28. S. W. Clark and D. I. Meyer, *Nature* **359**, 246 (1992); *J. Cell Biol.* **127**,129 (1994).
29. R. K. Herman, C. K. Cari, P. S. Hartman, *Genetics* **102**, 379 (1982).
30. R. J. Nelson, T. Ziegelhoffer, C. Nicolet, M. Werner-Washburne, E. A. Craig, *Cell* **71**, 97 (1992).
31. K. Normington, K. Kohno, Y. Kozutsumi, M. J. Gething, J. Sambrook, *ibid.* **57**, 1223 (1989).
32. M. F. P. Heschl and D. L Baillie, *Comp. Biochem. Physiol.* **96**, 633 (1990).
33. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
34. The mitochondrial ribosomal protein orthologs have been missed by the automatic comparison procedure primarily because they contain nonconserved NH_2-terminal import peptides as well as COOH-terminal tails (*28*).
35. R. Okimoto, J. L. Macfarlane, D. R. Wolstenholme, *J. Mol. Evol.* **39**, 598 (1994).
36. The domains were primarily from the SMART database [J. Schultz, F. Milpetz, P. Bork, C. P. Ponting, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857 (1998)], to which several domains were added. We have not attempted to cite the literature for each domain, but refer the reader to the SMART database. See also www.bork.embl-heidelberg.de/Modules_db/special_annotation_page.html
37. P. Bork, J. Schultz, C. P. Ponting, *Trends Biochem Sci.* **22**, 296 (1997).
38. To obtain robust counts for each of the domains in the yeast and worm protein sets, we compared representative sequences of each domain to the nonredundant protein database (National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD) using the PSI-BLAST program, and the resulting position-dependent weight matrices (profiles) were saved. The number of search iterations and the cutoff for inclusion of sequences in the profile were adjusted individually for each domain. For most widespread domains, several profiles were constructed to ensure complete coverage. The profiles were then compared separately to the yeast and worm protein databases. Typically, the random expectation value of 0.01 was used as the criterion for domain identification, but the search results were additionally scrutinized for the conservation of patterns typical of the respective domain, to ensure the elimination of any false positives. The profiles for each of the domains are available at the Web site. They can be obtained by FTP and used for PSI-BLAST searches.
39. T. M. Hall *et al.*, *Cell* **91**, 85 (1996).
40. J. A. Porter *et al.*, *ibid.* **86**, 21 (1996).
41. E. Dahl, H. Koseki, R. Balling, *Bioessays* **19**, 755 (1997); A. K. Ryan and M. G. Rosenfeld, *Genes Dev.* **11**, 1207 (1997).
42. G. Franz, T. G. Loukeris, G. Dialektaki, C. R. Thompson, C. Savakis, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4746 (1994).
43. Y. Nagai *et al.*, *FEBS Lett*, 418, 23 (1997); L. Aravind and E. V. Koonin, *J. Mol. Biol.*, in press.
44. J. M. Cherry *et al.*, *Nature* **387**, 67 (1997).
45. We thank J. Hodgkin, R. Horvitz, J. Kimble, and the editors of *Science* for the invitation to write this paper, D. Lipman for suggesting the collaborations, and K. Anders for helpful discussions. We are especially grateful to R. Durbin (Sanger Centre) and L. Hillier (Genome Sequencing Center) for providing sequence information and for their cooperation. The SGD is supported by a P41 national resources grant HG01315, from the National Human Genome Research Institute at the U.S. NIH. S.A.C. is supported by training grant PHS HG 00044. T.S. is supported by grant DE-FG02-98ER62558 from the Department of the Environment.

Neurobiology of the *Caenorhabditis elegans* Genome

Cornelia I. Bargmann

REVIEW

Neurotransmitter receptors, neurotransmitter synthesis and release pathways, and heterotrimeric GTP–binding protein (G protein)–coupled second messenger pathways are highly conserved between *Caenorhabditis elegans* and mammals, but gap junctions and chemosensory receptors have independent origins in vertebrates and nematodes. Most ion channels are similar to vertebrate channels but there are no predicted voltage-activated sodium channels. The *C. elegans* genome encodes at least 80 potassium channels, 90 neurotransmitter-gated ion channels, 50 peptide receptors, and up to 1000 orphan receptors that may be chemoreceptors. For many gene families, *C. elegans* has both conventional members and divergent outliers with weak homology to known genes; these outliers may provide insights into previously unknown functions of conserved protein families.

The nervous system of *C. elegans* has an unprecedented set of tools that are available for its analysis: a complete cell lineage that reveals the developmental origin of every neuron, a wiring diagram from serial section electron micrographs that describes all the synapses between neurons, and now the genome sequence with all the genes required to build the animal (*1, 2*) (Table 1). The nervous system contains about one-third of all the somatic cells in *C. elegans* and probably dominates a comparable portion of the genes, but at this point only a handful of those genes are understood.

Comparison of the predicted *C. elegans* genes with molecules in the vertebrate nervous system reveals many parallels and a few striking differences (*3*). Conserved gene systems include the neurotransmitter biosynthetic enzymes, synaptic release mechanisms, and neurotransmitter receptors, including both ligand-gated ion channels and G protein–coupled receptors. Most ion channels and second messenger pathways are also highly conserved. The gene for the voltage-activated sodium channel is absent, although voltage-gated potassium and calcium channels abound (*4*). *Caenorhabditis elegans* also lacks the rhodopsin molecules that are widely used for vertebrate and invertebrate vision. In two cases, gap junctions and olfactory receptors, the nematode and vertebrate gene families are unrelated in sequence but have similar properties.

Caenorhabditis elegans often has one recognizable member of a gene family where vertebrates have three or four very similar genes, consistent with the general model that the mammalian genome experienced two large-scale genome duplications early in chordate evolution (*5*). A few *C. elegans* gene families are represented by dozens or even hundreds of family members; these families are often quite large in vertebrates as well. The multigene families provide insight into the aspects of neuronal function that define different types of neurons or distinct functions within one neuron. The *C. elegans* genome highlights the special importance of diversity in sensory recognition (olfactory receptors), neuronal excitability (potassium channels), and information transfer and cell-cell recognition at the synapse (neurotransmitter receptors and gap junctions).

Regulation of Excitability

There is no voltage-activated sodium channel in the *C. elegans* genome, which is consistent with the absence of such an activity in electrophysiological recordings of *Ascaris* motor neurons (*4, 6*). Jellyfish are more primitive invertebrates than nematodes, and they do possess voltage-activated sodium channels, which suggests that the ability to generate a sodium-based action potential was lost during nematode evolution (*7*). *Caenorhabditis elegans* neurons are small and have high membrane resistance, and they are able to propagate signals efficiently without the large-scale amplification provided by the sodium channel (*8, 9*).

On the other hand, voltage-activated calcium channels and potassium currents have been observed in recordings from *C. elegans* neurons and muscles (*9, 10*). Voltage-activated calcium channels depolarize cells and allow calcium entry for synaptic exocytosis and muscle contraction. They consist of one $\alpha 1$ subunit that defines the

The author is at the Howard Hughes Medical Institute, Programs in Developmental Biology, Neuroscience, and Genetics, Departments of Anatomy and Biochemistry and Biophysics, University of California, San Francisco, CA 94143–0452, USA.

intrinsic properties of the channel and very different accessory α2/δ, β, and sometimes γ subunits that modify channel kinetics and density. *Caenorhabditis elegans* has five predicted α1 subunits, one related to $\alpha1_D$/L-type channels (EGL-19), one related to $\alpha1_A/\alpha1_B$/N/P/Q-type channels (UNC-2), one related to $\alpha1_G$/T-type channels, and two outliers that are most similar to one another and to a channel in the yeast *Schizosaccharomyces pombe*. This is one of the cases in which *C. elegans* encodes both conventional genes and outliers that are the most divergent known members of the gene family. There are also predicted α2 and β subunits and a potential ryanodine receptor that could release calcium from intracellular stores (*unc-68*) (*11, 12*).

Mutations are known for two α1 subunits, *egl-19* (also known as *pat-5* or *eat-12*) and *unc-2*. *egl-19* is likely to encode the muscle voltage-activated calcium channel: null mutants in *egl-19* are lethal and appear to lack almost all muscle contraction (*10*). *unc-2* is required for normal locomotion and acts in *C. elegans* neurons, where it might have a role at the synapse (*13*).

In contrast to the small number of calcium channel genes, there are many diverse potassium channel genes in the *C. elegans* genome (*14*). Potassium entry drives a cell to its resting potential, and potassium channels shape the electrical response of an excitable cell, determining its susceptibility to depolarization, its resting potential, and its recovery from depolarization.

There are three major classes of potassium channels: the inward rectifier channels, with two transmembrane domains (2 TM), the two pore/TWIK channels, with four transmembrane domains (4 TM), and the voltage-regulated potassium channels, with six transmembrane domains (6 TM) (*15*). For the well-characterized 2 TM and 6 TM proteins, a single channel is formed by four monomers encoded by one or more similar genes. Additional dissimilar β subunits can modify channel properties. The 6 TM channels can be regulated by voltage; in addition, the intracellular domains of all channels define their regulation by second messengers, their inactivation, and their multimerization. *Caenorhabditis elegans* has about 80 predicted potassium channels including members of 2 TM, 4 TM, and 6 TM families; within the 6 TM family there are members of five different subclasses—the voltage-gated Shaker-type potassium channels (Shak/shab/shaw/shal), the calcium-activated Slo potassium channels, the calcium-regulated SK potassium channels, the eag/erg channels with cyclic nucleotide binding domains, and the KQT channels. One of the Slo channels and all the inward rectifiers are potentially interesting outliers compared with the vertebrate genes. The 4 TM TWIK channels are the most striking group of outliers. These channels represent the largest worm family and include about 50 different genes. There are also many mammalian TWIK homologs in expressed sequence tag (EST) databases, but they have not been extensively studied. Some of the TWIK channels have unusual pore domains, so they may not all encode potassium channels (*16*).

One Shaw-like gene is mutated in *egl-36* mutants (*17, 18*), and its properties hint at the reason for potassium channel diversity. The *egl-36* mutation is a gain of function (gf) that allows potassium entry at inappropriate membrane potentials, resulting in constitutive inhibition of the egg-laying muscles that express the gene; its loss-of-function phenotype is subtle. The channel is expressed in only a few neurons and the egg-laying muscles, which suggests that its restricted function is due to restricted expression. Similarly, unpublished gf mutations in five additional potassium channel genes result in severe phenotypes that are much more striking than loss of function of the same genes. The simplest explanation for this result is that gf mutations can silence a cell entirely, whereas loss-of-function mutations affect the same cell slightly. The relatively mild loss-of-function phenotypes are probably due to a combination of redundancy of potassium channels within a given cell and restricted expression of each channel in small groups of cells.

Why are there so few calcium channels and so many potassium channels in both nematodes and vertebrates? Perhaps calcium channels provide a function that is regulated in similar ways in all cells, a common language of excitability. The potassium channels may provide the modulatory functions that make excitable cells different from one another.

Sensory Channels and Other Functions

The functions of some channels are best understood because of their roles in sensory transduction. The degenerin family of potential mechanosensory channels was identified by genetic analysis of touch sensation (*19*); this family also includes vertebrate epithelial sodium channels. These proteins have two predicted membrane-spanning domains and a predicted pore domain, together with large extracellular domains that could interact with extracellular signals. Two *C.*

Table 1. Genes that affect nervous system function (see text for references). PLC, phospholipase C; PDE, phosphodiesterase; ChAT, choline acetyltransferase; AChE, acetylcholinesterase; VChAT, vesicular ChAT.

	Approximate number of *C. elegans* genes
Voltage-regulated calcium channels	5 α1, 2 α2, 2β
Potassium channels	20 6TM (10 Shak/Shaw/Shab/Shal, 3 KQT, 2 EAG, 2 SLO, 4 SK), 3 2TM (IRK), 40+ TWIK
Chloride channels (CLC)	6
Classic neurotransmitter synthesis, degradation	1 each ChAT, GAD, tyrosine hydroxylase, etc. 4 AChE
Neurotransmitter transporters	1 each VChAT, VMAT, one outlier 12 GABA transporters/amino acid permeases 6 EAAT transporters
Neuropeptides	15 FMRF-amide related, 15 other
Neurotransmitter release/exocytosis	1 highly conserved each and 3 to 7 additional candidate synaptobrevin, syntaxin, synaptotagmin, *unc-18*, SNAP25; latrophilin; 1 rab3
Ligand-gated ion channels	10 excitatory glutamate receptors 42 acetylcholine receptors 37 GABA-A and inhibitory glutamate receptors, including outliers
PDZ domain proteins	30+
Heterotrimeric G proteins and targets	20 Gα subunits, 2 Gβ subunits, 2 Gγ, 12 RGS regulators, 3 adenylyl cyclases, 8 PLC, 4 cGMP PDE
G protein–coupled receptors	18 class A amine receptors 50 class A peptide receptors 4 class B peptide receptors 4 metabotropic glutamate receptors 3 GABA-B receptors
G protein–coupled orphan receptors (chemoreceptors)	700 *str* (ODR-10 related)/*stl/srd*, related groups; 150 *sra/srb/sre;* 40 *srg;* 80 class A orphan receptors
Innexin/gap junction proteins	24 genes
Degenerin/mechanosensory proteins	22 genes
Stomatin/*mec-2*–like regulatory proteins	9 genes
Receptor guanylyl cyclases	26 genes, also 5 soluble guanylyl cyclase genes
Cyclic nucleotide–regulated channels	6 genes (plus two *eag/erg*, K^+ channels)
TRP-related channels	11 genes
CREB and regulatory pathways	1 CREB, 300+ protein kinases: 2 Ca/CAM kinase, 2 protein kinase A, cGMP-dependent protein kinase

elegans degenerin genes required for touch sensation, *mec-4* and *mec-10*, can mutate to give dominant hyperactive channels that cause the touch cells to swell, degenerate, and die (thus degenerins) (*19, 20*). Three other degenerin genes might function as stretch receptors, a function analogous to touch sensation (*21–23*). Like the potassium channels, this family of channels appears to have a subtle or regulatory role in most cells that express it rather than an essential role.

The genetic analysis of touch sensation also identified *mec-2,* which is related to the erythrocyte band 7.2b integral membrane protein stomatin (*24*). There are eight other predicted stomatin-related genes, including *unc-24,* which is required for normal locomotion (*25*).

Cyclic nucleotide–gated channels are a prominent feature of vertebrate vision and olfaction, and two of the six *C. elegans* cyclic nucleotide–gated channel genes are essential for some forms of chemosensation and thermosensation *(tax-2* and *tax-4)*, consistent with a function in sensory transduction (*26, 27*). Other forms of chemosensation depend on the *trp*-like channel *osm-9,* whose closest vertebrate homolog is the capsaicin receptor expressed on pain-sensing nociceptive sensory neurons (*28, 29*). The capsaicin receptor is directly gated by noxious heat, whereas the *trp* channel is known from *Drosophila* phototransduction to be regulated by phospholipase C. *Caenorhabditis elegans* has 11 *trp/osm-9*-like genes, including 3 genes in a novel branch of the gene family. It also has six predicted chloride channels.

Neurotransmitters and Exocytosis

Like other animals, *C. elegans* uses the small molecule neurotransmitters acetylcholine, dopamine, serotonin, γ-aminobutyric acid (GABA), excitatory amino acids (glutamate), and the invertebrate catecholamine octopamine. At the neuromuscular junction, excitatory motor neurons release acetylcholine and inhibitory motor neurons release GABA (*30, 31*). The pharynx has specialized muscles that are also excited by acetylcholine but are inhibited by glutamate (*32–34*).

Caenorhabditis elegans and vertebrates appear to use similar molecules for neurotransmitter synthesis, packaging into synaptic vesicles, and reuptake or destruction. Mutations in genes for acetylcholine function are known: *C. elegans* has the biosynthetic enzyme choline acetyltransferase *(cha-1)* (*35*), a vesicle transporter (*unc-17)* (*36*), and cholinesterases that are highly similar to the vertebrate genes (*37, 38*) (indeed, *unc-17* was the first cloned vesicular acetylcholine transporter). GABA in *C. elegans* is presumably synthesized from glutamate by the enzyme glutamic acid decarboxylase (GAD) and packaged into vesicles by a novel transporter encoded by the *unc-47* gene (*39*). Vertebrate homologs of *unc-47* can package either GABA or glycine into synaptic vesicles (*39*). The genome includes several genes that are related to *unc-47,* and perhaps these molecules encode vesicular transporters for other small molecules.

The enzymes required for dopamine, serotonin, and octopamine synthesis (tyrosine hydroxylase, dopamine β-hydroxylase, aromatic acid decarboxylase, tryptophan hydroxylase) and vesicular transport (VMAT, the monoamine transporter) are present in the *C. elegans* genome. Candidate molecules to clear transmitters include monoamine oxidase, a putative cocaine-sensitive dopamine transporter, and a putative prozac-sensitive serotonin transporter. Glutamate reuptake from the synapse is mediated by the EAAT transporters, which may be encoded by six *C. elegans* genes.

In addition to small-molecule transmitters, *C. elegans* uses neuropeptides as transmitters. One widespread class of nematode neuropeptides has the sequence FMRF-amide (Phe-Met-Arg-Phe-amide) or closely related sequences at the COOH-terminus (*40, 41*). There are at least 15 genes in *C. elegans* that encode at least 56 potential FMRF-amide-like peptides (*42*). As is observed in other species, multiple peptides are likely to be produced from each peptide precursor gene by cleavage and processing. Mutations in one FMRF-amide gene, *flp-1,* cause defective locomotion and chemosensation (*43*). In addition to the FMRF-amide genes, there are at least 15 other genes that could encode 58 neuropeptides; most are nematode specific but some are related to snail peptides such as buccalin and pedal peptide. *Caenorhabditis elegans* does not have highly conserved homologs of the mammalian neuropeptide endorphins, enkephalins, dynorphins, substance P, substance K, and neuropeptide Y. These peptides are small and may be difficult to recognize after evolutionary divergence; the *C. elegans* genome does have many receptors that are related to the mammalian receptors for these peptides.

Many of the genes involved in neurotransmitter release have been identified because of their defects in cholinergic neurotransmission at the neuromuscular junction (*44–48*). The essential components of exocytosis are highly conserved between mammals and nematodes, with each important gene (syntaxin, synaptobrevin, *unc-18/nsec-1,* synaptotagmin, *rab-3)* represented by a highly conserved *C. elegans* homolog (*45, 49–53*). Most of the mutants in this pathway have severe uncoordinated phenotypes that are lethal or almost lethal, but a few like synaptotagmin and *rab-3* have relatively mild defects. The genome also encodes outlier proteins related to the exocytosis genes. Some of these genes probably have nonneuronal functions (*54*), but others might act in neurons. Their expression patterns should indicate whether they act in different types of neurons or whether they activate different classes of vesicles (for example, classic neurotransmitters versus neuropeptides).

Ligand-Gated Ion Channels Act in Neurons and at the Neuromuscular Junction

The ligand-gated ion channels are neurotransmitter receptors that open rapidly and desensitize rapidly, which makes them ideal for short-term signaling. Each channel is made up of four or five similar subunits that can be encoded by one to five different genes. *Caenorhabditis elegans* has about 100 genes for ligand-gated ion channels, including excitatory receptors for acetylcholine and glutamate and inhibitory receptors for GABA and glutamate.

At the nematode and vertebrate neuromuscular junction, excitatory motor neurons release acetylcholine, which drives muscle contraction (*30, 46*). In the nematode *Ascaris*, motor neurons also use acetylcholine to communicate with each other (*55*). There are about 40 predicted acetylcholine receptor subunits in *C. elegans* that fall into the ligand-gated (nicotinic) class. Three subunits of a levamisole-sensitive muscle acetylcholine receptor are encoded by the *unc-29, unc-38,* and *lev-1* genes (*56*). The roles of the other >35 genes are mostly unknown. There are other muscle acetylcholine receptors (*33, 46*). Dominant alleles in one neuronal acetylcholine receptor (*deg-3)* are associated with cell-autonomous sensory neuron degeneration, but this receptor has no obvious loss-of-function phenotype (*57*).

Excitatory glutamate receptors are unrelated in sequence to the acetylcholine receptors, although they share a similar multimeric subunit structure. *Caenorhabditis elegans* has six glutamate receptors of the AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid)/kainate class, for which mutations have been described in one: *glr-1* mutants have defects in their locomotor response to mechanosensory stimuli (*58, 59*). There are also two potential glutamate receptors in the *N*-methyl-D-aspartate (NMDA) receptor class, one type 1 and one type 2 subunit (NMDA receptors are typically type 1/type 2 heteromeric channels), and two unusual outlier genes related to glutamate receptors.

Caenorhabditis elegans has over 30 different genes that could encode inhibitory GABA-A, glycine, and glutamate receptors, which are distantly related to the excitatory nicotinic acetylcholine receptors and very closely related to one another. There is a major GABA receptor on nematode body muscle, but again the diversity of the gene family suggests other functions that are not known. Clear orthologs of glycine receptors were not found, and it is not known whether these will be present in *C. elegans*. The glutamate-gated chloride channels are the major molecular target for a class of nematocide drugs, the ivermectins, and they do not have known vertebrate homologs (*60*). Mutations in these channels have been identified in screens for high-level ivermectin resis-

tance; these mutants also have defects in pharyngeal function due to a loss of glutamate-mediated inhibition (*61*).

Several *C. elegans* ligand-gated ion channel genes are outliers that might encode previously unidentified receptor types. For example, the invertebrate histamine-gated chloride channel, which has not been cloned, might be one of these receptors.

The number of ligand-gated ion channel genes is unexpectedly large. One possible explanation for this diversity is that these channels may be localized to particular synapses or regions of the cell. In mammals, a peptide-binding motif called the PDZ domain binds to the COOH-terminus of many channels, including the glutamate-gated ion channels, and clusters them at synaptic regions. PDZ proteins also organize proteins in nonneuronal cells; the *C. elegans lin-2, lin-7,* and *lin-10* genes are required for localization of the epidermal growth factor receptor protein LET-23 in vulval precursor cells (*62*). PDZ-binding motifs are present in many *C. elegans* receptors and channels, including the glutamate receptor GLR-1, and the PDZ protein LIN-10 is involved in GLR-1 localization (*63*). There are at least 30 PDZ domain–containing proteins in *C. elegans*, which suggests that there are many different types of receptor clusters.

G Protein–Coupled Receptors

G protein–coupled receptors typically generate slower and longer-lasting changes in neuronal excitability than ligand-gated ion channels. G protein–coupled receptors have seven transmembrane domains, and they are associated with a cytoplasmic heterotrimeric G protein consisting of a guanosine triphosphate (GTP)– or guanosine diphosphate–binding α subunit and β and γ subunits. Upon ligand binding, the receptor causes GTP exchange and dissociation of the Gα subunit from Gβγ; both Gα and Gβγ subunits can interact with effectors in the cell.

Remarkably, about 5% of all *C. elegans* genes encode G protein–coupled receptors. The receptors fall into two groups: those with clear relationships to receptors identified in other animals (about 100 genes), and "orphan" worm-specific receptors (about 1000 genes). Most of the worm-specific receptors are thought to encode chemoreceptors. Most of the other genes are probably neurotransmitter receptors.

Both vertebrates and invertebrates have several families of conserved G protein–coupled receptors. Class A, the rhodopsin-like receptor family, has different subclasses with different ligands. The class A amine receptors interact with small molecule neurotransmitters such as acetylcholine, catecholamines, dopamine, and serotonin. The amine receptors are quite similar, so the ligands for the approximately 18 *C. elegans* genes in this family are uncertain. Based on sequence similarity there might be acetylcholine receptors, tyramine/octopamine receptors, serotonin receptors, and dopamine receptors. One of the predicted serotonin receptors has been shown to interact with serotonergic agonists and antagonists when expressed in mammalian cells (*64*).

Class A peptide receptors have peptide ligands. The 50 or more *C. elegans* genes in this group are less closely related to their vertebrate cousins than the amine receptors. Moreover, there are no clear *C. elegans* homologs of the vertebrate neuropeptides (see above). Nonetheless, it is possible to identify subgroups of the *C. elegans* receptors that have similarity to subgroups of vertebrate peptide receptors such as tachykinin receptors, neurotensin receptors, and neuropeptide Y receptors. No authentic rhodopsin is present.

The *C. elegans*–specific subfamily of about 80 orphan class A receptors is uncharacterized, but it might encode chemoreceptors like the other large orphan families of G protein–coupled receptors (see below) (*65*).

The class B or secretin receptor family of peptide receptors is completely unrelated to the class A family by sequence. This gene family has at least four members in *C. elegans*. By analogy to similar receptors in other animals, the *C. elegans* receptors may bind peptide hormones and perhaps have endocrine roles. The G protein–coupled metabotropic glutamate receptors are different in sequence from either class A or class B receptors. *Caenorhabditis elegans* has four genes that are similar to the glutamate receptors and related extracellular calcium receptors. The GABA-B receptor is distantly related to these receptors, and there are three *C. elegans* genes in the GABA-B subfamily. Each of the receptor families mentioned above includes at least one unusual *C. elegans* outlier gene.

Of the approximately 80 receptor genes described above, mutations have been described only for *npr-1*, a class A receptor most closely related to neuropeptide Y receptors. *npr-1* mutant animals associate with one another in groups, and they have subtle alterations in feeding behavior (*66*).

Chemosensory neurons make up about 10% of the *C. elegans* nervous system, and it is likely that many of the 1000 orphan G protein–coupled receptors act in chemosensation. A single olfactory receptor, *odr-10,* has been identified through mutation; it is predicted to be a G protein–coupled receptor in a novel sequence family (*67, 68*) The *C. elegans* genome contains over 600 genes that are *odr-10*–like as well as several additional families of orphan G protein–coupled chemoreceptors (*69*). Many but not all of these genes are expressed in chemosensory neurons (*69*), and about one-third have stop codons and frameshifts. These 1000 genes might encode 500 chemoreceptors, 200 genes with functions in other cells, and 300 pseudogenes.

The chemosensory receptors are often found in clusters that contain as many as 10 or 15 closely related genes. Clusters of two or three related genes are sometimes seen for other *C. elegans* genes, but the clustering of chemoreceptors is extreme. Clustering provides an opportunity for gene addition and loss by unequal sister chromatid exchange, perhaps explaining why these receptors evolve rapidly—a partial sequence from the nematode *Caenorhabditis briggsae* shows that as many as 20% of the genes may be present in only one of the two species (*68*). The nematode genes are not closely related to the three mammalian families of olfactory receptors, but the mammalian genes also accumulate in clusters.

Another proposed family of chemoreceptors is encoded by 26 predicted transmembrane receptor-type guanylyl cyclases (*70*). Some mammalian receptor-type guanylyl cyclases respond directly to peptide ligands, and the retinal receptor-type cyclases act as a constitutive source of guanosine 3′,5′-monophosphate (cGMP) for signaling by the G protein–coupled receptor rhodopsin. Many of the *C. elegans* cyclases are expressed in small subsets of chemosensory neurons, and they could subserve either an odorant-regulated or a constitutive role in those neurons (*70*). The one gene for which a mutation is known, *daf-11,* is required for responses to many odorants, but others could be more specific (*71*). One cyclase is expressed in the thermosensory AFD neurons, so it could function in thermosensation rather than chemosensation.

Caenorhabditis elegans encodes 20 different Gα subunits. These include one G_o-, G_q-, and G_s-like protein, as well as 17 GPA proteins that fall loosely within the G_i group. G_o and G_q proteins are expressed in neurons and affect locomotion and egg laying (*72–74*). The G_s-like protein is expressed in neurons and its loss of function phenotype is lethal at the early larval stage (*75*). Three G_i-like proteins act in olfaction and pheromone sensation (*76, 77*).

There are two Gβ subunits. One Gβ mutant is lethal, with neuronal defects that are reminiscent of G_q protein loss of function (*78*). In addition, Gβ orients cell divisions in the early embryo. Twelve potential RGS (regulators of G protein signaling) proteins are present, and mutations in one are found in the *egl-10* mutant, which has defective egg laying (*93*).

The best characterized targets of G proteins are adenylyl cyclase (G_s and G_i), phospholipase Cβ (G_q and Gβγ), potassium channels (G_i and Gβγ), calcium channels (G_o), and cGMP phosphodiesterase (G_i). The G protein–regulated second messengers cAMP, inositol trisphosphate, diacylglycerol, and cGMP can indirectly modulate many enzymes within the cell. *C. elegans* has homologs of all these major classes of G protein targets. Mutations in one adenylyl cyclase have been identified, but their phenotype is not visibly abnormal (*79*).

Gap Junctions

The *C. elegans* nervous system has about 600 highly reproducible neuronal gap junctions, or electrical synapses, in addition to the 5000 predicted chemical synapses (*2*). Neuronal gap junctions are important in locomotion and touch-withdrawal behaviors (*80*), and gap junctions also synchronize muscle cell contractions (*81*). However, *C. elegans* does not have any close relatives of the mammalian connexin gap junction gene family. Instead, its genome has over 20 innexin genes that are defined by similarity to the *Drosophila* Passover and l(1)ogre genes and the *C. elegans unc-7* and *eat-5* genes. Mutations in innexin genes can lead to a loss of electrical coupling between cells, and expression of a *Drosophila* innexin protein in paired *Xenopus* oocytes leads to electrical coupling between the two oocytes (*82*). There are no innexins in the existing vertebrate genomic sequence or EST sequences, which suggests that the gap junctions of invertebrates and vertebrates have independent and convergent origins. A detailed alignment of 24 predicted *C. elegans* innexin proteins has been published (*83*). The innexins are predicted to have four membrane-spanning domains, like the vertebrate connexins, with variable intracellular NH_2- and COOH-terminal domains that might have regulatory functions.

Mutations in the innexin *eat-5* gene affect pharyngeal muscle coupling (*81*), whereas mutations in *unc-7* and *unc-9* affect locomotion (*83, 84*). *unc-7* and *unc-9* alter neuronal gap junctions observed by electron microscopy, but these effects are subtle, which suggests that neither gene is essential for morphological gap junctions.

If each of the innexins can form a working gap junction only with a cell expressing certain other innexins, it is possible that the expression pattern of the 24 predicted innexin subunits specifies the pattern of gap junctions between cells. It is likely that some of the innexins function outside the nervous system. Innexin expression has been observed in the two-cell embryo, and many embryonic cells have gap junctions (*81*).

Does *C. elegans* Learn?

Behavioral plasticity in *C. elegans* has been described for thermotactic, chemotactic, and mechanosensory behaviors, but the genes and molecules responsible are mostly unknown (*85–88*). In other animals, protein kinases that respond to second messengers like calcium (CAM kinases), adenosine 3′,5′-monophosphate (cAMP) (protein kinase A), and cGMP (protein kinase G) are implicated in synaptic plasticity and learning, and each of these kinases is represented by one or two genes in the *C. elegans* genome. Because there are over 300 predicted protein kinases, it is likely that others also function in the nervous system. cAMP signaling pathways are important for learning in *Drosophila, Aplysia,* and mammals—in behavioral paradigms as well as in models of synaptic plasticity (*89, 90*). In particular, the transcription factor CREB (cAMP response element–binding protein) plays a role in long-term memories that depend on altered gene expression. *C. elegans* has one predicted CREB gene.

Synaptic plasticity in other animals can also be influenced by the NMDA-type glutamate receptors, which appear to be present in *C. elegans*, and nitric oxide synthase and neurotrophins, which are not present in a recognizable form. Further genetic analysis of these pathways would benefit from better assays for learning and assays for synaptic plasticity, which have not yet been developed in *C. elegans*.

Conclusions

The *C. elegans* genome reveals many potential targets for investigation. *Caenorhabditis elegans* homologs of highly conserved neuronal genes and human disease genes are open to standard methods for isolating mutations and characterizing gene networks by enhancer and suppressor analysis.

The nematode-specific systems are potential targets for nematode control. This principle has already worked for the ivermectin-sensitive glutamate-gated chloride channel and might also be useful for the innexin gap junctions and the chemosensory receptors.

Within a gene family, it is often true that *C. elegans* has genes that are more divergent than any of the known vertebrate members of the gene family. These outlier family members deserve special attention: in some cases, the highly divergent vertebrate genes probably exist but have not been found, although in other cases the *C. elegans* genes may be nematode-specific genes. Searching for functions of the outlier nematode sequences could be a fruitful way of exploring new functions in multigene families. Indeed, the mammalian T-type calcium channel was identified starting from the sequence of a divergent nematode channel that later proved to be T-type–like (*91*). Intriguing outliers include calcium channel subunits, potassium channels, ligand-gated ion channels, *trp*-like channels, GABA transporters, and neuropeptide receptors.

Although *C. elegans* is intensively studied, there are no known mutations in most of the receptor and channel genes. This absence is especially striking in the large gene families like the G protein–coupled receptors and the potassium channels. By contrast, numerous mutations are known in the small gene families such as calcium channel subunits and synaptic vesicle proteins. In the short term, analyzing the new genes will be facilitated by describing their patterns of expression, including subcellular localization. In the long term, the genome sequence points to the need for better behavioral and electrophysiological assays in *C. elegans* neurobiology. Most mutant screens conducted in the past have required a substantial defect in neuromuscular function such as uncoordinated movement. These screens revealed genes with widespread function and even weak mutations in lethal genes, but they overlooked most genes with subtle, modulatory, or cell-specific functions. The limiting steps for understanding gene function are defining the function of each neuron (still unknown for many *C. elegans* neurons) and devising better assays for neuronal function in vivo. In the end, instead of transcending neurobiology, the genome leads back to it.

References and Notes

1. J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, *Dev. Biol.* **100**, 64 (1983).
2. J. G. White *et al.*, *Philos. Trans. R. Soc. London Ser. B* **314**, 1 (1986).
3. *Caenorhabditis elegans* genes with relevance to the nervous system were sought by WU-BLAST 2.0a13 searches with channels, receptors, and signaling proteins to the *C. elegans* BLAST server (wormpep option), containing 18,452 predicted proteins, about 90% of the genome. Protein matches with a probability of $<10^{-4}$ were examined further, often by identifying the closest nonworm match to the *C. elegans* gene in a BLAST search to the NCBI nonredundant databases. For multigene families, multiple alignments of the conserved domains used in Pfam were generated (often essentially the whole protein), aligning all the worm proteins found to nonworm proteins in the family from Swiss-Prot/swtrembl. This analysis will miss some real similarities and falsely include some similarities based on a single shared protein domain, so it should be considered a first-pass approach rather than a complete description. A general overview and a list of the BLAST hits discussed here can be found at www.sciencemag.org/feature/data/c-elegans.shl and www.sciencemag.org/feature/data/986006.shl.
4. Although the true tetrodotoxin-sensitive sodium channel is absent from the genome, it is possible that one of the divergent calcium or potassium channel sequences actually encodes a voltage-activated sodium channel. However, this would represent an independent evolution of the channel rather than a true ortholog.
5. A. Sidow, *Curr. Opin. Genet. Dev.* **6**, 715 (1996).
6. R. E. Davis and A. O. W. Stretton, *J. Neurosci.* **9**, 415 (1989).
7. P. A. Anderson, M. A. Holman, R. M. Greenberg, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7419 (1993).
8. R. E. Davis and A. O. W. Stretton, *J. Neurosci.* **9**, 403 (1989).
9. M. B. Goodman, D. H. Hall, L. Avery, S. R. Lockery, *Neuron* **20**, 763 (1998).
10. R. Y. Lee, L. Lobel, M. Hengartner, H. R. Horvitz, L. Avery, *EMBO J.* **16**, 6066 (1997).
11. E. B. Maryon, R. Coronado, P. Anderson, *J. Cell Biol.* **134**, 885 (1996).
12. Y. Sakube, H. Ando, H. Kagawa, *J. Mol. Biol.* **267**, 849 (1997).
13. W. R. Schafer and C. J. Kenyon, *Nature* **375**, 73 (1995).
14. A. Wei, T. Jegla, L. Salkoff, *Neuropharmacology* **35**, 805 (1996).
15. The 6 TM channel family includes some channels that are not predominantly voltage gated, such as SK and BK channels.
16. J. H. Thomas, personal communication.
17. D. A. Elkes, D. L. Cardozo, J. Madison, J. M. Kaplan, *Neuron* **19**, 165 (1997).
18. D. B. Johnstone, A. Wei, A. Butler, L. Salkoff, J. H. Thomas, *ibid.*, p. 151.
19. M. Driscoll and M. Chalfie, *Nature* **349**, 588 (1991).
20. M. Huang and M. Chalfie, *ibid.* **367**, 467 (1994).
21. M. Chalfie and E. Wolinsky, *ibid.* **345**, 410 (1990).
22. J. Liu, B. Schrank, R. H. Waterston, *Science* **273**, 361 (1996).
23. N. Tavernarakis, W. Shreffler, S. Wang, M. Driscoll, *Neuron* **18**, 107 (1997).
24. M. Huang, G. Gu, E. L. Ferguson, M. Chalfie, *Nature* **378**, 292 (1995).

25. T. M. Barnes, Y. Jin, H. R. Horvitz, G. Ruvkun, S. Hekimi, *J. Neurochem.* **67**, 46 (1996).
26. C. M. Coburn and C. I. Bargmann, *Neuron* **17**, 695 (1996).
27. H. Komatsu, I. Mori, J.-S. Rhee, N. Akaike, Y. Ohshima, *ibid.*, p. 707.
28. M. J. Caterina *et al.*, *Nature* **389**, 816 (1997).
29. H. A. Colbert, T. L. Smith, C. I. Bargmann, *J. Neurosci.* **17**, 8259 (1997).
30. A. O. W. Stretton, R. E. Davis, J. D. Angstadt, J. E. Donmoyer, C. D. Johnson, *Trends Neurosci.* **8**, 294 (1985).
31. J. P. Walrond, I. S. Kass, A. O. W. Stretton, J. E. Donmoyer, *J. Neurosci.* **5**, 1 (1985).
32. L. Avery and H. R. Horvitz, *J. Exp. Zool.* **253**, 263 (1990).
33. D. M. Raizen, R. Y. N. Lee, L. Avery, *Genetics* **141**, 1365 (1995).
34. H. Li, L. Avery, W. Denk, G. P. Hess, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5912 (1997).
35. A. Alfonso, K. Grundahl, J. R. McManus, J. B. Rand, *J. Neurosci.* **14**, 2290 (1994).
36. A. Alfonso, K. Grundahl, J. S. Duerr, H. P. Han, J. B. Rand, *Science* **261**, 617 (1993).
37. C. D. Johnson *et al.*, *Genetics* **97**, 261 (1981).
38. M. Arpagaus *et al.*, *J. Biol. Chem.* **269**, 9957 (1994).
39. S. L. McIntire, R. J. Reimer, K. Schuske, R. H. Edwards, E. M. Jorgensen, *Nature* **389**, 870 (1997).
40. A. O. Stretton, C. Cowden, P. Sithigorngul, R. E. Davis, *Parasitology* **102**, S107 (1991).
41. C. Cowden, P. Sithigorngul, P. Brackley, J. Guastella, A. O. Stretton, *J. Comp. Neurol.* **333**, 455 (1993).
42. L. S. Nelson, L. Kim, J. E. Memmott, C. Li, *Mol. Brain Res.* **58**, 103 (1998).
43. L. S. Nelson, M. L. Rosoff, C. Li, *Science* **281**, 1686 (1998).
44. I. N. Maruyama and S. Brenner, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5729 (1991).
45. K. Gengyo-Ando *et al.*, *Neuron* **11**, 703 (1993).
46. E. M. Jorgensen and M. L. Nonet, *Semin. Dev. Biol.* **6**, 207 (1995).
47. J. B. Rand and C. D. Johnson, in *Methods in Cell Biology*. Caenorhabditis elegans: *Modern Biological Analysis of an Organism*, H. F. Epstein and D. C. Shakes, Eds. (Academic Press, San Diego, 1995), vol. 48, pp. 187–204.
48. K. Iwasaki, J. Staunton, O. Saifee, M. Nonet, J. H. Thomas, *Neuron* **18**, 613 (1997).
49. M. L. Nonet, K. Grundahl, B. J. Meyer, J. B. Rand, *Cell* **73**, 1291 (1993).
50. E. M. Jorgensen *et al.*, *Nature* **378**, 196 (1995).
51. M. L. Nonet *et al.*, *J. Neurosci.* **17**, 8061 (1997).
52. M. L. Nonet, O. Saifee, H. Zhao, J. B. Rand, L. Wei, *ibid.* **18**, 70 (1998).
53. O. Saifee, L. Wei, M. L. Nonet, *Mol. Biol. Cell* **9**, 1235 (1998).
54. For example, all the exocytosis proteins except synaptotagmin are conserved in yeast secretory pathways.
55. J. P. Walrond and A. O. W. Stretton, *J. Neurosci.* **5**, 9 (1985).
56. J. T. Fleming *et al.*, *ibid.* **17**, 5843 (1997).
57. M. Treinin and M. Chalfie, *Neuron* **14**, 871 (1995).
58. A. Hart, S. Sims, J. Kaplan, *Nature* **378**, 82 (1995).
59. A. V. Maricq, E. Peckol, M. Driscoll, C. I. Bargmann, *ibid.*, p. 78.
60. D. F. Cully *et al.*, *ibid.* **371**, 707 (1994).
61. J. A. Dent, M. W. Davis, L. Avery, *EMBO J.* **16**, 5867 (1997).
62. S. M. Kaech, C. W. Whitfield, S. K. Kim, *Cell* **94**, 761 (1998).
63. C. Rongo, C. W. Whitfield, A. Rodal, S. K. Kim, J. M. Kaplan, *ibid.*, p. 751.
64. B. Olde and W. R. McCombie, *J. Mol. Neurosci.* **8**, 53 (1997).
65. E. L. L. Sonnhammer and R. Durbin, *Genomics* **46**, 200 (1997).
66. M. de Bono and C. I. Bargmann, *Cell* **94**, 679 (1998).
67. P. Sengupta, J. C. Chou, C. I. Bargmann, *ibid.* **84**, 899 (1996).
68. H. M. Robertson, *Genome Res.* **8**, 449 (1998).
69. E. R. Troemel, J. H. Chou, N. D. Dwyer, H. A. Colbert, C. I. Bargmann, *Cell* **83**, 207 (1995).
70. S. Yu, L. Avery, E. Baude, D. L. Garbers, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3384 (1997).
71. D. Birnby and J. Thomas, unpublished data.
72. J. Mendel *et al.*, *Science* **267**, 1652 (1995).
73. L. Segalat, D. Elkes, J. Kaplan, *ibid.*, p. 1648.
74. L. Brundage *et al.*, *Neuron* **16**, 999 (1996).
75. H. C. Korswagen, J. H. Park, Y. Ohshima, R. H. Plasterk, *Genes Dev.* **11**, 1493 (1997).
76. R. R. Zwaal, J. E. Mendel, P. W. Sternberg, R. H. A. Plasterk, *Genetics* **145**, 715 (1997).
77. K. Roayaie, J. G. Crump, A. Sagasti, C. I. Bargmann, *Neuron* **20**, 55 (1998).
78. R. Zwaal *et al.*, *Cell* **84**, 619 (1996).
79. A. J. Berger, A. C. Hart, J. M. Kaplan, *J. Neurosci.* **18**, 2871 (1998).
80. M. Chalfie *et al.*, *ibid.* **5**, 956 (1985).
81. T. A. Starich, R. Y. N. Lee, C. Panzarella, L. Avery, J. E. Shaw, *J. Cell Biol.* **134**, 537 (1996).
82. P. Phelan *et al.*, *Nature* **391**, 181 (1998).
83. T. M. Barnes and S. Hekimi, *J. Neurochem.* **69**, 2251 (1997).
84. T. A. Starich, R. K. Herman, J. E. Shaw, *Genetics* **133**, 527 (1993).
85. E. M. Hedgecock and R. L. Russell, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4061 (1975).
86. C. H. Rankin, C. D. Beck, C. M. Chiba, *Behav. Brain Res.* **37**, 89 (1990).
87. H. A. Colbert and C. I. Bargmann, *Learn. Mem.* **4**, 179 (1997).
88. J. Y. Wen *et al.*, *Behav. Neurosci.* **111**, 354 (1997).
89. C. H. Bailey, D. Bartsch, E. R. Kandel, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13445 (1996).
90. A. J. Silva, J. H. Kogan, P. W. Frankland, S. Kida, *Annu. Rev. Neurosci.* **21**, 127 (1998).
91. E. Perez-Reyes *et al.*, *Nature* **391**, 896 (1998).

92. I thank J. Thomas for his ideas and insights into ion channels; R. Durbin and M. Finney for assistance with sequence analysis; and E. Sonnhammer, L. Salkoff, H. Robertson, D. Garbers, T. Barnes, and S. Hekimi for their analysis of various gene families. J. Thomas, V. Maricq, W. Davis, J. Dent, E. Lundquist, M. Finney, and B. Horvitz provided valuable comments on this manuscript.

The Taxonomy of Developmental Control in *Caenorhabditis elegans*

Gary Ruvkun* and Oliver Hobert

REVIEW

The *Caenorhabditis elegans* genome sequence was surveyed for transcription factor and signaling gene families that have been shown to regulate development in a variety of species. About 10 to 25 percent of the genes in most of the gene families already have been genetically analyzed in *C. elegans*, about half of the genes detect probable orthologs in other species, and about 10 to 25 percent of the genes are, at present, unique to *C. elegans*. *Caenorhabditis elegans* is also missing genes that are found in vertebrates and other invertebrates. Thus the genome sequence reveals universals in developmental control that are the legacy of metazoan complexity before the Cambrian explosion, as well as genes that have been more recently invented or lost in particular phylogenetic lineages.

Genetic analysis of development has been a traditional focus of *C. elegans* research. Approximately 200 of the 1600 loci that have been identified by genetic analysis cause the sort of cell fate transformations and patterning defects that attract developmental geneticists, and so far about 150 genes (almost 1% of the total genes) have been studied molecularly (*1*). This set of molecularly analyzed developmental control genes, while biased toward particular intensively studied pathways, represents genes that control a fairly random sample of developmental events. More than 90% are related to genes identified by analogous molecular and genetic analyses, especially in *Drosophila* and vertebrates. Most of the genes fit into the modern developmental control canon: growth factor signaling pathways (about 30% of the genes) and transcriptional regulatory cascades (about 25% of the genes). These sequence similarities allow developmental control to be described in molecular terms. Only 10% of these genes show no detectable sequence similarity to other genes in the databases. This is in contrast to the overall genome sequence, which reveals that about 50% of *C. elegans* genes encode novel proteins. The underrepresentation of novel genes in the set of developmental control genes identified by genetics, which is not biased toward any particular molecular feature, implies that a conserved set of genes regulates metazoan development.

Most of the gene families that include the genetically identified *C. elegans* control genes are large and contain members from many species; these families can be classified into dendrograms of relatedness (*2*) (Fig. 1). For example, the tree of 355 homeobox genes classifies the relatedness of an ancient, highly ramified gene family.

The authors are in the Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, Boston, MA 02114, USA.

*To whom correspondence should be addressed. E-mail: ruvkun@frodo.mgh.harvard.edu