

AMRFinder Class-Level Interpretation Documentation

This document includes script and expected output descriptions.

Script:

```
library(dplyr)
library(tidyr)
library(readr)

# Load TSV instead of CSV
amr <- read_tsv("/data/internship_data/nidhi/aba/output/amrfinder_output/merged/merged_cleaned.tsv")

# Keep key columns
amr <- amr %>%
  select(isolate = species_id,
         gene = `Element symbol`,
         class = Class,
         subclass = Subclass)

# Presence/absence matrix
amr_binary <- amr %>%
  distinct(isolate, class) %>%
  mutate(present = 1) %>%
  pivot_wider(names_from = class,
              values_from = present,
              values_fill = 0)

# Summary table: isolates per class
total_isolates <- length(unique(amr$isolate))

amr_summary <- amr_binary %>%
  select(-isolate) %>%
  summarise(across(everything(), sum)) %>%
  pivot_longer(cols = everything(),
              names_to = "AMR_class",
              values_to = "No_of_isolates") %>%
  mutate(Percentage = round((No_of_isolates/total_isolates)*100, 1))

# Subclass analysis (ESBLs / Carbapenemases)
carb_summary <- amr %>%
  filter(grepl("CARBAPENEM", subclass, ignore.case = TRUE)) %>%
  distinct(isolate) %>%
  summarise(No_of_isolates = n()) %>%
  mutate(Percentage = round(No_of_isolates/total_isolates*100, 1),
         Type = "Carbapenemase")

esbl_summary <- amr %>%
  filter(grepl("CEPHALOSPORIN|ESBL", subclass, ignore.case = TRUE)) %>%
  distinct(isolate) %>%
  summarise(No_of_isolates = n()) %>%
  mutate(Percentage = round(No_of_isolates/total_isolates*100, 1),
         Type = "ESBL")

# Save outputs
write_tsv(amr_binary, "AMR_class_presence_absence.tsv")
write_tsv(amr_summary, "AMR_class_summary.tsv")
write_tsv(bind_rows(carb_summary, esbl_summary), "beta_lactam_subclasses_summary.tsv")
```

Expected Output Files:

```
beta_lactam_subclasses_summary.tsv
AMR_class_summary.tsv
AMR_class_presence_absence.tsv
```