

NEARBY, a Computer Program for Phytogeographic Analysis Using Georeferenced Specimen Data

Daryl L. Lafferty & Leslie R. Landrum
Natural History Collections, School of Life Sciences
Arizona State University
Tempe, Arizona 85287-4108, U.S.A.

Author emails: daryl.lafferty@asu.edu; les.landrum@asu.edu

ABSTRACT: We present a computer program, NEARBY, that uses databases of georeferenced specimens to explore plant and lichen distributions and co-occurrences. NEARBY utilizes three SYMBIOTA databases: SEINet (mainly vascular plants of North America), NEOTROPICAL (mainly vascular plants of neotropical and temperate South America), and CNALH (lichens, mainly of the western hemisphere). A Primary species is entered into the program and a geographic area is defined. Parameters are chosen that limit the search for specimens of Secondary species that have been collected near the Primary species. The output consists of: a summary of the input data and how it was modified for the search; a list of the most commonly found Secondary species that occur with the Primary species in the defined area; and additional data and links to images for each species. These data can be manipulated in various ways or copied into another program for analysis. NEARBY includes a map option that allows the user to compare distributions of the Primary and Secondary species. An example of a search is discussed in detail and case studies that illustrate the use of the program are provided. An appendix describing the program function is provided.

INTRODUCTION

The distribution of organisms across landscapes, continents, and the globe have long been an interest of biologists, including Humboldt and Bonpland (1807; English translation 2009), Hooker (1853), and Darwin (1859; reprint 1985). Hooker emphasized how difficult a subject plant geography can be in his introduction to the *Flora of Novae-Zelandiae*:

“Of all the branches of botany there is none whose elucidation demands so much preparatory study, or so extensive an acquaintance with plants and their affinities, as that of the geographic distribution.”

In Hooker’s case, it meant traveling by sailing ship around the southern end of the world exploring southern South America, New Zealand, Tasmania, Australia, and various small and isolated islands. Until recently, discovering where plants grow and with which other species they are found involved much fieldwork and/or many visits to herbaria. It is a slow process and might require a great deal of work to obtain just a few plant distributions that can then be compared. With the recent advances in specimen databasing and georeferencing, preliminary distribution studies for many parts of the world may be made in minutes, thanks to the large databases of herbarium specimens (and museum records for other organisms) that have recently been amassed (e.g., GBIF, CRIA). We here present a computer program, NEARBY

(<https://serv.biokic.asu.edu/dlafferty/Nearby/index.php>), that uses three SYMBIOTA databases and allows one to rapidly discover what species of plants and/or lichens grow near one another and simultaneously compare the distributions of several species.

We previously presented two programs (PROXIMITY and CORRELATION) that allow a user to discover plant associations (vegetation types or continua) based on a list of plants, chosen because they are known to be dominant or conspicuous species in vegetation (Landrum & Lafferty 2015). The resulting associations depend on the list of plants one chooses in the operation of the program. With NEARBY the user starts with a single species (the Primary species) and then searches one or more databases for other species (the Secondary species) that grow near the Primary species using coordinate data (Figs. 1–3).

For this report we are using three databases established using the SYMBIOTA software package (Gries et al. 2014): the **SEINet** database (<http://symbiota.org/docs/seinet/>) of vascular plant herbarium specimens, with ca. 17.9 million specimen records, mainly from North America; the **NEOTROPICAL** database, mainly of vascular plants of neotropical and temperate South America (<https://serv.biokic.asu.edu/neotrop/plantae/index.php>) with ca. 3.6 million records; and the **CNALH** database of the Consortium of North American Lichen Herbaria (<https://lichenportal.org/cnalh/>) with ca. 2.6 million records, of lichens only, mainly of the Western Hemisphere. All databases are actively growing and these numbers are estimates made in August 2020. All use a combination of “live data” (which can be updated immediately by an administrator) and “snapshot data” (a copied database that is infrequently updated). The user should remember that some “snapshot data” may be weeks or months old.

The purpose of NEARBY is to quickly explore phytogeographic data. Fieldwork, including observations, collections, and systematic sampling of vegetation, will normally be a more complete and accurate way of sampling vegetation than databases, but much more time consuming. As a preliminary investigation of plant associations and plant geography we believe NEARBY will be quite valuable. In some cases, such as broad surveys, relying on databases that already exist will be the best strategy. NEARBY has the advantage of being able to use databases for different groups of organisms (e.g., lichens and vascular plants). We hope that insects, vertebrates, fungi, and bryophytes may eventually be added.

We have previously discussed the disadvantages and advantages of using databases of herbarium specimens for phytogeographic/phytosociological studies (Landrum & Lafferty 2015). In summary, the main disadvantages are that databases are never perfect: the specimens may be misidentified, or they may be georeferenced incorrectly or with a low degree of precision. These imperfections must be kept in mind. But we believe the large majority of data to be fairly accurate and thus those data will tend to “overwhelm” inaccurate data. The main advantages are that 1) the data we use already exist for millions of plant and lichen specimens (and are potentially available from other natural history collection databases) and only need to be analyzed—there is no new fieldwork required; 2) herbarium specimens are permanent vouchers of the existence of a species of plant or lichen at a specific place at a particular time, and therefore can be checked if there are doubts about their correct identification or georeferencing; and 3) herbarium specimens have been collected mainly for floristic, taxonomic, and ecological studies. Thus, the data they provide in databases, which in turn are used by NEARBY, are relatively unbiased in a phytogeographic study.

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

USING NEARBY

We will next briefly describe how to use NEARBY by “walking the reader through” a sample case. In the example we have selected *Pinus ponderosa* as the Primary species and the region of interest as Arizona.

Primary Species. The opening screen of NEARBY is where the user sets the parameters of a search. The upper half of the screen is shown in Fig. 1. The first step is to select a Primary species, in this case we will use *Pinus ponderosa*. Occasionally a species may be listed under more than one name in the database. The names *Rhus aromatica* and *R. trilobata* both appear in the SEINet database and we consider them to be synonyms and accept *R. aromatica* as the correct name. So, if it was a Primary species, we would list both names separated by commas, (*Rhus aromatica*,*Rhus trilobata*), with the accepted name first.

Companion Species. If the name of an optional Companion species is entered, then only Primary species specimens with a Companion species close by (defined by the Distance parameter discussed below) will be included. In other words, the original list of Primary species specimens is reduced to those that have a Companion species specimen close by. For instance, one might want to know what plant species grow near the Primary species *Pinus ponderosa* and its frequent associate (or companion) *Pseudotsuga menziesii*. As with Primary species, Companion species can have synonyms, separated by commas. In this example we will not use a Companion species.

Selecting States or Countries. There is a map of the American continents and associated islands. If you want to select particular states of the United States or Mexico, or provinces of Canada there are boxes below the map to check to toggle this option off or on. If these are not checked you will only be able to select these as whole countries. The larger Caribbean islands and countries (e.g., Cuba, Dominican Republic) can be selected. Individual small islands cannot be checked individually but groups of islands (e.g., Lesser Antilles) can be. Using the map, one clicks on the states or countries to toggle them on or off. Searches for Primary, Companion, and Secondary species specimens will be limited to the entities selected. If no states or countries are selected, the entire database will be used. It is possible to select states in one country (e.g., those of the United States bordering Mexico) and a whole country such as Mexico for a search. In this example we will check “Show US States” and select only Arizona.

Search Limits. There are several options that affect the scope of the search. In this example we will accept the default settings but will briefly discuss the options.

Distance: This value indicates how close specimens must be to be considered “nearby”. The default is 100 m, which defines a square 200 m on a side with the Primary at its center. Note that a Companion species or Secondary species specimen is considered “nearby” if it is within this square. Distance is not calculated radially, as that calculation would add significantly to server load. For the purposes of this program, square plots are just as appropriate. If in a search few Secondary species are found, the distance parameter can be increased.

Cull specimens if more than: Points (defined by coordinates) with numerous specimens georeferenced to them are suspect—the specimens have probably been georeferenced as a group, perhaps mechanically by being the middle of a city, state, or country. NEARBY allows the operator to eliminate these suspect specimens by setting a cull number: when the number of specimens georeferenced to a point exceeds this number, the specimens are excluded from the search. The default cull number is 50, but it can be set to other numbers up to 100. The program will never accept locations with more than 100 specimens georeferenced to a single point because that would cause it run too slowly. The specimens that pass this test (apparently not mechanically georeferenced) are then utilized.

Limit: This number sets an upper limit on the number of Primary species specimens included in the search. The default is 1000 and this number is rarely reached in our experience. Limiting to a lower number may produce a faster result in some cases. If the program does not produce a result in three minutes, lowering the Limit downward may help. When this limit is set to a number, for example 500, and the full number of Primary specimens is larger, for example 2000, the full list of Primary specimens is first obtained, in this case 2000, and randomized.

Truncate list of species at: Once all the Secondary species specimens are found for all the Primary species specimens, the data are sorted by % ***Primaries near this Secondary*** (this term defined below). The default is to consider only the most commonly found 50 species and truncate the list at that level. But numbers other than 50 may be used. This is just a way to keep the list reasonably short. It also reduces the calculation time somewhat. It is common for multiple species to have exactly the same percentage value, so increasing the list size to make sure you have all the species up to a certain value may be useful.

Don't allow Primary species specimens near other Primary specimens: If two Primary species specimens have exactly the same coordinates they are always counted just once. In addition, by leaving this default option on, if two Primary species specimens are close enough together to be considered as Secondaries of one another, only one is accepted. When Primary species specimens are few, it may be beneficial to turn off this option. When a genus name is chosen as the Primary we find that it is best to turn this option off, because species of the same genus are often collected together.

Stop after obtaining Primary species data: If this is checked, the program will stop after determining the number of Primary species specimens found in the region, or after determining the number limited by Companion species if it is included. This is often a good preliminary step to quickly check for the number of Primary species specimens in the selected area. When searching with just a Primary species it is probably best to find at least 100 individuals. Searches with 200 to 1000 are better. When searching for a Primary and a Companion species together, lower numbers are acceptable. The best way to increase the numbers is to include a wider search area by adding more states or countries.

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

Nearby Plant Finder

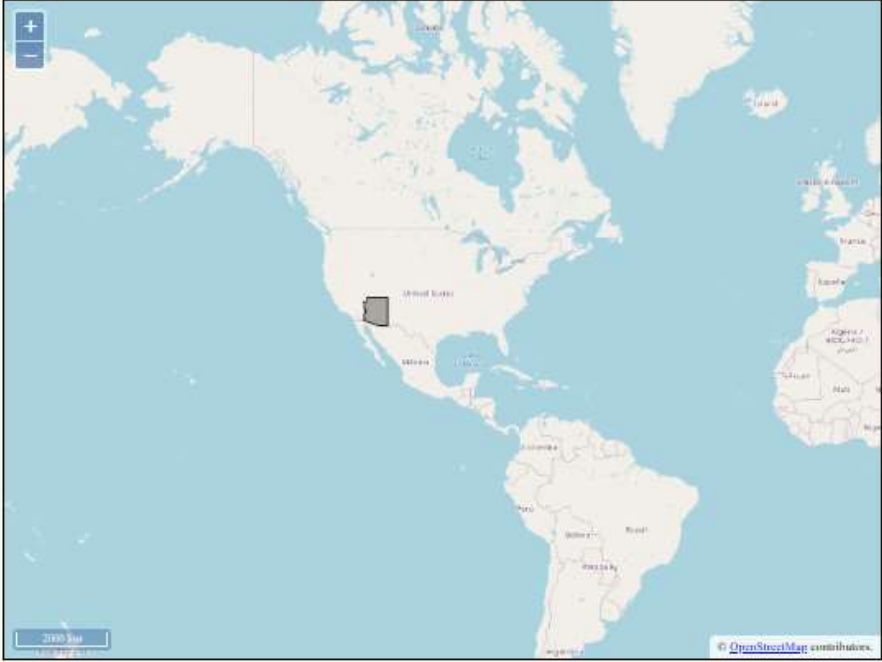
[Instructions](#)

Enter Primary species or genus. Separate synonyms by commas.

Optional Companion species:

Select States/Countries:

States:
Arizona



☒ Show US States
☐ Show Canadian Provinces
☐ Show Mexican Estados

Figure 1. Upper half of opening NEARBY screen. “Instructions” near top leads to help page on using NEARBY at this stage. In this case the user has chosen *Pinus ponderosa* as the Primary species. The option to “Show US States” has been toggled on and Arizona has been selected as the search area.

☒ Show US States
☐ Show Canadian Provinces
☐ Show Mexican Estados

☐ Separate database for Primary and Secondary specimens

Database:

☒ SEINet
☐ NEOTROPICAL
☐ CNALH (Lichens)

Collection:

SEINet - Academy of Natural Sciences of Drexel University
 SEINet - Ada Hayden Herbarium (Iowa State University)
 SEINet - Adams State University Herbarium
 SEINet - Albion College
 SEINet - Amistad National Recreation Area
 SEINet - Amway Herbarium
 SEINet - Andrews University Herbarium
 SEINet - Appalachian State University, I. W. Carpenter, Jr. Herbarium
 SEINet - Archbold Biological Station
 SEINet - Arizona State University Fruit and Seed Collection
 SEINet - Arizona State University Pollen Collection
 SEINet - Arizona State University Vascular Plant Herbarium

Distance: Meters
 Limit: Primary specimens
 Cull specimens if more than: Specimens at exact same coordinates
 Truncate list of species at : Secondary species
☒ Don't allow Primary specimens near other Primary specimens.
☐ Stop after obtaining primary species data

Figure 2. Screenshot of lower half of first screen of NEARBY with default parameters. Only the SEINet data base has been selected. See text for further explanation of how the parameters affect a search.

Output Display. Once all the numbers have been calculated, a table is displayed. A small portion of the table for this example is shown in Fig. 3. The table can be highlighted and copied into Excel, or similar program, for later reference if desired.

The output is divided into ten columns explained below. Clicking on the header for each column will sort the table by that column (e.g., by family or median elevation). Clicking on a scientific name will open a new page in SYMBIOTA showing details on that species. Clicking on “Show Images” at the top will add a thumbnail image of each Secondary species. Mousing over that thumbnail will pop-up a larger image and clicking on the thumbnail will open an even larger image in a separate window.

At the top of the page information about the search is given (e.g., the parameters used). In this case 276 specimens were found, but these were reduced to 222 “usable” specimen localities. Reasons for reducing specimens might be that two or more were found with exactly the same coordinates or were close enough to be considered Secondaries of each other or were found at a locality deemed to be mechanically georeferenced as a group. All numbers in the calculations below refer to “usable” specimens only.

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

Map: Just above the upper left-hand corner of the table is the word “Map.” By clicking on Map, a dot map of the Primary species is presented in a separate window. On the left side is a list of the Secondary species and by clicking on each name the species distribution is shown as dots, each with a distinctive color. This option will be discussed in more detail below.

Total Secondary Count: The total number of this Secondary species specimens collected and georeferenced in the defined region (e.g., 533 *Quercus gambelii* in Arizona).

Secondaries near this Primary: The number of georeferenced specimens of this species collected near a Primary species specimen (e.g., 50 *Quercus gambelii* were near a *Pinus ponderosa*).

% Secondaries near this Primary: The percent of georeferenced specimens of this species that were collected near a Primary species specimen, (e.g. $50/533 = 9.3\%$ of the *Quercus gambelii* were collected near a *Pinus ponderosa*). When the number of Primaries (set by the “Limit”) is less than the total in the selected region, then this percentage of Secondary species specimens in the defined area is adjusted to approximately compensate for the reduced sample. None of the other results need this adjustment.

Primaries near this Secondary: The number of Primary species specimens that have at least one of this Secondary species collected nearby (e.g. 39 of the 222 *Pinus ponderosa* specimens had a *Quercus gambelii* collected nearby).

%Primaries near this Secondary: The percent of the total Primary species specimens that have at least one specimen of this Secondary collected nearby (e.g. $39/222 = 17.5\%$ of the *Pinus ponderosa* had at least one *Quercus gambelii* collected nearby).

Scientific Name: Genus and specific epithet for the Secondary species, including all specimens identified as subspecies and varieties. Synonyms are not included for Secondary species. NEARBY uses the names as they appear in the databases, so Secondary species may appear under more than one name that the user considers synonyms or even common misspellings.

Family: Plant family name for each species.

Median Elevation: Median Elevation of the listed Secondary species. Median is determined by all the “usable” specimens in the defined region that have elevation data. Those specimens culled because of identical coordinates are not included.

The last columns are **Median Latitude** and **Median Longitude** of the listed Secondary species. These may prove useful in a broad geographic search.

Primary: *Pinus ponderosa*
 States: Arizona
 Databases: SEINet
 Meters: 100
 Maximum specimens same coordinates: 50
 Nearby primaries removed
 Primary Limit: 1000
 Found 276 total *Pinus ponderosa*
 Filtered for usable locations, retaining 222 [Pinus ponderosa](#) localities
 Median Elevation = 2134 meters
 Median Lat = 34.384, Median Long = -111.114
[Explanation of Output](#)
[Start over](#)
[Show Images](#)
[Map](#)



Total Secondary Count	Secondaries near this Primary	% Secondaries near this Primary*	Primaries near this Secondary	% Primaries near this Secondary	Scientific Name	Family	Median Elevation	Median Latitude	Median Longitude
533	50	9.3%	39	17.5%	Quercus gambelii	Fagaceae	1981	34.880	-111.449
256	44	17.1%	37	16.6%	Pseudotsuga menziesii	Pinaceae	2286	34.407	-110.892
1214	37	3%	32	14.4%	Erigeron divergens	Asteraceae	1760	34.486	-111.467
101	33	32.6%	31	13.9%	Pinus strobiformis	Pinaceae	2560	32.696	-110.723
389	36	9.2%	30	13.5%	Ceanothus fendleri	Rhamnaceae	2133	34.321	-111.022
163	28	17.1%	27	12.1%	Abies concolor	Pinaceae	2316	34.398	-110.982
847	36	4.2%	26	11.7%	Elymus elymoides	Poaceae	1830	34.711	-111.517
400	34	8.5%	25	11.2%	Juniperus deppeana	Cupressaceae	1700	33.394	-110.875
916	29	3.1%	25	11.2%	Poa fendleriana	Poaceae	1798	34.921	-111.677
148	29	19.5%	24	10.8%	Arceuthobium vaginatum	Santalaceae	2226	34.170	-110.988

Figure 3. Output of a query of *Pinus ponderosa* in Arizona. “Explanation of Output” near top leads to help page on using NEARBY at this stage. In this case the Primary species is *Pinus ponderosa* and the search area Arizona. The area around each Primary species specimen is defined by “Meters: 100” and refers to the distance from the Primary species specimens to the edges of a square sampling area. Each Primary species specimen defines a square 200 m on a side with the specimen at the center. Any locality with more than 50 specimens with exactly the same coordinates has been eliminated. When two Primaries are close enough to be considered Secondaries of each other, one is removed. After filtering for these parameters 222 “usable” Primaries were found. Only the 10 Secondary species with the highest “% Primaries near this Secondary” are shown. The default is set at 50 species, but the number can be changed.

By sorting on **Family** one can produce a systematic list of the plant species most likely to be encountered in a *P. ponderosa* forest—a mini-flora. And it can easily be illustrated by clicking on “**Show Images.**”

By sorting on **% Primaries near this Secondary** (the default sorting) we order the list by the most frequently associated species. We might want to choose one of these as a companion species to run another query that is more focused on a particular vegetation type.

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

By sorting on **% Secondaries near this Primary** we order the Secondary species as to how consistently they occur with the Primary species. If the total number of specimens of a Secondary species is small (e.g., less than 10), the percentage may be high, but probably of limited significance.

By sorting on **Median Elevation**, we may see how the Secondary species change with elevation. This can be interesting in areas with significant differences in elevation.

No matter how the Secondary species names are sorted they are reproduced as a list with family names at the end of the output file. They can easily be copied and pasted into another program such as Excel or Access for further analysis.

CASE STUDIES

We here present a few sample uses of NEARBY. These are not meant to be definitive studies but only examples of what might be done.

Secondary species commonly growing with a Primary. The objective of this example was to compile a list of the plants commonly found with *Kalmia latifolia* (Ericaceae), which was used as the Primary species. The search was made in the eastern states of Alabama, Florida, Georgia, Kentucky, Maryland, Mississippi, New Jersey, New York, North Carolina, Pennsylvania, South Carolina, Tennessee, Virginia, and West Virginia. The default settings were used except that a list of 100 Secondary species was obtained instead of 50. In this example 773 “usable” *Kalmia latifolia* specimens were found. The 100 Secondary species found belonged to 50 families, most with one, two, or three species. Cyperaceae and Pinaceae each had four species, and Rosaceae had 6 species. Only Ericaceae had more, with 14 species. If one includes *K. latifolia* itself, which was necessarily present by being the Primary, the number is 15. This is an unusual family distribution and is probably caused by the preference or tolerance of Ericaceae to acid soils. Ten of the species with the highest “**% Primaries near this Secondary**” are listed in Table 1.

Table 1. Sample of ten species found with *Kalmia latifolia* in eastern United States. These species had the highest “**% Primaries near this Secondary**”.

%	Species	Family
11.5	<i>Hamamelis virginiana</i>	(Hamamelidaceae)
8.5	<i>Vaccinium stamineum</i>	(Ericaceae)
8.4	<i>Tsuga canadensis</i>	(Pinaceae)
7.5	<i>Gaylussacia baccata</i>	(Ericaceae)
6.8	<i>Viburnum acerifolium</i>	(Adoxaceae)
6.8	<i>Acer rubrum</i>	(Sapindaceae)
6.7	<i>Vaccinium corymbosum</i>	(Ericaceae)
6.7	<i>Castanea dentata</i>	(Fagaceae)
6.7	<i>Sassafras albidum</i>	(Lauraceae)
6.7	<i>Gaultheria procumbens</i>	(Ericaceae)

Using the Map option. Here we discuss three studies with NEARBY that use the map option. In Fig. 3, just above the left-hand corner of the table, is the word “Map.” By clicking on “Map” a dot map of the Primary species will appear along with a list of Secondary species

ordered as to “% Primaries near this Secondary” (e.g., Fig. 4). By clicking on a Secondary species name it also appears in the same map with colored dots. Multiple species can be shown.

Map study 1: In this case we used both the SEINet and NEOTROPICAL databases. With a little investigation we found the *Escallonia myrtilloides* (Escalloniaceae) is a common high elevation species with a median elevation of 3250 m. We used it as the Primary species and found that its associated Secondary species mainly have median elevations between 3000 and 3800 m. In this case the distributions are often quite similar (Fig. 4), presumably because their habitat preferences are restricted to a narrow area of high elevations. The most common families among the first 50 associated species were Ericaceae and Melastomataceae with 7 species each.

Nearby Map -- Specimen Coordinates

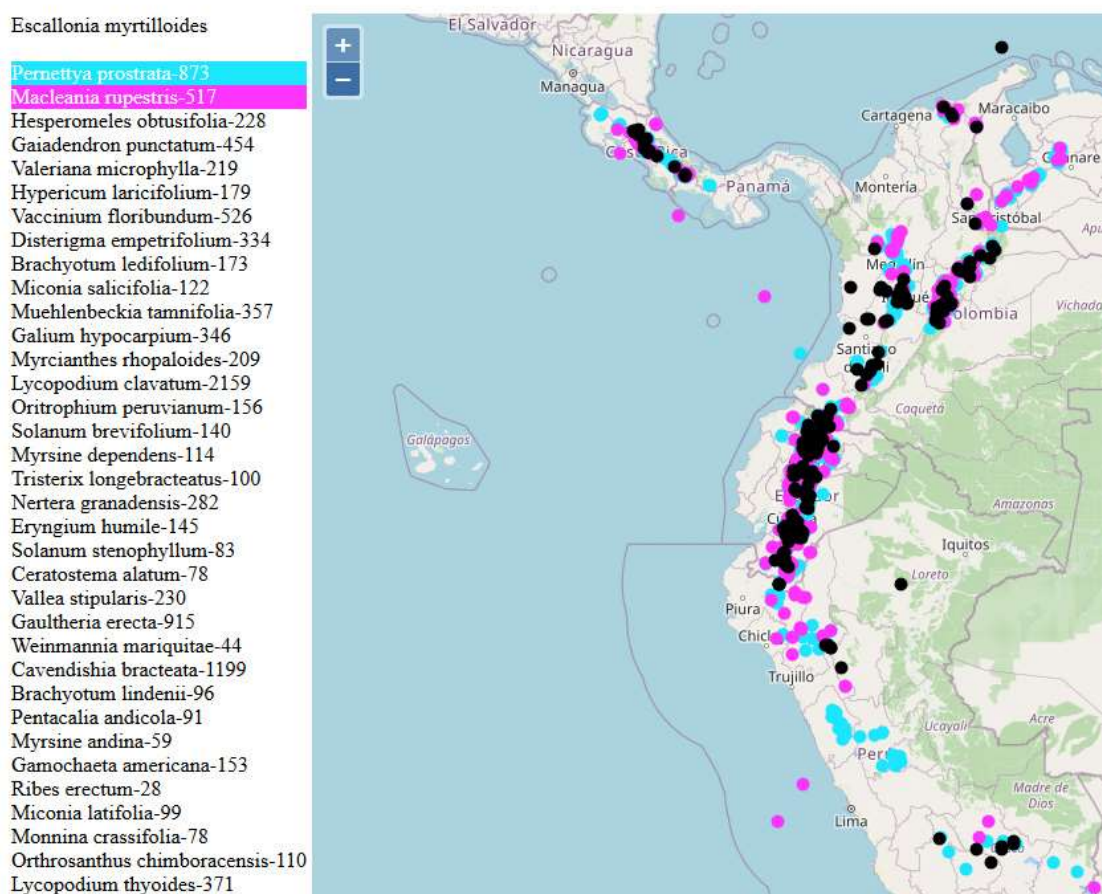


Figure 4. Distribution of *Escallonia myrtilloides* (Escalloniaceae; black dots) and its two most common Secondary species, both Ericaceae: *Pernettya prostrata* (light blue dots) and *Macleania rupestris* (pink dots). The few outlying dots in the ocean are examples of georeferencing errors and are a rough indication of the error rate.

Map study 2: Interior Chaparral vegetation in Arizona. Interior Chaparral (Brown 1994; Brown et al. 2007) is one of the common types of vegetation in Arizona at elevations of about 1000 to 1800 m. The dominant species are shrubs, often with

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

sclerophyllous leaves, that are adapted to frequent fires by sprouting from roots or stems below ground level. Chaparral is normally found above the Sonoran Desert scrub but lower than Oak Woodland or Ponderosa Pine forest (see Brown 1994 for vegetation types and Brown et al. 2007 for vegetation map). The lines between these kinds of vegetation are often indistinct. Chaparral is often found adjacent to or mixed with Pinyon-Juniper vegetation.

Among the most common dominant or co-dominant species in Interior Chaparral are *Arctostaphylos pringlei* and *A. pungens* (Ericaceae), *Ceanothus greggii* (Rhamnaceae), *Cercocarpus montanus* (Rosaceae), *Eriodictyon angustifolium* (Boraginaceae), *Fendlera rupicola* (Hydrangeaceae), *Garrya flavescens* and *G. wrightii* (Garryaceae), and *Quercus turbinella* (Fagaceae). We have noticed that in different regions chaparral vegetation may have different dominant species and these grow in a variety of associations.

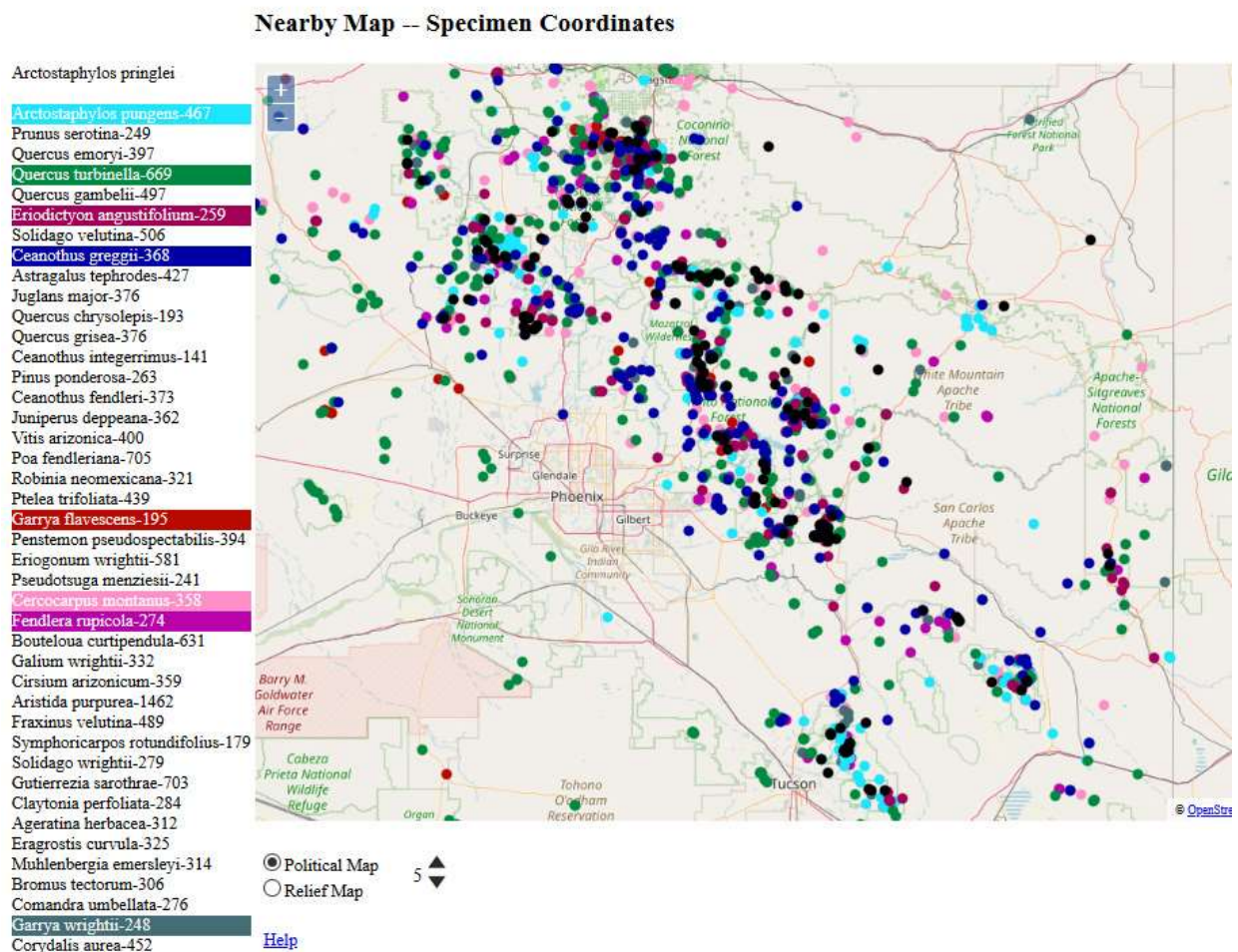


Figure 5. Map of nine Interior Chaparral species in central Arizona. The rich Interior Chaparral of central Arizona is a combination of several independent plant species distributions that happen to coincide in that area. Black dots indicate the Primary species, *Arctostaphylos pringlei*, and colored dots are for eight Secondary species. Chaparral species grow beyond central Arizona, but fewer species grow together. See Fig. 6 for an example.

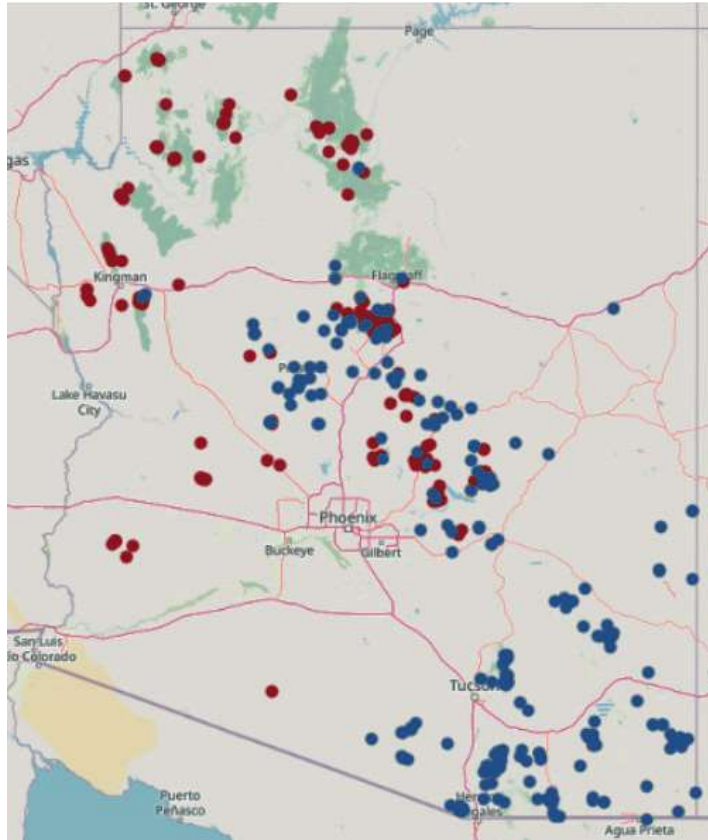


Figure 6. *Garrya flavescens* (red dots) extending mainly to north and west and *G. wrightii* (blue dots) mainly extending to east and south.

To investigate the variety of chaparral types in Arizona we started with *Arctostaphylos pringlei* as a Primary and found all the other species mentioned above appeared as Secondaries using the default settings. Using the map option, we mapped all these species on a single map (Fig. 5). All nine species are commonly found together in areas of chaparral between Flagstaff and Tucson. While using the map option as shown in Fig. 5, one can hide the black dots of the Primary species by hovering over its name at the top, and display any combination of species by clicking on the names (a single click toggles the distribution dots on or off). As this is done it becomes clear that the central Arizona highland supports a diverse set of chaparral species that overlap. But individually each species has a distinct distribution that may extend in various directions. For instance, if we reduce the species to *Garrya flavescens* and *G. wrightii* only (Fig. 6), we see that for much of their ranges they do not overlap. Thus, Interior Chaparral is most diverse and complex where the most species overlap in central Arizona. As one travels in any direction from that area, there will be fewer species and chaparral may have a notably different appearance.

Map study 3: When doing the chaparral species study we noticed that *Comandra umbellata* was often found with chaparral species. Unknown to many collectors of this species is that it is a hemiparasite. We thought we might find a host-parasite

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

relationship using NEARBY. Using *C. umbellata* as the Primary species we found seven of the nine chaparral species cited above appear as commonly found Secondaries. We used the map feature to compare its distribution in Arizona (Fig. 7) with the two most commonly found Secondary species (*Ceanothus greggii* and *Arctostaphylos pungens*). *Comandra umbellata* seems to be a common associate of these two chaparral species, but also grows where neither species is found. *Comandra umbellata* may grow with these species because of a similar habitat preference (rocky dry soils) rather than being a host specific hemiparasite of either. Through some literature research we found that *C. umbellata* is widespread in North America and in the Mediterranean region growing well beyond any chaparral species (Nickrent 2016). It is reported to have various hosts (Pielh 1965), so various chaparral and non-chaparral species may be among them.

Nearby Map -- Specimen Coordinates

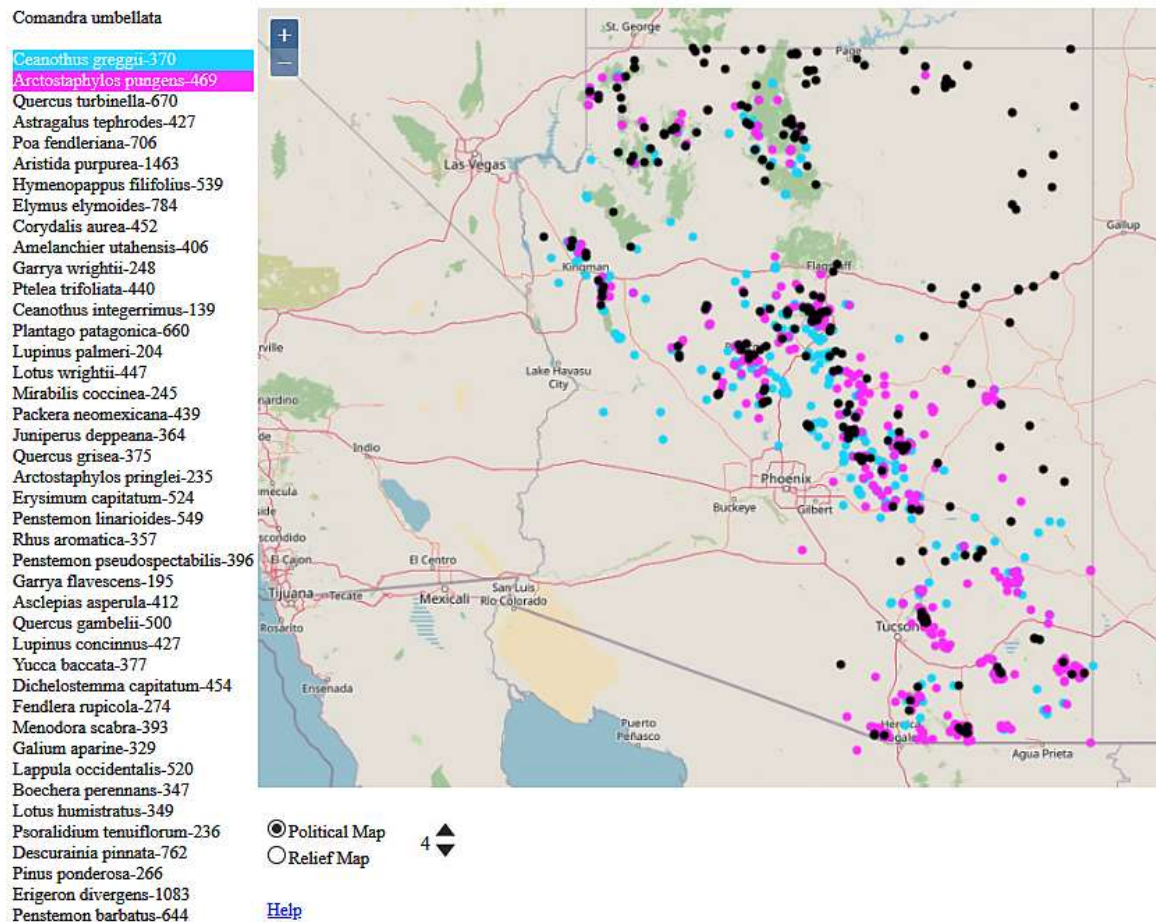


Figure 7. Map of *Comandra umbellata* in Arizona (black dots). Seven of the most commonly found Secondary species in left hand column are among the nine chaparral species we used in the study above. The first two Secondary species are represented by colored dots in this map: *Ceanothus greggii* (light blue), *Arctostaphylos pungens* (pink). *Comandra umbellata* seems to be mainly found in chaparral vegetation in Arizona but probably not restricted to it.

Using different databases for the Primary and Secondaries species. What vascular plants grow near the lichen *Xanthomendoza fallax* in Arizona and New Mexico? *Xanthomendoza fallax* generally grows on the bark of trees, so we might expect it to be associated with tree species. In this example we used different databases for the Primary and Secondary species. In the opening page of Nearby (Fig. 2), above the list of databases, is a box to check that facilitates doing this. We checked the box and selected the CNALH (Lichens) database for the Primary species and SEINet for the Secondary species. We raised the distance from the Primary to 500 m instead of the default 100 m to obtain a larger sample of Secondaries. We only searched for the first 25 vascular plant species, and among these were four tree species that might provide a substrate for *X. fallax*; they are listed with a map in Fig. 8. The high elevation *Quercus rugosa* indicates oak forests of southeastern Arizona and southwestern New Mexico; and *Fraxinus velutina*, *Acer negundo*, and *Juglans major* are species that grow near oak forest species but also along drainages at lower elevations. So, we can deduce that *X. fallax* probably grows in forests at higher elevations and extends to lower elevations along moist drainages.

Nearby Map -- Specimen Coordinates

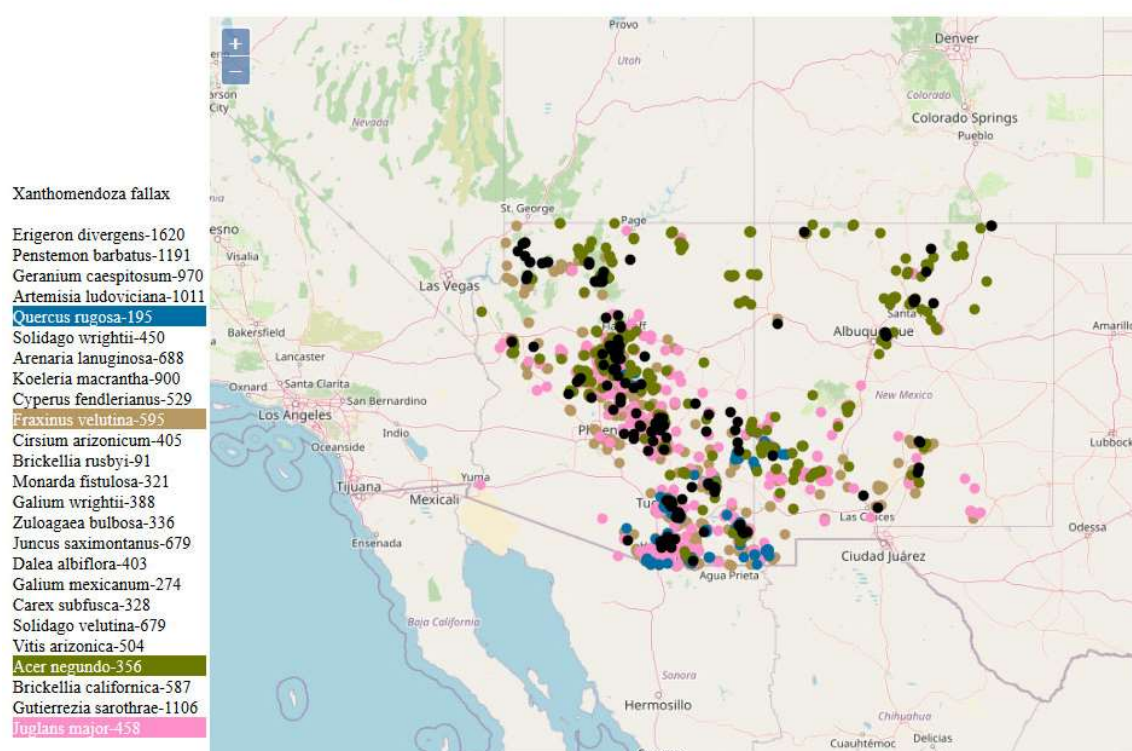


Figure 8. Map of lichen *Xanthomendoza fallax* as the Primary species using CNALH. Secondary species of vascular plants are drawn from SEINet. Black dots indicate sites for *X. fallax*, with four tree species indicated by colored dots. See text for explanation.

CONCLUSION

We believe that NEARBY offers a useful way to explore specimen databases for research and teaching. Perhaps the most broadly useful aspect will be making checklists of Secondary species commonly growing near one or a few Primary species. This should be a

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

useful tool for anyone exploring a particular kind of vegetation, especially before or soon after fieldwork. Comparing the checklists of different species in one general area, or the same species in different areas should be useful in phytosociological studies. The maps will be useful in discovering patterns of distribution, even in areas where the investigator has never been. We remind anyone using NEARBY that it is not a substitute for fieldwork nor in-depth studies of particular taxonomic groups or kinds of vegetation. We hope it will be used mainly as a tool for preparing for fieldwork and developing hypotheses for further investigation.

Biotic communities, including all organisms, are largely influenced by the plants that live in them. Plants, along with photosynthetic algae and bacteria, are the primary producers. Plants are often large and easily visible, so that when we describe a community of organisms, plants are often the principle entities mentioned. If databases of other organisms such as insects, birds, and fungi, are eventually added to NEARBY, we will be able to discover communities, not only of plants, but of a variety of organisms.

Databases are a good example of the power of working collectively. The larger the databases, and the broader their geographic and taxonomic coverage, the more useful they are. We hope that this program works as an incentive for additional collections to join the databases we have used here and for other databases to become part of NEARBY.

ACKNOWLEDGMENTS

This study would not have been possible without the three specimen databases we have used: SEINet plant database, NEOTROPICAL plant database, and CNALH database of the Consortium of North American Lichen Herbaria. Edward Gilbert facilitated access to the first two, and Frank Bungartz access to the third. Frank also offered valuable advice about lichen species to investigate. Work on this project was hosted at the Arizona State University herbarium where DLL is an adjunct professor and LRL is Curator Emeritus. Various NSF grants have supported the development of these databases and thousands of collectors have contributed millions of specimens. We are grateful to all. Sue Carnahan, Liz Makings, and Frank Bungartz read early versions of this manuscript and offered many helpful suggestions.

LITERATURE CITED

- Brown, D. E. (ed.). 1994. Biotic Communities: Southwestern United States and Northwestern Mexico by University of Utah Press, Salt Lake City. [Revised from Brown, D. E. (Ed.). 1982. The biotic communities of the American Southwest – United States and Mexico. *Desert Plants* 4 (1-4): 1-341.]
- Brown, D. E., T. C. Brennan, & P. J. Unmack. 2007. A Digitized Biotic Community Map. *CANOTIA* 3 (1): 1-12.
- CRIA-Centro de Referência em Informação Ambiental. 2020. *SpeciesLink*: simple search. Available at <http://www.splink.org.br/>. Accessed 2014–2020.
- Darwin, C. 1859. *Origin of Species by means of Natural Selection*. Reprint of first edition (1859). London: Penguin Books Ltd.
- GBIF.org. 2020. *GBIF Home Page*. Available from: <https://www.gbif.org>
- Gleason, H. A. 1926. The individualistic concept of the plant association. *Bull. Torrey Bot. Club* 53(1): 7–26. <http://dx.doi.org/10.2307/2479933>

- Gries, C., E. E. Gilbert, & N. Franz. 2014. Symbiota – A virtual platform for creating voucher-based biodiversity information communities. *Biodiversity Data Journal* 2: e1114. doi: 10.3897/BDJ.2.e1114
- Hooker, J. D. 1853. Introductory Essay, page xix, in *The Botany of the Antarctic Voyage, Of H. M. Discovery Ships Erebus and Terror in the Years 1839–1843 Under the Command of Captain Sir James Clark Ross. Flora Novae-Zelandiae - Flowering Plants.*
- Humboldt, A. von & A. Bonpland (English version edited by S. T. Jackson; translated from original French version of 1807 by S. Romanowski). 2009. *Essay on the Geography of Plants.* University of Chicago Press.
- Landrum, L. R. & D. Lafferty. 2015. PROXIMITY and CORRELATION: two new computer programs for Mining Phytosociological Information held in Herbarium Databases using central Arizona as a test case. *Taxon* 64: 998-1016.
- Nickrent, D. L. 2016. Comandraceae Nickrent & Dur, Bastard Toadflax Family. *Flora of North America* Vol. 12: 408–412.
- Piehl, M. A. 1965. The natural history and taxonomy of *Comandra* (Santalaceae). *Mem. Torrey Bot. Club* 22(1): 1–97.

Appendix A: Program Operation.

Primary Search. The MySQL query for the Primary species returns *latitude*, *longitude*, *elevation*, *collector*, *collection number*, and *catalog number*. The search includes all subspecies and varieties by appending the MySQL wildcard “%” to the *sciname* parameter but only uses the species binomial. Synonyms are not included unless entered by the user at the beginning of the program.

The search is limited to the geographic area (e.g., states, countries, whole world) selected by the user. It should be kept in mind that the databases mainly include specimens from the Western hemisphere and NEARBY should be restricted to searches in the Americas. If more Primary specimens are found than the user-defined limit, the list of Primary specimens is randomized before the next steps. This is to prevent systematic error resulting from the order in which the specimens were entered into the database.

Primary with Companion Search. If a Companion species has been entered, the next step is to query the database for any instances of a Companion specimen within the specified limit of (i.e., nearby) any Primary species specimen. The program iterates through each Primary species specimen one at a time, defining the square around the specimen and finding any Companion specimen within that square. If a Primary species specimen does not have a Companion species within the limit, that Primary species specimen is removed from the list. Note again, this only applies if a Companion species has been entered on the initial page of NEARBY.

The advantage of including a Companion species is that one can better define an environment—in particular, if the Primary species has a broad range. For example, if *Larrea tridentata* were used as the Primary, limiting the search with *Olneya tesota* as the Companion would focus on a more specific vegetation type, i.e., a subtype of Sonoran Desert vegetation.

Discovering Secondary Species. The next step is to iterate through each Primary species specimen and find all the other specimens within the defined limit of that Primary species specimen. All of these Secondary species specimens are collated and counted, creating a list

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

of all Secondary species found near any Primary species specimens. Secondary synonyms are not combined, but all subspecies and varieties are united with the binomial. If a limit on Primary specimens is set, this is where that limit is enforced. Once the program has found Secondary specimens for the limit number of Primary specimens, the iteration stops and the program continues to the next stage. Note that if the list of “usable” Primaries exceeds the limit parameter, they would have been randomized, so that repeated runs with exactly the same setting might yield slightly different results.

Sorting and Counting. Next the list of Secondary specimens is sorted by descending number of Primary specimens found near each Secondary species. Then only the Secondary species with the greatest number are retained, based on the number at “**Truncate list of Secondary species at.**” The default is 50.

Finally, another query is made of the database to find the total number of georeferenced specimens of the most common Secondary species found in the defined region (e.g., state or country, or worldwide). This number is used to determine what percent of Secondary specimens grow near a Primary species. The total number of Secondaries found nearby is often just an indication of how often it was collected, but by then dividing it by the total number found in the entire region, we get a good estimate of how frequently that Secondary grows near the Primary species.

Table 2. Summary of specimen selection. The process of selecting specimens to be used in a NEARBY search follows 7 steps that are summarized here.

1. A search is made for all georeferenced Primary species occurrences in the defined geographic area with coordinate data. If multiple specimens have identical coordinates, only one is used.
2. The upper limit of Primary specimens to be used is the “**Limit.**” The default **Limit** is 1000 but another number may be used. If the number of specimens found exceeds the **Limit**, they are randomized.
3. If a Primary species specimen has no specimens of any other species nearby, it is eliminated.
4. If the option “**Don’t allow Primary specimens near other Primary specimens**” is selected (the default is to select it), and any two Primary specimens are close enough to be considered Secondaries of each other, only one is used.
5. If a **Companion** species has been selected, only Primary specimens with Companion specimens close enough to be considered Secondary specimens are used.
6. “**Cull specimens if more than.**” The **Cull number** is a way of avoiding the use of specimens georeferenced mechanically to a common point, such as the middle of a town. The **Cull number** is set by the user, or the default (50) is accepted. In that case, if in a search for specimens, 50 or more specimens have exactly the same coordinates, they are eliminated. This applies to Secondaries alone and to each Primary and all its Secondaries counted together. Note that locations with more than 100 identically georeferenced specimens are always omitted from all searches, since these locations have almost certainly been programmatically georeferenced to a common point. Therefore, setting this number higher than 100 is the same as setting it to 100. By automatically eliminating these specimens we avoid known situations where hundreds, thousands or tens of thousands of specimens have exactly the same coordinates, which can cause the program to stall as it tries to analyze all the nearby relationships and will give meaningless results.
7. If after the above steps, the “**Limit**” is a number smaller than the remaining number of Primary specimens available, the randomized Primary specimens are reduced to the “**Limit**” number.

Appendix B: Alphabetical list of species names that appear in this paper.

Scientific Name	Scientific Name Authorship	Family
<i>Acer negundo</i>	L.	ACERACEAE
<i>Acer rubrum</i>	L.	ACERACEAE
<i>Ageratina herbacea</i>	(A. Gray) R. M. King & H. Rob.	ASTERACEAE
<i>Amelanchier utahensis</i>	Koehne	ROSACEAE
<i>Arctostaphylos pringlei</i>	Parry	ERICACEAE
<i>Arctostaphylos pungens</i>	H.B.K.	ERICACEAE
<i>Arcytophyllum nitidum</i>	(H.B.K.) Schltld.	RUBIACEAE
<i>Arenaria lanuginosa</i>	(Michx.) Rohrb.	CARYOPHYLLACEAE
<i>Aristida purpurea</i>	Nutt.	POACEAE
<i>Artemisia ludoviciana</i>	Nutt.	ASTERACEAE
<i>Asclepias asperula</i>	(Decne.) Woodson	APOCYNACEAE
<i>Astragalus tephrodes</i>	A. Gray	FABACEAE
<i>Baccharis tricuneata</i>	(L. fil.) Pers.	ASTERACEAE
<i>Boechera perennans</i>	(S. Watson) W. A. Weber	BRASSICACEAE
<i>Bouteloua curtipendula</i>	(Michx.) Torr.	POACEAE
<i>Brachyotum ledifolium</i>	(Desr.) Triana	MELASTOMATACEAE
<i>Brachyotum lindenii</i>	Cogn.	MELASTOMATACEAE
<i>Brickellia californica</i>	(Torr. & A. Gray) A. Gray	ASTERACEAE
<i>Brickellia rusbyi</i>	A. Gray	ASTERACEAE
<i>Bromus tectorum</i>	L.	POACEAE
<i>Bucquetia glutinosa</i>	(L. fil.) DC.	MELASTOMATACEAE
<i>Calamagrostis intermedia</i>	Steud.	POACEAE
<i>Carex subfusca</i>	W. Boott	CYPERACEAE
<i>Castanea dentata</i>	(Marshall) Borkh.	FAGACEAE
<i>Castilleja fissifolia</i>	L. fil.	OROBANCHACEAE
<i>Castratella piloselloides</i>	(Bonpl.) Naud.	MELASTOMATACEAE
<i>Cavendishia bracteata</i>	(Ruiz & Pav. ex J.St.Hil.) Hoerold	ERICACEAE
<i>Ceanothus fendleri</i>	A. Gray	RHAMNACEAE
<i>Ceanothus greggii</i>	A. Gray	RHAMNACEAE
<i>Ceanothus integerrimus</i>	Hook. & Arn.	RHAMNACEAE
<i>Ceratostema alatum</i>	(Hoerold) Sleumer	ERICACEAE
<i>Cercocarpus montanus</i>	Raf.	ROSACEAE
<i>Cirsium arizonicum</i>	(A. Gray) Petr.	ASTERACEAE
<i>Claytonia perfoliata</i>	Donn ex Willd.	MONTIACEAE
<i>Comandra umbellata</i>	(L.) Nutt.	SANTALACEAE
<i>Corydalis aurea</i>	Willd.	PAPAVERACEAE
<i>Cyperus fendlerianus</i>	Boeckeler	CYPERACEAE
<i>Dalea albiflora</i>	A. Gray	FABACEAE
<i>Descurainia pinnata</i>	(Walter) Britton	BRASSICACEAE
<i>Dichelostemma capitatum</i>	(Benth.) Alph. Wood	ASPARAGACEAE

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

<i>Disterigma empetrifolium</i>	(H.B.K.) Niedenzu ex Drude	ERICACEAE
<i>Elymus elymoides</i>	(Raf.) Swezey	POACEAE
<i>Eragrostis curvula</i>	(Schrad.) Nees	POACEAE
<i>Erigeron divergens</i>	Torr. & A. Gray	ASTERACEAE
<i>Eriodictyon angustifolium</i>	Nutt.	BORAGINACEAE
<i>Eriogonum wrightii</i>	Torr. ex Benth.	POLYGONACEAE
<i>Eryngium humile</i>	Cav.	APIACEAE
<i>Erysimum capitatum</i>	(Douglas ex Hook.) Greene	BRASSICACEAE
<i>Escallonia myrtilloides</i>	L. fil.	ESCALLONIACEAE
<i>Fendlera rupicola</i>	A. Gray	HYDRANGEACEAE
<i>Fraxinus velutina</i>	Torr.	OLEACEAE
<i>Gaiadendron punctatum</i>	G. Don	LORANTHACEAE
<i>Galium aparine</i>	L.	RUBIACEAE
<i>Galium hypocarpium</i>	(L.) Endl. ex Griseb.	RUBIACEAE
<i>Galium mexicanum</i>	H.B.K.	RUBIACEAE
<i>Galium wrightii</i>	A. Gray	RUBIACEAE
<i>Gamochaeta americana</i>	(Mill.) Wedd.	ASTERACEAE
<i>Garrya flavescens</i>	S. Wats.	GARRYACEAE
<i>Garrya wrightii</i>	Torr.	GARRYACEAE
<i>Gaultheria anastomosans</i>	(Mutis ex L.f.) H.B.K.	ERICACEAE
<i>Gaultheria erecta</i>	Vent.	ERICACEAE
<i>Gaultheria procumbens</i>	L.	ERICACEAE
<i>Gaylussacia baccata</i>	(Wangenh.) K. Koch	ERICACEAE
<i>Geranium caespitosum</i>	James	GERANIACEAE
<i>Gutierrezia sarothrae</i>	(Pursh) Britton & Rusby	ASTERACEAE
<i>Hamamelis virginiana</i>	L.	HAMAMELIDACEAE
<i>Heliomeris multiflora</i>	Nutt.	ASTERACEAE
<i>Hesperomeles obtusifolia</i>	(Pers.) Lindl.	ROSACEAE
<i>Hymenopappus filifolius</i>	Hook.	ASTERACEAE
<i>Hypericum laricifolium</i>	Juss.	HYPERICACEAE
<i>Juglans major</i>	(Torr.) Heller	JUGLANDACEAE
<i>Juncus saximontanus</i>	A. Nels.	JUNCACEAE
<i>Juniperus deppeana</i>	Steud.	CUPRESSACEAE
<i>Kalmia latifolia</i>	L.	ERICACEAE
<i>Koeleria macrantha</i>	(Ledeb.) Schult.	POACEAE
<i>Lappula occidentalis</i>	(S. Wats.) Greene	BORAGINACEAE
<i>Larrea tridentata</i>	(Sessé & Moc. ex DC.) Coville	ZYGOPHYLLACEAE
<i>Lotus humistratus</i>	Greene	FABACEAE
<i>Lotus wrightii</i>	(A. Gray) Greene	FABACEAE
<i>Lupinus concinnus</i>	J. G. Agardh	FABACEAE
<i>Lupinus palmeri</i>	S. Watson	FABACEAE
<i>Lycopodium clavatum</i>	L.	LYCOPODIACEAE

<i>Lycopodium thyoides</i>	Humb. & Bonpl. Ex Wild.	LYCOPODIACEAE
<i>Macleania rupestris</i>	(H.B.K.) A. C. Sm.	ERICACEAE
<i>Menodora scabra</i>	A. Gray	OLEACEAE
<i>Miconia latifolia</i>	Naudin	MELASTOMATACEAE
<i>Miconia salicifolia</i>	Naudin	MELASTOMATACEAE
<i>Mirabilis coccinea</i>	(Torr.) Benth. & Hook. f.	NYCTAGINACEAE
<i>Monarda fistulosa</i>	L.	LAMIACEAE
<i>Monnina crassifolia</i>	(Bonpl.) Kunth	POLYGALACEAE
<i>Muehlenbeckia tamnifolia</i>	(H.B.K.) Meisn.	POLYGONACEAE
<i>Muhlenbergia emersleyi</i>	Vasey	POACEAE
<i>Myrcianthes rhopaloides</i>	(H.B.K.) McVaugh	MYRTACEAE
<i>Myrsine andina</i>	(Mez) Pipoly	PRIMULACEAE
<i>Myrsine dependens</i>	(Ruiz & Pav.) Spreng.	PRIMULACEAE
<i>Nertera granadensis</i>	(L. fil.) Druce	RUBIACEAE
<i>Olneya tesota</i>	A. Gray	FABACEAE
<i>Oritrophium peruvianum</i>	Cuatrec.	ASTERACEAE
<i>Orthrosanthus chimboracensis</i>	(Kunth) Baker	IRIDACEAE
<i>Packera neomexicana</i>	(A. Gray) W.A. Weber & A. Löve	ASTERACEAE
<i>Penstemon barbatus</i>	(Cav.) Roth	PLANTAGINACEAE
<i>Penstemon linarioides</i>	A. Gray	PLANTAGINACEAE
<i>Penstemon pseudospectabilis</i>	M.E. Jones	PLANTAGINACEAE
<i>Pentacalia andicola</i>	(Turcz.) Cuatrec.	ASTERACEAE
<i>Pentacalia vaccinioides</i>	(H.B.K.) Cuatrec.	ASTERACEAE
<i>Pernettya prostrata</i>	Sleumer	ERICACEAE
<i>Pinus edulis</i>	Engelm.	PINACEAE
<i>Pinus ponderosa</i>	P. Lawson & C. Lawson	PINACEAE
<i>Pinus strobiformis</i>	Engelm.	PINACEAE
<i>Plantago patagonica</i>	Jacq.	PLANTAGINACEAE
<i>Poa fendleriana</i>	(Steud.) Vasey	POACEAE
<i>Prunus serotina</i>	Ehrh.	ROSACEAE
<i>Pseudotsuga menziesii</i>	(Mirbel) Franco	PINACEAE
<i>Psoralidium tenuiflorum</i>	(Pursh) Rydb.	FABACEAE
<i>Ptelea trifoliata</i>	L.	RUTACEAE
<i>Quercus chrysolepis</i>	Liebm.	FAGACEAE
<i>Quercus emoryi</i>	Torr.	FAGACEAE
<i>Quercus gambelii</i>	Nutt.	FAGACEAE
<i>Quercus grisea</i>	Liebm.	FAGACEAE
<i>Quercus rugosa</i>	Née	FAGACEAE
<i>Quercus turbinella</i>	Greene	FAGACEAE
<i>Rhus aromatica</i>	Aiton	ANACARDIACEAE
<i>Rhus trilobata</i>	Nutt.	ANACARDIACEAE
<i>Ribes erectum</i>	Freire-Fierro	SAXIFRAGACEAE

A COMPUTER PROGRAM FOR PHYTOGEOGRAPHIC ANALYSIS

<i>Robinia neomexicana</i>	A. Gray	FABACEAE
<i>Sassafras albidum</i>	(Nutt.) Nees	LAURACEAE
<i>Solanum brevifolium</i>	Dunal	SOLANACEAE
<i>Solanum stenophyllum</i>	Dunal	SOLANACEAE
<i>Solidago velutina</i>	DC.	ASTERACEAE
<i>Solidago wrightii</i>	A. Gray	ASTERACEAE
<i>Symphoricarpus rotundifolius</i>	A. Gray	CAPRIFOLIACEAE
<i>Tristerix longibracteatus</i>	(Desr.) Barlow & Wiens	LORANTHACEAE
<i>Tsuga canadensis</i>	(L.) Carrière	PINACEAE
<i>Vaccinium corymbosum</i>	L.	ERICACEAE
<i>Vaccinium floribundum</i>	H.B.K.	ERICACEAE
<i>Vaccinium stamineum</i>	L.	ERICACEAE
<i>Valeriana microphylla</i>	H.B.K.	CAPRIFOLIACEAE
<i>Vallea stipularis</i>	L. fil.	ELAEOCARPACEAE
<i>Viburnum acerifolium</i>	L.	ADOXACEAE
<i>Vitis arizonica</i>	Engelm.	VITACEAE
<i>Weinmannia mariquitae</i>	Szyzyl.	CUNONIACEAE
<i>Xanthomendoza fallax</i>	(Arnold) Søchting	TELOSCHISTACEAE
<i>Yucca baccata</i>	Torr.	ASPARAGACEAE
<i>Zuloagaea bulbosa</i>	(H.B.K.) Bess	POACEAE