bam2fq

Table of contents

Run bam2fq to convert BAM/CRAM to FASTQ.

This tool un-aligns a BAM file, reversing it from BAM to FASTQ format. This can be useful if the BAM needs to be re-aligned to a newer or different reference genome by applying bam2fq followed by fq2bam (BWA-MEM + GATK) with the new reference genome.

For paired reads, bam2fq will append "/1" to the 1st read name, and "/2" to the 2nd read name.

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 \ nvcr.io/nvidia/clara/clara-
parabricks:4.3.1-1 \ pbrun bam2fq \ --ref /workdir/${REFERENCE_FILE} \ --in-bam
/workdir/${INPUT_BAM} \ --out-prefix /workdir/${Prefix_for_output_fastq_files}
```

Compatible CPU-based BWA-MEM, GATK4 Commands

The command below is the bwa-0.7.15 and GATK4 counterpart of the Parabricks command above. The output from these commands will be identical to the output from the above command. See the Output Comparison page for comparing the results.

```
$ gatk SamToFastq \ -I <INPUT_DIR>/${INPUT_BAM} \ -F
<OUTPUT_DIR>/${OUTPUT_FASTQ_1} \ -F2 <OUTPUT_DIR>/${OUTPUT_FASTQ_2}
```

bam2fq Reference

Run bam2fq to convert BAM/CRAM to FASTQ.

Input/Output file options

--ref REF

Path to the reference file. This argument is only required for CRAM input. (default: None)

--in-bam IN_BAM

Path to the input BAM/CRAM file to convert to fastq.gz. (default: None)

Option is required.

--out-prefix OUT_PREFIX

Prefix filename for output fastq files. (default: None)

Option is required.

Tool Options:

--out-suffixF OUT_SUFFIXF

Output suffix used for paired reads that are first in pair. The suffix must end with ".gz". (default: _1.fastq.gz)

--out-suffixF2 OUT_SUFFIXF2

Output suffix used for paired reads that are second in pair. The suffix must end with ".gz". (default: _2.fastq.gz)

--out-suffixO OUT_SUFFIXO

Output suffix used for orphan/unmatched reads that are first in pair. The suffix must end with ".gz". If no suffix is provided, these reads will be ignored. (default: None)

--out-suffixO2 OUT_SUFFIXO2

Output suffix used for orphan/unmatched reads that are second in pair. The suffix must end with ".gz". If no suffix is provided, these reads will be ignored. (default: None)

--out-suffixS OUT_SUFFIXS

Output suffix used for single-end/unpaired reads. The suffix must end with ".gz". If no suffix is provided, these reads will be ignored. (default: None)

--rg-tag RG_TAG

Split reads into different fastq files based on the read group tag. Must be either PU or ID. (default: None)

--remove-qc-failure

Remove reads from the output that have abstract QC failure. (default: None)

Performance Options:

--num-threads NUM_THREADS

Number of threads to run. (default: 8)

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-petagene-dir WITH_PETAGENE_DIR

Full path to the PetaGene installation directory. By default, this should have been installed at /opt/petagene. 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.