IonCram

Introduction

IonCram is the first compression tool that efficiently compresses the Ion Torrent BAM files. IonCram extends the popular CRAM program by improving the compression of the flow signals. IonCram could improve the compression of CRAM by 13% achieving an overall space saving of about 45%.

Complilation

IonCram uses different compression techniques for compressing the flow signal(gzip, bzip2, xz, and zstd). To use zstd, Please install it(https://github.com/facebook/zstd) and make it system available.

Ubuntu

```
apt-get install g++ libncurses5-dev  parallel python3 make libbz2-dev zstd
zlib1g-dev liblzma-dev automake libtool samtools time
./configure --binDir <install directory>
make
make install
```

If install directory is not supplied "ioncram/build/" will be used

Centos

```
yum -y install https://centos7.iuscommunity.org/ius-release.rpm
yum  -y install  gcc-c++ compat-gcc-32 compat-gcc-32-c++ ncurses-devel parallel
xz-devel bzip2-devel make zlib-devel automake zstd file libtool samtools time
python34u

./configure --binDir <install directory>
make
make install
```

If install directory is not supplied "ioncram/build/" will be used

---

Test The System

```
gzip -d test_data/chrY.fa.gz
./ioncram compress -i test_data/test1.bam -r test_data/chrY.fa -c -o test1.ioncram
```

Usage

```
./ioncram compress -i input.bam -o compressed.IonCram -r hg19.fa

./ioncram decompress -i compressed.IonCram -o original.bam -r hg19.fa
```

Statistics scripts

Scripts to calculate statisitics about the flowsignal and compression ratio of CRAM and ioncram can be found under scripts/ folder.

Scripts take the folder as argument where the ioncram binaries are installed.

FlowSignal Data Size

```
scripts/flowSignalInBam.sh <binaries folder> <bam file>
```

CRAM Statisitcs

```
scripts/scrambleStatistics.sh <binaries folder> <bam file> <reference file>
```

ioncram Statisitcs

```
scripts/ionStatistics.sh <binaries folder> <bam file> <reference file> <compression tool>
```

Compression tools supported are : gzip, xz, and zstd.

Tools

ioncram has five commands compress, decompress, compare, version, and cite, You can list the available commands by "./ioncram --help". Every command has its own help page.

Compress

Tool for compressing SAM/BAM into IonCram format.

Usage

```
./ioncram compress [options] -i <input SAM/BAM> -r <input reference FASTA> -o
<outputfile>
```

Required arguments

- -i , --input-file :
  - File name of the input file can be SAM/BAM
- -o , --output-file:
  - File name of the output IonCram file
- -r , --reference:
  - Reference fasta file used by CRAM compression. Should be the same reference used in the
    mapping phase

Optional arguments

- -b , --bed-file:
  - reads overlapping this BED FILE
- -c , --check:
  - Check the compressed file and compare it with the original
- -z , --compression:
  - Compression technique used to compress Flow signal. Options are xz, gzip, bzip2, and zstd. xz
    is the default technique
- -v , --verbose:
  - Verbose mode prints the size used by every field in the l_data
- -l , --lossy:
  - Lossy mode discard the highly repeated reads and uses only one read to represent.
- -d , --deez:

* Use deez tool instead of samtools and CRAM

Decompress

Tool for decompressing IonCram into BAM format

Usage

```
./ioncram decompress -i <input IonCram> -r <input reference FASTA> -o <outputfile
bam>
```

Required arguments

- -i , --input-file :
  - File name of the input file can be SAM/BAM
- -o , --output-file:
  - File name of the output IonCram file

- -r , --reference:

* Reference fasta file used by CRAM compression. Should be the same reference used in the mapping phase

Compare

Tool for comparing SAM/BAM/IonCram files

Usage

```
./ioncram compare -i1 <input file1> -i2 <input file1> -r <input reference FASTA>
```

Required arguments

- -i1 , --input-file1 :
  - File name of the input file can be SAM/BAM/CRAM/IonCram
- -i2 , --input-file2 :
  - File name of the input file can be SAM/BAM/CRAM/IonCram
- -r , --reference:
  - Reference fasta file used by CRAM compression. Should be the same reference used in the mapping phase

Cite

Not Published yet

License

Opensource