Exploration of the patterns and processes driving lineage diversification

Hayden R. Davis

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

Reading Committee:

Adam D. Leaché, Chair
Sharlene E. Santana
Jeffrey A. Riffell

Program Authorized to Offer Degree:
Department of Biology

University of Washington

Abstract

Exploration of the patterns and processes driving lineage diversification

Hayden R. Davis

Chair of the Supervisory Committee:
Adam D. Leaché

Department of Biology

Lineages are constantly evolving, leading to the formation of distinct populations, and in some cases, species. In this dissertation, I explore three independent study systems that are at different stages of the speciation continuum, ranging from very recently diverged subpopulations to species-level diversity. This approach enables me to explore the genomic and life history characteristics driving diversification on multiple evolutionary levels. In Chapter 1, I focus on the genetic, morphological, and life history traits of a population of the Western Fence Lizard occurring at the northernmost extent of the distribution for the species. In Chapter 2, I focus on how interpopulation gene flow patterns correlate with distinct ecoregions in the southwestern United States and northern Mexico for the Western Banded Gecko. Lastly, in Chapter 3, I focus on the efficacy of current methodologies used for species delimitation in Southeast Asian geckos by testing these methods on three distinct species of bent-toed geckos occurring in Borneo. Exploring these three distinct study systems adds valuable insight into the evolution of natural organisms with unique evolutionary pressures across multiple time scales.

0.1 Acknowledgements

For Chapter 1, I am thankful to the Washington Department of Fish and Game for granting permission to collect specimens for scientific research (LIC #20-144). This project was supported by a National Science Foundation grant to A.D.L. (NSF-DEB-SBS-2023723). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by a NIH S10 OD018174 Instrumentation Grant.

For Chapter 2, I am thankful to the Arizona Game and Fish Department for permission to collect specimens (LIC#SP568189). This project was supported by a National Science Foundation grant to Adam D. Leaché (NSF-DEB-SBS-2023723). I am thankful to multiple individuals who helped collect *Coleonyx variegatus* samples. Jonathan Q. Richmond from the United States Geological Survey and his associated field team at the USGS Western Ecological Research Center, San Diego Field Station collected and sent specimens from poorly samples regions of California that were critical to defining populations in the study (Permit ID Number: SC-838). For fieldwork in Arizona, Lawrence L. C. Jones, John J. Wiens, and Wade C. Sherbrooke assisted during two nights of fieldwork in Tucson; Danny Martin assisted for two nights of collecting in Ajo; and Anthony J. Cobos and Aryeh H. Miller assisted with fieldwork in Mohave County, Arizona. I am also overly grateful to Wade who hosted Adam Leacheé and myself at his house while conducting research in Tucson. I appreciate the tenacity and willingness to help from Aryeh and Anthony who spent one week camping with me in the desert, as they stayed motivated despite being unable to find any geckos. Last but not least, my mother, Julie B. Davis, assisted for three nights of fieldwork camping in the deserts of Mohave County, Arizona. Despite not having camped in decades and never being involved in research, she found geckos, assisted with data collection, and proved a very hardy camper. I am thankful for the training and relief from teaching that I received through the Genome Training Grant program, which provided two years of fellowship funding for this project (NOT-OD-17-095). I am grateful to the donors who created the Kathryn C. Hahn Writing Fellowship, which provided me with a quarter of support to focus solely on writing. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by a NIH S10 OD018174 Instrumentation Grant.

For Chapter 3, I thank the Sarawak Forestry Department and the Sarawak Forestry Corporation for providing collection [NPW.907.4.4.(Jld.14)-79; (119)JHS/NCCD/600-7/2/107] and export permits (No. 18160). I am also grateful to Brett O. Butler and members of the Leaché Lab for providing comments and insight prior to submission. This chapter of my research was related to the research I conducted during my Master's program at Villanova University in the lab of Aaron M. Bauer. Aaron's research, philosophical, financial, and motivational support were key to establishing the long running research I have conducted in Sarawak, Malaysia, Borneo. Additionally, he connected me with Indraneil Das, who connected me with his student Izneil Nashriq. My work with Indraneil and Izneil has been some of the most motivating research to this

point in my career. Indraneil has been an extremely supportive colleague from our first introduction. Izneil and I spent months collecting geckos from across the state of Sarawak, and as a result, he is a dear friend and a great colleague. Benjamin R. Karin and Ian G. Brennan conducted multiple months of fieldwork, resulting in critical sampling for this study to take place. They have also been great coauthors on multiple prior projects. Funding for this work came from multiple sources. I am grateful to the donors who generated the following grant through UW Biology, from which I received the following funding: Heerensperger Award, Orians Awards for Tropical Studies, and the Margo and Tom Wyckoff Award. I also received critical funding for data collection from the Society of Systematic Biologists and the Association for Tropical Biology and Conservation. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by a NIH S10 OD018174 Instrumentation Grant.

I am extremely grateful for the supportive coauthors I have had on these projects who provided samples, conducted analyses, enhanced my writing, and helped synthesize results in a clear and concise manner. For chapter 1, Simone Des Roches collected and analyzed all of the morphology data and helped broaden the focus of the study, while greatly improving the writing. Roger A. Anderson provided $>$ 30 years of ecological data and knowledge on the Western Fence Lizards in the Puget Sound region. For chapter 2, Dean H. Leavitt conducted years of sampling for the Western Banded Gecko and provided his critical samples for use in our study. Atinuke Bandele was a highly motivated undergraduate who worked with me on the genomic library preparation and data analysis for this project. Sonal Singhal, a role model of mine, helped with the project conceptualization and writing. Julio A. Lemos-Espinal helped coordinate fieldwork in Mexico, that unfortunately fell through, and added samples from minimally sampled regions. Lastly for chapter 3, Henry T. Sanford was highly proficient undergraduate student who conducted multiple bioinformatic analyses and assisted with writing the manuscript. The contributions of the coauthors Indraneil Das and Izneil Nashriq are mentioned above.

My thesis committee has been very supportive of my research and I am very appreciate of their insight. My committee is comprised of Adam D. Leaché (committee chair), Sharlene E. Santana (committee member), Jeffrey A. Riffell (committee member), and Kelley Harris (graduate school representative). Due to COVID-19, my thesis plans changed drastically from year one to the present. My committee has been very amenable to me developing new projects and changing priorities throughout my PhD. They have been quick provide feedback and support when needed, and continually supported my career trajectory. My committee has helped me consider the bigger picture of my research conduct and helped frame the studies and questions accordingly. I am very thankful the support that they have provided.

My advisor, Adam D. Leaché, has elevated my science far beyond what I felt I was capable of. He has been extremely supportive both in and out of lab, and motivated me to be a better scientist. I am fortunate to have an advisor who helped from the permitting and fieldwork stages all the way through to

analysis and writing. He pushed me to continue making progress on my work through COVID-19 lockdowns, spent weeks in the field with me, and without fail reviewed written drafts of mine within a couple of days. He has also been supportive of my career trajectory, including my decision to do an internship at Phase Genomics, writing multiple reference letters, and sending me job positions that he felt I would be a good fit for. I can't express my appreciation enough.

Lastly, I want to thank my wife, Elena Stiles, for her continual support and motivation. The greatest thing that came out of my PhD is meeting her. She has helped me maintain a level head and allowed me to disconnect from a work mindset. The reason I am finishing my PhD on time and moving onto the next steps in my career is because of her positivity and motivational push. She's among the smartest people I've met, and I am excited to see what her future hold after her graduation.

0.2 Dedication

This thesis is dedicated to my lab mate and friend, Yu-Wei Hsiao. I will always remember the joy you brought to my life and lives of those around you. I will always miss you.

Chapter 1

Population expansion, divergence, and persistence in Western Fence Lizards (*Sceloporus occidentalis*) at the northern extreme of their distributional range

Hayden R. Davis, Simone Des Roches, Roger A. Anderson, & Adam D. Leaché

1.1 Abstract

Population dynamics within species at the edge of their distributional range, including the formation of genetic structure during range expansion, are difficult to study when they have had limited time to evolve. Western Fence Lizards (*Sceloporus occidentalis*) have a patchy distribution at the northern edge of their range around the Puget Sound, Washington, where they almost exclusively occur on imperiled coastal habitats. The entire region was covered by Pleistocene glaciation as recently as 16,000 years ago, suggesting that populations must have colonized these habitats relatively recently. We tested for population differentiation across this landscape using genome-wide SNPs and morphological data. A time-calibrated species tree supports the hypothesis of a post-glacial establishment and subsequent population expansion into the region. Despite a strong signal for fine-scale population genetic structure across the Puget Sound with as many as 8—10 distinct subpopulations supported by the SNP data, there is minimal evidence for morphological differentiation at this same spatiotemporal scale. Historical demographic analyses suggest that populations expanded and diverged across the region as the Cordilleran Ice Sheet receded. Population isolation, lack of dispersal corridors, and strict habitat requirements are the key drivers of population divergence in this system. These same factors may prove detrimental to the future persistence of populations as they cope with increasing shoreline development associated with urbanization.

1.2 Introduction

Population divergence occurs across a broad range of spatial and temporal scales. The expected pattern of genetic variation, however, depends partly on whether the populations are located at the periphery or core of the range [Wiley, 1988, Rundle and Nosil, 2005]. As a consequence of their smaller size, peripheral populations often have lower genetic diversity compared to core populations [Eckert et al., 2008, Assis et al., 2013]. Furthermore, the physiological constraints experienced by peripheral lineages can limit dispersal and gene flow among and between populations, which can restrict them to smaller geographic areas. As a result, peripheral populations can be relatively homogeneous with low genetic diversity due to low levels of dispersal and gene flow. Diversification at microgeographic scales is often correlated with either environmental heterogeneity or minor phenotypic changes (e.g., in life history) that contribute to reduction in gene flow among populations [Richardson et al., 2014, Langin et al., 2015, Mikles et al., 2020, Leaché et al., 2010]. Diversification at fine spatial scales can also occur when a species occupies a non-optimal niche or reaches the edge of a geographic range causing populations to become geographically fragmented, resulting in population structure [Angert et al., 2020, Sexton et al., 2009].

Intraspecific diversity is also a product of ongoing and historic environmental changes, including climate driven habitat change which can alter the genetic

composition of a population [Des Roches et al., 2020]. Populations near the edges or at the extremes of their preferred habitats and environments run up against physiological limitations or stressors, such as thermal maxima or minima, that can cause allele frequencies to shift as populations become smaller and more fragmented [Bowman et al., 2018, van Rensburg et al., 2018, Eckert et al., 2008]. The low genetic diversity of species already at their physiological limits further restricts adaptive potential in the face of abiotic stressors [Hoffmann and Blows, 1994, Pearson et al., 2009]. However, this central-peripheral hypothesis is not always supported in natural systems [Garner et al., 2004, Eckert et al., 2008], and there is even evidence of groups excelling at the periphery [Kilvitis et al., 2017]. Founder events and novel abiotic pressures can work together to shape patterns of genetic variation and population structure observed within species at the edge of their range.

The Western Fence Lizard (*Sceloporus occidentalis*) spans a large geographic area in the Western North America, extending from Baja California, Mexico into the Puget Sound region, Washington [Bouzid et al., 2022, Sites et al., 1992]. Across this relatively large distribution, at least five distinct genetic groups can be identified that are separated by major biogeographic barriers [Bouzid et al., 2022]. The Pacific Northwest (PNW) group comprises the northernmost populations, extending from Northern California into Oregon and Washington, and it shares a most recent common ancestor with populations from the Western Sierra Nevadas [Bouzid et al., 2022, Pereira and Singhal, 2022]. Throughout much of their central and southern range, populations of *S. occidentalis* are geographically continuous and successful in nearly all habitats, exploiting both urban and natural environments. Yet, populations in Washington have a fragmented distribution, occurring solely along the Columbia River near the Oregon border, in the Cascade Mountains in central Washington, and on shoreline habitats around the Puget Sound in Western Washington (Fig. 1.1).

Nowhere is the fragmented distribution of *S. occidentalis* more pronounced than at the northernmost extent of the PNW group around the Puget Sound where the species is restricted to small, isolated localities scattered across islands and coastal habitats (Figs. 1.1, S1.1–S1.2). With current temperatures in western Washington being lower than temperatures farther south in the species range, reduced fitness in the midst of a colder climate may be restricting their northern distribution, as well as limiting the number of suitable habitats [Sinervo and Huey, 1990, Sinervo, 1990, Sinervo and Adolph, 1994]. Past climatic events have also presumably played a large role in shaping the distribution of the species, particularly the expansion of the Cordilleran Ice Sheet, which covered the entirety of the Puget Sound as recently as 16 kya [Haugerud, 2020, Booth et al., 2003]. As a result, the colonization, or recolonization, of the Puget Sound by *S. occidentalis* has presumably been restricted until the Holocene glacial retreat. Therefore, *S. occidentalis* distributed around the Puget Sound region may lack genetic structure due to a combination of factors including recent colonization, range edge effects, and small population size.

In contrast to California populations, few studies have focused on the ecology and life history of *S. occidentalis* at the northern edge of their range. The species

is diurnal and a general insectivore with little indication of dietary specialization [Johnson, 1965]. However, the Washington populations differ substantially in many aspects of their life history. Washington populations hibernate between approximately October and March, leading to an increased standard metabolic rate during their active months [Tsuji, 1988]. The populations in Washington also have increased clutch sizes in comparison to California (12 eggs/clutch vs. 7 in California) that tend to hatch more quickly, presumably due to their smaller size Sinervo [1990]. Despite physiological comparisons of northern to central or southern populations, few studies have directly addressed the ecological or habitat requirements for the northern populations [Backus, 2016, Powers et al., 2018, McTernan, 2017]. In addition to investigating the genomic and morphological variation of *S. occidentalis* around the Puget Sound, we also aim to provide new information regarding their microhabitat requirements in this region.

We present the first comprehensive study of the distribution of *S. occidentalis* in the Puget Sound and use genome wide single nucleotide polymorphisms (SNPs) and morphological data to test whether population divergence is present in the region. We then describe patterns of genetic and morphological variation and use demographic models to understand how populations have evolved through time. Lastly, since *S. occidentalis* in the Puget Sound region is less abundant and more fragmented than other populations throughout the species distribution, we discuss the ecological requirements necessary for their success at the northernmost extent of their range. These new data provide strong support for multiple genetically distinct groups in the Puget Sound region, which has important implications for conservation and population management.

1.3 Methods

1.3.1 Ethics Statement

The research presented herein was conducted in accordance with the ethical standards and guidelines outlined by the University of Washington's Institutional Animal Care and Use Committee (IACUC protocol #4367-02). Samples were collected with permission from the Washington Department of Fish and Wildlife (WDFW permit 20-144).

1.3.2 Study System and Taxon Sampling.

We collected *Sceloporus occidentalis* individuals from 19 localities across the Puget Sound Region and four localities from surrounding regions in Eastern Washington, with most sampling being conducted in 2020 (Figure 1.1; Table 1.1). We captured lizards using a three meter fishing rod with a thin loop on the end, which was placed over the head of the lizard and subsequently tightened. In total, we incorporated 90 samples in our study with 78 of those from the Puget Sound Region (Table 1.1). Voucher specimens and tissue samples are accessioned at the Burke Museum of Natural History and Culture

(UWBM:HERP:10034–10111; Table S1.1). Our morphological and genomic datasets are mostly overlapping with a few exceptions: when the specimen wasn't collected or was a juvenile, we did not collect morphology data; when we had low sequencing coverage for an individual, we removed it from the genetic dataset.

1.3.3 Molecular Methods.

We extracted genomic DNA from liver biopsies using salt-extraction and then conducted double digest restriction-site associated DNA sequencing (ddRAD-seq) [Peterson et al., 2012, Aljanabi and Martinez, 1997]. We double-digested each sample using the digestion enzymes SbfI and MspI in CutSmart Buffer (New England Biolabs) for 7 hours at 37 °C. For fragment purification, we used Sera-Mag SpeedBeads. We then prepared a master mix for eight distinct barcodes to be ligated to the cut sites of the fragmented DNA. The libraries were size-selected (between 415 and 515 bp after accounting for adapter length) on a Blue Pippin Prep size fractionator (Sage Science). For the final library amplification, we used Phusion Hi-Fidelity DNA Polymerase and Illumina's indexed primers. We determined the concentration and size distribution of each indexed pool using an Agilent 2200 TapeStation. Lastly, we sent the quantified pools to QB3-Berkeley Genomics, UC Berkeley for qPCR to determine sequenceable library concentrations before multiplexing equimolar amounts of each pool for sequencing on one Illumina HiSeq 4000 lane (51-bp, single-end reads; 11 pools containing up to 8 samples each). The demultiplexed sequences are deposited at the Sequence Read Archive (NCBI-SRA; BioProject ID: PRJNA757434; Table S1.1).

1.3.4 Bioinformatics.

We demultiplexed each sample from their respective pool using their unique barcode sequence using ipyrad v.0.9.50 [Eaton and Overcast, 2020]. We conducted a reference-based assembly of the RAD loci using a draft of the *S. occidentalis* genome from Yosemite National Park, California (Table 1.2; [Genomic Resources Development Consortium et al., 2015, Westfall et al., 2021]). A sequence similarity threshold of 90% was used to cluster reads within samples and loci between samples. We removed consensus sequences with low coverage (< 6 reads), excessive undetermined or heterozygous sites ($> 5\%$), too many alleles for a sample (> 2 for diploids), or an excess of shared heterozygosity among samples (paralog filter $= 0.5$). For the final alignments we generated output files containing 0% missing data (1,037 loci) and 50% missing data (3,491 loci; Tables 1.2, S1.2). Downstream population genetic analyses used additional filtering to subsample one random SNP per locus, and those datasets are described below.

1.3.5 Genomic differentiation.

We conducted genetic clustering analyses to estimate population structure. Using the 50% missing dataset, we ran a principal component analysis (PCA) using the R package ADEGENET to establish general patterns of genetic diversity, and a discriminant analysis of principal components (DAPC) to determine if a sample's locality could be determined by genomic data [Jombart et al., 2010, Jombart and Ahmed, 2011]. We used the program ADMIXTURE to visualize ancestry within and among populations [Alexander and Lange, 2011]. This analysis used a reduced dataset containing one randomly sampled SNP from each locus. Due to the shallow divergence and microgeographic scale of the study system, estimating an optimal K value proved difficult. We repeated the ADMIXTURE analysis 10 times for each K value ranging from 1–10 and used the program's cross-validation (CV) procedure to test for the optimal number of subpopulations (lowest CV = optimal K-value). To visualize and compare results, we used the program CLUMPAK to produce structure barplots [Kopelman et al., 2015]. We repeated these procedures for datasets with 0% (236 SNPs) and 50% (707 SNPs total) datasets.

Using the genetic subpopulations identified by the population structure analyses, we quantified the extent of genetic differentiation (F_{ST}) between each subpopulation using the option --weir-fst-pop in the program VCFTOOLS [Danecek et al., 2011]. We used the 50% missing data dataset to maximize the number of loci in the analysis. Because the Puget Sound population is not geographically continuous throughout the region, our sampling was often confined to small, restricted geographic areas. To avoid including samples from siblings and other close familial relationships, we calculated the inbreeding coefficient among individuals using the option --relatedness2 in VCFTOOLS [Manichaikul et al., 2010]. Lastly, we used the program MEGA version X to calculate nucleotide diversity (π) among subpopulations [Kumar et al., 2018].

1.3.6 Phylogenetic analyses.

To estimate the phylogenetic relationships among samples and the timing of population divergence we used a combination of network, concatenated, and coalescent-based phylogenetic approaches. For the network analysis, we used the concatenated SNP data from the 50% missing dataset in the program SPLITSTREE v.4.16.1 with the Neighbor-Net algorithm [Bryant and Moulton, 2004, Huson, 1998]. To explore relationships between individuals and to test for monophyly among subpopulations, we concatenated the RAD loci and constructed a phylogeny using RAxML v8.2.10. For this, we expanded our 50% missing data dataset to include a broader representation of the PNW clade with samples from the Cascade Mountains, Columbia River, and Oregon (Tables 1.1 & S1.3). We used a GTR+GAMMA substitution model with 100 rapid bootstraps [Stamatakis, 2014]. Using the same expanded dataset but only including biallelic SNPs, we aimed to estimate divergence times using the multispecies coalescent model in the program SNAPP v1.5.0 [Bryant et al., 2012], implemented in BEAST2

v2.5.2 [Bouckaert et al., 2014]. To estimate the timing of diversification into the Puget Sound region, we calibrated the species tree using a secondary calibration for the age of the PNW clade of 100 kya [Bouzid et al., 2022]. We assigned a prior to calibrate the root of the species tree with a normal distribution, a mean = 100 kya, and a 95% confidence interval of ± 4 kya to accommodate estimation error. This analysis included three additional samples from Oregon, which were downloaded from the NCBI-SRA. We modified the input files for divergence dating in SNAPP using the snapp_prep scripts [Stange et al., 2018]. To decrease the computation time, we reduced the number of samples from the Puget Sound region to 11, each from unique localities (Table S1.3). We ran two separate analyses for 200,000 generations each (sampling every 50 generations) to check for convergence across independent runs. We then combined posterior distributions using LOGCOMBINER, and produced a maximum clade credibility (MCC) tree using TREEANNOTATOR after discarding the first 10% of samples as burn-in.

1.3.7 Demographic analyses.

To investigate population demographics, we analyzed samples from the Puget Sound region as a single population. We performed demographic analyses using the program MOMENTS [Jouganous et al., 2017] using one SNP sampled randomly from each locus (707 SNPs total) from the 50% missing data dataset. To maximize the number of segregating sites, we projected the data down to a smaller sample size ($N = 70$) using the program EASYSFS (https://github.com/isaacovercast/easySFS). We optimized four single-population demographic models using Python scripts developed in Portik et al. [2017]: 1) two-epoch model with instantaneous size change (two parameters): N_μ = ratio of contemporary to ancient population size and T = time in the past at which size change happened; 2) exponential growth model (two parameters): N_μ and T; 3) bottlegrowth with instantaneous size change followed by exponential growth (three parameters): N_μ, T, and $N_\mu B$ = ratio of population size after first change to ancient population size; 4) three epoch model with multiple population size changes (four parameters): N_μ, $N_\mu B$, T, and TB = duration of bottleneck. We performed four rounds of model optimization under each model with 50 replicates each and 25 maximum iterations. For each model, we used the parameters from the best-scoring replicate as starting values for the next round of optimization. After the final optimization, we used the replicate with the highest likelihood for each model to calculate AIC scores and perform model selection [Burnham et al., 2011]. For the top-ranked model, we conducted 100 replicate simulations to assess the goodness-of-fit of the model to the data. We tested the top-ranked model by comparing the empirical log-likelihood value to the values obtained from 100 parametric bootstrap replicates, with the expectation that the empirical value will fall within the range of simulated values. Finally, we obtained confidence intervals for parameters using bootstrapping (100 replicates) by re-sampling the SNP data with replacement, and then optimizing model parameters for each replicate using the same procedure described

above. We converted the unscaled population parameters to demographic terms as follows: the time parameter T used the equation $T = 2 * N_{ref} *$ generation time. We used a generation time of two years, as observed for the species [Jameson Jr and Allison, 1976]. To calculate N_{ref}, we used the equation $\theta/4\mu L$, where μ is the mutation rate and L is the number of loci multiplied by their length. θ was derived from the MOMENTS output and we used a generalized lizard μ of 7.7e-10 [Perry et al., 2018].

1.3.8 Morphological variation

To determine whether there are any patterns of morphological divergence in Puget Sound *S. occidentalis*, we collected morphometric and meristic data from adult individuals. We included most samples used in the genetic dataset, with additional samples from some localities (N = 80; Table S1.4). We included morphological traits which have been shown to be highly variable in *Sceloporus*, both across species and within populations [Smith, 1939, Putman et al., 2019]. In total, we collected 15 morphological characters with six morphometric measurements: snout-vent length (SVL), measured from the tip of the snout to the vent; tail length (TL), measured from vent to the end of the tail; head length (HL), measured from the parietal eye to the tip of the rostrum; head width (HW), measured at the widest part of the head; right and left longest toe length (RLL and LLL, respectively), measured from the base to the tip of the toe; right and left femur length (RFL and LFL, respectively), measured from the ventral midline to the distal part of the knee, and five meristic counts: left femoral pores (LFP); right femoral pores (RFP); left longest toe lamellae (LTL); right longest toe lamellae (RTL); medial scales (MS); and dorsal scales (DS).

To determine whether subpopulations within the Puget Sound show any morphometric variation, we performed multivariate statistics using a principal component analysis (PCA) and a linear discriminant analysis (LDA). We determined whether male and female individuals were significantly different by conducting a Mann-Whitney U test under the null assumption that the morphological distribution between the two groups are the same. The difference between males and females was statistically insignificant (p = 0.17), thus we analysed all individuals together. We made natural log transformations of all morphometric data. To account for allometric growth, we used R to size-correct the morphometric data by regressing each morphometric trait against SVL for all individuals and using the residuals in subsequent analyses (PCA and LDA) [Miller et al., 2021, R Core Team, 2021]. We normalized but did not size correct the discrete meristic data. We analyzed the normalized meristic and size-corrected and log-transformed morphometric data both separately and combined.

1.3.9 Natural History

From April 1999 through October 2018, ecological data was collected for 466 *Sceloporus occidentalis* individuals in the Puget Sound region. We collected the following data using visual encounter surveys: lizard body temperature, air

temperature, lighting conditions, and the microhabitat and substratum that the lizard was observed on. We collected lizard body temperature data using a cloacal thermometer and air temperature using various external thermometers. All data collected was combined with other ecological studies on the population [McTernan, 2017, Backus, 2016]. Ecological data were not collected for the same lizards used in the genetic study.

1.4 Results

1.4.1 Geographic distribution.

The geographic distribution of *Sceloporus occidentalis* is discontinuous throughout the Puget Sound region. The majority of sites sampled had subpopulations confined to small, geographically isolated stretches of habitat. All but two localities (Dewatto1 and Dewatto2; Figure 1.1) from which we found lizard assemblages were on coastlines, with most being south facing, presumably to maximize daily heat and sun exposure. The two Dewatto assemblages were found at deforested inland plots with high sun exposure (Figs. S1.1 & S1.2). The northernmost locality sampled is Chuckanut, WA; however, records indicate that this population is not naturally occurring and was transplanted from the Camano Island area [Brown, 1992]. Therefore, the northernmost naturally occurring locality is on Camano Island, although others may have been overlooked. In total, we detected 22 unique localities throughout the Puget Sound region (Figure 1.1).

1.4.2 Genomic differentiation.

Population structure analyses support multiple genetic clusters within the Puget Sound. For clarity in the following sections, we will refer to five subpopulations found in the Puget Sound region based primarily on the genomic data, while also considering their geographic distributions: Olympic Peninsula (OLY: Duckabush, Beckett Point, Anderson Island, Ketron Island), Puget Sound north (PUGn: Tulare Beach, Camano Island, Chuckanut, Spee-Bi-Dah), Puget Sound south (PUGs: Chambers Creek, Burien, Maury Island, Point Defiance), Kitsap Peninsula west (KITw: Belfair, Tahuya, Dewatto, Holly), and Kitsap Peninsula south (KITs). The PCA reveals at least four genetic groups within the Puget Sound region with most samples clustering with those from geographically proximate areas, except for the KITs group which spans the entire Puget Sound region (Fig. 1.2). The results from the ADMIXTURE analyses largely corroborate the results of the PCA (Figure 1.3). The best-supported cross-validation score for the 50% missing dataset supported a K-value of six; the best-supported K-value for the 0% missing data dataset was ten (Fig. S1.3). Despite the analyses not converging on the same "best" K-value, we use a K-value of five due to the consistency with the PCA results and the geographic pattern demonstrated (Fig. 1.2). When a K-value of four is used, samples from Beckett Point, Duck-

abush, and Ketron Island cluster with PUGs. When a K-value of five is used, Beckett point, Duckabush, and sometimes Ketron Island and Anderson Island comprise their own group, which is more consistent with the geographic proximity of the samples (Fig. 1.3). Intriguingly, as the K-value increases, samples from nearly all localities form separate clusters, thus enabling the identification of the specific locality many of the samples were taken from (Fig. 1.3B). In most cases, identifying the specific locality a sample originates from can also be determined from the DAPC using genomic data (Fig. S1.4). However, the unique localities comprising PUGn and KITw are difficult to distinguish using either ADMIXTURE with a K-value ≥ 10 or clusters from a DAPC.

Although the genomic data can be used to pinpoint the specific location of origin for most samples, the genetic diversity around the Puget Sound is, as expected, relatively low. The maximum genome-wide pairwise distance between any two genetic clusters is 0.089%, which is for the comparison between samples from opposite ends of the Puget Sound (Camano Island vs. Anderson Island. The minimum pairwise distance between any two genetic clusters is 0.030%, which is for the comparison between samples from Joemma (KITs) and Chambers Creek (PUGs); Table 1.3). Despite nominal pairwise distances, the F_{ST} values between subpopulations are high, for reasons discussed below (see Discussion). The highest F_{ST} value was between the PUGn and KITs groups (0.17), and lowest between the PUGs and OLY groups (0.065; Table 1.3). Despite the low genetic diversity and isolated regions from which we collected samples, the genomic data show no signals of first degree relatives, but second and third degree relatives are present (Fig. S1.5).

1.4.3 Phylogenetic analyses.

The phylogenetic analyses support the general patterns of genetic clustering found in the population structure analyses, wherein many samples are clustered according to their collection locality (Fig. S1.6). For example, the network analysis cluster samples by collection locality in the PUGs and OLY groups (Fig. 1.2B). The Camano Island samples are unique from the Tulare Beach and Spee-Bi-Dah samples — the latter two occur on the same stretch of beach and may have a continuous distribution. The Chuckanut population, which was established via human translocation, was derived from multiple source populations in the northern Puget Sound (Fig. 1.2B, Fig. S1.6). The samples from the KITs and KITw groups are far less structured in the network analyses, a pattern akin to the ADMIXTURE results.

The species tree topology estimated using SNAPP supports the monophyly of the Washington samples with strong support (posterior probability = 1.0; Fig. 1.4). The Washington clade diverged from the remainder of the PNW clade 17.21 kya (13.36–20.78 95% HPD), consistent with post-Pleistocene colonization as the Cordilleran Ice Sheet began to recede. The Puget Sound is weakly supported as sister to the North Cascades (posterior probability = 0.51), which share a most recent common ancestor (MRCA) 13.01 kya (10.13–15.96 95% HPD). This clade diverged from the Yakima population 14.50 kya (11.49–17.56

95% HPD), although this relationship is also weakly supported (posterior probability = 0.91). The divergence events in the Washington populations occurred rapidly, as reflected by their relatively short branch lengths (Fig. 1.4).

1.4.4 Demographic analyses.

Four demographic models were tested (two epoch, growth, bottlegrowth, and three epoch) to understand the population history of *S. occidentalis* in the Puget Sound. The differences in log-likelihood scores among the optimized models were small (< 0.2 units), and we ranked the models by their AIC scores (Table 1.4). The top-ranked model was the two epoch model (wAIC = 0.43), followed by the growth (wAIC = 0.36) and bottlegrowth (wAIC = 0.15) models (Table 1.4; Fig. S1.7). We present the results of the top two models (two epoch and growth models), which account for nearly 80% of the cumulative AIC. Both models infer a recent and substantial population size expansion, but they differ in the magnitude and timing (Table 1.5). The two epoch model infers a 25.83X population size increase at 10.15 kya (9.10–13.64 95% HPD) and the growth model supports a 16.71X expansion at 16.04 kya (13.95–24.40 95% HDP). The two epoch model indicates that there was a delay between the colonization of the Puget Sound region and the major expansion of the population; whereas the growth model indicates that the colonization and expansion of the population occurred simultaneously.

1.4.5 Morphological variation.

None of the morphological traits evaluated here can be used to clearly distinguish subpopulations from one another (Table S1.5). We present results for the combined analyses of the meristic and morphometric data (analyzing these characters separately produces similar results). Principal components one and two account for approximately half of the variation in the dataset (47.4%), with the heaviest loadings on PC1 being the right and left femur lengths, respectively; and the heaviest loading on PC2 being head width (Table S1.6). Nonetheless, the PCA is unable to distinguish any of the genetic subpopulations using morphometric and meristic data (Fig. 1.5). Despite lacking clear morphological distinction, the LDA demonstrates that the morphological data can be used to accurately identify which subpopulation a sample originates from with 61.5% accuracy (Table 1.6), suggesting that a complex combination of characters could be used to identify subpopulations with low accuracy.

1.4.6 Natural History

Using data collected from 466 lizards observed between 1999–2020, we provide updated information on the ecology and natural history of *S. occidentalis* in the Puget Sound region. The distribution of the Puget Sound population is primarily restricted to south-facing beaches and occasionally deforested areas that receive high sun exposure. Of the lizards recorded, 78.5% of individuals were in

open, sunlit area (no shading or light filtering). On coastlines, individuals are found on driftwood a majority of the time (57.7%), with most being found on shoreline that meets forested hills. Additionally, despite prior records stating that the populations hibernate from late September through mid May [Tsuji, 1988], we observed individuals basking from April through October. These activity windows are presumably dependent on annual climate patterns. Lastly, we observed body temperatures ranging from 19.8–37.8 °C, with a mean of 34.1 °C and median of 34.8 °C, consistent with body temperatures documented for the species [McGinnis, 1966].

1.5 Discussion

We investigated the geographic distribution, genetic and morphological variation, and phylogenetic and demographic history of *Sceloporus occidentalis* at the northern edge of the species range. The genomic data demonstrate that despite spanning a microgeographic scale and lacking any pronounced ecological differences, the Puget Sound region is comprised of multiple genetically identifiable subpopulations. Although genetic diversity is relatively low, in many cases there is sufficient genomic differentiation to identify the specific location from which a lizard originated.

Despite low genetic diversity (π) among subpopulations around the Puget Sound, F_{ST} values are relatively high (Table 1.3). The high F_{ST} values in the midst of low genomic differentiation is expected due to low heterozygosity in the Puget Sound population as a whole [Nei, 1973]. Considering the low genetic diversity observed in the group, the relatively high F_{ST} values denoted in Table 1.3 could indicate that minimal gene flow occurs among subpopulations in the Puget Sound region. This conclusion is also supported by the population structure and phylogenetic analyses, which indicate that many of the subpopulations examined are isolated from one another.

The post-glacial colonization of the Puget Sound by *S. occidentalis* combined with low levels of genetic diversity contribute to a lack of resolution on the optimal number of genomic subpopulations, and difficulty in identifying a consistent K-value. There is a large degree of difference between optimal K-value inferred for the 50% (K-value = 6) and 0% (K-value = 10) missing data datasets (Fig. S1.3). Inferring the true K-value for a population is a difficult procedure, and most models are expected to be somewhat inaccurate [Lawson et al., 2018]. We considered population structure models using differing amounts of missing data, and observed that increasing the number of populations in the model typically produced biologically realistic results well beyond any optimal K-value. In general, each incremental increase in K-value beyond the optimal value provided support for an additional unique sample location as distinct, suggesting that there is sufficient information in the SNP data to distinguish many of the sample locations. However, although we present results for a model with K-value = 5, we are not definitively stating that five genetic subpopulations occur within the region. Rather, given the data genomic data considered here, a K-value =

5 is a fairly conservative and biologically realistic estimate for the number of subpopulations. Substantially more loci are required to increase the resolution and confidence in the number of subpopulations in the Puget Sound region.

The phylogenetic and population demographic results are consistent with the hypothesis of a Holocene colonization of the Puget Sound region following the retreat of the Cordilleran Ice Sheet. Although the PNW clade split from the Sierra Nevada group approximately 100 kya [Bouzid et al., 2022], the opportunity for *S. occidentalis* to colonize the Puget Sound region was presumably limited until at least approximately 16 kya when the ice sheet began receding. The possibility remains that a population colonized the Puget Sound prior to the expansion of the ice sheet and was subsequently extirpated. Regardless, our results demonstrate that the colonization, or recolonization, of the Puget Sound region occurred approximately 13 kya and subsequently underwent a substantial population expansion (Fig. 1.4, Table 1.4). We expect that as the climate continues to warm, more habitat within the Puget Sound region could increase in suitability, thus promoting continued population expansion. However, although climate predictions may be favorable for *S. occidentalis* population expansion, the lack of suitable habitats may be prohibit further expansions. For example, extensive shoreline development [Johannessen et al., 2014] and urbanization surrounding the Seattle region has likely already limited and fragmented *S. occidentalis* habitats as it has done in the southern part of its range [Delaney et al., 2010]. Further research on the ability of *S. occidentalis* to colonize landscapes in the face of urbanization and shoreline development remains to be studied in this region.

By solely focusing on *S. occidentalis* in the Puget Sound region, we reveal a unique distribution wherein the group spans the majority of the region; however, the distribution is comprised of many isolated localities as opposed to being continuous. We suspect that the Camano Island locality is the northernmost naturally occurring location for the coastal PNW population. There are species records from farther north and west, but it is unclear whether these records represent introduced and/or extant populations [Farrell et al., 2020]. To our knowledge, *S. occidentalis* does not occur on the northernmost end of the west side of the Olympic Peninsula, as supported by other studies [Slater, 1963], yet one record exists [Owen, 1940]. If this record represents a natural population, it was likely extirpated. Additionally, in 2020, three records of *S. occidentalis* were made in British Columbia (B.C.), Canada, with two from the citizen-science platform iNaturalist and one from a scientific study [Farrell et al., 2020]. Older records also indicate the presence of *S. occidentalis* in B.C., but with no specific locality information [Van Denburgh, 1922]. We expect that these records represent either accidentally or purposefully translocated individuals, yet it remains unclear whether they have or will be able to establish sustainable populations. If more individuals are detected from the region, genomic testing could prove useful in determining their geographic origin.

Temperature can have a drastic effect on physiological function in *S. occidentalis* populations [Sinervo, 1990], which may provide an explanation for their fragmented distribution in the Puget Sound region. In addition to females dedi-

cating more resources to rearing offspring at higher latitudes [Sinervo, 1990], the northern populations also have substantially reduced physiological growth rates. In a laboratory-based experiment, hatchlings from Deschutes County, Oregon exposed to 34 °C versus 27 °C environments for 12 hours had a growth rate 1.4X greater. Given the same conditions but only 6 hours of exposure to the specified temperature, the growth rate increased to 1.7X that of the colder temperature [Sinervo, 1990]. Additionally, Oregon populations showed fewer hours of activity per day than those in California [Sinervo and Adolph, 1994]. The opportunity for Puget Sound subpopulations to have extra sunlight provided by inhabiting south facing beaches or full sun areas may be critical for their success. Further, the cool temperatures of the Puget Sound (average temperature for Seattle in August, the warmest month of the year, is 22 °C), may necessitate exposure to additional hours of warmth for survival. In our field surveys, we only detected two localities (Dewatto1 and Dewatto2, KITw group; Figure 1.1) that were not coastal, both of which were in deforested patches with full sun exposure for all hours of daylight. However, studies targeting both eastern and western Washington *S. occidentalis* populations have not found substantial variation in physiological function, despite a warmer climate for the eastern populations [McTernan, 2017, Powers et al., 2018]. Nonetheless, with the climate in the PNW expected to increase 0.1–0.6 °C per decade [Mote et al., 2008], it is likely that more habitat will become suitable for *S. occidentalis*, thus promoting favorable conditions for expansion.

Extensive shoreline development and urbanization could limit future population expansion of *S. occidentalis*, even as the climate becomes warmer and makes more habitat suitable. Though we are unaware of any urban or suburban *S. occidentalis* populations in the Puget Sound region, they are relatively common in habitats surrounding the metropolitan areas of San Francisco and Los Angeles. Increasing coastal development around the Puget Sound threatens to further reduce genetic diversity within subpopulations. Although patterns are highly taxon specific, urban fragmentation typically leads to increased genetic drift within fragmented populations and reduced gene flow among them [Miles et al., 2019, Bassitta et al., 2021]. This suggests that the future of the northern *S. occidentalis* populations could be contingent on a balance between increased suitable habitat with climate change and decreased access to or destruction of this habitat with anthropogenic development. The threat of coastal development is especially pertinent as it frequently involves shoreline armoring in the form of sea walls, revetments, and bulkheads. Armoring effectively limits the beaching of driftwood and logs, reduces the presence of beach wrack and its associated invertebrates (a food source for *S. occidentalis*), and removes the cover of riparian vegetation [Dethier et al., 2017]. Although no studies have directly examined impacts on lizard populations, armoring can have strong detrimental effects on species assemblages and abundance [Dugan et al., 2008, 2018]. Fortunately, armor removal can effectively restore these key elements of shoreline habitats [Lee et al., 2018, Toft et al., 2021].

Urban development may have already led to extirpation of local lizard assemblages, thus population reintroductions may become necessary. Historical

museum records indicate that *S. occidentalis* occurred one mile south of Lincoln Park, Seattle. We were unable to detect *S. occidentalis* in Lincoln Park. However, the locality from which the specimens were found is now developed with little to no natural habitat remaining, so it is unlikely that the population has persisted. As coastal development in the Puget Sound region continues, more isolated groups may, or already have, become threatened. As such, expanded studies on the ecology and population demographics of the species in the region will prove useful for potential reintroductions. Our study provides important genetic information for guiding the selection of source populations to be used for reintroductions by demonstrating that unique alleles are present in many subpopulations across the Puget Sound region. One approach could be to translocate gravid females from a locality that shares the distinct alleles associated with the extirpated group. Another approach could utilize translocated individuals from multiple subpopulations to increase the genetic diversity of the reintroduced group, considering the low genetic diversity within any given subpopulations. Lastly, and especially pertinent for locally extirpated groups with an unknown genetic heritage (e.g., Lincoln Park), reintroduction efforts could utilize individuals from Chuckanut, considering that the assemblage at this locality is known to be human introduced.

Although a warming climate may lead to natural expansion of the Puget Sound population, occupying the northernmost extent of the distributional range exposes them to threats that the remainder of the species does not face. We provide evidence for a population that has undergone a relatively recent and expansive growth, which could indicate continued expansion as the climate become more favorable. However, multiple compounding factors limit this success. The pattern of colonization limited to highly specific stretches of south-facing coastal habitats with high sun exposure, including islands and peninsulas, has likely limited gene flow and accelerated genetic drift producing the genetic patterns described herein. Further, geographic discontinuity coupled with expanded urban development may impede the group's ability to take advantage of the rapidly warming climate. The northernmost *S. occidentalis* population would be expected to continue a northward expansion, but the adaptive potential of the group may not be great enough to keep up with the rampant rate of anthropogenic change.

Data availability DNA sequence data generated for this study are deposited at the NCBI Sequence Read Archive (SRA); accession numbers SAMN20963029–20963135. Datasets and R scripts used in the study are available on Dryad (doi:10.5061/dryad.70rxwdc0f).

Table 1.1: Sample locations and sample sizes. Detailed voucher specimen information is provided in Table S1.1.

Location	Samples
Puget Sound Region	
Anderson Island	4
Beckett	4
Belfair	2
Burien	3
Camano Island	4
Cambers Creek	6
Chuckanut	4
Dewatto	6
Duckabush	4
Holly	4
Joemma	4
Ketron Island	4
Maury Island	5
Point Defiance	3
Port Townsend	4
SpeeBiDah	2
Tahuya	4
Tulare	7
Wauna	4
Eastern Washington	
Columbia River	4
Leavenworth	1
Swakane Canyon	5
Yakima	2
Oregon	
Skunk Hollow	2
Selma	2
Shaniko	2

Table 1.2: The number of SNPs obtained from the reference-based assembly of 78 samples of *Sceloporus occidentalis* from the Puget Sound Region. Allowing for more missing data (50%) results in more SNPs compared to no missing data (0%).

Chromosome	50% missing data	0% missing data
chr1	322	62
chr2	291	60
chr3	218	37
chr4	223	48
chr5	143	22
chr6	170	35
chr7	64	15
chr8	51	13
chr9	47	12
chr10	8	2
chr11	17	3
Total SNPs	1,554	309

Table 1.3: F_{ST} values between genetic groups (upper diagonal) and pairwise distances (lower diagonal) for *Sceloporus occidentalis* in the Puget Sound region. The lowest pairwise distance between any two samples from a given population are shown, rather than an average of all samples from a given locality. The five groups comprising the Puget Sound region are: OLY = Olympic Peninsula, KITs = Kitsap Peninsula south, KITw = Kitsap Peninsula west, PUGn = Puget Sound north, and PUGs = Puget Sound south.

	OLY	**PUGn**	**KITs**	**PUGs**	**KITw**
OLY		0.14	0.12	0.065	0.12
PUGn	7.6e-4		0.17	0.099	0.16
KITs	6.2e-4	5.2e-4		0.096	0.17
PUGs	4.7e-4	4.1e-4	3.0e-4		0.092
KITw	6.8e-4	5.5e-4	4.0e-4	3.2e-4	

Table 1.4: Demographic models ranked by AIC scores.

Model	**Parameters**	**log-likelihood**	**AIC**	**ΔAIC**	**RelativeL**	**wAIC**	**cumulativeAIC**
Two epoch	2	-35.38	74.76	0.00	1.00	0.43	0.43
Growth	2	-35.54	75.08	0.32	0.85	0.36	0.79
Bottlegrowth	3	-35.43	76.86	2.10	0.35	0.15	0.94
Three epoch	4	-35.32	78.64	3.88	0.14	0.06	1.00

Table 1.5: The top-ranked demographic models and their optimized model parameter estimates. The 95% confidence intervals were obtained using non-parametric bootstrapping. The time values have been converted into thousands of years using the equation provided in the Methods section.

Model	**Population Size Change (95% CI)**	**Time (95% CI)**
Two epoch	25.83X (15.49–46.83)	10.15 (9.10–13.64)
Growth	16.71X (16.35–44.67)	16.04 (13.95–24.40)

Table 1.6: Accuracy of correctly assigning a sample to its genetic group (assuming K-value = 5) using morphometric and meristic data in a linear discriminant analysis. The five groups tested from the Puget Sound region are: OLY = Olympic Peninsula, KITs = Kitsap Peninsula south, KITw = Kitsap Peninsula west, PUGn = Puget Sound north, and PUGs = Puget Sound south.

	OLY	**KITs**	**PUGs**	**KITw**	**PUGn**
OLY	4	0	1	1	0
KITs	0	6	2	0	1
PUGs	2	5	26	6	3
KITw	1	1	3	6	0
PUGn	0	0	2	0	6
Accuracy	0.57	0.50	0.76	0.46	0.60

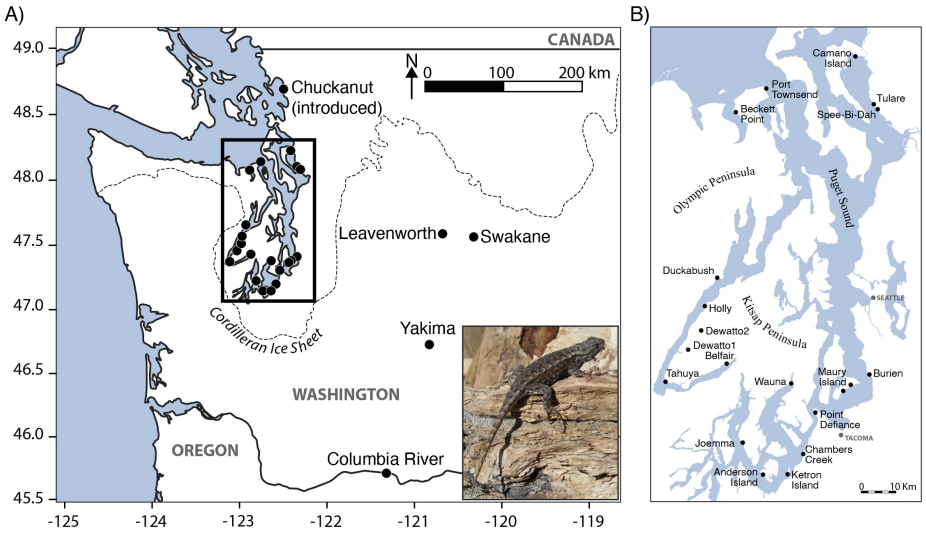

Figure 1.1: Map of study area. A) *Sceloporus occidentalis* sampled in Washington with the dashed line indicating the maximum extent of the Cordilleran Ice Sheet (16 kya; [Riedel, 2017]). Inset figure shows an individual of *S. occidentalis* on driftwood along the coastline of Tulare Beach. B) Detailed sampling map of the Puget Sound region. Sampled locations are labeled with black dots. Map generated using QGIS [QGIS, 2015].

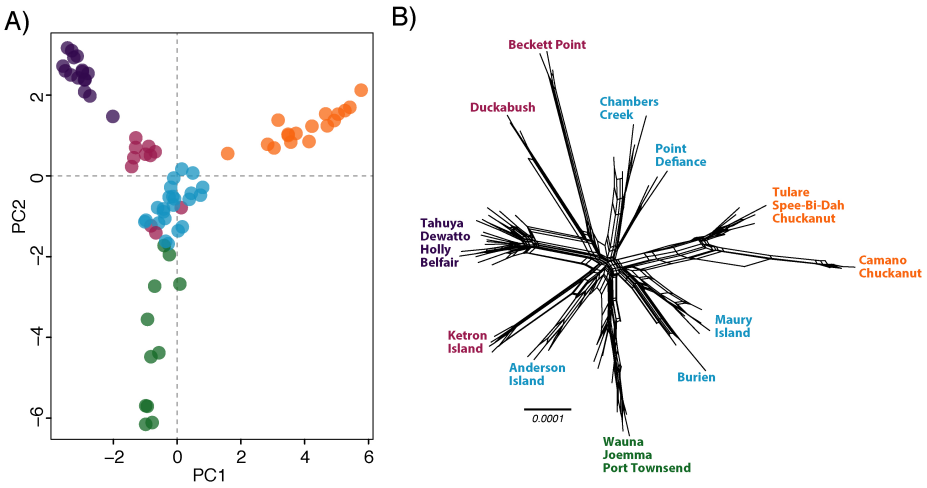

Figure 1.2: Genetic diversity of *Sceloporus occidentalis* in Western Washington. A) Principal components analysis (PCA) of genetic variation using 3,491 loci in the R package ADEGENET. B) Network analysis using the same dataset as the PCA demonstrates that samples from the same locality form distinct genetic clusters. Colors correspond to the population assignments from ADMIXTURE assuming K-value = 5

.

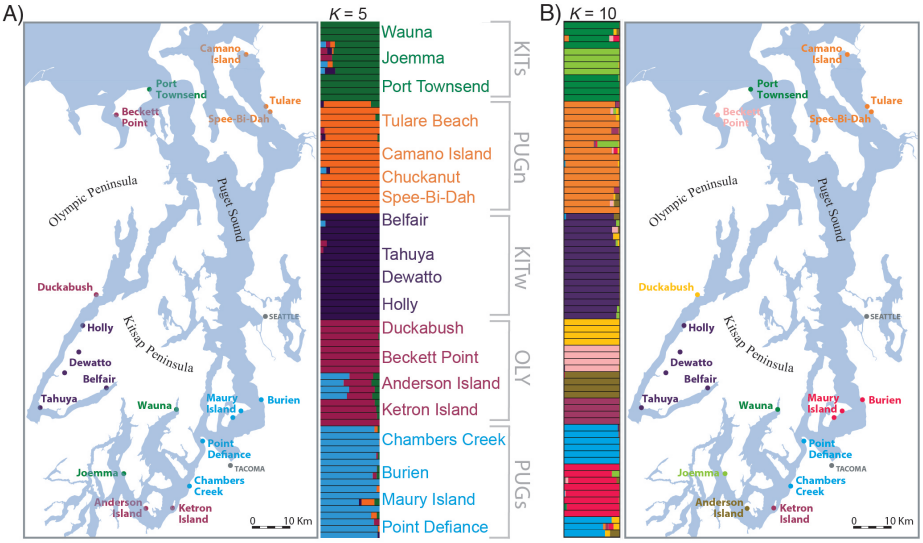

Figure 1.3: Geographic distribution of populations based on ADMIXTURE analyses assuming a K-value = 5 (A) and 10 (B). Each bar represents an individual, and the colors indicate the admixture proportions. Results are shown for the 50% missing data dataset (3,491 loci). Subpopulation names are assigned using a K-value = 5. Kitsap Peninsula south = KITs; Puget Sound north = PUGn; Kitsap Peninsula west KITw; Olympic Peninsula = OLY; and Puget Sound south = PUGs. Maps generated using QGIS [QGIS, 2015].

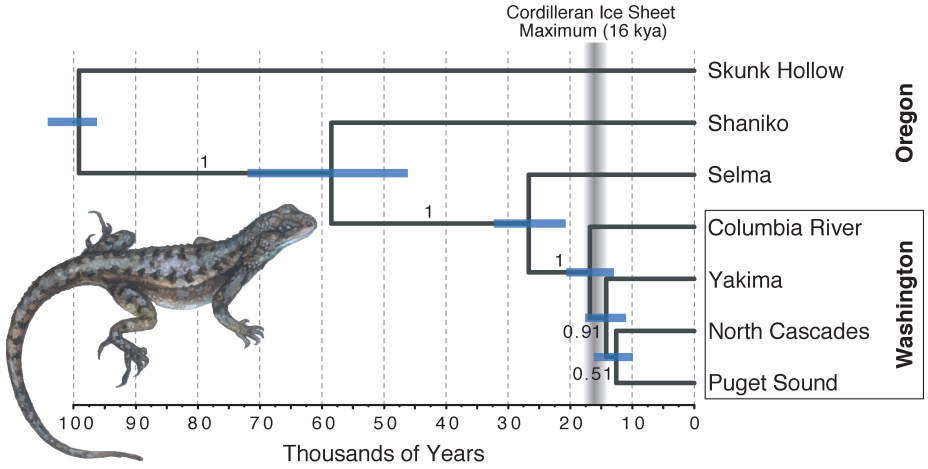

Figure 1.4: Species tree analysis of *Sceloporus occidentalis* from the Pacific Northwest based on a coalescent analysis of 953 biallelic SNPs. The species tree was calibrated assuming a root divergence time for the Pacific Northwest clade of 100 kya (± 4 kya; [Bouzid et al., 2022]). Posterior probability values are shown on branches, and node error bars show 95% highest posterior density (HPD) of divergence times. The timing of deglaciation of the Cordilleran Ice Sheet at 16 kya is shown with a vertical bar.

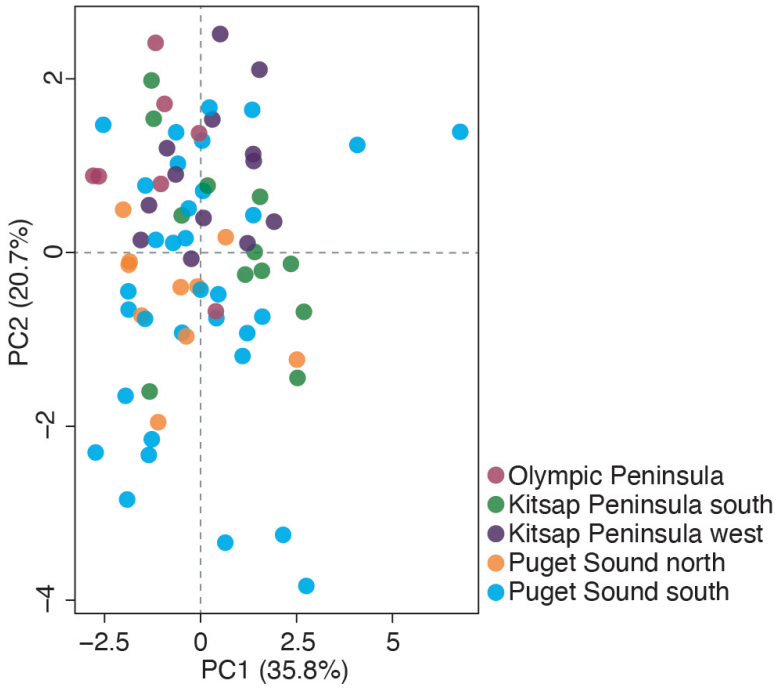

Figure 1.5: Principal component analysis of morphological data using size-corrected and log-transformed morphometric data and meristic counts for sub-populations from the Puget Sound region in the R package TIDYR. Colors correspond to the population assignments from ADMIXTURE assuming K-value = 5

.

Chapter 2

Gene flow spans ecoregions in the Western Banded Gecko

Hayden R. Davis, Atinuke Bandele, Dean H. Leavitt, Julio A. Lemos-Espinal, & Adam D. Leaché

2.1 Abstract

Desert ecosystems of southwestern North America harbor high levels of endemism and genetic diversity. Previous studies have found strong selection and well-defined phylogeographic patterns correlated with desert regions, but did these lineages evolve in isolation or along gradients with gene flow? In this study, we used genome wide SNP data to explore the genetic diversity of the gecko *Coleonyx variegatus* which spans the Sonora and Mojave deserts, and the entirety of the Baja Peninsula. The broad geographic distribution of *C. variegatus* across these unique ecosystems provides an opportunity to determine whether populations are primarily restricted to distinctive desert ecoregions, or if genetic diversity is driven by other factors. We apply population structuring analyses to identify population boundaries and then use the MSC-M model to directly test for gene flow among populations. Further, we apply multispecies coalescent methods to infer a species tree, which informs novel relationships and divergence dates among populations of *C. variegatus*. These results add to the growing body of phylogeographic literature for taxa spanning the region of focus. Combining our results with those from multiple sympatric species can help inform shared patterns of diversification across the broad, diverse region of southwestern North America.

2.2 Introduction

The deserts of the southwest United States and Northern Mexico contain many unique ecoregions with complex biogeographic and climatic histories [Riddle and Hafner, 2006]. Ecoregions are ecologically and geographically defined regions that are a smaller subset of a biological realm. These unique environments have been shown to correlate with the diversification of large communities of desert adapted organisms, although the phylogeographic patterns linking these ecoregions often differ, as inferred from phylogenetic studies of species [Wood et al., 2013, Riddle et al., 2000, Mulcahy, 2008, Buchalski et al., 2016, Massatti and Knowles, 2020, Sumarli et al., 2023, Gottscho et al., 2017, Pavón-Vázquez et al., 2024, Crews and Hedin, 2006]. We aim to explore these distinct ecoregions to determine their roles, if any, in the formation of barriers to reproductive isolation, as measured by gene flow between populations.

Increased accessibility to genomic data has broadened the opportunity to explore lineage diversification and demographic histories for non-model organisms with increased precision. Specifically, these large datasets enable the detection of gene flow and reticulation events, which have proven paramount for understanding species relationships and diversification patterns [Leaché et al., 2024, Thawornwattana et al., 2023, Ji et al., 2023, Barley et al., 2022, Burbrink et al., 2022, Portik et al., 2017]. In this study, we apply the multispecies coalescent model with migration (MSC-M) to explore the diversification of the Western Banded Gecko (*Coleonyx variegatus* (Baird, 1858)) in the deserts of the southwestern US and northern Mexico. In doing so, we aim to obtain estimates of gene flow between populations to better understand how the species has diversified across desert ecoregions. These geckos lack definitive morphological features among populations and multiple subspecies despite defined genetic structure, thus we sought to explore other factors that may be driving their diversification. *Coleonyx variegatus* is distributed across seven major ecoregions, making it an ideal system for testing the putative role of ecoregions as barriers to gene flow. In specific, the species spans: 1) Sonora Desert; 2) Mojave Desert; 3) Madrean Archipelago; 4) Eastern Sierra Madre; 5) California Coastal Sage; 6) Baja California Desert; and 7) Los Cabos Plains (Fig. 2.1).

Traditionally, *Coleonyx variegatus* contained seven subspecies based on morphology [Klauber, 1945, Dixon, 1970]. These subspecies are roughly distributed as follows: *C. v. variegatus* — Mojave and western half of US Sonora desert; *C. v. abbotti* — Baja California and the southern tip of California; *C. v. peninsularis* — Baja California Sur; *C. v. sonoriensis* — southern Sonora Desert in Mexico; *C. v. bogerti* — eastern half of Sonora Desert; *C. v. utahensis* — northern extent of the range in Utah; and *C. v. slevini* — Santa Inés Island, Mexico [Dixon, 1970]. Leavitt et al. [2020] conducted a phylogeographic study using mitochondrial DNA (mtDNA) and four nuclear loci to identify 17 clades within *C. variegatus*, many of which have distributions that are loosely correlated with ecoregions, although this was not a focus of their study. Considering the broad intraspecific diversity within *C. variegatus*, we seek to use genome-wide data to understand population diversification and whether that diversification cor-

relates with ecoregions. There is reason to suspect that nuclear data could support morphological subspecies and/or ecoregion diversification models. For example, a recent study using single nucleotide polymorphism (SNP) data on the congeneric species *C. elegans* found distinct genetic barriers that correspond with ecoregions, leading to the elevation of *C. nemoralis* from subspecies to full species [Butler et al., 2023]. In this case, the ecoregion boundaries are primarily defined by highlands dividing the two regions, however, at the Isthmus of Tehuantepec, no modern barrier to gene flow is detected despite a lack of introgression [Butler et al., 2023].

Using genome-wide SNP data and extensive sampling of *C. variegatus* from across their geographic distribution, we investigate the drivers and patterns of population diversification. Specifically, we address three primary questions: 1) Do population boundaries coincide with ecoregion boundaries?; 2) How much and in what geographic regions is gene flow occurring between populations?; and 3) What are the divergence times for populations of *C. variegatus*?

2.3 Methods

2.3.1 Sample Collection

In total, we obtained 189 tissue samples for a our study, including 162 samples used in a recent study and 27 new specimens obtained through fieldwork [Leavitt et al., 2020]. We also included nine samples of *C. brevis* as an outgroup. We conducted fieldwork in Arizona between the months of June and August 2021, and collected 31 samples from unique localities. We collected one specimen from Washington County, Utah in May 2022 and two samples of *C. brevis* from Durango, Mexico in June 2022. Finally, we collected seven specimens from the minimally sampled Imperial County, CA in September 2022. All animal research was approved by the University of Washington Institutional Animal Care and Use Committee (IACUC: 4367-03). Tissue samples (liver) were flash frozen in liquid nitrogen and stored long-term without solution at -80 °C. The two *C. brevis* samples were stored in pure ethanol (EtOH) immediately after excision and are stored long-term in EtOH at -80 °C. All new samples are accessioned at the Burke Museum of Natural History and Culture (UWBM).

2.3.2 Genomic Data Collection and Bioinformatics

We extracted genomic DNA from tissue samples using a salt extraction protocol [Aljanabi and Martinez, 1997], and conducted double digest restriction-site associated DNA sequencing (ddRADseq; [Peterson et al., 2012]). We double-digested each sample using the digestion enzymes *SbfI* and *MspI* in CutSmart Buffer (New England Biolabs) for 7 hours at 37 °C. For fragment purification, we used Sera-Mag SpeedBeads. We then ligated eight distinct barcodes with unique molecular identifiers (UMI) included to the cut sites of the fragmented DNA. After barcode ligation, we size-selected each library between 415 and 515

base pairs (bp) on a Blue Pippin Prep size fractionator (Sage Science). For final library amplification, we used Phusion Hi-Fidelity DNA Polymerase and Illumina index primers. We determined the concentration and size distribution of each pool using an Agilent 2200 TapeStation. Lastly, we sent the quantified libraries to QB3-Berkeley Genomics, UC Berkeley for qPCR and sequencing on two Illumina NovaSeq 6000 lanes (100-bp, single-end reads; 28 pools containing 8 samples). The demultiplexed data are deposited in the NCBI Sequence Read Archive (SRA; BioProject ID: XX; Table XX).

We demultiplexed and assembled the data using IPYRAD [Eaton and Overcast, 2020]. We applied a sequence similarity threshold of 85% to cluster reads within samples and loci between samples. We removed consensus sequences with low coverage (< 6 reads), excessive undetermined or heterozygous sites ($>$ 5%), too many alleles for a sample (> 2 for diploids), or an excess of shared heterozygosity among samples (paralog filter = 0.5). We generated one assembly with all individuals, including outgroups requiring a minimum 50% of individuals to share each locus. This assembly contains 198 individuals and 895 loci (coleonyx_all). From this assembly, we used the branching function in IPYRAD to remove outgroup samples and any samples with $<$ 500,000 raw reads. The final post-filtered assembly (coleonyx_assembly) includes 176 individuals and 1,483 loci (Table 2.1). From this assembly, we created additional branches for conducting species tree inference and gene flow estimation using BPP (details below).

2.3.3 Population Detection

To determine if the diversification of *Coleonyx variegatus* is linked to ecoregion boundaries, we needed to identify populations and their geographic ranges. Considering the difficulty associated with accurately detecting an optimal number of populations and the importance of a well-informed estimation for the purposes of this study, we analysed the coleonyx_assembly to identify population boundaries using three independent approaches. First, we analyzed the data by performing a principal components analysis (PCA) using the R package ADEGENET [Jombart, 2008]. Second, we used SplitsTree4 to construct a NeighborNet network to identify population clusters [Huson, 1998, Huson et al., 2008]. Lastly, we assessed population structure and estimated individual levels of admixture using ADMIXTURE [Alexander and Lange, 2011]. For ADMIXTURE, we filtered our dataset to one SNP per locus and applied a minor allele frequency of 0.05 using VCFtools. We conducted ten independent runs for each K-value ranging from 1–12, and considered the K-value with the lowest cross-validation (CV) score as the optimal population model.

To obtain general diversity metrics for population comparisons, we estimated the average genetic pairwise distance (p-distance) using MEGA X, and weighted F_{ST} values using VCFtools [Danecek et al., 2011, Stecher et al., 2020]. We also investigated whether the genetic diversity within *C. variegatus* correlates with geographic distance by testing for isolation-by-distance (IBD) using ADEGENET [Jombart, 2008]. We did this using two approaches: first, we tested for IBD

using all samples across the full range of the species; and second, we filtered the dataset to test for IBD within each identified population (detailed in the Results section).

2.3.4 Species Tree Estimation

We estimated a species tree for *Coleonyx variegatus* populations using the multispecies coalescent model in BPP v.4.7.0 [Yang, 2015]. Because BPP assumes no migration, we reduced the dataset to only include four samples per population by selecting individuals with no signatures of admixture. A description of the populations is provided in Results section. This reduced dataset contained 32 individuals and 8,675 loci (Table 2.1). To infer the tree, we set the θ and τ priors using an inverse gamma distribution: $\theta \sim (3, 0.02)$ with a mean = 0.01 and $\tau \sim (3, 0.05)$ with a mean of 0.025. We conducted three independent runs for 100,000 generations with a sample frequency of one.

Next, we used BPP to estimate divergence times (τ) and population sizes (θ) using the A00 option (fixed species tree option). For this analysis, we assembled a new dataset that combined the same samples used to infer the species tree together with one randomly selected admixed sample from each specific locality containing admixed samples. This assembly included 65 individuals and 4,117 loci (Table 2.1). We used the maximum pairwise distance estimated from MEGA to set a well-informed τ prior. To begin the analysis, we used inverse gamma priors for τ and θ: $\theta \sim (3, 0.012)$ with a mean of mean = 0.006 and $\tau \sim (3, 0.015)$ with a mean = 0.0075. We ran two independent analyses for 200,000 generations with a sample frequency of two. We assessed convergence by comparing the posterior probabilities for all of the parameters estimated from each of the three runs using TRACER v.1.7.1. For both the A01 and A00 analysis, we used the phase function within BPP to integrate over phase uncertainty [Huang et al., 2022].

To explore phylogenetic relationships within populations and assess differences between the species tree and a concatenated gene tree approach, we concatenated the RAD loci to estimated a maximum likelihood (ML) tree. To do so, we used the full dataset (coleonyx_all), which includes *C. brevis* samples as an outgroup. To infer the tree, we used IQ-TREE [Nguyen et al., 2015]. For model parameters, we determined the best-fit evolutionary model using MODELFINDER [Kalyaanamoorthy et al., 2017] and applied 1000 ultrafast bootstraps to assess node support [Hoang et al., 2018].

2.3.5 Divergence Time Estimation

To estimate divergence times, we converted species tree branch lengths into years using a germline mutation rate established for *Coleonyx brevis*. The rate of 3.17E-9 was established using whole genome sequencing of a known pedigree [Bergeron et al., 2023] and dividing the per-generation mutation rate by the parental age Bergeron et al. [2023]. Although Bergeron et al. [2023] suggested that this rate was underestimated based on their modeled rate of 1.96E10-8, we

are using the former because it is based on a two year generation time for the species, whereas the latter assumed a biologically unrealistic generation time of 3–4 months. We used the branch lengths (τ) from the BPP A00 analysis to estimate divergence times, which is a model that assumes no gene flow among groups. We acknowledge that divergence times can be overestimated if gene flow is ignored [Tiley et al., 2023], but analyzing the full dataset under the MSC-M model was computationally intractable.

2.3.6 Gene Flow Quantification

As a first step in exploring gene flow between *C. variegatus* populations, we tested for introgression using the D-statistic. To do so, we used the program DSUITE [Malinsky et al., 2018], and applied the DQUARTETS model, which calculates the D and F_4 statistics for all quartets of lineages in the dataset [Patterson et al., 2012]. We specified the relationships among populations by inputting the species tree estimated using BPP.

To measure gene flow across population boundaries, and to determine the relative strength of the barriers, we quantified gene flow using the MSC-M model in BPP [Flouri et al., 2023]. To reduce the computational burden of testing for gene flow between populations across the entire species tree, we used the branch function in IPYRAD to create assemblies that only included the two populations found on opposite sides of each barrier, and for ecoregions with more than one population within them. For example, in Table 2.1, the assembly bcn_bcs only includes samples belonging to the BCN and BCS groups, which are separated by the mid-peninsula of Baja California. For these assemblies, we imposed a more stringent missing data threshold that required 80% of individuals to have data at each locus.

With a fixed species tree and precisely estimated τ and θ values (Table 2.5), we quantified migration among current populations using the MSC-M (Figure 2.1). The MSC-M measures continuous migration over the entire history of a population without distinguishing between ancient and recent gene flow [Jiao et al., 2021]. The MSC-M measures the migration rate as $M = m\mathrm{N}$; M is measured in the expected number of migrants from the donor population to the recipient population per generation; N is the effective population size of the donor population, and m is the proportion of immigrants in the recipient population from the donor population every generation under the forward-in-time perspective [Flouri et al., 2023]. For each analysis, we assigned a gamma prior for M: $M \sim (2, 20)$ with a mean of 0.1 (e.g., mean of 10% migration per generation). We ran each model two times with 100,000 samples and a sample frequency of two. We checked for convergence and combined the posteriors across runs using TRACER v.1.7.1. If the resulting migration estimate or the 95% confidence intervals included zero, we assumed no gene flow between those populations.

Lastly, to evaluate the sensitivity of the MSC-M model to the migration prior, we reran our MSC-M models using a ten-fold smaller prior on the migration rate (all other analysis settings were held constant). Specifically, we

decreased the mean of the gamma prior by an order of magnitude: $M \sim$ (2, 200) with a mean of 0.01 (e.g., 1% migration per generation). We ran each model two times with the same sampling as above, and checked for convergence using TRACER.

2.4 Results

2.4.1 Population Detection

Considering the ADMIXTURE, PCA, and SPLITSTREE analyses, we are confident that these data can partition *C. variegatus* into eight biologically-relevant populations (Fig. 2.2). The composition of these populations are as follows: 1) Death Valley (DV); 2) Mojave Desert (MO); 3) Salton Trough (ST); 4) Arizona Mountains (AZM); 5) North Sonora Desert (NS); 6) South Sonora Desert(SS); 7) Baja California Norte (BCN); and 8) Baja California Sur (BCS).

Although we selected a model with $K = 8$ populations, CV scores from the ADMIXTURE analysis differ minimally between K-values of 6–12 (Fig. S2.1). Models with K-values > 9 consider highly admixed samples from the west end of the AZM group to be a biologically distinct genetic group, but we expect that this grouping stems from difficulty clustering those samples (Fig. S2.2). The PCA clusters five groups with minimal overlap, with NS, ST, and AZM being less distinguishable (Fig. 2.2B). The network analysis clusters seven groups, but the AZM and NS populations do not distinctly parse from one another (Fig. 2.2C). We assumed a model with $K = 8$ populations for use in subsequent analyses, as it is both well supported and biologically realistic.

Diversity metrics for *C. variegatus* highlight relatively low levels of genetic diversity and varying levels of fixation. The highest average p-distance between any two populations is 4.94E^{-3} (BCS vs. SS), and the lowest is 1.81E^{-3} (NS vs. ST; Table 2.2). F_{ST} values indicate that the population with the highest level of differentiation is BCS with values ranging from 0.514–0.595. The ST population is the least differentiated with F_{ST} values as low as 0.143 (Table 2.2).

Tests for IBD using all *C. variegatus* samples indicate IBD plays a role in the diversification of the species as a whole, whereas tests for IBD are not significant for most populations. Using the full dataset, a portion of the genomic diversity detected within the species is driven by geographic distance ($p = 0.001$; Fig. 2.3). When split into populations, the AZM group has support for IBD ($p = 0.028$), whereas none of the remaining populations do (p-values between 0.187–0.793; Fig. 2.3).

2.4.2 Species Tree Estimation

The species tree infers novel population relationships within *Coleonyx variegatus*. The species tree supports three major clades, corresponding to the Sonora Desert (AZM, NS, and ST), Mojave Desert (MO and DV), and Mexico (BCN,

BCS, and SS) populations. The Sonora and Mojave Desert clades are well supported (posterior probability (pp) = 1), but the Mexico clade lacks support due to incomplete lineage sorting (ILS; Fig. 2.5). The phylogenetic position of the SS populations lacks consensus across analyses, with all three runs lacking support for the relationship between the SS population and the remaining groups. Of the 300,000 trees inferred across three independent runs, 137,884 trees (45.96%) support a Mexico clade: ((BCN, BCS), SS); 136,288 (45.43%) support the SS population as an outgroup to all *C. variegatus* populations ((All), SS); and the remaining trees support the SS population as sister to the clades in the Mojave and Sonora Deserts (((Sonora, Mojave), SS), (BCN, BCS)). We used the tree that forms a Mexico clade as the topology for estimating τ and θ in the A00 analysis.

Within the three major clades, the majority of relationships have strong support. The Mojave clade contains MO and DV with strong support ($pp = 1$). The Sonora clade has strong support for the relationship between ST and the AZM + NS clade ($pp = 1$), but relationship between AZM and NS is weakly supported ($pp = 0.79$). Despite the weak support, a high majority of trees place AZM and NS as sister to one another. The BCS and BCN populations consistently form a well-supported clade ($pp = 1$).

The phylogeny inferred using concatenation lacks concordance with the species tree at most nodes. Neither a Mojave nor Sonora clade are recovered as monophyletic using the concatenated approach (Fig. S2.3). Of the populations within these groups, DV is the only population that has strong support for being monophyletic. All remaining Mojave and Sonora populations are weakly supported and form paraphyletic relationships. The two approaches are concordant in inferring a Mexico clade comprised of BCN, BCS, and SS, although this relationship is only well-supported in the concatenated tree (ultrafast bootstrap $\geq$ 95; Fig. S2.3).

2.4.3 Divergence Time Estimation

The divergence times estimated for *Coleonyx variegatus* support more recent population divergences than those estimated in previous studies [Leavitt et al., 2020]. Divergence time estimation using a germline mutation rate places an initial date of population diversification within the species at 561,514 years with relatively narrow confidence intervals (95% highest posterior density (HPD)) of 542,586–577,287 years. The divergence of the Mexico clade is estimated to have the same divergence time, as ILS prevents resolving the placement of SS with respect to the BCN and BCS populations with our current dataset. The Sonora and Mojave clades diverged from one another 188,012 (178,548–203,785) years ago. Subsequently, the Sonora population is estimated to have diverged 171,608 (161,829–179,810) years ago, and the most recent divergence is the Mojave clade at 137,223 (125,867–147,003) years (Fig. 2.5; Table 2.4).

2.4.4 Gene Flow Quantification

Introgression tests using the D statistic found 16 signals of introgession that are statistically significant with Z-scores ≥ 3.0.

We used the MSC-M model in BPP to test for migration in 11 population comparisons. Each gene flow model supported migration, and the migration prior had minimal influence on the posterior estimates (Figs. 2.1, 2.4); Table 2.5). The population pairs with the highest proportions of bidirectional gene flow ($M \geq 0.5$) under the 0.1 migration prior are: AZM $\leftrightarrow$ MO, AZM $\leftrightarrow$ NS, AZM $\leftrightarrow$ ST, and ST $\leftrightarrow$ NS. The populations occurring primarily in Mexico (BCN, BCS, and SS) have the lowest gene flow among all populations of *C. variegatus*, with $M \leq 0.134$ for all models tested (Table 2.5). Unexpectedly, the MO $\leftrightarrow$ DV model supports less gene flow ($M_{12} = 0.258$, $M_{21} = 0.245$) than the levels detected between any two adjacent populations in the US, despite occurring within the same ecoregion (Table 2.5; Fig. 2.4).

Varying the migration rate prior from 0.10 (10%) to 0.010 (1%) had minimal impact on the posterior estimation of M. Of the 22 gene flow models tested (11 bidirectional tests), seven of the mean posterior values for the 10% model are outside of the 95% HPD intervals of the 1% model (Table 2.5; Fig. 2.4). However, the DV $\rightarrow$ MO model has a mean M value (0.245) minimally outside of the 95% confidence interval (high end of the interval = 0.240). Only one population pair (ST $\leftrightarrow$ SS) has non-overlapping posterior estimates for both directions of gene flow. Importantly, none of the models using either prior included zero within the 95% HPD interval, thus supporting the finding no populations are fully reproductively isolated (Table 2.5; Fig. 2.4). These results suggest that our RAD data have enough information to overcome poorly fit priors to converge on the same or similar estimates of M using the MSC-M model.

2.5 Discussion

We tested whether genetically distinct populations of *Coleonyx variegatus* are correlated with specific ecoregions in the southwestern US and northern Mexico. We find most of the populations have geographical boundaries that correspond with the edges of the ecoregions, with the exception of the AZM and BCN populations that extend well beyond the boundaries of their respective ecoregions. Further, we find that regions with higher proportions of admixed individuals tend to occur on the edges of ecoregions. Not all populations from adjacent ecoregions have strong signatures of admixture and gene flow, but most of the gene flow occurs near ecoregions borders. The highest gene flow proportions occur in the intervening area among the AZM, MO, NS, and ST, which is near the junction of the Mojave and Sonora deserts (Figs. 2.2 & S2.2).

2.5.1 Species Tree Estimation and Time-Calibration

The species trees inferred herein differs from both the mtDNA and nuclear topologies inferred in Leavitt et al. [2020], although direct comparisons to the nuclear tree can not be made due to a different number of taxonomic groups used to infer the tree. The most discordant relationships between the species tree inferred herein, and those in Leavitt et al. [2020] stem from the populations comprising the Mexico clade. The Mexico clade recovered in the species tree (SS, BCN, and BCS) is not recovered with the mtDNA or concatenated nuclear loci tree in Leavitt et al. [2020]. The mtDNA approach places BCS as sister to all *C. variegatus*, and the nuclear tree forms a Baja clade (BCN + BCS) as an outgroup to all *C. variegatus* with SS being sister to the Mojave and Sonora clades. Further, both the mtDNA and nuclear approaches form paraphyletic relationships among the Mojave samples. With high gene flow among populations and patterns of ILS, we argue that more data-rich and MSC-based approaches provide the best evolutionary hypothesis for this group.

The diversification dates estimated using genome-based mutation rates are substantially younger than those inferred with mtDNA-based mutation rates. Using a mtDNA-based substitution rate of 1.55% per million years, the divergence between BCN and BCS is estimated at 4.22 mya and 5.14 mya for the divergence between SS and the remaining Sonora/Mojave populations [Leavitt et al., 2020]. Using the germline mutation rate for *C. brevis* of $3.17E^{-9}$ [Bergeron et al., 2023], we estimate a diversification date of 561,514 years for the split between SS and the BCN + BCS clade, and 470,031 years for the split between BCN and BCS. The divergence date estimated between the Mexico clade and the Mojave/Sonora clades is also 561,514 years, for the reasons mentioned in the Results. Mutation rates for taxonomic groups are highly variable and negatively correlate with effective population sizes [Lynch et al., 2023]. Despite mutation rates for reptiles being scarce in comparison to many other taxonomic groups [Wang and Obbard, 2023], we assume the rate inferred for *C. brevis* is the most biologically realistic rate for our dataset. This rate was estimated for the sister species to *C. variegatus* and was obtained using a minimally biased approach, wherein germline mutations are quantified from high-coverage genomes using parent-offspring trios [Bergeron et al., 2023].

2.5.2 Estimating Population Structure

Estimating an optimal number of population within a species is an important step in phylogeographic studies, which often requires a combination of analyses and biological information to accurately determine. For our study, we established and conducted subsequent species tree and gene flow estimations on an eight population model. To settle on a $K = 8$ model, we used a combination of analyses. CV scores for the ADMIXTURE analysis minimally differentiated between K-values from 6–12. Of note, $K = 12$ is the highest K-value we tested, so higher estimates may have also been supported. The network analysis identifies at least seven groups, although a drawback of this approach is the lack

of a statistical framework to evaluate network structures that assume different clusters or groups. The PCA only clusters five groups without overlap (Fig. 2.2). Lastly, as K increases to nine, a distinct population is recovered in the geographic range of the subspecies *C. v. utahensis*, although this pattern was not reflected by other programs. Because this population occurs at the northernmost extent of the geographic range for the species, this can lead to higher population structure and population fragmentation [Angert et al., 2020, Assis et al., 2013]. We consider that a $K = 8$ model accurately accounts for the high diversity within the species without artificially separating populations that are minimally divergent.

Although one of our primary questions asks whether population boundaries correspond to the boundaries of distinct ecoregions, we assigned samples to genetic populations for the MSC analyses using the ADMIXTURE results and not according to their ecoregion affiliation. To do so, any individual with admixture proportions $> 50\%$ was included with that respective population. The reason for assigning samples in this manner is because multiple genetically distinct populations can occur in a single ecoregion. Thus, comparing Sonora Desert to Mojave Desert would cause MO and DV to be a single group and NS, SS, and ST to be another, which the data do not support. Additionally, using ecoregions as population boundaries would substantially inflate gene flow metrics as we show ecoregion boundaries do not directly correlate with genetic boundaries. Our approach allows for more direct tests for gene flow among biological populations, whereas measuring gene flow across ecoregions (in a strict sense) would not provide biologically realistic results.

2.5.3 Patterns of Gene Flow

Prominent geographic and geologic breaks in the region of focus had varying impacts on the population structure and gene flow patterns for *C. variegatus*. Below, we provide an overview of the impact prominent barriers have on diversification in *C. variegatus* and other co-distributed reptiles.

The Colorado River serves as a well-defined barrier to gene flow for a variety of taxa [Dolby et al., 2019]. Despite phylogeographic breaks for many species occurring east and west of the river, this barrier does not appear to be a key factor limiting gene flow among *C. variegatus* populations (Table 2.5). We show that the samples occurring on one side of the river have high proportions of genetic material from populations on the other side, a pattern also seen in *Sceloporus magister* and *Phrynosoma mcallii* [Pavón-Vázquez et al., 2024, Gottscho et al., 2024].

The Baja California Peninsula has been a well-studied biogeographic region due to a high variety of lineages with with phylogenetic breaks occurring near the middle of the peninsular, splitting taxa into north and south groups. The Vizcaíno Seaway, and potentially other vicariant processes [Leaché et al., 2007], have been proposed to have separated the north and south margins of the peninsula. For *C. variegatus*, there is a phylogeographic break near the midpoint of the peninsular that correlates with low bidirectional gene flow rates between

the north (BCN) and south (BCS) of 0.0102 and 0.0140, respectively. Phylogeographic breaks occurring at, or near, the point where *C. variegatus* diverges are also detected in *Crotalus ruber*, *Callisaurus draconoides*, and *Urosaurus nigricaudus* [Harrington et al., 2018, Lindell et al., 2008, 2005].

The Salton Trough is a transition zone between the San Andreas Fault system and the spreading ridge complex of the Eastern Pacific Rise, which provides a stark divide between the California Coastal Sage ecoregion and the Sonora Desert. On the eastern margin of the Salton Trough region is the Bouse Formation, which contains the depositional record of a number of large bodies of water [Hafner et al., 2011]. The south and west ends of the Salton Trough form the margins of the ST population and correlate with low bidirectional gene flow between ST and BCN of 0.0705 and 0.134, respectively [Hafner et al., 2011]. In contrast, the NS population, which borders the Bouse Formation to the east, has high proportions of gene flow with the ST population (0.668 and 0.776, respectively). This pattern suggests that the eastern margin of the Salton Trough does not provide nearly as strong of a barrier to gene flow as the south and west margins for *C. variegatus*.

The populations with the highest gene flow are AZM and ST. As expected, these two populations only appear as genetically distinct populations with ADMIXTURE when K-values reach seven and eight, respectively (Fig. S1). The AZM population occurs at the northernmost extent of the Sonora Desert, extending into Eastern Sierra Madre and the Madrean Archipelago. This population has the highest gene flow of any population, particularly with the NS population (0.974 and 1.365, respectively).

2.5.4 Future directions

Although we find that a majority of population boundaries correlate with ecoregion boundaries, exploring additional factors driving the diversification of the species will prove valuable. Of particular interest is understanding the causes of limited gene flow between populations that share an ecoregion. For example, investigating factors leading to the comparatively low amounts of gene flow between MO and DV despite not having any clear barriers will be valuable for understanding parapatric population formation. Similarly, SS and NS have some of the lowest amounts of gene flow despite being within the same desert system and lacking clear geographic barriers to gene flow. Potential avenues of exploration include testing for variations in chemical signals released by each respective population, as these signals have been shown to inform population and species boundaries for morphologically cryptic lineages due to their direct impact on sexual selection [Zozaya et al., 2019, 2022]. Additionally, conducting extensive sampling in regions with signals of introgression, and in regions where sampling is limited, will help inform more precise geographic boundaries for each population, which may prove useful in detecting past or present geologic and geographic features that could have formed reproductive barriers. Lastly, directly testing for genomic signal such as positive or negative selection and signatures of second contact will help inform whether modern populations are

becoming more divergent or undergoing secondary contact. For example, on the Baja Peninsula, repeated flooding of the peninsula may have initiated the divergence between BCN and BCS, with the signatures of introgression being a recent emergence. Applying the population patterns shown herein to address these questions will be key to identifying drivers for the diversification of *Coleonyx variegatus* and potentially many other codistributed taxa.

Table 2.1: Assembly statistics for each ddRADseq datasets used in this study. Two clade-specific assemblies are shown (large and small Body), which we branched from to generate the species-specific assemblies. Assembly statistics represent an average of all samples from each respective assembly.

Assembly Name	Analyses	No. Ind.	No. Loci	Raw Reads	Clusters	Heterozygosity	Error
coleonyx	ML Tree	179	1,177	4,070,242.61	85,817.22	0.018	0.0016
variegatus	SPLITSTREE; PCA; ADMIXTURE	176	1,483	4,077,944.20	85,672.56	0.018	0.0016
variegatus_BPP	BPP: A01	32	8,675	4,302,402.22	90,065.63	0.017	0.0015
variegatus_A00	BPP: A00	65	4,117	4,325,276.88	90,776.31	0.018	0.0017
azm_ns	MSC-M	22	3,410	4,524,200.95	101,239.36	0.018	0.0019
bcn_bcs	MSC-M	11	5,901	3,664,455.45	71,728.18	0.015	0.0014
mo_azm	MSC-M	25	4,161	5,205,154.00	96,336.84	0.020	0.0015
mo_dv	MSC-M	18	4,164	4,656,155.22	81,740.78	0.020	0.0015
mo_st	MSC-M	19	4,060	4,216,639.63	86,444.84	0.020	0.0017
ns_ss	MSC-M	14	3,579	3,451,571.00	101,964.00	0.020	0.0022
ss_bcs	MSC-M	11	4,817	4,057,718.00	85,871.27	0.0017	0.0015
st_azm	MSC-M	20	3,867	4,500,712.35	95,563.55	0.020	0.0017
st_bcn	MSC-M	13	4,729	3,889,694.46	84,688.38	0.020	0.0018
st_ns	MSC-M	16	3,137	3,094,982.75	91,331.06	0.020	0.0022
st_ss	MSC-M	13	4,502	4,222,455.08	96,655.62	0.020	0.0019

Table 2.2: Population genetic diversity metrics for populations of *Coleonyx variegatus*. The bottom half of the diagonal is the average pairwise genetic distance; the top half is the F_{ST} values between the two populations. Acronyms are defined as follows: DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

	AZM	**BCN**	**BCS**	**DV**	**MO**	**NS**	**SS**	**ST**
AZM	-	0.434	0.496	0.4397	0.276	0.143	0.400	0.175
BCN	0.00361	-	0.498	0.568	0.458	0.484	0.553	0.443
BCS	0.00452	0.00360	-	0.595	0.515	0.541	0.592	0.514
DV	0.00402	0.00470	0.00532	-	0.322	0.48700	0.556	0.444
MO	0.00292	0.00395	0.00488	0.00295	-	0.340	0.443	0.273
NS	0.00183	0.00336	0.00437	0.00381	0.00286	-	0.447	0.186
SS	0.00348	0.00388	0.00494	0.00473	0.00394	0.00328	-	0.407
ST	0.00217	0.00331	0.00430	0.00371	0.00264	0.00181	0.00322	-

Table 2.3: D-statistics for populations of *Coleonyx variegatus*. The results shown are only for tests that had statistical support for introgression.

P1	P2	P3	P4	Dstatistic	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
MO	DV	BCN	ST	0.296	6.937	4.01E-12	0.164	101.547	50.435	27.413
NS	AZM	MO	ST	0.230	6.741	1.57E-11	0.176	64.402	55.117	34.4696
AZM	BCN	DV	MO	0.298	6.458	1.06E-10	0.188	99.809	53.741	29.044
AZM	BCN	DV	MO	0.298	6.458	1.06E-10	0.188	99.809	53.741	29.044
MO	DV	BCS	ST	0.302	6.231	4.62E-10	0.161	94.350	54.294	29.113
AZM	BCS	DV	MO	0.307	6.160	7.28E-10	0.195	99.551	55.492	29.435
NS	AZM	DV	ST	0.143	4.782	1.73E-06	0.0742	61.472	52.324	39.243
DV	MO	AZM	SS	0.191	4.425	9.65E-06	0.369	107.359	46.263	31.397
MO	DV	BCS	NS	0.231	4.222	2.42E-05	0.135	99.875	49.439	30.860
MO	DV	BCN	NS	0.214	4.134	3.57E-05	0.139	102.809	48.269	31.233
MO	DV	SS	ST	0.171	3.570	3.57E-04	0.122	106.586	42.577	30.133
BCN	BCS	NS	ST	0.174	3.332	8.62E-4	0.223	161.895	25.839	18.181
NS	AZM	SS	ST	0.133	3.320	9.01E-4	0.0852	57.635	45.708	34.955
NS	AZM	BCS	ST	0.165	3.217	1.30E-3	0.0857	52.683	50.511	36.172
BCN	BCS	AZM	ST	0.186	3.199	1.38E-3	0.746	150.923	28.758	19.735
AZM	DV	BCN	SS	0.189	3.056	2.24E-3	0.195	59.184	53.653	36.570
NS	AZM	BCN	ST	0.153	3.038	2.39E-3	0.0726	60.690	47.772	35.108

Table 2.4: Posterior probabilities estimated by the A00 analysis of BPP. τ estimates were converted to years using an estimated mutation rate. The 95% HPD columns indicate the confidence intervals for the 95% highest posterior probability. DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

	θ	**95% HPD**	τ	**95% HPD**	**Years**	**95% HPD**
MO	0.0151	0.0137, 0.0183	NA	NA	NA	NA
DV	1.31E-03	0.00121, 1.41E-3	NA	NA	NA	NA
AZM	0.0116	0.0101, 0.0137	NA	NA	NA	NA
NS	7.87E-03	0.00693, 8.82E-3	NA	NA	NA	NA
ST	6.80E-03	6.41E-3, 7.40E-3	NA	NA	NA	NA
SS	4.75E-03	4.53E-3, 4.97E-3	NA	NA	NA	NA
BCS	2.40E-03	2.28E-3, 2.52E-3	NA	NA	NA	NA
BCN	3.86E-03	3.64E-3, 4.10E-3	NA	NA	NA	NA
AZM-NS-ST-DV-MO-VCN-BCS-SS	5.58E-03	5.34E-3, 5.82E-3	1.78E-03	1.72E-3, 1.83E-3	561,514	542,586; 577,287
AZM-NS-ST-DV-MO	2.67E-02	0.024, 0.029	5.96E-04	5.66E-4, 6.46E-4	188,012	178,548; 203,785
AZM-NS-ST	3.27E-03	1.60E-3, 5.51E-3	5.44E-04	5.13E-4, 5.7E-4	171,608	161,829; 179,810
AZM-NS	2.81E-03	1.36E-3, 4.42E-3	4.82E-04	4.58E-4, 5.02E-4	152,050	144,479, 158,359
DV-MO	3.09E-03	2.13E-3, 4.30E-3	4.35E-04	3.99E-4, 4.66E-4	137,223	125,867, 147,003
BCN-BCS-SS	8.01E-03	1.60E-3, 0.0154	1.78E-03	1.71E-3, 1.83E-3	561,514	539,432, 577,287
BCN-BCS	1.70E-03	1.09E-3, 2.19E-3	1.49E-03	1.42E-3, 1.56E-3	470,031	447,949, 492,113

Table 2.5: Prior and posterior probabilities for the MSC-M analyses. The top half of the table shows the m estimates when using a gamma prior of 2,20 (10% migration per generation); and the bottom half shows estimates when using a prior of 2,200 (1%). m_{12} is the proportion of immigrants in population *1* from population *2*, and vice versa. The 95% HPD columns indicate the confidence intervals for the 95% highest posterior probability. Acronyms are defined as follows: DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

Populations	**θ Prior**	**τ Prior**	***M*1–*M*2**	**95% HPD**	***M*2–*M*1**	**95% HPD**
M Prior = 2, 20 (Mean 10%)						
BCN — BCS	3, 0.015	3, 0.0065	0.0102	0.00611, 0.0145	0.014	0.00788, 0.0204
MO — AZM	3, 0.03	3, 0.004	0.632	0.579, 0.690	0.607	0.535, 0.681
MO — DV	3, 0.03	3, 0.004	0.258	0.240, 0.277	0.245	0.209, 0.281
MO — ST	3, 0.02	3, 0.004	0.389	0.333, 0.446	0.518	0.451, 0.587
NS — SS	3, 0.013	3, 0.004	0.0796	0.0656, 0.0943	0.0561	0.0409, 0.0708
NS — AZM	3, 0.006	3, 0.0008	1.365	1.198, 1.530	0.974	0.797, 1.150
ST — AZM	3, 0.03	3, 0.004	0.672	0.582, 0.756	0.529	0.436, 0.623
ST — BCN	3, 0.02	3, 0.01	0.0705	0.0609, 0.0801	0.134	0.111, 0.158
ST — NS	3, 0.014	3, 0.01	0.776	0.632, 0.924	0.668	0.514, 0.833
ST — SS	3, 0.012	3, 0.01	0.122	0.0996, 0.144	0.0773	0.062, 0.0933
SS — BCN	3, 0.012	3, 0.01	5.68E-03	3.88E-3, 7.64E-3	0.0111	7.26E-3, 0.0149
M Prior = 2, 200 (Mean 1%)						
BCN — BCS	3, 0.015	3, 0.0065	9.19E-03	5.14E-3, 0.0133	0.0128	3.50E-3, 0.0259
MO — AZM	3, 0.03	3, 0.004	0.600	0.551, 0.653	0.495	0.431, 0.560
MO — DV	3, 0.03	3, 0.004	0.254	0.235, 0.272	0.206	0.176, 0.240
MO — ST	3, 0.02	3, 0.004	0.346	0.291, 0.403	0.437	0.375, 0.501
NS — SS	3, 0.013	3, 0.004	0.0747	0.0614, 0.0889	0.0497	0.0356, 0.0639
NS — AZM	3, 0.006	3, 0.0008	1.106	0.827, 1.420	0.631	0.347, 0.945
ST — AZM	3, 0.03	3, 0.004	0.602	0.521, 0.678	0.377	0.291, 0.467
ST — BCN	3, 0.02	3, 0.01	0.0702	0.0605, 0.0794	0.114	0.0935, 0.135
ST — NS	3, 0.014	3, 0.01	0.627	0.502, 0.749	0.384	0.250, 0.520
ST — SS	3, 0.012	3, 0.01	0.0728	0.0572, 0.0881	0.107	0.087, 0.128
SS — BCN	3, 0.012	3, 0.01	5.62E-03	3.83E-3, 7.48E-3	0.0105	7.11E-3, 0.0141

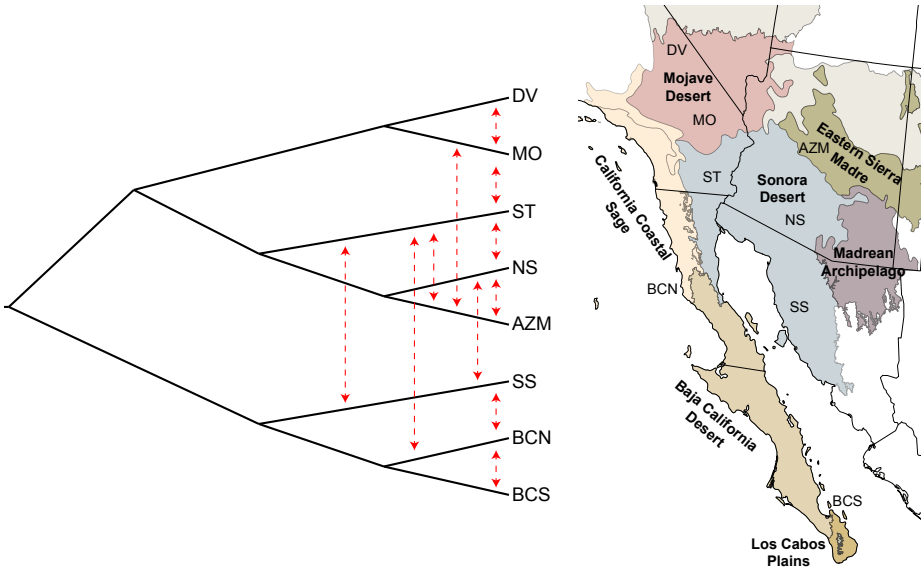

Figure 2.1: Model testing framework implemented in our study to explore gene flow patterns between *Coleonyx variegatus* populations. Double sided red arrows indicate the populations that we tested for signatures of gene flow. The position of the arrow on the branch does not correspond with testing for gene flow at that specific time in the lineages' evolutionary history. Colors on the map correlate with the ecoregion boundaries. Population acronyms are indicative of the general area that the populations occur in. Acronyms are defined as follows: DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

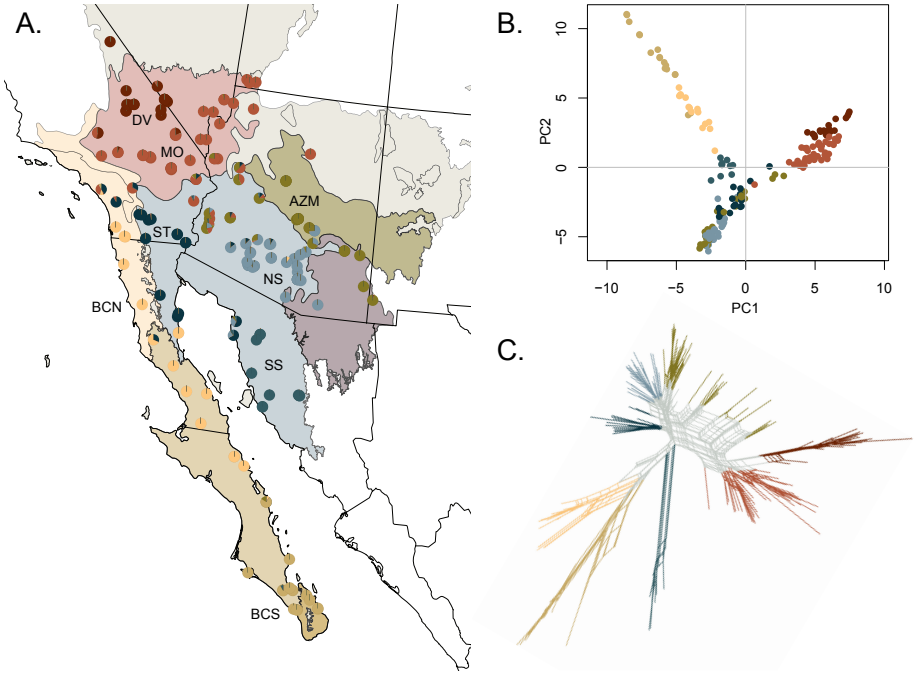

Figure 2.2: Population structure results established for *Coleonyx variegatus*. A) Map of the southwest US and northern Mexico, with colored regions indicating the distinct ecoregions in the region. Pie charts indicate the admixture proportions for each individual of *Coleonyx variegatus* included in our study. B) PCA of the ddRADseq data, with colors corresponding to genetic population the sample is assigned to based on the ADMIXTURE results. C) Network analysis of the dataset with colors corresponding to genetic population the sample is assigned to based on the ADMIXTURE results. Acronyms are defined as follows: DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

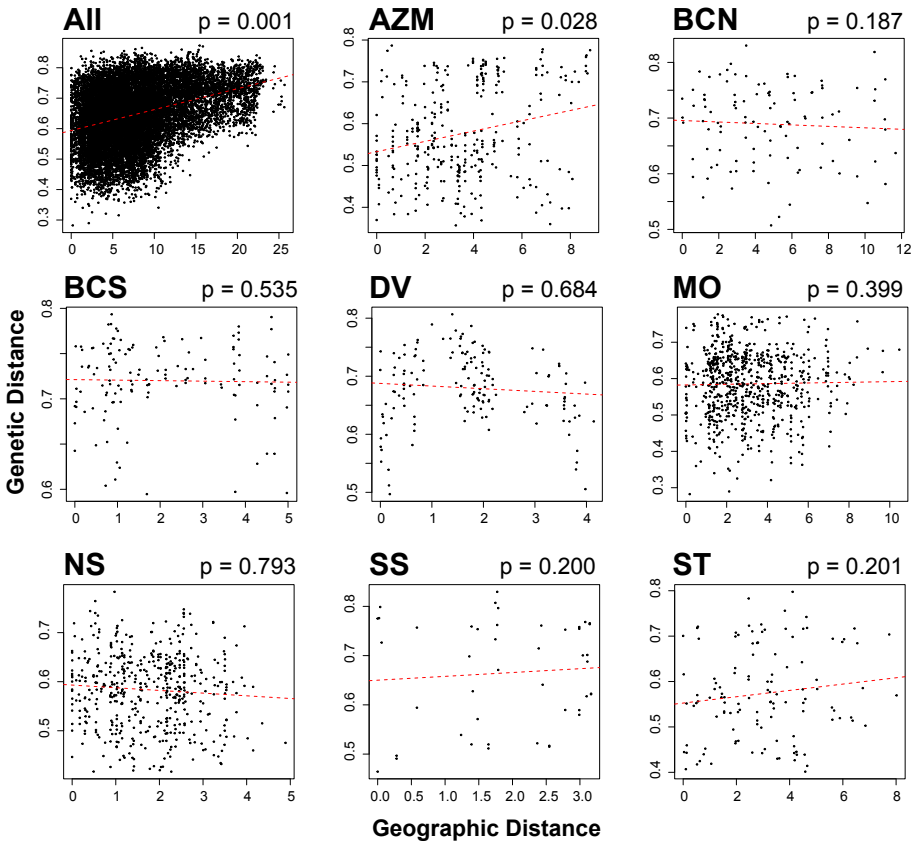

Figure 2.3: Isolation-by-distance plots for samples of *Coleonyx variegatus* spanning the geographic distribution of the species, and IBD plots for each population. X-axis for all plots is geographic distance and the y-axis is genetic distance. Red lines demonstrate the regression between geographic distance and genetic distance. P-values are shown assuming a null model of the genetic diversity not being driven by geographic distance. Acronyms are defined as follows: DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

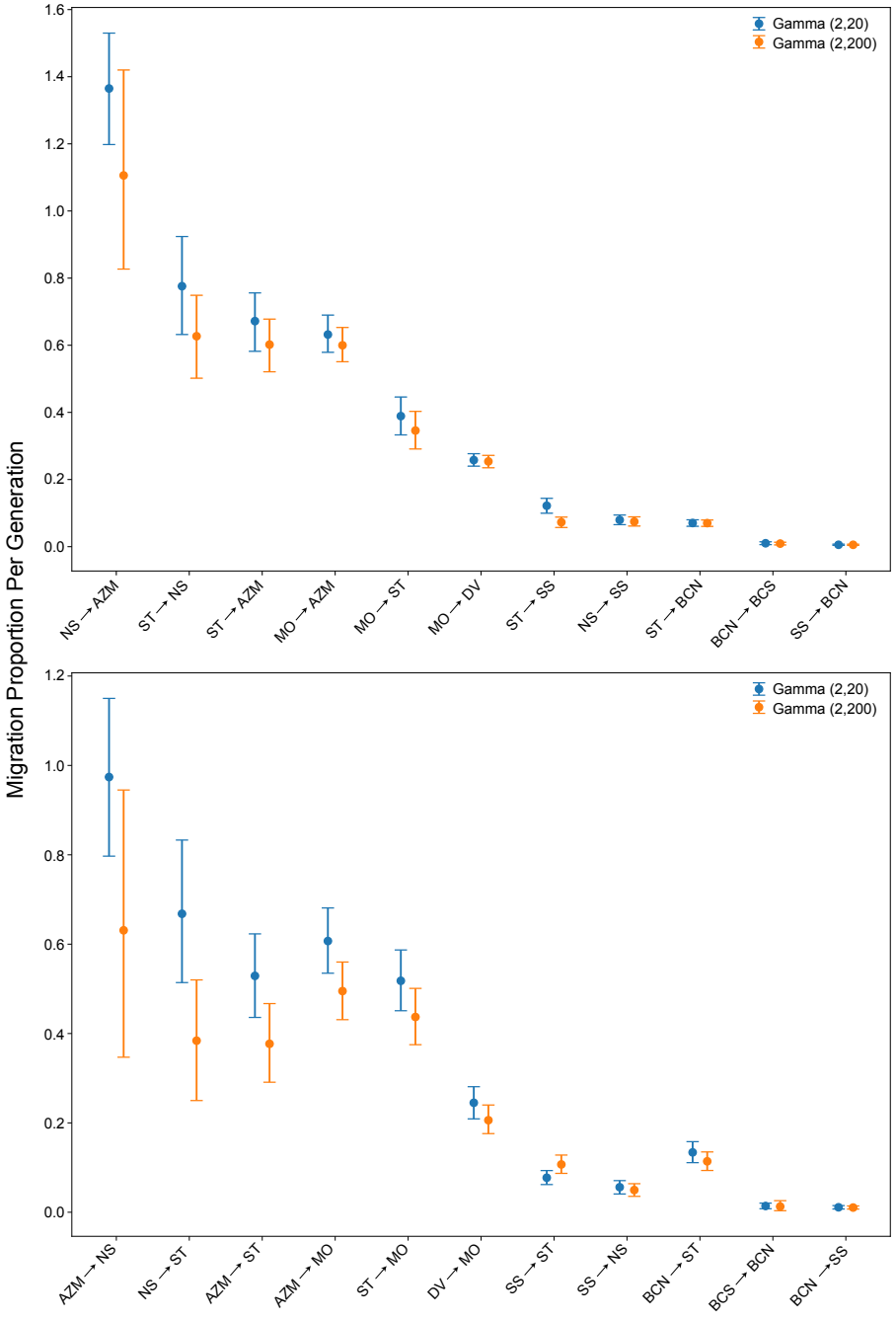

Figure 2.4: Bidirectional estimates of M estimated using the MSC-M model for *Coleonyx variegatus*. Labels on the X-axis indicate the populations and the arrow indicates the direction of gene flow. The y-axis is the migration proportion per generation. The circles in the plot indicate the mean m estimated for the given population comparison, and the bars show the 95% HPD intervals. Colors correlate with the two priors used to conduct each analysis. Acronyms are defined as follows: DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

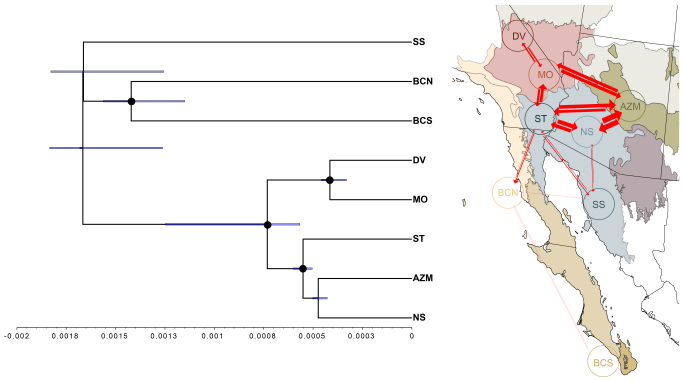

Figure 2.5: Left) Species tree for *Coleonyx variegatus* inferred using BPP with 4,117 RAD loci. Blue bars on the nodes indicate the 95% HPD intervals, and black circles denote nodes with strong support ($pp \geq 95$). Right) Migration estimates (*M)* from the MSC-M model are demonstrated using arrows where the width of the line is proportional to the amount of gene flow between the given populations. Acronyms are defined as follows: DV — Death Valley; MO — Mojave Desert; AZM — Arizona Mountains; ST — Salton Trough; NS — northern Sonora Desert; SS — southern Sonora Desert; BCN — Baja California Norte; and BCS — Baja California Sur.

Chapter 3

Establishing species boundaries in Bornean geckos

Hayden R. Davis, Henry T. Sanford, Indraneil Das, Izneil Nashriq, & Adam D. Leaché

3.1 Abstract

Species delimitation using mitochondrial DNA (mtDNA) remains an important and accessible approach for discovering and delimiting species. However, delimiting species with a single locus (e.g. DNA barcoding) is biased towards overestimating species diversity. The highly diverse gecko genus *Cyrtodactylus* is one such group where delimitation using mtDNA remains the paradigm. In this study, we use genomic data to test putative species boundaries established using mtDNA within three recognized species of *Cyrtodactylus* on the island of Borneo. We predict that multilocus genomic data will estimate fewer species than mtDNA, which could have important ramifications for the species diversity within the genus. We aim to 1) investigate the correspondence between species delimitations using mtDNA and genomic data; 2) infer species trees for each target species; and 3) quantify gene flow and identify migration patterns to assess population connectivity. We find that species diversity is overestimated and that species boundaries differ between mtDNA and nuclear data. This underscores the value of using genomic data to reassess mtDNA-based species delimitations for taxa lacking clear species boundaries. We expect the number of recognized species within *Cyrtodactylus* to continue increasing, but when possible, genomic data should be included to inform more accurate species boundaries.

3.2 Introduction

Species delimitation, the process of determining whether two populations belong to the same or distinct species, is an inherently complex task that depends on the operational, methodological, and philosophical species definitions. The proliferation of genomic data has reinvigorated the need for detailed investigations into the taxonomy of diverse species spanning the tree of life. This has coincided with the development of coalescent-based approaches for species delimitation that can reconcile genome-wide genealogical discordance [Leaché et al., 2014, O'Meara, 2010, Grummer et al., 2014, Yang, 2015, Sukumaran et al., 2021, Jackson et al., 2017b], thus advancing beyond methods that rely on single gene trees [Puillandre et al., 2021, Zhang et al., 2013, Fujisawa and Barraclough, 2013]. Together, these data and methods have advanced species delimitation on the operational and methodological fronts. Yet, identifying the boundary between populations and species for many taxa remains challenging [Maddison and Whitton, 2023], a problem often referred to as the 'grey zone' of speciation [Roux et al., 2016, Stankowski and Ravinet, 2021].

For genetically divergent yet morphologically cryptic lineages, species delimitation can prove particularly difficult as the sources of informative trait data for integrative approaches are limited [Singhal et al., 2018, Struck et al., 2018]. Due to a lack of diagnostic morphological characters in cryptic taxa, a heavier reliance is placed on genetic data [Fišer et al., 2018, Struck et al., 2018, Singhal et al., 2018]. However, for many researchers, the cost of genomic data remains financially intractable, resulting in most taxonomic studies of cryptic taxa to rely on DNA barcoding, or similar approaches, which have lower accuracy due to their reliance on the signal from a single locus [Chan et al., 2017, Burriel-Carranza et al., 2023, Hickerson et al., 2006, Yang and Rannala, 2017, Pedraza-Marrón et al., 2019, Barley and Thomson, 2016].

To explore whether putative species identified using mtDNA data are supported by genomic data, we focused on a group of related species within the highly diverse gecko genus *Cyrtodactylus*. The genus is the third most diverse among vertebrates with over 350 recognized species, yet few studies have explored evolutionary patterns within the group using genomic data [Oaks et al., 2019, Reilly et al., 2023, Davis et al., 2023]. We focus on three independently-derived cryptic species from the Southeast Asian island of Borneo (*C. consobrinus*, *C. miriensis*, and *C. pubisulcus*), where previous studies based primarily on mtDNA have uncovered evidence for potential additional unrecognized species [Davis et al., 2020, 2021]. We use this study as an opportunity to investigate the correspondence between mtDNA and genomic species delimitations.

Whether explicitly stated or not, the species concept held by a researcher plays a critical role in species delimitation. We view species as independent evolutionary lineages, which is a broadly-shared view compatible with most modern species concepts [De Queiroz, 1998, Mayden, 1997, Naomi, 2011], and we also see great value in considering species as reproductive communities emerging from the past [Maddison and Whitton, 2023]. In our study, we focus solely on genetic data, as former studies have demonstrated the lack of morphological

distinction within all three of the target species [Davis et al., 2021, 2023]. We analyze the mtDNA data using approaches often used to identify putative *Cyrtodactylus* species to provide comparison to current practices, and analyze the genomic data using multispecies coalescent methods that incorporate gene flow estimates. Using the demographic parameters from these analyses, we estimate the genealogical divergence index (*gdi*), a metric that helps disentangle species and population boundaries [Jackson et al., 2017b, Leaché et al., 2019]. Using these approaches, we demonstrate that mtDNA and genomic-based delimitations often differ, and our results emphasize the value of using genomic data to test species boundaries, especially for cryptic lineages.

3.3 Methods

Fieldwork was conducted in Borneo in the Malaysian state of Sarawak between 2014-2018 (Permits: NPW.907.4.4.(Jld.14)-79; (119)JHS/NCCD/600-7/2/107). We aimed to collect specimens of *Cyrtodactylus consobrinus*, *C. miriensis*, and *C. pubisulcus*, as the current taxonomy defines, from as many unique localities as possible, but heavy deforestation and minimal road access to field sites resulted in patchy sampling for these species. We sampled each species from multiple locations and obtained multiple specimens per locality (Fig. 3.1).

We extracted genomic DNA from liver biopsies using salt-extraction [Aljanabi and Martinez, 1997]. We supplemented the sequence data available on GenBank by sequencing individuals for the ND2 locus. We followed standard PCR amplification and sequencing protocols with primers used in a previous *Cyrtodactylus* study [Davis et al., 2019].

We generated an assembly for *Cyrtodactylus* species in Borneo using MAFFT [Katoh et al., 2002], comprising two evolutionarily distinct clades referred to herein as the small-bodied (*Cyrtodactylus cavernicolus, C. miriensis, C. pubisulcus, C. hantu*) and large-bodied (*C. consobrinus, C. hutan, C. kapitensis, C. malayanus*) clades. We included the additional non-target species to more accurately infer the topologies used for testing species delimitation models. The alignment comprises 83 taxa and 1,010 bp. Sample sizes for the target species are as follows: *C. miriensis* = 19, *C. pubisulcus* = 20, and *C. consobrinus* = 9. Novel sequences are available on GenBank (XXXXX). We used the assembly to apply commonly used species delimitation methods (bPTP; mPTP; GMYC; ASAP) and calculate pairwise-distances (p-distance) using SPDEL [Ramirez et al., 2023, Puillandre et al., 2021, Kapli et al., 2017, Fujisawa and Barraclough, 2013], and estimated genealogies using BEAST2 [Bouckaert et al., 2014].

To generate genomic data, we conducted double digest restriction-site associated DNA sequencing (ddRADseq) [Peterson et al., 2012]. We double-digested each sample using the digestion enzymes *SbfI* and *MspI* in CutSmart Buffer (New England Biolabs) for 7 hours at 37 °C For fragment purification, we used Sera-Mag SpeedBeads. We then ligated eight distinct barcodes to the cut sites of the fragmented DNA and subsequently size-selected (between 415 and 515

bp after accounting for adapter length) each library on a Blue Pippin Prep size fractionator (Sage Science). For the final library amplification, we used Phusion Hi-Fidelity DNA Polymerase and Illumina's indexed primers. We determine the concentration and size distribution of each indexed pool using an Agilent 2200 TapeStation. Lastly, we sent the quantified pools to QB3-Berkeley Genomics, UC Berkeley for qPCR to determine sequenceable library concentrations before multiplexing equimolar amounts of each pool for sequencing on two Illumina HiSeq 4000 lane (51-bp, single-end reads; 9 pools containing 8 samples). All sequences are uploaded to the Sequence Read Archive (SRA; BioProject ID: PRJNA1117503; Table S3.2). We combined these data with the ddRADseq data from [Davis et al., 2023].

Sample sizes in the assembly for the target species are as follows: *C. miriensis* = 26; *C. pubisulcus* = 32; and *C. consobrinus* = 15. To maximize the number of loci and reduce missing data, we produced separate assemblies for the small and large-bodied clades (Table S3.1). From the small and large-bodied assemblies, we used branching in IPYRAD [Eaton and Overcast, 2020] to generate species-specific assemblies from their respective clades (Table S3.1). We removed individuals from each assembly containing less than 500,000 raw reads. For all assemblies, we applied a sequence similarity threshold of 85% to cluster reads within samples and loci between samples. We removed consensus sequences with low coverage (< 6 reads), excessive undetermined or heterozygous sites ($> 5\%$), too many alleles for a sample (> 2 for diploids), or an excess of shared heterozygosity among samples (paralog filter = 0.5). For clade-specific assemblies, we required approximately 60% of individuals to share any given locus. To maximize the number of loci from the dataset, we generated separate assemblies for both Bornean clades (large- and small-bodied). For population genetic analyses, we filtered each dataset using VCFTOOLS [Danecek et al., 2011] to only allow one SNP per locus (`--thin 50`) and filtered out variable sites present in less than 5% of individuals (`--maf 0.05`).

Using a reduced dataset for computational efficiency, we inferred time-calibrated species trees using SNAPP within the BEAST2 framework for both the small and large-bodied datasets [Bouckaert et al., 2014]. The reduced dataset contained 32 small-bodied and 28 large-bodied individuals. We applied secondary calibrations to the roots of each tree using the snapp_prep ruby script, which allows the generation of an .xml file with a molecular clock [Matschiner, 2022]. For the large-bodied dataset, we constrained the crown age of the tree to 18.83 mya with a normal distribution ($\sigma = 2$), and for the small-bodied clade we constrained the crown age to 25.80 mya with a normal distribution ($\sigma = 2$). We applied dates inferred from a previous Bornean *Cyrtodactylus* study, which used secondary calibrations from a broader fossil-calibrated *Cyrtodactylus* phylogeny [Davis et al., 2020, Wood Jr et al., 2012]. We ran each dataset independently in BEAST2 without manually setting an outgroup for either analysis. We applied a strict clock model to the analysis and ran two independent runs for 500,000 generations each. We combined MCMCs from both runs using LogCombiner with a burn-in percentage of 20%. Finally, we checked for convergence by examining MCMC line plots and ensuring that ESS scores were over 200.

To test for phylogenetic structure within populations, we concatenated the RAD loci and inferred a phylogeny using IQ-TREE [Nguyen et al., 2015]. We determined the best-fit model of evolution using the embedded program MODELFINDER [Kalyaanamoorthy et al., 2017], applied 1000 ultrafast bootstraps to assess node support [Hoang et al., 2018], and tested branch support with an SH-like approximate likelihood ratio test [Guindon et al., 2010].

For the population-based analyses, we analyzed the SNP data using three analytical frameworks to explore the processes driving population divergence and to delimit species. These approaches included structure inference, estimation of migration surfaces, and species delimitation and gene flow estimation using the MSC-M model [Flouri et al., 2023]. We estimated population structure using principal component analyses (PCA) and ADMIXTURE [Alexander and Lange, 2011], and assessed models with varying population numbers (K) for biological reality. For the ADMIXTURE analyses, we replicated the analysis ten times to ensure consistency across runs and determined the optimal K-value by taking the lowest average cross-validation score. To visualize the results, we used the program CLUMPAK [Kopelman et al., 2015].

To identify geographic areas with increased migration (corridors) and reduced migration (barriers) we used Estimating Effective Migration Surface method (EEMS) [Petkova et al., 2016]. Quantifying migration and population connectivity are pertinent to the exploration of spatial patterns of genetic diversity. EEMS identifies deviations from population structure expected under a model of isolation-by-distance (IBD). For each species we performed 50 million iterations with a burn in of 10 million and 9,999 thinning iterations. We repeated the analysis at both 500 demes and 50 demes. We checked for convergence by examining MCMC line plots and visualized the results using the rEEMSplots package. Additionally, we directly tested for IBD using the R package ADEGENET [Jombart, 2008]. By considering the combined results from mtDNA genealogies, population structure, and migration surfaces, we established population boundaries used to conduct demographic modeling and phylogenetic tests for gene flow.

To conduct joint species delimitation and gene flow quantification, we used the Python wrapper HHSD [Kornai et al., 2023] to run BPP with the MSC-M model [Flouri et al., 2023], and subsequently used the demographic values to calculate *gdi* values [Leaché et al., 2019, Jackson et al., 2017a]. We started with species guide trees obtained by running the A01 step in BPP (when more than two tips were tested). We ran two independent A01 analyses to assess for convergence on the same tree. We then used HHSD to estimate population sizes (θ), species divergence times (τ), and gene flow rate (M; where $M = m\mathrm{N}$). Conducting these tests in BPP requires a priori populations to be defined, which we set using the ADMIXTURE and PCA results. Because HHSD iterates through populations and merges those that do not meet the *gdi* cutoff, we tested models where K was one higher than the optimal number, which also enabled us to test for gene flow among disjunct geographic regions. The *gdi* value is the probability that two alleles from a population will coalesce with one another before reaching the ancestral population. The *gdi* metric spans

the speciation continuum, ranging from 0 (panmictic) to 1 (genetically distinct) [Jackson et al., 2017a]. Values > 0.2 and < 0.7 are considered ambiguous. Using the combination of these methods, we identify putative species that may be representative of species level diversity. Lastly, we converted the τ values to years using a germline mutation rate of 4.46E-9 [Bergeron et al., 2023]. We derived the mutation rate by taking the mean of the average yearly mutation rate estimated for three gecko species: *Coleonyx brevis, Eublepharis macularius*, and *Sphaerodactylus inigoi* [Bergeron et al., 2023]. We divided τ parameter inferred using BPP by the mutation rate to obtain ages in years.

3.4 Results

3.4.1 mtDNA-Based Species Delimitation.

Using mtDNA species delimitation approaches, *Cyrtodactylus miriensis* is estimated to be comprised of 2–4 species, *C. pubisulcus* from 2–3 species, and C. consobrinus from 1–2. Within each of these species, a high amount of diversity is present (Fig. 3.1; Table 3.1). The pairwise distances (p-distance) within major lineages of C. miriensis range from 3.15% to 7.32% (min. – max.), 2.07% to 10.3% for lineages within C. pubisulcus, and 4.31% for the single major phylogenetic division within *C. consobrinus* (Fig. 3.1). These p-distances are within the current standards for what constitutes a species by other studies using similar types of data and methods. For example, recent taxonomic work within the highly diverse *C. pulchellus*, *C. intermedius*, and *C. khasiensis* species complexes has recognized new species with p-distances $\geq 6\%$, $\geq 3.5\%$, and $\geq 4.0\%$, respectively, for the ND2 locus [Wood Jr et al., 2020, Termprayoon et al., 2023, Lalremsanga et al., 2023].

3.4.2 Species Tree Inference.

Both the species trees inferred using SNAPP and the concatenated phylogeny have strong support for all species-level relationships (PP = 1; UFBoot ≥ 95; Fig. S1). The divergence dating analysis supports C. miriensis as the oldest lineage (25.6 mya [21.5–29.5 mya]), followed by *C. pubisulcus* (14.7 mya [12.2–16.9 mya]), and then C. consobrinus (5.1 mya [3.9–6.2 mya]). Of note, these dates are highly reliant on commonly used fossil calibrations, which are substantially older than the dates inferred using the demographic values and a molecular substitution rate established for geckos (see below).

3.4.3 Population Structuring and Demography.

The PCA analysis of the SNP data reveals that each species clusters primarily by geography (Fig. 3.2). The optimal number of populations (K-values) estimated using population structure inference supports fewer distinct clusters than revealed in the PCA plots. The optimal K-value for *C. miriensis* is uncertain with different analyses supporting $K = 1$–4. The optimal models for *C.*

consobrinus and *C. pubisulcus* are $K = 1$ (Fig. S3.2). The $K = 3$ model for *C. miriensis* structures populations geographically and closely matches the PCA groupings. Assuming a two population model for *C. pubisulcus* also provides a clear geographic division that is reflected in the PCA analysis (Fig. 3.2). For *C. consobrinus*, a population model with $K = 1$ is optimal, and assuming $K = 2$ demonstrate a lack of geographic structure (Fig. 3.2). The population structure and admixture results provide evidence of admixed samples in *C. pubisulcus* and *C. miriensis*. For example, within *C. pubisulcus*, one sample from the region northeast of Borneo Highlands is admixed between the Borneo Highlands and surrounding populations (Serian; Fig. 3.2).

For each species, significant levels of gene flow are detected between populations, although in some cases symmetric gene flow can be rejected (Table 3.2). Within *C. miriensis*, one instance of bidirectional gene flow is detected ($M = 0.09$ [5.0E-2, 0.11] from MIR and NIH; $M = 1.17$ [0.93, 1.48] from NIH and MIR) and two instances of unidirectional gene flow ($M = 0.56$ [0.37, 0.72] from MUL to MIR; M = 0.35 [0.29, 0.41] from NIH and MUL; Table S3). The remaining gene flow estimates include zero in the 95% credible interval and are therefore considered not significant (Table S3). Gene flow in *C. pubisulcus* ($M =$ 0.44 [0.33–0.57] from KUC to BH; $M = 0.21$ [0.003–0.34] from BH to KUC) and *C. consobrinus* ($M = 0.06$ [5.8E-05–0.37] from KUC to SER; $M = 0.14$ [8.3E-4–1.02] from SER to KUC) is bi-directional. Signatures of IBD are significant in *C. miriensis* and *C. consobrinus* (p-value = 0.004 and 0.025, respectively). The geographic pattern of diversity within C. pubisulcus is not driven by IBD (p-value = 0.434; Fig. 3.2).

3.4.4 Diversification Date Estimates.

Divergence dates estimated from the genomic data using a germline-based mutation rate are younger than fossil-based estimates (Table S3.3). For *C. miriensis*, the earliest divergence occurs at 665 kya (95% highest posterior density (HPD): 409–817 kya), followed by a more recent divergence between MIR and NIH at 146 kya (95% HPD: 108–188 kya). For *C. pubisulcus*, the earliest divergence occurs at 252 kya (95% HPD: 217–289 kya). Lastly, for *C. consobrinus*, the earliest divergence occurs at 60 kya, albeit with broad confidence intervals (95% HPD: 44–330 kya). Of note, as mentioned above, the best model for the *C. consobrinus* is as a single population.

3.4.5 SNP-Based Species Delimitation.

Even when mtDNA and nuclear data support the same number of species, we found species boundaries that differed. Species delimitation using the *gdi* metric indicates that C. miriensis may contain two unrecognized species ($gdi \geq 0.50$; Table 3.2). The *gdi* values for *C. pubisulcus* are low (0.18) to intermediate (0.32), resulting in dubious support for splitting these populations into two species. The two species model for *C. pubisulcus* differs from the two species delimited by mtDNA data by placing the population from Borneo Highlands as

the only distinct genetic group, whereas mtDNA forms two groups: 1) Borneo Highlands + Bau + Kuching; and 2) Gunung Pueh + Gunung Gading + Gunung Matang. Finally, there is weak evidence for splitting *C. consobrinus* into two species, with *gdi* values definitively within the single-species range (0.12 and 0.11; Table 3.2).

3.5 Discussion

Accurately delimiting species is an inherently complicated process, especially when dealing with cryptic taxa. Yet, species serve as fundamental units for conservation and biological research, thus there is a clear rationale for applying rigorous approaches and relevant data for species delimitation [Derkarabetian et al., 2022]. For taxonomically complicated groups, it is advantageous to leverage genomic data to explore evolutionary patterns and the processes giving rise to them. We recognize the immense value of using mtDNA loci in systematic studies, yet it is important to acknowledge the limitations imposed by the reliance on a single genetic locus when delimiting species [Chan et al., 2017, Burriel-Carranza et al., 2023, Hickerson et al., 2006, Yang and Rannala, 2017, Pedraza-Marrón et al., 2019, Barley and Thomson, 2016, Vences et al., 2024].

Although previous studies have shown that mtDNA tends to delimit more species than nuclear data, we find three distinct outcomes in our comparisons of mtDNA versus genomic species delimitation in *Cyrtodactylus* geckos: 1) oversplitting with mtDNA; 2) agreement on number of species, but differing population boundaries; and 3) concordant species hypotheses. In the case of C. miriensis, genomic data provide strong support for recognizing a new species from Lawas based on a combination of a high *gdi* value and a lack of gene flow (Tables 3.1, 3.2). However, we advise sampling of additional localities in the region prior to making any taxonomic changes to determine if this result is potentially biased by the small effective population sizes, which can skew *gdi* values towards supporting species-level designations [Leaché et al., 2021]. Furthermore, the strong signature of IBD in this group indicates that comprehensive sampling could reveal a geographic gradient of diversity. For *C. pubisulcus*, mtDNA and genomic data largely agree on a two-population model, yet the composition of the populations differs. Such disagreement can be caused by incomplete lineage sorting, mtDNA introgression, or gene flow [DeRaad et al., 2023, Toews and Brelsford, 2012, Firneno Jr et al., 2020]. The species delimitation comparison in *C. pubisulcus* is illustrative of the different conclusions that are reached when using a single locus vs. genomic data; specifically, the single locus methods clearly support two species ($> 10\%$ p-distance), whereas the genomic data support two populations connected by gene flow with weak support for a two-species model. Lastly, *C. consobrinus* is supported by both datasets as a single lineage, although some single-locus delimitation methods support a two-species model (Fig. 3.1).

The divergence dates estimated using a genomic rate calibration rather than a fossil calibration are substantially younger (Table S3.3). Fossil-based ap-

proaches estimate the earliest divergence time at 7.7 mya for C. miriensis and 8.3 mya for *C. pubisulcus* [Davis et al., 2020]. Using a rate calibration, these same divergence events are estimated at 665 kya and 252 kya, respectively. There are several sources of error that need to be considered for the fossil calibration approach. First, there are no crown or stem Gekkonidae fossils that can confidently be placed on the phylogeny [Daza et al., 2014]. Therefore, secondary calibrations from other studies must be used, which propagates errors across studies. More importantly, the fossil calibrations that are frequently applied to *Cyrtodactylus* depend on fossils for *Sphaerodactylus* which shares an MRCA with Gekkonidae > 100 mya. Alternatively, the rate calibration we used is estimated from germline mutations quantified from high-coverage genomes using parent-offspring trios [Bergeron et al., 2023]. Therefore, we suspect that the rate-based divergence times are more accurate than the fossil-derived dates.

A prominent issue facing many *Cyrtodactylus* — and other taxa with similar geographic ranges — is an inability to comprehensively sample between known populations, which can overestimate of species diversity. Extensive sampling from contact zones can better inform taxonomic decisions by revealing the patterns and processes of population divergence and/or population merging [Chambers et al., 2022, Carstens et al., 2013, Agnarsson and May-Collado, 2008, Nabhan and Sarkar, 2012, Heath et al., 2008]; whereas sampling isolated populations can inflate species numbers by accentuating the distinctiveness of populations, even under the MSC model [Opatova et al., 2023]. In regions such as Borneo, where extensive land is inaccessible due to private ownership, a lack of public access, and/or having undergone extensive deforestation, sampling gaps often reflect true absence data. However, these sampling gaps, whether biologically real or a byproduct of the landscape, can impact how we define a species boundary [Hedin et al., 2015, Mason et al., 2020]. Sampling for *Cyrtodactylus* on Borneo for genetic studies has only just begun. Using genetic samples available to us, we aimed to generate genomic data for as many unique localities as possible, while utilizing enough samples per locality to accurately infer demographic patterns under a MSC model. Although genomic data do not definitively resolve these issues, they provide an avenue for inferring demographic histories in a manner that is not possible with mtDNA data. This demography can translate into more accurate taxonomic inferences, especially in complex systems containing cryptic diversity [Pyron et al., 2023]. Despite our incomplete sampling, we demonstrate the utility of using genomic data to interrogate mtDNA-based delimitations.

The high diversity of *Cyrtodactylus* and its rank as the third most species rich vertebrate genus is not an artifact of oversplitting with mtDNA, but rather reflects an old origin, broad geographic distribution, high ecological diversity, or other factors [Grismer et al., 2020, 2022]. Yet, applying genomic data to delimit species within *Cyrtodactylus* more broadly may reduce the number of species in the group, especially for some of the recently described cryptic species based on mtDNA delimitation. We demonstrate how mtDNA alone may be insufficient to accurately infer species boundaries for cryptic lineages. Dependence on mtDNA will presumably continue for many years, as it provides a compar-

atively quick and inexpensive approach to documenting life's biodiversity. Yet, for *Cyrtodactylus*, we recommend reassessing mtDNA species boundaries as genomic data become more readily available, and ultimately utilizing genomic data to explore the impressive diversification within the genus.

Table 3.1: Number of species estimated by each of the species delimitation programs utilized in this study. For the single-locus mtDNA data, we use bPTP, mPTP, gmyc, and asap, and for genome-wide data we use the gdi metric. Min and max p-distances are for the mtDNA dataset. The min values correspond with the lowest p-distance among populations that resulted in a delimitation for any of the species. The maximum p-distance is the highest p-distance among populations that resulted in a delimitation for any of the species. The asterisk for *Cyrtodactylus pubisulcus gdi* corresponds to a different species boundary between mtDNA and SNP data.

Species	**Max p-distance**	**Min p-distance**	**bPTP**	**mPTP**	**gmyc**	**asap**	***gdi***
C. miriensis	7.30%	3.10%	4	4	4	2	≤ 3
C. pubisulcus	10.00%	2.10%	3	2	2	2	$\leq 2^*$
C. consobrinus	4.60%	4.60%	2	1	2	1	1

Table 3.2: Species delimitation results for *Cyrtodactylus* based on genomic data. Lineages with a $gdi \leq 0.2$ are considered conspecific and ≥ 0.7 are considered distinct species; the species status for lineages > 0.2 and < 0.7 is ambiguous. M 1 is the gene flow estimate from population 1 into population 2; M 2 is the gene flow estimate from population 2 into population 1, with the 95% confidence intervals in parentheses. MIR = Tinbarap + Lambir Hills +Lawas; NIH = Niah; MUL = Gunung Mulu; BH = Borneo Highlands.

Species	**Pop 1**	**Pop 2**	***gdi 1***	***gdi 2***	***M 1***	***M 2***
C. miriensis	MIR/NIH	MUL	0.3	0.59	1.90E-02 (0.0, 0.11)	5.70E-03 (0.0, 3.5E-2)
	MIR	NIH	0.19	0.5	0.09 (5.0E-2, 0.11)	1.17 (0.93, 1.48)
	MIR	MUL	NA	NA	0.04 (0.0, 9.5E-2)	0.56 (0.37, 0.72)
	MUL	NIH	NA	NA	5.30E-04 (0.0, 3.8E-3)	0.35 (0.29, 0.41)
C. pubisulcus	KUC	BH	0.18	0.32	0.44 (0.33, 0.57)	0.21 (3.0E-3, 0.34)
C. consobrinus	KUC	SER	0.12	0.11	0.06 (5.8E-05, 0.37)	0.14 (8.3E-4, 1.02)

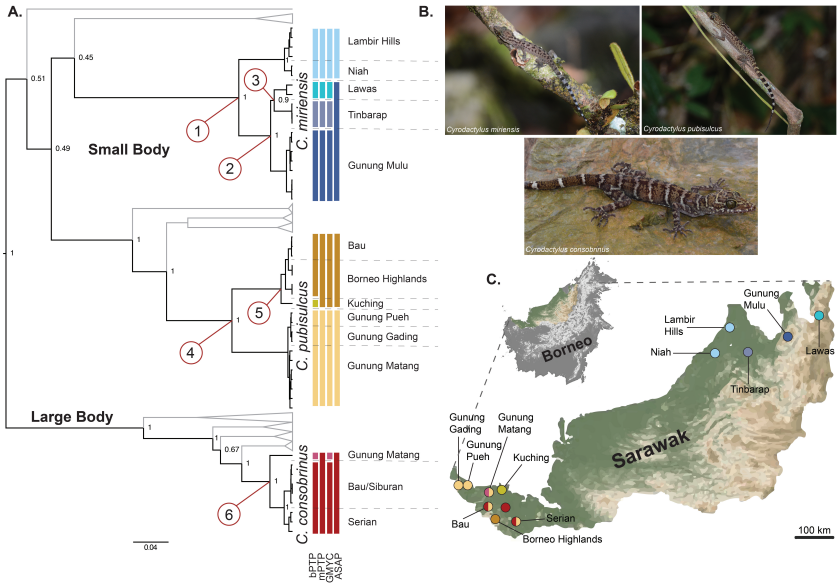

Figure 3.1: A) mtDNA gene tree highlighting the three *Cyrtodactylus* species of focus in this study. Values on the nodes are posterior probabilities. Vertical bars on the tips of the nodes correspond to the delimitations estimated by the single-locus delimitation programs: bPTP, mPTP, GMYC, and ASAP, where the colors of the bars denote the delimitation hypothesis. Node numbers 1–6 correspond to putative species of the most species rich delimitation hypotheses. Branches for species not discussed herein are greyed out. B) Photographs of live individuals of each of the three species of focus. C) Map of localities from which we have sequence data, with the colored dots corresponding to delimitations highlighted by nodes 1–6.

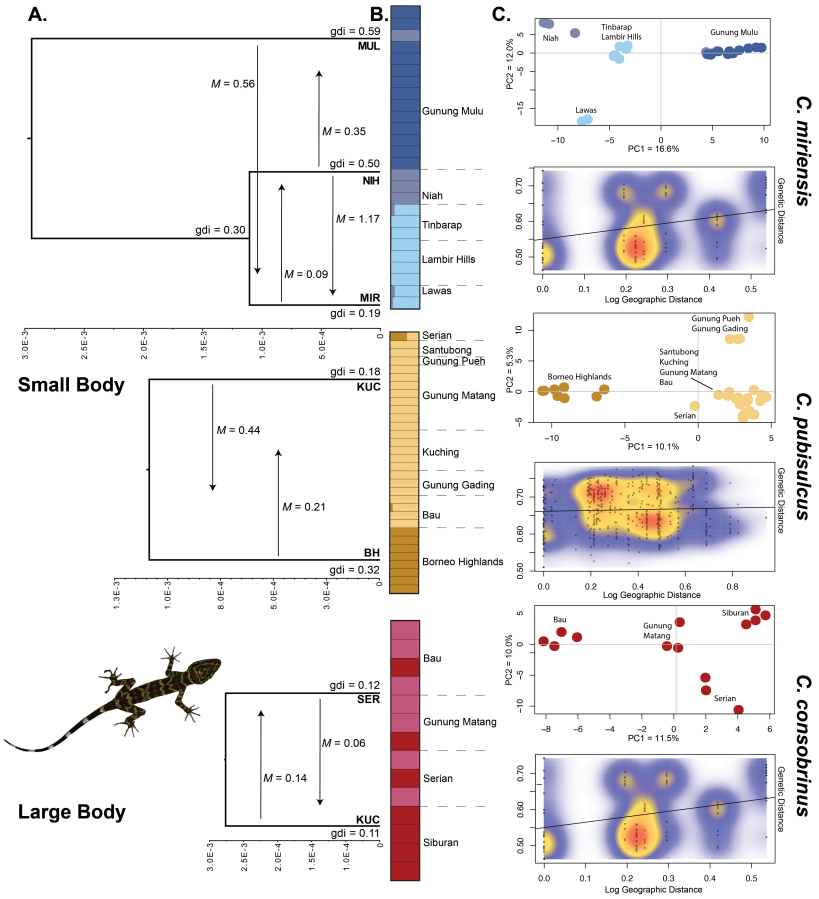

Figure 3.2: A) Species trees from for the three target species of *Cyrtodactylus* in this study. Arrows indicate the direction of gene flow, with the quantity of gene flow (M). Scale bars for respective species show the tau values. The population acronyms are shown on the tips are referred to throughout the text. B) Population structure analyses, showing the results from the K-value that corresponds with number of populations used for the species tree inference. C) Top: PCA of SNP data with colors corresponding to the results of the populations structure analyses. The colors on the PCA for *C. consobrinus* do not correspond with the ADMIXTURE, as both mtDNA and genomic data clearly infer a single species. Two populations are shown for the species tree to highlight the *gdi* and gene flow metrics. Percentages show the percent variation explained by the given principal component. Bottom: isolation-by-distance plots. Colors represent sampling density (white: low; red: high). Line shows the correlation between geographic and genetic distance.

Bibliography

Ingi Agnarsson and Laura J May-Collado. The phylogeny of *Cetartiodactyla*: the importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies. *Molecular Phylogenetics and Evolution*, 48(3):964–985, 2008.

David H Alexander and Kenneth Lange. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12:1–6, 2011.

Salah M Aljanabi and Iciar Martinez. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic acids research*, 25 (22):4692–4693, 1997.

Amy L Angert, Megan G Bontrager, and Jon Ågren. What do we really know about adaptation at range edges? *Annual Review of Ecology, Evolution, and Systematics*, 51:341–361, 2020.

Jorge Assis, Nelson Castilho Coelho, Filipe Alberto, Myriam Valero, Pete Raimondi, Dan Reed, and Ester Alvares Serrão. High and distinct range-edge genetic diversity despite local bottlenecks. *PLoS One*, 8(7):e68646, 2013.

Paul Backus. Use of patchy, early successional slope habitat along coastal sunfacing beaches by the Western Fence Lizard *Sceloporus occidentalis* at the species' northern geographic extreme. *Masters Thesis*, 2016.

Anthony J Barley and Robert C Thomson. Assessing the performance of dna barcoding using posterior predictive simulations. *Molecular Ecology*, 25(9): 1944–1957, 2016.

Anthony J Barley, Adrián Nieto-Montes de Oca, Norma L Manríquez-Morán, and Robert C Thomson. The evolutionary network of whiptail lizards reveals predictable outcomes of hybridization. *Science*, 377(6607):773–777, 2022.

Marta Bassitta, Richard P Brown, Ana Pérez-Cembranos, Valentín Pérez-Mellado, José A Castro, Antònia Picornell, and Cori Ramon. Genomic signatures of drift and selection driven by predation and human pressure in an insular lizard. *Scientific Reports*, 11(1):1–13, 2021.

Lucie A Bergeron, Søren Besenbacher, Jiao Zheng, Panyi Li, Mads Frost Bertelsen, Benoit Quintard, Joseph I Hoffman, Zhipeng Li, Judy St. Leger, Changwei Shao, et al. Evolution of the germline mutation rate across vertebrates. *Nature*, 615(7951):285–291, 2023.

Derek B Booth, Kathy Goetz Troost, John J Clague, and Richard B Waitt. The Cordilleran Ice Sheet. *Developments in Quaternary Sciences*, 1:17–43, 2003.

Remco Bouckaert, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A Suchard, Andrew Rambaut, and Alexei J Drummond. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4):e1003537, 2014.

Nassima M Bouzid, James W Archie, Roger A Anderson, Jared A Grummer, and Adam D Leaché. Evidence for ephemeral ring species formation during the diversification history of western fence lizards (*Sceloporus occidentalis*). *Molecular Ecology*, 31(2):620–631, 2022.

Larry L Bowman, Elizaveta S Kondrateva, Maxim A Timofeyev, and Lev Y Yampolsky. Temperature gradient affects differentiation of gene expression and SNP allele frequencies in the dominant Lake Baikal zooplankton species. *Molecular Ecology*, 27(11):2544–2559, 2018.

H. A. Brown. Studies of three translocated pouplations of the Western Fence Lizard in northern Puget Sound. *Northwest Science*, 66(2):127–127, 1992.

David Bryant and Vincent Moulton. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21(2):255–265, 2004.

David Bryant, Remco Bouckaert, Joseph Felsenstein, Noah A Rosenberg, and Arindam RoyChoudhury. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8):1917–1932, 2012.

Michael R Buchalski, Benjamin N Sacks, Daphne A Gille, Maria Cecilia T Penedo, Holly B Ernest, Scott A Morrison, and Walter M Boyce. Phylogeographic and population genetic structure of bighorn sheep (*Ovis canadensis*) in North American deserts. *Journal of Mammalogy*, 97(3):823–838, 2016.

Frank T Burbrink, Justin M Bernstein, Arianna Kuhn, Marcelo Gehara, and Sara Ruane. Ecological divergence and the history of gene flow in the nearctic milksnakes (*Lampropeltis triangulum* complex). *Systematic Biology*, 71(4): 839–858, 2022.

Kenneth P Burnham, David R Anderson, and Kathryn P Huyvaert. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1): 23–35, 2011.

Bernat Burriel-Carranza, Maria Estarellas, Gabriel Riaño, Adrián Talavera, Héctor Tejero-Cicuéndez, Johannes Els, and Salvador Carranza. Species boundaries to the limit: integrating species delimitation methods is critical to avoid taxonomic inflation in the case of the hajar banded ground gecko (*Trachydactylus hajarensis*). *Molecular Phylogenetics and Evolution*, page 107834, 2023.

Brett O Butler, Lydia L Smith, and Oscar Flores-Villela. Phylogeography and taxonomy of *Coleonyx elegans* gray 1845 (Squamata: Eublepharidae) in Mesoamerica: The Isthmus of Tehuantepec as an environmental barrier. *Molecular Phylogenetics and Evolution*, 178:107632, 2023.

Bryan C Carstens, Tara A Pelletier, Noah M Reid, and Jordan D Satler. How to fail at species delimitation. *Molecular Ecology*, 22(17):4369–4383, 2013.

E Anne Chambers, Thomas L Marshall, and David M Hillis. The importance of contact zones for distinguishing interspecific from intraspecific geographic variation. *Systematic Biology*, 72(2):357–371, 2022.

Kin Onn Chan, Alana M Alexander, L Lee Grismer, Yong-Chao Su, Jesse L Grismer, Evan SH Quah, and Rafe M Brown. Species delimitation with gene flow: a methodological comparison and population genomics approach to elucidate cryptic species boundaries in malaysian torrent frogs. *Molecular Ecology*, 26(20):5435–5450, 2017.

Sarah C Crews and Marshal Hedin. Studies of morphological and molecular phylogenetic divergence in spiders (Araneae: *Homalonychus*) from the American southwest, including divergence along the Baja California Peninsula. *Molecular Phylogenetics and Evolution*, 38(2):470–487, 2006.

Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A Albers, Eric Banks, Mark A DePristo, Robert E Handsaker, Gerton Lunter, Gabor T Marth, Stephen T Sherry, et al. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, 2011.

Hayden R Davis, Aaron M Bauer, Todd R Jackman, Izneil Nashriq, and Indraneil Das. Uncovering karst endemism within Borneo: two new *Cyrtodactylus* species from Sarawak, Malaysia. *Zootaxa*, 4614(2):331–352, 2019.

Hayden R Davis, Kin Onn Chan, Indraneil Das, Ian G Brennan, Benjamin R Karin, Todd R Jackman, Rafe M Brown, Djoko T Iskandar, Izneil Nashriq, L Lee Grismer, et al. Multilocus phylogeny of Bornean bent-toed geckos (Gekkonidae: *Cyrtodactylus*) reveals hidden diversity, taxonomic disarray, and novel biogeographic patterns. *Molecular Phylogenetics and Evolution*, 147:106785, 2020.

Hayden R Davis, Indraneil Das, Adam D Leaché, Benjamin R Karin, Ian G Brennan, Todd R Jackman, Izneil Nashriq, Kin Onn Chan, and Aaron M Bauer. Genetically diverse yet morphologically conserved: hidden diversity

revealed among Bornean geckos (Gekkonidae: *Cyrtodactylus*). *Journal of Zoological Systematics and Evolutionary Research*, 59(5):1115–1135, 2021.

Hayden R Davis, Izneil Nashriq, Kyra S Woytek, Shanelle A Wikramanayake, Aaron M Bauer, Benjamin R Karin, Ian G Brennan, Djoko T Iskandar, and Indraneil Das. Genomic analysis of Bornean geckos (Gekkonidae: *Cyrtodactylus*) reveals need for updated taxonomy. *Zoologica Scripta*, 52(3):249–263, 2023.

Juan D Daza, Aaron M Bauer, and Eric D Snively. On the fossil record of the Gekkota. *The Anatomical Record*, 297(3):433–462, 2014.

Kevin De Queiroz. The general lineage concept of species, species criteria, and the process of speciation. *Endless Forms: Species and Speciation*, page 57–78, 1998.

Kathleen Semple Delaney, Seth PD Riley, and Robert N Fisher. A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. *Plos ONE*, 5(9):e12767, 2010.

Devon A DeRaad, Jenna M McCullough, Lucas H DeCicco, Paul M Hime, Leo Joseph, Michael J Andersen, and Robert G Moyle. Mitonuclear discordance results from incomplete lineage sorting, with no detectable evidence for gene flow, in a rapid radiation of *Todiramphus* kingfishers. *Molecular Ecology*, 32 (17):4844–4862, 2023.

Shahan Derkarabetian, James Starrett, and Marshal Hedin. Using natural history to guide supervised machine learning for cryptic species delimitation with genetic data. *Frontiers in Zoology*, 19(1):8, 2022.

Simone Des Roches, Michael A Bell, and Eric P Palkovacs. Climate-driven habitat change causes evolution in Threespine Stickleback. *Global Change Biology*, 26(2):597–606, 2020.

Megan N Dethier, Jason D Toft, and Hugh Shipman. Shoreline armoring in an inland sea: science-based recommendations for policy implementation. *Conservation Letters*, 10(5):626–633, 2017.

James R Dixon. *Coleonyx variegatus*. *Catalogue of American Amphbians and Reptiles*, 96:1–4, 1970.

Greer A Dolby, Rebecca J Dorsey, and Matthew R Graham. A legacy of geoclimatic complexity and genetic divergence along the lower Colorado River: Insights from the geological record and 33 desert-adapted animals. *Journal of Biogeography*, 46(11):2479–2505, 2019.

Jenifer E Dugan, David M Hubbard, Iván F Rodil, David L Revell, and Stephen Schroeter. Ecological effects of coastal armoring on sandy beaches. *Marine Ecology*, 29:160–170, 2008.

Jenifer E Dugan, Kyle A Emery, Merryl Alber, Clark R Alexander, James E Byers, AM Gehman, Natalie McLenaghan, and Sarah E Sojka. Generalizing ecological effects of shoreline armoring across soft sediment environments. *Estuaries and Coasts*, 41(1):180–196, 2018.

Deren AR Eaton and Isaac Overcast. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36(8):2592–2594, 2020.

CG Eckert, KE Samis, and SC Lougheed. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, 17(5):1170–1188, 2008.

Ron Farrell, Gavin Hanke, and David Veljacic. First verified sighting of a Western Fence Lizard (*Sceloporus occidentalis*) in British Columbia, Canada. *The Canadian Field-Naturalist*, 134(3):210–212, 2020.

Thomas J Firneno Jr, Justin R O'Neill, Daniel M Portik, Alyson H Emery, Josiah H Townsend, and Matthew K Fujita. Finding complexity in complexes: assessing the causes of mitonuclear discordance in a problematic species complex of Mesoamerican toads. *Molecular Ecology*, 29(18):3543–3559, 2020.

Cene Fišer, Christopher T Robinson, and Florian Malard. Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27 (3):613–635, 2018.

Tomáš Flouri, Xiyun Jiao, Jun Huang, Bruce Rannala, and Ziheng Yang. Efficient Bayesian inference under the multispecies coalescent with migration. *Proceedings of the National Academy of Sciences*, 120(44):e2310708120, 2023.

Tomochika Fujisawa and Timothy G Barraclough. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology*, 62(5):707–724, 2013.

Trenton WJ Garner, Peter B Pearman, and Sonia Angelone. Genetic diversity across a vertebrate species' range: a test of the central–peripheral hypothesis. *Molecular Ecology*, 13(5):1047–1053, 2004.

Genomic Resources Development Consortium, Wolfgang Arthofer, BL Banbury, Miguel Carneiro, Francesco Cicconardi, Thomas F Duda, RB Harris, David S Kang, Adam D Leaché, et al. Genomic resources notes accepted 1 August 2014–30 September 2014. *Molecular Ecology Resources*, 15(1):228–229, 2015.

Andrew D Gottscho, Dustin A Wood, Amy G Vandergast, Julio Lemos-Espinal, John Gatesy, and Tod W Reeder. Lineage diversification of fringe-toed lizards (Phrynosomatidae: *Uma notata* complex) in the Colorado Desert: delimiting species in the presence of gene flow. *Molecular Phylogenetics and Evolution*, 106:103–117, 2017.

Andrew D Gottscho, Daniel G Mulcahy, Adam D Leaché, Kevin de Queiroz, and Robert E Lovich. Population genomics of Flat-tailed Horned Lizards (*Phrynosoma mcallii*) informs conservation and management across a fragmented Colorado Desert landscape. *Molecular Ecology*, page e17308, 2024.

L Lee Grismer, Perry L Wood Jr, Minh Duc Le, Evan SH Quah, and Jesse L Grismer. Evolution of habitat preference in 243 species of bent-toed geckos (genus *Cyrtodactylus* gray, 1827) with a discussion of karst habitat conservation. *Ecology and Evolution*, 10(24):13717–13730, 2020.

L Lee Grismer, Nikolay A Poyarkov, Evan SH Quah, Jesse L Grismer, and Perry L Wood Jr. The biogeography of bent-toed geckos, *Cyrtodactylus* (Squamata: Gekkonidae). *PeerJ*, 10:e13153, 2022.

Jared A Grummer, Robert W Bryson Jr, and Tod W Reeder. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology*, 63 (2):119–133, 2014.

Stéphane Guindon, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3):307–321, 2010.

David J Hafner, Brett R Riddle, P Upchurch, AJ McGowan, and SCS Slater. Boundaries and barriers of North American warm deserts. *Palaeogeography and Palaeobiogeography Biodiversity in Space and Time*, pages 75–114, 2011.

Sean M Harrington, Bradford D Hollingsworth, Timothy E Higham, and Tod W Reeder. Pleistocene climatic fluctuations drive isolation and secondary contact in the Red Diamond Rattlesnake (*Crotalus ruber*) in Baja California. *Journal of Biogeography*, 45(1):64–75, 2018.

Ralph A Haugerud. Deglaciation of the Puget Lowland, Washington. *Untangling the Quaternary period - A legacy of Stephen C. Porter: Geological Society of America Special Paper*, 548, 2020.

Tracy A Heath, Shannon M Hedtke, and David M Hillis. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, 46(3):239–257, 2008.

Marshal Hedin, Dave Carlson, and Fred Coyle. Sky island diversification meets the multispecies coalescent–divergence in the Spruce-fir Moss Spider (*Microhexura montivaga*, araneae, mygalomorphae) on the highest peaks of southern appalachia. *Molecular Ecology*, 24(13):3467–3484, 2015.

Michael J Hickerson, Christopher P Meyer, and Craig Moritz. DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology*, 55(5):729–739, 2006.

Diep Thi Hoang, Olga Chernomor, Arndt Von Haeseler, Bui Quang Minh, and Le Sy Vinh. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2):518–522, 2018.

Ary A Hoffmann and Mark W Blows. Species borders: ecological and evolutionary perspectives. *Trends in Ecology & Evolution*, 9(6):223–227, 1994.

Jun Huang, Jeremy Bennett, Tomáš Flouri, Adam D Leaché, and Ziheng Yang. Phase resolution of heterozygous sites in diploid genomes is important to phylogenomic analysis under the multispecies coalescent model. *Systematic Biology*, 71(2):334–352, 2022.

Daniel H Huson. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics (Oxford, England)*, 14(1):68–73, 1998.

Daniel H Huson, Tobias Kloepper, and David Bryant. Splitstree 4.0-Computation of phylogenetic trees and networks. *Bioinformatics*, 14:68–73, 2008.

Nathan D Jackson, Bryan C Carstens, Ariadna E Morales, and Brian C O'Meara. Species delimitation with gene flow. *Systematic Biology*, 66(5): 799–812, 2017a.

Nathan D Jackson, Ariadna E Morales, Bryan C Carstens, and Brian C O'Meara. PHRAPL: phylogeographic inference using approximate likelihoods. *Systematic Biology*, 66(6):1045–1053, 2017b.

EW Jameson Jr and Allen Allison. Fat and breeding cycles in two montane populations of *Sceloporus occidentalis* (Reptilia, Lacertilia, Iguanidae). *Journal of Herpetology*, pages 211–220, 1976.

Jiayi Ji, Donavan J Jackson, Adam D Leaché, and Ziheng Yang. Power of Bayesian and heuristic tests to detect cross-species introgression with reference to gene flow in the *Tamias quadrivittatus* group of North American chipmunks. *Systematic Biology*, 72(2):446–465, 2023.

Xiyun Jiao, Tomáš Flouri, and Ziheng Yang. Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. *National Science Review*, 8(12):nwab127, 2021.

J Johannessen, A MacLennan, A Blue, J Waggoner, S Williams, W Gerstel, R Barnard, R Carman, and H Shipman. Marine shoreline design guidelines. Technical report, Washington Department of Fish and Wildlife, Olympia, WA, 2014.

Clifford Ray Johnson. The diet of the Pacific Fence Lizard, *Sceloporus occidentalis occidentalis* (Baird and Girard), from northern California. *Herpetologica*, 21(2):114–117, 1965.

Thibaut Jombart. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11):1403–1405, 2008.

Thibaut Jombart and Ismaïl Ahmed. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21):3070–3071, 2011.

Thibaut Jombart, Sébastien Devillard, and François Balloux. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1):1–15, 2010.

Julien Jouganous, Will Long, Aaron P Ragsdale, and Simon Gravel. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics*, 206(3):1549–1567, 2017.

Subha Kalyaanamoorthy, Bui Quang Minh, Thomas KF Wong, Arndt Von Haeseler, and Lars S Jermiin. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6):587–589, 2017.

Paschalia Kapli, Sarah Lutteropp, Jiajie Zhang, Kassian Kobert, Pavlos Pavlidis, Alexandros Stamatakis, and Tomas Flouri. Multi-rate poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, 33(11):1630–1638, 2017.

Kazutaka Katoh, Kazuharu Misawa, Kei-ichi Kuma, and Takashi Miyata. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30(14):3059–3066, 2002.

Holly J Kilvitis, Haley Hanson, Aaron W Schrey, and Lynn B Martin. Epigenetic potential as a mechanism of phenotypic plasticity in vertebrate range expansions. *Integrative and Comparative Biology*, 57(2):385–395, 2017.

Laurence Monroe Klauber. The geckos of the genus *Coleonyx*: With descriptions of new subspecies. San Diego Society of Natural History, 1945.

Naama M Kopelman, Jonathan Mayzel, Mattias Jakobsson, Noah A Rosenberg, and Itay Mayrose. Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, 15(5):1179–1191, 2015.

Daniel Kornai, Tomas Flouri, and Ziheng Yang. Hierarchical heuristic species delimitation under the multispecies coalescent model with migration. *bioRxiv*, pages 2023–09, 2023.

Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6):1547, 2018.

Hmar Tlawmte Lalremsanga, Zosiamliana Colney, Mathipi Vabeiryureilai, Fanai Malsawmdawngliana, Sanath Chandra Bohra, Lal Biakzuala, Lal Muansanga, Madhurima Das, and Jayaditya Purkayastha. It‘s all in the name: another

new *Cyrtodactylus* gray (squamata: Gekkonidae) from northern Mizoram, North-east India. *Zootaxa*, 5369(4):553–575, 2023.

Kathryn M Langin, T Scott Sillett, W Chris Funk, Scott A Morrison, Michelle A Desrosiers, and Cameron K Ghalambor. Islands within an island: repeated adaptive divergence in a single population. *Evolution*, 69(3):653–665, 2015.

Daniel J Lawson, Lucy Van Dorp, and Daniel Falush. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature communications*, 9(1):1–11, 2018.

Adam D Leaché, Sarah C Crews, and Michael J Hickerson. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. *Biology Letters*, 3(6):646–650, 2007.

Adam D Leaché, Der-Shing Helmer, and Craig Moritz. Phenotypic evolution in high-elevation populations of Western Fence Lizards (*Sceloporus occidentalis*) in the Sierra Nevada Mountains. *Biological Journal of the Linnean Society*, 100(3):630–641, 2010.

Adam D Leaché, Matthew K Fujita, Vladimir N Minin, and Remco R Bouckaert. Species delimitation using genome-wide SNP data. *Systematic Biology*, 63(4): 534–542, 2014.

Adam D Leaché, Tianqi Zhu, Bruce Rannala, and Ziheng Yang. The spectre of too many species. *Systematic Biology*, 68(1):168–181, 2019.

Adam D Leaché, Hayden R Davis, Sonal Singhal, Matthew K Fujita, Megan E Lahti, and Kelly R Zamudio. Phylogenomic assessment of biodiversity using a reference-based taxonomy: an example with horned lizards (*Phrynosoma*). *Frontiers in Ecology and Evolution*, 9:678110, 2021.

Adam D Leaché, Hayden R Davis, Chris R Feldman, Matthew K Fujita, and Sonal Singhal. Repeated patterns of reptile diversification in western North America supported by the Northern Alligator Lizard (*Elgaria coerulea*). *Journal of Heredity*, 115(1):57–71, 2024.

Dean H Leavitt, Bradford D Hollingsworth, Robert N Fisher, and Tod W Reeder. Introgression obscures lineage boundaries and phylogeographic history in the Western Banded Gecko, *Coleonyx variegatus* (Squamata: Eublepharidae). *Zoological Journal of the Linnean Society*, 190(1):181–226, 2020.

Timothy S Lee, Jason D Toft, Jeffery R Cordell, Megan N Dethier, Jeffrey W Adams, and Ryan P Kelly. Quantifying the effectiveness of shoreline armoring removal on coastal biota of Puget Sound. *PeerJ*, 6:e4275, 2018.

Johan Lindell, Fausto R Méndez-de la Cruz, and Robert W Murphy. Deep genealogical history without population differentiation: discordance between mtdna and allozyme divergence in the Zebra-tailed Lizard (*Callisaurus draconoides*). *Molecular Phylogenetics and Evolution*, 36(3):682–694, 2005.

Johan Lindell, Fausto R Mendez-De La Cruz, and Robert W Murphy. Deep biogeographical history and cytonuclear discordance in the Black-tailed Brush Lizard (*Urosaurus nigricaudus*) of Baja California. *Biological Journal of the Linnean Society*, 94(1):89–104, 2008.

Michael Lynch, Farhan Ali, Tongtong Lin, Yaohai Wang, Jiahao Ni, and Hongan Long. The divergence of mutation rates and spectra across the Tree of Life. *EMBO Reports*, 24(10):e57561, 2023.

Wayne P Maddison and Jeannette Whitton. The species as a reproductive community emerging from the past. *Bulletin of the Society of Systematic Biologists*, 2(1):1–35, 2023.

Milan Malinsky, Hannes Svardal, Alexandra M Tyers, Eric A Miska, Martin J Genner, George F Turner, and Richard Durbin. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology & Evolution*, 2(12):1940–1955, 2018.

Ani Manichaikul, Josyf C Mychaleckyj, Stephen S Rich, Kathy Daly, Michèle Sale, and Wei-Min Chen. Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22):2867–2873, 2010.

Nicholas A Mason, Nicholas K Fletcher, Brian A Gill, W Chris Funk, and Kelly R Zamudio. Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity*, 18 (3):269–280, 2020.

Rob Massatti and L Lacey Knowles. The historical context of contemporary climatic adaptation: a case study in the climatically dynamic and environmentally complex southwestern United States. *Ecography*, 43(5):735–746, 2020.

Michael Matschiner. Species tree inference with SNP data. In *Plant Comparative Genomics*, pages 23–44. Springer, 2022.

RL Mayden. A hierarchy of species concepts: The denouncement of the saga of the species problem. *Species: The Unit of Biodiversity*, 1997.

Samuel M McGinnis. *Sceloporus occidentalis*: preferred body temperature of the Western Fence Lizard. *Science*, 152(3725):1090–1091, 1966.

Matthew R McTernan. How do lizards us behavior and physiology to inhabit different climate zones. *Masters Thesis*, 2017.

Chloe S Mikles, Stepfanie M Aguillon, Yvonne L Chan, Peter Arcese, Phred M Benham, Irby J Lovette, and Jennifer Walsh. Genomic differentiation and local adaptation on a microgeographic scale in a resident songbird. *Molecular Ecology*, 29(22):4295–4307, 2020.

Lindsay S Miles, L Ruth Rivkin, Marc TJ Johnson, Jason Munshi-South, and Brian C Verrelli. Gene flow and genetic drift in urban environments. *Molecular Ecology*, 28(18):4138–4151, 2019.

Aryeh H Miller, George R Zug, Guinevere OU Wogan, Justin L Lee, and Daniel G Mulcahy. Phylogeny, diversity, and distribution of *Micryletta* (Anura: Microhylidae) in Myanmar. *Ichthyology & Herpetology*, 109(1):245–257, 2021.

Philip W Mote, Eric P Salathé, Valérie Dulière, and Emily Jump. Scenarios of future climate change for the Pacific Northwest. Report, Climate Impacts Group, University of Washington, 2008.

Daniel G Mulcahy. Phylogeography and species boundaries of the western North American Nightsnake (*Hypsiglena torquata*): revisiting the subspecies concept. *Molecular phylogenetics and evolution*, 46(3):1095–1115, 2008.

Ahmed Ragab Nabhan and Indra Neil Sarkar. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Briefings in Bioinformatics*, 13(1):122–134, 2012.

Shun-Ichiro Naomi. On the integrated frameworks of species concepts: Mayden's hierarchy of species concepts and de queiroz's unified concept of species. *Journal of Zoological Systematics and Evolutionary Research*, 49(3):177–184, 2011.

Masatoshi Nei. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12):3321–3323, 1973.

Lam-Tung Nguyen, Heiko A Schmidt, Arndt Von Haeseler, and Bui Quang Minh. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1): 268–274, 2015.

Jamie R Oaks, Cameron D Siler, and Rafe M Brown. The comparative biogeography of Philippine geckos challenges predictions from a paradigm of climate-driven vicariant diversification across an island archipelago. *Evolution*, 73(6): 1151–1167, 2019.

Brian C O'Meara. New heuristic methods for joint species delimitation and species tree inference. *Systematic Biology*, 59(1):59–73, 2010.

Vera Opatova, Kellie Bourguignon, and Jason E Bond. Species delimitation with limited sampling: An example from rare trapdoor spider genus *Cyclocosmia* (Mygalomorphae, Halonoproctidae). *Molecular Ecology Resources*, 2023.

Robert P Owen. A list of the reptiles of Washington. *Copeia*, 1940(3):169–172, 1940.

Nick Patterson, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. Ancient admixture in human history. *Genetics*, 192(3):1065–1093, 2012.

Carlos J Pavón-Vázquez, Qaantah Rana, Keaka Farleigh, Erika Crispo, Mimi Zeng, Jeevanie Liliah, Daniel Mulcahy, Alfredo Ascanio, Tereza Jezkova, Adam D Leaché, et al. Gene flow and isolation in the arid nearctic revealed by genomic analyses of Desert Spiny Lizards. *Systematic Biology*, page syae001, 2024.

Gareth A Pearson, Asuncion Lago-Leston, and Catarina Mota. Frayed at the edges: selective pressure and adaptive response to abiotic stressors are mismatched in low diversity edge populations. *Journal of Ecology*, 97(3):450–462, 2009.

Carmen del R Pedraza-Marrón, Raimundo Silva, Jonathan Deeds, Steven M Van Belleghem, Alicia Mastretta-Yanes, Omar Domínguez-Domínguez, Rafael A Rivero-Vega, Loretta Lutackas, Debra Murie, Daryl Parkyn, et al. Genomics overrules mitochondrial dna, siding with morphology on a controversial case of species delimitation. *Proceedings of the Royal Society B*, 286 (1900):20182924, 2019.

Ricardo J Pereira and Sonal Singhal. A lizard with two tales: what diversification within *Sceloporus occidentalis* teaches us about species formation, 2022.

Blair W Perry, Daren C Card, Joel W McGlothlin, Giulia IM Pasquesi, Richard H Adams, Drew R Schield, Nicole R Hales, Andrew B Corbin, Jeffery P Demuth, Federico G Hoffmann, et al. Molecular adaptations for sensing and securing prey and insight into amniote genome diversity from the garter snake genome. *Genome Biology and Evolution*, 10(8):2110–2129, 2018.

Brant K Peterson, Jesse N Weber, Emily H Kay, Heidi S Fisher, and Hopi E Hoekstra. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS ONE*, 7(5): e37135, 2012.

Desislava Petkova, John Novembre, and Matthew Stephens. Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, 48(1):94–100, 2016.

Daniel M Portik, Adam D Leaché, Danielle Rivera, Michael F Barej, Marius Burger, Mareike Hirschfeld, Mark-Oliver Rödel, David C Blackburn, and Matthew K Fujita. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Molecular Ecology*, 26(19):5245–5263, 2017.

Sean D Powers, Matthew R McTernan, Donald R Powers, and Roger A Anderson. Energetic consequences for a northern, range-edge lizard population. *Copeia*, 106(3):468–476, 2018.

Nicolas Puillandre, Sophie Brouillet, and Guillaume Achaz. ASAP: assemble species by automatic partitioning. *Molecular Ecology Resources*, 21(2):609–620, 2021.

Breanna J Putman, Maria Gasca, Daniel T Blumstein, and Gregory B Pauly. Downsizing for downtown: limb lengths, toe lengths, and scale counts decrease with urbanization in Western Fence Lizards (*Sceloporus occidentalis*). *Urban Ecosystems*, 22(6):1071–1081, 2019.

Alex Pyron, Anvith Kakkera, David Beamer, and Kyle O'Connell. Discerning structure versus speciation in phylogeographic analysis of Seepage Salamanders (*Desmognathus aeneus*) using demography, environment, geography, and phenotype. *Molecular Ecology*, 2023.

Development Team QGIS. Qgis geographic information system. *Open source geospatial Foundation project*, 2015.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

Jorge L Ramirez, Paola Valdivia, Ulises Rosas-Puchuri, and Nereida L Valdivia. SPdel: A pipeline to compare and visualize species delimitation methods for single-locus datasets. *Molecular Ecology Resources*, 23(8):1959–1965, 2023.

Sean B Reilly, Alexander L Stubbs, Benjamin R Karin, Evy Arida, Umilaela Arifin, Amir Hamidy, Hinrich Kaiser, Ke Bi, Awal Riyanto, Djoko T Iskandar, et al. Bewildering biogeography: waves of dispersal and diversification across southern Wallacea by bent-toed geckos (genus: *Cyrtodactylus*). *Molecular Phylogenetics and Evolution*, page 107853, 2023.

Jonathan L Richardson, Mark C Urban, Daniel I Bolnick, and David K Skelly. Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution*, 29(3):165–176, 2014.

Brett R Riddle and David J Hafner. A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. *Journal of Arid Environments*, 66 (3):435–461, 2006.

Brett R Riddle, David J Hafner, Lois F Alexander, and Jef R Jaeger. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences*, 97(26):14438–14443, 2000.

JL Riedel. Deglaciation of the North Cascade Range, Washington and British Columbia, from the last glacial maximum to the Holocene. *Cuadernos de investigación geográfica/Geographical Research Letters*, (43):467–496, 2017.

Camille Roux, Christelle Fraisse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier, and Nicolas Bierne. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14(12):e2000234, 2016.

Howard D Rundle and Patrik Nosil. Ecological speciation. *Ecology Letters*, 8 (3):336–352, 2005.

Jason P Sexton, Patrick J McIntyre, Amy L Angert, and Kevin J Rice. Evolution and ecology of species range limits. *Annu. Rev. Ecol. Evol. Syst.*, 40:415–436, 2009.

Barry Sinervo. Evolution of thermal physiology and growth rate between populations of the Western Fence Lizard (*Sceloporus occidentalis*). *Oecologia*, 83 (2):228–237, 1990.

Barry Sinervo and Stephen C Adolph. Growth plasticity and thermal opportunity in *Sceloporus* lizards. *Ecology*, 75(3):776–790, 1994.

Barry Sinervo and Raymond B Huey. Allometric engineering: an experimental test of the causes of interpopulational differences in performance. *Science*, 248(4959):1106–1109, 1990.

Sonal Singhal, Conrad J Hoskin, Patrick Couper, Sally Potter, and Craig Moritz. A framework for resolving cryptic species: a case study from the lizards of the Australian wet tropics. *Systematic Biology*, 67(6):1061–1075, 2018.

Jack W Sites, James W Archie, Charles J Cole, and Oscar Flores-Villela. A review of phylogenetic hypotheses for lizards of the genus *Sceloporus* (Phrynosomatidae): implications for ecological and evolutionary studies. *Bulletin of the AMNH*, 213, 1992.

J R Slater. Distribution of Washington reptiles. *Occasional Papers, Department of Biology, University of Puget Sound*, 24:212–233, 1963.

Hobart Muir Smith. *The Mexican and Central American lizards of the genus Sceloporus*, volume 26. Field Museum of Natural History, 1939.

Alexandros Stamatakis. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313, 2014.

Madlen Stange, Marcelo R Sánchez-Villagra, Walter Salzburger, and Michael Matschiner. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Systematic Biology*, 67(4):681–699, 2018.

Sean Stankowski and Mark Ravinet. Quantifying the use of species concepts. *Current Biology*, 31(9):R428–R429, 2021.

Glen Stecher, Koichiro Tamura, and Sudhir Kumar. Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular Biology and Evolution*, 37 (4):1237–1239, 2020.

Torsten H Struck, Jeffrey L Feder, Mika Bendiksby, Siri Birkeland, José Cerca, Vladimir I Gusarov, Sonja Kistenich, Karl-Henrik Larsson, Lee Hsiang Liow, Michael D Nowak, et al. Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, 33(3):153–163, 2018.

Jeet Sukumaran, Mark T Holder, and L Lacey Knowles. Incorporating the speciation process into species delimitation. *PLoS Computational Biology*, 17 (5):e1008924, 2021.

Alexandra Sumarli, Bradford D Hollingsworth, Jorge H Valdez-Villavicencio, and Tod W Reeder. Phylogenomic data reveal cryptic diversity and deep phylogeographical structure within the Common Chuckwalla, *Sauromalus ater* (Squamata: Iguanidae). *Biological Journal of the Linnean Society*, 141(4): 572–588, 2023.

Korkhwan Termprayoon, Attapol Rujirawan, L Lee Grismer, Perry L Wood Jr, and Anchalee Aowphol. Two new karst-adapted species in the *Cyrtodactylus pulchellus* group (Reptilia, Gekkonidae) from southern Thailand. *ZooKeys*, 1179:313–352, 2023.

Yuttapong Thawornwattana, Fernando Seixas, Ziheng Yang, and James Mallet. Major patterns in the introgression history of Heliconius butterflies. *ELife*, 12:RP90656, 2023.

George P Tiley, Tomáš Flouri, Xiyun Jiao, Jelmer W Poelstra, Bo Xu, Tianqi Zhu, Bruce Rannala, Anne D Yoder, and Ziheng Yang. Estimation of species divergence times in presence of cross-species gene flow. *Systematic Biology*, 72(4):820–836, 2023.

David PL Toews and Alan Brelsford. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16):3907–3930, 2012.

Jason D Toft, Megan N Dethier, Emily R Howe, Emily V Buckner, and Jeffery R Cordell. Effectiveness of living shorelines in the Salish Sea. *Ecological Engineering*, 167:106255, 2021.

Joyce S Tsuji. Seasonal profiles of standard metabolic rate of lizards (*Sceloporus occidentalis*) in relation to latitude. *Physiological Zoology*, 61(3):230–240, 1988.

John Van Denburgh. *The Reptiles of Western North America: An Account of the Species Known to Inhabit California and Oregon, Washington, Idaho, Utah, Nevada, Arizona, British Columbia, Sonora and Lower California.* Number 10. California Academy of Sciences, 1922.

Alexandra Jansen van Rensburg, Maria Cortazar-Chinarro, Annsi Laurila, and Josh Van Buskirk. Adaptive genomic variation associated with environmental gradients along a latitudinal cline in *Rana temporaria*. *BioRxiv*, page 427872, 2018.

Miguel Vences, Aurélien Miralles, and Christophe Dufresnes. Next-generation species delimitation and taxonomy: Implications for biogeography. *Journal of Biogeography*, 2024.

Yiguan Wang and Darren J Obbard. Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis. *Evolution Letters*, 7 (4):216–226, 2023.

Aundrea Westfall, Rory S Telemeco, Mariana B Grizante, Damien S Waits, Amanda D Clark, Dasia Y Simpson, Randy L Klabacka, Alexis P Sullivan, George H Perry, Jamie R Oaks, Michael W Sears, Christian L Cox, Robert M Cox, Matt E Gifford, Henry B John-Alder, Tracy Langkilde, Michael J Angilletta Jr, Adam D Leaché, Marc A Tollis, Kenro Kusumi, and Tonia S Schwartz. A chromosome-level genome assembly for the Eastern Fence Lizard (*Sceloporus undulatus*), a reptile model for physiological and evolutionary ecology. *GigaScience*, 2021.

Edward O Wiley. Vicariance biogeography. *Annual Review of Ecology and Systematics*, 19(1):513–542, 1988.

Dustin A Wood, Amy G Vandergast, Kelly R Barr, Rich D Inman, Todd C Esque, Kenneth E Nussear, and Robert N Fisher. Comparative phylogeography reveals deep lineages and regional evolutionary hotspots in the Mojave and Sonoran Deserts. *Diversity and Distributions*, 19(7):722–737, 2013.

Perry L Wood Jr, Matthew P Heinicke, Todd R Jackman, and Aaron M Bauer. Phylogeny of bent-toed geckos (*Cyrtodactylus*) reveals a west to east pattern of diversification. *Molecular Phylogenetics and Evolution*, 65(3):992–1003, 2012.

Perry L Wood Jr, L Lee Grismer, Mohd Abdul Muin, Shahrul Anuar, Jamie R Oaks, et al. A new potentially endangered limestone-associated bent-toed gecko of the *Cyrtodactylus pulchellus* (Squamata: Gekkonidae) complex from northern Peninsular Malaysia. *Zootaxa*, 4751(3):437–460, 2020.

Ziheng Yang. The BPP program for species tree estimation and species delimitation. *Current Zoology*, 61(5):854–865, 2015.

Ziheng Yang and Bruce Rannala. Bayesian species identification under the multispecies coalescent provides significant improvements to dna barcoding analyses. *Molecular Ecology*, 26(11):3028–3036, 2017.

Jiajie Zhang, Paschalia Kapli, Pavlos Pavlidis, and Alexandros Stamatakis. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22):2869–2876, 2013.

Stephen M Zozaya, Megan Higgie, Craig Moritz, and Conrad J Hoskin. Are pheromones key to unlocking cryptic lizard diversity? *The American Naturalist*, 194(2):168–182, 2019.

Stephen M Zozaya, Luisa C Teasdale, Craig Moritz, Megan Higgie, and Conrad J Hoskin. Composition of a chemical signalling trait varies with phylogeny and precipitation across an Australian lizard radiation. *Journal of Evolutionary Biology*, 35(7):919–933, 2022.