

Documentation: MatesToGff v0.2.03

1 Program

1.1 Program Name *(required)*

MatesToGff.java

1.2 Program Version *(required)*

v0.2.03

2 Description

2.1 Application *(required)*

MatesToGff converts a FASTA file, *maFile*, of mated, mapped reads (and optionally an associated quality value file) to a SOLiD GFF v0.2 file, *gffFile*. A "read" is a dibase color encoded string.

2.2 Description

n/a

2.3 Revision Information

n/a

2.4 Algorithm/Script Description

See the file: GFF_format_for_SOLID_v0.2.03.doc

3 Usage

3.1 Usage Parameters *(required)*

```
Usage: MatesToGff <matesIn.mates>
      [--f3qv=<F3_QV.qual>]
      [--r3qv=<R3_QV.qual>]
      [--ref=<ref.fasta>]
      [--fout=<sortedFragments.gff>]
      [--mout=<sortedMates.gff>]
      [--tints=a[g[y[r]]]]
      [--b]
      [--cn]
      [--tempdir=<temporaryDirectory>]
```

where:

`matesIn.mates` the required input mates file.

`--f3qv` the quality value file for the F3 fragment color calls.
Not added if not specified.

- `--r3qv` the quality value file for the R3 fragment color calls.
Not added if not specified.
- `--ref` the file containing the reference sequence. MatesToGff will
annotate the color calls only if `--ref` is specified.
- `--fout` the output file for all Fragments in
order of increasing start point.
- `--mout` the output file for all Mates in
order of increasing start point of the F3 fragment.
The fragment mated to an F3 fragment appears immediately
on the next line. MatesToGff writes these data to standard
out if and only if BOTH `--mout` and `--fout` are not specified.
- `--tints` the annotations to use for color mismatches. See
AnnotateChanges for details.
- `--b` causes MatesToGff to add a corrected base sequence to every
fragment.
No effect if `--ref` is not specified.
- `--cn` Contig Names. Writes a table-of-contents to the head of the
file, with lines like `##contig 3 name-of-the-third-contig
(i=3)`
- `--tempdir` the temporary directory that MatesToGff should use.
Defaults to `/scratch`

3.2 Usage Example *(required)*

```
MatesToGff <matesIn.mates>  
    [--f3qv=<F3_QV.qual>]  
    [--r3qv=<R3_QV.qual>]  
    [--ref=<ref.fasta>]  
    [--fout=<sortedFragments.gff>]  
    [--mout=<sortedMates.gff>]  
    [--tints=a[g[y[r]]]]  
    [--b]  
    [--cn]  
  
    [--tempdir=<temporaryDirectory>]
```

3.3 List of programs/scripts included *(required)*

MatesToGff.java

3.4 Other scripts called by this program

None

3.5 Path constraints

Scratch space with large available memory must be defined.

3.6 System Input Files *(required)*

xx.ma and xx.qual .

3.7 Input File Versions supported *(required)*

Compatible with v1 and v2 versions of .ma and .qual files.

3.8 Additional Input Files

None

3.9 Input File Comments

n/a

3.10 Output File(s) *(required)*

xx.gff

See the documentation in the file: GFF_format_for_SOLID_v0.2.03.doc for full information on the format of the gff file.

3.11 Output File(s) Comments

See the documentation in the file: GFF_format_for_SOLID_v0.2.03.doc for full information on the format of the gff file.

3.12 Sample Input File(s) *(required)*

See the files in MaToGffSampleData.0.2.03.zip for sample input files test_S1_F3_QV.qual.txt, test_S1_R3_QV.qual.txt, test_S1_F3.csfasta.ma.20.3 and test_S1_R3.csfasta.ma.20.3.

Use the following commands to create a mate-pair GFF file (this is without a translation of each mapped read from colorspace to predicted base calls – see MaToGff documentation for an example containing a translation from colorspace to basespace:

```
MaToGff.java --sort --qvs=test_S1_F3_QV.qual.txt --convert=beads --first --tempdir=./tmp  
test_S1_F3.csfasta.ma.20.3
```

```
sample-results\F3-beads-first.v2.gff
```

```
MaToGff.java --sort --qvs=test_S1_R3_QV.qual.txt --convert=beads --first --tempdir=./tmp  
test_S1_R3.csfasta.ma.20.3
```

```
sample-results\R3-beads-first.v2.gff
```

```
MatesToGff --f3qv=./reads1/test_S1_F3_QV.qual --r3qv=./reads2/test_S1_R3_QV.qual --  
ref=/data/results/RegressionDriver/CaseManager/knownData/validatedReference/matchingPipeli  
ne/ecoli_k12_MG1655.fasta --  
fout=/data/results/RegressionDriver/CaseManager/results/r12/integration/case0002/F3_R3_all_s  
orted.v2.gff --  
mout=/data/results/RegressionDriver/CaseManager/results/r12/integration/case0002/F3_R3_pair  
ed.v2.gff --tempdir=/scratch F3_R3.mates
```

```
sample-results\F3_R3_paired-sortByStart.v2.gff  
sample-results\F3_R3_paired.v2.gff
```

3.13 Sample Output File(s) *(required)*

See the following files in MaToGffSampleData.0.2.03.zip .
sample-results\F3_R3_paired-sortByStart.v2.gff and
sample-results\F3_R3_paired.v2.gff

These files should be identical to output files created using the commands above.

3.14 Supports AB kit or protocol or sample prep method

n/a.

4 Other

4.1 Development language *(required)*

Java.

4.2 Compiled for: *(required)*

Any OS.

4.3 PBS is Required? *(required)*

No

4.4 Comments

None.

4.5 Date *(required)*

May 30, 2008