

Results Supplement

NGN Team

20221019

Field Sampling

Laboratory Processing

Bioinformatics Processing

Quality Controls

Positive controls were included on each sequencing run to monitor for cross contamination that might have occurred in the laboratory or due to “tag jumping”. With 13 MiSeq runs, we included one sample of kangaroo tissue on each run and then measured how many reads of kangaroo were found in environmental samples and how many reads of non-kangaroo were found in kangaroo samples.

Annotation

We first used a tree-based annotation method (insect package) and then followed up with a BLAST search for all ASVs that were not annotated to species level by insect.

```
{r FIG_annotationstats, fig.cap="\label{fig:annotationstats}  
Number of reads and ASVs annotated to species level.", fig.path=here("Output", "Figures", "annotationstats.pdf"),  
echo=F, message=F, warning=F, include = T, fig.height=4, dpi =  
300} # knitr::include_graphics(here("Output", "Figures", "annotationstats.pdf"))  
auto_pdf = TRUE) #
```

Correcting metabarcoding data for amplification bias

Using our six mock communities (three different taxa compositions x two proportions [even and skewed]), we can first check how well the quantitative metabarcoding model corrects for amplification bias. In one case, we consider the even mock communities as the mock community data and the skewed mock communities as unknown. We can then re-create what the model believes to be the original starting proportions of the skewed mock community given the proportions of reads found in the skewed mock communities and the proportion of DNA as compared to the proportion of reads found in the even mock communities. We can also do the same treating the skewed mock communities as known and even mock communities as unknown.

Correcting DNA concentrations for varying flow in creeks

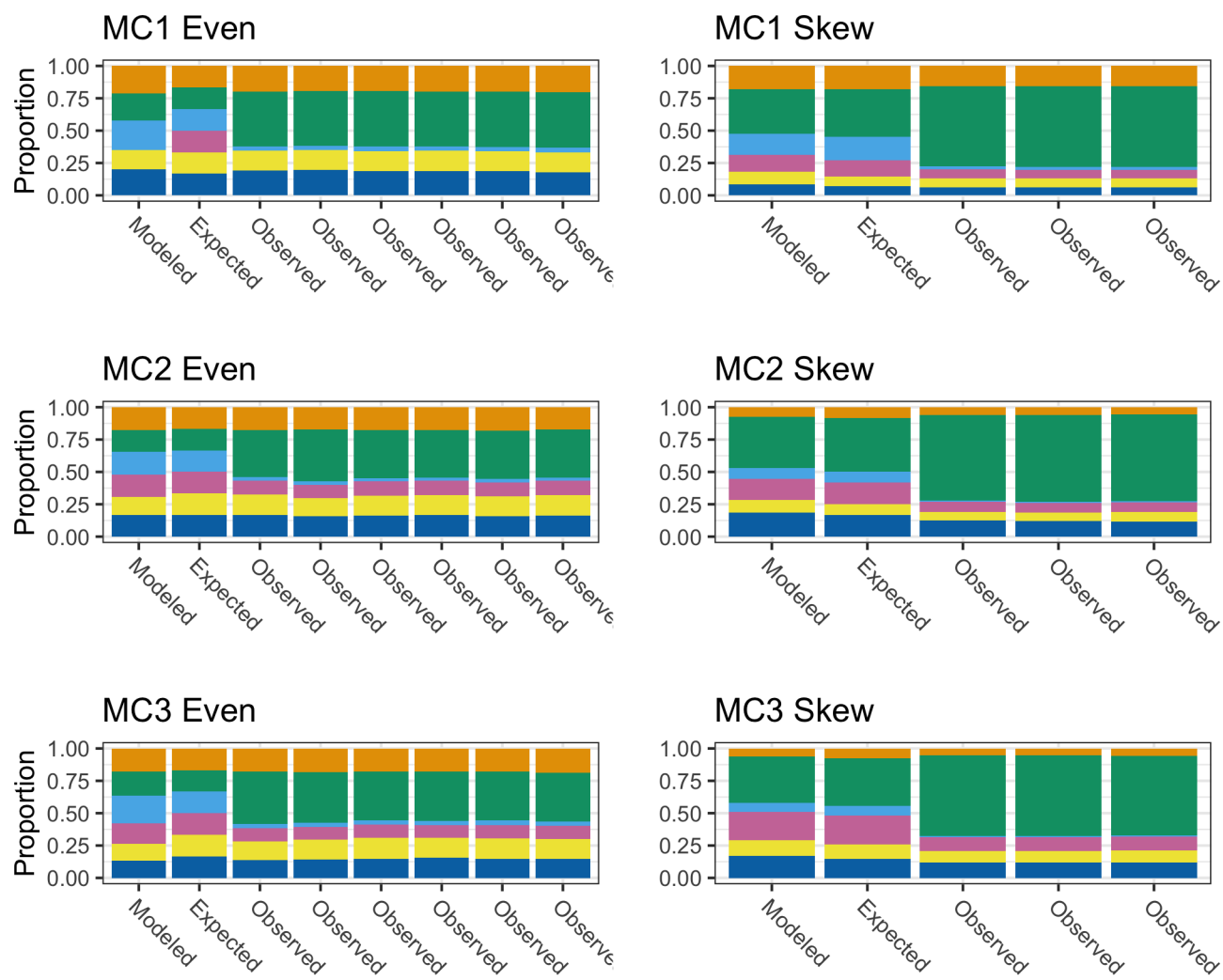


Figure 1: Intercalibration of mock communities used to correct environmental samples for amplification bias.

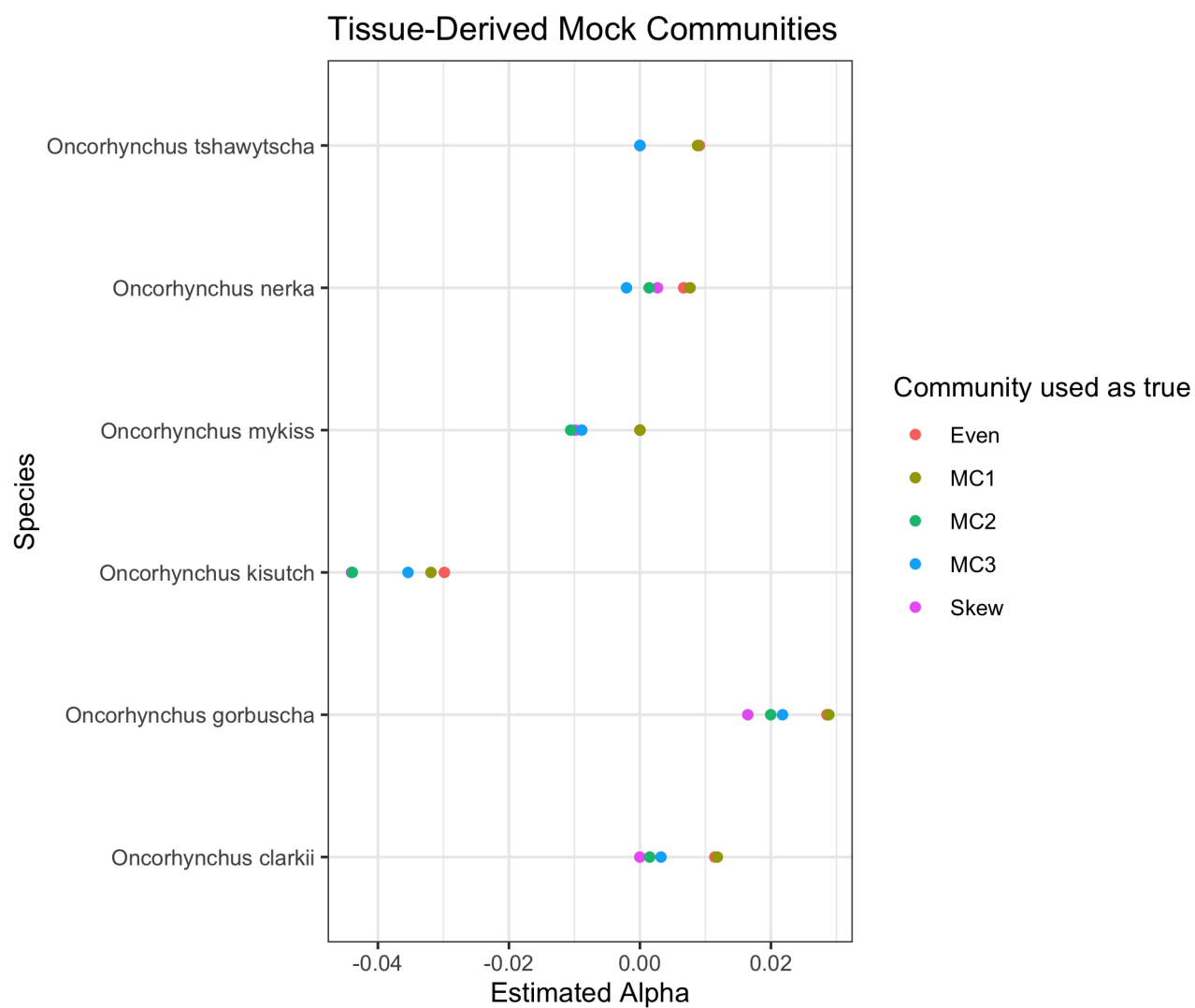


Figure 2: Estimated alpha values of salmonid species with different calibrations of the mock communities. Each color represents a different subset of mock community data treated as 'true' to calibrate the remainder of the mock community data.

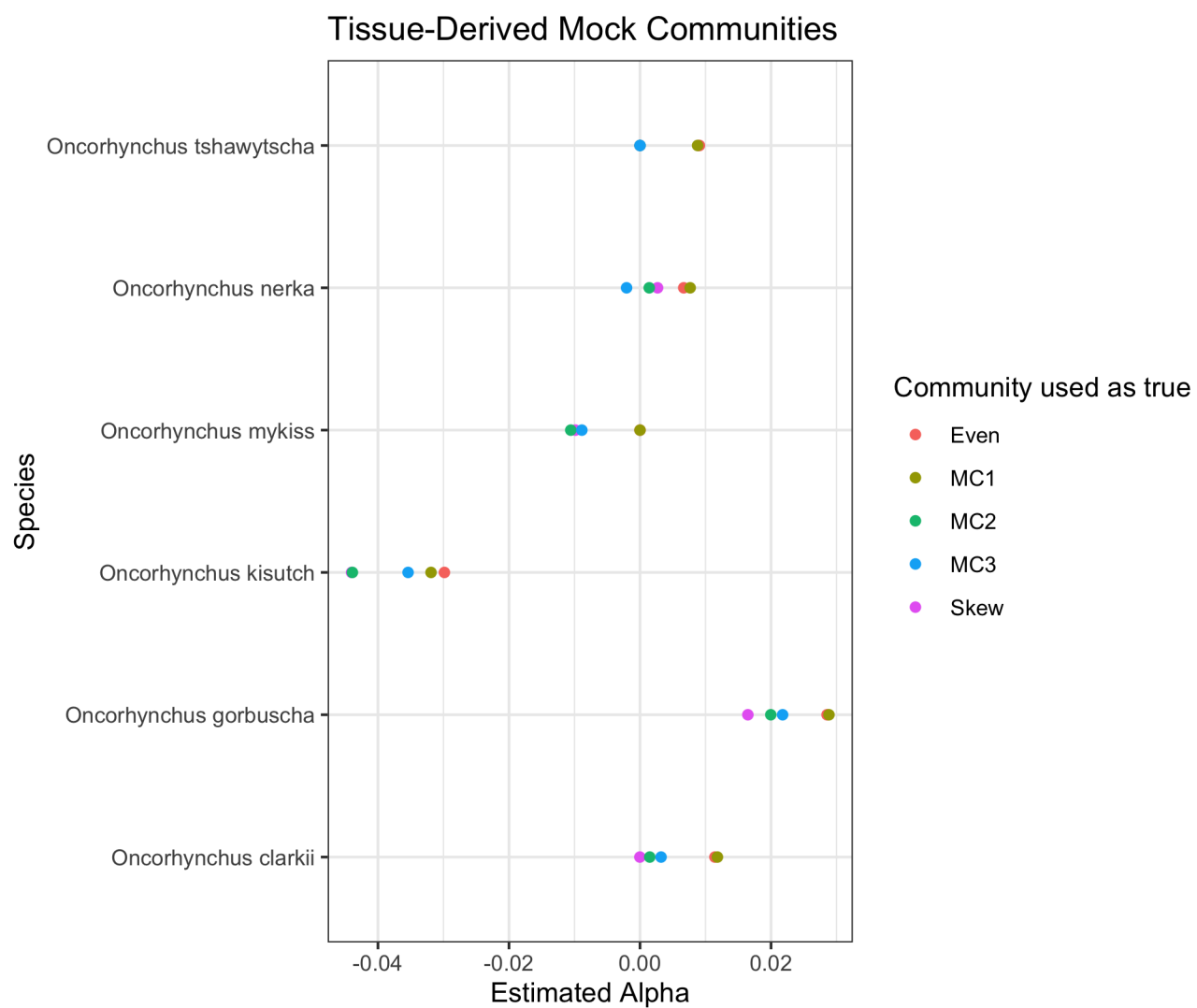


Figure 3: Estimated alpha values of salmonid species with different calibrations of the mock communities. Each color represents a different subset of mock community data treated as 'true' to calibrate the remainder of the mock community data.