Sequencing Data Report

Project: F23A430001415_GRAqrimR
Date: 2024.4.7
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 37 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	97.30;97.87	91.75;93.33	45.00;44.87	39,038,053	11,711,415,900
11	150;150	97.32;97.84	91.94;93.60	45.32;45.37	44,944,669	13,483,400,700
12	150;150	97.68;97.99	92.73;93.88	44.48;44.32	39,561,203	11,868,360,900
13	150;150	97.63;98.23	92.64;94.49	44.81;44.63	48,379,149	14,513,744,700
15	150;150	97.77;98.27	93.02;94.57	44.92;44.84	64,385,103	19,315,530,900
16	150;150	97.85;98.35	93.28;94.80	44.87;44.77	53,656,211	16,096,863,300
17	150;150	97.39;96.99	91.73;90.67	44.78;44.64	25,143,279	7,542,983,700
18	150;150	97.35;97.85	91.90;93.34	44.90;44.73	34,636,313	10,390,893,900
19	150;150	97.87;98.17	93.32;94.46	44.01;43.84	31,540,851	9,462,255,300
2	150;150	97.23;98.08	91.68;94.02	45.41;45.28	43,933,668	13,180,100,400
20	150;150	97.75;98.02	92.90;93.86	44.34;44.14	43,434,435	13,030,330,500
21	150;150	97.85;98.33	93.28;94.78	44.74;44.57	37,242,787	11,172,836,100
22	150;150	97.86;98.30	93.27;94.72	44.68;44.55	37,843,742	11,353,122,600
23	150;150	97.60;97.88	92.46;93.32	44.90;44.78	47,213,046	14,163,913,800
24	150;150	97.96;98.59	93.67;95.57	44.81;44.65	43,293,327	12,987,998,100
26	150;150	97.61;98.20	92.69;94.40	44.57;44.47	35,997,065	10,799,119,500
29	150;150	97.54;98.08	92.48;94.07	44.65;44.51	35,370,858	10,611,257,400
32	150;150	97.56;98.39	92.57;94.95	44.31;44.21	47,158,489	14,147,546,700
38	150;150	97.51;98.12	92.42;94.24	44.53;44.38	37,059,023	11,117,706,900
4	150;150	97.48;98.29	92.35;94.67	44.98;44.83	37,033,811	11,110,143,300
7	150;150	97.34;97.82	91.84;93.19	45.44;45.30	28,994,332	8,698,299,600
7369_Nod2_1	150;150	98.30;98.53	94.69;95.21	49.53;49.37	2,739,172	821,751,600
8	150;150	97.30;97.97	91.73;93.54	45.60;45.47	42,878,374	12,863,512,200
9	150;150	97.24;97.84	91.64;93.42	44.73;44.59	26,165,345	7,849,603,500
ES_RIGS	150;150	98.25;98.98	94.06;96.29	54.25;41.35	41,478	12,443,400
ES_RIGS1	150;150	98.44;98.94	94.68;96.17	53.83;42.47	65,341	19,602,300

ES_RIGS10	150;150	97.75;99.03	92.63;96.48	54.05;41.58	11,562	3,468,600
ES_RIGS11	150;150	98.53;99.05	94.98;96.53	54.00;41.35	71,985	21,595,500
ES_RIGS12	150;150	97.91;99.01	93.32;96.37	54.18;41.35	65,338	19,601,400
ES_RIGS2	150;150	98.17;98.94	93.90;96.11	53.93;41.46	67,947	20,384,100
ES_RIGS3	150;150	98.19;99.05	93.95;96.51	53.69;43.06	65,756	19,726,800
ES_RIGS5	150;150	98.34;98.91	94.39;96.09	53.93;41.61	62,794	18,838,200
ES_RIGS7	150;150	97.89;98.92	93.09;96.05	54.12;41.54	20,763	6,228,900
ES_RIGS8	150;150	98.28;99.01	94.16;96.39	54.02;41.83	73,386	22,015,800
ES_RIGS9	150;150	98.52;98.96	94.96;96.24	54.19;42.08	66,834	20,050,200
Shur1_e	150;150	97.76;97.61	92.09;91.21	42.33;39.00	22	6,600
Zap_e2	150;150	97.12;96.54	91.01;88.62	42.26;42.28	8,061,847	2,418,554,100

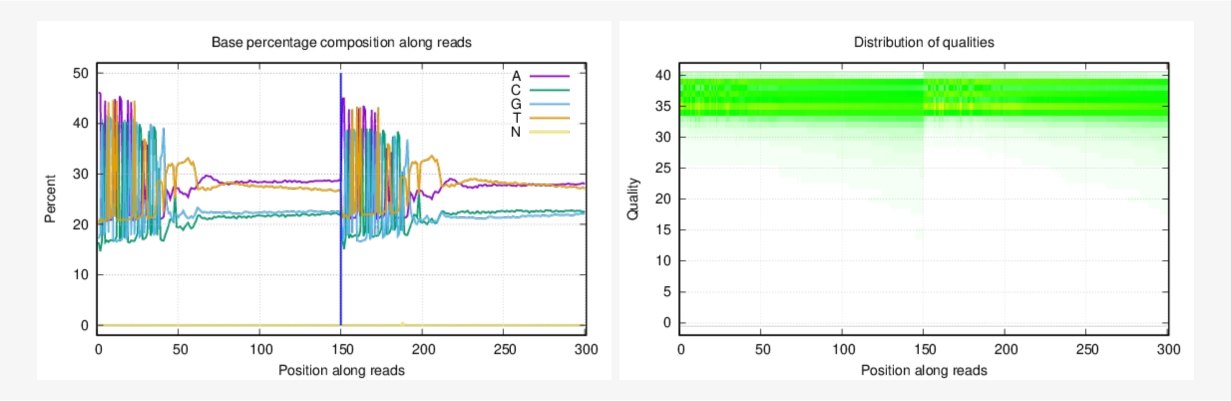
Table Format:

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

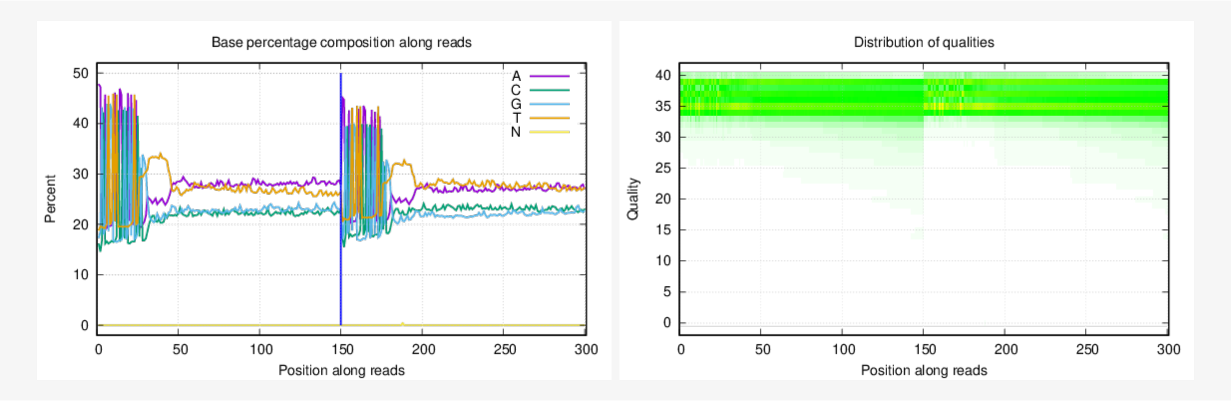
Data Quality Control

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.

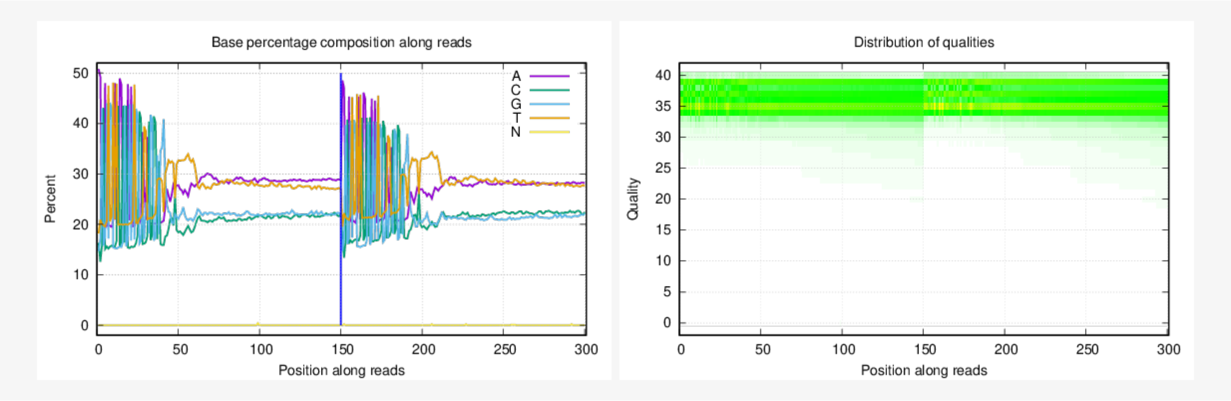
1



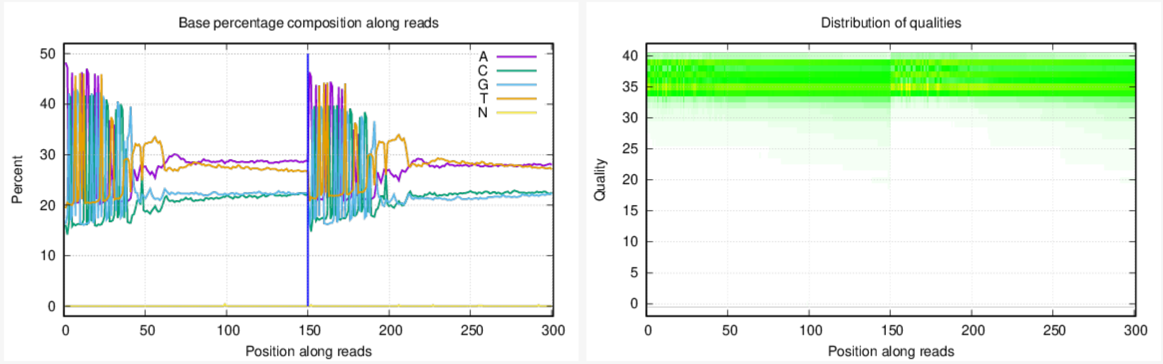
11



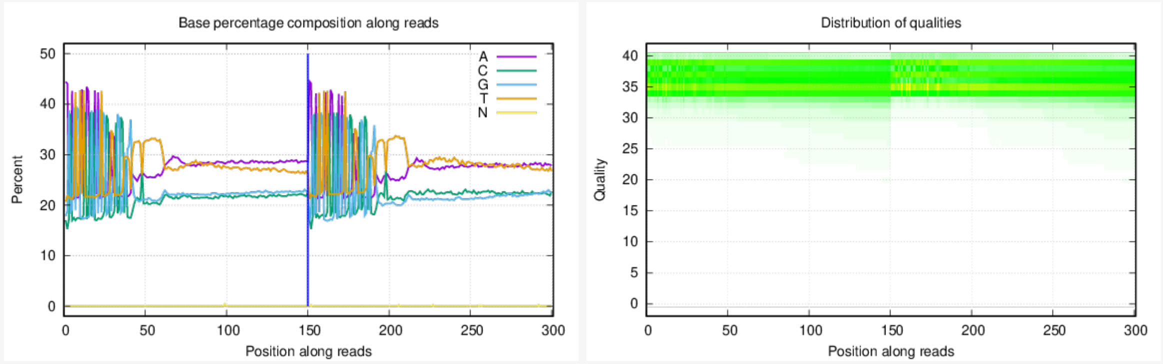
12



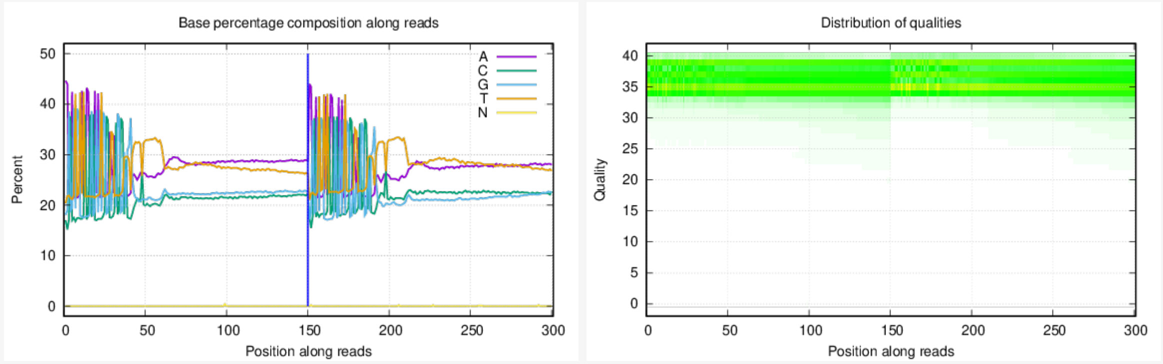
13


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

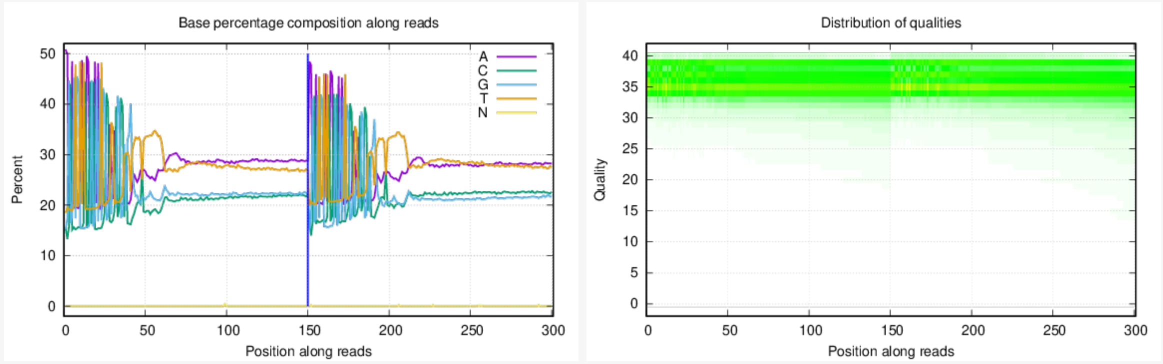
15


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

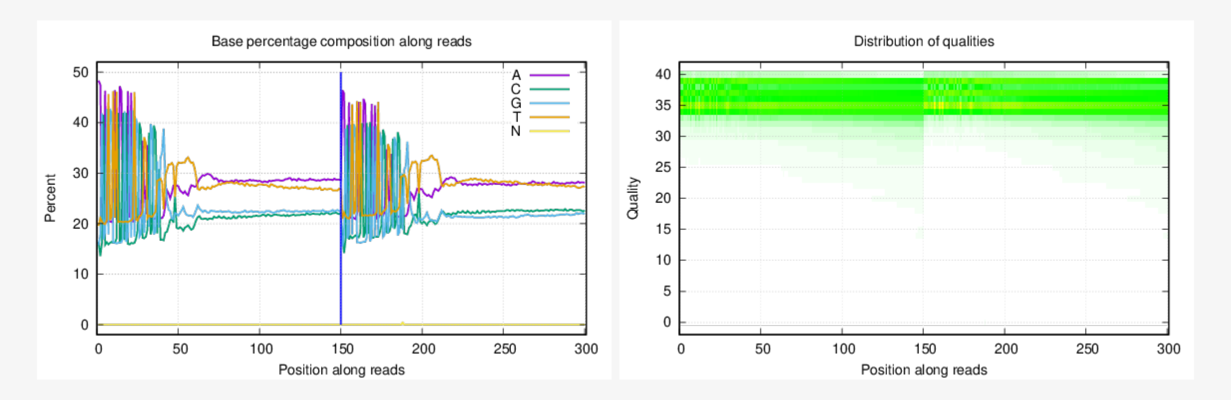
16


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

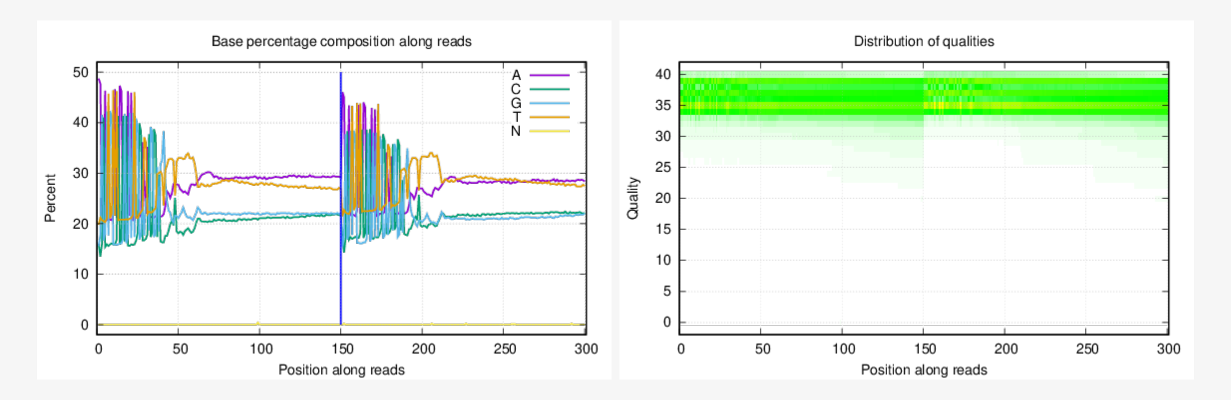
17


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

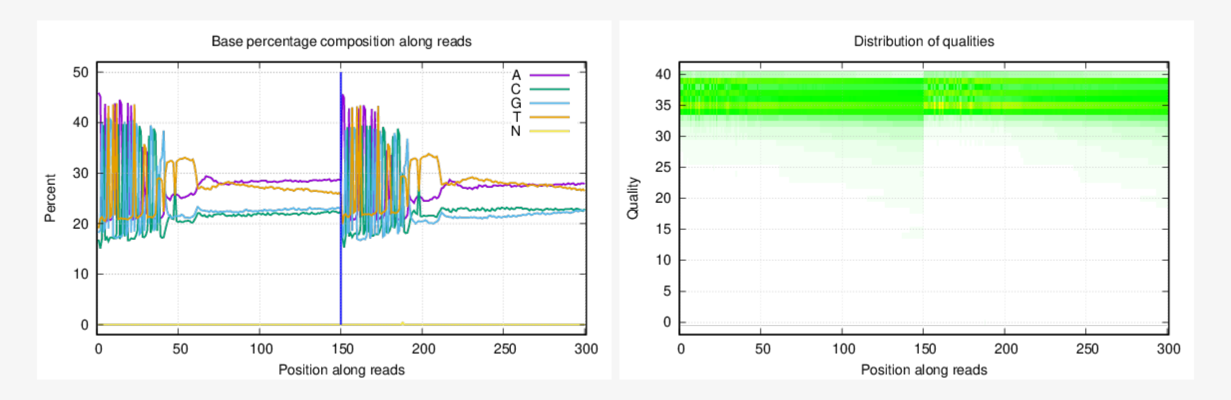
18


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

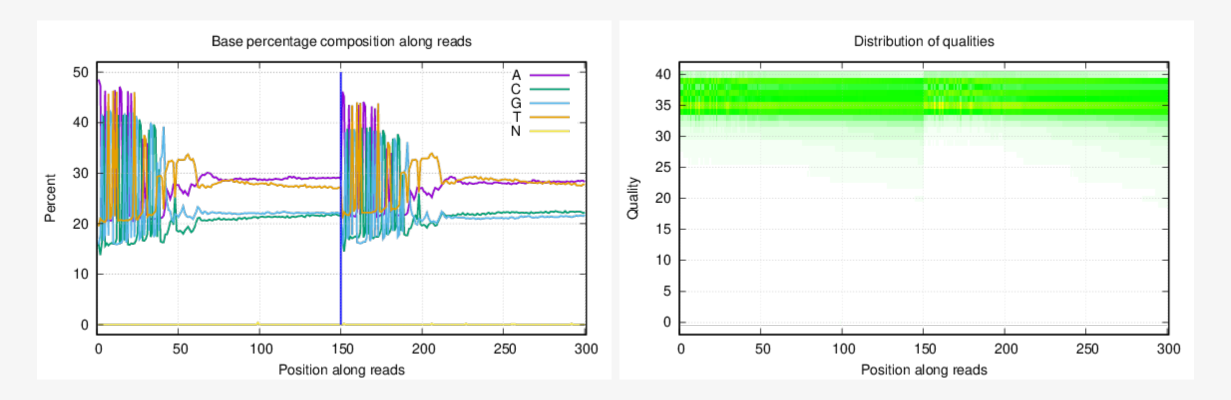
19


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

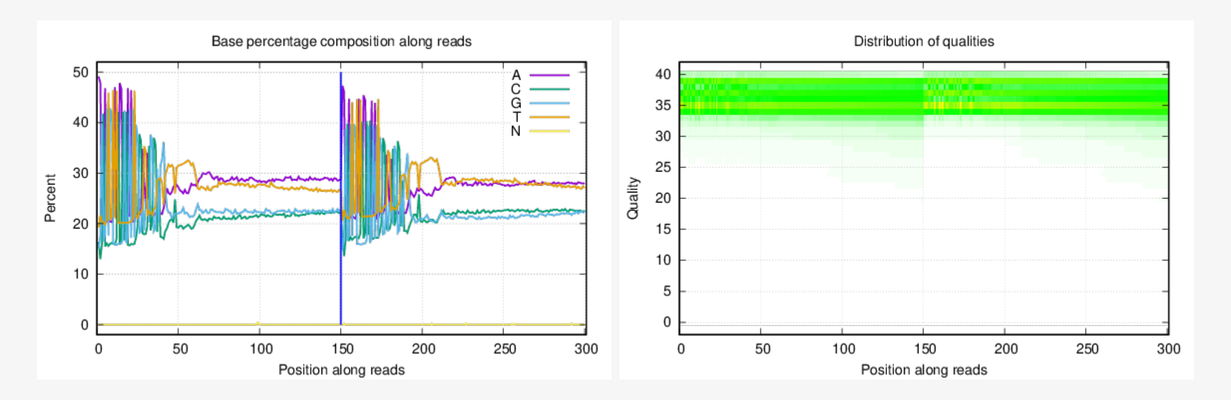
2


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

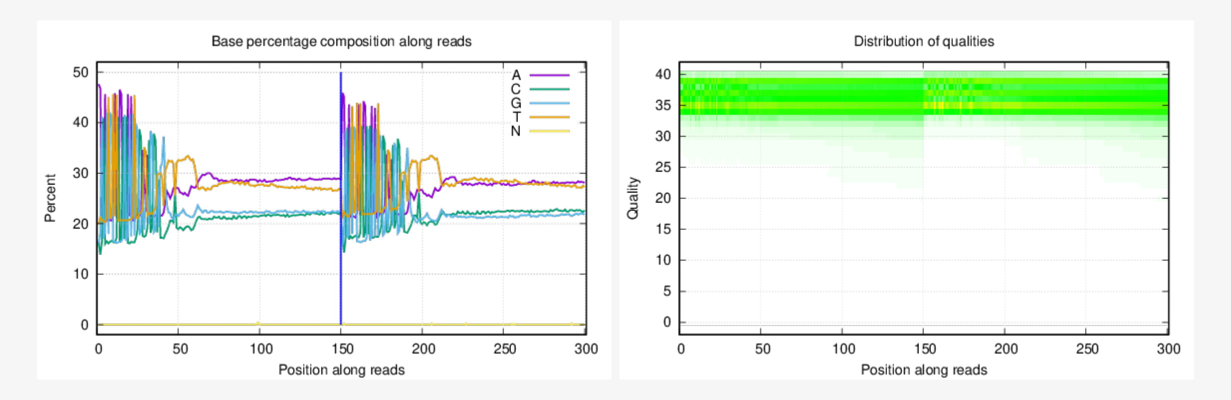
20


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

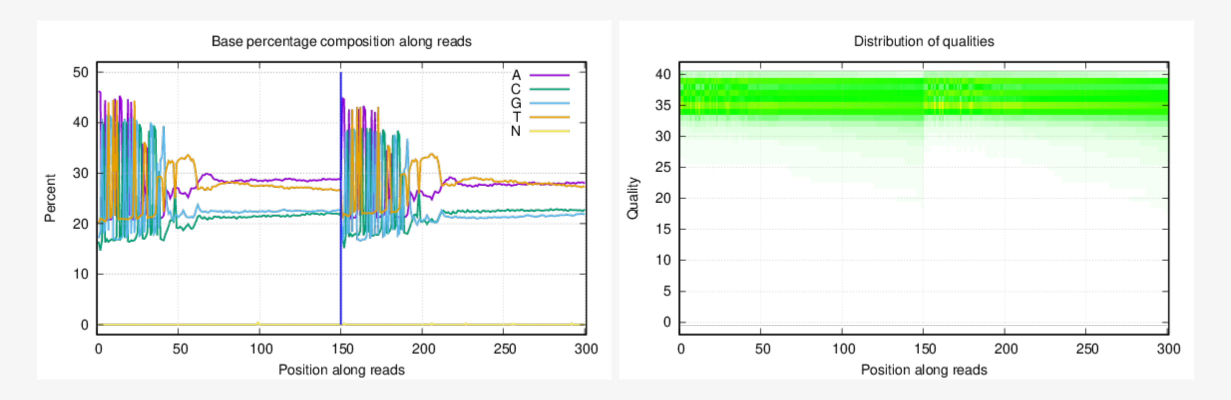
21


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

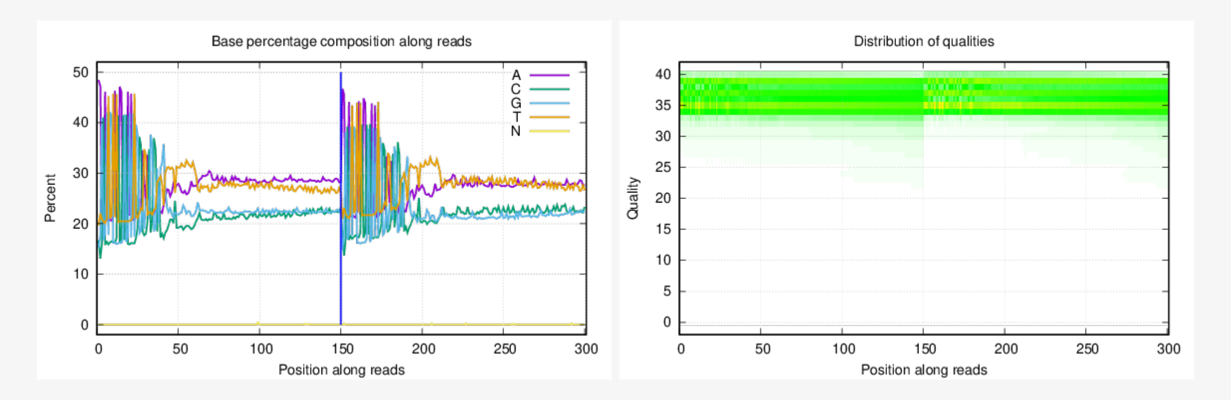
22


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

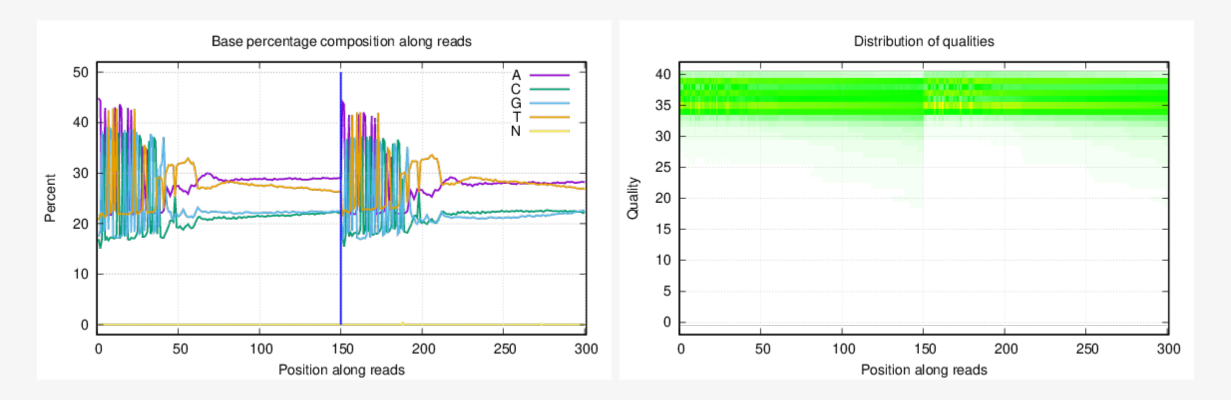
23


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

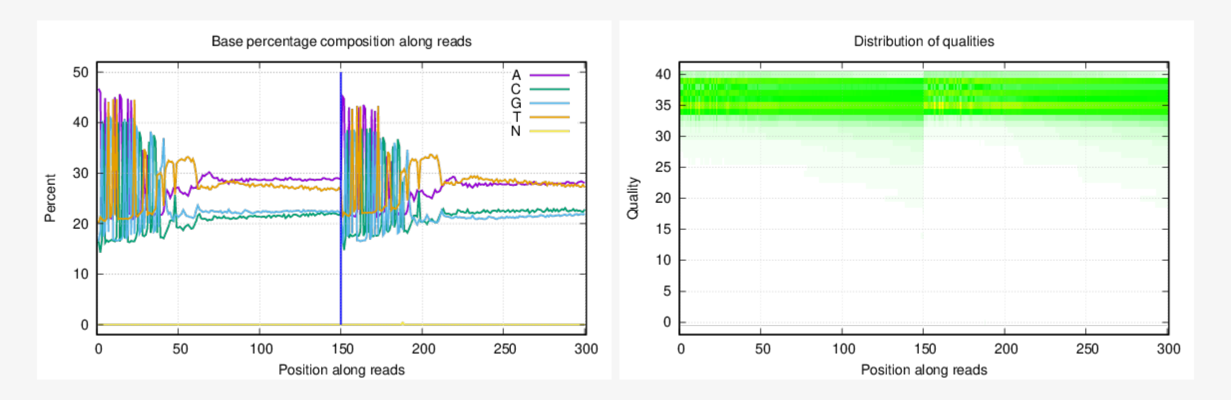
24


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

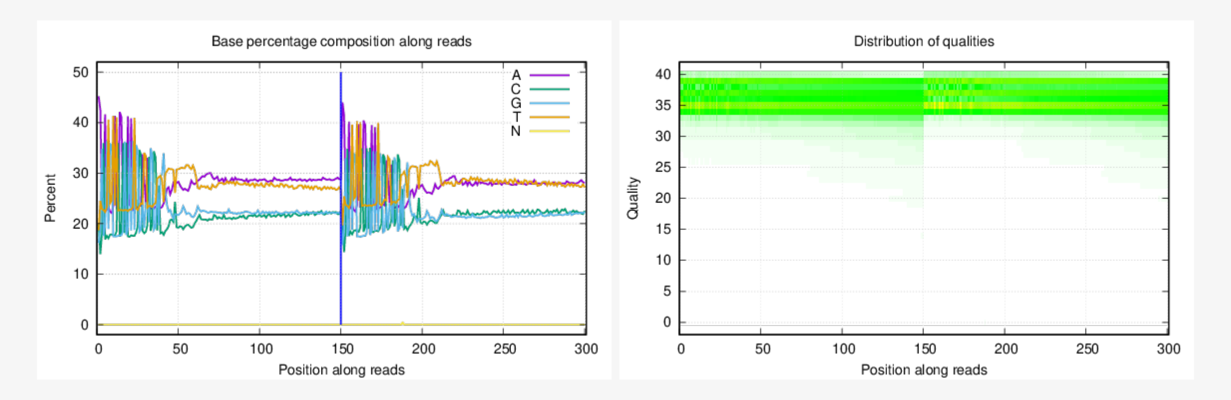
26


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

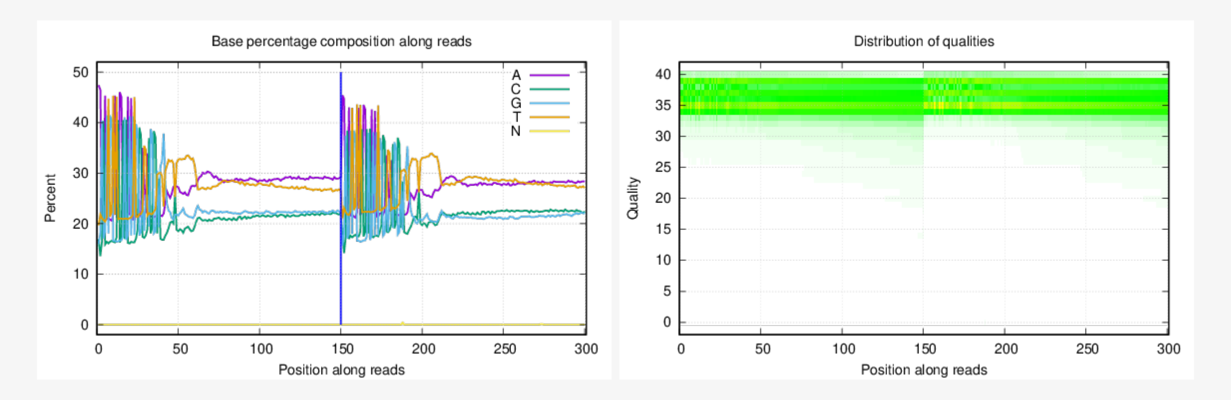
29


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

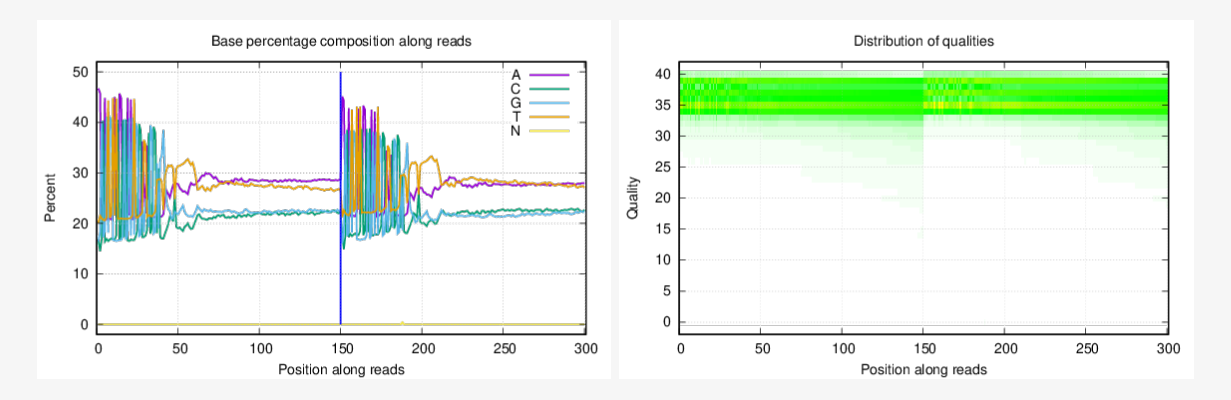
32


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

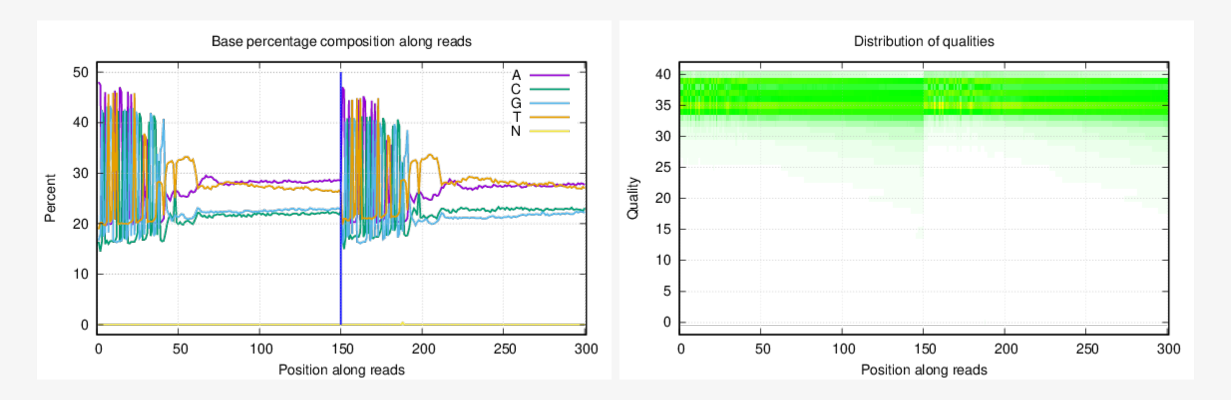
38


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

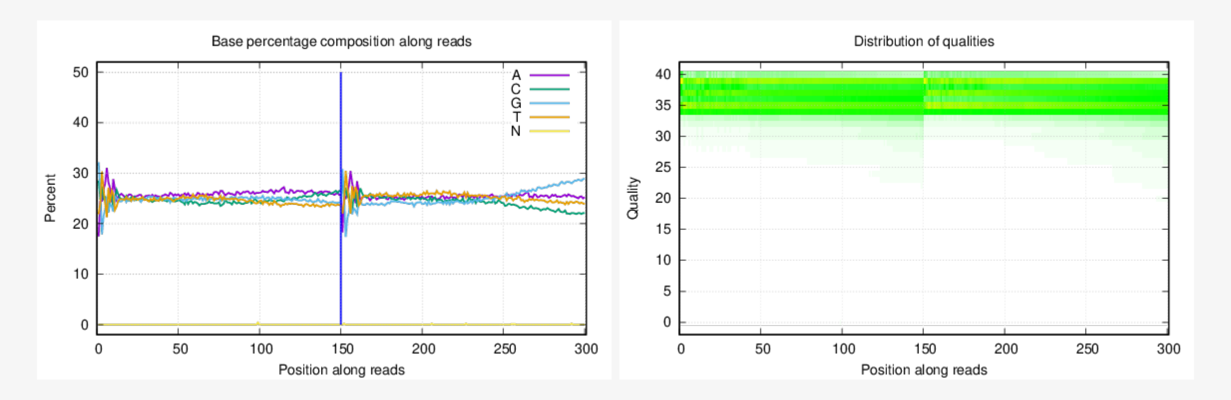
4



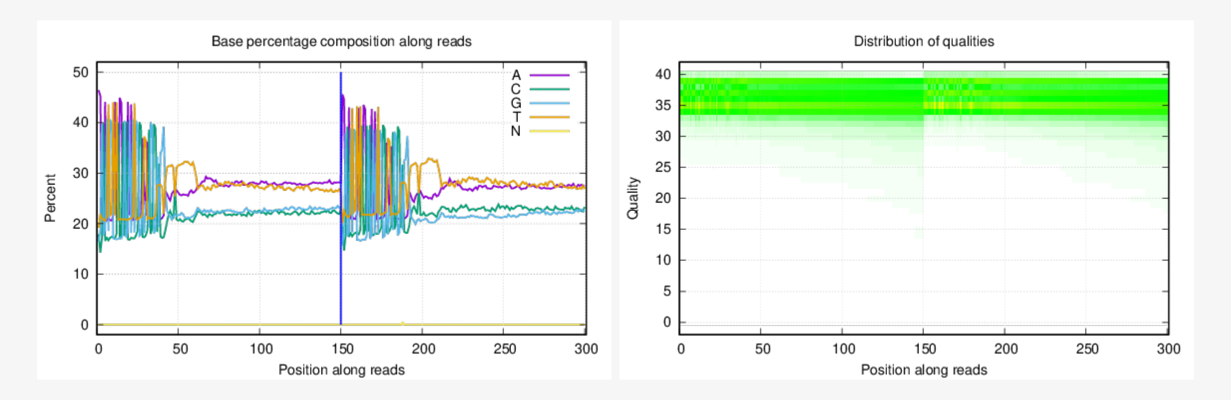
7



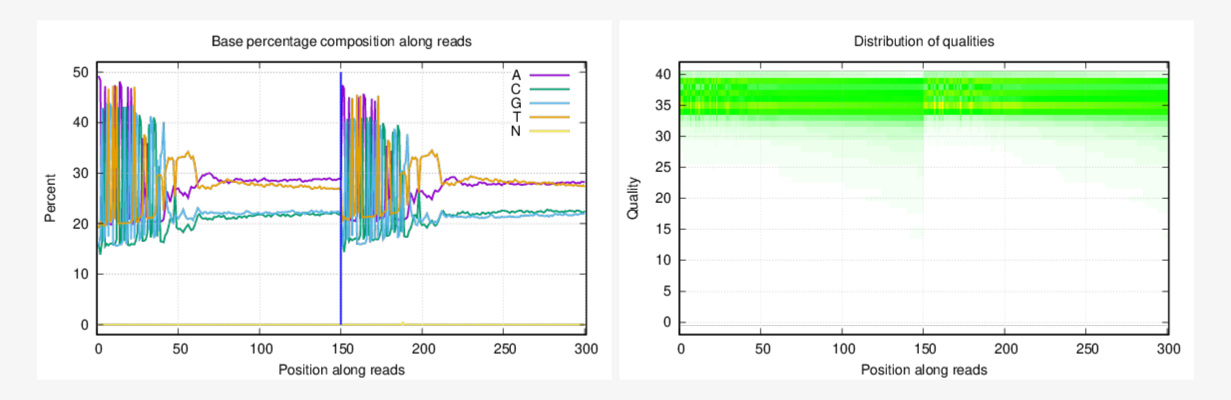
7369_Nod2_1



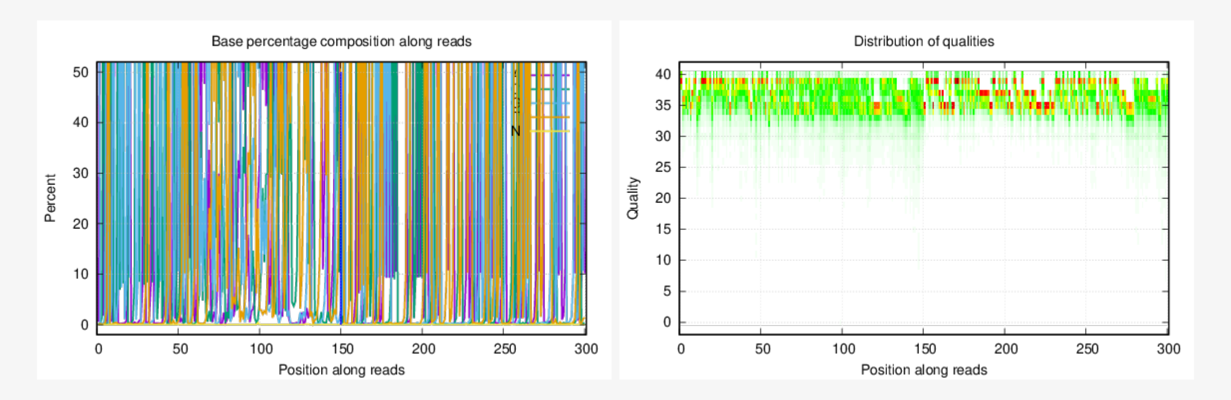
8



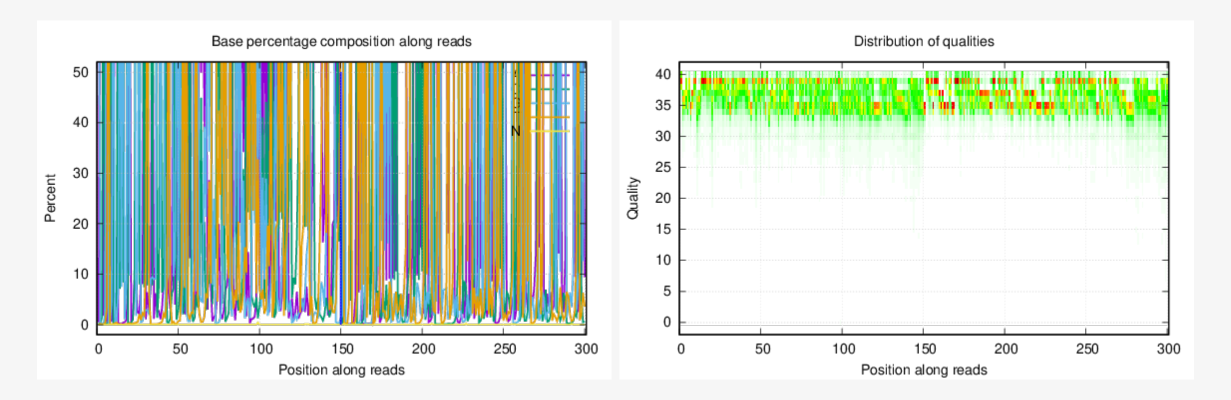
9



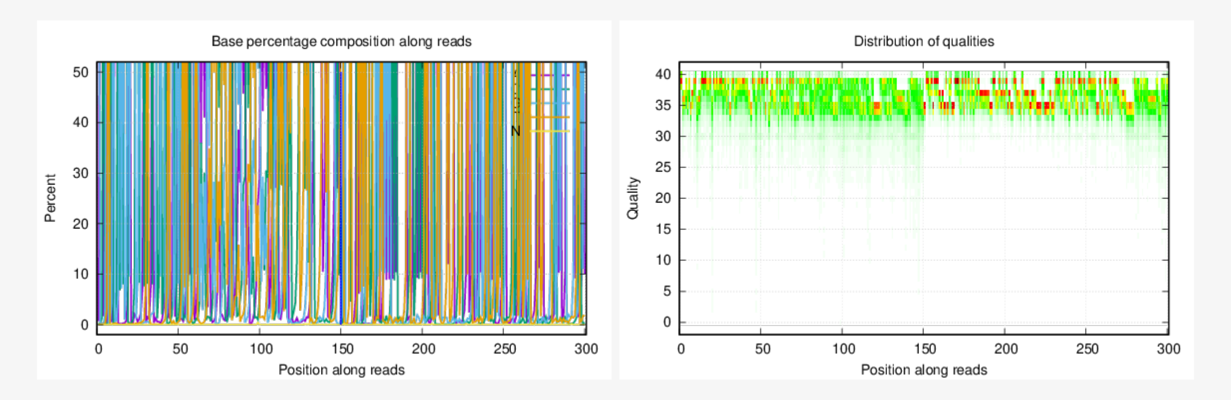
ES_RIGS



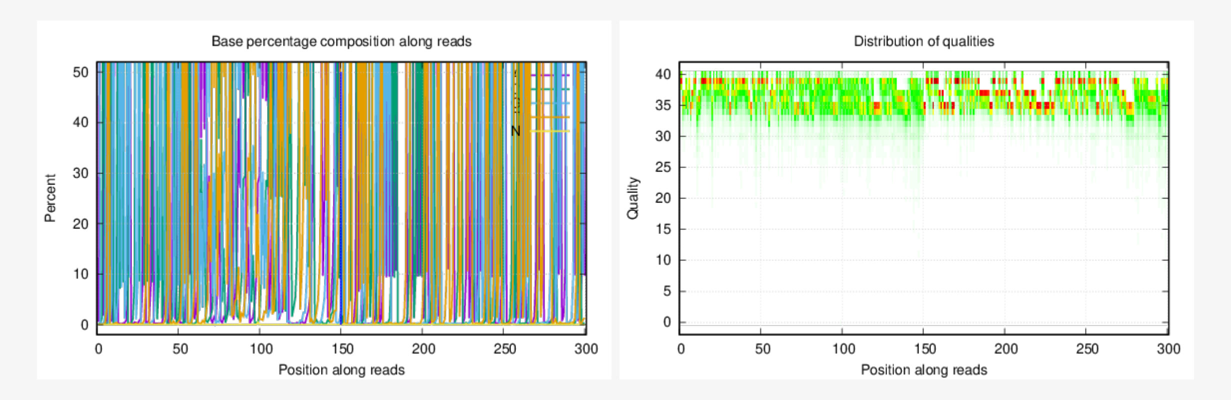
ES_RIGS1



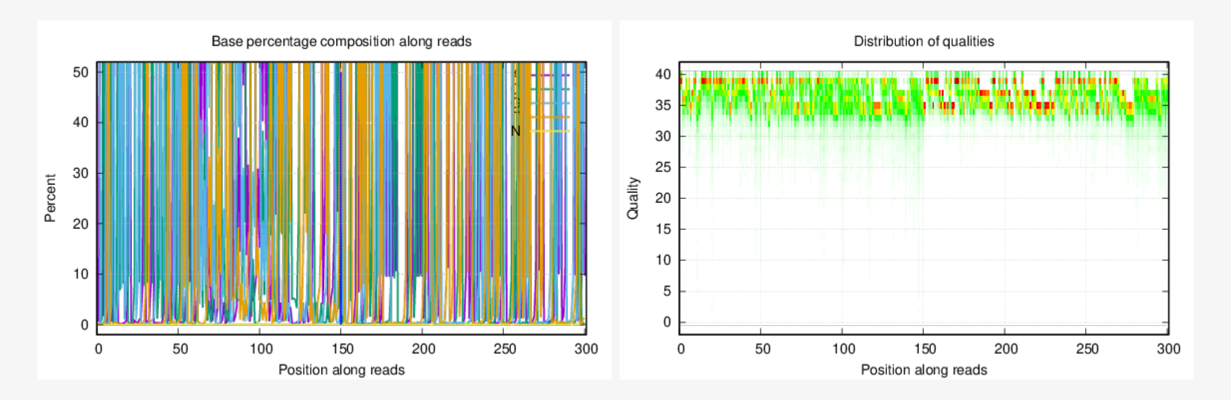
ES_RIGS10



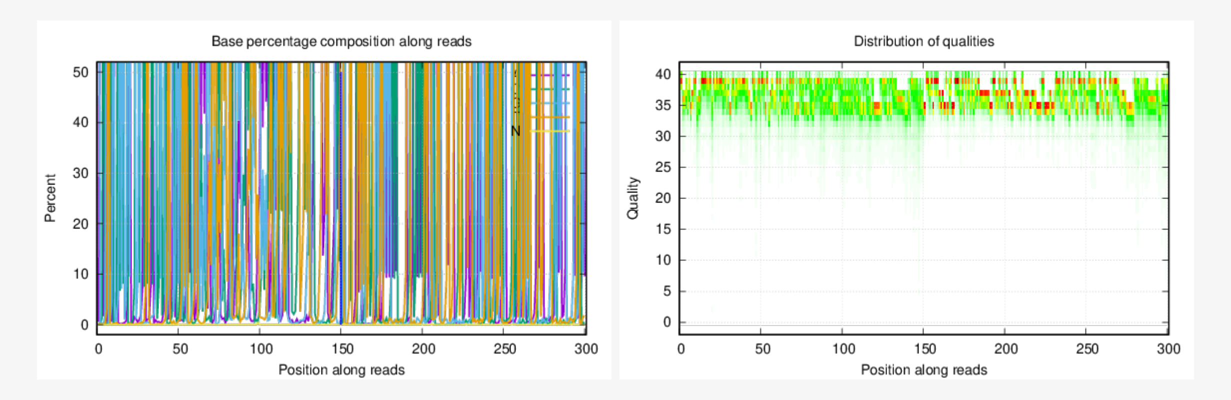
ES_RIGS11



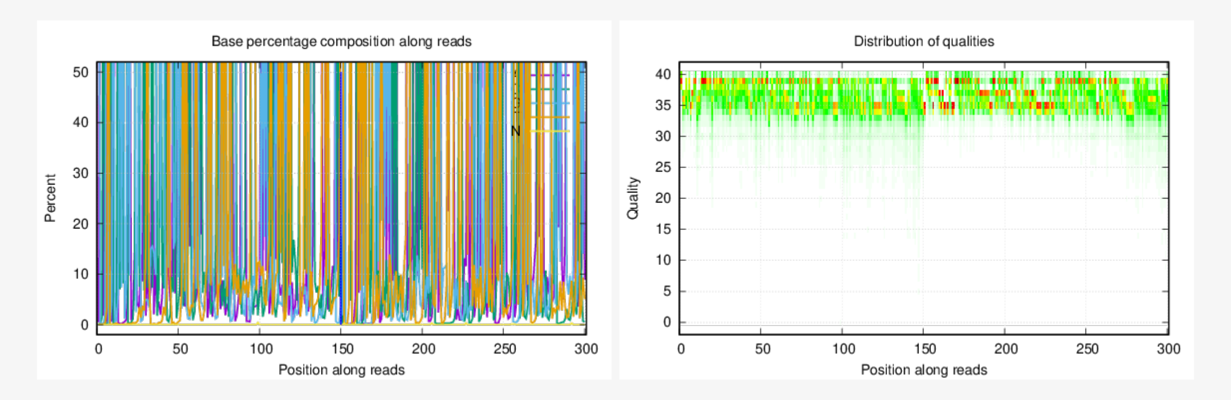
ES_RIGS12



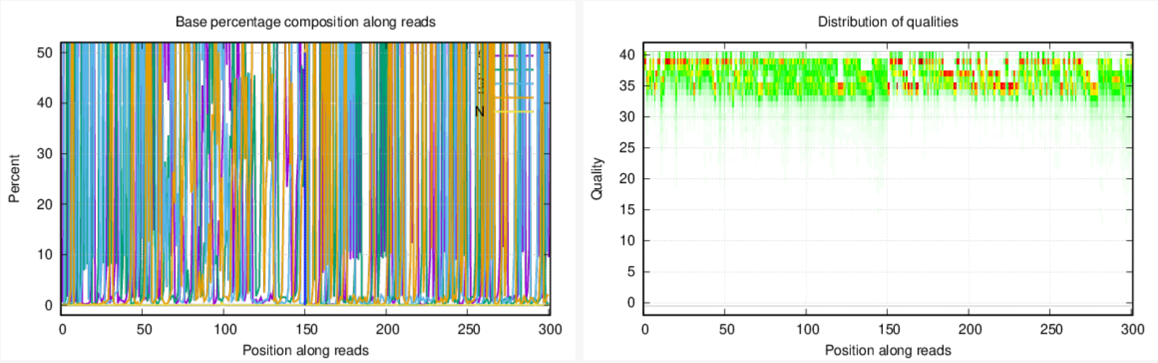
ES_RIGS2



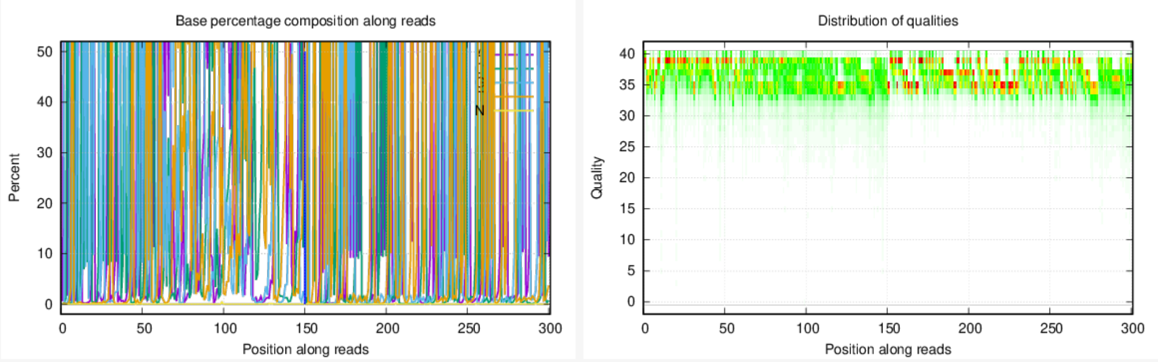
ES_RIGS3



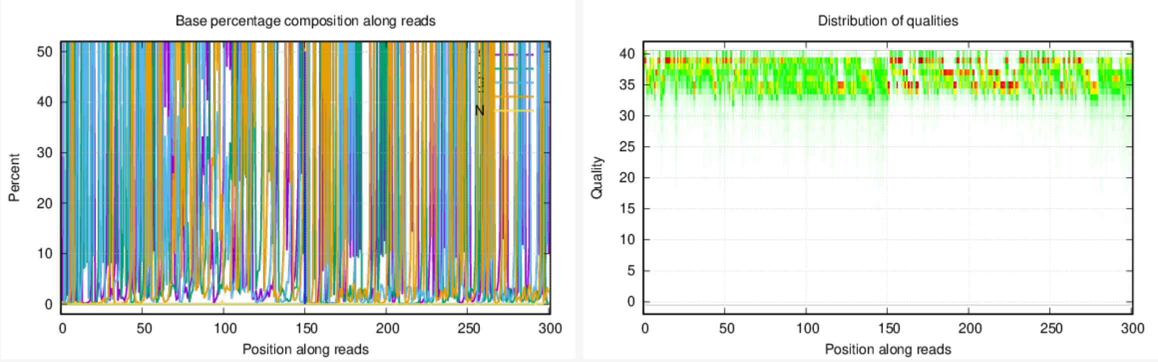
ES_RIGS5



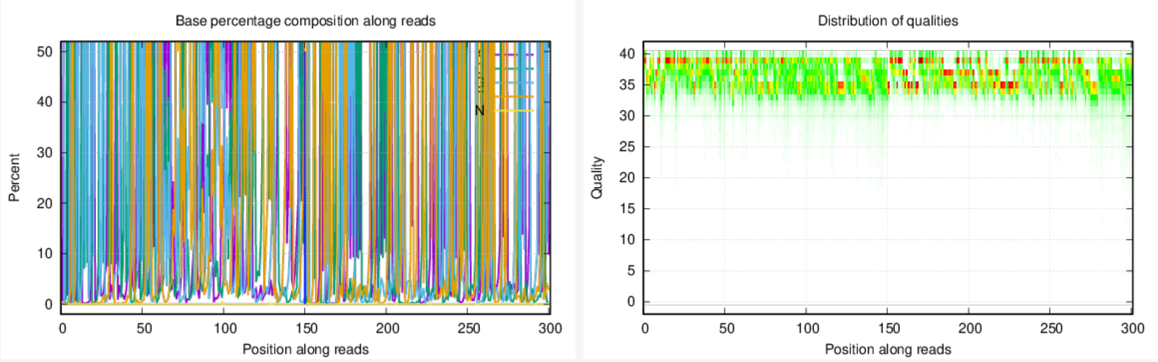
ES_RIGS7



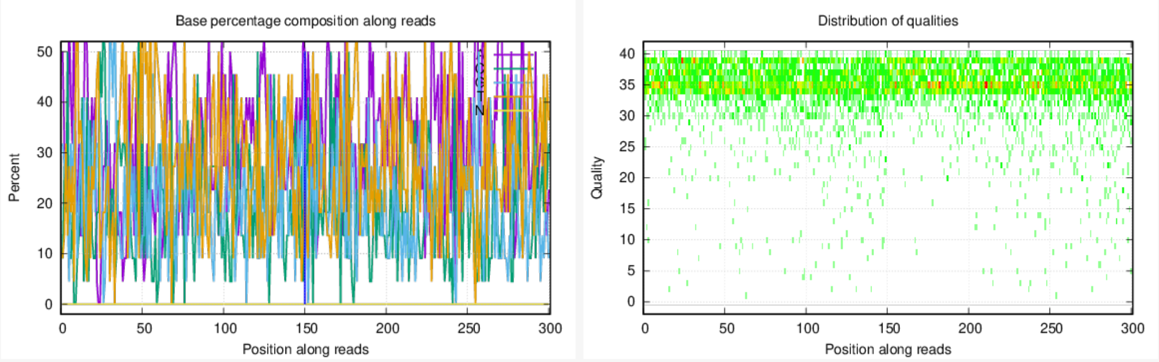
ES_RIGS8



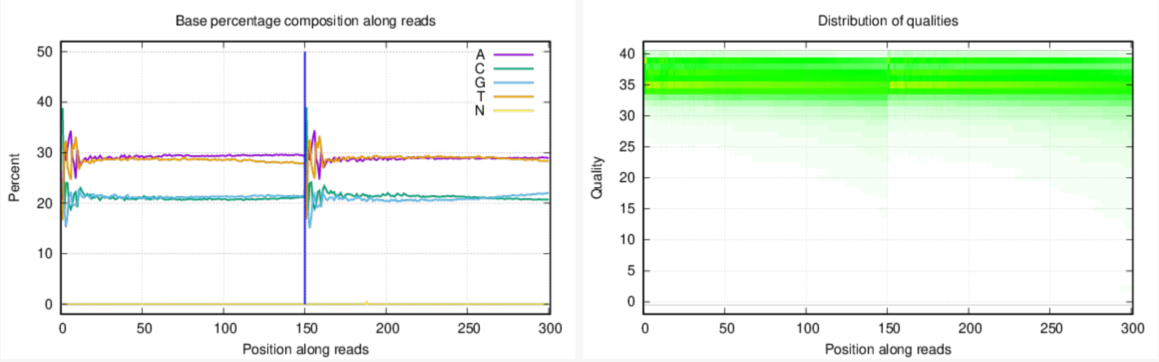
ES_RIGS9



Shur1_e


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

Zap_e2


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