BGI Sequencing-only Report

2022/5/6

Table of Contents

1 Data Statistics

Raw reads produced from sequencer contain adapters, unknown or low quality bases. The statistics of raw data is shown below.

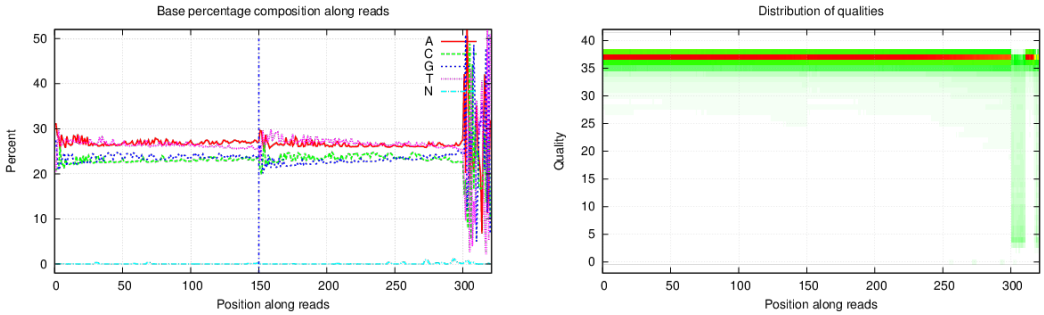
Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
A	150;170	98.27;94.96	94.98;90.16	46.59;46.90	923,919,048	147,827,047,680

Table Format:
1. Sample: The name of sample
2. Length: The Length of raw reads
3. Q20 (%): The proportion of nucleotides with quality value larger than 20
4. Q30 (%): The proportion of nucleotides with quality value larger than 30
4. GC Content(%): The proportion of bases G and C
5. Total Reads: The total number of raw reads
6. Total Bases: The total nucleotides number of raw reads

2 Data Quality Control

The distribution of base percentage and qualities along reads in data filtering are shown as following(If example.fqcheck sample has multiple lanes, only one of them will be displayed). The left picture is base percentage distribution along reads after filtering of the sample, the right picture is distribution of qualities along reads after filtering of the sample.

Quality control of sample A



3 Help Document

The original image data is transferred into sequence data via base calling, which is defined as raw data or raw reads and saved as FASTQ file. In each FASTQ file, every read is described by four lines, listed as follows:

@V300029029L1C001R0010000210/1
GCGACCCCAGGTCAGTCGGGACTACCCGCTGAAGTCGGAGGCCAAGCGGT
+
FFFCFFFFFFFFFDFEFFFFEFEF0FFFFEFFFFFFFEFFFFFECGFFFF

The first and third lines are sequences' names generated by the sequence analyzer; the second line is sequence; the fourth line is sequencing quality value, in which each letter corresponds to the base in line 2; the base's quality is equal to ASCII value of the character in line 4 minus 33 (Phred +33 quality system), e.g. the ASCII value of A is 65, then its base quality value is 32.

The relationship between DNBseq sequencer sequencing error rate and the sequencing quality value is shown in the following formula. Specifically, if the sequencing error rate is

denoted as "E", DNBseq sequencer base quality value is denoted as "sQ", the relationship is as follows:

$$sQ = -10\log_{10} E$$

Sequencing error rate	Sequencing quality value	Character of Phred +33 quality system
5%	13	.
1%	20	5
0.1%	30	?