/\*Commands used for Avocado variant caller program

\*

\* Installation: https://github.com/bigdatagenomics/avocado

\* Avocado-submit

\* Using SPARK\_SUBMIT=/usr/bin/spark-submit

Argument "READS" is required

READS : ADAM read-oriented data

REFERENCE : ADAM or FASTA reference genome data

VARIANTS : ADAM variant output

CONFIG : avocado configuration file

-debug : If set, prints a higher level of debug output.

-fragment\_length N : Sets maximum fragment length. Default value is 10,000. Values greater than 1e9 should be avoided.

-h (-help, --help, -?) : Print help

-parquet\_block\_size N : Parquet block size (default = 128mb)

-parquet\_compression\_codec [UNCOMPRESSED | SNAPPY | GZIP | LZO] : Parquet compression codec

-parquet\_disable\_dictionary : Disable dictionary encoding

-parquet\_logging\_level VAL : Parquet logging level (default = severe)

-parquet\_page\_size N : Parquet page size (default = 1mb)

-print\_metrics : Print metrics to the log on completion

\* avocado-submit <input files> <reference file> <output directory> <config file(local file system)>

\*

\* there are 2 sample config files that comes with avocado package:

\* avocado-sample-configs/basic.properties

\* avocado-sample-configs/snap-basic.properties # for unaligned reads

\*<https://groups.google.com/forum/#!msg/adam-developers/QoVJLJwHqxs/nK7k1evsBwAJ>

\*

\* While input adam files, reference file, output directory need to be on HDFS, the config files can be accessed locally, do not use HDFS directory

\*/

//===========Map assembled ADAM with Reference Genome===========

avocado-submit mrsa/adam/adamq.ERR064898 mrsa/adam/refseq/ASM58453v1\_genomic.fa mrsa/avocado/variant-map.ERR0464898 /home/biotools/avocado/avocado-sample-configs/basic.properties

/\* ERRORS

ERROR executor.Executor: Exception in task 2.0 in stage 7.0 (TID 8)org.apache.parquet.io.InvalidRecordException: Parquet/Avro schema mismatch. Avro field 'readNum' not found.

\*/