Sequencing Data Report

Project: F20FTSEUHT0946-02_MUShuavR2
Date: 2024.7.24
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 56 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	97.67;97.91	92.93;93.10	55.53;54.66	3,856	1,156,800
10	150;150	97.53;97.53	92.31;92.18	46.97;46.87	6,563	1,968,900
11	150;150	97.83;97.82	93.18;93.01	43.93;43.80	7,155	2,146,500
12	150;150	97.64;97.79	92.56;92.80	46.71;46.57	4,252	1,275,600
13	150;150	97.70;97.55	92.87;92.26	53.05;52.48	2,114	634,200
14	150;150	97.91;97.87	93.66;93.21	53.77;53.03	763	228,900
146-2_A	150;150	97.41;96.73	92.08;90.48	45.07;45.00	20,678,063	6,203,418,900
146-2_B	150;150	97.06;94.68	90.91;84.49	44.88;44.81	8,390,511	2,517,153,300
146-2_V1	150;150	97.27;96.05	91.72;88.65	45.07;44.97	16,470,055	4,941,016,500
146-2_C1	150;150	97.78;97.52	93.23;92.82	43.99;43.87	15,640,546	4,692,163,800
15	150;150	98.02;97.78	93.79;92.89	54.25;53.49	9,012	2,703,600
16	150;150	97.51;97.59	92.42;92.30	45.44;45.18	7,192	2,157,600
17	150;150	97.70;97.73	92.88;92.81	50.11;50.07	2,066	619,800
18	150;150	97.56;97.52	92.47;92.11	52.96;52.41	683	204,900
19	150;150	97.84;97.74	93.14;92.78	49.49;48.88	2,716	814,800
2	150;150	97.14;97.60	91.32;92.26	53.82;52.96	11,619	3,485,700
20	150;150	97.32;97.76	91.85;92.76	47.59;47.24	9,970	2,991,000
21	150;150	97.71;97.60	92.86;92.32	46.82;46.37	4,873	1,461,900
22	150;150	97.58;97.49	92.52;92.01	49.80;49.17	7,639	2,291,700
23	150;150	97.74;97.80	92.81;92.84	47.85;47.29	4,412	1,323,600
24	150;150	97.65;97.74	92.57;92.66	45.44;45.14	4,536	1,360,800
25	150;150	97.91;97.74	93.49;92.72	55.66;54.46	5,931	1,779,300
26	150;150	97.84;97.81	93.19;92.87	49.58;48.77	2,539	761,700
27	150;150	98.06;97.87	93.76;93.03	50.81;50.28	5,416	1,624,800
28	150;150	97.73;97.66	92.91;92.46	52.50;51.23	3,929	1,178,700
29	150;150	98.09;97.95	93.91;93.31	55.68;54.56	4,892	1,467,600

© 2024 BGI Genomics Co.,Ltd.

3	150;150	97.76;97.83	92.92;93.04	50.36;49.54	8,059	2,417,700
30	150;150	97.70;97.61	92.78;92.36	50.56;50.02	6,957	2,087,100
31	150;150	98.09;98.00	93.87;93.41	52.16;51.02	4,007	1,202,100
3_a	150;150	97.10;94.54	90.90;84.00	45.46;45.41	15,547,758	4,664,327,400
4	150;150	96.91;97.75	90.62;92.74	45.45;45.33	9,631	2,889,300
5	150;150	97.78;97.70	93.10;92.61	53.37;52.24	5,659	1,697,700
5_2_D	150;150	97.14;95.58	91.24;87.08	44.84;44.75	12,870,284	3,861,085,200
5_2_E	150;150	97.01;95.58	90.87;87.02	45.01;44.95	12,654,870	3,796,461,000
5_2_V	150;150	96.52;92.24	89.25;77.32	45.72;45.72	1,845,821	553,746,300
5_2_C5	150;150	89.25;92.60	73.44;81.11	42.84;41.72	11,332	3,399,600
5_2_C6	150;150	97.64;97.20	92.79;91.92	44.09;43.97	19,900,977	5,970,293,100
6	150;150	97.30;97.68	91.75;92.52	52.19;51.49	11,721	3,516,300
6069_1_K	150;150	97.69;95.18	92.54;85.17	43.08;42.97	8,628,551	2,588,565,300
7	150;150	97.94;97.91	93.51;93.24	52.42;51.51	7,603	2,280,900
7_12_A	150;150	97.52;96.69	92.38;90.43	44.24;44.08	13,489,589	4,046,876,700
7_12_V1	150;150	97.22;95.97	91.51;88.43	44.93;44.84	8,876,419	2,662,925,700
7_12_V2	150;150	97.09;95.60	91.13;87.20	44.76;44.65	13,230,101	3,969,030,300
7_12_V3	150;150	97.70;97.01	92.88;91.25	45.02;44.96	15,961,155	4,788,346,500
8	150;150	97.06;97.74	91.08;92.71	49.19;48.66	7,401	2,220,300
8388_3_K	150;150	97.76;95.32	92.78;85.61	43.06;42.96	5,678,140	1,703,442,000
87c_Rep2	150;150	97.59;96.95	92.54;90.61	44.47;44.42	205,910	61,773,000
9	150;150	97.85;97.75	93.32;92.60	54.25;53.39	758	227,400
Arvalis	150;150	98.83;98.91	96.31;96.56	46.01;46.10	76,098,171	22,829,451,300
Lim1_K	150;150	97.74;95.36	92.72;85.80	44.55;44.41	4,992,887	1,497,866,100
Oeconomus	150;150	98.89;99.04	96.49;96.95	46.04;46.16	62,921,613	18,876,483,900
P10-six_S1	150;150	98.78;98.84	96.09;96.24	51.11;51.23	5,620,698	1,686,209,400
P142-six_S2	150;150	98.75;98.79	95.97;96.06	49.41;49.51	5,815,461	1,744,638,300
P143-six_S3	150;150	98.92;98.84	96.50;96.23	50.08;50.20	4,770,621	1,431,186,300
P181-six_S4	150;150	98.86;98.95	96.31;96.52	50.54;50.68	6,584,286	1,975,285,800
TA5	150;150	97.38;95.41	91.73;85.83	34.35;34.18	32,354,642	9,706,392,600

Table Format:

© 2024 BGI Genomics Co.,Ltd.

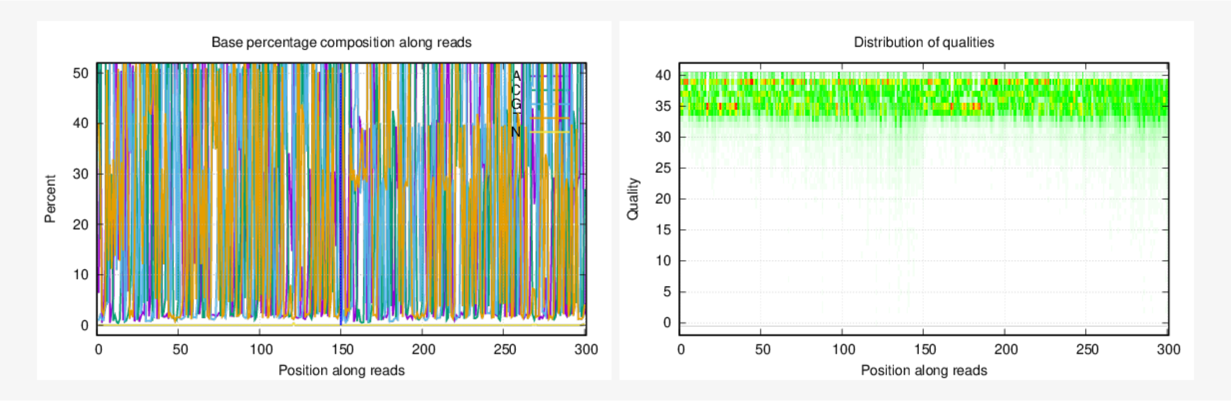
1. Sample: The name of sample
2. Length: The Length of reads
3. Q20 (%): The proportion of nucleotides with quality value larger than 20
4. Q30 (%): The proportion of nucleotides with quality value larger than 30
5. GC Content(%): The proportion of bases G and C
6. Total Reads: The total number of read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

© 2024 BGI Genomics Co.,Ltd.

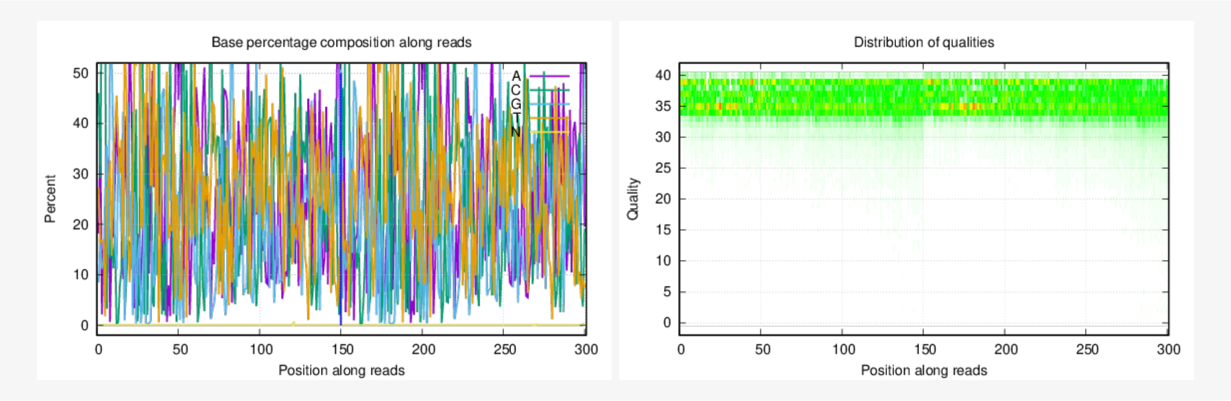
Data Quality Control

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.

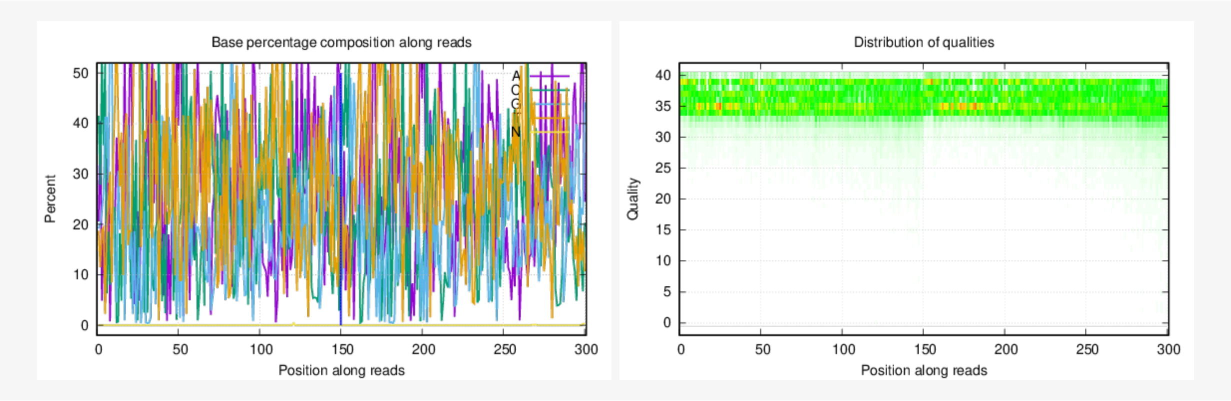
1



10

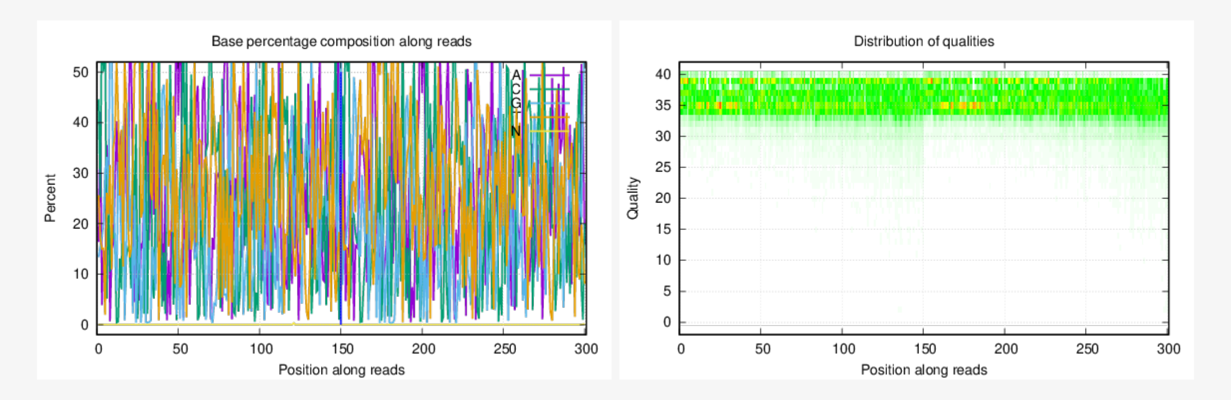


11

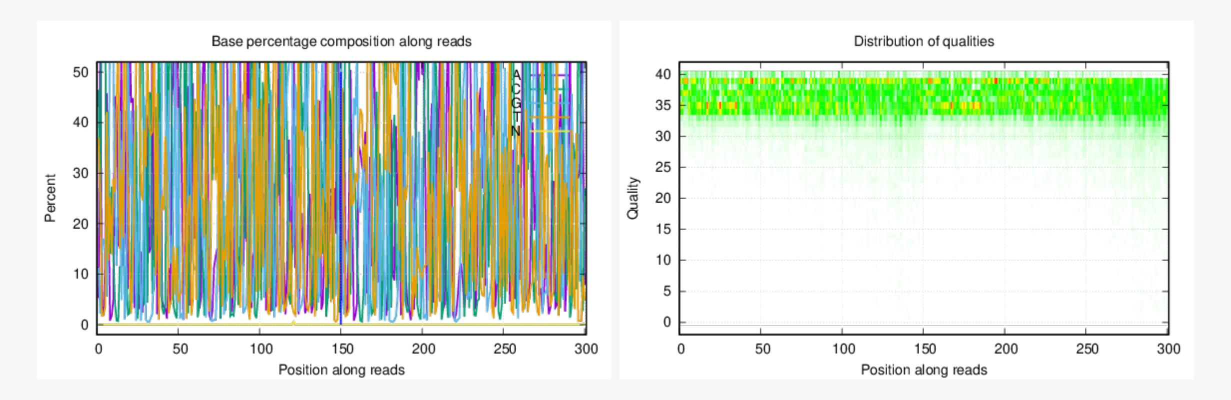


12

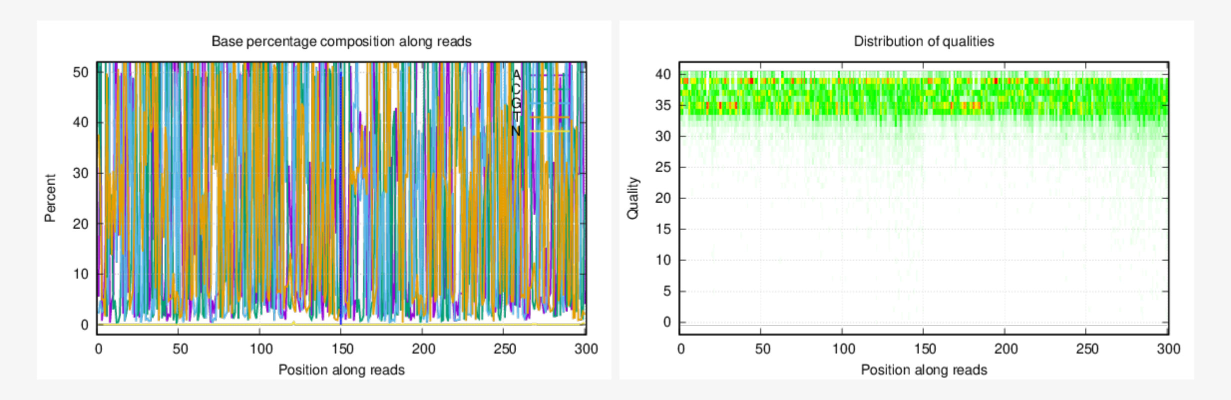
© 2024 BGI Genomics Co.,Ltd.



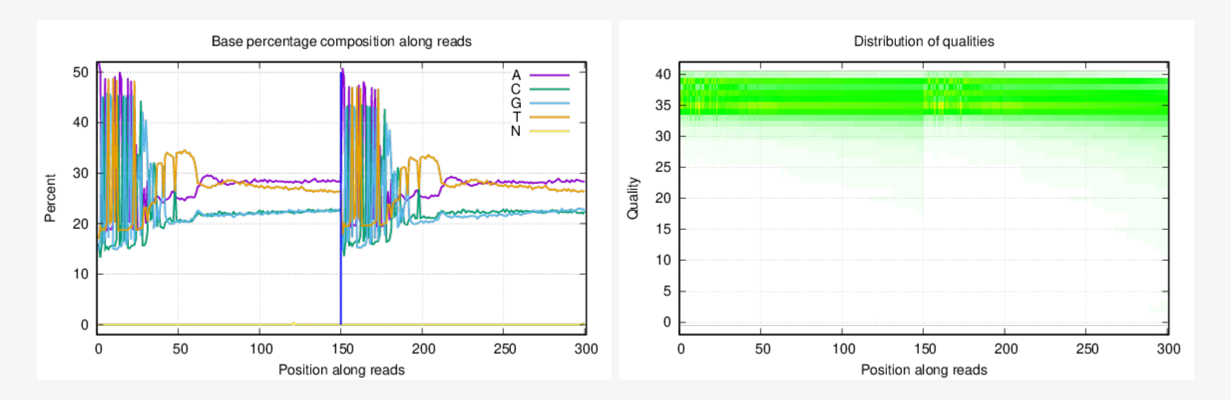
13



14

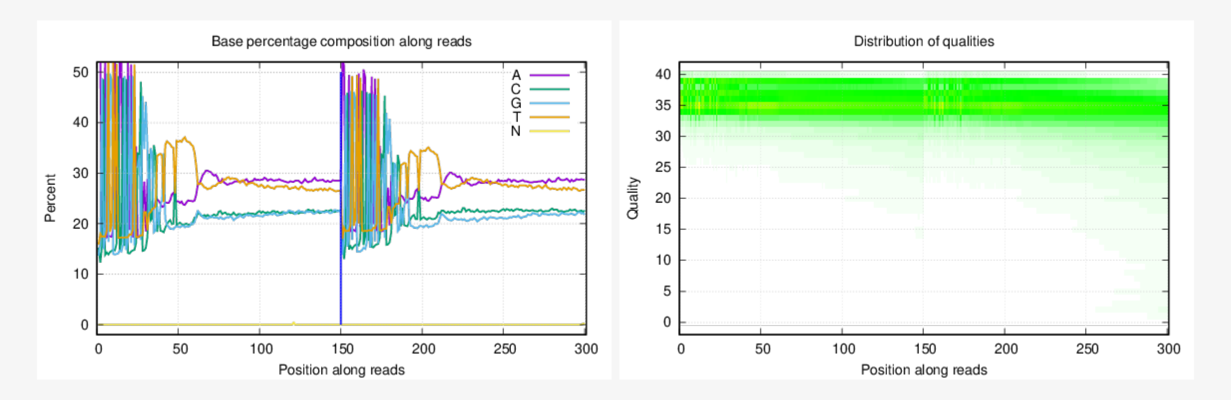


146-2_A

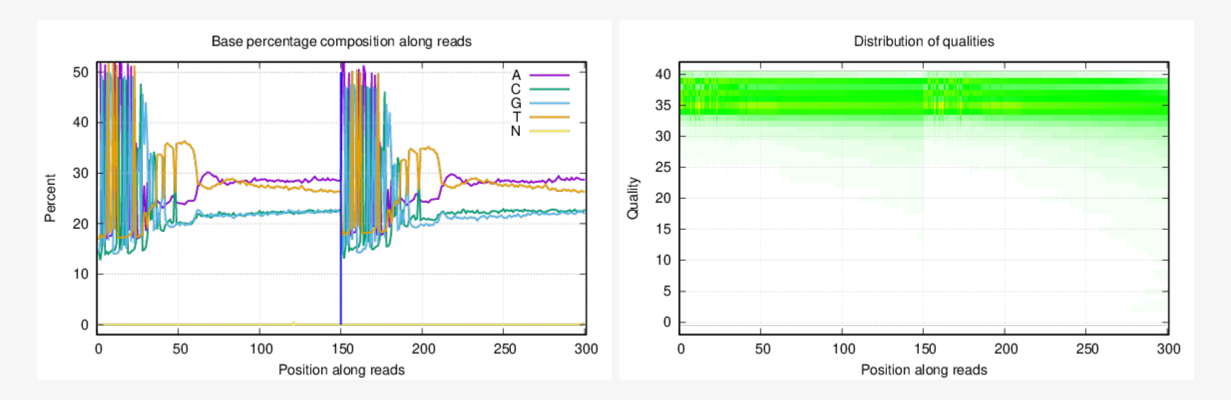


146-2_B

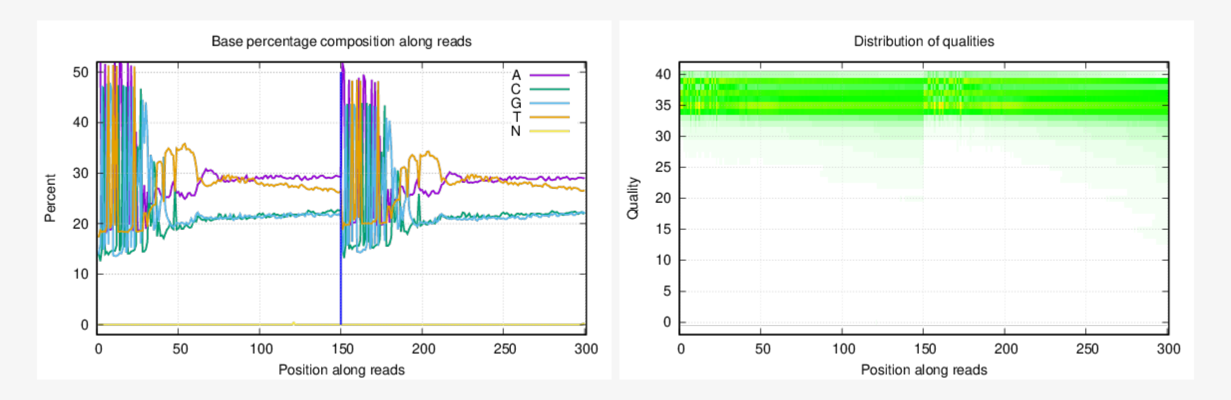
© 2024 BGI Genomics Co.,Ltd.



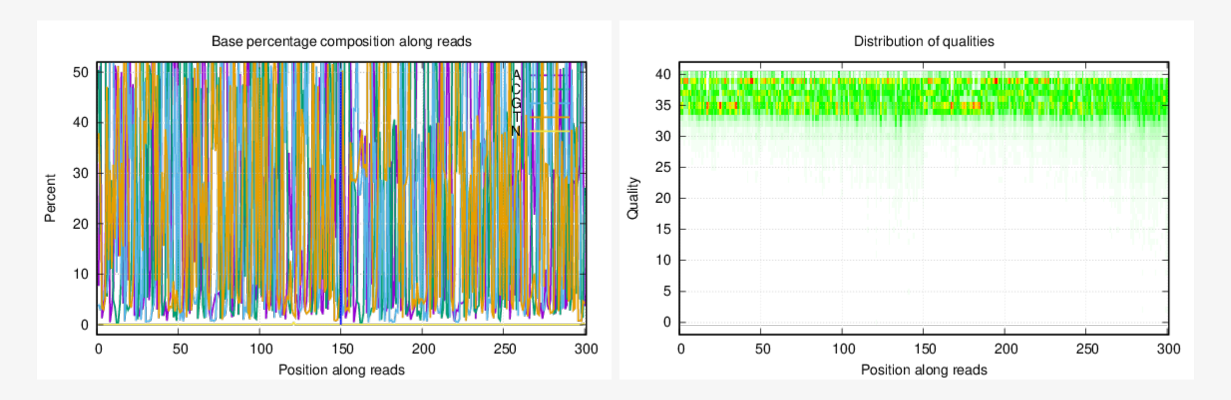
146-2_V1



146-2_C1

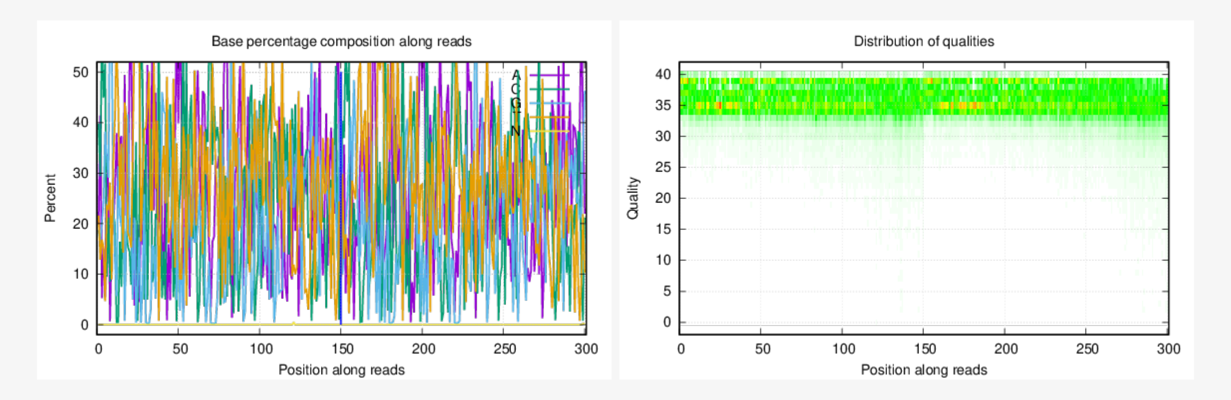


15

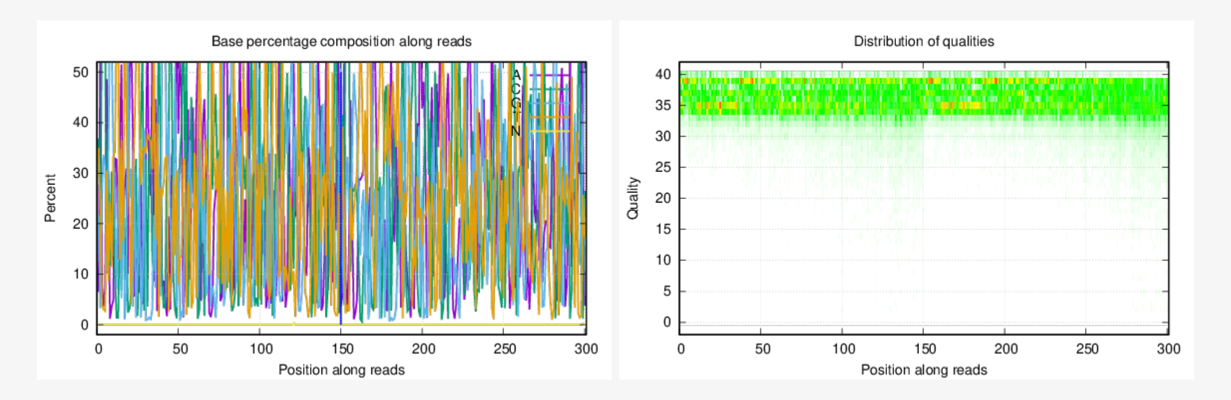


16

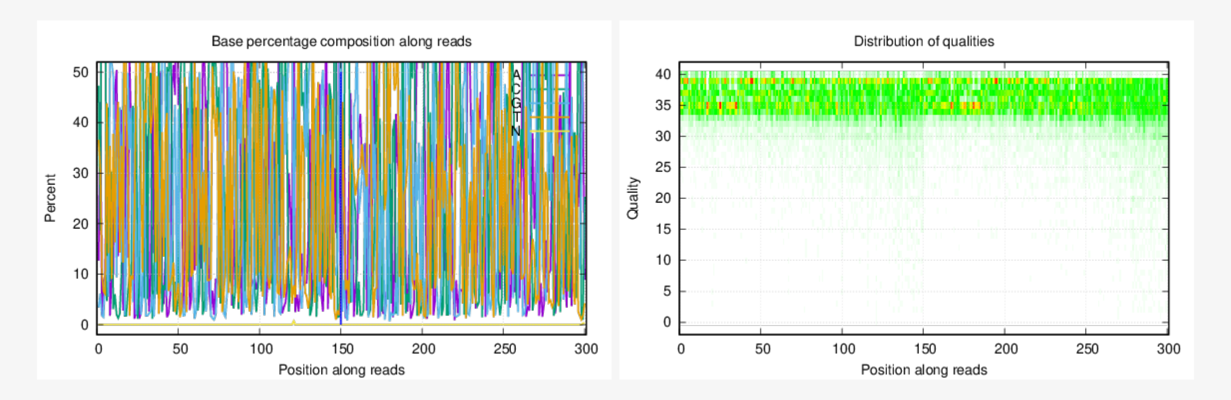
© 2024 BGI Genomics Co.,Ltd.



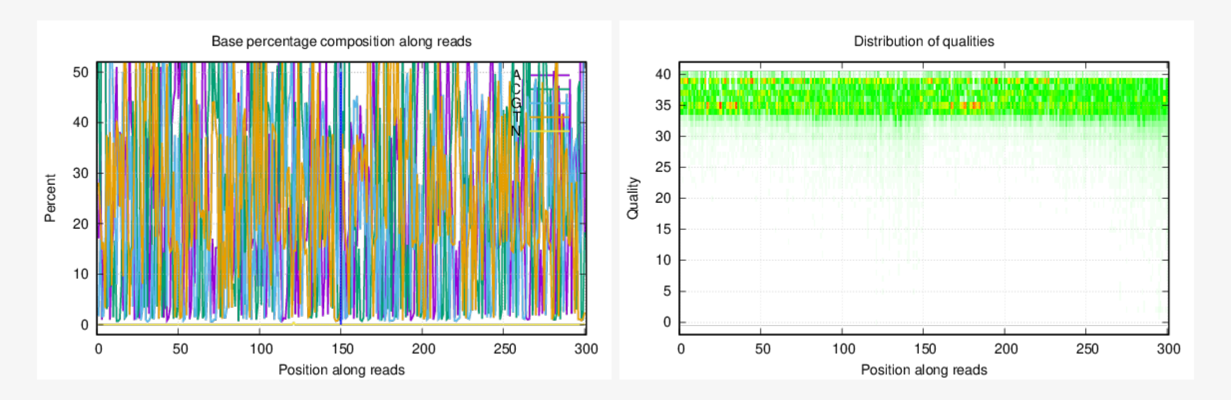
17



18

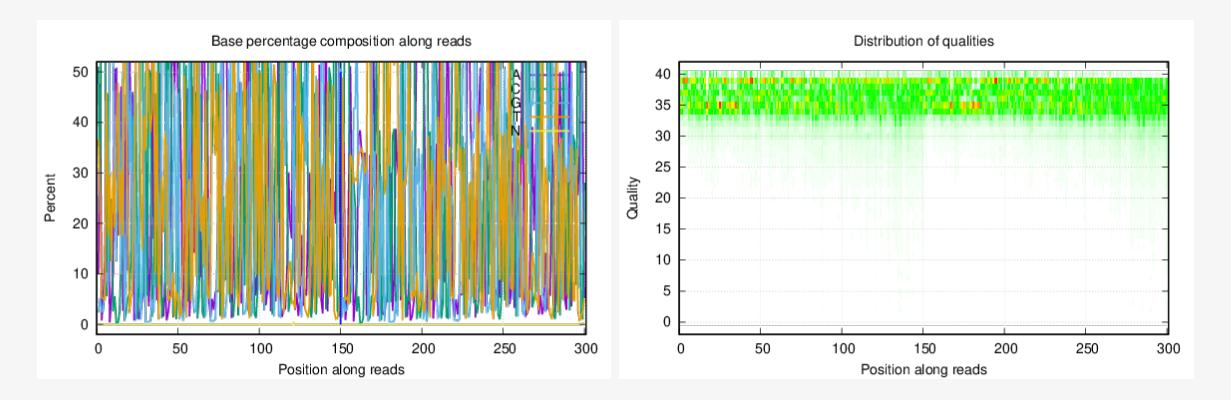


19

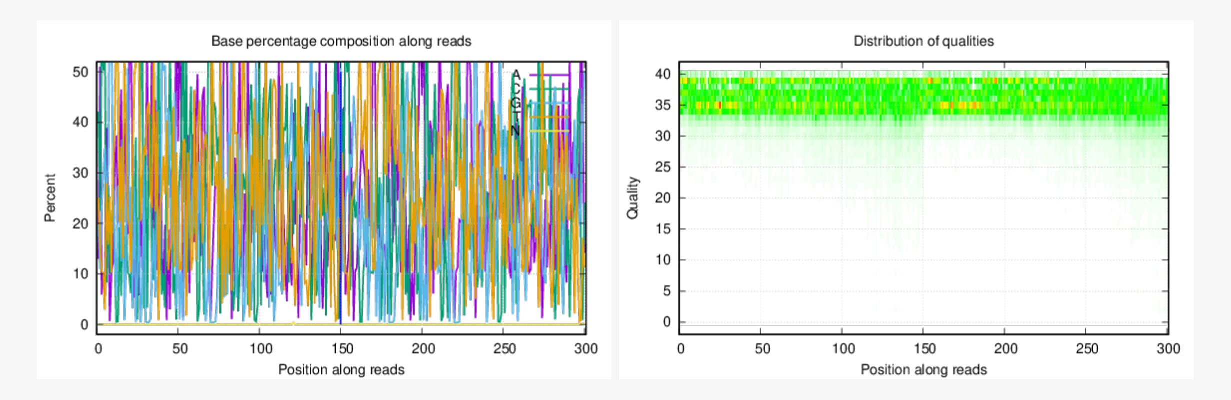


2

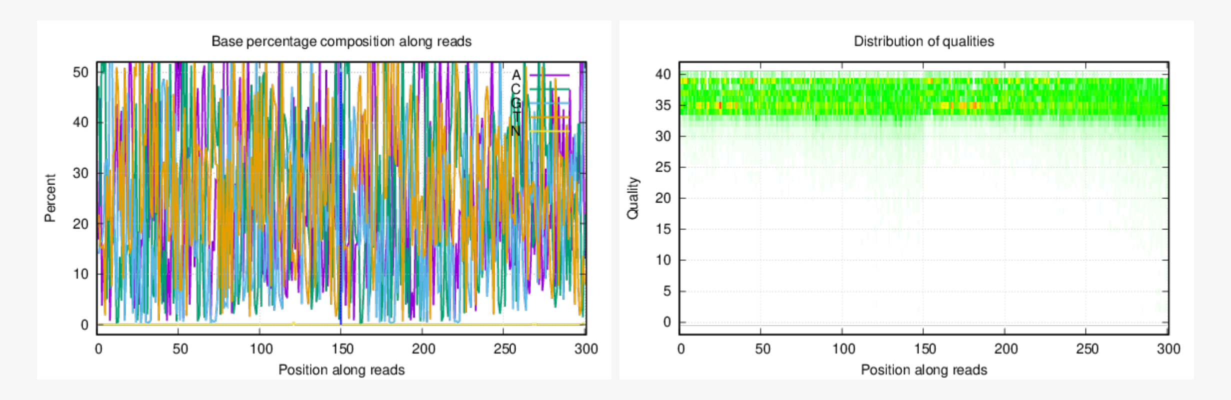
© 2024 BGI Genomics Co.,Ltd.



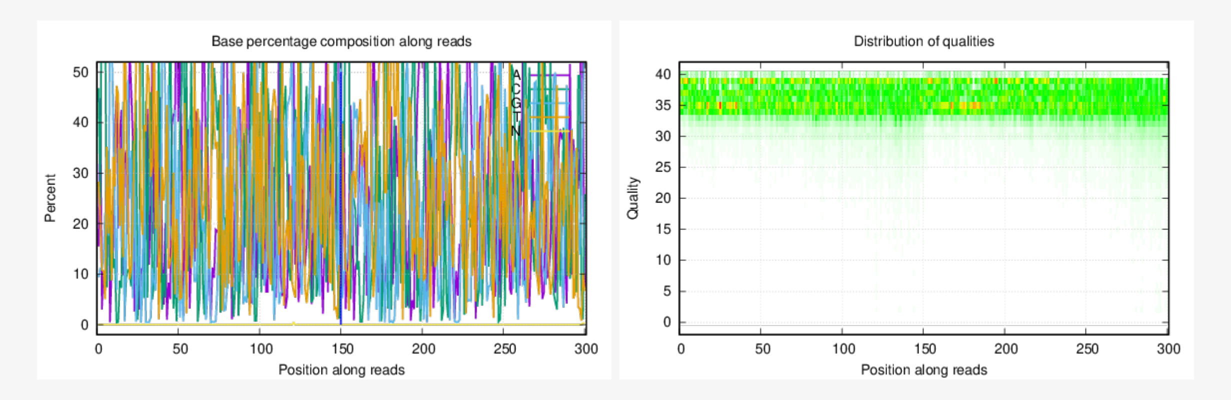
20



21

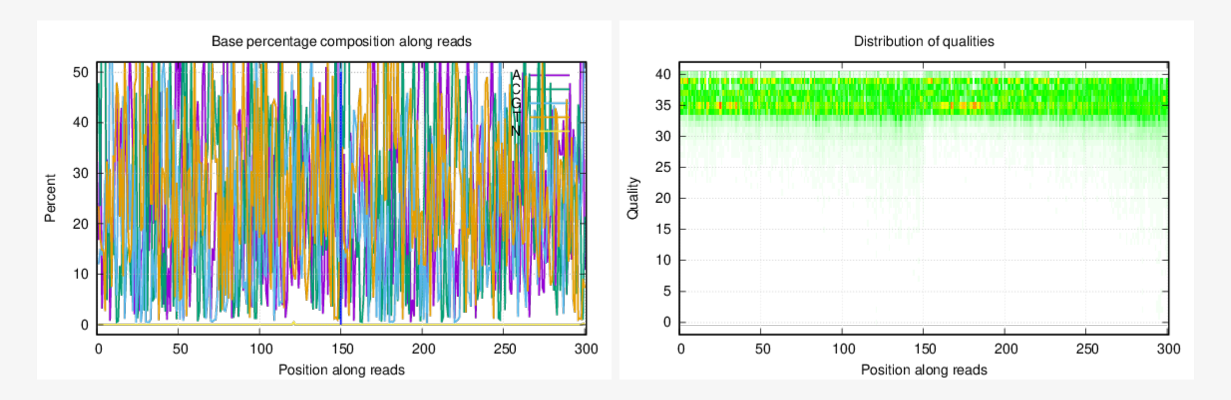


22

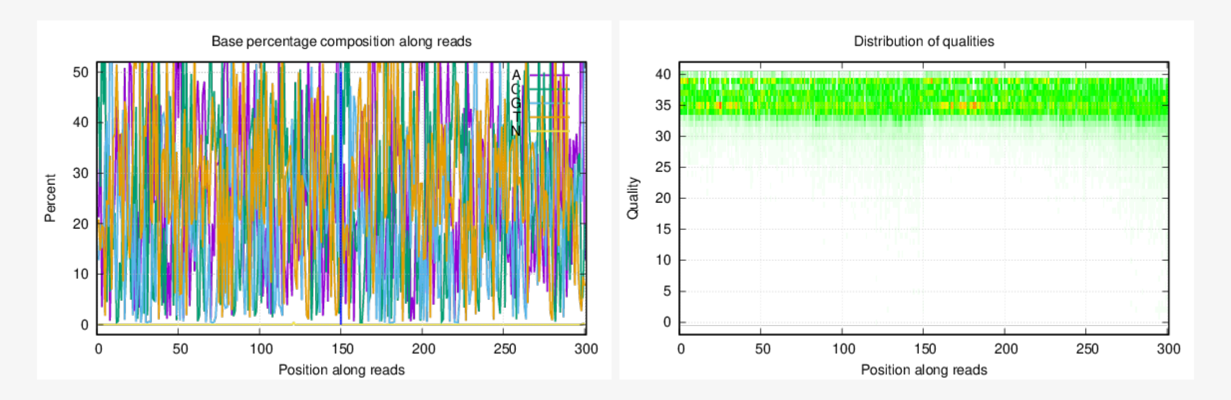


23

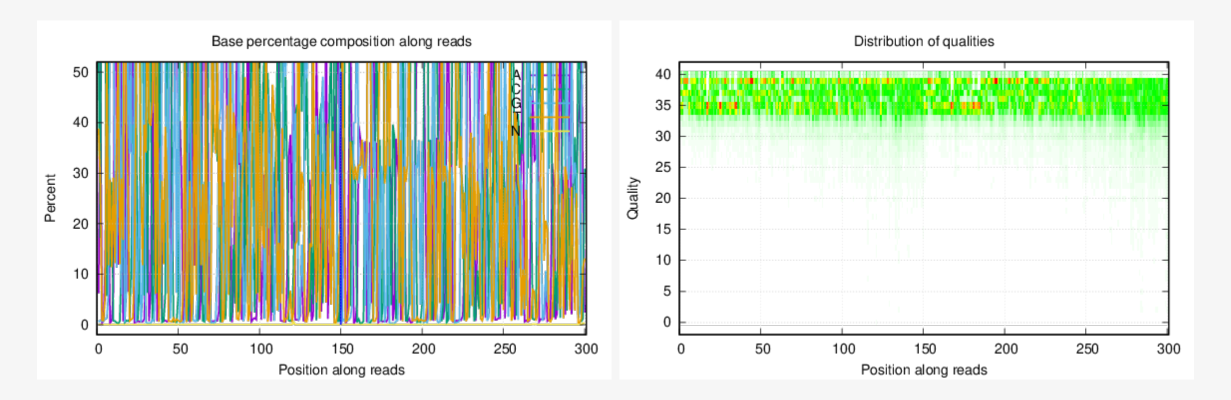
© 2024 BGI Genomics Co.,Ltd.



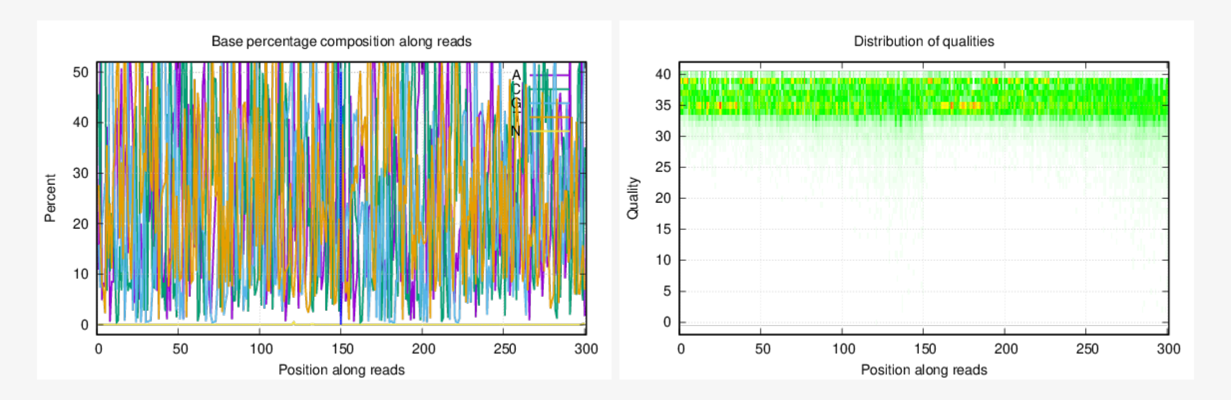
24



25

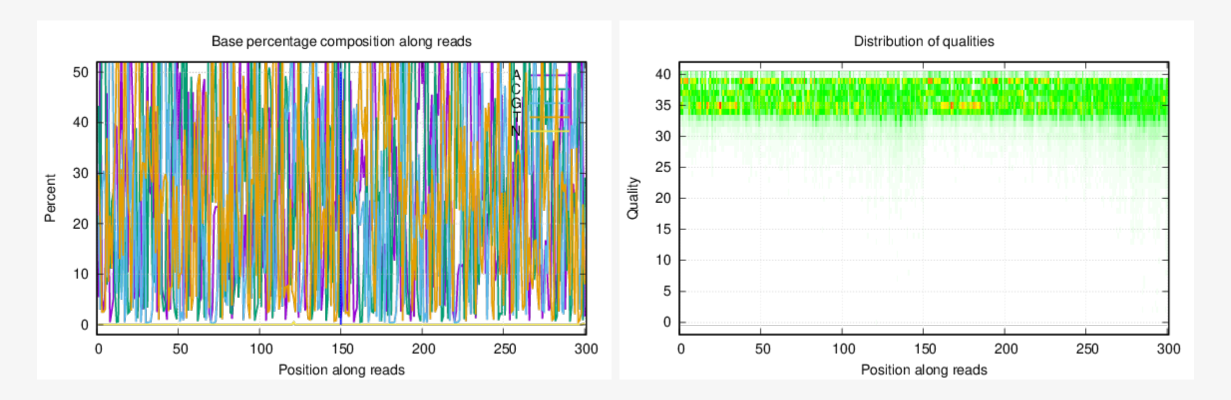


26

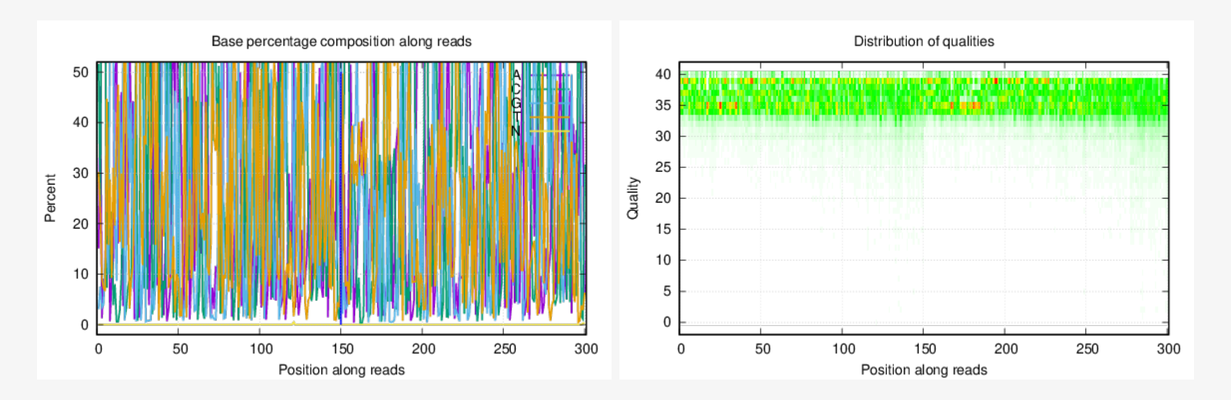


27

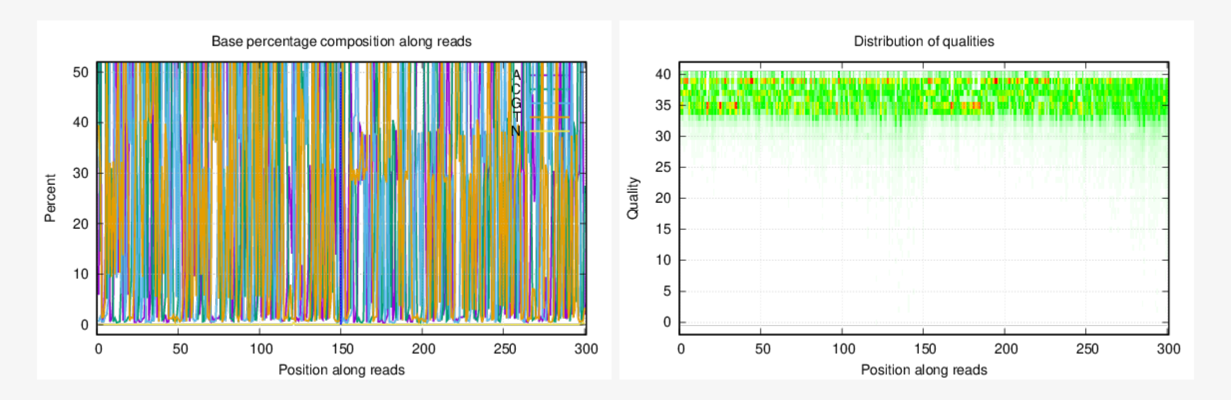
© 2024 BGI Genomics Co.,Ltd.



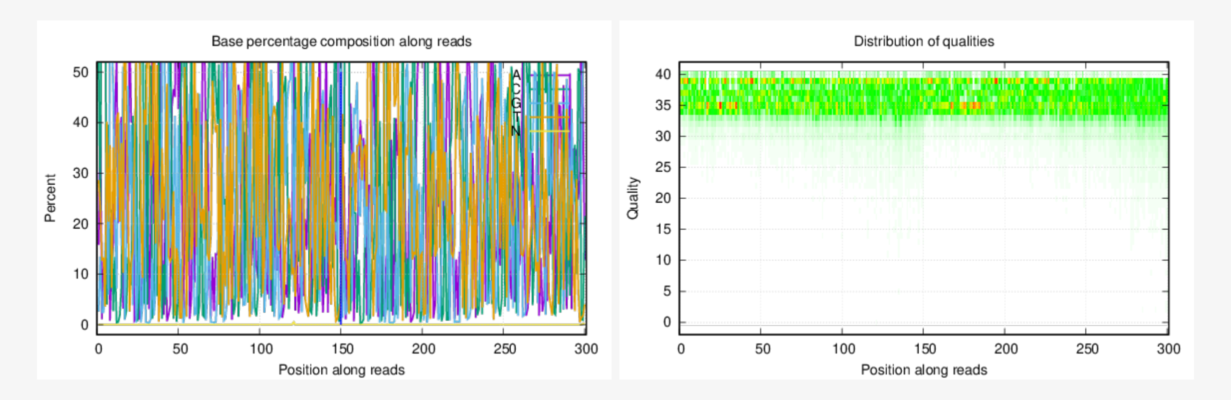
28



29

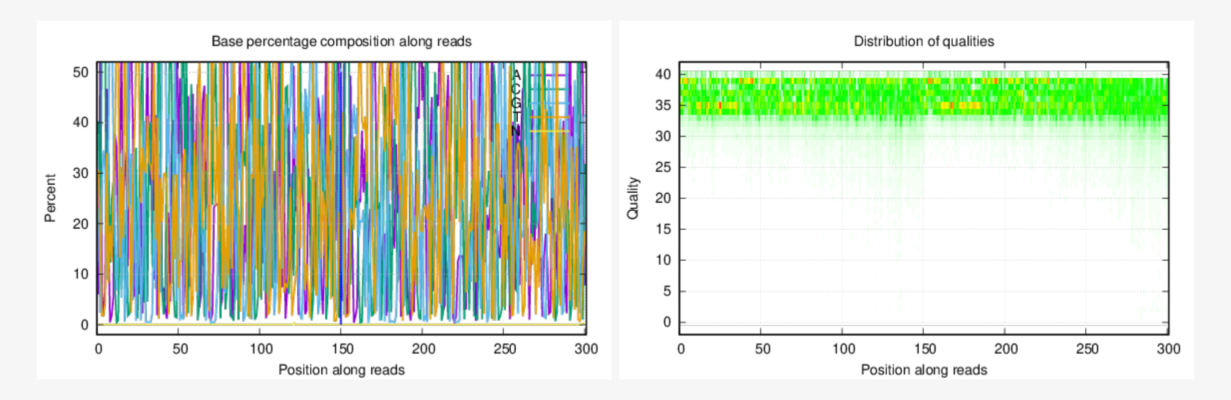


3

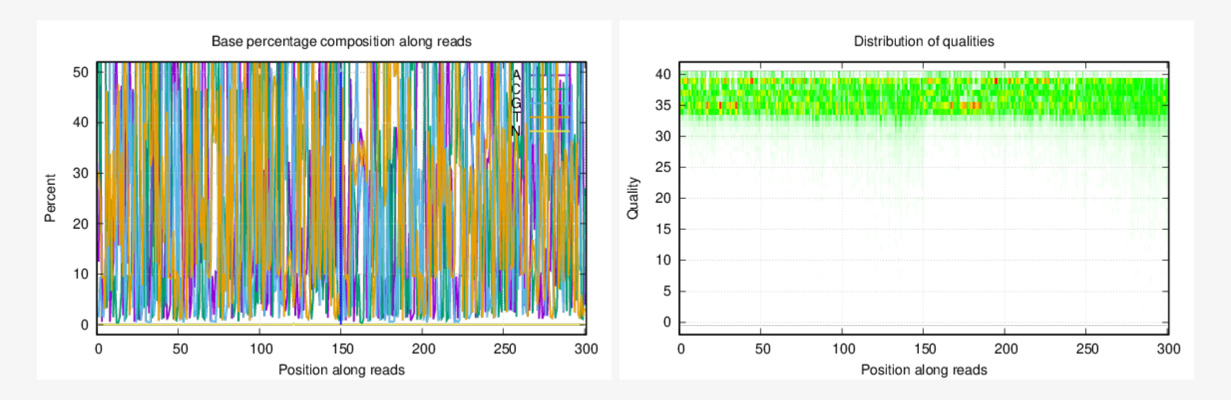


30

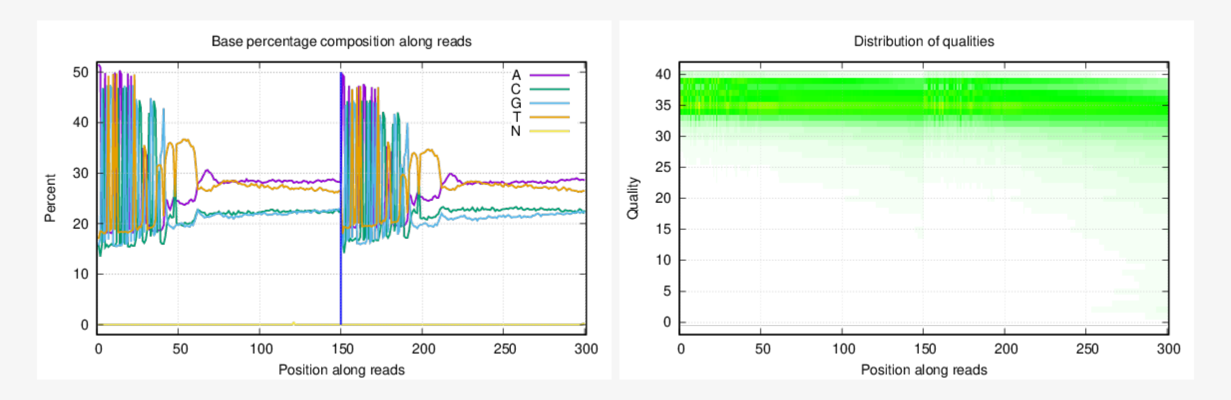
© 2024 BGI Genomics Co.,Ltd.



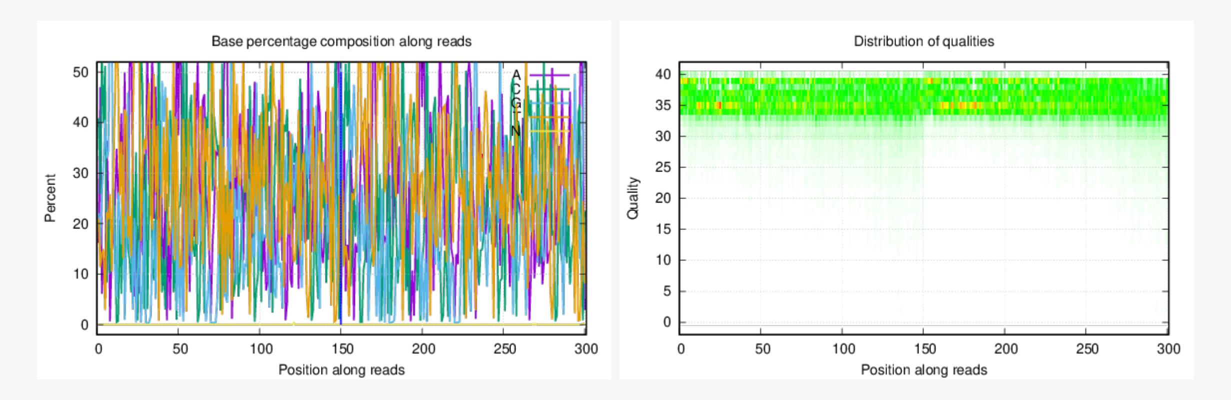
31



3_a

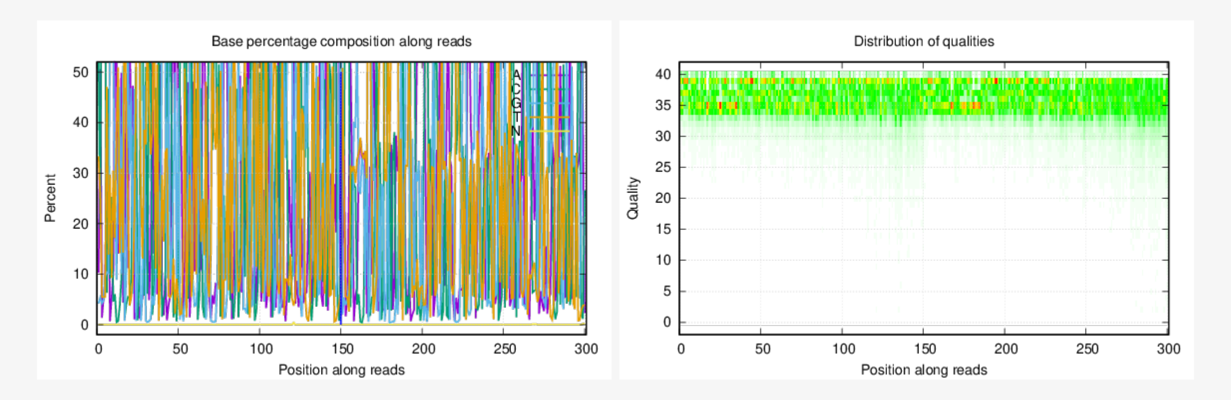


4

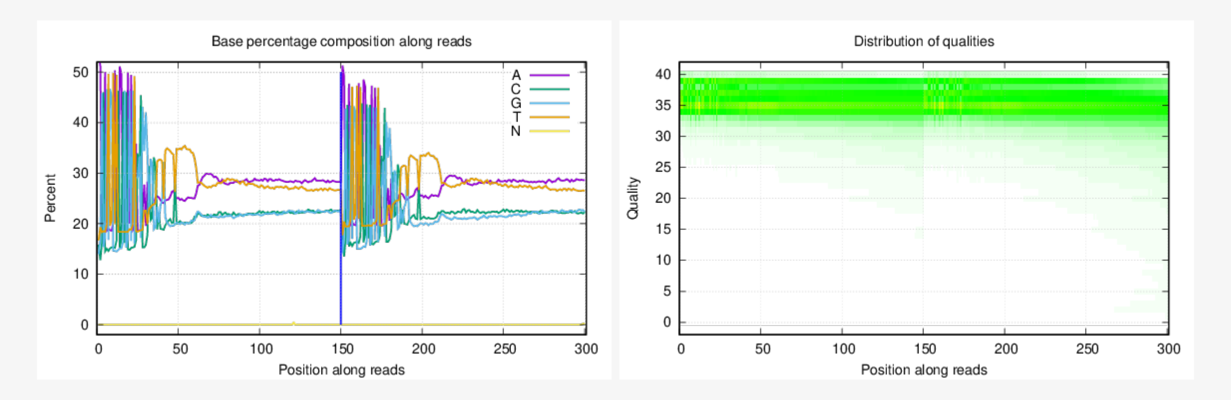


5

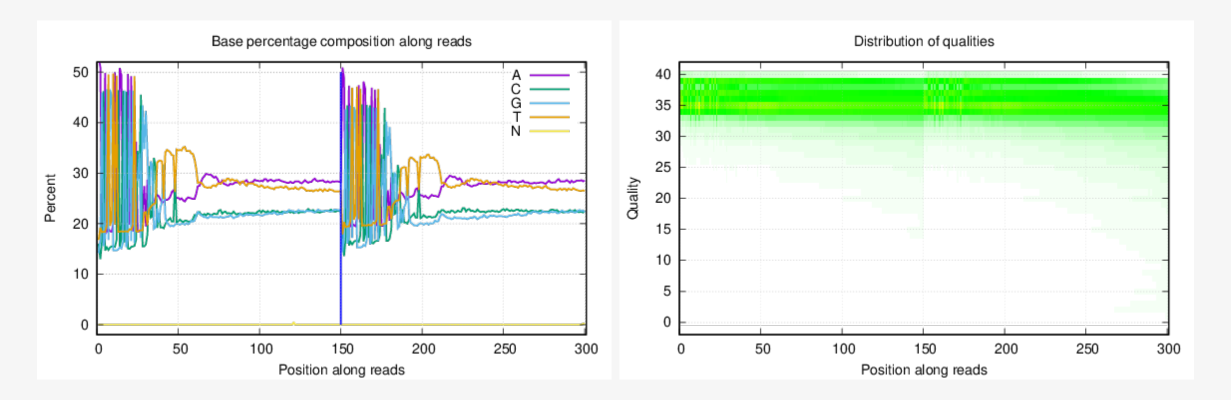
© 2024 BGI Genomics Co.,Ltd.



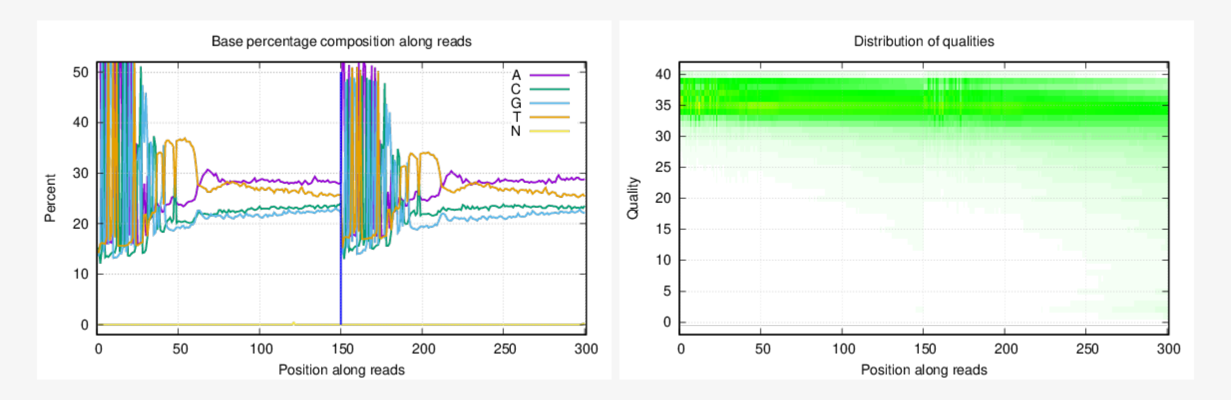
5_2_D



5_2_E

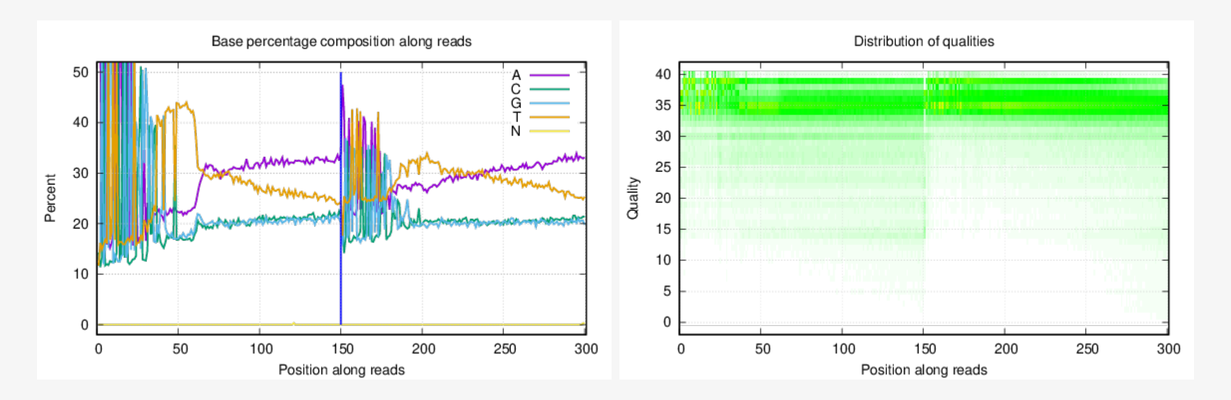


5_2_V

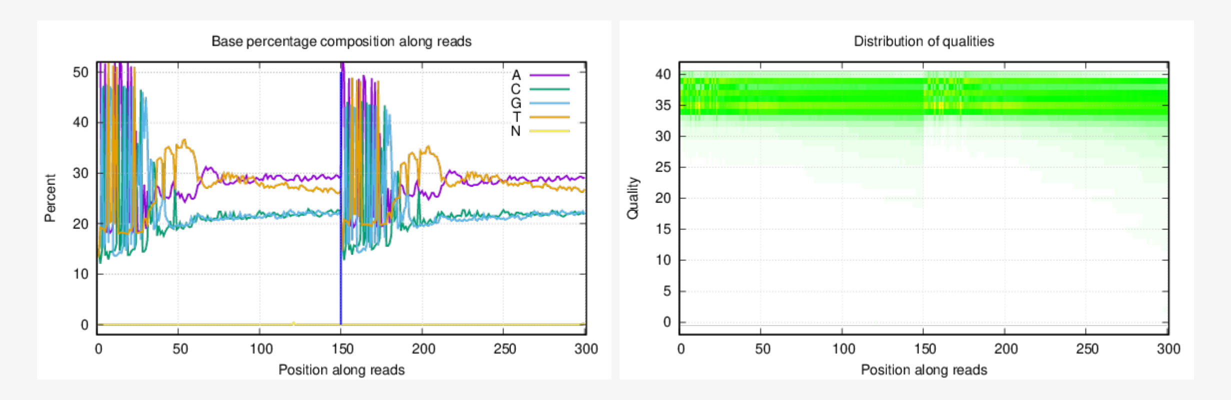


5_2_C5

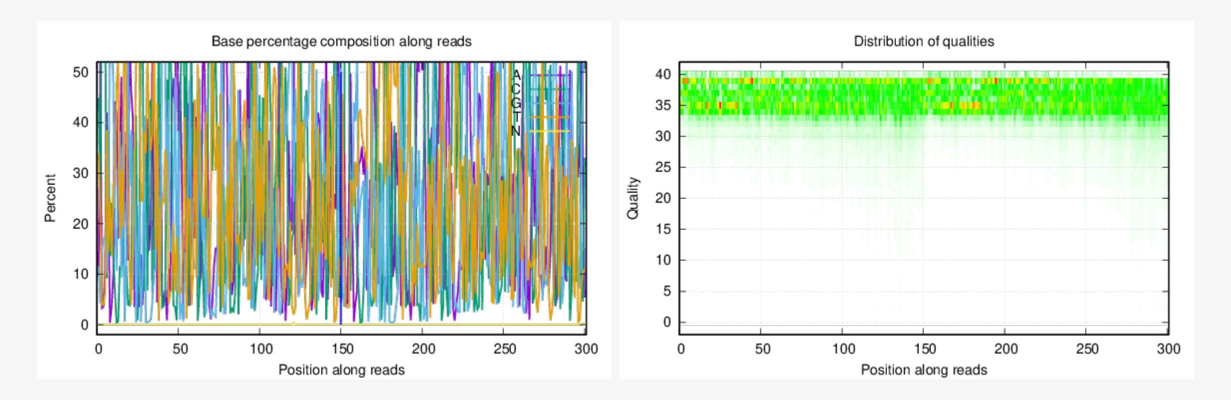
© 2024 BGI Genomics Co.,Ltd.



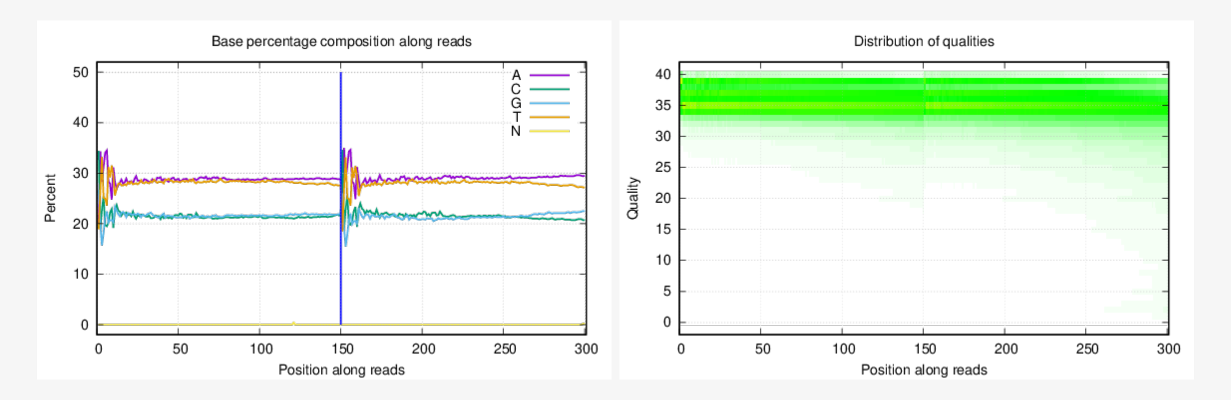
5_2_C6



6

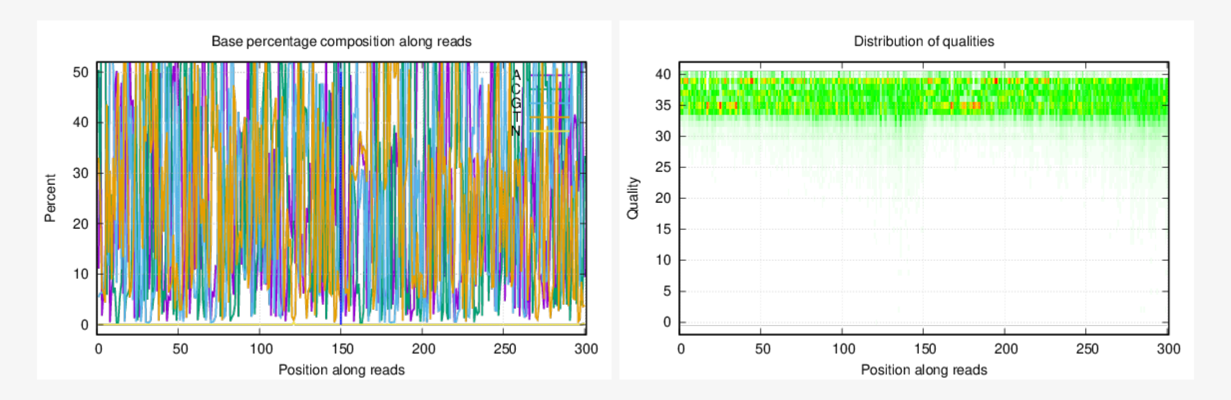


6069_1_K

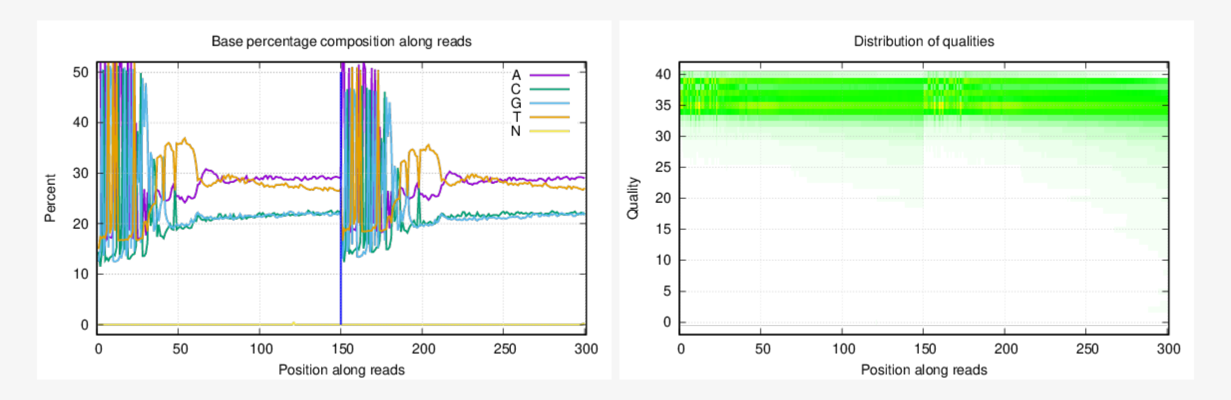


7

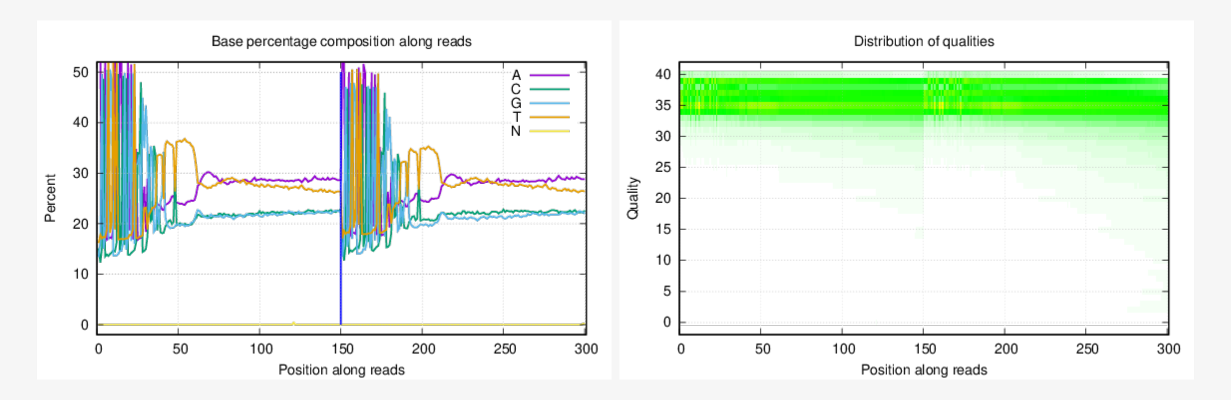
© 2024 BGI Genomics Co.,Ltd.



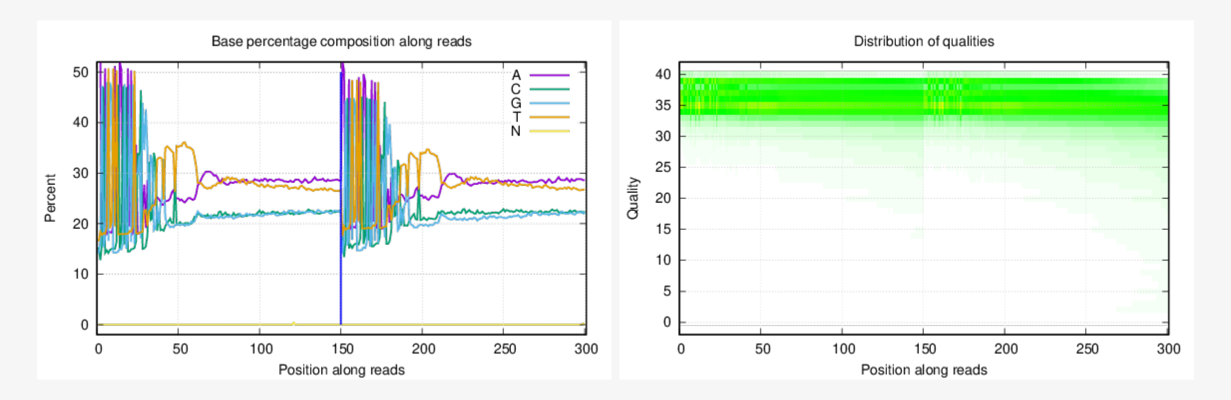
7_12_A



7_12_V1

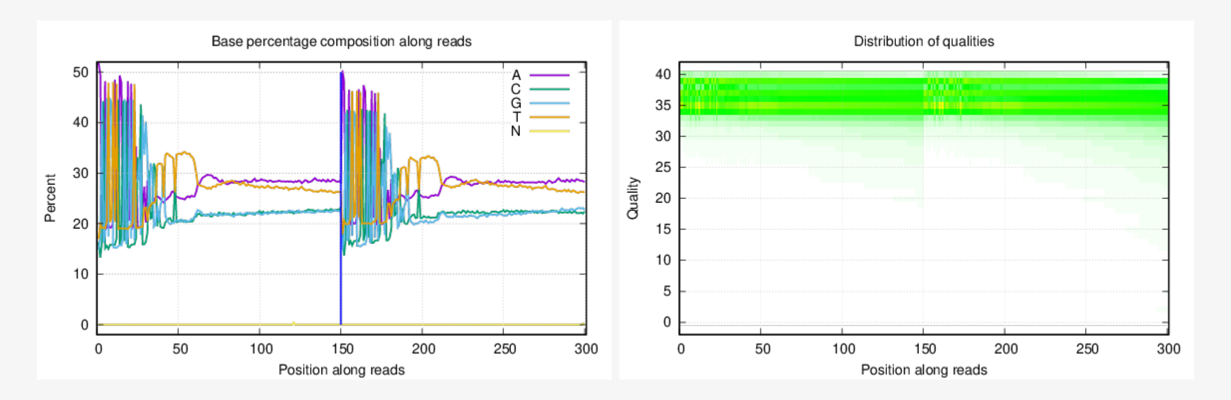


7_12_V2

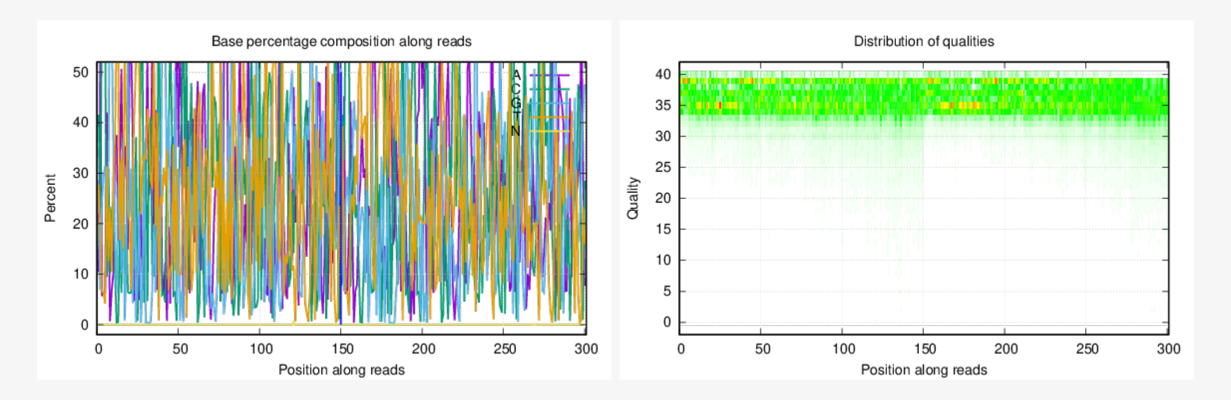


7_12_V3

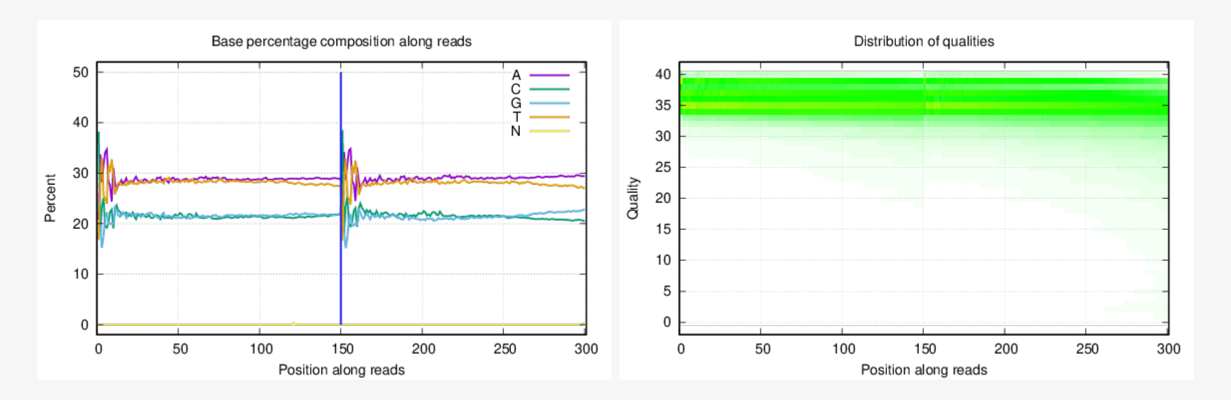
© 2024 BGI Genomics Co.,Ltd.



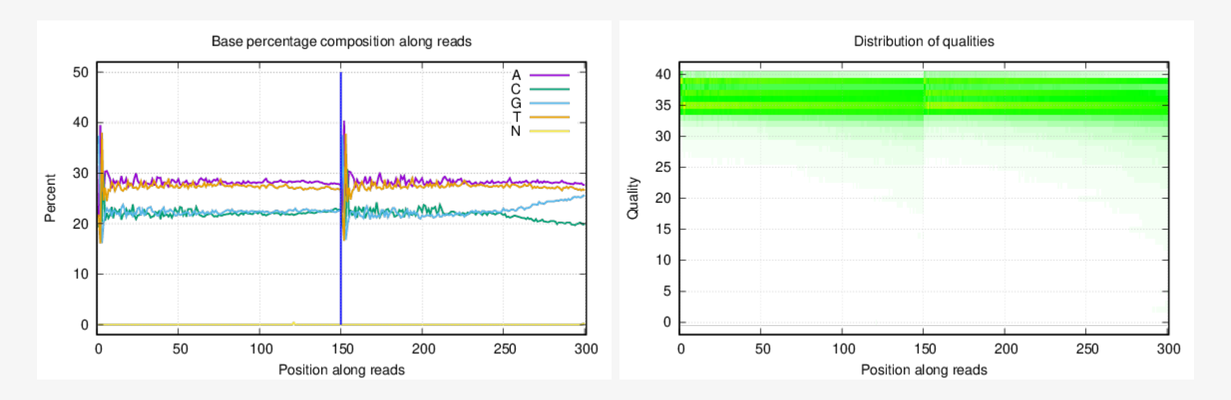
8



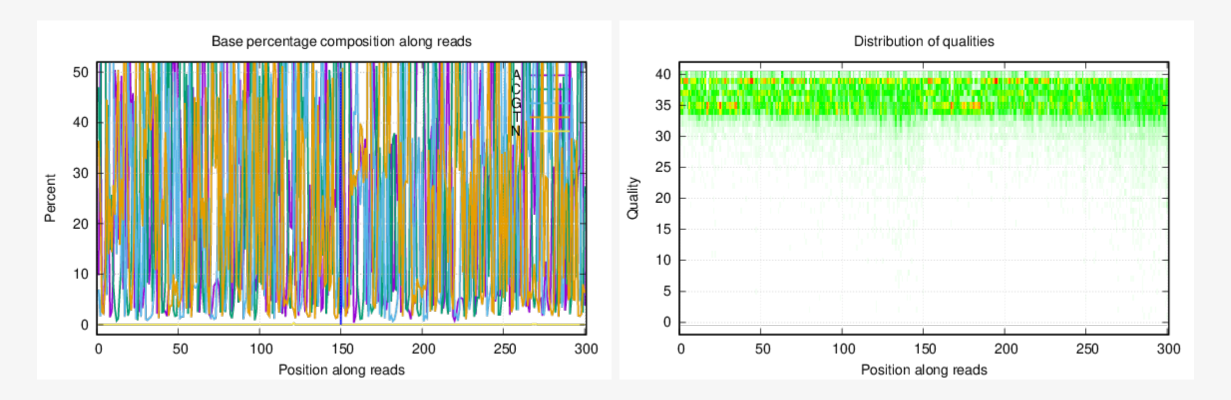
8388_3_K



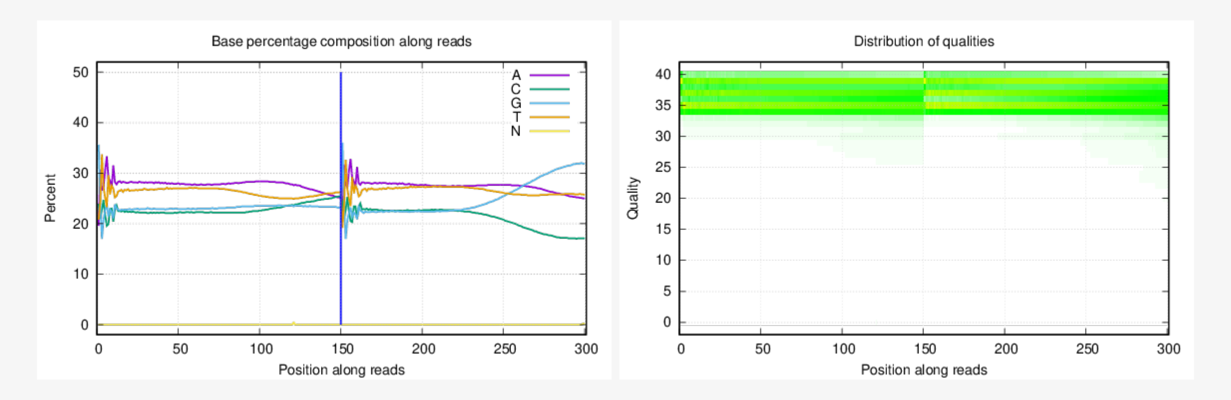
87c_Rep2



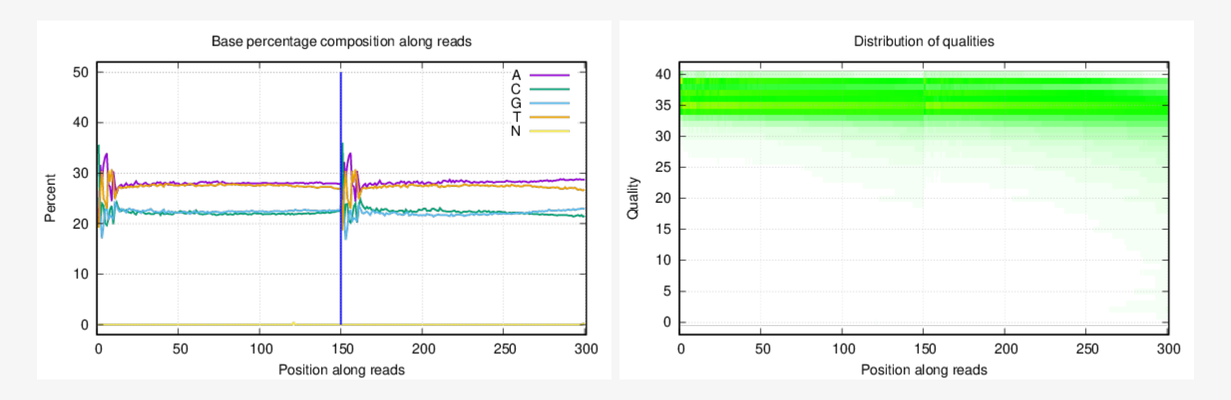
9



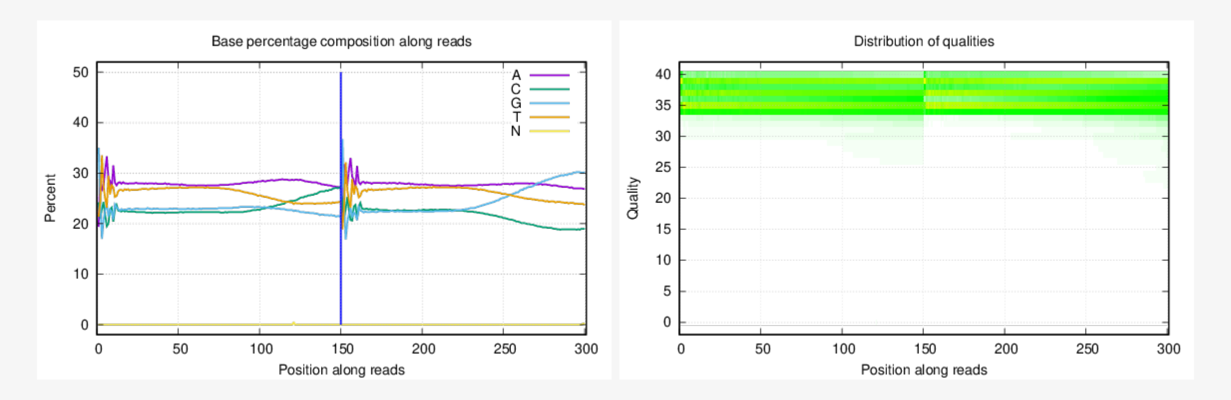
Arvalis



Lim1_K

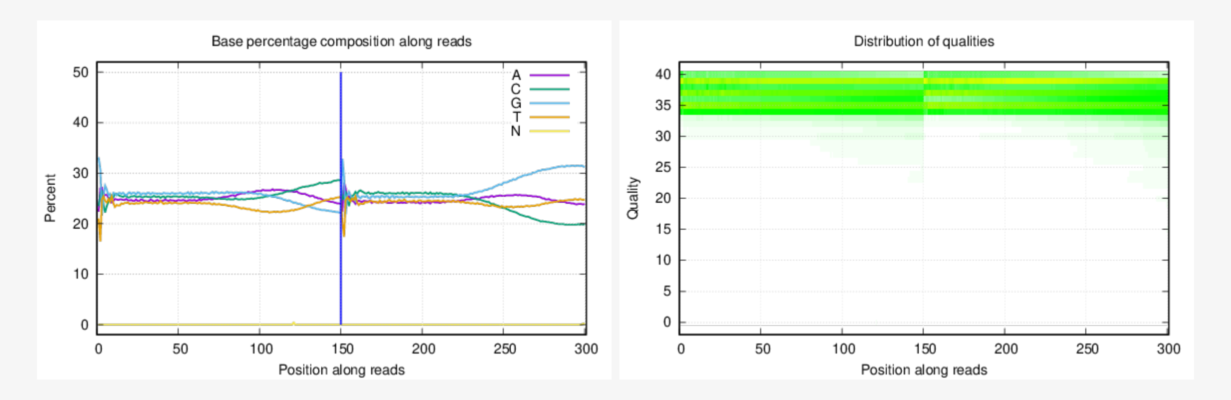


Oeconomus

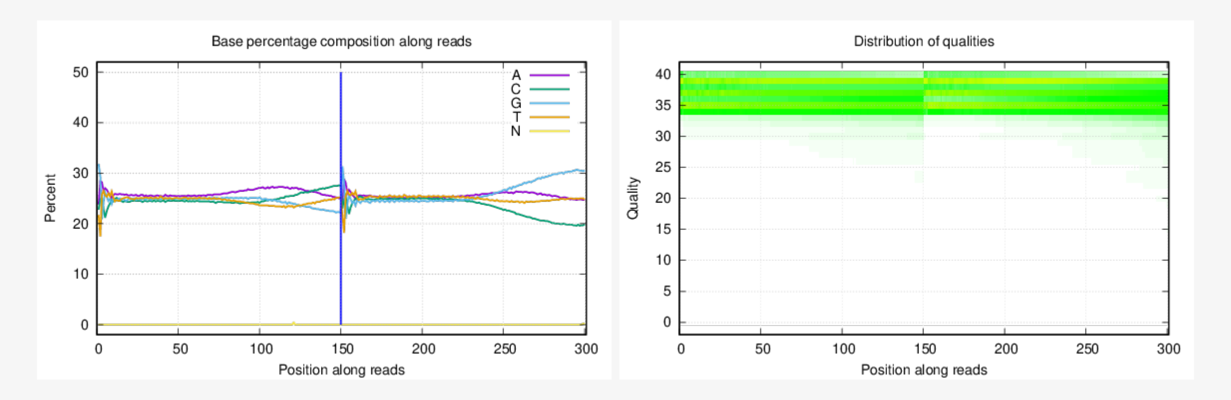


P10-six_S1

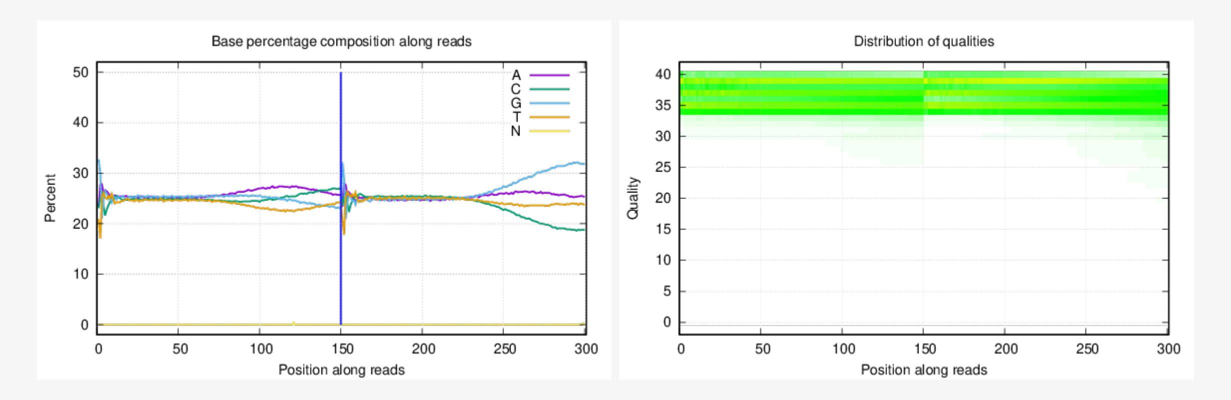
© 2024 BGI Genomics Co.,Ltd.



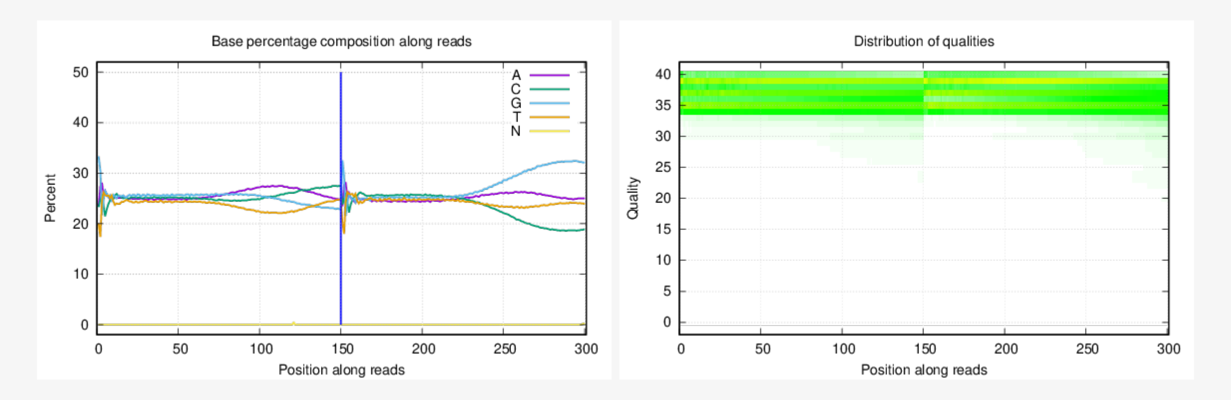
P142-six_S2



P143-six_S3

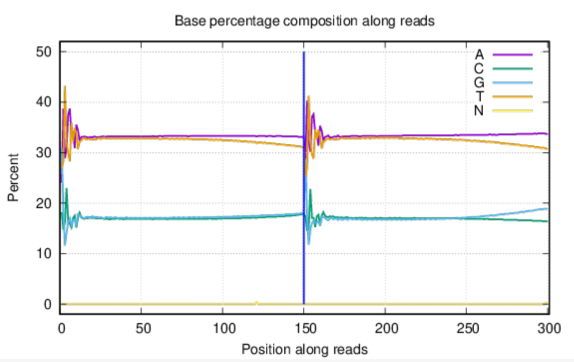


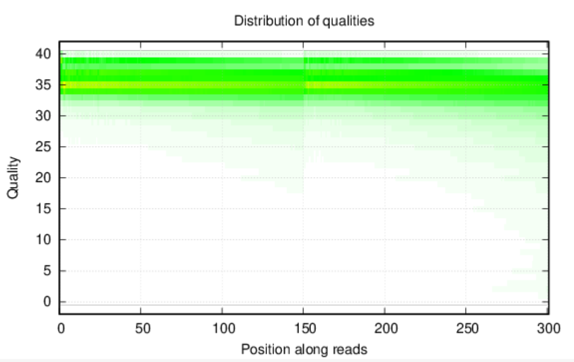
P181-six_S4



TA5

© 2024 BGI Genomics Co.,Ltd.





© 2024 BGI Genomics Co.,Ltd.

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	?

© 2024 BGI Genomics Co.,Ltd.

Omics For All

Contact us

Website: www.bgi.com
E-mail: 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.