STAR Transgenes

Input FASTA File Requirements

1. If you want to incorporate 3 trangenes into a genome, you must provide 3 FASTA files, one for each transgene. DO NOT combine them into a single FASTA file.
2. Each FASTA file must conform to the standard FASTA file format specification.
3. The entire sequence must be in one line.
4. The last line of the FASTA file must end with a newline character (i.e. `\n`)
5. The FASTA filename must have a extension `.fa` (not `.fasta`)

Bad Example 1

```
$ cat egfp.fa
>EGFP
AGCAAGGGCGAGGAGCTGTTCACCGGGGTG
GTGCCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGC
GAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGTTCATCTGCACCACCGGC
AAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTC
AGCCGCTACCCCGACCACATGAAGCAGCAC
GACTTCTTCAAGTCCGCCATGCCCGAAGGC
TACGTCCAGGAGCGCACCATCTTCTTCAAG
GACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAG
GAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACTACAACAGCCACAACGTCTAT
ATCATGGCCGACAAGCAGAAGAACGGCATC
AAGGTGAACTTCAAGATCCGCCACAACATC
GAGGACGGCAGCGTGCAGCTCGCCGACCAC
TACCAGCAGAACACCCCCATCGGCGACGGC
CCCGTGCTGCTGCCCGACAACCACTACCTG
AGCACCCAGTCCGCCCTGAGCAAAGACCCC
AACGAGAAGCGCGATCACATGGTCCTGCTG
GAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAG
```

- Why: The whole sequence must not be splitted into multiple lines.
- How to correct: The entire sequence must be in one line as shown below:

```
>EGFP
AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG
```

Bad Example 2

```
$ cat mCherry.fa
>mCherry
ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAaTTtATGCGCTTCAAaGTtCACATGGAGGGCTCCGTGAACGGCG
```

- Why: The sequence line does not end with a newline character (`\n`). Your bash prompt (i.e. `$`) is displayed at the end of the sequence when you run the `cat` command to display the contents of the file.
- How to correct: Add a new line character (`\n`) at the end of the sequence as shown below. Your bash prompt must show up at the next line when you run the `cat` command:

```
$ cat mCherry.fa
>mCherry
ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAaTTtATGCGCTTCAAaGTtCACATGGAGGGCTCCGTGAACGGCC

$
```

Output Verification

Once the custom genome/index have been built, make sure that you find your transgenes in the following output files:

- `annotations.gtf`
- `geneInfo.tab`
- `chrNameLength.txt`