

Supplemental Information Figure 4. A subset of MMETSP samples.

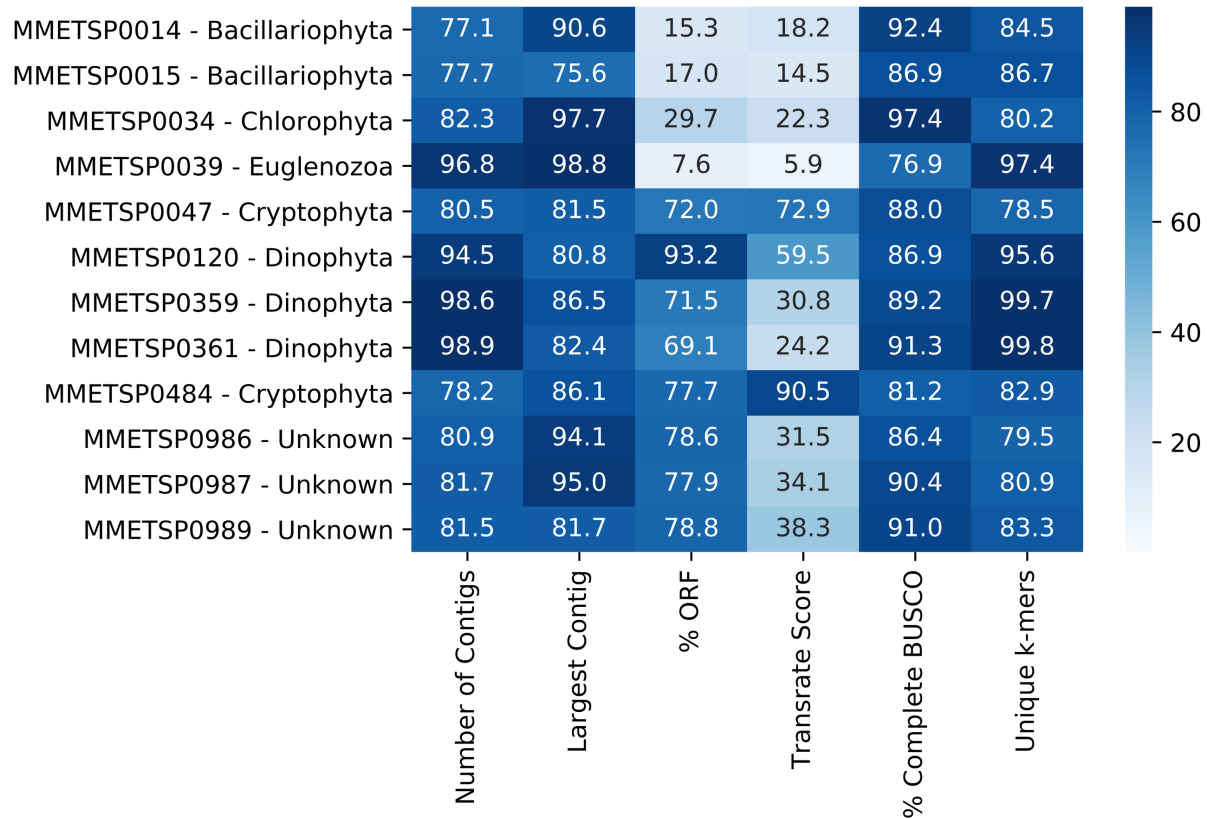
Based on the evaluation metrics collected in Johnson et al. 2018, we created a subset of the MMETSP collection which includes low and high performing samples. The idea is that this subset can act as a benchmarking dataset which can be used to assess future pipelines. Although the evaluation metrics collected from the collection was useful in setting a reasonable expectation for an expected distribution of evaluation metrics from a project of this size, since the entire MMETSP collection consists of 678 samples, it is not feasible for it to be used as a benchmarking dataset.

The relative evaluation metrics mentioned in Johnson et al. 2018 (Figures 2, 3, 4, 5, and 7) are one method for assessing the quality of the transcriptome. To create this subset, we used the absolute evaluation metrics, regardless of their improvement compared to the original NCGR assembly.

Methods

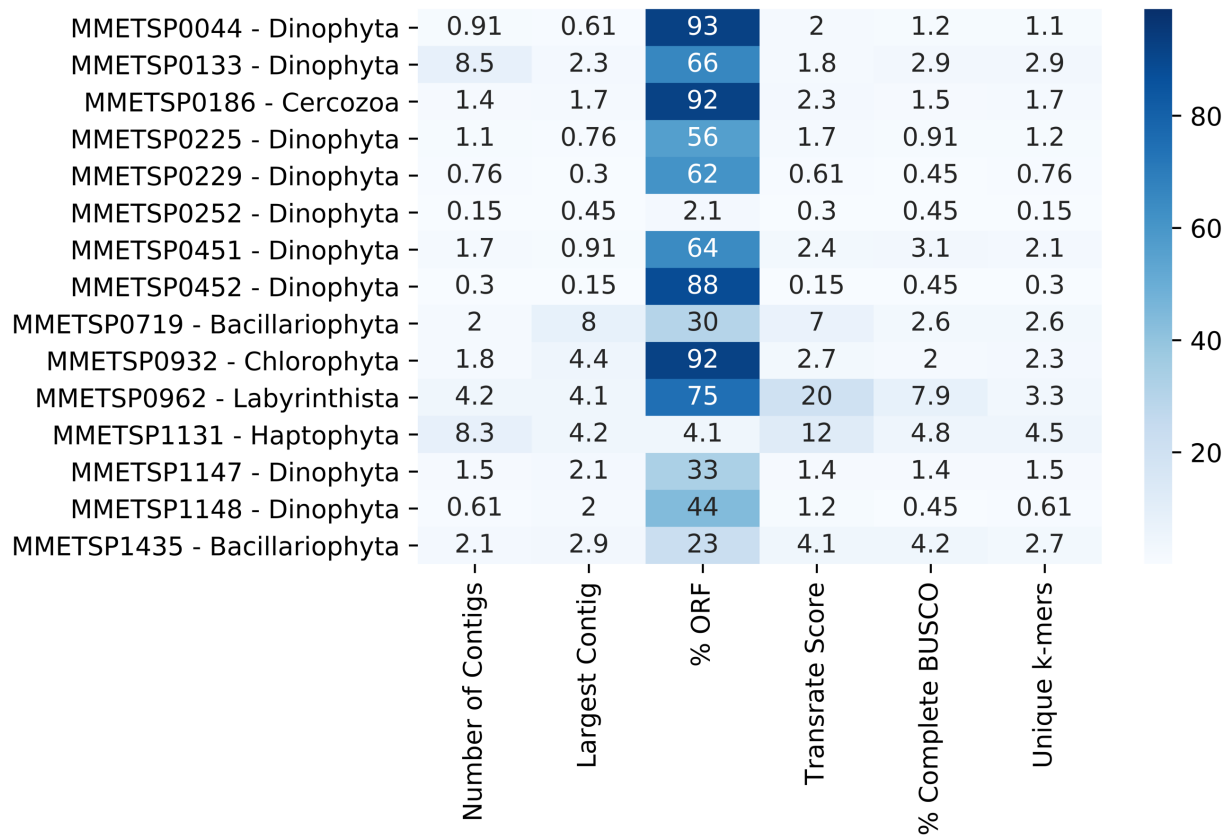
The evaluation metrics table (Supplemental Table 1) was used. MMETSP ID with more than one sequencing sample were excluded, leaving 660 samples with metrics. Python pandas (version 0.22.0) was used to calculate the rank of each sample for the following four metrics: number of contigs, largest contig (bases), % Complete BUSCO (eukaryotic), and unique k -mers ($k=25$). There were 12 samples ranking more than the 75% percentile across all four of these metrics, included in the “High Performing” subset (Figure 4a). There were 15 samples ranking less than the 25th percentile across all four of these metrics, included in the “Low Performing” subset (Figure 4b). The % Open Reading Frame (ORF) and Transrate score metrics were also included in Figures 4a and 4b to show that these metrics were not as consistent across the samples as the other four metrics.

A. High Performing MMETSP ID - Phylum



Supplemental Figure 4a. Subset of MMETSP ID. High performing MMETSP Samples: MMETSP0014, MMETSP0015, MMETSP0034, MMETSP0039, MMETSP0047, MMETSP0120, MMETSP0359, MMETSP0361, MMETSP0484, MMETSP0986, MMETSP0987, MMETSP0989.

B. Low Performing MMETSP ID - Phylum



Supplemental Figure 4b. Low performing MMETSP Samples: MMETSP0044, MMETSP0133, MMETSP0186, MMETSP0225, MMETSP0229, MMETSP0252, MMETSP0451, MMETSP0452, MMETSP0719, MMETSP0932, MMETSP0962, MMETSP1131, MMETSP1147, MMETSP1148, MMETSP1435.