 DOI: 10.1086/695984

Appendix A from A. L. Angert et al., "Testing Range-Limit Hypotheses Using Range-Wide Habitat Suitability and Occupancy for the Scarlet Monkeyflower (*Erythranthe cardinalis*)" (Am. Nat., vol. 191, no. 3, p. E76)

Expanded Methods

Training Data: Treatment of Herbarium Records

We conducted extensive preliminary data cleaning in R to remove exact duplicate records (−91), records within 100 m of a locality used as part of the testing data set (−131), and records with fewer than 3 decimal degrees of precision in latitude or longitude (−10). We also removed records that were in close proximity (<600 m) of one another (−187), retaining the record with the more recent collection date. The 600-m filter was chosen on the basis of the approximate aggregation of points around popular recreation sites and viewpoints, which were routinely oversampled in the raw data set. Records in this proximity to each other were often collected by the same individual on the same date and had identical values for climatic variables at this scale, making them essentially duplicate records. All remaining records were individually inspected in Google Earth to verify that the mapped location corresponded to the description of the locality and fell within appropriate riparian habitat. We removed poorly georeferenced points (e.g., falling on city centers; −33 records), those with vague descriptions that were georeferenced with false precision (e.g., east side of a mountain range; −13 records), and other instances of mismatches between description and mapped location (−3 records). In 18 cases, points were manually adjusted when the record's description provided unequivocal information to do so. The 432 retained records (available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.bg15b [Angert et al. 2018]) were subjected to thinning to minimize bias due to oversampled regions. We chose this approach because we knew that several regions (particularly San Diego County and Yosemite National Park) had been disproportionately targeted by surveys. We used OccurrenceThinner (Verbruggen 2012; Verbruggen et al. 2013), which identifies oversampled regions on the basis of a two-dimensional kernel density grid. We generated the kernel density grid using CrimeStat (Levine 2010) with a fixed bandwidth distance of 80 km (equal to 1 SD). We used CrimeStat because it is designed to generate kernel density estimates across spherical (i.e., geospatial) data. We ran OccurrenceThinner 10 times using default settings to ensure that results were not sensitive to the inclusion or exclusion of certain records, generating 10 pseudoreplicate sets of records ($N = 187$–204; fig. A1). These are not true replicates because they tend to share points from undersampled areas but have different records from oversampled areas that experienced heavier thinning. All model evaluation and hypothesis testing were performed on the 10 individual pseudoreplicates within each model type and for the average across pseudoreplicates. For simplicity, we present results using averages.

Training Data: Pseudoabsences

We constrained the elevation range available for sampling to fall within the maximum elevation sampled by the presence-absence data set at any given latitude (fig. A2). We also sampled pseudoabsences in proportion to the relative abundance of true absences across US Environmental Protection Agency (2013) level 3 ecoregions (table A1). Because of the configuration of the landscape, sampling background records from a fixed distance to presence records would have resulted in a large portion of records falling in desert regions, which were purposely undersampled in the testing data and whose inclusion would likely result in inflated model evaluation scores. Thus, adjusting the sampling of pseudoabsence records to reflect bioclimatic regions was more appropriate for this study over a solely distanced-based approach.

Testing Data: Sampling Design

To define climatic strata, we first created a minimum convex polygon (MCP) surrounding primary occurrence records using the Hawth's tools extension (Beyer 2004) for ArcGIS V9.3 (ESRI, Redlands, CA) and then added an ~100 km (1 decimal degree) buffer around the MCP. Within the buffered MCP, we extracted annual mean temperature and annual precipitation (bio1 and bio12, respectively; Hijmans et al. 2005) at 30-s resolution and divided each variable into five

equal intervals. Climatic strata were determined by the 25 possible combinations of each of the five levels of bio1 and bio12, but only 18 combinations existed within the buffered MCP. The buffered MCP, which reached from 29.54993° to 45.052307°N, was also divided into five equal latitudinal intervals. Because of US State Department travel advisories for northern Mexico at the time of field sampling, the southernmost stratum was not sampled. Not all climatic strata were available within each latitudinal stratum, resulting in a total of 49 unique climatic × latitudinal strata. Within each latitudinal stratum, we drew random points equal to 40 times the number of climatic strata. Because the random points landed in climatic strata in proportion to the land area covered by each stratum, some climatic × latitudinal strata contained hundreds of candidate points, while seven strata with extremely small area did not contain any random points, bringing the number of sampling strata down to 42. These random points were treated as candidates for field surveys but were subjected to additional filtering on the basis of the availability of public lands and access to riparian habitat. Because we were concerned with both spatial and environmental coverage, we strove to include two to 10 survey points from each stratum but required points within a stratum to be separated by at least 20 km. In 2010, surveys began in the central Sierra Nevada Mountains, proceeded northward into Oregon and then southward along the coast, and ended in southern California. In 2011, surveys began in southern California and proceeded northward into the Sierras and then northward along the coast into Oregon. Surveyed points are available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.bg15b (Angert et al. 2018).

Testing Data: Assignment to Groups

Latitudinal groups (fig. A3) were assigned on the basis of US Environmental Protection Agency (2013) level 3 ecoregions, resulting in 97 points assigned to the northern group, 89 assigned to the central group, and 54 assigned to the southern group. We explored alternative groupings on the basis of geography (i.e., Cascade vs. Sierra Nevada Mountains in the north and Transverse vs. Sierra Nevada Mountains in the south), the 25th and 75th percentile latitudes of presence records, equal division of points (e.g., $n = 80$ per group), and population genetic structure (J. R. Paul, unpublished manuscript). Using alternate breakpoints between groups did not alter results. Elevational groups were assigned on the basis of a prior reciprocal transplant experiment (Angert and Schemske 2005): <400 m = low margin, 400–1,200 m = center, and >1,200 m = high margin.

Building and Evaluating ENMs

Using the glm function (R base package), we conducted stepwise variable reduction to select logistic regression models of presence/pseudoabsence versus environmental predictor variables, hereafter referred to as GLM. Linear and quadratic terms were included. We used the package gam version 1.09 (Hastie 2013) to build generalized additive models (GAMs) of presence/pseudoabsence versus environmental predictor variables. We built GAMs in a stepwise manner by proposing a scope list for each possible predictor that contained options to exclude it, include it as a linear term, or include it as a smoothed term with possible smoothing parameter values ranging from 2 to 5. We built random forest (RF) models, which fit many replicate classification trees to bootstrapped subsamples of the data (De'Ath and Fabricius 2000; Cutler et al. 2007), using the package randomForest v. 4.6–12 (Liaw and Wiener 2002) and default settings (number of trees = 500, number of variables randomly sampled at each split = square root of number of predictors, sampling conducted with replacement, minimum size of terminal nodes = 1, maximum number of terminal nodes unconstrained, and fivefold cross-validation). We explored alternative numbers of trees (50, 1,000, 1,500), variables sampled at each split (two, three, or four and using the tuneRF optimization function), and portion of data split for cross-validation (three- or 10-fold), but model performance and downstream results were insensitive to these alternatives. To build boosted regression trees (BRTs), which combine classification trees with machine learning (boosting; De'ath 2007; Elith et al. 2008), we used the packages gbm version 2.1 (Ridgeway et al. 2013) and dismo version 1.0–12. We tested a range of learning rates and tree complexities and selected those that best balanced explanatory power (minimizing residual deviance) with parsimony (using the fewest trees). Final settings were learning rate = 0.1, tree complexity = 3, and step size = 5. Finally, we ran maximum entropy models using MaxEnt (MAX; Phillips et al. 2006) via the dismo package version 1.0–12. We used the ENMeval package version 0.1.1 (Muscarella et al. 2014) with the random k-fold method to select a value of the regularization multiplier (2.5) and types of features (linear, quadratic, hinge, and product). For GLM, GAM, BRT, and MAX models, internal accuracies were calculated using the entire training data set (resubstitution) and with a fivefold cross-validation procedure. For RF, we report internal accuracy as the out-of-bag estimate from the training data set, which is analogous to a cross-validated estimate.

Table A1: US Environmental Protection Agency (2013) level 3 ecoregions assigned to latitudinal groups of testing data points (surveyed presence and absence records)

Latitude	*n*	Ecoregion
North	93	Cascade
		Coast Range[a]
		Eastern Cascades Slopes and Foothills
		Klamath Mountains/California High North Coast Range
Center	93	Central Basin and Range
		Central California Foothills and Coastal Mountains
		Central California Valley
		Sierra Nevada
South	54	Mojave Basin and Range
		Sonoran Basin and Range
		Southern California Mountains
		Southern California/Northern Baja Coast

[a] Four points that fell within the extreme southern end of the Coast Range ecoregion, south of San Francisco Bay, were manually reassigned to the central group.

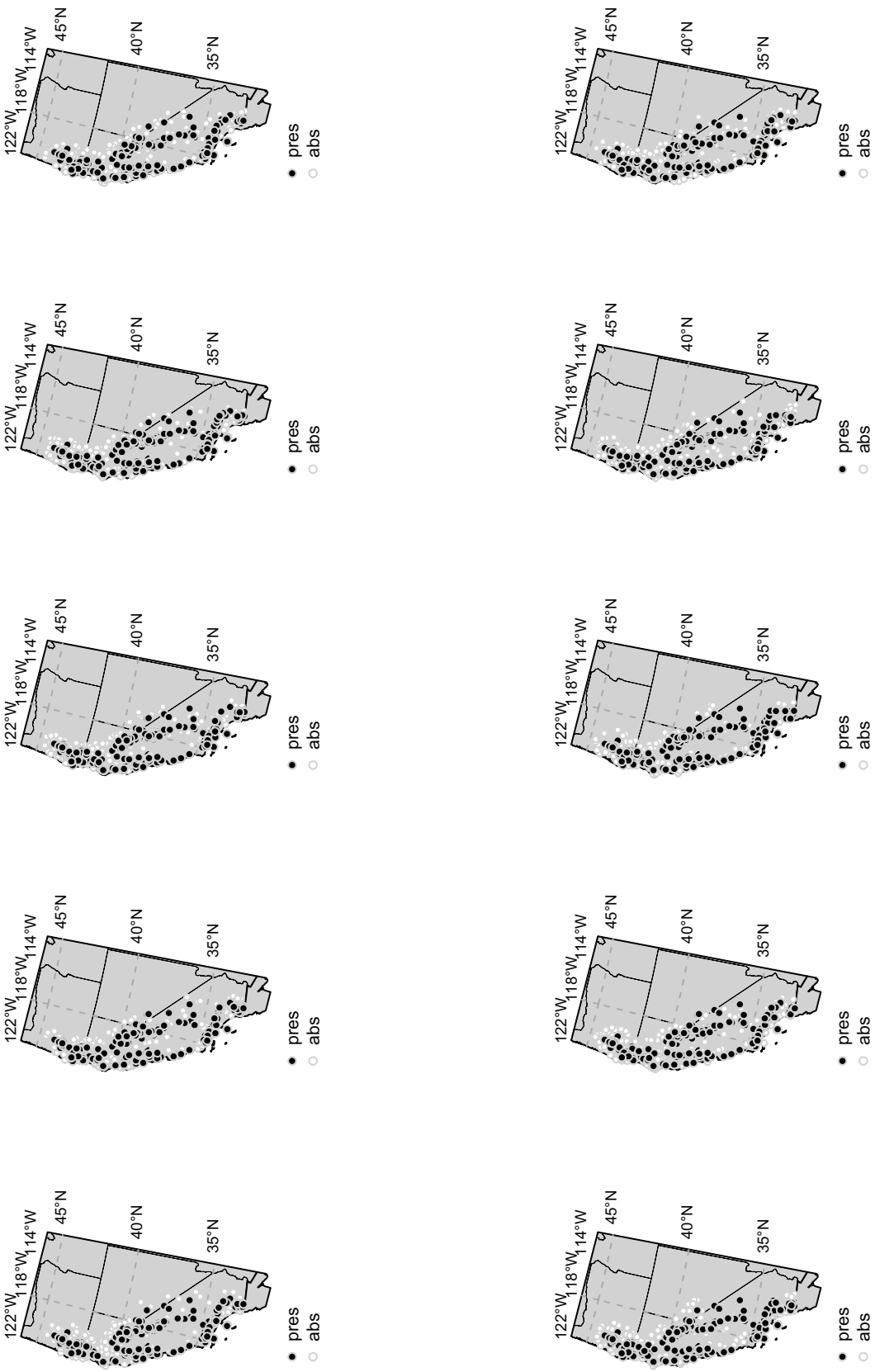

Figure A1: Maps of 10 pseudoreplicate training data sets showing randomly thinned presence records (pres; black circles) and associated pseudoabsences (abs; white circles).

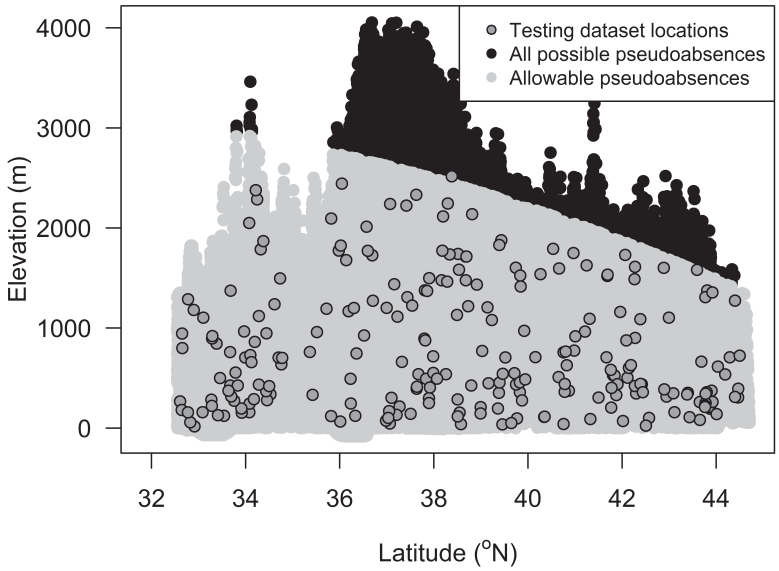

Figure A2: Adjusted elevation limits for sampling pseudoabsences at different latitudes. Dark gray circles show sampling locations within the testing data set. Black circles show all possible pseudoabsences when drawn randomly, while light gray circles show the allowable pseudoabsences after being constrained to fall within the maximum elevation sampled by the testing data at a given latitude.

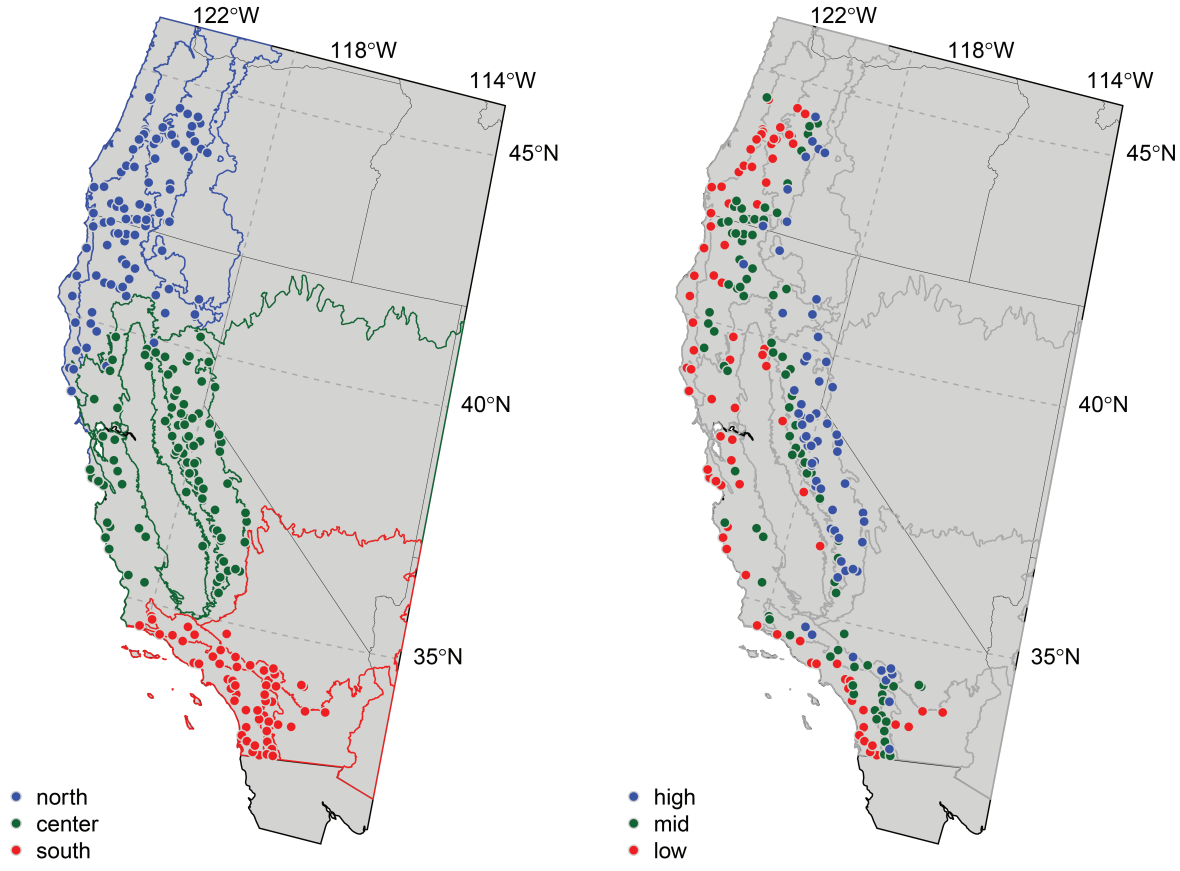

Figure A3: Map showing groups used for hypothesis testing. Testing data points split into north, center, and south latitudinal groups (*left*), determined on the basis of US Environmental Protection Agency level 3 ecoregions, and high-, mid-, and low-elevational groups (*right*).

 DOI: 10.1086/695984

Appendix B from A. L. Angert et al., "Testing Range-Limit Hypotheses Using Range-Wide Habitat Suitability and Occupancy for the Scarlet Monkeyflower (*Erythranthe cardinalis*)" (Am. Nat., vol. 191, no. 3, p. E76)

Summary of Climatic Predictor Variables

Table B1: Pearson correlations among predictor variables

	Bio2	Bio3	Bio4	Bio10	Bio11	Bio12	Bio14	Bio15
Bio2								
Bio3	.42							
Bio4	.32	−.70						
Bio10	.30	.02	.25					
Bio11	.00	.56	−.56	.66				
Bio12	−.39	−.11	−.21	−.46	−.25			
Bio14	−.29	−.45	.23	−.50	−.62	.52		
Bio15	.10	.47	−.42	.23	.54	.02	−.71	

Note: Bioclimatic variables (Hijmans et al. 2005) were derived from monthly temperature and precipitation values that were downloaded from ClimateWNA (Wang et al. 2012). See text for definitions of variables.

 DOI: 10.1086/695984

Appendix C from A. L. Angert et al., "Testing Range-Limit Hypotheses Using Range-Wide Habitat Suitability and Occupancy for the Scarlet Monkeyflower (*Erythranthe cardinalis*)" (Am. Nat., vol. 191, no. 3, p. E76)

Exploration of Nonstationarity

To test whether the effect of climatic variables on probability of presence varies across the range, we created generalized linear models that included linear and quadratic terms for each bioclimatic variable as well as their interactions with latitude (or elevation). We compared results from these models to those including only linear and quadratic terms for each bioclimatic variable. We did not conduct stepwise variable reduction for these models. For each model type (i.e., with vs. without latitude or elevation interactions), we then conducted quadratic regressions of average predicted suitability versus latitude (or elevation) and calculated slopes and intercepts of calibration curves for northern, central, and southern latitudinal regions (or low-, mid-, and high-elevation regions), as described in "Hypothesis Testing."

Interactions with latitude were detected for each bioclim variable in at least one pseudoreplicate data set, and they were prevalent for bio2 (mean diurnal range), bio11 (mean temperature of coldest quarter), and bio15 (precipitation seasonality; table C1). Models including interactions with latitude had slightly higher discrimination (mean increase in area under the curve [AUC] for internal validation via resubstitution = 0.060; mean increase in AUC for external validation on testing data = 0.046). When interactions with latitude were included, average modeled suitability peaked just south of the range center and declined toward both range edges (occupied sites: adjusted $R^2 = 0.25$; latitude, $b = 0.65$, $P < .001$; latitude^2, $b = -0.009$, $P < .001$; unoccupied sites: adjusted $R^2 = 0.23$; latitude, $b = 0.46$, $P < .001$; latitude^2, $b = -0.006$, $P < .001$). This relationship was more strongly hump shaped with respect to latitude compared with models not including latitude (fig. C1). The inclusion of interactions with latitude did not greatly affect the slopes of calibration curves (slopes with vs. without latitude: north, 0.60 vs. 0.51; center, 1.09 vs. 1.19; south, 0.71 vs. 0.77). Intercepts of calibration curves were slightly decreased by the inclusion of interactions with latitude in the center and south (with vs. without latitude: center, −0.58 vs. −0.39; south, 0.19 vs. 0.27), while intercepts became less negative in the north (−1.09 vs. −1.27). Nonetheless, the qualitative pattern of difference in calibration among regions was relatively unaffected by the inclusion of interactions with latitude (fig. C2).

Interactions with elevation were most prevalent for bio4 (temperature seasonality; table C2). Models including interactions with elevation had minimally higher discrimination on internal validation (mean increase in AUC for training data = 0.034) but slightly lower discrimination on external validation (mean change in AUC on testing data = −0.028). The relationship between suitability and elevation was qualitatively unaffected by the inclusion of interactions with latitude (fig. C3). As for models without interactions with elevation, average modeled suitability peaked at mid elevation and declined strongly toward high elevations for occupied sites (adjusted $R^2 = 0.17$; elevation, $b = -3.23\text{E}{-}04$, $P < .05$; elevation^2, $b = -2.26\text{E}{-}07$, $P < .001$) and declined linearly (though not significantly so) with elevation for unoccupied sites (adjusted $R^2 = 0.17$; elevation, $b = -2.85\text{E}\ {-}05$, $P > .10$; elevation^2, $b = -5.51$, $P > .10$). The inclusion of interactions with elevation decreased the slopes of the calibration curves (with vs. without elevation: high, 0.95 vs. 1.16; mid, 0.55 vs. 0.66; low, 0.70 vs. 0.99). Intercepts of calibration curves were unaffected at mid elevation, slightly more negative at low elevation, and slightly less negative at high elevation (high, −0.48 vs. −0.63; mid, −0.23 vs. −0.22; low, −0.71 vs. −0.60). However, the qualitative pattern of difference in calibration among regions was unaffected by the inclusion of interactions with elevation (fig. C4).

Table C1: Summary of generalized linear models (GLMs) including interactions between bioclimatic variables and latitude for each of 10 pseudoreplicate data sets of thinned herbarium presence records and associated pseudoabsences

Term	1	2	3	4	5	6	7	8	9	10
Bio2	***	*	**	***	**	−	−	***	*	***
Bio3	−	−	−	−	−	−	−	−	−	*
Bio4	−	−	−	−	−	−	−	+	*	*
Bio10	−	−	−	−	+	−	−	*	+	*
Bio11	*	*	*	−	*	−	−	**	***	**
Bio12	**	−	**	−	*	−	−	−	−	−
Bio14	−	+	−	−	*	*	*	+	−	−
Bio15	+	*	*	*	*	**	*	*	+	**
$Bio2^2$	***	*	***	***	**	−	−	***	*	***
$Bio3^2$	−	−	−	−	−	−	−	−	−	*
$Bio4^2$	−	−	−	−	−	−	−	−	−	−
$Bio10^2$	−	−	−	−	*	−	−	*	*	**
$Bio11^2$	**	+	−	*	−	−	−	−	*	**
$Bio12^2$	**	+	*	−	+	−	−	−	−	−
$Bio14^2$	−	−	+	−	−	−	−	−	−	−
$Bio15^2$	−	*	+	*	*	**	*	*	−	*
Latitude:										
Bio2	***	*	***	***	**	+	+	***	**	***
Bio3	−	−	−	−	−	−	−	−	−	*
Bio4	−	−	−	−	−	−	−	−	*	+
Bio10	−	−	−	−	+	−	−	+	+	*
Bio11	*	*	*	−	*	−	−	**	***	**
Bio12	**	−	*	−	+	−	−	−	−	−
Bio14	−	+	−	−	*	*	*	+	−	−
Bio15	+	*	*	*	*	**	*	*	+	*
$Bio2^2$	***	*	***	***	**	+	+	***	*	***
$Bio3^2$	−	−	−	−	−	−	−	−	−	+
$Bio4^2$	−	−	−	−	−	−	−	−	−	−
$Bio10^2$	−	−	+	−	*	−	−	*	*	**
$Bio11^2$	**	+	−	*	−	−	−	+	*	**
$Bio12^2$	*	−	*	−	+	−	−	−	−	−
$Bio14^2$	−	−	−	−	+	−	−	−	−	−
$Bio15^2$	−	*	+	*	*	**	*	*	−	*

− $P > .10$.
+ $P < .10$.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

Table C2: Summary of generalized linear models (GLMs) including interactions between bioclimatic variables and elevation for each of 10 pseudoreplicate data sets of thinned herbarium presence records and associated pseudoabsences

Term	1	2	3	4	5	6	7	8	9	10
Bio2	***	***	*	***	*	**	**	+	***	*
Bio3	*	**	*	***	**	*	**	*	**	**
Bio4	*	+	**	**	+	*	+	*	**	***
Bio10	−	−	−	−	−	−	−	−	−	−
Bio11	−	−	−	−	−	−	−	−	−	*
Bio12	−	*	***	+	*	*	+	−	**	−
Bio14	*	−	−	−	−	+	−	−	−	−
Bio15	−	−	−	−	−	−	−	−	−	−
$Bio2^2$	***	**	*	***	*	**	**	*	***	*
$Bio3^2$	*	**	*	***	**	*	**	+	**	**
$Bio4^2$	+	*	**	***	**	**	**	*	***	**
$Bio10^2$	−	−	−	−	−	−	−	−	−	−
$Bio11^2$	*	−	+	*	−	−	−	+	*	+
$Bio12^2$	−	*	***	+	*	*	+	−	**	−
$Bio14^2$	**	−	−	−	−	+	−	−	−	−
$Bio15^2$	−	−	−	−	−	−	−	−	−	−
Elevation:										
Bio2:	*	**	−	*	−	−	−	−	**	−
Bio3:	−	−	−	−	−	−	−	−	−	−
Bio4:	+	*	**	*	*	*	*	*	*	***
Bio10:	−	−	−	−	+	−	−	−	−	−
Bio11:	+	−	*	−	+	*	+	−	−	**
Bio12:	−	−	*	−	+	*	−	−	−	−
Bio14:	*	−	−	−	−	−	−	−	−	−
Bio15:	−	−	−	−	−	−	−	−	−	−
$Bio2^2$:	**	**	−	*	−	+	−	+	**	−
$Bio3^2$:	−	−	−	−	−	−	−	−	−	−
$Bio4^2$:	−	−	*	−	+	−	−	−	−	*
$Bio10^2$:	−	−	−	−	+	−	−	−	−	*
$Bio11^2$:	−	−	−	−	−	−	−	−	−	−
$Bio12^2$:	−	−	*	−	+	*	−	−	−	−
$Bio14^2$:	*	−	−	−	−	−	−	−	−	−
$Bio15^2$:	−	−	−	−	−	−	−	−	−	−

− $P > .10$.
+ $P < .10$.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

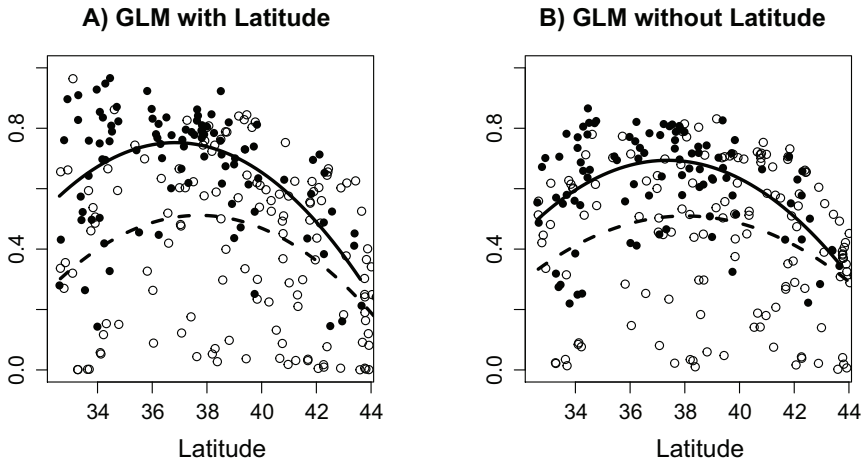

Figure C1: Latitudinal variation in predicted suitability for generalized linear models (GLMs) with (*A*) versus without (*B*) interactions between latitude and bioclimatic variables. Solid line, filled symbols = presences; dashed line, open symbols = absences.

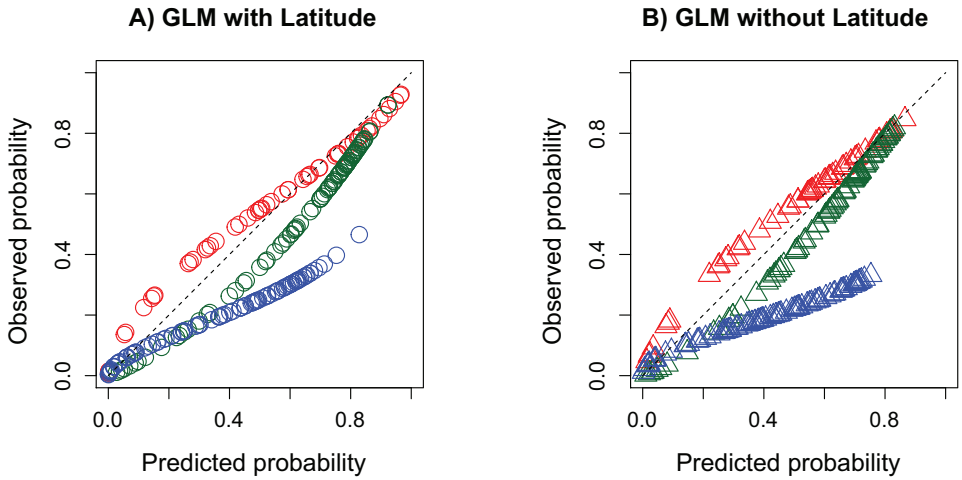

Figure C2: Latitudinal variation in calibration (a proxy for occupancy) for generalized linear models (GLMs) with (*A*) versus without (*B*) interactions between latitude and bioclimatic variables. Blue = north; green = center; red = south.

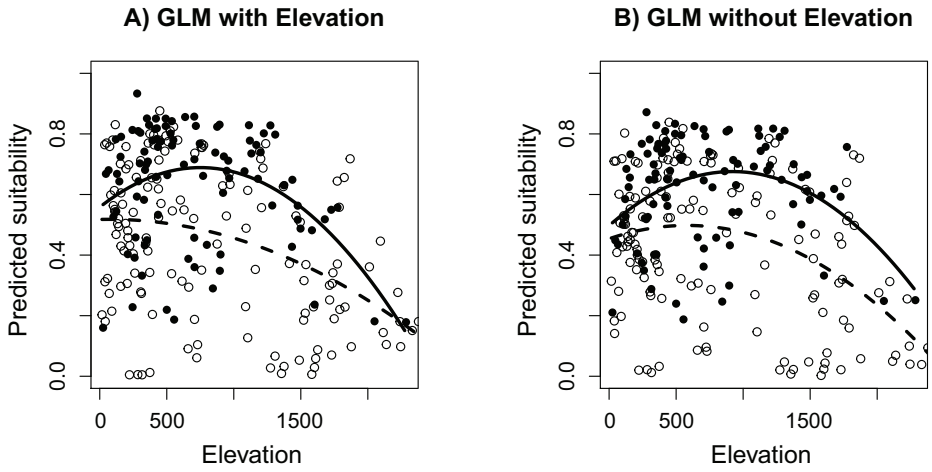

Figure C3: Elevational variation in predicted suitability for generalized linear models (GLMs) with (*A*) versus without (*B*) interactions between elevation and bioclimatic variables. Solid line, filled symbols = presences; dashed line, open symbols = absences.

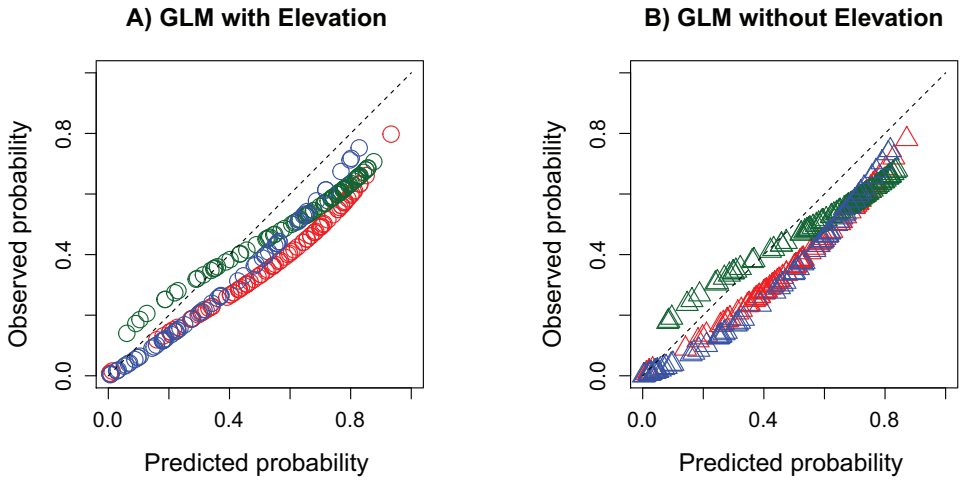

Figure C4: Elevational variation in calibration (a proxy for occupancy) for generalized linear models (GLMs) with (*A*) and without (*B*) interactions between elevation and bioclimatic variables. Blue = low elevation; green = mid elevation; red = high elevation.

Appendix D from A. L. Angert et al., "Testing Range-Limit Hypotheses Using Range-Wide Habitat Suitability and Occupancy for the Scarlet Monkeyflower (*Erythranthe cardinalis*)" (Am. Nat., vol. 191, no. 3, p. E76)

Expanded Results of Model Evaluation

(Semi)-parametric models (GLM and GAM) tended to have lower AUC determined from internal validation than nonparametric models (RF, BRT, and MAX; table D1). Internal AUC calculated from resubstitution were slightly (GLM, MAX) or substantially (GAM, BRT) higher than AUC determined from internal cross-validation (table D1). On average across the 10 pseudoreplicate training data sets, internal AUC calculated from cross-validation ranked in descending order as follows: BRT (0.871), RF (0.840), MAX (0.756), GLM (0.732), and GAM (0.674). However, AUC determined from external validation (external AUC) of the (semi)-parametric models increased by up to 0.11 compared with internal cross-validation, while AUC of the nonparametric models decreased by up to −0.17 (table D1). Mean external AUC decreased in the following rank order when pseudoreplicate models were applied to the testing data: RF (0.760), MAX (0.758), BRT (0.753), GAM (0.743), and GLM (0.733).

GLM, GAM, and RF had high performance when reliability was assessed internally (i.e., on training data); slopes and intercepts of calibration curves never (GLM) or only sometimes (GAM, RF) deviated from expectation (table D2). BRT and MAX also rarely showed bias (intercepts not usually different from 0), but slopes were much >1 (table D2), indicating that predicted values < 0.5 overestimated the true probability of occurrence while predictions > 0.5 underestimated the probability of occurrence. During external validation (i.e., on testing data), GLM, GAM, RF, and BRT models showed negative bias (intercepts < 0) but little spread (slopes usually not different from 1; table D2). Negative bias indicates that the models predict greater prevalence across the entire range of suitability values than is reflected in actual presences in the testing data. However, this does not undermine hypothesis tests about occupancy across the range because the important contrasts are relative differences between the range center and range edges. MAX retained significant spread (slopes > 1), but intercepts rarely differed from 0. On the basis of these model evaluation steps, we emphasize results from the best parametric model (GAM) and the best nonparametric model (RF), but note that outcomes of hypothesis tests were largely insensitive to model choice.

Table D1: Model discrimination, as measured by area under the receiver operating curve (AUC), for each model type and pseudoreplicate training data set (rep) of bias-corrected presence records from herbaria and associated pseudoabsences

	GLM			GAM			RF		BRT			MAX		
Rep	Int_R	Int_C	Ext	Int_R	Int_C	Ext	Int_R	Ext	Int_R	Int_C	Ext	Int_R	Int_C	Ext
1	.780	.739	.766	.802	.697	.767	.867	.798	1.000	.893	.762	.789	.769	.767
2	.749	.720	.762	.783	.688	.774	.808	.759	.987	.845	.763	.757	.717	.776
3	.773	.733	.729	.826	.645	.750	.837	.792	.999	.860	.780	.749	.776	.767
4	.803	.754	.708	.837	.711	.716	.875	.753	.998	.881	.723	.797	.777	.742
5	.753	.712	.717	.798	.648	.718	.824	.733	.997	.872	.740	.749	.745	.735
6	.759	.729	.729	.806	.670	.743	.839	.772	.998	.880	.764	.756	.750	.764
7	.784	.763	.733	.820	.680	.739	.819	.758	1.000	.858	.752	.769	.739	.756
8	.743	.700	.743	.801	.654	.747	.830	.738	.996	.879	.749	.742	.736	.762
9	.777	.730	.713	.839	.667	.737	.854	.737	1.000	.835	.754	.784	.797	.747
10	.771	.745	.732	.804	.685	.738	.852	.764	1.000	.907	.739	.749	.751	.765
Mean	.769	.732	.733	.812	.674	.743	.840	.760	.998	.871	.753	.764	.756	.758

Note: Model types include generalized linear models (GLMs), generalized additive models (GAMs), random forests (RFs), boosted regression trees (BRTs), and MaxEnt (MAX). For GLM, GAM, BRT, and MAX, internal metrics (Int) are calculated in two ways: via resubstitution (R) of the entire training data set and via fivefold cross-validation (C) of training data. For RF, internal metrics are out-of-bag estimates, which are analogous to cross-validation. For all models, external metrics (Ext) are calculated for independent testing data from systematic surveys of presence and absence. Greater values indicate greater discrimination.

Table D2: Parameters of calibration curves for training (internal) and testing (external) data sets

	Internal		External	
Model	Intercept	Slope	Intercept	Slope
GLM	0	1	−.606***	1.054
	0	1	−.527***	1.061
	0	1	−.419**	.740+
	0	1	−.424***	.584**
	0	1	−.539***	.949
	0	1	−.445***	.832
	0	1	−.457***	.832
	0	1	−.530***	1.127
	0	1	−.470***	.776
	0	1	−.429**	.842
GAM	−.026	1.219	−.679***	1.179
	−.021	1.201	−.619***	1.176
	−.034	1.406**	−.560***	.927
	−.043	1.307*	−.534***	.728+
	−.044	1.351*	−.599***	1.104
	−.033	1.300*	−.508***	1.016
	−.030	1.299*	−.507***	.891
	−.031	1.291*	−.536***	1.036
	−.047	1.395**	−.560***	.911
	−.044	1.285*	−.468***	.936
RF	−.021	1.34**	−.733***	.867
	−.039	1.05	−.71***	.833
	−.011	1.259*	−.807***	.973
	−.015	1.238*	−.689***	.658**
	.026	1.247*	−.705***	.718*
	−.007	1.242*	−.796***	.835
	−.042	1.129	−.634***	.683**
	.023	1.138	−.649***	.654**
	.002	1.327**	−.67***	.656**
	−.015	1.314*	−.7***	.864
BRT	.393	7.393	−.7***	.919
	.057	3.087	−.694***	.923
	−.102	4.608	−.894***	1.149
	−.281	4.381	−.66***	.772
	−.103	4.132	−.787***	1.034
	.32	4.222	−.813***	.977
	15.508	257.506	−.686***	.938
	−.046	3.622	−.771***	.935
	−.683	6.222	−.686***	.873
	−.231	5.715	−.797***	.936
MAX	.479*	2.624***	−.132+	2.570***
	.367+	2.423***	−.178+	2.802***
	.331	2.614***	−.166+	2.932***
	.475*	2.611***	−.089	2.181***
	.318	2.580***	−.193*	2.428***
	.365+	2.454***	−.134+	2.912***
	.402+	2.45***	−.105+	2.505***
	.294	2.441***	−.191+	2.978***
	.417*	2.525***	−.105	2.341***
	.311+	2.367***	−.105	3.008***

Note: Significant deviations from slopes of 1 and intercepts of 0 were tested via likelihood ratio tests of deviance from models with and without constrained parameters. Model abbreviations are as follows: generalized linear models (GLMs), generalized additive models (GAMs), random forests (RFs), boosted regression trees (BRTs), and MaxEnt (MAX).

+ $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.