CAENOR

THE SYSTEM

The nematode *C. elegans* is an important mod
genetic analysis. Furthermore, *C. elegans* is t
made in mapping and sequencing the genome, and some of the interesting fea
side of the chart are highlighted some of the paths that can be taken and some

PROGRESS ON THE GENOME

Statistics and Resources

Haploid genome size	100,000,000 bases (100 Mb)
Chromosomes	5 autosomes (I-V), 1 sex chromosome (X)
Repeat content	17%, mostly in small families

Sequence Statistics

Total finished	21.3 Mb from 660 clones
II - 6.6 Mb	III - 7.1 Mb IV - 0.1 Mb X - 7.5 Mb
Gene density	
Autosome gene clusters	1 per 4.8 kb
Sex chromosome	1 per 6 kb
Predicted total genes	13,100

Physical Map

Mapped cosmids	17,500
Mapped YACs	3000
Total coverage	>95%
Total genes covered	>99%
Remaining gaps	7

URLs for Internet resources for *Caenorhabditis*:

Genome Sequencing Center, St. Louis:	http://genome.wustl.edu/gsc/gschmpg.html
Sanger Centre, Cambridge:	http://www.sanger.ac.uk
ACEDB database:	ftp://ncbi.nlm.nih.gov/repository/acedb
Caenorhabditis Genetics Center:	gopher://elegans.cbs.umn.edu:70
Leon Avery's World Wide Web page:	http://eatworms.swmed.edu

Other resources are accessible through Leon Avery's World Wide Web page.

cDNAs

Partial
provid
sequences. The cDN
genomic sequence, pa
plicated splice pattern
analyzed. A group is
the same 3' untranslat

Unique cDNA	
Differential 3'-ends	
Alternative splicing	
Similarity to known seque	
Identical to known *C. eleg*	
Mapped onto YAC arrays	
Nonuniform embryonic expression	>16/ by in situ hy

*Members of a pair are de
from the same gene but h
different polyA addition si
**Similarity was defined b
BLASTX >100

As of 1 June, 1995

SCIENCE

GENOME MAPS VI

RHABDITIS E

nt model system for studies of development, cell biology, and neurobiology. It has a short life c
ans is transparent, and the fate of individual cells can be reproducibly mapped in the developing
ing features that have appeared. The map and sequence are tools that can be used to increase ou
d some examples of the biological questions that can be addressed.

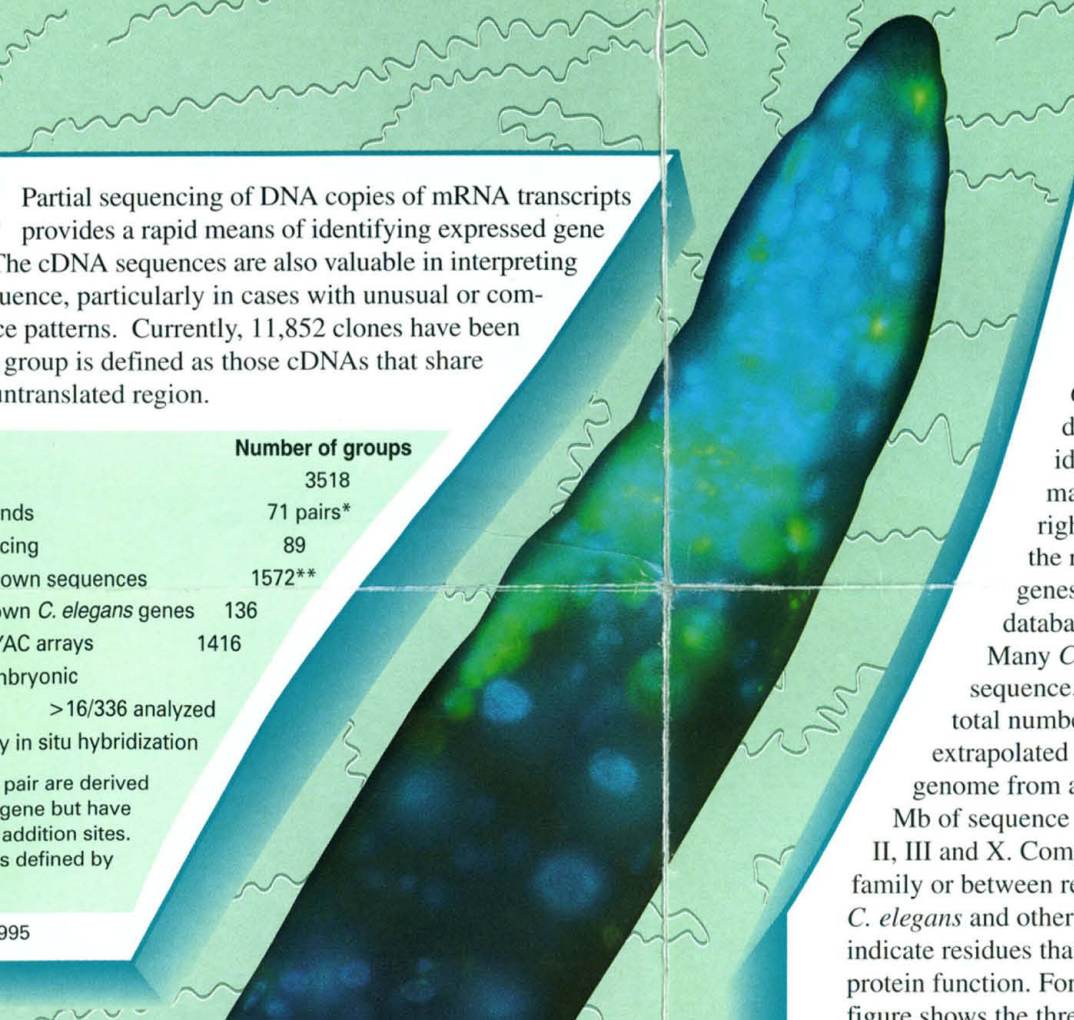
FROM S

Partial sequencing of DNA copies of mRNA transcripts provides a rapid means of identifying expressed gene
The cDNA sequences are also valuable in interpreting
uence, particularly in cases with unusual or com-
ce patterns. Currently, 11,852 clones have been
group is defined as those cDNAs that share
ntranslated region.

	Number of groups
	3518
nds	71 pairs*
cing	89
own sequences	1572**
wn *C. elegans* genes	136
AC arrays	1416
nbryonic	
	>16/336 analyzed
y in situ hybridization	

pair are derived
gene but have
addition sites.
s defined by

995



Predicted G

have significant simila
organisms. These simil
sequences shared by all
only in Metazoa. To illu
C. elegans as a model sy
determine that 32 of the 4
identified by positional cl
matches to worm genes (*P*
right provides some exampl
the recently discovered early
genes, the *C. elegans* gene re
database match.

Many *C. elegans* genes also belo
sequence. The table at the bottom
total number of members in several
extrapolated for the whole
genome from an analysis of 10
Mb of sequence on chromosomes
II, III and X. Comparisons within a
family or between related sequences in
C. elegans and other organisms can
indicate residues that are important to
protein function. For example, the
figure shows the three-dimensional

ELEGANS

rt life cycle and reproduces by self- and cross-fertilization, which facilitates eloping organism. This chart shows on the left the progress that has been ease our understanding of fundamental biological processes. On the right

M SEQUENCE TO BIOLOGY

ed Genes More than 40% of *C. elegans* genes nt similarity to genes from other ese similarities range from ed by all organisms to those found a. To illustrate the potential utility of model system, BLASTX was used to 2 of the 44 human disease genes tional cloning had significant genes ($P<0.05$). The table at the top e examples. In some cases, such as red early-onset Alzheimer's disease gene represents the only significant

also belong to families of related e bottom right shows the predicted n several prominent gene families,

Human Disease Gene (Gene symbol)	*C. elegans* Gene
▼ Putative early-onset Alzheimer's *(S182)*	*spe-4*
▼ Achondroplastic dwarfism *(FGFR3)*	ZK938.5
▼ Amyotrophic lateral sclerosis *(SOD1)*	F55H2.1
▼ Cystic fibrosis *(CFTR)*	DH11.3
▼ Breast and ovarian cancer susceptibility *(BRCA1)*	T02C1.1
▼ Duchenne muscular dystrophy *(DMD)*	yk26d3.5
▼ Lowe oculocerebrorenal syndrome *(OCRL)*	C50C3.7
▼ Dominant myotonia congenita *(CLC1)*	E04F6.11
▼ Polycystic kidney disease 1 *(PKD1)*	ZK945.9
▼ Tuberous sclerosis *(TSC2)*	T14B4.7
▼ Wilson disease *(ATP7B)*	yk29a9.5
▼ Klein-Waardenburg syndrome *(PAX3)*	F26C11.2
▼ Wilm's tumor *(WT1)*	F54H5.4
▼ Neurofibromatosis type 1 *(NF1)*	C07B5.1

Similarity: ▼strong ▼moderate ▼weak

Gene family	Estimated total
Protein kinase	350
RNA recognition motifs [RRM]	140
Homeodomain	110
EF-Tu/Ras GTPase superfamily	110
[Novel, *C. elegans*-specific?]	90
Annexin	80

Total finished	21.3 Mb from 660 clones
II - 6.6 Mb	III - 7.1 Mb IV - 0.1 Mb X - 7.5 Mb
Gene density	
Autosome gene clusters	1 per 4.8 kb
Sex chromosome	1 per 6 kb
Predicted total genes	13,100

Mapped cosmids	17,300
Mapped YACs	3000
Total coverage	>95%
Total genes covered	>99%
Remaining gaps	7

Unique cDNA
Differential 3'-ends
Alternative splicing
Similarity to known seque
Identical to known *C. eleg*
Mapped onto YAC arrays
Nonuniform embryonic expression >16/ by in situ hy

*Members of a pair are de from the same gene but h different polyA addition si
**Similarity was defined b BLASTX >100

As of 1 June, 1995

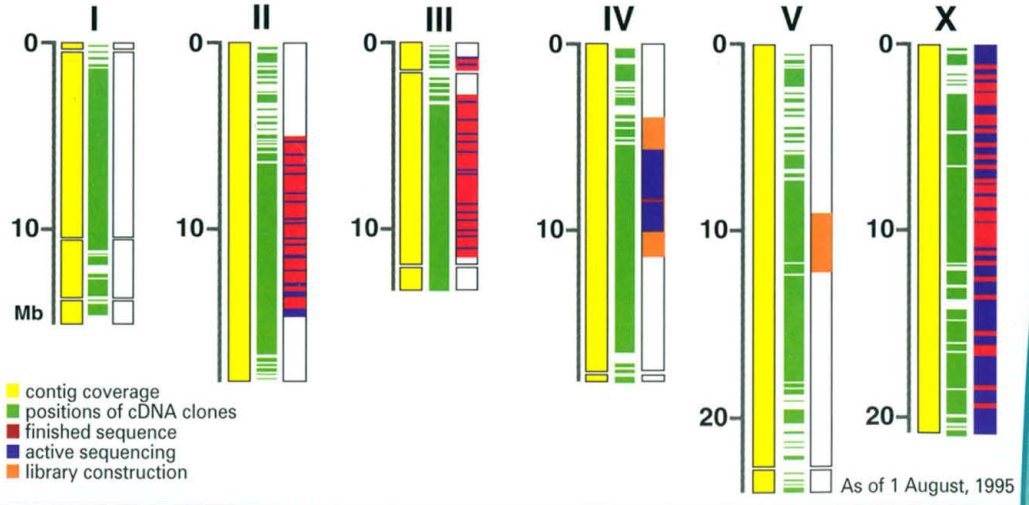
URLs for Internet resources for *Caenorhabditis*:

Genome Sequencing Center, St. Louis:	http://genome.wustl.edu/gsc/gschmpg.html
Sanger Centre, Cambridge:	http://www.sanger.ac.uk
ACEDB database:	ftp://ncbi.nlm.nih.gov/repository/acedb
Caenorhabditis Genetics Center:	gopher://elegans.cbs.umn.edu:70
Leon Avery's World Wide Web page:	http://eatworms.swmed.edu

Other resources are accessible through Leon Avery's World Wide Web page.

Physical and Sequence Maps

Only two to three contigs are currently required to cover each chromosome, with II and X entirely spanned by just a single contig (leftmost column under each chromosome). The position of cDNA clones (middle) illustrates the nonuniform distribution of genes,with the majority of genes clustered in ~half of each autosome. Sequencing (right) is virtually complete for the clusters on chromosomes III and II and just starting on chromosome V.



As of 1 August, 1995

Wealth of Information

Vast amounts of data are available on *C. elegans* and the sequence has a high information content. For example, the more than 20 Mb of completed sequence, printed at the size of the sequences shown below, would fill 150 wall charts. To give some sense of the scale, the chromosome overview of physical and sequence maps is expanded through increasingly refined views until it reaches the sequence itself.

	3518
nds	71 pairs*
cing	89
own sequences	1572**
own *C. elegans* genes	136
/AC arrays	1416
nbryonic	
	>16/336 analyzed
y in situ hybridization	

pair are derived
gene but have
addition sites.
s defined by

995

identified by positional cl
matches to worm genes (*P*-
right provides some exampl
the recently discovered early
genes, the *C. elegans* gene re
database match.

Many *C. elegans* genes also belo
sequence. The table at the bottom
total number of members in several
extrapolated for the whole
genome from an analysis of 10
Mb of sequence on chromosomes
II, III and X. Comparisons within a
family or between related sequences in
C. elegans and other organisms can
indicate residues that are important to
protein function. For example, the
figure shows the three-dimensional
structure of human H-RAS and the
residues (pink) with >70% identity in
an alignment of the *C. elegans* EF-
Tu/Ras GTPase family. The conserved
residues are clustered around the GTP-
binding active site (yellow).

Forward and Reverse Genetics

recognizable phenotype is generally fairly simple. After
are introduced as transgenes into the worm's germ
line (*A*) and assayed for their ability to rescue the
mutation. For reverse genetics, given the sequence
of a gene, one can isolate a mutant animal in which
that gene is inactivated in two steps: identification
in an ordered, frozen, mutant library (*B*) of a worm
that carries a transposon insertion near the gene,
followed by isolation of a deletion derivative. With
a mutant in hand, a hunt for enhancers and
suppressors of the original phenotype can point
the way to other factors that interact with the gene.

A

Expression Patterns

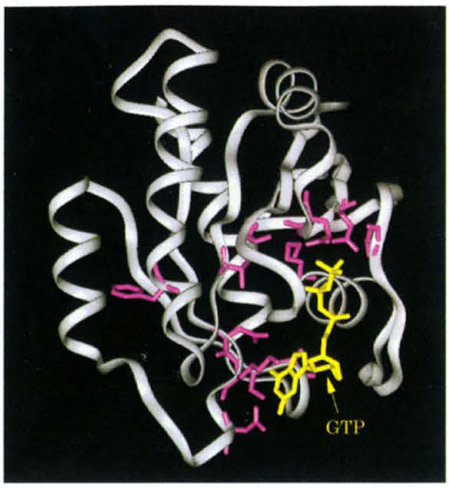
Cell-
have
tagging a promoter by fusion with genes
encoding LacZ or GFP, or immunocytochem
istry. (*A*) RNA in situ hybridization of *myo-*

tional cloning had significant
genes ($P<0.05$). The table at the top
e examples. In some cases, such as
red early-onset Alzheimer's disease
gene represents the only significant

also belong to families of related
e bottom right shows the predicted
n several prominent gene families,

▼ Dominant myotonia congenita *(CLC1)*	E04F6.11
▼ Polycystic kidney disease 1 *(PKD1)*	ZK945.9
▼ Tuberous sclerosis *(TSC2)*	T14B4.7
▼ Wilson disease *(ATP7B)*	yk29a9.5
▼ Klein-Waardenburg syndrome *(PAX3)*	F26C11.2
▼ Wilm's tumor *(WT1)*	F54H5.4
▼ Neurofibromatosis type 1 *(NF1)*	C07B5.1

Similarity: ▼strong ▼moderate ▼weak

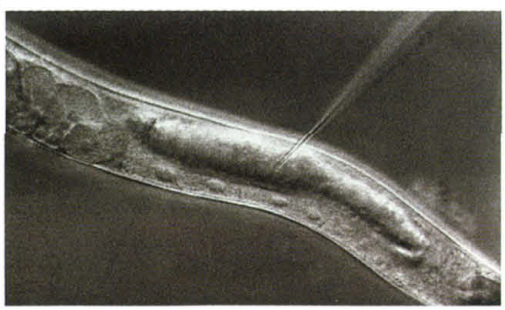


Gene family	Estimated total
Protein kinase	350
RNA recognition motifs [RRM]	140
Homeodomain	110
EF-Tu/Ras GTPase superfamily	110
[Novel, *C. elegans*-specific?]	90
Annexin	80
Serine protease inhibitor	80
Protein phosphatase	80
Ubiquitin	80
DEAD-box RNA helicase	70
Reverse transcriptase	70
Cytochrome P450	60
[Novel, *C. elegans*-specific?]	60
G protein-coupled receptor	60

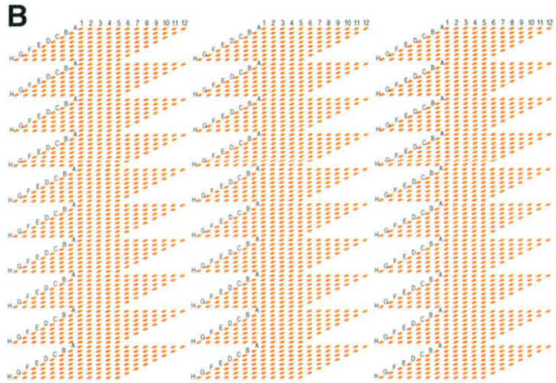
etics

The analysis of a gene's function requires knowledge of mutant phenotypes. Both forward and reverse genetics are possible. Positional cloning of a gene that has mutated to cause a e. After genetic mapping of the gene, available cosmid or YAC clones from that region of the genome

A



B



s

Cell- or organ-specific patterns of expression have been revealed by in situ hybridization, genes cytochem- of *myo-2*

Mutant Phenotype

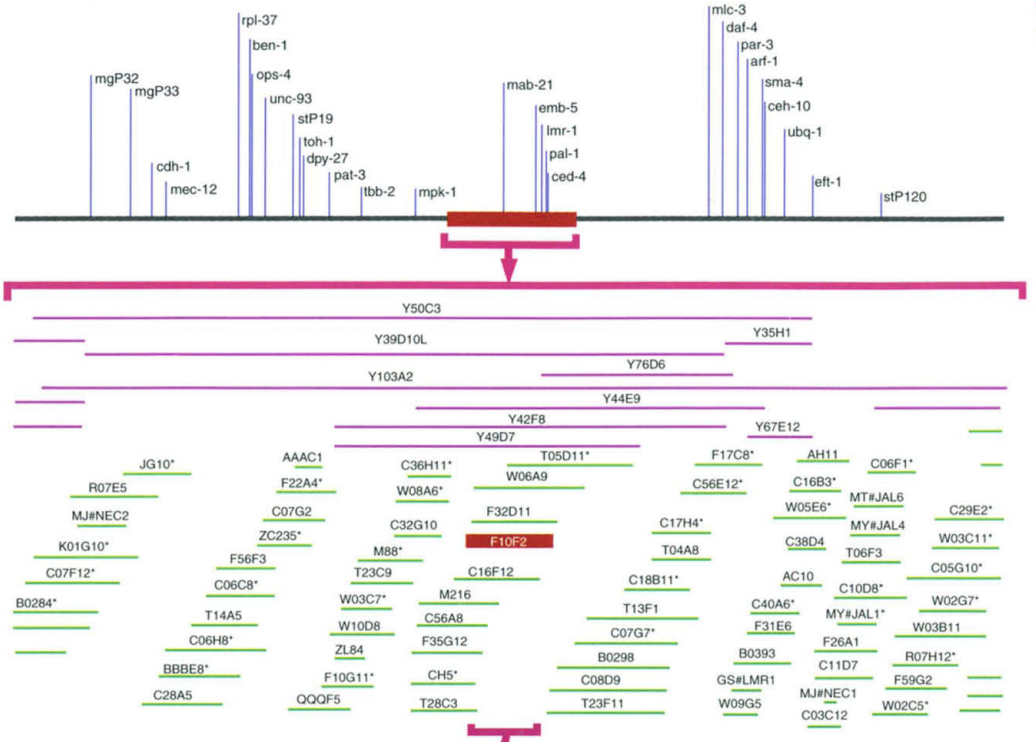
Mutants of *C. elegans* exhibit a variety of morphological and behavioral phenotypes. This *unc-6* mutant shows

finished sequence
active sequencing
library construction

As of 1 August, 1995

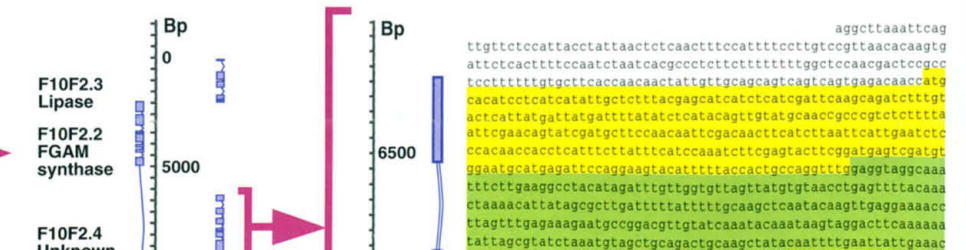
Wealth of Information

Vast amounts of data are available on *C. elegans* and the sequence has a high information content. For example, the more than 20 Mb of completed sequence, printed at the size of the sequences shown below, would fill 150 wall charts. To give some sense of the scale, the chromosome overview of physical and sequence maps is expanded through increasingly refined views until it reaches the sequence itself.



The region in the upper portion of the diagram spans roughly 3 Mb of chromosome III, showing some of the and polymorphisms used to correlate positions on the physical and genetic maps. The bar at the center is ex below to show the clone coverage of an ~500-kb region, including the cosmid F10F2. Asterisks indicate site additional mapped clones are available.

A fragment has been expanded to show the five genes within the large introns of gene F10F2.2. These genes are present on the strand opposite to that of F10F2.2. The portion containing the gene F10F2.4 has been expanded to the resolution of the DNA sequence.



that carries a transposon insertion near the gene, followed by isolation of a deletion derivative. With a mutant in hand, a hunt for enhancers and suppressors of the original phenotype can point the way to other factors that interact with the gene.

cDNAs

Expression Patterns

Cell- have

tagging a promoter by fusion with genes encoding LacZ or GFP, or immunocytochem istry. (*A*) RNA in situ hybridization of *myo-* (pharyngeal muscle-specific heavy chain myosin) in a larva. (*B*) The expression patte of alternate forms of *mec-9* as shown by *lac* staining. Left: the product of the long transcript, which is expressed in the touch cells. Right: the short transcript is expressed in the ventral cord motor neurons and groups of cells in the head. (*C*) GFP fluores- cence from a fusion with *glr-1* (glutamate receptor subunit). (*D*) Treatment with an antibody to MEC-7; staining shows outlines of touch receptor neurons.

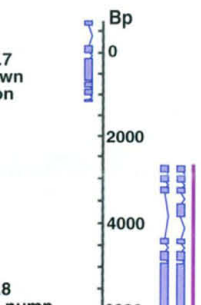
A

— YACs
— cosmids

me of the genes
nter is expanded
licate sites where

Fate and Function

All differentiated cells in developing organisms arise from multipotent precu and control of cell fate and function is central problem of biology. In *C. elega* availability of cell lineage maps and ge combined with increased sequence information, allows analysis of networ interactions that control development. example illustrates the formation and differentiation of five types of motor ne (VA, VB, VC, AS, and VD) that are pa central nervous system in *C. elegans*. E the 10 precursor cells (ventral cord neuroblasts) divides in the same, stereo way. The fates of the resulting neurons (bottom diagram) differ according to th patterns of genes they express; this patt



As another example of the compact organization of genes in *C. elegans* this region of chromosome III, which is 3 Mb from F10F2.2, illustrates an operon. Genes

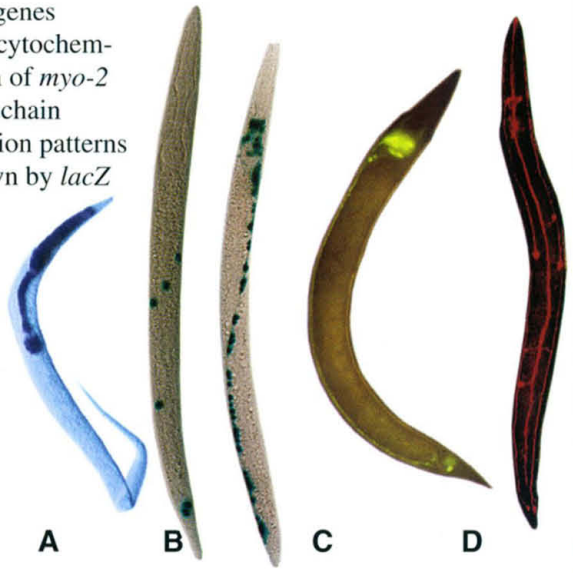
ith

ne.

s

Cell- or organ-specific patterns of expression have been revealed by in situ hybridization,

genes
cytochem-
of *myo-2*
chain
ion patterns
n by *lacZ*

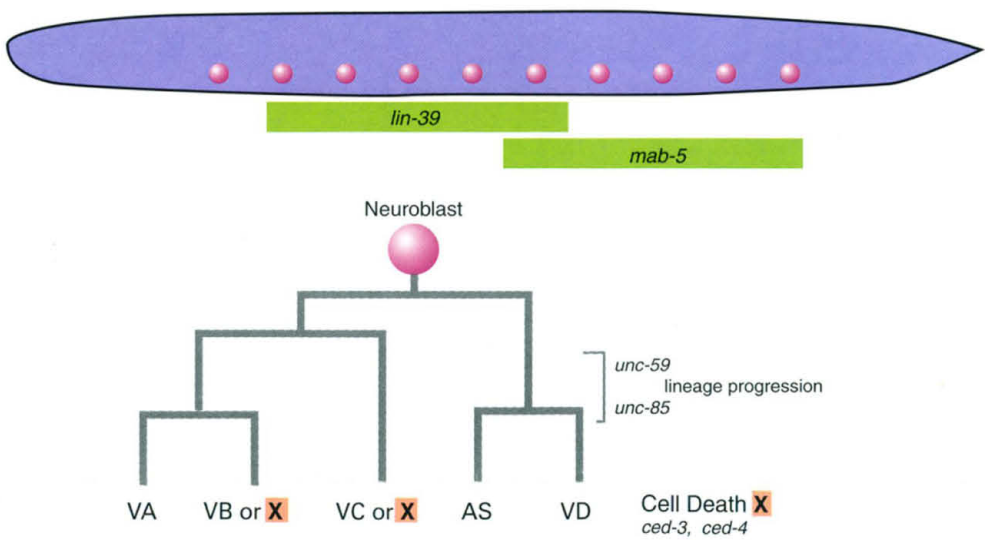


Mutant Phenotype

Mutants of *C. elegans* exhibit a variety of morphological and behavioral phenotypes. This *unc-6* mutant shows a severe coiling phenotype as a result of neuronal miswiring.

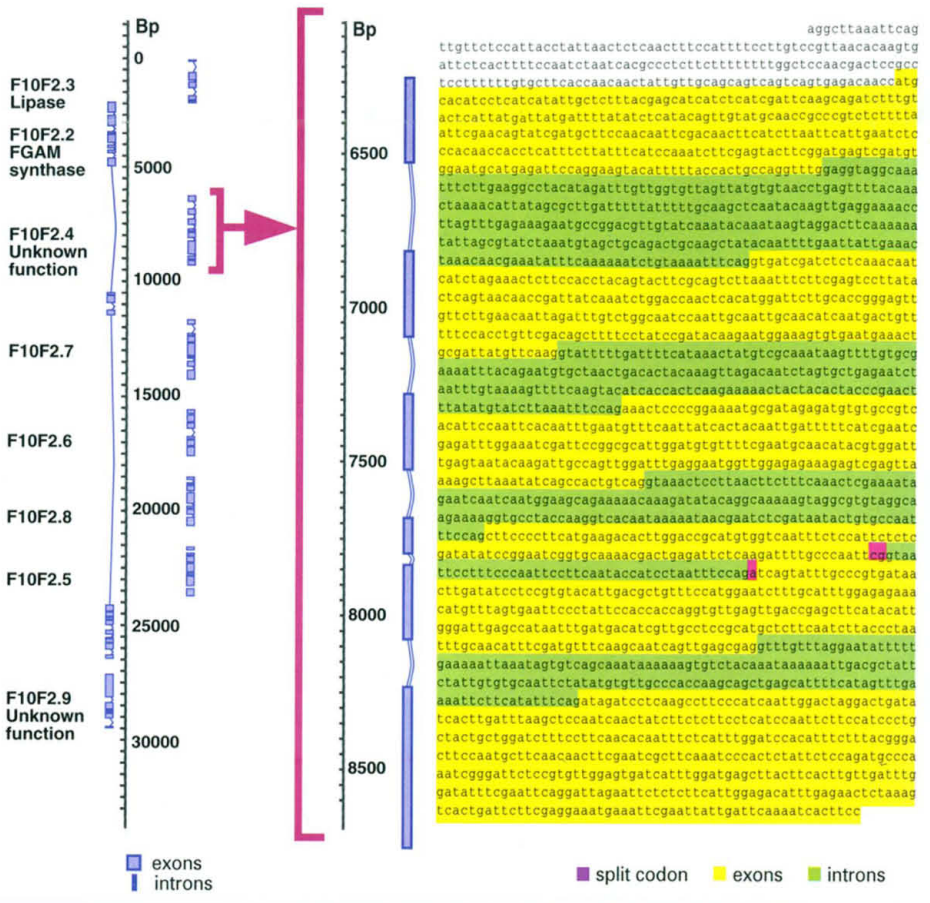
n

loping
ent precursors,
nction is a
C. elegans,
s and genetics,
ence
f networks of
opment. This
ion and
motor neurons
at are part of the
legans. Each of
cord
ne, stereotypic
neurons
ding to the
this pattern can

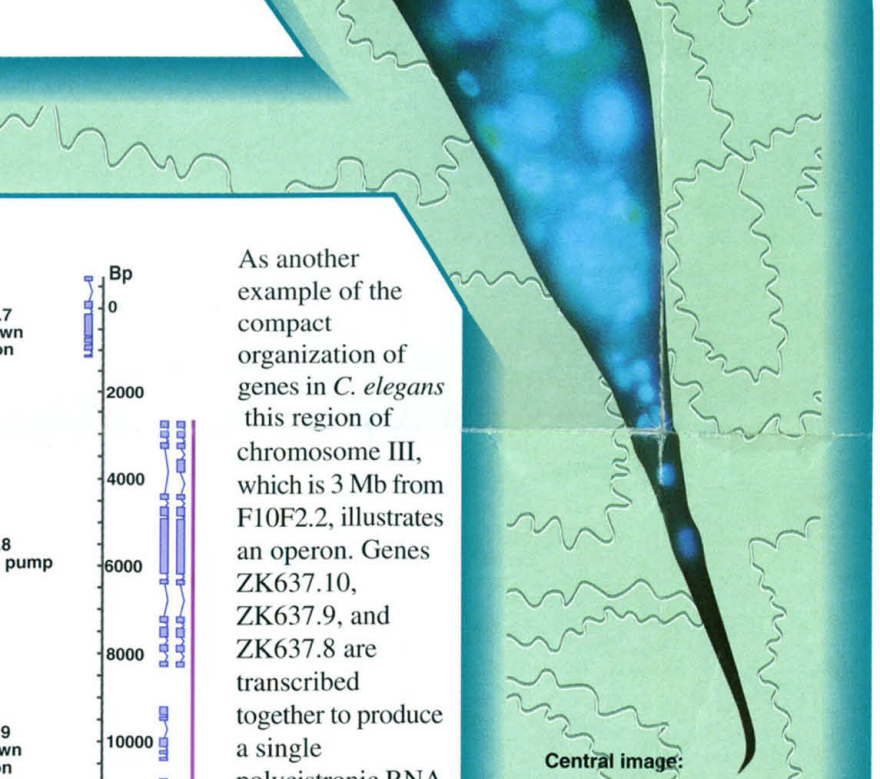


additional mapped clones are available.

A fragment has been expanded to show the five genes within the large introns of gene F10F2.2. These genes are present on the strand opposite to that of F10F2.2. The portion containing the gene F10F2.4 has been expanded to the resolution of the DNA sequence.



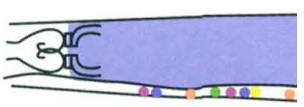
Credits: *Science* **Coordinator:** Barbara R. Jasny ■ **Authors:** Martin Chalfie, Columbia University, New York, NY, USA; Sean Eddy, Washing H.A. Plasterk, Netherlands Cancer Institute, Amsterdam, The Netherlands; Robert H. Waterston, Washington University School of Medicine, St Louis, M Nowoslawski. ■ **Photographic Credits:** Central image, Morris F. Maduro and David B. Pilgrim, University of Alberta, Edmonton, Alberta, Canada. I University Press, UK]; (panel B), Ronald Plasterk, Netherlands Cancer Institute, Amsterdam, The Netherlands; Expression Patterns, (panel A) Hiroaki T University, New York, NY [from S. Mitani *et al.*, *Development* **119**, 773 (1993); with permission, The Company of Biologists, LTD, UK]. Mutant Phenoty



Central image:
C. elegans that has been stained blue with DAPI and contains an *unc-119*::GFP fusion (GFP, green fluorescent protein). The figure has been composited and retouched.

As another example of the compact organization of genes in *C. elegans* this region of chromosome III, which is 3 Mb from F10F2.2, illustrates an operon. Genes ZK637.10, ZK637.9, and ZK637.8 are transcribed together to produce a single polycistronic RNA molecule. After transcription, the polycistronic message is processed to form individual mRNA molecules for each of the three genes. ZK673.8 forms two alternatively-spliced products.

All differentiated cells in developing organisms arise from multipotent precu and control of cell fate and function is central problem of biology. In *C. elega* availability of cell lineage maps and ge combined with increased sequence information, allows analysis of networ interactions that control development. example illustrates the formation and differentiation of five types of motor ne (VA, VB, VC, AS, and VD) that are par central nervous system in *C. elegans*. E the 10 precursor cells (ventral cord neuroblasts) divides in the same, stereo way. The fates of the resulting neurons (bottom diagram) differ according to th patterns of genes they express; this patt be affected by the position of the precu cell in the worm. Early determination fate results in part from regional expres two of the homeobox genes *lin-39* and Mutations in the genes *unc-59* and *unc* block cell divisions within the lineage s not all of the mature motor neurons are produced. Some cells become VB or V if they are in one location, but if they a positioned elsewhere they will undergo programmed cell death, a process that r *ced-3* and *ced-4*.The functions of the m motor neurons, are governed by the exp of specific genes, including those repre in the table.



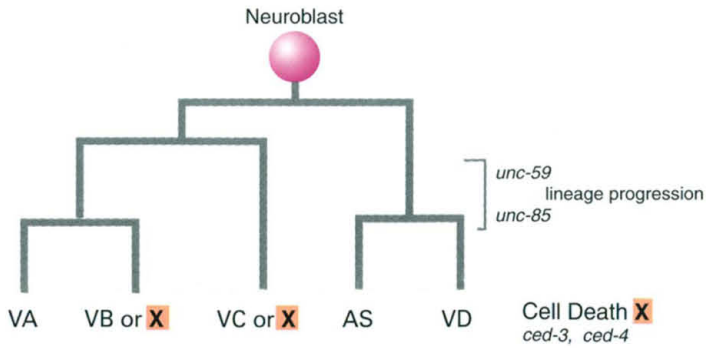
Further information on *C. elegans* can be found in the accompanying article (J. Hodgkin *et al.*) in the 20 October, 1995 issue of *Science*.

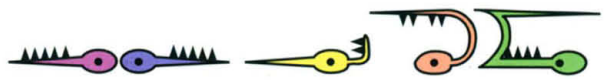
dy, Washington University School of Medicine, St. Louis, MO, USA; Michael O. Hengartner, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; Jonath St Louis, MO, USA; John G. White, University of Wisconsin, Madison, WI, USA. ■ **Reviewers:** Tom Blumenthal, Indiana University, Bloomington, IN, USA; Wil a, Canada. Predicted Genes, Sean Eddy, Washington University School of Medicine, St. Louis, MO, USA. Forward and Reverse Genetics, (panel A) Craig C. Mello A) Hiroaki Tabara and Yuji Kohara, National Institute of Genetics, Mishima, Japan; (panel B) Hongping Du, Columbia University, New York, NY, USA; (panel C tant Phenotype, William G. Wadsworth, Robert Wood Johnson Medical School, Piscataway, NJ, USA.

loping
ent precursors,
nction is a
C. elegans,
os and genetics,
ence
f networks of
opment. This
ion and
motor neurons
at are part of the
legans. Each of
cord
ne, stereotypic
neurons
ding to the
this pattern can
he precursor
nination of cell
al expression of
-39 and *mab-5*.
and *unc-85*
lineage so that
urons are
VB or VC cells
if they are
undergo
ess that requires
of the mature
y the expression
ose represented

lin-39

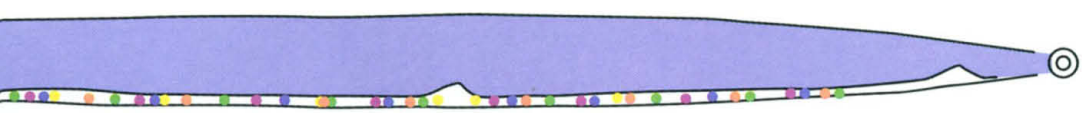
mab-5





Gene	Function*	VA	VB	VC	AS	VD
unc-30	Cell fate determination					+
unc-4	Synaptic specificity	+				
unc-5	Neuronal guidance				+	+
unc-2	Voltage-sensitive Ca channel			+		
unc-25	GABA synthesis					+
cha-1	Production of ACh	+	+	+	+	
unc-17	Storage of ACh	+	+	+	+	
unc-104	Synaptic vesicle location	+	+	+	+	+
snt-1	Synaptic release	+	+	+	+	+
unc-18	Synaptic release	+	+	+	+	+

*GABA, γ-aminobutyric acid; ACh, acetylcholine



USA; Jonathan Hodgkin, Medical Research Council Center, Cambridge, UK; Yuji Kohara, National Institute of Genetics, Mishima, Japan; Ronald N, USA; William B. Wood, University of Colorado, Boulder, CO, USA. ■ **Art**: DIRECTOR, Amy Decker Henry; DESIGN and ILLUSTRATION, Susan ig C. Mello,University of Massachusetts, Worcester, MA, USA [from C.C. Mello *et al.*, *EMBO J.* **10**, 3959 (1991); with permission, the Oxford A; (panel C) Andres Villu Maricq, University of California at San Francisco, CA, USA; (panel D), Hongping Du and Martin Chalfie, Columbia ssociation for the Advancement of Science.