

MLST IC Analysis Documentation

This document includes the full MLST IC analysis script and expected output files.

Expected Output Files:

MLST_with_IC.csv
IC_distribution.csv

Script:

```
library(dplyr)
library(readr)

input_file <- "/data/internship_data/nidhi/aba/output/mlst_output/mlst_results.csv"
mlst <- read_csv(input_file, col_names = FALSE)

colnames(mlst) <- c(
  "file", "species", "ST",
  "cpn60", "fusA", "gltA", "pyrG", "recA", "rplB", "rpoB"
)

mlst$ST <- as.character(mlst$ST)

ic_lookup <- list(
  IC1 = c("ST1", "ST10", "ST20", "ST81"),
  IC2 = c("ST2", "ST195", "ST208"),
  IC3 = c("ST3", "ST145"),
  IC4 = c("ST15"),
  IC5 = c("ST25"),
  IC6 = c("ST78"),
  IC7 = c("ST79"),
  IC8 = c("ST85"),
  IC9 = c("ST158")
)

assign_ic <- function(st_value) {
  st_label <- paste0("ST", st_value)
  for (ic in names(ic_lookup)) {
    if (st_label %in% ic_lookup[[ic]]) {
      return(ic)
    }
  }
  return("Unassigned")
}

mlst$IC <- sapply(mlst$ST, assign_ic)

write.csv(mlst, "MLST_with_IC.csv", row.names = FALSE)

ic_table <- mlst %>%
  count(IC, name = "Count") %>%
  mutate(Percentage = round((Count / sum(Count)) * 100, 2))

write.csv(ic_table, "IC_distribution.csv", row.names = FALSE)

cat("\nIC assignment complete.\n")
cat("Generated files:\n")
cat(" - MLST_with_IC.csv\n")
cat(" - IC_distribution.csv\n")
```