
Analysis Report

Version 0.2

Raw Data Service

TBD210213

Theragen Etex Bioinformatics Team
www.theragenetex.com

Contents

General Information	1
1.1 Sample Information	1
1.2 Analysis Contents	10
Workflow	11
2.1 Library Construction Process	11
2.2 Sequencing Method	12
Data Analysis Results	13
3.1 Quality Controls	13
3.1.1 FastQC Results	13
Appendix	41
4.1 Figure Description	41
4.1.1 FastQC	41
4.2 Explanation of Term Regarding Sequencing Statistics Result	44
4.3 Directory Structure	45

General Information

1.1 Sample Information

Sample ID (TBI ID)	Total reads	Total bases (bp)	Total bases (Gbp)	GC Percent	Q30 MoreBasesRat e
4_AGCGATAG-ACGTCTG (TN2103L0520--AGCGATAG-ACGTCTG)	262,360	65,852,360	0.07	48.30%	91.23%
4_AGCGATAG-AGGATAGG (TN2103L0520--AGCGATAG-AGGATAGG)	117,978	29,612,478	0.03	57.06%	90.50%
4_AGCGATAG-AGGCTATA (TN2103L0520--AGCGATAG-AGGCTATA)	347,668	87,264,668	0.09	51.31%	91.77%
4_AGCGATAG-CTTCGCCT (TN2103L0520--AGCGATAG-CTTCGCCT)	100,174	25,143,674	0.03	70.49%	92.12%
4_AGCGATAG-GCCTCTAT (TN2103L0520--AGCGATAG-GCCTCTAT)	232,144	58,268,144	0.06	55.19%	90.88%
4_AGCGATAG-GTCAGTAC (TN2103L0520--AGCGATAG-GTCAGTAC)	222,054	55,735,554	0.06	53.60%	91.64%
4_AGCGATAG-TAAGATT (TN2103L0520--AGCGATAG-TAAGATT)	44,026	11,050,526	0.01	53.49%	93.05%
4_AGCGATAG-TCAGAGCC (TN2103L0520--AGCGATAG-TCAGAGCC)	281,608	70,683,608	0.07	50.69%	89.50%

4_ATTACTCG-ACGTCTG (TN2103L0520--ATTACTCG-ACGTCTG)	271,036	68,030,036	0.07	50.56%	88.87%
4_ATTACTCG-AGGATAGG (TN2103L0520--ATTACTCG-AGGATAGG)	175,774,968	44,119,516,968	44.12	49.58%	90.89%
4_ATTACTCG-AGGCTATA (TN2103L0520--ATTACTCG-AGGCTATA)	366,814	92,070,314	0.09	50.73%	90.44%
4_ATTACTCG-CTTCGCCT (TN2103L0520--ATTACTCG-CTTCGCCT)	5,989,472	1,503,357,472	1.50	49.63%	89.46%
4_ATTACTCG-GCCTCTAT (TN2103L0520--ATTACTCG-GCCTCTAT)	303,346	76,139,846	0.08	54.06%	89.78%
4_ATTACTCG-GTCAGTAC (TN2103L0520--ATTACTCG-GTCAGTAC)	309,802	77,760,302	0.08	50.36%	89.77%
4_ATTACTCG-TAAGATTA (TN2103L0520--ATTACTCG-TAAGATTA)	360,724	90,541,724	0.09	56.38%	87.84%
4_ATTACTCG-TCAGAGCC (TN2103L0520--ATTACTCG-TCAGAGCC)	297,554,310	74,686,131,810	74.69	49.63%	90.66%
4_ATTCAGAA-ACGTCTG (TN2103L0520--ATTCAGAA-ACGTCTG)	183,310	46,010,810	0.05	47.98%	91.19%
4_ATTCAGAA-AGGATAGG (TN2103L0520--ATTCAGAA-AGGATAGG)	138,542	34,774,042	0.03	55.56%	89.52%
4_ATTCAGAA-AGGCTATA (TN2103L0520--ATTCAGAA-AGGCTATA)	288,846	72,500,346	0.07	46.81%	90.67%
4_ATTCAGAA-CTTCGCCT (TN2103L0520--ATTCAGAA-CTTCGCCT)	86,986	21,833,486	0.02	62.39%	90.49%

4_ATTCAGAA-GCCTCTAT (TN2103L0520--ATTCAGAA-GCCTCTAT)	265,350	66,602,850	0.07	54.10%	90.70%
4_ATTCAGAA-GTCAGTAC (TN2103L0520--ATTCAGAA-GTCAGTAC)	287,950	72,275,450	0.07	47.99%	91.20%
4_ATTCAGAA-TAAGATT (TN2103L0520--ATTCAGAA-TAAGATT)	70,956	17,809,956	0.02	61.91%	92.03%
4_ATTCAGAA-TCAGAGCC (TN2103L0520--ATTCAGAA-TCAGAGCC)	159,516	40,038,516	0.04	57.52%	89.58%
4_CGCTCATT-ACGTCTG (TN2103L0520--CGCTCATT-ACGTCTG)	217,280	54,537,280	0.05	50.03%	91.33%
4_CGCTCATT-AGGATAGG (TN2103L0520--CGCTCATT-AGGATAGG)	4,111,028	1,031,868,028	1.03	49.71%	89.49%
4_CGCTCATT-AGGCTATA (TN2103L0520--CGCTCATT-AGGCTATA)	316,026	79,322,526	0.08	47.24%	91.81%
4_CGCTCATT-CTTCGCCT (TN2103L0520--CGCTCATT-CTTCGCCT)	1,819,796	456,768,796	0.46	50.82%	89.73%
4_CGCTCATT-GCCTCTAT (TN2103L0520--CGCTCATT-GCCTCTAT)	256,344	64,342,344	0.06	52.18%	90.67%
4_CGCTCATT-GTCAGTAC (TN2103L0520--CGCTCATT-GTCAGTAC)	248,922	62,479,422	0.06	48.20%	91.41%
4_CGCTCATT-TAAGATT (TN2103L0520--CGCTCATT-TAAGATT)	512,570	128,655,070	0.13	55.25%	90.61%
4_CGCTCATT-TCAGAGCC (TN2103L0520--CGCTCATT-TCAGAGCC)	4,064,872	1,020,282,872	1.02	50.82%	89.23%

4_CGGCTATG-ACGTCTG (TN2103L0520--CGGCTATG-ACGTCTG)	248,496	62,372,496	0.06	49.59%	90.32%
4_CGGCTATG-AGGATAGG (TN2103L0520--CGGCTATG-AGGATAGG)	106,244	26,667,244	0.03	57.10%	89.61%
4_CGGCTATG-AGGCTATA (TN2103L0520--CGGCTATG-AGGCTATA)	287,628	72,194,628	0.07	50.28%	91.14%
4_CGGCTATG-CTTCGCCT (TN2103L0520--CGGCTATG-CTTCGCCT)	104,656	26,268,656	0.03	66.05%	91.83%
4_CGGCTATG-GCCTCTAT (TN2103L0520--CGGCTATG-GCCTCTAT)	235,050	58,997,550	0.06	51.38%	90.45%
4_CGGCTATG-GTCAGTAC (TN2103L0520--CGGCTATG-GTCAGTAC)	233,630	58,641,130	0.06	52.35%	91.00%
4_CGGCTATG-TAAGATTA (TN2103L0520--CGGCTATG-TAAGATTA)	43,260	10,858,260	0.01	53.88%	92.81%
4_CGGCTATG-TCAGAGCC (TN2103L0520--CGGCTATG-TCAGAGCC)	110,772	27,803,772	0.03	59.32%	88.26%
4_CTGAAGCT-ACGTCTG (TN2103L0520--CTGAAGCT-ACGTCTG)	2,787,986	699,784,486	0.70	54.54%	85.41%
4_CTGAAGCT-AGGATAGG (TN2103L0520--CTGAAGCT-AGGATAGG)	806,840	202,516,840	0.20	50.58%	86.68%
4_CTGAAGCT-AGGCTATA (TN2103L0520--CTGAAGCT-AGGCTATA)	325,984	81,821,984	0.08	47.62%	91.36%
4_CTGAAGCT-CTTCGCCT (TN2103L0520--CTGAAGCT-CTTCGCCT)	140,664	35,306,664	0.04	58.81%	91.68%

4_CTGAAGCT-GCCTCTAT (TN2103L0520--CTGAAGCT-GCCTCTAT)	259,490	65,131,990	0.07	52.55%	91.22%
4_CTGAAGCT-GTCAGTAC (TN2103L0520--CTGAAGCT-GTCAGTAC)	2,620,688	657,792,688	0.66	54.13%	85.93%
4_CTGAAGCT-TAAGATT (TN2103L0520--CTGAAGCT-TAAGATT)	43,656	10,957,656	0.01	53.94%	93.02%
4_CTGAAGCT-TCAGAGCC (TN2103L0520--CTGAAGCT-TCAGAGCC)	263,542	66,149,042	0.07	55.04%	87.47%
4_GAATTCGT-ACGTCCTG (TN2103L0520--GAATTCGT-ACGTCCTG)	197,206	49,498,706	0.05	46.47%	91.46%
4_GAATTCGT-AGGATAGG (TN2103L0520--GAATTCGT-AGGATAGG)	99,508	24,976,508	0.02	55.39%	90.41%
4_GAATTCGT-AGGCTATA (TN2103L0520--GAATTCGT-AGGCTATA)	289,816	72,743,816	0.07	48.26%	91.72%
4_GAATTCGT-CTTCGCCT (TN2103L0520--GAATTCGT-CTTCGCCT)	68,256	17,132,256	0.02	61.11%	90.70%
4_GAATTCGT-GCCTCTAT (TN2103L0520--GAATTCGT-GCCTCTAT)	225,432	56,583,432	0.06	50.39%	91.18%
4_GAATTCGT-GTCAGTAC (TN2103L0520--GAATTCGT-GTCAGTAC)	234,426	58,840,926	0.06	52.34%	91.92%
4_GAATTCGT-TAAGATT (TN2103L0520--GAATTCGT-TAAGATT)	55,168	13,847,168	0.01	62.21%	92.57%
4_GAATTCGT-TCAGAGCC (TN2103L0520--GAATTCGT-TCAGAGCC)	143,952	36,131,952	0.04	56.50%	90.57%

4_GAGATTCC-ACGTCTG (TN2103L0520--GAGATTCC-ACGTCTG)	147,280	36,967,280	0.04	48.25%	91.25%
4_GAGATTCC-AGGATAGG (TN2103L0520--GAGATTCC-AGGATAGG)	2,641,076	662,910,076	0.66	49.97%	89.66%
4_GAGATTCC-AGGCTATA (TN2103L0520--GAGATTCC-AGGCTATA)	253,728	63,685,728	0.06	48.46%	90.83%
4_GAGATTCC-CTTCGCCT (TN2103L0520--GAGATTCC-CTTCGCCT)	1,148,714	288,327,214	0.29	51.93%	89.95%
4_GAGATTCC-GCCTCTAT (TN2103L0520--GAGATTCC-GCCTCTAT)	204,478	51,323,978	0.05	51.70%	90.54%
4_GAGATTCC-GTCAGTAC (TN2103L0520--GAGATTCC-GTCAGTAC)	248,174	62,291,674	0.06	48.15%	91.20%
4_GAGATTCC-TAAGATTA (TN2103L0520--GAGATTCC-TAAGATTA)	1,273,864	319,739,864	0.32	53.51%	90.93%
4_GAGATTCC-TCAGAGCC (TN2103L0520--GAGATTCC-TCAGAGCC)	2,351,316	590,180,316	0.59	52.63%	88.96%
4_TAATGCGC-ACGTCTG (TN2103L0520--TAATGCGC-ACGTCTG)	2,443,514	613,322,014	0.61	54.55%	85.04%
4_TAATGCGC-AGGATAGG (TN2103L0520--TAATGCGC-AGGATAGG)	342,916	86,071,916	0.09	58.44%	83.14%
4_TAATGCGC-AGGCTATA (TN2103L0520--TAATGCGC-AGGCTATA)	278,036	69,787,036	0.07	48.20%	91.28%
4_TAATGCGC-CTTCGCCT (TN2103L0520--TAATGCGC-CTTCGCCT)	144,072	36,162,072	0.04	61.39%	91.91%

4_TAATGCGC-GCCTCTAT (TN2103L0520--TAATGCGC-GCCTCTAT)	195,440	49,055,440	0.05	53.42%	91.04%
4_TAATGCGC-GTCAGTAC (TN2103L0520--TAATGCGC-GTCAGTAC)	2,467,680	619,387,680	0.62	54.33%	85.53%
4_TAATGCGC-TAAGATT (TN2103L0520--TAATGCGC-TAAGATT)	40,560	10,180,560	0.01	53.53%	92.66%
4_TAATGCGC-TCAGAGCC (TN2103L0520--TAATGCGC-TCAGAGCC)	615,608	154,517,608	0.15	52.13%	86.29%
4_TCCGCGAA-ACGTCTG (TN2103L0520--TCCGCGAA-ACGTCTG)	253,766	63,695,266	0.06	49.01%	90.34%
4_TCCGCGAA-AGGATAGG (TN2103L0520--TCCGCGAA-AGGATAGG)	167,558	42,057,058	0.04	56.86%	89.90%
4_TCCGCGAA-AGGCTATA (TN2103L0520--TCCGCGAA-AGGCTATA)	333,200	83,633,200	0.08	50.86%	91.46%
4_TCCGCGAA-CTTCGCCT (TN2103L0520--TCCGCGAA-CTTCGCCT)	118,146	29,654,646	0.03	66.04%	92.02%
4_TCCGCGAA-GCCTCTAT (TN2103L0520--TCCGCGAA-GCCTCTAT)	315,038	79,074,538	0.08	52.03%	90.59%
4_TCCGCGAA-GTCAGTAC (TN2103L0520--TCCGCGAA-GTCAGTAC)	249,372	62,592,372	0.06	52.30%	91.06%
4_TCCGCGAA-TAAGATT (TN2103L0520--TCCGCGAA-TAAGATT)	38,216	9,592,216	0.01	53.77%	92.74%
4_TCCGCGAA-TCAGAGCC (TN2103L0520--TCCGCGAA-TCAGAGCC)	288,908	72,515,908	0.07	51.52%	88.94%

4_TCCGGAGA-ACGTCTG (TN2103L0520--TCCGGAGA-ACGTCTG)	310,526	77,942,026	0.08	51.83%	87.91%
4_TCCGGAGA-AGGATAGG (TN2103L0520--TCCGGAGA-AGGATAGG)	172,497,276	43,296,816,276	43.30	49.71%	90.45%
4_TCCGGAGA-AGGCTATA (TN2103L0520--TCCGGAGA-AGGCTATA)	441,708	110,868,708	0.11	54.45%	89.12%
4_TCCGGAGA-CTTCGCCT (TN2103L0520--TCCGGAGA-CTTCGCCT)	4,178,792	1,048,876,792	1.05	50.71%	88.86%
4_TCCGGAGA-GCCTCTAT (TN2103L0520--TCCGGAGA-GCCTCTAT)	314,648	78,976,648	0.08	53.52%	88.70%
4_TCCGGAGA-GTCAGTAC (TN2103L0520--TCCGGAGA-GTCAGTAC)	348,696	87,522,696	0.09	50.35%	88.28%
4_TCCGGAGA-TAACGATTA (TN2103L0520--TCCGGAGA-TAACGATTA)	475,594	119,374,094	0.12	56.43%	87.46%
4_TCCGGAGA-TCAGAGCC (TN2103L0520--TCCGGAGA-TCAGAGCC)	357,914,374	89,836,507,874	89.84	49.84%	90.17%
4_TCTCGCGC-ACGTCTG (TN2103L0520--TCTCGCGC-ACGTCTG)	266,532	66,899,532	0.07	48.45%	91.01%
4_TCTCGCGC-AGGATAGG (TN2103L0520--TCTCGCGC-AGGATAGG)	168,560	42,308,560	0.04	64.41%	90.75%
4_TCTCGCGC-AGGCTATA (TN2103L0520--TCTCGCGC-AGGCTATA)	348,150	87,385,650	0.09	48.83%	91.94%
4_TCTCGCGC-CTTCGCCT (TN2103L0520--TCTCGCGC-CTTCGCCT)	114,754	28,803,254	0.03	64.22%	92.56%

4_TCTCGCGC-GCCTCTAT (TN2103L0520--TCTCGCGC-GCCTCTAT)	292,792	73,490,792	0.07	51.34%	90.64%
4_TCTCGCGC-GTCAGTAC (TN2103L0520--TCTCGCGC-GTCAGTAC)	274,568	68,916,568	0.07	50.57%	91.29%
4_TCTCGCGC-TAAGATT (TN2103L0520--TCTCGCGC-TAAGATT)	39,048	9,801,048	0.01	53.86%	92.99%
4_TCTCGCGC-TCAGAGCC (TN2103L0520--TCTCGCGC-TCAGAGCC)	255,040	64,015,040	0.06	50.83%	89.37%

- Sample ID (TBI ID): ID provided by customer (Temporary ID assigned by Theragen Etex)
- Total reads: Total number of produced reads
- Total bases: Total number of produced nucleotides
- Total bases (Gbp): The total bases expressed by giga unit
- GC Percent: GC ratio of the total bases
- Q30 MoreBasesRate: Proportion of nucleotides greater than Phred quality score 30 of the total bases

1.2 Analysis Contents

■ Basic sequence analysis list of Theragen Etex

- Quality control for sequencing data

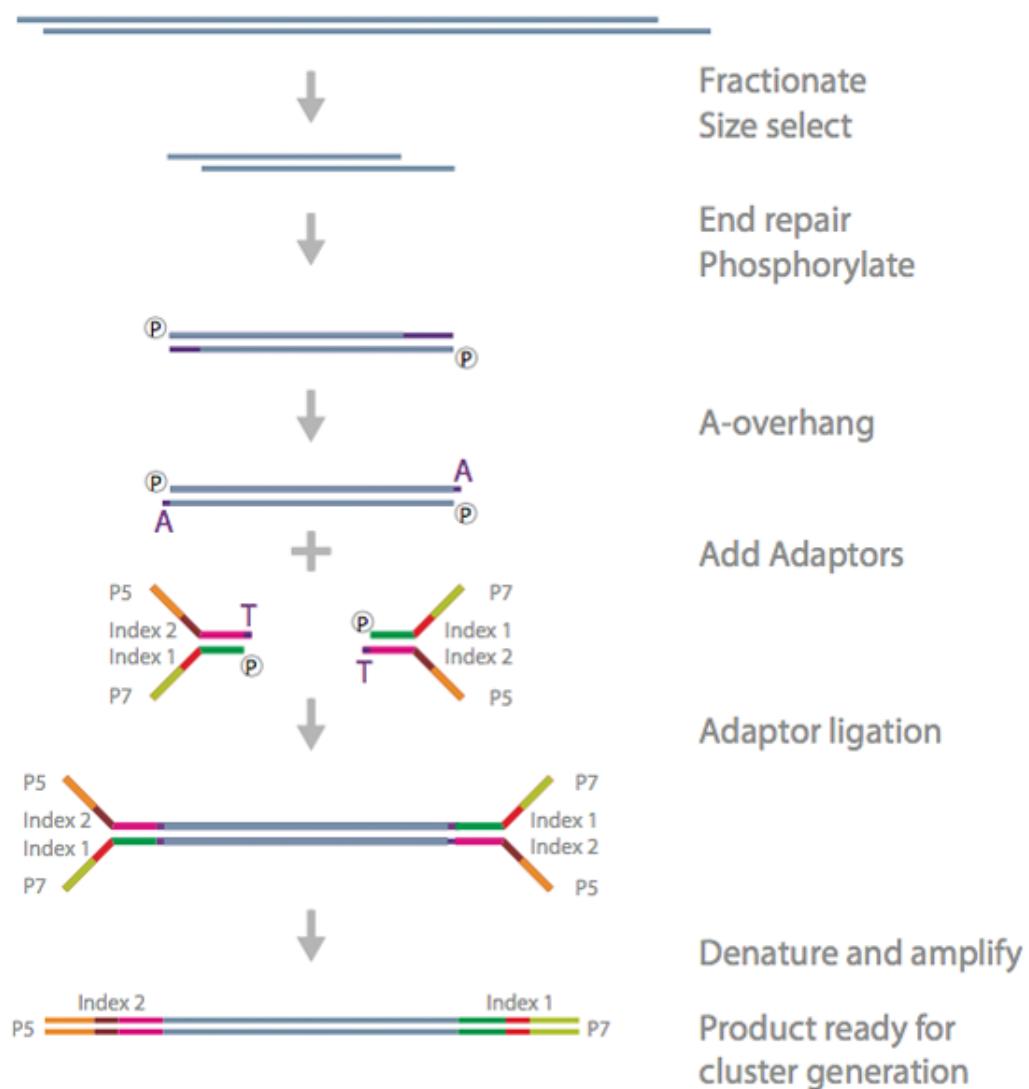
※ Important Information:

- If you need another advanced analysis, additional cost may be requested. Please notify us by E-mail.
- Your analysis data will be stored for 4 months at ftp server. If you want to extend the period, please notify us by E-mail 3 weeks prior to the expiration date.
- Additional fee may be requested for storage time extension.

Workflow

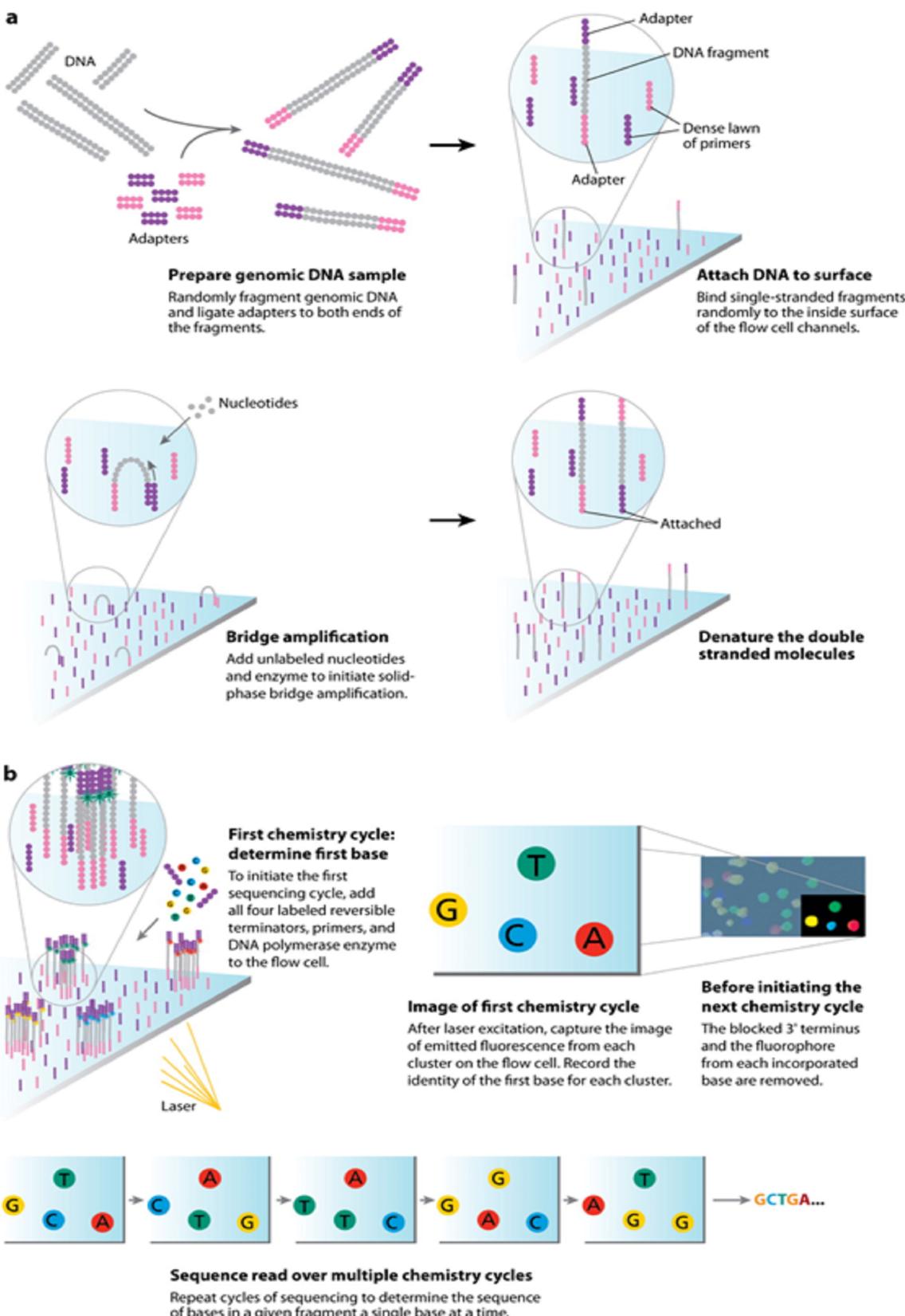
2.1 Library Construction Process

Library construction process



FOR RESEARCH USE ONLY
©2015 Illumina, Inc. All rights reserved.

2.2 Sequencing Method

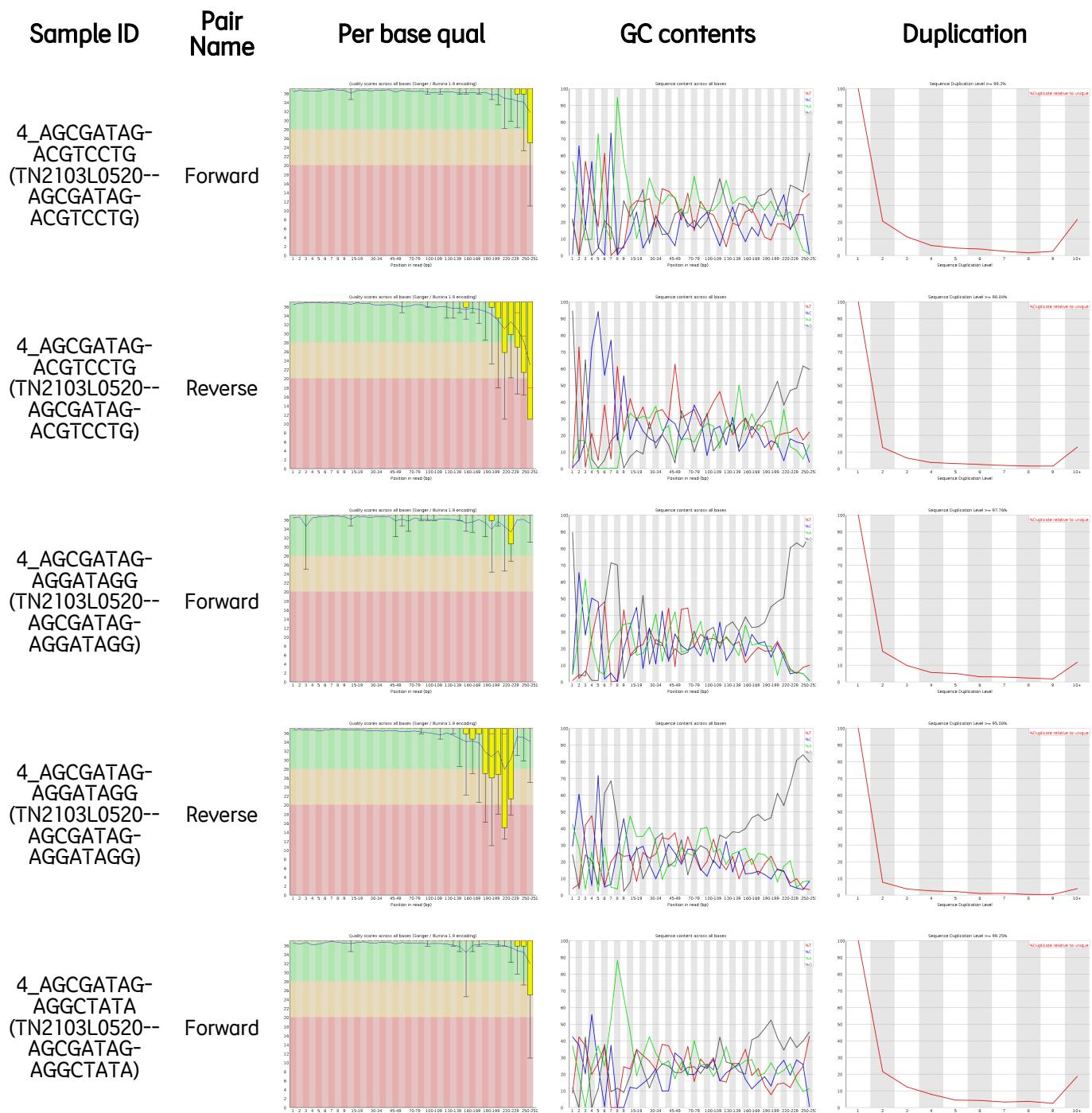


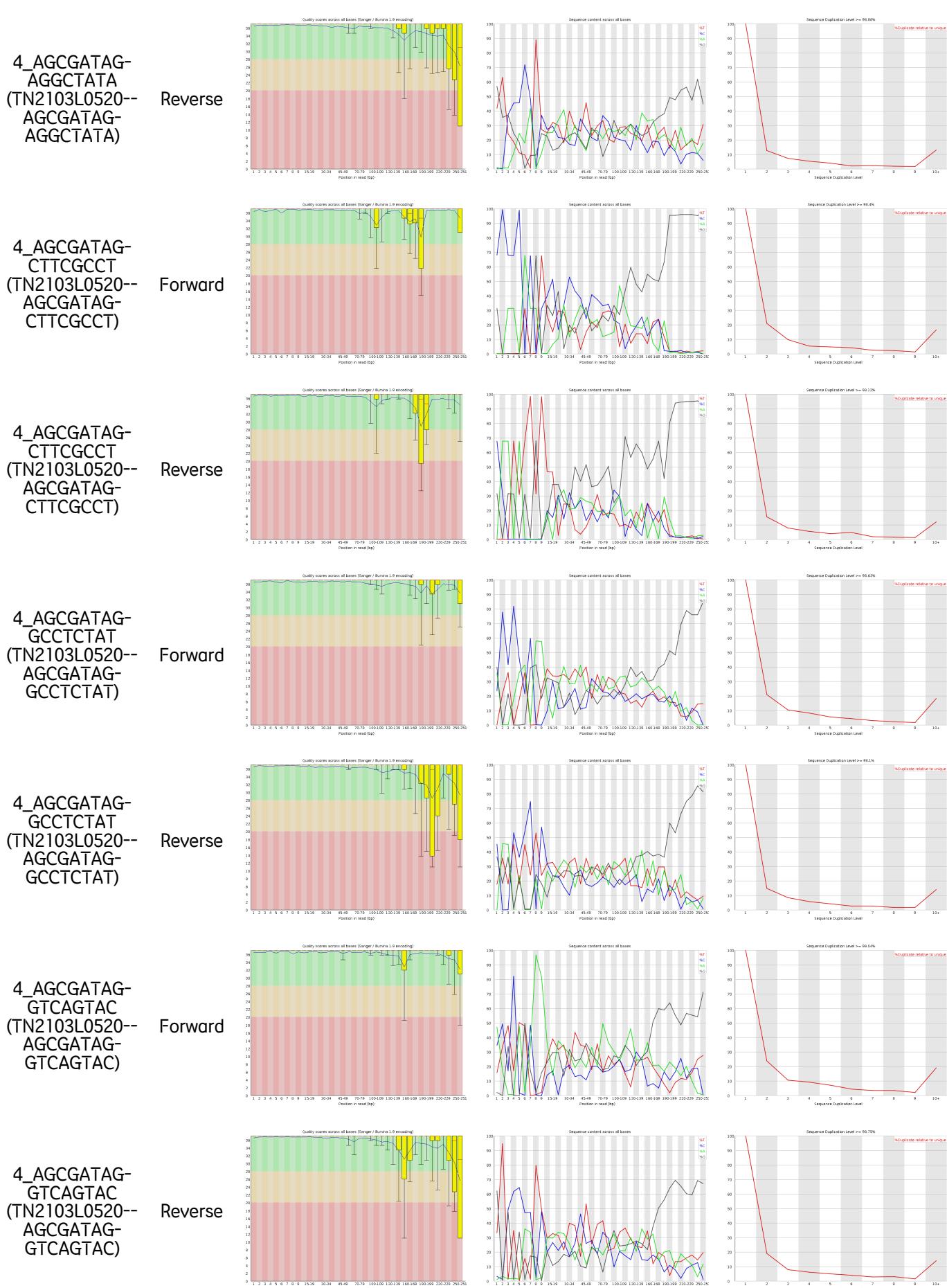
Annual Review of Genomics and Human Genetics Vol. 9: 387-402

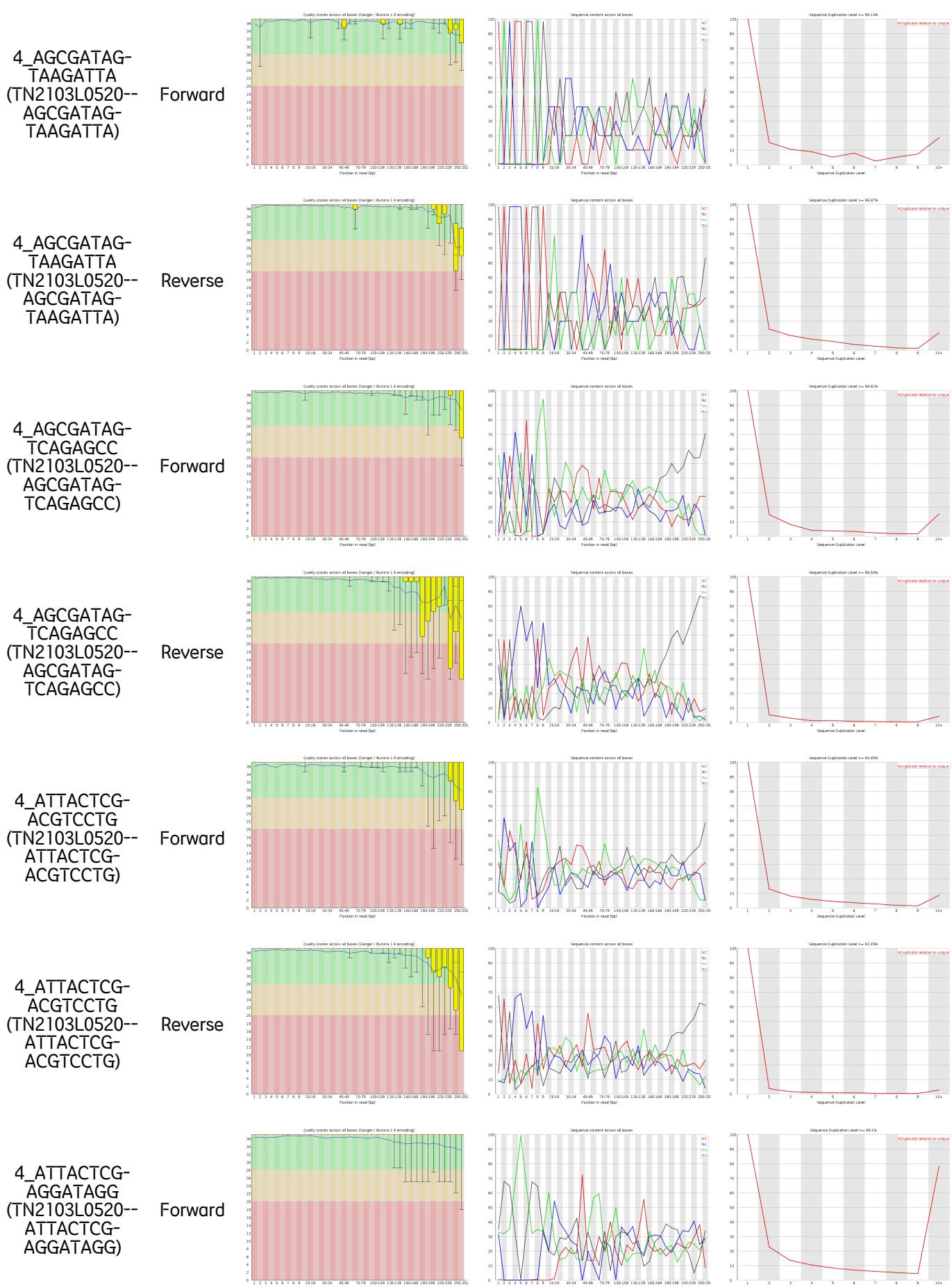
Data Analysis Results

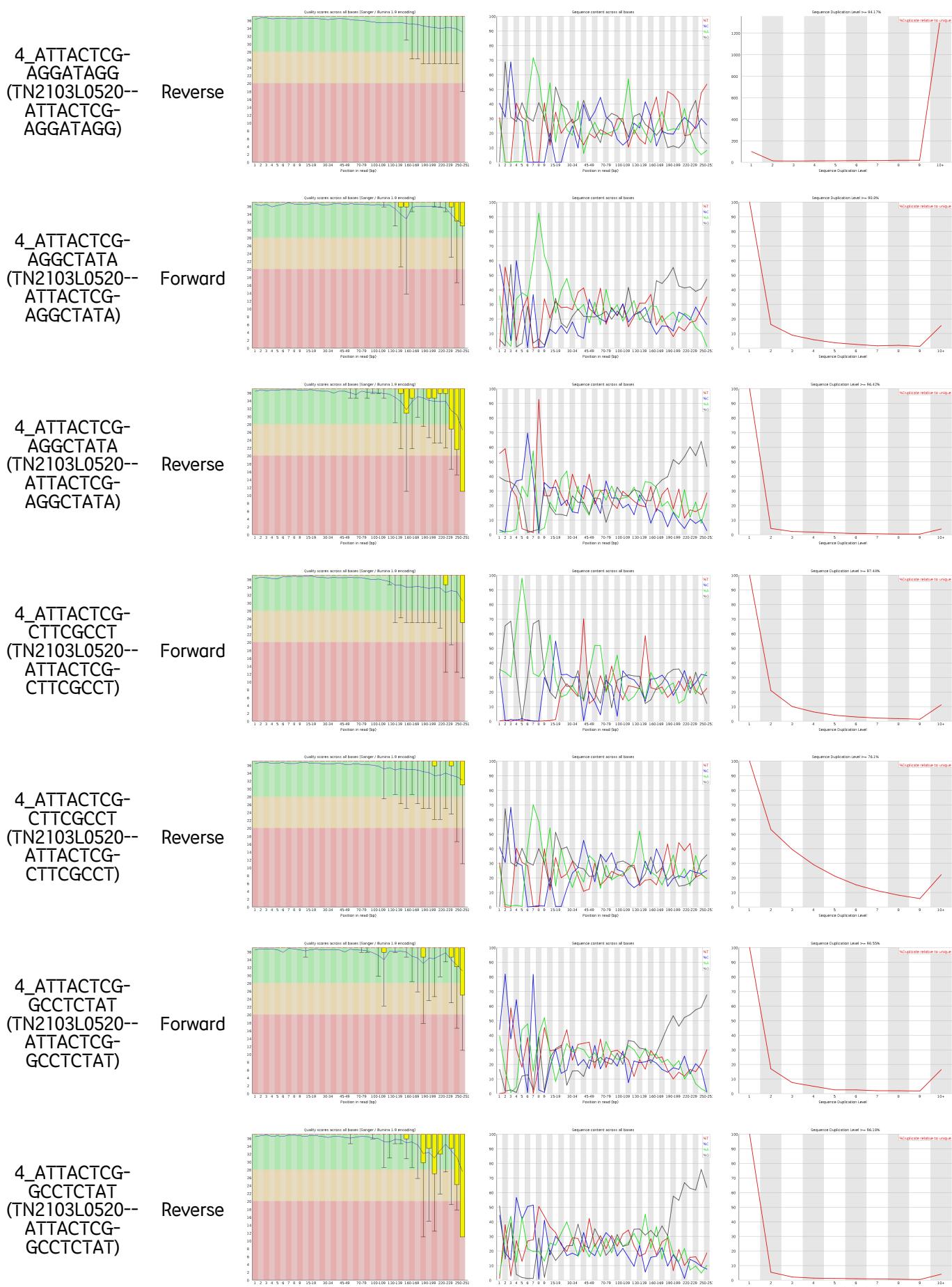
3.1 Quality Controls

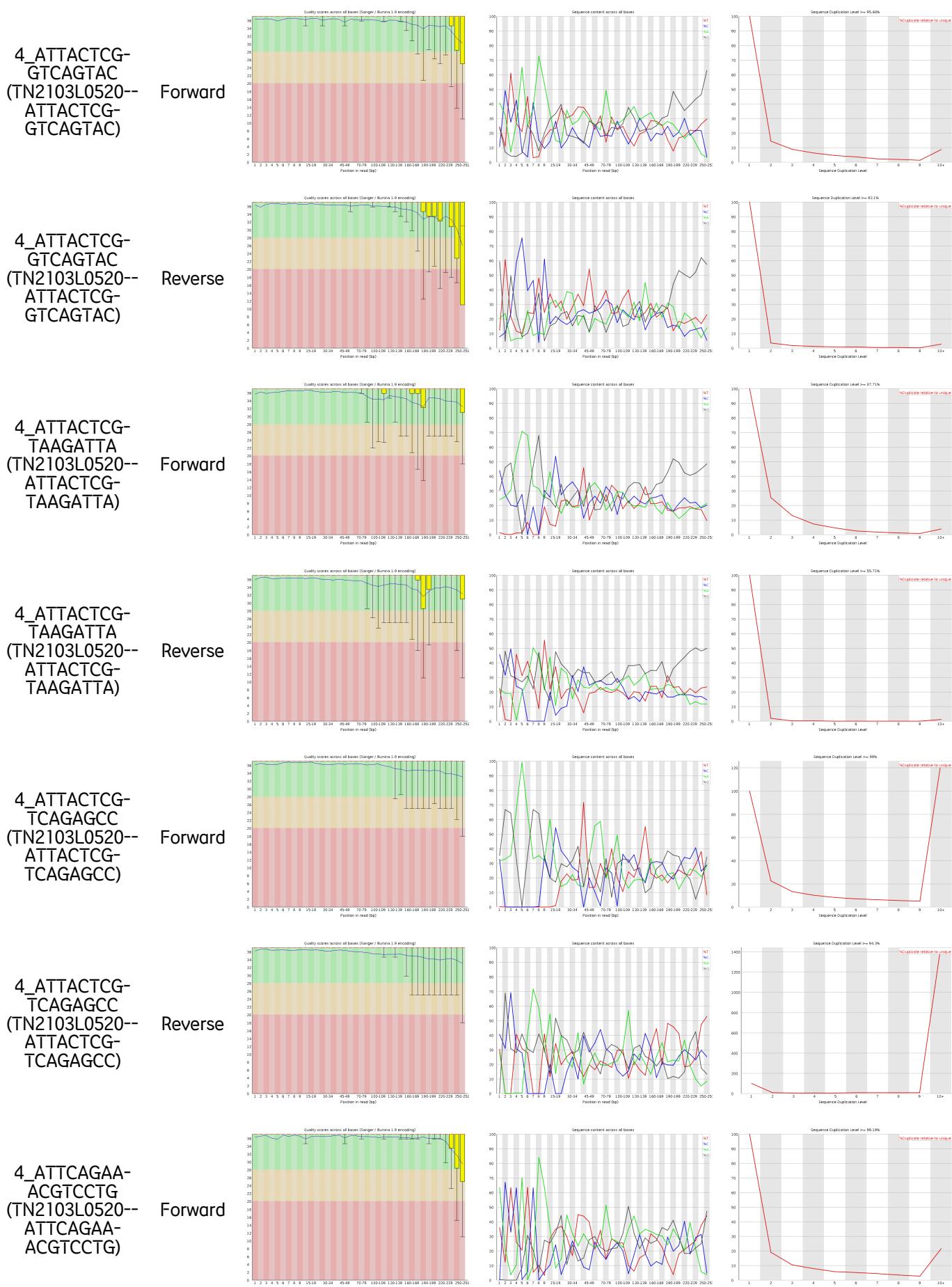
3.1.1 FastQC Results

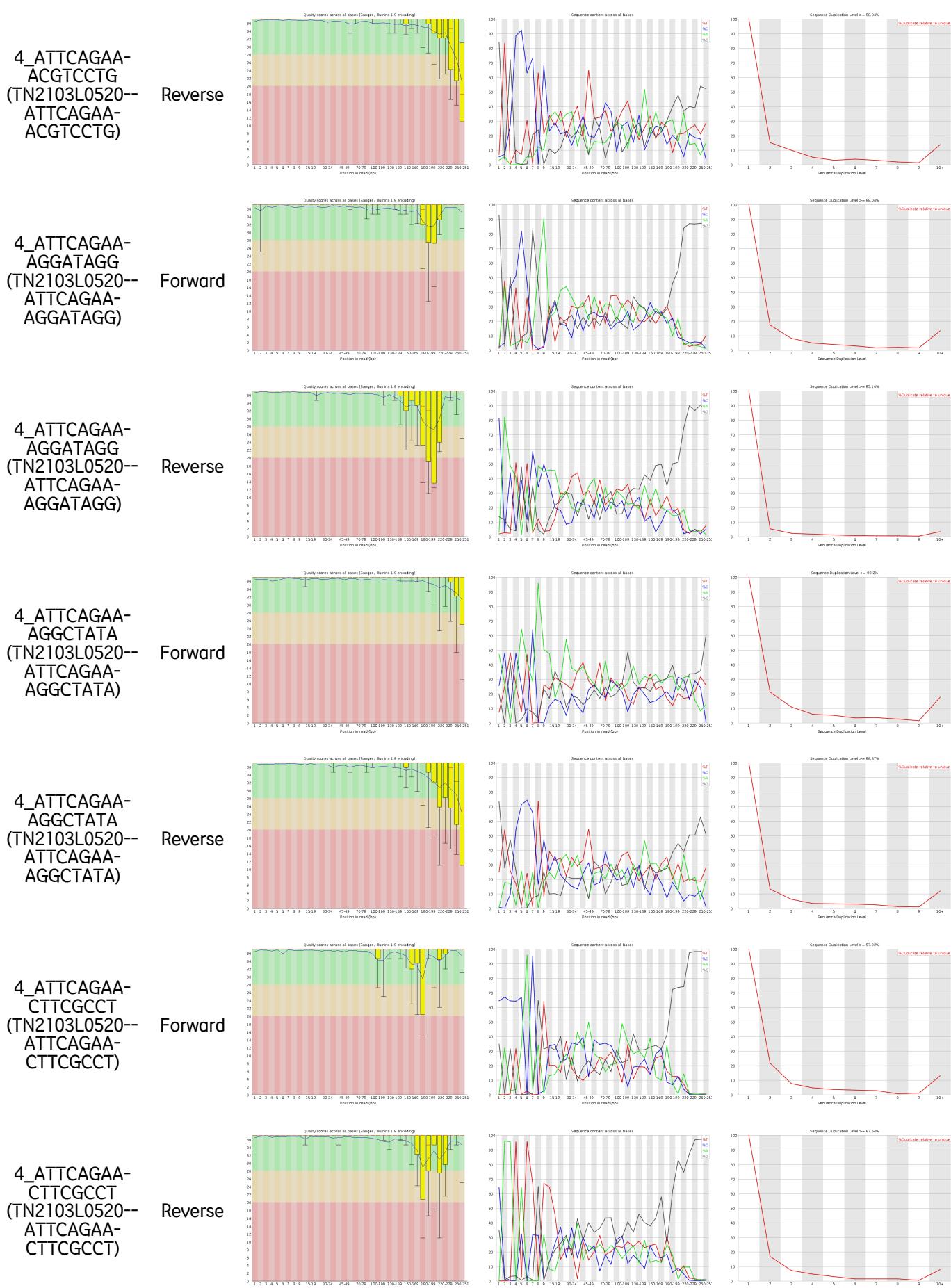


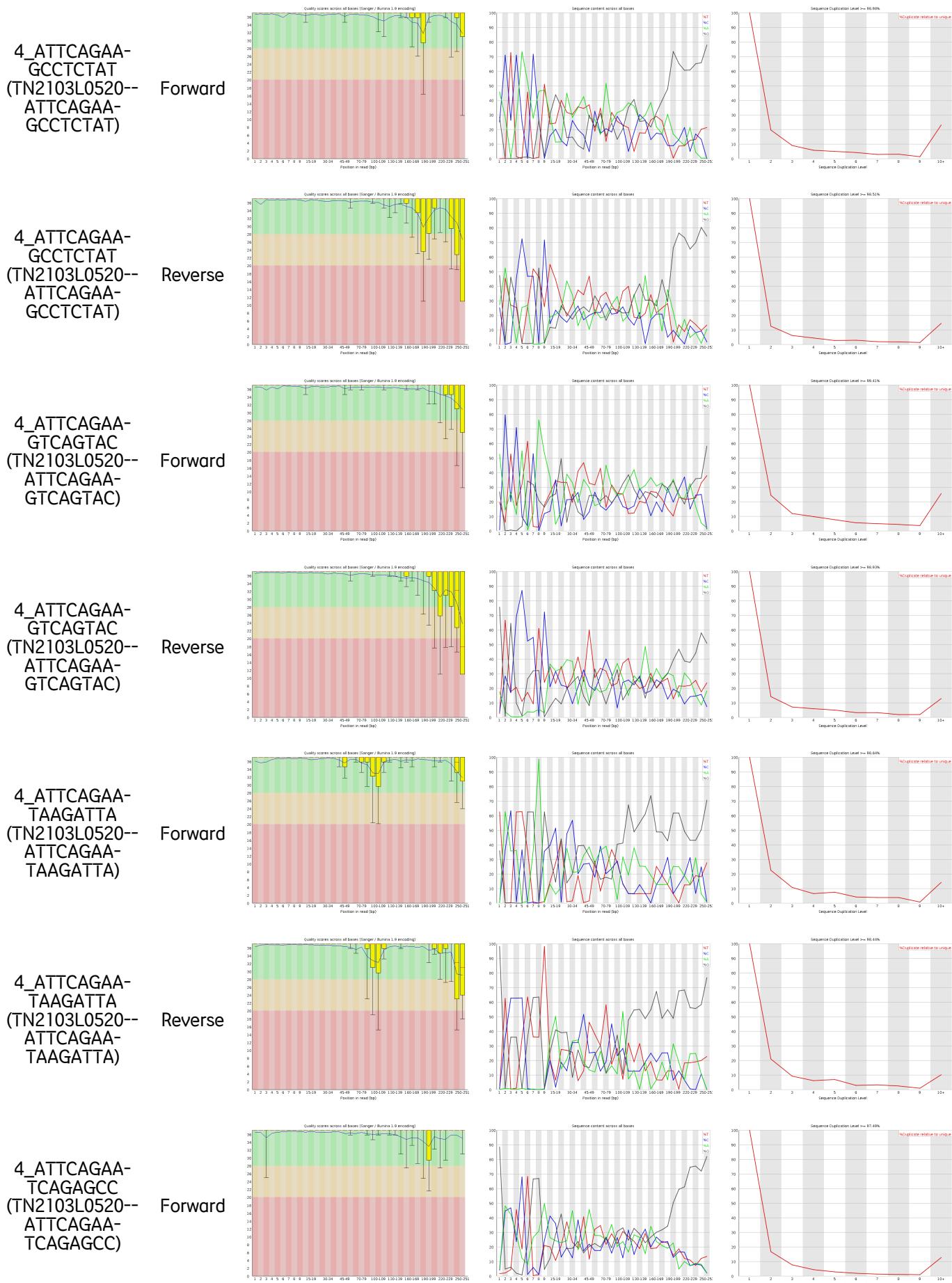


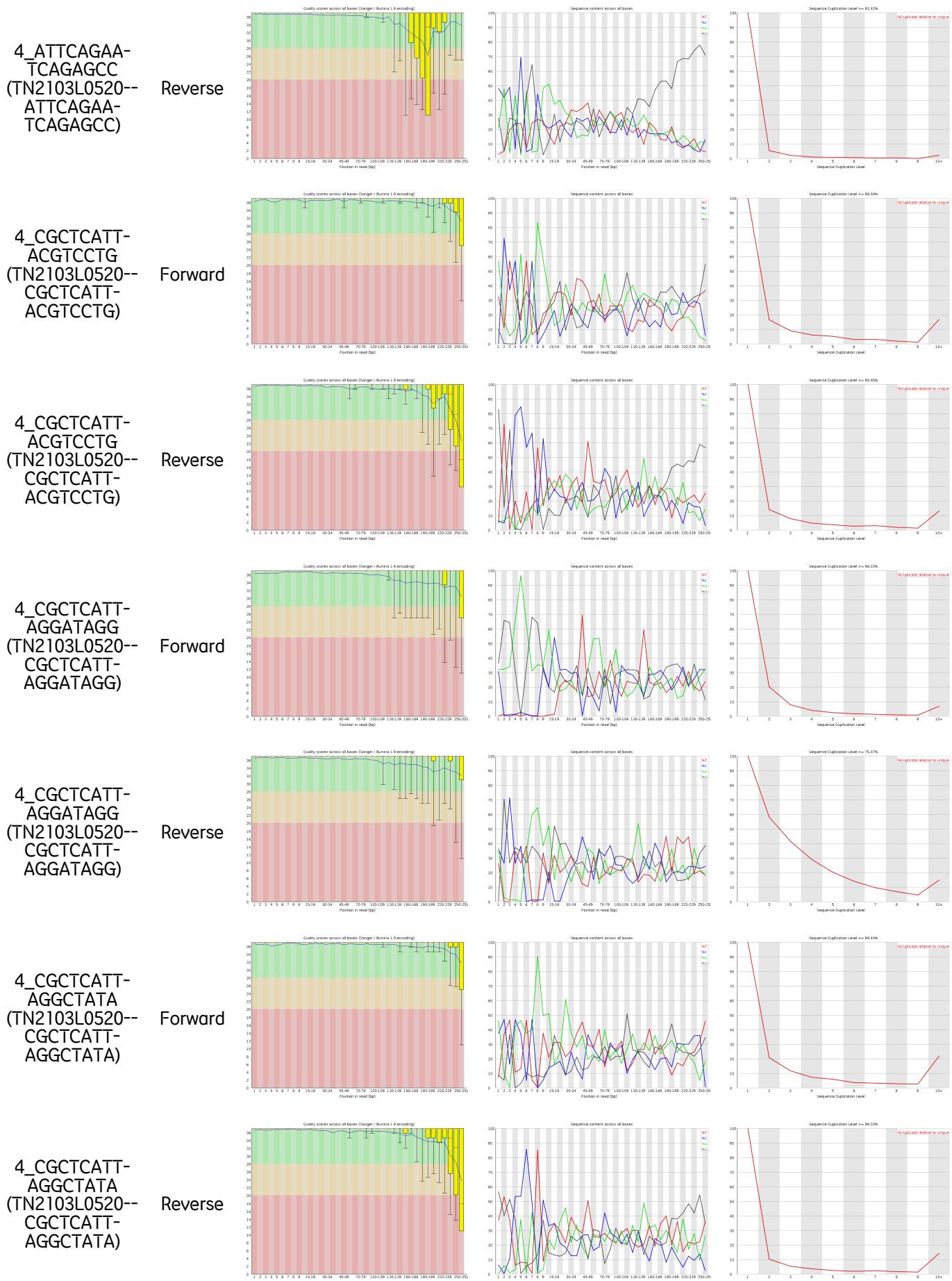


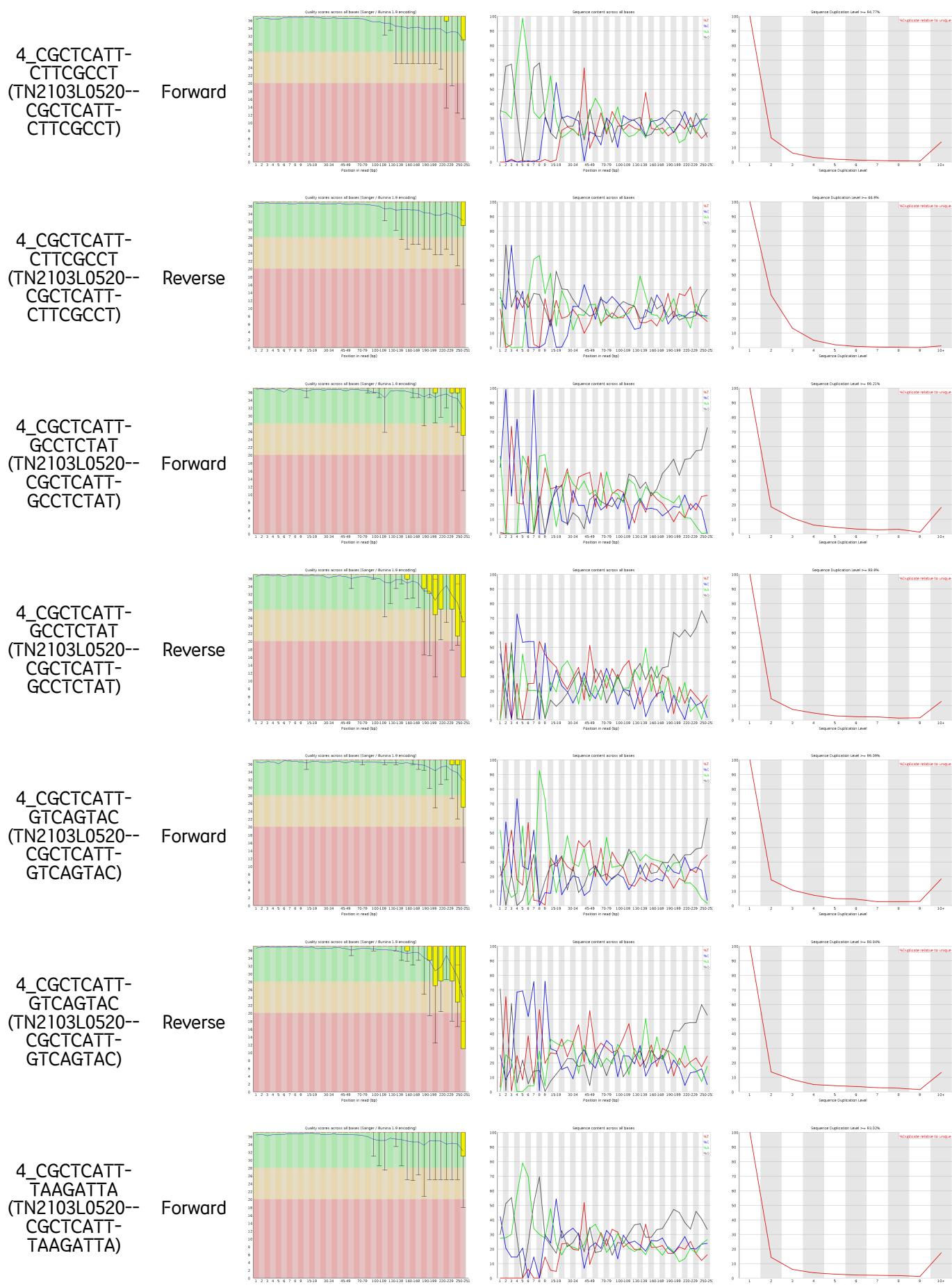


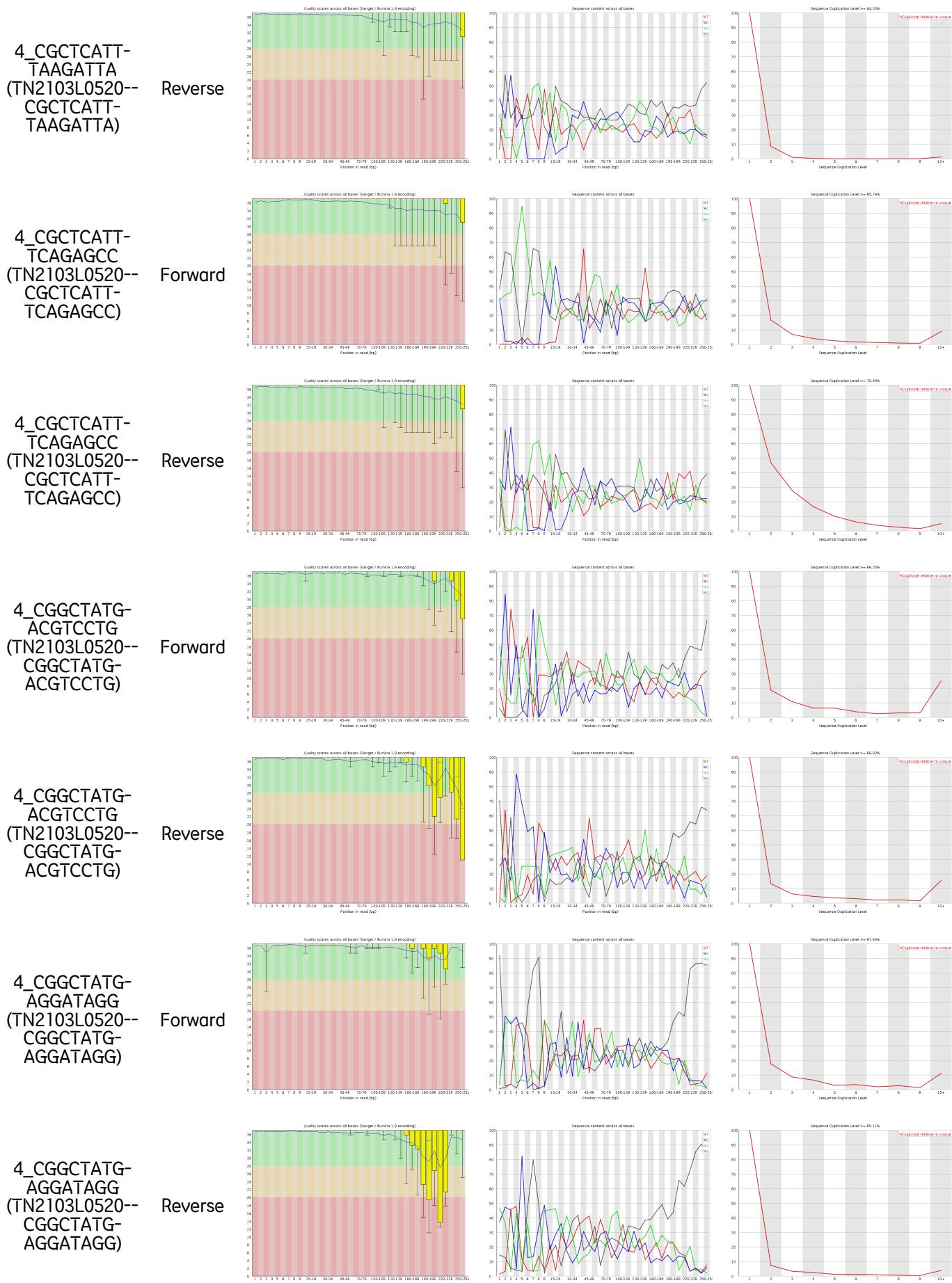


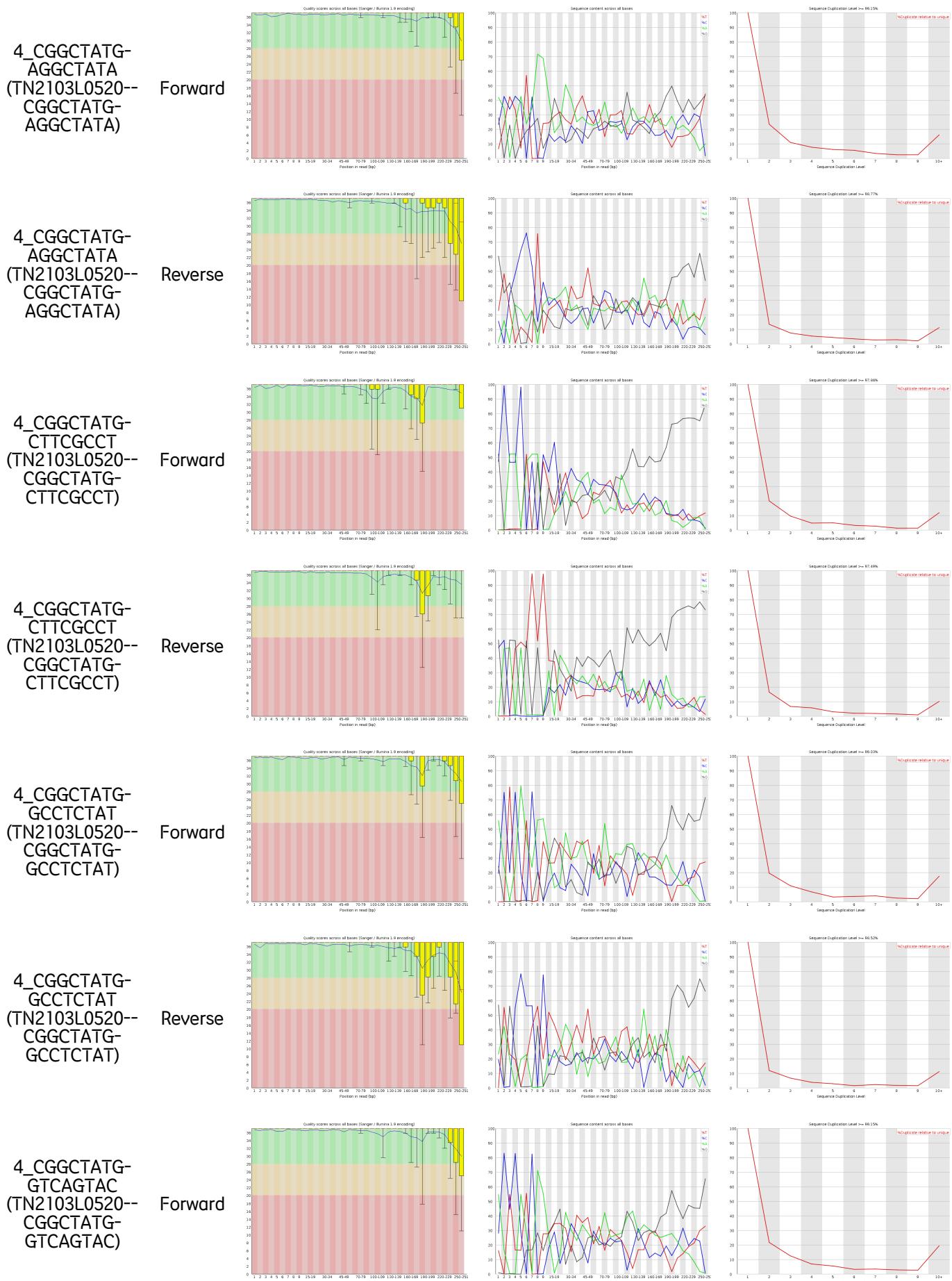


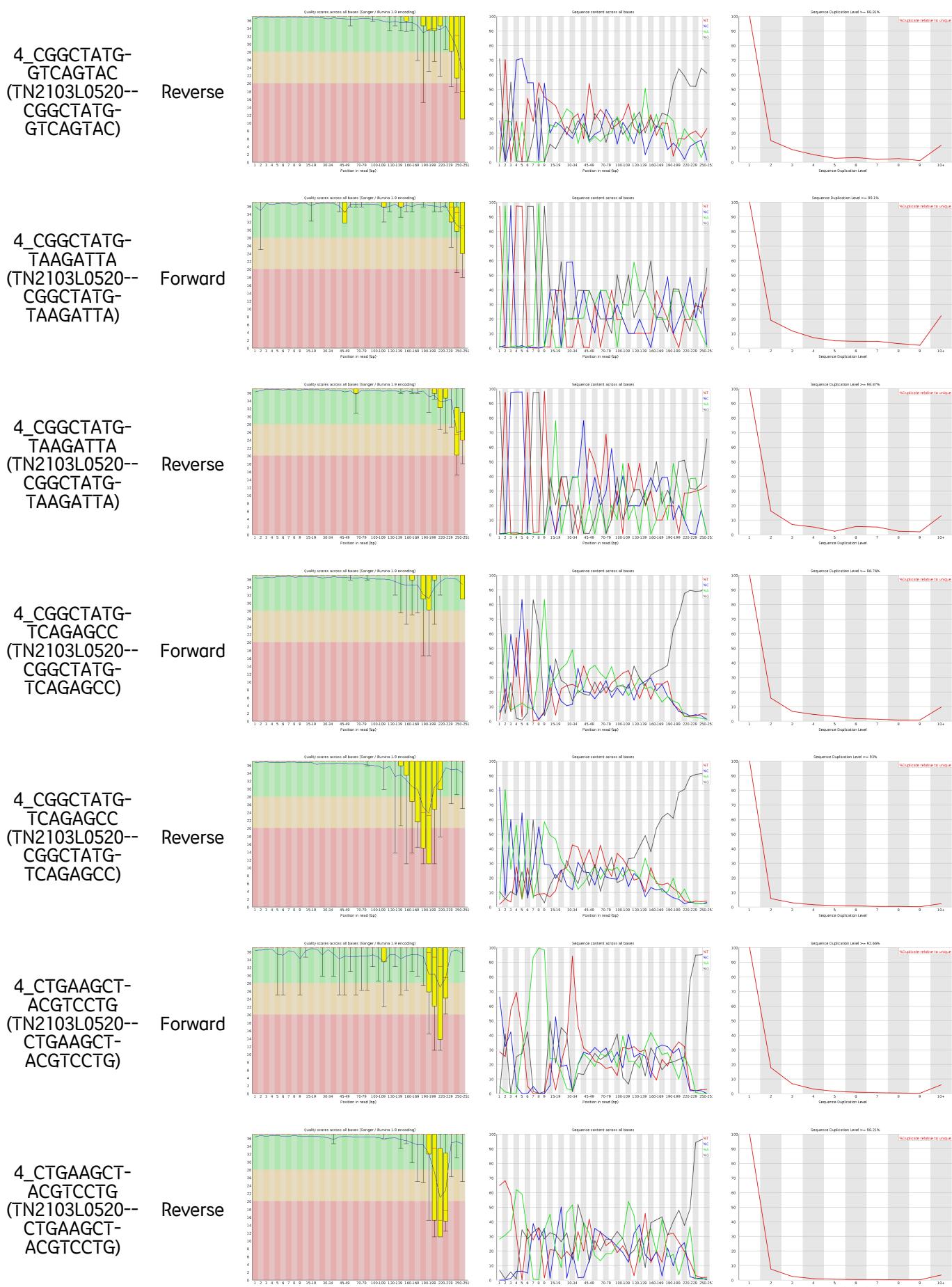


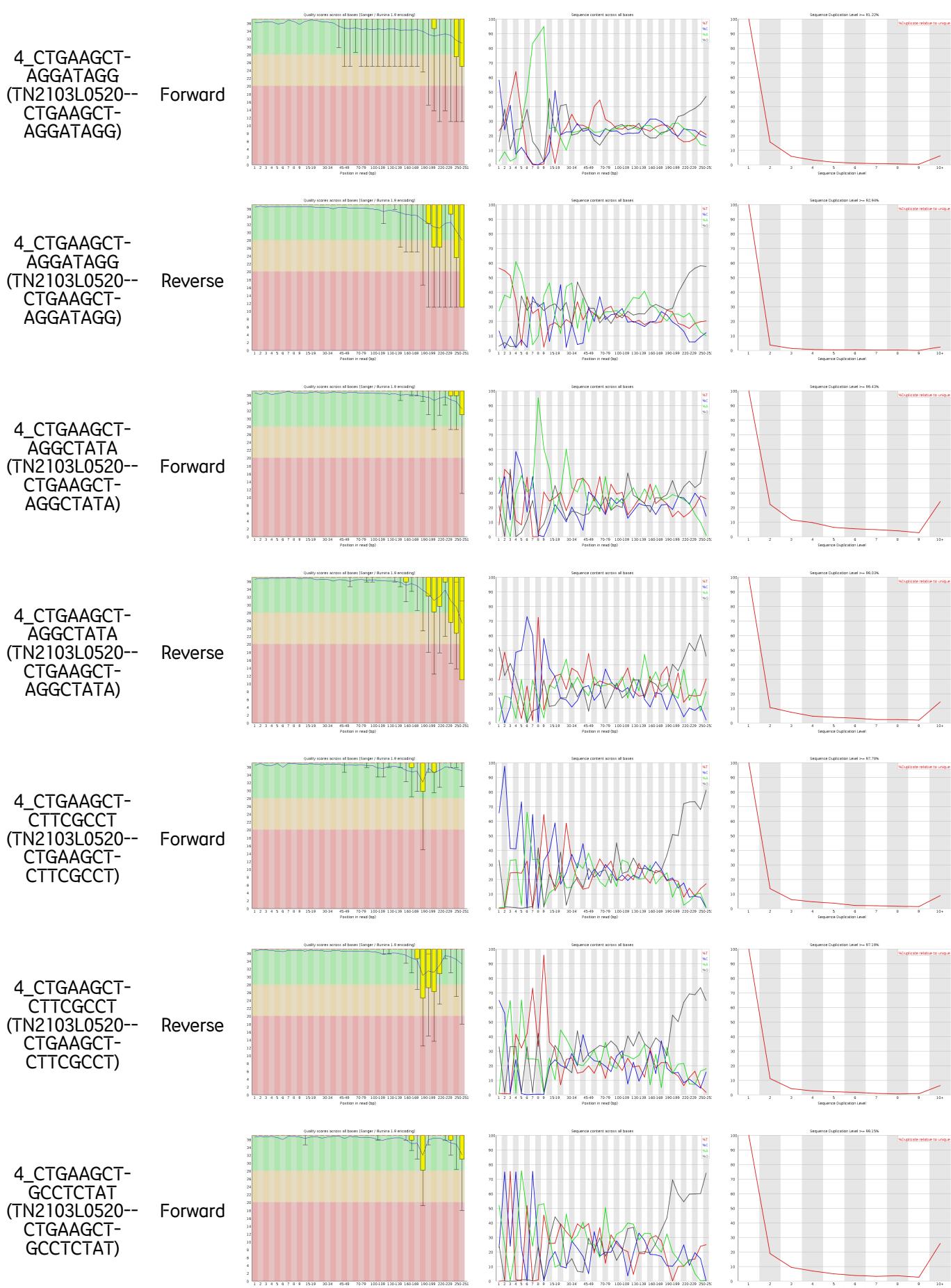


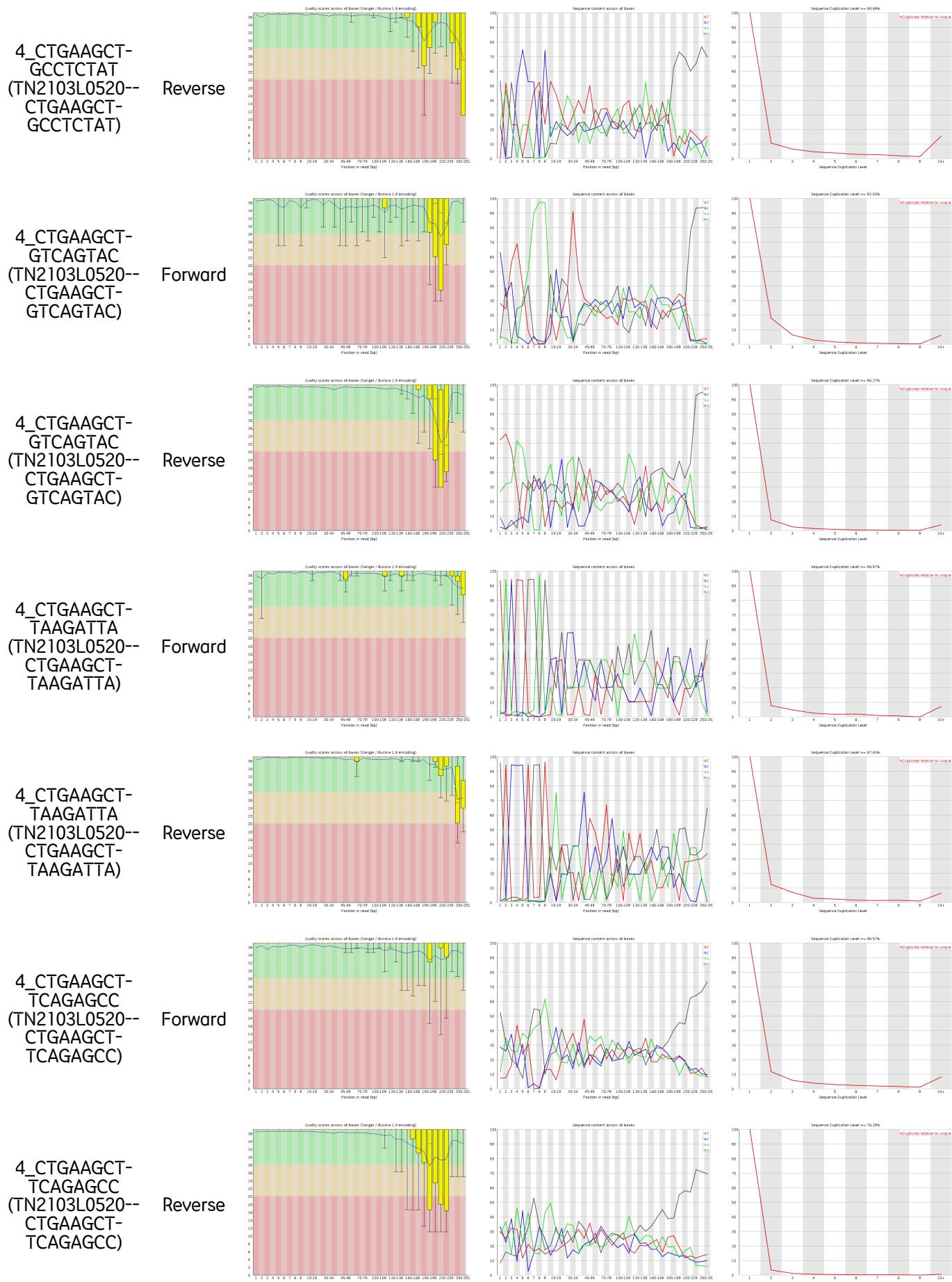


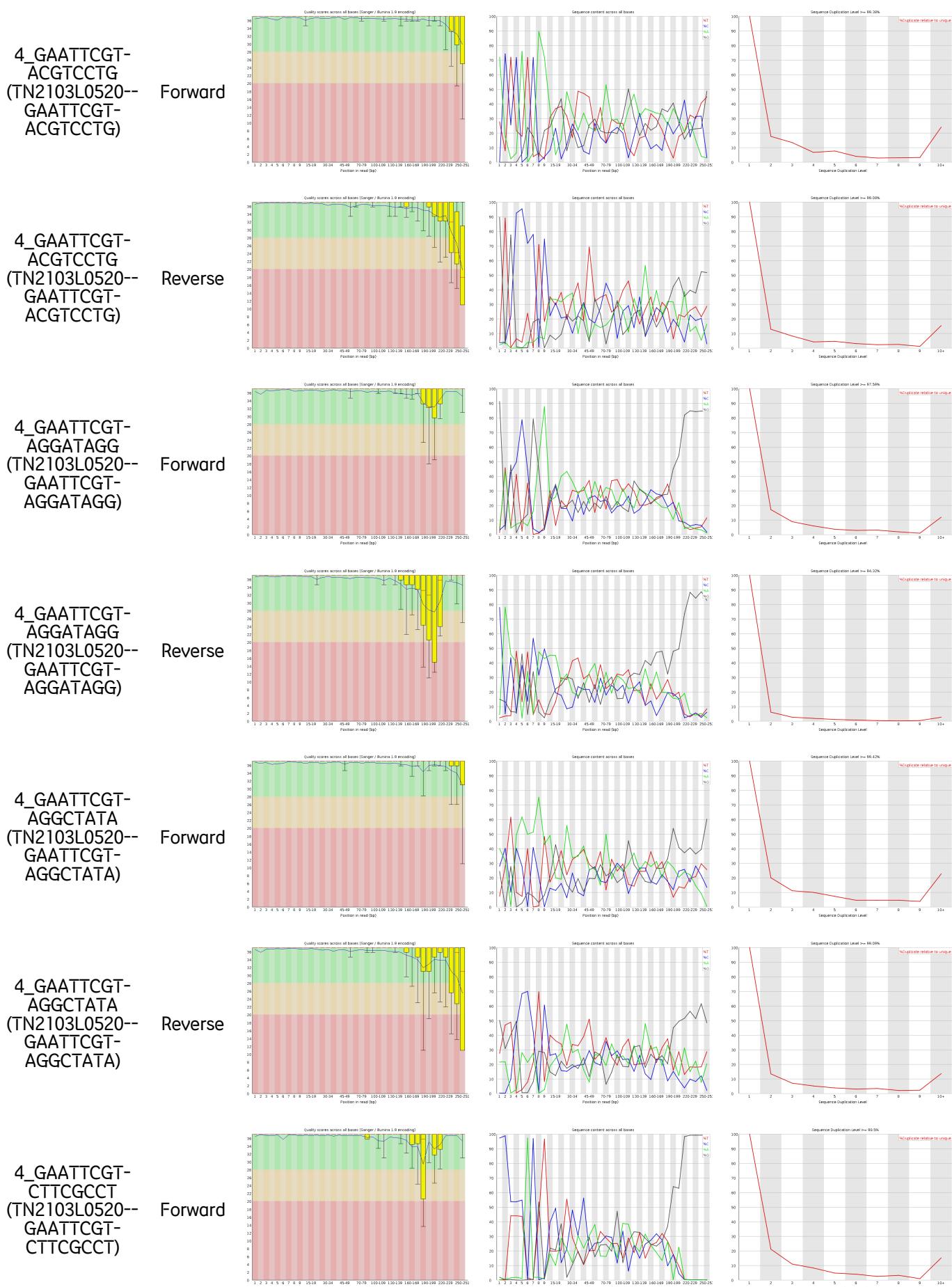


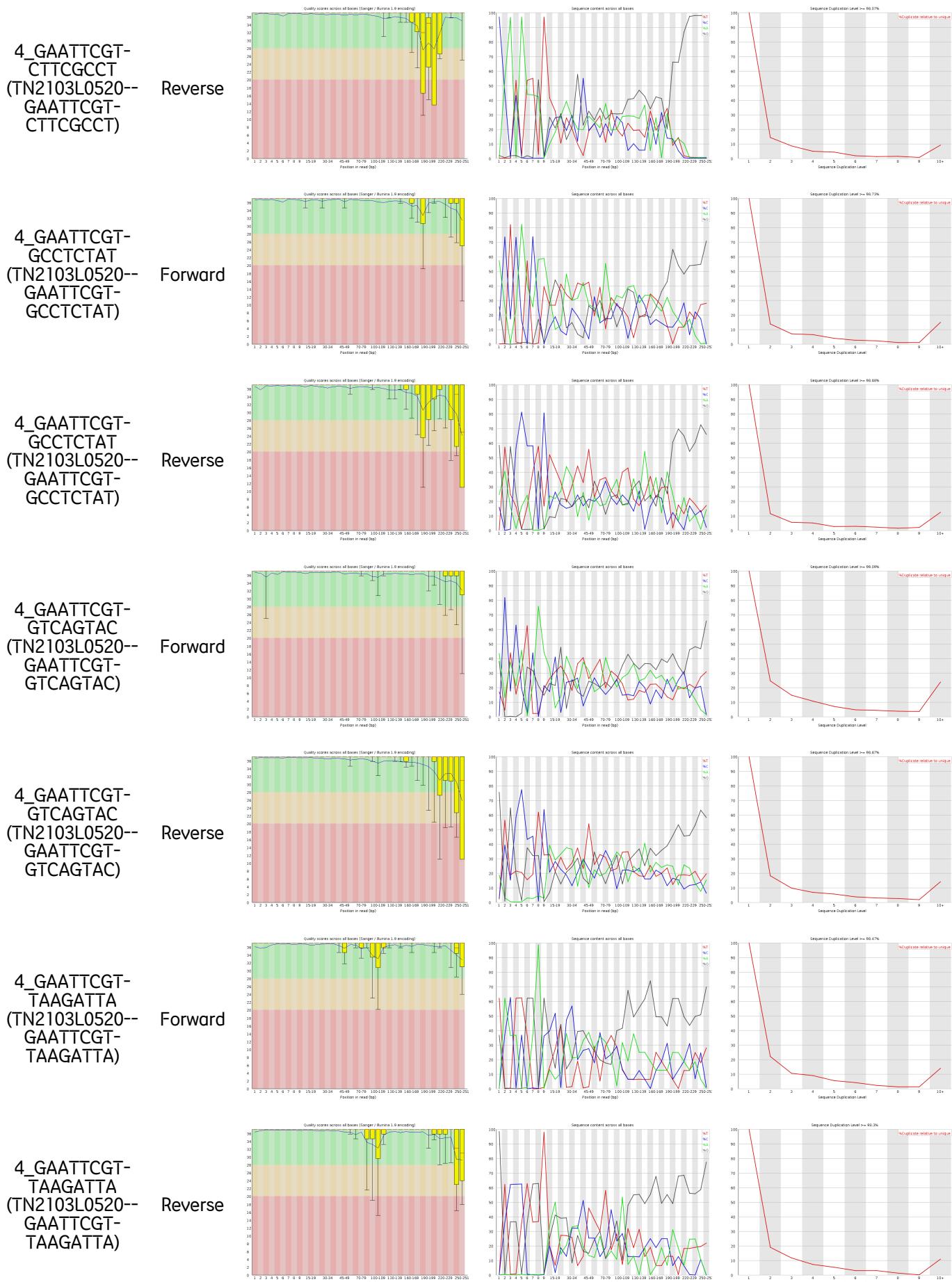


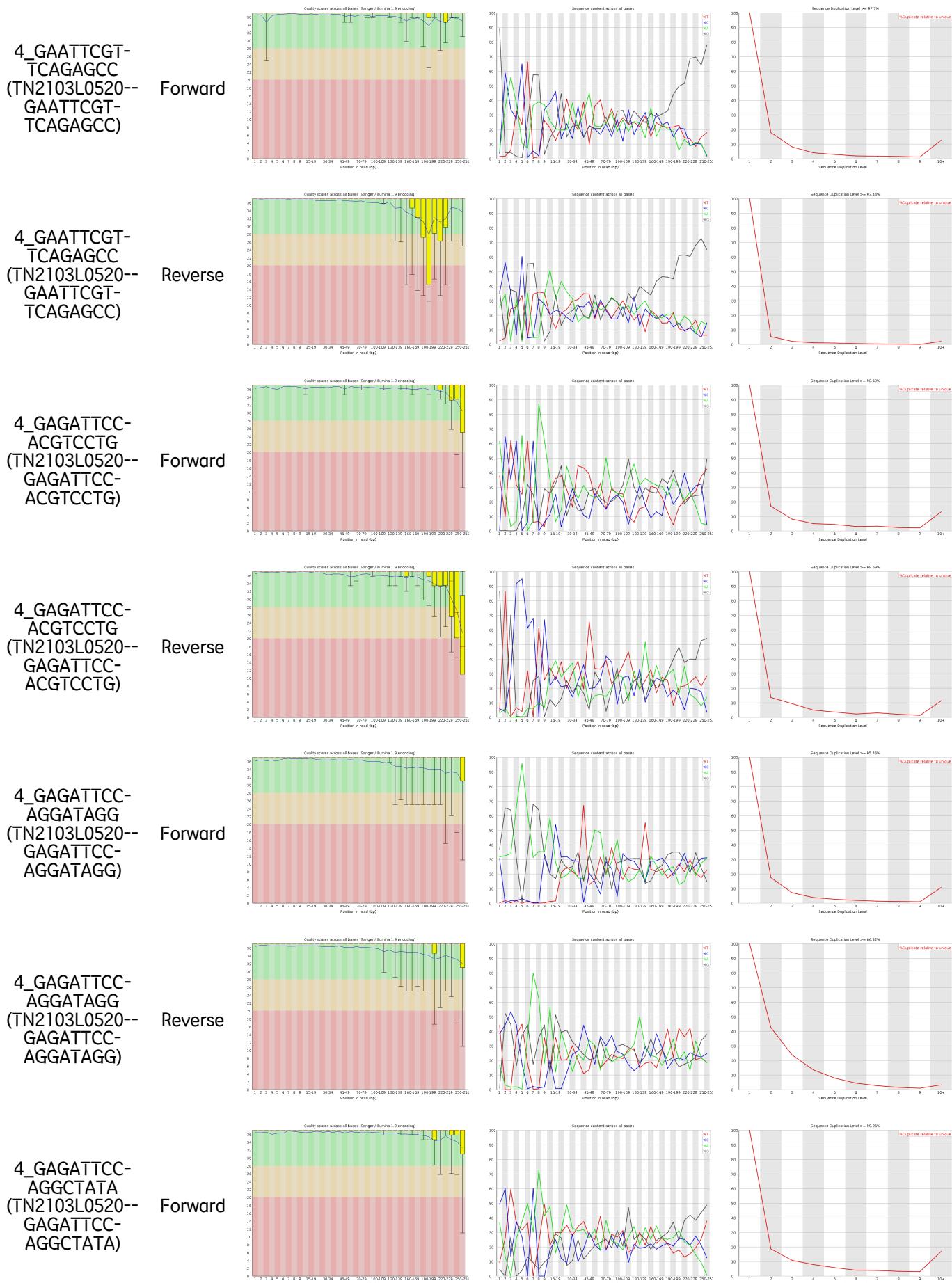


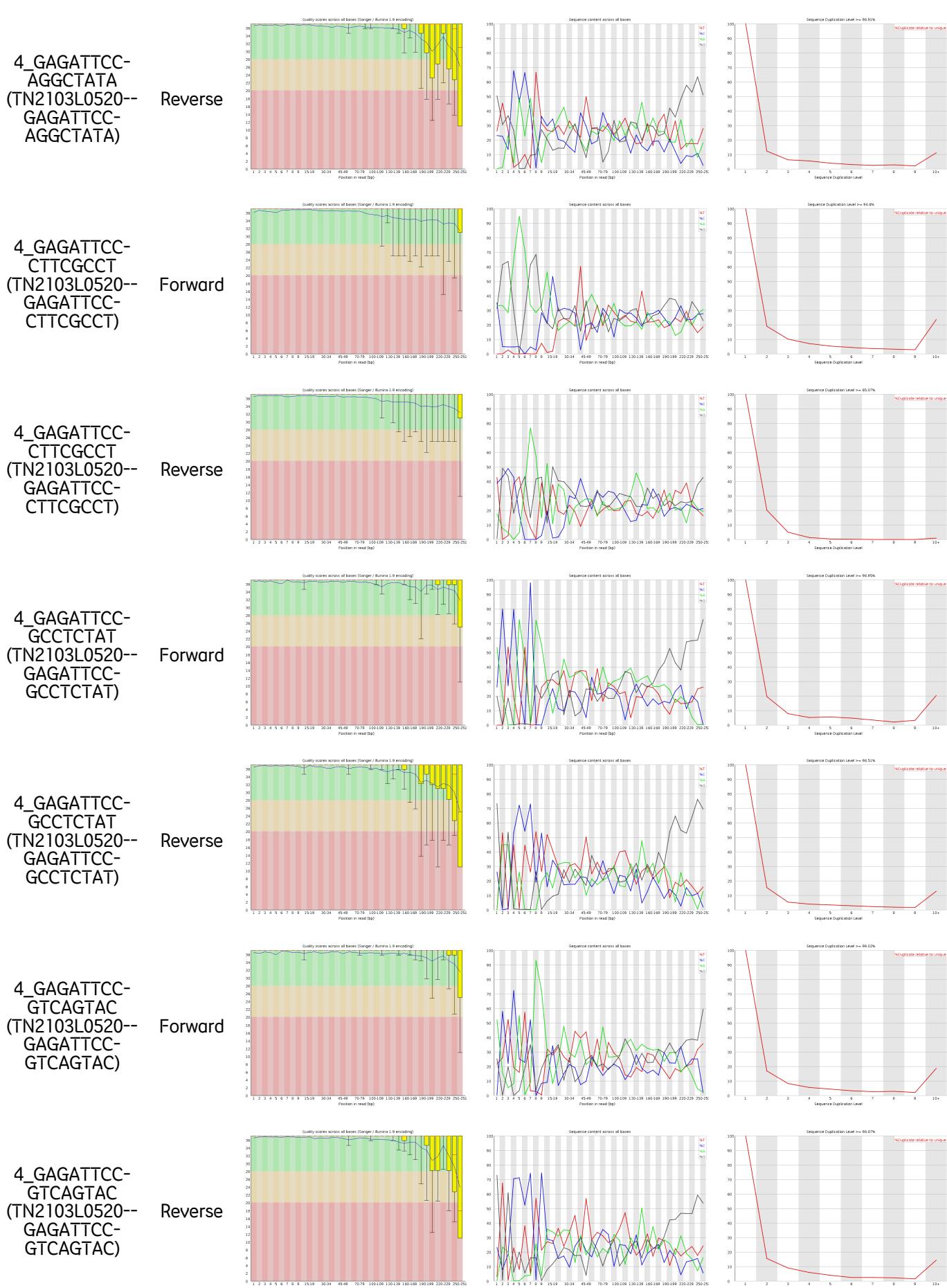


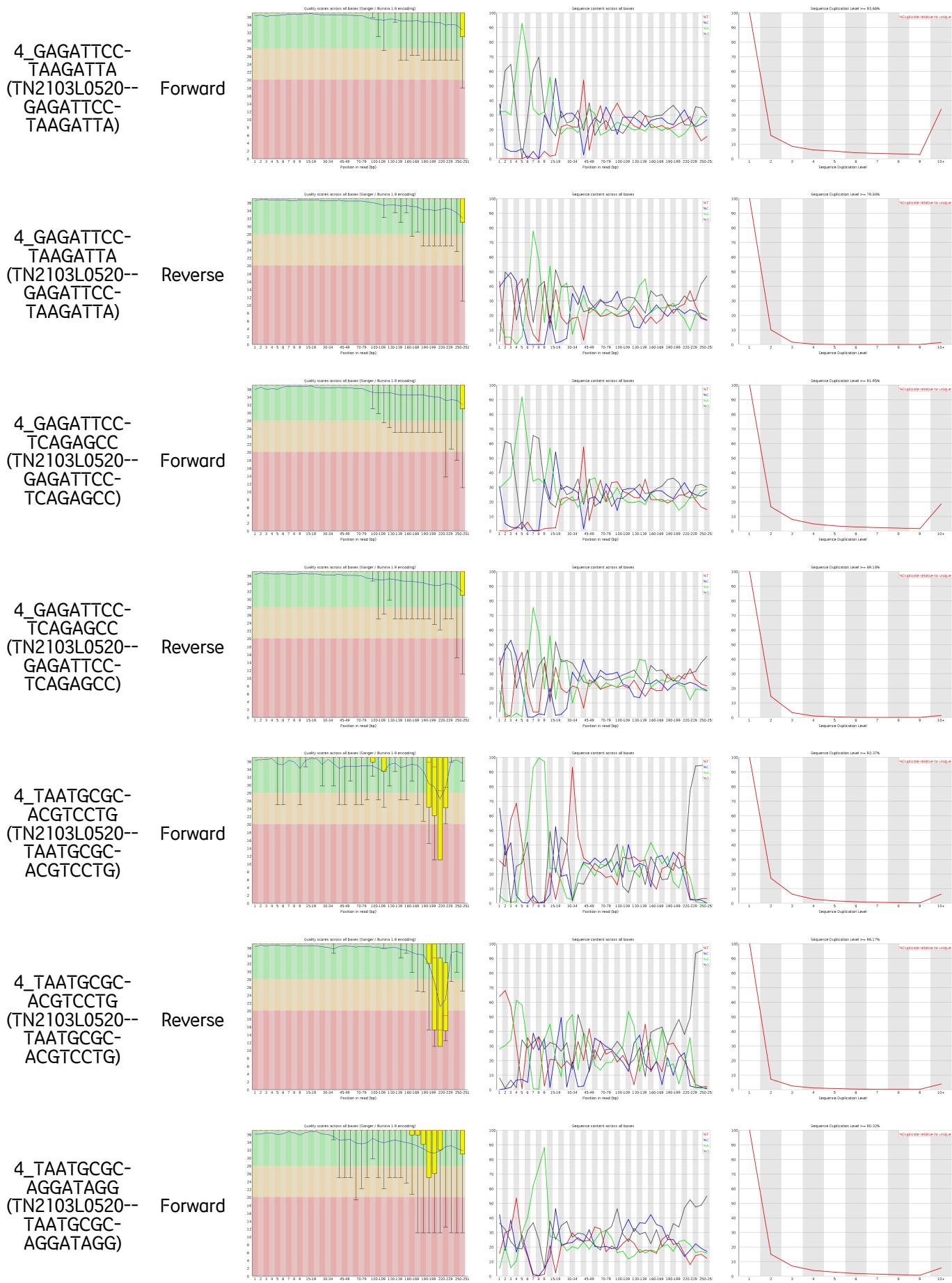


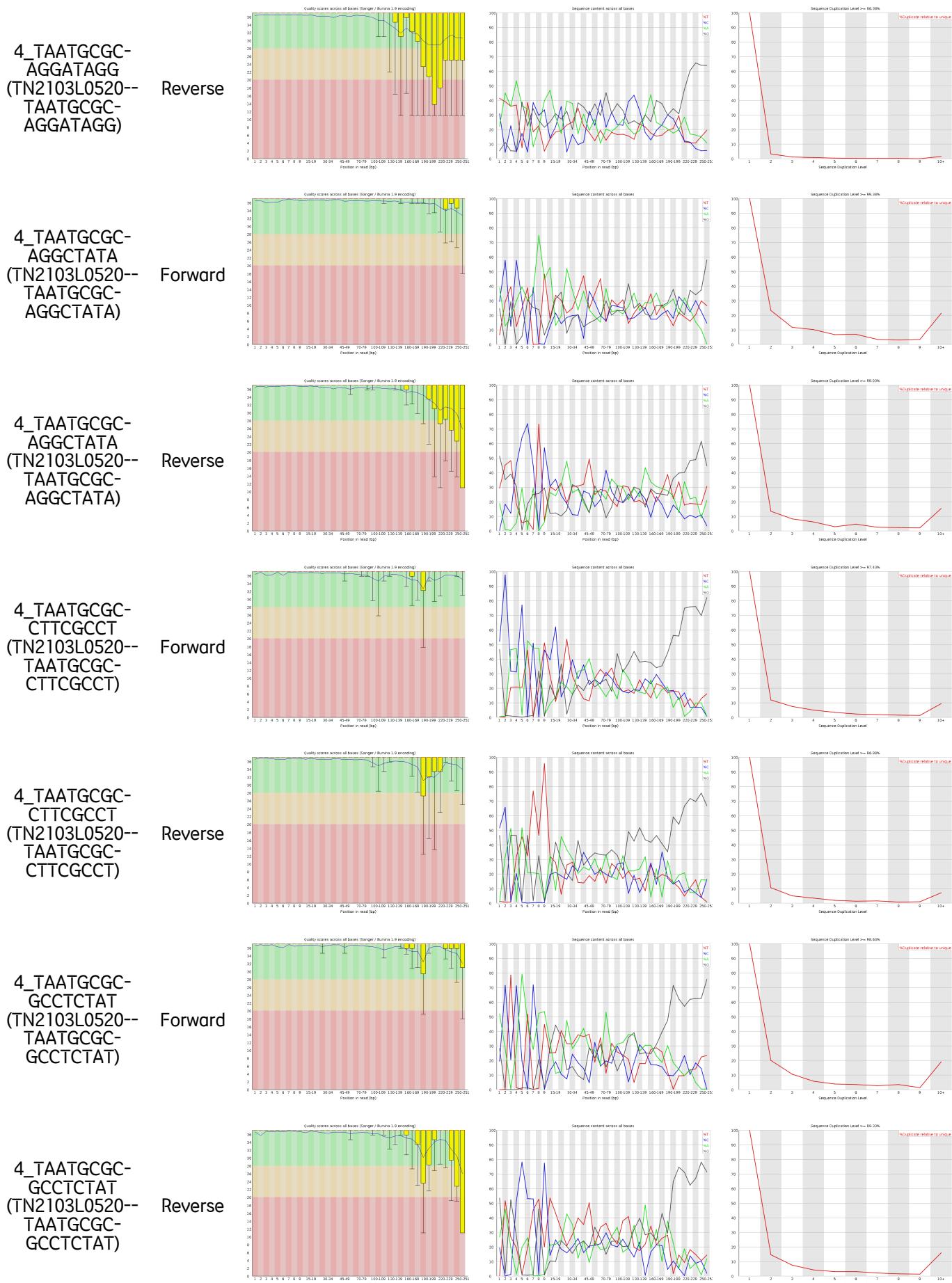


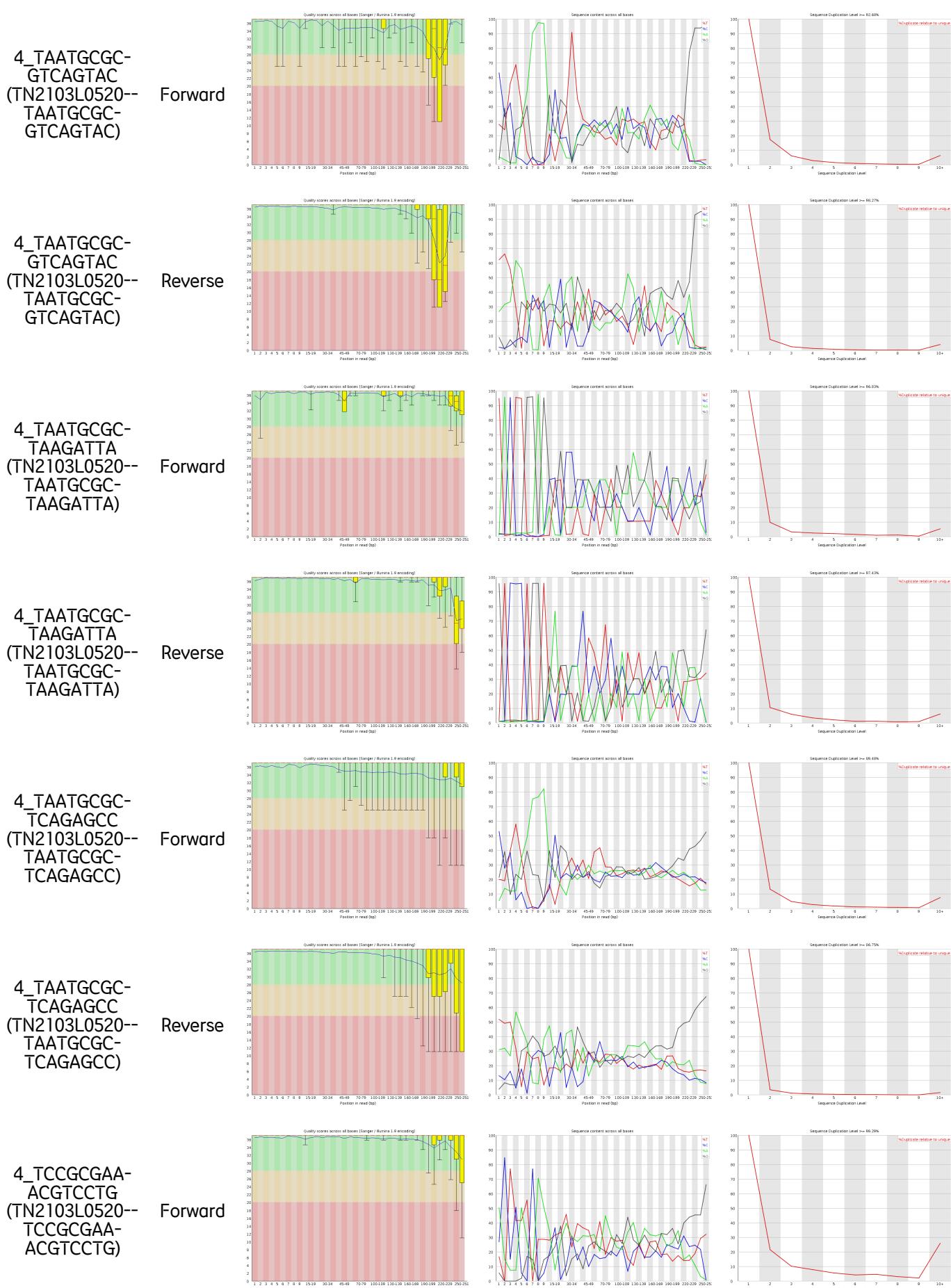


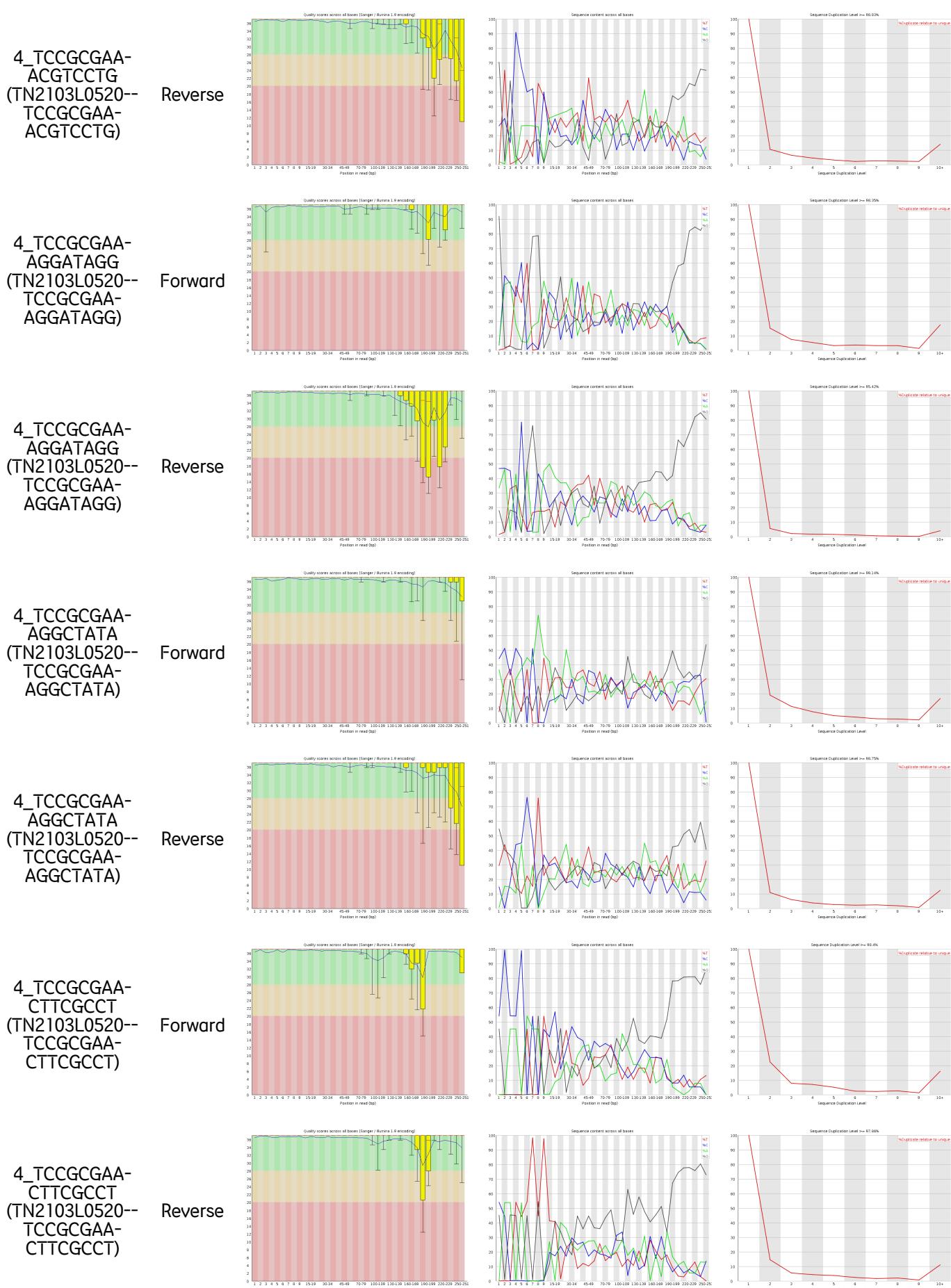


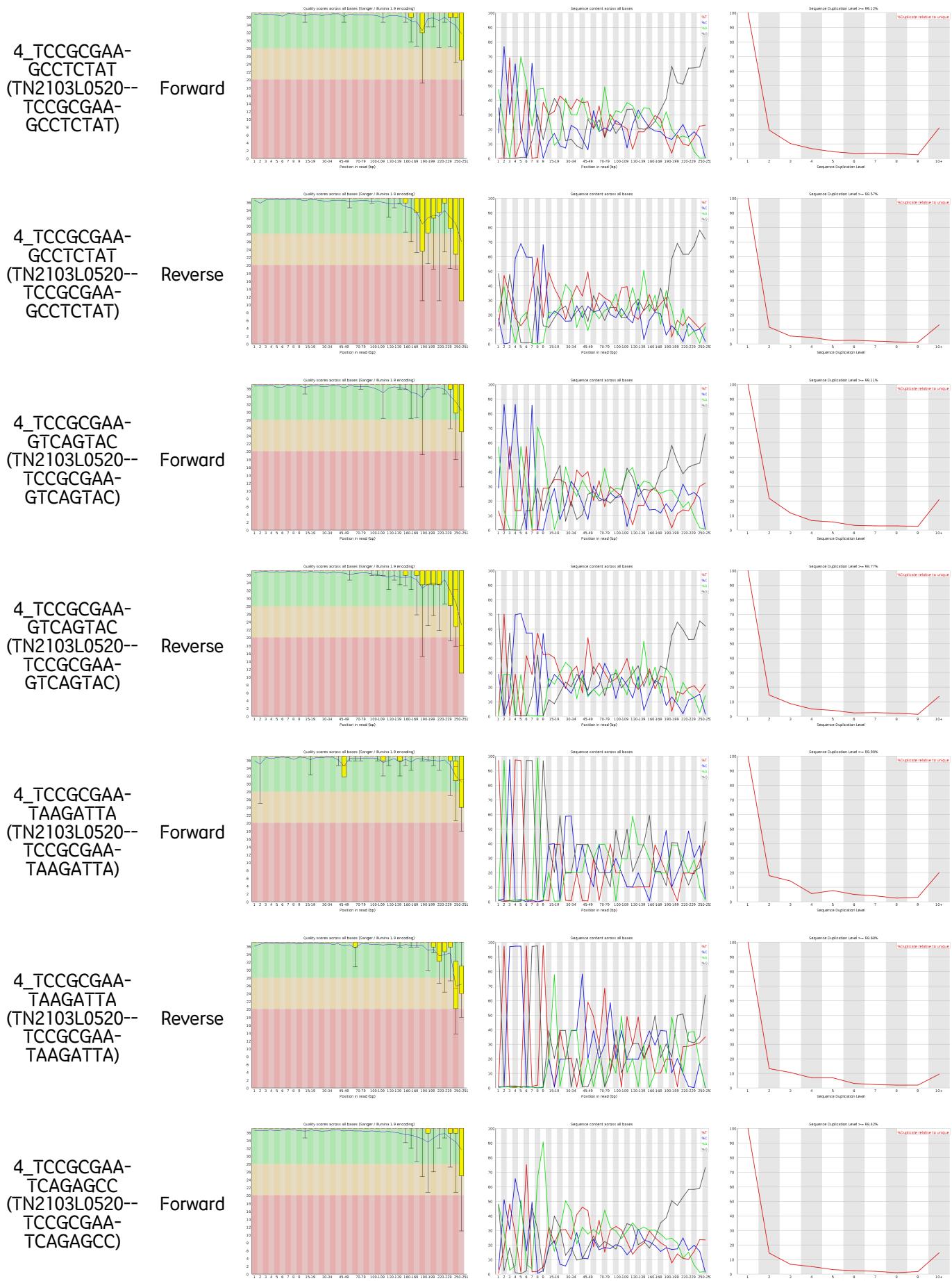


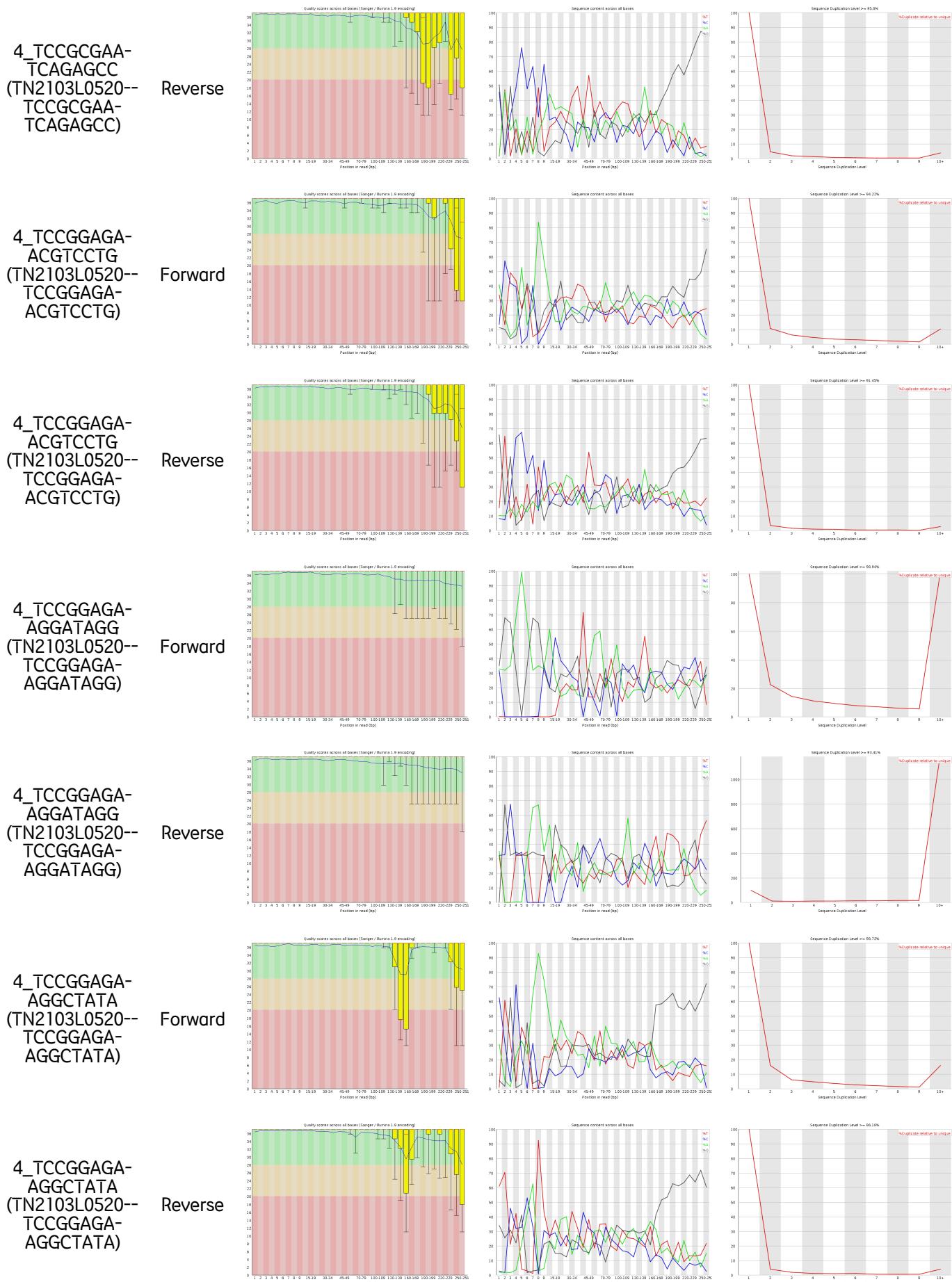


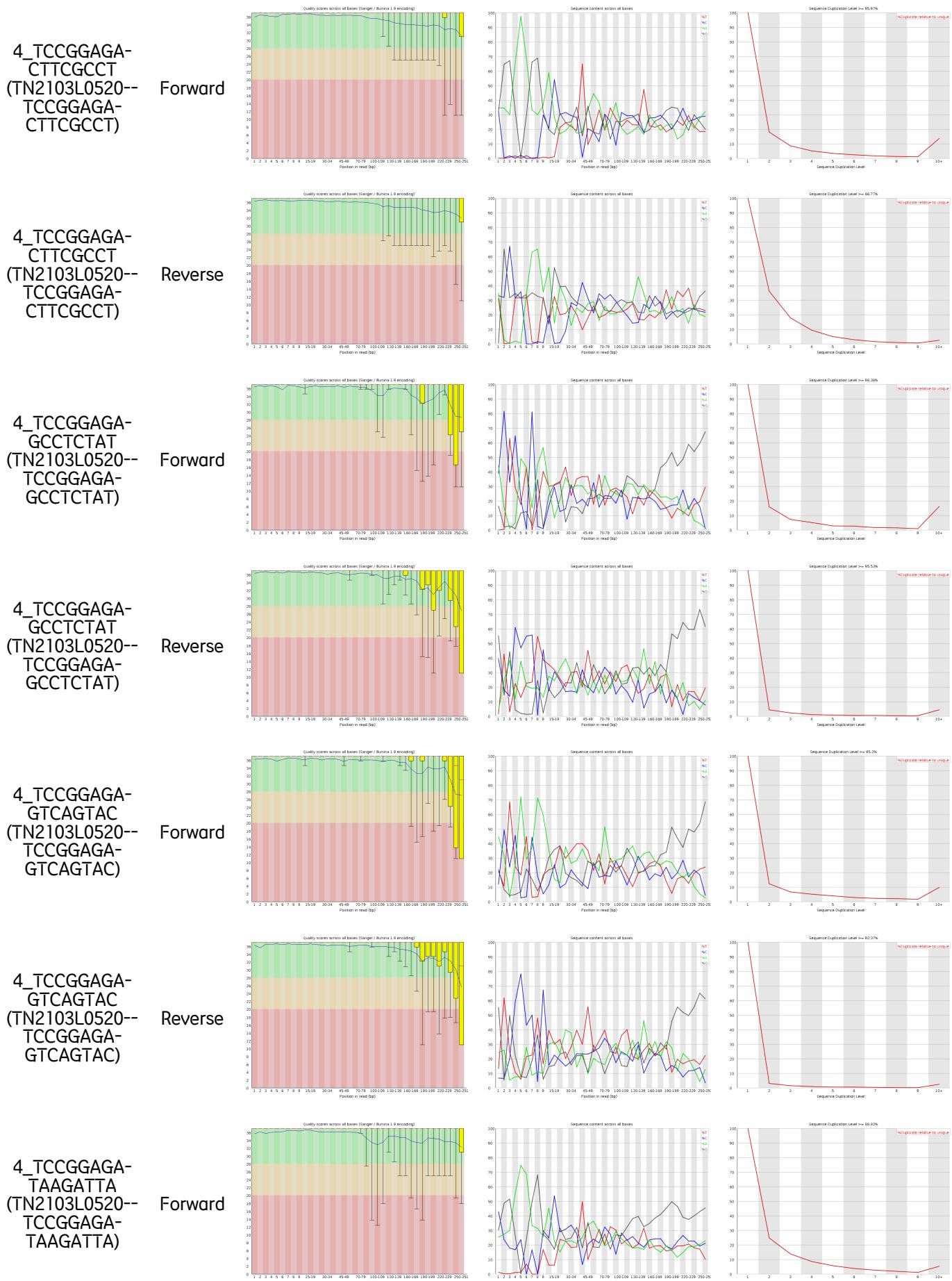


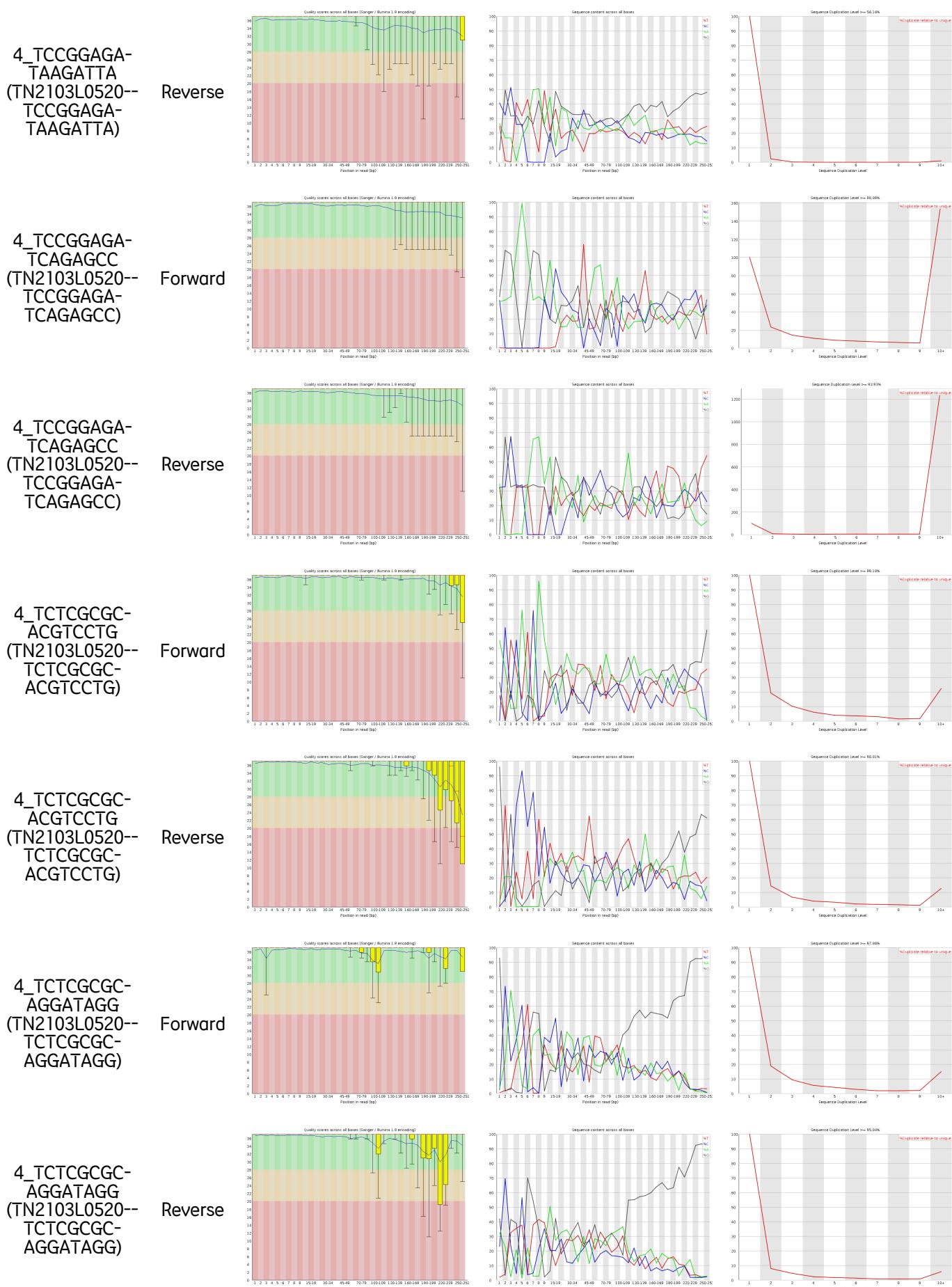


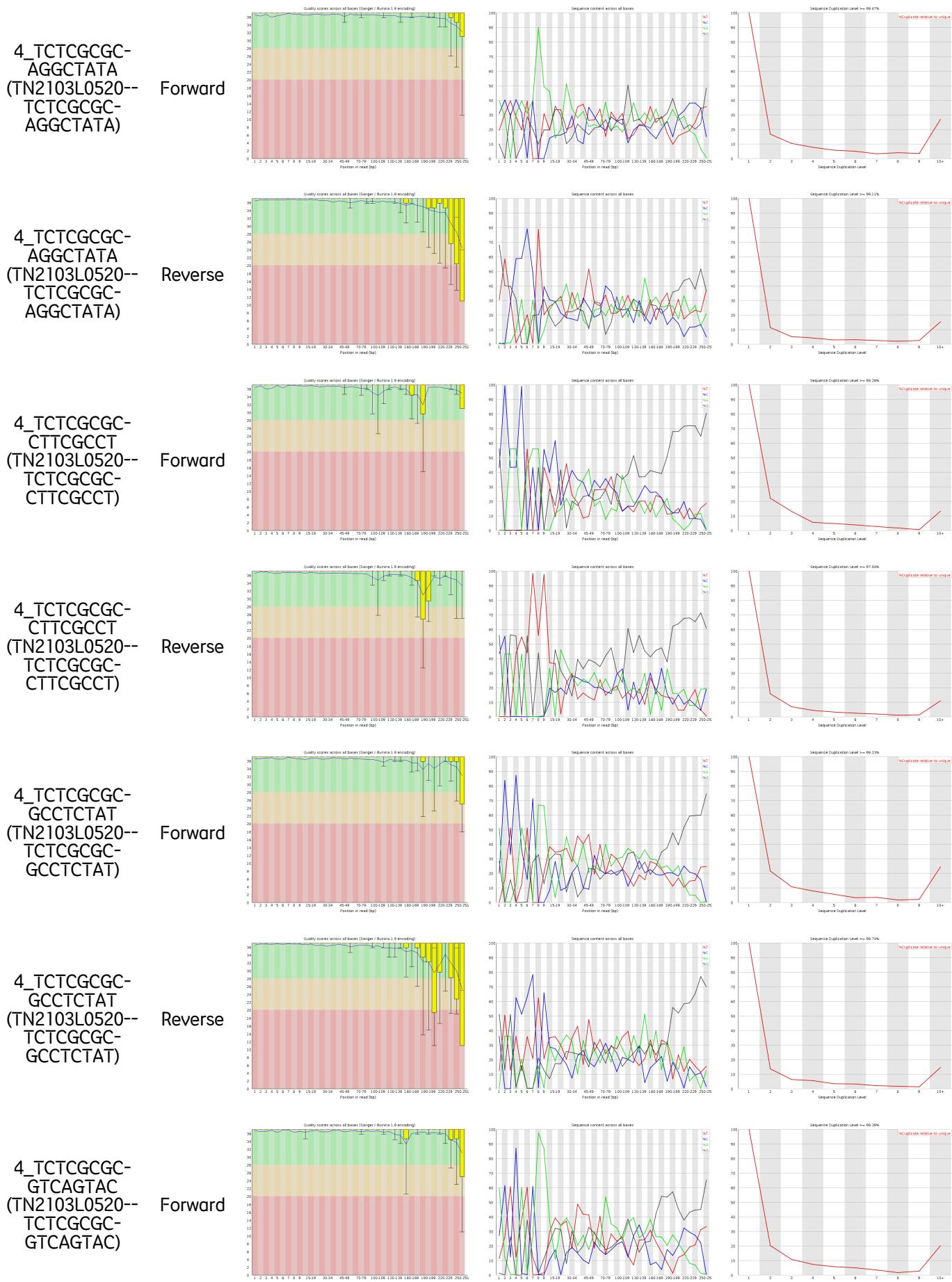


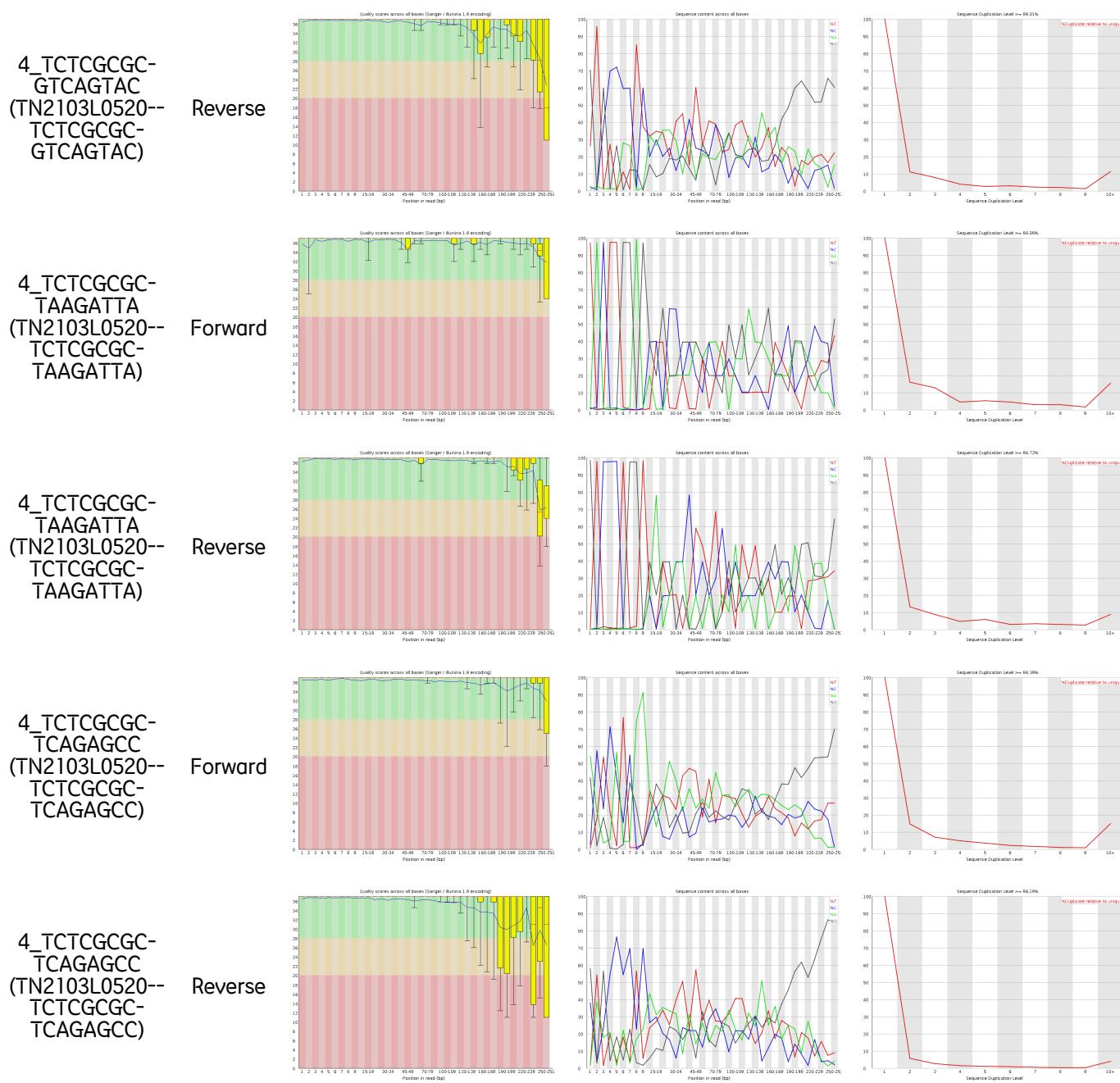












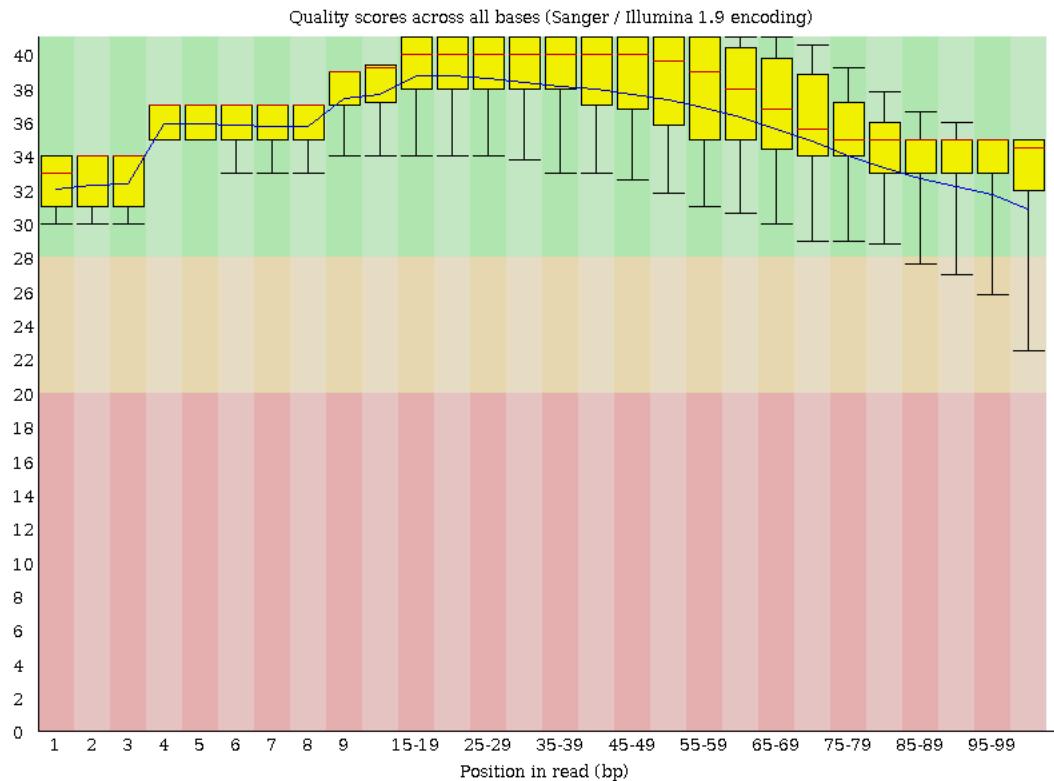
- The QC results were generated using FastQC software
- Sample ID: ID provided by customer
- Pair Name: Forward, Reverse for sequencing strand
 - Paired End(PE): Both of Forward and Reverse types are provided
 - Single End(SE): Forward type is only provided
- Per base qual: Quality values across all bases at each position
- GC contents: GC content of each base position in a sample
- Duplication level: Proportion of sequence duplication level

Appendix

4.1 Figure Description

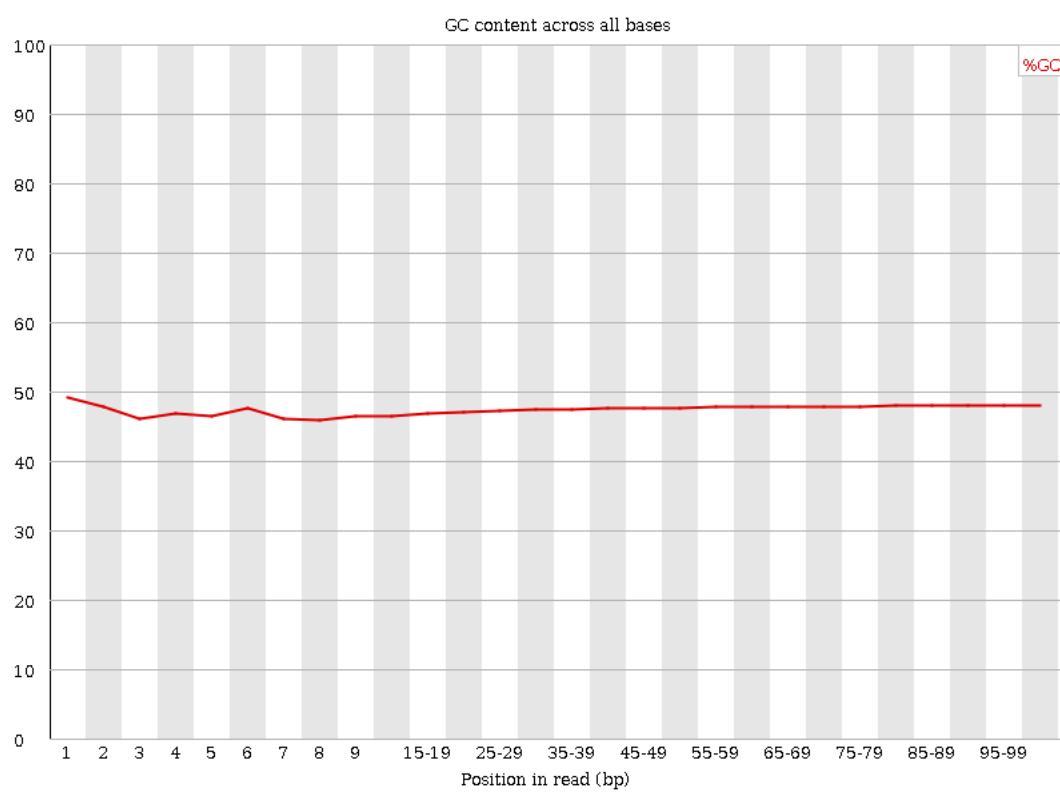
4.1.1 FastQC

Quality score across all bases



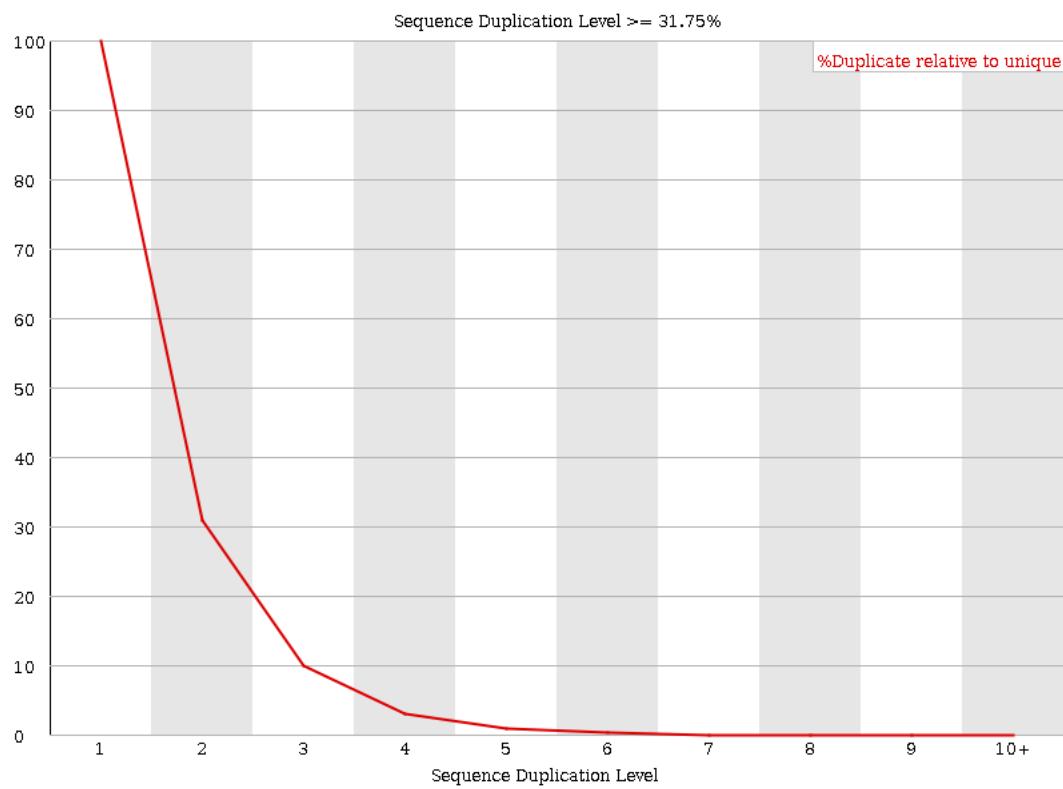
- This figure shows an overview of the range of quality values across all bases at each position in the fastq file.
- X-axis : Position in reads(bp)
- Y-axis : Quality score of sequence reads
- Green region : Very good quality calls
- Orange region: Calls of reasonable quality
- Red region : Calls of poor quality
- Central Red Line : Median Value
- Yellow Box: Inter-Quartile Range (25~75%)
- Upper and Lower whiskers : 10% and 90% points
- Blue Line : mean quality
- Warning: A warning will be issued if the lower quartile for any base is less than 10, or if the median for any base is less than 25.

GC content across all bases



- This figure shows an overview of the GC content of each base position in a file.
- X-axis: Position in reads (bp)
- Y-axis: GC rate of each base position

Sequence Duplication Level



- This figure shows an overview of the degree of duplication for every sequence in the set and the relative number of sequences with different degrees of duplication.
- X-axis: Sequence Duplication Level
- Y-axis: Duplicate relative to unique
- Warning: This module will issue a warning if non-unique sequences make up more than 20% of the total.

4.2 Explanation of Term Regarding Sequencing Statistics Result

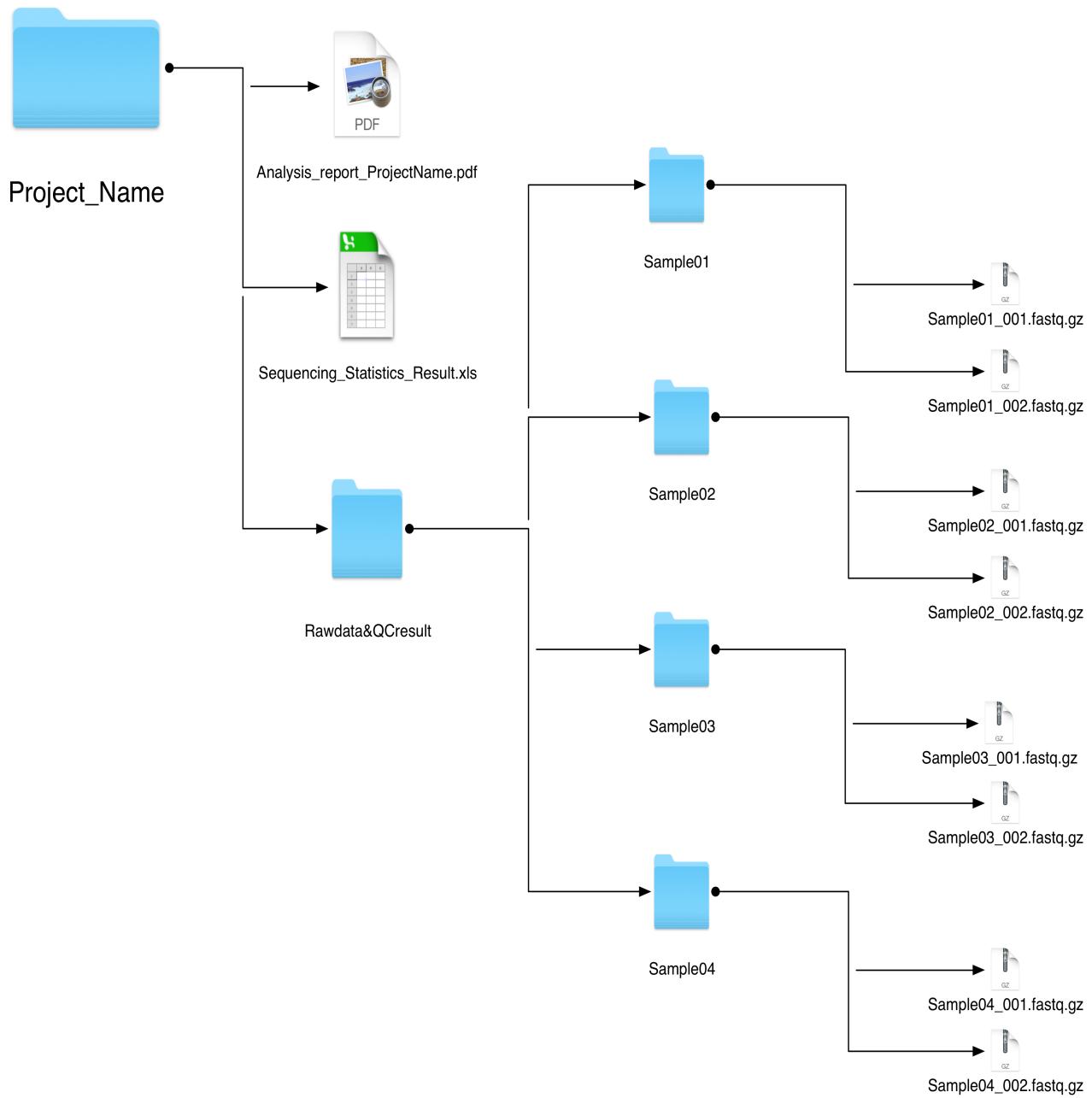
Term Explanation

Term	Explanation
Total Reads	Total number of produced reads
Total Bases	Total number of produced nucleotides
Total Bases(Gb)	The total bases expressed by giga unit
GC Count	Number of both guanine(G) and cytosine(C) of the total bases
GC Rate	GC ratio of the total bases
N ZeroReads	Number of reads not including N(any base) nucleotide
N ZeroReadsRate	Proportion of the N ZeroReads of the total reads
N5 LessReads	Number of reads including less than 5 times of N(any base) nucleotide in each read
N5 LessReadsRate	Proportion of the N5 LessReads of the total reads
N Count	Total number of N(any base) nucleotides
N Rate	Proportion of the N Count of total bases
Q30 MoreBases	Number of nucleotides greater than Phred quality score 30 (nucleotide accuracy: 99.9%) of the total bases
Q30 MoreBasesRate	Proportion of the Q30 MoreBases of the total bases
Q20 MoreBases	Number of nucleotides greater than Phred quality score 20 (nucleotide accuracy: 99%) of the total bases
Q20 MoreBasesRate	Proportion of the Q20 MoreBases of the total bases

Explanation of Q-Score

Quality Score	Error rate	Nucleotide accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

4.3 Directory Structure



Location of total results (i.e. rawdata, fastqc, statistics etc) is followed by the above figure. To access our FTP server, please refer to the FTP related information (i.e. server IP, ID, PW) provided by Theragen Etex.

TotalOmics

Sequencing for Everyone