

.FASTA/FASTQ Read Sequences

The Flux Simulator uses FASTA/FASTQ sequences at different points; for the (optional) input of a genomic sequence to (optionally) produce read sequences. Genomic references are expected to provide one single FASTA file per reference sequence (i.e., chromosome, scaffold, etc.), as described in the [Sequencing Section](#).

The read sequence output is a multi-FASTA file, where each fasta block contains a description line that starts with a ">" ("greater than") symbol and the following one or multiple lines containing the read sequence. If a quality/error-model is provided, the very related FASTQ file format is produced, where the ">" identifier is replaced by the "@" symbol, and a quality block is following the fasta block, which uses a "+" separator and subsequently provides the qualities of the read sequences. The description line contains the read identifier as described in the [Sequencing Section](#).

Example

FASTQ

```
@chr1:4847775-4887990W:NM_001159750:1:2668:917:1137/1
AAGAGATGAGGAAAAACCTGACCAAAGAAGCCATCAGGGAGCATCAGATGGCCAAGACTGGTGGGACCCAGACTGA
+
IEEIIIGIIIIIF<GGEHHHD4<D@147=;7*+BDBGACDGGHIIIIHHHDGGDB@@FEGGD9DGIHHHHI@BDG
@chr1:4847775-4887990W:NM_001159750:1:2668:917:1137/2
CCAATTCTTCCAAACTCAACAGAACTTCCACCGATTTCACATTCATTACATACAACAAATGTTGTCATTGGTTCA
+
G:GB78?:9>;?EGGGGHIDGDD=EBFGIIHHGGGIIIIHHIIIIHHIGEIIIIHIFCBFIHGD@@@BBEIC
```

FASTA

```
>chr1:4847775-4887990W:NM_001159750:1:2668:917:1137/1
AAGAGATGAGGAAAAACCTGACCAAAGAAGCCATCAGGGAGCATCAGATGGCCAAGACTGGTGGGACCCAGACTGA
>chr1:4847775-4887990W:NM_001159750:1:2668:917:1137/2
CCAATTCTTCCAAACTCAACAGAACTTCCACCGATTTCACATTCATTACATACAACAAATGTTGTCATTGGTTCA
```