



Mapping Report for Case A Read Mapping
Type: Read mapping
Reference count: 1
Generated by: sevin
Date: Sat Aug 15 10:48:54 MDT 2020
Software: CLC Genomics Workbench 20.0.4

Table of contents

1. Summary	3
2. References	3
2.1 Reference coverage	3
2.2 Coverage statistics	3
2.3 Coverage level distribution	4
2.4 Coverage level distribution (excl. zero coverage regions)	5
2.5 Zero coverage regions	6
2.6 Coverage vs GC Content Plot	6
3. Mapped reads	6
3.1 All mapped reads	7
3.2 Non-specific matches	7
3.3 Non-perfect matches	8
3.4 Paired reads	9
3.5 Read length distribution	10
3.6 Insertion length distribution	11
3.7 Deletion length distribution	11
3.8 Nucleotide differences in reads relative to reference	11
3.9 Nucleotide Mapping	12
3.10 Quality for match distribution	15
3.11 Quality for mismatch distribution	16
3.12 Mismatch fraction per read position	16
3.13 Unaligned ends	16

1. Summary

Reference count	1
Type	Read mapping
Total reference length	29,903
GC contents in %	37.97
Total read count	961,359
Mean read length	75.45
Total read length	72,531,806

2. References

2.1 Reference coverage

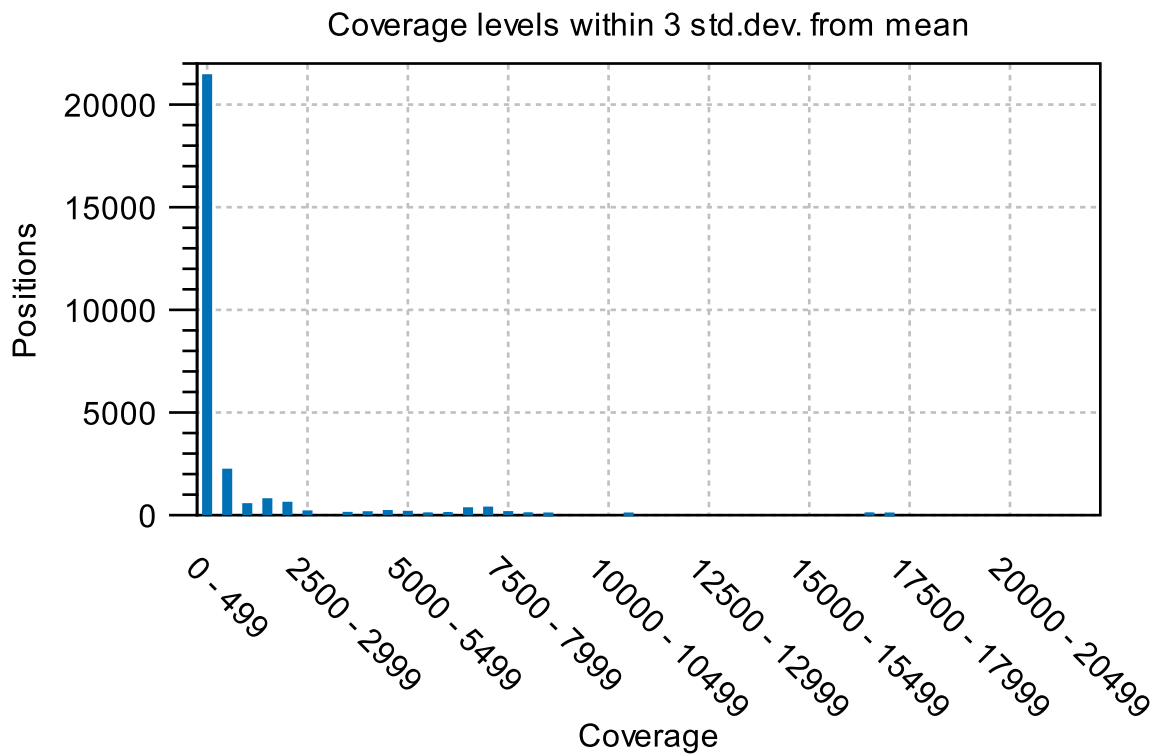
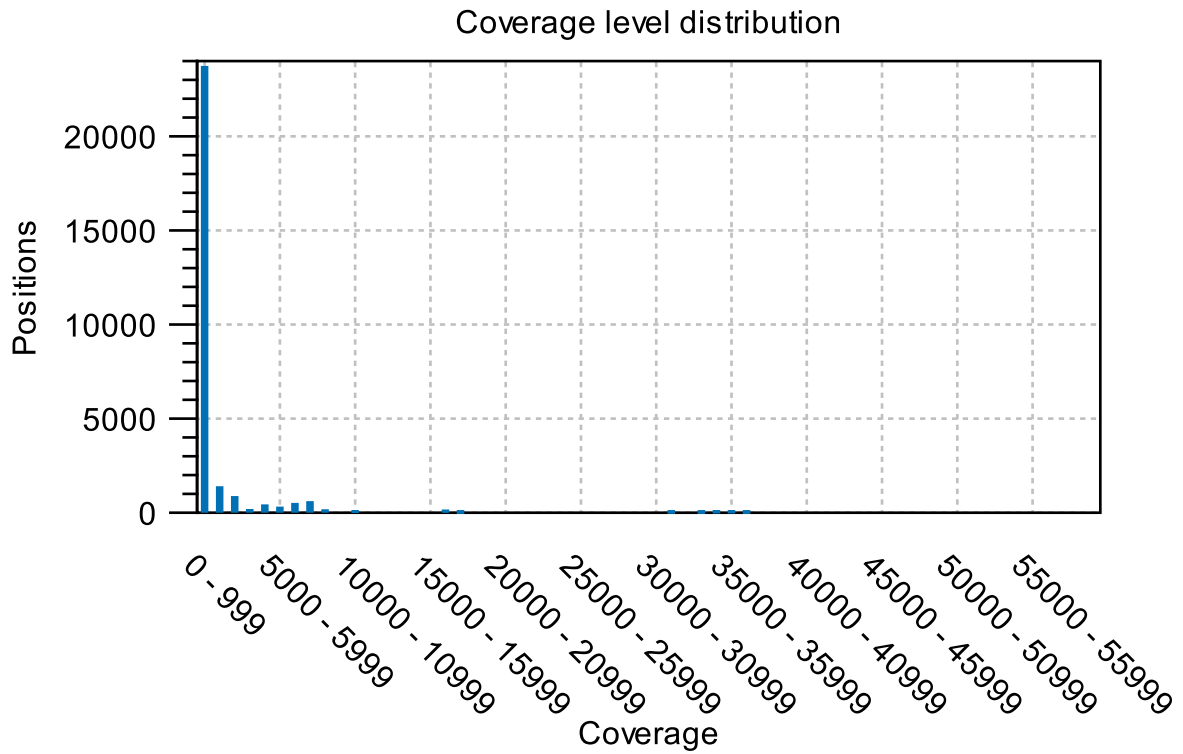
Total reference length	29,903
% GC	37.97
Fraction of reference covered	1.00

2.2 Coverage statistics

Total reference length	29,903
Minimum coverage	0
Maximum coverage	59,819
Median coverage	246.00
Average coverage	2,257.96
Standard deviation	6,761.18
Minimum excl. zero coverage regions	5
Median excl. zero coverage regions	246.00
Average excl. zero coverage regions	2,258.42
Standard deviation excl. zero coverage regions	6,761.78

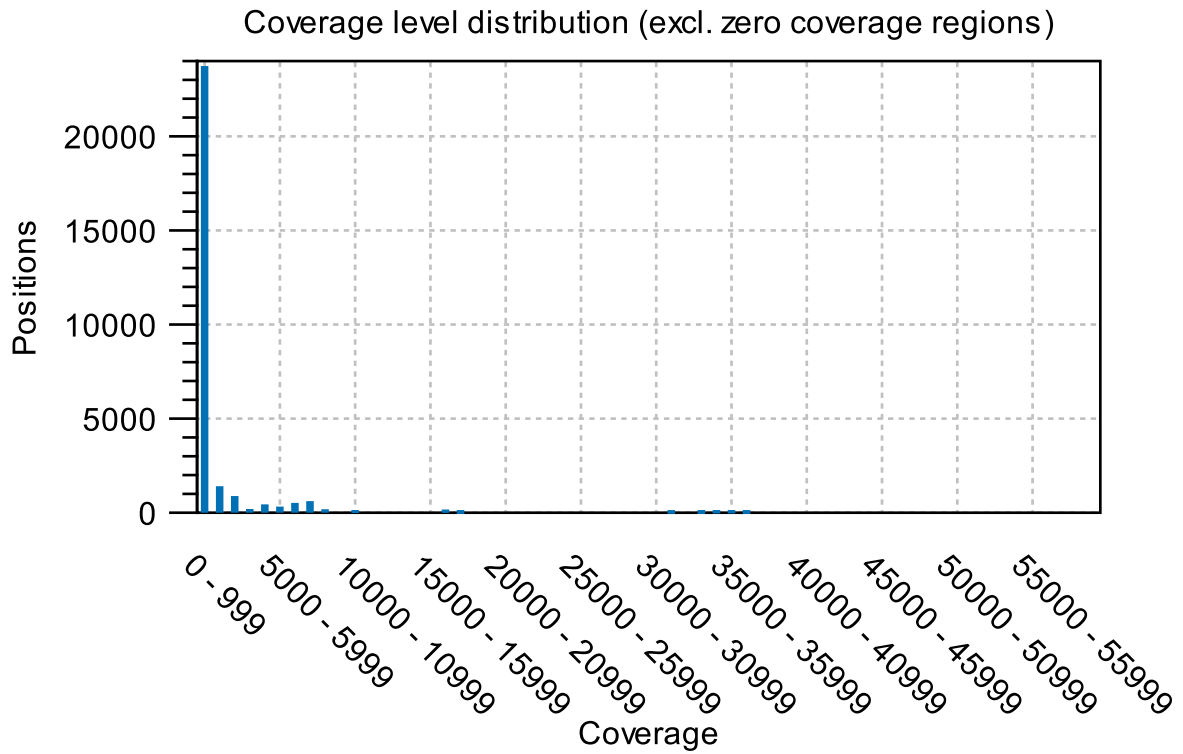
Note that positions with an ambiguous nucleotide in the reference (i.e. not A,C,T or G), count as zero coverage regions, regardless of the number of reads mapping across them.

2.3 Coverage level distribution

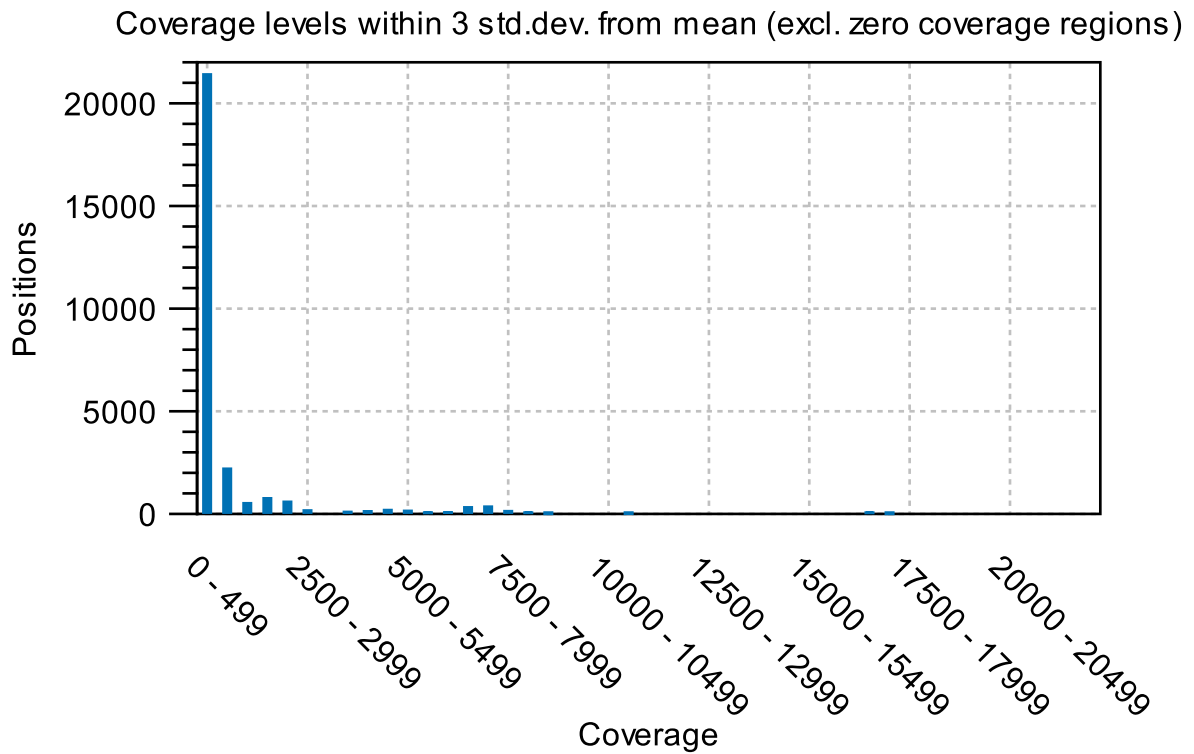


28,962 positions have coverage between 0 and 22,475.
941 positions have coverage above 22,475 (not shown in graph).

2.4 Coverage level distribution (excl. zero coverage regions)



6 positions have zero coverage (not shown in graph).



6 positions have coverage below 5 (not shown in graph).
 28,956 positions have coverage between 5 and 22,475.
 941 positions have coverage above 22,475 (not shown in graph).

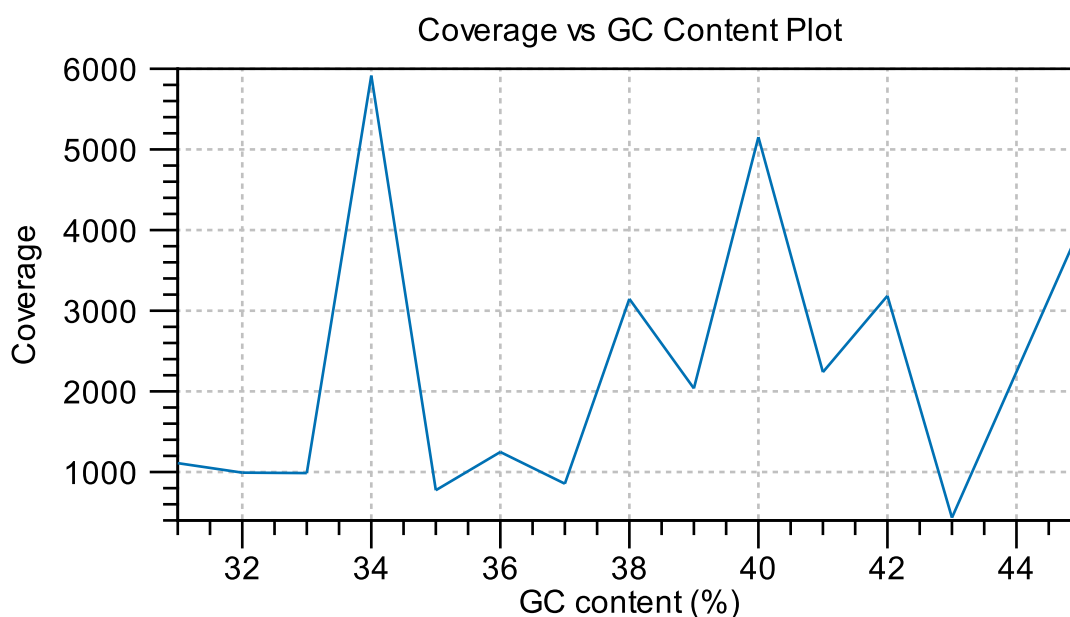
Note that positions with an ambiguous nucleotide in the reference (i.e. not A,C,T or G), count as zero coverage regions, regardless of the number of reads mapping across them.

2.5 Zero coverage regions

Count	1
Minimum length	6
Maximum length	6
Mean length	6.00
Standard deviation	0.00
Total length	6

Note that positions with an ambiguous nucleotide in the reference (i.e. not A,C,T or G), count as zero coverage regions, regardless of the number of reads mapping across them.

2.6 Coverage vs GC Content Plot



The plot displays, for each GC content level (0-100%), the mean read coverage of 100bp reference segments with that GC content.

3. Mapped reads

3.1 All mapped reads

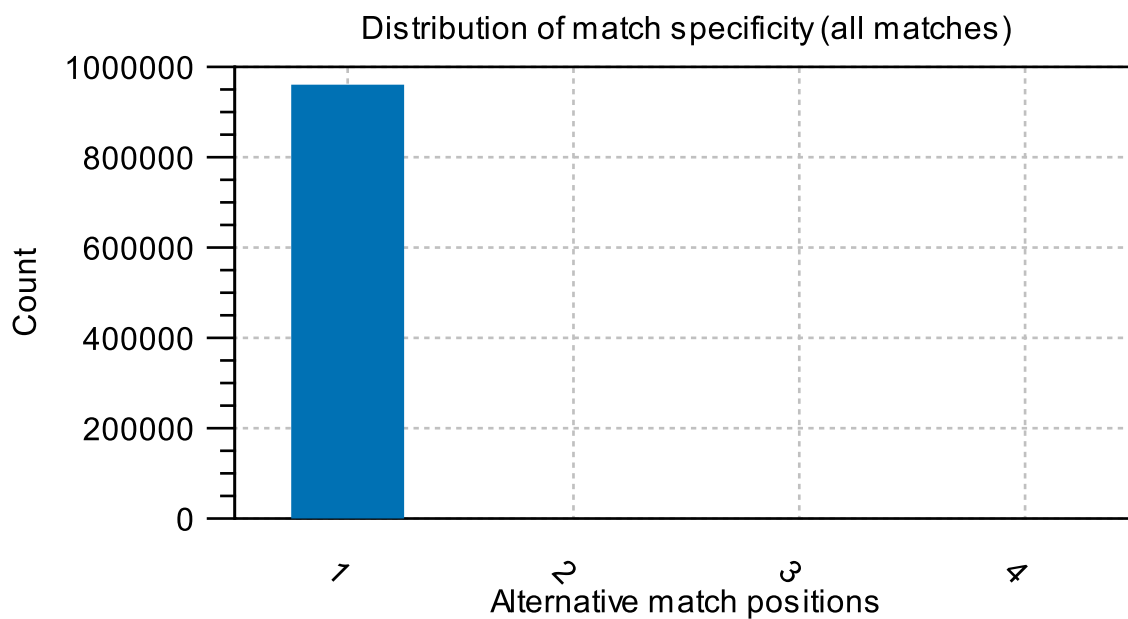
Read count	961,359
Mean read length	75.45
Total read length	72,531,806

3.2 Non-specific matches

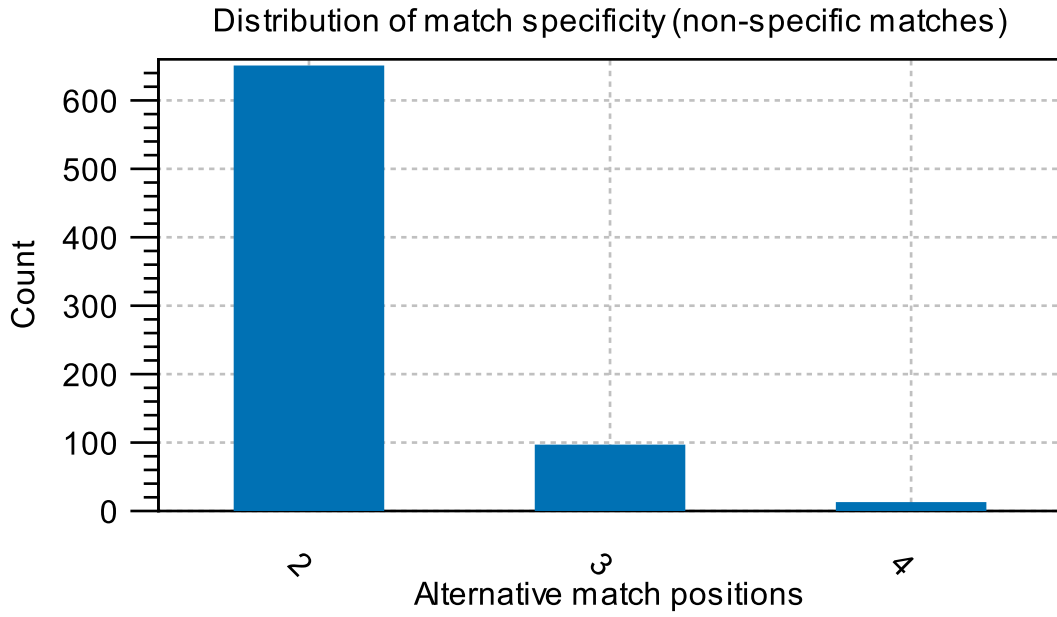
Non-specific matches

Read count	761
% of all mapped reads	0.08
Mean read length	75.11
Total read length	57,156

Distribution of match specificity (all matches)



Distribution of match specificity (non-specific matches)

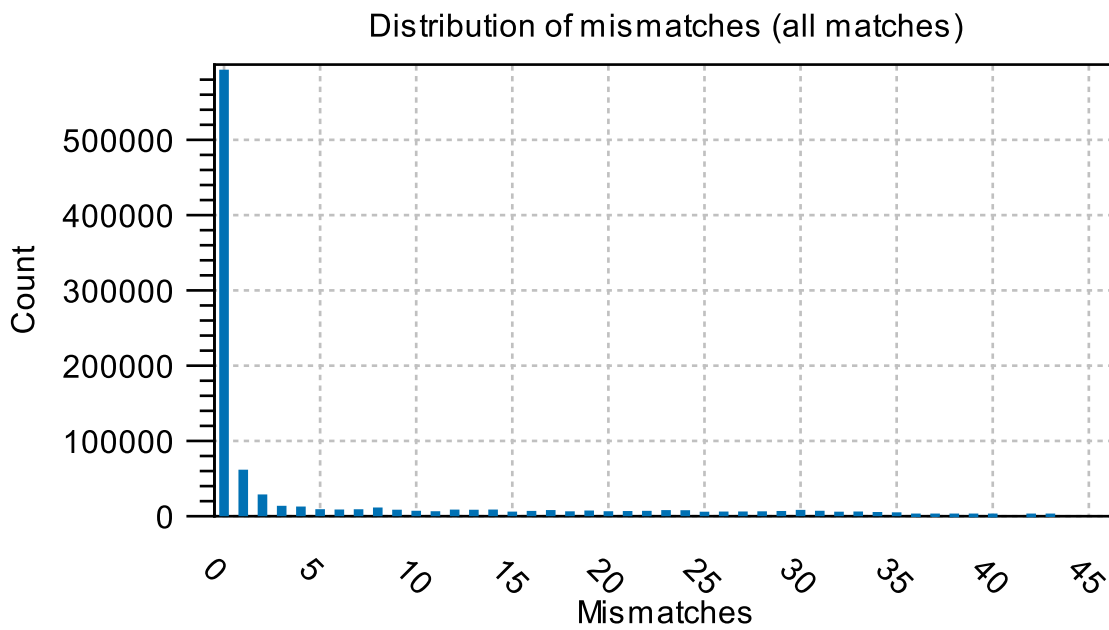


3.3 Non-perfect matches

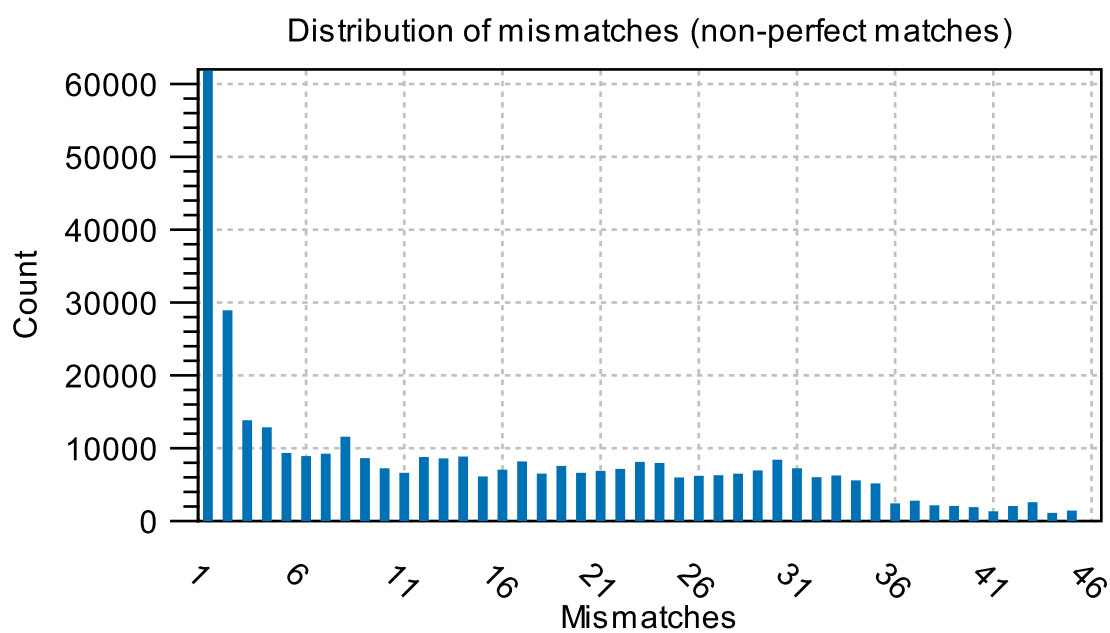
Non-perfect matches

Read count	368,147
% of all mapped reads	38.29
Mean read length	75.41
Total read length	27,762,444

Distribution of mismatches (all matches)

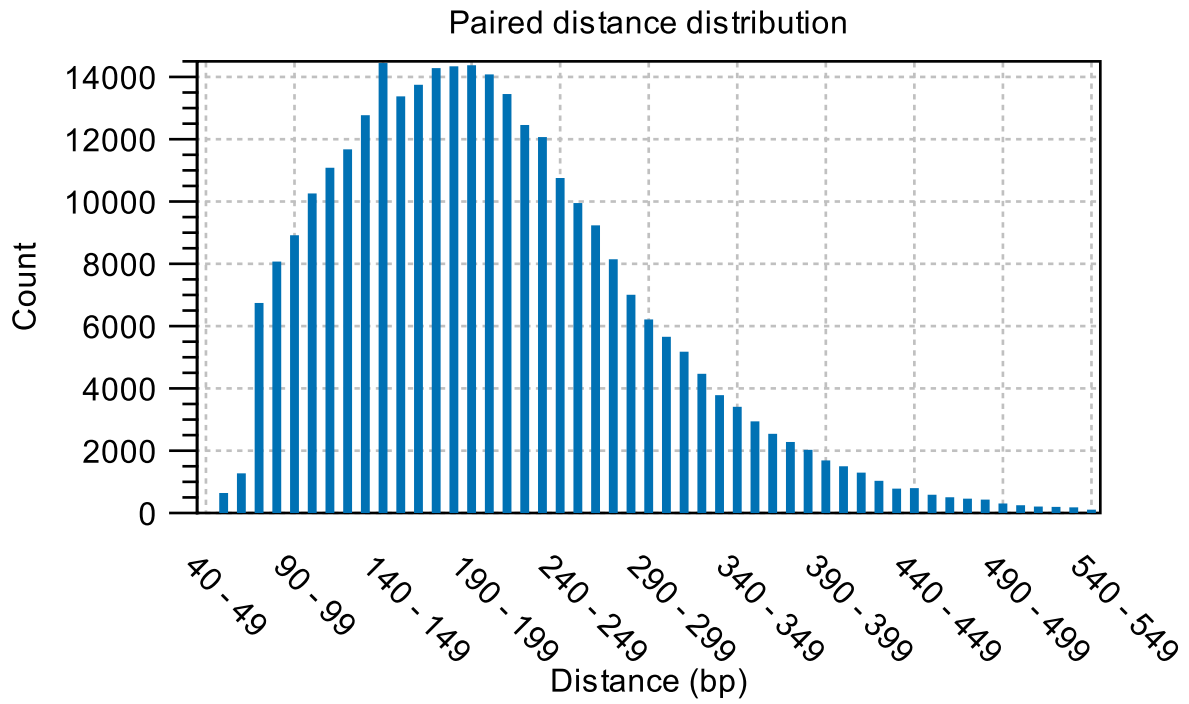


Distribution of mismatches (non-perfect matches)



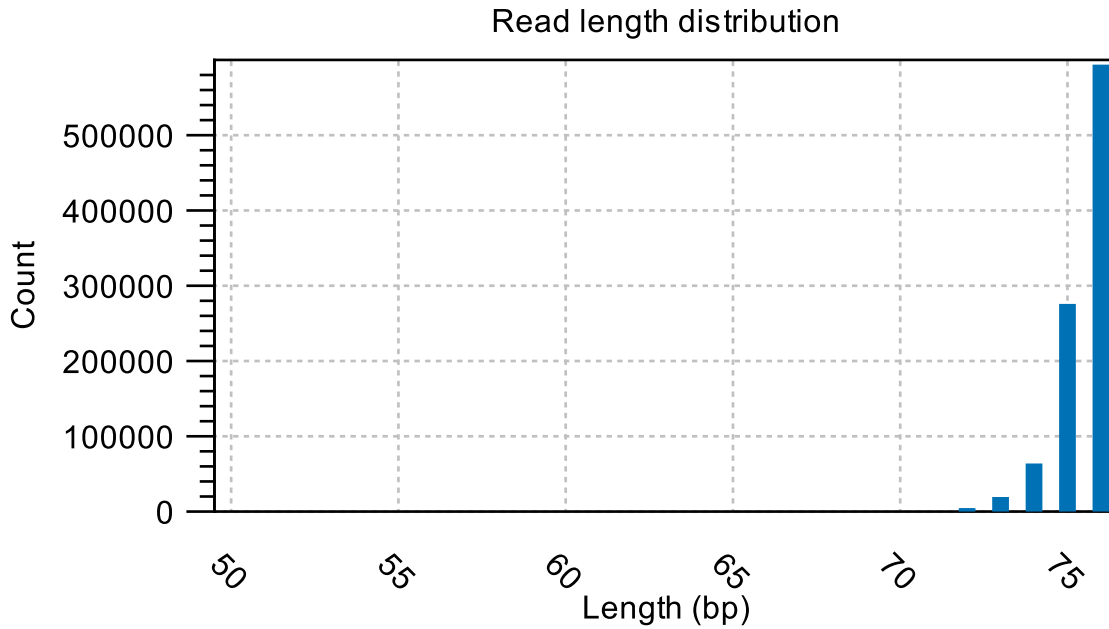
3.4 Paired reads

Reads in aligned pairs	603,982
Reads in broken pairs: wrong distance or mate inverted	314,324
Reads in broken pairs: mate on other contig	0
Reads in broken pairs: mate not mapped	43,053
Mean distance	206.70
Standard deviation	85.51

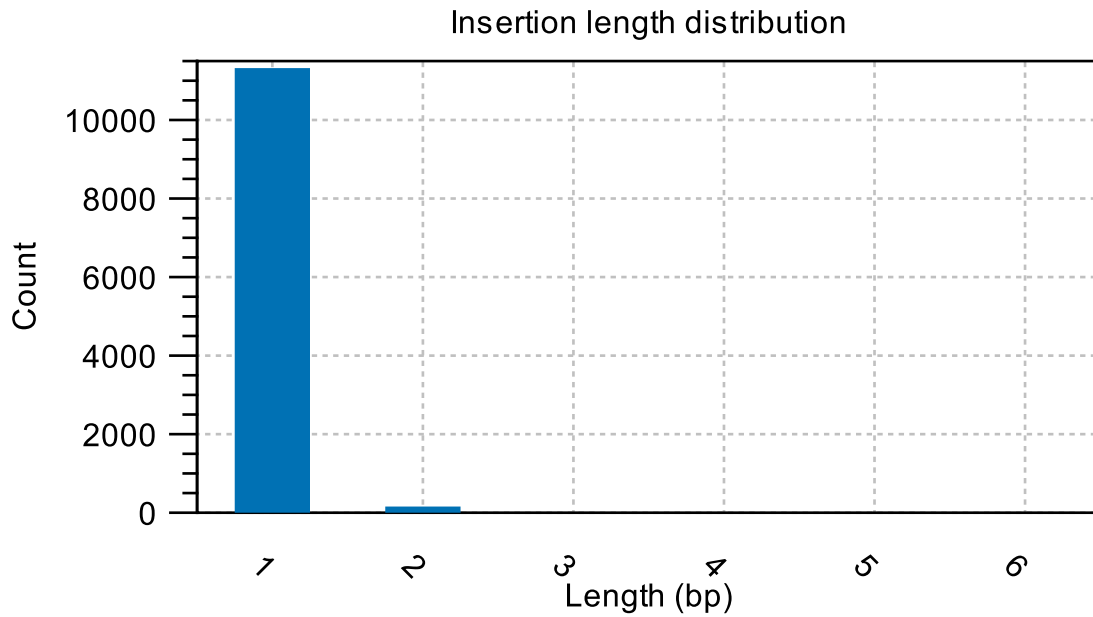


Paired distance for intact pairs (including unaligned ends)

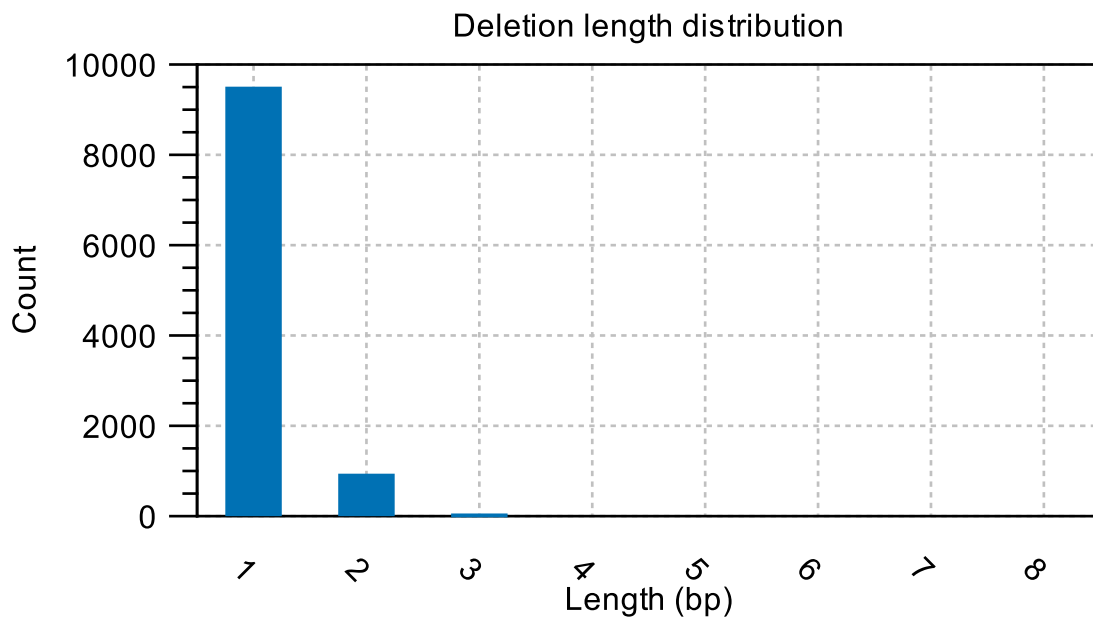
3.5 Read length distribution



3.6 Insertion length distribution



3.7 Deletion length distribution



3.8 Nucleotide differences in reads relative to reference

Nucleotide in reference	% read bases that differ
A	0.48
C	0.42

Nucleotide in reference	% read bases that differ
G	0.65
T	0.48
-	0.07
Total	0.42

3.9 Nucleotide Mapping

Counts

	Read: A	Read: C	Read: G	Read: T
Reference: A	20,258,417	13,102	28,965	50,560
Reference: C	25,887	12,445,641	12,412	13,537
Reference: G	20,536	35,909	13,676,963	29,715
Reference: T	53,527	30,032	14,454	20,798,594
Reference: -	3,363	1,823	1,508	5,035
Total	20,361,730	12,526,507	13,734,302	20,897,441

Read: -	Total
4,807	20,355,851
1,155	12,498,632
3,519	13,766,642
2,067	20,898,674
16,406,022	16,417,751
16,417,570	83,937,550

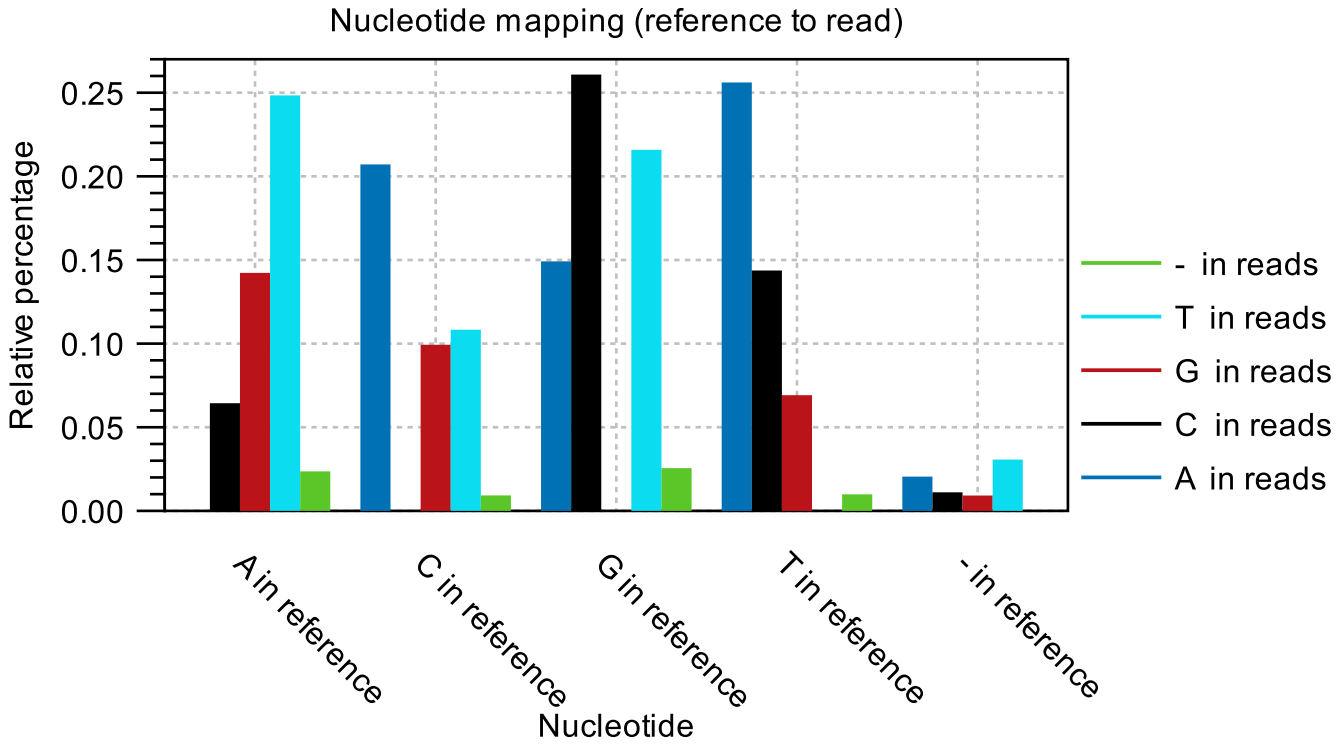
Percentages

	Read: A	Read: C	Read: G	Read: T
Reference: A	24.14	0.02	0.03	0.06
Reference: C	0.03	14.83	0.01	0.02
Reference: G	0.02	0.04	16.29	0.04
Reference: T	0.06	0.04	0.02	24.78
Reference: -	0.00	0.00	0.00	0.01
Total	24.26	14.92	16.36	24.90

Read: -	Total
0.01	24.25
0.00	14.89
0.00	16.40
0.00	24.90
19.55	19.56

Read: -	Total
19.56	100.00

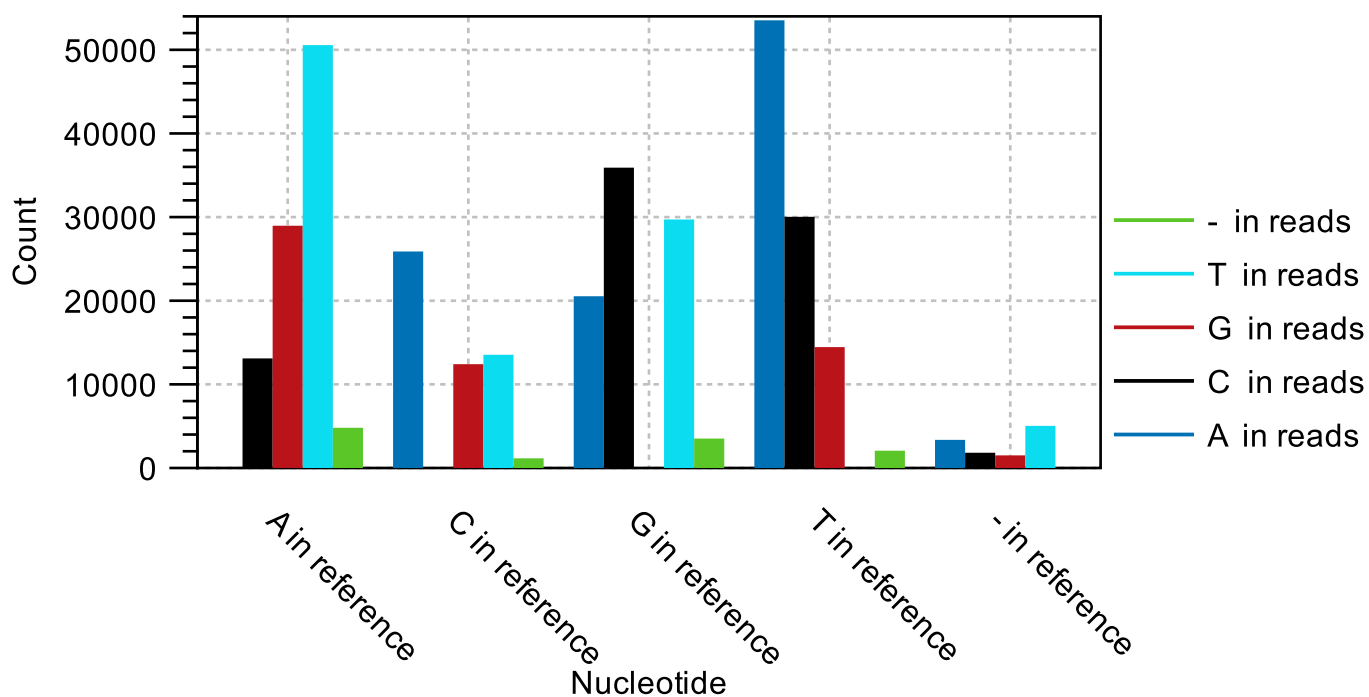
Relative errors (reference to read)



Bars depict off-diagonal element counts divided by counts in the column with header "total".

Error counts (reference to read)

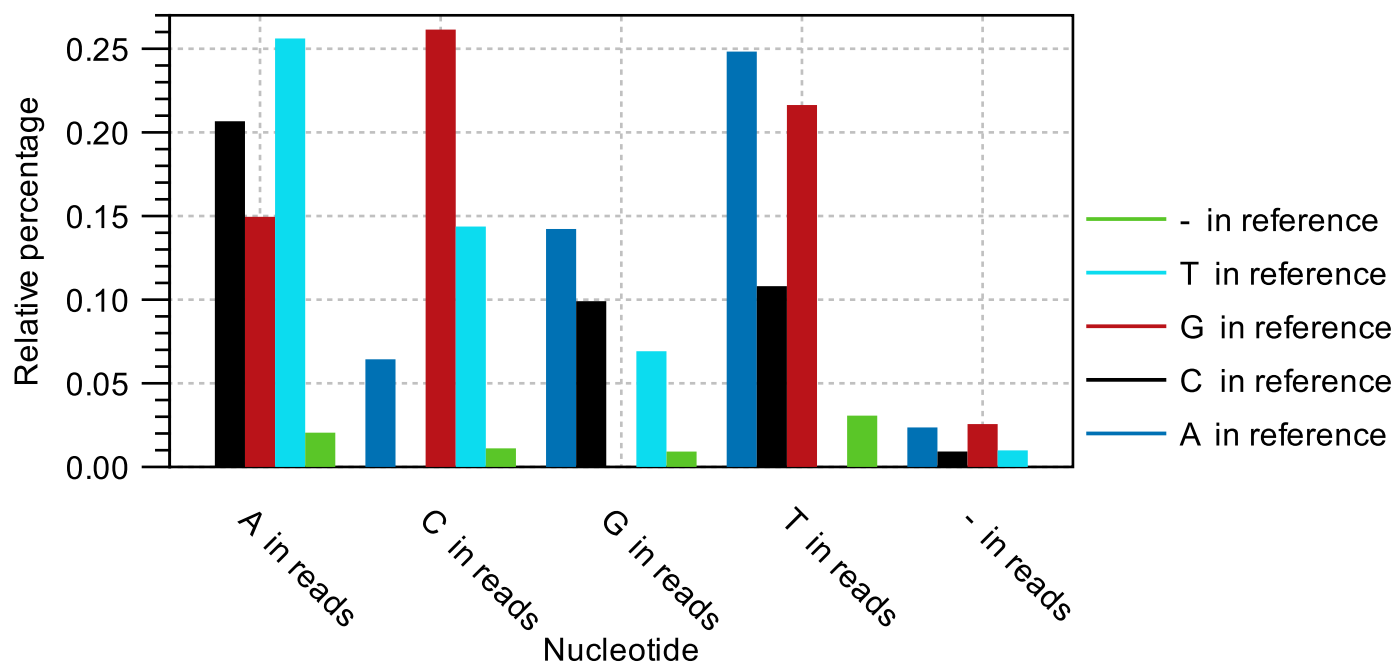
Nucleotide mapping (reference to read)



Bars depict off-diagonal element counts.

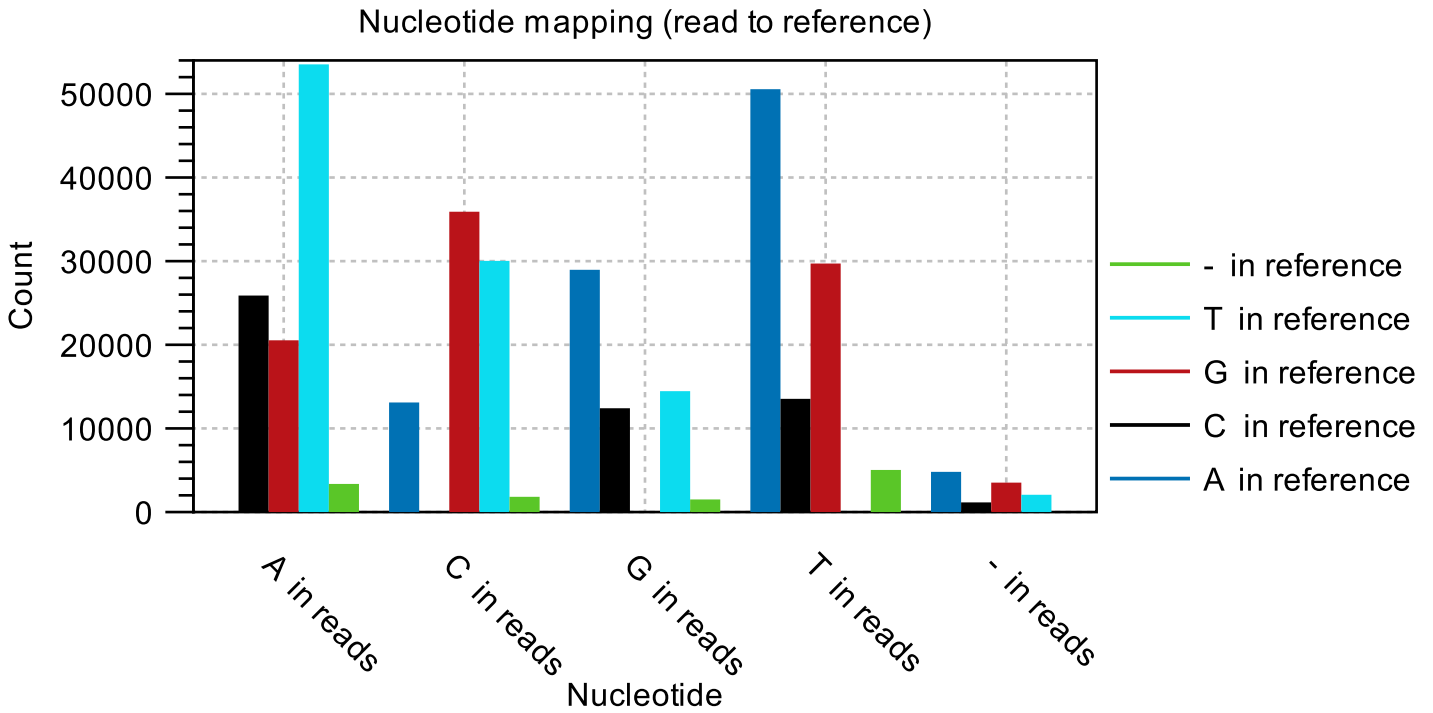
Relative errors (read to reference)

Nucleotide mapping (read to reference)



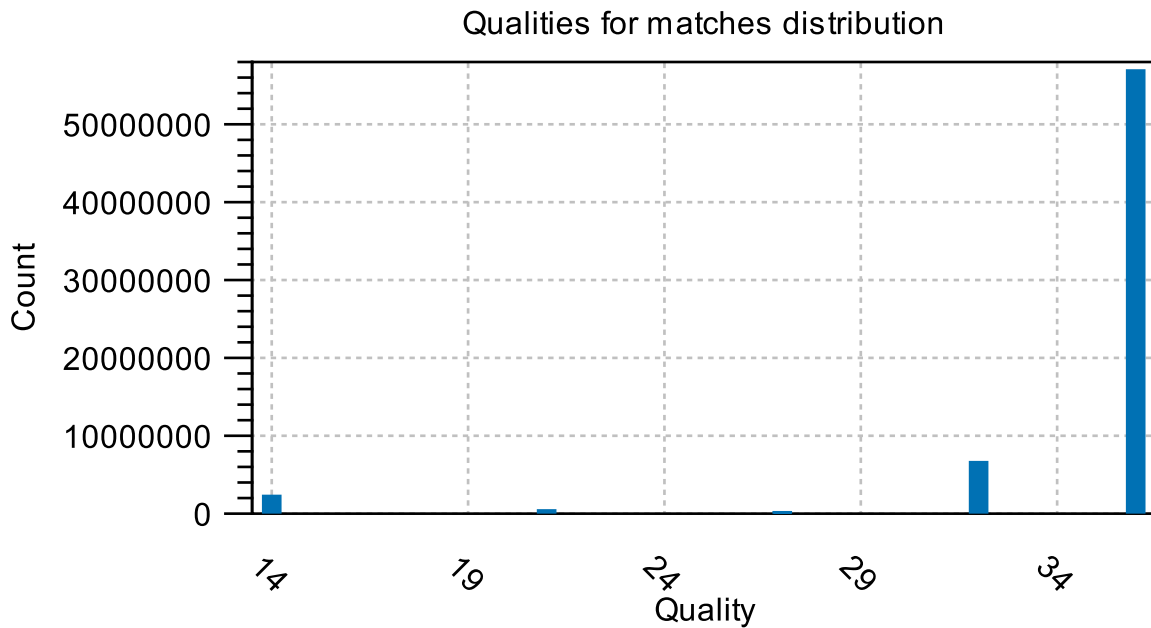
Bars depict off-diagonal element counts divided by counts in the row with header "total".

Error counts (read to reference)

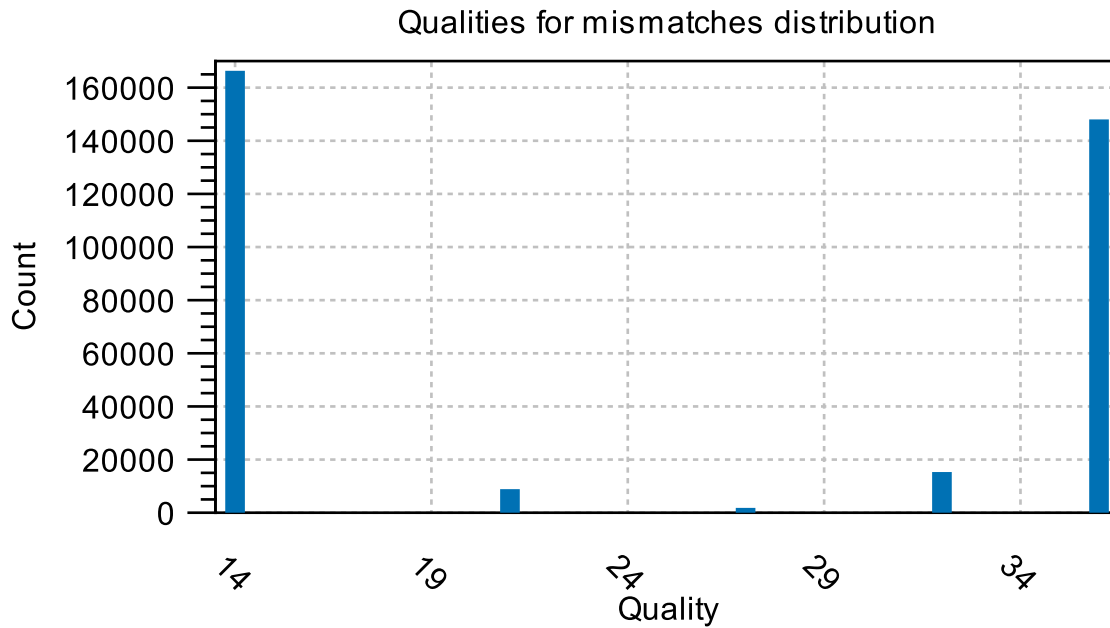


Bars depict off-diagonal element counts.

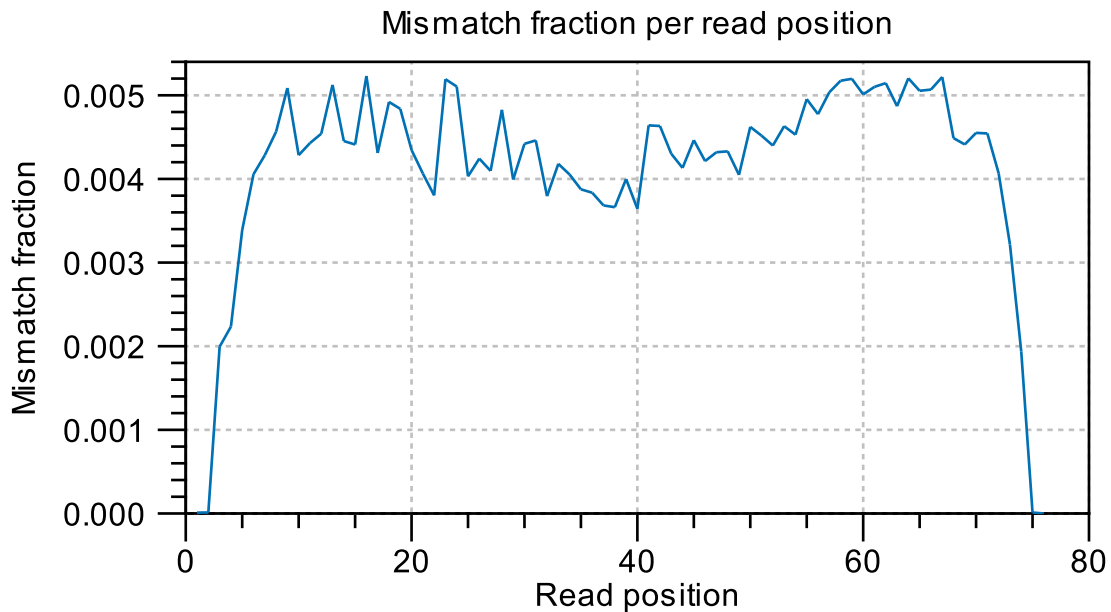
3.10 Quality for match distribution



3.11 Quality for mismatch distribution



3.12 Mismatch fraction per read position



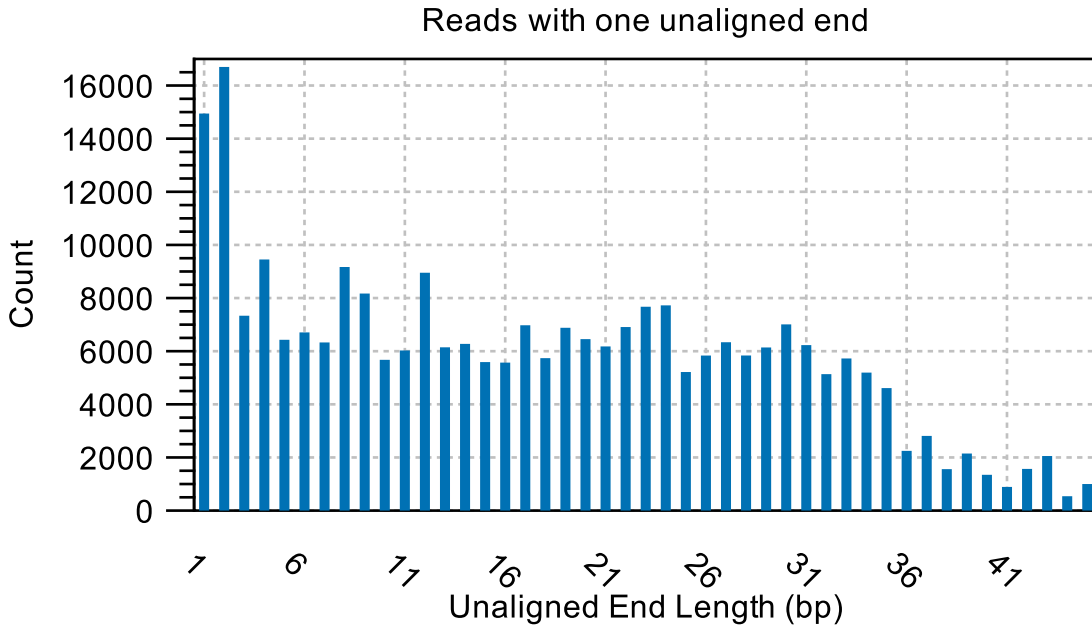
The plot displays the mismatch fraction per read position in the sequenced segments

3.13 Unaligned ends

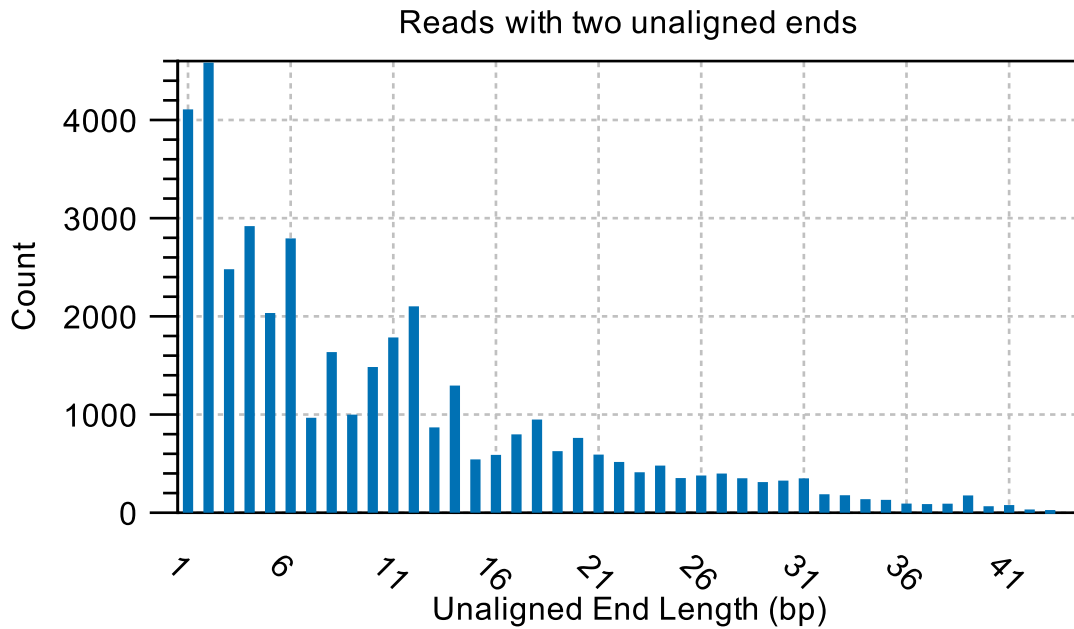
Read count	303,523
------------	---------

% of all mapped reads	31.57
Positions covered	13,487
Positions covered in % of reference	45.10
Positions covered in % of bases covered	45.11

Reads with one unaligned end



Reads with two unaligned ends



History for:

Case A_QC_Mapping

QC for Read Mapping 1.5

15 Aug 2020 10:48:54

Version: CLC Genomics Workbench 20.0.4
Modified by: sevin
Create separate table with statistics for each mapping Yes
resources-key-available-threads 16
Comments:

Local Realignment 2.1

15 Aug 2020 10:48:47

Version: CLC Genomics Workbench 20.0.4
Modified by: sevin
Realign unaligned ends No
Multi-pass realignment 2
Guidance-variant track Case A (InDel, guidance track)
Allow guidance insertion mismatches Yes
Maximum guidance-variant length 200
Force realignment to guidance-variants No
Output mode Create reads track
Output track of realigned regions No
Comments:

Prepare Guidance Variant Track 1.2 [Biomedical Genomics Analysis 20.1.1]

15 Aug 2020 10:48:26

Version: CLC Genomics Workbench 20.0.4
Modified by: sevin
Structural variants Case A (SV)
Reference sequence Not set
Comments:

InDels and Structural Variants 1.9

15 Aug 2020 10:48:26

Version: CLC Genomics Workbench 20.0.4
Modified by: sevin
P-Value threshold 0.0001
Maximum number of mismatches 3
Minimum quality score 0
Minimum relative consensus coverage 0.0
Filter variants No
Ignore broken pairs Yes
Restrict calling to target regions Not set
Create report No
Create breakpoints No
Create InDel variants Yes
Create structural variations Yes

Comments:

Map Reads to Reference 1.7

15 Aug 2020 10:47:40

Version:	CLC Genomics Workbench 20.0.4
Modified by:	sevin
References	MN908947.3 (Genome)
Masking mode	No masking
Match score	1
Mismatch cost	2
Cost of insertions and deletions	Linear gap cost
Insertion cost	3
Deletion cost	3
Length fraction	0.5
Similarity fraction	0.8
Global alignment	No
Auto-detect paired distances	Yes
Non-specific match handling	Map randomly
Output mode	Create reads track
Create report	Yes
Collect unmapped reads	No
Comments:	Reads mapped: 961,490 of 1,471,824 Estimated paired distance range(s): Case A (trimmed pairs): 0 to 549 bp
Originates from:	



MN908947.3 (Genome)

Trim Reads 2.4

15 Aug 2020 10:47:30

Version:	CLC Genomics Workbench 20.0.4
Modified by:	sevin
Quality trim	Yes
Quality limit	0.05
Ambiguous trim	Yes
Ambiguous limit	2
Automatic read-through adapter trimming	Yes
Remove 5' terminal nucleotides	No
Remove 3' terminal nucleotides	No
Fixed length trimming	No
Maximum length	150
Trim from side	3'-end
Discard short reads	Yes
Minimum number of nucleotides in reads	50
Discard long reads	No
Save discarded sequences	No
Save broken pairs	No
Create report	No
Comments:	Processed a total of 1,477,234 sequences. 4,608 sequences have been completely removed during trimming. 195,571 nucleotides have been trimmed, altogether.

Originates from:



Case A