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

VectorBase Population Biology Browser, PopBio: Simple Search and Browse

Note, the URLs given in this page are provided for confirmation and reassurance only. It should be possible to follow the workshop without having to type each URL into your browser.

Data in the PopBio resource is organized as follows:

There are three main "items" in the database:

1. Projects
2. Samples
3. Assays

Projects are like bags, containing samples and assays. Projects can be thought of as studies, experiments or publications. Sometimes a sample or assay might be re-used in different projects.

Let's take a look at a project page in the PopBio browser.

The project is called "*Susceptibility of Aedes aegypti larvae to the insecticide temephos in the Federal District, Brazil*" and its VectorBase stable ID is **VBP0000017**. You can search with either the title, words from the title, or the ID in the main top-right VectorBase search box and eventually you will come to this page:

https://www.vectorbase.org/popbio/project/?id=VBP0000017

The basic information is listed. Some words and phrases are in green text. Click on one, perhaps "compound treatment design" to see what happens.

These are ontology terms. From the popup, you can browse parent and child terms and search VectorBase for data annotated with the term or its children.

PopBio uses a lot of ontologies - experimental factor ontology (EFO, as you saw above), Units Ontology (UO) to name just two. VectorBase also maintains some specialist ontologies for disease transmission and insecticide resistance. These plus a few third party ontologies are available to browse in the Ontology Browser in the Tools menu.

Find the term "pyrethroid" in the mosquito insecticide resistance ontology (MIRO)

The ontology browser also tells you how many search results there are within VectorBase for the current term.

1. How many Population Biology "hits" are there for pyrethroid?

2. Expand the hierarchy to show the children of pyrethroid. Click each of the children, one by one from the top until you see one with data in "Population Biology". What is the insecticide?

3. Right-click the search link to open a new tab with those results.
You will see the same number of "Assay" results and "Insecticide resistance assay" results. This is because all IR assays are also assays.
Write down which species and projects the results come from:

Species	Projects

Click the link "Culex quinquefasciatus" to "drill down" into results only for that species.

Click on the first result. This takes you to an assay page…

Check that you are viewing this assay: VBA0171764
(If not, search for it using site search.)
Again this is a **VectorBase Stable ID** it will always refer to this assay, even if we make changes or corrections to the database.

Again there are various tables of text information and green ontology term links.

4. Assays usually have one or more protocols to formalize what has been done, and one or more results. This is a phenotype assay and its results are phenotypes. In a few words, what is the phenotype from this assay?

5. Which sample and project does this assay belong to?

6. Click on the sample IRBS68/VBS0057362.
Note the familiar layout of the sample page. Basic info is at the top, with tables and lists of associated items below.
You have now seen three types of VectorBase stable ID for PopBio data. Name the three types of ID below:

VBPnnnnnnn VBSnnnnnnn VBAnnnnnnn

The field collection and assays are the main points of interest on the sample page.

Quickly take a look at the field collection page
https://www.vectorbase.org/popbio/assay/?id=VBA0171758
Note: field collections are stored inside PopBio as a subtype of assay. This is why the URL for field collections contains the word "assay" and field collections have VBA stable IDs. This will be useful to know when doing Advanced Search of PopBio data.

Let's go back to the project we were looking at before:
https://www.vectorbase.org/popbio/project/?id=VBP0000017

Scroll down to the "Graphical summaries" section. Here you can see a map showing where the collection sites are, and if you click on the tab, also a plot of percent mortality vs time for different collection sites.

Let's look at the map. We'll call these "old maps" because we have a "new map" that you're going to see in another session. The old maps and plots are *custom configured for each project* by VectorBase staff to show data of interest. Let's look at a few more of these maps and plots to get a better idea about this.

Find the "Projects" link in the "breadcrumb trail" near the top of the page, below the main VectorBase site menu.

Scroll down to VBP0000009 "ANVR 1st dataset" and click on it (opening a new browser tab is best). Explore the different maps. The first three aim to summarize the data broadly. Choose the first map (species) and zoom in and out.

7. What happens to the pie charts when you zoom in?

Zoom in a bit and click on a circle or pie chart with a total number greater than one and less than ten. A new browser tab will open with a site search for the items within that circle/pie. From the search results page we can see that these pie charts are showing the number of field collections for each species. You could click through to get more details. For the phenotype maps, a similar functionality exists.

Go back to the list of projects page and scroll down to VBP0000016 and open up that project page. Here the map is configured to show resistance genotypes (in the second tab) for one of the study countries.

Explore the data and answer the question.

8. Which genotype (in terms of gene and mutated amino acid number) is shown on the map?
Hint: the easiest way to find an assay page for a genotyped sample is to click on a pie chart in the genotypes map.

9. Go back to the list of projects page and scroll down to VBP0000023 (or use site search to find it). Explore the plots of LC50 and LC99. You can click on plot data points to find out which samples/assays they represent. There appears to be a resistant phenotype from Rome (Roma). Check that this phenotype comes from the same sample for both LC50 and LC99. What is the field collection stable ID and collection date?

Note that a lot of clicking was required there. The new map interface and advanced search will make it easier to explore some types of data.

Finally we'll do a simple four-click search for some data of interest to you all. Go to the VectorBase homepage (or any page with a search box top-right) and type DDT and click "go". Now click on the links on the left hand side as follows

1. Population Biology
2. Sample
3. *Anopheles nuneztovari* (scroll down until the hit count is 5)

10. You can see on the left hand side that only one project is represented (VBP0000117). Click through to a sample and then to the project page. Who is the first author of the paper?

Wrap up

Now you know what kind of data is available in PopBio and how it is stored and displayed as collections of projects, samples and assays.