

# MapVEu Exercise

## Introduction

MapVEu is part of our ongoing efforts to integrate genomic, phenotypic and population data into an interactive and user-friendly tool. Data for the map primarily comes from manual curation of genomic and phenotypic data extracted from published papers on population biology of vectors of human disease. Surveillance data is often acquired from US mosquito abatement districts. We would like to encourage researchers to actively deposit their data with us. The more data that is aggregated on the map, the more valuable the resource. At present, the map is limited by the level of granularity of data in published papers. It is a snapshot of data in a location and at a specific time. Users must keep in mind that data in MapVEu is not a comprehensive overview of the species present, absent, resistance profiles, or abundance across the globe.

This exercise uses VectorBase.org data as an example, and demonstrates an innovative new version of the MapVEu tool that we have recently released. The new mapping tool is in its infancy and updates to improve usability will be made during the coming months. We welcome your comments and feedback!

## Learning objectives:

In this exercise, we will use data from MapVEu to answer a scientific question about the spatiotemporal distribution of insecticide-resistance mutations in disease vectors. You will learn how to:

- Start a new analysis on MapVEu
- Browse variables and filter (select) data of interest
- Adjust map markers to view the geospatial distribution of any variable
- Make a supporting plot
- Download the raw data
- Find the data sources

## Background

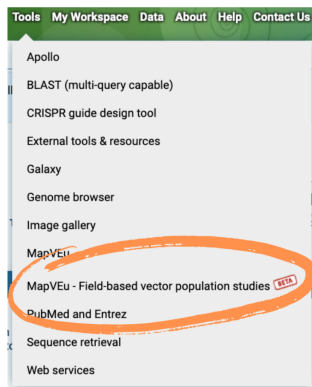
- Malaria is caused by parasites of the *Plasmodium* spp., that are transmitted by mosquitoes of *Anopheles* spp., with high incidence in the African continent, where *Plasmodium* spp. are mainly transmitted by mosquitoes of the *Anopheles gambiae* complex.
- Malaria prevention relies mostly on vector control interventions based on the use of chemical approaches, such as long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) but extensive and indiscriminate use of these insecticides has resulted in the emergence of a wide range of resistance in wild *Anopheles* populations.
- Two primary mechanisms have been reported underlying resistance to insecticides in these populations: target-site insensitivity and metabolic resistance. Target-site insensitivity in *Anopheles* is caused mainly by point mutations present in the voltage-gated sodium channel and they are commonly known as knockdown resistance (*kdr*) mutations. Two *kdr* mutations have been detected in the voltage-gated sodium channel gene (at position 1014 of the encoded protein) of *Anopheles* populations: the replacement of leucine by phenylalanine (L1014F) and the replacement of leucine by serine (L1014S). These mutations have been used as molecular markers for insecticide resistance in *Anopheles* populations.

## Exploring genotypes in the map

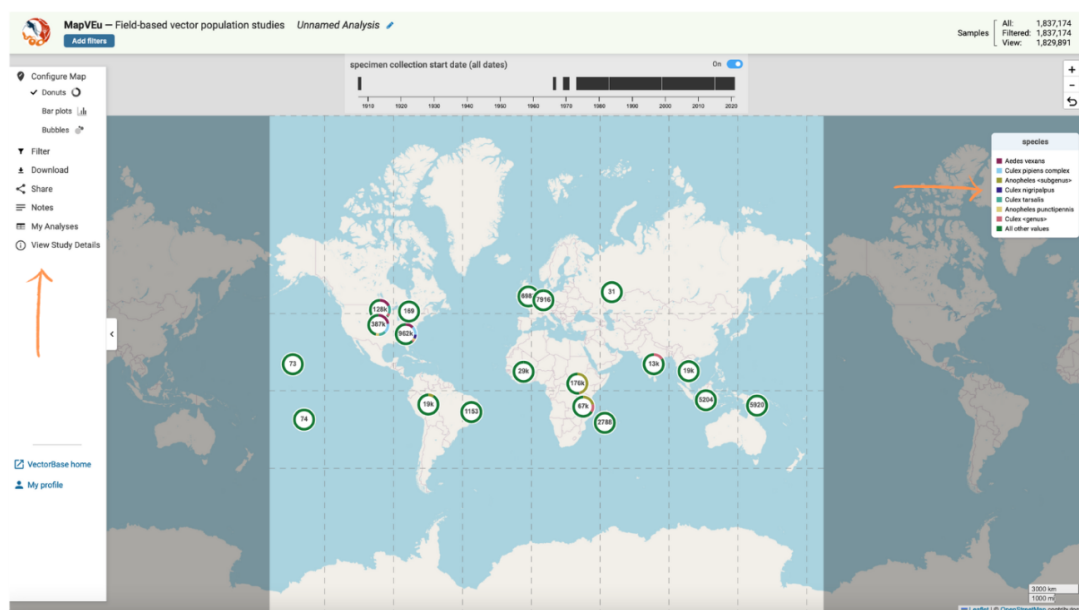
Use MapVEu data to answer:

- Where on the African continent are the two *kdr* mutations L1014S and L1014F most prevalent?
- When did these mutations first appear in the vector *Anopheles gambiae*?

1. Navigate to Vectorbase.org > **Tools** drop-down menu in header > open **MapVEu - Field-based vector population studies (beta)**



2. This opens a new analysis showing the distribution of data- in this case, *species*, as the legend indicates. On the left is a menu of tools to explore the data.

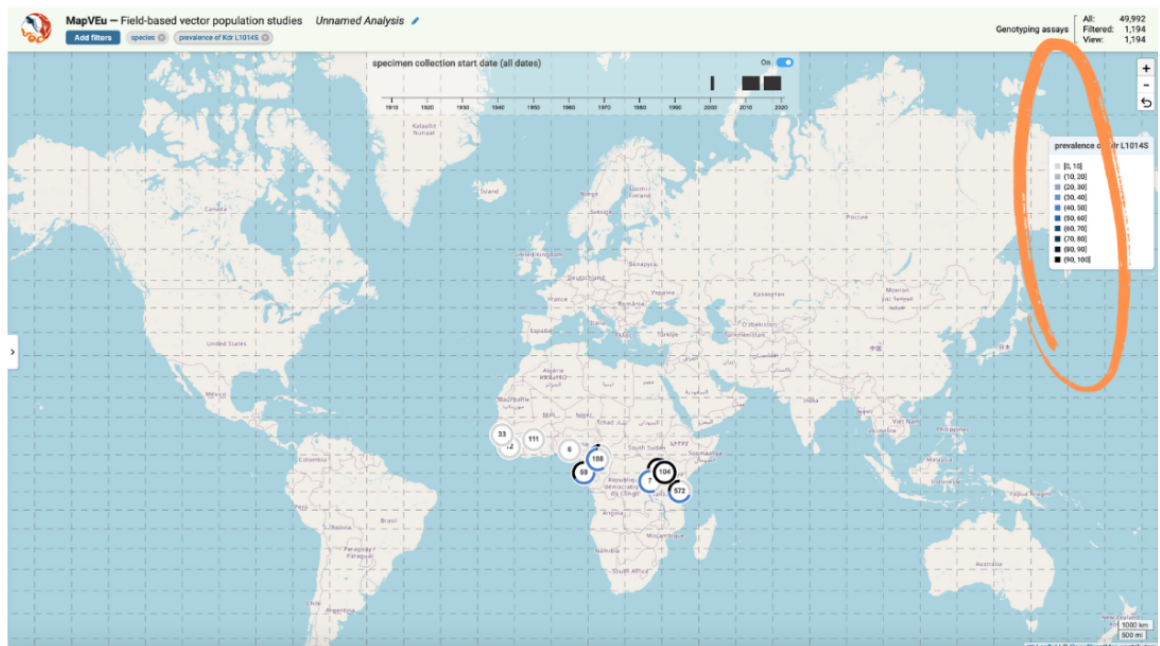


3. Start by selecting the species of interest, in this case *Anopheles gambiae*.
  - a. Click on “**Filter**” in the menu- all variables in the data are displayed in hierarchical categories
  - b. The *Species* category is listed in the Featured variables box at the top, you can also type the word ‘*Species*’ in the variable tree search panel. It will be listed under the Sample category.
  - c. Click on *Species* to see the frequency distribution of species in the data
  - d. Find *Anopheles gambiae* in the list and click the checkbox to restrict the data to this species
  - e. See the effect of filtering on the dataset diagram at the top of this page





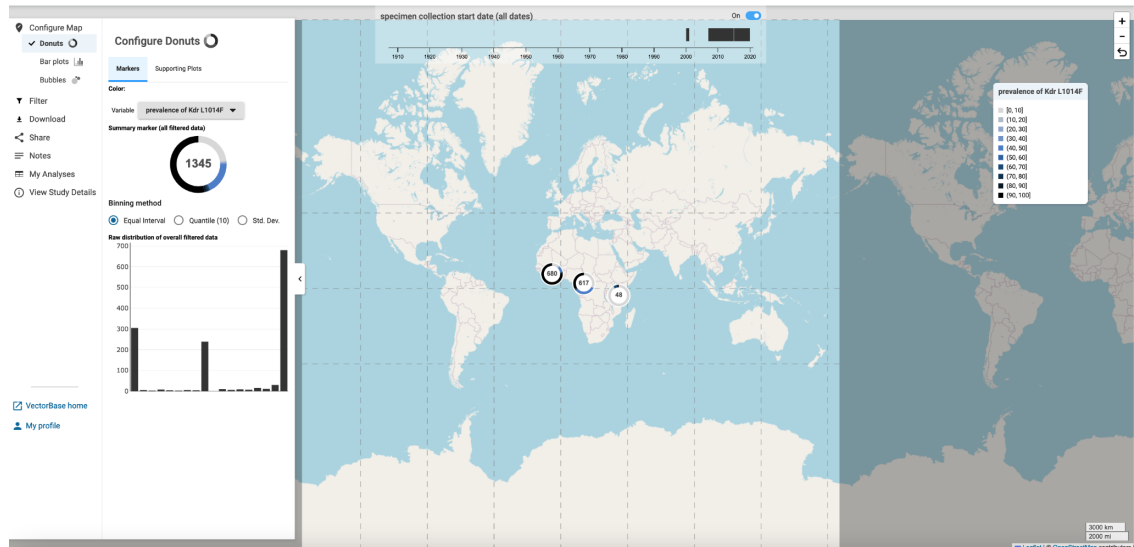
6. Observe the map and answer the following
  - a. What part of Africa has a higher prevalence of Kdr L1014S- east or west?
  - b. Name a country where some of the assays conducted show a prevalence of Kdr L1014S of 90 - 100% (*hint: zoom in*). When looking at the legend on the right you see a grey box with [0, 10] next to it. This indicates 0 - 10% prevalence of Kdr L1014S in the samples you have filtered on. The darker the colour gets the higher the % prevalence of Kdr L1014S.



7. Next, do the same for the other Kdr mutation of interest: Kdr L1014F
  - a. Remove the Kdr L1014S filter at the top of the page, click x to remove filter
  - b. Find the *prevalence of Kdr L1014F* variable and select all samples that have data for this variable
  - c. Change the donut marker to *prevalence of Kdr L1014F*



8. Observe the map and answer the following
  - a. What part of Africa has a higher prevalence of Kdr L1014F- east or west?
  - b. Name a country where some of the assays conducted show a prevalence of Kdr L1014F of 90 - 100%



9. Our next goal is to explore the timeline of the appearance of the Kdr L1014F mutation in *Anopheles gambiae*. We can do this using line plots
  - a. Change the donut marker to *Species*
  - b. Go to **Supporting plots** under Configure donuts
  - c. Supporting plots → Draw a line plot
  - d. Select the following variable to denote time on the X axis: *specimen collection start date*
  - e. Select the following variable to show mutation prevalence on the Y axis: *prevalence of Kdr L1014F*
  - f. Overlay → *Species*
  - g. Auto-zoom
  - h. Change X axis binning to 1 Year





## Appendix: A note on how to submit data to MapVEu and how to cite MapVEu data in your publications

### 1. How to submit data to MapVEu

- a. To submit your data for integration use the [Contact Us](#) link to send a brief description (two or three sentences) of your data and we will be able to help you out with submission!

### 2. How to reference your own data within your paper

**Note:** Please use the VectorBase website for the most up to date information on how to [cite](#) VectorBase and VEuPathDB.

- a. The paper with the PubMed ID [29260658](#), is a great example of how to reference the data **you** deposit with VectorBase. Under the section ‘Data availability’ the following statement can be found:

#### Data availability.

Microsatellite and SNP genotypes were deposited in VectorBase<sup>[48](#), [49](#)</sup>, PopBio projects: VBP0000201 (new data), and VBP0000138, VBP0000176-177 (previously published data). Sequencing data were deposited in NCBI under accession numbers [MF371160-MF371174](#) and [MG241351-MG241354](#).

### 3. Using and citing data submitted by others within your paper

If you would like to cite data in VectorBase that other data providers have deposited, please see an example in the paper with the DOI: [10.1038/s41598-023-30751-4](#).

Please remember it is important to cite *both* VectorBase and the original data generator(s). It is especially important to cite the original data generators when you are using data from primarily one or a small number of studies.

Mosquito control surveillance trap data were downloaded from VectorBase PopBio data repository (<https://vectorbase.org/popbio-map/web/>)<sup>[36](#)</sup>. The data set included multiple fields, which are outlined in an extensive data dictionary provided through the data repository. For the purpose of this study, we included latitude and longitude coordinates of the collection site, the collection date, trap type, attractant, and an abundance value for individual species. Data were filtered to include only New Jersey light traps with visible light as the attractant between the years 2000–2020.

## Answers

6. Observe the map and answer the following

- a. What part of Africa has a higher prevalence of Kdr L1014S- east or west?
  - i. East
- b. Name a country where some of the assays conducted show a prevalence of Kdr L1014S of 90 - 100%
  - i. Kenya, Tanzania, Gabon (any country with black in the donut)

8. Observe the map and answer the following

- a. What part of Africa has a higher prevalence of Kdr L1014F- east or west?
  - i. West
- b. Name a country where some of the assays conducted show a prevalence of Kdr L1014F of 90 - 100%
  - i. Burkina Faso, Sierra Leone, Gabon etc (any country with black in the donut)

10. Look at the resulting graph. What was the year in which the prevalence of L1014F first reached 100% in *Anopheles gambiae*?

- a. 2008