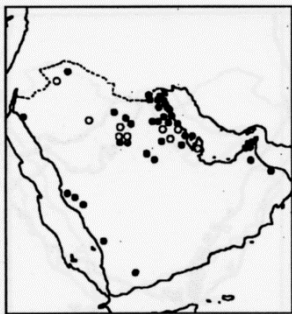


130 **Table A3.** The distribution of Sahara mustard over its native range recorded in the literature and
131 GBIF database.

Sources	Described distribution of Sahara mustard
Jalas 1996	Along the coastal area of Mediterranean Europe, including continental coast and islands of Spain, France, Italy, and Greece.
Zohary 1966 and Zohary et al. 1980	In Egypt: Nile Delta, Nile Valley, western desert, oasis, northwest, northeast, Northern, Central and Southern Sinai. In Saudi Arabia: Hejaz, eastern Arabia, central Arabia. Bahrein. Kuwait. In Israel, Palestine and Jordan: Mediterranean Littorals, northern, western and central Negev Desert, Acco Plain, Sharon Plain, Philistean Plain, desert of Edom, Jordan Mts, East Jordan Desert, Southern Jordan Desert. In Syria and Lebanon: coastland, Lebanon Mts, Jebel Druze, Northern Mts. Northern, southern and eastern Cyprus. In Turkey: Western Anatolia, Mesopotamian Anatolia, Aegean Islands. In Iraq: mountain region, lower Mesopotamia, northern plains and foothills, western and southern desert. In Iran: northern, southwestern mts, and central, and southern Iran.
Townsend and Guest 1980	In Iraq: occasional in the steppe region. Common in southern sector of the desert region.
Miller and Cope 1996	 Saudi Arabia, Southern Yemen, Oman, UAE, Qatar, Bahrain, Kuwait, S&W Europe, N Africa and SW Asia. On sand and gravel in deserts: 0 – 2400 m

Maire 1965	Coastal and interior dunes of North Africa. Oasis in M'zab (Algeria) of northern Sahara. The High Plateau. Saharan Atlas Mountains range. Oasis in Ahaggar Mountains (Algeria).
Rechinger 1968	Western and southern Europe, North Africa, western Syria, Iraq, Anatolia, Cyprus, Iran and Armenia.
Jafri 1977	N. Africa, S. Europe, eastwards to Pakistan. Recorded collections in coastal Libya.
Battandier and Trabut 1888-90	Coastal Algeria, High Plateau, Sahara. Mediterranean region.
Global Biodiversity Information Facility (data.gbif.org)	Records found in the following countries: Europe (Greece, Cyprus, France, Greece, Italy, Portugal, Spain, Turkey) Africa (Algeria, Burkina Faso, Egypt, Libya, Morocco, Tunisia) Asia (Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Pakistan, Qatar, Saudi Arabia, Syria, United Arab Emirates)

132

133 **References**

- 134 Battandier, J. A. and L. Trabut. 1888-90. *Flore de L'Algerie: contenant la description de toutes*
135 *les plantes singnalee's jusqu a ce jour comme spontanees en Algerie et catalogue des*
136 *plants du Maroc, Dicotyledones*. Librairie Adolphe Jourdan. Alger.
- 137 Jafri, S.M.H. 1977. *Flora of Libya: Brassicaceae*. 23. Al Faatech University. Tripoli.
- 138 Jalas, J., J. Suominen, and R. Lampinen. 1996. *Atlas Florae Europaeae: Distribution of Vascular*
139 *Plants in Europe*. The Committe for Mapping the Flora of Europe and Societas Biologica
140 Fennica Vanamo, Helsinki.

- Maire, R. 1965. *Flore de L'Afrique du Nord*. Paul Lechevalier, Paris.
- Miller, A. G. and T. A. Cope. 1996. *Flora of the Arabian Peninsula and Socotra*. Edinburgh University Press, Edinburgh.
- Rechinger, Karl. H (ed.). 1968. *Flora Iranica: Flora des iranischen Hochlandes und der umrahmenden Gebirge. Cruciferae. 57*. Akademische Druck – u. Verlagsanstalt, Graz.
- Townsend, C. C. and E. Guest. 1980. *Flora of Iraq*. Ministry of Agriculture and Agrarian Reform, Baghdad.
- Zohary, M. 1966. *Flora Palaestina*. The Israel Academy of Sciences and Humanities, Jerusalem.
- Zohary, M., C. C. Heyn, and D. Hiller. 1980. *Conspectus Florae Orientalis: An Annotated Catalogue of the Flora of the Middle East*. The Israel Academy of Sciences and Humanities, Jerusalem.

Appendix 4

Using MaxEnt to model the distribution of range-expanding species

The expansion of a non-native species means that its range does not reflect its stable relationship with the invaded environments (Elith et al. 2010a). This lack of equilibrium presents challenges for modeling potential distribution using data from current distribution.

Solutions to this problem include using less complex models and comparing models based on different background samples (Elith et al. 2010a).

We reduced the complexity of our models in three ways. First, we used only four climatic variables that are most biologically relevant to our focal species. Second, we used only the hinge

and quadratic features in MaxEnt. Choosing the two features means that the modeled distribution is constrained by the mean and variance of the given climatic variables, may have piecewise linear response to any of them, but is not constrained by any interaction between them (Elith et al. 2010b). Third, we increased the regularization parameter in MaxEnt (from the default value of 1 to 2.5) to reduce the complexity of the surface of fitted models. Hence, our models excluded complicated detail response of species distribution to climate, which is more appropriate for species that have formed a stable relationship with its environment (Elith et al. 2010a).

To further account for ongoing range expansion, we also allowed our models to provide Sahara mustard with more potential space for expansion. We did so by adding models based on a much larger background than those using our standard background. Our standard background was a polygon that consists of the majority of southwestern North America. The enlarged background was a rectangular region containing all lower 48 states of the U.S. and the entire territory of Mexico. By choosing the standard background, we asked why Sahara mustard was only found in certain areas of the Southwest given the spatial climatic variation within the region and whether it could further expand in the Southwest. By choosing the enlarged background, we asked why this species was only found in the Southwest given the climatic conditions across North America and whether it could expand beyond the Southwest. To provide approximately equal spatial density of sampling for our models, we drew 10,000 random samples from the standard background and 25,000, from the enlarged background.

We found that models based on both backgrounds allow us to reach the same conclusion that 1) Sahara mustard in North America is restricted by its climatic envelope (Fig. A3) and 2) the

climate in the invaded range generally predicts the native distribution (Fig. A4). The model predictions also allow us to infer the climatic range under which Sahara mustard is likely to be present in both its invaded and native range (Fig. A5).

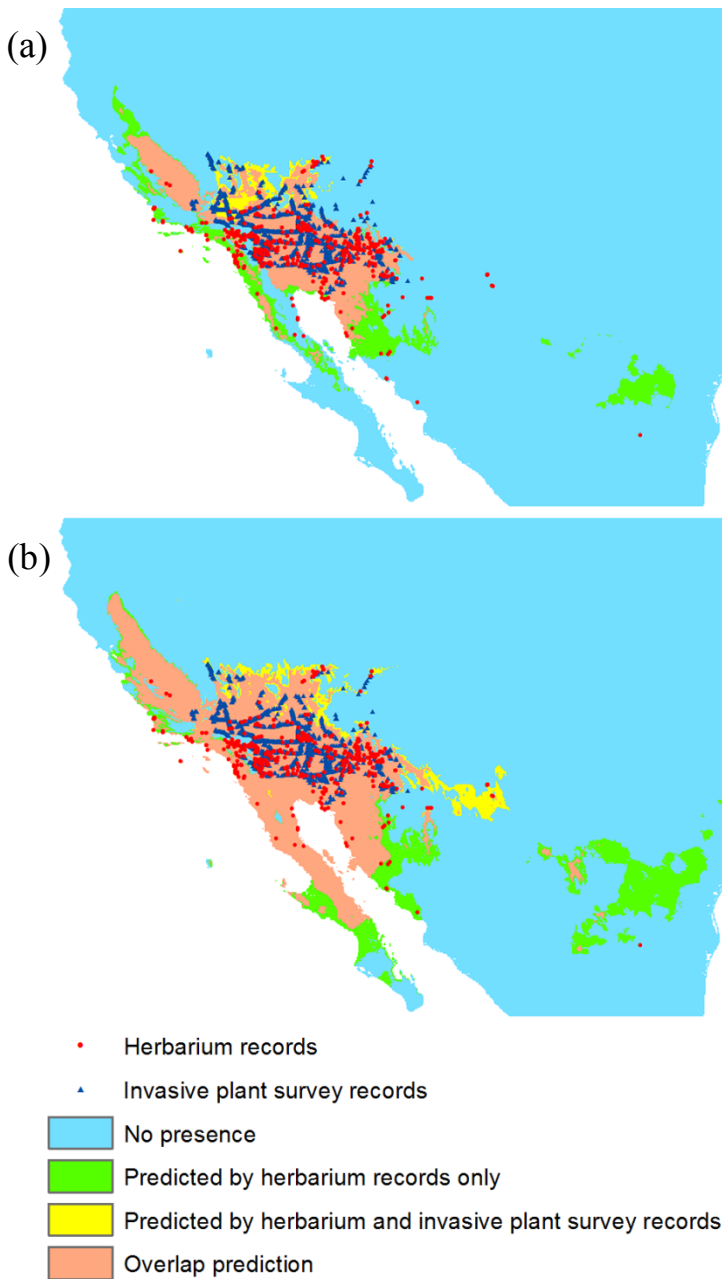


Figure A3. Distribution of Sahara mustard within its climatic niche in North America predicted by SDMs. We used background samples from (a) southwestern North America (SWNA) and (b) North America (NA) to build the models. For each background scenario, we trained two models: one using only

herbarium records and the other, herbarium and invasive plant survey records combined. We then derived an ensemble from the two models. Each ensemble shows the area predicted by both models (peach) and by each model alone (green or yellow). The maps also show the occurrence of Sahara mustard recorded by herbarium collections (red circles) and invasive plant surveys (blue triangles).

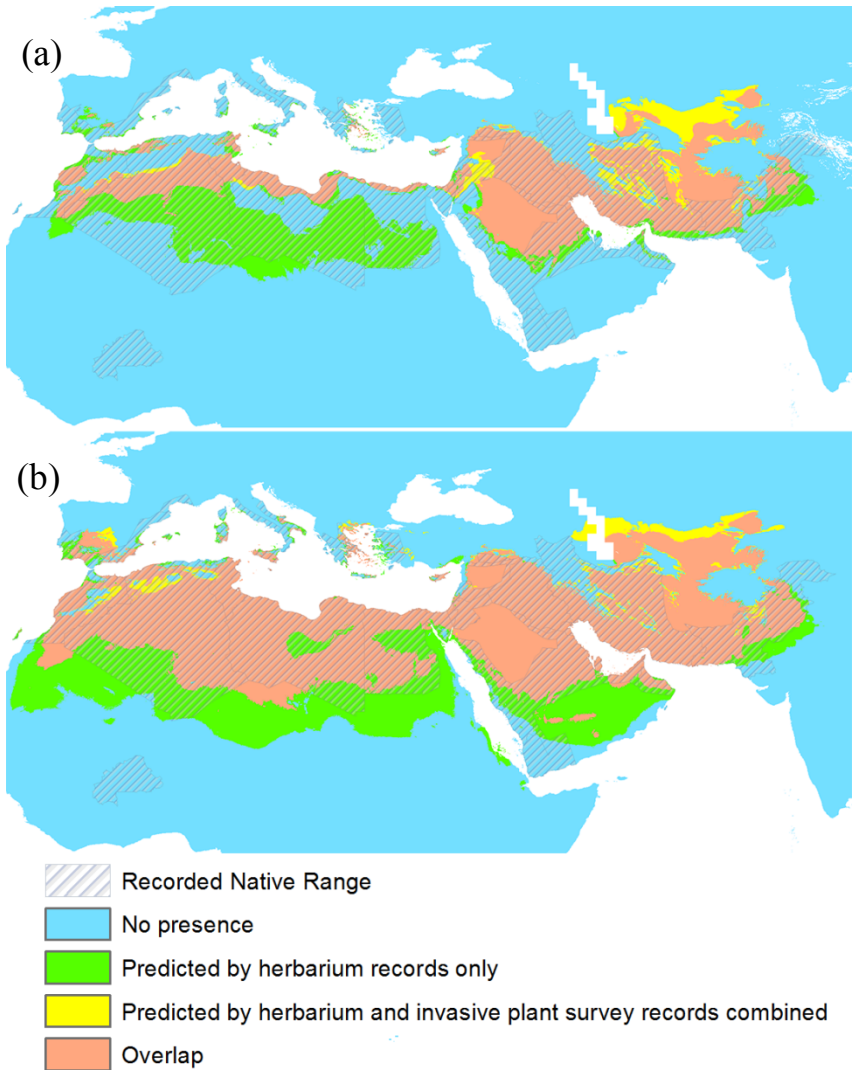


Figure A4. Projected distribution of Sahara mustard in its native continents projected by SDMs based on its invaded range in North America. We used background samples from (a) southwestern North America (SWNA) and (b) North America (NA) to build the models. The building of the models was the same as described in Fig. A3. Shaded areas represent its native range estimated from the literature and the GBIF records.

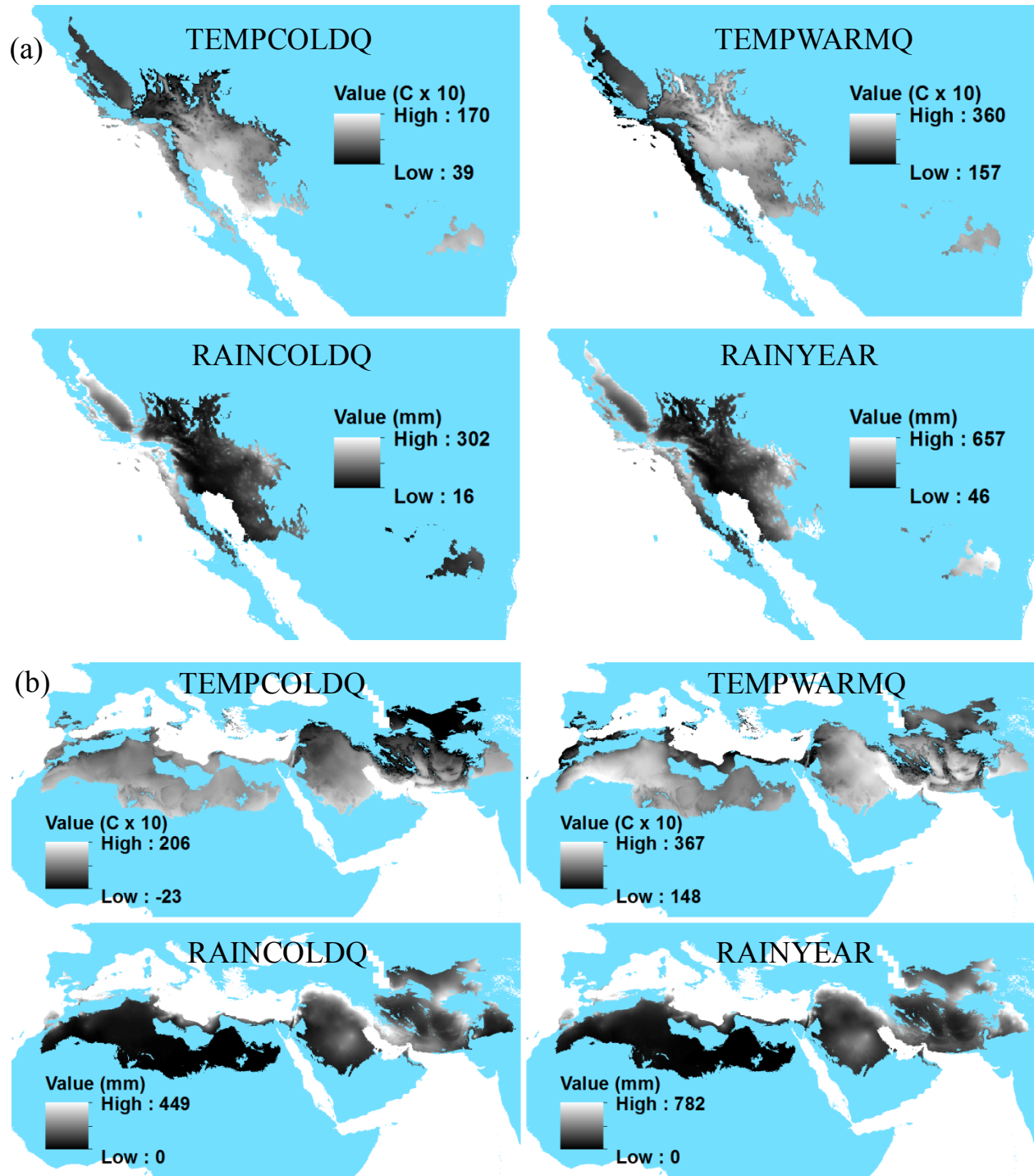


Figure 5A. Range of climatic variables in areas where Sahara mustard is predicted to be present in (a) North America and (b) its native continents by the SDMs. The four variables are mean temperature of the coldest quarter (TEMPCOLDDQ), mean temperature of the warmest quarter (TEMPWARMQ), precipitation of the coldest quarter (RAINCOLDQ), and annual precipitation (RAINYEAR). Temperature values are shown as degree Celsius x 10 and precipitation values, millimeters. The SDMs use southwest North America as the background.

Climatic variables used in the model and their contribution

None of the four climatic variables used in the two background regions were overly correlated (if $|R| > 0.85$) with each other (Table A4); therefore we included all of them in each model.

To understand which variable is most important in limiting the species' distribution, we evaluated the contribution of each variable to a model using MaxEnt's built-in evaluation algorithm (Table A5). Models based on SWNA background include summer temperature as the most influential variable, whereas models based on NA background include annual precipitation as the most influential variable. MaxEnt's evaluation of each variable's contribution to a model is sensitive to correlation between variables (though the model itself is not). In our models, summer temperature is correlated with winter temperature, and annual precipitation, with winter precipitation (Table A4). Therefore, the results can only allow us to suggest that temperature variation drives the species distribution within the Southwest, whereas precipitation is more important in limiting its range to the Southwest.

Models based on the SWNA background predicted a smaller range than those based on the NA background (Fig. A3). Since temperature is a more influential variable in SWNA background models, this stronger climatic restriction suggests that Sahara mustard would have a broader range in the Southwest if temperature were not a limiting factor. Given that regional distributions are likely to shift following a changing global climate, Sahara mustard might be predicted to expand particularly in response to elevated temperatures.

Table A4. Pearson coefficient of the four climatic variables used for building SDMs drawn from the two background regions: a polygon that consists of the majority of southwestern region in North America (SWNA) and a rectangular region containing all lower 48 states of the United States and the entire territory of Mexico (NA).

	TEMPCOLDQ		RAINCOLDQ		RAINYEAR	
	SWNA	NA	SWNA	NA	SWNA	NA
TEMPCOLDQ						
RAINCOLDQ	-0.1605	0.0523				
RAINYEAR	-0.0470	0.2770	0.7382	0.7383		
TEMPWARMQ	0.7582	0.7257	-0.4071	-0.0943	-0.4198	0.1275

Table A5. The estimated influence of climatic variables on each SDM. To determine the “percent contribution”, in each iteration of the training algorithm, the increase in regularized gain of the model is added to the contribution of the corresponding variable, or subtracted from it if the gain is negative. To determine the “permutation importance”, each variable was selected in turn, the values of that variable on training presence and background data are randomly permuted. The model is re-evaluated on the permuted data, and the resulting drop in training AUC (normalized to percentages) is shown.

		Training Data					
		Herbarium records			Herbarium and invasive plant survey records combined		
		Variable	Percent contribution	Permutation importance	Variable	Percent contribution	Permutation importance
Backg ground	SWN	TEMPWARMQ	42.2	33.4	TEMPWARMQ	51.6	42.1
		TEMPCOLDQ	30.6	10.7	RAINYEAR	18.8	33.7
		RAINCOLDQ	14.5	27.2	TEMPCOLDQ	17.6	6.3

		RAINYEAR	12.6	28.7	RAINCOLDQ	12	17.8
	NA	RAINYEAR	41.9	71.8	RAINYEAR	61.1	67.7
		TEMPCOLDQ	30.7	14.5	TEMPCOLDQ	27.8	18.6
		TEMPWARMQ	19.1	4.1	RAINCOLDQ	6.5	6.4
		RAINCOLDQ	8.4	9.6	TEMPWARMQ	4.6	7.3

Model validation

We tested our SDM models using 10-fold cross validation and used the test score of the Area Under the receiver operating characteristic Curve (AUC) to judge whether each model performed reasonably well (Table A6). An AUC of 0.5 means the model prediction is no better than random, whereas a value closer to the maximum achievable AUC indicates better performance of a model. The maximum achievable AUC is $1-a/2$, where a is the prevalence of the species over the sampled region (Phillips et al. 2006). This means that models trained by records of lower prevalence (e.g. fewer records or larger background) will have a higher maximum achievable AUC values. Therefore, models with higher AUC scores in the table are not necessarily better. Unfortunately, using presence-only data means that the prevalence of a species is unknown (no information on absence) and thus the maximum achievable AUC cannot be estimated. In our study, we used the test AUC to affirm that each model performed reasonably well ($AUC > 0.5$) but did not judge how much the model deviates from its theoretical optimum. We note that MaxEnt has been repeatedly shown to produce some of the most robust SDMs using presence-only data when compared with other modeling methods (Elith et al. 2006; Phillips et al. 2006).

If a model passed the test, we trained the model used the entire dataset and accepted its

logistic output as the relative probability of the species' presence in a spatial unit. We used the logistic output threshold at which the model achieved maximum training sensitivity plus specificity (minimizing the error rate for both positive and negative observations (Freeman & Moisen 2008)) and treated any logistic value under this threshold as an indication of absence.

Table A6. The AUC score of the MaxEnt species distribution model based on two different backgrounds (SWNA and NA) and using two different datasets: herbarium records only (H) and by combining herbarium and invasive plant survey records (H+IS).

		Training Data	
		H	H+IS
Background	SWNA	0.903+0.016	0.879+0.006
	NA	0.977+0.004	0.963+0.002

References

- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129 – 151.
- Elith, J. et al. 2010a. The art of modelling range-shifting species. - *Methods in Ecology and Evolution* 1: 330–342.

276 Elith, J. et al. 2010b. A statistical explanation of MaxEnt for ecologists. - Diversity and Distributions 17: 43–
277 57.

278 Freeman, E. A. and Moisen, G. G. 2008. A comparison of the performance of threshold criteria for binary
279 classification in terms of predicted prevalence and kappa. - Ecological Modelling 217: 48–58.

280 Phillips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic
281 distributions. *Ecological Modelling*, 190, 231 – 259.

282

Appendix 5.

Examining the influence of decadal cold-season precipitation on the local expansion of Sahara mustard

Introduction

As a winter annual plant, Sahara mustard experiences population boom in response to high precipitation over the cold (late fall-winter-early spring) season. Moreover, precipitation in fall and early winter may favor its growth more strongly than late cold-season precipitation because early winter rainfall followed by later precipitation events provides a larger time window for the species to outperform native annual plants (Barrows et al. 2009; Marushia et al. 2010). Therefore, years of high (early) cold-season precipitation provide potential temporal niche opportunities for its expansion. To examine whether such niche opportunities exist, we assessed the relationship between the species' local expansion and cold-season precipitation at the decadal scale.

Methods

We located local expansion hotspots using the same box-counting method (see Method in main text). We first evenly divided the southwestern North America into 100 km squares as our focal cells. We then divided each focal cell into 1, 5 and 10 km square cells and calculated the expansion rate within a focal cell (not adjusted for sampling effort) at those three local scales. We chose the maximum expansion rate among the three as the expansion rate in a focal cell and located cells that experienced high rates. We considered those focal cells as local expansion hotspots in that decade (Fig. A6). The calculation was done for each of the five decades (the

1960s – the 2000s).

We evaluated the decadal cold-season precipitation in those hotspots using weather data from the United States Historical Climatology Network. We acquired monthly precipitation data from weather stations that are located either within or adjacent to the hotspots (Table A7). We do not suggest that these stations faithfully represent the climate of each hotspot. Our interest is in revealing the general trend of winter precipitation between decades, which we believe is consistent over large spatial scales and therefore should be comparable between the hotspots and their correspondent weather stations. For each station, we calculated the annual cold-season precipitation as the mean monthly precipitation from October to April. We calculated its long-term mean averaged over year 1929-2010. We then calculated the abnormality of decadal cold-season precipitation (cold-season A_d) as the percent deviation of the decadal mean from the long-term mean. For instance, a cold-season A_d of 0.2 means the cold-season precipitation averaged over that decade is 20% higher than the long-term mean. We used Wilcox signed rank test to determine whether the cold-season A_d averaged over those hotspots in each decade was significantly above zero.

We also located focal cells that were hotspots in one decade but experienced negligible expansion in the following decade. We considered those cells as local expansion coldspots for that following decade and calculated cold-season A_d in those cells. We compared the cold-season A_d averaged over hot- and coldspots of the same decade (Wilcox rank sum test) to examine whether the former received significantly more rainfall. We only compared hot- and coldspots in the 1970s – 1990s, the three decades in which similar number of hot- and coldspots were

detected.

Moreover, since high amount of rainfall early in the winter season may give Sahara mustard, a species characterized by rapid phenology, an advantageous start, we performed the same analyses using early cold-season A_d , namely the decadal abnormality of mean monthly precipitation from October to December.

We repeated all the above analyses using precipitation data over the decade prior to any local expansion to indicate whether a local expansion was a delayed response to historical conditions. Finally we applied False Discovery Rate analysis (QVALUE in R package) to control the increased chance of listing a false positive test in this multiple-hypotheses test.

Results

Neither the cold-season nor the early cold-season A_d in local expansion hotspots were consistently positive across the five decades (Table A8). The cold-season A_d was significantly positive in the 1970s, 80s and 90s, but not so in the 1960s and significantly negative in the 2000s. The early cold-season A_d was significantly positive in the 1960s and 80s, but not so in the 1970s and 90s, and significantly negative in the 2000s. Comparing early cold-season A_d between hot- and coldspots, we found that those in hotspots were not significantly higher than in coldspots in any decade.

Among the three decades (the 1960s, 70s and 2000s) in which Sahara mustard achieved the most rapid local expansion (Fig. 1 in main text; Fig. A6), only the 1960s had local expansion hotspots experiencing a substantial increase (+37.5%) in early cold-season precipitation but no increase in cold-season precipitation (Table A8). The hotspots in the 1970s experienced a modest

increase (+5.9%) in cold-season precipitation, but the coldspots in the same decade received similar amount of rainfall. The 2000s saw the highest rate of local expansion and the largest number of hotspots. However, hotspots in this decade experienced a significant decline in cold-season (-17.7%) and early cold-season precipitation (-19.1%).

Neither did results suggest that the local expansion was a delayed response to previous decade's high precipitation. Among the three decades (the 1970s, 90s and 2000s) in which either cold-season or early cold-season A_d was significantly positive in the previous decade, none had hotspots exceeding coldspots by their previous-decade cold-season A_d (Table A8).

Discussion

Our results do not support the hypothesis that decadal variation in climate explains the decadal change in local expansion of Sahara mustard. Neither the rapid local expansion nor the difference between local expansion hotspots and coldspots can be consistently explained by higher cold-season precipitation averaged over each decade or the previous decade.

The local population growth of Sahara mustard may respond to climatic fluctuation at a much shorter time scale. One or two years of heavy seasonal rainfall is sufficient to trigger a local population boom of an annual species, strongly increasing the local occurrences recorded in a decade, but not raising the average precipitation over the same decade. For example, herbarium records of Sahara mustard in 2005 (78 records) make up more than a quarter of the 272 records in the 2000s as a result of an extremely wet winter and spring in the 2004-2005. However, the cold-season precipitation averaged over the 2000s is significantly below the long-term average.

Population data of higher temporal resolution is needed to investigate whether the local expansion of this species tracks the climate at a much finer temporal scale.

References

Barrows, C. W. et al. 2009. Effects of an invasive plant on a desert sand dune landscape. - *Biological Invasions* 11: 673–686.

Marushia, R. G. et al. 2010. Phenology as a basis for management of exotic annual plants in desert invasions. - *Journal of Applied Ecology* 47: 1290–1299.

Table A7. Locations of weather stations in the United States Historical Climatology Network

that are either in or adjacent to local expansion hotspots and coldspots in each decade from 1960s to 2000s. The weather data from those weather stations were used to calculate the cold-season precipitation in each hot- or coldspot.

Decade	1960s	1970s	1980s	1990s	2000s
Locations of weather stations that are in or adjacent to local expansion hotspots	Ajo, AZ	Ajo, AZ	Buckeye, AZ	Chandler height, AZ	Ajo, AZ
	Buckeye, AZ	Buckeye, AZ	Parker, AZ	Miami, AZ	Buckeye, AZ
	Chandler Heights, AZ	Kingman, AZ	Yuma, AZ	Pearce Sunsites, AZ	Chandler Heights, AZ
	Miami, AZ	Prescott, AZ	Tucson WFO, AZ	Prescott, AZ	Childs, AZ
	Parker, AZ	Wickenburg, AZ	Blythe, CA	Roosevelt 1 WNW, AZ	FT Valley (Flagstaff), AZ
	Roosevelt 1 WNW, AZ	Yuma, AZ	Chula Vista, CA	Tombstone, AZ	Grand Canyon NP, AZ
	Sacaton, AZ	Blythe, CA	Cuyamaca, CA	Wickenburg, AZ	Kingman, AZ
	Tucson WFO, AZ	Indio, CA	Pasadena, CA	Blythe, CA	Miami, AZ
	Brawley, CA	Redlands, CA	Redlands, CA	Brawley, CA	Parker, AZ
	Indio, CA	Santa Barbara, CA		Indio, CA	Prescott, AZ
	Redlands, CA	Tustin Irvine RCH, CA		Needles, CA	Roosevelt 1 WNW, AZ
		Boulder City, NV		Ojai, CA	Sacaton, AZ
		El Paso, TX		Santa Barbara, CA	Safford
				Jornada Exp Range, NM	Agricultural center, AZ
				Orogrande, NM	Tucson WFO, AZ
				NM State	Wickenburg, AZ
				University (Las Cruces), NM	Williams, AZ
					Blythe, CA
					Brawley, CA
					Cuyamaca, CA
					Indio, CA
					Needles, CA
					Newport Beach Harbar
					(Huntington Beach), CA
					Ojai, CA
					Pasadena, CA
					Santa Barbara, CA
					CA

					Tustin Irvine RCH, CA Boulder City, NV Kanab, UT St George, UT Zion NP, UT
Locations of weather stations that are in or adjacent to local expansion coldspots		Buckeye, AZ Chandler Heights, AZ Miami, AZ Parker, AZ Roosevelt 1 WNW, AZ Tucson WFO, AZ Brawley, CA Indio, CA	Ajo, AZ Kingman, AZ Prescott, AZ Wickenburg, AZ Yuma, AZ Indio, CA Santa Barbara, CA Boulder City, NV El Paso, TX	Buckeye, AZ Parker, AZ Tucson WFO, AZ Yuma, AZ Chula Vista, CA Cuyamaca, CA Pasadena, CA Redlands, CA	Pearce Sunsites, AZ Tombstone, AZ Jornada Exp Range, NM Orogrande, NM NM State University (Las Cruces), NM