

## **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 Processing

Date: May 25, 2012

Details: Report on removing duplicated reads and trimming of RNAseq data prior to assembling the transcriptome contigs

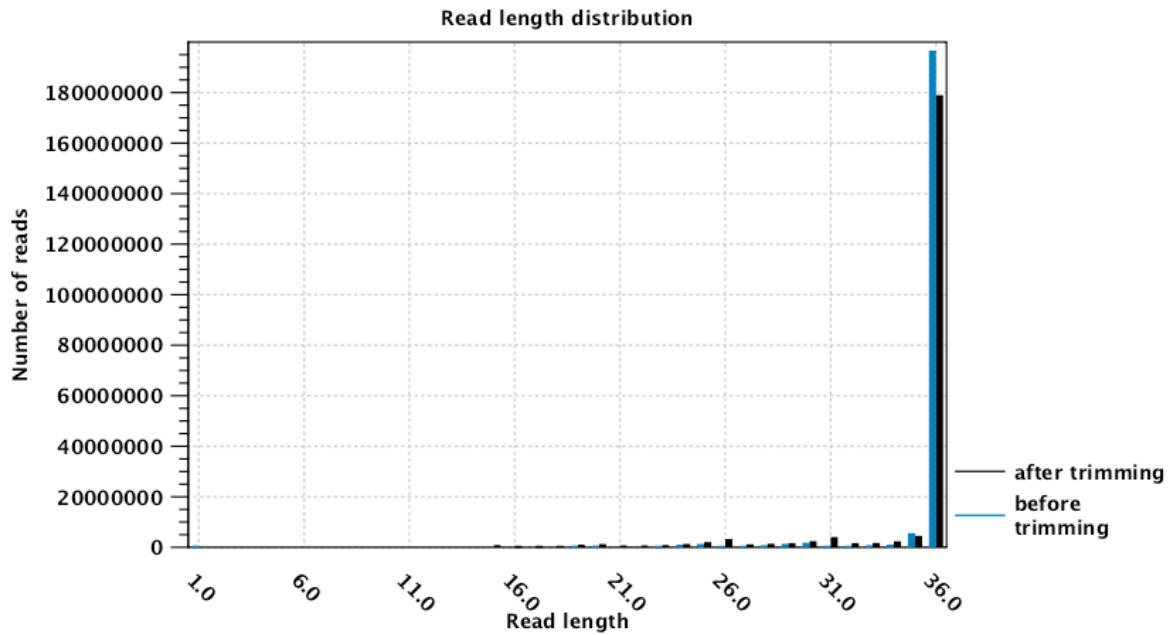
### RNAseq removal of duplicated reads

Name	Input sequences	Duplicates	Remaining sequences	Duplicates %	Remaining %
Hetero	29,558,009	12,671,503	16,886,506	42.87	57.13
OD2a	25,299,236	11,880,086	13,419,150	46.96	53.04
OD2b	28,674,844	12,984,042	15,690,802	45.28	54.72
OD2c	29,514,213	15,488,082	14,026,131	52.48	47.52
OD4A	28,115,213	11,745,949	16,369,264	41.78	58.22
OD4B	26,151,032	11,402,984	14,748,048	43.60	56.40
OD6A	30,859,737	14,467,341	16,392,396	46.88	53.12
OD6B	34,208,959	16,713,756	17,495,203	48.86	51.14
OD7Ra	30,531,209	15,952,524	14,578,685	52.25	47.75
OD7Ra2	30,630,285	15,741,332	14,888,953	51.39	48.61
OD7Rb	28,761,862	14,318,428	14,443,434	49.78	50.22
OD8a	30,182,048	15,311,846	14,870,202	50.73	49.27
OD8b	29,259,262	15,104,991	14,154,271	51.62	48.38
OD8c	30,491,089	12,864,811	17,626,278	42.19	57.81

## 1 Trim summary

Name	Number of reads	Avg.length	Number of reads after trim	Percentage trimmed	Avg.length after trim
Hetero without duplicates	16,886,506	34.8	16,362,979	96.9%	33.4
OD2a without duplicates	13,419,150	34.6	12,941,044	96.44%	34.0
OD2b without duplicates	15,690,802	35.4	15,615,033	99.52%	35.2
OD2c without duplicates	14,026,131	35.7	13,947,253	99.44%	35.4
OD4A without duplicates	16,369,264	35.6	16,217,215	99.07%	35.4
OD4B without duplicates	14,748,048	35.6	14,622,209	99.15%	35.1
OD6A without duplicates	16,392,396	35.4	16,254,022	99.16%	34.9
OD6B without duplicates	17,495,203	35.4	17,361,612	99.24%	34.9
OD7Ra without duplicates	14,578,685	35.3	14,480,964	99.33%	35.0
OD7Ra2 without duplicates	14,888,953	35.2	14,742,420	99.02%	34.9
OD7Rb without duplicates	14,443,434	35.4	14,348,281	99.34%	35.2
OD8a without duplicates	14,870,202	35.2	14,727,321	99.04%	34.8
OD8b without duplicates	14,154,271	35.5	14,081,897	99.49%	35.2
OD8c without duplicates	17,626,278	35.0	17,185,562	97.5%	33.4

## 2 Read length before / after trimming



## 3 Trim settings

- Removal of low quality sequence. (limit = 0.01).
- Removal of ambiguous nucleotides: No ambiguous nucleotides allowed.
- Removal of sequences on length: minimum length 15 nucleotides.

## 4 Detailed trim results

Trim	Input reads	No trim	Trimmed	Nothing left or Discarded
Trim on quality	215,589,323	188,773,337	26,815,959	27
Ambiguity trim	215,589,296	214,139,407	1,398,027	51,862
Filter on length	215,537,434	212,887,812	0	2,649,622