Comparative Morphological Analysis of a Leopard Frog Complex in Colorado

Honors Thesis

Presented in Partial Fulfillment of the Requirements for the
University Honors Program
Colorado State University

By

Finnegan E. Ward
Department of Fish, Wildlife, and Conservation Biology

Dr. Larissa Bailey, Department of Fish, Wildlife, and Conservation Biology
Elena Newmark, Department of Biology

Spring 2026

Abstract

Hybridization is a unique conservation issue that can lead to reduced fitness and loss of local adaptation and may influence management practices, especially for species of conservation concern. The ability to distinguish hybrids with morphological features is helpful for wildlife managers but often requires confirmation via genomic information. The Northern Leopard Frog (*Lithobates pipiens*; NLF) and Plains Leopard Frog (*Lithobates blairi*; PLF) are two native species of leopard frogs in Colorado which are known to hybridize in areas where their ranges overlap in the southeastern region of the state. NLFs are listed as a Tier 1 species of conservation concern in Colorado, with significant population declines throughout the state but especially on the Front Range. In contrast, PLF populations are thought to be stable, and the species is not of conservation concern (CPW, 2026). The goal of this study is to determine the morphological characteristics that distinguish between NLFs, PLFs, and putative hybrid individuals. The Funk Lab and CPW biologists collected tissue samples and photographs of leopard frogs throughout Colorado. I randomly selected 20 individuals of each species and 4 putative hybrids for this study. Characteristics from sampled individuals were categorically scored on a variety of traits that are hypothesized to distinguish between the species. I used two analyses to evaluate traits that best predicted species identity: logistic regression and Random Forest computer learning algorithm. Both logistic regression and Random Forest suggest the tympanum spot is the best predictor of the NLF species, with the dorsolateral fold also showing importance for identification. Additionally, the Random Forest shows strong predictive ability between species, but accuracy is negatively influenced with the inclusion of putative hybrids. Eventually, we will be able to use genetic data to determine true species identity and whether hybrids can be

identified morphologically. If they can, we will be able to establish the best combination of traits for field identification of species.

Introduction

Landscape changes and range expansion can increase sympatry between closely related species, leading to interbreeding and potential hybridization (Garroway et al. 2010). Identifying natural or anthropogenic factors influencing range expansion, such as assisted translocation or climate change, is important for determining the conservation management for species and hybrids (Allendorf et al. 2001). Hybridization can have positive effects for biodiversity, such as hybrid vigor, increased genetic diversity, and evolution of traits better adapted for specific environments (Arce-Valdés & Sánchez-Guillén 2022). It can also have negative effects, especially when one species is rare or of conservation concern. Hybridization and introgression can lead to genetic swamping, where hybrids replace the threatened species via decreased population growth rates and loss of adapted traits, leading to local extinction (Austin et al. 2011). Since hybridization can have a variety of effects on wildlife populations, it is important to know where hybridization happens and determine management specific for each case. However, it is often difficult to distinguish hybrids via morphology, as hybrids can resemble a single parent, show intermediate traits between the two parents, or express transgressive traits, where the morphology is outside the variation of either parent species (Stelkens & Seehausen 2009). In many cases, genetics are needed to identify hybridization in wildlife populations.

The Northern (*Lithobates pipiens*, NLF) and Plains (*Lithobates blairi*, PLF) leopard frogs are two native species in Colorado. NLFs are relatively widespread across North America, ranging throughout the northeastern United States, Great Lakes Region, northern Great Plains, and Intermountain West (Hammerson et al. 1999). O'Donnell & Mock (2012) found that the Mississippi River acts as a genetic boundary and populations west of the Mississippi have faced significant declines. In Colorado, NLFs are listed as a Tier 1 Species of conservation concern due to wetland destruction from urban development, chytrid fungus (*Batrachochytrium dendrobatidis*), bullfrog introduction, and drought (CPW 2026). NLFs are found everywhere in Colorado except the southeastern corner of the state. PLFs are native to the southern Great Plains region of the United States, and their range extends into the southeastern corner of Colorado (Hammerson 1999) and other areas where they have been accidentally introduced. PLFs are generally adapted to drier (Blais et al. 2023) and warmer (Post 1972) habitats than the NLF, and their populations are thought to be stable in Colorado (CPW 2026).

Hybridization between NLFs and PLFs has been previously documented in watersheds in southeastern Colorado, where their ranges overlap (Gillis 1975). Additionally, Gillis (1975) documented hybridization between the two species in the lab, where hybrids showed intermediate morphological characteristics with greater similarity to PLF traits. Post (1972) also bred hybrids in the lab and found they had significantly lower survival rates. Post (1972) examined the breeding characteristics of populations found in these overlap zones, documenting differences in breeding phenology, vocalizations, and habitat. NLF generally bred in permanent ponds, and PLF in ephemeral ponds. However, one overlap zone exhibited a lack of difference in these prezygotic reproductive barriers, which is where they hybridized. There is a possibility that increases in drought could lead to range expansion and higher rates of hybridization between the NLF and PLF.

Given that NLF are a species of greatest conservation need and the species-specific risks of hybridization with PLF in Colorado are not fully known, it is important for wildlife managers to know where, if at all, wild populations are hybridizing. If morphology alone can be used to identify hybridization between the two species, this would enable more efficient use of conservation resources because genetic testing would not be necessary, saving limited funds and time. In this thesis, I aim to address the questions of whether NLFs, PLFs, and putative hybrids (HYB) are distinguishable by morphology alone, and what combination of morphological traits best predict species identification.

Methods

Sampling

A total of 705 individuals were captured from 63 sites during breeding and rearing season, April-October 2024 and 2025. Sites were selected throughout the state based on recent documentation of leopard frog presence. Tissue samples, i.e. toe or tadpole tail clips, were collected from captured individuals, along with measurements of snout-vent length and total body length. Photos of each frog's dorsum, legs, and sides were taken in the field with a color standard for morphology and identification. Individuals were identified to the species level (NLF, PLF, or putative hybrid) in the field based on capture location and traits from the literature. After processing, individuals were released at their captured location. All sampling was conducted under an approved Institutional Animal Care and Use Committee protocol (IACUC #6773) and with an approved Scientific Collections permit from CPW.

Trait Selection

A literature review was conducted to determine what morphological traits have been used to distinguish between the two species. Hybrids between the two focal species are hypothesized to have intermediate morphological traits (Gillis 1975). Eleven morphological traits were found in the literature to distinguish between the two species, and seven were selected to examine in this preliminary test. We categorized phenotypes within these remaining traits based on previous literature and field observations. We then tested consensus by asking multiple expert observers (N = 6) to score these traits for a subset of 20 frogs, and to provide an identification of species for each frog based on these traits (Table 1). Observer responses were compared using a Fleiss' Kappa test to assess agreement with the irr package in R (Gamer et al. 2012). Traits with poor agreement (Fleiss' Kappa values < 0.4) were omitted from future analyses; specifically, snout bluntness.

Data Collection and Analyses

20 adult individuals of each species (NLF, PLF) were randomly selected for morphological analysis, which were different than the individuals used for the trait selection tests above. Four putative hybrid adults were also included. Adults were individuals with a snout-vent length ≥ 51mm (Blais et al. 2013). All 44 individuals were scored with the six morphological traits that had high observer agreement (Figure 1). Each trait category included three possibilities: the expected trait characteristic for NLF or PLF and an "ambiguous" category (Figure 2).

I converted all species field identification to a binary response variable, where '1' indicated a NLF and '0' was non-NLF. I fit multiple logistic regression models with univariate relationships to the six explanatory variables (traits) and used corrected Akaike's Information Criterion (AICc) to determine which trait was the best predictor of NLF identification. Models were fit and evaluated with the AICcmodavg package in R (Mazerolle 2017). The same models were fit to datasets without putative HYB individuals included, with ‘1’ indicating NLF and ‘0’ for PLF (Figure 3).

I also used a Random Forest computer learning algorithm, using the R packages randomForest (Liaw & Wiener 2002) and caret (Kuhn 2008) to create the algorithm and calculate metrics. Similar to the logistic regression analysis above, the Random Forest algorithm was first run with data from just the two species, and the analysis was repeated with the hybrids included. Data was randomly subset with 70% of individuals grouped for the training data (to train the algorithm) and the remaining 30% of individuals were used for testing the trained algorithm. 125 trees were used based on the results of Oshiro et al. (2012), and nodes contained three splits. Finally, an NMDS plot was created using the vegan and ggplot2 packages (Oksanen et al. 2013, Wickham 2011) with a distance matrix calculated by the Random Forest.

Results

My literature review revealed 11 potential traits used to differentiate between the NLF and PLF: dorsolateral fold insetness and brokenness, tympanum spot, nose spot, inner thigh pattern, upper lip stripe, skin smoothness, color, dorsal spot border color, spot shape, and snout bluntness. The dorsolateral fold was split into two traits, the brokenness and the insetness of the fold at the posterior end of the body (Figures 1 & 2). Traits including color were removed, specifically the color of the frog and the dorsal spot border color, because the color standards proved to be insufficient to evaluate these color characteristics. Skin texture and spot shape were also removed because of difficulty categorizing the traits.

Consistency among the remaining traits were evaluated using multiple expert observers and the Fleiss’ Kappa test. Snout bluntness had a low Kappa value (0.08: Table 1), suggesting poor agreement for this trait among observers. I removed this trait from subsequent analyses. The remaining traits all had Kappa values > 0.4 (Table 1) and were used in subsequent logistic regression and Random Forest analyses. Additionally, traits with many "ambiguous" individuals generally yielded lower kappa values than traits where most individuals displayed classic phenotypes (Table 1 & Figure 2).

Univariate logistic regression models fit to the dataset without hybrids suggested the tympanum spot was the best predictor of NLF identification (AIC weight = 0.90; Figure 3). When hybrid individuals were added to the data set, the tympanum spot was still the best fitting model, and AIC weight increased to 0.95 for distinguishing NLFs from PLF or HYB individuals (Figure 3).

Fitting the random forest algorithm to the dataset without hybrids had an estimated out-of-bag error rate of 0% and tested accuracy of 100%. Mean decrease accuracy is a measure of how many individuals the model would incorrectly identify if a trait was removed. The tympanum spot was the most significant predictor of species, with a scaled mean decrease value of 10.73 (Figure 4). The upper lip stripe, thigh pattern, and nose spot all had scaled values <2. Mean decrease gini is a measure of node purity, or the predictive power of a specific trait, and showed similar results as the mean decrease accuracy metric (Figure 4). When the four putative hybrids were added to the dataset, the random forest analysis yielded an estimated out-of-bag error rate of 3.23% and a tested accuracy of 76.9%. The confusion matrix (Table 2) showed that when hybrid individuals were used to train the model, the model identified two individuals as NLF, where field identification would suggest they were PLF. Similarly, the model identified the single field-called hybrid individual in the test data set as a PLF. The tympanum spot was still the best predictor, with a scaled mean decrease accuracy value of 10.78. The same three traits (nose spot, thigh pattern, and upper lip stripe) had values close to zero, meaning they have very low influence on frog identification in the model. The dorsolateral fold insetness and brokenness also had scaled mean decrease values closer in value to the tympanum of 8.89 and 7.72, respectively.

The proximity matrix calculated by the Random Forest can be used to create a distance matrix, providing an illustration of the ‘distance’ between individuals with the observed collections of trait characteristics through an NMDS plot (Figure 5). It had a stress value of 0.064, suggesting the plot has a good fit to the multidimensional data. The NMDS plot illustrates NLF clustering on the bottom left side of the plot, and the PLF clustering on the other side at the bottom. HYB individuals were closer to the top of the graph but have a small sample size. Based on the results of the plot, it seems there is morphological distinction between the three groups.

Discussion

Results from the Random Forest analysis suggest that NLFs can be distinguished from PLF and putative hybrids via morphology. The NMDS plot (Figure 5) shows that field identified NLF and PLF individuals cluster on opposite ends of the plot, indicating they are morphologically distinguishable. HYB individuals may exhibit a mixture of (or ambiguous) characteristics, but one individual clustered with PLF (Figure 5). A larger sample size with paired genetic data is needed to determine if putative HYB individuals are true hybrids of NLF and PLF species. The random forest confusion matrices reinforced these results (Table 2). The accuracy of the species assignment via the random forest algorithm was 100%, when hybrids were not included, which demonstrates the ability to morphologically differentiate between the two leopard frog species.

When the HYB individuals were included, however, the Random Forest algorithm had more difficulty correctly identifying individuals. Some of the individuals incorrectly identified had traits leaning towards the other species in combination with other ambiguous traits, which is likely an issue with the categorization of individuals and photo quality. Many of the traits have a fair amount of natural intraspecific variation, which is difficult to capture in three categories. Evaluating combined characteristics in the field may lead to different species assignments, as certain traits could be interpreted differently in a photo without the benefit of observing the trait from multiple angles (Akhtar et al. 2024). A combination of multiple ambiguous traits could lead to discrepancies between field identification and model-based predictions using photographs alone.

The Random Forest analysis found that the tympanum spot, dorsolateral fold brokenness and insetness had the best ability to identify individuals as both species and hybrids. This is consistent with the literature for identifying the two species (Powell et al. 2019, Gillis 1975, Fischer et al. 1999, Conant & Collins 1991). Similarly, the logistic regression analyses suggested the tympanum was the best predictor of NLF identification, followed by the dorsolateral fold brokenness, using univariate models (Figure 3). While tympanum spot comes out as the best predictor for both analyses, this may be due to the ability to easily categorize the trait. The tympanum spot is generally less ambiguous than other tested traits, with a dot usually being distinct and present or absent. However, the dorsolateral fold brokenness has more variation and is more often categorized as ambiguous. It can be partially broken or difficult to determine if it is broken, and it can also be broken on just one side. In this data, 4 individuals had an ambiguous tympanum spot and 8 had ambiguous dorsolateral fold brokenness. The dorsolateral fold insetness is also difficult to distinguish in photos, as different angles are important for determining the extent of insetness. This could lead to errors when categorizing the trait from photos and would be less of an issue in the field.

This study could be improved in a couple of ways. The underlying assumption is that the field identification is correct, but it needs to be confirmed with genetic information. This is especially true for our PLF and putative HYB individuals, but all individuals will be genetically analyzed after this thesis. PLF individuals were mostly collected at one location where they were also introduced into NLF range, but it is currently unknown whether there were NLF at the particular location, previously. A couple of putative HYB individuals were also collected at this site. In the future, it would be good to include more PLF individuals from pure PLF range. This would also increase the number of parameters able to fit the dataset in the logistic regression, as Stoltzfus (2011) recommends having at least 10 individuals per independent variable for a fit logistic regression. From our dataset, a maximum of two variables could be used for the model, but identifying the species may use multiple traits for a holistic view of the individual before making a call in the field.

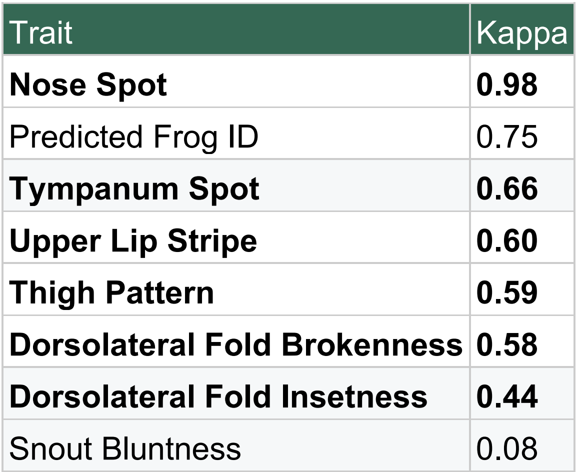
The ability to identify a threatened species from potential hybrid individuals in the field provides many opportunities to implement effective conservation strategies efficiently. Colorado Parks and Wildlife is interested in this information to improve NLF management for a variety of reasons. Managers want to understand the range of NLFs, the location and extent of the hybrid zone, and occupancy trends for each species. Additionally, accurate species identification can allow for studies of survival probabilities for hybrid individuals and facilitate their removal if hybridization results in poor survival outcomes for NLFs (e.g., genetic swamping) (Buggs 2007, Hending 2025). Accurate species identification could also prevent undesired translocations of HYB individuals into a genetically pure species population (Scriber 2013). Without the ability to distinguish species from hybrids via morphology, genetics are required to identify individuals, which would take a significant amount of time and funds before conservation implementation, potentially leading to worse outcomes for NLFs. These preliminary results suggest that identifying NLF from other species or hybrids is feasible without genetics using the tympanum spot, dorsolateral fold brokenness and insetness, and future work with genetic information will improve model accuracy. Once these analyses are finished, these results should be able to inform Colorado Parks and Wildlife for their management to improve NLF populations across the state.

Acknowledgements

I’d like to thank my advisors Dr. Larissa Bailey and Elena Newmark for giving me this opportunity and all of the help on this project throughout the last few semesters. In addition, I thank the Funk Lab for being such a welcoming and constructive space. For help in sampling efforts, we thank Colorado Parks and Wildlife and everyone who collected samples throughout the last couple field seasons.

Figures

Table 1. Fleiss' Kappa results for remaining traits after preliminary removal from a literature review. Kappa values depend on the consensus between observers in categorizing certain traits. Kappa values are seen as excellent agreement for values >0.8, good for 0.6-0.8, and moderate for 0.4-0.6. Bolded traits were kept for the final analysis. Predicted frog ID is what was tested in the analyses as the dependent variable but additionally tested for consensus between observers in the Fleiss' Kappa to examine how they use traits in combination to make an identification call.



Trait	Kappa
Nose Spot	**0.98**
Predicted Frog ID	0.75
Tympanum Spot	**0.66**
Upper Lip Stripe	**0.60**
Thigh Pattern	**0.59**
Dorsolateral Fold Brokenness	**0.58**
Dorsolateral Fold Insetness	**0.44**
Snout Bluntness	0.08

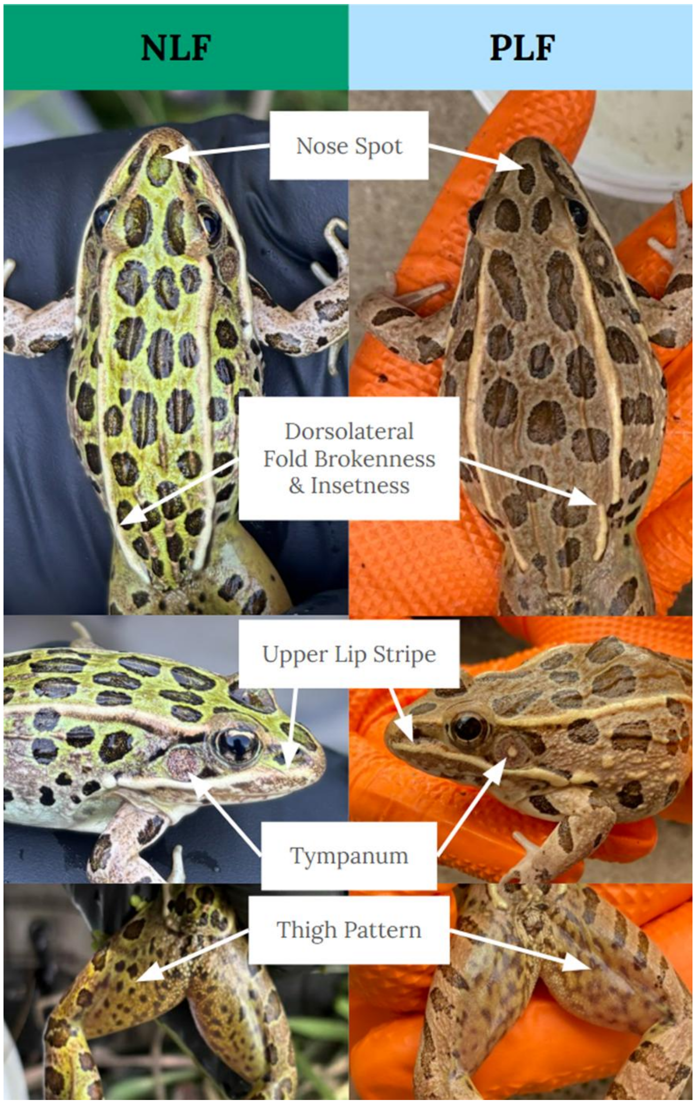


Figure 1. Final traits selected for analyses compared on NLF and PLF individuals. These include nose spot presence, dorsolateral fold brokenness, dorsolateral fold insetness, the prominence of the upper lip stripe in front of the eye, tympanum spot presence, and the thigh pattern.

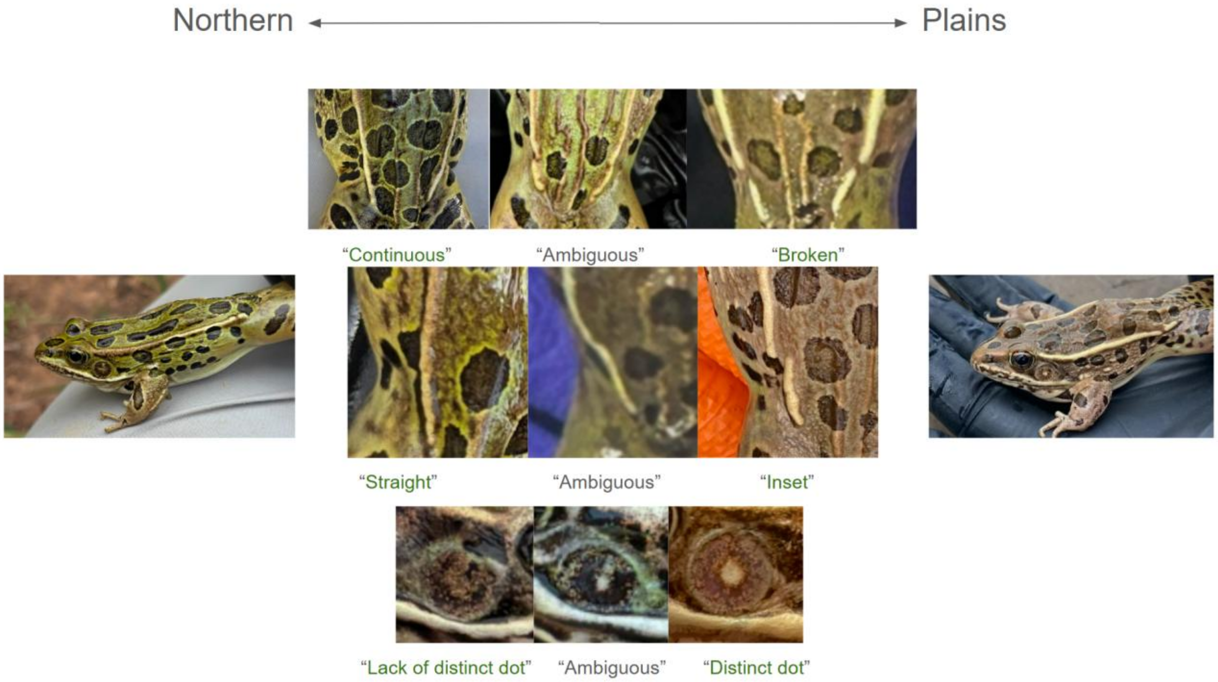


Figure 2. Example of the spectrum of traits that can be observed in the dorsolateral fold brokenness, insetness, and tympanum spot, with their corresponding category names below. NLF traits are on the left side and PLF traits are on the right, with ambiguous traits in the middle.

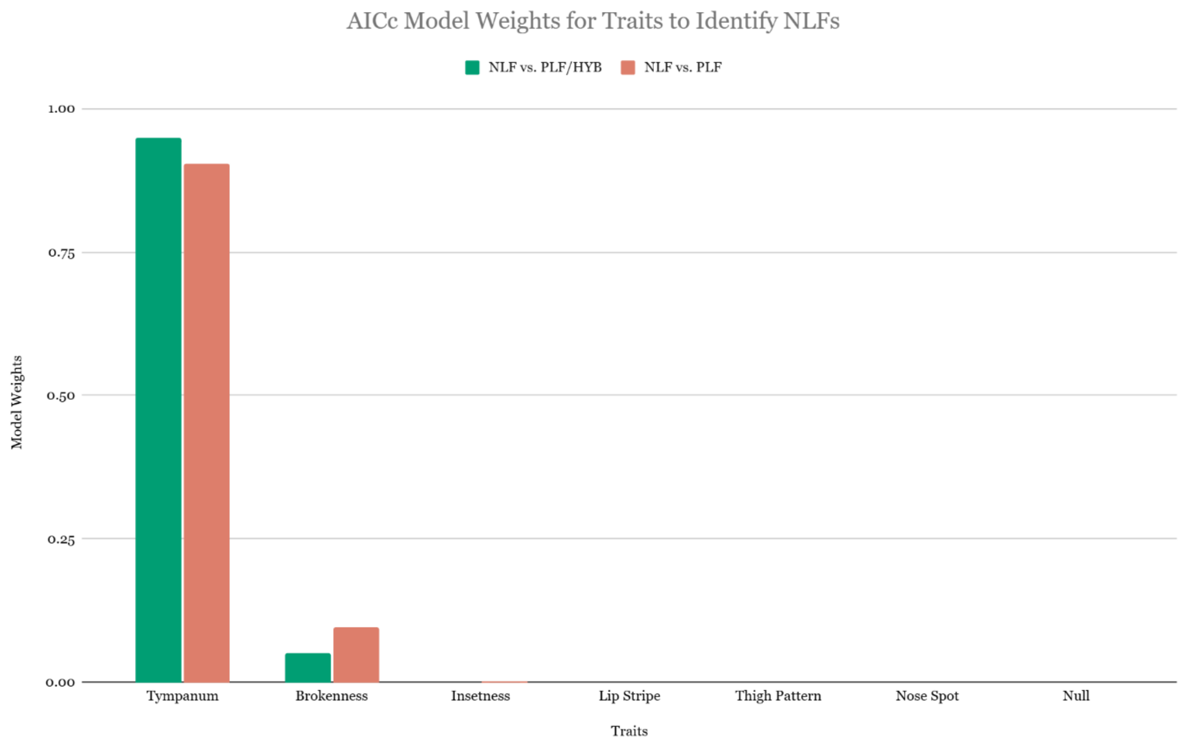


Figure 3. Model weights for traits distinguishing NLFs. Tympanum had the largest AIC model weight of 0.95 for hybrids included and 0.91 for just species. Dorsolateral fold brokenness had the second largest weight for both, being the remaining weight in the model selection with hybrids included. In the model selection with just species, dorsolateral fold insetness had a small model weight of 0.001.

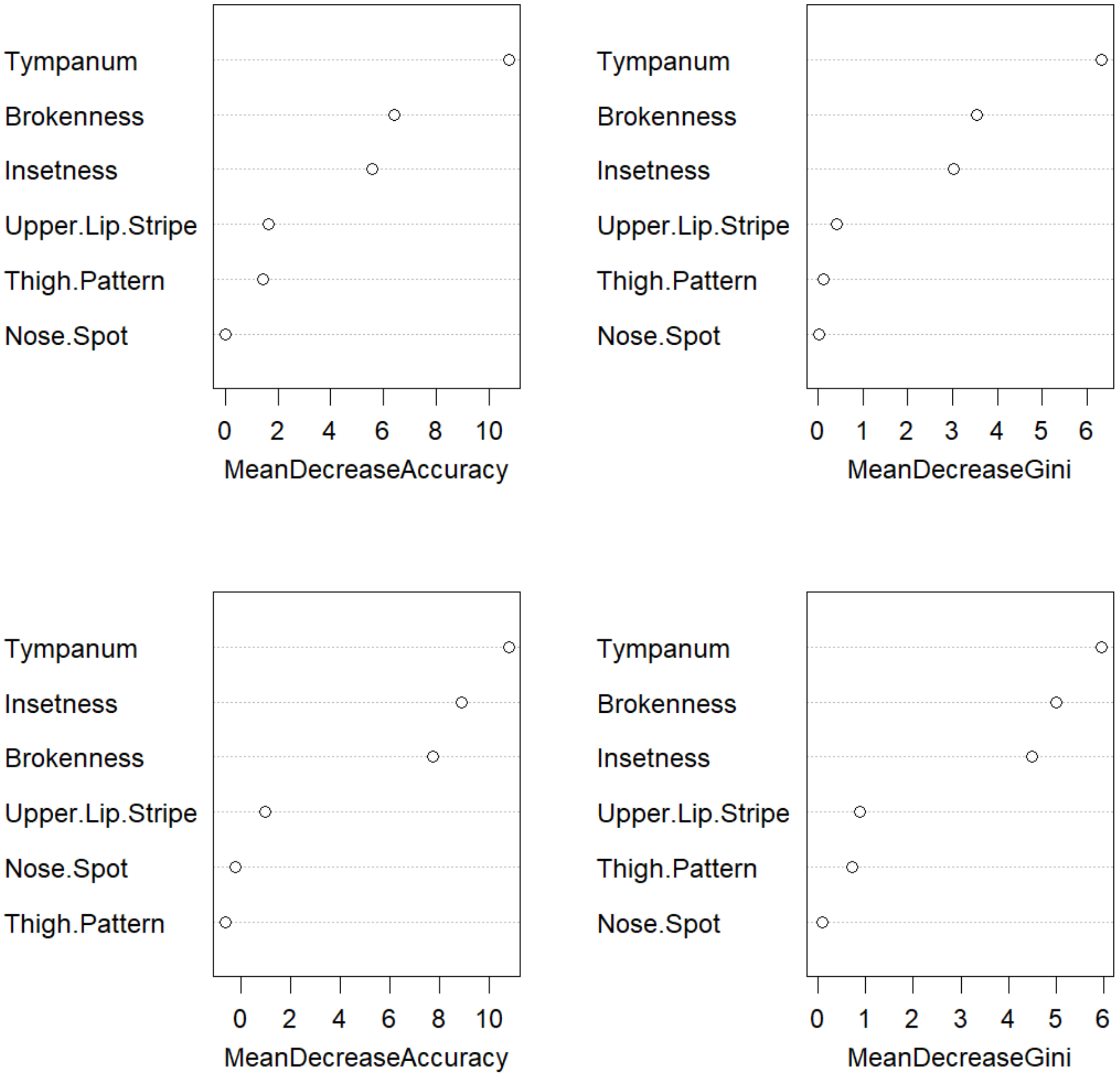


Figure 4. Mean decrease accuracy and mean decrease gini values for the random forests of just species (top) and species with hybrids (bottom). Tympanum spot was the best predictor for both models and metrics. Dorsolateral fold brokenness and insetness followed with higher predicting abilities than the rest of the traits tested.

Table 2. Confusion matrix for the test dataset in the random forest including HYB. Ideally, non-zero values would be on a diagonal line from the upper left to the lower right. The rows of the matrix represent the model prediction for the individual's identification, and the columns are the identifying call we used for analyses, given in the field.

Random Forest ID	Field ID		
	HYB	NLF	PLF
HYB	0	0	0
NLF	0	6	2
PLF	1	0	4

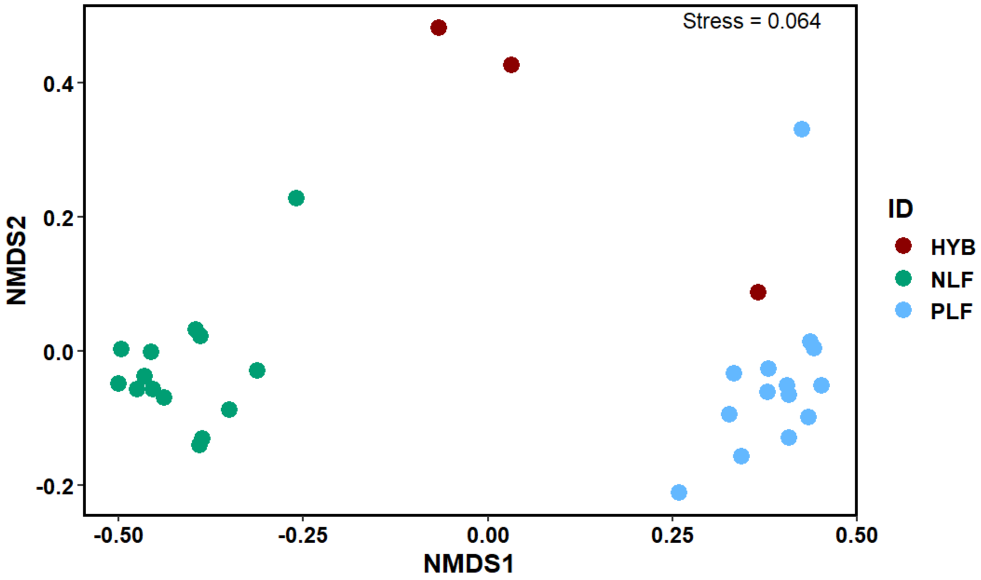


Figure 5. NMDS plot of individuals from the training data in the random forest. NLFs seem to cluster on the left side of the plot, and PLF clusters more on the right side, implying that NLF separates morphologically from PLF. The HYB may potentially cluster on their own at the top of the plot, but a larger sample size is needed to see similarities between NLF and PLF. It is important to mention that some individuals had the same exact traits, and points overlap in this case.

Literature Cited

Akhtar, M. S., Zafar, Z., Nawaz, R., & Fraz, M. M. (2024). Unlocking plant secrets: A systematic review of 3D imaging in plant phenotyping techniques. Computers and Electronics in Agriculture, 222, 109033.

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: setting conservation guidelines. *Trends in ecology & evolution, 16*(11), 613-622.

Arce-Valdés, L. R., & Sánchez-Guillén, R. A. (2022). The evolutionary outcomes of climate-change-induced hybridization in insect populations. *Current Opinion in Insect Science, 54*, 100966.

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity, 99*(3), 301-312.

Conant, R. & Collins, J. T. (1991). *Peterson Field Guide to Reptiles and Amphibians of Eastern/Central North America, 2nd Edition*. Houghton Mifflin Company.

CPW (Colorado Parks and Wildlife). (2026). State Wildlife Action Plan. https://cpw.state.co.us/state-wildlife-action-plan.

Fischer, T. D., Backlund, D. C., Higgins, K. F., & Naugle, D. E. (1999). *A field guide to South Dakota amphibians*. Research Bulletins of the South Dakota Agricultural Experiment Station (1887-2011). 736.

Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Package 'irr'. *Various coefficients of interrater reliability and agreement, 22*, 1-32.

Garroway, C. J., Bowman, J., Cascaden, T. J., Holloway, G. L., Mahan, C. G., Malcolm, J. R., ... & Wilson, P. J. (2010). Climate change induced hybridization in flying squirrels. *Global Change Biology, 16*(1), 113-121.

Gillis, J. E. (1975). *Characterization of a Hybridizing Complex of Leopard Frogs*. [Doctoral dissertation, Colorado State University, Fort Collins]. ProQuest Dissertations and Theses database.

Hammerson, G. A. (1999). *Amphibians and Reptiles in Colorado.* University Press of Colorado.

Hending, D. (2025). Cryptic species conservation: a review. *Biological Reviews, 100*(1), 258-274.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of statistical software, 28*, 1-26.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18-22.

Mazerolle, M. J., & Mazerolle, M. M. J. (2017). Package 'AICcmodavg'. *R package, 281*, 1-220.

O'Donnell, R. P., & Mock, K. E. (2012). Two frog species or one? A multi-marker approach to assessing the distinctiveness of genetic lineages in the northern leopard frog, Rana pipiens. *Conservation Genetics, 13*(5), 1167-1182.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Oksanen, M. J. (2013). Package 'vegan'. *Community ecology package, version, 2*(9), 1-295.

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012, July). How many trees in a random forest?. In *International workshop on machine learning and data mining in pattern recognition* (pp. 154-168). Berlin, Heidelberg: Springer Berlin Heidelberg.

Post, D. D. (1972). *Species differentiation in the Rana pipiens complex*. [Doctoral dissertation, Colorado State University, Fort Collins]. ProQuest Dissertations and Theses database.

Powell, R., Collins, J. T., & Hooper Jr, E. D. (2019). *Key to the herpetofauna of the continental United States and Canada*. University Press of Kansas.

Scriber, J. M. (2013). Climate-driven reshuffling of species and genes: potential conservation roles for species translocations and recombinant hybrid genotypes. *Insects, 5*(1), 1-61.

Stelkens, R., & Seehausen, O. (2009). Genetic distance between species predicts novel trait expression in their hybrids. *Evolution, 63*(4), 884-897.

Wickham, H. (2011). ggplot2. *Wiley interdisciplinary reviews: computational statistics, 3*(2), 180-185.