

## **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)  
Highly repeated portion that is >1000x

## 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	113
Type	Reference mapping
Total reference length	73,940
GC contents in %	58.48
Total read count	4,281,083
Mean read length	89.07
Total read length	381,302,782

# 2. References

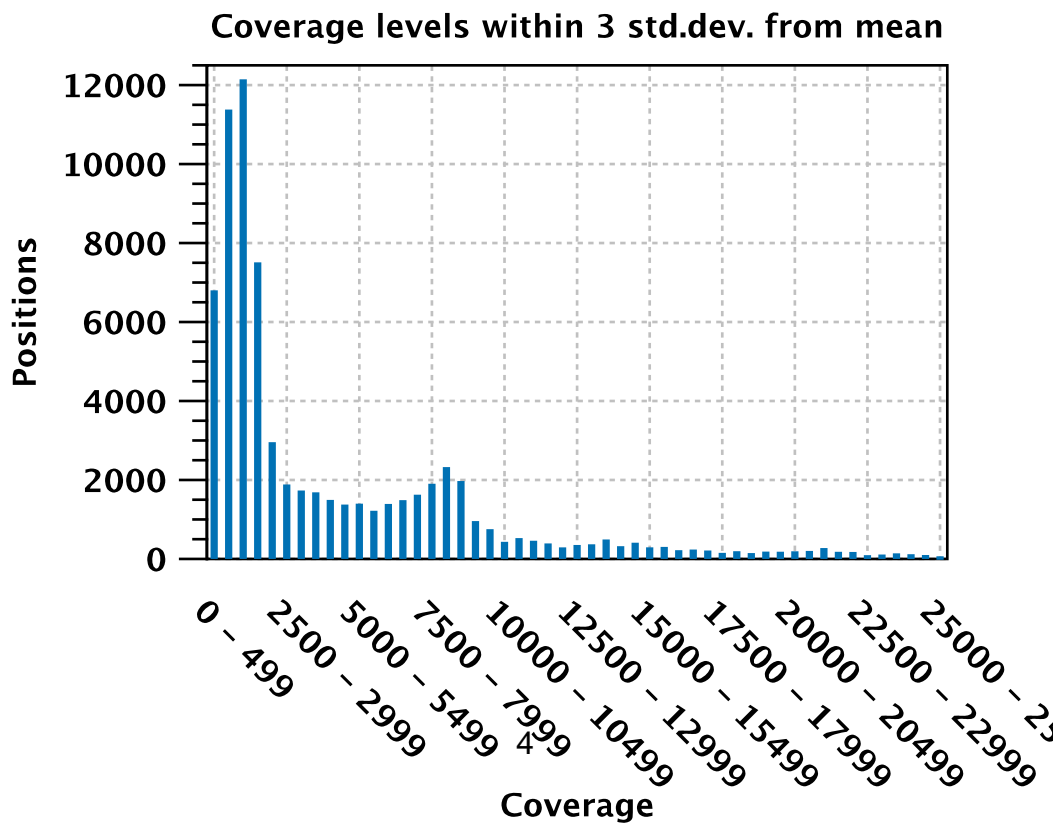
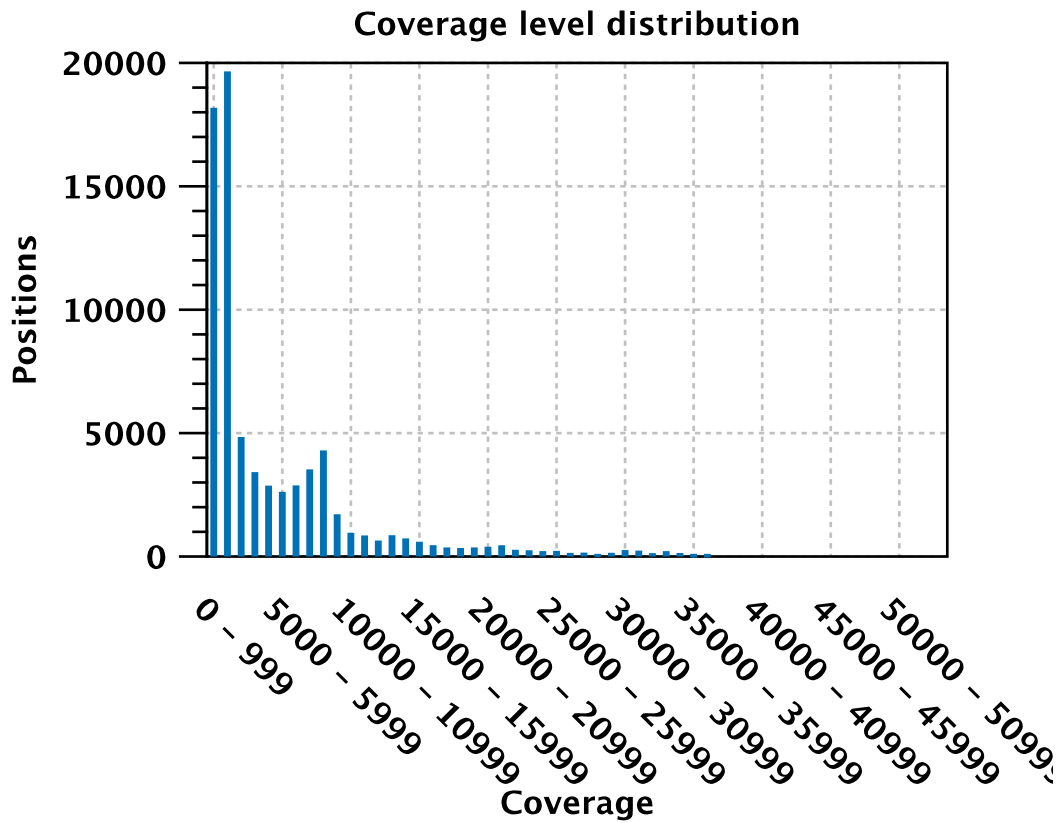
## 2.1 Reference coverage

Total reference length	73,940
% GC	58.48
Total consensus length	72,935
Fraction of reference covered	0.99

## 2.2 Coverage statistics

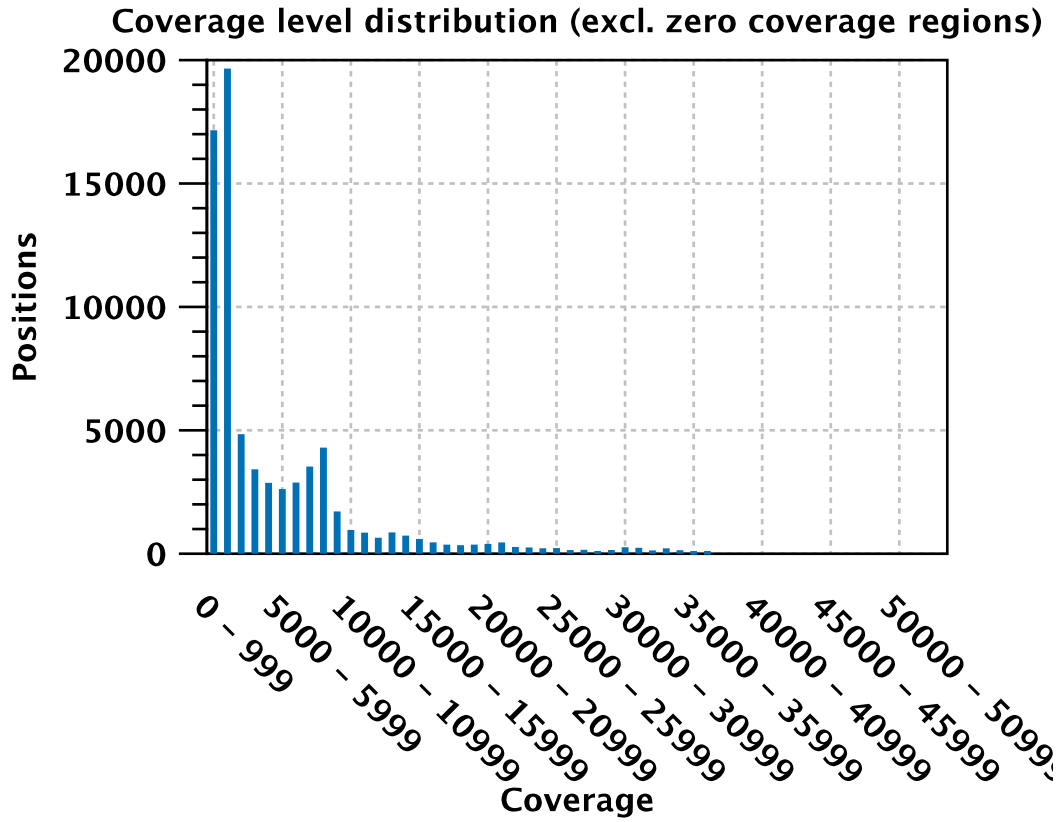
Total reference length	73,940
Minimum coverage	0
Maximum coverage	53,218
Average coverage	5,093.20
Standard deviation	6,701.47
Minimum excl. zero coverage regions	1
Average excl. zero coverage regions	5,164.72
Standard deviation excl. zero coverage regions	6,720.93

## 2.3 Coverage level distribution



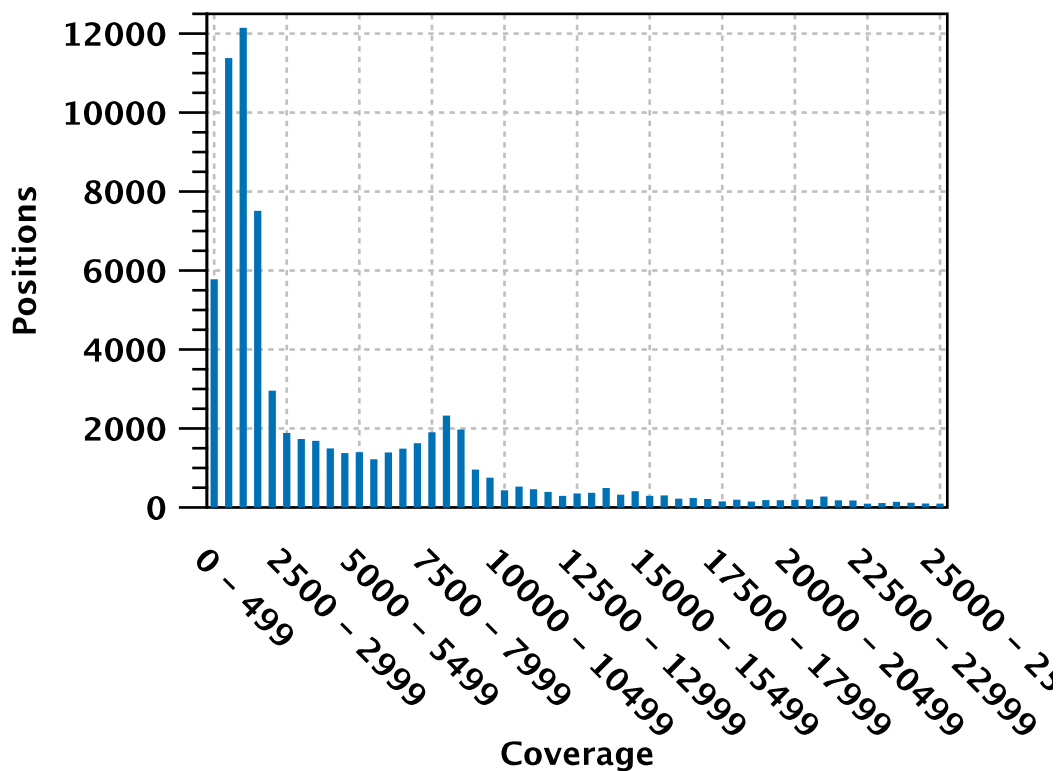
71,869 positions have coverage between 0 and 25,194.  
2,071 positions have coverage above 25,194 (not shown in graph).

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



1,024 positions have zero coverage (not shown in graph).

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



1,024 positions have coverage below 1 (not shown in graph).

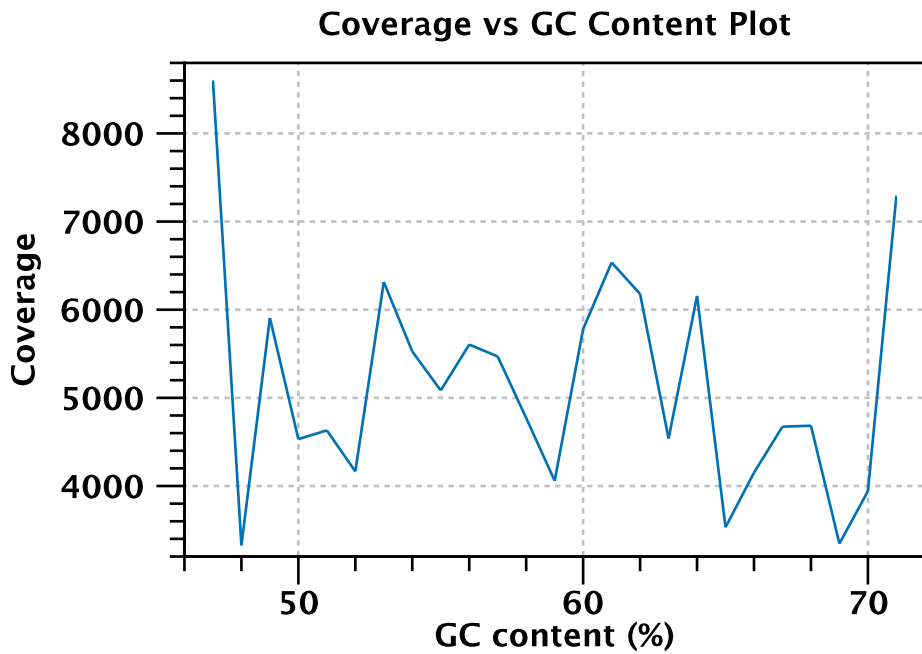
70,895 positions have coverage between 1 and 25,324.

2,021 positions have coverage above 25,324 (not shown in graph).

## 2.5 Zero coverage regions

Count	27
Minimum length	3
Maximum length	104
Mean length	37.93
Standard deviation	28.48
Total length	1,024

## 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

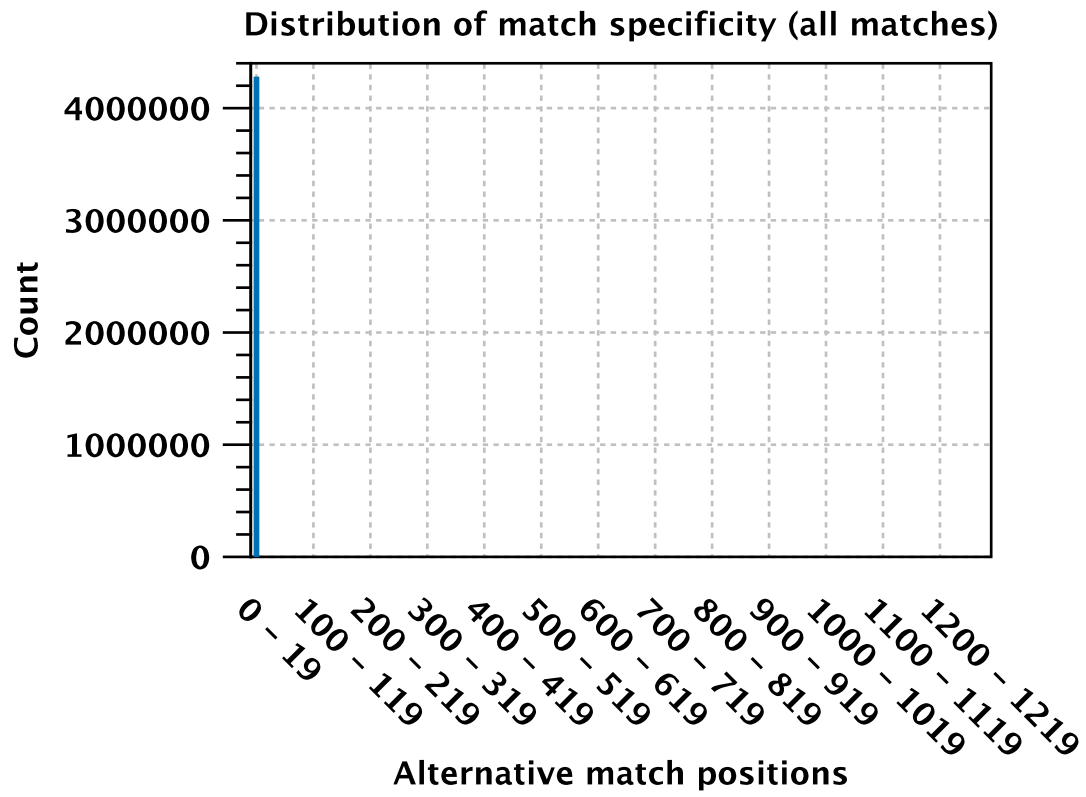
Read count	4,281,083
Mean read length	89.07
Total read length	381,302,782

### 3.2 Non-specific matches

#### *Non-specific matches*

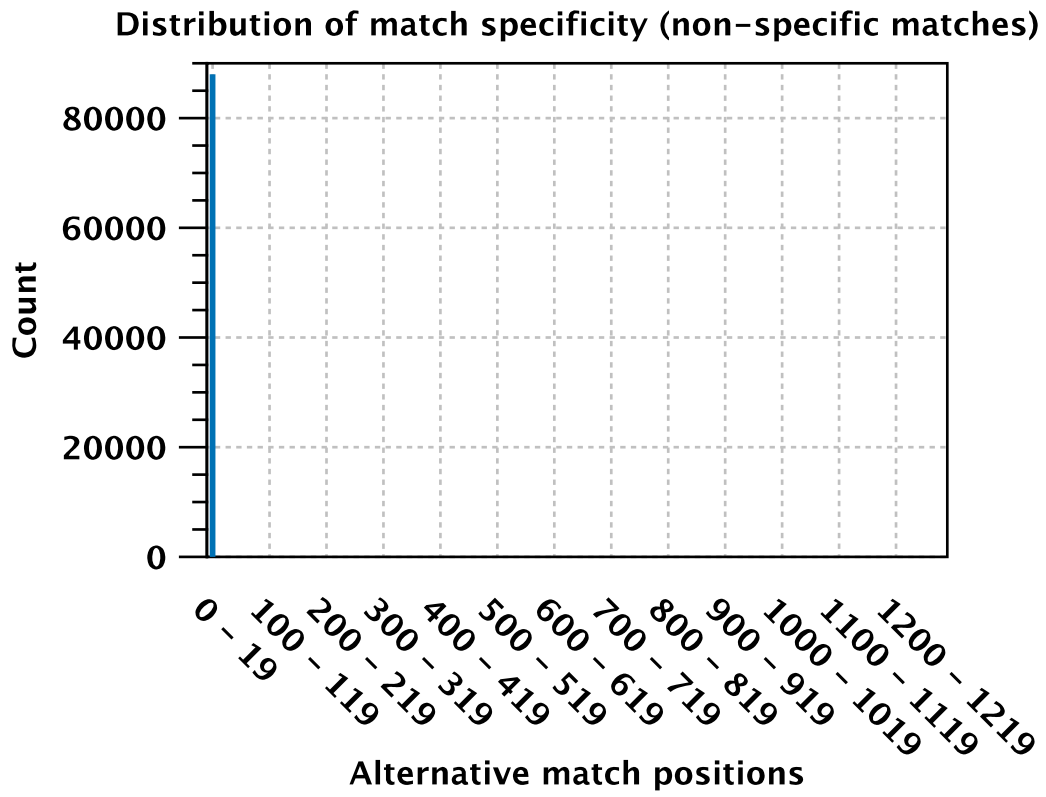
Read count	88,008
Mean read length	57.32
Total read length	5,044,210

*Distribution of match specificity (all matches)*

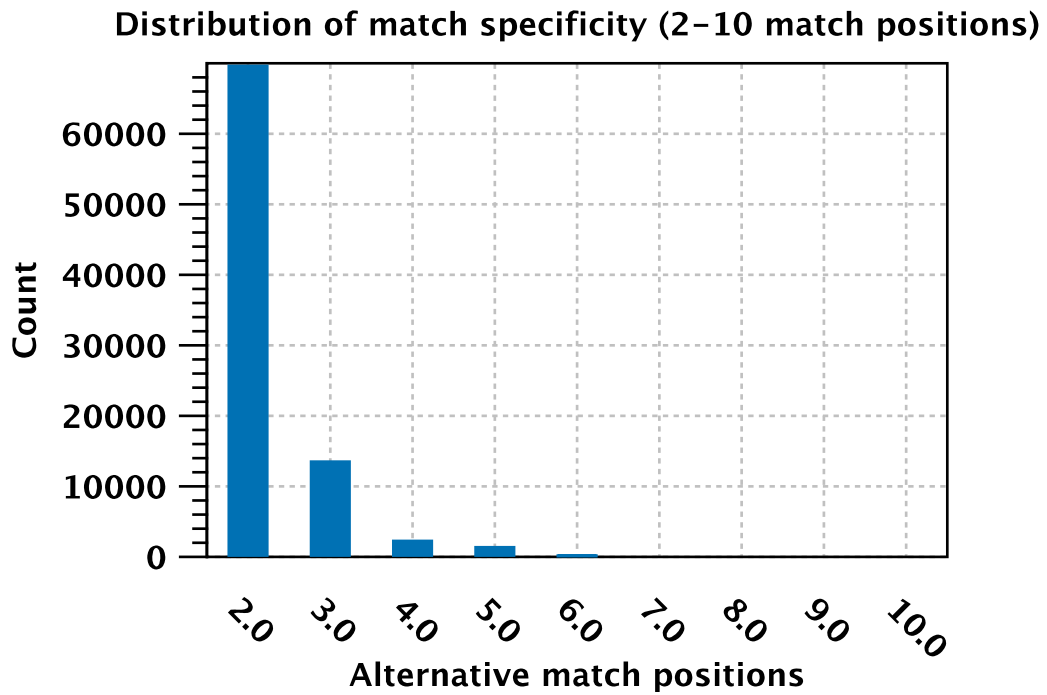




*Distribution of match specificity (non-specific matches)*



## Distribution of match specificity (2-10 match positions)

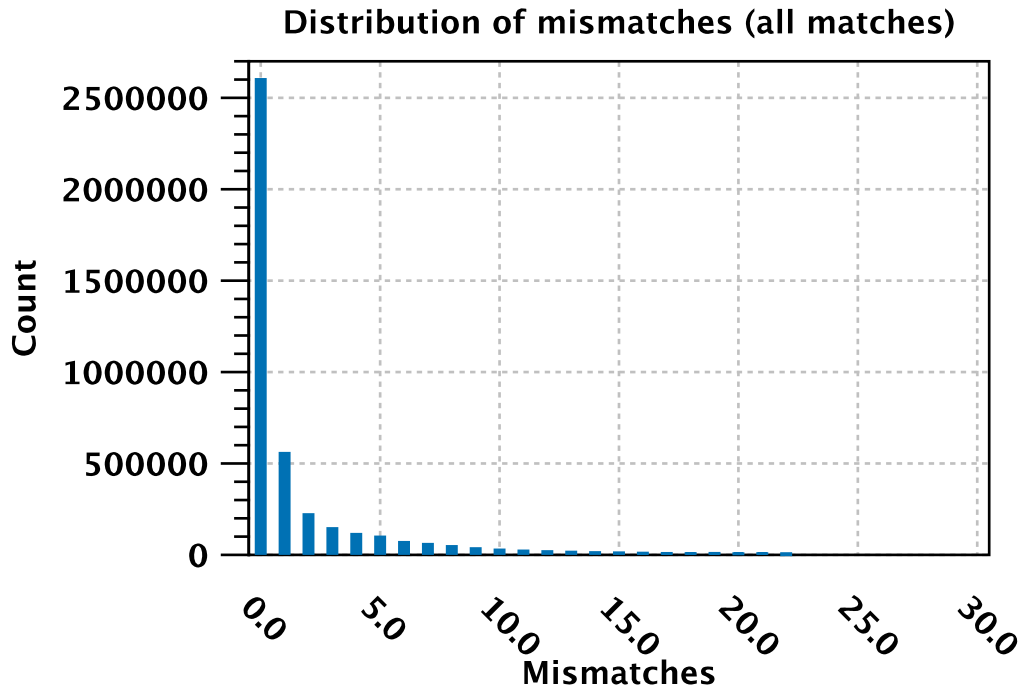


### 3.3 Non-perfect matches

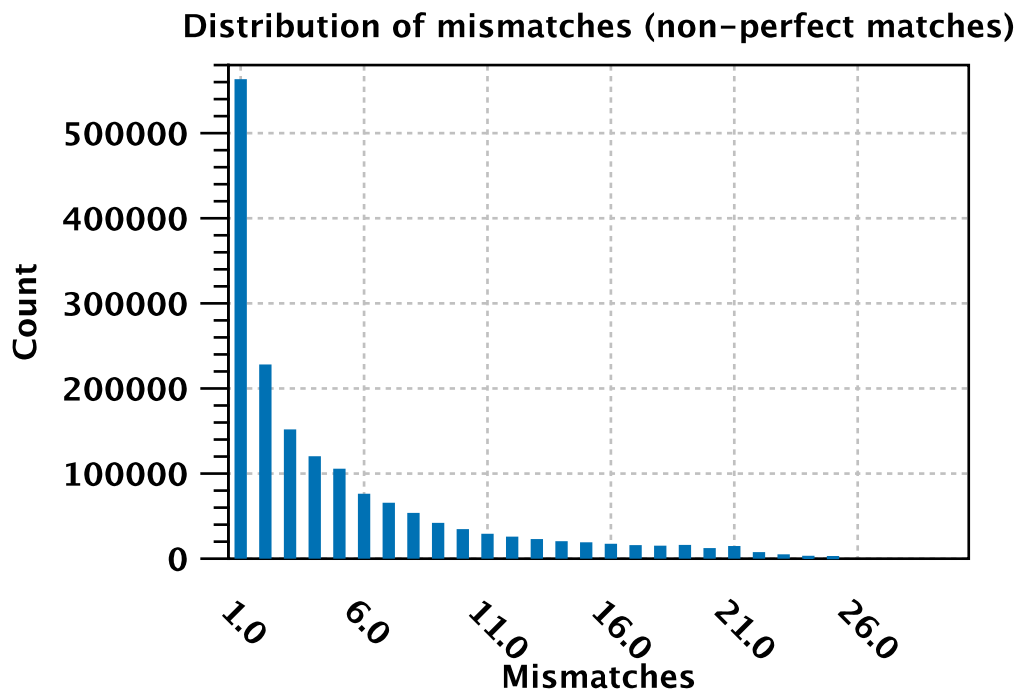
#### *Non-perfect matches*

Read count	1,672,733
Mean read length	89.36
Total read length	149,472,831

*Distribution of mismatches (all matches)*

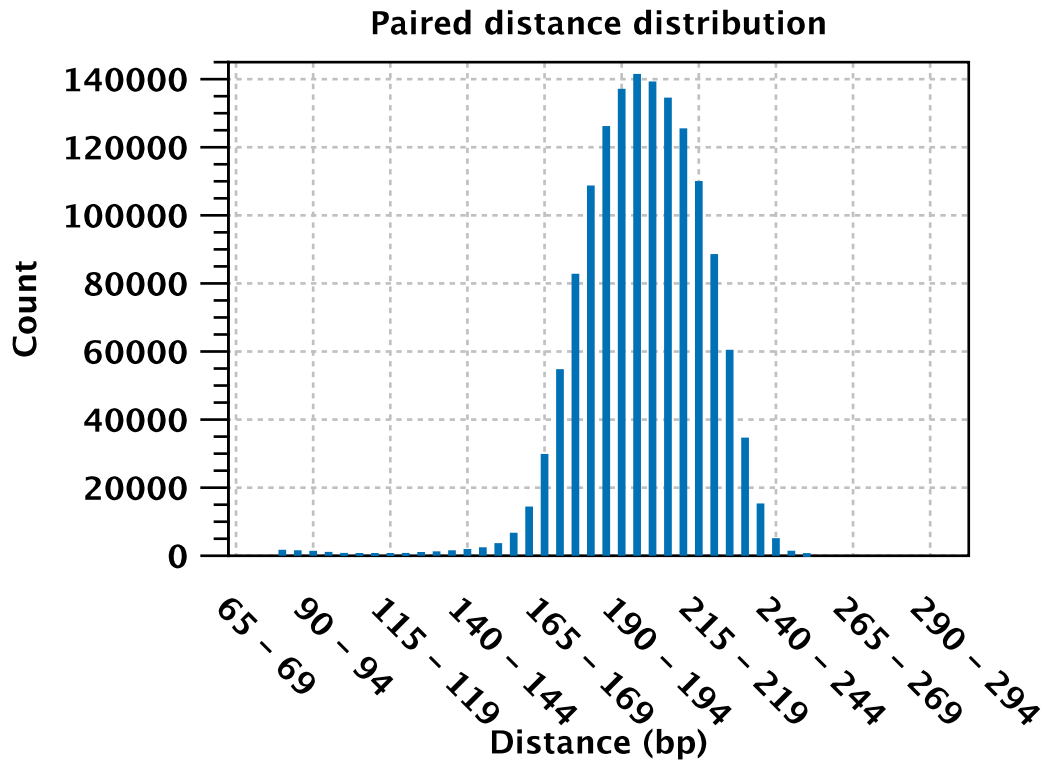


*Distribution of mismatches (non-perfect matches)*



### 3.4 Paired reads

Reads in aligned pairs	2,883,154
Reads in broken pairs: wrong distance or mate inverted	87,694
Reads in broken pairs: mate on other contig	639,670
Reads in broken pairs: mate not mapped	670,565
Mean distance	198.70
Standard deviation	20.11



### 3.5 Read length distribution

