Sequencing Data Report

Project: F20FTSEUHT0946-02_MUSvixyR
Date: 2024.4.9
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 59 samples in this project, the statistics of fastq data is shown below.

Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
1	150;150	100.00;0.00	98.67;0.00	47.67;46.00	2	600
11	150;150	97.84;98.85	93.15;95.84	42.33;43.45	18,275	5,482,500
12	150;150	98.30;98.74	94.63;95.56	54.53;53.59	14,337	4,301,100
33	150;150	98.05;98.68	94.02;96.01	45.18;45.29	37,127,234	11,138,170,200
36	150;150	97.47;98.17	92.15;94.41	43.90;43.79	29,220,414	8,766,124,200
40	150;150	97.16;97.96	91.23;93.69	44.40;44.28	25,790,735	7,737,220,500
4C_Ph-MEF-Pct-Bgl-1	150;150	96.74;97.17	89.86;90.59	43.73;44.43	1,426,010	427,803,000
4C_Ph-MEF-Pct-Bgl-2	150;150	97.03;97.84	90.74;92.73	43.76;44.39	1,066,622	319,986,600
7369_Nod2_2	150;150	97.66;98.17	92.95;94.02	56.22;55.93	1,213,020	363,906,000
8	150;150	99.33;0.00	98.00;0.00	59.33;52.67	1	300
Boc_K4	150;150	97.05;97.28	90.85;90.88	42.80;42.78	2,996,247	898,874,100
Boc_e4	150;150	96.88;96.41	90.20;88.03	42.46;42.52	3,552,525	1,065,757,500
CS1	150;150	98.12;98.48	94.11;94.99	46.41;46.41	123,089	36,926,700
CS2	150;150	98.12;98.90	94.18;96.34	45.02;45.05	353,085	105,925,500
CS3	150;150	96.56;94.66	88.92;81.32	40.52;40.06	3,218,621	965,586,300
CS4	150;150	96.52;94.32	88.71;80.09	40.27;39.88	1,325,714	397,714,200
CS5	150;150	92.99;94.60	78.98;80.82	40.19;39.99	17,801	5,340,300
CS6	150;150	93.33;94.18	79.93;79.63	40.12;40.08	1,891	567,300
Casian	150;150	98.12;99.07	94.09;96.87	42.00;42.04	292,070	87,621,000
Cherry_R3	150;150	97.69;98.68	92.87;95.62	46.16;46.12	31,206,317	9,361,895,100
Cherry_a_R3	150;150	98.08;98.69	93.93;95.58	44.63;44.62	48,414,954	14,524,486,200
F1613_19	150;150	98.24;99.07	94.49;96.91	42.77;42.79	224,472	67,341,600
G1250_15	150;150	98.00;99.08	93.87;96.98	44.02;44.03	214,131	64,239,300
G1359_20	150;150	98.46;99.03	95.15;96.88	44.87;44.39	278,840	83,652,000
G2073_21	150;150	98.38;99.19	94.95;97.34	43.93;44.22	906,018	271,805,400
G278_20	150;150	98.16;99.08	94.32;96.96	42.96;42.74	101,899	30,569,700

Gri2_K	150;150	97.67;98.03	92.73;93.39	48.69;48.67	49,293	14,787,900
Kra1_e	150;150	96.73;97.60	90.06;92.00	43.49;43.53	107,049	32,114,700
Kra_e	150;150	97.28;97.57	91.52;91.95	43.28;43.39	104,582	31,374,600
NOD_15	150;150	97.54;97.39	92.25;91.07	50.87;51.02	91,485	27,445,500
NOD_7	150;150	97.21;97.87	91.27;92.45	47.49;47.66	137,149	41,144,700
NOD_8	150;150	97.68;97.14	92.47;90.35	52.03;51.40	140,180	42,054,000
Prime_contr	150;150	95.99;97.52	87.98;91.62	43.02;42.78	192,889	57,866,700
S2_719_R1	150;150	97.13;97.36	91.11;91.10	43.81;43.66	133,385	40,015,500
S2_719_R2	150;150	95.16;95.86	85.71;86.09	43.92;43.88	17,741	5,322,300
S2_R1	150;150	96.95;97.05	90.57;90.06	44.22;44.18	46,095	13,828,500
S2_R2	150;150	96.83;96.89	90.21;89.58	43.99;43.96	76,122	22,836,600
Shen1_K	150;150	97.29;97.47	91.53;91.58	43.44;43.51	3,955,641	1,186,692,300
Shen2_e	150;150	97.21;97.49	91.31;91.65	43.27;43.30	115,732	34,719,600
Shen3_K	150;150	97.53;97.75	92.22;92.52	43.58;43.66	60,089	18,026,700
Shen3_e	150;150	97.50;97.83	92.24;92.88	44.93;44.87	75,390	22,617,000
Zap_K2	150;150	91.37;89.76	77.44;73.51	47.11;47.07	99	29,700
Zap_e3	150;150	98.25;98.02	94.41;94.52	42.79;43.10	116	34,800
brCTCF_R1	150;150	97.97;98.49	93.59;94.94	43.63;43.57	394,062	118,218,600
brCTCF_a_R1	150;150	97.82;98.54	93.15;95.07	43.55;43.56	382,245	114,673,500
dmCTCF_R1	150;150	98.25;98.97	94.51;96.55	45.12;45.13	376,362	112,908,600
dmCTCF_a_R1	150;150	98.14;98.84	94.12;96.08	44.68;44.67	340,672	102,201,600
ggCTCF_R2	150;150	97.80;98.46	93.08;94.83	45.53;45.49	38,178,646	11,453,593,800
ggCTCF_a_R1	150;150	97.79;98.72	93.06;95.66	45.12;45.09	35,098,537	10,529,561,100
ggCTCF_a_R2	150;150	97.43;98.60	92.04;95.23	44.05;43.98	33,574,651	10,072,395,300
hCTCF_a_R3	150;150	97.74;98.41	92.89;94.64	44.89;44.83	21,467,865	6,440,359,500
iCS1	150;150	97.71;98.82	92.94;95.98	42.01;42.04	460	138,000
iCS2	150;150	96.85;98.09	90.94;93.92	45.86;45.65	303	90,900
iCS3	150;150	96.30;94.46	88.40;80.32	47.66;47.21	1,563,900	469,170,000
iCS4	150;150	96.93;95.08	90.39;82.34	53.67;52.94	2,011,089	603,326,700
mmCTCF_R1	150;150	98.04;98.72	93.74;95.65	44.89;44.87	38,295,918	11,488,775,400
mmCTCF_R2	150;150	97.60;98.58	92.47;95.14	43.46;43.44	25,891,621	7,767,486,300
mmCTCF_a_R1	150;150	97.86;98.56	93.23;95.09	44.16;44.09	35,202,979	10,560,893,700

mmCTCF_a_R2	150;150	97.82;98.65	93.19;95.44	46.41;46.34	35,750,841	10,725,252,300

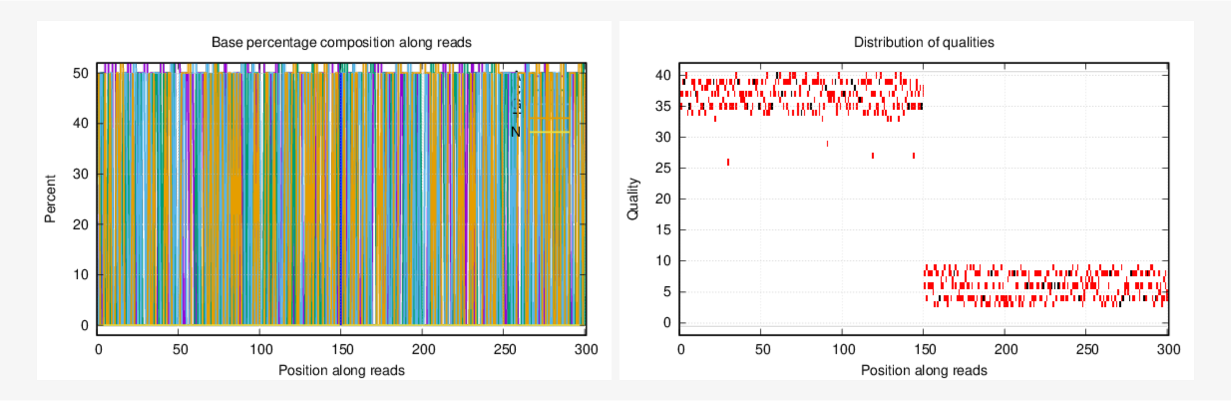
Table Format:

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

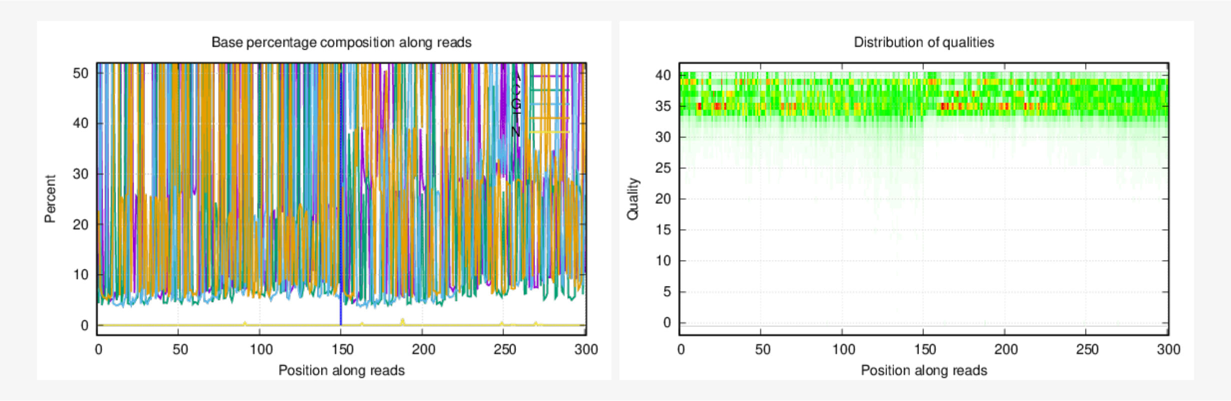
Data Quality Control

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.

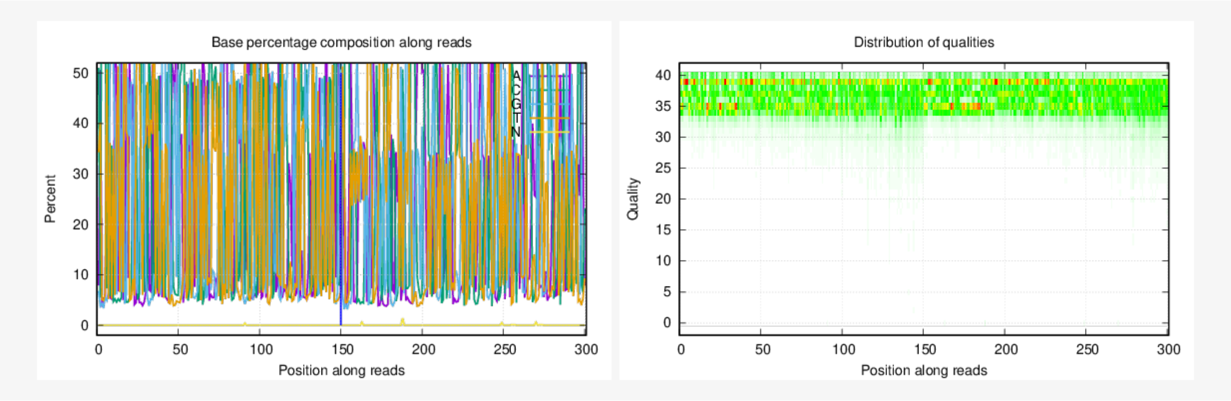
1



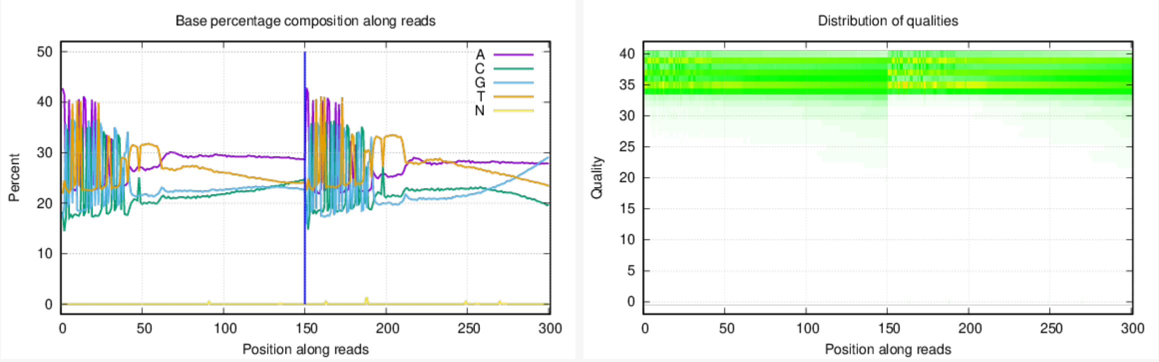
11



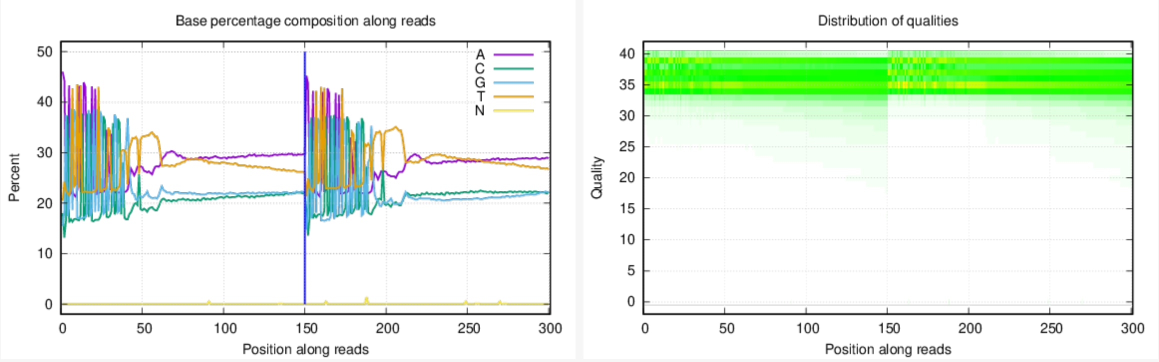
12



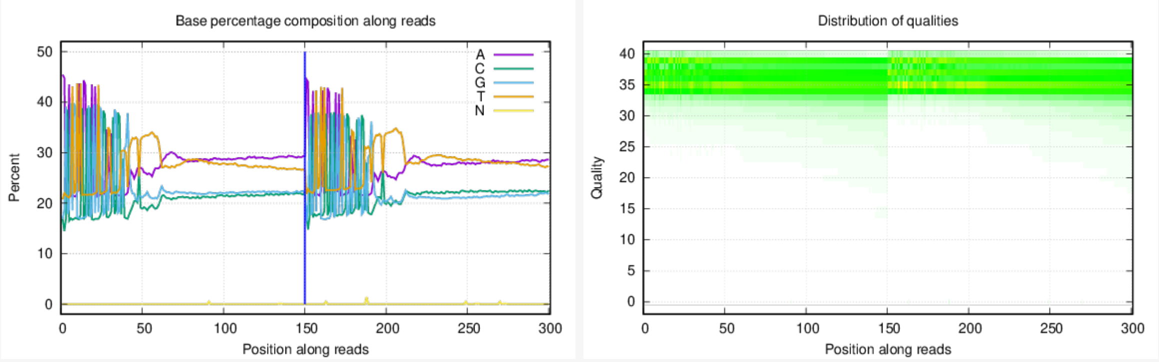
33



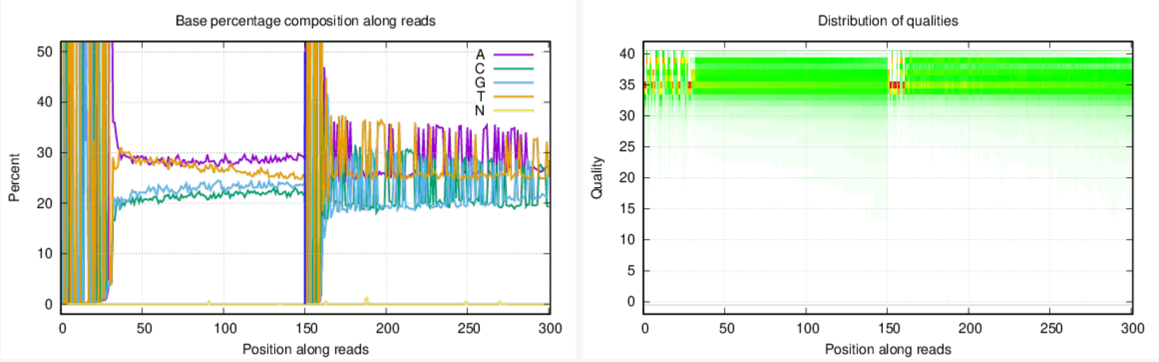
36



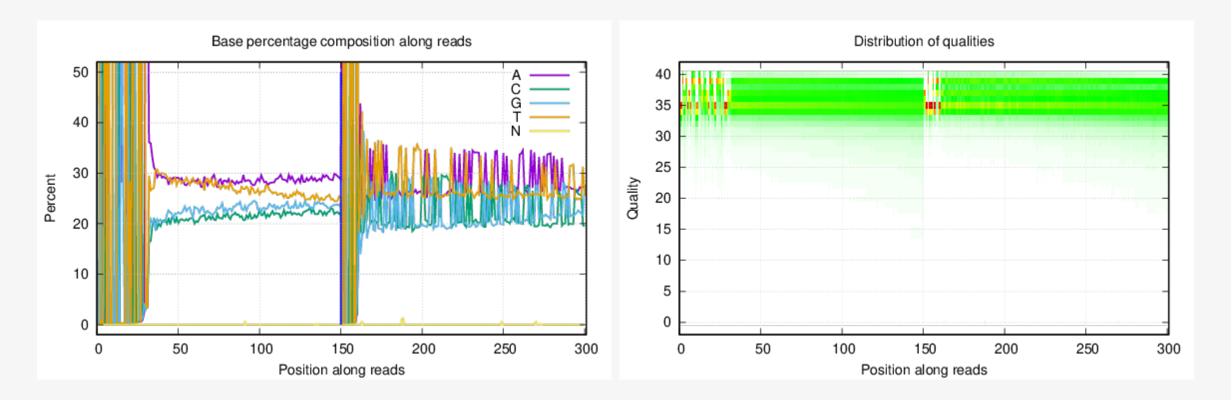
40



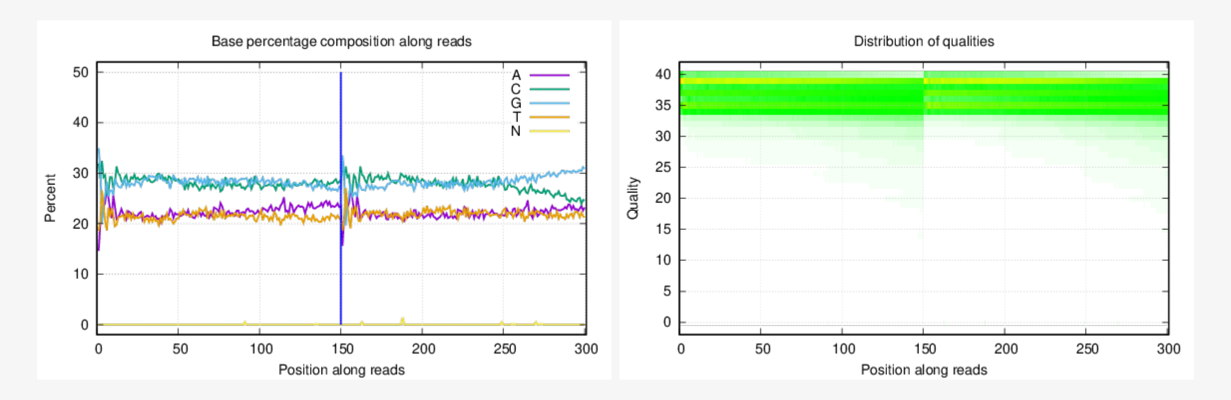
4C_Ph-MEF-Pct-Bgl-1



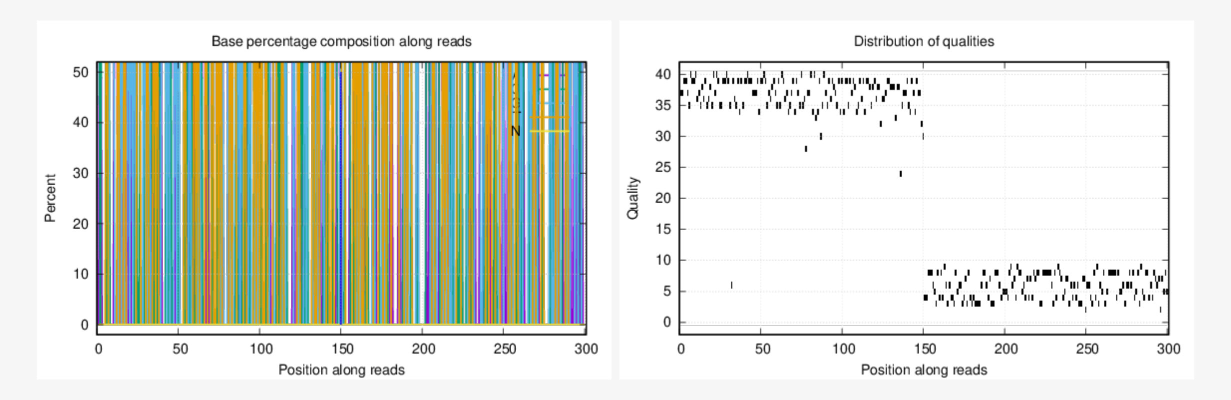
4C_Ph-MEF-Pct-Bgl-2



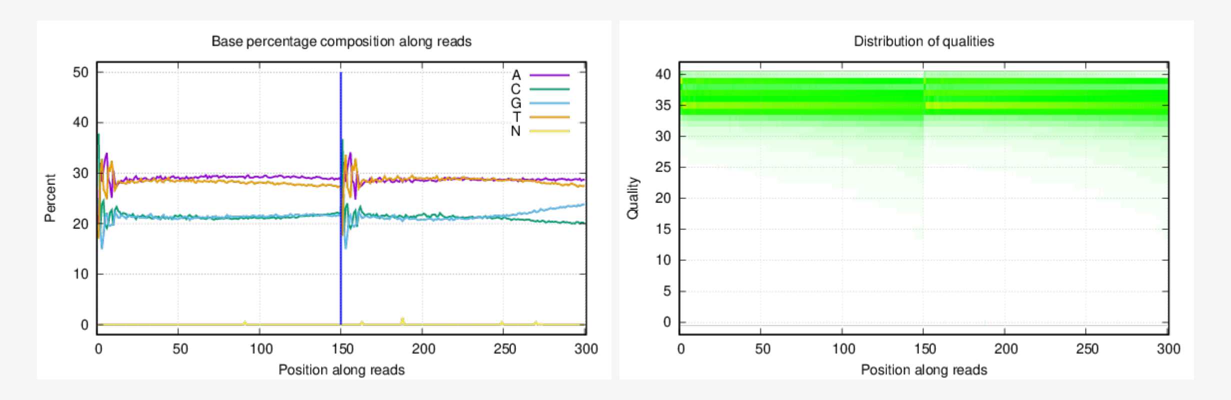
7369_Nod2_2



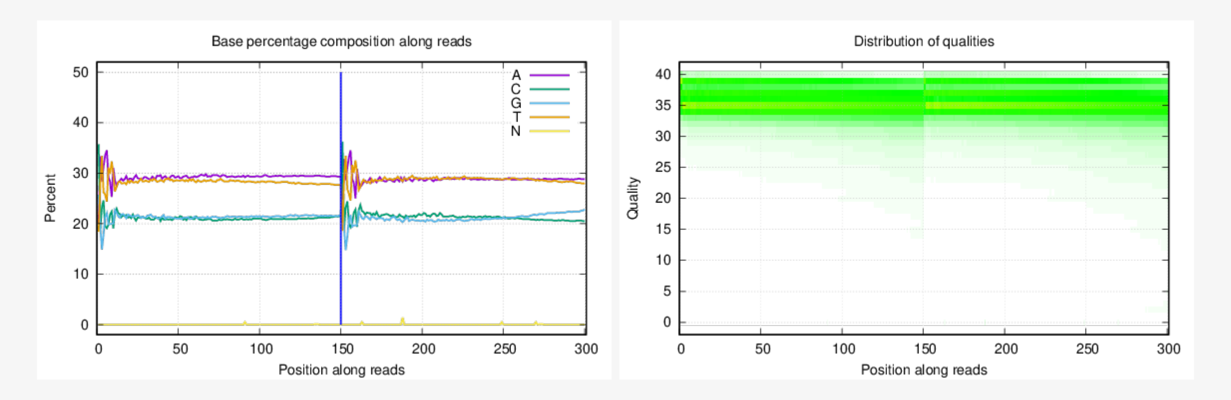
8



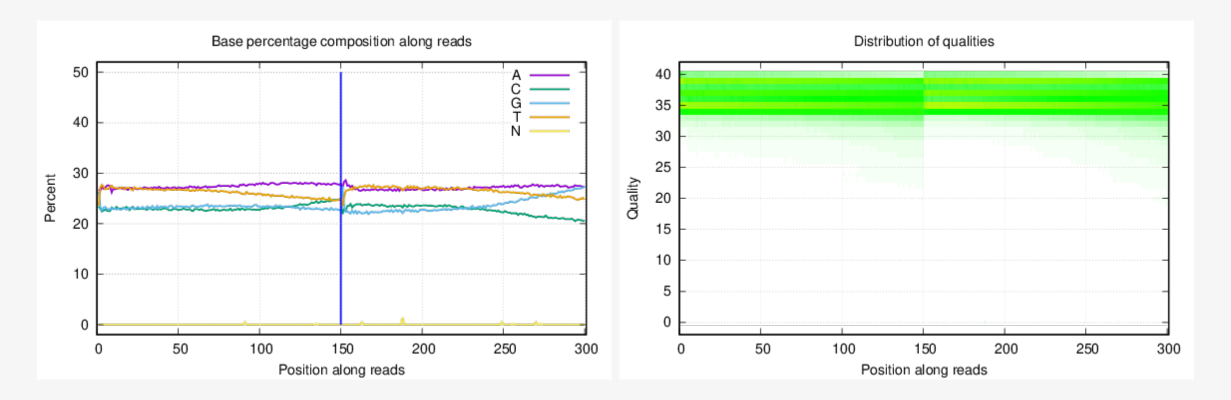
Boc_K4



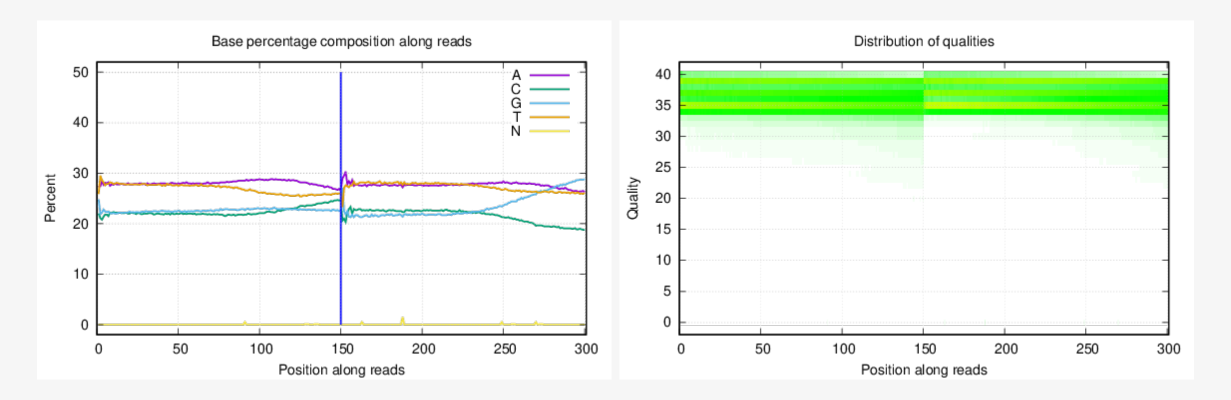
Boc_e4



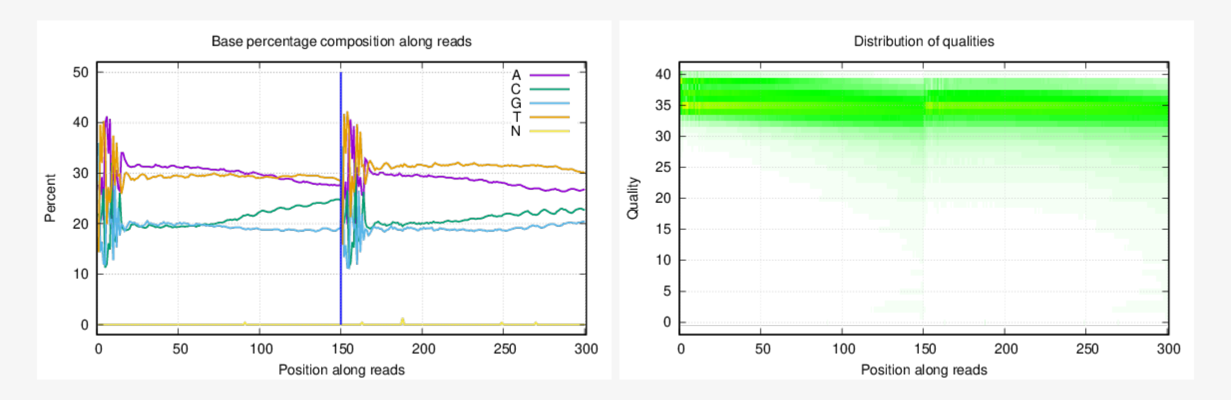
CS1



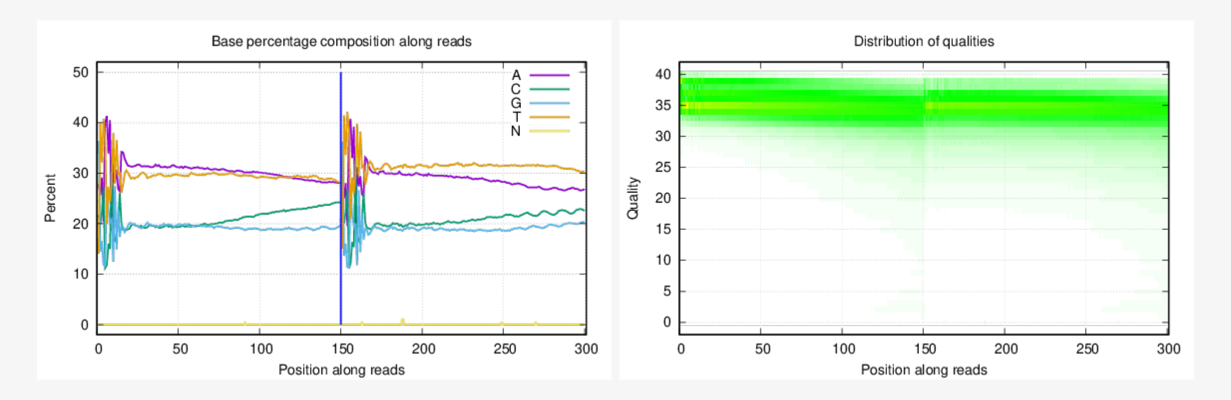
CS2



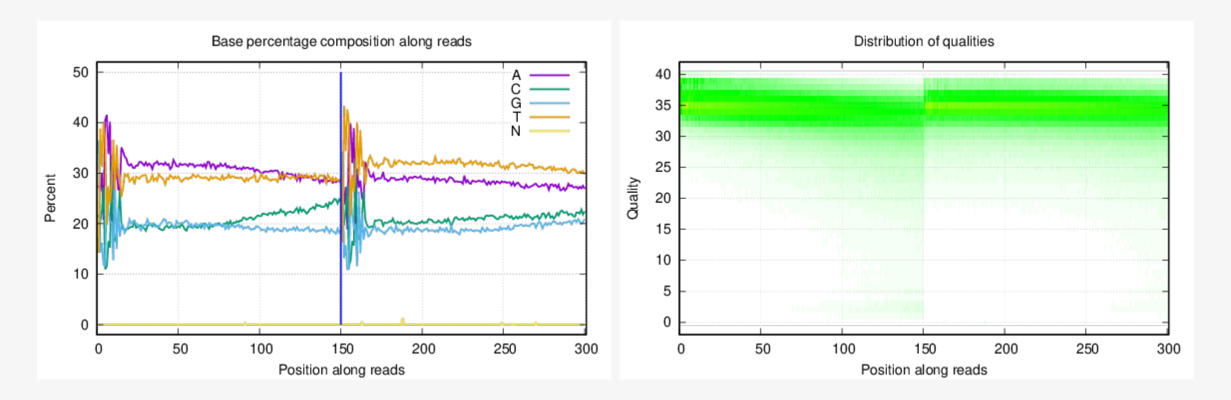
CS3



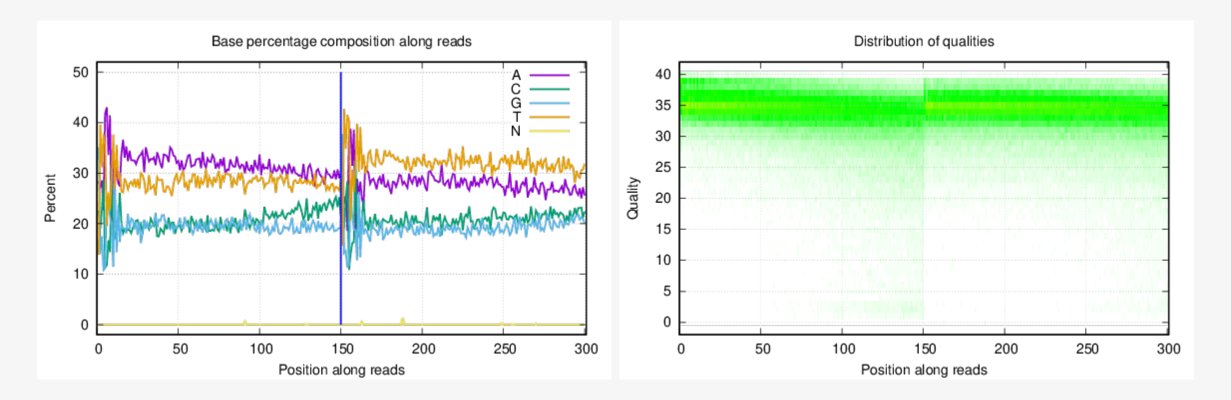
CS4



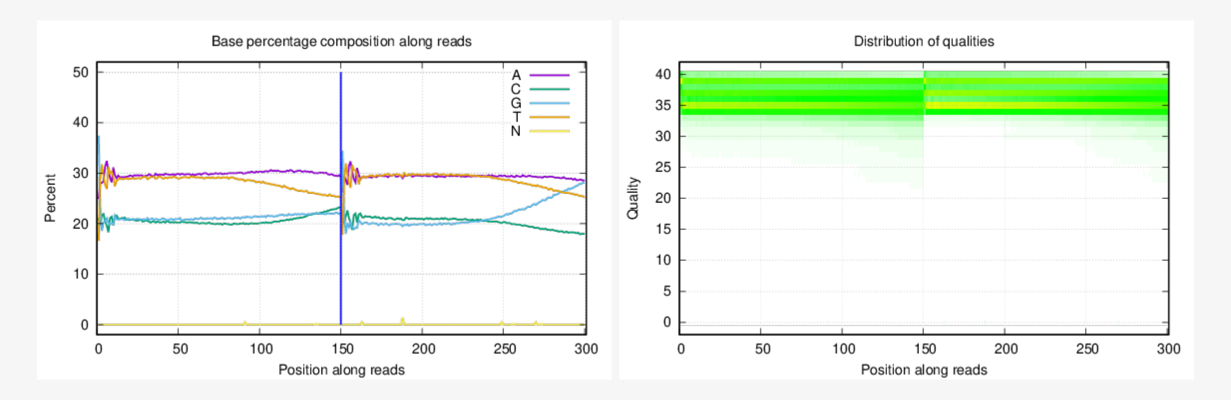
CS5



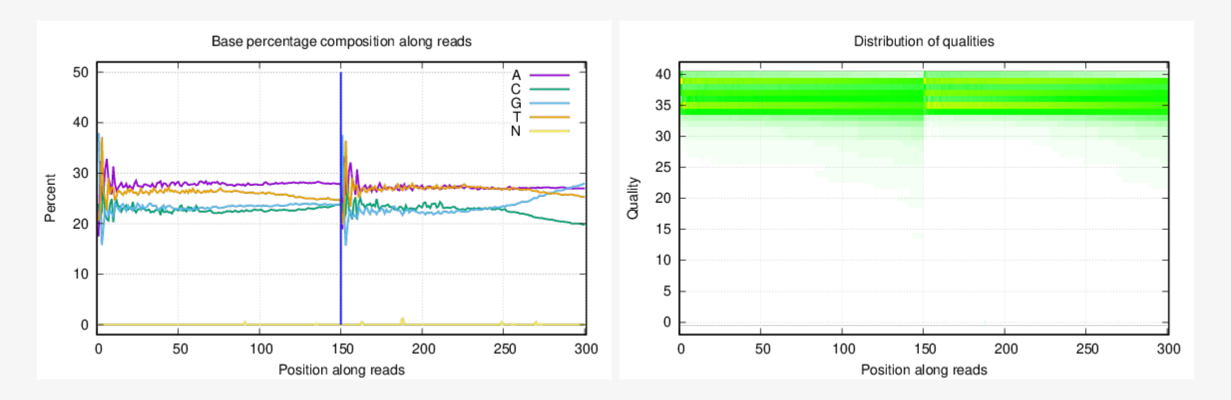
CS6



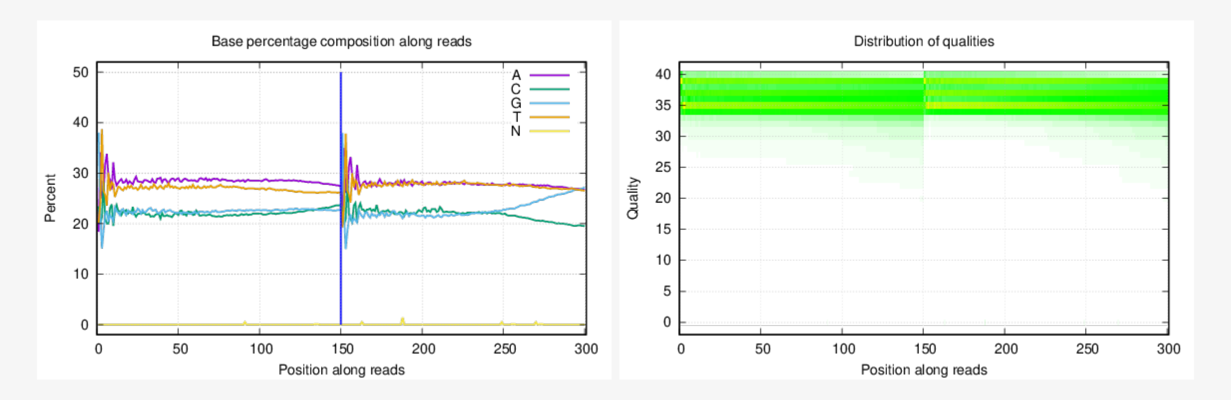
Casian



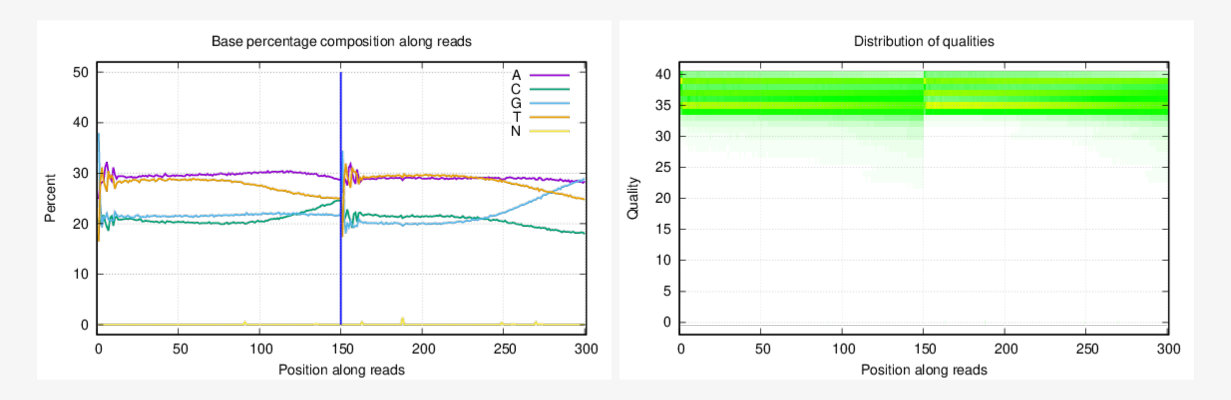
Cherry_R3



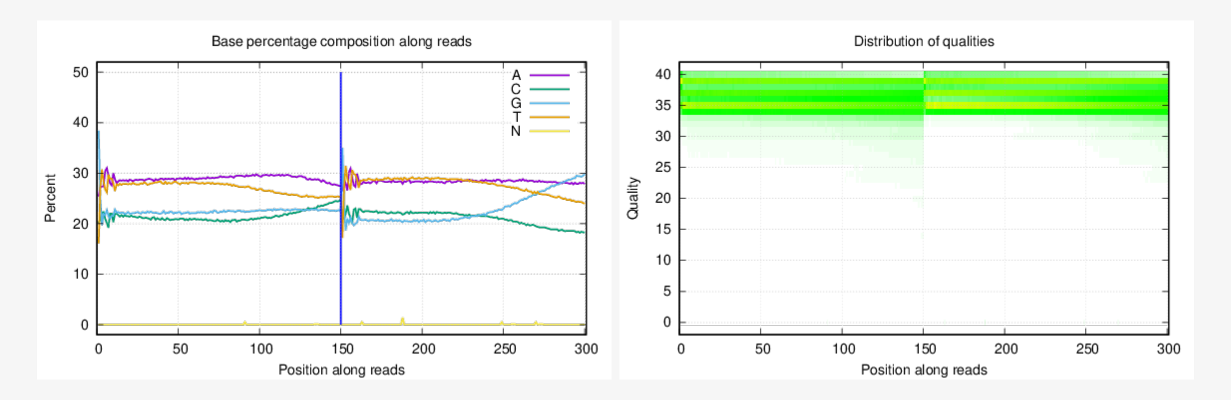
Cherry_a_R3



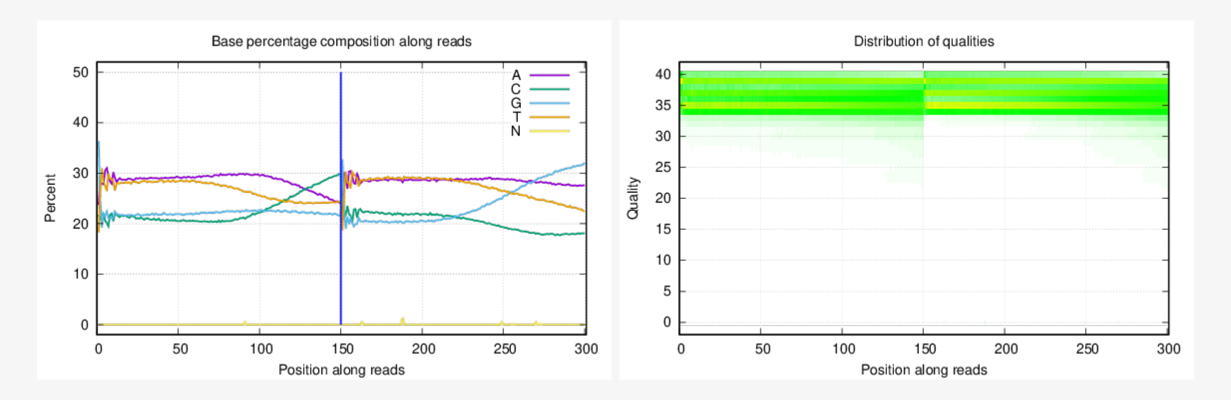
F1613_19



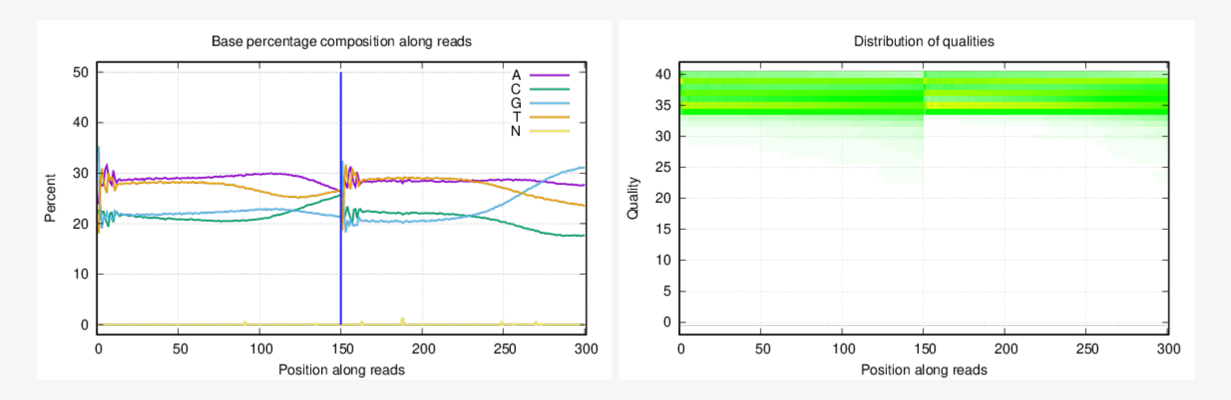
G1250_15



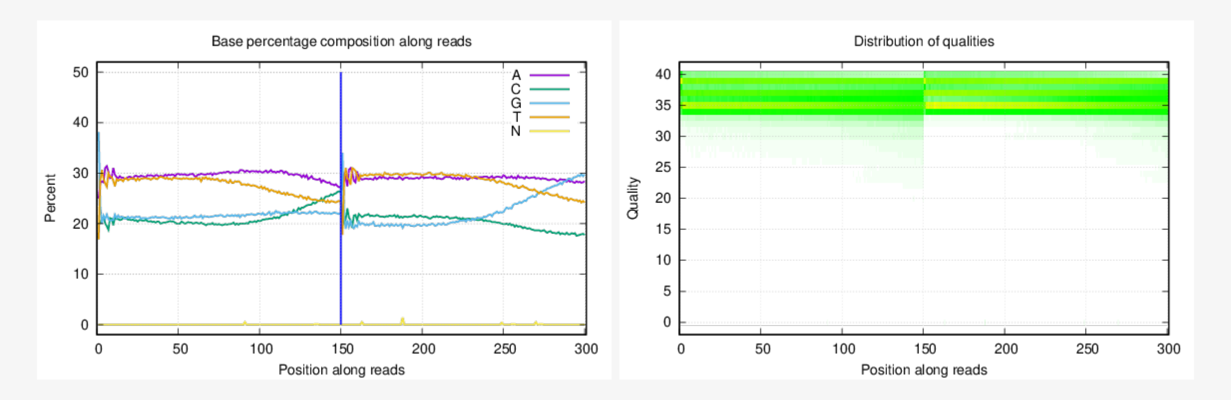
G1359_20



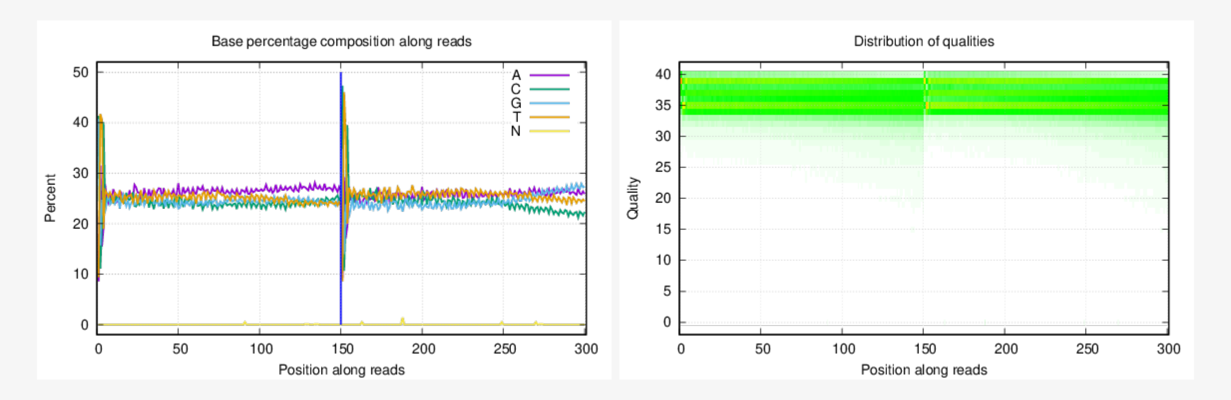
G2073_21



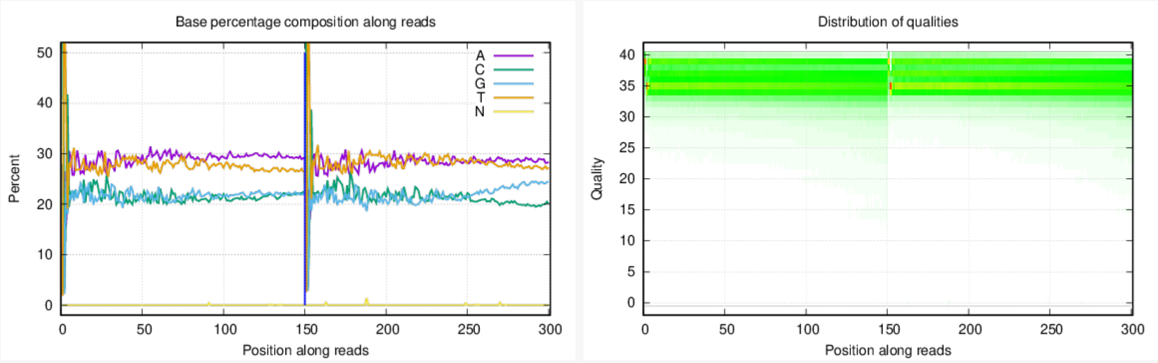
G278_20



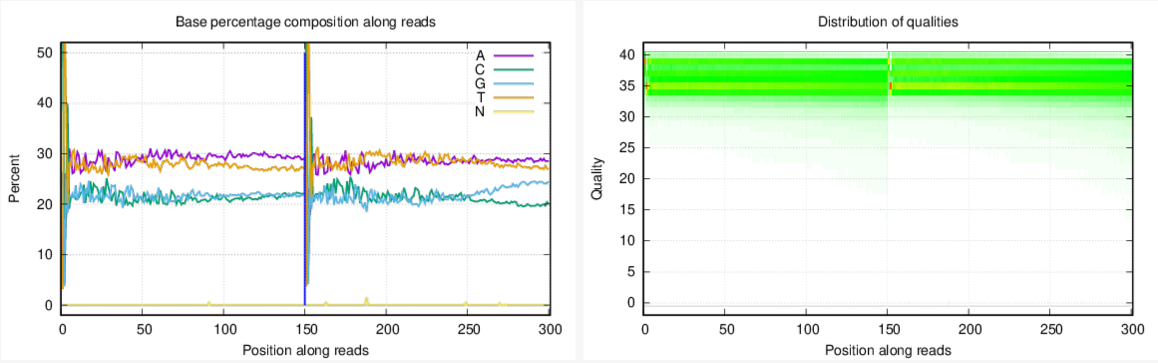
Gri2_K



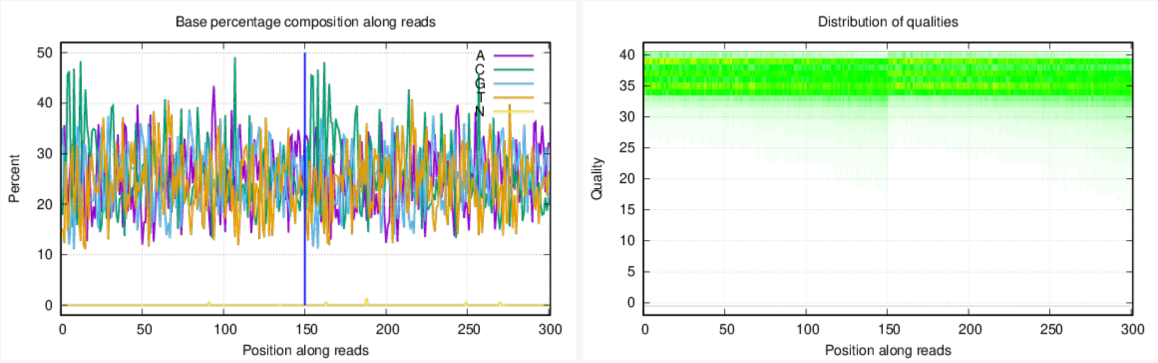
Kra1_e



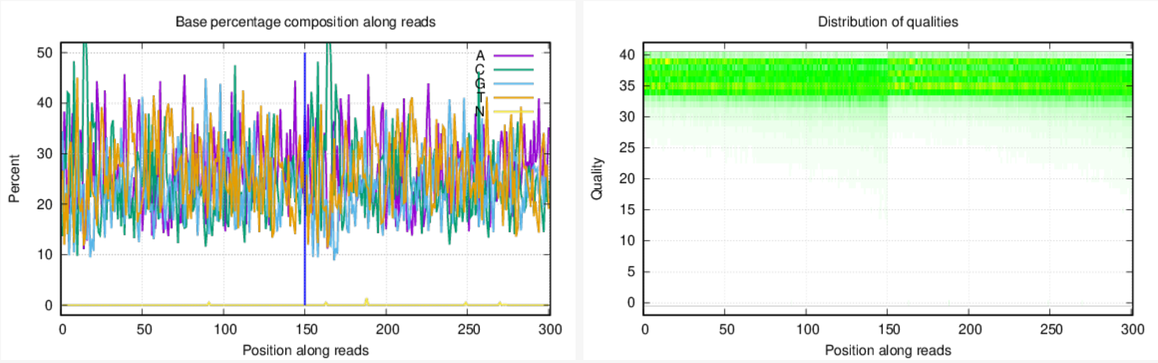
Kra_e



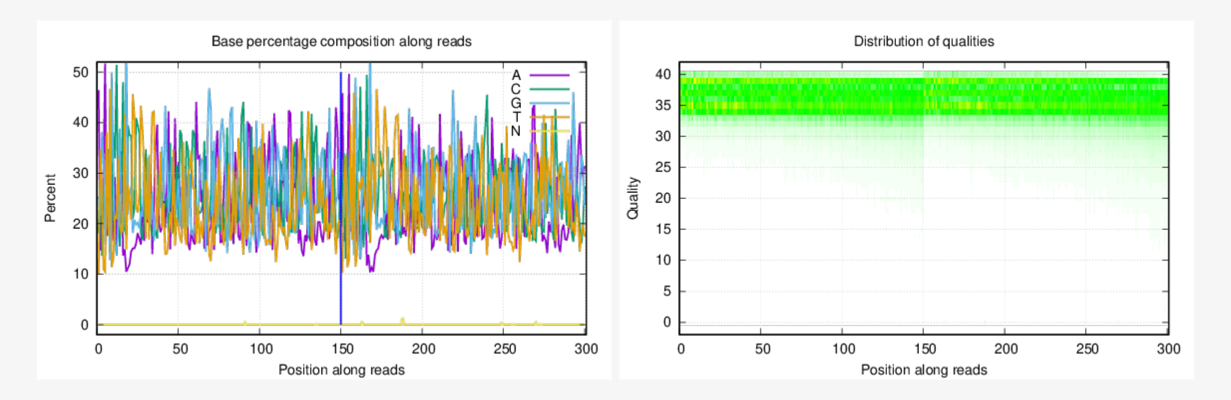
NOD_15



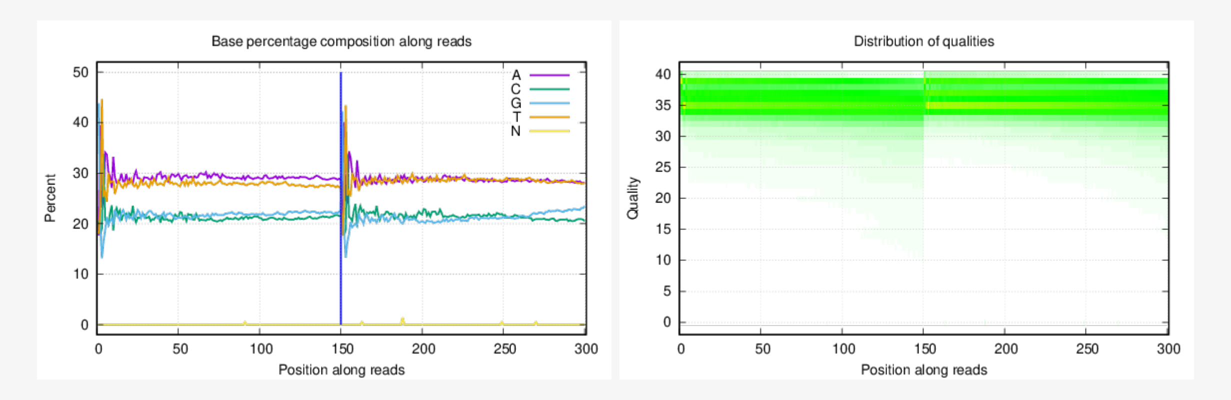
NOD_7



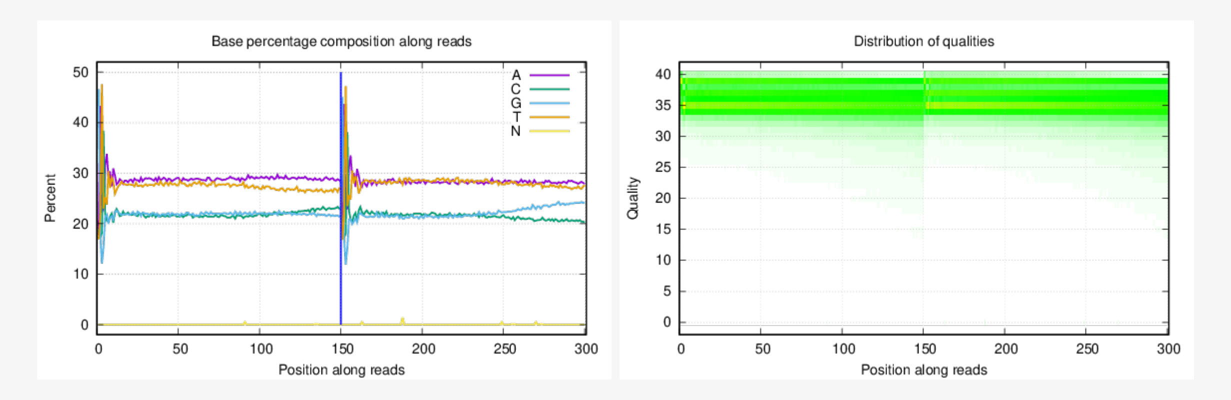
NOD_8



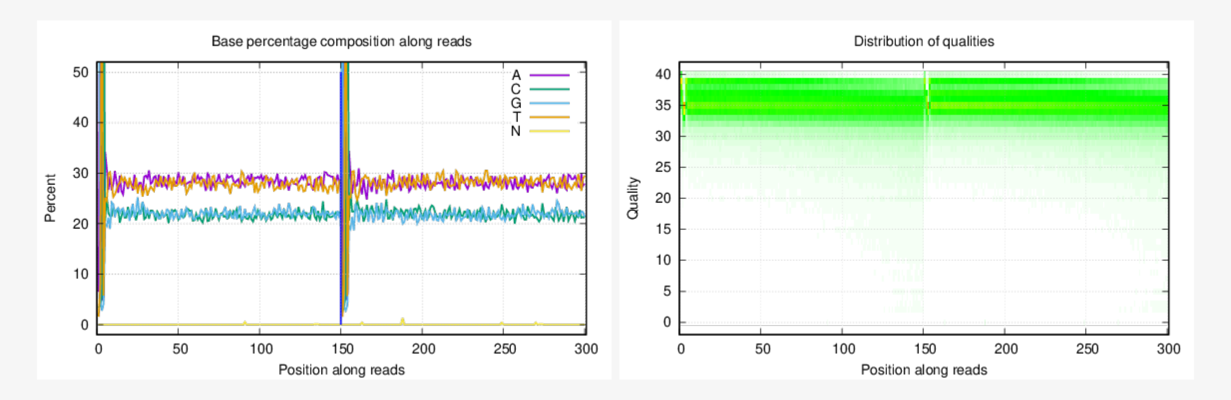
Prime_contr



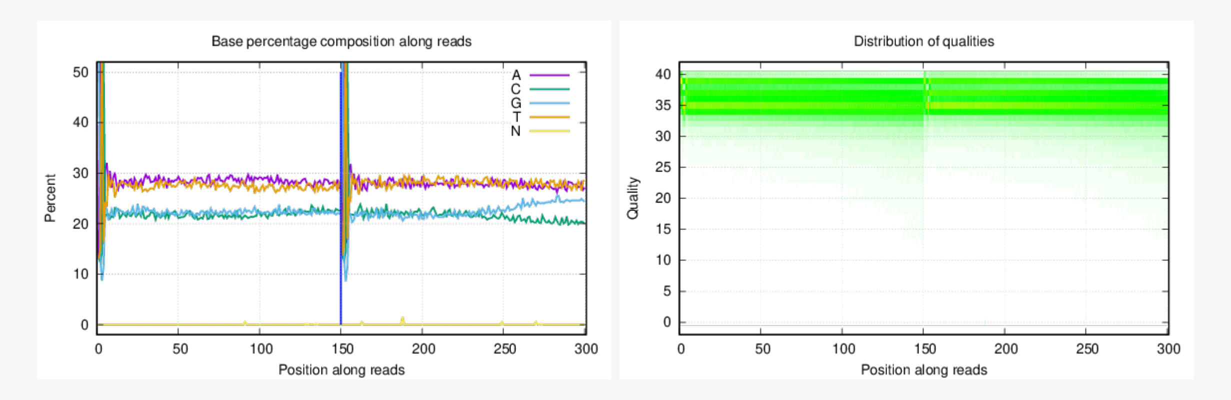
S2_719_R1



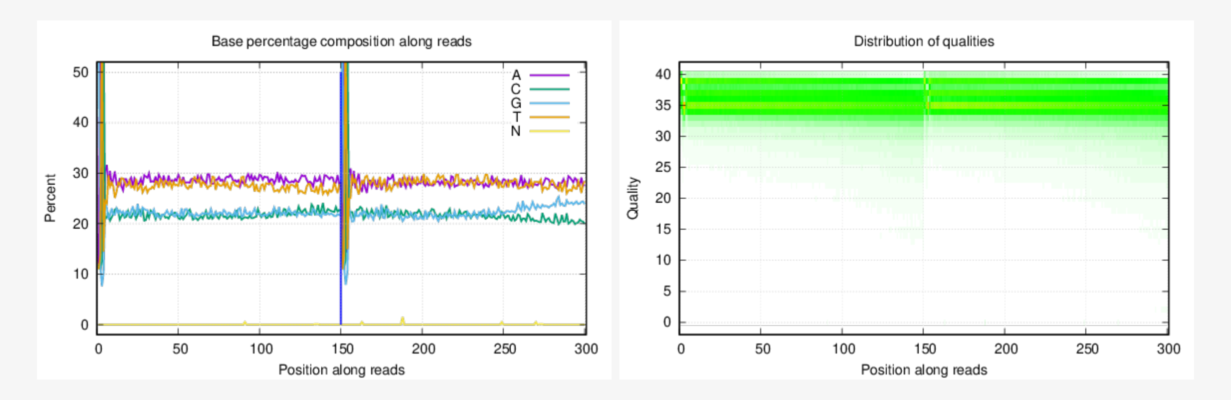
S2_719_R2



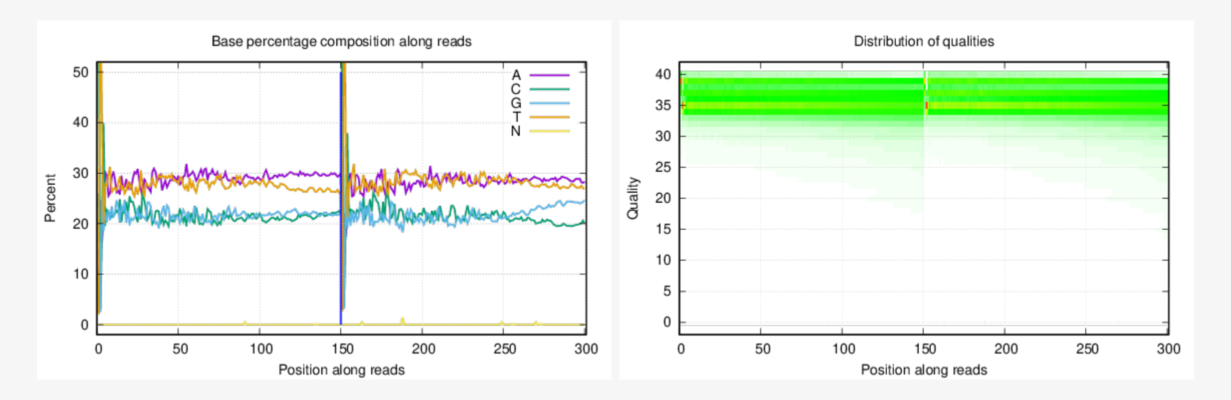
S2_R1



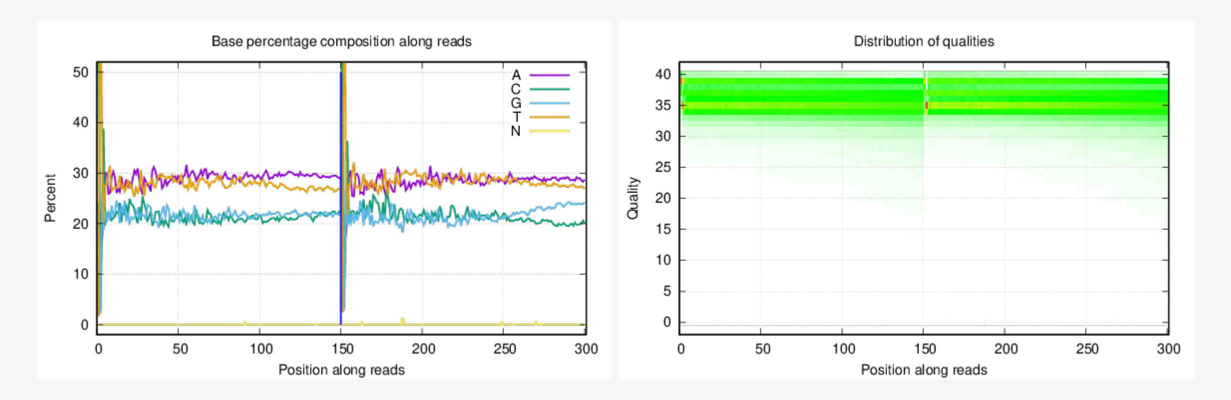
S2_R2



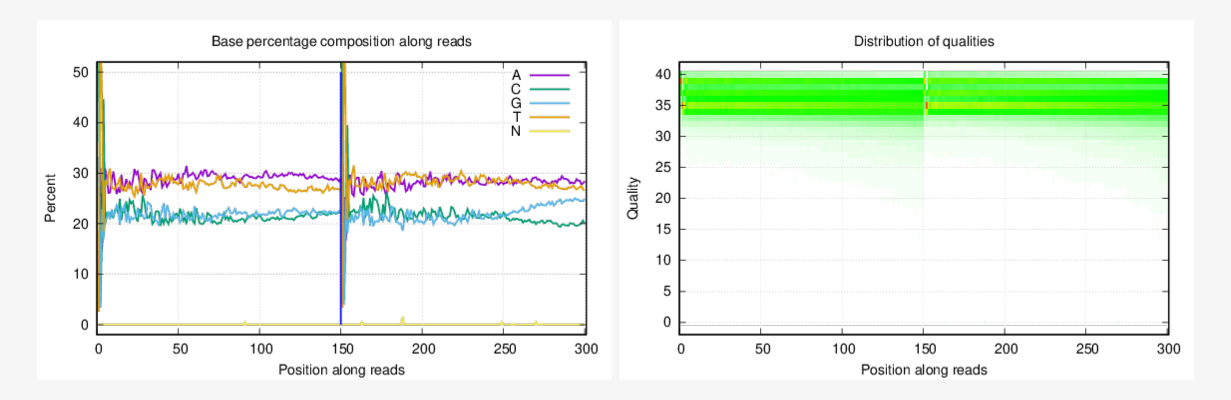
Shen1_K



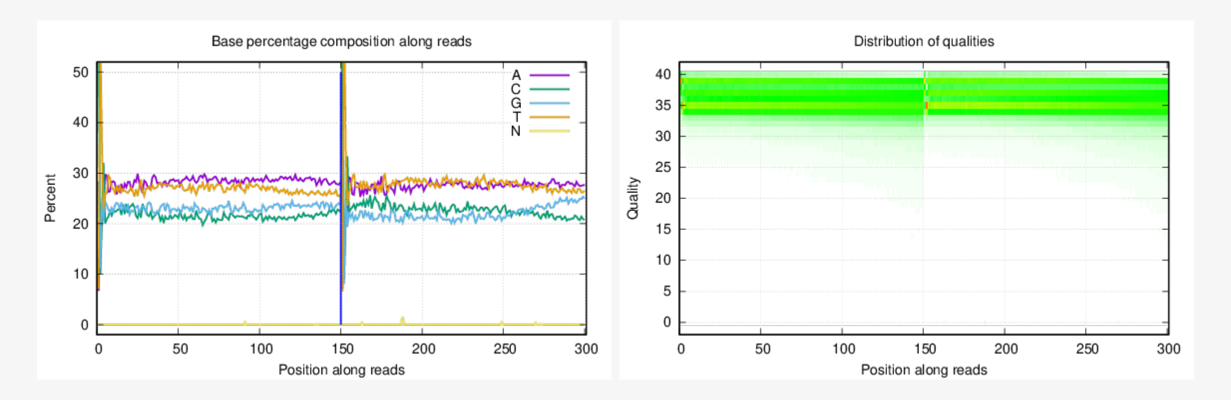
Shen2_e



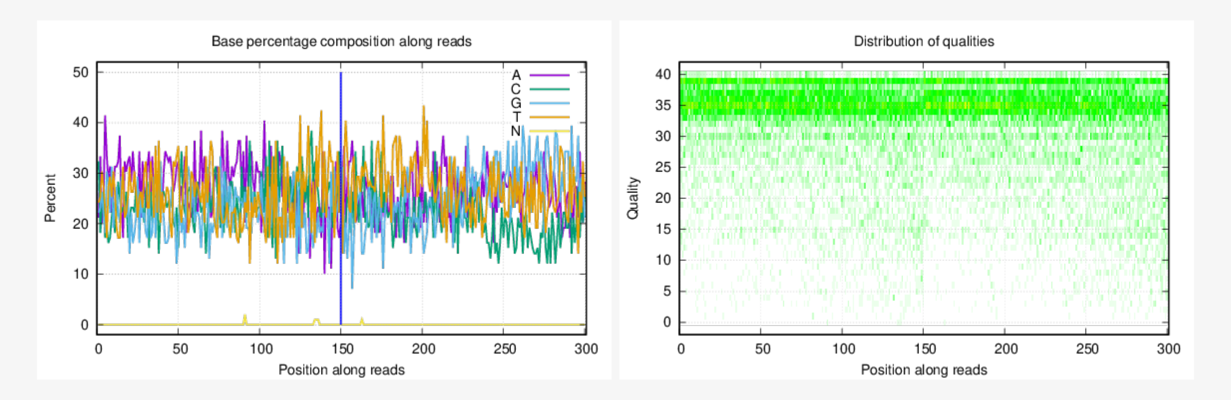
Shen3_K



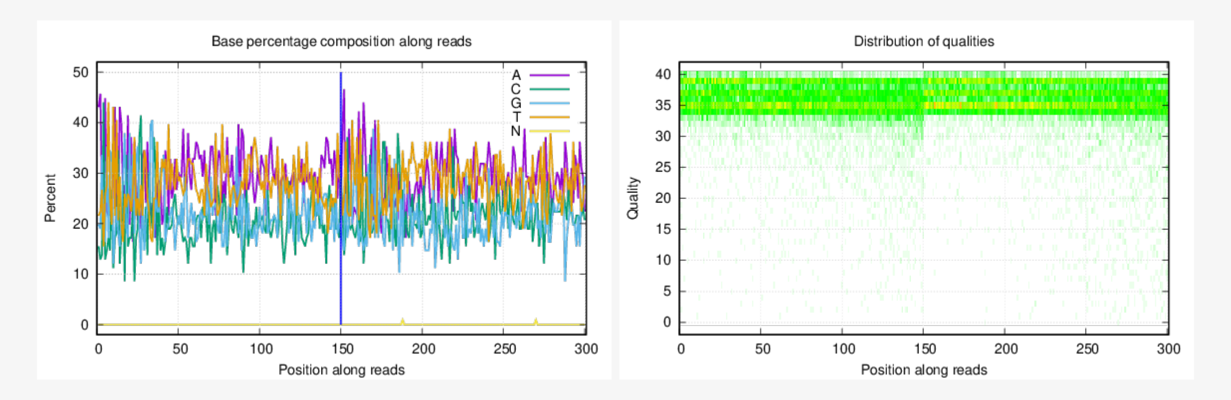
Shen3_e



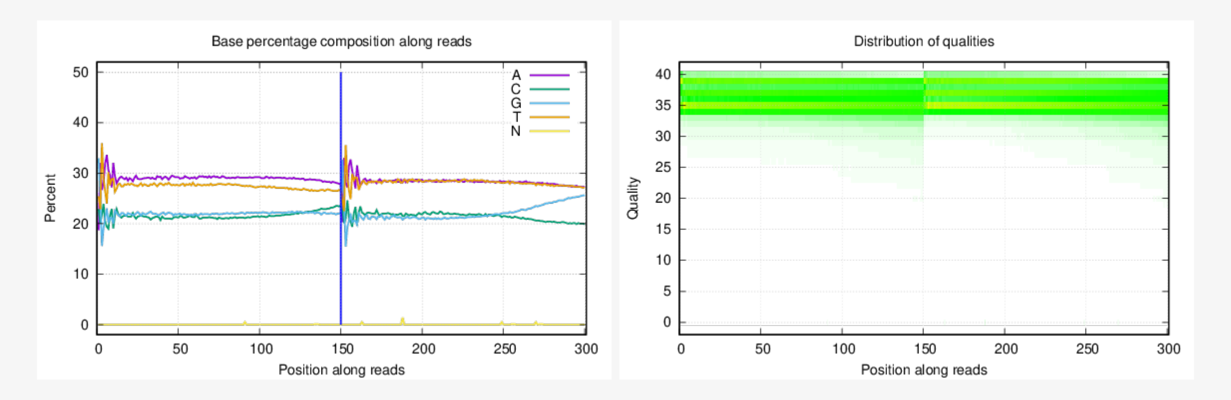
Zap_K2



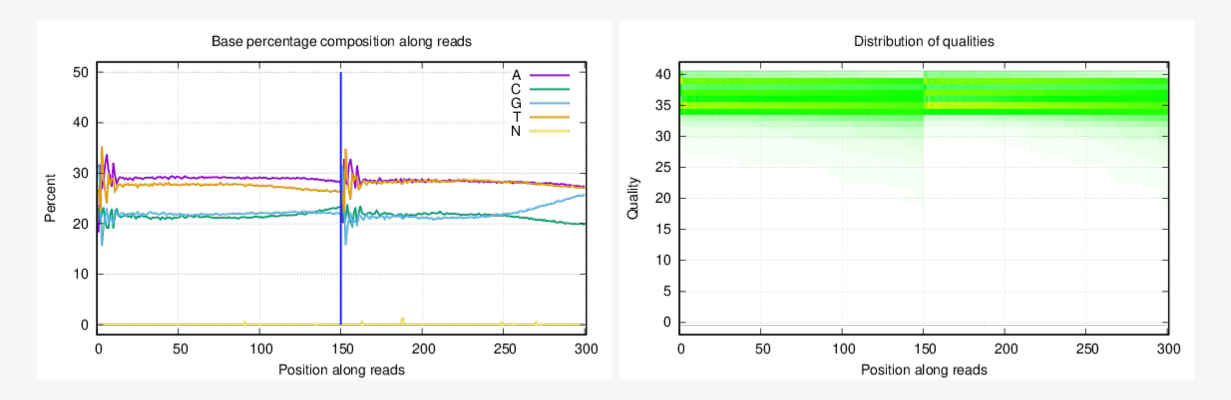
Zap_e3



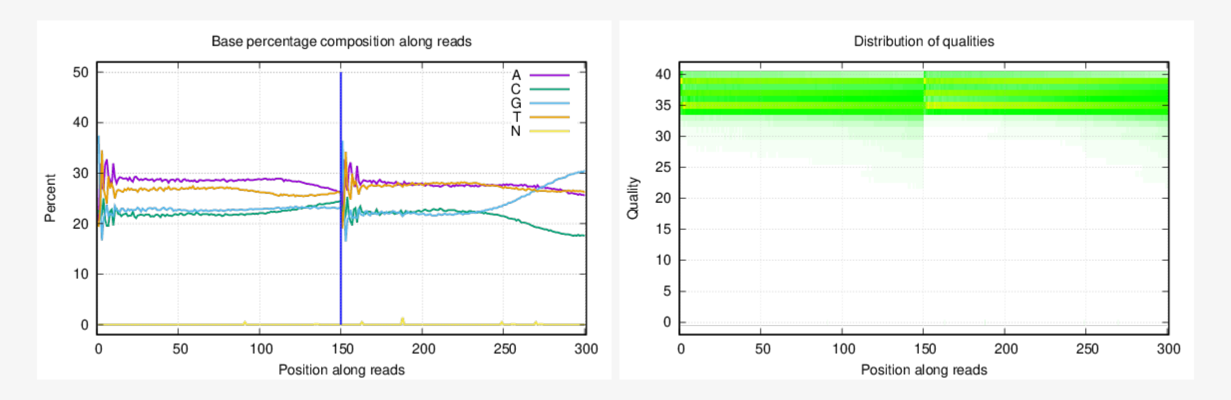
brCTCF_R1



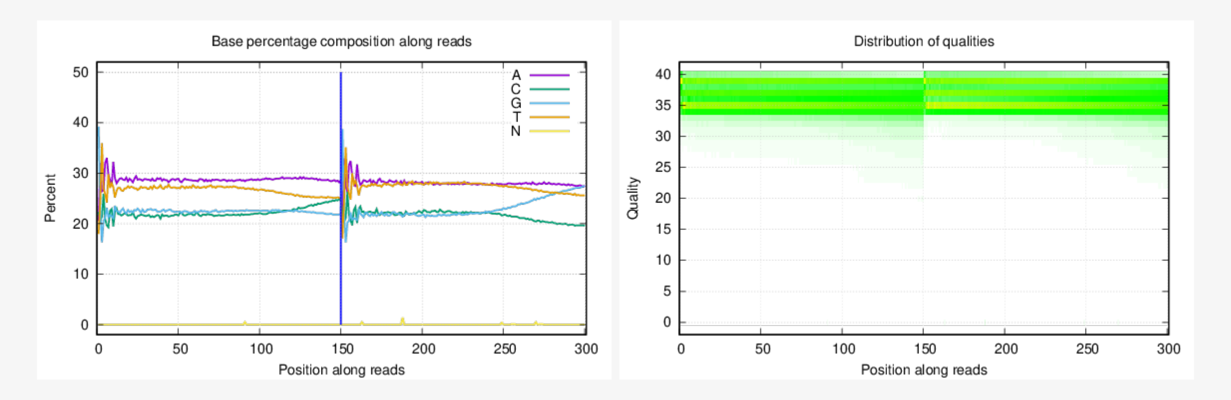
brCTCF_a_R1



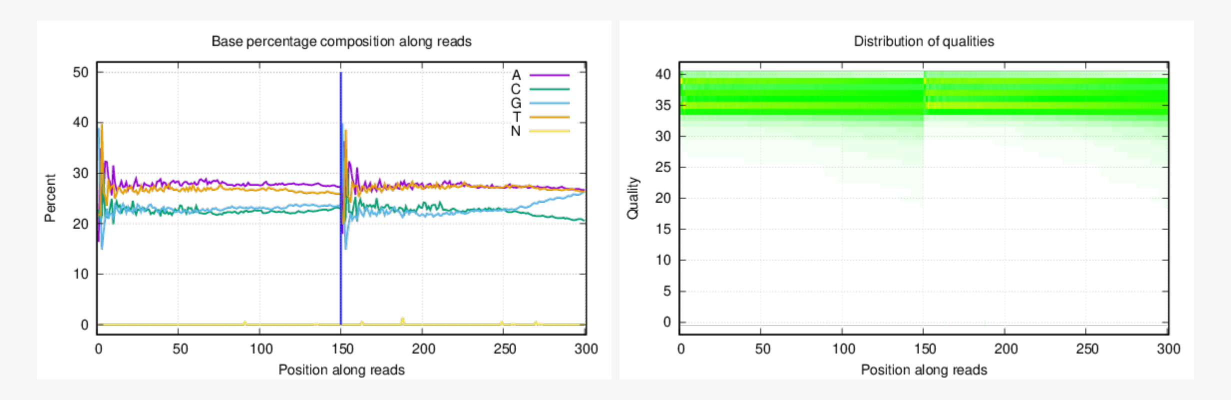
dmCTCF_R1



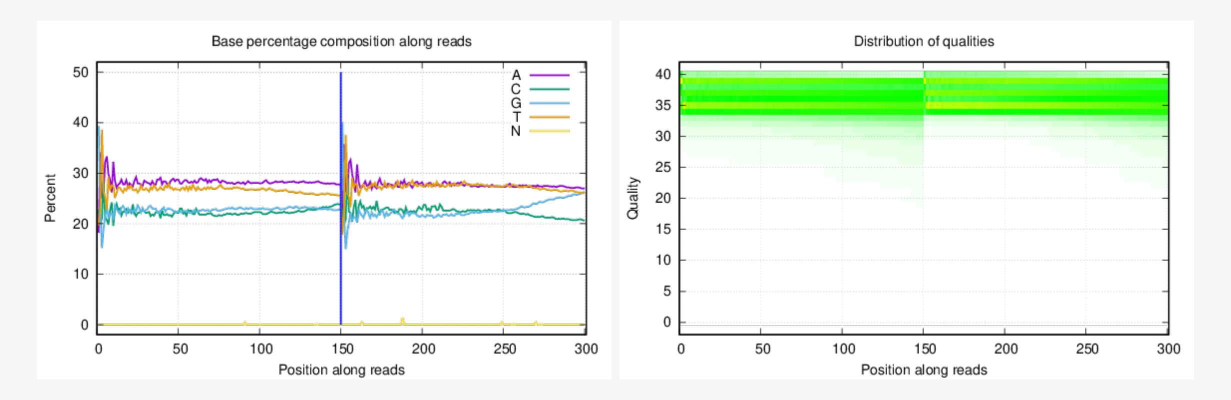
dmCTCF_a_R1



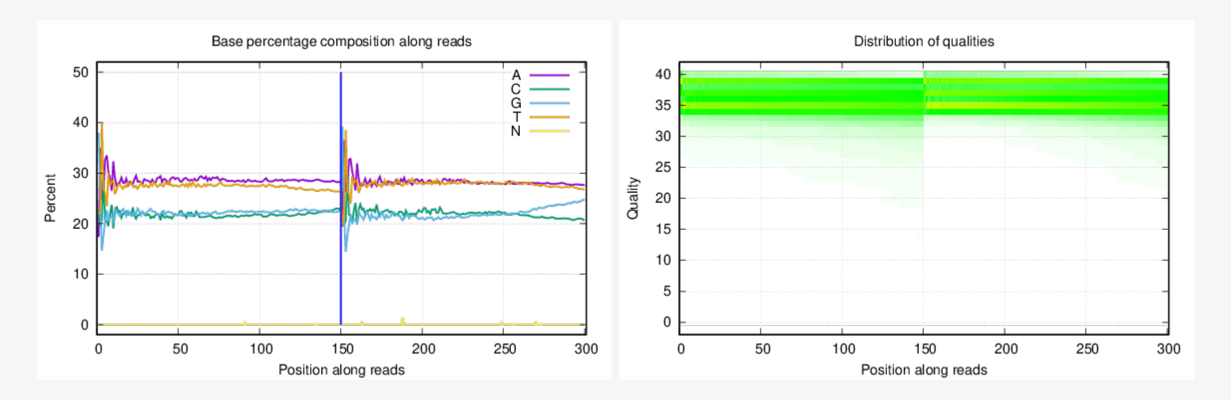
ggCTCF_R2



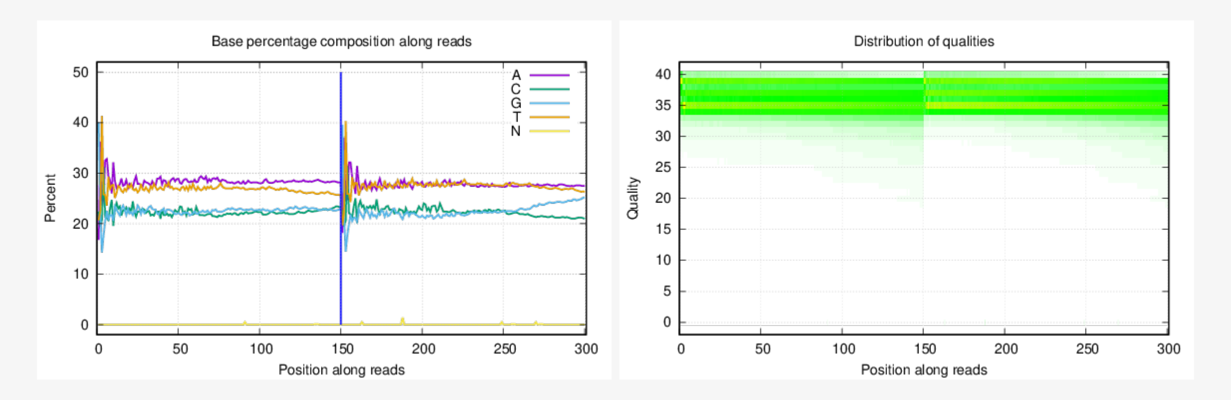
ggCTCF_a_R1



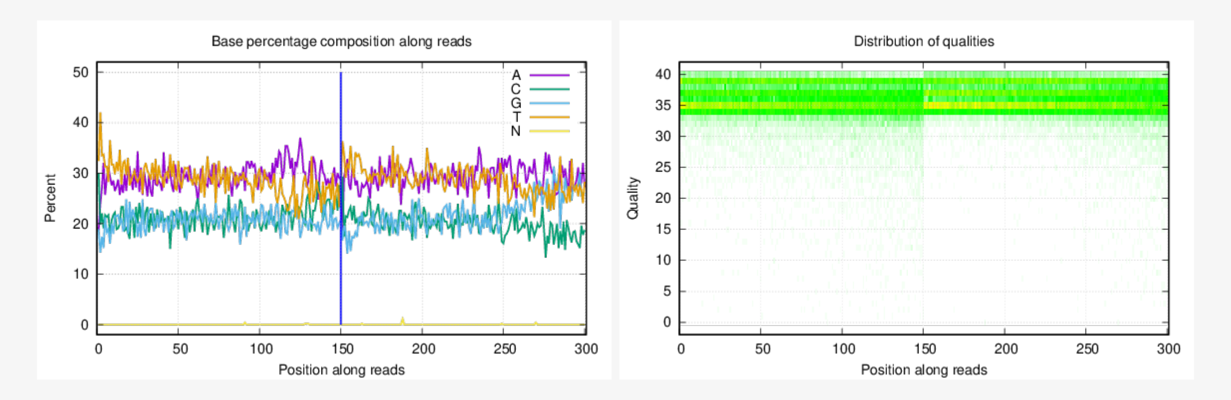
ggCTCF_a_R2



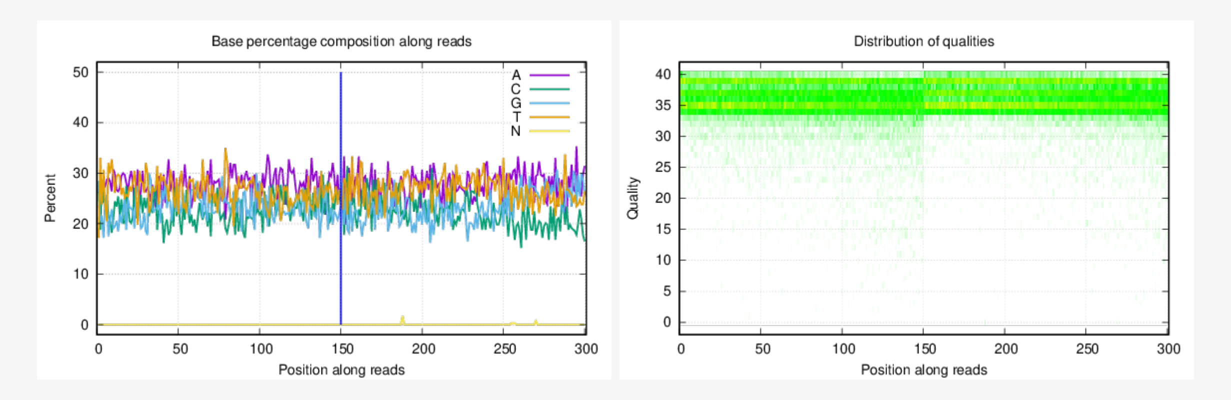
hCTCF_a_R3



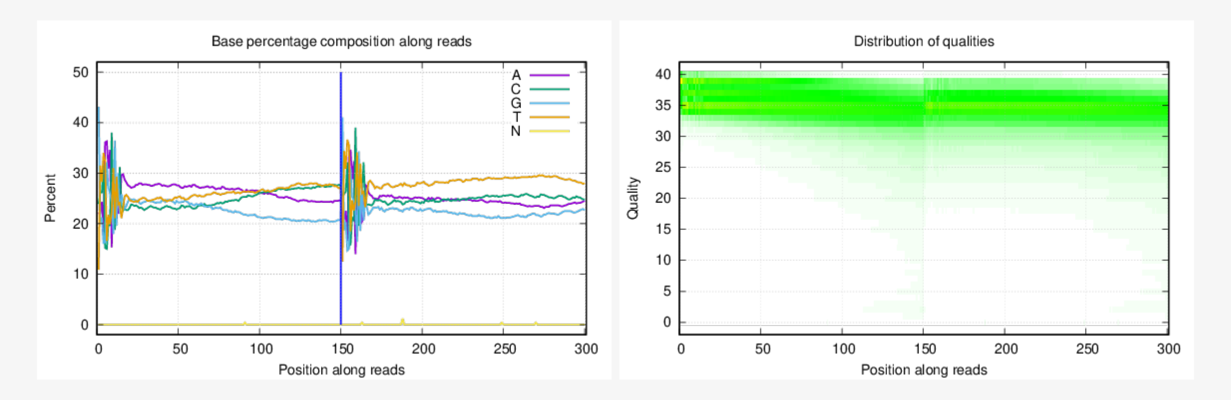
iCS1



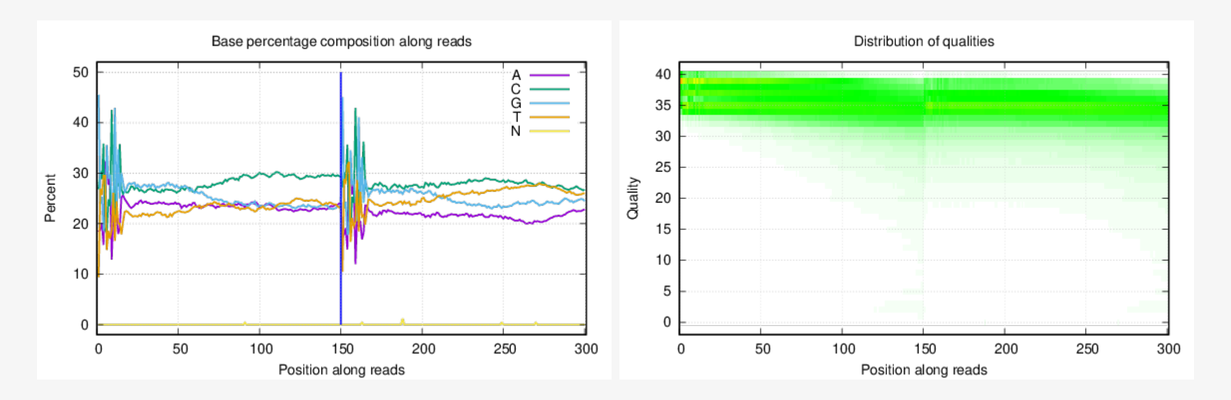
iCS2



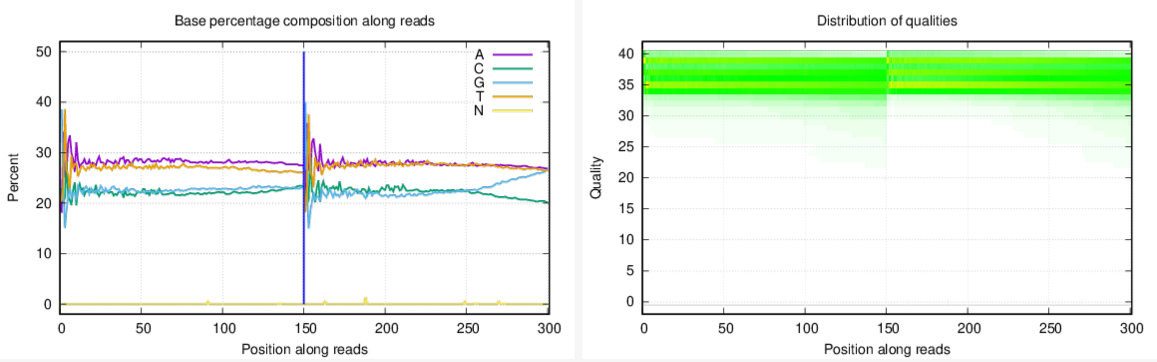
iCS3



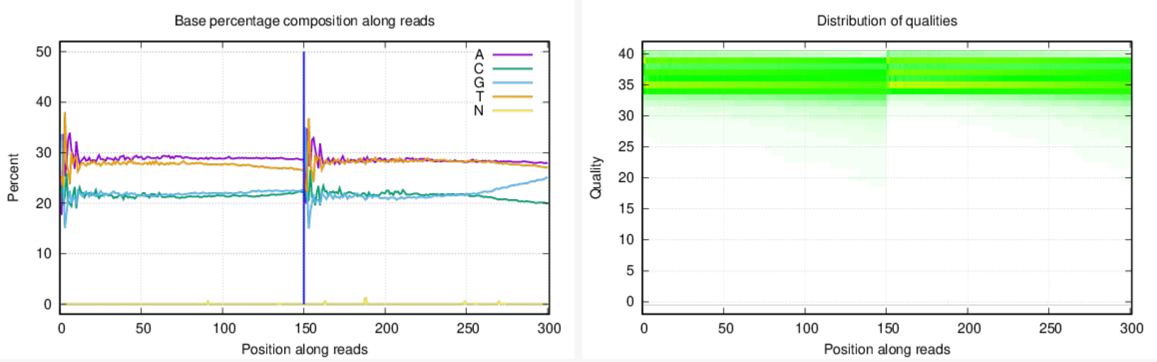
iCS4



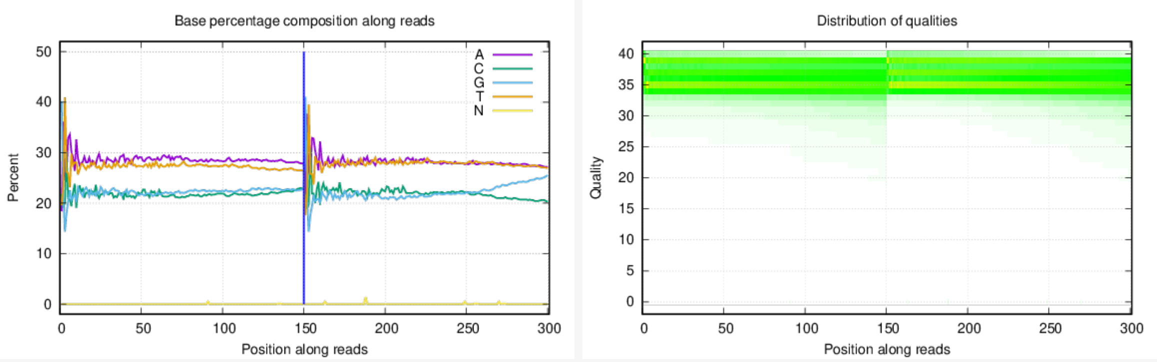
mmCTCF_R1



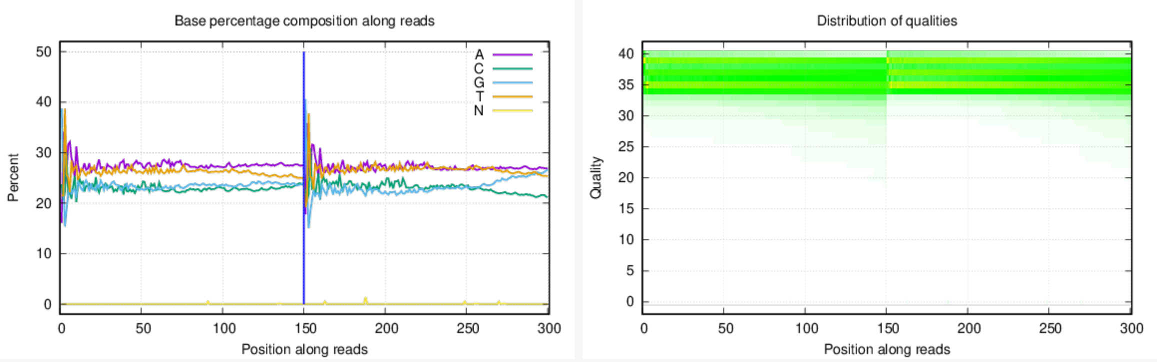
mmCTCF_R2



mmCTCF_a_R1



mmCTCF_a_R2



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