phixscreen_jigsaw

Jigsaw report 02/13/2014

1. Read Statistics

Illumina Std PE Statistics

File name	TNNW.7707.1.80834.CCGTC.nophix.contam.artifact.clean.fastq
Number of reads	5,822,354
Sequencing depth †	292X
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	6,083,750	100
Contam removed	158	0.0
Artifact removed	261,238	4.3
Total removed	261,396	4.3
Total remaining	5,822,354	95.7

List of Contaminants Removed

Description	Num Reads	Pct Reads
gi|357579577|Canis_lupus_familiaris_chr3	138	0.00
human_chr2	116	0.00
human_chr15	16	0.00
human_chr16	2	0.00
human_chr4	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

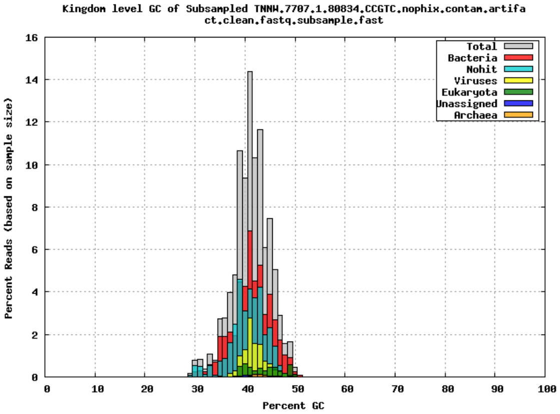
Description	Num Reads	Pct Reads
Input	6,083,750	100
Contam identified	156	0.0

List of Contaminants Identified

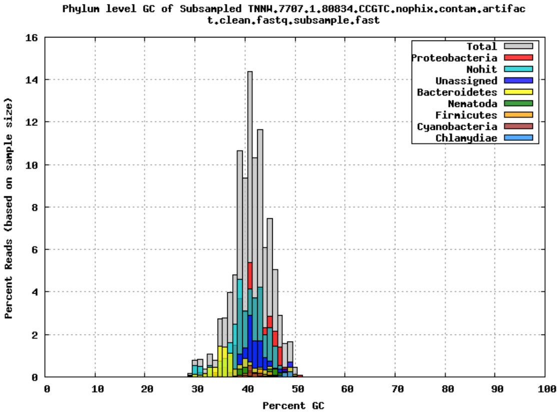
Description	Num Reads	Pct Reads
Escherichia	124	0.00
Shigella	32	0.00

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

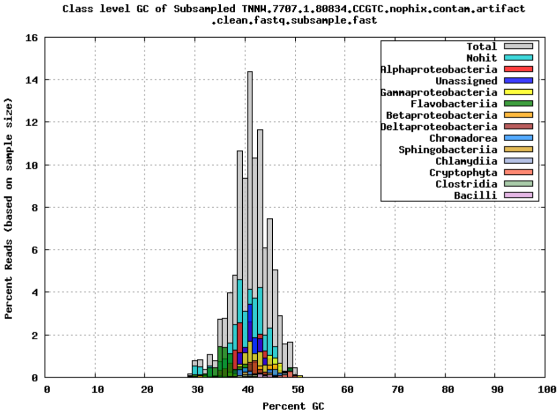
Kingdom



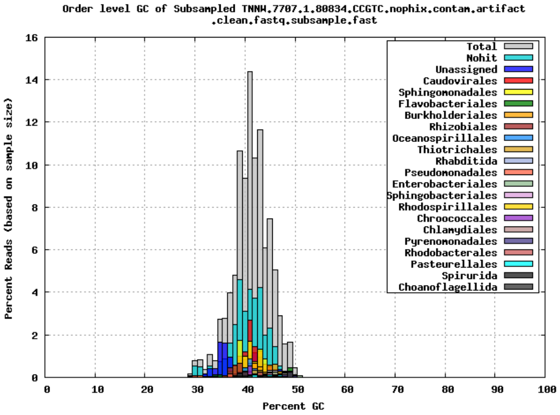
Phylum



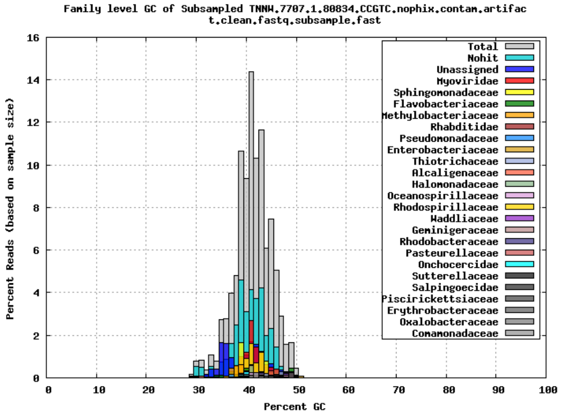
Class



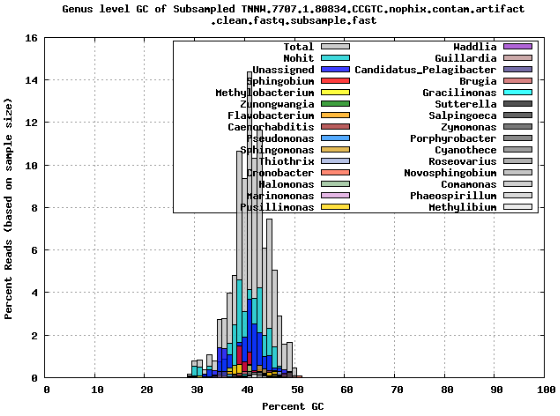
Order



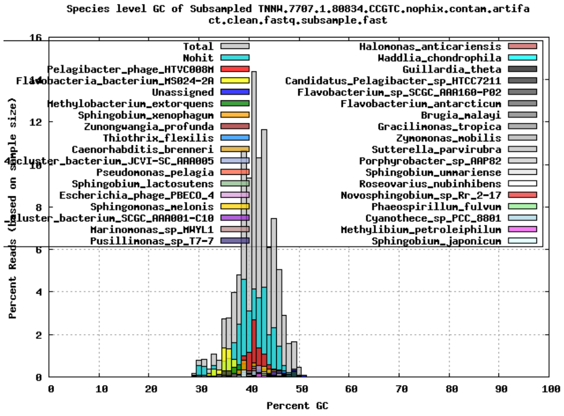
Family



Genus



Species



3. Assembly Statistics

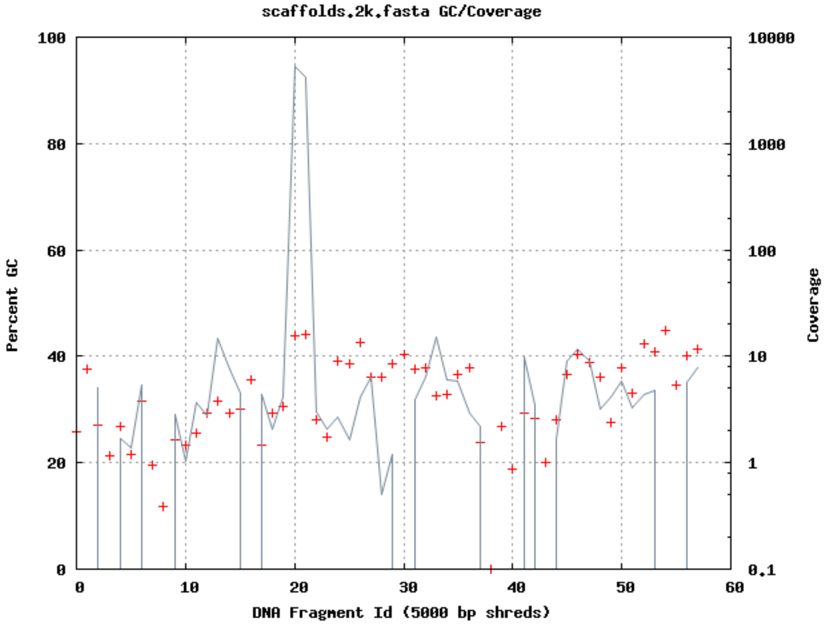
Assembly method	SPAdes
Scaffold total	17
Contig total	54
Scaffold sequence length	83.4 kb
Contig sequence length	83.3 kb (0.0% gap)
Scaffold N/L50	4/8.8 kb
Contig N/L50	8/2.8 kb
Largest Contig	15.3 kb
Number of scaffolds >50 kb	0
Pct of genome in scaffolds >50 kb	0.0

4. Assembly QC Results

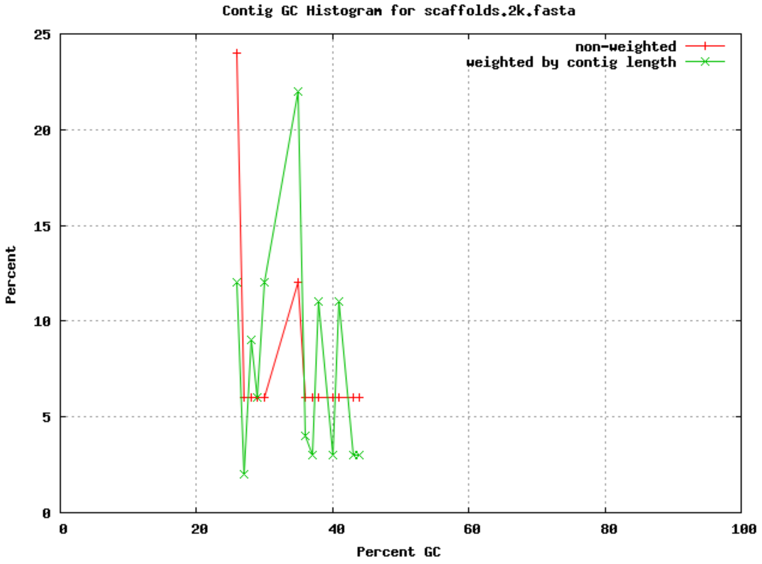
GC vs coverage based on GC of NCBI nt and Greengenes 16S rRNA gene hits to the assembly using megablast, shown for different taxonomic levels.

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_5_length_7658_cov_5.63738_ID_20599558, NODE_6_length_4855_cov_3.63771_ID_20599560, NODE_9_length_2902_cov_2.7752_ID_14802298, NODE_10_length_2558_cov_3.39233_ID_10506577, NODE_13_length_2389_cov_2.43016_ID_6720931, NODE_16_length_2102_cov_2.12702_ID_13742346

	NODE_17_length_2069_cov_2.87934_ID_12147015
30	NODE_2_length_9835_cov_8.38732_ID_12494981
35	NODE_1_length_15343_cov_6.34066_ID_14242651, NODE_3_length_9236_cov_7.34125_ID_13828858, NODE_7_length_3222_cov_4.40006_ID_6804787, NODE_8_length_2982_cov_4.94807_ID_14055459 NODE_15_length_2200_cov_2.39441_ID_14169452
40	NODE_4_length_8805_cov_5.09783_ID_14782348, NODE_11_length_2514_cov_7873.85_ID_20651256, NODE_12_length_2463_cov_3.00374_ID_7447577, NODE_14_length_2241_cov_7919.28_ID_20647746

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