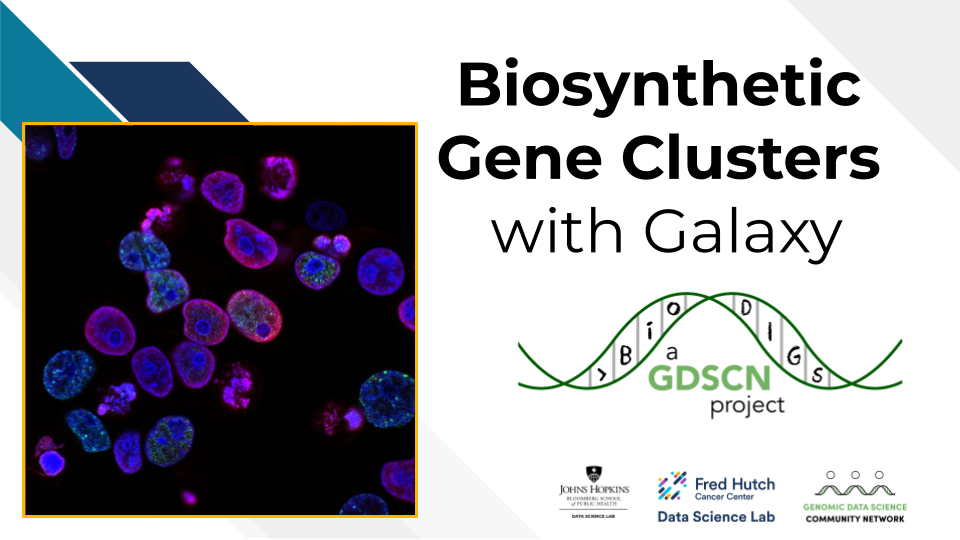
BioDIGS: Biosynthetic Gene Clusters

May 05, 2025

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 explore **microbiome data for the presence of biosynthesis genes**. They will run analyses using **antiSMASH 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*  
**Beginner**: some genetics knowledge needed

*Programming skills*  
**Novice**: no programming experience needed

## 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/).

# 1 Student Guide

## 1.1 Activity One

You might want to create a student guide that contains a different subset of Rmd files from your book, or renders to a different output format (e.g. word document). You can specify the output and Rmd files that will be used for the student guide using the \_output.yml and \_bookdown.yml files in the student-guide directory.

## 1.2 Activity Two

Steps of the guide could go here.

# 2 References