

# Trimmomatic

Trimmomatic is one of the many software tools for clipping low quality regions as well as adapter sequences from DNA sequencing reads. It is part of the Trinity standard workflow for RNA seq assembly. However, please note that this is by no means a guarantee that Trimmomatic is the all time optimal tool for sequence read pre-processing. Trimmomatic can be run both in single end and in paired end mode. Figure 1 outlines the workflow for the paired end mode.

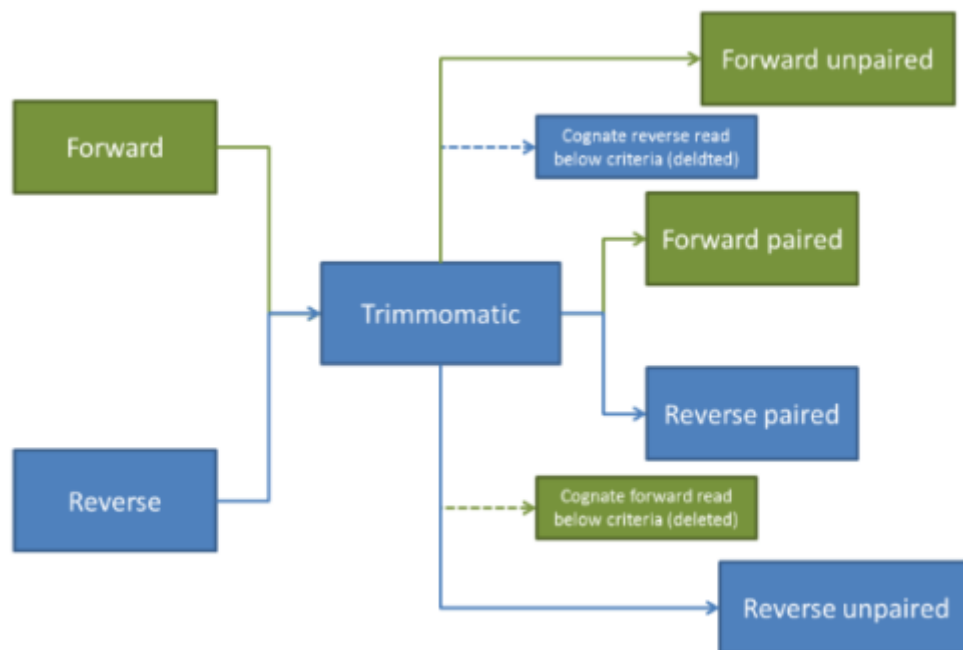


Figure 1: Flow of reads in Trimmomatic paired end mode. The software expects the two reads of a read pair to be stored in separate files. Once the trimming is completed, Trimmomatic will generate four output files, again separate files for the forward and the reverse reads. Figure source: [Trimmomatic manual V0.32](#).

## Running Trimmomatic

You can start your Trimmomatic run via a command line call where you specify the names of the input files together with the tasks and the parameter settings that tell Trimmomatic how the clipping should be performed. **Be careful, the tasks that you pass on to Trimmomatic must not be in an arbitrary order.** This is, because Trimmomatic will perform the tasks exactly in the order as it is specified in the command line call. If this is not obeyed, the trimming results might deviate in part substantially from what you had in mind.

### Example program call

```

trimmomatic PE -phred33 <Inputfile-Forward>.fq <Inputfile-Reverse>.fq
<Forward_trimm_paired>.fq.gz <Forward_trimm_unpaired>.fq.gz
<Reverse_trimm_paired>.fq.gz <Reverse_trimm_unpaired>.fq.gz
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
MINLEN:36
  
```

## The options

Below you find a number of common options. For full details please have a look at the [Trimmomatic manual](#).

- **PE**: you provided reads of a paired end library.
- **-phred33**: Tells trimmomatic to use the phred+33 encoding. If the encoding is not specified, trimmomatic will try to determine it directly from the data.
- **ILLUMINACLIP:TruSeq3-PE.fa:2:30:10**: Clip adapters from the 3 primed end of the sequence. The adapter sequences are provided in the file TruSeq3-PE.fa. See the documentation for the meaning of 2:30:10.
- **LEADING:<QualValue>**: trim bases with a quality value of <QualValue> or below from the 5' end of the read. Stop trimming as soon as a base with a higher quality is encountered.
- **TRAILING:<QualValue>**: trim bases with a quality value of <QualValue> or below from the 3' end of the read. Stop trimming as soon as a base with a higher quality is encountered.
- **SLIDINGWINDOW:<WindowSize>:<MeanQuality>**: scan the read starting at the 5' end with a sliding window of length <WindowSize>. Once the program encounters a window where the average quality value of the base calls in the window falls below the limit specified with <MeanQuality>, the read is clipped and the window plus the remainder of the read towards the 3' end is removed. Typical values are SLIDINGWINDOW:4:15.
- **MINLEN:<MinLength>**: specifies the minimum length of a read after trimming and clipping that is required to keep the read. If the read length falls below this limit, the read is discarded. Its read pair, if it exists, will then remain as an unpaired read.

From:  
<https://applbio.biologie.uni-frankfurt.de/teaching/wiki/> - Teaching

Permanent link:  
<https://applbio.biologie.uni-frankfurt.de/teaching/wiki/doku.php?id=general:computerenvironment:software:trimmomatic>

Last update: 2024/05/19 13:40

