



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

VectorBase Map-based search and analysis of insecticide resistance data

A completely new map-based search and browse web interface is being developed for VectorBase's population biology data. A beta-release version will be made available as part of the August or October 2015 release (VectorBase adds new features and data every two months). For now you can access the page at the following URL:

<http://funcgen.vectorbase.org/popbio-map-preview/>

1. Your first task is to play with the map for five minutes and explore the data. We would like to know your first impressions, and if anything doesn't work as you expect. Is anything confusing? Are any major features missing? (Note: we will be adding date-based search very soon.) Please type and send your message via VectorBase contact form: <https://www.vectorbase.org/contact>

Did you notice the two types of view?

- Samples view
- Insecticide resistance phenotypes view

If not, you can select these view types from the menu just to the right of the search box inside the map.

Before we continue, reset the map by reloading the page in your web browser, and make sure that the "samples view" is selected.

Then, if you have a small screen, it's a good idea to click on the "full screen" icon, top right. Make sure you are in "Samples view" for the following section.

Search

Only the samples that match the current search will be shown on the map.

Start typing "*Anopheles darlingi*" in the search box. Stop when you see "*Anopheles darlingi* in Taxonomy" in the autocomplete suggestions below where you are typing. Choose this option and see how the display changes. The total number of displayed items (samples) is shown at the bottom of the map. It should say 927 samples in current view.

Clear the search by clicking the small "x" next to "*Anopheles darlingi*" or the larger red "X" next to the search box.

Try the same thing again, but this time just start typing "*darlingi*" and you should get the same autocomplete suggestion.

Cancel/clear the search again and reset the zoom level (globe icon below the "+" and "-" icons, top-right).

Now we will learn what the blue globe button next to the red 'X' button does.

Type "*darl*" into the search box again and you will see the "*Anopheles darlingi*" suggestion. Now zoom the map in to show only India or Africa. Click the blue globe icon (so that it goes white) and try typing "*darl*" again - there are no search term suggestions any more. If you start typing the name of a species endemic to India or Africa you will see suggestions.

2. In your own words, explain the behavior of the autocomplete search suggestions with and without the globe button pressed:

Reset the search and zoom level so that you can view the whole world, and make sure the globe icon is blue.

3. You can also search using synonyms of species names. Try "*costalis*". Note that the autocomplete suggestion has an icon. Hover over the icon with your mouse and see what the "tooltip" explanation is. Write it down here:

Select the autocomplete suggestion for "*Anopheles costalis* in Taxonomy". Click on some of the map markers and observe the species pie charts that appear on the left hand side. If you don't see any map markers you may need to zoom out.

4. *Anopheles costalis* is a synonym for a well known species complex, what is it?

Now we're going to search for two different species at the same time. Clear the search again and reset the zoom level. Start typing and selecting the "in Taxonomy" suggestions for both of these species names (do one and then the other):

- *Anopheles culicifacies*
- *Anopheles fluviatilis*

Note that when searching for two different things within the same category (in this case "in Taxonomy") then the search terms are combined with a logical "or". Later we will search for terms in different categories - which are combined with "and".

5. Assuming that all the data shown on the map was collected in an unbiased way, describe briefly where the two species co-occur.

Digging deeper

Keep the two-species search from the previous section. Zoom into the southern tip of India until you see several markers with two "samples", one from each species. What is going on here? Why is there exactly one of each species at these collection sites? We can find out more about what data is shown on the map by clicking (once) on a marker. By default, a species pie chart is shown on the left hand side. Click the table icon (third from the top on the left hand side) so it says "Marker details" and shows information panels about each sample. An important piece of information is shown in parentheses after the sample stable ID. Most of the samples in this part of India/Sri Lanka are either "population" or "pool" samples. Click around on various markers until you see both of these sample types.

You will notice that the "population" samples come from project VBP0000029. You can click on this and it will open up the project page in another browser tab. There you can see that this is the Asia data from the **Malaria Atlas Project**. The "old map" in the "graphical summaries" section of the page takes a long time to load and could stress your computer - so click through to the publication page ("Journal" link is the quickest) and you can read there that each data

point represents the *presence* of a local population of a particular species on a particular date (or date range) using one or more specific collection protocols.

If you wanted to show only the "population" samples in the PopBio map display, you can refine your search by typing "population" into the search box and selecting "population *in Sample type*" from the autocomplete suggestions. Stay zoomed in to southern India and cancel the two species search terms (click the small 'x' next to them). When no species are searched for, *all* species are shown.

6. Several collection sites show the presence of eight species. Look at the table and pie chart information panels to see what these species are and write them below:

7. The marker near the town of Dindigul^{*} contains 13 population samples, but only two species. Can you explain why there are 12 *A. stephensi* samples?

--

Insecticide resistance data

Clear the search, reset the zoom and switch to **Insecticide resistance phenotypes view**.

Zoom in to Central and South America and make sure you can see a blue marker with 66 samples over Puerto Rico, and a red marker with 6 samples in North East Brazil.

Click the "i" icon in the bottom right corner of the map. This shows the species color key (also available in Sample view mode) and a resistance color scale from blue (susceptible) to red (resistant). More info on the color scale in the panel below:

RESISTANCE COLOR SCALE

VectorBase's PopBio resource contains insecticide resistance data from different assay protocols and reported in a variety of measures and units, such as percent mortality, lethal concentration (e.g. LC50) and lethal time (e.g. LT95). Not all of these protocols

^{*} Sorry there is currently no place name search facility like in Google/Bing maps at the moment. Dindigul is just north of Madurai at the tip of Southern India.

and measures have WHO-recognized thresholds for categorizing into “susceptible” or “resistant” classes.

To aid the user in discovering geographical regions of resistance we have rescaled all comparable data (see below) between 0 (susceptible) and 1 (resistant) after discarding data outside the 2nd and 98th percentiles, and inverting value ranges where appropriate. These rescaled values are used to color the map markers (from blue to red).

COMPARABLE DATA

By comparable data we mean all phenotypes recorded with the same measurement type and units. For example, LC50 measured in parts per million or LT95 measured in minutes. Note that comparable data currently includes assays performed using different concentrations of insecticides. Thus the data is not strictly comparable. However the insecticide concentrations are displayed in info panel popups for plots showing individual data points. A search facility to limit displayed data to particular insecticide concentrations will be added soon.

Mouse-over the Puerto Rican '66' marker. The grey box that appears shows the GPS coordinates of the region in which the 66 samples are found. Double click on the marker to zoom in to show the 66 markers... You may want to hide the color legend panel if it is hiding some markers.

8. How many different species from Puerto Rico have IR phenotypes in VectorBase at the moment?

9. Briefly switch to "Samples view". How many different species are present in this region including any non-IR samples?

Switch back to IR phenotypes view. Don't be alarmed that the species color scheme changes. It is supposed to do this.

Now click on the "10" marker on Puerto Rico's northern coast and select the graph/plot icon on the left hand panel. You will see IR phenotype data for the clicked "selection" marker plotted against "background" data. The background data can be defined in three different ways but we will come back to that. Leave it on the default "phenotypes matching search".

The marker contained 10 phenotypes but only 5 are plotted - LC50 milligram per liter. The second drop-down menu takes you to another plot for LC95 milligram per liter. You can see that in total there are plots for 10 phenotypes.

10. Mouse-over the points on the right hand plot for LC50. What is the difference between the 5 points?

11. Now you are going to restrict all displayed data to those assays using malathion as the insecticide. Figure it out for yourself and write down briefly how you did it.

Now the map shows the relative picture of malathion resistance in the area.

You can also search for insecticides using higher order chemical classes such as "organophosphate" and "pyrethroid". Try that out before going back to a simple search for "malathion".

View the plots (LC50 and LC90) for the two phenotypes with the highest resistance (the lighter blue marker in the same place on the north coast as before). You can now see that the background plots only show malathion data. There is less data than before, so now the plots can show individual data points.

12. At least for LC50, you can see that this assay suggests relatively high resistance. Only two other samples in the database have higher LC50 values. Where and when are those samples from?

13. What is the collection date of the Puerto Rican sample?

Note that clicking on the data points makes the popup info panels sticky.

Note that the LC50 background data points include assays performed on other species. Restrict the search to "*Aedes aegypti*"

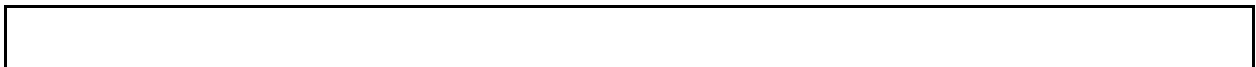
Now you can see that there are relatively few (n=7) background data points with the same combination of measurement type (LC50, milligrams per liter), species and insecticide.

Click on the 3 marker at the east end of the island. Now you can see data for LC50, LC90 and LC95, all reported in parts per million. Again this assay suggests relatively high levels of resistance and this is reflected by the lighter blue coloring of the marker.

At this point you may be wondering if it is meaningful to compare Central American assay results with those from Australia. Luckily, you can restrict the background data to those shown in the current map view (and therefore also matching the current search) by selecting "phenotypes visible on map" in the upper menu of the plot panel. As expected, this also reduces the number of data points plotted quite drastically. A third background data option "all phenotypes" ignores both the search query and the current view and shows all "compatible data". This option is probably not very useful.

Now take a moment to reflect on how the IR plots can change when you move around the map and change your search terms.

14. Clear the search and zoom out to show Brazil again. Zoom into the area around the red 6 until the yellow 7 below it splits up into smaller markers. Which insecticide is showing high levels of resistance? **Hint:** mouse-over the points in the plot to see this or consult the marker table display



Note: we have noticed a problem with the display of insecticide concentrations (they are missing in some cases, such as these Brazilian assays, but you can click through to the PopBio browser to see them).

Restrict the search to this insecticide and also *Aedes aegypti* and zoom out to see the global picture.

Reset the search and add *Aedes albopictus* as a search term. You can also search/filter on the type of insecticide resistance assay performed. Enter "dose" and select the general term "dose response test in *Protocols*". This will now show only lethal concentration-type (e.g. LC50) results.

As you can see, after a few keystrokes and mouse clicks, you can get a picture of insecticide resistance for the specific species and/or insecticides of interest. When we add more up to date

data from the **President's Malaria Initiative (PMI)** and other sources, and provide date-based search and filtering, it will be a very simple and powerful tool.

Other map features you can try

1. You can change the map background style (see the layers icon). In the future we will add layers for vegetation type, climate and epidemiology.
2. Zoom to a rectangle with shift-click-drag
3. Search on collection protocols (try typing "catch" or "human")
4. Text-based search on geographical regions e.g. "Europe" (but note that some data points may be missed due to issues with data annotation quality)

Coming soon

1. As mentioned already, date search
2. More data from PMI
3. Color coded pie charts based on other annotations (not just species), such as sample type, insecticide, protocols
4. Integration with "normal" VectorBase site search