Sequencing Data Report

Project: F24A430001331_MUSnxmoR
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 29 samples in this project, the statistics of fastq data is shown below.

Sample	Length	Q20(%)	Q30(%)	GC Content(%)	Total Reads	Total Bases
4C_dm6_nla_7_lig1000_1	150;150	97.90;98.00	93.25;93.39	45.78;46.82	648,806	194,641,800
4C_dm6_nla_7_lig1000_2	150;150	98.04;98.10	93.68;93.74	45.84;46.72	666,863	200,058,900
4C_dm6_nla_7_lig100_01	150;150	97.76;97.62	92.76;92.25	47.76;48.17	365,087	109,526,100
4C_dm6_nla_7_lig100_1_7_17	150;150	98.07;97.82	93.76;92.95	48.95;48.84	100,514	30,154,200
4C_dm6_nla_7_lig100_2_12_12	150;150	97.74;97.48	92.69;91.95	50.85;50.11	167,870	50,361,000
4C_dm6_nla_7_lig100_3_0_24	150;150	97.00;97.42	90.53;91.53	46.01;47.20	497,384	149,215,200
4C_dm6_nla_7_lig100_T4_1	150;150	97.93;97.84	93.31;92.88	45.81;46.91	801,306	240,391,800
4C_pct_cmv_prc2_dm6_nla_6	150;150	97.41;97.68	91.71;92.34	46.05;46.97	239,460	71,838,000
4C_pct_cmv_prc2_dm6_nla_7	150;150	97.52;97.53	92.10;92.01	48.03;48.65	291,792	87,537,600
74C_R2	150;150	97.38;96.96	91.50;89.68	42.46;42.47	21,960,276	6,588,082,800
P201_wgs_A1	150;150	98.05;98.52	93.78;94.97	42.02;42.27	344,100,927	103,230,278,100
P201_wgs_UIA	150;150	97.65;98.66	92.58;95.39	41.20;41.43	32,912	9,873,600
Prime_contr	150;150	97.47;98.33	91.87;94.17	42.84;42.74	57,231,024	17,169,307,200
Vul_219	150;150	98.50;99.14	95.30;97.23	47.84;47.94	1,299,908	389,972,400
Vul_238	150;150	98.50;99.30	95.22;97.66	45.70;45.85	2,266,031	679,809,300
Vul_343	150;150	98.53;99.23	95.36;97.52	47.18;47.28	1,583,806	475,141,800
Vul_377	150;150	98.56;99.19	95.41;97.40	46.77;46.91	1,796,630	538,989,000
Vul_384	150;150	98.58;99.30	95.46;97.65	46.57;46.78	2,072,080	621,624,000
Vul_389	150;150	98.54;99.23	95.39;97.51	47.65;47.85	1,337,582	401,274,600
Vul_408	150;150	98.67;99.30	95.73;97.69	46.42;46.62	1,523,438	457,031,400
Vul_491	150;150	98.46;99.23	95.14;97.48	47.29;47.42	1,608,678	482,603,400
Vul_492	150;150	98.67;99.21	95.82;97.44	48.23;48.46	1,241,968	372,590,400
Vul_496	150;150	98.50;99.04	95.26;96.94	48.54;48.48	1,665,259	499,577,700
Vul_497	150;150	98.61;99.17	95.67;97.36	47.50;47.46	1,594,820	478,446,000
Vul_517	150;150	98.53;99.22	95.36;97.50	48.13;48.26	1,410,003	423,000,900
Vul_519	150;150	98.63;99.18	95.63;97.34	47.13;47.20	1,446,997	434,099,100
Vul_nso1	150;150	98.35;99.14	94.84;97.23	47.41;47.53	1,043,534	313,060,200
Vul_nso2	150;150	98.20;99.18	94.32;97.39	46.08;46.18	1,889,072	566,721,600

Vul_nso3	150;150	98.48;99.16	95.20;97.34	46.26;46.38	969,718	290,915,400

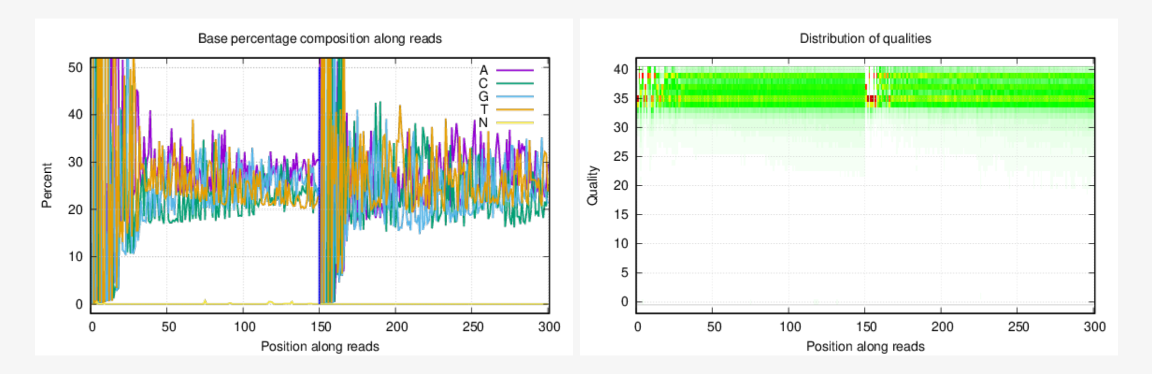
Table Format:

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

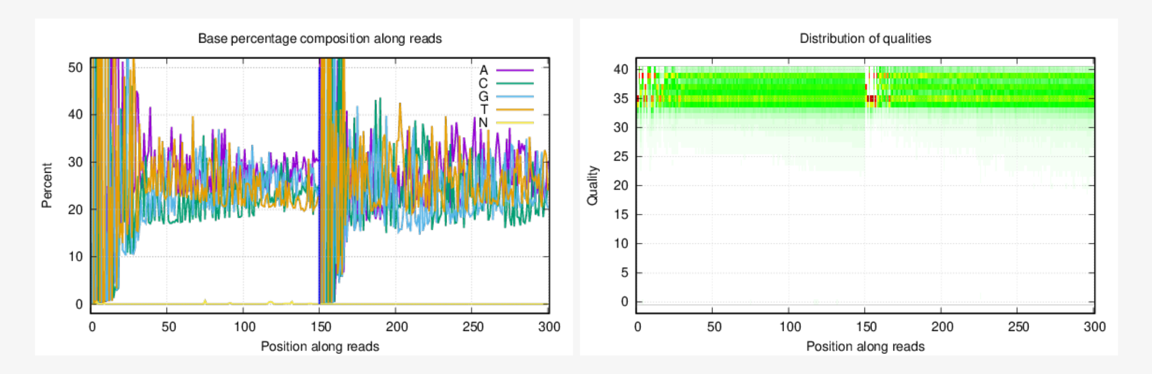
Data Quality Control

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.

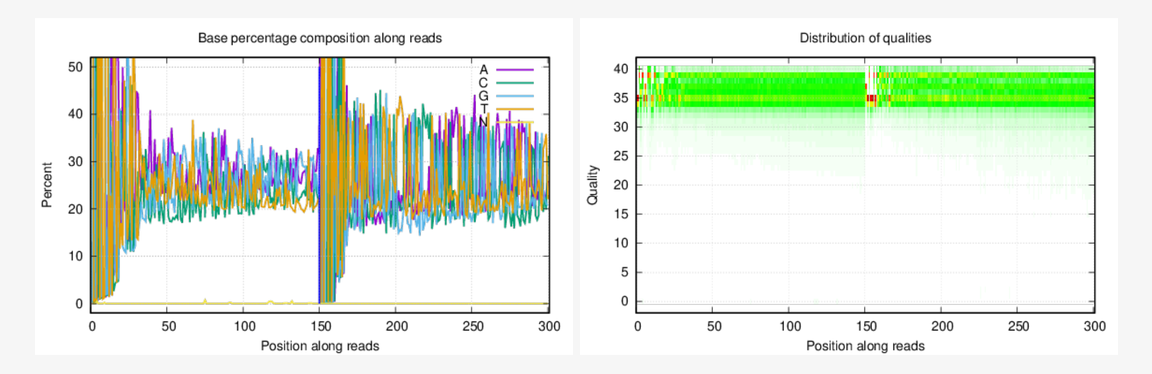
4C_dm6_nla_7_lig1000_1



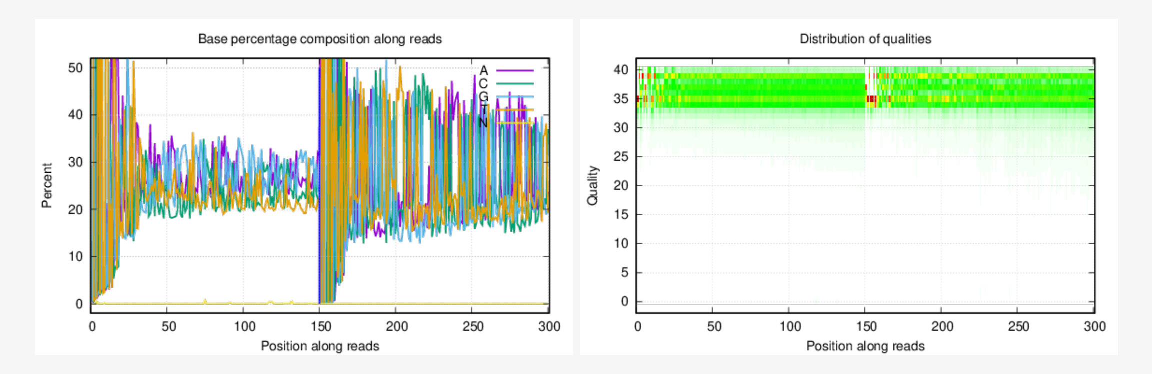
4C_dm6_nla_7_lig1000_2



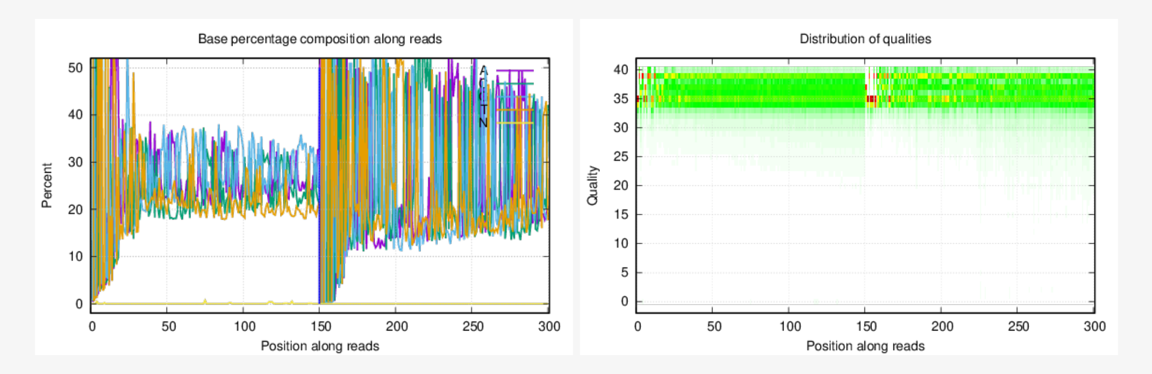
4C_dm6_nla_7_lig100_01



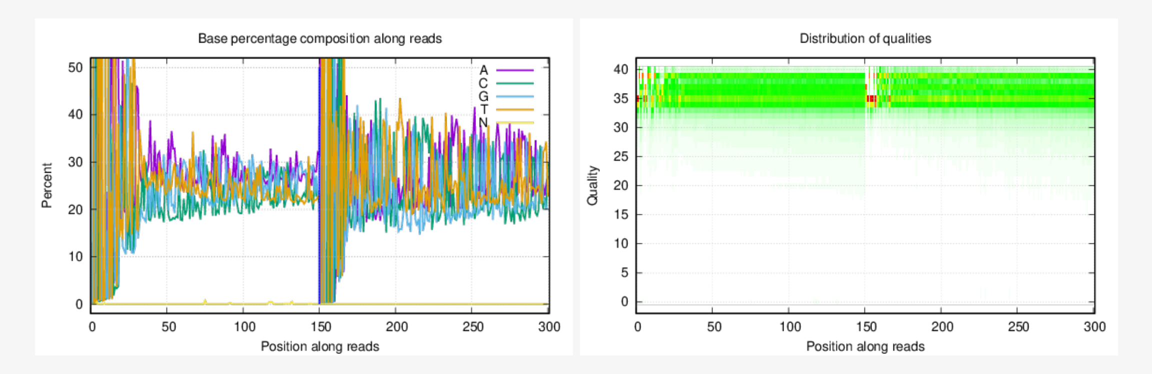
4C_dm6_nla_7_lig100_1_7_17



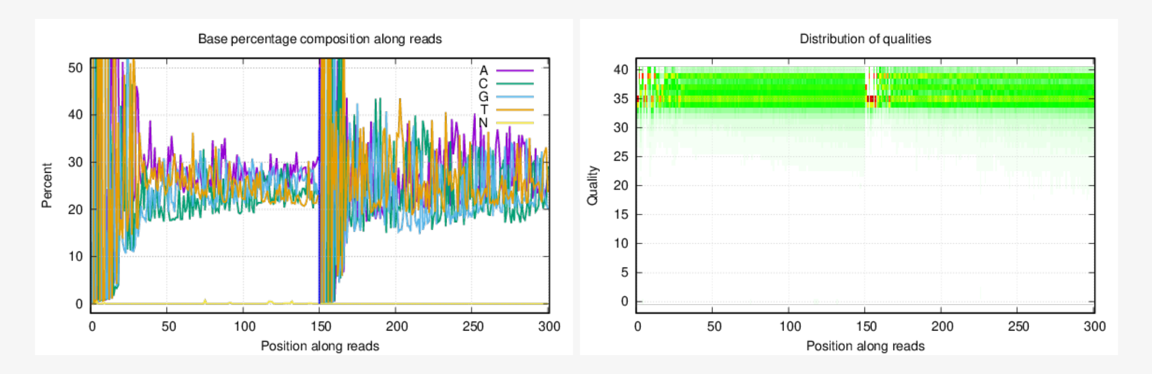
4C_dm6_nla_7_lig100_2_12_12



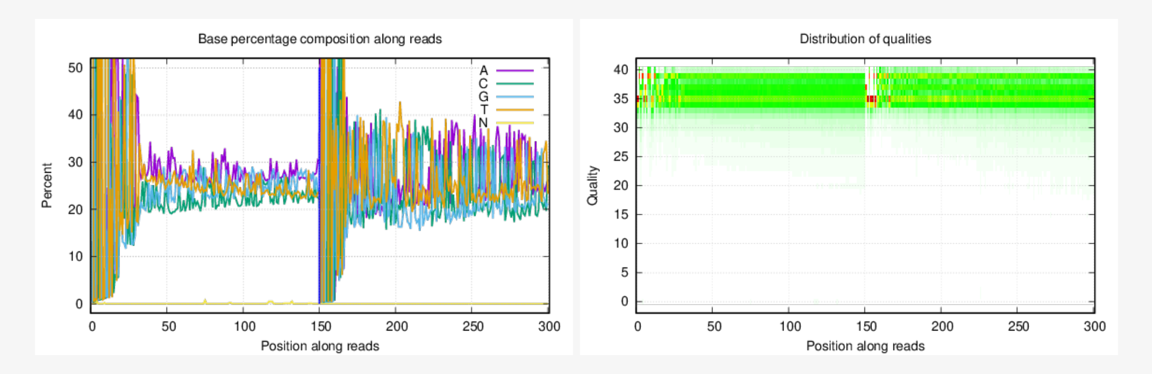
4C_dm6_nla_7_lig100_3_0_24



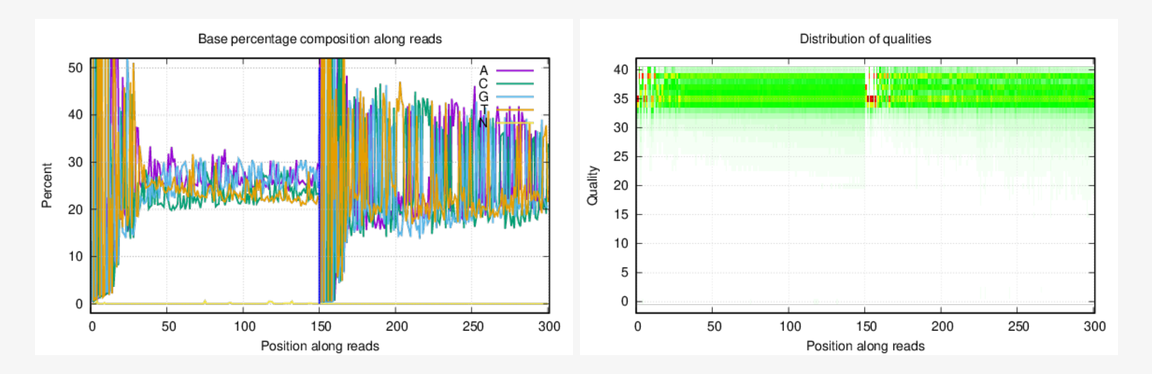
4C_dm6_nla_7_lig100_T4_1



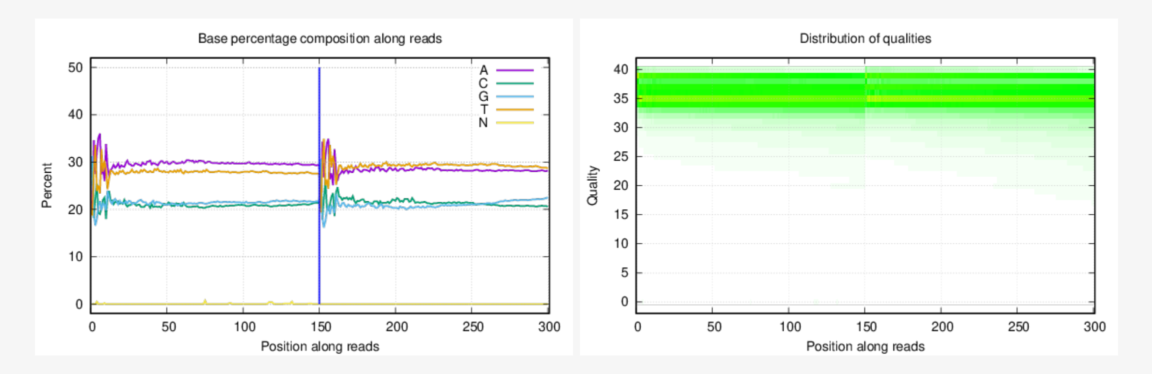
4C_pct_cmv_prc2_dm6_nla_6



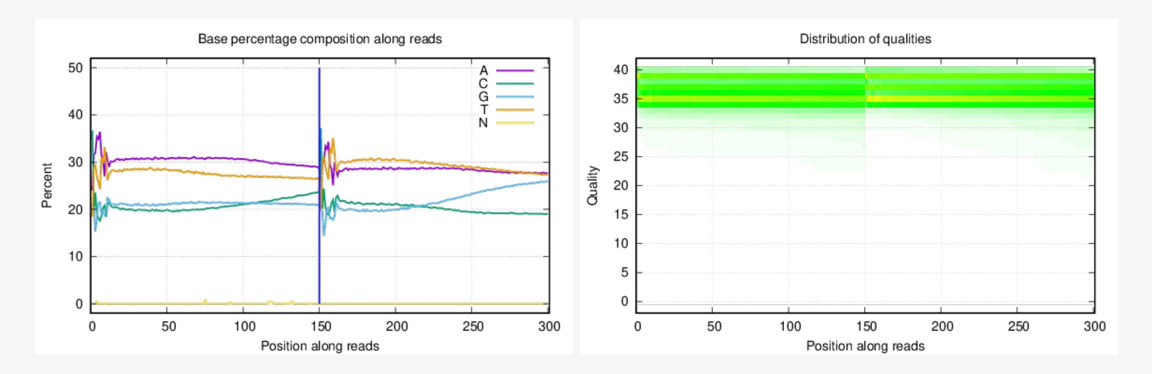
4C_pct_cmv_prc2_dm6_nla_7



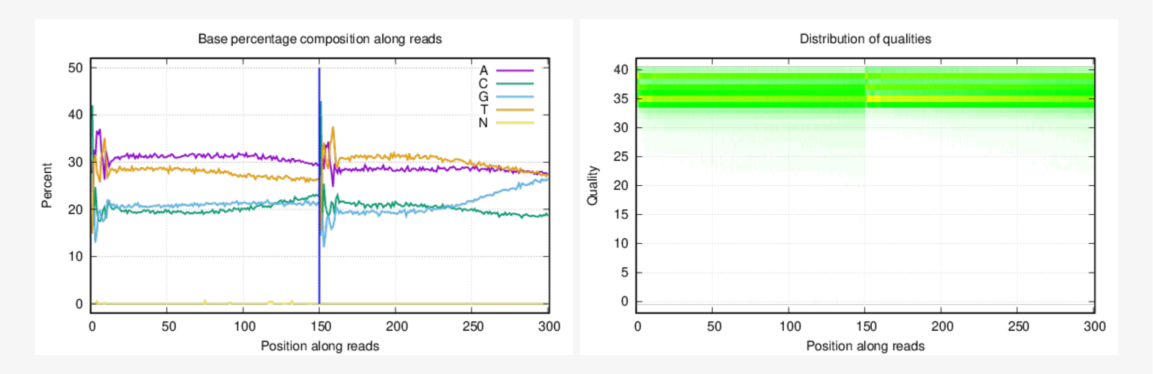
74C_R2



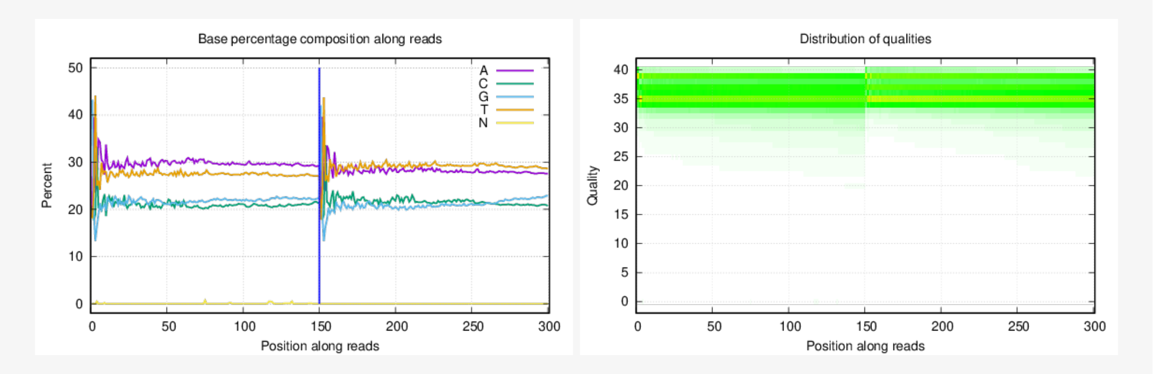
P201_wgs_A1



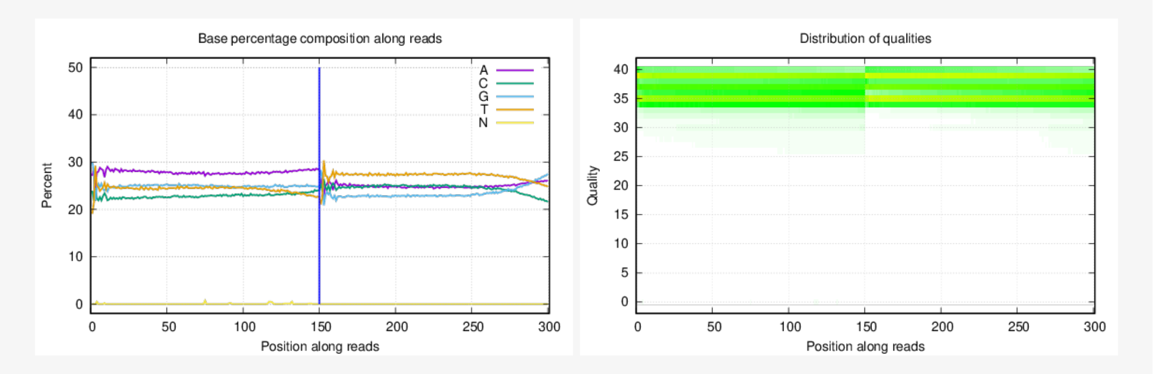
P201_wgs_UIA



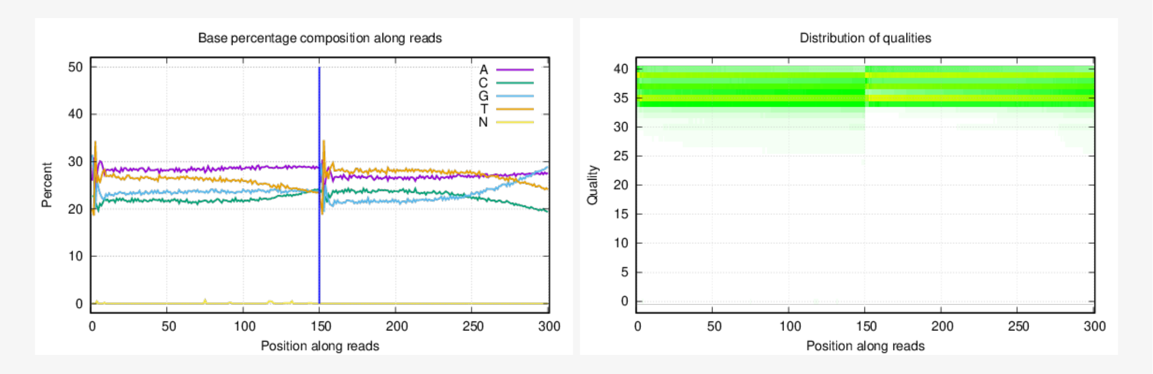
Prime_contr



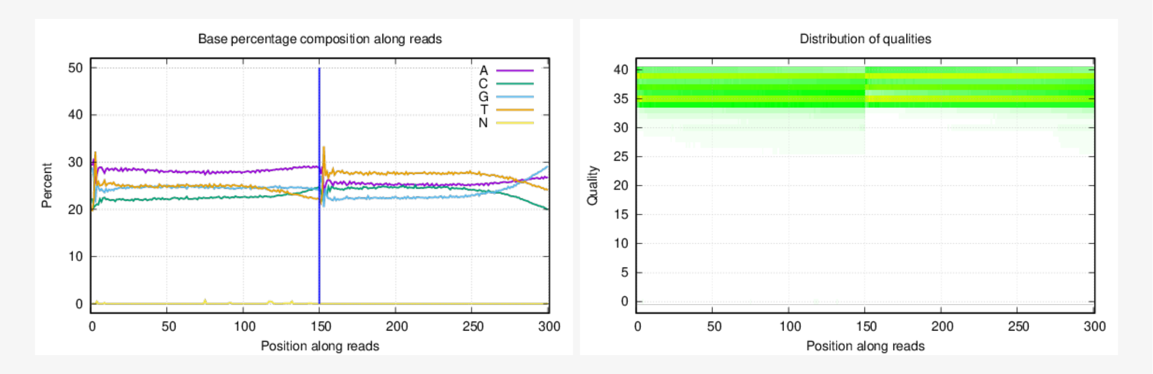
Vul_219



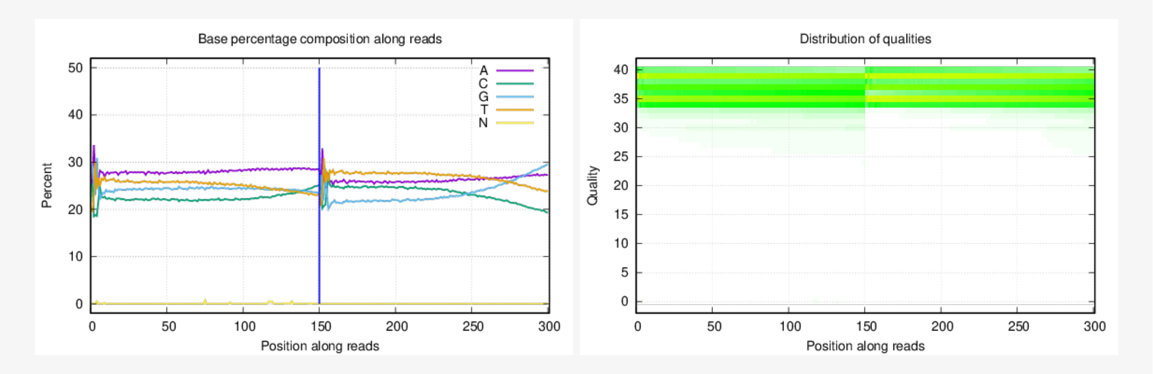
Vul_238



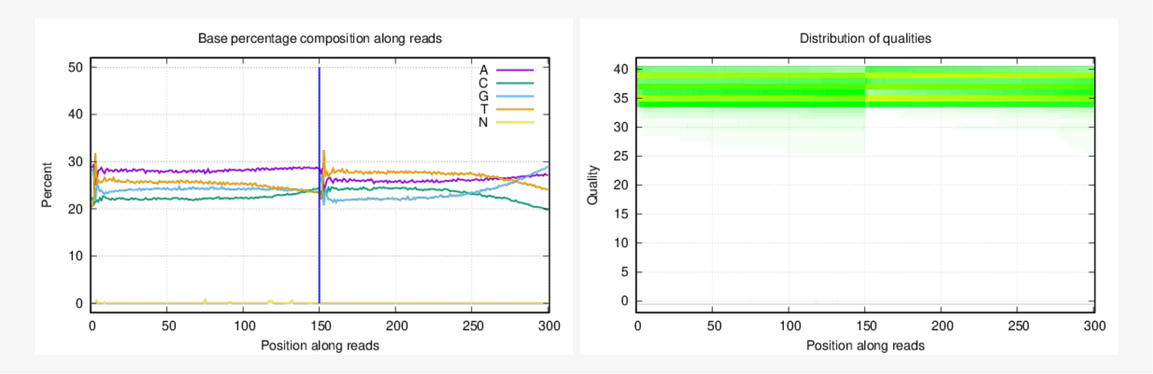
Vul_343



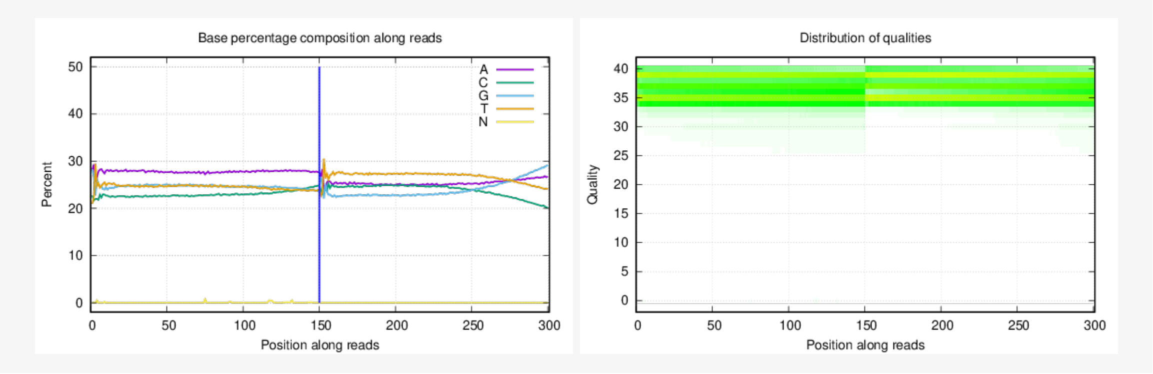
Vul_377



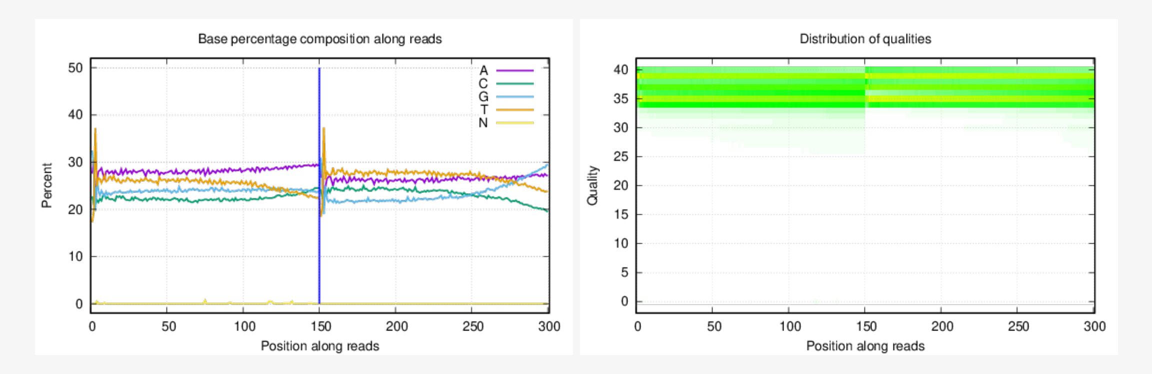
Vul_384



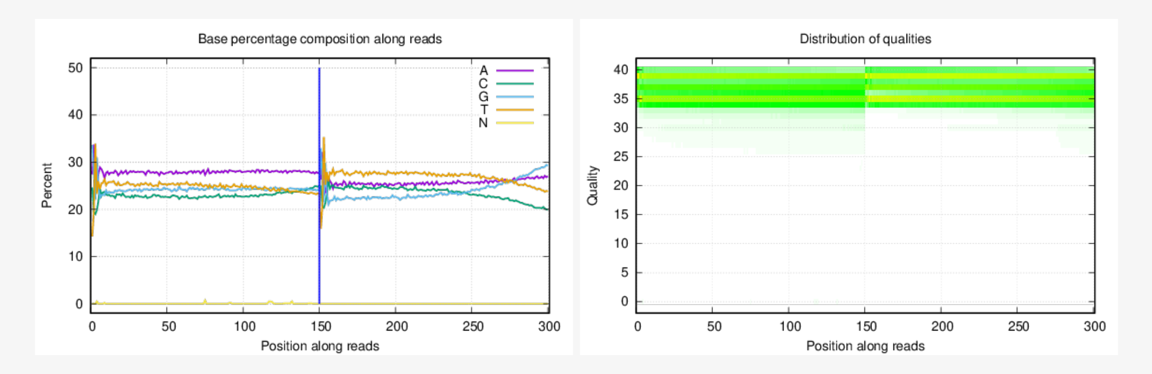
Vul_389



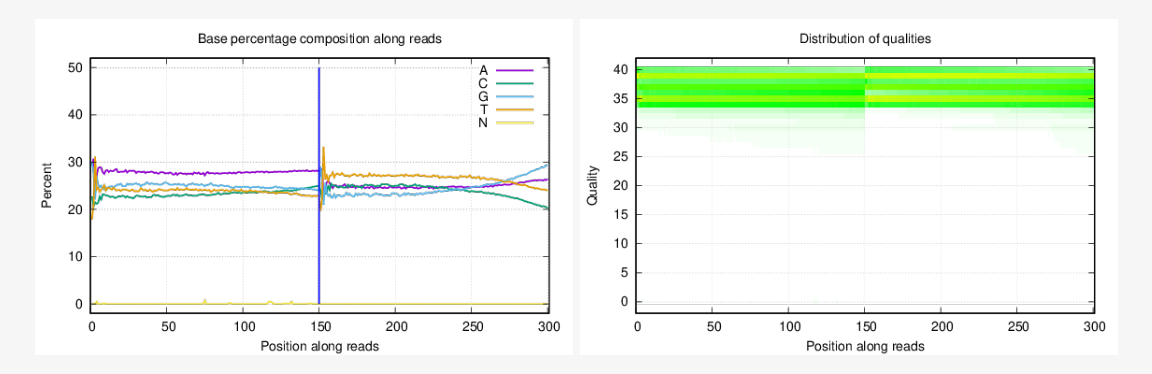
Vul_408



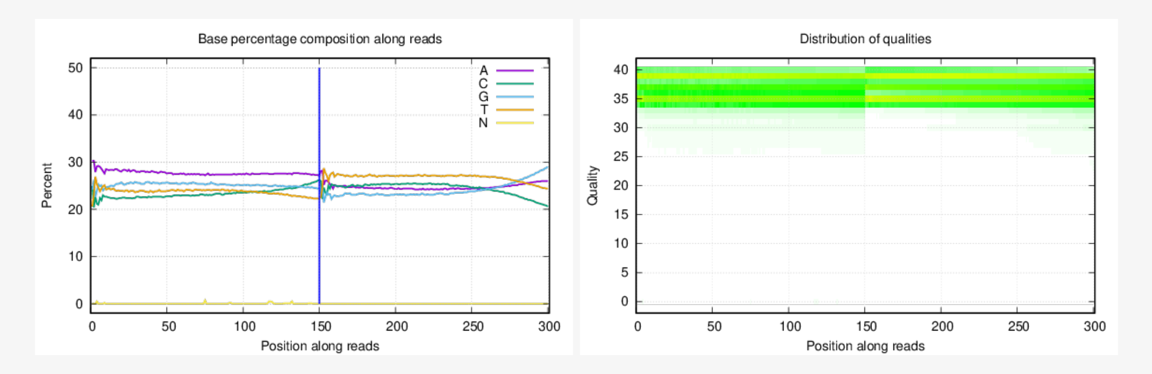
Vul_491



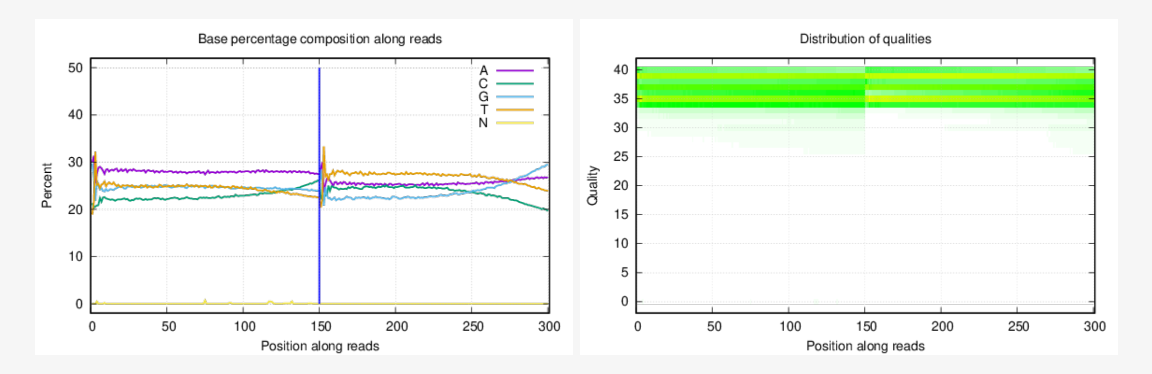
Vul_492



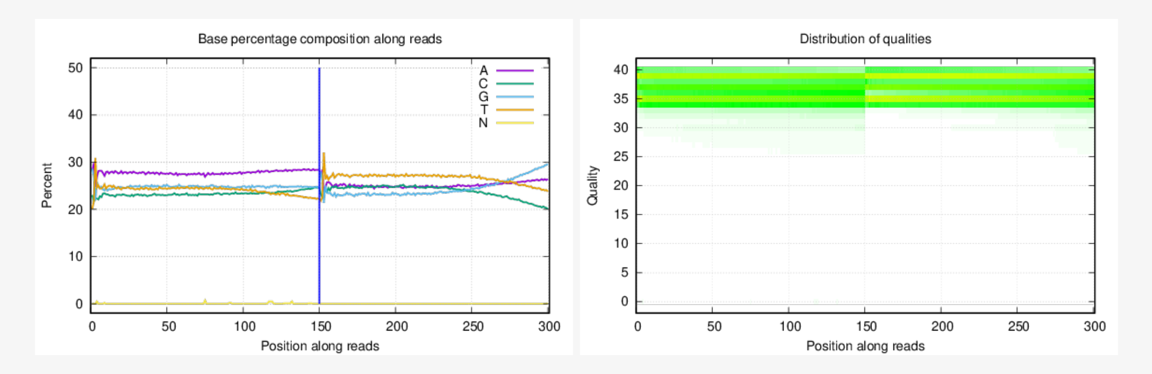
Vul_496



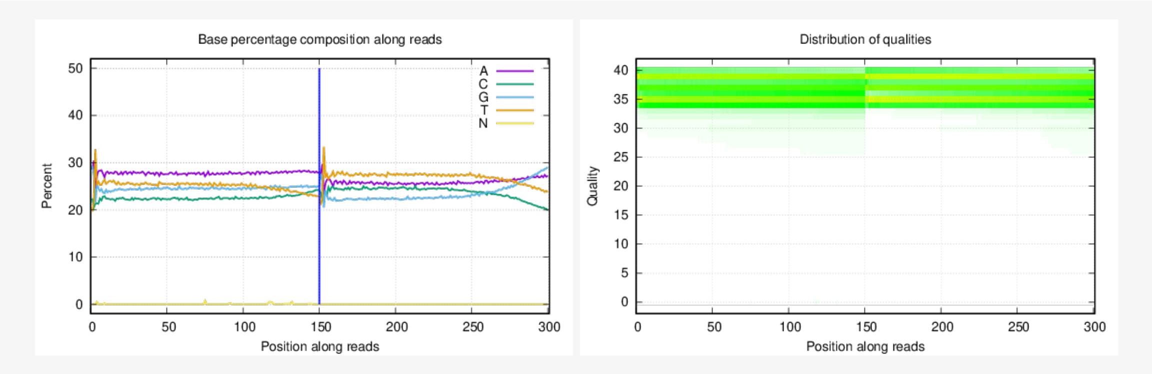
Vul_497



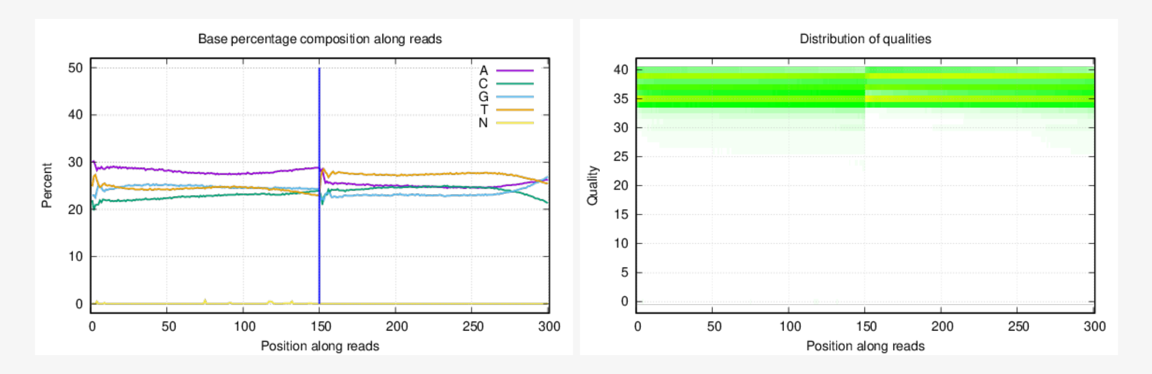
Vul_517



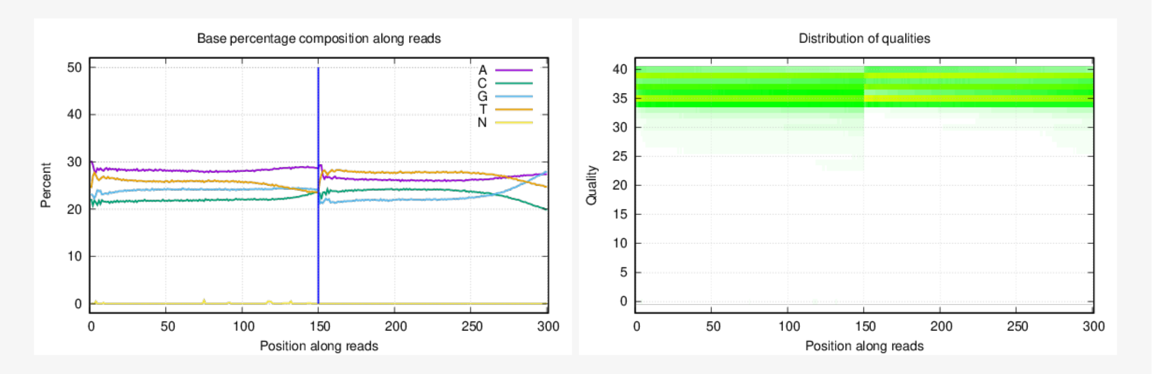
Vul_519



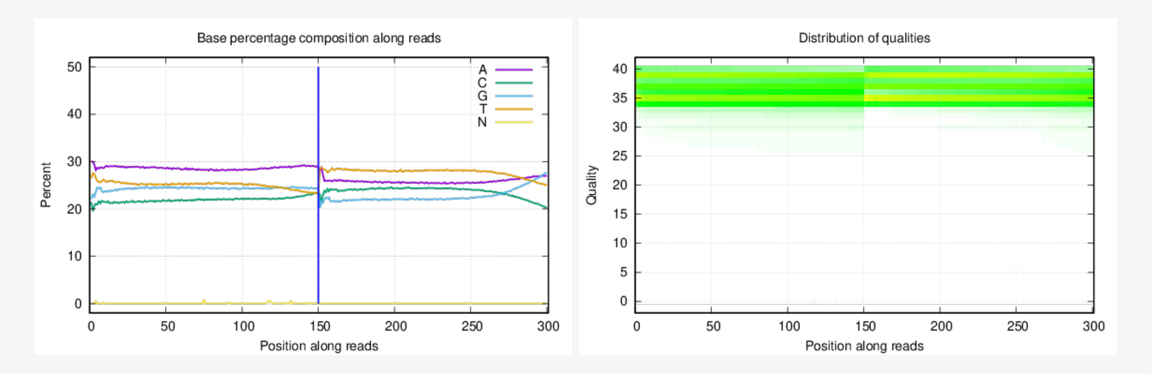
Vul_nso1



Vul_nso2


Base percentage composition along reads
Percent
Position along reads
A
C
G
T
N
0
10
20
30
40
50
0
50
100
150
200
250
300
Distribution of qualities
Quality
Position along reads
0
5
10
15
20
25
30
35
40

Vul_nso3


Base percentage composition along reads
Percent
Position along reads
A
C
G
T
N
0
10
20
30
40
50
0
50
100
150
200
250
300
Distribution of qualities
Quality
Position along reads
0
5
10
15
20
25
30
35
40

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