46121	-88.18728	X		X
044	INHS 19127	<i>fowleri</i>	41.58247	-88.07273			
045	INHS 21799	<i>americanus</i>	46.01258	-94.26710	X		X
046	KAC 016	<i>terrestris</i>	30.54819	-86.93067	X		X
061	KAC 053	<i>fowleri</i>	32.78044	-86.73877			
062	KAC 060	<i>speciosus</i>	27.69185	-99.71955	X		
063	KAC 062	<i>punctatus</i>	29.43603	-103.50564	X		
064	KAC 063	<i>speciosus</i>	29.29522	-103.92916	X		
065	KAC 064	<i>speciosus</i>	29.29522	-103.92916	X		
066	KAC 065	<i>terrestris</i>	30.43282	-81.64088	X		
067	KAC 066	<i>terrestris</i>	30.43282	-81.64088			
068	KAC 067	<i>terrestris</i>	30.43282	-81.64088			
069	KAC 070	<i>americanus</i>	34.79963	-84.57678	X		X
071	KAC 074	<i>terrestris</i>	30.77430	-85.22690	X		X
130	KAC 137	<i>fowleri</i>	33.01461	-86.60953	X		X
150	KAC 157	<i>fowleri</i>	32.43769	-85.63620	X		X
158	KAC 165	<i>fowleri</i>	32.66356	-85.48498	X		X
159	KAC 166	<i>fowleri</i>	32.66356	-85.48498	X		X
163	KAC 174	<i>fowleri</i>	32.62938	-85.63828	X		X
164	KAC 175	<i>fowleri</i>	32.64849	-85.64711	X		X
167	KAC 178	<i>fowleri</i>	32.38644	-85.23561			

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
168	KAC 179	<i>fowleri</i>	32.38644	-85.23561	X		X
169	KAC 180	<i>fowleri</i>	32.38644	-85.23561	X		X
175	KAC 186	<i>fowleri</i>	32.38579	-85.23565	X		X
190	KAC t2018-02-17-01	<i>americanus</i>	33.55274	-85.82913	X		X
191	KAC t2018-02-17-04	<i>americanus</i>	33.48548	-85.88857	X		X
196	KAC t2018-03-10-2	<i>fowleri</i>	32.93116	-86.08465	X		X
200	KAC t2018-08-18-1	<i>terrestris</i>	30.66902	-81.44013	X		X
201	KAC t2018-08-18-2	<i>terrestris</i>	30.66902	-81.44013			
202	KAC t2018-08-18-3	<i>terrestris</i>	30.43282	-81.64088	X	X	X
203	KAC t2018-08-18-4	<i>terrestris</i>	30.66902	-81.44013	X		X
205	KAC t2019-08-25-2	<i>fowleri</i>	34.21852	-87.36662	X		X
206	KAC 202	<i>fowleri</i>	33.25104	-86.43850	X		X
207	KAC 203	<i>fowleri</i>	32.62294	-85.49660	X		X
208	KAC 204	<i>fowleri</i>	32.62294	-85.49660	X		X
229	KAC 226	<i>fowleri</i>	32.48119	-85.79838	X		X
230	KAC 230	<i>terrestris</i>	30.80933	-86.77686	X		X
231	KAC 232	<i>terrestris</i>	30.80922	-86.78994	X		X
231	KAC 232	<i>terrestris</i>	30.80922	-86.78994	X		X
232	KAC 233	<i>terrestris</i>	30.80922	-86.78994	X		X
233	KAC 234	<i>terrestris</i>	30.80922	-86.78994	X		X
234	KAC 236	<i>terrestris</i>	30.82632	-86.80258	X		X
235	KAC 237	<i>terrestris</i>	30.83733	-86.77630	X		X
236	KAC 238	<i>terrestris</i>	30.82433	-86.76284	X		X
237	KAC 239	<i>terrestris</i>	30.80162	-86.76659	X		X
238	KAC 240	<i>fowleri</i>	32.64328	-85.37114	X		X
239	KAC 241	<i>fowleri</i>	32.64328	-85.37114	X		X
240	KAC 242	<i>americanus</i>	34.50446	-85.63768	X		X
241	KAC 243	<i>nebulifer</i>	30.39140	-90.62049	X	X	

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
242	KAC 244	<i>fowleri</i>	32.89261	-93.88756	X		X
243	MSB 100793	<i>microscaphus</i>	37.27154	-114.46478	X	X	
244	MSB 100800	<i>woodhousii</i>	36.73612	-114.21972	X	X	X
245	MSB 100913	<i>microscaphus</i>	33.28038	-108.08868		X	
246	MSB 104548	<i>woodhousii</i>	36.49094	-103.20838			
247	MSB 104570	<i>fowleri</i>	34.00087	-95.38229			
248	MSB 104571	<i>americanus</i>	34.00917	-95.38058			
249	MSB 104608	<i>americanus</i>	34.00367	-94.82670			
250	MSB 104644	<i>americanus</i>	36.95124	-94.27782	X		X
251	MSB 104677	<i>cognatus</i>	46.39834	-97.20927	X	X	
252	MSB 104681	<i>hemiophrys</i>	46.47076	-97.04604	X	X	
253	MSB 104731	<i>woodhousii</i>	42.61091	-100.65607	X	X	X
254	MSB 75646	<i>woodhousii</i>	33.36365	-104.34282	X	X	X
255	MSB 92689	<i>baxteri</i>	41.21182	-105.82558			
256	MSB 92691	<i>baxteri</i>	41.21182	-105.82558	X	X	
257	MSB 92692	<i>baxteri</i>	41.21182	-105.82558	X	X	
258	MSB 96528	<i>debilis</i>	32.58239	-107.46348			
259	MSB 98058	<i>woodhousii</i>	32.83360	-108.60900			
260	MSB 98065	<i>cognatus</i>	32.63240	-108.73800		X	
261	KAC t1020	<i>terrestris</i>	31.10783	-86.62247	X		X
264	KAC t2004	<i>americanus</i>	33.58295	-85.73524	X		X
265	KAC t2015	<i>americanus</i>	33.58435	-85.74064	X		X
267	KAC t2040	<i>americanus</i>	33.58295	-85.73539	X		X
269	KAC t3040	<i>fowleri</i>	32.38644	-85.23561			
270	UTEP 18705	<i>woodhousii</i>	32.45198	-106.88317	X	X	X
271	UTEP 19941	<i>fowleri</i>	34.79137	-88.95715	X	X	X
272	UTEP 19943	<i>fowleri</i>	33.81998	-88.29533	X		X
273	UTEP 19947	<i>terrestris</i>	31.22432	-88.77548	X	X	X

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
274	UTEP 20105	<i>woodhousii</i>	33.62853	-103.08198	X		X
275	UTEP 20482	<i>woodhousii</i>	32.90708	-94.74945	X		X
276	UTEP 20921	<i>americanus</i>	35.55405	-91.83443	X		X
277	UTEP 21284	<i>debilis</i>	31.25968	-105.33402		X	
278	UTEP 21286	<i>speciosus</i>	31.70140	-105.47958			
279	UTEP 21724	<i>speciosus</i>	31.26087	-104.60168			
280	UTEP 21881	<i>cognatus</i>	35.53600	-100.44035		X	
281	UTEP 21884	<i>speciosus</i>	32.75472	-101.43208	X		
282	UTEP 21885	<i>speciosus</i>	32.20195	-100.34345	X		X
283	UTEP 21886	<i>woodhousii</i>	35.07800	-100.43392	X		X