

Contents

Table of Contents	2
List of Figures	3
List of Tables	4
1 Phylogeography	5
1.1 Introduction	5
1.2 Methods	9
1.2.1 Sampling and DNA Isolation	9
1.2.2 RADseq Library Preparation	10
1.2.3 Phylogenetic Data Processing	11
1.2.4 Maximum Likelihood	11
1.2.5 Multispecies Coalescent	12
1.2.6 Introgression	13
1.2.7 Population Structure Data Processing	13
1.2.8 Population Structure	14
1.2.9 Recent <i>A. fowleri</i> x <i>A. woodhousii</i> hybridization	15
1.3 Results	15
1.3.1 Assembly and alignment with <i>ipyrad</i>	15
1.3.2 Maximum Likelihood Phylogeny	15
1.3.3 Coalescent Phylogeny	16
1.3.4 Introgression	16

1.3.5	Population Structure	17
1.3.6	Hybridization between <i>A. fowleri</i> and <i>A. woodhousii</i>	17
1.4	Discussion	17
1.4.1	Phylogenetic relationships	17
1.4.2	Divergence Time	19
1.4.3	Hybridization and Introgression	20
1.4.4	Population Structure	23
1.4.5	Conclusion	25
	References	25
1.5	Figures	31
1.6	Tables	44

List of Figures

1.1.	Distribution map of <i>americanus</i> group samples	32
1.2.	Iqtree	33
1.3.	Iqtree	34
1.4.	Iqtree collapsed	35
1.5.	Multispecies coalescent phylogeny	36
1.6.	<i>f</i> -branch statistics	37
1.7.	<i>A. americanus</i> population structure, K=3	38
1.8.	<i>A. americanus</i> population structure, K=2	39
1.9.	<i>A. fowleri</i> population structure	40
1.10.	<i>A. terrestris</i> population structure	41
1.11.	<i>A. woodhousii</i> population structure	42
1.12.	Estimate of admixture between <i>A. fowleri</i> and <i>A. woodhousii</i>	43

List of Tables

1.1 Samples used in this study	45
------------------------------------------	----

Chapter 1

Phylogeography

1.1 Introduction

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. A critical step to understanding these processes is to obtain an accurate reconstruction of the evolutionary history of organisms.

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 *americanus* group

composed of *A. americanus*, *A. baxteri*, *A. fowleri*, *A. hemiophrys*, *A. houstonensis*, *A. microscaphus*, *A. terrestris*, and *A. woodhousii*. Two phylogenetic studies inferred trees with *A. fowleri* forming a polytomy making them inconsistent with the current taxonomy of *Anaxyrus* (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 potential source of discordance among datasets which include different loci that arises from real biological processes and impacts phylogenetic inference (Kubatko & Degnan, 2007). Incomplete lineage sorting could also produce the polytypic relationship among *A. fowleri*.

Gene flow is another potential source of discordance among genes which could drive the differences in inferred relationships among studies using different loci and could also produce the pattern seen in *A. fowleri* (Degnan & Rosenberg, 2009). 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. In the previous chapter I presented the results of a study on the hybrid zone between *A. americanus* and *A. terrestris*. Meacham, 1962 presented compelling evidence on the basis of morphological variation for the existence of a hybrid zone between *A. fowleri* and *A. woodhousii* in East Texas. 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 many populations divergence and may lead to subsequent or concurrent gene flow. Therefore, environmental variables are also important context for making inferences from hybrid zones in addition to understanding the process of diversification more generally. To date, there have not been any studies conducted to understand how the environment as driven diversification in North American toads. 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 impact multiple species simultaneously (Oaks, 2019). There has been recent development in methods to infer these events (Oaks, 2019; Oaks et al., 2022). The identification of multiple pairs of lineages that underwent divergence at the same time could provide evidence about environmental changes driving diversification. Present day population structure could also provide further understanding by revealing environmental factors that reduce gene flow assuming the biological limits of present day species have not evolved dramatically from the ancestral condition.

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 maximum likelihood analysis of a concatenated dataset of many broadly distributed samples in addition to using a multispecies coalescent analysis of a subset of the data. I also test for the presence of shared divergence times which might suggest *Anaxyrus* diversification has been driven by the same environmental changes and also estimate the absolute timing of all divergences within the genus. With the robust estimate of phylogenetic relationships among *Anaxyrus* species, I test for the presence of ongoing and historic introgression among *Anaxyrus* species. In order to identify the types of environmental factors that might have played a role in isolating populations that would eventually diverge as species, I investigate population structure within a subset of *Anaxyrus* species. Finally, I estimate proportions admixture between *A. fowleri* and *A. woodhousii* to test the hypothesis that these species form a hybrid zone in the central United States where their ranges meet.

1.2 Methods

1.2.1 Sampling and DNA Isolation

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 *Anaxyrus* as possible. I also included one *Rhinella marina* and one *incilius nebulifer* for as outgroups for phylogenetic analyses.

I isolated DNA from tissues by first lysing a piece of tissue approximately the size of a grain of rice in 300 μL 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 μL 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/ μL .

1.2.2 RADseq Library Preparation

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 μ L of a solution with 1X CutSmart Buffer (New England Biolabs, USA; NEB), 10 units of XbaI, 10 units of EcoRI, 0.33 μ M XbaI compatible adapter, 0.33 μ M EcoRI compatible adapter, and nuclease free water with a 1 hour incubation at 37°C. I then immediately added 5 μ 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 μ 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 μ L of TLE solution.

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 μ L containing 1X Kapa Hifi Buffer (Kapa Biosystems, USA; Kapa), 0.3 μ M iTru5-8N Primer, 0.3 mM dNTP, 1 unit Kapa HiFi DNA Polymerase, 10 μ 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 μ L TLE. I added the remaining adapter and index sequences unique to each plate with four PCR reactions with a total volume of 50 μ L containing 1X Kapa Hifi (Kapa), 0.3 μ M iTru7 Primer, 0.3 μ M P5 Primer, 0.3 mM dNTP, 1 unit of Kapa Hifi DNA Polymerase (Kapa), 10 μ 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 μ 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 μ M P5 Primer, 0.3 μ M P7 Primer, 0.3 mM dNTP, 1 unit of Kapa HiFi DNA Polymerase (Kapa), 10 μ 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 μ 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 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 *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 (Hoang et al., 2018) under the GTR substitution model.

1.2.5 Multispecies Coalescent

In order to account for incomplete lineage sorting in the inference of phylogenetic relationships and to infer shared divergence times, I used the program *phycoeval*. I selected a subset of up to four samples from each species due to the infeasible run times for *phycoeval* with greater numbers of samples (see table 1). I excluded sample 006 from consideration due it having an anomalous position in the maximum likelihood tree. I used *ipyrad* to filter loci not present in at least 75% of samples. Using a custom script I filtered the phylip alignment file produced by *ipyrad* to exclude sites with more than two characters and output the filtered alignment to nexus format with a biallelic character encoding. I ran *phycoeval* with state frequencies fixed at 0.5. I set the mutation rate equal to one so that divergence times are in units of expected substitutions per site. I set the prior on the age of the root as an exponential distribution with a mean of 0.01. I ran *phycoeval* with the assumption of a single effective population size shared across all of the branches of the tree. The prior on the effective population size was a gamma prior with a shape of four and mean of 0.0005 I ran five independent MCMC chains for 10,000 generations, sampling every 10 generations. Each chain was started with a comb tree topology with all branches sharing the same divergence time. I summarized the posterior sample of tree topologies and parameters using *sumphycoeval*. To assess convergence and mixing, I used *sumphycoeval* to calculate the potential scale reduction factor (PSRF) and the effective sample size (ESS). I discarded the first 100 samples from each chain as burnin. I used *sumphycoeval* to rescale the branch lengths of the maximum

a posteriori (MAP) tree produced by *sumphycoeval* so that the posterior mean root age was 16.5 million years ago based on the estimate of Feng et al., 2017.

1.2.6 Introgression

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

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

1.2.7 Population Structure Data Processing

I processed reads differently for the analysis of population structure following PCR duplicate filtering. I demultiplexed individual samples, trimmed adapter sequence, and filtered reads with low quality scores as well as reads with any uncalled bases using the *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 *POPHELP* v2.3.1 (Francis, 2017) to combine iterations for each value of K and to select the model producing the largest ΔK 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.

1.3.2 Maximum Likelihood Phylogeny

The full majority rule consensus tree inferred by *IQ-TREE* is presented in Fig. 1.2. All species were inferred as a single monophyletic group with the exception of *A. fowleri*. A single *A. fowleri* sample (sample 006) does not form a monophyletic group with other *A. fowleri* samples but is instead sister to the branch containing *A. woodhousii* and *A. fowleri* samples. A representation of the tree inferred by *IQ-TREE* with the tips within species specific clades collapsed is presented in Fig. 1.4. 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 *Anaxyrus* and *A. punctatus* along with the outgroup *Incilius nebulifer* has an ultrafast bootstrap support value of 99%. The sister branch to *A. terrestris*, which contains the spurious *A. fowleri* sample (sample 006) and the clade containing *A. fowleri* and *A. woodhousii*, has an ultrafast bootstrap support value of 96%. The lowest ultrafast

bootstrap support value is found on the branch sister to the *A. cognatus/A. speciosus* clade with a value of only 70%.

1.3.3 Coalescent Phylogeny

The maximum a posteriori (MAP) tree inferred under the multispecies coalescent model using *phycoeval* has a topology differs from the maximum likelihood topology inferred by *IQ-TREE* Fig. 1.5. The MAP tree produced by *phycoeval* 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 *IQ-TREE* and the MAP tree inferred by *phycoeval* is that the MAP tree has one multifurcation. This multifurcation happens at the ancestor of the *A. quercicus*, *A. speciosus/A. cognatus*, and *A. americanus* 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 *Anaxyrus* have occurred in the past 3.5 million years and most diversification within the *A. americanus* group is less than 2.5 million years old.

1.3.4 Introgression

I used the program *dsuite* to compute the *f*-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 *phycoeval* for estimating the *f*-branchstatistics. The *f*-branchestimates for each species pair are presented with a heat map in figure Fig. 1.6. Most *f*-branchestimates produced by *dsuite* were zero or very near zero. Only 24 out of 112 *f*-branchestimates were greater than 0 and 11 of those were greater than 0.05 Fig. 1.6. *A. americanus* and *A. woodhousii* had the largest number of estimates greater than zero associated with them with nearly every pairwise comparison

greater than 0 Fig. 1.6. The highest *f*-branchstatistic values are between *A. americanus* and two other species: *A. hemiophrys* (0.24) and *A. baxteri* (0.22) Fig. 1.6. The values associated with *A. woodhousii* are appreciably lower with none exceeding 0.1 Fig. 1.6. The highest being between *A. americanus* and *A. woodhousii* with a value of 0.098 Fig. 1.6. The *A. woodhousii* *f*-branchvalues for *A. baxteri* and *A. hemiophrys* are 0.082 and 0.086 respectively Fig. 1.6. The *f*-branchvalue between *A. woodhousii* and *A. microscaphus* is 0.05. Finally, the smallest *A. woodhousii* *f*-branchvalues are in the tests with *A. cognatus* and *A. speciosus* at 0.023 and 0.029 respectively.

1.3.5 Population Structure

1.3.6 Hybridization between *A. fowleri* and *A. woodhousii*

1.4 Discussion

1.4.1 Phylogenetic relationships

The maximum likelihood tree inferred by *IQ-TREE* Figs. 1.2 and 1.4 differs from trees inferred in previous studies of the relationships among *Anaxyrus* (Fontenot et al., 2011; Graybeal, 1997; Masta et al., 2002; Portik et al., 2023; Pramuk et al., 2007; Pyron & Wiens, 2011). Even among these previous studies there has been a great deal of inconsistency in the inferred relationships except in for the position of a few taxa. As in all previous studies, the maximum likelihood tree inferred in this study places *A. punctatus* sister to all other *Anaxyrus*. I also found the *americanus* group to be monophyletic with *A. microscaphus* sister to all other *americanus* group species which is consistent with most previous studies. Two previous studies have inferred trees which do not place *A. fowleri* samples into a single monophyletic group (Fontenot et al., 2011; Masta et al., 2002). A single *A. fowleri* sample included in this study does not fall within a monophyletic group with the remaining *A. fowleri* samples but is instead sister to the clade containing all *A. fowleri* and *A. woodhousii* samples Fig. 1.4.

All of these studies have included different species, individuals, and loci, and also used

different methods for alignment and phylogenetic inference. These differences in study design could result in the observed topology differences. The choice of locus in particular has a high likelihood of being the cause of these differences. Due to incomplete lineage sorting, the true histories of each gene may in fact differ from one another and not reflect the history of the species (Kingman, 1982). The practice of concatenating multiple loci as all of the previous studies of *Anaxyrus* evolutionary relationships have done, can produce erroneous trees with high statistical support (Kubatko & Degnan, 2007). Despite the inappropriateness of concatenated analysis with genome-wide data, it was reassuring to find that all but one individual clustered with members of its own species. *Anaxyrus* can be challenging to identify, particularly in a preserved state. The maximum likelihood tree does not suggest that any samples in the dataset have been misidentified which could be problematic for other analyses.

To account for incomplete lineage sorting, I also inferred phylogenetic relationships among *Anaxyrus* species using the multispecies coalescent method *phycoeval* along with a subset of individuals used for the maximum likelihood tree due to increased computational demands of multispecies coalescent methods. The topology of the *phycoeval* tree is substantially different from the maximum likelihood tree inferred in this study as well as trees from previous studies Fig. 1.5 (Fontenot et al., 2011; Graybeal, 1997; Masta et al., 2002; Portik et al., 2023; Pramuk et al., 2007; Pyron & Wiens, 2011). Unlike in any previous study or in the maximum likelihood tree, *A. americanus* and *A. terrestris* are placed sister to one another, whereas in all other trees it has had closer affinity to the *A. hemiophrys/A. baxteri* clade Fig. 1.4 (Portik et al., 2023; Pyron & Wiens, 2011). In the *phycoeval* tree, the *A. hemiophrys/A. baxteri* clade is instead sister to the *A. americanus/A. fowleri/A. terrestris/A. woodhousii* clade.

An unusual feature of *phycoeval* is that it can allow for multifurcations in inferred topologies (Oaks et al., 2022). This feature proved relevant for in this study as the inferred tree included one multifurcation at the ancestral node of *A. quercicus*, the *A. cognatus/A. speciosus* clade, and the *americanus* group. Previous studies have produced trees with quite short internode branches at this part of the tree as did the *IQ-TREE*

analysis in this study which is somewhat consistent with this. These methods can only produce bifurcations and thus would force any true multifurcation into bifurcations and estimate some branch length between them which would be expected to be short. In the *phycoeval* tree, the posterior probability of this split is low (0.51) so it may not be a perfect representation of the history of these lineages Fig. 1.5. More data may be necessary to have full resolution in this part of the tree. But it is clear that these three lineages diverged at least in very rapid succession if not simultaneously.

1.4.2 Divergence Time

Only three previous studies have produced estimates for age of the *Anaxyrus* lineage Feng et al., 2017; Frazão et al., 2015; Portik et al., 2023. The Frazão et al., 2015 phylogeny places *Incilius* sister to *Rhinella* rather than *Incilius* which is not supported by most recent studies making the approximately 23 mya estimate for the origin of the genus questionable Feng et al., 2017; Portik et al., 2023; Pyron and Wiens, 2011. Portik et al., 2023 estimate the split between *Anaxyrus* and *Incilius* to be 20.3 mya (95% HPD: 17.8-22.5) whereas Feng et al., 2017 estimate a much earlier age of 16.5 mya (95% CI: 14.0-19.4). The dataset from Feng et al., 2017 included near complete coverage from 95 nuclear loci whereas the Portik et al., 2023 has a higher degree of missing data (95%) and includes both mitochondrial as well as nuclear loci. For these reasons I consider the Feng et al., 2017 estimate to be the most reliable and chose it for the rescaling the branch lengths of the *phycoeval* tree.

Scaling the root of the *phycoeval* tree I estimated with the Feng et al., 2017 estimate puts the time since the most recent common ancestor (MRCA) of extant *Anaxyrus* some time between 11.9 mya when *A. punctatus* diverged from other *Anaxyrus* and 16.5 mya when *Anaxyrus* split from *Incilius* Fig. 1.5. This range is not inconsistent with the estimate of 12.3 mya (95% CI: 9.7-15.2) made by Feng et al., 2017. But it would suggest that it must have happened almost immediately before the split leading to *A. punctatus*. Portik et al., 2023 estimate the age of MRCA of *Anaxyrus* to be approximately halfway between the 14.7 mya *A. punctatus* split and the 20.3 mya split with *Incilius* at 16.7

mya. This study and the previous ones, have a high degree of uncertainty around the ages of these basal splits in the *Anaxyrus* tree. But it seems that that the split between *Incilius* and *Anaxyrus* likely happened somewhere around the start or just before the middle of the Miocene epoch. The MRCA of *Anaxyrus* and the split between the *boreas* group with a Western distribution, likely occurred prior to the middle of the Miocene. My estimate for the split between *A. punctatus* and other *Anaxyrus* would be right at the middle of the Miocene at a time when both precipitation and temperature underwent a decline in the North American interior and there was expansion of grasslands (Morales-García et al., 2020). The timing of the multifurcation of the *A. quercicus*, *A. cognatus/A. speciosus*, and *americanus* group lineages coincides with a previously identified shift in the ecomorphology of ungulate mammals inhabiting North America (Morales-García et al., 2020).

I estimate that diversification of the *americanus* group has all happened in the past 3.4 million years. This accounts for a large portion of the diversity of *Anaxyrus* and includes two additional un-sampled species which other studies have found to be nested within this clades Portik et al., 2023; Pyron and Wiens, 2011. Those being *A. houstonensis* and *A. californicus*. This means that most diversification within *Anaxyrus* took place just before and during the Pleistocene 2.58 million to 11,700 years ago. This is a period marked by extreme climatic variation and repeated glacial cycles that transformed the climate and geography of the North American continent (Holman, 1995, 2003). Surprisingly, there is no evidence from the *phycoeval* analysis that any single one of these cycles was a driver of multiple diversification events and instead each event occurred independently during this period of *Anaxyrus* evolution.

1.4.3 Hybridization and Introgression

There are numerous reports of hybridization among many different pairs of *Anaxyrus* species. However, the consequences of this hybridization are largely unknown. Using the *f*-branchtest, I found support for a modest level of introgression among several pairs of species which have been previously reported to hybridize. Most of which presently exist

in sympatry with one another. The highest *f*-branchstatistics were calculated between *A. americanus* and *A. hemiophrys* and between *A. americanus* and *A. baxteri* with values of 0.24 and 0.22 respectively Fig. 1.6. One known *Anaxyrus* hybrid zones is one that exists between *A. americanus* and *A. hemiophrys*. Green, 1983 reported clinal variation of allozyme alleles at five different loci across an approximately 100 km transect in southeastern Manitoba, Canada. This sharp cline of variation suggests that reproductive isolation between these species is quite high. It could be that introgression is occurring beyond this narrow hybrid zone but the sample included in this study was sampled from a location in close proximity to the range of *A. americanus* Fig. 1.1 (Conant & Collins, 1998). Thus, it is difficult to say if they detected introgression is shared by *A. hemiophrys* as a whole or is only present within a hybrid zone. Interestingly, there is also a high *f*-branchscore between *A. americanus* and *A. baxteri* which do no have ranges that are close to on another. It is possible that introgression from *A. americanus* occurred before the divergence of *A. hemiophrys* and *A. baxteri*. *A. baxteri* is believed to be a relict of a more southerly distribution of *A. hemiophrys* during a recent Pleistocene glacial period (Henrich, 1968). Unfortunately, it is not possible to directly test for this scenario with *dsuite* due to limitations of the *f*-branchtest (Malinsky et al., 2021).

Several other *f*-branchtests returned non-zero values although these were much lower. More than half of the *A. woodhousii* *f*-branchstatistics were greater than zero *A. americanus* Fig. 1.6. Hybridization between all of these species is plausible as *A. woodhousii* occurs in sympatry at some part of its range with nearly all of them. Contemporary hybridization involving *A. woodhousii* has been reported with *A. americanus*, *A. cognatus*, *A. microscaphus*, and *A. speciosus* (Sullivan, 1986). There is presently little to no overlap between *A. woodhousii* and *A. hemiophrys* however there could have been in the recent past as due to Pleistocene glaciation pushing the range of *A. hemiophrys* further south (Henrich, 1968). The two non-zero *f*-branchvalues for *A. quercicus* with *A. punctatus* and *A. speciosus* are perplexing. The distribution of *A. quercicus* is confined to the pine woodlands of the Southeastern United States whereas the other two species are found in the short arid grasslands and deserts of the Southwest (Conant & Collins, 1998). The *f*-

branchstatistic for the comparison between *A. punctatus* and the common ancestor of *A. speciosus* and *A. cognatus* is more plausible given their broadly overlapping distributions in the present day (Conant & Collins, 1998).

Unfortunately, there were many comparisons among *Anaxyrrus* that could not be made using the *f*-branchtest due to the structure of the phylogeny. Particularly among ancestral species between which past introgression could be driving the pattern seen across the extant diversity which is a known caveat with the *f*-branchtest (Malinsky et al., 2021). The results presented here are consistent with introgression being an important factor in the evolutionary history of *Anaxyrrus* but gaps remain. These are gaps that could potentially be addressed with more powerful likelihood or pseudo-likelihood methods.

The d-statistic class of methods for detecting introgression are not able to test for introgression between sister species so could not shed any light on putative hybridization between *A. fowleri* and *A. woodhousii*. In order to test for introgression between *A. fowleri* and *A. woodhousii* I used the program structure along with PCA. The results of both *STRUCTURE* and PCA are consistent with the existence of a hybrid zone between these two species. Two *A. woodhousii* samples, one from Arkansas and the other from Texas, have large proportions of inferred ancestry from *A. fowleri*. Several *A. fowleri* samples have large admixture proportions from *A. woodhousii* as well. The transition of ancestry proportions forms a steady East West gradient with one outlier present in Louisiana Fig. 1.12. The PCA results largely corroborate the results of the *STRUCTURE* analysis with *A. woodhousii* samples clustered tightly together, most *A. fowleri* samples clustering tightly with a few deviating toward the center of the first principal component axis, and finally two samples right in the center of the first principal component axis. These results suggest the hybrid zone between *A. fowleri* and *A. woodhousii* is quite wide, on the order of hundreds of kilometers Fig. 1.12.

This brings the number of confirmed *Anaxyrrus* hybrid zones to three along with the *A. americanus/A. terrestris* and *A. americanus/A. hemiophrys* hybrid zones. Based on the *phycoeval* phylogeny, all of these species emerged within the past 2.5 million years. This important context sheds light on the tempo of diversification within *Anaxyrrus*. The

sister species pairs *A. fowleri/A. woodhousii* and *A. americanus/A. terrestris* diverged only 0.7 and 1.0 mya respectively Fig. 1.5. Neither of these species pairs has evolved a degree of reproductive isolation and/or character displacement that permits them to exist in sympatry with one another. Hybridization between species with older divergence times appears to be much less frequent with the exception of hybridization between *A. americanus* and *A. hemiophyrs*. *A. fowleri* occurs in sympatry across large regions with both *A. americanus* and *A. terrestris* and *A. woodhousii* overlaps significantly with *A. americanus* (Conant & Collins, 1998). This is likely possible due to a higher degree of reproductive isolation that has evolved between these species pairs in the form of differences in advertisement call and timing of reproduction ??Citation, possibly as a result of reinforcement. This pattern suggests that pre-zygotic reproductive incompatibility tends to evolve more rapidly in *Anaxyrus*. However, the hybrid zone between *A. americanus* and *A. hemiophyrs* complicates things as these two species are much more divergent from one another. Perhaps they have more recently come into secondary contact and there has not been sufficient time to evolve pre-zygotic barriers to reproduction. This would support the idea that mating barriers between sympatric species have evolved through reinforcement and not under some other selective pressure.

1.4.4 Population Structure

An examination of population structure can potentially provide clues about the historical factors that have shaped a species' diversification. The geographic barriers that result in the reduction of gene flow within species could be the same type of barrier that has resulted in past speciation events involving a species or its close relatives. The *STRUCTURE* analysis of the *A. americanus* samples reveals population structure within this species which materializes as a gradient along three geographic axis at the center of which there is a high degree of admixture Fig. 1.7. This variation corresponds loosely with the proposed subspecies *A. anaxyrus charlesmithi* with a range that includes the Southwestern part of the *A. americanus* range in Oklahoma, Arkansas, and Missouri (Bragg, 1954). This population structure, separates *A. americanus* into a Western pop-

ulation, a Southern population, and a Northeastern population Fig. 1.7. At the center of the range are highly admixed samples and with distance from this center, samples differentiate along the three different geographic axes. This pattern is recapitulated in the PCA plots showing the same three axes of variation. Using the ΔK method for choosing the best fitting *STRUCTURE* model favors the model with a K of two which separates *A. americanus* into Western and Eastern populations. However, the K of three fits well with the results of the PCA, makes sense in a geographic context, and shows a more stark transition of admixture proportions from the Western group and Eastern groups.

The population *STRUCTURE* analyses conducted for *A. fowleri*, *A. terrestris*, and *A. woodhousii* did not yield evidence for the existence of distinct populations within these species. Despite previous studies finding evidence for two distinct *A. woodhousii* populations Masta et al., 2003; Shannon and Lowe, 1955, I found *A. woodhousii* no evidence of this in my dataset.

Masta et al., 2003 found two divergent clades of *A. woodhousii* using mitochondrial data. Conant and Collins, 1998 two distinct populations of *A. woodhousii* have been recognized on the basis of morphological differences with considering the southwest population to be a subspecies *Anaxyrus woodhousii australis* Shannon and Lowe, 1955 subspecies *A. woodhousii australis* described on the basis of morphology.

It is possible that I simply did not include samples from the range of a distinct population. I may not have sampled from this clade.

Differentiated samples of *A. woodhousii* may not have been included.

For *A. americanus* there is evidence for the existence of three somewhat differentiated groups that emerge along three different geographic axis at the center of which there is a high degree of admixture.

subspecies *A. anaxyrus charlesmithi* (Bragg, 1954) subspecies *A. woodhousii australis*

Population structure in *A. woodhousii* with two overlapping mtDNA clades with one more associated with the Southwest and one more associated with the great plains (Masta et al., 2003)

1.4.5 Conclusion

The evolutionary history of *Anaxyrus* inferred in this study has some important implications for understanding the contemporary hybridization in *Anaxyrus*.

Given the great interest in hybridization within *Anaxyrus* and the promise of this group for furthering our understanding speciation, it is important to consider the implications of the evolutionary history inferred in this study.

References

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelandsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., Bentley, K. E., Finger, J. W., Louha, S., Troendle, N., Diaz-Jaimes, P., Mauricio, R., & Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*, 7, e7724. <https://doi.org/10.7717/peerj.7724>
- Blair, W. F. (1963). Intragroup genetic compatibility in the *Bufo americanus* species group of toads. *The Texas Journal of Science*, 13, 15–34.
- Blair, W. F. (1972). *Evolution in the genus Bufo*. University of Texas Press.
- Bragg, A. N. (1954). *Bufo terrestris charlesmithi*, a new subspecies from Oklahoma. *The Wasmann Journal of Biology*, 12, 245–254.
- Conant, R., & Collins, . (1998). *Reptiles and amphibians: Eastern/Central North America* (3rd ed.). Houghton Mifflin Co., Boston, MA.

- Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, 24(6), 332–340. <https://doi.org/10.1016/j.tree.2009.01.009>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Feng, Y.-J., Blackburn, D. C., Liang, D., Hillis, D. M., Wake, D. B., Cannatella, D. C., & Zhang, P. (2017). Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. *Proceedings of the National Academy of Sciences*, 114(29). <https://doi.org/10.1073/pnas.1704632114>
- Fontenot, B. E., Makowsky, R., & Chippindale, P. T. (2011). Nuclear–mitochondrial discordance and gene flow in a recent radiation of toads. *Molecular Phylogenetics and Evolution*, 59(1), 66–80. <https://doi.org/10.1016/j.ympev.2010.12.018>
- Francis, R. M. (2017). POPHELPER: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17(1), 27–32. <https://doi.org/10.1111/1755-0998.12509>
- Frazão, A., da Silva, H. R., & Russo, C. A. d. M. (2015). The Gondwana Breakup and the History of the Atlantic and Indian Oceans Unveils Two New Clades for Early Neobatrachian Diversification. *PLOS ONE*, 10(11), e0143926. <https://doi.org/10.1371/journal.pone.0143926>
- Graybeal, A. (1997). Phylogenetic relationships of bufonid frogs and tests of alternate macroevolutionary hypotheses characterizing their radiation. *Zoological Journal of the Linnean Society*, 119(3), 297–338. <https://doi.org/10.1111/j.1096-3642.1997.tb00139.x>
- Green, D. M. (1983). Allozyme Variation through a Clinal Hybrid Zone between the Toads *Bufo americanus* and *B. hemiophrys* in Southeastern Manitoba. *Herpetologica*, 39(1), 28–40.

- Green, D. M. (1996). The bounds of species: Hybridization in the *Bufo americanus* group of North American toads. *Israel Journal of Zoology*, 42, 95–109.
- Henrich, T. W. (1968). Morphological Evidence of Secondary Intergradation between *Bufo hemiophrys* Cope and *Bufo americanus* Holbrook in Eastern South Dakota. *Herpetologica*, 24(1), 1–13.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Holman, J. A. (1995). *Pleistocene Amphibians and Reptiles in North America*. Oxford University Press.
- Holman, J. A. (2003). *Fossil frogs and toads of North America*. Indiana university press.
- Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, 33(6), 1635–1638. <https://doi.org/10.1093/molbev/msw046>
- Jombart, T. (2008). Adegenet : A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kingman, J. (1982). The coalescent. *Stochastic Processes and their Applications*, 13(3), 235–248. [https://doi.org/10.1016/0304-4149\(82\)90011-4](https://doi.org/10.1016/0304-4149(82)90011-4)
- Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence (T. Collins, Ed.). *Systematic Biology*, 56(1), 17–24. <https://doi.org/10.1080/10635150601146041>
- Lyman, R. A., & Edwards, C. E. (2022). Revisiting the comparative phylogeography of unglaciated eastern North America: 15 years of patterns and progress. *Ecology and Evolution*, 12(4), e8827. <https://doi.org/10.1002/ece3.8827>
- Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595. <https://doi.org/10.1111/1755-0998.13265>

- Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, 2(12), 1940–1955. <https://doi.org/10.1038/s41559-018-0717-x>
- Masta, S. E., Laurent, N. M., & Routman, E. J. (2003). Population genetic structure of the toad *Bufo woodhousii* : An empirical assessment of the effects of haplotype extinction on nested cladistic analysis. *Molecular Ecology*, 12(6), 1541–1554. <https://doi.org/10.1046/j.1365-294X.2003.01829.x>
- Masta, S. E., Sullivan, B. K., Lamb, T., & Routman, E. J. (2002). Molecular systematics, hybridization, and phylogeography of the *Bufo americanus* complex in Eastern North America. *Molecular Phylogenetics and Evolution*, 24(2), 302–314. [https://doi.org/10.1016/S1055-7903\(02\)00216-6](https://doi.org/10.1016/S1055-7903(02)00216-6)
- Meacham, W. R. (1962). Factors Affecting Secondary Intergradation Between Two Allopatic Populations in the *Bufo Woodhousei* Complex. *American Midland Naturalist*, 67(2), 282. <https://doi.org/10.2307/2422709>
- Morales-García, N. M., Säilä, L. K., & Janis, C. M. (2020). The Neogene Savannas of North America: A Retrospective Analysis on Artiodactyl Faunas. *Frontiers in Earth Science*, 8.
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nuñez, L. P., Gray, L. N., Weisrock, D. W., & Burbrink, F. T. (2023). The phylogenomic and biogeographic history of the gartersnakes, watersnakes, and allies (Natricidae: Thamnophiini). *Molecular Phylogenetics and Evolution*, 186, 107844. <https://doi.org/10.1016/j.ympev.2023.107844>
- Oaks, J. R. (2019). Full Bayesian Comparative Phylogeography from Genomic Data (L. Kubatko, Ed.). *Systematic Biology*, 68(3), 371–395. <https://doi.org/10.1093/sysbio/syy063>

- Oaks, J. R., Wood, P. L., Siler, C. D., & Brown, R. M. (2022). Generalizing Bayesian phylogenetics to infer shared evolutionary events. *Proceedings of the National Academy of Sciences*, 119(29), e2121036119. <https://doi.org/10.1073/pnas.2121036119>
- Portik, D. M., Streicher, J. W., & Wiens, J. J. (2023). Frog phylogeny: A time-calibrated, species-level tree based on hundreds of loci and 5,242 species. *Molecular Phylogenetics and Evolution*, 188, 107907. <https://doi.org/10.1016/j.ympev.2023.107907>
- Pramuk, J. B., Robertson, T., Sites, J. W., & Noonan, B. P. (2007). Around the world in 10 million years: Biogeography of the nearly cosmopolitan true toads (Anura: Bufonidae). *Global Ecology and Biogeography*, 0(0), 070817112457001–???. <https://doi.org/10.1111/j.1466-8238.2007.00348.x>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Pyron, R. A., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, 61(2), 543–583. <https://doi.org/10.1016/j.ympev.2011.06.012>
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489–494. <https://doi.org/10.1038/nature08365>
- Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, 152(2), 713–727.
- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28(21), 4737–4754. <https://doi.org/10.1111/mec.15253>
- Shannon, F. A., & Lowe, F. A. (1955). A new subspecies of *Bufo woodhousei* from the inland Southwest. *Herpetologica*, 11, 185–190.

- Sullivan, B. K. (1986). Hybridization between the Toads *Bufo microscaphus* and *Bufo woodhousei* in Arizona: Morphological Variation. *Journal of Herpetology*, 20(1), 11–21. <https://doi.org/10.2307/1564120>
- Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution*, 37(2), 599–603. <https://doi.org/10.1093/molbev/msz240>
- Wickam, H. (2016). Ggplot2: Elegant graphics for data analysis. *Springer-Verlag*. Accessed March, 16, 2021.
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. <https://doi.org/10.1111/2041-210X.12628>

1.5 Figures

Sampling Distribution

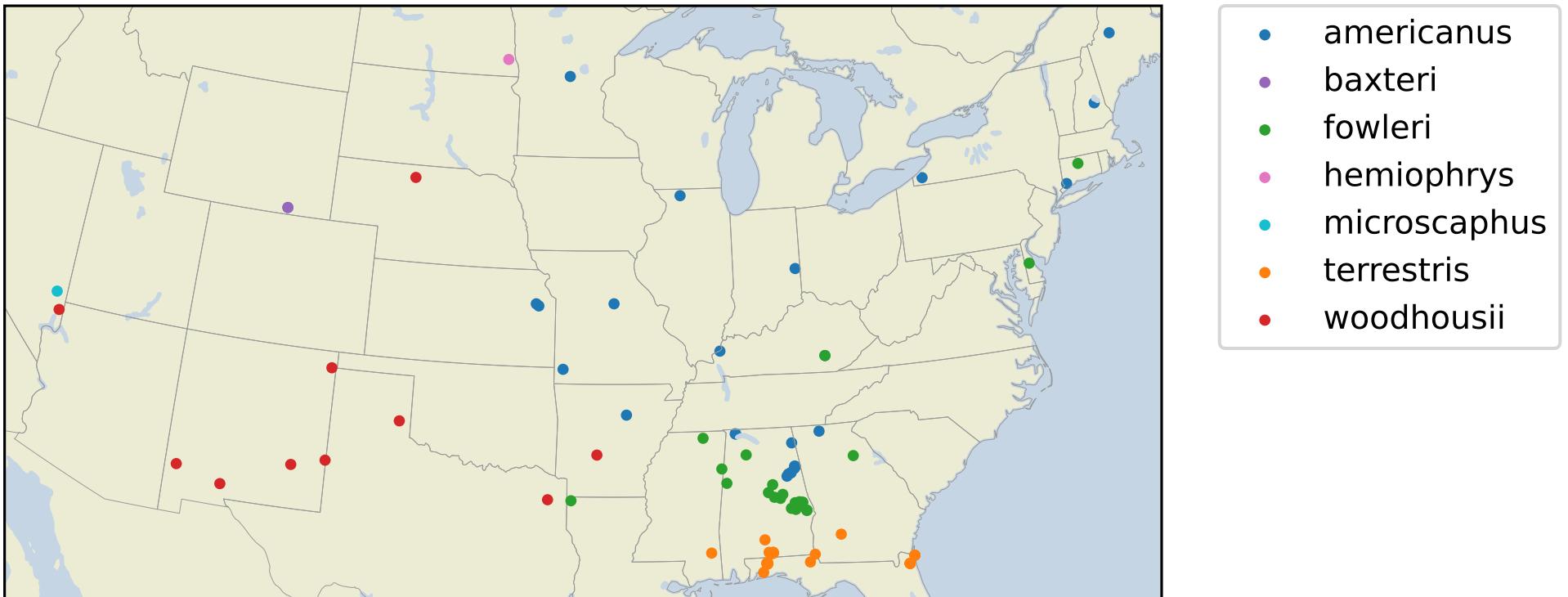


Figure 1.1. Distribution map of *americanus* group samples sequenced for this study.

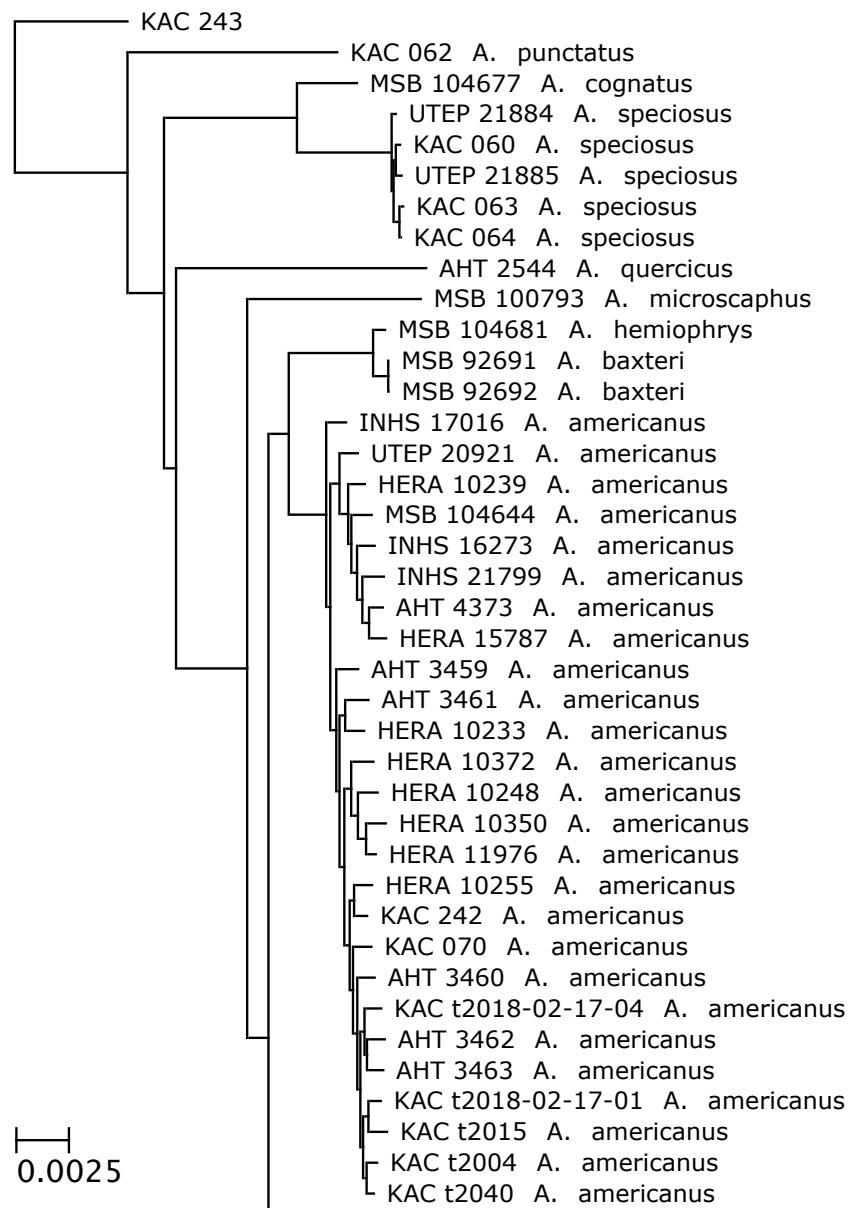


Figure 1.2. Maximum Likelihood Phylogeny Plotted using ETE 3.1.2 (Huerta-Cepas et al., 2016).

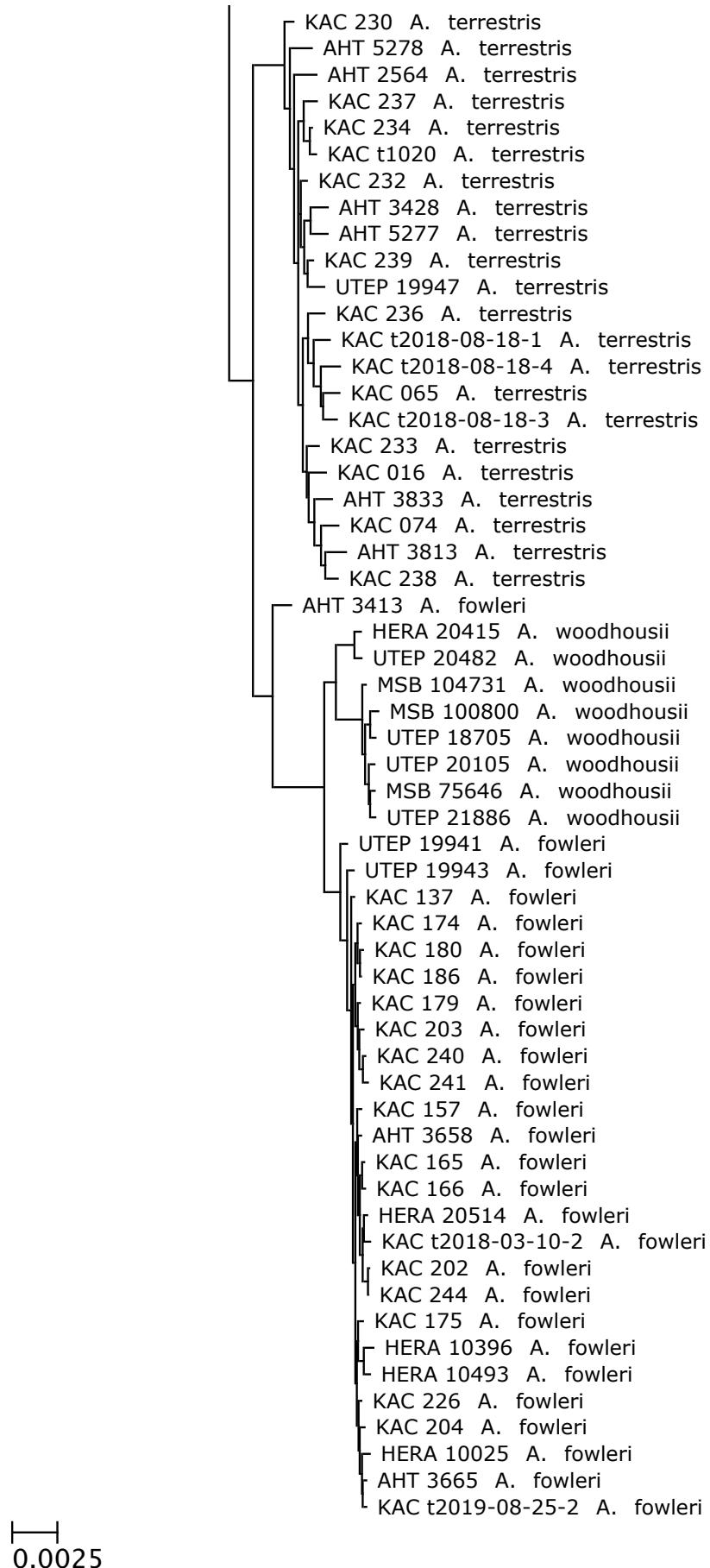


Figure 1.3. Maximum Likelihood Phylogeny Plotted using ETE 3.1.2 (Huerta-Cepas et al., 2016).

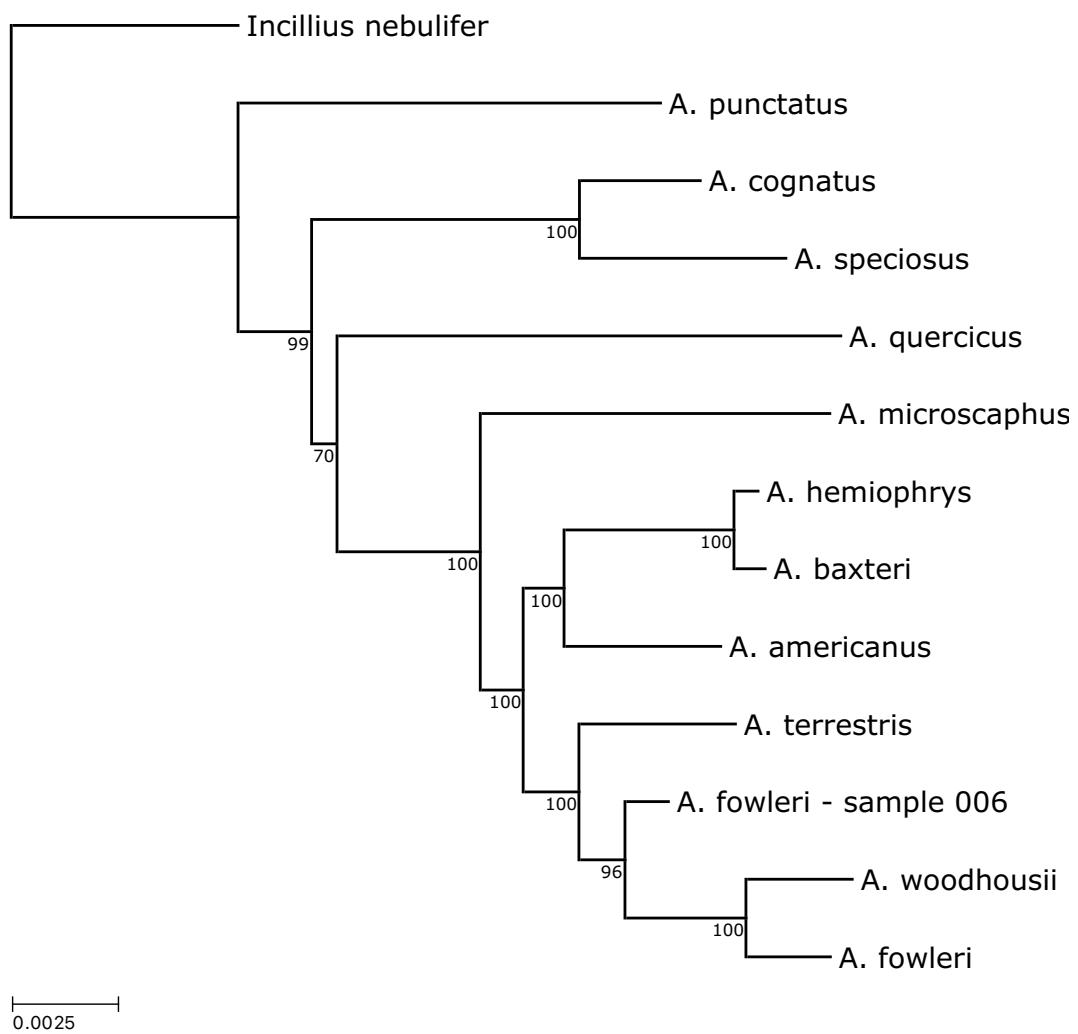


Figure 1.4. Maximum Likelihood Phylogeny with species clades collapsed. The lengths of tip branches are equal to the mean height of all collapsed tips from the base the collapsed clade. Plotted using ETE 3.1.2 (Huerta-Cepas et al., 2016).

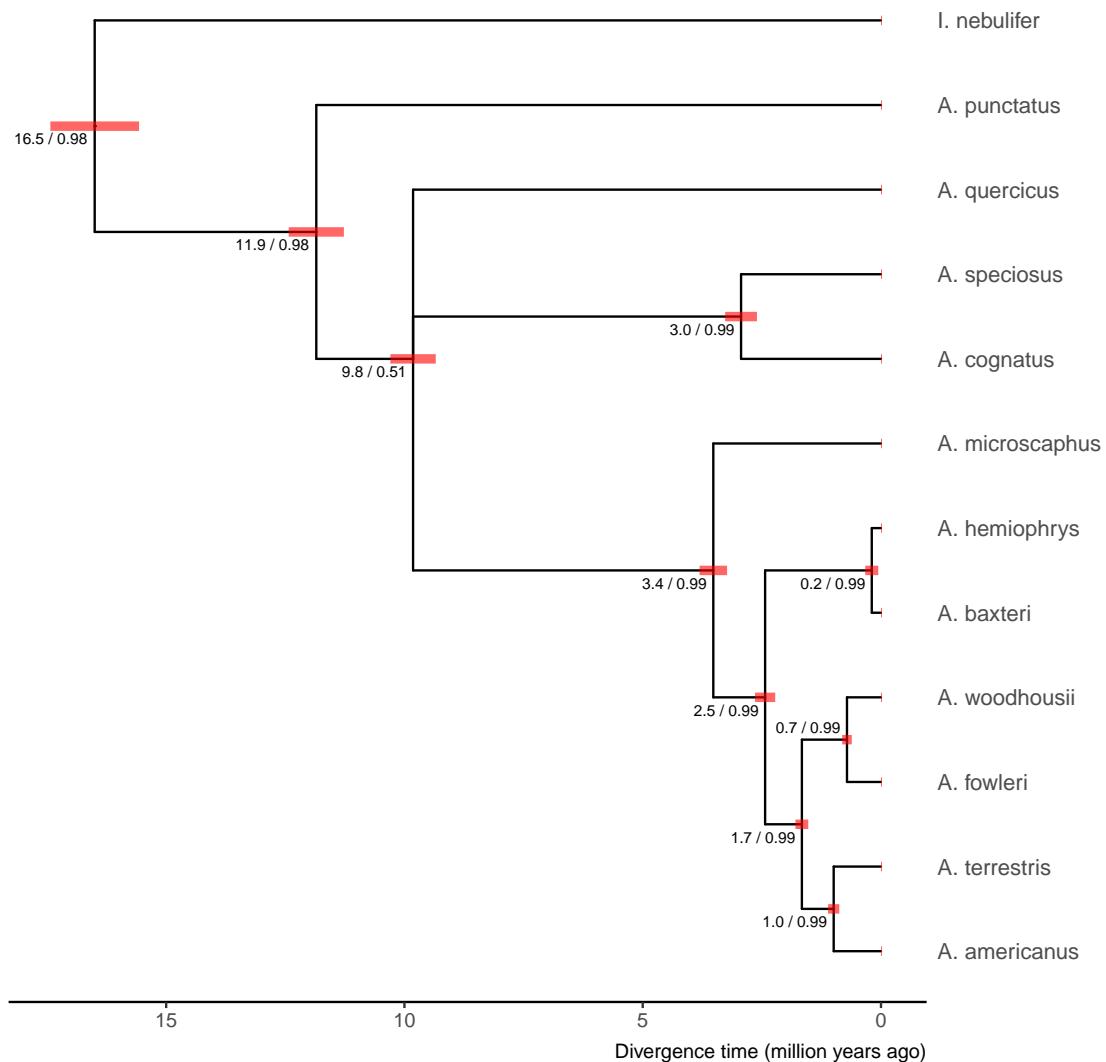


Figure 1.5. The maximum a posteriori tree inferred under a multispecies coalescent model by *phycoeval*. Branch lengths are rescaled from expected substitutions per site to millions of years using secondary time calibrations (*Materials and Methods*). Numbers displayed at each node are the mean posterior 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 (Wickam, 2016), ggtree (Yu et al., 2017), and treeio (Wang et al., 2020)

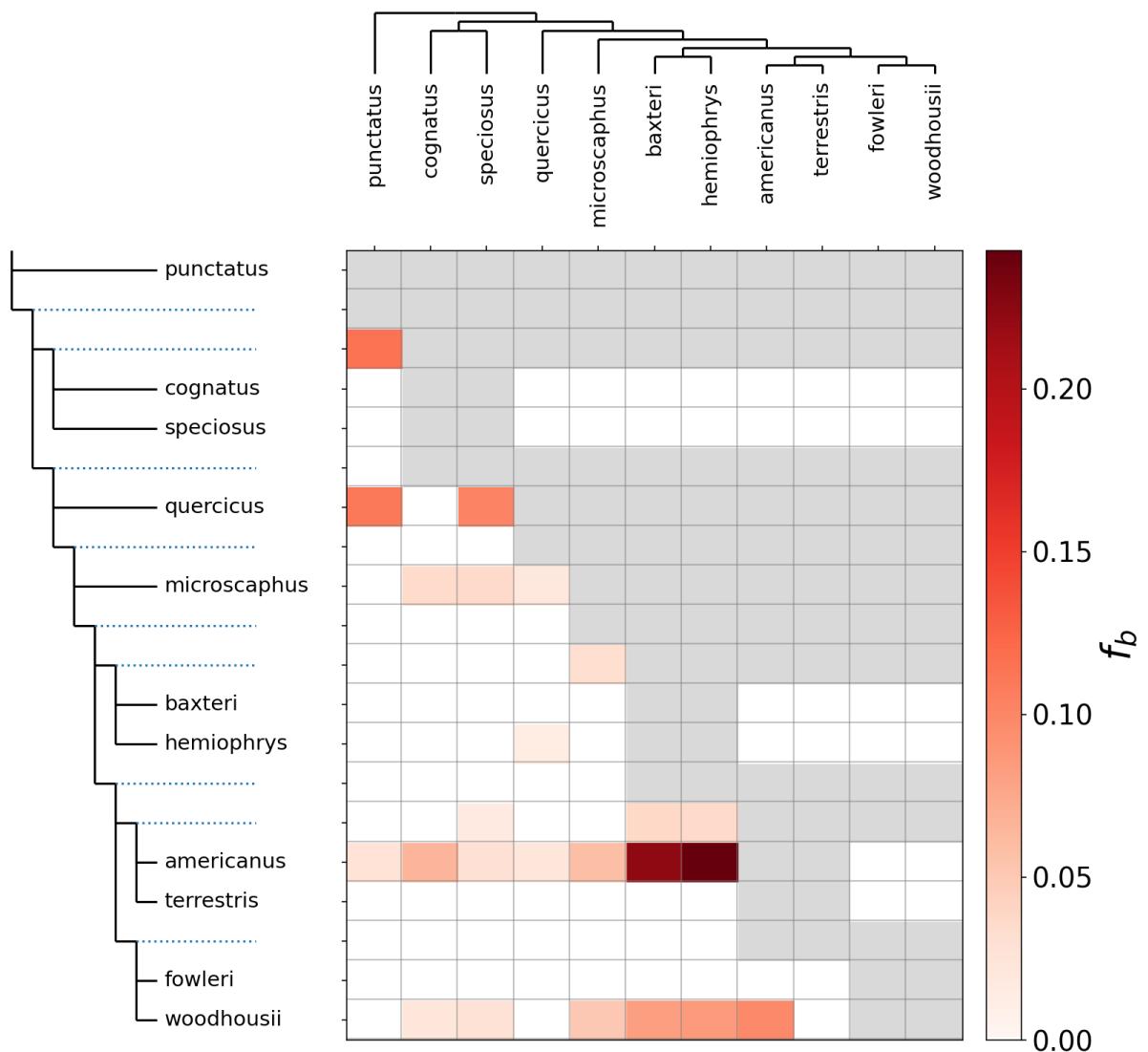


Figure 1.6. Heatmap showing the value of the f -branch statistic computed for all pairs of *Anaxyrus* species. The f -branch statistic indicates the proportion of excess allele sharing between a species on the x-axis and branch on the y-axis (relative to its sister branch). Excess allele sharing between species identifies possible gene flow between them. Grey boxes indicate that the given tips cannot be tested by Dsuite for the given tree topology.

A. americanus Population Structure, K=3

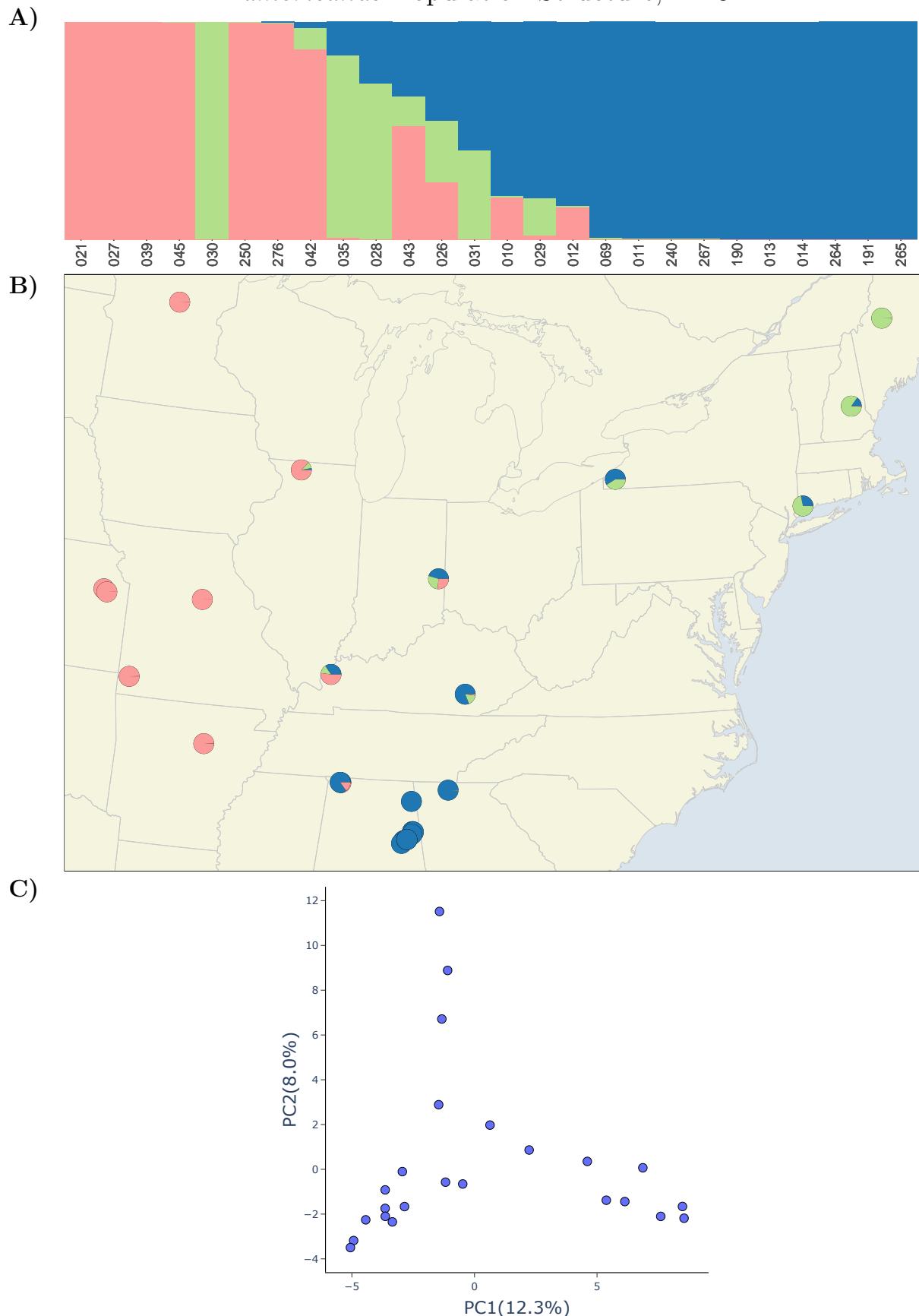


Figure 1.7. *A. americanus* population structure with K=3. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with K=3. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. americanus* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. americanus Population Structure, K=3

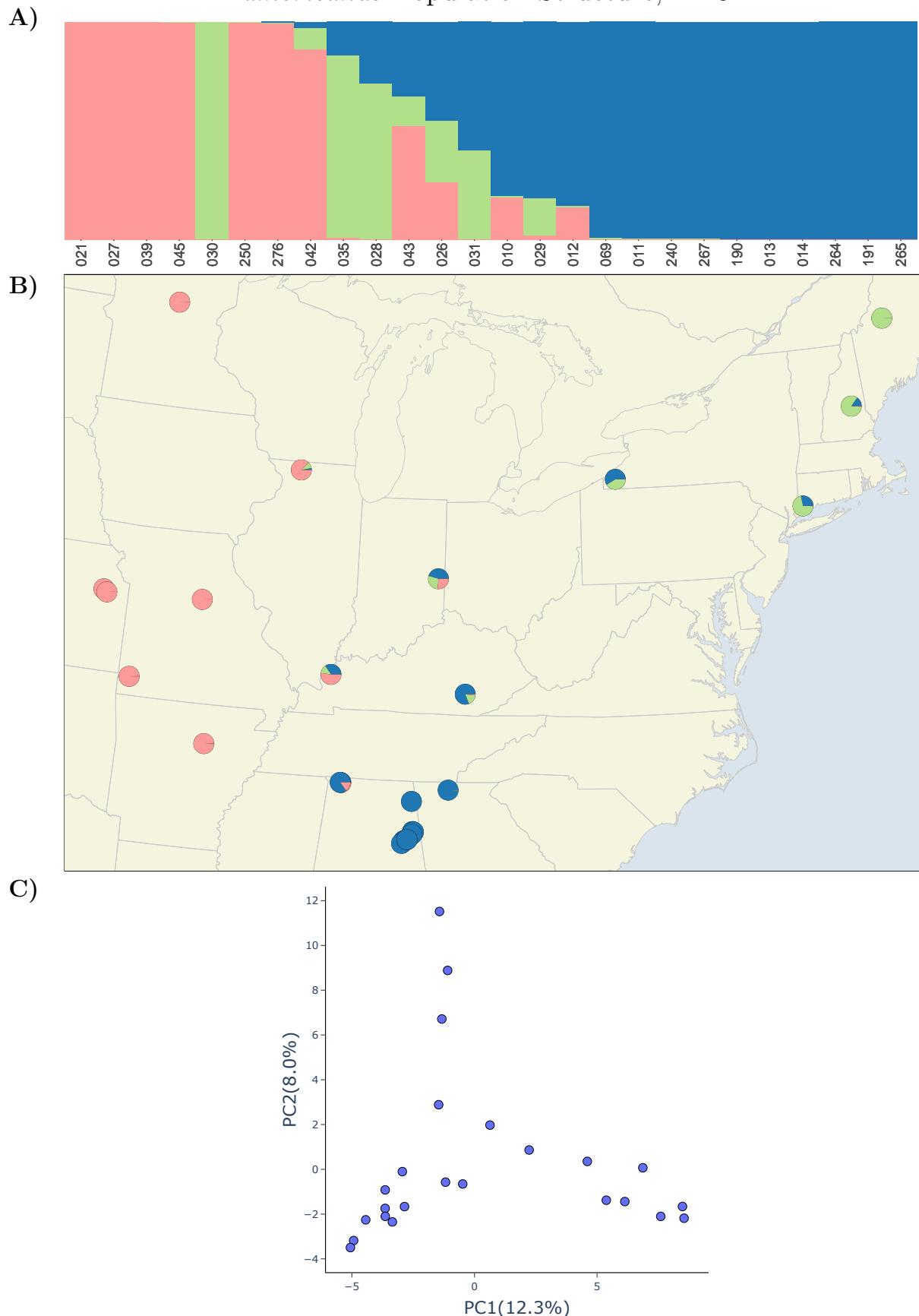


Figure 1.8. *A. americanus* population structure with K=3. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with K=3. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. americanus* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. americanus Population Structure, K=2

A)



B)

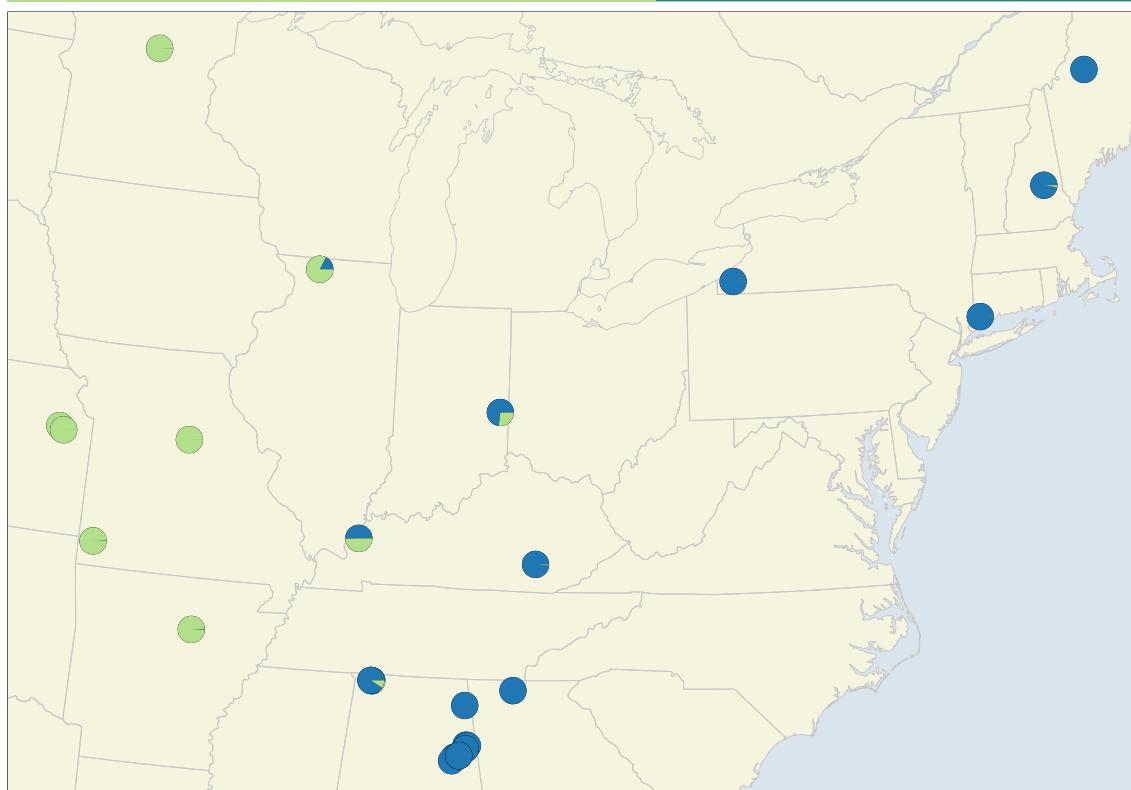
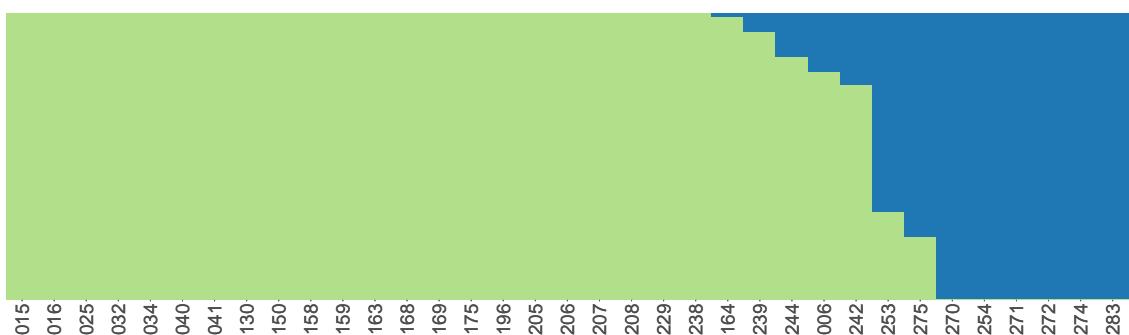


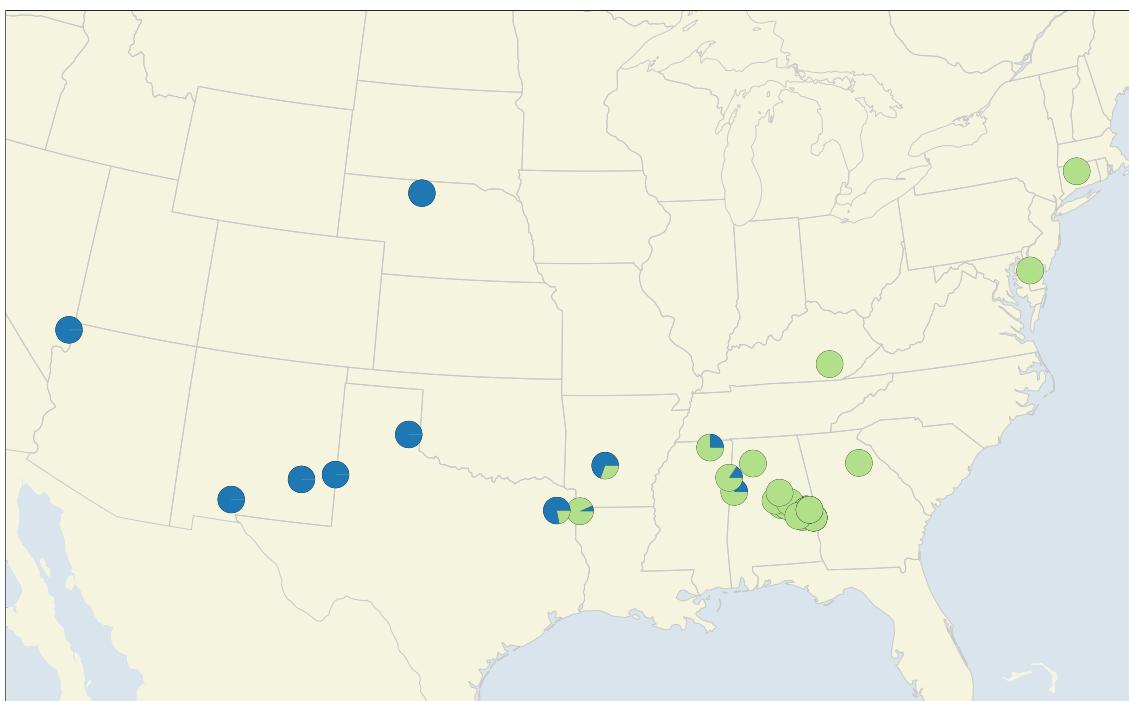
Figure 1.9. *A. americanus* population structure with K=2. A) Barplot with admixture coefficients from the STRUCTURE analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. americanus* samples.

A. fowleri + A. woodhousii Population Structure

A)



B)



C)

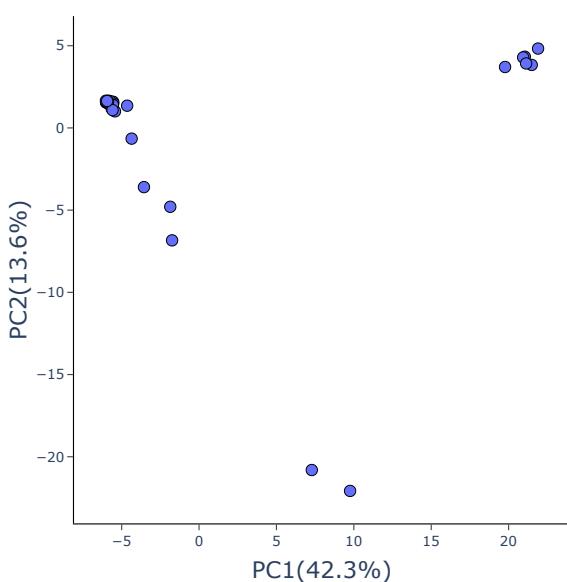


Figure 1.10. Estimates of admixture between *A. fowleri* and *A. woodhousii*. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. woodhousii* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. fowleri Population Structure

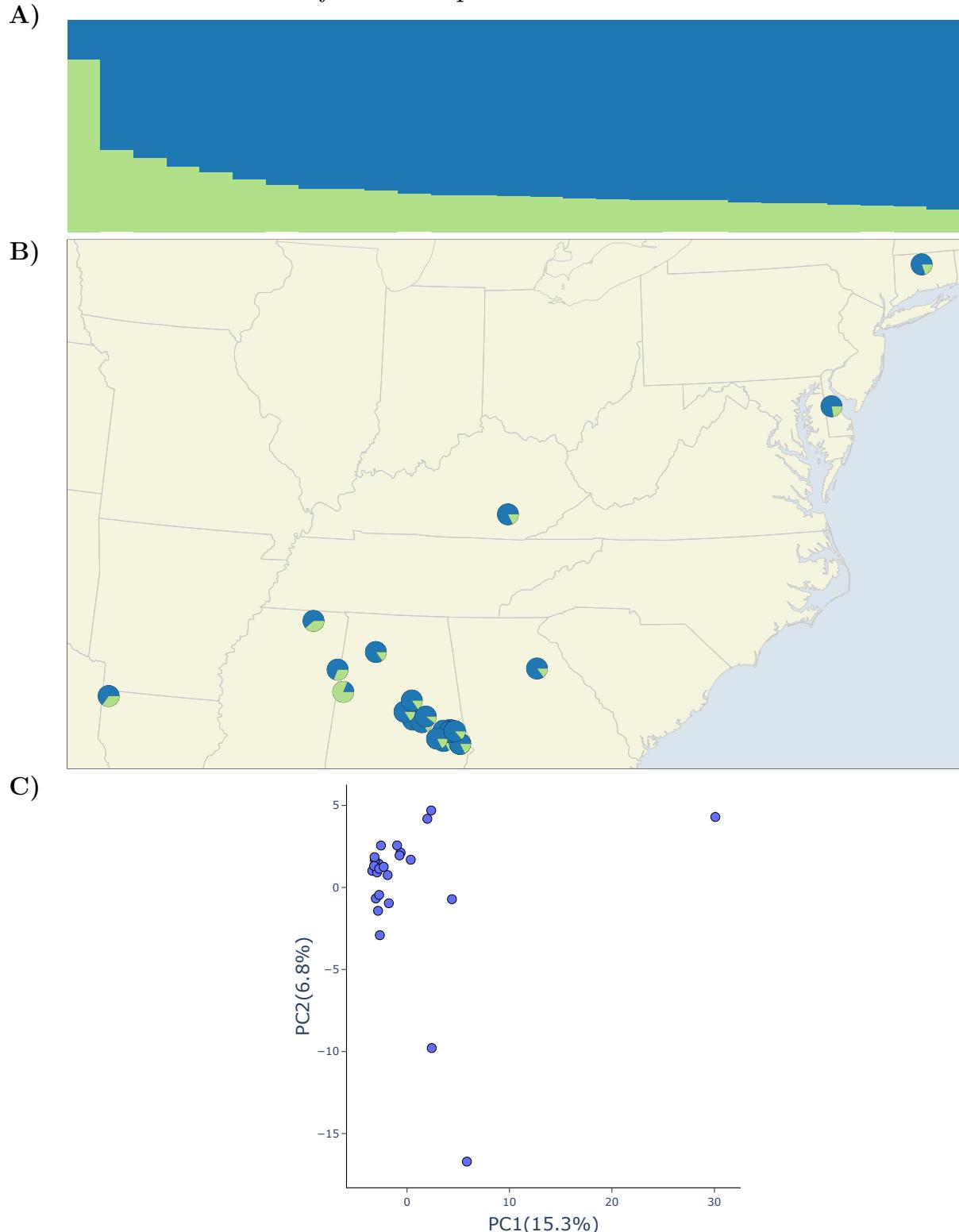


Figure 1.11. *A. fowleri* population structure. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with $K=2$. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. fowleri* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. terrestris Population Structure

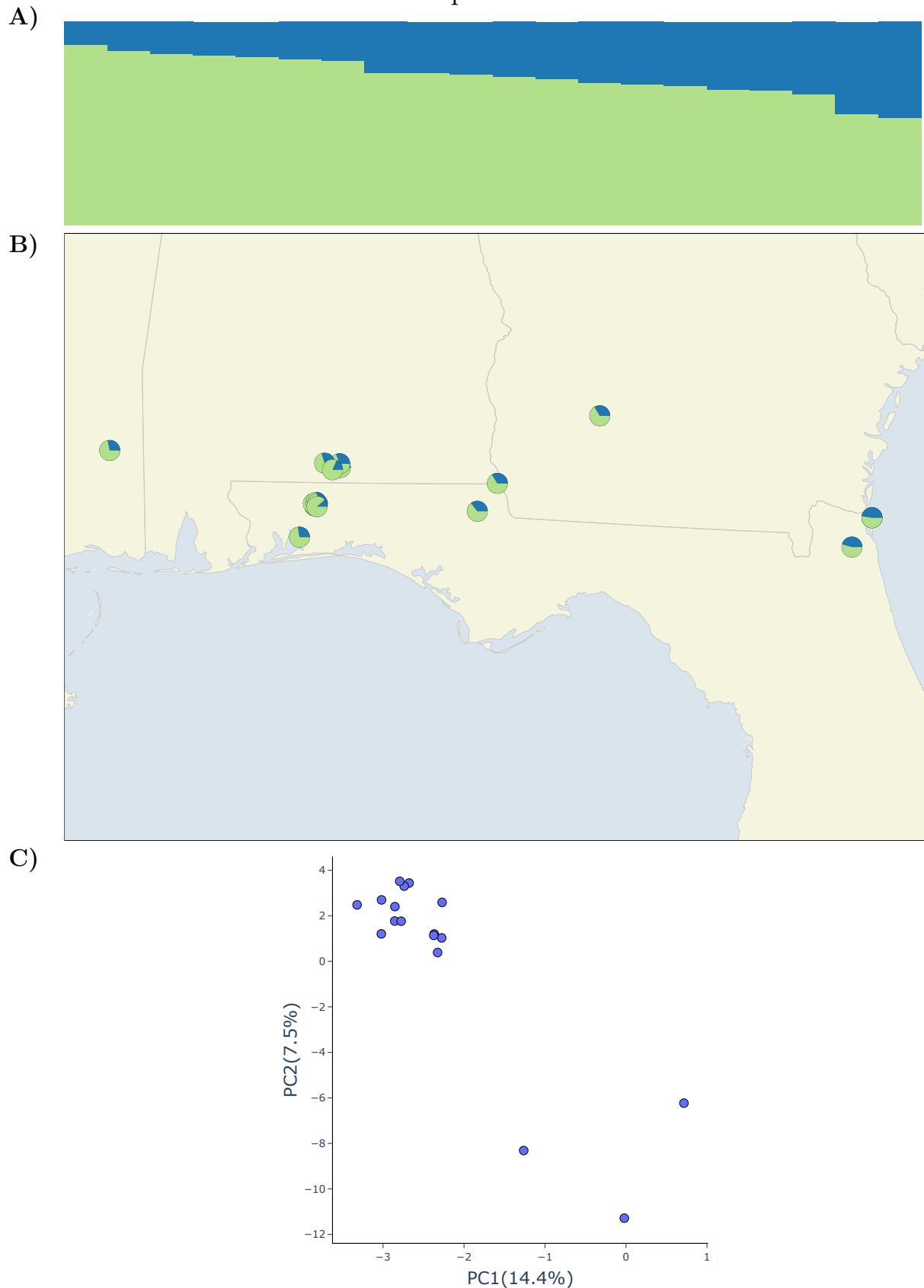


Figure 1.12. *A. terrestris* population structure. A) Barplot with admixture coefficients from the STRUCTURE analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. terrestris* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. woodhousii Population Structure

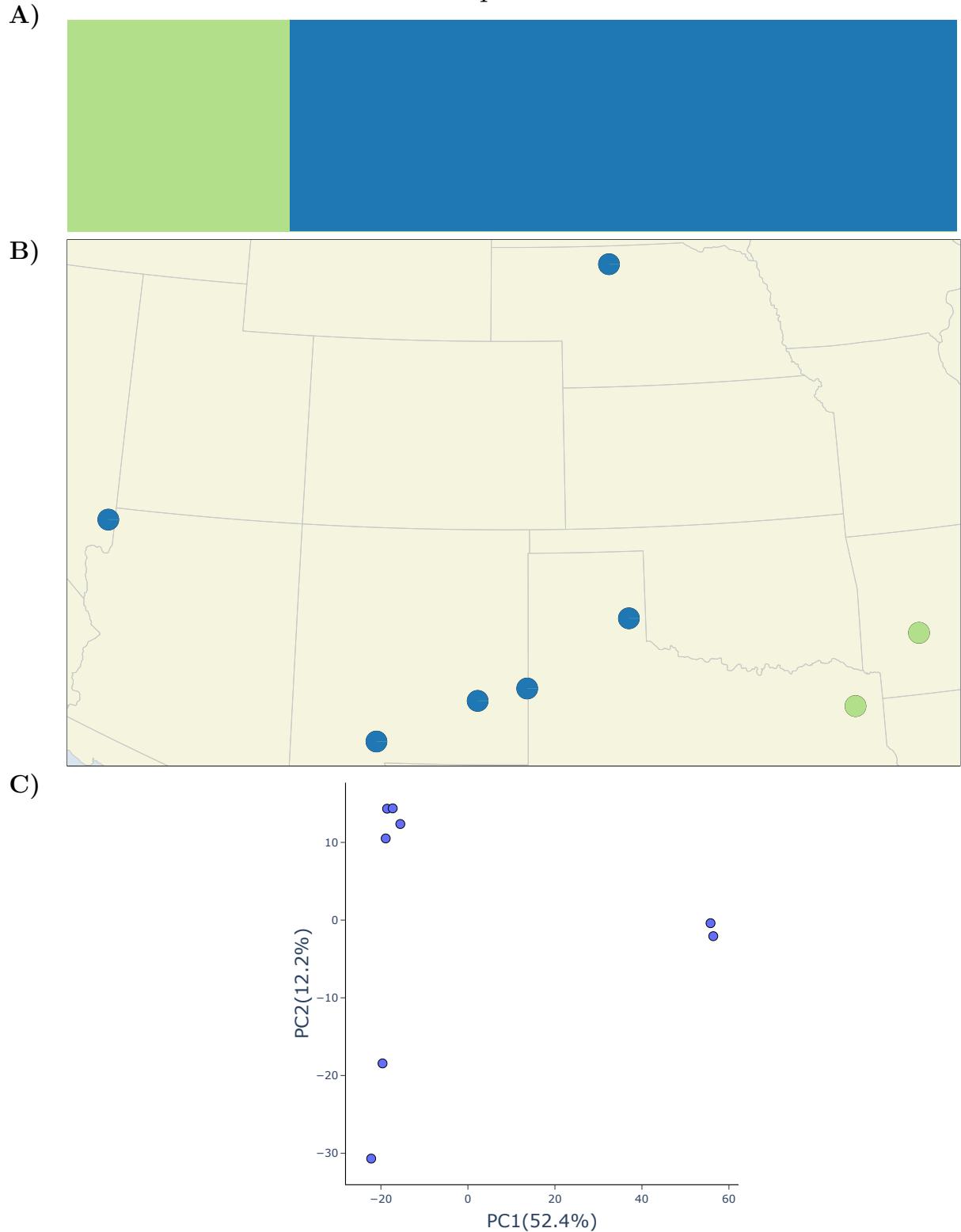


Figure 1.13. *A. woodhousii* population structure. A) Barplot with admixture coefficients from the STRUCTURE analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. woodhousii* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. fowleri + A. woodhousii Population Structure

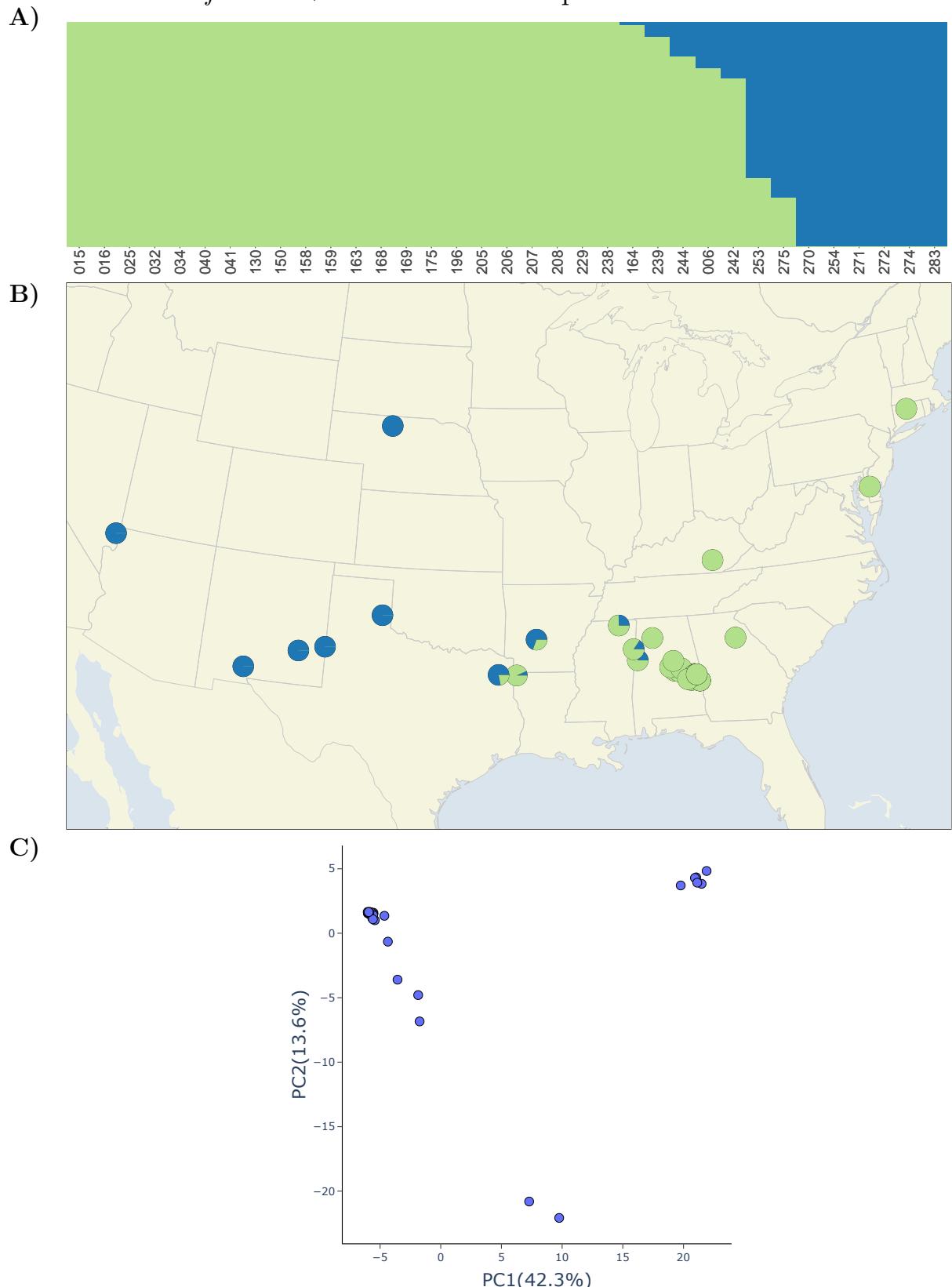


Figure 1.14. Estimates of admixture between *A. fowleri* and *A. woodhousii*. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. woodhousii* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

1.6 Tables

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

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
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
	033	HERA 10484	<i>marina</i>	25.61296	-80.56606		
	034	HERA 10493	<i>fowleri</i>	39.08588	-75.56844	X	X
	035	HERA 11976	<i>americanus</i>	43.51819	-71.42336	X	X
	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

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
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

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
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			
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			

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
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

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
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	
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

Continued on next page

Table 1.1 – continued from previous page

ID	Sample ID	Species	Latitude	Longitude	Passed Filtering	Phycoeval	Structure
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