

Lightning - Bug #12245

pasta rotini-fastj does not handle tagset without ending newline

09/13/2017 04:30 PM - Abram Connelly

Status:	New	Start date:	09/13/2017
Priority:	Normal	Due date:	
Assigned To:		% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	Tiling 1.0		

Description

There is a bug in the pasta rotini-fastj conversion where if the tagset provided does not have a trailing newline, the last tag will be ignored.

For example, the following should work but doesn't:

```
pasta -action rotini-fastj \  
-start 0 \  
-tilepath 0000 \  
-chrom chr1 \  
-build hg19 \  
-i stage/hu034DB1-GS00253-DNA_A02/0000.pa \  
-assembly <( 17g assembly assembly.00.hg19.fw.gz 0000 ) \  
-tag <( samtools faidx tagset.fa.gz 0000.00 | egrep -v '^>' | tr -d '\n' | fold -w 24 ) > stage/  
hu034DB1-GS00253-DNA_A02/0000.fj
```

The culprit being the

```
samtools faidx tagset.fa.gz 0000.00 | egrep -v '^>' | tr -d '\n' | fold -w 24
```

As a workaround, adding an extra newline will correct the issue:

```
pasta -action rotini-fastj \  
-start 0 \  
-tilepath 0000 \  
-chrom chr1 \  
-build hg19 \  
-i stage/hu034DB1-GS00253-DNA_A02/0000.pa \  
-assembly <( 17g assembly assembly.00.hg19.fw.gz 0000 ) \  
-tag <( cat <( samtools faidx tagset.fa.gz 0000.00 | egrep -v '^>' | tr -d '\n' | fold -w 24 ) <  
( echo "" ) ) > stage/hu034DB1-GS00253-DNA_A02/0000.fj
```

The place to look is the [readTag function in pasta_fastj.go](#) and where the [g.TagFinished flag is referenced](#) but the details of what's wrong need to be investigated.

This bug specifically happens when there is a variant on the next to last tag in the stream being converted, as is the case for data set hu034DB1-GS00253-DNA_A02 in tilepath 0000.

History

#1 - 04/10/2018 05:58 PM - Sarah Zaranek

Abram has this been handled?

#2 - 05/15/2018 06:28 AM - Abram Connelly

Note this happens with pasta version 0.2.6 and lower.

#3 - 04/30/2019 05:43 PM - Sarah Zaranek

Currently the cwl pipeline using 0.2.6. Need to insure that this fix in all relevant code.

#4 - 04/30/2019 05:49 PM - Jiayong Li

In convert2fastj of gvcf src code, there's the line

```
$pasta -action rotini-fastj -start $tilepath_start0 -tilepath $tilepath -chrom $chrom -build $ref \  
-i $odir/$tilepath.pa \  
-assembly <( $tile_assembly tilepath $afn $tilepath ) \  
-tag <( cat <( samtools faidx $tagdir $tilepath.00 | egrep -v '^>' | tr -d '\n' | fold -w 24 ) <( echo )  
) > $odir/$tilepath.fj
```

We need to verify 'echo' is functionally the same as 'echo ""'.

#5 - 04/30/2019 05:50 PM - Jiayong Li

- Target version set to Tiling 1.0

#6 - 04/30/2019 06:07 PM - Forrest Hartley

Pasta 0.2.8 can be found in the l7g abramVersion branch. This needs to be tested thoroughly.