



# Sequencing Data Report

Project: F25A430000584\_MUSEkmhR

Date: 2025.12.19

Note: For Research Use Only.



---

## Table of Contents

Data Statistics	1
Data Quality Control	3
Help Document	11

## Data Statistics

Raw reads produced from sequencer contain adapters, unknown or low quality bases.

There are 29 samples in this project, the statistics of fastq data is shown below.

Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
CTCF1	150;150	97.13;98.6	91.57;94.94	43.5;43.58	16,684,669	5,005,400,700
CTCF1_input	150;150	96.64;98.36	90.47;94.37	42.46;42.56	9,558,542	2,867,562,600
CTCF2	150;150	96.88;98.57	90.9;94.81	44.2;44.05	19,218,701	5,765,610,300
CTCF2_input	150;150	96.54;98.24	90.24;94.0	42.05;42.23	9,341,216	2,802,364,800
H3K27Ac1	150;150	96.56;97.55	89.9;91.24	45.35;45.49	10,016,823	3,005,046,900
H3K27Ac1_input	150;150	96.32;98.05	89.44;93.06	42.91;42.97	6,661,968	1,998,590,400
H3K27Ac2	150;150	94.07;98.17	83.85;93.54	45.32;45.36	30,627,001	9,188,100,300
H3K27Ac2_input	150;150	96.85;98.47	90.93;94.74	42.0;42.01	11,333,420	3,400,026,000
MoPh15_DR	150;150	96.83;97.14	90.67;90.41	48.35;48.4	17,500,483	5,250,144,900
MoPh15_TR	150;150	96.5;96.71	89.65;88.54	47.77;47.89	10,380,718	3,114,215,400
MoPh7_DR	150;150	96.8;97.54	90.67;91.78	47.22;47.27	15,963,500	4,789,050,000
MoPh7_HBG_DR	150;150	96.47;97.46	89.82;91.7	48.18;48.29	16,954,472	5,086,341,600
MoPh7_HBG_R1	150;150	97.44;97.7	92.34;92.37	47.84;47.93	21,819,886	6,545,965,800
MoPh7_HBG_R2	150;150	96.99;97.47	90.99;91.22	45.28;45.28	17,533,275	5,259,982,500
MoPh7_HBG_TR	150;150	95.39;96.68	86.81;88.4	47.29;47.37	5,290,831	1,587,249,300
MoPh7_TR	150;150	94.93;97.26	85.8;90.36	47.47;47.46	7,647,543	2,294,262,900
i_MoPh15_DR	150;150	96.95;98.02	91.09;93.1	44.08;44.09	17,641,011	5,292,303,300
i_MoPh7_DR	150;150	96.64;98.25	90.33;93.89	43.88;44.05	20,015,280	6,004,584,000
i_MoPh7_HBG_DR	150;150	95.72;98.14	87.98;93.57	45.05;45.17	18,601,740	5,580,522,000
i_MoPh7_HBG_R1	150;150	96.9;98.1	91.03;93.46	44.1;44.2	19,665,973	5,899,791,900
i_MoPh7_HBG_R2	150;150	96.45;97.82	89.66;92.2	42.82;42.95	17,036,552	5,110,965,600
i_MoPh7_HBG_TR	150;150	96.66;97.55	90.15;91.15	43.43;43.58	11,139,873	3,341,961,900
i_MoPh7_TR	150;150	96.12;97.21	88.73;89.88	43.96;44.0	7,497,375	2,249,212,500
rigs_11	150;150	97.03;99.11	91.48;96.72	49.9;50.96	417,070	125,121,000
rigs_14	150;150	97.35;98.79	92.25;95.69	50.5;51.13	246,140	73,842,000
rigs_17	150;150	97.45;99.17	92.52;96.96	49.76;50.14	373,261	111,978,300

### Data Statistics

---

rigs_2	150;150	97.32;99.06	92.13;96.51	50.05;51.41	609,315	182,794,500
rigs_20	150;150	97.56;99.24	92.87;97.25	49.84;50.2	621,115	186,334,500
rigs_3	150;150	97.06;99.08	91.65;96.63	49.74;50.39	225,802	67,740,600

---

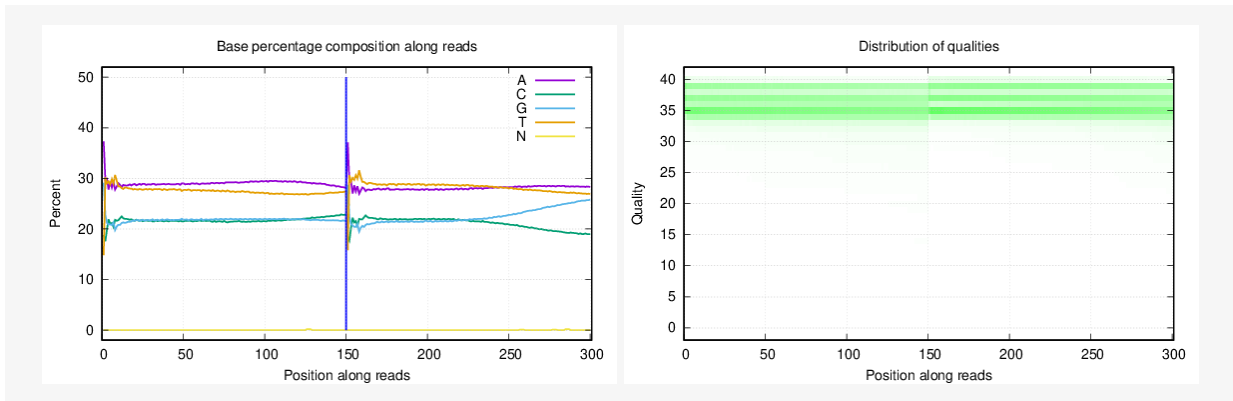
#### Table Format:

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
5. GC Content(%): The proportion of bases G and C
6. Total Reads: The total number of read pairs
7. Total Bases: The total nucleotides number of reads

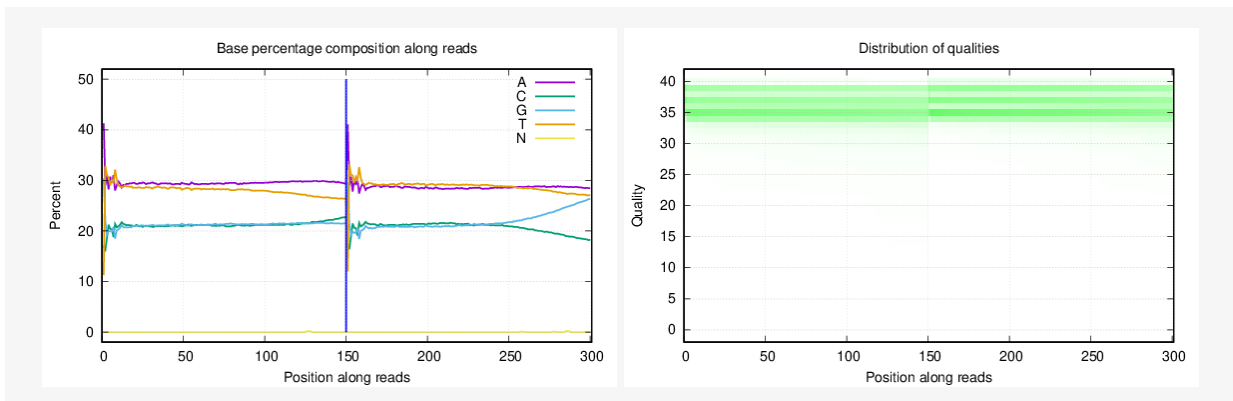
## Data Quality Control

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.

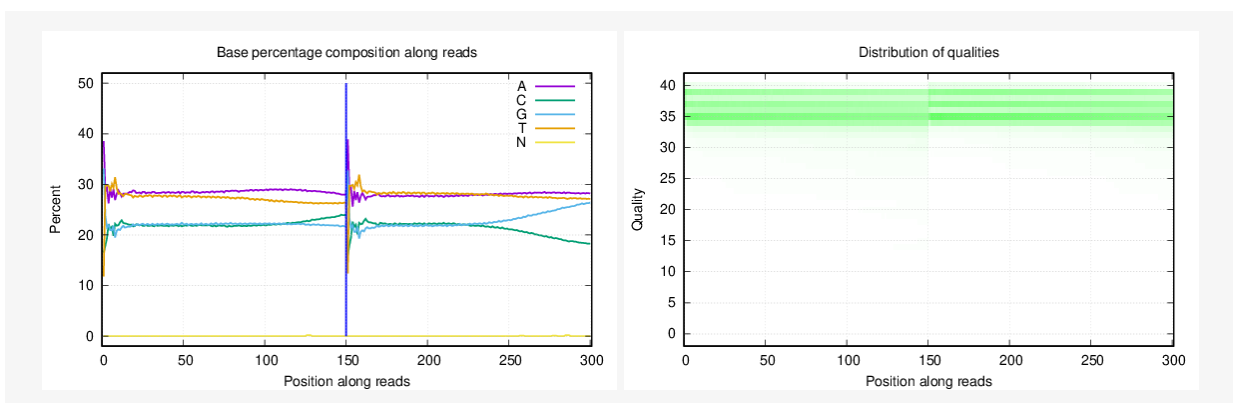
### CTCF1



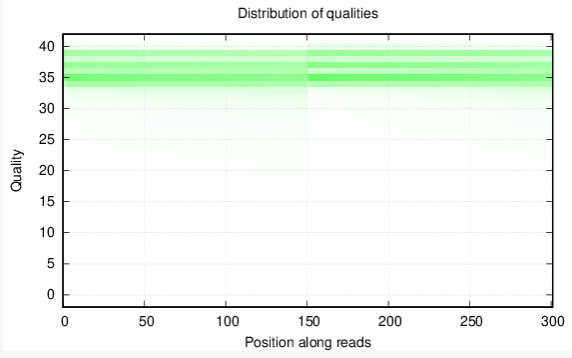
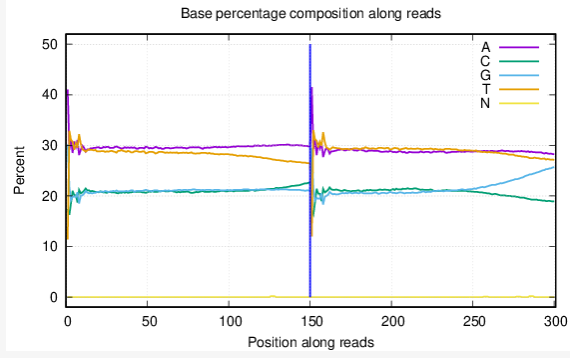
### CTCF1\_input



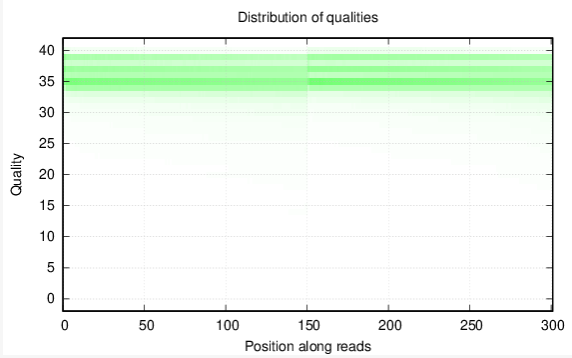
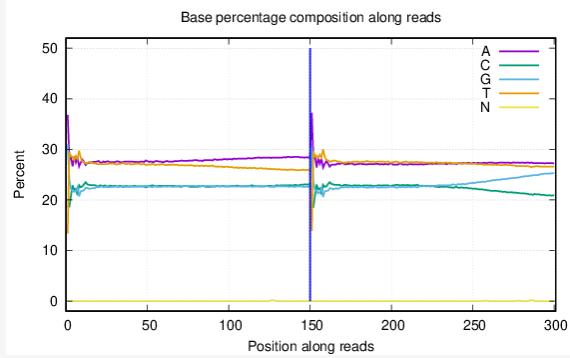
### CTCF2



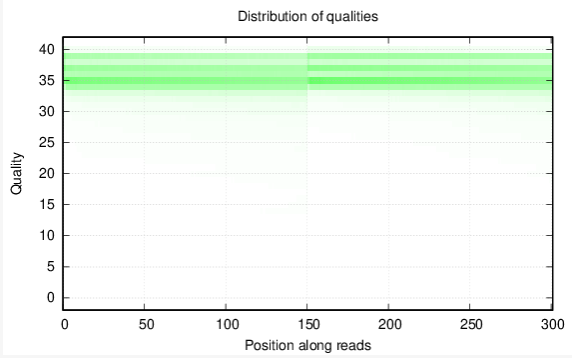
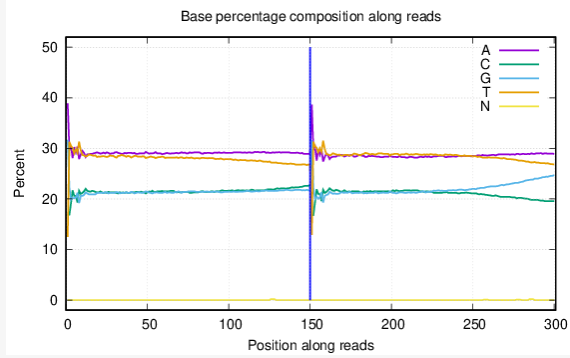
### CTCF2\_input



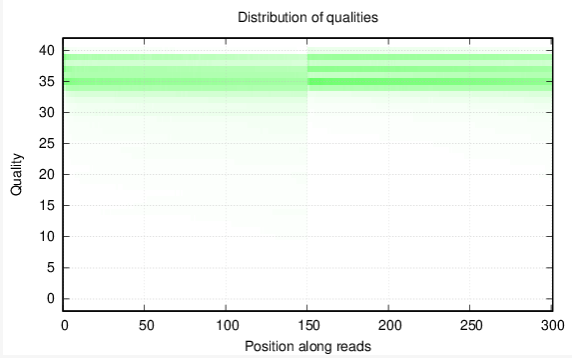
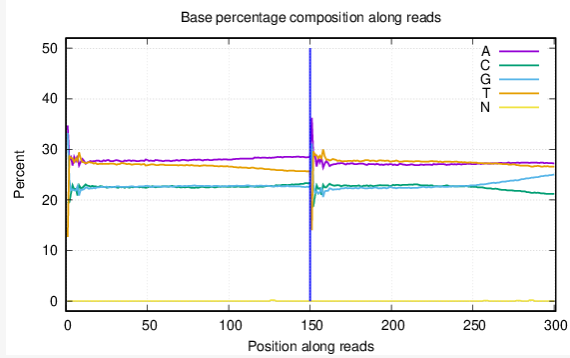
H3K27Ac1



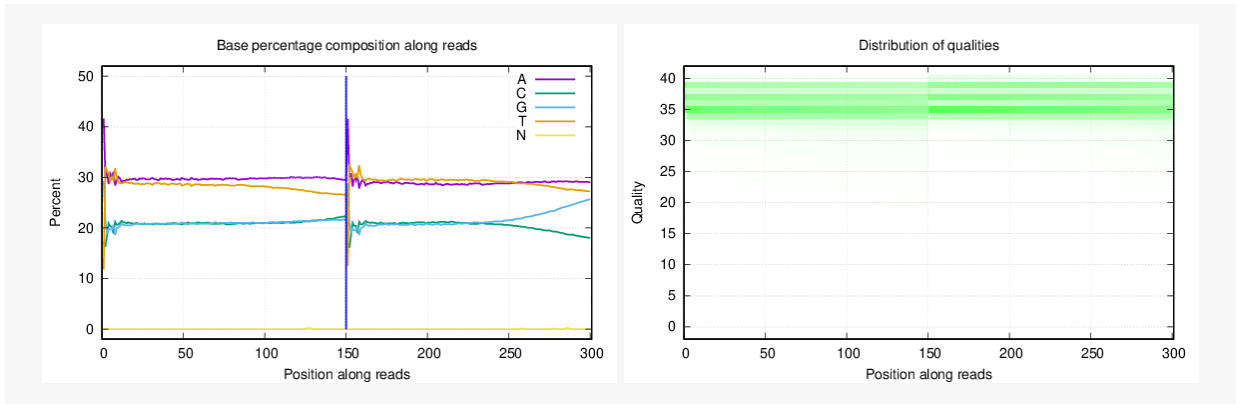
H3K27Ac1\_input



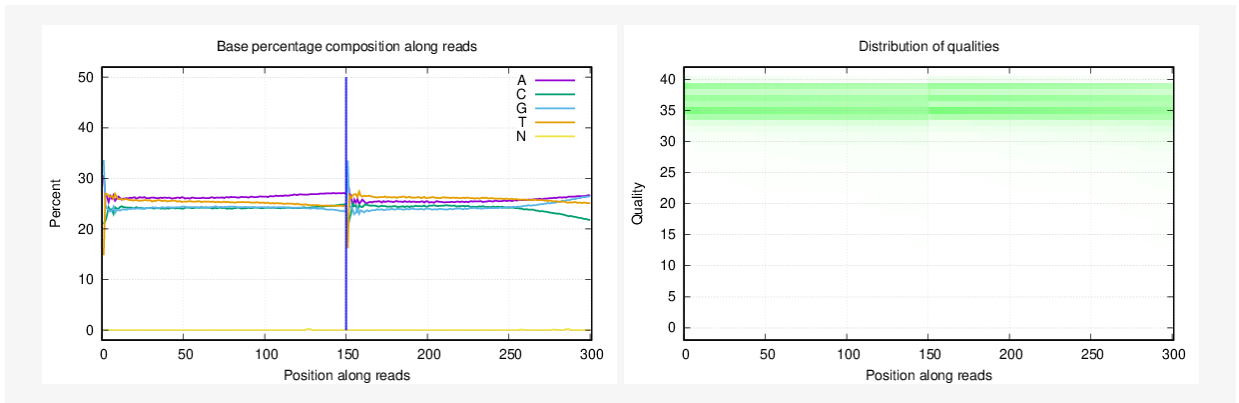
H3K27Ac2



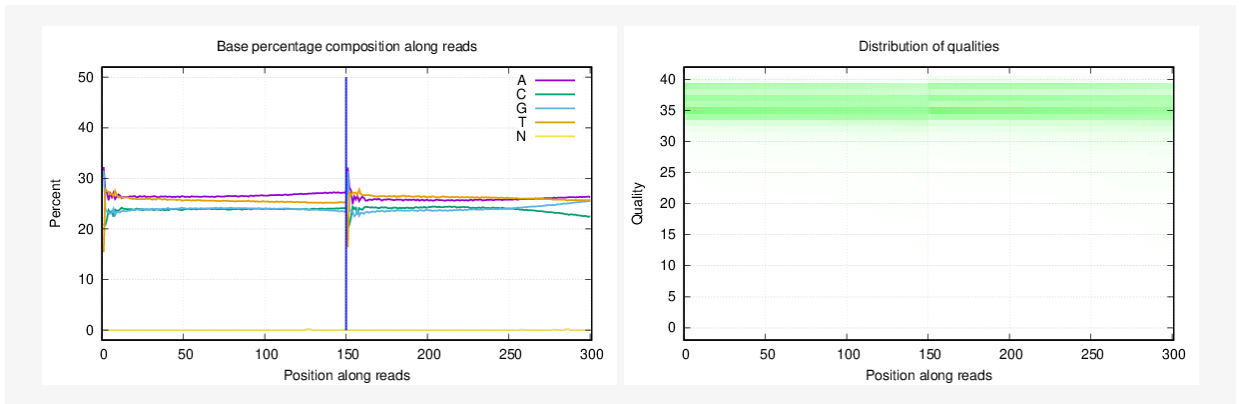
H3K27Ac2\_input



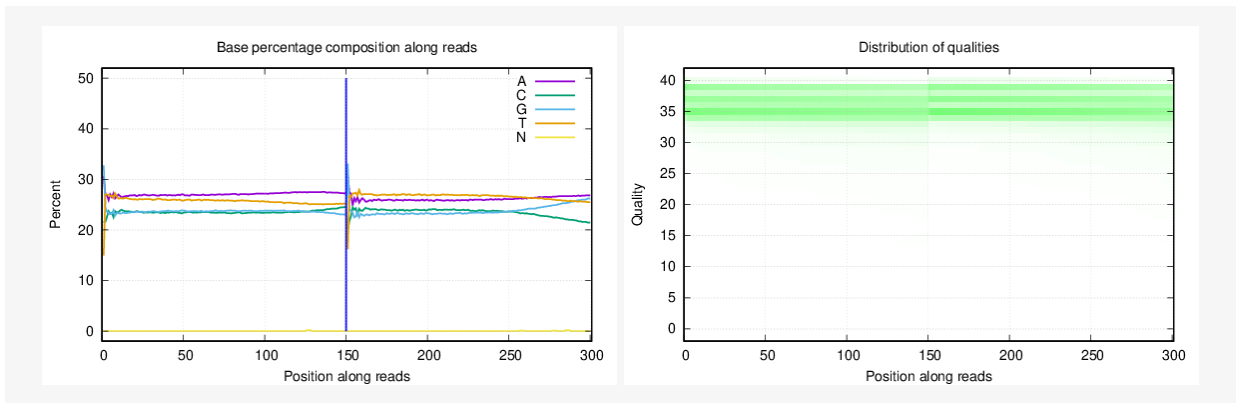
MoPh15\_DR



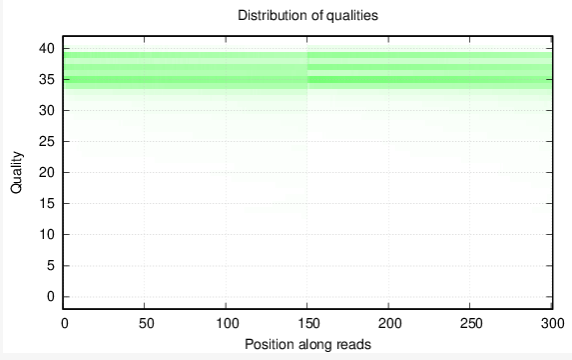
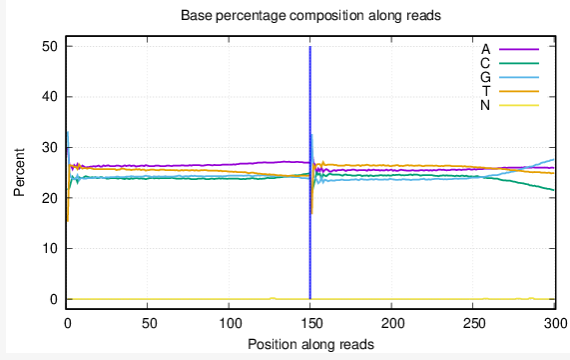
MoPh15\_TR



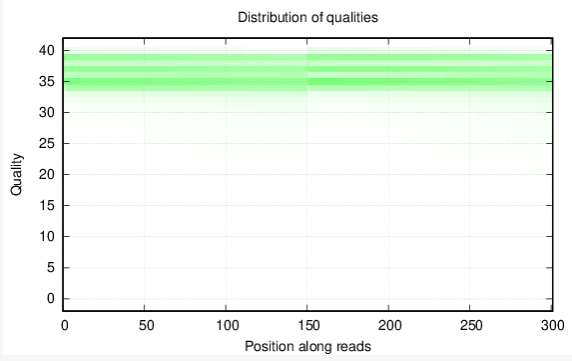
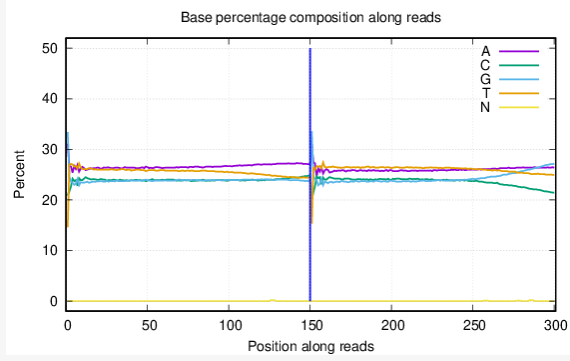
MoPh7\_DR



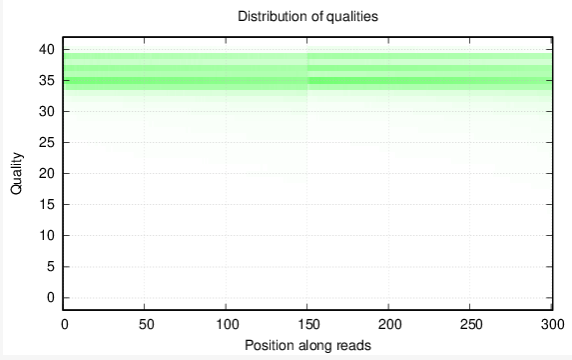
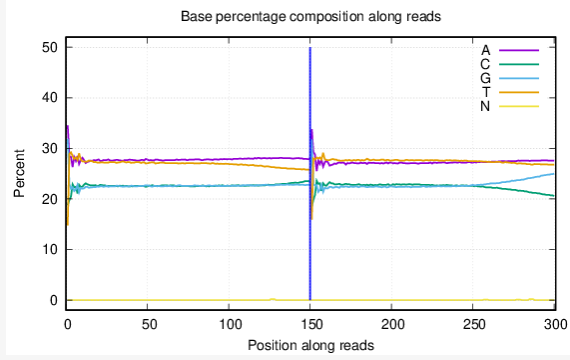
MoPh7\_HBG\_DR



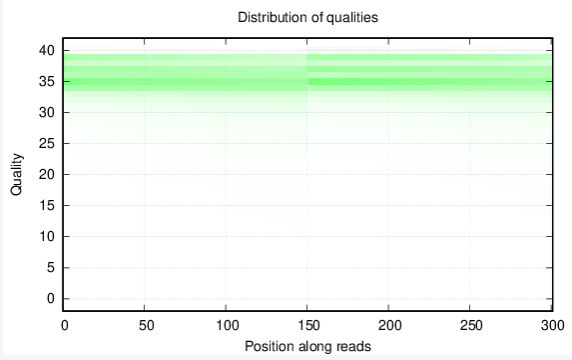
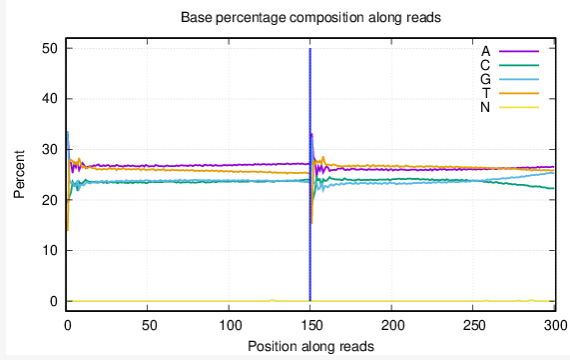
MoPh7\_HBG\_R1



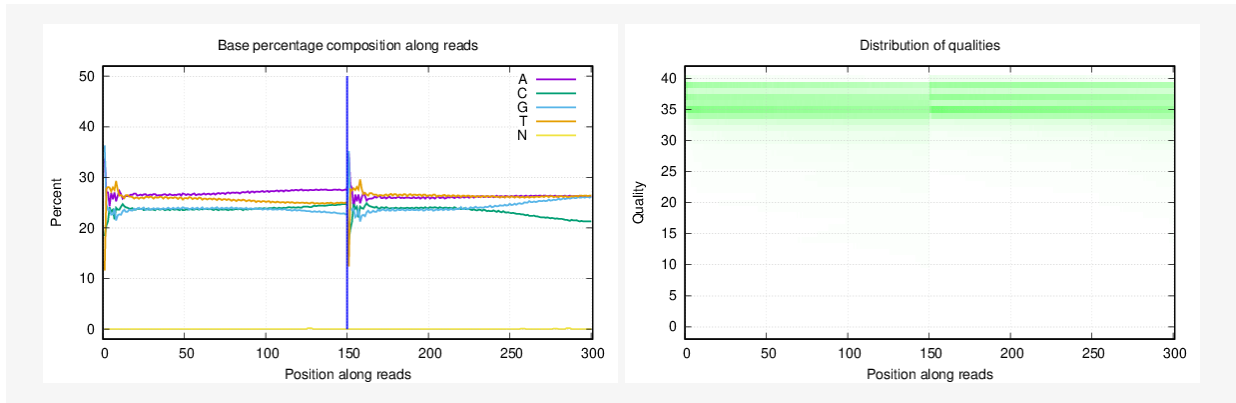
MoPh7\_HBG\_R2



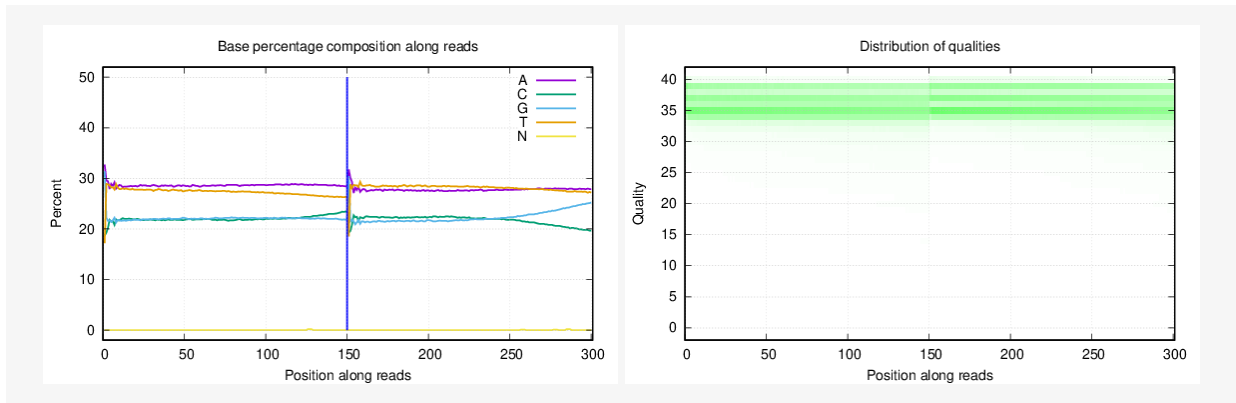
MoPh7\_TR



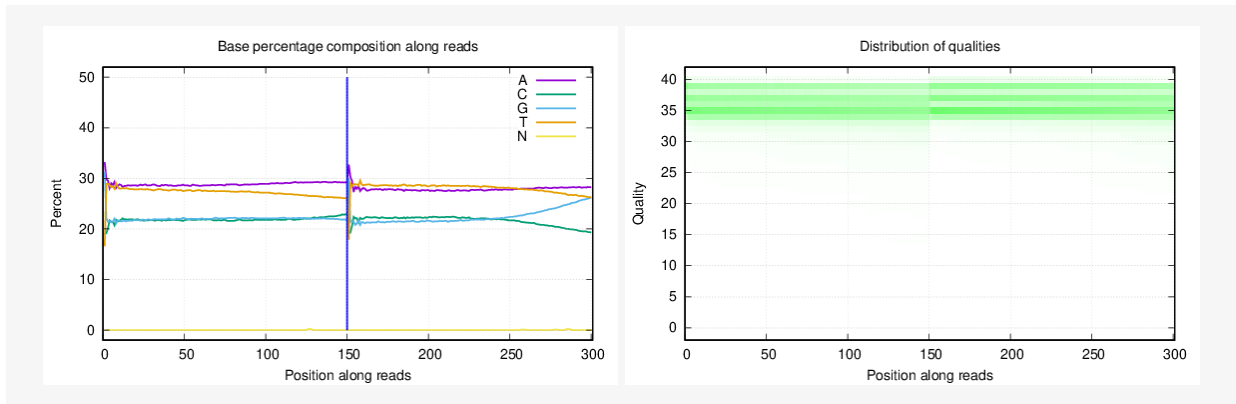
MoPh7\_TR



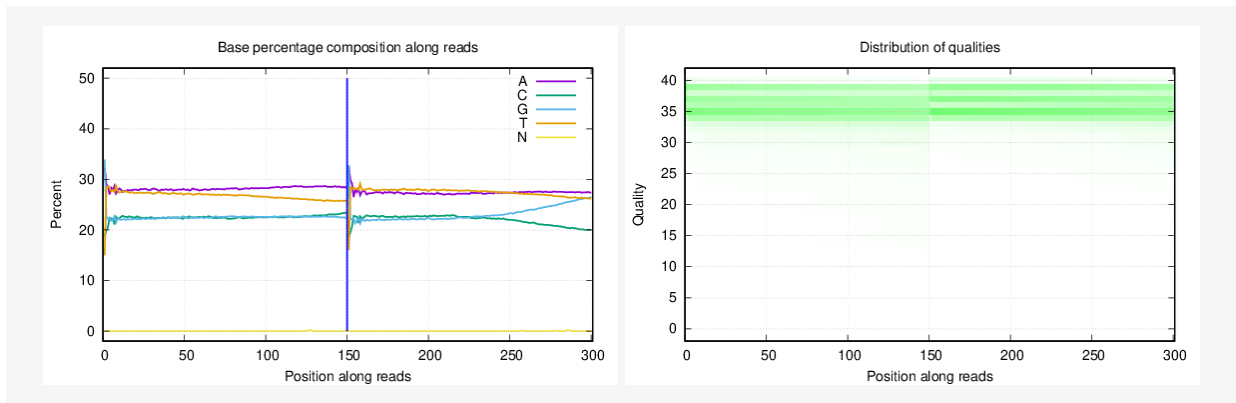
**i\_MoPh15\_DR**



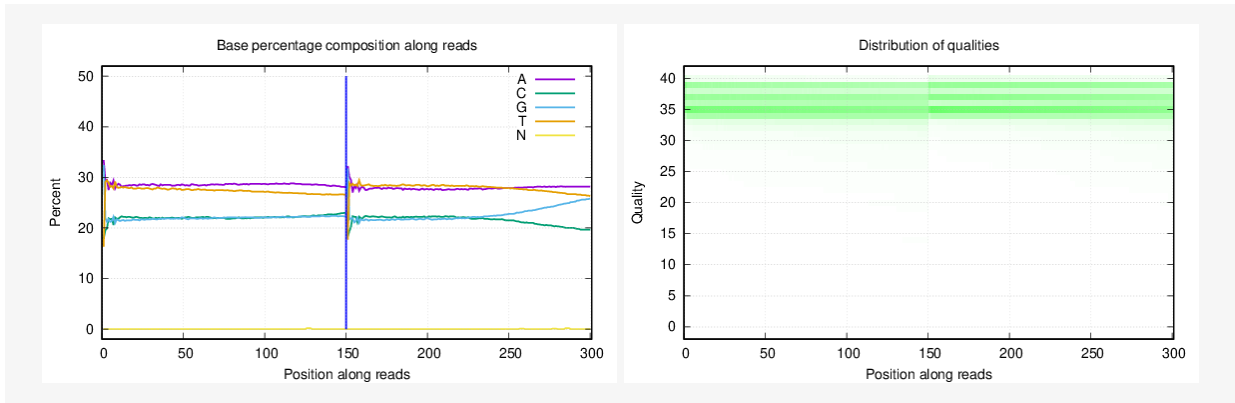
**i\_MoPh7\_DR**



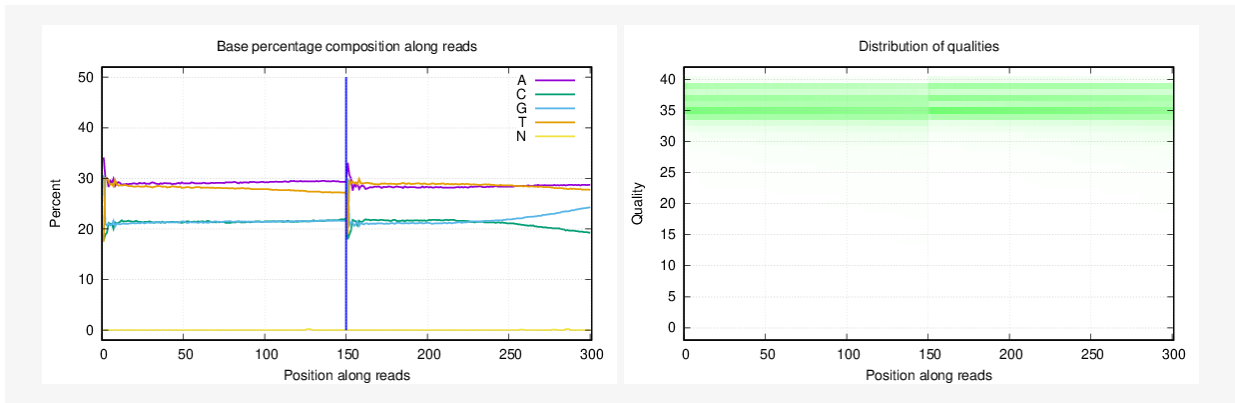
**i\_MoPh7\_HBG\_DR**



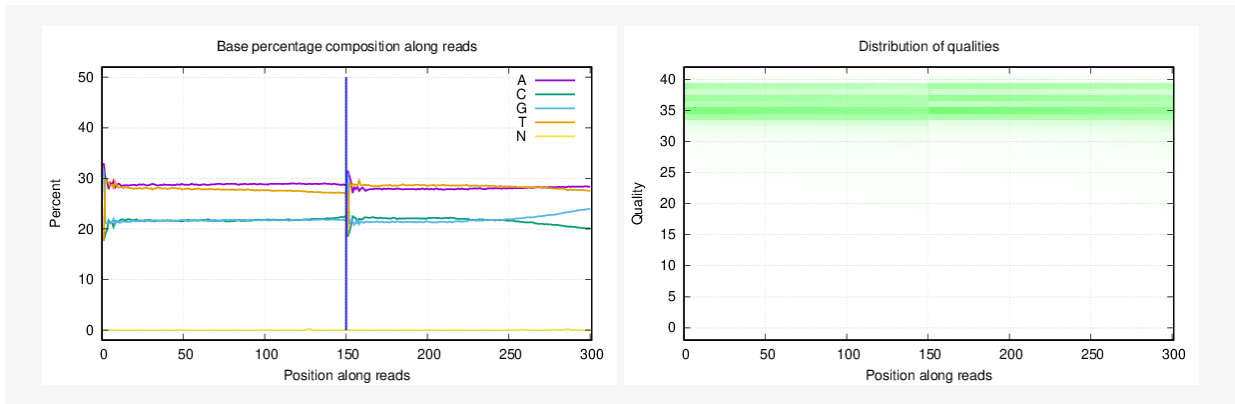
**i\_MoPh7\_HBG\_R1**



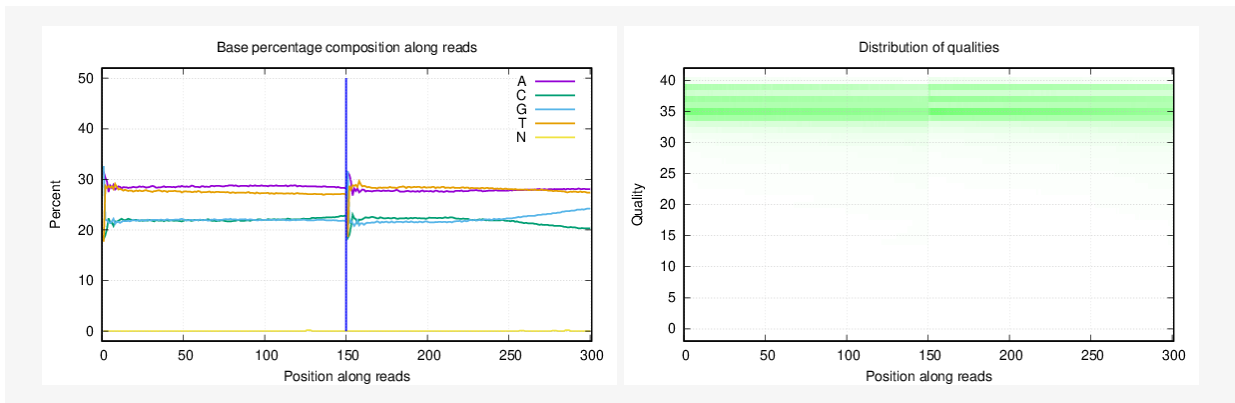
i\_MoPh7\_HBG\_R2



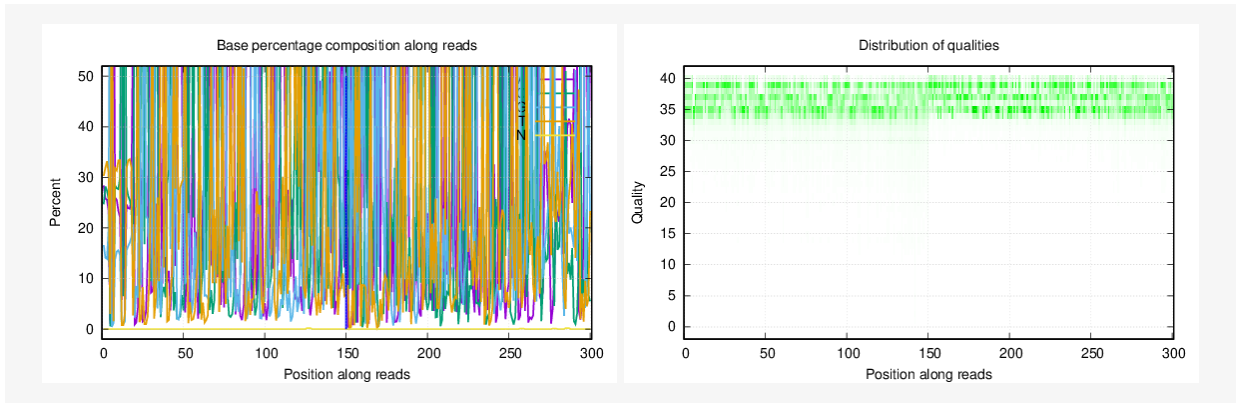
i\_MoPh7\_HBG\_TR



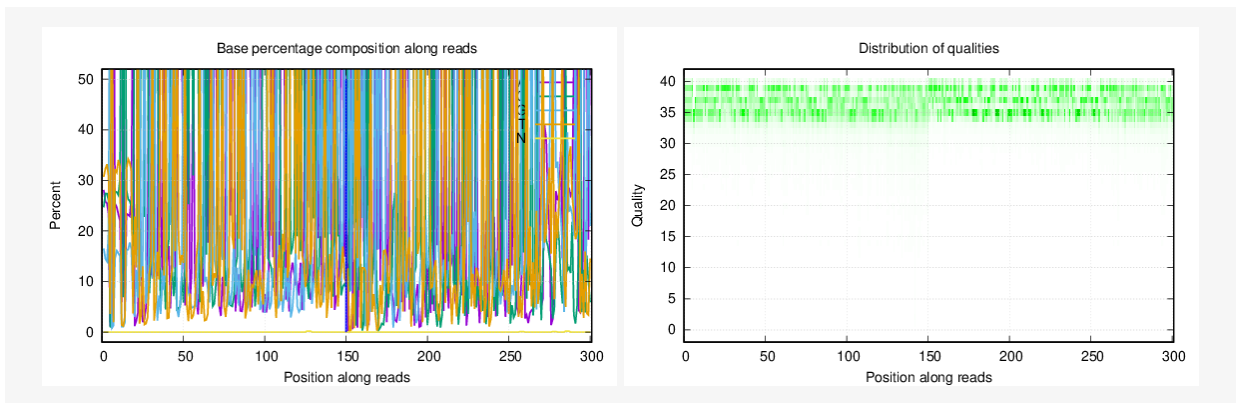
i\_MoPh7\_TR



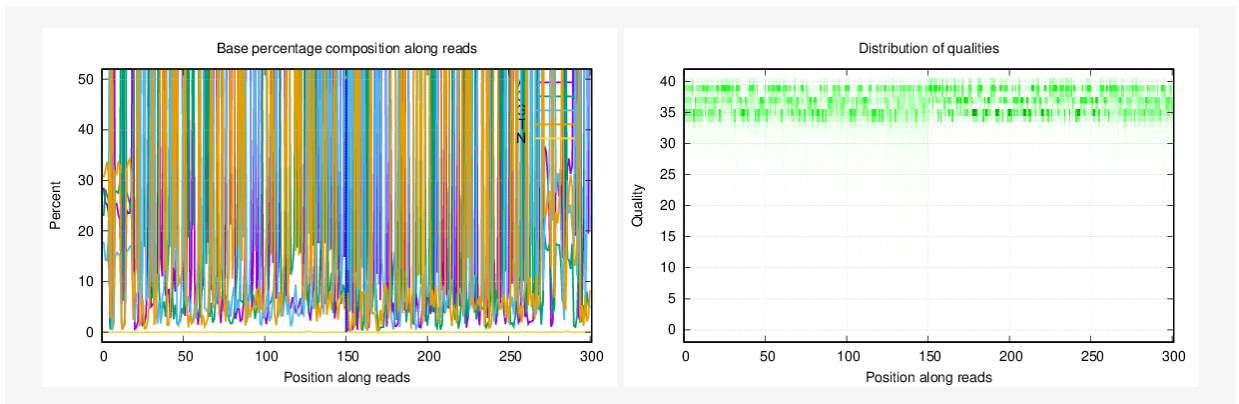
rigs\_11



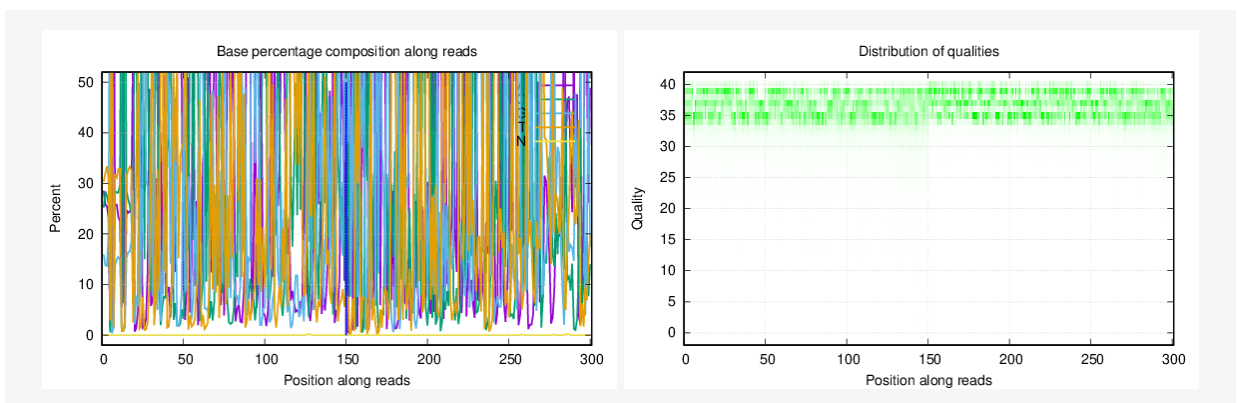
rigs\_14



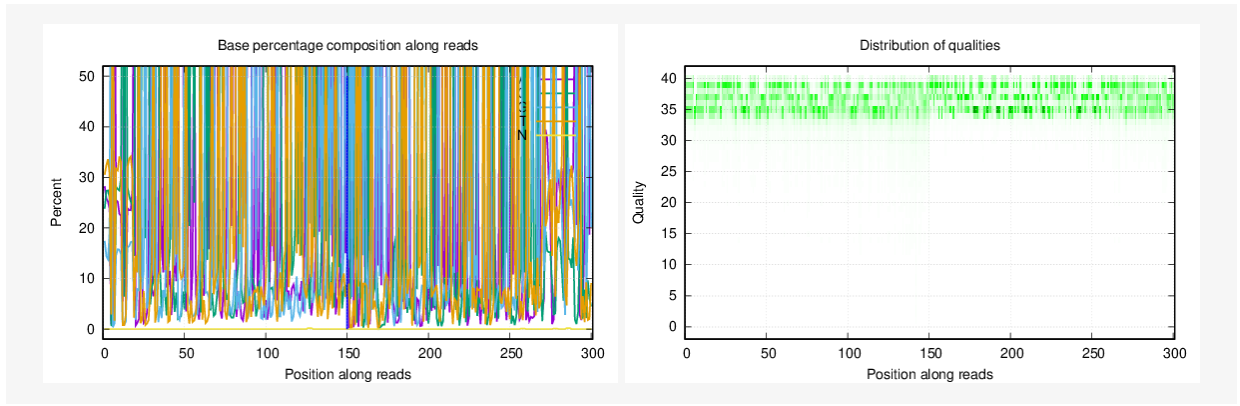
rigs\_17



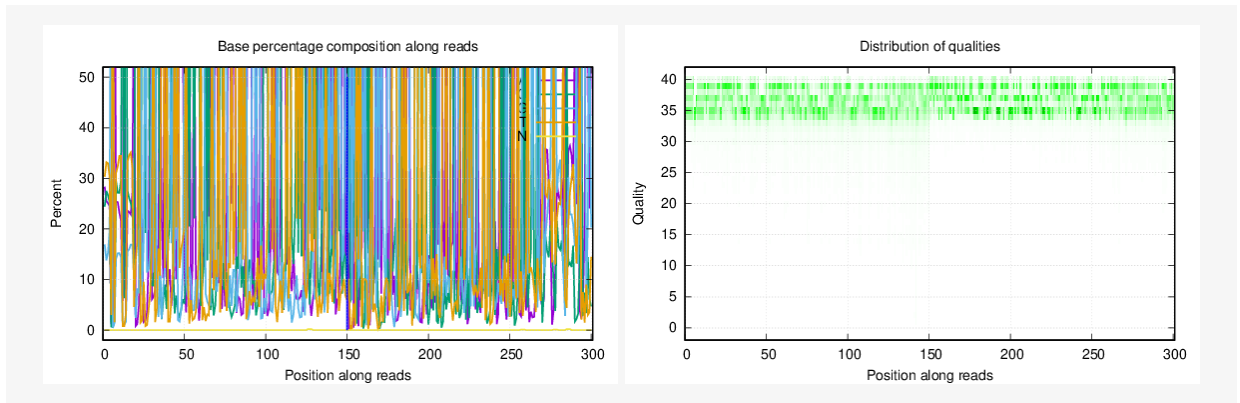
rigs\_2



rigs\_20



rigs\_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. 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
+
FFFCFFFFFFFFDFEFFFFFFEF0FFFFFFFFFFFFFFFFEFCGFFFF
```

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	?

# BGI·Tech

## Omics For All

### | Contact us

Website: [www.bgi.com](http://www.bgi.com)

E-mail: [info@bgi.com](mailto:info@bgi.com)

**For Research Use Only. Not for use in diagnostic procedures.**

© 2023 BGI Genomics Co.,Ltd. All right reserved. All trademarks are the property of BGI, or their respective owners.