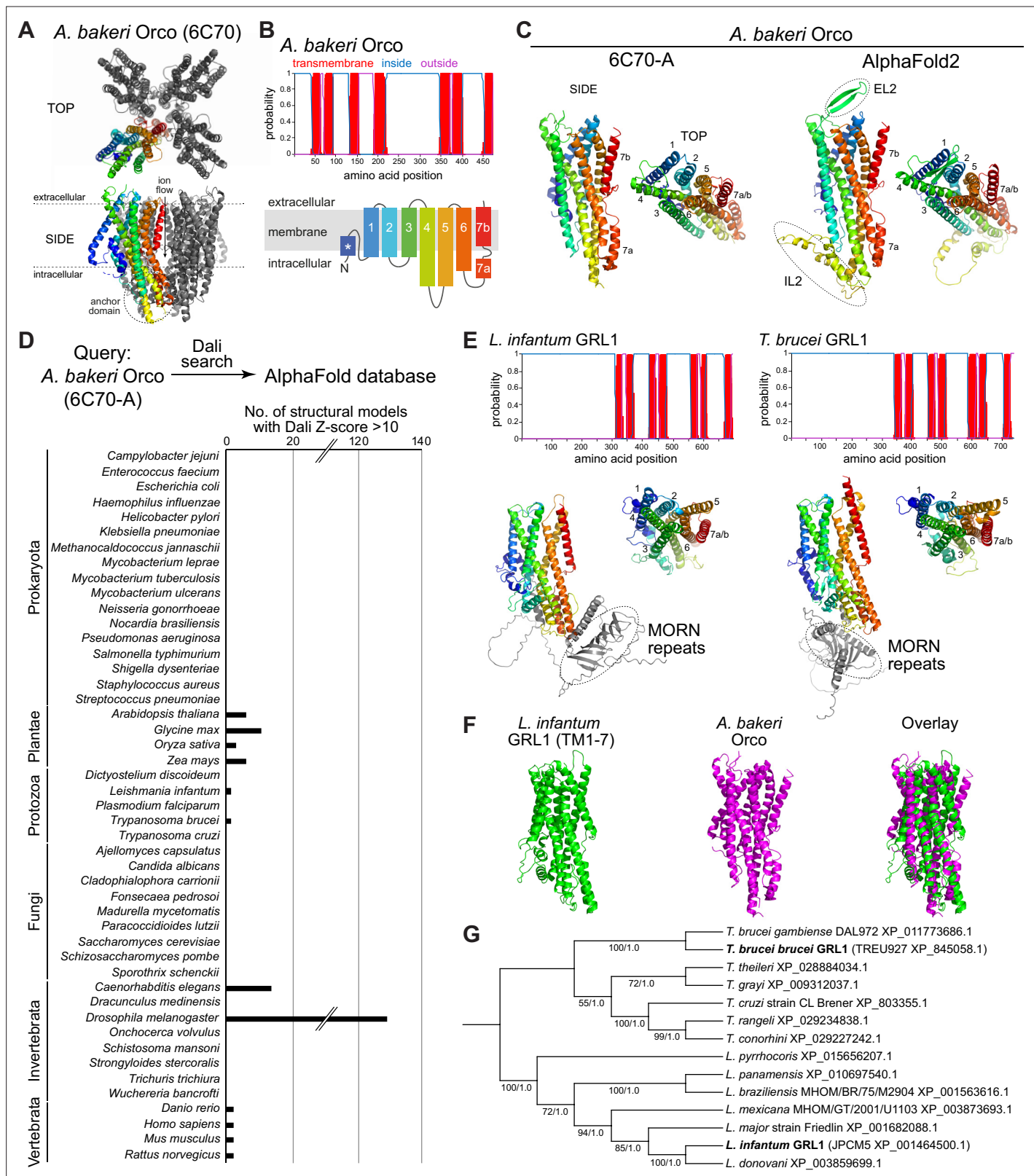


---

## Figures and figure supplements

Structural screens identify candidate human homologs of insect chemoreceptors and cryptic *Drosophila* gustatory receptor-like proteins

**Richard Benton and Nathaniel J Himmel.**

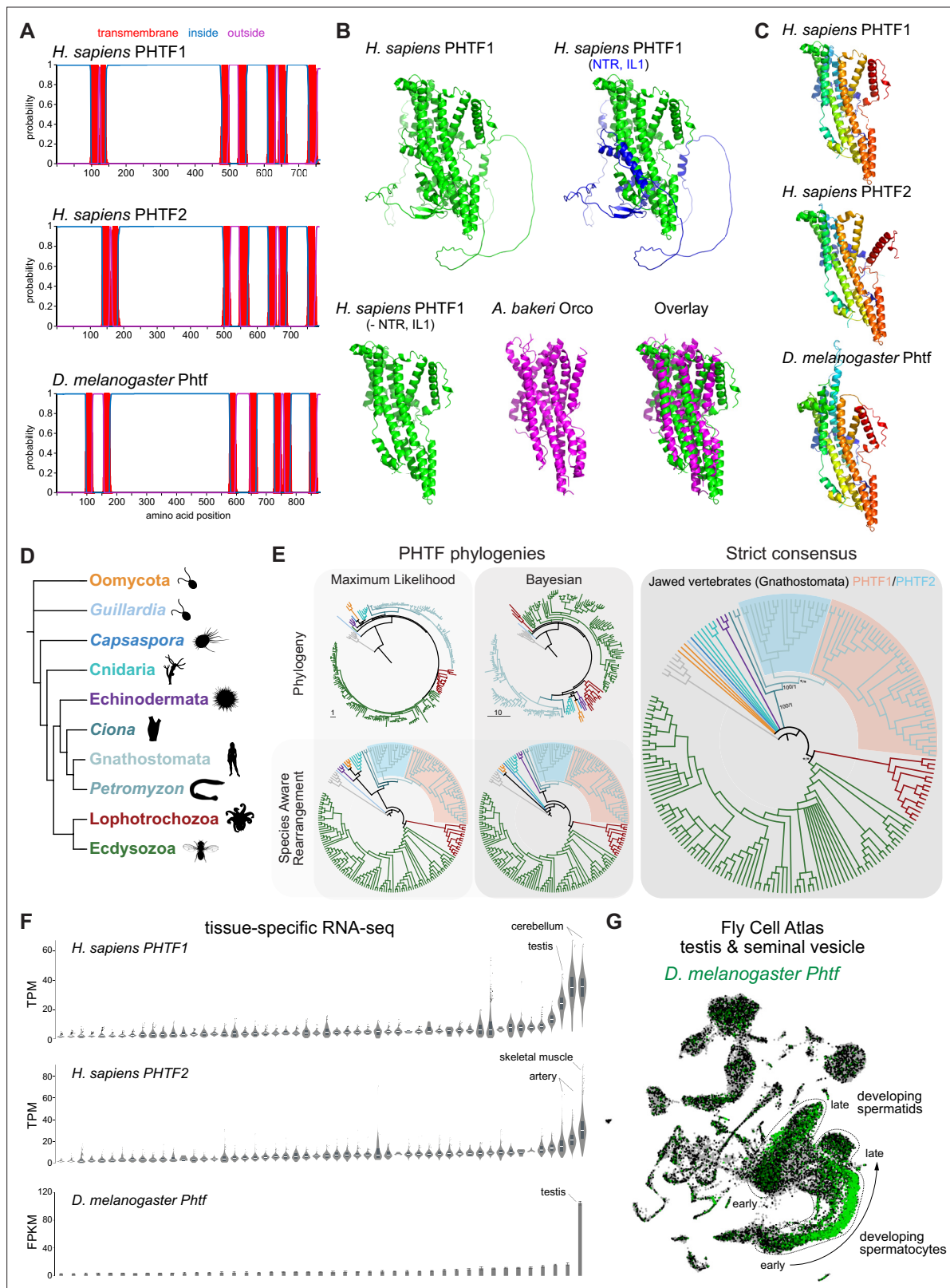


**Figure 1.** Structure-based screening for seven transmembrane domain ion channel (7TMIC) homologs. **(A)** Top view of a cryo-electronic microscopic (cryo-EM) structure of the homotetramer of Or co-receptor (Orco) from *A. bakeri* (derived from PDB 6C70; [Butterwick et al., 2018](#)), in which one subunit has a spectrum coloration (N-terminus [blue] to C-terminus [red]). The ion channel pore is formed at the interface of the four subunits. A side view is shown below. The anchor domain, comprising the cytoplasmic projections of TM4-6 and TM7a, forms most of the inter-subunit interactions in

Figure 1 continued on next page

## Figure 1 continued

odorant receptors (Ors) (**Butterwick et al., 2018**; **Del Marmol et al., 2021**). **(B)** Top: output of transmembrane topology predictions of DeepTMHMM (**Hallgren et al., 2022**) for *A. bakeri* Orco. Bottom: schematic of the membrane topology of an Orco monomer, with the same spectrum coloration as in **(A)**, reproduced from Figure 1a from **Benton et al., 2020**. Note that the seventh predicted helical region is divided into two in the cryo-EM structure: TM7a (located in the cytosol) and TM7b (located in the membrane). **(C)** Comparisons of side and top views of the cryo-EM structure of an *A. bakeri* Orco subunit (6C70-A) (left) and an AlphaFold2 protein structure prediction of *A. bakeri* Orco. Helical regions are numbered in the top views. Note the model contains the extracellular loop 2 (EL2) and intracellular loop 2 (IL2) regions that were not able to be accurately visualized in the cryo-EM structure (**Butterwick et al., 2018**). Quantitative comparisons of structures are provided in **Table 1**. **(D)** Summary of the results of the screen for Orco/Or-like protein folds in the AlphaFold Protein Structure Database for the indicated species using Dali (**Holm, 2022**). The threshold of Dali Z-score >10 was informed by inspection of the results of the screen (see Results). Raw outputs of the screen are provided in Source data 2. **(E)** Top: transmembrane topology predictions of the single screen hits from the *Trypanosoma* species *Leishmania infantum* and *Trypanosoma brucei brucei*. Bottom: AlphaFold2 structural models of these proteins, displayed as in **(C)**. The long N-terminal region contains tandem Membrane Occupation and Recognition Nexus (MORN) repeats and sequence of unknown structure (gray); these are masked in the top view of the models. **(F)** Visual comparison of the *L. infantum* GRL1 AlphaFold2 model (the N-terminal region is masked) with the *A. bakeri* Orco structure, aligned with Coot (**Emsley et al., 2010**). Quantitative comparisons of structures are provided in **Table 1**. **(G)** Consensus phylogeny of putative trypanosome homologs. The primary sequence database was assembled using *L. infantum* GRL1 (XP\_001464500.1) and *T. brucei brucei* GRL1 (XP\_845058.1) as query sequences (highlighted in bold). Branch support values refer to maximum likelihood UFboot/Bayesian posterior probabilities. Note that although the *Trypanosoma cruzi* homolog (XP\_803355.1) was not identified in the original Dali screen, visual inspection of the corresponding AlphaFold2 model (A0A2V2WL40) revealed the same global fold.

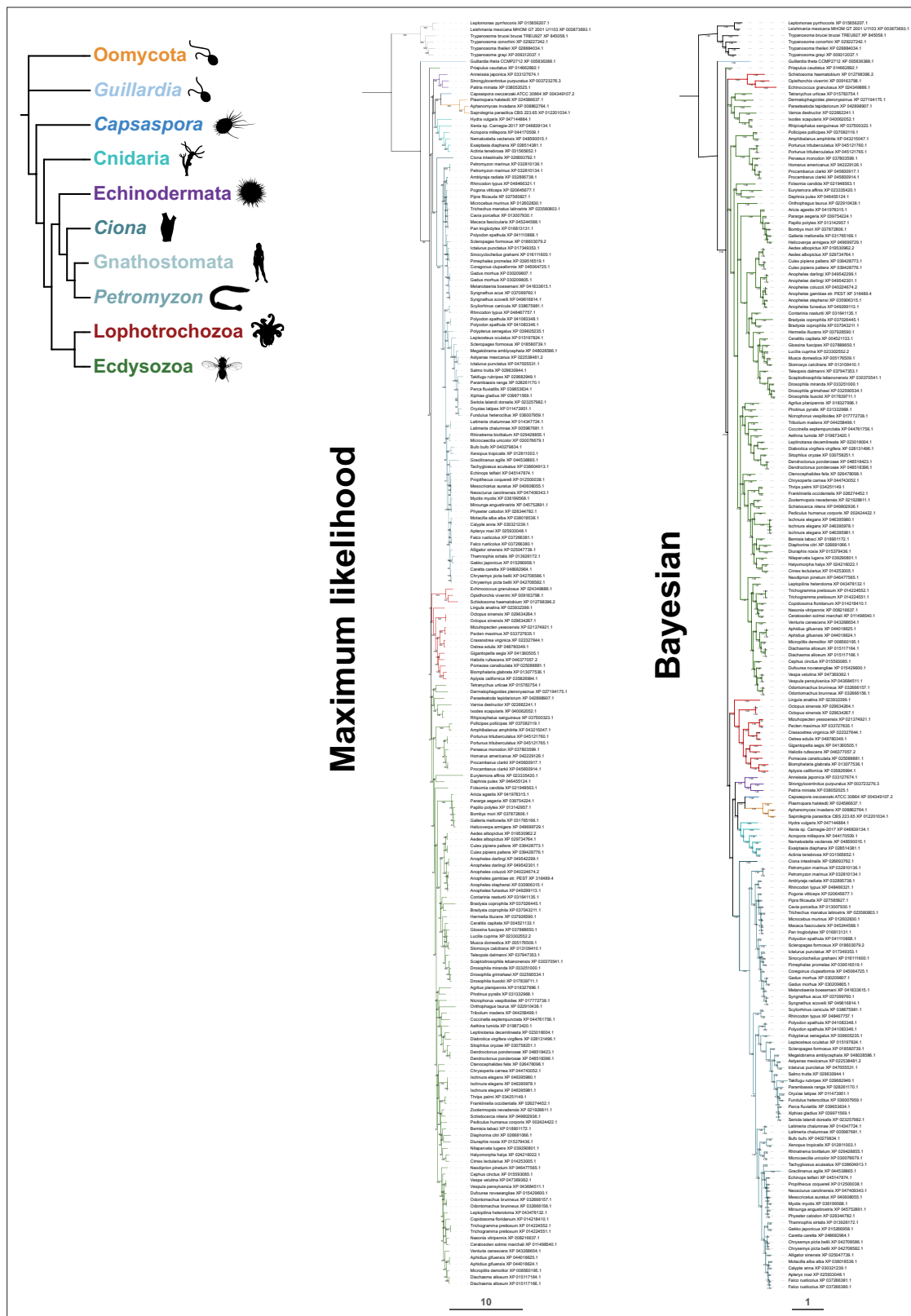


**Figure 2.** PHTF proteins are candidate vertebrate seven transmembrane domain ion channels (7TMICs). **(A)** DeepTMHMM-predicted transmembrane topology of PHTF proteins. **(B)** Top: AlphaFold2 predicted structure of *H. sapiens* PHTF1; in the image on the right the long N-terminal region (NTR) and intracellular loop 1 (IL1) are highlighted in blue; these sequences contain a few predicted helical regions but are of largely unknown structure. Bottom: visual comparison of the *H. sapiens* PHTF1 AlphaFold2 structure (in which the NTR and IL1 are masked) with the *A. bakeri* Or co-receptor (Orco)

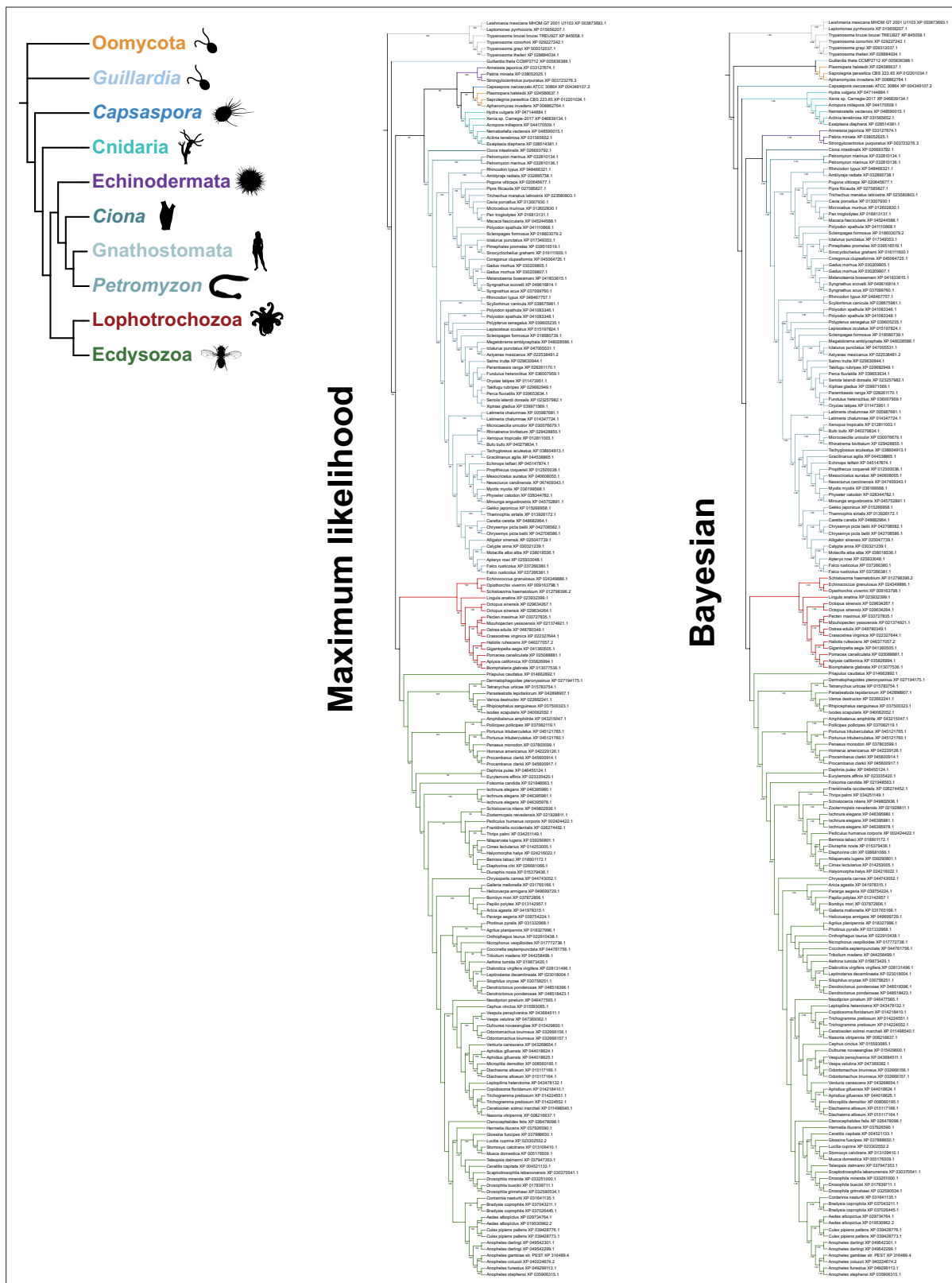
Figure 2 continued on next page

## Figure 2 continued

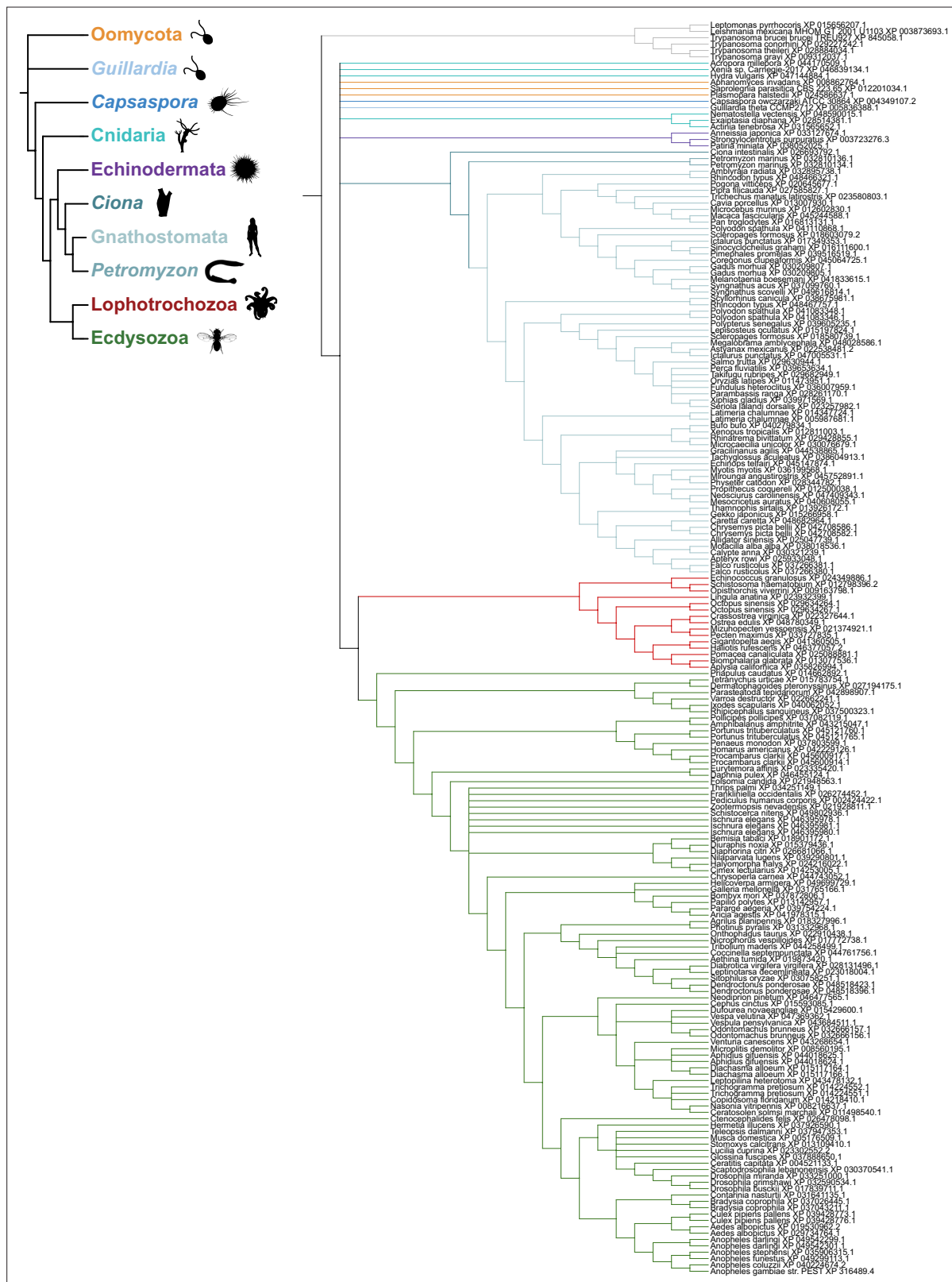
structure. (C) AlphaFold2 structures of PHTF proteins in which the NTR and IL1 are masked. Quantitative comparisons of these structures to the cryo-electronic microscopic (cryo-EM) Orco structure are provided in **Table 1**. (D) Major taxa/species in which a PHTF homolog was identified (see sequence databases in **Figure 2—source data 1**). Silhouette images in this and other figures are from Phylopic ([phylopic.org/](https://phylopic.org/)). (E) Phylogenies of a representative set of PHTF sequences. The sequence database was constructed using the *D. melanogaster* and *H. sapiens* PHTF query sequences. Top left: maximum likelihood phylogeny (JTT + R10 model) and Bayesian phylogeny. The scale bars represent the average number of substitutions per site. Bottom left: phylogenies where weakly supported branches (<95/0.95) have been rearranged and polytomies resolved in a species tree-aware manner. Right: strict consensus of the species tree-aware phylogenies. There is a single eukaryotic PHTF clade and the PHTF1-2 split occurred in the jawed vertebrate lineage. However, this interpretation relies on the rearrangement of the weakly supported jawless vertebrate PHTF branch. Therefore, an alternative but weakly supported hypothesis is that the duplication occurred in a common vertebrate ancestor and a single PHTF copy was lost in jawless vertebrates. Select branch support values are present on key branches and refer to maximum likelihood UFboot/Bayesian posterior probabilities. Asterisks indicate that branch support was below the threshold for species-aware rearrangement. The fully annotated trees are available in **Figure 2—figure supplements 1–3**. (F) Summary of tissue-enriched RNA expression of *H. sapiens* PHTF1 and PHTF2 (data are from the GTex Portal; the fully annotated dataset is provided in **Figure 2—figure supplement 4**) and *D. melanogaster* Phtf (data from the Fly Atlas 2.0; the fully annotated dataset is provided in **Figure 2—figure supplement 5**). (G) Left: Uniform Manifold Approximation and Projection (UMAP) representation of RNA-seq datasets from individual cells of the *D. melanogaster* testis and seminal vesicle generated as part of the Fly Cell Atlas (10× relaxed dataset) ([Li et al., 2022](#)) colored for expression of *Phtf*. Simplified annotations of cell clusters displaying the highest levels of *Phtf* expression are adapted from [Li et al., 2022](#); unlabeled clusters represent non-germline cell types of the testis.



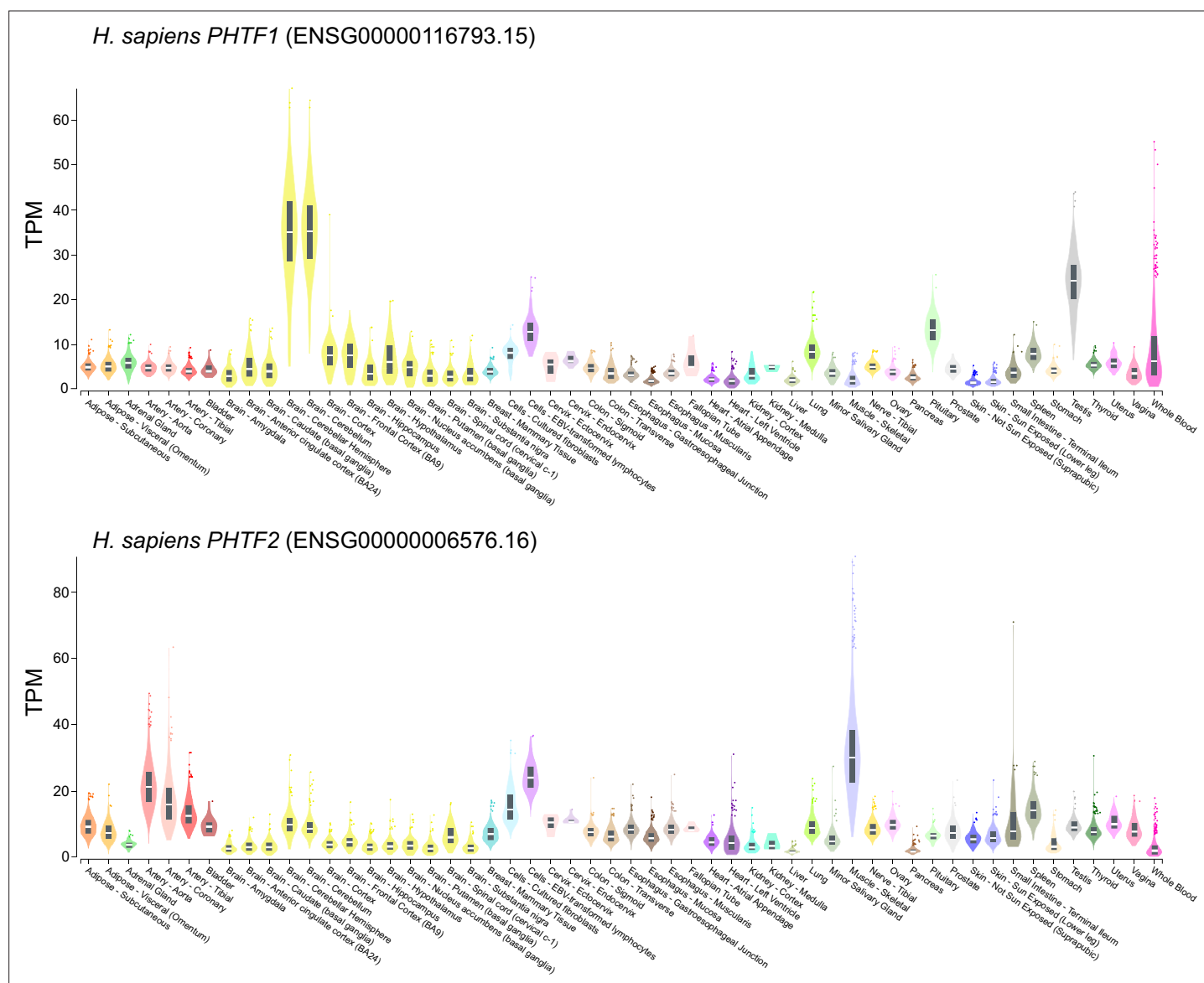
**Figure 2—figure supplement 1.** Fully annotated phylogenetic trees for PHTF homologs. Sequences are from the protein sequence database generated using *D. melanogaster* Phtf and *H. sapiens* PHTF1/2, and are representatives of clusters of 90% sequence identity. For maximum likelihood, the tree was generated using a JTT + R10 substitution model. Branch support values for maximum likelihood (UFboot) and Bayesian analyses (posterior probability) are shown at the branches. The scale bars represent the average number of substitutions per site.



**Figure 2—figure supplement 2.** Fully annotated species-aware trees for PHTF homologs. Trees are based on the maximum likelihood (left) and Bayesian (right) trees. Branches without support values were eligible for rearrangement.



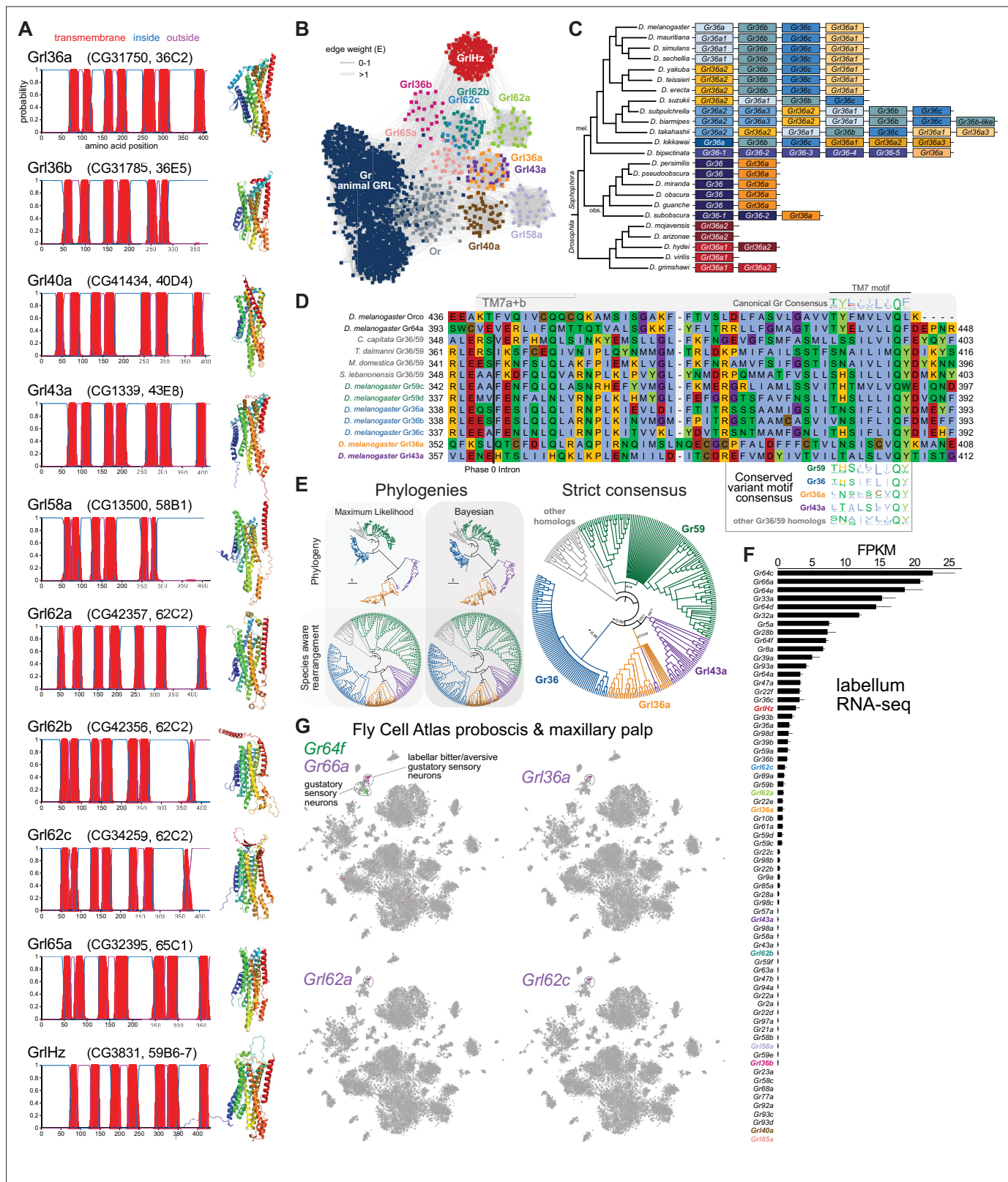
**Figure 2—figure supplement 3.** Strict consensus of the species-aware trees for PHTF homologs.



**Figure 2—figure supplement 4.** Tissue-specific RNA expression of *H. sapiens PHTF1* and *PHTF2*. Plot of RNA expression levels (transcripts per million [TPM]) from the indicated tissues is from the GTEx Portal (GTEx Analysis Release V8 [dbGaP Accession phs000424.v8.p2]).

		<i>Phtf</i>		<i>Grl36a</i>	<i>Grl36b</i>	<i>Grl40a</i>	<i>Grl43a</i>	<i>Grl58a</i>	<i>Grl62a</i>	<i>Grl62b</i>	<i>Grl62c</i>	<i>Grl65a</i>	<i>GrlHz(A)</i>	<i>GrlHz(C)</i>	FKPM
		FKPM													
Adult male	Head	8.19		0.06	2.54	0.03	0.06	0.09	0.07	0.03	0.00	0.90	3.20	0.22	
	Eye	11.53	120	0.00	3.29	0.00	0.02	0.00	0.00	0.01	0.04	0.48	1.75	1.37	40
	Brain	15.33	80	0.09	5.14	0.00	0.02	0.17	0.00	0.03	0.01	0.05	1.99	0.82	35
	Thoracoabdominal ganglion	10.32	40	0.04	5.30	0.03	0.01	0.28	0.01	0.02	0.02	0.18	3.39	0.96	30
	Crop	4.39	0	0.00	0.37	0.00	0.06	0.00	0.00	0.02	0.00	0.04	1.38	0.00	25
	Midgut	4.78		0.00	0.27	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.01	1.30	20
	Hindgut	5.2		0.00	0.16	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.36	0.39	15
	Malpighian Tubules	5.41		0.00	0.30	0.00	0.01	0.00	0.00	0.00	0.00	0.00	2.03	0.77	10
	Fat body	4.63		0.00	0.29	0.35	0.00	0.00	0.00	0.00	0.00	1.37	0.49	1.09	5
	Salivary gland	4.91		0.02	0.31	0.02	0.12	0.00	0.00	0.00	0.00	0.18	2.64	0.10	0
	Heart	10.29		0.00	0.23	0.00	0.04	0.00	0.00	0.00	0.00	0.64	8.76	0.00	
	Testis	104.27		0.00	1.32	1.05	0.03	0.03	0.00	0.00	0.07	0.13	9.04	1.00	
	Accessory glands	4.48		0.00	0.17	0.00	0.21	0.00	0.00	0.00	0.00	0.08	1.04	0.00	
	Carcass	5.71		0.00	1.01	0.04	0.04	0.05	0.00	0.00	0.00	0.15	2.54	0.11	
	Rectal pad	4.26		0.00	0.26	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.40	0.50	
	Whole body	16.35		0.00	1.28	0.06	0.08	0.11	0.03	0.04	0.00	0.09	2.52	0.23	
	Head	7.71		0.02	2.49	0.00	0.00	0.01	0.06	0.06	0.01	1.03	3.73	0.83	
Adult female	Eye	9.02		0.00	2.38	0.00	0.00	0.00	0.00	0.00	0.01	0.50	3.98	0.67	
	Brain	13.3		0.07	5.47	0.02	0.01	0.12	0.02	0.00	0.03	0.01	2.77	0.32	
	Thoracoabdominal ganglion	11.08		0.05	5.55	0.07	0.02	0.33	0.03	0.01	0.01	0.19	4.86	0.38	
	Crop	3.34		0.00	0.16	0.00	0.01	0.02	0.00	0.00	0.00	0.02	1.93	0.00	
	Midgut	2.75		0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.60	
	Hindgut	5.63		0.00	0.16	0.00	0.02	0.00	0.03	0.00	0.01	0.06	1.11	0.06	
	Malpighian Tubules	4.74		0.00	0.42	0.01	0.01	0.00	0.00	0.02	0.00	0.02	1.35	0.31	
	Fat body	3.69		0.00	0.08	0.00	0.07	0.00	0.00	0.00	0.00	3.25	4.53	0.00	
	Salivary gland	4.54		0.00	0.12	0.02	0.17	0.02	0.00	0.00	0.02	0.08	5.94	0.16	
	Heart	10.83		0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	1.14	26.40	0.15	
	Ovary	6.86		0.00	0.10	0.01	0.02	0.00	0.14	0.11	0.16	0.00	22.91	0.19	
	Virgin Spermatheca	2.63		0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	2.58	2.10	0.45	
	Mated Spermatheca	2.86		0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	3.12	2.96	0.00	
	Carcass	5.57		0.00	0.68	0.00	0.01	0.00	0.00	0.00	0.00	0.15	6.20	0.58	
	Rectal pad	4.43		0.00	0.18	0.00	0.06	0.00	0.00	0.00	0.00	0.00	1.35	0.35	
	Whole body	6.04		0.00	0.15	0.03	0.00	0.00	0.06	0.12	0.05	0.10	14.30	0.30	
Larva	Central nervous system	5.18		0.05	1.40	0.00	0.07	0.03	0.08	0.21	0.14	0.02	4.97	0.15	
	Midgut	5.49		0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.85	0.09	
	Hindgut	3.5		0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.82	0.73	
	Malpighian Tubules	2.88		0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	5.31	0.34	
	Fat body	5.64		0.00	0.00	0.00	0.05	0.07	0.13	0.07	0.00	0.07	8.60	3.22	
	Salivary gland	3.5		0.00	0.02	0.00	0.09	0.00	0.00	0.00	0.03	0.03	1.33	0.55	
	Trachea	3.12		0.00	2.71	0.00	0.01	0.00	0.00	0.00	0.00	0.00	1.64	0.24	
	Carcass	2.57		0.08	0.32	0.00	0.00	0.00	0.00	0.00	0.03	0.00	2.16	0.34	
	Garland cells	5.8		0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	36.62	1.00	
	Whole body	3.1		0.00	0.22	0.00	0.03	0.00	0.00	0.00	0.00	0.05	3.86	0.64	

**Figure 2—figure supplement 5.** Tissue-specific RNA expression of *D. melanogaster* *Phtf* and *Grls*. Heatmap plot of the expression of *D. melanogaster* *Phtf* and *Grls* in the indicated tissues/life stages/sexes determined by bulk RNA-seq; fragments per kilobase of exon per million mapped fragments (FPKM) values are shown; data are from the Fly Atlas 2.0 (Krause et al., 2022).

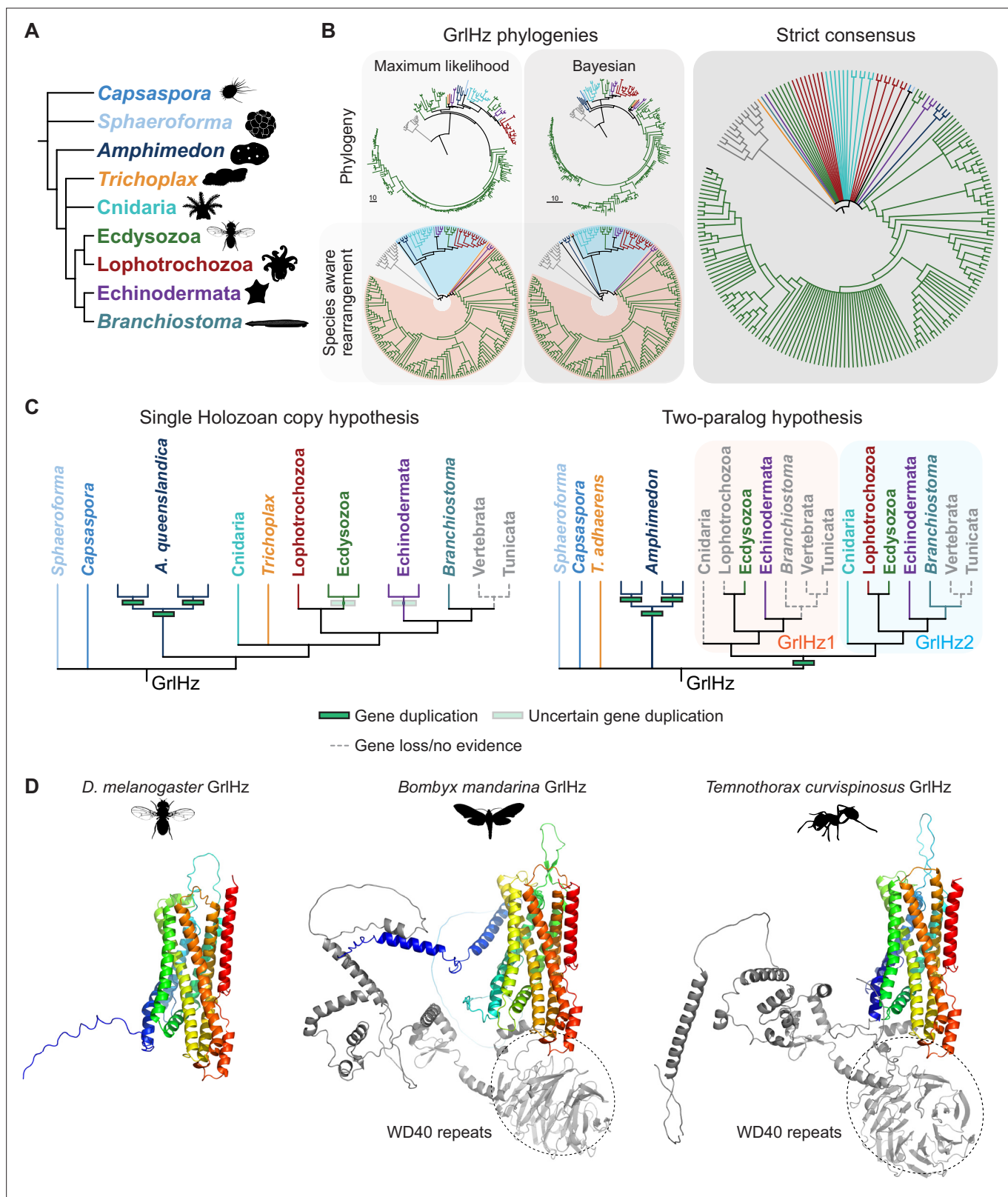


**Figure 3.** Insect Grs are highly divergent, candidate chemosensory receptors. **(A)** Proposed nomenclature of *D. melanogaster* Grs (the original gene name and cytological location are in parentheses), with corresponding DeepTMHMM-predicted transmembrane topologies and AlphaFold2 structural models. Note that TM7 is not predicted for Gr136b and Gr158a by DeepTMHMM, but is predicted – with the characteristic TM7a/7b split – in the structural model (as well as predicted by Phobius [data not shown]). Quantitative comparisons of these structures to the cryo-electronic microscopic

Figure 3 continued on next page

## Figure 3 continued

(cryo-EM) Or co-receptor (Orco) structure are provided in **Table 1**. **(B)** Sequence similarity network of Grs, Grs, and Ors (including Orco). The network was generated using an all-to-all comparison made by MMSeqs2 as implemented by gs2. The connections represent E-values where the weakest connections (arbitrarily defined as edge weights >1) are colored in lighter gray. Lack of connection between two nodes indicates that those two sequences could not be identified as having any significant sequence similarity under the most sensitive MMSeqs2 settings. Nodes and edges are arranged in a prefuse force-directed layout. The graph splitting tree is visualized in **Figure 3—figure supplement 5**; however, we do not place high confidence in the phylogenetic accuracy of the tree due to the likely effects of long branch attraction. The evolution of GrHolozoa (GrHlz) is described in **Figure 3—figure supplement 1**, with detailed phylogenies in **Figure 3—figure supplements 2–4**. **(C)** Schematic of the gene arrangement of *Gr36a* and *Gr36* homologs in drosophilids. Color coding reflects relatedness with respect to major speciation and gene duplication events; colors match the phylogenetic tree branches in **Figure 3—figure supplement 6B–C**. The *Drosophila* subgenus entirely lacks *Gr36* homologs (see **Figure 3—figure supplement 6**). **(D)** Alignment of the C-terminal region of *D. melanogaster* Orco, *Gr64a*, select insect *Gr36/Gr59* homologs, and *D. melanogaster* *Gr36a* and *Gr43a*, extracted from a larger alignment available in **Figure 3—source data 5**. The black bar shows the common location of a phase 0 intron, which is presumably homologous in different sequences. The canonical TM7 motif of the Gr family (represented as relative amino acid frequencies extracted from WebLogo) is shown above the sequence, and the variant motifs of different Gr or Grl ortholog groups are shown below. **(E)** Phylogenies of *Gr36*, *Gr59c/d*, *Gr36a*, *Gr43a* and homologous non-drosophilid sequences (color-coded as in **(D)**). The sequence database was assembled using *D. melanogaster* *Gr36a*, *Gr36a*, and *Gr43a* as the query sequences. Top left: maximum likelihood phylogeny (JTT + F + R7 model) and Bayesian phylogeny. The scale bars represent average number of substitutions per site. Bottom left: phylogenies where weakly supported branches (<95/0.95) have been rearranged and polytomies resolved in a species tree-aware manner. Right: strict consensus of the species tree-aware phylogenies. These analyses support that *Gr36* and *Gr36a/43a* are sister clades, which likely split after *Gr59c/d* diverged from the ancestral lineage. Sequences are colored as in **(D)**. Select branch support values are present on key branches and refer to maximum likelihood UFboot and Bayesian posterior probabilities, in this order. Asterisks indicate that branch support was below the threshold for species-aware rearrangement. A simplified schematic of gene duplication and loss is illustrated in **Figure 3—figure supplement 6F**. The fully annotated trees are available in **Figure 3—figure supplements 7–9**. **(F)** Histogram of Gr and Grl expression levels in adult proboscis and maxillary palps determined by bulk RNA-sequencing (RNA-seq). Mean values  $\pm$  SD of fragments per kilobase of transcript per million mapped reads (FPKM) are plotted; n=3 biological replicates. Data is from **Dweck et al., 2021**. **(G)** Left: t-distributed stochastic neighbor embedding (tSNE) representation of RNA-seq datasets from individual cells of the *D. melanogaster* proboscis and maxillary palp – generated as part of the Fly Cell Atlas (10 $\times$  stringent dataset) (**Li et al., 2022**) – colored for expression of the indicated genes. *Gr64f* and *Gr66a* are broad markers of ‘sweet/appetitive’ and ‘bitter/aversive’ gustatory sensory neurons, respectively. Transcripts for three Grs are detected in subsets of bitter/aversive neurons. Annotations of cell clusters are adapted from **Li et al., 2022**; unlabeled clusters represent other non-gustatory sensory neuron or non-neuronal cell types of this tissue.

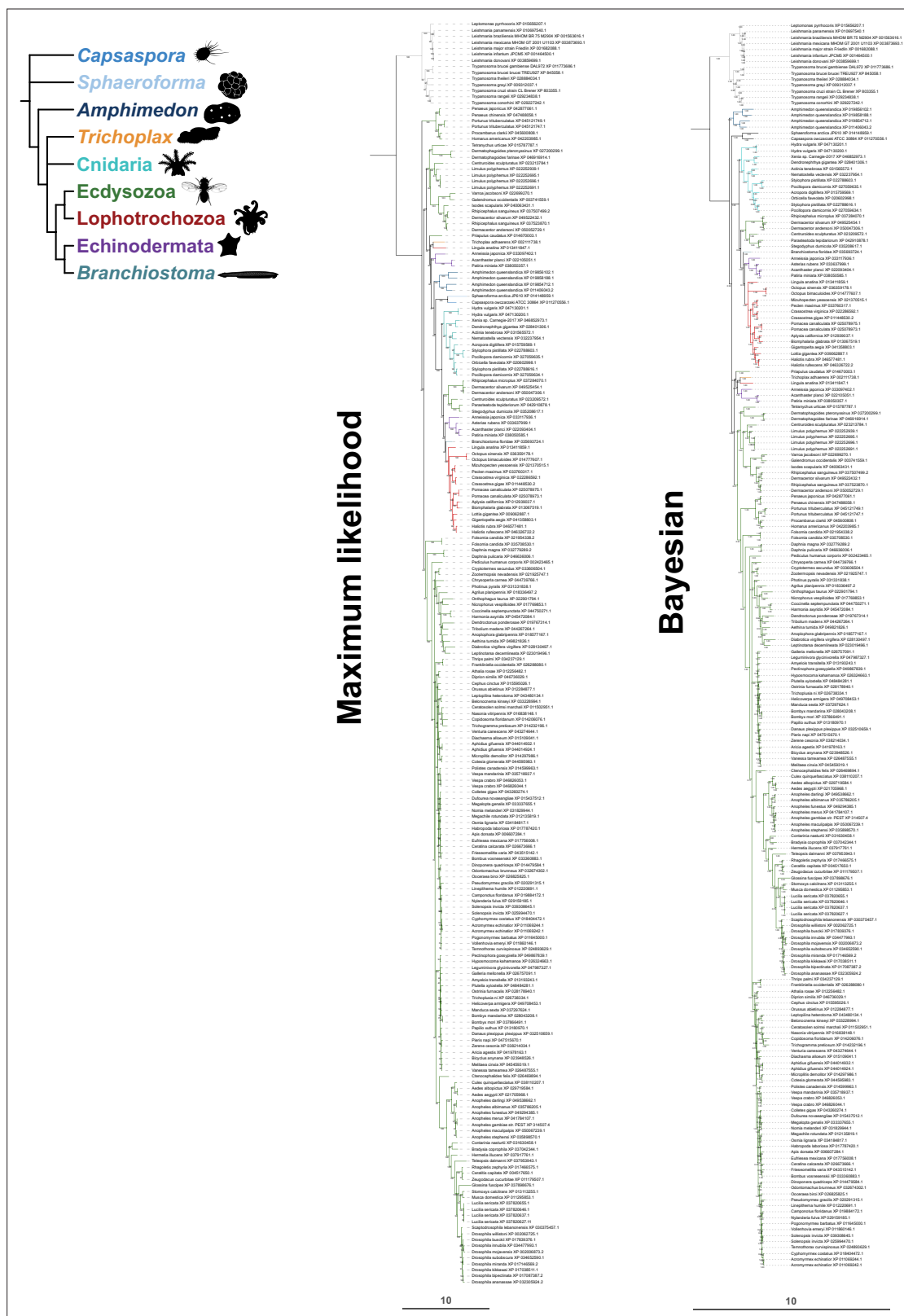


**Figure 3—figure supplement 1.** Evolution of GrlHolozoa (GrlHz), a family of Grl seven transmembrane domain ion channel (7TMIC) not restricted to flies. (A) Major taxa/species for which a GrlHz homolog was recovered. (B) Phylogenies of a representative set of GrlHz sequences (clustered by 70% sequence identity). The sequence database was assembled using *D. melanogaster* GrlHz as the query sequence. Top: maximum likelihood phylogeny and Bayesian phylogeny. The scale bars represent the average number of substitutions per site. Bottom: phylogenies where weakly supported branches

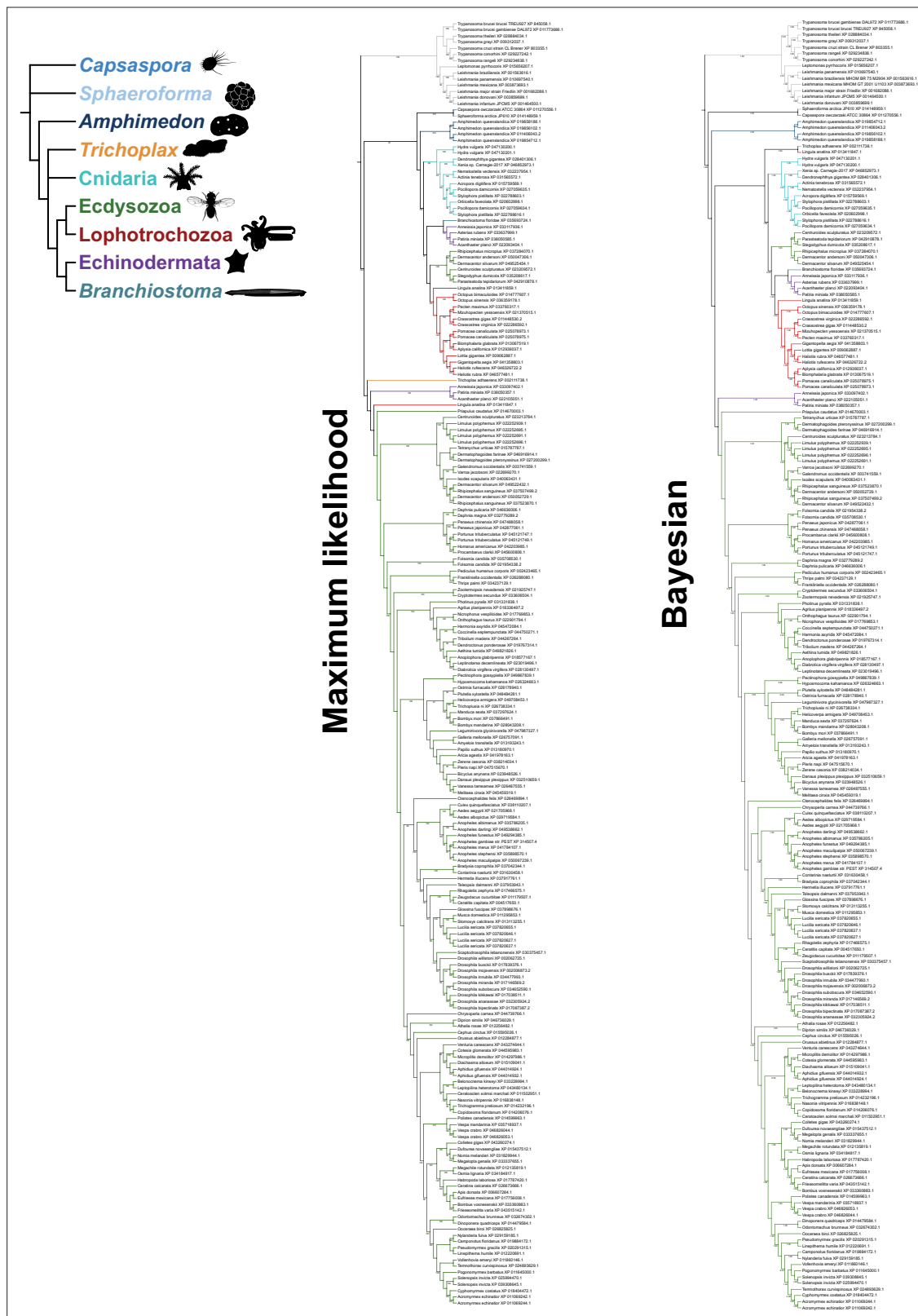
Figure 3—figure supplement 1 continued on next page

*Figure 3—figure supplement 1 continued*

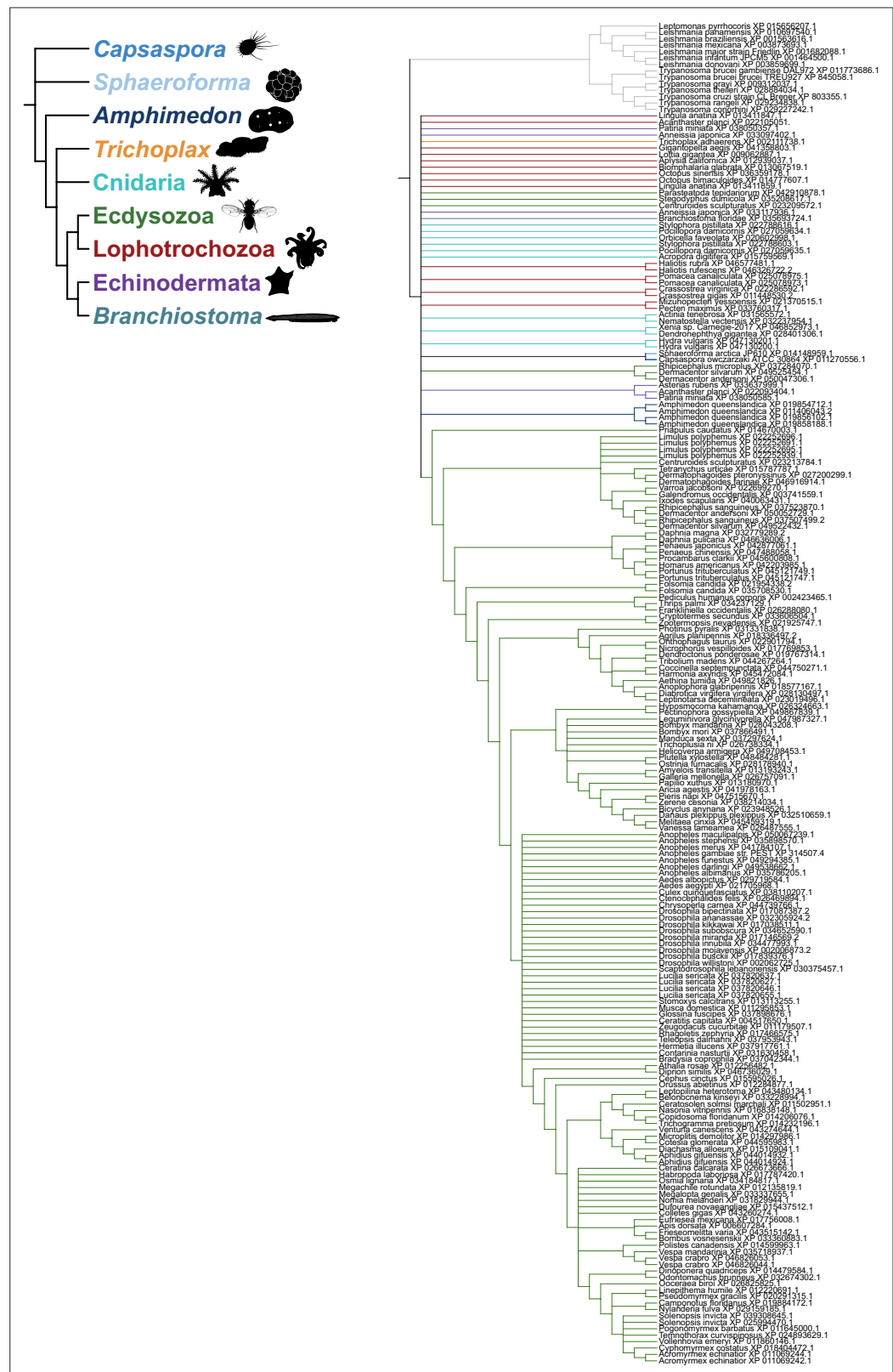
(<95/0.95) have been rearranged and polytomies resolved in a species tree-aware manner. Right: strict consensus of the species tree-aware phylogenies. The fully annotated trees are visualized in **Figure 3—figure supplements 2–4**. **(C)** Left: the single holozoan copy hypothesis of GrHz evolution. Under this scenario, a single GrHz is widely conserved across Holozoa, but has been independently duplicated/lost several times in various taxa. Right: the two-paralog hypothesis of GrHz evolution. As both the maximum likelihood and Bayesian phylogenies provide evidence for two GrHz clades, and because some species have two substantially divergent GrHz sequences, it is possible that there was a gene duplication event early in the evolution of Holozoa. **(D)** Examples of GrHz structures. Of 196 representative sequences, 31 sequences (mostly from Hymenoptera and Lepidoptera) bear N-terminal WD40 repeats.



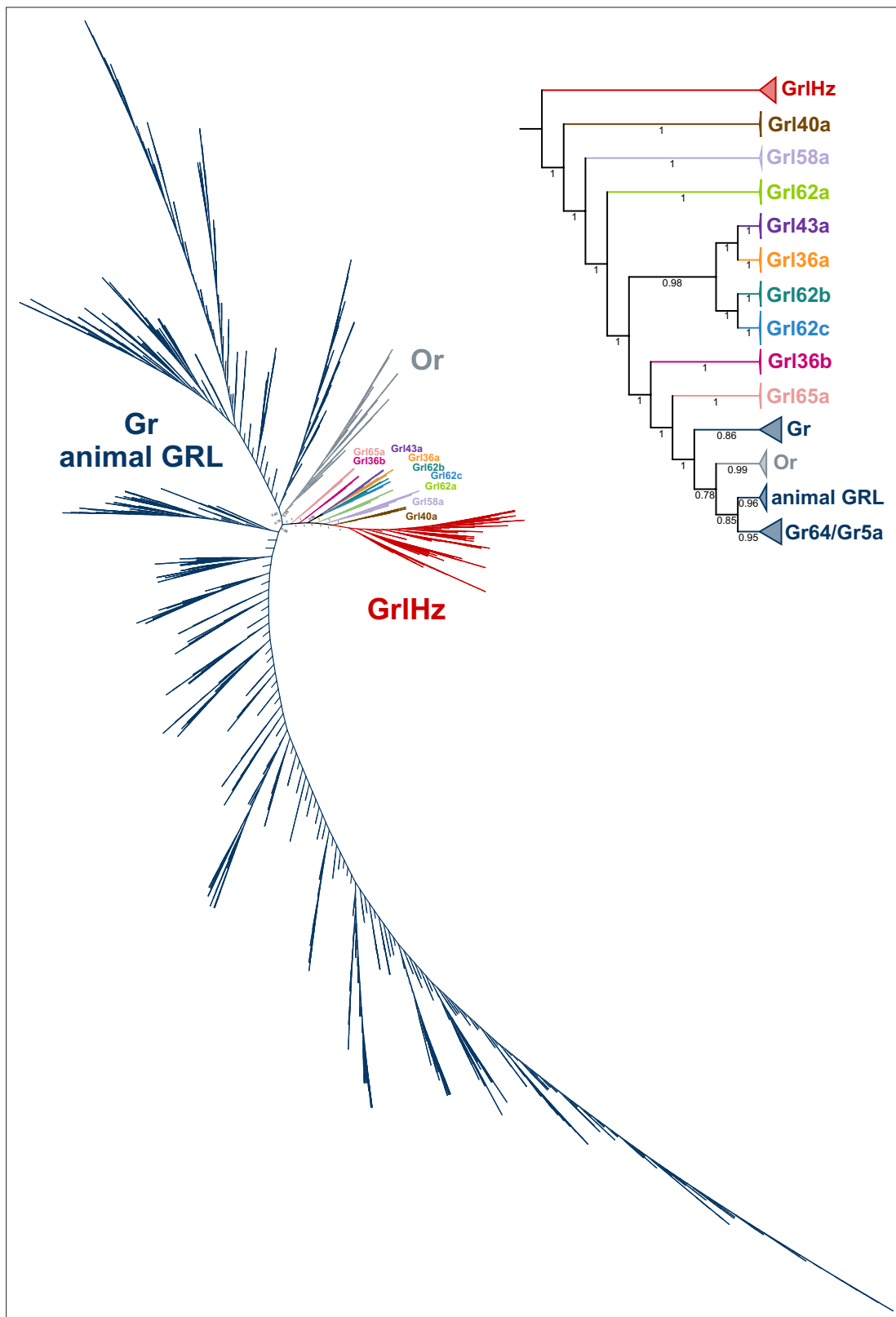
**Figure 3—figure supplement 2.** Fully annotated phylogenetic trees for GrHolozoa (GrlHz) homologs. For maximum likelihood, the tree was generated using a JTT + F + R9 substitution model. Branch support values for maximum likelihood (UFboot) and Bayesian analyses (posterior probability) are shown at the branches. The scale bars represent the average number of substitutions per site.



**Figure 3—figure supplement 3.** Fully annotated species-aware trees for GrlHolozoa (GrlHz) homologs. Trees are based on the maximum likelihood (left) and Bayesian (right) trees. Branches without support values were eligible for rearrangement.



**Figure 3—figure supplement 4.** Strict consensus of the species-aware trees for Gr1Holozoa (Gr1Hz) homologs.

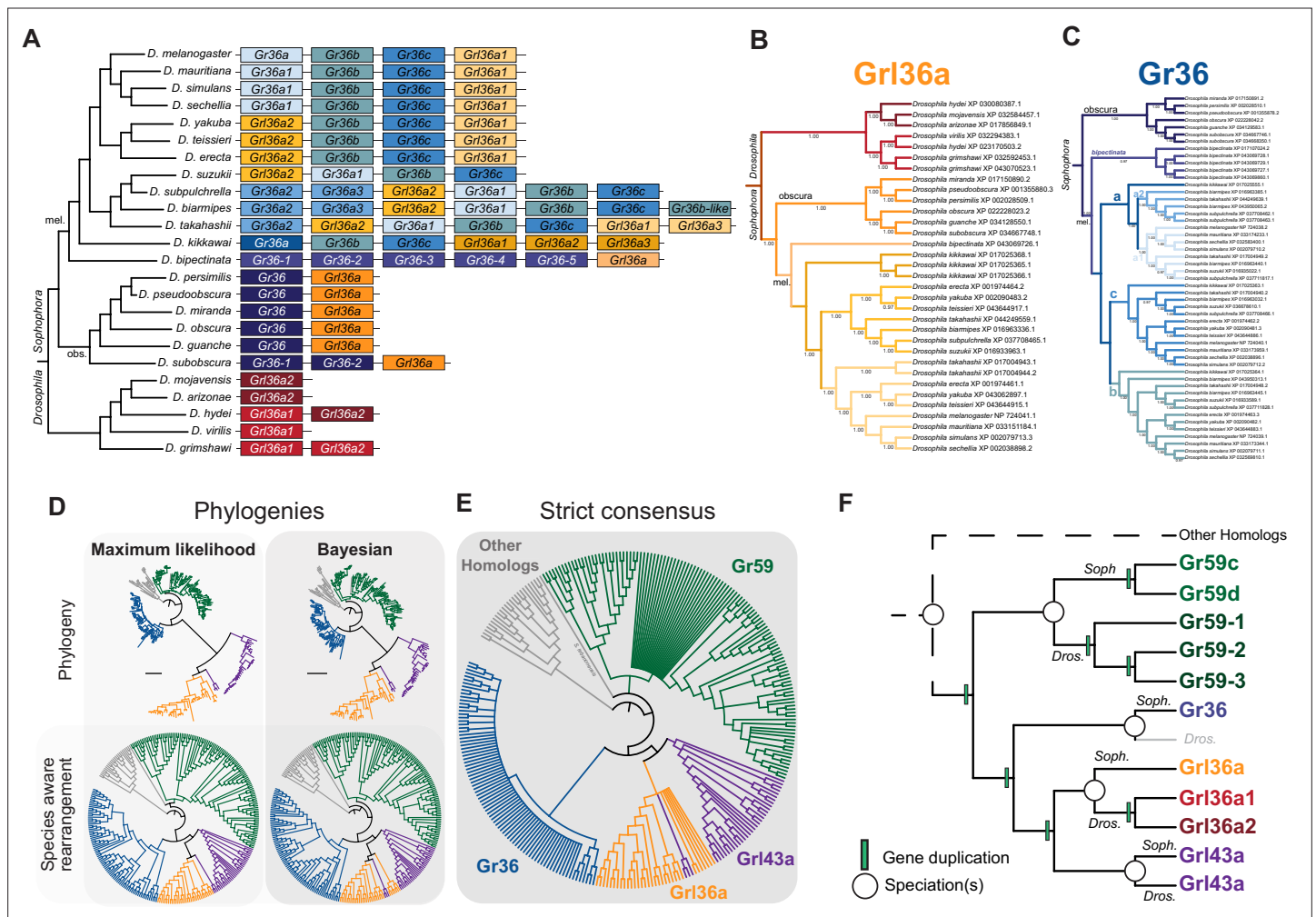


**Figure 3—figure supplement 5.** Fully annotated graph splitting tree for odorant receptors (Ors), gustatory receptors (Grs), and GRLs. Key edge perturbation support values are visible on branches. The primary sequence databases were assembled using each of the *D. melanogaster* GRLs as query sequences. *D. melanogaster* Or and Gr sequences were manually collected from FlyBase. Sequences from *M. hubei* (jumping bristletail), *Thermobia domestica* (firebrat), *Ladona filva* (dragonfly), and *Ephemera danica* (green drake mayfly) were added, following the proposal that canonical Ors may

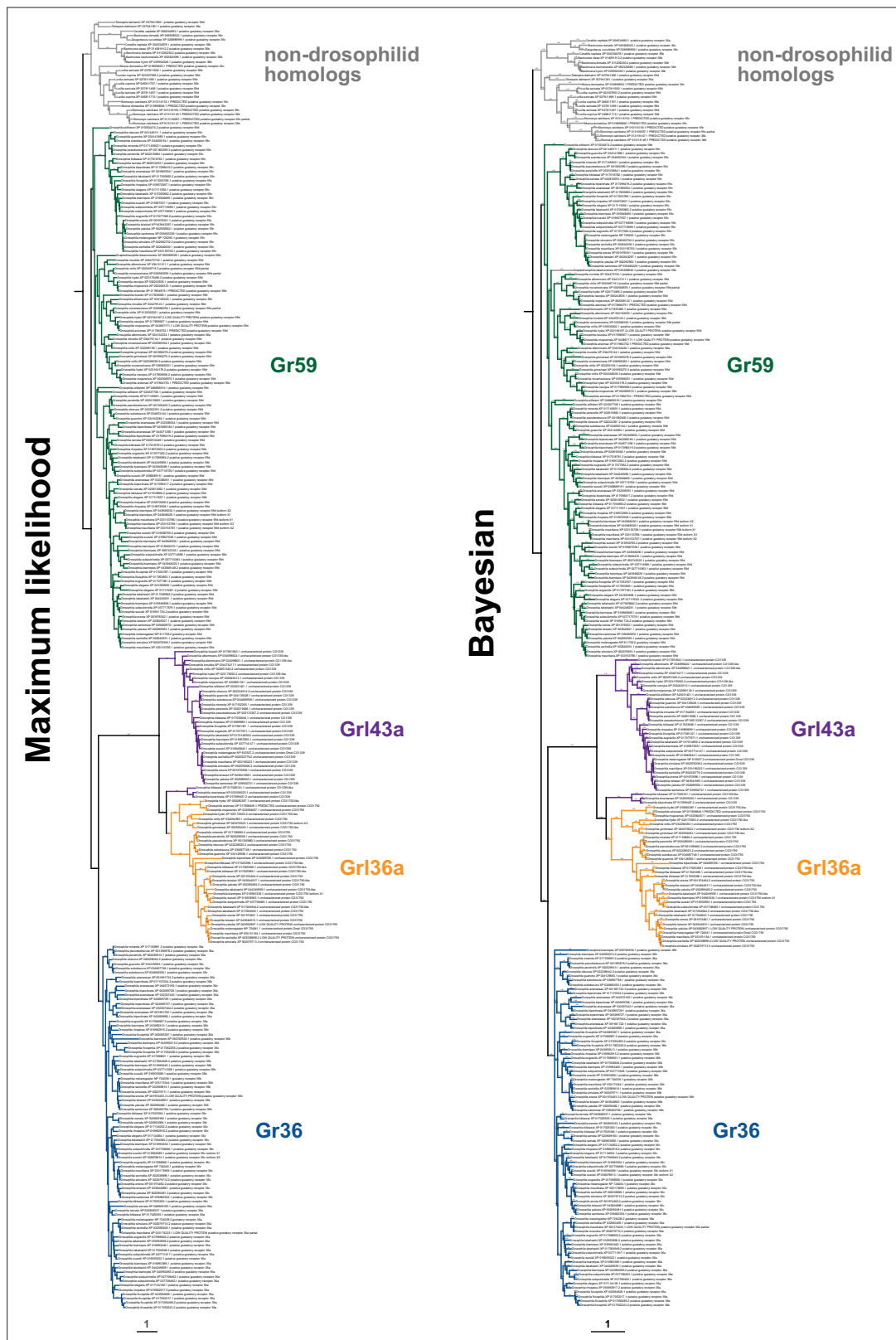
Figure 3—figure supplement 5 continued on next page

*Figure 3—figure supplement 5 continued*

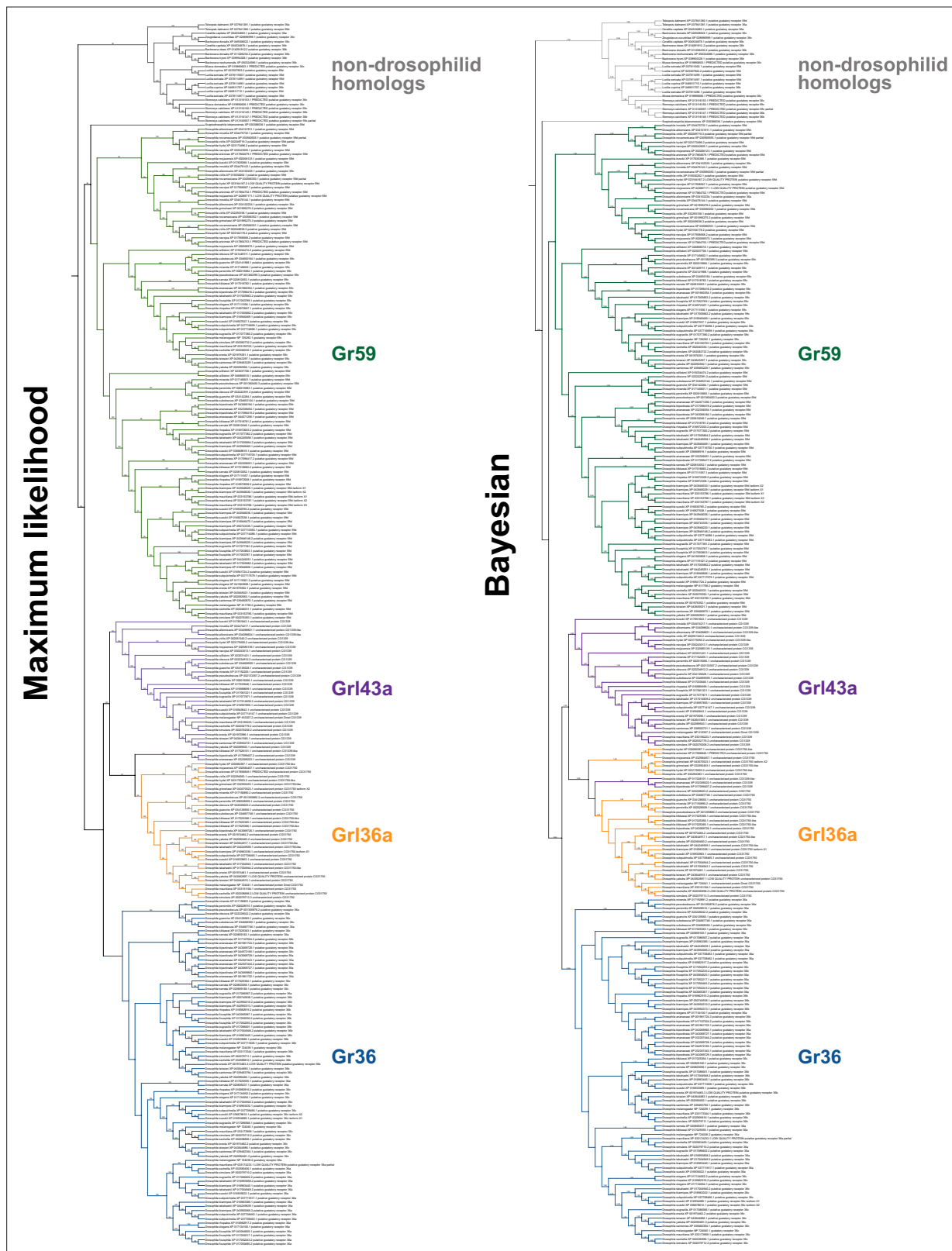
have diversified after the emergence of Neoptera (most winged insects) (**Brand et al., 2018**); 2498 additional sequences were collected using the *N. vectensis* GRL1 query sequence (XP\_048580785.1); the PSI-BLAST searches were stopped at four iterations, as the search had substantially recovered insect Gr sequences, and further searches returned tens of thousands of sequences. The basal placement of the Grs is unusual given their conservation in flies, as this would suggest they diversified in a common animal ancestor and that the Grs were lost in all animal taxa except flies. This hypothesis seems unlikely given the extreme number of independent gene loss events this would require, and we therefore suspect that this tree topology represents a phylogenetic error, for example, long branch attraction (**Bergsten, 2005**). The inset shows major collapsed clades, where the tip node is sized proportionally to the number of sequences collapsed.



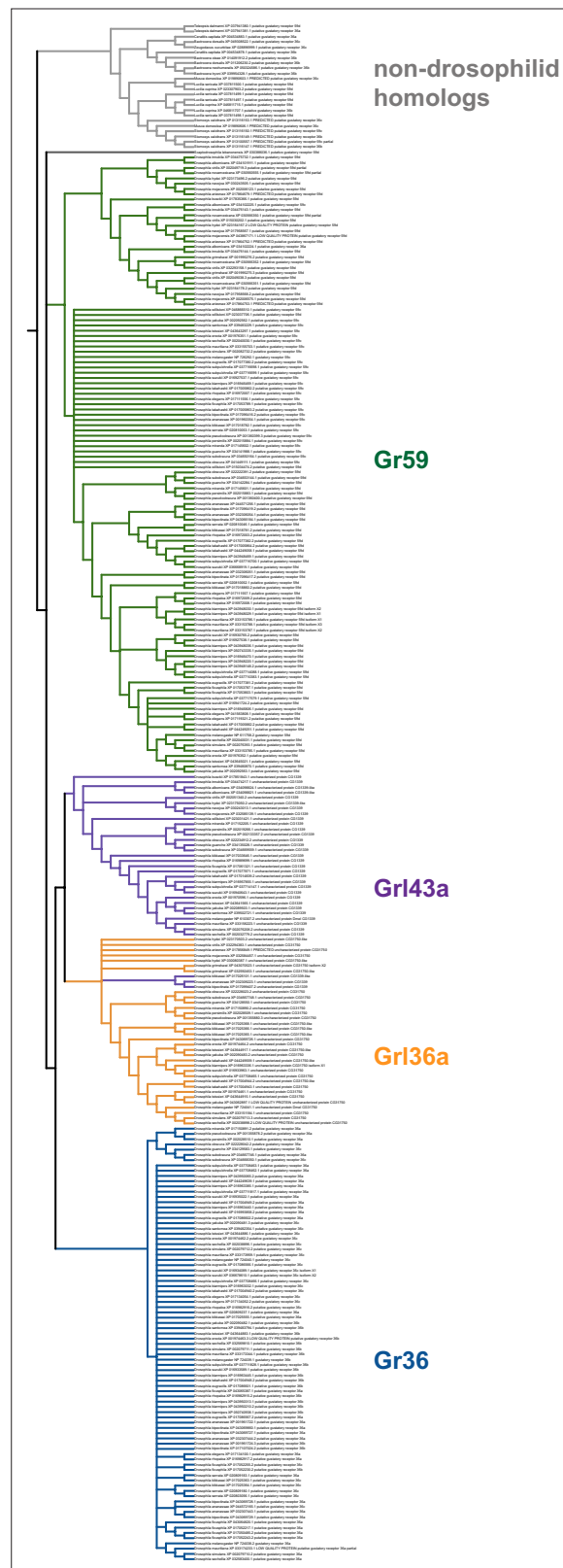
**Figure 3—figure supplement 6.** The evolution of *Gr36*, *Gr59*, *Grl36a*, and *Grl43a*. **(A)** Schematic of the gene arrangement of *Grl36a* and *Gr36* homologs in drosophilids, with colors matching trees in **(B)** and **(C)**. This panel is reproduced from **Figure 3C**. **(B)** Species-aware Bayesian phylogeny of *Grl36*. **(C)** Species tree-aware Bayesian phylogeny of *Gr36*. **(D)** Phylogenies of *Gr36*, *Gr59*, *Grl36a*, *Grl43a*, and other homologous sequences. The sequence database was assembled using *D. melanogaster* *Gr36a*, *Grl36a*, and *Grl43a* as query sequences. Top: maximum likelihood phylogeny and Bayesian phylogeny. The scale bars represent the average number of substitutions per site. Bottom: phylogenies where weakly supported branches (<95/0.95) have been rearranged and polytomies resolved in a species tree-aware manner. **(E)** Strict consensus of the species tree-aware phylogenies. These analyses support that *Gr36* and *Grl36a/43a* are sister clades, which likely split after the *Gr59* split. **(F)** Proposed model of *Gr36*, *Gr59*, *Grl36a*, and *Grl43a* evolution.



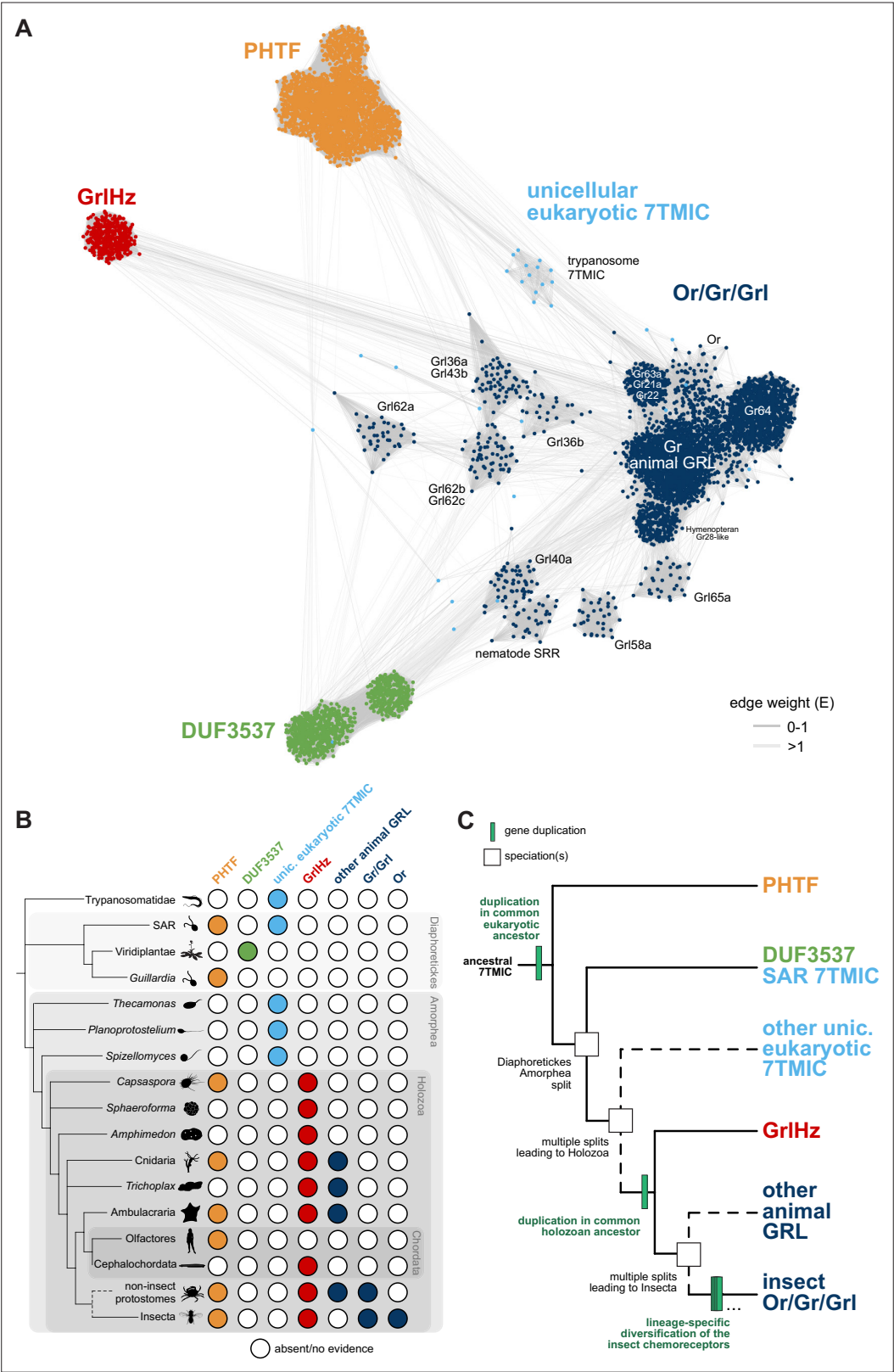
**Figure 3—figure supplement 7.** Fully annotated phylogenetic trees for Gr36, Gr59, Grl36a, and Grl43a homologs. For maximum likelihood, the tree was generated using a JTT + F + R7 substitution model and is rooted. Branch support values for maximum likelihood (UFboot) and Bayesian analyses (posterior probability) are shown at the branches. Non-drosophilid sequences are assumed to be the outgroup. The scale bars represent the average number of substitutions per site.



**Figure 3—figure supplement 8.** Fully annotated species-aware trees for Gr36, Gr59, Grl36a, and Grl43a homologs. Trees are based on the maximum likelihood (left) and Bayesian (right) trees. Branches without support values were eligible for rearrangement.

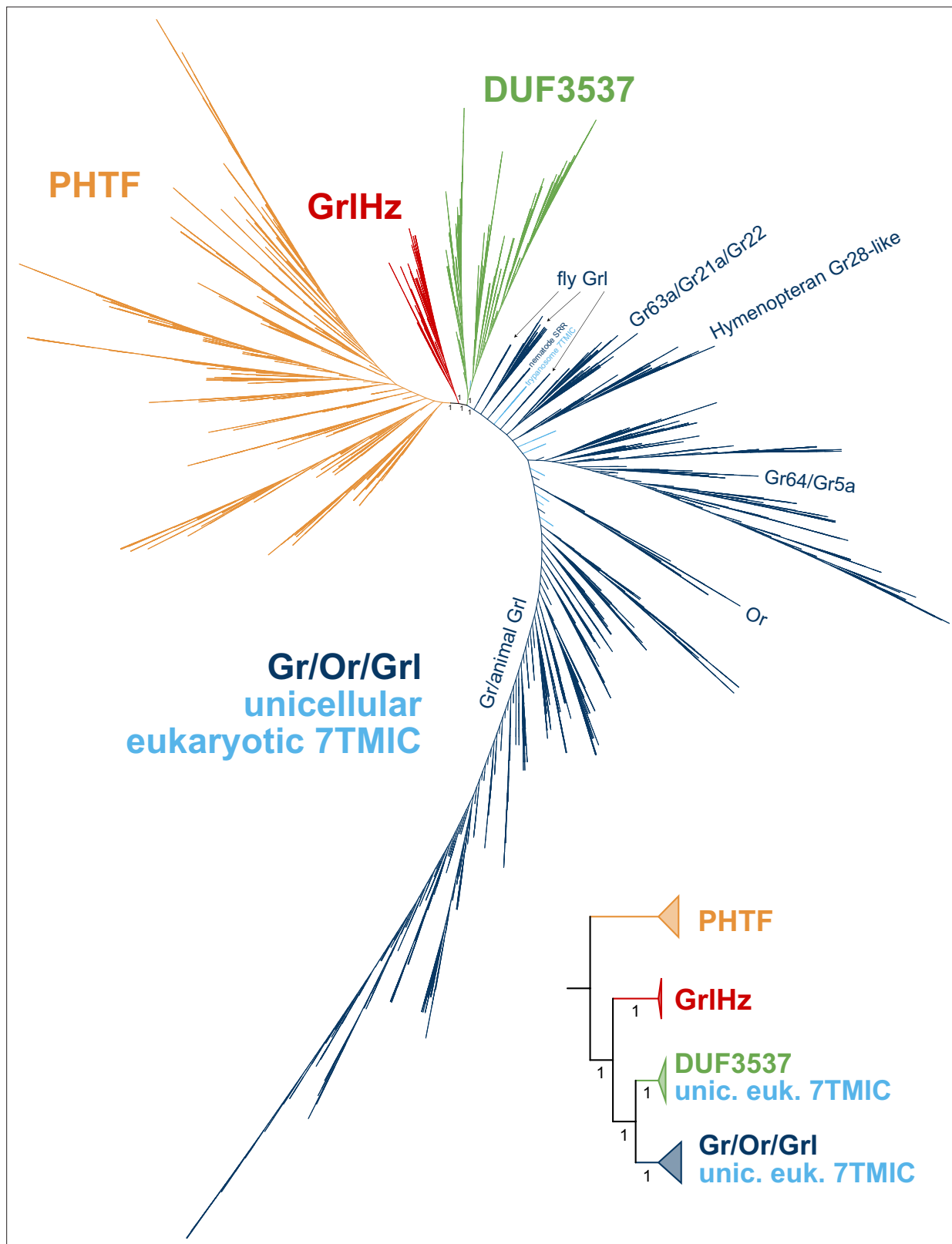


**Figure 3—figure supplement 9.** Strict consensus of the species-aware trees for Gr36, Gr59, GrI36a, and GrI43a homologs. Although the consensus tree has a polytomy near the emergence of Gr59, this is strictly due to disagreement as to whether the lone *Scaptodrosophila* sequence is a Gr59 homolog or an outgroup to all other *Drosophila*/*Sophophora* sequences shown here.



## Figure 4 continued

*H. sapiens* PHTF1 and PHTF2, *Arabidopsis thaliana* Domain of Unknown Function (DUF) 3537, *C. elegans* SRRs and trypanosome GRLs. The network was generated and visualized as in **Figure 3B**. The graph splitting tree is visualized in **Figure 4—figure supplement 1**. **(B)** Presence and absence of 7TMICs across taxa: 'other animal GRL' refers to GRLs in non-insect animal species previously identified by primary sequence similarity (**Benton, 2015; Robertson, 2015; Saina et al., 2015**) and nematode SRRs. The dashed branch represents several collapsed paraphyletic clades. **(C)** Model of 7TMIC superfamily evolution. The dashed branches represent several collapsed paraphyletic clades and speciation events. The trypanosome 7TMICs are unplaced due to the currently unresolved taxonomy of trypanosomes (**Burki et al., 2020**).



**Figure 4—figure supplement 1.** Graph splitting tree for the proposed seven transmembrane domain ion channel (7TMIC) superfamily. Key edge perturbation support values are visible on branches. The inset shows major collapsed clades, where the triangular tip is sized proportionally to the number of sequences collapsed. This tree suggests a different branching pattern than the hypothesis in **Figure 4C**, consistent with a more complex duplication/loss history for the 7TMIC superfamily. However, as in **Figure 3—figure supplement 5**, we suspect long branch attraction is present in this analysis, at least for the fly GrIs and nematode proteins.