

ARTICLE

Vegetation Ecology

# Synthesizing historical plant collections to identify priorities for future collection efforts and research applications

Elizabeth M. Lombardi<sup>1,2,3</sup>  | Harpo Faust<sup>1,2</sup> | Hannah E. Marx<sup>1,2,4</sup>

<sup>1</sup>Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA

<sup>2</sup>Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico, USA

<sup>3</sup>Department of Biology, University of Minnesota Duluth, Duluth, Minnesota, USA

<sup>4</sup>Section of Plant Biology and the L.H. Bailey Hortorium, School of Integrative Plant Science, Cornell University, Ithaca, New York, USA

**Correspondence**

Elizabeth M. Lombardi  
 Email: [emlombardi@unm.edu](mailto:emlombardi@unm.edu)

**Handling Editor:** Charles D. Canham

## Abstract

To understand how and where biodiversity is threatened, it is imperative to build historical baselines that accurately characterize the present and past states of biodiversity across environments. Botanical collections provide important ecological, evolutionary, and biogeographic information on the diversity and distributions of plant taxa, yet biases in collection efforts across spatial, temporal, and taxonomic scales are well known. Here, we characterize and quantify trends in botanical collections made from across different abiotic, biotic, and sociopolitical boundaries within the present-day state of New Mexico. Using a biodiversity informatics approach applied toward a regional case study, we identify opportunities for efficiently improving natural history collection coverage and analyses of botanical diversity. Accurate representation of botanical biodiversity, preserved for future generations through vouchered plant specimens deposited in herbaria, depends on collection decisions made now. This work aims to provide a useful workflow for synthesizing digitized regional botanical collections as researchers prioritize current and future resources in the face of global change.

## KEY WORDS

biodiversity, botanical collections, digitization, herbarium, reproducible science, spatial phylogenetics, taxonomic harmonization

## INTRODUCTION

To measure and predict the consequences of climate and land use change on biodiversity, it is critical to understand the gaps or biases in baseline data. Historical plant collections held in herbaria offer a wealth of biodiversity information to understand species distributions and adaptation across environments (Funk, 2003, 2018; Lendemer et al., 2020) but are also limited by uneven and uncertain temporal, taxonomic, and spatial collecting

(Meyer et al., 2016). Digitization of collection information from preserved physical specimens has made biodiversity data more accessible (Heberling et al., 2019), and novel methods facilitate data use beyond traditional systematics (Davis, 2023; Heberling & Isaac, 2017). However, modern collection rates are historically low (Prather et al., 2004; Tewksbury et al., 2014). As a result, our present era is on course to be yet another source of uncertainty and limited sampling (Daru et al., 2018; Meyer et al., 2016) against which biodiversity metrics will be measured,

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Ecosphere* published by Wiley Periodicals LLC on behalf of The Ecological Society of America.

unless collection and curation interventions are made now. Unevenly collected taxonomic groups and regions require prompt attention in order to revise insights built from biased records and, critically, to improve biodiversity baselines that are necessary for monitoring change going forward.

Herbarium specimens generally represent inconsistent and opportunistic approaches to botanical collecting (Daru et al., 2018; Lang et al., 2019; Lavoie, 2013). Effort focused on outputs such as site-specific floras, land surveys, long-term monitoring of focal plots or taxa and focused systematic and taxonomic research (D'Antraccoli et al., 2022; Daru et al., 2018; Troudet et al., 2017) have provided botanical information that is irreplaceable and also incomplete. As a result, analyses of gaps in records of species diversity and distributions, predictions regarding the impacts of climate change on plant biodiversity (Heberling et al., 2019; Lang et al., 2019), and other modern uses of herbarium data are therefore also biased and incomplete (Cornwell et al., 2019). Rapidly changing environments and associated biodiversity crises lend urgency to synthesizing botanical collections made across spatial and environmental clines and require inclusive, accessible methods for efficient data synthesis.

In this paper, we take a fundamental approach to data exploration of museum collections by synthesizing trends in digitized herbarium data across temporal, taxonomic, and spatial scales from botanical collections made in the contemporary state of New Mexico. As a case study, New Mexico's botanical diversity and history of collections is compelling due to its unique floristic assemblage, rich environmental heterogeneity, density of rare and endemic species, exceptionally high level of taxonomic diversity, but inadequate collection coverage. Our main objectives are to (1) quantitatively synthesize plant collections across time, taxonomic lineages, and space (ecological, climatic, and sociopolitical categories) and (2) provide data management and geocomputational workflows that can be leveraged for biodiversity analyses in other areas.

Despite uncertain environmental conditions and a growing biodiversity crisis that necessitates greater monitoring and protection, there has been a decline in support for important scientific and collections-based research (Edwards et al., 2024; Lavoie, 2013; Thiers et al., 2024). Demonstrating the value of biodiversity collections is becoming a central duty of curatorial positions in order to highlight the nearly limitless research applications of natural history collections and ensure continued support for curation and protection of these irreplaceable data resources. While retroactive collecting with the goal of creating a more accurate and robust baseline is not possible, we can learn from existing collections through data

synthesis and analysis and by leveraging these insights into targeted research going forward. We discuss how our strategic synthesis can be used to identify opportunities for past collections and prioritize future collecting efforts for biodiversity research.

## MATERIALS AND METHODS

### Case study area

To demonstrate the usefulness of collections data synthesis and to provide useful information for a botanically diverse region, we analyzed all botanical records for a particular area of interest—the contemporary state of New Mexico (c. 1812). Located in the southwestern region of North America, this area is characterized by high taxonomic and habitat diversity (Stein, 2002), but there is a comparative deficit of vouchered herbarium records relative to other western states with similar histories regarding statehood and collections-based science (Taylor, 2014). The climate is characterized as arid, mixed high-elevation grasslands and rocky desert environments with many diverse habitats. Across the region, conditions are variable in terms of seasonal temperature and precipitation (Notaro et al., 2012; Thomas et al., 2023) with characteristic monsoonal precipitation patterns delivering moisture in a bimodal pattern across New Mexico and parts of Arizona (Higgins et al., 1999; Sheppard et al., 2002; Vera et al., 2006). Global climate models generally predict ecologically consequential changes to monsoonal cycles, such as delayed spring and summer precipitation and increased fall precipitation (Cook & Seager, 2013; Hernandez & Chen, 2022) declines in taxonomic biodiversity across eastern grasslands and throughout the southern Rocky Mountain mountains associated with warming and highly variable precipitation (Notaro et al., 2012).

### Workflow overview

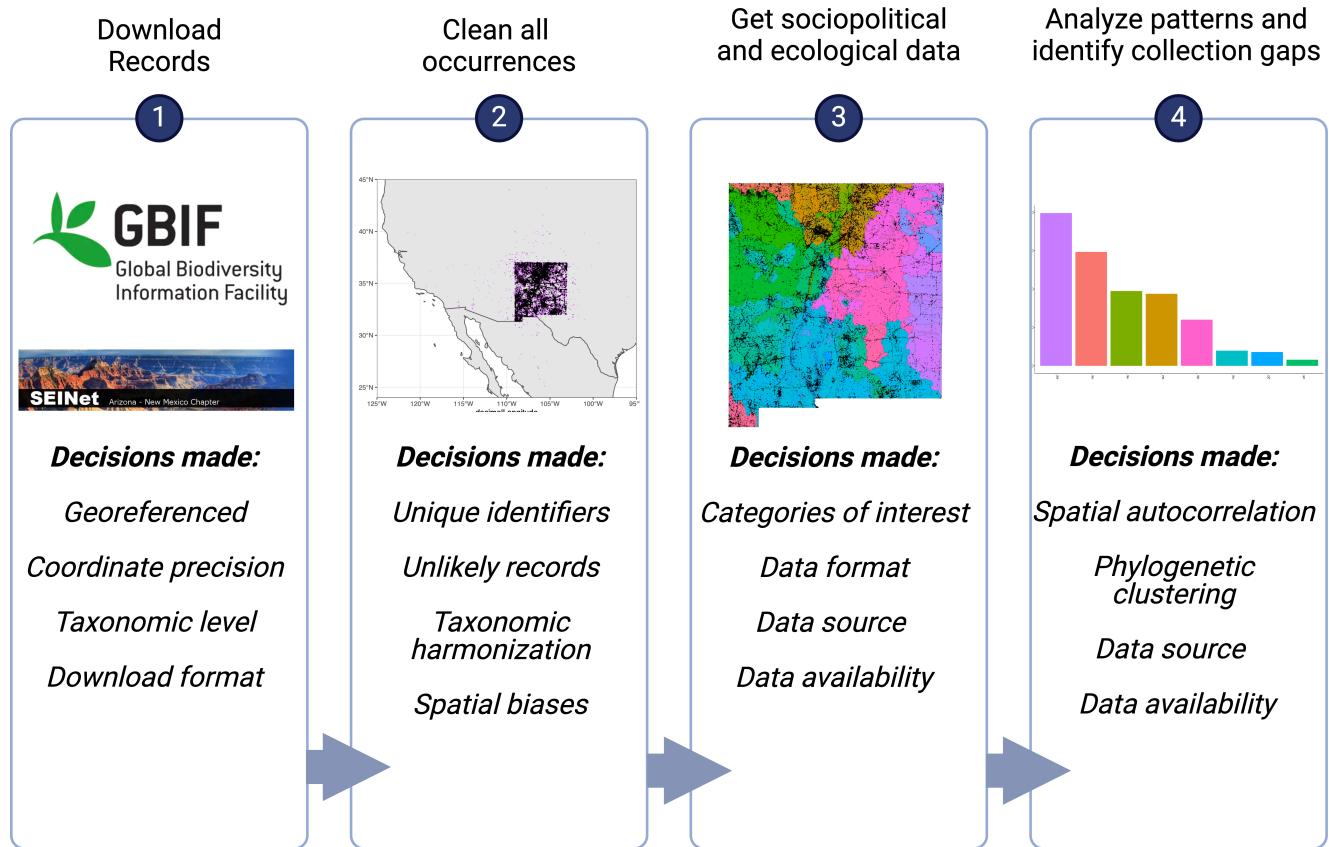
The data management and geocomputational workflow we developed to synthesize historical collections records from New Mexico consists of the following major steps: (1) downloading botanical collection records, (2) biodiversity occurrence cleaning, (3) spatial data selection and cleaning, and (4) temporal, taxonomic, and spatial pattern analyses. A brief description of the programs and parameters used for each step follows below; greater detail regarding computational resources and methodological approaches for each of the steps is provided in an extended methods section (Appendix S2).

All analyses were conducted in R version 4.2.2 (R Core Team). Scripts and data for all analyses are available from Zenodo (<https://doi.org/10.5281/zenodo.10778245>) as used in the presented workflow (see Figure 1) and with input and output requirements (see data tables in Zenodo project repository; <https://doi.org/10.5281/zenodo.10778245>). Figure aesthetics were improved for clarity using BioRender (BioRender.com).

## Biodiversity data selection and occurrence cleaning

All botanical collection records for vascular plants (Tracheophyta) were downloaded from both SEINet (SEINet [July 2023]) and the Global Biodiversity Information Facility (GBIF; GBIF Occurrence Download <https://doi.org/10.15468/dl.qyt96y>) on July 21, 2023. The original downloaded data included all digitized occurrence records that were classified as “Preserved Specimen” in the column “basisOfRecord.” There were 309,262 records for vascular plants that matched these parameters downloaded from GBIF and 465,003 records from SEINet. We removed duplicate records based on distinct collector

record number, scientific name and collection date (removed  $N = 17,755$  duplicate records from GBIF and 72,430 duplicate records from SEINet) in order to eliminate as many single specimens that were split between institutions and have multiple records. After cleaning each dataset separately, we standardized column names, combined them, and removed duplicate records shared between SEINet and GBIF based on catalog number. From this concatenated dataframe, we standardized categorical variables (“stateProvince” and “basisOfRecord”) and removed any records that had values other than “New Mexico” or “Physical specimen,” respectively ( $N = 11$  and  $N = 1247$  records removed). To clean temporal data, we removed records with uncertain or incorrect collection dates, such as those with no year data or collection dates in the future ( $N = 14,324$  records removed during temporal cleaning). Taxonomic dissimilarities were minimized by conducting harmonization between data sources at two separate times in our workflow. First, we harmonized species-level taxonomies for each of the original datasets (GBIF and SEINet) individually using the TNRS() function in the R package TNRS (version 0.3.4) (Maitner & Boyle, 2023) specifying only the World Catalog of Vascular Plants database as the taxonomic backbone



**FIGURE 1** Workflow diagram describing the main analytical stages used to characterize the history of botanical collections across biotic, abiotic, and sociopolitical environments in New Mexico. All resources are open access. Created with BioRender.com.

(Grenié et al., 2023). The scientific name with infraspecies information was harmonized without authority. Once the two datasets were concatenated, we removed any records without critical taxonomic information. This included removing 7198 records that did not have a matched specific epithet (i.e., result only had family and genus level taxonomic information) after harmonization and 831 records that did not have genus information after harmonization. Furthermore, improper taxonomic classification at the phylum level was an issue only for data downloaded from the SEINet repository, and we removed 4178 misclassified nonvascular specimen records. The temporally and taxonomically cleaned dataset is available as data object “dat” in the Rdata stack “datastack\_manuscript.Rdata” available through Zenodo (<https://doi.org/10.5281/zenodo.10778245>).

Spatial coordinate cleaning of records with georeferenced data was conducted in concurrence with best practices outlined in Zizka et al. (2021). For downstream spatial analyses, we retained records that met the following criteria: (1) included both latitude and longitude coordinates, (2) indicated coordinate precision below 1000-m accuracy when this information was available, and (3) coordinates that fell within the boundary delimiting the contemporary state of New Mexico. The georeferenced and spatially cleaned dataset is available as data object “env.v” in the Rdata stack “datastack\_manuscript.Rdata” available through Zenodo (<https://doi.org/10.5281/zenodo.10778245>). While revision of large-scale biodiversity data may be automated to identify and clean many sources of error, other issues arise due to irregularities, entry mistakes, or other non-systematic data inconsistencies. Keen manual revision and data exploration are necessary to identify these issues, which we did using R packages lubridate (version 1.9.2) (Grolemund & Wickham, 2011), dplyr (version 1.1.3) (Wickham, François, et al., 2023), and tidyr (version 1.3.0) (Wickham, Vaughan, et al., 2023) to identify unlikely or unusual trends or values in otherwise regular variables. We also leveraged taxonomic information from the Flora Neomexicana (hereafter “FNM”) (Allred & Jercinovic, 2020b).

## Environmental data selection and cleaning

Spatial data layers were collected and collated to characterize different biotic, abiotic, and sociopolitical environments across which collections have been made in New Mexico (Figure 1). We only used spatial data sources that are openly available either through direct API access or via downloads from open GIS data repositories (e.g., The Nature Conservancy). All spatial layers were projected into WGS84 and geolocation was checked with the R package

*rnaturrearth* (version 0.3.4) (South, n.d.) for alignment to the New Mexico state boundary. In order to compare abiotic information in raster format to the discrete vector data available for biotic and sociopolitical data, we evenly split the full range of each continuous variable (precipitation, temperature, and elevation) across New Mexico into 10 zones of equal bin width. Zonal minimum and maximum values for each elevation, precipitation, and temperature zones are defined in a supplemental data table (see data tables in Zenodo project repository; <https://doi.org/10.5281/zenodo.10778245>).

## Temporal, taxonomic, and spatial pattern analyses

Temporal collection evenness was characterized by tallying the total number of cleaned occurrences each year from 1800 through May 2022 and calculating summary statistics (minimum, maximum, mean, and median number of collections). Annual classification as “wet” or “dry” was based on whether or not the monthly average precipitation each year was greater or less than the 1901–2022 average of 35.15 centimeters/month, as calculated by the National Oceanic and Atmospheric Administration (NOAA, 2023, accessed on 12 September 2023). Intra-annual temporal variation was quantitatively described by tallying the number of collections made on each ordinal date (1–365) and calculating mean, median, and range.

Broad taxonomic trends were described using all temporally and taxonomically cleaned records (“dat” dataset) by summarizing the number of collections that have been made for each family, genus, and species ever collected from New Mexico. Due to the heavily right-skewed distribution, we used a logarithmic transformation to visually compare collection frequency for each taxon, and filtered for singly collected taxa based on uniqueness in post-harmonization species and genus data. Singly collected records are included in the full dataset.

We used the georeferenced dataset (“env.v”) to quantify spatially explicit patterns in plant diversity and collection. For each environment in each layer of the abiotic, biotic, and sociopolitical categories, we calculated the total number of records, the total number of species collected from each environment, the relative species diversity of collections in each category (relative species diversity is the total number of species in each environment divided by the total number of species across all environments in each layer), and Shannon diversity index (Shannon, 1948; Spellerberg & Fedor, 2003), which is the negative sum of the relative diversity in each environment multiplied by the natural log of the relative diversity in each environment.

To describe the spatial distribution of records across New Mexico, we first created a hexagonal grid across the state and calculated the number of records per grid cell with 10-km diameter. Categorical bins of collection coverage across grid cells (i.e., bin widths) were defined manually to best visualize spatial spread. We also calculated the degree to which all collection locations conform to spatial randomness using Ripley's  $K$  (Diggle et al., 2007). To determine how collection patterns vary by spatial category, we summarized statistics including counts of the total number of collection records and the number of species taxa collected across spatial categories. We calculated the minimum nearest neighbor distance for every record in our database and normalized these values within each spatial category such that the most to least clustered areas fall between 1 (most clustered) and 0 (least clustered). Similarly, the number of records and the number of species present across spatial categories was calculated both in absolute metrics and in terms of value per square kilometers in each category. To compare descriptive spatial statistics across areas of different size, we used the *sf* (version 1.0-14) (Pebesma, 2018) package and *spatstat* (version 3.0-6) (Baddeley, 2008) to calculate spatial areas and spatial statistics for each category. The best-collected and most species-rich taxonomically diverse areas resulted in a value of 1.0 in the normalized value scale, while the least well-collected and species taxonomically depauperate areas resulted in values closer to zero.

We used a megaphylogenetic approach to quantify phylogenetic structure trends overall and across different spatial layers. To obtain a hypothesis of phylogenetic relationships among all species in the combined collections dataset, we used the function *phylo.maker()* in the R package *U.Phylomaker* version 0.1.0 (Jin & Qian, 2023). The input megatree was based on the megatree of seed plants in Smith and Brown (2018) and the pteridophyte megatree (Zanne et al., 2015) and downloaded from GitHub ([https://github.com/megatrees/plant\\_20221117](https://github.com/megatrees/plant_20221117)). The phylogeny backbone for taxa in the dataset matching the input megaphylogeny was retained and species not in our collection dataset were dropped. Species missing from the input megaphylogeny were added to the backbone using “scenario 3” in the function *phylo.maker()*, which binds the tip of a new taxa halfway along the branch of the next higher taxonomic level based on a lookup table of genus family relationships provided through the *U.Phylomaker* documentation (<https://github.com/jinyizju/genus.family.relationship>).

For each spatial layer (e.g., southwestern biotic communities), the species in that area were transformed into a “community matrix” with the presence/absence of taxa across spatial subcategory variables, and the

megaphylogeny was pruned to taxa within the given spatial layer using the function *comparative.comm()* in the package *pez* version 1.2-4 (Pearse et al., 2015). Classic metrics of phylogenetic structure, including mean nearest taxon distance (MNTD) and mean pairwise distance (MPD) (Tucker et al., 2017), were calculated using *pez.dispersion()* in *pez*. Faith's phylogenetic diversity (PD) was calculated with the function *ses.pd()* in *picante* version 1.8.2 (Kembel et al., 2010). Standardized effect sizes (SES) for all metrics were calculated from a null distribution generated by swapping taxon labels and 999 replicates within each spatial layer. Positive SES values (SES MNTD, SES MPD, or SES PD > 0) indicate phylogenetic overdispersion, or a greater phylogenetic distance among co-occurring species than expected, while negative SES values indicate phylogenetic clustering, or small phylogenetic distances among co-occurring species than expected (Kembel et al., 2010). Finally, the number of taxa that were missing from the backbone phylogeny were recorded to summarize the percentage of species missing phylogenetic information across space.

Statistical differences in patterns of collecting and species diversity across categorical environments was determined using multiple linear regression models. To examine relationships between environments and across all layers, species diversity was treated as the response variable and the total number of records, clustering, and factorized environmental layers were independent variables. For each exemplary layer in the main analyses, we also tested simple linear relationships between the number of records and number of species collected from across independent environments in each layer separately. All multiple linear regressions were done using the R package *lme4* (version 1.1.35.1; Bates et al., 2015).

## RESULTS

Results regarding temporal and taxonomic trends in vascular plants collected from the state of New Mexico are based on 349,986 digitized specimens that were retained following initial cleaning and harmonization. Of these records, 219,216 (62.6%) had location coordinates of sufficient accuracy and were used in spatial analyses. The average number of digitized New Mexico specimens held across 252 institutions is 1383 (range 1–93,900 records per institution) with an average georeferencing rate of 35% (see data tables in Zenodo project repository; <https://doi.org/10.5281/zenodo.10778245>). We report that the 349,986 digitized specimens represent 7210 species and infraspecific taxa across 1688 genera and 167 families

(see data tables in Zenodo project repository; <https://doi.org/10.5281/zenodo.10778245>). There is spatial clustering and uneven collection across space with areas that have relatively high collection rates (e.g., the Chihuahuan Desert) surrounded by relatively low collections rates (e.g., eastern New Mexico) (Figure 2).

## Temporal trends

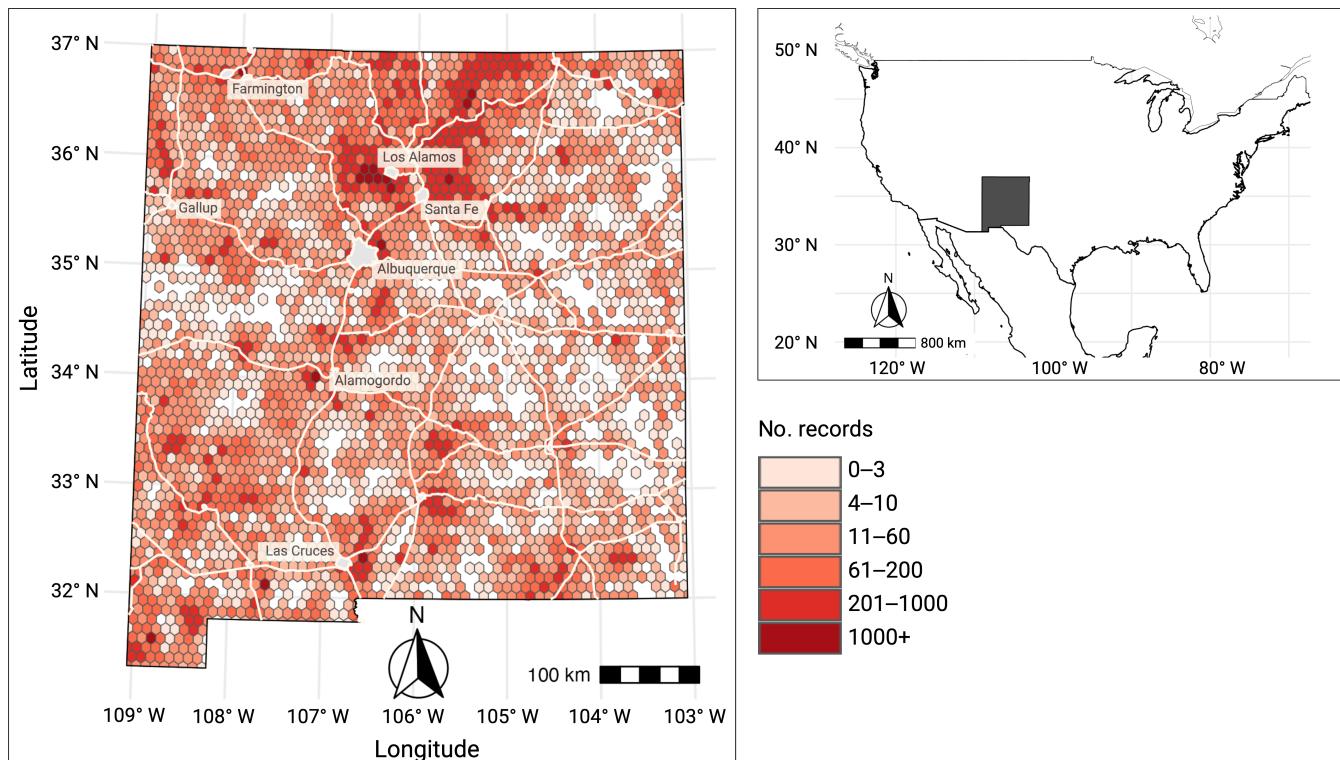
The earliest data display initial collections occurring in 1804 and 1805, but manual review suggests that these very early records were entered incorrectly into digital databases. We conclude that the first valid collections were made by William Gambel in 1841, as has been previously reported (Allred & Jercinovic, 2020b; Dickerman, 1985). Sparse collections were recorded until approximately 1890, after which the number of individual plants and collected species increases but fluctuates (Figure 3A). Between 1800 and 2022 a mean of 1699 collection records were made each year (median 1049 records/year). The greatest number of individual collections made in a single year was in 2005 ( $N = 10,586$ ), which corresponded with the greatest number of collections made per decade in the early 2000s (from 2000 to

2010; Figure 3). There has been a steady and steep decrease in collections since the early 2000s (Figure 3A) with a recent inflection and subsequent rise in annual collections made since 2022.

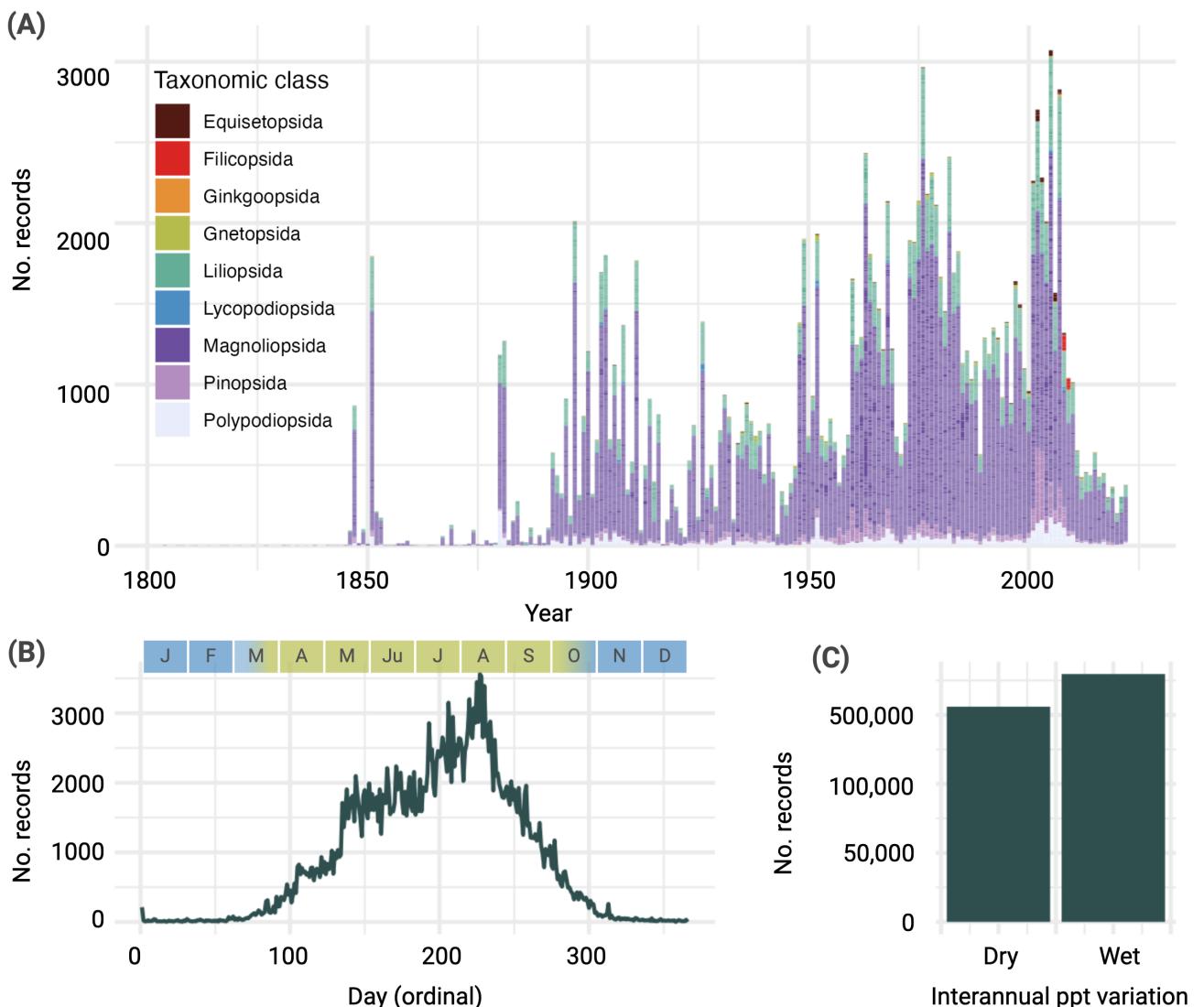
Intra-annual patterns in collections track the typical southwestern regional growing season from approximately April through September (Figure 3B). Average collections made per month were highest in August ( $N = 501.81$  records on average in August), with August 15 as the best-collected date throughout history ( $N = 2563$  records). Variation in collections across summer is normally distributed with no discernible peaks across the growing season. We did not find any peaks in collecting on holidays and a slight but insignificant increase in collections made on weekends.

Interannual variation in the number of records suggests more specimens are collected during wet years than dry years ( $N = 179,616$  collections in wet years since 1895 and  $N = 155,950$  in dry years since 1895). On average, there were 2897 records collected in each wet year compared with an average of 2363 records collected in dry years, but the difference is not statistically significant over time ( $p = 0.2124$ ; Figure 3C).

When compared with neighboring states over the full annual range considered here (1800–2022),



**FIGURE 2** Georeferenced botanical collections across the contemporary state of New Mexico, USA. Colors correspond to the normalized number of collections per 10-km<sup>2</sup> area. Urban areas and roads are demarcated in light yellow. White areas indicate regions without any collections that have been georeferenced. *Map Source:* Natural Earth data accessed via the R package rnaturrearth (Massicotte & South, 2023).



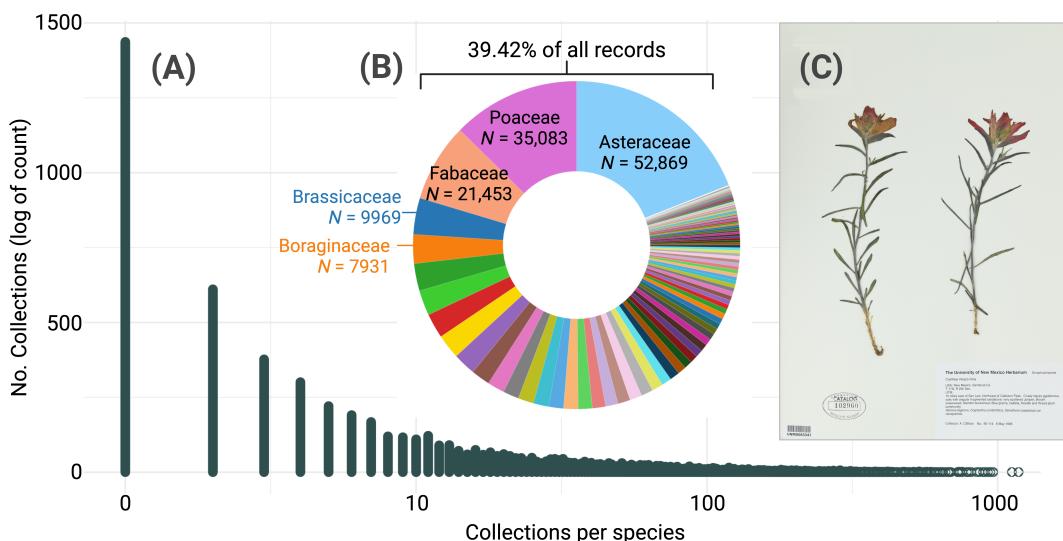
**FIGURE 3** Digitized botanical collections made across New Mexico over time. The total number of collections made each year (1800 through 2022) fluctuates, as does (A) the taxonomic class to which these collections belong. (B) More collections are made during spring through mid-fall each year, with a peak in August. (C) There were slightly more collections made in wet years compared with dry years, but this difference is not significant ( $p = 0.2124$ ). Interannual precipitation (ppt) variation categories are based on monthly mean precipitation data from NOAA.

New Mexico's collections trends indicate historic and current under-collection. Fewer records per square kilometer have been made in New Mexico than the number of records per square kilometer that have been made in the region as a whole ( $0.0046$  records/km $^2$  in NM compared with  $0.0069$  records/km $^2$  regionally;  $p = 0.0036$ ), though there is variation between regional states and across time (Appendix S1: Figure S3).

## Taxonomic and phylogenetic trends

While New Mexico has rich biodiversity, some taxonomic groups had many records made across time and habitat

space, while other groups had only a few records that provide limited information about the history and spatial extent of the taxa. The three best-collected plant families (Asteraceae, Poaceae, and Fabaceae) account for nearly 40% of all plant specimens on record (Figure 4B), while there are 1436 taxa in this dataset that are represented by only a single collection, a notably high number we discuss at length in the discussion (Figure 4A; Appendix S1: Figure S4A). Generally, singly collected taxa represent non-native species, misidentified specimens, records digitized with older nomenclature that failed to harmonize, or are taxa not included in the regional flora that do represent true first reports of rare species (Allred & Jercinovic, 2020b).



**FIGURE 4** Taxonomic diversity of digitized botanical collections made across New Mexico at the (A) species level (bar chart) and (B) the family level (pie chart). (C) The best-represented species in all collections from New Mexico is *Castilleja integra*. This specimen was collected by A. Clifford and is held at the UNM herbarium (UNM0063347).

Taxa from New Mexico are represented by an average number of 48.64 collection records/species and range from 1 to 1179 records/species. *Castilleja integra* was the most collected species with 1179 records, followed by *Quercus gambelii* (1118 records) and *Bouteloua gracilis* (964 records). Similarly, there was variability in the number of collections made at the genus level. The total number of genera in our dataset is 1688 with an average of 207.34 records/genus. There were 253 genera represented by a single collection. *Astragalus* was the most collected genus with 6830 collections made from across the 113 species represented in our dataset.

The number of taxa at every level, from subspecies to family, was higher in our dataset compared with an expert-compiled state flora, FNM (Allred & Jercinovic, 2020b). FNM came out in 2020 and documents 4184 vascular plant taxa from decades of study led by Kelly Allred and multiple New Mexico botanists. It is the most accurate expert-curated tally available that can be confidently used as a baseline. Recent additions include 9 species reported in addendums to the flora (Allred, 2021a, 2022b); 2 species from (Ackerfield, 2022); 11 species from (Rogers et al., 2023); 3 species from (Sivinski, 2023, 2024); and 1 species from Faust (2024), adding at least 24 taxa to the state flora for an updated total of 4209 verified taxa in the state. We identified 108 species listed only in the FNM and not represented in our cleaned collection synthesis, none of which were the aforementioned recent introductions, which is due to a variety of factors described in our discussion of taxonomic errors and cleaning methods. These species were evenly phylogenetically dispersed across the entire dataset ( $PD = 68,571.474$ ,  $pd.obs.z = 0.4450122$ ,  $pd.$

$obs.p = 0.675$ ) (Appendix S1: Figure S5). There were 3126 taxa represented in the collections but not listed in FNM. These taxa were also evenly phylogenetically dispersed across the entire flora ( $PD = 83,789.474$ ;  $pd.obs.z = 0.6628910$ ,  $pd.obs.p = 0.761$ ) (Appendix S1: Figure S5). The FNM represented fewer taxa with a lower PD ( $n = 4267$ ;  $PD = 68,571.474$ ) than the complete collections dataset ( $n = 7210$ ;  $PD = 100,805.776$ ).

## Spatial trends overview

More collections were made near cities and roads than in rural, less populated areas (Appendix S1: Figure S2). A disproportionate number of collections per area have been made in central/northern New Mexico, which is also reflected in elevated species richness that occurs in the southern Rocky Mountains/Sangre de Cristo mountains (Figure 2). Secondary collection hotspots include the Sacramento Mountains, the northern part of Navajo Nation near Farmington, and the area surrounding Las Cruces, particularly from the Organ Mountains.

Overall, the number of species collected from each categorical environment in New Mexico was positively correlated with the number of records collected ( $p = 2.2E-16$ ,  $R^2 = 0.8265$ ) (Appendix S1: Figure S6), and we detected clear spatial structure that deviated from spatial randomness and indicates significant spatial clustering in collection locations ( $p = 0.0009$ ) (Appendix S1: Figure S7). In comparing collections from across categorical environments, we found variation in the number of records collected from across unique environments in spatial layers.

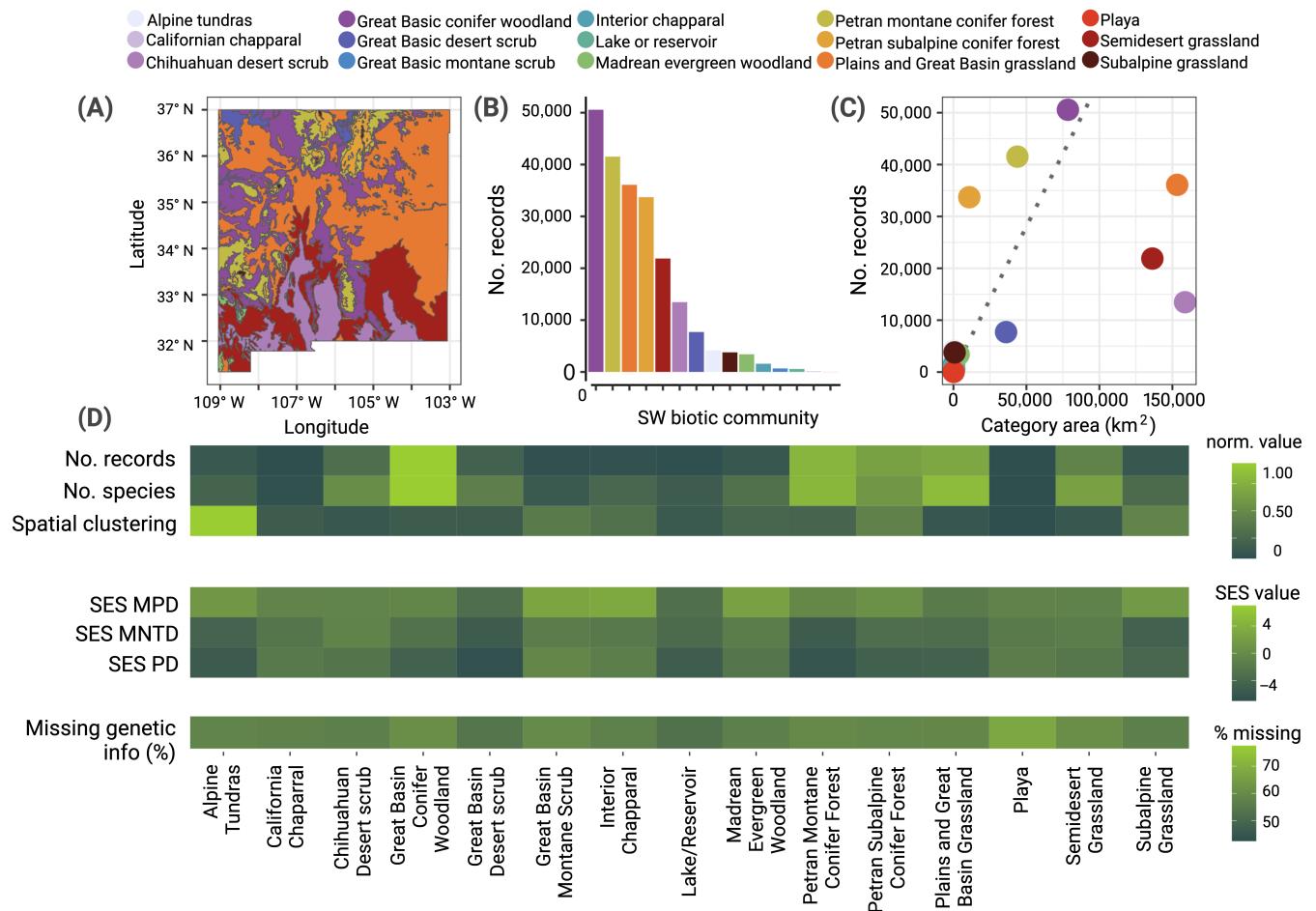
In particular, the number of collections made from any area with legal protection is remarkably low ( $p = 7.32E-09$ , SE = 4492), and there was statistically uneven sampling across environments in all other layers except for the evenly defined climatic and elevational zones. Similarly, the number of species from across environments varied by layer, and there was a significant positive correlation between the number of species and the number of records collected in each environment ( $p = 2.2E-16$ ,  $R^2 = 0.8802$ ). Statistically significant variation suggests uneven species representation across collections made from protected areas ( $p = 0.000925$ , SE = 1.585E+02) and physiographic regions ( $p = 0.035432$ , SE = 2.389E+02).

We examined a number of spatial point patterns from different biotic (ecoregions, physiographic regions, and Southwestern biotic communities), abiotic (major

geological processes, precipitation, and temperature) and sociopolitical (land surface ownership and political conservation protections) categories. Here, we focus on results from three exemplary spatial layers. Figures and model results that include the other spatial categories are available in the appendices (Appendix S1: Figures S8–S15).

## Spatial biotic category: Southwestern biotic communities

The Chihuahuan desert and dry grassland in the southern and eastern parts of New Mexico (Figure 5) were disproportionately under-represented by collections given their spatial area while montane and forest communities, particularly in the central and northern part of New Mexico,



**FIGURE 5** (A) Patterns in georeferenced collections made from southwestern biotic communities across New Mexico, as defined by The Nature Conservancy (TNC) southwestern biotic community analyses. (B) Uneven collection coverage across biotic communities is displayed in the bar graph displaying the total number of records per community and (C) via proportional representation of botanical records made across southwestern biotic communities relative to the area covered by each community. (D) Heatmaps compare normalized values across communities for the number of records, species, and spatial distances between records (top section); standardized effect sizes for mean pairwise distance (SES MPD), mean nearest taxon distance (SES MNTD), and phylogenetic diversity (SES PD) (middle section); and the proportion of species in each biotic community that were missing phylogenetic information (bottom section). Source: The Nature Conservancy, Biotic Communities of the Southwest GIS Layer (McCauley, 2021).

were comparatively over-represented (Figure 5B,C). Great Basin Conifer Woodland communities are the most represented in collections overall. Petran montane conifer forests were second best-represented in the number of collections made and were proportionally over-represented when you consider the area of New Mexico that hosts this forest type (Figure 5B,C). Variation in the number of records was also reflected with similar normalized values for the number of species collected from each biotic community (Figure 5D). Alpine tundras were not especially well collected in terms of total number of collections made, nor are they large areas. However, the distribution of collections made from alpine communities was highly spatially clustered (Figure 5D).

Standard effect sizes for PD and MNTD were negative for collections made from across all southwestern biotic communities, indicating that there is phylogenetic clustering within biotic communities across New Mexico. That said, there are some environments (interior chaparral, Great Basin Montane Scrub, Madrean evergreen woodland, subalpine grassland, alpine tundras, and Petran subalpine conifer forests) from which mean phylogenetic distance (MPD) was positive, indicating greater evenness across the phylogenetic tips of some communities despite clustering in PD and MNTD. Overall, the degree of phylogenetic clustering varied between southwestern biotic communities with the greatest phylogenetic distance between species collected from Great Basin Montane Scrub ( $PD = -0.4737669$ ,  $N = 673$ ) and the most clustering between species collected from the Great Basin Desert Scrub ( $PD = -6.5920885$ ,  $N = 7702$ ). There was a marginally significant positive correlation between the number of records collected in each biotic

community and the standard effect size of the PD across taxa in each biotic community ( $p = 0.0533$ ,  $SE = 2.7E-05$ ), and no significant effect of spatial clustering on the PD in each biotic community ( $p = 0.2625$ ,  $SE = 2.157E-02$ ). Overall, variation in the PD of collections made from across southwestern communities could not be explained by the distance between collections or the number of collections in each environment (Table 1;  $p = 0.1138$ ,  $R^2 = 0.1879$ ).

That said, we did find a positive correlation between total record numbers and the number of species collected from each biotic environment ( $p < 0.001$ ,  $R^2 = 0.8984$ ). Great Basin Conifer Woodland is the biotic community with the greatest relative species diversity of collections (3712 species from across 50,568 records) while the geographically small Playa community in southern New Mexico has the fewest records with the lowest relative species diversity (20 species from across 22 records).

The proportion of species in each biotic community that did not have available phylogenetic information ranged from 50.12% in Lake or Reservoir aquatic plant communities to 58.34% in semidesert grassland plant communities (Figure 5D).

## Spatial abiotic category: Elevational zones

The total number of records collected from across elevational zones was slightly skewed toward greater representation of mid to lower elevations (Figure 6B). However, proportional to the area in each band, mid to higher elevation areas were comparatively better represented than lower elevation areas (Figure 6C) and

**TABLE 1** Results of multiple linear regression to determine how collecting rates and spatial clustering impact the phylogenetic diversity of collections made from each environment in three exemplary categories (Southwestern biotic communities, elevation zones and land ownership types).

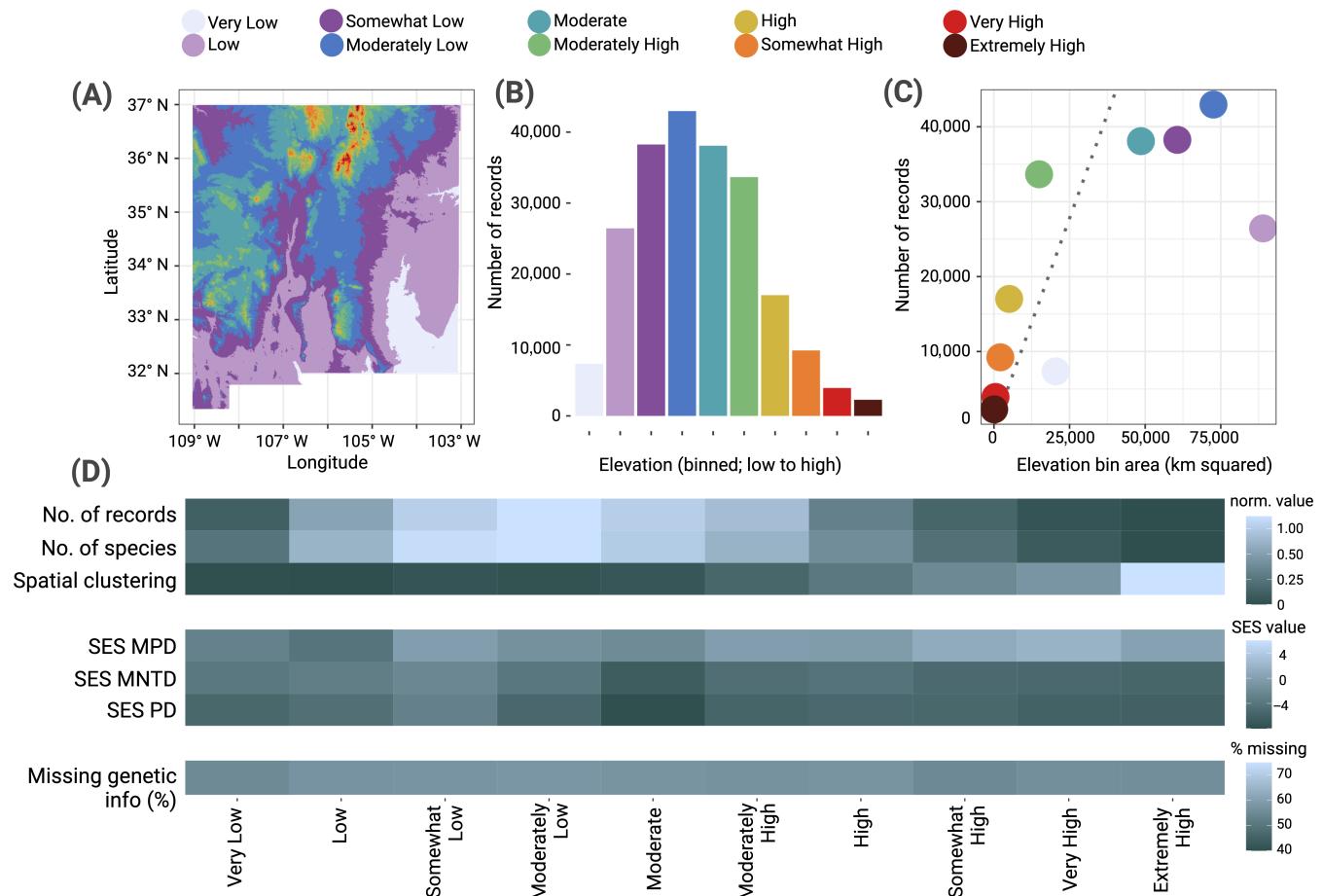
Model	<b>p</b>	<b>R</b> <sup>2</sup> adj	Parameters	Estimate	SE	<b>t</b>	Pr(> t )
Southwestern biotic communities	0.1138	0.1879	Intercept	-2.27	8.00E-01	-2.838	0.015 <sup>b</sup>
			<i>N</i>	-5.78E-05	2.70E-05	-2.143	0.0533 <sup>a</sup>
			Clustering	-2.54E-02	2.16E-02	-1.176	0.2625
Elevation zones	0.8257	-0.2173	Intercept	-4.11E+00	1.48E+00	-2.767	0.0278 <sup>b</sup>
			<i>N</i>	-3.76E-06	4.05E-05	-0.093	0.9287
			Clustering	-1.14E-02	2.24E-02	-0.511	0.6251
Land ownership type	0.6082	-0.0864	Intercept	-4.22E+00	9.82E-01	-4.303	0.0015 <sup>c</sup>
			<i>N</i>	3.09E-06	2.32E-05	0.133	0.89683
			Clustering	4.24E-02	4.16E-02	1.018	0.33264

Note: *N* is the number of records collected from each environment. Overall model performance for each layer is reported.

<sup>a</sup> $p = 0.1$ ;

<sup>b</sup> $p = 0.05$ ;

<sup>c</sup> $p = 0.001$ .



**FIGURE 6** (A) Distribution of georeferenced collections made from across the elevational gradient that occurs in New Mexico. Elevation was binned into 10 elevational bands of equal width. Bins are defined as follows: very low (865–1176 m), low (1177–1488 m), somewhat low (1489–1800 m), moderately low (1801–2112 m), moderate (2113–2424 m), moderately high (2425–2736 m), high (2737–3048 m), somewhat high (3049–3360 m), very high (3361–3671 m), and extremely high (3676–3984 m). (B) Evenness of collection coverage across elevation is displayed in the bar graph displaying the total number of records per zone and (C) via proportional representation of botanical records made across elevation relative to the area covered by each zone. (D) Heatmaps compare normalized values across elevation bins for the number of records, species, and spatial distances between records (top section); standardized effect sizes for mean pairwise distance (SES MPD), mean nearest taxon distance (SES MNTD), and phylogenetic diversity (SES PD) (middle section); and the proportion of species in each elevation bin that were missing phylogenetic information (bottom section). Source: Amazon Web Service Terrain Tiles (accessed 24 April 2023).

spatial clustering increased with elevation (Figure 6D). The lower elevational zones occur in southern and eastern New Mexico and the highest elevational zones are in the northern part of the state, which is the southern termini of the Rocky Mountains and the San Juan Mountain ranges (Figure 6A).

SES for phylogenetic relationships among species were negative for both PD and MNTD across all 10 elevation zones (Figure 6D) indicating phylogenetic clustering among plant communities that occur in the same approximate elevational bands. However, MPD values indicated variation in tree-wide phylogenetic patterns with positive MPD scores for all but the four lowest elevation zones despite negative PD and MNTD (Figure 6D). Overall, the

degree of phylogenetic clustering varied across elevation zones. The greatest SES for PD was from “moderate” elevations ( $PD = -5.5656$ ,  $N = 38,083$ ) while the lowest was found in “somewhat low” elevation zones ( $PD = -1.8984$ ,  $N = 38,256$ ). As with biotic community analyses, variation in the PD of collections made from across elevation zones could not be explained by the distance between collections or the number of collections in each environment (Table 1;  $p = 0.8257$ ,  $R_{adj}^2 = -0.2173$ ).

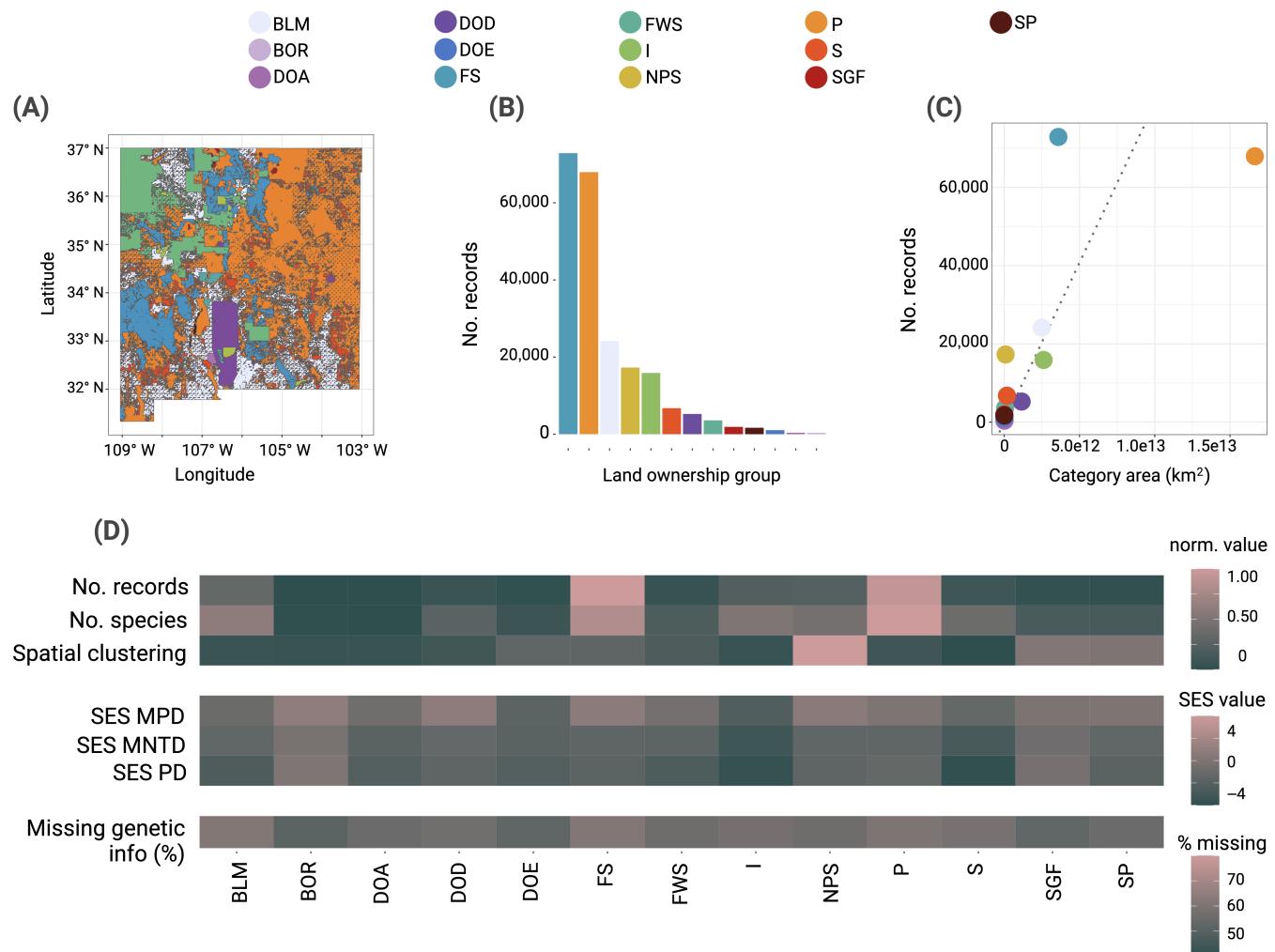
As seen across southwestern biotic communities, the number of species collected from each elevation zone was also positively correlated with the number of records made in each zone ( $p < 0.001$ ,  $SE = 0.0019$ ). The “moderately low” elevation zone had both the greatest number

of records and the highest relative taxonomic diversity of collections (3587 species from across 42,966 records). The highest elevation zone, “extremely high,” was represented by the fewest total records and the lowest relative taxonomic diversity (444 species from across 2275 records). There was an effect of the degree of spatial clustering within each elevational zone on the number of species collected ( $p = 0.021$ ,  $SE = 1.977$ ) (see data tables in Zenodo project repository; <https://doi.org/10.5281/zenodo.10778245>), which could also be seen qualitatively in the density of collections made in the northern, mountainous part of New Mexico (Figure 2).

The percentage of species in each elevation zone without phylogenetic information ranged from 54.74% of plants in “somewhat high” elevation zones to 57.65% of plants in “moderately low” elevation zones (Figure 6D).

## Spatial sociopolitical category: Land ownership

The number of collections made across different land ownership areas was skewed toward certain management types (Figure 7). Forest Service lands (“FS”),



**FIGURE 7** (A) Distribution of georeferenced collections made from across the surface land ownership types in New Mexico. (B) Evenness of collection coverage across land ownership is displayed in the bar graph displaying the total number of records per type and (C) via proportional representation of botanical records made across land ownership types relative to the area of New Mexico managed by each ownership type. (D) Heatmaps compare normalized values across ownership for the number of records, species and spatial distances between records (top section); standardized effect sizes for mean pairwise distance (SES MPD), mean nearest taxon distance (SES MNTD), and phylogenetic diversity (SES PD) (middle section); and the proportion of species in each ownership bin that were missing phylogenetic information (bottom section). Land ownership category abbreviations are as follows: Bureau of Land Management (BLM), Bureau of Reclamation (BOR), Department of Agriculture (DOA), Department of Defense (DOD), Department of Energy (DOE), Forest Service (FS), Fish and Wildlife Service (FWS), tribal lands (I), National Park Service (NPS), private lands (P), New Mexico state owned land (S), New Mexico State Game and Fish (SGF), New Mexico State Park (SP). Source: U.S. Bureau of Land Management (BLM) and New Mexico State Office (accessed 24 April 2023).

which occur across the state but most commonly in northern and western New Mexico, were the best-collected ownership category (Figure 7B) and proportionately best-represented in botanical collections by area (Figure 7C). Private ownership ("P") areas were well collected by the total number of records made (Figure 7B) but were disproportionately under-represented when you consider the large area of the state that is privately owned (Figure 7C). All other land ownership categories were smaller in area and have been collected less often than FS and P lands, though lands managed by the National Park Service (NPS) were relatively well represented given the small total area that falls in this category and the high degree of spatial clustering (Figure 7C,D). Across all ownership types, however, spatial clustering did not significantly impact the number of records collected ( $p = 0.907$ , SE = 540.04).

SES for phylogenetic relationships were negative for both PD and MNTD across all land ownership types except for Bureau of Reclamation (BOR) lands, which had a slightly positive SES for PD and the least negative SES for MNTD (Figure 7D). Like results from analyses across southwestern biotic communities and elevation zones, most plant communities were overall phylogenetically clustered regardless of ownership category (Table 1;  $p = 0.6082$ ,  $R_{adj}^2 = -0.0864$ ). Similarly, we found positive MPD that indicated overdispersion in tree-wide phylogenetic patterns for plant communities collected from some land ownership types (BOR, Department of Defense [DOD], FS, and NPS), despite otherwise clustered phylogenetic patterns (Figure 7D). The greatest SES for PD was in plant communities collected from New Mexico state owned land ("S") (PD =  $-6.7268$ ,  $N = 6762$ ), which was also the land ownership category with the second-strongest SES for MNTD after tribal lands ("I").

As reported in comparisons of taxonomic diversity across biotic communities and elevation zones, there was a positive and significant correlation between the number of species recorded and the number of collections made from each land ownership type ( $p < 0.001$ , SE = 0.00674) (see data tables in Zenodo project repository; <https://doi.org/10.5281/zenodo.10778245>), with the greatest relative species richness from collections made from private lands (4651 species from across 6799 records) and the lowest relative species diversity recorded for collections made on Department of Agriculture (DOA) lands (255 species from across 391 records). There was no impact of spatial clustering on the number of species detected ( $p = 0.89150$ , SE = 12.083031).

Species with missing phylogenetic information ranged from 58.56% on BLM (U.S. Bureau of Land Management) lands to 49.63% on BOR lands (Figure 7D).

## DISCUSSION

### Temporal trends

Since 1800, temporal patterns in botanical collections made from across New Mexico reflect periods of highs and lows that conform to national and regional oscillations (Appendix S1: Figure S3; Meyer et al., 2016), with an additional peak in the early 21st century. Our temporal results reflect similar peaks and troughs in collecting as those from nearby regions, though our data also suggest that there was a regionally specific recent increase in collecting between 2000 and 2010 and a relatively late decline ( $N = 475$  collections made in 2020). The period of 2000–2010 aligns with the height of a floristic studies push in New Mexico made by students at the Rocky Mountain Herbarium (RM) during that time. A few significant regional floristic projects contributed over 40,000 collections from the northern/central New Mexico mountains (Fowler et al., 2014; Larson et al., 2014; Larson, 2008; Legler, 2010; Taylor, 2014). These collection-based research projects relied on access to a mix of Forest Service and Private lands and illustrate the impact of collector bias on temporal consistency of collections and spatial coverage.

This most recent decline in collecting corresponded with the beginning of the SARS-CoV-2 pandemic, which may have disrupted collection trips. We saw similar dips in collection rates during other times of global insecurity, such as WWI and WWII ( $N = 69$  and  $N = 163$  new records added in 1917 and 1943, respectively), though other peaks and troughs lasted longer (e.g., relatively low collections in the late 1980s through early 1990s). These multiyear patterns in collecting corresponded with political decisions that supported collecting (e.g., the Endangered Species Act in 1976 was followed by a rise in the number of collections made from New Mexico) and national funding priorities (e.g., collections-based research was not funded often in the 1990s) that influenced availability of resources for collecting expeditions (Daru et al., 2018; Lipscomb et al., 2000). That said, temporal trends in herbarium collections made in neighboring states differ from the trends in New Mexico, which suggests that some patterns detected in our regional analysis were governed by state or local factors (Appendix S1: Figure S3). Our results implicate human societal unrest, policy, research priorities, and funding as factors that have impacted temporal evenness across this history of New Mexico's botanical collections (Crawford & Hoagland, 2009).

Intra-annual collection trends corresponded largely with seasonality across southwestern desert environments of the northern hemisphere, which is to say that the

collecting season begins in spring and ends in mid-autumn each year. We found no evidence that more collections are made on holidays or weekends, which suggests that professional botanists, collectors, land managers, and other natural historians without traditional work week limitations have been critical in building botanical collections across New Mexico. This trend emphasizes the importance of providing sustained support for formal and structured collections-based research as a way to maintain and improve biodiversity information.

The final temporal trend that we considered was interannual variation in collections based on annual precipitation. In New Mexico and the North American southwest, plant growth and reproduction are mostly seasonal from spring to fall but depend heavily upon monsoon rains that deliver much of the annual moisture (Notaro et al., 2012; Theobald et al., 2013; Thomas et al., 2023). Given the sensitivity of flowering onset to precipitation for some taxa (Crimmins et al., 2013; Zachmann et al., 2021), and the fact that plant collections are most often made during flowering (Lavoie, 2013), we predicted that wetter years would be better represented in herbarium records. We did find that those years with above-normal precipitation had more collections compared with dry years (Figure 3C), but the difference was not statistically significant. We conclude that botanists are not meaningfully persuaded against making field collecting trips in drier years or based on monsoonal seasonality.

While we cannot expect to retroactively address temporal unevenness across existing collections, synthesized trends provide useful information that may compel action today. First and foremost, the current downtrend in collecting requires funding, resources, and botanists to be actively trained and collecting (Davis, 2023). Maintaining collection momentum and consistency through time is critical for monitoring shifts in biodiversity and predicting how landscaped will change in the future. Second, periods of time with ample collections—such as the early aughts—may be useful for targeted spatiotemporal research questions that focus on phenological shifts, non-native species dynamics, or local adaptation. Finally, climatic shifts (e.g., failing seasonal monsoons) may alter the representation of taxa that are collected in the future. We encourage consideration of past and projected climate conditions in structuring targeted collections efforts going forward.

## Taxonomic and phylogenetic trends

Taxonomic trends in botanical collections made from New Mexico reflect the dominance of a few diverse and abundant families and also identify some groups that

would benefit from targeted efforts to improve collection coverage. As would be expected based on similar work in natural history collections (Deng et al., 2015), we found that there was a positive relationship between the number of collections made and the number of unique species recorded. In short, detecting taxonomic diversity across New Mexico depends on sufficient state-wide collecting. Furthermore, when we categorically assessed the impact of different environments on this relationship, we consistently found positive correlative relationships between the number of species and the number of records made from each environment. However, the number of species detected across biotic and abiotic categories increased with a logarithmic growth curve (Appendix S1: Figure S6), which we interpret as evidence that there is a point at which increased sampling effort does not necessarily increase the taxonomic diversity of collections (Bevilacqua et al., 2018). Species accumulation curves varied across categories and we note that some areas are still characterized by fully linear collecting rates with no evidence of saturation.

We also recognize that the taxonomic distribution of botanical collections is dependent upon the goals of specific collectors or collection projects (Daru et al., 2018; Meyer et al., 2016). Understanding when and how specific collector goals have impacted the diversity of collections made is not something that we address in detail here, but we point to known sources of observer bias (e.g., project goals, proximity to roads, and trails; see Appendix S1: Figure S2) as probable sources of bias that may impact which taxa are collected (Daru et al., 2018; Hortal et al., 2008; Loiselle et al., 2007). We confirm that sampling impacts taxonomic coverage and thus that uneven sampling (particularly in specific environmental categories) may bias taxonomic insights built from existing herbarium data.

An immediate taxonomic use of code resources provided here is automating laborious curatorial tasks for updating, revising, and improving digitized biodiversity data across natural history collections. Through harmonization and comparison with complementary resources, stakeholders and researchers may create new taxonomic checklists across habitats and taxa that may be updated regularly and efficiently with an automated process that is optimized through manual curation. Furthermore, synthesis of taxonomic data is an important step for identifying focal groups that are suitable for specific research questions. Well-collected groups that are evenly represented across space and time are good candidates for temporal research projects while singly collected genera and species warrant curatorial attention and future collections. Additionally, our results point to opportunities for curatorial inquiry to better understand bias in

rare versus common species, and how each is reflected in collections.

## Spatial trends

Clustering versus over-dispersal of collections made across New Mexico suggests that there are spatially explicit considerations that impact collection trends, which we explore by analyzing records within and between discrete environments. There are multiple environmental variables that likely structure the spatial patterns in collections, and mechanisms could be explored with experiments and field work but are beyond the scope of this paper. One example that we specifically point to as an example of covariance in our spatial data is the relatively high degree of clustering in alpine biotic communities that fall into our “extremely high” elevational zone, but which represent low overall number of species and records. It is possible that low taxonomic diversity of collections represents a true diversity decline in higher elevation environments that is in line with the elevational diversity gradient hypothesis (Sanders & Rahbek, 2012), or that sampling effort has been insufficient at these highest, most inaccessible peaks. Patterns of spatial clustering in these small areas would, if considered alone, suggest that these are not fruitful areas to target for further collecting. With further analyses and additional ecological and environmental context, however, we show that there is more to be learned about the diversity of plants that grow in New Mexico’s highest, rarest alpine environments. Indeed, phylogenetic analysis of collections made in these same alpine environments demonstrate relatively high phylogenetic clustering. Targeted collecting of alpine plant communities can improve overall taxonomic coverage and determine what role environmental filtering, evolutionary history, or ecological interactions have had in assembling alpine plant communities across New Mexico.

PD trends reflect species richness trends, and therefore, these two metrics of diversity are nonindependent processes in our workflow, but considering relatedness of collected species provides an evolutionary lens through which to consider spatial distributions and botanical diversity. Across environments, plant communities in New Mexico are phylogenetically clustered (SES PD), but variation in MPD (SES MPD) compared with consistently clustered mean nearest taxon distance (SES MNTD) points to a nested pattern of turnover in early diverging lineages across space and from within these clades coexistence of closely related lineages. Identifying mechanisms driving these spatial phylogenetic patterns are beyond the scope of this paper, but some of our results (e.g., PD

of alpine communities compared with grassland communities) could be fruitfully leveraged to test the lasting influence of historical processes (e.g., glaciation) that impact community assembly across contemporary plant communities (Blonder et al., 2015; Douda et al., 2018; He et al., 2023). Spatial phylogenetics is a relatively newly defined and growing subdiscipline thanks to development in data and computing resources (Mishler, 2023; Mishler et al., 2020), and our data demonstrate the importance of synthesizing collections data to understand evolutionary processes across spatial scales.

In more recent evolutionary time, plant communities across New Mexico have been impacted by disturbance (Boggie et al., 2018; Haire et al., 2017) and development (Pierre et al., 2020), with variation in the strength of these impacts depending on land use and management types. While taxonomic and phylogenetic patterns across land ownership boundaries do not reflect underlying eco-evolutionary processes that govern community assembly, sociopolitical processes have consequences for detection, monitoring, and protection of New Mexican plants. In our analysis, we found that private land ownership is the second most often collected ownership category yet largest by area but is still comparatively under-represented in collections by area (Figure 7C). Even so, taxonomic diversity is highest for collections made from private land compared with all other ownership categories, emphasizing how important community science and diverse stakeholder involvement is in biodiversity research. Similarly, public lands managed by the United States Forest Service are the best-collected ownership category in New Mexico by total number of records (Figure 7B). Access and permission to collect specimens from public lands has been critical for biodiversity research in the context of postcolonial scientific efforts and will undoubtedly remain important as environments change with climate and development.

Further investing in public lands and maintaining multi-institutional research networks is, based on our results, especially important for protecting plant communities given the comparatively low number of records collected from areas with legal protections. Forest Service and Bureau of Land Management ownership types have been historically amenable to collecting and have therefore contributed significantly to biodiversity information about the region. In addition to improving collection coverage in biotic and abiotic areas that are relatively less-well represented in existing collections, we also encourage consideration of habitats for which conservation protections may be justified (Appendix S1: Figure S8). Currently, 3% (7479 records) of all botanical records from New Mexico were collected from areas that are classified as protected (Appendix S1: Figure S10)

and only 1.5% of the state falls within the boundaries of a protection area (wilderness areas, study areas, areas of environmental concern, and conservation areas) (Appendix S1: Figure S8). While development of public lands is possible, it is less likely or common than on privately owned land, which is rapidly being converted for commercial use across the North American west (Pierre et al., 2020). We also encourage co-development of collections projects with protected and private land stakeholders and encourage stakeholders and policymakers to consider if and where additional protections are needed to preserve New Mexico's botanical diversity.

Results from each separate environmental land category can also be interpreted in the context of other categorical results, and we found agreement across and between broad biotic, abiotic, and sociopolitical trends. In general, some plant communities were better represented than others, which may be due, in part, to permitting requirements (Figure 7) or physical accessibility of certain areas (Appendix S1: Figure S2; Daru et al., 2018). Similarly, trends indicated that biotic community representation was reflected across elevational zones. Integrative perspectives from across environmental layers specifically identify low collection coverage from certain grassland and desert areas, which provides a base from which future research efforts can build targeted outcomes.

As with many things of an ecological nature, context matters, and we urge consideration of multivariate environments in interpreting our results and targeting new efforts. Clearly, multiple interacting environmental processes qualify our ability to identify the “most” important variables that structured historical collecting across the state, but each layer provides a specific lens through which we might interpret botanical trends. Synthesizing trends in historical data and placing results into relevant environmental contexts is an outcome of this work that we hope will support efficient and effective research going forward, particularly given the complexity of how built and natural environments are changing in New Mexico and beyond.

Practical uses of spatial results include informed and targeted new collections from low coverage environments, phylogenetically informed species distribution monitoring and streamlining curatorial processes that are spatially explicit. We hope that providing data and scripts encourages future analyses that will also improve products built from these data (e.g., maps, field guides, and models).

## Challenges and limitations

Large synthesis projects using biodiversity records from across institutions are often plagued by messy data and

should be analyzed with consideration for specific project goals and requirements (Boakes et al., 2010; Daru et al., 2018; Isaac et al., 2014). In this study, our aim was inclusive, and we retained as many records as possible while minimizing sources of error using automated and manual workflows to identify errors in temporal, taxonomic, and spatial data.

With erroneous temporal information, there are limits to how much improvement can be made. Obvious digitization errors (e.g., year data indicating a collection date that is in the future) can be managed automatically through culling or correction (Appendix S1: Figure S1). Less obvious errors, such as unlikely seasonality, require close inspection and, in some instances, may not be resolvable or will only be addressed through manual curation. This is particularly true for older collections, which may lack specific dates, clear year data, or may have the wrong label ascribed to the specimen. In our study, for example, the earliest valid digitized botanical collections from within the state boundaries of contemporary New Mexico were made in the early 1840s by William Gambel. However, multiple records in our full dataset predate these confirmed records but do not provide full temporal information. Errors in dating 19th century collections make it difficult to know for sure when the first collection was made, and New Mexico had different borders prior to statehood in 1912 (Allred & Jercinovic, 2020a). Through manual curation, we confirmed that annual dates for pre-1841 specimens were mis-entered on their respective databases, and we chose to exclude them from temporal analyses.

Taxonomic revision of natural history collections is both critical and difficult to do when handling large and highly variable datasets, and there are many emerging tool sets that practitioners may find useful (Grenié et al., 2023). Specific aims and goals should be considered when researchers select specific resources and workflows, but here we describe some taxonomic issues that are more consequential than others when analyzing large and noisy biodiversity data. We found that overall taxonomic trends above the family level do not shift dramatically pre- versus post-harmonization, but that accurate identification of genus- and species- level trends requires taxonomic standardization prior to thesaurus-informed harmonization. Comparison of our results with expert-compiled checklists indicates that we describe taxonomic trends to the species and sometimes infra-species levels but detect far some species not previously included on curated botanical checklists for New Mexico (Allred & Jercinovic, 2020b). This is largely expected as our approach to characterizing species richness was inclusive of non-native horticultural collections, and no automated taxonomic harmonization can catch all issues that come with digital collections data.

Reasons for mismatching taxonomic lists are numerous and include mis-identified specimens, incorrectly entered data, evolving nomenclature, inclusion of non-native crop and ornamental species and novel species occurrences.

While taxonomic curation and cleaning improved the quality of our data, it was not possible to eliminate idiosyncratic errors completely, which is particularly consequential when considering those taxa with few or single records. Multiple records of singly collected taxa (Figure 4; Appendix S1: Figure S4) are actually members of better represented species but have errors in identification, harmonization, or digitization. While TNRS and most other tools are able to correct minor misspellings and identify synonyms, not every issue was captured through automated cleaning and we encourage involving taxonomists and other experts in verifying taxonomic results, particularly in targeted research projects. Using taxonomic trend analysis to generate a list of singly collected records, as we did (Appendix S1: Figure S4), is imperative for identifying truly novel occurrences as well as problematic specimens that require revision or annotation. This approach was possible thanks to the availability of a state-wide expert-compiled flora (Allred & Jercinovic, 2020a), which we used as a resource against which we compared taxonomic results.

For more targeted taxonomic or habitat studies, researchers should explore existing regional resources, such as checklists or taxon-specific thesauruses, to increase specificity and accuracy of automated taxonomic cleaning (Grenié et al., 2023; König et al., 2019). A recent example of region-specific collections analysis in New Mexico was completed for the highly biodiverse Organ Mountains of New Mexico, which demonstrated more local application of data synthesis and trend analysis (Mancillas, 2023). For the purposes of our state-wide analysis, it was enough to clean with automated functions for most common errors and then spot-check through manual inspection of cleaned records. This latter step, though laborious and time-intensive, improved the quality of records included in analyses. While validation of every record was not reasonable for such a large dataset, we still integrated expert knowledge and cleaned for errors through iterative revision and trend analyses using regional expert-compiled botanical resources.

Many issues we encountered in cleaning georeferenced records can be attributed to coordinate incongruity, which limits the spatial precision of results and, unfortunately, cannot be entirely improved through post hoc cleaning. The history of spatial accuracy in natural history records has varied with shifts from narrative to technological geolocation methods, and as high-precision GPS units have become more accessible. While some records, particularly those accessed through GBIF, are

assigned measurements of spatial accuracy, too few have these data to reasonably use as a standard cleaning filter, except to a fairly coarse degree. Nonetheless, we eliminated records with common spatial problems and considered the coarse resolution of georeferenced records to be appropriate given the similarly coarse vector inputs for environmental categories used to interpret broad trends. Again, we encourage further work to verify and improve spatial results and acknowledge that microhabitat variation structures plant communities at scales far smaller than the 4-km resolution used here (Angert et al., 2011; Kephart & Paladino, 1997; Massatti & Lacey Knowles, 2014; Niittynen et al., 2020).

Finally, it is well-documented that the number of collections that have been digitized is far below the number of specimens that exist (Ariño, 2010; Sweeney et al., 2018). Furthermore, of those specimens that have been digitized, not all have been georeferenced. In our study, only 64% of all cleaned records had coordinate data sufficient for analyses. We expect that some of our results may shift or change as more specimens are digitized and georeferenced, and we built code and data resources with the explicit hope that reanalyzing these trends can be done quickly and efficiently in the future. We look forward to revisiting patterns in collections from New Mexico in the future as more records are collected, processed, digitized, and georeferenced.

## Conclusions

The future of herbaria depends, to some extent, on justifying improved resources and research support toward their protection, use and growth. Here we show that there is ongoing need for new collections in order to build biodiversity baselines in low coverage areas and to more accurately monitor the impacts of environmental change across taxa. We also model ways to further utilize biodiversity specimen data that already exists in order to generate new insights with practical and research implications.

Access to resources and training, as well as physical access to under-collected areas, is a common hurdle facing botanists and collectors. We encourage those training the next generation of biodiversity collectors to consider optimizing existing collections, especially when research interests intersect with the history of colonization of resources and permissions across varying land ownerships (Park et al., 2023; Vargas et al., 2024). We also encourage those training the next generation to take a proactive role in mitigating barriers to collecting plant specimens, from baseline exposure to the field, to confronting ableism and unsafe field conditions (Demery & Pipkin, 2021; Kottler, n.d.). Engaging head-on in the hurdles that make

collections-based research inaccessible is necessary to ensure that future biodiversity data, research and conservation approaches can continue.

Fostering trust and building capacity with diverse stakeholders and land managers is a long-term effort that is crucial to sustaining equitable and collaborative collections-based research, and trust is built on honesty. We acknowledge the history of oppression and land dispossession across New Mexico and the reality that many contemporary collections-based projects perpetuate inequality in research. Fortunately, we also show that multi-stakeholder capacity building is a critical step for improving collections research going forward and suggest that those interested in readdressing indigenous dispossession in their collections work with the organization Local Contexts (<https://localcontexts.org>; “Grounding Indigenous Rights,” n.d.) to properly label specimens and extend access to collections research beyond traditional institutional boundaries. Large synthesis projects like the present study are an important first step toward leveraging data to produce meaningful knowledge, but channeling results into conservation actions requires advocacy, inclusive community collaboration, and motivated stakeholders. We hope this study provides useful resources in supporting urgent botanical research, management, and protection practices in New Mexico and beyond.

## AUTHOR CONTRIBUTIONS

Elizabeth M. Lombardi, Hannah E. Marx, and Harpo Faust jointly conceived of and designed the study. Elizabeth M. Lombardi and Harpo Faust performed data management and cleaning. All authors analyzed the data and performed statistical analyses. All authors interpreted results. Elizabeth M. Lombardi led the manuscript writing. Harpo Faust and Hannah E. Marx contributed.

## ACKNOWLEDGMENTS

We thank the University of New Mexico Herbarium and the Museum of Southwestern Biology community, past and present. New Mexico botanists collected and curated the data used for this publication, and we specifically acknowledge the important contributions of Bob Sivinski, Kelly Allred, and Cecilia Alexander for their roles in building New Mexico’s botanical capacity for analyses like ours. The R community continues to provide exceptional computational resources and support, and we thank all those who have contributed to the development of Global Biodiversity Information Facility and SEINet data infrastructure.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Data (Lombardi, 2024) are available from Zenodo: <https://doi.org/10.5281/zenodo.10778245>.

## ORCID

Elizabeth M. Lombardi  <https://orcid.org/0000-0001-8256-5037>

## REFERENCES

- Ackerfield, J. R. 2022. “Hiding in Plain Sight: Two New Species of Alpine Thistles, *Cirsium culebraensis* and *C. funkiae* (Cardueae) from the Southern Rocky Mountains (United States).” *The International Compositae Alliance* 1: 1–15. <https://doi.org/10.53875/capitulum.01.2.01>.
- Allred, K. W., and E. M. Jercinovic. 2020b. *Flora Neomexicana*, 2nd ed., Vol. II–III. Self published.
- Allred, K. W., and E. M. Jercinovic. 2020a. “Flora Neomexicana: Annotated Checklist. An Introduction to the Vascular Plants of New Mexico, with Synonymy and Bibliography.” <https://play.google.com/store/books/details?id=LC6IzgEACAAJ>.
- Allred, K. W. 2021a. Post Publication Updates. “FNM III part 1 updates, 22 Jan 2021.” <https://floraneomexicana.org/wp-content/uploads/2021/01/fnm-iii-part-1-updates.pdf>.
- Allred, K. W. 2021b. Post Publication Updates. “FNM III part 2 updates, 23 July 2021.” <https://floraneomexicana.org/wp-content/uploads/2021/07/fnm-iii-part-2-updates.pdf>.
- Angert, A. L., S. N. Sheth, and J. R. Paul. 2011. “Incorporating Population-Level Variation in Thermal Performance into Predictions of Geographic Range Shifts.” *Integrative and Comparative Biology* 51(5): 733–750. <https://doi.org/10.1093/icb/icr048>.
- Ariño, A. H. 2010. “Approaches to Estimating the Universe of Natural History Collections Data.” *Biodiversity Informatics* 7 (2): 81–92. <https://doi.org/10.17161/bi.v7i2.3991>.
- Baddeley, A. 2008. “Analysing Spatial Point Patterns in R.” 171 pp. <https://bjpcjp.github.io/pdfs/r/spatial-pattern-analysis-in-r.pdf>.
- Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67(1): 611–636. <https://doi.org/10.18637/jss.v067.i01>.
- Bevilacqua, S., K. I. Ugland, A. Plicanti, D. Scuderi, and A. Terlizzi. 2018. “An Approach Based on the Total-Species Accumulation Curve and Higher Taxon Richness to Estimate Realistic Upper Limits in Regional Species Richness.” *Ecology and Evolution* 8(1): 405–415. <https://doi.org/10.1002/ece3.3570>.
- Blonder, B., D. Nogués-Bravo, M. K. Borregaard, J. C. Donoghue 2nd, P. M. Jørgensen, N. J. B. Kraft, J.-P. Lessard, et al. 2015. “Linking Environmental Filtering and Disequilibrium to Biogeography with a Community Climate Framework.” *Ecology* 96(4): 972–985. <https://doi.org/10.1890/14-0589.1>.
- Boakes, E. H., P. J. K. McGowan, R. A. Fuller, D. Chang-qing, N. E. Clark, K. O’Connor, and G. M. Mace. 2010. “Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data.” *PLoS Biology* 8(6): e1000385. <https://doi.org/10.1371/journal.pbio.1000385>.
- Boggie, M. A., D. P. Collins, J. Patrick Donnelly, and S. A. Carleton. 2018. “Land Use, Anthropogenic Disturbance, and Riverine Features Drive Patterns of Habitat Selection by a Wintering

- Waterbird in a Semi-Arid Environment." *PLoS One* 13(11): e0206222. <https://doi.org/10.1371/journal.pone.0206222>.
- Cook, B. I., and R. Seager. 2013. "The Response of the North American Monsoon to Increased Greenhouse Gas Forcing." *Journal of Geophysical Research* 118(4): 1690–99. <https://doi.org/10.1002/jgrd.50111>.
- Cornwell, W. K., W. D. Pearse, R. L. Dalrymple, and A. E. Zanne. 2019. "What We (Don't) Know about Global Plant Diversity." *Ecography* 42(11): 1819–31. <https://doi.org/10.1111/ecog.04481>.
- Crawford, P. H. C., and B. W. Hoagland. 2009. "Can Herbarium Records Be Used to Map Alien Species Invasion and Native Species Expansion over the Past 100 Years?" *Journal of Biogeography* 36(4): 651–661. <https://doi.org/10.1111/j.1365-2699.2008.02043.x>.
- Crimmins, T. M., M. A. Crimmins, and C. David Bertelsen. 2013. "Spring and Summer Patterns in Flowering Onset, Duration, and Constancy across a Water-Limited Gradient." *American Journal of Botany* 100(6): 1137–47. <https://doi.org/10.3732/ajb.1200633>.
- D'Antraccoli, M., G. Bedini, and L. Peruzzi. 2022. "Next Generation Floristics: A Workflow to Integrate Novel Methods in Traditional Floristic Research." *Plant Biosystems – An International Journal Dealing with all Aspects of Plant Biology* 156(2): 594–97. <https://doi.org/10.1080/11263504.2022.2056650>.
- Daru, B. H., D. S. Park, R. B. Primack, C. G. Willis, D. S. Barrington, T. J. S. Whitfeld, T. G. Seidler, et al. 2018. "Widespread Sampling Biases in Herbaria Revealed from Large-Scale Digitization." *The New Phytologist* 217(2): 939–955. <https://doi.org/10.1111/nph.14855>.
- Davis, C. C. 2023. "The Herbarium of the Future." *Trends in Ecology & Evolution* 38(5): 412–423. <https://doi.org/10.1016/j.tree.2022.11.015>.
- Demery, A.-J. C., and M. A. Pipkin. 2021. "Safe Fieldwork Strategies for At-Risk Individuals, Their Supervisors and Institutions." *Nature Ecology & Evolution* 5(1): 5–9. <https://doi.org/10.1038/s41559-020-01328-5>.
- Deng, C., T. Daley, and A. D. Smith. 2015. "Applications of Species Accumulation Curves in Large-Scale Biological Data Analysis." *Quantitative Biology (Beijing, China)* 3(3): 135–144. <https://doi.org/10.1007/s40484-015-0049-7>.
- Dickerman, C. 1985. "Mid-Nineteenth-Century Botanical Exploration in New Mexico." *New Mexico Historical Review* 60(2): 3 <https://digitalrepository.unm.edu/nmhr/vol60/iss2/3>.
- Diggle, P. J., V. Gomez-Rubio, P. E. Brown, A. G. Chetwynd, and S. Gooding. 2007. "Second-Order Analysis of Inhomogeneous Spatial Point Processes Using Case-Control Data." *Biometrics* 63(2): 550–57. <https://doi.org/10.1111/j.1541-0420.2006.00683.x>.
- Douda, J., A. Havrdová, J. Doudová, and B. Mandák. 2018. "Legacy of Post-Glacial Colonization Affects  $\beta$ -Diversity: Insights into Local Community Assembly Processes." *Journal of Biogeography* 45(7): 1604–15. <https://doi.org/10.1111/jbi.13235>.
- Edwards, E. J., B. D. Mishler, and C. D. Davis. 2024. "University Herbaria Are Uniquely Important." *Trends in Plant Science* 29(8): 825–826. <https://doi.org/10.1016/j.tplants.2024.06.002>.
- Faust, H. 2024. "Shorter Note." *American Fern Journal* 114(4): 332–33. <https://doi.org/10.1640/0002-8444-114.4.332>.
- Fowler, J. F., B. E. Nelson, and R. L. Hartman. 2014. "Vascular Plant Flora of the Alpine Zone in the Southern Rocky Mountains, USA." *Journal of the Botanical Research Institute of Texas* 8(2): 611–636. <https://www.jstor.org/stable/26549415>.
- Funk, V. A. 2003. "100 Uses for an Herbarium: Well at Least 72." *American Society of Plant Taxonomists Newsletter*, 2003.
- Funk, V. A. 2018. "Collections-Based Science in the 21st Century." *Journal of Systematics and Evolution* 56(3): 175–193. <https://doi.org/10.1111/jse.12315>.
- Grenié, M., E. Berti, J. Carvajal-Quintero, G. M. L. Dädlow, A. Sagouis, and M. Winter. 2023. "Harmonizing Taxon Names in Biodiversity Data: A Review of Tools, Databases and Best Practices." *Methods in Ecology and Evolution/British Ecological Society* 14(1): 12–25. <https://doi.org/10.1111/2041-210X.13802>.
- Grolemund, G., and H. Wickham. 2011. "Dates and Times Made Easy with Lubridate." *Journal of Statistical Software* 40: 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Haire, S. L., J. D. Coop, and C. Miller. 2017. "Characterizing Spatial Neighborhoods of Refugia Following Large Fires in Northern New Mexico USA." *Land* 6(1): 19. <https://doi.org/10.3390/land6010019>.
- He, X., M. Arif, J. Zheng, X. Ni, Z. Yuan, Q. Zhu, J. Wang, D. Ding, and C. Li. 2023. "Plant Diversity Patterns along an Elevation Gradient: The Relative Impact of Environmental and Spatial Variation on Plant Diversity and Assembly in Arid and Semi-Arid Regions." *Frontiers of Environmental Science & Engineering in China* 11: 1021157. <https://doi.org/10.3389/fenvs.2023.1021157>.
- Heberling, J. M., and B. L. Isaac. 2017. "Herbarium Specimens as Exaptations: New Uses for Old Collections." *American Journal of Botany* 104(7): 963–65. <https://doi.org/10.3732/ajb.1700125>.
- Heberling, J. M., L. A. Prather, and S. J. Tonsor. 2019. "The Changing Uses of Herbarium Data in an Era of Global Change: An Overview Using Automated Content Analysis." *BioScience* 69(10): 812–822. <https://doi.org/10.1093/biosci/biz094>.
- Hernandez, M., and L. Chen. 2022. "Future Land Precipitation Changes over the North American Monsoon Region Using CMIP5 and CMIP6 Simulations." *Journal of Geophysical Research* 127(9): e2021JD035911. <https://doi.org/10.1029/2021jd035911>.
- Higgins, R. W., Y. Chen, and A. V. Douglas. 1999. "Interannual Variability of the North American Warm Season Precipitation Regime." *Journal of Climate* 12(3): 653–680. [https://doi.org/10.1175/1520-0442\(1999\)012<0653:IVOTNA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<0653:IVOTNA>2.0.CO;2).
- Hortal, J., A. Jiménez-Valverde, J. F. Gómez, J. M. Lobo, and A. Baselga. 2008. "Historical Bias in Biodiversity Inventories Affects the Observed Environmental Niche of the Species." *Oikos* 117(6): 847–858. <https://doi.org/10.1111/j.0030-1299.2008.16434.x>.
- Isaac, N. J. B., A. J. van Strien, and T. A. August. 2014. "Statistics for Citizen Science: Extracting Signals of Change from Noisy Ecological Data." *Methods in Ecology and Evolution/British Ecological Society* 5: 1052–60. <https://doi.org/10.1111/2041-210X.12254>.
- Jin, Y., and H. Qian. 2023. "U.PhyloMaker: An R Package that Can Generate Large Phylogenetic Trees for Plants and Animals." *Plant Diversity* 45(3): 347–352. <https://doi.org/10.1016/j.pld.2022.12.007>.

- Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. "Picante: R Tools for Integrating Phylogenies and Ecology." *Bioinformatics* 26(11): 1463–64. <https://doi.org/10.1093/bioinformatics/btq166>.
- Kephart, S., and C. Paladino. 1997. "Demographic Change and Microhabitat Variability in a Grassland Endemic, *Silene douglasii* var. *oraria* (Caryophyllaceae)." *American Journal of Botany* 84(2): 179. <https://doi.org/10.2307/2446079>.
- König, C., P. Weigelt, J. Schrader, A. Taylor, J. Kattge, and H. Kreft. 2019. "Biodiversity Data Integration—the Significance of Data Resolution and Domain." *PLoS Biology* 17(3): e3000183. <https://doi.org/10.1371/journal.pbio.3000183>.
- Kottler, E. n.d. "TGNC Fieldwork Alliance." <https://ezrakottler.wixsite.com/field>.
- Lang, P. L. M., F. M. Willems, J. F. Scheepens, H. A. Burbano, and O. Bossdorf. 2019. "Using Herbaria to Study Global Environmental Change." *The New Phytologist* 221(1): 110–122. <https://doi.org/10.1111/nph.15401>.
- Larson, J., B. Reif, B. E. Nelson, and R. L. Hartman. 2014. "Floristic Studies in North Central New Mexico, U.S.A. the Sangre De Cristo Mountains." *Journal of the Botanical Research Institute of Texas* 8(1): 271–303. <https://www.jstor.org/stable/26549368>.
- Larson, J. E. 2008. *A Floristic Inventory of Vascular Plants of the Carson National Forest and Vicinity, North-Central New Mexico*. Edited by R. L. Hartman. Ann Arbor, MI: University of Wyoming.
- Lavoie, C. 2013. "Biological Collections in an Ever Changing World: Herbaria as Tools for Biogeographical and Environmental Studies." *Perspectives in Plant Ecology, Evolution and Systematics* 15(1): 68–76. <https://doi.org/10.1016/j.ppees.2012.10.002>.
- Legler, B. S. 2010. "Additions to the Vascular Flora of New Mexico." *Journal of the Botanical Research Institute of Texas* 4(2): 777–784. [http://www.jstor.org/stable/41972110](https://www.jstor.org/stable/41972110).
- Lendemer, J., B. Thiers, A. K. Monfils, J. Zaspel, E. R. Ellwood, A. Bentley, K. LeVan, et al. 2020. "Corrigendum: The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education." *BioScience* 70(2): 195. <https://doi.org/10.1093/biosci/biz165>.
- Lipscomb, B. L., J. J. Pipoly, and R. W. Sanders. 2000. *Floristics in the New Millennium: Proceedings of the Flora of the Southeast US Symposium*. Fort Worth, TX: BRIT Press. <https://play.google.com/store/books/details?id=GUWx4g0Sii0C>.
- Local Contexts. n.d. "Grounding Indigenous Rights." <https://localcontexts.org/>.
- Loiselle, B. A., P. M. Jørgensen, T. Consiglio, I. Jiménez, J. G. Blake, L. G. Lohmann, and O. M. Montiel. 2007. "Predicting Species Distributions from Herbarium Collections: Does Climate Bias in Collection Sampling Influence Model Outcomes?" *Journal of Biogeography* 35: 105–116. <https://doi.org/10.1111/j.1365-2699.2007.01779.x>.
- Lombardi, E. 2024. "EMLgit/Herbarium-Patterns-and-Gaps: New Mexico State-Wide Preserved Plant Data." Zenodo. <https://doi.org/10.5281/ZENODO.13936206>.
- Maitner, B., and B. Boyle. 2023. "TNRS: Taxonomic Name Resolution Service." <https://CRAN.R-project.org/package=TNRS>.
- Mancillas, M. E. 2023. *Plant Biodiversity Data Synthesis and Species-Landscape Relationships through Time (1849–2021) in the Organ Mountains, New Mexico*. Edited by S. Fuentes-Soriano and L. Prihodko. Ann Arbor, MI: New Mexico State University.
- Massatti, R., and L. Lacey Knowles. 2014. "Microhabitat Differences Impact Phylogeographic Concordance of Codistributed Species: Genomic Evidence in Montane Sedges (*Carex* L.) from the Rocky Mountains." *Evolution; International Journal of Organic Evolution* 68(10): 2833–46. <https://doi.org/10.1111/evo.12491>.
- Massicotte, P., and A. South. 2023. "Rnaturalearth: World Map Data from Natural Earth." <https://CRAN.R-project.org/package=rnaturalearth>.
- McCauley, L. 2021. "Biotic Communities of the Southwest." <https://tnc.maps.arcgis.com/home/item.html?id=d4a6ae739df745b2ae faa1386e759c02>.
- Meyer, C., P. Weigelt, and H. Kreft. 2016. "Multidimensional Biases, Gaps and Uncertainties in Global Plant Occurrence Information." *Ecology Letters* 19(8): 992–1006. <https://doi.org/10.1111/ele.12624>.
- Mishler, B. D. 2023. "Spatial Phylogenetics." *Journal of Biogeography* 50(8): 1454–63. <https://doi.org/10.1111/jbi.14618>.
- Mishler, B. D., R. Guralnick, P. S. Soltis, S. A. Smith, D. E. Soltis, N. Barve, J. M. Allen, and S. W. Laffan. 2020. "Spatial Phylogenetics of the North American Flora." *Journal of Systematics and Evolution* 58(4): 393–405. <https://doi.org/10.1111/jse.12590>.
- Niittynen, P., R. K. Heikkinen, J. Aalto, A. Guisan, J. Kemppinen, and M. Luoto. 2020. "Fine-Scale Tundra Vegetation Patterns Are Strongly Related to Winter Thermal Conditions." *Nature Climate Change* 10(12): 1143–48. <https://doi.org/10.1038/s41558-020-00916-4>.
- NOAA. 2023. "National Centers for Environmental Information, Statewide Time Series." <https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/statewide/time-series>.
- Notaro, M., A. Mauss, and J. W. Williams. 2012. "Projected Vegetation Changes for the American Southwest: Combined Dynamic Modeling and Bioclimatic-Envelope Approach." *Ecological Applications: A Publication of the Ecological Society of America* 22(4): 1365–88. <https://doi.org/10.1890/11-1269.1>.
- Park, D. S., X. Feng, S. Akiyama, M. Ardiyani, N. Avendaño, Z. Barina, B. Bärtschi, et al. 2023. "The Colonial Legacy of Herbaria." *Nature Human Behaviour* 7(7): 1059–68. <https://doi.org/10.1038/s41562-023-01616-7>.
- Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker, and M. R. Helmus. 2015. "Pez: Phylogenetics for the Environmental Sciences." *Bioinformatics* 31(17): 2888–90. <https://doi.org/10.1093/bioinformatics/btv277>.
- Pebesma, E. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10: 439. <https://doi.org/10.32614/RJ-2018-009>.
- Pierre, J. P., J. R. Andrews, M. H. Young, A. Y. Sun, and B. D. Wolaver. 2020. "Projected Landscape Impacts from Oil and Gas Development Scenarios in the Permian Basin, USA." *Environmental Management* 66(3): 348–363. <https://doi.org/10.1007/s00267-020-01308-2>.
- Prather, L. A., O. Alvarez-Fuentes, M. H. Mayfield, and C. J. Ferguson. 2004. "The Decline of Plant Collecting in the United States: A Threat to the Infrastructure of Biodiversity Studies." *Systematic Botany* 29(1): 15–28. <https://doi.org/10.1600/036364404772974185>.

- Rogers, Z. S., S. Fuentes-Soriano, and R. Spellenberg. 2023. "Noteworthy New Angiosperm Plant Distribution Records for New Mexico and Colorado, U.S.A." *Journal of the Botanical Research Institute of Texas* 17(1): 323–337. <https://doi.org/10.17348/jbrit.v17.i1.1299>.
- Sanders, N. J., and C. Rahbek. 2012. "The Patterns and Causes of Elevational Diversity Gradients." *Ecography* 35(1): 1–3. <https://doi.org/10.1111/j.1600-0587.2011.07338.x>.
- Shannon, C. E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27(3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Sheppard, P. R., A. C. Comrie, G. D. Packin, K. Angersbach, and M. K. Hughes. 2002. "The Climate of the US Southwest." *Climate Research* 21: 219–238. <https://doi.org/10.3354/cr02129>.
- Sivinski, R. C. 2024. "Noteworthy Collections of New Mexico Flora." *The New Mexico Botanist Newsletter*, February. <https://floraneomexicana.org/wp-content/uploads/2024/02/tnmb-76.pdf>.
- Sivinski, R. C. 2023. "Oreocarya worthingtonii (Boraginaceae): A New Species from Southeastern New Mexico, U.S.A." *Journal of the Botanical Research Institute of Texas* 17(1): 1–8. <https://doi.org/10.17348/jbrit.v17.i1.1286>.
- Smith, S. A., and J. W. Brown. 2018. "Constructing a Broadly Inclusive Seed Plant Phylogeny." *American Journal of Botany* 105(3): 302–314. <https://doi.org/10.1002/ajb2.1019>.
- South, A. n.d. "Rnaturalearth: World Map Data from Natural Earth. R Package Version 0.1.0." <https://CRAN.R-Project.Org/package=Rnaturalearth>.
- Spellerberg, I. F., and P. J. Fedor. 2003. "A Tribute to Claude Shannon (1916–2001) and a Plea for More Rigorous Use of Species Richness, Species Diversity and the 'Shannon–Wiener' Index." *Global Ecology and Biogeography: A Journal of Macroecology* 12(3): 177–79. <https://doi.org/10.1046/j.1466-822x.2003.00015.x>.
- Stein, B. A. 2002. *States of the Union: Ranking America's Biodiversity*. Arlington, VA: The Nature Conservancy.
- Sweeney, P. W., B. Starly, P. J. Morris, X. Yiming, A. Jones, S. Radhakrishnan, C. J. Grassia, and C. C. Davis. 2018. "Large-Scale Digitization of Herbarium Specimens: Development and Usage of an Automated, High-Throughput Conveyor System." *Taxon* 67(1): 165–178. <https://doi.org/10.12705/671.10>.
- Taylor, D. W. 2014. "Large Inequalities in Herbarium Specimen Density in the Western United States." *Phyton* 53(June): 1–8.
- Tewksbury, J. J., J. G. T. Anderson, J. D. Bakker, T. J. Billo, P. W. Dunwiddie, M. J. Groom, S. E. Hampton, et al. 2014. "Natural History's Place in Science and Society." *BioScience* 64(4): 300–310. <https://doi.org/10.1093/biosci/biu032>.
- Theobald, D. M., W. R. Travis, M. A. Drummond, E. S. Gordon, and M. Bettsill. 2013. "The Changing Southwest." Assessment of Climate Change in the Southwest United States: A Report Prepared for the National Climate Assessment. <https://www.resolutionmineis.us/sites/default/files/references/garfjardine-merideth-black-leroy-2013.pdf>.
- Thiers, B., T. E. Roberts, R. J. Rundell, G. M. Spellman, G. A. Fischer, G. Nelson, J. Bates, et al. 2024. "Duke's Herbarium Merits Continued Enhancement, Not Dissolution." *BioScience* 74(8): 507–508. <https://doi.org/10.1093/biosci/biae031>.
- Thomas, K. A., B. A. Stauffer, and C. J. Jarchow. 2023. "Decoupling of Species and Plant Communities of the U.S. Southwest: A CCSM4 Climate Scenario Example." *Ecosphere* 14(2): e4414. <https://doi.org/10.1002/ecs2.4414>.
- Troudet, J., P. Grandcolas, A. Blin, R. Vignes-Lebbe, and F. Legendre. 2017. "Taxonomic Bias in Biodiversity Data and Societal Preferences." *Scientific Reports* 7(1): 9132. <https://doi.org/10.1038/s41598-017-09084-6>.
- Tucker, C. M., M. W. Cadotte, S. B. Carvalho, T. Jonathan Davies, S. Ferrier, S. A. Fritz, R. Grenyer, et al. 2017. "A Guide to Phylogenetic Metrics for Conservation, Community Ecology and Macroecology." *Biological Reviews of the Cambridge Philosophical Society* 92(2): 698–715. <https://doi.org/10.1111/brv.12252>.
- Vargas, C. A., M. Bottin, T. Sarkinen, J. E. Richardson, M. Celis, B. Villanueva, and A. Sanchez. 2024. "How to Fill the Biodiversity Data Gap: Is It Better to Invest in Fieldwork or Curation?" *Plant Diversity* 46(1): 39–48. <https://doi.org/10.1016/j.pld.2023.06.003>.
- Vera, C., W. Higgins, J. Amador, T. Ambrizzi, R. Garreaud, D. Gochis, D. Gutzler, et al. 2006. "Toward a Unified View of the American Monsoon Systems." *Journal of Climate* 19(20): 4977–5000. <https://doi.org/10.1175/JCLI3896.1>.
- Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughan. 2023. "Dplyr: A Grammar of Data Manipulation." <https://CRAN.R-project.org/package=dplyr>.
- Wickham, H., D. Vaughan, and M. Girlich. 2023. "Tidyr: Tidy Messy Data." <https://CRAN.R-project.org/package=tidyr>.
- Zachmann, L. J., J. F. Wiens, K. Franklin, S. D. Crausbay, V. A. Landau, and S. M. Munson. 2021. "Dominant Sonoran Desert Plant Species Have Divergent Phenological Responses to Climate Change." *Madrono: A West American Journal of Botany* 68(4): 473–486. <https://doi.org/10.3120/0024-9637-68.4.473>.
- Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, et al. 2015. "Corrigendum: Three Keys to the Radiation of Angiosperms into Freezing Environments." *Nature* 521(7552): 380. <https://doi.org/10.1038/nature14371>.
- Zizka, A., A. Antonelli, and D. Silvestro. 2021. "Sampbias, A Method for Quantifying Geographic Sampling Biases in Species Distribution Data." *Ecography* 44(1): 25–32. <https://doi.org/10.1111/ecog.05102>.

## SUPPORTING INFORMATION

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

**How to cite this article:** Lombardi, Elizabeth M., Harpo Faust, and Hannah E. Marx. 2024. "Synthesizing Historical Plant Collections to Identify Priorities for Future Collection Efforts and Research Applications." *Ecosphere* 15(12): e70102. <https://doi.org/10.1002/ecs2.70102>