Mol Gen Genet (1998) 260: 280–288 © Springer-Verlag 1998

ORIGINAL PAPER

**H. I. Stewart · N. J. O'Neil · D. L. Janke · N. W. Franz**
**H. M. Chamberlin · A. M. Howell · E. J. Gilchrist**
**T. T. Ha · L. M. Kuervers · G. P. Vatcher**
**J. L. Danielson · D. L. Baillie**

# Lethal mutations defining 112 complementation groups in a 4.5 Mb sequenced region of *Caenorhabditis elegans* chromosome III

Received: 23 April 1998 / Accepted: 18 August 1998

**Abstract** The central gene cluster of chromosome III was one of the first regions to be sequenced by the *Caenorhabditis elegans* genome project. We have performed an essential gene analysis on the left part of this cluster, in the region around *dpy-17III* balanced by the duplication *sDp3*. We isolated 151 essential gene mutations and characterized them with regard to their arrest stages. To facilitate positioning of these mutations, we generated six new deficiencies that, together with preexisting chromosomal rearrangements, subdivide the region into 14 zones. The 151 mutations were mapped into these zones. They define 112 genes, of which 110 were previously unidentified. Thirteen of the zones have been anchored to the physical sequence by polymerase chain reaction deficiency mapping. Of the 112 essential genes mapped, 105 are within these 13 zones. They span 4.2 Mb of nucleotide sequence. From the nucleotide sequence data, 920 genes are predicted. From a Poisson distribution of our mutations, we predict that 234 of the genes will be essential genes. Thus, the 105 genes constitute 45% of the estimated number of essential genes in the physically defined zones and between 2 and 5% of all essential genes in *C. elegans*.

**Key words** *Caenorhabditis elegans* · Essential genes · *sDp3* · Genome organization

Communicated by D. Finnegan

H. I. Stewart (✉) · N. J. O'Neil · D. L. Janke · N. W. Franz
H. M. Chamberlin[1] · A. M. Howell · E. J. Gilchrist · T. T. Ha
L. M. Kuervers · G. P. Vatcher · J. L. Danielson · D. L. Baillie
Institute of Molecular Biology and Biochemistry
Simon Fraser University, Burnaby, B.C.
Canada, V5A 1S6
Tel.: Fax: +1-604-291-4597

*Present address*:
[1]Department of Genetics, University of Washington
Seattle, WA 98195, USA

## Introduction

Understanding the mechanism by which contiguous blocks of genes contribute to the development and maintenance of an organism requires knowledge of the identity, organization and function of genes within the genome. Toward this goal, the genomes of a number of genetically amenable organisms are being sequenced in their entirety. The first metazoan to be completely sequenced will be the nematode, *Caenorhabditis elegans*. The present estimated completion date is 1998 (Wilson et al. 1994; Waterston and Sulston 1995). To date, almost all of the 100 Mb genome has been sequenced and 16 400 protein-encoding genes have been predicted from sequence data (Waterston et al. 1997; S. J. M. Jones, personal communication). While sequence data may allow for the identification of many of the genes, the sequence alone is not enough to determine their biological relevance. Approximately 40% of the predicted genes share significant sequence similarities with genes of known function in other organisms. Such comparative sequence analysis provides useful information but does not provide a complete understanding of how the information encoded in sequence data correlates with biological function. In addition, the remaining 60% of predicted genes either share no significant similarity with genes in the database or have similarity to genes of unknown function. To understand fully how the information encoded in the nucleotide sequence contributes to the development and maintenance of an organism, functions must be attributed to these genes. The analysis of genetic mutations in genes positioned within sequenced regions supplies this functional information. The information is useful not only for understanding the functions of the genes in the *C. elegans* genome but also provides insight into the functions of homologous genes in other organisms.

Such functional analysis has already begun on the sequenced genome of *Saccharomyces cerevisiae* (Goffeau et al. 1996). The techniques that are employed in func-

tional analysis in yeast are not available for the *C. elegans* genome as homologous recombination techniques have not been developed for use in *C. elegans*. However, many other techniques are available with which to perform functional analyses in *C. elegans*. One approach is the use of RNA interference to manipulate gene expression (Fire et al. 1998). This involves the microinjection of double-stranded RNA (dsRNA) complementary to the coding region of the gene to be studied. The dsRNA injection results in a decrease in production of the protein. However, this method does not always result in a detectable phenotype, and it is difficult to establish whether such negative results indicate that the gene, when interfered with, gives no obvious phenotype or the RNA interference method does not knockout the protein product of this gene. A second approach is mutational analysis, either random mutagenesis and subsequent positional cloning, or gene-selected mutagenesis. The selected mutagenesis of specific genes is often referred to as reverse genetics. Gene-selected mutagenesis employs the polymerase chain reaction (PCR) to identify deletions or Tc1 transposon insertions into a specific gene in DNA pooled from a library of mutagenized animals (Jansen et al. 1997). The PCR-based screening process allows the isolation of putative null mutants of specific genes. However, this method will likely be applied preferentially to genes that share sequence similarity with known genes. The random mutagenesis approach, on the other hand, does not rely on the identification of sequence homologies or similarities to known genes. The screen criterion determines what types of gene mutations will be isolated. In addition, random mutagenesis can identify novel, non-null mutations in genes. These mutations can provide functional information beyond that available if the function of the gene is eliminated.

To correlate the sequence data with biological function, we performed a mutagenesis screen for lethal mutations using the duplication *sDp3* as a balancer. Lethal mutations comprise the largest class of easily identified mutations in *C. elegans*. It is estimated that 25–33% of *C. elegans* genes are essential (Johnsen and Baillie 1997). Large collections of lethal mutations in essential genes exist for many regions of the *C. elegans* genome: the *unc-22* region of chromosome IV (Rogalski et al. 1982; Rogalski and Baillie 1985; Clark et al. 1988), the *dpy-10* to *rol-5* region of chromosome II (Sigurdson et al. 1984), the left portion of chromosome V balanced by the translocation *eT1 (III;V)* (Rosenbluth et al. 1988; Johnsen and Baillie 1991), and the left portion of chromosome I balanced by the duplication *sDp2* (Howell and Rose 1990; McKim et al. 1992). In this paper we report the generation of 151 lethal mutations defining 112 essential genes, as well as six new deficiencies, within a 7 map unit (mu), 4.5 Mb region around *dpy-17* on chromosome III. Only two of the genes have been previously identified genetically. The lethal mutations allow us to estimate the total number of essential genes in the *C. elegans* genome by extrapolation.

The *dpy-17* region was chosen for study because it was one of the first regions to be sequenced by the *C. elegans* Genome Sequencing Consortium and is now almost entirely sequenced. According to the 1997 *C. elegans* Genetic Map (Hodgkin et al. 1997), 75 genetic markers were available for the region, of which only 31 were correlated with the physical map. We have mapped the new lethal mutations to 14 distinct zones and have anchored 13 of these zones to the physical map using PCR deficiency mapping. This will facilitate the assignment of genetically identified genes to predicted gene sequences. Analyses of the mutant phenotypes will then assist in correlating biological function with sequence data. The mutational identification and physical positioning of 112 genes reported here is a significant contribution toward such a correlation in the 4.5 Mb region around *dpy-17*.

## Materials and methods

### Nematode strains and culture conditions

Nematodes were maintained and mated on petri plates containing Nematode Growth Medium, seeded with *Escherichia coli* OP50 (Brenner 1974). Nomenclature follows the uniform system adopted for *C. elegans* (Horvitz et al. 1979). The wild-type N2 strain var. Bristol and the mutant strains used were obtained from our own laboratory or through the *C. elegans* Genetic Stock Center unless otherwise stated. We define a recessive lethal mutation as one that when homozygous prevents the worm from developing into a fertile adult producing viable progeny [lethals, steriles, maternal-effect lethals (Mels)]. Maternal-effect lethal mutations are those that have no obvious effect on the homozygote, but that result in the progeny of the homozygote exhibiting a lethal phenotype. All experiments were carried out at 20° C. The duplications *sDp3* (Rosenbluth et al. 1985) and *sDp8* (Stewart et al. 1991) were obtained in previous studies.

### Isolation of lethal mutations

Our mutagenesis and screening protocol makes use of the fact that *sDp3* covers *dpy-17* but not *unc-32* and that it does not recombine with the normal LGIII chromosome. Thus homozygous lethal mutations in the covered region can be maintained in strains without being lost by recombination, i.e., they are "balanced" by *sDp3*.

$P_o$ *dpy-17(e164) unc-32(e189)/dpy-17(e164) lin-12(n941)III; sDp3(III;f)* hermaphrodites were phenotypically wild type since *sDp3* covers *dpy-17*. They were exposed to one of the following mutagen treatments. For ethyl methanesulfonate (EMS) mutagenesis, worms were suspended in 0.018 M EMS in M9 buffer for 4 h [as by Brenner (1974), but with a lower EMS concentration]. For ultraviolet (UV) irradiation mutagenesis, worms were irradiated with 110 J/$m^2$ UV irradiation in uncovered petri plates, from a distance of 43 cm using a 30 W Germacidal lamp (254 nm) preequilibrated for 15 min (Stewart et al. 1991). For gamma-ray mutagenesis, worms were irradiated with 2000 R, using a Cobalt-60 source for 8 min 26 s. In some cases of EMS mutagenesis, the *dpy-17 unc-32* marked chromosome included the mutation *ncl-1(e1865)* between the two markers.

After exposure to a mutagen, the worms were allowed to recover for 1–2 h before picking young gravid hermaphrodites (one per plate) and allowing these to lay eggs for 24 h. They were then transferred to fresh plates and after another 24 h the $P_o$ worms were removed. The recovery period was used to ensure that the in utero embryos present during mutagen treatment were expelled, and not included in the F1 samples.

For screening, wild-type F1s were individually allowed to self-fertilize. If no lethal mutation had been induced on the *dpy-17 unc-32* marked chromosome, they should have produced fertile Dpy-17 Unc-32s among their F2 progeny (i.e., *dpy-17 unc-32* homozygotes not carrying *sDp3*). Plates on which no more than two gravid Dpy Unc F2s were detected were retained, and Unc-32s were picked to fresh plates (i.e., *dpy-17 unc-32; sDp3*, with a putative lethal mutation on the marked chromosome). The presence of a lethal mutation was established if these failed to produce any fertile Dpy Uncs. These lethal mutations ("lethals") were presumed to be within the region covered by *sDp3*. They were maintained as homozygotes with *sDp3*, by picking Unc-32s.

Lethals to the left of *dpy-17* could have crossed off in the F1 heterozygote to produce some gravid Dpy Unc F2s (*let-x dpy-17 unc-32/+ dpy-17 unc-32*). Thus, those mutations that were more than 4 mu to the left were not expected to be recovered.

### Positioning lethal mutations relative to *dpy-17*

Lethal-bearing Unc-32 hermaphrodites were crossed to wild-type (N2) males and the F2 progeny from wild-type F1 hermaphrodites were scored. Recovery of fertile Dpy non-Unc recombinants indicated the lethal was to the right of *dpy-17*, while fertile Dpy Unc recombinants indicated it was to the left.

### Complementation tests

Tests between lethal mutations "*sx*" and "*sy*" isolated in this study were performed by first crossing Unc-32 hermaphrodites from the stock plates carrying lethal *sx* to N2 males. Phenotypically wild-type F1 males were then crossed to Unc-32s carrying lethal *sy*. The presence or absence of fertile Dpy Unc F2 hermaphrodites established either complementation or failure to complement, respectively. Other appropriate protocols were devised to test our lethals for complementation with *mup-4(ar60)* (Gatewood and Bucher 1997) and *let-972(ar65)* (both received from E. A. Bucher), visible morphological mutations and *nDf20*, as well as for coverage by *sDp8*, since these were in a variety of genetic backgrounds.

If two mutations isolated in this study complemented, the ratio of fertile Dpy Unc hermaphrodites/wild-type males was expected to be 0.5. However, the low viability of some of the deficiency heterozygotes often gave lower ratios even with EMS mutations mapping far away. Therefore establishing failure to complement required that there be no fertile Dpy Unc hermaphrodites in the presence of at least 30 wild-type males. On the other hand, in the case of lethals mapping to the left of *dpy-17*, crossing over could occur in the male heterozygote. Thus, + *dpy-17 unc-32/+ + +* sperm would give Dpy Unc hermaphrodite progeny even if the mutations failed to complement and appear as a false-positive complementation. Based on the recombination experiments with *dpy-17*, we calculated the ratio of fertile Dpy Unc hermaphrodites/wild-type males expected, for a given mutation, simply due to recombination. Complementation, then, was recorded as positive only if the ratio of Dpy Uncs hermaphrodites/wild-type males was larger than this ratio.

### Identification of deficiencies

UV- and gamma ray-induced lethals failing to complement two or more mutations that complement each other, were defined as deficiencies (*sDf*) deleting at least two genes.

### Establishing zones and identifying essential genes

Positioning with respect to *dpy-17*, together with complementation tests, mapped the lethal mutations with respect to chromosomal rearrangements (*sDfs*, *nDf20* and *sDp8*). The rearrangement breakpoints, thus determined, were refined relative to the physical map by PCR mapping (see below). This, as well as left-right positioning with respect to *dpy-17*, was used to divide the *sDp3*-covered region into separate zones (Fig. 1). Inter se complementation tests between lethals mapping to the same zone identified individual genes.

### Phenotypic analysis of EMS-induced lethal mutations

Unless otherwise noted in Table 1, animals were staged as follows: adult parental animals (*let-x dpy-17 unc-32; sDp3* or *let-x dpy-17 unc-32/+* males × *let-x dpy-17 unc-32; sDp3)* were placed on fresh plates. Three to four days later Dpy Unc animals were selected and prepared for observation under Nomarski optics as described by Sulston and Horvitz (1977). The animals were staged based on the extent of gonad development (Kimble and Hirsh 1979), and the presence or absence of cuticular alae, which are present in L1 and adult cuticle (Cox et al. 1981). The most mature animals were assumed to represent the mutant arrest stage.

### PCR mapping of deficiency endpoints

As homozygotes, in the absence of *sDp3,* the *sDf* deficiencies are embryonic lethals. Fifteen hermaphrodites of the genotype *dpy-17(e164) sDf x unc-32(e189); sDp3(III;f)* were transferred to a seeded plate and allowed to lay eggs for 8 h. Adult hermaphrodites were removed and the plates kept at 20° C for 48 h to allow viable eggs (carrying *sDp3*) to hatch. Eggs that did not hatch were considered to be of the genotype *dpy-17 sDf x unc-32*. Template DNA was extracted from homozygous deficiency eggs according to Barstead et al. (1991) with the modifications of Williams et al. (1992). For each reaction one or two eggs were treated for 3 min with chitinase and transferred to worm lysis buffer (50 mM KCl, 10 mM TRIS-HCl, pH 8.2, 2.5 mM $MgCl_2$, 0.45% NP-40, 0.45% Tween 20, 0.01% gelatin, 120 µg/ml Proteinase K). Eggs in lysis buffer were then digested at 65° C for 45 min and the enzyme inactivated at 95° C for 15 min. Thermo-cycling reactions were performed in an Idaho Technology model 1605 Air Thermo-cycler using "Low" (20 mM $Mg^{2++}$) 10× reaction buffer obtained from Idaho Technologies. *Taq* polymerase was obtained from BioCan Scientific. Nucleotide triphosphates were obtained from Pharmacia. Thermal profiles for all PCRs were: 1 min at 94° C, followed by 30 cycles of 10 s at 94° C, 20 s at 53°–60° C, 40 s at 72° C, followed by a final extension of 2 min at 72° C. Ramp speed was set at 9 (fastest). Primers were designed based on cosmid sequences available from GenBank. All PCRs were run in duplicate with positive control primers. PCR products were run on 1% agarose gels. Experimental primers that produced no PCR product indicate that the sequence from which the primer pairs were designed is deleted by the deficiency.

## Results

### Recovery and mapping of lethal mutations

From the UV- or gamma ray-treated $P_o$ hermaphrodites we retained 36 lethal mutations. These were left-right positioned with respect to *dpy-17*. Only those that mapped tightly to *dpy-17* or to its right were retained for complementation testing with EMS lethals.

The screening of 10 401 EMS-treated chromosomes for the presence of *sDp3*-balanced lethals yielded 151 mutations. Complementation tests with the retained UV and gamma-ray lethals identified six deficiencies: *sDfs 125, 127* and *130*, induced by UV irradiation; *sDfs 121, 128 and 135*, from the gamma-ray screen.

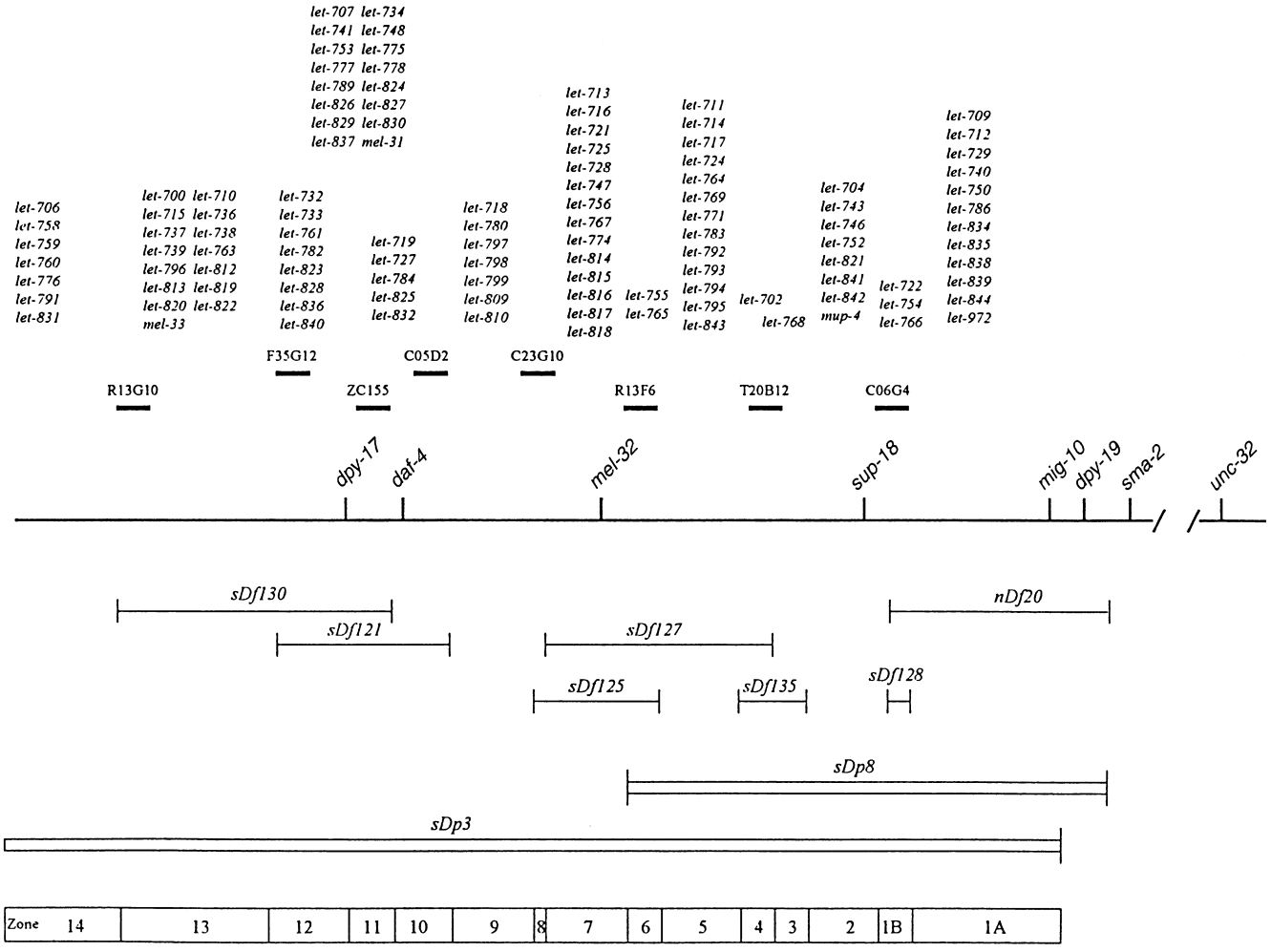


**Fig. 1** Partial genetic and physical map of the *dpy-17* region. Essential genes are clustered by zone. Genes are arbitrarily ordered within each zone by gene name. Cosmids shown are the outmost cosmids known to be deleted by the deficiency shown beneath. Scale is based on the physical positions of the cosmids (Pmap) as listed in the public database ACeDB (available by anonymous FTP from ncbi.nlm.nih.gov, directory repository/acedb). The deletion of *daf-4* by *sDf121* is based on personal communication from H. Honda and I. Mori. The positions of the other non-*let* genes as well as *nDf20* are based on ACeDB

The EMS-induced lethal mutations were positioned with respect to 14 chromosomal zones determined by rearrangement breakpoints and left-right positioning with respect to *dpy-17* (Fig. 1). Inter se tests within each zone then established that the 151 mutations represented 112 genes (Table 1). Of these, two had already been identified by others: *mup-4* (Gatewood and Bucher 1997), and *let-972* (E. A. Bucher, personal communication). Owing to the large number of lethal mutations generated, we did not perform extensive complementation tests against the 75 genetic markers previously identified. Therefore, it is possible that some of these lethal mutations may be allelic with existing lethal or visible mutations in this region. Thirty of the genes were hit multiple times in this study: twenty-two had two hits, seven had three, and one had five hits.

## Distribution of essential genes along the chromosome

The left *sDp3* breakpoint is between *daf-7* and *dpy-1*, while the right-hand one is between *mig-10* and *dpy-19* (Fig. 1). The duplication covers about 17 mu to the left of *dpy-17* and only 2 mu to its right.

Zone 14 is defined as the region balanced by *sDp3* and to the left of *sDf130*. The lethal gene closest to *dpy-17* in this region is only approximately 2 mu away from *dpy-17*. Thus zone 14 comprises about 15 mu or 80% of the balanced recombination distance. However, only 5% of the genes for which we recovered EMS-induced mutations fall into this zone. Furthermore, none of them map more than 5 mu from *dpy-17*. That is, they lie within the right-most 3 mu of zone 14. Two reasons probably account for this. First, as pointed out in Materials and methods, lethals to the left of *dpy-17* could have crossed off in the F1 heterozygotes during the screening protocol. Second, the already established genetic map shows a clustering of other genes in the *dpy-17* to *unc-32* region, relative to the rest of the left arm of chromosome III (Hodgkin et al. 1997).

Ninety-five percent of the essential genes are distributed among zones 1–13 (Fig. 1). Genes in zones 11 and 12 are deleted by both *sDf121* and *sDf130*, but are

**Table 1** Ethyl methane sulfonate induced lethal mutations recovered in the *dpy-17 III* region of *Caenorhabditis elegans*

| Zone[a] | Gene | Allele | Phenotype of lethal arrest[b] | Zone[a] | Gene | Allele | Phenotype of lethal arrest[b] | Zone[a] | Gene | Allele | Phenotype of lethal arrest[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1A | *let-709* | *s2444* | Embryo[c] | | *let-771* | *s2442* | Sterile | | *let-775* | *s2484* | Early larval |
| | *let-712* | *s2439* | Early larval | | *let-783* | *s2601* | Embryo[c] | | | *s2594* | Early larval |
| | | *s2598* | Early larval | | *let-792* | *s2798* | Early larval[c] | | | *s2840* | Early larval[c] |
| | | *s2841* | Early larval[c] | | *let-793* | *s2825* | Maternal effect[c] | | *let-777* | *s2469* | Late larval |
| | *let-729* | *s2579* | Embryo[c] | | *let-794* | *s2833* | Sterile[c] | | *let-778* | *s2584* | Early larval |
| | *let-740* | *S2624* | Sterile | | *let-795* | *s2849* | Sterile[c] | | *let-789* | *s2606* | Mid larval |
| | | *s2795* | Sterile[c] | | *let-843* | *s2627* | Maternal effect | | *let-824* | *s2793* | Late larval[c] |
| | *let-750* | *s2446* | Late larval | 6 | *let-755* | *s2600* | Mid larval | | *let-826* | *s2811* | Sterile[c] |
| | *let-786* | *s2585* | Early larval | | *let-765* | *s2575* | Mid larval | | | *s2820* | Sterile[c] |
| | | *s2631* | Early larval | | | *s2630* | Early larval | | *let-827* | *s2827* | Sterile[c] |
| | *let-834* | *s2815ts* | Sterile[c] | 7 | *let-713* | *s2449* | Late larval | | *let-829* | *s2826* | Early larval[c] |
| | *let-835* | *s2802* | Mid larval[c] | | | *s2470* | Late larval | | *let-830* | *s2828* | Late larval[c] |
| | *let-838* | *s2839* | Mid larval[c] | | *let-716* | *s2457* | Early larval | | *let-837* | *s2834* | Mid larval[c] |
| | *let-839* | *s2821* | Early larval[c] | | | *s2626* | Early larval | | *mel-31* | *s2438* | Maternal effect[c] |
| | *let-844* | *s2597* | Early larval[c] | | *let-721* | *s2447* | Maternal effect[c] | 12 | *let-732* | *s2611* | Early larval[c] |
| | *let-972* | *s2441* | Early larval | | | *s2812* | Late larval[c] | | | *s2808* | Maternal effect[c] |
| | | *s2583* | Early larval | | *let-725* | *s2454* | Sterile | | *let-733* | *s2467* | Sterile[c] |
| 1B | *let-722* | *s2448* | Early larval | | | *s2800* | Mid larval[c] | | | *s2621* | Mid larval |
| | | *s2450* | Early larval | | *let-728* | *s2573* | Mid larval | | *let-761* | *s2461* | Mid larval |
| | | *s2588* | Mid larval | | *let-747* | *s2456* | Mid larval | | *let-782* | *s2591* | Late larval |
| | | *s2816* | Early larval[c] | | | *s2807* | Early larval[c] | | *let-823* | *s2792* | Early larval[c] |
| | | *s2842* | Early larval[c] | | *let-756* | *s2613* | Early larval[d] | | *let-828* | *s2797* | Early larval[c] |
| | *let-754* | *s2622* | Mid larval | | | *s2809* | Early larval[d] | | *let-836* | *s2803* | Late larval[c] |
| | *let-766* | *s2463* | Late larval | | *let-767* | *s2464* | Late larval[c] | | *let-840* | *s2846* | Early larval[c] |
| 2 | *let-704* | *s2458* | Late larval[c] | | | *s2819* | Mid larval | 13 | *let-700* | *s2478* | Mid larval |
| | *let-743* | *s2581* | Early larval | | *let-774* | *s2615* | Early larval | | | *s2571* | Mid larval |
| | | *s2830* | Early larval[c] | | *let-814* | *s2829* | Mid larval[c] | | *let-710* | *s2572* | Mid larval |
| | *let-746* | *s2609* | Early larval | | *let-815* | *s2799* | Early larval[c] | | *let-715* | *s2618* | Sterile |
| | *let-752* | *s2619* | Mid larval | | *let-816* | *s2818* | Mid larval[c] | | *let-736* | *s2577* | Mid larval |
| | *let-821* | *s2804* | Early larval[c] | | *let-817* | *s2823* | Mid larval[c] | | | *s2590* | Late larval |
| | *let-841* | *s2814* | Sterile[c] | | *let-818* | *s2835* | Early larval[c] | | | *s2832* | Mid larval[c] |
| | *let-842* | *s2851* | Early larval[c] | 9 | *let-718* | *s2445* | Early larval | | *let-737* | *s2477* | Mid larval |
| | *mup-4* | *s2433* | Mid larval | | *let-780* | *s2576* | Sterile | | *let-738* | *s2435* | Embryo[c] |
| | | *s2440* | Embryo[c] | | *let-797* | *s2824* | Early larval[c] | | *let-739* | *s2593* | Early larval |
| | | *s2574* | Embryo[c] | | *let-798* | *s2850* | Embryo[c] | | *let-763* | *s2475* | Early larval |
| 3 | *let-768* | *s2482* | Sterile | | *let-799* | *s2852* | Early larval[c] | | | *s2580* | Early larval |
| | | *s2592* | Late larval[c] | | *let-809* | *s2844* | Early larval[c] | | *let-796* | *s2822* | Mid larval[c] |
| | | *s2628* | Late larval | | *let-810* | *s2836* | Early larval[c] | | *let-812* | *s2794* | Late larval[c] |
| 4 | *let-702* | *s2483* | Early larval[c] | 11 | *let-719* | *s2455* | Sterile | | *let-813* | *s2801* | Late larval[c] |
| | | *s2813* | Early larval[c] | | *let-727* | *s2589* | Early larval | | *let-819* | *s2843* | Early larval[c] |
| 5 | *let-711* | *s2473* | Early larval | | *let-784* | *s2602* | Late larval | | *let-820* | *s2845* | Sterile[c] |
| | | *s2474* | Mid larval | | *let-825* | *s2806* | Late larval[c] | | *let-822* | *s2848* | Mid larval[c] |
| | | *s2587* | Early larval | | *let-832* | *s2817* | Late larval[c] | | *mel-33* | *s2805* | Maternal effect[c] |

| | | | 11/12 | | | | 14 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *let-714* | *s2465* | Sterile | | *let-707* | *s2462* | Early larval | | *let-706* | *s2480* | Early larval |
| | *s2481* | Sterile | | *let-734* | *s2479* | Early larval | | *let-758* | *s2607* | Sterile |
| | *s2582* | Early larval | | | *s2837* | Early larval[c] | | *let-759* | *s2595* | Mid larval |
| *let-717* | *s2471* | Early larval | | *let-741* | *s2476* | Early larval | | *let-760* | *s2603* | Late larval[c] |
| *let-724* | *s2437* | Early larval[c] | | | *s2810* | Early larval[c] | | *let-776* | *s2453* | Mid larval |
| *let-764* | *s2443* | Mid larval | | *let-748* | *s2578* | Early larval[c] | | *let-791* | *s2838as* | Sterile[c] |
| | *s2616* | Early larval | | *let-753* | *s2605* | Early larval | | *let-831* | *s2853* | Early larval[c] |
| *let-769* | *s2432* | Late larval | | | *s2614* | Early larval | | | | |

[a] Refer to Fig. 1 for the location of each zone on the map
[b] Worms were characterized as described in Materials and methods unless otherwise noted
[c] Characterized by observation under dissecting microscope after 3–4 days
[d] Worms are very slow developers

distinguished from each other by virtue of their position with respect to *dpy-17*. Sixteen genes were not separated from *dpy-17* by recombination mapping (Fig. 1).

None of the genes in this study maps to zones 8 or 10. Physical markers define both zones. In the case of zone 8, the left-hand breakpoints of *sDf125* and *sDf127* are distinguishable. Three primer pairs were tested from C23G10. *sDf127* deletes the two sets of right-hand primers but not the left-hand primer pair, while *sDf125* deletes all three primer pairs. Neither *sDf125* nor *sDf127* deletes the primer pair from F42A10. In the case of zone 10, the primer pair from cosmid C05D2 distinguishes the right breakpoints of *sDf121* and *sDf130*. The set of primers is deleted by *sDf121* but not by *sDf130*.

Genes in zones 2 and 9 are to the right of *dpy-17*, but are not deleted by any of the deficiencies used in this study. Those in zone 2 are covered by *sDp8*, while those in zone 9 are not. This may imply that there are haplo-insufficient loci in these zones. However, *nDf17* deletes these regions and is viable as a heterozygote (Ellis et al. 1991). Assuming *nDf17* is not discontinuous, this suggests that there are no haplo-insufficient loci in the two zones.

### PCR deficiency endpoint mapping

The region in which mutations were recovered spans approximately 4.5 Mb of sequence. PCR deficiency mapping was used to anchor the genetically defined zones to the physical sequence map. All 14 zones were assigned physical breakpoints, with the closest cosmids from which primers were designed acting as physical markers. The data are tabulated in Table 2. For each *sDf* the cosmids at the ends are listed. "Out" refers to the primer sequence being present in the homozygous *sDf* embryo. "In" means that the sequence is deleted.

### Phenotypes

The developmental arrest stages of the homozygous EMS-induced lethal mutations included embryos, larvae, sterile adults, as well as some cases of Mel (Table 1). The larval lethals are the most common with 76% of the lethals recovered arresting as larvae. Generally, different alleles of the same gene arrest at approximately the same stage. In a few cases different alleles arrest at significantly different stages, e.g., *let-722* (zone 1A), *let-714* (zone 5) and *let-732* (zone 12). The distribution of the earliest arrest stages of the genes is displayed in Fig. 2.

## Discussion

We have recovered a set of 151 EMS-induced recessive lethal mutations under the balancer *sDp3*. These mutations identify 112 essential genes in a 7 mu, 4.5 Mb region to the left of *unc-32III*. Of these genes, only two

**Table 2** Polymerase chain reaction (PCR) deficiency mapping data. Numbers below cosmids correspond to base pair positions on Genbank cosmid entry

| Deficiency | Left | | Right | |
|---|---|---|---|---|
| | Out[a] cosmid | In[b] cosmid | Out cosmid | In cosmid |
| *sDf130* | C32A3<br>3416-4060 | R13G10<br>4233-5506 | C05D2<br>30876-31406 | ZC155<br>1227-1792 |
| *sDf121* | F10G11<br>891-1513 | F35G12<br>35188-35567 | F54E7<br>6725-7529 | C05D2<br>30876-31406 |
| *sDf127* | C23G10a<br>1778-2723 | C23G10b<br>2626-3117 | R13A5[c] | T20B12<br>6500-6838 |
| *sDf125* | F42A10<br>18575-19251 | C23G10a<br>1778-2723 | K04C2<br>3805-4495 | R13F6<br>4747-5611 |
| *sDf135* | B0361<br>10625-11306 | T20B12<br>6500-6838 | R13A5[c] | T20B12<br>6500-6838 |
| *sDf128* | C29E4<br>11250-12030 | C06G4<br>4540-5157 | F44B9<br>7092-7950 | C06G4<br>4540-5157 |

[a] "Out": the PCR site is not deleted by the deficiency
[b] "In": the site is deleted
[c] No Genbank entry. Primer sequence R13-F (5′-TTTCGGATAGGGGTCAGA-3′), R13-R(5′-GGTTTTCATTTCGTTCAT-3′)

were previously identified. Assuming that our region is representative of the entire *C. elegans* genome and that all essential genes are equally mutable, it is possible to estimate the total number of essential genes in *C. elegans* from our data. We predict from a Poisson distribution of our mutations in the region spanning zones 1–13, the zones with distinct physical boundaries, that this region contains a minimum of 234 essential genes. As all genes are not equally mutable this number is at best a rough estimate. The sequence data predict a total of 920 genes in the region. This indicates that at least 25% of these genes are essential under laboratory conditions. Extrapolating this value to the 16400 genes predicted in *C. elegans*, we estimate that there are approximately 4200 essential genes in the *C. elegans* genome. This number is not in accordance with the 2850 predicted by Johnsen and Baillie (1991), and is considerably higher than the 3300 estimated by Howell and Rose (1990) and the 3500 estimated by Clark (1990). Several factors may explain why our predicted number is much higher. It is possible that this region is not representative of the rest of the *C. elegans* genome or is dissimilar to the regions studied in other experiments. Genes (both those predicted by sequencing as well as genetically identified genes) are not distributed evenly across the chromosomes. There is an obvious clustering of genes in the central regions of the chromosomes and fewer genes on the arms. Such positional effects may therefore lead to different gene estimates when based on data from different regions.

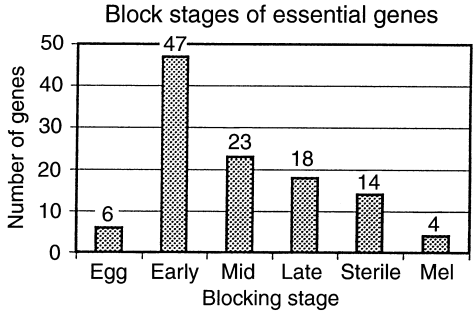


**Fig. 2** Histogram of the number of genes whose most severe alleles block development at the indicated developmental stages. The stages are egg lethal, early, mid, late larval lethal, sterile adults and maternal-effect lethal (*Mel*)

The distribution of essential genes across the *dpy-17* region is relatively uniform. Most zones have a comparable number of essential gene hits per unit of physical distance. This is due in part to the chromosomal position of the region studied. Almost all of the 14 zones are within the central cluster of chromosome III as defined by Barnes et al. (1995). The authors denoted *pal-1* as the left-most marker within the cluster; *pal-1* falls within zone 12. Therefore all essential genes to the right of *pal-1* are in the gene-rich cluster. This explains the scarcity of essential genes in zone 14 in addition to the fact that they may have been lost during the screening process. Zone 13, although not in the cluster, has approximately the same essential gene density as those zones within the cluster. This is most likely due to the fact that the cluster boundary is gradual (Barnes et al. 1995). Interestingly, the essential gene density is greatest in zones 11 and 12, while immediately to the right, no essential genes were identified in zone 10, which is of a comparable size. Whether this is statistically significant has yet to be determined.

All developmental block stages are represented in our study but most of the essential gene mutations block as early larvae. This is in accordance with the essential gene mutation studies performed on chromosome V (Johnsen and Baillie 1990) and on chromosome IV (Clark and Baillie 1992). The low number of Mels is most likely a representation of our screen criteria. The screen was not designed to isolate Mels. The Mels that were isolated were probably picked up as sterile adults. They were not identified as Mels until further phenotypic characterization was performed, since they laid very few eggs and did not have a particularly gravid appearance.

Phenotypes of essential gene mutations may be more complex than simply the arrest stage of development. We report here only the block stages but for some we have observed more detailed phenomena, such as, defective cell divisions, improper gonadogenesis, clear, small, and very slowly developing worms. An example of a more specific phenotype is *let-733(s2621)* in which there is abnormal cytokinesis and cell division. Therefore, within this set of lethal mutations specific developmental defects may lead to the overall lethal phenotype.

Different alleles of a gene generally block at approximately the same developmental stage. We suspect that this reflects the zygotic null phenotype for these genes. EMS predominantly induces G/C → A/T transitions and therefore would generate numerous nonsense alleles, resulting in truncated protein products. The few genes with alleles that arrest at significantly different stages are interesting. The differing arrest stages suggest that the gene may have separate functions at different developmental times. It is also possible that we have generated mutations in the same gene with varying degrees of severity. The actual nature of these different arrest stages will be determined when these lethal alleles are further characterized.

In conclusion, we have divided the *dpy-17* region of the genome into 14 zones using chromosomal rearrangements – primarily deficiencies isolated in the UV and gamma-ray lethal screens reported in this paper. These zones have subsequently been anchored to the physical map by PCR-based deficiency endpoint mapping. Zones are delineated both genetically and physically. We mapped the 112 genes into these 14 physically defined zones and thus have narrowed the candidate sequences for each. Their mutations provide us with tools from which it will be possible to positionally clone these essential genes. Using systematic rescue experiments (Janke et al. 1997), we have already positioned some genes to cosmids and even to the correct coding element (Vatcher et al. 1998).

Our collection of mutations for 112 essential genes is a small part of the next step in the *C. elegans* genomic sequencing project. These genes constitute between 2.5 and 4.5% of all essential genes in *C. elegans*. As the *C. elegans* genome sequencing project nears the completion of the entire genomic sequence, researchers must focus on post-sequencing strategies to correlate the genomic sequence of predicted genes with biological function. Mutational analysis is an excellent method with which to achieve this goal.

**Acknowledgements** The authors thank Raja Rosenbluth for helpful discussion and comments on the manuscript. This work was supported by a grant from the Medical Research Council (MRC) of Canada to A. Rose and D. Baillie and a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to D. Baillie. The deficiencies were generated by The Genetic Toolkit project, which is funded by a grant from the NIH National Center for Research Resources (NCCR)(USA) to A. Rose, D. Baillie and D. Riddle. H.M.C. was a fellow of the Helen Hay Whitney Foundation.

## References

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

Barstead RJ, Kleiman L, Waterston RH (1991) Cloning, sequencing, and mapping of an alpha-actinin gene from the nematode *Caenorhabditis elegans*. Cell Motil Cytoskeleton 20:69–78

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

Clark DV (1990) The *unc-22(IV)* region of *Caenorhabditis elegans*: genetic analysis and molecular mapping. PhD Thesis, Simon Fraser University, Burnaby, BC, Canada

Clark DV, Baillie DL (1992) Genetic analysis and complementation by germ-line transformation of lethal mutations in the *unc-22 IV* region of *Caenorhabditis elegans*. Mol Gen Genet 232:97–105

Clark DV, Rogalski TM, Donati LM, Baillie DL (1988) The *unc-22(IV)* region of *Caenorhabditis elegans*: genetic analysis of lethal mutations. Genetics 119:345–353

Cox GN, Staprans S, Edgar RS (1981) The cuticle of C. elegans. II. Stage-specific changes in ultrastructure and protein composition during postembryonic development. Dev Biol 86:456–470

Ellis RE, Jacobson DM, Horvitz HR (1991) Genes required for the engulfment of cell corpses during programmed cell death in *Caenorhabditis elegans*. Genetics 129:79–94

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Gatewood BK, Bucher EA (1997) The mup-4 locus in *Caenorhabditis elegans* is essential for hypodermal integrity, organismal morphogenesis and embryonic body wall muscle position. Genetics 146:165–183

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:563–567

Hodgkin J, Durbin R, Martinellie S (1997) 1997 genetic map of *Caenorhabditis elegans*. Caenorhabditis Genetics Center, St. Paul, Minn

Horvitz HR, Brenner S, Hodgkin J, Herman RK (1979) A uniform genetic nomenclature for the nematode *C. elegans*. Mol Gen Genet 175:129–133

Howell AM, Rose AM (1990) Essential genes in the *hDf6* region of chromosome I in *Caenorhabditis elegans*. Genetics 126:583–592

Janke DL, Schein JE, Ha T, Franz NW, O'Neil NJ, Vatcher GP, Stewart HI, Kuervers LK, Baillie DL, Rose AM (1997) Interpreting a sequenced genome: toward a cosmid transgenic library of *Caenorhabditis elegans*. Genome Res 7:974–985

Jansen G, Hazendonk E, Thijssen KL, Plasterk RHA (1997) Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. Nat Genet 17:119–121

Johnsen RC, Baillie DL (1991) Genetic analysis of a major segment [LGV(left)] of the genome of *Caenorhabditis elegans*. Genetics 129:735–752

Johnsen RC, Baillie DL (1997) Mutation. In: Riddle D, Blumenthal T, Meyer B, Priess J (eds) *C. elegans II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 79–95

Kimble J, Hirsh D (1979) The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. Dev Biol 70:396–417

McKim KS, Starr T, Rose AM (1992) Genetic and molecular analysis of the *dpy-14* region in *Caenorhabditis elegans*. Mol Gen Genet 233:241–251

Rogalski TM, Baillie DL (1985) Genetic organization of the *unc-22 IV* gene and the adjacent region in *C. elegans*. Mol Gen Genet 201:409–414

Rogalski TM, Moerman DG, Baillie DL (1982) Essential genes and deficiencies in the *unc-22 IV* region of *C. elegans*. Genetics 102:725–736

Rosenbluth RE, Cuddeford C, Baillie DL (1985) Mutagenesis in *Caenorhabditis elegans*. II. A spectrum of mutational events induced with 1500 R of $\gamma$-radiation. Genetics 109:493–511

Rosenbluth RE, Rogalski TM, Johnsen RC, Addison LM, Baillie DL (1988) Genomic organization in *Caenorhabditis elegans*: deficiency mapping on linkage group V(left). Genet Res 52:105–118

Sigurdson DC, Spanier GJ, Herman RK (1984) *Caenorhabditis elegans* deficiency mapping. Genetics 108:331–345

Stewart HI, Rosenbluth RE, Baillie DL (1991) Most ultraviolet irradiation induced mutations in the nematode *Caenorhabditis elegans* are chromosomal rearrangements. Mutat Res 249:37–54

Sulston J, Horvitz HR (1977) Postembryonic lineages of the nematode *Caenorhabditis elegans*. Dev Biol 56:110–156

Vatcher GP, Thacker CM, Kaletta T, Schnabel H, Schnabel R, Baillie DL (1998) Serine hydroxymethyltransferase is maternally essential in *Caenorhabditis elegans*. J Biol Chem 273:6066–6073

Waterston RH, Sulston JE (1995) The genome of *Caenorhabditis elegans*. Proc Natl Acad Sci USA 92:10836–10840

Waterston RH, Sulston JE, Coulson AR (1997) The genome. In: Riddle D, Blumenthal T, Meyer B, Priess J (eds) *C. elegans II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 23–47

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

Wilson R, Ainscough R, Anderson K, Baynes C, Berks M, Burton J, Connell M, Bonfield J, Copsey T, Cooper J (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature 368:32–38