

Thu Oct 13 21:46:44 2005

## REPORT FOR *galdieria* (This file is evaluate.brief.ps.)

run = 'Build\_2005-10-13-A/work'

We compared our contigs with 0 known contigs, of total length 0.

### SOURCE DATA

The genome size was estimated to be 16000000.

Coverage by trimmed production reads = 8.82X.

Coverage by bases of quality  $\geq 20$  in trimmed production reads = 8.1X.

Percentage of trimmed production reads which are paired: 93.6%.

### DEPTH OF COVERAGE

This section gives plate-by-plate statistics for apparent depth of coverage. Its purpose is to facilitate the identification and removal of plates from other species.

2330 plates

overall average depth of coverage = 8.49X

82% of the plates lie between 85% and 115% of the average

There are no plates which lie below 50% and have 20 or more reads.

There are no plates which lie above 200% and have 20 or more reads.

### LENGTHS OF OUR SUPERCONTIGS

There are 894 supercontigs of lengths:

374968 364776 334351 325379 305035 298641 293734 285644 282925 273291 245434  
243541 238216 226412 216452 195119 195029 192517 191898 191787 ...

The N50 supercontig length is 158738.

The largest 10 supercontigs account for 21% of the supercontig length.

The largest 50 supercontigs account for 66.6% of the supercontig length.

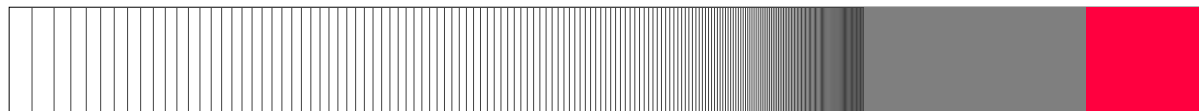
The largest 100 supercontigs account for 82% of the supercontig length.

The largest 500 supercontigs account for 95.7% of the supercontig length.

### LENGTHS OF OUR CONTIGS

There are 1079 arachne contigs, of lengths:

304861 294127 223188 206028 191799 181017 177836 174205 169763 159123 154406  
154036 153558 150233 138705 137000 134679 134130 133385 132261 ...



10% uncovered

The mean length is 13306.

The expected length of a contig containing a random base (chosen from one of these contigs) is 86739.

The N50 contig length is 80518.

The largest 10 contigs account for 14.5% of the contig length.  
The largest 50 contigs account for 46.9% of the contig length.  
The largest 100 contigs account for 67.2% of the contig length.  
The largest 500 contigs account for 92% of the contig length.  
The largest 1000 contigs account for 99.5% of the contig length.

Contigs longer than 250000 account for 4.17% of contig length  
Contigs longer than 100000 account for 39% of contig length  
Contigs longer than 25000 account for 70.3% of contig length  
Contigs longer than 10000 account for 79.6% of contig length  
Contigs longer than 5000 account for 88.6% of contig length  
Contigs longer than 2000 account for 97% of contig length

fraction of genome in gap contigs: 0%

## BASE QUALITY IN OUR CONTIGS

These numbers are computed by comparing with the known (known) contigs, and hence include errors in them. In addition, if there are errors in our alignment of our contigs with the known contigs, these show up too.

The known contigs are too short to evaluate our error rates.

Fraction of bases in reads appearing in contigs: 96.6%.

(Note: The reads referred to are the reads which are passed to the layout stage. They have been trimmed and possibly formed into larger reads prior to layout.)  
(See also reads.unutilized.)

## COVERAGE BY OUR CONTIGS

The total length of our contigs is 14357686.

The estimated total length of our supercontigs is 14932726.

(To the extent that our contigs or supercontigs overlap, these figures are inflated.)

## LONG INSERT INFORMATION

After trimming and deletion of suspicious reads, reads (from both ends) of 0 ~40k inserts were available to us. Of these, 0 were fully utilized, 0 were half-utilized (i.e. one of the end reads is in the assembly), and 0 were not utilized at all.

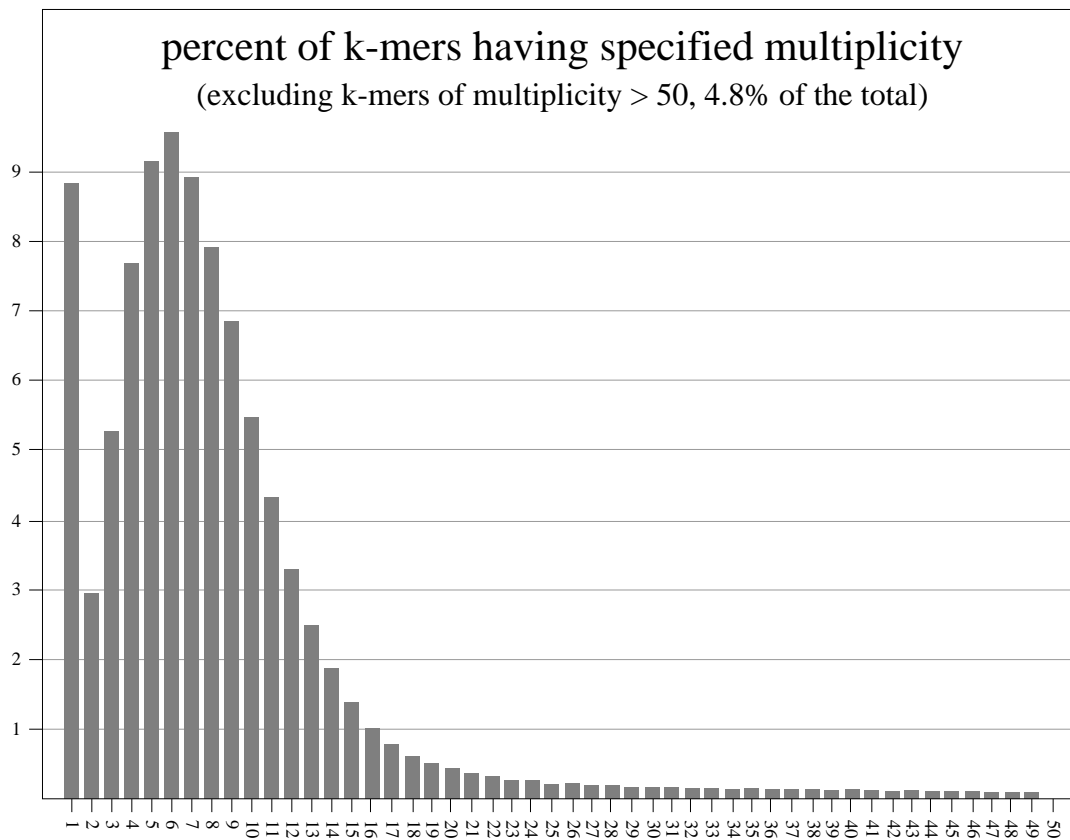
## UNDERCOLLAPSING AND FAILURE TO JOIN CONTIGS

See the data shown below!

## K-MER ANALYSIS (k = 24)

A total of 75622747 occur with multiplicity > 2.

Assuming a genome size of 16000000, this would imply a coverage of 4.73X.



## CONTIG BY CONTIG ANALYSIS

We show the correlation between the known contigs and our merged contigs.

Known contigs which are completely covered by a merged contig are not shown.

For an account of what contigs went into a given merged contig, see `evaluate.order`.

For a list of all original reads that appear in the merged contigs which are shown, see `select.humanlocs`

## SUPERCONTIG COVERAGE

## MISASSEMBLIES

None identified!

## CONTIGS OF OURS, ASSIGNED TO MULTIPLE KNOWN CONTIGS

None found!