

Bioinformatic Analysis Report

Genome Project Solutions
1024 Promenade Street
Hercules, CA 94547

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

Project title: Genome Sequence of Nannochloropsis
Prepared for: Dr. Randor Radakovits and Dr. Matthew Posewitz
Type of report: Detailed Reference Alignment Report
Date: September 1, 2011
Details: All pooled RNA sequencing reads aligned to 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	9
3.4 Aligned pairs	9
3.5 Read length distribution	11

1. Summary

Reference count	37,055
Type	Reference mapping
Total reference length	17,024,428
GC contents in %	56.29
Total read count	16,094,133
Mean read length	50.51
Total read length	812,838,793

2. References

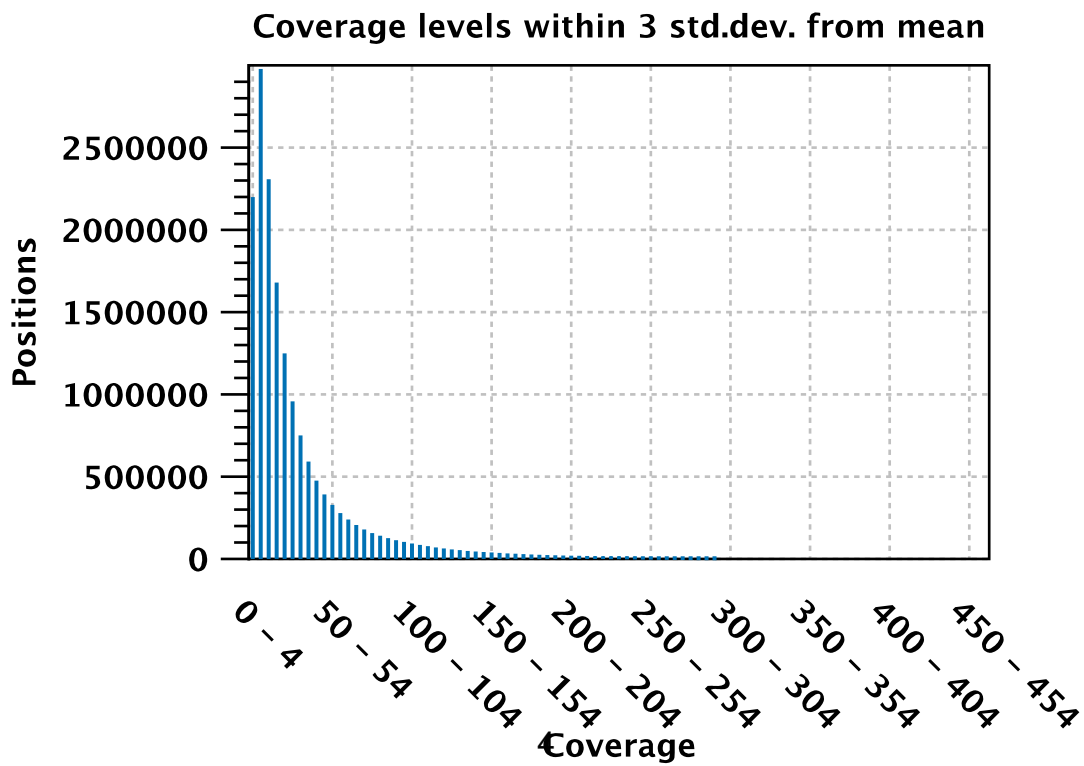
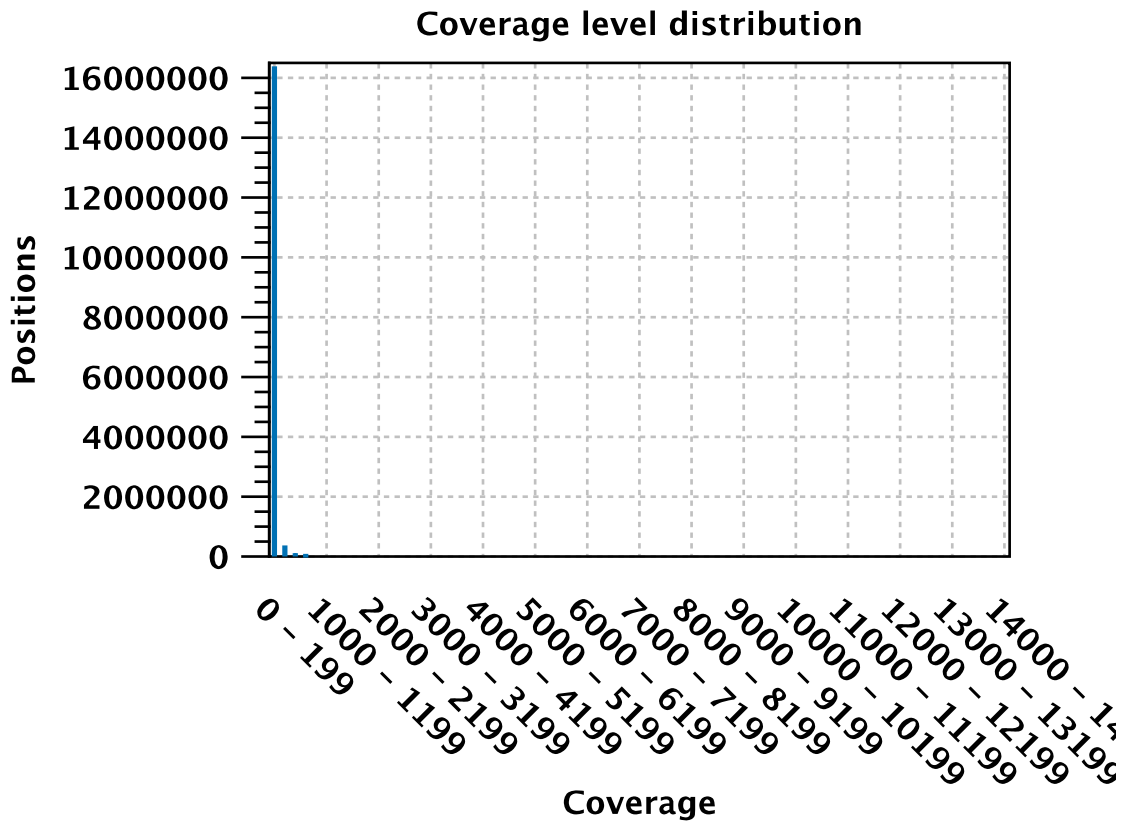
2.1 Reference coverage

Total reference length	17,024,428
% GC	56.29
Total consensus length	16,992,901
Fraction of reference covered	1.00

2.2 Coverage statistics

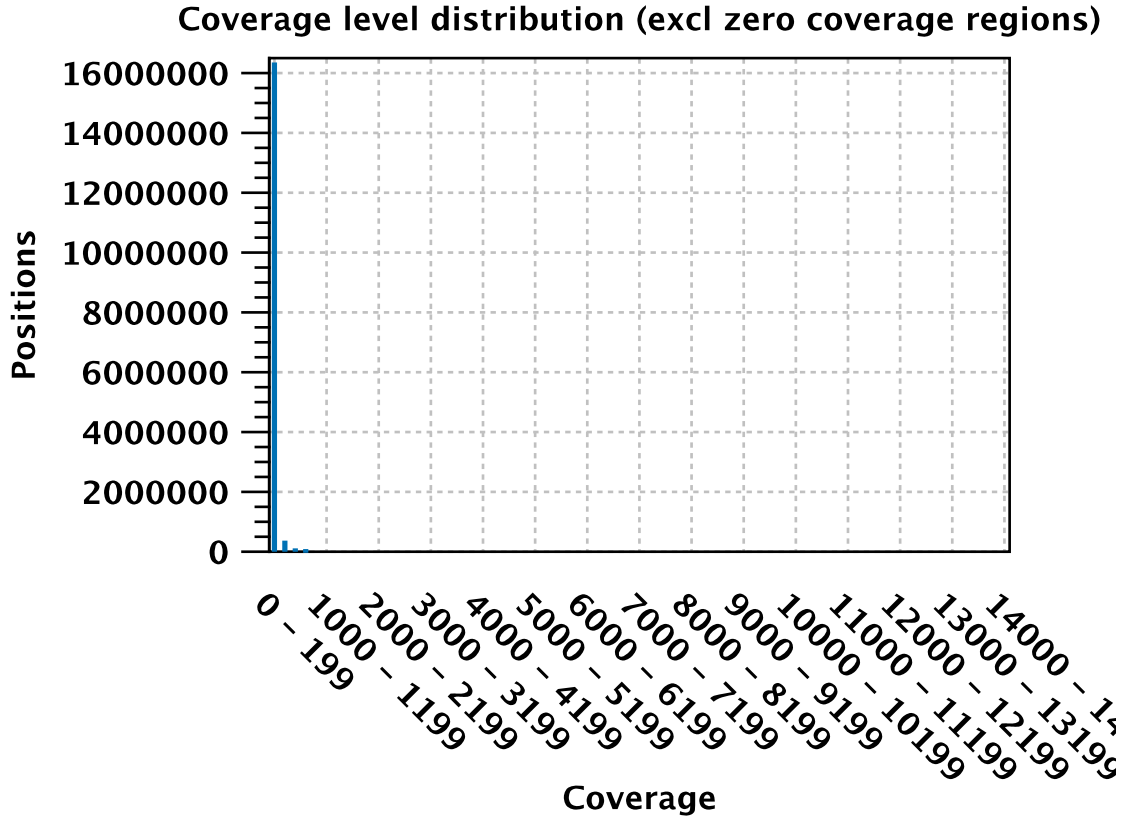
Total reference length	17,024,428
Minimum coverage	0
Maximum coverage	14,068
Average coverage	47.57
Standard deviation	137.71
Minimum excl zero coverage regions	1
Average excl. zero coverage regions	47.66
Standard deviation excl- zero coverage regions	137.82

2.3 Coverage level distribution



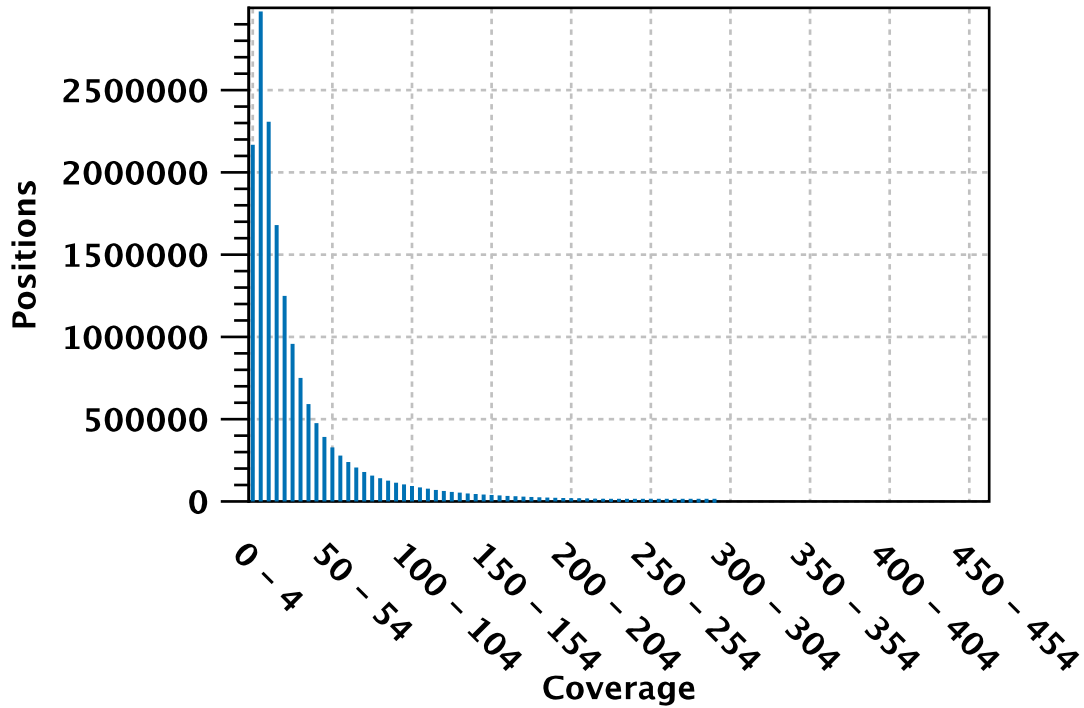
16,802,609 positions have coverage between 0 and 460.
221,819 positions have coverage above 460 (not shown in graph).

2.4 Coverage level distribution (excl zero coverage regions)



31,527 positions have zero coverage (not shown in graph).

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

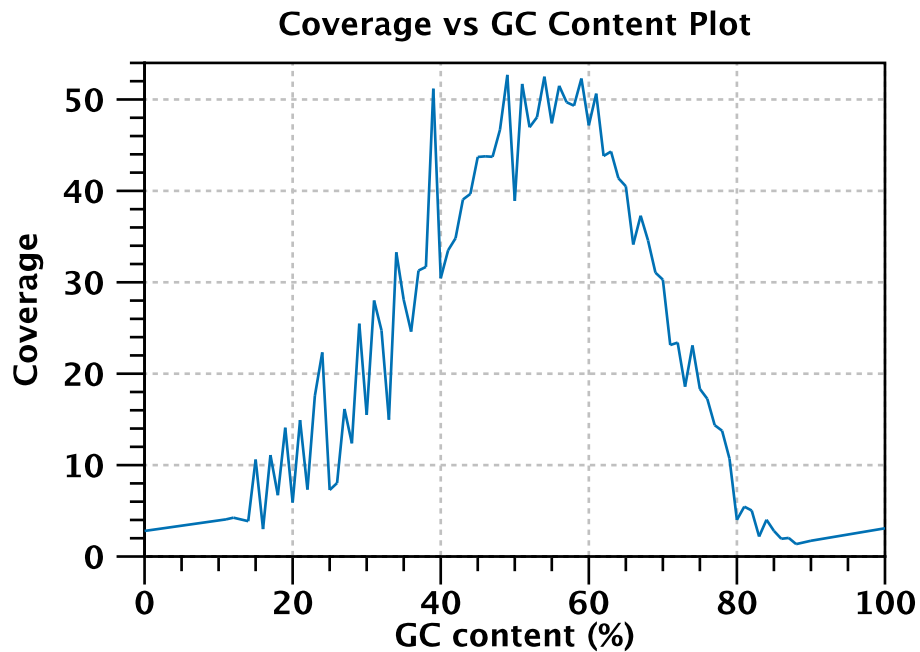


31,527 positions have coverage below 1 (not shown in graph).
 16,771,744 positions have coverage between 1 and 461.
 221,157 positions have coverage above 461 (not shown in graph).

2.5 Zero coverage regions

Count	3,642
Minimum length	1
Maximum length	60
Mean length	8.66
Standard deviation	9.62
Total length	31,527
Regions	

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

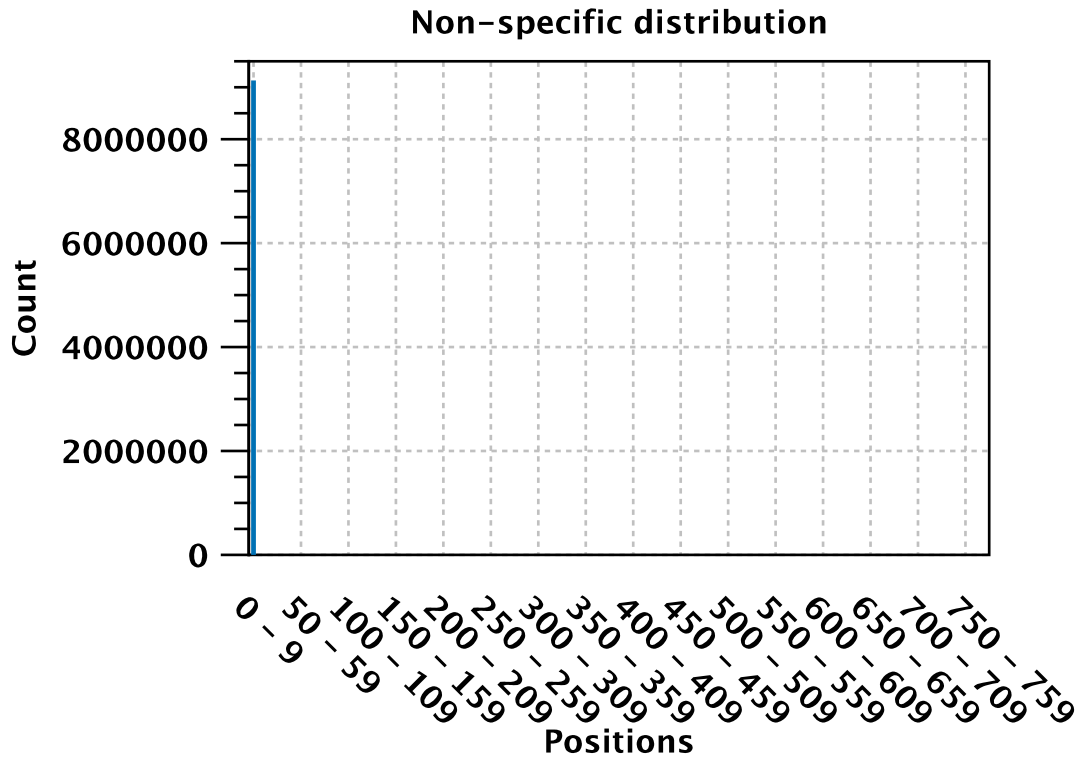
Read count	16,094,133
Mean read length	50.51
Total read length	812,838,793

3.2 Non-specific matches

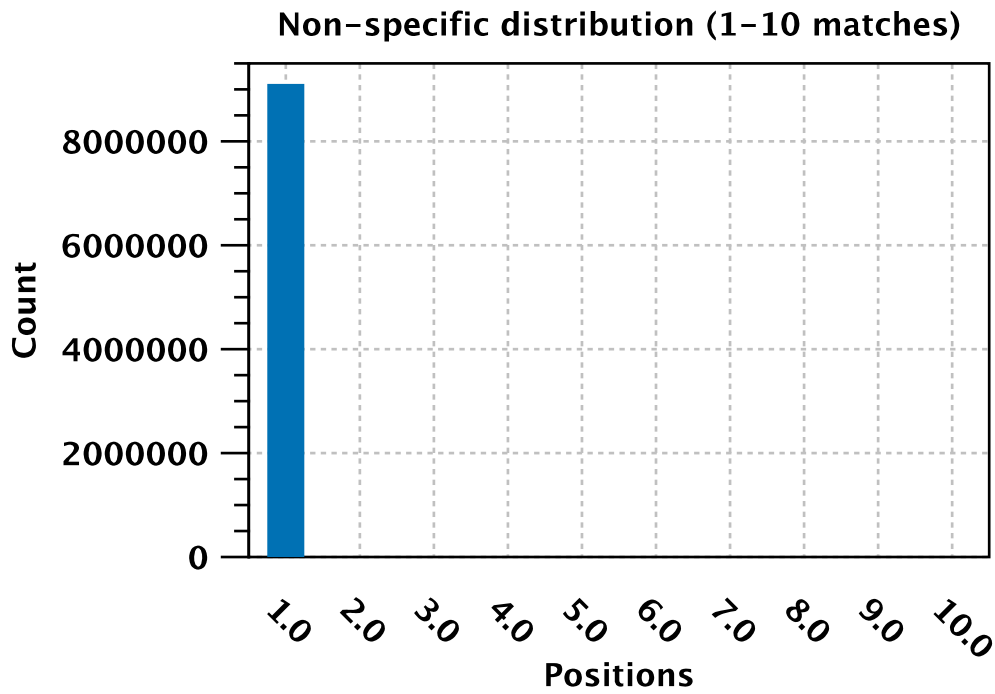
Non-specific matches

Read count	27,961
Mean read length	46.91
Total read length	1,311,613

Non-specific distribution



Non-specific distribution (1-10 matches)

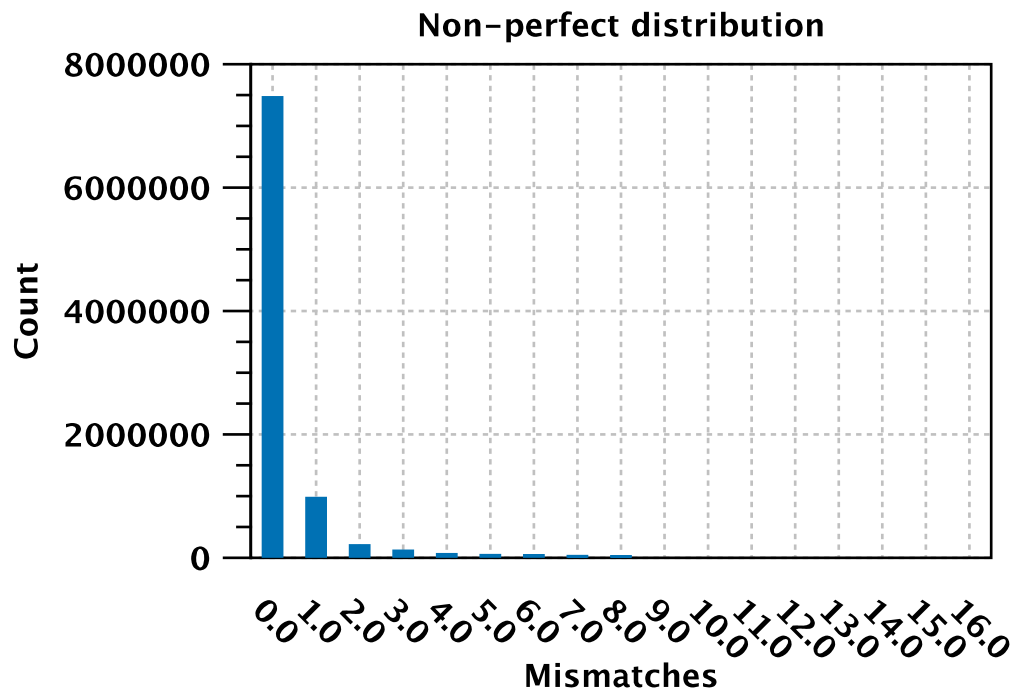


3.3 Non-perfect matches

Non-perfect matches

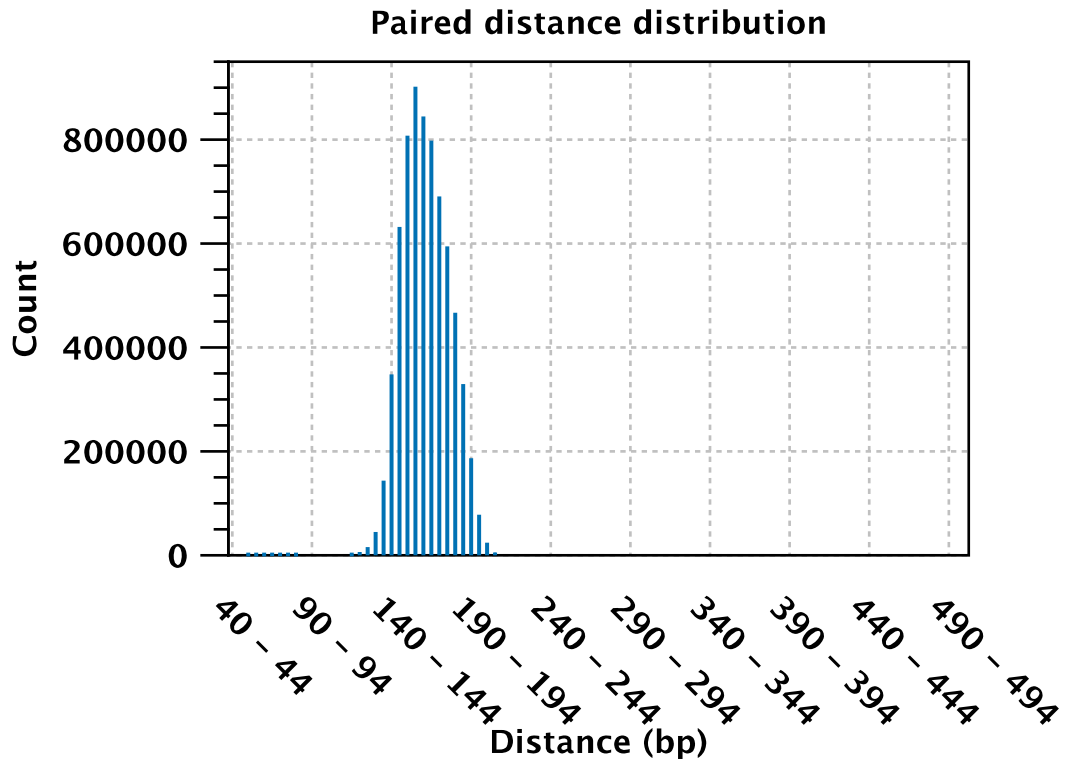
Read count	1,649,107
Mean read length	85.67
Total read length	141,283,280

Non-perfect distribution



3.4 Aligned pairs

Aligned pairs	6,960,747
Mean distance	163.40
Standard deviation	15.94
Reads from broken pairs: wrong distance	144,858
Reads from broken pairs: mate inverted	118,656
Reads from broken pairs: mate on other contig	895,922



3.5 Read length distribution

