

1. Read Statistics

Illumina Std PE Statistics

File name	AD-726-N03_pe.fastq
Number of reads	8,406,756
Sequencing depth [†]	234X
Read type	2x139 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 (human, dog, cat) and artifact sequences. Pairs of matching reads ($\geq 95\%$ ID) were removed from the dataset and subsampled to 20,000,000 reads.

Illumina Std PE Read Filter Statistics

Description	Num Reads	Pct Reads
Input	8,406,756	100
Contam removed	0	0.0
Artifact removed	2,078	0.0
Total removed	2,078	0.0
Total remaining	8,404,678	100.0

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	8,406,756	100
Contam identified	2	0.0

List of Contaminants Identified

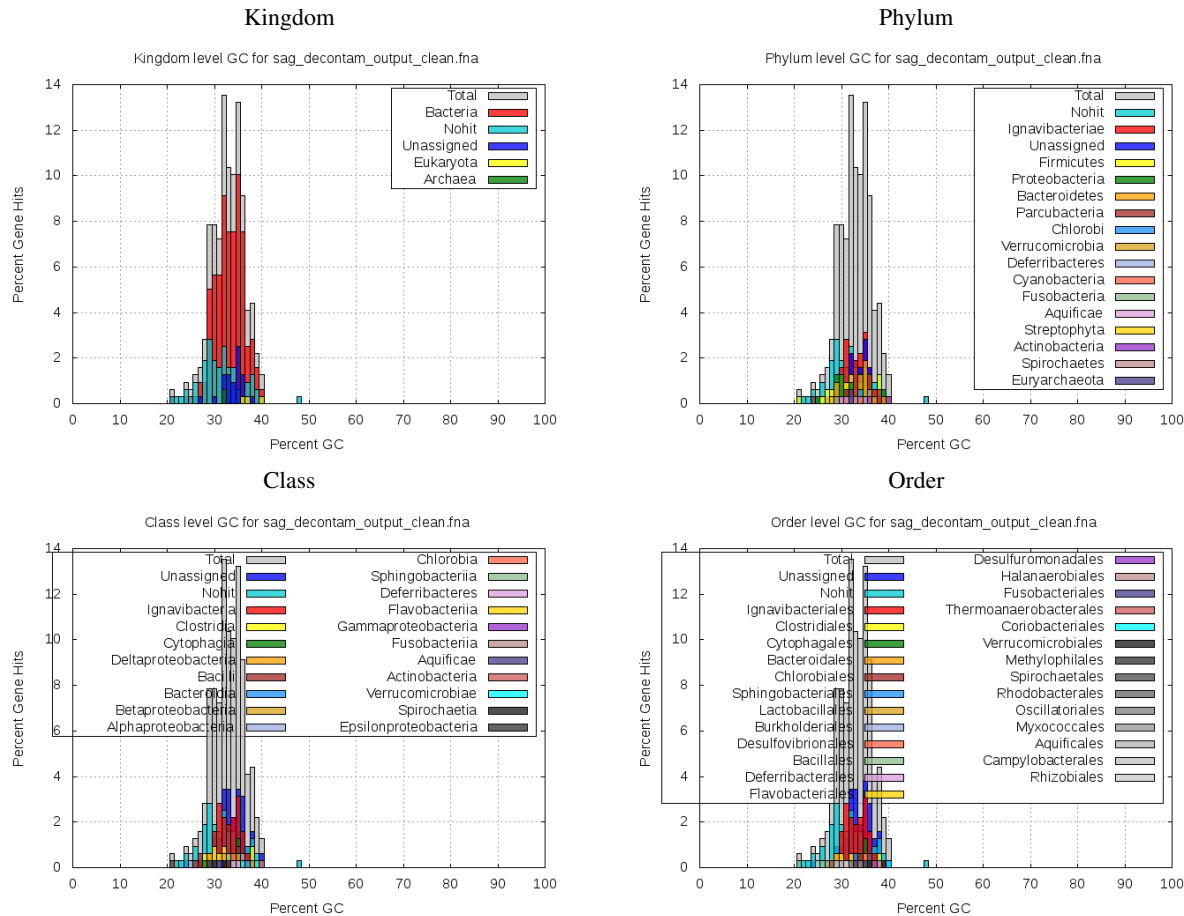
Description	Num Reads	Pct Reads
<i>Ralstonia</i>	2	0.00

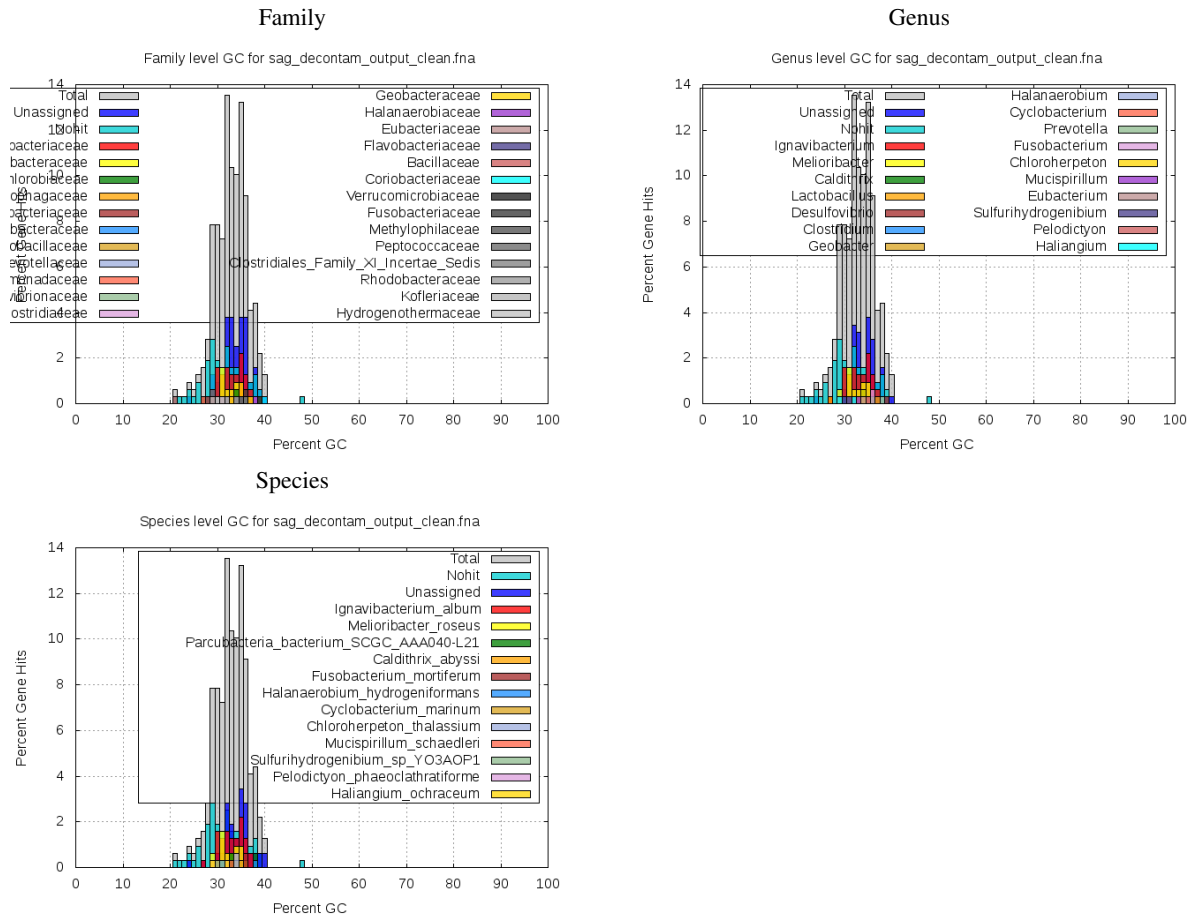
3. Screened Assembly Statistics

Assembly method	SPAdes with auto decontamination (ProDeGe)
Scaffold total	16
Contig total	16
Scaffold sequence length	0.329 Mb
Contig sequence length	0.329 Mb 0.000% gap
Scaffold N/L50	6/21.652 kb
Contig N/L50	6/21.652 kb
Largest Contig	38.1 kb
Number of scaffolds >50 kb	0
Pct of genome in scaffolds >50 kb	0.00
Pct of reads assembled (raw)	82.7
Pct of reads assembled (decontam)	79.4

4. Screened Assembly QC Results

GC histogram of the predicted genes on each scaffold, overlaid with GC of hits based on LAST, shown for different taxonomic levels.

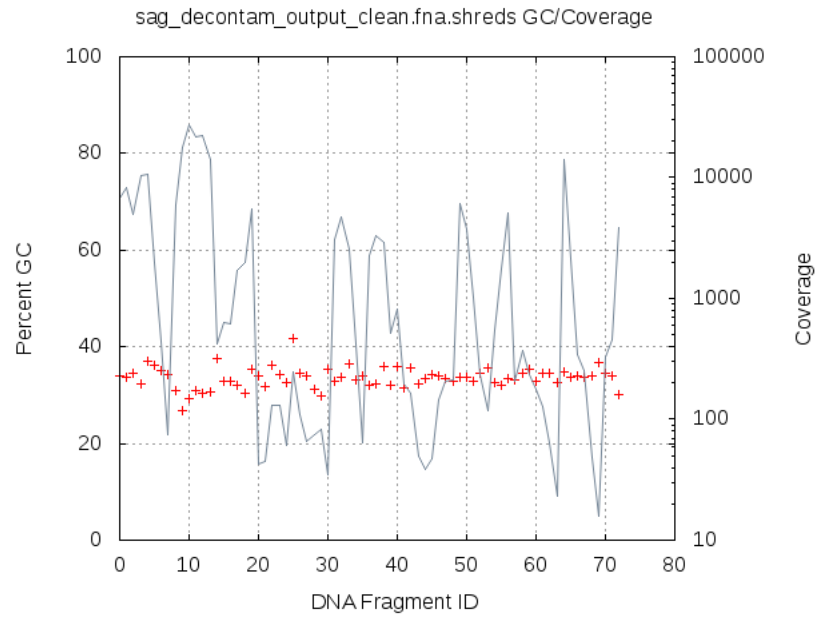




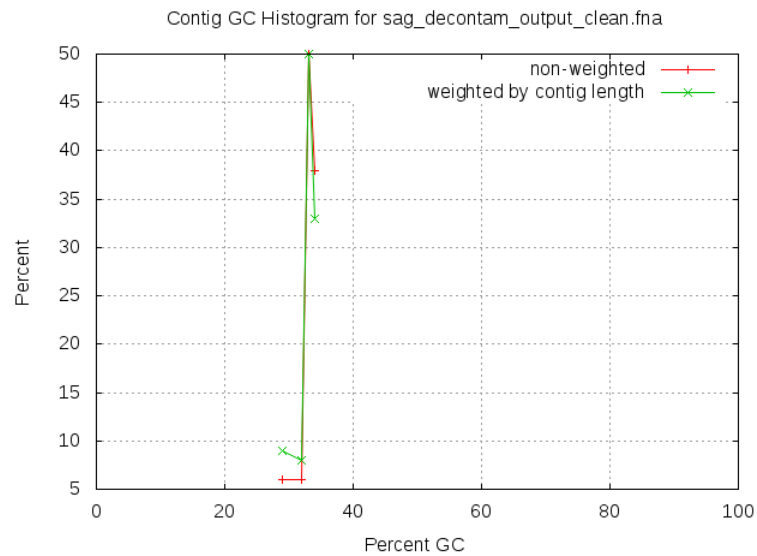
GC vs coverage based on GC of NCBI nt and Greengenes 16S rRNA gene hits to the assembly using megablast, shown for different taxonomic levels.

No hits found.

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



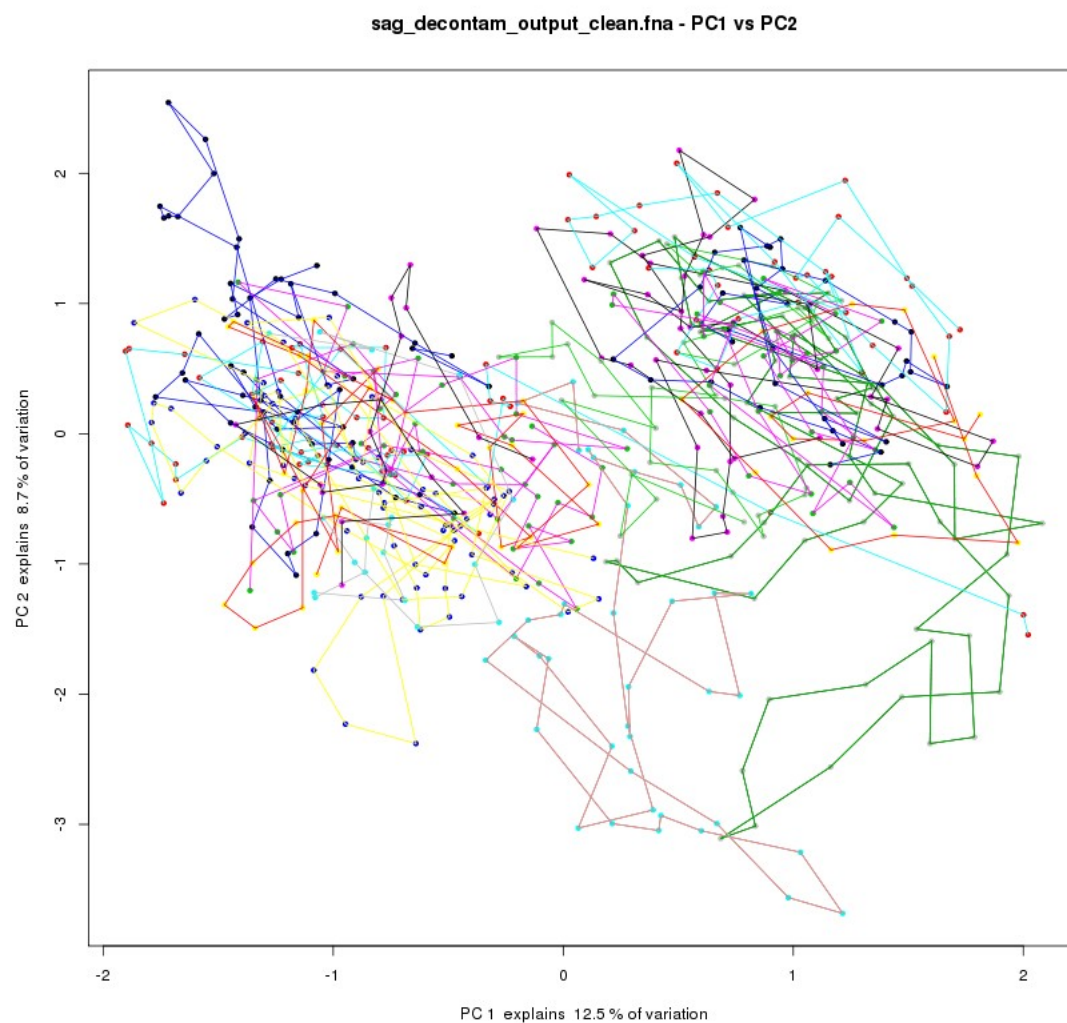
GC histogram of the scaffolds, including scaffold length weighted distribution.



List of the top scaffold megablast hits against 16S ribosomal RNA genes.

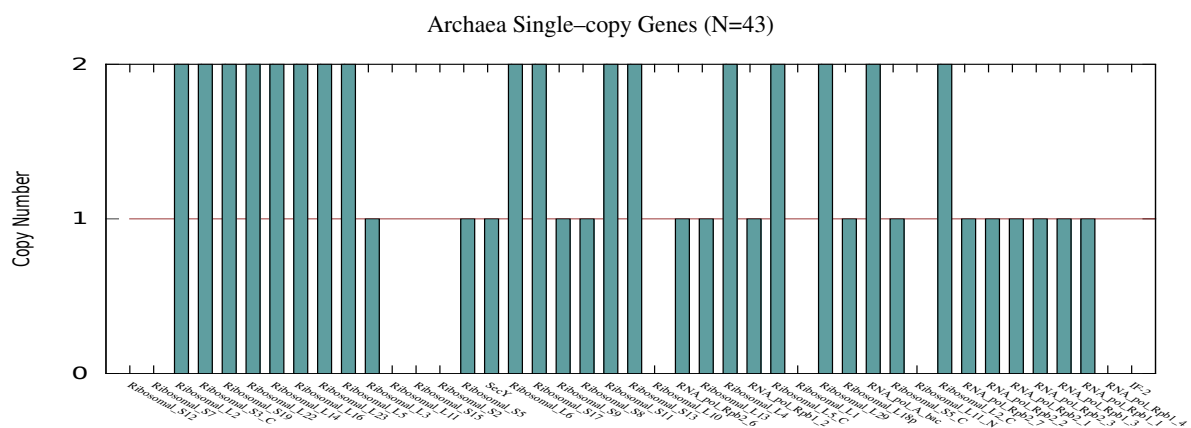
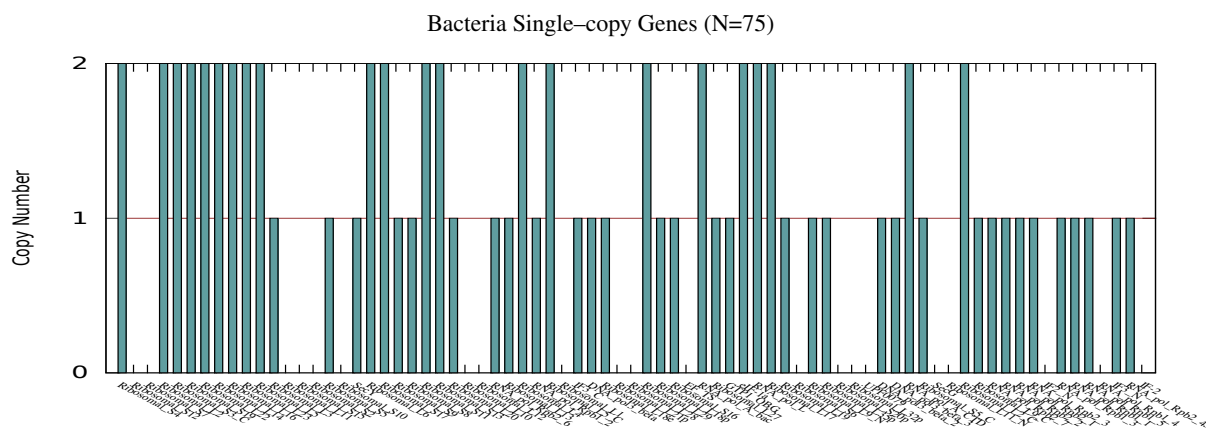
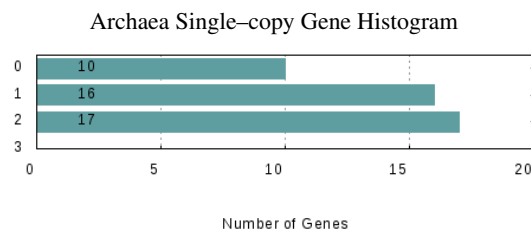
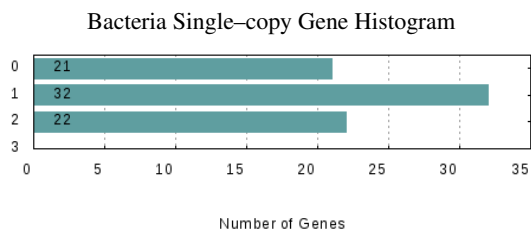
Organism	Align Length (bp)	Pct Id	Contig Name
<i>AB193914.1.1558.Bacteria;Candidate_division_OD1;uncultured_bacterium</i>	108	96.30	AD-726-N03_NODE_3.length.27243_cov.50.6775.ID_5

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



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

HMM	Pct Recovered
archaea	85.27 %
bacteria	80 %

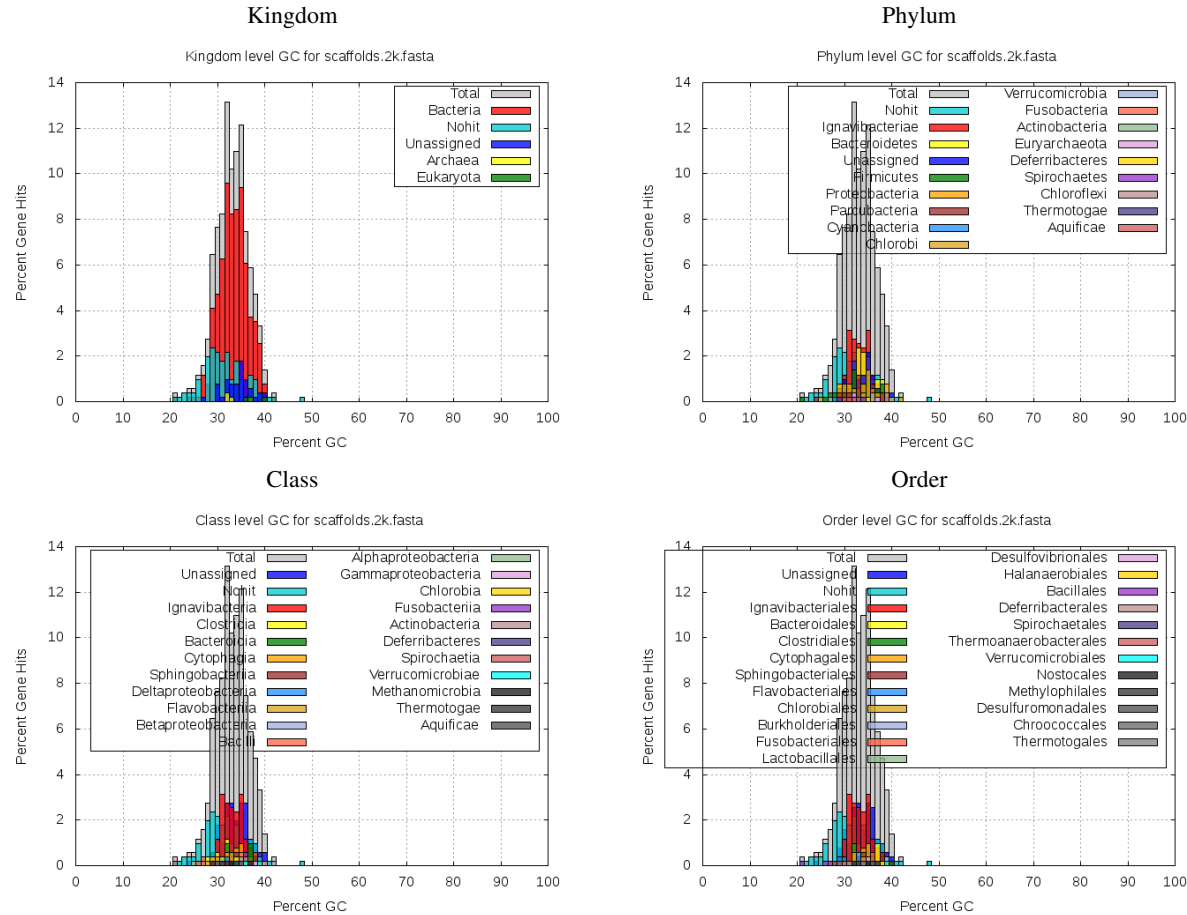


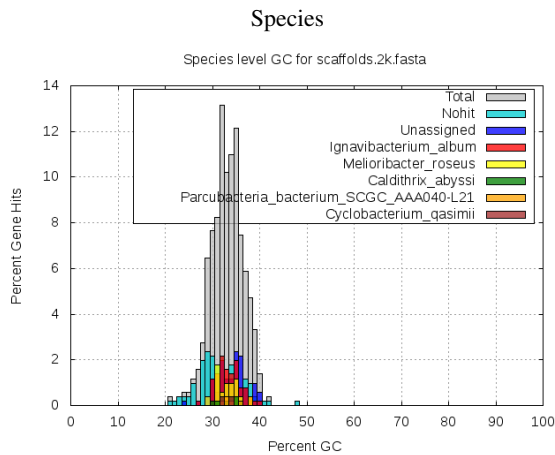
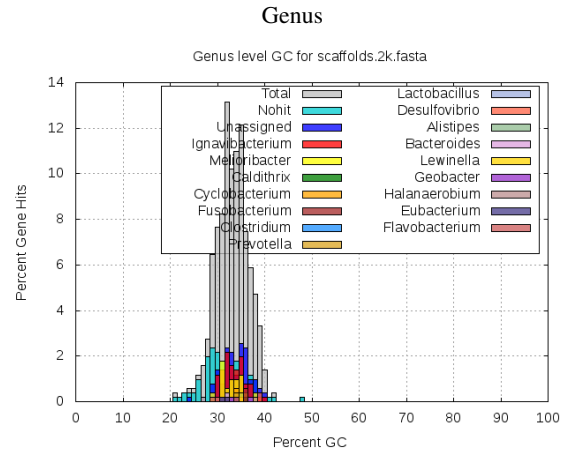
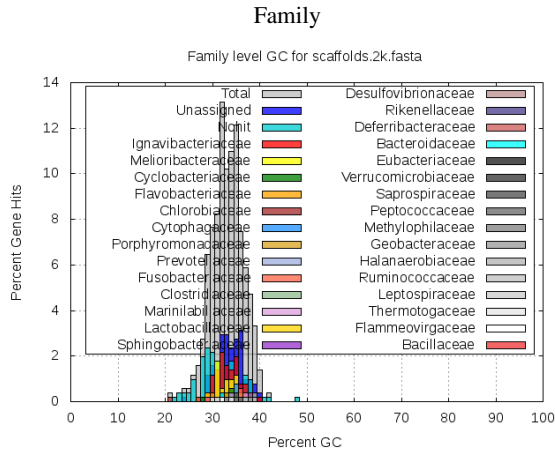
5. Unscreened Assembly Statistics

Assembly method	SPAdes
Scaffold total	63
Contig total	63
Scaffold sequence length	0.523 Mb
Contig sequence length	0.523 Mb 0.000% gap
Scaffold N/L50	11/18.447 kb
Contig N/L50	11/18.447 kb
Largest Contig	38.1 kb
Number of scaffolds >50 kb	0
Pct of genome in scaffolds >50 kb	0.00
Pct of reads assembled	82.7

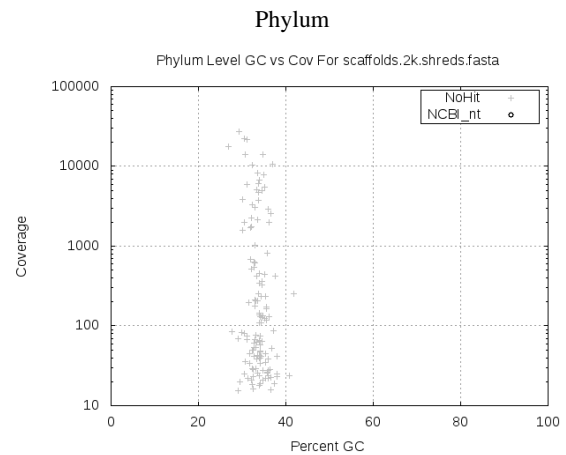
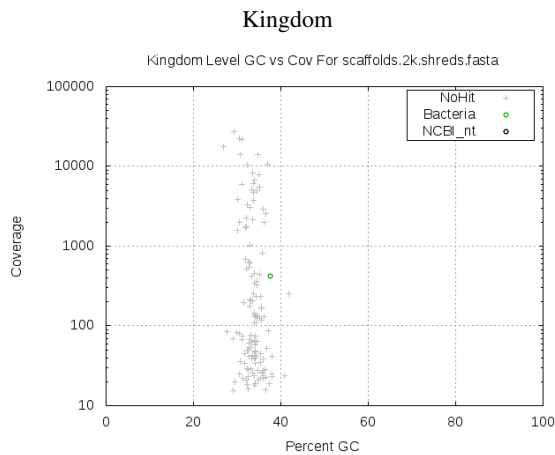
6. Unscreened Assembly QC Results

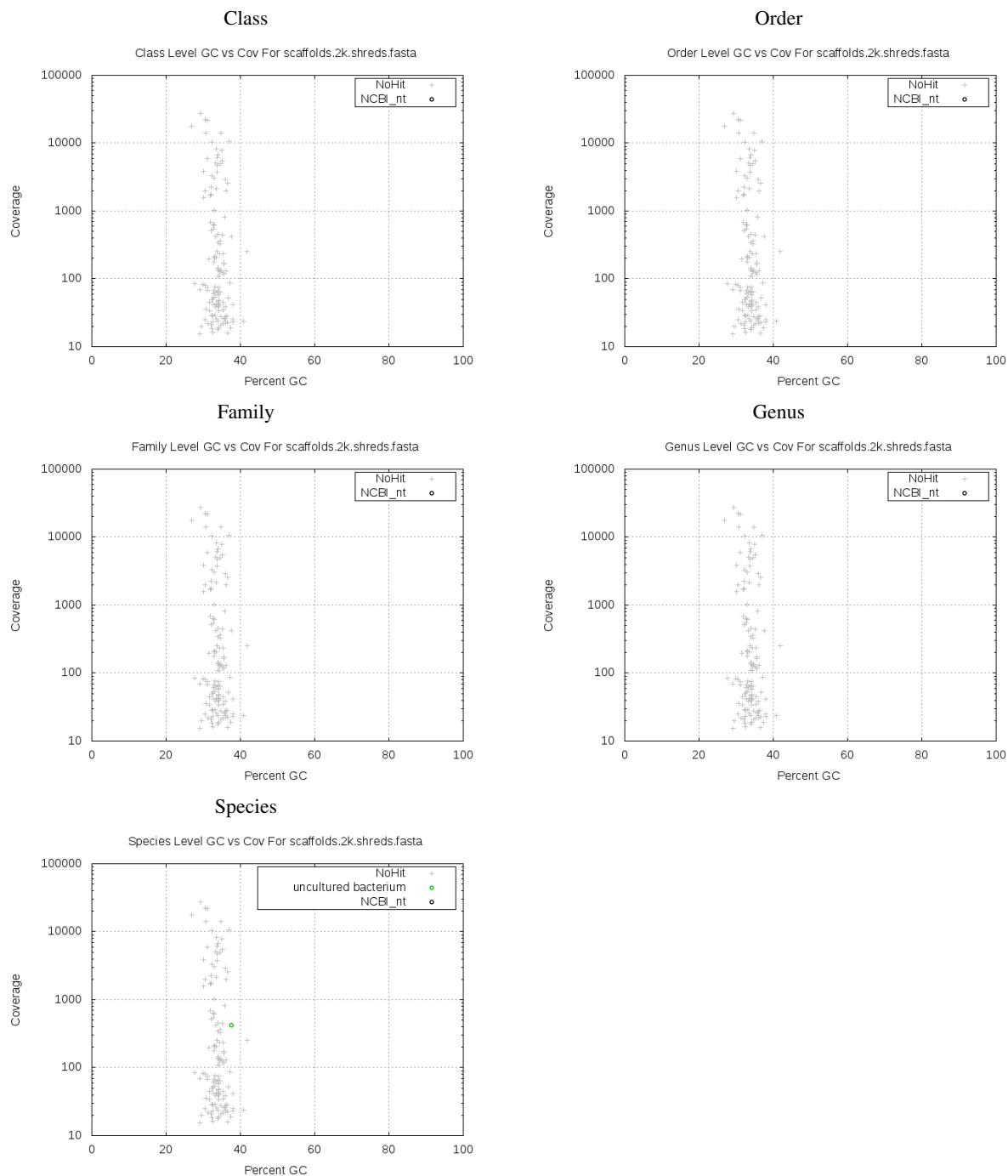
GC histogram of the predicted genes on each scaffold, overlaid with GC of hits based on LAST, shown for different taxonomic levels.



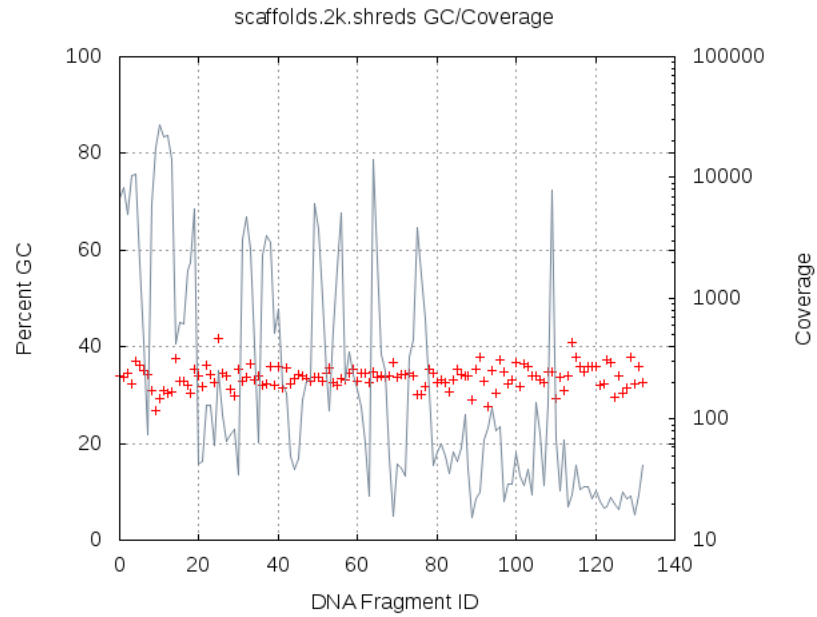


GC vs coverage based on GC of NCBI nt and Greengenes 16S rRNA gene hits to the assembly using megablast, shown for different taxonomic levels.

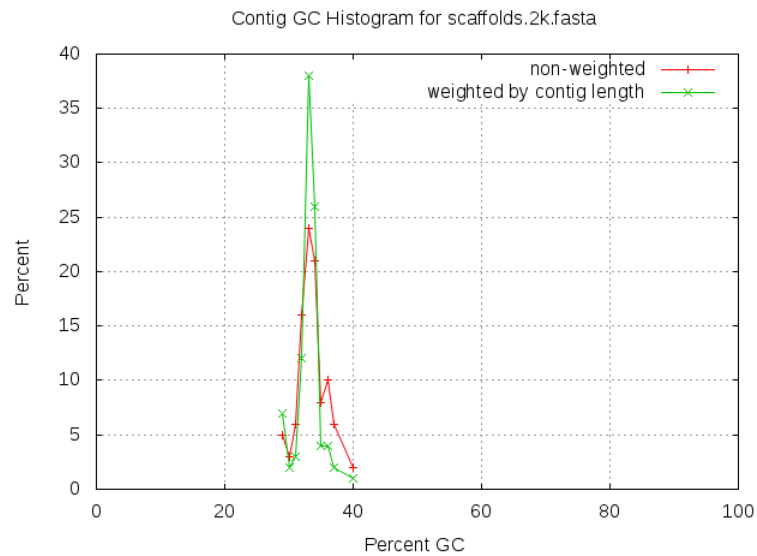




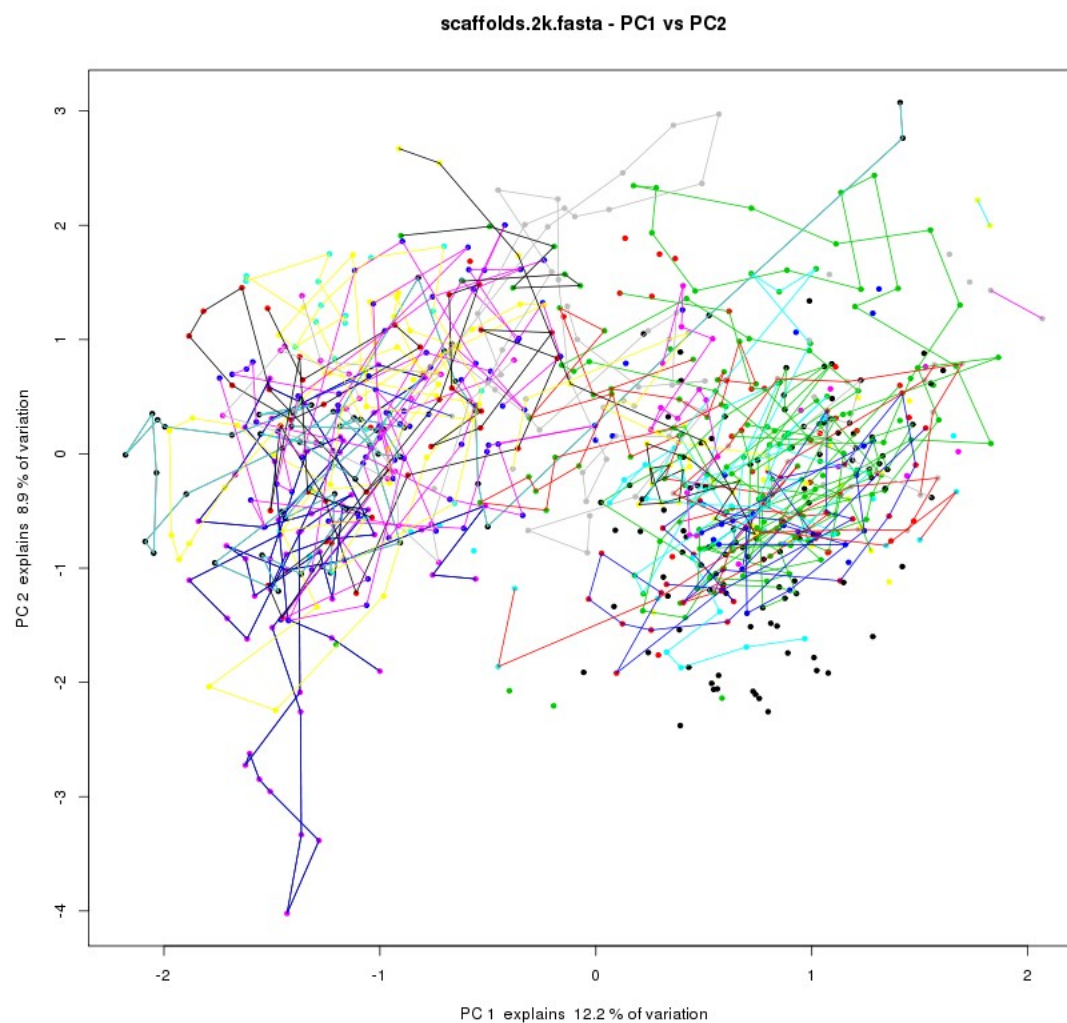
Coverage vs GC. Scaffolds were shredded into non-overlapping 5 kbp fragments and the GC content of each shred was plotted as a point, colored by scaffold id. Coverage was calculated by mapping the fragment library to the final assembly and plotted as connected points.



GC histogram of the scaffolds, including scaffold length weighted distribution.



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



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

HMM	Pct Recovered
archaea	87.86 %
bacteria	82.96 %

