# Hetrosigosity measurment

# Load required libraries

library(vcfR)

library(ggplot2)

# Function to calculate heterozygosity

calculate\_heterozygosity <- function(vcf\_file\_path) {

vcf <- read.vcfR(vcf\_file\_path)

geno <- extract.gt(vcf)

het <- apply(geno, 2, function(x) sum(x == "0/1", na.rm = TRUE) / length(x))

return(het)

}

# File paths

fish\_vcf\_path <- "data/Fish/final.50.recode.vcf"

copepod\_vcf\_path <- "data/Copepod/final.50.recode.vcf"

# Calculate heterozygosity for fish and copepods

fish\_het <- calculate\_heterozygosity(fish\_vcf\_path)

copepod\_het <- calculate\_heterozygosity(copepod\_vcf\_path)

# Shapiro-Wilk Normality Test

cat("Shapiro-Wilk Normality Test Results:\n")

fish\_shapiro <- shapiro.test(fish\_het)

copepod\_shapiro <- shapiro.test(copepod\_het)

cat("Fish heterozygosity p-value:", fish\_shapiro$p.value, "\n")

cat("Copepod heterozygosity p-value:", copepod\_shapiro$p.value, "\n")

# Non-parametric Wilcoxon Rank-Sum Test

cat("\nWilcoxon Rank-Sum Test Results:\n")

wilcox\_test <- wilcox.test(fish\_het, copepod\_het, exact = FALSE)

print(wilcox\_test)

# Combine the data for plotting

group <- c(rep("Fish", length(fish\_het)), rep("Copepod", length(copepod\_het)))

heterozygosity <- c(fish\_het, copepod\_het)

data <- data.frame(Group = group, Heterozygosity = heterozygosity)

# Violin plot with boxplot overlay

ggplot(data, aes(x = Group, y = Heterozygosity, fill = Group)) +

geom\_violin(trim = FALSE) +

geom\_boxplot(width = 0.1, fill = "white", outlier.shape = NA) +

theme\_minimal() +

labs(title = "Heterozygosity Distribution: Fish vs. Copepods",

x = "Group",

y = "Heterozygosity") +

scale\_fill\_manual(values = c("Fish" = "skyblue", "Copepod" = "salmon"))

# Boxplot

ggplot(data, aes(x = Group, y = Heterozygosity, fill = Group)) +

geom\_boxplot(outlier.color = "red", outlier.shape = 16, notch = TRUE) +

theme\_minimal() +

labs(title = "Boxplot: Heterozygosity of Fish vs. Copepods",

x = "Group",

y = "Heterozygosity") +

scale\_fill\_manual(values = c("Fish" = "skyblue", "Copepod" = "salmon"))