

# Variant calling

```
bwa index GCF_000146045.2_R64_genomic.fna
samtools faidx GCF_000146045.2_R64_genomic.fna
gatk CreateSequenceDictionary -R GCF_000146045.2_R64_genomic.fna
```

```
samtools faidx GCF_000146045.2_R64_genomic.fna
gatk CreateSequenceDictionary -R GCF_000146045.2_R64_genomic.fna
[bwa_index] Pack FASTA... 0.08 sec
[bwa_index] Construct BWT for the packed sequence...
[bwa_index] 5.53 seconds elapse.
[bwa_index] Update BWT... 0.06 sec
[bwa_index] Pack forward-only FASTA... 0.05 sec
[bwa_index] Construct SA from BWT and Occ... 2.72 sec
[main] Version: 0.7.19-r1273
[main] CMD: bwa index GCF_000146045.2_R64_genomic.fna
[main] Real time: 9.826 sec; CPU: 8.440 sec
Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
  java -Dsamjdk.use_async_io_read_samtools=false -Dsamjdk.use_async_io_write_samtools=true -Dsamjdk.use_async_io_write_tribble=false -Dsamjdk.compression_level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar CreateSequenceDictionary -R GCF_000146045.2_R64_genomic.fna
INFO 2025-10-16 22:36:19 CreateSequenceDictionary Output dictionary will be written in /home/ibab/SEM3/RL/101625/varcal/GCF_000146045.2_R64_genomic.dict
22:36:19.467 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
[Thu Oct 16 22:36:19 IST 2025] CreateSequenceDictionary --REFERENCE GCF_000146045.2_R64_genomic.fna --TRUNCATE_NAMES_AT_WHITESPACE true --NUM_SEQUENCES 2147483647 --VERBOSITY INFO --QUIET false --VALIDATION_STRINGENCY STRICT --COMPRESSION_LEVEL 2 --MAX_RECORDS_IN_RAM 500000 --CREATE_INDEX false --CREATE_MD5_FILE false --help false --version false --showHidden false --USE_JDK_DEFLATER false --USE_JDK_INFLATER false
[Thu Oct 16 22:36:19 IST 2025] Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux 6.14.0-29-generic amd64; OpenJDK 64-Bit Server VM 17.0.16+8-Ubuntu-0ubuntu124.04.1; Deflater: Intel; Inflater: Intel; Provider GCS is available; Picard version: Version:4.6.1.0
[Thu Oct 16 22:36:19 IST 2025] picard.sam.CreateSequenceDictionary done. Elapsed time: 0.01 minutes.
Runtime.totalMemory()=119537664
Tool returned:
0
```

```
bwa mem GCF_000146045.2_R64_genomic.fna conA_rep1.fq > conA_rep1.sam
bwa mem GCF_000146045.2_R64_genomic.fna conA_rep2.fq > conA_rep2.sam
```

```
bwa mem GCF_000146045.2_R64_genomic.fna conA_rep2.fq > conA_rep2.sam
[M::bwa_idx_load_from_disk] read 0 ALT contigs
[M::process] read 99010 sequences (10000010 bp)...
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.133 CPU sec, 5.054 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.111 CPU sec, 4.990 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.345 CPU sec, 5.233 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 6.117 CPU sec, 6.097 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.801 CPU sec, 5.684 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.491 CPU sec, 5.410 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.418 CPU sec, 5.309 real sec
[M::process] read 7920 sequences (799920 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.173 CPU sec, 5.105 real sec
[M::mem_process_seqs] Processed 7920 reads in 0.493 CPU sec, 0.420 real sec
[main] Version: 0.7.19-r1273
[main] CMD: bwa mem GCF_000146045.2_R64_genomic.fna conA_rep1.fq
[main] Real time: 44.162 sec; CPU: 44.107 sec
[M::bwa_idx_load_from_disk] read 0 ALT contigs
[M::process] read 99010 sequences (10000010 bp)...
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.321 CPU sec, 5.254 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.948 CPU sec, 5.957 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.366 CPU sec, 5.245 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.852 CPU sec, 5.779 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.355 CPU sec, 5.234 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.447 CPU sec, 5.328 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.555 CPU sec, 5.436 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.454 CPU sec, 5.326 real sec
[M::process] read 8910 sequences (899910 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.597 CPU sec, 5.528 real sec
[M::mem_process_seqs] Processed 8910 reads in 0.561 CPU sec, 0.489 real sec
[main] Version: 0.7.19-r1273
[main] CMD: bwa mem GCF_000146045.2_R64_genomic.fna conA_rep2.fq
[main] Real time: 50.924 sec; CPU: 50.576 sec
```

```
bwa mem GCF_000146045.2_R64_genomic.fna conB_rep1.fq > conB_rep1.sam
bwa mem GCF_000146045.2_R64_genomic.fna conB_rep2.fq > conB_rep2.sam
```

```
(base) ibab@IBAB-MSc14-Comp011:~/SEM3/RL/101625/varcal$ bwa mem GCF_000146045.2_R64_genomic.fna conB_rep1.fq > conB_rep1.sam
bwa mem GCF_000146045.2_R64_genomic.fna conB_rep2.fq > conB_rep2.sam
[M::bwa_idx_load_from_disk] read 0 ALT contigs
[M::process] read 99010 sequences (10000010 bp)...
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.248 CPU sec, 5.166 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.148 CPU sec, 5.033 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.541 CPU sec, 5.427 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.274 CPU sec, 5.155 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.596 CPU sec, 5.599 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.429 CPU sec, 5.302 real sec
[M::process] read 6930 sequences (699930 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.538 CPU sec, 5.461 real sec
[M::mem_process_seqs] Processed 6930 reads in 0.487 CPU sec, 0.413 real sec
[main] Version: 0.7.19-r1273
[main] CMD: bwa mem GCF_000146045.2_R64_genomic.fna conB_rep1.fq
[main] Real time: 39.236 sec; CPU: 38.378 sec
[M::bwa_idx_load_from_disk] read 0 ALT contigs
[M::process] read 99010 sequences (10000010 bp)...
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.578 CPU sec, 5.518 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.608 CPU sec, 5.482 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.550 CPU sec, 5.421 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.459 CPU sec, 5.338 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.639 CPU sec, 5.549 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.513 CPU sec, 5.389 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.772 CPU sec, 5.653 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.556 CPU sec, 5.437 real sec
[M::process] read 99010 sequences (10000010 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.583 CPU sec, 5.466 real sec
[M::process] read 9900 sequences (999900 bp)...
[M::mem_process_seqs] Processed 99010 reads in 5.271 CPU sec, 5.212 real sec
[M::mem_process_seqs] Processed 9900 reads in 0.635 CPU sec, 0.563 real sec
```

```
samtools view -bS conA_rep1.sam | samtools sort -o conA_rep1_sorted.bam
samtools view -bS conA_rep2.sam | samtools sort -o conA_rep2_sorted.bam
samtools view -bS conB_rep1.sam | samtools sort -o conB_rep1_sorted.bam
samtools view -bS conB_rep2.sam | samtools sort -o conB_rep2_sorted.bam
samtools merge conA_merged.bam conA_rep1_sorted.bam conA_rep2_sorted.bam
samtools merge conB_merged.bam conB_rep1_sorted.bam conB_rep2_sorted.bam
```

```
java -jar picard.jar MarkDuplicates \
  I=conA_merged.bam \
  O=conA_merged_markdup.bam \
  M=conA_markdup_metrics.txt
```

```
samtools index conA_merged_markdup.bam
```

```
java -jar picard.jar MarkDuplicates \
  I=conB_merged.bam \
  O=conB_merged_markdup.bam \
  M=conB_markdup_metrics.txt
```

```
samtools index conB_merged_markdup.bam
```

```
(base) ibab@IBAB-MSc14-Comp011:~/SEM3/RL/101625/varcal$ java -jar picard.jar MarkDuplicates \
I=conA_merged.bam \
O=conA_merged_markdup.bam \
M=conA_markdup_metrics.txt
INFO 2025-10-16 22:50:53 MarkDuplicates

***** NOTE: Picard's command line syntax is changing.
*****
***** For more information, please see:
*****
https://github.com/broadinstitute/picard/wiki/Command-Line-Syntax-Transition-For-Users-(Pre-Transition)
*****
***** The command line looks like this in the new syntax:
*****
***** MarkDuplicates -I conA_merged.bam -O conA_merged_markdup.bam -M conA_markdup_metrics.txt
*****

22:50:53.754 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/101625/varcal/picard.jar!/com/intel/gkl/native/libgkl
_compression.so
[Thu Oct 16 22:50:53 IST 2025] MarkDuplicates INPUT=[conA_merged.bam] OUTPUT=conA_merged_markdup.bam METRICS_FILE=conA_markdup_metrics.txt MAX_SEQUENCES_F
OR_DISK_READ_ENDS_MAP=50000 MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=8000 SORTING_COLLECTION_SIZE_RATIO=0.25 TAG_DUPLICATE_SET_MEMBERS=false REMOVE_SEQUENCING_DUPL
ICATES=false TAGGING_POLICY=DontTag CLEAR_DT=true DUPLEX_UMI=false FLOW_MODE=false FLOW_DUP_STRATEGY=FLOW_QUALITY_SUM_STRATEGY FLOW_USE_END_IN_UNPAIRED_READS
=false FLOW_USE_UNPAIRED_CLIPPED_END=false FLOW_UNPAIRED_END_UNCERTAINTY=0 FLOW_UNPAIRED_START_UNCERTAINTY=0 FLOW_SKIP_FIRST_N_FLOWS=0 FLOW_Q_IS_KNOWN_END=false
FLOW_EFFECTIVE_QUALITY_THRESHOLD=15 ADD_PG_TAG_TO_READS=true REMOVE_DUPLICATES=false ASSUME_SORTED=false DUPLICATE_SCORING_STRATEGY=SUM_OF_BASE_QUALITIES
PROGRAM_RECORD_ID=MarkDuplicates PROGRAM_GROUP_NAME=MarkDuplicates READ_NAME_REGEX=<optimized capture of last three ':' separated fields as numeric values>
OPTICAL_DUPLICATE_PIXEL_DISTANCE=100 MAX_OPTICAL_DUPLICATE_SET_SIZE=300000 VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RE
CORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false USE_JDK_DEFLATER=false USE_JDK_INFLATER=false
[Thu Oct 16 22:50:54 IST 2025] Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux 6.14.0-29-generic amd64; OpenJDK 64-Bit Server VM 17.0.16+8-Ubuntu-0u
buntu124.04.1; Deflater: Intel; Inflater: Intel; Provider GCS is available; Picard version: 3.4.0
INFO 2025-10-16 22:50:54 MarkDuplicates Start of doWork freeMemory: 44774128; totalMemory: 56623104; maxMemory: 4181721088
INFO 2025-10-16 22:50:54 MarkDuplicates Reading input file and constructing read end information.
INFO 2025-10-16 22:50:54 MarkDuplicates Will retain up to 15151163 data points before spilling to disk.
INFO 2025-10-16 22:50:56 MarkDuplicates Read 1,000,000 records. Elapsed time: 00:00:02s. Time for last 1,000,000: 2s. Last read position: N
C_001143.9:370,781
INFO 2025-10-16 22:50:56 MarkDuplicates Tracking 0 as yet unmatched pairs. 0 records in RAM.
INFO 2025-10-16 22:50:57 MarkDuplicates Read 1669600 records. 0 pairs never matched.
INFO 2025-10-16 22:50:57 MarkDuplicates After buildSortedReadEndLists freeMemory: 613427408; totalMemory: 868220928; maxMemory: 4181721088
INFO 2025-10-16 22:50:57 MarkDuplicates Will retain up to 130678784 duplicate indices before spilling to disk.
INFO 2025-10-16 22:50:58 MarkDuplicates Traversing read pair information and detecting duplicates.
INFO 2025-10-16 22:50:58 MarkDuplicates Traversing fragment information and detecting duplicates.
INFO 2025-10-16 22:50:58 MarkDuplicates Sorting list of duplicate records.
INFO 2025-10-16 22:50:58 MarkDuplicates After generateDuplicateIndexes freeMemory: 1042025576; totalMemory: 2101346304; maxMemory: 4181721088
INFO 2025-10-16 22:50:58 MarkDuplicates Marking 788150 records as duplicates.
INFO 2025-10-16 22:50:58 MarkDuplicates Found 0 optical duplicate clusters.
INFO 2025-10-16 22:50:58 MarkDuplicates Reads are assumed to be ordered by: coordinate
```

```
java -jar picard.jar AddOrReplaceReadGroups \
I=conA_merged_markdup.bam \
O=conA_merged_markdup_rg.bam \
RGID=1 RGLB=lib1 RGPL=illumina RGPU=unit1 RGSM=conA
samtools index conA_merged_markdup_rg.bam
java -jar picard.jar AddOrReplaceReadGroups \
I=conB_merged_markdup.bam \
O=conB_merged_markdup_rg.bam \
RGID=2 RGLB=lib1 RGPL=illumina RGPU=unit1 RGSM=conB
samtools index conB_merged_markdup_rg.bam
```

```
at org.broadinstitute.hellbender.Main.main(Main.java:308)
(base) ibab@IBAB-MSc14-Comp011:~/SEM3/RL/101625/varcal$ java -jar picard.jar AddOrReplaceReadGroups \
I=conA_merged_markdup.bam \
O=conA_merged_markdup_rg.bam \
RGID=1 RGLB=lib1 RGPL=illumina RGPU=unit1 RGSM=conA
samtools index conA_merged_markdup_rg.bam

java -jar picard.jar AddOrReplaceReadGroups \
I=conB_merged_markdup.bam \
O=conB_merged_markdup_rg.bam \
RGID=2 RGLB=lib1 RGPL=illumina RGPU=unit1 RGSM=conB
samtools index conB_merged_markdup_rg.bam
INFO 2025-10-16 22:55:32 AddOrReplaceReadGroups

***** NOTE: Picard's command line syntax is changing.
*****
***** For more information, please see:
*****
https://github.com/broadinstitute/picard/wiki/Command-Line-Syntax-Transition-For-Users-(Pre-Transition)
*****
***** The command line looks like this in the new syntax:
*****
***** AddOrReplaceReadGroups -I conA_merged_markdup.bam -O conA_merged_markdup_rg.bam -RGID 1 -RGLB lib1 -RGPL illumina -RGPU unit1 -RGSM conA
*****

22:55:32.779 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/101625/varcal/picard.jar!/com/intel/gkl/native/libgkl
_compression.so
[Thu Oct 16 22:55:32 IST 2025] AddOrReplaceReadGroups INPUT=conA_merged_markdup.bam OUTPUT=conA_merged_markdup_rg.bam RGID=1 RGLB=lib1 RGPL=illumina RGPU=uni
t1 RGSM=conA VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=fals
e USE_JDK_DEFLATER=false USE_JDK_INFLATER=false
[Thu Oct 16 22:55:32 IST 2025] Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux 6.14.0-29-generic amd64; OpenJDK 64-Bit Server VM 17.0.16+8-Ubuntu-0u
buntu124.04.1; Deflater: Intel; Inflater: Intel; Provider GCS is available; Picard version: 3.4.0
INFO 2025-10-16 22:55:32 AddOrReplaceReadGroups Created read-group ID=1 PL=illumina LB=lib1 SM=conA

INFO 2025-10-16 22:55:39 AddOrReplaceReadGroups Processed 1,000,000 records. Elapsed time: 00:00:06s. Time for last 1,000,000: 6s. Last rea
d position: NC_001143.9:370,781
[Thu Oct 16 22:55:44 IST 2025] picard.sam.AddOrReplaceReadGroups done. Elapsed time: 0.19 minutes.
Runtime.totalMemory()=266338304
INFO 2025-10-16 22:55:45 AddOrReplaceReadGroups

***** NOTE: Picard's command line syntax is changing.
*****
***** For more information, please see:
*****
https://github.com/broadinstitute/picard/wiki/Command-Line-Syntax-Transition-For-Users-(Pre-Transition)
*****
***** The command line looks like this in the new syntax:
*****
```

## gatk HaplotypeCaller \

```
-R GCF_000146045.2_R64_genomic.fna \
-I conA_merged_markdup_rg.bam \
-O conA_variants.vcf
```

## gatk HaplotypeCaller \

```
-R GCF_000146045.2_R64_genomic.fna \
-I conB_merged_markdup_rg.bam \
-O conB_variants.vcf
```

```
(base) ibab@IBAB-MS14-Comp011:~/SEM3/RL/101625/varcal$ gatk HaplotypeCaller \
-R GCF_000146045.2_R64_genomic.fna \
-I conA_merged_markdup_rg.bam \
-O conA_variants.vcf

gatk HaplotypeCaller \
-R GCF_000146045.2_R64_genomic.fna \
-I conB_merged_markdup_rg.bam \
-O conB_variants.vcf

Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
  java -Dsamjdk.use_async_io_read_samtools=false -Dsamjdk.use_async_io_write_samtools=true -Dsamjdk.use_async_io_write_tribble=false -Dsamjdk.compression_level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar HaplotypeCaller -R GCF_000146045.2_R64_genomic.fna -I conA_merged_markdup_rg.bam -O conA_variants.vcf
22:56:53.790 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@ad3f70a]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@54ccb3]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@ad3f70a]
22:56:53.909 INFO HaplotypeCaller - -----
22:56:53.915 INFO HaplotypeCaller - The Genome Analysis Toolkit (GATK) v4.6.1.0
22:56:53.915 INFO HaplotypeCaller - For support and documentation go to https://software.broadinstitute.org/gatk/
22:56:53.915 INFO HaplotypeCaller - Executing as ibab@IBAB-MS14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic amd64
22:56:53.915 INFO HaplotypeCaller - Java runtime: OpenJDK 64-Bit Server VM v17.0.16+8-Ubuntu-0ubuntu124.04.1
22:56:53.915 INFO HaplotypeCaller - Start Date/Time: October 16, 2025 at 10:56:53 PM IST
22:56:53.916 INFO HaplotypeCaller - -----
22:56:53.916 INFO HaplotypeCaller - -----
22:56:53.916 INFO HaplotypeCaller - HTSJDK Version: 4.1.3
22:56:53.917 INFO HaplotypeCaller - Picard Version: 3.3.0
22:56:53.917 INFO HaplotypeCaller - Built for Spark Version: 3.5.0
22:56:53.919 INFO HaplotypeCaller - HTSJDK Defaults.COMPRESSION_LEVEL : 2
22:56:53.920 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
22:56:53.920 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
22:56:53.920 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
22:56:53.920 INFO HaplotypeCaller - Deflater: IntelDeflater
22:56:53.920 INFO HaplotypeCaller - Inflater: IntelInflater
22:56:53.920 INFO HaplotypeCaller - GCS max retries/reopens: 20
22:56:53.921 INFO HaplotypeCaller - Requester pays: disabled
22:56:53.921 INFO HaplotypeCaller - Initializing engine
22:56:54.097 INFO HaplotypeCaller - Done initializing engine
22:56:54.109 INFO NativeLibraryLoader - Loading libgkl_utils.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_utils.so
22:56:54.110 INFO NativeLibraryLoader - Loading libgkl_smithwaterman.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_smithwaterman.so
22:56:54.111 INFO SmithWatermanAligner - Using AVX accelerated SmithWaterman implementation
22:56:54.114 INFO HaplotypeCallerEngine - Disabling physical phasing, which is supported only for reference-model confidence output
22:56:54.124 INFO NativeLibraryLoader - Loading libgkl_pairhmm_omp.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_pairhmm_omp.so
22:56:54.142 INFO IntelPairHmm - Flush-to-zero (FTZ) is enabled when running PairHMM
22:56:54.142 INFO IntelPairHmm - Available threads: 4
22:56:54.143 INFO IntelPairHmm - Requested threads: 4
22:56:54.144 INFO PairHMM - Using the OpenMP multi-threaded AVX-accelerated native PairHMM implementation
22:56:54.186 INFO ProgressMeter - Starting traversal
22:56:54.186 INFO ProgressMeter - Current Locus Elapsed Minutes Regions Processed Regions/Minute
22:56:54.519 WARN InbreedingCoeff - InbreedingCoeff will not be calculated at position NC_001133.9:376 and possibly subsequent; at least 10 samples must have called genotypes
22:57:04.191 INFO ProgressMeter - NC_001134.8:252784 0.2 2750 16493.4
22:57:14.194 INFO ProgressMeter - NC_001136.10:152629 0.3 8370 25100.0
22:59:09.596 INFO HaplotypeCaller - -----
22:59:09.596 INFO HaplotypeCaller - The Genome Analysis Toolkit (GATK) v4.6.1.0
22:59:09.596 INFO HaplotypeCaller - For support and documentation go to https://software.broadinstitute.org/gatk/
22:59:09.596 INFO HaplotypeCaller - Executing as ibab@IBAB-MS14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic amd64
22:59:09.597 INFO HaplotypeCaller - Java runtime: OpenJDK 64-Bit Server VM v17.0.16+8-Ubuntu-0ubuntu124.04.1
22:59:09.597 INFO HaplotypeCaller - Start Date/Time: October 16, 2025 at 10:59:09 PM IST
22:59:09.597 INFO HaplotypeCaller - -----
22:59:09.597 INFO HaplotypeCaller - -----
22:59:09.598 INFO HaplotypeCaller - HTSJDK Version: 4.1.3
22:59:09.598 INFO HaplotypeCaller - Picard Version: 3.3.0
22:59:09.598 INFO HaplotypeCaller - Built for Spark Version: 3.5.0
22:59:09.598 INFO HaplotypeCaller - HTSJDK Defaults.COMPRESSION_LEVEL : 2
22:59:09.598 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
22:59:09.598 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
22:59:09.598 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
22:59:09.598 INFO HaplotypeCaller - Deflater: IntelDeflater
22:59:09.598 INFO HaplotypeCaller - Inflater: IntelInflater
22:59:09.598 INFO HaplotypeCaller - GCS max retries/reopens: 20
22:59:09.598 INFO HaplotypeCaller - Requester pays: disabled
22:59:09.598 INFO HaplotypeCaller - Initializing engine
22:59:09.598 INFO HaplotypeCaller - Done initializing engine
22:59:09.675 INFO NativeLibraryLoader - Loading libgkl_utils.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_utils.so
22:59:09.676 INFO NativeLibraryLoader - Loading libgkl_smithwaterman.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_smithwaterman.so
22:59:09.678 INFO SmithWatermanAligner - Using AVX accelerated SmithWaterman implementation
22:59:09.681 INFO HaplotypeCallerEngine - Disabling physical phasing, which is supported only for reference-model confidence output
22:59:09.692 INFO NativeLibraryLoader - Loading libgkl_pairhmm_omp.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_pairhmm_omp.so
22:59:09.707 INFO IntelPairHmm - Flush-to-zero (FTZ) is enabled when running PairHMM
22:59:09.708 INFO IntelPairHmm - Available threads: 4
22:59:09.709 INFO IntelPairHmm - Requested threads: 4
22:59:09.709 INFO PairHMM - Using the OpenMP multi-threaded AVX-accelerated native PairHMM implementation
22:59:09.755 INFO ProgressMeter - Starting traversal
22:59:09.756 INFO ProgressMeter - Current Locus Elapsed Minutes Regions Processed Regions/Minute
22:59:10.187 WARN InbreedingCoeff - InbreedingCoeff will not be calculated at position NC_001133.9:27210 and possibly subsequent; at least 10 samples must have called genotypes
22:59:19.767 INFO ProgressMeter - NC_001134.8:259895 0.2 2810 16841.5
22:59:29.770 INFO ProgressMeter - NC_001136.10:201844 0.3 8720 26143.0
22:59:39.776 INFO ProgressMeter - NC_001136.10:938592 0.5 13110 26202.5
22:59:49.801 INFO ProgressMeter - NC_001137.3:348660 0.7 18560 27808.7
22:59:59.814 INFO ProgressMeter - NC_001139.9:232705 0.8 23110 27699.9
23:00:09.818 INFO ProgressMeter - NC_001140.6:410107 1.0 29940 29900.6
```

```
gatk BaseRecalibrator \
-R GCF_000146045.2_R64_genomic.fna \
-I conA_merged_markdup_rg.bam \
--known-sites conA_variants.vcf \
-O conA_BQSR.table
```

```
gatk ApplyBQSR \
-R GCF_000146045.2_R64_genomic.fna \
-I conA_merged_markdup_rg.bam \
--bqsr-recal-file conA_BQSR.table \
-O conA_merged_markdup_rg_bqsr.bam
```

```
(base) ibab@IBAB-MSc14-Comp011:~/SEM3/RL/101625/varcal$ gatk BaseRecalibrator \
-R GCF_000146045.2_R64_genomic.fna \
-I conA_merged_markdup_rg.bam \
--known-sites conA_variants.vcf \
-O conA_BQSR.table

gatk ApplyBQSR \
-R GCF_000146045.2_R64_genomic.fna \
-I conA_merged_markdup_rg.bam \
--bqsr-recal-file conA_BQSR.table \
-O conA_merged_markdup_rg_bqsr.bam

Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
java -Dsamjdk.use_async_io_read_samtools=false -Dsamjdk.use_async_io_write_samtools=true -Dsamjdk.use_async_io_write_tribble=false -Dsamjdk.compression_level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar BaseRecalibrator -R GCF_000146045.2_R64_genomic.fna -I conA_merged_markdup_rg.bam --known-sites conA_variants.vcf -O conA_BQSR.table
23:04:35.428 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@6812c8cc]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@3457cc8d]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@6812c8cc]
23:04:35.552 INFO BaseRecalibrator - -----
23:04:35.556 INFO BaseRecalibrator - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:04:35.556 INFO BaseRecalibrator - For support and documentation go to https://software.broadinstitute.org/gatk/
23:04:35.556 INFO BaseRecalibrator - Executing as ibab@IBAB-MSc14-Comp011:ibab.ac.in on Linux v6.14.0-29-generic amd64
23:04:35.556 INFO BaseRecalibrator - Java runtime: OpenJDK 64-Bit Server VM v17.0.16+8-Ubuntu-0ubuntu124.04.1
23:04:35.557 INFO BaseRecalibrator - Start Date/Time: October 16, 2025 at 11:04:35 PM IST
23:04:35.557 INFO BaseRecalibrator - -----
23:04:35.557 INFO BaseRecalibrator -
23:04:35.558 INFO BaseRecalibrator - HTSJDK Version: 4.1.3
23:04:35.558 INFO BaseRecalibrator - Picard Version: 3.1.0
23:04:35.558 INFO BaseRecalibrator - Built for Spark Version: 3.5.0
23:04:35.560 INFO BaseRecalibrator - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:04:35.561 INFO BaseRecalibrator - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:04:35.561 INFO BaseRecalibrator - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:04:35.561 INFO BaseRecalibrator - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:04:35.561 INFO BaseRecalibrator - Deflater: IntelDeflater
23:04:35.561 INFO BaseRecalibrator - Inflater: IntelInflater
23:04:35.562 INFO BaseRecalibrator - GCS max retries/reopens: 20
23:04:35.562 INFO BaseRecalibrator - Requester pays: disabled
23:04:35.562 INFO BaseRecalibrator - Initializing engine
23:04:35.710 INFO FeatureManager - Using codec VCFCodec to read file file:///home/ibab/SEM3/RL/101625/varcal/conA_variants.vcf
23:04:35.736 INFO BaseRecalibrator - Done initializing engine
23:04:35.741 INFO BaseRecalibrationEngine - The covariates being used here:
23:04:35.741 INFO BaseRecalibrationEngine - ReadGroupCovariate
23:04:35.742 INFO BaseRecalibrationEngine - QualityScoreCovariate
23:04:35.742 INFO BaseRecalibrationEngine - ContextCovariate
23:04:35.742 INFO BaseRecalibrationEngine - CycleCovariate
23:04:35.744 INFO ProgressMeter - Starting traversal
23:04:35.745 INFO ProgressMeter - Current Locus Elapsed Minutes Reads Processed Reads/Minute
23:04:45.760 INFO ProgressMeter - NC_001145.3:389883 0.2 607000 3637271.5
```

```
gatk BaseRecalibrator \
-R GCF_000146045.2_R64_genomic.fna \
-I conB_merged_markdup_rg.bam \
--known-sites conB_variants.vcf \
-O conB_BQSR.table
```

```
gatk ApplyBQSR \
-R GCF_000146045.2_R64_genomic.fna \
-I conB_merged_markdup_rg.bam \
--bqsr-recal-file conB_BQSR.table \
-O conB_merged_markdup_rg_bqsr.bam
```



```
(base) tbab@IBAB-MScl4-Comp011:~/SEM3/RL/101025/varcal$ gatk BaseRecalibrator \
-R GCF_000146045.2_R64_genomic.fna \
-I conB_merged_markdup_rg.bam \
--known-sites conB_variants.vcf \
-O conB_BQSR.table

gatk ApplyBQSR \
-R GCF_000146045.2_R64_genomic.fna \
-I conB_merged_markdup_rg.bam \
--bqsr-recal-file conB_BQSR.table \
-O conB_merged_markdup_rg_bqsr.bam
Using GATK jar /home/tbab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
  java -Dsanjdk.use_async_io.read.santools=false -Dsanjdk.use_async_io.write.santools=true -Dsanjdk.use_async_io.write.tribble=false -Dsanjdk.compression.level=2 -jar /home/tbab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar BaseRecalibrator -R GCF_000146045.2_R64_genomic.fna -I conB_merged_markdup_rg.bam --known-sites conB_variants.vcf -O conB_BQSR.table
23:06:52.627 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/tbab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@6812c8cc]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@3457cc8d]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@6812c8cc]
23:06:52.746 INFO BaseRecalibrator - .....
23:06:52.749 INFO BaseRecalibrator - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:06:52.749 INFO BaseRecalibrator - For support and documentation go to https://software.broadinstitute.org/gatk/
23:06:52.749 INFO BaseRecalibrator - Executing as tbab@IBAB-MScl4-Comp011.tbab.ac.in on Linux v6.14.0-29-generic amd64
23:06:52.749 INFO BaseRecalibrator - Java runtime: OpenJDK 64-Bit Server VM v17.0.16+8-Ubuntu-0ubuntu124.04.1
23:06:52.749 INFO BaseRecalibrator - Start Date/Time: October 16, 2025 at 11:06:52 PM IST
23:06:52.749 INFO BaseRecalibrator - .....
23:06:52.749 INFO BaseRecalibrator - HTSJDK Version: 4.1.3
23:06:52.750 INFO BaseRecalibrator - Picard Version: 3.3.0
23:06:52.750 INFO BaseRecalibrator - Built for Spark Version: 3.5.0
23:06:52.754 INFO BaseRecalibrator - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:06:52.755 INFO BaseRecalibrator - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SANTOOLS : false
23:06:52.755 INFO BaseRecalibrator - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SANTOOLS : true
23:06:52.755 INFO BaseRecalibrator - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:06:52.755 INFO BaseRecalibrator - Deflater: IntelDeflater
23:06:52.755 INFO BaseRecalibrator - Inflater: IntelInflater
23:06:52.755 INFO BaseRecalibrator - GCS max retries/reopens: 20
23:06:52.756 INFO BaseRecalibrator - Requester pays: disabled
23:06:52.756 INFO BaseRecalibrator - Initializing engine
23:06:52.912 INFO FeatureManager - Using codec VCFCodec to read file file:///home/tbab/SEM3/RL/101025/varcal/conB_variants.vcf
23:06:52.937 INFO BaseRecalibrator - Done initializing engine
23:06:52.942 INFO BaseRecalibrationEngine - The covariates being used here:
23:06:52.942 INFO BaseRecalibrationEngine - ReadGroupCovariate
23:06:52.942 INFO BaseRecalibrationEngine - QualityScoreCovariate
23:06:52.942 INFO BaseRecalibrationEngine - ContextCovariate
23:06:52.942 INFO BaseRecalibrationEngine - CycleCovariate
23:06:52.945 INFO ProgressMeter - Starting traversal
23:06:52.946 INFO ProgressMeter - Current Locus Elapsed Minutes Reads Processed Reads/Minute
23:07:02.952 INFO ProgressMeter - NC_001145.3:479994 0.2 582000 3490952.7
```

**gatk HaplotypeCaller \**

**-R GCF\_000146045.2\_R64\_genomic.fna \**  
**-I conA\_merged\_markdup\_rg\_bqsr.bam \**  
**-O conA\_final\_variants.vcf**

**gatk HaplotypeCaller \**

**-R GCF\_000146045.2\_R64\_genomic.fna \**  
**-I conB\_merged\_markdup\_rg\_bqsr.bam \**  
**-O conB\_final\_variants.vcf**

```

-R GCF_000146045.2_R64_genomic.fna \
-I conA_merged_markdup_rg_bqsr.bam \
-O conA_final_variants.vcf

gatk HaplotypeCaller \
-R GCF_000146045.2_R64_genomic.fna \
-I conB_merged_markdup_rg_bqsr.bam \
-O conB_final_variants.vcf
Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
  java -Dsanjdk.use_async.io.read.santools=false -Dsanjdk.use_async.io.write.santools=true -Dsanjdk.use_async.io.write.tribble=false -Dsanjdk.compression.level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar HaplotypeCaller -R GCF_000146045.2_R64_genomic.fna -I conA_merged_markdup_rg_bqsr.bam -O conA_final_variants.vcf
23:09:15.583 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@ad3f70a]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@54ccb3]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@ad3f70a]
23:09:15.698 INFO HaplotypeCaller - .....
23:09:15.700 INFO HaplotypeCaller - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:09:15.701 INFO HaplotypeCaller - For support and documentation go to https://software.broadinstitute.org/gatk/
23:09:15.701 INFO HaplotypeCaller - Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic amd64
23:09:15.701 INFO HaplotypeCaller - Java runtime: OpenJDK 64-Bit Server VM v17.0.16-8-Ubuntu-0ubuntu124.04.1
23:09:15.701 INFO HaplotypeCaller - Start Date/Time: October 16, 2025 at 11:09:15 PM IST
23:09:15.701 INFO HaplotypeCaller - .....
23:09:15.701 INFO HaplotypeCaller - .....
23:09:15.702 INFO HaplotypeCaller - HTSJDK Version: 4.1.3
23:09:15.702 INFO HaplotypeCaller - Picard Version: 3.3.0
23:09:15.702 INFO HaplotypeCaller - Built for Spark Version: 3.5.0
23:09:15.704 INFO HaplotypeCaller - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:09:15.705 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:09:15.705 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:09:15.705 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:09:15.705 INFO HaplotypeCaller - Deflater: IntelDeflater
23:09:15.705 INFO HaplotypeCaller - Inflater: IntelInflater
23:09:15.705 INFO HaplotypeCaller - GCS max retries/reopens: 20
23:09:15.705 INFO HaplotypeCaller - Requester pays: disabled
23:09:15.706 INFO HaplotypeCaller - Initializing engine
23:09:15.881 INFO HaplotypeCaller - Done initializing engine
23:09:15.893 INFO NativeLibraryLoader - Loading libgkl_utils.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_utils.so
23:09:15.895 INFO NativeLibraryLoader - Loading libgkl_smithwaterman.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_smithwaterman.so
23:09:15.896 INFO SmithWatermanAligner - Using AVX accelerated SmithWaterman implementation
23:09:15.899 INFO HaplotypeCallerEngine - Disabling physical phasing, which is supported only for reference-model confidence output
23:09:15.911 INFO NativeLibraryLoader - Loading libgkl_pairhmm_omp.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_pairhmm_omp.so
23:09:15.928 INFO IntelPairHMM - Flush-to-zero (FTZ) is enabled when running PairHMM
23:09:15.928 INFO IntelPairHMM - Available threads: 4
23:09:15.929 INFO IntelPairHMM - Requested threads: 4
23:09:15.929 INFO PairHMM - Using the OpenMP multi-threaded AVX-accelerated native PairHMM implementation
23:09:15.957 INFO ProgressMeter - Starting traversal
23:09:15.957 INFO ProgressMeter - Current Locus Elapsed Minutes Regions Processed Regions/Minute
23:09:16.246 WARN InbreedingCoeff - InbreedingCoeff will not be calculated at position NC_001133.9:376 and possibly subsequent; at least 10 samples must have called genotypes
23:09:26.015 INFO ProgressMeter - NC_001134.8:302436 0.2 2850 17001.4

```

```

Runtime.totalMemory()=538968064
Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
  java -Dsanjdk.use_async.io.read.santools=false -Dsanjdk.use_async.io.write.santools=true -Dsanjdk.use_async.io.write.tribble=false -Dsanjdk.compression.level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar HaplotypeCaller -R GCF_000146045.2_R64_genomic.fna -I conB_merged_markdup_rg_bqsr.bam -O conB_final_variants.vcf
23:11:08.584 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@ad3f70a]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@54ccb3]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@ad3f70a]
23:11:08.707 INFO HaplotypeCaller - .....
23:11:08.712 INFO HaplotypeCaller - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:11:08.712 INFO HaplotypeCaller - For support and documentation go to https://software.broadinstitute.org/gatk/
23:11:08.712 INFO HaplotypeCaller - Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic amd64
23:11:08.712 INFO HaplotypeCaller - Java runtime: OpenJDK 64-Bit Server VM v17.0.16-8-Ubuntu-0ubuntu124.04.1
23:11:08.712 INFO HaplotypeCaller - Start Date/Time: October 16, 2025 at 11:11:08 PM IST
23:11:08.713 INFO HaplotypeCaller - .....
23:11:08.713 INFO HaplotypeCaller - .....
23:11:08.713 INFO HaplotypeCaller - HTSJDK Version: 4.1.3
23:11:08.714 INFO HaplotypeCaller - Picard Version: 3.3.0
23:11:08.714 INFO HaplotypeCaller - Built for Spark Version: 3.5.0
23:11:08.716 INFO HaplotypeCaller - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:11:08.716 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:11:08.716 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:11:08.717 INFO HaplotypeCaller - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:11:08.717 INFO HaplotypeCaller - Deflater: IntelDeflater
23:11:08.717 INFO HaplotypeCaller - Inflater: IntelInflater
23:11:08.717 INFO HaplotypeCaller - GCS max retries/reopens: 20
23:11:08.717 INFO HaplotypeCaller - Requester pays: disabled
23:11:08.718 INFO HaplotypeCaller - Initializing engine
23:11:08.878 INFO HaplotypeCaller - Done initializing engine
23:11:08.887 INFO NativeLibraryLoader - Loading libgkl_utils.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_utils.so
23:11:08.889 INFO NativeLibraryLoader - Loading libgkl_smithwaterman.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_smithwaterman.so
23:11:08.890 INFO SmithWatermanAligner - Using AVX accelerated SmithWaterman implementation
23:11:08.893 INFO HaplotypeCallerEngine - Disabling physical phasing, which is supported only for reference-model confidence output
23:11:08.996 INFO NativeLibraryLoader - Loading libgkl_pairhmm_omp.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_pairhmm_omp.so
23:11:08.924 INFO IntelPairHMM - Flush-to-zero (FTZ) is enabled when running PairHMM
23:11:08.924 INFO IntelPairHMM - Available threads: 4
23:11:08.924 INFO IntelPairHMM - Requested threads: 4
23:11:08.924 INFO PairHMM - Using the OpenMP multi-threaded AVX-accelerated native PairHMM implementation
23:11:08.952 INFO ProgressMeter - Starting traversal
23:11:08.953 INFO ProgressMeter - Current Locus Elapsed Minutes Regions Processed Regions/Minute
23:11:09.373 WARN InbreedingCoeff - InbreedingCoeff will not be calculated at position NC_001133.9:27210 and possibly subsequent; at least 10 samples must have called genotypes
23:11:19.061 INFO ProgressMeter - NC_001134.8:269173 0.2 2580 15314.6
23:11:29.077 INFO ProgressMeter - NC_001136.10:427272 0.3 8870 26446.0
23:11:39.087 INFO ProgressMeter - NC_001137.3:15736 0.5 14500 28072.0
23:11:49.110 INFO ProgressMeter - NC_001138.5:103507 0.7 18860 28173.8
23:11:59.122 INFO ProgressMeter - NC_001140.6:121440 0.8 25210 30150.1
23:12:09.170 INFO ProgressMeter - NC_001142.9:189779 1.0 30810 30699.0
23:12:19.176 INFO ProgressMeter - NC_001144.5:207464 1.2 37860 32348.4
23:12:29.208 INFO ProgressMeter - NC_001145.3:217471 1.3 43650 32633.5
23:12:39.216 INFO ProgressMeter - NC_001146.8:194915 1.5 48640 32332.2

```

## gatk SelectVariants \

```
-R GCF_000146045.2_R64_genomic.fna \
-V conA_final_variants.vcf \
--select-type-to-include SNP \
-O conA_snps.vcf
```

#### gatk VariantFiltration \

```
-R GCF_000146045.2_R64_genomic.fna \
-V conA_snps.vcf \
--filter-name "QD_filter" --filter-expression "QD < 2.0" \
--filter-name "FS_filter" --filter-expression "FS > 60.0" \
--filter-name "MQ_filter" --filter-expression "MQ < 40.0" \
-O conA_snps_filtered.vcf
```

```
-R GCF_000146045.2_R64_genomic.fna \
-V conA_final_variants.vcf \
--select-type-to-include SNP \
-O conA_snps.vcf

gatk VariantFiltration \
-R GCF_000146045.2_R64_genomic.fna \
-V conA_snps.vcf \
--filter-name "QD_filter" --filter-expression "QD < 2.0" \
--filter-name "FS_filter" --filter-expression "FS > 60.0" \
--filter-name "MQ_filter" --filter-expression "MQ < 40.0" \
-O conA_snps_filtered.vcf
Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
  java -Dsanjdk.use_async_io.read_samtools=false -Dsanjdk.use_async_io.write_samtools=true -Dsanjdk.use_async_io.write_tribble=false -Dsanjdk.compression_level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar SelectVariants -R GCF_000146045.2_R64_genomic.fna -V conA_final_variants.vcf --select-type-to-include SNP -O conA_snps.vcf
23:14:07.711 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
SLF4J(W): Found provider [ch.qos.logback.classic.jul.LoggerServiceProvider@63d677f5]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
23:14:07.825 INFO SelectVariants - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:14:07.828 INFO SelectVariants - For support and documentation go to https://software.broadinstitute.org/gatk/
23:14:07.828 INFO SelectVariants - Executing as ibab@IBAB-M5C14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic and64
23:14:07.828 INFO SelectVariants - Java runtime: OpenJDK 64-Bit Server VM v17.0.16-8-Ubuntu-0ubuntu124.04.1
23:14:07.828 INFO SelectVariants - Start Date/Time: October 16, 2025 at 11:14:07 PM IST
23:14:07.829 INFO SelectVariants -
23:14:07.829 INFO SelectVariants - HTSJDK Version: 4.1.3
23:14:07.830 INFO SelectVariants - Picard Version: 3.3.0
23:14:07.830 INFO SelectVariants - Built for Spark Version: 3.5.0
23:14:07.832 INFO SelectVariants - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:14:07.833 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:14:07.833 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:14:07.833 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:14:07.833 INFO SelectVariants - Deflater: IntelDeflater
23:14:07.833 INFO SelectVariants - Inflater: IntelInflater
23:14:07.834 INFO SelectVariants - GCS max retries/reopens: 20
23:14:07.834 INFO SelectVariants - Requester pays: disabled
23:14:07.834 INFO SelectVariants - Initializing engine
23:14:07.970 INFO FeatureManager - Using codec VCFCodec to read file file:///home/ibab/SEM3/RL/101625/varcal/conA_final_variants.vcf
23:14:08.003 INFO SelectVariants - Done initializing engine
23:14:08.030 INFO ProgressMeter - Starting traversal
23:14:08.030 INFO ProgressMeter - Current Locus Elapsed Minutes Variants Processed Variants/Minute
23:14:08.550 INFO ProgressMeter - NC_001147.6:294553 0.0 12917 1519647.1
23:14:08.551 INFO ProgressMeter - Traversal complete. Processed 12917 total variants in 0.0 minutes.
23:14:08.550 INFO SelectVariants - Shutting down engine
[October 16, 2025 at 11:14:08 PM IST] org.broadinstitute.hellbender.tools.walkers.variantutils.SelectVariants done. Elapsed time: 0.02 minutes.
```

#### gatk SelectVariants \

```
-R GCF_000146045.2_R64_genomic.fna \
-V conB_final_variants.vcf \
--select-type-to-include SNP \
-O conB_snps.vcf
```

#### gatk VariantFiltration \

```
-R GCF_000146045.2_R64_genomic.fna \
-V conB_snps.vcf \
--filter-name "QD_filter" --filter-expression "QD < 2.0" \
--filter-name "FS_filter" --filter-expression "FS > 60.0" \
--filter-name "MQ_filter" --filter-expression "MQ < 40.0" \
-O conB_snps_filtered.vcf
```



```

-R GCF_000146045.2_R64_genomic.fna \
-V conB_final_variants.vcf \
--select-type-to-include SNP \
-O conB_snps.vcf

gatk VariantFiltration \
-R GCF_000146045.2_R64_genomic.fna \
-V conB_snps.vcf \
--filter-name "QD_filter" --filter-expression "QD < 2.0" \
--filter-name "FS_filter" --filter-expression "FS > 60.0" \
--filter-name "MQ_filter" --filter-expression "MQ < 40.0" \
-O conB_snps_filtered.vcf
Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
java -Dsanjdk.use_async_io_read_samtools=false -Dsanjdk.use_async_io_write_samtools=true -Dsanjdk.use_async_io_write_tribble=false -Dsanjdk.compression_level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar SelectVariants -R GCF_000146045.2_R64_genomic.fna -V conB_final_variants.vcf --select-type-to-include SNP -O conB_snps.vcf
23:15:15.697 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@63d677f5]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
23:15:15.811 INFO SelectVariants - -----
23:15:15.813 INFO SelectVariants - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:15:15.814 INFO SelectVariants - For support and documentation go to https://software.broadinstitute.org/gatk/
23:15:15.814 INFO SelectVariants - Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic amd64
23:15:15.814 INFO SelectVariants - Java runtime: OpenJDK 64-Bit Server VM v17.0.16-8-Ubuntu-0ubuntu124.04.1
23:15:15.814 INFO SelectVariants - Start Date/Time: October 16, 2025 at 11:15:15 PM IST
23:15:15.814 INFO SelectVariants - -----
23:15:15.815 INFO SelectVariants - HTSJDK Version: 4.1.3
23:15:15.815 INFO SelectVariants - Picard Version: 3.3.0
23:15:15.815 INFO SelectVariants - Built for Spark Version: 3.5.0
23:15:15.818 INFO SelectVariants - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:15:15.818 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:15:15.819 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:15:15.819 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:15:15.819 INFO SelectVariants - Deflater: IntelDeflater
23:15:15.819 INFO SelectVariants - Inflater: IntelInflater
23:15:15.819 INFO SelectVariants - GCS max retries/opens: 20
23:15:15.820 INFO SelectVariants - Requester pays: disabled
23:15:15.820 INFO SelectVariants - Initializing engine
23:15:15.950 INFO FeatureManager - Using codec VCFCodec to read file file:///home/ibab/SEM3/RL/101625/varcal/conB_final_variants.vcf
23:15:15.981 INFO SelectVariants - Done Initializing engine
23:15:16.028 INFO ProgressMeter - Starting traversal
23:15:16.028 INFO ProgressMeter - Current Locus Elapsed Minutes Variants Processed Variants/Minute
23:15:16.569 INFO ProgressMeter - NC_061147.6:293132 0.0 12901 1436103.9
23:15:16.570 INFO ProgressMeter - Traversal complete. Processed 12901 total variants in 0.0 minutes.
23:15:16.597 INFO SelectVariants - Shutting down engine
[October 16, 2025 at 11:15:16 PM IST] org.broadinstitute.hellbender.tools.walkers.variantutils.SelectVariants done. Elapsed time: 0.02 minutes.

```

## gatk SelectVariants \

```

-R GCF_000146045.2_R64_genomic.fna \
-V conA_final_variants.vcf \
--select-type-to-include INDEL \
-O conA_indels.vcf

```

## gatk VariantFiltration \

```

-R GCF_000146045.2_R64_genomic.fna \
-V conA_indels.vcf \
--filter-name "QD_filter" --filter-expression "QD < 2.0" \
--filter-name "FS_filter" --filter-expression "FS > 200.0" \
-O conA_indels_filtered.vcf

```

```
(base) ibab@IBAB-MSc14-Comp011:~/SEM3/RL/101625/varcal$ gatk SelectVariants \
-R GCF_000146045.2_R64_genomic.fna \
-V conA_final_variants.vcf \
--select-type-to-include INDEL \
-O conA_indels.vcf

gatk VariantFiltration \
-R GCF_000146045.2_R64_genomic.fna \
-V conA_indels.vcf \
--filter-name "QD_filter" --filter-expression "QD < 2.0" \
--filter-name "FS_filter" --filter-expression "FS > 200.0" \
-O conA_indels_filtered.vcf

Using CATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
java -Dsamjdk.use_async_io_read_samtools=false -Dsamjdk.use_async_io_write_samtools=true -Dsamjdk.use_async_io_write_tribble=false -Dsamjdk.compression_level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar SelectVariants -R GCF_000146045.2_R64_genomic.fna -V conA_final_variants.vcf --select-type-to-include INDEL -O conA_indels.vcf
23:16:16.142 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@63d677f5]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
23:16:16.258 INFO SelectVariants - 
23:16:16.261 INFO SelectVariants - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:16:16.261 INFO SelectVariants - For support and documentation go to https://software.broadinstitute.org/gatk/
23:16:16.261 INFO SelectVariants - Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic amd64
23:16:16.261 INFO SelectVariants - Java runtime: OpenJDK 64-Bit Server VM v17.0.16+8-Ubuntu-0ubuntu124.04.1
23:16:16.262 INFO SelectVariants - Start Date/Time: October 16, 2025 at 11:16:16 PM IST
23:16:16.262 INFO SelectVariants - 
23:16:16.262 INFO SelectVariants - 
23:16:16.263 INFO SelectVariants - HTSJDK Version: 4.1.3
23:16:16.263 INFO SelectVariants - Picard Version: 3.1.0
23:16:16.263 INFO SelectVariants - Built for Spark Version: 3.5.0
23:16:16.265 INFO SelectVariants - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:16:16.266 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:16:16.266 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:16:16.266 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:16:16.266 INFO SelectVariants - Deflater: IntelDeflater
23:16:16.266 INFO SelectVariants - Inflater: IntelInflater
23:16:16.266 INFO SelectVariants - GCS max retries/reopens: 20
23:16:16.266 INFO SelectVariants - Requester pays: disabled
23:16:16.267 INFO SelectVariants - Initializing engine
23:16:16.397 INFO FeatureManager - Using codec VCFCodec to read file file:///home/ibab/SEM3/RL/101625/varcal/conA_final_variants.vcf
23:16:16.431 INFO SelectVariants - Done initializing engine
23:16:16.461 INFO ProgressMeter - Starting traversal
23:16:16.461 INFO ProgressMeter - Current Locus Elapsed Minutes Variants Processed Variants/Minute
23:16:16.728 INFO ProgressMeter - unmapped 0.0 736 170501.9
23:16:16.728 INFO ProgressMeter - Traversal complete, Processed 736 total variants in 0.0 minutes.
23:16:16.725 INFO SelectVariants - Shutting down engine
[October 16, 2025 at 11:16:16 PM IST] org.broadinstitute.hellbender.tools.walkers.variantutils.SelectVariants done. Elapsed time: 0.01 minutes.
Runtime.totalMemory()=119537664
```

**gatk SelectVariants \**

**-R GCF\_000146045.2\_R64\_genomic.fna \**

**-V conB\_final\_variants.vcf \**

**--select-type-to-include INDEL \**

**-O conB\_indels.vcf**

**# Apply hard filters to conB indels**

**gatk VariantFiltration \**

**-R GCF\_000146045.2\_R64\_genomic.fna \**

**-V conB\_indels.vcf \**

**--filter-name "QD\_filter" --filter-expression "QD < 2.0" \**

**--filter-name "FS\_filter" --filter-expression "FS > 200.0" \**

**-O conB\_indels\_filtered.vcf**

```
(base) ibab@IBAB-MSc14-Comp011:~/SEM3/RL/101625/varcal$ # Select indels from conB
gatk SelectVariants \
  -R GCF_000146045.2_R64_genomic.fna \
  -V conB_final_variants.vcf \
  --select-type-to-include INDEL \
  -O conB_indels.vcf

# Apply hard filters to conB indels
gatk VariantFiltration \
  -R GCF_000146045.2_R64_genomic.fna \
  -V conB_indels.vcf \
  --filter-name "QD_filter" --filter-expression "QD < 2.0" \
  --filter-name "FS_filter" --filter-expression "FS > 200.0" \
  -O conB_indels_filtered.vcf
Using GATK jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar
Running:
  java -Dsamjdk.use_async_io_read_samtools=false -Dsamjdk.use_async_io_write_samtools=true -Dsamjdk.use_async_io_write_tribble=false -Dsamjdk.compression_level=2 -jar /home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar SelectVariants -R GCF_000146045.2_R64_genomic.fna -V conB_final_variants.vcf --select-type-to-include INDEL -O conB_indels.vcf
23:16:57.789 INFO NativeLibraryLoader - Loading libgkl_compression.so from jar:file:/home/ibab/SEM3/RL/gatk-4.6.1.0/gatk-package-4.6.1.0-local.jar!/com/intel/gkl/native/libgkl_compression.so
SLF4J(W): Class path contains multiple SLF4J providers.
SLF4J(W): Found provider [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
SLF4J(W): Found provider [ch.qos.logback.classic.spi.LogbackServiceProvider@63d677f5]
SLF4J(W): See https://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J(I): Actual provider is of type [org.apache.logging.slf4j.SLF4JServiceProvider@52ba685a]
23:16:57.905 INFO SelectVariants - .....
23:16:57.908 INFO SelectVariants - The Genome Analysis Toolkit (GATK) v4.6.1.0
23:16:57.908 INFO SelectVariants - For support and documentation go to https://software.broadinstitute.org/gatk/
23:16:57.909 INFO SelectVariants - Executing as ibab@IBAB-MSc14-Comp011.ibab.ac.in on Linux v6.14.0-29-generic amd64
23:16:57.909 INFO SelectVariants - Java runtime: OpenJDK 64-Bit Server VM v17.0.16+8-Ubuntu-0ubuntu124.04.1
23:16:57.909 INFO SelectVariants - Start Date/Time: October 16, 2025 at 11:16:57 PM IST
23:16:57.909 INFO SelectVariants - .....
23:16:57.909 INFO SelectVariants - .....
23:16:57.910 INFO SelectVariants - HTSJDK Version: 4.1.3
23:16:57.910 INFO SelectVariants - Picard Version: 3.3.0
23:16:57.910 INFO SelectVariants - Built for Spark Version: 3.5.0
23:16:57.912 INFO SelectVariants - HTSJDK Defaults.COMPRESSION_LEVEL : 2
23:16:57.913 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_READ_FOR_SAMTOOLS : false
23:16:57.913 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_SAMTOOLS : true
23:16:57.913 INFO SelectVariants - HTSJDK Defaults.USE_ASYNC_IO_WRITE_FOR_TRIBBLE : false
23:16:57.913 INFO SelectVariants - Deflater: IntelDeflater
23:16:57.913 INFO SelectVariants - Inflater: IntelInflater
23:16:57.913 INFO SelectVariants - GCS max retries/reopens: 20
23:16:57.913 INFO SelectVariants - Requester pays: disabled
23:16:57.914 INFO SelectVariants - Initializing engine
23:16:58.041 INFO FeatureManager - Using codec VCFCodec to read file file:///home/ibab/SEM3/RL/101625/varcal/conB_final_variants.vcf
23:16:58.069 INFO SelectVariants - Done initializing engine
23:16:58.100 INFO ProgressMeter - Starting traversal
23:16:58.100 INFO ProgressMeter - Current Locus Elapsed Minutes Variants Processed Variants/Minute
23:16:58.381 INFO ProgressMeter - unmapped 0.0 720 154285.7
23:16:58.381 INFO ProgressMeter - Traversal complete, Processed 720 total variants in 0.0 minutes.
23:16:58.386 INFO SelectVariants - Shutting down engine
[October 16, 2025 at 11:16:58 PM IST] org.broadinstitute.hellbender.tools.walkers.variantutils.SelectVariants done. Elapsed time: 0.01 minutes.
Runtime.totalMemory()=335544320
```

```
awk '!/^#/ {print $1}' conA_variants_final.filtered.PASS.vcf | sort -u | head
```

```
cat > rename.txt << 'EOF'
```

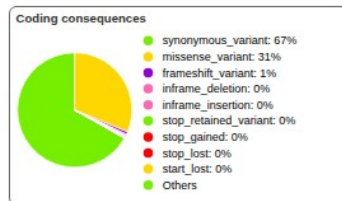
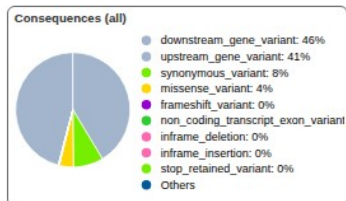
```
NC_001133.9   chrI
NC_001134.8   chrII
NC_001135.5   chrIII
NC_001136.10  chrIV
NC_001137.3   chrV
NC_001138.5   chrVI
NC_001139.9   chrVII
NC_001140.6   chrVIII
NC_001141.2   chrIX
NC_001142.9   chrX
NC_001143.9   chrXI
NC_001144.5   chrXII
NC_001145.3   chrXIII
NC_001146.8   chrXIV
NC_001147.6   chrXV
NC_001148.4   chrXVI
NC_001224.1   chrM
EOF
```

```
bcftools annotate --rename-chrs rename.txt conA_variants_final.filtered.PASS.vcf -Ov -o conA_variants_renamed.vcf
```

## In Ensembl VEP selecting *Saccharomyces cerevisiae*










### Summary statistics

Category	Count
Variants processed	9432
Variants filtered out	0
Novel / existing variants	1042 (11.0) / 8390 (89.0)
Overlapped genes	3278
Overlapped transcripts	3278
Overlapped regulatory features	-



### Results preview

Navigation (per variant)

Page:   1 of 1887   | Show:      variants

Filters

Uploaded variant  is  defined 

 Download 

All: [VCF](#) [VEP](#) [TXT](#)

BioMart: [Variants](#) [Genes](#)