Manual for trim_fastq2.pl

Minou Nowrousian

April 2014

Summary:

The Perl program trim_fastq.pl takes fastq files of single or paired-end sequence reads and trims low-quality bases from the 5' and 3' ends, also it removes any reads that contain at least one "N" (undefined base) after trimming. Trimming is done from the 5' and 3' ends until the first base with a quality score above a certain threshold is reached or the read gets shorter than a certain minimum length. In case of paired-end reads, pairs where both reads meet the criteria are kept in a pairs file, and in case only one read of a pair meets the criteria, this read is kept in a separate singlets file.

Usage: `trim_fastq2.pl <options> <file(s)>`

Options:

- *-a* type of analysis:
  - *s* for single reads
  - *t* for paired-end reads with pairs in two files
  - *p* for paired-end reads with pairs in one file
- *-o* offset for the quality scores (default: 33, for certain types of Illumina pipelines it can be 64, then this option must be set)
- *-l* minimum length of read that is kept (default: 40)
- *-q* minimum base quality that must be reached at the 3' endfor the trimming to stop (default: 10)
- *-r* core name for results file(s). For paired-ends, will be extended by _pairs.fastq, for single-ends by _singlets.fastq, default trim_fastq. Must contain only letters, numbers, - or _

Important: Bases and quality sequences must be only ONE line each (i.e. four lines containing the information for each read).

Output files:

trim_fastq_ singlets.fastq:

when trimming single reads: all reads that were kept

when trimming paired-end reads: all reads that meet the quality criteria but where the mate does not meet the quality criteria

trim_fastq_ pairs.fastq

only written when trimming paired-end reads, contains all read pairs that meet the quality criteria, in format *pair1_read1 pair1_read2 pair2_read1 pair2_read2* etc.

trim_fastq_output_stats.txt

output statistics, gives an overview over the number of input reads, the number of reads that were kept in the single and paired files and the number of reads of length x that were kept

Examples:

trim single reads in file reads.fastq, output reads_trimmed_singlets.fastq and reads_trimmed_output_stats.txt

```
perl trim_fastq.pl -a s reads.fastq -r reads_trimmed
```

trim single reads in file reads.fastq with offset 64, minimum quality score at 3' end 20, and minimum length of 50

```
perl trim_fastq.pl -a s -o 64 -l 50 -q 20 reads.fastq
```

trim paired-end reads with both reads in file read_pairs.fastq (reads that belong to one pair must be directly after each other, the file must contain an even number of reads, i.e. no mixture of single and paired-end reads)

```
perl trim_fastq.pl -a p reads_pairs.fastq
```

trim paired-end reads with the first read of each pair in file read1.fastq and the second read in each pair in file read2.fastq (reads must be in the same order in the files)

```
perl trim_fastq.pl -a t reads1.fastq read2.fastq
```

Licensing information