**Steps Taken for the Creation of One Unified Dataset from the Files Containing Information Regarding the Frequency of Antimicrobial Resistance Genes in Each Bacterial Isolate, and the Frequency of Prophage Communities in Each Bacterial Isolate**

Step 1: Preliminary Review

Objective: Familiarise yourself with the content and structure of both Excel files.

Action:

1. Open both files separately.
2. Note the key columns, especially those for genome identifiers and the data columns (AMGs, non-AMGs, phage communities) you intend to combine.

Step 2: Data Cleaning

Objective: Ensure the data is clean and structured for easy merging

Actions:

1. In each file, review and remove any extraneous columns or rows that are not needed for analysis.
2. Ensure that genome identifiers are consistent across both files (e.g. formatting, prefixes).

Step 3: Data Preparation

Objective: Organise your data to facilitate the merge   
Actions:

1. Create a new Excel workbook for your combined dataset
2. Copy the contents of the AMG dataset into a sheet in the new workbook. Label this sheet appropriately (e.g. AMGs Data)

Step 4: Merging Data

Objective: Combine the AMG data with the Phage Communities data

Actions:

1. Open the Phage Communities datasheet
2. Copy the relevant phage community data
3. Paste this data into a new sheet in your combined dataset workbook, and label it (e.g. ‘Phage Communities’).
4. Use Excel’s VLOOKUP or INDEX/MATCH functions to align the who databases based on genome identifiers

EXAMPLE: =VLOOKUP([Genome Identifier], [Phage Communities Range], [Column Number], FALSE)

Replace ‘[Genome Identifier]’ with the cell reference for the Isolate IDs in your AMG data sheet. ‘[Phage Communities Range]’ should be the range of cells in the Phage Communities sheet that includes both Isolate IDs and the data you wish to merge. ‘[Column Number]’ is the number of the column in the range that contains the data you want to pull into your AMG sheet.

Step 5: Final Adjustments

Objective: Ensure the combined dataset is accurate and complete

Actions:

1. Review the combined dataset for any errors or omissions, particularly I rows where the VLOOKUP formula may not have found a match.
2. Adjust any formulas as needed to correct errors
3. Remove any temporary or helper columns that were used only for the merge process

Step 6: Saving and Sharing

Objective: Make your combined dataset accessible

Actions:

1. Save your workbook with an appropriate name (e.g. CombinedData)