BioDIGS: Antimicrobial Resistance

October 16, 2024

Table of Contents

# About this Book

This is a companion training guide for BioDIGS, a GDSCN project that brings a research experience into the classroom. In this module, students will investigate whether metagenomic samples from the BioDIGS project contain either antimicrobial resistance genes or biosynthesis genes. They will run analyses using ABRicate as implemented on Galaxy. Visit the BioDIGS (BioDiversity and Informatics for Genomics Scholars) website [here](https://biodigs.org/) for more information about this collaborative, distributed research project, including how you can get involved!

The GDSCN (Genomics Data Science Community Network) is a consortium of educators who aim to create a world where researchers, educators, and students from diverse backgrounds are able to fully participate in genomic data science research. You can find more information about its mission and initiatives [here](https://www.gdscn.org/home).



## 0.1 Skills Level

The activities in this guide are written for undergraduate students and beginning graduate students.

*Genetics*

*Programming skills*

## 0.2 Platform

The activities in this guide are demonstrated on NHGRI’s [AnVIL](https://anvilproject.org/) cloud computing platform. AnVIL is the preferred computing platform for the GDSCN. However, all of these activities can be done using your personal installation of R or using the online [Galaxy](usegalaxy.org) portal.

Please check out our full collection of AnVIL and related resources: <https://hutchdatascience.org/AnVIL_Collection/>

## 0.3 Data

The data generated by the BioDIGS project is available through the [BioDIGS website](biodigs.org), as well as through an [AnVIL workspace](https://anvilproject.org).

Data about the soil itself as well as soil metal content was generated by the [Delaware Soil Testing Program](https://www.udel.edu/canr/cooperative-extension/environmental-stewardship/soil-testing/) at the University of Delaware. Sequences were generated by the [Johns Hopkins University Genetic Resources Core Facility](https://grcf.jhmi.edu/) and by [PacBio](https://www.pacb.com/).

## 0.4 AnVIL Account Setup

In order to run your analyses, you will use the [AnVIL cloud computing platform](https://anvilproject.org/). The AnVIL (Analysis Visualization and Informatics Lab-space) platform is specially designed for analyzing biological data, and is used by scientists doing all sorts of biological research.

**AnVIL in a nutshell**

* Behind the scenes, AnVIL relies on Google Cloud Platform to provide computing infrastructure. Basically, AnVIL lets you “rent” computers through the internet. The analysis is run on the rented computer. AnVIL lets you see the results in your browser.
* AnVIL uses [Terra](https://anvil.terra.bio/) to provide many computational tools useful for biological data analysis, such as [RStudio](https://www.rstudio.com/products/rstudio/), [Galaxy](https://usegalaxy.org/), and [Jupyter Notebooks](https://jupyter.org/). Terra takes care of installing these tools so you can use them right away.

### 0.4.1 Create Google Account

First, you will need to set up a (free) Google account.

A Google account usually looks like “myname@gmail.com”. Alternatively, you can enable Google for an existing non-Gmail email address using [these instructions](https://support.terra.bio/hc/en-us/articles/360029186611).

If you do not already have a Google account that you would like to use for accessing AnVIL, [create one now](https://accounts.google.com/SignUp).

### 0.4.2 Log In to Terra

Next, make sure you can log in to Terra – you will use Terra to perform computations on AnVIL.

You can access Terra by going to [anvil.terra.bio](https://anvil.terra.bio/). Open Terra, and you should be prompted to sign in with your Google account.

### 0.4.3 Share Username

Make sure your instructor has your Google account username (e.g. myname@gmail.com), so they can give you access to everything you need.

* Make sure there are no typos!
* If you have multiple Google accounts, make sure you give them the username that you will be using to log in to anvil.terra.bio.

It is *very important* that you share the Google account you will be using to access AnVIL with with your instructor! Otherwise, the instructor cannot add you to Billing Projects or Workspaces, and you will be unable to proceed with your assignments.

## 0.5 Cloning the AnVIL Workspace

## Warning: Chunk option fig.alt is not supported for docx output  
## Warning: Chunk option fig.alt is not supported for docx output

This **will not work** until your instructor has given you permission to spend money to “rent” the computers that will power your analyses (by adding you to a “Billing Project”).

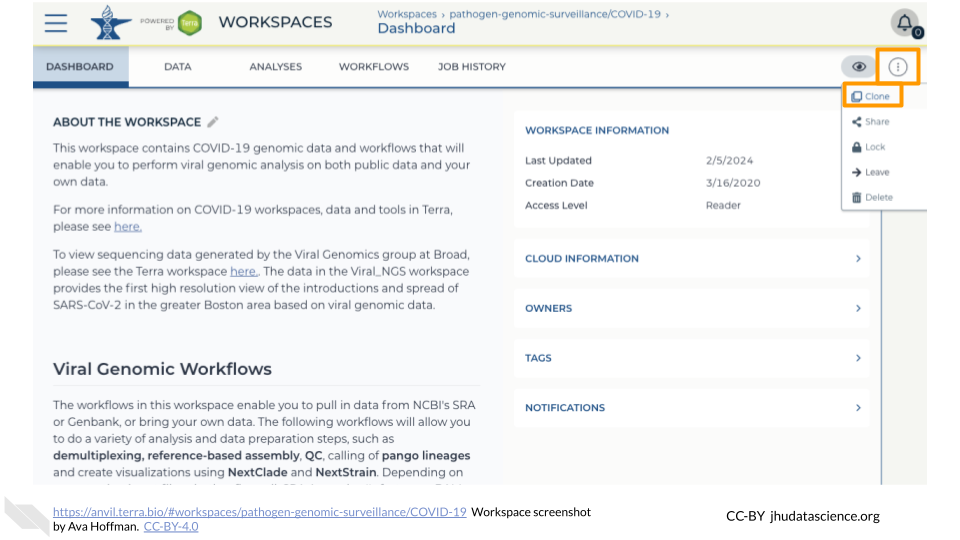
On AnVIL, you access files and computers through **Workspaces**. Each Workspace functions almost like a mini code laboratory - it is a place where data can be examined, stored, and analyzed. The first thing we want to do is to copy or “clone” a Workspace to create a space for you to experiment. This will give you access to

* the files you will need (data, code)
* the computing environment you will use

**Tip** At this point, it might make things easier to open up a new window in your browser and split your screen. That way, you can follow along with this guide on one side and execute the steps on the other.

To clone an AnVIL Workspace:

1. Go to the Workspace by clicking this link: <https://anvil.terra.bio/#workspaces/gdscn-exercises/Investigating-Antimicrobial-Resistance>.
2. Clone the workspace by clicking the teardrop button (). Select “Clone”. See the screenshot below on a different Workspace:

* 

1. You will see a popup box appear, asking you to configure your Workspace
   1. Give your Workspace clone a name by adding an underscore (“\_“) and your name. For example, "Investigating-Antimicrobial-Resistance\_Firstname\_Lastname".
   2. Select the Billing Project provided by your instructor.
   3. Leave the bottom two boxes as-is.
   4. Click “CLONE WORKSPACE”.

* See the screenshot below on a different Workspace:
* 

1. The new Workspace should now show up under <https://anvil.terra.bio/#workspaces>. This is your own copy of the Workspace to work in.

# 1 References