seqkit

如果要处理序列，应该首先想到此软件
下载地址：

1、amplicon

2、bam

3、common

4、concat

5、convert

6、duplicate

7、faidx

8、fish

9、fq2fa

10、fx2tab

11、genautocomp

12、grep

13、head

14、help

15、locate

16、mutate

17、range

18、rename

19、replace

```
./software/seqkit.exe replace -h
```

replace name/sequence by regular expression.

Note that the replacement supports capture variables.
e.g. $1 represents the text of the first submatch.
ATTENTION: use SINGLE quote NOT double quotes in *nix OS.

Examples: Adding space to all bases.

    seqkit replace -p "(.)" -r '$1 ' -s

Or use the \ escape character.

    seqkit replace -p "(.)" -r "\$1 " -s

more on: http://bioinf.shenwei.me/seqkit/usage/#replace

Special replacement symbols (only for replacing name not sequence):

    {nr}   Record number, starting from 1
    {kv}   Corresponding value of the key (captured variable $n) by key-value file,
            n can be specified by flag -I (--key-capt-idx) (default: 1)

Usage:
  seqkit replace [flags]

Flags:
  -s, --by-seq   replace seq
  -h, --help   help for replace
  -i, --ignore-case   ignore case
  -K, --keep-key   keep the key as value when no value found for the key (only for sequence name)
  -I, --key-capt-idx int   capture variable index of key (1-based) (default 1)
  -m, --key-miss-repl string replacement for key with no corresponding value
  -k, --kv-file string   tab-delimited key-value file for replacing key with value when using "{kv}" in -r (--replacement)
      --nr-width int   minimum width for {nr} in flag -r/--replacement. e.g., formating "1" to "001" by --nr-width 3 (de
  -p, --pattern string   search regular expression
  -r, --replacement string   replacement. supporting capture variables. e.g. $1 represents the text of the first submatch. AT

Global Flags:
      --alphabet-guess-seq-length int   length of sequence prefix of the first FASTA record based on which seqkit guesses the se
      --id-ncbi   FASTA head is NCBI-style, e.g. >gi|110645304|ref|NC_002516.2| Pseud...
      --id-regexp string   regular expression for parsing ID (default "^(\\S+)\\s?")
  -w, --line-width int   line width when outputing FASTA format (0 for no wrap) (default 60)
  -o, --out-file string   out file ("-" for stdout, suffix .gz for gzipped out) (default "-")
      --quiet   be quiet and do not show extra information
  -t, --seq-type string   sequence type (dna|rna|protein|unlimit|auto) (for auto, it automatically detect by the f
  -j, --threads int   number of CPUs. (default value: 1 for single-CPU PC, 2 for others) (default 2)

20、restart

21、rmdup

22、sample

23、sana

24、seq

25、shuffle

26、sliding

27、sort

28、split

29、split2

30、stats

31、subseq

32、tab2fx

33、translate

34、version

35、watch