NAME

```
cutter.pl
```

SYNOPSIS

% perl cutter.pl <Lmer> <#bp per chunk> <#bp overlap between chunks > <last chunk must be this % chunk size> e.g. % perl cutter.pl 4 10000 3000 0.5

means chunks of 10000bp, each subsequent chunk will start 3000bp downstream, and ensure that the last chuck is at least 50% of the present chunk size; this means the last chunk(file) is in the size range of: [50% of chunk size .. 1.5x chunk size)

DESCRIPTION

SUMMARY

```
 Cutter.pl ("Script #1") is the second of a suite of scripts designed to
assist in the analysis of DNA.  This particular script breaks a large DNA
sequece down into several smaller chunks of user-determined size.  The
script then shifts downstream by a user-determined amount (note that we
recommend chunk size and shift size should be the same) and repeats the
process until it reaches the end of the sequence.  If the last chunk is
large enough in proportion to the previous chunk (again, this is determined
 by the user) it will be treated as its own separate chunk; otherwise it
will be added on to the previous chunk.  This process can be repeated
multiple times for several files as needed.

 The "chunks" produced by this script can be passed to motifCounts.pl
("Script #2") for analysis.
```

INPUT

```
 Run Cutter.pl using FNA files extracted from your database using the
script extract_ALL_chr.pl.
```

OUTPUT

```
 Cutter.pl produces several FNA files containing discrete "chunks" of DNA
for use with  Script #2 (motifCounts.pl).  These smaller chunks are much
more manageable (and therefore more preferable) than using entire genomes
at a time.
```

AUTHORS

```
 Christina Nelson
 Amos C. Jones
 Mark LeBlanc
```

MODIFICATION HISTORY

June 1, 2010 (nkf) --

```
 Fixed some potential problems with regex across platforms.
```

May 31, 2010 (nkf) --

```
 Improved the internal documentation.  Wrote some of the readme
documentation.
 Fixed a bug that caused "tailing" chunks to get deleted instead of added
to the end of the previous chunk.
 Modified file I/O code to implement cross-platform functionality.
```

Jan 08, 2009 (mdl) --

```
added separate output directory for each chromosome
for example:  in ~split_texts/data_all_chr/
Agrobacterium_tumefaciens_C58_Cereon__chr01/
Agrobacterium_tumefaciens_C58_Cereon__chr02/
```

Nov 18, 2008 (mdl) --

```
changed to bp rather than #words
```

October 01, 2008 (cn) --

```
reworked for DNA
```

June 17, 2008 (acj) --

```
changed output file names to include genre and manuscript and text numbers

reworked algorithm to work work with chunk sizes and shift sizes that are
not multiples of each other
```

June 16, 2008 (acj) --

```
finished commenting and wrote pod

fixed the last chunk problem
```

June 12, 2008 (acj) --

```
wrote and debugged script (for anglo-saxon files =:o  )
```

COPYRIGHT INFORMATION