



Mapping Report for Case B Read Mapping  
Type: Read mapping  
Reference count: 1  
Generated by: sevin  
Date: Sat Aug 15 10:50:21 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	951,644
Mean read length	75.45
Total read length	71,804,424

# 2. References

## 2.1 Reference coverage

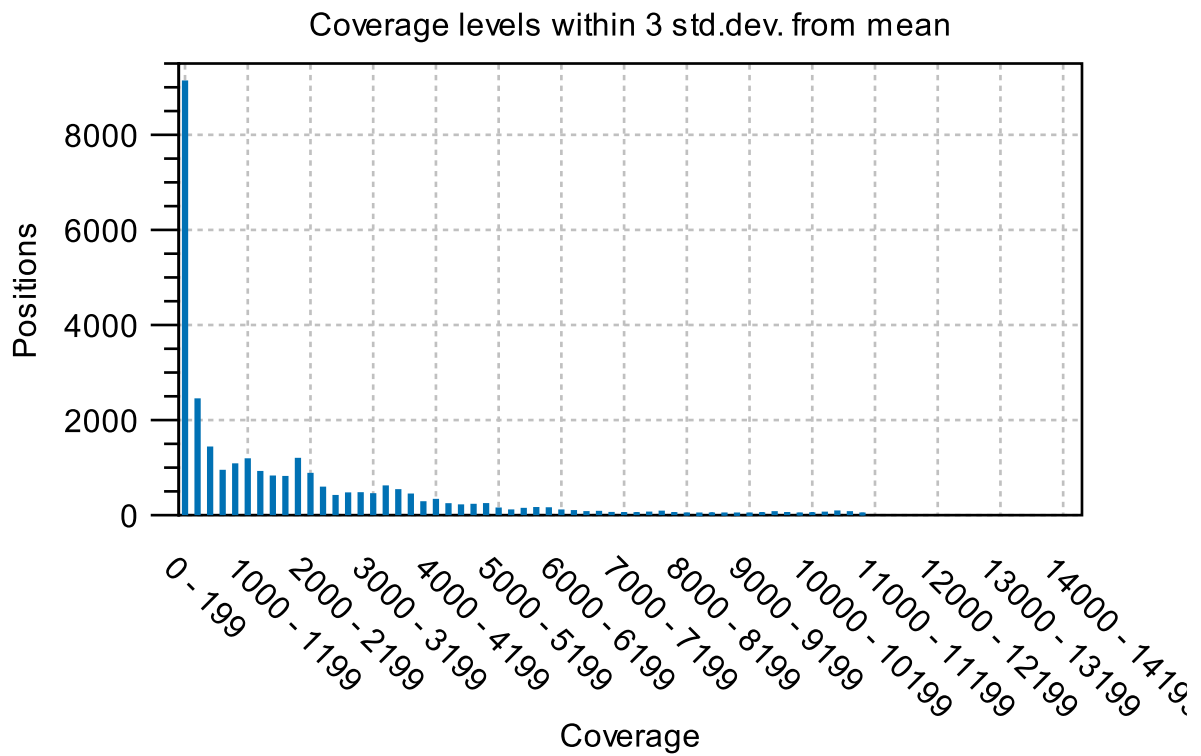
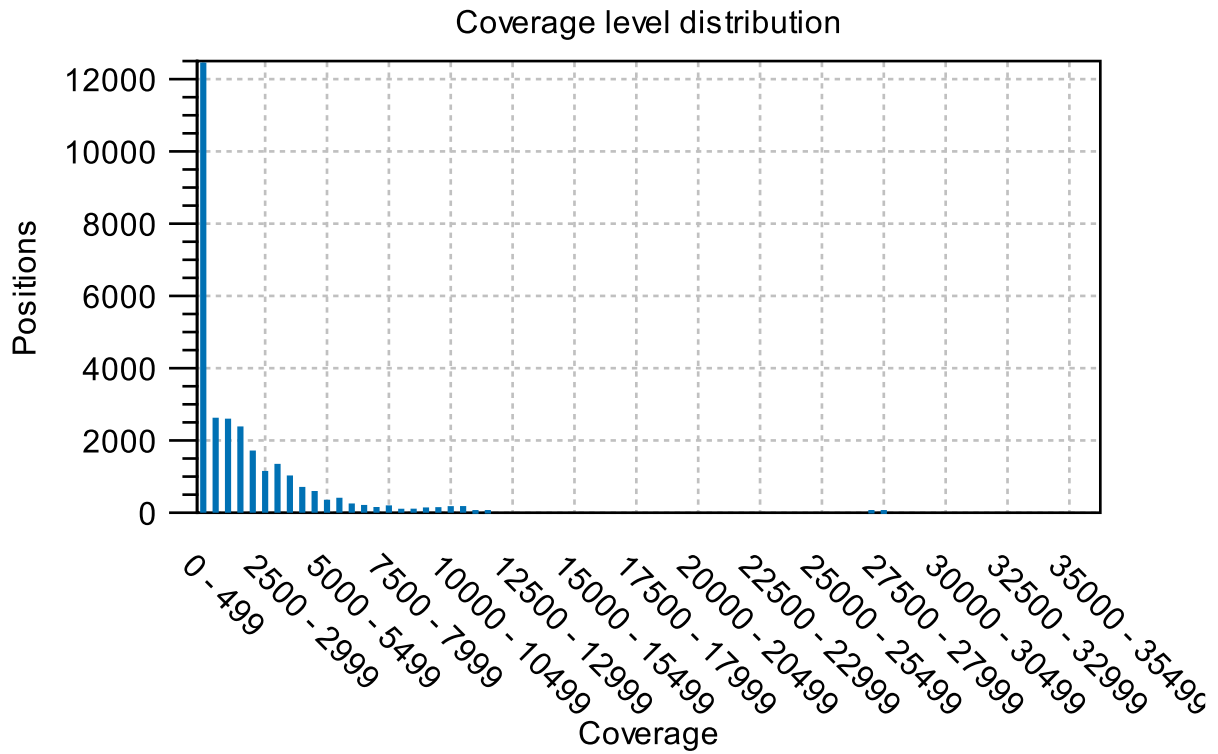
Total reference length	29,903
% GC	37.97
Fraction of reference covered	0.97

## 2.2 Coverage statistics

Total reference length	29,903
Minimum coverage	0
Maximum coverage	36,041
Median coverage	978.00
Average coverage	2,251.02
Standard deviation	4,012.71
Minimum excl. zero coverage regions	1
Median excl. zero coverage regions	1,046.00
Average excl. zero coverage regions	2,329.54
Standard deviation excl. zero coverage regions	4,059.64

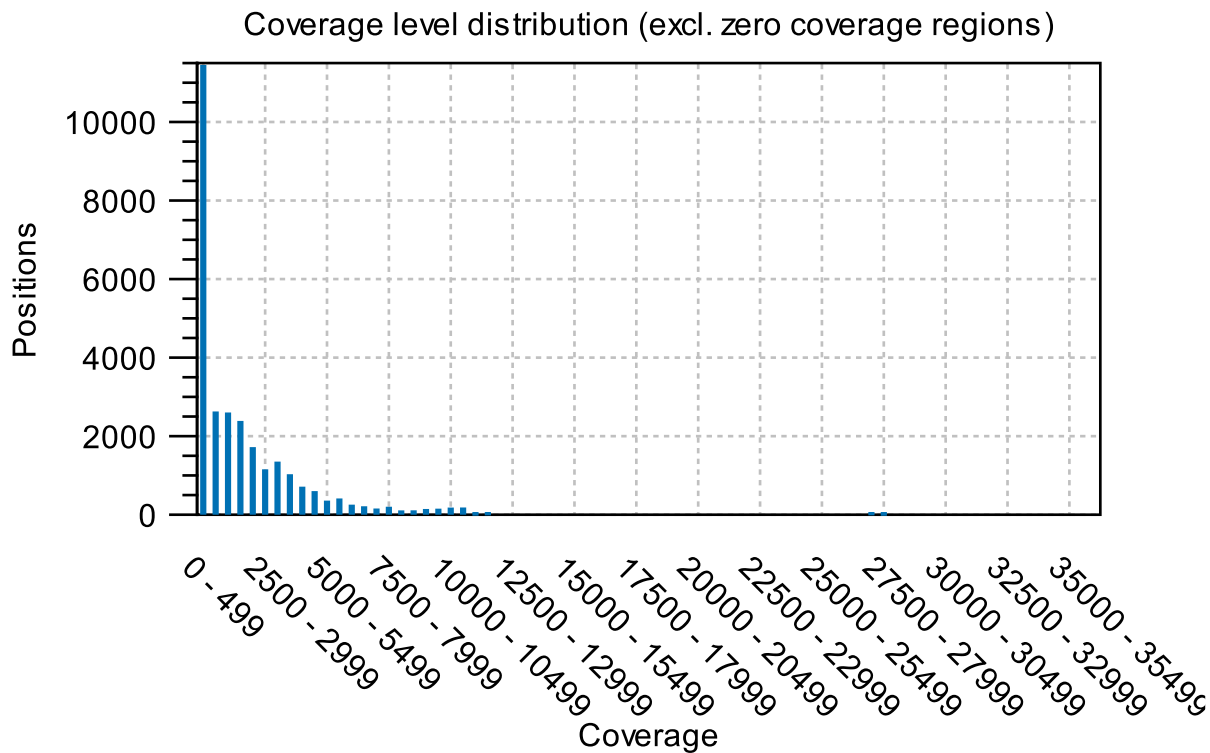
Note that positions with an ambiguous nucleotide in the reference (i.e. not A,C,T or G), count as zero coverage regions, regardless of the number of reads mapping across them.

## 2.3 Coverage level distribution

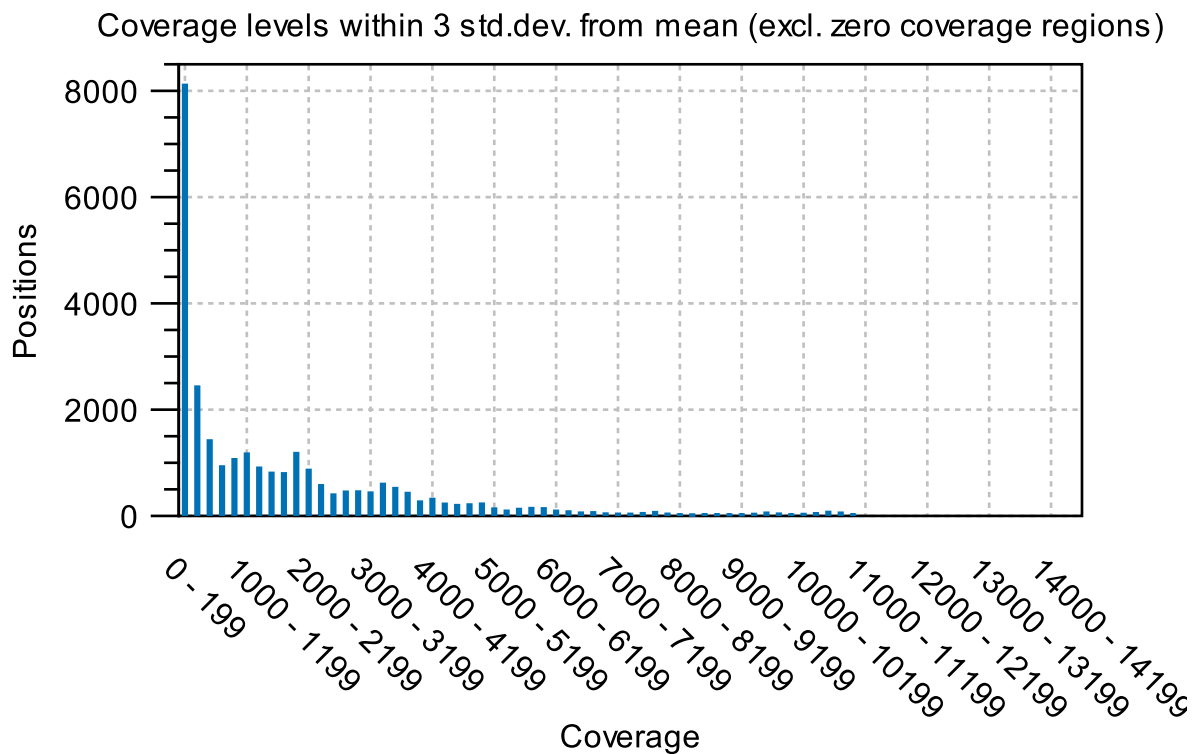


29,279 positions have coverage between 0 and 14,277.  
624 positions have coverage above 14,277 (not shown in graph).

## 2.4 Coverage level distribution (excl. zero coverage regions)



1,008 positions have zero coverage (not shown in graph).



1,008 positions have coverage below 1 (not shown in graph).  
 28,280 positions have coverage between 1 and 14,506.  
 615 positions have coverage above 14,506 (not shown in graph).

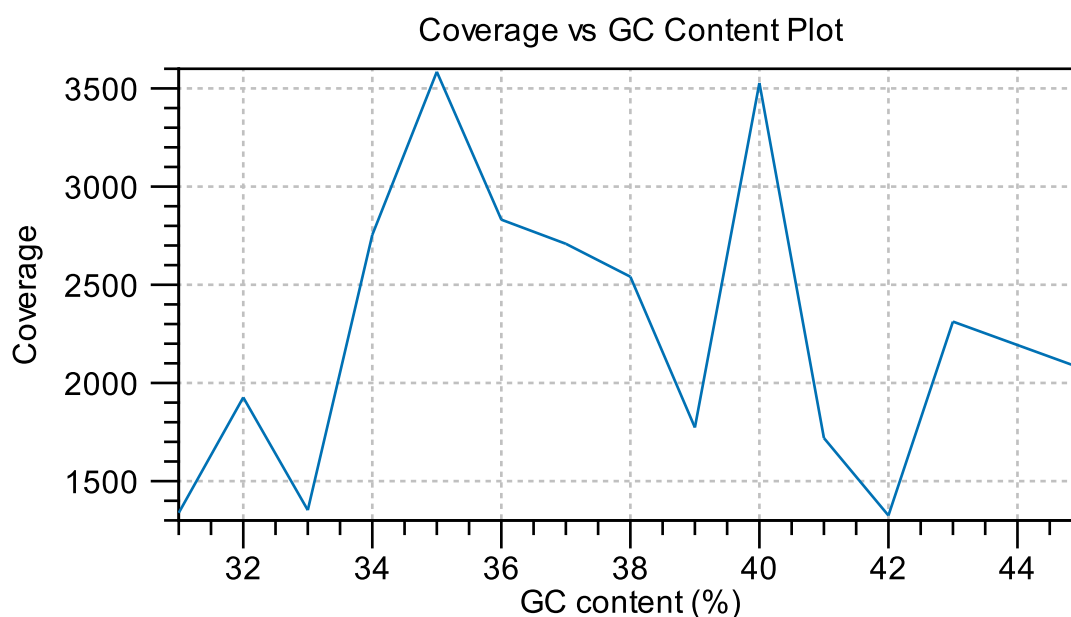
Note that positions with an ambiguous nucleotide in the reference (i.e. not A,C,T or G), count as zero coverage regions, regardless of the number of reads mapping across them.

## 2.5 Zero coverage regions

Count	23
Minimum length	1
Maximum length	122
Mean length	43.83
Standard deviation	34.04
Total length	1,008

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

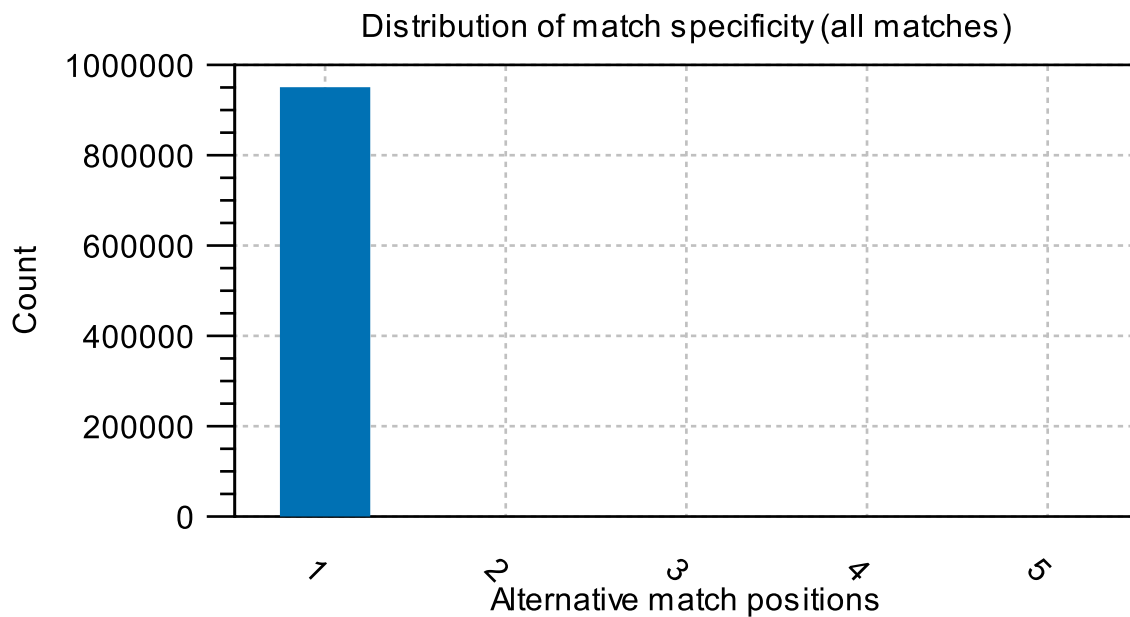
Read count	951,644
Mean read length	75.45
Total read length	71,804,424

### 3.2 Non-specific matches

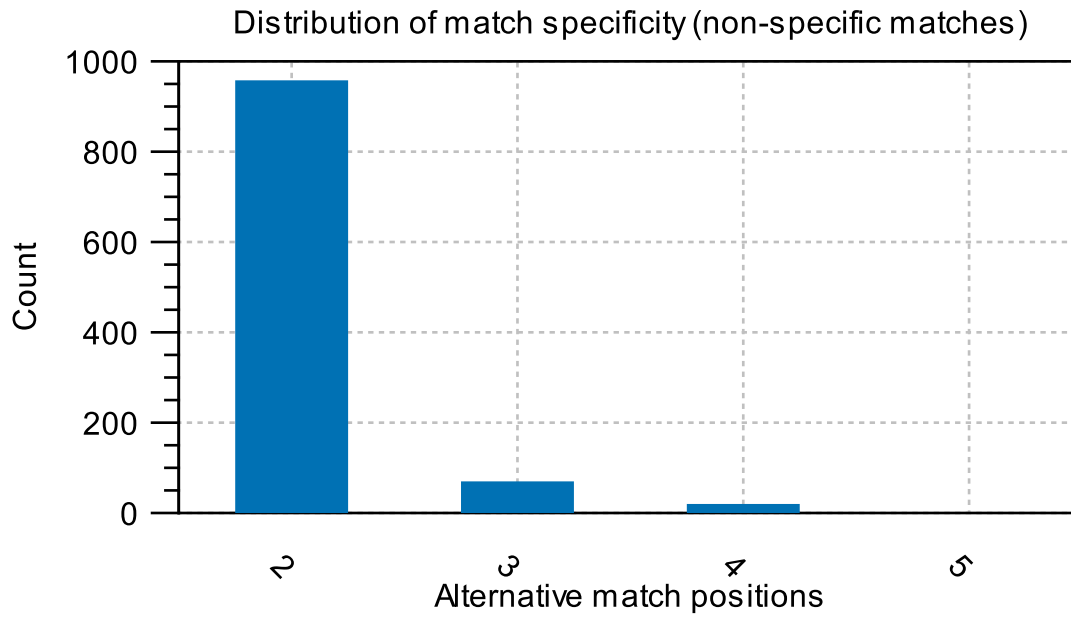
#### *Non-specific matches*

Read count	1,049
% of all mapped reads	0.11
Mean read length	75.46
Total read length	79,162

#### *Distribution of match specificity (all matches)*



#### *Distribution of match specificity (non-specific matches)*

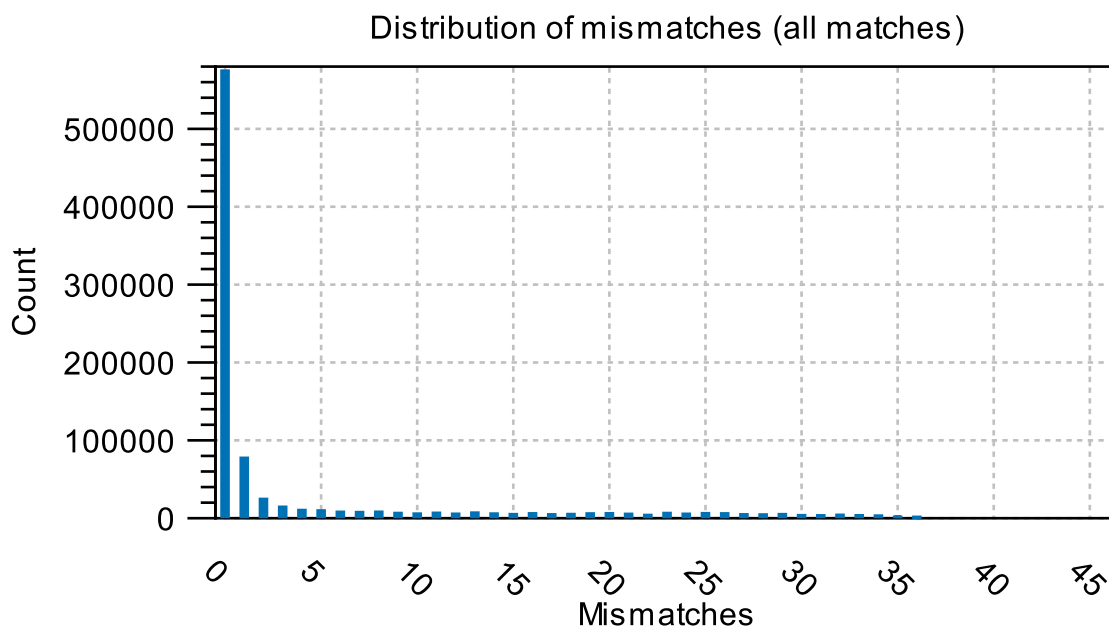


### 3.3 Non-perfect matches

#### *Non-perfect matches*

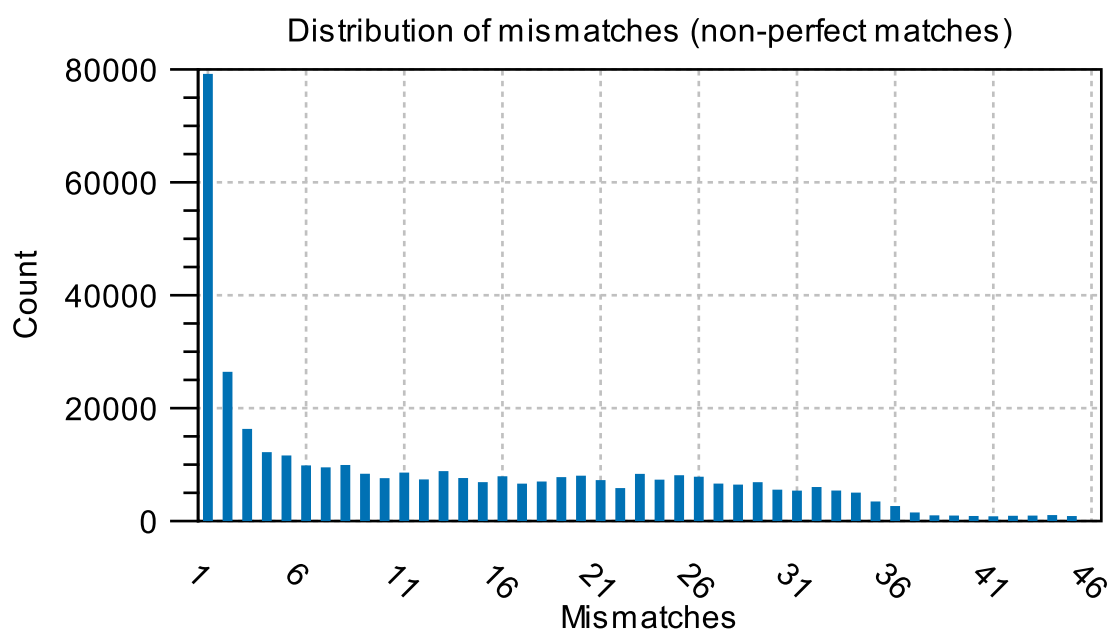
Read count	375,213
% of all mapped reads	39.43
Mean read length	75.44
Total read length	28,307,147

#### *Distribution of mismatches (all matches)*



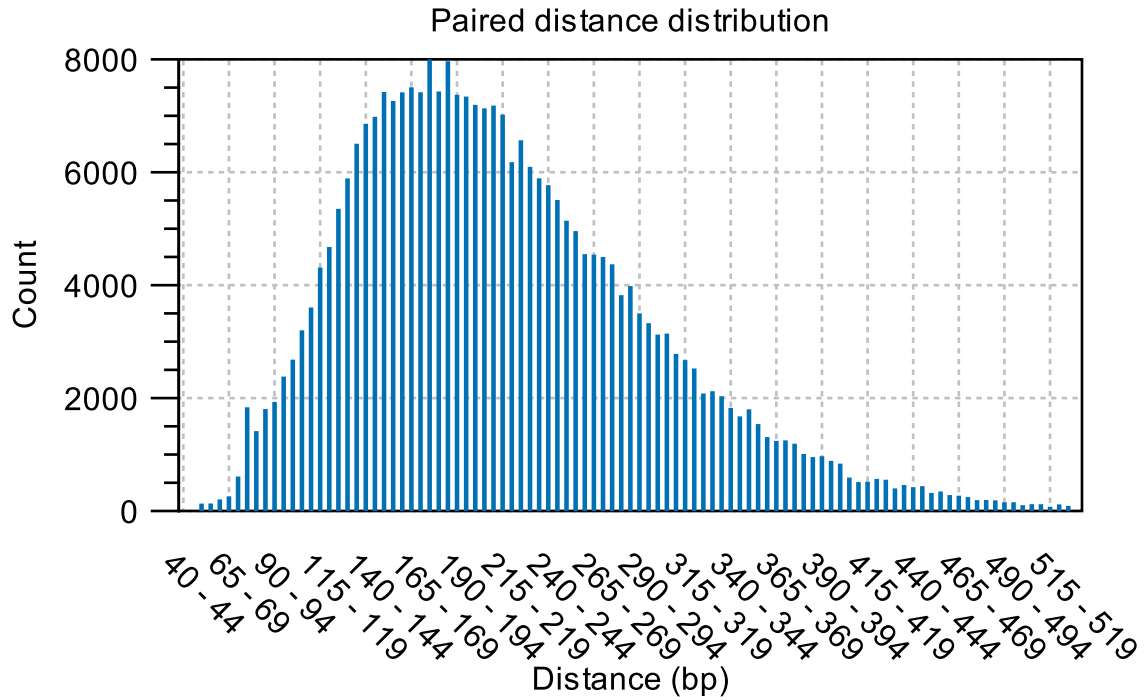


## Distribution of mismatches (non-perfect matches)



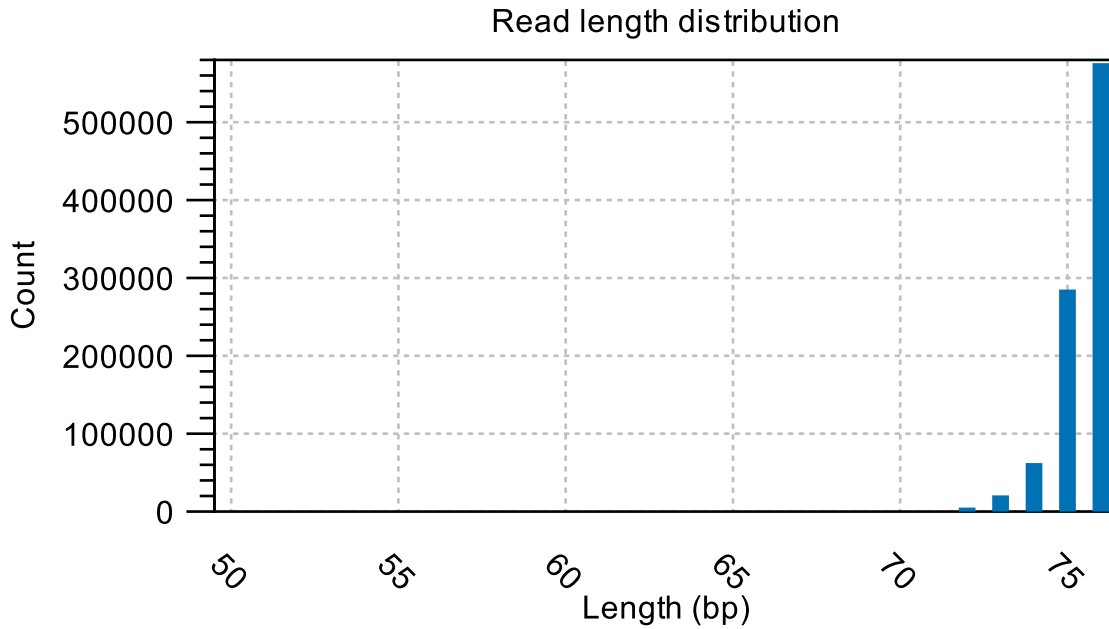
### 3.4 Paired reads

Reads in aligned pairs	575,180
Reads in broken pairs: wrong distance or mate inverted	322,204
Reads in broken pairs: mate on other contig	0
Reads in broken pairs: mate not mapped	54,260
Mean distance	217.07
Standard deviation	80.99

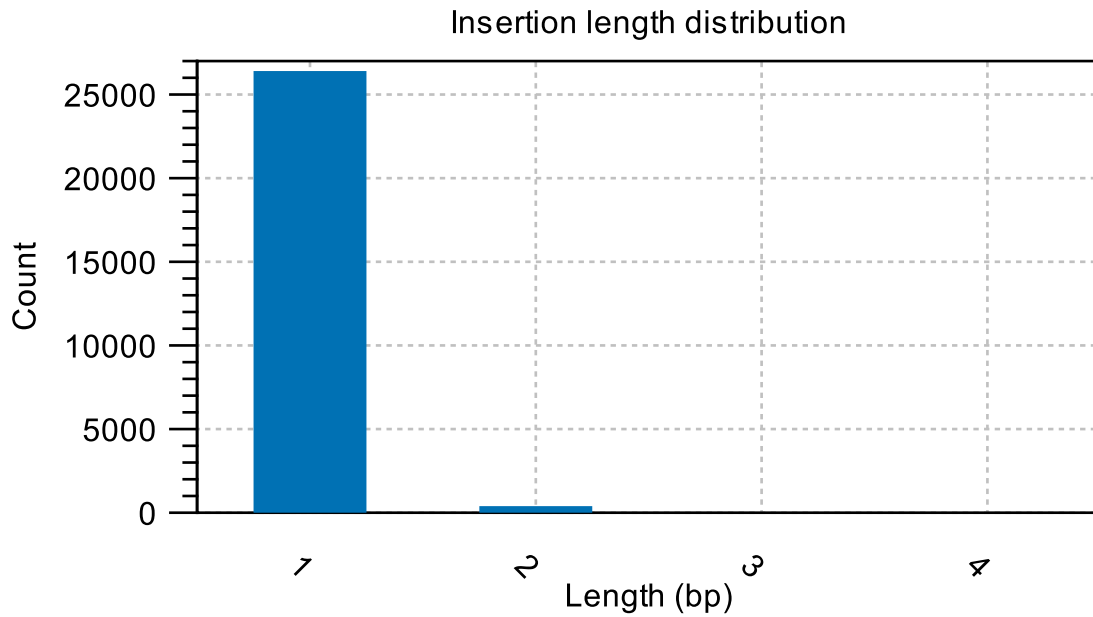


Paired distance for intact pairs (including unaligned ends)

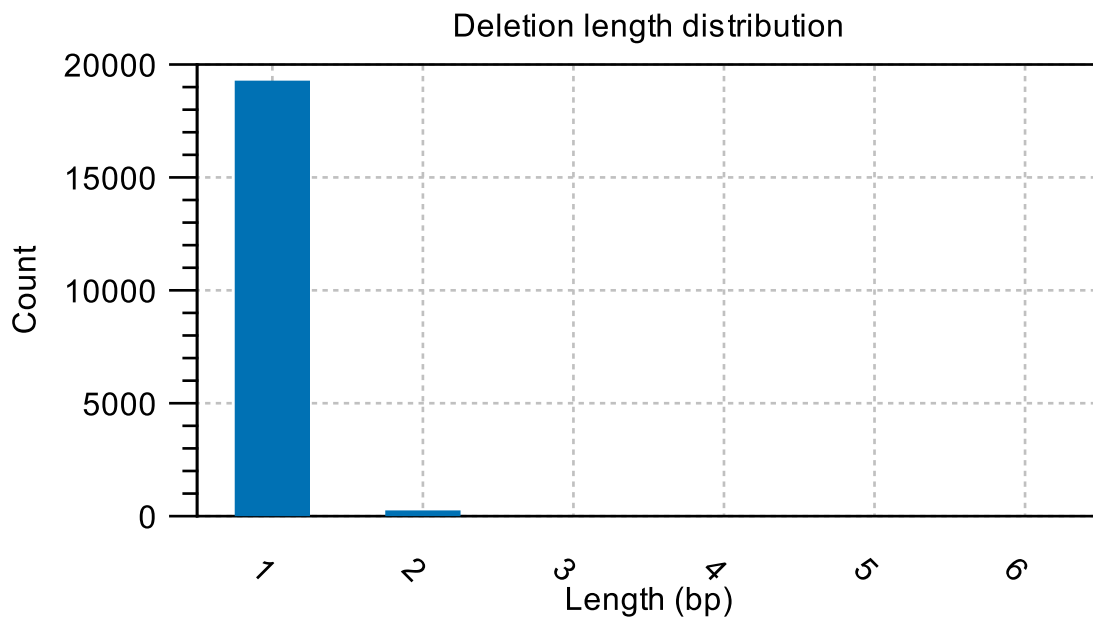
### 3.5 Read length distribution



### 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	0.72
C	0.63

Nucleotide in reference	% read bases that differ
G	0.54
T	0.47
-	0.29
Total	0.55

### 3.9 Nucleotide Mapping

#### Counts

	Read: A	Read: C	Read: G	Read: T
Reference: A	19,840,603	22,928	45,866	67,268
Reference: C	27,184	12,495,633	15,921	32,126
Reference: G	22,557	15,436	12,797,480	29,058
Reference: T	56,800	19,596	21,779	21,781,855
Reference: -	12,455	4,557	2,703	7,633
Total	19,959,599	12,558,150	12,883,749	21,917,940

Read: -	Total
7,881	19,984,546
3,769	12,574,633
2,805	12,867,336
5,486	21,885,516
9,460,523	9,487,871
9,480,464	76,799,902

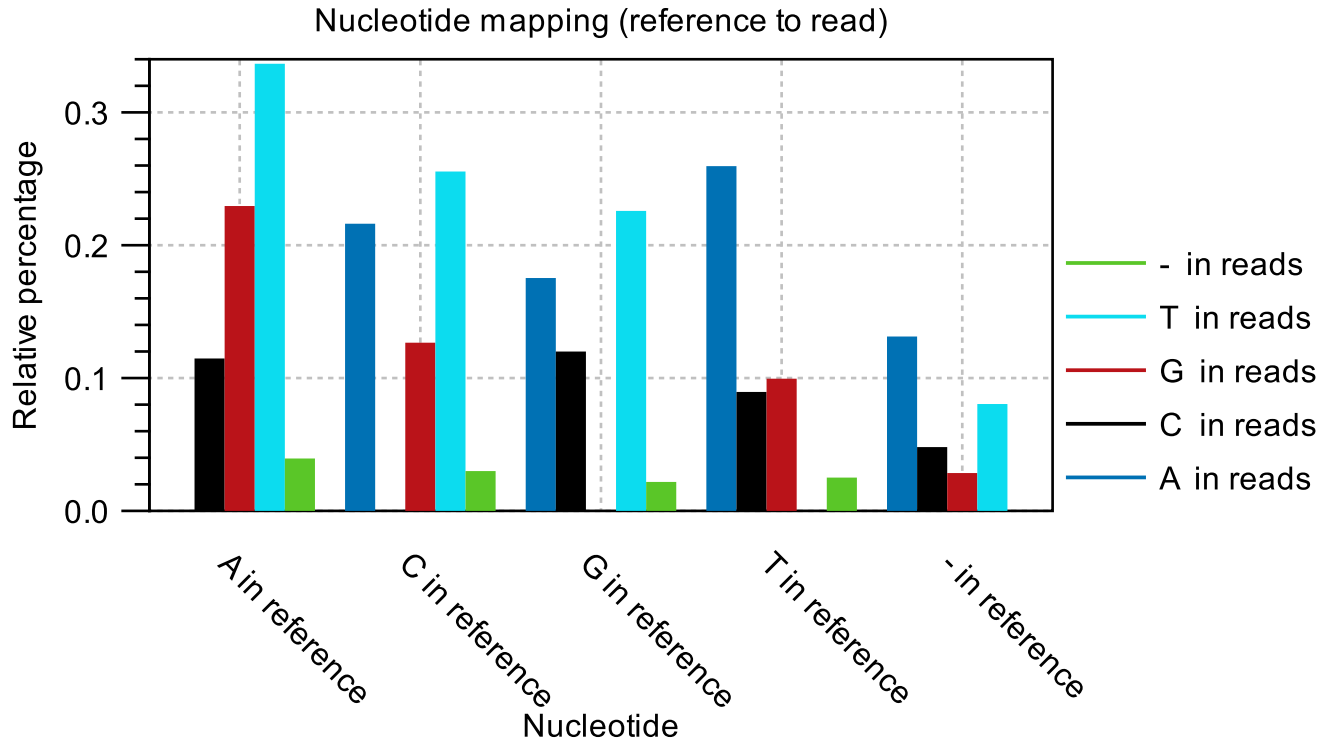
#### Percentages

	Read: A	Read: C	Read: G	Read: T
Reference: A	25.83	0.03	0.06	0.09
Reference: C	0.04	16.27	0.02	0.04
Reference: G	0.03	0.02	16.66	0.04
Reference: T	0.07	0.03	0.03	28.36
Reference: -	0.02	0.01	0.00	0.01
Total	25.99	16.35	16.78	28.54

Read: -	Total
0.01	26.02
0.00	16.37
0.00	16.75
0.01	28.50
12.32	12.35

Read: -	Total
12.34	100.00

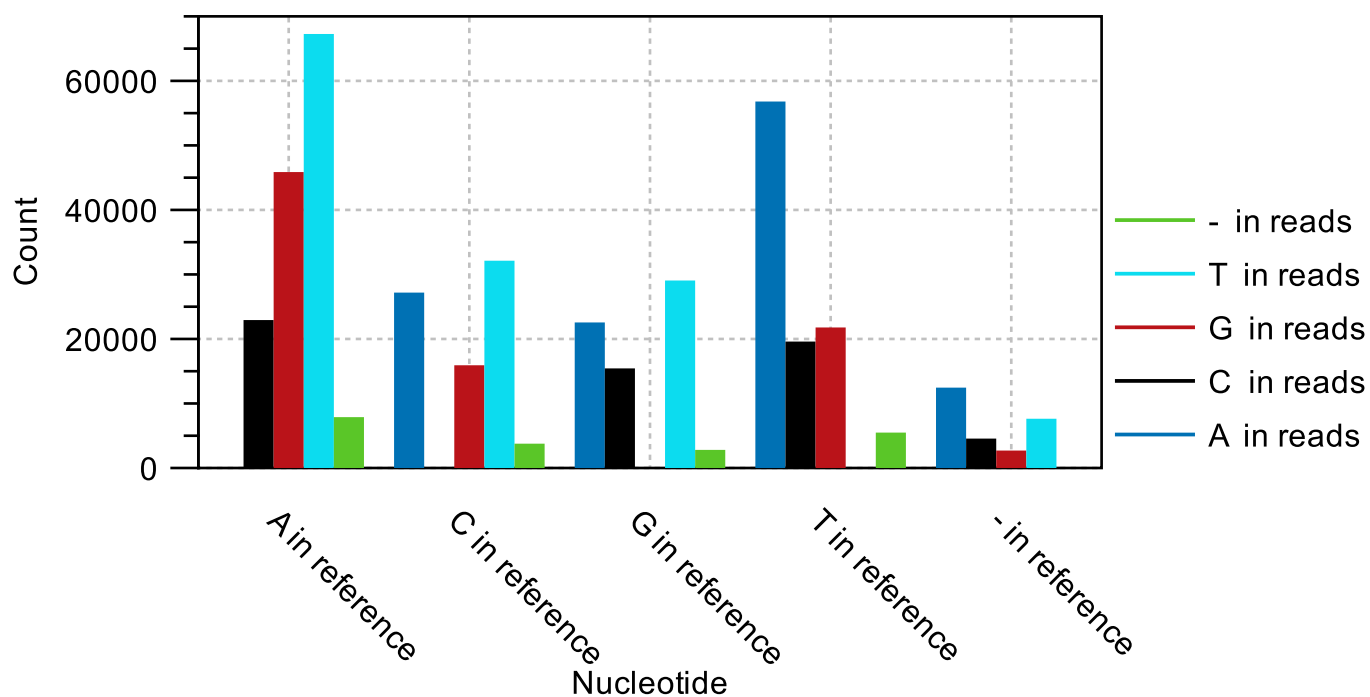
*Relative errors (reference to read)*



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

*Error counts (reference to read)*

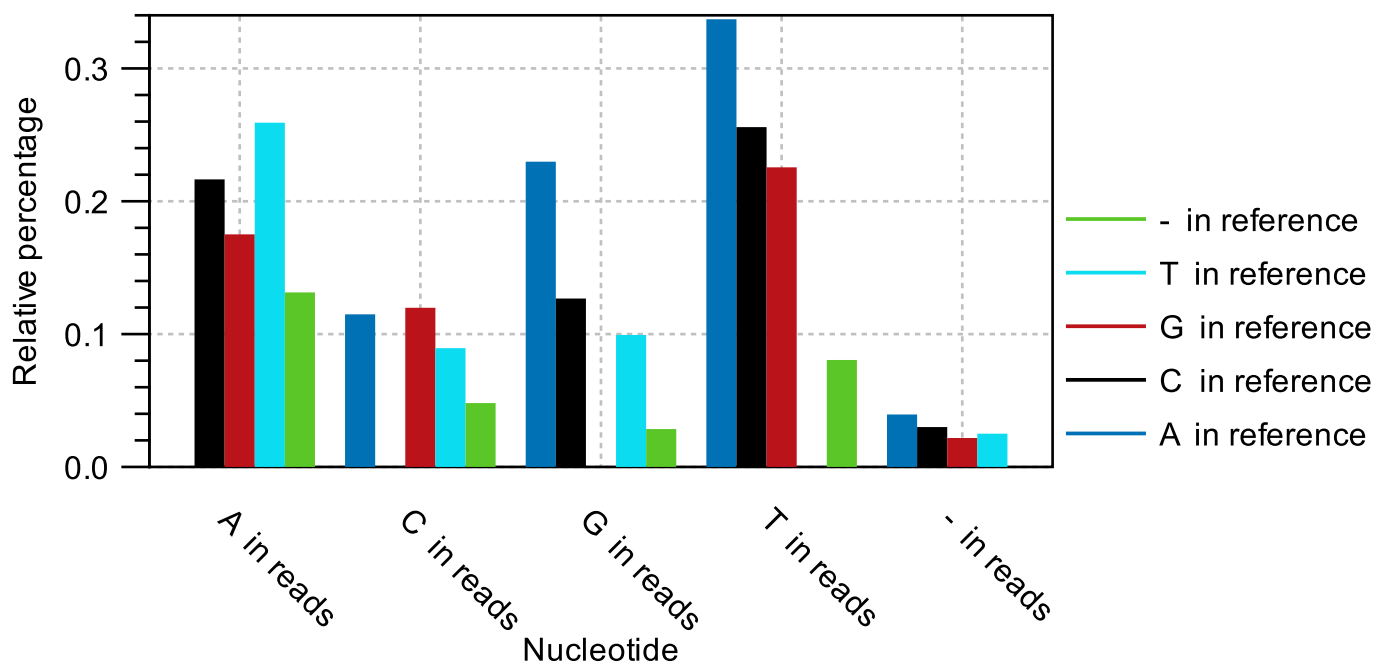
Nucleotide mapping (reference to read)



Bars depict off-diagonal element counts.

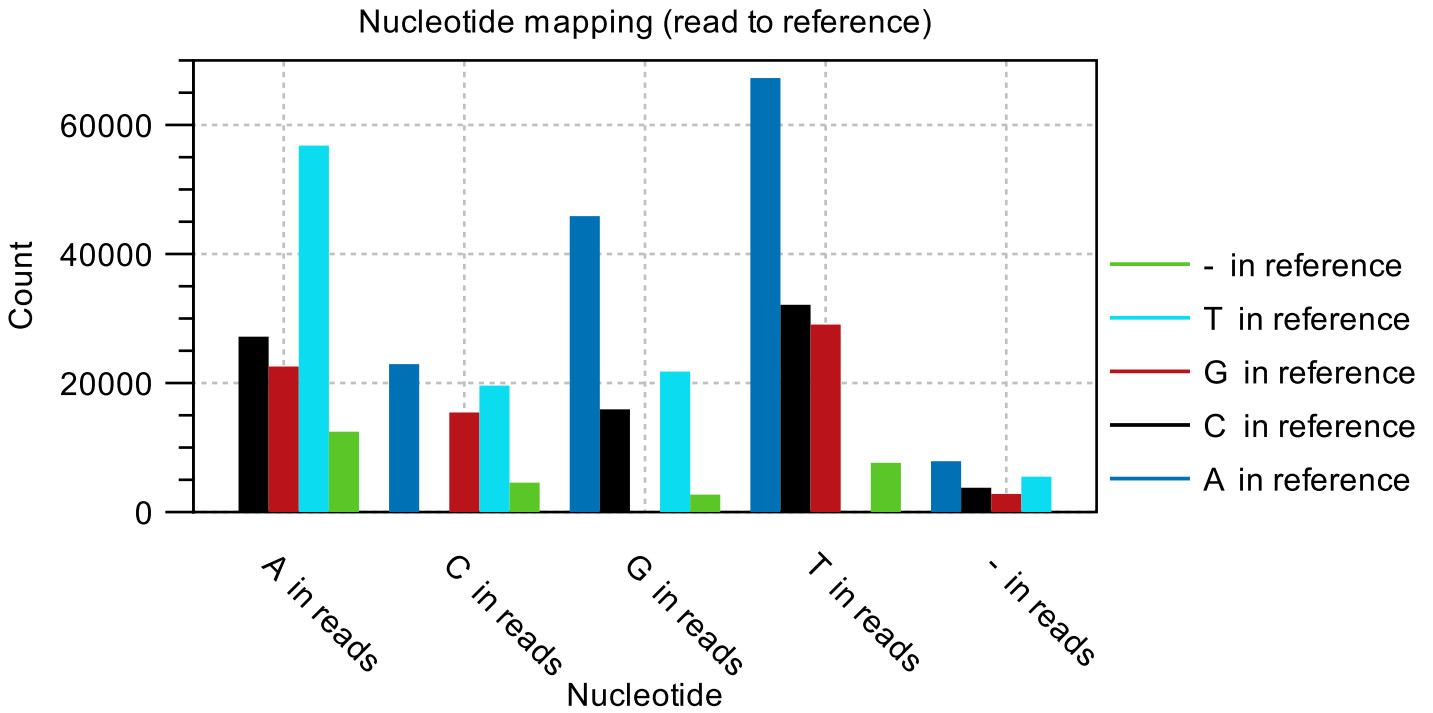
*Relative errors (read to reference)*

Nucleotide mapping (read to reference)



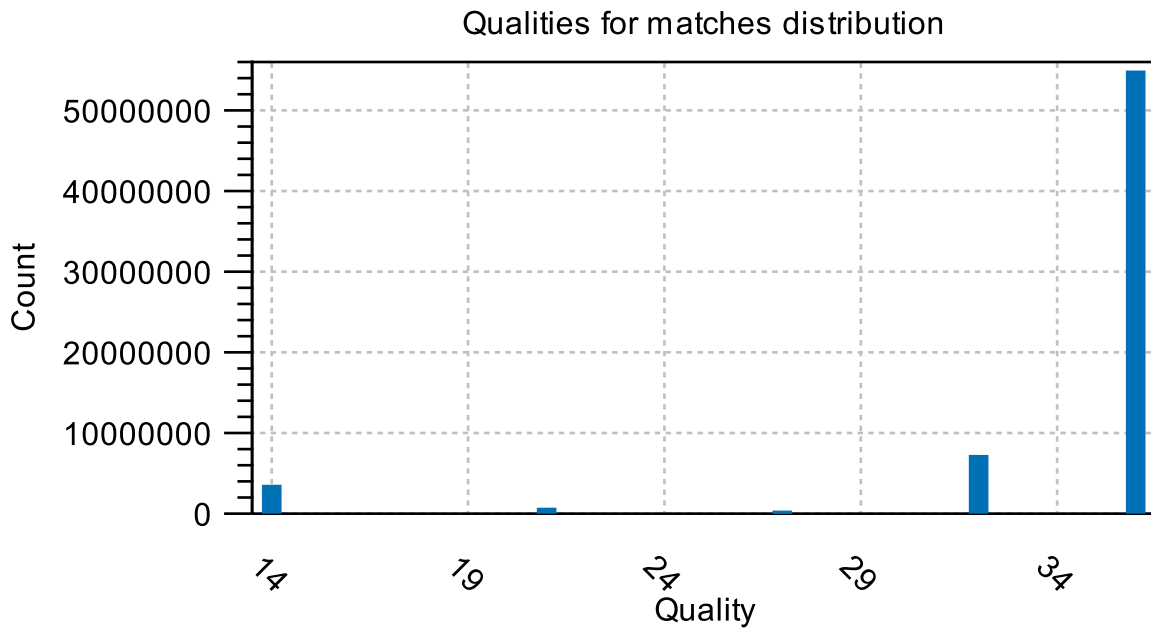
Bars depict off-diagonal element counts divided by counts in the row with header "total".

## Error counts (read to reference)

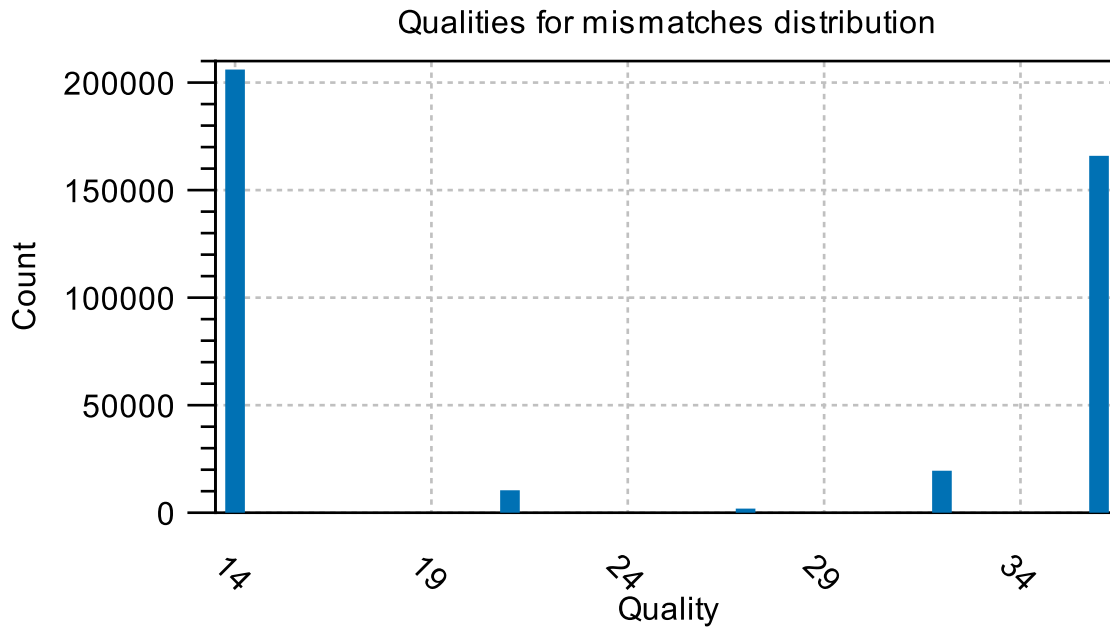


Bars depict off-diagonal element counts.

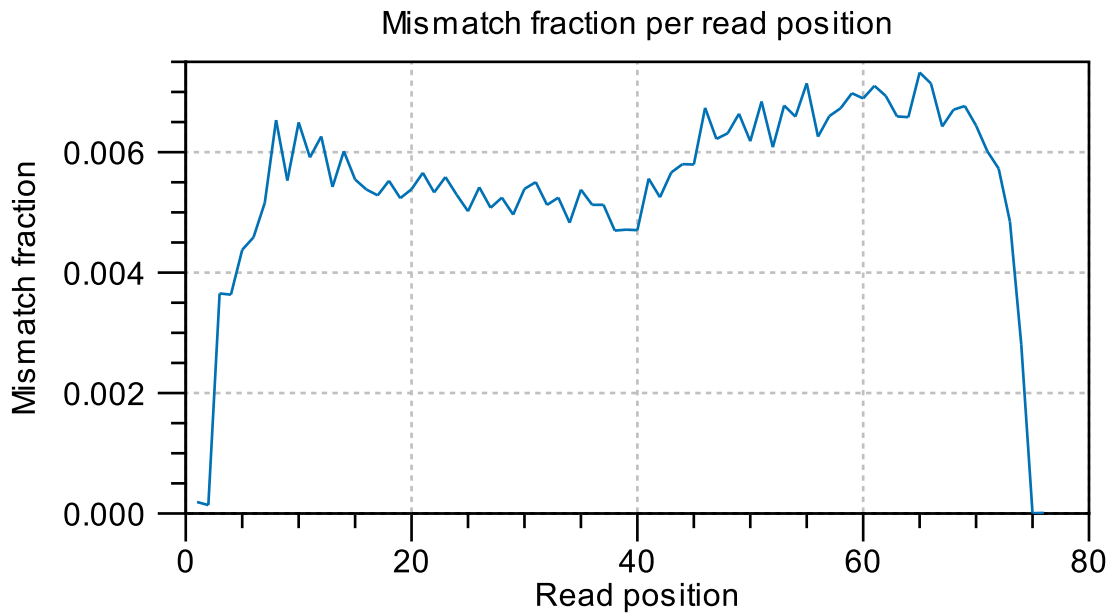
## 3.10 Quality for match distribution



### 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

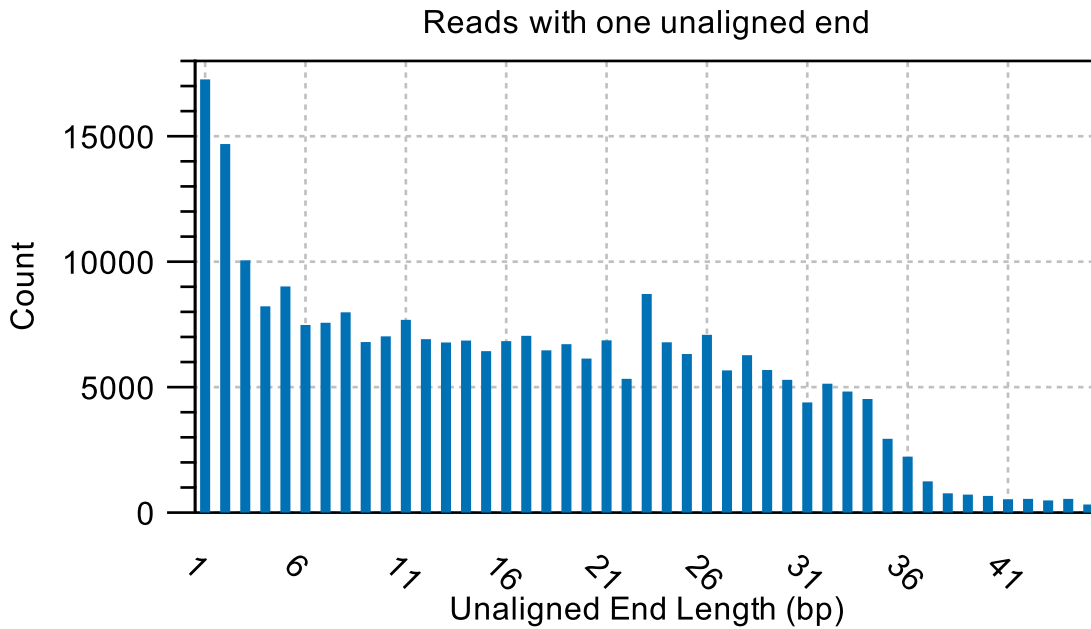
### 3.13 Unaligned ends

Read count	289,673
------------	---------

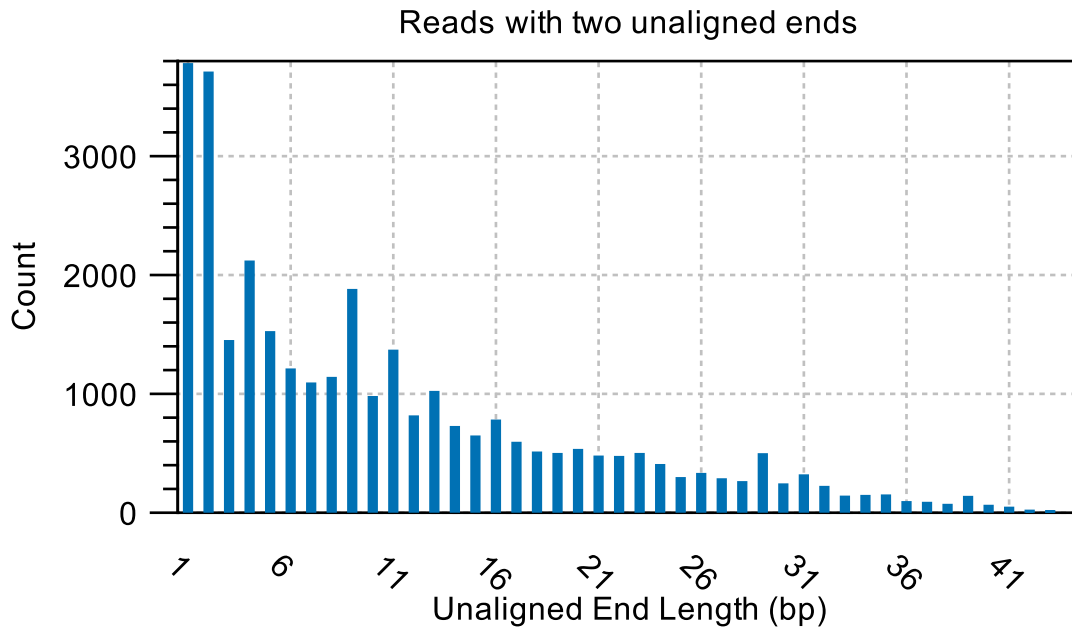


% of all mapped reads	30.44
Positions covered	12,487
Positions covered in % of reference	41.76
Positions covered in % of bases covered	43.22

*Reads with one unaligned end*



*Reads with two unaligned ends*



## History for:

### Case B\_QC\_Mapping

#### QC for Read Mapping 1.5

15 Aug 2020 10:50:21

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 10:50:15

Version: CLC Genomics Workbench 20.0.4  
Modified by: sevin  
Realign unaligned ends No  
Multi-pass realignment 2  
Guidance-variant track Case B (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 10:49:58

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

#### InDels and Structural Variants 1.9

15 Aug 2020 10:49:58

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 10:49:17

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: 951,644 of 2,796,090 Estimated paired distance range(s): Case B (trimmed pairs): 0 to 530 bp
Originates from:	



MN908947.3 (Genome)

### Trim Reads 2.4

15 Aug 2020 10:49:02

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 2,821,770 sequences. 22,490 sequences have been completely removed during trimming. 906,683 nucleotides have been trimmed, altogether.

Originates from:

	Case B
---	--------