

1. Project Information

Program	Microbial
JGI Sequencing Project ID	1042295
Sequencing Project Name	Nostoc sp. sp2

2. Read Statistics

	Raw Reads	Filtered SubReads	Error Corrected Reads
Reads	1,492,529	681,096	10,127
Bases	5,246,454,896	2,719,947,908	111,422,226
Avg Read Length	3,515.1 +/- 3,817.1	3,993.5 +/- 2,957.1	11,002.5 +/- 6,130.9
Reads >5 kbp	324,238	154,974	7,509
Bases, reads >5 kbp	3,017,401,994	1,301,770,773	106,693,466
Avg Read Length, reads >5 kbp	9,306.1 +/- 4,114.9	8,399.9 +/- 3,112.7	14,208.7 +/- 3,216.9

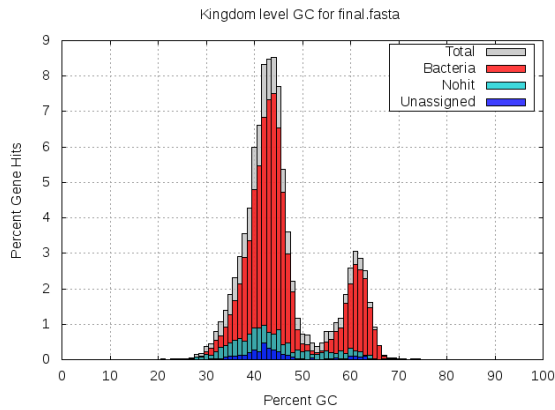
3. Assembly Statistics

Scaffold total	286
Contig total	286
Scaffold sequence length	18.430 mb
Contig sequence length	18.430 mb 0.000% gap
Scaffold N/L50	11/302.205 kb
Largest Contig	4,101.0 kbp
Number of scaffolds >50 kb	52
Pct of genome in scaffolds >50 kb	76.77%

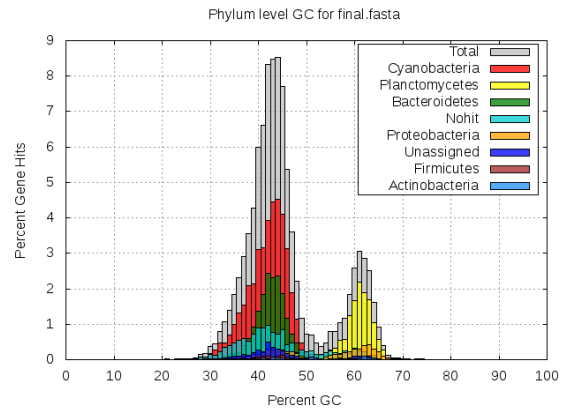
4. Assembly QC Results

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

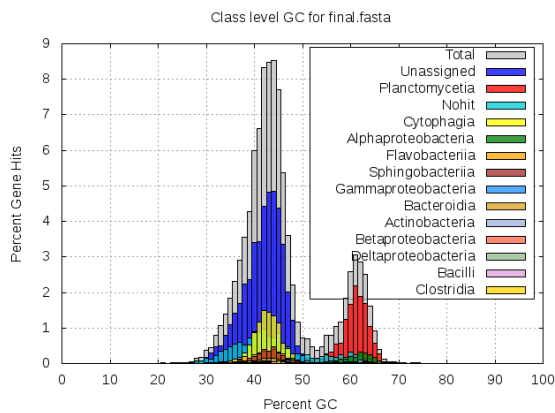
Kingdom



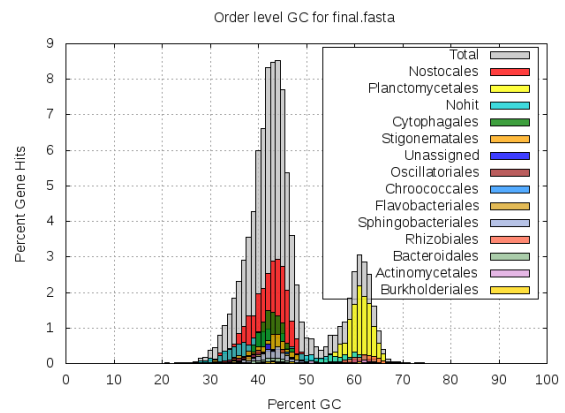
Phylum



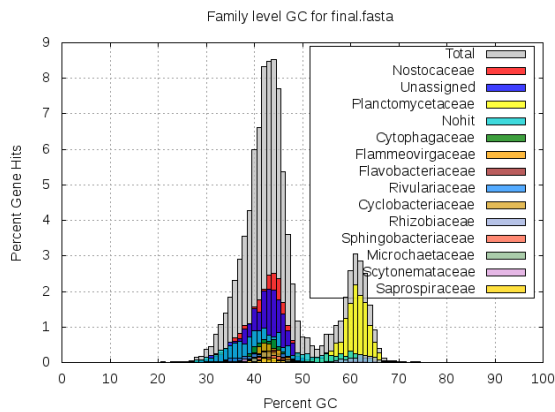
Class



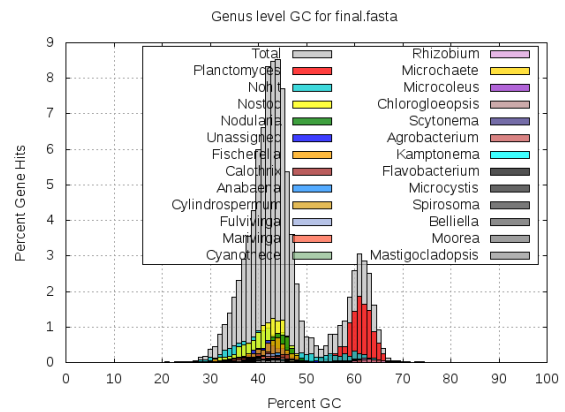
Order

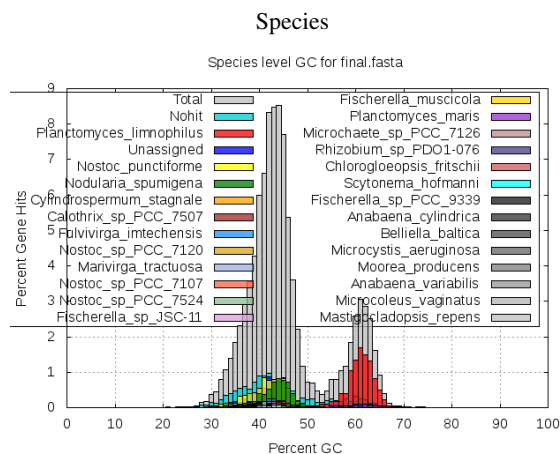


Family

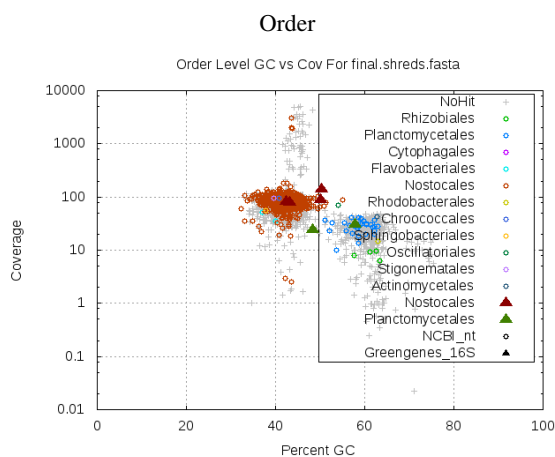
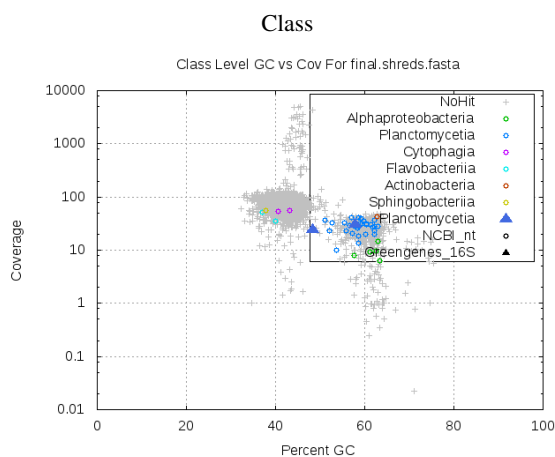
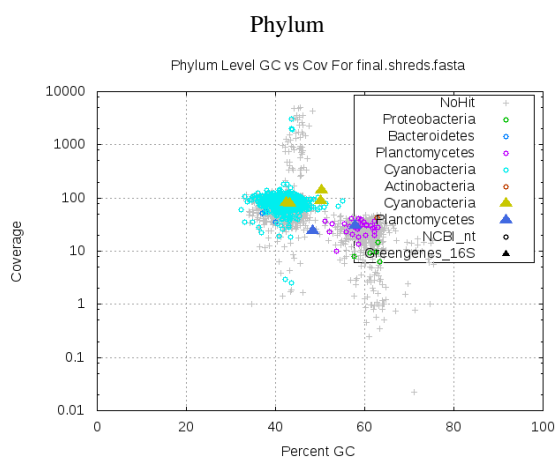
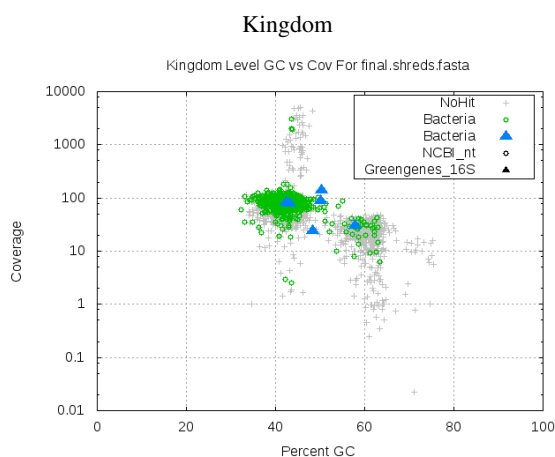


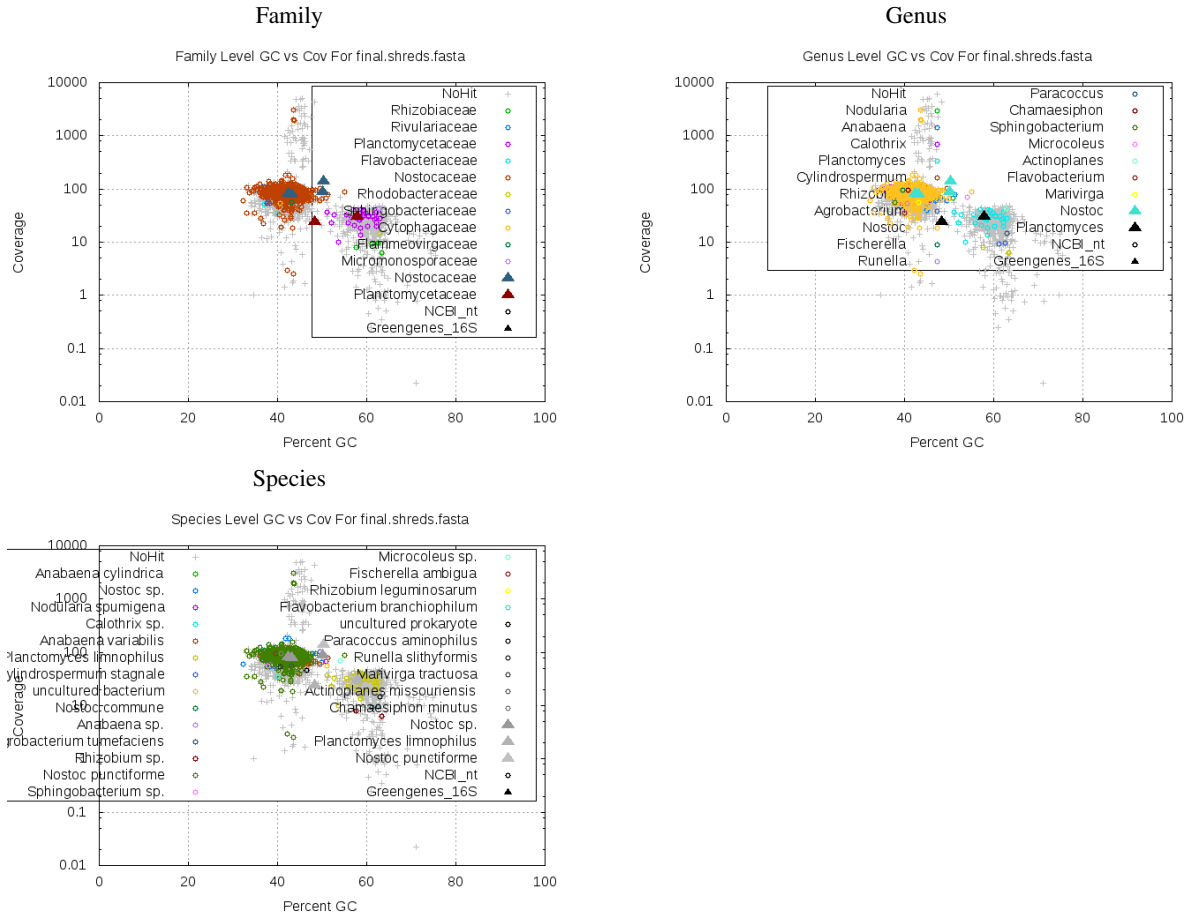
Genus



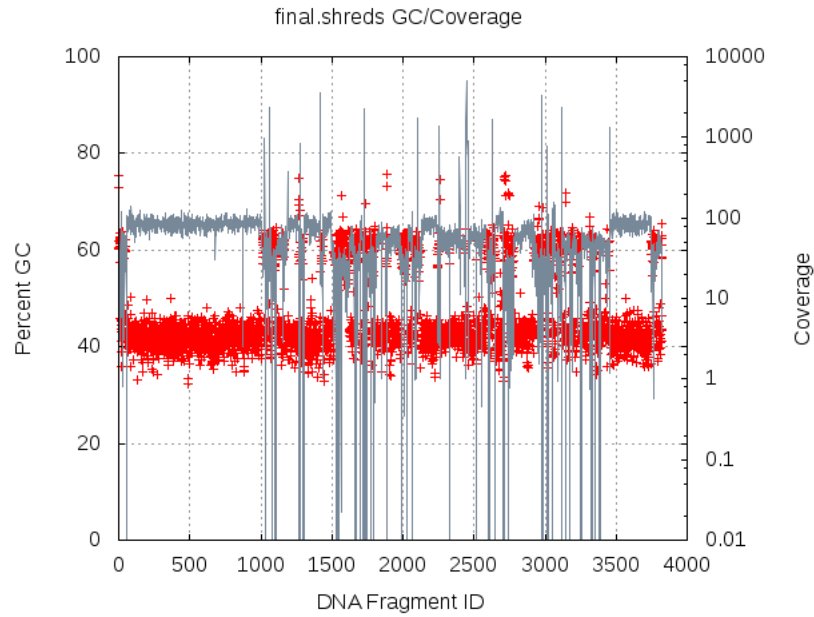


GC vs coverage based on GC of NCBI nt and Greengenes 16S rRNA gene hits to the assembly using megablast, shown for different taxonomic levels.

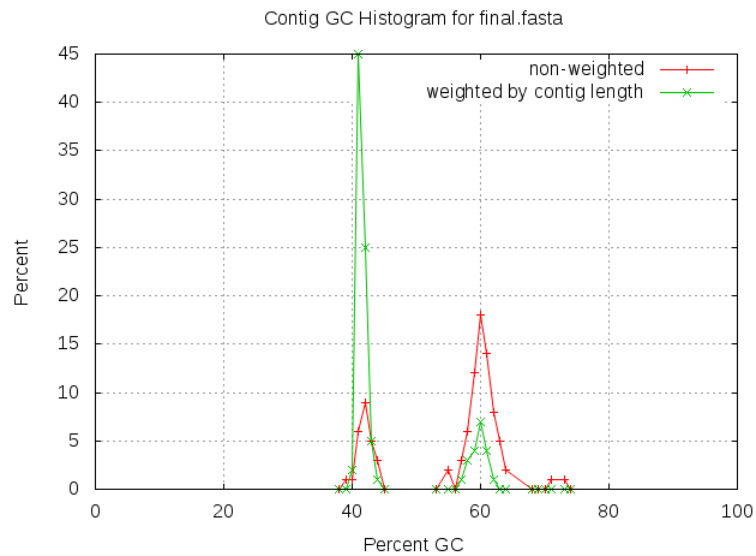




Coverage vs GC. Contigs were shredded into non-overlapping 5kbp and the GC of each shred was plotted as a point, colored by scaffold id. Coverage was calculated by mapping the fragment library to the final assembly and plotted as connected points.



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



List of contigs and average percent GC bin:

Pct GC Bin	Contig Name
35	unitig_8 quiver, unitig_68 quiver, unitig_98 quiver
40	unitig_0 quiver, unitig_1 quiver, unitig_2 quiver, unitig_3 quiver, unitig_4 quiver, unitig_5 quiver, unitig_6 quiver, unitig_7 quiver, unitig_9 quiver, unitig_10 quiver, unitig_11 quiver, unitig_12 quiver, unitig_13 quiver, unitig_14 quiver, unitig_15 quiver, unitig_16 quiver, unitig_17 quiver, unitig_18 quiver, unitig_19 quiver, unitig_20 quiver, unitig_21 quiver, unitig_23 quiver, unitig_24 quiver, unitig_25 quiver, unitig_26 quiver, unitig_27 quiver, unitig_28 quiver, unitig_29 quiver, unitig_30 quiver, unitig_31 quiver, unitig_32 quiver, unitig_33 quiver, unitig_34 quiver, unitig_35 quiver, unitig_36 quiver, unitig_37 quiver, unitig_38 quiver, unitig_39 quiver, unitig_40 quiver, unitig_41 quiver, unitig_42 quiver, unitig_43 quiver, unitig_44 quiver, unitig_45 quiver, unitig_46 quiver, unitig_47 quiver,

	unitig_48 quiver, unitig_49 quiver, unitig_50 quiver, unitig_51 quiver, unitig_54 quiver, unitig_56 quiver, unitig_62 quiver, unitig_63 quiver, unitig_76 quiver, unitig_77 quiver, unitig_82 quiver, unitig_83 quiver, unitig_87 quiver, unitig_115 quiver, unitig_171 quiver, unitig_188 quiver, unitig_191 quiver, unitig_839 quiver, unitig_840 quiver, unitig_842 quiver, unitig_843 quiver, unitig_844 quiver, unitig_846 quiver, unitig_847 quiver, unitig_848 quiver, unitig_851 quiver
45	unitig_841 quiver
50	unitig_493 quiver
55	unitig_52 quiver, unitig_55 quiver, unitig_67 quiver, unitig_70 quiver, unitig_71 quiver, unitig_75 quiver, unitig_78 quiver, unitig_79 quiver, unitig_80 quiver, unitig_81 quiver, unitig_88 quiver, unitig_92 quiver, unitig_93 quiver, unitig_94 quiver, unitig_95 quiver, unitig_97 quiver, unitig_100 quiver, unitig_101 quiver, unitig_104 quiver, unitig_105 quiver, unitig_107 quiver, unitig_108 quiver, unitig_112 quiver, unitig_114 quiver, unitig_118 quiver, unitig_121 quiver, unitig_124 quiver, unitig_126 quiver, unitig_127 quiver, unitig_128 quiver, unitig_129 quiver, unitig_133 quiver, unitig_140 quiver, unitig_142 quiver, unitig_146 quiver, unitig_148 quiver, unitig_152 quiver, unitig_164 quiver, unitig_175 quiver, unitig_178 quiver, unitig_184 quiver, unitig_187 quiver, unitig_192 quiver, unitig_202 quiver, unitig_204 quiver, unitig_208 quiver, unitig_211 quiver, unitig_213 quiver, unitig_214 quiver, unitig_216 quiver, unitig_220 quiver, unitig_227 quiver, unitig_231 quiver, unitig_256 quiver, unitig_266 quiver, unitig_289 quiver, unitig_306 quiver, unitig_339 quiver, unitig_377 quiver, unitig_512 quiver, unitig_528 quiver, unitig_587 quiver, unitig_673 quiver, unitig_703 quiver unitig_704 quiver
60	unitig_53 quiver, unitig_57 quiver, unitig_58 quiver, unitig_59 quiver, unitig_60 quiver, unitig_61 quiver, unitig_64 quiver, unitig_65 quiver, unitig_66 quiver, unitig_69 quiver, unitig_72 quiver, unitig_73 quiver, unitig_74 quiver, unitig_84 quiver, unitig_85 quiver, unitig_86 quiver, unitig_89 quiver, unitig_90 quiver, unitig_91 quiver, unitig_96 quiver, unitig_99 quiver, unitig_102 quiver, unitig_103 quiver, unitig_106 quiver, unitig_109 quiver, unitig_110 quiver, unitig_111 quiver, unitig_113 quiver, unitig_116 quiver, unitig_117 quiver, unitig_119 quiver, unitig_120 quiver, unitig_122 quiver, unitig_123 quiver, unitig_125 quiver, unitig_130 quiver, unitig_131 quiver, unitig_132 quiver, unitig_134 quiver, unitig_135 quiver, unitig_136 quiver, unitig_137 quiver, unitig_138 quiver, unitig_139 quiver, unitig_141 quiver, unitig_147 quiver, unitig_149 quiver, unitig_151 quiver, unitig_153 quiver, unitig_155 quiver, unitig_156 quiver, unitig_157 quiver, unitig_158 quiver, unitig_159 quiver, unitig_160 quiver, unitig_161 quiver, unitig_162 quiver, unitig_163 quiver, unitig_166 quiver, unitig_167 quiver, unitig_168 quiver, unitig_169 quiver, unitig_170 quiver, unitig_172 quiver, unitig_174 quiver, unitig_176 quiver, unitig_177 quiver, unitig_179 quiver, unitig_180 quiver, unitig_181 quiver, unitig_182 quiver, unitig_183 quiver, unitig_185 quiver, unitig_186 quiver, unitig_189 quiver, unitig_190 quiver, unitig_194 quiver, unitig_195 quiver, unitig_196 quiver, unitig_197 quiver, unitig_198 quiver, unitig_199 quiver, unitig_200 quiver, unitig_201 quiver, unitig_203 quiver, unitig_205 quiver, unitig_206 quiver, unitig_207 quiver, unitig_209 quiver, unitig_210 quiver, unitig_212 quiver, unitig_215 quiver, unitig_217 quiver, unitig_218 quiver, unitig_219 quiver, unitig_221 quiver, unitig_222 quiver, unitig_223 quiver, unitig_224 quiver, unitig_225 quiver, unitig_226 quiver, unitig_228 quiver, unitig_263 quiver, unitig_290 quiver, unitig_309 quiver, unitig_333 quiver, unitig_355 quiver, unitig_378 quiver, unitig_379 quiver, unitig_439 quiver, unitig_441 quiver, unitig_485 quiver, unitig_545 quiver, unitig_591 quiver, unitig_617 quiver, unitig_618 quiver, unitig_620 quiver, unitig_659 quiver, unitig_701 quiver, unitig_715 quiver, unitig_719 quiver, unitig_728 quiver, unitig_746 quiver, unitig_757 quiver, unitig_776 quiver, unitig_781 quiver, unitig_782 quiver, unitig_784 quiver, unitig_808 quiver, unitig_810 quiver, unitig_832 quiver, unitig_833 quiver, unitig_836 quiver
65	unitig_173 quiver, unitig_821 quiver
70	unitig_143 quiver, unitig_144 quiver, unitig_145 quiver, unitig_150 quiver, unitig_154 quiver, unitig_193 quiver, unitig_229 quiver, unitig_391 quiver, unitig_452 quiver

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

Organism	Align Length (bp)	Pct Id	Contig Name
244486 <i>Nostoc sp. KVI20 EU022731.1</i>	1,482	99.60	unitig_1 quiver
2845 <i>Planctomyces limnophilus X62911.1</i>	1,528	99.08	unitig_80 quiver

5. Data Access

The following sequence fasta files can be downloaded from our JGI portal website.

<http://www.jgi.doe.gov/genome-projects>

The annotation of the assembled contigs can be found within IMG.

<http://img.jgi.doe.gov>

6. Methods

Isolate Improved Draft

Genome sequencing and assembly

The draft genome of *Nostoc sp.* was generated at the DOE Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology [1]. A >10kpb Pacbio SMRTbell™ library was constructed and sequenced on the PacBio RS platform, which generated 681,096 filtered subreads totaling 2.7 Gbp. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov>. The raw reads were assembled using HGAP (version: 2.3.0, protocol version=2.3.0 method=RS_HGAP_Assembly.3, smrtpipe.py v1.87.139483,) [2]. The final draft assembly contained 286 contigs in 286 scaffolds, totaling 18.430 Mbp in size. The input read coverage was 129.9X.

Genome annotation

Genes were identified with Prodigal [3], followed by one round of manual curation using GenePRIMP [4] for genomes in fewer than 10 scaffolds. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAscanSE tool [5] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [6]. Other non-coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL [7]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform [8] developed by the Joint Genome Institute, Walnut Creek, CA, USA [9].

1. Eid John, et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 2008
2. Chin C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013
3. Hyatt D, Chen GL, Lacascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; 11:119.
4. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; 7:455–457.
5. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; 25:955–964.
6. Pruesse E, Quast C, Knittel, Fuchs B, Ludwig W, Peplies J, Glckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nuc Acids Res* 2007; 35: 2188–7196.
7. INFERNAL. Inference of RNA alignments. <http://infernal.janelia.org>.
8. The Integrated Microbial Genomes (IMG) platform. <http://img.jgi.doe.gov>.
9. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; 25:2271–2278.