

# High throughput Python pipeline to identify Horizontal Gene Transfer

Student: Arun Dash, University of Arizona  
arundash@arizona.edu

Mentor: Dr. Simon Delattre, Penn State University

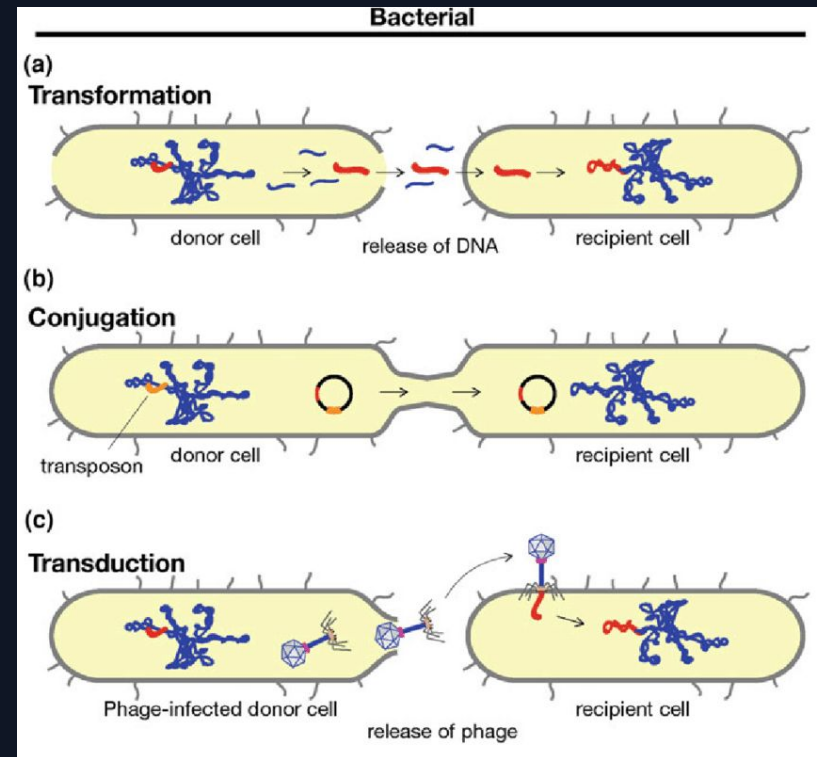
Researcher: Dr. Vinayak Mathur, Cabrini University

3/8/2023



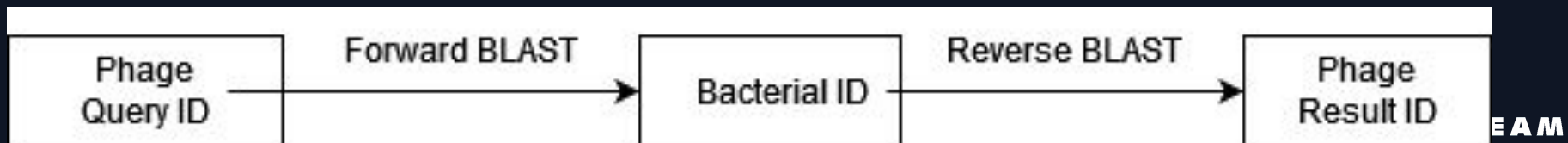
# Project Description

- The purpose was to use BLAST to identify instances of horizontal gene transfer between bacteria and bacteriophages.



# Project Goals

- The main goal of the project was to speed up the process.
  - This goal was accomplished by locally downloading the database and using a tool based on the command line to query it.



# Project Timeframe

- Start: September 2022
- End: March 2023



# Project Accomplishments

- The major accomplishment of this project was making the analysis pipeline work faster with Diamond.
- The original version of the pipeline was a long Python script that took about 5 minutes per query.



# Lessons Learned

- What went well? What could we have done differently?
  - Increased familiarity with working in HPC environment, using command line tools and bash commands.
  - Should have asked for more help early and often.



# Next Steps in Continuation of Research

- Filtered searches of nr database with diamond, to limit them by taxonomic groups.
- Analysis of the large list of pairs of bacteria and bacteriophages generated by the pipeline.

