RESEARCH ARTICLE

Utilizing digitized occurrence records of Midwestern feral *Cannabis sativa* to develop ecological niche models

Tori Ford[1] 🖸 | Ademola Aina[2] | Shelby Ellison[2] | Tyler Gordon[1] | Zachary Stansell[1]

[1]USDA-Agricultural Research Service, Plant Genetic Resources Unit, Geneva, New York, USA

[2]Department of Plant and Agroecosystem Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence
Tori Ford and Zachary Stansell, USDA-Agricultural Research Service, Plant Genetic Resources Unit, Geneva, NY, USA.
Email: tori.ford@usda.gov and zachary.stansell@usda.gov

Abstract

Hemp (*Cannabis sativa* L.) has historically played a vital role in agriculture across the globe. Feral and wild populations have served as genetic resources for breeding, conservation, and adaptation to changing environmental conditions. However, feral populations of Cannabis, specifically in the Midwestern United States, remain poorly understood. This study aims to characterize the abiotic tolerances of these populations, estimate suitable areas, identify regions at risk of abiotic suitability change, and highlight the utility of ecological niche models (ENMs) in germplasm conservation. The Maxent algorithm was used to construct a series of ENMs. Validation metrics and MOP (Mobility-oriented Parity) analysis were used to assess extrapolation risk and model performance. We also projected the final projected under current and future climate scenarios (2021–2040 and 2061–2080) to assess how abiotic suitability changes with time. Climate change scenarios indicated an expansion of suitable habitat, with priority areas for germplasm collection in Indiana, Illinois, Kansas, Missouri, and Nebraska. This study demonstrates the application of ENMs for characterizing feral Cannabis populations and highlights their value in germplasm conservation and breeding efforts. Populations of feral *C. sativa* in the Midwest are of high interest, and future research should focus on utilizing tools to aid the collection of materials for the characterization of genetic diversity and adaptation to a changing climate.

KEYWORDS
climate, ecological niche modeling, feral cannabis, germplasm

TAXONOMY CLASSIFICATION
Theorectical ecology

1 | INTRODUCTION

Hemp, (*Cannabis sativa* L.) is among the oldest cultivated crops with documented usage dating to 6000 BCE, and has been widely and globally dispersed, broadly following human migration patterns (Clarke & Merlin, 2016; Jiang et al., 2016; Liu et al., 2017; McPartland & Small, 2020; Ren et al., 2021). Historic uses of the plant include fiber, grain, fuel, and secondary metabolites (Kobayashi et al., 2008; McPartland et al., 2019). During the United States' colonial period, industrial hemp was often produced via enslaved labor as a fiber commodity (Johnson, 2019), with large regions of the Southeast and Midwest dedicated to cordage and textile production. Policymakers

and media entities linked "marihuana" with migrant labor communities and initiated policies culminating in 33 state-wide bans by 1933, followed by effective full criminalization in the Marihuana Tax Act of 1937 (Johnson, 2019; Solomon, 2020). State and Federal political campaigns urging citizens to eradicate feral hemp populations were moderately successful; by 1938, approximately 26,000 tons of hemp were eradicated from 15,132 acres in 23 states, largely accomplished by civilians (Conrad, 1999; Johnson, 2019). Federal research continued despite these efforts; for example, communication between the Treasury Department, Narcotics Laboratory, and the University of Illinois Chemistry Department compared *Cannabis* with opiates and the Narcotics Laboratory provided large amounts of "wild" *Cannabis* for evaluation. During World War II, the U.S. Military faced significant long-line fiber shortages and initiated the Federal "Hemp for Victory" Program (Filer, 2020; Johnson, 2019) to increase fiber supply. In 1942, American production was briefly revitalized from nearly zero to 226,000 acres by 1943 (Table S1; Hudson, 1942; Johnson, 2019). These efforts quickly diminished due to a lack of subsidies and federal regulation (Harmon, 2022; Johnson, 2019).

In the fall of the American hemp industry, remnants of the production lines remained as a growing population of feral escaped hemp gained recognition as "ditchweed." Woods et al. (2023) determined that escaped American feral *C. sativa* are genetically distinct from other *C. sativa* populations. Populations collected from Nebraska, Kansas, and Colorado exhibit evidence of significant genetic variation indicative of local adaptation to abiotic and biotic factors (Woods et al., 2023), although evaluation by Carlson et al. (2021) indicated that some American feral populations most closely cluster with older European fiber landraces. Ren et al. (2021) also found that other American feral populations, from Kansas and Nebraska, closely cluster with landraces and feral germplasm from China.

Feral and wild populations of agronomic crops are typically valuable genetic resources to identify biotic and abiotic resistance, introgress unique alleles within crop breeding systems, and to address bottlenecks in genetic diversity (Henry & Nevo, 2014). These adapted populations could prove useful within breeding efforts for locally adapted cultivars. Plant breeding relies on collections of diverse germplasm to meet persistent challenges such as climate change, changing consumer demands, and evolving pathogen pressures (Byrne et al., 2018; Frankel, 1974; Mascher et al., 2019; Milner et al., 2019). For instance, East Asian rice (*Oryza sativa indica*) landraces held in germplasm repositories contributed dwarfing genes that allowed significant genetic gains during the Green Revolution (Hedden, 2003). Repositories also provided landrace wheat accessions with new sources of resistance (Gordon et al., 2021) to emerging races of a fungal pathogen of wheat, *Puccinia graminis* Pers.:Pers. f. sp. *tritici* Erikss. and E. Henning which had quickly evolved virulence on wheat cultivars with resistance genes that had been effective for decades (Singh et al., 2011). Conserving ex situ landraces, breeding lines, and feral accessions has the same potential in hemp. Maintaining diverse germplasm will allow for current and future breeding challenges to be addressed by the hemp breeding community. Current conservation efforts to collect, characterize, and conserve feral germplasm

are ongoing (Ellison, 2021). These materials will be deposited within the newly formed Hemp Germplasm Collection administered by the United States Department of Agriculture National Plant Germplasm System which is now the largest repository of hemp genetic resources globally holding over 500 accessions, 128 of which are feral (USDA-NPGS, Accessed Aug 8, 2023).

Ecological niche models (ENMs) estimate the dimensions of ecological niche space to predict geographic species distribution (Soberon & Peterson, 2005). ENMs have applications within paleoecology, distributional ecology, and conservation biology (Chiarenza et al., 2019; Martínez-Freiría et al., 2020; Regos et al., 2021). ENMs generate predictive models by comparing species occurrence data against environmental predictors, where every point corresponds to a set of environmental suitability conditions (Hutchinson, 1957).

An ecological niche may be defined with various methods. Hutchinson delineates the niche into two categories: fundamental and realized niche, where the fundamental niche is described as the n-dimensional hypervolume where every point corresponds to conditions deemed suitable for a species to persist, and the realized niche as the points of occurrence in physical space that match the fundamental niche (Hutchinson, 1957). Soberon and Peterson (2005) sought to alleviate the discrepancies in the definition and application of niches to ENMs with their biotic-abiotic-movement (BAM) diagram, defining the M region as the area accessible to the species since origin (Soberon & Peterson, 2005). ENMs are typically constructed to understand ecological niche on a species level; however, recent inquiries have been made questioning whether species level modeling was appropriate. Species level models tend to ignore locally adaptive response and advise that informing niche models using evolutionary relationships is key to truly understanding ecological niche space (Bothwell et al., 2021; Smith et al., 2019).

There are three main methodologies generating ENMs, presence–absence (PA), presence-background (PB), and presence-only (PO) and each relies on algorithms unique to the sampling methods. PA algorithms use presence–absence field samples to discriminate between environments of occupancy and nonoccupancy, providing probabilities of a species occurrence within an area. GAMs (generalized additive models), GLMs (generalized linear models), BRTs (boosted regression trees), and RFs (random forests) are PA models. PB algorithms only require presence occurrence data and that the study extent (accessible area) is large enough for a background sample to be representative of a species' niche; resulting models distinguish suitable and unsuitable habitat. Maxent (maximum entropy), GARP (genetic algorithm for rule-set production), and ENFA (ecological niche factor analysis) are PB Models. PO requires only presence observations within an envelope algorithm, which defines potential niche but does not identify sinks properly, therefore making it less viable to identify areas of low suitability. Overlap Analysis, Bioclim, Domain, Habitat, and Mahalanobis Distance are PO models.

In practice, ENMs are mostly used for endangered species to make predictions of occurrence or persistence where the response to climate change remains unknown, and losing more germplasm would lead to critical endangerment (Borokini et al., 2023; Sony

et al., 2018; Wani et al., 2021). However, this technology has been less explored in the context of diversifying crop germplasm for use in herbaria, ex situ conservation, and plant breeding.

Here, we develop an ecological niche model of a large subpopulation of Midwestern feral *C. sativa* to (i) characterize the abiotic conditions correlated with population occurrence, (ii) estimate the areas with the highest abiotic suitability for feral *C. sativa*, (iii) identify areas at the highest risk of abiotic suitability change in the short- (2021–2040) and mid-term (2061–2080) future, (iv) highlight the utility of ENMs to germplasm collectors, and (v) utilize this technology to ultimately aid germplasm curators and plant breeders.

2 | METHODS

2.1 | Observance coordinates and data cleaning

Herbaria or occurrence coordinates were sourced from multiple online repositories (Table 1): Atlas of Living Australia (Belbin et al., 2021), Botanical Information and Ecology Network (Maitner, 2023), Centro de Referência em Informação Ambiental (Lima et al., 2021), Global Biodiversity Information Facility (Chamberlain et al., 2023), Genesys (Obreza, 2023), Germplasm Resources Information Network (USDA-NPGS, accessed 02/2023), the Smithsonian National Museum of Natural History (2023, Smithsonian NMNH, accessed 01/2023), and vPlants (2001, accessed 04/2023). All online repositories were accessed through the APIs provided per source. Additional occurrence records were provided by collaborators at Cornell University, Kansas State University, and the University of Wisconsin Madison.

The occurrence records do not include absence data, so we selected a presence-background algorithm such that the records defined the study extent across the Midwest. Occurrence records were cleaned to eliminate spatial biases by scrubbing duplicates, within and across repositories. Outliers, invalid coordinates, and records

TABLE 1 Number of occurrences for each repository, before and after the cleaning process.

Repository	Precleaned occurrences	Final occurrences
University of Wisconsin- Madison	49	37
Botanical Information and Ecology Network (BIEN)	29	15
Cornell University	42	27
Global Biodiversity Information Facility (GBIF)	680	620
Germplasm Resources Information Network (GRIN-Global)	86	53
Kansas State University	353	295
Smithsonian National Museum of Natural History	59	25
vPlants	590	121

with institutional coordinates were removed. Occurrences were filtered to exclude any records not contained within the 12 Midwestern States (Figure 1, Table S2; Legind, 2018). Once restricted, occurrence records were thinned by 500 m (Sillero et al., 2021), resulting in 1193 cleaned, restricted, and thinned records (Table 1). Accessible area was limited to a 5 km buffered convex hull surrounding occurrences.

2.2 | Ecological predictors

Ecological predictors represent the n-dimensions in the Hutchinsonian fundamental niche and were sampled as rasters at varied resolutions. Ninety-one variables from WorldClim (Fick & Hijmans, 2017), 18 Environmental Rasters for Ecological Modeling (ENVIREM, Title & Bemmels, 2018), and three variables from the U.S. Geological Survey Soil Properties dataset (Boiko et al., 2021) (Table S3) were downloaded. We calculated local daylength (the longest and shortest day) by latitude using the R package *geosphere* (Forsythe et al., 1995; Hijmans, 2022) as additional predictors to account for photoperiod sensitivity. In total, 115 environmental predictors were selected and clipped to the calibration and projection region.

2.3 | Model development

Variance inflation factor (VIF) was used to detect multicollinearity within these ecological predictors. The *USDM*, "vifstep()" function was used to compute a VIF analysis for all predictors, reducing variables included in the final model to 19 (Table S4). All predictors with a VIF score over 10 were removed. AICc was selected as the optimal model-selection method to explain the abiotic conditions suitable for this subpopulation, stronger transferability to project future climate scenarios, and overfitting resiliency (Warren et al., 2014). We also sought expert feedback on predictors selection from scientists collecting and evaluating these populations (A. Aina, personal communication, May 2023; Table S3). The Maxent algorithm was best to prioritize abiotic tolerance rather than the probability of occurrence. ENMs were constructed using the R package *ENMeval* (Kass et al., 2021) using the MaxEnt v3.4.4 (Phillips et al., 2006, 2017) algorithm. Occurrence records were used as presences and ecological predictors. Background points (*N* = 10,000) were sampled randomly within the study area as contingent absences resulting in a model that predicts potential niche (Lobo et al., 2010). Occurrence records and background records were spatially partitioned into equal quadrants called "blocks." These blocks are spatially determined by the MaxEnt algorithm. The partitioned occurrence and background blocks serve as both testing and training datasets to generate average model validation statistics. Spatial partitions are recommended by Kass et al., specifically when processing large datasets, as random partitions can lead to artificially overperforming models (2021). Block, and other spatial partitions like checkerboards, provide spatial cross-validation which increases

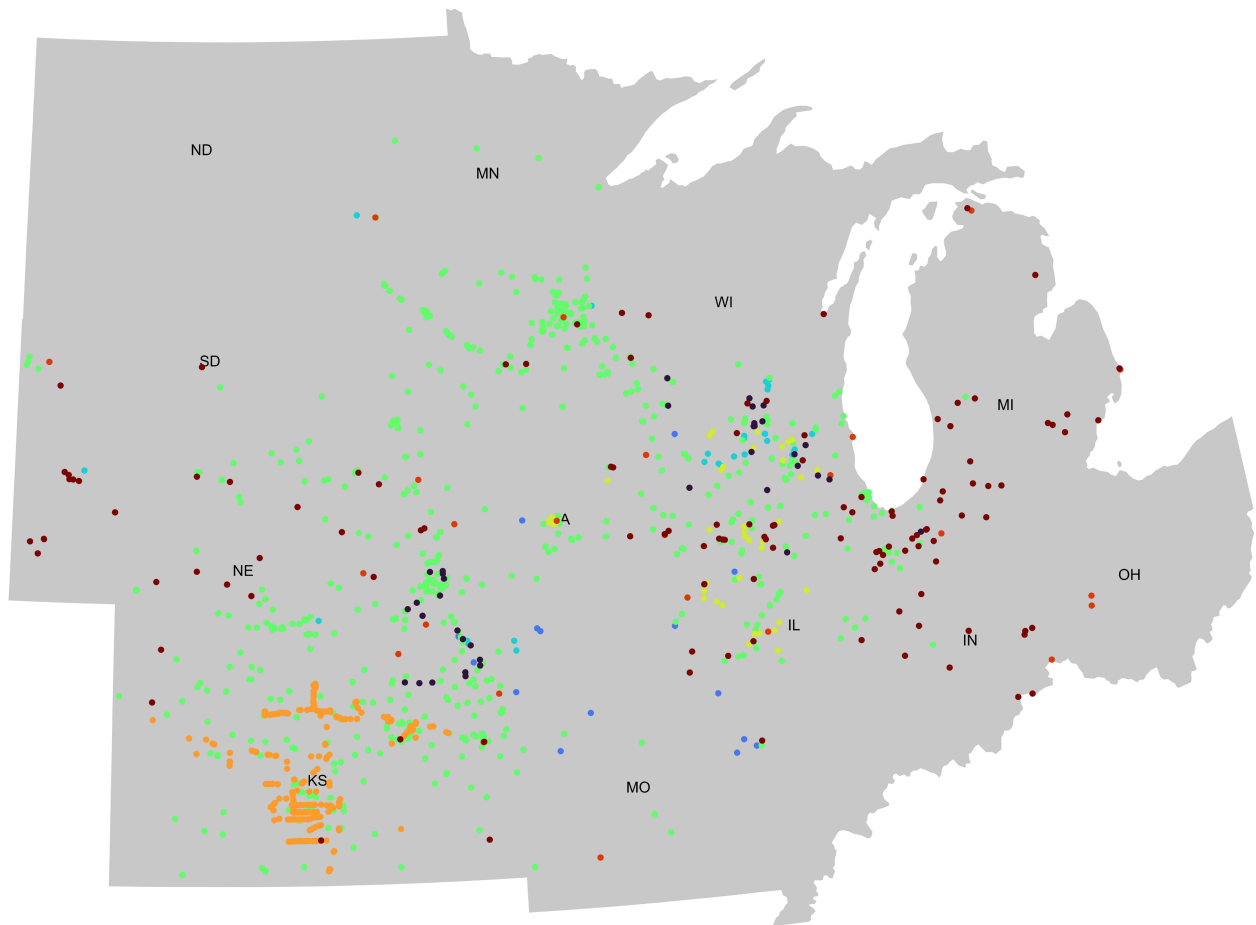## Occurrence Records of *Cannabis sativa* in the Midwestern United States

FIGURE 1 Map of *Cannabis sativa* occurrences in the Midwestern United States.

environmental extrapolation in the model that leads to models with higher transferability to new conditions; particularly useful when extrapolating to climate models (Kass et al., 2021). The Maxent algorithm returned models detailing the abiotic suitability of feral hemp across the study area. A total of 35 Maxent models were generated, with seven feature classes (L * Q * H) and five regularization (1:5) multipliers. Delta AICc scores were generated through ENMeval to compare the models.

The "ENMnulls()" function from *ENMeval* 2.0 (Kass et al., 2021) was used to compute 200 null model iterations. The validation statistics (AUC$_{test}$, AUC$_{train}$, AUC$_{diff}$, OR$_{10p}$, and CBI$_{val}$) were visualized using evalplot.nulls() and values from the calibrated model were used for performance validation. The optimal model was projected to the Midwest using the *dismo* "predict()" function (Hijmans et al., 2022), and binary suitability maps were generated using MPA (Minimal Predicted Area) from the *ecospat* package (Di Cola et al., 2017). We set suitability thresholds for the percentage of included occurrences.

Area suitability categories were defined by low to high ramping of the suitability threshold range. We generated response curves to describe the relationship of the environmental predictor to presence. MOP analysis was accomplished via the *kuenm* package (Cobos, Peterson, Barve, et al., 2019; Cobos, Peterson, Osorio-Olvera, et al., 2019), "kuenm_mop()" by comparing environmental predictors between the calibration region and the Midwest United States.

2.4 | Climate change scenarios

The optimal model was projected into the GISS-E2-1-H climate scenario during two time periods, under two SSPs (Shared Socio-Economic Pathways). SSPs represent future climate scenarios as a result of the intersection of political climate and social change (Hausfather, 2018). We projected the best model against two climate-change scenarios using the *dismo's* "predict()" function and

GISS-E2-1-H climate model against SSP1 (+2.6°C; "Sustainability—Taking the Green Road"), SSP5 (+8.5°C; "Fossil-fueled Development—Taking the Highway") (Riahi et al., 2017) and we ran these models under two time periods: 2021–2040 and 2061–2080. The five models (current, normal, negligible, terrible, and meltdown) were converted to binary and compared with *ecospat's* "BIOMOD_RangeSize()" function, and the resulting comparisons were used to visualize trends in abiotic suitability across climate-change scenarios.

3 | RESULTS

3.1 | Model validation

Of 35 models output by Maxent, the optimal model selected has an LH feature class with a regularization multiplier of 1. Accordingly, the AUC_{train} of the null models is significant (Figure S1). The AUC of the optimal model is 0.73, indicating good classification capacity. The CBI of the model is 0.64, indicating high model calibration and the null models indicate that CBI is significant. OR_{10p} did not meet the 0.05 significance threshold while AUC_{diff} surpassed the threshold (Figure S1). The larger pool of AUC_{diff} scores ranged from 0.031 to 0.098 and the optimal model had an AUC_{diff} of 0.089, which does not indicate severe overfitting (Bohl et al., 2019). We confidently state that the optimal model has good discriminatory abilities with some potential of being overfit.

3.2 | Projection

When projecting the model across the Midwest, the highest suitability areas lie along low-lying river channels and tributaries, specifically the Upper Mississippi, Ohio, Missouri, and Platte Rivers. In fact, the only suitable area at the highest suitability threshold lies along river channels and the coasts of the Great Lakes (Figure 2a). The largest continuous region of abiotic unsuitability was located within northern Minnesota, predominately within Aitkin, Beltrami, Koochiching, and Itasco counties. Another large region of abiotic unsuitability was located within North Dakota (Haakon, Jones, Lyman, and Stanley counties). MOP analysis indicated the areas under the highest risk of unreliable prediction (extrapolation) are the Northern half of North Dakota, Minnesota, Wisconsin, and Michigan's Upper Peninsula and Ohio River regions (southern Missouri, Illinois, and Indiana) (Figure 2b).

The optimal habitat of feral *C. sativa* occurrence is defined by the most important ecological predictors of occurrence determined by the model. The predictor most effective for predicting the occurrence of feral hemp was BIO8 (mean temperature of the driest quarter); the driest quarter is the winter months, December, January, and February (MRCC, 2024). Response curves indicate that feral hemp occurrence is higher in areas characterized by temperature extremes during these months. The second-best ecological predictor of occurrence

was fc_gNATSGO_US (soil field capacity). Response curves indicate that lower field capacity (e.g., the water content retained after excess water drainage) predicted feral hemp occurrence. The third-best ecological predictor of occurrence was PREC09 (average precipitation in September). Response curves indicate that high precipitation in September predicted feral hemp occurrence. In the Midwest, September is the last month of the year averaging more than 1 day of moderate rainfall (over 0.5 inches; MRCC, 2024). As drought conditions persist in the Midwest (National Oceanic and Atmospheric Administration, 2023), September's precipitation is crucial for providing season-extending irrigation for nonirrigated or drought-stricken plant populations (Goldstein, 2023).

3.3 | Climate change predictions

Generally, climate change scenario models suggest an increase in total abiotic suitable habitat for *Cannabis* and an expansion of that suitable area over time. Across all scenarios, primary states for habitat loss, and priority for germplasm collection, are Indiana, Illinois, Kansas, Missouri, and Nebraska. Generally, this means that these areas will observe a change in the ecological predictors associated with the persistence of feral hemp populations. Gains in suitability vary per climate scenario (Figure 3). For SSP1-2.6°C, abiotic suitability was mainly gained in Iowa and temporarily in Ohio, with suitability drastically increasing into the far future. For SSP5-8.5°C, the average predicted gain across the Midwest is 34% suitability from 2021 to 2080. The SSP5-8.5°C when compared with SSP1-2.6°C resulted in a greater increase in suitable land area. SSP5 gains more suitability than SSP1 in short- to mid-term temporal projections (Table 3).

4 | DISCUSSION

4.1 | Study population

Hemp landraces have been poorly characterized and maintained which has likely led to the loss of genetic diversity. Admixture of feral, grain, fiber, and secondary metabolite cultivars may exacerbate genetic erosion (Torkamaneh & Jones, 2021). We have created an ecological niche model that describes the abiotic tolerances of an important subpopulation of *C. sativa*. We focused on the United States. Midwest feral hemp subpopulation for several reasons: significant population stratification is present in *C. sativa* and artificial selective pressure is considered to be high within distinct pools of *C. sativa* germplasm (e.g., temperate or equatorial populations have likely adapted to divergent abiotic conditions). To our knowledge—through queries on Google Scholar and Web of Science—there have been no previous attempts to create an ecological niche model of feral *C. sativa*, and these efforts can be used to inform ongoing feral collection efforts and enhance the conservation of hemp genetic

(a) Predicted Habitat Suitability of *Cannabis sativa* in the Midwestern United States
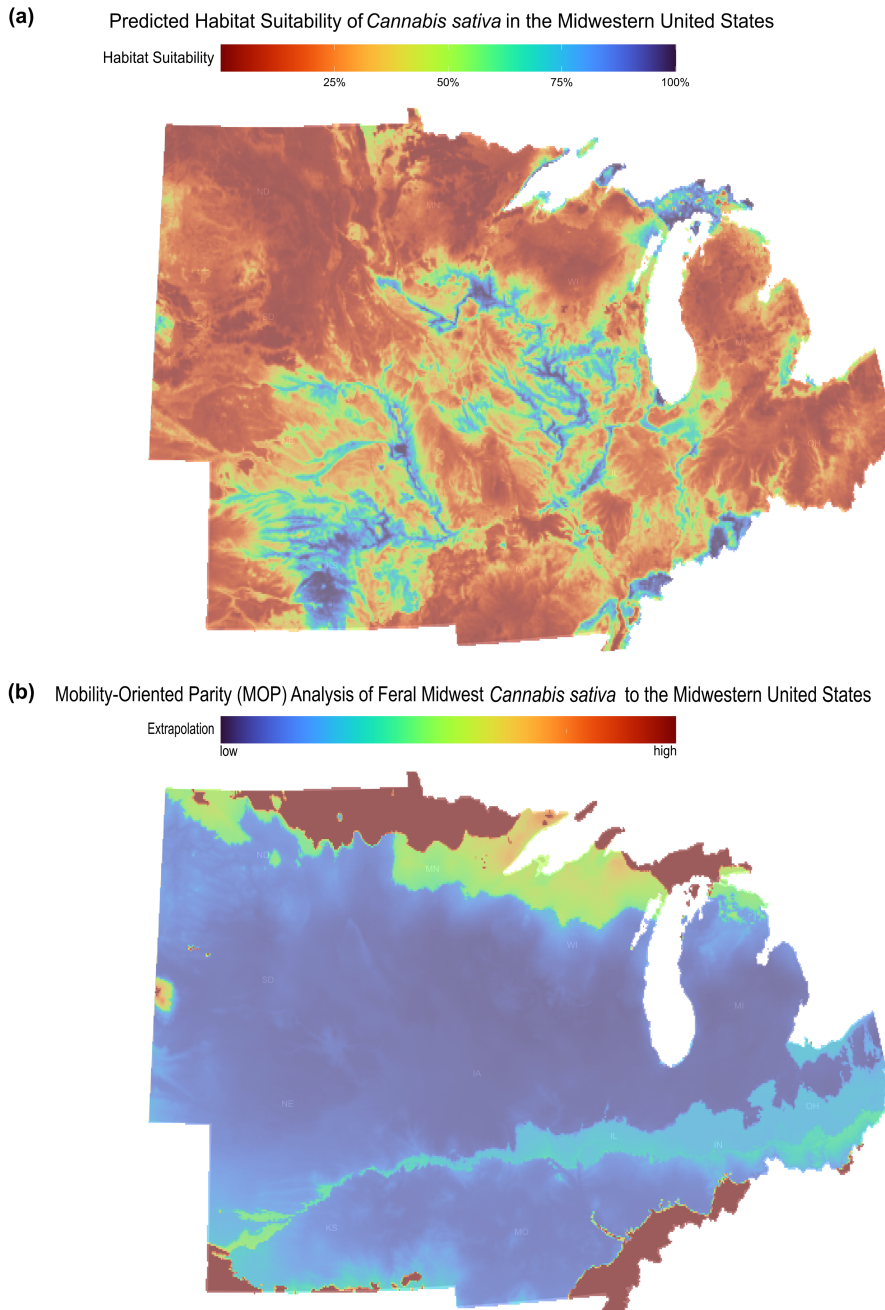
Habitat Suitability

25%   50%   75%   100%

FIGURE 2 (a) Abiotic suitability heatmap for *Cannabis sativa* in the Midwestern United States. Areas in blue represent high abiotic suitability, and areas in red represent low abiotic suitability. (b) Mobility-oriented Parity Analysis of the Calibration Region to the Entire Midwestern United States. Areas in blue and green represent low extrapolation, and areas in orange and red represent high extrapolation.

(b) Mobility-Oriented Parity (MOP) Analysis of Feral Midwest *Cannabis sativa* to the Midwestern United States

Extrapolation

low   high

resources. This currently understudied population of Midwestern *C. sativa* should be a top priority for conservation efforts.

## 4.2 | Diversifying the approach

Although Maxent has been widely utilized within the ecological niche model space, computing these models under multiple other algorithms is a best practice. Follow-up field sampling of these suitability predictions could be used as inputs within a subsequent presence–absence approach. Using predictive models to inform collections is an underutilized practice (Amici et al., 2014) and feral plants are an important group to study given that dispersal is partially driven through anthropological means.

4.3 | Model results and implications

The current projection of the potential niche depicts high abiotic suitability mainly along river channels. The most influential environmental predictors concur with previous inquiries into the environmental conditions most suitable for the "naturalization" of hemp in Iowa. Haney and Kutscheid (1975) remark that "naturalized," or feral hemp grows larger when in soils that are well-drained with ample moisture. They also note that site disturbance was a predictor of feral hemp, observing populations along waste areas like fence rows, stream banks, ditches, and abandoned fields (Haney & Kutscheid, 1975), which likely corresponds with favored areas often corresponding with higher human population density.

Predicted Range Change of *Cannabis sativa* in the Midwestern United States
using the GISS-E2-1-H Climate Model in SSP1-2.6°C from 2021 to 2040: "Normal"

Predicted Range Change of *Cannabis sativa* in the Midwestern United States
using the GISS-E2-1-H Climate Model in SSP1-2.6°C from 2061 to 2080: "Negligible"
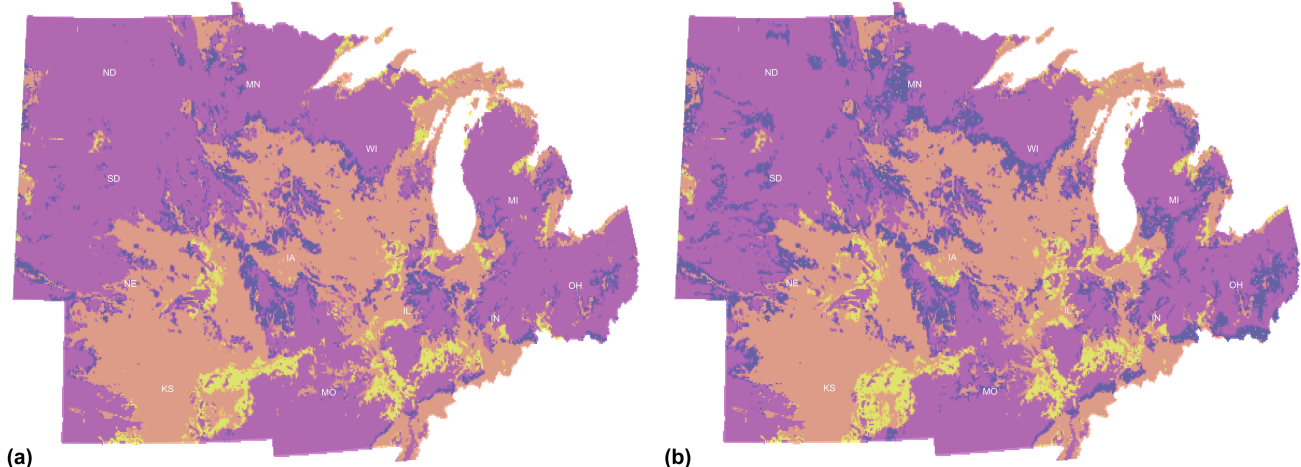
(a)

(b)

Predicted Range Change of *Cannabis sativa* in the Midwestern United States
using the GISS-E2-1-H Climate Model in SSP5-8.5°C from 2021 to 2040: "Terrible"

Predicted Range Change of *Cannabis sativa* in the Midwestern United States
using the GISS-E2-1-H Climate Model in SSP5-8.5°C from 2061 to 2080: "Meltdown"

(c)

(d)
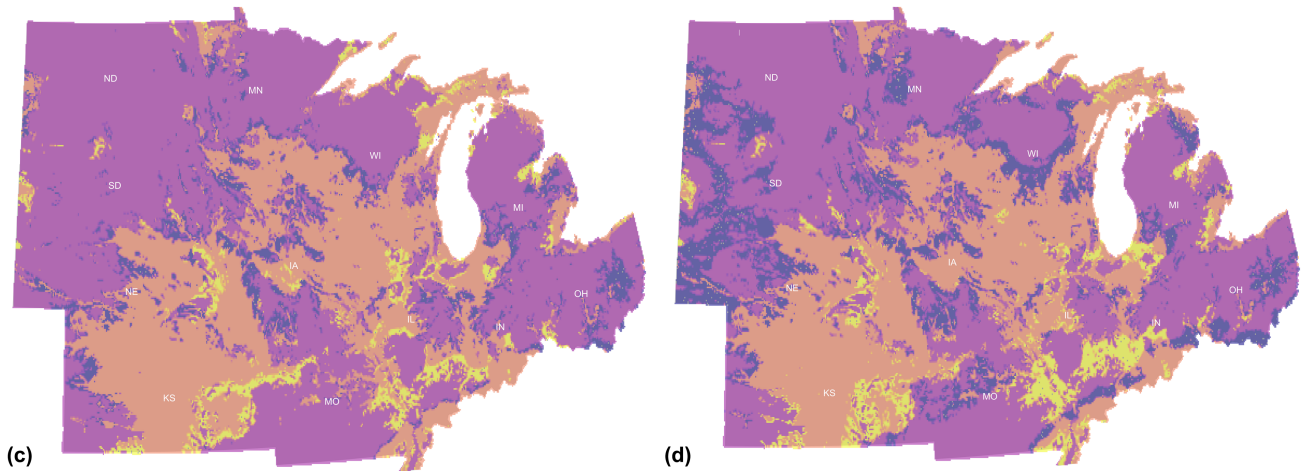
Range Change

loss   remain occupied   remain unoccupied   gain

FIGURE 3 Range comparisons of loss, occupation, inoccupation, and gain under the four climate change scenarios—normal (a), negligible (b), terrible (c), and meltdown (d)—compared to the current conditions.

Cannabis sativa cultivation in the Midwest typically entails a late May or early June transplantation. Young hemp plants are prone to low establishment if moisture and weed-pest pressure are high. For a successful transplantation, *C. sativa* plants should be put in warm, well-drained mostly loamy soils (Phillip Alberti, 2021). A review by Clarke and Merlin (2015) outlines the physiological needs of feral or "ruderal" hemp. They remark that wild-type hemp generally is more tolerant to a wider range of climates than its domestic counterparts. They propose that "ruderal" plants rely on vastly deep root architecture and loose-textured soils to withstand drought conditions and persist without the inputs of conventional agriculture. From our model, we observe the high contribution of soil field capacity on the model as a predictor (Table 2). Our model reports that the lower the field capacity, meaning soils retain less water, the higher the predicted occurrence of feral hemp. The model does not state that low field capacity is causative of feral hemp occurrence, it is correlated with higher occurrence.

4.4 | Models and their limitations

Models are not and should not be the end goal but are a useful tool to inform conservation and policy decisions. Our model evaluation statistics indicate the optimal model sufficiently predicts areas of abiotic suitability for *C. sativa* across the Midwest, however the limitations of this model must be further evaluated. Ecological predictors, WorldClim's bioclimatic variables, are well documented for being highly multicollinear (Arif et al., 2007). Some studies address this issue by removing multicollinear variables (Cobos, Peterson, Barve, et al., 2019; Cobos, Peterson, Osorio-Olvera, et al., 2019) entirely while other studies suggest using a priori-determined selection method (Zeng et al., 2016). A purely algorithmic-based selection could select variables that result in a predictively accurate model with variables that are realistically uninfluential (Smith & Santos, 2020). A priori selection could result in variables that are highly multicollinear,

Variable name	Contribution (%)	Permutation importance
Mean temperature of wettest quarter (BIO8)	41.26	8.83
Soil field capacity (fc_gNATSGO)	23.13	20.76
Average precipitation in month 09 (prec_09)	7.50	4.44
Average precipitation in month 01 (prec_01)	4.90	19.73
Sum of months with a mean temperature over 5°C×Number of days (growingDegDays5)	4.85	1.91
Elevation	3.67	16.91
Average solar radiation in month 04 (prec_04)	3.56	1.96
Mean monthly potential evapotranspiration of the driest quarter (PETDriestQuarter)	2.21	2.68
Terrain roughness index (tri)	2.15	3.55
Soil available water capacity (awc_gNATSGO)	1.74	2.83

Note: Percent contribution is the increase in regularized gain, per every iteration of developing the optimal model, converted to percentage. Permutation importance measures the decrease in AUC_{train} by permuting values of the training points in each variable after the model has been generated and converted to percentage (Phillips, 2017).

resulting in an ultimately ineffective model (Cobos, Peterson, Barve, et al., 2019; Cobos, Peterson, Osorio-Olvera, et al., 2019). We followed alternative recommendations by Smith and Santos (2020) to first select variables a priori, followed by algorithmic selection to remove multicollinear variables. By employing the VIF (variance inflation factor) analysis for algorithmic selection, we identified the variance induced when an ecological predictor is manipulated and allowed for comparisons of predictor collinearity (Cobos, Peterson, Barve, et al., 2019; Cobos, Peterson, Osorio-Olvera, et al., 2019). Generally, VIF >10 is a signal of a model's excessive multicollinearity problem (Naimi et al., 2014).

From the MOP analysis, we also know that the optimal model does have highly extrapolative regions. Extrapolation in the model specifically refers to the model's capability to provide reliable predictions in areas that are too different from the areas in which the model is trained. This indicates that there is a gap in knowledge of the ability of feral hemp to grow in these areas, increasing sampling in these areas would lead to better predictive models.

Also, we did not include any anthropogenic factors in this study. Even though suitability expands with more extreme climate change scenarios, anthropogenic population shifts will likely occur during these times as well. Given that the dispersal of new feral populations is largely anthropologically driven, our models are not able to account for dispersal events by humans. Also, with few genotypes available within these populations, there is no way to distinguish what populations are remnants of historical hemp cultivation in the Americas, or European fiber, illicit high-THC types, etc. For example, we did not evaluate the extent of the correlation between predicted biologically suitable habitat and historical hemp production regions. We also are not able to account for biotic interactions within the ENM, how those biotic factors would respond to climate change, and how those interactions shift in reaction to climate change.

4.5 | Future models and their implications

Under short- and long-term predictions of low and severe climate scenarios, there is always an expansion in the potential niche for feral hemp. We observe range change in two ways: by examining range size in the future with and without dispersal. Dispersal refers specifically to the expansion of current feral populations and new anthropologically driven feralization events. Range size with dispersal allows us to describe the potential niche of feral *Cannabis* under every scenario (Table 3). At full dispersal, the potential niche of feral *Cannabis* expands under every scenario but expands most during the "meltdown" climate change scenario, SSP5-8.5°C from 2061 to 2080 (Figure 3d). Based on predictions of the potential expansion in abiotic suitability, this population could allow us to observe the contested Niche Variation Hypothesis, where populations with wider niches tend toward greater genetic variation (Van Valen, 1965). We can begin to ask questions such as: in areas where models predict niche expansion, do we observe more genotypic variation? The niche variation hypothesis asserts that there will be. As we continue to generate models, we should consider how niche variation is informed by genetics and use that lens to develop more complex models and hypotheses. As current research is underway to genotype and characterize the Midwestern feral *C. sativa* type (Ellison, 2021), these efforts will allow for more utilitarian predictive models.

The regions of highest priority for germplasm collection are the regions that experience suitability loss under both the "normal" and "terrible" climate change scenarios. The main areas of concern for feral Cannabis population loss over the next 20 years are in (i) Kansas—in the Flint Hills, South of Topeka and East of Wichita, (ii) Illinois and Missouri—along the Mississippi River, around Chester and St. Louis, respectively, (iii) Illinois—along the Fairfield basin and the cities of Naperville, Aurora, Rockford, Rochelle, LaSalle, and Rock Falls, (iv) Indiana—along the East Fork White River, and (v) the

TABLE 3 Pixel-to-pixel comparison of each climate scenario to the current predicted suitable area in the Midwest.

Climate model	Loss (pixels)	Remain occupied (pixels)	Remain unoccupied (pixels)	Gain (pixels)	% Loss	% Gain	% Species range change	Current range size (pixels)	Future range size with no dispersal	Future range size with full dispersal
GISS-E2-1-H 2021-2040 SSP1-2.6C	6286	71,051	39,878	10,173	13.62	22.04	+8.42	46,164	39,878	50,051
GISS-E2-1-H 2061-2080 SSP1-2.6C	6950	63,386	39,214	17,838	15.06	38.64	+23.59	46,164	39,214	57,052
GISS-E2-1-H 2021-2040 SSP5-8.5C	5721	70,402	40,443	10,822	12.39	23.44	+11.05	46,164	40,443	51,265
GISS-E2-1-H 2061-2080 SSP5-8.5C	6992	60,874	39,172	20,350	15.15	44.08	+28.94	46,164	39,172	59,522

city of Bloomington, South Dakota—surrounding the Black Hills. Collections of these populations are critical to saving this germplasm in the case of a climate "meltdown". Lower priority for germplasm collection regions is within Michigan, Minnesota, Ohio, and Iowa.

Future hemp breeding efforts will rely on diverse genetic resources conserved in seed banks to achieve genetic gain and address important agricultural challenges including climate change and emerging pest and disease pressures. Milner et al. (2019) demonstrated that affordable high-throughput genotyping provided defining characteristics of an accession and population structure analyses differentiated accessions according to domestication status and geographic origin within the German barley collection. This information can define subpopulation pools for breeding, and combined with extensive phenotyping of accessions, can be used to identify potential duplicate accessions and germplasm gaps within collections. Therefore, genotyping accessions as they become available will be an early priority for newly acquired feral accessions identified in habitats through this modeling work.

Typically, seedbank accessions include passport records with geographic origin information and information describing the taxonomy and breeding history of each accession. However, standardized phenotyping protocols, like the newly developed USDA Hemp Descriptor and Phenotyping Handbook (https://www.ars.usda.gov/northeast-area/geneva-ny/plant-genetic-resources-unit-pgru/docs/hemp-descriptors/), will also be essential tools for accession characterization. Phenotyping will be required for new germplasm acquisitions to identify highly heritable phenotypic traits that will be the focus of future breeding objectives. With the ongoing characterization of feral Cannabis, the germplasm collected could also aid in the development of standardization of phenotyping methods and sequencing protocols. In creating standards, breeders could accelerate their breeding programs.

5 | CONCLUSIONS AND QUESTIONS

Due to the unique history of *Cannabis sativa* and its interaction with humans, we underscore the importance of this subpopulation as a useful case study to illuminate anthropogenic, genetic, environmental, and climate-change interactions. We believe that the tools implemented in this work are currently underutilized by germplasm collectors, curators, and breeders and have applications within ongoing collection, conservation, characterization, and breeding efforts. Specifically, these ENMs or other predictive models could be critical to informing feral germplasm collection strategies, protocols, and priorities. Delineation of current and future species niche can inform cultural production practices and breeding objectives for enhanced abiotic tolerance.

AUTHOR CONTRIBUTIONS

Tori Ford: Conceptualization (lead); methodology (lead); validation (lead); visualization (lead); writing – original draft (equal); writing – review

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on GitHub at [https://github.com/tr-ford/USDA_Cannabis_sativa_Midwest_Feral_Biogeo.git], reference: 25, Ford and Stansell (2023).

ORCID

Tori Ford https://orcid.org/0000-0002-9764-6726

REFERENCES

Aina, A. (2023, May). Personal communication.

Alberti, P. (2021). Midwestern hemp database: 2020 research report. Commercial Agriculture Educator, University of Illinois. Retrieved February 19, 2021, from https://farmdoc.illinois.edu/field-crop-production/crop_production/midwestern-hemp-database-2020-research-report.html

Amici, V., Geri, F., Bonini, I., & Rocchini, D. (2014). Ecological niche modelling with herbarium data: A framework to improve Natura 2000 habitat monitoring. *Applied Ecology and Environmental Research*, *12*(3), 645–659.

Arif, S., Adams, D. C., & Wicknick, J. A. (2007). Bioclimatic modelling, morphology, and behaviour reveal alternative mechanisms regulating the distributions of two parapatric salamander species. *Evolutionary Ecology Research*, *9*, 843–854.

Belbin, L., Wallis, E., Hobern, D., & Zerger, A. (2021). The Atlas of Living Australia: History, current state and future directions. *Biodiversity Data Journal*, *9*, e65023. https://doi.org/10.3897/BDJ.9.e65023

Bohl, C. L., Kass, J. M., & Anderson, R. P. (2019). A new null model approach to quantify performance and significance for ecological niche models of species distributions. *Journal of Biogeography*, *46*, 1101–1111. https://doi.org/10.1111/jbi.13573

Boiko, O., Kagone, S., & Senay, G. B. (2021). Soil properties dataset in the United States: U.S. Geological Survey data release. https://doi.org/10.5066/P9TI3IS8

Borokini, I. T., Nussear, K., Petitpierre, B., Dilts, T. E., & Weisberg, P. J. (2023). Iterative species distribution modeling results in the discovery of novel populations of a rare cold desert perennial.

Endangered Species Research, *50*, 47–62. https://doi.org/10.3354/esr01218

Bothwell, H. M., Evans, L. M., Hersch-Green, E. I., Woolbright, S. A., Allan, G. J., & Whitham, T. G. (2021). Genetic data improves niche model discrimination and alters the direction and magnitude of climate change forecasts. *Ecological Applications*, *31*(3), e02254. https://doi.org/10.1002/eap.2254

Byrne, P. F., Volk, G. M., Gardner, C., Gore, M. A., Simon, P. W., & Smith, S. (2018). Sustaining the future of plant breeding: The critical role of the USDA-ARS National Plant Germplasm System. *Crop Science*, *58*, 451–468. https://doi.org/10.2135/cropsci2017.05.0303

Carlson, C. H., Stack, G. M., Jiang, Y., Taşkıran, B., Cala, A. R., Toth, J. A., Philippe, G., Rose, J. K. C., Smart, C. D., & Smart, L. B. (2021). Morphometric relationships and their contribution to biomass and cannabinoid yield in hybrids of hemp (*Cannabis sativa*). *Journal of Experimental Botany*, *72*(22), 7694–7709. https://doi.org/10.1093/jxb/erab346

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2023). rgbif: Interface to the global biodiversity information facility API. R package version 3.7.5. https://CRAN.R-project.org/package=rgbif

Chiarenza, A. A., Mannion, P. D., Lunt, D. J., Farnsworth, A., Jones, L. A., Kelland, S. J., & Allison, P. A. (2019). Ecological niche modelling does not support climatically-driven dinosaur diversity decline before the Cretaceous/Paleogene mass extinction. *Nature Communications*, *10*, 1091. https://doi.org/10.1038/s41467-019-08997-2

Clarke, R., & Merlin, M. (2016). *Cannabis: Evolution and ethnobotany*. University of California Press.

Clarke, R. C., & Merlin, M. D. (2015). Letter to the Editor: Small, Ernest. 2015. Evolution and classification of *Cannabis sativa* (marijuana, hemp) in relation to human utilization. Botanical Review 81(3): 189-294. *Botanical Review*, *81*(4), 295–305.

Cobos, M. E., Peterson, A. T., Barve, N., & Osorio-Olvera, L. (2019). kuenm: An R package for detailed development of ecological niche models using Maxent. *PeerJ*, *7*, e6281. https://doi.org/10.7717/peerj.6281

Cobos, M. E., Peterson, A. T., Osorio-Olvera, L., & Jiménez-García, D. (2019). An exhaustive analysis of heuristic methods for variable selection in ecological niche modeling and species distribution modeling. *Ecological Informatics*, *53*, 100983. https://doi.org/10.1016/j.ecoinf.2019.100983

Conrad, C. (1999). Legislator blasts hemp eradication programs. Hempology: Hemp History Library and Museum. http://hempology.org/CURRENT%20HISTORY/1996%20DEARADICATION.html

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N., & Guisan, A. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, *40*, 774–787. https://doi.org/10.1111/ecog.02671

Ellison, S. (2021). Collection, pre-characterization, and preservation of American hemp germplasm. USDA: REEIS. https://portal.nifa.usda.gov/web/crisprojectpages/1027159-collection-pre-characterization-and-preservation-of-american-hemp-germplasm.html

Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315. https://doi.org/10.1002/joc.5086

Filer, C. N. (2020). Minnesota wild hemp: A crucial botanical source in early cannabinoid discovery. *Journal of Cannabis Research*, *2*, 25. https://doi.org/10.1186/s42238-020-00031-3

Ford, T. F., & Stansell, Z. J. (2023). USDA Cannabis sativa Midwest feral ecological niche modeling scripts. GitHub. https://github.com/tr-ford/USDA_Cannabis_sativa_Midwest_Feral_Biogeo.git

Forsythe, W. C., Rykiel, E. J., Jr., Stahl, R. S., Hsin-i, W., & Schoolfield, R. M. (1995). A model comparison for daylength as a function of latitude and day of the year. *Ecological Modelling*, *80*, 87–95.

Frankel, O. H. (1974). Genetic conservation: Our evolutionary responsibility. *Genetics*, *78*, 53–65.

Goldstein, B. (2023, June 23). Midwest drought: Corn and soybeans suffer as forecasters expect no quick relief for farmers. *Wisconsin Watch*. https://wisconsinwatch.org/2023/06/midwest-drought-corn-and-soybeans-suffer-as-forecasters-expect-no-quick-relief-for-farmers/

Gordon, T., Jin, Y., Gale, S., Rouse, M., Stoxen, S., Wanyera, R., Macharia, G., Randhawa, M., Bhavani, S., Brown-Guedira, G., & Marshall, D. (2021). Identification of winter habit bread wheat landraces in the national small grains collection with resistance to emerging stem rust pathogen variants. *Plant Disease, 105*, 3998–4005.

Haney, A., & Kutscheid, B. B. (1975). An ecological study of naturalized hemp (*Cannabis sativa* L.) in East-Central Illinois. *The American Midland Naturalist*, *93*(1), 1–24. https://doi.org/10.2307/2424101

Harmon, J. (2022). Hemp for Victory!: The history of hemp in America, embodied in an educational artifact. *International Textile and Apparel Association Annual Conference Proceedings*, *78*(1), 1–4. https://doi.org/10.31274/itaa.13836

Hausfather, Z. (2018, April 4). Explainer: How 'shared socioeconomic pathways' explore future climate change. *CarbonBrief*. https://www.carbonbrief.org/explainer-how-shared-socioeconomic-pathways-explore-future-climate-change/

Hedden, P. (2003). The genes of the Green revolution. *Trends in Genetics*, *19*, 5–9.

Henry, R. J., & Nevo, E. (2014). Exploring natural selection to guide breeding for agriculture. *Plant Biotechnology Journal*, *12*(6), 655–662. https://doi.org/10.1111/pbi.12215

Hijmans, R. (2022). geosphere: Spherical trigonometry. R package version 1.5-18. https://CRAN.R-project.org/package=geosphere

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2022). dismo: Species distribution modeling. R package version 1.3-9. https://CRAN.R-project.org/package=dismo. https://doi.org/10.1038/s41588-018-0266-x

Hudson, J. B. (1942, November 6). Letter to the Secretary: Prices under 1943 growers' hemp contract. WWII - HEMP FOR VICTORY ARCHIVE! Hempology: Hemp History Library and Museum.

Hutchinson, G. E. (1957). Concluding remarks–cold spring harbor symposia on quantitative biology. Reprinted in 1991: Classics in theoretical biology. *Bulletin of Mathematical Biology*, *53*(1507), 193–213.

Jiang, H., Wang, L., Merlin, M. D., Clarke, R. C., Pan, Y., Zhang, Y., Xiao, G., & Ding, X. (2016). Ancient cannabis burial shroud in a central Eurasian cemetery. *Economic Botany*, *70*, 213–221. https://doi.org/10.1007/s12231-016-9351-1

Johnson, N. (2019). American weed: A history of cannabis cultivation in the United States. *EchoGéo*, *48*, 1–22. https://doi.org/10.4000/echogeo.17650

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C., Pinilla-Buitrago, G. E., Boria, R. A., Soley-Guardia, M., & Anderson, R. P. (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. *Methods in Ecology and Evolution*, *12*, 1602–1608. https://doi.org/10.1111/2041-210X.13628

Kobayashi, M., Momohara, A., Okitsu, S., Yanagisawa, S., & Okamoto, T. (2008). Fossil hemp fruits in the earliest Jomon period from the Okinoshima site, Chiba Prefecture, Japan. *Japanese Journal of Historical Botany*, *16*(1), 11–18.

Legind, J. K. (2018, November 22). Using shapefiles on GBIF data with R. *GBIF: Data blog*. https://data-blog.gbif.org/post/shapefiles/

Lima, R. A. F., Sánchez-Tapia, A., Mortara, S. R., ter Steege, H., & Siqueira, M. F. (2021). plantR: An R package and workflow for managing species records from biological collections. *Methods in Ecology and Evolution*, *14*, 332–339. https://doi.org/10.1111/2041-210X.13779

Liu, F.-H., Hu, H.-R., Du, G.-H., Deng, G., & Yang, Y. (2017). Ethnobotanical research on origin, cultivation, distribution and utilization of hemp (*Cannabis sativa* L.) in China. *Indian Journal of Traditional Knowledge*, *16*(2), 235–242.

Lobo, J. M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, *33*, 103–114. https://doi.org/10.1111/j.1600-0587.2009.06039.x

Maitner, B. (2023). BIEN: Tools for accessing the botanical information and ecology network database. https://CRAN.R-project.org/package=BIEN

Martínez-Freiría, F., Freitas, I., Zuffi, M. A., Golay, P., Ursenbacher, S., & Velo-Antón, G. (2020). Climatic refugia boosted allopatric diversification in western Mediterranean vipers. *Journal of Biogeography*, *47*(8), 1698–1713. https://doi.org/10.1111/jbi.13861

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J. C., & Stein, N. (2019). Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nature Genetics*, *51*, 1076–1081. https://doi.org/10.1038/s41588-019-0443-6

McPartland, J. M., Hegman, W., & Long, T. (2019). Cannabis in Asia: Its center of origin and early cultivation, based on a synthesis of subfossil pollen and archaeobotanical studies. *Vegetation History and Archaeobotany*, *28*, 691–702. https://doi.org/10.1007/s00334-019-00731-8

McPartland, J. M., & Small, E. (2020). A classification of endangered high-THC cannabis (*Cannabis sativa* subsp. indica) domesticates and their wild relatives. *PhytoKeys*, *144*, 81–112. https://doi.org/10.3897/phytokeys.144.46700

Midwestern Regional Climate Center. (2024, March 18). Precipitation summary for station USW00014837 - MADISON DANE RGNL AP, WI. Purdue: MRCC. https://mrcc.purdue.edu/mw_climate/climateSummaries/climSummOut_pcpn?stnId=USW00014837

Milner, S. G., Jost, M., Taketa, S., Mazón, E. R., Himmelbach, A., Oppermann, M., Weise, S., Knüpffer, H., Basterrechea, M., König, P., & Schüler, D. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nature Genetics*, *51*, 319–326.

Naimi, B., Na, H., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling. *Ecography*, *37*, 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

National Oceanic and Atmospheric Administration. (2023, May 6). Drought severity index division weekly value for period ending May 06, 2023: Long term palmer. Midwest Regional Climate Center. https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/regional_monitoring/palmer.gif

Obreza, M. (2023). genesysr: Genesys PGR client. https://CRAN.R-project.org/package=genesysr

Phillips, S. J. (2017). A brief tutorial on Maxent. Retrieved May 17, 2023, from http://biodiversityinformatics.amnh.org/open_source/maxent/

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*, 887–893.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Regos, A., Arenas-Castro, S., Tapia, L., Domínguez, J., & Honrado, J. P. (2021). Using remotely sensed indicators of primary productivity to improve prioritization of conservation areas for top predators. *Ecological Indicators*, *125*, 107503. https://doi.org/10.1016/j.ecolind.2021.107503

Ren, G., Zhang, X., Li, Y., Ridout, K., Serrano-Serrano, M. L., Yang, Y., Liu, A., Ravikanth, G., Nawaz, M. A., Mumtaz, A. S., Salamin, N., & Fumagalli, L. (2021). Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa. Science Advances*, *7*(29), eabg2286. https://doi.org/10.1126/sciadv.abg2286

Riahi, K., Vuuren, D. V., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K. V., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., ... Tavoni, M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Sillero, N., Arenas-Castro, S., Enriquez-Urzelai, U., Vale, C., Guedes, D., Martínez-Freiría, F., Real, R., & Barbosa, A. M. (2021). Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. *Ecological Modelling*, *456*, 109671. https://doi.org/10.1016/j.ecolmodel.2021.109671

Singh, R. P., Hodson, D. P., Huerta-Espino, J., Jin, Y., Bhavani, S., Njau, P., Herrera-Foessel, S., Singh, P. K., Singh, S., & Govindan, V. (2011). The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annual Review of Phytopathology*, *49*, 465–481.

Smith, A. B., Godsoe, W., Rodríguez-Sánchez, F., Wang, H. H., & Warren, D. (2019). Niche estimation above and below the species level. *Trends in Ecology & Evolution*, *34*(3), 260–273. https://doi.org/10.1016/j.tree.2018.10.012

Smith, A. B., & Santos, M. J. (2020). Testing the ability of species distribution models to infer variable importance. *Ecography*, *43*, 1801–1813. https://doi.org/10.1111/ecog.05317

Smithsonian National Museum of Natural History. (Smithsoanian NMNH; 2023, January 27). Botany collections. https://collections.nmnh.si.edu/search/botany/

Soberon, J., & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, *2*, 1–10. https://doi.org/10.17161/bi.v2i0.4

Solomon, R. (2020). Racism and its effect on cannabis research. *Cannabis and Cannabinoid Research*, *5*(1), 2–5. https://doi.org/10.1089/can.2019.0063

Sony, R. K., Sen, S., Kumar, S., Sen, M., & Jayahari, K. M. (2018). Niche models inform the effects of climate change on the endangered Nilgiri Tahr (*Nilgiritragus hylocrius*) populations in the southern Western Ghats, India. *Journal of Ecological Engineering*, *120*, 355–363. https://doi.org/10.1016/j.ecoleng.2018.06.017

The vPlants Project. (2001). vPlants: A virtual herbarium of the Chicago region. http://www.vplants.org

Title, P. O., & Bemmels, J. B. (2018). ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, *41*, 291–307. https://doi.org/10.1111/ecog.02880

Torkamaneh, D., & Jones, A. M. P. (2021). Cannabis, the multibillion dollar plant that no genebank wanted. *Genome*, *65*(1), 1–5. https://doi.org/10.1139/gen-2021-0016

United States Department of Agriculture - Agricultural Research Service. (USDA-ARS; 2023, February 6). U.S. national plant germplasm system. GRIN-Global. https://npgsweb.ars-grin.gov/gringlobal/search

Van Valen, L. (1965). Morphological variation and width of ecological niche. *The American Naturalist*, *99*(908), 377–390.

Wani, I. A., Verma, S., Mushtaq, S., Alsahli, A. A., Alyemeni, M. N., Tariq, M., & Pant, S. (2021). Ecological analysis and environmental niche modelling of *Dactylorhiza hatagirea* (D. Don) Soo: A conservation approach for critically endangered medicinal orchid. *Saudi Journal of Biological Sciences*, *28*(4), 2109–2122. https://doi.org/10.1016/j.sjbs.2021.01.054

Warren, D. L., Wright, A. N., Seifert, S. N., & Shaffer, H. B. (2014). Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. *Diversity and Distributions*, *20*, 334–343. https://doi.org/10.1111/ddi.12160

Woods, P., Price, N., Matthews, P., & McKay, J. K. (2023). Genome-wide polymorphism and genic selection in feral and domesticated lineages of *Cannabis sativa*. *G3: Genes, Genomes, Genetics*, *13*(2), jkac209. https://doi.org/10.1093/g3journal/jkac209

Zeng, Y., Low, B. W., & Yeo, D. C. (2016). Novel methods to select environmental variables in MaxEnt: A case study using invasive crayfish. *Ecological Modelling*, *341*, 5–13. https://doi.org/10.1016/j.ecolmodel.2016.09.019

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.