Plasmid Prediction and Gene Regulation Inference

PhD Research Progress and Results

Christian D. Basile

April 11, 2025

Plasmid Prediction Workflow

Steps Taken:

- Integrated PlasmidFinder results with BLAST hits and sequence metrics.
- Generated DataFrames for single (DF) and paired contigs (DFpaired).
- Computed total sequence and BLAST-based costs.
- Developed rules to predict plasmid count and arrangement across bacterial strains.

Strain-Level Analysis

KH5 Strain:

- Paired plasmid structure evaluated using combined metrics.

KH2_17 Strain:

- Rule-based predictions refined.

AL_12 Strain:

- Working on prediction validation with incomplete BLAST and replicon data.

Prediction Strategy Overview

Key Ideas:

- Normalize BLAST and sequence scores by contig length.
- Apply distance-based equations to assess contig pairs.
- Enable fallback strategies if BLAST or replicon hits are missing.

Future Enhancements

- Integrate k-mer frequency data and statistical priors.
- Encode gene order and orientation into matching rules.
- Partial scoring for fragmented or ambiguous contigs.
- Move toward generalization across bacterial species.

Gene Regulatory Network Inference

GNN-Based Approach:

- Developed a Graph Neural Network (GNN) to infer gene regulatory relationships.
- Used transcriptomic and perturbation data as graph node features.
- Outperformed rule-based methods like cBONITA in accuracy and generalizability.
- Captures nonlinear interactions and topology-aware patterns.

Model Advantages

- Learns structure-function links from perturbation experiments.
- Improves inference quality in noisy or sparsely labeled datasets.
- Potential to generalize across bacterial and eukaryotic systems.

Summary

- Workflow built to predict plasmid structure using DNA sequencing and rule-based logic.
- GNN approach shows promise in capturing regulatory networks from expression data.
- Continued work focuses on improving prediction robustness and extending generalizability.

Thank You

Thank you for your attention.
Questions or feedback welcome!

`christian.daniel.basile@gmail.com`