OpenSites: exploring their abilities

Tab 2, part 2

Specifying, using text, *which* records to locate from the selected collections

This is the page reached after selecting the collections you wish to include in your search. It is for text-based searches. Another way to search for records is Map-based. The link for map-based searches is tab 3 of the home page.

There are five sets of criteria for selecting records via a text-based search: Taxonomic (Arrows 2-4), Locality (Arrow 5), Latitude and longitude (Arrows 6-9), Collector information (Arrows 10-12) and Specimen Criteria (Arrows 13-14). Each of these is discussed below the figure.

One selection criterion must be entered, but entering more than one is fine.

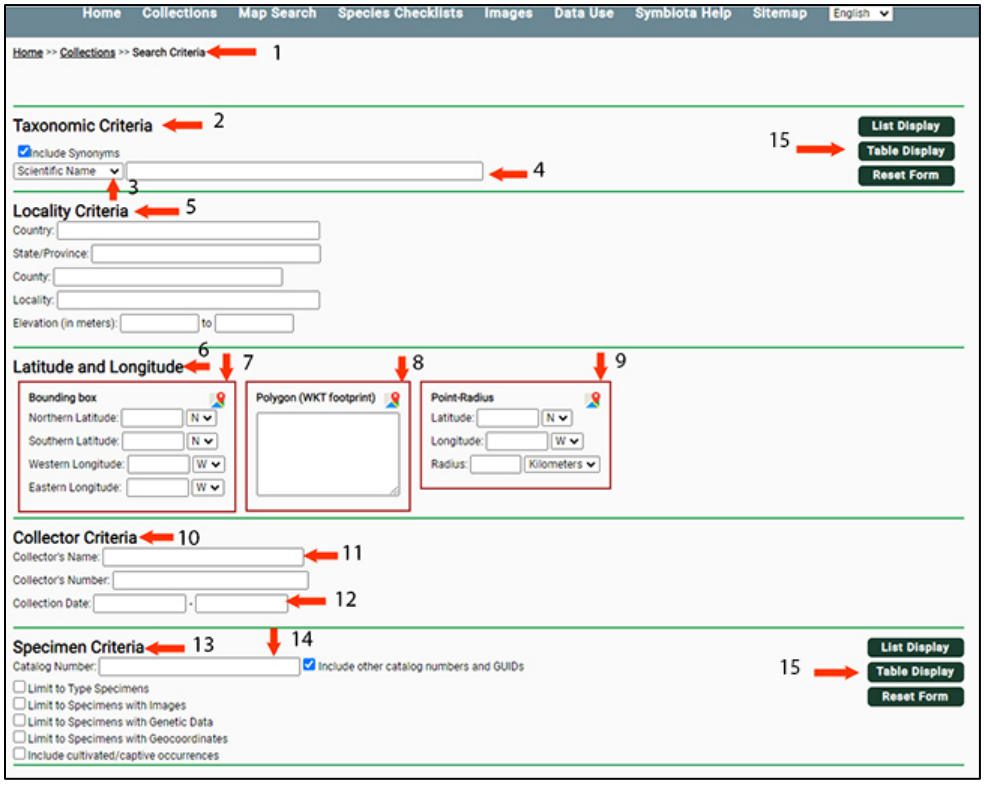


Fig. 3. Screen shot of page for text-based searches

Arrow 1

This shows how you reached this page. Many web sites provide such information at the top of their pages. On this page, it offers an easy way to go back to the page for selecting the collections to be searched – just click “Collections” in this line shown.

Taxonomic criteria (Arrows 2 – 4)

By default, the page is set to search for genera or species, but searching for other ranks is possible. To see the options, click the down arrow above arrow 3. The options are "scientific name" (should be "genus or lower rank"), Family, Taxonomic Group, and "Common name". It is also possible to search for several names so long as they are all at the same rank. The "Common name" option does not work in OpenHerbarium because I have entered very few common names.

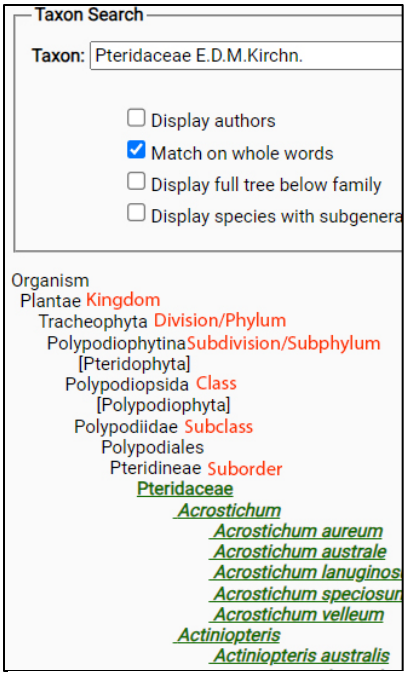
Scientific name: To search for records of families, genera and lower ranks, enter the name involved in the space by arrow 4. The name will auto complete *if* it is in the site's taxonomic tables. If it does not auto complete, first check your spelling. The abbreviation for form is "f.", for variety "var.", and for subspecies "subsp.". If the problem is not a spelling problem, email me (Mary Barkworth) with a request to add the name you are looking for. In your request, please tell me where you found the name.

If looking for a subfamily, enter its name, for example "Faboideae"; do not enter it as "Fabaceae subf. Faboideae" but for ranks at the generic level and below on must enter the including name, name of lower rank, epithet of lower rank, for example as "Pteridium aquilinum subsp. latiusculum". In this example, the including name is *Pteridium aquilinum*, the rank of the taxa required is subsp. [aka subpecies] and the epithet of the lower rank is "*latiusculum*". The reason for this is that names of ranks between family and genus have obligatory endings ("-oideae", for example, means subfamily when dealing with plants); names of ranks below species do not have any designated endings so one must tell people and computers the name of the rank ine is looking for.

Family: Start entering the name of the family that interests you in the box by Arrow 4. As with the names of genera and lower ranks, your entry will auto complete if it is in the taxonomic table. If not, follow the same procedure as given above: first check your spelling and, if that is not the problem, ask me (Mary Barkworth), to add the name. In your request, please tell me where you found the name.

Taxonomic group refers to ranks above the rank of family. Again, the name will autocomplete, but make sure you select the ending that will find the rank that interests you. For example, if looking for the order than includes the *Liliaceae,* select *Liliales*, not *Liliidae.*

Because there is less agreement about the groups to recognize at higher taxonomic ranks, it may be necessary to use the Taxonomy viewer option to see what name is being used for the taxonomic group you are looking for. To do this, open the Sitemap (the far right choice on the Home Page) and scroll down to "Taxonomic Tree Viewer". Click on it and enter the name of a genus or family that is in the group that interests you. For example, if you want to know what class the *Pteridaceae* is included in, enter "Pteridaceae" in the tree viewer. The result will be as shown to the right. I added the names of the higher ranks because I tend to forget which endings go with which of the higher level ranks.



Taxonomic tree viewer for *Pteridaceae*

Searching for multiple taxa: To search for multiple taxa at the same rank, select the appropriate option in box 3, and then enter the names, separated by commas, in box 4. For example, selecting "All collections" from the list of collections and then selecting "Scientific name" from the options offered by arrow 3 and entering "Pinus wallichiana, Pinus roxburghii, Pinus gerardiana, Pinus halepensis" into box 4 retrieves 47 records and the distribution map shown below.

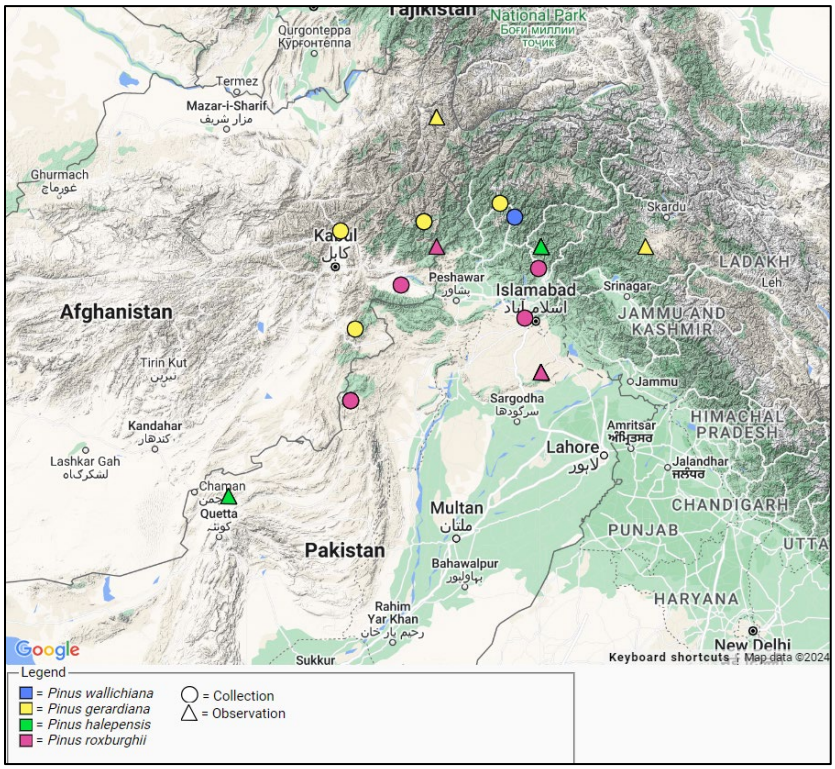
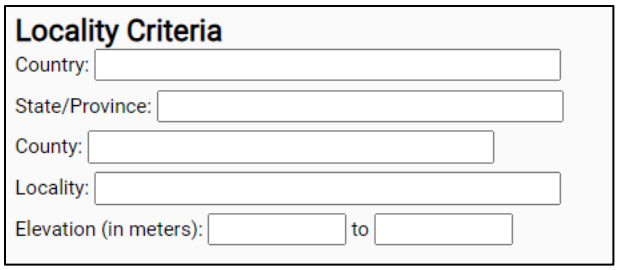


Fig. 5. Map generated in OpenHerbarium showing the distribution of four different species of Pinus.

Arrow 5

The boxes in the locality field use refer to different administrative levels within a country, but they do so by the names used in North America, particularly the US.



Country (Admin level 1): For most parts of the world, this is straightforward, but there are exceptions. For example, the OpenSites treat Somaliland and Palestine as countries. Also, of course, Ukraine (including the Crimean Peninsula).

State/Province (Admin level 2): In Somaliland, Ethiopia, and Ukraine, the corresponding entities are called *Regions*, in Egypt *Governorates*, and in Seychelles *Districts*.

County (admin level 3): In Somaliland and Pakistan, the corresponding units are called *Districts*, in Ethiopia *Zones*, in Egypt *Markaz* (sing. *Marakiz*), and in Ukraine *Raions*. Seychelles does not have a third administrative level.

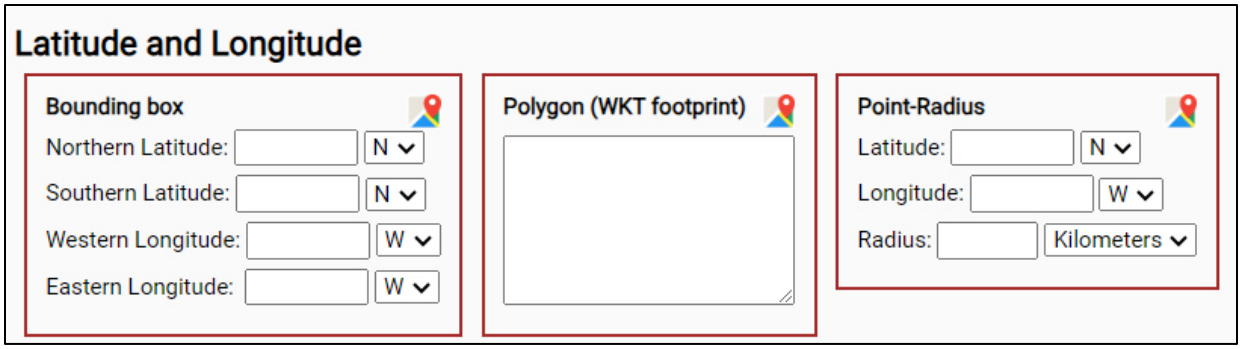
Locality: Use this for a distinctive term that is likely to be included in the written locality information, for example the name of a town, mountain, or notable monument (such as one that is written up in Wikipedia).

Elevation: Notice that the numbers entered should be in meters, not feet.

Multiple search criteria: You can enter search criteria in all the fields or only some of them. More significantly, you can search for multiple names within an administrative level by separating the name with a comma. For example, entering Pakistan in the country and "Upper South Waziristan, Lower South Waziristan, South Waziristan, Tank, Dera Ismail Khan" in the County box will bring up all records from the southern districts of Khyber Pakhtunkhwa. I included both the District "South Waziristan" and the two districts into which it has been split because most existing records come from collections made before the split.

Latitude and Longitude (Arrows 6-9)

There are three options for finding records using latitude and longitude data. The first two, "Bounding box and "Polygon" specify an area of interest by providing 4 (Bounding box) or more (Polygon) points around the area. The third option, Point-Radius, requires only one pair of coordinates but one also specify the radius of a circle around that location to be sure to locate all the records of interest.



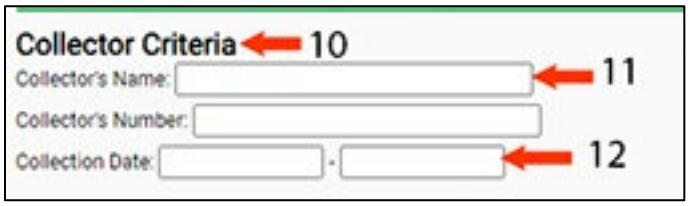
Boxes for setting latitude and longitude search criteria

Bounding box: This requires specifying 2 latitudes and 2 longitudes that will include the area of interest. Enter them in decimal format, not a combination of degrees, minutes, and seconds.

Polygon: This requires pasting in a WKT footprint into the box. "WKT" stands for "Well known text". Well known perhaps to computer programmers, not your average collector. It is a list of pairs of coordinates that specify a vertex (corner) of a polygon surrounding the area of interest to you. For information on how to create a WKT specification for an area in the OpenSites, click here.

Point-radius: With this option, one species just one coordinate but then specifies how large an area around it to search. This is important because the latitude/longitude data on collection records need to be regarded as approximate. *It is up to users to decide how serious the problem is **for their purpose***. Collectors cannot do more than their equipment, time, and knowledge permit. Unfortunately, many collectors (including me) started collecting before they were aware of the importance of stating the "uncertainty" of their latitude/longitude data.

Collector criteria (Arrows 10-12)



Collector's name: This can be any part of a collector's name, but if there are many different people who have a name, the computer cannot determine which one you are looking for. For example, in some parts of the world, "Mary" would bring up records from many different people. On the other hand, I seem (at present) to be the only collector with "Barkworth" as part of their name. Entering all parts of a person's name is better, people are often not consistent in how they write their names. For instance. I may appear as M.E. Barkworth, Mary E. Barkworth, or M. Barkworth. A computer would consider these different names.

Collector's number: This, if known, is useful. Despite being referred to as a number, it may be any combination of letters and digits, but collectors should NOT include commas in their "numbers" because these are used as field separators in csv files and in, in the OpenSites,

searching for multiple items within a search block. Even if a collector used “pure” numbers, one cannot search for a range of numbers by entering “5-500”; each number must be entered separately (see below).

Collection date: Dates can be entered in the standard international format (YYYY-MM-DD) or as 1 May 2002. One can also enter them as XX-XX-YYYY BUT such numbers will be interpreted as if following the US convention for dates: MM-DD-YYYY, not the European convention of DD-MM-YYYY. It is not possible to search for collections made in the same month but different years.

Multiple collectors of collector’s numbers: to search for multiple entries in either of these two fields, enter each name or number separated by commas. For example, “15, 25, 16, 17, 17b”. Note that their order does not matter.

Specimen criteria (Arrows 13-14)

Catalog number: This must be the whole “number”, including any letters. To search for multiple catalog numbers, enter each one and separate them by commas.

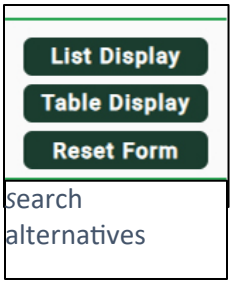
Geocoordinates: This means “with latitude and/or longitude data”.

Cultivated/captivated: This looks for records that have the Cultivated/captive button on the data entry form checked. If it is not checked, the specimen will not be found by this search even if the locality or habitat field indicates the specimen was a cultivated plant or captive animal.

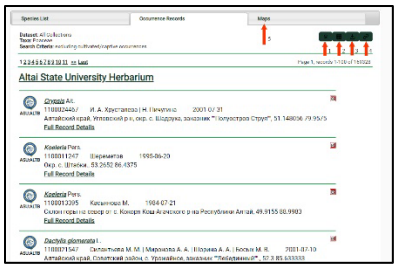
The other fields are self-explanatory.

Search (finally) (Arrow 15)

There are three buttons for searching. Clicking “Reset form” will clear all your search criteria. The other two buttons will search for records meeting the criteria you set.



List display



This will display the records in the format shown on the right. One cannot sort the records, but the dark buttons on the top right (arrows 1-4) provide some useful options if you wish to work with the data.

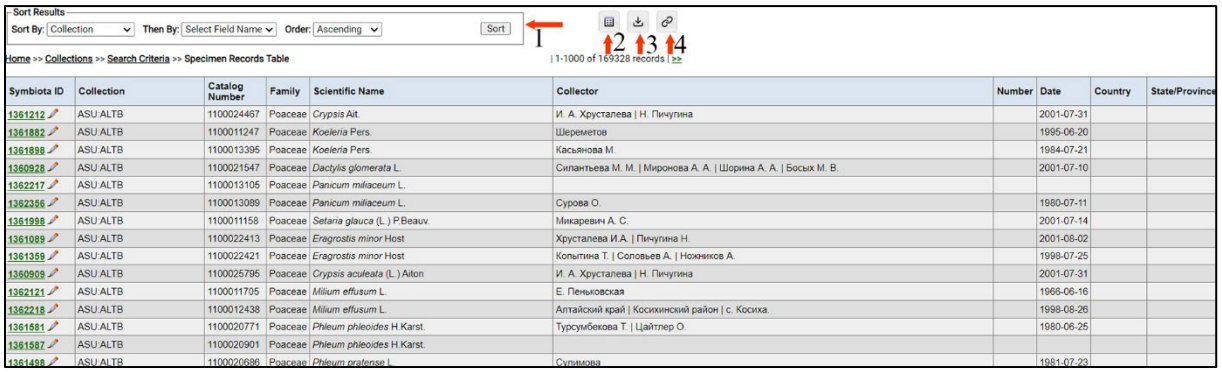
Arrow 1 Create a dataset (or add to an existing dataset). I have not used this tool yet. I think it would be valuable if working on a paper that will require submission of your data as a supplemental file. For more information about datasets and their uses see, https://biokic.github.io/symbiota-docs/user/dataset/edit/.

Arrow 2 This button is for switching to the Table Display.

Arrow 3. Allows you to download the records as a csv file so you can use them in other analyses. If you do so, remember to keep both the SymbiotaID and the CatalogNumber. Both are (supposedly unique identifiers, but some observations will not have a catalog number. It is safer to keep both.

Arrow 4. This button allows you to copy the URL for your search. You can then save it, making regenerating the search easy, and/or share it with a colleague so you can both work on the records in it. You will not be able to modify any of the records unless you have been given permission to do so and that will require having an account in the portal concerned.

Table display



Symbiota ID	Collection	Catalog Number	Family	Scientific Name	Collector	Number	Date	Country	State/Province
1361212	ASU:ALTB	1100024467	Poaceae	*Crypsis* Ait.	И. А. Хрусталева | Н. Пичугина		2001-07-31		
1361882	ASU:ALTB	1100011247	Poaceae	*Koeleria* Pers.	Шереметов		1995-06-20		
1361898	ASU:ALTB	1100013395	Poaceae	*Koeleria* Pers.	Касьянова М.		1984-07-21		
1360928	ASU:ALTB	1100021547	Poaceae	*Dactylis glomerata* L.	Силантьева М. М. | Миронова А. А. | Шорина А. А. | Босых М. В.		2001-07-10		
1362217	ASU:ALTB	1100013105	Poaceae	*Panicum miliaceum* L.					
1362356	ASU:ALTB	1100013089	Poaceae	*Panicum miliaceum* L.	Сурова О.		1980-07-11		
1361998	ASU:ALTB	1100011158	Poaceae	*Setaria glauca* (L.) P.Beauv.	Микаревич А. С.		2001-07-14		
1361089	ASU:ALTB	1100022413	Poaceae	*Eragrostis minor* Host	Хрусталева И.А. | Пичугина Н.		2001-08-02		
1361359	ASU:ALTB	1100022421	Poaceae	*Eragrostis minor* Host	Копытина Т. | Соловьев А. | Ножников А.		1998-07-25		
1360909	ASU:ALTB	1100025795	Poaceae	*Crypsis aculeata* (L.) Aiton	И. А. Хрусталева | Н. Пичугина		2001-07-31		
1362121	ASU:ALTB	1100011705	Poaceae	*Milium effusum* L.	Е. Пеньковская		1966-06-16		
1362218	ASU:ALTB	1100012438	Poaceae	*Milium effusum* L.	Алтайский край | Косихинский район | с. Косиха.		1998-08-26		
1361581	ASU:ALTB	1100020771	Poaceae	*Phleum phleoides* H.Karst.	Турсумбекова Т. | Цайтлер О.		1980-06-25		
1361587	ASU:ALTB	1100020901	Poaceae	*Phleum phleoides* H.Karst.					
1361498	ASU:ALTB	1100020686	Poaceae	*Phleum pratense* L.	Сулимова		1981-07-23		

Table display from the same search

Table display, unlike List display, allows sorting the records by up to two fields, using the three boxes indicated by arrow 1. The buttons above arrows 2, 3, & 4 are the same as those in List display (or see below). What table display does not have is a button for switching to List display.

Arrow 2 Create a dataset (or add to an existing dataset). I have not used this tool yet. I think it would be valuable if working on a paper that will require submission of your data as a supplemental file. For more information about datasets and their uses see, https://biokic.github.io/symbiota-docs/user/dataset/edit/.

Arrow 3. Allows you to download the records as a csv file so you can use them in other analyses. If you do so, remember to keep both the SymbiotaID and the CatalogNumber. Both are (supposedly unique identifiers, but some observations will not have a catalog number. It is safer to keep both.

Arrow 4. This button allows you to copy the URL for your search. You can then save it, making regenerating the search easy, and/or share it with a colleague so you can both work on the records in it. You will not be able to modify any of the records unless you have been given permission to do so and that will require having an account in the portal concerned.