

Contents

Table of Contents	3
List of Figures	5
List of Tables	6
1 Hybrid Zone	7
1.1 Introduction	7
1.2 Methods	11
1.2.1 Sampling and DNA Isolation	11
1.2.2 RADseq Library Preparation	12
1.2.3 Data Processing	14
1.2.4 Genetic Clustering & Ancestry Proportions	14
1.2.5 Genomic Cline Analysis	15
1.2.6 Genetic differentiation and Introgression	16
1.3 Results	17
1.3.1 Sampling and Data Processing	17
1.3.2 Genetic Clustering & Ancestry Proportions	17
1.3.3 Patterns of Introgression	18
1.3.4 Genomic Differentiation	19
1.4 Discussion	19
1.4.1 Evidence for ongoing hybridization	19
1.4.2 Variability of introgression	21

1.4.3	Relationship between introgression and differentiation	23
1.4.4	Conclusion	23
References	24
1.5	Figures	31
1.6	Tables	37
2	Evolutionary History	45
2.1	Introduction	45
2.2	Methods	49
2.2.1	Sampling and DNA Isolation	49
2.2.2	RADseq Library Preparation	50
2.2.3	Phylogenetic Data Processing	51
2.2.4	Maximum Likelihood	51
2.2.5	Multispecies Coalescent	52
2.2.6	Introgression	53
2.2.7	Population Structure	53
2.3	Results	54
2.3.1	Assembly and alignment with <i>ipyrad</i>	54
2.3.2	Maximum Likelihood Phylogeny	55
2.3.3	Coalescent Phylogeny	55
2.3.4	Introgression	56
2.3.5	Population Structure	57
2.4	Discussion	58
2.4.1	Phylogenetic relationships	58
2.4.2	Divergence Time	60
2.4.3	Hybridization	62
2.4.4	Population Structure	65
2.4.5	Conclusion	67
References	67
2.5	Figures	74

2.6 Tables	97
----------------------	----

List of Figures

1.1.	Evanno method for optimal value of K in <i>STRUCTURE</i>	31
1.2.	<i>STRUCTURE</i> iterations	32
1.3.	Summarized <i>STRUCTURE</i> results for each value of K.	33
1.4.	Genetic evidence of hybridization between <i>A. americanus</i> and <i>A. terrestris</i>	34
1.5.	Shape of genomic clines	35
1.6.	Patterns of genomic divergence.	35
1.7.	Relationship between genetic divergence and introgression.	36
2.1.	Distribution map of <i>americanus</i> group samples	75
2.2.	Maximum likelihood tree part 1	76
2.3.	Maximum likelihood tree part 2	77
2.4.	Simplified maximum likelihood tree	78
2.5.	Multispecies coalescent phylogeny	79
2.6.	<i>f</i> -branch statistics	80
2.7.	<i>A. americanus</i> population structure, K=3	81
2.8.	<i>A. americanus</i> population structure, K=2	82
2.9.	<i>A. fowleri</i> population structure	83
2.10.	<i>A. terrestris</i> population structure	84
2.11.	<i>A. woodhousii</i> population structure	85
2.12.	Estimate of admixture between <i>A. fowleri</i> and <i>A. woodhousii</i>	86
2.13.	Delta K for <i>americanus</i>	87
2.14.	Delta K for <i>fowleri</i>	88
2.15.	Delta K for <i>terrestris</i>	89

2.16. Delta K for woodhousii	90
2.17. Delta K for woodhousii x fowleri	91
2.18. All <i>STRUCTURE</i> runs americanus.	92
2.19. All <i>STRUCTURE</i> runs fowleri.	93
2.20. All <i>STRUCTURE</i> runs terrestris.	94
2.21. All <i>STRUCTURE</i> runs woodhousii.	95
2.22. All <i>STRUCTURE</i> runs fowleri x woodhousii.	96

List of Tables

1.1	Samples collected for this study	37
1.2	Samples loaned from museums	44
2.1	Samples used in this study	98

Chapter 1

Hybrid Zone

1.1 Introduction

Speciation is the process by which genetic divergence leads to reproductive isolation between divergent lineages. It is a continuous process during which there may be ongoing gene flow or introgression via hybridization following a period of isolation and subsequent secondary contact (Mallet, 2008; Wu, 2001). Introgression is possible because genetic barriers to introgression that accumulate within the genome are a property of genomic regions rather than a property of the entirety of the genome (Gompert, Parchman, et al., 2012; Wu, 2001). Natural hybridization between divergent lineages has become increasingly appreciated as a widespread phenomenon in recent years (Mallet, 2005; Moran et al., 2021). It is a phenomenon that can have important evolutionary consequences. Hybridization can be a source of adaptive variation (Hedrick, 2013). It can also introduce deleterious genetic load which persists long term within a population (Moran et al., 2021). Hybridization can create conditions where selection favors the evolution of traits that enhance assortative mating and reduce the production of unfit hybrid offspring which drives further genetic divergence and reinforcement of reproductive barriers between lineages (Servedio & Noor, 2003). If hybrids do not suffer any negative fitness effects, hybridization could lead to the erosion of differences between divergent populations (Taylor et al., 2006). Potentially resulting in populations that are genetically distinct

from either parent species which can themselves eventually evolve reproductive isolation from the parent species (Moran et al., 2021).

Aside from having important evolutionary consequences which need to be understood, hybridization is also an excellent opportunity to investigate the processes that result in the evolution of reproductive incompatibility and divergence between evolutionary lineages. Hybrid zones are particularly suitable due to the production of a large numbers of recombinant genomes carrying many possible combinations of genomic elements from parent species resulting from many generations of backcrossing (Rieseberg et al., 1999). The many generations of backcrossing and recombination make it possible to distinguish between the effects of closely linked genes (Rieseberg et al., 1999). The many generations and large number of individuals producing these genetic combinations are not feasible to produce experimentally in the vast majority of species (Rieseberg et al., 1999). Furthermore, the combination of genes produced are exposed to selection under natural conditions. This is important as the effect of hybrid incompatibilities can be dependent on environmental conditions and can only truly understood in this context (Miller & Matute, 2016).

Despite being a fundamental evolutionary process, our understanding of speciation is far from complete (Butlin et al., 2011). Only a few loci, in a few species, have been pinpointed as the direct cause of reproductive incompatibility between species (Blackman, 2016; Nosil & Schlüter, 2011). Consequently, our understanding of the processes that drive the evolution of loci resulting in reproductive incompatibility is limited (Butlin et al., 2011). Studies of introgression within hybrid zones have identified highly variable rates of introgression among loci (Barton & Hewitt, 1985; Gompert et al., 2017). This heterogeneity can arise via genetic drift occurring within hybrid zones, but will also be caused by differences among loci in the strength of selection against them in a hybrid genomic background (Barton & Hewitt, 1985; Gompert et al., 2017). It has also been observed that the levels of genetic divergence between species are highly variable across the genome (Nosil et al., 2009). Much of this heterogeneity is the result of divergent selection acting on each species independently (Nosil et al., 2009). Regions with particularly high

levels of divergence between closely related species have been coined "genomic islands of divergence" (Wolf & Ellegren, 2017). It is assumed, particularly in the case of speciation with gene flow, that these genomic island harbor genes that reduce interbreeding between species. When speciation occurs with gene flow, divergent selection can cause adaptive divergence in habitat use, phenology, or mating signals, and reduce the frequency or success of interspecific matings. When species diverge in geographic isolation, divergent selection and reproductive isolation could be decoupled and reproductive isolation could just be a result of genetic drift. Whether loci under divergent selection between two species also contribute to reproductive isolation has not been widely explored. A handful of studies have found evidence for a modest relationship between genetic divergence and selection against introgression (Gompert, Lucas, et al., 2012; Larson et al., 2013; Nikolakis et al., 2022; Parchman et al., 2013). How consistent and widespread this pattern is remains to be seen. At least one study has found no association (Jahner et al., 2021).

In this study I investigate hybridization between the American toad (*Anaxyrus americanus*) and Southern toad (*Anaxyrus terrestris*) at a suspected hybrid zone in the Southern United States to assess the extent of introgression between them and test for a relationship between introgression and genetic divergence. This suspected hybrid zone has not been investigated with genetic data previously but it bears many hallmarks of a tension zone (Barton & Hewitt, 1985). Under the tension zone model of hybridization, species boundaries are maintained by a balance between dispersal and selection against individuals carrying incompatible hybrid genotypes (Barton & Hewitt, 1985). The ranges of *A. americanus* and *A. terrestris* abut with an abrupt transition and no apparent overlap along a long contact zone which from Louisiana to Virginia. This contact zone closely corresponds with a prominent physiographic feature known as the "fall line" (Mount, 1975; Powell et al., 2016). The Fall line is the boundary between the Southern coastal plain to the South and the Appalachian Highlands to the North (Shankman & Hart, 2007). These regions differ in their underlying geology, topography, and elevation (Shankman & Hart, 2007). The distribution of *A. terrestris* is restricted to the coastal plain extending from the Mississippi River in the West to Virginia in the East (Fig. 1.4).

The distribution of the American Toad encompasses nearly all of the Eastern North American with the exception of the Southern coastal plain (Fig. 1.4). Tension zones are expected to correspond with natural features that reduce dispersal or abundance (Barton, 1979). Such a sudden transition is difficult to explain if not the result of the processes characteristic of tension zones. For there to be no mutually hospitable areas permitting some range overlap is implausible without there being an extreme level of competition or extreme degree of adaptation by each species to their respective environments. The two species have only slight differences in male advertisement call and in morphological appearance (Cocroft & Ryan, 1995; Weatherby, 1982). They only differ slightly in the timing of their spawn (Mount, 1975) However, there is some overlap in the spawning period and male Bufonidae are famously indiscriminate in their choice of mates (Đorđević & Simović, 2014; Weatherby, 1982). They have also been shown to have a degree of reproductive compatibility through laboratory crossing experiments which produced viable F_2 offspring (Blair, 1963). Analysis of morphological variation in central Alabama by Weatherby, 1982 suggests there has been introgression between them.

The "true toads" in the family Bufonidae, to which *A. americanus* and *A. terrestris* belong, have been a prominent group of organisms in the literature on hybridization. W.F. Blair and colleagues performed a remarkable 1,934 separate experimental crosses to quantify the degree of reproductive incompatibility between species pairs within this family (Blair, 1972; Malone & Fontenot, 2008). These experiments demonstrated a high degree of compatibility between some closely related species pairs in which hybrids were capable of producing viable backcross or F_2 hybrid offspring (Blair, 1963). Furthermore, numerous cases of natural hybridization among toad species have been reported with several apparent or clear hybrid zones (Colliard et al., 2010; Green, 1996; Van Riemsdijk et al., 2023; Weatherby, 1982). Despite the interest in and appreciation for hybridization in Bufonidae, only a small amount of work has been done to understand patterns of introgression within hybrid zones. A clinal pattern of admixture at 26 allozyme loci has been shown within the *Anaxyrus americanus* X *Anaxyrus hemiophrys* hybrid zone in Ontario, Canada(Green, 1983). Almost no admixture was detected at 7 microsatellite loci

within the suspected *Bufo siculus* X *Bufo balearicus* hybrid zone in Sicily, Italy (Colliard et al., 2010). The most comprehensive study of introgression within a Bufonidae hybrid zone found significant levels of genome wide admixture, fitting a clinal pattern, at two separate transects at either end of the *Bufo bufo* x *Bufo spinosus* hybrid zone in Southern France (Van Riemsdijk et al., 2023).

The suspected *A. americanus*, *A. terrestris* hybrid zone has great potential to expand our understanding of speciation. This will be dependent on the degree of ongoing introgression, if any, between these species. In this study I use genome-wide sequence data to characterize patterns of introgression within the hybrid zone using model based inference of admixture proportions, bayesian genomic cline analysis, and estimates of parental population differentiation. With these approaches I specifically address the following questions: 1) Is there evidence of ongoing hybridization and admixture between the two species, 2) Do any loci have outstanding patterns of introgression consistent with them being linked to reproductive incompatibility, and 3) Is there any relationship between patterns of introgression and levels of genetic differentiation between parental lineages?

1.2 Methods

1.2.1 Sampling and DNA Isolation

I collected genetic samples from *A. americanus* and *A. terrestris* by driving roads during rainy nights between 2017 and 2020 in a region of central Alabama where hybridization has previously been inferred from the presence of morphological intermediates (Weatherby, 1982). I euthanized individuals with immersion in buffered MS-222. I removed liver and/or toes and preserved them in 100% ethanol and fixed specimens with 10% Formalin solution. Genetic samples and formalin fixed specimens were deposited at the Auburn Museum of Natural History. Additional samples were also provided by museums (see ??).

I isolated DNA by lysing a small piece of liver or toe approximately the size of a grain

of rice in 300 μ L of a solution of 10mM Tris-HCL, 10mM EDTA, 1% SDS (w/v), and nuclease free water along with 6 mg Proteinase K and incubating for 4-16 hours at 55°C. 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 removing and 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 thoroughly mixing the beads with 100 μ 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, collected the TLE solution, and discarded the beads. I quantified DNA in the TLE solution with a Qubit fluorometer (Life Technologies, USA) and diluted samples with additional TLE solution to bring the concentration to 20 ng/ μ L.

1.2.2 RADseq Library Preparation

I prepared RADseq libraries using the 2RAD approach developed 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 and combined the two pools of cleaned ligation product.

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 which were 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 Data Processing

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. To remove PCR duplicates, I used the *clone_filter* command from *Stacks*. I demultiplexed inline sample barcodes, trimmed adapter sequence, and filtered reads with low quality scores as well as reads with any uncalled bases using the *process_radtags* command again and allowed for the rescue of restriction site sequence as well as barcodes with up to two mismatches. I built alignments from the processed reads using the *Stacks* pipeline. 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.

1.2.4 Genetic Clustering & Ancestry Proportions

To cluster individuals and characterize patterns of genetic differentiation and admixture between clusters, I used the Bayesian inference program *STRUCTURE* v2.3.4 (Pritchard et al., 2000) with *STRUCTURE*'s admixture model which returns an estimate of ancestry proportions for each sample. To evaluate the assumption that samples are best modeled as inheriting their genetic variation just two groups corresponding to the species identification made in the field, I ran *STRUCTURE* under four different models,

each with a different number of assumed clusters of individuals (K parameter) ranging from 1 to 4. For each value of K, I ran 20 iterations for 100,000 total steps with the first 50,000 as burnin. 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 ΔK which is the the model that has the greatest increase in likelihood score from the model with one fewer populations as described by (Evanno et al., 2005). I also examined genetic clustering and evidence of admixture using a non-parametric approach with a principal component analysis (PCA) implemented in the R package *adegenet* v2.1.10 (Jombart, 2008). I visualized the relationship between the first principal component axis and the estimated admixture proportion for each individual to check for agreement between the parametric *STRUCTURE* analysis and the non-parametric PCA analysis.

1.2.5 Genomic Cline Analysis

To investigate patterns of introgression across the hybrid zone I used the bayesian genomic cline inference tool *BGC* v1.03 (Gompert & Buerkle, 2012) to infer parameters under a genomic cline model. Explain a bit how *BGC* works??? . . . I classified a sample as being admixed if it had an inferred admixture proportion of <95% for one species under the model with a K of two in the *STRUCTURE* analysis. I used *VCFtools* vXX.XX.XX to filter all non-biallelic sites from the the VCF file produced by the *populations* command in *Stacks*. I converted the VCF formatted data into the *BGC* format using *bgc_utils* v0.1.0, a *Python* package that I developed for this project github.com/kerrycobb/bgc_utils. I ran *BGC* with 5 independent chains, each for 1,000,000 steps and sampling every 1000. I visualized MCMC output, discarded samples from the posterior as burnin, combined the independent chains, summarized the posterior samples, and identified outlier loci with *bgc_utils*. A primary goal of *BGC* analysis is to identify loci which have exceptional patterns of introgression. These loci, or loci in close linkage to them, are expected to be enriched for genetic regions affected by selection due to reproductive incompatibility between the two species. I identified loci with exceptional patterns of introgression using two approaches described by Gompert and Buerkle, 2011. (1) If locus specific introgres-

sion differed from the genome-wide average which I will refer to as "excess ancestry" following Gompert and Buerkle, 2011. More specifically, I classified a locus as having excess ancestry if the 90% highest posterior density interval (HPDI) for the alpha or beta parameter did not cover zero. (2) If locus specific introgression is statistically unlikely relative to the genome-wide distribution of locus specific introgression which I will refer to as "outliers" following Gompert and Buerkle, 2011. I classified a locus as an outlier if the median of the posterior sample for the α or β parameters for a locus were not contained the interval from 0.05 to 0.95 of the probability density functions $Normal(0, \tau_\alpha)$ or $Normal(0, \tau_\beta)$ respectively, where τ_α and τ_β are the median values from the posterior sample for the conditional random effect priors on τ_α and τ_β . These conditional priors describe the genome-wide variation of locus specific α and β . I further classified outlier α parameter estimates for a locus based on whether the median of the posterior sample was positive or negative. Positive estimates of α mean there is a greater probability of *A. americanus* ancestry in individuals at the locus relative to their hybrid index whereas negative estimates of α mean there is a greater probability of *A. terrestis* ancestry.

1.2.6 Genetic differentiation and Introgression

To test for a relationship between patterns of introgression and genetic divergence I used *VCFtools* to calculate the Weir and Cockerham, 1984 F_{ST} between each species using only the samples inferred through the *STRUCTURE* analysis to have >95% ancestry for one species under the model with a K of two (Danecek et al., 2011). The Weir and Cockerham F_{ST} is calculated per site and I calculated the per site F_{ST} for the same sites as those used in the *BGC* analysis. To determine if patterns of introgression are correlated with population differentiation a performed a Pearson Correlation to test if F_{ST} correlates with either the α or β parameters. I ran the correlation test with the absolute value of the median of the posterior sample for the α parameter and the median of the posterior sample for the β parameter. I binned loci based on their status as α parameter outliers. I categorized loci as being α outliers greater expected *A. americanus* ancestry, α outliers with greater expected *A. terrestis* ancestry, and estimates of α

that are not outliers. I performed a Kruskal-Wallis test using *SciPy* v1.10.1 to test whether there were significant differences in values of F_{ST} at each locus between these groups (Virtanen et al., 2020). I then performed Mann-Whitney tests between all pairs of groups using *scikit-posthocs* to test which groups differ significantly from each other (github.com/maximtrp/scikit-posthocs).

1.3 Results

1.3.1 Sampling and Data Processing

I prepared reduced-representation sequencing libraries from 173 samples collected for this study (Table 1.1) and 19 samples available from existing collections (Table 1.2)). The *Stacks* pipeline assembled reads into 432,336 loci with a mean length of 253.31 bp. Prior to filtering the mean coverage per sample was 32X. After filtering loci missing from greater than 5% of samples, filtering sites with minor allele counts less than 3, filtering individuals with greater than 90% missing loci, and randomly sampling a single SNP from each locus, 1194 sites remained and 43 samples were excluded from further analyses leaving a total of 149. For the included samples, 56 had been identified as most closely resembling *A. americanus* and 93 had been identified as most closely resembling *A. terrestris*.

1.3.2 Genetic Clustering & Ancestry Proportions

A visual inspection of the *STRUCTURE* results shows that each iteration with same value for K converged on very similar results (Fig. 1.2). The *STRUCTURE* model with the largest ΔK was the model with a K of two (Fig. 1.1). Furthermore, individuals are inferred as having ancestry derived largely from only two ancestral groups even for K values of three and four. For these values of K, only a small amount of ancestry is attributed to the third or fourth ancestral groups for any individual sample (Fig. 1.3) Using a 95% estimated ancestry proportion as a cutoff for considering individuals to have pure ancestry, 36 samples were classified as pure *A. americanus*, 75 as pure *A. terrestris*,

and 38 as being admixed. The proportions of admixture among the samples shows a clear gradient between 0 and 1 which is consistent with many individuals being the product of advanced generation hybrids beyond the F_1 generation. The transition of admixture proportions from one species to the other increase with distance from the locations of pure individuals with proportions closest to 0.5 being found in the center of this transition (Fig. 1.4).

1.3.3 Patterns of Introgression

Visualization of the MCMC output with trace plots and histograms of each parameter indicated that each of the five chains run in *BGC* converged on the same parameter space and that each chain quickly reached stationarity. I conservatively discarded the first 10% of samples as burnin. The median of the posterior sample for α ranged from -0.525-0.494. The β parameter was less variable and ranged from -0.158-0.220. I identified 16 loci with excess ancestry for the α parameter relative to the genome wide average; i.e., the 90% HDPI does not cover 0. Of these, the median of the posterior sample for 5 of these loci was negative and for 11 loci was positive. Negative values represent a greater probability of *A. americanus* ancestry at a locus relative to the hybrid index whereas positive values represent a greater probability of *A. terrestris* ancestry. I did not identify any loci for which the estimates of β were outliers relative to the genome-wide average. I identified 116 loci as outliers for the α parameter relative to the genome-wide distribution of locus specific introgression. Of these, the median of the posterior sample for 24 of these loci was negative and for 92 loci was positive (Fig. 1.5). I did not identify any loci for which the estimates of β were outliers relative to the genome-wide distribution of locus specific introgression. All 16 of the loci identified as having excess ancestry for the α parameter relative to the genome-wide average were also identified as outliers relative to the genome-wide distribution of locus specific introgression.

1.3.4 Genomic Differentiation

Genetic differentiation between *A. americanus* and *A. terrestris* was highly variable (Fig. 1.6). Locus-specific F_{ST} between non-admixed *A. americanus* and *A. terrestris* had a mean of 0.07. F_{ST} values for 249 loci were 0. Only a single locus had fixed differences between species with an F_{ST} of 1.0. There is little apparent relationship between α or β and F_{ST} except at that the highest α and β estimates have non-zero F_{ST} estimates (Fig. 1.7). The Pearson correlation test estimates a weak correlation between α and F_{ST} ($r=0.29$, $p=1.62e-23$) a week correlation between β and F_{ST} ($r=0.32$, $p = 8.28 \times 10^{-30}$). The result of the Kruskal-Wallis test are consistent with there being significant differences between the F_{ST} values of loci with outlier α estimates and non-outlier α estimates on average ($p = 1.32 \times 10^{-40}$) (Fig. 1.6). The results of the post hoc pairwise Mann-Whitney tests are consistent with both categories of loci with outlier α estimates having greater F_{ST} values on average than the non-outlier estimates of α . The difference between non-outlier loci and loci with greater probability of *A. americanus* ancestry was slightly higher ($p = 2.72 \times 10^{-38}$) than the difference between non-outlier loci and loci with greater *A. terrestris* ancestry ($p = 8.16 \times 10^{-6}$).

1.4 Discussion

1.4.1 Evidence for ongoing hybridization

With the genome-wide sequence data obtained in this study, I find evidence of substantial gene flow across the hybrid zone of these two species. The *STRUCTURE* analysis inferred 38 out of 149 samples as having a proportion of ancestry of at least 5% of sites attributable to admixture (Fig. 1.4). The admixture proportions inferred in the *STRUCTURE* analysis range from 0.05%-0.5% which is consistent with hybrids being viable, fertile, and capable of backcrossing over multiple generations [CITATION NEEDED!] (Fig. 1.4). When backcrossing occurs over multiple generations in combination with migration of hybrid progeny and selection against introgressing alleles, a cline will form across the hybrid zone with introgressing alleles becoming more uncommon with distance

from the cline center (Barton & Hewitt, 1985). The results of the *STRUCTURE* analysis are largely consistent with this. Inferred admixture coefficients are highest at the center of the hybrid zone and decrease and approach zero with distance from the center (Fig. 1.4).

Admixed samples were located quite far from the center of the hybrid zone. In fact samples with greater than 5% admixture proportions are located all the way at the North-eastern and Southwestern edges of the sampling area. The width of a hybrid zone is a product of the strength of selection for or against introgression and the average dispersal distance of individuals within their reproductive lifespan (Barton & Hewitt, 1985). Breeden, 1987 estimated that 27% of individual *A. fowleri* breed at non-natal breeding ponds with some individuals dispersing at least as much as 2 km. Female *A. americanus* can migrate at least 1 km between breeding sites and post-breeding locations (Forester et al., 2006) Invasive cane toads (*Rhinella marina*) in Australia are estimated to have expanded their range at a rate of 10-15 km per year shortly after their introduction (Urban et al., 2008). The presence of samples with little to no admixture in close proximity to toads with high proportions of admixture shows that dispersal has an important roll in shaping the patterns of this hybrid zone. Individuals would be expected to appear more like their neighbors if dispersal rates and distances were very low. It is also likely that this hybrid zone may be more appropriately described as a mosaic hybrid zone rather than a more simple tension zone (Harrison, 1986). However, the sampling for this study is too sparse and irregular to definitively test this. Another possibility is that some of this inferred admixture is the result of a statistical artifact or due to error. Some reassurance is provided by the result of the PCA which is largely consistent with the *STRUCTURE* results although it is possible that they could be affected by the same bias or error introduced at by data collection and processing (Fig. 1.4) [CITATION NEEDED!].

The tension zone model of hybrid zones predicts that location of hybrid zones centers will be dependent on the effects of selection along with population density and natural dispersal barriers (Barton, 1979). The *STRUCTURE* results show that in two areas, there is a clear transition from samples with primarily *A. americanus* ancestry to samples with

primarily *A. terrestris* ancestry corresponding with the locations of streams and rivers. In the Northern part of the sampling area, transitions occur at the Coosa River and at Waxahatchee Creek (Fig. 1.4). In the Southern part, they occur at Sougahatchee Creek (Fig. 1.4). Clearly these are not impassable boundaries as there has been introgression beyond them. However, they likely reduce dispersal and as a result that the center of the hybrid zone is caught in this location as described by Barton, 1979.

1.4.2 Variability of introgression

There are two primary parameters of interest in a genomic cline model that can be interpreted in the evolutionary context of hybrid zones. The α parameter specifies the center of the cline and is dependent on the increase or decrease in the probability of locus-specific ancestry from one of the parental populations. The β parameter specifies the rate of change in probability of ancestry along the genome-wide admixture gradient. Extreme estimates of these parameters may be associated with loci that cause reproductive incompatibility between hybridizing species. The Bayesian genomic cline analysis of the genome-wide data in this study yielded extreme estimates for α at some sites. Sites were classified as having extreme values in two ways. First, sites could be classified as having excess ancestry if the HDPI does not cover zero and is therefore extreme relative to the genome-wide average of cline parameter estimates. Second, sites could be classified as being outliers if they are extreme relative to the genome-wide distribution of locus specific effects under the cline model. A greater number of sites qualified as outliers for estimates of α than qualified as having excess ancestry. There were 116 loci classified as outliers which make up 9.7% of the total number of sites. Of those, 16 were also classified as having excess ancestry making up 1.3% of all sites. This difference is consistent with other studies using both simulated and empirical data which typically find more outlier loci than excess ancestry loci (Gompert & Buerkle, 2012). Both of these methods can produce false positives as these extreme values can be produced solely by genetic drift rather than by selection (Gompert & Buerkle, 2012). So not all sites with extreme estimates will be associated with incompatibility loci. The false positive rate is exacer-

bated when there are many loci with small effects on compatibility. However, these sites should be enriched for loci associated with modest to strong reproductive incompatibility and thus provide an upper estimate of the number of sites that are associated with these modest to strong barriers to gene flow (Gompert & Buerkle, 2012).

None of the estimates for β were classified as either outliers or as having excess ancestry. Simulations have demonstrated that the α parameter is more impacted by selection against hybrid genotypes than the β parameter (Gompert, Lucas, et al., 2012). Other studies have also found no extreme estimates of β (Gompert, Lucas, et al., 2012; Nikolakis et al., 2022)[CITATION NEEDED!]. One possible interpretation of the absence of extreme values of β is that selection is only strong enough to have a significant impact on α but it is not strong enough to have a large impact on β . Unlike for α , there is not a strong relationship between locally positive selection favoring introgressed geneotypes and β (Gompert, Lucas, et al., 2012). Therefore, some of the extreme values for α could be due to adaptive introgression which does have much impact on estimates of β . This is plausible given the large extent of introgression which is potentially due to adaptive introgression. There is a negative relationship between β and dispersal rate (Gompert, Lucas, et al., 2012). It is also plausible that high dispersal rates, rather than selection is the cause of lower β values that do not reach the threshold to qualify as extreme.

Of the 9.7% of sites that qualified as α outliers, a substantially larger proportion had positive values which represent greater *A. americanus* ancestry than expected at those sites in admixed individuals. Negative α estimates represent a greater probability of *A. terrestris* ancestry at a site in within admixed individuals. Sites with positive outlier estimates for α made up 7.7% of all sites whereas those with negative outlier estimates made up just 2%. This asymmetry suggests that introgression flows more in the direction of *A. americanus* than it does in the direction of *A. terrestris*. This result is consistent with a pattern evident upon visual inspection of the mapped *STRUCTURE* results. Samples collected from sites adjacent to sites with admixed samples appear to have a greater proportion of *A. americanus* ancestry than *A. terrestris* ancestry (Fig. 1.4). Taken to-

gether, these observations suggest that introgression at this hybrid zone is asymmetric (Yang et al., 2020). Asymmetries in introgression can arise for multiple reasons. There could differences in mate choice which make females of one species more selective than females of the other (Baldassarre et al., 2014). There can also be species differences in dispersal tendencies [CITATION NEEDED!]. Reciprocal-cross differences in reproductive isolation, termed Darwin’s Corollary, are very common (Turelli & Moyle, 2007). If one of the sexes is more prone to dispersal, introgression will flow more freely in one direction than it would in the other. It is possible that this observation is just an artifact of sampling. Particularly if this is a highly mosaic hybrid zone. However, many more samples with primarily *A. terrestris* ancestry were collected than samples with primarily *A. americanus* ancestry.

1.4.3 Relationship between introgression and differentiation

Patterns of genetic differentiation and genomic introgression between *A. americanus* and *A. terrestris* are consistent with the hypothesis that regions of the genome experiencing divergent selection also affect hybrid fitness. As predicted, there is a positive association between locus specific estimates of F_{ST} and both the absolute value of the α and the β parameter estimates. Although this correlation supports the hypothesis that introgression outliers are linked to loci under selection, the association is only a modest one. Despite this, it is notable all of the outlier α estimates as well as the highest β estimates have non-zero F_{ST} estimates. Whereas sites with lower α and β estimates span the entire range from zero to one. This is consistent with expectations of secondary contact where not all loci that have undergone genomic divergence will necessarily result in reproductive isolation. A tighter coupling of divergence and resistance to gene flow would be expected under a scenario of divergence with gene flow.

1.4.4 Conclusion

In conclusion, the genome-wide sequence data analysis conducted in this study has provided compelling evidence of significant gene flow across the hybrid zone of two dis-

tinct species. The STRUCTURE analysis reveals that a substantial number of samples exhibit evidence of admixture, with the proportion of ancestry attributed to hybridization ranging from 0.05% to 0.5%. These findings suggest that hybrids are not only viable and fertile but also capable of backcrossing over multiple generations. Furthermore, the spatial distribution of admixture coefficients suggests the formation of a cline, with the highest levels of admixture at the hybrid zone's center gradually diminishing with distance. Notably, some samples with substantial admixture are located at the edges of the hybrid zone, underscoring the influence of dispersal on shaping hybrid patterns. While the tension zone model predicts hybrid zone centers based on factors like selection, population density, and natural dispersal barriers, the study reveals that transitions between species dominance align with the presence of streams and rivers. Although the data identifies outlier loci associated with reproductive incompatibility, it also underscores the complexities of hybridization, suggesting that not all divergent loci result in reproductive isolation. In summary, this research provides valuable insights into the dynamics of hybrid zones and the role of selection, dispersal, and gene flow in shaping the genetic landscape of these two species.

References

- Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). GENOMIC AND MORPHOLOGICAL ANALYSIS OF A SEMIPERMEABLE AVIAN HYBRID ZONE SUGGESTS ASYMMETRICAL INTROGRESSION OF A SEXUAL SIGNAL. *Evolution*, 68(9), 2644–2657. <https://doi.org/10.1111/evo.12457>
- Barton, N. H. (1979). The dynamics of hybrid zones. *Heredity*, 43(3), 341–359. <https://doi.org/10.1038/hdy.1979.87>
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review*, 16, 113–148.
- 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>
- Blackman, B. (2016). Speciation Genes. *Encyclopedia of Evolutionary Biology* (pp. 166–175). Elsevier. <https://doi.org/10.1016/B978-0-12-800049-6.00066-4>
- 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.
- Breden, F. (1987). The Effect of Post-Metamorphic Dispersal on the Population Genetic Structure of Fowler's Toad, *Bufo woodhousei fowleri*. *Copeia*, 1987(2), 386–395. <https://doi.org/10.2307/1445775>
- Butlin, R., Debelle, A., Kerth, C., Snook, R. R., Beukeboom, L. W., RF, C. C., Diao, W., Maan, M. E., Paolucci, S., Weissing, F. J., et al. (2011). What do we need to know about speciation? *Trends in ecology & evolution*, 27(1), 27–39.
- Cocroft, R. B., & Ryan, M. J. (1995). Patterns of advertisement call evolution in toads and chorus frogs. *Animal Behaviour*, 49(2), 283–303. <https://doi.org/10.1006/anbe.1995.0043>
- Colliard, C., Sicilia, A., Turrisi, G. F., Arculeo, M., Perrin, N., & Stöck, M. (2010). Strong reproductive barriers in a narrow hybrid zone of West-Mediterranean green toads (*Bufo viridissubgroup*) with Plio-Pleistocene divergence. *BMC Evolutionary Biology*, 10(1), 232. <https://doi.org/10.1186/1471-2148-10-232>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Đorđević, S., & Simović, A. (2014). STRANGE AFFECTION: MALE BUFO BUFO (ANURA: BUFONIDAE) PASSIONATELY EMBRACING A BULGE OF MUD. *Ecologica Montenegrina*, 1(1), 15–17. <https://doi.org/10.37828/em.2014.1.4>

- 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>
- Forester, D. C., Snodgrass, J. W., Marsalek, K., & Lanham, Z. (2006). Post-Breeding Dispersal and Summer Home Range of Female American Toads (*Bufo americanus*). *Northeastern Naturalist*, 13(1), 59–72. [https://doi.org/10.1656/1092-6194\(2006\)13\[59:PDASHR\]2.0.CO;2](https://doi.org/10.1656/1092-6194(2006)13[59:PDASHR]2.0.CO;2)
- 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>
- Gompert, Z., & Buerkle, C. A. (2012). Bgc: Software for Bayesian estimation of genomic clines. *Molecular Ecology Resources*, 12(6), 1168–1176. <https://doi.org/10.1111/1755-0998.12009.x>
- Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines: BAYESIAN GENOMIC CLINES. *Molecular Ecology*, 20(10), 2111–2127. <https://doi.org/10.1111/j.1365-294X.2011.05074.x>
- Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). GENOMIC REGIONS WITH A HISTORY OF DIVERGENT SELECTION AFFECT FITNESS OF HYBRIDS BETWEEN TWO BUTTERFLY SPECIES: GENOMICS OF SPECIATION. *Evolution*, 66(7), 2167–2181. <https://doi.org/10.1111/j.1558-5646.2012.01587.x>
- Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of Population Genomic Data from Hybrid Zones. *Annual Review of Ecology, Evolution, and Systematics*, 48(1), 207–229. <https://doi.org/10.1146/annurev-ecolsys-110316-022652>
- Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 439–450. <https://doi.org/10.1098/rstb.2011.0196>

- 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.
- Harrison, R. G. (1986). Pattern and process in a narrow hybrid zone. *Heredity*, 56(3), 337–349. <https://doi.org/10.1038/hdy.1986.55>
- Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18), 4606–4618. <https://doi.org/10.1111/mec.12415>
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jahner, J. P., Parchman, T. L., & Matocq, M. D. (2021). Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition. *Molecular Ecology*, 30(17), 4245–4258. <https://doi.org/10.1111/mec.16056>
- 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>
- Larson, E. L., Andrés, J. A., Bogdanowicz, S. M., & Harrison, R. G. (2013). DIFFERENTIAL INTROGRESSION IN A MOSAIC HYBRID ZONE REVEALS CANDIDATE BARRIER GENES. *Evolution*, 67(12), 3653–3661. <https://doi.org/10.1111/evo.12205>
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20(5), 229–237. <https://doi.org/10.1016/j.tree.2005.02.010>
- Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506), 2971–2986. <https://doi.org/10.1098/rstb.2008.0081>

- Malone, J. H., & Fontenot, B. E. (2008). Patterns of Reproductive Isolation in Toads (R. DeSalle, Ed.). *PLoS ONE*, 3(12), e3900. <https://doi.org/10.1371/journal.pone.0003900>
- Miller, C. J. J., & Matute, D. R. (2016). The Effect of Temperature on Drosophila Hybrid Fitness. *G3: Genes/Genomes/Genetics*, 7(2), 377–385. <https://doi.org/10.1534/g3.116.034926>
- Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization (P. J. Wittkopp, Ed.). *eLife*, 10, e69016. <https://doi.org/10.7554/eLife.69016>
- Mount, R. H. (1975). *The Reptiles and Amphibians of Alabama*. The University of Alabama Press.
- Nikolakis, Z. L., Schield, D. R., Westfall, A. K., Perry, B. W., Ivey, K. N., Orton, R. W., Hales, N. R., Adams, R. H., Meik, J. M., Parker, J. M., Smith, C. F., Gompert, Z., Mackessy, S. P., & Castoe, T. A. (2022). Evidence that genomic incompatibilities and other multilocus processes impact hybrid fitness in a rattlesnake hybrid zone. *Evolution*, 76(11), 2513–2530. <https://doi.org/10.1111/evo.14612>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Nosil, P., & Schlüter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, 26(4), 160–167. <https://doi.org/10.1016/j.tree.2011.01.001>
- Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. a. C., Zhang, G., Jarvis, E. D., Schlinger, B. A., & Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology*, 22(12), 3304–3317. <https://doi.org/10.1111/mec.12201>
- Powell, R., Conant, R., & Collins, J. T. (2016). *A Field Guide to Reptiles & Amphibians: Eastern and Central North America* (4th ed.). Houghton Mifflin Harcourt.

- 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>
- 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>
- Servedio, M. R., & Noor, M. A. (2003). The Role of Reinforcement in Speciation: Theory and Data. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), 339–364. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132412>
- Shankman, D., & Hart, J. L. (2007). The Fall Line: A Physiographic-Forest Vegetation Boundary. *Geographical Review*, 97(4), 502–519. <https://doi.org/10.1111/j.1931-0846.2007.tb00409.x>
- Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, 15(2), 343–355. <https://doi.org/10.1111/j.1365-294X.2005.02794.x>
- Turelli, M., & Moyle, L. C. (2007). Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule. *Genetics*, 176(2), 1059–1088. <https://doi.org/10.1534/genetics.106.065979>
- Urban, M. C., Phillips, B. L., Skelly, D. K., & Shine, R. (2008). A Toad More Traveled: The Heterogeneous Invasion Dynamics of Cane Toads in Australia. *The American Naturalist*, 171(3), E134–E148. <https://doi.org/10.1086/527494>
- Van Riemsdijk, I., Arntzen, J. W., Bucciarelli, G. M., McCartney-Melstad, E., Rafajlović, M., Scott, P. A., Toffelmier, E., Shaffer, H. B., & Wielstra, B. (2023). Two transects

reveal remarkable variation in gene flow on opposite ends of a European toad hybrid zone. *Heredity*. <https://doi.org/10.1038/s41437-023-00617-6>

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

Weatherby, C. A. (1982). INTROGRESSION BETWEEN THE AMERICAN TOAD, BUFO AMERICANUS, AND THE SOUTHERN TOAD, B. TERRESTRIS, IN ALABAMA.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2408641>

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100. <https://doi.org/10.1038/nrg.2016.133>

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6), 851–865. <https://doi.org/10.1046/j.1420-9101.2001.00335.x>

Yang, W., Feiner, N., Laakkonen, H., Sacchi, R., Zuffi, M. A. L., Scali, S., While, G. M., & Uller, T. (2020). Spatial variation in gene flow across a hybrid zone reveals causes of reproductive isolation and asymmetric introgression in wall lizards*. *Evolution*, 74(7), 1289–1300. <https://doi.org/10.1111/evo.14001>

1.5 Figures

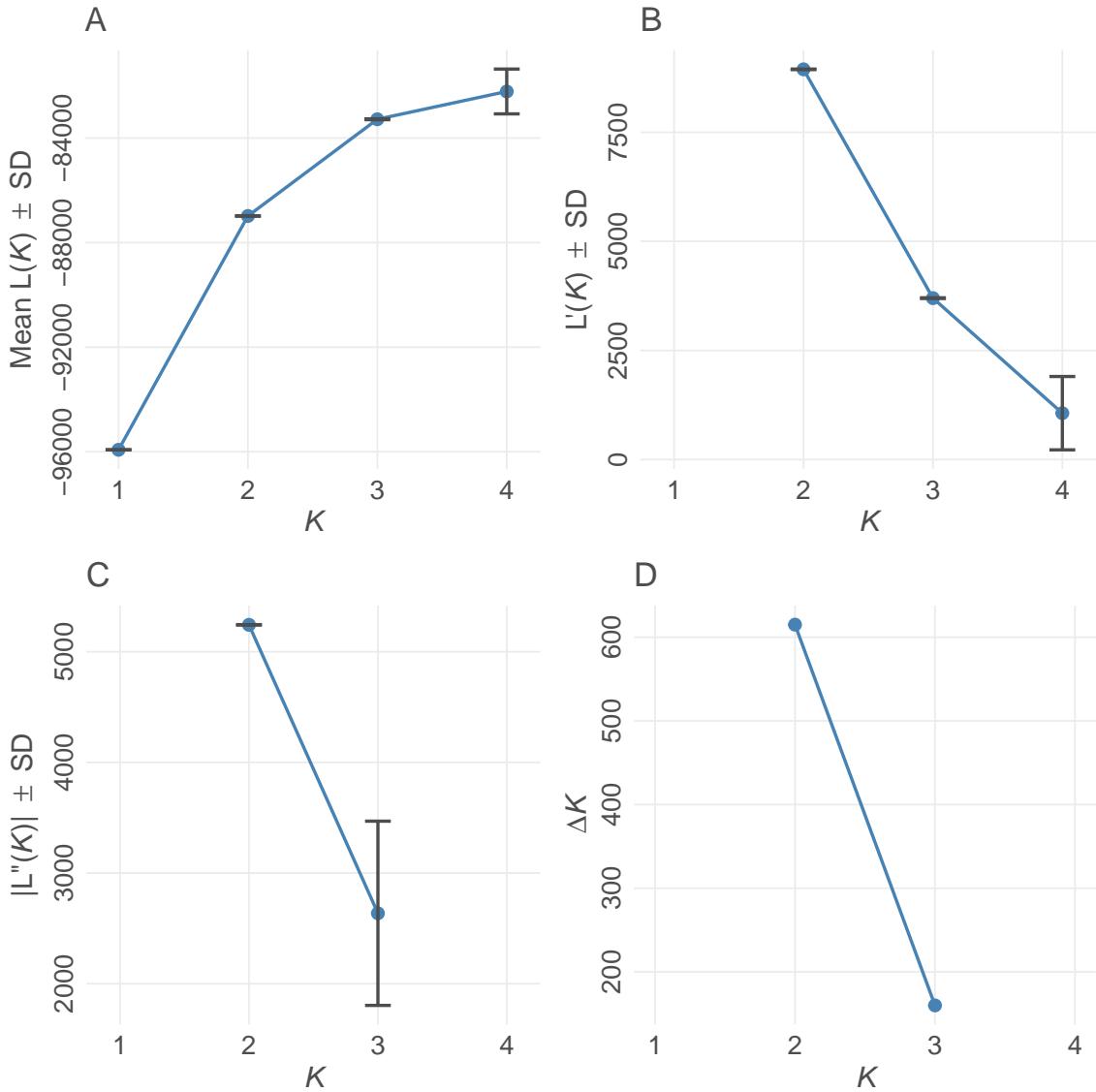


Figure 1.1. Evanno method for optimal value for K in *STRUCTURE* (Evanno et al., 2005). K refers to the number of populations for each of the different *STRUCTURE* models examined. (A) Mean estimated \ln probability of data over 10 iterations for each value of $K \pm SD$. (B) Rate of change of the likelihood distribution (mean $\pm SD$) (C) Absolute values of the second order rate of change of the likelihood distribution (mean $\pm SD$) (D) ΔK . The modal value of this distribution is considered the true value of K for the data. Plot created using *POPHELP* (Francis, 2017).

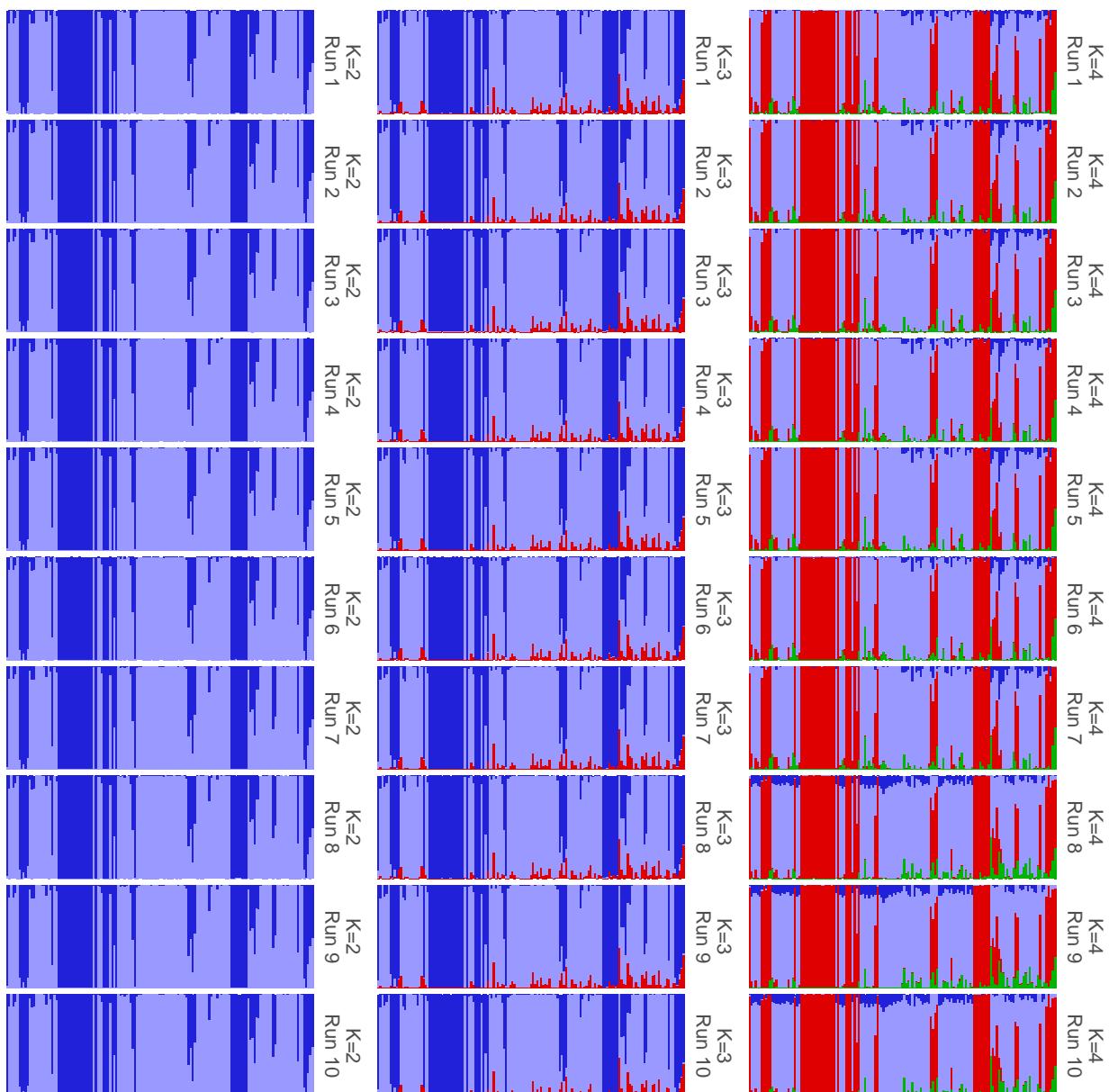


Figure 1.2. Results of each iteration of *STRUCTURE* showing convergence among iterations within runs having the same value for K. Plot was created with *POPHELPER* (Francis, 2017).

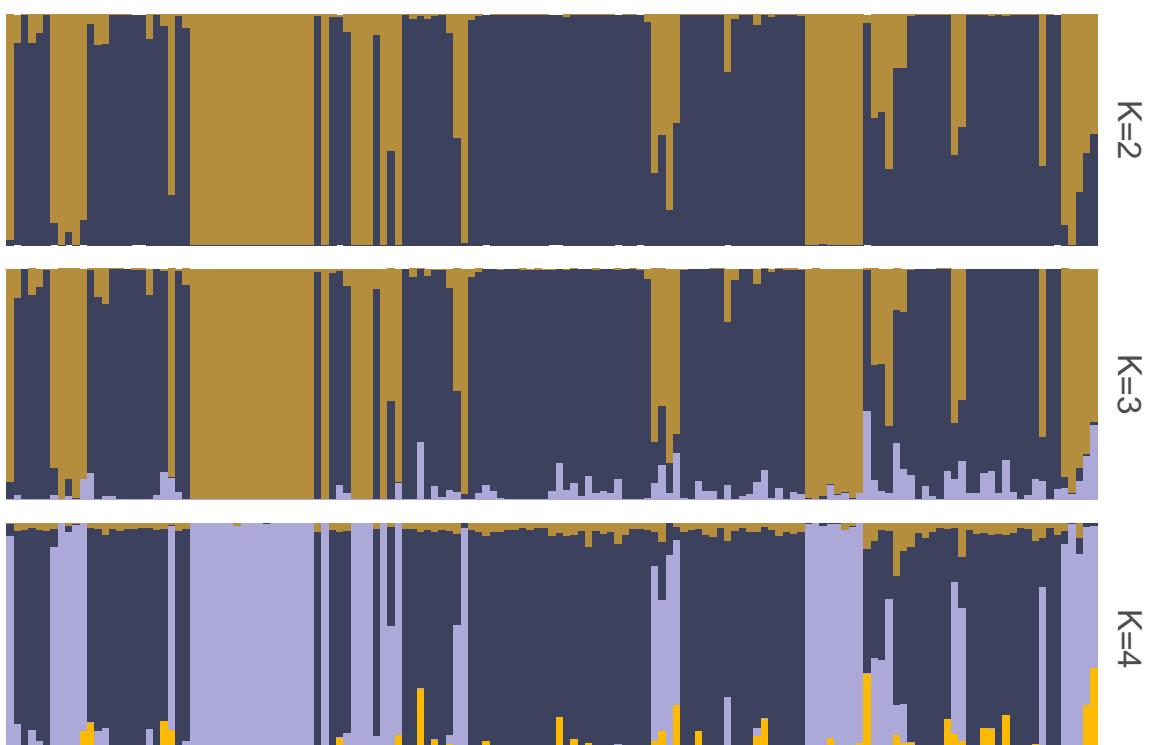


Figure 1.3. Summarized *STRUCTURE* results for each value of K. Ancestry proportions shown are the mean of ancestry proportions across all iterations. Summarization and plotting done using *POPHELP* (Francis, 2017).

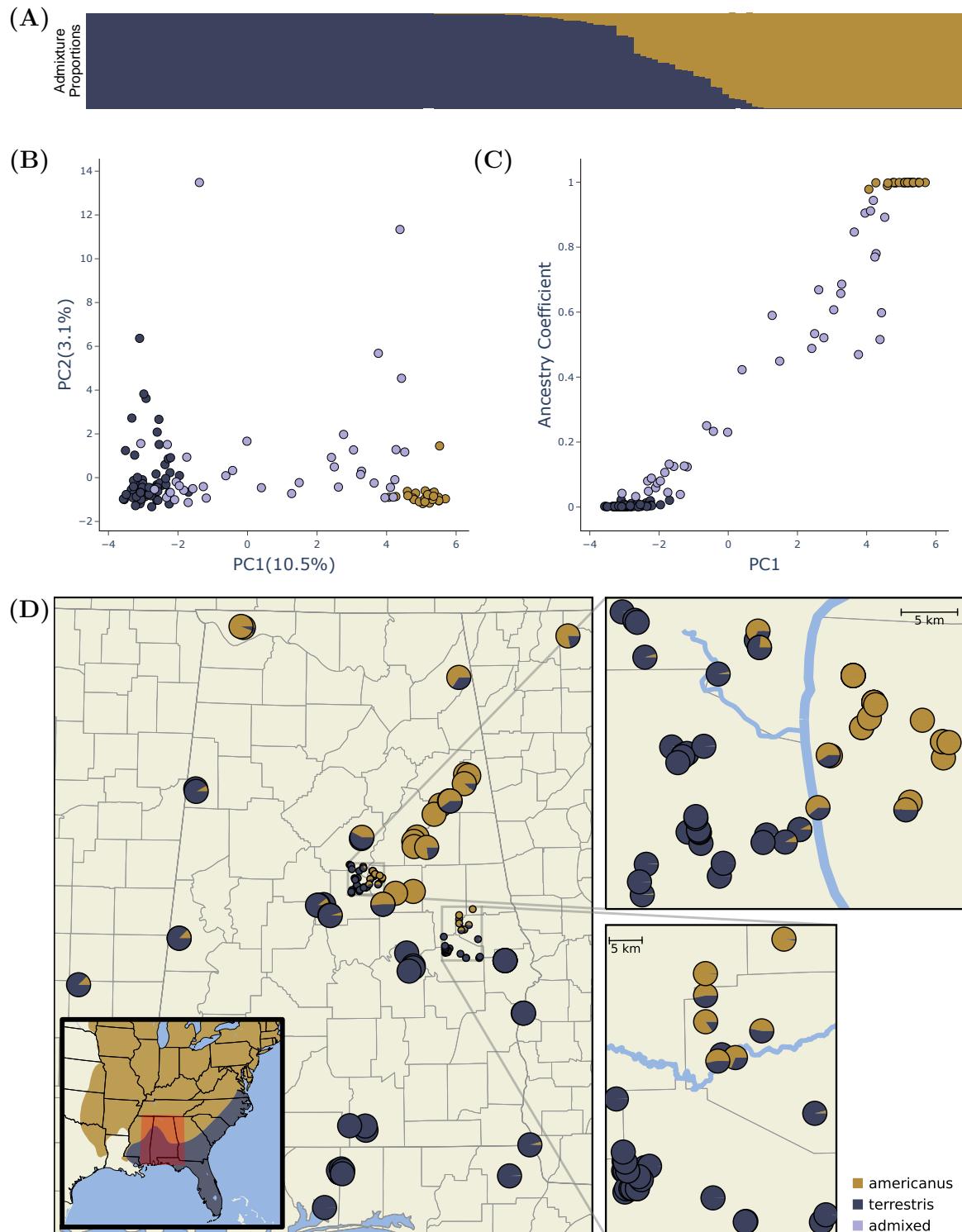


Figure 1.4. Genetic evidence of hybridization between *A. americanus* and *A. terrestris*. (A) Bar plot with the ancestry coefficients estimated with *STRUCTURE*. (B) Summary of population genetic structure based on the principal component axes one (PC1) and two (PC2). These axes explain 10.5% (PC1) and 3.1% (PC2) of the genetic variation among individuals. (C) Relationship between the first principal component axis and the admixture proportions estimated with *STRUCTURE*. (D) Sample map showing the sampling location and estimated ancestry coefficients of each sample. The inset map shows the approximate ranges of each species and the study area highlighted in red. Figure created using *POPHelper* (Francis, 2017) and *Matplotlib* (Hunter, 2007)

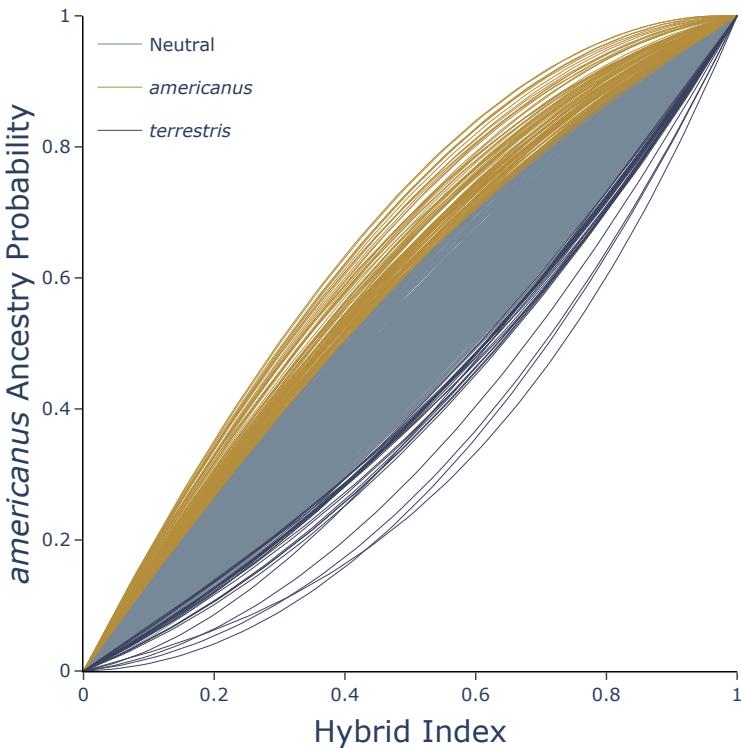


Figure 1.5. Shape of genomic clines estimated for each locus with BGC. Outliers are highlighted with XX.

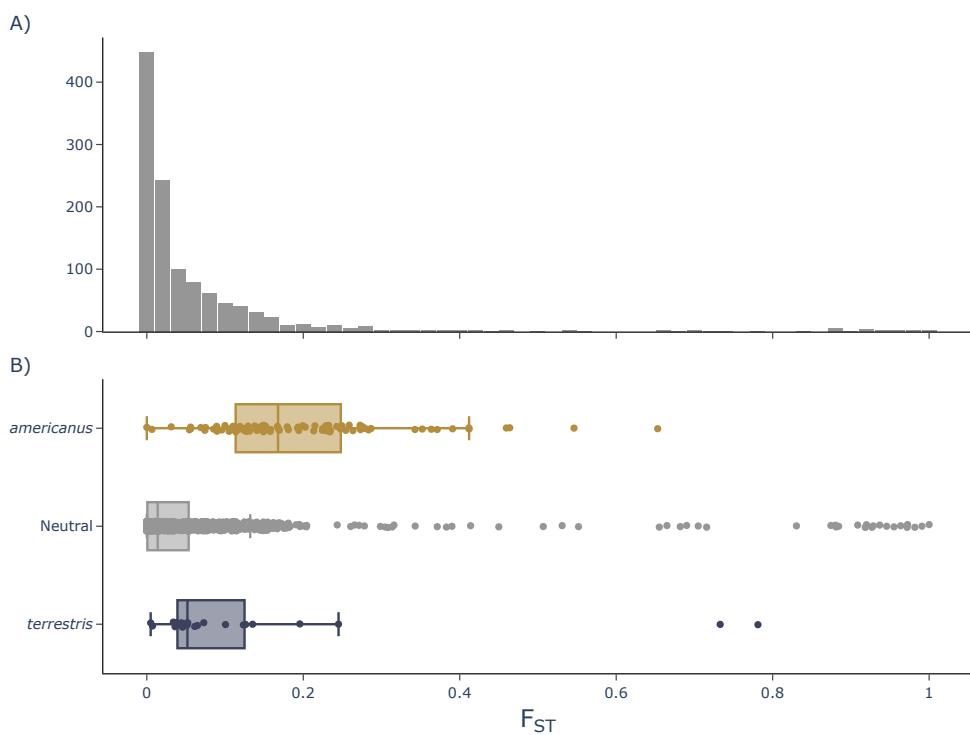


Figure 1.6. A) Distribution of per site F_{ST} estimates. B) Box plots showing the distribution and mean of F_{ST} for three categories of α estimates, outliers with greater than expected *A. americanus* ancestry (gold), outliers with greater than expected *A. terrestris* ancestry (violet), and non-outliers (gray).

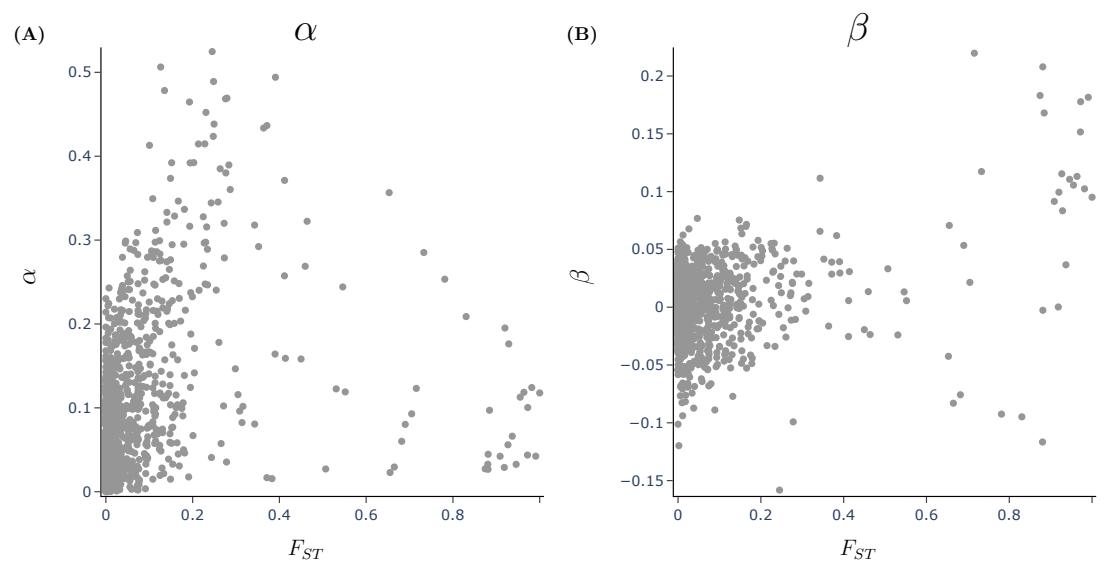


Figure 1.7. Relationship between genetic divergence measured with Weir and Cockerham, 1984 F_{ST} and BGC cline parameters A) α and B) β .

1.6 Tables

Table 1.1. Samples collected for this study

Sample ID	Species	Latitude	Longitude	Passed Filtering
KAC 016	<i>terrestris</i>	30.54819	-86.93067	X
KAC 038	<i>terrestris</i>	32.81470	-86.93968	X
KAC 039	<i>terrestris</i>	32.81094	-86.98967	X
KAC 040	<i>terrestris</i>	32.80985	-86.99795	X
KAC 042	<i>terrestris</i>	32.82406	-86.99314	
KAC 043	<i>terrestris</i>	32.82406	-86.99314	
KAC 044	<i>terrestris</i>	32.80450	-87.03078	
KAC 045	<i>terrestris</i>	32.76703	-87.07073	
KAC 046	<i>terrestris</i>	32.76592	-87.07184	
KAC 047	<i>terrestris</i>	32.78932	-86.90850	
KAC 048	<i>terrestris</i>	32.73575	-86.88149	X
KAC 049	<i>terrestris</i>	32.73291	-86.87707	X
KAC 050	<i>terrestris</i>	32.74822	-86.79806	
KAC 051	<i>terrestris</i>	32.78742	-86.75847	
KAC 052	<i>terrestris</i>	32.78044	-86.73877	
KAC 070	<i>americanus</i>	34.79963	-84.57678	X
KAC 071	<i>terrestris</i>	32.43478	-85.64630	
KAC 074	<i>terrestris</i>	30.77430	-85.22690	X
KAC 075	<i>terrestris</i>	32.94778	-86.63224	X
KAC 076	<i>terrestris</i>	32.94970	-86.52687	
KAC 077	<i>terrestris</i>	32.94970	-86.52687	
KAC 078	<i>americanus</i>	33.00267	-86.38960	X
KAC 079	<i>americanus</i>	33.01205	-86.47872	
KAC 080	<i>americanus</i>	33.04456	-86.45547	
KAC 081	<i>americanus</i>	33.04456	-86.45547	X

Continued on next page

Table 1.1 – continued from previous page

Sample ID	Species	Latitude	Longitude	Passed Filtering
KAC 082	<i>americanus</i>	33.04456	-86.45547	X
KAC 083	<i>americanus</i>	33.04456	-86.45547	X
KAC 084	<i>americanus</i>	33.04456	-86.45547	X
KAC 085	<i>americanus</i>	33.04456	-86.45547	
KAC 086	<i>americanus</i>	33.04456	-86.45547	X
KAC 087	<i>americanus</i>	33.01484	-86.39040	X
KAC 089	<i>americanus</i>	33.01484	-86.39040	X
KAC 090	<i>americanus</i>	33.06472	-86.47496	X
KAC 091	<i>americanus</i>	33.06472	-86.47496	X
KAC 092	<i>americanus</i>	33.06472	-86.47496	
KAC 093	<i>americanus</i>	33.06472	-86.47496	X
KAC 094	<i>americanus</i>	33.06472	-86.47496	X
KAC 095	<i>americanus</i>	33.06472	-86.47496	X
KAC 096	<i>americanus</i>	33.06472	-86.47496	X
KAC 097	<i>americanus</i>	33.06472	-86.47496	X
KAC 098	<i>americanus</i>	33.02572	-86.46711	X
KAC 099	<i>americanus</i>	33.02572	-86.46711	X
KAC 100	<i>terrestris</i>	32.92374	-86.67199	X
KAC 101	<i>americanus</i>	33.03283	-86.45975	X
KAC 102	<i>terrestris</i>	32.94544	-86.55777	X
KAC 103	<i>terrestris</i>	32.94947	-86.52630	X
KAC 104	<i>terrestris</i>	32.94947	-86.52630	X
KAC 105	<i>americanus</i>	33.04278	-86.45377	X
KAC 106	<i>americanus</i>	33.00464	-86.49692	X
KAC 107	<i>americanus</i>	33.01416	-86.38417	X
KAC 108	<i>terrestris</i>	32.94013	-86.54004	X

Continued on next page

Table 1.1 – continued from previous page

Sample ID	Species	Latitude	Longitude	Passed Filtering
KAC 109	<i>terrestris</i>	32.94173	-86.55787	
KAC 110	<i>americanus</i>	33.03099	-86.40941	X
KAC 111	<i>americanus</i>	33.00518	-86.49895	X
KAC 112	<i>terrestris</i>	32.95011	-86.53723	
KAC 113	<i>americanus</i>	33.00528	-86.38897	
KAC 114	<i>americanus</i>	33.01617	-86.40318	
KAC 115	<i>americanus</i>	32.98218	-86.40488	
KAC 116	<i>americanus</i>	32.96964	-86.42137	X
KAC 117	<i>terrestris</i>	32.97146	-86.52901	
KAC 121	<i>terrestris</i>	32.44120	-85.65386	X
KAC 122	<i>terrestris</i>	32.85411	-86.76619	
KAC 123	<i>terrestris</i>	32.90084	-86.67587	X
KAC 124	<i>terrestris</i>	32.91060	-86.67850	X
KAC 125	<i>terrestris</i>	32.91715	-86.68208	
KAC 126	<i>terrestris</i>	32.92717	-86.67407	
KAC 127	<i>terrestris</i>	32.97159	-86.62516	
KAC 128	<i>terrestris</i>	33.00585	-86.63703	
KAC 129	<i>terrestris</i>	33.00797	-86.64210	
KAC 130	<i>terrestris</i>	33.00818	-86.64333	
KAC 131	<i>terrestris</i>	33.01508	-86.64937	
KAC 132	<i>terrestris</i>	33.02034	-86.66651	
KAC 133	<i>terrestris</i>	33.01163	-86.64759	X
KAC 134	<i>terrestris</i>	33.00537	-86.63652	X
KAC 135	<i>terrestris</i>	33.00644	-86.63368	X
KAC 136	<i>terrestris</i>	33.00673	-86.63316	X
KAC 138	<i>americanus</i>	32.70224	-85.66196	X

Continued on next page

Table 1.1 – continued from previous page

Sample ID	Species	Latitude	Longitude	Passed Filtering
KAC 139	<i>americanus</i>	32.73042	-85.66173	X
KAC 140	<i>terrestris</i>	32.62553	-85.63684	X
KAC 141	<i>terrestris</i>	32.41032	-85.60107	X
KAC 142	<i>terrestris</i>	32.57011	-85.80888	X
KAC 143	<i>terrestris</i>	32.47773	-85.79824	X
KAC 144	<i>terrestris</i>	32.47707	-85.79577	X
KAC 145	<i>terrestris</i>	32.48128	-85.76354	X
KAC 146	<i>terrestris</i>	32.48291	-85.75622	X
KAC 147	<i>terrestris</i>	32.45001	-85.79652	X
KAC 148	<i>terrestris</i>	32.45420	-85.79408	X
KAC 149	<i>terrestris</i>	32.45449	-85.78664	X
KAC 150	<i>terrestris</i>	32.45449	-85.78664	X
KAC 151	<i>terrestris</i>	32.45451	-85.78416	X
KAC 152	<i>terrestris</i>	32.45423	-85.77634	X
KAC 153	<i>terrestris</i>	32.45423	-85.77634	X
KAC 154	<i>terrestris</i>	32.46574	-85.76977	X
KAC 155	<i>terrestris</i>	32.46961	-85.77369	X
KAC 156	<i>terrestris</i>	32.47709	-85.79175	X
KAC 158	<i>terrestris</i>	32.47709	-85.79175	X
KAC 159	<i>terrestris</i>	32.49000	-85.79741	X
KAC 160	<i>terrestris</i>	32.40809	-85.47857	X
KAC 161	<i>terrestris</i>	32.41744	-85.47117	X
KAC 162	<i>terrestris</i>	32.35417	-86.09838	X
KAC 163	<i>terrestris</i>	32.33994	-86.09946	X
KAC 164	<i>terrestris</i>	32.31562	-86.13789	X
KAC 167	<i>terrestris</i>	33.06620	-86.60328	X

Continued on next page

Table 1.1 – continued from previous page

Sample ID	Species	Latitude	Longitude	Passed Filtering
KAC 172	<i>americanus</i>	32.62171	-85.61467	X
KAC 173	<i>americanus</i>	32.61751	-85.64335	X
KAC 176	<i>americanus</i>	32.66836	-85.66233	X
KAC 177	<i>americanus</i>	32.65571	-85.57134	X
KAC 181	<i>terrestris</i>	32.38644	-85.23561	X
KAC 182	<i>terrestris</i>	32.38579	-85.23565	X
KAC 183	<i>terrestris</i>	32.38579	-85.23565	X
KAC 184	<i>terrestris</i>	32.38579	-85.23565	X
KAC 185	<i>terrestris</i>	32.38579	-85.23565	X
KAC 187	<i>americanus</i>	32.64548	-85.55135	
KAC 188	<i>terrestris</i>	32.40976	-85.60208	X
KAC 189	<i>terrestris</i>	33.09152	-86.56686	X
KAC 190	<i>terrestris</i>	33.11298	-86.69434	X
KAC 191	<i>terrestris</i>	33.10659	-86.68228	X
KAC 192	<i>terrestris</i>	33.10509	-86.68014	X
KAC 193	<i>terrestris</i>	33.07896	-86.67286	X
KAC 194	<i>terrestris</i>	32.93933	-86.62008	X
KAC 195	<i>terrestris</i>	32.94745	-86.62146	X
KAC 196	<i>terrestris</i>	32.94829	-86.62190	X
KAC 197	<i>terrestris</i>	32.94929	-86.62241	X
KAC 198	<i>terrestris</i>	32.95077	-86.62306	
KAC 199	<i>terrestris</i>	32.95794	-86.62477	X
KAC 200	<i>terrestris</i>	32.95940	-86.62489	X
KAC 205	<i>terrestris</i>	32.54852	-85.48692	X
KAC 206	<i>americanus</i>	33.30759	-86.58201	X
KAC 207	<i>americanus</i>	33.31685	-86.57596	X

Continued on next page

Table 1.1 – continued from previous page

Sample ID	Species	Latitude	Longitude	Passed Filtering
KAC 208	<i>americanus</i>	33.09829	-86.56529	X
KAC 209	<i>terrestris</i>	33.08600	-86.56394	X
KAC 210	<i>terrestris</i>	33.08600	-86.56394	X
KAC 211	<i>terrestris</i>	33.01464	-86.60995	
KAC 212	<i>terrestris</i>	33.01208	-86.61707	X
KAC 213	<i>terrestris</i>	33.00435	-86.63710	X
KAC 214	<i>terrestris</i>	32.99991	-86.64181	X
KAC 215	<i>terrestris</i>	32.99605	-86.64526	
KAC 216	<i>terrestris</i>	33.01346	-86.60960	
KAC 217	<i>terrestris</i>	32.91470	-86.60270	X
KAC 218	<i>terrestris</i>	32.92432	-86.59895	X
KAC 219	<i>terrestris</i>	32.93987	-86.56113	X
KAC 220	<i>americanus</i>	32.96579	-86.50892	X
KAC 221	<i>americanus</i>	32.96389	-86.42549	X
KAC 223	<i>terrestris</i>	32.53362	-85.79839	
KAC 224	<i>terrestris</i>	32.48869	-85.79555	X
KAC 225	<i>terrestris</i>	32.50159	-85.79860	X
KAC 230	<i>terrestris</i>	30.80933	-86.77686	X
KAC 232	<i>terrestris</i>	30.80922	-86.78994	X
KAC 233	<i>terrestris</i>	30.80922	-86.78994	X
KAC 234	<i>terrestris</i>	30.80922	-86.78994	X
KAC 236	<i>terrestris</i>	30.82632	-86.80258	X
KAC 237	<i>terrestris</i>	30.83733	-86.77630	X
KAC 238	<i>terrestris</i>	30.82433	-86.76284	X
KAC 239	<i>terrestris</i>	30.80162	-86.76659	X
KAC 242	<i>americanus</i>	34.50446	-85.63768	X

Continued on next page

Table 1.1 – continued from previous page

Sample ID	Species	Latitude	Longitude	Passed Filtering
KAC t1020	<i>terrestris</i>	31.10783	-86.62247	
KAC t1030	<i>terrestris</i>	31.99042	-85.07423	X
KAC t1040	<i>terrestris</i>	31.99016	-85.07046	X
KAC t2004	<i>americanus</i>	33.58295	-85.73524	X
KAC t2015	<i>americanus</i>	33.58435	-85.74064	X
KAC t2018-02-17-01	<i>americanus</i>	33.55274	-85.82913	X
KAC t2018-02-17-04	<i>americanus</i>	33.48548	-85.88857	X
KAC t2018-02-17-05	<i>americanus</i>	33.31649	-86.05293	X
KAC t2018-02-17-06	<i>americanus</i>	33.28443	-86.08443	X
KAC t2018-02-17-07	<i>americanus</i>	33.24576	-86.08168	X
KAC t2018-03-10-1	<i>americanus</i>	32.91057	-86.09272	X
KAC t2018-03-10-3	<i>americanus</i>	32.95104	-86.14539	
KAC t2018-03-10-4	<i>americanus</i>	32.89787	-86.26061	X
KAC t2018-03-10-5	<i>americanus</i>	32.81642	-86.38018	X
KAC t2019-08-25-1	<i>americanus</i>	34.21852	-87.36662	
KAC t2020	<i>americanus</i>	33.23853	-85.96270	X
KAC t2040	<i>americanus</i>	33.58295	-85.73539	X
KAC t2043	<i>americanus</i>	32.81642	-86.38018	X

Table 1.2. Samples loaned from museums

Sample ID	Species	Latitude	Longitude	Passed Filtering
AHT 1975	<i>americanus</i>	32.77356	-85.53325	X
AHT 2456	<i>terrestris</i>	32.19494	-89.23629	X
AHT 2885	<i>terrestris</i>	32.45090	-86.15934	X
AHT 3419	<i>terrestris</i>	33.67290	-88.16068	X
AHT 3421	<i>terrestris</i>	33.65420	-88.15580	X
AHT 3428	<i>terrestris</i>	31.12679	-86.54755	X
AHT 3459	<i>americanus</i>	34.88028	-87.71849	X
AHT 3460	<i>americanus</i>	33.78013	-85.58421	X
AHT 3461	<i>americanus</i>	34.88779	-87.74103	X
AHT 3462	<i>americanus</i>	33.77001	-85.55434	X
AHT 3463	<i>americanus</i>	33.71125	-85.59762	X
AHT 3813	<i>terrestris</i>	31.13854	-86.53906	
AHT 3833	<i>terrestris</i>	31.00422	-85.03427	X
AHT 3997	<i>terrestris</i>	32.55607	-88.29975	X
AHT 3998	<i>terrestris</i>	32.55607	-88.29975	X
AHT 5276	<i>terrestris</i>	31.55613	-86.82514	
AHT 5277	<i>terrestris</i>	31.15830	-86.55430	X
AHT 5278	<i>terrestris</i>	31.16105	-86.69868	X
UTEP 19947	<i>terrestris</i>	31.22432	-88.77548	

Chapter 2

Evolutionary History

2.1 Introduction

Many factors are hypothesized 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 (Hua & Wiens, 2013). Changes in these environmental variables can alter the distributions of organisms or alter the distribution of variation within species and as a result, change patterns of gene flow within a species (Coyne & Orr, 2004). Spatially separated 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 (Coyne & Orr, 2004). Local adaptation in response to the environment may lead assortative mating among populations of a species which in time could result in complete reproductive isolation (Mallet, 2008). Another important process, which itself will often be tied to environmental changes is hybridization. Environmental changes can reestablish migration between previously isolated populations resulting in hybridization and potentially introgression between species (**abbot2013**). 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 obtaining an accurate reconstruction of the evolutionary history of organisms. Knowing the order of divergences along with their

timing, it may be possible to identify the underlying environmental factors driving them. Revealing signatures of introgression in the past paves the way for understanding its consequences and role in the diversification process.

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. terrestris*, and *A. woodhousii*. Two phylogenetic studies have 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). The conflicting results produced by these studies 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* (Green, 1983). Meacham, 1962 presented compelling evi-

dence on the basis of morphological variation for the existence of a hybrid zone between *A. fowleri* and *A. woodhousii* in East Texas although this has never been investigated with genetic data. In the previous chapter, I presented evidence for extensive hybridization between *A. americanus* and *A. terrestris*. Furthermore, numerous laboratory crosses have been performed between pairs of *Anaxyrus* species that occur in sympatry (Blair, 1963, 1972). Some of which produce viable and fertile backcross progeny (Blair, 1963, 1972). These studies suggest that gene flow could very well have played a role in shaping patterns of diversity in *Anaxyrus*. However, these studies provide only a snapshot in time with no indication of the long term evolutionary consequences if any. There are many potential lasting 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. Inferring the structure of phylogenetic networks is much more computationally demanding than inferring more simple bifurcating phylogenetic trees (Wen et al., 2018). 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 making.

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 (Rieseberg et al., 1999). 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 hybrid zones. Context such as the phylogenetic relationships of hybridizing species, the amount of genetic divergence between them, the time since divergence, and the bio-

geographic processes driving initial divergence and subsequent secondary contact in cases of allopatric divergence. This important context is currently missing for *Anaxyrus* which limits our understanding of hybridization within the genus.

Ultimately, changes in the environment are what drive speciation and hybridization and so it is important to identify these. To date, there have not been any studies conducted to understand how the environment has 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; Xue & Hickerson, 2015). 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 simultaneously can provide clues as to the cause. 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 their ancestral state.

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 sampling that encompasses a large portion of the ranges of *A. americanus*, *A. fowleri*, *A. terrestris*, and *A. woodhousii*. With these data I conduct the first inference the evolutionary relationships among these species using genome wide sequence 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.

2.2 Methods

2.2.1 Sampling and DNA Isolation

I obtained tissue samples from museum tissue collections as well as individuals that 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 *incilius nebulifer* as an outgroup for phylogenetic analyses.

I isolated sample DNA from liver or muscle tissue by lysing a piece approximately the size of a grain of rice in a 300 μL solution of 10mM Tris-HCL, 10mM EDTA, 1% SDS (w/v), 6 mg Proteinase K, and nuclease free water incubated for 4-12 hours at 55°C. To purify the DNA, I mixed the lysis solution 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 after beads had collected on the side of the tube. I then performed two ethanol washes with 1 mL of 70% ETOH added to the beads while still placed in the magnet stand and allowed the sample to stand for 5 minutes before discarding the ethanol. After discarding all ethanol from the second wash, I removed the tube from the magnet stand and allowed the sample to dry for 1 minute. I then mixed the beads with 100 μL of TLE solution containing 10 mM Tris-HCL, 0.1 mm EDTA, and nuclease free water. After allowing this mixture to stand at room temperature for 5 minutes I returned the beads to the magnet stand and separated the DNA solution from the beads. I quantified DNA with a Qubit fluorometer (Life Technologies, USA) and diluted samples with TLE solution to bring all sample concentrations to 20 ng/ μL .

2.2.2 RADseq Library Preparation

I prepared RADseq libraries using the 2RAD approach outlined by Bayona-Vásquez et al., 2019. On 96 well plates, I digested 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 ligated the adapter by adding 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 into equal volumes. I purified each pool of libraries with a 1X volume of SPRI bead 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 volume and purified the product with a 2X SPRI bead 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. These pooled libraries were pooled with other projects and sequenced on an Illumina HiSeqX by Novogene (China) to obtain paired end, 150 base pair sequences.

2.2.3 Phylogenetic Data Processing

To produce alignments for phylogenetic analysis, I first demultiplexed the iTru7 indexes (identifying the 96 well plates) 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.

2.2.4 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 conducted maximum likelihood phylogenetic inference in order to infer a phylogeny with all of the sequenced samples in order to identify samples that may be problematic for other methods due to recent admixture, data quality, or misidentification. I conducted the maximum likelihood phylogenetic inference with *IQ-TREE* v1.6.12 (Nguyen et al., 2015) with the *ipyrad* alignment as input. I ran *IQ-TREE* with 1000 ultrafast bootstrap replicates (Hoang et al., 2018) under the GTR substitution model.

2.2.5 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* (Oaks et al., 2022). I selected a subset of up to four samples from each species due to infeasible run times for *phycoeval* with greater numbers of samples (see ??). 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 in the 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* which is packaged with *phycoeval* (Oaks et al., 2022). 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.

2.2.6 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 sample 006 due to it's anomalous position in the maximum likelihood phylogeny. 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).

2.2.7 Population Structure

I processed reads differently for the investigation of population structure within *A. americanus*, *A. fowleri*, *A. terrestris*, and *A. woodhousii* as well as for the investigation of hybridization between *A. fowleri* and *A. woodhousii*. Starting from the decloning step of the data processing for the phylogenetic analyses, I demultiplexed individual samples using the *process_radtags* program in *STRUCTURE*. I also trimmed adapter sequence and filtered reads with low quality scores as well as reads with any uncalled bases with *process_radtags* 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.

I ran the program *STRUCTURE* v2.3.4 (Pritchard et al., 2000) with its admixture model for each species separately and with *A. fowleri* and *A. woodhousii* samples combined in order to cluster individuals and estimate their ancestry proportions. For the *STRUCTURE* analyses of individuals from a single species, I ran *STRUCTURE* under five different models with each assuming a different number of populations (K parameter) ranging from one to five. For the *STRUCTURE* analysis of the combined *A. fowleri* and *A. woodhousii* samples, I ran *STRUCTURE* under four different models with K ranging from one to four. I ran 10 independent runs of *STRUCTURE* for each value of K for a total of 100,000 steps and burnin of 50,000 for each run. I used the R package *POPHelper* v2.3.1 (Francis, 2017) to combine runs for each value of K and to select the model resulting in the largest ΔK which is the the model that has the greatest increase in likelihood score from the previous model which assumed one less population 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).

2.3 Results

2.3.1 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 (see ??). Mean sequence read coverage of the loci

passing filter was 54x. The final alignment contained a total of 184,453 sites and 20,361 SNPs with 14.96% of sites and 14.71% of SNPs missing.

2.3.2 Maximum Likelihood Phylogeny

The full majority rule consensus tree inferred by *IQ-TREE* is presented in Figs. 2.2 and 2.3. 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 (Figs. 2.2 and 2.3). A representation of the tree inferred by *IQ-TREE* with the tips within species specific clades collapsed is presented in (Fig. 2.4). Each species specific clade for which there are at least two representatives samples all have ultrafast bootstrap support values of 100% (Fig. 2.4). All branches below the level of the species specific clades have ultrafast bootstrap support values ranging from 70-100% with the majority being 100% (Fig. 2.4). 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% (Fig. 2.4). 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% (Fig. 2.4). 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% (Fig. 2.4).

2.3.3 Coalescent Phylogeny

The maximum a posteriori (MAP) tree inferred under the multispecies coalescent model using *phycoeval* has a topology that differs from the maximum likelihood topology inferred by *IQ-TREE* (Fig. 2.5). The MAP tree produced by *phycoeval* does not have any shared divergence times among any of the 10 internal nodes of the tree (Fig. 2.5). 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 occurring 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 (Fig. 2.5). However, this node has a low posterior probability of only 0.51 (Fig. 2.5). All other branches in the MAP tree have high posterior probabilities of 0.98 or more (Fig. 2.5). Most divergence events within *Anaxyrus* have occurred in the past 3.5 million years and all diversification within the *A. americanus* group is less than 2.5 million years old (Fig. 2.5).

2.3.4 Introgession

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. 2.6. Most *f*-branchestimates produced by *dsuite* were zero or very near zero. Only 24 out of 112 *f*-branchestimates were greater than 0 and just 11 of those were greater than 0.05 Fig. 2.6. *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. 2.6. The highest *f*-branchstatistic values are between *A. americanus* and two other species: *A. hemiophrys* (0.24) and *A. baxteri* (0.22) Fig. 2.6. The values associated with *A. woodhousii* are appreciably lower with none exceeding 0.1 Fig. 2.6. The highest being between *A. americanus* and *A. woodhousii* with a value of 0.098 Fig. 2.6. The *A. woodhousii* *f*-branchvalues for *A. baxteri* and *A. hemiophrys* are 0.082 and 0.086 respectively Fig. 2.6. 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.

2.3.5 Population Structure

For the *STRUCTURE* analysis of each species and the analysis of the *A. fowleri* and *A. woodhousii* samples combined, a visual inspection of the 10 independent *STRUCTURE* runs performed for each value of K, shows that each independent run converged on a nearly identical result for all runs for a given K value (Figs. 2.18–2.21). For the *A. americanus*, *A. fowleri*, *A. woodhousii*, and *A. fowleri* + *A. woodhousii* analyses, the *STRUCTURE* model with the highest ΔK was the model with a K of two. (Figs. 2.13, 2.14 and 2.16 and ??). For the *A. terrestris* analysis, the *STRUCTURE* model with the highest ΔK was the model with a K of three (Fig. 2.15).

The *STRUCTURE* analysis with a K of two for *A. americanus* produced a western and eastern cluster of individuals with four admixed samples in the center of the species range (Fig. 2.8). There was a large increase in likelihood between the model with a K of two and the model with a K of three although it was not large enough to be identified as the best model using the method described by Evanno et al., 2005. Therefore, the *STRUCTURE* results for the model with a K of three are also presented (Fig. 2.7). The analysis performed with a K of three shows the same East/West division but also shows a gradient from North to South in the eastern half of the *A. americanus* range (Fig. 2.7). The PCA for *A. americanus* also shows three non-discrete groupings of individuals *A. americanus* samples which more closely matches the *STRUCTURE* analysis with a K of three.

The ancestry coefficients inferred in the *STRUCTURE* analysis for *A. terrestris* fall into three categories. Individuals in the first category, which includes all but four individuals, have admixture proportions attributed to two different source populations (Population 1 and Population 2) with the majority of ancestry attributed to Population 1 (Fig. 2.10). The second category of individuals, which includes the two easternmost samples, have ancestry proportions attributed to Population 1 and a third population (Population 3) (Fig. 2.10). The third category, which includes the next two easternmost samples, has ancestry proportions attributed to all three. These samples resemble the first category except that they have a small amount of ancestry attributable to population 3.

The PCA result for *A. terrestris* is fairly consistent with the *STRUCTURE* results with most individuals clustering tightly together and three samples forming another Fig. 2.10.

The results of the *A. fowleri* or *A. woodhousii* *STRUCTURE* analysis and PCA do not show any obvious population structure or pattern in the distribution of genetic diversity Figs. 2.9 and 2.11. Two *A. woodhousii* samples have ancestry coefficients of 1.0 for a separate population than the remaining samples Fig. 2.9. However, when analyzing the *A. fowleri* and *A. woodhousii* samples together, these same two samples have a high proportion of *A. fowleri* ancestry and are also located in the center of the two ranges Fig. 2.12. Several other samples in the combined *A. fowleri* and *A. woodhousii* analysis have mixed ancestry with a small proportion of *A. woodhousii* ancestry and these too are located in the center of the two ranges Fig. 2.12. Again, the PCA results are consistent with the *STRUCTURE* results. The PCA plot shows *A. woodhousii* samples clustered tightly, two samples right in the middle of principal component one which captures 42% of variation in the data. The *A. fowleri* samples are also tightly clustered except for four samples which gravitate towards the center of principal component one Fig. 2.12.

2.4 Discussion

2.4.1 Phylogenetic relationships

The maximum likelihood tree inferred by *IQ-TREE* Figs. 2.2 and 2.4 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 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 and sister to *A. microscaphus* 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). In this study, 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. 2.4.

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 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 in my analysis. In my experience, *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*. *phycoeval* was run 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. 2.5 (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, *A. americanus* has had closer affinity to the *A. hemiophrys/A. baxteri* clade Fig. 2.4 (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. The placement of *A. fowleri* and *A. woodhousii* sister to one another by *phycoeval* is also

unlike any previous study or the maximum likelihood tree.

An unusual feature of *phycoeval* is that it can allow for multifurcations in inferred topologies (Oaks et al., 2022). This feature proved relevant 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. Most phylogenetic methods can only produce bifurcations and thus would force any true multifurcation into bifurcations and then have to estimate some branch length between nodes. In the *phycoeval* tree, the posterior probability of this split is low (0.51) so may not be a perfect representation of the history of these lineages Fig. 2.5. 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.

2.4.2 Divergence Time

Only three previous studies have produced estimates for the age of *Anaxyrus* or any of its members 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 their 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 with the Feng et al., 2017 estimate puts the time since the most recent common ancestor (MRCA) of extant *Anaxyrus* or the split between the *boreas* group and rest of *Anaxyrus*, some time between 11.9 mya when *A. punctatus*

diverged from other *Anaxyrus* and 16.5 mya when *Anaxyrus* split from *Incilius* Fig. 2.5. This range is consistent with the 12.3 mya estimate (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 the MRCA of all *Anaxyrus* to be approximately halfway between their 14.7 mya estimate for the *A. punctatus* split and the 20.3 mya estimate for the split with *Incilius* at 16.7 mya meaning. So not immediately preceding the split between *A. punctatus* and all other *Anaxyrus*. One thing this study and others have in common is a much higher degree of uncertainty around the ages of these basal splits in the *Anaxyrus* tree. But they are all in support of the split between *Incilius* and *Anaxyrus* happening 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 which coincided with an 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 2.5 million years. This accounts for a large portion of the diversity of *Anaxyrus* and including *A. houstonensis* which was not included in this study and others have found to be nested within this clade Portik et al., 2023; Pyron and Wiens, 2011. This means that a large amount of 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 that each event occurred independently during this

period of *Anaxyrus* evolution.

2.4.3 Hybridization

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 species pairs 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. 2.6. A hybrid zone is known to exist 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. The steep cline observed by Green, 1983 over a relatively short distance suggests that reproductive isolation between these species is quite high. It is possible that introgression is occurring beyond this narrow hybrid zone but the sample that I included here was sampled from a location in close proximity to the range of *A. americanus* Fig. 2.1 (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* and this is why both have high *f*-branchscores. *f*-branch values can actually be indicative of introgression occurring between more basal branches of a tree in some circumstances (Malinsky et al., 2021). This scenario is plausible as *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 and without wider sampling from the range of *A. hemiophrys* it is not possible to rule out recent introgression (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. 2.6. 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 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 *Anaxyrus* that could not be made using the *f*-branchtest due to the input tree topology. Particularly among ancestral species, between which past introgression could be driving the pattern seen across the extant diversity as this 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 *Anaxyrus* but gaps remain.

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* (Meacham, 1962). In order to test for admixture between *A. fowleri* and *A. woodhousii* I used the program *STRUCTURE* along with PCA. The results of both *STRUCTURE* and the PCA are consistent with the existence of a hybrid zone between these two species Fig. 2.12. Two *A. woodhousii* samples, one from Arkansas and the other from Texas, have large proportions of inferred ancestry from *A. fowleri* Fig. 2.12. 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. 2.12. 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 Fig. 2.12. These results suggest the hybrid zone between *A. fowleri* and *A. woodhousii* is quite wide, possibly on the order of hundreds of kilometers Fig. 2.12.

This brings the number of confirmed *Anaxyrus* 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 *Anaxyrus*. The sister species pairs *A. fowleri/A. woodhousii* and *A. americanus/A. terrestris* diverged only 0.7 and 1.0 mya respectively Fig. 2.5. Within this timeframe 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. Introgression across these hybrid zones extends a surprisingly long distance. Introgression across the hybrid zone of *A. americanus* and *A. hemiophrys*, two species with much older divergence times, appears to be much more limited on the other hand (Green, 1983). Despite having more recent divergence times, *A. fowleri* occurs in sympatry across a large area 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 (Blair, 1974). Why more recently diverged species exist in sympatry and have evolved pre-zygotic isolating mechanisms whereas a *A. americanus* and *A. hemiophrys* with much greater time since divergence do not is interesting. Perhaps they have recently come into secondary contact and have not had sufficient time to evolve pre-zygotic barriers to reproduction. This would lend support to reinforcement being the driving force behind the evolution of pre-zygotic isolation in

these taxa.

2.4.4 Population Structure

An examination of population structure can potentially provide clues about the environmental factors that have shaped the evolutionary history of a species as population divergence is an early stage along the speciation continuum (Mallet, 2008). The geographic barriers that result in the reduction of gene flow within species could be the same types of barriers that have resulted in past speciation events involving a species or its close relatives. In my analysis of population structure in *A. americanus*, *A. fowleri*, *A. terrestris*, and *A. woodhousii*, none of the *STRUCTURE* analyses reveal any discrete populations within these species. The most abrupt transitions in admixture coefficients are seen within *A. americanus*. The *STRUCTURE* model with the highest *DeltaK* in the *A. americanus* analysis was the model with a K of two which separates *A. americanus* into western and eastern populations Fig. 2.13 (Evanno et al., 2005). However, I will focus the discussion of the *STRUCTURE* results for *A. americanus* on the analysis run with a K of three despite it producing a likelihood that was only marginally better Fig. 2.13. The results from the model with a K of three correspond well with the results of the PCA, make sense in a geographic context, and still show a stark transition of admixture coefficients from East to West as seen in the analysis with a K of 2 Figs. 2.7 and 2.8. The *STRUCTURE* analysis for *A. americanus* reveals a fairly abrupt transition from East to West beginning at the Mississippi River Fig. 2.7. All samples West of the Mississippi River have an admixture coefficient of one for the Western cluster of samples. This cluster of individuals has a range that loosely corresponds with a proposed subspecies *A. anaxyrus charlesmithi* which was said to exist in parts of Oklahoma, Arkansas, Missouri, and along the margins of some bordering states (Bragg, 1954). East of the Mississippi River, admixture coefficients associated with this cluster shrink with increasing distance from the river. In the Southeastern direction, the admixture coefficient associated with samples at the most Southeastern extent increase to one along this axis. In the Northeastern direction is a more gradual transition from samples with a

fairly balanced proportion of admixture from all three clusters to samples with a mixture of Northeastern and Southeastern ancestry, to finally a single sample with an admixture coefficient of one for the cluster of samples associated with this direction. Samples in Eastern part of the *A. americanus* range appear to only vary with distance from one another and do not have any patterns of variation that are associated with any geographic feature as they are in the West.

The *STRUCTURE* results for *A. fowleri*, *A. terrestris*, and *A. woodhousii* show very little if any differentiation within species. *A. terrestris* shows the greatest level of differentiation among these three with eastern samples having ancestry attributed to a population not represented in any of the western samples (Fig. 2.10). It is difficult to interpret this result with the extent and the size of the current dataset but it suggests there may be a gradient of genetic variation across the range of *A. terrestris* much like was found in *A. americanus*. Broader sampling which includes more samples from a greater extent of the *A. terrestris* range may shed light on this.

The PCA for *A. fowleri* shows one tightly clustered group with 4 samples that stand out from the rest (Fig. 2.9). This corresponds to the number of samples which were inferred as having a proportion of *A. woodhousii* ancestry in the combined *A. fowleri* and *A. woodhousii* *STRUCTURE* analysis. The same four samples inferred as having *A. woodhousii* ancestry have the highest proportion of ancestry from the secondary population in the *A. fowleri* *STRUCTURE* analysis (Fig. 2.9). Apart from these individuals, the admixture proportions inferred for *A. fowleri* samples are highly uniform across the range of *A. fowleri* (Fig. 2.9).

Previous studies of *A. woodhousii* have identified two distinct groups of *A. woodhousii*. Shannon and Lowe, 1955 described the subspecies *A. woodhousii australis* on the basis of morphological differentiation. This subspecies was said to be distributed across the southern parts of Arizona and New Mexico (Shannon & Lowe, 1955). Masta et al., 2003 found two divergent clades of *A. woodhousii* in a phylogeny inferred from a single mitochondrial locus. Samples from one of these clades matched the distribution of the subspecies described by although the distribution of samples from the two inferred clades

overlapped to some extent Masta et al., 2003; Shannon and Lowe, 1955. Sampling of *A. woodhousii* in this study is not as comprehensive as for other species so may not be adequate to detect population structure consistent with previous findings of differentiation. However, the sampling does include one sample from Southwest New Mexico and the *STRUCTURE* analysis does not differentiate it from other samples Fig. 2.1. There are two samples assigned to a different population however these are the same two samples found to be highly admixed with *A. fowleri*.

2.4.5 Conclusion

Presented the first genome-wide dataset investigating the evolutionary history of *Anaxyrus*.

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., Brelsford, 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.
- Blair, W. F. (1974). Character Displacement in Frogs. *American Zoologist*, 14(4), 1119–1125. <https://doi.org/10.1093/icb/14.4.1119>
- 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.
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.
- 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.
- Hua, X., & Wiens, J. J. (2013). How Does Climate Influence Speciation? *The American Naturalist*, 182(1), 1–12. <https://doi.org/10.1086/670690>
- 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>
- Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506), 2971–2986. <https://doi.org/10.1098/rstb.2008.0081>
- 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>
- Wen, D., Yu, Y., Zhu, J., & Nakhleh, L. (2018). Inferring Phylogenetic Networks Using PhyloNet. *Systematic Biology*, 67(4), 735–740. <https://doi.org/10.1093/sysbio/syy015>

- Wickam, H. (2016). Ggplot2: Elegant graphics for data analysis. *Springer-Verlag*. Accessed March, 16, 2021.
- Xue, A. T., & Hickerson, M. J. (2015). The aggregate site frequency spectrum for comparative population genomic inference. *Molecular Ecology*, 24(24), 6223–6240. <https://doi.org/10.1111/mec.13447>
- 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>

2.5 Figures

Sampling Distribution

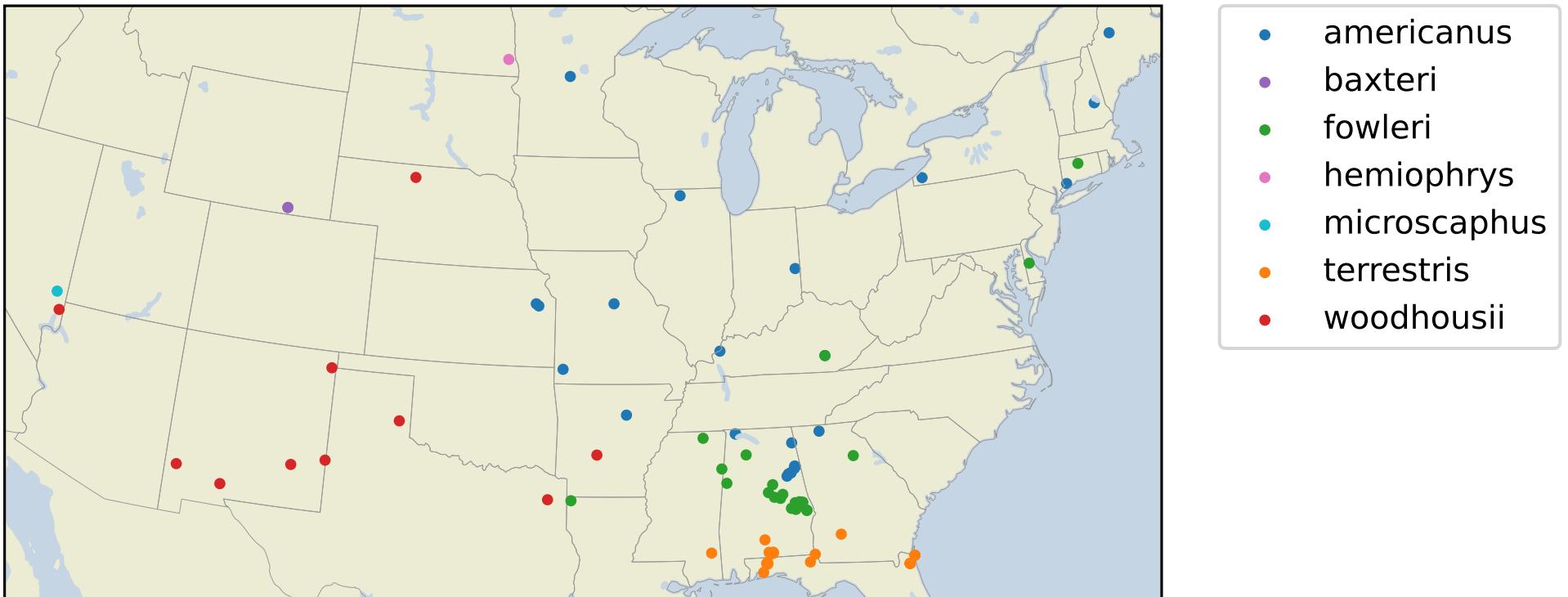


Figure 2.1. Map showing the distribution of the *americanus* group samples sequenced for this study.

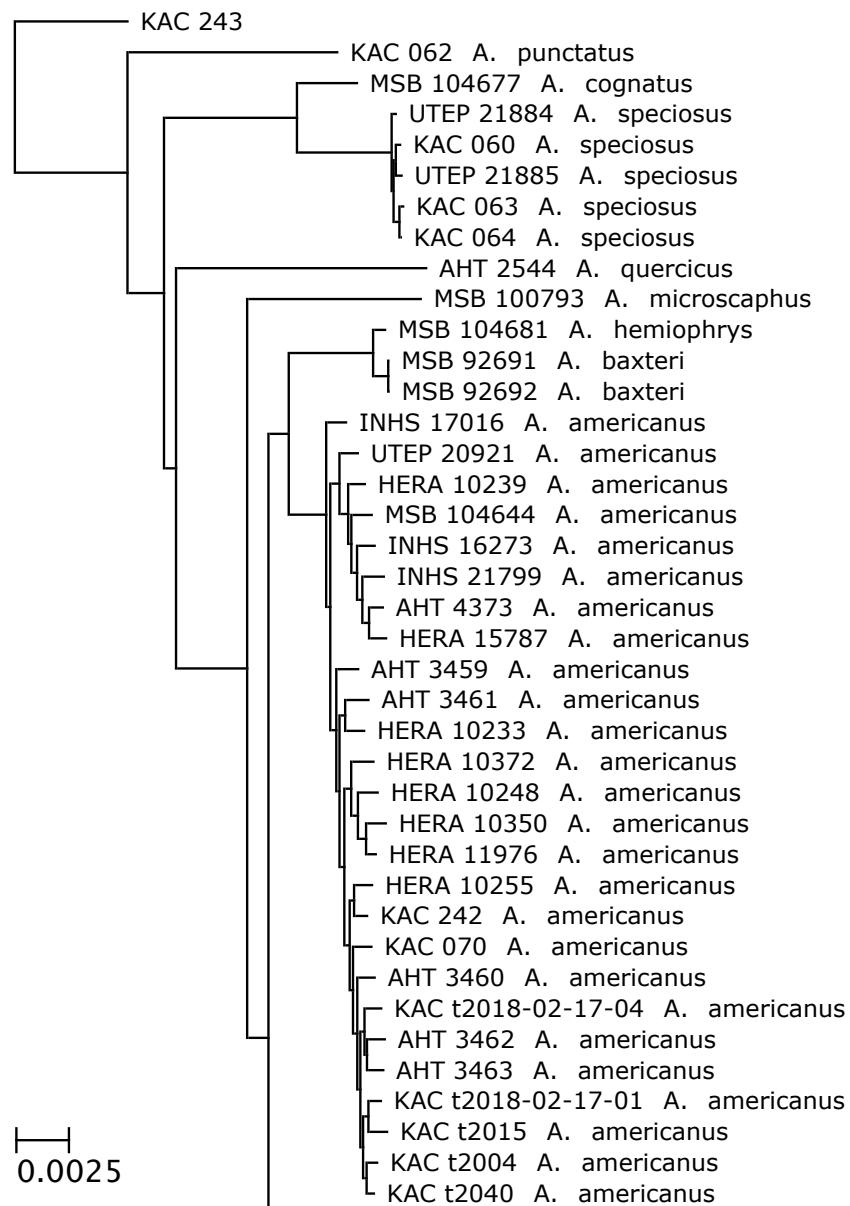


Figure 2.2. Part of maximum likelihood phylogeny inferred with *IQ-TREE*. The values associated with nodes are the ultra fast bootstrap support values rounded down to the nearest whole number. The tree was plotted using ETE 3.1.2 (Huerta-Cepas et al., 2016).

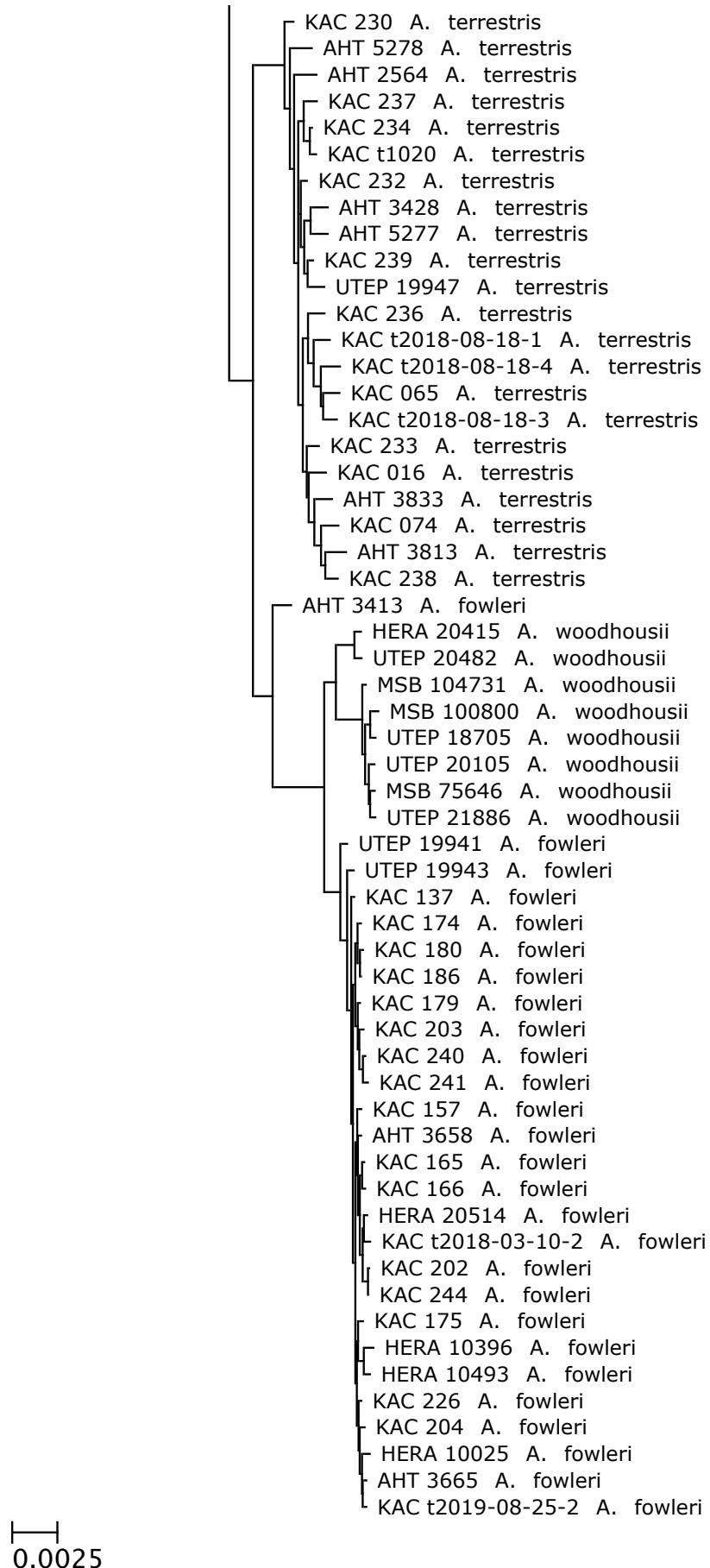


Figure 2.3. Part of maximum likelihood phylogeny inferred with IQ-TREE. The values associated with nodes are the ultra fast bootstrap support values rounded down to the nearest whole number. The tree was plotted using ETE 3.1.2 (Huerta-Cepas et al., 2016).

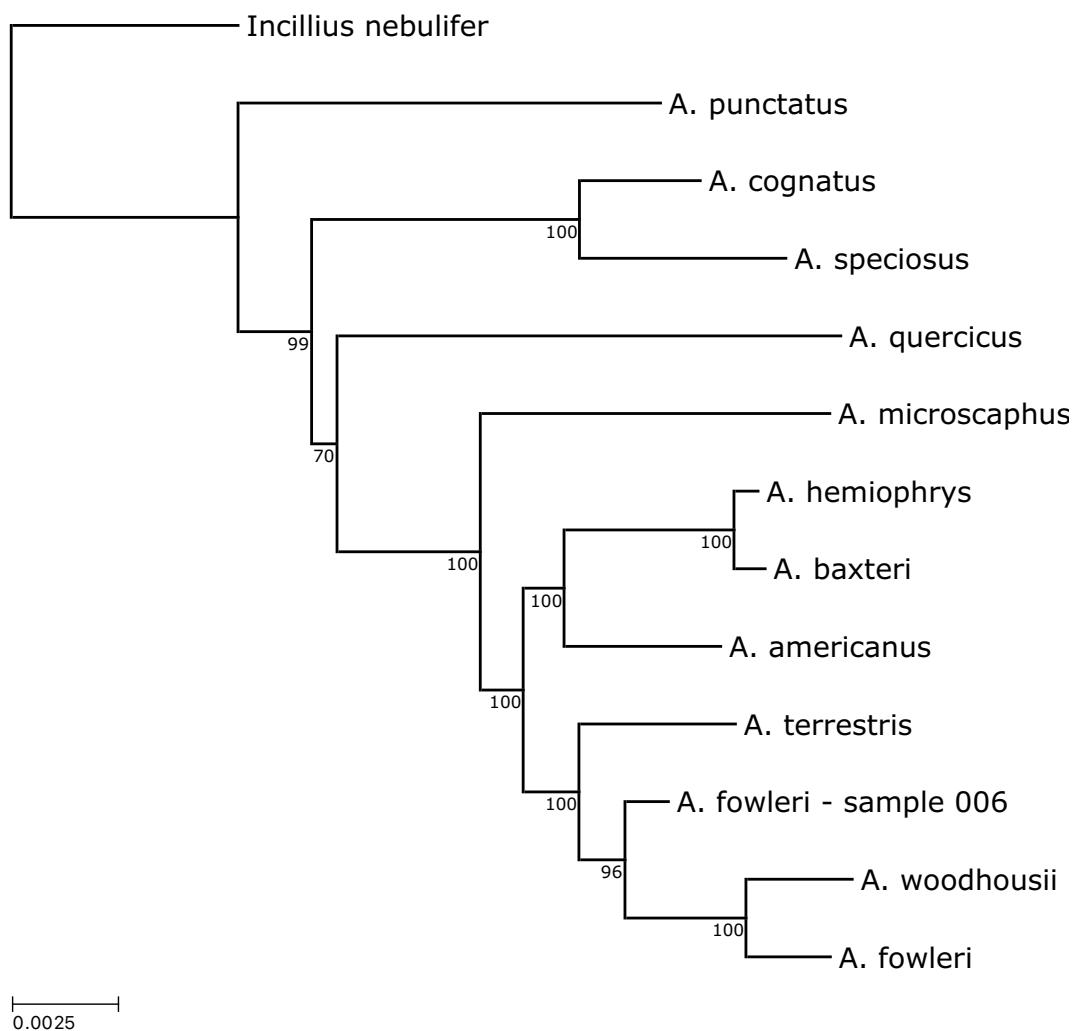


Figure 2.4. Maximum likelihood phylogeny inferred with *IQ-TREE* with clades for each species collapsed. The values associated with nodes are the ultra fast bootstrap support values rounded down to the nearest whole number. The tree was plotted using ETE 3.1.2 (Huerta-Cepas et al., 2016).

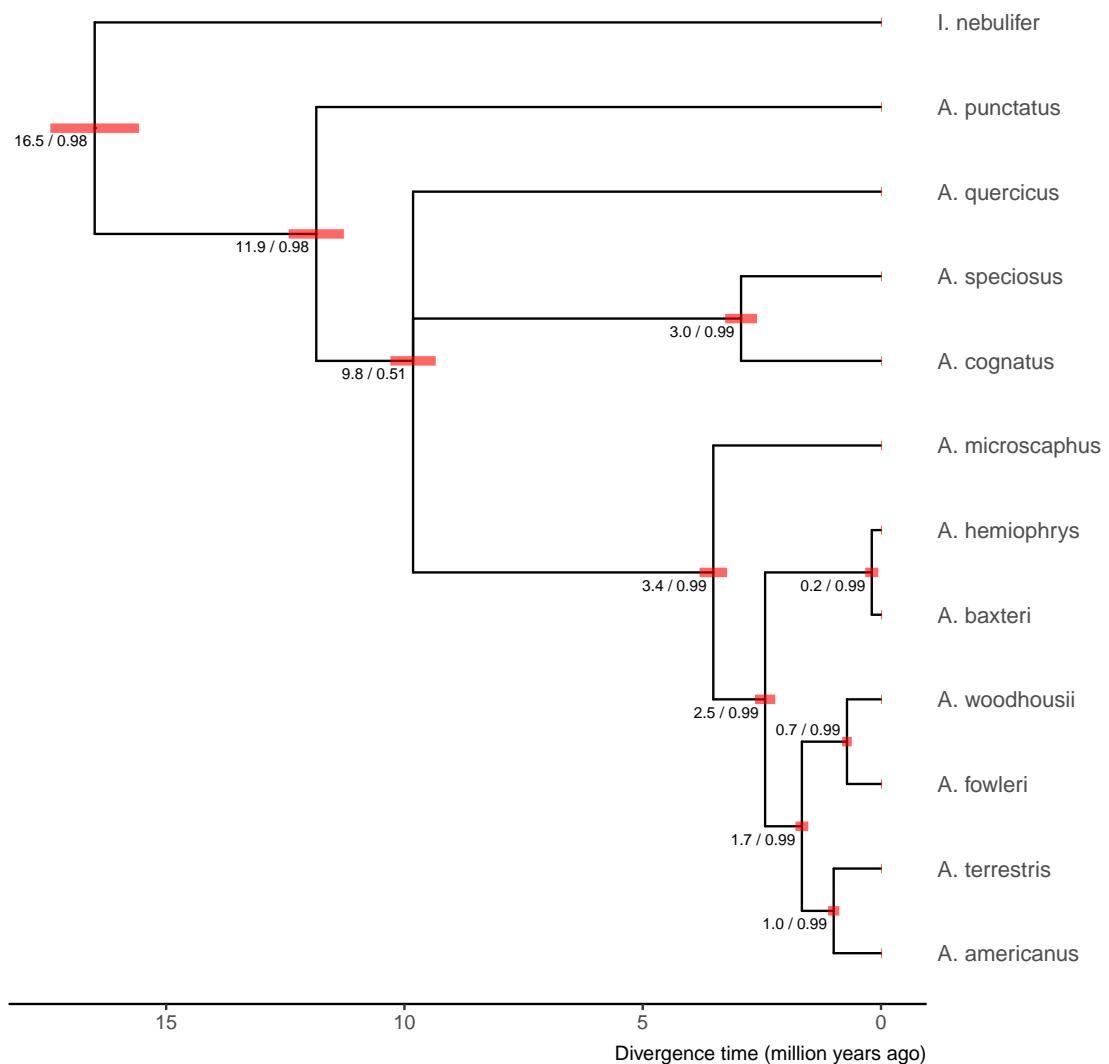


Figure 2.5. The maximum a posteriori tree inferred under a multispecies coalescent model by *phycoeval*. Branch lengths are rescaled from expected substitutions per site to millions of years using secondary time calibrations (*Materials and Methods*). Numbers displayed at each node are the mean posterior 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).

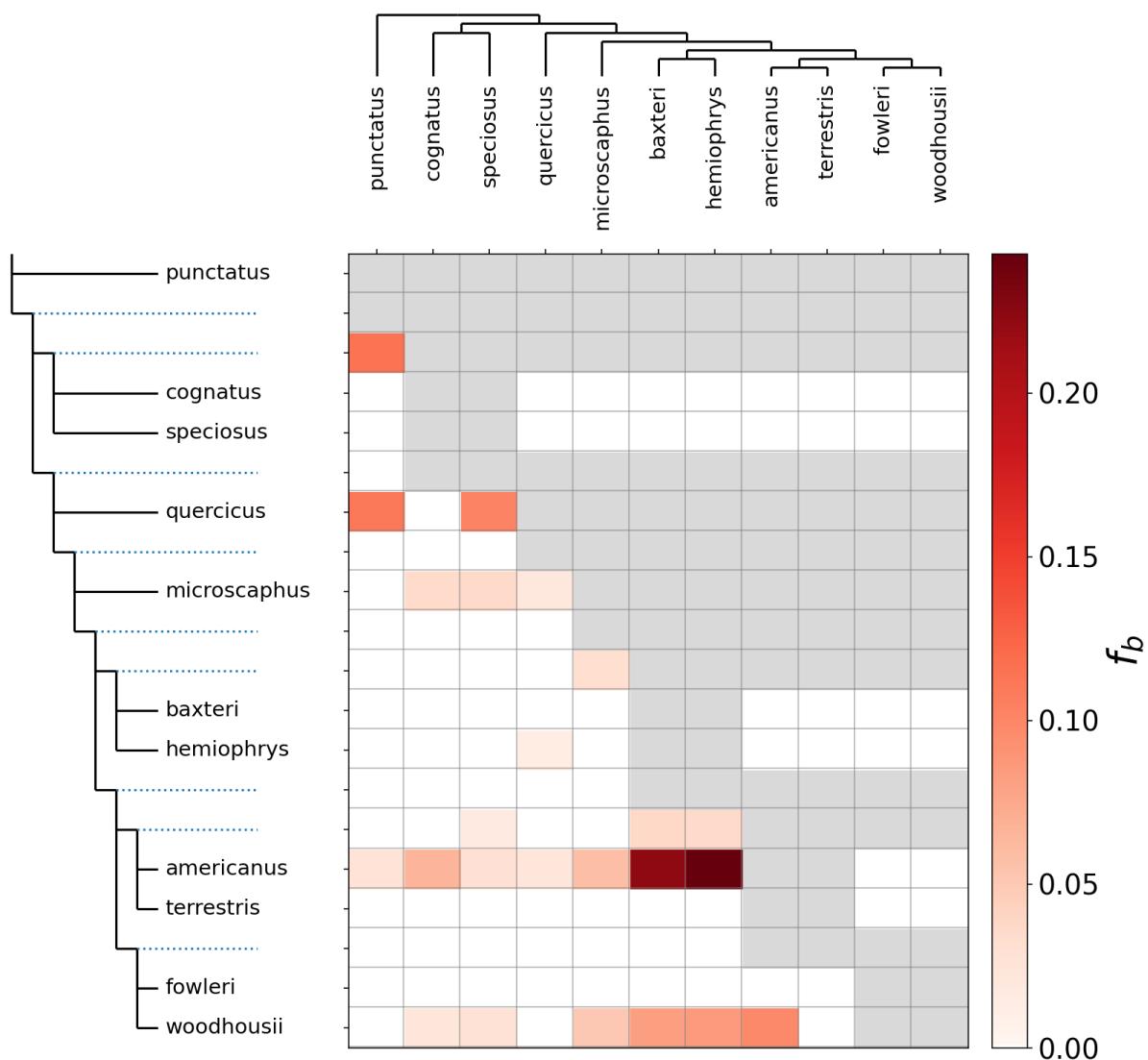


Figure 2.6. Heatmap showing the value of the f -branch statistic computed for all pairs of *Anaxyrus* species. The f -branch statistic indicates the proportion of excess allele sharing between a species on the x-axis and branch on the y-axis (relative to its sister branch). Excess allele sharing between 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

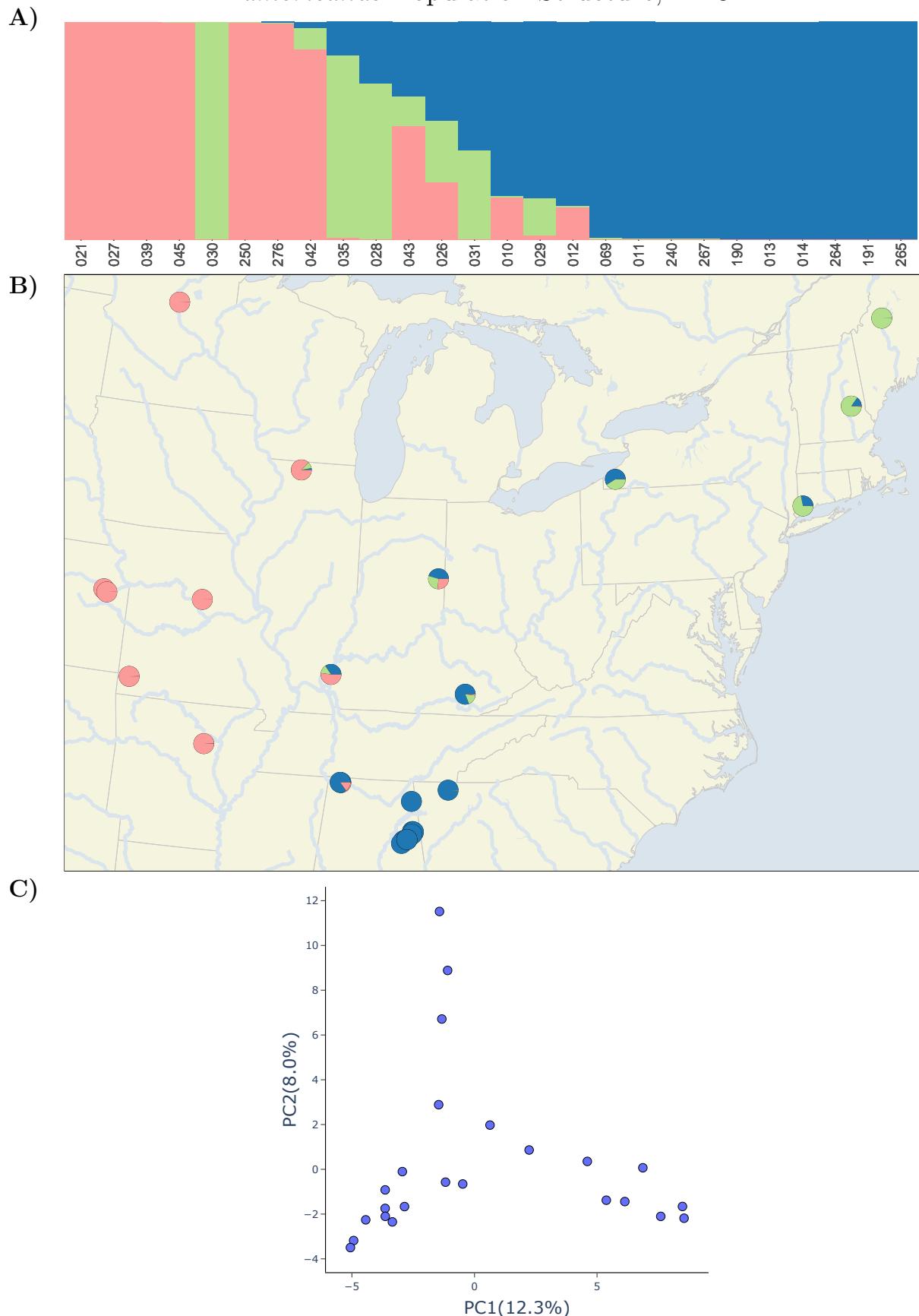


Figure 2.7. *A. americanus* population structure with K=3. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with K=3. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. americanus* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. americanus Population Structure, K=2

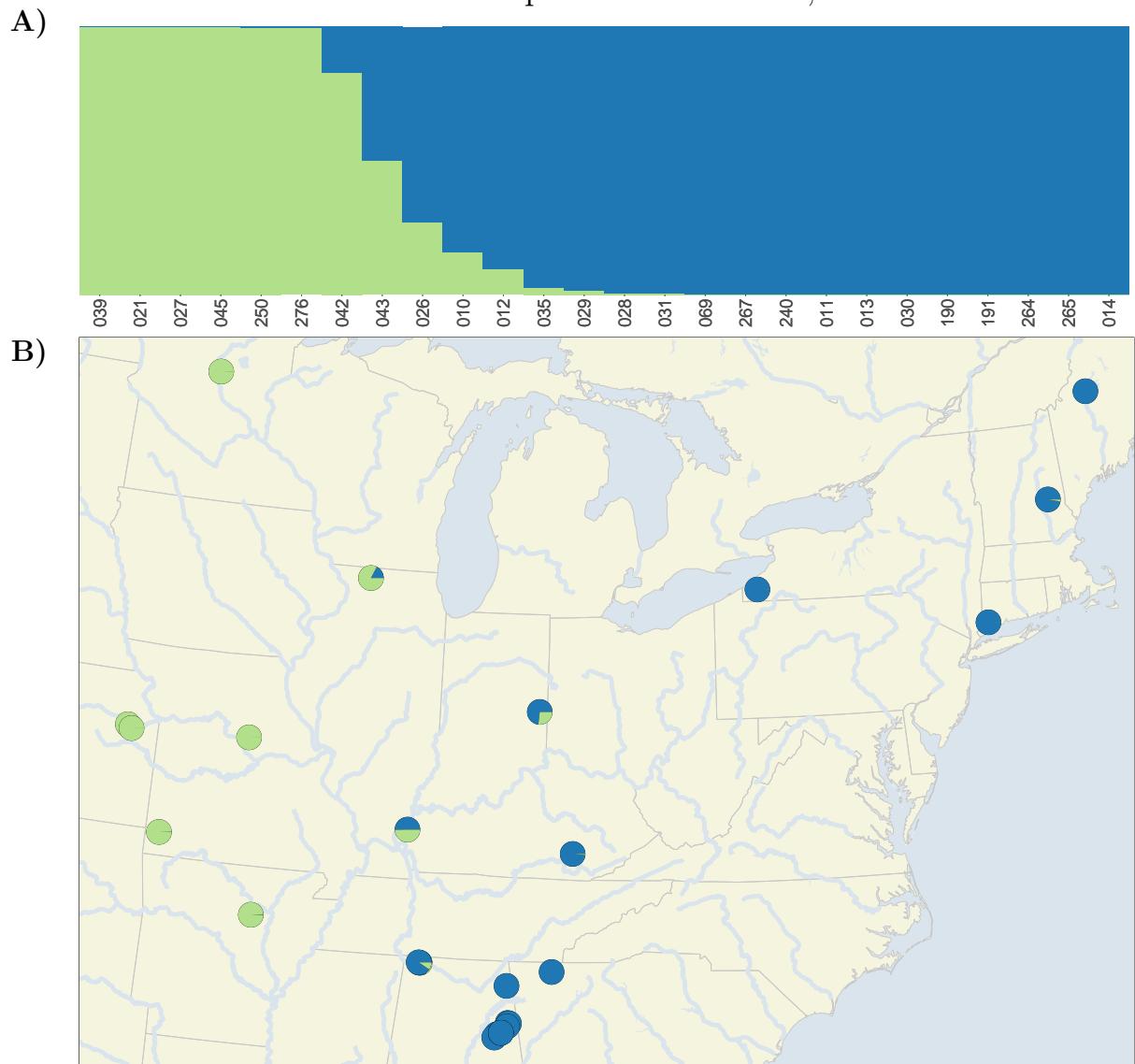


Figure 2.8. *A. americanus* population structure with K=2. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. americanus* samples.

A. fowleri Population Structure

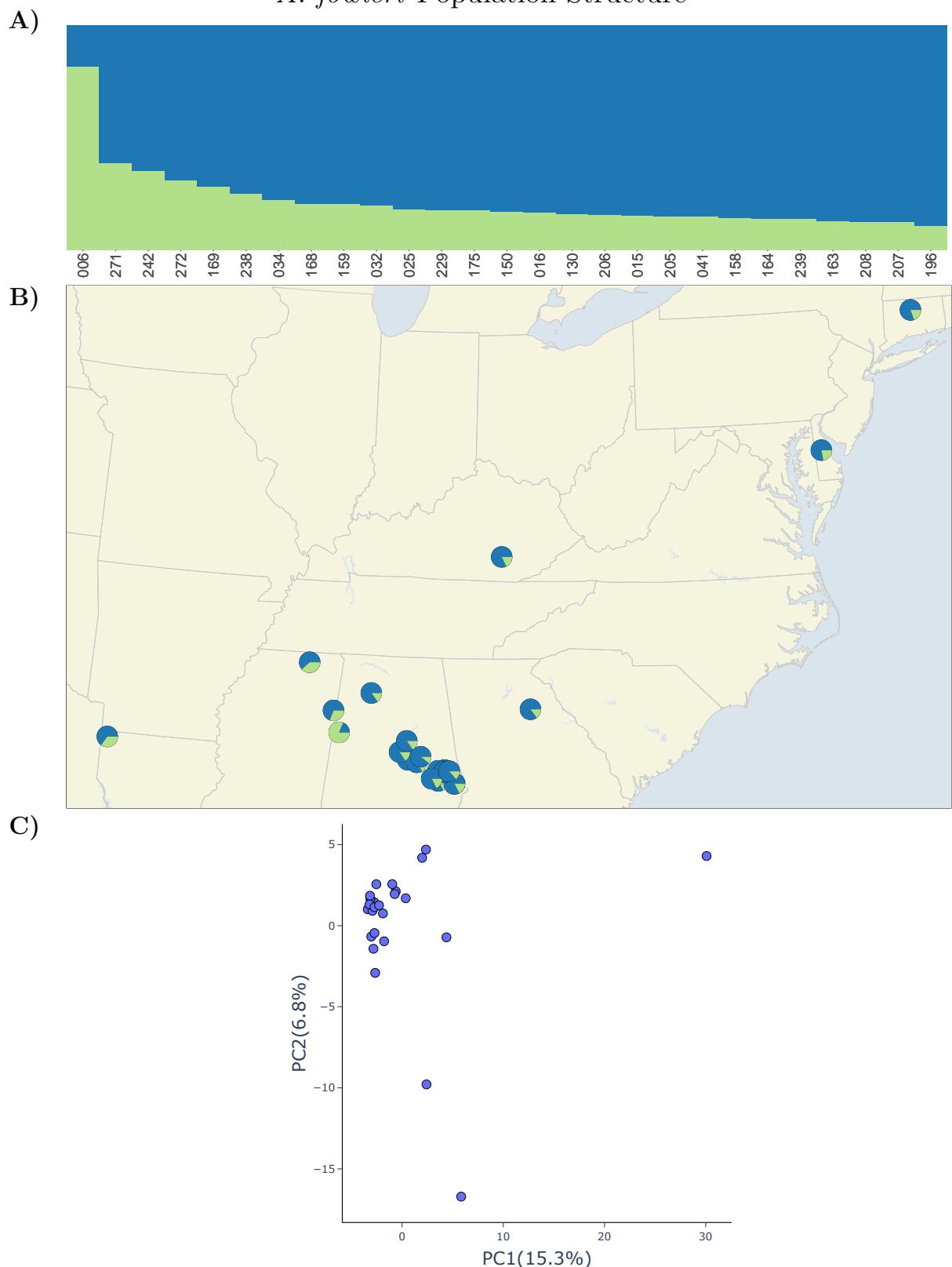


Figure 2.9. *A. fowleri* population structure. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with $K=2$. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. fowleri* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. terrestris Population Structure

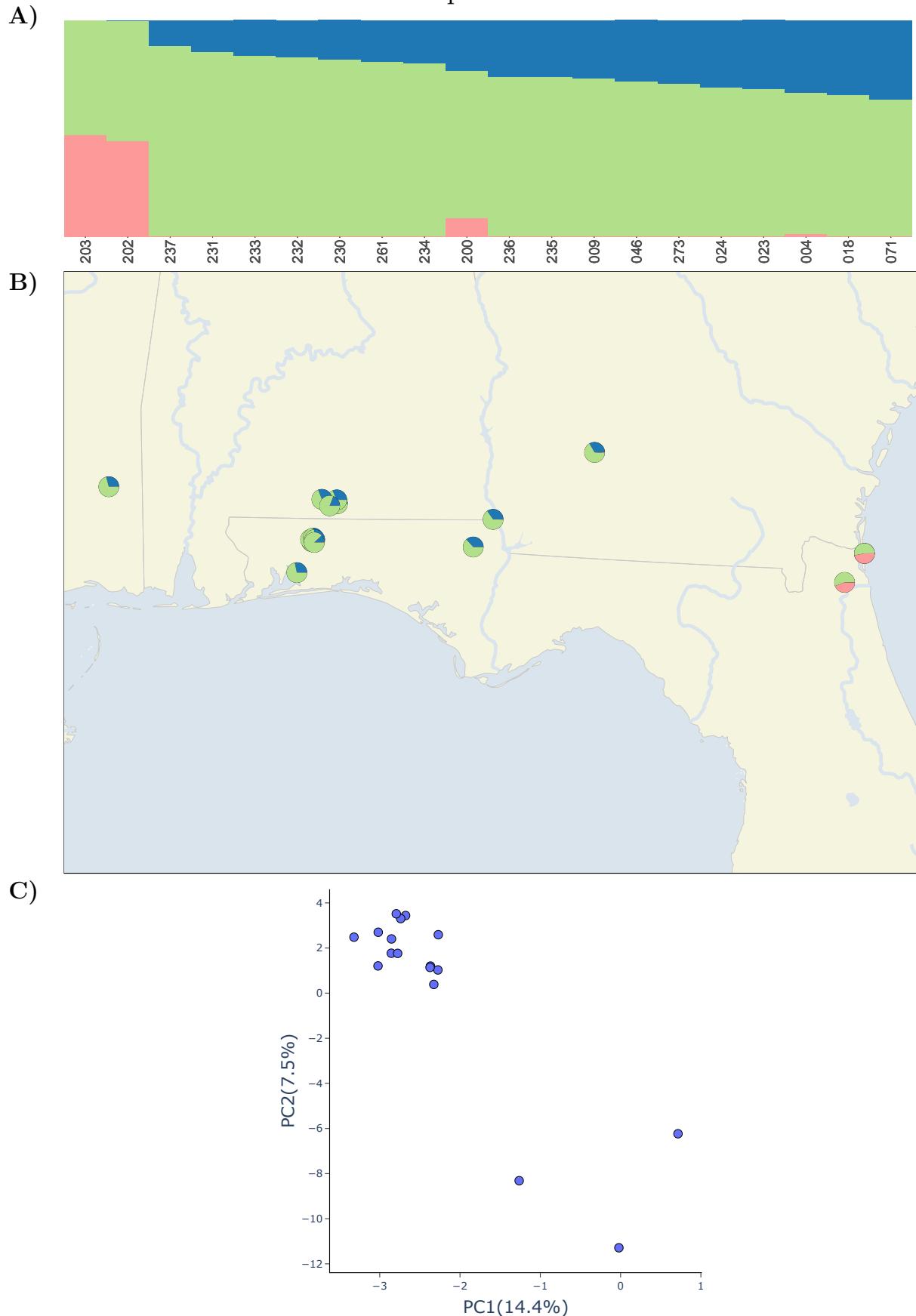


Figure 2.10. *A. terrestris* population structure. A) Barplot with admixture coefficients from the STRUCTURE analysis with $K=2$. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. terrestris* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. woodhousii Population Structure

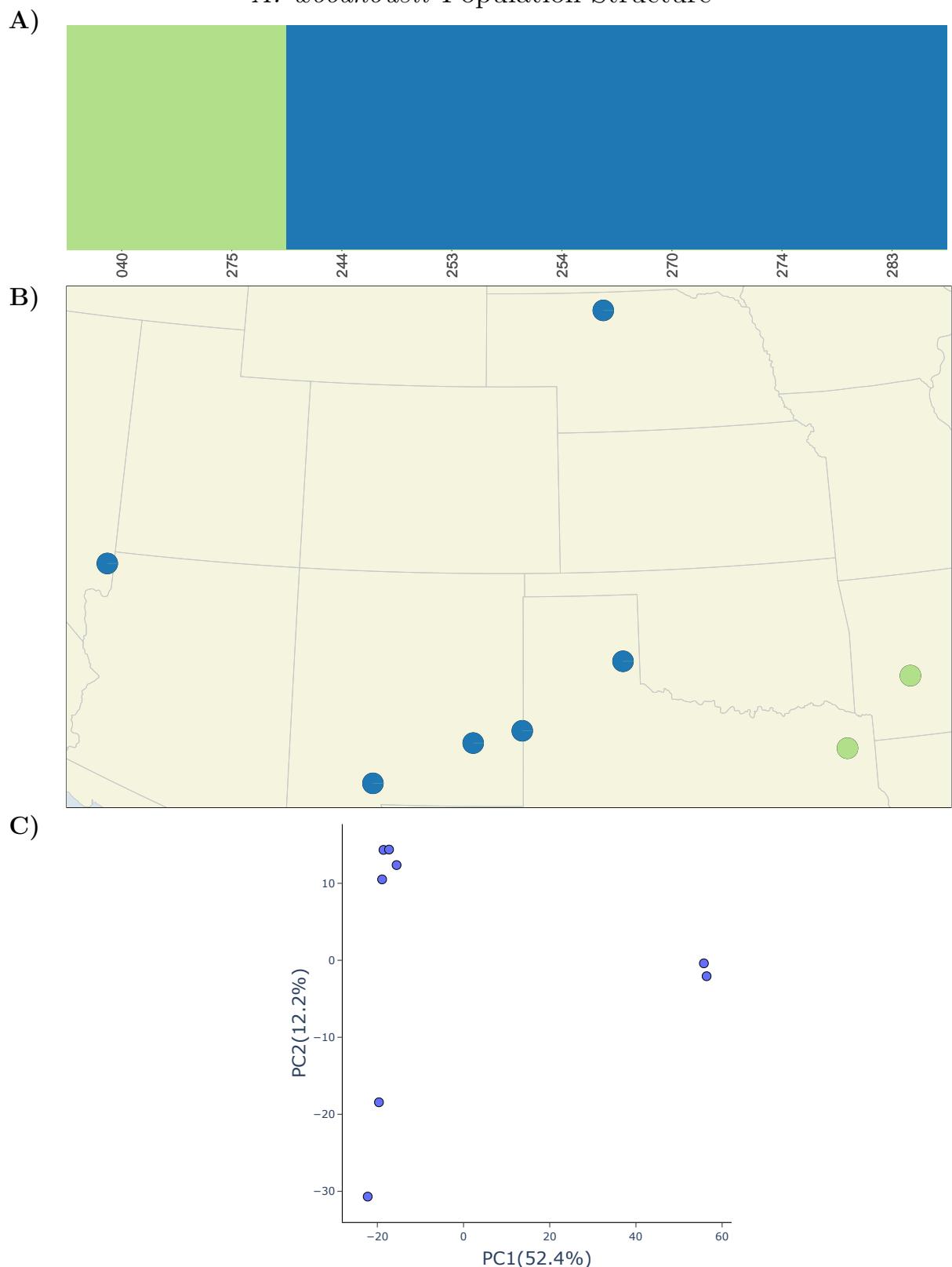
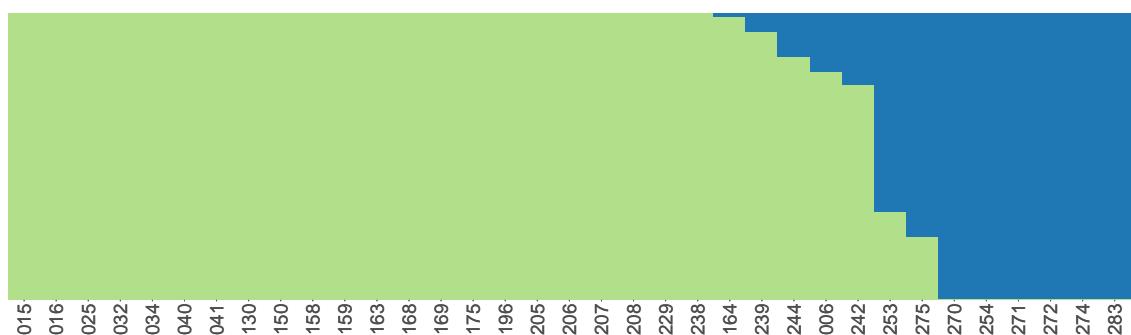


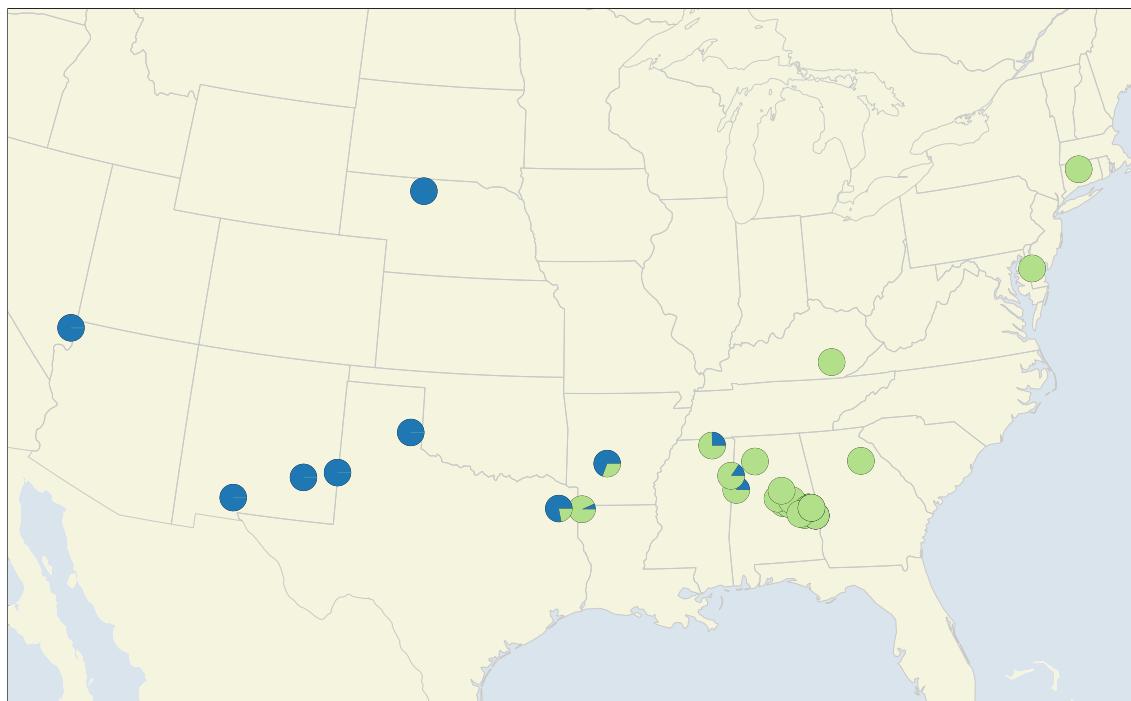
Figure 2.11. *A. woodhousii* population structure. A) Barplot with admixture coefficients from the STRUCTURE analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. woodhousii* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

A. fowleri + A. woodhousii Population Structure

A)



B)



C)

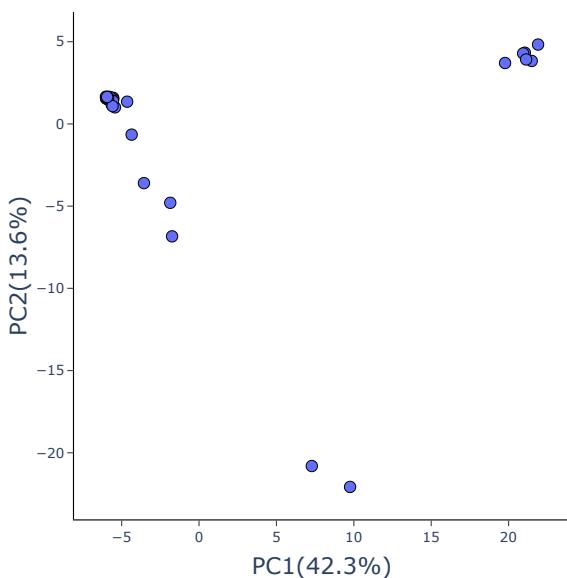


Figure 2.12. Estimates of admixture between *A. fowleri* and *A. woodhousii*. A) Barplot with admixture coefficients from the *STRUCTURE* analysis with K=2. B) Sample map with pie chart markers showing the sampling location and estimated ancestry coefficients of *A. woodhousii* samples. C) Plot showing principal component one and two from the PCA performed on SNP data.

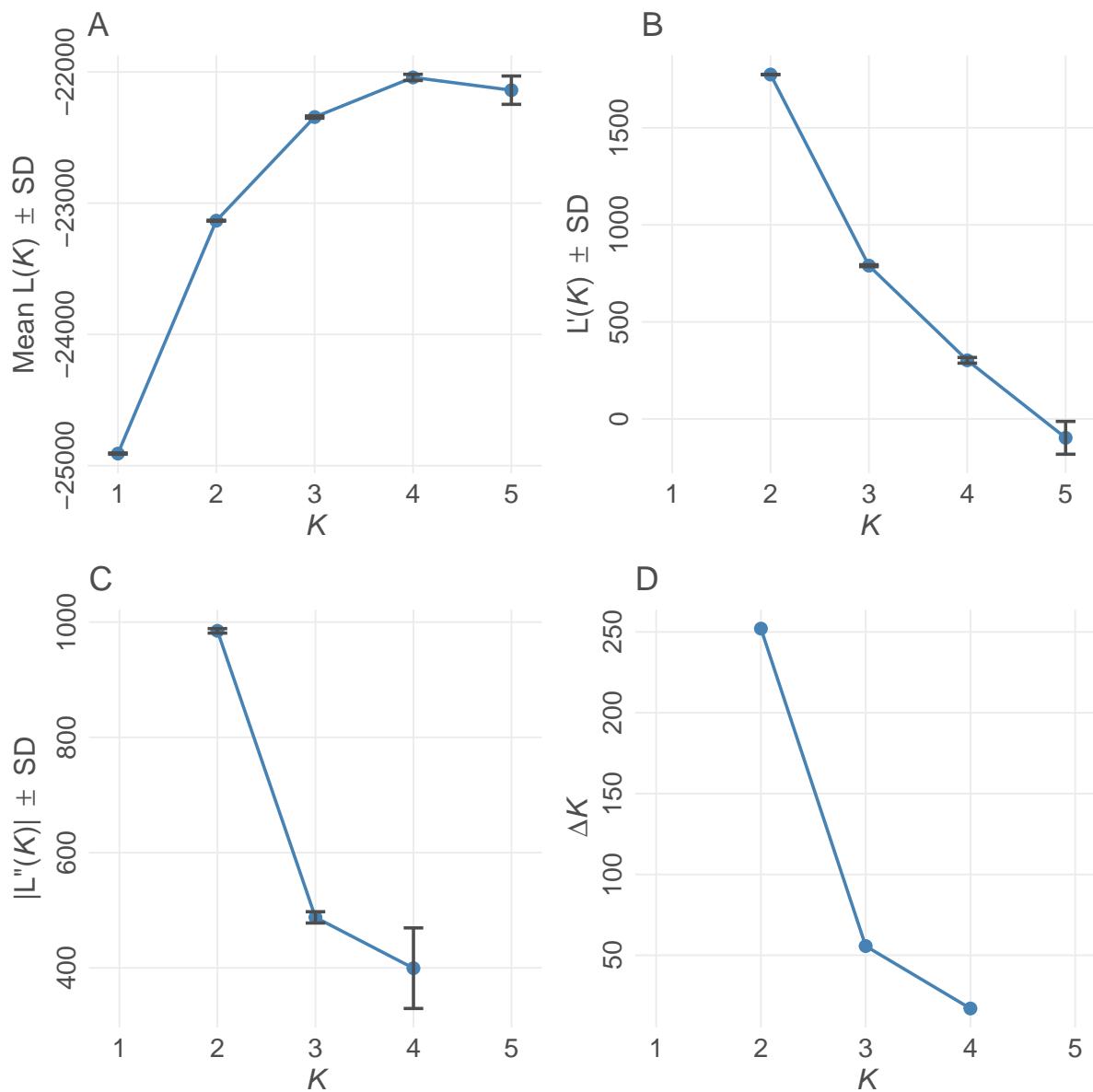


Figure 2.13.
Delta K for americanus

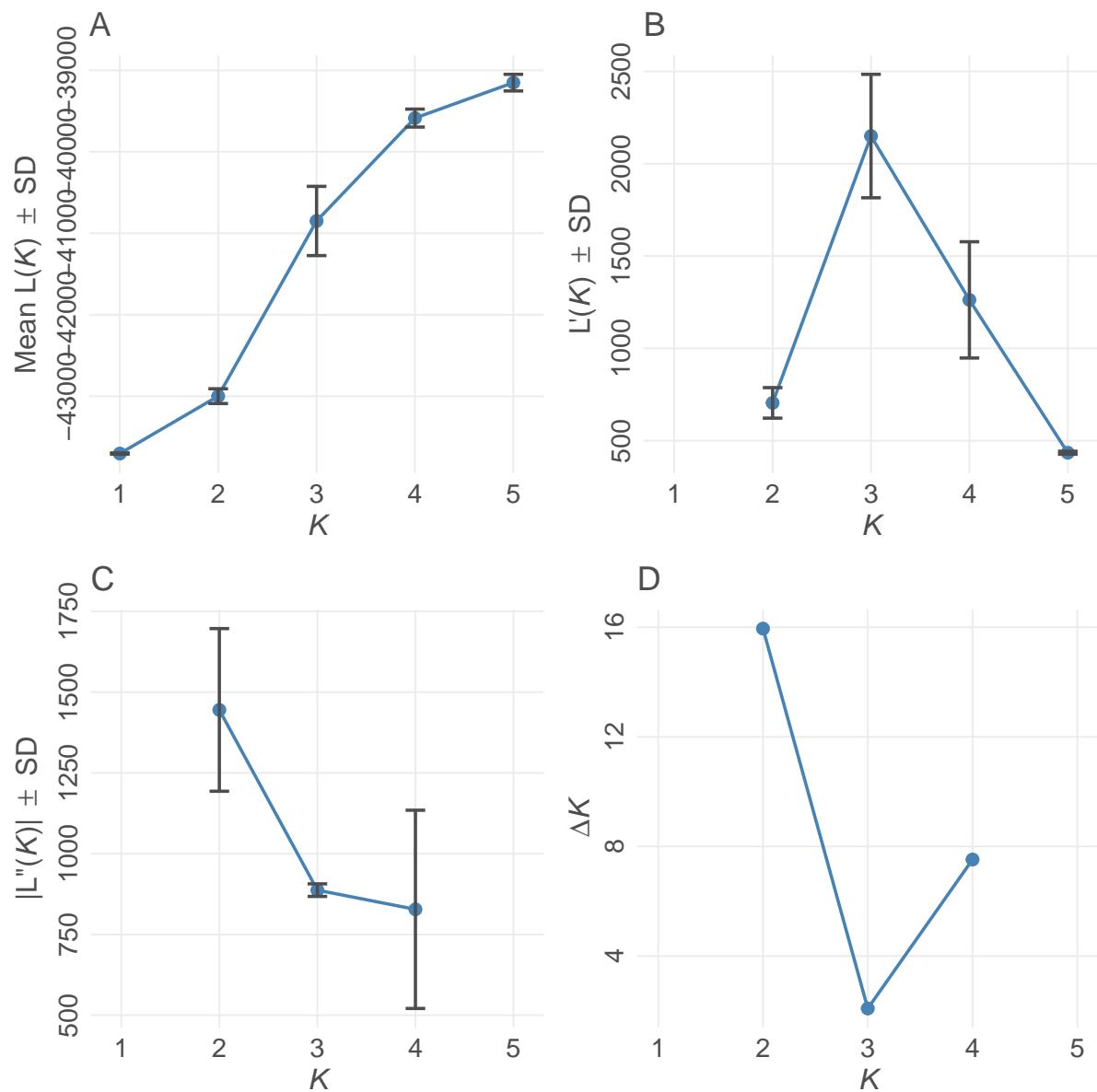


Figure 2.14.
Delta K for fowleri

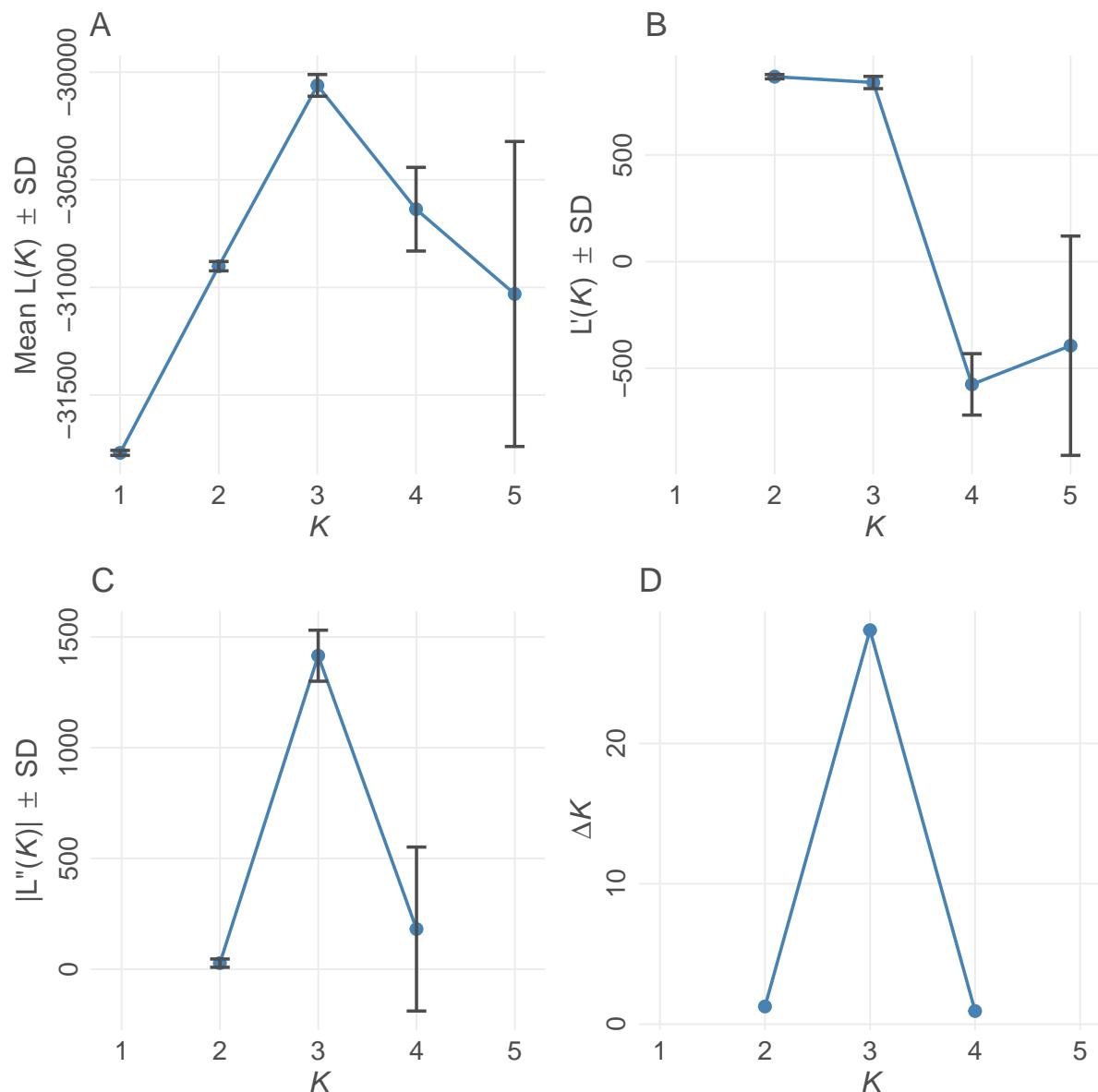


Figure 2.15.
Delta K for terrestris

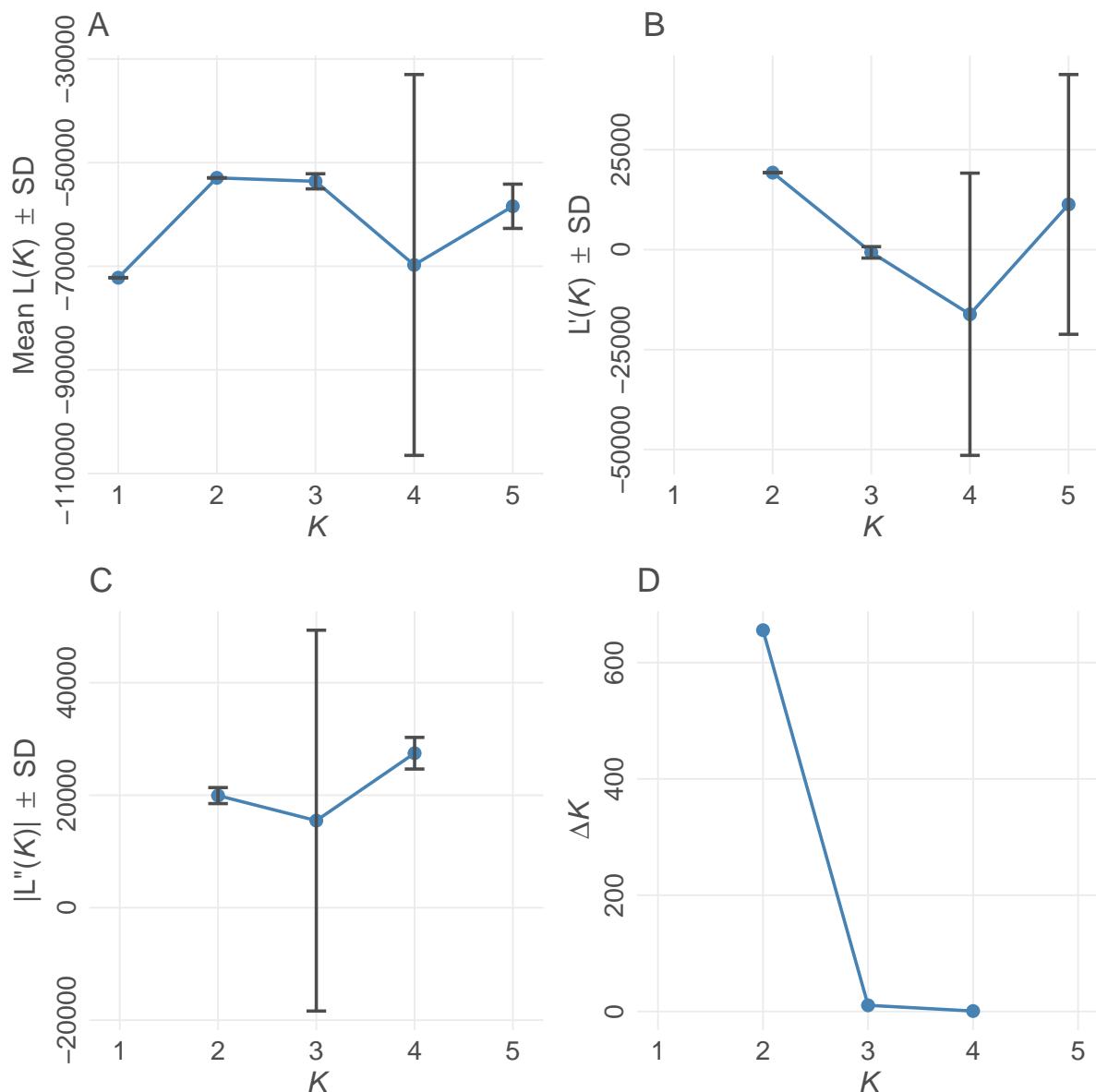


Figure 2.16.
Delta K for woodhousii

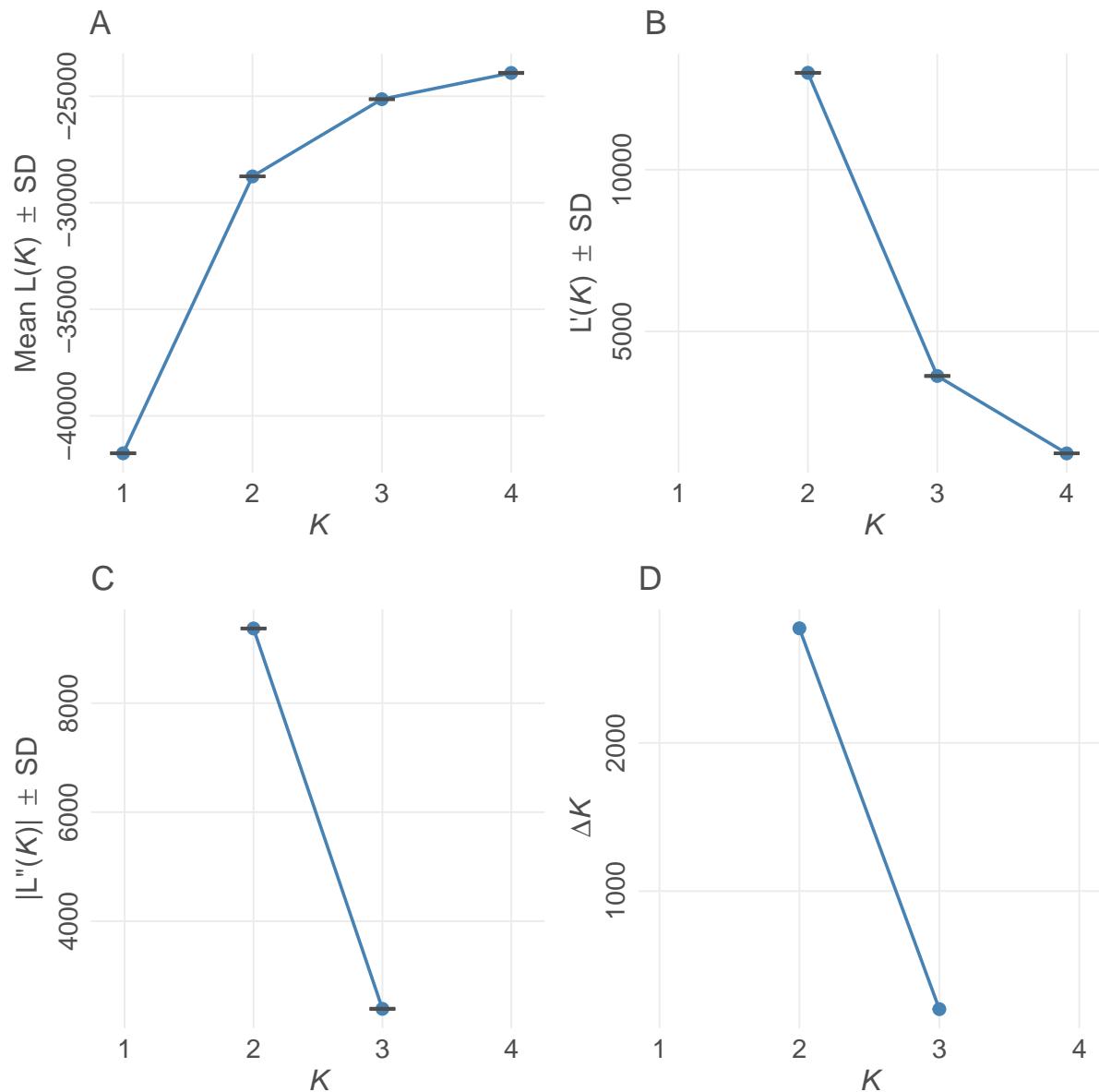


Figure 2.17. Delta K for *woodhousii x fowleri*

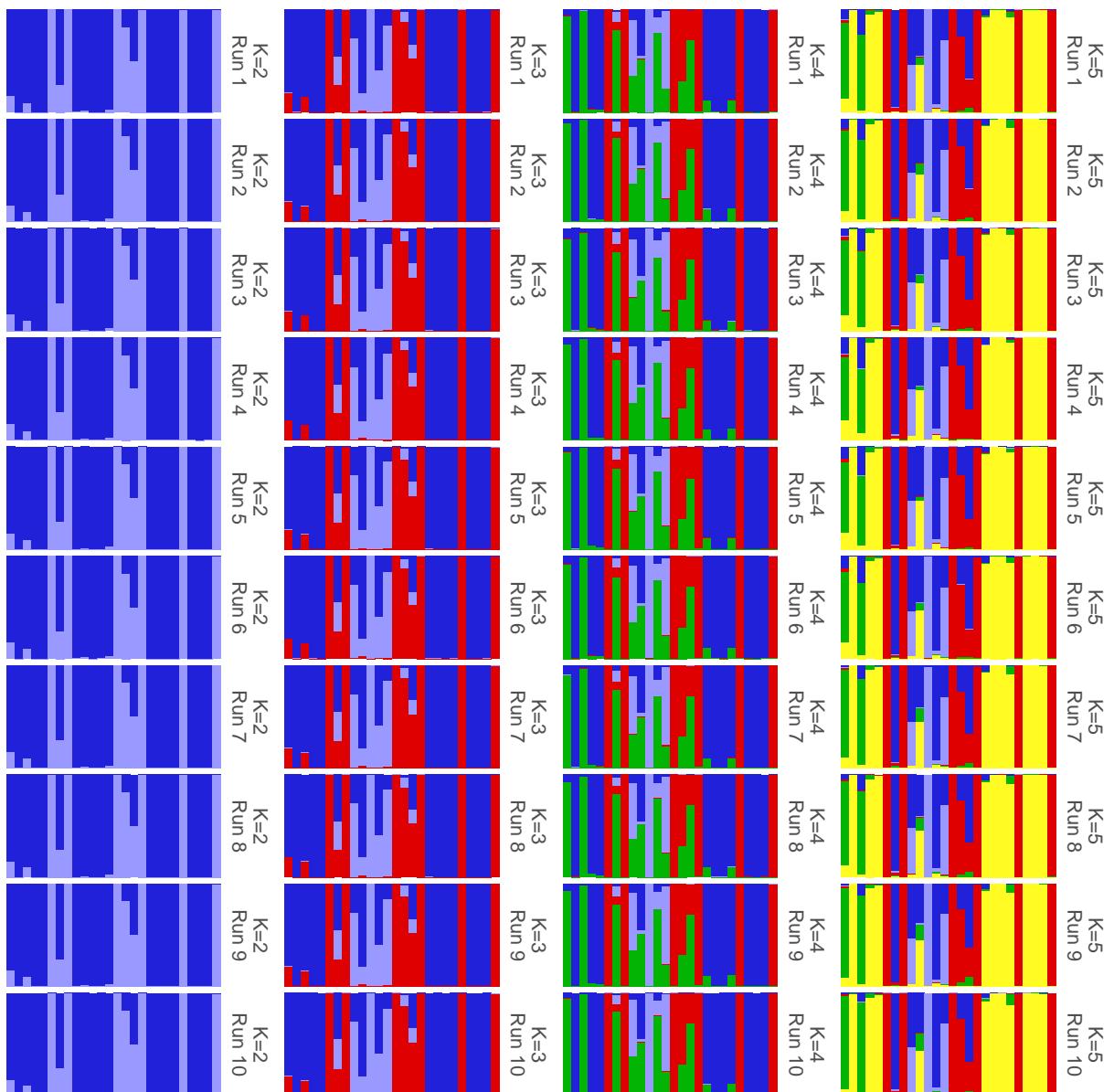


Figure 2.18.
All *STRUCTURE* runs *americanus*.

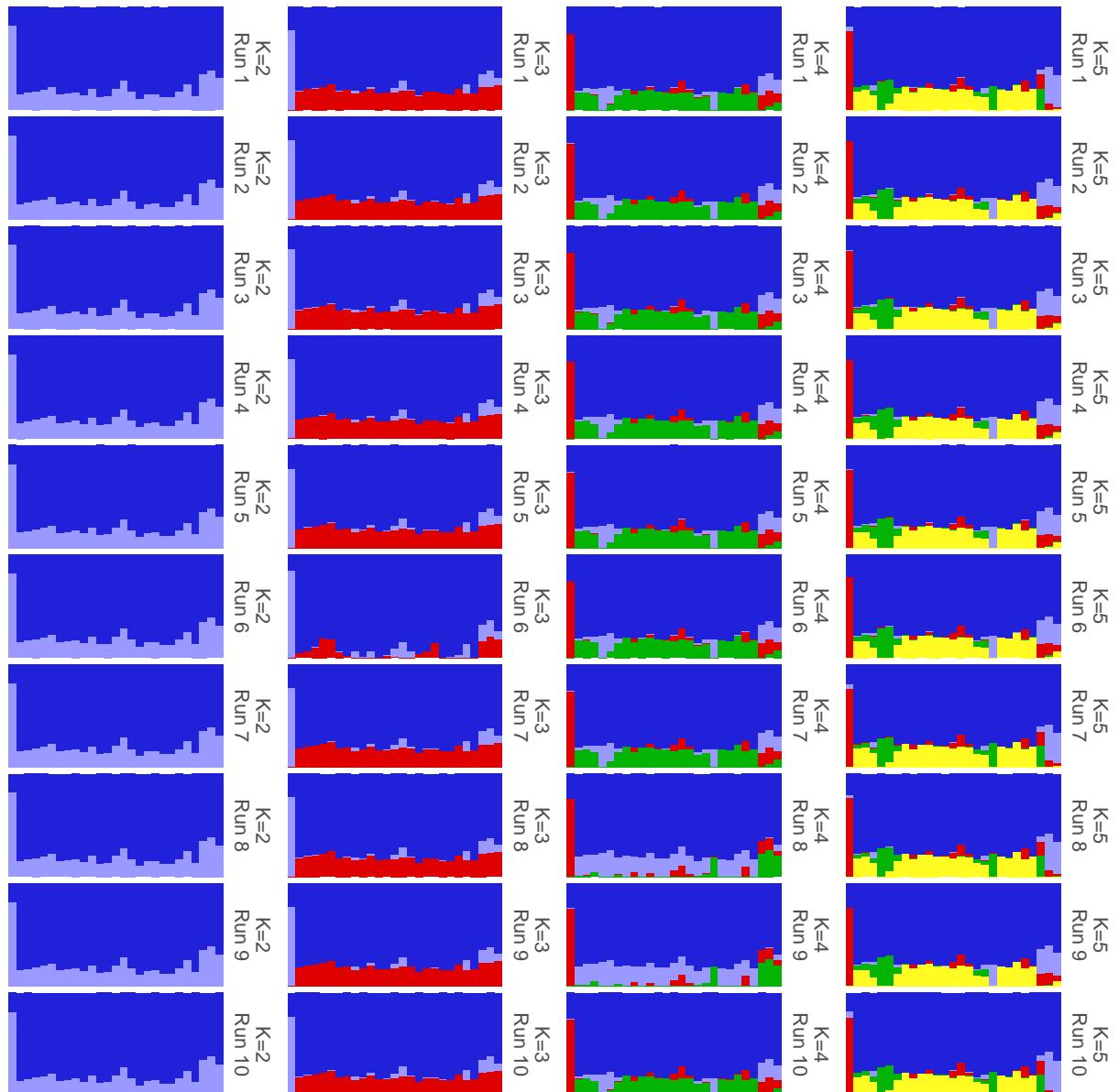


Figure 2.19.
All *STRUCTURE* runs *fowleri*.

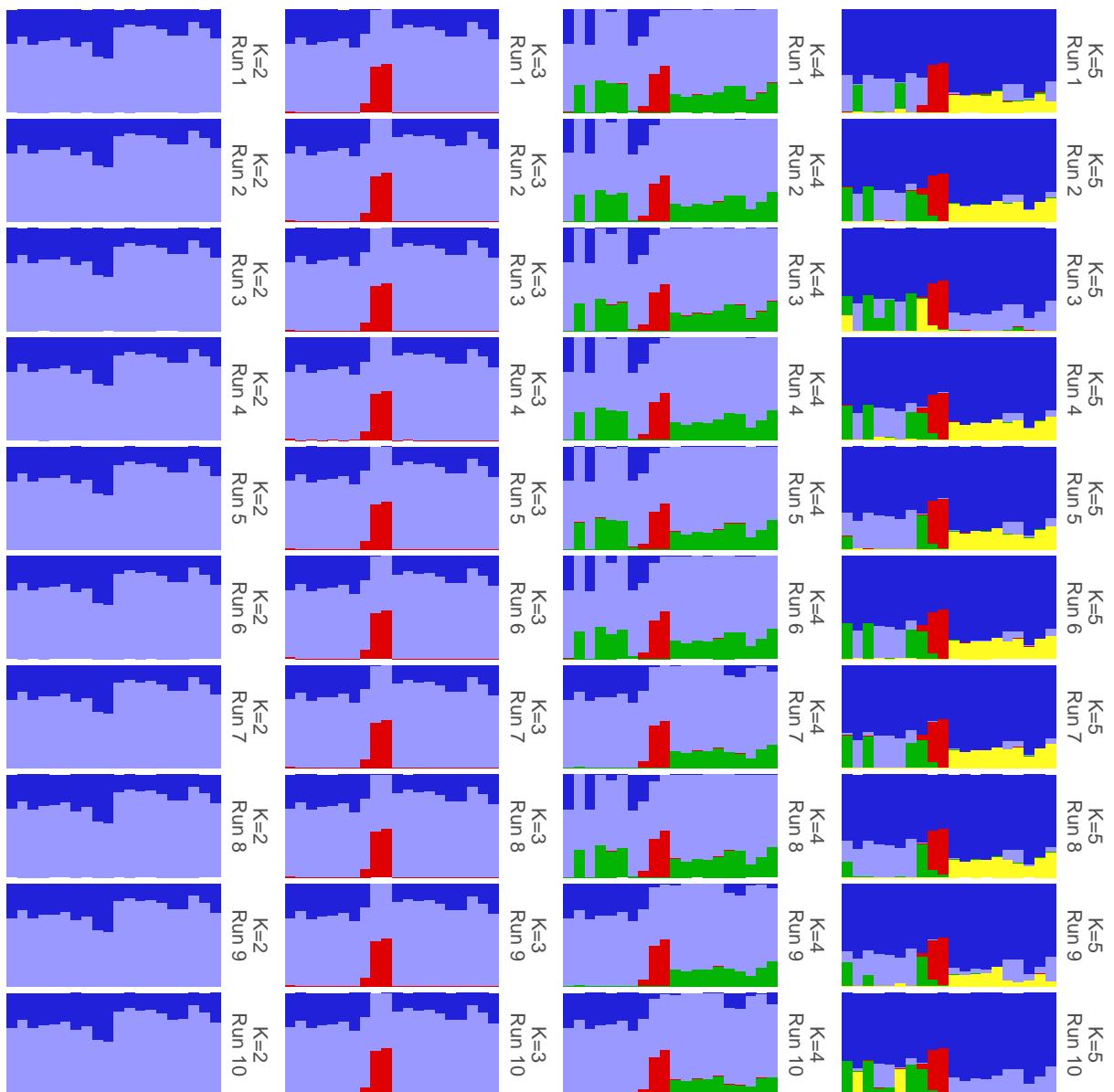


Figure 2.20.
All *STRUCTURE* runs terrestris.

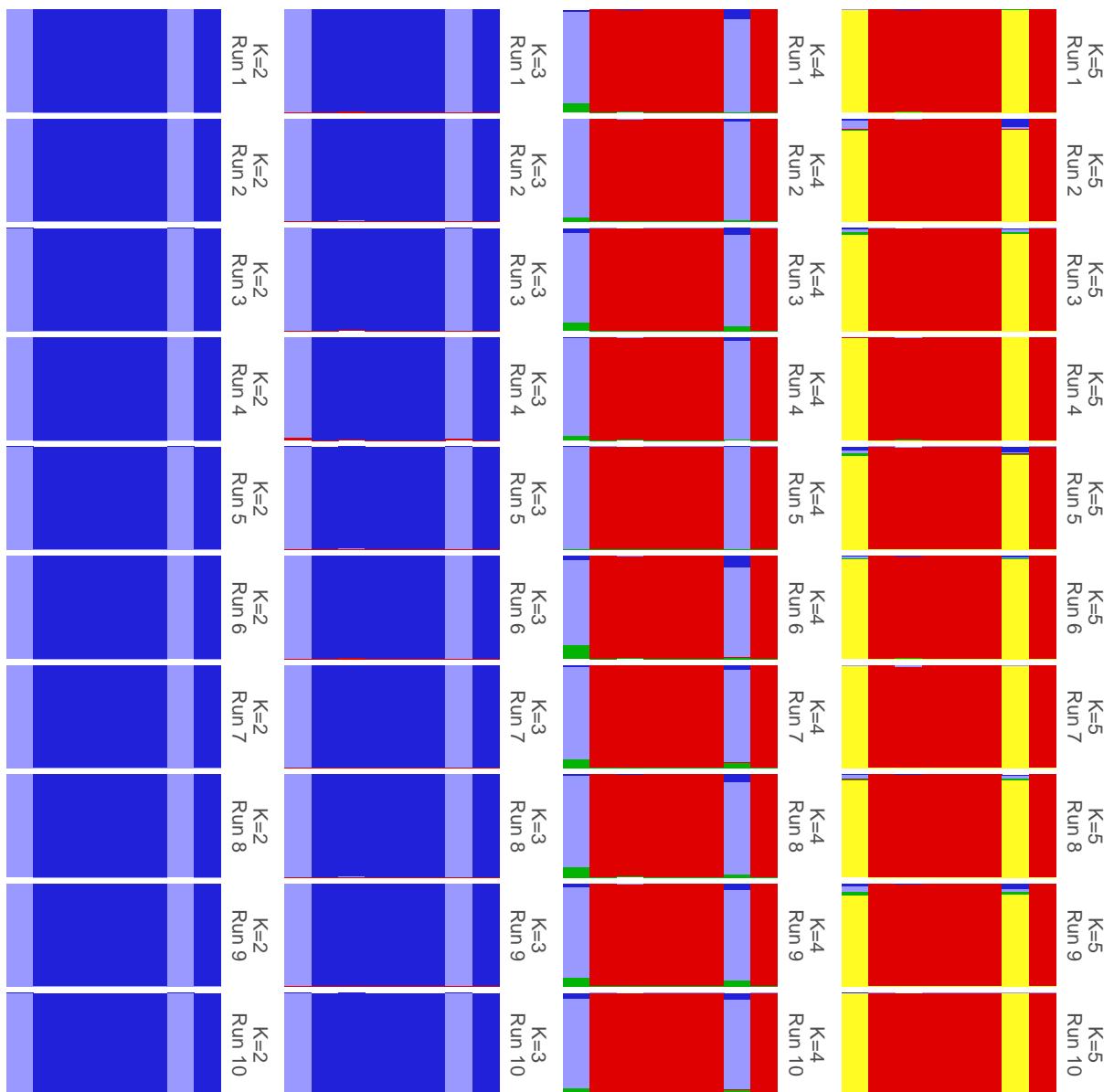


Figure 2.21.
All *STRUCTURE* runs woodhousii.

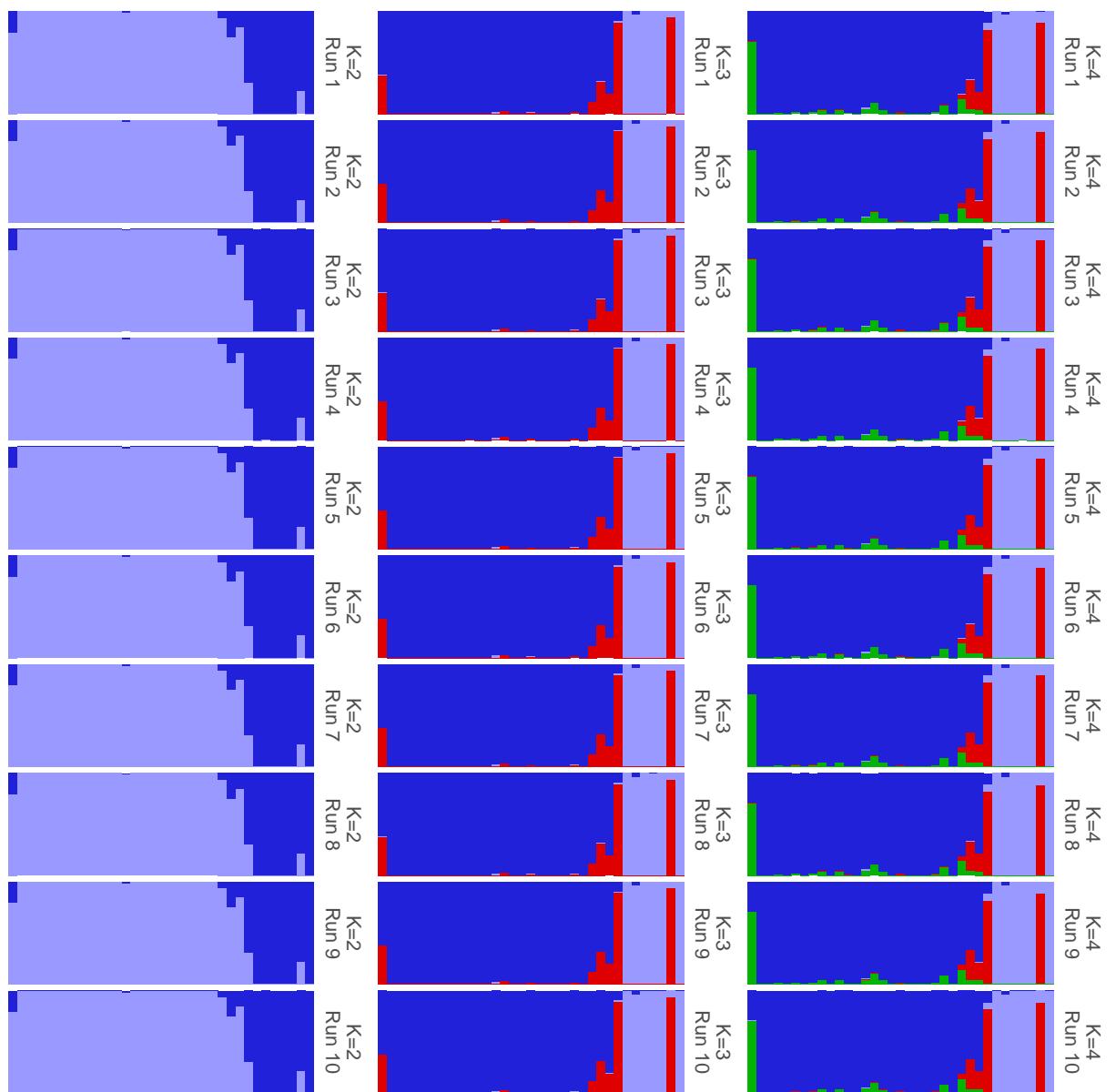


Figure 2.22.
All *STRUCTURE* runs fowleri x woodhousii.

2.6 Tables

Table 2.1. 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 2.1 – 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	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
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

Continued on next page

Table 2.1 – 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
100	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 2.1 – 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 2.1 – 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 2.1 – 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 2.1 – 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 2.1 – 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