The putative chemoreceptor families of *C. elegans*[*]

Hugh M. Robertson[§], Department of Entomology, University of Illinois at Urbana, IL 61801 USA

James H. Thomas[§], Department of Genome Sciences, University of Washington, Seattle, WA 98195-7360 USA

Table of Contents

Abstract

Chemoreception of environmental stimuli is a major sensory system in small soil nematodes like *C. elegans*. As in other animals, chemoreception is mediated in *C. elegans* by members of the seven-transmembrane G-protein-coupled receptor class (7TM GPCRs). We summarize the many large putative chemoreceptor gene families, including the *str* family (which includes *odr-10,* the only receptor with an identified ligand), and the *sra, srab, srb, srbc, srd, sre, srg, srh, sri, srj, srm, srr, srsx, srt, sru, srv, srw, srx, srxa,* and *srz* families. Together these comprise ±1280 apparently intact genes and ±420 apparent pseudogenes, about 7% of the total gene count of *C. elegans*. These genes are unusually clustered on chromosomes, both within and between families, and are enigmatically concentrated on the large chromosome V. Comparative studies with *C. briggsae* have revealed extraordinary divergence of the

[§]To whom correspondence should be addressed. E-mail: hughrobe@uiuc.edu or jht@u.washington.edu

chemoreceptor repertoire between the two species, including frequent amplifications of subfamilies in *C. elegans* and positive selection in the *srz* family. The size and complexity of the chemoreceptor gene families also facilitate studies of promoter elements using paralogous and orthologous comparisons, as well as other aspects of gene family and genome evolution.

1. Introduction

Chemoperception is a central sense of soil nematodes like *Caenorhabditis elegans*, as it is for most other animals (reviewed in Mori, 1999; Sengupta, 1997; Bargmann, 1997). Following the Nobel-Prize-winning discovery of olfactory receptors (ORs) in mammals (Buck and Axel, 1991) as members of the rhodopsin superfamily of G-protein-coupled receptors (GPCRs) with seven transmembrane domains (7TM), Troemel et al. identified a series of disparate GPCRs as candidate olfactory receptors in *C. elegans* (Troemel et al., 1995). These identifications were made from some of the earliest sequences reported by the genome project and proved to be the tip of the iceberg. The genes were classified into small gene families and given names starting with sr for serpentine receptor (*sra*, *srb*, *srd*, *sre*, and *sro*). Their probable chemosensory function was based on transgene expression patterns in one or more known pairs of chemosensory neurons. Subsequently, Sengupta et al. used a genetic screen to identify *odr-10*, which they showed encodes a GPCR that mediates olfactory response to diacetyl (Sengupta et al., 1996). With completion of the genome sequence it became possible to flesh out the true extent of these and related gene families encoding GPCRs that are candidate chemoreceptors. Robertson (Robertson, 1998; Robertson, 2001) worked up the relatives of *odr-10* as the large *str* family along with the related *sri* and *srj* families, and then also worked up the large related *srh* family (Robertson, 2000). Annotation and naming of the smaller *sra* family was completed by HMR as part of the comparison with *C. briggsae* (Stein et al., 2003), while Chen et al. described the related *srab* family (Chen et al., 2005), and Thomas et al. described the *srz* family (Thomas et al., 2005) and the *srt* family (Thomas, 2006 [in press]). A decade after their first recognition, we describe here all the remaining families and relate them in a series of superfamilies that comprise a remarkable array of roughly 1300 potentially functional candidate chemoreceptor genes, along with about 400 apparent pseudogenes. Together they comprise roughly 7% of the gene complement of this tiny animal, more than any other sequenced genome. This large genetic investment might result from an extreme dependence on chemosensory abilities in the absence of visual and auditory systems. On the basis of genetic complexity, this system probably constitutes the bulk of signal transduction that occurs in *C. elegans*.

We first discuss these families below in superfamily groupings, loosely defined as consisting of readily recognizable groupings based on protein sequence similarity, as well as sometimes shared intron locations in their genes. We then discuss various aspects of their biology, including new expression pattern results that support their likely function as chemoreceptors. For the rest of this chapter, while fully acknowledging the formal status of all but *odr-10* as putative or candidate chemoreceptors, we will refer to them as chemoreceptors. Genetic analysis of sensory-neuron specific G proteins indicates that olfaction, nociception, and pheromone responses are mediated predominantly by GPCRs (e.g., Hilliard et al., 2004; Jansen et al., 1999; Jansen et al., 2002; Roayaie et al., 1998) and it is difficult to see what else these huge families might generally mediate. Nevertheless, it seems likely that some of the genes have been subsequently recruited to other roles such as monitoring of internal chemistry, so the function of any specific gene should be interpreted with caution.

2. Summary of the chemoreceptor families

Table 1 lists the named chemoreceptor families by superfamily group with the numbers of apparently intact genes and of apparent pseudogenes in each. Annotation of the genes is not nearly as straightforward as for mammalian olfactory receptors, because the *C. elegans* genes have complex gene structures with 1–8 introns each, and the intron locations are commonly different between gene families. EST data are of little assistance due to uniformly low expression levels (e.g., of ~160,000 ESTs on WS138, only 81 were from chemoreceptor genes). In the absence of functional information it is not possible to be confident of the functionality of any specific gene, but each has been annotated to encode a seemingly intact full-length protein that aligns well with other family members. It is always possible that particular amino acid changes or small in-frame insertions or deletions might inactivate an otherwise apparently intact receptor. On the other hand, amongst the apparent pseudogenes from N2 is a subset with probable functional alleles in other wild strains of *C. elegans* (Stewart et al., 2005). In part because of this we have chosen to name both genes and pseudogenes in these families. We have arbitrarily chosen to ignore gene fragments that encode less than 50% of the amino acid length typical for a gene family. The curated annotations and gene names for all families should be correct in WormBase release WS141, though ongoing studies will undoubtedly refine their structure.

Table 1. Gene and pseudogene numbers in chemoreceptor families and superfamilies

Superfamily	Family	Genes	Pseudogenes	% Pseudo	Total loci
Str	*srh*	217	91	30	308
	str	192	75	28	267
	sri	60	18	23	78
	srd	64	12	16	76
	srj	39	18	32	57
	srm	5	1	17	6
	srn	1	0	0	1
	all Str	578	215	27	793
Sra	*sre*	51	5	9	56
	sra	32	7	18	39
	srab	22	5	19	27
	srb	14	5	26	19
	all Sra	119	22	16	141
Srg	*srx*	94	44	32	138
	srt	59	15	20	74
	srg	59	9	13	68
	sru	39	9	19	48
	srv	30	6	17	36
	srxa	17	2	11	19
	all Srg	298	85	22	383
Solo	*srw*	100	45	31	145
Solo	*srz*	64	39	38	103
Solo	*srbc*	71	13	15	84
Solo	*srsx*	37	3	8	40
Solo	*srr*	9	1	10	10
All family totals:		1276	423	25	1699

Pseudogenes and functional genes are distinguished by the presence of apparent defects in coding potential (mostly deletions, stop codons, and frameshifts) and are occasionally subjective; these numbers should be taken as current best estimates. The *srbc* and *srsx* families are probably distantly related to each other and perhaps to the *srx* family (JHT, unpublished), but these relationship are not yet sufficiently clear to justify inclusion in the table. An additional ~150 seven-pass receptors may also function as chemoreceptors, but they are either in very small families or are unique in the genome and their status is unclear.

2.1. The Str superfamily

The Str superfamily consists of the *str*, *srd*, *srh*, *sri*, *srj*, *srm*, and *srn* genes, with a total of 795 genes and pseudogenes. The superfamily includes the *str* and *srh* families, which are the two largest gene families in *C. elegans*. Robertson was able to extract considerable information about gene family, intron, and chromosomal evolution from studies of these large families (Robertson, 1998; Robertson, 2000; Robertson, 2001). Similar analysis has subsequently been extended to all of the other families. The high frequency of pseudogenes in these families was novel at the time for *C. elegans*, but is now more commonly recognized in the genome (e.g., Mounsey et al., 2002). Rampant gain and loss of introns was inferred from phylogenetic analysis and has been confirmed by

genome-wide and comparative analyses with *C. briggsae* (e.g., Coghlan and Wolfe, 2004; Kiontke et al., 2004). Processes of genome flux have moved these genes around a lot, although somewhat mysteriously retaining most of them on the large chromosome V, which might reflect functional clustering with each other and other chemoreceptor genes, similar to that inferred for several other gene classes (e.g., Miller et al., 2004). Comparisons across the entire superfamily show that the srd family is a highly divergent and basal family (Figure 1), with just one intron location ancestrally shared in the family and across the entire superfamily. A few genes and pseudogenes have been added to the larger of these families since their original description with completion of the genome sequence.
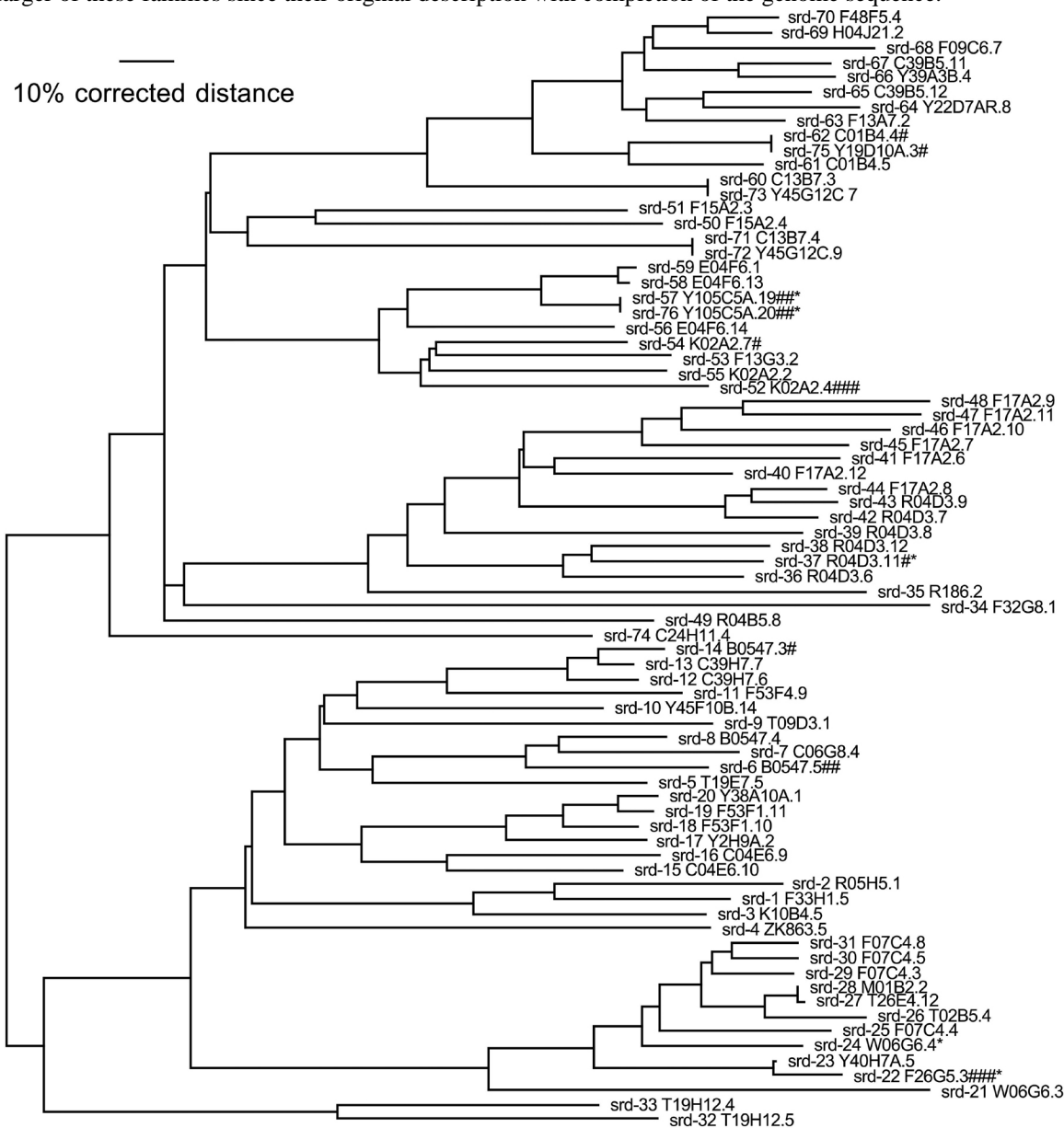
Figure 1. Phylogenetic tree of the *srd* family, rooted at the midpoint. Receptor amino acid sequences were aligned using CLUSTALX (Chenna et al., 2003) after refinement of gene annotations. Phylogenetic analysis was performed using the minimum distance heuristic algorithm in PAUP * v4.0b10 (Swofford, 1998), with distances corrected for multiple changes using the maximum likelihood model in TREE-PUZZLE v5 (Schmidt et al., 2002) with the BLOSUM62 matrix (Henikoff and Henikoff, 1992). Pseudogenes are indicated by the following symbols after their cosmid gene names (# - frameshifting or large indel or truncation; * - in-frame stop codon; ? -unacceptable intron boundary or missing start codon). Right click or control-click for larger image.

2.2. The Sra superfamily

The Sra superfamily consists of the *sra*, *srab*, *srb*, and *sre* families, totaling 141 genes and pseudogenes. Comparison of the *sra* and *srab* families with their homologs in *C. briggsae* revealed that these families are

considerably expanded in *C. elegans* (Chen et al., 2005; Stein et al., 2003), something that appears to be true for many of the chemoreceptor gene families. Using computational gene-building methods, Chen et al. (Chen et al., 2005) did not note any pseudogenes in the *srab* family in *C. elegans*, although they did note them in the *srab* family in *C. briggsae* and the *sra* family in both species. Chen and HMR have reevaluated this family in *C. elegans*, and now recognize four pseudogenes and added one gene. Phylogenetic analysis of the superfamily reveals that the *sre* family is the most divergent, with the *sra*, *srab*, and *srb* families being clustered together. The last three named members of the *sra* family (*sra*-37-39, T21H8.2-4) cluster with the *srb* family in some analyses, but are best left in the *sra* family with this caveat.

2.3. The Srg superfamily

The Srg superfamily consists of the *srg*, *srt*, *sru*, *srv*, *srx*, and *srxa* families, totaling 310 genes and pseudogenes, with the latter being the most divergent. Evolutionary analysis of these families has been limited, and virtually nothing is known about their specific functions, though expression patterns are consistent with a role in chemosensation (Table 2).

Table 2. Summaries of known GFP-fusion expression patterns for chemoreceptors

| Family	Expressed	Unexpressed	Sensory	Amphid	Phasmid	Other sensory	Other neurons	Non-neuronal
sra	10	3	7/10	5	1	3	4	2
srab	5	0	5/5	3	1	1	1	0
srd	5	4	4/5	4	1	1	1	0
srg	4	0	3/4	2	1	0	0	1
srh/sri	15	1	12/15	12	1	3	3	8
str/srj	16	1	10/16	10	0	0	7	2
other	8	5	7/8	7	1	1	2	1
total	63	14	48/63	43	6	9	18	14

Some genes are expressed in more than one listed tissue. Data were gathered from Battu et al., 2003; Chen et al., 2005; McCarroll et al., 2005; Sengupta et al., 1996; Troemel et al., 1995; Wenick and Hobert, 2004, and generous personal communications from Cori Bargmann and Piali Sengupta.

2.4. The *srw* family

The *srw* family of 145 genes and pseudogenes is unusual among the candidate chemoreceptors because its origins from another family of otherwise conserved GPCRs can be recognized. This family of highly divergent genes and many pseudogenes is related to families of apparent FMRFamide and other peptide receptors, which have relatives in vertebrates and insects (e.g., Egerod et al., 2003; Meeusen et al., 2002; Mertens et al., 2004; Figure 2). The candidate FMRFamide/peptide receptors are dispersed in the genome, but almost 90% of *srw* genes are on chromosome V, most in large clusters on the chromosomal arms (Figure 3B). Given the similarities of their gene family evolution to that of the other families, and their common genomic organization, we hypothesize that the *srw* family represents a novel origin of chemoreceptors within the nematode lineage, perhaps serving as environmental peptide receptors.

WormBook.org
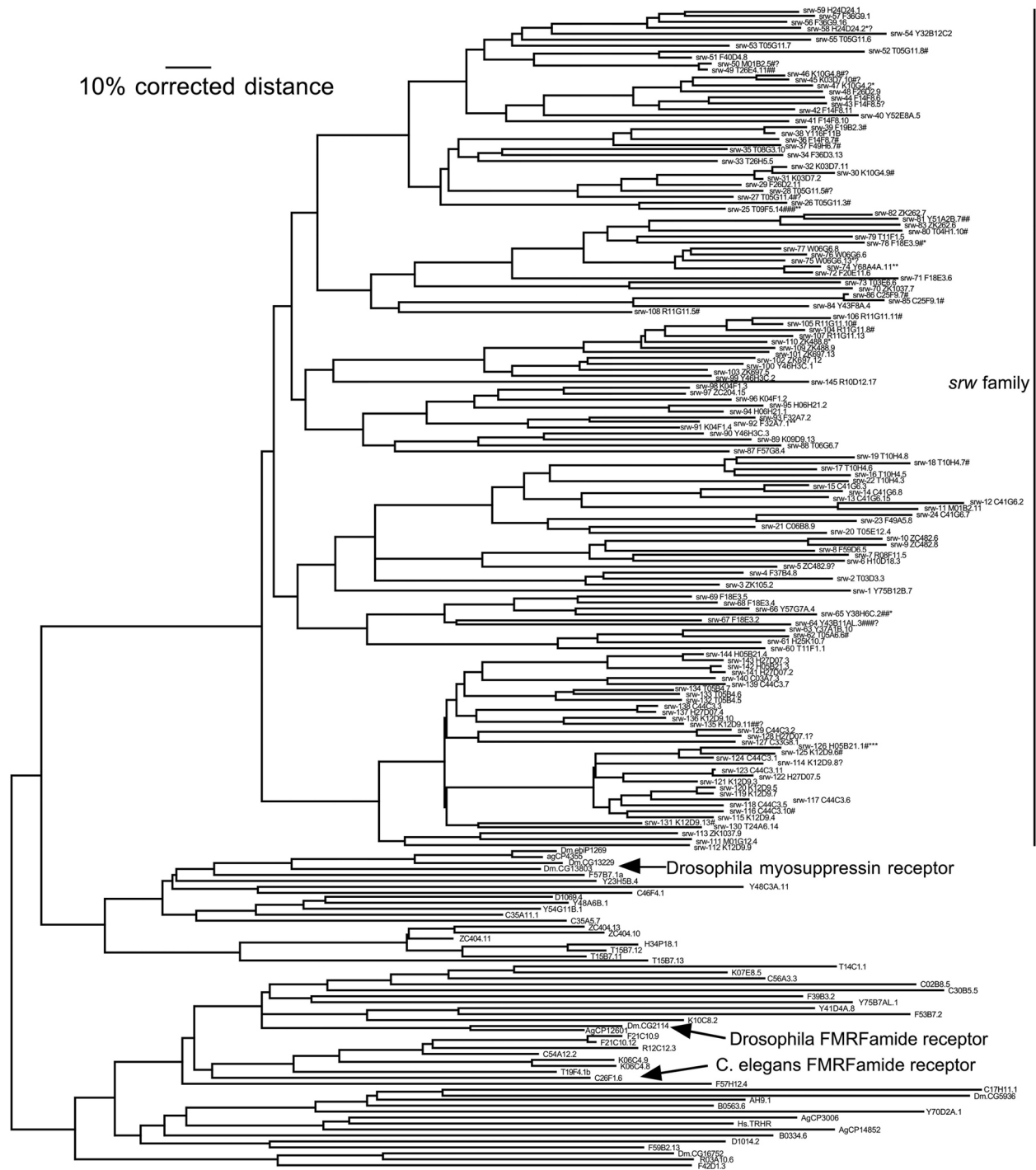
Figure 2. Phylogenetic tree of the srw family, rooted with a group of FMRF-amide and other peptide receptors from nematodes and insects. Large arrows indicate a *Drosophila* FRMFamide receptor (Meeusen et al., 2002), a *Drosophila* myosuppressin receptor (Egerod et al., 2003), and one of the three *C. elegans* FMRFamide receptors (Mertens et al., 2004). See Figure 1 legend for details of methods and symbols. Right click or control-click for larger image.

A all families

B *srh*, *str*, *srw* families
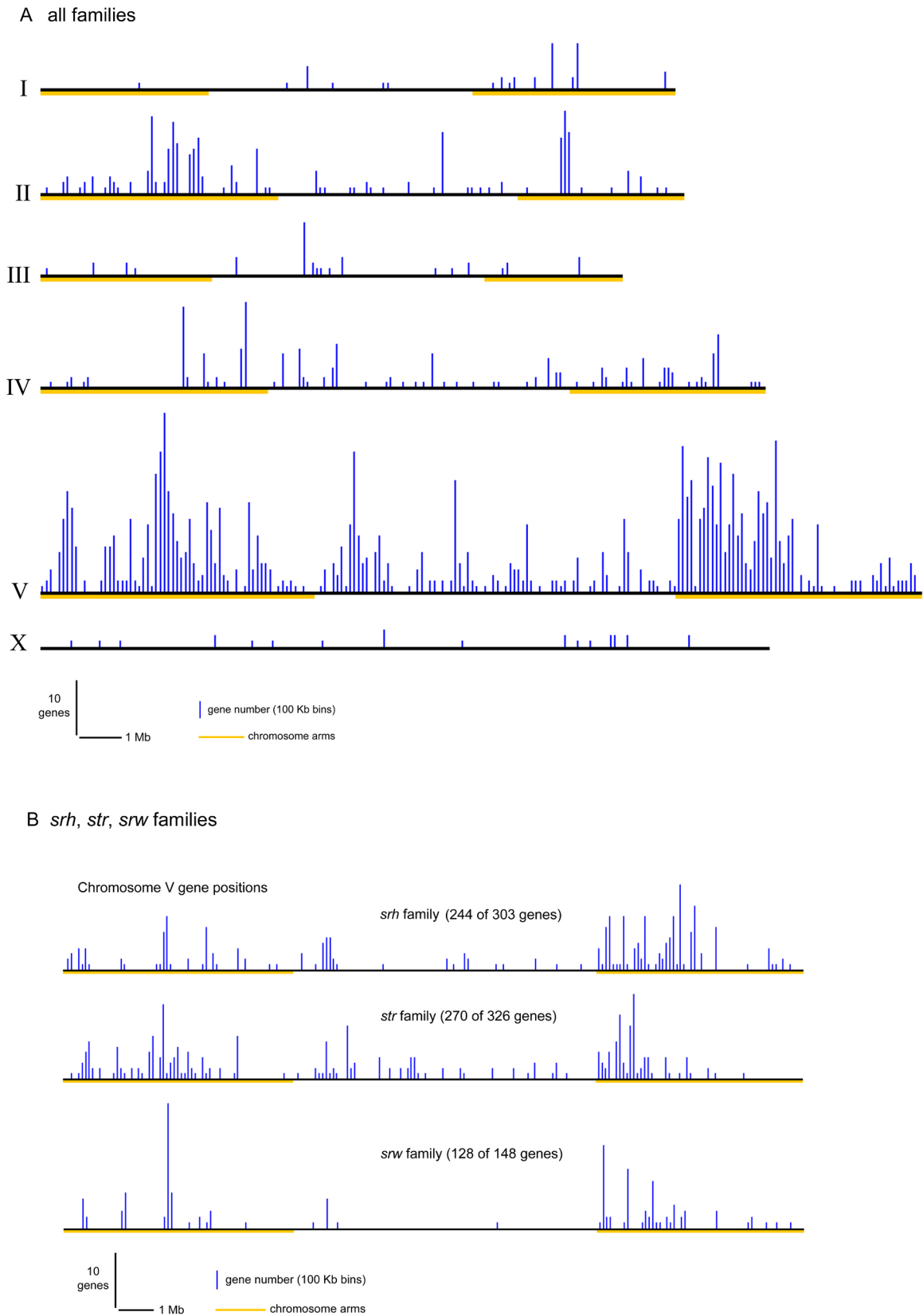
Figure 3. SR genome positions. Panel A shows the positions in the entire genome of all analyzed members of chemoreceptor families (those listed in Table 1). Panel B shows the positions of all members of three specific families on chromosome V.

WormBook.org

3. Gene expansions

An unusual characteristic of the chemoreceptor families is their highly dynamic gene number, reflected within *C. elegans* by frequent duplications and gene loss. Systematic comparisons with *C. briggsae* are only beginning, but evidence thus far indicates that the familiar concept of orthology is largely useless for these families due to their dynamic evolution. Figure 4 shows a phylogenetic tree of two of the three families for which a completed *C. briggsae* annotation is available (Chen et al., 2005). Several large expansions in gene number in *C. elegans* are apparent and only 7 pairs of genes are convincing one-to-one orthologs. Several additional reciprocal best match pairs (marked "a") are substantially more divergent in protein sequence, which probably indicates that they duplicated well before *elegans-briggsae* speciation (followed by loss of one copy in both species) and are not orthologs in the strict sense. Similar dynamics are even more extreme in the *srz* family (Thomas et al., 2005) and the *srh* families (JHT, unpublished).
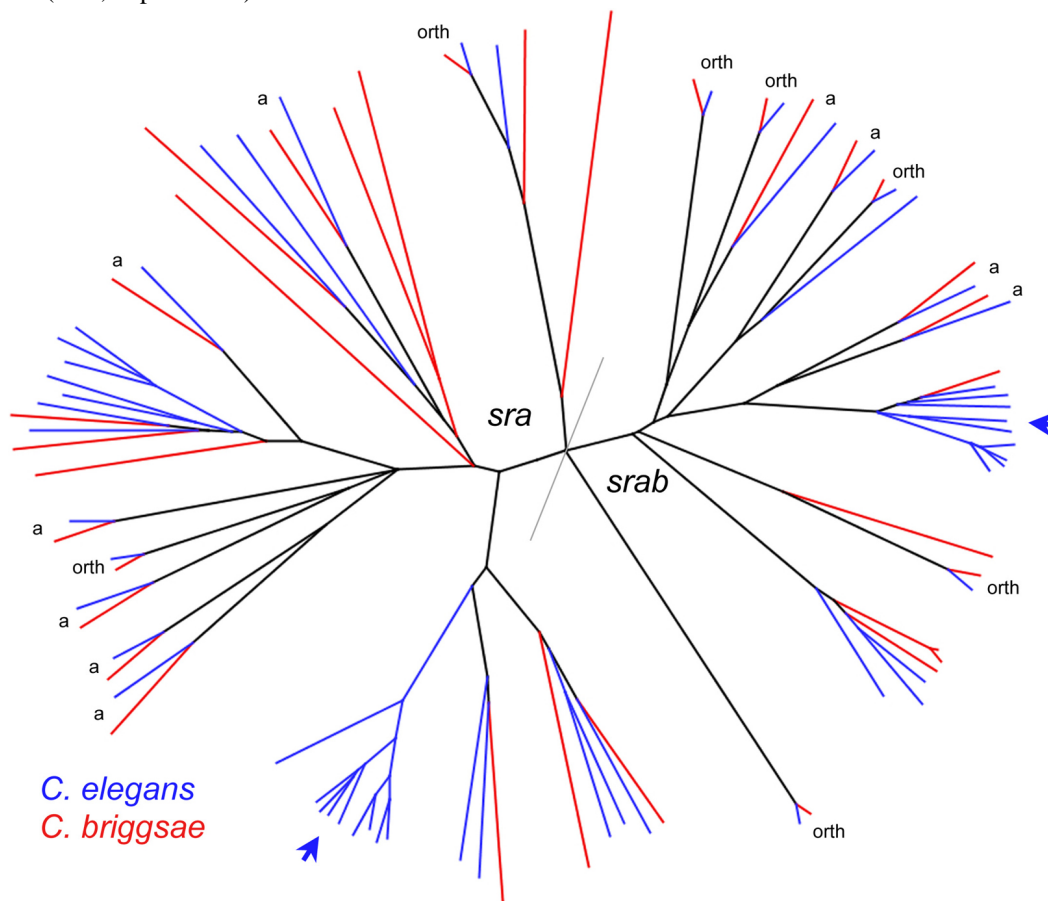
Figure 4. *sra* and *srab* protein tree. Unrooted neighbor-joining tree of all putative functional SRA and SRAB proteins from *C. elegans* (blue) and *C. briggsae* (red). Data are adapted from Chen et al. (2005). Arrows point to *C. elegans* gene expansions. "orth" = putative one-to-one orthologs. "a" = possible paralogs with reciprocal gene loss. The distinction between orthologs and paralogs is based only on protein distance and is somewhat arbitrary, though in approximate agreement with the average divergence of apparent orthologs in other chemoreceptor families.

Much work remains to be completed, but the broad picture is that the chemoreceptor gene families are in a dynamic state, with fixation of new gene duplicates approximately balancing gene loss (reflected in the large number of defective genes, and in the evidence for ultimate deletion of many genes entirely). Over relatively short periods this dynamic is probably near equilibrium, resulting in a stable number of genes. However, over longer periods unknown factors can change the relative rates of gene gain and gene loss, giving rise to the dramatically different gene numbers in *C. elegans* and *C. briggsae*. This pattern is strikingly reminiscent of recent findings in mammalian odorant and vomeronasal gene families, in which gene duplications, gene loss, and lineage specific expansions are common (Del Punta et al., 2000; Glusman et al., 2001; Matsunami et al., 2000). It will be interesting to determine the chemosensory gene distributions for the other *Caenorhabditis* species currently in the genome sequencing pipeline.

4. Gene clusters

The putative chemosensory genes are strongly clustered in the genome (Figure 3). Most families are heavily enriched on chromosome V, and significant enrichment is also found on chromosomes II and IV. Furthermore, the positions of genes on all of these chromosomes are strongly biased toward the chromosome arms (as defined by increased recombination frequency; Barnes et al., 1995). In addition to these broad patterns of gene clustering, there is a strong tendency for closely related genes to be physically close in the genome (Chen et al., 2005; Robertson, 2000; Robertson, 2001). These patterns are likely attributable to patterns of gene duplication in *C. elegans*. New duplications are almost always local, often tandem or inverted repeats with little or no intervening sequence between the duplicates (Katju and Lynch, 2003; Semple and Wolfe, 1999). The current gene positions are consistent with a history of frequent local gene duplication and rare large chromosomal rearrangements, which occasionally break up clusters of local related genes. These large rearrangements are rarely if ever interchromosomal translocations; we suspect that most are inversions.

Detailed analysis of the distribution of *srh* and *str* genes not on chromosome V showed that small interchromosomal transposition events occur occasionally (Robertson, 1998; Robertson, 2000; Robertson, 2001). These events are followed in some cases by a duplication and diversification process in the new genomic location. As with chromosome V, residence on other chromosomal arms seems to favor the duplication process. The mechanisms underlying this pattern are not known, but are likely to involve the higher frequency of complex DNA repeats found on almost every autosomal arm.

5. Function and expression patterns

Analysis of the function of *C. elegans* chemoreceptors has been remarkably resistant to genetic methods. Only *odr-10* has been identified in classical forward genetic screens (Sengupta et al., 1996) and the fraction of genes assigned gross phenotypes in the initial global RNAi screen is close to zero (Kamath et al., 2003). This resistance may result from a combination of the expected narrow phenotype range for the genes (and a relative paucity of genetic screens directed toward this) and genetic redundancy among closely related receptors. A few additional chemoreceptor genes have more limited phenotypic information. Deletion or overexpression of *sra-13* affects olfactory response to isoamyl alcohol and also affects vulval induction, probably as part of a sensory response to starvation (Battu et al., 2003). One sign of hope is that about 25 chemoreceptor genes were identified as affecting fat accumulation in a global RNAi screen testing specifically for this phenotype (Ashrafi et al., 2003). In addition, RNAi knockdown of *str-2* is reported to extend lifespan, though the effect is small (Alcedo and Kenyon, 2004). These RNAi results are uncorroborated by other approaches, but it is possible that environmental chemoperception or internal physiological monitoring by chemoreceptors plays a prominent role in energy resource allocation.

The cellular expression patterns of the putative chemosensory receptor genes are of interest for two reasons. First, their classification as chemosensory receptors is currently best assessed by testing whether expression is predominantly in chemosensory neurons. Second, because of the large set of genes, their patterns of expression provide a remarkable opportunity to study the relationship between cell fate, transcription factors, and *cis* regulatory sequences. Only one gene in the entire SR class has been rigorously shown to function as a chemosensory receptor (Sengupta et al., 1996; Zhang et al., 1997). Are the other ~1,300 genes really chemosensory receptors, or are they a mixture of chemoreceptors and other types of receptors? Transgenic GFP-fusion expression data bearing on this question are summarized in Table 2. Though the coverage of the superfamily is insufficient to draw detailed inferences, a broad picture can be painted. Many genes are expressed only in chemosensory neurons, consistent with a strictly chemosensory function. There are also exceptions to this rule, which might result from real function in other types of neurons and non-neuronal tissue or from transgenic misexpression in ectopic tissues, possibilities that it is currently not possible to distinguish. In addition to GFP fusion data, one study used microarrays to examine gene expression from cell-sorted AWB neurons (Colosimo et al., 2004), a method that is likely to gain importance in the future. SR genes were prominent among the genes identified, including *str-220*, *srd-16*, *srd-23*, *sru-38*, and *srsx-5*. Analysis of SR expression by microarray studies using total worm mRNA is problematic because of their generally low abundance transcript levels; though some genes were assigned to expression mountains (Kim et al., 2001), these assignments should be viewed with some skepticism because absolute expression levels are typically at or close to noise levels.

A final method that shows promise in assigning expression patterns to SR genes is identification and analysis of specific promoter sequence motifs that may be binding sites for transcription factors. Using this approach, one study identified a well-defined E-box motif in the promoters of about 200 SR genes, mostly from the *srh* and *sri*

9

families (McCarroll et al., 2005). It was further shown that the E-box is correlated with expression in the ADL chemosensory neuron, and that it was sufficient to direct expression of *odr-10* in ADL.

6. Acknowledgements

7. References

Alcedo, J., and Kenyon, C. (2004). Regulation of *C. elegans* longevity by specific gustatory and olfactory neurons. Neuron *41*, 45–55. Abstract Article

Ashrafi, K., Chang, F.Y., Watts, J.L., Fraser, A.G., Kamath, R.S., Ahringer, J., and Ruvkun, G. (2003). Genome-wide RNAi analysis of *Caenorhabditis elegans fat* regulatory genes. Nature *421*, 268–272. Abstract Article

Bargmann, C.I., Mori, I. (1997). Chemotaxis and thermotaxis. In *C. elegans II*, D.L. Riddle, Blumenthal, T., Meyer, B.J., Priess, J.R., eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 717–737.

Barnes, T.M., Kohara, Y., Coulson, A., and Hekimi, S. (1995). Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. Genetics *141*, 159–179. Abstract

Battu, G., Hoier, E.F., and Hajnal, A. (2003). The *C. elegans* G-protein-coupled receptor SRA-13 inhibits RAS/MAPK signalling during olfaction and vulval development. Development *130*, 2567–2577. Abstract Article

Buck, L., and Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell *65*, 175–187. Abstract Article

Chen, N., Pai, S., Zhao, Z., Mah, A., Newbury, R., Johnsen, R.C., Altun, Z., Moerman, D.G., Baillie, D.L., and Stein, L.D. (2005). Identification of a nematode chemosensory gene family. Proc. Natl. Acad. Sci. USA *102*, 146–151. Abstract Article

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. *31*, 3497–3500. Abstract Article

Choi, S.S., and Lahn, B.T. (2003). Adaptive evolution of MRG, a neuron-specific gene family implicated in nociception. Genome Res. *13*, 2252–2259. Abstract Article

Coghlan, A., and Wolfe, K.H. (2004). Origins of recently gained introns in *Caenorhabditis*. Proc. Natl. Acad. Sci. USA *101*, 11362–11367. Abstract Article

Colosimo, M.E., Brown, A., Mukhopadhyay, S., Gabel, C., Lanjuin, A.E., Samuel, A.D., and Sengupta, P. (2004). Identification of thermosensory and olfactory neuron-specific genes via expression profiling of single neuron types. Curr. Biol. *14*, 2245–2251. Abstract Article

Del Punta, K., Rothman, A., Rodriguez, I., and Mombaerts, P. (2000). Sequence diversity and genomic organization of vomeronasal receptor genes in the mouse. Genome Res. *10*, 1958–1967. Abstract Article

Egerod, K., Reynisson, E., Hauser, F., Cazzamali, G., Williamson, M., and Grimmelikhuijzen, C.J. (2003). Molecular cloning and functional expression of the first two specific insect myosuppressin receptors. Proc. Natl. Acad. Sci. USA *100*, 9808–9813. Abstract Article

Glusman, G., Yanai, I., Rubin, I., and Lancet, D. (2001). The complete human olfactory subgenome. Genome Res. *11*, 685–702. Abstract Article

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA *89*, 10915–10919. Abstract

Hilliard, M.A., Bergamasco, C., Arbucci, S., Plasterk, R.H., and Bazzicalupo, P. (2004). Worms taste bitter: ASH neurons, QUI-1, GPA-3 and ODR-3 mediate quinine avoidance in*Caenorhabditis elegans*. EMBO J. *23*, 1101–1111. Abstract Article

Jansen, G., Thijssen, K.L., Werner, P., van der Horst, M., Hazendonk, E., and Plasterk, R.H. (1999). The complete family of genes encoding G proteins of *Caenorhabditis elegans*. Nat. Genet. *21*, 414–419. Abstract Article

Jansen, G., Weinkove, D., and Plasterk, R.H. (2002). The G-protein gamma subunit *gpc-1* of the nematode *C.elegans* is involved in taste adaptation. EMBO J. *21*, 986–994. Abstract Article

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature *421*, 231–237. Abstract Article

Katju, V., and Lynch, M. (2003). The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. Genetics *165*, 1793–1803. Abstract

Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N., and Davidson, G.S. (2001). A gene expression map for *Caenorhabditis elegans*. Science *293*, 2087–2092. Abstract Article

Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., and Fitch, D.H. (2004).*Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. USA *101*, 9003–9008. Abstract Article

Matsunami, H., Montmayeur, J.P., and Buck, L.B. (2000). A family of candidate taste receptors in human and mouse. Nature *404*, 601–604. Abstract Article

McCarroll, S.A., Li, H., and Bargmann, C. (2005). Identification of transcriptional regulatory elements in chemosensory receptor genes by probabilistic segmentation. Curr. Biol. *15*, 347–352. Abstract Article

Meeusen, T., Mertens, I., Clynen, E., Baggerman, G., Nichols, R., Nachman, R. J., Huybrechts, R., De Loof, A., and Schoofs, L. (2002). Identification in *Drosophila melanogaster* of the invertebrate G protein-coupled FMRFamide receptor. Proc. Natl. Acad. Sci. USA *99*, 15363–15368. Abstract Article

Mertens, I., Vandingenen, A., Meeusen, T., Janssen, T., Luyten, W., Nachman, R. J., De Loof, A., and Schoofs, L. (2004). Functional characterization of the putative orphan neuropeptide G-protein coupled receptor C26F1.6 in *Caenorhabditis elegans*. FEBS Lett. *573*, 55–60. Abstract Article

Miller, M.A., Cutter, A.D., Yamamoto, I., Ward, S., and Greenstein, D. (2004). Clustered organization of reproductive genes in the *C. elegans* genome. Curr. Biol. *14*, 1284–1290. Abstract Article

Mori, I. (1999). Genetics of chemotaxis and thermotaxis in the nematode *Caenorhabditis elegans*. Annu. Rev. Genet. *33*, 399–422. Abstract Article

Mounsey, A., Bauer, P., and Hope, I.A. (2002). Evidence suggesting that a fifth of annotated*Caenorhabditis elegans* genes may be pseudogenes. Genome Res. *12*, 770–775. Abstract Article

Roayaie, K., Crump, J.G., Sagasti, A., and Bargmann, C.I. (1998). The Gα protein ODR-3 mediates olfactory and nociceptive function and controls cilium morphogenesis in *C. elegans* olfactory neurons. Neuron *20*, 55–67. Abstract Article

Robertson, H.M. (1998). Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res. *8*, 449–463. Abstract Abstract

WormBook.org

Robertson, H.M. (2000). The large*srh* family of chemoreceptor genes in*Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. Genome Res. *10*, 192–203. Abstract Article

Robertson, H.M. (2001). Updating the*str* and *srj* (*stl*) families of chemoreceptors in*Caenorhabditis* nematodes reveals frequent gene movement within and between chromosomes. Chem. Senses *26*, 151–159. Abstract Article

Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics *18*, 502–504. Abstract Article

Semple, C., and Wolfe, K.H. (1999). Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. J. Mol. Evol. *48*, 555–564. Abstract

Sengupta, P. (1997). Cellular and molecular analyses of olfactory behavior in *C. elegans*. Semin. Cell Dev. Biol. *8*, 153–161. Abstract Article

Sengupta, P., Chou, J.H., and Bargmann, C.I. (1996). *odr-10* encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. Cell *84*, 899–909. Abstract Article

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. *1*, E45. Abstract Article

Stewart, M.K., Clark, N.L., Merrihew, G., Galloway, E., and Thomas, J.H. (2005). High genetic diversity in the chemoreceptor superfamily of *C. elegans*. Genetics *169*, 1985–1996. Article

Swanson, W.J., and Vacquier, V.D. (2002). The rapid evolution of reproductive proteins. Nat. Rev. Genet. *3*, 137–144. Abstract Article

Swofford, D.L. (1998). *PAUP∗: Phylogenetic analysis using parsimony and other methods, Version 4* (New York, NY: Sinauer Press).

Thomas, J.H., Kelley, J.L., Robertson, H.M., Ly, K., and Swanson, W.J. (2005). Adaptive evolution in the SRZ chemoreceptor families of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Proc. Natl. Acad. Sci. USA *102*, 4476–4481. Abstract Article

Thomas, J.H. (2006). Analysis of Homologous Gene Clusters in *C. elegans* Reveals Striking Regional Cluster Domains. Genetics. In press

Troemel, E.R., Chou, J.H., Dwyer, N.D., Colbert, H.A., and Bargmann, C.I. (1995). Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell *83*, 207–218. Abstract Article

Wenick, A.S., and Hobert, O. (2004). Genomic *cis*-regulatory architecture and *trans*-acting regulators of a single interneuron-specific gene battery in *C. elegans*. Dev. Cell *6*, 757–770. Abstract Article

Zhang, Y., Chou, J.H., Bradley, J., Bargmann, C.I., and Zinn, K. (1997). The *Caenorhabditis elegans* seven-transmembrane protein ODR-10 functions as an odorant receptor in mammalian cells. Proc. Natl. Acad. Sci. USA *94*, 12162–12167. Abstract Article