# QFilter

## *Introduction*

The FASTQ (.fastq) file is a common bioinformatics format displaying, among other things, the sequence quality for individiual nucleotides. These quality scores are coded as a series of ASCII characters, each denoting a quality value.

## *The Program*

QFilter is a python-based program will allow you to input a FASTQ file, along with a given a quality threshold value. It either creates a file of the sequence ID's that meet the given threshhold value, or print these ID's directly to the terminal. You can also set the maximum number of sequnce ID's to print.

**For FASTQ files with a large number of sequences, use the output file option.**

## *Usage*

The inputs and options for this program are:

- An input FASTQ file
- `-q` or `-qual` is the given quality threshold as an floating point (i.e. 80% = .80) [**Required**]
- `-o` or `-outfile` is the output file name. If not specified, it will print to the terminal. [Optional]
- `-n` or `-num` is the maximum number of reads to print. If left blank, it will find all sequences that meet the quality threshold. [Optional]

For example, to find the sequences within a file "seqreads.fastq" that have a quality of over 90%, the syntax is:

```
./qfilter.py seqreads.fastq -q .90
```

This will return:

```
"Done. 4 sequences had a quality of 90% or greater."  ID: diad_3479283749 = 94.3%
ID: diad_3049384747 = 98.1% ID: diad_172398399 = 90.3% ID: diad_120983753 = 92.9%
```

When using an output file:

```
./qfilter.py seqreads.fastq -q .90 -o goodreads.txt
```

This will return:

```
"Done. 4 sequences had a quality of 90% or greater written to 'goodreads.txt'."
```

If only a certain number of quality reads are needed (in this example, 20):

```
./qfilter.py seqreads.fastq -q .90 -o goodreads.txt -n 20
```

This will return:

> *"Done. 20 sequences had a quality of 90% or greater written to 'goodreads.txt'."*