



**bamsort**

# Table of contents

Quick Start

---

Compatible Picard Command

---

bamsort Reference

---

Sort BAM files.

This tool can sort the reads within a BAM file in a variety of ways, including by position in the genome (coordinate) or read name (queryname). This enables compatibility with the requirements of different downstream tools.

Five sort modes are supported:

- coordinate (Picard-compatible)
- coordinate (fgbio-compatible)
- queryname (Picard-compatible)
- queryname (fgbio-compatible)
- template coordinate sort (fgbio-compatible)

Allowed values for **--sort-order** are as follows:

- coordinate [default]
- queryname
- templatecoordinate

Allowed values for **--sort-compatibility** are as follows:

- picard [default]
- fgbio

*coordinate* and *queryname* sorting can be done in either *picard* or *fgbio* mode. *templatecoordinate* can only be done in *fgbio* mode.

## Quick Start

```
$ # This command assumes all the inputs are in INPUT_DIR and all the outputs go to  
OUTPUT_DIR. docker run --rm --gpus all --volume INPUT_DIR:/workdir --volume  
OUTPUT_DIR:/outputdir \ --workdir /workdir \ nvc.io/nvidia/clara/clara-  
parabricks:4.3.1-1 \ pbrun bamsort \ --ref /workdir/${REFERENCE_FILE} \ --in-bam  
/workdir/${INPUT_BAM} \ --out-bam /outputdir/${OUTPUT_BAM} \ --sort-order  
coordinate
```

## Compatible Picard Command

The command below is the Picard counterpart of the Parabricks command above. The output from this command will be identical to the output from the above command.

```
$ java -Xmx30g -jar picard.jar SortSam \ I=<INPUT_DIR>/${INPUT_BAM} \ O=  
<OUTPUT_DIR>/${OUTPUT_BAM}
```

## bamsort Reference

Sort BAM files. There are five modes: Coordinate sort (Picard-compatible), Coordinate sort (fgbio-compatible), queryname sort (Picard-compatible), queryname sort (fgbio-compatible), and template coordinate sort (fgbio-compatible).

### Input/Output file options

--in-bam IN\_BAM

Path of BAM/CRAM for sorting. This option is required. (default: None)

Option is required.

--out-bam OUT\_BAM

Path of BAM/CRAM file after sorting. (default: None)

Option is required.

--ref REF

Path to the reference file. (default: None)

Option is required.

## **Pipeline Options:**

--sort-order SORT\_ORDER

Type of sort to be done. Possible values are {coordinate,queryname,templatecoordinate}. (default: coordinate)

--sort-compatibility SORT\_COMPATIBILITY

Sort comparator compatibility to be used for compatibility with other tools. Possible values are {picard,fgbio}. TemplateCoordinate will only use fgbio. (default: picard)

## **Performance Options:**

--num-zip-threads NUM\_ZIP\_THREADS

Number of CPUs to use for zipping BAM files in a run (default 16 for coordinate sorts and 10 otherwise). (default: None)

--num-sort-threads NUM\_SORT\_THREADS

Number of CPUs to use for sorting in a run (default 10 for coordinate sorts and 16 otherwise). (default: None)

--max-records-in-ram MAX\_RECORDS\_IN\_RAM

Maximum number of records in RAM when using a queryname or template coordinate sort mode; lowering this number will decrease maximum memory usage. (default: 65000000)

--mem-limit MEM\_LIMIT

Memory limit in GBs during sorting and postsorting. By default, the limit is half of the total system memory. (default: 62)

## **Common options:**

--logfile LOGFILE

Path to the log file. If not specified, messages will only be written to the standard error output. (default: None)

--tmp-dir TMP\_DIR

Full path to the directory where temporary files will be stored.

--with-petogene-dir WITH\_PETAGENE\_DIR

Full path to the PetaGene installation directory. By default, this should have been installed at /opt/petogene. Use of this option also requires that the PetaLink library has been preloaded by setting the LD\_PRELOAD environment variable. Optionally set the PETASUITE\_REFPATH and PGCLOUD\_CREDPATH environment variables that are used for data and credentials (default: None)

--keep-tmp

Do not delete the directory storing temporary files after completion.

--no-seccomp-override

Do not override seccomp options for docker (default: None).

--version

View compatible software versions.

## **GPU options:**

--num-gpus NUM\_GPUS

Number of GPUs to use for a run. GPUs 0..(NUM\_GPUS-1) will be used.

© Copyright 2024, Nvidia.. PDF Generated on 06/05/2024