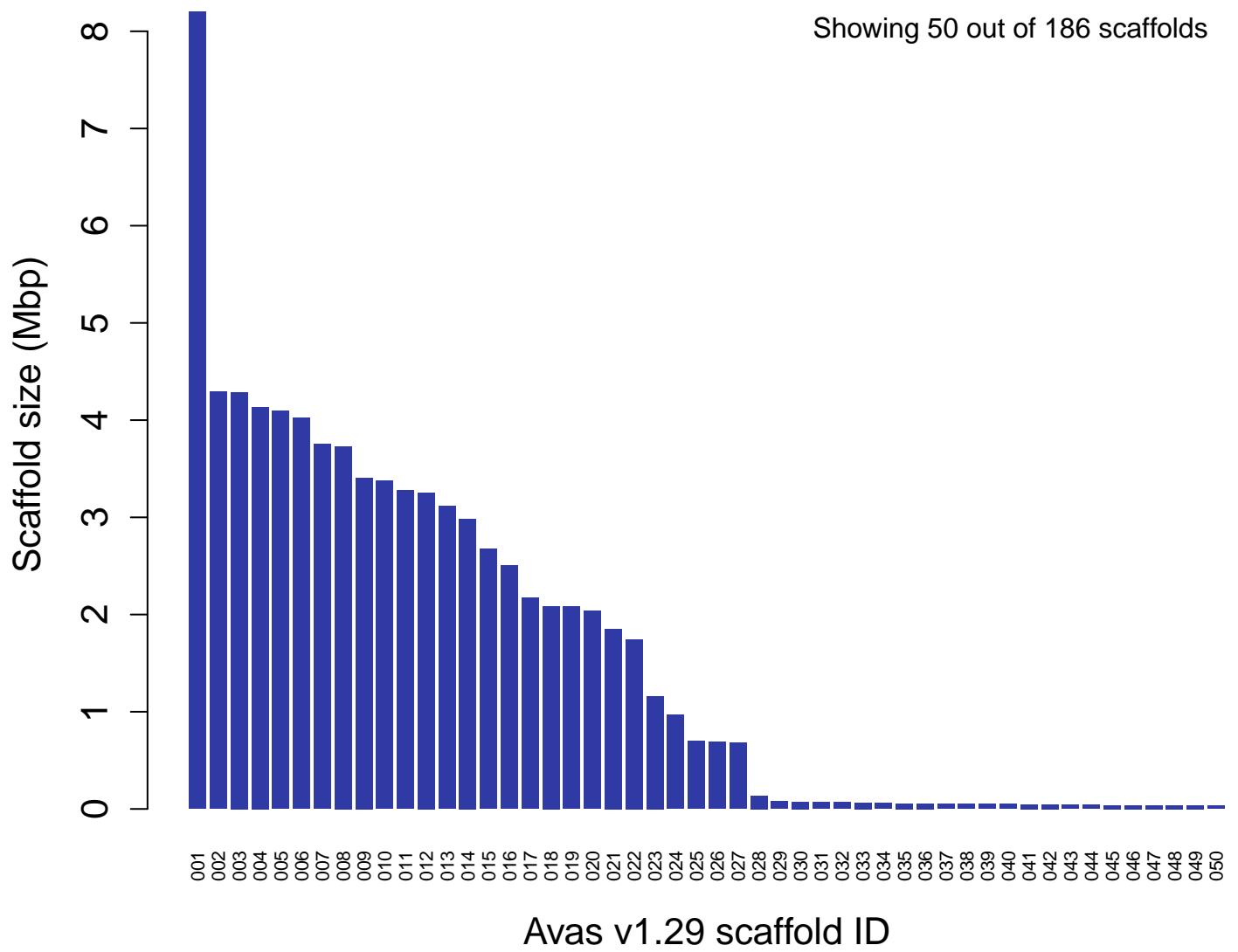
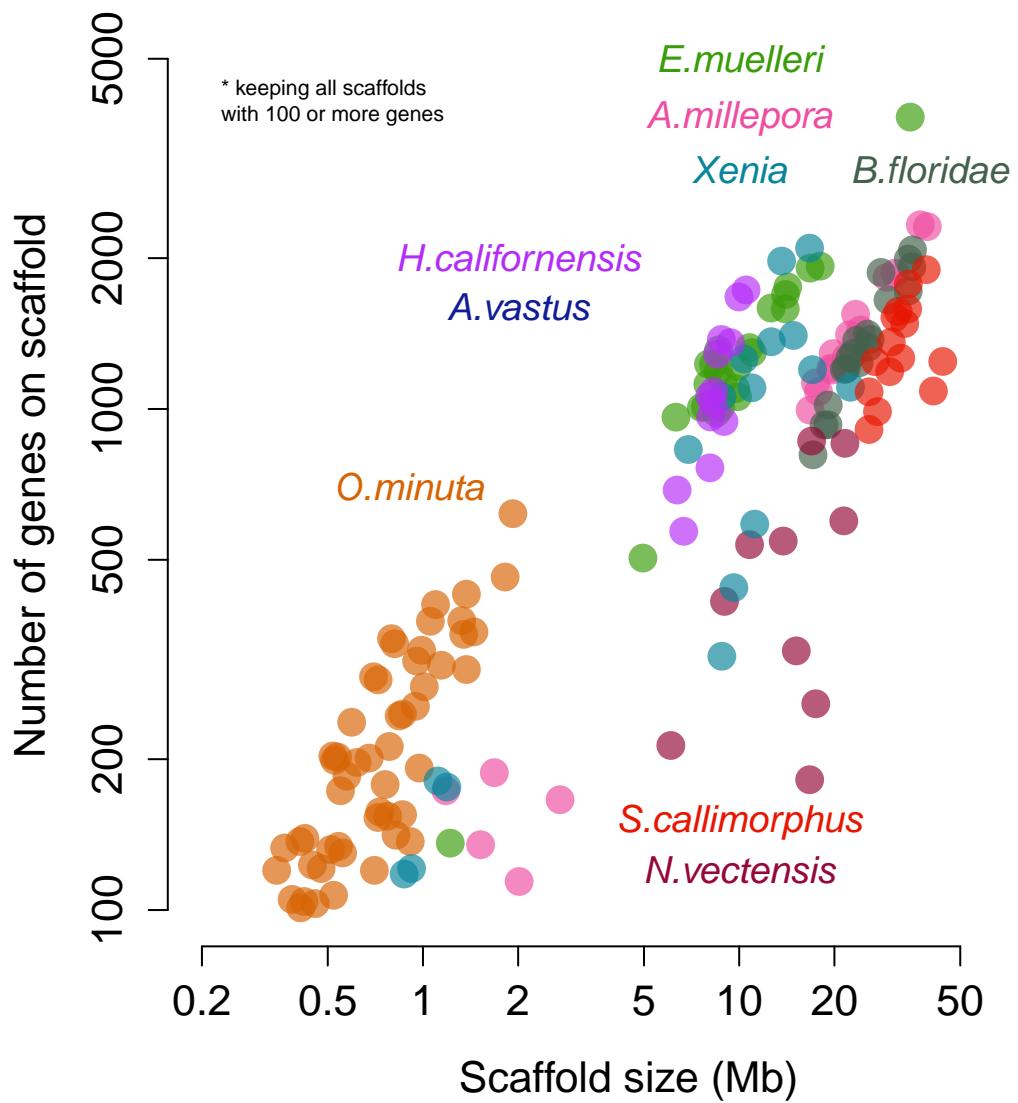


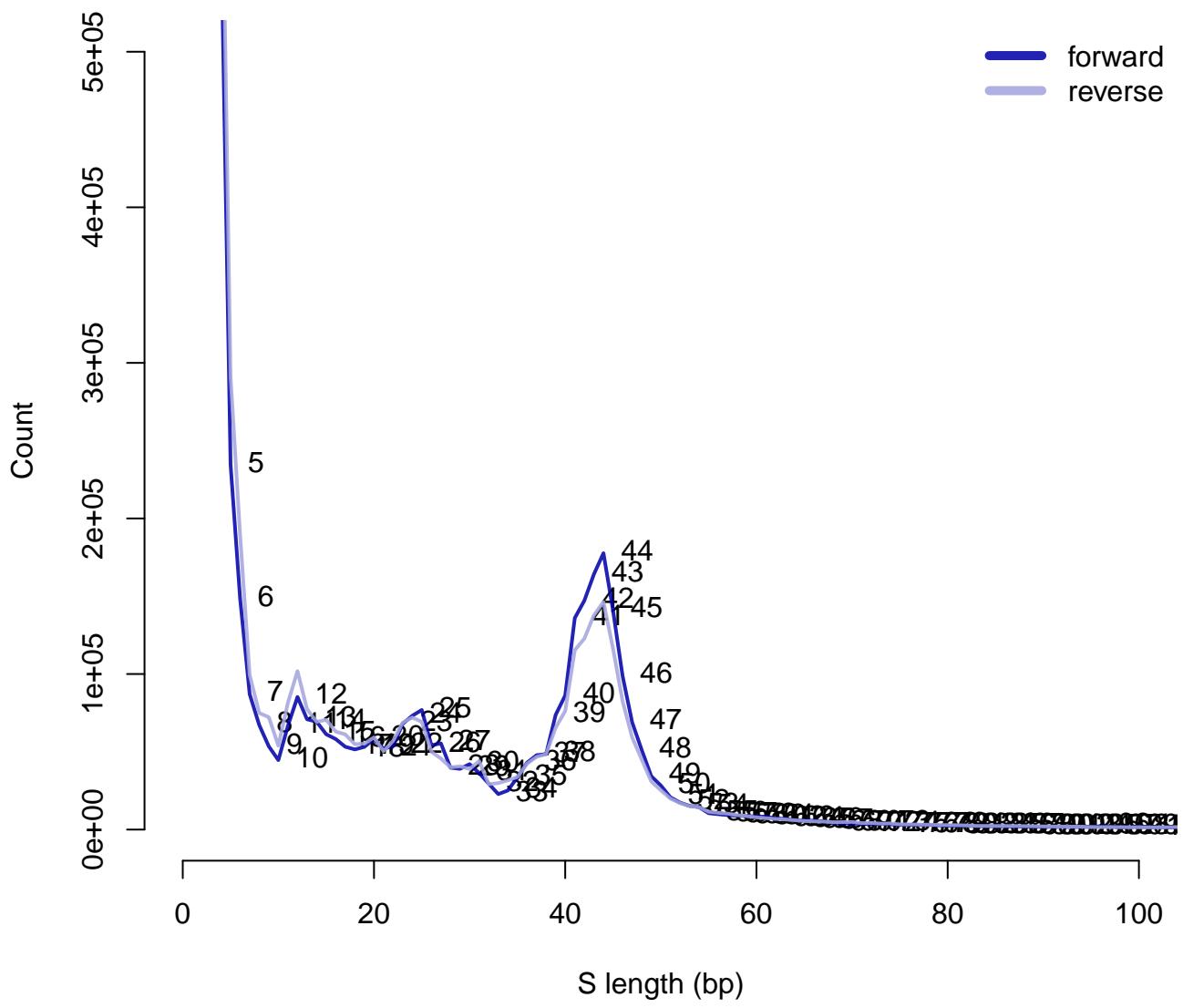
**Supplemental Figure 1:**  
**Mapping coverage histogram,**  
**showing average coverage of 112-fold.**



**Supplemental Figure 2: Barplot of scaffold lengths,  
showing only the top 50.  
The remaining scaffolds are 30kb or less.**

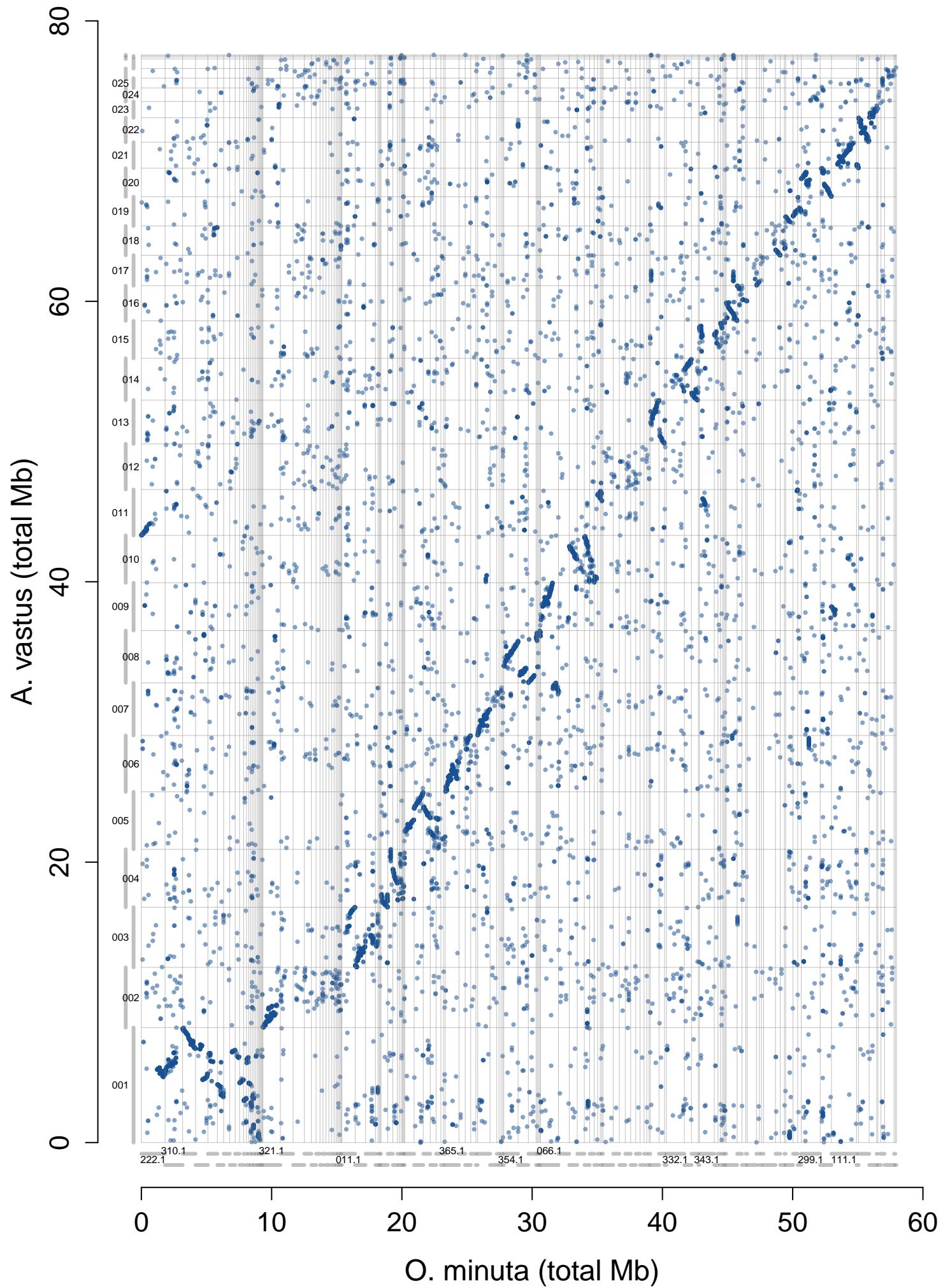


**Supplemental Figure 3:**  
**Plot of numbers of genes per scaffold by length**  
**across several chromosome-scale animal genomes.**

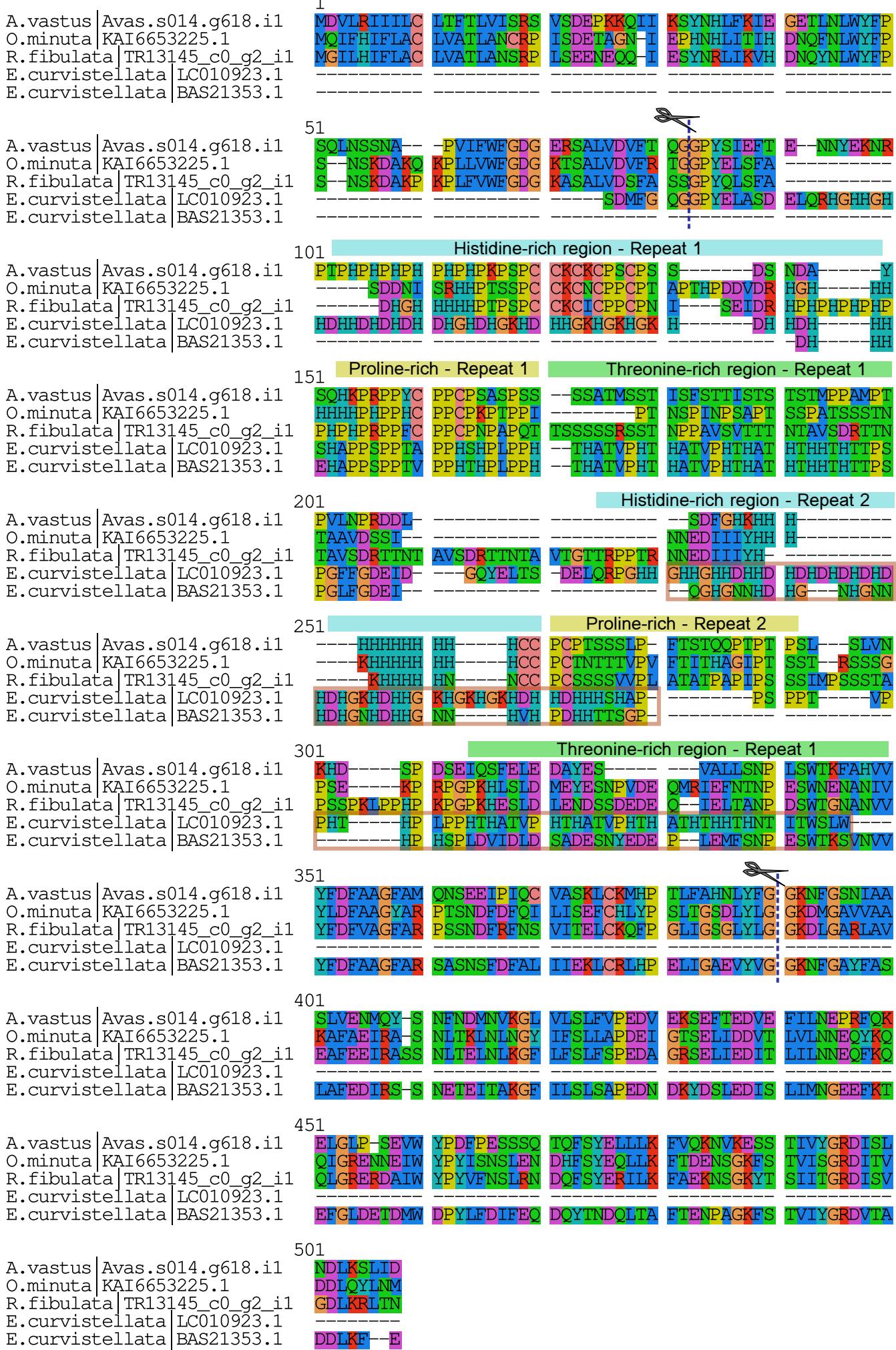


**Supplemental Figure 4:**  
**Histogram of read skip lengths from long RNA reads.**

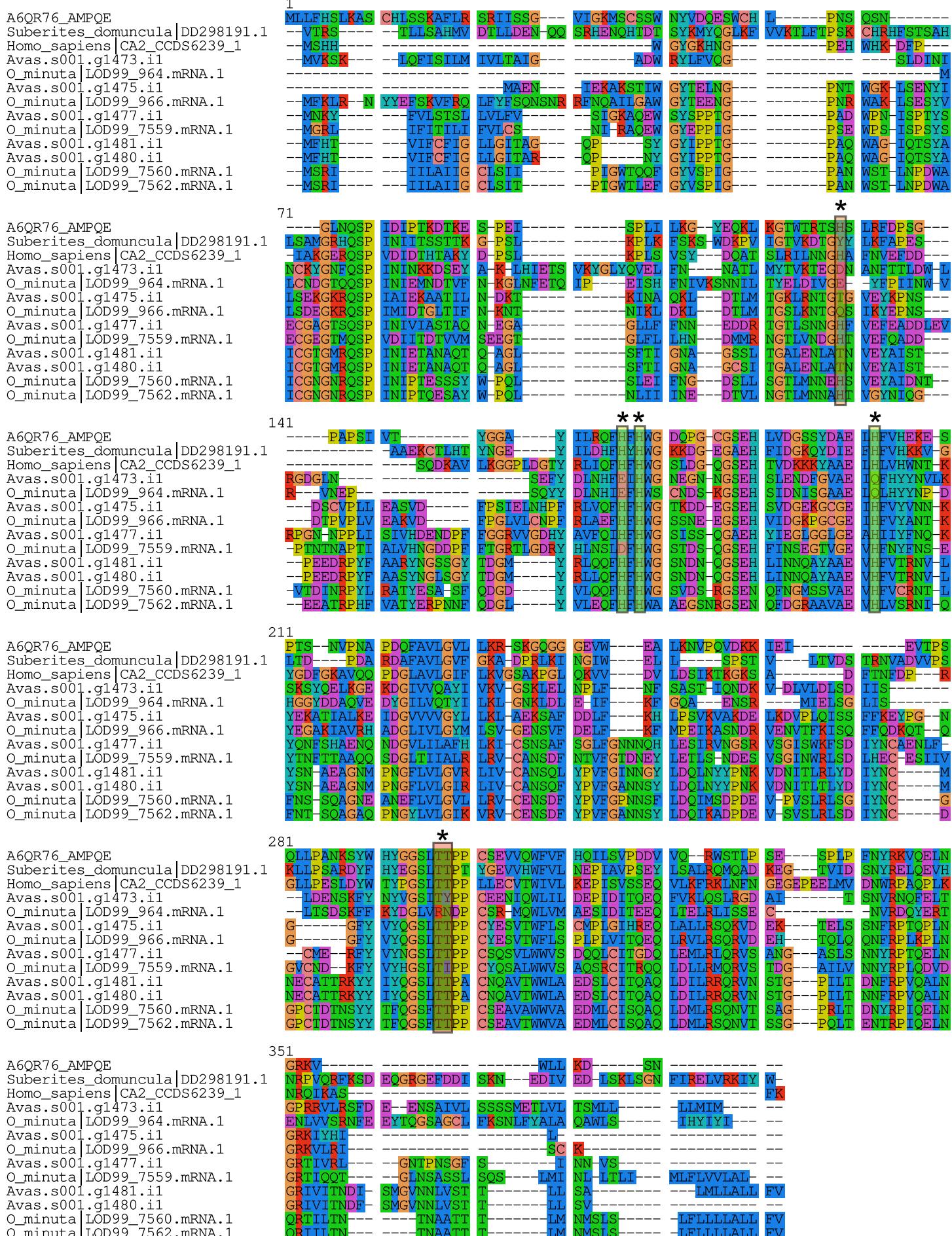
Avas.1.29\_vs\_oopsacas\_gb.scaffold2d\_points.local.tab



A



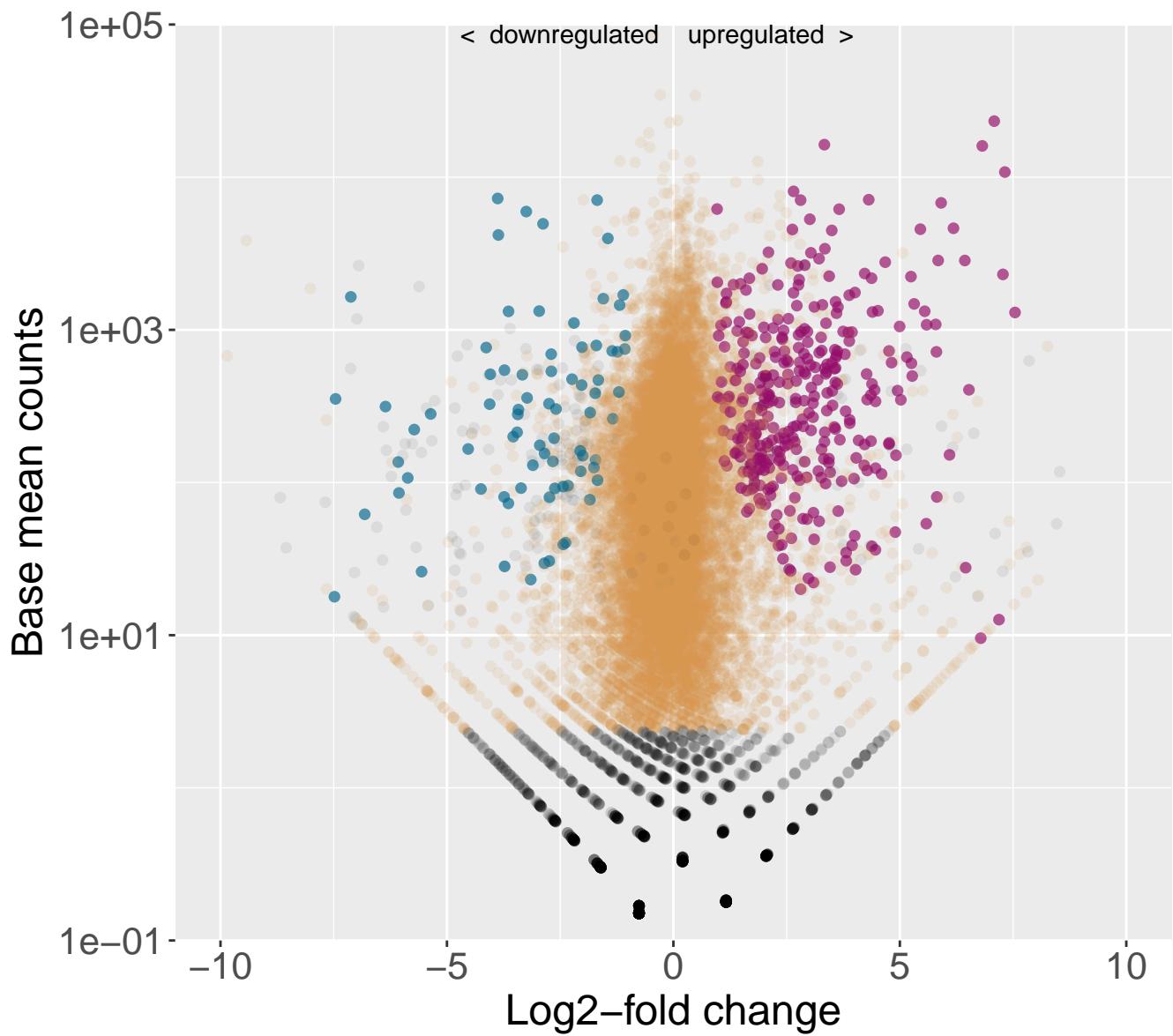
A



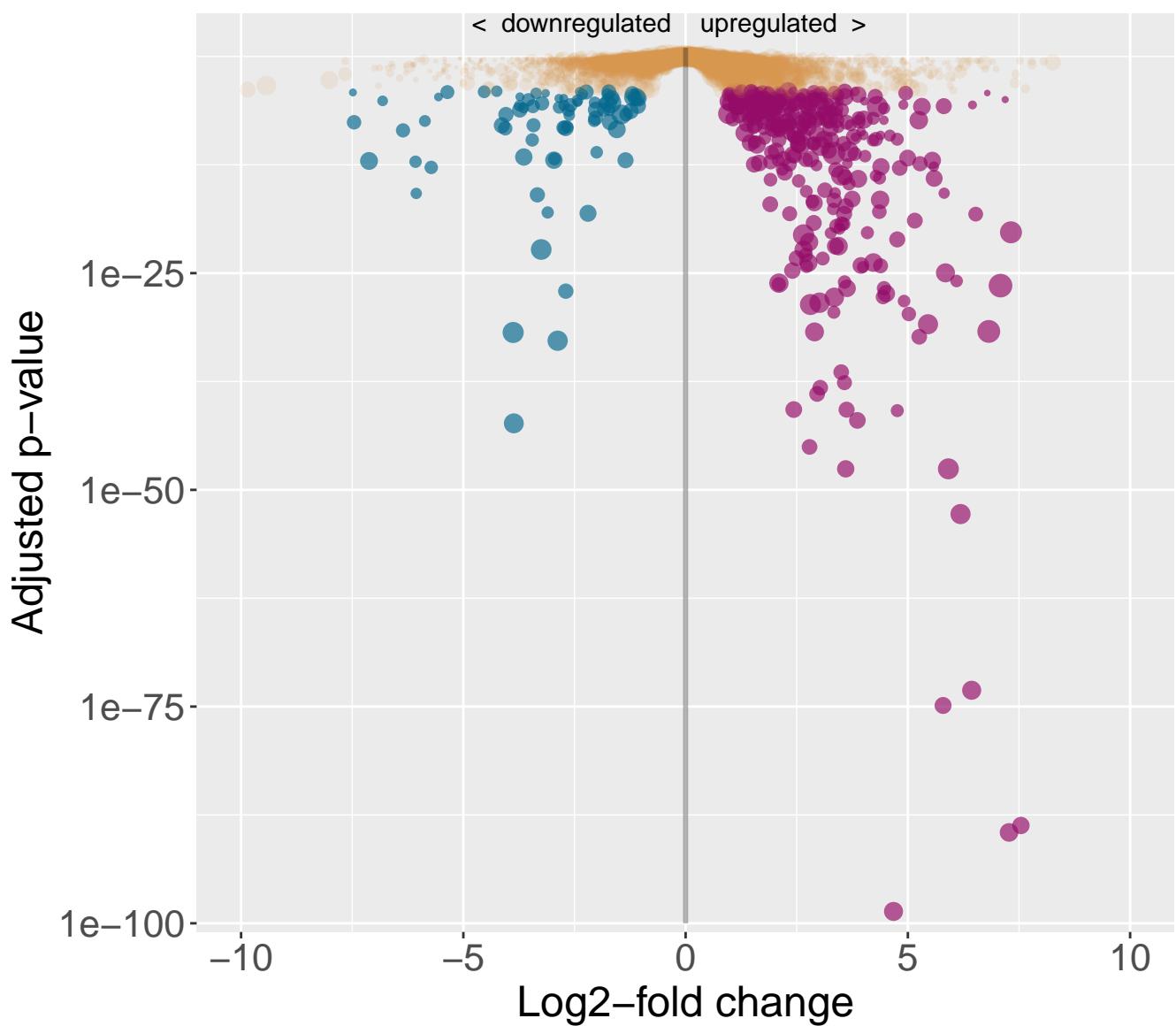
**Supplemental Figure 5:** Large synteny dot plot between *Aphrocallistes* and *Opsacas*. This is the same plot as Figure 2A, made larger for clarity.

**Supplemental Figure 6:** Multiple sequence alignment of glassin homologs, showing the same 5 sequences as in Figure 3. The repeat regions, as defined by the *Euplectella* sequence, are indicated above the sequences. The region of low sequence identity between the two *Euplectella* copies is indicated by the red boxes. Two possible cathepsin cleavage sites at G'G are indicated by dotted lines. The complete alignment with all species can be found in Supplemental Alignment 02.

**Supplemental Figure 7:** Multiple sequence alignment of carbonic anhydrases, showing the additional sequences from *Opsacas*. Stars above columns indicate showing blocks of the binding pocket residues, as shown in Figure 5.

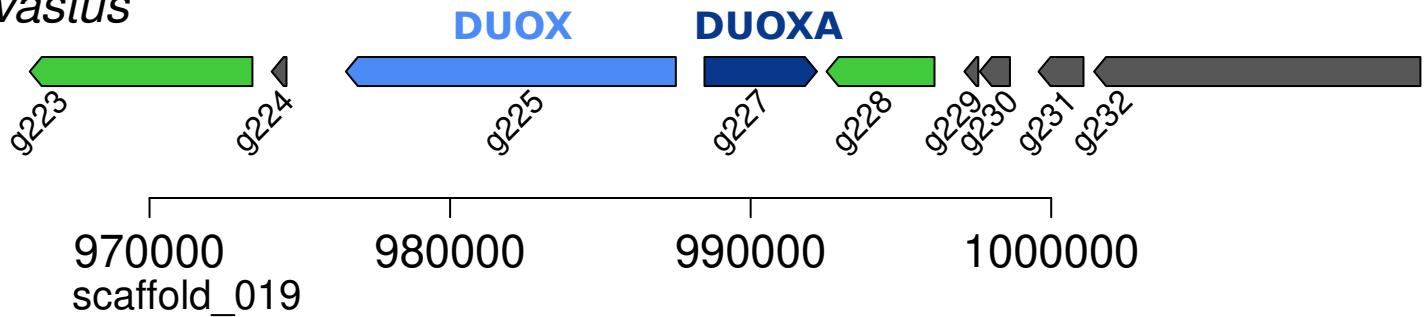


Supplemental Figure 08: Differential gene expression between body and osculum. With p-value threshold of  $1e-4$ , 419 genes were significantly differentially expressed, with 342 upregulated (purple) and 77 downregulated (blue).

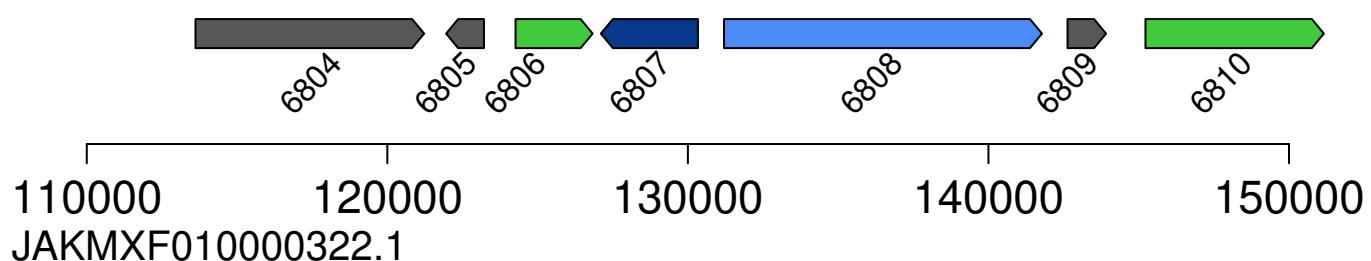


Supplemental Figure 09: Differential gene expression between body and osculum. With p-value threshold of  $1e-4$ , 419 genes were significantly differentially expressed, with 342 upregulated (purple) and 77 downregulated (blue).

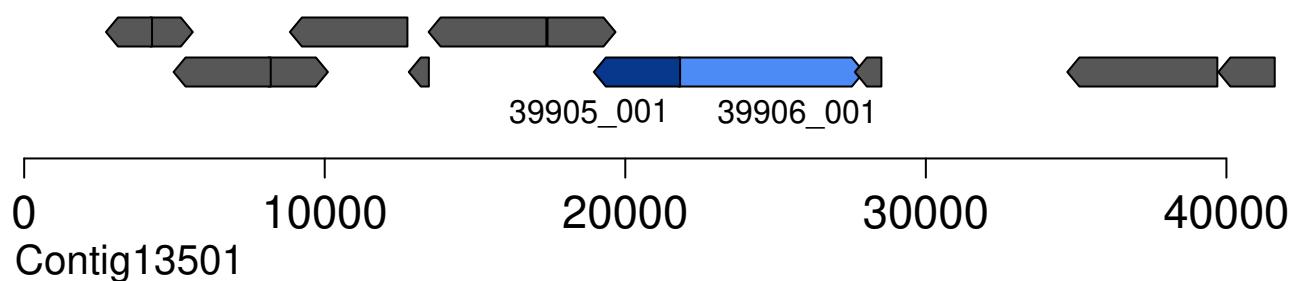
### *A.vastus*



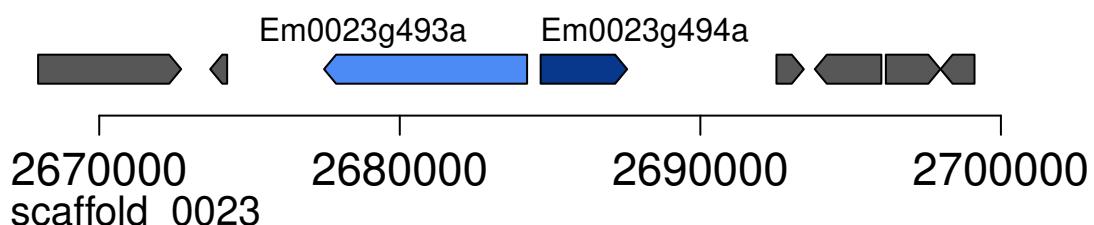
### *O.minuta*



### *A.queenslandica*



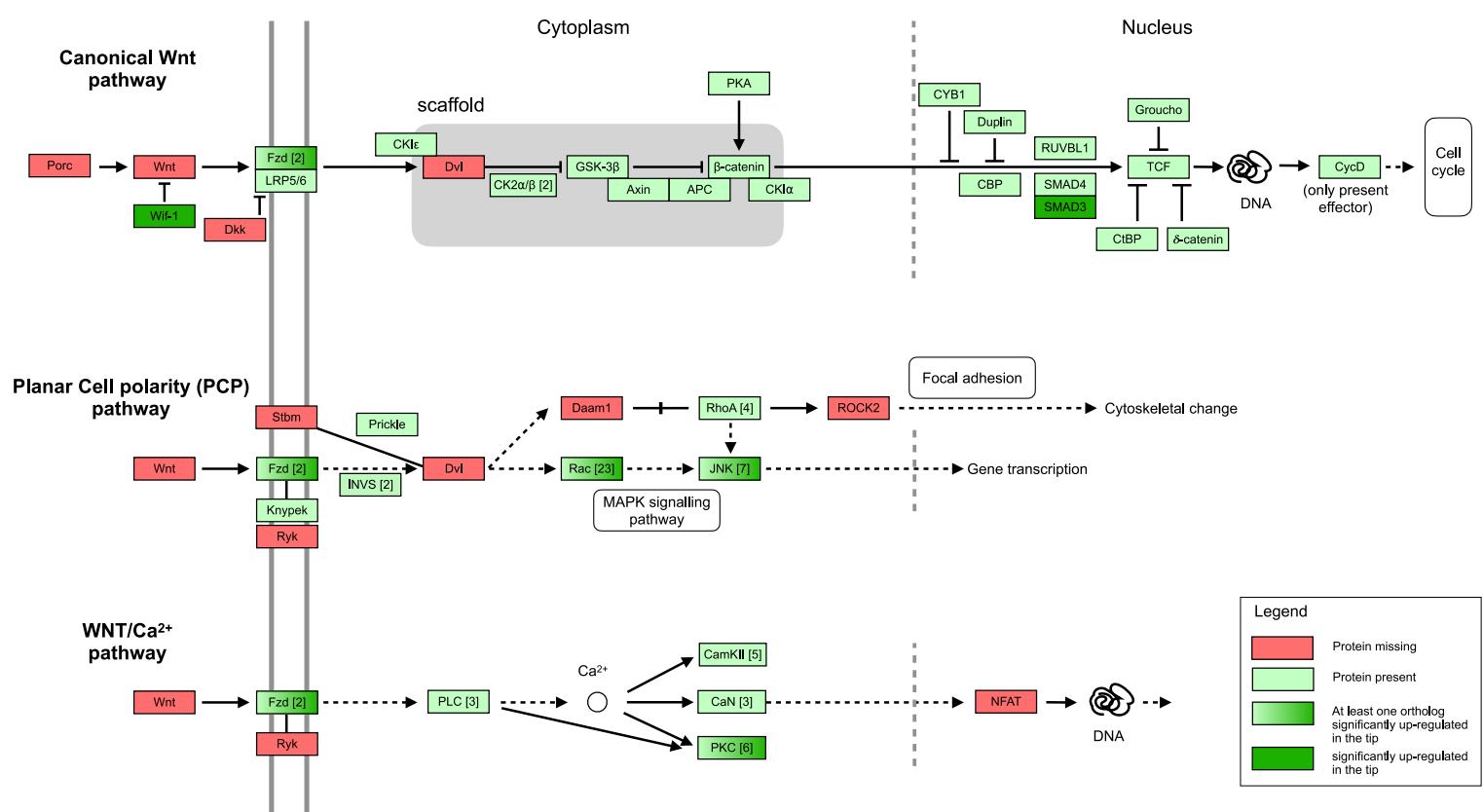
### *E.muelleri*



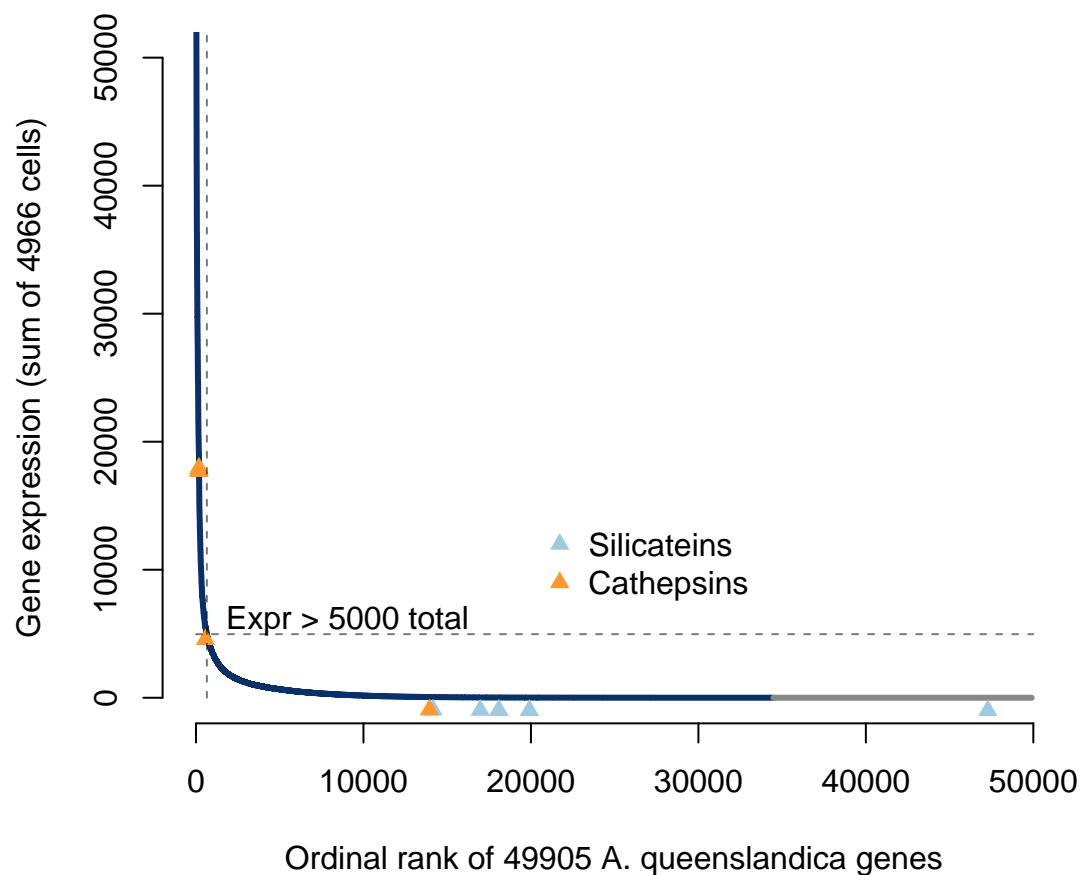
### *S.rosetta*



**Supplemental Figure 10: Schematic of DUOX loci in several sponges and the choanoflagellate *Salpingoeca rosetta*. All panels show a 40kb span. Green-colored genes indicate orthologs between *A. vastus* and *O. minuta*.**

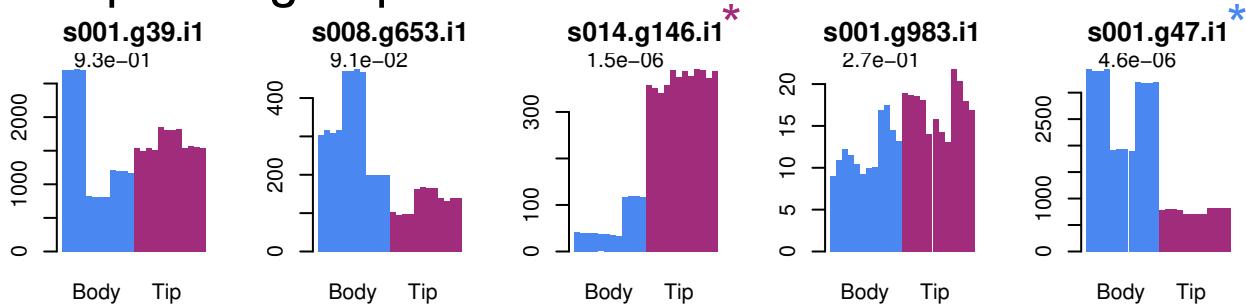


Supplemental Figure 11: Identified components of the Wnt pathway. See Supplemental Table 10 for gene numbers.

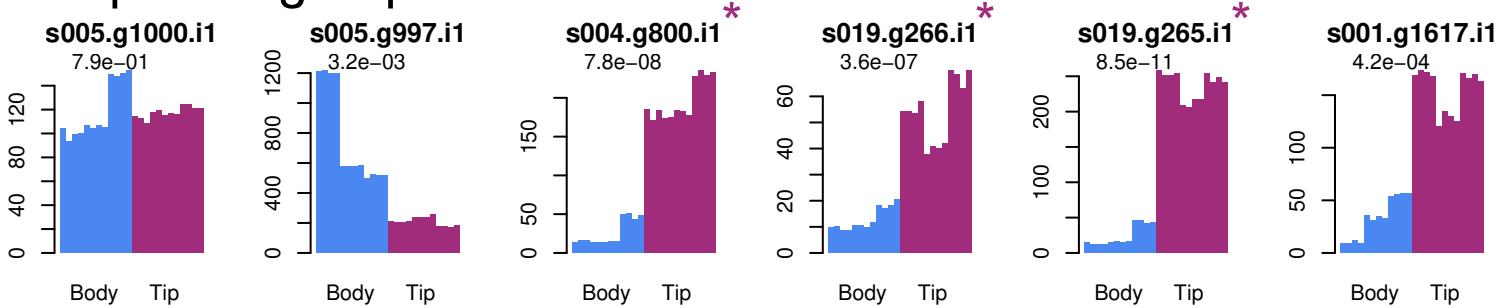


**Supplemental Figure 12:**  
**Plot of *A. queenslandica* body single cell RNAseq**  
**from Sebe-Pedros et al (2018).**

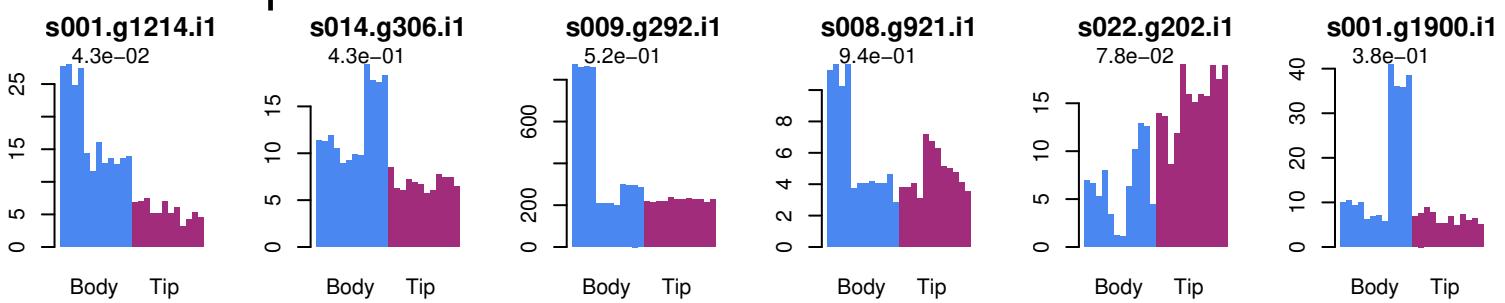
## Cathepsin-L group



## Cathepsin-B group



## Other Cathepsins



Supplemental Figure 13: Expression profiles of cathepsins between the body (blue) and osculum (purple). Genes with an asterisk \* are considered significant in our analyses.

**Supplemental Table 1 ( additional file): Sequencing libraries for DNA and RNA, including the RNA libraries for the additional hexactinellid transcriptomes.**

**Supplemental Table 2 ( additional file): Details of candidate assemblies.**

**Supplemental Table 3: user-annotated problems during manual annotation.**

<b>Annotation problem</b>	<b>Number of genes/transcripts</b>	<b>Percent total</b>
Split gene/ fragmented across 2 or more	1362	6.9%
Wrong exons	512	2.6%
Removed isoform, contained unreal isoform, or part of another gene	441	2.2%
Incomplete UTR only (no problem with CDS)	418	2.1%
Fused, two or more genes joined into one	332	1.6%
Missing exons, gene model excludes internal exons	191	0.9%
No RNA support, gene does not appear supported by any RNaseq	102	0.5%
Frameshift corrected, likely manually edited gene to deal with frameshift in the assembly, or misassembled region	32	0.1%
Wrong starting exon, 5' exon not supported by RNaseq	19	0.1%
All other issues	84	0.4%
Total	3493	17.8%

**Supplemental Table 4: evidence sources for final transcripts used in the manual annotation.**

<b>Gene/transcript source used in final</b>	<b>Count</b>	<b>Percent</b>	<b>Comments about usage</b>
BRAKER2 with ONT-RNA	16416	81.8%	default
PINFISH, whole pipeline	1129	5.6%	Had the most transcript options
BRAKER2_ref_proteins_aln	1054	5.2%	
StringTie, long mode with ONT-RNA	586	2.9%	Tended to over-fuse genes at overlapping UTRs
Manually edited genes	294	1.4%	
BRAKER2_PE-RNA	225	1.1%	
STRG_PE-RNA_hisat2	180	0.9%	
PINFISH_stranded	134	0.6%	
STRG_PE-RNA_stranded	30	0.1%	
PINFISH_stepwise	9	<0.1%	
STRG_ONT-RNA_minimap2	3	<0.1%	

**Supplemental Table 5 (additional file): Ortholog cluster BUSCO results. Summary of BUSCO scores for candidate metazoan genomes or transcriptomes, the 14 hexactinellid transcriptomes, and other genomes of outgroups, that were going to be used for the downstream ortholog clustering.**

**Supplemental Table 6: Final counts of genes used in the ortholog clustering analysis. Total genes refers to all proteins from that taxon. BLAST hits refers to those with any non-self hit (i.e. those not hitting the same protein, but would allow other proteins from the same species). One-to-one clusters refers to proteins that ended up in a cluster with no more than 1 protein per species; this criterion did not require that all species were represented.**

Species	code	Total genes	BLAST hits	Without BLAST hits	one-to-one-clusters
<i>Amphimedon queenslandica</i>	AQUE	43615	29742	13873	1175
<i>Aphrocallistes vastus</i>	AVAS	19514	16474	3040	4631
<i>Branchiostoma floridae</i>	BFLO	26529	15415	11114	1772
<i>Caenorhabditis elegans</i>	CELE	20191	8428	11763	345
<i>Ephydatia muelleri</i>	EMUE	39329	29422	9907	971
<i>Hormiphora californensis</i>	HCAL	14375	5616	8759	473
<i>Eurete sp</i>	HEX06-EUSP	20491	16162	4329	6062
<i>Farrea occa</i>	HEX07-FOCC	21569	16732	4837	6104
<i>Hoilungia hongkongensis</i>	HHON	12010	10599	1411	4548
<i>Homo sapiens</i>	HSAP	20375	12608	7767	1646
<i>Monosiga brevicollis</i>	MBRE	10864	3247	7617	380
<i>Nematostella vectensis</i>	NVEC	15355	8262	7093	1740
<i>Rhopilema esculentum</i>	RESC	17219	9931	7288	1494
<i>Sycon ciliatum</i>	SCIL	32061	9882	22179	753
<i>Trichoplax adhaerens</i>	TADH	12633	10157	2476	4560
<b>Totals</b>	<b>Totals</b>	326130	202677	123453	36654

**Supplemental Table 7 (additional file): Overview statistics from 224 genomes of various animals, used for Figure 1C.**

**Supplemental Table 8 (additional file): Sub-table of Supplemental Table 9, showing only the 419 differentially expressed genes, using p-adj > 1\*10<sup>-4</sup>.**

**Supplemental Table 9 (additional file): Complete version of Supplemental Table 8, showing all genes, as well as predicted functional annotations based on multiple pipelines. Columns were combined from the DEG analysis pipeline (columns would include base mean counts, log2fold change, P-value, and adjusted P-value ; see Methods) with the functional predictions from InterPro, EggNOG, and KEGG.**

**Supplemental Table 10 (additional file): Components and gene IDs of the Wnt signaling pathway, shown in Supplemental Figure 11.**