Oct 29, 2025

Computational workflow for the study “PBP2 as the Predominant Siderophore Recognizer in Bacillus”

DOI
https://dx.doi.org/10.17504/protocols.io.eq2ly45dmlx9/v1

Linlong Yu[1]

[1]Peking University

LinlongYu

Linlong Yu

DOI: https://dx.doi.org/10.17504/protocols.io.eq2ly45dmlx9/v1

Protocol Citation: Linlong Yu 2025. Computational workflow for the study “PBP2 as the Predominant Siderophore Recognizer in Bacillus”. **protocols.io** https://dx.doi.org/10.17504/protocols.io.eq2ly45dmlx9/v1

Protocol status: Working
We use this protocol and it's working

Created: October 29, 2025

Last Modified: October 29, 2025

Protocol Integer ID: 231023

Keywords: siderophore biosynthetic gene cluster, pbp2 as the predominant siderophore recognizer, biosynthetic gene cluster, predominant siderophore recognizer, bacillus, genome retrieval, genome mining, sequence clustering, siderophore, phylogenetic reconstruction, pbp2, motif scanning, taxonomic classification, computational workflow for the study

Abstract

This protocol describes the complete computational workflow used to identify, classify, and analyze siderophore biosynthetic gene clusters (BGCs) and their cognate receptors. All analyses were performed using version-controlled scripts and publicly available databases to ensure full reproducibility. The workflow includes genome retrieval, taxonomic classification, genome mining, phylogenetic reconstruction, sequence clustering, mutual information analysis, and motif scanning.

Guidelines

4. Define feature sites as residues with MI values greater than 85% of the theoretical maximum MI in Bacillus.

Step 7: Motif scanning

1. Construct a Fur-binding PWM using 19 validated *Bacillus subtilis* motifs via **MEME Suite**. Obtain the DmdR1 PWM from the **LogoMotif** database.
2. Scan upstream regions (150 bp) of all genes using **FIMO** ($p < 0.001$).
   - Fur PWM applied to *Bacillus* genomes
   - DmdR1 PWM applied to *Streptomyces* genomes

Materials

Category	Tool / Database	RRID
Genome database	NCBI RefSeq	SCR_003496
Taxonomy assignment	GTDB-Tk (release R220)	SCR_019136
Phylogenetic reconstruction	PhyloPhlAn3	SCR_013082
Genome mining	antiSMASH v7.0.0	SCR_022060
Known cluster comparison	MIBiG database	SCR_023660
Domain identification	HMMER v3.0 / Pfam	SCR*005305 / SCR*004726
Sequence alignment	Clustal Omega	SCR_001591
Statistical analysis	MATLAB R2024a	SCR_001622
Motif discovery / scanning	MEME Suite v5.5.7	SCR_001783
Structure annotation	DSSP	SCR_016067

Procedure

1 Retrieve complete genomes from the NCBI RefSeq database (as of January 21, 2025).
Assign taxonomy using GTDB-Tk (release R220).

2 Run antiSMASH v7.0.0 on all annotated genomes.
Identify BGCs annotated as *NRP-metallophore* or *NIS-siderophore*.
Compare predicted clusters against the MIBiG database (similarity ≥80%) using the knownclusterblast module.

3 Use PhyloPhlAn3 to extract and align 400 universal marker genes.
Build a phylogenetic tree and visualize it with the R package *ggtree*.

4 Extract all core biosynthetic genes from siderophore BGCs.
Compute pairwise *p*-distance values.
Apply hierarchical clustering; determine the optimal threshold using silhouette scores.

5 Compute pairwise distance matrices for biosynthetic and candidate genes.
Calculate Pearson correlation coefficients in MATLAB using the *corr* function.
Define feature sites as residues with MI values 3e85% of the theoretical maximum in *Bacillus*.

6 Collect six experimentally validated siderophore receptors from Bacillus species based on published literature.
Use BLASTP (RRID:SCR_001010) to identify homologous receptor sequences across all Bacillus genomes.
Perform multiple sequence alignment and calculate mutual information (MI) between alignment positions and receptor identity labels.
Define feature sites as residues with MI values greater than 85% of the theoretical maximum MI in *Bacillus*.

7 Construct a Fur-binding PWM using 19 validated *Bacillus subtilis* motifs via **MEME Suite**.
Obtain the DmdR1 PWM from the **LogoMotif** database.
Scan upstream regions (150 bp) of all genes using **FIMO** (p 3c 0.001).
Fur PWM applied to *Bacillus* genomes
DmdR1 PWM applied to *Streptomyces* genomes

Acknowledgements

Data and Code Availability

All genomes are available from NCBI RefSeq (RRID:SCR_003496).
All scripts and analysis pipelines are publicly available on GitHub: https://github.com/Linlong-Yu/PBP2-as-the-Predominant-Siderophore-Recognizer

Expected Output

- Annotated siderophore BGCs
- Hierarchical clustering dendrograms of biosynthetic genes
- Coevolutionary correlation matrices
- Predicted Fur- and DmdR1-binding site lists and motif logos

Version Control

All analyses were executed using version-controlled scripts with fixed random seeds to ensure reproducibility.