

Relative Abundance Error Rate

Nate Olson

2017-10-24

NOTE Not sure about interpretation of RCOV when using non-absolute feature-level error rate, how to interpret positive and negative values.

Overall agreement between the inferred and observed relative abundance was high for all individuals and bioinformatic pipelines (Fig. 1A). The pre- and post-exposure estimated relative abundance and inferred theta values were used to calculate titration and feature level error rates. Only features observed in all pre- and post-exposure PCR replicates and pre- and post-exposure specific features were included in the analysis (Table ??). Pre- and post-exposure specific features were defined as present in all four PCR replicates of the pre-exposure or post-exposure PCR replicates, respectively, but none of the PCR replicates for the other unmixed sample. There is lower confidence in the relative abundance of a feature in the pre- or post-exposure unmixed samples when the feature is observed in some of the 4 PCR replicates, therefore these features were not included in the error analysis. The deviation from the expected value on the low end varies by biological replicate and pipeline. Outliers are observed for all pipelines and individuals.

Next we evaluated the quantitative accuracy of the relative abundance values by comparing the distribution of the feature-level median error and feature-level robust coefficient (RCOV=(maximum - minimum)/median) of variation for the relative abundance error rate across pipelines (Fig. 1). Feature-level median error rates and RCOV were compared across pipelines and individuals using a mixed effects model. Large error rates and RCOV values were observed for all pipelines (Table 1). Features with large error rates, defined as $1.5 \times IQR$ from the median, were excluded from the analysis to prevent outliers from biasing the comparison. The mean feature-level error rate by pipeline was negative for all three pipelines (DADA2 -0.05, Mothur -0.06, and QIIME -0.12). Multiple comparisons test (Tukey) was used to test for significant differences in feature-level error between pipelines. A one-sided alternative hypothesis to determine which pipelines had a smaller, closer to zero, feature-level error rate. The Mothur and DADA2 mean feature-level error rate were closer to zero and significantly different from the QIIME pipeline, (Mothur v. QIIME $t = -4.2$, $p = NA$; DADA2 v. QIIME $t = -4.57$, $p = NA$). Though the Mothur and DADA2 mean feature-level error rates were not significantly different from each other, $t = -0.68$, $p = NA$. Unlike feature-level error rates, large RCOV was observed for all individuals and pipelines (Table 1). Outlier values were also excluded from the RCOV analysis. The feature-level RCOV was not significantly different between pipelines, Mothur = -3.34, QIIME = 1 and DADA2 = -1 (Fig. 1C).

Table 1: Maximum feature-level error rate bias (median error rate) and variance (robust COV) by pipeline and individual.

Metric	Pipeline	E01JH0004	E01JH0011	E01JH0016	E01JH0017	E01JH0038
Median	dada2	2.37	2.55	17.03	4.34	0.54
	mothur	5.43	6.76	19.24	4.15	1.93
	qiime	3.99	6.07	8.83	4.80	1.25
	unclustered	6.45	7.24	16.85	4.37	3.23
RCOV	dada2	10.30	165.34	25.38	48.45	25.34
	mothur	17.94	20.63	43.74	117.16	175.95
	qiime	470.91	409.21	58.33	144.63	107.27
	unclustered	3442.72	3835.29	249.11	65.58	167.95

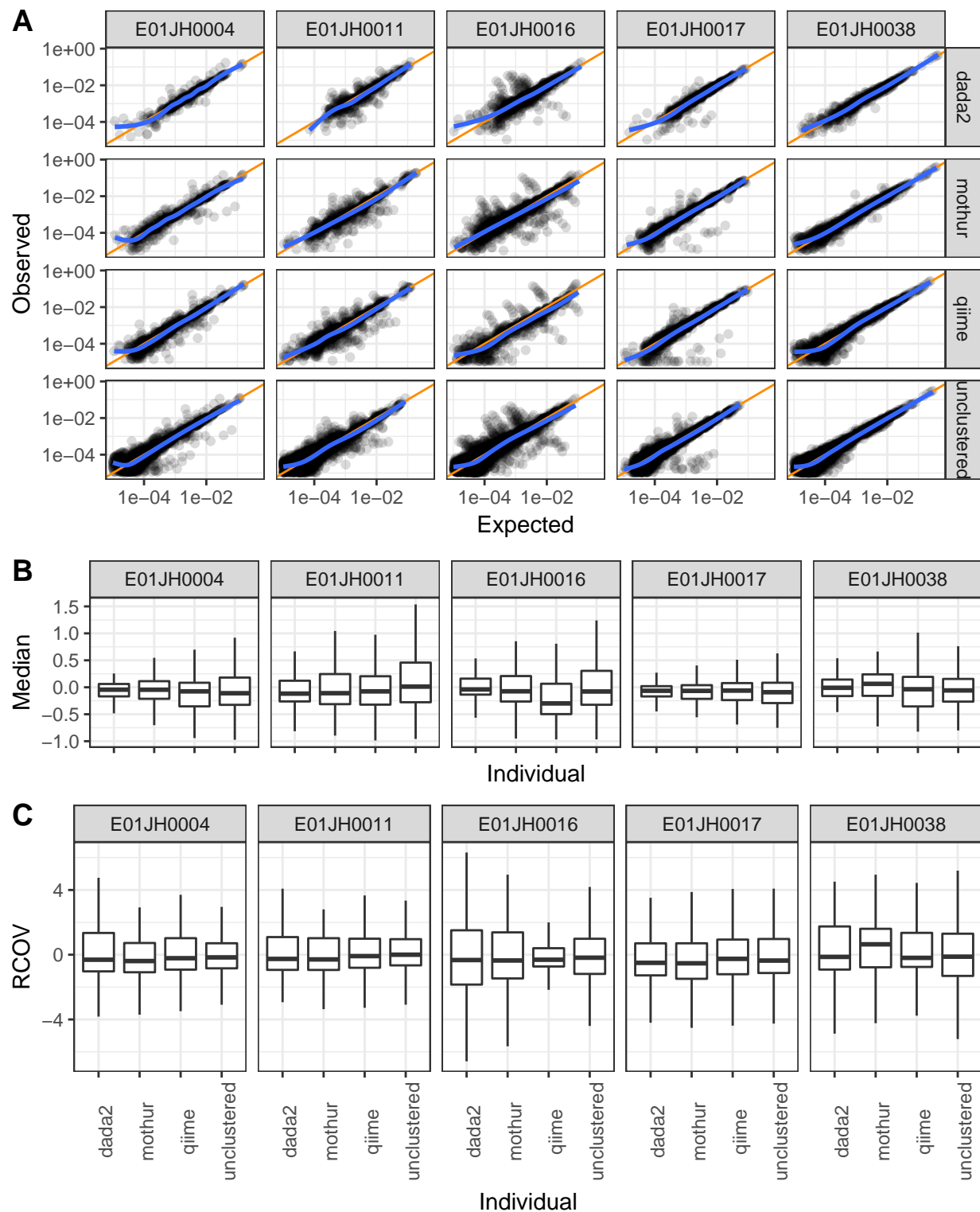


Figure 1: (A) Expected and observed count relationship. Orange line indicates expected 1-to-1 relationship. Blue line a smoothed regression line of the observed and expected value relationship. Distribution of feature-level relative abundance (B) median error rates and (C) robust coefficient of variation (RCOV) by individual and pipeline.