GENETICS, 2023, 223(4), iyad026
https://doi.org/10.1093/genetics/iyad026
Advance Access Publication Date: 18 February 2023
Neurogenetics

Transcriptomic profiling of sex-specific olfactory neurons reveals subset-specific receptor expression in *Caenorhabditis elegans*

Douglas K. Reilly,[1,6†] Erich M. Schwarz,[2,†] Caroline S. Muirhead,[1] Annalise N. Robidoux,[1,3] Anusha Narayan,[4,7] Meenakshi K. Doma,[4,8] Paul W. Sternberg,[4] Jagan Srinivasan[1,5,*]

[1]Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, MA 01605, USA
[2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[3]Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, MA 01605, USA
[4]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[5]Bioinformatics and Computational Biology Program, Worcester Polytechnic Institute, Worcester, MA 01605, USA
[6]Present address: Biology Department, Tufts University, Medford, MA 02155, USA
[7]Present address: Merck and Co., 2000 Galloping Hill Rd, Kenilworth, NJ 07033, USA
[8]Present address: Individualized Cell and Gene Therapies, Genentech, South San Francisco, CA 94080, USA

*Corresponding author: Email: jsrinivasan@wpi.edu
[†]These authors contributed equally to this work.

Abstract

The nematode *Caenorhabditis elegans* utilizes chemosensation to navigate an ever-changing environment for its survival. A class of secreted small-molecule pheromones, termed ascarosides, play an important role in olfactory perception by affecting biological functions ranging from development to behavior. The ascaroside #8 (ascr#8) mediates sex-specific behaviors, driving avoidance in hermaphrodites and attraction in males. Males sense ascr#8 via the ciliated male-specific cephalic sensory (CEM) neurons, which exhibit radial symmetry along dorsal–ventral and left–right axes. Calcium imaging studies suggest a complex neural coding mechanism that translates stochastic physiological responses in these neurons to reliable behavioral outputs. To test the hypothesis that neurophysiological complexity arises from differential expression of genes, we performed cell-specific transcriptomic profiling; this revealed between 18 and 62 genes with at least twofold higher expression in a specific CEM neuron subtype vs both other CEM neurons and adult males. These included two G protein–coupled receptor (GPCR) genes, *srw-97* and *dmsr-12*, that were specifically expressed in nonoverlapping subsets of CEM neurons and whose expression was confirmed by GFP reporter analysis. Single CRISPR-Cas9 knockouts of either *srw-97* or *dmsr-12* resulted in partial defects, while a double knockout of both *srw-97* and *dmsr-12* completely abolished the attractive response to ascr#8. Together, our results suggest that the evolutionarily distinct GPCRs SRW-97 and DMSR-12 act nonredundantly in discrete olfactory neurons to facilitate male-specific sensation of ascr#8.

Keywords: *Caenorhabditis elegans*, pheromone, sex-specific, G protein–coupled receptors, transcriptomes, single-cell

Introduction

The ability of an organism to find a mate is critical to the survival of a species. Many species utilize small-molecule pheromones to signal mate location (Pungaliya *et al.* 2009; Narayan *et al.* 2016), sexual maturity (Aprison and Ruvinsky 2015, 2017), and receptivity (Houck *et al.* 2007; Jang *et al.* 2017), which are sensed and processed via the nervous system, driving behavioral and developmental responses.

The nematode *Caenorhabditis elegans* communicates with conspecifics almost exclusively using pheromones called ascarosides (Ludewig *et al.* 2019; McGrath and Ruvinsky 2019), a structurally conserved, modular class of small-molecule pheromones (von Reuss and Schroeder 2015; Zhang *et al.* 2017) composed of a core ascarylose sugar and a fatty-acid–derived side chain (Butcher *et al.* 2007; von Reuss *et al.* 2012; Ludewig *et al.* 2019). Ascarosides signal a host of environmental and developmental information, including the sexual maturity and location of potential mates (Narayan *et al.* 2016; Aprison and Ruvinsky 2017), and this information is often conserved across nematodes (Choe *et al.* 2012; Ragsdale *et al.* 2013; Dong *et al.* 2016, 2018; Reilly *et al.* 2019).

While most male-specific neurons are located in the tail, the major contributor to male-specific chemosensory-driven behaviors is the ciliated male-specific cephalic sensory (CEM) neurons, part of the cephalic sensilla. Previous studies have shown these four radially symmetric neurons to be involved in sensing ascarosides #3 and #8 (ascr#3 and ascr#8) and two pheromones that serve to attract males to mates (Pungaliya *et al.* 2009; Narayan *et al.* 2016; Reilly *et al.* 2017). Ascr#3 also plays a role in dauer formation, functioning alongside ascr#2 and ascr#4 as the "dauer pheromone" (Butcher *et al.* 2008). Ascr#8 is unique in that it contains a *p*-aminobenzoic acid moiety on its terminus (Pungaliya *et al.* 2009; Artyukhin *et al.* 2018). In animals lacking CEM, the

Received: December 22, 2022. **Accepted:** February 13, 2023

attractive response to ascr#8 is abolished, while further ablation of the ASK is required for complete loss of the ascr#3 behavioral response (Narayan *et al.* 2016). Ablating three of the four CEM neurons, while leaving one CEM intact, revealed that the four CEM neurons cooperate to drive a tuned attractive response to an intermediate (1 μM) concentration of both ascr#3 and ascr#8 (Narayan *et al.* 2016).

Calcium imaging and electrophysiological experiments demonstrated that CEM neurons show variable responses to stimuli not only between different animals but also within a single animal (Narayan *et al.* 2016; Reilly *et al.* 2017). While neurons such as AWC exhibit variable calcium responses between different animals, such responses are consistent between the left and right AWC neurons of a single animal (Cochella *et al.* 2014). In contrast, the responses of individual CEM neurons are stochastic within single animals, and yet four CEM neurons within one animal consistently generate proper behavioral responses (Narayan *et al.* 2016). To understand this pattern of stochastic neuronal activity yielding consistent behavioral outputs, it is imperative to uncover genes encoding components of the CEM response.

A recent study in *C. elegans* hermaphrodites generated transcriptomic landscapes of 118 neuronal classes from 302 neurons in order to link functional and anatomical properties of individual neurons with their molecular identities (Taylor *et al.* 2021). Discrete neuronal classes were successfully identified via their combinations of expressed neuropeptides and neuropeptide receptor genes. However, a similar feat has yet to be performed on male *C. elegans*, and the transcriptomic profiles of individual male neurons remain enigmatic.

In a more focused transcriptomic approach, gene expression profiles of extracellular vesicle-releasing neurons (EVNs) were identified (Wang *et al.* 2015; Kaletsky *et al.* 2016). The CEM neurons are a subset of the 27 EVN neurons present in the male; therefore, the CEM transcriptomic data are embedded in these data, since there were thousands of EVNs in each replicate diluting CEM-specific expression values.

To begin understanding how CEM neurons achieve stochastic yet reliable physiological responses to ascr#8, we performed single-cell RNA-seq (Schwarz *et al.* 2012) on CEM neurons and uncovered a small number of highly enriched genes encoding G protein–coupled receptors (GPCRs). Given that all ascaroside receptors identified to date have been GPCRs (Kim *et al.* 2009; McGrath *et al.* 2011; Park *et al.* 2012; Greene, Dobosiewicz, *et al.* 2016; Greene, Brown, *et al.* 2016; Chute *et al.* 2019), we tested whether any of these enriched genes contribute to male *C. elegans* sensation of ascr#8. We identified two distantly related GPCR genes that contribute to the ascr#8 behavioral response, *srw-97* and *dmsr-12*. Loss of each receptor resulted in partially deficient behavioral responses to ascr#8. Following generation of a double *srw-97;dmsr-12* mutant, we found that loss of both receptors results in a complete loss of behavioral response to ascr#8. Phylogenetic analysis further indicates that both receptors are homologous to closely related receptors present throughout the *Caenorhabditis* genus, and robust ascr#8 responses may be a trait recently evolved in *C. elegans*.

Methods

Single-cell isolation and RT-PCR

Individual CEM neurons were isolated from adult males of the *C. elegans* strain CU607 (smIs23 [*ppkd-2*::GFP + pBX]; *him-5*(e1490)), as previously described (Narayan *et al.* 2016). This strain expresses an integrated GFP transgene that labels male-specific EVNs (*ppkd-2*::GFP [Fig. 1a]). Cells were separated by anatomical identity [i.e. CEM dorsal left (CEM_DL), CEM dorsal right (CEM_DR), CEM ventral left (CEM_VL), and CEM ventral right (CEM_VR)]. Microdissection and single-cell RT-PCR of individual CEM_DL, CEM_DR, CEM_VL, and CEM_DR neurons were performed essentially as described (Schwarz *et al.* 2012). For all four neuronal subtypes, single-end 50-nucleotide (50-nt) to 100-nt RNA-seq was performed on an Illumina HiSeq 2000. To identify so-called housekeeping genes and genes primarily active outside the nervous system, we compared our results from CEM_DL, CEM_DR, CEM_VL, and CEM_DR to equivalent reanalyses of published single-end 38-nt RNA-seq data from mixed-stage whole *C. elegans* hermaphrodite larvae (Schwarz *et al.* 2012) and of published single-end 36-nt RNA-seq data from adult whole *C. elegans* males and females (Thomas *et al.* 2012).

Transcriptional analysis

Reads from RT-PCR–amplified CEM neurons and larvae were quality-filtered as follows: neuronal reads that failed chastity filtering were discarded (chastity filtering had not been available for the larval reads); raw 38-nt larval reads were trimmed 1–37 nt; all reads were trimmed to remove any indeterminate ("N") residues or residues with a quality score of <3; and larval reads that had been trimmed below 37 nt were deleted, as were neuronal reads that had been trimmed below 50 nt. For uniformity of analysis, all RNA-seq reads of 51–100 nt in length were trimmed to 50 nt before further RNA-seq analysis. This left a total ranging from 55,902,297 (CEM_VL) to 73,412,777 (CEM_VR) filtered reads for analysis of each neuronal subtype, vs 23,369,056 filtered reads for whole larvae (Supplementary Table 2). Previously published RNA-seq reads (Thomas *et al.* 2012) from three biological replicates of adult males (accession numbers SRR580386, SRR580387, and SRR580388) and adult females (accession numbers SRR580383, SRR580384, and SRR580385) were downloaded from the European Nucleotide Archive (ftp://ftp.sra.ebi.ac.uk) and used without further trimming or filtering.

We used RSEM version 1.2.17 (Li and Dewey 2011) with Bowtie2 version 2.2.3 (Langmead and Salzberg 2012) and SAMTools version 1.0 (Li *et al.* 2009) to map filtered reads to a *C. elegans* gene index and generate read counts (Supplementary Table 1) and gene expression levels in transcripts per million (TPM). To create the *C. elegans* gene index, we ran RSEM's *rsem-prepare-reference* with the arguments "*--bowtie2 --transcript-to-gene-map*" upon a collection of coding DNA sequences (CDSes) from both protein-coding and nonprotein-coding *C. elegans* genes in WormBase release WS245 (Howe *et al.* 2016). The CDSes were obtained from ftp://ftp.sanger.ac.uk/pub2/wormbase/releases/WS245/species/c_elegans/PRJNA13758/c_elegans.PRJNA13758.WS245.mRNA_transcripts.fa.gz and ftp://ftp.sanger.ac.uk/pub2/wormbase/releases/WS245/species/c_elegans/PRJNA13758/c_elegans.PRJNA13758.WS245.ncRNA_transcripts.fa.gz. For each RNA-seq data set of interest, we computed mapped reads and expression levels per gene by running RSEM's *rsem-calculate-expression* with the arguments "*--bowtie2 -p 8 --no-bam-output --calc-pme --calc-ci --ci-credibility-level 0.99 --fragment-length-mean 200 --fragment-length-sd 20 --estimate-rspd --ci-memory 30000*." These arguments, in particular "*--estimate-rspd*," were aimed at dealing with single-end data from 3′-biased RT-PCR reactions; the argument "*--phred64-quals*" was used for the larval reads, while "*--phred33-quals*" was used for all other reads (neuronal, adult males, and adult females). We computed posterior mean estimates (PMEs) both for read counts and for gene expression levels and rounded PMEs of read counts down to the nearest lesser integer. We also computed

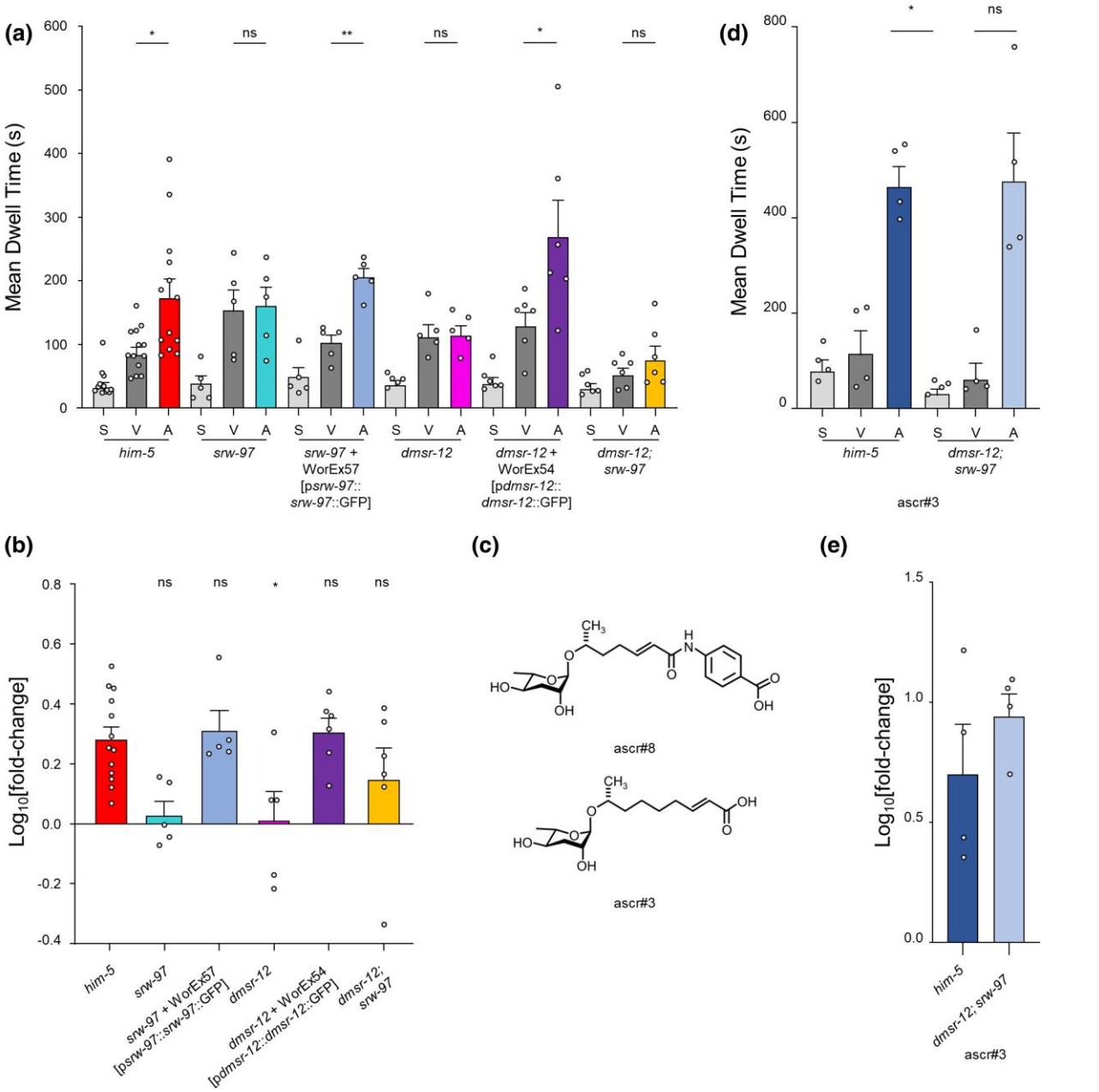

Fig. 1. The GPCRs, SRW-97 and DMSR-12, act in nonoverlapping CEM neurons to aid in driving the ascr#8 response. a) Raw dwell times in the centers of spatial control (S), vehicle-only control (V), and ascaroside a) wells and b) Log(fold-change) of A vs V for *srw-97* knockout and transgenic rescue (WorEx57), *dmsr-12* and transgenic rescue (WorEx54), as well as double *srw-97;dmsr-12* mutant animals. c) Structural comparison of ascr#8 (left) and ascr#3 (right). d) Raw dwell times and e) log(fold-change) of *him-5* and *dmsr-12;srw-97* animals in response to the structurally related attracted pheromone, ascr#3. Error bars denote SEM. $n \geq 5$. a, d) RM-ANOVA followed by Dunnett's post hoc test or Friedman test followed by Dunn's (within strain), b) ANOVA followed by Bonferroni's post hoc test [between strain log(fold-change) values]. e) Student's *t*-test between strain log(fold-change) values $^{*}P < 0.05$, $^{**}P < 0.01$.

99% credibility intervals (Cis) for expression data, so that we could use the minimum value in the 99% Ci for TPM as a robust minimum estimate of a gene's expression (minTPM).

We observed the following overall alignment rates of reads to the WS245 *C. elegans* gene index: 47.16% for the CEM_DL read set, 27.37% for the CEM_DR read set, 40.20% for the CEM_VL read set, 17.99% for the CEM_VR read set, and 76.41% for the larval read set (Supplementary Table 2). A similar discrepancy between lower alignment rates for hand-dissected linker cell RNA-seq reads vs higher alignment rates for whole larval RNA-seq reads was previously observed and found to be due to a much higher rate of human contaminant RNA-seq in the hand-dissected linker cells (Schwarz *et al.* 2012). For previously published adult male and female reads, we observed mapping rates of 88.20–96.11%; this may reflect greater difficulty of getting mappable reads from our RT-PCR-amplified cDNAs than from unamplified cDNA syntheses.

We defined detectable expression for a gene in a given RNA-seq data set by that gene having an expression level of 0.1 TPM; we defined robust expression by that gene having a minimum estimated expression level (termed minTPM) of at least 0.1 TPM in a Ci of 99% (i.e. ≥0.1 minTPM). The numbers of genes being scored as expressed in a given neuronal subtype above background levels, for various data sets, are given in Supplementary Table 3. Other results from RSEM analysis are given in Supplementary Table 5.

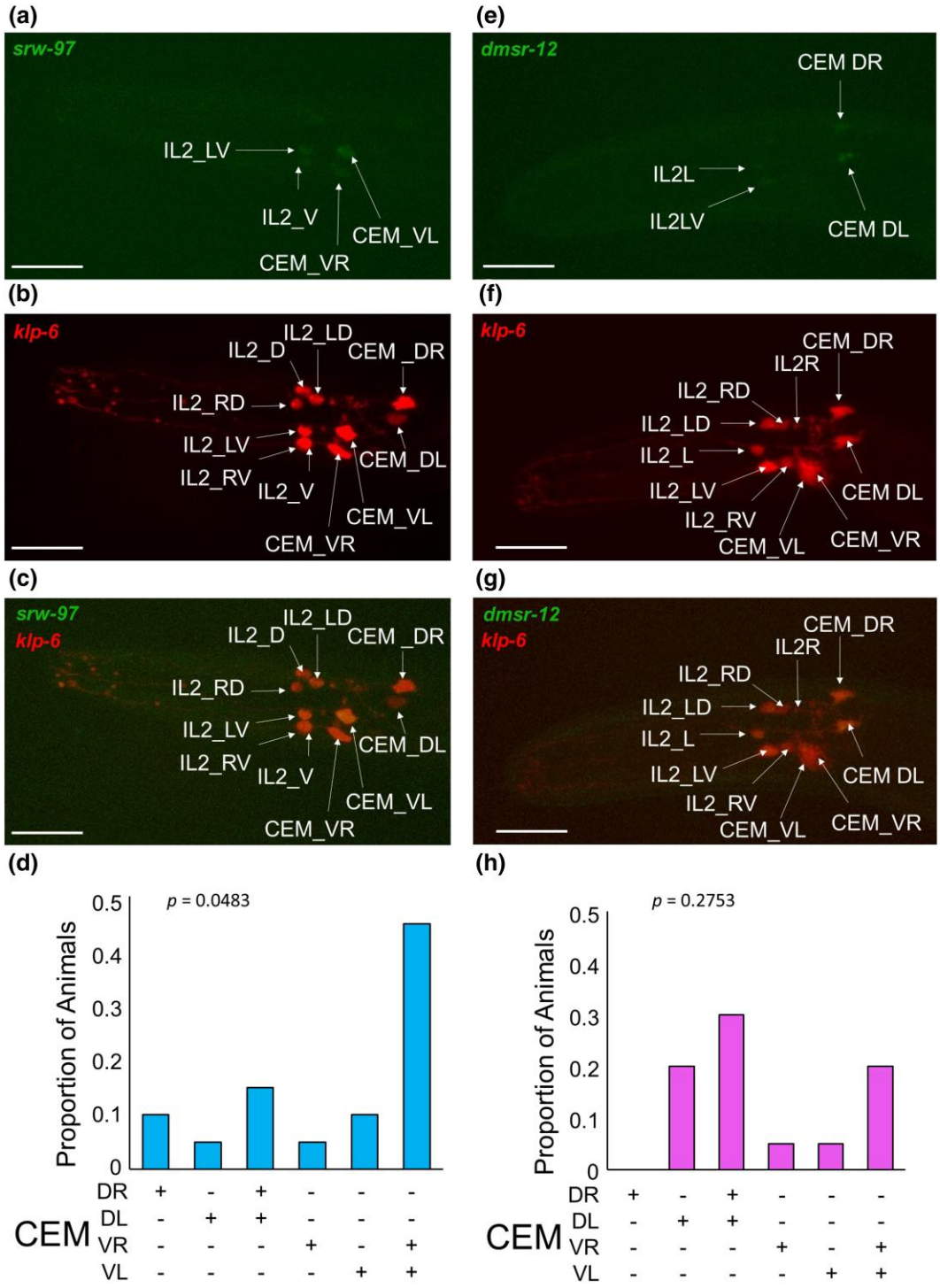

Fig. 2. Expression of *srw-97* and *dmsr-12* within CEM neurons. a–c) Expression of WorEx61 (p*srw-97*::*srw-97*::GFP) in myIs20 (p*klp-6*::tdTomato), which expresses in b, f) EVN neurons. The receptor fusion is observed in a, c) CEM_VR, CEM_VL, IL_LV, and IL_V. e–g) Expression of WorEx60 (p*dmsr-12*::*dmsr-12*::GFP) in myIs20. The receptor fusion is observed in e, g) CEM_DR, CEM_DL, IL_L, and IL2_LV. d, h) Quantification of expression within CEM of d) *srw-97* and h) *dmsr-12*. *P* values are the results of a chi-squared analysis testing the null hypothesis of equal expression between different combinations of CEM subtypes. Images captured at ×63 using an oil lens. Scale bars denote 22 μm.

We annotated *C. elegans* genes and the encoded gene products in several ways (Supplementary Table 5). For the products of protein-coding genes, we predicted classical signal and transmembrane sequences with Phobius 1.01 (Käll *et al.* 2004), regions of low sequence complexity with pseg (SEG for proteins, from ftp://ftp.ncbi.nlm.nih.gov/pub/seg/pseg) (Wootton 1994), and coiled-coil domains with ncoils (from http://www.russell.embl-heidelberg.de/coils/coils.tar.gz) (Lupas 1996). Protein domains from Pfam 31.0 (Finn *et al.* 2016) were detected with HMMER 3.1b2/*hmmsearch* (Eddy 2009), using the argument "*--cut_ga*" to invoke model-specific reporting thresholds. Memberships of genes in orthology groups from eggNOG 3.0 (Powell *et al.* 2012) were

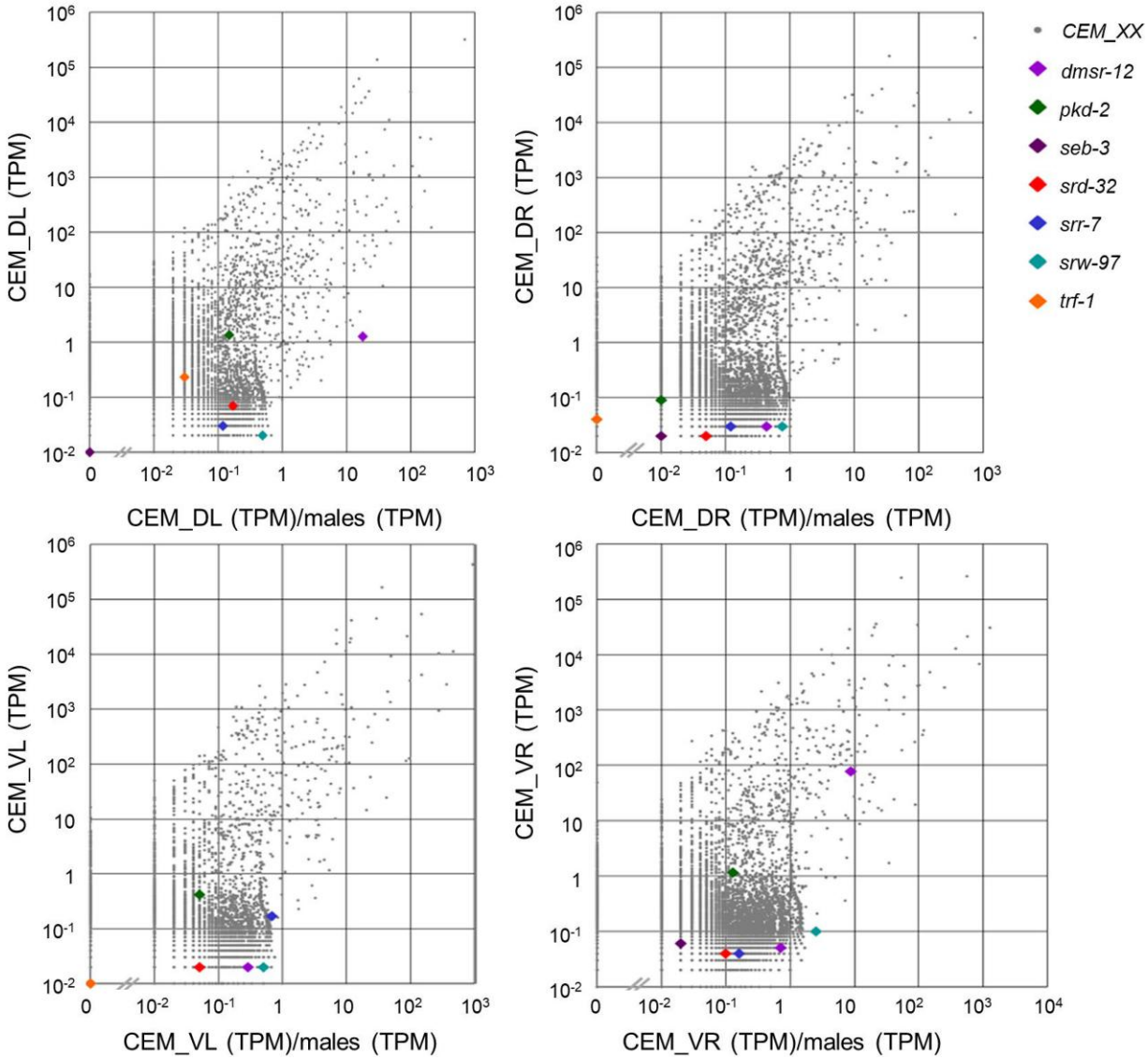

Fig. 3. Transcriptomic landscapes of the CEM neurons. a–d) Expression plots of genes expressed in individual a) CEM DL, b) CEM DR, c) CEM VL, and d) CEM VR neurons. For each gene, its X-axis position displays the ratio of its expression in an individual CEM neuron divided by its highest expression in three replicates of adult *C. elegans* males; its Y-axis position displays expression in an individual CEM neuron subtype. Expression levels are measured in TPM. Genes of interest are denoted by colored symbols, defined in the legend.

extracted from WormBase WS245 with the TableMaker function of ACEDB 4.3.39. Genes with likely housekeeping status (based on ubiquitous expression in both larvae and linker cells) were as identified in our previous work (Schwarz *et al.* 2012). Genes were predicted to encode GPCRs on the basis of their encoding a product containing one or more of the following Pfam-A protein domains: 7tm_1 [PF00001.16], 7tm_2 [PF00002.19], 7tm_3 [PF00003.17], 7tm_7 [PF08395.7], 7TM_GPCR_Srab [PF10292.4], 7TM_GPCR_Sra [PF02117.11], 7TM_GPCR_Srbc [PF10316.4], 7TM_GPCR_Srb [PF02175.11], 7TM_GPCR_Srd [PF10317.4], 7TM_GPCR_Srh [PF10318.4], 7TM_GPCR_Sri [PF10327.4], 7TM_GPCR_Srj [PF10319.4], 7TM_GPCR_Srsx [PF10320.4], 7TM_GPCR_Srt [PF10321.4], 7TM_GPCR_Sru [PF10322.4], 7TM_GPCR_Srv [PF10323.4], 7TM_GPCR_Srw [PF10324.4], 7TM_GPCR_Srx [PF10328.4], 7TM_GPCR_Srz [PF10325.4], 7TM_GPCR_Str [PF10326.4], ABA_GPCR [PF12430.3], Sre [PF03125.13], and Srg [PF02118.16]. By this criterion, we identified 1,615 genes encoding GPCRs in the WS245 version of the *C. elegans* genome; this resembles a previous estimate of ~1,470 *C. elegans* genes encoding chemoreceptors and other GPCRs, identified through extensive computational and manual analysis (Hobert 2013). The memberships of genes in orthology groups from eggNOG 3.0 (Powell *et al.* 2012) were extracted directly from WormBase WS245 with the TableMaker function of ACEDB 4.3.39. Genes with likely housekeeping status (based on ubiquitous expression in both larvae and linker cells) were as identified in our previous work (Schwarz *et al.* 2012). Gene Ontology (GO) annotations for *C. elegans* genes were extracted from WormBase-computed annotations in ftp://ftp.wormbase.org/pub/wormbase/releases/WS245/ONTOLOGY/gene_association.WS245.wb.c_elegans; human-readable text descriptions for GO term IDs were extracted from *term.txt* in the GO archive *http://archive.geneontology.org/full/2014-07-01/go_201407-termdb-tables.tar.gz*.

GFP reporter construction

Reporter fusion constructs were generated using previously described techniques (Boulin *et al.* 2006). Approximately 2–3 kb of

upstream promoter region of each gene was included in construct generation, as well as a portion of the coding sequence (Supplementary Table 6). This was then fused to GFP (from the Fire Vector Kit plasmid, pPD95.75), via PCR fusion (Boulin *et al.* 2006). Primers were designed using Primer 3 and ordered from IDT (Integrated DNA Technologies). Primer sequences are available in Supplementary Table 9. Successful fusion was confirmed via gel electrophoresis prior to injection.

Reporter fusion constructs were injected into the gonads of *pha-1(e2123ts);lite-1(ce314);him-5(e1490)* animals, along with a co-injection marker of pBX(*pha-1*(+)). In this manner, positive array animals will propagate normally at 20°C. Strains were confirmed via GFP expression, with multiple array lines being generated per injection (Supplementary Table 8). Injections were performed by InVivo Biosystems, with strain isolation being performed in house.

Imaging

Animals were imaged for GFP expression using previously described techniques. GFP + young adult male animals were mounted on a 1% agarose pad and immobilized with sodium azide. Animals were then imaged on a spinning disk confocal microscope at ×63 magnification. Z-stack imaging was performed, generating 3D reconstructions of the heads of the imaged animals. Central/optimal z-plane images were used to generate the images used to verify expression (Fig. 2; Supplementary Figs. 1 and 2).

RNAi feeding

VH624 (*rhIs13* [*unc-119*::GFP + *dpy-20*(+)]; *nre-1(hd20);lin-15B(hd126)*) animals (Schmitz *et al.* 2007; Poole *et al.* 2011) were crossed with *him-5(e1490)* animals to integrate male production into a strain hypersensitive to neuronal RNAi knockdown, generating JSR44 (*nre-1(hd20);lin-15B(hd126);him-5(e1490)*). During the cross, insertion of the *him-5(e1490)* allele displaced the integrated array *rhIs13*, suggesting location of the array on Chromosome V. Presence of *lin-15B(hd126)* in JSR44 was confirmed via sequencing. The non-annotated *nre-1(hd20)* is linked with *lin-15B*, being retained alongside *lin-15B* (Schmitz *et al.* 2007; Poole *et al.* 2011).

RNAi clones were grown overnight in cultures of LB containing 50 µg/mL ampicillin. Cultures were then diluted to an OD_{600} of 1.0 before plating on NGM agar plates containing 50 µg/mL ampicillin and 1 mM IPTG (isopropyl-β-D-thiogalactoside) to select for RNAi clones and induce expression. Lawns were allowed to grow at room temperature for 8–16 h, before JSR44 young adult hermaphrodites were placed on the plates and left to propagate at 16°C. Young adult males of the F1 progeny were then selected for behavioral testing suing the *spot retention assay*. Empty vector controls (VC-1 clone) were run alongside every targeted knockdown experiment.

Spot retention assay

Following previously described methods, young adult males were isolated from hermaphrodites 5–16 h prior to testing (Pungaliya *et al.* 2009; Narayan *et al.* 2016). In short, at the time of the assay, 0.6 µL of either vehicle control (–) or ascaroside #8 (+) was added to the NGM plates covered in a thin lawn of OP50 *Escherichia coli*. Ten males were then divided between two pre-marked spots on the agar, equidistant from the cues. The plate was then recorded for 20 min. The time spent of each visit in either vehicle or ascaroside #8 (if >10 s) was scored and averaged. Plates in which the average was >2 standard deviations removed from the population average were removed from the final analysis as outliers. To compare between strains or conditions, the vehicle was subtracted from the ascaroside dwell time for each plate. The average of these differences was then compared statistically.

Single worm behavioral assay

Following previously described methods (Pungaliya *et al.* 2009; Narayan *et al.* 2016; Reilly *et al.* 2021), animals were isolated and prepared in an identical manner to the spot retention assay. The two outside rings of wells in a 48-well tissue culture plate were filled with 200 µL of NGM agar, which was then seeded with 65 µL of OP50 *E. coli*. The plates and lawns were then dried at 37°C for 4 h. Alternating wells were then prepared as spatial controls ("S," nothing done), vehicle controls ("V," 0.85 µL of dH_2O was placed in the center of the well), or ascaroside well ("A," 0.85 µL of ascaroside was placed in the center of the well). This was performed over four quadrants. Animals were scored for their visits and duration to the center of the well and/or the cue.

The average duration of each worm's visits was calculated, and these values were again averaged together to generate a mean dwell time in seconds for each plate. When comparing across strains or conditions, the spatial controls were then compared for statistical difference. If none was observed, the log(fold-change) A/V was then calculated by taking the log of the ascaroside mean dwell time divided by the vehicle mean dwell time for each plate. The number of times each worm visited the center was averaged to generate the visit counts.

The percent attraction values were calculated by first determining the "attractive" cutoff as two standard deviations above the vehicle average. Any visit longer than this was deemed "attractive" and scored as a "1"; non-attractive visits were scored as "0." The percent attraction was then calculated for each worm, which was the percent of visits scored as "1." The average was then calculated across the plate to determine percent attraction.

CRISPR design and strain generation

A novel null mutation was generated for *srw-97* by InVivo Biosystems. The *srw-97(knu456)* allele was generated in a *him-5(e1490)* strain using two sgRNAs (TTTAGTAGAGCAGAAATTAA and TACAGCTTTAACTTTCAAC) to generate a 1,620-nt deletion which removed the start codon and left only the terminal exon intact. The *knu456* knockout allele was generated by donor homology using the pNU1361odn oligo: (TTTTCTTGTATTTCCAAAAATTGTAAAAACCTTTATGAAAGTTAAAGCTGTAAGGATTTTCAGACATTTA). Following generation of a homozygous deletion by InVivo Biosystems, the line was then backcrossed twice into *him-5(e1490)*.

The *dmsr-12(tm8706)* allele, provided by the National BioResource Group (NRBP), contains a 118-nt deletion that was generated by Dr. Mitani of the NRBP. The allele was crossed into a *him-5(e1490)* background prior to testing. The deletion spans intron 2 and exon 3 of the coding sequence. Whether this results in a correctly spliced gene remains unknown, although the expected remaining coding sequence remains in frame.

Phylogenetic analyses

For phylogenetic analysis of selected CEM genes, we downloaded proteomes for *C. elegans* and related *Caenorhabditis* nematodes from WormBase (release WS275), the Blaxter *Caenorhabditis* database (release 1), or our unpublished work, as listed in Supplementary Table 4. From each proteome, we extracted the longest predicted isoform for each gene with *get_largest_isoforms.pl* (*https://github.com/SchwarzEM/ems_perl/blob/master/fasta/get_largest_isoforms.pl*). We observed that the predicted isoform for *dmsr-12* in the WormBase WS275 release of *C. elegans* proteome

Table 1. Expression levels for genes of interest in CEM neurons.

Gene	CEM DL	CEM DR	CEM VL	CEM VR	CEM/ others	CEM/ males
srw-97	0.02	0.03	0.02	0.10	3.33	2.50
dmsr-12	**1.26**	0.03	0.02	0.05	25.2	18.0
seb-3	0.01	0.02	0.01	**0.06**	3.00	0.02
srd-32	**0.07**	0.02	0.02	0.04	1.75	0.17
srr-7	0.03	0.03	**0.17**	0.04	4.25	0.68
pkd-2	**1.36**	0.09	0.42	1.15	1.18	0.15
trf-1	0.23	0.04	0.01	**78.02**	339.22	8.72
srw-98	0.02	0.03	0.02	**0.05**	1.67	0.20
dmsr-13	0.02	0.03	0.02	**0.05**	1.67	0.45
dmsr-10	0.02	0.04	0.03	**0.07**	1.75	1.40
dmsr-11	0.01	0.02	0.02	**0.04**	2.00	0.33
dmsr-16	0.02	0.04	0.03	**0.06**	1.50	0.67

For each gene, observed RNA-seq expression levels in each CEM neuron subtype are given in TPM. Two measures of gene enrichment and specificity are also shown: first, the ratio of a gene's expression for the highest expressing CEM neuron subtype divided by that gene's expression in the highest expressing of the other three CEM neuron subtypes ("CEM/others"; "CEM_any/CEM_non_any.max" in Supplementary Table 5); second, the ratio of that gene's highest expressing CEM neuron subtype divided by that gene's expression in the highest expressing of three replicates of adult *C. elegans* males ("CEM/males"; "CEM_all.max_TPM/male.max_TPM" in Supplementary Table 5). The first ratio gives the skew for a gene's expression towards a single CEM subtype; the second ratio gives the degree to which that gene's expression cannot be explained merely by its being a male-specific gene, perhaps expressed throughout the male body. Genes listed here are of interest either because we tested them for transgenic GFP expression or biological function, because they have previously known CEM specificity, or because they are evolutionary related to *srw-97* and *dmsr-12*. For each gene, the CEM with the highest expression are shown in bold.

was shorter than past versions of *dmsr-12* and that the WS275 isoform omitted exons that our transgenic expression data (based on older gene models for *dmsr-12*) indicated were likely to be real. We therefore manually replaced the WS275 version of *dmsr-12* with an older version (extracted from the *C. elegans* proteome in the WS250 release of WormBase). We then computed orthology groups for *C. elegans* and its related species with OrthoFinder version 2.3.11 (Emms and Kelly 2015; Emms and Kelly 2019), using the arguments "*-a 1 -S diamond --og*." We identified which orthology groups contained the *C. elegans* genes *srr-7*, *srw-97*, and *dmsr-12* and extracted their sequences from a concatenation of all 11 proteomes via *extract_fasta_subset.pl* (*https://github.com/SchwarzEM/ems_perl/blob/master/fasta/extract_fasta_subset.pl*). For each orthogroup's member sequence, we aligned the sequences with MAFFT version 7.455 (Katoh and Standley 2013) and filtered the alignments twice with trimAl version 1.4.rev15 (Capella-Gutiérrez *et al.* 2009), using first the argument "*-automated1*" and then the arguments "*-resoverlap 0.50 -seqoverlap 50*." From the filtered alignments, we computed maximum-likelihood protein phylogenies with IQ-TREE version 2.0-rc1 (Nguyen *et al.* 2015; Kalyaanamoorthy *et al.* 2017), using the arguments "*-m MFP -b 100 --tbe*." In particular, we used transfer bootstrap expectation ("*--tbe*") which provides more reliable confidence values than classic bootstrapping (Lemoine *et al.* 2018). We visualized the resulting phylogenies with FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Statistical analyses

Prior to any statistical analyses, outliers were identified and removed. Outliers were defined as any data points >2 standard deviations removed from the average. All data were then tested for normality using a Shapiro–Wilk normality test. This test was chosen over the more conventional D'Agostino–Pearson normality test as many data sets were below 10 in number (due to the statistical power offered by the single-worm behavioral assay; Reilly *et al.* 2021).

The spot retention assay data were analyzed using paired *t*-tests or Wilcoxon Matched–Pairs signed rank tests to compare vehicle control and ascaroside dwell times within strains following tests for normality (Supplementary Fig. 3). When comparing the values of multiple conditions or strains, the data were first normalized to account for vehicle dwell time variation between plates using a base-2 exponentiation, as described previously, to transform all data points into nonzero values. This allows for the calculation of the fold-change as the log(base2) of the ascaroside dwell time divided by the vehicle dwell time. These normalized values were then compared using a Mann–Whitney test or a one-way analysis of variance (ANOVA) followed by a Dunnett's multiple comparisons test (Supplementary Fig. 3).

The single-worm behavioral assay was first analyzed within each strain by performing either a repeated measures ANOVA followed by a Dunnett's multiple comparisons test or Friedman test followed by Dunn's correction, comparing both spatial control and ascaroside dwell times to that of the vehicle control. The log(fold-change) of raw ascaroside and vehicle dwell time values were calculated and compared using either a one-way ANOVA followed by a Bonferroni's multiple corrections test or Friedman test followed by Dunn's correction (Fig. 1). The ascr#3 log(fold-change) values were compared using a Student's *t*-test (Fig. 1e). Visit counts were compared in the same manner as the mean dwell time data, while the percent attraction data were analyzed using paired *t*-tests or Wilcoxon matched pairs signed rank tests to compare the attractive values of the vehicle and ascaroside (Supplementary Figs. 3, 4, and 6).

The expression patterns of srw-97 and *dmsr-12* were quantified in a binary manner, as present (1) or not present (0) within each CEM neuron (Fig. 2d and h). Their quantified expression was then analyzed using a chi-squared analysis.

Results

The transcriptomic landscape of CEM neurons is variable

Individual CEM neurons were isolated from *C. elegans* expressing an integrated GFP labeling male-specific EVNs (*ppkd-2*::GFP [Supplementary Fig. 1a]), as previously described (Goodman *et al.* 1998; Narayan *et al.* 2011, 2016). Cells were separated by anatomical identity [i.e. CEM dorsal left (DL), dorsal right (DR), ventral left (VL), and ventral right (VR)]; single-cell CEM cDNA libraries were constructed and used for RNA-seq (Schwarz *et al.* 2012).

Enriched genes in each CEM neuron were identified by comparing RNA-seq profiles of distinct CEM subtypes to one another and to whole adult *C. elegans* males (Supplementary Tables 1 and 5). We observed 267 genes with at least 2-fold higher expression in CEM neurons than in adult males. Of these, we observed 164 genes that were specific to individual CEM subtypes, defining specificity as being at least 2-fold higher expression in one subtype than in any other subtype: 62 genes specific to CEM DL, 40 to CEM DR, 18 to CEM VL, and 44 to CEM VR. Uniquely mapped reads ranged from 5.2 to 9.2 million per CEM subtype, matching the trend for alignment rates of each CEM neuron, which ranged from 18.0% to 47.2%, (Supplementary Table 2) with an average of 1,426 genes showing robust RNA expression in each neuron (Supplementary Table 3).

Five GPCR-encoding genes expressed in CEMs (*seb-3*, *srr-7*, *srw-97*, *dmsr-12*, and *srd-32*) were selected for study; of these, *srw-97* and *dmsr-12* showed specific expression in CEM VR and CEM DL. Four of these GPCR genes were uncharacterized; one (*seb-3*) has previously been shown to play roles in locomotion, stress

response, and ethanol tolerance (Jee *et al.* 2013). *dmsr-12* is related to *daf-37* (Robertson and Thomas 2006), a previously identified ascaroside receptor gene (Park *et al.* 2012), although it is more closely related to the neuropeptide receptor gene *dmsr-1*, and more distantly to *srw-97* (Robertson and Thomas 2006). *srd-32* belongs to a divergent branch of the SRD phylogeny (Robertson and Thomas 2006). *srr-7* belongs to the one of the smallest families of *C. elegans* chemoreceptor genes (Robertson and Thomas 2006).

seb-3 and *srw-97* exhibited similar enrichment profiles across the CEM neurons, with 3.0- and 3.3-fold higher expression in CEM VR than in any other CEM neuron (Table 1; Fig. 3d; Supplementary Table 5). *dmsr-12* showed 25-fold higher expression in CEM DL vs any other CEM neuron (Table 1; Fig. 3a; Supplementary Table 5), while *srd-32* was only 1.8-fold higher in CEM DL (Table 1; Fig. 3a; Supplementary Table 5). *srr-7* showed 4.2-fold higher expression in CEM VL vs any other CEM neuron (Table 1; Fig. 3c; Supplementary Table 5), which correlated with previous transcriptomic analyses that found *srr-7* to be enriched in *C. elegans* EVNs (Wang *et al.* 2015).

CEM-specific receptor expression patterns

To confirm our single-cell RNA-seq results, we generated transgenic GFP fusions for the five receptor genes (Boulin *et al.* 2006). Roughly 3 kb of promoter region upstream of the start codon was included in these constructs, along with the majority of the coding sequence (Supplementary Table 6); this would have automatically included any large 5′-ward introns that might contain cis-regulatory elements of these genes (Fuxman Bass *et al.* 2014). The GFP coding sequence was cloned from the Fire Kit vector pPD95.75 (Boulin *et al.* 2006). *pha-1;lite-1;him-5* animals were injected with reporter constructs and the co-injection marker pBX (*pha-1*(+)). We isolated GFP$^+$ strains and imaged GFP$^+$ males for expression at ×63 magnification (Supplementary Figs. 1 and 2). An integrated p*pkd-2*::GFP line was used as a CEM-specific control (Supplementary Fig. 1a).

The previously characterized GPCR gene, *seb-3*, displayed a non-CEM-specific expression pattern matching that previously described (Supplementary Fig. 1b). The other four receptor genes showed transgenic expression patterns similar to their RNA-seq data (Supplementary Fig. 1c–e). SRR-7 was observed in CEM VR, as well as another neuron that is likely to be the mechanoreceptor CEP VR (Sawin *et al.* 2000) (Supplementary Fig. 1f). Both *seb-3* and *srr-7* were excluded from further analyses, as we aimed to identify CEM-specific regulators of the ascr#8 response.

We also observed non-GPCR genes to be specifically expressed in single CEM neurons, such as *trf-1*, which encodes a TNF receptor homolog (Tenor and Aballay 2008) expressed in CEM VR (Table 1); *trf-1* has an EVN-specific promoter (Wang *et al.* 2015), providing an alternative to the *pkd-2* and *klp-6* promoters typically used to drive transgenes in EVNs (Supplementary Fig. 2) (Peden and Barr 2005; Bae *et al.* 2006).

RNAi-mediated knockdown of CEM receptors

To test whether these CEM-enriched genes encode receptors that are required in ascr#8 sensation, we fed dsRNA to a strain that is hypersensitive to neuronal RNA interference, *nre-1; lin-15B* (Schmitz *et al.* 2007; Poole *et al.* 2011).

To confirm that our system for male neuronal RNAi could affect behavioral phenotypes, we first fed animals *osm-3* and *osm-9* dsRNA clones from the Ahringer Library (Fraser *et al.* 2000; Kamath *et al.* 2003). Using a spot retention assay (Narayan *et al.* 2016), we assayed animals for their behavioral dwell time in ascr#8 (Supplementary Fig. 3a). Animals fed *osm-3* dsRNA showed significantly defective responses to ascr#8, as observed for loss-of-function alleles (Supplementary Fig. 3b). In contrast, animals fed *osm-9* showed only a slight decrease in their times spent within ascr#8 (Supplementary Fig. 3a). Because we could successfully abolish ascr#8 attraction through male neuronal RNAi of *osm-3*, we were confident that RNAi would be effective at functionally verifying CEM-enriched genes encoding components of the ascr#8 response.

We fed animals dsRNA clones targeting three CEM-enriched receptor genes: *srd-32*, *dmsr-12*, and *srw-97*. A *srd-32* clone was unavailable in the Ahringer Library (Fraser *et al.* 2000; Kamath *et al.* 2003) but was available from the Vidal Library (Rual *et al.* 2004). Each library uses the same backbone vector, allowing the same control to be used for both.

RNAi of *srd-32* caused no defect in ascr#8 responses (Supplementary Fig. 3c). Both *dmsr-12* or *srw-97* dsRNA caused partial defects of ascr#8 dwell time (Supplementary Fig. 3c). Neither knockdown statistically lowered the dwell time in ascr#8 (Supplementary Fig. 3d), though they did abolish the statistically significant increase in time spent in ascr#8 over vehicle controls (Supplementary Fig. 3c).

CRISPR-generated null mutants of candidate receptors

For phenotypic analysis of null mutants, we used our single-worm attraction assay (SWAA), as previously described (Reilly *et al.* 2021). Wild-type (*him-5*) animals were strongly attracted to ascr#8 in the SWAA, replicating previous observations (Fig. 1a and b; Supplementary Fig. 4) (Reilly *et al.* 2021).

srw-97(knu456) males displayed partially defective ascr#8 attraction: their ascr#8 dwell time was no different than that of vehicle, nor was it different than *him-5* animals (Fig. 1a and b; Supplementary Fig. 4). We were able to restore normal ascr#8 attraction in *srw-97(knu456)* males through transgenesis with a translational fusion construct (p*srw-97*::*srw-97*::GFP) (Fig. 1, Supplementary Fig. 5). Expression of the rescue transgene matched that of our initial fusion, with GFP visible in both ventral CEM neurons (Fig. 2a–d; Supplementary Fig. 1e).

Unlike *srw-97(knu456)*, *dmsr-12(tm8706)* males exhibited completely defective responses to ascr#8 (Fig. 1a and b; Supplementary Fig. 4). Similarly, the expression profile of the p*dmsr-12*::*dmsr-12*::GFP rescue construct (Fig. 2e–g) matched its earlier reporter expression (Supplementary Fig. 1c), and completely restored attraction to ascr#8 (Fig. 1a and b; Supplementary Figs. 4 and 5).

Given that *dmsr-12* is expressed in dorsal CEM neurons and *srw-97* is expressed in ventral CEM neurons (Fig. 2e–g), we speculated that single mutants of either receptor partially retain the ability to sense and respond to ascr#8, with the remaining CEM-specific receptor gene conferring some residual response to ascr#8. We thus generated a double *dmsr-12;srw-97* mutant strain and assayed animals for their ability to respond to ascr#8. These double mutants were not attracted to ascr#8 compared to the vehicle control (Fig. 1a; Supplementary Fig. 4), though they were not significantly defective in their attraction compared to wild-type animals either (Fig. 1b). However, this is likely a result of lower vehicle dwell times in the double mutants. This defect was specific to ascr#8; double mutants showed no defect in their responses to ascr#3 (Fig. 1d and e; Supplementary Fig. 6).

Both the *srw-97* and *dmsr-12* rescues (Supplementary Fig. 5) are propagated within these strains as extrachromosomal arrays. Expression being limited to a small number of neurons makes it difficult to isolate based on CEM expression, and so animals analyzed were selected through the co-injection marker (*unc-122*::RFP). To understand how the mosaic expression of *srw-97* and

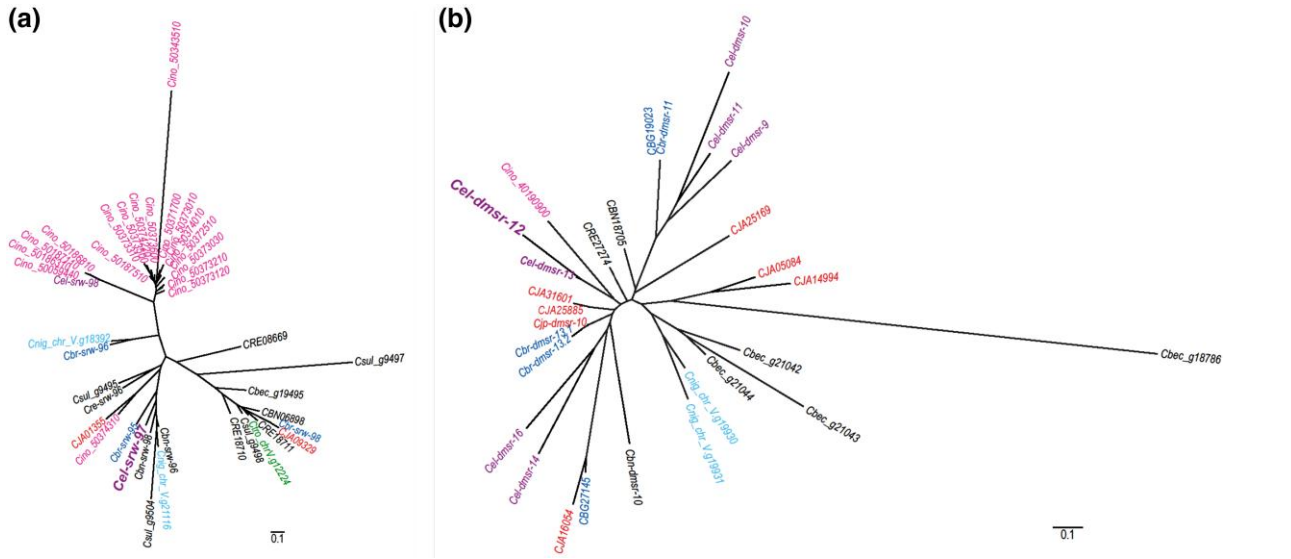

Fig. 4. Phylogenetic analysis of ascr#8 receptor candidate paralogs across the *Caenorhabditis* genus. a) The phylogeny of *Cel_srw-97* reveals a *C. inopinata*-specific amplification of *Cel_srw-98*. b) The phylogeny of *Cel_dmsr-12* shows conserved numbers of orthologs across the genus. Genes for *C. elegans* are denoted in purple, *C. inopinata* in pink, *C. nigoni* in light blue, *C. briggsae* in dark blue, *C. japonica* in red, and *C. tropicalis* in dark green. Distance reference bars (0.1) depict substitutions per site.

dmsr-12 rescues between individuals resulted in the population studied in Fig. 1, we analyzed the expression patterns across animals and found SRW-97 to be enriched within the CEM_VL neuron and to a lesser extent CEM_VR and CEM_DL neurons. These differences in expression levels between different sets of CEM subtypes were statistically significant (chi-squared analysis, $P = 0.0483$). DMSR-12 was observed within CEM_DL and CEM_DR neurons, with weaker differences in expression levels between sets of CEM subtypes (chi-squared analysis, $P = 0.2753$) (Supplementary Table 7).

Together, these data suggest that at least two GPCR receptors are expressed in discrete pairs of CEM neurons and act in parallel during sensation of ascr#8.

Phylogenetic analyses of ascr#8 receptors reveal likely gene duplication events

The ability to attract mates through pheromones is often essential for a species' survival. We have recently observed that different *Caenorhabditis* species show variable levels of attraction to ascr#8 (Reilly *et al.* 2019). To better understand variability in levels of attraction, we analyzed the evolution of both *srw-97* and *dmsr-12*.

We identified orthologs of *srw-97* from *C. elegans* (*Cel-srw-97*) in other *Caenorhabditis* proteomes via OrthoFinder (Fig. 4a). A closely related *C. elegans* paralog, *Cel-srw-98*, underwent a species-specific expansion within *C. inopinata*. The CDSes of both *Cel-srw-97* and *Cel-srw-98* are 66.2% identical to one another, while their amino acid sequences are only 57.1% identical (Madeira *et al.* 2019). *Cel-srw-98* shows enriched expression in CEM VR that is similar to (though weaker than) that of *Cel-srw-97* (Table 1; Supplementary Table 5).

C. tropicalis, like *C. elegans* and *C. briggsae*, is a hermaphroditic *Caenorhabditis* species (Noble *et al.* 2021), which encodes a reduced number of ascr#8 receptor gene paralogs, with no *Cel-dmsr-12* paralogs, and only one *Cel-srw-97* paralog (Fig. 4). We previously observed that *C. tropicalis* fails to be attracted to ascr#8, and it is possible that this loss of receptor genes is one reason for that failure (Reilly *et al.* 2019).

In contrast, the other candidate ascr#8 receptor, *dmsr-12*, does not appear in our phylogenetic analysis to have undergone any species-specific expansion (Fig. 4b). Rather, there are generally fewer paralogs per non-*C. elegans* species.

Discussion

Pheromones are important for mating in many animals. In the nematode *C. elegans*, males are attracted to hermaphrodites as possible mates by the small-molecule pheromone ascaroside #8 (Pungaliya *et al.* 2009; Narayan *et al.* 2016). While the ascaroside class of small-molecule pheromones utilized by nematodes is rapidly being elucidated (with over 230 known ascaroside structures so far; https://smid-db.org) (Artyukhin *et al.* 2018), the neuronal receptor proteins that mediate pheromone signals remain largely unknown. For only a select few ascarosides have sensory components been identified at the cellular (Greene, Dobosiewicz *et al.* 2016; Greene, Brown, *et al.* 2016; Chute *et al.* 2019), receptor (Kim *et al.* 2009; McGrath *et al.* 2011; Park *et al.* 2012), or signal transduction levels (Zwaal *et al.* 1997).

Here, we identify two novel GPCRs as active, required components in sensory and behavioral responses to the mating pheromone ascr#8. Transcripts for the two GPCRs, *dmsr-12* and *srw-97*, are enriched in single CEM neurons (Table 1), and they express in nonoverlapping subsets of the male-specific chemosensory neurons (Figs. 2 and 3, Supplementary Fig. 1). There are, however, other receptors present in these same neurons that contribute to the navigation of a vast and ever-changing array of environmental cues, such as the widely expressed ethanol sensor *seb-3* (Supplementary Fig. 1b) (Jee *et al.* 2013).

One mechanism for sensory flexibility in CEM neurons may be heterodimerization of receptor proteins. Previous work identified two receptors for ascr#2 in the ASK neuron: DAF-37 and DAF-38 (Park *et al.* 2012). While both are required for proper ascr#2-induced dauer formation, only DAF-38 is involved in the sensation of other ascarosides (Park *et al.* 2012). Such heterodimers may also exist for pheromone receptors in CEM neurons.

Another mechanism for sensory flexibility in CEM neurons is to have diverse receptor proteins with distinct specificities co-expressed within a single neuron. Multiple ascarosides are sensed

by the male-specific CEM neurons, including ascr#8 and ascr#3 (Narayan *et al.* 2016). The two receptors identified here may function as ascr#8-specific receptors, as double *dmsr-12;srw-97* mutant animals do not exhibit defective responses to ascr#3 (Fig. 1d and e). The related receptors *dmsr-10*, *dmsr-13*, *dmsr-16*, and *srw-98* are enriched in ventral CEM neurons and may heterodimerize with *srw-97*, to result in fine-tuned sensation of ascr#8. Similarly, there may be other receptors that may be enriched in the dorsal CEM neurons to heterodimerize with *dmsr-12*. The role of DAF-37 as an ascr#2-specific receptor further supports our hypothesis, as *dmsr-12* is related to DAF-37 (Robertson and Thomas 2006). The ventral CEM receptor SRW-97 falls within the same large family of GPCRs as well (Krishnan *et al.* 2014), while the promiscuous DAF-38 does not (Robertson and Thomas 2006).

The initial goal of our single-cell RNA-seq analysis of male-specific chemosensory CEM neurons was to identify GPCRs encoding pheromone receptors. However, further analysis of other genes with CEM-enriched expression should also uncover specific and novel promoter profiles, which will enable optogenetic manipulation and calcium imaging of defined individual CEM neurons (Narayan *et al.* 2016).

There are some limitations to the profiling method used in this study. For example, the gene *trf-1* was found to be enriched in CEM VR (Table 1). However, while the gene is enriched 78-fold in CEM VR over the remaining CEM subtypes, the 340-fold enrichment in pooled CEM compared to the entire animal is what is observed in our expression studies (Supplementary Fig. 2), as well as in EVN-pooled transcriptomics (Wang *et al.* 2015). In addition, *pkd-2* expression in this study was relatively low, despite being one of the most highly expressed genes in EVNs. This may be due to differences between the developmental stage of *pkd-2* peak transcriptional expression (L1 larvae) and the stage at which we collected CEMs (L4s and young adults) (Gerstein *et al.* 2010). This may also be affected by the small number of cells analyzed, the number of replicates within each subtype, and the comparison to a male-only read set (Thomas *et al.* 2012). As such, future studies should use individual CEM subtype transcriptomes in parallel with translational reporter studies to confirm expression patterns of candidate cell-specific genes.

The development of advanced transgenic reagents will also permit chemical biology of CEM neurons. We have recently developed an active ascr#8 bioaffinity probe to employ in the targeted elucidation of ascr#8 receptors (Zhang *et al.* 2019). Use of this probe will lead to further elucidation of the identity of acsr#8 receptors, by confirming either SRW-97 or DMSR-12 as receptors or identifying their heterodimeric partners. The combination of these technologies should clarify how heterogeneous CEM neurons achieve homogeneous sensory responses.

Data availability

RNA-seq reads have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA781271. Supplementary Table 5 is available at figshare: https://doi.org/10.25386/genetics.22057019.

Supplemental material available at GENETICS online.

Acknowledgments

We thank the *Caenorhabditis* Genetics Center, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD01044), as well as the National BioResource Project, Ding Xue (University of Colorado, Boulder), L. Rene Garcia (Texas A&M University), and Douglas Portman (University of Rochester Medical Center) for providing strains. We also thank Igor Antoshechkin, Caltech Genomics facility for sequencing, and InVivo Biosystems for generating transgenic and CRISPR knockout animals. The synthetic ascr#8 utilized in this study was provided by Frank Schroeder (Cornell University).

Funding

The research reported in this publication was supported by National Institutes of Health R01 DC016058 (J.S.), R01 GM084389 (P.W.S.), the Howard Hughes Medical Institute (P.W.S.), Moore Foundation Grant No. 4551 (E.M.S.), and Cornell startup funding (E.M.S.). We thank Titus Brown and the Michigan State University High-Performance Computing Center (supported by U.S. Department of Agriculture grant 2010-65205-20361 and NIFA–National Science Foundation (NSF) grant IOS-0923812) for computational support; additional computing was enabled by start-up and research allocations from NSF XSEDE (TG-MCB180039 and TG-MCB190010).

Conflict of interest

The authors declare no conflict of interest.

Literature cited

Aprison EZ, Ruvinsky I. Sex pheromones of *C. elegans* males prime the female reproductive system and ameliorate the effects of heat stress. PLoS Genet. 2015;11(12):e1005729. doi:10.1371/journal.pgen.1005729.

Aprison EZ, Ruvinsky I. Counteracting ascarosides act through distinct neurons to determine the sexual identity of *C. elegans* pheromones. Curr Biol. 2017;27(17):2589–2599.e3. doi:10.1016/j.cub.2017.07.034.

Artyukhin AB, Zhang YK, Akagi AE, Panda O, Sternberg PW, Schroeder FC. Metabolomic "dark matter" dependent on peroxisomal beta-oxidation in *Caenorhabditis elegans*. J Am Chem Soc. 2018;140(8):2841–2852. doi:10.1021/jacs.7b11811.

Bae Y-K, Qin H, Knobel KM, Hu J, Rosenbaum JL, Barr MM. General and cell-type specific mechanisms target TRPP2/PKD-2 to cilia. Development. 2006;133(19):3859–3870. doi:10.1242/dev.02555.

Boulin T, Etchberger JF, Hobert O. Reporter gene fusions. WormBook. 2006:1–23. doi:10.1895/wormbook.1.106.1.

Butcher RA, Fujita M, Schroeder FC, Clardy J. Small-molecule pheromones that control dauer development in *Caenorhabditis elegans*. Nat Chem Biol. 2007;3(7):420–422. doi:10.1038/nchembio.2007.3.

Butcher RA, Ragains JR, Kim E, Clardy J. A potent dauer pheromone component in *Caenorhabditis elegans* that acts synergistically with other components. Proc Natl Acad Sci USA. 2008;105(38):14288–14292. doi:10.1073/pnas.0806676105.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics (Oxford, England). 2009;25(15):1972–1973. doi:10.1093/bioinformatics/btp348.

Choe A, von Reuss S, Kogan D, Gasser R, Platzer E, Schroeder F, Sternberg P. Ascaroside signaling is widely conserved among nematodes. Curr Biol. 2012;22(9):772–780. doi:10.1016/j.cub.2012.03.024.

Chute CD, DiLoreto EM, Zhang YK, Reilly DK, Rayes D, Coyle VL, Choi HJ, Alkema MJ, Schroeder FC, Srinivasan J. Co-option of neurotransmitter signaling for inter-organismal communication in *C.*

elegans. Nat Commun. 2019;10(1):3186. doi:10.1038/s41467-019-11240-7.

Cochella L, Tursun B, Hsieh Y-W, Galindo S, Johnston RJ, Chuang C-F, Hobert O. Two distinct types of neuronal asymmetries are controlled by the *Caenorhabditis elegans* zinc finger transcription factor *die-1*. Genes Dev. 2014;28(1):34–43. doi:10.1101/gad.233643.113.

Dong C, Dolke F, von Reuss SH. Selective MS screening reveals a sex pheromone in *Caenorhabditis briggsae* and species-specificity in indole ascaroside signalling. Org Biomol Chem. 2016;14(30):7217–7225. doi:10.1039/C6OB01230B.

Dong C, Reilly DK, Bergame C, Dolke F, Srinivasan J, von Reuss SH. Comparative ascaroside profiling of *Caenorhabditis* exometabolomes reveals species-specific (ω) and (ω - 2)-hydroxylation downstream of peroxisomal beta-oxidation. J Org Chem. 2018;83(13):7109–7120. doi:10.1021/acs.joc.8b00094.

Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform. 2009;23(1):205–211. doi:10.1142/9781848165632_0019.

Emms DM, Kelly S. Orthofinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157. doi:10.1186/s13059-015-0721-2.

Emms DM, Kelly S. Orthofinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. doi:10.1186/s13059-019-1832-y.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, *et al.* The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1):D279–D285. doi:10.1093/nar/gkv1344.

Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature. 2000;408(6810):325–330. doi:10.1038/35042517.

Fuxman Bass JI, Tamburino AM, Mori A, Beittel N, Weirauch MT, Reece-Hoyes JS, Walhout AJM. Transcription factor binding to *Caenorhabditis elegans* first introns reveals lack of redundancy with gene promoters. Nucleic Acids Res. 2014;42(1):153–162. doi:10.1093/nar/gkt858.

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science. 2010;330(6012):1775–1787. doi:10.1126/science.1196914.

Goodman MB, Hall DH, Avery L, Lockery SR. Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. Neuron. 1998;20(4):763–772. doi:10.1016/S0896-6273(00)81014-4.

Greene JS, Brown M, Dobosiewicz M, Ishida IG, Macosko EZ, Zhang X, Butcher RA, Cline DJ, McGrath PT, Bargmann CI. Balancing selection shapes density-dependent foraging behaviour. Nature. 2016;539(7628):254–258. doi:10.1038/nature19848.

Greene JS, Dobosiewicz M, Butcher RA, McGrath PT, Bargmann CI. Regulatory changes in two chemoreceptor genes contribute to a *Caenorhabditis elegans* QTL for foraging behavior. Elife. 2016;5:e21454. doi:10.7554/eLife.21454.

Hobert O. The neuronal genome of *Caenorhabditis elegans*. WormBook. 2013:1–106. doi:10.1895/wormbook.1.161.1.

Houck LD, Palmer CA, Watts RA, Arnold SJ, Feldhoff PW, Feldhoff RC. A new vertebrate courtship pheromone, PMF, affects female receptivity in a terrestrial salamander. Anim Behav. 2007;73(2):315–320. doi:10.1016/j.anbehav.2006.07.008.

Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, Done J, Down T, Gao S, Grove C, *et al.* Wormbase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. 2016;44(D1):D774–D780. doi:10.1093/nar/gkv1217.

Jang YH, Chae HS, Kim YJ. Female-specific myoinhibitory peptide neurons regulate mating receptivity in *Drosophila melanogaster*. Nat Commun. 2017;8(1):1630. doi:10.1038/s41467-017-01794-9.

Jee C, Lee J, Lim JP, Parry D, Messing RO, McIntire SL. SEB-3, a CRF receptor-like GPCR, regulates locomotor activity states, stress responses, and ethanol tolerance in *C. elegans*. Genes, Brain and Behavior. 2013;12(2):250–262. doi:10.1111/j.1601-183X.2012.00829.x.

Kaletsky R, Lakhina V, Arey R, Williams A, Landis J, Ashraf J, Murphy CT. The *C. elegans* adult neuronal IIS/FOXO transcriptome reveals adult phenotype regulators. Nature. 2016;529(7584):92–96. doi:10.1038/nature16483.

Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338(5):1027–1036. doi:10.1016/j.jmb.2004.03.016.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. Modelfinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–589. doi:10.1038/nmeth.4285.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature. 2003;421(6920):231–237. doi:10.1038/nature01278.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780. doi:10.1093/molbev/mst010.

Kim K, Sato K, Shibuya M, Zeiger DM, Butcher RA, Ragains JR, Clardy J, Touhara K, Sengupta P. Two chemoreceptors mediate developmental effects of dauer pheromone in *C. elegans*. Science. 2009;326(5955):994–998. doi:10.1126/science.1176331.

Krishnan A, Almén MS, Fredriksson R, Schiöth HB. Insights into the origin of nematode chemosensory GPCRs: putative orthologs of the srw family are found across several phyla of protostomes. PLoS ONE. 2014;9(3):e93048. doi:10.1371/journal.pone.0093048.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359. doi:10.1038/nmeth.1923.

Lemoine F, Domelevo Entfellner J-B, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature. 2018;556(7702):452–456. doi:10.1038/s41586-018-0043-0.

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC bioinformatics. 2011;12(1):323–323. doi:10.1186/1471-2105-12-323.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Ludewig AH, Artyukhin AB, Aprison EZ, Rodrigues PR, Pulido DC, Burkhardt RN, Panda O, Zhang YK, Gudibanda P, Ruvinsky I, *et al.* An excreted small molecule promotes *C. elegans* reproductive development and aging. Nat Chem Biol. 2019;15(8):838–845. doi:10.1038/s41589-019-0321-7.

Lupas A. Prediction and analysis of coiled-coil structures. Meth Enzymol. 1996;266:513–525. doi:10.1016/S0076-6879(96)66032-7.

Madeira F, Park Ym, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47(W1):W636–W641. doi:10.1093/nar/gkz268.

McGrath PT, Ruvinsky I. A primer on pheromone signaling in *Caenorhabditis elegans* for systems biologists. Curr Opin Syst Biol. 2019;13:23–30. doi:10.1016/j.coisb.2018.08.012.

McGrath PT, Xu Y, Ailion M, Garrison JL, Butcher RA, Bargmann CI. Parallel evolution of domesticated *Caenorhabditis* species targets

pheromone receptor genes. Nature. 2011;477(7364):321–325. doi:10.1038/nature10378.

Narayan A, Laurent G, Sternberg PW. Transfer characteristics of a thermosensory synapse in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 2011;108(23):9667–9672. doi:10.1073/pnas.1106617108.

Narayan A, Venkatachalam V, Durak O, Reilly DK, Bose N, Schroeder FC, Samuel ADT, Srinivasan J, Sternberg PW. Contrasting responses within a single neuron class enable sex-specific attraction in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 2016;113(10):E1392-1401. doi:10.1073/pnas.1600786113.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–274. doi:10.1093/molbev/msu300.

Noble LM, Yuen J, Stevens L, Moya N, Persaud R, Moscatelli M, Jackson JL, Zhang G, Chitrakar R, Baugh LR, *et al.* Selfing is the safest sex for *Caenorhabditis tropicalis*. Elife. 2021;10:e62587. doi:10.7554/eLife.62587.

Park D, O'Doherty I, Somvanshi RK, Bethke A, Schroeder FC, Kumar U, Riddle DL. Interaction of structure-specific and promiscuous G-protein-coupled receptors mediates small-molecule signaling in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 2012;109(25):9917–9922. doi:10.1073/pnas.1202216109.

Peden EM, Barr MM. The KLP-6 kinesin is required for male mating behaviors and polycystin localization in *Caenorhabditis elegans*. Curr Biol. 2005;15(5):394–404. doi:10.1016/j.cub.2004.12.073.

Poole RJ, Bashllari E, Cochella L, Flowers EB, Hobert O. A genome-wide RNAi screen for factors involved in neuronal specification in *Caenorhabditis elegans*. PLoS Genet. 2011;7(6):e1002109. doi:10.1371/journal.pgen.1002109.

Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, *et al.* eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res. 2012;40(D1):D284–D289. doi:10.1093/nar/gkr1060.

Pungaliya C, Srinivasan J, Fox BW, Malik RU, Ludewig AH, Sternberg PW, Schroeder FC. A shortcut to identifying small molecule signals that regulate behavior and development in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 2009;106(19):7708–7713. doi:10.1073/pnas.0811918106.

Ragsdale EJ, Muller MR, Rodelsperger C, Sommer RJ. A developmental switch coupled to the evolution of plasticity acts through a sulfatase. Cell. 2013;155(4):922–933. doi:10.1016/j.cell.2013.09.054.

Reilly DK, Lawler DE, Albrecht DR, Srinivasan J. Using an adapted microfluidic olfactory chip for the imaging of neuronal activity in response to pheromones in male *C. elegans* head neurons. J Vis Exp. 2017;(127):e56026. doi:10.3791/56026.

Reilly DK, McGlame EJ, Vandewyer E, Robidoux AN, Muirhead CS, Northcott HT, Joyce W, Alkema MJ, Gegear RJ, Beets I, *et al.* Distinct neuropeptide-receptor modules regulate a sex-specific behavioral response to a pheromone. Communications Biology. 2021;4(1):1018. doi:10.1038/s42003-021-02547-7.

Reilly DK, Randle LJ, Srinivasan J. Evolution of hermaphroditism decreases efficacy of ascaroside#8-mediated mate attraction in *Caenorhabditis* nematodes. MicroPubl Biol. 2019. doi:10.17912/micropub.biology.000134.

Robertson HM, Thomas JH. The putative chemoreceptor families of *C. elegans*. WormBook. 2006:1–12. doi:10.1895/wormbook.1.66.1.

Rual J-F, Ceron J, Koreth J, Hao T, Nicot A-S, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den Heuvel S, *et al.* Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. Genome Res. 2004;14-(10b):2162–2168. doi:10.1101/gr.2505604.

Sawin ER, Ranganathan R, Horvitz HR. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. Neuron. 2000;26(3):619–631. doi:10.1016/S0896-6273(00)81199-X.

Schmitz C, Kinge P, Hutter H. Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive *Caenorhabditis elegans* strain *nre-1(hd20) lin-15b(hd126)*. Proc Natl Acad Sci USA. 2007;104(3):834–839. doi:10.1073/pnas.0510527104.

Schwarz EM, Kato M, Sternberg PW. Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 2012;109(40):16246–16251. doi:10.1073/pnas.1203045109.

Taylor SR, Santpere G, Weinreb A, Barrett A, Reilly MB, Xu C, Varol E, Oikonomou P, Glenwinkel L, McWhirter R, *et al.* Molecular topography of an entire nervous system. Cell. 2021;184(16):4329–4347.e23. doi:10.1016/j.cell.2021.06.023.

Tenor JL, Aballay A. A conserved toll-like receptor is required for *Caenorhabditis elegans* innate immunity. EMBO Rep. 2008;9(1):103–109. doi:10.1038/sj.embor.7401104.

Thomas C, Li R, Smith H, Woodruff G, Oliver B, Haag E. Simplification and desexualization of gene expression in self-fertile nematodes. Curr Biol. 2012;22(22):2167–2172. doi:10.1016/j.cub.2012.09.038.

von Reuss SH, Bose N, Srinivasan J, Yim JJ, Judkins JC, Sternberg PW, Schroeder FC. Comparative metabolomics reveals biogenesis of ascarosides, a modular library of small-molecule signals in *C. elegans*. J Am Chem Soc. 2012;134(3):1817–1824. doi:10.1021/ja210202y.

von Reuss SH, Schroeder FC. Combinatorial chemistry in nematodes: modular assembly of primary metabolism-derived building blocks. Nat Prod Rep. 2015;32(7):994–1006. doi:10.1039/C5NP00042D.

Wang J, Kaletsky R, Silva M, Williams A, Haas LA, Androwski RJ, Landis JN, Patrick C, Rashid A, Santiago-Martinez D, *et al.* Cell-specific transcriptional profiling of ciliated sensory neurons reveals regulators of behavior and extracellular vesicle biogenesis. Curr Biol. 2015;25(24):3232–3238. doi:10.1016/j.cub.2015.10.057.

Wootton JC. Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem. 1994;18(3):269–285 doi:10.1016/0097-8485(94)85023-2.

Zhang YK, Reilly DK, Yu J, Srinivasan J, Schroeder FC. Photoaffinity probes for nematode pheromone receptor identification. Org Biomol Chem. 2019;18(1):36–40. doi:10.1039/c9ob02099c.

Zhang YK, Sanchez-Ayala MA, Sternberg PW, Srinivasan J, Schroeder FC. Improved synthesis for modular ascarosides uncovers biological activity. Org Lett. 2017;19(11):2837–2840. doi:10.1021/acs.orglett.7b01009.

Zwaal RR, Mendel JE, Sternberg PW, Plasterk RH. Two neuronal G proteins are involved in chemosensation of the *Caenorhabditis elegans* dauer-inducing pheromone. Genetics. 1997;145(3):715–727. doi:10.1093/genetics/145.3.715.

Editor: A. Barrios