
cvc Documentation

Release 1.0.0

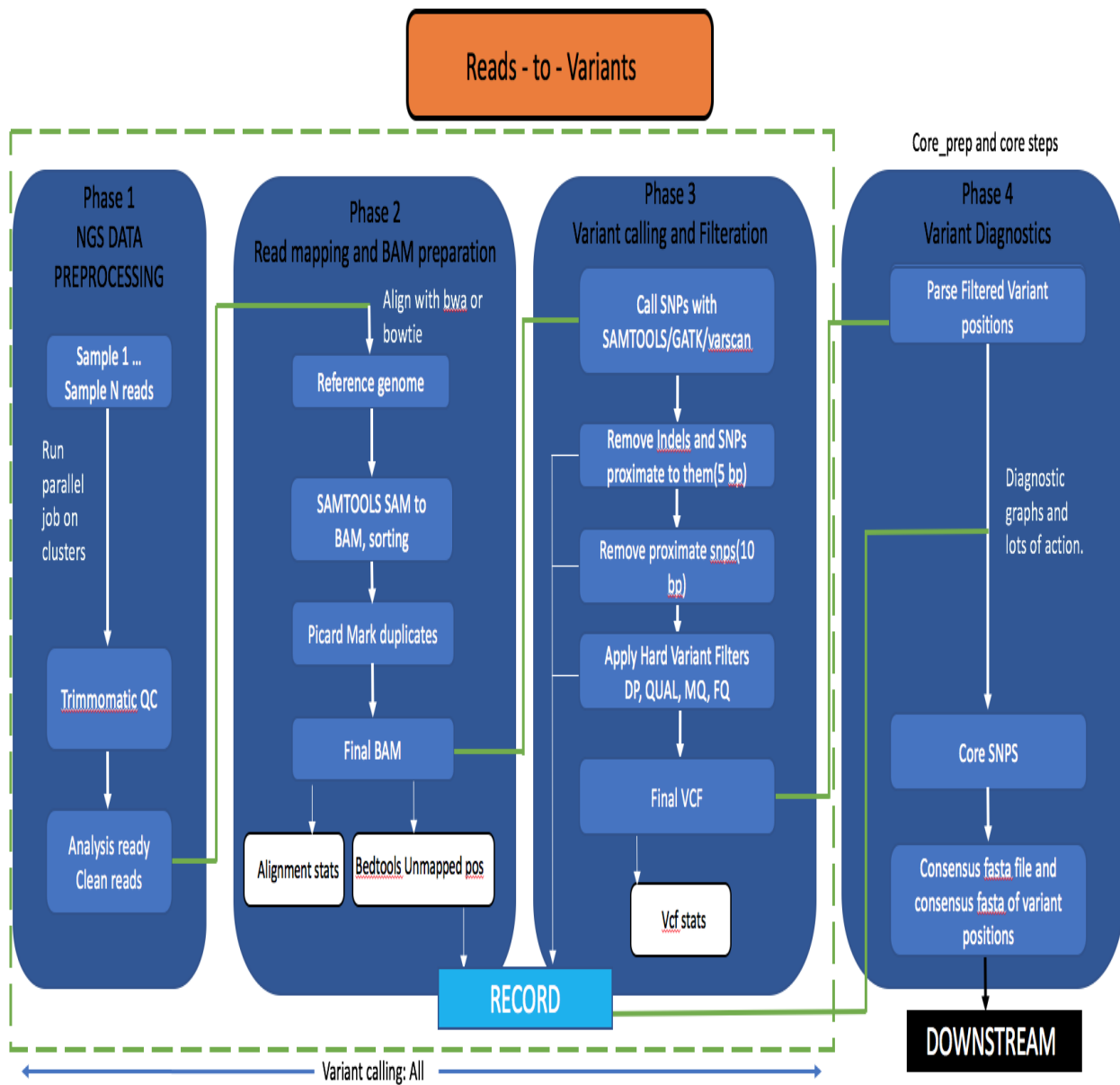
Ali Pirani

Feb 11, 2022

CONTENTS

1	Steps	3
----------	--------------	----------

CVC pipeline calls variants from Illumina PE/SE raw data and generates core SNP consensus fasta files that can be used as an input for downstream phylogenetic analysis.



There are three main steps to generate core SNPs which can be provided with -steps argument and should be run in

sequential order.

1. Variant Calling: This step will run all the standard variant calling steps on sample read files residing in input reads directory. The possible options are:

Option ***All*** : This will run all the variant calling steps starting from cleaning the reads to variant calling and generating intermediate files that will be used in later steps.

Option ***clean,align,post-align,varcall,filter,stats*** : This option will run variant calling steps starting from cleaning to variant calling, variant filtering and generating stats.

You can modify this argument to run a part of the pipeline. For example, if you want to restart the pipeline from alignment step, you can do it by supplying the following option: “align,post-align,varcall,filter,stats” which will skip the Trimmomatic cleaning part.

Note: The order of variant calling steps needs to be sequential to avoid any errors. While skipping any of the steps, make sure the results for skipped steps are already present in your output folder.

2. Preparing files for Core SNP extraction and diagnostics purposes:

Option ***core_prep*** : Run this step before running the last core steps. This will prepare all the intermediate data required for the last core step i.e generating the consensus fasta file out of core snps. This can be a time consuming step depending on how closely related the samples are to the reference genome.

3. Generate Core SNP consensus and data matrix for diagnostics plots:

Option ***core*** : This step will generate core SNP consensus fasta file and a consensus fasta of only core variant positions. Various data matrices will be generated at this step that can be used later for diagnostics purposes.

4. Generate report for the pipeline:

Option ***report*** : This step will generate final core results directory and various reports that will summarize the alignment and core SNP results.