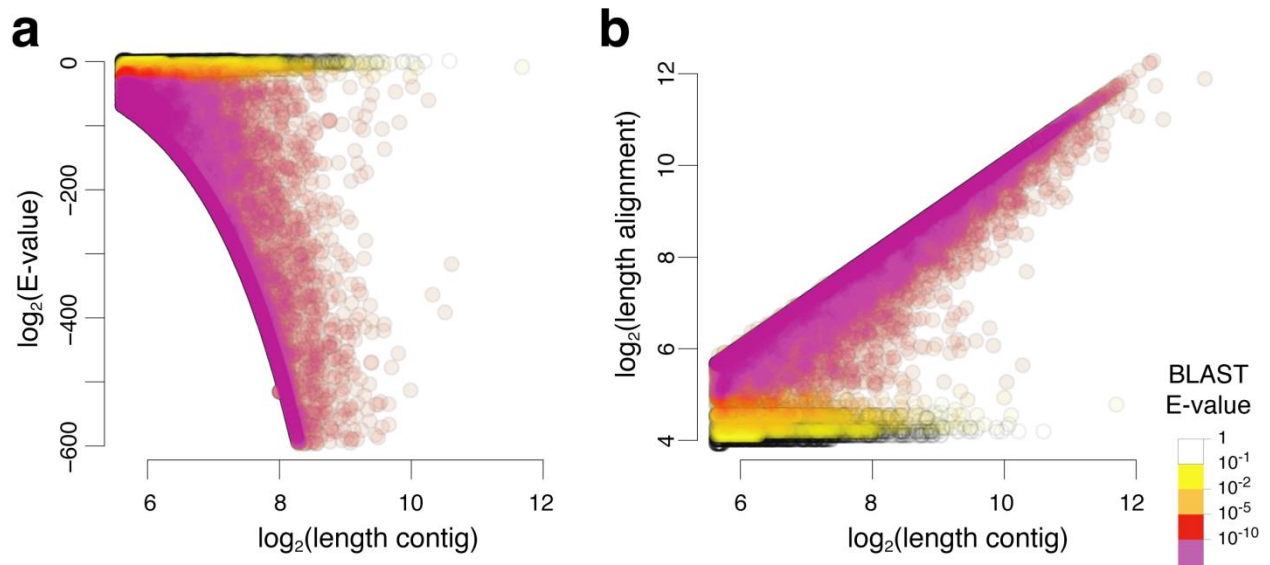


**Current Biology, Volume 22  
Supplemental Information**

**Simplification and Desexualization  
of Gene Expression  
in Self-Fertile Nematodes**

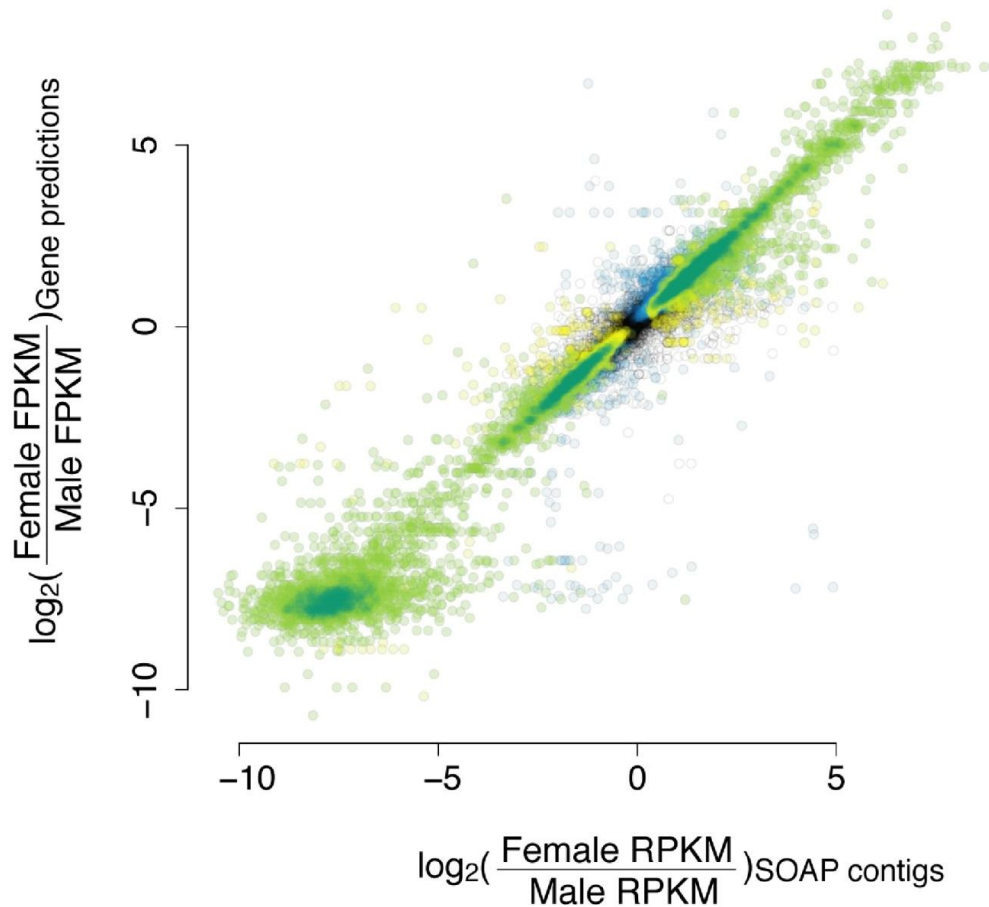
**Cristel G. Thomas, Renhua Li, Harold E. Smith, Gavin C. Woodruff, Brian Oliver, and Eric S. Haag**



**Figure S1. *C. elegans* SOAPdenovo assembly comparison to *C. elegans* Gene Prediction**

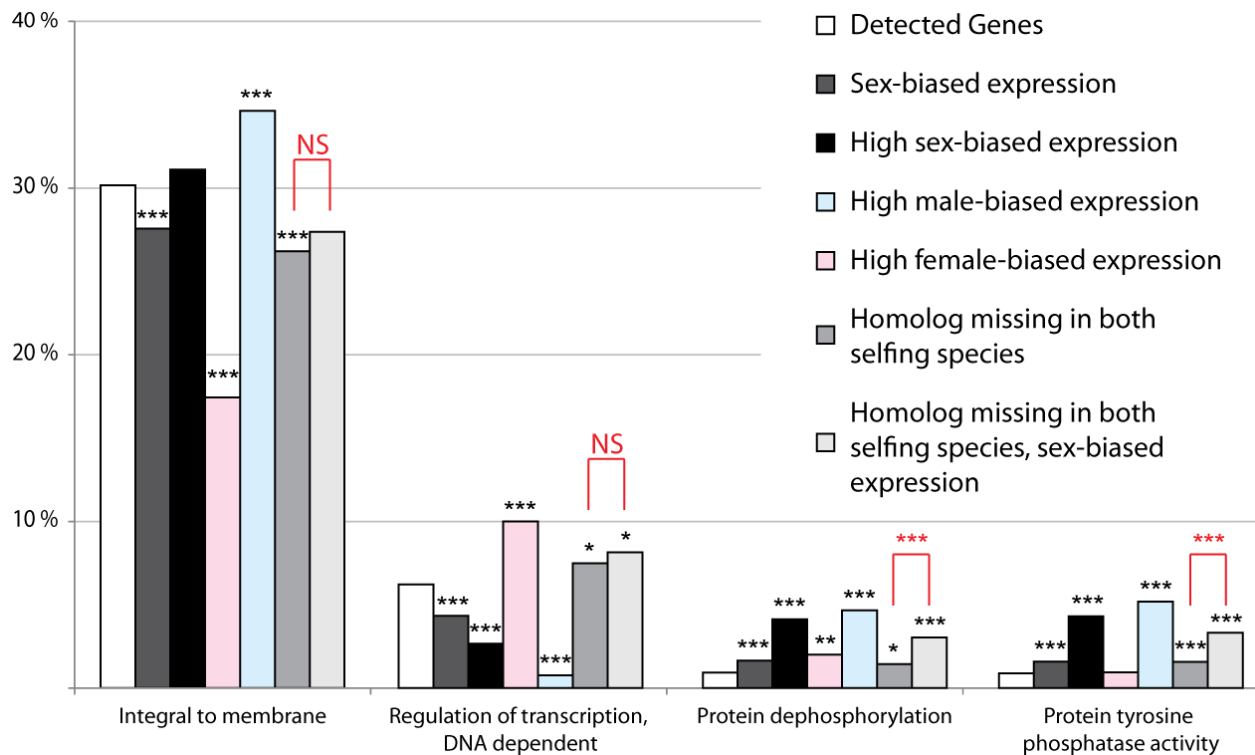
(a) Contig length against BLASTn alignment E-value.

(b) Contig length against BLASTn alignment length. Hits for which the BLASTn E-value was below  $10^{-10}$  are plotted in purple, between  $10^{-10}$  and  $10^{-5}$  in red, between  $10^{-5}$  and  $10^{-2}$  in orange and between  $10^{-2}$  and  $10^{-1}$  in yellow.



**Figure S2. Consistency of Inferred Sex Bias in Two Analytical Pipelines**

For each *C. elegans* cDNA contig assembled by SOAPdenovo with a significant BLAST hit to the *C. elegans* genome ( $E\text{-val} < 10^{-10}$ ), the value of the expression differential between the female and male datasets is plotted against the value of the expression differential of its corresponding best BLAST hit in the set of predicted *C. elegans* genes. Yellow dots represent transcript pairs in which the expression of the SOAPdenovo contig is significantly sex-biased, blue dots those for which the expression of the predicted gene is sex-biased, and green dots those for which both are significantly sex-biased.



**Figure S3. Gene Ontology Analysis of Genes with Different Patterns of Conservation and Sex Bias**

The fraction of genes predicted to code for proteins integral to membrane, involved in regulation of transcription, protein dephosphorylation and protein tyrosine phosphatase are indicated for all *C. remanei* genes whose expression was detected (white), sex-biased (dark grey), highly sex-biased (black), highly female-biased (pink) and highly male-biased (blue), as well as for genes whose homologs were missing in both selfing species (grey) and sex-biased (light grey). Significance of difference between the latter two (red) or to detected number of genes (black) was assessed by Chi square test (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ).

**Table S1. SOAPdenovo Transcriptome Assembly Statistics**

	<i>C. japonica</i>	<i>C. elegans</i>	<i>C. brenneri</i>	<i>C. remanei</i>	<i>C. briggsae</i>	<i>C. briggsae</i> ( <i>C. elegans</i> nb of reads)
Total number of reads	236,262,428	152,031,884	228,536,165	195,998,192	234,183,830	152,031,884
Estimated Transcriptome size	43.2 Mb	27 Mb	41.2 Mb	36.1 Mb	28.1Mb	
Estimated sequencing depth	197X	203X	200X	196X	300X	195X
Total number of contigs	73,739	24,965	78,540	50,841	22,636	16,024
Total assembly size (bp)	11,694,030	6,161,094	12,204,742	9,473,757	6,496,394	4,761,219
N50 (bp)	246	489	219	331	602	661
Consistently detected contigs	72,956	24,399	76,496	50,427	22,480	15,920
Differentially expressed contigs	30,802 (42.2%)	11,394 (46.7%)	22,930 (30%)	23,767 (47.1%)	10,608 (47.2%)	7,369 (46.3%)
Female-biased	17,781 (23.7%)	5,061 (20.2%)	8,926 (11.0%)	11,817 (22.3%)	5,739 (25.5%)	3,983 (25.0%)
Female-specific	500 (0.7%)	138 (0.6%)	530 (0.7%)	570 (1.1%)	79 (0.3%)	71 (0.4%)
Male-biased	13,021 (17.8%)	6,333 (26%)	14,004 (18.3%)	11,950 (23.7%)	4,790 (21.3%)	3,315 (20.8%)
Male-specific	3,238 (4.4%)	836 (3.4%)	6,675 (8.7%)	4,601 (9.1%)	885 (4.0%)	761 (4.8%)

Transcriptome size is estimated by extrapolating *C. elegans* exon content [1] to publicly available genome sequence assembly sizes [available at <http://genome.wustl.edu/pub/organism/Invertebrates>], corrected for heterozygosity [2]. The percentages in parentheses are fraction of consistently detected contigs. The right-most column provides an alternative assembly for *C. briggsae* with the same number of reads used for *C. elegans* for comparison.

**Table S2. *C. elegans* SOAPdenovo Transcriptome BLASTn Statistics**

E-value cutoff	Number of SOAPdenovo contigs	Fraction of total number of contigs
0	4,528	18.1%
$10^{-10}$	20,647	82.7%
$10^{-5}$	21,234	85.05%
$10^{-2}$	22,047	88.31%
$10^{-1}$	22,843	91.5%

**Table S3. Comparison of Analysis Pipelines**

Species	SOAPdenovo/Bowtie	TopHat/Cufflinks with genome annotation
	<i>C. elegans</i>	<i>C. elegans</i>
Consistently detected	24,399	19,050
Differentially expressed	11,394 (46.7%)	8,534 (44.8%)
Female-biased	5,061 (20.7%)	3,444 (18.1%)
Female-specific	138 (0.6%)	18 (0.1%)
Male-biased	6,333 (26%)	5,090 (26.7%)
Male-specific	836 (3.4%)	271 (1.4%)

The numbers represent numbers of transcribed units for the de novo assembly-based and gene annotation-based analysis pipelines. The percentages in parentheses are fraction of consistently detected transcripts.

**Table S4. *C. elegans* SOAPdenovo Transcriptome and TopHat/Cufflinks Comparison**

BLASTn E-value Threshold	1	$10^{-10}$
Detected in both pipelines	23,091	20,060
<i>Total in SOAPdenovo pipeline</i>	23,662	20,331
<i>Total in TopHat / Cufflinks pipeline</i>	23,612	20,372
Differentially expressed in both pipelines	9,646	8,991
<i>Total differentially expressed in SOAPdenovo pipeline</i>	11,180	10,003
<i>Total differentially expressed in TopHat / Cufflinks pipeline</i>	13,555	11,802

The number of pairs of contigs/corresponding BLASTn hits in each category are indicated.

**Table S5. *C. remanei* genes with highly sex-biased expression lacking orthologs in both *C. elegans* and *C. briggsae***

Gene names highlighted in blue have a strongly male-biased expression, while that of those highlighted in pink is strongly female-biased. The absence of orthologs was determined by the modENCODE Project's Ortholog List

(<http://compbio.mit.edu/modencode/orthologs/modencode-orths-2011-06-20/README.html>).

Transcription factors with highly female-biased expression mentioned in the main text are noted in the comments column. In addition, three genes in the female-biased set judged from TreeFam (<http://www.treefam.org>) and another study [3] to be false positives are noted in comments highlighted in brown.

[see separate Excel document]

### Supplemental References

1. The *C. elegans* Sequencing Consortium, (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012-2018.
2. Barriere, A., Wang, S., Pekarek, E., Thomas, C., Haag, E., and Ruvinsky, I. (2009). Detecting heterozygosity in shotgun genome assemblies: Lessons from obligately outcrossing nematodes. *Genome Res* 19, 470-480.
3. Liu, Q., Stumpf, C., Wickens, M., and Haag, E.S. (2012). Context-dependent function of a conserved translational regulatory module. *Development* 139, 1509-1521.