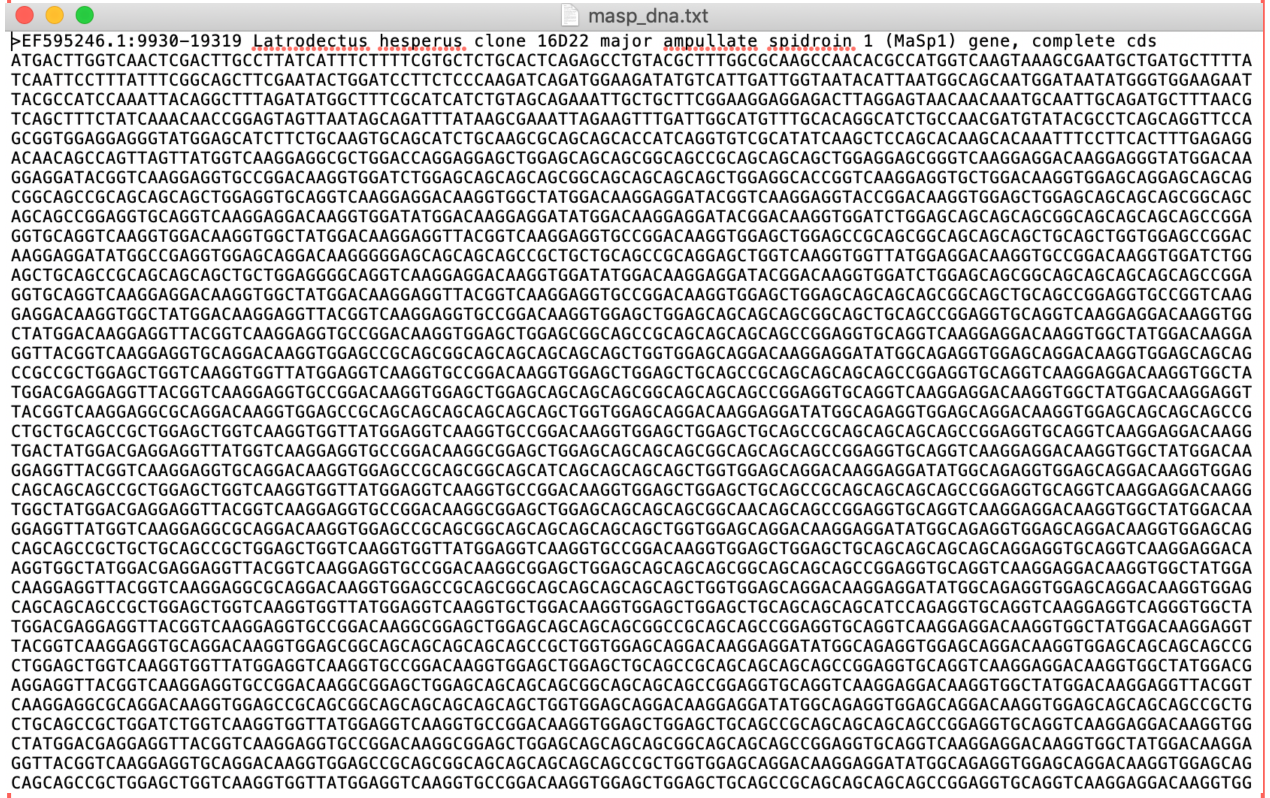
Project one

Fasta Maker

Fasta



```
>EF595246.1:9930-19319 Latrodectus hesperus clone 16D22 major ampullate spidroin 1 (MaSp1) gene, complete cds
ATGACTTGGTCAACTCGACTTGCCTTATCATTTCTTTTCGTGCTCTGCACTCAGAGCCTGTACGCTTTGGCGCAAGCCAACACGCCATGGTCAAGTAAAGCGAATGCTGATGCTTTTA
TCAATTCCTTTATTTCGGCAGCTTCGAATACTGGATCCTTCTCCCAAGATCAGATGGAAGATATGTCATTGATTGGTAATACATTAATGGCAGCAATGGATAATATGGGTGGAAGAAT
TACGCCATCCAAATTACAGGCTTTAGATATGGCTTTCGCATCATCTGTAGCAGAAATTGCTGCTTCGGAAGGAGGAGACTTAGGAGTAACAACAAATGCAATTGCAGATGCTTTAACG
TCAGCTTTCTATCAAACAACCGGAGTAGTTAATAGCAGATTTATAAGCGAAATTAGAAGTTTGATTGGCATGTTTGCACAGGCATCTGCCAACGATGTATACGCCTCAGCAGGTTCCA
GCGGTGGAGGAGGGTATGGAGCATCTTCTGCAAGTGCAGCATCTGCAAGCGCAGCAGCACCATCAGGTGTCGCATATCAAGCTCCAGCACAAGCACAAATTTCCTTCACTTTGAGAGG
ACAACAGCCAGTTAGTTATGGTCAAGGAGGCGCTGGACCAGGAGGAGCTGGAGCAGCAGCGGCAGCCGCAGCAGCAGCTGGAGGAGCGGGTCAAGGAGGACAAGGAGGGTATGGACAA
GGAGGATACGGTCAAGGAGGTGCCGGACAAGGTGGATCTGGAGCAGCAGCAGCGGCAGCAGCAGCAGCTGGAGGCACCGGTCAAGGAGGTGCTGGACAAGGTGGAGCAGGAGCAGCAG
CGGCAGCCGCAGCAGCAGCTGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAAGGAGGATACGGTCAAGGAGGTACCGGACAAGGTGGAGCTGGAGCAGCAGCAGCGGCAGC
AGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGATATGGACAAGGAGGATATGGACAAGGAGGATACGGACAAGGTGGATCTGGAGCAGCAGCAGCGGCAGCAGCAGCAGCCGGA
GGTGCAGGTCAAGGTGGACAAGGTGGCTATGGACAAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGTGGAGCTGGAGCCGCAGCGGCAGCAGCAGCTGCAGCTGGTGGAGCCGGAC
AAGGAGGATATGGCCGAGGTGGAGCAGGACAAGGGGGAGCAGCAGCAGCCGCTGCTGCAGCCGCAGGAGCTGGTCAAGGTGGTTATGGAGGACAAGGTGCCGGACAAGGTGGATCTGG
AGCTGCAGCCGCAGCAGCAGCTGCTGGAGGGGCAGGTCAAGGAGGACAAGGTGGATATGGACAAGGAGGATACGGACAAGGTGGATCTGGAGCAGCGGCAGCAGCAGCAGCAGCCGGA
GGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGTGGAGCTGGAGCAGCAGCAGCGGCAGCTGCAGCCGGAGGTGCCGGTCAAG
GAGGACAAGGTGGCTATGGACAAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGTGGAGCTGGAGCAGCAGCAGCGGCAGCTGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGG
CTATGGACAAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGTGGAGCTGGAGCGGCAGCCGCAGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAAGGA
GGTTACGGTCAAGGAGGTGCAGGACAAGGTGGAGCCGCAGCGGCAGCAGCAGCAGCAGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAGCAGCAG
CCGCCGCTGGAGCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCCGGACAAGGTGGAGCTGGAGCTGCAGCCGCAGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTA
TGGACGAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGTGGAGCTGGAGCAGCAGCAGCGGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAAGGAGGT
TACGGTCAAGGAGGCGCAGGACAAGGTGGAGCCGCAGCAGCAGCAGCAGCAGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAGCAGCAGCAGCCG
CTGCTGCAGCCGCTGGAGCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCCGGACAAGGTGGAGCTGGAGCTGCAGCCGCAGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGG
TGACTATGGACGAGGAGGTTATGGTCAAGGAGGTGCCGGACAAGGCGGAGCTGGAGCAGCAGCAGCGGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAA
GGAGGTTACGGTCAAGGAGGTGCAGGACAAGGTGGAGCCGCAGCGGCAGCATCAGCAGCAGCAGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAG
CAGCAGCAGCCGCTGGAGCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCCGGACAAGGTGGAGCTGGAGCTGCAGCCGCAGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGG
TGGCTATGGACGAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGCGGAGCTGGAGCAGCAGCAGCGGCAACAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAA
GGAGGTTATGGTCAAGGAGGCGCAGGACAAGGTGGAGCCGCAGCGGCAGCAGCAGCAGCAGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAGCAG
CAGCAGCCGCTGCTGCAGCCGCTGGAGCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCCGGACAAGGTGGAGCTGGAGCTGCAGCAGCAGCAGCAGGAGGTGCAGGTCAAGGAGGACA
AGGTGGCTATGGACGAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGCGGAGCTGGAGCAGCAGCAGCGGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGA
CAAGGAGGTTACGGTCAAGGAGGCGCAGGACAAGGTGGAGCCGCAGCGGCAGCAGCAGCAGCAGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAG
CAGCAGCAGCCGCTGGAGCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCTGGACAAGGTGGAGCTGGAGCTGCAGCAGCAGCATCCAGAGGTGCAGGTCAAGGAGGTCAGGGTGGCTA
TGGACGAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGCGGAGCTGGAGCAGCAGCAGCGGCCGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAAGGAGGT
TACGGTCAAGGAGGTGCAGGACAAGGTGGAGCGGCAGCAGCAGCAGCAGCCGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAGCAGCAGCAGCCG
CTGGAGCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCCGGACAAGGTGGAGCTGGAGCTGCAGCCGCAGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACG
AGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGCGGAGCTGGAGCAGCAGCAGCGGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAAGGAGGTTACGGT
CAAGGAGGCGCAGGACAAGGTGGAGCCGCAGCGGCAGCAGCAGCAGCAGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAGCAGCAGCAGCCGCTG
CTGCAGCCGCTGGATCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCCGGACAAGGTGGAGCTGGAGCTGCAGCCGCAGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGG
CTATGGACGAGGAGGTTACGGTCAAGGAGGTGCCGGACAAGGCGGAGCTGGAGCAGCAGCAGCGGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGGCTATGGACAAGGA
GGTTACGGTCAAGGAGGTGCAGGACAAGGTGGAGCCGCAGCGGCAGCAGCAGCAGCAGCCGCTGGTGGAGCAGGACAAGGAGGATATGGCAGAGGTGGAGCAGGACAAGGTGGAGCAG
CAGCAGCCGCTGGAGCTGGTCAAGGTGGTTATGGAGGTCAAGGTGCCGGACAAGGTGGAGCTGGAGCTGCAGCCGCAGCAGCAGCAGCCGGAGGTGCAGGTCAAGGAGGACAAGGTGG
```

- To be a fasta file :
  - Need header ">Gene info" full first line
  - Followed by DNA sequence of gene
  - Can have several genes within single fasta file
  - File is known by suffix ".fasta"

Project goals

- Create two functions:
  - 1. Creates random DNA sequence that replicates spider silk sequence structure.
    - Sequence must be 300 bp long TOTAL.
    - Sequence must contain 2 large repeat units that are composed of 2 small repeat units (consistently).
    - Both units must be generated at random (cannot hard code units).
- EX:
  - AAATTTGGGATATATAAATTTGGGATATAT

Cont.

- 2. create function that writes output from first function to a file.
  - File must have all properties that make a true fasta file.
  - File must be created as name given from user (cannot hard code).
  - File must contain full sequence.

Programs you *can* use

- FileIO:
  - To create and write variables (repeat units) to fasta file
  - In addition to try block and “with open”
- ArgParse:
  - To get user information (ie file name, header name, etc)
- Random:
  - Generate random strings of small repeat unit to stitch together
- +:
  - adding strings together