GENETIC MAPPING OF *CAENORHABDITIS ELEGANS* COLLAGEN GENES USING DNA POLYMORPHISMS AS PHENOTYPIC MARKERS

GEORGE N. COX,[1] STEPHEN CARR, JAMES M. KRAMER[2] AND DAVID HIRSH

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309

Manuscript received May 7, 1984
Revised copy accepted November 8, 1984

ABSTRACT

In *Caenorhabditis elegans* collagens comprise a dispersed family of 40–150 genes, the majority of which probably code for collagen proteins found in the animal's cuticle. The conserved $(Gly\text{-}X\text{-}Y)_n$ triple helix coding sequence of collagen genes has facilitated the isolation of a large number of *C. elegans* collagen genes by recombinant DNA methods. We have begun a study of the chromosomal organization of these genes by screening laboratory strains of *C. elegans* for DNA polymorphisms in the regions surrounding collagen genes. Polymorphisms near seven genes have been identified and have been used as phenotypic markers in genetic crosses to assign the genes to linkage groups II, III, IV, and X. Four genes are shown by multifactor crosses to map to a 2–3 map unit interval between *unc-24* and *unc-22* on chromosome *IV*.

CUTICLE formation in the nematode *Caenorhabditis elegans* is an attractive system for studying gene regulation and morphogenesis (EDGAR *et al.* 1982). The cuticle is a multilayered extracellular structure that is composed primarily of protein and that is sloughed off and reformed *de novo* at each of four postembryonic molts. Although the gross morphological and biochemical features of each molt appear superficially similar (SINGH and SULSTON 1978; COX *et al.* 1981), comparisons of cuticles formed at the different molts indicate that they differ considerably in ultrastructure and protein composition (COX, STAPRANS and EDGAR 1981).

As is the case for other nematodes, the major structural proteins of the *C. elegans* cuticle are collagens (COX, STAPRANS and EDGAR 1981; BIRD 1971). Recently, we described the isolation of a number of *C. elegans* collagen genes by recombinant DNA methods (KRAMER, COX and HIRSH 1982; COX, KRAMER and HIRSH 1984). Southern blot hybridization experiments and recombinant phage library screenings indicate that collagens are a large gene family in *C. elegans*, comprising at least 40 and as many as 150 members. In contrast to vertebrate collagen genes which encode large 150,000-dalton proteins (BORNSTEIN and

[1] Present address: Synergen, Inc., 1885 33rd Street, Boulder, Colorado 80301.
[2] Present address: Abbott Laboratories, Molecular Biology Laboratory, North Chicago, Illinois 60064.

SAGE 1980), most *C. elegans* collagen genes code for 30,000- to 40,000-dalton proteins, which are probably cuticle collagens (KRAMER, COX and HIRSH 1982; POLITZ and EDGAR 1984). The chromosomal organization of these genes is presently unknown.

As part of our study of cuticle formation in *C. elegans*, we would like to identify mutations in cuticle collagen genes. Genetic analysis of collagen genes should provide insights into their physiological roles and regulation in the nematode. Since collagens appear to be ubiquitous components of metazoan organisms, the information gained from such analyses should be applicable to collagen gene systems in other organisms. To date, mutations in more than 35 genes that affect cuticle morphology have been described for *C. elegans* (HIGGINS and HIRSH 1977; COX *et al.* 1980; EDGAR *et al.* 1982). The genes are located on all six linkage groups and do not exhibit extensive clustering. Whether these genes code for structural components of the cuticle or for ancillary enzymatic functions required for proper cuticle assembly is not known. Biochemical analysis of the mutants is complicated because of the large number of posttranslational modifications, such as propeptide processing, glycosylation and hydroxylation of specific amino acid residues and interchain covalent cross-linking, that occur during collagen biosynthesis (FESSLER and FESSLER 1978).

If the chromosomal locations of collagen genes were known, then efforts could be directed at mutations in these regions to screen for defects in collagen genes. Toward this end we initiated a study to position cloned collagen genes on the *C. elegans* genetic map. The two common laboratory strains of *C. elegans*, Bristol and Bergerac, exhibit a considerable degree of restriction enzyme site polymorphism (EMMONS, KLASS and HIRSH 1979). DNA polymorphisms provide phenotypic markers that can be analyzed in standard genetic crosses to localize genes on the genetic map, in much the same way that protein electrophoretic differences have been used to map genetic loci in other systems (GOLDSMITH and BASEHOAR 1978; PETERSON, MOLLER and MITCHELL 1979; ROSE *et al.* 1982; FILES, CARR and HIRSH 1983). We have screened the Bristol and Bergerac strains of *C. elegans* for DNA polymorphisms in the regions surrounding collagen genes. Polymorphisms near seven genes have been identified and used in genetic crosses to assign the genes to four different linkage groups. Multifactor cross data indicate that four genes are located in a 2- to 3-map unit interval on chromosome *IV*, raising the possibility that this region may represent a collagen gene cluster.

MATERIALS AND METHODS

Nematode strains and culture: *Caenorhabditis elegans* var. Bristol (N2) and *C. elegans* var. Bergerac (LY) were from the Boulder stock collection and are descended from nematodes originally obtained from S. BRENNER (in 1972) and from J. BRUN (in 1977), respectively. Mutant genes and alleles of the Bristol strain used in genetic crosses were *dpy-5(e61)I, dpy-10(e128)II, dpy-1(e1)III, dpy-18(e364)III, dpy-13(e184)IV, unc-24(e138)IV, daf-14(m77)IV, daf-15(m81)IV, dpy-20(e1282)IV, unc-22(ct13)IV, dpy-4(e1166)IV, dpy-11(e224)V, lon-2(e678)X and unc-3(e151)X* (BRENNER 1974; HERMAN, HORVITZ and RIDDLE 1980). Nematodes were grown at 16° or 20° on NGM plates (BRENNER 1974) seeded with *E. coli* strain OP50. Crosses involving temperature-sensitive mutants were analyzed at 25°.

Genetic crosses: The procedures used for genetic crosses were those described by BRENNER (1974) as adapted for interstrain crosses and DNA polymorphism mapping by FILES, CARR and HIRSH (1983). Heterozygous Bristol males were used to introduce genetic markers into the F_1 hybrids

because Bergerac LY males mate poorly. Cross progeny hermaphrodites were identified by individually plating immature fourth larval stage (L4) hermaphrodite progeny from plates containing many male progeny (an indication that mating had occurred), and those F_2 animals segregating the Bristol markers were retained for further analysis. For multifactor crosses, the two F_2 recombinant classes, *e.g.*, Dpy non-Unc and Unc non-Dpy, were chosen from different plates to ensure independence of crossover events. Typically, 12–15 progeny of each recombinant were individually plated, and those animals segregating only the Bristol marker were assumed to be homozygous for the recombinant chromosome; only one such hermaphrodite was retained for each recombination event. When plates containing these animals reached saturation, the population was transferred to several fresh plates and grown until saturated again. Eggs were then obtained from gravid adults by treatment with hypochlorite (EMMONS, KLASS and HIRSH 1979) and hatched in sterile M9 buffer (BRENNER 1974). First larval stage (L1) juveniles were collected the next day by low speed centrifugation.

DNA preparation and hybridization conditions: Methods for the preparation of L1 juvenile DNA, restriction enzyme digests and Southern blot hybridizations were essentially as described by EMMONS, KLASS and HIRSH (1979). In cases in which intact recombinant phage or plasmids containing subcloned collagen genes were used as hybridization probes, the posthybridization washes were performed at 50°–55° to eliminate cross-hybridization between collagen genes (COX, KRAMER and HIRSH 1984). Restriction enzymes used to search for DNA polymorphisms were *Bam*HI, *Bgl*II *Eco*RI, *Hin*dIII, *Pvu*II, *Sal*I, *Xba*I and *Xho*I. Bristol and Bergerac DNAs were digested with each of these enzymes, size-fractionated on agarose gels, Southern blotted onto nitrocellulose filters (SOUTHERN 1975) and hybridized to ^{32}P-labeled DNA probes. Only *C. elegans* Bristol DNA fragments were used as hybridization probes. Specific restriction fragments were eluted from agarose gels using strips of DEAE cellulose membrane (NA-45, Schleicher and Schuell; DRETZEN *et al.* 1981) and either labeled directly or first subcloned into pBR322 or pBR325 by conventional procedures. Plasmid DNAs were prepared by the method of CLEWELL (1972), and DNA probes were labeled with ^{32}P by nick-translation as described by RIGBY *et al.* (1977).

Positioning morphological markers on linkage group IV: Two-factor crosses (BRENNER 1974) were used to position *unc-24* relative to *dpy-13* and *unc-22*. Segregation of Dpy and Unc recombinants from *dpy-13 unc-24*/++ hermaphrodites yielded a map distance (±95% confidence limits) of 6.7 ± 1.3 map units (55 *dpy-13* and 43 *unc-24* recombinants of 1505 total progeny). Segregation of *unc-24* and *unc-22* recombinants from *unc-24 unc-22*/++ hermaphrodites yielded a map distance of 2.7 ± 0.8 map units (27 *unc-24* and 18 *unc-22* recombinants of 1666 total progeny). Crosses were performed at 20°, and all progeny were scored.

Three-factor crosses (BRENNER 1974) were used to position *dpy-20*, *daf-14* and *daf-15* between *unc-24* and *unc-22*. *dyp-20:* 42 of 46 *unc-24* recombinants from *unc-24 unc-22/dpy-20* hermaphrodites segregated *dpy-20*. *daf-14:* seven of nine *unc-24* recombinants and six of 12 *dpy-20* recombinants from *unc-24 dpy-20/daf-14* hermaphrodites segregated *daf-14*. *daf-15:* seven of 13 *unc-24* recombinants and five of nine *dpy-20* recombinants from *unc-24 dpy-20/daf-15* hermaphrodites segregated *daf-15*.

RESULTS

Identification of collagen gene DNA polymorphisms: We have described previously the isolation of approximately 25 collagen genes from *C. elegans* Bristol DNA libraries cloned in λ Charon 10 or λ1059 (KRAMER, COX and HIRSH 1982; COX, KRAMER and HIRSH 1984). The average size of the *C. elegans* DNA inserts in these phages is about 15 kb. The genomic organization of these collagen genes and their surrounding regions were compared in Bristol N2 and Bergerac LY worms by Southern blot hybridization experiments using as hybridization probes either intact recombinant phage DNA or specific restriction fragments of these phage (see MATERIALS AND METHODS). Eight different restriction endonucleases were routinely used in these experiments. Seven phage, which defined five collagen genes, *col-2*, *col-3*, *col-6*, *col-8* and *col-9*, hybridized to polymorphic

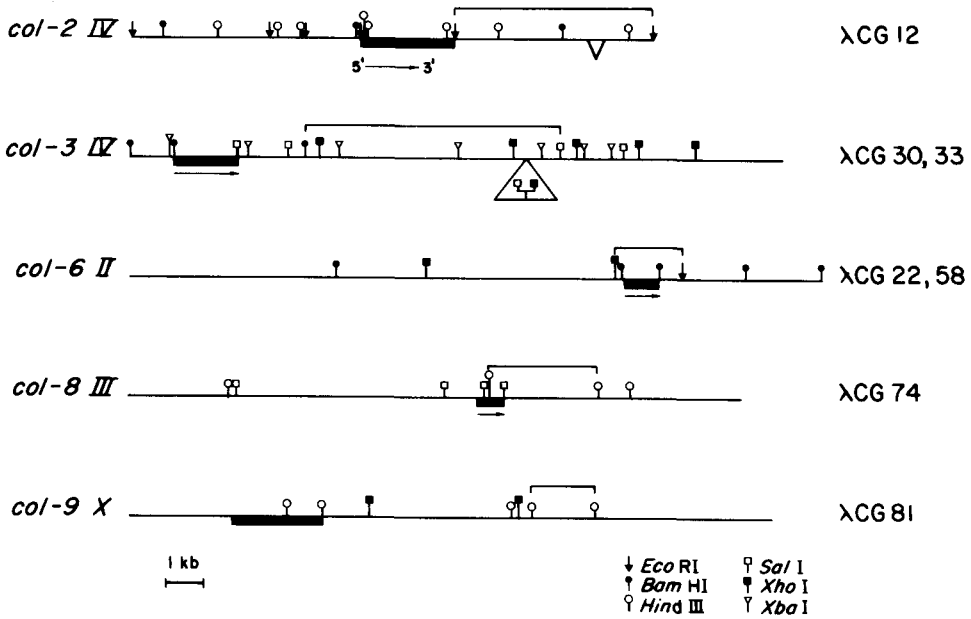

FIGURE 1.—Restriction maps of *C. elegans* collagen genes and flanking regions. *C. elegans* Bristol DNA inserts in recombinant phage are shown. Black boxes indicate positions of collagen genes. The limits of the boxes indicate the positions of the closest restriction enzyme sites used in the mapping experiments and, therefore, do not necessarily represent the true limits of the genes. The direction of transcription for each gene (where known) is indicated by a horizontal arrow. Triangles represent DNA polymorphisms. An upward pointing triangle indicates a DNA insertion present in the Bergerac strain (relative to the Bristol strain), whereas a downward pointing triangle indicates a DNA deletion in the Bergerac strain. Restriction fragments subcloned into plasmids and later used as hybridization probes to reveal polymorphisms are indicated with brackets. Not all restriction enzyme sites used in the mapping experiments are shown for each gene. The cloned region surrounding *col-6* contains additional *Eco*RI cleavage sites which are not shown.

restriction fragments between the two strains. We have verified that the polymorphic DNA regions lie within or adjacent to the DNA regions cloned in the phages by using single-copy DNA fragments or collagen gene DNA fragments that hybridize to single restriction fragments under stringent hybridization conditions (see MATERIALS AND METHODS) as diagnostic probes for the polymorphisms. Restriction maps and polymorphic regions (where known) of these phages are shown in Figure 1.

A brief description of each polymorphism will be given and is summarized in Table 1.

col-2: The *Eco RI* fragment adjacent to the 3′ end of *col-2* is approximately 200 base pairs (bp) larger in Bristol worms than in Bergerac worms (4.9 *vs.* 4.7 kb, Figure 2). When this *Eco*RI fragment is hybridized to whole genome Southern blots of Bristol and Bergerac DNAs digested with *Bam*HI or *Hin*dIII, the hybridizing bands also show a size difference of approximately 200 bp (Table 1, Figure 2). This result indicates that the observed polymorphism is due to the presence of an additional 200 bp of DNA in the Bristol strain that is absent from this region in the Bergerac strain rather than due to a single base pair change.

TABLE 1

Collagen gene DNA polymorphisms

Gene	Chromosome	Restriction fragment length differences[a]	
		N2 > LY	N2 < LY
col-2	*IV*	4,950 *vs.* 4,700 (*Eco*RI)	
		4,370 *vs.* 4,170 (HindIII)[b]	
		3,370 *vs.* 3,200 (*Bam*HI)[c]	
col-3	*IV*	7,500 *vs.* 7,500 + 1,480 (*Sal*I)	14,300 *vs.* 16,600 (*Bam*HI)[b]
		1,880 *vs.* 1,970 + 1,460 (XhoI)	1,000 *vs.* 2,650 (*Hin*dIII)
			2,290 *vs.* 3,870 (*Xba*I)
col-4	*IV*		4,700 *vs.* 5,300 (*Msp*I)[b]
col-5	*IV*		3,300 *vs.* 3,600 (*Msp*I)[b]
col-6	*II*	23,100 *vs.* 16,00 (*Xho*I)[b]	
col-8	*III*	20,400 *vs.* 9,100 (*Sal*I)[b]	13,000 *vs.* 14,600 (*Pvu*II)
			5,100 *vs.* 7,200 (*Xba*I)
col-9	*X*	18,200 *vs.* 15,000 (*Xho*I)[b]	

[a] Lengths in base pairs of restriction fragments exhibiting size differences between N2 (Bristol) and LY (Bergerac) strains are given. Polymorphisms in which the N2 band is larger than the LY band are listed in the left-hand column, whereas polymorphisms in which the N2 band is smaller than the LY band are listed in the right-hand column. Restriction enzymes used to reveal polymorphisms are shown in parentheses.

[b] These restriction enzymes and polymorphisms were routinely used for mapping experiments.

[c] *Bam*HI was used to reveal the *col-2* polymorphism for the *unc-24 unc-22* multifactor cross.

The Bristol 4.9-kb *Eco*RI fragment contains a repeated sequence that hybridizes weakly to about 15 other bands on whole genome Southern blots. We do not know whether the repeated sequence in this fragment is the additional 200-bp insert.

col-3: The *col-3* polymorphism resides in the 3′ flanking region of the gene and also is due to an insertion/deletion of a DNA segment rather than to a single base pair change. In this case, however, the extra DNA is present in the Bergerac strain, as evidenced by the larger size of the Bergerac *Bam*HI, *Xba*I and *Hin*dIII bands relative to their Bristol counterparts (Table 1, Figure 2). With each of these enzymes the polymorphic Bergerac band is approximately 1.6 kb larger than its Bristol counterpart. The locations of the *Xba*I and *Hin*dIII restriction fragments showing Bristol/Bergerac size differences position the 1.6-kb Bergerac insert within the region shown in Figure 1. The 1.6-kb Bergerac insert must contain restriction sites for *Sal*I and *Xho*I because two, rather than one, new bands are seen after digestion of Bergerac DNA with these enzymes (Figure 2). The sum of the lengths of the two new Bergerac bands equals the sum of the sole Bristol band plus 1.6 kb.

col-6: Two phage, λCG22 and λCG58, which contain inserts that overlap *col-6*, hybridize to an *Xho*I restriction fragment that is approximately 7 kb larger in the Bristol strain than in the Bergerac strain (Table 1, Figure 3). None of the other enzymes tested exhibited polymorphic hybridizing bands between the two strains when probed with these phage, indicating that the DNA region responsible for the *Xho*I polymorphism lies outside of the region cloned on the phages.

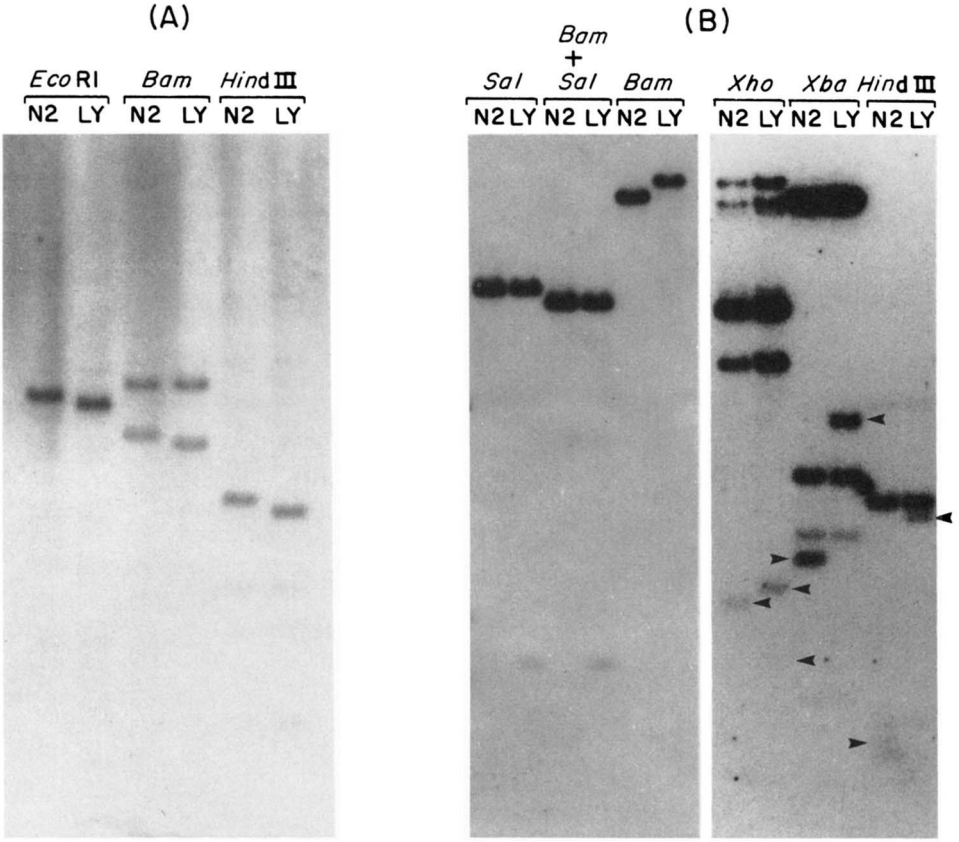

FIGURE 2.—Localization of DNA polymorphisms surrounding *col-2* (A) and *col-3* (B). Bristol (N2) and Bergerac (LY) DNAs were digested with the indicated restriction enzymes, size-fractionated on 0.7% agarose gels, blotted onto nitrocellulose filters and hybridized to appropriate plasmids (see Figure 1 and Table 1) to reveal polymorphisms. Sizes of polymorphic hybridization bands are given in Table 1. Faint polymorphic bands are indicated with arrowheads.

The positions of *Xho*I restriction sites on the two phages indicate that the polymorphic region must lie off the 3′ end of the *col-6* gene.

col-8: λCG74, which contains *col-8*, hybridizes to polymorphic DNA bands when used as the hybridization probe to Southern blots of Bristol and Bergerac DNAs digested with *Xba*I, *Pvu*II and *Sal*I (*Xho*I was not tested). The polymorphic Bergerac band is 1.6–2.2 kb larger than the corresponding Bristol band when *Xba*I or *Pvu*II is used as the restriction enzyme but approximately 11 kb smaller when *Sal*I is used as the restriction enzyme (Table 1). The 2.3-kb *Hin*dIII fragment containing *col-8* hybridizes only to this polymorphic *Sal*I restriction fragment under stringent hybridization conditions, indicating that the polymorphic region lies to the 3′ side of the gene, outside of the region cloned on λCG74.

col-9: The single-copy 1.8-kb *Hin*dIII fragment adjacent to *col-9* in λCG81 hybridizes to an *Xho*I fragment that is 3.2 kb larger in Bristol worms than in

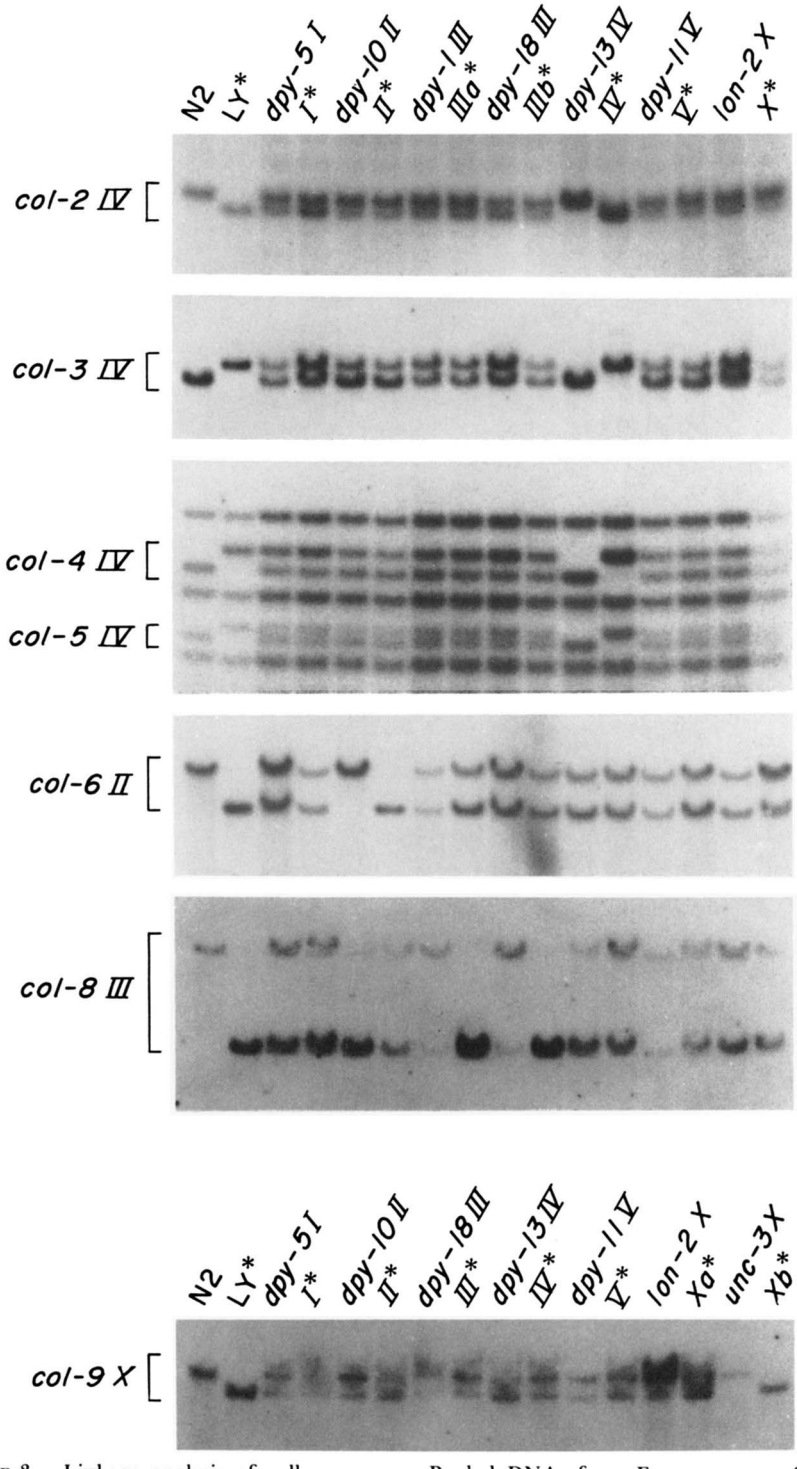

FIGURE 3.—Linkage analysis of collagen genes. Pooled DNAs from F_2 segregants of crosses between Bristol (N2) and Bergerac (LY) worms, utilizing the Bristol genetic markers indicated, were digested with restriction enzymes (see Table 1), size-fractionated on 0.7% agarose gels, blotted onto nitrocellulose filters and hybridized to appropriate plasmids (see Figure 1) to reveal polymorphisms. F_2 Bristol segregants are indicated by the genetic marker used in the cross, whereas F_2 Bergerac segregants are indicated by a chromosome number marked with an asterisk.

Bergerac worms (Table I, Figure 3). None of the other restriction enzymes tested yielded polymorphic hybridization bands, indicating that the DNA polymorphism lies outside of the region cloned in λCG81. The direction of transcription for *col-9* has not been determined.

col-4 and col-5: In addition to searching for DNA polymorphisms among cloned collagen genes, we also searched for other collagen gene DNA polymorphisms by comparing Southern blot hybridization patterns of restriction endonuclease-digested Bristol and Bergerac DNAs hybridized to collagen gene probes under conditions in which all collagen genes cross-hybridize. No obvious differences were observed for most restriction enzymes tested, although the complexity of the hybridization pattern (greater than 30 bands) may have obscured any such differences. However, two clear polymorphisms were observed when *Msp*I (or *Hpa*II) was used as the restriction enzyme (Figure 3). The collagen genes defined by these polymorphisms have been named *col-4* and *col-5*. For both genes, the hybridizing band in Bergerac worms is slightly larger (300–600 bp) than in Bristol worms (Table 1). Digestion of all of the collagen gene-containing phage in our collection with *Msp*I suggests that *col-4* and *col-5* are distinct from any of the collagen genes cloned in these phage, although we cannot be certain in cases in which collagen genes are located near the phage break points.

Linkage analysis of collagen genes: The DNA polymorphisms described for *col-2* through *col-9* were used as phenotypic markers to assign the genes to linkage groups. Linkage analyses were performed using DNAs derived from F_2 segregants of crosses between Bristol and Bergerac worms. The origin of these DNAs and detailed methods for their preparation are described by FILES, CARR and HIRSH (1983). Briefly, matings were performed between Bristol and Bergerac worms, each cross utilizing a different Bristol genetic marker [either Dpy (dumpy), Lon (long) or Unc (uncoordinated)] specific for one of the six *C. elegans* linkage groups. Approximately 50 homozygous Dpy (Bristol) or wild-type (Bergerac) F_2 segregants from each cross were individually plated, grown into large populations and their DNAs isolated and pooled.

The hybridization patterns of the pooled F_2 DNAs were used to assign linkage groups to the DNA polymorphisms in question. If the polymorphism is closely linked to one of the genetic markers used in the crosses, then the pooled F_2 Dpy (Lon or Unc) DNA for that chromosome will show *only* the Bristol polymorphic band, whereas F_2 wild-type DNA for that chromosome will show *only* the Bergerac polymorphic band (assuming no recombination between the marker gene and the polymorphism). Dpy and wild-type DNAs for all other linkage groups will show *both* the Bristol and Bergerac polymorphic bands in about equal intensities due to random assortment of chromosomes during meiosis. The farther away a polymorphism lies from the linked genetic marker, the greater will be the recombination frequency between the polymorphism and the marker gene and less clear-cut will be the ratios (2:0 *vs.* 1:1) of the hybridizing polymorphic bands in the pooled F_2 DNAs.

The results of the linkage experiments are shown in Figure 3 and summarized in Table 1. Four genes, *col-2*, *col-3*, *col-4* and *col-5*, showed tight linkage to *dyp-13 IV*. The remaining genes showed linkage to different chromosomes: *col-6*

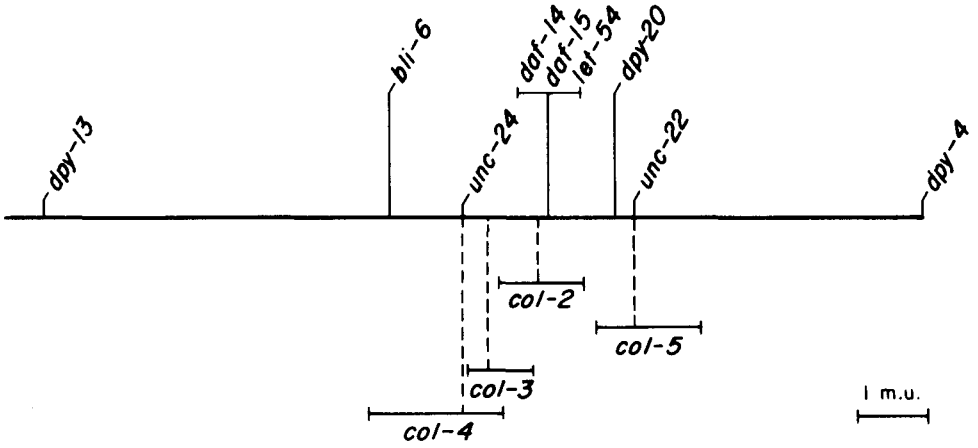

FIGURE 4.—Summary of map locations of chromosome *IV* collagen genes. Positions of morphological markers used in mapping experiments or discussed in text are indicated above the line, whereas positions of collagen genes are indicated below the line. Dashed lines indicate map positions of collagen genes based on recombinant ratios, and parentheses indicate extents of 95% confidence intervals. Positioning of several morphological markers is described in MATERIALS AND METHODS. *daf-14*, *daf-15* and *let-54* have not been ordered with respect to each other.

showed linkage to *dpy-10 II*, and *col-8* showed linkage to the chromosome *III* markers *dpy-1* and *dpy-18*. *col-9* exhibited weak linkage to *lon-2 X*; therefore, a second *X*-linked marker, *unc-3*, which is located about 25 maps units to the right of *lon-2*, was used subsequently for linkage analysis. *col-9* showed tight linkage to *unc-3*, indicating that *col-9* is *X* linked and probably closer to *unc-3* than to *lon-2*.

Multifactor cross mapping of chromosome IV *collagen genes*: Multifactor crosses were used to position *col-2* through *col-5* relative to morphological markers on chromosome *IV*, a portion of which is shown in Figure 4. These crosses also allowed *col-2* through *col-5* to be positioned relative to each other. In the first multifactor cross, Bristol males of genotype *dpy-13 unc-24*/++ were mated to Bergerac hermaphrodites and five *dpy-13* and seven *unc-24* recombinants were picked from among the F_2 segregants. The recombinant chromosome *IV* in these animals was made homozygous by selfing, and each worm grown into a large population. The DNA from each recombinant population was then isolated and analyzed for all four collagen gene DNA polymorphism patterns. The results of these experiments (Figure 5) were that all seven *unc-24* recombinants retained the Bristol polymorphic band for all four collagen genes, whereas all five *dpy-13* recombinants gained the Bergerac band and lost the Bristol band for *col-3* and *col-4*. Two *dpy-13* recombinants, no. 4 and no. 5 in Figure 5, exhibited both the Bristol and Bergerac polymorphic bands in about equal intensities for *col-2* and *col-5*. The simplest explanation for this result is that these recombinants gained the Bergerac band during the initial recombination event, but that a second crossover event occurred between *unc-24* and these genes while the recombinant chromosome was being made homozygous. This interpretation is consistent with the map order of the genes as determined below. The results of this cross, then,

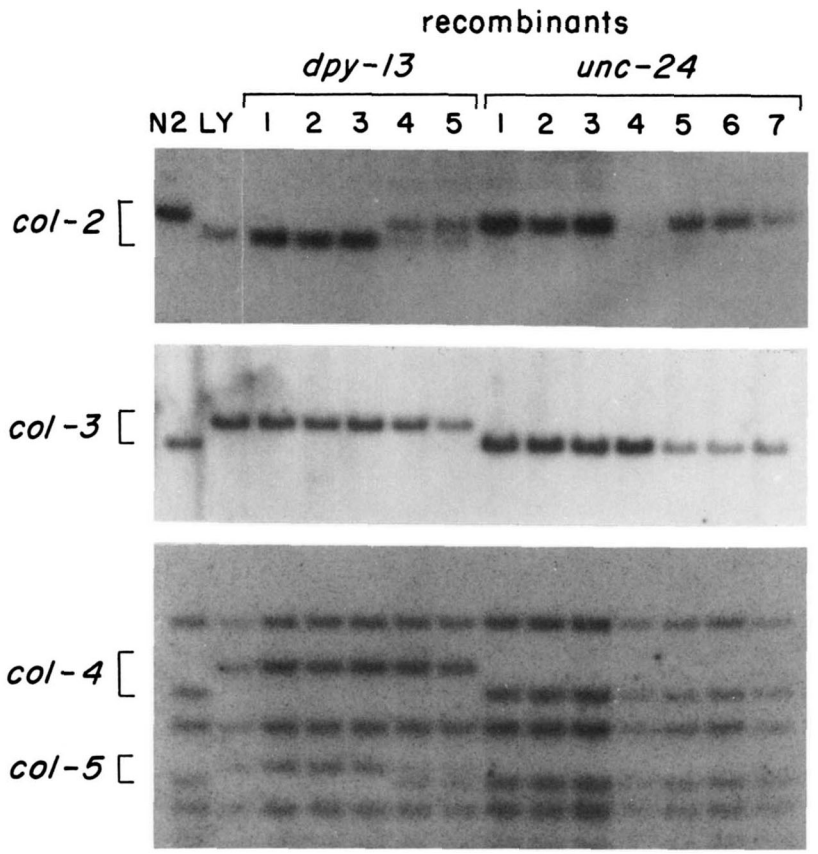

FIGURE 5.—Analysis of chromosome *IV* collagen gene polymorphism patterns of individual recombinants from the *dpy-13 unc-24* multifactor cross. Individual *dpy-13* and *unc-24* recombinants are numbered and were analyzed in the same order for all three hybridization experiments. Individual recombinants were grown into large populations and their DNAs isolated, digested with restriction enzymes (see Table 1), size-fractionated on 0.7% agarose gels, blotted onto nitrocellulose filters and hybridized with appropriate plasmids (see Figure 1) to reveal polymorphisms. *unc-24* recombinant no. 4 exhibits a faint Bristol (N2) band for the *col-2* polymorphisms. Note that *dpy-13* recombinants no. 4 and no. 5 exhibit both the Bristol and Bergerac (LY) hybridizing bands for the *col-2* and *col-5* polymorphisms (see text for explanation).

indicated that all four collagen genes were located to the right of *unc-24* or very close to it on the left.

The second multifactor cross was performed in the same way but utilized the linked Bristol markers *unc-22 dpy-4*. As shown in Figure 6, all six *unc-22* recombinants retained the Bristol polymorphic band for all four collagen genes, whereas all seven *dpy-4* recombinants gained the Bergerac band and lost the Bristol band for all four genes. This result indicated that all four collagen genes lie to the left of *unc-22* or very close to the right of it.

The final multifactor cross performed utilized the linked Bristol markers *unc-24 unc-22*. The phenotypes of both recombinant classes are distinct from the double mutant and from each other. The results of this cross (Figure 7) indicated

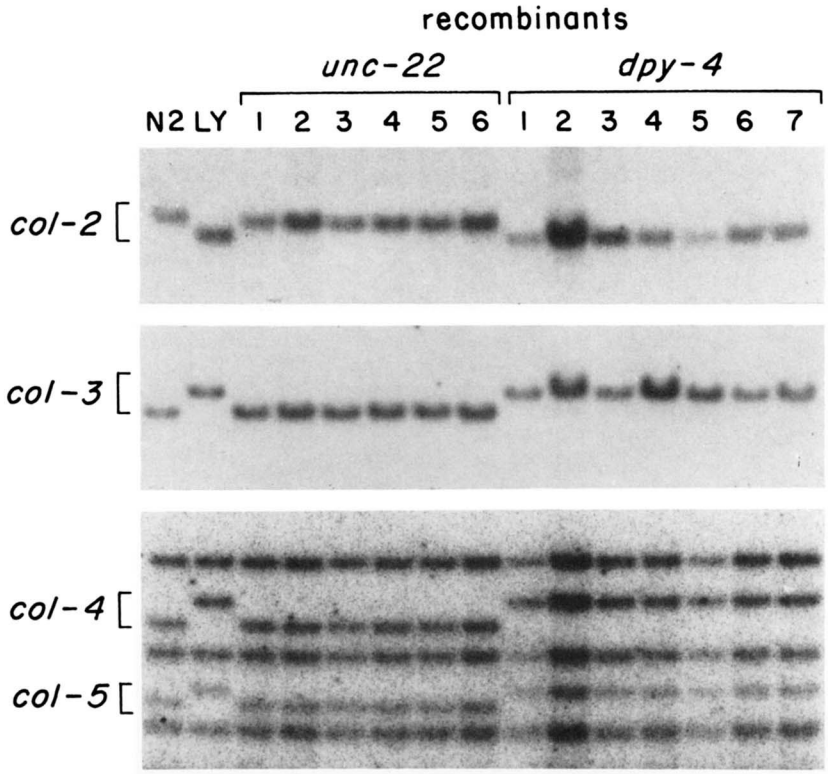

FIGURE 6.—Analysis of chromosome *IV* collagen gene polymorphism patterns of individual recombinants from the *unc-22 dpy-4* multifactor cross. Experimental procedures and figure legend are as for Figure 5.

that all four collagen genes map to different positions on chromosome *IV*. *col-4* must be very close to *unc-24* because all seven *unc-24* recombinants retained the *col-4* Bristol band, whereas all six *unc-22* recombinants lost the Bristol band and gained the Bergerac band. Just the opposite result was obtained for *col-5*, indicating that this gene must be located very close to *unc-22*. *col-2* and *col-3* must be located between *unc-24* and *unc-22* because certain recombinants of both classes retained the Bristol hybridizing band, whereas other recombinants showed only the Bergerac hybridizing band. The numbers observed were that three of seven *unc-24* recombinants and two of six *unc-22* recombinants retained the *col-2* Bristol band, whereas six of seven *unc-24* recombinants and one of six *unc-22* recombinants retained the *col-3* Bristol band. From these data we conclude that two of the 13 recombination events occurred between *col-4* and *col-3* (*unc-24* recombinant no. 7 and *unc-22* recombinant no. 6), four recombination events occurred between *col-3* and *col-2* (*unc-24* recombinants nos. 1, 2, 3 and *unc-22* recombinant no. 3) and seven recombination events occurred between *col-2* and *col-5* (*unc-24* recombinants nos. 4, 5, 6 and *unc-22* recombinants 1, 2, 4 and 5). These results position *col-2* near the middle of the *unc-24 unc-22* interval and *col-3* between *unc-24* and *col-2*. Since the same recombinant DNAs were used to

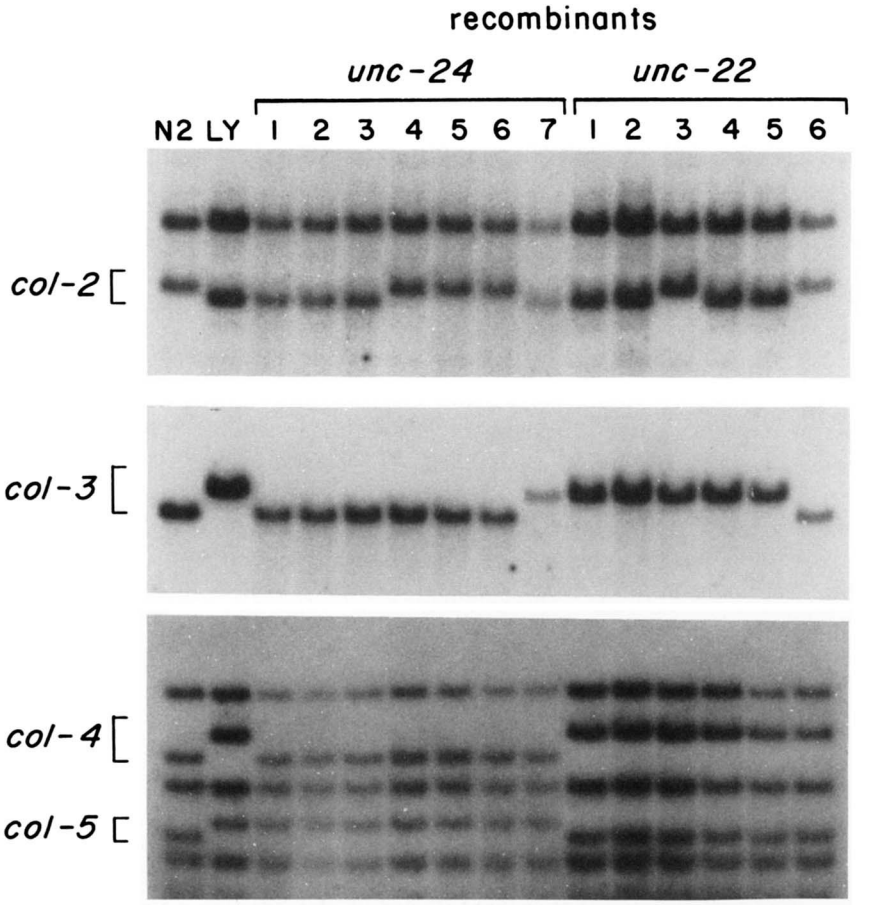

FIGURE 7.—Analysis of chromosome *IV* collagen gene polymorphism patterns of individual recombinants from the *unc-24 unc-22* multifactor cross. Experimental procedures and figure legend are as for Figure 5.

map all four genes, recombinant ratios and individual recombinants can be compared and give the gene order in this region as (*unc-24 col-4*) *col-3 col-2* (*col-5 unc-22*). Genes listed within parentheses cannot be ordered with respect to each other based on the multifactor mapping data presented here.

A summary of our findings is shown in Figure 4.

DISCUSSION

Previous studies on the physical mapping of recombinant phage containing *C. elegans* collagen genes left open the question of the chromosomal organization of these genes (KRAMER, COX and HIRSH 1982; COX, KRAMER and HIRSH 1984). Most phage analyzed in these studies contained only a single collagen gene within the average 15 kb of *C. elegans* DNA cloned per phage and did not overlap end fragments. Thus, although these data implied that *C. elegans* collagen genes were not organized into tandem arrays as is typical of histone and ribosomal genes

(KEDES 1979; LONG and DAWID 1980), it remained possible that these genes were all part of a single, large cluster, but separated from one another by blocks of DNA larger than 5–10 kb. Such a genomic organization has been determined for several other multigene families such as those encoding silkmoth chorion genes (EICKBUSH and KAFATOS 1982) and mouse transplantation antigen and immunoglobulin genes (STEINMETZ *et al.* 1982; SHIMIZU *et al.* 1982). The existence of DNA polymorphisms in the regions surrounding collagen genes in different strains has allowed us to show here by genetic analysis that *C. elegans* collagen genes are located on at least four of the six linkage groups and are thus a chromosomally dispersed multigene family.

Multifactor cross data indicate that the four collagen genes that map to chromosome *IV* are located within a 2- to 3-map unit interval. The distribution of collagen genes on this chromosome (50 total map units; HERMAN, HORVITZ and RIDDLE 1980) is not random and raises the possibility that this region may represent a collagen gene cluster. The size of such a cluster would be on the order of several hundred kilobases, assuming an equal recombination frequency throughout the genome (8.7×10^7 bp/haploid genome and 300 total map units). Although quite large, clusters of this size have been described for several eukaryotic gene families (EICKBUSH and KAFATOS 1982; STEINMETZ *et al.* 1982). Alternatively, the close proximity of collagen genes on chromosome *IV* may simply reflect a nonrandom distribution of all genes and/or DNA polymorphisms on the genetic map of this chromosome (possibly due to regional differences in recombination frequencies). We are presently investigating the organization of collagen genes in this region in detail using chromosome "walking" techniques. Studies so far have revealed the existence of a fifth collagen gene in this region, approximately 20 kb away from *col-3* (J. KRAMER, unpublished results).

The physical distances between collagen genes in the *unc-24 unc-22* region of chromosome *IV* may actually be smaller than we anticipate based on the observed recombination frequencies. In the *dpy-13 unc-24* multifactor cross, two of five *dpy-13* recombinants appear to have undergone a second recombination event in the interval between (*unc-24*) *col-3* and *col-2* during the selfing process involved in making the recombinant chromosome homozygous. Second recombination events such as these would not have been detected among the *unc-24* recombinants because of the map order of the genes. The recombination frequency between *col-3* and *col-2* in these crosses is much higher than would be expected for two genes that are less than 2 map units apart and may indicate a recombination "hot-spot" in the *col-3* to *col-2* interval, at least in these interstrain crosses. Such a hot spot would give an overestimate of the true physical distances between *col-3* and *col-2* and other genes in this region.

The localization of several collagen genes to the *unc-24 unc-22* region of chromosome *IV* permits us to begin detailed molecular and genetic analyses of mutations in this region to screen for defects in collagen genes. When the collagen genes unlinked to chromosome *IV* are mapped more precisely on their respective linkage groups it will be possible to extend these studies to other chromosomal regions as well. Several previously identified genes in the *unc-24 unc-22* region of chromosome *IV* have phenotypes that suggest these genes may

code for cuticle collagen genes or for other genes involved in cuticle formation. Mutations in *bli-6* and *dpy-20* (see Figure 4) cause animals to manifest cuticular abnormalities (M. KUSCH, personal communication, unpublished results). *bli-6* exhibits a high frequency of dominant alleles, a property often found with genes coding for structural proteins (KEMPHUES *et al.* 1979; MACLEOD *et al.* 1977) or for proteins involved in complex aggregates such as ribosomes (DAVIES and NOMURA 1972). Mutations in a third gene, *let-54* (Figure 4) cause animals to die at the L1 to L2 molt (ROGALSKI, MOERMAN and BAILLIE 1983). The existence of several deficiencies surrounding *unc-22* (ROGALSKI, MOERMAN and BAILLIE 1983), coupled with chromosome walking data, should aid in establishing what relationship, if any, *bli-6*, *dpy-20*, *let-54* and other mutations have with collagen genes in this region.

The *unc-24 unc-22* region of chromosome *IV* also contains two genes, *daf-14* and *daf-15*, required for proper entry into the dauer larva stage (RIDDLE, SWANSON and ALBERT 1981, see Figure 4). These genes map near *col-2* which is expressed at a high level only in animals molting into dauer larvae and presumably codes for a major structural component of the dauer cuticle (J. M. KRAMER, unpublished results). Although *daf-14* and *daf-15* are presumed to affect sensory transduction (RIDDLE, SWANSON and ALBERT 1981) and are not thought to be directly involved in cuticle formation, the close proximity of these three genes involved in dauer formation is evolutionarily intriguing.

The chromosomal organization of *C. elegans* collagen genes does not appear to be strictly related to their temporal programs of expression during development (G. N. COX and J. M. KRAMER, unpublished results). Although *col-2* and *col-3* are closely linked on chromosome *IV*, *col-3* is expressed at high levels at all molts, whereas *col-2* is expressed at high levels only in animals molting into dauer larvae. Furthermore, *col-6*, which has an identical pattern of expression as *col-2*, is located on chromosome *II*.

The nature of the DNA rearrangements responsible for the observed DNA polymorphisms could be ascertained for only *col-2* and *col-3*. In both cases the DNA rearrangements involved the insertion/deletion of a DNA segment rather than to single base pair changes. This is the case for all other *C. elegans* strain DNA polymorphisms analyzed in detail (EMMONS *et al.* 1983; LIAO, ROSENZWEIG and HIRSH 1983). The *col-2* polymorphism differs from others described in that it involves only 200 bp of DNA and by the fact that the additional DNA is present in the Bristol rather than the Bergerac strain. The *col-3* polymorphism is likely to be the result of the presence of the transposable element, Tc*1*, near this gene in the Bergerac strain (EMMONS *et al.* 1983; ROSENZWEIG, LIAO and HIRSH 1983). The size of Tc*1* (1610 nucleotides) and its restriction enzyme cleavage pattern (internal sites for *Sal*I and *Xho*I and absence of sites for *Eco*RI, *Bam*HI, *Xba*I and *Hin*dIII) are similar to the characteristics of the *col-3* DNA polymorphism. The DNA rearrangements responsible for the *col-6*, *col-8* and *col-9* polymorphisms could not be analyzed in detail because the DNA regions responsible for the polymorphisms reside outside of the regions cloned in the recombinant phages containing these genes. However, the *col-8* polymorphism is seen with at least two restriction enzymes and also must be due to a DNA

rearrangement rather than to a single base pair change. The restriction fragment length differences and enzyme specificities of the *col-6*, *col-8* and *col-9* polymorphisms are consistent with the notion that these polymorphisms also may result from the presence of Tc*1* elements located near these genes in the Bergerac strain.

This work was supported by grants GM26515 and GM19851 from the National Institutes of Health. G.N.C. and J.M.K. were supported by fellowships from the Jane Coffin Childs Memorial Fund for Medical Research and the Helen Hay Whitney Foundation, respectively. Some of the nematode strains used in this study were obtained from the Caenorhabditis Genetics Center which is supported by contract N01-AG9-2113 between the National Institutes of Health and the Curators of the University of Missouri.

LITERATURE CITED

BORNSTEIN, P. and H. SAGE, 1980 Structurally distinct collagen types. Annu. Rev. Biochem. **49:** 957–1003.

BIRD, A. F., 1971 *The Structure of Nematodes.* Academic Press, New York.

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

CLEWELL, D. B., 1972 Nature of Col E_1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. J. Bacteriol. **110:** 667–676.

COX, G. N., J. M. KRAMER and D. HIRSH, 1984 Number and organization of collagen genes in *Caenorhabditis elegans. Mol. Cell. Biol.* **4:** 2389–2395.

COX, G. N., M. KUSCH, K. DENEVI and R. S. EDGAR, 1981 Temporal regulation of cuticle synthesis during development of *Caenorhabditis elegans.* Dev. Biol. **84:** 277–285.

COX, G. N., J. S. LAUFER, M. KUSCH and R. S. EDGAR, 1980 Genetic and phenotypic characterization of roller mutants of *Caenorhabditis elegans.* Genetics **95:** 317–339.

COX, G. N., S. STAPRANS and R. S. EDGAR, 1981 The cuticle of *Caenorhabditis elegans.* II. Stage-specific changes in ultrastructure and protein composition during postembryonic development. Dev. Biol. **86:** 456–470.

DAVIES, J. and M. NOMURA, 1972 The genetics of bacterial ribosomes. Annu. Rev. Genet. **6:** 203–234.

DRETZEN, G., M. BELLARD, P. SASSONE-CORSI and P. CHAMBON, 1981 A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. **112:** 295–298.

EDGAR, R. S., G. N. COX, M. KUSCH and J. C. POLITZ, 1982 The cuticle of *Caenorhabditis elegans.* J. Nematol. **14:** 248–258.

EICKBUSH, T. H. and F. C. KAFATOS, 1982 A walk in the chorion locus of *Bombyx mori.* Cell **29:** 633–643.

EMMONS, S. W., M. R. KLASS and D. HIRSH, 1979 Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans.* Proc. Natl. Acad. Sci. USA **76:** 1333–1337.

EMMONS, S. W., L. YESNER, K. RUAN and D. KATZENBERG, 1983 Evidence for a transposon in *Caenorhabditis elegans.* Cell **32:** 55–65.

FESSLER, J. H. and L. I. FESSLER, 1978 Biosynthesis of procollagen. Annu. Rev. Biochem. **47:** 129–162.

FILES, J. G., S. CARR and D. HIRSH, 1983 Actin gene family of *Caenorhabditis elegans.* J. Mol. Biol. **164:** 355–375.

GOLDSMITH, M. R. and G. BASEHOAR, 1978 Organization of the chorion genes of *Bombyx mori*, a multigene family. I. Evidence for linkage to chromosome 2. Genetics **90:** 291–310.

HERMAN, R. K., H. R. HORVITZ and D. L. RIDDLE, 1980 The nematode *Caenorhabditis elegans.* Genet. Maps **1:** 183–193.

HIGGINS, B. J. and D. HIRSH, 1977 Roller mutants of the nematode *Caenorhabditis elegans.* Mol. Gen. Genet. **150:** 63–72.

KEDES, L. H., 1979 Histone genes and histone messengers. Annu. Rev. Biochem. **48:** 837–870.

KEMPHUES, K. J., R. A. RAFF, T. C. KAUFMAN and E. C. RAFF, 1979 Mutation in a structural gene for a β-tubulin specific to testes in Drosophila melanogaster. *Proc. Natl. Acad. Sci. USA* **76:** 3991–3995.

KRAMER, J. M., G. N. COX and D. HIRSH, 1982 Comparisons of the complete sequences of two collagen genes from *Caenorhabditis elegans.* Cell **30:** 599–606.

LIAO, L. W., B. ROSENZWEIG and D. HIRSH, 1983 Analysis of a transposable element in *Caenorhabditis elegans.* Proc. Natl. Acad. Sci. USA **80:** 3585–3589.

LONG, E. O. and I. B. DAWID, 1980 Repeated genes in eukaryotes. Annu. Rev. Biochem. **49:** 727–764.

MACLEOD, S., R. H. WATERSON, R. M. FISHPOOL and S. BRENNER, 1977 Identification of the structural gene for a myosin heavy chain in Caenorhabditis elegans. J. Mol. Biol. **114:** 133–140.

PETERSON, N. S., G. MOLLER and H. K. MITCHELL, 1979 Genetic mapping of the coding regions for three heat-shock proteins in *Drosophila melanogaster.* Genetics **92:** 891–902.

POLITZ, J. C. and R. S. EDGAR, 1984 Overlapping stage-specific sets of numerous small collagenous polypeptides are translated *in vitro* from *Caenorhabditis elegans* RNA. Cell **37:** 853–860.

RIDDLE, D. L., M. M. SWANSON and P. S. ALBERT, 1981 Interacting genes in nematode dauer larva formation. Nature **290:** 668–671.

RIGBY, P. W. J., M. DIECKMANN, C. RHODES and P. BERG, 1977 Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113:** 237–251.

ROGALSKI, T. M., D. G. MOERMAN and D. L. BAILLIE, 1983 Essential genes and deficiencies in the *unc-22 IV* region of *Caenorhabditis elegans.* Genetics **102:** 725–736.

ROSE, A. M., D. L. BAILLIE, E. P. M. CANDIDO, K. A. BECKENBACH and D. NELSON, 1982 The linkage mapping of cloned restriction fragment length differences in *Caenorhabditis elegans.* Mol. Gen. Genet. **188:** 286–291.

ROSENZWEIG, B., L. W. LIAO and D. HIRSH, 1983 Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Res. **11:** 4201–4209.

SHIMIZU, A., N. TAKAHASHI, Y. YAOITA and T. HONJO, 1982 Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. Cell **28:** 499–506.

SINGH, R. N. and J. E. SULSTON, 1978 Some observations on moulting in *Caenorhabditis elegans.* Nematologica **24:** 63–71.

SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

STEINMETZ, M., A. WINOTO, K. MINARD and L. HOOD, 1982 Clusters of genes encoding mouse transplantation antigens. Cell **28:** 489–498.

Communicating editor: R. K. HERMAN