Sample and Genotype Explorers

We recommend using a mouse with a track wheel rather than the inbuilt trackpad

1) Compare variants for one *Anopheles funestus* exon between samples in different African countries, and export VCF files for each population

a) In the Tools menu select the Genotype Explorer. Add a species as a reference genome, e.g., *Anopheles funestus*, and a query, e.g., AFUN001649
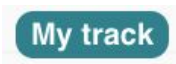
ADD NEW REFERENCE

Click on this icon. A pop-out window will display.

Type a keyword and notice the autocomplete suggestions. Select the suggestion that matches your query. Selected terms are display below the query line in green. Click on 'ADD'.

b) *Hover with your mouse* to discover the names/functions of all this tool icons. Complete this table with the information you learn. Come back again to this question once you finish all five exercises in this tutorial to complete or improve the information.

Icon	Name/Function
SAVE LOAD	
ADD NEW REFERENCE	

🖊 RESET	
↺ UNDO	
➕	
✛	
✖	
My track	
← − + →	
◯	
≡	
↓ ↑	
✖	
🖊	
AGAP004707-RA	

c) add samples - search for 'burkina faso'
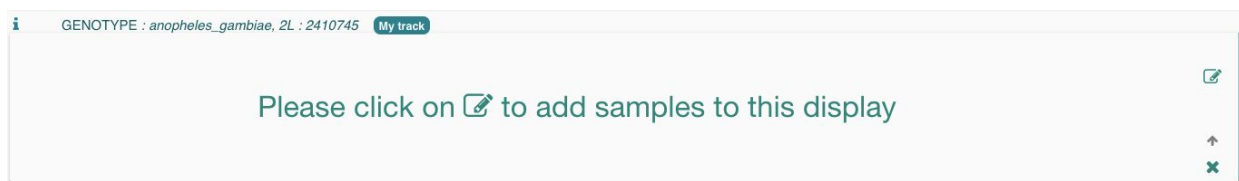
Click on the pencil icon.
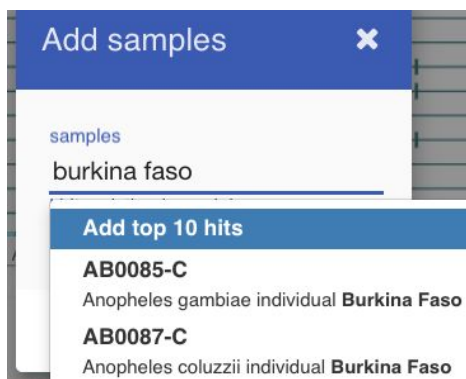
Type the query. From the autocomplete suggestions select 'Add top 10 hits'.
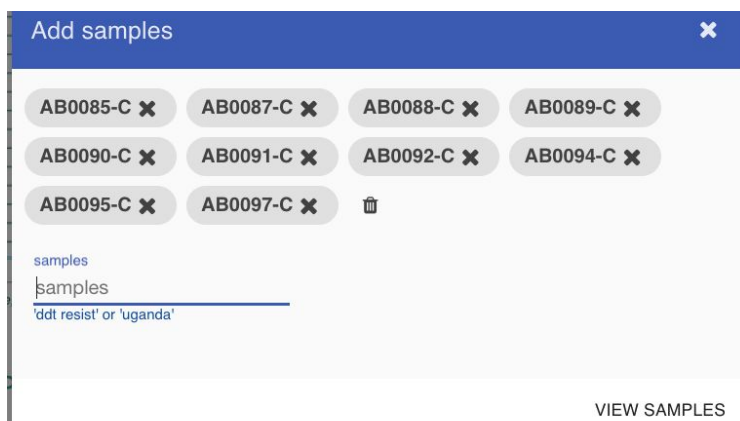
Click on 'VIEW SAMPLES'.

d) pick an exon or location, e.g., KB668221:2514743 and zoom in until you can view genotypes
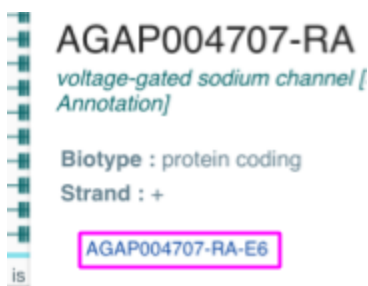
Locate the exon and click on it. A pop-out window confirm you are in the right position

Use the plus/minus and arrowheads to zoom and browse in the right position.

e) add a new track (click on + icon on top left display bar and select genotype)

f) scroll down to the new track and add samples

g) export VCF for each population (click on the VCF icon in the right hand vertical bar item in each of the genotype tracks)

Keep this query results for the next exercise.

2) display protein altering variants

a) return to the transcript track and scroll out so that you can view the entire AFUN001649 gene (Click on the 'Optimise view' icon)

b) select one of the transcripts (click on any exon or intron) and on the pop up box select the 'Protein' (soon to be changed to "Protein domains") button. This will add a new track showing the protein domains and the variants in the context of the protein sequence for each of the previously selected populations.
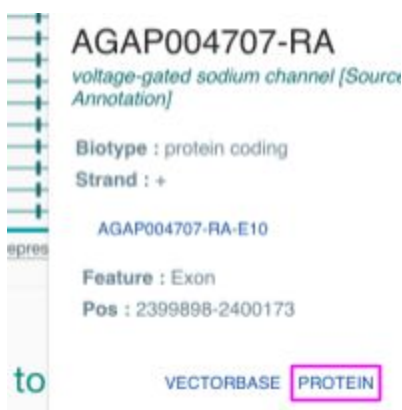
c) Click on the 'Show overview' icon to show/hide features

The 'Show overview' icon turns orange when you have hide features from the display.
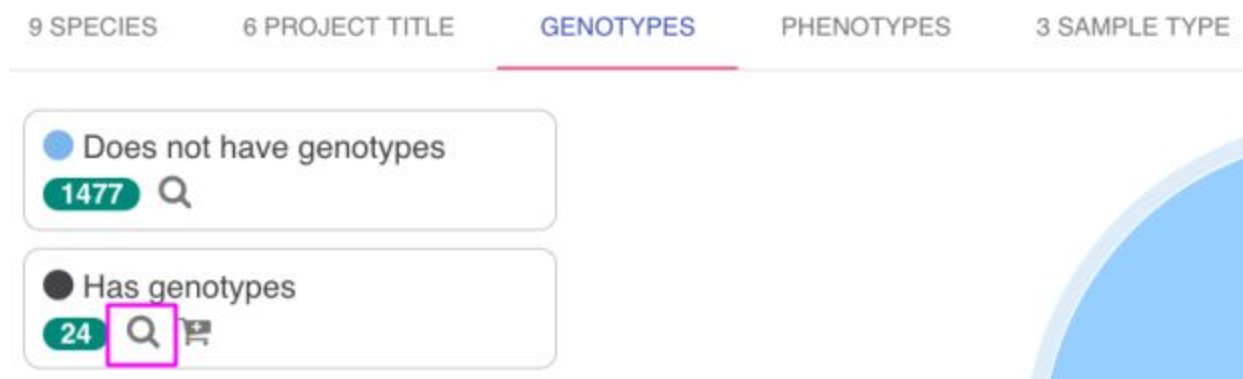
3) Sample Explorer[1]

a) In the Tools menu select the Sample Explorer. Type the term Senegal. How many species have data for this country?

| |
| |

b) Which of these species has genotype data?

| |
| |

Select the 'Genotypes' tab. Click on the magnifying glass icon on the has genotypes box

---

[1] Sample Explorer tutorial, https://www.vectorbase.org/tutorials/tools-and-resources-tutorials/sample-explorer

Click on the species tab

c) What sample types are these genotyping data from Senegal? Individuals or pools? (Click on the 'Sample type' tab)

---

4) Use the *Sample Explorer* to find all the *Anopheles funestus* Burkina Faso samples collected in 2005 with genotypes for the AFUN001544-RA transcript, and display them in the *Genotype Explorer*

a) in Sample Explorer query for "anopheles funestus" + has genotypes + "burkina" + "2005"
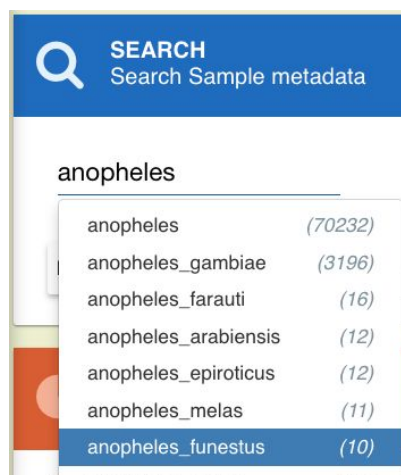
*When typing queries, select the available options from the autocomplete suggestions.*

Select the 'Genotypes' tab. In the has genotypes box select magnifying glass.

anopheles_funestus ✖   true *(has genotypes)* ✖   burkina ✖   200
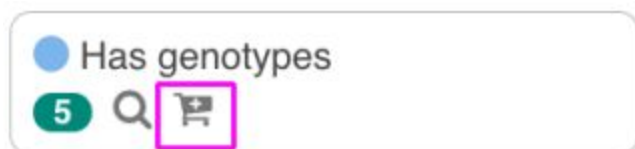
| 2005 | (5) |

Notice how the query is constructed (gray bubbles) to obtain the data of interest.
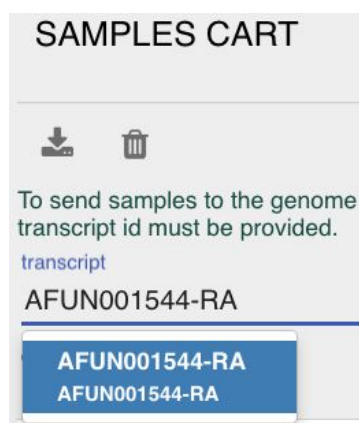
b) scroll to the sample cards and click on the 2005 next to the calendar icon. This extra step helps you filter for data <u>collected on</u> year 2005 (and discard the data <u>published on</u> 2005).
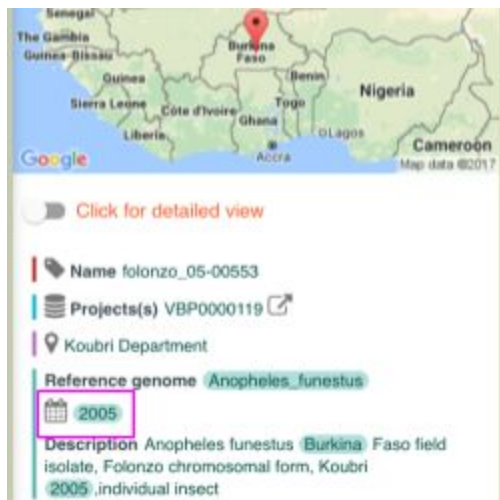
b) scroll up again, to the overview panel. add the samples (5) to the shopping cart

c) go to the shopping cart at the top of the page and add the transcript you wish to view (AFUN001544-RA)

d) click on the '[BETA] view in LERA' menu icon (3rd along from left, the compass)

Keep this query results for the next exercise.

5) compare the *Anopheles funestus* AFUN001544 gene to its ortholog in *Anopheles epiroticus* (note this is an example of cross species genome browsing rather than looking variation data)

a) using the display from from 4(d) click on the AFUN001544-RA transcript object and select "Orthologs"

b) type "epiroticus" into the search box and select the supplied transcript

c) zoom out (or click on 'Optimise view') in the transcript tracks of each species and compare the genomic organisation (optional: close the 'genotype' tracks)

c) does the AFUN016088-RA gene (*Anopheles funestus*) have an ortholog in *Anopheles epiroticus*? (If in any doubt, confirm with a gene tree or other displays from VectorBase comparative genomics [2])

d) the AEPI002805-RA gene has an exon that overlaps the AEPI002806-RA gene - does it have an *An. funestus* homolog, and how does this relate to the genomic structure shown in An funestus.

If you need help with any question and its answer contact us at info@vectorbase.org. Because VectorBase data, tools and resources are updated every two months (6 release cycles per year), answers to these exercises will change too.

---

[2] Comparative Genomics tutorial,
https://www.vectorbase.org/tutorials/tools-and-resources-tutorials/comparative-genomics