CABBI
CENTER FOR ADVANCED BIOENERGY AND BIOPRODUCTS INNOVATION

# pyFLANK, a Graph Neural Network Based Null Distribution Inference Model for $F_{ST}$ Outlier Detection

## Background/Objective

Detecting genomic regions under selection is a fundamental goal in population genetics, yet it is often confounded by complex demographic histories. The objective was to develop pyFLANK, a Graph Convolutional Network (GCN) based model that infers a more robust null distribution by accounting for the local genomic landscape. By modeling the relationships through embedding space among variations, the tool aims to reduce false positives and improve the accuracy of $F_{ST}$ outlier detection.

## Approach

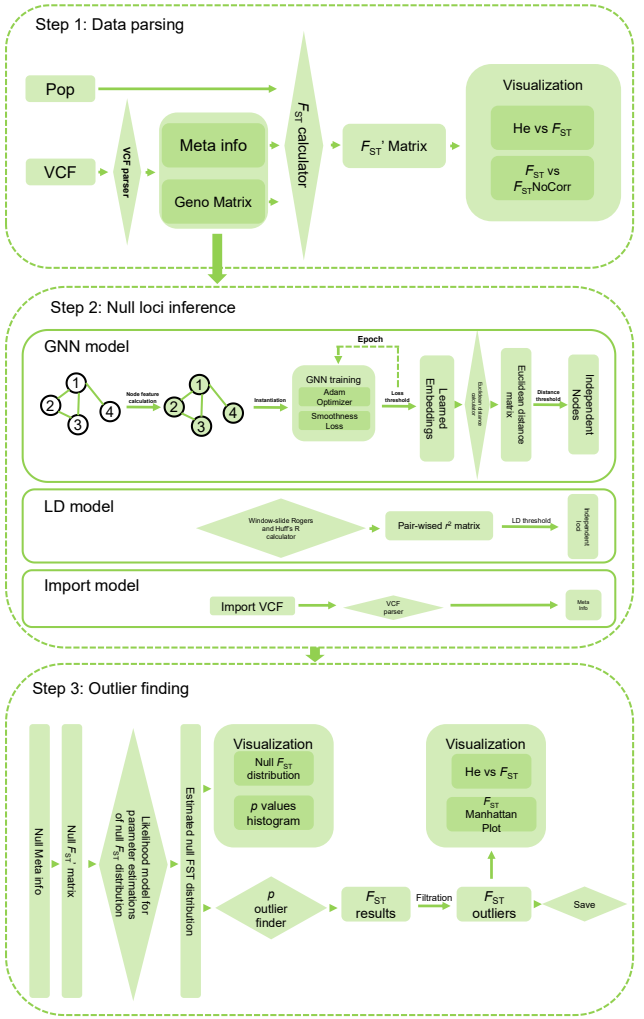
- Researchers developed pyFLANK, a model using GCN to infer the null distribution of $F_{ST}$ values.
- Unlike traditional outlier detection methods that rely on strict prior assumptions, pyFLANK uses a weighted undirected graph where genetic variations are nodes and correlation between the corresponding genotypes are edges.
- The model was validated using simulated datasets and real-world genomic data from cattle and yeast populations.

## Results

- pyFLANK achieved up to 1700× faster execution, lower CPU usage than traditional methods, and lower false positive rates while maintaining a comparable detection power.
- The tool successfully identified well-known selective sweeps in cattle and 84 adaptive loci in yeast, capturing functional signals often missed by conventional approaches.

## Significance/Impacts

Our evaluation demonstrates the power of pyFLANK to accurately identify the $F_{ST}$ outliers, providing a robust and user-friendly tool for identifying loci under divergent selection. By robustly inferring the quasi-independent loci, it enables efficient identification of adaptive variants in crops, livestock, and microorganisms. Its scalability and speed provide a transformative solution for high-throughput population genomics, allowing researchers to pinpoint functional traits critical for environmental adaptation and agricultural improvement.




**Overview of the pyFLANK workflow.**

Zhang et al. 2026. "pyFLANK, a Graph Neural Network Based Null Distribution Inference Model for $F_{ST}$ Outlier Detection." *BMC Bioinformatics*. DOI: 10.1186/s12859-026-06430-2.