

Agaricus bisporus 8x Release [project ID 4084801 / March 2009]

JGI have recently announced that the 8X sequencing & draft assembly for *Agaricus bisporus* (H97, lead homokaryon) is complete.

This assembly will now undergo a machine annotation with Igor Grigoriev's JGI team (ca. 3-4 months). The draft assembly/annotation will then transfer to the community for hands-on detailed annotations (ca. 6 months). When this steps are complete, the *Agaricus bisporus* draft genome assembly and annotation will become publicly available via the JGI web portal, and both the assembly & annotation will be deposited in GenBank.

An extensive *Agaricus bisporus* U1 (H97 x H39 heterokaryon) EST data set has also been clustered to the genome.

Jeremy Schmutz of the JGI has provided an assembly report <PDF>.

Library construction for the second homokaryon, *Agaricus bisporus* var. *burnettii* JB137-S8 is in progress.

----- Original Message -----

Subject: [asm_release] Agaricus bisporus 8x
Release [project ID 4084801]
Date: Fri, 13 Mar 2009 17:21:26 -0500
From: Jeremy Schmutz <jschmutz@hudsonalpha.org>
To: asm_release@quagmire.jgi-psf.org

[asm_release] Agaricus bisporus 8x Release
[project ID 4084801]

Assembly Stats:

Main genome scaffold total: 27
Main genome contig total: 253
Main genome scaffold sequence total: 30.3 MB
Main genome contig sequence total: 30.0 MB (-> 0.7% gap)
Main genome scaffold N/L50: 5/2.4 MB
Main genome contig N/L50: 35/263.8 KB
Number of scaffolds > 50 KB: 17
% main genome in scaffolds > 50 KB: 99.7%

Minimum Scaffold Length	Number of Scaffolds	Number of Contigs	Total Scaffold Length	Total Contig Length	Scaffold Contig Coverage
All	27	253	30,252,850	30,047,564	99.32%
1 kb	27	253	30,252,850	30,047,564	99.32%
2.5 kb	27	253	30,252,850	30,047,564	99.32%
5 kb	22	246	30,233,799	30,029,562	99.32%
10 kb	19	242	30,214,738	30,010,676	99.32%
25 kb	17	239	30,175,376	29,971,556	99.32%
50 kb	17	239	30,175,376	29,971,556	99.32%
100 kb	17	239	30,175,376	29,971,556	99.32%
250 kb	15	233	29,728,699	29,528,316	99.33%
500 kb	15	233	29,728,699	29,528,316	99.33%
1 mb	11	207	27,079,627	26,897,561	99.33%
2.5 mb	4	87	14,558,022	14,522,480	99.76%
5 mb	1	36	5,440,081	5,425,440	99.73%

Coverage Stats:

LIB	COV.	INSERT	STDDEV
FWBC	3.84x	2904 +/-	405
FWBB	3.92x	6367 +/-	572
FWBA	0.74x	39552 +/-	2828
Total	8.50x		

Completeness Stats:

To evaluate the completeness of the genomic sequence we placed the clustered ESTs. We used all clusters > 400bp. Of the 12,621 clusters 12,050 (97.54%) are well placed (at 90% ID, 85% Coverage) in the genome. 267 clusters are identified as artifacts. 78 (0.63%) are found at > 50% coverage and 226 (1.83%) cluster sequences are not found. The 226 not found clusters are mostly not found in the reads either and appear to probably be poorly sequenced clusters built from 454 reads that disagree with the genome. It doesn't appear that these indicate a problem in the genome sequence.

The scaffolds are divided into the following sets:

- 1) Main genome
- 2) Mitochondrion
- 3) Repeat
- 4) Excluded scaffolds

Production Stats:

LIB	INSERT	READS	FAILED(%)	VECTOR(%)
FWBC	2904	191136	5634(2.9)	7242(3.8)
FWBB	6367	179424	6821(3.8)	2035(1.1)
FWBA	39552	36192	2416(6.7)	167(0.5)
Total	N/A	406752	14871(3.7)	9444(2.3)

LIB	UNPAIRED(%)	PAIRED(%)	GOODP20	STDDEV
FWBC	3846(2.0)	174414(91.3)	772.60	+/-99.35
FWBB	4060(2.3)	166508(92.8)	772.51	+/-102.36
FWBA	1535(4.2)	32074(88.6)	724.65	+/-106.89
Total	9441(2.3)	372996(91.7)	768.42	+/-102.24