Predicting ARG composition in Effluent samples based on Influents in WWTP

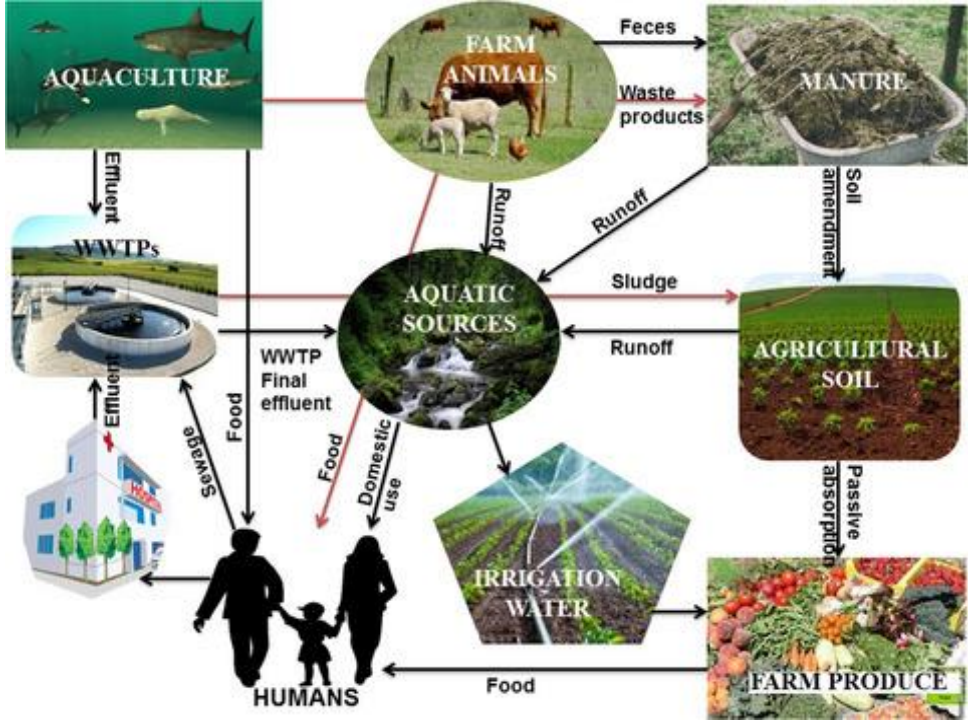
Presented by

Monjura Afrin Rumi

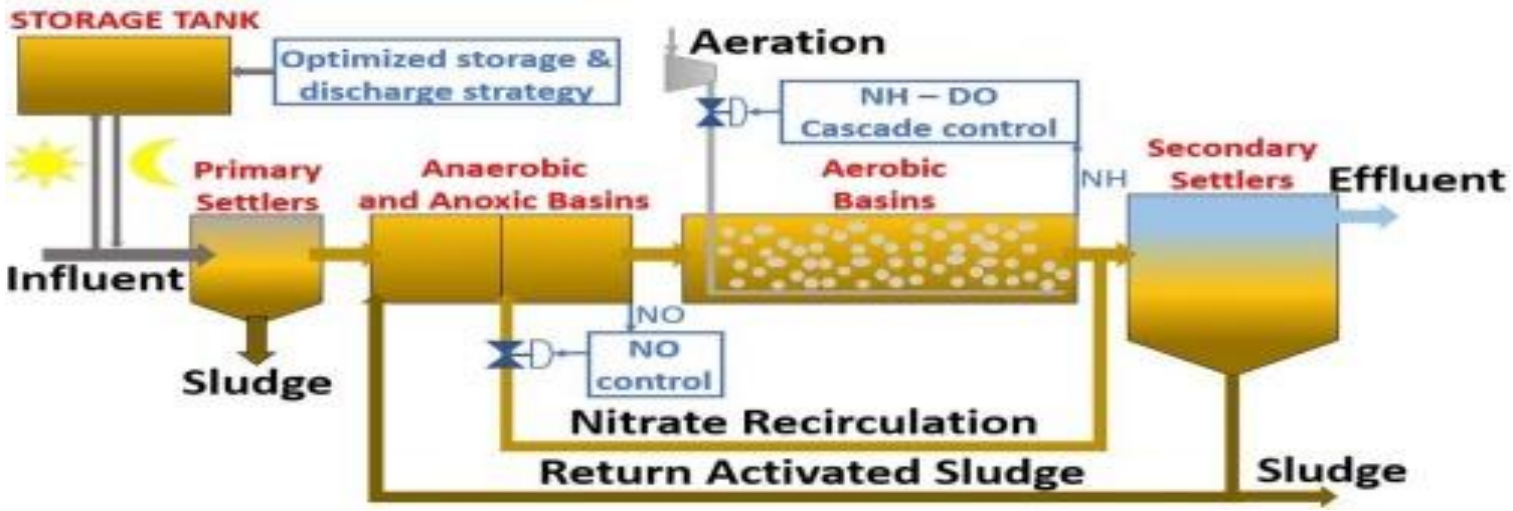
Background

- ARG – antibiotic resistance gene
- WWTP – wastewater treatment plant

1



2



1. Chidozie D. Iwu,Lise Korsten,Anthony I. Okoh. The incidence of antibiotic resistance within and beyond the agricultural ecosystem: A concern for public health. Microbiology, Volume9, Issue9, September 2020.
2. Melinda Simon-Várhelyi, Vasile Mircea Cristea, Alexandra Veronica Luca. Reducing energy costs of the wastewater treatment plant by improved scheduling of the periodic influent load. Journal of Environmental Management, Volume 262,2020.

Motivation

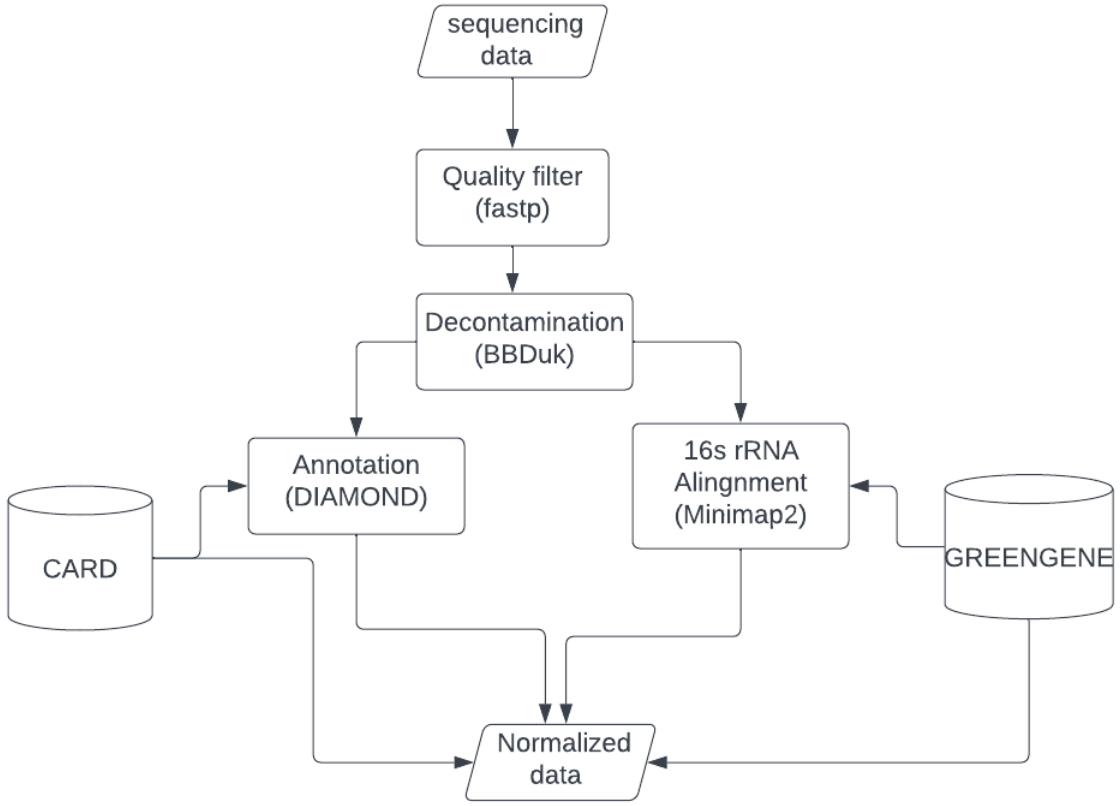
- Sequencing is expensive
- Influent vs effluent
- Derive ARG relationship in influent vs effluent
- Find ARG relationship with other environmental properties
- **Target: predict ARG abundances in effluent from influent**

Data Collection

- Christianburg, VA
- October, 2020 – September, 2021
- 192 effluent samples and 224 influent samples

Data preprocessing

$$\text{Abundance} = \sum_{1}^{n} \frac{N_{\text{ARG-like sequence}} \times L_{\text{reads}} / L_{\text{ARG reference sequence}}}{N_{\text{16S sequence}} \times L_{\text{reads}} / L_{\text{16S sequence}}}$$



Learning objective

- Target: predict ARG abundances in effluent from influent
  - Number of sample ~200
  - Number of genes ~200

- Baseline: linear regression
- Improvement
  - Predict a subset of genes
  - SVM, RF, ANN

Questions