BAM - Mapping Locations

SAM (Sequence Alignment/Map) format is a generic format for storing large nucleotide sequence alignments. BAM format is the compressed binary version of SAM format. Further information about these formats can be found in the SAM format specifications document.

The Flux Capacitor needs the SAM input file to be indexed. The index is needed to access every locus indipendently, without the need to sequentially read the whole file. For this reason only **BAM** files are supported.

The BAM file must be sorted by **the reference ID and then the leftmost coordinate before indexing** (see SAM format specification above). The index file has to be placed in the same folder as the BAM file.

The BAM file cannot contain multiple alignments per entry/line ("compact format" NOT allowed).

Example.

The following example shows a valid mapped read-pair in SAM format:

```
ID:1:2:3    129    chr1    127926    1    75M    =    128047    122
CTACCAGGGCCGCTGGGAGCTGGGCAGGAGCTGAGTCCAAAGACGTTGTTGGGACCTGGAGTCGGGCCAGAGTCCG
@@@FFFFFHDHFFGGIIGHGIIJIIIFGFEEFHECDHGCBHIGIIDCACA(;5?@?ED@;?;C?688;?(82::>?
ID:1:2:3    65    chr1    128047    1    75M    =    127926    -122
CCGGGAGGCTGCAAGTGGGTCTGAGAGGCCAACTTGAGGAGGCCTGGCCTCTGCCTCCCACATTGCCCAGCTGTTC
@@@FFADFGHHHHGIGHGCGGIIIGGHCHHIJJJIJIGD?FDGHIGHIIIIJAHGHHHGFD?DECCCCE?DCC>@C
```

Multiple alignments.

The Flux Capacitor supports input files in SAM format containing multiple mappings. These alignments should be represented in an extended format, that is each line contains a single alignment and flag *256* have to be used to specify that the alignment is **secondary**. The usage of optional fields for representing multiple alignments on a single line is currently not supported.

The following example shows the required representation for multiple alignments:

```
ID:1:2:3    385    chr1    135712    1    75M    =    135833    122
CTACCAGGGCCGCTGGGAGCTGGGCAGGAGCTGAGTCCAAAGACGTTGTTGGGACCTGGAGTCGGGCCAGAGTCCG
@@@FFFFFHDHFFGGIIGHGIIJIIIFGFEEFHECDHGCBHIGIIDCACA(;5?@?ED@;?;C?688;?(82::>?
ID:1:2:3    321    chr1    135833    1    75M    =    135712    -122
CCGGGAGGCTGCAAGTGGGTCTGAGAGGCCAACTTGAGGAGGCCTGGCCTCTGCCTCCCACATTGCCCAGCTGTTC
@@@FFADFGHHHHGIGHGCGGIIIGGHCHHIJJJIJIGD?FDGHIGHIIIIJAHGHHHGFD?DECCCCE?DCC>@C
ID:1:2:3    385    chr1    662078    1    75M    =    662199    122
CTACCAGGGCCGCTGGGAGCTGGGCAGGAGCTGAGTCCAAAGACGTTGTTGGGACCTGGAGTCGGGCCAGAGTCCG
@@@FFFFFHDHFFGGIIGHGIIJIIIFGFEEFHECDHGCBHIGIIDCACA(;5?@?ED@;?;C?688;?(82::>?
ID:1:2:3    321    chr1    662199    1    75M    =    662078    -122
CCGGGAGGCTGCAAGTGGGTCTGAGAGGCCAACTTGAGGAGGCCTGGCCTCTGCCTCCCACATTGCCCAGCTGTTC
@@@FFADFGHHHHGIGHGCGGIIIGGHCHHIJJJIJIGD?FDGHIGHIIIIJAHGHHHGFD?DECCCCE?DCC>@C
```

Using SAM tools to pre-process input files for the Flux Capacitor.

SAM Tools provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format. Please see more here. These tools can be used to prepare a BAM file which can be used as the input file for the Flux Capacitor.

1.Convert SAM file to BAM file:

```
samtools view -Sb file.sam > file.bam
```

2.Sort the BAM file:

```
samtools sort file.bam file_sorted
```

3.Create the index:

to create the index, the BAM file has to be sorted by genomic position. Then you could run:

```
samtools index file_sorted.bam
```