phixscreen_jigsaw

Jigsaw report 02/13/2014

1. Read Statistics

Illumina Std PE Statistics

File name	TNNZ.7707.1.80834.GTGAA.nophix.contam.artifact.clean.fastq
Number of reads	6,859,678
Sequencing depth †	344X
Read type	2x251 bp

† A genome size of 5.0 Mbp was assumed in this calculation.

2. Read QC Results

The following are the results of reads screened against contaminants. Pairs of matching reads were removed from the dataset.

Illumina Std PE Read Filter Statistics

Description	Num Reads	Pct Reads
Input	7,287,298	100
Contam removed	230	0.0
Artifact removed	427,390	5.9
Total removed	427,620	5.9
Total remaining	6,859,678	94.1

List of Contaminants Removed

Description	Num Reads	Pct Reads
gi|357579577|Canis_lupus_familiaris_chr3	178	0.00
human_chr2	144	0.00
human_chr15	18	0.00
human_chr3	6	0.00
human_chr13	4	0.00
human_chr10	4	0.00
human_chr9	4	0.00
human_chr1	2	0.00
human_chr18	2	0.00
human_chr17	2	0.00
human_chr14	2	0.00
human_chr16	2	0.00
human_chr8	2	0.00
human_chr4	2	0.00
human_chr6	2	0.00

The following are the results of reads screened against potential reagent and process contaminants but were not removed from the dataset.

Illumina Std PE Contamination Identification Statistics

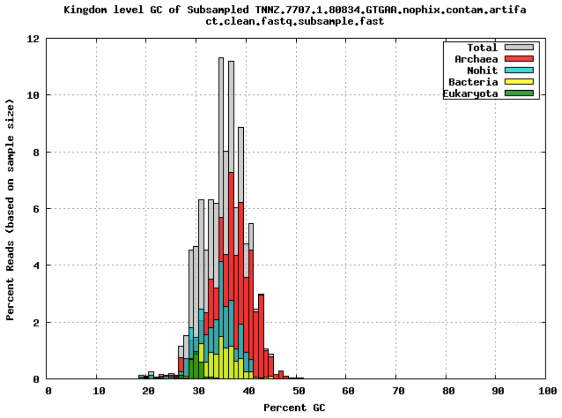
Description	Num Reads	Pct Reads
Input	7,287,298	100
Contam identified	60	0.0

List of Contaminants Identified

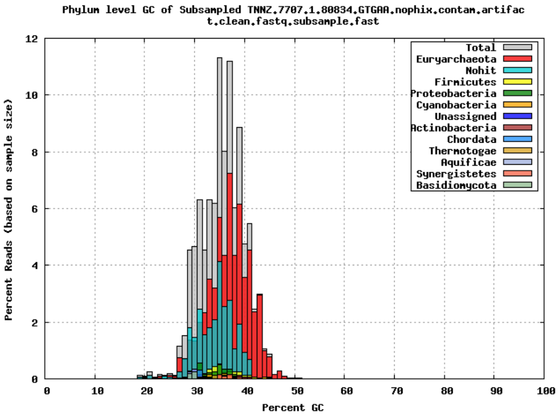
Description	Num Reads	Pct Reads
Escherichia	52	0.00
Shigella	6	0.00
Ralstonia	2	0.00

GC histogram of the reads subsampled to 10k, overlaid with GC of hits based on BLASTX, shown for different taxonomic levels.

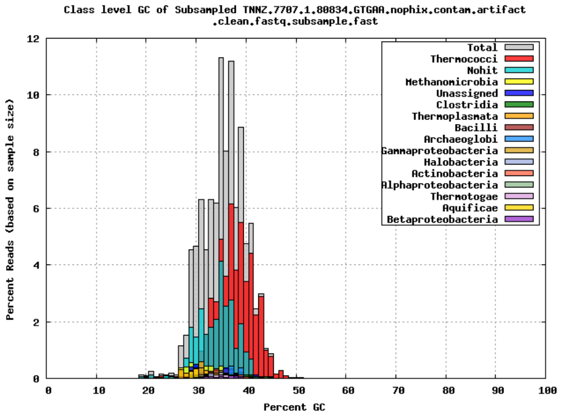
Kingdom



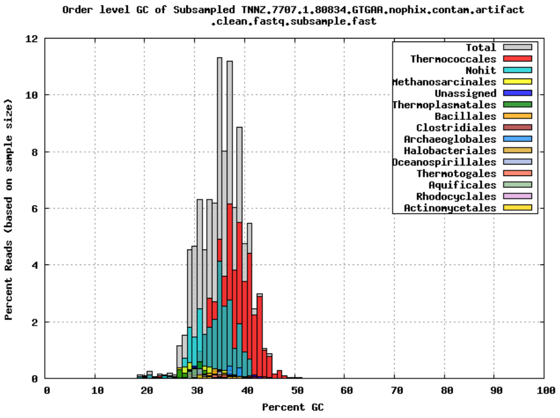
Phylum



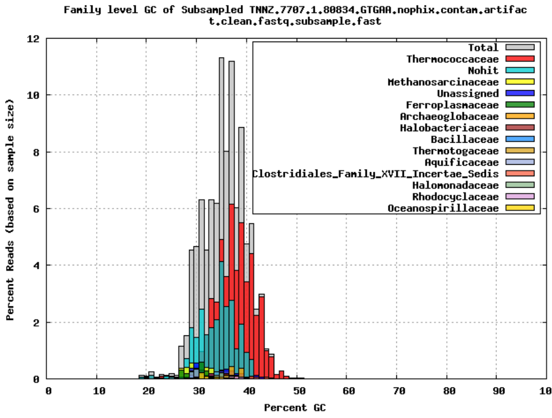
Class



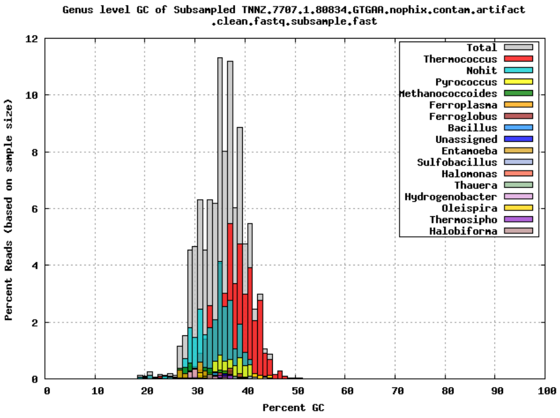
Order



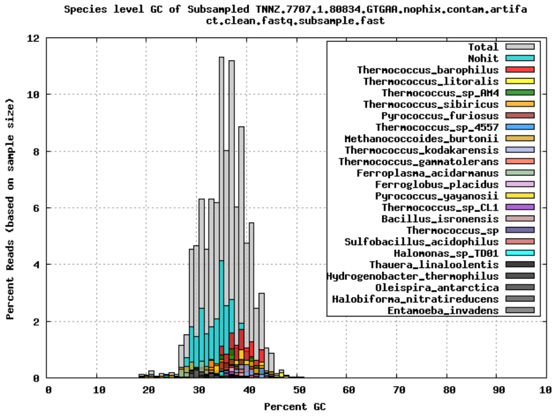
Family



Genus



Species



3. Assembly Statistics

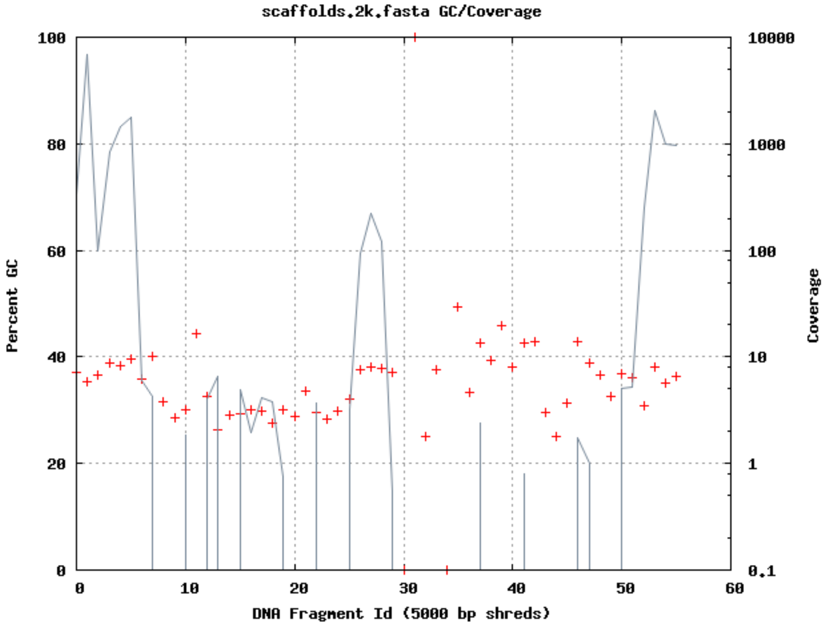
Assembly method	SPAdes
Scaffold total	12
Contig total	50
Scaffold sequence length	65.0 kb
Contig sequence length	65.0 kb (0.1% gap)
Scaffold N/L50	3/8.1 kb
Contig N/L50	3/8.1 kb
Largest Contig	15.6 kb
Number of scaffolds >50 kb	0
Pct of genome in scaffolds >50 kb	0.0

4. Assembly QC Results

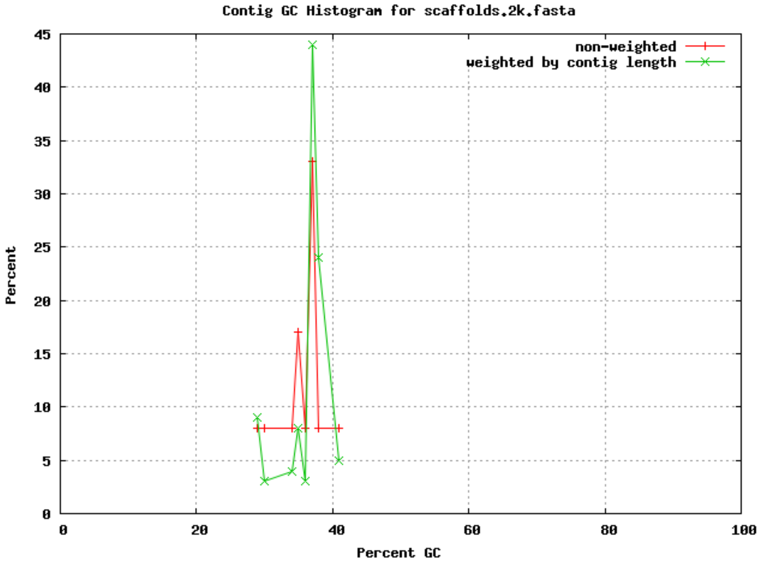
GC vs coverage based on GC of NCBI nt and Greengenes 16S rRNA gene hits to the assembly using megablast, shown for different taxonomic levels.

No hits found.

Coverage vs GC. Contigs were shredded into non-overlapping 5kbp and the GC of each shred was plotted as a point, colored by scaffold id. Coverage was calculated by mapping the fragment library to the final asssembly and plotted as connected points.



GC histogram of the contigs, including contig length weighted distribution.

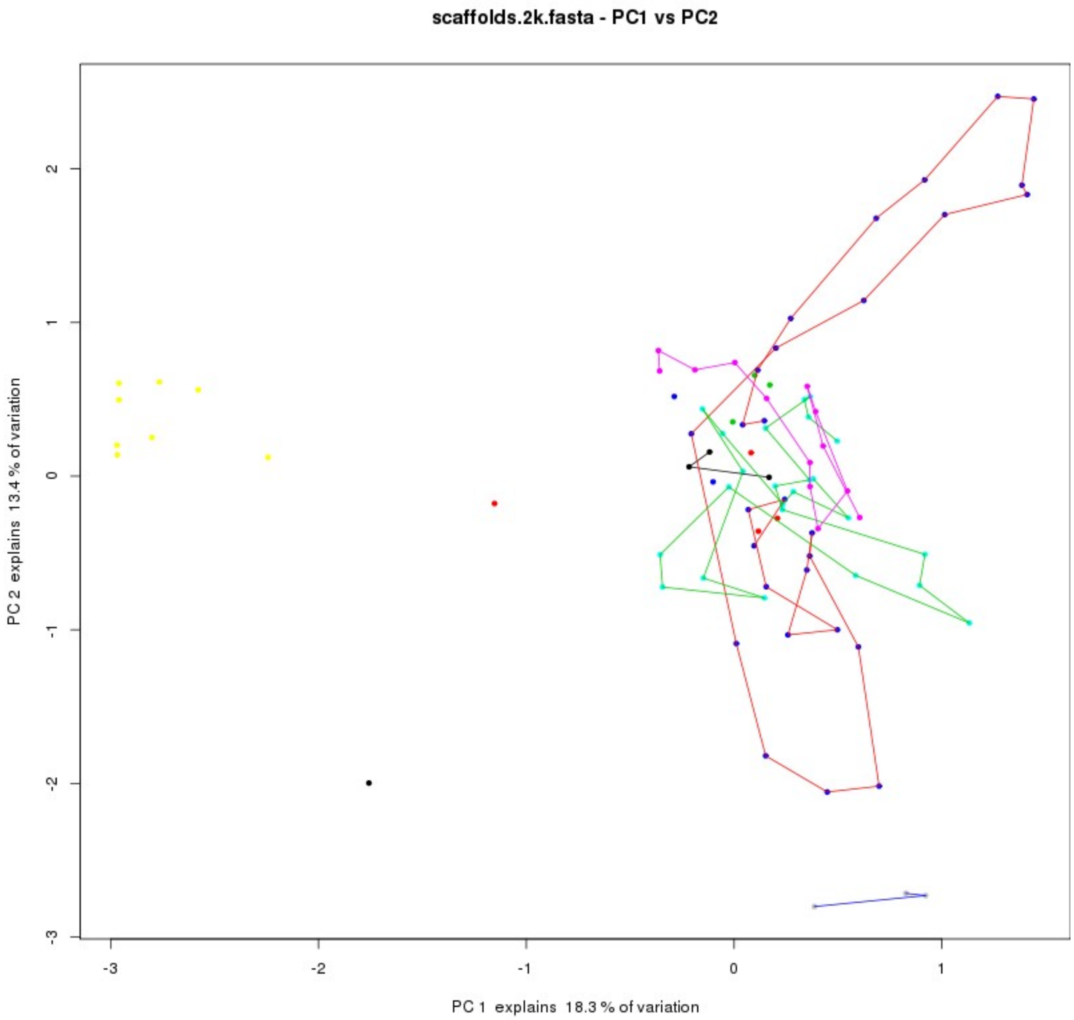


List of contigs and average percent GC, grouped in bins of 5:

Pct GC Bin	Contig Name
25	NODE_4_length_5847_cov_5.13553_ID_21688097
30	NODE_10_length_2363_cov_4.08016_ID_22695669, NODE_11_length_2180_cov_422.553_ID_29029524

35	NODE_1_length_15588_cov_714.003_ID_28956096, NODE_2_length_13854_cov_122.426_ID_23597839, NODE_3_length_8066_cov_1353.07_ID_24075324, NODE_6_length_3238_cov_347.547_ID_29029170, NODE_7_length_3187_cov_4.33269_ID_23411286, NODE_8_length_2517_cov_5.0723_ID_29017804, NODE_9_length_2509_cov_5979.98_ID_29022048, NODE_12_length_2148_cov_884.403_ID_28956250
40	NODE_5_length_3503_cov_2.24362_ID_21735021

Principal component analysis of tetramer frequencies of contigs. Detectable variations are highlighted in color.



Estimated genome recovery derived from analysis of universal single-copy genes detected in final assembly.

HMM	Pct Recovered
bacteria	0.8 %
archaea	0.69 %