Additional file 1

Supplementary Figures S1–S24

Perets D, Yakir E, Papathanos PA, Bohbot JD. Appendage identity dominates chemosensory transcriptomic organization in the black soldier fly *Hermetia illucens*.

Figures S1–S8: maximum-likelihood phylogenies of the chemosensory gene families. Figure S9: consensus appendage-biased gene sets. Figures S10–S13: Gene Ontology over-representation and domain composition of appendage-biased genes. Figures S14–S20: expression heatmaps of the seven chemosensory families. Figures S21 and S22: Gene Ontology over-representation among sex-biased and mating-responsive genes. Figure S23: volcano plots of every contrast. Figure S24: mated-female-specific expression against constitutive sex bias.

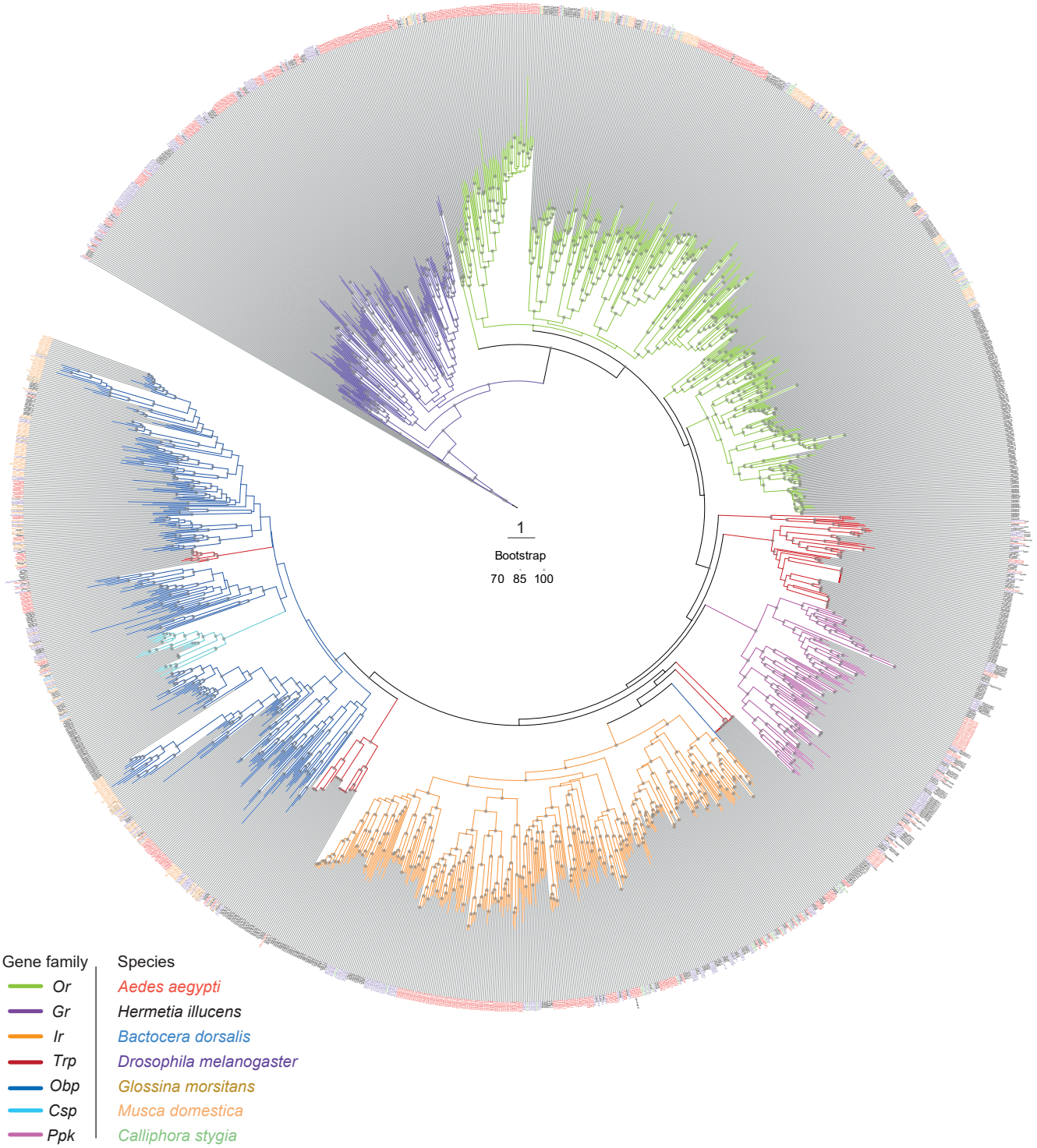


Figure S1. Phylogenetic reconstruction of the *H. illucens* chemosensory repertoire. Circular maximum-likelihood phylogeny of all 1,632 chemosensory proteins analyzed in this study, comprising 567 annotated *H. illucens* sequences and 1,065 homologous proteins from *Aedes aegypti*, *Drosophila melanogaster*, *Musca domestica*, *Calliphora stygia*, *Bactrocera dorsalis*, and *Glossina morsitans*. Branches are color-coded according to the seven conserved chemosensory protein families: odorant receptors (OR), gustatory receptors (GR), ionotropic receptors (IR), transient receptor potential channels (TRP), odorant-binding proteins (OBP), chemosensory proteins (CSP), and pickpocket channels (PPK). Because the seven families are not homologous to one another, the combined tree is an overview of repertoire composition, and relationships between families are not interpreted. Terminal labels identify individual proteins, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (70–100%), with node size proportional to support; values below 70 are omitted. This overview summarizes the phylogenetic relationships among all annotated chemosensory proteins and provides the basis for the detailed family-specific phylogenies presented in **Figures S2–S8**.

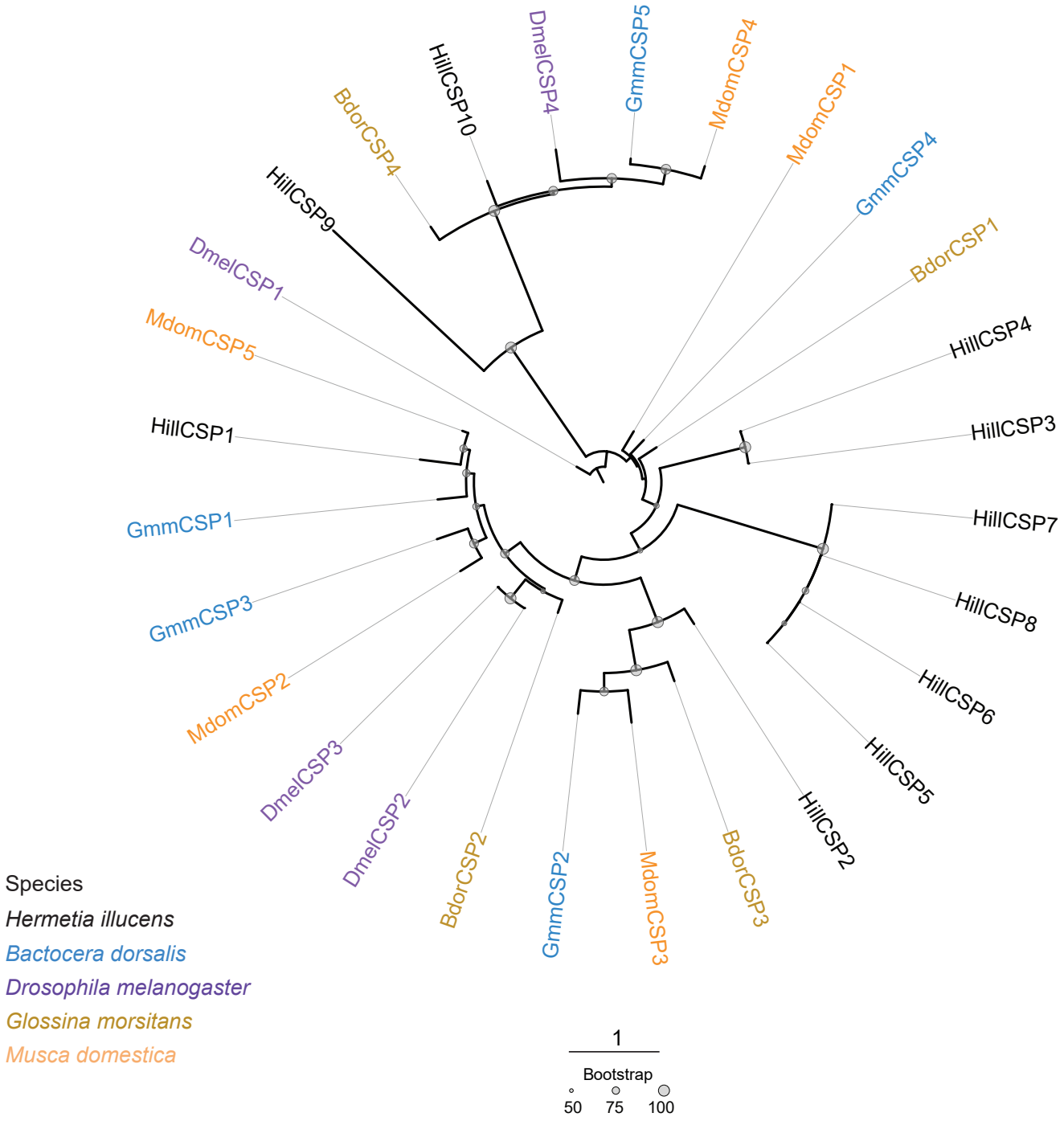


Figure S2. Maximum-likelihood phylogeny of the *H. illucens* chemosensory protein (CSP) family. Circular maximum-likelihood phylogeny of the 10 annotated *H. illucens* CSP proteins reconstructed together with homologous CSP sequences from *Drosophila melanogaster*, *Musca domestica*, *Bactrocera dorsalis*, and *Glossina morsitans*. Terminal labels identify individual proteins, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (50–100%), with node size proportional to support; values below 50 are omitted. The phylogeny identifies orthologous relationships and lineage-specific expansions within the CSP family, providing the basis for the annotation and nomenclature of *H. illucens Csp* genes.

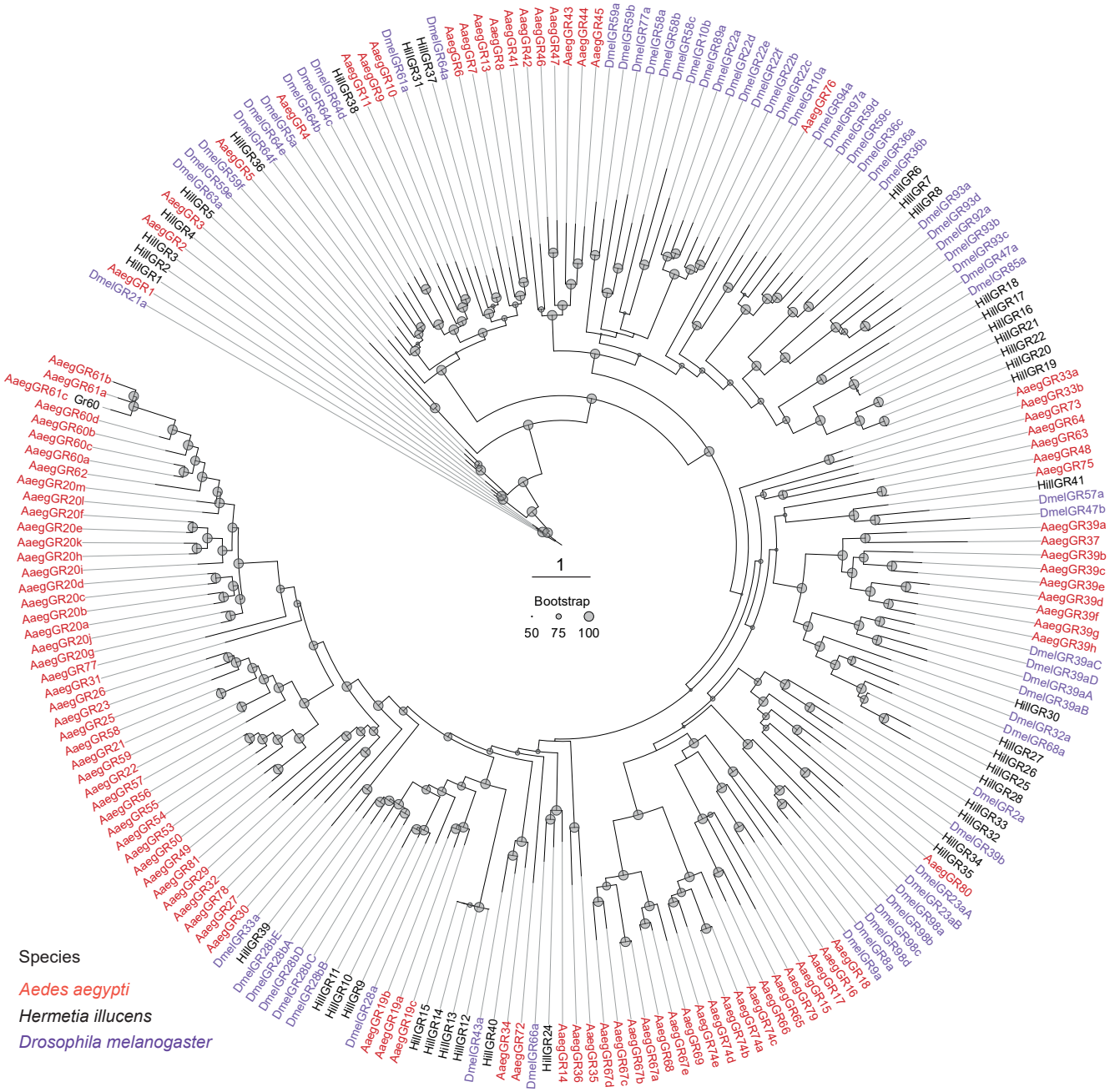


Figure S3. Maximum-likelihood phylogeny of the *H. illucens* gustatory receptor (GR) family. Circular maximum-likelihood phylogeny of the 39 annotated *H. illucens* gustatory receptor (GR) proteins reconstructed together with homologous GR sequences from *Aedes aegypti* and *Drosophila melanogaster*. Terminal labels identify individual receptors, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (50–100%), with node size proportional to support; values below 50 are omitted. The phylogeny identifies orthologous relationships between *H. illucens* and well-characterized dipteran GRs, including the conserved carbon dioxide receptor clade (HillGR1–HillGR5), the fructose receptor ortholog HillGR40 (*Gr43a*), and the bitter receptor orthologs HillGR24 (*Gr66a*) and HillGR39 (*Gr33a*), providing the basis for functional annotation of the *H. illucens* GR repertoire.



Figure S4. Maximum-likelihood phylogeny of the *H. illucens* ionotropic receptor (IR) family. Circular maximum-likelihood phylogeny of the 101 annotated *H. illucens* ionotropic receptor (IR) proteins reconstructed together with homologous IR sequences from *Aedes aegypti*, *Drosophila melanogaster*, and *Calliphora stygia*. Terminal labels identify individual receptors, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (53–100%), with node size proportional to support; values below 53 are omitted. The phylogeny identifies conserved co-receptors (*Ir8a*, *Ir25a* and *Ir76b*), antennal tuning receptors, and lineage-specific expansions within the *H. illucens* IR repertoire, providing the basis for functional annotation and gene nomenclature.

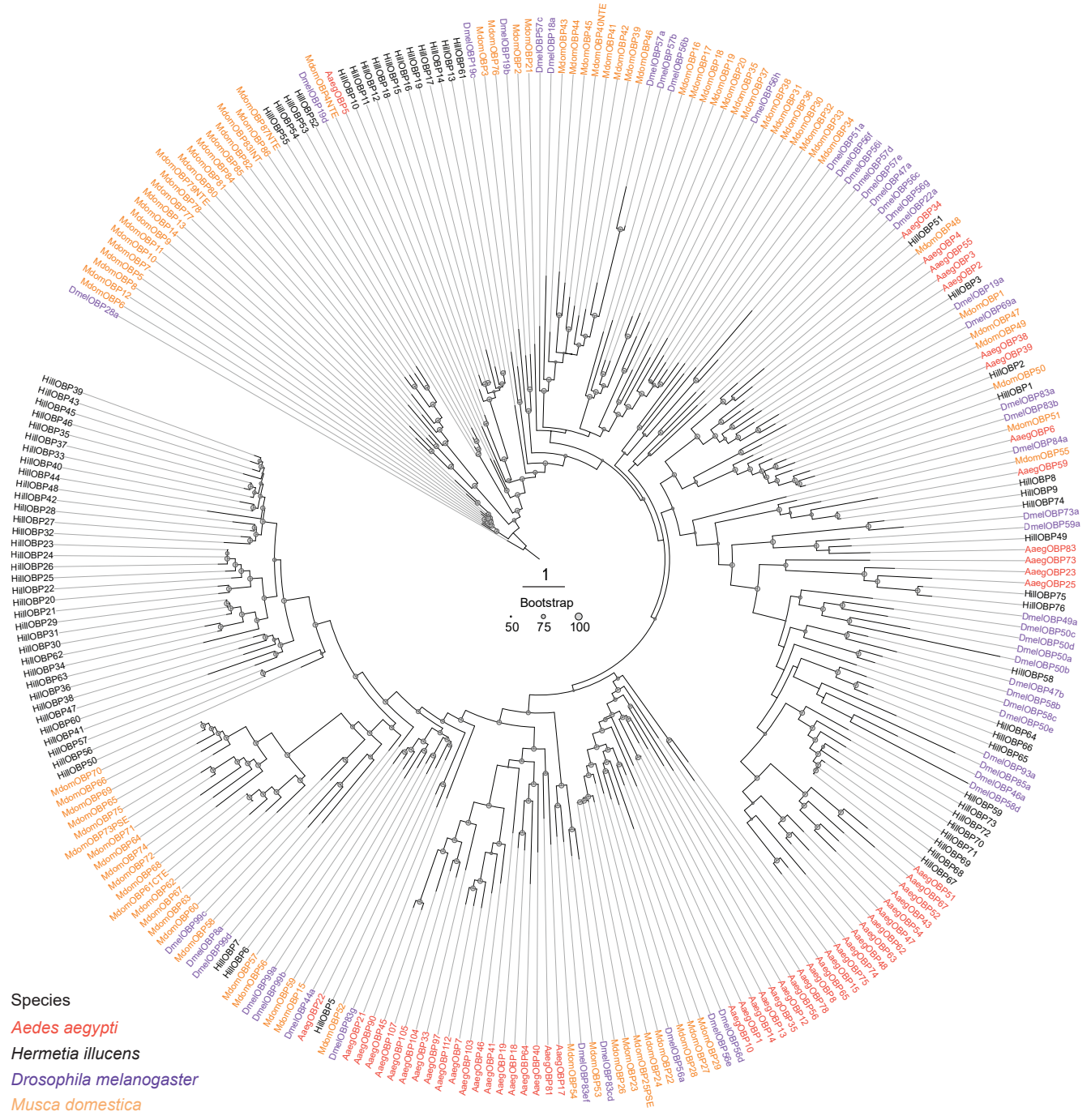


Figure S5. Maximum-likelihood phylogeny of the *H. illucens* odorant-binding protein (OBP) family. Circular maximum-likelihood phylogeny of the 75 annotated *H. illucens* odorant-binding protein (OBP) sequences reconstructed together with homologous OBPs from *Aedes aegypti*, *Drosophila melanogaster*, and *Musca domestica*. Terminal labels identify individual proteins, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (50–100%), with node size proportional to support; values below 50 are omitted. The phylogeny identifies orthologous relationships and lineage-specific expansions within the *H. illucens* OBP family and provides the basis for gene annotation and nomenclature.

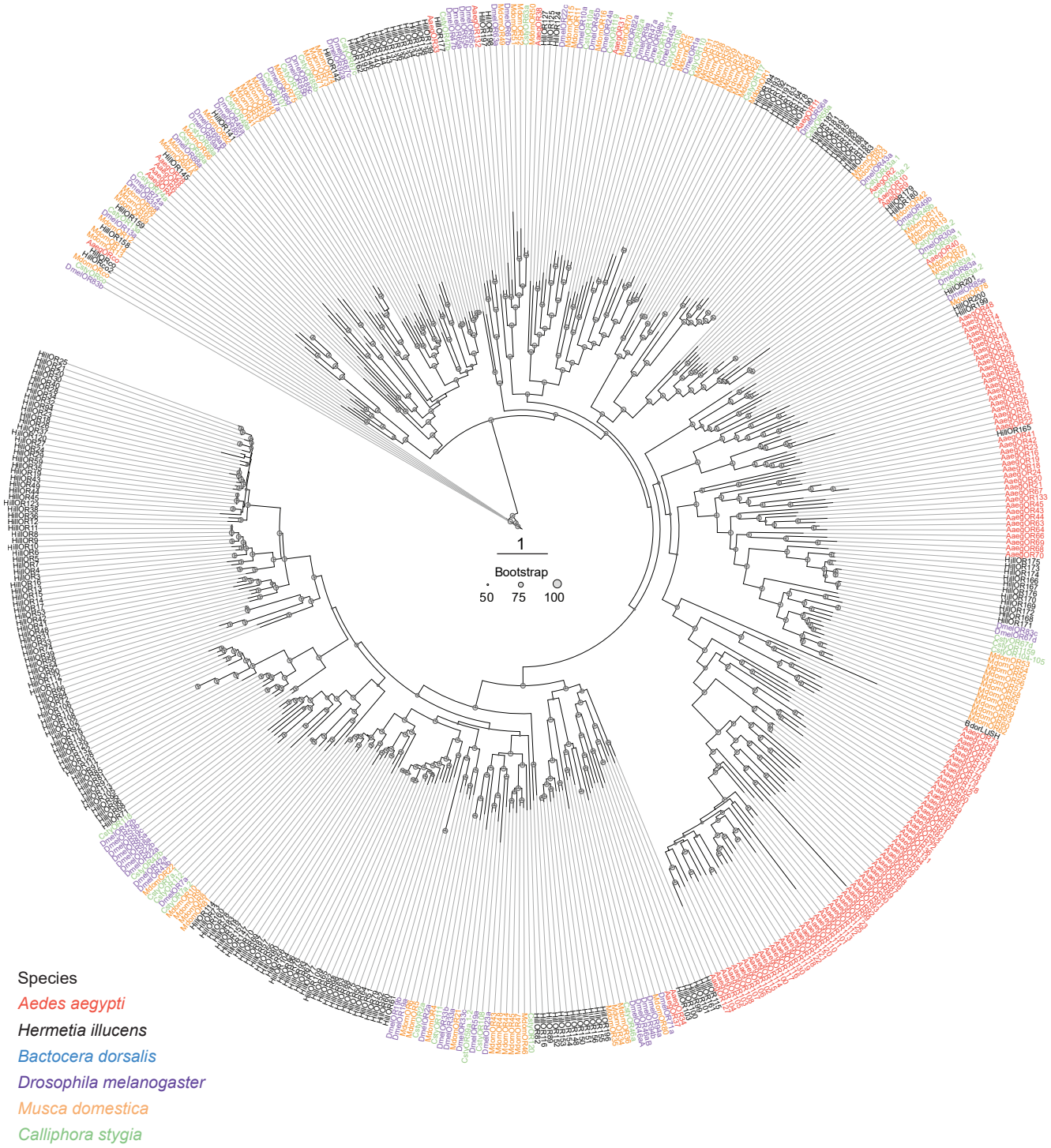


Figure S6. Maximum-likelihood phylogeny of the *H. illucens* odorant receptor (OR) family. Circular maximum-likelihood phylogeny of the 192 annotated *H. illucens* odorant receptor (OR) proteins reconstructed together with homologous OR sequences from *Aedes aegypti*, *Drosophila melanogaster*, *Musca domestica*, and *Calliphora stygia*. Terminal labels identify individual receptors, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (50–100%), with node size proportional to support; values below 50 are omitted. The phylogeny identifies the conserved odorant receptor co-receptors (HillORco and HillORco2) together with extensive lineage-specific expansion of the OR repertoire, providing the basis for functional annotation and evolutionary comparisons.



Figure S7. Maximum-likelihood phylogeny of the *H. illucens* pickpocket (PPK) family. Circular maximum-likelihood phylogeny of the 39 annotated *H. illucens* pickpocket (PPK) proteins reconstructed together with homologous PPK sequences from *Aedes aegypti* and *Drosophila melanogaster*. Terminal labels identify individual proteins, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (50–100%), with node size proportional to support; values below 50 are omitted. The phylogeny identifies orthologous relationships within the PPK family, including the *ppk23* pheromone-sensing lineage and the *Nach* lineage, providing the basis for functional annotation of the *H. illucens* PPK repertoire.

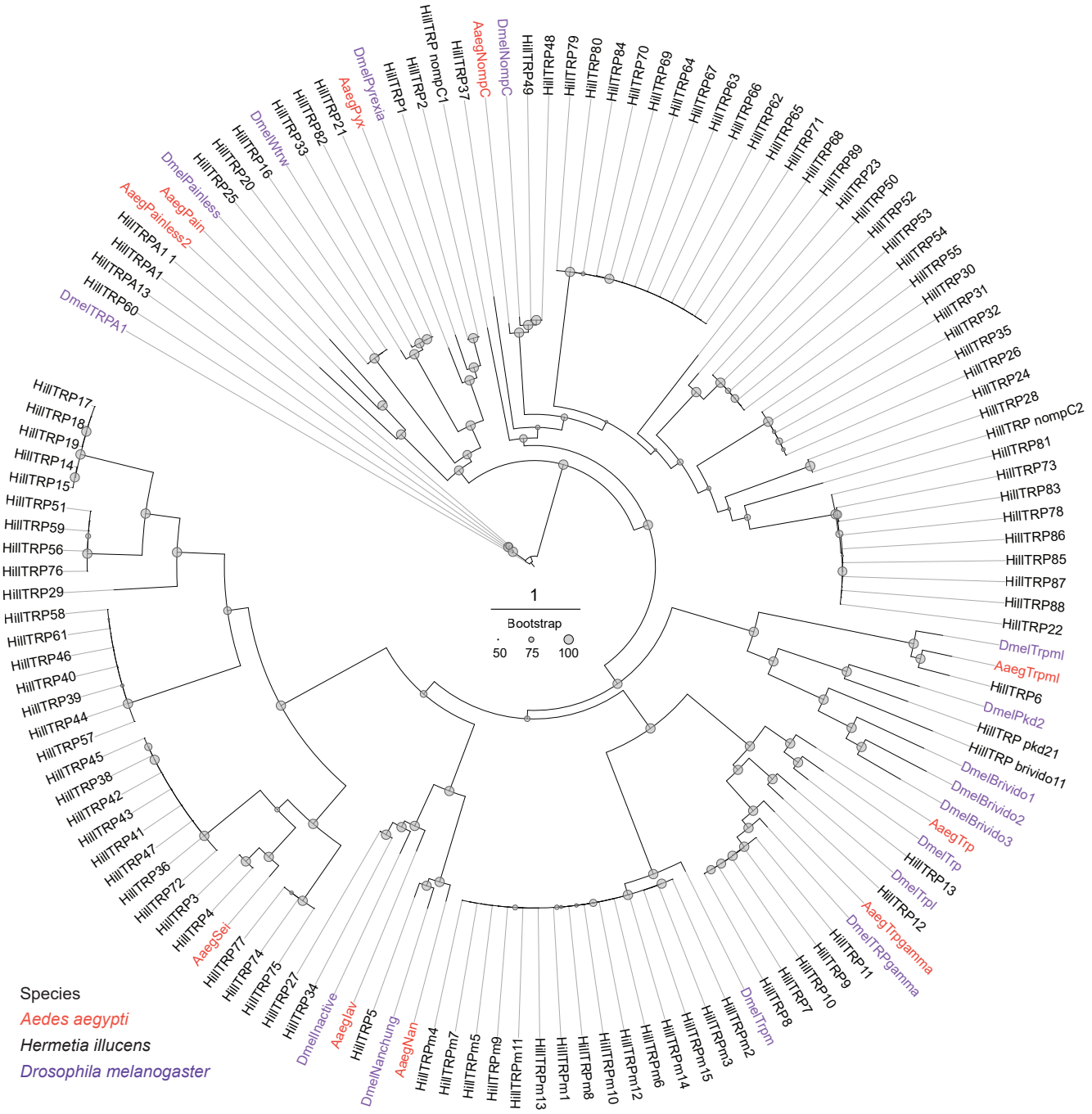


Figure S8. Maximum-likelihood phylogeny of the *H. illucens* transient receptor potential (TRP) family. Circular maximum-likelihood phylogeny of the 111 annotated *H. illucens* transient receptor potential (TRP) proteins reconstructed together with homologous TRP sequences from *Aedes aegypti* and *Drosophila melanogaster*. Terminal labels identify individual proteins, with species indicated by label color. Gray circles at internal nodes indicate bootstrap support (50–100%), with node size proportional to support; values below 50 are omitted. The phylogeny resolves the major TRP subfamilies, including TRPA, TRPM, TRPN (NompC), TRPC, TRPV, TRPML, PKD, and Brivido channels, highlighting lineage-specific expansions within the *H. illucens* TRP repertoire and providing the basis for functional annotation.

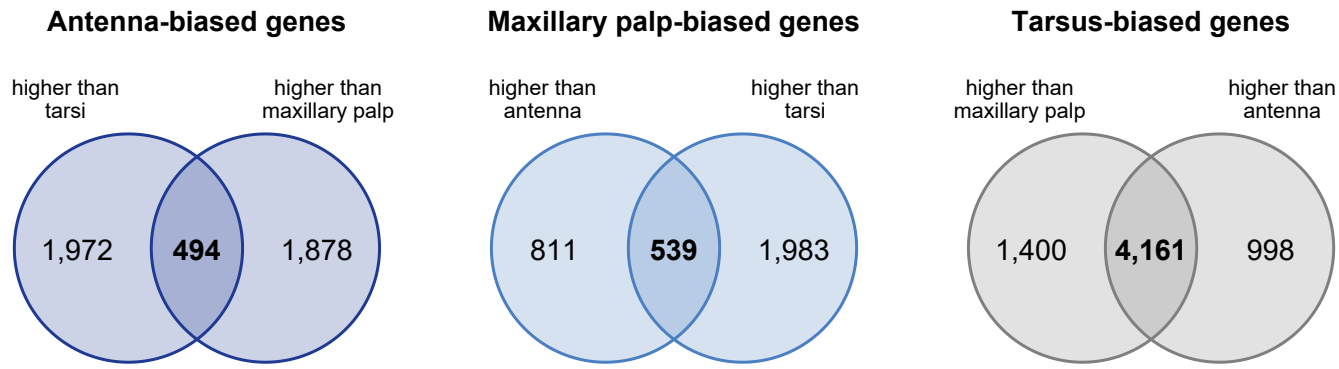


Figure S9. Consensus appendage-biased gene sets. For each appendage, the genes significantly higher in that appendage than in each of the other two (adjusted $p < 0.001$, $|\log_2 FC| \geq 1$) are compared between its two pairwise contrasts. The intersection defines the consensus appendage-biased set: 494 antenna-biased, 539 maxillary palp-biased and 4,161 tarsus-biased genes. These sets are tested for Gene Ontology over-representation in **Figs. S10–S12**.

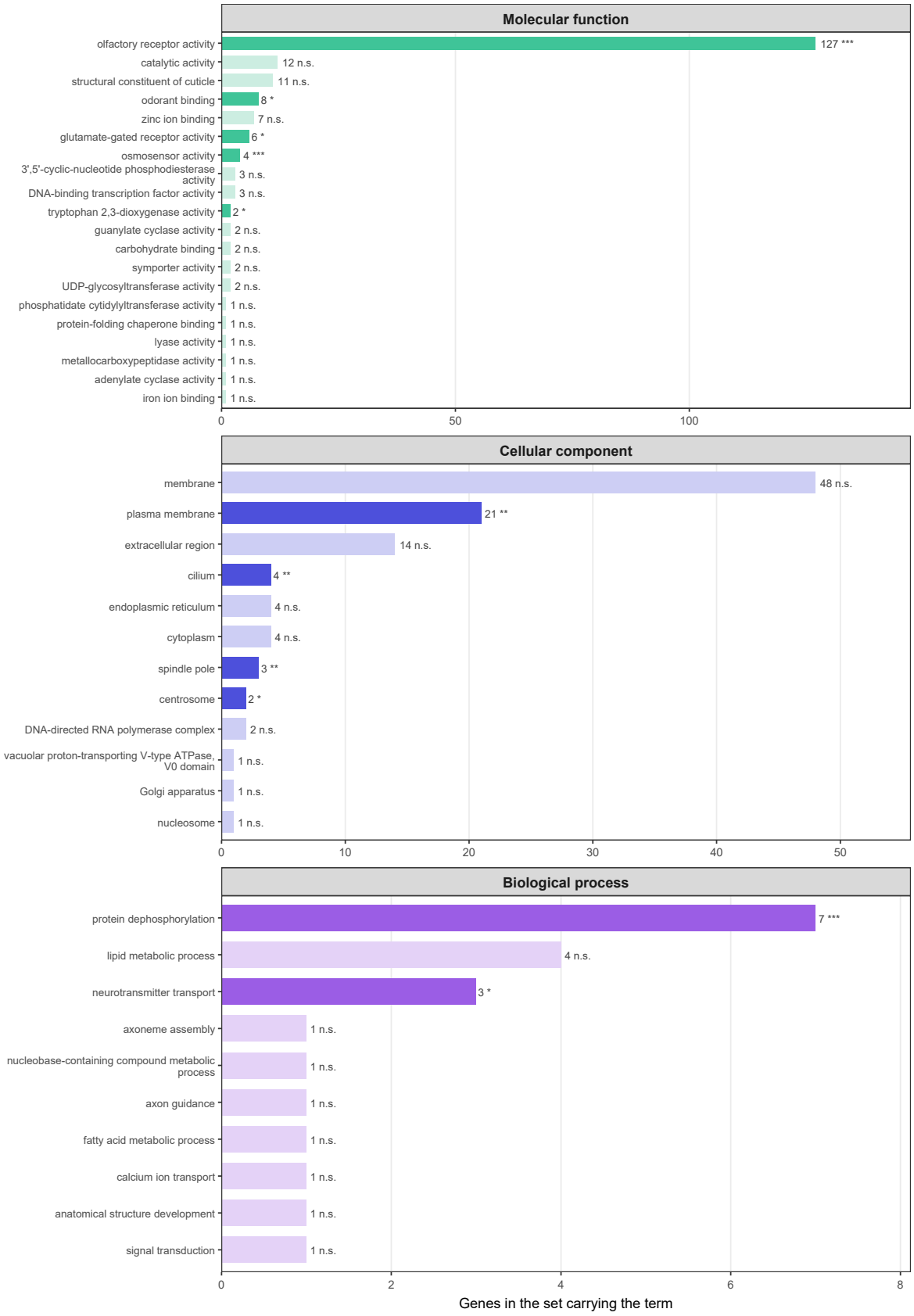


Figure S10. Gene Ontology over-representation among antenna-biased genes. The 494 consensus antenna-biased genes, those significantly higher in the antenna than in both other appendages, were tested for GO term over-representation. Terms were tested with a one-sided Fisher's exact test for over-representation against all annotated genes tested in all three pairwise appendage contrasts; each gene carries a single GO term (**Additional file 39**), and *p* values were corrected by the Benjamini–Hochberg procedure across every background term within each GO domain. The 20 terms with the lowest adjusted p per domain are shown. Bar length, and the number beside each bar, is the count of genes in the set carrying the term; stars give the adjusted p (*** < 0.001, ** < 0.01, * < 0.05, n.s. not significant). Bars are colored by GO domain (Molecular function green, Cellular component blue, Biological process purple) and drawn in a pale tint where the term did not reach q < 0.05. Terms are ordered by gene count.

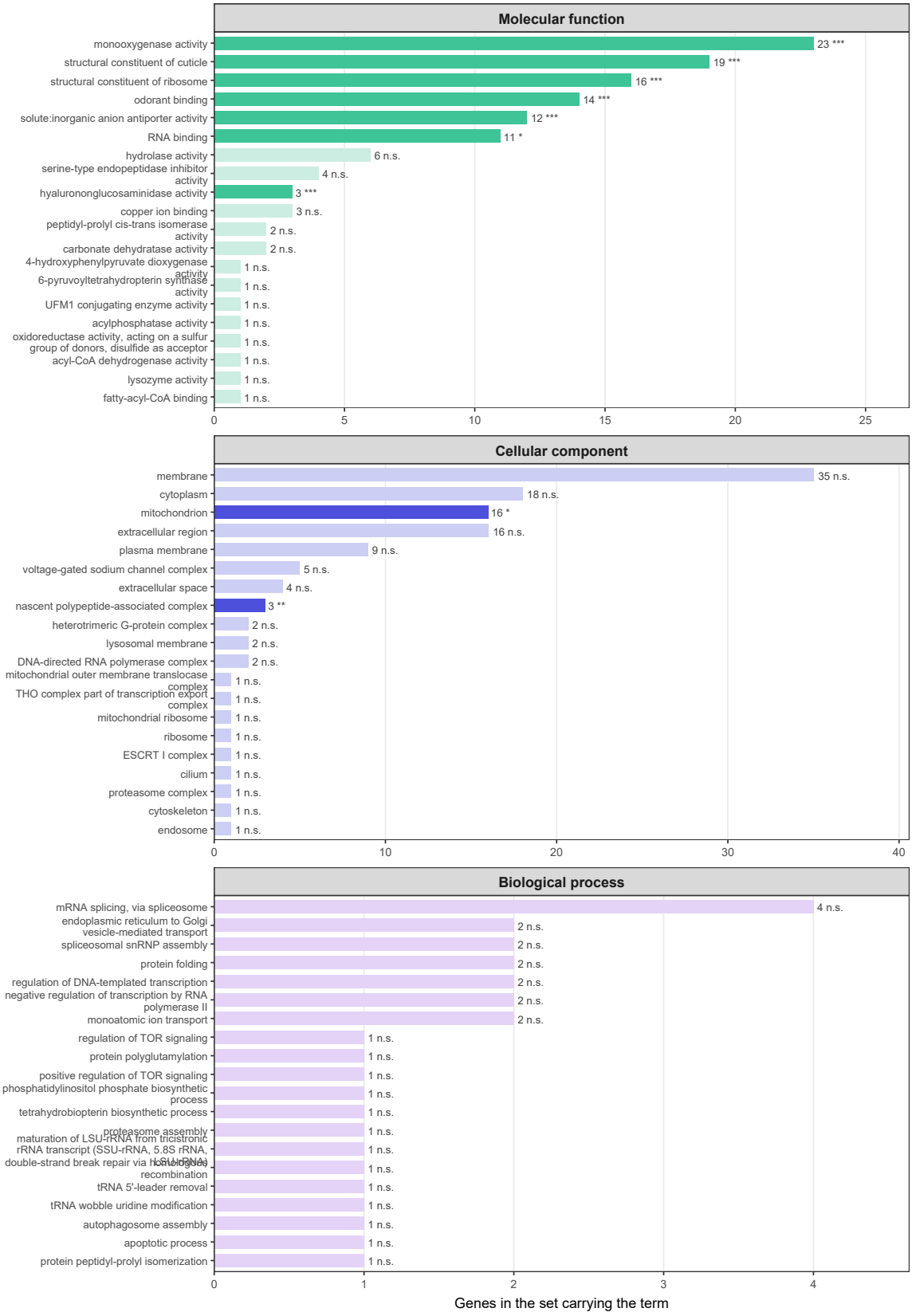


Figure S11. Gene Ontology over-representation among maxillary palp-biased genes. The 539 consensus maxillary palp-biased genes were tested for GO term over-representation. Terms were tested with a one-sided Fisher's exact test for over-representation against all annotated genes tested in all three pairwise appendage contrasts; each gene carries a single GO term (**Additional file 39**), and *p* values were corrected by the Benjamini–Hochberg procedure across every background term within each GO domain. The 20 terms with the lowest adjusted p per domain are shown. Bar length, and the number beside each bar, is the count of genes in the set carrying the term; stars give the adjusted p (*** < 0.001, ** < 0.01, * < 0.05, n.s. not significant). Bars are colored by GO domain (Molecular function green, Cellular component blue, Biological process purple) and drawn in a pale tint where the term did not reach q < 0.05. Terms are ordered by gene count. Unlike the antenna, the maxillary palp shows a heterogeneous functional profile: the most strongly over-represented terms are solute:inorganic anion antiporter activity and odorant binding, with structural constituents of the ribosome and cuticle and monooxygenase activity also enriched.

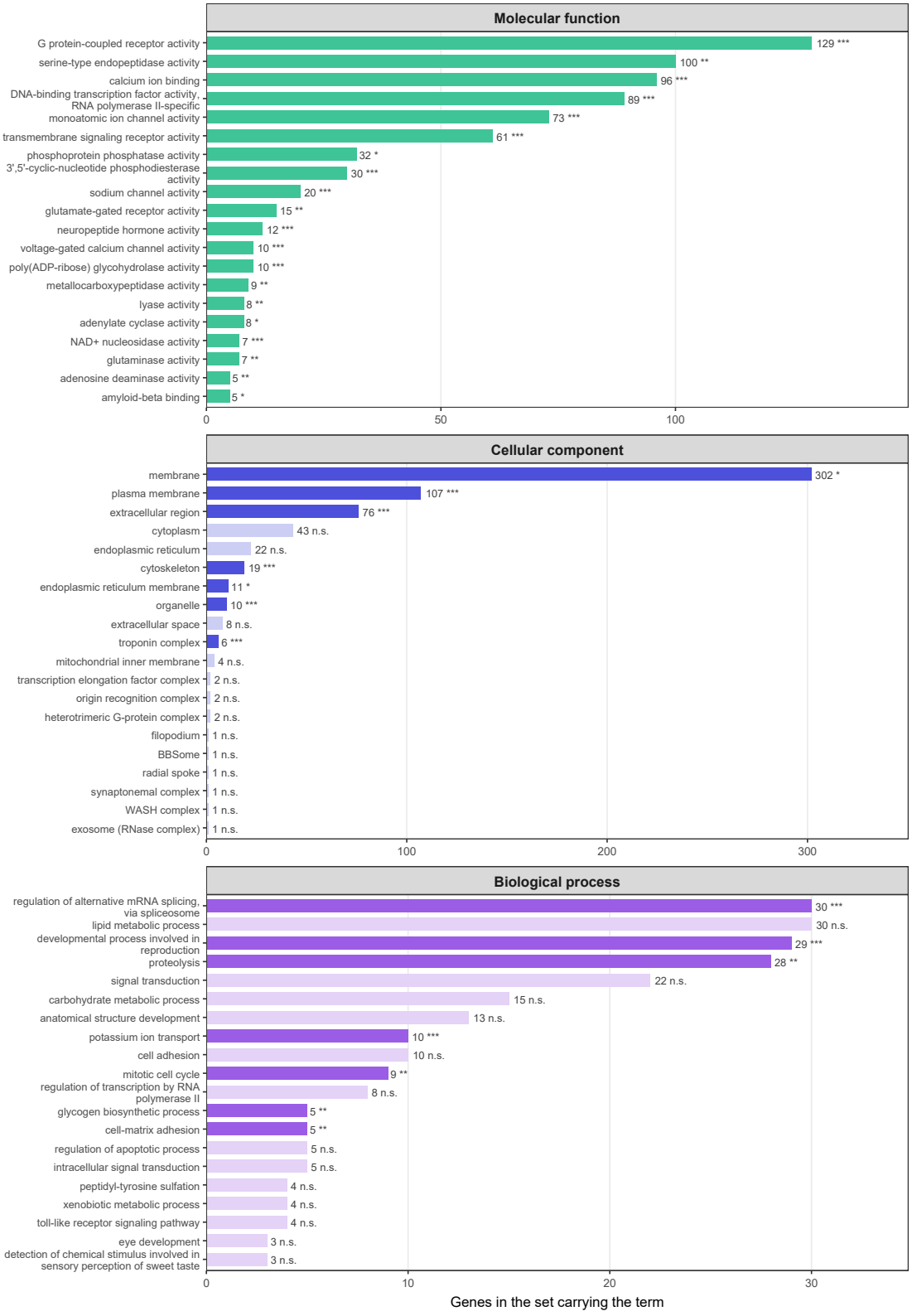


Figure S12. Gene Ontology over-representation among tarsus-biased genes. The 4,161 consensus tarsus-biased genes were tested for GO term over-representation. Terms were tested with a one-sided Fisher's exact test for over-representation against all annotated genes tested in all three pairwise appendage contrasts; each gene carries a single GO term (**Additional file 39**), and *p* values were corrected by the Benjamini–Hochberg procedure across every background term within each GO domain. The 20 terms with the lowest adjusted p per domain are shown. Bar length, and the number beside each bar, is the count of genes in the set carrying the term; stars give the adjusted p (*** < 0.001, ** < 0.01, * < 0.05, n.s. not significant). Bars are colored by GO domain (Molecular function green, Cellular component blue, Biological process purple) and drawn in a pale tint where the term did not reach q < 0.05. Terms are ordered by gene count. The tarsus has the largest appendage-biased repertoire, but over-representation is confined to signaling and regulatory functions, principally G protein-coupled receptor activity, DNA-binding transcription factor activity, regulation of alternative mRNA splicing and neuropeptide hormone activity. Protein binding and nucleotide binding are the numerically largest categories in this set but are not over-represented against the annotated background (q = 0.77 and q = 1.00).

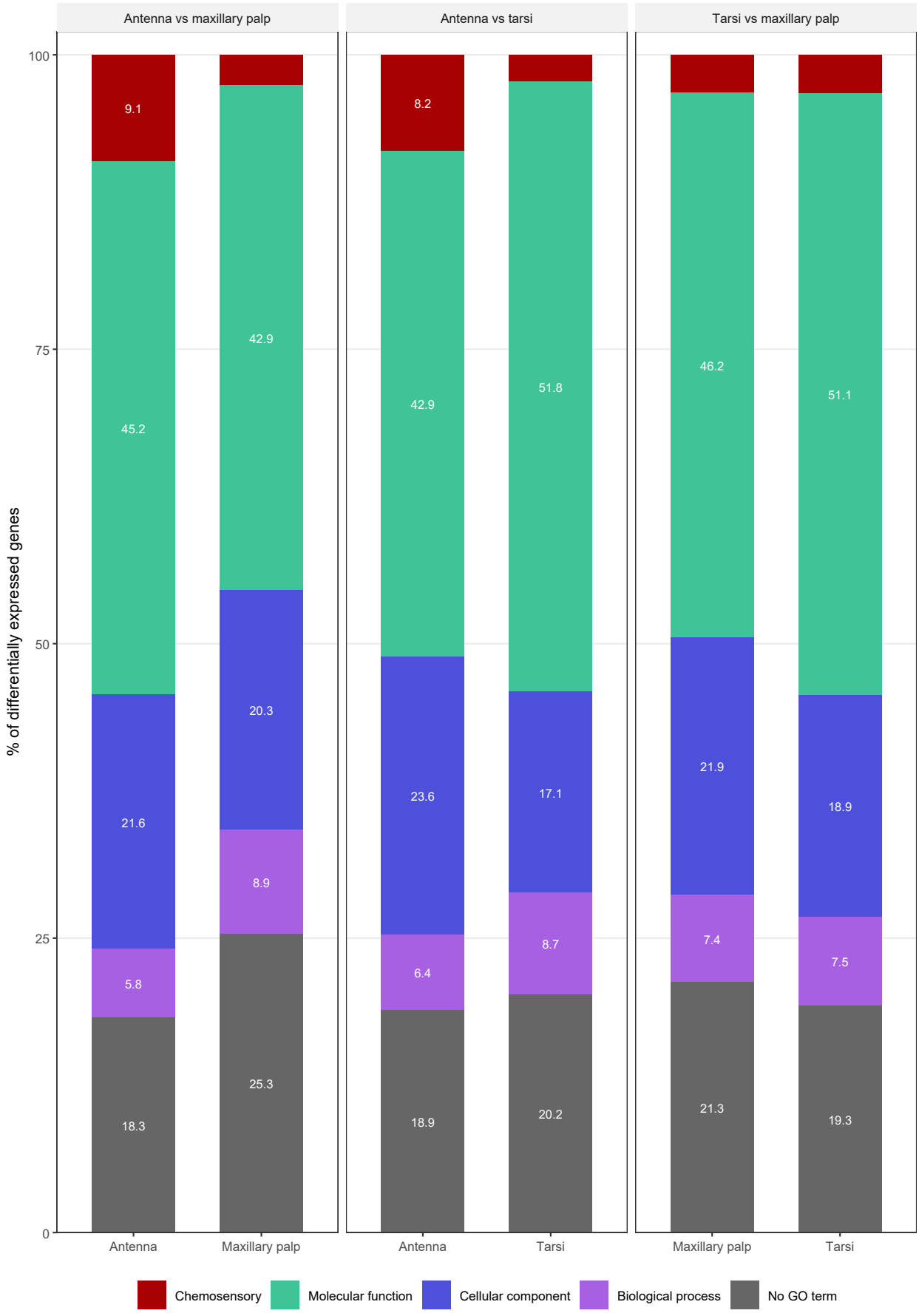


Figure S13. Gene Ontology domain composition of appendage-biased genes. For each pairwise appendage comparison, the differentially expressed genes biased to either appendage (adjusted $p < 0.001$, $|\log_2 FC| \geq 1$) are broken down by Gene Ontology domain, with chemosensory genes shown as a separate category. This breakdown is descriptive: **Additional file 39** assigns exactly one GO term to each transcript, so the domain split reflects how the genome was annotated rather than a tested property of the gene sets. Statistically tested over-representation is reported in **Figures S10–S12**, and the chemosensory enrichment tests, with odds ratios and 95% confidence intervals, in **Additional file 13**.

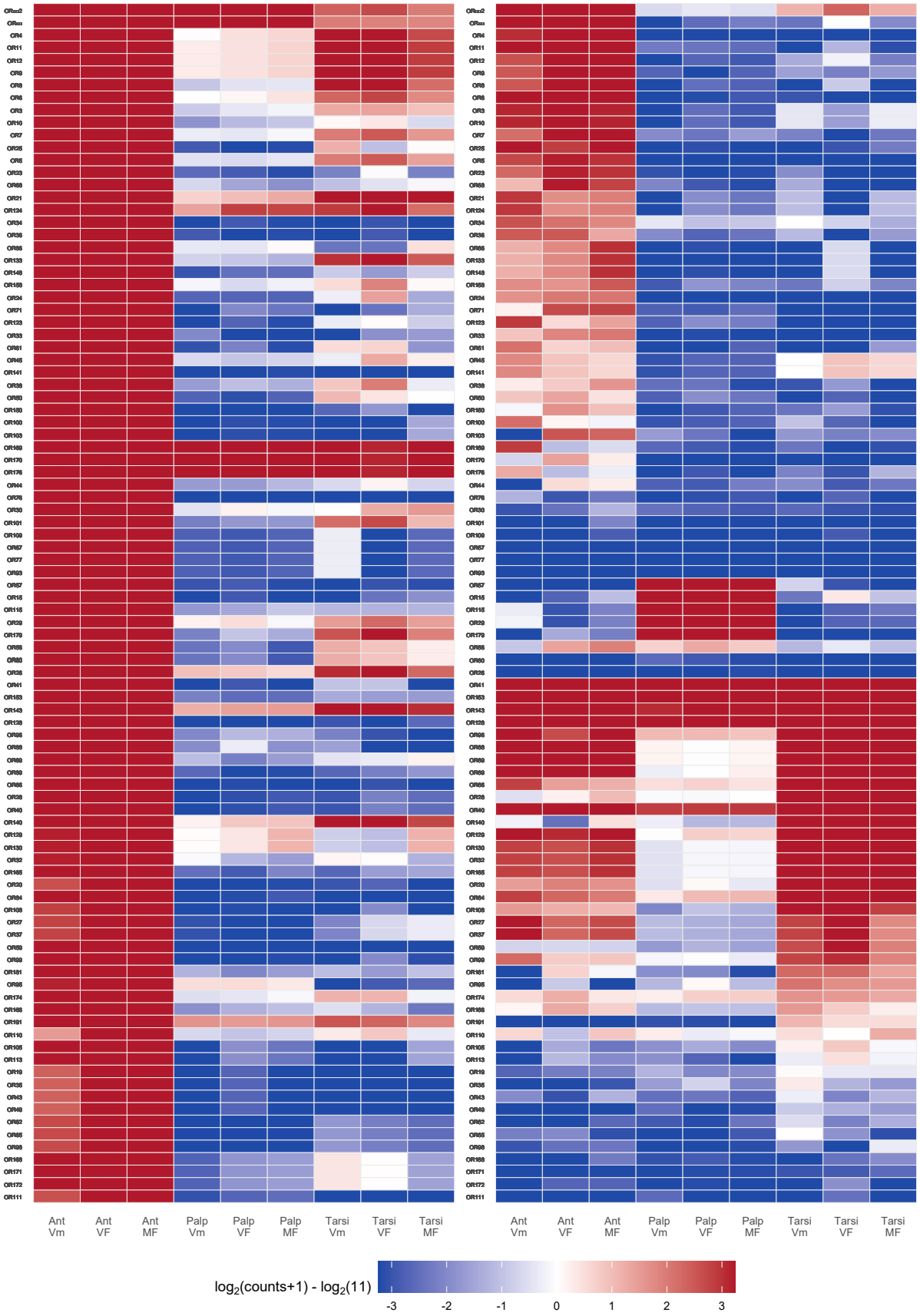


Figure S14. Expression heatmap of the *H. illucens* odorant receptor (*Or*) family. Heatmap showing normalized expression of the 192 annotated odorant receptor (*Or*) transcript models across the antennae, maxillary palps, and tarsi of virgin males (Vm), virgin females (VF), and mated females (MF). Expression values are displayed as $\log_2$(counts + 1) and centered on a threshold of 10 normalized counts (white = 10 counts, blue = lower expression, red = higher expression). Transcript models are ordered according to the appendage showing the highest mean expression (antenna → maxillary palp → tarsi), and subsequently by expression level within each appendage. The heatmap illustrates the pronounced enrichment of *Or* expression in the antennae, with a smaller subset of appendage-specific receptors expressed in the maxillary palps and relatively few enriched in the tarsi.

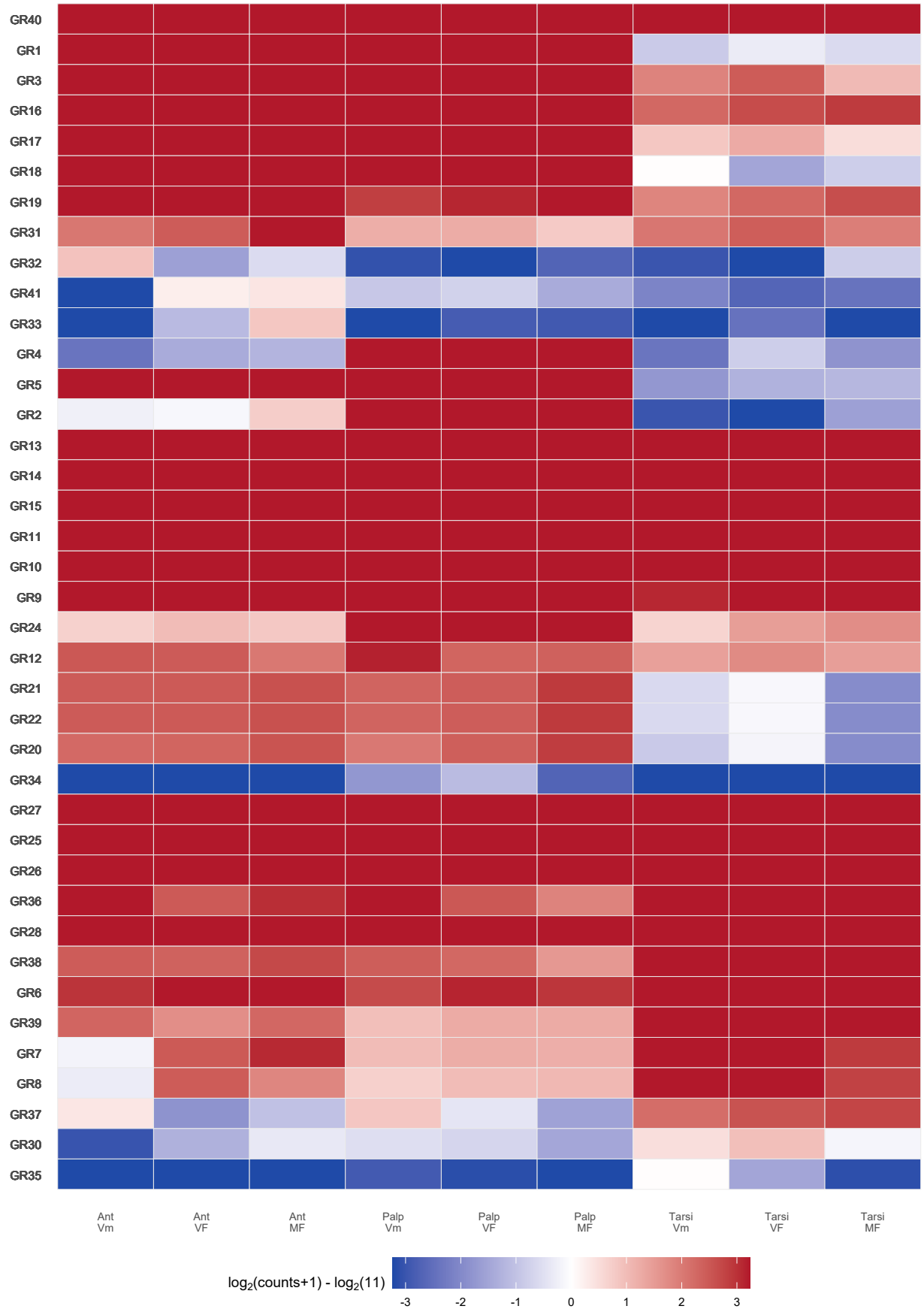


Figure S15. Expression heatmap of the *H. illucens* gustatory receptor (*Gr*) family. Heatmap showing normalized expression of the 39 annotated gustatory receptor (*Gr*) transcript models across the antennae, maxillary palps, and tarsi of virgin males (Vm), virgin females (VF), and mated females (MF). Expression values are displayed as $\log_2$(counts + 1) and centered on a threshold of 10 normalized counts (white = 10 counts). Transcript models are ordered by appendage of highest mean expression and then by expression level. The heatmap highlights distinct appendage-specific *Gr* subsets, including receptors preferentially expressed in the maxillary palps and tarsi, consistent with their roles in CO_2 detection and contact chemosensation.

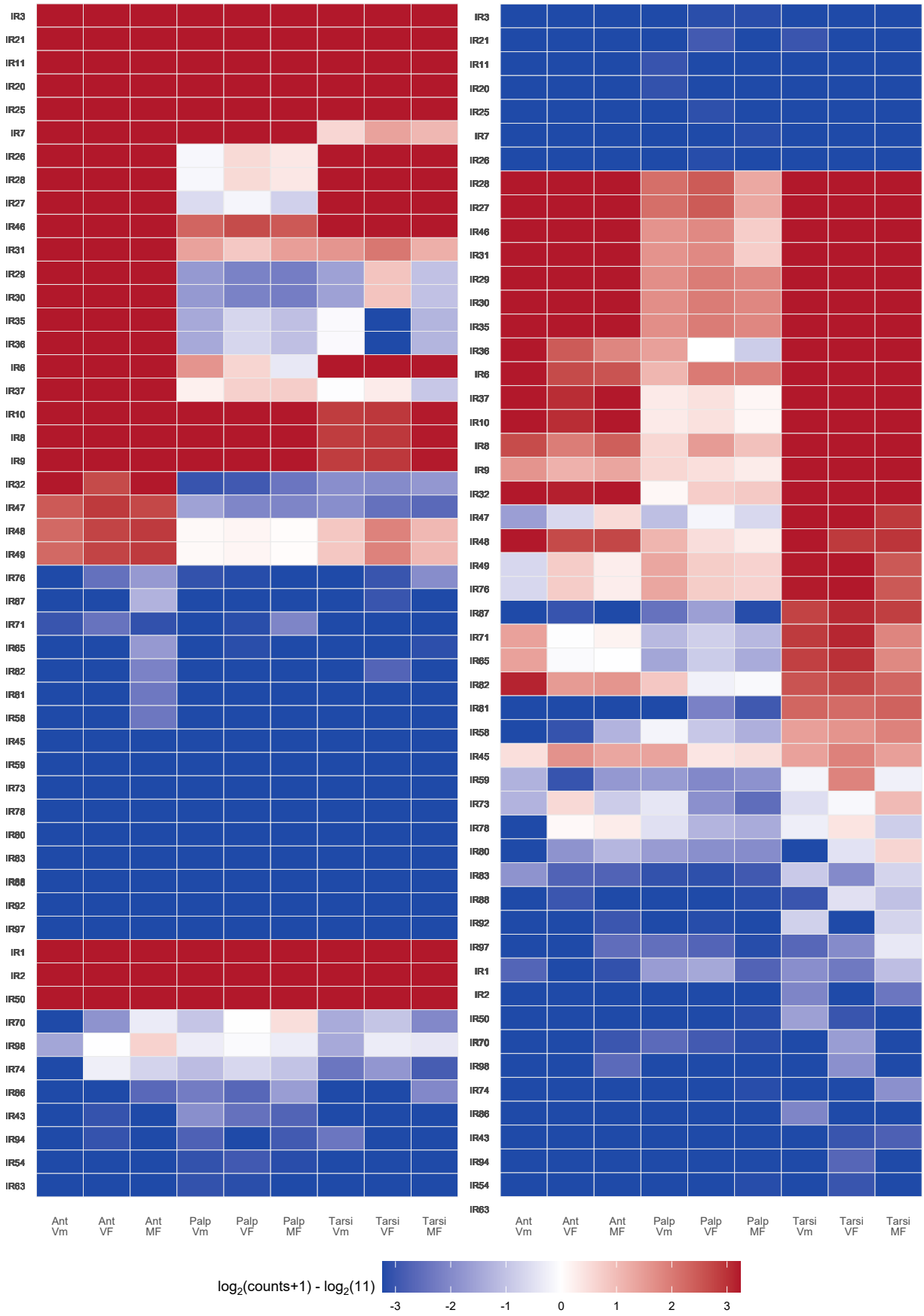


Figure S16. Expression heatmap of the *H. illucens* ionotropic receptor (*Ir*) family. Heatmap showing normalized expression of the 101 annotated ionotropic receptor (*Ir*) transcript models across the antennae, maxillary palps, and tarsi of virgin males (Vm), virgin females (VF), and mated females (MF). Expression values are displayed as $\log_2$(counts + 1) and centered on a threshold of 10 normalized counts (white = 10 counts). Transcript models are ordered by appendage of highest mean expression and then by expression level. The heatmap shows that the conserved co-receptors *Ir8a* and *Ir25a* are expressed in all three appendages, most highly in the antenna, together with appendage-specific receptor subsets enriched in the antennae and tarsi.

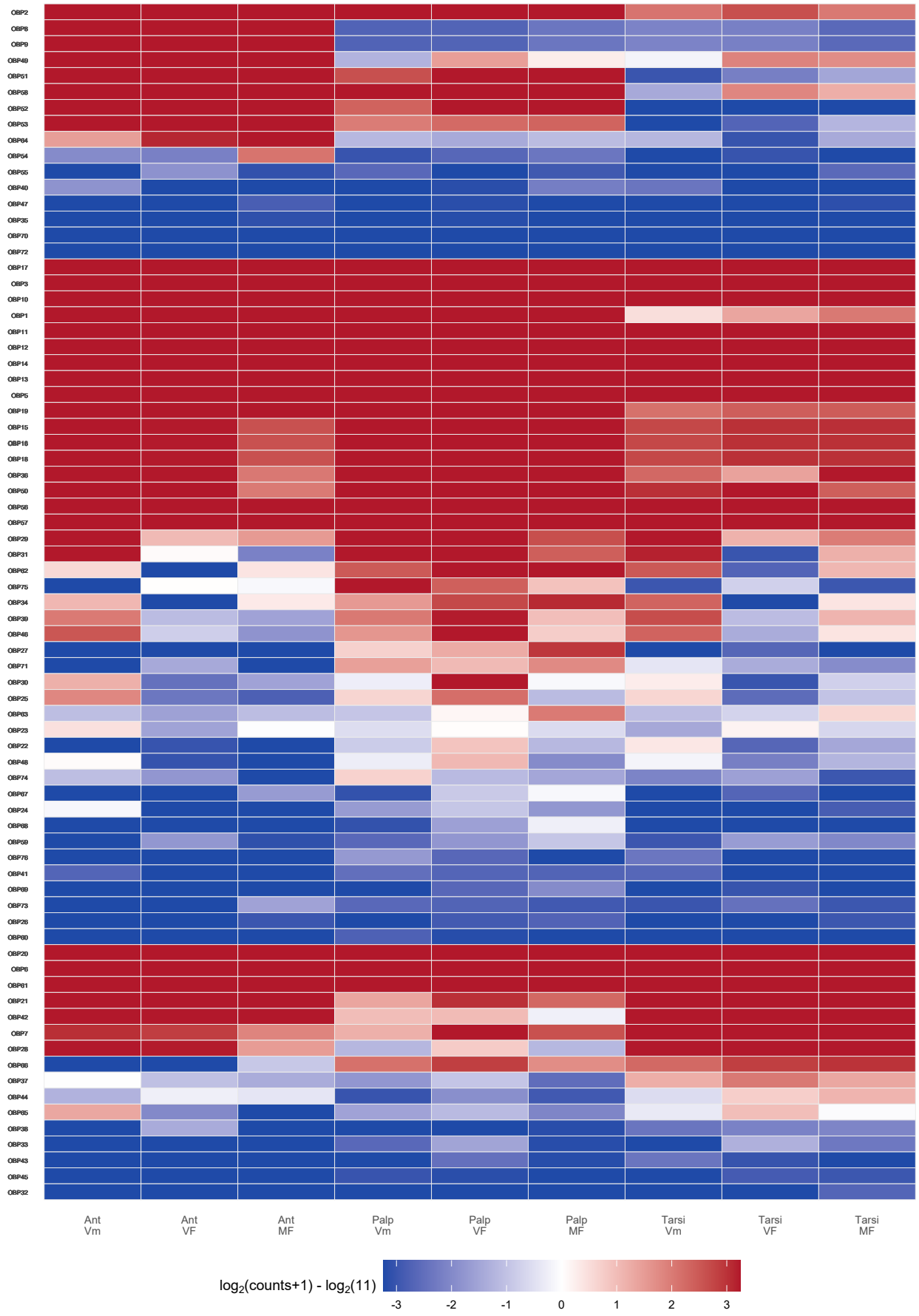


Figure S17. Expression heatmap of the *H. illucens* odorant-binding protein (*Obp*) family. Heatmap showing normalized expression of the 75 annotated odorant-binding protein (*Obp*) transcript models across the antennae, maxillary palps, and tarsi of virgin males (Vm), virgin females (VF), and mated females (MF). Expression values are displayed as $\log_2$(counts + 1) and centered on a threshold of 10 normalized counts (white = 10 counts). Transcript models are ordered by appendage of highest mean expression and then by expression level. Distinct antennal-, maxillary palp-, and tarsus-enriched *Obp* clusters are apparent, illustrating the extensive appendage-specific diversification of odorant-binding proteins.

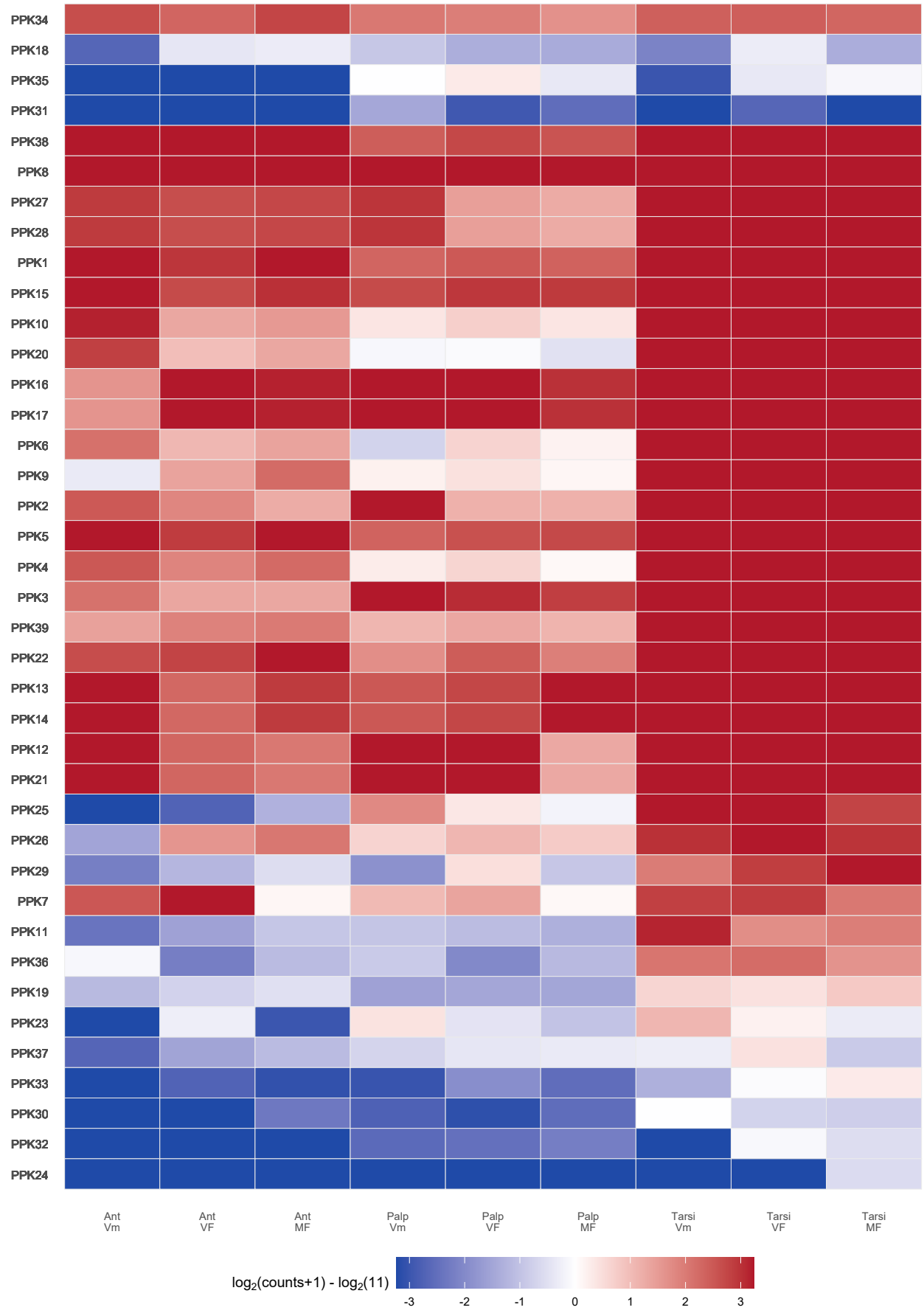


Figure S18. Expression heatmap of the *H. illucens* pickpocket (*Ppk*) family. Heatmap showing normalized expression of the 39 annotated pickpocket (*Ppk*) transcript models across the antennae, maxillary palps, and tarsi of virgin males (Vm), virgin females (VF), and mated females (MF). Expression values are displayed as $\log_2$(counts + 1) and centered on a threshold of 10 normalized counts (white = 10 counts). Transcript models are ordered by appendage of highest mean expression and then by expression level. Most *Ppk* genes exhibit preferential expression in the tarsi, consistent with their established roles in contact chemosensation and mechanosensory function.

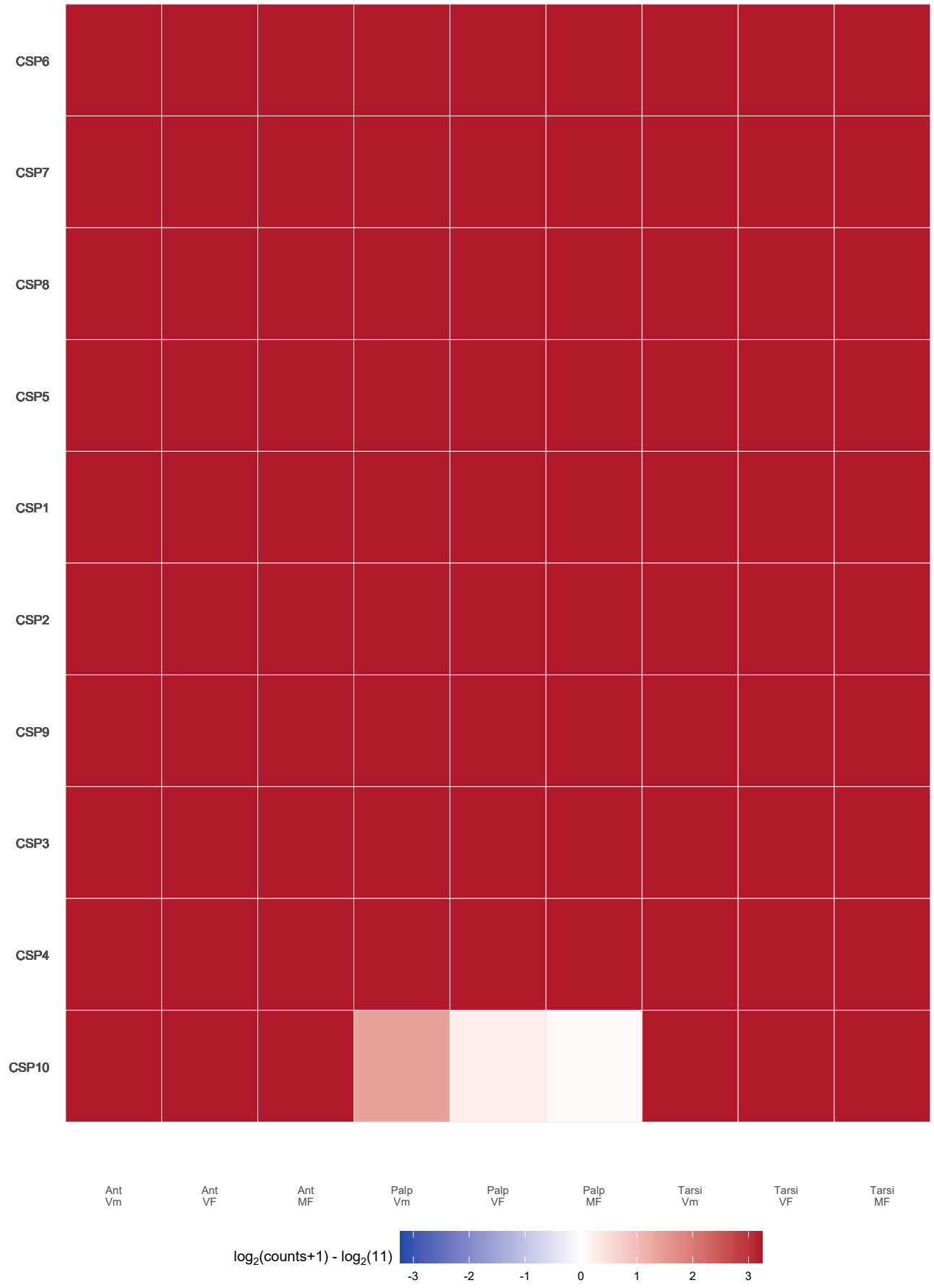


Figure S19. Expression heatmap of the *H. illucens* chemosensory protein (*Csp*) family. Heatmap showing normalized expression of the 10 annotated chemosensory protein (*Csp*) transcript models across the antennae, maxillary palps, and tarsi of virgin males (Vm), virgin females (VF), and mated females (MF). Expression values are displayed as $\log_2$(counts + 1) and centered on a threshold of 10 normalized counts (white = 10 counts). Transcript models are ordered by appendage of highest mean expression and then by expression level. In contrast to other chemosensory families, most *Csp* genes exhibit broad expression across all appendages, suggesting more generalized functions in the peripheral sensory system.

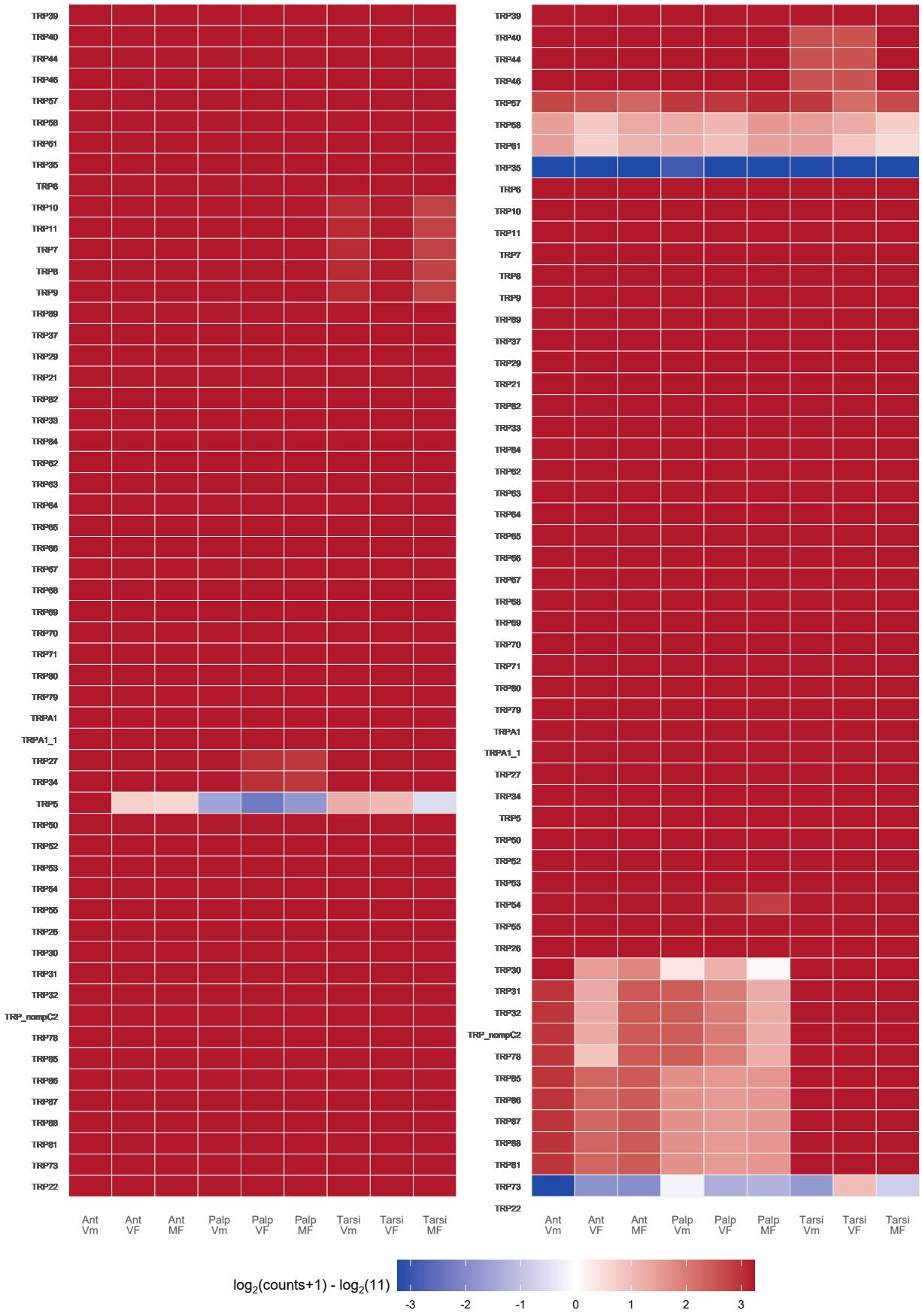


Figure S20. Expression heatmap of the *H. illucens* transient receptor potential (*Trp*) family. Heatmap showing normalized expression of the 111 annotated transient receptor potential (*Trp*) channel transcript models across the antennae, maxillary palps, and tarsi of virgin males (Vm), virgin females (VF), and mated females (MF). Expression values are displayed as $\log_2$(counts + 1) and centered on a threshold of 10 normalized counts (white = 10 counts). Transcript models are ordered by appendage of highest mean expression and then by expression level. The heatmap reveals widespread expression of *Trp* channel genes together with appendage-specific subsets, particularly within the tarsi, consistent with roles in thermo-, mechano-, and polymodal sensory signaling.

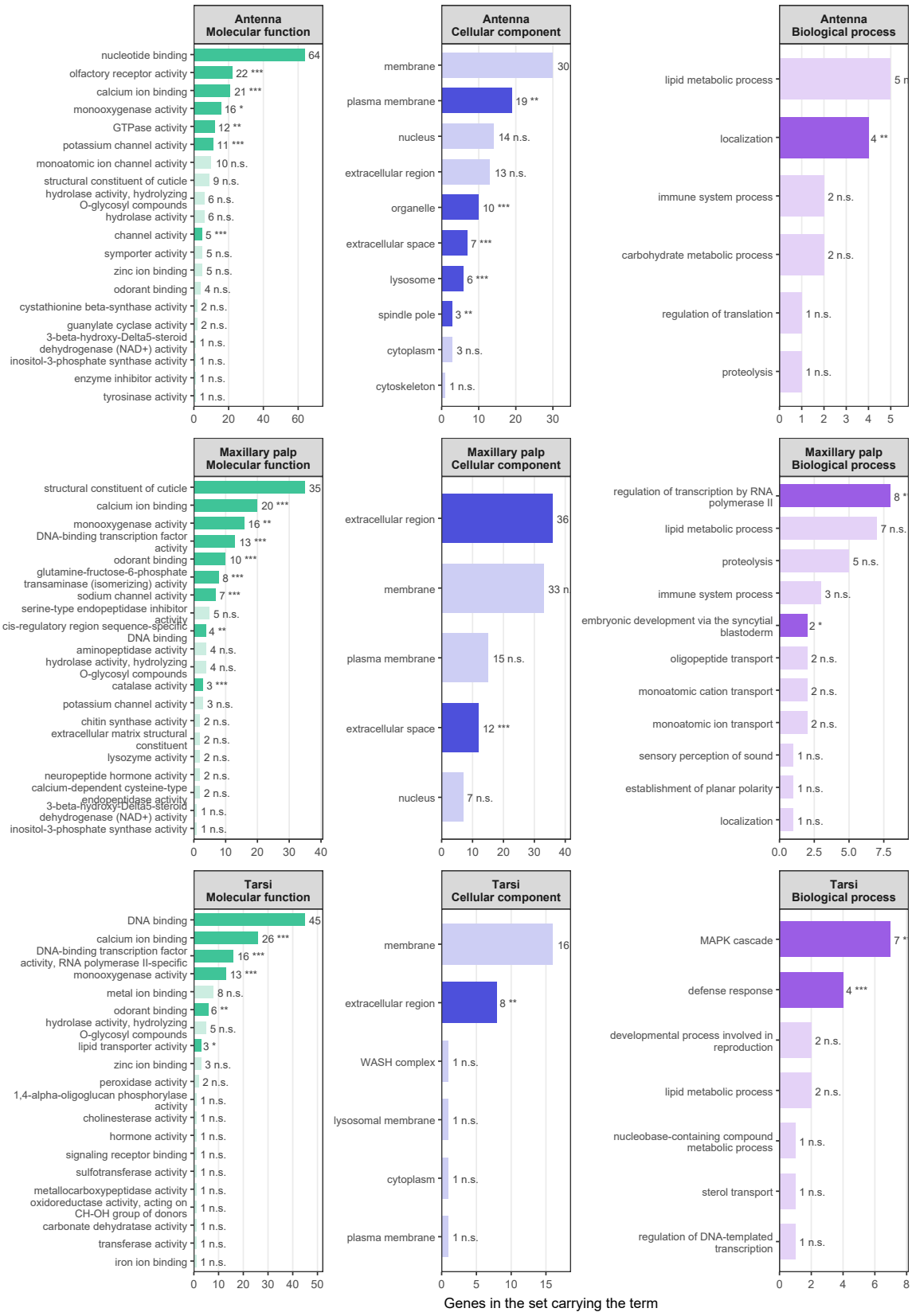


Figure S21. Gene Ontology over-representation among sex-biased genes across chemosensory appendages. Genes differentially expressed between virgin females and virgin males (adjusted $p < 0.001$, $|\log_2 FC| \geq 1$) within each appendage were tested for GO term over-representation. Terms were tested with a one-sided Fisher's exact test for over-representation against all annotated genes tested in the same contrast; each gene carries a single GO term (**Additional file 39**), and p values were corrected by the Benjamini–Hochberg procedure across every background term within each panel. The 20 terms with the lowest adjusted p are shown per panel, with the appendage in rows (antennae, maxillary palps, tarsi) and the GO domain in columns (Molecular function, Cellular component, Biological process). Bar length, and the number beside each bar, is the count of genes in the set carrying the term; stars give the adjusted p (*** < 0.001, ** < 0.01, * < 0.05, n.s. not significant). Bars are colored by GO domain (Molecular function green, Cellular component blue, Biological process purple) and drawn in a pale tint where the term did not reach q < 0.05. Terms are ordered by gene count.

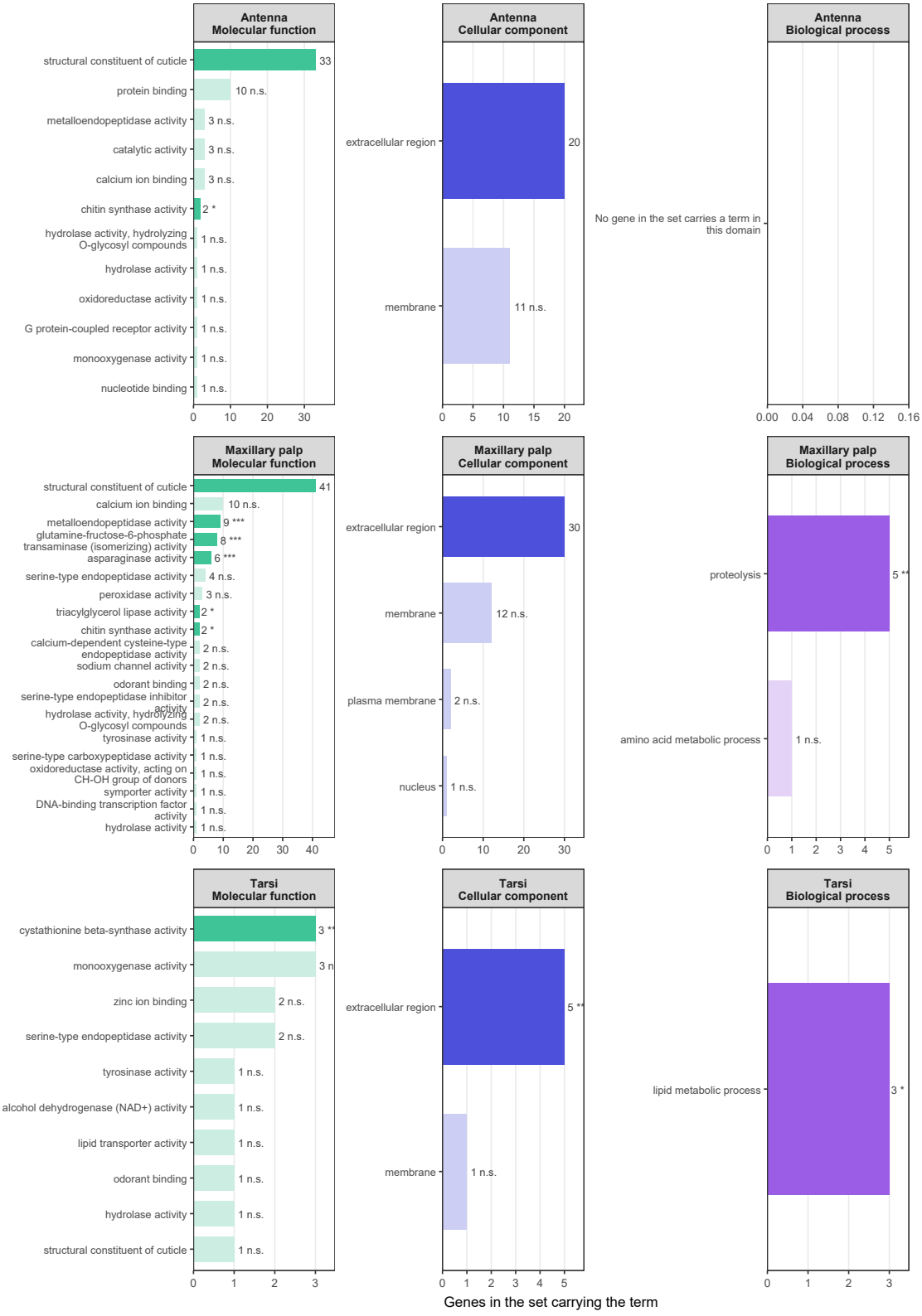


Figure S22. Gene Ontology over-representation among mating-responsive genes across chemosensory appendages. Genes responding to mating, defined as significant in the same direction in both the mated-female versus virgin-female and the mated-female versus virgin-male comparisons (adjusted $p < 0.001$, $|\log_2 FC| \geq 1$), were tested for GO term over-representation within each appendage. Terms were tested with a one-sided Fisher's exact test for over-representation against all annotated genes tested in the same contrast; each gene carries a single GO term (**Additional file 39**), and p values were corrected by the Benjamini–Hochberg procedure across every background term within each panel. The 20 terms with the lowest adjusted p are shown per panel, with the appendage in rows (antennae, maxillary palps, tarsi) and the GO domain in columns (Molecular function, Cellular component, Biological process). Bar length, and the number beside each bar, is the count of genes in the set carrying the term; stars give the adjusted p (*** < 0.001, ** < 0.01, * < 0.05, n.s. not significant). Bars are colored by GO domain (Molecular function green, Cellular component blue, Biological process purple) and drawn in a pale tint where the term did not reach $q < 0.05$. Terms are ordered by gene count. No antennal mating-responsive gene carries a Biological process term, so that panel is empty.

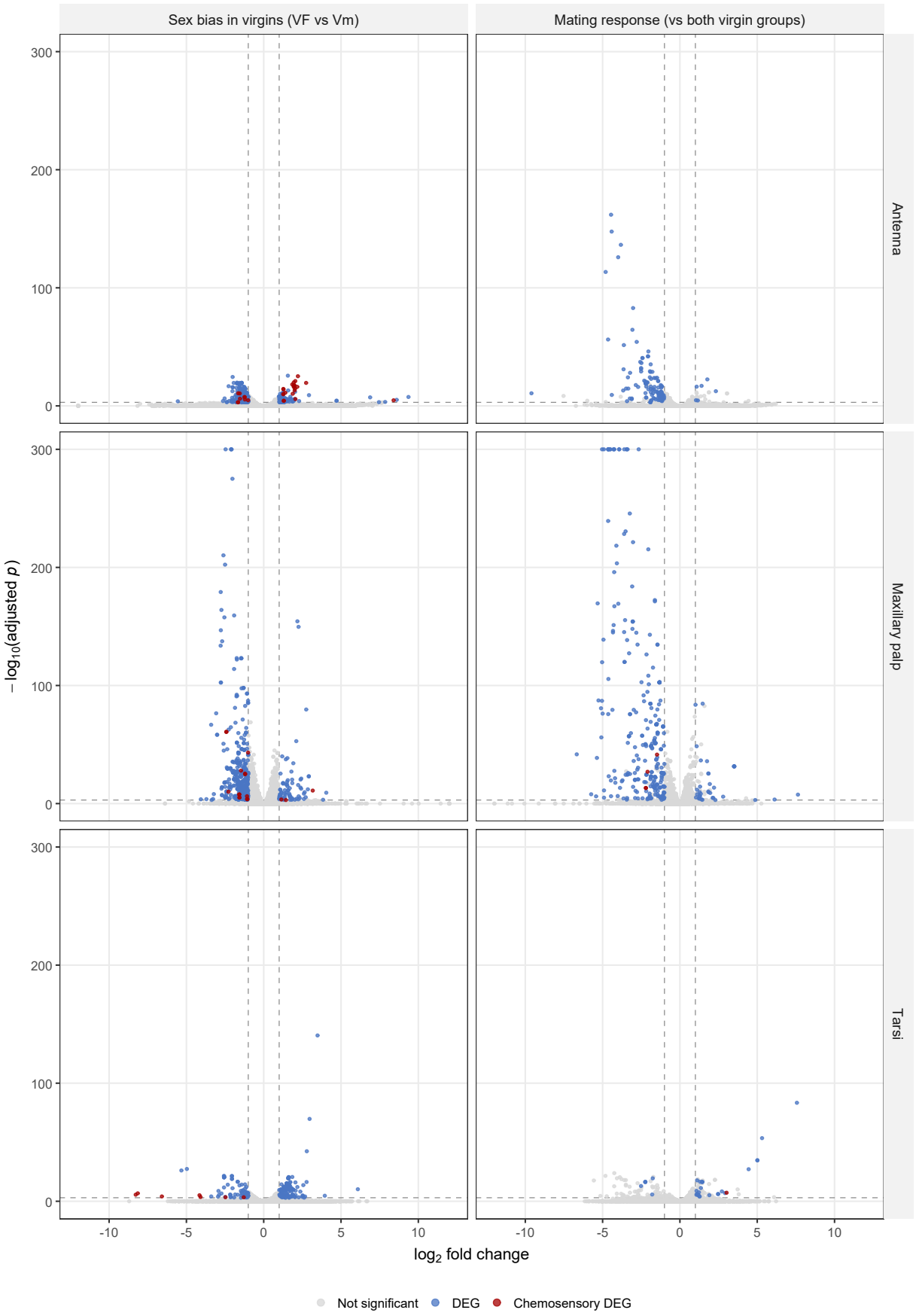


Figure S23. Differential expression across appendages, by sex and by mating. Volcano plots for every gene tested in each appendage. Left column, sex bias in virgins (virgin female versus virgin male); right column, the mating response, which requires significance in the same direction against both virgin females and virgin males. Points are colored gray where not significant, blue for differentially expressed genes (DEG) and red for differentially expressed genes of the seven chemosensory families (chemosensory DEG). Dashed lines mark the significance thresholds (adjusted $p < 0.001$, $|\log_2 FC| \geq 1$). Axes are clipped at $|\log_2 FC|$ 12 and $-\log_{10}$(adjusted p) 300; 38 of 148,968 points lie beyond and are drawn on the boundary. These are the distributions underlying the counts in **Fig. 4**.

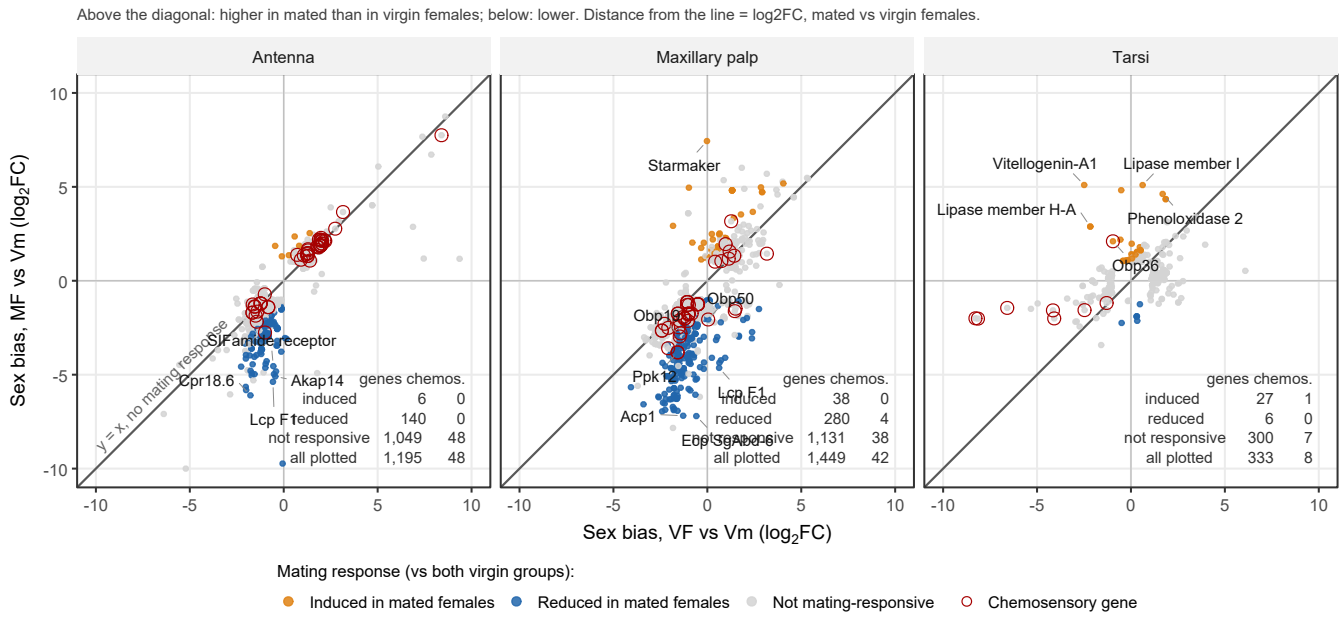


Figure S24. Mated-female-specific expression against constitutive sex bias. Axes as in **Fig. 4D**: sex bias measured in virgins (virgin female versus virgin male) against sex bias measured in mated females (mated female versus virgin male). Because both axes share virgin males as the reference, vertical displacement from the diagonal equals the $\log_2$ fold change between mated and virgin females; genes on the line show sex bias unchanged by mating. Points are colored by mating response, defined in the Methods as significance in the same direction against both virgin groups: orange where induced in mated females, blue where reduced and gray where not mating-responsive. Red rings mark chemosensory genes. Counts in each panel give the number of genes plotted in each category (induced, reduced, not responsive, all plotted) and the number of chemosensory genes among them. Axes are clipped at ±10 $\log_2$FC. Because mated females were older than virgin males at collection, this contrast carries the same age confound as the mating comparison (Methods).