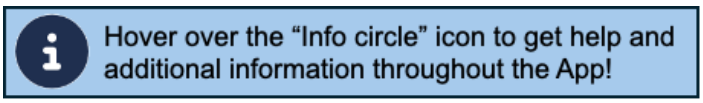
# NatureServe Rapid Analysis of Rarity and Endangerment Conservation Assessment Tool (RARECAT) Documentation

*Version*: 2.1.1 (released Mar 31, 2025)

*Documentation Last Updated*: Mar 31, 2025

To report issues or provide feedback, please submit tickets by either by logging into the RARECAT Help Desk or by emailing rarecatsupport@natureserve.org, which will automatically create a ticket.

*Suggested citation:*
NatureServe (2025). RARECAT version 2.1.1. Available from https://natureserve.shinyapps.io/RARECAT. Accessed [Date].




## Table of Contents

# 1. Background

NatureServe Conservation Status Assessments quantify Rarity, Threat, and Trend Factors for target taxa to produce Conservation Assessment Status Ranks at global (G Rank), national (N Rank), or subnational scales (S Rank). NatureServe's RARECAT (Rapid Analysis of Rarity and Endangerment Conservation Assessment Tool) facilitates extraction, vetting, and filtering of distribution data, visualization of data quality and quantity over space and time, and calculation of the three fundamental factors of Rarity – Range Extent (also known as Extent of Occurrence/EOO), Area of Occupancy (AOO), and Number of Occurrences – for single taxa or multiple taxa concurrently. Doing so, RARECAT supports higher quality, efficiency, and standardization in Conservation Assessment Status ranking practice for both common and rare taxa. RARECAT can be accessed at **https://natureserve.shinyapps.io/RARECAT**.

To cite RARECAT, we suggest the following citation template:

NatureServe (2025). RARECAT version 2.1.1. Available from https://natureserve.shinyapps.io/RARECAT. Accessed [Date].

# 2. How-to Guide

## *2.1. Application*

RARECAT is an R Shiny application hosted on NatureServe's shinyapps.io account (natureserve.shinyapps.io). Use your browser to navigate to the stable up-to-date version of RARECAT at the following URL: **https://natureserve.shinyapps.io/RARECAT**.

## *2.2. Modes*

RARECAT has two modes of use: "Single Species Mode" and "Multispecies Mode".

RARECAT Single Species mode enables users to focus on a single taxon, providing the greatest degree of flexibility and specificity in the exact data included and vetted, the spatial and temporal filters applied, and the outputs generated. RARECAT Single Species mode should be considered when the target taxon is known to have patchy or uncertain data of varying quality, or unusual geographical, ecological, or life history attributes.

RARECAT Multispecies mode enables users to select a shared set of input data and filters to run relatively quick assessments across up to 100 taxa concurrently. RARECAT Multispecies mode should be considered when users might benefit from a relatively quick and preliminary assessment across multiple taxa, before drilling down into the details of a subset of them.

A powerful feature of RARECAT is that its two modes of use interact: Multispecies mode inputs and outputs can be fed back into Single Species mode for specific taxa in need of review; following review, Single Species mode output can then be sent back into Multispecies mode.

## *2.3. RARECAT Single Species*

### *2.3.1. User Interface*

RARECAT “Single Species” mode follows a relatively simple user interface with a single page including 6 components (Figure 1): 1) the Assessment menu, 2) the Data and Analysis panel, 3) the Map, 4) the Filters panel, 5) the Records Table, and 6) the Temporal Change panel.

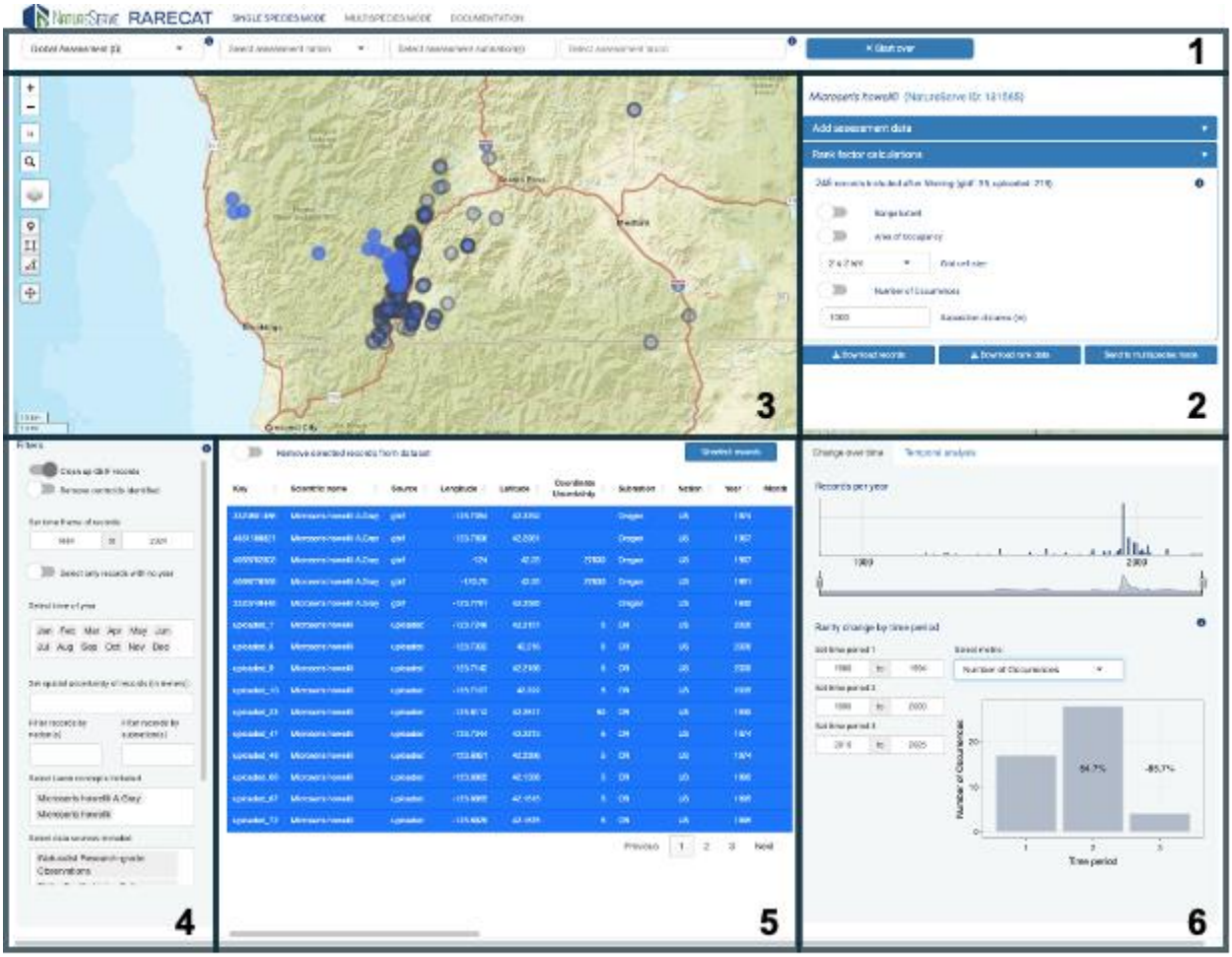



**Figure 1**: The RARECAT Single Species user interface comprises 6 components: 1) the Assessment menu, 2) the Data and Analysis panel, 3) the Map, 4) the Filters panel, 5) the Records Table, and 6) the Temporal Change panel.

### *2.3.1.1. Assessment Menu*

The Assessment menu appears on the top of the page and prompts the user to select the two primary parameters needed to run a rank assessment: the assessment geography and the assessment taxon. The first three dropdown menus to the left of the Assessment menu enable users to select the geography of the rank assessment at one of three levels: global, national, or subnational.

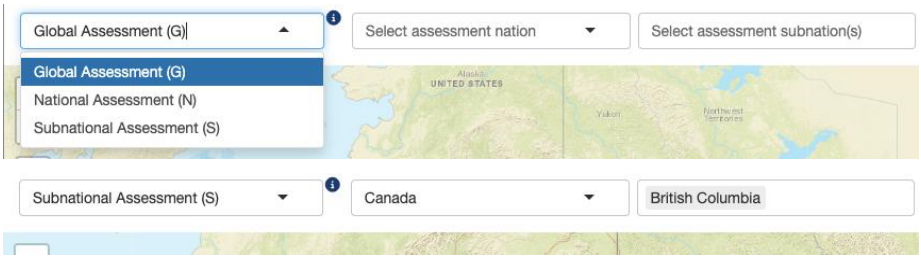



**Figure 2**: The Assessment menu enables users to select the geography of the rank assessment at one of three levels: global, national, or subnational. Dropdown menus 2 and 3 will only show corresponding options if a national or subnational assessment is selected.

The fourth component of the Assessment menu enables users to select the taxon to be assessed and the associated data to be used in the assessment. To begin, type a taxon name in the text box labeled “Select assessment taxon”; this will open up the taxon options panel. Currently, RARECAT suggests taxonomic names included within the NatureServe Taxonomy Backbone served via the NatureServe Explorer API and the Global Biodiversity Information Facility (GBIF) Taxonomic Backbone served via the GBIF API (Figure 3A). The taxon selected from the first taxon options panel will become the overall taxon associated with this assessment and its name will appear on the top of the Data and Analysis Panel (Figure 5).

After an assessment taxon name is selected, the taxon options panel will display all taxonomic concepts listed by GBIF as being associated with the assessment taxon name selected (Figure 3B). The user is prompted to select all taxon concepts that apply and should be included in the assessment. Taxon concepts included will be highlighted in blue; click to deselect and the row will no longer be highlighted in blue. Once all relevant taxon concepts are selected, users can take one of two routes to move forward with selecting relevant GBIF data to be included in the assessment: to “Start the assessment with all records (up to 5000)” or to “Select specific datasets” (Figure 3B).

Clicking on “Start the assessment with all records (up to 5000)” will prepare RARECAT to load all available records from GBIF up to 5000 records; if the taxon concepts selected have records that total less than 5000 records, all records will be included; if the taxon concepts selected have records that total more than 5000 records, RARECAT will include up to 5000 records spread over the last 5 decades (see “Methodology” section for additional details). Clicking on “Select specific datasets” will open a new panel that displays the exact tallies of records across all GBIF datasets, divided in two separate categories: “Occurrences”, refers to all records that are not labeled as basisOfRecord “HUMAN OBSERVATION” in GBIF (primarily specimen data from biocollections);

“Human Observations” refers to all records labeled as basisOfRecord “HUMAN OBSERVATION” in GBIF (primarily data from citizen and community science projects like iNaturalist or eBird). From both categories, users can chose to “Include all datasets”, “Exclude all datasets”, or select individual datasets by clicking/unclicking them. In addition, a “Maximum records to include per dataset” text box allows users to select a limit number of records to be loaded per dataset which will be shared across all datasets in that category (Figure 4).


**Figure 3**: The taxon options panel enables users to select all relevant taxonomic concepts that should be included in the assessment. A) First, all taxon names in the NatureServe and GBIF taxonomic backbones associated with the name typed in the “Select assessment taxon” text box are returned. B) After a taxon is selected from either the NatureServe or GBIF taxonomy, all taxonomic concepts linked to it within the GBIF taxonomic backbone are returned; taxonomic concepts highlighted in blue will be included in the assessment.

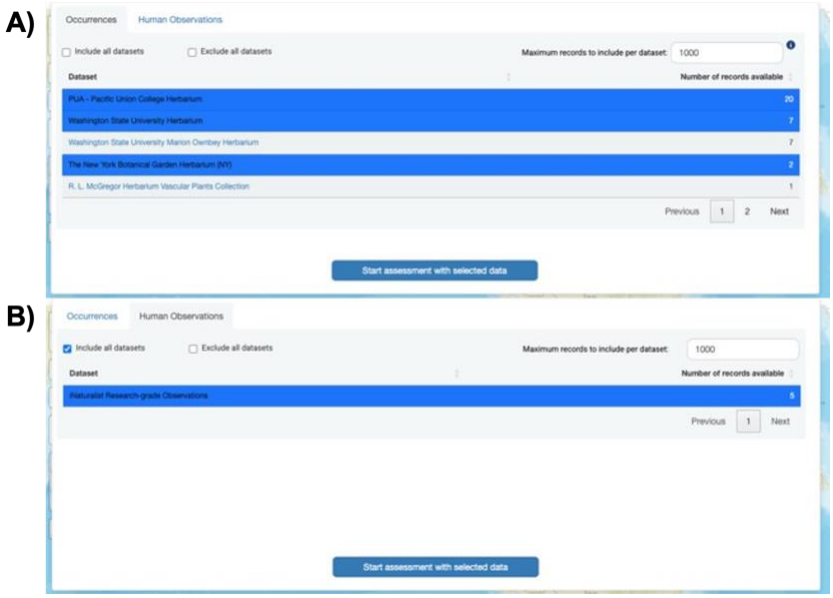



**Figure 4**: Clicking on “Select specific datasets” enables users to select records of the assessment taxon associated with given datasets of “Occurrences” (primarily data from biocollections) or “Human observations” (primarily data from citizen and community science projects).

### *2.3.1.2. Data and Analysis Panel*

The Data and Analysis panel summarizes the main inputs into and outputs from the ranking assessment (Figures 5 and 6). The top section of this panel, titled “Add assessment data”, enables users to add records for the taxon under assessment to the analysis from one or more of two sources: 1) open online data repositories including the Global Biodiversity Information Facility and associated sources queried via respective Application Programming Interfaces, 2) one or more comma-separated values (CSV) files added by the user. On the left of the “Add assessment data” section, the “Add records from CSV” toggle switch enables users to upload a one or more CSV files containing records for the assessment taxon (Figure 5). CSV files downloaded from widespread biodiversity data platforms and repositories such as GBIF, iNaturalist, SEINet, Biotics are all accepted (see Methodology for additional details on accepted input CSV formats). One the right of the “Add assessment data” section, the “Add records from GBIF” toggle switch enables users to add to the assessment the GBIF records which were selected from the taxon options panel, either using the general GBIF data request or by selecting specific datasets (Figure 4). The box below the “Add records from GBIF” toggle switch will display either the specific datasets and records selected or the number of records to be downloaded from GBIF overall (Figure 5).

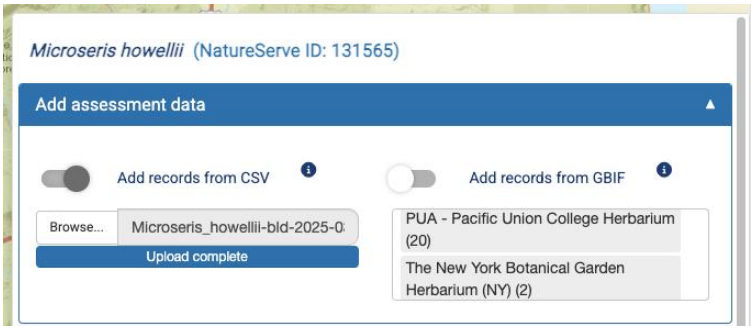



**Figure 5**: The top section of the “Data and Analysis” Panel displays the assessment taxon and the data inputs selected and/or loaded by the user. Distribution records for the target taxon can be added to an assessment using the Data and Analysis Panel via two main ways: by turning on the “Add GBIF records” toggle to import records from GBIF and associated online data sources, and/or by turning on the “Add records from CSV” and browsing to a local folder on the user’s machine.

Sliding the respective toggle switch will add the preloaded records to the analysis. Below, we discuss a third way for users to add records to the assessment, which involves drawing point records directly on the map using the “Draw a point record” tool (see Figure 7).

Once records are added to the assessment and map by sliding one or both toggle switches, the “Add assessment data” section is automatically collapsed and the “Rank factor calculations” section is displayed and opened. The “Rank factor calculations” section displays the number and source of records added to the analysis and enables the user to turn on the calculation for one or more of the three metrics of Rarity: Range Extent, Area of Occupancy, and Number of Occurrences (Figure 6). Sliding the respective toggle switch will enable the calculation of each metric: the metric’s value will appear to the right of the metric’s toggle and the metric’s spatial

representation will be displayed on the map (Figure 6). Symbology for the three Rarity metric spatial layers is as follows: Range Extent is displayed as a large gray polygon surrounding all mapped point records; Area of Occupancy is displayed as sets of blue square/rectangular grid cells of the chosen grid cell size; and the Number of Occurrences is displayed as red circular buffers around each point records, intersecting red circular buffers are considered part of the same "Occurrence". See Methodology section to learn more about how the three metrics of Rarity are quantified.

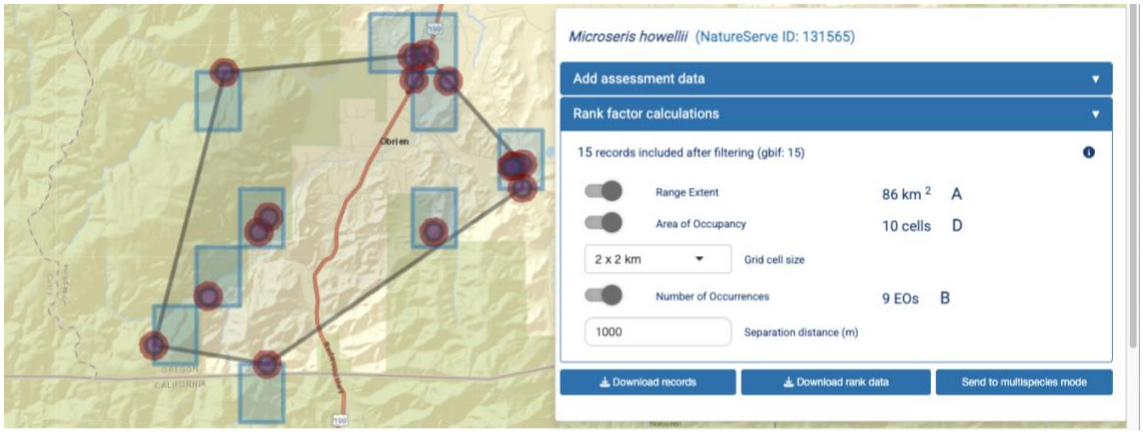



**Figure 6**: Once records are added to the map, the three metrics of Rarity – Range Extent, Area of Occupancy, and Number of Occurrences – can each be calculated and mapped by turning on the respective toggle in the "Rank factor calculations" section. Range Extent is displayed as a large gray polygon surrounding all mapped point records; Area of Occupancy is displayed as sets of blue square/rectangular grid cells of the chosen grid cell size; and the Number of Occurrences is displayed as red circular buffers around each point records, intersecting red circular buffers are considered part of the same "Occurrence".

The imported, selected, and filtered records can be exported from RARECAT using the "Download records" button; the Rarity metric values calculated, as well as additional information about the target taxon (derived from the NatureServe Explorer API), can be downloaded as a comma-separated values (CSV) file formatted as a NatureServe Rank Calculator using the "Download rank data" button (see Outputs section for more details). The "Send to multispecies mode" button can be used to send newly calculated outputs to Multispecies mode and we will discuss its use in the "RARECAT Multispecies" section of this document.

#### *2.3.1.3. Map*

The interactive map enables the user to explore, select, and filter records that have been added to the ranking assessment for the target taxon. Once added, records are displayed on the map as gray circle markers that scale with the map's zoom level. Current map functionality includes: zooming in/out buttons, restoring map to full extent (i.e. North America), changing basemap (toggle between ESRI World Street Map, ESRI World Terrain, ESRI World Imagery, and OpenStreetMap), drawing point record, drawing selection box, drawing selection polygon, and dragging drawn shapes. The buttons available to support map functionality are described in Figure 7.

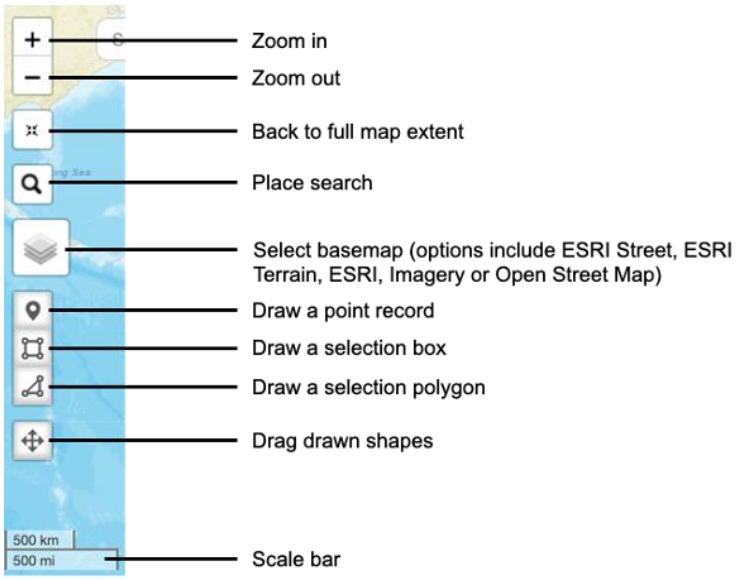



**Figure 7**: Functionality available within the interactive mapping interface.

Of particular note, the “Draw a point record” button enables users to create data or add to uploaded data by selecting particular locations on the map, while the “Draw a selection box” and “Draw a selection polygon” buttons enable users to select subsets of records to be examined more closely and/or removed (Figure 8).

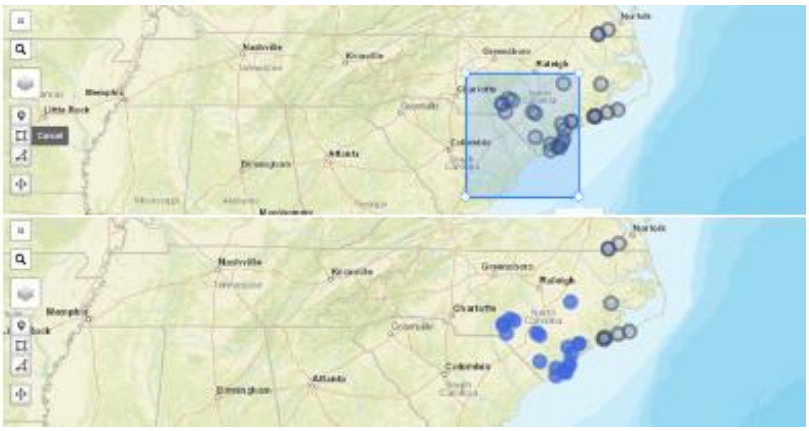


**Figure 8**: The “Draw a selection box” and “Draw a selection polygon” buttons enable the user to select a subset of mapped records to be sent to the interactive Records Table.

### *2.3.1.4. Filters Panel*

The Filters Panel in the bottom left of the Single Species page enables selecting filters that are applied wholesale to the Rank Assessment dataset, thus providing users with greater flexibility and efficiency in narrowing down the most desirable dataset to quantify Rarity metrics. Available filters are described in Figure 9. Find additional details in the Filters subsection of the Methodology section.


**Figure 9**: Description of filters available from the Filters panel.

### *2.3.1.5. Occurrences Table*

The interactive Occurrences Table enables the user to examine basic metadata on the subset of point records selected from the map using the The "Draw a selection box" and "Draw a selection polygon" buttons (see Figure 4). The basic metadata reported for each record are: 1) the unique ID or Key (which hyperlinks to the available online page for the record, if recognized); 2) the Scientific Name associated with the record; 3) the Source of the record in RARECAT (one of gbif, uploaded, or drawn); 4) the longitude and latitude of the record as unprojected coordinates; 5) the spatial uncertainty of longitude/latitude coordinates, if available (DarwinCore field coordinateUncertaintyInMeters); 6) the state, province, or NatureServe Subnation (i.e. the US State/territory or Canadian province) the record overlaps; 7) the country or NatureServe Nation (US or Canada) the record overlaps; 8) the year the record was

collected; 9) the month the record was collected; 10) the name of the dataset in which the record is included; 11) the Institution or Organization which owns or provides the record; 12) the type or basis of the record (i.e. specimen, occurrence, human observation, etc.); 13) the rank of the element occurrence (Biotics field EORANK) if the record has been uploaded from a Biotics input file; 13) the record source or url for additional details. The interactive Occurrences Table enables removing single records at a time (by clicking on the corresponding row) or all records by turning on the “Remove selected records from dataset” toggle. Selected records can be unselected by clicking the “Unselect records” button.

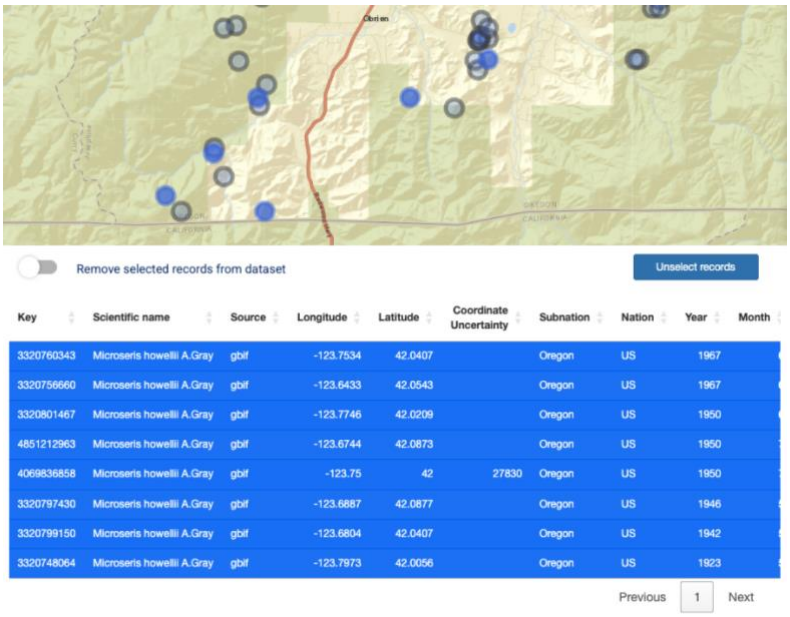



**Figure 10:** The interactive Occurrences Table reports basic metadata for the records selected from the map and enables removing single records at a time (by clicking on the corresponding row) or all records by turning on the “Remove selected records from dataset” toggle. Selected records can be unselected by clicking the “Unselect records” button.

### *2.3.1.6. Temporal Change Panel*

The Temporal Change panel provides a temporal breakdown and summary of the records included and rarity metrics calculated within the Rank Assessment. The starting dataset underlying the graphs and analyses in the Temporal Change Panel corresponds to the entire set of records being included on the Map, but additional temporal filters can be applied to these analyses specifically. The Temporal Change panel includes two tabs: the “Change over time” tab (Figure 11) and the “Temporal analysis” (Figure 12) tab. The “Change over time” tab includes two visualizations: 1) the “Records per year” barchart shows the number of records per unit time (Figure 11) across all years in included in the Rank Assessment time frame (as specified in the time frame filter of the Filters Panel, see Figure 7); 2) the “Rarity change by time period” barchart summarizes the percentage change in each metric of rarity across up to 3 multi-year time periods (Figure 12). The “Change over time” tab allows interactivity by enabling the user to update the time frame of the “Records per year” barchart using the date range slider

below the barchart, as well as updating the years shown in the year range text boxes for Time periods 1-3 in the “Rarity change by time period” barchart. The metric displayed in the “Rarity change by time period” barchart can be updated using the dropdown menu.

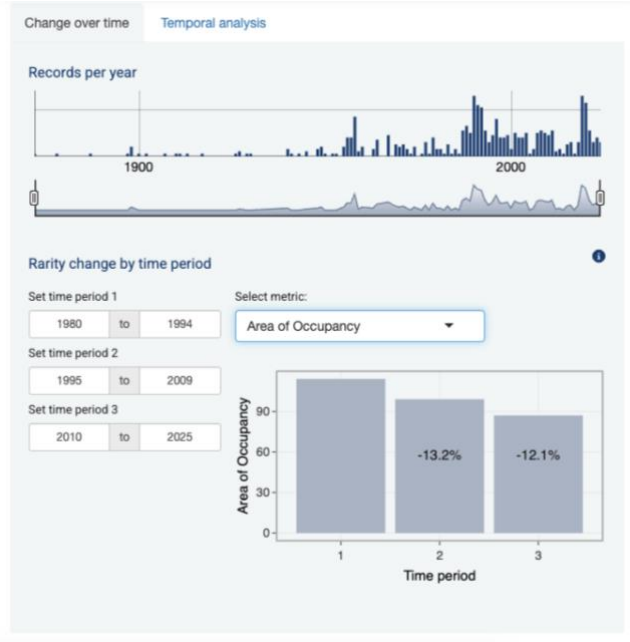



**Figure 11:** The ”Change over time” tab of the “Temporal Change” panel provides a breakdown of the number of records per unit time (“Records per year” visualization) and estimates of percentage change in each metric of rarity across up to 3 multi-year time periods (“Rarity change by time period” visualization).

The “Temporal analysis” tab allows users to generate bias-corrected estimates of the rate and modeled probability of observation for the assessment taxon that takes into account the intensity of recording across time and space for a broader reference taxon which contains the assessment taxon (such as the genus, family, or class the assessment taxon belongs to). To run this analysis, users are prompted to select the taxonomic resolution of the reference taxon (one of “genus”, “family”, “order”, “class”, “phylum”, “kingdom”) and the start year to determine how far back temporal estimates should stretch. Visualizations in the “Temporal analysis” tab summarize 1) the yearly number of observations for the target taxon, 2) the yearly number of observations of the reference taxon, 3) the yearly proportion of observations of the target taxon as a function of observations of the reference taxon, and 4) the modeled probability of observation of target taxon across years given the spatial and temporal rate of observation of the reference taxon. Visualizations 3 and 4 provide are different approaches to the same question: can we get closer to the true the ecological signal of yearly change in local abundance for the assessment taxon once we correct for variation in recording intensity across space and time? For more details in these analyses, see Methodology section.


**Figure 12:** The ”Temporal Analysis” tab of the “Temporal Change” panel enables users to generate bias-corrected estimates of the rate and probability of observation for the assessment taxon in light of observations for a broader reference taxon which the assessment taxon belongs to (see Methodology for more details on these analyses).

### *2.3.2. Workflow*

*Step 1. Select assessment geography*

Begin a rank assessment by selecting the geography of the assessment from the “Assessment menu” on the top left of the page. Global (G) Rank assessments are selected by default and no change is needed from the user if moving forward with a global assessment. For National (N) and Subnational (S) assessments, use the respective Dropdown menus to select the desired nation/subnation (see Figure 2).

*Step 2. Select assessment taxon*

Once the correct assessment geography is selected, proceed to select the assessment taxon using the “Select assessment taxon” search bar on the right of the “Assessment menu” on the top of the page: if the taxon being ranked has an existing recognized NatureServe Global Element ID (i.e. EGT ID) available on NatureServe Explorer or a recognized GBIF taxon available from GBIF.org, the user should search for and select it using the “Select assessment taxon” search bar (see Figure 3); else, the assessment can move forward without the selection of an existing NatureServe Element or GBIF taxon and the taxon being ranked will simply be “New taxon”.

*Step 3. Select and preload input data*

In the process of selecting the assessment taxon from the “Assessment menu”, work through the steps in the taxon options panel superimposed on the Map. This will enable you to preload all GBIF records (up to 5000) available for the assessment taxon or dive in and select/deselect the specific datasets including records of the assessment taxon (see Figure 4). The preloaded GBIF data selected will appear under the “Add GBIF records” toggle switch in the “Add

assessment data" section of the "Data and Analysis" panel (see Figure 5). Preload additional records from a comma-delimited values (CSV) file by browsing your local files using the "Browse" button under the "Add records from CSV" toggle switch in the "Add assessment data" section of the "Data and Analysis" panel (see Figure 5).

*Step 4. Load and map data*

Load and map data preloaded in the "Add assessment data" section of the "Data and Analysis" panel by sliding the respective toggle switch (see Figure 5). These data are now included in the Map for further assessment and filtering, and will be used in the assessment and calculation of rarity metrics and temporal summaries.

*Step 5. Explore, vet, and filter records*

After adding data to the Map, proceed to verify that the data loaded accurately represent the assessment taxon's distribution and range to the best of your knowledge. Unreliable, erroneous, or irrelevant records should be removed from the assessment. Removal of records can be done using by navigating to and/or selecting single or multiple records through the Map interface (see Figure 7 and Figure 8) or, by applying filters to the whole dataset using the "Filters" panel (see Figure 9). Verify and seek additional details on the records selected from the map using the "Occurrences table" (see Figure 10). In addition, the "Temporal Change" panel can also be used to verify the breakdown of records over years to further inform filtering over time.

*Step 6. Quantify rarity metrics*

Once you have narrowed down a suitable and accurate set of point records for the assessment taxon, you can now proceed to calculate one or more of the rarity metrics and associated Rank factors by sliding on or off the corresponding toggle switch from the "Data and Analysis" panel (see Figure 6), and choosing the desired parameters for each (e.g. grid cell size for "Area of Occupancy" or separation distance for "Number of occurrences").

*Step 7. Refine assessment inputs and rarity metrics*

Further updates to the underlying set of point records included in the assessment will lead to automatic updates to the value of each rarity metric included in the assessment.

*Step 8. Explore changes over time*

Use the "Change over time" tab of the "Temporal Change" panel to explore variation in the number of records per unit time for the assessment taxon and estimate percentage changes in rarity metrics over up to three user-specified time periods (see Figure 11). Use the "Temporal analysis" tab of the "Temporal Change" panel to generate bias-corrected estimates of the change in the rate and probability of observing the assessment taxon across years given the overall rate of observation of a broader reference taxon such as the family or order to which the assessment taxon belongs (see Figure 12).

*Step 9. Download outputs*

Once you are happy with the input data and have obtained reliable values of each rarity metric for the assessment taxon, you can proceed to download from the "Data and Analysis" panel: 1)

the filtered set of point records and associated metadata via the “Download records” button; 2) the formatted rank calculator Excel file including the values of the rarity metrics quantified, as well as additional information about the assessment, via the “Download rank data” button. See Outputs for additional details.

*Step 10. Clear map and start new assessment*
Once you have downloaded the all necessary outputs for the current assessment, start a new assessment by clicking on the “Start over” button to the top right of the “Assessment menu”. Alternatively, a new assessment may be started by refreshing the App in your browser window.

### *2.3.3. Outputs*

Two types of tabular output can be generated from RARECAT: 1) a comma-separated values (CSV) file with all filtered and vetted records included in the assessment (that is, all records added to the Map); 2) an Excel workbook (.xlsx) file with all Rank Factor values calculated for the assessment taxon, formatted as a NatureServe Conservation Rank Calculator file, plus additional information about the assessment in a separate tab.

#### *2.3.3.1. Analysis Records*

Records included in the assessment – that is the records added to the map at the time – can be downloaded by clicking the “Download records” button. This will prompt your browser to download a CSV file with the following naming convention: target taxon’s scientific name, “RARECAT_assessment_records”, present date. The downloaded CSV includes a table where each row represents a single record included in the assessment and the following fields are reported for each record: key (DarwinCore field occurrenceID), scientificName (DarwinCore field scientific Name), prov (the data source or “provenance” for the record, one of: gbif, uploaded, drawn), longitude (DarwinCore field decimalLongitude), latitude (DarwinCore field decimalLatitude), coordinateUncertaintyInMeters (DarwinCore field coordinateUncertaintyInMeters), stateProvince (DarwinCore field stateProvince), countryCode (DarwinCore field countryCode), year (DarwinCore field year), month (DarwinCore field month), datasetName (DarwinCore field datasetName), institutionCode (DarwinCore field institutionCode), basisOfRecord (DarwinCore field basisOfRecord), EORANK (Biotics field EO Rank), references (the underlying URL for the record), egt_uid (NatureServe Global Element Unique Identifier), el_code (NatureServe ELCODE), collectionCode (DarwinCore field collectionCode), recordedBy (DarwinCore field recordedBy), recordNumber (DarwinCore field recordNumber), accessRights (Dublin Core field accessRights), taxonKey (DarwinCore field taxonID), samplingProtocol (DarwinCore field samplingProtocol).

#### *2.3.3.2. Rank Calculator*

The results of the assessment can be downloaded as an Excel workbook (.xlsx) file with two tabs: a Rank Calculator tab and an Assessment Details tab. The Rank Calculator tab is

formatted as a NatureServe Conservation Rank Calculator file, including all factors necessary to conduct a Conservation Rank Assessment, with the rank factor values calculated by RARECAT included in the respective rarity metrics fields. For more information on the NatureServe Conservation Rank Calculator, please visit https://www.natureserve.org/products/conservation-rank-calculator. The Assessment Details tab summarizes most of the details of the assessment, including taxon concepts and IDs included, filters applied to the input data, and parameters of the calculations.

## *2.4. RARECAT Multispecies*

### *2.4.1. User Interface*

RARECAT "Multispecies" mode can be accessed via the main navigation menu and consists of a simple single-page interface with two components: the "Assessment Parameters" panel, and the "Assessment Results" table.

#### *2.4.1.1. Assessment Parameters Panel*

The Assessment Parameters panel is open by default when first entering Multispecies mode. The panel summarizes in one single view all of the main selections and parameters available to the user to run a new assessment for multiple taxa concurrently (Figure 13). Specifically, the Assessment Parameters panel is divided into 4 sections which mirror the parameters and filters available to users in Single Species mode: 1. Set assessment geography, 2. Set assessment taxa, 3. Set additional filters, and 4. Set calculation parameters.

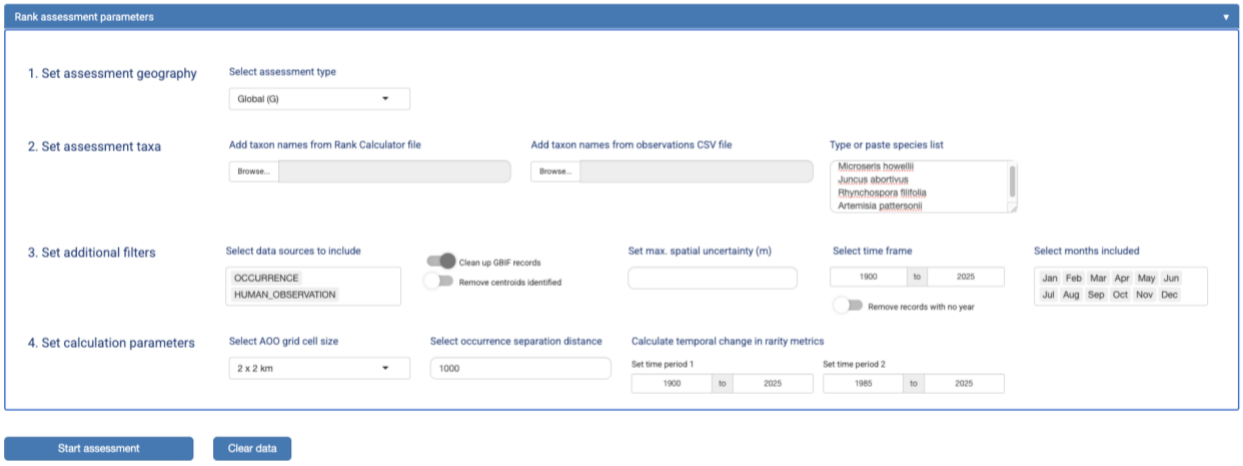



**Figure 13:** The Assessment Parameters panel summarizes in one single view all of the main selections available to the user to run a new assessment for multiple taxa concurrently, including assessment geography and taxa, data inputs and filters, and rank factor calculation parameters.

In the first section, users are prompted to select the assessment geography in a manner akin to Single Species mode, via one or more of three dropdown menus to "Select assessment type, "Select assessment nation", and "Select assessment subnation". The second and third dropdown menus only appear if selecting National or Subnational assessment types.

In the second section, users are prompted to select the list of taxa to be assessed concurrently. There are three ways to upload a list of taxa in Multispecies mode: 1) via an existing Rank Calculator comma-separated values (CSV) file, where each row represents a different taxon to be assessed; 2) via one or more CSV files including observations of the multiple taxa to be assessed formatted similarly to CSV files used to upload observations in Single Species mode; 3) by typing and/or pasting a list of names from a spreadsheet or other source of text directly into a text box. See Methodology section for more details about acceptable formats for uploads.

In the third section, users can select additional filters to be applied broadly across all taxa in the assessment. These filters are a subset of the filters available from the Filters panel in Single Species mode, including filters to clean up GBIF records (GBIF filter and Centroids filter), set the maximum spatial uncertainty of records, the time frame and collection months of records to be included. For more information on how these filters work, see the Filters subsection of the Methodology section. An important distinction with Single Species modes is that users in Multispecies mode only have the option to select GBIF records from one or both of two sets of data sources: "OCCURRENCE" of "HUMAN_OBSERVATION". These data sources each correspond to the respective tab users are able to select in Single Species mode when choosing to "Select specific datasets" for the assessment taxon (see Figure 4).

In the fourth and last section, users can set the parameters for the calculation of rarity and change in rarity to be used broadly across all taxa in the assessment. Specifically, users can set the resolution of grid cells to be used to quantify Area of Occupancy and the separation distance to be used to define Element Occurrences. In addition, users can set the time frame of two time periods which will be used to calculate percentage changes in rarity metrics between time periods.

After inputting the assessment geography and a list of taxa to be assessed, as well as updating any of the data inputs, filters, or other parameters, the Multispecies assessment can be kicked off by clicking the "Start assessment" button.

#### *2.4.1.2. Assessment Results Table*

The Assessment Results table summarizes the main outputs from the Multispecies assessment, with each row in the table corresponding to each assessment taxon. The Assessment Results table is comprised of 12 columns (Figure 14): the Scientific Name (Assessment) column lists the names of the taxa included in the multispecies assessment; the Number of Records Included column reports the total number of point records used to calculate rank factor values for each taxon; the Range Extent (New) column reports the value and associated rank factor letter for the Range Extent metric newly calculated by RARECAT; the Range Extent (Previous) column includes a flag comparing the newly calculated Range Extent metric to the previously calculated Range Extent metric available via either NatureServe Explorer or a user-uploaded Rank Calculator file; the AOO (New) column reports the value and associated rank factor letter for the Area of Occupancy metric newly calculated by RARECAT; the AOO (Previous) column includes a flag comparing the newly calculated Area of Occupancy metric to the previously calculated

Area of Occupancy metric available via either NatureServe Explorer or a user-uploaded Rank Calculator file; the Occurrence Count (New) column reports the value and associated rank factor letter for the Occurrence Count metric newly calculated by RARECAT; Occurrence Count (Previous) column includes a flag comparing the newly calculated Occurrence Count metric to the previously calculated Occurrence Count metric available via either NatureServe Explorer or a user-uploaded Rank Calculator file; the Range Extent Change, AOO Change, and Occurrence Count Change columns report the estimates of percentage change in the respective rarity metric between the two time periods selected by users in the Assessment Parameters panel; the Reviewed column provides a flag for whether or not the corresponding taxon's rank assessment has been reviewed by the user in Single Species mode.

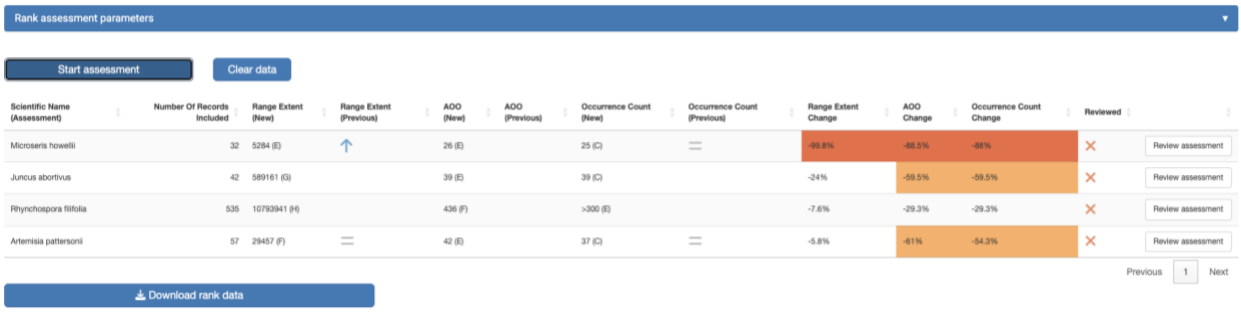



**Figure 14:** The Assessment Results table summarizes the main outputs from the Multispecies assessment, with each row in the table corresponding to each assessment taxon.

The symbols and colors displayed in the Assessment Results table provide visual cues to highlight potential concerns with either the validity of the newly calculated factors or causes to consider additional information in a given taxon's rank assessment. In the Range Extent (Previous), AOO (Previous), and Occurrence Count (Previous) columns, the symbology is as follows: a grey equal sign indicates that the newly calculated value matches the previously calculated value from either NatureServe Explorer or a user-uploaded Rank Calculator file; a blue upward arrow indicates that the newly calculated value is higher than the previously calculated value, thus suggesting a potential change towards lower rarity; a red downward arrow indicates that the newly calculated value is lower than the previously calculated value, thus suggesting a potential increase in rarity and cause for concern. In the Range Extent Change, AOO Change, and Occurrence Count Change columns, more negative percentage change values are highlighted with warmer colors: -30% to -50% in yellow, -51% to -70% in orange, below -71% in red. In the Reviewed column, an X mark indicates rows that have not been reviewed by the user in Single Species mode while a check mark indicates rows that have been reviewed by the user in Single Species mode.

### *2.4.2. Workflow*

#### *Step 1. Select assessment geography*

Begin a rank assessment by selecting the geography of the assessment from the top section of the "Assessment Parameters". Global (G) Rank assessments are selected by default and no change is needed from the user if moving forward with a global assessment. Selecting National

(N) and Subnational (S) assessments will reveal the respective dropdowns which you can use to select the assessment nation(s) and subnation(s).

*Step 2. Select assessment taxa*

Once the correct assessment geography is selected, proceed to input a list of up to 100 taxa to be assessed concurrently from the second section of the "Assessment Parameters". The list of taxa can be entered via one of three ways: 1) using an existing Rank Calculator comma-separated values (CSV) file; 2) using one or more CSV files including observations of the multiple taxa to be assessed; 3) by typing and/or pasting it directly into a text box.

*Step 3. Select additional filters and parameters*

Use the next two sections of the "Assessment Parameters" panel to select spatial and temporal filters, as well as rarity metric calculation parameters, to be applied broadly across all taxa in the assessment.

*Step 4. Start assessment*

Click the "Start assessment" button to start the multispecies assessment.

*Step 5. Review results table*

Once the Multispecies assessment has finished running, RARECAT will display the "Review results" table. Each row in the table corresponds to a taxon from the multispecies list.

*Step 6: Send data to Single Species mode*

To review a specific taxon, use the "Review" button from the corresponding table row in the "Review results" table to send the taxon's assessment data inputs, filters, and parameters to Single Species mode. Use the full functionality of Single Species mode to review and update the taxon's assessment inputs and outputs.

*Step 7: Send data back Multispecies mode*

Once you have reviewed and updated the taxon's assessment inputs and outputs to the desired degree using Single Species mode, you can send the resulting parameters and outputs back to Multispecies mode using the "Send to multispecies mode" button in the "Data and Analysis" panel.

*Step 8: Download output*

To download data summarizing the results of the multispecies assessment, click on the "Download rank data" button. Results are formatted as a rank calculator Excel file where every row corresponds to a taxon in the multispecies assessment list.

*Step 9: Clear data and start fresh*

To start a new multispecies assessment click the "Clear data" button. Alternatively, a new assessment may be started by refreshing the App in your browser window.

### *2.4.3. Outputs*

#### *2.4.3.1. Rank Calculator*

The results of the Multispecies assessment can be downloaded as an Excel workbook (.xlsx) file with two tabs: a Rank Calculator tab and an Assessment Details tab. The Rank Calculator tab is formatted as a NatureServe Conservation Rank Calculator file, including all factors necessary to conduct a Conservation Rank Assessment, with the rank factor values calculated by RARECAT included in the respective rarity metrics fields. For more information on the NatureServe Conservation Rank Calculator, please visit https://www.natureserve.org/products/conservation-rank-calculator. The Assessment Details tab summarizes most of the details of the assessment, including taxon concepts and IDs included, filters applied to the input data, and parameters of the calculations. In both tabs, each row of the represents a different taxon in the Multispecies assessment.

## 3. Methodology

### *3.1. Taxonomy*

RARECAT recognizes taxon concepts based on two sources: the NatureServe Taxonomic Backbone (accessed via the NatureServe Explorer API) and the GBIF Taxonomic Backbone (accessed via the GBIF API).

In Single Species mode, entering some text in the “Select assessment taxon” search bar in the “Assessment menu” returns taxon suggestions from both NatureServe and GBIF in the table below. More information on each taxon concept displayed can be accessed by clicking on the corresponding taxon ID (which provides a link to the relevant taxon page on NatureServe Explorer or gbif.org). The user is prompted to select a target taxon for the assessment by clicking on the relevant table row, which becomes highlighted in blue. Once a target taxon is selected for assessment, its name and ID appear on the top of the Data and Analysis panel: this will be the overall taxon selected for the current rank assessment (or Assessment Scientific Name). However, since a given taxon concept may have several synonyms or related infra-species listed by both NatureServe and GBIF, RARECAT provides an explicit breakdown of all records available from GBIF across all scientific names connected with the target taxon selected. This information is provided in a second table below the “Assessment menu”. Specifically, the options displayed in this table correspond to all scientific names provided by the GBIF API as synonyms or infra-species of either 1) the GBIF taxon concept selected as target taxon, 2) the NatureServe element selected as target taxon, as well as all of the synonyms provided for it by the NatureServe Explorer API. The user is prompted to select all relevant scientific names related to the target taxon for which records ought to be included in the assessment.

In Multispecies mode, RARECAT a process akin to Single Species mode is carried out implicitly, whereby each taxon entered by the user in the multispecies set is matched to both the NatureServe and GBIF taxonomies, and all taxonomic concepts linked to it by either source are

included in the full set of taxonomic concepts which contribute point records to the assessment (these are labeled as Source Scientific Names).

### *3.2. Importing Data*

Data can be imported into RARECAT from one or both of the following sources: 1) from the Global Biodiversity Information Facility (GBIF); or 2) via a comma-delimited values (CSV) file uploaded by the user from their local directory.

GBIF records are imported into RARECAT using the GBIF API via the R packages rgbif and SPOCC. In Single Species mode, users can select which data are imported from GBIF using the "Select assessment taxon" search bar in the "Assessment menu": entering text in the search bar will reveal the taxon options panel and provide users with the ability to select taxon concepts and respective records available from various GBIF data sources. In Multispecies mode, users can use the "Select data sources to include" box to choose to include either or both all relevant "OCCURRENCE" data (primarily biocollection data) or all relevant "HUMAN_OBSERVATION" data (primarily citizen and community science projects data) across all taxa in the multispecies assessment.

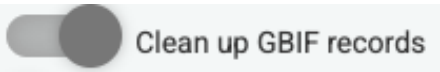
One or more CSV files can be uploaded in both Single Species and Multispecies modes to supplement assessment datasets. RARECAT accepts various CSV file formats that include at least a scientific name column (e.g. DarwinCore field scientific Name), a longitude column (DarwinCore field decimalLongitude), and a latitude column (DarwinCore field decimalLatitude). Additional fields that RARECAT will recognize are coordinateUncertaintyInMeters (DarwinCore field coordinateUncertaintyInMeters), stateProvince (DarwinCore field stateProvince), countryCode (DarwinCore field countryCode), year (DarwinCore field year), month (DarwinCore field month), datasetName (DarwinCore field datasetName), institutionCode (DarwinCore field institutionCode), basisOfRecord (DarwinCore field basisOfRecord). RARECAT has some built in flexibility in the naming and inclusion of these fields: files downloaded from GBIF, Biotics, iNaturalist, SEINet, and similar sources are readily accepted.

### *3.3. Mapping*

Mapping in RARECAT is done using the Leaflet package for R (Cheng et al. 2024). The Leaflet package expects all point data to be specified in latitude and longitude using WGS 84 (EPSG:4326). By default, when displaying this data, Leaflet projects everything to Web Mercator (also known as WGS 84/Pseudo-Mercator) projection (EPSG:3857). All records loaded directly from GBIF are always in WGS 84. Please note that RARECAT assumes that any longitude and latitude coordinate values uploaded by the user from a CSV file are in WGS 84. No spatial transformations are implemented in RARECAT prior to mapping records for exploration; errors are likely to emerge from the upload of records with longitude and latitude values not in WGS 84.

### *3.4. Filters*

#### *3.4.1. GBIF Filter*

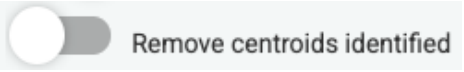



By default, the following filters are applied to GBIF records for the target taxon imported via the “Add GBIF records” toggle:

- Records have coordinates
- Records have some value at least for fields latitude, longitude, and basisOfRecord
- institutionCode does not equal “iNaturalist” (iNaturalist records are loaded independently from the iNaturalist API)
- latitude and longitude are both not 0
- basisOfRecord does not equal FOSSIL_SPECIMEN, LIVING_SPECIMEN, MATERIAL_SAMPLE
- occurrenceStatus equals PRESENT
- coordinateUncertaintyInMeters does not equal 999, or 9999
- samplingProtocol does not equal "from a cultivated plant of known (indirect) wild origin", or "grown"
- Only the first record is kept if there are duplicate records that have the exact same combination of latitude, longitude, speciesKey, and datasetKey

These default filters can be cleared by turning the “Clean up GBIF records” toggle off and reloading all GBIF records. When the “Clean up GBIF records” toggle is off, all available GBIF records are imported so long as they are have both latitude and longitude coordinates and both do not equal 0.

#### *3.4.2. Centroids Filter*



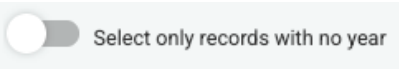

Turning on the “Remove centroids identified” toggle will filter out all records identified as corresponding to jurisdiction centroids (e.g. city, county, state, or country centroids). Specifically, RARECAT does this by applying the following filters to GBIF records:

- Remove records that have georeferenceRemarks including the words “centroid”, “Centroid”, or “CENTROID”
- Removing records where uncertainty equals 301 or 3036, which are values known to be used to identify various kinds of jurisdiction centroids

#### *3.4.3. Time Frame Filter*

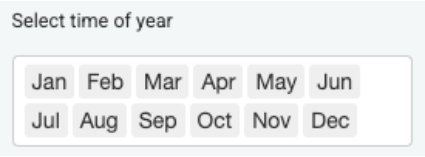

A time frame filter can be applied to exclude records from the rank assessment. This can be done by entering the start and end years for the desired analysis time frame in the respective text boxes under the “Set time frame of records” section of the “Filters” panel. It is important to note that records which do not have a value available for year or date will remain included in the analysis regardless of the time frame specified.

### *3.4.4. No Year Filter*



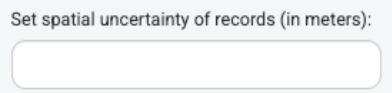

All records with a missing value for Year can be selected at once by turning on the “Select only records with no year” toggle. This enables the user to further assess all records which do not have a year value, and could therefore be erroneous or misleading, by using the Map and the Records Table.

### *3.4.5. Months filter*




A Months filter can be applied to constrain the assessment dataset to records collected in specific months of the year. This can be done using the “Select time of year” filters in Single Species and Multispecies modes. By default, all months will be selected. To deselect a month, click on it and hit the “delete” key.

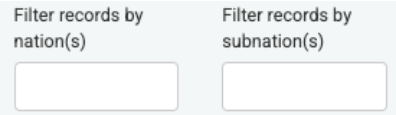
### *3.4.6. Spatial Uncertainty Filter*




Records can be excluded from the rank assessment based on a specified level of uncertainty in the precision of spatial coordinates. This can be done by entering the desired level of coordinate uncertainty in meters above which records should be excluded. It is important to note that

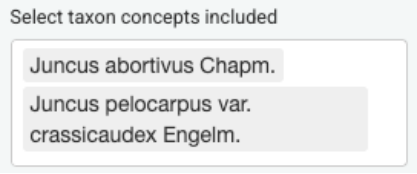
records which do not have a value available for spatial uncertainty will remain included in the analysis regardless of the coordinate uncertainty level specified.

### *3.4.7. Nation and Subnation Filters*



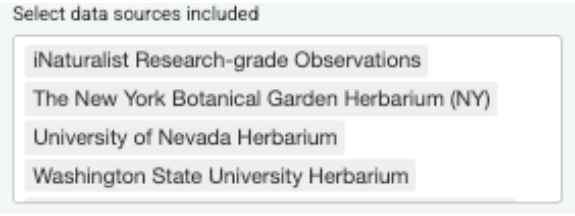

Rank assessments can be limited to a given Nation (either the United States of America or Canada) or Subnation (US state/territory or Canadian province/territory; https://www.natureserve.org/ns-network-directory) by selecting the desired geographical unit using the “Filter records by nation(s)” and “Filter records by subnation(s)” dropdown menus in the “Filters” section. Note that only nations and subnations corresponding to records added to the map will be available for selection in the dropdown menus.

### *3.4.8. Taxon Filter*




Rank assessments can be limited to a subset of taxon concepts to be included within the overall assessment taxon. All taxon concepts are included by default. One or more taxon concepts can be excluded by deselecting them from the “Select taxon concepts included” box in Single Species mode. To deselect a taxon, click on it and hit the “delete” key.

### *3.4.9. Data Source Filter*

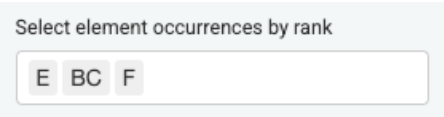



Rank assessments can be limited to particular GBIF datasets. All GBIF datasets with relevant records or all GBIF datasets explicitly selected by the user in Single Species mode are selected by default. To exclude records from given datasets from the assessment, deselect one or more of the data sources from the “Select data sources included” box. To deselect a data source, click on it and hit the “delete” key.

### *3.4.10. Data Type Filter*


Rank assessments can be limited to particular GBIF data types (DarwinCore field basisOfRecord). Records from the following GBIF data types are included by default: “PRESERVED_SPECIMEN”, “HUMAN_OBSERVATION”, “OBSERVATION”, “MACHINE_OBSERVATION”, and “OCCURRENCE”. Records corresponding to given data types can be excluded from the assessment using the “Select record types included” box. To deselect a data type, click on it and hit the “delete” key.

### *3.4.11. Element Occurrence Rank Filter*




By default, this filter has no effect on the assessment if records are included from sources outside the NatureServe Network. However, if a CSV file is uploaded by the user with records stemming from the NatureServe Network, such as from Biotics, the file may contain an EO Rank field with populated Element Occurrence Rank values. In this case, RARECAT will recognize this field and enable filtering those particular records by EO Rank. All other records, such as those coming in from GBIF or other sources will remain unaffected by the filter. For more information on the EO Rank field and relevant values see

## *3.5. Rarity Metrics*

### *3.5.1. Range Extent*

Range Extent is one of the spatial metrics used in NatureServe Rank Assessments to quantify the rarity and endangerment of the species assessed. NatureServe’s definition of this metric follows the IUCN Red List of Threatened Species definition for Extent Of Occurrence (EOO): the area within the shortest continuous line that can be drawn to include all known, inferred, or projected locations of a species, excluding vagrancy. In practice, Range Extent is therefore calculated by drawing a minimum convex polygon around a minimum of three point records.

RARECAT calculates and maps Range Extent using the following steps:

1. Unprojected longitude and latitude coordinates in WGS 84 are reprojected to a suitable area projection, that is either Cylindrical Equal-Area projection for latitudes below 70

degrees or Lambert Azimuthal Equal-Area projection for polar areas (latitudes above 70 degrees).

2. A minimum convex polygon is drawn around all point records included in the assessment (provided there are at least 3 point records; a value of NA will be returned otherwise).
3. The area of the minimum convex polygon is calculated and returned in square kilometers in the “Data and Analysis Panel” to the right of the “Range Extent” toggle.
4. A minimum convex polygon is also calculated from the unprojected longitude and latitude coordinates (in WGS 84) for the purpose of mapping and is added to the map as a grey polygon.

Please note that the mapped Range Extent polygon and the Range Extent polygon used to calculate the Range Extent value are not equivalent as they are calculated over unprojected and projected space, respectively. The Range Extent polygon that is mapped could potentially be distorted or fail to encapsulate all point records mapped, especially for very large-ranging and/or circumpolar taxa. Irrespective of potential web mapping issues, the returned value for Range Extent represents the best approximation of the area of the minimum convex polygon in an equal-area projection. Additionally, the spatial configuration of some point records datasets may lead to issues in the calculation of a minimum convex polygon and, therefore, a Range Extent value; this is particularly the case for taxa where few outliers are geographically distant from the majority of point records.

### *3.5.2. Area of Occupancy*

Area of Occupancy (AOO) is one of the spatial metrics used in NatureServe Rank Assessments to quantify the rarity and endangerment of the species assessed. NatureServe’s definition of this metric follows the IUCN Red List of Threatened Species definition: a statistic to represent the area of suitable habitat that a taxon occupies within its range extent. In practice, AOO is calculated by placing a grid over all selected taxon records within a user-defined area, with a default grid size of 2 km (alternatively, grid size can be set at 1km).

RARECAT calculates and maps Area of Occupancy using the following steps:

1. Unprojected longitude and latitude coordinates in WGS 84 are reprojected to Universal Transverse Mercator - a suitable area projection. Please note that all coordinates are reprojected to the UTM zone that corresponds to the minimum longitude value in the dataset.
2. The number of unique equal-area grid cells overlapped by the reprojected point records is counted, given the grid cell size specified by the user (2 x 2 km, by default, or 1 x 1 km).
3. Area of Occupancy is then calculated by multiplying the number of cells by 4 to obtain a value for area in kilometers squared.
4. For the purposes of mapping and visualization, a grid of the approximate user-specified cell size is created in WGS 84, cells that overlap points in WGS 84 are identified and added to the Map as blue square/rectangular cells. Note that because the web map is in

unprojected Web Mercator all cells appear to be of the same size when mapped but will actually be of different areas depending on the longitude on which they fall. This may lead to slight differences in the number of AOO cells mapped versus the accurate AOO metric value displayed in the “Data and Analysis” panel.

### *3.5.3. Number of Occurrences*

Number of Occurrences is one of the spatial metrics used in NatureServe Rank Assessments to quantify the rarity and endangerment of the species assessed. Based on the NatureServe methodology, an Occurrence (or Element Occurrence) is an area of land and/or water in which a species is, or was, present and should have practical conservation value for the Element as evidenced by potential continued (or historical) presence and/or regular recurrence at a given location. Therefore, for the purposes of ranking, it is necessary to derive occurrences from the point records at hand. In order to do so, individual point records within a certain “separation distance” of each other are grouped as part of the same Element Occurrence (EO); the occurrence may reflect a portion of a population (e.g., long distance dispersers) or a group of nearby populations (e.g., metapopulation).

RARECAT calculates and maps Number of Occurrences using the following steps:

1. Point records in WGS 84 are buffered by half of the specified separation distance (i.e. 500 meters if the user-specified separation distance is 1000 meters), such that an intersection between buffered point polygons would indicate a distance equal to or less than the specified separation distance between the unbuffered point records.
2. Buffered point polygons that cross each other are spatially dissolved into the same occurrence; note that two buffered point polygons will become part of the same occurrence even if they do not directly overlap but they both share overlap with a third buffered point polygon.
3. The number of individual spatially dissolved polygons is counted.
4. For the purposes of mapping, undissolved buffered point polygons are added to the map as red polygons.

### *3.5.4. Change in Rarity Metrics Across Time Periods*

RARECAT provides estimates of change in all three rarity metrics across time periods in both Single Species and Multispecies modes. In Single Species mode, the ”Change over time” tab of the “Temporal Change” panel provides estimates of change in the Number of Records, Range Extent, Area of Occupancy, and Number of Occurrences across two or three time periods, with the start year and end year of each time period being set by the user. RARECAT runs new calculations of each rarity metric in each time period and also calculates the percentage change in each metric between each time periods and the previous.

### *3.5.5. Temporal Analysis*

Point records derived from the Global Biodiversity Information Facility (GBIF) and similar sources are primarily opportunistic in nature and not derived from structured surveys. This means that variation in the number of records for a taxon over space and time is not simply the result of its distribution and ecology but also results from variation in the intensity of recording and behavior of recorders over space and time. As a result, deriving ecological signals from opportunistic and unstructured data taken at face value is subject to significant uncertainty. Fortunately, a lot of research has gone into potential ways to control for variation in the intensity of recording when deriving metrics of change from opportunistic biodiversity data (Young et al. 2017; Rapacciuolo et al. 2021; Johnston et al. 2022)

RARECAT includes a built-in module to derive estimates of yearly rate of recording and probability of detection for the assessment taxon that take into account the intensity of recording over years across the assessment geography (following methods in Young et al. 2017 and Rapacciuolo et al. 2021). Users can find this functionality in the bottom right "Temporal change" panel, by clicking on the "Temporal Analysis" tab. Two parameters are needed to run the analysis: to select a "Reference taxon" and to select a "Start year". The Reference taxon represents the broader taxon related to the assessment taxon (say the genus or family to which the assessment taxon belongs) to be used in order to understand large-scale patterns in the intensity of recording over years and across the assessment geography. The choice of a broader taxon is necessary so that sufficient data of relevance to the assessment taxon can be used to understand overall variation in recording intensity; the default is the "genus" in which the assessment taxon is included. Choice of a "Start year" will determine how far back in time this analysis will run, defaulting to the year 1980.

The temporal analysis can be started by clicking on the "Calculate temporal change" button and, if successful, should result in four plots (Figure 15):

- Plot 1 summarizes the number of records for the assessment taxon from the selected start year to the current year.
- Plot 2 summarizes the number of records for the selected reference taxon from the chosen start year to the current year.
- Plot 3 shows the proportion of records of the assessment taxon relative to the total number of records of the reference taxon in each year. This plot aims to answer the following question: given a certain number of records collected for the reference taxon, how many records of the assessment taxon were made? Years where a higher number of records of the reference taxon were collected but very few or no records of the assessment taxon were collected may imply a higher likelihood that the assessment taxon existed at lower local abundances. Plot 3 estimates provide a more useful approximation of the ecological signal of change in local abundance for the assessment taxon, once accounted for the overall pattern of recording for the broader reference taxon.
- Plot 4 displays the results of a linear mixed-effects model (LMER) which aims to model the probability of detecting the assessment taxon each year across the assessment geography given the number of records collected and the number of taxa detected during those collections. The response variable in this model is the detection-

nondetection of the assessment taxon in a given year in a given area; the predictor variables in the model are the year, the number of records collected, the number of taxa detected, and the given area. The model is run only across areas of the assessment geography relevant to the assessment taxon; specifically, the 2 x 2 (or 1 x 1) $km^2$ grid cells derived from the calculation of Area of Occupancy. The pattern in Plot 4 provides an alternative estimate of change in the local abundance of the assessment taxon over years across the assessment geography to be compared with Plot 3.

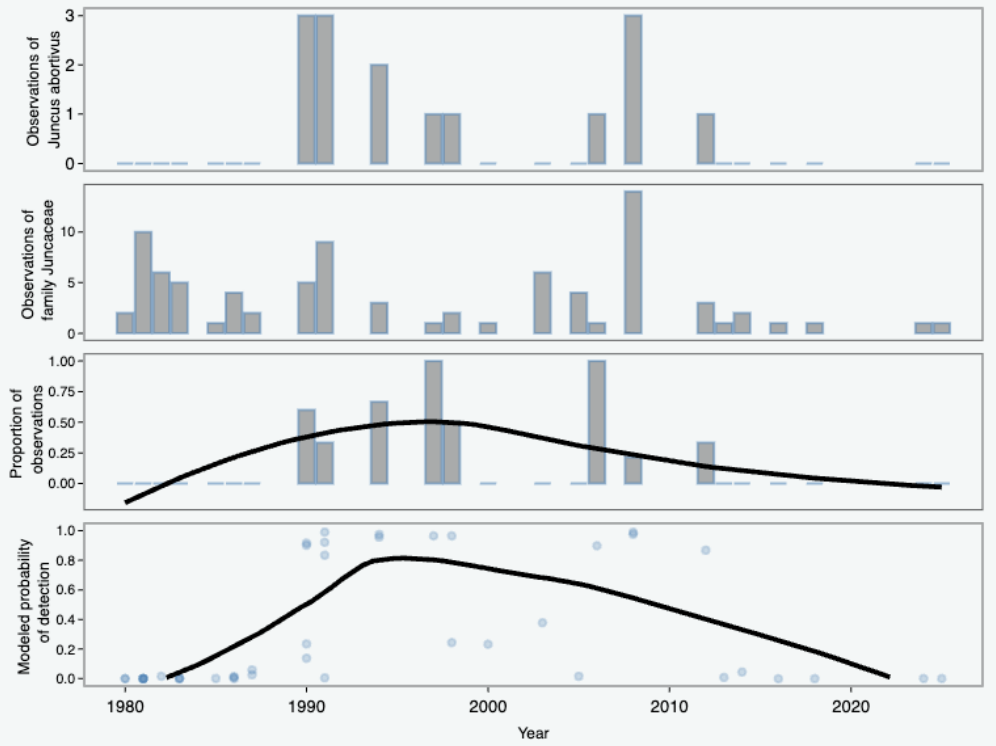



**Figure 15**: The output of a “Temporal Analysis” generating estimates of rate or probability of observation of the assessment taxon adjusted by the intensity of recording over years across the assessment geography.