Species Characterization and Hybrid Investigation in Juvenile Spiny Lizards (Sceloporus spp.) by Genetic Sequencing

Malia Loustalot

Biology 499 (OPIHI Internship)
Mentors: Robert Thomson and Van Wishingrad

Identifying species accurately can be difficult. This can be especially true in groups that have many species, or multiple species that are similar in how they look or where they live. The focus of this study is on lizards in the genus (group marked by similar characteristics) Sceloporus*, or spiny lizards, which is the most diverse genus in the family Phrynosomatidae. Given the number of species in the genus, and areas of range overlap, species identification can be difficult. Here I demonstrate how molecular tools can be used to identify juvenile (individuals who have not yet met sexual maturity) and hatchling* Sceloporus *species using different sources of molecular data, and how they can be analyzed to reliably identify unknown species and investigate possible novel hybrid individuals. I sequenced nuclear and mitochondrial DNA, and analyzed genomic-scale data to inform species identification in this group. Genetic sequencing of a mitochondrial and nuclear gene revealed the identity of the unknown species as well as the identity of a possible novel hybrid (offspring produced from parents of different species), since different species have subtle differences in their gene sequences. The most parsimonious conclusion from these results is the unknown, possible hybrid, individual is a Sagebrush lizard* (S. graciosus) *given that both gene trees place the unknown specimen within this group, rather than one gene from each of the two species. These methods and techniques can be used in the genus broadly to reliably identify species at stages of development when morphological features unique to each species have not yet developed. The consistent data produced indicates the usefulness of these methods in future studies where species identity is in question, and relatively inexpensive and straightforward single-gene phylogenies may be enough for reliable identification.*

I grew up on the island of O'ahu where I attended Kaiser High School then went on to graduate from the University of Hawai'i at Mānoa with a BS in Molecular Cell Biology in the spring of 2020. I have always wanted to become a physician; I will be attending St. Georges University School of Medicine this fall in order to pursue my dream. Working on the article, "Species characterization and hybrid investigation in juvenile spiny lizards (*Sceloporus* spp.) by genetic sequencing" has been an incredible journey that I started my senior year at UH so I am thrilled to see it through to the finish line. I have had the privilege of working with amazing mentors that have aided and encouraged me throughout this past year as I encountered each hurdle in the process of publishing my first article. I have learned invaluable skills from benchwork, to data analysis, to putting together a polished piece of work, that proved useful in obtaining a position as a laboratory technician for the Associate Dean for Research at John A. Burns School of Medicine. I am so grateful for this experience and opportunity to become a published author, as it has allowed me to develop into a more competent investigator of science.

Horizons, Vol. 6, 2021, pp. 32–39

Squamates are the largest order of reptiles, and include lizards, snakes and amphisbaenians (worm lizards) (Vitt & Caldwell, 2014). There are 10,954 species of squamates, and most of these (6,905) are lizards (Uetz, 2010). Groups with many species can be difficult to tell apart because there are so many possible options for what they could be. The focus of this study is on lizards in the genus *Sceloporus*, or spiny lizards, which is the most diverse genus in the family Phrynosomatidae (Leaché, 2010). Given the large number of species in the genus, and areas of range overlap, identification can be difficult, especially in early life stages. For example, there are usually clear observable variations between adult Western Fence (*S. occidentalis*) and Sagebrush (*S. graciosus*) lizards that allow for reliable identification in nature using a number of traits such as size, scale type, and scale color (Stebbins & McGinnis, 2012). The scales on the back of the thighs of the Western Fence lizard are keeled (rough to the touch), while the Sagebrush lizard's scales are granular (smooth to the touch) (Stebbins & McGinnis, 2012). However, due to similar body size, shape, and color, identification by these means can prove difficult and sometimes unreliable as it can be very similar between these two species—especially in the early life stages (Stebbins & McGinnis, 2012).

When morphology (how they look) cannot be used to differentiate between species, biologists who study reptiles and amphibians often rely on geographic location to differentiate between species. However, identification based on geographic location could lead to misidentifications due to range overlap and range changes, as is the case for Western Fence lizards and Sagebrush lizards who inhabit much of the same range throughout western North America (Stebbins & McGinnis, 2012). Additionally, in some cases, individuals may resemble two or more species. In these cases, it may be due to unreliable taxonomic keys (guides to identify species) used to identify species based on how they look, variable appearance between different individuals, or the formation of hybrid individuals. There are several examples of spiny lizard species forming hybrids in nature, such as the Florida Scrub lizard (*S. woodi*) and the Eastern Fence lizard (*S. undulatus*) in Florida (Jackson, 1973), and the Southwestern Fence lizard (*S. cowlesi*) and the Plateau Fence lizard (*S. tristichus*) in Arizona (Leaché & Cole, 2007), so it is possible hybrids could also be formed from Western Fence lizards and Sagebrush lizard parents. In these cases, molecular tools can be used to reliably identify species if it cannot be determined based on their body form and color or the geographic location where they are found. The most well-known and commonly used method of animal molecular identification relies on sequencing the cytochrome oxidase subunit 1 (CO1) "barcode" gene region (i.e., DNA barcoding; Hebert & Gregory, 2005). This gene region is called a barcode region because its unique sequence is often used to differentiate between species—like a barcode in a supermarket. However, sequence data for this barcode region was not readily available for many spiny lizard species, but I was able to find several studies which used other gene regions, including ND1 and TRAF6. I therefore investigated the utility of these gene regions for the objective to identify unknown individuals.

It is important to have accurate species data for several reasons. Any studies in this group rely on accurate species identification, making methods for reliable species identification important. Furthermore, natural history and basic knowledge such as species ranges are only accurate if species are identified reliably. Accurate species identification also provides useful information to determine whether hybridization is occurring within the two species. In this study, I demonstrate how molecular tools can be used to identify spiny lizard species using different sources of molecular data, and how they can be analyzed to reliably identify unknown species and investigate a possible novel hybrid individual.

Methods Summary

Specimen Collected and Morphological Identification

Twenty-nine lizards were captured in the Sierra Nevada mountain range (Figure 1). Of these, three adults could be reliably

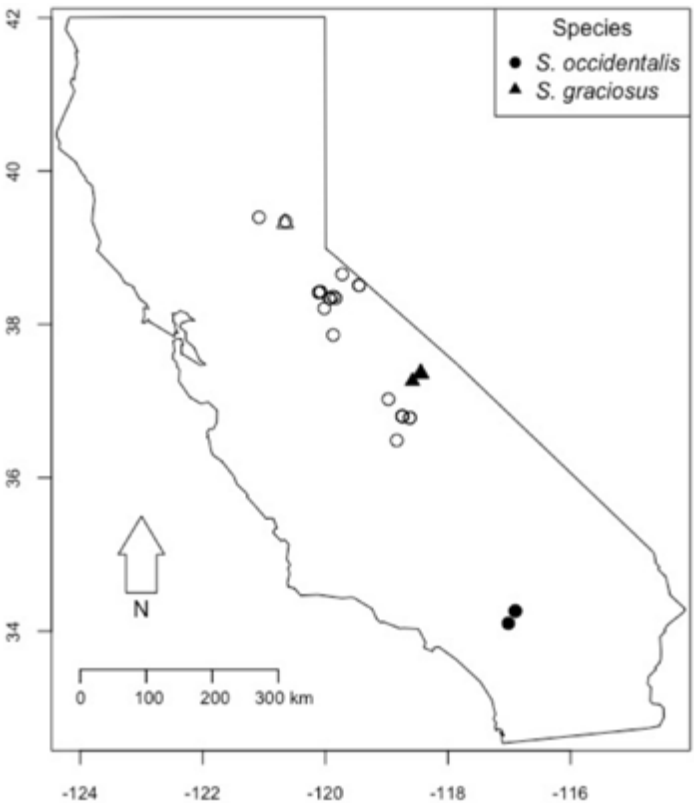

Figure 1. Map of sampled lizards in California, USA. Individuals that could be reliably identified to species are shown using filled symbols, while those that were not identifiable in the field and were identified using molecular methods are shown as open (clear) symbols.

identified as the Western Fence lizard (*S. occidentalis*; IDs: VW004, VW007, VW011) and two adults were positively identified as Sagebrush lizards (*S. graciosus*, Figure 2b; IDs: VW121 & VW146). Species were identified on the basis of their morphology following the identification keys outlined in Stebbins and McGinnis (2012) and the resources found at caliorniaherps .com (Nafis, 2000–2020). These two species have several morphological characters that can help differentiate them. Western Fence lizards (Figure 2a) are usually larger as adults, tend to have yellow/orange on the underside of their legs. They also have scales that are more keeled (rough), especially on the top side of the thighs, and more extensive blue coloration on the throat in males (Stebbins & McGinnis, 2012). Sagebrush lizards (Figure 2b) on the other hand, lack the yellow coloration on the underside of their limbs, have more granular scales—especially on the back of the thighs—and have less extensive blue throat coloration (Stebbins & McGinnis, 2012). Sagebrush lizards also usually have rusty colored axilla (armpit) and a black bar on the shoulder. However, these distinguishing traits are not always present, especially in juveniles, making these identifications difficult. Additionally, one spiny lizard collected had a unique morphology characteristic of both Western Fence lizards and Sagebrush lizards, plus brightly fluorescent orange, yellow, green and blue dorsal scale coloration (Figure 3). As no other Spiny lizards species are present in this area, I explored whether this individual was a Western Fence lizard x Sagebrush lizard hybrid using molecular techniques.

I sequenced nuclear and mitochondrial DNA, and analyzed genomic-scale data to inform species identification in this group. Different species have small differences in their gene sequences, and by looking at the gene sequences one can determine the identity of unknown species. These methods and techniques may be used in the genus broadly to reliably identify species. Briefly, I extracted DNA from a small piece of the tail of each of the twenty-nine lizards collected. I then amplified two genes (increase the number of copies), ND1 from the mitochondrial genome and TRAF6 from the nuclear genome, for each individual using polymerase chain reaction (PCR). Then, I carried out DNA sequencing of these two genes for each individual to conduct analyses: (1) a BLAST (Basic Local Alignment Search Tool) analysis to compare the results to a database; and (2) a phylogenetic reconstruction for each gene tree (to infer the evolutionary relatedness of the individuals tested). Additionally, I carried out a phylogenetic reconstruction using a larger dataset of genomic data from a ddRAD (double-digest restriction site associated DNA sequencing) which captures data from across the entire genome (similar to thousands of genes), rather than within a single gene fragment. To investigate if the possible hybrid individual was indeed a hybrid, I tested if it had increased heterozygosity (i.e., having two different copies of genes in many locations) which can be indicative of hybrid individuals since hybrids would inherit different gene copies from different parental species. For detailed information about the methods used, please refer to the supplementary information document.

All research was conducted under Scientific Collecting Permit SC-13472 issued by the California Department of Fish and Wildlife. Euthanasia methods followed those in protocol 16-2384, approved by the Institutional Animal Care and Use Committee. Lizards were then preserved in a formalin solution and are now curated in the herpetology collection at the Los Angeles Museum of Natural History where they can be held for hundreds of years for future research and as a record of natural history during this time.

Results and Discussion

Genetic sequences

The ND1 gene sequences recovered were 934 nucleotides in length for 20 lizards with 629 (68.1%) identical sites. A sample

Figure 2. Images of adult specimens of two lizard species: (a) Western Fence lizard (*S. occidentalis*) and (b) Sagebrush lizard (*S. graciosus*). Note that while the differences between these two lizards is evident in these adult specimens, they are not visually distinguishable as juveniles. Similarly, while these two specimens differ in overall color, these differences are not fixed between species and color varies widely in each species.

Figure 3. (a) Dorsal and (b) ventral images of the spiny lizard (*Sceloporus*) individual collected which had a unique morphology. It was characteristic of both a Western Fence lizard (*S. occidentalis*) and Sagebrush lizards (*S. graciosus*), and also exhibited brightly fluorescent orange, yellow, green and blue dorsal scale coloration.

of the sequence alignment for bases 100–150 is shown in Figure 4a, and the number of nucleotide differences between each sequence is shown in Figure 4b. The TRAF6 gene sequences recovered were 510 nucleotides in length for 14 lizards, with 485 (95.1%) identical sites. A sample of the sequence alignment for bases 100–150 is shown in Figure 5a, and the number of nucleotide differences between each sequence is shown in Figure 5b.

BLAST (Basic Local Alignment Search Tool) from the (National Center for Biotechnology Information) Database

The BLAST tool compares a sequence of DNA to all sequences in a NCBI (National Center for Biotechnology Information) database to identify the species from which the DNA sequence is most similar to. Using the BLAST tool, I identified the percent identity of all the bases in the possible hybrid sequence in comparison to the individuals with known identities. I found that both ND1 and TRAF6 genes from lizard VW120 were more similar to the Sagebrush lizard sequences than any others in the database. The ND1 and TRAF6 genes produced similarities of 97.17% with an E-value (error value) of 0, and 98.5% with an E-value of 0, respectively. This means the gene sequences of this individual are very similar to existing database samples for known species, suggesting it is not a hybrid. Through the BLAST results I saw that the ND1 gene only contained Sagebrush lizards in the top ten matches, while the TRAF6 gene included ten different species in the top ten matches. All these matches had a >97% identity and E-values indistinguishable from zero. This suggests BLAST may be a useful tool for identification using ND1 but may be less reliable for genes such as TRAF6 where many different species have high similarity values. Species identification using this method will also be limited to species in the database and therefore should be interpreted with caution.

Nuclear and Mitochondrial Gene Trees

Phylogenetic trees use gene sequence information to show the way the gene sequences are evolving and describe the evolutionary relationships between individuals. The ND1 phylogenetic tree shows two major clades (groups of related individuals), one including known Western Fence lizard individuals who were positively identified in the field based on morphological traits, and the other clade including Sagebrush lizards positively identified in the field. The unknown individuals fell into these two clades, with no intermediate groupings or additional clades. Therefore, the results suggest lizard individuals VW120 and

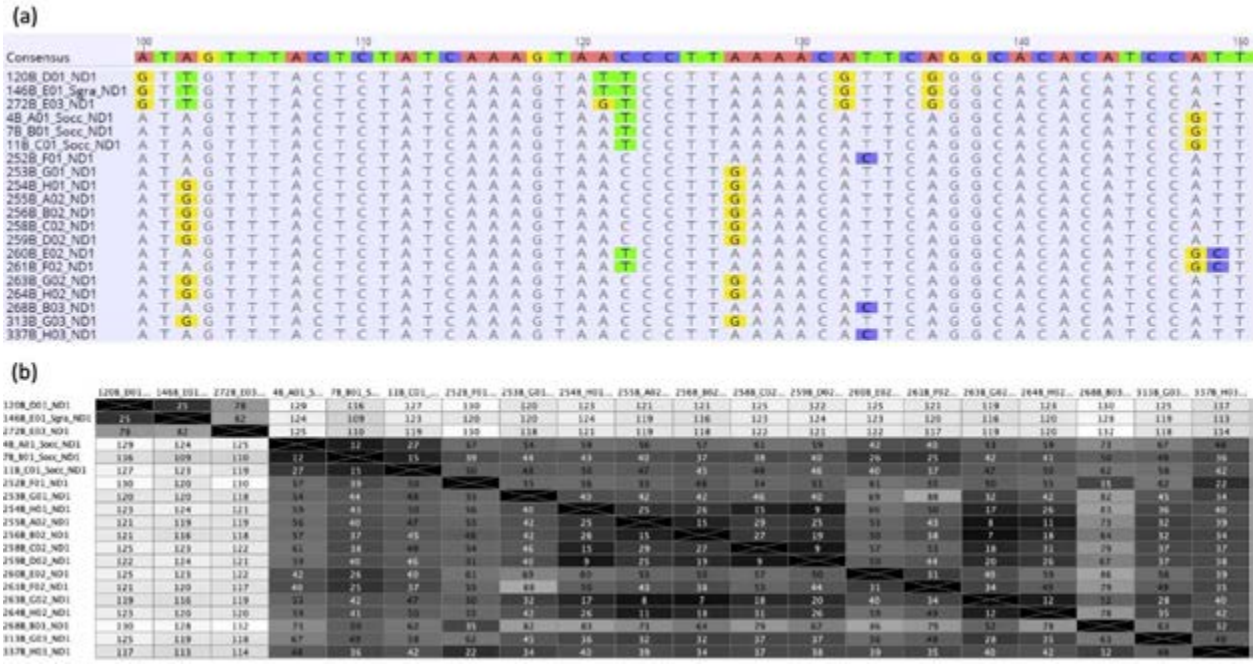

Figure 4. (a) Nucleotide view of the ND1 gene sequences for each lizard. The top "consensus" sequence shows the most common sequence of nucleotides, which each sequence below shows that of each individual listed in the left column for bases 100–150 out of 934 total nucleotides as an example. Highlighted colors indicate where each sequence differs from the reference "consensus" sequence. (b) Number of nucleotide differences between each of the sequences for the entire ND1 gene sequence. Darker colors indicate fewer differences, while lighter colors indicate more differences. Here, lizards VW120, VW146 (Sagebrush lizard, *S. graciosus*), and VW272 are more identical to each other than they are to the other sequences.

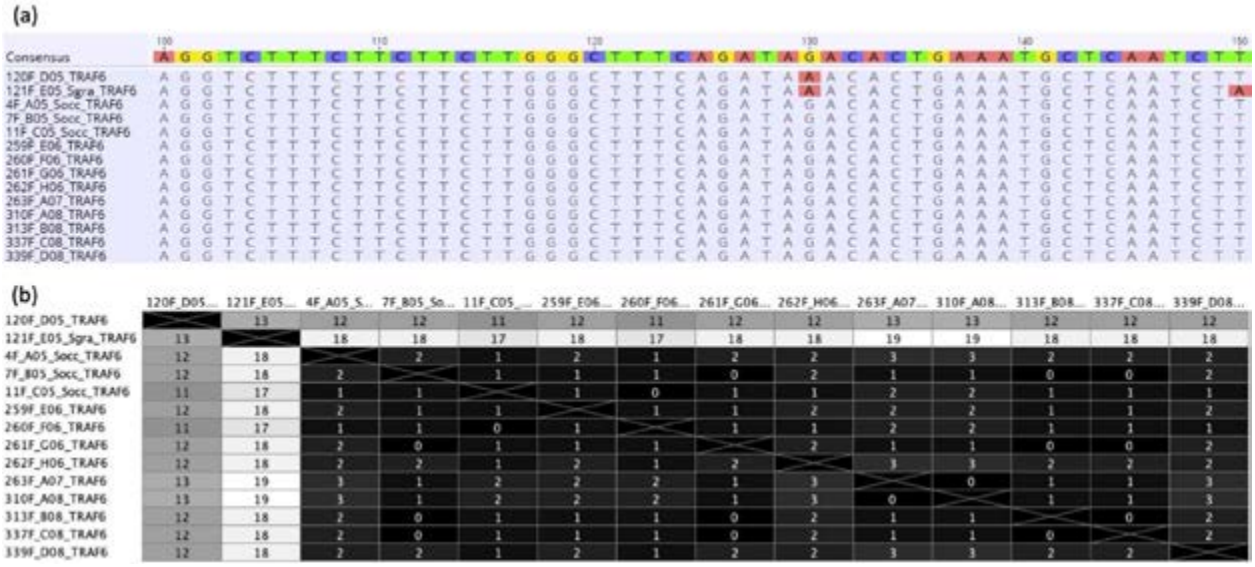

Figure 5. (a) Nucleotide view of the TRAF6 gene sequences for each lizard. The top "consensus" sequence shows the most common sequence of nucleotides, which each sequence below shows that of each individual listed in the left column for bases 100–150 out of 510 total nucleotides as an example. Highlighted colors indicate where each sequence differs from the reference "consensus" sequence. (b) Number of nucleotide differences between each of the sequences for the entire TRAF6 gene sequence. Darker colors indicate fewer differences, while lighter colors indicate more differences. Here, lizards VW120 and VW121 (Sagebrush lizard, *S. graciosus*) are more identical to each other than they are to the other sequences.

VW272 are Sagebrush lizards, while the remaining unknown lizards are Western Fence lizards (Figure 6).

Similar to the ND1 gene phylogenetic tree, the TRAF6 gene phylogenetic tree also shows a distinction between two main clades of Western Fence lizards and Sagebrush lizards. This phylogenetic tree confirms the information provided in the ND1 tree, indicating that VW120 is a Sagebrush lizard, while the remaining unknowns are Western Fence lizards (Figure 7).

Genomic (ddRAD) Phylogenetic Tree

The genome-wide ddRAD tree also recapitulates the results shown on each of the ND1 and TRAF6 mitochondrial and nuclear gene trees. There are two main genetic groups, one of which contains all known Western Fence lizard individuals, and the other contains all known Sagebrush lizard individuals. These results indicate lizards VW120 and VW272 are Sagebrush lizards (*S. graciosus*), while the remaining unknowns are Western Fence lizards (*S. occidentalis*) (Figure 8).

Heterozygosity Estimates

Increased heterozygosity (i.e., having two different copies of genes in many locations) which can be indicative of hybrid individuals since hybrids would inherit different gene copies from different parental species. I estimated heterozygosity using the ipyrad program version 0.9.42 (Eaton & Overcast, 2020). I found that VW120, the individual with striking intermediate morphology and traits consistent with both Western Fence lizards and Sagebrush lizard species, had no excess of heterozygous sites as would be expected for a hybrid individual (Figure 9).

Conclusions

Genetic sequencing of mitochondrial and nuclear genes, as well as analysis of genomic-scale ddRAD data provided the identity of the unknown individuals as well as the identity of the potential novel hybrid in question. The most parsimonious conclusion from these results is the suspected hybrid individual is not a hybrid, but a Sagebrush lizard given that both gene trees and the ddRAD tree place the unknown specimen within this clade. These results demonstrate the essential role that molecular tools play in reliable species identification when a definitive answer cannot be determined by observation alone.

It is noted that the BLAST results for the TRAF6 gene included several other species aside from Sagebrush species, however this does not affect my conclusions about the species identification. Of the several species included, there was only one that had a range overlap nearby (Desert Spiny lizard, *S. magister*) but was dismissed due to the contrasting phenotypic differences (much larger, darker, and having large keeled scales). Through these results, it can be seen that both nuclear as well as mitochondrial genetic sequencing provided a reliable output for the identification of the unknown species. The results are further verified by the genome-scale data. These are

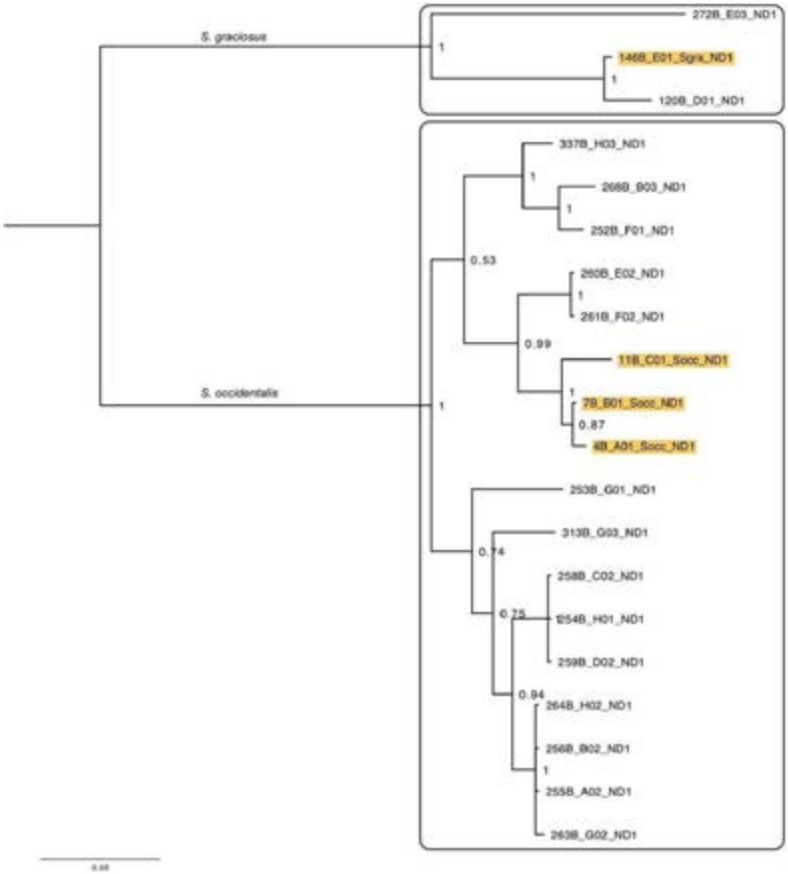

Figure 6. Phylogenetic tree from the mitochondrial gene region ND1 from a total of 934 nucleotides. Each unknown individual is represented by a unique code alphanumeric code, and known individuals include either "Socc" (Western Fence lizard, *S. occidentalis*) or "Sgra" (Sagebrush lizard, *S. graciosus*) in their code. Numbers on nodes indicate Bayesian posterior probabilities, where values closer to 1 indicate higher confidence in each split. Highlighted tips are individuals identified in the field based on morphological traits. The two circled groups delineate the two species: Sagebrush lizards (*S. graciosus*), top, and Western Fence lizards (*S. occidentalis*), bottom.

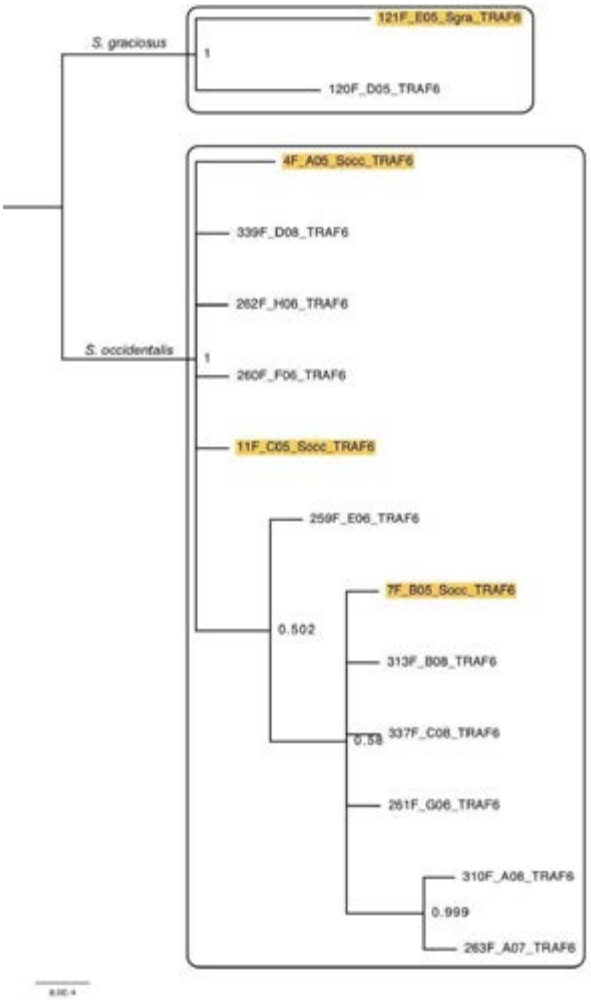

Figure 7. Phylogenetic tree from the nuclear gene region TRAF6 from a total of 510 nucleotides, and each locus has data for at least 50% of the individuals. Each unknown individual is represented by a unique code alphanumeric code, and known individuals include either "Socc" (Western Fence lizard, *S. occidentalis*) or "Sgra" (Sagebrush lizard, *S. graciosus*) in their code. Numbers on nodes indicate Bayesian posterior probabilities, where values closer to 1 indicate higher confidence in each split. Highlighted tips are individuals positively identified in the field based on morphological traits. The two circled groups delineate the two species: Sagebrush lizards (*S. graciosus*), top, and Western Fence lizards (*S. occidentalis*), bottom.

promising results as it implies that these methods produce consistent results which can be used in future studies. Using the molecular tools mentioned, the possible hybrid species was able to be identified as a Sagebrush lizard, rather than a hybrid, as the phenotypic observations had suggested. These methods can be used in future studies for reliably identifying morphologically intermediate species using mitochondrial and nuclear genetic sequencing.

The different methods each have benefits and drawbacks. The BLAST method is the simplest method, but least reliable as several other species showed up on the top matches, likely because BLAST relies on measures of similarity rather than phylogenetic relatedness. Phylogenetic tree methods on the other hand seem to be the most reliable as they attempt to measure evolutionary relatedness directly. Either way, having a good *a priori* hypothesis, a hypothesis made prior to beginning research, about the possible species is always beneficial as this reduces the number of possible matches to choose from and increases the probability of a correct identification.

Spiny lizards are a highly diverse genus that contains many species that encompass a large range in North America (Leaché, 2010). As a result of this, traditional methods of identification may fall short when identifying juveniles leading to misidentified species. Molecular tools have shown they can fill the gap as the unknown individuals as well as the possible hy-

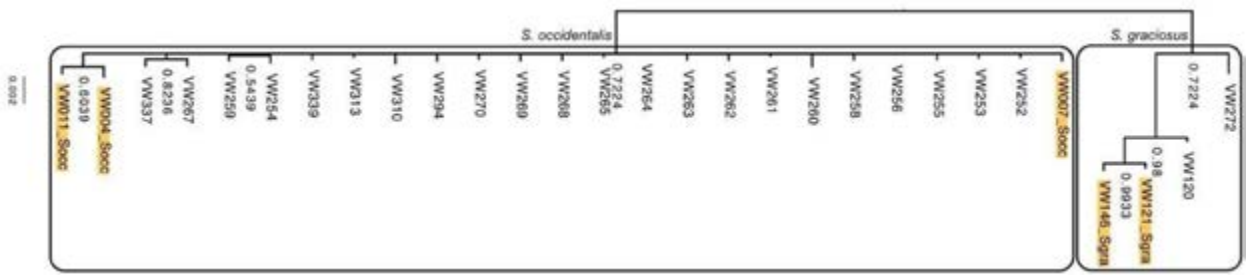

Figure 8. Phylogenetic tree based on ddRAD sequence data from a total of 93,109 nucleotides. Each unknown individual is represented by a unique code alphanumeric code, and known individuals include either "Socc" (Western Fence lizard, *S. occidentalis*) or "Sgra" (Sagebrush lizard, *S. graciosus*) in their code. Numbers on nodes indicate Bayesian posterior probabilities, where values closer to 1 indicate higher confidence in each split. Highlighted tips are individuals positively identified in the field based on morphological traits. The two circled groups delineate the two species: Western Fence lizards (*S. occidentalis*), left, and Sagebrush lizards (*S. graciosus*), right.

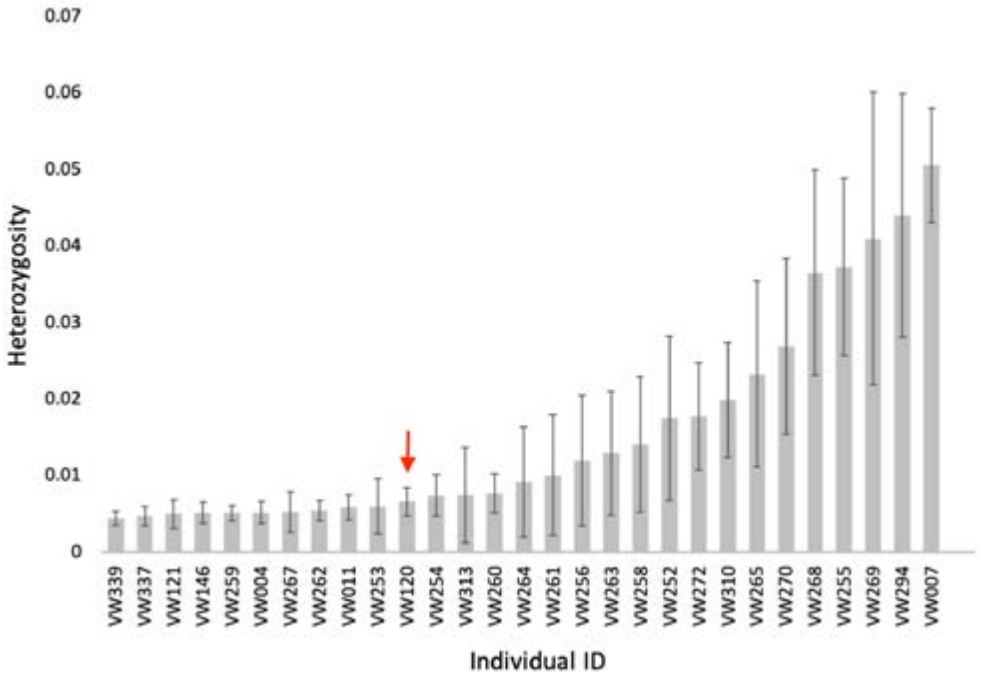

Figure 9. Heterozygosity measurements for all samples. The red arrow indicates the suspected hybrid individual, while the other samples are known non-hybrids (adults positively identified based on morphological traits). Very high heterozygosity (beyond the maximum value of ~0.05 observed here) would be evidence of a hybrid individual—such an individual would be expected to have an excess of different genes (increased heterozygosity) due to inheriting different copies from different parental species. Here, the suspected hybrid does not have an excess of heterozygosity, suggesting that it is not a hybrid individual.

brid were able to be identified using this approach. On a larger scale, these methods could be used to distinguish between lizard species outside the Spiny lizard genus whose key morphological features necessary for identification are not yet present.

This method can provide important insight for species identification more generally. In Hawai'i, for example, there are numerous species—both terrestrial (land) and marine—that can be difficult to identify visually. The single-gene phylogeny method I use here for species identification may represent an ideal tradeoff between research cost and effort and species identification accuracy for organisms that cannot be identified through traditional methods such as their appearance.

Supplemental Information

Additional details about the methods used here may be found at http://doi.org/10.5281/zenodo.5068309

References

Ammundsen, B. & Duran, C. (2015). Geneious v8.1.9 (https://www.geneious.com)

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*. doi: 10.1016/S0022-2836(05)80360-2

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology, 22*(11), 3124–3140. doi: 10.1111/mec.12354

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). JModelTest 2: More models, new heuristics and parallel computing. *Nature Methods, 9*(8), 772. doi: 10.1038/nmeth.2109

Eaton, D. A. R., & Overcast, I. (2020). Ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*. doi: 10.1093/bioinformatics/btz966

Harris, R., Banbury, B., & Leache, A. (2014). Comparative genomic resources for spiny lizards (genus Sceloporus) in Genomic Resources Notes accepted 1 August 2014–30 September 2014. *Molecular Ecology Resources*, (August 2014), 1–9. doi: 10.5061/dryad.n2q7f

Hebert, P. D. N., & Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. *Systematic Biology, 54*(5), 852–859. doi: 10.1080/10635150500354886

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. doi: 10.1093/bioinformatics/17.8.754

Jackson, J. F. (1973). The Phenetics and Ecology of a Narrow Hybrid Zone. *Evolution, 27*(1), 58. doi: 10.2307/2407119

Leaché, A. D. (2010). Species trees for spiny lizards (Genus Sceloporus): Identifying points of concordance and conflict between nuclear and mitochondrial data. *Molecular Phylogenetics and Evolution, 54*(1), 162–171. doi: 10.1016/j.ympev.2009.09.006

Leaché, A. D., & Cole, C. J. (2007). Hybridization between multiple fence lizard lineages in an ecotone: Locally discordant variation in mitochondrial DNA, chromosomes, and morphology. *Molecular Ecology, 16*(5), 1035–1054. doi: 10.1111/j.1365-294X.2006.03194.x

Nafis, G. (2020). A Guide to the Amphibians and Reptiles of California. Retrieved March 12, 2020, from http://www.californiaherps.com/

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE, 7*(5). doi: 10.1371/journal.pone.0037135

Rambaut, A. (2009). *FigTree v1. 3.1: Tree figure drawing tool*. Edinburgh, UK.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology, 67*(5), 901–904. doi: 10.1093/sysbio/syy032

Stebbins, R. C., & McGinnis, S. M. (2012). *Field Guide to Amphibians and Reptiles of California*. University of California Press. Retrieved from http://www.jstor.org/stable/10.1525/j.ctt1pn65t

Uetz, P. (2010) The original descriptions of reptiles. Zootaxa 2334: 59–68

Wiens, J. J., Kuczynski, C. A., Arif, S., & Reeder, T. W. (2010). Phylogenetic relationships of phrynosomatid lizards based on nuclear and mitochondrial data, and a revised phylogeny for Sceloporus. *Molecular Phylogenetics and Evolution, 54*(1), 150–161. doi: 10.1016/j.ympev.2009.09.008