

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. Michael Guarnieri
Type of report: Assembly Detailed Report
Date: October 22, 2011
Details: Detailed Report for Assembly v.1.0 (with scaffolding)
Most reliable portion that is from 100x to 1000x coverage

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 Paired reads	12
3.5 Read length distribution	13

1. Summary

Reference count	3,600
Type	Reference mapping
Total reference length	37,342,286
GC contents in %	61.49
Total read count	158,482,233
Mean read length	88.72
Total read length	14,060,504,145

2. References

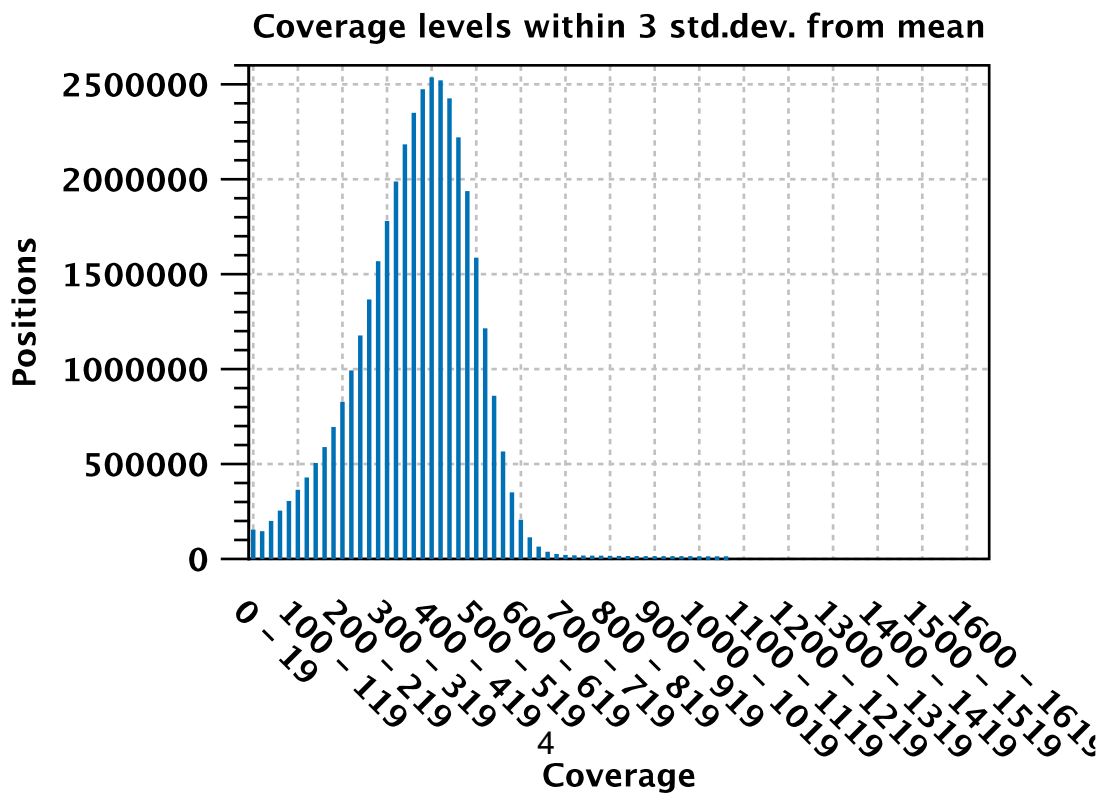
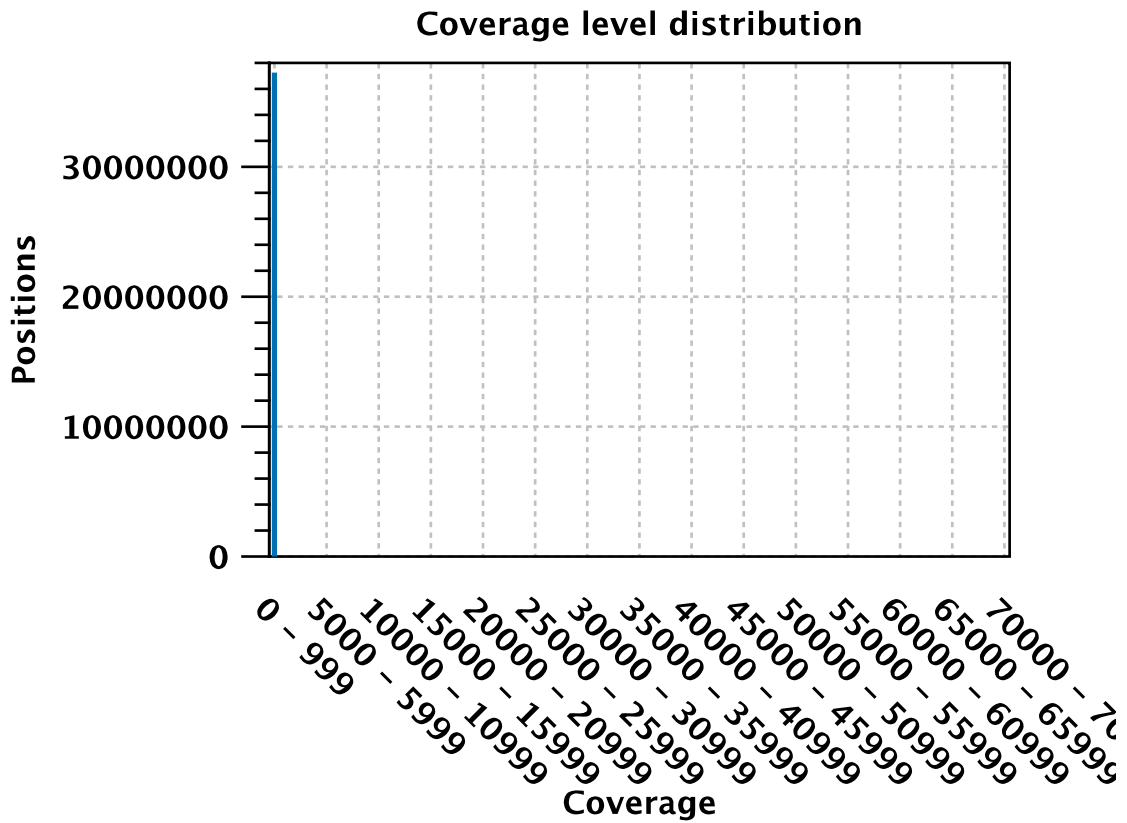
2.1 Reference coverage

Total reference length	37,342,286
% GC	61.49
Total consensus length	37,297,950
Fraction of reference covered	1.00

2.2 Coverage statistics

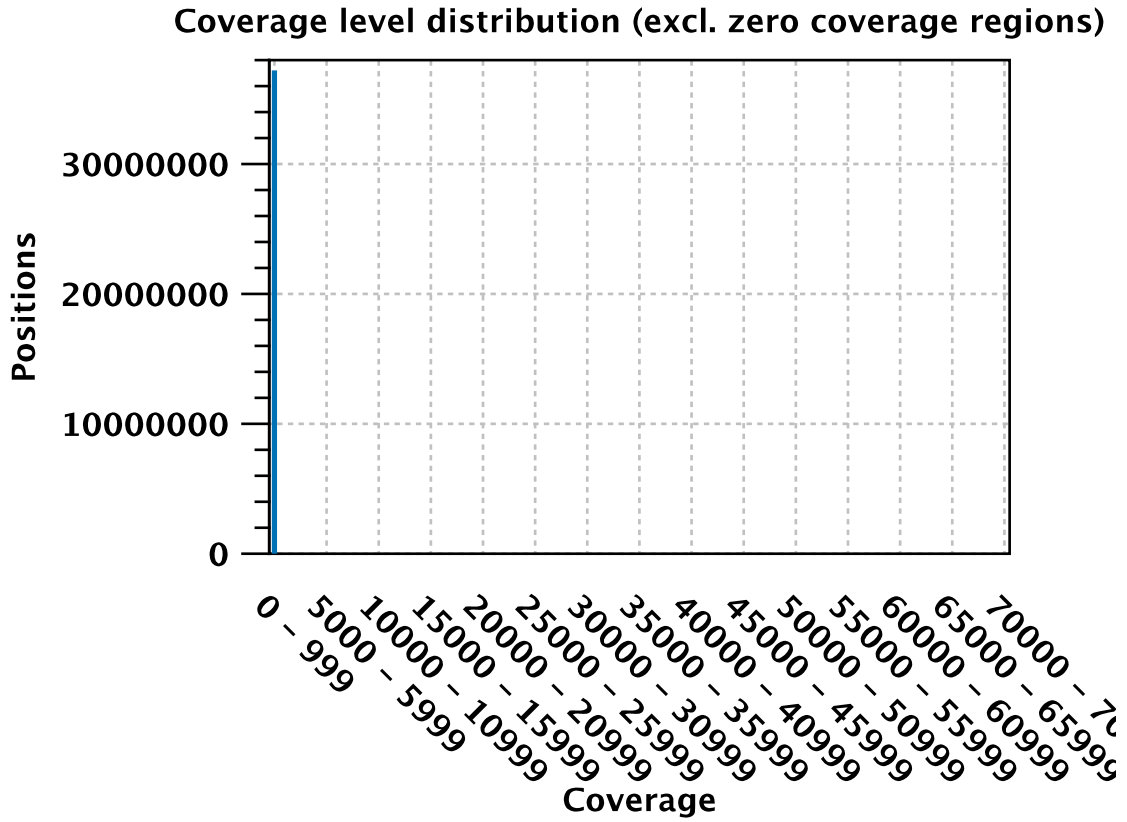
Total reference length	37,342,286
Minimum coverage	0
Maximum coverage	70,887
Average coverage	376.19
Standard deviation	424.00
Minimum excl. zero coverage regions	1
Average excl. zero coverage regions	376.65
Standard deviation excl. zero coverage regions	424.27

2.3 Coverage level distribution



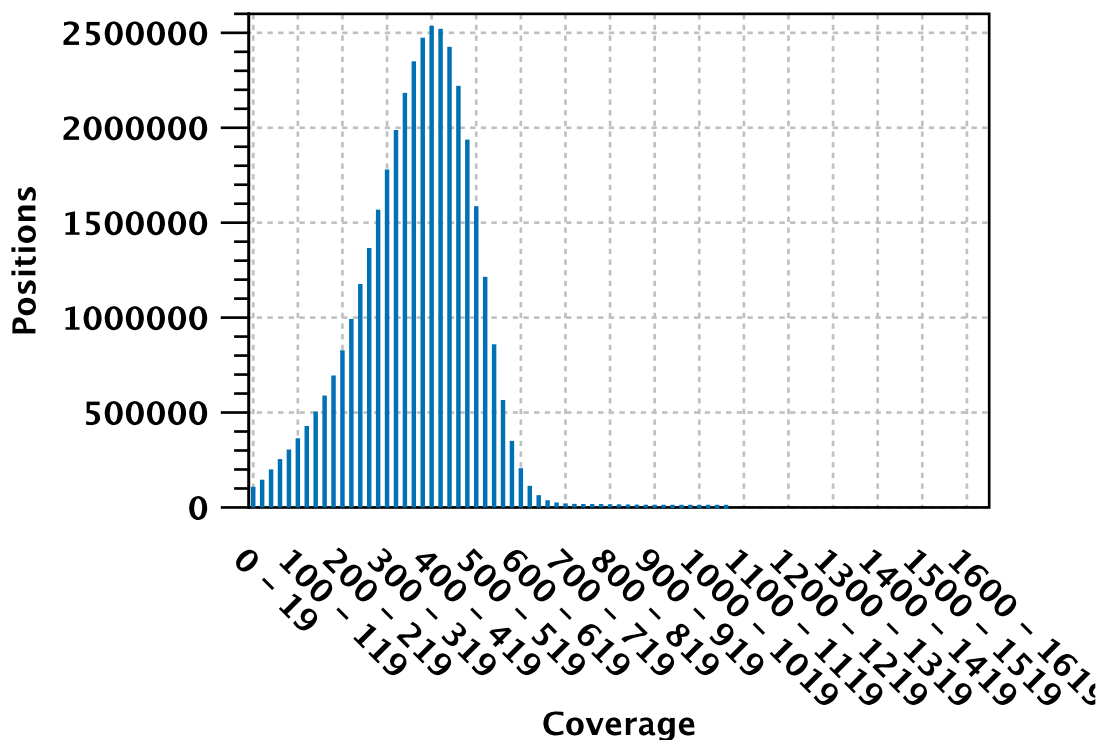
37,330,069 positions have coverage between 0 and 1,648.
12,217 positions have coverage above 1,648 (not shown in graph).

2.4 Coverage level distribution (excl. zero coverage regions)



45,520 positions have zero coverage (not shown in graph).

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

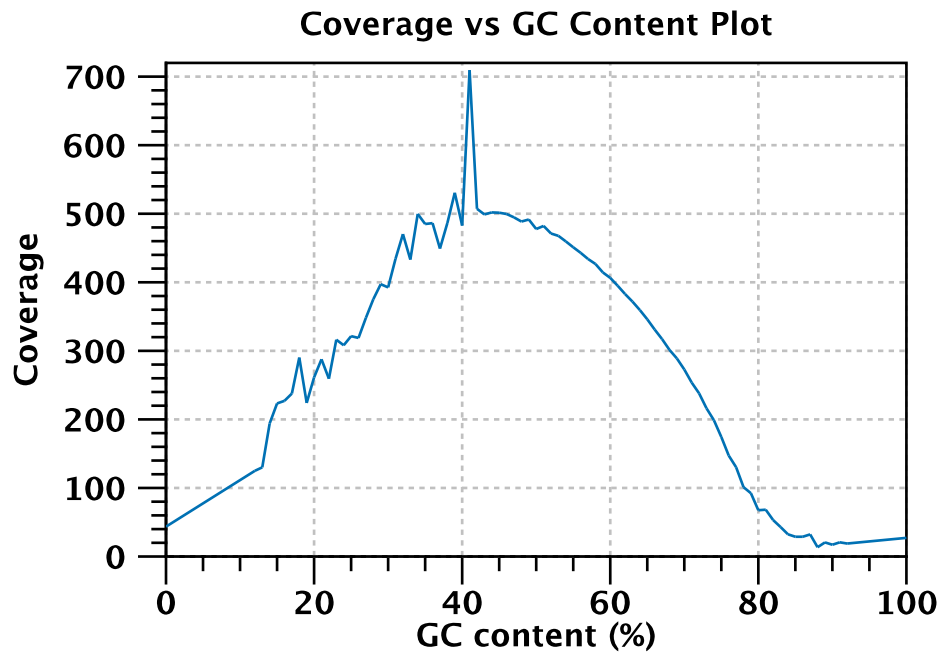


45,520 positions have coverage below 1 (not shown in graph).
 37,284,569 positions have coverage between 1 and 1,649.
 12,197 positions have coverage above 1,649 (not shown in graph).

2.5 Zero coverage regions

Count	2,438
Minimum length	1
Maximum length	169
Mean length	18.67
Standard deviation	25.20
Total length	45,520

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

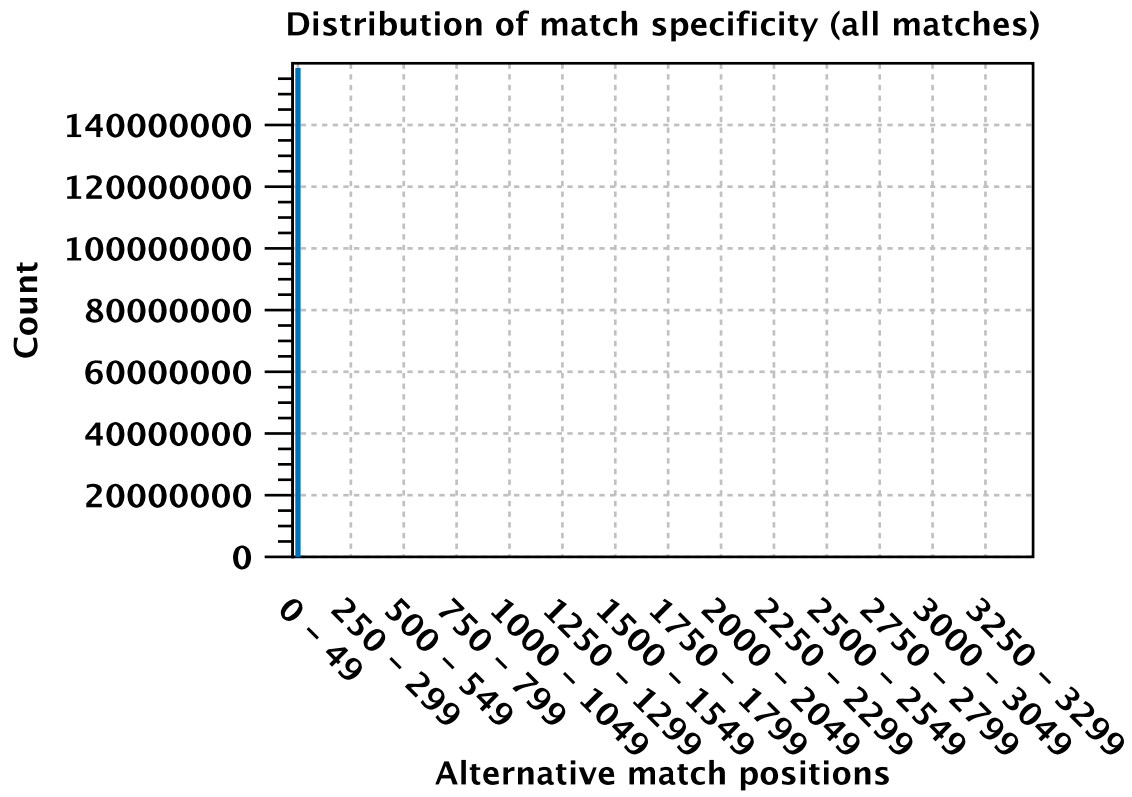
Read count	158,482,233
Mean read length	88.72
Total read length	14,060,504,145

3.2 Non-specific matches

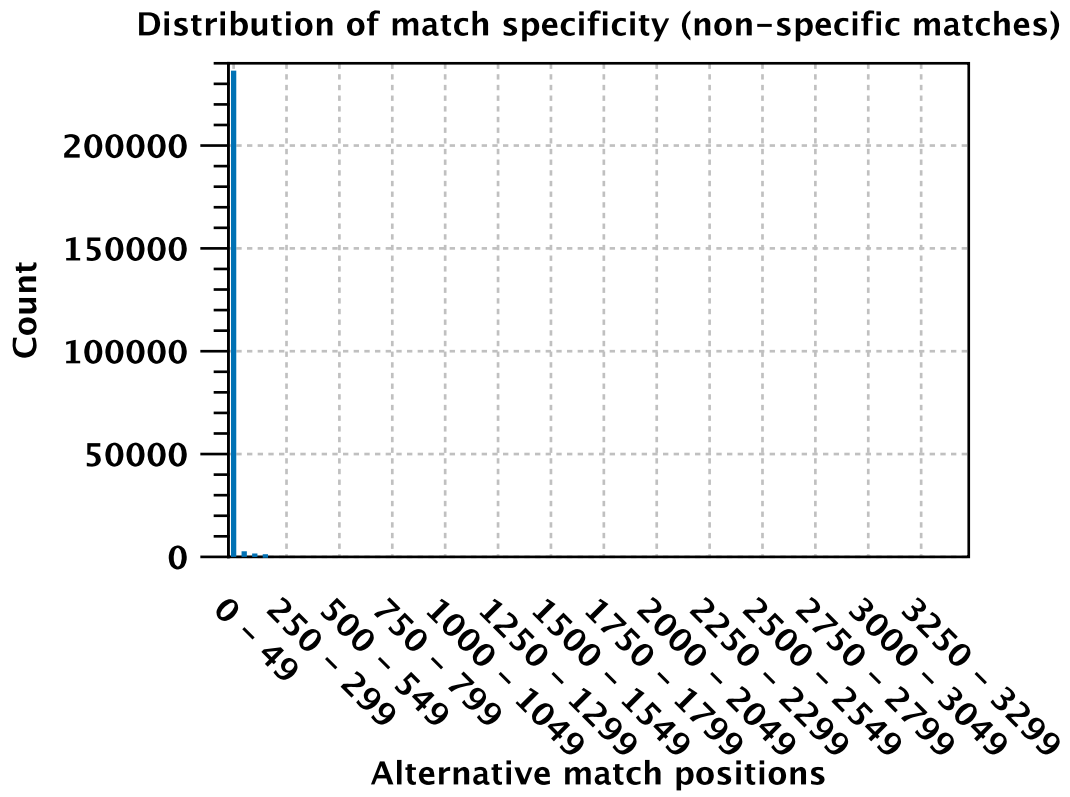
Non-specific matches

Read count	249,251
Mean read length	53.45
Total read length	13,323,032

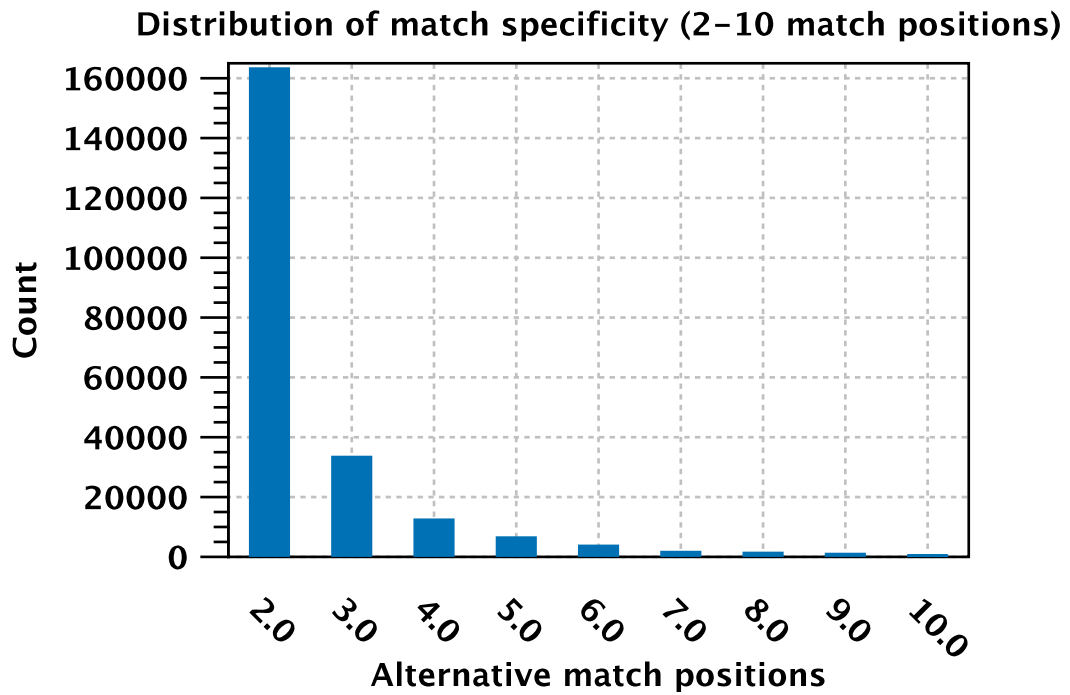
Distribution of match specificity (all matches)



Distribution of match specificity (non-specific matches)



Distribution of match specificity (2-10 match positions)

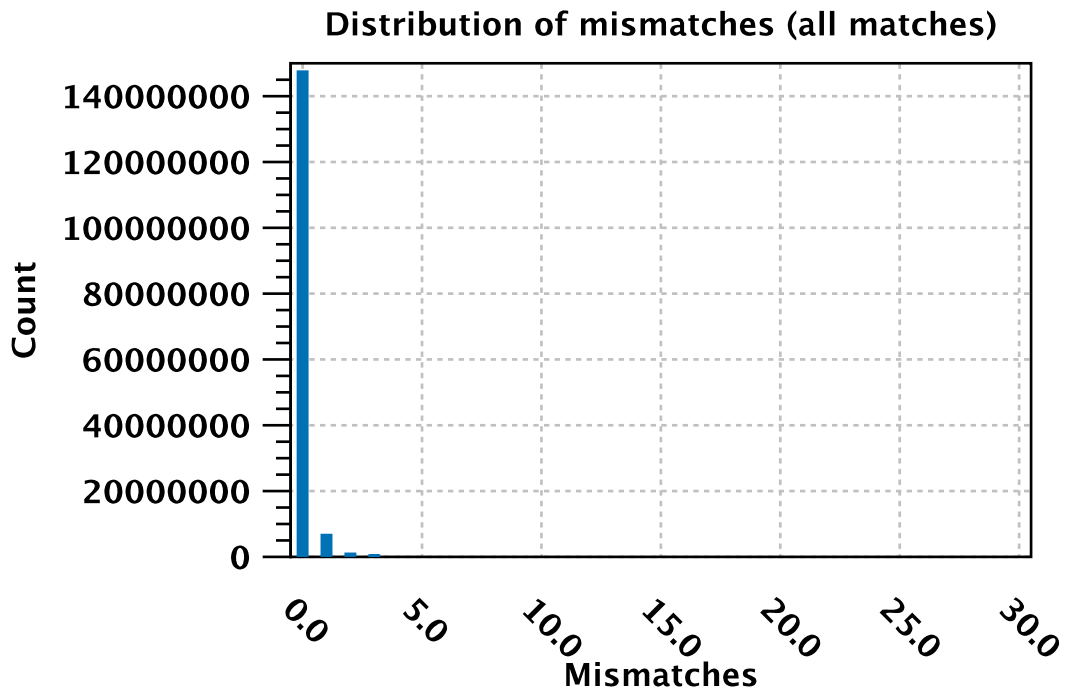


3.3 Non-perfect matches

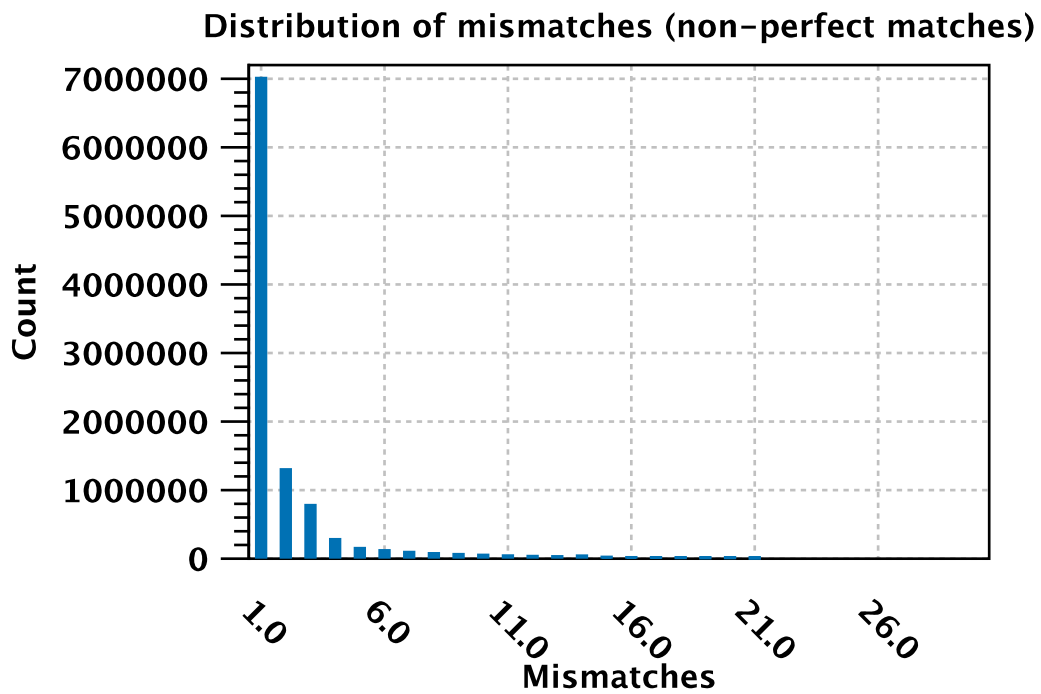
Non-perfect matches

Read count	10,633,015
Mean read length	87.61
Total read length	931,592,150

Distribution of mismatches (all matches)

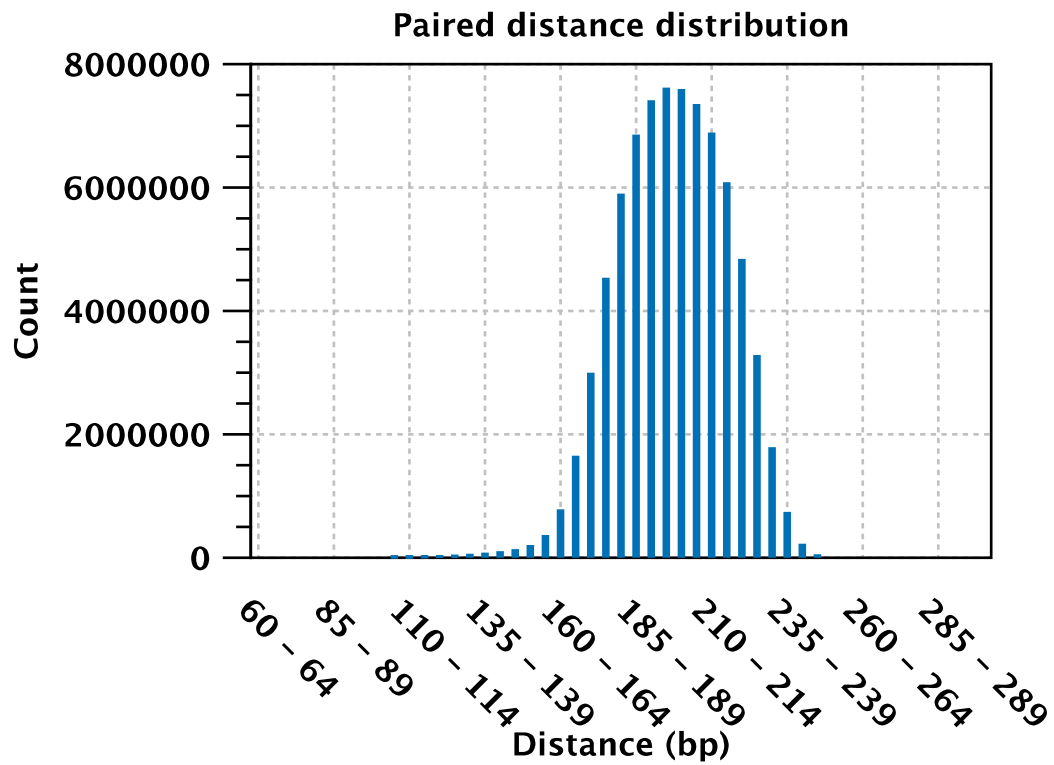


Distribution of mismatches (non-perfect matches)



3.4 Paired reads

Reads in aligned pairs	155,738,624
Reads in broken pairs: wrong distance or mate inverted	138,872
Reads in broken pairs: mate on other contig	1,544,572
Reads in broken pairs: mate not mapped	1,060,165
Mean distance	199.09
Standard deviation	199.73



3.5 Read length distribution

