

Open Genomics Engine - Command Documentation

Lee Baker
Virginia Bioinformatics Institute
lee@leecbaker.com

April 12, 2012

1 Introduction

2 Commands

OpenGE consists of a collection of operations that can be invoked using the OpenGE should be called with the following syntax:

`openge command [options]`

where *command* is one of the commands listed in the following table, and *options* are various flags and parameters that may be supplied depending on the command:

Command	Description
convert	Convert a SAM file to a BAM file
count	Count alignments in a BAM file
dedup	Mark duplicates
head	Create a BAM file with the first <i>n</i> alignments of another BAM file
help	Get help for any OpenGE command
mergesort	Merges and sorts multiple bamfiles

2.1 General

The following flags are common between for all commands. Additional flags and parameters are available for individual commands, depending on the command.

Flag	Long flag	Description
-i <i>filename</i>	--in <i>filename</i>	Input filename. Input files can also be specified without this flag. Optional.
-o <i>filename</i>	--out <i>filename</i>	Output filename. Defaults to stdout if omitted.
-v	--verbose	Display additional information to stderr while processing, including progress indicators and informational messages. Optional.
-t	--threads	Select the number of threads to be used when doing parallel processing. Optional.
-d	--nothreads	Disable multithreaded processing.

When input or output files are required, stdin or stdout may be used by simply omitting the input filename or output filename. For instance, the following command:

```
openge mergesort
```

by itself will read a BAM file from command line, process it, and print the results to stdout. This enables commands to be chained together. For instance:

```
openge convert in.sam | openge count
```

will count the number of lines in a SAM file. Console redirection can be used as well. The following two commands are equivalent in a bash shell:

```
openge mergesort a.bam -o c.bam
```

```
openge mergesort < a.bam > c.bam
```

2.2 convert

Converts a SAM file to a BAM file.

Example:

```
openge convert in.sam -o out.bam
```

2.3 count

The count command returns the number of alignments contained in one or more BAM files. No parameters (other than input files) are required.

Example:

```
openge count a.bam
```

```
openge count b.bam c.bam d.bam
```

2.4 dedup

To be completed at a later date.

2.5 head

A single bamfile is generated from the first n alignments in the provided file. The header information saved is the same as in the input file.

For example,

```
openge head -o output.bam infile.bam -n 500
```

generates a BAM file called output.bam that contains the header and the first 500 alignments contained in infile.bam.

Parameters:

Flag	Long flag	Description
-n	-n	Number of alignments to include in the generated file

2.6 help

The help command shows available options for any available command.

```
openge help [command]
```

For example:

```
openge help mergesort
```

```
openge help count
```

Available commands can be seen in the table at the top of this section.

2.7 mergesort

This command generates a single bamfile from one or more input files. The supplied input files are merged and then sorted by alignment position or region.

Parameters:

Flag	Long flag	Description
-c <i>level</i>	--compression <i>level</i>	Compression level- defaults to 6. Valid levels are 0-9, and correspond to zlib's deflate compression levels.
-r <i>region</i>	--region <i>region</i>	Genomic region to use
-n <i>alignments</i>	--n <i>alignments</i>	Number of alignments to use per temporary file. Defaults to 500,000.
-C	--compresstempfiles	Compress temp files. By default, temporary files are uncompressed.

The number of alignments per temporary file has a large effect on speed when processing large datasets. Adjusting the number of alignments upwards produces larger (and fewer) temporary files at the expense of memory, but reduces the processing time.

3 License

©2012 Virginia Bioinformatics Institute. All rights reserved.

Note: License may change at a future date.