Supporting Information Legends

Figure S1. Additional analysis of ortholog gene expression

(A) The graph depicts the number of randomly selected genes with orthologs found in the genomes of 0, 1, 2, 3 or 4 of the examined species. (B) Graph depicts a reanalysis of the percentage of AE and random genes not expressed in each dataset using a higher threshold for gene expression (>5 FPKM). (C) Analysis of a second antennal transcriptome dataset from *Tribolium* revealed that a higher percentage of random gene orthologs are not expressed than AE gene orthologs. This is independent of the expression detection threshold (3 vs 5 FPKM). (D) Similar results are seen upon analysis of an antennal transcriptome from *Bombyx mori.*

File S1. Summary of RNA-Seq datasets

This dataset contains six spreadsheets summarizing our RNA-Seq data.

1) The spreadsheet "HTSeq" lists each of the 17,706 genes identified by HTSeq. The genes are listed by FlyBase symbol. The number of read fragments mapping to each gene is listed for each of three *amos* and three CS samples.

2) Highly expressed auditory organ genes were identified by averaging the mean microarray fluorescence intensities from the six "control" replicates of each of the auditory organ genes described in Table S2 of Senthilan et al., 2012. The 20 most highly expressed auditory organ genes are listed in the spreadsheet "auditory gene expression". We examined if each was expressed in our CS antennal samples, using our standard criteria of being detected at > 1 RPM in each of three samples. Seven of the 20 genes were not expressed, suggesting that the antennal second (auditory)

segment did not substantially contaminate our collection of the olfactory third antennal segments.

3) The spreadsheet "amos EdgeR analysis" summarizes the EdgeR differential expression analysis. First, the 9,462 genes analyzed by EdgeR are listed. These only include genes that are expressed >1 CPM in *amos* and/or CS antennae. The next column contains the log fold change (logFC) of expression in CS and *amos* samples and each gene's average expression in all six samples ($\log_2$(CPM). Finally, the results of the EdgeR differential expression analysis are provided with each gene's p-value for differential expression between *amos* and CS samples and the associated false discovery rate (FDR). The section highlighted in orange reports the 187 genes whose antennal expression was significantly reduced (FDR <0.01) in *amos* flies to a level more than 4-fold lower than CS flies. The section highlighted in blue reports the 154 genes whose antennal expression was significantly upregulated (FDR <0.01) in *amos* flies to a level more than 4-fold higher than CS flies. This includes 68 genes whose expression was only increased in *amos* sample 3, as described in the Methods.

4) The spreadsheet "*amos*-upregulated genes" lists the predicted molecular functions of the 86 genes consistently upregulated in *amos* mutants.

5) The spreadsheet "Cufflinks" lists each of the 17,615 genes identified by Cufflinks. The genes are listed by FlyBase symbol and FBgn ID. The estimated expression (FPKM) in each of three samples and their average expression is listed. Of these, 4,130 genes were expressed at >1 FPKM in all three samples and had average expression of >10 FPKM. These were used for identifying antennal-enriched genes.

6) The spreadsheet "antennal enriched genes" lists the 141 AE genes, their Cufflinks FPKM expression in our antennal transcriptome, and their expression in the seven whole body transcriptomes (yellow) and tissue transcriptomes (orange) generated by ModEncode.

Figure S1

A
Orthologs in genome

B
>5 FPKM

C
***Tribolium* Set 2**

D
Bombyx