BGI Sequencing Data Report

2023/5/15

Table of Contents

1 Project Information

Project code: F20FTSEUHT0946-01_MUSvwagR_poolB
Sample number: 36

2 Data Statistics

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

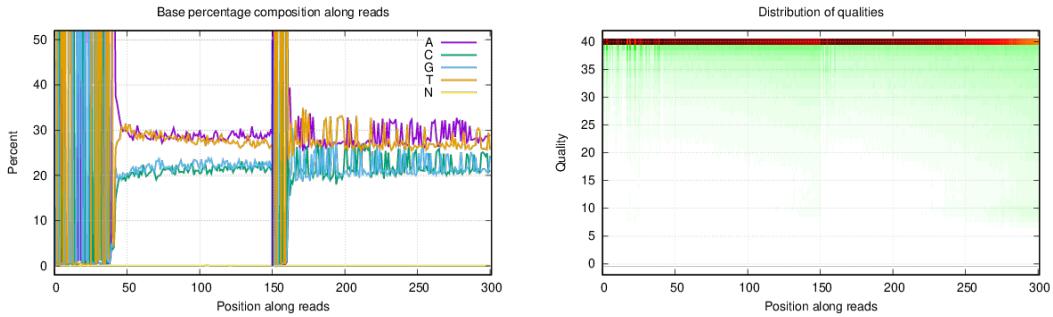
Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
10_2	150;150	93.56;92.24	83.67;77.21	51.14;50.92	756,314	226,894,200
10_2_d	150;150	93.49;92.59	83.64;77.95	50.14;49.89	1,108,873	332,661,900
1_2	150;150	99.00;75.33	97.67;55.00	42.33;48.00	2	600
7_1	150;150	97.33;95.59	92.39;88.89	50.13;49.88	177	53,100
CS-2	150;150	97.04;92.95	90.40;81.20	42.67;43.28	19,256	5,776,800
DM1N	150;150	93.28;92.05	83.73;77.70	52.15;51.65	13,373	4,011,900
KAPA_1_30	150;150	97.99;95.68	93.69;88.22	49.02;48.87	326	97,800
Lop	150;150	98.73;97.05	96.01;92.45	49.63;49.62	47,421,766	14,226,529,800
MoPh11	150;150	97.85;94.16	93.45;85.56	53.15;53.14	58,061,758	17,418,527,400
MoPh14	150;150	97.49;94.37	92.10;85.44	47.96;48.08	5,470,229	1,641,068,700
MoPh15	150;150	98.33;86.33	93.00;64.00	52.67;53.67	2	600
MoPh7	150;150	97.88;94.96	93.25;86.68	49.32;49.37	13,176,081	3,952,824,300
NCS-2	150;150	97.11;93.45	90.67;82.55	43.58;43.83	22,084	6,625,200
NP_Ph_1_1603	150;150	97.53;94.94	92.20;86.66	45.96;46.77	156,674	47,002,200
NP_Ph_1_1803	150;150	98.35;98.24	95.21;95.40	46.35;46.80	121,788	36,536,400
NP_Ph_1_DNAse_1603	150;150	97.35;94.70	91.66;85.96	45.52;46.22	192,228	57,668,400
NP_Ph_2	150;150	97.40;95.19	91.78;87.26	45.73;46.67	10,496	3,148,800
NP_Ph_DNAse_1803	150;150	98.21;97.93	94.71;94.48	46.16;46.52	430,294	129,088,200
NP_Ph_DNAse_3	150;150	97.06;93.14	90.83;82.18	46.12;47.41	49,964	14,989,200
NP_Ph_Dnase2	150;150	97.48;94.04	91.80;84.42	45.73;47.15	28,844	8,653,200
P125	150;150	98.37;96.35	94.78;90.41	48.70;48.66	14,246,808	4,274,042,400
P126	150;150	98.27;96.31	94.51;90.29	50.07;50.02	89,703,021	26,910,906,300
P131	150;150	98.25;96.40	94.39;90.53	46.96;46.93	12,951,026	3,885,307,800
P132	150;150	98.43;96.32	94.92;90.34	48.13;48.12	15,479,216	4,643,764,800
P133	150;150	98.45;96.55	95.02;90.87	48.11;48.08	11,057,432	3,317,229,600
P135	150;150	98.40;96.85	94.90;91.60	47.34;47.30	12,342,928	3,702,878,400
P136	150;150	98.53;97.01	95.22;91.95	45.91;45.92	10,715,181	3,214,554,300
P137	150;150	98.39;96.51	94.87;90.83	48.22;48.20	19,050,267	5,715,080,100
P139	150;150	98.42;96.46	94.89;90.63	47.95;47.93	15,162,288	4,548,686,400
PICO8_8_cells	150;150	94.06;92.27	87.44;83.76	36.88;36.92	44	13,200
mix10	150;150	98.36;95.84	94.75;89.17	49.74;49.71	8,167,546	2,450,263,800
mix25	150;150	98.42;96.05	94.93;89.70	49.94;49.90	7,838,618	2,351,585,400
mix50	150;150	98.31;96.06	94.57;89.66	49.43;49.40	5,774,736	1,732,420,800
mix75	150;150	98.25;96.17	94.37;89.90	48.21;48.21	9,764,587	2,929,376,100
new_CS-2_EcoRI	150;150	97.20;94.17	91.25;84.68	44.39;44.56	26,160	7,848,000
CS-2_EcoRI	150;150	97.07;94.38	90.76;85.10	43.42;43.67	25,433	7,629,900

Table Format:
1. Sample: The name of sample
2. Length: The Length of 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 read pairs
6. Total Bases: The total nucleotides number of raw reads

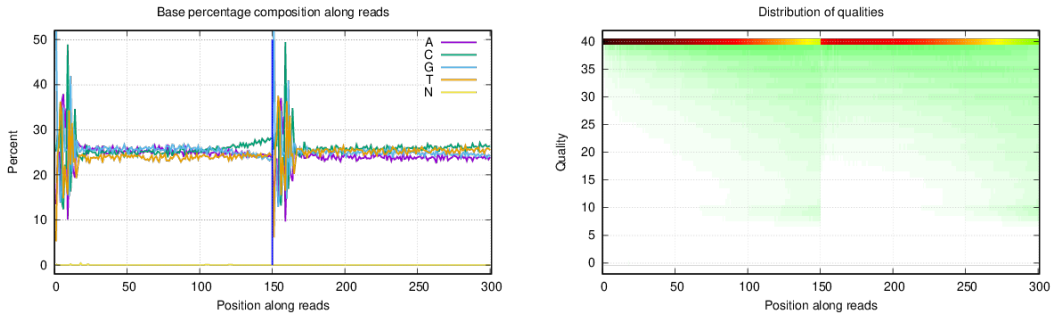
3 Data Quality Control

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

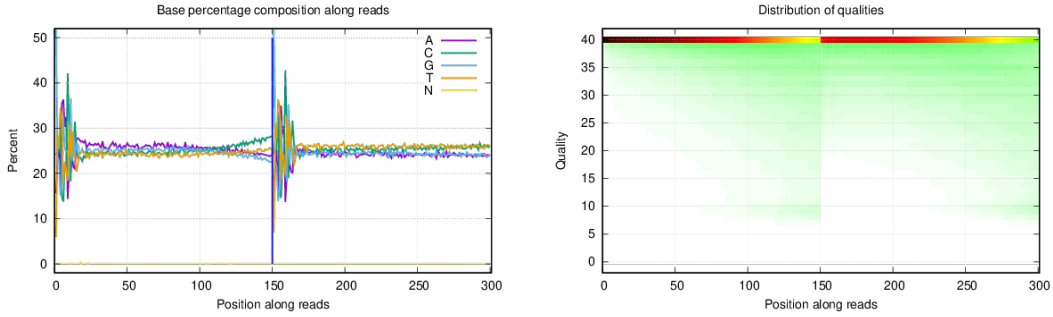
Quality control of sample CS-2_EcoRI



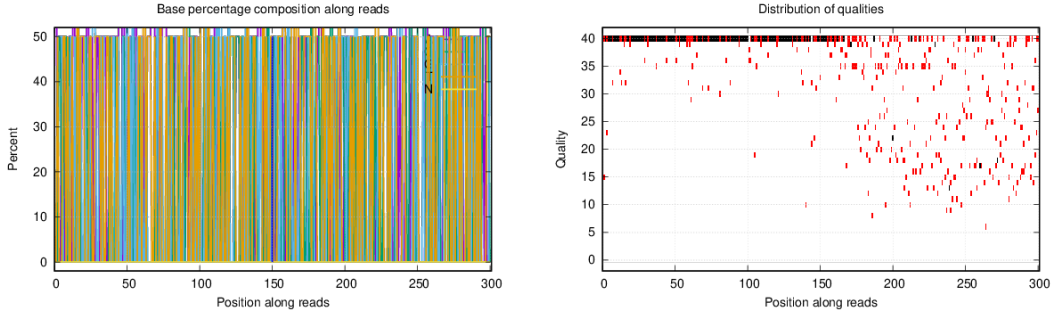
Quality control of sample 10_2



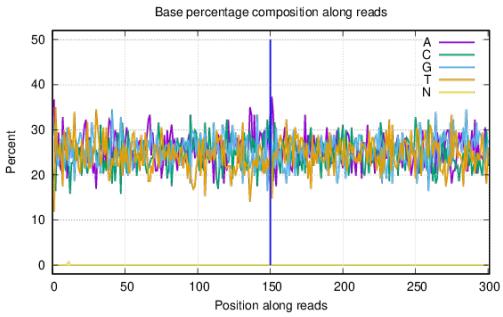
Quality control of sample 10_2_d

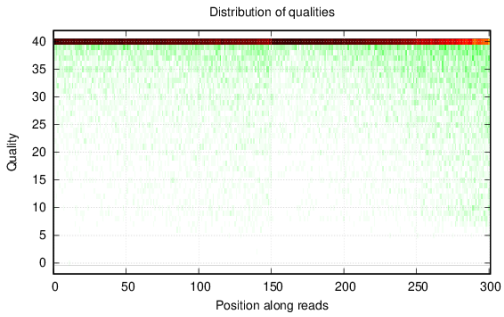


Quality control of sample 1_2

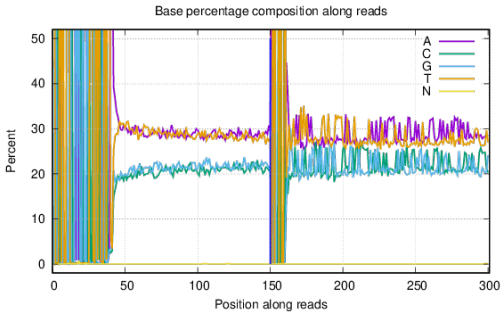


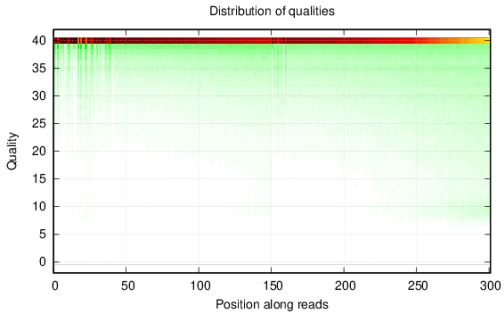
Quality control of sample 7_1



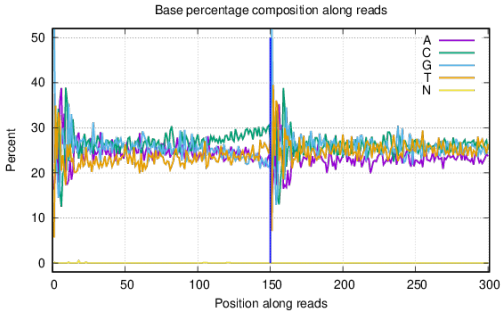


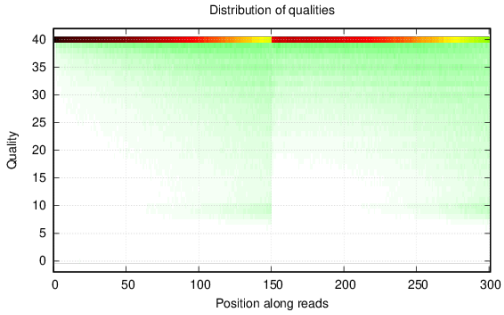
Quality control of sample CS-2



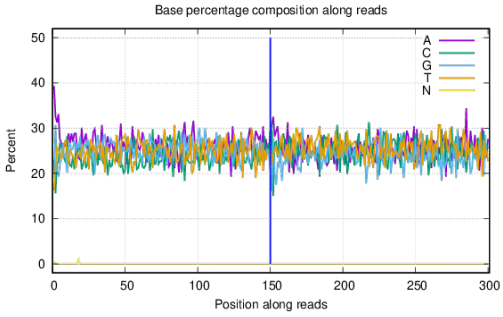


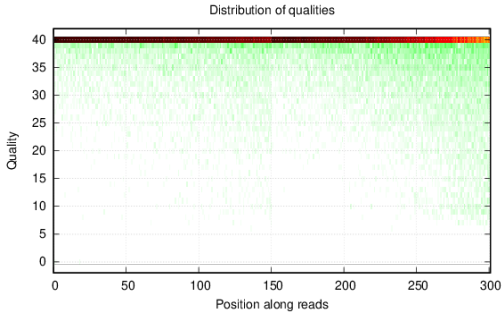
Quality control of sample DM1N



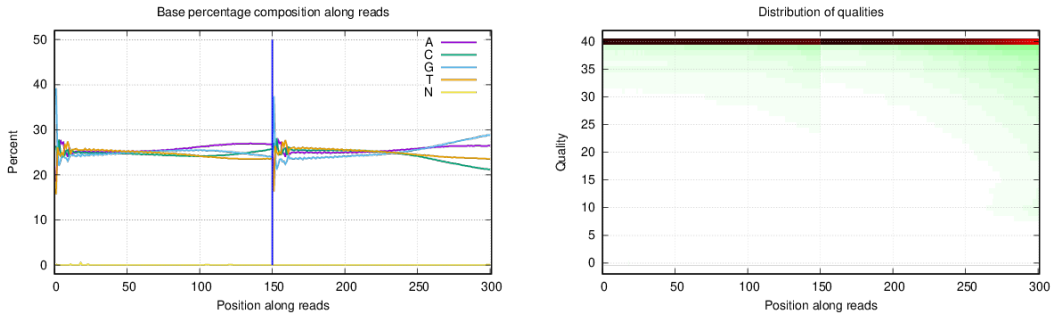


Quality control of sample KAPA_1_30

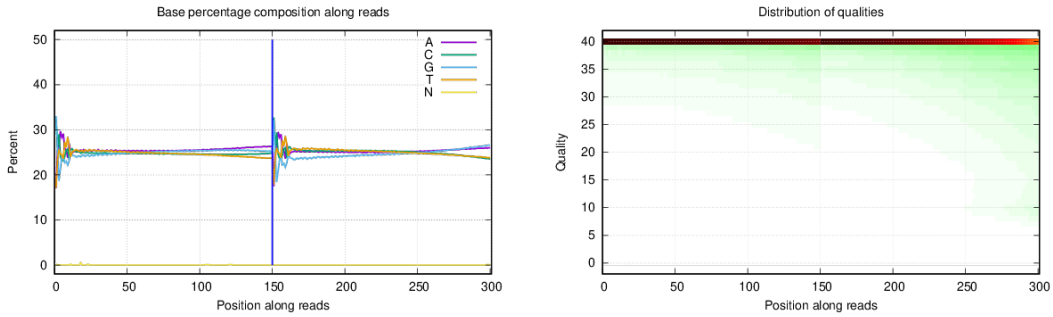




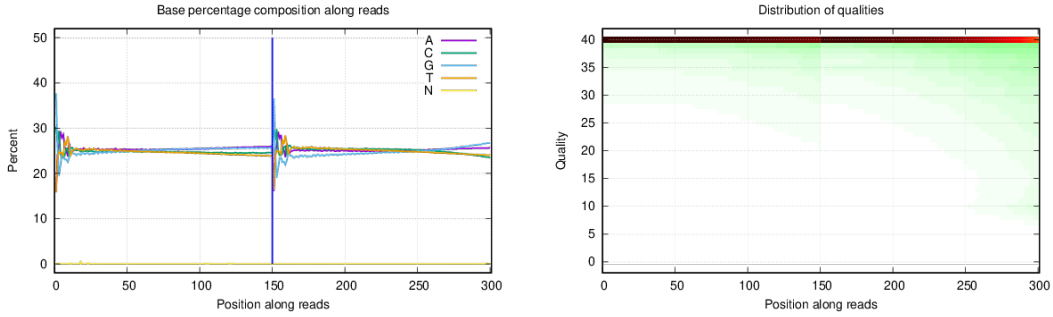
Quality control of sample Lop



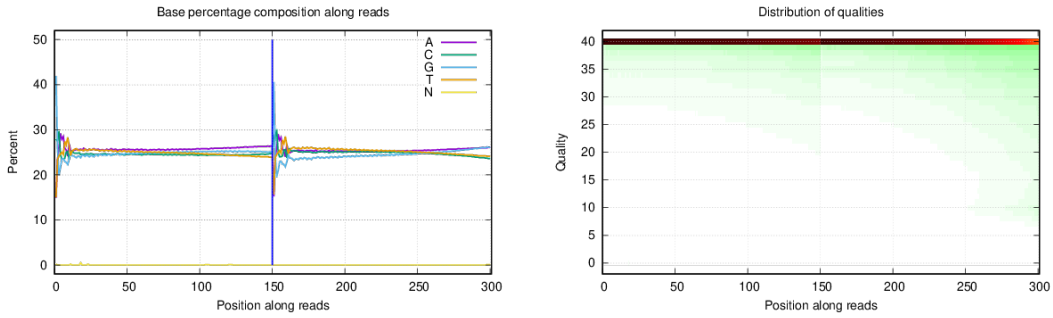
Quality control of sample mix10



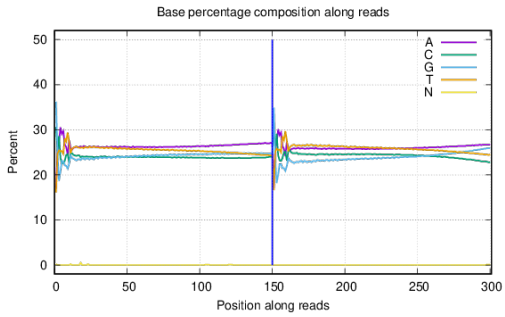
Quality control of sample mix25

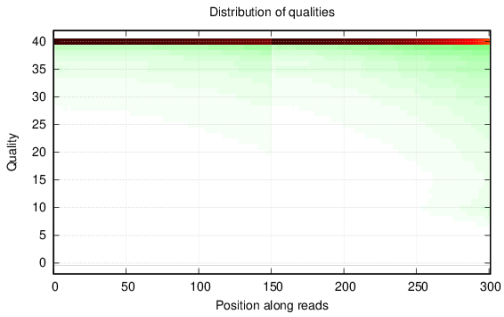


Quality control of sample mix50

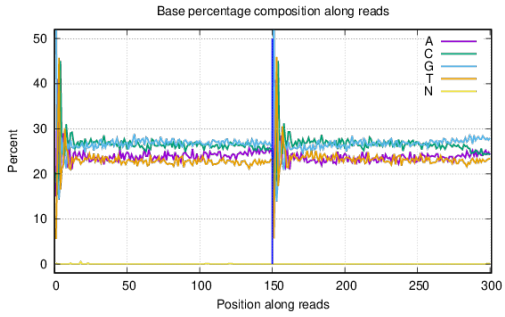


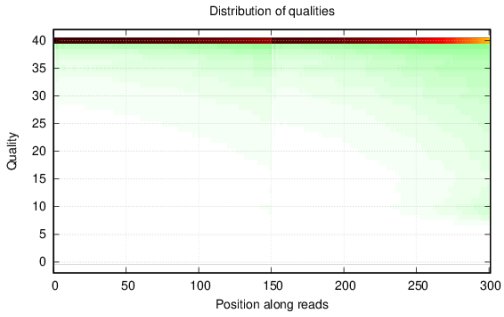
Quality control of sample mix75



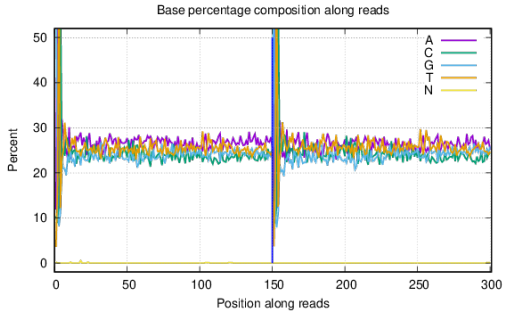


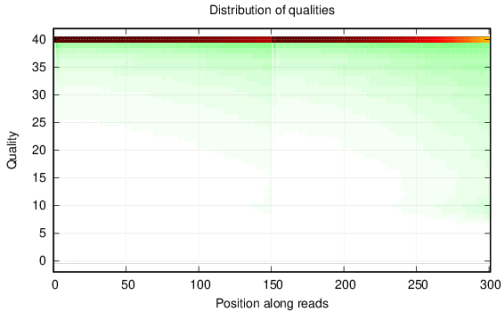
Quality control of sample MoPh11



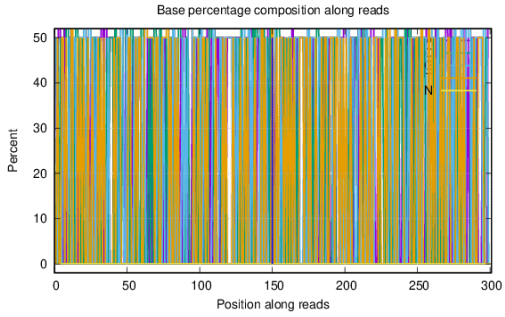


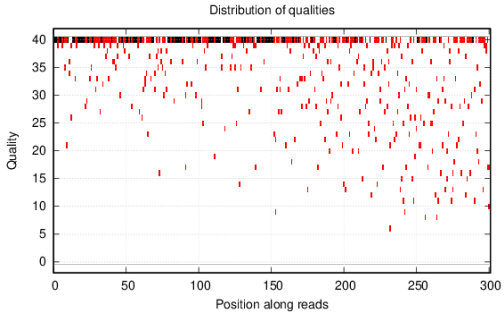
Quality control of sample MoPh14



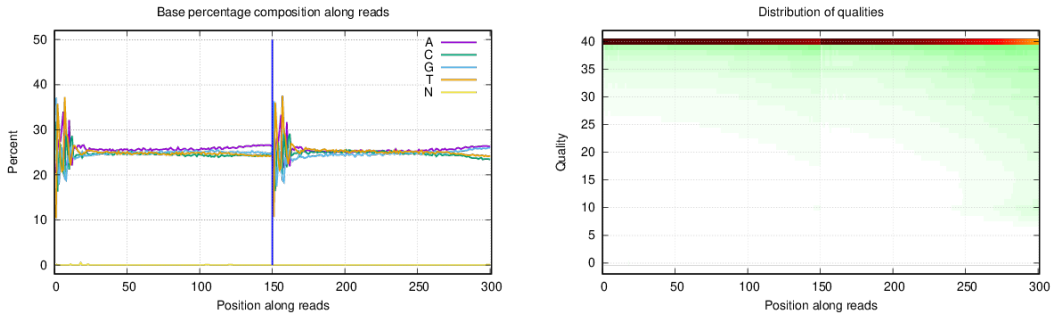


Quality control of sample MoPh15

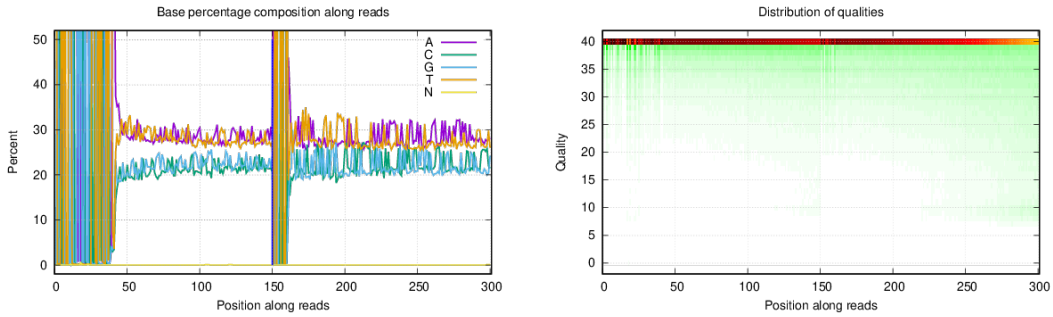




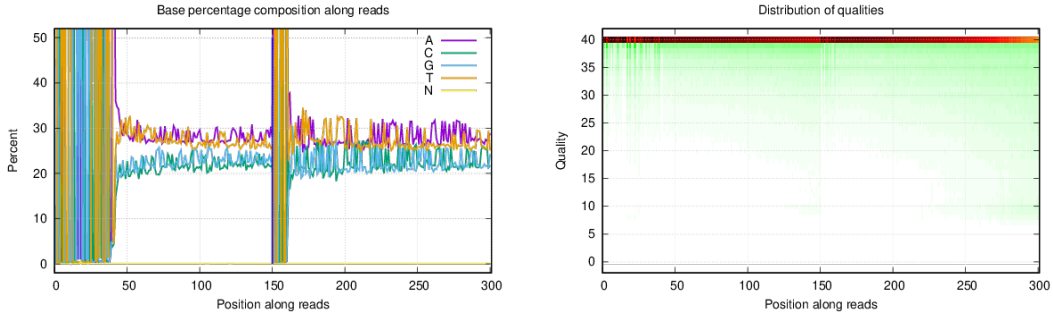
Quality control of sample MoPh7



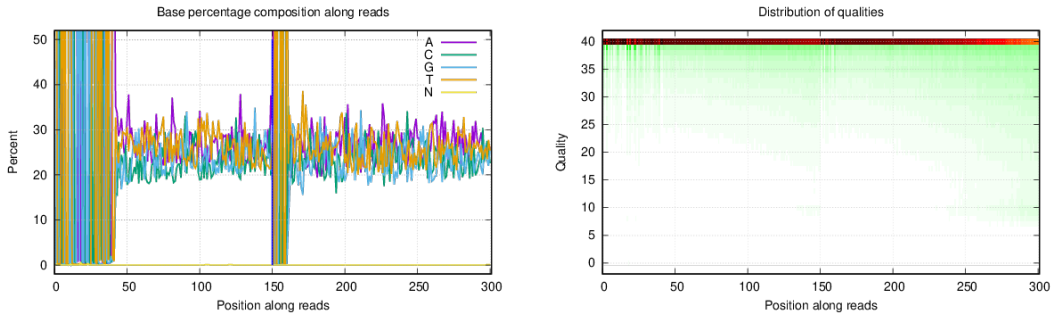
Quality control of sample NCS-2



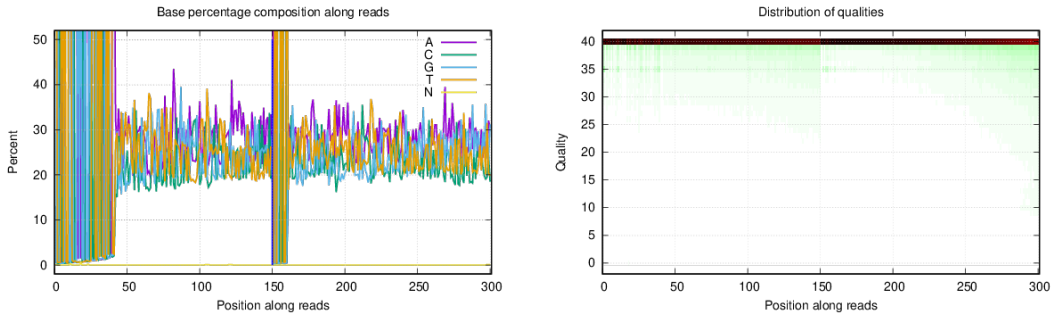
Quality control of sample new_CS-2_EcoRI



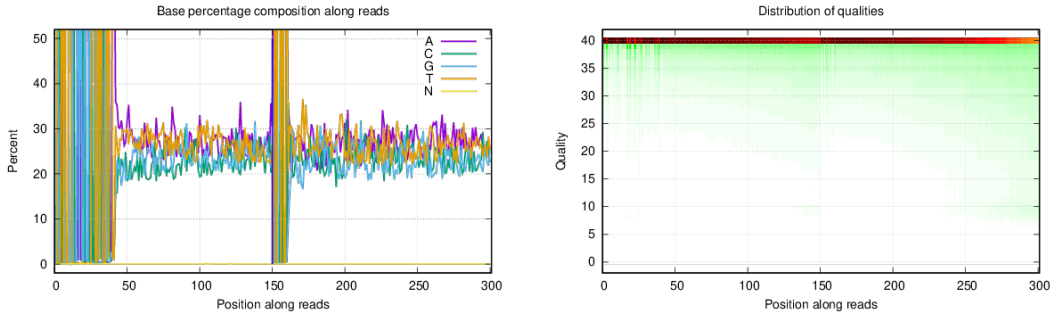
Quality control of sample NP_Ph_1_1603



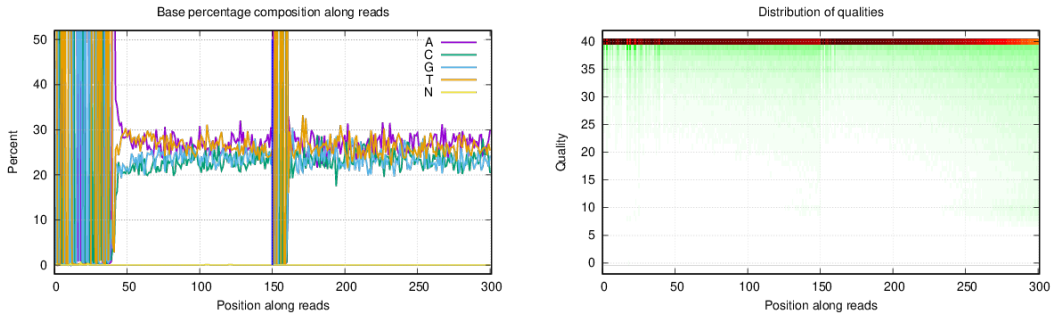
Quality control of sample NP_Ph_1_1803



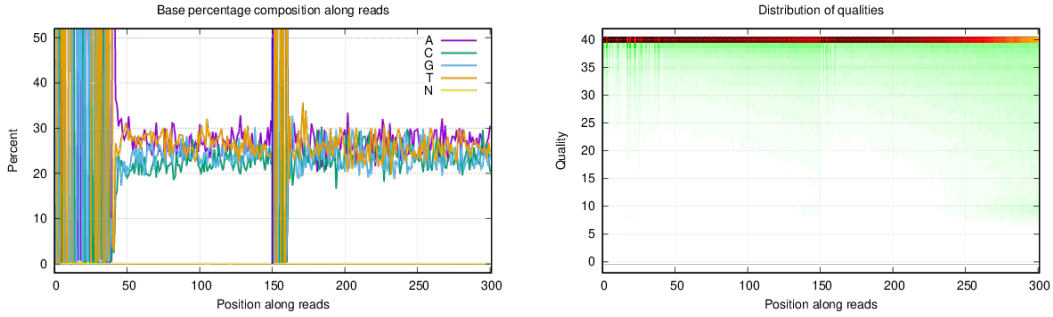
Quality control of sample NP_Ph_1_DNAse_1603



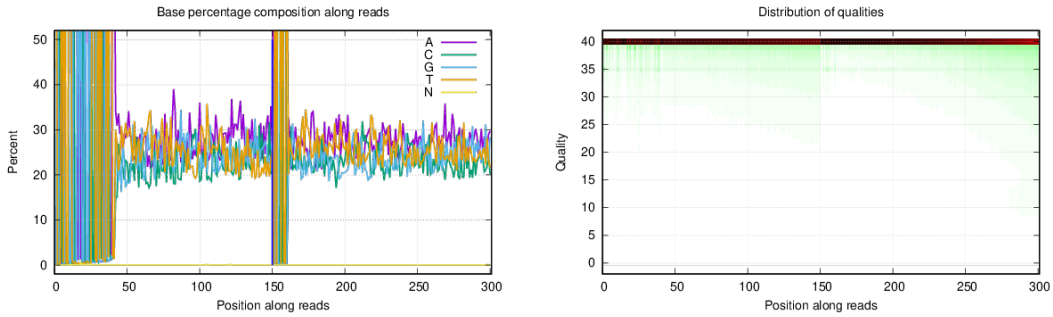
Quality control of sample NP_Ph_2



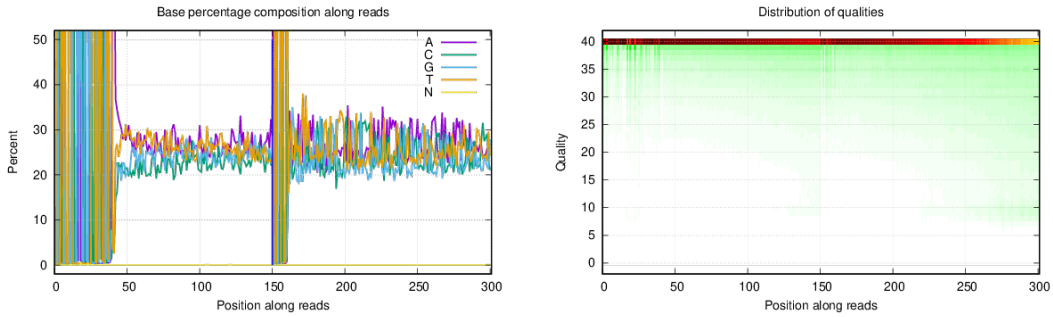
Quality control of sample NP_Ph_Dnase2



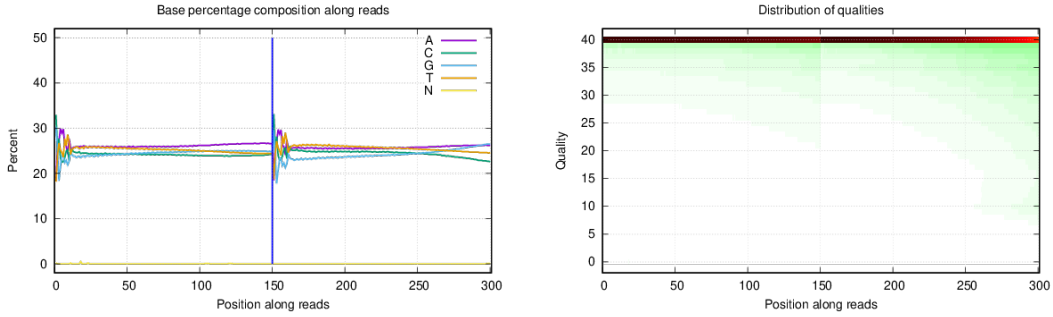
Quality control of sample NP_Ph_DNAse_1803



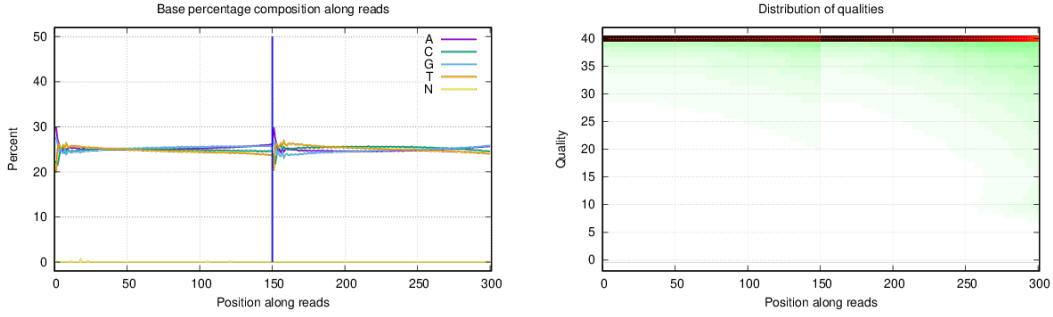
Quality control of sample NP_Ph_DNAse_3



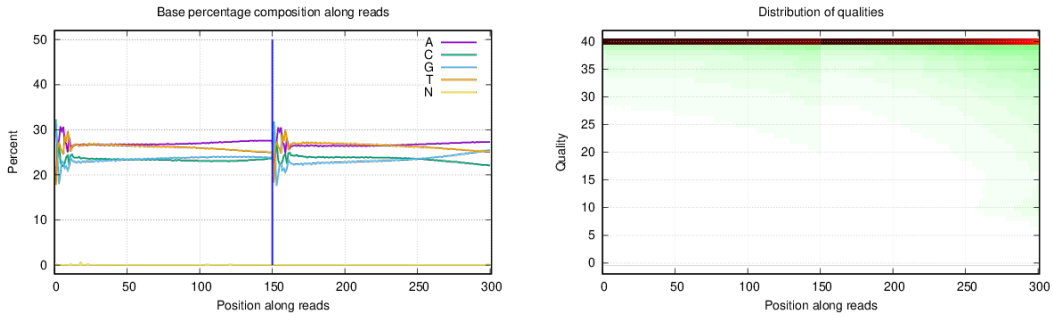
Quality control of sample P125



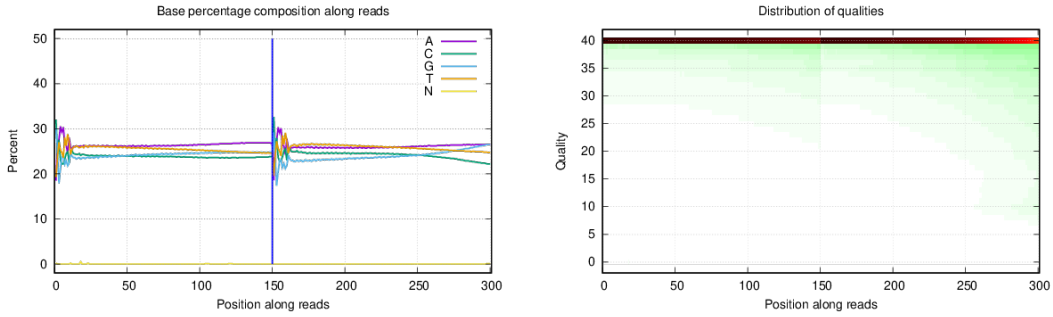
Quality control of sample P126



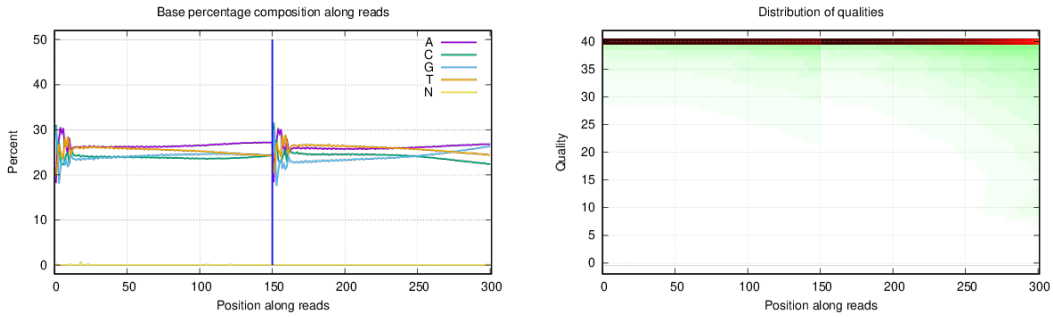
Quality control of sample P131



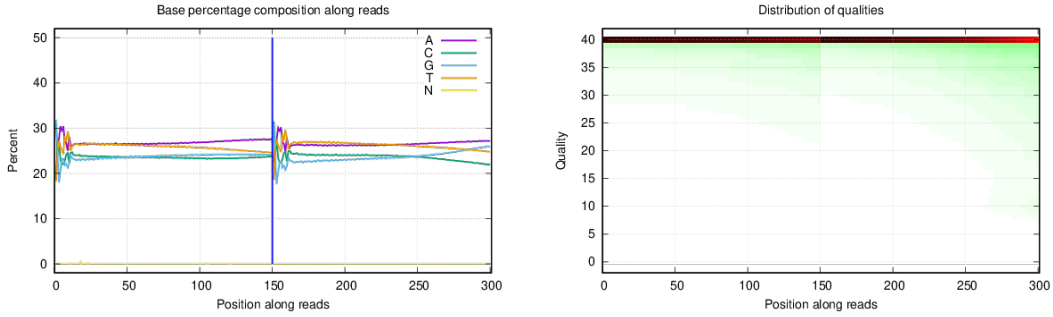
Quality control of sample P132



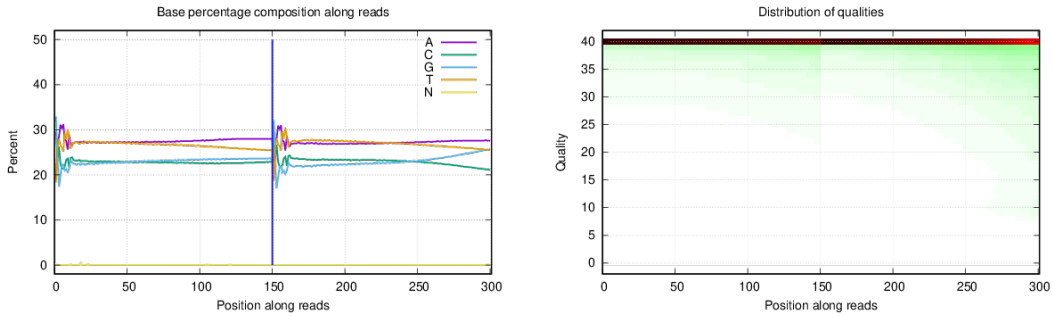
Quality control of sample P133



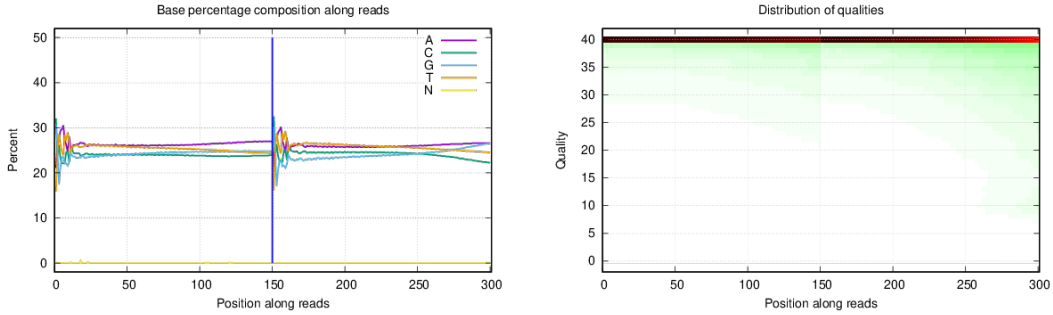
Quality control of sample P135



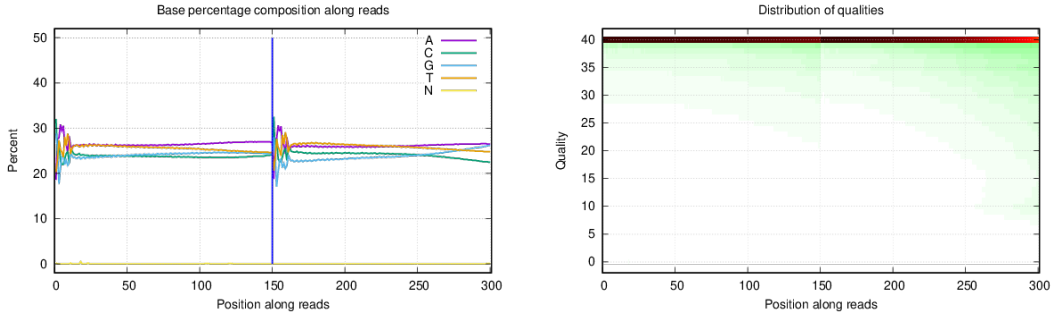
Quality control of sample P136



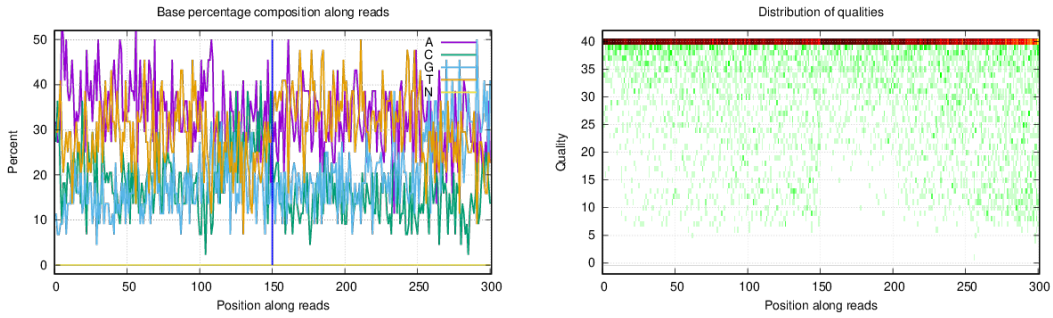
Quality control of sample P137



Quality control of sample P139



Quality control of sample PICO8_8_cells



4 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. Each entry in a FASTQ files consists of 4 lines:

1. A sequence identifier with information about the sequencing run and the cluster. The exact contents of this line vary by based on the BCL to FASTQ conversion software used.
2. The sequence (the base calls; A, C, T, G and N).
3. A separator, which is simply a plus (+) sign.
4. The base call quality scores. These are Phred +33 encoded, using ASCII characters to represent the numerical quality scores.

Here is an example of a single entry in a FASTQ file:

@V300029029L1C001R0010000210/1
GCGACCCCAGGTCAGTCGGGACTACCCGCTGAAGTCGGAGGCCAAGCGGT
+
FFFCFFFFFFFFFDFEFFFFEFEF0FFFFEFFFFFFFEFFFFFECGFFFF

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	?