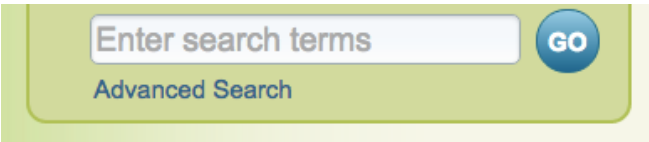
VectorBase Hands-on Workshop
August 16, 2015
Intercontinental Hotel
Cali, Colombia

VectorBase PopBio: Advanced Search

We've already introduced the basics of search, now let's see how to find more specific information and data.

Click the "Advanced Search" link below the search box in the top-right corner of any page in VectorBase. You will land on a page showing every single search result available in VectorBase.



Familiarize yourself with the "Filter Results" section on the left hand side. You can "drill down" into subsections of VectorBase data by clicking on the links in the filter section, and "back out" again by clicking the "Reset Filter" links. Drill down into the Population Biology subsection (click on "Population Biology").

1. How many samples are in the database?

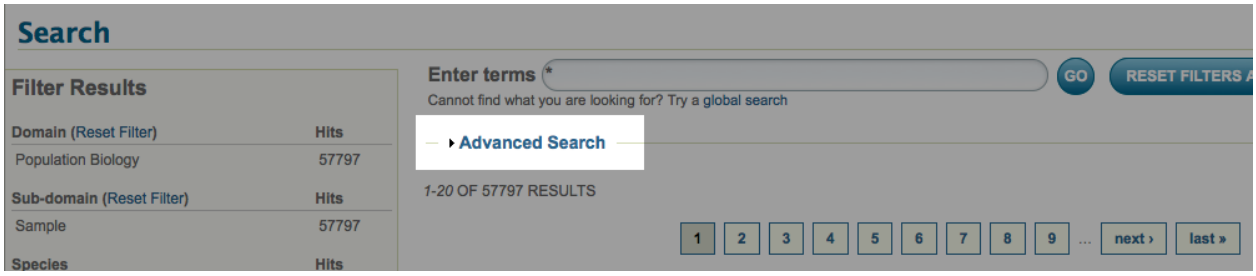
Sample search

Drill down into the Samples and scroll down the filter section to see links for restricting further by species and project.

2. Which project has the most samples?

3. Drill down into that and see which species it contains data for.

Reset the project filter, and open up the Advanced Search (see image below) panel so that you can see the "Add field" link.



Click on "Add field" and check the boxes for "Collection or assay date" and "Collection site". Click anywhere outside the popup menu to dismiss it.

We're going to search for samples collected during the year 2000 in different regions of Africa. First enter the date range* 2000-01 to 2000-12, then "West Africa" (with or without quotes).

*in some browsers a graphical "date picker" will appear, in others you may have to type the YYYY-MM date.

4. Do searches for both "West Africa" and "East Africa" and write down the number of results for each:

Practice adding and removing more advanced search filters. For example you could restrict results to "population" in "Sample type" (other common values for Sample type are "pool" and "individual").

5. How many "individual" type samples were collected in Mali? (no date restriction).

6. Add a "Search by geo-location" to restrict results to those collected within 100km of coordinates 11.314, -5.695 (Sikasso, Mali). How many results are there now?

The main search box

Note that while using advanced search it is normal to ignore the main search box (labelled "Enter terms") just above the advanced search fields. Just leave the asterisk there and use the advanced search fields only. You will get more predictable results this way.

Assay search

Start a completely new advanced search (link in top-right of any VectorBase page). In the advanced search menu, select "Population Biology - Assays". If you click "GO" now, it's the same as if you had "drilled down" using the search filters into PopBio Assays, but with fewer clicks.

Now add an advanced search field for "Protocols/Method" and enter "aspiration" and click "GO". Inspect the results, you will see that were looking for the collection protocol "catch by aspiration". You can alter your search to "catch by aspiration" or even "MIRO:30000043" which is the accession code/ID for the ontology term that is stored in the PopBio database.

How would you find out this ontology term accession? Any easy way, is to click on the green links in the PopBio browser. Open one of the search results in a new tab, click on the green "catch by aspiration" text and you will see the accession code.

Assay types

Using the "Add field" menu, remove the Protocols/Method field and add the "Assay type" field. The values you can enter here are

- field collection
- species identification
- genotype
- phenotype

Let's search for "genotype" assays…

7. How many are there in total?

This number includes several different kinds of assay so let's restrict it to one kind, using the Protocols/Method field.

How you know what to type in the box? At the moment the best way is to use simple search and browse, and also the new map search tool, to find an example data item of the type you are looking for and get the protocol name or ontology accession from there.

For example, you could do a simple search for "karyotype" and drill down into PopBio Assays and look at assay VBA0108013 - from there you can copy the text "cytological chromosome examination" or its ontology accession "MIRO:30000037".

8. How many genotype assays that used protocol "cytological chromosome examination"" (MIRO:30000037) are there?

Now let's add the "Genotype" field and search for a specific chromosomal inversion genotype. Again we need to know what we are looking for, so go to the page for VBA0108013 (via simple search) and pick one of the genotype descriptions, such as "2Rc heterozygous inverted" and paste that into the Genotype field.

9. How many samples have the "2Rc heterozygous inverted" genotype?

Downloading results

You can download a text file of comma-separated results using the "Export results" button. This feature is currently under development and a wider choice of fields/columns will be available soon.

Current limitations

In the "Assay search" facility, some searches are not possible. For example, try adding "Collection site = Africa" to the previous genotype query. You will get zero results. The reason is that only field collections (which are a kind of assay, remember) have collection site information. (For similar reasons it is also not possible to filter genotype assay results by species.)

To get around this, we have included genotype and phenotype information in the Sample entries in our search index. Take the above search with zero results and change the Domain/sub-domain to "PopBio - Samples" (say "yes" to the warning about removing some fields that are not available for Samples). Now you get same number of results back as before, although this time each result is a Sample, not an Assay.

10. How many Samples from Mali have the "2Rc heterozygous inverted" genotype?

However, we can no longer filter our results based on assay protocols. There will be changes to the PopBio search index coming soon so watch this space.

Insecticide resistance search

Phenotype assays that measure insecticide resistance are treated specially in VectorBase site search at present. For example, unlike with other assays, as mentioned above, the species, collection protocols and other information is "flattened" into the search index records for insecticide resistance assays.

11. Switch the Domain/sub-domain to "Insecticide Resistance Assays". The collection site field should still contain "Mali" - click "GO" to see what happens, then change "Mali" to "Colombia". How many results do you get?

12. Look at the species filter on the left hand side… something is strange about the species listed there, what is it?

This is due to an annotation error in the database, which we will fix for the October release. Please ignore it for now. If you like, you can view the problem more clearly by searching for "Colombia *in Geography*" in the PopBio map. http://funcgen.vectorbase.org/popbio-map-preview/

Back to insecticide resistance advanced search…

Add the "Insecticide" field and type "pyrethroid" - you can try other insecticides and insecticide classes.

In a forthcoming VectorBase release we will replace the current assay-centric search with a phenotype-centric alternative. This will allow much more powerful searches, for example based on numeric ranges of insecticide concentrations and resistance measures.

Project search

Finally we will show you a useful way to get a list of all insecticide resistance projects in the PopBio database.

Change the Domain/sub-domain to "Population Biology - Projects", add the "Study design" field and enter "compound treatment" into the box (and click "GO").

Wrap up

Now you have had a comprehensive introduction to Advanced Search of PopBio data in VectorBase. Similar principles apply to other data domains, such as expression and genome annotation data. Remember to look out for the improved and hopefully simplified PopBio search coming soon.

Homework Exercises

1. Use Advanced Search and/or the new PopBio Map Search to compare the species collected using animal vs. human/man baited catch methods. Choose one geographical area for this analysis. It may be useful to filter on "population"-type Samples. In the PopBio map, you can select each of the animal-related catch protocols or you can type "animal" and select the "~7721 *in Collection protocols*" from More suggestions.

2. As above, use Advanced Search and/or the PopBio Map to compare species distributions between two neighboring geographical regions (where some species will be shared). Filter on sample type = "population" or "individual" (but not both). If using the map, you may want to zoom out fully.