Abstract

Sequence data manipulation and preparation for analysis requires multiple trimming and filtering steps to remove unwanted low quality scores, adapter sequences, amongst other miscellaneous information that disrupt data processing. After sequence data has been assembled and converted into a BAM file, a processing step to remove low mapping quality scores for the sequence reads aligned to a reference genome is required so that only high confidence reads are used for analysis. After running bam.py, a quality check is necessary to ensure that the outputted file contains an ideal number of reads after filtering. This script can fit into the larger bioinformatics workflow of converting raw sequencing data into a BAM or variant call format to view specific regions of interest.