

kmercount

Writes in an output file the number of times a given k-mer is found in each DNA sequence from a FASTA file, using regular expressions, considering overlapping and including the reverse complement. It is able to perform the processing of a big file by its automatic splitting in smaller parts whenever is necessary.

Syntax:

```
kmercount -f FILE -k KMER -o FILE [-r -d -g FILE -s SPLIT_SIZE]
```

Options:

- **-f, --fasta FILE**
FILE is the filename of a FASTA file containing the sequences to be analyzed.
- **-k, --kmer KMER**
KMER is a regular expressions surrounded by quotes that represents the k-mer to be searched into the sequences. It must use IUPAC symbols for DNA sequences.
- **-o, --output FILE**
FILE is the filename of an output file with the results of the counting of the k-mer in each sequence. Each line in this file follows this pattern:
SEQUENCE ID: NUMBER OF OCCURRENCES
- **-r, --reverse_complement**
If this option is given, then the reverse complement of each sequence will be also inspected.
- **-d, --disable_overlapping**
If this option is given, then the searching of the k-mer will be done without considering overlapping of subsequences.
- **-s, --split_size SPLIT_SIZE**
Whenever the file is so big to be processed due to memory limits, is better to split it into fragments of size SPLIT_SIZE (gigabytes). The program does this task automatically.
- **-g, --debug FILE**
FILE is the filename where the debugging output will be write. There is no debugging operation if this option is not given.