Report

Current Biology

Nematode nuclear receptors as integrators of sensory information

Highlights

- Nuclear hormone receptor expansion correlates with free-living nematode lifestyle
- GPCR and insulin genes have co-expanded with NHRs in free-living nematode species
- *C. elegans* sensory neurons show enriched expression of NHR, GPCR, and insulin genes
- NHRs may integrate sensory or internal cues to regulate GPCR and insulin expression

Authors

Surojit Sural, Oliver Hobert

Correspondence

ss5950@columbia.edu (S.S.),
or38@columbia.edu (O.H.)

In brief

Sural and Hobert put forward a hypothesis about the function of the vastly expanded nuclear hormone receptor family in nematodes, proposing that they present an adaptation to complex sensory environments and act as sensory receptors to provide a feedback system for the modulation of sensory and metabolic responses.

Sural & Hobert, 2021, Current Biology *31*, 4361–4366
October 11, 2021
https://doi.org/10.1016/j.cub.2021.07.019

CellPress

Current Biology

CellPress

Report

Nematode nuclear receptors as integrators of sensory information

Surojit Sural[1,*] and Oliver Hobert[1,2,*]
[1]Department of Biological Sciences, Howard Hughes Medical Institute, Columbia University, New York, NY, USA
[2]Lead contact
*Correspondence: ss5950@columbia.edu (S.S.), or38@columbia.edu (O.H.)
https://doi.org/10.1016/j.cub.2021.07.019

SUMMARY

More than 20 years ago, the sequencing of the genome of the nematode *Caenorhabditis elegans* uncovered a still unparalleled abundance of C4-zinc finger orphan nuclear hormone receptors, encoded by 267 different *nhr* genes.[1,2] Only less than 20 of them are conserved throughout the animal kingdom; all the remaining genes are the results of an expansion of the HNF4-subtype of nuclear receptors.[3,4] Strikingly, even though most of the receptors contain predicted ligand binding domains, no ligand has since been identified for any of the non-conserved, *C. elegans*-expanded *nhr* genes. Based on an analysis of more than 100 nematode genome sequences, as well as the mining of recently established nervous system-wide gene expression patterns, we propose here that *nhr* family expansion is a manifestation of adaptation of free-living nematodes to complex sensory environments and that NHR proteins may function as sensory receptors for external or internal sensory cues to modulate the animal's sensory responses to environmental cues as well as its internal metabolic state.

RESULTS AND DISCUSSION

The availability of more than 100 nematode genome sequences[5,6] prompted us to undertake a comparative analysis of *nhr*-encoding genes. According to the latest genome sequence release, the *C. elegans* genome encodes 267 C4-zinc finger nuclear hormone receptors (Figure 1A; Data S1A). We find that the expansion of *nhr* genes greatly varies in different nematodes species; while some species encode more than 300 nhr genes, others encode as little as 20 *nhr* genes (Figure 1A; Data S1A). Another gene family that varies greatly in size in nematodes are olfactory-type G protein-coupled serpentine receptors (GPCRs).[7,8] A systematic survey of GPCR-encoding genes shows a wide variability of gene number across nematodes, up to more than 2,000 genes in some nematodes down to as little as 60 *gpcr* genes in others (Figure 1A; Data S1A).[9] Chemosensory GPCRs display high sequence divergence and are under strong positive selection in wild *C. elegans* isolates, and we observe these patterns also for *nhr* genes (Figure S1A; Table S1).[10]

Strikingly, we find a tight correlation between the expansion of *nhr* genes and GPCR-encoding genes (Figure 1A), a correlation that cannot be explained by the evolutionary relationship of these nematode species or their genome sequence assembly quality (Figures S1B and S1C). While other large nematode transcription factor families (e.g., homeobox genes or C2H2 zinc fingers) also vary in size (Data S1A), the size variation for these transcription factor families does not correlate with the expansion of *nhr* or *gpcr* genes (Figure S1D). Intriguingly, the co-expansion of *nhr* and GPCR-encoding genes correlates with nematode lifestyle, such that free-living species display more such expansions than parasitic nematodes (Figures 1A and 1B). Free-living nematodes experience a much richer external sensory environment compared to parasitic nematodes, a notion also reflected in a more simplified sensory anatomy of parasitic nematodes.[11,12] Parasitic lifestyle has evolved independently in the different Nematoda clades, suggesting that the common ancestor of all nematode species was free living.[13] To determine whether such free-living ancestral species had an already expanded family of *nhr* genes, we compared the similarity between the C4-zinc finger DNA binding domains of all *nhr* genes in two distantly related free-living nematode species, *Caenorhabditis elegans* (clade V) and *Acrobeloides nanus* (clade IV). While both these free-living species have undergone some independent expansions of the *nhr* gene family, many orthologous relationships between expanded family members are apparent, indicating that *nhr* genes were independently lost after adapting a parasitic lifestyle (Figures 2 and S2; Data S1A).

The correlation of *nhr* and *gpcr* diversification both across and within species and their association with the complexity of their sensory environment suggest that *nhr* genes may play a role in some aspect of sensory biology. This notion is further strengthened by a recently established scRNA-based gene-expression atlas of the entire *C. elegans* nervous system, which shows a striking enrichment of *nhr* gene expression in sensory neurons.[14] Many sensory neurons co-express several dozens, some almost 100 *nhr* genes (Figures 1C and 1D). Some *nhr* genes are uniquely expressed in single sensory neuron classes, but 46 *nhr*s are expressed in more than 20 sensory neuron classes. Other large transcription factor families do not show such sensory neuron enrichment (Figures 1D and S1F).[14]

Previous reporter analysis,[8,15] as well as the new transcriptome atlas of the entire *C. elegans* nervous system,[14] show

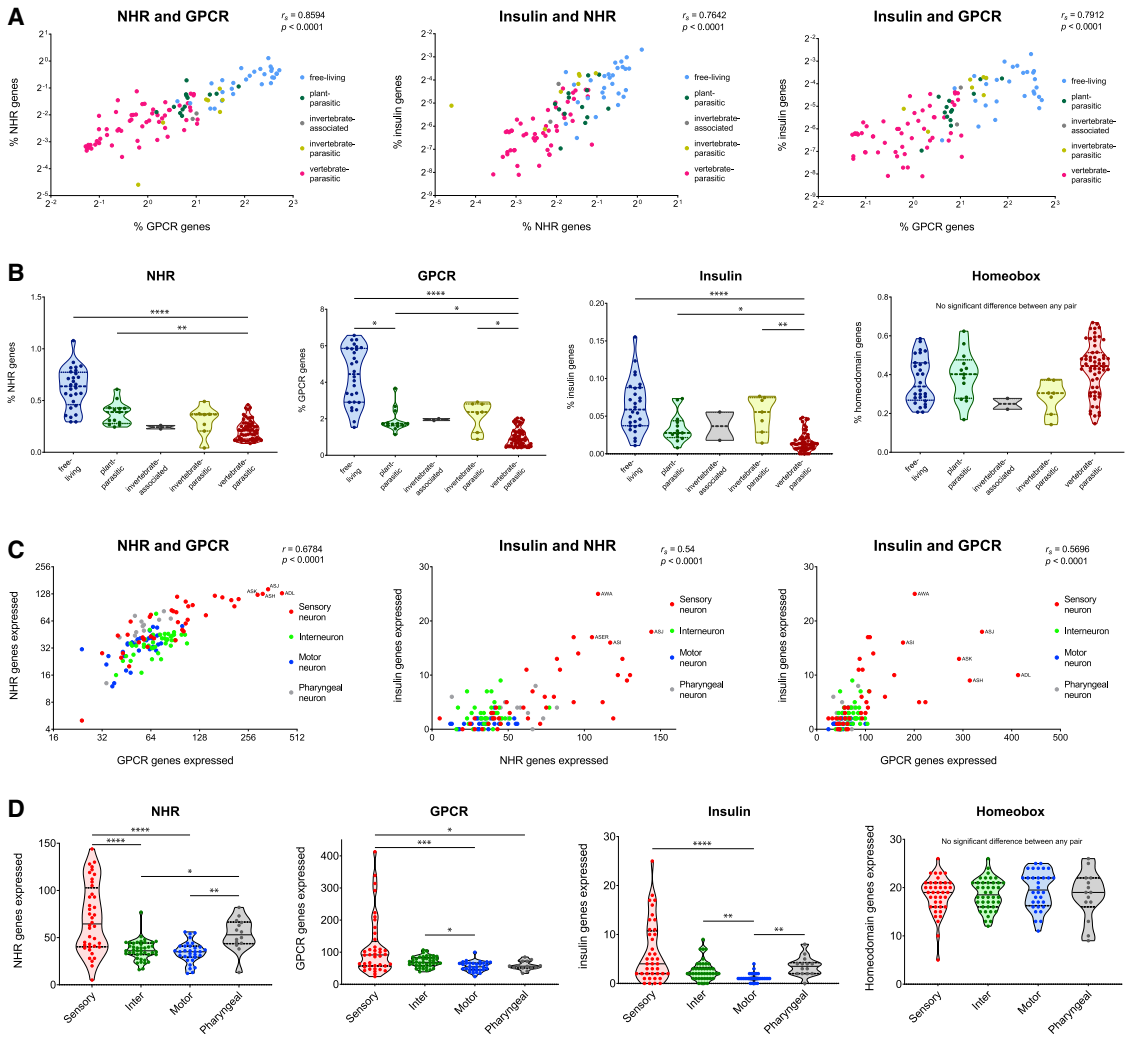

Figure 1. Strongly correlated gene expansions and expression patterns of NHR, GPCR, and insulin-like peptide genes

(A) Scatterplots for proportion of NHR, GPCR, and insulin-like peptide-coding genes in the genomes of 109 nematode species. Colors indicate lifestyle of the species.

(B) Proportions of NHR, GPCR, insulin-like peptide, and homeobox gene families in the genomes of 109 nematode species divided according to lifestyle. *p < 0.05, **p < 0.01, and ****p < 0.0001, Dunn's multiple comparisons test performed after Kruskal-Wallis one-way ANOVA.

(C) Scatterplots for number of NHR, GPCR, and insulin-like peptide genes expressed in 128 distinct *C. elegans* neuron classes as identified from the CeNGEN scRNA-based gene-expression atlas. Colors indicate functional type of each neuron class. Spearman correlation values (r_s) are shown.

(D) Number of NHR, GPCR, insulin-like peptide, and homeobox genes expressed in 128 distinct *C. elegans* neuron classes as identified from the CeNGEN scRNA-based gene-expression atlas. Neurons are divided according to their functional type. *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001, Dunn's multiple comparisons test performed after Kruskal-Wallis one-way ANOVA.

See also Figure S1, Table S1, and Data S1.

that many individual sensory neurons coexpress >100 GPCR-encoding genes (Figures 1C and 1D). Intriguingly, there is again a correlation of the number of *nhr* genes expressed by a sensory neuron with the number of GPCR-encoding genes (Figure 1C). This is most evident in the amphid and phasmid sensory neurons. For example, the ASH, ADL, and ASJ neurons express more than 300 GPCR-encoding genes and more than 120 *nhr* genes (Figure 1C).

It has previously been noted that many *Caenorhabditis* species also show a very large complement of other specific gene families, including C-type lectins, F-box proteins, and kinases.[16–18] We find that the expansion of two of these families, C-type lectins and F-box proteins, again correlates with free-living lifestyle (Figure S1E; Data S1A). However, these gene families do not show a similar enrichment in terms of sensory neuron expression (Figure S1F), underscoring the notion that the co-expansion of *gpcr* and *nhr* genes and their correlated co-expression in sensory neurons speaks to a specific functional association of these two gene families in nematodes.

Based on these findings, we propose the following four hypotheses. (1) The NHR family expanded in free-living nematodes as a reflection of a complex sensory environment and, hence,

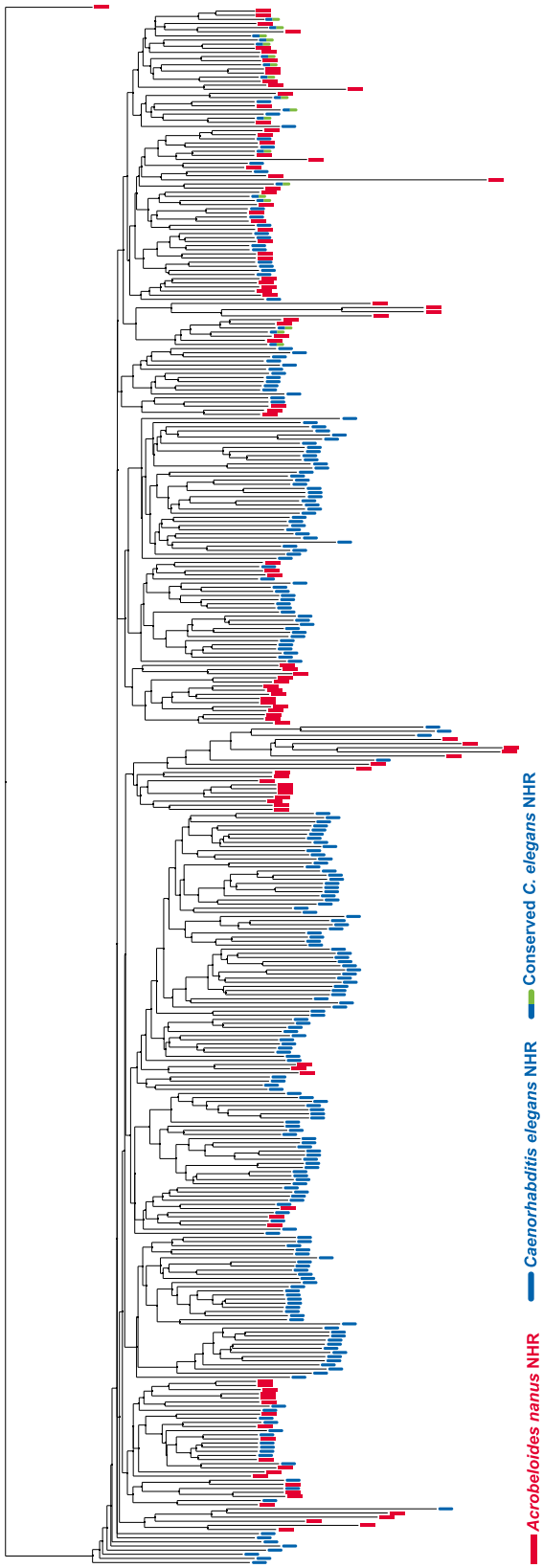

Figure 2. Phylogenetic tree for NHR genes in two distantly related free-living nematode species

The phylogenetic tree shows similarity between C4-zinc finger DNA binding domain sequences of all NHR genes in the two nematode species *Caenorhabditis elegans* (clade V) and *Acrobeloides nanus* (clade IV). *C. elegans* (blue) and *A. nanus* (red) NHR genes are shown with different colors. *C. elegans* NHR genes that show broad conservation across evolutionary taxa are also labeled in green.[3] Gene names are shown in Figure S2.
See also Figure S2.

complex food sources. Decreases in the complexity and variability of sensory environments and food sources that came with the acquisition of parasitic lifestyles may have led to a contraction of the *nhr* and *gpcr* gene families. (2) NHR proteins are sensors for either external environmental food signals that diffuse through the cuticle or are sensors of metabolites from ingested bacterial food sources. (3) The striking enrichment of *nhr* gene expression in sensory neurons suggests that *nhrs* have a role in altering sensory responses of the worm. (4) We propose that NHR transcription factors may modulate the expression of GCPR-encoding genes in sensory neurons. In support of this notion, there are many correlations in the cellular specificity of expression patterns of NHR- and GPCR-encoding genes (Figure 3; Table S2), which are observed between many NHRs and their predicted *gpcr* targets (Figure S1G; Data S2). This similarity in cellular expression is higher than what would be expected by chance and is not observed between other large gene families (Figures S1H and S1I). The *nhr* and *gpcr* genes with the most strongly correlated cellular expression patterns are not located in adjoining genomic regions, hence excluding the possibility that these genes might be co-regulated via the same *cis-regulatory* elements (Table S3). In addition, the correlated expression of *nhr*s and *gpcr*s is not restricted only to genes in the hyperdivergent regions of the *C. elegans* genome (Figures S3A–S3G).[19] However, we find that the sensory neuron-enriched expression of *gpcr* genes more strongly resembles the expression pattern of *C. elegans*-expanded *nhr*s compared to that of evolutionarily conserved *nhr*s (Figures S3H–S3M).[3]

Though correlated expression of genes does not necessarily imply an underlying functional relationship, it is consistent with our hypothesis that individual nuclear receptors potentially modulate the expression of GPCR-encoding genes to dictate an animal's sensory experience in a particular environment. The expression of GPCR-encoding genes is known to be responsive to food signals,[15,20] and we propose that, by binding to food signals or food-derived signals and ensuing activation, NHRs are the conveyors of such nutrient signals (Figure 4). Such regulatory relationship may ensure sensory feedbacks (Figure 4). For example, ingestion of a metabolite considered to be of good nutritional value may modulate the expression of sensory receptors to augment the attraction of animals to a specific bacterial food source. Such a feedback system, either positive or negative in nature, may have a fundamental impact on animal navigation and feeding behavior. Indeed, previous studies have demonstrated that, when *C. elegans* first encounters pathogenic bacteria, its olfactory preference changes from attraction to aversion within a few hours of exposure to the new food source, and this is associated with transcriptional changes for a large number of GPCR-encoding genes.[21,22] Experimental validation of potential regulatory relationships between *nhr*s and

Figure 3. Heatmap for Jaccard similarity between expression of NHR and GPCR genes

Heatmap showing Jaccard similarity indices for nervous system-wide binary expression patterns of *nhr* and *gpcr* gene pairs as identified from the CeNGEN scRNA-based gene-expression atlas. Clustering of rows and columns in the heatmap is based on Euclidean distance.
See also Figures S1 and S3, Tables S2 and S3, and Data S2.

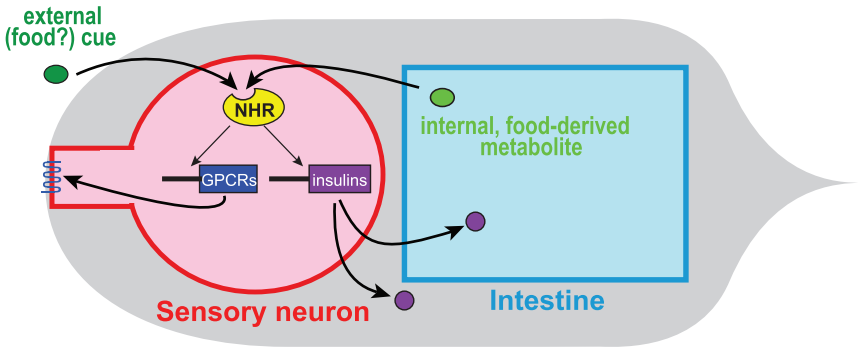

Figure 4. Hypothesis based on our genome sequence and expression pattern analysis

External cues or intestinally derived metabolites from ingested food that are indicative of specific food sources (green circles) activate a cognate, specific member of the NHR family, which in turn modulates (up- or downregulates) the expression of specific GPCRs and insulin-encoding genes. Up- or downregulation of GPCRs in sensory neurons may result in an altered (attractive or repulsive) behavioral response of the animals in relation to a specific cue, thereby generating a reinforcing feedback loop. In parallel, insulin-like peptides (purple circles) secreted from sensory neurons may act to modulate the internal metabolic state of the animal and/or could regulate developmental arrest (e.g., dauer formation) in response to specific sensory cues.

*gpcr*s will require the identification of ligands for NHRs, followed by an assessment whether ligand addition/removal indeed results in changes in *gpcr* gene expression. Regulatory relationships may not necessarily involve on/off switches, but quantitative modulation of expression levels in order to fine-tune sensory responses.

The potential regulation of GPCR-type sensory receptors may not be the only function of NHR proteins in sensory neurons. Another gene family that has undergone an expansion in the *C. elegans* genome compared to that of other invertebrate and vertebrate genomes is the insulin gene family, of which there are 40 members in *C. elegans*.[23] Strikingly, we find that this gene family expansion also correlates with free living versus parasitic lifestyle, and it correlates with the expansion of both the *nhr* and *gpcr* gene family (Figure 1A). Moreover, mining of the scRNA expression atlas[14] reveals that insulin gene expression is also highly biased toward sensory neurons (Figures 1C and 1D). We propose that NHR transcription factors may also control the expression of insulin genes in sensory neurons in response to metabolite signals (Figure 4). Consistent with this notion, the expression of most insulin genes is controlled by nutrient availability.[24,25] Through the regulation of insulin genes, NHRs may couple food signals to changes in internal metabolic states and thus control development arrest in nematodes via modulating insulin-dependent dauer formation.

Testing the above hypotheses will require a systematic analysis of GPCR and insulin gene expression in a number of different physiological food environments and will require biochemical and metabolomic efforts to identify ligands for NHR proteins. The enormous structural diversity of potential NHR ligands, and their connection to nutrient signals, is already greatly appreciated.[26,27] These approaches have the potential to yield novel strategies for the control of parasitic nematodes, which continue to represent a major impediment to human and livestock health as well as crop production.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.07.019.

ACKNOWLEDGMENTS

We thank Erich Schwarz, Molly B. Reilly, and Lori Glenwinkel for invaluable help with genome, scRNA, and TargetOrtho analysis, respectively. O.H. is an Investigator with the Howard Hughes Medical Institute, which funded this work.

AUTHOR CONTRIBUTIONS

S.S. and O.H. conceived this project, S.S. conducted all the computational analysis, and S.S. and O.H. interpreted results and co-wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 9, 2021
Revised: June 11, 2021
Accepted: July 8, 2021
Published: August 3, 2021

REFERENCES

1. The C. elegans Sequencing Consortium (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012–2018.
2. Sluder, A.E., Mathews, S.W., Hough, D., Yin, V.P., and Maina, C.V. (1999). The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. Genome Res. *9*, 103–120.

3. Bodofsky, S., Koitz, F., and Wightman, B. (2017). Conserved and exapted functions of nuclear receptors in animal development. Nucl. Receptor Res. *4*, 101305.
4. Robinson-Rechavi, M., Maina, C.V., Gissendanner, C.R., Laudet, V., and Sluder, A. (2005). Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. J. Mol. Evol. *60*, 577–586.
5. Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite - a comprehensive resource for helminth genomics. Mol. Biochem. Parasitol. *215*, 2–10.
6. Howe, K.L., Bolt, B.J., Cain, S., Chan, J., Chen, W.J., Davis, P., Done, J., Down, T., Gao, S., Grove, C., et al. (2016). WormBase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. *44* (D1), D774–D780.
7. Bargmann, C.I. (2006). Comparative chemosensation from receptors to ecology. Nature *444*, 295–301.
8. 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.
9. Wheeler, N.J., Heimark, Z.W., Airs, P.M., Mann, A., Bartholomay, L.C., and Zamanian, M. (2020). Genetic and functional diversification of chemosensory pathway receptors in mosquito-borne filarial nematodes. PLoS Biol. *18*, e3000723.
10. Ma, F., Lau, C.Y., and Zheng, C. (2021). Large genetic diversity and strong positive selection in F-box and GPCR genes among the wild isolates of Caenorhabditis elegans. Genome Biol. Evol. *13*, evab048.
11. Goater, T.M., Goater, C.P., and Esch, G.W. (2013). Nematoda: the roundworms. In Parasitism (Cambridge University Press).
12. Fraser, L.M., Madriz, R.I., Srinivasan, D., Zamanian, M., Bartholomay, L.C., and Kimber, M.J. (2018). Chemosensory structure and function in the filarial nematode, Brugia malayi. bioRxiv. https://doi.org/10.1101/427229.
13. Blaxter, M., and Koutsovoulos, G. (2015). The evolution of parasitism in Nematoda. Parasitology *142* (*Suppl 1*), S26–S39.
14. Taylor, S.R., Santpere, G., Weinreb, A., Barrett, A., Reilly, M.B., Xu, C., Varol, E., Oikonomou, P., Glenwinkel, L., McWhirter, R., et al. (2021). Molecular topography of an entire nervous system. Cell. Published online July 2, 2021. https://doi.org/10.1016/j.cell.2021.06.023.
15. Vidal, B., Aghayeva, U., Sun, H., Wang, C., Glenwinkel, L., Bayer, E.A., and Hobert, O. (2018). An atlas of Caenorhabditis elegans chemoreceptor expression. PLoS Biol. *16*, e2004218.
16. Schwarz, E.M. (2005). Genomic classification of protein-coding gene families. WormBook. Published online September 23, 2005. https://doi.org/10.1895/wormbook.1.29.1.
17. Thomas, J.H. (2006). Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. Genome Res. *16*, 1017–1030.
18. Bauters, L., Naalden, D., and Gheysen, G. (2017). The distribution of lectins across the phylum nematoda: a genome-wide search. Int. J. Mol. Sci. Published online January 4, 2017. https://doi.org/10.3390/ijms18010091.
19. Lee, D., Zdraljevic, S., Stevens, L., Wang, Y., Tanny, R.E., Crombie, T.A., Cook, D.E., Webster, A.K., Chirakar, R., Baugh, L.R., et al. (2021). Balancing selection maintains hyper-divergent haplotypes in Caenorhabditis elegans. Nat. Ecol. Evol. *5*, 794–807.
20. Peckol, E.L., Troemel, E.R., and Bargmann, C.I. (2001). Sensory experience and sensory activity regulate chemosensory receptor gene expression in Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA *98*, 11032–11038.
21. Zhang, Y., Lu, H., and Bargmann, C.I. (2005). Pathogenic bacteria induce aversive olfactory learning in Caenorhabditis elegans. Nature *438*, 179–184.
22. Moore, R.S., Kaletsky, R., and Murphy, C.T. (2019). Piwi/PRG-1 Argonaute and TGF-beta mediate transgenerational learned pathogenic avoidance. Cell *177*, 1827–1841.
23. Duret, L., Guex, N., Peitsch, M.C., and Bairoch, A. (1998). New insulin-like proteins with atypical disulfide bond pattern characterized in Caenorhabditis elegans by comparative sequence analysis and homology modeling. Genome Res. *8*, 348–353.
24. Chen, Y., and Baugh, L.R. (2014). Ins-4 and daf-28 function redundantly to regulate C. elegans L1 arrest. Dev. Biol. *394*, 314–326.
25. Ritter, A.D., Shen, Y., Fuxman Bass, J., Jeyaraj, S., Deplancke, B., Mukhopadhyay, A., Xu, J., Driscoll, M., Tissenbaum, H.A., and Walhout, A.J. (2013). Complex expression dynamics and robustness in C. elegans insulin networks. Genome Res. *23*, 954–965.
26. Sladek, F.M. (2011). What are nuclear receptor ligands? Mol. Cell. Endocrinol. *334*, 3–13.
27. Tao, L.J., Seo, D.E., Jackson, B., Ivanova, N.B., and Santori, F.R. (2020). Nuclear hormone receptors and their ligands: metabolites in control of transcription. Cells *9*, 2606.
28. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.
29. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.
30. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.
31. Glenwinkel, L., Taylor, S.R., Langebeck-Jensen, K., Pereira, L., Reilly, M.B., Basavaraju, M., Rafi, I., Yemini, E., Pocock, R., Sestan, N., et al. (2021). In silico analysis of the transcriptional regulatory logic of neuronal identity specification throughout the *C. elegans* nervous system. eLife *10*, e64906.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Code for analysis of genomic and scRNA data	This paper; Zenodo	https://doi.org/10.5281/zenodo.5087531
Software and algorithms		
GraphPad Prism 9	GraphPad Software	https://www.graphpad.com
R version 4.0.4	R Project	https://www.r-project.org
ComplexHeatmap Package	[28]	https://github.com/jokergoo/ComplexHeatmap
Clustal Omega	[29]	https://www.ebi.ac.uk/Tools/msa/clustalo
Jalview version 2.11.1.4	[30]	https://www.jalview.org
TargetOrtho2	[31]	https://github.com/loriglenwinkel/TargetOrtho2.0
Other		
Wormbase Parasite	[5]	http://parasite.wormbase.org/index.html
C. elegans Neuronal Gene Expression Map & Network (CeNGEN) scRNA atlas	[14]	http://www.cengen.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources used in this study should be directed to and will be fulfilled by the lead contact, Oliver Hobert (or38@columbia.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data. The source identifiers for these datasets are listed in the key resources table.
- All original code will be deposited at Zenodo and will be made publicly available as of the date of publication. DOIs will be listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Gene numbers for different gene families

Number of genes for each gene family described in this study was obtained from WormBase Parasite (https://parasite.wormbase.org) for 109 available nematode genomes and are included in Data S1A.[5] Pfam or InterPro protein domain terms that were used to obtain the gene lists for different gene families are listed in Data S1B. Gene numbers in a family are reported as proportions of the total number of genes in a species to account for genome duplication events during Nematoda evolution. Spearman correlations between gene proportions of different gene families across nematode species were calculated using GraphPad Prism 9.

Analysis of CeNGEN scRNA atlas data

Average Transcripts Per Million (TPM) values per neuron class for all *C. elegans* genes were downloaded from http://www.cengen.org for the threshold of 2 (medium stringency). Data were converted to binary format to identify whether a gene is expressed in a particular neuron class. Spearman correlations between number of genes of different gene families expressed across 128 distinct neuron classes were calculated using GraphPad Prism 9. For generating the heatmap for similarity in gene expression patterns, Jaccard similarity indices were calculated for each *nhr:gpcr* pair using the formula: Jaccard index for gene A and gene B = (number of neuron classes that express both gene A and gene B) / (number of neuron classes that express either gene A or gene B). A Jaccard index of one would indicate that the two genes are expressed in an identical set of neuron classes. The heatmap was generated using the 'ComplexHeatmap' function in R with the following modifications: the color gradient was linearly ramped from 'white' (= 0) to 'red'

(= 1).[28] Euclidean distance method was used for the clustering of rows and columns in the heatmap. To calculate Jaccard similarity indices for simulated *nhr:gpcr* datasets, the binary expression pattern for each *nhr* or *gpcr* gene is randomized among the 128 distinct neuron classes, while maintaining the total number of neuron classes each gene is expressed in as identical to that in the CeNGEN scRNA-based gene expression atlas. The percentile values of Jaccard similarity indices for any two gene families were calculated in R and plotted using GraphPad Prism 9.

Phylogenetic tree for *nhr* genes

Protein sequences for NHR genes in the two species, *Caenorhabditis elegans* (Clade V) and *Acrobeloides nanus* (Clade IV), were obtained from WormBase Parasite (http://parasite.wormbase.org/index.html) using the Pfam protein domain term PF00105.[5] The protein sequences (one protein per gene) were used for multiple sequence alignment using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo).[29] Jalview version 2.11.1.4 was used to generate a phylogenetic tree for only the C4-zinc finger DNA binding domain sequences with distances calculated from aggregate BLOSUM62 scores and using the neighbor joining method.[30] *C. elegans* NHR genes that are conserved across multiple evolutionary phyla are labeled in the tree.[3]

Identifying NHR and GPCR genes in hyperdivergent genomic regions

The genomic coordinates for all members of the NHR and GPCR gene families were compared to that of the previously identified hyperdivergent regions of the *C. elegans* genome.[19] A gene was identified to be in a hyperdivergent region if at least 10% of the gene's length resides within hyperdivergent regions of the *C. elegans* genome, which cover ~20% of the *C. elegans* reference genome.[19] Unlike for NHR (55%) and GPCR (52%) genes, only 17.5% of all insulin genes reside in hyperdivergent regions and hence they were excluded from this analysis. Expression analyses for NHR and GPCR genes were performed using data from the CeNGEN scRNA atlas, as described above.

Nucleotide polymorphism and positive selection

Data for nucleotide polymorphism and measures of positive selection for all *C. elegans* genes were obtained from a recent study.[10] As in the original study, all genes that display high nucleotide polymorphism were identified using a cutoff score of $Pi > 0.0025$. Similarly, genes that show signs of strong positive selection were identified based on fulfillment of both of the following criteria: (1) Tajima's *D* score < -2, and (2) Fay and Wu's *H* value < -20. The proportion of NHR genes was calculated among the 1,685 genes with high nucleotide polymorphism and the 432 genes that are under strong positive selection. The fold-enrichment values were determined relative to the total number of protein-coding genes in the *C. elegans* genome. The fold-enrichment values for chemosensory GPCR genes were reported from the previous study.[10] The statistical significance for fold-enrichment was determined using hypergeometric tests performed in R.

Predicting GPCR target genes of NHR transcription factors

TargetOrtho2 was used to predict target genes for NHR transcription factors with a classifier probability cutoff score of 0.5.[31] The Jaccard similarity indices for nervous system-wide expression patters of NHRs and their target GPCR genes were determined using data from the CeNGEN scRNA atlas, as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses and graphing were performed in GraphPad Prism 9 and R, unless otherwise stated. Details of statistical tests and the corresponding *p* values are included in the figure legends. Statistical significance was determined using a *p* value cutoff of 0.05.