

**How big is your Y? A genome-sequence based estimate of the size of the male-specific region in *Megaselia scalaris***

Kenneth B. Hoehn<sup>\*§</sup> and Mohamed A. F. Noor<sup>§</sup>

<sup>\*</sup>Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom

<sup>§</sup>Biology Department, Duke University, Durham, NC 27708

DOI: [10.1534/g3.114.015057](https://doi.org/10.1534/g3.114.015057)

<b>Page</b>	<b>Contents</b>
2 SI	Table S1: Summary statistics of <i>Megaselia</i> male assembly
3 SI	Figure S1: <i>Megaselia</i> 15-mers vs contig size
4 SI	Figure S2: <i>Megaselia</i> MSP weighted by minimum contig size and prior weight
5 SI	Figure S3: <i>D. virilis</i> MSP weighted by minimum contig size and prior weight

**Table S1 Assembly statistics from *Megaselia* male genome assembly: Output from SOAPdenov2.**

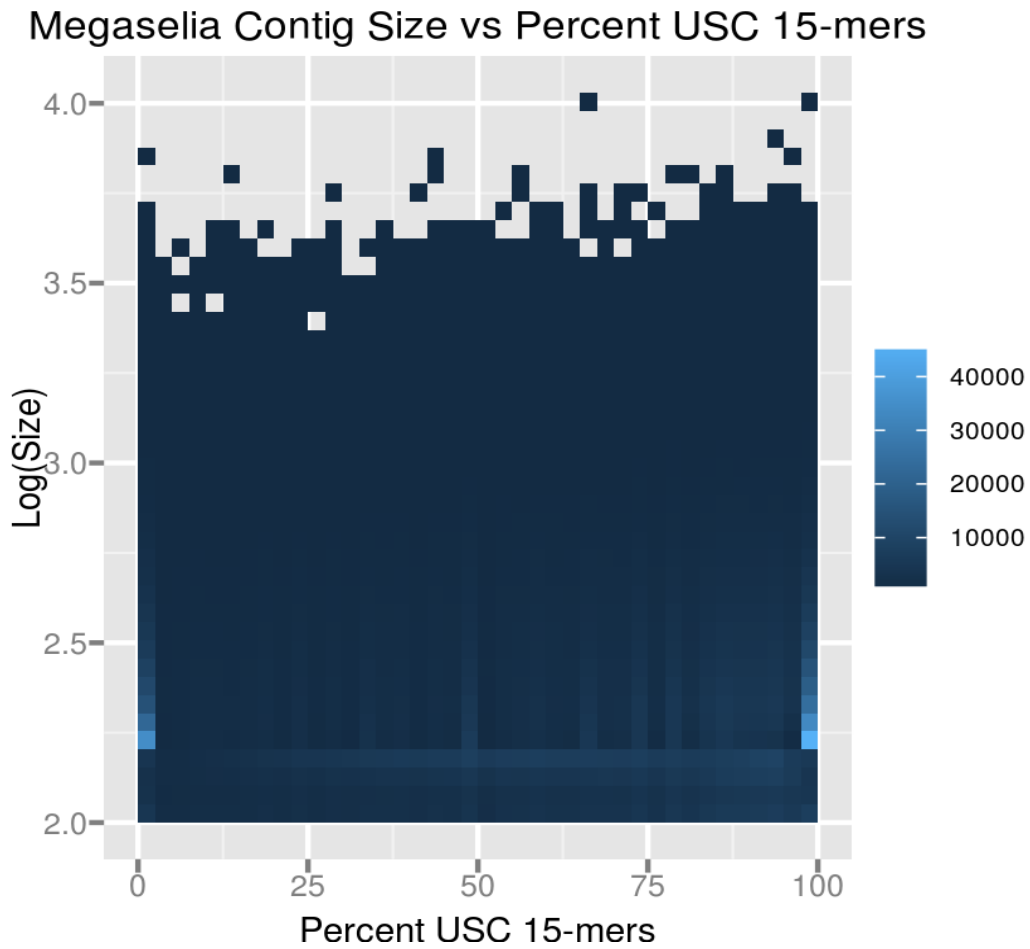
**Assembly Statistics: *Megaselia* Male**

<b>Total Size (bp)</b>	453444693
<b>Scaffold Number</b>	1755386
<b>Mean Size</b>	258
<b>Median Size</b>	183
<b>Longest Sequence</b>	11221
<b>Shortest Sequence</b>	100
<b>Singleton Number</b>	1755386

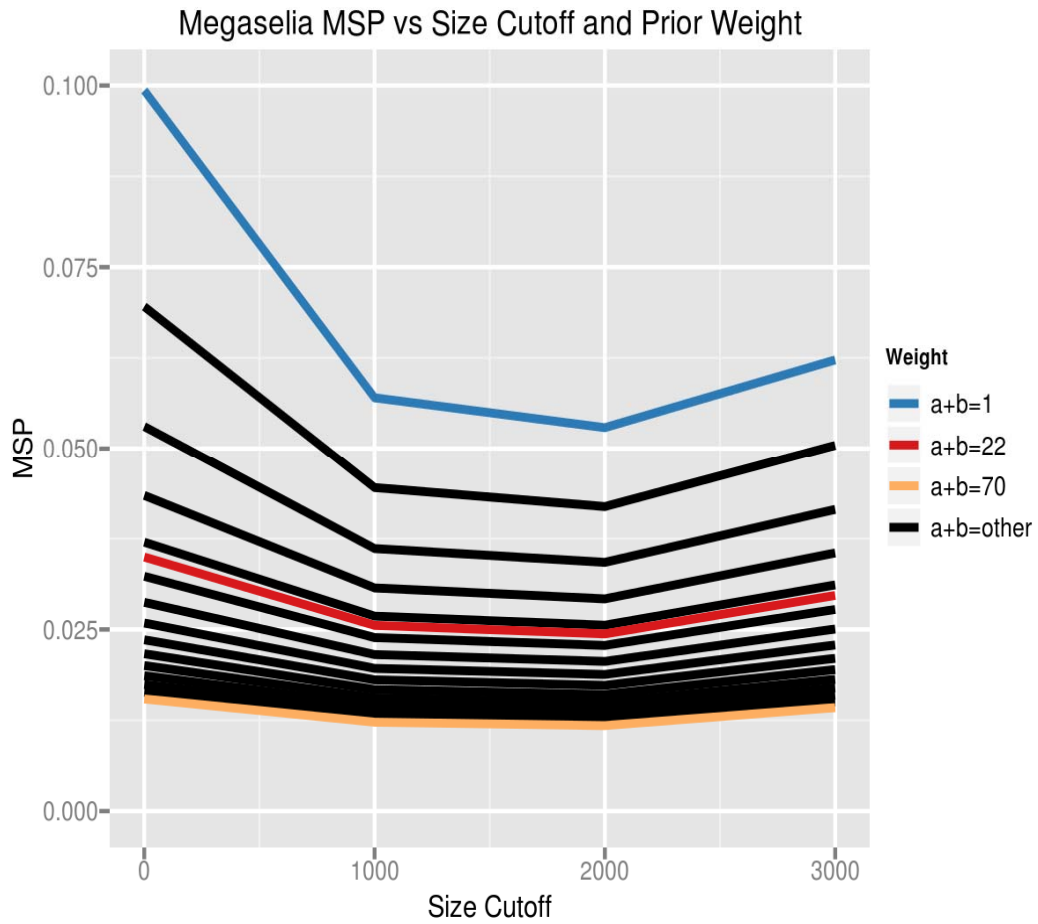
<b>Scaffolds &gt; Length (bp)</b>	<b>Number</b>	<b>Percentage</b>
scaffolds>100	1743801	99.34%
scaffolds>500	156140	8.89%
scaffolds>1K	29777	1.70%
scaffolds>10K	2	0.00%
scaffolds>100K	0	0.00%
scaffolds>1M	0	0.00%

<b>Length Quantiles</b>	<b>Length</b>	<b>Number</b>
N10	978	31746
N20	643	90350
N30	476	173071
N40	368	281942
N50	292	420610
N60	236	593838
N70	193	806946
N80	160	1065221
N90	136	1373198

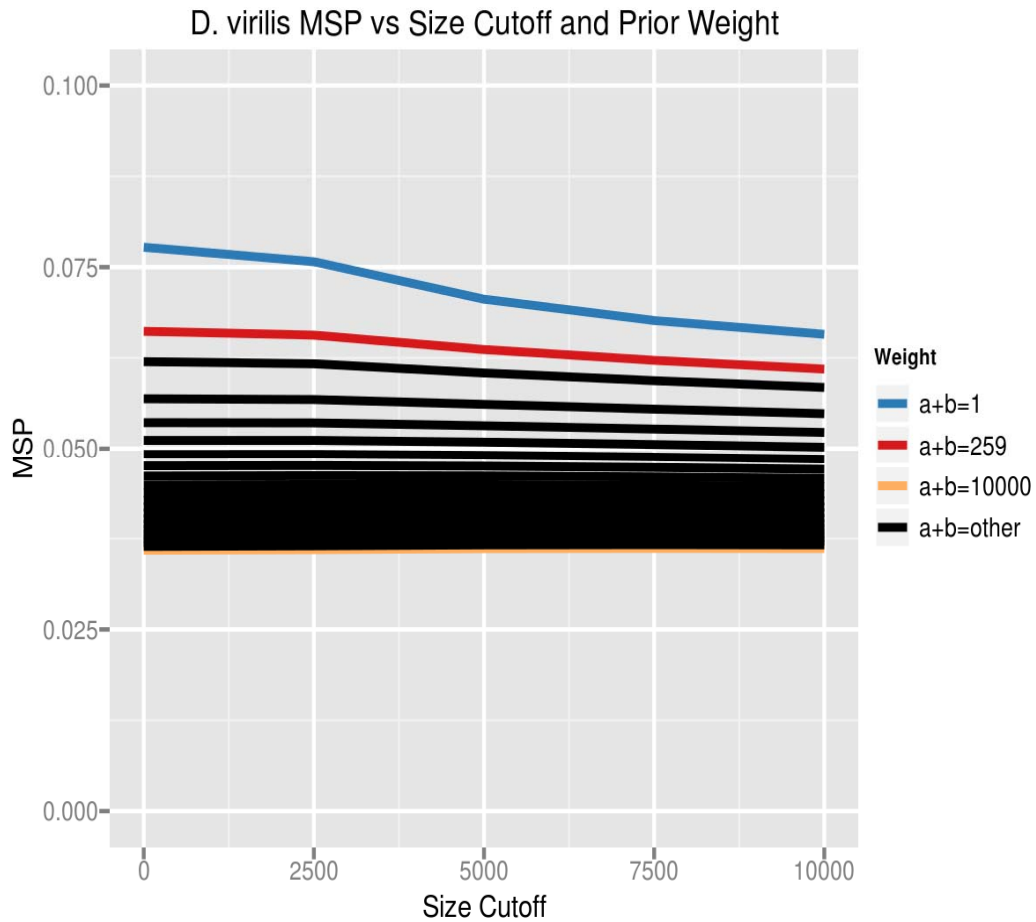
<b>Nucleotide Content</b>	<b>Number</b>	<b>Percentage</b>
Nucleotide A	153644768	33.88%
Nucleotide C	76208688	16.81%
Nucleotide G	74994329	16.54%
Nucleotide T	148596908	32.77%
GC Content	33.35%	



**Figure S1** Heatmap showing count of *Megaselia* 15-mers by  $\log_{10}(\text{Contig Size})$  and percent of single copy 15-mers that were unmatched (USC) by female short read data. Note peaks at 0% USC and 100% USC 15-mers. These are inferred to be caused by autosomal vs. male-specific contigs, respectively.



**Figure S2** Estimates of male specific portion (MSP) of *M. scalaris* by minimum contig size and prior weight (a+b). The least stringent estimate (a+b=1), median estimate (a+b=22), and most stringent estimate (a+b=70) are colored. Prior weight values are otherwise selected in intervals of 5.



**Figure S3** Estimates of male specific portion (MSP) of *D. virilis* by minimum contig size and prior weight (a+b). The least stringent estimate (a+b=1), median estimate (a+b=259), and most stringent estimate (a+b=10000) are colored. Prior weight values are otherwise selected in intervals of 500. These values are much higher than in the *Megaselia* estimation because the contigs of the *D. virilis* assembly are much larger (median contig size of 189 vs 4605, respectively).