Methods and Results Outline

INTRO?

The goal was to obtain the species predicted distribution to be able to assess broad global climatic patterns that might be driving microhabitats in lungless salamanders and specifically the arboreal microhabitat.

Inspiration:

You need only share the *decisions* you made when doing the methods.

To test this assertion, we employed two methodological frameworks of comparing the climactic differences between arboreal and terrestrial species ranges. First, we extract the climactic variables from each species ranges to give each species a summary of climatic variables. Using these data, we conduct phylogenetic ANOVAs and model comparison to test whether and how arboreal and terrestrial species differ in the climate they experience. Second, we employ ecological niche modeling methods to define the \_\_\_\_\_ Using these two approaches, we are able to show from several different perspectives that arboreal species inhabit a significantly different area of climate space. This has implications\_\_\_\_.

**Methods**

Phylogeny

Microhabitat Use

Species Polygons [ Lauren’s verbiage]

To obtain species polygons from which we could extract climate variables, we downloaded the predicted geographic species ranges for all new world species available on the International Union for Conservation of Nature’s (IUCN) Red List of Threatened Species database (# species). We supplemented the available species polygons by constructing similar species ranges, using reliable occurrence localities from VertNet (website/citation) which increased our sample size to \_\_\_\_ new world species.

To utilize these extra points from VertNet, we mirrored IUCN protocol to recreate species polygons using the VertNet occurrence data. We were able to construct the distributional polygon ranges by similar standards as does IUCN, given the data available by making the Extent of Occurrence (EOO), Area of Occupancy (AOO) and overall the Extent of Suitable Habitat (ESH). By mirroring the EOO function, we were using the alpha.hull as the range method which allowed us to change the alpha-convex hull value that reflected the connectivity between the points. This resulted in a polygon that represented the EOO in square kilometers that could be manipulated to exclude regions that did not include species presence, such as the ocean or industrial topologies. IUCN suggests that 2 is a good starting point for an alpha hull value but may need to be reduced to find a compromise between minimizing the potential bias associated with incomplete sampling of outlying occurrences and minimizing the departure from a convex hull. We found that an alpha value of 1 was suitable for our distributional range maps. We combined the study-made polygons with the IUCN polygons to get one distribution for each microhabitat. These polygons were used to assess the breadth of climate within them to see the range of conditions each species is likely to experience within their predicted distribution.

Data for PCA

The different types of data: Worldclim (<http://www.worldclim.org/>) variables were derived from this website (<https://deepblue.lib.umich.edu/data/concern/data_sets/gt54kn05f>).

Bioclim variables were derived (<https://www.worldclim.org/bioclim>) from this website specifically (<http://worldclim.org/version2>).

Cloud cover from IPCC were derived from this website (<https://visibleearth.nasa.gov/view.php?id=85843>). Paper that used it before <https://www.sciencedirect.com/science/article/pii/S0304380016304665#bib0195> .

Elevation is derived from 2.5 arcminute altitude grids from the Worldclim v1.4 data repository (http://biogeo.ucdavis.edu/data/climate/worldclim/ 1\_4/grid/cur/alt\_2-5m\_bil.zip).

From all these variables, the ones we ended up using for ENM were: elevation, annual mean temperature, Max temp of warmest month, min temp of coldest month, annual precipitation, precipitation of wettest quarter, precipitation of driest quarter, climate moisture index, PET driest quarter, PET wettest quarter, and cloud cover.

Min, 1q, mean, 3q max of lots of things

Polygons

PCA, and ANOVA Analyses

ANOVApgls for PCA

Loadings PCA axes

ANOVA on those variables

Arb-Terr

Data for Ecological Niche Modeling [Lauren will work on this verbiage]

CLIMATE VARIABLES - And the world climate variable details – how much will be discussed in Erica’s section and how much is needed here?

WorldClim and BioClim variables shown to be important, and *what they actually mean*.

We clipped the raster files to the extent of the arboreal and terrestrial polygons and extracted *climatic information* for each microhabitat type within their distributional polygon. This was the climate data that we chose to use in the maxent model.

POINT DATA -

The polygons were used as a proxy to make the point data for the Environmental Niche Model (ENM) because the point/occurrence data was not available through IUCN that specified which points were used to make their polygons. Under the assumption that the entire range distribution equally represented where the species was able to go, we wanted to use that extent for the global microhabitat climate models. To use in a suitable model, pseudo-occurrence sample points from polygons were generated (1) that are representative of the predicted distributional range for each species. The point data were created by first gridding the polygon by a given resolution and then taking the centroid coordinates of each gridded cell to become our sample points for each microhabitat. This heavily depended on the resolution used and resulted in a range of possible points. Using the resolution that our climate data was at, gave us too many points to use in the Maxent model and caused the model to overpredict based on those points. The opposite caution would be to use too few points and then risk that the model overfit. To alleviate these problems, we tested the impact of the resolution parameters and the resulting Area Under the Curve (AUC) score of each maxent model resolution. (*Maybe take a sec to explain what the AUC score is and why it is used but idk where?*) To determine the resolution to grid the polygons by each microhabitat-defined species range, we did a series of tests. We ran a 10-fold cross validation maxent model with multiple resolutions and analyzed the AUC scores associated with them. To maximize the AUC score and to minimize the uncertainty, we found a resolution that compromised the model needs for both arboreal and terrestrial distributions. This resolution was (\_\_) which resulted in the microhabitat distributions being represented by (\_\_\_) arboreal points and (\_\_\_) terrestrial points.

Resolution to grid the polygons/points by tests: The resolution indicated by asterisk\* is the resolution of the climate data.

HEY LAUREN RE-DO TESTS WITH A CV OF 10-FOLD

Analyses for Ecological Niche Modeling [Lauren will work on this section]

Many types of methods have been developed and implemented for Ecological Niche Modeling (ENM). Among them, the maximum entropy algorithm implemented in the Maxent software (Phillips et al. 2004, 2006; Elith et al. 2011) generally performs better than other algorithms (Elith et al. 2006; Phillips et al. 2006; Ortega-Huerta and Peterson 2008). Maximum entropy is a machine-learning technique that predicts species distributions by using detailed environmental variables associated with species occurrence. Maxent was desirable in this study because it is less sensitive to sample size (Wisz et al. 2007) and can be applied to sample sizes as small as five while still having high predictive power (Pearson et al. 2007).

We aimed to model the potential geographic and environmental niche of the microhabitat groups to see what climate variables are related to their distribution. We created ecological niche models for each microhabitat group and classification type in Maxent v.3.4.1 implemented in the dismo package (citation). We performed a 10-fold cross validation (2,3,4) model (n) times for stability. Each model was evaluated using the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC). The AUC is a measure of model performance that compares the model’s predictive ability to a random prediction of background points. The AUC values should be interpreted as the ability of the model to differentiate presence from a random prediction**. It is important to note that the AUC calculation in Maxent has been modified for use without true absence data by using the fraction of the total area predicted present (fractional predicted area) instead of the commonly used commission rate.** Each model output showed good performance compared to random expectation if it was above 0.75 AUC score. After evaluating for sufficiency, we combined each model into a (weighted) consensus model for each microhabitat group and classification scheme. We also measured the importance of each environmental variable by training with each environmental variable first omitted, then used in isolation, which also gave us the percent contribution of each environmental variable to the model. The model predicted continuous maps of microhabitat-defined species’ suitability by setting the output to logistic, which expresses the suitability in terms of probability values ranging from 0 (non-suitable) to 1 (highly suitability). The final projected maps were to the geographic extent where Plethodontidae are found in North and South America (20.0° S:70.0° N; 140.0° W:11.33° W).

Another evaluation step we took was to measure the niche overlap of the various microhabitats predicted suitability using two metrics to quantify niche overlap. The first metric is Schoener’s statistic for niche overlap (1968), defined as *Schoener’s D.* The second metric is Warren’s *I* statistic. The Schoener’s *D* (Schoener 1968) and Warren’s *I* statistic (Warren et al. 2008) were used because they were directly based on suitability scores and have been widely used for niche overlap measurements for both environmental and geographic space (McCormack et al. 2010; Hawlitschek et al. 2011; Peterson 2011, FOURCADE). The metrics *D* and *I* were calculated by taking the difference between species in suitability score at each grid cell. The two metrics ranged from 0 (species have completely discordant ENM) to 1 (species have identical ENM) (Warren et al. 2010) (5). We evaluated the statistical significance of I and D using permutation procedures defined by \_\_\_\_ [Erica will write this]. Zhu et al 2013 did the same thing of asking if they are different or not

To test whether arboreal species live in habitats suitable for terrestrial life and vice versa, we calculated a reciprocal suitability score for each relationship of distributional range and suitability model with a suitability cutoff of 0.5. In order to do so, we calculated the suitable area for each microhabitat type in square kilometers. We then calculated the geographic area of the distribution of species for the other microhabitat that overlaps with the suitable habitat.

Robustness analyses

1000 posterior trees

5 alternative microhabitat classification schemes

Non-maxent models

Other algorithms used are: if doing multiple models, I recommend the ones in bold because those are the most popular in the literature.

- **Generalized linear model (GLM)**

- Generalized additive model (GAM)

- **Multivariate adaptive regression splines (MARS)**

- Generalized boosted regressions model (GBM)

- Classification tree analysis (CTA)

- **Random forest (RF)**

- Artificial neural network (ANN)

- Support vector machines (SVM)

Different degrees of suitability cutoff for reciprocal suitability analyses

**Results**

Species Polygons

Give us some numbers/stats? What do other papers do?

- report the number of testing and training points for each model (the exact points will change because of the random seeds)

- report the resolution (2.5 arc minute and the gridded testing resolution of \_\_\_\_\_) and the implications of that

- report how many species fit the microhabitat definitions per classification scheme and how many were used/conglomerated to make the microhabitat polygon

Report number of pseudo occurrences from polygons

As of right now with a 2.5 arc-minute resolution, the number of pseudo-occurrences without thinning are:

- \_\_\_\_ arboreal

- \_\_\_\_ terrestrial

- \_\_\_\_ aquatic

- \_\_\_\_ cave

- \_\_\_\_\_ fossorial

- \_\_\_\_\_ saxicolous

ANOVA Results

To test whether species that utilize different microhabitats live in different broad scale climates, we treated all climatic variables as a single multidimensional trait. As different variables use different units, we used standard normal deviates for each climate variable, scaling all variables to have a mean of 0 and a standard deviation of 1. The phylogenetic ANOVA revealed significant differences in mean climate across microhabitat types (R2 = 0.0436, F = 2.6974, Z = 2.9415, P = 0.003). Pairwise comparisons revealed that arboreal and terrestrial species also differ significantly (Z = 5.1671707, p = 0.001).

The first six PCA axes explain 91.93% of all climactic variation and thus, we tested whether these axes differed significantly between microhabitat type. The phylogenetic ANOVA for PCA1 again revealed that arboreal and terrestrial species differ (Z = 7.3862, p < 0.01). Although similar analyses for PCA2 also demonstrated significant differences between these microhabitat type, these analyses were not robust to microhabitat classification or phylogenetic uncertainty (Appendix).

The raw climatic variables that loaded heavily on PCAs 1 and 2 included average temperature of the coldest month (BIOL\_\_\_), annual precipitation (BIOL\_\_\_\_), precipitation during the wettest quarter (BIOL\_\_\_\_), temperature of the warmest month (BIOL\_\_\_\_), and annual potential evapotranspiration (BIOL\_\_\_; Table \_\_\_\_). The loadings for the first six PCA axes showed all climate variables loading heavily except climatic moisture (CM).

Specific Climate Variables Results

Hab~Precip

Hab~Temp

Hab~PET

Arb-Terr

Ecological Niche Modeling [Lauren will work on this section]

MaxEnt AUC values for each consensus model ranged from (\_\_ to \_\_\_). The ROC values associated for each consensus model ranged from (\_\_\_ to \_\_\_). The models were evaluated using randomly generated training and testing data. After determining if each Maxent replicate was deemed acceptable, the mean of the replicates was used to make one consensus model. This consensus model was used to predict to the New World.

Final maps – possibly showing the overlap of distribution/suitability and making the 0.5 blue and everything else non-blue.

Report additional usage of CV, BS, or SS here

Report additional usage of different models here as well

Report the fractional predicted area for each model as well

Report these and the significance of them by the identity test:

I:

D:

Terr species in Arb suitable land: (area or overlap?)

Arb species in Terr suitable land: (area or overlap?)

- \_\_\_ species present where \_\_\_\_ can live

Robustness analyses

1000 posterior trees

5 alternative microhabitat classification schemes

Non-maxent models

Different degrees of suitability cutoff for reciprocal suitability analyses



LAURENS CITATIONS SO FAR

1. Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling, Elia Axinia Machado-Machado, 2012, Applied Geography. Doi: 10.1016/j.apgeog.2011.06.011

2. <http://abc.museucienciesjournals.cat/files/ABC_41-2_pp_217-225.pdf>

3.

[https://royalsocietypublishing.org/doi/full/10.1098/rsos.180513](https://royalsocietypublishing.org/doi/full/10.1098/rsos.180513 4)

[4](https://royalsocietypublishing.org/doi/full/10.1098/rsos.180513 4).

[https://www.researchgate.net/profile/Clementine\_Preau/publication/325036491\_Modeling\_potential\_distributions\_of\_three\_european\_amphibian\_species\_comparing\_enfa\_and\_Maxent\_Preau\_et\_al-MaxEnt\_and\_ENFA\_modeling\_on\_three\_amphibian\_species/links/5af2b07eaca272bf4259e46b/Modeling-potential-distributions-of-three-european-amphibian-species-comparing-enfa-and-Maxent-Preau-et-al-MaxEnt-and-ENFA-modeling-on-three-amphibian-species.pdf](https://www.researchgate.net/profile/Clementine_Preau/publication/325036491_Modeling_potential_distributions_of_three_european_amphibian_species_comparing_enfa_and_Maxent_Preau_et_al-MaxEnt_and_ENFA_modeling_on_three_amphibian_species/links/5af2b07eaca272bf4259e46b/Modeling-potential-distributions-of-three-european-amphibian-species-comparing-enfa-and-Maxent-Preau-et-al-MaxEnt-and-ENFA-modeling-on-three-amphibian-species.pdf 5)

[5](https://www.researchgate.net/profile/Clementine_Preau/publication/325036491_Modeling_potential_distributions_of_three_european_amphibian_species_comparing_enfa_and_Maxent_Preau_et_al-MaxEnt_and_ENFA_modeling_on_three_amphibian_species/links/5af2b07eaca272bf4259e46b/Modeling-potential-distributions-of-three-european-amphibian-species-comparing-enfa-and-Maxent-Preau-et-al-MaxEnt-and-ENFA-modeling-on-three-amphibian-species.pdf 5). Zhu et al 2013