

Resource Summary Report

Generated by [dkNET](#) on Sep 10, 2026

Steps to Create FASTQ of CCS Overlapping Genomic SSR - CCS ROI

DOI:10.17504/protocols.io.9i7h4hn

Type: Protocol

Proper Citation

Gregory Harhay 2019. Steps to Create FASTQ of CCS Overlapping Genomic SSR - CCS ROI . protocols.io dx.doi.org/10.17504/protocols.io.9i7h4hn

Protocol Information

URL: <https://dx.doi.org/10.17504/protocols.io.9i7h4hn>

Authors: Gregory Harhay

Summary: The virulence and pathogenicity of bacterial pathogens are related to their adaptability to changing environments. One process enabling adaptation is based on minor changes in genome sequence, as small as a few base pairs, within segments of genome called simple sequence repeats (SSRs) that consist of multiple copies of a short sequence (from one to several nucleotides), repeated in series. SSRs are found in eukaryotes as well as prokaryotes, and variation in them occurs at frequencies up to a million-fold higher than the average bacterial mutation rate through a process of slipped stranded mispairing (SSM) by DNA polymerase during replication. The characterization of SSR length by standard sequencing methods is complicated by the appearance of length variation introduced during the sequencing process that does not accurately quantify lower-abundance repeat number variants in a population. Here we report a computational approach to correct for process-induced artifacts, validated for tetranucleotide repeats by use of synthetic constructs of fixed, known length. We apply this method to a laboratory culture of *Histophilus somni*, prepared from a single colony, and demonstrate that the culture consists of populations of distinct sequence phase and read length variants at individual tetranucleotide SSR loci. **Input requirements:** Closed Genome - It is recommended that only organisms with closed genomes be the subject of the analyses described here. Mapping repetitive reads to contigs of non-closed genomes may map to multiple locations, complicating the analysis. Mapping CCS (circular consensus sequence) with repetitive sequence to closed genomes are guaranteed to map to a single locus if sufficient unique flanking sequence is used to confirm the unique mapping. Consequently, long CCS with high base quality are the most desirable input into this workflow.

Affiliations: United States Department of Agriculture

Version: 5

Publication Date: 2019

Proper Citation: Gregory Harhay 2019. Steps to Create FASTQ of CCS Overlapping Genomic SSR - CCS ROI . protocols.io [dx.doi.org/10.17504/protocols.io.9i7h4hn](https://doi.org/10.17504/protocols.io.9i7h4hn)

Record Creation Time: 20210329T220532+0000

Record Last Update: 20210329T220959+0000

Data and Source Information

Source: