

Manual for fastq_qual_stats.pl

Minou Nowrousian

May 2012

Summary:

The program fastq_qual_stats.pl takes a fastq file and calculates summary statistics for the quality scores. For each base position, the mean quality score is calculated and the quality scores are sorted in six bins from 0-9, 10-19, 20-29, 30-39, 40-49, >=50. The mean quality score, the number of bases, and percent of total bases at that position in each bin are given as output.

Usage: fastq_qual_stats.pl <option> <file>

Options: -o offset for the quality scores (default: 33, for certain types of Illumina pipelines it can be 64, then this option must be set)

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

The following information is given in the output file (named file.qualstats):

tab-delimited text file with the following columns:

- 1) base number, starting with first base of reads
- 2) total number of bases that were counted for all reads at this position
- 3) mean quality score for base at this position
- 4) no. of bases counted in bin for quality scores 0-9
- 5) % of bases at this position that are present in bin for quality scores 0-9
- 4) no. of bases counted in bin for quality scores 10-19
- 6) % of bases at this position that are present in bin for quality scores 10-19
- 7) no. of bases counted in bin for quality scores 20-29
- 8) % of bases at this position that are present in bin for quality scores 20-29
- 9) no. of bases counted in bin for quality scores 30-39
- 10) % of bases at this position that are present in bin for quality scores 30-39
- 11) no. of bases counted in bin for quality scores 40-49
- 12) % of bases at this position that are present in bin for quality scores 40-49
- 13) no. of bases counted in bin for quality scores >=50
- 14) % of bases at this position that are present in bin for quality scores >=50

Licensing information

Copyright (C) 2012 Minou Nowrousian

The program fasta_stats_N50.pl is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details (<http://www.gnu.org/licenses/>).