Nov 17, 2019

BAM Alignment to CP018802 for Submission to GenBank V.2

Gregory Harhay[1]

[1]United States Department of Agriculture

1 Works for me dx.doi.org/10.17504/protocols.io.862hzge

Gregory P. Harhay
United States Department of Agriculture

ABSTRACT

Create BAM Alignment FIles including necessary tags that designate library info for submission into the Sequence Read Archive (SRA) at NCBI. Parking the BAM file at the SRA is convenent way to provide alignments to reproduce results at Steps to Create FASTQ of CCS Overlapping Genomic SSR - CCS ROI

Software Requirements

1

Geneious 11.1.5
by Biomatters Ltd

or another read mapper such as BowTie2

samtools 1.9
source by http://htslib.org/

Download genome from GenBank

2 https://www.ncbi.nlm.nih.gov/nuccore/CP018802.1

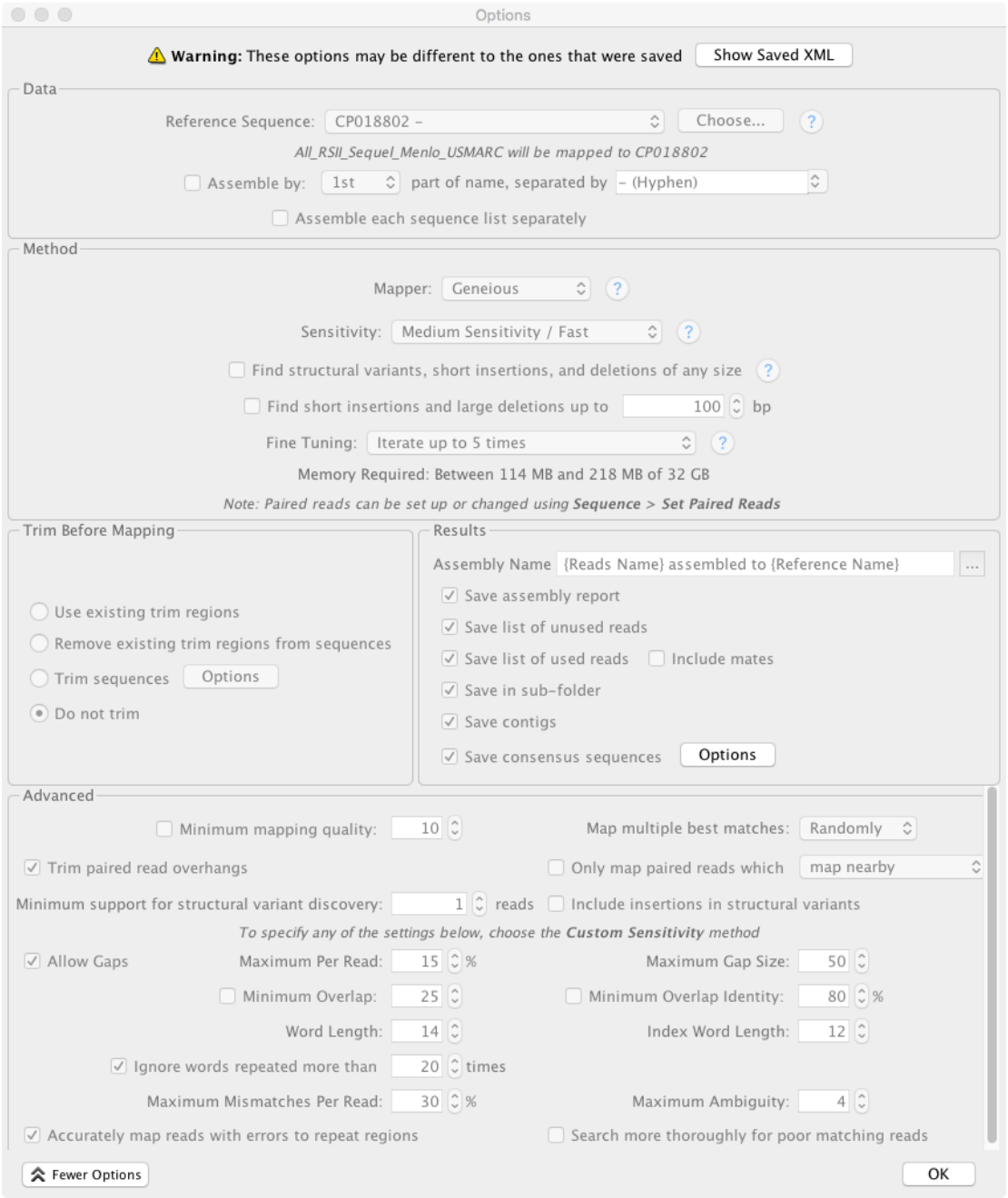
Align reads in each sequencing run to reference (CP018802)

3 Nine PacBio CCS libraries were mapped to CP018802

L.17494
L.17495
L.17836
L.18256_1

L.18256_2
L.18257
L.19770
L.19771
L.19981

Mapping Parameters: Geneious Assembler, Medium Sensitiivity/Fast, iterate up to 5 times, Map multiple best matched randomly for 9 libraries



- Export each alignment to SAM

Edit metadata in each SAM - NCBI requires library identifier metadata in read groups (@R)

4 ▪ Edit each SAM with SAM tools to append library as read group (RG) ID and append sample RG tag to each of the SAM alignments generated

samtools addreplacerg -r ID:LB:17494 -r SM:63250 -m overwrite_all -o L.17494_GeneMap_CP018802_RG.sam L.17494_GeneMap_CP018802.sam

Edit SAM read group (@RG) metadata, append library (LB) and sample (SM) metadata

▪ for each SAM, need to remove original @RG header line, for example

change

@HD VN:1.0 SO:unsorted
@SQ SN:CP018802 LN:2110642
@RG ID:Unpaired_reads_assembled_against_CP018802 SM:L.17495_GeneMap_CP018802
@RG ID:LB:17495 SM:63250

to

@HD VN:1.0 SO:unsorted
@SQ SN:CP018802 LN:2110642
@RG ID:LB:17495 SM:63250

and save file

Merge SAM files into a single BAM

5

samtools merge H_somni.L.101617.S.63250.bam L.19981_GeneMap_CP018802_RG.sam L.19771_GeneMap_CP018802_RG.sam L.19770_GeneMap_CP018802_RG.sam L.18257_GeneMap_CP018802_RG.sam L.18256_2_GeneMap_CP018802_RG.sam L.18256_1_GeneMap_CP018802_RG.sam L.17836_GeneMap_CP018802_RG.sam L.17495_GeneMap_CP018802_RG.sam L.17494_GeneMap_CP018802_RG.sam

Merge SAM files into a single BAM

Upload to SRA with appropriate metadata

6 BAM file available at https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?run=SRR8080935 , the download BAM file is named SRR8080935_CP018802.bam