

## Script to filter a fasta file by minimum and maximum size using the awk command.

The following command filters a fasta file by minimum and maximum sequence size using the awk command and puts the sequences that pass the filter in a new file. It runs from the Linux Ubuntu command line. In this case, the minimum size was set to 300bp and the maximum size is 900bp but these values can be customized by changing them in the first line. The input file is called 1all.fasta while the output file is called 1all\_szfilt.fasta. These can be changed by editing the final line.

```
-----  
awk -v MIN_SIZE=300 -v MAX_SIZE=900 '  
/^>/ {  
  if (seqlen && seqlen >= MIN_SIZE && seqlen <= MAX_SIZE)  
    print header "\n" sequence  
    header=$0; seqlen=0; sequence=""  
  }  
/^[\^>]/ {  
  seqlen += length($0); sequence = sequence $0  
  }  
END {  
  if (seqlen >= MIN_SIZE && seqlen <= MAX_SIZE)  
    print header "\n" sequence  
}' 1all.fasta > 1all_szfilt.fasta  
-----
```