SUPPLEMENTARY MATERIALS

Evolutionary History of Chemosensory-Related Gene Families Across the Arthropoda

Seong-il Eyun[1], Ho Young Soh[2], Marijan Posavi[3], James B. Munro[4], Daniel S.T. Hughes[5], Shwetha C. Murali[5], Jiaxin Qu[5], Shannon Dugan[5], Sandra L. Lee[5], Hsu Chao[5], Huyen Dinh[5], Yi Han[5], HarshaVardhan Doddapaneni[5], Kim C. Worley[5], Donna M. Muzny[5], Eun-Ok Park[6], Joana C. Silva[4], Richard A. Gibbs[5], Stephen Richards[5], and Carol Eunmi Lee[3]

[1]Center for Biotechnology, University of Nebraska-Lincoln, Lincoln, NE 68588, USA
[2]Faculty of Marine Technology, Chonnam National University, Yeosu 550-749, Korea
[3]Center of Rapid Evolution (CORE), University of Wisconsin, Madison, WI 53706, USA
[4]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA
[5]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA
[6]Fisheries Science Institute, Chonnam National University, Yeosu 550-749, Korea

SUPPLEMENTAL FIGURES AND TABLES

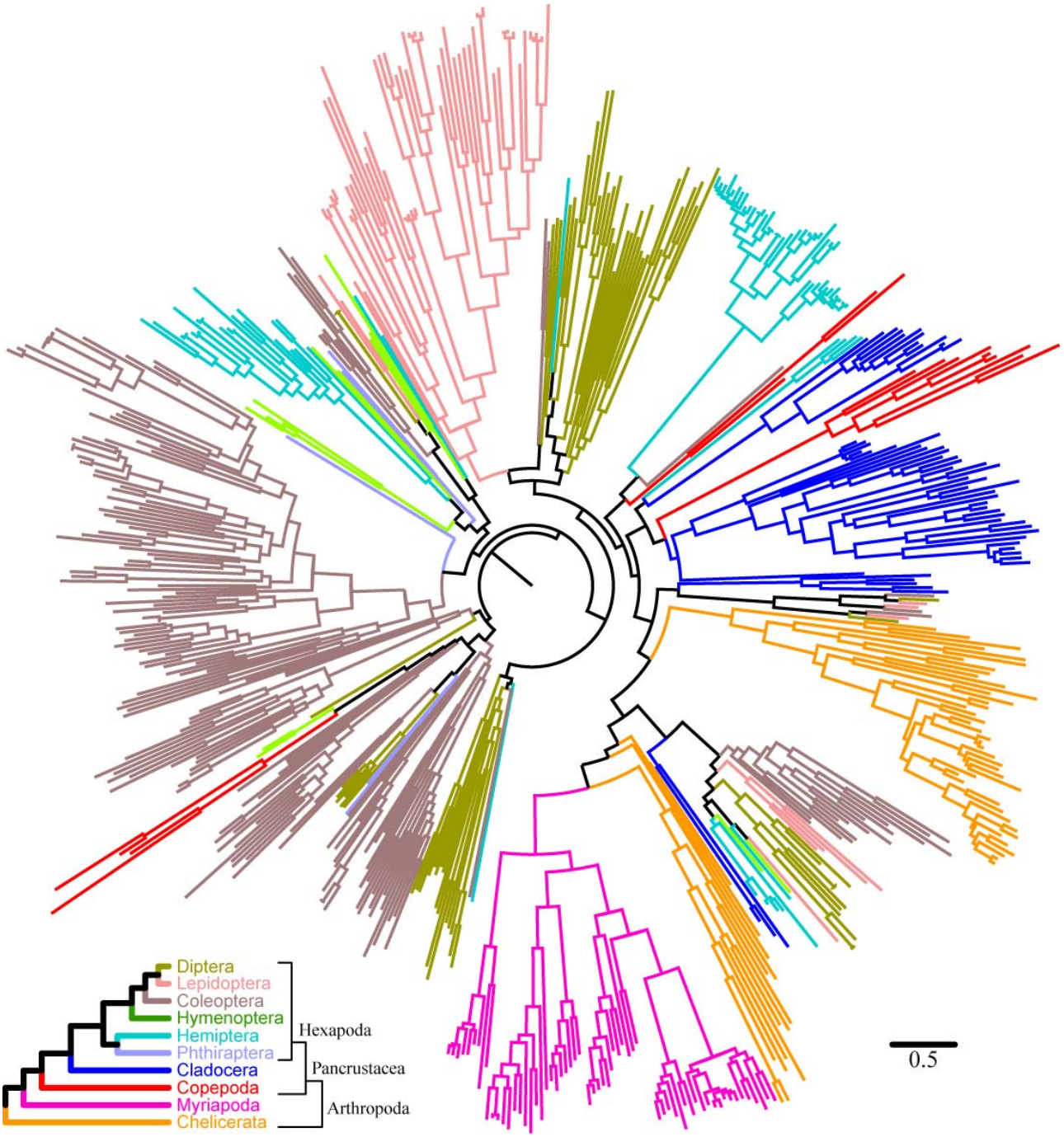


Figure S1. Expansions and contractions of gustatory receptors (GRs) in the Arthropoda. The phylogeny was reconstructed using maximum-likelihood for GR genes with a sequence alignment length of 1,916 amino acids. The bootstrap support values for a phylogeny using five

species (major subphyla of Arthropoda) are shown in Figure 2. The following representative species were used in this analysis: the fruitfly *Drosophila melanogaster* (Diptera, olive), the silkworm moth *Bombyx mori* (Lepidoptera, pink), the red flour beetle *Tribolium castaneum* (Coleoptera, brown), the honeybee *Apis mellifera* (Hymenoptera, dark green), the pea aphid *Acyrthosiphon pisum* (Hemiptera, cyan), the human body louse *Pediculus humanus* (Phthiraptera, slate-blue), the waterflea *Daphnia pulex* (Cladocera, blue), the common copepod *Eurytemora affinis* (Copepoda, red), the tide pool copepod *Tigriopus californicus* (Copepoda, red), the centipede *Strigamia maritima* (Myriapoda, magenta), and the black legged tick *Ixodes scapularis* (Chelicerata, orange). The scale bar represents the number of amino acid substitutions per site. The high-resolution phylogenetic tree and the Newick format can be available at: the Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.ts747.

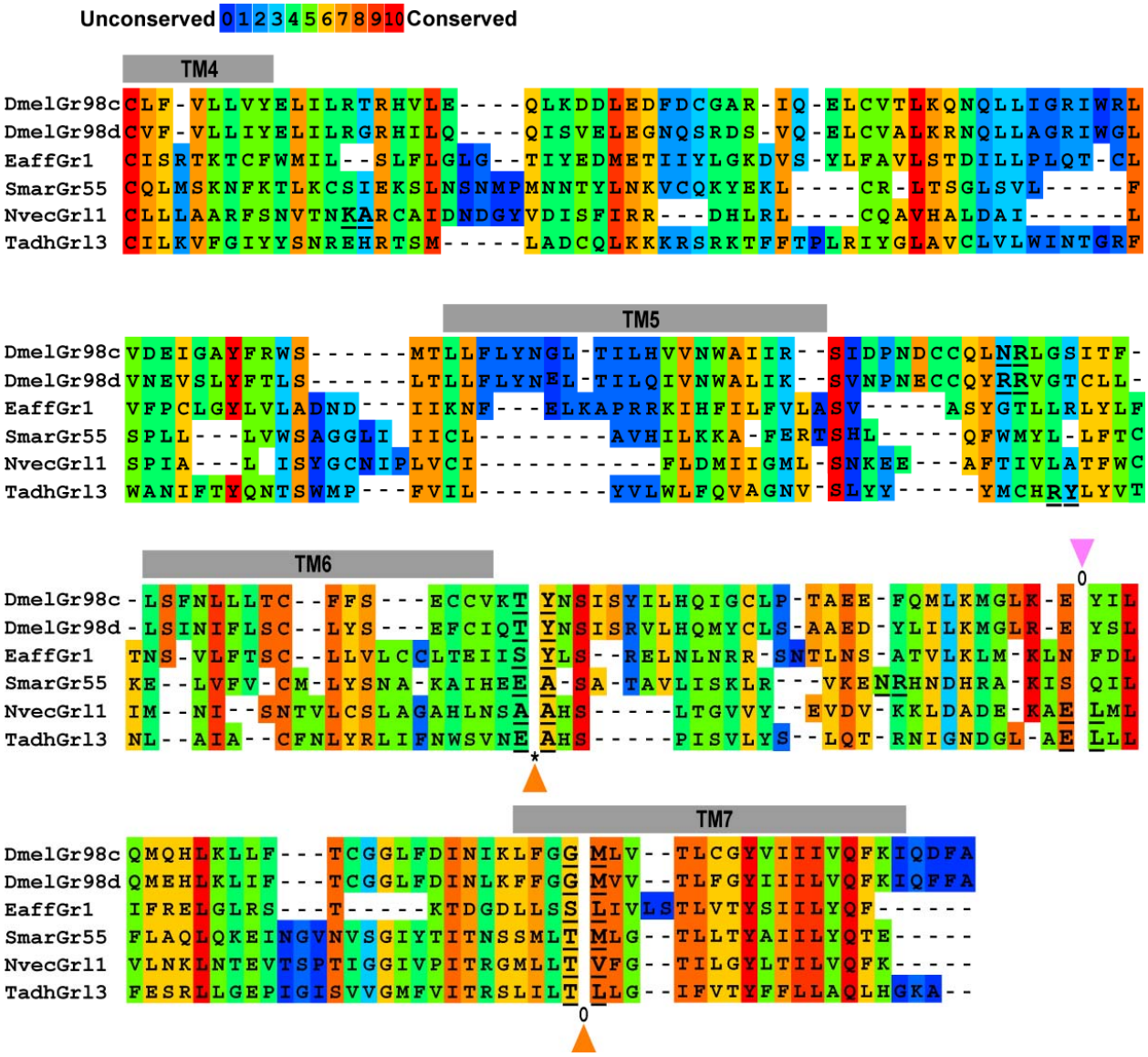


Figure S2. Partial sequence alignment and locations of introns of selected *Drosophila melanogaster* GRs and four other GR candidates from four eumetazoan species. The number of introns is two in EaffGr1 (*Eurytemora affinis* Gr1), three in DmelGr98c (NP_524531.2), DmelGr98d (NP_733214.1), SmarGr55 (*Strigamia maritima* Gr55), and four in NvecGrl1 (*Nematostella vectensis* AJP16415) and TadhGrl3 (*Trichoplax adhaerens* XP_002110324). The positions of the introns are marked with underlines (between two amino acids). The two orange arrows below the alignment indicate the introns that are shared among six genes and a pink arrow above the alignment indicates that an intron that is found only in the NvecGrl1 and TadhGrl3 genes. The numbers above the orange and pink arrows show the phase. An asterisk indicates phase 0 in DmelGr98c, DmelGr98d, EaffGr1, NvecGrl1 and phase 2 in

TadhGrl3. In this position of TadhGrl3, a non-GT-AG intron was found. The degree of conservation for each amino acid is indicated by color according to the index shown above the alignment (from blue to red on the basis of increasing conservation score by PRALINE) (Pirovano et al. 2008). The transmembrane regions (TM4, TM5, TM6, and TM7) are predicted using TMHMM (ver. 2.0) (Krogh et al. 2001).

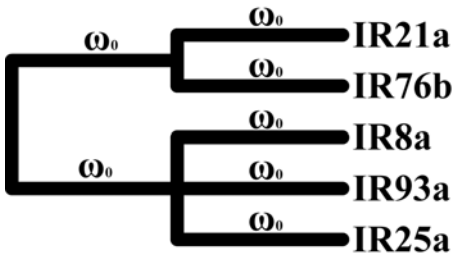
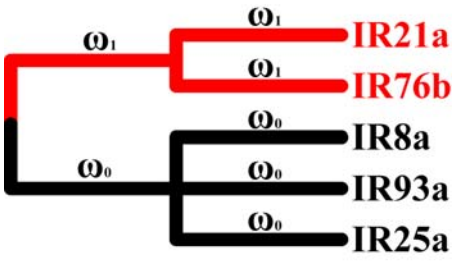
H_0	H_1	$2\Delta lnL$
		4.2722 (**P = 0.0387**)
ω_0 = 0.024 (lnL = -36259.75602)	ω_0 = 0.0131, ω_1 = 0.0249 (lnL = -36257.6199)	

Figure S3. Tests for signatures of selection in antennal IRs, which show elevated gene expression in males of the copepod *Eurytemora affinis*. PAML branch-model tests were performed between significant male-biased IRs (IR25a, IR93a-1, and IR8a) and unbiased IRs (IR76b and IR21a). This test was performed comparing the two hypotheses: H_0 (a single ω for all branches) and H_1 (two independent ω's: ω_0 for the black lineages and ω_1 for the red lineage), where ω is the ratio of nonsynonymous to synonymous substitution rates. For the likelihood ratio test statistics, $2\Delta lnL$, *P*-values (shown in parentheses) were obtained based on a χ^2 distribution with d.f. = 1. Therefore, male-biased expressed IRs showed significant signatures of purifying selection relative to the unbiased IRs (P = 0.0387).

(a) Amino acid sequences of chemosensory proteins (CSP) for six copepod species

```
CrogCSP1  ---------------------------------------------MRGIAVFLLVCCVGLSLGQ
CrogCSP2  ---------------PPPPPVRSRQAPPRPLPPRRRPPPPPRRPGPPRRGPPSKSGIAGIGSSI
LsalCSP   ------------------------------------------------------------------
TcalCSP   PRQPLPPQAPPRNFDQKGEELPLPPPPPPPGFQRRPRPRPRPRPNQGGSVLGSIGTRISNFGNNI
LcypCSP1  ----------------------------------PYRSQPRPPPVRSSEENGGILAGIGSGLTGI
LcypCSP2  ---------------------------------------------MKCPILCCLLVALLGLSMVTAQ
AfosCSP1  ------------------------------------------------MAKVLVVLICLVTISAAM
AfosCSP2  ------------------------------------------------MMAKGFLLLLIIAYTSAM
AfosCSP3  ----------------------------------------MSPKTVLLYLLLAVCLLYADGMSL
AfosCSP4  ----------------------------------------MLRQLTVTVLVTLMTVSLVQCQG
EaffCSP1  ------------------------------------------------MFRVFLLSSLLQLVFSE
EaffCSP2  ----------------------------------MLNLSYISFTISCFLFSAENSAANPHRLG
EaffCSP3  ---------------------------------------MFKFCIFCVVSLTSLTSSERSY
EaffCSP4  ------------------------MNLMLMFNVQVNMLKLRFGLILLLSFYLSTSLAQLEIDL

CrogCSP1  LC-PKDAAIEQKLK--DSKFIQKQVKCIL-GTGKCNRLGGRLKRKVPEAIR-GVCPKP-CKPCDQ
CrogCSP2  SCFTEDVAADYRLS--DKNYMQAQIKCVL-NEGPCDKVGGAIKRLAPEIFR-GRCPHP-CNPCKK
LsalCSP   --------ADYRLS--DKNYMQSQIKCVL-NEGPCDKIGGAIK--APEIFR-GRCPHP-CNPCKK
TcalCSP   KCAAEDAAANVKLN--DRAFIKQQVQCVL-DQGPCDETGRSIKLLAPEVMK-GRCPPP-CDPCKQ
LcypCSP1  KCAAEEWAAESKLK--DETFVRQQLDCVM-DRGPCDDFGLRIKRLAPDVMR-GLCPHP-CDQCVK
LcypCSP2  FC-LEDNYVDEKLN--NAPLVKAQLNCVLHEEAQCDQLGHQLRAIAPDVLR-GNCFPP-CTACIK
AfosCSP1  KC-GVDRDVRFILNNFNPTAVKKNVDCVV-GAGACDSLGSRLRAEARPAVRQGRCGKS-CT-CDQ
AfosCSP2  KC-SVDSEVRFILNNFNRGAVTRNVDCVV-GSGPCDNLGNRLKAEAPNAVRQGRCGSS-CT-CDQ
AfosCSP3  KC-KPDKEVRFILRNMNPVAVRKEVDCVV-GIGKCDKLGVKLKAEAPNAVRRRKCRKGECN-C--
AfosCSP4  RC-DIDERVKLALRTMNHQTVPAVVNCVL-GLSSCDSTGSWIKQNARDGACGRRCGSN-CT-CRE
EaffCSP1  TC-VVDREARFILDNMTAAAVKKEVDCII-GQGACDNLGFRLKEVAGKAVREGRCAGN-CD-CEQ
EaffCSP2  AC-KPTNGVENILRHMSAESVRKRVDCAV-GIGRCDSIGKKLKEEGPKALQAAKCGNN-CT-CEE
EaffCSP3  RC-QVDTEVKFILNNMTPLAVRKEVDCVV-GVGSCDALGRRIKARAADVVQRPRCAKP-CS-CEE
EaffCSP4  AC-PQELALLPQLE--NKSFMNAQIACML-DEGPCDPIGKKAKRIGADVLS-GHCVGP-CDKCTK

CrogCSP1  KQIQRVIQVMQEKYPSEWASILSNYNGK-----
CrogCSP2  KQIQKIMARVSKEYPYEWTQMIKIFG-------
LsalCSP   KQIQKIMARVSKEYPYEWAQMIKVFG-------
TcalCSP   KQIQKIMATLSKEYSQEWGQIL-----------
LcypCSP1  GKIQKVMSFIARNY------------------
LcypCSP2  KQVRKVVITIQTKYPKEFSQLVNKYTKKGRR--
AfosCSP1  IQVRLVVNKLKREYPSQWSRVVSHFS-------
AfosCSP2  IQVRLVVNKLKREYPGQWTRVVNRFS-------
AfosCSP3  FQVRRVVNKLQRDYRDQWQRVVDYFSK------
AfosCSP4  VQIRLIVRKMKQEFRDQWARIENHYRC------
EaffCSP1  IQVRLVVNKVKRLYPEEWKRVQDYYSK------
EaffCSP2  KRLRRLVNMMKRKHPEQWRRIEEYFNL------
EaffCSP3  INVRLVVNKMKRNHMDQWRRVENHFR-------
EaffCSP4  QAIRKIISKMQNNYPKAWNRIQQKLISERRPGK
```

(b) Conserved CSP signature motifs in Arthropods

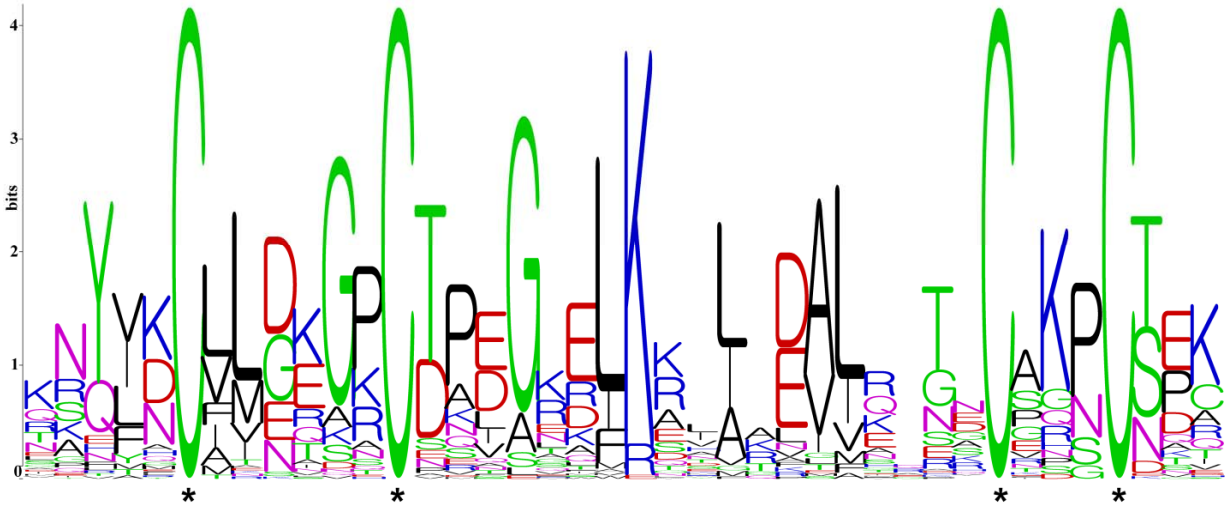


Figure S4. Chemosensory protein (CSP) in Arthropods. (a) Amino acid alignment for CSP genes from six copepod species. Fourteen copepod CSP protein sequences were identified in this study. Species abbreviations are given in fig. 3**.** The signal peptide sequences were predicted using SignalP (ver. 4.1) (Petersen et al. 2011) and are indicated with grey shading. Species labels for partial sequences are shown in italics. The transmembrane regions, which are underlined, correspond to the helices of the moth *Mamestra brassicae* CSP (MbraCSPA6, PDB: 1N8V, see supplementary fig. S7). The highly conserved four cysteines forming two disulfide bridges are highlighted in red (Forêt et al. 2007). Note that two additional cysteines (highlighted in cyan) are found only in copepods among arthropod CSPs. The number of disulfide bonds was predicted using DISULFIND (Lin et al. 2013) and DiANNA (ver. 1.1) (Ferre and Clote 2006). The median number of disulfide bonds is 2 in insect CSPs and 3 in copepod CSPs. **(b) Conserved CSP signature motifs in Arthropods.** To construct the CSP signature motifs, ninety five CSP sequences from seventeen arthropod species were used. Conserved amino acid patterns, based on the multiple sequence alignments from positions 44 – 81 (numbering according to *Drosophila melanogaster* CSP1: NP_611990), are shown using the sequence logo (http://weblogo.berkeley.edu) (Crooks et al. 2004). The height of each amino-acid letter is proportional to its frequency of occurrence at a given position. The highly conserved four cysteines forming two disulfide bridges are marked with asterisks (Forêt et al. 2007).

```
                                                                                                  <---ATD
DmelIR25a  ------------------------MILMNPKTSKILWLLGFLSLLSSFSLEIAAQTTQNINVLFI
DpulIR25a  ------------------------------------MRSFQLLLVLGLAFAVQSADPIRVLLV
AcalIR25a  MKAPRPQPESTPRRNYYMQTFLQRLGKLTIMAPNINTFTSIVFLVVVLTLVKIGAFLTQPTVVVVI
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  ------------------------------------------------------------------
EaffIR25a  ------------------------------------MVLQLLAGLPSILLLITPTFALNVYVV
LcypIR25a  ------------------------------------------------------------------
CrogIR25a  ----------------------------MPLLHLMMWAYNILRVSLVLSLLLGSLESAEVNVLWV
TcalIR25a  ---------------------------------MWAHAVTCGLWVLVTLNLGLVQSQ--KILWV

           ------------------------------ATD----------------------------------
DmelIR25a  NEVDNEPAAKAVEVVLTYLKKNIRYGLSV-QLDSIEANKSDAKVLLEAICNKYATSIEKKQTPHLI
DpulIR25a  YETNNVDADRAFTAVQSYLERTKVQGLSLGNVTRVTLDSTQKYLTVDDVCSVYDKSIDAGTPPHIV
AcalIR25a  DQDIDLYYNQNLKHIFENMHKFADRDKKKQEVVFTVVEARNDTDVLDEVCALLHSGA------VAL
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  -------------------------AKVAKKSVVEVEGTEAYVTVEKVCGELDSMLDGGTPPDLV
EaffIR25a  SDKNNGIADDSTKVAFDYLASRPTDPVTINAKKFSTLEGTDPKVVIDKVCGDLDAMIDGGQTPDLL
LcypIR25a  ------------------------------------------------------------------
CrogIR25a  NDGGMPVVKDSVNTAIEFVKR--EKGVSVPSLESVMMNENELENSTNAVCSKLRTMLDAGTPPDMI
TcalIR25a  NDMRNKIADESVDRIFD---DNKGTGLEEAELIKVDI-GEDPTESVTNICNALEGEV--GETISAI

           ------------------------------ATD----------------------------------
DmelIR25a  LDTTKSGIASETVKSFTQALGLPTISASYGQQGDLRQWRDLDEAKQKYLLQVMPPADIIPEAIRSI
DpulIR25a  LDLTWSGLSSEVMKALTRNLGLPTISGSYGGIGDLKHWSNIDGNQTKYLVQVMPPSDIIPQLVALI
AcalIR25a  IDMS-TPSLSPLLRSTCSSMGLTYISVV------DRSYFNHPIESKGTELMIEPPASQMLKVVADI
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  LDLTKIGISSEVVKSLSLTLGLPTVSAALGEEGDIREWRELTDAQKQYLIQVRPPGDLLPNIIRDL
EaffIR25a  LDLTRGGVNSEVVKSLSLTLGIPTVTATYGSVGDIREWKELTMEQSKYLIQVRPPGDTLPDIIRSL
LcypIR25a  -----------------*ELGLPTVST------NPRTWKELTADEAEFLVHIKPPIMGMLASLSEL
CrogIR25a  LDTTSAGALSEVVKSLSLSLGIPTLSTSYGDMQDVREWRNLTMNQEKYLIRVRPPGDILPNIIRSI
TcalIR25a  YDTTYGGKSTEALKSLALSMGIPMLSFNFLKYG---EWGELSDNQMKYLITVNPPGHTLVRAIRNI

           ------------------------------ATD----------------------------------
DmelIR25a  VIHMNITNAAILYDDSFVMDHKY-KSLLQNIQTRHVITAIAKDGKREREEQIEKLRNLDINNFFIL
DpulIR25a  TSMQNMTNAAILYDDSFDMLNKY-KSLLKNRPIRHMFSKI----ETNINTQIRRLEDMDIVNFFVL
AcalIR25a  AKKEQLNNIAIIYDETFVL------SGVCIAAQRGLSASTLPDHNRNINMPVDEA-----------
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  AMASNITNAGILYDKTFVMDHQY-KSLLLNMPVRHIINKV-------------------------
EaffIR25a  VTDLNITNAGILYDESFVMDHKY-RSLLLNLPTRHIIKPL-ETTDAGIKDQLTRLESVDINNFFVL
LcypIR25a  ALGLDLRGAVIIHEGASSEHKKYIQSLLQGAGVFYLTQAL---NRKSVEKLVIKIQNSHIKNLFIL
CrogIR25a  ITTDNITNAAVLFDDSFVMHYKY-RSLLQNIPVRHMIQKI-RMGKNELKAQLLRMTEVDLNNYFVF
TcalIR25a  AEEGNVTNAAIVFDETFDMYNQY-GTLLLNVPTRHLIQPM-KKTMDEMKRQLTFFDEIDLHNIFLV

          ------------------------------ATD----------------------------------
DmelIR25a  GTLQSIRMVLES-VKPAYFERNFAWHAITQNE----GEISSQRDNATIMFMKPMAYTQYRDRLGLL
DpulIR25a  GKIDRINQVLMSAAQENYFGKKYSWTAISKDGS---AEPFVRTENGSILFAVPTVNPDVAN--GIL
AcalIR25a  ------------------------------------------------------------------
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  ------------------------------------------------------------------
EaffIR25a  ASLANLKMVYGVASRKGMTGEKYAWFAGTKDSGQEFDQTCCERMKVAFFHASPIPSNIMRS---LK
```

```
LcypIR25a  GSADTVNFVTQTLSTDKQ-NSSFAWFAVTKDKI-------SHSVVKTLLHLYPVSRYQ--------
CrogIR25a  GSVESIQRVVGLANELKLYGTKYSWFAITKESD---FNFKCSCEKLDLIFIKP---AKPGNGALIS
TcalIR25a  GSAENIKLYMAAANSMDMFGIKYSWFAVSKDP----YIANCDCSEVTLLSVNP---EVLES---EK

                              ATD
           ------------------------------ATD--------------------------------
DmelIR25a  RTTYNLNEE--PQLSSAFYFDLALRSFLTIKEMLQSGAWPKDMEYLNCDDFQGGNTP--QRNLDLR
DpulIR25a  FKTSGLNNG--YSVDTGFYFDLILRAITTVKNMLDGNTWPVDMSYSKCSMTNITAVT--RNNFDLR
AcalIR25a  -----------LAFDIGRIVTLALEAVPNVQNIVK----------VSCDNGTDPTPASLKQSAELT
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  ----------------------------------------------------KSXI--DRSIDLI
EaffIR25a  TMKNGLTGM--PEVDAGFYFDLAVRAATSMKKLMAGGSLP-NYSLSKCSEYTETTPV--ARSIDLL
LcypIR25a  ----TMGLR--SHIDSDFYFNVAVSAFTAISNMKASGMWPVTDGNPKCGQNMSSPFVG-RSDVSLR
CrogIR25a  SIVSGSDLK--PKLDAGFYFEVAARGLVGVSDAKDGSNWP-DFTIQKCSEFGNAPPTE-RSNLDLM
TcalIR25a  SLSSALDTSTLPPTDVGFYHATMKALLQAVSALTTEGKLPVTLE-QTCATFSNENPT--DRDIDLM

           ------------ATD------------>
DmelIR25a  DYFTKITEP--TSYGTFDLVTQSTQPFNGHSFMKFEMDINVLQIRGGSSVNSKSIGKWISGLN-SE
DpulIR25a  KAFADTNVG--STFANMILN-G-----NGKSYPQVEMTINQMNFKNSRLESKNALGIWKAGMP-GE
AcalIR25a  NELTVVSIP--GVLGPLSWSEQ-----SQALRYNMTLLLSEMFFESGILKSKDQVANWTQA---GG
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  SSIKSVSMP--DSFGSMMLV-S-----NGLSYMSFDLSVVTYEIMSGKIANREDIGTWSSGTP-GT
EaffIR25a  SVFKVDNPGMSYAYGPMKIS-S-----NGLSMMNFKVNLLSYTIESMKVADRKEIGSWEWG---SS
LcypIR25a  QQMVLAQS---ATYGPIILP-A-----TGSGCFDFQVNVSKLEFEEGKLVNSTLMGTWTSRRP-SA
CrogIR25a  AHFKDANVP--SVYVPLSYGMK-----DDEIYMKFGVSIEKHKYLNSKLVKKEYIGDWTAELSEDS
TcalIR25a  GAMSAAI----SAGSPSSLMIM-----DGKPQYNVNVVLRRLEMKSKKLVLVSDIGMSSPN---EV


DmelIR25a  LIVKDE--EQMKNLTADTVYRIFTVVQ-APFIMRDE----TAPKGYKGYCIDLINEIAAIVHFDYT
DpulIR25a  ISFSPG--QSLRPFQVISVFKIAVVVQ-APFIMKRK---ANGTVTFYGYCVDLIKDIQAIMGFEYE
AcalIR25a  LQLDV---PTLQKANKKKRYRIVTVAGIPPFVYKEEPINSTGPPVYKGYCIDLLERIAQDMNFDYE
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  MKFNAGMEETIVKFRAKTVYRVVTVVQ-PPFMLWNE-----TTKKYEGYCADLIEELTKIMDFEYE
EaffIR25a  IMYSPGMMNSMSGFRATTVYRVVTVLQ-APFMMMDE------QGEYSGYCVDLINELKDMMGFQYE
LcypIR25a  ISLEPNATQTLAKLFIKPFYRIVTVVQ-PPFVYKHQ--DPGSKVTFSGYCIDLIKDLAKLANFDYE
CrogIR25a  IQYKTG--MSLQGVGAKTIFRVVTVLQ-PPFVTKEM---VDGKPVYTGYCIDLINELKNLIEFEYE
TcalIR25a  INYKTG--KALGSVAATTVLRVVVVHQ-PPFVERV------GDGEFIGYCIDLINELKQLMGFEFE

                                 <------------------S1--*-------------------->
DmelIR25a  IQEVEDGKFGNMDENGQWNGIVKKLMDKQADIGLGSMSVMAERREIVIDFTVPYYDLVGITIMMQRP
DpulIR25a  LYEVPDGKYGNMDSKMNWNGMIKELMEKRADIGLGALAVMAERRENVIDFTVPYYDLVGISILMAKP
AcalIR25a  IHDVE--IVGSMDDDGNWSGVIKELIDNKADIAVGPISVMAERRENVIDFTVPYYDLVGLTILMRKP
AfosIR25a  ------------------------------------------------------------------
CsinIR25a  IYEAPDGKYGRMDENEEWDGLIAELVNKKADMAVGALSVMAERRENVVDFTVPYYDLVGITIMMKKP
EaffIR25a  IKVVED--YGNMDMDMSWNGMIRELVDKRADIALGALSVMAERRENVVDFTVPYYDLVGISILMKQP
LcypIR25a  LYEAPDGQYGRLSESLTWNGMIKELLDHKADIGLGALYVMGERREMAVDFTVPYYDQVGMAILMKKP
CrogIR25a  IYEAPDGQFGRMDENEDWNGMIKELIDKKADIALGALSVMAERRENVIDFTVPYYDLVGIIIMMKKP
TcalIR25a  LYEAPDAKYGRMDEEMNWNGMVKELVDKKADVALGALSVMAERRENVIDFTVPYYDLVGITILMKKP

                           <--------M1-------->
DmelIR25a  SSPSSLFKFLTVLETNVWLCILAAYFFTSFLMWIFDRWSPYSYQNNREKYKDD--EEKREFNLKEC
DpulIR25a  QVSTSLFKFLTVLENDVWGCILAAYFFTSILLWIFDRWSPYSYQNNKEKYADDPEEEKREFSLKES
AcalIR25a  RFDYSLVKFLNVLDEEVWGCIIGAFFLFSILICVFDKLSPFSYQNRKNQWKSS-GSEPRVFTLKEG
```

```
AfosIR25a  ------------------------------------------------------------LFDLKEC
CsinIR25a  KVPTSLFKFLSVLEDSVWGCILAAYFVTSVLMWIFDRWSPYSYQNNMEKYEDD--DEKRYFNFKES
EaffIR25a  QVPTSLFKFLSVLEDSVWFCILAAYVVTSVIMWIFDRWSPYSYQNNMEKYIDD--DEKRYFNLKEC
LcypIR25a  SNPTSLF-----------------------------------------------------------
CrogIR25a  KLPTSLFKFLSVLEDQVWGCILAAYFATSFLMWIFDKYSPYSYQNNIEKYEDD--DEKRYFNFKEC
TcalIR25a  KVPTSLFKFLSVLEDSVWGCILLAYFVTSLLMWLFDKFSPYSYQNNMEKYEDD--DEKRYFNLKEC

           <------P---->          <----------M2--------->
DmelIR25a  LWFCMTSLTPQGGGEAPKNLSGRLVAATWWLFGFIIIASYTANLAAFLTVSRLDTPVESLDDLAKQ
DpulIR25a  LWFCMTSLTPQGGGESPKHLSGRLIAATWWLFGFIIIASYTANLAAFLTVSRLDSPVNSLDDLSTQ
AcalIR25a  VWFCMMSLTPQGGGETPKALSGRLIAATWWLFGFIIIATYTANLAAFLTVSRLETPIESLDDLSEQ
AfosIR25a  LWFCMTSLTPQGGGEAPKNLSGRLVAATWWLFGFIIIASYTANLAAFLTVSRLETPVESLEDLAKQ
CsinIR25a  LWFCMTSLTPQGGGEAPKNLSGRLVAATWWLFGFIIIASYTANLAAFLTVSRLETPVESLEDLAKQ
EaffIR25a  LWFCMTSLTPQGGGEAPKNLSGRLVAATWWLFGFIIIASYTANLAAFLTVSRLETPVESLEDLAKQ
LcypIR25a  ------------------------------------------------------------------
CrogIR25a  LWFCLTSLTPQGGGEAPKNLSGRLVAATWWLFGFIIIASYTANLAAFLTVSRLETPVESLEDLSKQ
TcalIR25a  LWFCMTSLTPQGGGEAPKNLSGRLVAATWWLFGFIIIASYTANLAAFLTVSRLETPVESLEDLSKQ

               <--------*----------------------------S2----------------------
DmelIR25a  YKILYAPLNGS-SA-MTYFERMSNIEQMFYEIWKDLSLNDSLTAVERSKLAVWDYPVSDKYTKMWQ
DpulIR25a  YKVQYAPQNGTDVA--TYFERMAYIEKRFYEIWKDMSLNDSMNEVERSKLAVWDYPVSDKYTKIWQ
AcalIR25a  FKVQYAPMNGS-TA-MIYFKRMAHIEHRFYEIWKNMSLNDNLAAVERAQLAVWDYPVSDKYTKLWQ
AfosIR25a  YKIKYAPKNGSASA--TYFERMAYIEERFYEIWKDMSLDDSLTELERSKLAVWDYPVSDKYTKIWQ
CsinIR25a  YKIQYAPQNGSATA--MYFERMAYIEERFYTIWKEMSLNDSLSELERSKLAVWDYPVSDKYTKIWQ
EaffIR25a  YKIKYAPKNGSSAA--VYFQRMAYIEERFYEIWKDMSLNDSLNELERSKLAVWDYPVSDKYTKIWQ
LcypIR25a  ------------------------------------------------KLAIRTELVKDPHDK---
CrogIR25a  YKIQYAPMNGS--ATMTYFERMAYIEDKFYEIWKDMSLNDSLSELEKSKLAVWDYPVSDKYTKMWQ
TcalIR25a  YKIKYSPMNGS--ATMTYFERMAYIENKFYEIWKDMSLNDSLSELERSKLAVWDYPVSDKYTKMWQ

           -------S2------------------------*
DmelIR25a  AMQEAKLPATLDEAVARVRNS--TAATGFAFLGDATDIRYLQLTNCDLQVVGEEFSRKPYAIAVQQ
DpulIR25a  SMQDAGLPHTFDEALTRVRASNAENNDGFAFLGDATDIRYQVLVNCDLQMVGEEFSRKPYAVAVQE
AcalIR25a  TMQDSHFPSNKTEAVHRVL------NEDFAFISDATTNKYQTLINCDLWQVGEEFSRKPYALAVQE
AfosIR25a  SMQEAELPLTDEEAYARVRAS-PSSSEGFAFLGDATDIRYQTLINCDLQMVGEEFSRKPYALAVQQ
CsinIR25a  SMKEANLPATFEEALKRVRES-PSSSEGFAFLGDATDIRYQTYTNCDLQMVGDEFSRKPYALAVQQ
EaffIR25a  SMQEAKLPDTDEEAYARVRAS-PSSSEGFAFLGDATDIRYQTLTNCDLQMVGEEFSRKPYALAVQQ
LcypIR25a  ----------------------SSCEKG------------QXYTNCDLQMVGEEFSRKPYSIAVRH
CrogIR25a  SMKEANLPQTFTEALERVRSS-PSSSEGFAFLGDATDIRYQVLTNCDLQMVGEEFSRKPYAIAVQQ
TcalIR25a  SMQEAGLPQSFDQALEWVRDS-PSSSEGYAFLGDATDIRYQVLTNCDLQMVGEEFSRKPYALAVQQ

                                                                <-------M3----
DmelIR25a  GSHLKDQFNNAILTLLNKRQLEKLKEKWWKNDEALAKCDKPEDQSDGISIQNIGGVFIVIFVGIGM
DpulIR25a  GSPLRDLLNDAILRLLNQRRLETLKERWWTDNPEKQECGDTNDQSDGISIQNIGGVFIVIFVGIFL
AcalIR25a  GSPLRSQLSNIILKLINQRALEEMKTKWWKEDE--KECPKLENETDGISIRNIGGVFLVIVIGSGL
AfosIR25a  GSPLKDQLNDAILKLLNQRKLENLKEKWWNQNPLKKTCDDAEGESGGISIHNIGGVFIVVFVGIGL
CsinIR25a  GSPLKDQFNDAILKLLNQRKLETLKERWWNQKPSKE--------------------------GL
EaffIR25a  GSPLKDQLNDAILKLLNQRKLETLKERWWNQNPKKKVCDANEEDGGGISIHNIGGVFIVIFVGIGL
LcypIR25a  GSLLRDVLNDGILKLLNQRKLDYYKQSWWNLNPLRAICTDEEDSSGGISIYNIGGVFIVIFFGMAL
CrogIR25a  GSPLKDQFNDAILKLLNQRKLETLKERWWNGNRNKKNCEEDDDSGGGISIYNIGGVFIVIFVGIGL
TcalIR25a  GSPLKDQFNDAILKLLNQRKLETLKERWWNQNPKKKVCDDDDDSSGGISIYNIGGVFIVIFVGIGL

           ---------->
```

```
DmelIR25a  ACITLVFEYWWYRYRK--NPRIIDVAEANAER--SNAADHPGK--LVDGVILGHSGEKFEKSKAAL
DpulIR25a  ACVTLAIEYCYFKVRR--NPEGDEVVSTPESR----NNAAGNRNKLDDAYVKNSQKPFVLDDNSKD
AcalIR25a  SLITLAFECYWYRLR----PERKTLSKMYNGR----SKDTNSQGQLTTSGTATSVLASDSQMSKEN
AfosIR25a  AIITLAVEYWYYKSKR---PGSQVGSITKV------NPKFDDKLEYMRRDDKGVAVVANPW-----
CsinIR25a  ------------------------------------------------------------------
EaffIR25a  AIITLGIEYWYYKYKKPLSRVSSSAQNIIDKKGHAFDKSYGGRGSADVVPVGNPW-----------
LcypIR25a  AFITLAFEYCYFKYYKKATPKDNDSAQVFVVKEMAMKKASN-KIGIITD----------------
CrogIR25a  AIIILIFEYWYYKYKVPQNQISDKNLKILQVK----KAHIRDKTDFSSGFK---------------
TcalIR25a  AIITLIFEYWYYKNKKPVSPTGSQTKK-LEVKQALRDVNFGNKSDYETEYKNSGGGIAPIGNPW--


DmelIR25a  RPRFNQYPATFKPRF-----------------------------
DpulIR25a  PYAGDYGYYGAKKELELEGDGPRPRKAW----------------
AcalIR25a  QRNNKLETGGLDSGFVNTGFELNGGGLDSGFTSDRIIERTHTEWF
AfosIR25a  -----------------------------------------
CsinIR25a  -----------------------------------------
EaffIR25a  -----------------------------------------
LcypIR25a  -----------------------------------------
CrogIR25a  -----------------------------------------
TcalIR25a  -----------------------------------------
```

Figure S5. Alignment of IR25a protein sequences identified from six copepod transcriptomes. Six copepod IR25a protein sequences were aligned against those from three other species, namely the fruit fly *Drosophila melanogaster* (DmelIR25a, ADU79032.1), the waterflea *Daphnia pulex* (DpulIR25a, EFX86214), and the California sea hare *Aplysia californica* (AcalIR25a, XP_005102425.1). Abbreviations for the six copepod species used a four-letter code: Afos (*Acartia fossae*), Cfin (*Calanus finmarchicus*), Eaff (*Eurytemora affinis*), Lcyp (*Lernaea cyprinacea*), Crog (*Caligus rogercresseyi*), and Tcal (*Tigriopus californicus*). Species labels for partial sequences are shown in italics. Potential ligand binding amino acid residues are highlighted in red (Benton et al. 2009). Approximate regions for three transmembrane domains (M1-M3), the ion channel pore domain (P), and two ligand-binding domains (S1 and S2) are indicated above each alignment block. These regions were based on Benton et al. (2009) except for amino-terminal domain (ATD), M1, and M3 regions (shown with gray highlighting). The ATD region was derived from the Pfam domain (PF01094) (Finn et al. 2014). M1 and M3 regions were predicted using Phobius (Kall et al. 2007).

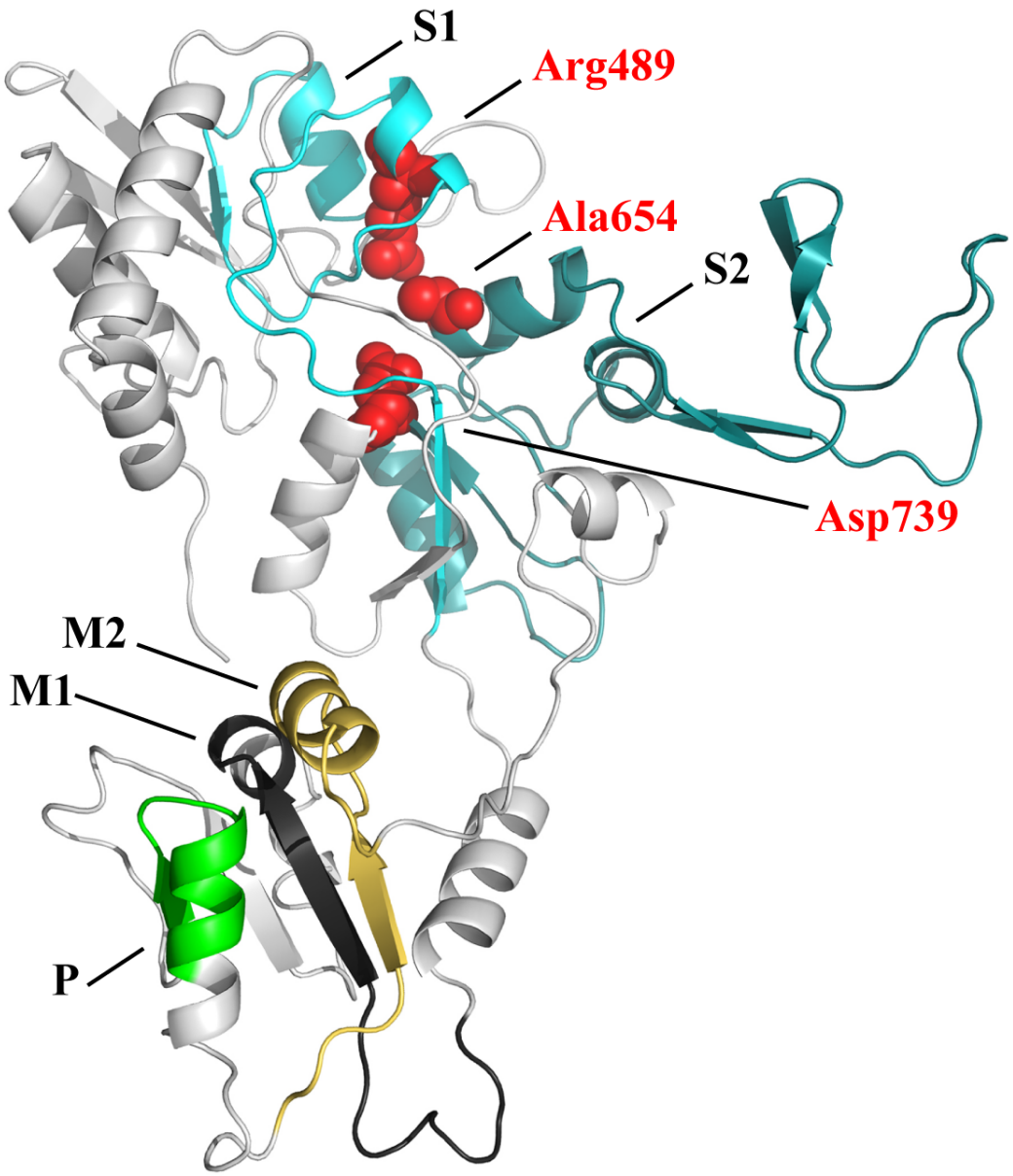


Figure S6. The 3D-structural model of the IR25a protein for the tide pool copepod *Tigriopus californicus*. The protein model construction above utilized the template for the known protein structure of the B-chain of the rat (*Rattus norvegicus*) ionotropic glutamate receptor (PDB: 1YAE, NP_062182.1), as selected by SWISS-MODEL (http://swissmodel.expasy) (Arnold et al. 2006). Three potential ligand binding amino acid residues (Arg489, Ala654, and Asp739) were shown in red (Benton et al. 2009). See supplementary fig. S5 for the alignment and more detailed information. The graphical representation was created using PyMOL (version 1.3) (DeLanoScientific, San Carlos, CA).

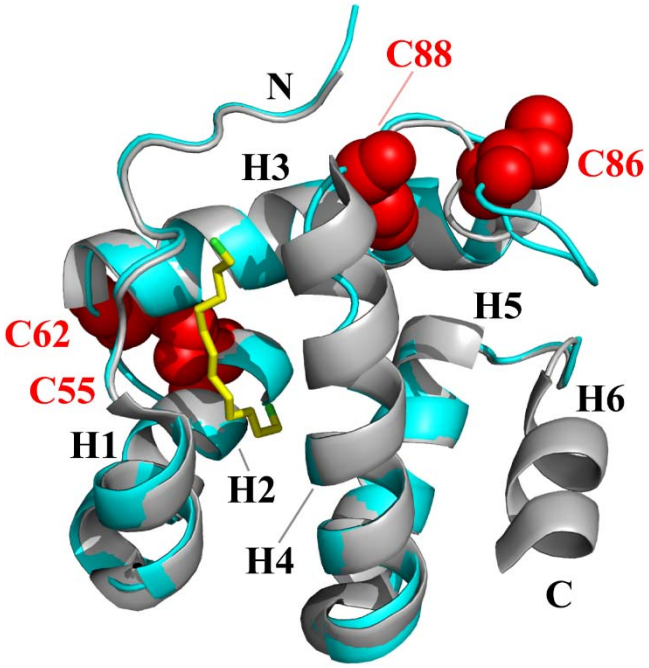


Figure S7. The 3D-structural model of the common estuarine copepod *Eurytemora affinis* CSP2 protein (EaffCSP2, cyan) superimposed with the moth *Mamestra brassicae* CSP (MbraCSPA6, gray). The template for constructing the protein model above (*Mamestra brassicae* CSP, PDB: 1N8V, AAF71289.1) was selected using SWISS-MODEL (http://swissmodel.expasy) (Arnold et al. 2006). The sequence similarity against *M. brassicae* CSP was 25.93% and the QMEAN4 Z-score given by SWISS-MODEL was -1.62. The conserved four cysteines are indicated with red. The ligand 12-bromo-dodecanol, BrC12OH, is shown with the stick model. The graphical representation was created using PyMOL (version 1.3) (DeLanoScientific, San Carlos, CA). The transmembrane regions were predicted using Phobius (Kall et al. 2007) and were marked with H1-H6.

Table S1. *E. affinis* sequencing, assembly, annotation statistics and accession numbers

Bio Projects	i5K Pilot NCBI Bio-project	PRJNA163973 http://www.ncbi.nlm.nih.gov/bioproject/163973
	Eurytemora affinis NCBI Bio-project	PRJNA203087 http://www.ncbi.nlm.nih.gov/bioproject/203087
	NCBI Bio-sample	SAMN02302763 http://www.ncbi.nlm.nih.gov/biosample/02302763
Genome Sequence	180bp insert paired end library	1 Illumina HiSeq 2000 run: 59.4M read pairs, 11.9 Gbp
	500bp insert paired end library	2 Illumina HiSeq 2000 runs: 53.9M read pairs, 10.8 Gbp
	3kb insert mate pair library	1 Illumina HiSeq 2000 run: 54.0M read pairs, 10.8 Gbp
	8kb insert mate pair library	1 Illumina HiSeq 2000 run: 27.0M read pairs, 5.4 Gbp
	180bp insert NCBI SRA Accession	SRX387234 http://www.ncbi.nlm.nih.gov/sra/SRX387234
	500bp insert NCBI SRA Accession	SRX387237 http://www.ncbi.nlm.nih.gov/sra/SRX387237
	3kb insert NCBI SRA Accession	SRX387236 http://www.ncbi.nlm.nih.gov/sra/SRX387236
	8kb insert NCBI SRA Accession	SRX387235 http://www.ncbi.nlm.nih.gov/sra/SRX387235
Genome Assembly	Number of contigs	122,625
	Contig N50	5,738 bp
	Number of scaffolds	6,899
	Scaffold N50	862,645 bp
	Size of final assembly	494,890,867 bp
	Size of final assembly - without gaps	387,574,754 bp
	NCBI Genome Assembly Accession	GCA_000591075.1 http://www.ncbi.nlm.nih.gov/assembly/GCA_000591075.1
RNAseq data	Adult Male RNAseq reads	1 Illumina HiSeq 2000 run: 29.2M read pairs, 5.9 Gbp
	Adult Female RNAseq reads	1 Illumina HiSeq 2000 run: 28.9M read pairs, 5.8 Gbp
	Juvenile RNAseq reads	1 Illumina HiSeq 2000 run: 22.2M read pairs, 4.5 Gbp
	Adult Male RNAseq SRA Accession	SRX903140 http://www.ncbi.nlm.nih.gov/sra/SRX903140
	Adult Female RNAseq SRA Accession	SRX903138 http://www.ncbi.nlm.nih.gov/sra/SRX903138
	Juvenile RNAseq reads SRA Accession	SRX903139 http://www.ncbi.nlm.nih.gov/sra/SRX903139
Automated Genome Annotation (Eaff_0.5.3)	Genes (Eaff_0.5.3)	29,783
	Average Transcript length	1,220 bp
	Average CDS length	821.2 bp (273.3 aa)
	Exons per gene	6.34
	Genome Annotation Link	National Agricultural Library https://i5k.nal.usda.gov/Eurytemora_affinis

Table S2. Next-generation sequencing assemblies and accession numbers for twelve crustacean species used in this study

Order and Family	Species names	Accession number	Sequence Platform
Siphonostomatoida			
Caligidae	*Lepeophtheirus salmonis*	ERX222692, ERX222693, ERX222694, ERX222695, ERX222696, ERX222697, ERX222698, ERX222699, ERX222700, ERX222701, ERX222702, ERX222703, ERX222704, ERX222705, ERX222706, ERX222707, ERX222708, ERX222709, ERX222710, ERX222711, ERX222712, ERX222713, ERX222714, ERX222715	Illumina
Caligidae	*Caligus rogercresseyi*	SRX425308	Illumina
Cyclopoida			
Cyclopidae	*Mesocyclops edax*	SRX246444, SRX246445	454 GS
Lernaeidae	*Lernaea cyprinacea*	SRX423300, SRX433579	Illumina
Harpacticoida			
Harpacticidae	*Tigriopus californicus*	SRX151926, SRX151927, SRX151928, SRX151929	Illumina
Calanoida			
Calanidae	*Calanus sinicus*	SRX214617	454 GS
Calanidae	*Calanus finmarchicus*	SRX456026	Ion Torrent
Acartiidae	*Acartia fossae*	SRX485026	Illumina
Temoridae	*Eurytemora affinis*	see supplementary table S1	
Thecostraca, Sessilia			
Balanidae	*Amphibalanus amphitrite*	SRX120025	Illumina
Eumalacostraca, Decapoda			
Penaeidae	*Penaeus monodon*	SRX110652	Illumina
Phyllopoda, Anostraca			
Artemiidae	*Artemia franciscana*	SRX565005-SRX565014	Illumina

Table S3. Genomes and ESTs of species from basal metazoan, fungal, and protistan phyla used in this study

Species Name	Phylum	Common Name	Source[a]	Version
Nematostella vectensis	Cnidaria	starlet sea anemones	JGI	version 1.0
Hydra magnipapillata	Cnidaria	polyp hydra	NCBI	version 1.0
Acropora millepora	Cnidaria	cnidarian	NCBI	EST (15,413)
Trichoplax adhaerens	Placozoa		JGI	version 1.0
Amphimedon queenslandica	Porifera	sponge	JGI	version 1.0
Mnemiopsis leidyi	Ctenophora	comb jelly	NHGRI	version 2.2
Monosiga brevicollis	Choanozoa	choanoflagellate	JGI	version 1.0
Saccharomyces cerevisiae RM11-1a	Ascomycota	baker's yeast	BI	ASM14936v1
Sporobolomyces roseus	Basidiomycota	red mirror yeast	JGI	version 1.0
Dictyostelium purpureum QSDP1	Mycetozoa	slime mold	JGI	version 1.0
Naegleria gruberi	Percolozoa	amoeboflagellate	JGI	version 1.0
Trichomonas vaginalis G3	Metamonada		JGI	version 1.0
Giardia intestinalis	Metamonada	giardia	GDB	version 1.0

[a]Websites from which the genome sequences were downloaded: JGI (Joint Genome Institute, http://www.jgi.doe.gov), NCBI (the National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov), NHGRI (National Human Genome Research Institute, http://www.genome.gov), and GDB (GiardiaDB; http://giardiadb.org).

Table S4. A summary of the species list used in this study

Phylum	Species Name	Common Name	Sequencing Source
Metamonada	*Giardia intestinalis*	giardia	Genome
	Trichomonas vaginalis G3		Genome
Percolozoa	*Naegleria gruberi*	amoeboflagellate	Genome
Mycetozoa	*Dictyostelium purpureum* QSDP1	slime mold	Genome
Basidiomycota	*Sporobolomyces roseus*	red mirror yeast	Genome
Ascomycota	*Saccharomyces cerevisiae* RM11-1a	baker's yeast	Genome
Choanozoa	*Monosiga brevicollis*	choanoflagellate	Genome
Ctenophora	*Mnemiopsis leidyi*	comb jelly	Genome
Porifera	*Amphimedon queenslandica*	sponge	Genome
Placozoa	*Trichoplax adhaerens*		Genome
Cnidaria	*Nematostella vectensis*	starlet sea anemones	Genome
	Hydra magnipapillata	polyp hydra	Genome
	Acropora millepora	cnidarian	Transcriptome
Onychophora	*Euperipatoides rowelli*	velvet worm	Genome
Chelicerata	*Ixodes scapularis*	black-legged tick	Genome
	Centruroides exilicauda	bark scorpion	Genome
	Latrodectus hesperus	black widow spider	Genome
	Loxosceles reclusa	brown recluse spider	Genome
Myriapoda	*Strigamia maritima*	centipede	Genome
Copepoda	*Lepeophtheirus salmonis*	salmon louse	Genome
	Caligus rogercresseyi	sea louse	Transcriptome
	Mesocyclops edax	freshwater cyclopoid	Genome
	Lernaea cyprinacea	anchor worm	Transcriptome
	Tigriopus californicus	tide pool copepod	Genome and Transcriptome
	Calanus sinicus	Asian Pacific copepod	Transcriptome
	Calanus finmarchicus	North Atlantic copepod	Genome
	Acartia fossae	Oceanic shelf copepod	Transcriptome
	Eurytemora affinis	common estuarine copepod	Genome and Transcriptome
Thecostraca	*Amphibalanus amphitrite*	purple acorn barnacle	Transcriptome
Eumalacostraca	*Hyalella azteca*	amphipod	Genome
	Penaeus monodon	giant tiger prawn	Transcriptome
Phyllopoda	*Artemia franciscana*	brine shrimp	Transcriptome

	Daphnia pulex	water flea	Genome
Hexapoda[a]	*Drosophila melanogaster*	fruit fly	Genome
	Bombyx mori	Silkworm moth	Genome
	Tribolium castaneum	red flour beetle	Genome
	Apis mellifera	honeybee	Genome
	Acyrthosiphon pisum	pea aphid	Genome
	Pediculus humanus	body louse	Genome

[a]Information from these species (highlighted in grey) was taken from the references (see Fig. 1). See further details in supplementary tables S2 and S3.

Table S5. Expression analysis[a] of GR and IR genes for four copepod species and the purple acorn barnacle (*Amphibalanus amphitrite*)

Gene name	Sequence length (bp)	Reads Mapped[b]	RPKM[c]
Caligus rogercresseyi IR8a	847	6	0.41
Caligus rogercresseyi IR21a	801	11	0.80
Caligus rogercresseyi IR25a	2733	116	2.47
Caligus rogercresseyi IR93a	2133	73	1.99
Lernaea cyprinacea GR1	161	2	0.18
Lernaea cyprinacea GR2	271	30	1.64
Lernaea cyprinacea IR8a-1	1369	41	0.44
Lernaea cyprinacea IR8a-2	819	32	0.58
Lernaea cyprinacea IR25a	1707	51	0.44
Lernaea cyprinacea IR76b	407	7	0.26
Tigriopus californicus GR1	1080	0	0.00
Tigriopus californicus GR2	1008	2	0.10
Tigriopus californicus GR3	807	0	0.00
Tigriopus californicus GR4	732	0	0.00
Tigriopus californicus GR5	687	2	0.14
Tigriopus californicus GR6	465	5	0.53
Tigriopus californicus GR7	438	160	17.98
Tigriopus californicus GR8	648	5	0.38
Tigriopus californicus GR9	758	3	0.19
Tigriopus californicus GR10	630	2	0.16
Tigriopus californicus IR8a-1	949	53	2.69
Tigriopus californicus IR8a-2	1455	81	2.68
Tigriopus californicus IR8a-3	808	71	4.24
Tigriopus californicus IR25a	2703	483	8.62
Tigriopus californicus IR93a	1452	210	6.97
Acartia fossae IR25a	963	10	0.60
Acartia fossae IR93a	1524	61	2.33
Amphibalanus amphitrite GR1	240	415	64.51

[a]GR and IR genes for which RPKM values were less than 0.3 were considered to not be expressed and are marked with grey shading (see Materials and Methods).

[b]The number of reads mapped to their genomes or transcriptomes. The total numbers of single-end reads are 26,806,389 in *Artemia franciscana*, 67,324,348 in *Lernaea cyprinacea*, 6,248,088 in *Caligus rogercresseyi*, 20,735,302 in *Tigriopus californicus*, and 17,192,509 in *Acartia fossae*.
[c]Reads per kilobase per million mapped reads.

Table S6. Protein family classification[a] and numbers of transmembrane helices of putative *Trichoplax* (Placozoa) and *Nematostella* (Cnideria) Gustatory Receptors (GRs)

GR genes	NCBI accession number	Sequence length (aa)	CDD[b]	Panther[c]	HHpred[d] (> 100 aa)	Phobius		TMHMM		HMMTOP	
						#[e]	C[f]	#[e]	C[f]	#[e]	C[f]
TadhGrl1	XP_002117254.1	413	No hit	No hit	PF06151 ($1.5e^{-4}$)	4	In	6	In	6	In
TadhGrl2	XP_002110323.1	325	No hit	PTHR21421 ($3.7e^{-23}$)	PF06151 ($2.1e^{-4}$)	4	In	4	IN	5	In
TadhGrl3	XP_002110324.1	267	No hit	PTHR21421 ($2.8e^{-31}$)	PF08395 ($2.2e^{-4}$)	5	In	4	IN	5	In
NvecGrl1	KP294348	456	cl19976 ($1.08e^{-9}$)	PTHR21421 ($2.6e^{-58}$)	PIRSF038981 ($3.3e^{-45}$)	9	Out	6	In	8	Out
NvecGrl2	KP294349	417	cl19976 ($1.21e^{-9}$)	PTHR21421 ($1.5e^{-34}$)	PIRSF038981 ($2.8e^{-70}$)	6	in	6	In	8	Out
NvecGrl3		446	cl19976 ($2.28e^{-7}$)	PTHR21421 ($1.4e^{-33}$)	PF06151 ($1.5e^{-42}$)	6	In	9	Out	9	Out
NvecGrl4		359	cl19976 ($3.88e^{-5}$)	No hit	PIRSF038981 ($1.8e^{-57}$)	4	Out	3	Out	5	Out

[a]Family group description
cl19976: 7tm Chemosensory receptor superfamily
PTHR21421: Gustatory Receptor 5a for Trehalose-related
PF08395: gustatory and odorant receptors
PF06151: Trehalose receptor
PIRSF038981: gustatory receptor protein

[b]RPS-BLAST hit to conserved Domain Database (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (Marchler-Bauer et al. 2015)

[c]The top scoring HMM to PANTHER system (http://www.pantherdb.org) (Mi et al. 2013)

[d]HHsearch, a hidden Markov model (HMM) based remote homolog detection method (http://toolkit.tuebingen.mpg.de/hhpred) (Söding et al. 2005). The sequence coverage of more than 100 amino acids in length was considered to determine the protein family.

[e]Prediction of the number of transmembrane helices

[f]The location of the C-terminus

Table S7. Number of Ionotrophic Glutamate Receptor (iGluR) genes identified in nine copepod[a] assemblies

Gene subfamily[b]	Genes	Lsal	Crog	Meda	Lcyp	Tcal	Csin	Cfin	Afos	Eaff
Antennal	IR25a	0	1	0	1	1	1	0	1	1
IR	IR76b	0	0	0	1	0	0	0	0	1
	IR93a	0	1	0	0	1	2	0	1	2
	IR8a	1	1	0	2	3	0	0	0	1
	IR21a	0	1	0	0	0	0	0	0	1
Divergent IRs		0	0	0	1	2	1	0	0	3
KR	CG3822	3	0	0	4	0	1	0	0	3
	CG5621	0	1	0	1	2	2	1	1	2
	CG11155	0	1	1	3	1	0	0	0	2
	GluR2	1	1	0	1	1	0	0	1	1
	GluRCS1	0	0	0	1	3	1	0	0	0
	GluRCS2	0	0	0	0	0	0	0	0	2
AMPAR	GluR1	0	1	0	1	1	1	0	0	1
NMDAR	NMDAR1	0	1	0	1	0	0	0	0	1
	NMDAR2	0	0	1	6	0	0	0	0	2
Total		5	9	2	22	13	8	1	4	19

[a]Species abbreviations used a four-letter code: Lsal (*Lepeophtheirus salmonis*), Crog (*Caligus rogercresseyi*), Meda (*Mesocyclops edax*), Lcyp (*Lernaea cyprinacea*), Tcal (*Tigriopus californicus*), Csin (*Calanus sinicus*), Cfin (*Calanus finmarchicus*), Afos (*Acartia fossae*), and Eaff (*Eurytemora affinis*).

[b]iGluR gene candidates are divided into four major categories: IR (ionotrophic receptor, including antennal and divergent IRs), NMDAR (*N*-methyl-D-aspartate receptor), AMPAR (α-amino-3-hydroxyl-5-methyl-4-isoxazole receptor), and KR (Kaninate receptor).

Table S8. Sex-specific expression[a] of the copepod *E. affinis* GR, IR, and CSP genes identified in this study

Genes	Sequence length (bp)	Male					Female							
		Sample 1[b] (181,352,803)		Sample 2[b] (195,268,926)		S.D.[d]	Sample 3[b] (172,614,735)		Sample 4[b] (190,634,532)		S.D.[d]	FC[e,f]	*P*-values[f]	Prob.[g]
		read[c]	RPKM	read[c]	RPKM		read[c]	RPKM	read[c]	RPKM				
GRs														
EaffGR1	654	19	0.16	39	0.31	0.11	34	0.30	28	0.22	0.06	0.17	0.7041	0.17
EaffGR2	957	37	0.21	78	0.42	0.15	25	0.15	43	0.24	0.06	-0.67	0.2363	0.44
EaffGR3	969	11	0.06	91	0.48	0.30	80	0.48	66	0.36	0.08	0.84	0.2422	0.44
EaffGR4	957	19	0.11	25	0.13	0.01	17	0.10	12	0.07	0.02	-0.64	0.4944	0.33
EaffGR5	954	20	0.12	23	0.12	0	17	0.10	17	0.09	0.01	-0.32	0.4018	0.22
EaffGR6	963	0	0.00	0	0.00	0	0	0.00	0	0.00	0	-1.28	1	NA
EaffGR7	987	0	0.00	4	0.02	0.01	13	0.08	1	0.01	0.05	2.08	0.4082	0.44
EaffGR8I	552	18	0.18	11	0.10	0.06	0	0.00	9	0.09	0.06	-1.86	0.1839	0.44
EaffGR9I	666	0	0.00	10	0.08	0.06	11	0.10	7	0.06	0.03	1.22	0.4351	0.39
EaffGR10	819	9	0.06	10	0.06	0	10	0.07	11	0.07	0	0.06	1	0.05
Antennal IRs														
EaffIR8a	1560	1108	3.92	1559	5.12	0.85	170	0.63	223	0.75	0.08	-2.80	**<0.0001**	**0.87**
EaffIR21a	1270	266	1.15	263	1.06	0.06	96	0.44	114	0.47	0.02	-1.35	0.0907	0.50
EaffIR25a	2724	17036	34.49	13221	24.86	6.81	3286	6.99	2961	5.70	0.91	-2.31	**0.0001**	**0.82**
EaffIR76b	498	3	0.03	52	0.53	0.35	35	0.41	23	0.24	0.12	-0.13	0.9458	0.00
EaffIR93a-1	2733	7707	15.55	5871	11.00	3.22	2043	4.33	1899	3.64	0.49	-1.81	**0.0009**	**0.79**

EaffIR93a-2	392	84	1.18	63	0.82	0.25	20	0.30	35	0.47	0.12	-1.37	0.2307	0.37
Divergent IRs														
EaffIRCS2-1	773	49	0.35	48	0.32	0.02	13	0.10	33	0.22	0.08	-1.01	0.4881	0.29
EaffIRCS2-2	831	128	0.85	92	0.57	0.20	20	0.14	46	0.29	0.11	-1.66	0.0764	0.47
EaffIRCS2-3	954	168	0.97	381	2.05	0.76	187	1.14	296	1.63	0.35	-0.23	0.7794	0.16
CSPs														
EaffCSP1	318	1195	20.72	200	3.22	12.37	24785	451.53	16025	264.34	132.4	4.89	**<0.0001**	**0.97**
EaffCSP2	354	66	1.03	73	1.06	0.02	0	0.00	23	0.34	0.24	-2.42	0.1643	0.53
EaffCSP3	343	146	2.35	239	3.57	0.86	45	0.76	187	2.86	1.48	-0.67	0.5097	0.34
EaffCSP4	396	805	11.21	1058	13.68	1.75	973	14.23	687	9.10	3.63	-0.28	0.4495	0.16

[a]GR, IR and CSP genes for which RPKM values are less than 0.3 are considered to not be expressed and are marked with grey shading (see Materials and Methods).

[b]Total numbers of single-end reads before filtering are shown in parentheses.

[c]Number of single-end reads mapped to their genomes. Note that although the reads were not filtered, we used 0 mismatch to map the reads to the assembled transcriptome. Thus, reads that included any ambiguity, including unknown nucleotide 'N's, were not counted.

[d]S.D. indicates the standard deviation for the two biological replicates.

[e]FC (Fold change) refers to log2 ratio between female *versus* male samples. For example, a positive value indicates greater gene expression in female samples, whereas a negative value indicates greater expression in male samples.

[f]FC and *P*-values (shown in boldface when $P < 0.001$) were taken from an edgeR (ver.3.8.5) analysis (Zhou et al. 2014).

[g]Probability values (shown in boldface when Prob. > 0.70) were taken from a NOISeq analysis (Tarazona et al. 2011).

Table S9. RPKM expression levels for five representative housekeeping genes of *E. affinis* in males and females, two replicates each (~220 individual copepods per replicate)

Selected housekeeping genes	Male			Female			MFC[b]
	Sample1	Sample2	S.D.[a]	Sample1	Sample2	S.D.[a]	
EaffCyclophilin33	29.1	25.9	2.26	52.4	48.9	2.47	2.02
Eaff42A	269.7	249.1	15.56	192.6	213.7	38.89	1.40
EaffHsp83	330.1	308.1	14.57	309.5	364.5	14.92	1.18
EaffGapdh1	594.0	571.1	16.19	381.1	369.8	7.99	1.60
EaffRPL32	658.2	795.2	96.87	598.3	861.8	186.3	1.44

[a]S.D. indicates the standard deviation for the two biological replicates.

[b]MFC represents 소

e maximum RPKM value divided by the minimum RPKM value.

Table S10. Number of chemosensory proteins (CSPs) identified in six copepod assemblies

Species	Number of CSP genes[a]
Lepeophtheirus salmonis[b]	[1]
Caligus rogercresseyi[c]	1 [1]
Lernaea cyprinacea[c]	1 [1]
Tigriopus californicus[b,c]	1
Acartia fossae[c]	4
Eurytemora affinis[b,c]	3 [1]

[a]CSP gene candidates are divided into two categories: intact and incomplete. The first number shown is that of "intact" genes, which contain full-length open reading frames. If the sequences are more than 110 amino acids (aa) in length, the genes are considered intact, given that the average length of copepod CSPs is 112.5 amino acids. The number of "incomplete" genes are given in square brackets. The "incomplete" genes are less than 110 aa.

[b,c]The source(s) is (are) from genomic sequences and from transcriptomic sequences (Table 1 and supplementary table S2).

Table S11. Numbers of predicted α-helical structures of copepod CSPs

CSP genes	Sequence length (aa)	Numbers of predicted α-helical structures[a]
CrogCSP1	106	6
CrogCSP2	135	6
LsalCSP	76	4
TcalCSP	147	6
LcypCSP1	105	6
LcypCSP2	113	6
AfosCSP1	105	6
AfosCSP2	105	6
AfosCSP3	111	6
AfosCSP4	111	6
EaffCSP1	105	6
EaffCSP2	117	6
EaffCSP3	109	6
EaffCSP4	131	6

[a]The sequence-based secondary structure were predicted using Jpred 4 (Drozdetskiy et al. 2015).

SUPPLEMENTAL FILES

Supplementary File S1. GR protein sequences in FASTA format. Six are from *Eurytemora affinis*, two from *Tigriopus californicus*, two from *Lernaea cyprinacea*, and 1 from *Amphibalanus amphitrite*.

Supplementary File S2. Copepod IR protein sequences in FASTA format. These IR sequences and alignments are also available at:
the Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.ts747.

Supplementary File S3. Copepod CSP protein sequences in FASTA format. These CSP sequences and alignments are also available at
the Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.ts747.

SUPPLEMENTAL MOVIES

Supplementary Movie S1. Male Copepod Tracking of a Female Scent Trail.

https://www.youtube.com/watch?v=OQSqN3IGsuQ

Supplementary Movie S2. Male copepod follows another female scented trail.

https://www.youtube.com/watch?v=ALOhBxtmxHE

These movies are also available at:

http://bioinformatics.unl.edu/eyun/Copepoda_CRG.

SUPPLEMENTAL REFERENCES

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 22:195-201.

Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. 2009. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell*. 136:149-162.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res*. 14:1188-1190.

Drozdetskiy A, Cole C, Procter J, Barton GJ. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*. 43:W389-W394.

Ferre F, Clote P. 2006. DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res*. 34:W182-W185.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. *Nucleic Acids Res*. 42:D222-D230.

Forêt S, Wanner KW, Maleszka R. 2007. Chemosensory proteins in the honey bee: insights from the annotated genome, comparative analyses and expressional profiling. *Insect Biochem Mol Biol*. 37:19-28.

Kall L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res*. 35:W429-432.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1. *J Mol Biol*. 305:567-580.

Lin HH, Hsu JC, Hsu YN, Pan RH, Chen YF, Tseng LY. 2013. Disulfide connectivity prediction based on structural information without a prior knowledge of the bonding state of cysteines. *Comput Biol Med*. 43:1941-1948.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, et al. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 43:D222-D226.

Mi H, Muruganujan A, Thomas PD. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 41:D377-D386.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth*. 8:785-786.

Pirovano W, Feenstra KA, Heringa J. 2008. PRALINE™: a strategy for improved multiple alignment of transmembrane proteins. *Bioinformatics*. 24:492-497.

Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 33:W244-W248.

Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A. 2011. Differential expression in RNA-seq: a matter of depth. *Genome Res*. 21:2213–2223.

Zhou X, Lindsay H, Robinson MD. 2014. Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Res*. 42:e91.