

Supplemental Table ST3. 16S rRNA gene read count

Statistics. Raw reads were processed using a custom pipeline developed at the Joint Genome Institute. The number of total reads in the dataset remaining after each analysis step in this pipeline, as well as the percentage of the initial reads that these values represent, are displayed below.

Analysis Step	Read counts	Percentage
Input	57,380,917	100%
Filtered Contaminants	57,131,680	99.57%
Primer-Trimmed	56,898,358	99.16%
Extended	56,271,063	98.07%
Length-Filtered	56,155,628	97.86%
High Quality	45,977,229	80.13%
Clustered	26,087,863	45.46%
Classified	25,956,957	45.24%
Tax-Filter Pass	25,694,125	44.78%