

# Feature-Level Relative Abundance Error Analysis

- Address issues with var-mean: NA cats, and 0 values for rcov\_error
- summarize results
- feature count table

Outlier feature-level error-rates had lower median relative abundance, but not RCOV (Fig. 1). The median relative abundance was significantly lower for features identified as outliers based on the feature-level error rate, but only when considering features with positive error rates. This is likely due to the more extreme error rates all being positive. For the RCOV the feature-level median relative abundance values were not significantly different between the outlier and non-outlier features.

The error rate is dependent on the accuracy of the relative abundance estimates for the unmixed pre- and post-exposure samples. The feature-level median error-rate and RCOV was compared to the the unmixed sample variance/mean relative abundance to determine if extreme error-rate and RCOV values could be attributed to variability in relative abundance between PCR replicates for the unmixed samples. The variance/mean for the unmixed samples was lower for the outliers compared to the non-outliers for both the feature-level error rate and

Investigation of relationship between phylogeny and feature-level relative abundance error metrics (Fig. 3).

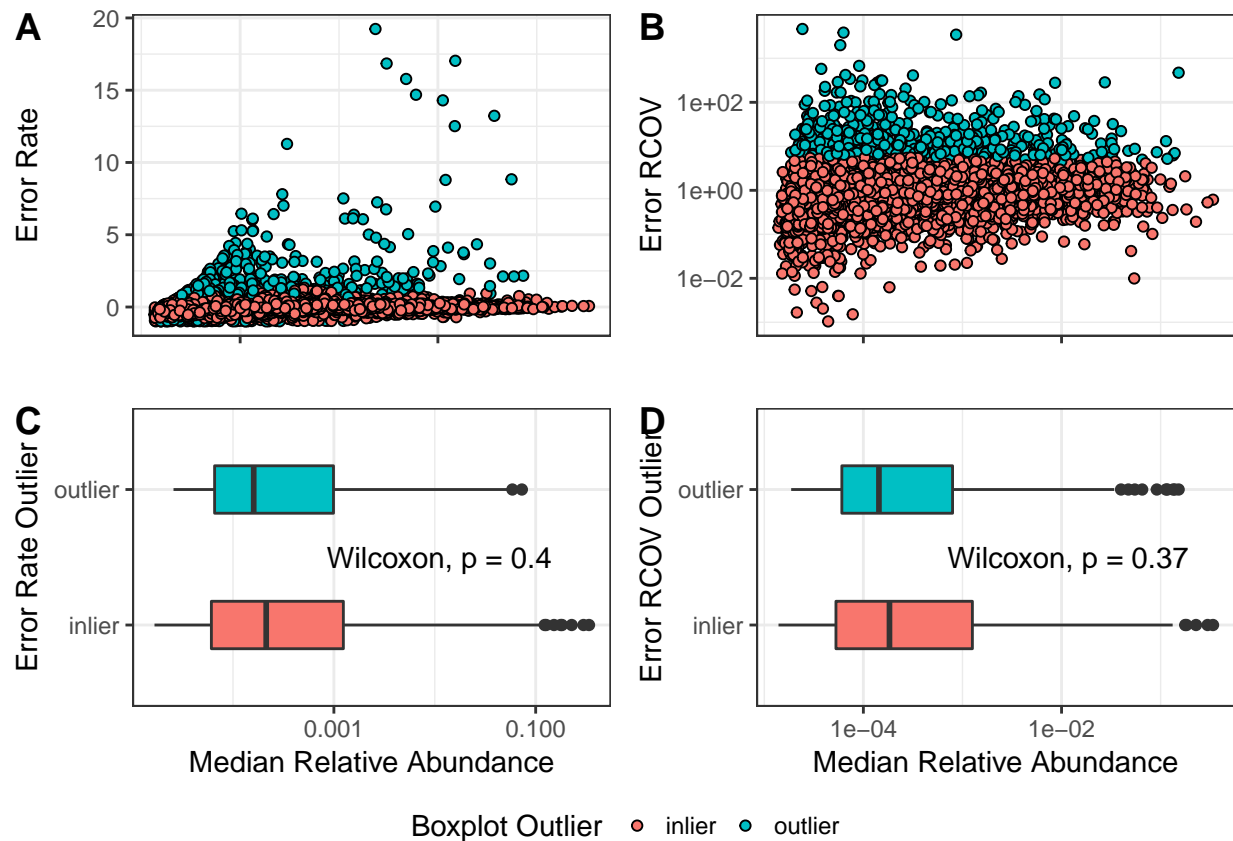


Figure 1: Feature-level median relative abundance is lower for error rate outlier features but not for RCOV outlier features. (A) Relationship between median relative abundance to the feature-level (A) median error rate and (B) error RCOV. Boxplots summarizing the feature-level median relative abundance between (C) error rate and (D) RCOV outlier and non-outlier features for features. Only features with median error rates greater than 0 included in the error rate boxplot.

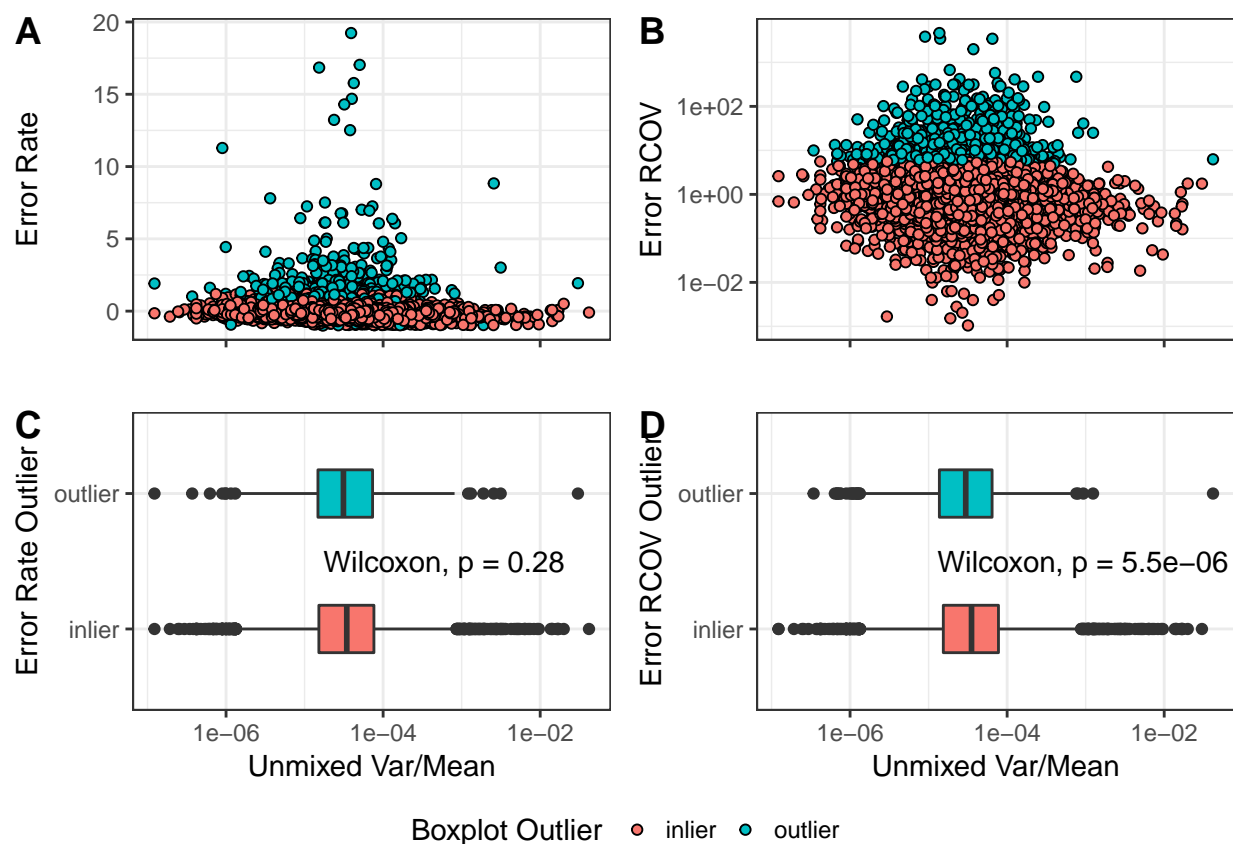


Figure 2: The variance-mean relationship for the unmixed sample relative abundance values does not explain outlier median and RCOV features. (A) Relationship between unmixed sample relative abundance variance/mean to the feature-level (A) median error rate and (B) error RCOV. Boxplots summarizing the unmixed sample relative abundance variance/mean between (C) error rate and (D) RCOV outlier and non-outlier features for features. Only features with median error rates greater than 0 included in the error rate boxplot.

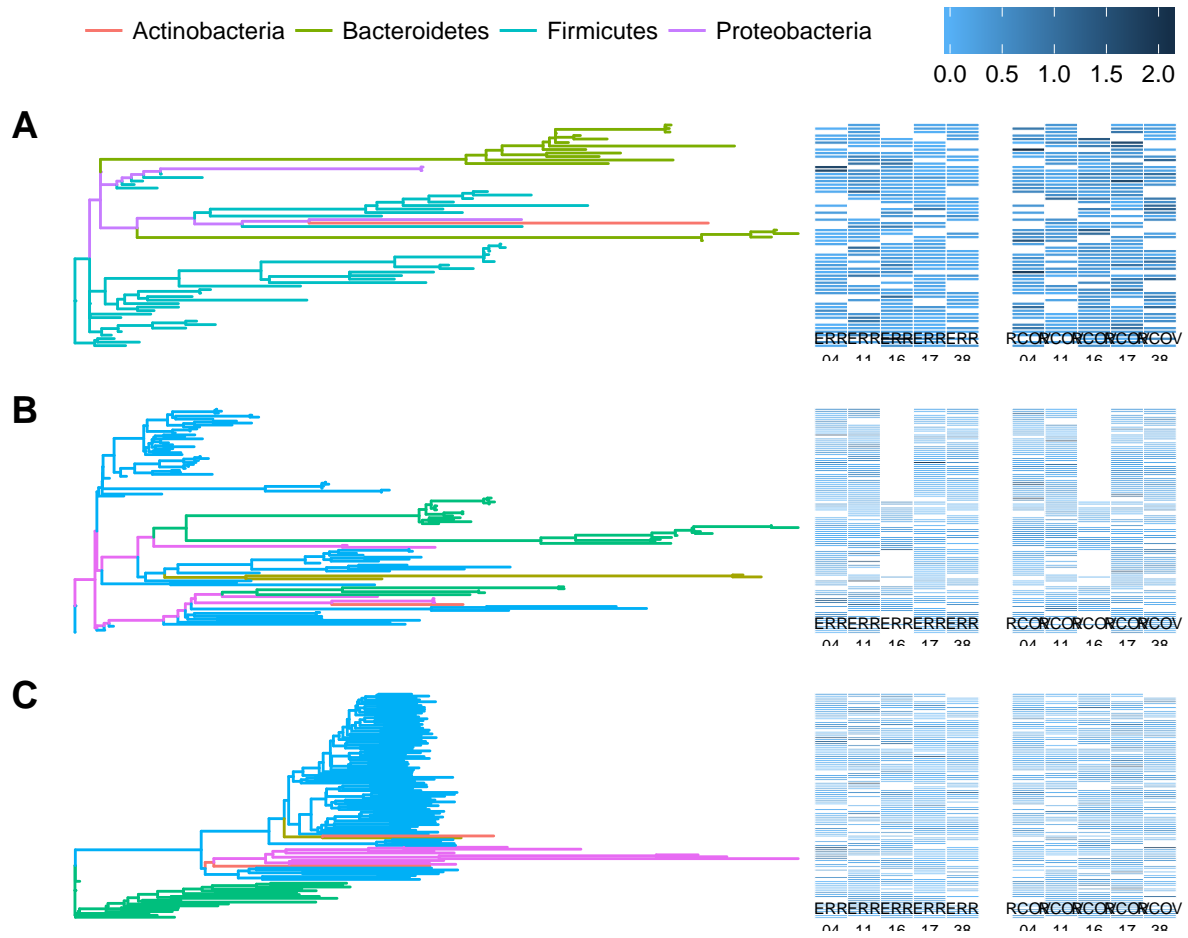


Figure 3: Phylogenetic analysis of feature-level relative abundance error metrics, as heatmaps across individuals. Subplots for individual pipelines (A) DADA2, (B) QIIME, (C) Mothur. Feature phylum assignment is indicated by branch color.