Sequencing Data Report

Project: F20FTSEUHT0946-02_MUSrwmqR
Date: 2025.2.20
Note: For Research Use Only.

BGI Genomics Co.,Ltd.

Table of Contents

Data Statistics

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

There are 36 samples in this project, the statistics of fastq data is shown below.

Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
34oocyt	150;150	96.65;98.09	90.28;93.88	43.64;43.78	2,919,092	875,727,600
Arab_34	150;150	97.87;98.95	93.58;96.42	43.75;43.56	428,629	128,588,700
Arab_54	150;150	97.82;98.97	93.41;96.48	44.57;44.99	411,905	123,571,500
CTCF1	150;150	97.25;98.09	91.75;93.89	52.28;52.81	90,245	27,073,500
CTCF2	150;150	96.99;98.21	91.10;94.07	49.09;49.44	75,455	22,636,500
Cenp_B1	150;150	97.75;98.54	93.26;95.18	50.31;50.37	6,272,986	1,881,895,800
Cenp_B2	150;150	97.18;98.26	91.68;94.31	46.67;46.86	10,243,720	3,073,116,000
Cenp_P2	150;150	96.81;98.14	90.62;93.93	44.32;44.44	12,162,958	3,648,887,400
Ch_R1	150;150	97.18;98.52	91.63;95.17	42.97;43.09	14,734,040	4,420,212,000
Ch_R2	150;150	97.87;98.96	93.63;96.58	43.67;43.64	20,824,876	6,247,462,800
Ch_aux_R1	150;150	97.78;99.03	93.38;96.85	44.06;44.20	13,139,713	3,941,913,900
Ch_aux_R2	150;150	96.46;99.05	89.58;96.86	44.21;44.29	16,119,868	4,835,960,400
Dik_e	150;150	96.79;96.88	90.46;90.08	43.60;43.74	6,217,119	1,865,135,700
Dik_e1	150;150	96.56;97.29	89.84;91.24	41.97;42.11	15,574,766	4,672,429,800
Fed1_e	150;150	96.52;96.69	89.68;89.41	42.60;42.68	4,848,353	1,454,505,900
K11	150;150	96.68;97.20	90.15;90.95	42.49;42.68	9,367,596	2,810,278,800
K8	150;150	96.54;96.29	89.65;88.15	42.57;42.71	7,986,104	2,395,831,200
Sach1_K_2110	150;150	96.78;96.83	90.39;89.88	43.34;43.47	6,327,443	1,898,232,900
Say3_K	150;150	96.88;97.08	90.69;90.65	42.43;42.57	5,913,914	1,774,174,200
Shur3_K	150;150	96.44;96.45	89.41;88.77	42.46;42.56	5,702,246	1,710,673,800
Vla1_K	150;150	96.64;97.09	90.09;90.69	42.45;42.41	6,815,479	2,044,643,700
Vla2_K	150;150	96.57;97.39	89.94;91.60	42.86;42.85	5,625,188	1,687,556,400
Vla6_K	150;150	96.58;96.61	89.88;89.18	43.06;43.21	5,486,739	1,646,021,700
Vla9_e	150;150	96.44;96.52	89.44;88.85	42.17;42.22	4,731,633	1,419,489,900
Vor1_K	150;150	96.38;97.15	89.38;90.88	43.19;43.28	6,659,146	1,997,743,800
br_RAD21	150;150	98.05;98.31	93.95;94.31	55.24;54.93	74,722	22,416,600

eb186_10d	150;150	97.47;98.39	92.50;94.82	43.29;43.09	39,854,461	11,956,338,300
eb186_10d_i	150;150	96.08;96.50	89.09;90.07	46.12;46.41	8,177,738	2,453,321,400
eb1_36_10d	150;150	97.54;98.71	92.74;95.85	42.67;43.09	26,462,245	7,938,673,500
eb1_36_10d_i	150;150	96.53;97.15	89.98;91.31	46.37;46.40	10,057,225	3,017,167,500
hsRad_R1	150;150	97.14;98.46	91.55;95.01	43.14;43.33	16,702,464	5,010,739,200
hsRad_R2	150;150	98.12;98.93	94.38;96.53	42.84;43.00	19,993,062	5,997,918,600
hsRad_aux_R1	150;150	97.71;98.75	93.24;95.97	43.13;43.12	16,431,097	4,929,329,100
hsRad_aux_R2	150;150	97.87;98.79	93.66;96.07	44.40;44.26	26,592,156	7,977,646,800
st_CTCF_L	150;150	97.57;98.68	92.61;95.37	44.49;44.16	52,203	15,660,900
st_CTCF_R	150;150	97.55;98.63	92.54;95.23	43.83;44.50	60,175	18,052,500

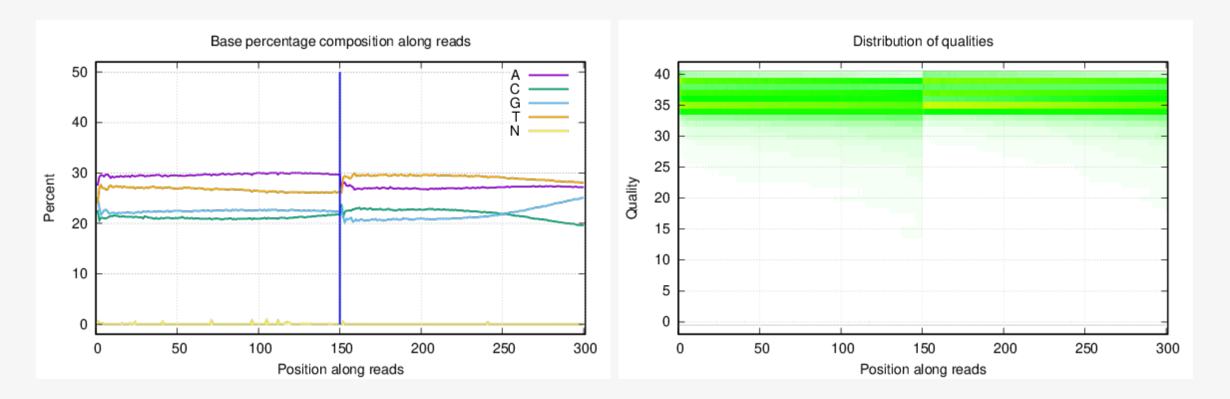
Table Format:

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

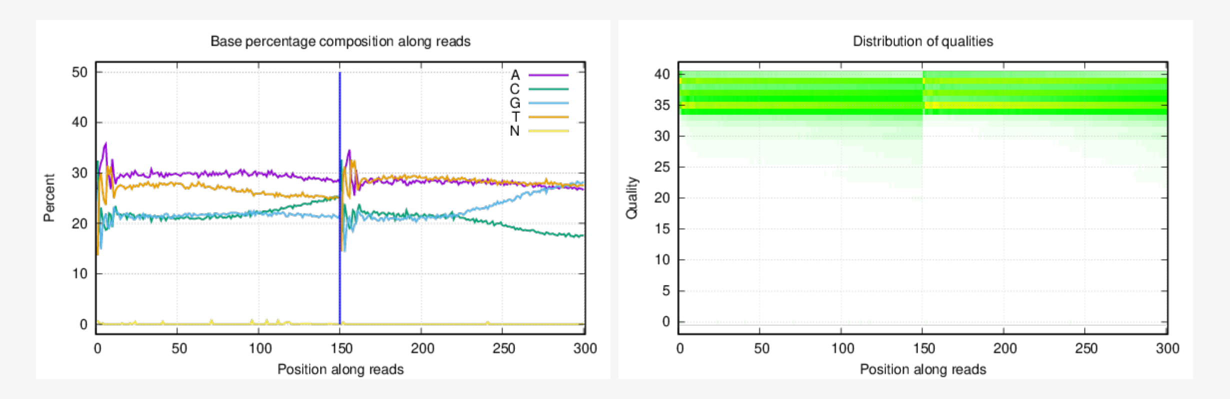
Data Quality Control

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.

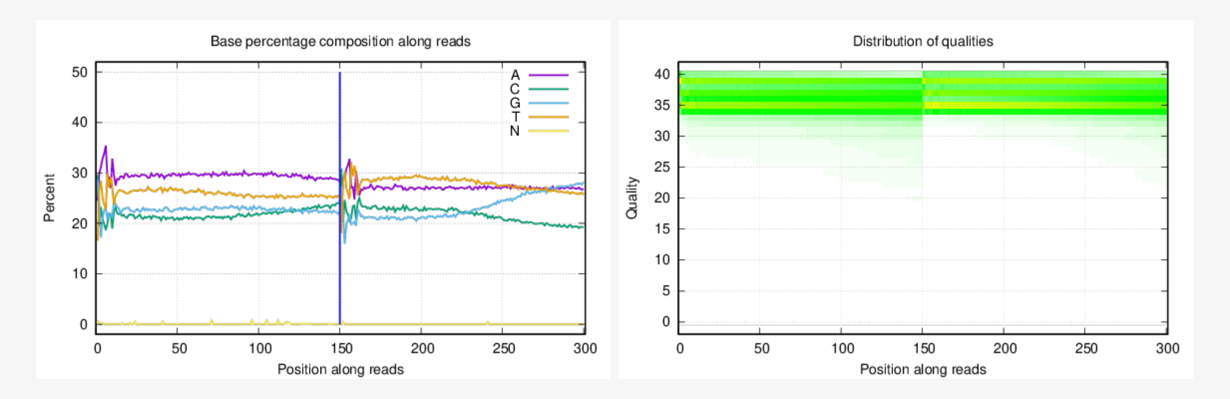
34oocyt



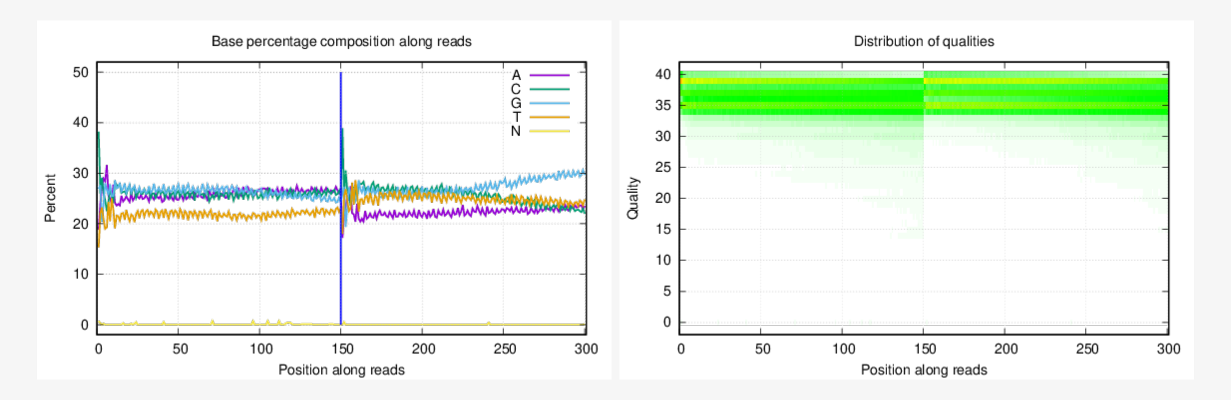
Arab_34



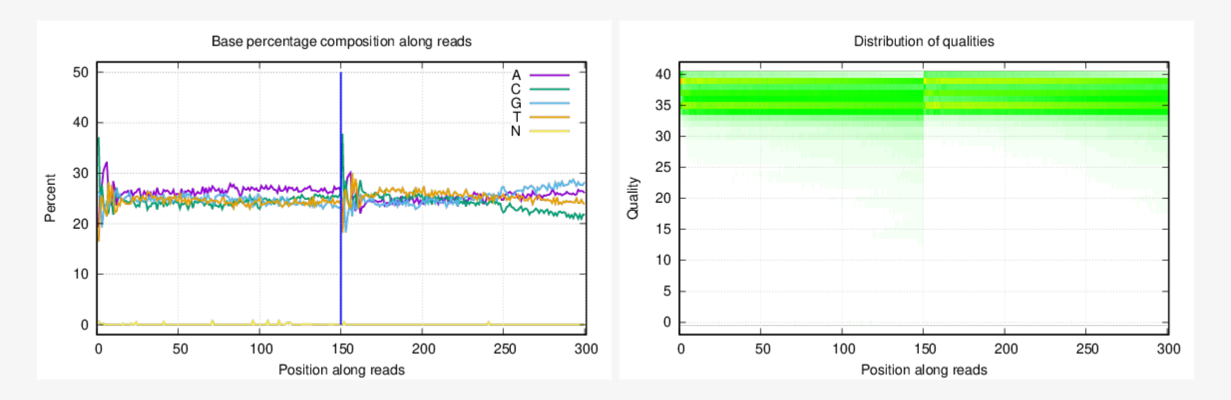
Arab_54



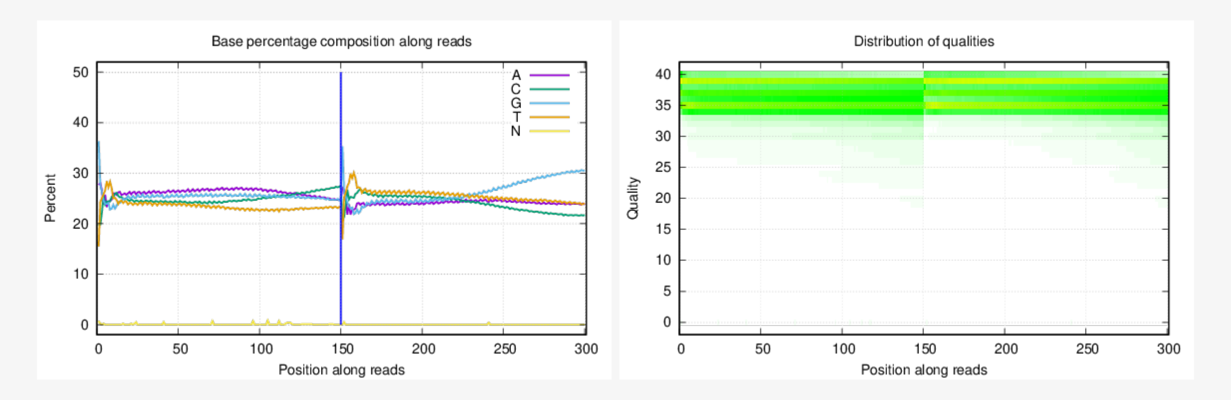
CTCF1



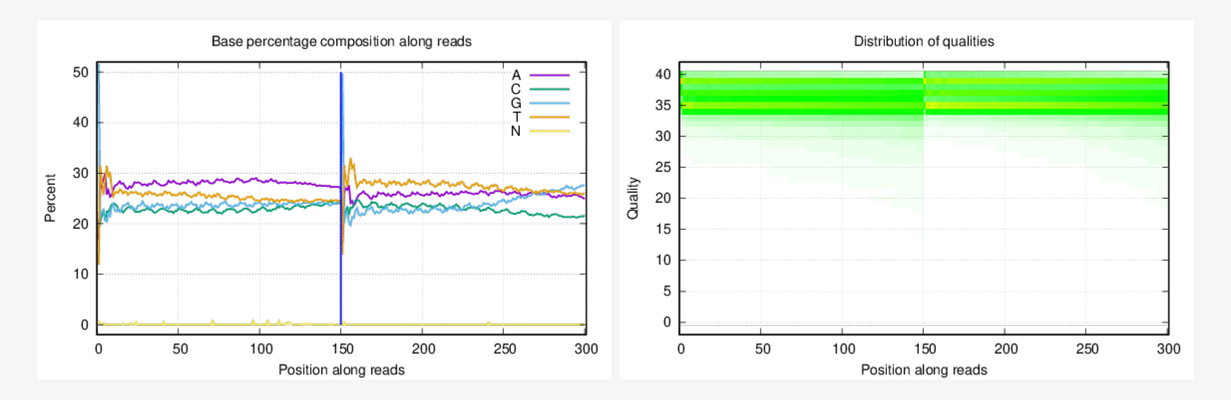
CTCF2



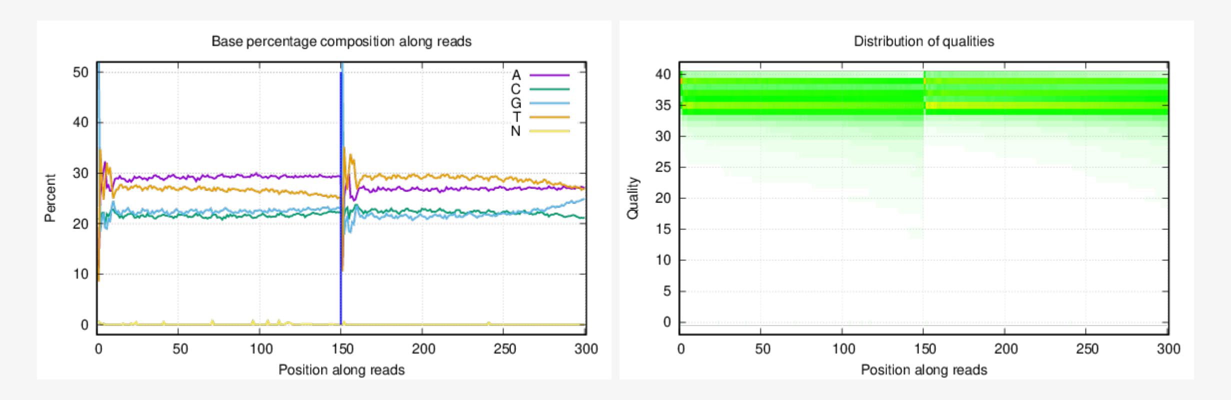
Cenp_B1



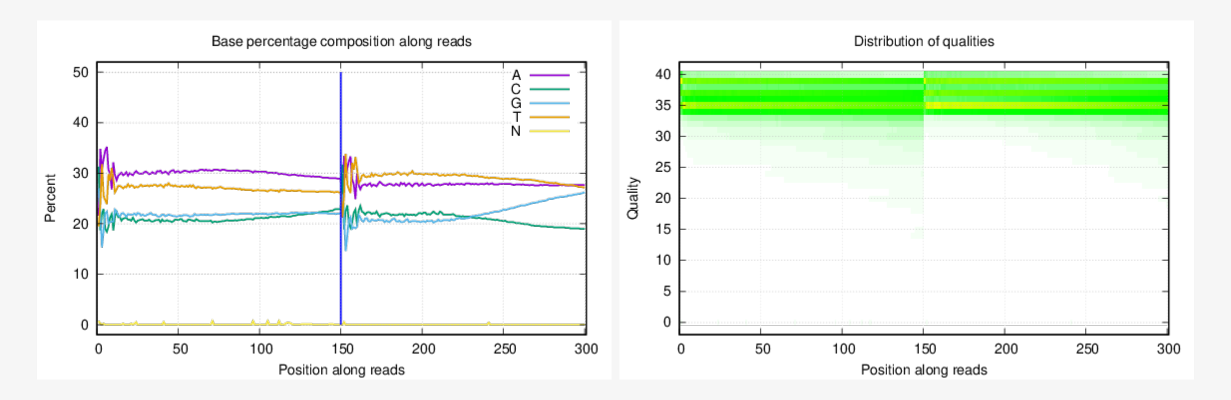
Cenp_B2



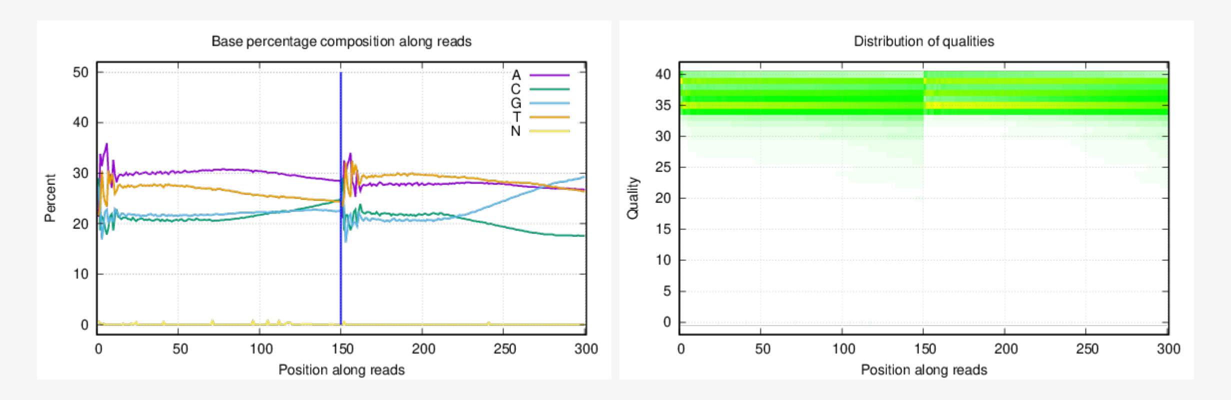
Cenp_P2



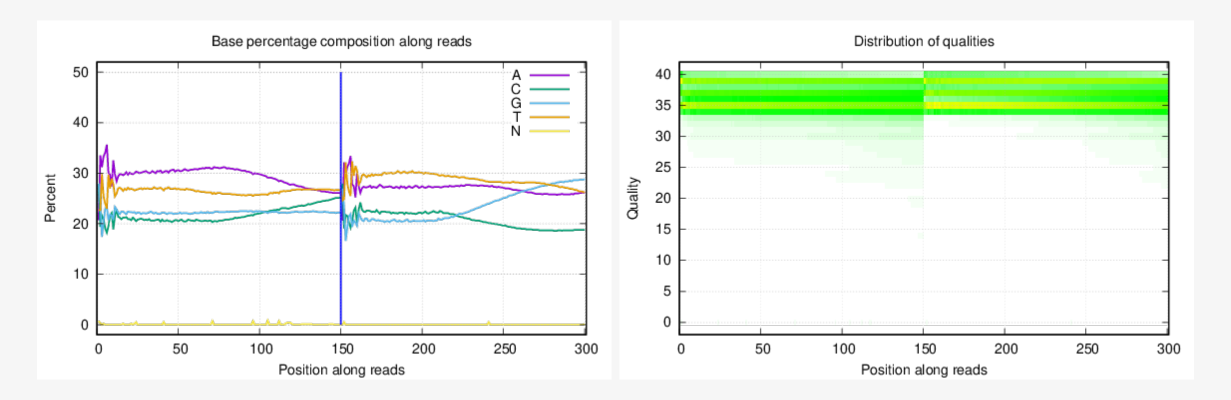
Ch_R1



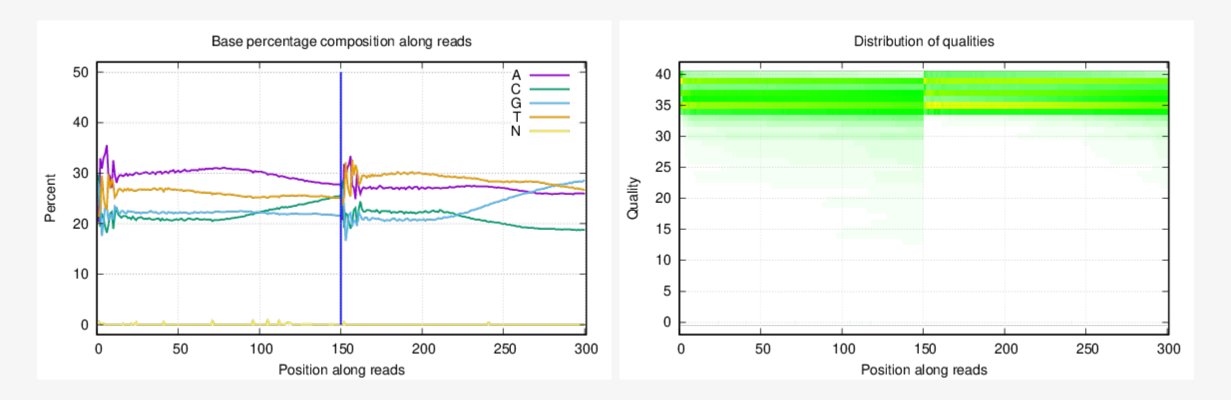
Ch_R2



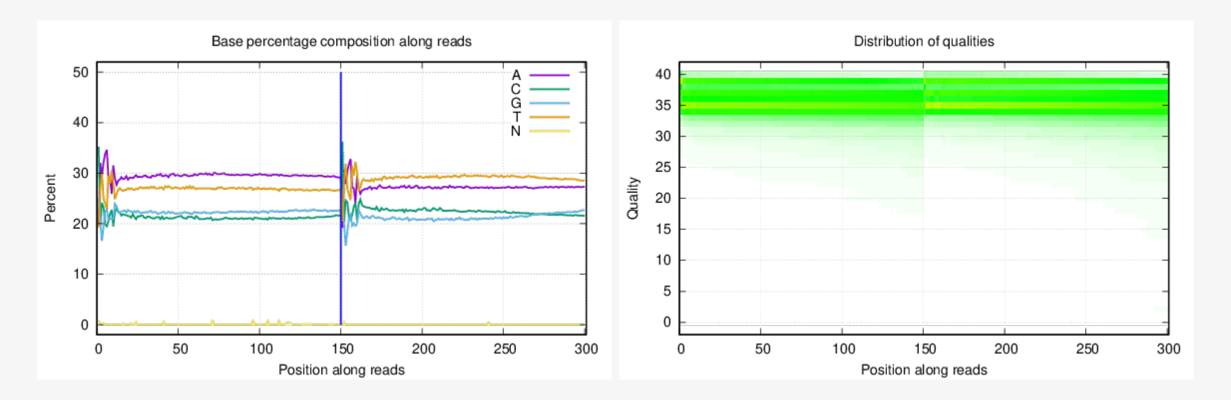
Ch_aux_R1



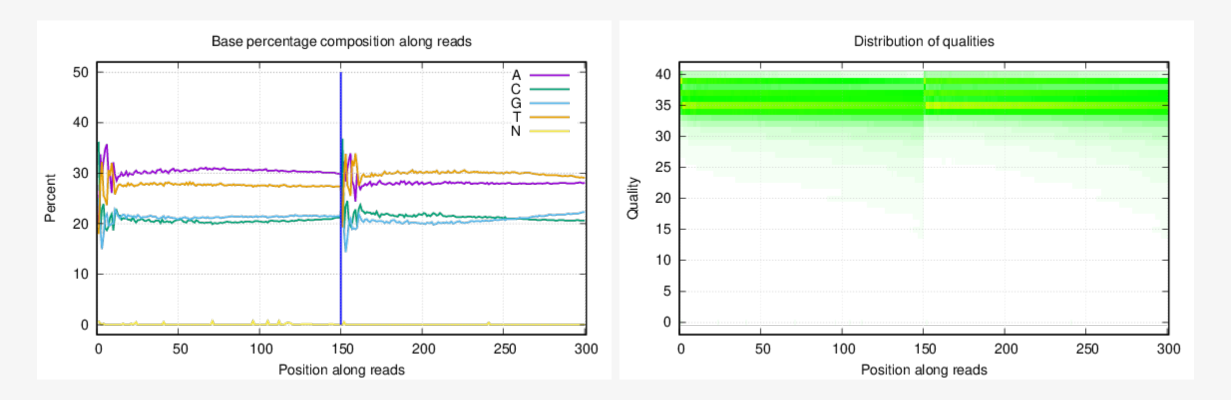
Ch_aux_R2



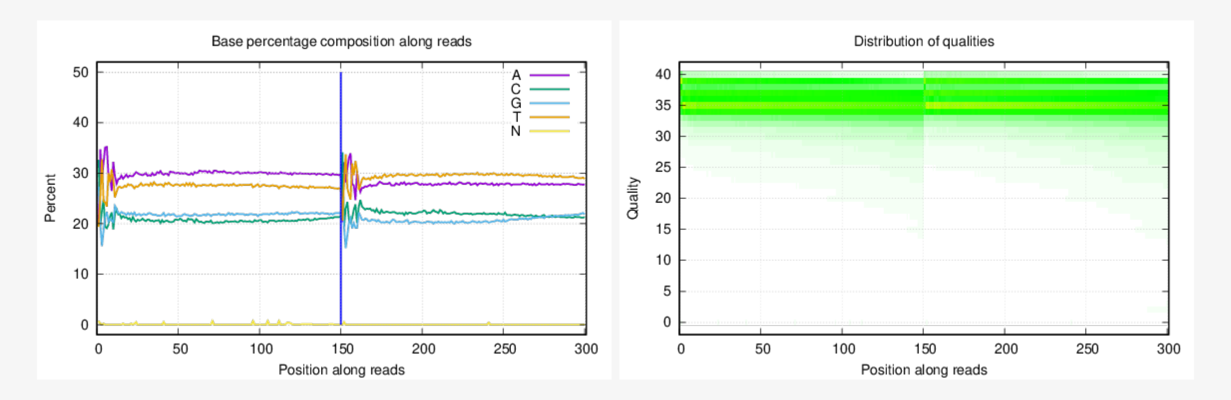
Dik_e



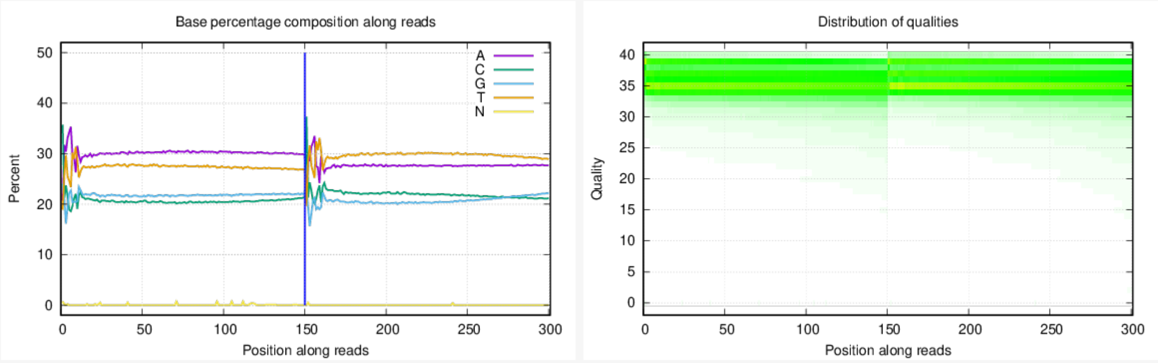
Dik_e1



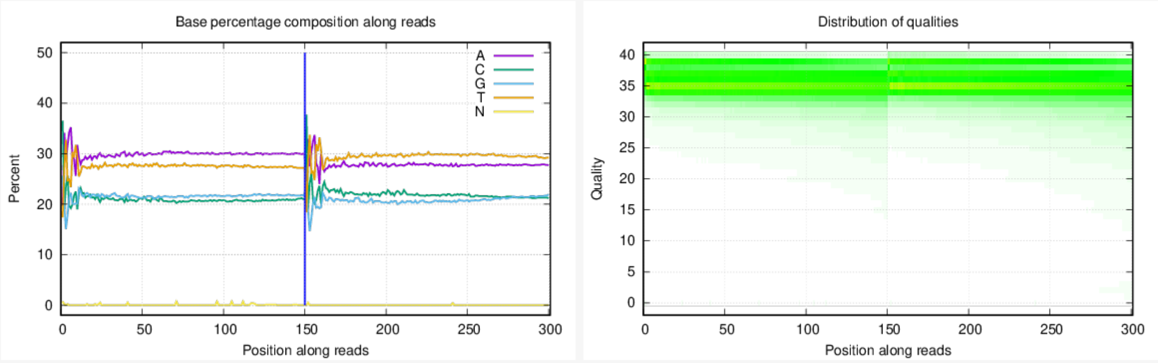
Fed1_e



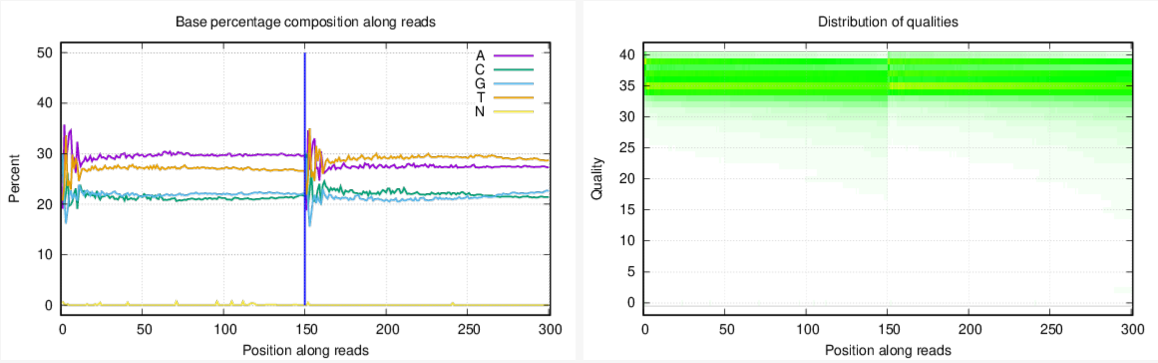
K11



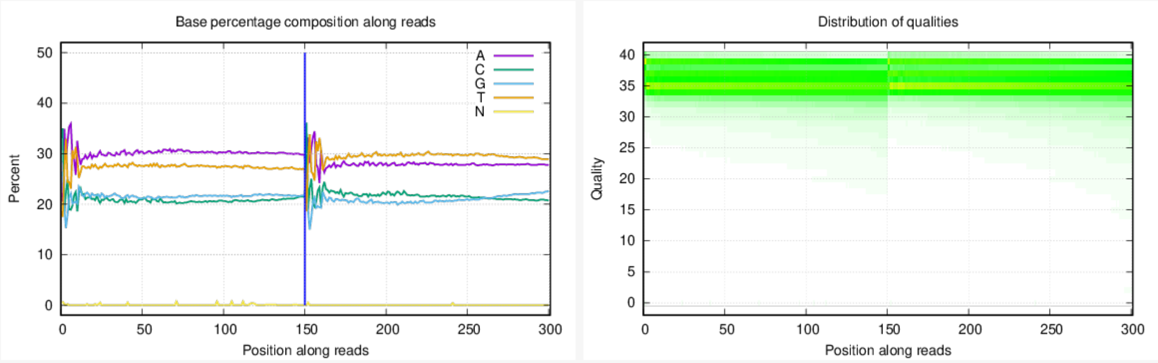
K8



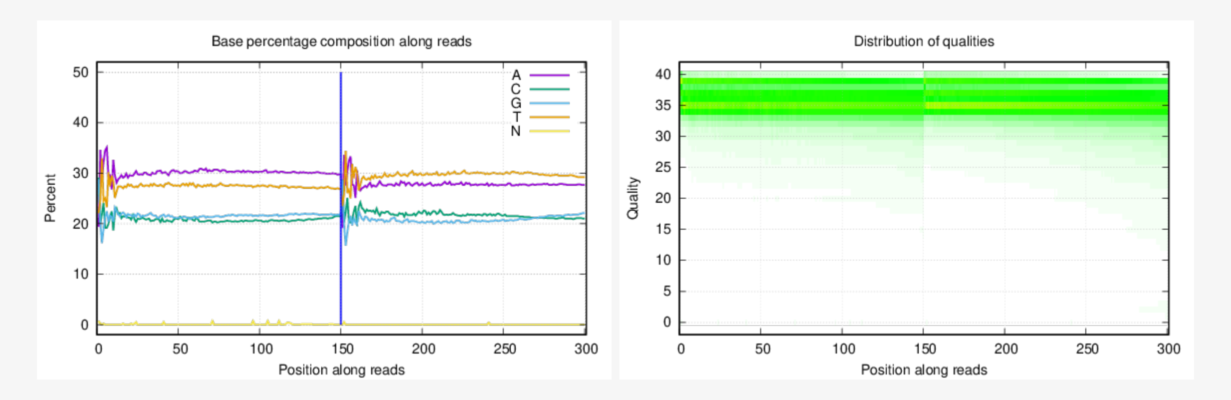
Sach1_K_2110



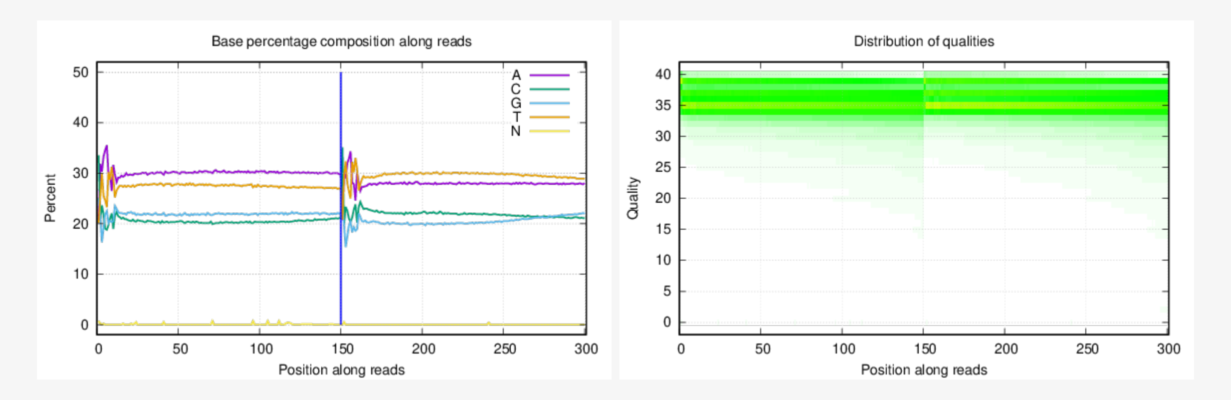
Say3_K



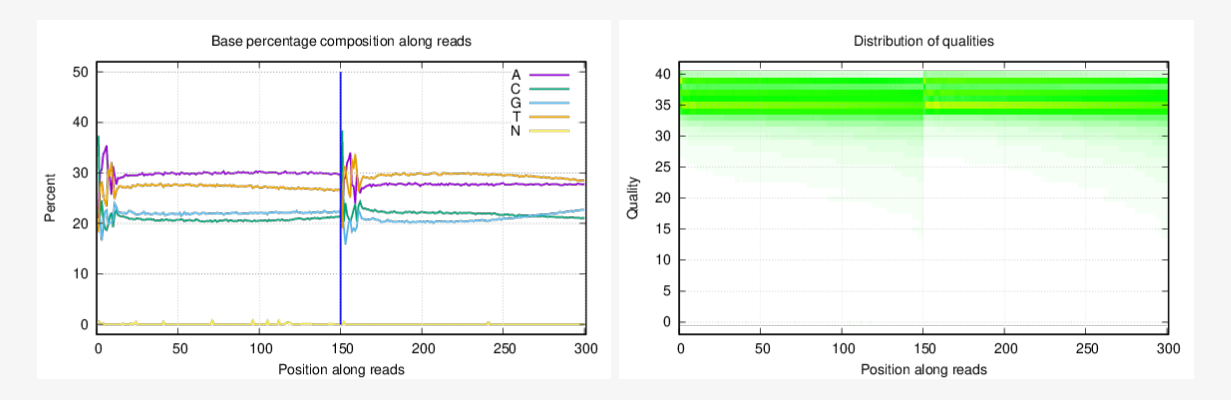
Shur3_K



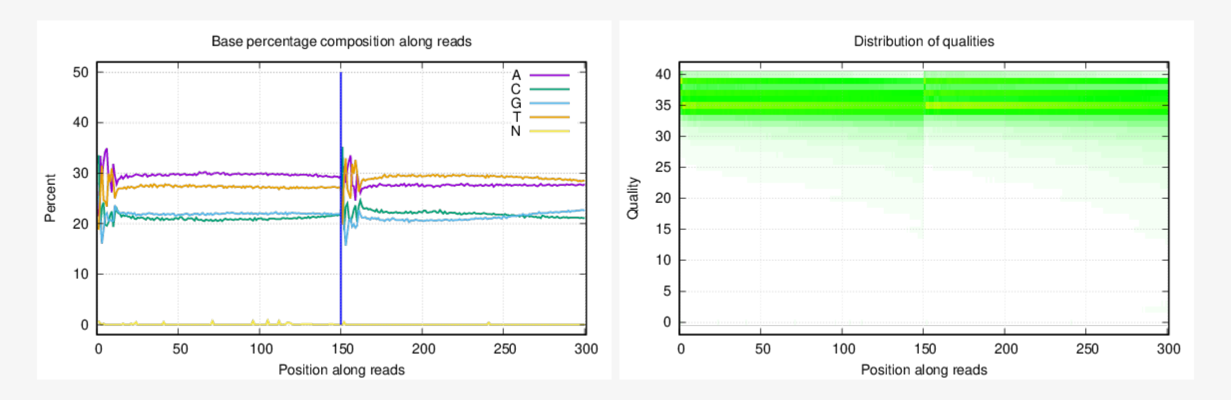
Vla1_K



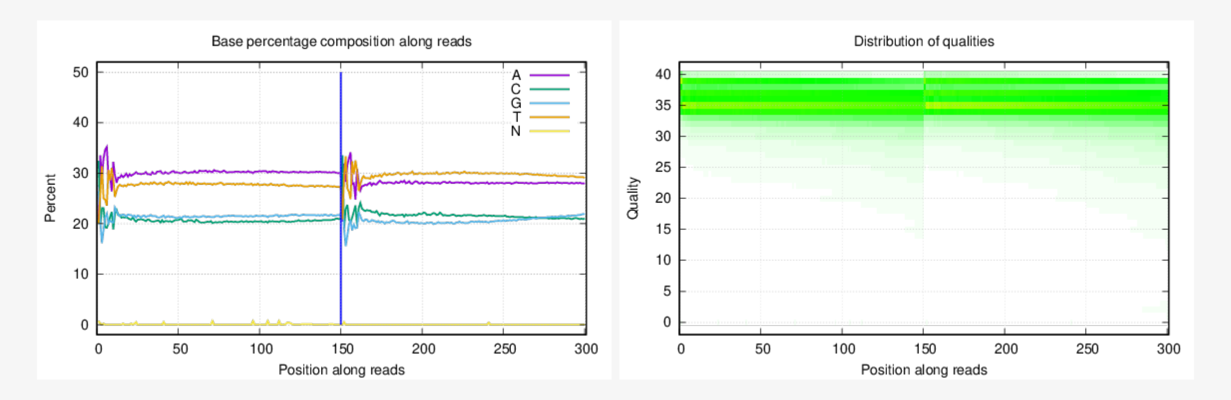
Vla2_K



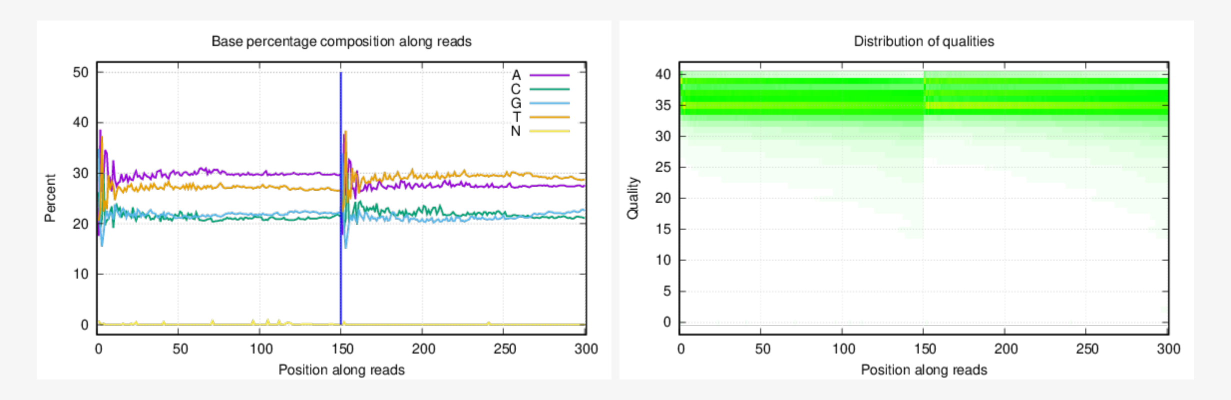
Vla6_K



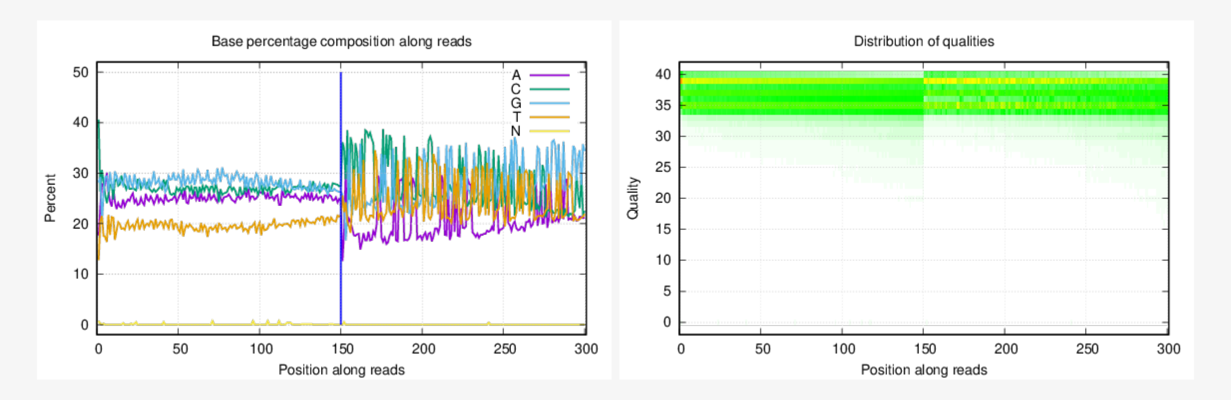
Vla9_e



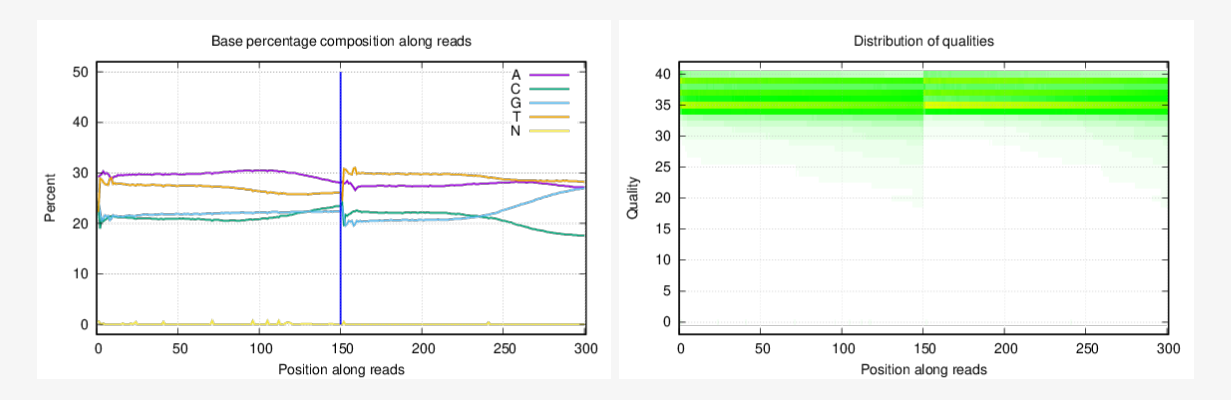
Vor1_K



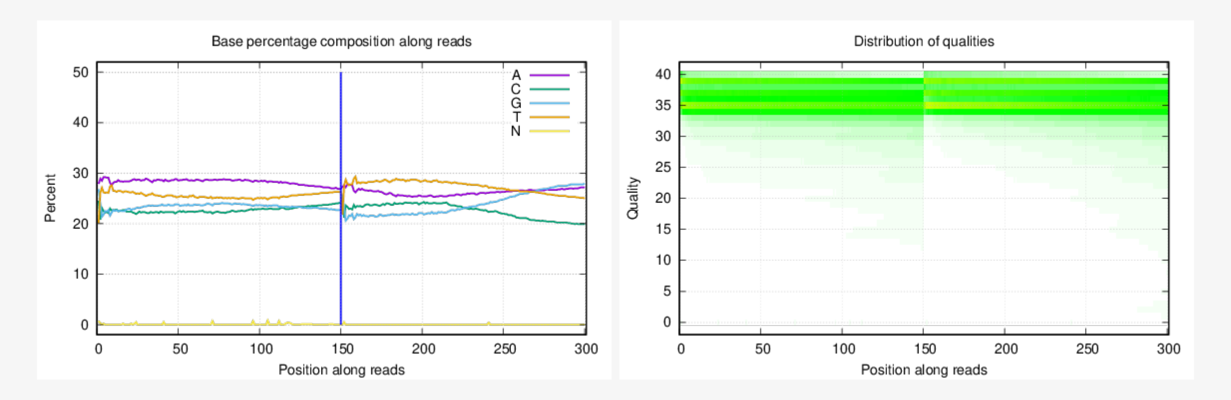
br_RAD21



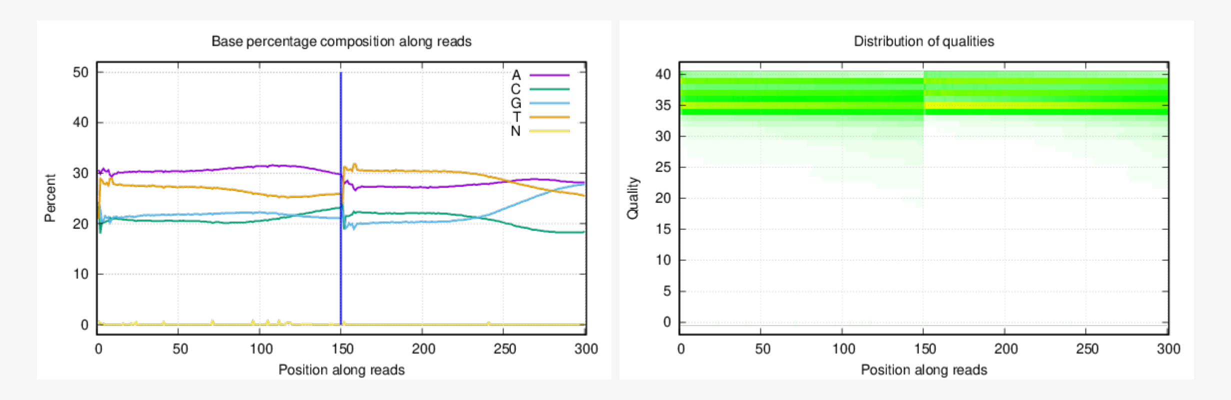
eb186_10d



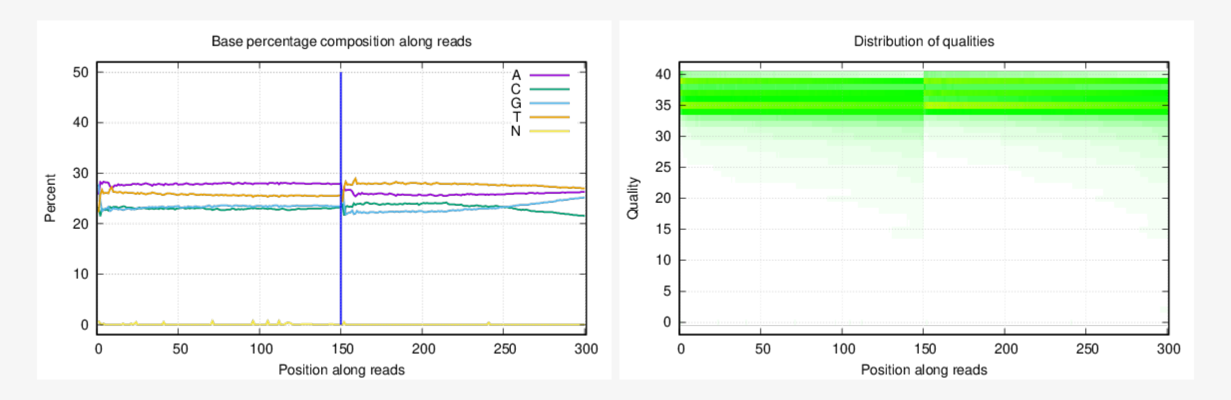
eb186_10d_i



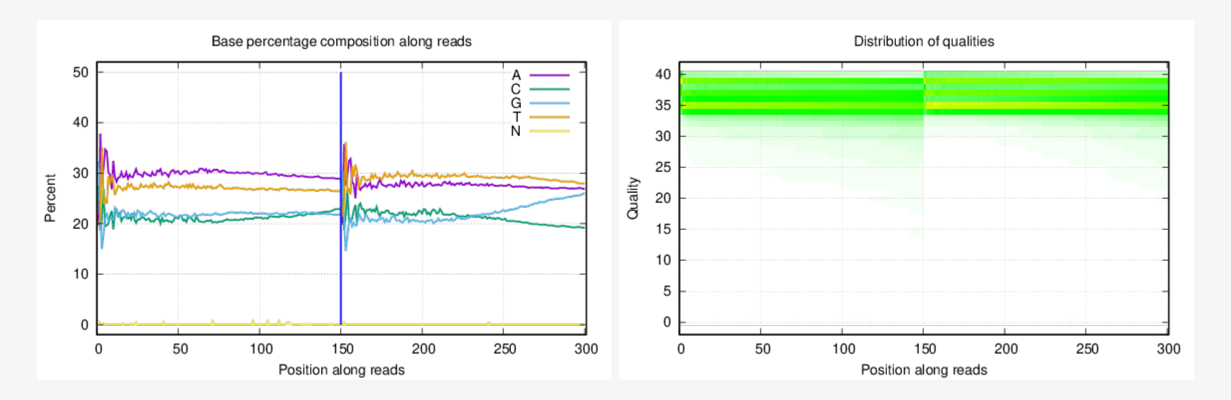
eb1_36_10d



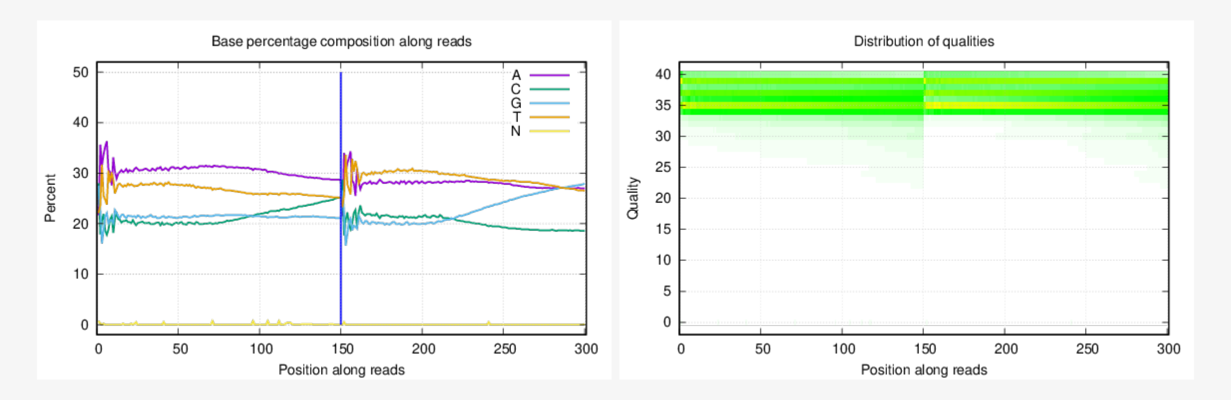
eb1_36_10d_i



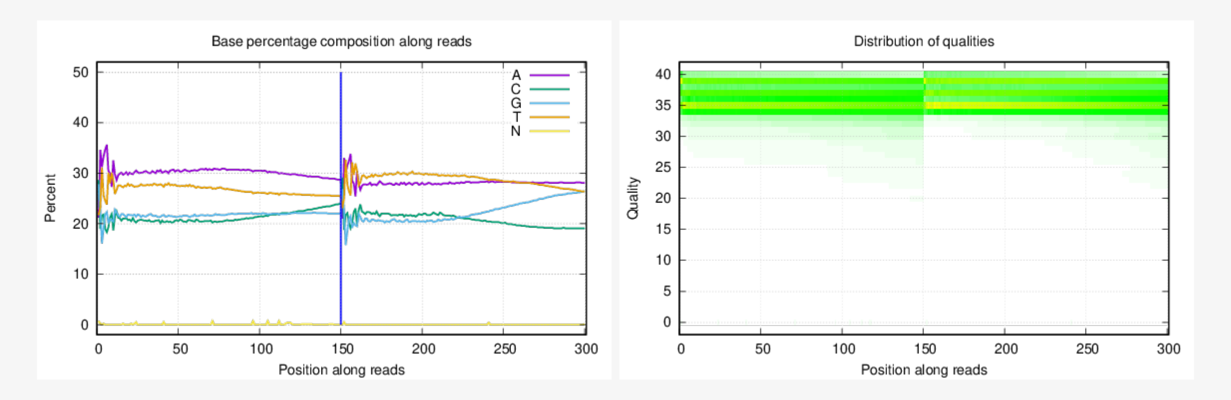
hsRad_R1



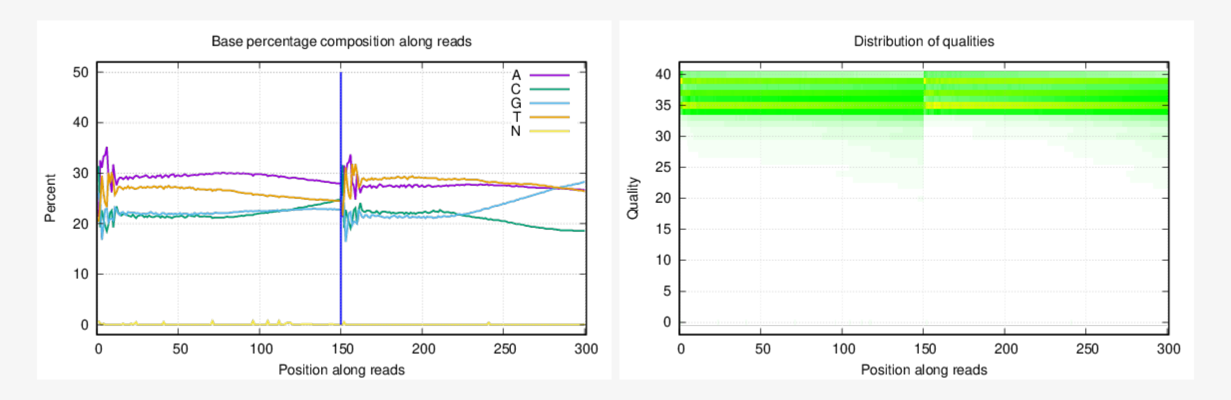
hsRad_R2



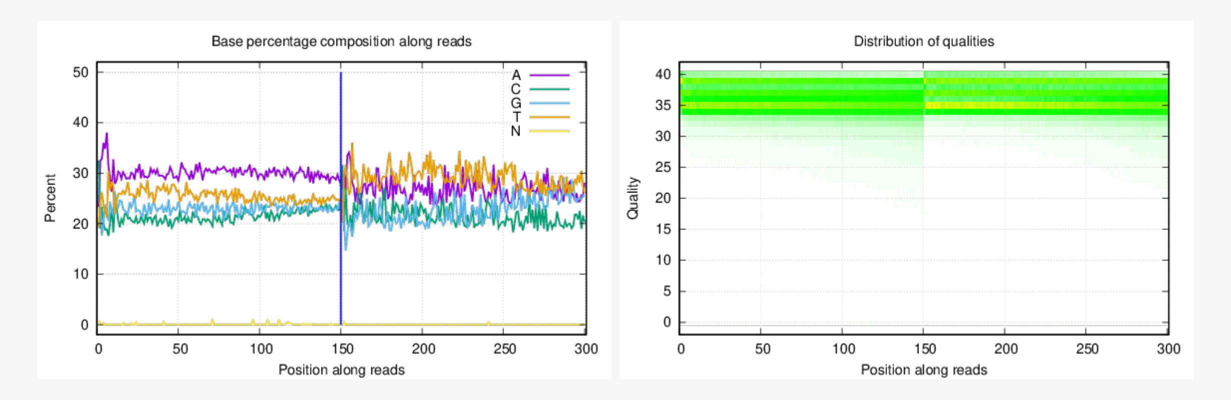
hsRad_aux_R1



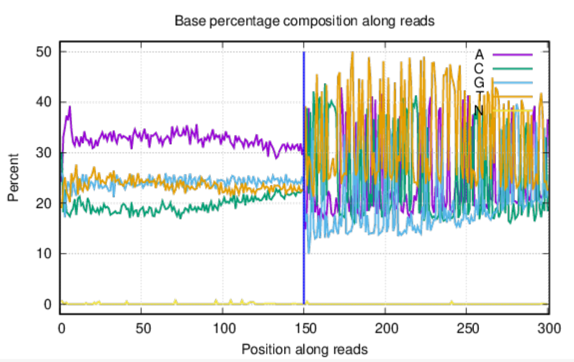
hsRad_aux_R2

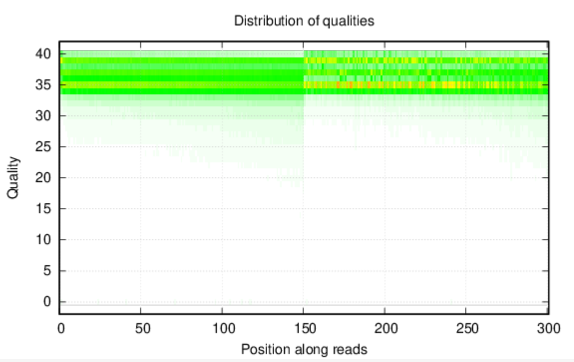


st_CTCF_L



st_CTCF_R


Base percentage composition along reads
Percent
Position along reads
0
10
20
30
40
50
0
50
100
150
200
250
300
A
C
G
T
N


Distribution of qualities
Quality
Position along reads
0
5
10
15
20
25
30
35
40
0
50
100
150
200
250
300

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	?

Omics For All

Contact us

Website: www.bgi.com

E-mail: info@bgi.com