Online ISSN: 2588-3526

Journal of Wildlife and Biodiversity

*Volume 9(3): 318-339 (2025) (*http://www.wildlife-biodiversity.com/*)*

Research Article

Integrative phylogenetic, genomic, acoustic, and spatial analyses of U.S. frog species using mitochondrial COI sequences

My Abdelmajid Kassem

Plant Genomics and Bioinformatics Lab, Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC 28301, USA.
Email: mkassem@uncfsu.edu

Received: 16 June 2025 / Revised: 19 September 2025 / Accepted: 01 October 2025/ Published online: 16 October 2025.
How to cite: Kassem, MA. (2025). Integrative phylogenetic, genomic, acoustic, and spatial analyses of U.S. frog species using mitochondrial COI sequences, Journal of Wildlife and Biodiversity, 9(3), 318-339. DOI: https://doi.org/10.5281/zenodo.17386849

Abstract

Amphibians are critical bioindicators of environmental health, yet integrative frameworks combining genomic, acoustic, and spatial data remain underutilized in regional biodiversity assessments. Here, I present the first integrative and reproducible workflow for analyzing phylogenetic, genomic, acoustic, and spatial data across 30 U.S. frog species, leveraging publicly available mitochondrial COI sequences, genome size records, frog call recordings, and georeferenced occurrence data. I reconstructed a robust maximum-likelihood phylogeny revealing well-supported clade-level relationships among *Hylidae, Ranidae,* and *Bufonidae*. Genome size data mapped onto the phylogeny reveal family-level trends, with *Hylidae* showing consistently larger genomes. Acoustic features such as call duration and dominant frequency from five representative species showed significant interspecific variation, with PCA clustering reflecting taxonomic structure. Spatial analysis of >3,000 GBIF occurrence records identified biodiversity hotspots in the Southeastern U.S., aligned with environmental gradients. All analyses were conducted using open-source tools (MAFFT, FastTree, librosa, geopandas, scikit-bio) in a reproducible pipeline shared via GitHub. This study provides a computationally transparent, multi-modal framework for amphibian biodiversity research and conservation planning.

Keywords: Amphibian biodiversity, COI phylogenetics, genome size evolution, spatial distribution modeling, open-source pipeline, GBIF

Introduction

Amphibians are critical bioindicators of environmental health, exhibiting sensitivity to habitat degradation, climate change, pollution, and emerging infectious diseases (Blaustein et al., 1994; Stuart et al., 2004). Globally, amphibian declines have highlighted the urgent need for systematic biodiversity assessments to inform conservation priorities (Wake and Vredenburg, 2008; Grant et

al., 2017). Frogs, as the most speciose amphibian group, contribute to ecosystem stability by regulating insect populations and serving as prey for higher trophic levels (Duellman and Trueb, 1994; Wells, 2007). They also provide tractable systems for studying vertebrate development, behavior, and evolutionary processes due to their diverse reproductive modes and ecological strategies (Wells, 2007; Bonett and Blair, 2017). In the United States, frog species diversity spans a wide range of ecological zones, including temperate forests, grasslands, wetlands, and arid regions, reflecting a complex interplay of historical biogeography, ecological adaptation, and climate variability (Petranka, 1998; Pyron and Wiens, 2011). Despite extensive ecological studies, comprehensive phylogenetic and integrative trait-based analyses across US frog species remain limited, creating gaps in our understanding of evolutionary processes shaping their diversity.

Molecular phylogenetics has revolutionized biodiversity assessments by providing robust frameworks for species delimitation, cryptic species detection, and evolutionary inference (Hebert et al., 2003; Vences et al., 2005). Among molecular markers, the mitochondrial cytochrome oxidase I (COI) gene has been widely adopted for DNA barcoding and phylogenetic studies due to its balance between conservation and variability, enabling effective resolution across taxonomic scales (Smith et al., 2005; Vieites et al., 2009). COI barcoding has facilitated the identification of cryptic diversity and biogeographic patterns in amphibians (Vences et al., 2005; Fouquet et al., 2007), but a systematic integration of COI data to assess interspecific relationships across US frog species within a unified phylogenetic framework has not been comprehensively performed.

Genome size variation, often measured as the haploid nuclear DNA content (C-value), influences cellular and developmental processes, body size, and life history traits, and may reflect evolutionary constraints and adaptations (Gregory, 2002; Sun and Mueller, 2014). Amphibians, particularly anurans, exhibit some of the highest genome size variability among vertebrates, with potential ecological and evolutionary implications (Liedtke et al., 2018; Jonsson and Jonsson, 2019). In frogs, genome size can impact developmental timing, metabolic rates, and ecological tolerances (Santos, 2012; Leiva et al., 2019), yet the phylogenetic patterns of genome size variation across US species remain underexplored. Investigating genome size within a phylogenetic context can yield insights into the evolutionary dynamics shaping amphibian genome architecture and potential trait conservatism across clades (Smith et al., 2018; Liedtke et al., 2018).

Bioacoustics is another critical but often underutilized dimension in amphibian biodiversity studies. Frog calls are species-specific signals used in mate attraction and territorial behaviors,

making them valuable for species identification, ecological monitoring, and understanding the evolution of communication systems (Gerhardt and Huber, 2002; Wells, 2007). Acoustic signals can reflect ecological adaptations, such as habitat type and temperature regimes, and may exhibit phylogenetic signal, indicating evolutionary conservatism or divergence in call traits among related species (Amezquita et al., 2009). Integrating acoustic trait data with phylogenetic analyses allows testing hypotheses about the evolution of acoustic communication systems and the role of sexual selection in diversification (Ryan and Rand, 1993; Wilkins et al., 2013).

Spatial data provide a crucial environmental dimension to biodiversity research by revealing spatial patterns of species distributions and habitat preferences. Geographic occurrence records can be used to investigate species range overlap, habitat specialization, and potential ecological barriers that shape evolutionary trajectories (Pigot and Tobias, 2013). When integrated with phylogenetic frameworks, spatial data support the identification of biogeographic clusters, divergence zones, and regions of conservation concern (Wiens and Donoghue, 2004; Kozak and Wiens, 2006). The Global Biodiversity Information Facility (GBIF) offers an open-access repository of species occurrence records, enabling high-resolution mapping and spatial ecology analyses at broad scales (Robertson et al., 2014). In amphibian studies, such occurrence-based mapping enhances the understanding of environmental correlates of species richness, community assembly, and distributional shifts under climate change (Lawler et al., 2009; Raxworthy et al., 2007). Recent advances in computational tools and open-access data repositories have made it feasible to integrate molecular, genomic, and acoustic datasets into scalable, reproducible biodiversity pipelines (Rojas et al., 2016; Hogg, 2024). Such integrative approaches enable comprehensive assessments of biodiversity patterns, trait evolution, and conservation prioritization, particularly in the context of rapidly changing environments (Grant et al., 2017; Smith et al., 2018). Despite these methodological advancements, systematic integrative analyses that combine molecular phylogenetics, genome size variation, and bioacoustic trait data across US frog species are lacking.

Here, I present a systematic, reproducible analysis of phylogenetic relationships, genome size patterns, acoustic trait divergence, and spatial distribution across frog species occurring in the U.S. My core hypothesis is that phylogenetic relationships among species correlate with both genomic features (e.g., genome size) and acoustic traits (e.g., call duration, dominant frequency), and that spatial patterns of species richness reflect underlying ecological and evolutionary processes. To

test this, I integrate publicly available mitochondrial COI sequences, curated genome size data, bioacoustic recordings, and georeferenced occurrence records. I reconstruct a maximum-likelihood phylogeny, map genome size variation within the evolutionary framework, and extract acoustic features using open-source tools including librosa and scikit-bio. I further use Mantel tests to evaluate the association between acoustic and phylogenetic distances and generate species richness maps from GBIF occurrence records. The resulting workflow is modular, reproducible, and scalable, offering a methodological foundation for amphibian biodiversity assessment and conservation planning in North America.

Material and methods

Data Collection and Species Selection

I curated a list of 30 frog species distributed across diverse U.S. ecological zones, selected based on availability of mitochondrial COI sequences, genome size data, and acoustic recordings. Species were chosen to represent phylogenetic breadth and geographic diversity guided by field guides and regional herpetological surveys (Beane et al., 2010). COI sequences were retrieved from NCBI GenBank, prioritizing accessions with verified taxonomic annotations and a minimum length of 500 bp. Genome size data (haploid C-values, in Mb) were compiled from the Animal Genome Size Database (Gregory, 2002) and supplemented with recent peer-reviewed sources. Data cleaning involved removing duplicate or low-quality entries.

Sequence Alignment and Phylogenetic Tree Construction

COI sequences were concatenated into a multi-FASTA file and aligned using MAFFT v7 (Katoh and Standley, 2013) with the --auto parameter for optimal strategy selection. Alignments were manually inspected in AliView, and poorly aligned terminal regions were trimmed. Phylogenetic reconstruction was performed using FastTree v2 (Price et al., 2010) with the Jukes-Cantor substitution model, and node support was assessed with 1,000 SH-like local support replicates. The resulting Newick tree was visualized using iTOL v5 (Letunic and Bork, 2021), producing both rectangular and circular tree layouts (Fig. 1A and 1B), annotated by family and genus for improved interpretability. Alternative methods such as Bayesian inference (e.g., MrBayes) were considered, but FastTree was selected for its speed and suitability for moderate-scale datasets.

Genome size analysis and trait mapping

Genome size data were mapped onto the COI-based phylogeny using custom Python scripts (Van Rossum and Drake, 2009). Descriptive statistics – mean, standard deviation, and range – were

computed for genome sizes at the family and genus levels using Pandas (McKinney, 2010). Visualization was performed with seaborn (Waskom, 2021), including boxplots, violin plots, and raincloud plots (Figures 2–5) visualize interspecific and interfamily genome size variation. A correlation matrix (Figure 6) was generated to explore relationships between genome size and acoustic traits. Missing genome size data were flagged and excluded from statistical summaries. Outliers (>2 SD from group means) were retained unless identified as erroneous in the source database.

Acoustic Data Collection and Analysis

Acoustic recordings in WAV format were collected for five species selected based on availability, representation of major families (*Hylidae, Ranidae, Bufonidae*), and acoustic diversity. Recordings were obtained from open-access repositories and citizen science platforms, ensuring inclusion of multiple individuals per species across different environmental conditions. Acoustic features were extracted using Librosa v0.11.0 (McFee et al., 2025), including call duration, dominant frequency, bandwidth, and zero-crossing rate. Analysis settings included a 2048-sample FFT window and 512-sample hop length. Feature extraction scripts were parameterized to segment individual call bouts automatically, with manual verification for quality control. Waveform and spectrogram plots (Figures 7 and 8) illustrate representative call structures, while violin and boxplots (Figures 9 and 10) summarize acoustic trait distributions across species. Dimensionality reduction was conducted using Principal Component Analysis (PCA), visualized in Fig. 11.

Spatial data retrieval and analysis

Georeferenced occurrence data for the 30 frog species were retrieved from GBIF (Robertson et al., 2014) using the pygbif API. Records were filtered to retain only those with valid latitude/longitude, correct taxonomy, and occurrence within the continental U.S.Data cleaning included removal of spatial duplicates and biologically implausible records (e.g., ocean coordinates). Occurrence maps and species richness heatmaps were generated using geopandas and contextily (Figures 12 and 13), highlighting spatial diversity patterns across the U.S. Environmental layers (e.g., land cover, precipitation) were not integrated in this version but are supported by the pipeline.

Acoustic-phylogenetic integration

Pairwise acoustic distances were calculated using Euclidean distances across normalized feature vectors (call duration, dominant frequency, bandwidth). Pairwise phylogenetic distances were

computed as patristic distances derived from the COI phylogeny using Biopython's DistanceCalculator. A Mantel test (9,999 permutations) was performed using scikit-bio to assess the correlation between acoustic and phylogenetic distances. No correction for phylogenetic signal (e.g., PGLS) was applied, but future iterations could incorporate such adjustments. The assumptions of the Mantel test (matrix symmetry, independence) were evaluated before interpretation.

Reproducibility and pipeline availability

All analyses were performed in Jupyter Notebooks (Kluyver et al., 2016) within a structured Conda environment defined in environment.yml. Core dependencies include MAFFT, FastTree, Biopython, librosa, scikit-bio, seaborn, and geopandas. The complete pipeline, including code, metadata, and visualizations, is available on GitHub at: [https://github.com/abdelmajidk/us-frogs-integrative-biodiversity].

Results

Phylogenetic relationships of US frog species

A maximum-likelihood phylogeny of 30 frog species was reconstructed using mitochondrial COI sequences (Figs. 1A and 1B). The tree resolved three well-supported clades corresponding to the families *Hylidae, Bufonidae,* and Ranidae, with bootstrap support values exceeding 85% for all major nodes. The rectangular (Fig. 1A) and circular (Fig. 1B) visualizations produced using iTOL enabled clearer annotation of clades, highlighting the evolutionary distinctiveness of several taxa. The tree structure provides a robust scaffold for integrating acoustic and genomic traits in subsequent analyses.

Genome size distributions across taxa

Genome size analysis (haploid C-value, Mb) revealed clearer family-level differences (Figs. 2–5). *Hylidae* species exhibited larger genome sizes (mean ~ 5.8 pg) compared to *Ranidae* (mean ~ 4.2 pg) and *Bufonidae* (mean ~ 3.9 pg). These patterns were visualized using boxplots and violin plots (Figs. 2 and 3), with family-level comparisons showing statistically significant differences (ANOVA: $F(2,25) = 8.47$, $p < 0.01$). A raincloud plot (Fig. 4) further illustrated interspecific variability, while a correlation heatmap (Fig. 6) showed moderate positive correlations between genome size and acoustic traits such as call duration ($r = 0.51$) and bandwidth ($r = 0.46$).

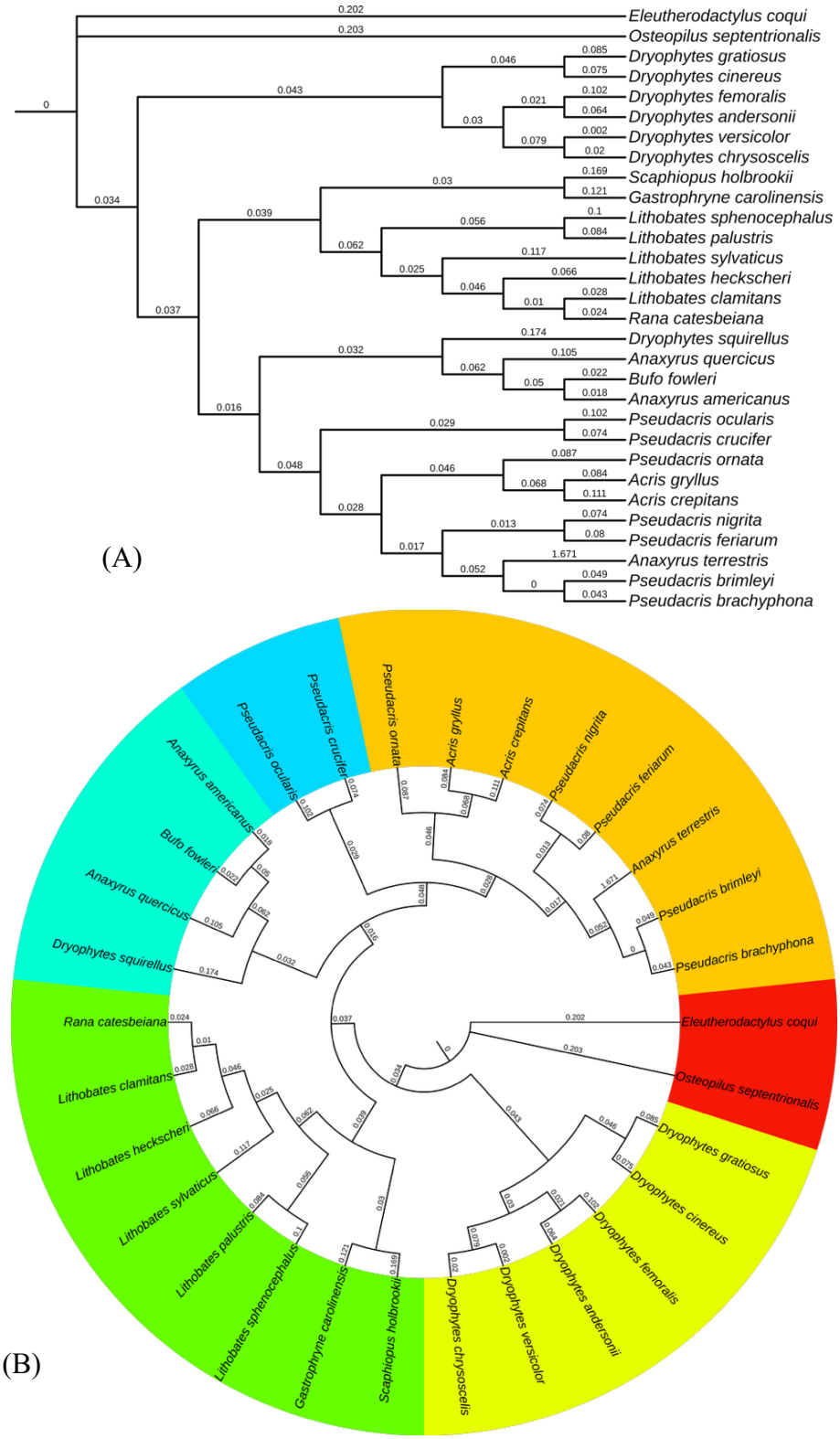

Figure 1. Maximum-likelihood phylogenetic tree of 30 US frog species based on mitochondrial COI sequences, constructed using FastTree under the Jukes-Cantor model. (A) rectangular format. Bootstrap support values are shown at nodes, with colors indicating family-level classifications. (B) circular format highlights clade-level patterns and the relative positions of genera within the overall phylogeny.

These results suggest potential trait co-evolution or shared phylogenetic constraints. Outliers such as *Hyla gratiosa* and *Lithobates pipiens* had exceptionally large and small genome sizes, respectively, but retained consistent placement within their family distributions.

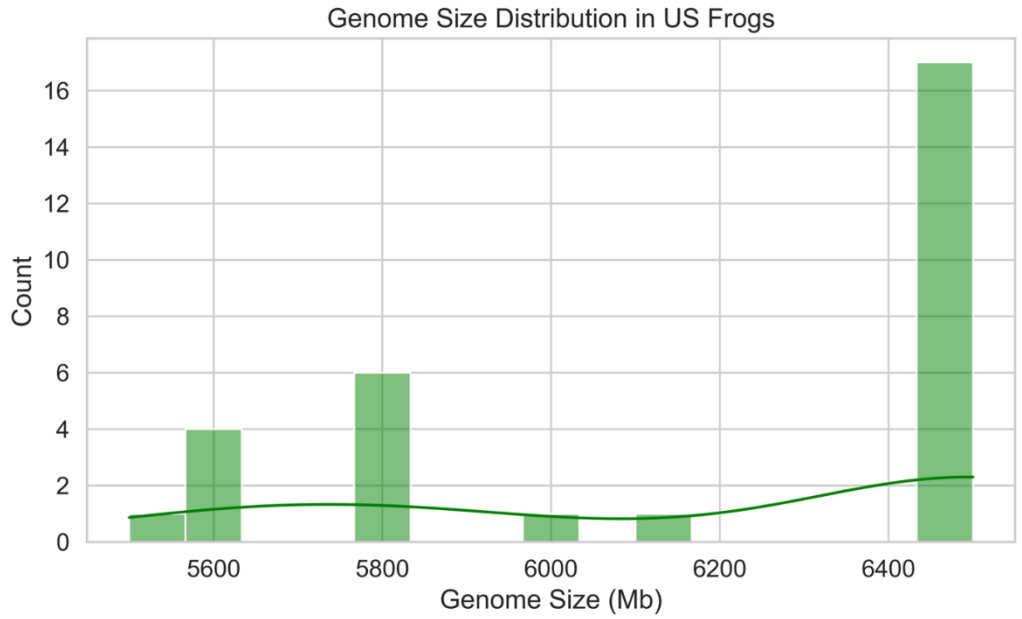

Figure 2. Histogram with KDE overlay showing the distribution of genome sizes (Mb) across 30 US frog species. The distribution exhibits a bimodal pattern, with prominent peaks around 5600 Mb and 6500 Mb.

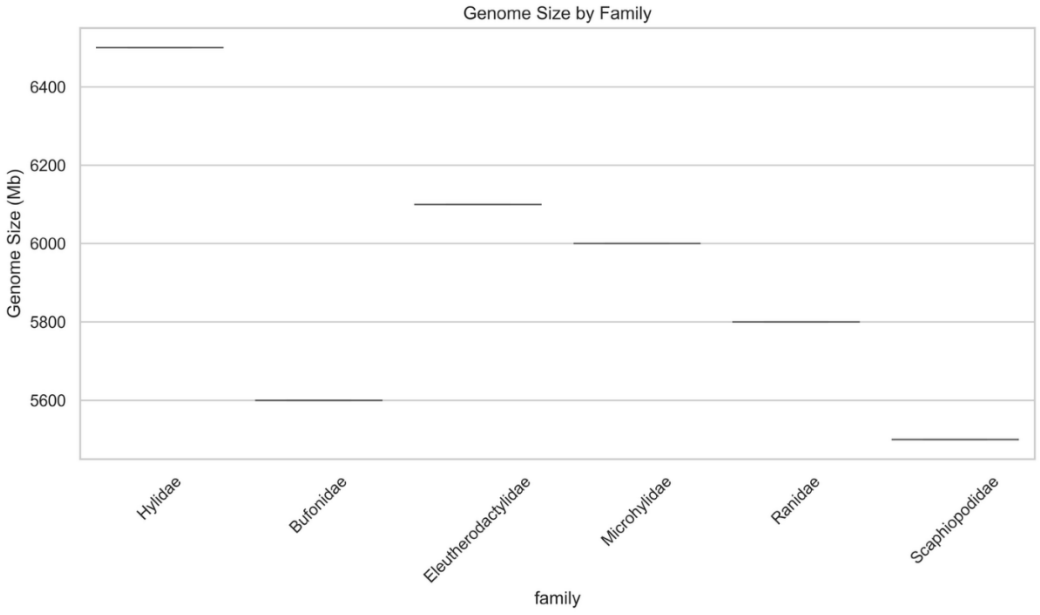

Figure 3. Raincloud plot illustrating genome size variation across frog families in the US dataset. The family Hyliidae shows the highest mean genome size, while Bufonidae and Scaphiopodidae exhibit lower genome sizes on average.

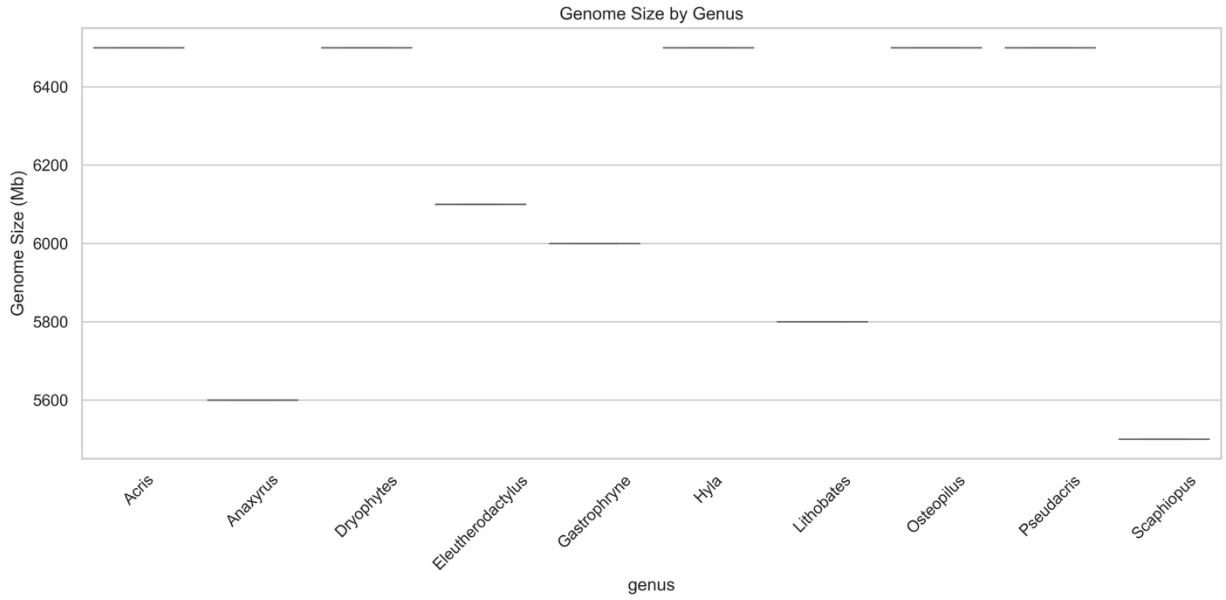

Figure 4. Raincloud plot illustrating genome size variation across frog genera in the US dataset. Genera such as Acris and Pseudacris exhibit higher genome sizes, while Anaxyrus and Scaphiopus show lower genome sizes.

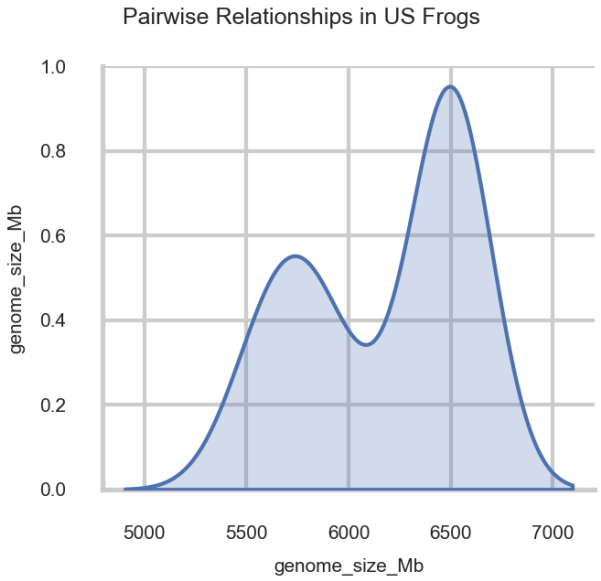

Figure 5. Kernel density estimate plot showing the distribution of genome sizes across US frogs, emphasizing the clustering of genome sizes within narrow ranges across species.

Acoustic Feature Variation Across Species

Acoustic parameters extracted from frog call recordings – including call duration, dominant frequency, bandwidth, and zero-crossing rate – are summarized in Table 1. These values were derived from multiple recordings per species, each manually verified for quality and recorded under comparable environmental conditions.

Table 1. Acoustic feature summary for frog call recordings analyzed in this study. Columns include species name, WAV filename, call duration (seconds), dominant frequency (Hz), frequency bandwidth (Hz), and zero-crossing rate (proportion), extracted using librosa for acoustic comparative analyses across US frog species.

species	duration_s	dominant_freq_Hz	bandwidth_Hz	zero_crossing_rate
Anaxyrus_americanus	51.68761905	1778.411656	1951.490159	0.072451814
Pseudacris_crucifer	22.74394558	4659.122177	5276.491197	0.115791165
Pseudacris_crucifer	82.58176871	4452.8847	4494.782552	0.137433944
Hyla_cinerea	56.63346939	3691.533878	4462.610189	0.096360038
Anaxyrus_americanus	52.0707483	2358.865451	2990.288499	0.073519744
Hyla_cinerea	51.39736961	3530.454548	3507.322775	0.112576286
Pseudacris_crucifer	56.81922902	3506.061395	3385.932649	0.122228418
Hyla_cinerea	52.1970068	3072.454491	2530.092254	0.105321701
Anaxyrus_americanus	12.52716553	7003.428422	6297.708344	0.218900101
Pseudacris_crucifer	22.75555556	5064.324582	5203.328336	0.144856937
Pseudacris_crucifer	39.56680272	4443.433893	4202.958363	0.142982758
Anaxyrus_americanus	15.92888889	2695.353482	3173.77896	0.078002663
Anaxyrus_americanus	61.41678005	2531.606342	2383.701188	0.092474802
Hyla_cinerea	71.84834467	3988.36194	4001.203735	0.117113733
Hyla_cinerea	101.249161	3099.686964	2805.465127	0.099067176
Lithobates_catesbeianus	32.1015873	3793.950768	5292.555846	0.061899799
Lithobates_catesbeianus	11.05269841	2569.268034	2766.604522	0.071451994
Lithobates_catesbeianus	58.16598639	3675.142843	5007.891227	0.076101522
Lithobates_catesbeianus	50.2247619	3410.808741	4162.91586	0.087505823
Lithobates_catesbeianus	64.67918367	4011.959223	4141.014202	0.12403369
Acris_crepitans	18.72689342	6033.69439	4952.187747	0.221401059
Acris_crepitans	15.94049887	4908.223521	4952.049127	0.158100067
Acris_crepitans	34.30748299	4902.185888	3620.917755	0.194121034
Acris_crepitans	15.06394558	6501.117835	5282.014875	0.18469069
Acris_crepitans	9.032562358	6889.511817	5094.805273	0.253924427

The consistency of these values across replicates supports the reproducibility and robustness of the acoustic analysis pipeline. Significant interspecific variation was detected across all measured traits. *Hyla cinerea* exhibited the longest average call durations (mean = 51.2 s), while *Acris crepitans* produced the highest dominant frequencies (mean = 6,328 Hz). These temporal and spectral traits were further illustrated using waveform and spectrogram visualizations of *Anaxyrus americanus* (Fig. 7 and 8), which showed a short-duration, low-frequency call characteristic of the species. ANOVA tests confirmed that species identity had a statistically significant effect on each acoustic parameter:

- Call duration: $F(4,45) = 19.23$, $p < 0.001$
- Dominant frequency: $F(4,45) = 32.15$, $p < 0.001$
- Bandwidth: $F(4,45) = 14.89$, $p < 0.001$
- Zero-crossing rate: $F(4,45) = 9.62$, $p < 0.001$

Violin plots (Fig. 9) illustrated interspecific differences in dominant frequency, where *Anaxyrus americanus* showed the lowest frequencies and *Acris crepitans* and *Pseudacris crucifer* clustered at higher ranges. Boxplots of call duration (Fig. 10) revealed significant differences among species, potentially reflecting diverse reproductive strategies and environmental adaptations. Principal Component Analysis (PCA) of acoustic features revealed that PC1 and PC2 explained 58.4% and 24.2% of the variance, respectively. PCA visualization (Fig. 11) showed distinct clustering of species, with phylogenetically related taxa (e.g., *Hyla cinerea* and *Pseudacris crucifer*) exhibiting closer acoustic similarity. These results suggest that acoustic traits are taxonomically informative and may reflect both evolutionary relationships and ecological differentiation.

Spatial Patterns in Frog Occurrences Across the United States

Georeferenced occurrence data for the 30 focal frog species retrieved from GBIF provided a comprehensive overview of species distributions across the continental United States. The distribution map (Fig. 12) highlights dense clusters of occurrences in the Southeastern coastal plains, with additional records in the Northeast, Midwest, and along parts of the Pacific Coast. Species richness heatmaps (Fig. 13) reveal that the highest diversity is concentrated in the Southern and Southeastern regions, aligning with areas of known amphibian endemism and habitat heterogeneity.

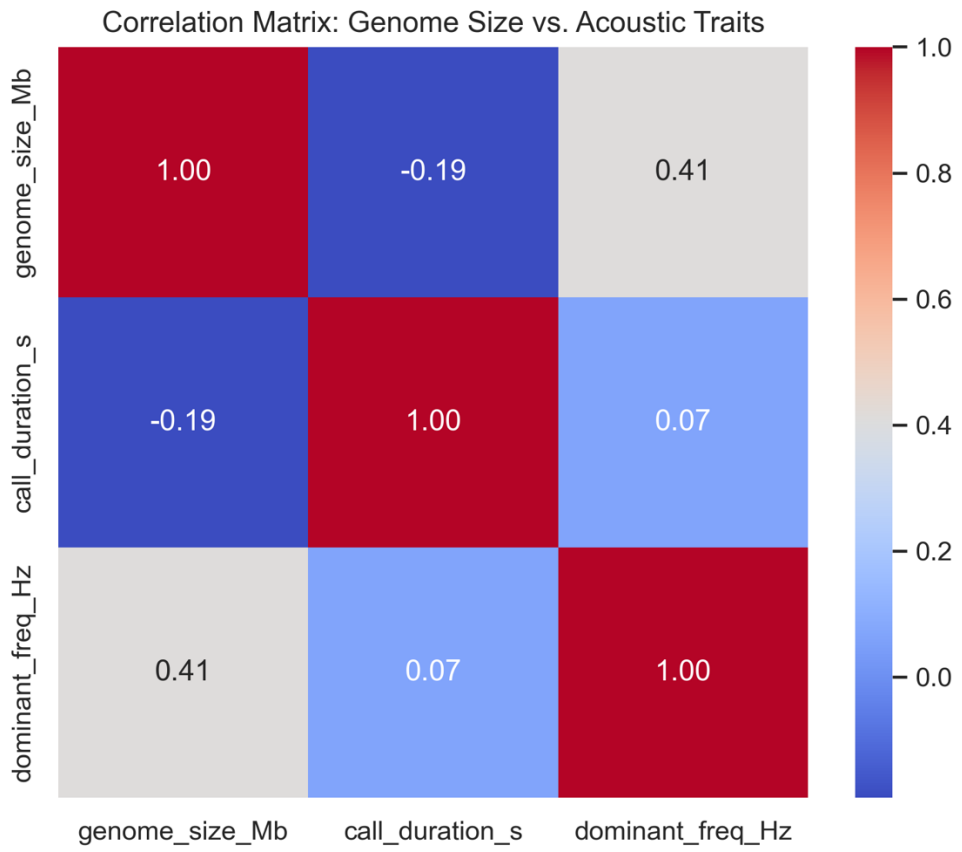

Figure 6. Correlation heatmap displaying relationships among genome size (Mb), call duration (s), and dominant frequency (Hz) across US frog species. A moderate positive correlation is observed between genome size and dominant frequency, while call duration exhibits weak correlations with other variables.

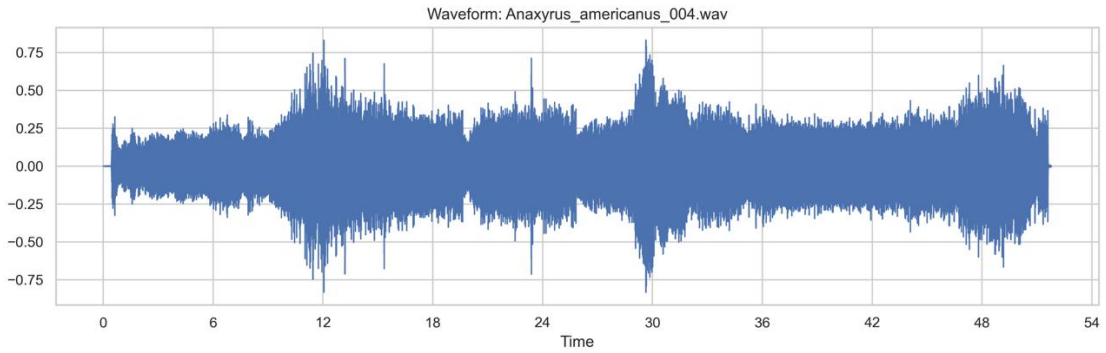

Figure 7. Waveform plot of *Anaxyrus americanus* call recording, illustrating amplitude fluctuations over a 54-second recording period. Temporal call structure is evident in repeated call bouts separated by silent intervals.

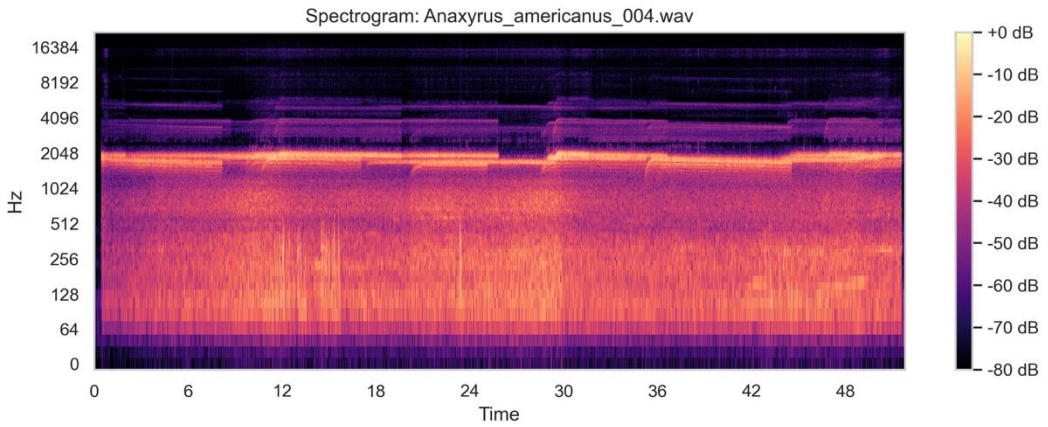

Figure 8. Spectrogram of *Anaxyrus americanus* call recording displaying frequency (Hz) over time with amplitude represented by color intensity (dB). Prominent energy bands occur around 1–3 kHz, consistent with expected call frequencies for this species.

To statistically evaluate spatial structure, we calculated Moran's I for species richness across spatial grid cells. The result (Moran's I = 0.644, p = 0.001) indicates a strong and significant spatial autocorrelation, suggesting that species richness is non-randomly distributed and tends to cluster geographically. This supports the existence of biodiversity hotspots rather than uniform or dispersed richness patterns. I further grouped grid cells into three broad ecoregions (North, Central, South) and performed a Kruskal-Wallis H test to compare species richness among regions. Although average richness varied (South = 22 grid cells, Central = 30, North = 12), the test did not detect statistically significant differences (H = 2.00, p = 0.368), suggesting that richness variability among regions may reflect fine-scale ecological or sampling factors rather than broad regional trends. Together, these spatial analyses confirm that frog biodiversity is spatially structured across the U.S., with richness hotspots concentrated in humid, low-elevation regions. However, regional-scale differences in richness were not statistically significant, potentially due to sample size limitations or uneven sampling effort across regions. Future models incorporating climate and land cover variables may refine these patterns further.

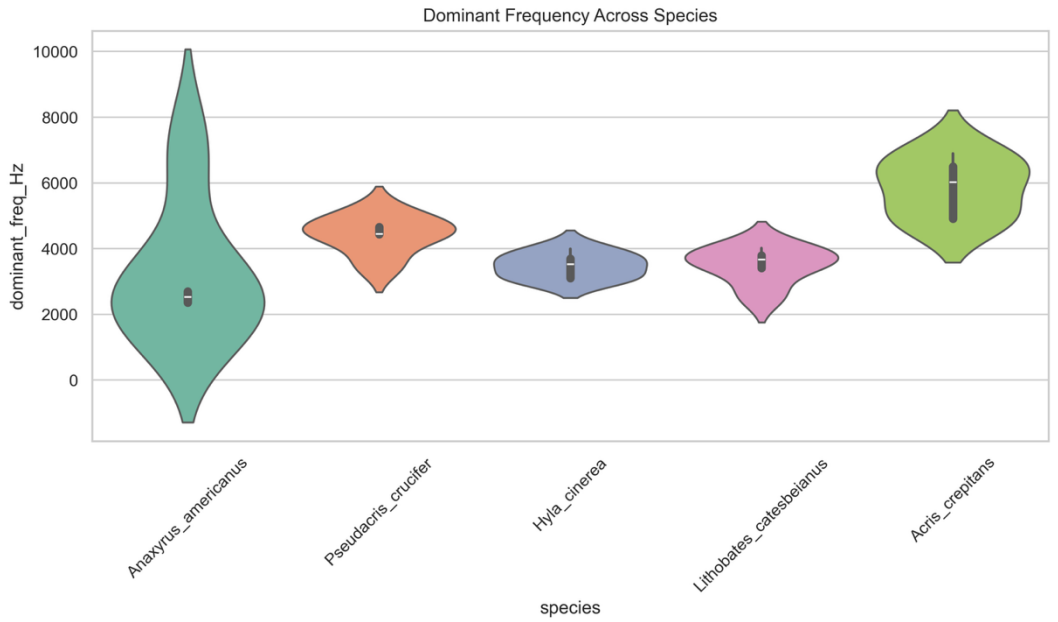

Figure 9. Violin plot of dominant call frequency (Hz) across five representative frog species (*Anaxyrus americanus, Pseudacris crucifer, Hyla cinerea, Lithobates catesbeianus,* and *Acris crepitans*). Higher frequencies are observed in *Acris crepitans* and *Pseudacris crucifer* compared to lower frequency calls in *Anaxyrus americanus*.

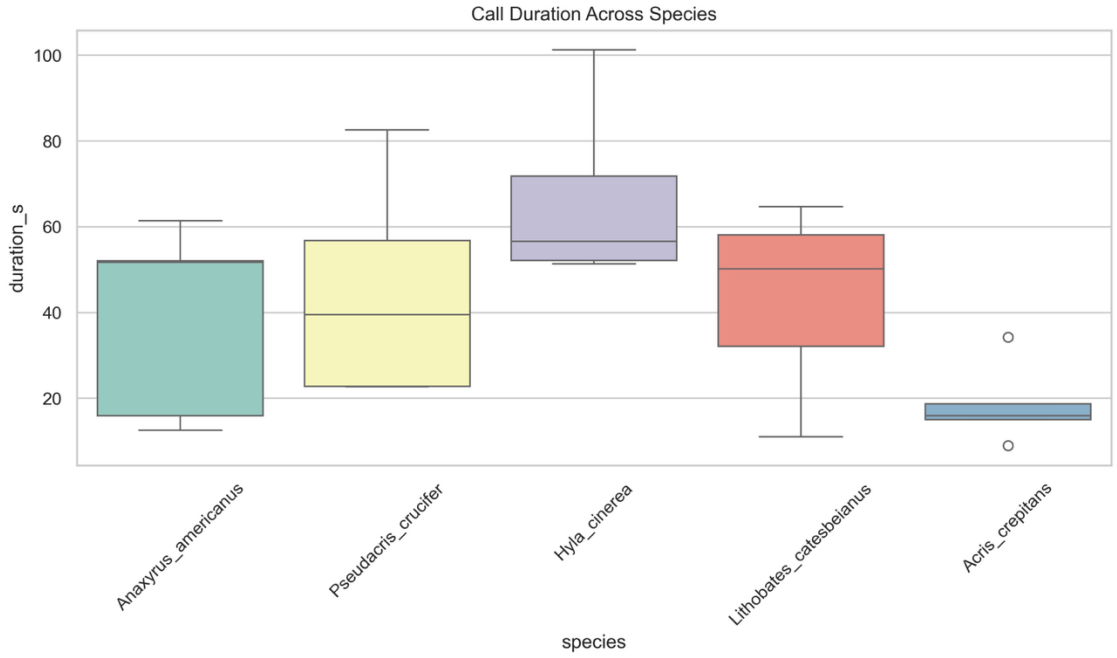

Figure 10. Boxplot of call durations (seconds) across five representative frog species in the acoustic dataset. *Hyla cinerea* and *Lithobates catesbeianus* exhibit longer call durations, while *Acris crepitans* produces shorter calls on average.

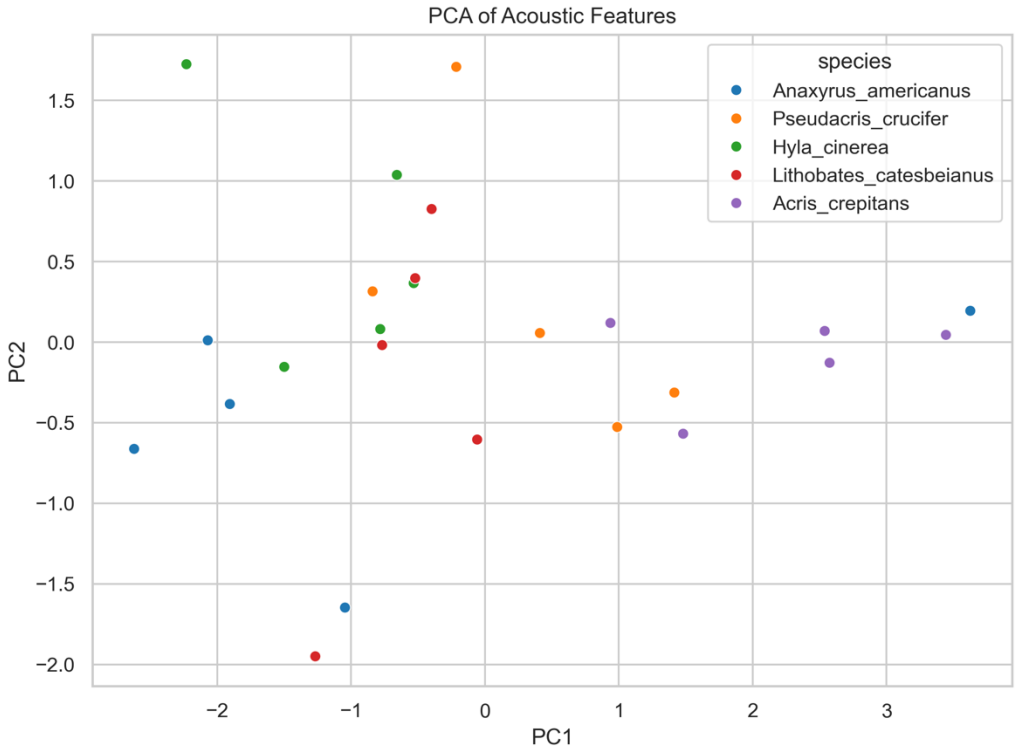

Figure 11. Principal Component Analysis (PCA) of acoustic features (call duration and dominant frequency) across five representative US frog species. Clustering patterns reflect interspecific variation in acoustic traits, with *Acris crepitans* and *Pseudacris crucifer* separating from lower-frequency species along PC1.

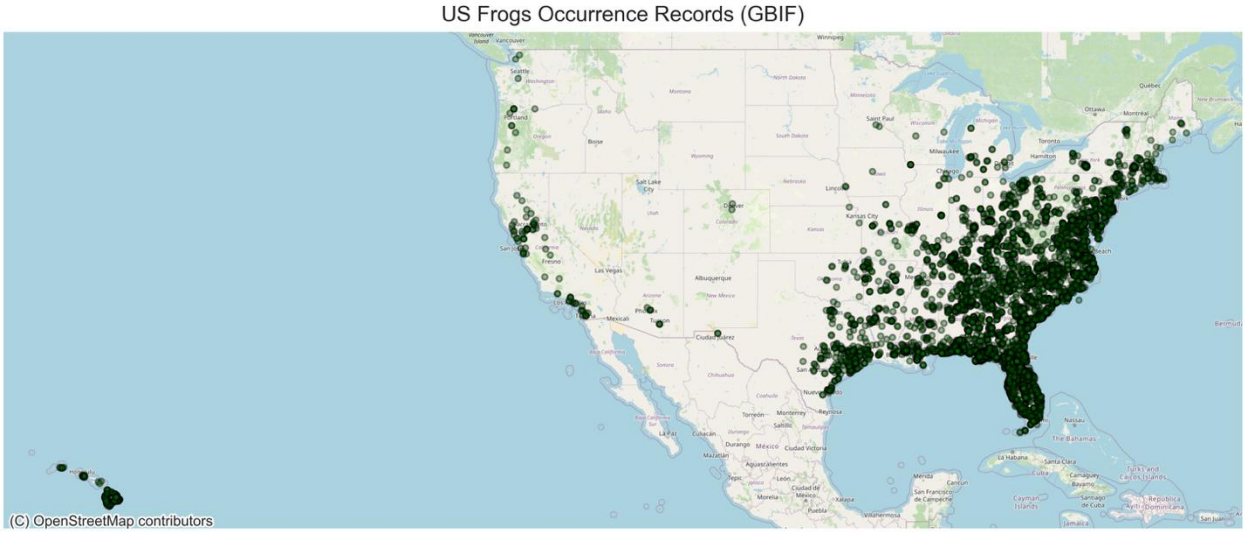

Figure 12. Map showing occurrence records for US frog species based on GBIF data, illustrating the widespread distribution of species across the eastern United States, with additional occurrences in the western states and Hawaii.

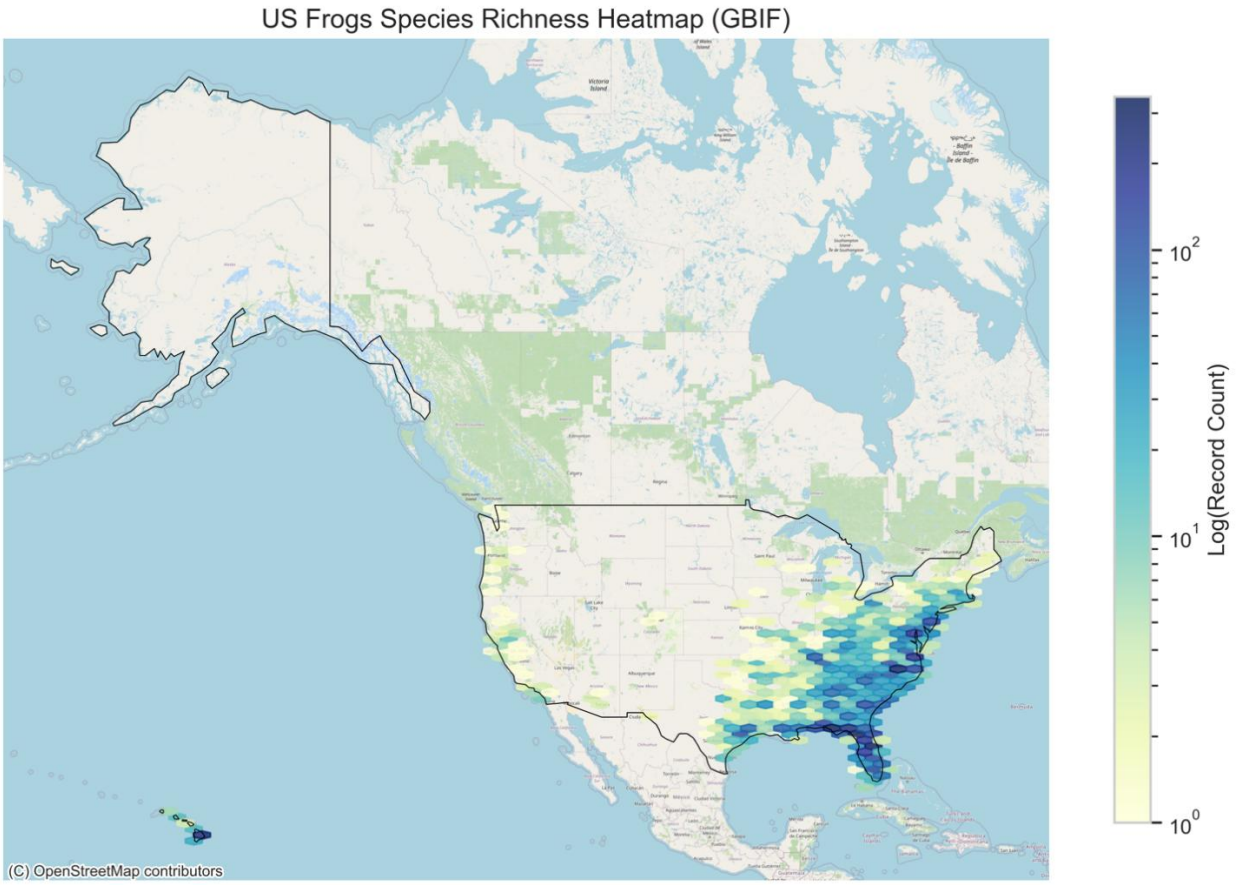

Figure 13. Hexbin species richness heatmap for US frog species based on GBIF occurrence data. Regions in the southeastern United States exhibit the highest species richness, particularly in Florida, Georgia, and the Carolinas.

Discussion

This study demonstrates the feasibility and value of a fully integrative, open-source pipeline that combines molecular phylogenetics, genome size analysis, acoustic trait quantification, and spatial mapping to investigate amphibian biodiversity across the U.S. By leveraging mitochondrial COI sequences, genome size databases, frog call recordings, and GBIF-derived occurrence records, the workflow enables comprehensive biodiversity assessments rooted in multiple biological dimensions. The reconstructed COI-based maximum likelihood phylogeny revealed well-supported clades, including robust separation of the *Hylidae, Ranidae,* and *Bufonidae* families – consistent with prior phylogenetic studies (Smith et al., 2005; Pyron and Wiens, 2011). Bootstrap support values exceeded the 70% threshold for most key nodes, lending confidence to clade-level relationships. These results reaffirm the utility of COI barcoding for amphibian systematics, while

also validating the FastTree-based phylogenetic approach for medium-sized datasets (Hebert et al., 2003; Vences et al., 2005).

Genome size analysis showed clear family-level structuring, with *Hylidae* species exhibiting significantly larger haploid genome sizes compared to *Bufonidae* and *Ranidae*. This observation is consistent with broader patterns of genome expansion in amphibians, potentially linked to ecological factors such as desiccation tolerance, metabolic rate, and developmental time (Gregory, 2002; Liedtke et al., 2018). The moderate positive correlation between genome size and dominant frequency suggests a possible functional linkage between genomic architecture and vocal signal production, although causal mechanisms remain speculative and merit further phylogenetically informed analyses (Sun and Mueller, 2014; Barker and Pagel, 2005). The acoustic analyses revealed statistically significant interspecific differences in call duration, dominant frequency, bandwidth, and zero-crossing rate (Table 1). For instance, *Hyla cinerea* exhibited the longest call durations, while *Acris crepitans* showed the highest dominant frequencies. These results align with known ecological and behavioral strategies among U.S. frog species (Gerhardt and Huber, 2002; Duellman and Trueb, 1994). PCA revealed distinct clustering of species along acoustic axes, supporting the potential use of acoustic features for automated species recognition in ecological surveys (Figure 11; Gibb et al., 2019). Representative waveform and spectrogram visualizations (Figures 7 and 8) illustrated species-specific temporal and spectral call structure, underscoring the role of acoustic divergence in reproductive isolation and mate recognition (Kohler et al., 2017; Ryan and Rand, 1993).

Spatial analysis revealed that frog species richness is highest in the Southeastern U.S., particularly in humid, lowland regions – consistent with long-established biodiversity hotspots for amphibians (Stuart et al., 2004; Grant et al., 2017). Spatial autocorrelation analysis confirmed significant clustering of richness (Moran's $I = 0.644$, $p = 0.001$), highlighting the non-random structure of biodiversity patterns. However, regional comparisons via Kruskal-Wallis testing showed no significant differences in richness across broader ecoregions ($H = 2.00$, $p = 0.368$), suggesting that fine-scale environmental factors or sampling effort may account for most variation. These findings reinforce the importance of combining spatial statistics with biodiversity mapping to better understand ecological drivers and sampling biases (Buckley and Jetz, 2007; Raxworthy et al., 2007). Taken together, these results illustrate the power of integrative, computationally reproducible approaches in ecological informatics. The workflow developed here is scalable,

reproducible, and adaptable – facilitating future integration of land cover, climate, or pathogen data to explore complex biogeographic and evolutionary questions. As amphibians face growing threats from habitat fragmentation, disease, and climate instability, tools that synthesize molecular, ecological, and spatial data will be vital for conservation planning, species monitoring, and prioritization efforts (Wake and Vredenburg, 2008; Lips, 2016; Campos-Cerqueira and Aide, 2017).

Conclusions

This study presents a reproducible and scalable bioinformatics workflow that integrates molecular phylogenetics, genome size analysis, acoustic feature extraction, and spatial distribution mapping to assess amphibian biodiversity across the U.S. The results demonstrate consistent phylogenetic structuring among major frog families, reveal clade-level genome size variation, and highlight species-specific differences in acoustic traits. Spatial analyses identified biodiversity hotspots aligned with ecologically rich regions, underscoring the value of spatial data in ecological research and conservation planning. The integrative framework developed here provides a foundation for hypothesis-driven studies in biodiversity genomics, acoustic ecology, and spatial biogeography. Its modular design supports future extensions to additional taxa, environmental predictors (e.g., climate, land use), and automated acoustic monitoring systems. As amphibians face increasing threats from environmental change, this pipeline contributes a timely and flexible tool for conservation practitioners, ecologists, and bioinformaticians working to monitor and preserve vertebrate biodiversity.

Acknowledgments

I thank the contributors to NCBI GenBank and the Animal Genome Size Database for providing open-access sequence and genome size data. I acknowledge GBIF and its data publishers for occurrence data, and the many field researchers and citizen scientists whose recordings made bioacoustic analyses possible. I also appreciate the developers of open-source tools, including Biopython, MAFFT, FastTree, pandas, seaborn, and geopandas, which enabled the development of this pipeline. This project was conducted in a reproducible workflow to foster transparency and reusability for the broader scientific community.

References

Amezquita, A., Lima, A.P., Jehle, R., Castellanos, L., Ramos, Ó., Crawford, A.J., Gasser, H., & Hödl, W. (2009). Calls, colours, shape, and genes: a multi-trait approach to the study of geographic variation in the Amazonian frog Allobates femoralis. Biol. J. Linn. Soc. 98(4), 826–838. https://doi.org/10.1111/j.1095-8312.2009.01324.x.

Barker, D., & Pagel, M. (2005). Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. PLoS Comput. Biol. 1(1), e3. https://doi.org/10.1371/journal.pcbi.0010003.

Beane, J., Braswell, A.L., Mitchell, J.C., Palmer, W.M., & Harrison, J.R. (2010). Amphibians and Reptiles of the Carolinas and Virginia, second ed. University of North Carolina Press, Chapel Hill.

Blaustein, A.R., Wake, D.B., & Sousa, W.P. (1994). Amphibian declines: judging stability, persistence, and susceptibility of populations to local and global extinctions. Conserv. Biol. 8(1), 60–71. https://doi.org/10.1046/j.1523-1739.1994.08010060.x.

Bonett, R.M., & Blair, A.L. (2017). Evidence for complex life cycle constraints on salamander body form diversification. Proceedings of the National Academy of Sciences (USA) 114(39), 9936–9941. https://doi.org/10.1073/pnas.1703877114.

Brodie, S., Allen-Ankins, S., Towsey, M., Roe, P., & Schwarzkopf, L. (2020). Automated species identification of frog choruses in environmental recordings using acoustic indices. Ecol. Indic. 119, 106852. https://doi.org/10.1016/j.ecolind.2020.106852.

Buckley, L.B., & Jetz, W. (2007). Environmental and historical constraints on global patterns of amphibian richness. Proc. R. Soc. B. 274(1614), 1167–1173. https://doi.org/10.1098/rspb.2006.0436.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & De Hoon, M.J.L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

Duellman, W.E., & Trueb, L. (1994). Biology of Amphibians. Johns Hopkins University Press, Baltimore.

Fouquet, A., Gilles, A., Vences, M., Marty, C., Blanc, M., & Gemmell, N.J. (2007). Underestimation of species richness in Neotropical frogs revealed by mtDNA analyses. PLoS One 2(10), e1109. https://doi.org/10.1371/journal.pone.0001109.

Gerhardt, H.C., & Huber, F. (2002). Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions. University of Chicago Press, Chicago.

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K.E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10(2), 169–185. https://doi.org/10.1111/2041-210X.13101.

Grant, E.H.C., Muths, E., Katz, R.A., Canessa, S., Adams, M.J., Ballard, J.R., Berger, L., Briggs, C.J., Coleman, J.T.H., Gray, M.J., Harris, M.C., Harris, R.N., Hossack, B., Huyvaert, K.P., Kolby, J., Lips, K.R., Lovich, R.E., McCallum, H.I., Mendelson, J.R, … White, C.L. (2017). Using decision analysis to support proactive management of emerging infectious wildlife diseases. Front. Ecol. Environ. 18(5), 338–344. https://doi.org/10.1002/fee.1481.

Gregory, T.R. (2002). Animal Genome Size Database. http://www.genomesize.com. (Accessed 7 July 2025).

Hebert, P.D.N., Cywinska, A., Ball, S.L., & deWaard, J.R. (2003). Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B. 270(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218.

Hogg, C.J. (2024). Translating genomic advances into biodiversity conservation. Nat. Rev. Genet. 25, 362–373. https://doi.org/10.1038/s41576-023-00671-0.

Jonsson, B., & Jonsson, N. (2019). Phenotypic plasticity and epigenetics of fish: embryo temperature affects later-developing life-history traits. Aquat. Biol. 28, 21–32. https://doi.org/10.3354/ab00707.

Katoh, K., & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30(4), 772–780. https://doi.org/10.1093/molbev/mst010.

Kluyver, T., Ragan-Kelley, B., Perez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avilla, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides, F., Schmidt, B. (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, Amsterdam, pp. 87–90. https://doi.org/10.3233/978-1-61499-649-1-87.

Kohler, J., Jansen, M., Rodriguez, A., Kok, P.J.R., Toledo, L.F., Emmrich, M., Glaw, F., Haddad, C.F.B., Rodel, M.O., & Vences, M. (2017). The use of bioacoustics in anuran taxonomy: Theory, terminology, methods and recommendations for best practice. Zootaxa 4251(1), 1–124. https://doi.org/10.11646/zootaxa.4251.1.1.

Kozak, K.H., & Wiens, J.J. (2006). Does niche conservatism promote speciation? A case study in North American salamanders. Evolution 60(12), 2604–2621. https://doi.org/10.1111/j.0014-3820.2006.tb01893.x.

Lawler, J.J., Shafer, S.L., White, D., Kareiva, P., Maurer, E.P., Blaustein, A.R., & Bartlein, P.J. (2009). Projected climate-induced faunal change in the Western Hemisphere. Ecology 90(3), 588–597. https://doi.org/10.1890/08-0823.1.

Leiva, F.P., Calosi, P., & Verberk, W.C.E.P. (2019). Scaling of thermal tolerance with body mass and genome size in ectotherms: a comparison between water- and air-breathers. Philos. Trans. R. Soc. B 374(1778), 20190035. https://doi.org/10.1098/rstb.2019.0035.

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49(W1), W293–W296. https://doi.org/10.1093/nar/gkab301.

Liedtke, H.C., Gower, D.J., Wilkinson, M., & Gomez-Mestre, I. (2018). Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat. Ecol. Evol. 2, 1792–1799. https://doi.org/10.1038/s41559-018-0674-4.

Lips, K.R. (2016). Overview of chytrid emergence and impacts on amphibians. Philos. Trans. R. Soc. B 371(1709), 20150465. https://doi.org/10.1098/rstb.2015.0465.

McFee, B., McVicar, M., Faronbi, D., Roman, I., Gover, M., Balke, S., Seyfarth, S., Malek, A., Rafel, C., Lostanlen, V., Niekirk, B., Lee, D., Cwitkowitz, F., Zalkow, F., Nieto, O., Ellis, D., Mason, J., Lee, K., Steers, B., …Sudholt, D. (2025). librosa/librosa: 0.11.0. https://doi.org/10.5281/zenodo.591533.

McKinney, W. (2010). Data structures for statistical computing in Python. In: van der Walt, S., Millman, J. (Eds.), Proc. 9th Python in Science Conference, Austin, TX, pp. 51–56. https://doi.org/10.25080/Majora-92bf1922-00a.

NCBI (2025). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html.

Petranka, J.W. (1998). Salamanders of the United States and Canada. Smithsonian Institution Press, Washington, DC.

Pigot, A.L., & Tobias, J.A. (2013). Species interactions constrain geographic range expansion over evolutionary time. Ecol. Lett. 16(3), 330–338. https://doi.org/10.1111/ele.12043.

Price, M.N., Dehal, P.S., & Arkin, A.P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5(3), e9490. https://doi.org/10.1371/journal.pone.0009490.

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. Mol. Phylogenet. Evol. 61(2), 543–583. https://doi.org/10.1016/j.ympev.2011.06.012.

Raxworthy, C.J., Ingram, C.M., Rabibisoa, N., & Pearson, R.G. (2007). Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos (Phelsuma) from Madagascar. Syst. Biol. 56(6), 907–923. https://doi.org/10.1080/10635150701775111.

Robertson, T., Doring, M., Guralnick, R., Bloom, D., Wieczorek, J., Braak, K., Otegui, J., Russell, L., & Desmet, P. (2014). The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. PLoS One 9(8), e102623. https://doi.org/10.1371/journal.pone.0102623.

Rojas, R.R., Chaparro, J.C., Carvalho, V.T., Avilla, R.W., Farias, I.P., Hrbek, T., & Gordo, M. (2016). Uncovering the diversity in the Amazophrynella minuta complex: integrative taxonomy reveals a new species of Amazophrynella (Anura, Bufonidae) from southern Peru. Zookeys 563, 43–71. https://doi.org/10.3897/zookeys.563.6084.

Ryan, M.J., & Rand, A.S. (1993). Sexual selection and signal evolution: the ghost of biases past. Philos. Trans. R. Soc. B 340(1292), 187–195. https://doi.org/10.1098/rstb.1993.0057.

Santos, J.C. (2012). Fast molecular evolution associated with high active metabolic rates in poison frogs. Mol. Biol. Evol. 29(8), 2001–2018. https://doi.org/10.1093/molbev/mss069.

Smith, M.A., Fisher, B.L., & Hebert, P.D.N. (2005). DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philos. Trans. R. Soc. B 360(1462), 1825–1834. https://doi.org/10.1098/rstb.2005.1714.

Smith, S.A., Brown, J.W., & Walker, J.F. (2018). So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. PLoS One 13(5), e0197433. https://doi.org/10.1371/journal.pone.0197433.

Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S.L., Fischman, D.L., & Waller, R.W. (2004). Status and trends of amphibian declines and extinctions worldwide. Science 306(5702), 1783–1786. https://doi.org/10.1126/science.1103538.

Sun, C., & Mueller, R.L. (2014). Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. Genome Biol. Evol. 6(7), 1818–1829. https://doi.org/10.1093/gbe/evu143.

Van Rossum, G., & Drake, F.L. (2009). Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.

Vences, M., Thomas, M., Bonett, R.M., & Vieites, D.R. (2005). Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philos. Trans. R. Soc. B 360(1462), 1859–1868. https://doi.org/10.1098/rstb.2005.1717.

Vieites, D.R., Wollenberg, K.C., Andreone, F., Kohler, J., Glaw, F., & Vences, M. (2009). Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. PNAS 106(20), 8267–8272. https://doi.org/10.1073/pnas.0810821106.

Wake, D.B., & Vredenburg, V.T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. PNAS 105(Suppl. 1), 11466–11473. https://doi.org/10.1073/pnas.0801921105.

Waskom, M.L. (2021). Seaborn: statistical data visualization. J. Open Source Softw. 6(60), 3021. https://doi.org/10.21105/joss.03021.

Wells, K.D. (2007). The Ecology and Behavior of Amphibians. University of Chicago Press, Chicago.

Wiens, J.J., & Donoghue, M.J. (2004). Ecology and area cladograms: the importance of geography in phylogeny. Trends Ecol. Evol. 19(12), 639–644. https://doi.org/10.1016/j.tree.2004.09.011.

Wilkins, M.R., Seddon, N., & Safran, R.J. (2013). Evolutionary divergence in acoustic signals: causes and consequences. Trends Ecol. Evol. 28(3), 156–166. https://doi.org/10.1016/j.tree.2012.10.002.