

# Assignment 6

For this assignment you will write a java program that will count the number of 3-mers in a DNA sequences (which is a string of the letter A,C,G,T). A 3-mers is a substring of the sequence for example all the 3-mer for the sequence ACTCAGG are ACT,CTC,TCA,AGG. Your program should take in a fasta file which is a file that contains DNA sequences it is formatted like this.

```
>species name
Sequence
>species name
Sequence
.
.
.
```

With the fasta file your program should count the number of all the 3-mers in the file. And write those counts but in a fun way to a file called output. See example below of output and the file that is made.

```
What file would you like to use?
Fasta.fa
Processing Compsosisymbrium reconditcladus
Processing Pseudogranum aridacinus
Processing Piosisymbrium myriacarpus
Processing Punicparilla aphelopomus
Processing Anemosisymbrium aridradox
Processing Redolsisymbrium myriasiliqua
Processing Streptophytum xerocaulis
Processing Gonimogranum mutablastus
Processing Fluviavitis poliocaulis
Processing Humescrhops euthyspora
```

Where each X is 10 and the remainder.

```
Output.txt
AAA : XXXXXXXXXXXXX+9
CAA : XXXXXXXXXXXXX+5
GAA : XXXXXXXXXXXXX+6
TAA : XXXXXXXXXXXXX+3
ACA : XXXXXXXXXXXX+5
CCA : XXXXXXXXXXXX+0
```

GCA : XXXXXXXXXXXXX+3  
TCA : XXXXXXXXXXXXXXX+6  
AGA : XXXXXXXXXXXXX+7  
CGA : XXXXXXXXXXXXX+5  
GGA : XXXXXXXXXXX+2  
TGA : XXXXXXXXXXXXX+0  
ATA : XXXXXXXXXXXXX+6  
CTA : XXXXXXXXXXXXX+7  
GTA : XXXXXXXXXXXXX+5  
TTA : XXXXXXXXXXXXXXX+0  
AAC : XXXXXXXXXXXXX+0  
CAC : XXXXXXXXXXXXXXX+3  
GAC : XXXXXXXXXXXXX+5  
TAC : XXXXXXXXXXXXX+1  
ACC : XXXXXXXXXXXXX+9  
CCC : XXXXXXXXXXXXX+6  
GCC : XXXXXXXXXXXXX+1  
TCC : XXXXXXXXXXXXX+6  
AGC : XXXXXXXXXXXXX+9  
CGC : XXXXXXXXXXXXX+0  
GGC : XXXXXXXXXXXXX+6  
TGC : XXXXXXXXXXXXX+2  
ATC : XXXXXXXXXXXXX+7  
CTC : XXXXXXXXXXXXXXX+3  
GTC : XXXXXXXXXXXXXXX+3  
TTC : XXXXXXXXXXXXXXX+9  
AAG : XXXXXXXXXXXXXXX+5  
CAG : XXXXXXXXXXXXX+5  
GAG : XXXXXXXXXXXXX+9  
TAG : XXXXXXXXXXXXX+5  
ACG : XXXXXXXXXXXXX+8  
CCG : XXXXXXXXXXXXX+6  
GCG : XXXXXXXXXXXXX+4  
TCG : XXXXXXXXXXXXXXX+7  
AGG : XXXXXXXXXXXXX+0  
CGG : XXXXXXXXXXXXX+3  
GGG : XXXXXXXXXXXXX+5  
TGG : XXXXXXXXXXXXX+4  
ATG : XXXXXXXXXXXXXXX+6  
CTG : XXXXXXXXXXXXX+9  
GTG : XXXXXXXXXXXXX+1  
TTG : XXXXXXXXXXXXX+1  
AAT : XXXXXXXXXXXXXXX+8  
CAT : XXXXXXXXXXXXX+9  
GAT : XXXXXXXXXXXXX+5  
TAT : XXXXXXXXXXXXX+9  
ACT : XXXXXXXXXXXXX+7  
CCT : XXXXXXXXXXXXX+9  
GCT : XXXXXXXXXXXXX+1  
TCT : XXXXXXXXXXXXX+3  
AGT : XXXXXXXXXXXXXXX+8  
CGT : XXXXXXXXXXXXX+8  
GGT : XXXXXXXXXXXXX+9  
TGT : XXXXXXXXXXXXX+1  
ATT : XXXXXXXXXXXXX+3

CTT : XXXXXXXXXXXXX+0  
GTT : XXXXXXXXXXXXX+7  
TTT : XXXXXXXXXXXXX+2

## Submissions

- Check that all of your programs work the way they are intended to work by running them with different inputs
- Submit to D2L a zip/archive file containing the .java files you have written and all the files needed to run the program this means the .fa files too
- If you do not zip your file, you will receive **ZERO** (<http://www.wikihow.com/Zip-Files-Together>)

.