A Local, High-Density, Single-Nucleotide Polymorphism Map Used to Clone *Caenorhabditis elegans cdf-1*

Janelle Jakubowski and Kerry Kornfeld

Department of Molecular Biology and Pharmacology, Washington University School of Medicine, St. Louis, Missouri 63110

Manuscript received March 29, 1999
Accepted for publication July 2, 1999

ABSTRACT

Ras-mediated signaling is required for induction of vulval cell fates during *Caenorhabditis elegans* development. By screening for suppressors of the multivulva phenotype caused by constitutively active *let-60 ras*, we identified the mutation *n2527*. To clone the gene affected by *n2527*, we developed a method for high-resolution mapping. We took advantage of the genomic DNA sequence of the N2 strain by using DNA sequencing to scan for single-nucleotide polymorphisms (SNPs) at defined genomic positions of the RC301 strain. An average of one polymorphism per 1.4 kb was detected in predicted intergenic regions. Because of this high frequency, DNA sequencing is an efficient method to scan for SNPs. By alternating between identifying SNPs and mapping *n2527* using selected recombinants, we generated an SNP map of progressively higher density. An intensive search for SNPs resulted in a local map with an average marker spacing of ~4 kb. This was used to map *n2527* to a 9.6-kb interval. The small size of this interval made it feasible to use DNA sequencing to identify the molecular lesion. In principle, this approach can be used for high-resolution mapping of any *C. elegans* mutation. Furthermore, this approach can be applied to other species as the genomic sequence becomes available. The *n2527* mutation affects a previously uncharacterized gene that we named *cdf-1*, as it encodes a predicted protein with significant similarity to members of the cation diffusion facilitator family.

SCREENING for mutants is an important method that has been used to identify genes that mediate a wide variety of biological processes. Because cloning these genes is essential for a comprehensive analysis, improved cloning methods are extremely useful. Mutations are usually induced in *Caenorhabditis elegans* using ethyl methanesulfonate (EMS) or other chemical mutagens that primarily cause single-base substitutions (Brenner 1974; Coulondre and Miller 1977); by contrast to gross DNA changes or the insertion of a transposable element, these subtle changes do not facilitate cloning.

Because the genetic and physical maps of the *C. elegans* genome are well characterized, a gene affected by a chemically induced mutation is typically identified using a positional cloning approach that involves the following three phases:

1. The mutation is positioned on the physical map (Coulson *et al.* 1988). This defines an interval that contains the gene.
2. Transgenic animals containing genomic DNA from this interval cloned in cosmid or YAC vectors are generated, and assays for rescue of the mutant phenotype are conducted (Mello *et al.* 1991). This approach is used to search for a DNA fragment that contains the mutated gene and then to define a minimal rescuing fragment.
3. Candidate open reading frames (ORFs) are sequenced positioned on the minimal rescuing fragment using DNA from mutant animals to identify the nucleotide change that causes the mutant phenotype.

In practice, precise mapping reduces the difficulty of identifying a rescuing fragment, and precise definition of a minimal rescuing fragment reduces the difficulty of identifying the nucleotide alteration.

A mutation is initially positioned on the physical map relative to genome-wide systems of markers. These include mutations that cause a phenotype and affect a cloned gene, deletions with endpoints that can be characterized (Barstead *et al.* 1991), and polymorphic-sequence-tagged sites caused by Tc1 transposons present in *C. elegans* strains that have diverged from the wild-type N2 strain (Williams *et al.* 1992; Korswagen *et al.* 1996). After genome-wide systems of markers are utilized, the interval containing the mutation is often relatively large, and additional markers that could be used for further mapping are desirable. However, the genome-wide systems of markers cannot usually be expanded in defined genomic regions because these markers are typically identified in genome-wide searches. Additional markers have been identified in defined genomic regions by using the physical map to search for

Corresponding author: Kerry Kornfeld, Department of Molecular Biology and Pharmacology, Washington University School of Medicine, 660 S. Euclid, St. Louis, MO 63110.
E-mail: kornfeld@molecool.wustl.edu

restriction fragment length polymorphisms (RFLPs; Ruvkun *et al.* 1989; Hodgkin 1993; Kornfeld *et al.* 1995a,b; Kimura *et al.* 1997; Ogg *et al.* 1997). While this approach has been extremely useful, it has two significant limitations. First, it is likely that many polymorphisms cannot be detected as RFLPs because they do not affect a restriction enzyme site or grossly alter the DNA. Second, the mapping resolution that can be achieved with RFLPs is limited because the precise position of the nucleotide change(s) responsible for an RFLP is generally unknown. Because of these limitations, mutations are not generally mapped to high resolution, and a significant effort is often required for the transformation rescue phase of positional cloning.

Single-nucleotide polymorphisms (SNPs), a term we will use to refer to the substitution or insertion/deletion of one or a small number of nucleotides, appear to be the most common type of polymorphism in vertebrates (Kwok *et al.* 1996). Significant efforts are being directed at the development of high-throughput approaches to scan for and score SNPs (Lai *et al.* 1998; Landegren *et al.* 1998; Wong *et al.* 1998). To develop a method for high-resolution mapping in *C. elegans*, we took advantage of the complete genomic sequence by using DNA sequencing to scan for SNPs at defined genomic positions. We found that SNPs are abundant in the RC301 strain compared to the N2 strain. By alternating between scanning for SNPs and mapping, we generated a local SNP map with an average interval size of ~4 kb and mapped the *cdf-1(n2527)* mutation to a 9.6-kb interval. The molecular lesion in the affected gene was identified by sequencing this interval. This approach is likely to facilitate the positional cloning of any gene that is affected by a mutation in *C. elegans*, and may be applicable to other organisms as genomic sequence becomes available.

MATERIALS AND METHODS

General methods and strains: *C. elegans* strains were cultured as described by Brenner (1974) and grown at 22.5° unless otherwise noted. The parent strain of all the mutant strains was N2, a wild-type isolate from Bristol, England. RC301 is a wild-type isolate from Freiburg, Germany (Hodgkin 1993). *C. elegans* strains maintained in the laboratory are inbred by self-fertilization and homozygous at essentially all loci. Unless otherwise noted, the mutations used in this study are described by Riddle *et al.* (1997) and are as follows. LGIV: *let-60(n1046gf); dpy-20(e1282)*. LGX: *lon-2(e678); mup-2(n2346ts); cdf-1(n2527)* (this study); *unc-6(e78); unc-10(e102); dpy-7(e88); uDf1*.

Genetic analyses: We previously described a screen for suppressors of the *let-60(n1046gf)* multivulva (Muv) phenotype (Lackner *et al.* 1994; Kornfeld *et al.* 1995a,b; Jacobs *et al.* 1998). In brief, we mutagenized *let-60(n1046gf)* hermaphrodites with EMS, placed 2794 F_1 self-progeny on separate Petri dishes, and examined the F_2 self-progeny for non-Muv animals at 22.5°. We identified 33 independently derived mutations that reduced the penetrance of the Muv phenotype from 93% to <10%, including the *n2527* mutation. Ten additional suppressor mutations that meet these criteria were identified in a related screen (Beitel *et al.* 1990).

The suppression of the *let-60(gf)* Muv phenotype caused by *n2527* displayed linkage to *lon-2* on chromosome *X* (data not shown). The seven other suppressor mutations that are positioned on chromosome *X* complemented the *n2527* suppression of *let-60(gf)* Muv phenotype, indicating that the complementation group defined by *n2527* contains only one mutation (data not shown). Three-factor crosses were used to more precisely map *n2527*. Of uncoordinated (Unc) non-dumpy (Dpy) self-progeny from *let-60(gf); unc-10 dpy-6/ n2527* hermaphrodites, 0/10 segregated *n2527*. From *let-60(gf); unc-6 dpy-7/n2527* hermaphrodites, 0/17 Unc non-Dpy self-progeny and 17/18 Dpy non-Unc self-progeny segregated *n2527*. These data suggest *n2527* is positioned left of *unc-10* and *unc-6*. Of Unc non-muscle positioning abnormal (non-Mup) self-progeny from *let-60(gf); mup-2 unc-6/n2527* hermaphrodites, 18/25 segregated *n2527*. These data suggest *n2527* is positioned between *mup-2* and *unc-6*, an ~0.6-map-unit interval (Figure 1A).

To generate a *lon-2 n2527 unc-6* chromosome, we first identified a *lon-2 n2527* chromosome by selecting long (Lon) non-Muv non-Unc self-progeny of *let-60(gf); lon-2 unc-6/n2527* hermaphrodites. Second, we identified a *n2527 unc-6* chromosome by selecting Unc non-Muv non-Mup self-progeny of *let-60(gf); mup-2 unc-6/n2527* hermaphrodites. Third, we identified a *lon-2 n2527 unc-6* chromosome by selecting Lon self-progeny of *let-60(gf); lon-2 n2527/n2527 unc-6* hermaphrodites and identifying an animal that segregated Unc progeny.

To map *n2527* relative to SNPs, we mated RC301 males and *let-60(gf); lon-2 n2527 unc-6* hermaphrodites, placed cross-progeny on separate Petri dishes, and then identified self-progeny of genotype *let-60(gf); lon-2 n2527 unc-6/RC301*. We picked 201 Lon non-Unc self-progeny to separate Petri plates, identified self-progeny homozygous for the recombinant chromosome, and scored the penetrance of the Muv phenotype in these strains.

Polymerase chain reaction and DNA sequencing: Unless otherwise noted, molecular biology techniques were performed as described by Sambrook *et al.* (1989). The genomic DNA sequence of the N2 strain (*C. elegans* Sequencing Consortium 1998) was used to design oligonucleotide primers that were used to scan for polymorphisms and sequence the region between *amP4* and *amP11*. Table 1 shows oligonucleotide primers that can be used to identify polymorphisms. Genomic DNA was prepared and PCR amplified as described by Williams *et al.* (1992). Amplification products were fractionated by electrophoresis in low-melting-temperature agarose, purified using β-agarase (New England Biolabs, Beverly, MA), and used as templates for cycle sequencing reactions that were analyzed using an automated ABI 373A sequencer (Applied Biosystems, Foster City, CA). DNA sequence data were analyzed using the Sequencher 3.0 computer program (Gene Codes Corp., Ann Arbor, MI).

DNA cloning: To subclone the region containing the predicted open reading frame C15B12.7, we digested cosmid C15B12 (received from A. Coulson, Sanger Center) with *Spe*I and *Sac*I to generate a 6239-bp DNA fragment that extends from 1612 bp upstream of the predicted START codon to 384 bp downstream from the predicted STOP codon of C15B12.7; this fragment contains no coding sequences from adjacent predicted genes. This fragment was ligated into pBluescript digested with *Spe*I and *Sac*I to create pJJ4. To delete the majority of the C15B12.7 open reading frame, we digested pJJ4 with *Xba*I, purified the plasmid backbone, and ligated to recircularize. This plasmid was named pJJ5. This procedure removed a 3704-bp *Xba*I fragment that contains predicted exons 2–7 of C15B12.7. The *n2527* mutation, a G-to-A transition at cosmid

TABLE 1

Polymorphisms detected by DNA sequencing

Polymorphism[a]	Cosmid[b]	Primer sequence[c] (5′ → 3′)	N2 product size (bp)[d]	Sequence at site of polymorphism[e]			Type of change[f]
				N2	RC301	*n2527*	
amP5	T22E5	GAA AGA CTC TGT GAA GAT CAG CGT ACA ATG CGT TTC ATC TG	1059	22807-TTACC	TTCCC	N2	S
amP4	C15B12	GTG TAA TAC TGT AGT TCA TGG GCA AAA TGC CTC AGT TCC G	955	34363-TCCAG	TCAAG	N2	S
amP14	C15B12	ACC GAA TGT TTG CAG GTA GG TTG TTG GTG TAT CCA AAG TGG	1041	37455-AATTT	AA-TT	N2	D
amP15	C15B12	ATC CAG TAG TCG CAC ATT GG CCA CAG AAA CTA TCA GCG ATG	1274	39073-TTGAA	N2	TTAAA	S
amP9	C15B12	CCG TCG TCA ACT GCC GTC G ACG TGT GTT CGT ATG TGT GC	1245	41695-ATATT	ATGTT	N2	S
amP12[g]				41961-CAAAA	CAGAA	N2	S
amP10	C15B12	CCT CCA CTT AAC TCA AGA GC CTG CGA GCA AAT CTA CTT CC	1161	43943-GGAAG	GGTAG	N2	S
amP16	F22A3	GGA CGG AAT ATG TTG TAC AG CGT TGA AGA AAG TGG TTG TG	1181	2635-TGCTT	N2	TGTTT	S
amP11	F22A3	AAG GAA TGA GTG AGA AGG AG CTT GTT GCT CAG TGT CGT CG	1059	9019-GTTTTT	GTTACG[h]	N2	I
amP8	F22A3	GTC TTT GCC AAC GAT AAG CG AGT ACA CAA CCG TCC ATA AAG	981	18227-TTAAA	TTTAA	N2	S
amP7	T14E8	CTT CAC TCG ACT CTA CTG CTC GTC ACC GCA ACT GAT AAA CG	1080	14106-GCCAG	GCAAG	N2	S
amP6	T28B4	ACT CCA TTG CAG ACG ACA TG CCA GAT GAC AAC ACT CAG C	1298	12179-AAGTT	AATTT	N2	S
amP13	F38B6	GGC AGT TAC TAG GTT TCA TC CTT CAG GGC CGT ATG CTA AC	1219	24233-AATAG	AATTAG	N2	I
amP3[i]				24606-TTAGGG	TTGCGG	N2	S
amP2	R03E9	CCA ATG TGA CCA TCT TCT CG CTC CGT CAC TCA AGT GCA TC	1185	14595-GTTTAA	GT--AA	N2	D
amP1	R03E9	AGA GTT CTA CTG GGT TGA CG GCC AAC TTA TAT CTG CGG GC	1205	32254-TCCAA	TCAAA	N2	S

[a] Following the standardized *C. elegans* nomenclature, the DNA sequence of the N2 strain is defined as wild type, and a sequence difference in the RC301 or *n2527* strain was assigned an allele name (Horvitz *et al.* 1979). This name is composed of a laboratory designation (*am*), an indication that the difference is physical and does not cause a visible phenotype (*P*), and a number. Polymorphisms are ordered from the left arm of chromosome *X* (top) to the center (bottom).

[b] Names of cosmids containing N2 genomic DNA that were sequenced by the *C. elegans* Sequencing Consortium.

[c] Oligonucleotide primers were designed on the basis of the corresponding cosmid DNA sequence.

[d] A pair of oligonucleotide primers will PCR amplify a product of the indicated size in base pairs (bp) from N2 genomic DNA.

[e] The genomic sequence of N2 was determined by the *C. elegans* Sequencing Consortium. Numbers refer to the position in the cosmid of the first nucleotide listed. We determined the sequence of RC301 and a strain containing the *n2527* mutation [*let-60(gf); lon-2 n2527 unc-6*]. N2 indicates that the sequence was identical to the N2 strain.

[f] S, substitution; I, insertion; D, deletion.

[g] *amP12* and *amp9* were present in the same amplified product.

[h] *amP11* is an insertion of ~300 bp; only the first three nucleotides are shown.

[i] *amP3* and *amP13* were present in the same amplified product.

position 39081, was engineered into pJJ4 by replacing a 1427-bp *Nco*I (cosmid position 38840)/*Kas*I (cosmid position 40267) with the equivalent *Nco*I/*Kas*I fragment derived from the PCR-amplified DNA of *let-60(gf); n2527* animals. This plasmid was named pJJ6. DNA sequencing of the *Nco*I/*Kas*I fragment of pJJ6 revealed no additional changes compared to wild type.

Germ-line transformation experiments: Germ-line transformation experiments were performed as described by Mello *et al.* (1991). We coinjected *let-60(gf); n2527* mutants with plasmid pRF4 (80–90 μg/ml), which contains the dominant mutation *rol-6(su1006)*, and cosmid C15B12 (50 μg/ml) or plasmids pJJ4, pJJ5, or pJJ6 (10–20 μg/ml). We established an independently derived transgenic strain from each F_1 animal that displayed the Rol phenotype and segregated F_2 progeny that displayed the Rol phenotype. If >60% of the Rol animals from a transgenic strain were Muv, then we concluded that the introduced DNA rescued the *n2527* suppression of *let-60(gf)* Muv phenotype.

RESULTS

Isolation of *n2527* and mapping using genome-wide systems of markers: The *C. elegans* vulva, a specialized epidermal structure used for egg laying and sperm entry, is formed by the descendants of P5.p, P6.p, and P7.p (reviewed by Horvitz and Sternberg 1991). In wild-type hermaphrodites, the anchor cell of the somatic gonad activates a conserved receptor tyrosine kinase-Ras-mitogen-activated protein (MAP) kinase signaling pathway in P6.p, causing P6.p to adopt the 1° vulval cell fate (eight descendants, reviewed by Kornfeld 1997). P6.p then signals to P5.p and P7.p, causing these cells to adopt the 2° vulval cell fate (seven descendants). P3.p, P4.p, and P8.p are capable of adopting vulval fates, but appear to receive neither of these signals; thus, these cells adopt the nonvulval 3° cell fate (two descendants). In hermaphrodites with a gain-of-function mutation that constitutively activates the *let-60 ras* gene, P3.p, P4.p, and P8.p often inappropriately adopt vulval cell fates; the resulting ectopic tissue forms a series of protrusions along the ventral side of the animal, which is called the multivulva (Muv) phenotype. To identify genes that are involved in this Ras-mediated signaling event, we screened for mutations that suppress the *let-60(n1046gf)* Muv phenotype (see materials and methods). We isolated 43 mutations that define 21 complementation groups, including the mutation *n2527*.

The *n2527* mutation is an effective suppressor of the phenotype caused by constitutively active Ras, since it reduced the penetrance of the Muv phenotype from 84 to 2% (Table 2). The *n2527* mutation is weakly semidominant. Mutants containing *n2527* in *trans* to a deficiency displayed a phenotype similar to mutants homozygous for *n2527* (Table 2), indicating that *n2527* is a loss-of-function mutation. In a wild-type genetic background, *n2527* did not cause a significant penetrance of vulval defects or other visible phenotypes.

To clone the gene affected by *n2527*, we first positioned it on the genetic and physical maps using mutations that cause visible phenotypes and affect cloned genes. The *n2527* mutation displayed linkage to chromosome *X*, and three-factor mapping experiments indicated that *n2527* is positioned between *mup-2* and *unc-6* (Figure 1A; see materials and methods). An analysis of the completed genomic DNA sequence of this region indicates that *mup-2* and *unc-6* are separated by ~450 kb (*C. elegans* Sequencing Consortium 1998). We did not utilize *odr-10*, the one remaining visible marker in this interval that has been cloned, because *odr-10* mutations cause a behavioral phenotype that is not readily scored.

TABLE 2

Genetic analysis of *cdf-1(n2527)*

Genotype[a]	% Muv[b]	*n*[c]
N2	0	Many
let-60(gf)	84	519
let-60(gf); cdf-1(n2527)	2	580
let-60(gf); cdf-1(n2527)/+[d]	65	468
let-60(gf); cdf-1(n2527)/uDf1[e]	8	384
let-60(gf); +/uDf1[f]	87	396
let-60(gf); cdf-1(n2527); amEx28[g]	83	303
let-60(gf); cdf-1(n2527); amEx29[g]	8	257
let-60(gf); cdf-1(n2527); amEx31[g]	16	332

[a] *let-60(gf)* refers to *let-60(n1046)*.

[b] Percentage of adult hermaphrodites that displayed the multivulva phenotype, one or more protrusions displaced from the position of the vulva. Unless otherwise noted, animals were raised at 20°.

[c] *n*, number of hermaphrodites examined.

[d] Non-Lon non-Mup non-Unc self-progeny of *let-60(gf) dpy-20; lon-2 cdf-1(n2527)/mup-2 unc-6* hermaphrodites.

[e] Non-Lon self-progeny of *let-60(gf) dpy-20; lon-2 cdf-1 (n2527)/uDf1* hermaphrodites. The deficiency *uDf1* is likely to delete *cdf-1* because it fails to complement *unc-6* (Savage *et al.* 1989) and *mup-2* (data not shown), which are positioned right and left of *cdf-1*, respectively. *uDf1* does not delete *lon-2*, and *uDf1* homozygotes die as embryos (Savage *et al.* 1989).

[f] Non-Lon self-progeny of *let-60(gf) dpy-20; lon-2/uDf1* hermaphrodites.

[g] *amEx28*, *amEx29*, and *amEx31* are extrachromosomal arrays that contain the transformation marker pRF4 and pJJ4, pJJ5 and pJJ6, respectively (see materials and methods; Figure 2). Multiple, independently derived, transgenic lines were generated in each case; these data are from one representative line of each genotype. Transgenic animals were raised at 22.5°.

Identification of SNPs at defined genomic positions: We next considered two approaches: generating a collection of transgenic animals containing fragments of genomic DNA spanning this interval cloned in cosmid vectors to identify a fragment that can rescue the *n2527* mutant phenotype, or additional mapping to more precisely position *n2527* on the physical map. We rejected the first approach because it is relatively laborious and there was a chance that it would not succeed, since some of the DNA in the interval was not present in cosmid vectors. To pursue the second approach, we needed to

identify many additional markers in this interval, since only two polymorphisms had been reported. We hypothesized that we could exploit our knowledge of the genomic DNA sequence of this interval to scan for SNPs at defined positions by sequencing small fragments of DNA from an evolutionarily diverged wild-type strain. Compared to scanning for RFLPs, this approach has the advantages that any sequence difference can be detected and the position of the polymorphism is known precisely. However, SNPs must occur relatively frequently for this approach to be practical.

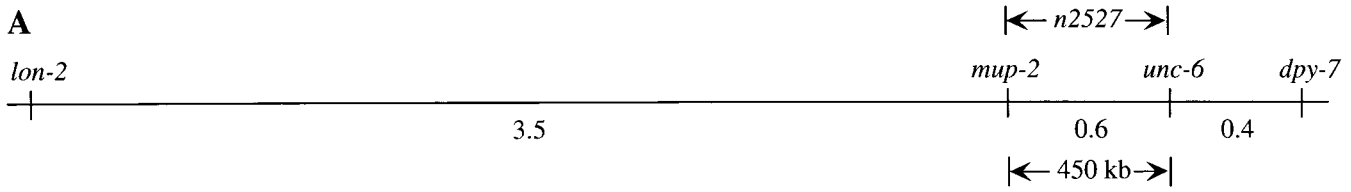
The *n2527* mutation was generated in a strain derived from the N2 strain, a wild-type isolate from England (Brenner 1974). We chose to look for polymorphisms in the RC301 strain, a wild-type isolate from Germany, since this strain has been used successfully to identify RFLPs in multiple regions of the genome (Hodgkin 1993; Kornfeld *et al.* 1995a,b; Kimura *et al.* 1997; Ogg *et al.* 1997). Our standard approach to identify a polymorphism was to use the genomic sequence of the N2 strain to design a pair of oligonucleotide primers that amplify a DNA fragment at a defined genomic position. In the first phase of the mapping experiment, primers were designed to amplify predicted intergenic regions, since we reasoned that polymorphisms would be more frequent in noncoding sequence. In the second phase, predicted exons and introns were also scanned to maximize the density of SNPs. The primers were composed of 20 bases and ≥50% G/C. Products were amplified using DNA from both RC301 and the N2-derived *let-60(n1046gf); lon-2 n2527 unc-6* strain, purified from agarose gels, and used as templates for sequencing using the amplification primers and an ABI automated sequencer. We designed primers that were separated by ~1.2 kb to fully exploit the capability of the automated sequencer—up to 600 bases per primer.

Generation of a progressively higher density SNP map by alternating between identifying SNPs and mapping *n2527*: In general, a crossover that occurs between *n2527* and a polymorphism can be used to determine marker order. To select crossovers near *n2527*, we used *lon-2* and *unc-6*, visible markers that flank *n2527* and are separated by ~1900 kb (Figure 1A). Although genetic mapping positioned *n2527* right of *mup-2*, the more distal marker *lon-2* was used to select recombinants because the *mup-2* phenotype is not suitable for this procedure. From *let-60(n1046gf); lon-2 n2527 unc-6/RC301* animals, we selected 201 Lon non-Unc self-progeny and then identified hermaphrodites homozygous for the recombinant chromosome. A total of 164 strains displayed the Muv phenotype, indicating that they lost *n2527* and the crossover occurred between *lon-2* and *n2527*, while 37 strains displayed the non-Muv phenotype, indicating that they contained *n2527* and the crossover occurred between *n2527* and *unc-6*. To define a small interval that contains *n2527*, we pursued two goals: First, the identification of the recombinants in which the crossover breakpoint occurred closest to the left and right of *n2527*; second, the identification of polymorphisms positioned close to but outside of these closest crossovers. These polymorphisms define an interval that contains *n2527*.

To identify a polymorphism close to and left of *n2527*, we first scanned for an SNP near *mup-2* to distinguish crossovers that occurred right of *mup-2* and, thus, close to *n2527* from crossovers that occurred left of *mup-2*. Two pairs of primers were used to amplify and sequence ~2 kb of DNA; one A-to-C substitution was detected in RC301 and designated *amP5* (Table 1). *amP5* is 2 kb left of *mup-2*. DNA sequencing was used to score *amP5* in 145 Lon non-*n2527* non-Unc recombinants; 140 recombinants contained *amP5* (the RC301 sequence), indicating that these crossovers occurred left of *amP5*, whereas 5 had the wild-type (N2) sequence, indicating that these crossovers occurred right of *amP5* (Figure 1, B and C). These results identify five crossovers between *amP5* and *n2527*, and they indicate that *n2527* is right of *amP5*. We next scanned for a polymorphism positioned ~90 kb right of *amP5* by sequencing ~1.4 kb of DNA; one C-to-A substitution was detected in RC301 and designated *amP4* (Table 1). DNA sequencing was used to score *amP4* in the five recombinants with crossovers right of *amP5;* three contained *amP4* and two had the wild-type sequence (Figure 1, B and C). These results identify two crossovers that occurred between *amP4* and *n2527*, and they indicate that *n2527* is right of *amP4*. Having established that *n2527* is right of *amP4*, we scored *amP4* in the remaining 19 Lon non-*n2527* non-Unc recombinants to determine if any of these crossovers occurred right of *amP4*. However, all 19 contained *amP4* (the RC301 sequence; Figure 1, B and C).

To identify a polymorphism close to and right of *n2527*, we scanned for polymorphisms left of *unc-6*. Four pairs of primers were used to sequence ~4.2 kb of DNA, and four SNPs were detected in RC301: a C-to-A substitution designated *amP1*, a 2-bp deletion designated *amP2*, an AG-to-GC substitution designated *amP3*, and a 1-bp insertion designated *amP13* (Table 1). *amP3* is positioned ~170 kb left of *unc-6*. DNA sequencing was used to score *amP3* in the 37 Lon *n2527* non-Unc recombinants; 16 had *amP3* and 21 had the wild-type sequence (Figure 1, B and C). These results identify 16 crossovers that occurred between *n2527* and *amP3*, and they indicate that *n2527* is left of *amP3*. Since *amP1*, *amP2*, and *amP13* are close to or right of *amP3*, we did not score these polymorphisms. We next scanned for polymorphisms left of *amP3*. About 3.6 kb of DNA was sequenced, and two SNPs were detected in RC301: a G-to-T substitution designated *amP6*, and a C-to-A substitution designated *amP7* that is positioned 133 kb left of *amP3* (Table 1). A total of 10 of the 16 recombinants had *amP7*, and 6 had the wild-type sequence (Figure 1, B and C). Thus, *n2527* is left of *amP7*. We next sequenced ~2 kb of DNA and identified an A-to-T substi-

A



B

Genotype of Heterozygote	Selected Recombinant		Sequence at Position of Polymorphism (N, N2-derived; R, RC301) *amP5*	*amP4*	*amP14*	*amP15*	*amP10*	*amP16*	*amP11*	*amP8*	*amP7*	*amP3*
let-60(gf); lon-2 n2527 unc-6 / RC301	Lon non-Unc Muv	140/164	R	nd	nd	nd	nd	nd	nd	nd	nd	nd
		19/164	nd	R	nd	nd	nd	nd	nd	nd	nd	nd
		3/164	N	R	nd	nd	nd	nd	nd	nd	nd	nd
		1/164	N	N	R	nd	nd	nd	nd	nd	nd	nd
		1/164	N	N	N	R	nd	nd	nd	nd	nd	nd
	Lon non-Unc non-Muv	4/37	nd	nd	nd	N	N	R	R	R	R	R
		6/37	nd	nd	nd	nd	nd	nd	nd	N	R	R
		6/37	nd	nd	nd	nd	nd	nd	nd	nd	N	R
		21/37	nd	nd	nd	nd	nd	nd	nd	nd	nd	N

C

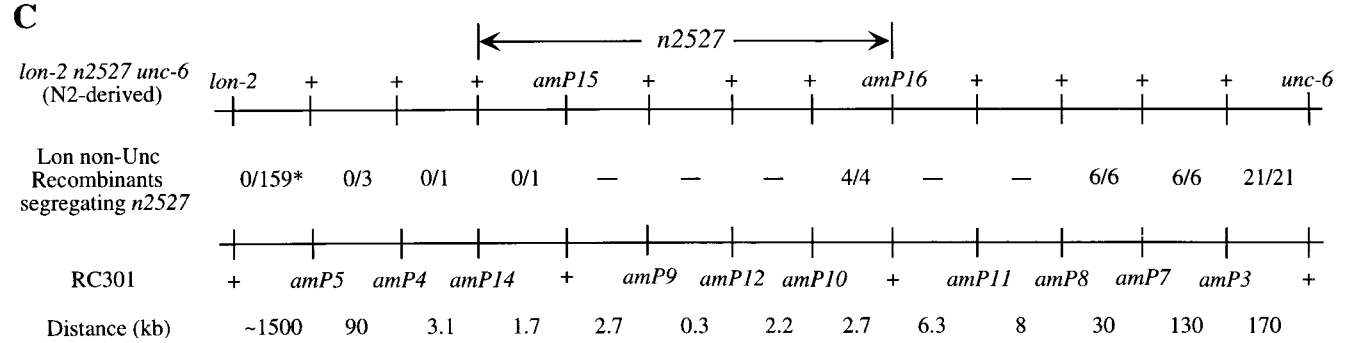


Figure 1.—High-resolution mapping of *n2527* relative to SNPs. (A) A portion of the genetic map of chromosome *X*. Loci defined by mutations that cause visible phenotypes are indicated above, and approximate distances in map units are shown below. Three-factor mapping experiments were used to position *n2527* between *mup-2* and *unc-6*. (B) A total of 201 Lon non-Unc self-progeny were selected from the indicated heterozygote. A total of 164 strains displayed the Muv phenotype (top), indicating they lack *n2527*, and 37 strains displayed the non-Muv phenotype (bottom), indicating they contain *n2527*. Each group was subdivided by analyzing the DNA sequence at the position of SNPs. N indicates the sequence of the N2-derived *lon-2 n2527 unc-6* strain, and R indicates the sequence of the RC301 strain. Scoring *amP9* gave the same results shown for *amP10*. nd, not determined. (C) Horizontal lines represent the physical map; vertical lines indicate cloned genes and polymorphisms (not drawn to scale). The approximate distance between markers is shown in kilobases. +, wild-type N2 sequence. Polymorphisms are changes in the RC301 strain (shown below), except *amP15* and *amP16*, which are changes in the *lon-2 n2527 unc-6* strain (shown above); at these two positions, the RC301 strain has the wild-type (+) sequence. The data shown in B were used to determine the number of crossovers that occurred in each interval. These data position *n2527* between *amP14* and *amP16*. *, of these 159 strains, 19 were scored at *amP4* but not *amP5* (B, line 2); thus, the crossovers in these strains may have occurred between *amP5* and *amP4*.

tution in RC301, designated *amP8*, that is positioned 28 kb left of *amP7* (Table 1). A total of 4 of the 10 recombinants had *amP8*, and 6 had the wild-type sequence (Figure 1, B and C). These data indicate that *n2527* is left of *amP8*, and they identify four crossovers that occurred between *n2527* and *amP8*. We next sequenced ~1.5 kb and identified an insertion of ~300 bp in RC301, designated *amP11*, that is positioned 8 kb left of *amP8*. Because this insertion is relatively large, it could be scored by gel electrophoresis of PCR products. All four recombinants had *amP11*, indicating that *n2527* is left of *amP11* (Figure 1, B and C). To summarize, of 164 crossovers that occurred between *lon-2* and *n2527*, we identified 2 that occurred between *amP4* and *n2527*. Of 37 crossovers that occurred between *n2527* and *unc-6*, we identified 4 that occurred between *n2527* and *amP11*. Thus, *amP4* and *amP11* define an ~19-kb interval that contains *n2527*.

Generation of a local, high-density SNP map and identification of the *n2527* molecular lesion: Thus far, we had

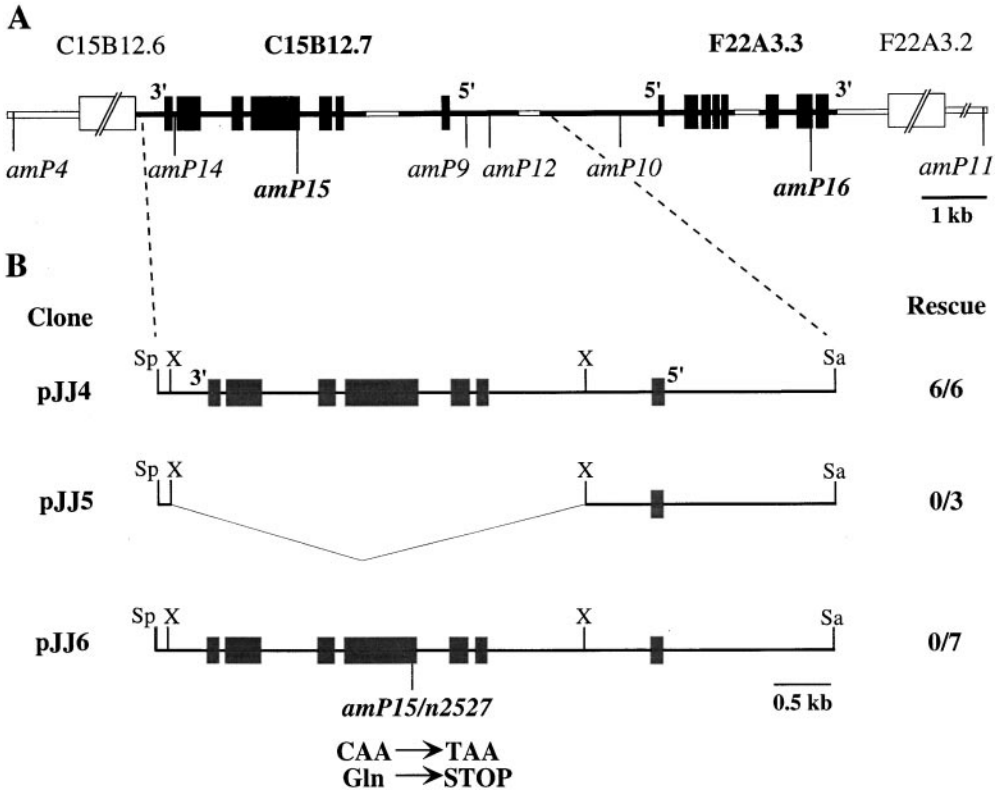


Figure 2.—The gene affected by the *n2527* mutation is predicted ORF C15B12.7. (A) The horizontal line represents the physical map; thick boxes are predicted exons, and thin lines are predicted introns or intergenic regions. Except for regions marked by diagonal lines, the map uses the scale shown below. The four predicted open reading frames in this region are labeled above; the 5′ and 3′ ends of C15B12.7 and F22A3.3 are indicated. We determined the DNA sequence of regions shown in black using DNA from the *n2527* and RC301 strains; vertical lines indicate the positions of SNPs. Polymorphisms in regular type were identified in RC301; *amP15* and *amP16* were identified in the *n2527* strain (shown lower in bold type). (B) Plasmids listed on the left contained the indicated fragment of genomic DNA. Vertical lines indicate restriction enzyme cleavage sites used for cloning: Sp, *Spe*I; X, *Xba*I; Sa, *Sac*I. pJJ5 lacks the *Xba*I fragment, and pJJ6 contains a C-to-T substitution that changes predicted codon 186 from Gln to STOP. The number of independently derived transgenic strains that displayed rescue of the *n2527* suppression of *let-60(gf)* Muv phenotype and the total number of strains analyzed are indicated (Table 2; see materials and methods).

generated an SNP map of progressively higher density centered on the *n2527* mutation. We next used DNA sequencing to pursue two goals: first, the generation of an extremely high density map that could be used for further mapping; second, the identification of the *n2527* molecular lesion. The 19-kb interval between *amP4* and *amP11* is predicted to contain four ORFs by the GeneFinder computer program. We determined the sequence of the coding regions and introns of the two centrally positioned, predicted ORFs, designated C15B12.7 and F22A3.3, as well as most of the intergenic region using DNA from both RC301 and the *let-60(n1046gf); lon-2 n2527 unc-6* strain (Figure 2A). After analyzing ~2.8 kb of sequence from coding regions, we identified two polymorphisms in the *n2527* strain: a G-to-A substitution designated *amP15* that changes the predicted codon 186 of C15B12.7 to STOP, and a C-to-T substitution designated *amP16* that affects a codon in exon 7 of F22A3.2 but does not result in an amino acid change (Table 1 and Figure 2A). These nucleotide changes were candidates for the *n2527* molecular lesion. After analyzing ~6.1 kb of sequence from introns and intergenic regions, we identified four SNPs in RC301: a 1-bp deletion designated *amP14*, an A-to-G substitution designated *amP9*, an A-to-G substitution designated *amP12*, and an A-to-T substitution designated *amP10* (Table 1 and Figure 2A).

We used the polymorphisms from both strains to further position *n2527*. We analyzed *amP14* in the two recombinants that had crossovers between *amP4* and *n2527*; one contained *amP14* and one had the wild-type sequence (Figure 1, B and C). These results indicate that *n2527* is right of *amP14*, and they identify one crossover that occurred between *amP14* and *n2527*. We analyzed *amP16* in the four recombinants that had crossovers between *n2527* and *amP11*. All four contained the RC301 allele, indicating that *n2527* is left of *amP16* (Figure 1, B and C) and that *amP16* is not the *n2527* molecular lesion. These results positioned *n2527* in a 9.6-kb interval defined by *amP14* and *amP16* that contains portions of two predicted ORFs, C15B12.7 and F22A3.3 (Figure 2A). The remaining four polymorphisms in this interval, *amP15*, *amP9*, *amP12*, and *amP10*, were not separated from *n2527* by any of the five closest crossovers (Figure 1, B and C). These results do not establish the order of *n2527* relative to these polymorphisms.

The following evidence suggests that *amP15* is the *n2527* mutation: (1) *amP15* is positioned in the 9.6-kb interval between *amP14* and *amP16* that contains the *n2527* mutation, and *amP15* was not separated from *n2527* by any crossover; (2) no other changes were detected in the *n2527* strain after determining the sequences of all the predicted exons and most of the

introns and intergenic regions between *amP14* and *amP16;* (3) *amP15* creates a premature STOP codon at predicted amino acid 186 of C15B12.7, resulting in a truncated protein lacking the C-terminal two-thirds of the protein; and (4) *amP15* is a G-to-A transition, the most common mutation induced by EMS (Coulondre and Miller 1977), the mutagen used to generate *n2527*.

DNA containing C15B12.7 rescues the *n2527* mutant phenotype in transgenic animals: To test the hypothesis that the predicted ORF C15B12.7 is the gene affected by the *n2527* mutation, we used a transformation rescue assay. *let-60(n1046gf); n2527* animals, which are non-Muv, were transformed with cosmid C15B12 and a plasmid that contains a dominant *rol-6* mutation as a transformation marker. Six independently derived transgenic strains that displayed the Rol phenotype were obtained. All six strains displayed the Muv phenotype, indicating that C15B12 rescued the *n2527* suppression of the Muv phenotype (data not shown). To determine if predicted ORF C15B12.7 is sufficient to rescue the *n2527* mutant phenotype, we constructed a plasmid (pJJ4) that contains the complete predicted ORF C15B12.7 and no other predicted ORFs. All six independently derived transgenic strains containing pJJ4 displayed the Muv phenotype, indicating that C15B12.7 is sufficient for rescuing activity (Figure 2B and Table 2). Plasmids containing a deletion of predicted exons 2–7 of C15B12.7 (pJJ5) or the base change detected in the *n2527* strain at predicted codon 186 of C15B12.7 (pJJ6) were used as controls. All three independently derived transgenic strains containing pJJ5 displayed the non-Muv phenotype (Figure 2B and Table 2), indicating that an intact version of C15B12.7 is necessary for rescuing activity. All seven independently derived transgenic strains containing pJJ6 displayed the non-Muv phenotype (Figure 2B and Table 2). This failure to rescue indicates that a nonsense change at predicted codon 186 reduces gene activity, and it supports the hypothesis that this G-to-A nucleotide substitution is the *n2527* mutation.

The *C. elegans* Sequencing Consortium noted that the predicted ORF C15B12.7 encodes a protein that is similar to *Saccharomyces cerevisiae* COT1. Paulsen and Saier (1997) analyzed the sequence of the predicted C15B12.7 protein, COT1, and 11 other similar proteins that comprise the cation diffusion facilitator (CDF) protein family. This analysis showed that the predicted C15B12.7 protein has significant similarity to all the members of this family and contains all six predicted transmembrane domains that are characteristic of CDF proteins. Thus, we named the gene defined by the *n2527* mutation *cdf-1*. Several CDF proteins have been shown to regulate the intracellular concentration of heavy metal ions (Kamizono *et al.* 1989; Conklin *et al.* 1992; Palmiter and Findley 1995; Palmiter *et al.* 1996). Interestingly, CDF proteins have not been reported to be involved in Ras-mediated signaling. Further characterization of *cdf-1* is required to determine whether CDF-1 regulates the concentration of heavy metal ions and thereby modulates Ras signaling, or whether CDF-1 has a different mechanism of action.

DISCUSSION

SNPs are abundant in RC301: The data presented here can be used to estimate the abundance and nature of polymorphisms in the strain RC301 compared to N2. We identified 13 polymorphisms in ~18 kb of predicted intergenic DNA—an average of one polymorphism per 1.4 kb. One polymorphism was detected in 3.7 kb of predicted intron DNA, and no polymorphisms were detected in 2.8 kb of predicted exon DNA. The polymorphisms are primarily subtle changes—substitutions or insertions/deletions of one or two nucleotides—although one is a 300-bp insertion. It is likely that polymorphisms occur at a similar frequency throughout the RC301 genome, since RFLPs have been detected in many genomic regions of RC301 and we have begun identifying SNPs at a similar frequency on chromosome *II* (Hodgkin 1993; Kornfeld *et al.* 1995a,b; Kimura *et al.* 1997; Ogg *et al.* 1997; J. Jakubowski and K. Kornfeld, unpublished observations). The frequency of SNPs in other strains of *C. elegans* has yet to be determined. The *C. elegans* Sequencing Consortium has initiated a large-scale effort to identify randomly positioned SNPs in the strain CB4856, which was isolated in Hawaii (http://genome.wustl.edu/gsc/). Although many candidate SNPs have been identified, it is difficult to estimate the frequency on the basis of these data, since candidate SNPs have not yet been verified.

These findings have two important implications. First, they demonstrate that DNA sequencing is a practical and efficient method to scan for SNPs at defined genomic positions. Based on an average of one polymorphism per 1.4 kb, only about three oligonucleotide primers and three sequencing reactions are necessary to detect a polymorphism in RC301. Second, these results suggest that a genome-wide search for SNPs could yield a map containing thousands of markers. Such a map would have a significantly higher marker density than the existing genome-wide polymorphism maps, which contain several hundred polymorphisms caused by insertions of Tc1 transposable elements (Williams *et al.* 1992; Korswagen *et al.* 1996). The *C. elegans* Sequencing Consortium has initiated such an effort. A genome-wide, moderate-density SNP map could be used to position newly identified mutations to a reasonably small chromosomal interval using unselected recombinants. Such a map could also provide a starting point for the generation of a local, high-density SNP map that can be used for high-resolution mapping with selected recombinants.

A method for high-resolution mapping: The method presented here involves three main steps: The first is

the generation of recombinants with crossovers near the mutation. We used *cis*-linked visible markers to select recombinants. In our experience, this is the most laborious step. Second, a progressively higher density SNP map is generated by alternating between identifying SNPs at defined genomic locations and mapping. This is both efficient and relatively rapid. By identifying SNPs in the center of the interval, relatively few are required; we used six SNPs to narrow the 450-kb interval between *mup-2* and *unc-6* to a 19-kb interval between *amP4* and *amP11*. The effort necessary to analyze recombinants decreases with each new SNP as the useful crossovers that occurred closely to the mutation are identified. The third step involves the generation of a local, high-density SNP map and the simultaneous search for the molecular lesion.

Mapping resolution depends on the density of crossovers and markers (in this case, SNPs). We analyzed 201 crossovers between *lon-2* and *unc-6*. This interval is ~1900 kb, and thus we can calculate that crossovers occurred on average every 9.5 kb. We generated a local map with an average interval between SNPs of ~4 kb, considering only RC301 polymorphisms between *amP14* and *amP11* (Figure 1C). In the region from *amP14* to *amP16* that was investigated intensely, the average interval between SNPs was ~2.4 kb (Figure 1C). These reagents were used to map the *n2527* mutation to a 9.6-kb interval. This resolution is much finer than what has been reported previously in *C. elegans*. Because the local SNP map was denser than the average interval between crossovers, it was not surprising that the mapping resolution was limited by the density of crossovers rather than by the density of SNPs: three polymorphisms in RC301 could not be positioned relative to *n2527*. Since *C. elegans* has an average gene density of about one gene per 5 kb (*C. elegans* Sequencing Consortium 1998), these results indicate that this method may make it possible to map many mutations to a single gene.

High-resolution mapping is useful and important because it significantly reduces the difficulty of subsequent cloning steps. Because *n2527* was mapped to an interval of only 9.6 kb, it was practical to identify the molecular lesion by DNA sequencing and bypass the need for the standard procedure of transformation of mutant worms with genomic DNA to identify a rescuing fragment. This is important because transformation can be laborious and is prone to both false-negative and false-positive results. It is likely that this high-resolution mapping can be used to analyze any *C. elegans* mutation. High-resolution mapping will be particularly useful for positionally cloning genes identified by mutations that cannot be rescued by injection of wild-type DNA, *e.g.*, gain-of-function mutations or mutations that affect genes that function in the germ line, a tissue in which transformed genes are not expressed efficiently. Furthermore, this approach is likely to be useful in other organisms, such as Drosophila and zebrafish, as genomic sequences become available.

We thank Jennifer Boots and Drew Syder for assistance with mapping and sequencing and Tim Schedl and Steve Johnson for advice about the manuscript. Some strains were provided by the Caenorhabditis Genetics Center (St. Paul, MN), which is funded by the National Center for Research Resources of the National Institutes of Health. This research was supported by a grant from Monsanto-Searle/Washington University Biomedical Program (K.K.). K.K. is a recipient of the Burroughs Wellcome Fund New Investigator Award in the Basic Pharmacological Sciences.

LITERATURE CITED

Barstead, R. J., L. Kleiman and R. H. Waterston, 1991 Cloning, sequencing, and mapping of an α-actinin gene from the nematode *Caenorhabditis elegans*. Cell Motil. Cytoskeleton **20:** 69–78.

Beitel, G. J., S. G. Clark and H. R. Horvitz, 1990 *Caenorhabditis elegans ras* gene *let-60* acts as a switch in the pathway of vulval induction. Nature **348:** 503–509.

Brenner, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71–94.

C. elegans Sequencing Consortium, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012–2018.

Conklin, D. S., J. A. McMaster, M. R. Culbertson and C. Kung, 1992 *COT1*, a gene involved in cobalt accumulation in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:** 3678–3688.

Coulondre, C., and J. H. Miller, 1977 Genetic studies of the *lac* repressor. IV. Mutagenic specificity in the *LacI* gene of *Escherichia coli*. J. Mol. Biol. **117:** 577–606.

Coulson, A., R. Waterston, J. Kiff, J. Sulston and Y. Kohara, 1988 Genome linking with yeast artificial chromosomes. Nature **355:** 184–186.

Hodgkin, J., 1993 Molecular cloning and duplication of the nematode sex-determining gene *tra-1*. Genetics **133:** 543–560.

Horvitz, H. R., and P. W. Sternberg, 1991 Multiple intercellular signalling systems control the development of the *Caenorhabditis elegans* vulva. Nature **351:** 535–541.

Horvitz, H. R., S. Brenner, J. Hodgkin and R. K. Herman, 1979 A uniform genetic nomenclature for the nematode *Caenorhabditis elegans*. Mol. Gen. Genet. **175:** 129–133.

Jacobs, D., G. J. Beitel, S. G. Clark, H. R. Horvitz and K. Kornfeld, 1998 Gain-of-function mutations in the *Caenorhabditis elegans lin-1* ETS gene identify a C-terminal regulatory domain phosphorylated by ERK MAP kinase. Genetics **149:** 1809–1822.

Kamizono, A., M. Nishizawa, Y. Teranishi, K. Murata and A. Kimura, 1989 Identification of a gene conferring resistance to zinc and cadmium ions in the yeast *Saccharomyces cerevisiae*. Mol. Gen. Genet. **219:** 161–167.

Kimura, K. D., H. A. Tissenbaum, Y. Liu and G. Ruvkun, 1997 *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science **277:** 942–926.

Kornfeld, K., 1997 Vulval development in *Caenorhabditis elegans*. Trends Genet. **13:** 55–61.

Kornfeld, K., K.-L. Guan and H. R. Horvitz, 1995a The *Caenorhabditis elegans* gene *mek-2* is required for vulval induction and encodes a protein similar to the protein kinase MEK. Genes Dev. **9:** 756–768.

Kornfeld, K., D. B. Hom and H. R. Horvitz, 1995b The *ksr-1* gene encodes a novel protein kinase involved in Ras-mediated signaling in C. elegans. Cell **83:** 903–913.

Korswagen, H. C., R. M. Durbin, M. T. Smits and R. H. A. Plasterk, 1996 Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. Proc. Natl. Acad. Sci. USA **93:** 14680–14685.

Kwok, P.-Y., Q. Deng, H. Zakeri, S. L. Taylor and D. A. Nickerson, 1996 Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. Genomics **31:** 123–126.

Lackner, M. L., K. Kornfeld, L. M. Miller, H. R. Horvitz and S. K. Kim, 1994 A MAP kinase homolog, *mpk-1*, is involved in

ras-mediated induction of vulval cell fates in *Caenorhabditis elegans*. Genes Dev. **8:** 160–173.

Lai, E., J. Riley, I. Purvis and A. Roses, 1998 A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. Genomics **54:** 31–38.

Landegren, U., M. Nilsson and P.-Y. Kwok, 1998 Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res. **8:** 769–776.

Mello, C. C., J. M. Kramer, D. Stinchcomb and V. Ambros, 1991 Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration. EMBO J. **10:** 3959–3970.

Ogg, S., S. Paradis, S. Gottlieb, G. I. Patterson, L. Lee *et al.*, 1997 The Forkhead transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. Nature **389:** 994–999.

Palmiter, R. D., and S. D. Findley, 1995 Cloning and functional characterization of a mammalian zinc transporter that confers resistance to zinc. EMBO J. **14:** 639–649.

Palmiter, R. D., T. B. Cole and S. D. Findley, 1996 ZnT-2, a mammalian protein that confers resistance to zinc by facilitating vesicular sequestration. EMBO J. **15:** 1784–1791.

Paulsen, I. T., and M. H. Saier, Jr., 1997 A novel family of ubiquitous heavy metal ion transport proteins. J. Membr. Biol. **156:** 99–105.

Riddle, D. L., T. Blumental, B. J. Meyer and J. R. Priess (Editors), 1997 *C. elegans II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ruvkun, G., V. Ambros, A. Coulson, R. Waterston, J. Sulston *et al.*, 1989 Molecular genetics of the *Caenorhabditis elegans* heterochronic gene *lin-14*. Genetics **121:** 501–516.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Savage, C., M. Hamelin, J. G. Culotti, A. Coulson, D. G. Albertson *et al.*, 1989 *mec-7* is a β-tubulin gene required for the production of 15-protofilament microtubules in *Caenorhabditis elegans*. Genes Dev. **3:** 870–881.

Williams, B. D., B. Schrank, C. Huynh, R. Shownkeen and R. H. Waterston, 1992 A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. Genetics **131:** 609–624.

Wong, D. G., J.-B. Fan, C.-J. Siao, A. Berno, P. Young *et al.*, 1998 Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science **280:** 1077–1082.

Communicating editor: I. Greenwald