

FASTA Sequence format

White Space (Space or Tab)

> Sequence_Name_1 Sequence Annotation

MDSKGSSQKGSRLLLLLLVVSNLLLCQGVVSTPVCPNGPGNCQVSLRDLFDRAVMVSHYIHDLS
EMFNEFDKRYAQGKGFITMALNSCHTSSLPTPEDKEQAQQTHHEVLMSLILGLLRSWNPPLYHL
VTEVRGMKGAPDAILSR AIEIEEENKRLLEGMEMIFGQVIPGAKETEPYPVWSGLPSLQTKDED
ARYSAFYNLLHCLRRDSSKIDTYLKLLNCRIIYNNNC

> Sequence_Name_2 Sequence Annotation

ADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTID
FPEFLTMMARKMKD TDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREA
DIDGDGQVNYEEFVQMMTAK

> Sequence_Name_3 Sequence Annotation

LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPY
IGTNLVEWIWGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIP
FHPYYTIKDFLGLLILILLLLLLLLALLSPDMLGDPDNHMPADPLNTPLHIKPEWYFLFAYAILRS
VPNKLGGVLALFLSIVILYGLMPFLHTSKHRSMMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYP
YTIIGQMASILYFSIILAFLPIAGXIENY

FASTA Sequence format

> Sequence_Name_1 Sequence Annotation

MDSKGSSQKGSRLLLLVSNNLLCQGVVSTPVCPNGPGNCQVSLRDLFDRAVMVSHYIHDLS
EMFNEFDKRYAQGKGFI TMAINSCHTSSLPTPEDKEQAQQTHHEVLMSLILGLLRSWNP
VTEVRGMKGAPDAILSRATIEIEEENKRLLLEGMEMIFGQVIPGAKETEPYPVWSGLPSLQTKDE
ARYSAFYNNLLHCLRRDSSKIDTYLKLLNCRIIYNNNC

> Sequence_Name_2 Sequence Annotation

ADQLTEEQIAEFKEAFSLFDKKGDTETTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTID
FPEFLTMMARKMKD TDSEEEI REAFRVFDKBNGYISAAELRHVMTNLGEKLTDEEVDEMIREA
DIDGDGQVNYEEFVQMMTAK

> Sequence_Name_3 Sequence Annotation

LCLYTHIGRNIYYGSYLYSETWNTGIMLLITMATAFMGYVLPWGQMFWGATVITNLFSAIPY
IGTNLVEWIWGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIP
FHPYYTIKDFLGLLILILLLLLLLLALLSPDMLGDPDNHMPADPLNTEFLHIKPEWYFLFAYAILRS
VPNKLGGVLALFLSIVILYGLMPFLHTSKHRSMLRPLSQALFWTLTMDLLTLTWIGSQPVEYP
YTIIGQMASILYFSIILAFLPIAGXIENY

FASTQ Sequence format

```
@ Sequence_Name_1 Sequence Annotation ... ..  
MDSKGSSQKGSRLLLLLLVVSNLLLCQGVVSTPVCPNGPGNCQVSLRDLFDRAVMVSHYIHDLS  
EMFNEFDKRYAQGKGFITMALNSCHTSSLPTPEDKEQAQQTHHEVLMSLILGLLRSWNP  
VTEVRGMKGAPDAILSR AIEIEEENKRLLEGMEMIFGQVIPGAKETEPYPVWSGLPSLQTK  
DEDA  
ARYSAFYNLLHCLRRDSSKIDTYLKLLNCRIIYNNNC  
+ MORE Sequence Annotation ... ..
```

FASTA Format is designed to store **DNA, RNA, Protein** Sequences with minimal **Annotation**.

FASTQ Format is an adaptation designed to store **Sequencing Reads** with minimal **Annotation**.

The transition from **FASTA** to **FASTQ** Format includes a couple of cosmetic “enhancements”:

The initial “>” becomes an “@”

An extra **Annotation** line beginning with a “+” is added.

The opportunity to include **Annotation** in this line is rarely used.

FASTQ Sequence format

```
@ Sequence_Name_1 Sequence Annotation ... ..  
CCGCTAGCTGGGGTATCATCAGCATGCATGGCATGAGCGTTCTTAATTCTCAGGGACTCGGAGCAGGGGCATCGAG ...  
+
```

Sequences in a **FASTA** file can be **DNA**, **RNA** or **Protein**.

Allowing **DNA** & **Protein** **Ambiguity Codes**, makes it difficult to ascertain **Sequence Type**.

Sequence Type is not specified in the **FASTA** format. Distinction is left to the software.

Each sequence in a **FASTQ** file is always the that of a single **Sequencing Read**.

Thus it can only ever be **DNA** sequence.

Currently, the length of an individual read is modest. Generally, the sequence will be entirely contained in a single line.

FASTQ Sequence format

```
@ Sequence_Name_1 Sequence Annotation ... ..  
+  
CCGCTAGCTGGGGTATCATCAGCATGCATGGCATGAGCGTTCTTAATTCTCAGGGACTCGGAGCAGGGGCATCGAG ...  
! ' ' * ( ( ( * * * + ) ) % % % + + ) ( % % % % ) . 1 * * * - + * ' ' ) * * 5 5 C C F > > > > > C C C C C C C 6 5
```

The **Q** in **FASTQ** stands for **Quality**. A definitive purpose of the **FASTQ** Format is to record the **Base Call Quality** of each element of a **Sequencing Read**.

Base Call Qualities are recorded in a fourth line of the **FASTQ** file. Each **Base Call Quality** is recorded as a single printable character.

Currently, the length of an individual read is modest. Generally, the sequence will be entirely contained in a single line.

