# Detailed Data Requirements for PhageMatch Prototype

*Note: The genomic data (complete sequences and annotations) is already available. The following specifications focus on the complementary data needed.*

## 1 Phage–Host Interaction Data

### 1.1 Interaction Matrix/Database

- **Format:** A structured file (e.g., CSV or Excel) linking phage IDs to bacterial strain IDs.
- **Columns to Include:**
  - Phage ID
  - Bacterial Strain ID
  - nteraction Outcome (e.g., binary effective/ineffective or quantitative measures like Efficiency of Plating [EOP] or Plaque Forming Units [PFU])
  - Statistical data (e.g., number of replicates, p-values/confidence intervals for interaction significance)

### 1.2 Experimental Protocol Details

- **Assay Description:** Detailed protocol of how phage–host interactions were measured (e.g., plaque assays).
- **Key Parameters:**
  - Multiplicity of Infection (MOI)
  - Incubation time and temperature
  - Media composition and any additives
- **Instrumentation**
  **Analysis:**
  - Details of any imaging or automated plaque counting systems used
  - Information on the sample size and replicates, including control experiments

# 2 Proteomic Data

## 2.1 Protein Sequence Data

- **Format:** FASTA files for protein sequences from both phages and bacteria.
- **Annotations:** Include protein names and functional descriptions where available.

## 2.2 Protein Expression Data

- **Experimental Method:** Quantitative data obtained through proteomics (e.g., LC-MS/MS).
- **Raw**
  **Processed Data:**
  - Raw data files (such as mzML or vendor-specific formats)
  - Processed data tables (CSV/Excel) detailing protein abundance levels, including label-free quantification (LFQ) values or ratios from isobaric tagging (iTRAQ, TMT)
- **Experimental Details:**
  - Sample preparation (e.g., trypsin digestion protocols, fractionation details)
  - Chromatography conditions and mass spectrometer model (e.g., Orbitrap or Q-TOF)
  - Information on controls (non-infected samples) and time-course data if available
- **Data Processing:**
  - Software and normalization methods used in data processing (e.g., MaxQuant, Proteome Discoverer)

# 3 Transcriptomic Data (If Available)

## 3.1 RNA Sequencing Data

- **Raw Data:** FASTQ files from RNA-seq experiments capturing bacterial response during phage infection.
- **Processed Data:**
  - Count matrices or normalized expression values (TPM, FPKM)
  - Differential expression analysis results (e.g., log fold changes, adjusted p-values)

### 3.2 Experimental Conditions Metadata

- **Design Details:**
  - Clear differentiation between control and phage-infected conditions
  - Time points of sample collection (e.g., early, mid, and late infection stages)
  - Number of biological replicates per condition
- **Library Preparation Sequencing:**
  - Protocols used for RNA extraction and library preparation
  - Sequencing platform details (e.g., Illumina HiSeq or NovaSeq) and targeted read depth
- **Data Analysis Pipeline:**
  - Brief notes on alignment (e.g., using STAR) and differential expression tools (e.g., DESeq2)

## 4 Environmental and Experimental Metadata

### 4.1 Assay Conditions

- **Detailed Experimental Parameters:**
  - Temperature, pH, and media composition during experiments
  - Incubation times and any specific environmental stressors (e.g., osmotic or oxidative stress)
  - Bacterial growth phase (log vs. stationary) at the time of infection
  - Specific details regarding MOI and any pre-treatment protocols

### 4.2 Protocol Documentation

- **Standard Operating Procedures (SOPs):**
  - Documents or files that describe sample handling, storage conditions, and overall experimental workflows
  - Metadata files that annotate how each data point was generated and any normalization or quality control steps applied

## 5 Additional Phenotypic Data (Optional but Valuable)

### 5.1 Bacterial Phenotypic Profiles

- **Data Types:**

  - Antibiotic resistance profiles, growth curves, or any observable morphological changes under phage infection

- **Format:**
  - Tables or spreadsheets clearly correlating experimental conditions with phenotypic outcomes
  - Descriptive notes on measurement methods and any standardized scales used