

Bioinformatic Analysis Report

Genome Project Solutions
1024 Promenade Street
Hercules, CA 94547

Informatics@GenomeProjectSolutions.com
<http://www.GenomeProjectSolutions.com>
877-867-0146

Project title: Chlorella Genome Sequence

Prepared for: Dr. Micheal Guarnieri

Type of report: RNAseq mapping

Date: May 25, 2012

Details: Detailed report on mapping all RNAseq raw reads to the transcript contigs of the transcriptome assembly

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	7
3. Mapped reads	7
3.1 All mapped reads	7
3.2 Non-specific matches	7
3.3 Non-perfect matches	10
3.4 Read length distribution	12

1. Summary

Reference count	27,303
Type	Reference mapping
Total reference length	20,780,790
GC contents in %	62.39
Total read count	345,520,531
Mean read length	35.56
Total read length	12,285,158,508

2. References

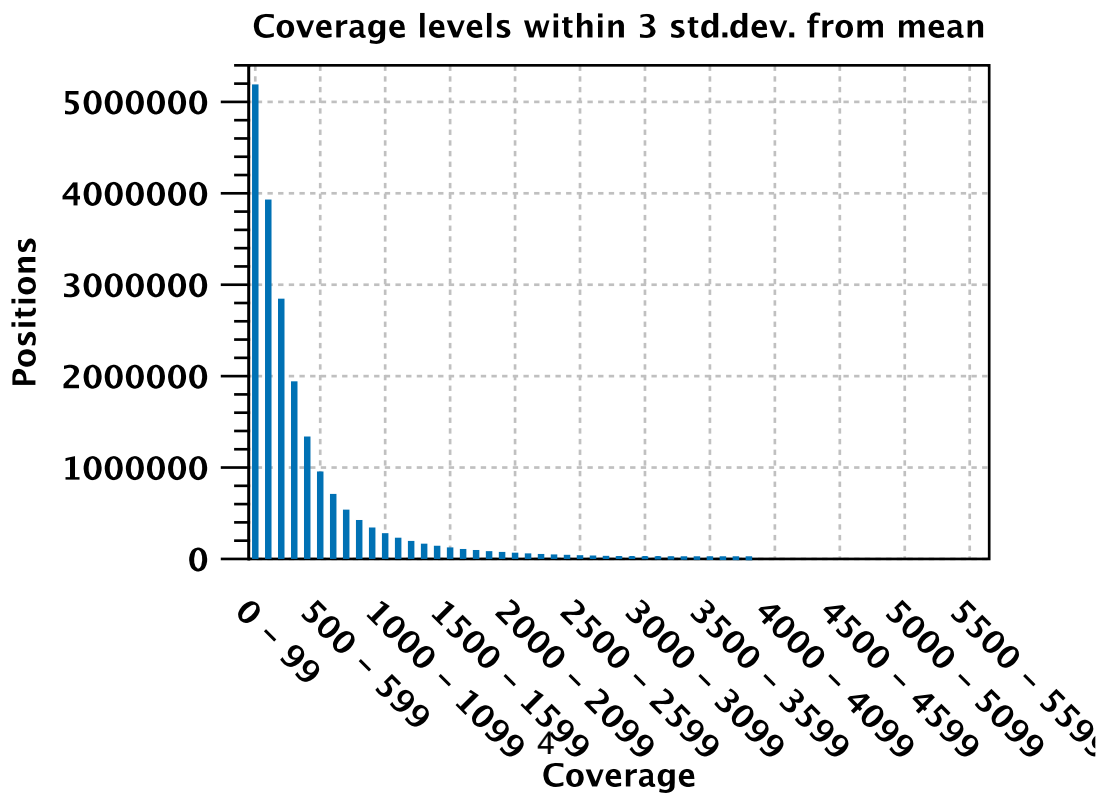
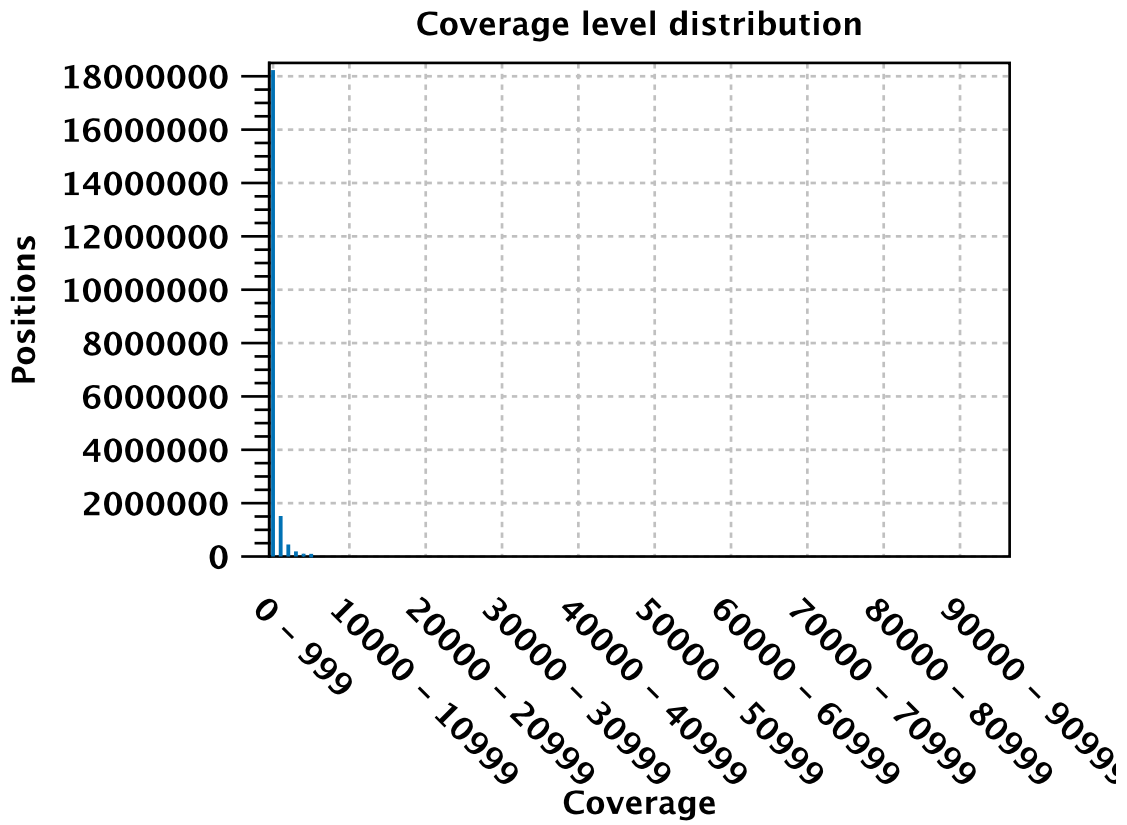
2.1 Reference coverage

Total reference length	20,780,790
% GC	62.39
Total consensus length	20,779,827
Fraction of reference covered	1.00

2.2 Coverage statistics

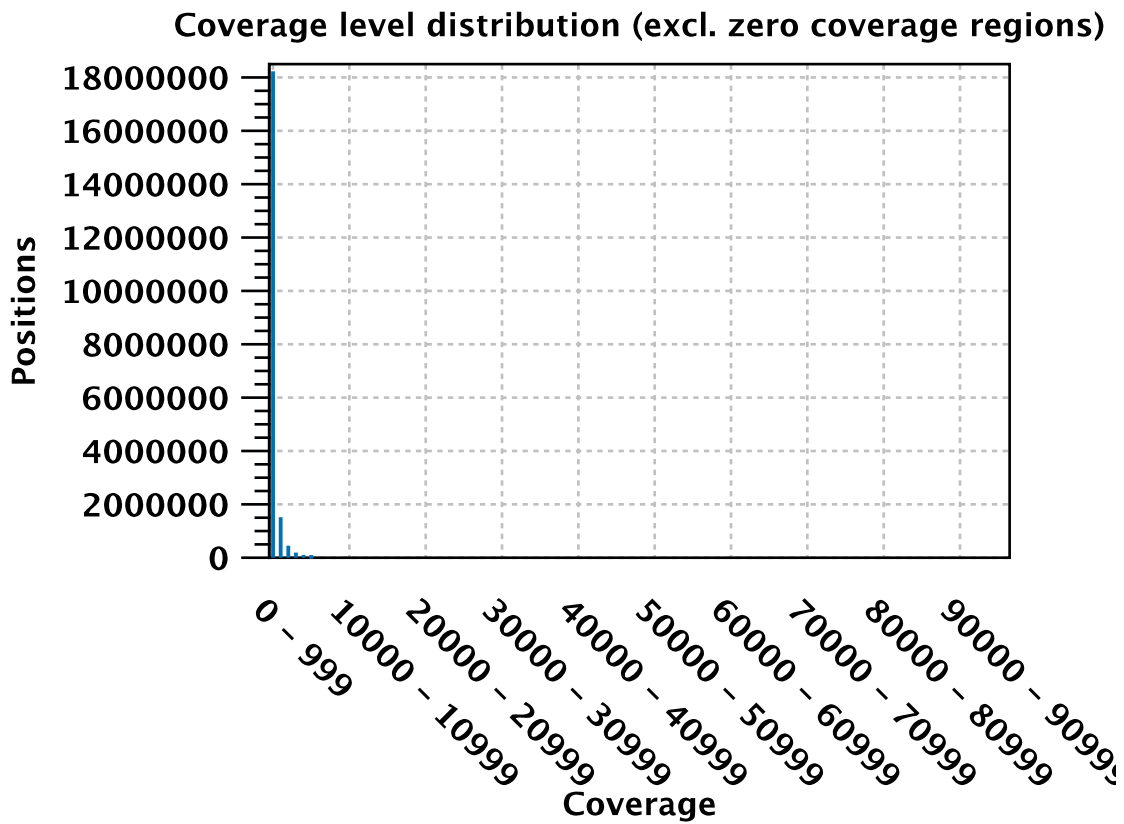
Total reference length	20,780,790
Minimum coverage	0
Maximum coverage	96,800
Average coverage	587.67
Standard deviation	1,691.23
Minimum excl. zero coverage regions	1
Average excl. zero coverage regions	587.70
Standard deviation excl. zero coverage regions	1,691.27

2.3 Coverage level distribution



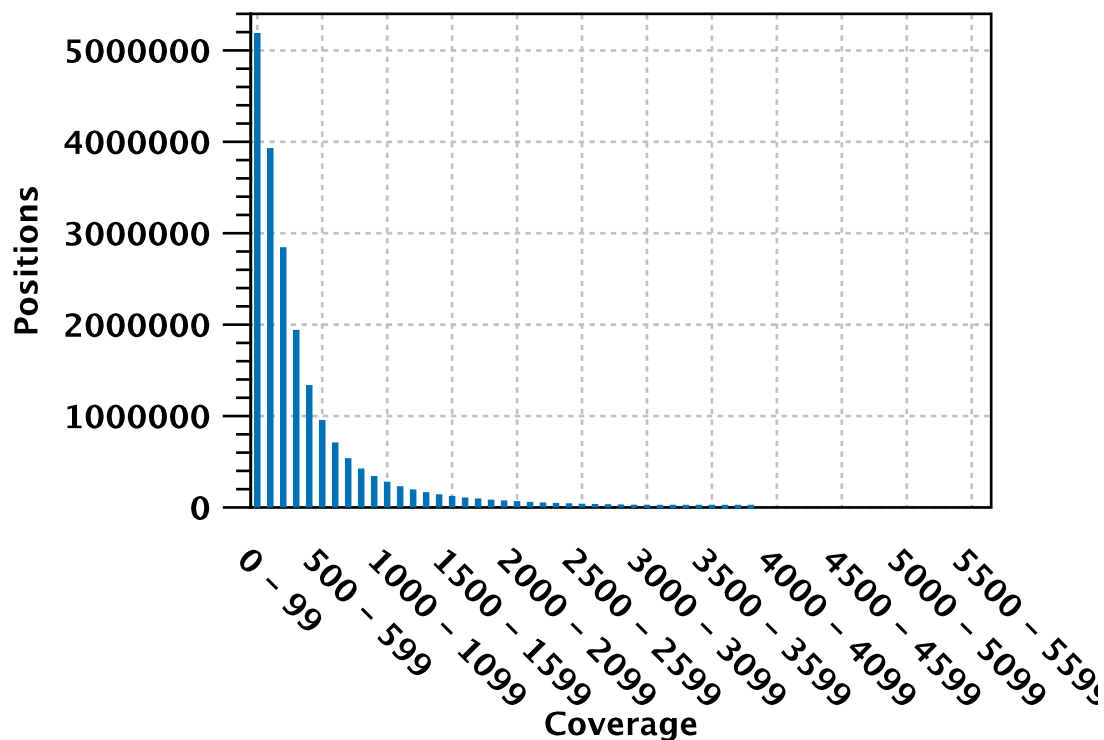
20,542,916 positions have coverage between 0 and 5,661.
237,874 positions have coverage above 5,661 (not shown in graph).

2.4 Coverage level distribution (excl. zero coverage regions)



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

Coverage levels within 3 std.dev. from mean (excl. zero coverage)

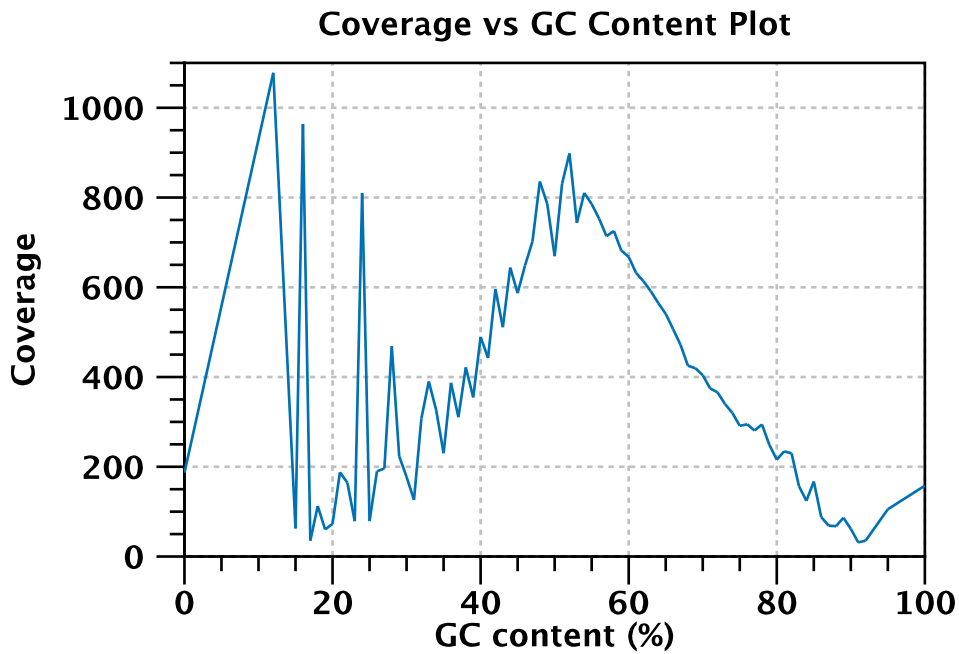


963 positions have coverage below 1 (not shown in graph).
 20,541,953 positions have coverage between 1 and 5,661.
 237,874 positions have coverage above 5,661 (not shown in graph).

2.5 Zero coverage regions

Count	454
Minimum length	1
Maximum length	22
Mean length	2.12
Standard deviation	1.89
Total length	963

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

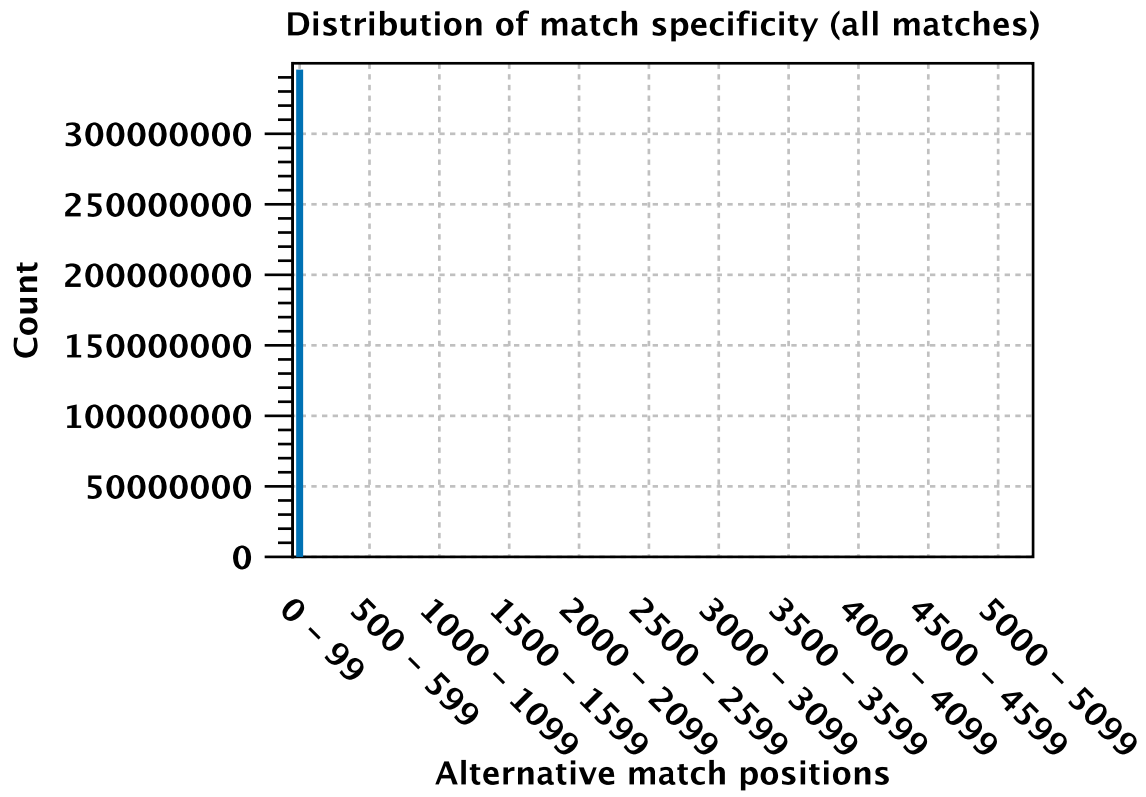
Read count	345,520,531
Mean read length	35.56
Total read length	12,285,158,508

3.2 Non-specific matches

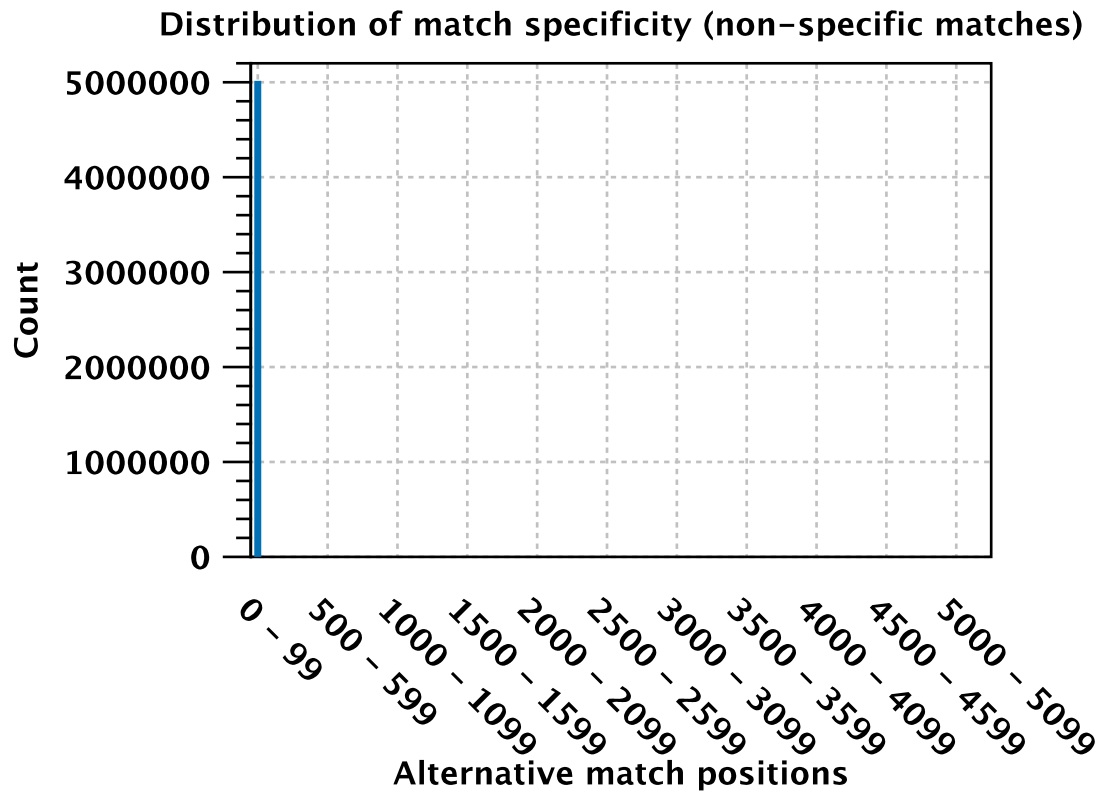
Non-specific matches

Read count	5,027,863
Mean read length	34.42
Total read length	173,054,958

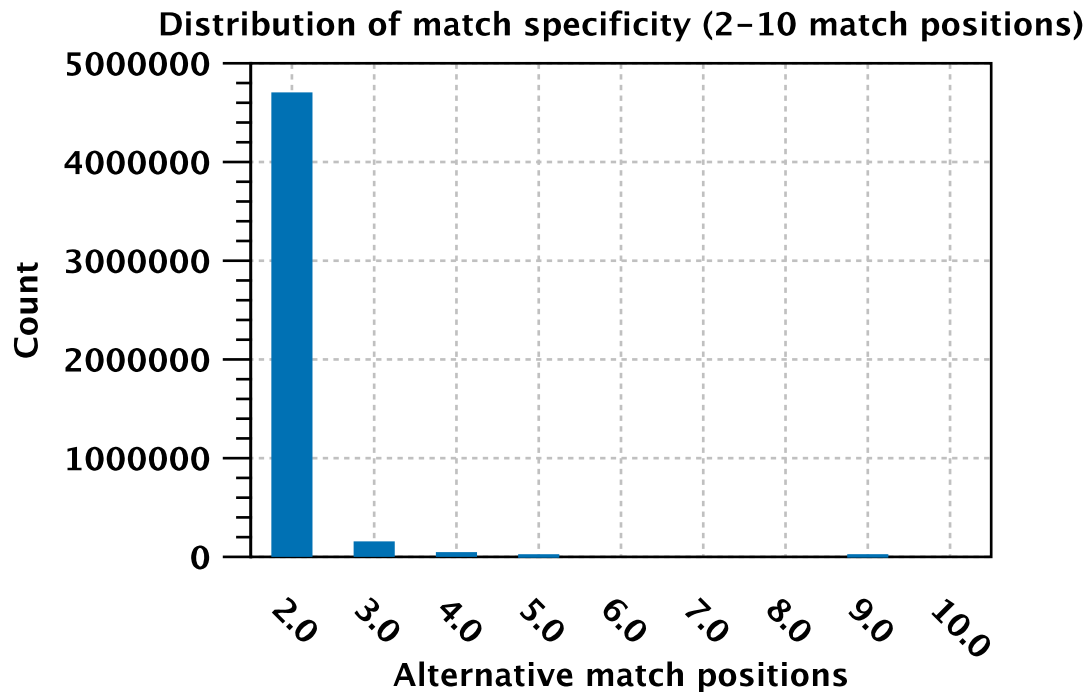
Distribution of match specificity (all matches)



Distribution of match specificity (non-specific matches)



Distribution of match specificity (2-10 match positions)

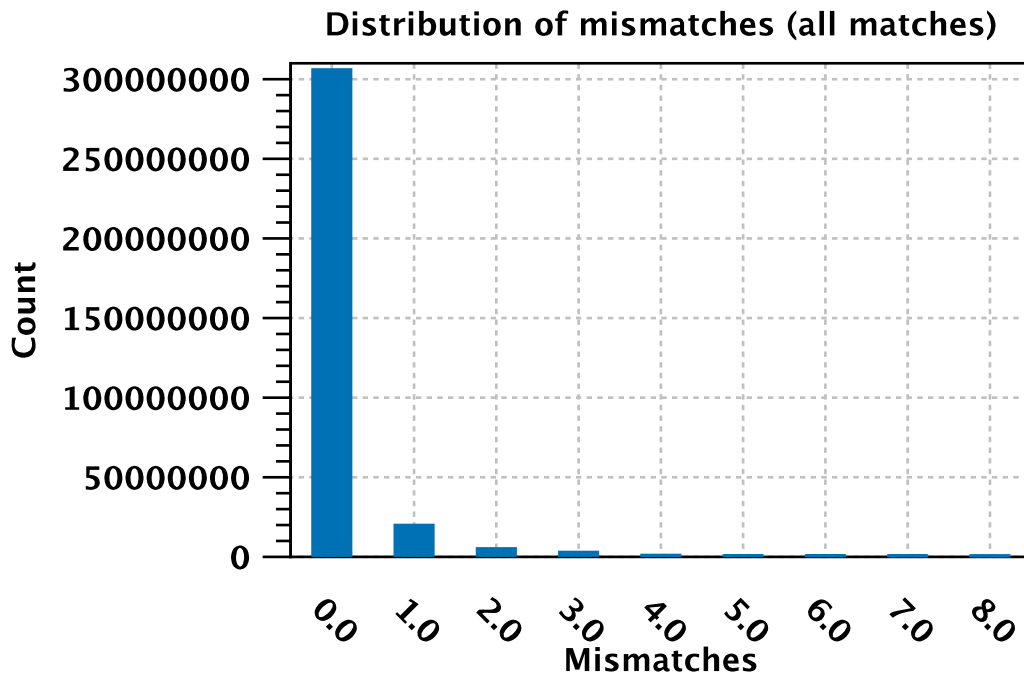


3.3 Non-perfect matches

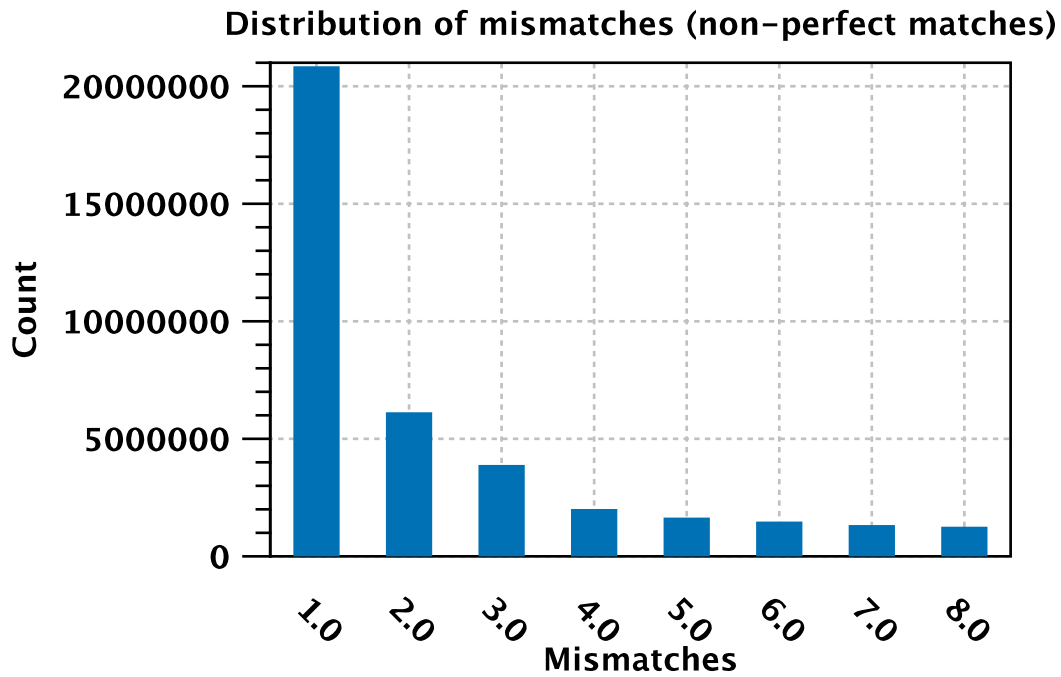
Non-perfect matches

Read count	38,603,662
Mean read length	35.33
Total read length	1,363,801,355

Distribution of mismatches (all matches)



Distribution of mismatches (non-perfect matches)



3.4 Read length distribution

