Meiotic Recombination, Noncoding DNA and Genomic Organization in *Caenorhabditis elegans*

T. M. Barnes,* Y. Kohara,† A. Coulson‡ and S. Hekimi*

**Department of Biology, McGill University, Montréal, Quebec, H3A 1B1 Canada, †Gene Library Lab, National Institute of Genetics, Mishima 411, Japan, and ‡Sanger Center, Hinxton Hall, Hinxton, Cambridgeshire, CB10 1RQ, England*

Manuscript received February 22, 1995
Accepted for publication May 31, 1995

ABSTRACT

The genetic map of each *Caenorhabditis elegans* chromosome has a central gene cluster (less pronounced on the *X* chromosome) that contains most of the mutationally defined genes. Many linkage group termini also have clusters, though involving fewer loci. We examine the factors shaping the genetic map by analyzing the rate of recombination and gene density across the genome using the positions of cloned genes and random cDNA clones from the physical map. Each chromosome has a central gene-dense region (more diffuse on the *X*) with discrete boundaries, flanked by gene-poor regions. Only autosomes have reduced rates of recombination in these gene-dense regions. Cluster boundaries appear discrete also by recombination rate, and the boundaries defined by recombination rate and gene density mostly, but not always, coincide. Terminal clusters have greater gene densities than the adjoining arm but similar recombination rates. Thus, unlike in other species, most exchange in *C. elegans* occurs in gene-poor regions. The recombination rate across each cluster is constant and similar; and cluster size and gene number per chromosome are independent of the physical size of chromosomes. We propose a model of how this genome organization arose.

THE haploid genome of the nematode *Caenorhabditis elegans* consists of some 100 megabase pairs (mb) of DNA distributed among five autosomes and an *X* chromosome all of roughly similar size (BRENNER 1974; SULSTON and BRENNER 1974; ALBERTSON and THOMSON 1982). Of the ~12,500 genes currently estimated in this genome (WATERSTON *et al.* 1992; WILSON *et al.* 1994; S. JONES, personal communication), more than 1000 are defined by mutations and have been genetically mapped (BRENNER 1974; EDGLEY and RIDDLE 1993). Since the construction of the first genetic map for this organism (BRENNER 1974), it has been noted that the distribution of genes on the genetic map is strikingly nonuniform: most of the known genes on each autosome map to a central tenth of the genetic map (the cluster), with a minority of genes mapping outside this region (the arms). This bias is less pronounced for the *X* chromosome (BRENNER 1974; EDGLEY and RIDDLE 1993). In addition, on the genetic map clustering of a much smaller number of loci is often seen at the chromosomal termini [the terminal clusters; (EDGLEY and RIDDLE 1993)]. The central clusters are thought to arise for two reasons. First, the rate of recombination per base pair has been shown to be reduced with respect to the genomic average in selected small regions of the central clusters on chromosomes *I, III* and *IV* (GREENWALD *et al.* 1987; PRASAD and BAILLIE 1989; STARR *et al.* 1989). Second, the physical location of 670 random cDNA clones from a normalized library has suggested that there is a higher density of genes in the central portions of the autosomes (WATERSTON *et al.* 1992).

These observations, however, lead to many more questions about the organization of the genome in this species that call for a more global analysis of the relationship between rates of recombination and gene density. In this paper, we address these questions by undertaking a systematic quantitative analysis of recombination rates and gene density across the whole genome. We conclude that most recombinational exchange in *C. elegans* occurs in noncoding DNA. In contrast, recombination has been found to occur predominantly in gene-rich regions in other organisms (IKEMURA and WADA 1991; CIVARDI *et al.* 1994; WU and LICHTEN 1994). We also conclude that central clusters are discrete physical entities with similar properties in terms of size, gene density and recombination rates, and presumably confer some property on autosomes. The *X* chromosome has no such cluster, but we suggest that its organization is related to that found on the autosomes. Furthermore, terminal clusters appear to be distinct from central clusters in having a recombination rate similar to the arms in which they reside. We point out that the pattern of repetitive DNA in *C. elegans* conforms to theoretical predictions for a genome where most genetic exchange occurs in gene-poor regions. Also, we find that the number of genes per chromosome is strikingly similar, despite a near twofold variation in chromosome length. We present a model to suggest how clusters might have arisen. Using this model, it is possible to explain the uniformity

Corresponding author: T. M. Barnes, 1205 Dr. Penfield Ave., Montréal PQ H3A 1B1, Canada. E-mail: tomb@notung.biol.mcgill.ca

TABLE 1

Summary of recombination rate analysis

	Physical length (mb)[a]					Genetic length (cM)					Metric (kb/cM)				
			Arm					Arm					Arm		
Region	Whole	Cluster[b]	Both	Left	Right	Whole	Cluster[b]	Both	Left	Right	Whole	Cluster	Both	Left	Right
I	13.4	6.9	6.5	2.1	4.4	50.5	6.5	44.0	19.9	24.1	265	1062	147	106	182
II	16.1	7.7	8.5	5.5	3.0	50.5	4.6	45.9	23.9	22.0	319	1663	185	230	135
III	11.8	4.8	7.0	3.8	3.2	49.9	2.6	47.3	25.3	22.0	237	1846	149	150	147
IV	16.0	6.1	9.9	5.8	4.1	48.3	3.1	45.2	32.7	12.5	332	1968	220	177	331
V	21.3	7.7	13.6	7.0	6.6	50.0	5.6	44.4	21.3	23.1	426	1371	307	330	286
X	18.5	—	—	—	—	48.7	—	—	—	—	379	—	—	—	—
Unlinked[c]	2.8	—	2.8	—	—	—	—	—	—	—	—	—	—	—	—
Autosomes	81.5	33.1	48.4	—	—	249.2	22.4	226.8	—	—	327	1479	213	—	—
X + A[d]	100.0	—	—	—	—	297.9	—	—	—	—	336	—	—	—	—

[a] The physical lengths have been rounded.

[b] The genetic and physical lengths of each cluster were inferred from the Marey maps, by extrapolating the cluster slope partly into the next interval in some cases.

[c] The *X* chromosome and all the clusters are represented by single contigs. Hence the unlinked contigs probably belong in the autosomal arms, and their total physical length was included in the totals for autosomal arms and for whole autosomes.

[d] X + A refers to the whole genome.

of cluster rates of recombination and of gene number per chromosome, if we propose that the modern chromosomes arose from more compact and uniformly gene-dense chromosomes. We consider the possible function of clusters, and the basis of the differences between the *X* chromosome and the autosomes.

MATERIALS AND METHODS

Data: The source of the data described in this paper was version 2.0 of ACEDB, a *C. elegans* database (R. DURBIN and J. THIERRY-MIEG, unpublished) running data releases up to 2–14. The data and source code are available by anonymous ftp from ncbi.nlm.nih.gov (130.14.20.1) in repository/acedb, from cele.mrc-lmb.cam.ac.uk (131.111.84.1) in pub/acedb, or from lirmm.lirmm.fr (193.49.104.10) in pub/acedb. These data include the current genetic map and the genetic mapping data on which it is based (J. HODGKIN, R. DURBIN and M. O'CALLAHAN, unpublished); a near-complete physical map of the genome (COULSON *et al.* 1986, 1988, 1991), which includes the location of nearly all cloned genes and some repetitive elements (NACLERIO *et al.* 1992; CANGIANO and LA VOLPE 1993); an extensive expressed sequence tag (EST) database (WATERSTON *et al.* 1992; Y. KOHARA, unpublished results); and the emerging genomic sequence (SULSTON *et al.* 1992; WILSON *et al.* 1994).

Construction of Marey maps: *Genetic coordinates:* To analyze the rate of recombination across the genome, we have employed the method of CHAKRAVARTI (1991), using a Cartesian representation he terms "Marey maps". Using ACEDB, we first compiled a list of loci that had been cloned. We then selected those loci that formed a set such that the genetic distance between adjacent pairs could be reliably calculated from the mapping data in ACEDB. Where there were just single mapping experiments, typically, we required two-factor data where ≥1000 progeny were scored or multifactor data where ≥30 recombinants were scored. Where there were multiple data pertaining to an interval, these limits were relaxed, but consistency was required. All data that directly or indirectly could be used to provide some mapping data on an interval were reviewed. Consequently, this involved most of the 3700 quantitative mapping data in ACEDB. Because many of the loci selected for the Marey maps are commonly used mapping loci, there was often a large data set to review for any one locus. Care was taken to exclude data based on recombination in males or which were performed at a nonstandard temperature. Anomalous single data, which included those that would invert gene order or those which were quantitatively inconsistent, were not included. The remaining mapping data subset pertinent to each locus were then used to rederive the genetic distances between loci, using 95% confidence limits for the data to arrive at the most consensual value. The cumulative constraints that this places on the possible values for a genetic interval are usually large. Thus each genetic interval length is based on many mapping data and represents a consensus of all two-factor and multifactor data. Given these constraints, and the number of progeny scored for individual data, we estimate that all intervals are correct to within 10%. Because most distances are quite small, we have allowed that the recombinant fraction (*RF*) = distance in centimorgans (cM). Only for three intervals is the distance based on single measurements >10% *RF* (left arm of chromosome *I*, left arm of chromosome *II*, and right arm of chromosome *III*), though all are <25% *RF*. In any case, we argue (see RESULTS) that multiple exchanges are rare in *C. elegans*, so that percent *RF* should closely approximate distance in centimorgans up to at least 25% *RF*. Setting the conventional left end of each linkage group at coordinate 0 cM, the coordinate of each locus in the set was then derived by summing genetic distances. The termini of the linkage groups were arbitrarily equated with the most extreme known loci. This assumption will underestimate the genetic length of the ultimate intervals, and so these are drawn with a dashed line (see Figure 1, and Figures 4 and 5). However, the proximity of these loci, where cloned, to the physical map termini, suggests that the underestimate will generally be modest.

Physical coordinates: To derive the physical coordinates of the loci in the set, it was necessary to make some assumptions concerning the physical map. First, the length unit of the *C. elegans* physical map is the fingerprint band, which is not strictly a physical measure (COULSON *et al.* 1986). However,

one can convert from bands to kilobase pairs (kb) using the following conversion factor, which should be accurate over longer distances. From 2,491,050 bp of sequence in ACEDB spanning 1360 bands in the central part of chromosome *III*, one can derive an average of 1.83 kb per band, similar to the value obtained over the whole genome by dividing the average kilobase pair per cosmid (40) by the average number of bands per cosmid (23) (COULSON *et al.* 1986). Second, not all chromosomes are represented by a single contig (*i.e.*, a set of overlapping clones). For each unjoined contig pair, we assumed an arbitrary physical interval of 100 fingerprint bands (183 kb), as this is likely to be the minimum size of such regions. Indeed, some of these gaps are probably the location of those contigs not yet assigned to a chromosome (see below). The genetic coordinates of the contig ends were then calculated by linear interpolation. Third, many parts of the physical map are spanned only by YAC clones (so called "YAC bridges"), where the size of the gap, in terms of band number, is unknown (COULSON *et al.* 1988). These are drawn on the physical map as spanning 10 bands (18 kb) based on approximate measurements at several such bridges (COULSON *et al.* 1988), but may in some instances be much greater (BARNES 1991). This distance was not adjusted for this analysis. Finally, there remain some 16 contigs composed of cosmid and YAC clones in data release 2–14, which are not assigned to chromosomes, encompassing some 2.8 mb (3% of the genome). Because all autosomal clusters are represented by an unbroken contig, and the *X* chromosome is represented by a single contig, these unlinked contigs probably derive from the autosomal arms, and were included in the appropriate totals of Tables 1 and 2. Thus the leftmost band of the leftmost contig of each chromosome was assigned the coordinate 0 kb, and using these assumptions, the kilobase pair coordinate of the midpoint of the molecular clones for each locus in the set was derived. The loci, and their genetic and physical coordinates, are given in Table A1 in the APPENDIX.

The physical map data used in this paper are subject to updating, and in a number of places the physical map has been assembled only tentatively (A. COULSON, unpublished), so one concern is its reliability over time. However, the effect of any revisions will be only on the physical lengths of affected intervals, as all the cloned loci in Table A1 of the APPENDIX obviously have a fixed genetic location. Furthermore, the density and redundancy of clones in the clusters (both cosmid and YAC) suggests that such revisions would occur mostly in the arms, for which we have fewer loci and for which the total physical or genetic lengths are of more relevance to us here rather than single intervals. Thus none of the major conclusions reached in this paper should change.

Construction of cDNA density plots: Two sets of cDNA clones (the "cm" set and the "YK" set) have been positioned on the physical map (WATERSTON *et al.* 1992; Y. KOHARA, unpublished results). The positioned cm set comprises 1268 clones representing >1000 cDNA species, derived from a normalized cDNA library (WATERSTON *et al.* 1992). The positioned YK set in data release 2–14 comprises 1248 species derived from a normalized mixed stage library (Y. KOHARA, unpublished data). These cDNAs have been positioned on the physical map by hybridization mainly to a reference set of 958 overlapping YACs that span the genome (COULSON *et al.* 1991). Of these 958 YACs, 15 do not have a physical map location and 46 have no physical length data from pulsed-field gels in ACEDB (precluding measuring cDNA density). Hence these YACs were excluded. In addition, six YACs (Y45D4, Y44H9, Y44G7, Y44H11, Y46A6 and Y45D12) may be mislabeled (A. COULSON, unpublished observations), and were also excluded. Finally, 11 YACs (Y9C2U, Y24H10, Y41C8, Y42B4, Y42H4, Y43A10, Y45D1, Y45D2, Y48A12, Y58A7 and

TABLE 2

Summary of gene density and *NotI* site analysis

Region	Physical length (Mb)[a]					No. cDNA hits				Gene density (cDNA hits/Mb)				*NotI* sites			*NotI* sites/Mb		
			Arm					Arm				Arm							
	Whole	Cluster[b]	Both	Left	Right	Whole	Cluster[b]	Both	Left	Right	Whole	Cluster	Both	Left	Right	Whole	Cluster[b]	Arm	Whole	Cluster	Arm
I	13.4	6.9	6.5	2.1	4.4	480	343	137	51	86	35.9	49.7	21.1	24.3	19.6	37	16	21	2.8	2.3	3.2
II	16.1	7.7	8.5	5.5	3.0	466	331	135	78	57	28.9	43.3	15.9	14.2	19.2	49	23	26	3.0	3.0	3.1
III	11.8	6.8	5.0	3.1	1.9	472	353	119	60	59	39.9	51.9	23.7	19.4	30.6	33	24	9	2.8	3.5	1.8
IV	16.0	7.2	8.8	4.7	4.1	498	316	182	83	99	31.0	43.9	20.6	17.7	23.9	39	15	24	2.4	2.1	2.7
V	21.3	8.5	12.8	6.2	6.6	496	298	198	95	103	23.3	35.1	15.5	15.3	15.6	53	21	32	2.5	2.5	2.5
X	18.5	10.0	8.5	5.9	2.6	479	307	172	123	49	25.9	30.7	20.3	20.8	19.0	55	28	27	3.0	2.8	3.2
Unlinked[c]	2.8	—	2.8	—	—	39	—	39	—	—	13.8	—	13.8	—	—	11	—	11	3.9	—	3.9
Autosomes	81.5	37.0	44.5	—	—	2451	1641	810	—	—	30.1	44.3	18.2	—	—	222	99	123	2.7	2.7	2.8
X + A[d]	100.0	47.0	53.0	—	—	2930	1948	982	—	—	29.3	41.4	18.5	—	—	277	127	150	2.8	2.7	2.8

[a] The physical lengths have been rounded.
[b] The physical length of each cluster was inferred from Figure 7 and is thus based on gene density.
[c] The *X* chromosome and all the clusters are represented by a single contig. Hence the unlinked contigs probably belong in the autosomal arms, and their physical length, cDNA, and *NotI* site counts were included in the totals for autosomal arms and for whole autosomes.
[d] X + A refers to the whole genome.

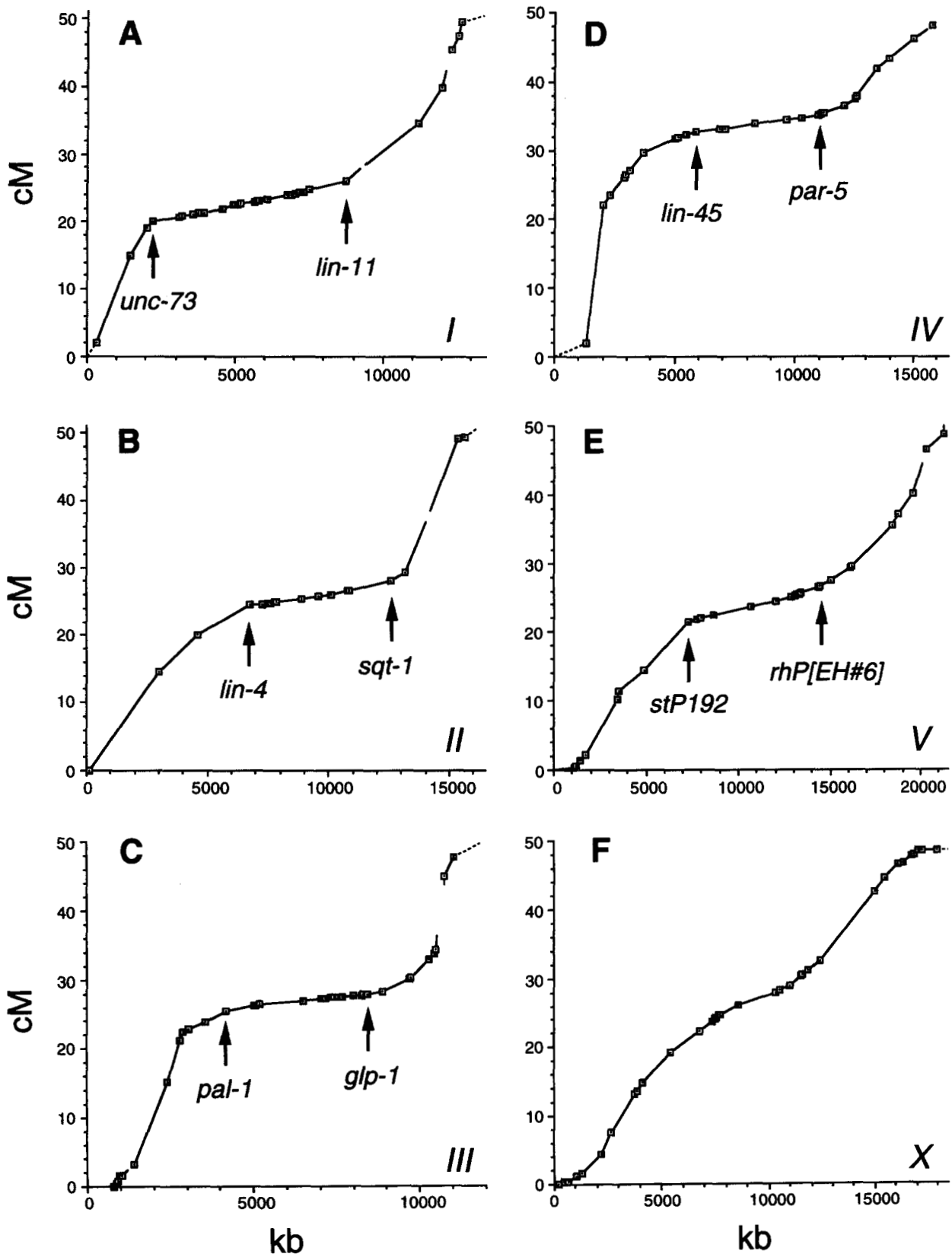

FIGURE 1.—Marey maps for the six chromosomes of *C. elegans*. (A–F) These correspond to chromosomes *I–X*. Some cluster boundary regions are drawn in expanded scale in Figure 2 to show that each cluster appears to have a discrete boundary. The arrows indicate the purely linear extents of the cluster regions. There is no apparent cluster for the *X* chromosome. Note that the physical scale for each plot is different to enhance visibility, although the genetic scale is the same. In each panel, each contig is represented by a single line; breaks in the line correspond to contig breaks (see Table A1 of the APPENDIX for

Y64G5) had clearly anomalously few hits and were excluded. These discrepancies are probably all false negatives, arising from poor growth or replication of these clones in the reference set (A. COULSON, unpublished). While their inclusion would not affect any major conclusion reached here, they would distort the gene density values in their vicinity because of the smoothing function (see below). The remaining 880 YACs were distributed as follows: 111 on chromosome *I*, 148 on *II*, 99 on *III*, 147 on *IV*, 193 on *V*, 162 on *X*, and 21 to unassigned contigs. The average resolution of positioning of cDNAs by hybridization afforded by this YAC set is ~100 kb (WATERSTON *et al.* 1992), and well in excess of 99% of cDNAs become satisfactorily located on the physical map in this way.

For each of these 880 YACs, the number of cDNA hits (corrected for redundant clones) and the length of the YAC (estimated from pulsed-field gel mobility data in ACEDB) were tabulated. These YACs ranged in size from 50 to 680 kb (mean 261) with the central two quartiles between 190 and 330 kb. The smallest and largest YAC for each chromosome was in the range 50–70 and 580–680, respectively. The number of hits per YAC ranged from 0 to 35 (mean 7.5), with the central two quartiles between 2 and 11. To minimize high-frequency noise in the data while retaining larger-scale features, the values were smoothed by summing hits and lengths over a window of three YACs, which were then divided to obtain the cDNA density, expressed as hits/megabase pair. To determine the physical coordinate of each YAC, the midpoint was calculated (from the physical map coordinates) and converted to the same kilobase pair scale as used for the Marey maps, as described above. All broad peaks and valleys were checked in detail to see if they could possibly be interpreted as sampling artifacts, and all appear to be real.

cDNA counts: The cDNA counts in Table 2 were obtained by dividing each chromosome into three regions, the left arm, the right arm and the cluster, based on the density-defined boundaries (see RESULTS). Within each region the total number of cDNA species that hybridized to a YAC in the region was enumerated. All YACs except the 11 anomalous ones and the six possibly mislabeled ones (see previous section) were included.

*NotI sites: Not*I restriction endonuclease sites (GCGGCCGC) should occur rarely in a genome with a low (G+C) content. The (G+C) content of the 100 Mb *C. elegans* genome is 36% (SULSTON and BRENNER 1974), so that for a random sequence of this composition, there should be only one *Not*I site per 900 kb, or 110 sites in the genome. A *C. elegans* library of *Not*I/*Sau*3AI genomic DNA fragments inserted into a loristseries cosmid vector has been described (GIBSON *et al.* 1987). These clones have been given a "W" prefix. The 867 positioned W clones are mostly found to cluster around or flank certain points in the genome, consistent with these points being presumptive *Not*I sites. Such clustered W clones comprise 751 of the 867 total, and identify 183 likely *Not*I sites. The remaining 116 W clones are solitary, providing a first estimate of 183 + 116 = 299 sites. In several instances, solitary or multiple W clones flank YAC gaps. It is possible that in such instances, these pairs of clone groups define single *Not*I sites, rather than a pair of neighboring sites. To obtain a minimum *Not*I site count, then, we can allow that such paired sites are really a single *Not*I site. These paired sites account for 27 solitary clones and 17 groups of clones. Thus the estimated minimum total of *Not*I sites is 299 − (27 + 17)/2 = 277 sites, including 89 solitary clones. The fact that this is larger than the predicted 110 sites suggests (not surprisingly) that the 100 Mb *C. elegans* genome is not well-modeled by random sequence.

This estimate of the total number of *Not*I sites can be refined. We can ask: do some *Not*I sites have no associated W clone (increasing the estimate); and, are there some W clones not bounded by a *Not*I site (decreasing the estimate)? In the genomic sequence available for the cluster on chromosome *III*, there are 10 *Not*I sites present. All of these occur near W clones or W clone clusters, accounting for 29 W clones. In another region between *unc-38* and *dpy-5* on chromosome *I*, there are nine W clones associated with two *Not*I sites that define a 340-kb *Not*I fragment (H. BROWNING and J. PAULSEN, personal communication). This suggests that nearly all *Not*I sites in the genome (12/12 in these examples, accounting for 38 clones) will indeed have an associated W clone. However, there are two solitary W clones in the sequenced region (W04B2 and W06F7) that are not near *Not*I sites. For each, however, one end is near a sequence that matches 7/8 bases of a *Not*I site. There is also one solitary W clone (W01F6) within the same 340-kb *Not*I fragment mentioned above, thus clearly not being associated with a *Not*I site. This means that some W clones (roughly 3/41 as for these examples) may have been created by "star" activity of the *Not*I restriction enzyme during the creation of the library. These "star clones" may be a significant fraction of the solitary clones, because in these regions considered here, three of six solitary clones were star clones. So of the 89 remaining solitary clones, if half of these are star clones, the estimate of the minimum *Not*I site number can be corrected to 277 − (89/2) = 233, and therefore the total number in the genome, allowing also for a few missed sites, will probably be in the range 240 ± 10. However, rather than arbitrarily excluding certain *Not*I sites, the unrefined estimate of 277 sites was used. The positions of these inferred sites were used to compile the data in Table 2.

RESULTS

Comparison of physical and genetic maps by Marey maps: To determine the rate of recombination across the genome, we plotted the genetic map position of a set of loci for each chromosome as a function of the physical location of each locus (see MATERIALS AND METHODS). In this Cartesian representation, the origin corresponds to a chromosome terminus (or the first locus in a set), one axis corresponds to the genetic position, and the other axis corresponds to the physical position. Thus each locus can be plotted as a unique point. The points are connected by straight lines, and thus the rate of recombination across an interval is given by the slope of the connecting line. The advantage of this format [a "Marey map"; (CHAKRAVARTI 1991)] is that only the primary data are plotted, allowing a better assessment of the effects of uncertainties in physical or genetic location on the recombination rate, as discussed by CHAKRAVARTI (1991). As also pointed out, it is possible to display different data sets for the same, or even different, chromosomes on the same pair of axes, which greatly facilitates comparative analysis. Each locus used here has a physical location

coordinates). The genetic coordinates of the contig breaks were determined by linear interpolation. The ultimate intervals for each chromosome are drawn as dashed lines, because the genetic distance from the last known locus to the true genetic terminus is unknown (see MATERIALS AND METHODS).

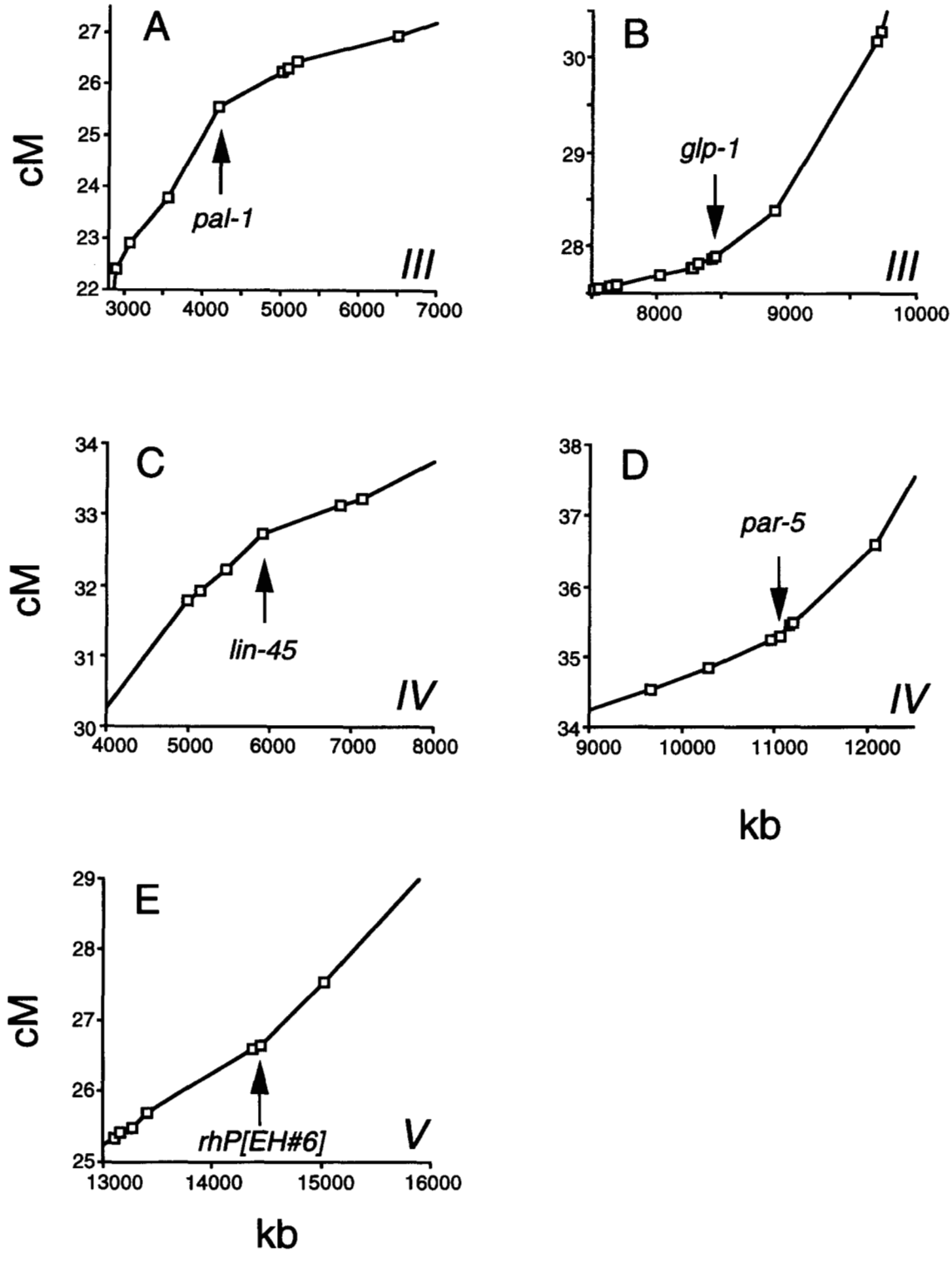

FIGURE 2.—Cluster boundaries where the change in recombination rate is more gradual. (A and B) The left and right boundaries of chromosome *III*; (C and D) The right and left boundaries for chromosome *IV*; and (E) The right boundary of chromosome *V*. The arrow indicates the last locus judged to be in the cluster, adjacent to which there is a distinct change in recombination rate.

and has been sufficiently well mapped so that the recombinational distance to both neighbors in the set can be reliably quantified (see MATERIALS AND METHODS). The 178 loci satisfying these criteria (see APPENDIX) were assembled into the maps shown in Figure 1, which will be discussed below. In these maps, the curves are in principle monotonically increasing (due to the colinearity of the physical and genetic maps). Regions with

a small gradient have low rates of recombination, while regions with a large gradient have high rates of recombination. The shapes of the Marey map curves are based on two independent types of interval measurement: physical and genetic. Physical interval lengths have been computed using certain assumptions (see MATERIALS AND METHODS), but any effects of these assumptions are likely to be systematic or present in only high resolution maps. Genetic interval lengths used here are mostly based on multiple mapping experiments which converge on a value and should be accurate to ~ 10% (see MATERIALS AND METHODS). Thus features of the Marey maps discussed below should be stable over time. Indeed, these features have been evident from the earliest maps constructed (in 1992) with a much more limited data set (T. M. BARNES, unpublished data).

All of the genetic map distances in this paper [and in the standard genetic map (EDGLEY and RIDDLE 1993)] are based on recombination in the hermaphrodite germ line. Thus all of the conclusions here pertain to this tissue. Those mapping data in ACEDB that measured recombination in the male germ line (there are only a handful) were not considered in this analysis. Recombination in males is reduced with respect to the standard map for most intervals in which it has been examined (ZETKA and ROSE 1990), and it appears that the regulation and distribution of exchanges is different in this sex (HODGKIN *et al.* 1979; ZETKA and ROSE 1990).

In this paper we use the term "rate of recombination" nominally to mean the rate of genetic change with respect to physical distances, which could be expressed as centimorgans/kilobase pair (or centimorans/megabase pair, more practically). However, for numerical comparisons, we prefer to present the values as the "metric", or the length of DNA necessary to traverse a constant genetic distance, expressed as kilobase pairs/centimorgan, which is more useful for positional cloning. Thus a high metric means a low rate of recombination and vice versa.

Autosomal clusters have the same uniform rate of recombination: We shall first consider the autosomes. An examination of Figure 1 shows that each of the autosomes has a well defined central region with lower-than-average rates of recombination (the average rate of recombination is defined by a straight line connecting the origin with the rightmost point). We will call these regions "metrically defined" clusters to distinguish them from the "genetically defined" clusters as marked on the genetic map (EDGLEY and RIDDLE 1993) (when unqualified, we will hereafter mean metrically defined). At a number of cluster boundaries, there is a sharp transition in recombination rate (*i.e., I* left, *II* right and *V* left; see Figure 1). For others, the transition appears more gradual. In some gradual transitions, the transition involves a single, apparently junctional, interval that is part cluster and part arm (*i.e., I* right and *II* left), thus probably reflecting the limited resolution of the maps in these regions. In the remaining cases, the rate of change appears to be truly gradual (*i.e.,* both boundaries of *III* and *IV*, and *V* right), so by metric alone, the precision in the placement of these boundaries is less than for the sharper transitions. However, in these regions, there still is a point at which the slope begins to increase monotonically away from cluster values (Figure 2). Wherever the precise point, clusters are therefore discrete, implying that there is a point at which the property responsible for cluster uniformity ceases to hold.

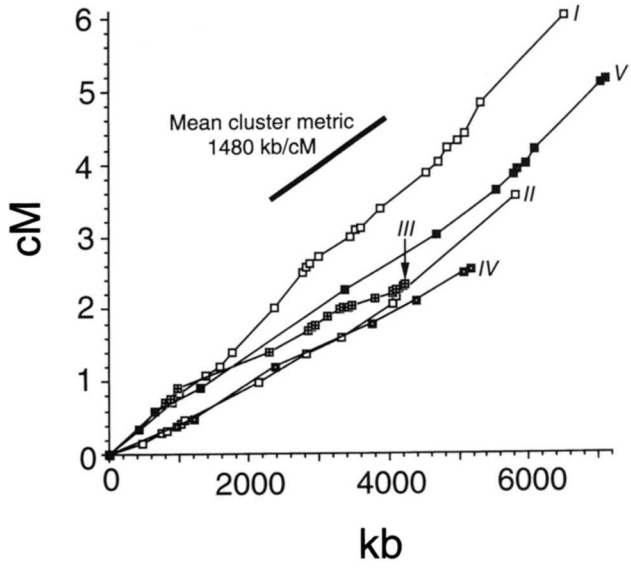

FIGURE 3.—Marey maps for the five autosomal clusters. Note that the overall slope of each cluster is very similar (~1500 kb/cM) and that most subintervals have slopes similar to this (the mean value is from Table 1). This plot only includes the purely linear intervals of the cluster (*i.e.,* the regions between the arrows in Figure 1). For any given autosome, the cluster may extend partially into the next interval.

When the slopes of these clusters are amplified and compared with each other (Figure 3), two major points emerge. First, the Marey map plot for each cluster is very close to linear, implying that there is a constant rate of recombination across each region. This observation has significant practical consequences for those pursuing a positional cloning strategy for genes located in clusters. That is, using flanking markers that have both well-defined physical and genetic locations, three-factor genetic data can be used to accurately interpolate the physical location of the gene of interest regardless of the size of the interval in question. Second, the actual rate of recombination exhibited by the different clusters is similar (1500 kb/cM), all being within a factor of two (Table 1). This suggests that the particular recombination rate of clusters arises for the same reason, and that their genomic organization is coherent across the cluster.

Cluster length and position: We have used the linearity of the clusters' rate of recombination to estimate the position of the cluster boundaries, and hence the

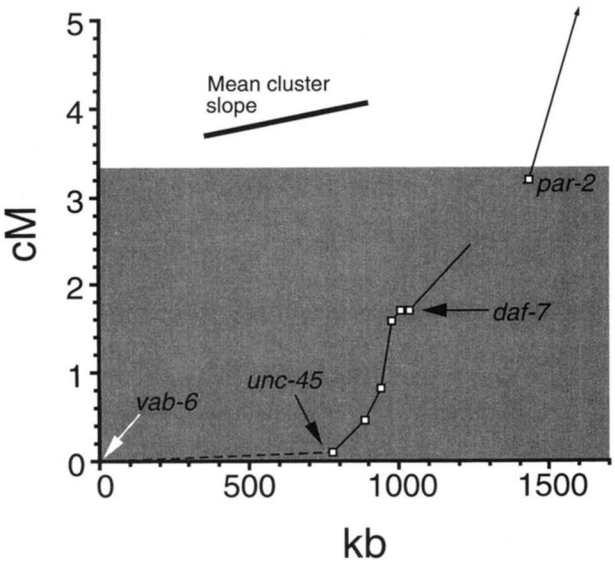

FIGURE 4.—The left terminus of chromosome *III*. An expanded view showing the genetically defined terminal cluster, extending from *vab-6* (the leftmost locus of the chromosome, plotted as the terminus) to *par-2*. *vab-6* is not yet cloned, but probably lies physically much closer to *unc-45* than the figure suggests. Hence it is indicated with a white arrow. The shaded region corresponds to the location of the genetic cluster, when projected onto the ordinate. The ultimate interval is drawn as a dashed line (see Figure 1 legend and MATERIALS AND METHODS for explanation). The line break corresponds to a contig break, and is drawn as spanning 183 kb, as explained in MATERIALS AND METHODS. The slope in the terminal cluster is clearly more similar to that of the adjacent arm segment (extending rightwards from *par-2*) than it is to the sample line drawn at the mean cluster slope.

lengths of the clusters (Table 1). From these values, it is clear that the physical length of the cluster does not covary with the physical length of the whole autosome. For example, chromosome *V* is >50% longer than chromosome *I*, but the cluster is essentially the same length (Table 1; Figure 3). In fact, apart from chromosome *III*, the physical lengths of clusters are quite similar at 7 ± 1 mb (Table 1; Figure 3 shows only those intervals fully contained in the cluster). Thus differences in autosome physical length arise principally from differences in the physical lengths of arms (Table 1).

From Figure 1, it can be seen that the physical position of the clusters on the autosomes is generally toward the center, although there are exceptions (*e.g.*, chromosome *II*). Thus clusters are central, but not necessarily centered on the physical chromosome.

Autosomal arms vary in physical rather than genetic length: Each of the six chromosomes sums to a total genetic length of ~50 cM (Figure 1; Table 1), implying one crossover per bivalent per meiosis on average in hermaphrodites [similar conclusions have been reached previously by others using an independent type of genetic argument (HERMAN and KARI 1989; MCKIM *et al.* 1993)]. With a mean of 1 event, if achiasmate bivalents are rare, then doubly chiasmate bivalents must also be rare, and thus there will mostly be only a single crossover per bivalent per meiosis [see HERMAN and KARI (1989) for a similar argument for the *X* chromosome]. This would predict strong recombinational interference in *C. elegans*. For the single chromosome for which interference has been measured in the hermaphrodite (the *X* chromosome), interference was complete [no double crossovers were seen (HODGKIN *et al.* 1979)].

Given this similarity in total genetic length, the combined genetic length of the arms of each chromosome is also found to be similar, at ~45 cM (Table 1). Also, on the genetic map the clusters are generally quite centered, so the genetic length of each arm is typically 20–25 cM (Figure 1; Table 1). The exception is chromosome *IV*, the left arm of which is nearly three times the genetic length of the right arm (32 *vs.* 12 cM, respectively; Figure 1D; Table 1). The asymmetry of the two arms of chromosome *IV* arises principally from a single genetic interval (*daf-1* to *lin-1*), which spans 20 cM (and <1 Mb, for a metric of <50 kb/cM; see Table A1). The Marey map plot to the right of *lin-1* is much more symmetrical with respect to the cluster (Figure 1D). This suggests that there is a profound hotspot for recombination in this interval. Interestingly, this is a subinterval of one of two map intervals shown to have reduced recombination in homozygous *him-1* hermaphrodites [the other is on the *X* chromosome (HODGKIN *et al.* 1979)].

Considering now the rates of recombination in the autosomal arms, it is found that they vary considerably (Table 1), but this variation arises principally from the variation in physical length of the arms. That is, each arm receives roughly the same number of recombination events as argued above, but for the longer chromosomes these events are distributed across much greater physical lengths, reducing the apparent rate of recombination. Thus, the metric is ~100 kb/cM for the left arm of chromosome *I*, which is physically short, and is >300 kb/cM for the left arm of chromosome *V*, which is physically three times longer (Table 1).

Although the recombination rates in chromosome arms are unequal, we can ask if the rate of recombination is uniform across an arm (other than for the left arm of chromosome *IV*). A simple examination of Figure 1 shows that for almost all arms with a number of points sufficient for a meaningful statement, this does not appear to be the case (*e.g.*, both arms of chromosome *III* and the right arm of chromosome *IV*). This conclusion is consistent with a higher resolution study of the right arm of chromosome *IV* (BARNES 1991; T. M. BARNES, unpublished results).

Terminal clusters are not the same as central clusters: Apart from the central clusters, most chromosomes exhibit genetic clustering of genes at the termini of the genetic map. Such terminal clusters usually consist of 5–10 identified loci in the most terminal 1–2 cM of the genetic map (EDGLEY and RIDDLE 1993). They are evi-

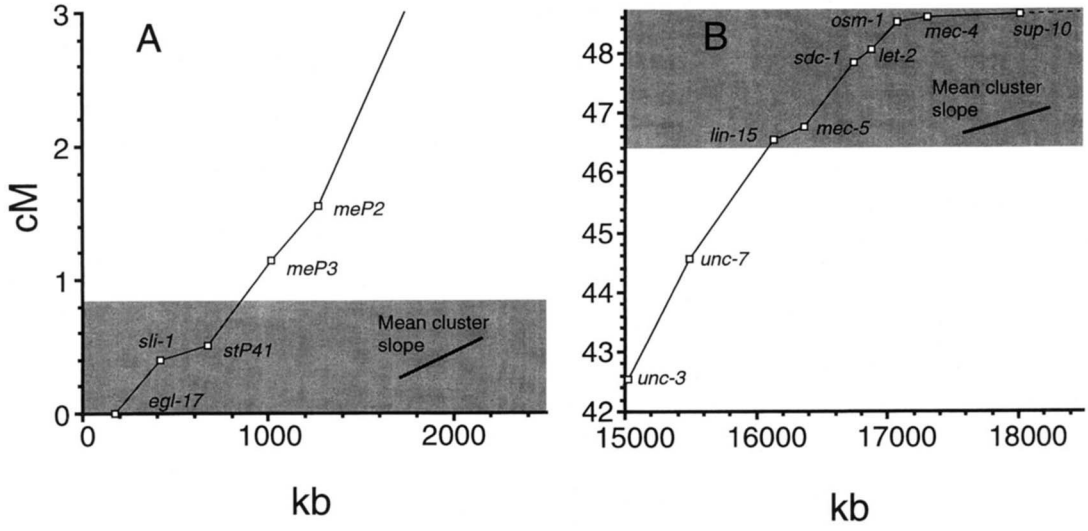

FIGURE 5.—The terminal clusters of the *X* chromosome. (A and B) The left and right terminal clusters, respectively. The left cluster extends from *egl-17* (the leftmost locus) to *unc-1* (not yet cloned), and the right cluster extends from *lin-15* to *let-6* (the rightmost locus, not yet cloned). Other drawing conventions as for Figure 4. Both termini have rates of recombination greater than that found in the autosomal clusters, except for the rightmost 0.2 cM, which has rates of recombination much less than that found in autosomal clusters.

dent for example at the right terminus of chromosome *I* (*IR*), *IIR*, *IIIL*, *IVL*, and both ends of chromosomes *V* and *X*. Do these terminal genetically defined clusters also have reduced rates of recombination with respect to the genomic average, as do the central clusters? Unfortunately, only one autosomal terminus is available in sufficient resolution [*IIIL*; data from PILGRIM (1993)] given the criteria for locus selection that we have used (see MATERIALS AND METHODS). Figure 4 shows the region of *IIIL* that encompasses the terminal cluster [from the uncloned leftmost marker, *vab-6*, to *par-2* (EDGLEY and RIDDLE 1993)]. It is clear that the rate of recombination in the terminal cluster is more similar to that of arms than that of central clusters. This analysis implies that the genetic clustering here reflects the actual distribution of genes; that is, the gene density is higher here than elsewhere in the arm. This also appears to be true for the left terminal cluster on the *X* chromosome (see next section), and therefore suggests that it might be true for all terminal clusters. However, *VR* may be an exception (see gene density analysis below).

The *X* chromosome has no central cluster: The *X* chromosome has a Marey map remarkably different from that of the autosomes (Figure 1F). There appears to be no central cluster (peak metric in the central portion is 1000 kb/cM in a single interval, and no other intervals are >750 kb/cM). Instead, the rate is much more uniform, with a mean metric (excluding the right terminus) of 350 kb/cM, comparable with the genomic average of 336 kb/cM (Table 1). Thus whatever produces the autosomal clusters does not occur on the *X* chromosome. It is clear, however, that there is an excess of loci on the genetic map of *X* in the central third, between *lon-2* and *sma-5* (EDGLEY and RIDDLE 1993), corresponding roughly to physical coordinates 4500–12,000 kb. As there is no pronounced reduction in recombination rate here (Figure 1F), this excess must reflect a real bias in the physical distribution of genes (which appears to be the case; see gene density analysis below).

Both ends of the genetic map of *X* have terminal clusters, with the right terminal cluster being quite pronounced. Similarly to the left terminal cluster of chromosome *III*, the left terminal cluster here does not appear to have reduced rates of recombination compared to the adjacent arm (Figure 5A). This suggests as before that gene density is higher here. At the right terminal cluster, in contrast, there is a region with strongly reduced rates of recombination (Figure 5B). This region, extending from coordinate 17,000 kb to the end of the chromosome, has an order of magnitude less recombination than in clusters (>10,000 kb/cM in the last measurable interval). However, the genetically defined terminal cluster (EDGLEY and RIDDLE 1993) appears to begin well to the left of this region (Figure 5B), and thus occurs mostly in a region with a recombination rate not much different to the rest of the *X* chromosome. Therefore the clustering from 16,000 to 17,000 kb must also be mostly attributable to a higher gene density, but adjacent to a recombinationally suppressed region, which involves only the very tip of the genetic map. Given its proximity to the end of the chromosome and the intensity of its effects, the cause of this reduction in recombination rate is probably distinct from that of the autosomal clusters.

Gene density across the genome: While the reduced recombination rate in clusters could in principle fully account for genetic clustering, this observation in fact

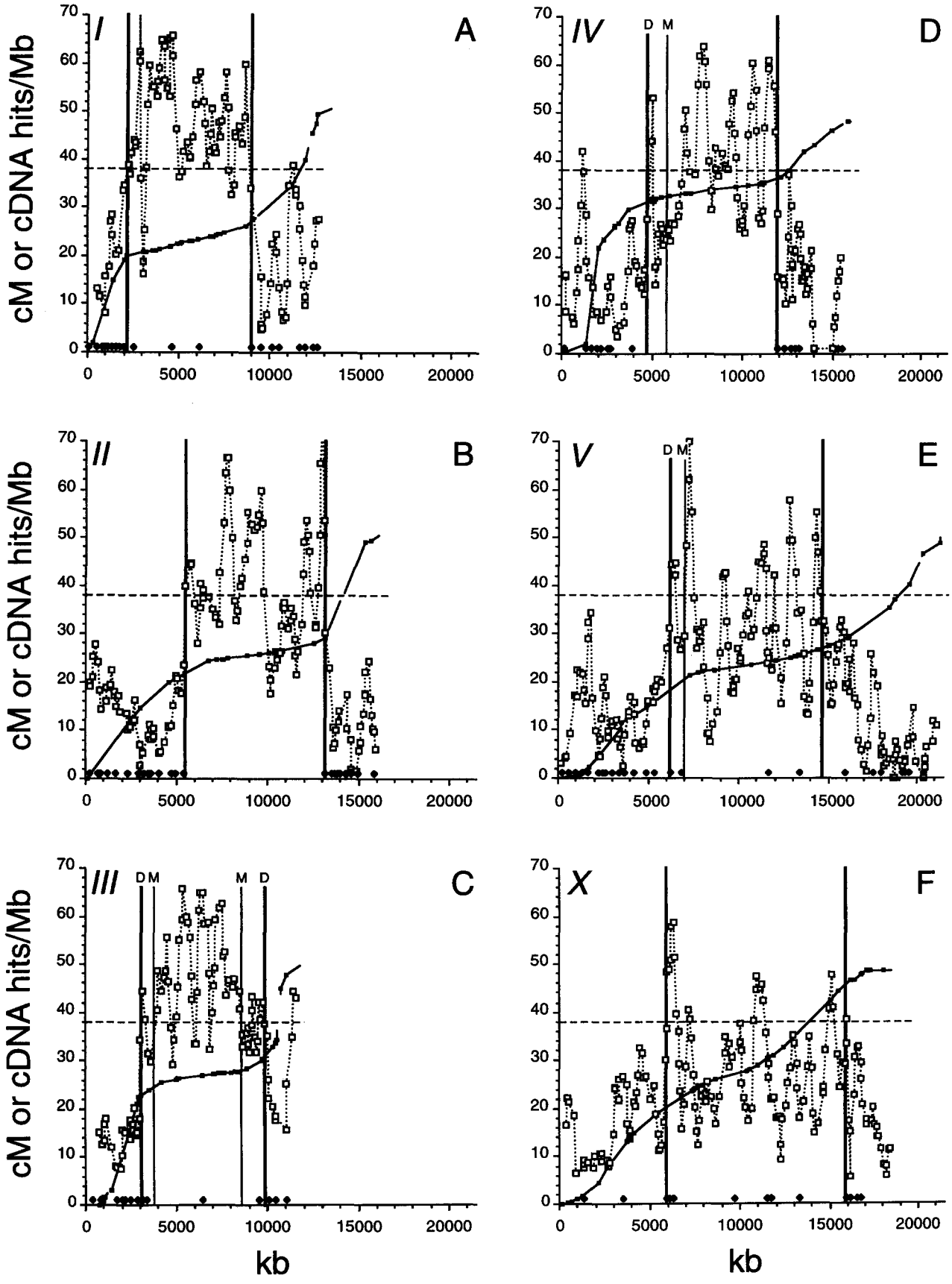

FIGURE 6.—Comparison of gene density, recombination rate, and CeRep3 distribution. (A–F) These correspond to chromosomes *I–X*. In this diagram, both horizontal and vertical axes are at the same scale for each panel. For gene density, each □ represents the mean hit rate for one YAC of the reference averaged with its two neighbors (see MATERIALS AND METHODS). At

does not reject the alternative hypothesis, namely that gene density is higher in the clusters. Indeed, it tells us nothing about the underlying physical distribution of genes. From the distribution of several hundred cDNA clones, it appeared that there was in fact an excess of genes in central autosomal regions (WATERSTON *et al.* 1992). We have now derived a map of gene density from the more advanced data obtained from the *C. elegans* cDNA mapping and sequencing projects (WATERSTON *et al.* 1992; Y. KOHARA, unpublished results). More than 2600 different cDNAs from normalized libraries have been positioned on the physical map by hybridization to a reference set of yeast artificial chromosomes (YACs) (WATERSTON *et al.* 1992; Y. KOHARA, unpublished results), which account for ~20% of the currently estimated ~12,500 total genes (WATERSTON *et al.* 1992; WILSON *et al.* 1994; S. JONES, personal communication), and should be representative of the genome. To obtain a map of gene density, we plotted the reference YACs as a function of the number of cDNA hits received, normalized to the length of the YAC (see MATERIALS AND METHODS for details). The results are presented in Figure 6 (□ and · · ·). This figure also contains several other features, which will be discussed in turn; for now, however, we will discuss only the density plots.

Autosomes: Each autosome possesses a central region of high gene density, flanked by regions of lower gene density (Figure 6). Also, the gene dense regions generally have similar physical lengths (particularly chromosomes *I* to *IV*) and well-defined boundaries. These boundaries are so distinct that we can define a single threshold density (38 hits/mb in this data set) above which in general only the central, gene-dense regions rise. Using this threshold, we can derive unambiguous boundaries for the gene-rich regions (bold vertical lines in Figure 6). We will call these regions the density-defined clusters. The average density of these clusters ranges from 35 to 52 cDNA hits/Mb (Table 2), and in shorter regions can be as high as 65 hits/mb (*e.g.*, between coordinates 3500 and 5000 kb of chromosome *I*; Figure 6A).

What actual gene densities do these values correspond to in the genome? From the genomic sequencing project, which has completed one-third (>2 mb) of the gene-rich region on chromosome *III* (WILSON *et al.* 1994) (corresponding to physical coordinates 7100–9250 kb in Figure 6C), the average gene spacing was found to be one every 4.5 kb, or 220 genes/mb. In Figure 6C, the mean hit rate in this region is 44 hits/mb, suggesting that one can convert to real values by multiplying the hit rate by 5. Thus the threshold corresponds to 38 × 5 = 190 genes/mb, or roughly one every 5 kb. Furthermore, in the 2.2 mb of sequence for chromosome *III*, 48% of the sequence was found to lie between the predicted start and stop codons of predicted genes (WILSON *et al.* 1994), showing that the density-defined clusters are very compact and will probably contain <50% intergenic DNA (that is, genes on average will probably be longer than 2.5 kb).

The mean gene densities of the autosomal arms range from 14 to 30 hits/mb (Table 2). Regions with the lowest average density (*e.g.*, 5 hits/mb between coordinates 18,000 and 19,500 on chromosome *V*; Figure 6E) are thus predicted to have an actual density of ~5 × 5 = 25 genes/mb (using the conversion factor from above), or one gene every 40 kb. Thus the difference in mean gene density between dense and sparse regions of the genome can be more than 10-fold (5–65 hits/mb). The variation between the arms in mean arm cDNA density (Table 2) arises primarily from the variation in arm length (as for arm recombination rates) because most arms have similar numbers of cDNAs. This will be discussed further below.

It was inferred above that the likely cause of the autosomal terminal clusters was a local increase in gene density. From Figure 6C, it can be seen that the left terminal region of chromosome *III* from coordinate 0 to 1500 kb (see Figure 4) does indeed have elevated gene density with respect to the adjacent arm. This appears to be true of almost all terminal regions (*e.g.*, the right end of *II*, and the left ends of chromosomes *IV* and *V*, which also correspond to terminal clusters). However, such a peak is not evident at the right terminus of chromosome *V*, which does have a terminal cluster; and the peak at the left end of chromosome *II* has no corresponding genetic cluster. This lack of correspondence in the latter case may simply be through the incomplete representation of genes in the available mutant set, but in the former case could possibly suggest reduced recombination, or alternatively a limitation in the cDNA data set for this region.

The X chromosome: The pattern of gene density on the *X* chromosome is somewhat different from that of the autosomes (Figure 6F). Formally, most of the chromosome (from 5950 to 15,950 kb) would qualify as a gene-

two points (left end of the cluster on chromosome *I* and right end of the cluster on chromosome *II*), the plot just exceeds 70 hits/mb. The horizontal line is drawn at 38 hits/Mb in each panel and can be used across the genome as a threshold to define the gene-rich central regions (bold vertical lines), and corresponds to a density of ~190 genes/mb on the chromosome (see RESULTS). The Marey map plots are the same as in Figure 1. Where unlabeled, the cluster boundaries are coincident by both gene density and recombination rate criteria. Where the boundary positions differ between these two methods, the density-defined cluster boundary is a bold line designated D; the metrically defined boundary is a thinner line designated M. The CeRep3 positions are indicated (◆) just above the abscissa. In chromosome *V* (E), the left metrically defined boundary is positioned to exclude two CeRep3 elements, as the density-defined boundary would be admissible by purely metrical criteria. For the *X* chromosome (F), the vertical lines indicate only density-defined boundaries.

rich region by the threshold criterion. From the density plots alone, chromosomes *V* and *X* seem more similar than they do to the other chromosomes (Figure 6, E and F). Both chromosomes *V* and *X* have longer and less dense clusters, although the *X* is the more extreme of the two.

The excess of loci in the central part of the genetic map of the *X* chromosome was inferred above to reflect a real bias in gene distribution. From Figure 6F, we can see that this region (~4500–12,000 kb) does indeed correspond to roughly the highest density part of the *X* chromosome. To the left of this region, the density plot resembles an autosomal arm, peaking again at the left terminus. This terminal peak corresponds to the left terminal genetic cluster (see Figure 5A), which was also predicted above to have a higher gene density.

The right terminal genetic cluster (coordinates 16,000–18,500 kb, which is essentially the whole right "arm") corresponds to a region of declining, rather than increasing, density (Figure 6F). However, only the leftmost 1000 kb of this region is outside the region of extremely low recombination rates (see Figure 5B). This region is not well resolved from adjacent regions but corresponds somewhat to a local peak in density (Figure 6F). The recombination rate over the remainder of the terminal cluster is adequately low to offset the declining density to produce genetic clustering: the gene density drops several fold, but the recombination rate decreases by an order of magnitude. Thus, in having a density-defined cluster (albeit diffuse), and the left terminal cluster, the *X* chromosome appears related to the autosomes (particularly chromosome *V*) in physical organization.

All chromosomes have a similar number of genes: An unexpected conclusion from the cDNA hit totals was that all chromosomes seem to have a very similar number of genes, even though the chromosome lengths vary over a twofold range (Table 2). Even more remarkably, the density-defined clusters, including that of the *X* chromosome, also have similar hit totals (and thus so do the arms; Table 2). This is a striking observation which cannot be simply explained. If one considers either the cm cDNA set or the YK cDNA set separately (see MATERIALS AND METHODS), one obtains the same result (not shown), so it is unlikely to reflect a bias inherent in the data. Furthermore, the total number of genetically defined loci for each chromosome in ACEDB (excluding *let* genes, which have been mostly defined only in genetically balanced regions of the genome) is 136, 132, 143, 106, 109, 126 for chromosomes *I* to *X* respectively, which shows this similarity again in a completely different type of data set. It thus seems that both the crossover number per meiotic bivalent and the gene number are not dependent on the physical length of the chromosomes, but are a fixed value for all chromosomes, including the *X* chromosome.

Gene-dense regions correlate imperfectly with recombination-poor regions: How do the density-defined clusters correspond to the metrically defined clusters? The data are compared in Figure 6 (the ordinal values in each panel are in units of centimorgans for the Marey maps and hits/megabase pair for the density plots). We can see that the boundaries are largely coincident; chromosome *I* is a good example (Figure 6A). Does this coincidence mean that high gene density causes low recombination rates? A consideration of Figure 6 shows that this conclusion is not supported. First, there are places within clusters where the gene density is quite low (*e.g.*, it drops below 20 hits/mb on chromosome *V* in several places), which do not produce a corresponding anomaly in the linear Marey map plots. Also, the right half of the cluster on chromosome *II* has a mean density different from the left half, without a corresponding difference in recombination rates. Second, the lack of correlation between the boundaries of gene density and recombination rate at some places is clear, for example the left boundary region of chromosome *IV* (Figure 6D). In this region, there are actually two feasible places to put the boundary based on density (at 4700 and 6500 kb), but one is well within the metrically defined cluster and the other well outside (we have assigned it to 4700 kb as this preserves the uniformity of both cluster length and number of cDNA hits). The region where the metric begins to change corresponds to an extended region of low gene density. Finally, the similarity of the physical organization of the *X* chromosome with at least chromosome *V* clearly illustrates that gene density does not directly influence recombination rate, because the *X* chromosome has no extended central metrically defined cluster.

These counterexamples strongly suggest firstly that regions of low gene density are not recombination-promoting *per se*, and secondly that a high density of genes is insufficient to reduce recombination rates to cluster levels. Rather, regions with low rates of exchange may be the preferred location for genes in *C. elegans*, or alternatively these features may occur in the same region for independent reasons.

Noncoding DNA and recombination rates: Sequences influencing recombination rate could arise by chance, as a consequence of differing average nucleotide composition due to the constraints imposed by encoding genes. Alternatively, the differences could arise from the presence and maintenance of particular sequence elements, which could be associated with noncoding DNA (if they promoted recombination). Below we examine simple indicators to try to distinguish between these alternatives.

There is no difference between cluster and arm in (G+C) content: The (G+C) content of the whole *C. elegans* genome is 36% (SULSTON and BRENNER 1974), while in 2.5 Mb of sequence of only the cluster on chromosome *III* (coordinates 7000–9500 kb), the (G+C) content is 37% (data from ACEDB). If this is typical of cluster DNA composition, then because the clusters occupy

more than a third of the genome (37 of 100 mb; Table 2), there can be only a small difference in (G+C) content between cluster and arm on average.

A second way to address (G+C) content is to consider the distribution of *Not*I restriction endonuclease sites. *Not*I recognizes an octamer site composed only of G·C base pairs. Regions with higher-than-average (G+C) content should have a higher density of *Not*I sites than regions with a lower-than-average (G+C) content. The W series of *C. elegans* cosmid clones were constructed such that they should all terminate in a *Not*I site (GIBSON *et al.* 1987). Correlation with the available sequence and the redundancy of the clones suggests that this set probably accounts for nearly all such sites in the genome (see MATERIALS AND METHODS). They can be used to infer the existence of 277 candidate *Not*I sites, whose occurrence in various genomic regions is given in Table 2. There is a uniformity of *Not*I site incidence regardless of the genomic region analyzed, averaging 2.7, 2.8 and 3.0 sites/mb for autosomal clusters, autosomal arms, and the *X* chromosome, respectively (the genomic average being 2.8 sites/mb, or one site every 360 kb).

In summary these results suggest that recombination rate differences are probably not secondary consequences of differences in nucleotide composition.

One known repetitive element is a marker for autosomal arm DNA: C. elegans has a significant number (100–1000) of small families (10–10^4 copies) of repetitive elements (SULSTON and BRENNER 1974; EMMONS *et al.* 1980), some of which could conceivably promote or inhibit exchanges. Several classes of these elements (in addition to known transposons) have been molecularly characterized (EMMONS *et al.* 1980; FELSENSTEIN and EMMONS 1988; LA VOLPE *et al.* 1988; NACLERIO *et al.* 1992; CANGIANO and LA VOLPE 1993), and the genomic distribution of seven classes of repeat has been described (NACLERIO *et al.* 1992; CANGIANO and LA VOLPE 1993). Five classes were distributed uniformly (NACLERIO *et al.* 1992), but two classes, the CeRep3 class (FELSENSTEIN and EMMONS 1988) and the RcS5 class (CANGIANO and LA VOLPE 1993), were found to occur preferentially outside the clusters as defined on the genetic map (CANGIANO and LA VOLPE 1993). To compare the distribution of these elements to the various genome features already described in this paper, we plotted the locations of the CeRep3 element on Figure 6 (small diamonds above the abscissa). We find that the distribution of this element respects the arm/cluster boundaries defined here in a strikingly consistent way. The transition from arm to cluster by this measure appears as sharp as that inferred from gene density. In addition, we notice that the *X* chromosome has fewer and more uniformly positioned elements than the autosomes. Consistent with the inference that there is little difference in nucleotide frequency between cluster and arm, the CeRep3 element has a (G+C) content of 34% (FELSENSTEIN and EMMONS 1988). Overall this shows that indeed there are repetitive element classes that obey the metrical arm/cluster boundary as strictly as other measures.

Most of the extra DNA in the arms will not be repetitive, however, by the following argument. If there are ~12,500 genes in *C. elegans* (WATERSTON *et al.* 1992; WILSON *et al.* 1994; S. JONES, personal communication), and each chromosome has an equal number, then there are ~2100 genes per chromosome. Each autosome has one-third of its genes in the arms, so that the five pairs of autosomal arms, which comprise 45 mb (Table 2), will contain 3500 genes. If we assume that the average mRNA length of arm genes is no different from that of cluster genes, then 3500 genes should span no more than 7000 kb (WILSON *et al.* 1994), leaving >35 mb of noncoding DNA. In contrast, the total repetitive fraction in *C. elegans* genomic DNA is 17% (SULSTON and BRENNER 1974), or ~17 mb. Even if all of this material was located in the arms and none in the clusters, which is not the case (NACLERIO *et al.* 1992; SULSTON *et al.* 1992; CANGIANO and LA VOLPE 1993; WILSON *et al.* 1994), over half of the noncoding DNA in the arm would be single copy.

Is CeRep3 a recombination-promoting element?: CeRep3 has the distribution expected for a recombination-promoting element, including a more uniform (low) abundance on the *X* chromosome to yield a more uniform recombination rate. However, there are only ~120 CeRep3 elements in the genome (FELSENSTEIN and EMMONS 1988; CANGIANO and LA VOLPE 1993), and so it is unlikely that this element would account for all of the recombination in the arms in any case (as it would imply each element is responsible for ~2 cM of recombination). Even with a more diffuse effect of this element, its accounting for all recombination seems unlikely, because there are four nearby elements on the *X* chromosome at 16,000–16,800, without an apparent effect on the Marey map curve. Also, the more isolated element at 3900 in the left arm of chromosome *IV* is not associated with a peak. Furthermore because there are CeRep3 elements in the clusters of chromosomes *I, III* and *V*, CeRep3 cannot be sufficient to promote arm-like rates of recombination. Nonetheless, it does remain possible that CeRep3, along with other elements such as RcS5 and interstitial telomeric repeats (CANGIANO and LA VOLPE 1993), does contribute to promoting recombination in the arms, as suggested by CANGIANO and LA VOLPE (1993). Interestingly, a cloned CeRep3 element has been shown to confer replication (*ARS*) and segregation (*CEN*) functions on DNA in *S. cerevisiae* (FELSENSTEIN and EMMONS 1988), which shows that it is biologically active.

Defining the cluster: Using the criteria of recombination rate, gene density and CeRep3 distribution, can we derive a consensus for cluster boundaries for each chromosome? For both chromosomes *I* and *II* these three criteria converge on unique positions (Figure 6). Similarly, using all three criteria, the right boundaries

on chromosomes *IV* and *V* can be unambiguously defined (Figure 6). The left boundaries of *IV* and *V*, and both boundaries of *III*, have two possible locations, depending on which criteria are used. Indeed, we have appropriately used metrically defined clusters for Table 1 and density-defined clusters for Table 2. Although it is true that two of the three measures (higher rates of recombination and CeRep3 distribution) correlate quite tightly, the third measure (gene density) would allow all chromosomes to be included. We therefore feel that all definitions need to be taken into account.

DISCUSSION

The organization of the genome: Using an extensive, publicly available data set, we have analyzed the organization of the genome in *C. elegans*. Each linkage group contains a very similar number of genes, which have a similar arrangement on all chromosomes (Figure 6; Table 2). Two-thirds of the genes on each chromosome are bundled into a gene-rich central cluster; the remaining third of the genes are spread out on the arms amongst mostly single-copy noncoding DNA. On the autosomes, the rate of recombination in the gene-rich regions is reduced, so that over most of the length of the density-defined cluster the rate of recombination is uniformly low and very similar between clusters (Figures 1, 3 and 6). The rate of recombination is not reduced in the density-defined cluster on the *X* chromosome (Figure 6). Most terminal regions of the chromosomes have local maxima of gene density, but the rate of recombination does not appear to be reduced in these regions (Figures 4 and 5).

The rate of recombination and the gene density in the autosomal arms varies principally because the physical length of each arm varies. Because there is one exchange per chromosome per meiosis on average in hermaphrodites and clusters are central on the genetic map (except for chromosome *IV*), all arms receive a similar number of recombination events. Furthermore, they have only one-third of the genes. Thus long arms (*e.g.*, on chromosome *V*) have generally lower gene densities and rates of recombination, while short arms (*e.g.*, on chromosome *I*) generally have higher gene densities and rates of recombination (Tables 1 and 2).

Cluster lengths do not correlate with chromosome length. Rather, there appears to be a preferred size of ~7 mb, although by the criterion of recombination rate the cluster on chromosome *III* is distinctly shorter physically, and by the criterion of gene density, the clusters on chromosomes *V* and *X* are physically longer (Tables 1 and 2). The cluster/arm boundaries are all discrete; the transition away from cluster-like properties is sharp by gene density, but varies by recombination rate, from sharp (left boundary on chromosome *I*) to gradual (right boundary of chromosome *V*; Figures 2 and 6).

Variations in gene density do not appear to drive variations in recombination rate: while the plot of gene density is quite noisy (Figure 6), this is not true for the Marey maps, especially for the *X* chromosome. A much tighter correlation with recombination rate is manifested by a class of repetitive element, which shows that both coding (cDNA) and some noncoding (CeRep3) DNA respect the cluster boundary (Figure 6). However, the differing amount of coding DNA in the clusters and arms does not appear to have skewed DNA composition appreciably (Table 2), and at least half of the noncoding DNA in the arms is nonrepetitive.

How is the metrically defined cluster specified?: One possibility is that the cluster is not specified by a genomic sequence feature, but is simply the central part of an autosome. For example, if crossovers preferentially occur about one-third of the way in from either chromosome end, then there will be a central clustering. However, clusters have discrete boundaries and uniform recombination rates; these metrical boundaries are mostly coincident with different physical boundaries such as coding and certain noncoding DNA; the cluster length is not a function of chromosome length (rather, the clusters are similar in length); and the cluster position on the physical map can be asymmetrical, all of which are not predicted by such a model.

The other possibility requires the unequal distribution between cluster and arm of genetic elements that have effects on recombination. The question of whether the arms contain a recombination-promoting element, or whether the clusters contain a recombination-suppressing element is not semantic, but embodies mechanistic differences given the evidence for a single exchange per meiosis. We shall, however, consider only the former case.

In this model, recombination in the arms promoted by one or several types of noncoding DNA element. Because the arms have more noncoding DNA, such elements could have expanded in number and be tolerated there. In this view, the effect on recombination from any one element would be small, so that complete exclusion from the cluster would not be required, but they should be largely absent there. Any sequence lacking such elements could still recombine, but at a lower frequency. Nonuniform recombination rates in the arms would thus reflect the uneven distribution of such elements. To explain the shape of the Marey map for the *X* chromosome, the *X* could be either one large arm, dotted with such elements, or one large cluster, mostly devoid of them. It would be more consistent with the density-defined clusters on other chromosomes if the *X* chromosome were one large cluster. In either case, because there is only one exchange per meiosis on average, the probability of exchange would be relatively uniform across the whole *X* chromosome. It is interesting to note that the element CeRep3 enjoys all these properties, and in addition is underrepresented on the

X chromosome. Thus the current evidence favors this model. However, as noted above (see RESULTS), the CeRep3 is unlikely to be the only such element.

Clusters, arms, and evolutionary processes: The cluster, carrying two-thirds of all the genes on the chromosome, undergoes exchange only ~10% of the time, and thus is inherited mostly as a unit. It has been argued that positive or negative selection on new mutations in such regions will tend to reduce heterozygosity at neutral sites over an extended physical region, because they are strongly linked genetically ["the hitchhiking effect" and "background selection"; (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989; CHARLESWORTH *et al.* 1993)]. Given the dramatic genome organization in *C. elegans*, hitchhiking effects will thus extend to most of the genes on any chromosome. This predicts that variation at neutral sites in autosomal clusters will be very low in natural populations, but should be higher for the *X* chromosome.

Theoretical and experimental approaches have shown that satellite DNA will accumulate in regions of the genome with low recombination that can tolerate physical expansion (CHARLESWORTH *et al.* 1986; STEPHAN 1986). In *C. elegans*, the genomic regions with fewer apparent constraints on size (the autosomal arms) have high rates of recombination, while regions with reduced recombination (the clusters) probably are under selective constraints for compactness. Therefore there is no predicted compartment that would favor satellite DNA, which might account for its absence in this species (SULSTON and BRENNER 1974; EMMONS *et al.* 1980). In contrast, the type of repeat favored in euchromatic DNA by theoretical models is more moderately repetitive and more variable between elements and is more likely to be lost between closely related species (CHARLESWORTH *et al.* 1994). This could account for the observed diversity of *C. elegans* repetitive elements (SULSTON and BRENNER 1974; EMMONS et al. 1980), and their absence from related nematode species (*e.g.*, CANGIANO and LA VOLPE 1993; KENNEDY *et al.* 1993).

Why are cluster recombination rates similar? In the model above, there is no *a priori* reason why the residual recombination in the cluster should match its physical length to produce similar rates of recombination for all the autosomes. One possible explanation is that originally, the metric and gene density may have been more uniform in the autosomes. During physical expansion of the arms, a recombination-promoting element may have arisen and proliferated, sequestering a growing fraction of the exchanges. For some reason, the element was not tolerated on the *X* chromosome. In theory, such expansion could have continued until all exchanges occurred in the arms, and none in the cluster. However, background selection against the entire cluster would then be very high against new deleterious mutations arising anywhere in the cluster (CHARLESWORTH *et al.* 1993), which carries two-thirds of the genes. Thus we propose that the "selfish" expansion of arm-specific recombination-promoting elements would have been tolerated only up to some critical point beyond which exchange in the cluster would be so low as to compromise the fitness of the chromosome. In this view, there is no obligation for a precisely identical rate of recombination in the clusters, and indeed, there is a twofold variation in recombination rates for the different clusters (Figure 3; Table 1).

Why is the gene number per chromosome similar? It is unclear how constancy of gene number could be ensured. One possibility is that it is a historical relic. Originally, the metric and gene density may have been more uniform in the autosomes (*i.e.*, cluster-like). If there had been a requirement for uniform chromosome size at this time, then gene number would obviously have been equalized. Subsequently, there may have been an expansion of noncoding DNA to different degrees in various chromosomes to give the current arrangement. It would, however, remain odd that during this time, no chromosomal rearrangement disturbed this arrangement. A fusion of chromosomes *IV* and *X* is viable as a homozygote (SIGURDSON *et al.* 1986), suggesting that unequal gene number *per se* is not acutely problematic.

What is the relationship between recombination rate and gene density? In many organisms, there is a strong correlation between the distribution of genes across the genome and the distribution of meiotic exchanges. However, in these organisms, exchanges occur preferentially in gene-rich regions, unlike in *C. elegans*. For example, in maize, most crossover events are within genes, with few crossovers occurring in intergenic DNA (CIVARDI *et al.* 1994). In vertebrates, the genome is organized into large domains (isochores) of uniform (G+C) content (BERNARDI 1993). Most genes and chiasmata are found in the highest (G+C) isochores (IKEMURA and WADA 1991; MOUCHIROUD *et al.* 1991). In Drosophila, exchange is mostly absent from centric heterochromatin, which is gene-poor (ASHBURNER 1989). Furthermore, within the gene-rich euchromatic arms, exchanges are not distributed uniformly (ASHBURNER 1989), whereas they are in *C. elegans* clusters (Figure 3). In *S. cerevisiae*, the genome is in general extremely compact (OLIVER *et al.* 1992; DUJON *et al.* 1994; JOHNSTON *et al.* 1994), so all regions are gene rich. However, most, if not all, exchanges in budding yeast can be accounted for by recombination at the open chromatin of gene promoters (WU and LICHTEN 1994), suggesting a strict association of genes and recombination. Not only are exchanges in yeast gene-associated, but in the fission yeast, suppression of recombination and transcriptional silencing are mediated by single gene products that presumably affect chromatin configuration (THON *et al.* 1994).

The situation in *C. elegans* is the converse of this. Two-

thirds of the genes on the autosomes (over half of those in the genome) are crowded into regions where exchange is modest. Thus it appears that *C. elegans* is an unusual example of an organism that prefers noncoding DNA for crossing over.

To account for cluster boundaries that do not coincide by metrical and density criteria, we propose that the different boundaries correspond to the different extents to which recombinationally inert and recombination-promoting noncoding DNA penetrated the cluster. As for the terminal clusters, these can be viewed as areas that have become populated with recombination-promoting noncoding DNA while resisting the proliferation of recombinationally inert noncoding DNA. Chromosome termini are typically associated with the nuclear envelope, and thus may represent a distinct environment within the chromosome (DE LANGE 1992).

Recombination is almost fully excluded from the right terminal cluster on the *X* chromosome. The means by which this is achieved may make gene expression difficult, which could account for the low gene density there. This would then be analogous to the "cold spot" of recombination described at the *S. pombe* mating-type locus, which is a region of genomic "silencing" (THON *et al.* 1994).

What function might clusters and arms have? In terms of the model proposed above, we presume it was for functional reasons that the cluster did not tolerate expansion or recombination. The absence of a cluster on the *X* chromosome implies that this function is not required for the *X*. Might the cluster correspond to the centromere? Cytologically, the chromosomes in *C. elegans* have no defined centromere for spindle attachment. Rather, spindles attach to a diffuse (holocentric) kinetochore in mitosis (ALBERTSON and THOMSON 1982), and directly to the chromatin at the ends of bivalents in meiosis (ALBERTSON and THOMSON 1993). Other studies have shown that in *C. elegans* there is a single site on each chromosome required for proper homolog pairing in meiosis, a function usually carried out by the centromere (ROSENBLUTH and BAILLIE 1981; HERMAN *et al.* 1982; MCKIM *et al.* 1988, 1993; HERMAN and KARI 1989). This is confined, however, to one chromosomal terminus for each chromosome (see MCKIM *et al.* 1988 for a summary). Hence the traditional functions of the centromere seem to be specified by multiple regions of the chromosome, and their distribution is different between mitosis and meiosis. The organization of autosomes into cluster and arms may reflect some of these functions. Interestingly, *C. elegans* is the first species with holocentric chromosomes to have its genome analyzed in such detail. Perhaps the unusual correspondence of low rates of recombination with high gene density will be a feature of such organisms.

Why is the *X* chromosome different from the autosomes? There are several known differences in function between the *X* chromosome and the autosomes. First, the *X* chromosome dose determines sex (MADL and HERMAN 1979) and is dosage-compensated in hermaphrodites (HODGKIN 1983; MEYER and CASSON 1986). Recently, it has been shown that the product of the dosage compensation gene *dpy-27* is specifically associated with the *X* chromosome, but only in *XX* animals, and thus presumably reduces expression of *X*-linked genes in hermaphrodites (CHUANG *et al.* 1994). This may be relevant, as genetic map distances derive from hermaphrodites, and thus the Marey map for the *X* chromosome pertains to dosage-compensated homologs. Second, in both *XX* and *XO* spermatocytes, the *X* bivalent has a distinct spatial behavior in meioses I and II with respect to the autosomes (ALBERTSON and THOMSON 1993). Finally, and perhaps most interestingly, there appear to be recombination functions that specifically affect *X* chromosome disjunction when absent (HODGKIN *et al.* 1979; HERMAN and KARI 1989). The interdependence of these features is unknown; however, it seems reasonable that the distinct organization of the *X* chromosome is a consequence of at least one of the above functional distinctions.

We would like to thank the *C. elegans* community for their continued sharing of unpublished genetic and molecular data, which make ACEDB such a rich information resource. We particularly thank A. ALFONSO, J. AHRINGER, L. BERKOWITZ, H. BROWNING, R. HORVITZ, Y. JIN, C. KENYON, S. KIM, J. PAULSEN, W. SCHAFER, J. SIMSKE and S. STROME for sharing data not yet in ACEDB, and colleagues for comments on the manuscript. T.M.B. and S.H. were supported by a grant from the Medical Research Council of Canada to S.H. T.M.B. was also supported by a Helen Hay Whitney Fellowship during the early stages of this work. Y.K. was supported by Grants-in-aid for Creative Basic Research (Human Genome Program) and for Scientific Research on Priority Areas (Genome Informatics) from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

ALBERTSON, D. G., and J. N. THOMSON, 1982 The kinetochores of *Caenorhabditis elegans*. Chromosoma **86:** 409–428.

ALBERTSON, D. G., and J. N. THOMSON, 1993 Segregation of holocentric chromosomes at meiosis in the nematode, *Caenorhabditis elegans*. Chromosome Res. **1:** 15–26.

ASHBURNER, M., 1989 Mapping and exchange, pp. 451–501 in *Drosophila. A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

BARNES, T. M., 1991 Molecular analysis of the *tra-3* region of the *Caenorhabditis elegans* genome. Ph.D. Thesis, University of Cambridge, England.

BERNARDI, G., 1993 The isochore organization of the human genome and its evolutionary history—a review. Gene **135:** 57–66.

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

CANGIANO, G., and A. LA VOLPE, 1993 Repetitive DNA sequences located in the terminal portion of the *Caenorhabditis elegans* chromosomes. Nucleic Acids Res. **21:** 1133–1139.

CHAKRAVARTI, A., 1991 A graphical representation of genetic and physical maps: the Marey map. Genomics **11:** 219–222.

CHARLESWORTH, B., C. H. LANGLEY and W. STEPHAN, 1986 The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics **112:** 947–62.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

CHARLESWORTH, B., P. SNIEGOWSKI and W. STEPHAN, 1994 The evo-

lutionary dynamics of repetitive DNA in eukaryotes. Nature **371:** 215–220.

CHUANG, P.-T., D. G. ALBERTSON and B. J. MEYER, 1994 DPY-27: a chromosome condensation homolog that regulates *C. elegans* dosage compensation through association with the *X* chromosome. Cell **79:** 459–474.

CIVARDI, L., Y. XIA, K. J. EDWARDS, P. S. SCHNABLE and B. J. NIKOLAU, 1994 The relationship between genetic and physical distances in the cloned *a1-sh2* interval of the *Zea mays* L. genome. Proc. Natl. Acad. Sci. USA **91:** 8268–8272.

COULSON, A., J. SULSTON, S. BRENNER and J. KARN, 1986 Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **83:** 7821–7825.

COULSON, A., R. WATERSTON, J. KIFF, J. SULSTON and Y. KOHARA, 1988 Genome linking with yeast artificial chromosomes. Nature **335:** 184–186.

COULSON, A., Y. KOZONO, B. LUTTERBACH, R. SHOWNKEEN, J. SULSTON *et al.*, 1991 YACs and the *C. elegans* genome. Bioessays **13:** 413–417.

DE LANGE, T., 1992 Human telomeres are attached to the nuclear matrix. EMBO J. **11:** 717–724.

DUJON, B., D. ALEXANDRAKI, B. ANDRE, W. ANSORGE, V. BALADRON *et al.*, 1994 Complete DNA sequence of yeast chromosome XI. Nature **369:** 371–378.

EDGLEY, M. L., and D. L. RIDDLE, 1993 The nematode *Caenorhabditis elegans*, pp. 3.281–3.318 in *Genetic Maps: Locus Maps of Complex Genomes, Lower Eukaryotes*, edited by S. J. O'BRIEN. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

EMMONS, S. W., B. ROSENZWEIG and D. HIRSH, 1980 Arrangement of repeated sequences in the DNA of the nematode *Caenorhabditis elegans*. J. Mol. Biol. **144:** 481–500.

FELSENSTEIN, K. M., and S. W. EMMONS, 1988 Nematode repetitive DNA with *ARS* and segregation function in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8:** 875–883.

GIBSON, T. J., A. ROSENTHAL and R. H. WATERSTON, 1987 Lorist6, a cosmid vector with *Bam* HI, *Not* I, *Sca* I and *Hind* III cloning sites and altered neomycin phosphotransferase gene expression. Gene **53:** 283–286.

GREENWALD, I., A. COULSON, J. SULSTON and J. PRIESS, 1987 Correlation of the physical and genetic maps in the *lin-12* region of *Caenorhabditis elegans*. Nucleic Acids Res. **15:** 2295–2307.

HERMAN, R. K., and C. K. KARI, 1989 Recombination between small *X* chromosome duplications and the *X* chromosome in *Caenorhabditis elegans*. Genetics **121:** 723–737.

HERMAN, R. K., C. K. KARI and P. S. HARTMAN, 1982 Dominant *X*-chromosome nondisjunction mutants of *Caenorhabditis elegans*. Genetics **102:** 379–400.

HODGKIN, J., 1983 *X* chromosome dosage and gene expression in *C. elegans*: two unusual dumpy genes. Mol. Gen. Genet. **192:** 452–458.

HODGKIN, J., H. R. HORVITZ and S. BRENNER, 1979 Nondisjunction mutants of the nematode *Caenorhabditis elegans*. Genetics **91:** 67–94.

IKEMURA, T., and K. WADA, 1991 Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers: relation between nucleotide sequence data and cytogenetic data. Nucleic Acids Res. **19:** 4333–4339.

JOHNSTON, M., S. ANDREWS, R. BRINKMAN, J. COOPER, H. DING *et al.*, 1994 Complete nucleotide sequence of Saccharomyces cerevisiae chromosome VIII. Science **265:** 2077–2082.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

KENNEDY, B. P., E. J. AAMODT, F. L. ALLEN, M. A. CHUNG, M. F. HESCHL *et al.*, 1993 The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J. Mol. Biol. **229:** 890–908.

LA VOLPE, A., M. CIARAMELLA and P. BAZZICALUPO, 1988 Structure, evolution and properties of a novel repetitive DNA family in *Caenorhabditis elegans*. Nucleic Acids Res. **16:** 8213–8231.

MADL, J. E., and R. K. HERMAN, 1979 Polyploids and sex determination in *Caenorhabditis elegans*. Genetics **93:** 393–402.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

MCKIM, K. S., A. M. HOWELL and A. M. ROSE, 1988 The effects of translocations on recombination frequency in *Caenorhabditis elegans*. Genetics **120:** 987–1001.

MCKIM, K. S., K. PETERS and A. M. ROSE, 1993 Two types of sites required for meiotic chromosome pairing in *Caenorhabditis elegans*. Genetics **134:** 749–768.

MEYER, B. J., and L. P. CASSON, 1986 *Caenorhabditis elegans* compensates for the difference in *X* chromosome dosage between the sexes by regulating transcript levels. Cell **47:** 871–881.

MOUCHIROUD, D., G. D'ONOFRIO, B. AISSANI, G. MACAYA, C. GOUTIER *et al.*, 1991 The distribution of genes in the human genome. Gene **100:** 181–187.

NACLERIO, G., G. CANGIANO, A. COULSON, A. LEVITT, V. RUVOLO *et al.*, 1992 Molecular and genomic organization of clusters of repetitive DNA sequences in *Caenorhabditis elegans*. J. Mol. Biol. **226:** 159–168.

OLIVER, S. G., Q. J. VAN DER AART, M. L. AGOSTONI-CARBONE, M. AIGLE, L. ALBERGHINA *et al.*, 1992 The complete DNA sequence of yeast chromosome III. Nature **357:** 38–46.

PILGRIM, D., 1993 The genetic and RFLP characterization of the left end of linkage group *III* in *Caenorhabditis elegans*. Genome **36:** 712–724.

PRASAD, S. S., and D. L. BAILLIE, 1989 Evolutionarily conserved coding sequences in the *dpy-20-unc-22* region of *Caenorhabditis elegans*. Genomics **5:** 185–198.

ROSENBLUTH, R. E., and D. L. BAILLIE, 1981 The genetic analysis of a reciprocal translocation, *eT1(III; V)*, in *Caenorhabditis elegans*. Genetics **99:** 415–428.

SIGURDSON, D. C., R. K. HERMAN, C. A. HORTON, C. K. KARI and S. E. PRATT, 1986 An *X*-autosome fusion chromosome of *Caenorhabditis elegans*. Mol. Gen. Genet. **202:** 212–218.

STARR, T., A. M. HOWELL, J. MCDOWALL, K. PETERS and A. M. ROSE, 1989 Isolation and mapping of DNA probes within the linkage group *I* gene cluster of *Caenorhabditis elegans*. Genome **32:** 365–372.

STEPHAN, W., 1986 Recombination and the evolution of satellite DNA. Genet. Res. **47:** 167–74.

SULSTON, J. E., and S. BRENNER, 1974 The DNA of *Caenorhabditis elegans*. Genetics **77:** 95–104.

SULSTON, J., Z. DU, K. THOMAS, R. WILSON, L. HILLIER *et al.*, 1992 The *C. elegans* genome sequencing project: a beginning. Nature **356:** 37–41.

THON, G., A. COHEN and A. J. KLAR, 1994 Three additional linkage groups that repress transcription and meiotic recombination in the mating-type region of *Schizosaccharomyces pombe*. Genetics **138:** 29–38.

WATERSTON, R., C. MARTIN, M. CRAXTON, C. HUYNH, A. COULSON *et al.*, 1992 A survey of expressed genes in *Caenorhabditis elegans*. Nature Genet. **1:** 114–123.

WILSON, R., R. AINSCOUGH, K. ANDERSON, C. BAYNES, M. BERKS *et al.*, 1994 2.2 Mb of contiguous nucleotide sequence from chromosome *III* of *C. elegans*. Nature **368:** 32–38.

WU, T. C., and M. LICHTEN, 1994 Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science **263:** 515–518.

ZETKA, M. C., and A. M. ROSE, 1990 Sex-related differences in crossing over in *Caenorhabditis elegans*. Genetics **126:** 355–63.

Communicating editor: R. K. HERMAN

APPENDIX

Table A1 contains the data used in plotting the Marey maps. The loci are listed in canonical genetic left-to-right order, from *I* to *X*. For the terminal loci of each linkage group, the suffixes :L end or :R end are used to indicate that these are the most extreme known loci, and have been arbitrarily equated with the ends of the corresponding contigs. Where the last known locus is included as a real point (*e.g.*, *stP100* at the left end of linkage group *II*), the most extreme locus is given the dummy name L end or R end, a genetic coordinate identical to the last locus, and a physical coordinate equal to the end of the corresponding contig. L end and R end within a list for a linkage group refer to the interpolated positions of

TABLE A1

Marey map data for Figure 1

LG	Locus	Genetic map position	Genetic coordinate (cM)	Contig	Physical map position (bands)	Physical coordinate (kb)
I	*spe-13*:L end	−21.20	0.00	ctg465	−4494	0
	sup-34	−19.20	2.00		−4323	313
	nP59	−6.20	15.00		−3696	1460
	unc-11[a]	−2.20	19.00		−3385	2029
	unc-73	−1.20	20.00		−3289	2205
	unc-38	−0.50	20.70		−2790	3118
	mes-3	−0.36	20.84		−2737	3215
	spe-11	−0.13	21.07		−2527	3600
	dpy-5	0.00	21.20		−2427	3783
	unc-40	0.20	21.40		−2321	3977
	bli-4	0.80	22.00		−1993	4577
	smg-5	1.30	22.50		−1775	4976
	unc-87	1.37	22.57		−1748	5025
	dpy-14	1.42	22.62		−1717	5082
	smg-1	1.53	22.73		−1656	5194
	eat-5	1.79	22.99		−1402	5658
	unc-15	1.88	23.08		−1359	5737
	unc-13	1.92	23.12		−1325	5799
	gld-1	2.17	23.37		−1170	6083
	mei-1	2.67	23.87		−806	6749
	unc-120	2.82	24.02		−717	6912
	fer-1	3.02	24.22		−646	7042
	sup-17	3.12	24.32		−570	7181
	unc-29	3.22	24.42		−506	7298
	mel-26	3.62	24.82		−390	7510
	lin-11	4.82	26.02		269	8716
	R end	6.62	27.82		548	9227
	L end	7.22	28.42	ctg714	−508	9410
	unc-101	13.32	34.52		451	11165
	hP4	18.82	40.02		895	11977
	R end	21.22	42.42		975	12124
	L end	24.22	45.42	ctg1	−154	12307
	ZK1209	24.32	45.52		−150	12314
	unc-54	26.32	47.52		−21	12550
	ces-2	28.32	49.52		36	12654
	eDf24:R end	29.32	50.52		436	13386
II	L end	−24.82	0.00	ctg369	−6160	0
	stP100	−24.82	0.00		−6104	102
	stP196	−10.32	14.50		−4517	3007
	lin-31	−4.82	20.00		−3642	4608
	lin-4	−0.32	24.50		−2469	6755
	lin-23	−0.17	24.65		−2202	7243
	mua-1	−0.02	24.80		−2060	7503
	dpy-10	0.00	24.82		−2014	7587
	tra-2	0.10	24.92		−1901	7794
	unc-104	0.15	24.97		−1880	7832
	zyg-11	0.65	25.47		−1305	8885
	rol-6	1.05	25.87		−928	9575
	let-23	1.28	26.10		−649	10085
	unc-4	1.73	26.55		−254	10808
	egl-43	1.83	26.65		−227	10857
	sqt-1	3.23	28.05		713	12578
	lin-29	4.43	29.25		1040	13176
	R end	12.03	36.85		1493	14005
	L end	13.73	38.55	ctg743	−274	14188

TABLE A1

Continued

LG	Locus	Genetic map position	Genetic coordinate (cM)	Contig	Physical map position (bands)	Physical coordinate (kb)
II	*lin-7*[a]	24.43	49.25		366	15359
	unc-52	24.63	49.45		530	15659
	mog-4:R end	25.73	50.55		782	16120
III	*vab-6*:L end	−27.77	0.00	ctg325	−498	0
	unc-45	−27.67	0.10		−72	780
	eP93	−27.31	0.46		−15	884
	eP64	−26.95	0.82		16	941
	eP91	−26.18	1.59		36	977
	eP92	−26.07	1.70		52	1006
	daf-7	−26.07	1.70		66	1032
	R end	−25.32	2.45		176	1233
	L end	−24.62	3.15	ctg377	−3775	1416
	par-2	−24.57	3.20		−3767	1431
	mec-12	−12.57	15.20		−3215	2441
	ben-1	−6.57	21.20		−3003	2829
	unc-93	−5.37	22.40		−2965	2899
	dpy-27	−4.87	22.90		−2865	3082
	mpk-1	−3.97	23.80		−2599	3568
	pal-1	−2.22	25.55		−2245	4216
	daf-4	−1.52	26.25		−1805	5022
	par-3	−1.47	26.30		−1768	5089
	sma-4	−1.32	26.45		−1703	5208
	mec-14	−0.83	26.94		−997	6500
	lin-13	−0.53	27.24		−689	7064
	mab-5	−0.48	27.29		−660	7117
	egl-5	−0.45	27.32		−639	7155
	lin-36	−0.33	27.44		−534	7347
	unc-36	−0.23	27.54		−439	7521
	unc-86	−0.22	27.55		−420	7556
	mig-10	−0.20	27.57		−373	7642
	unc-116	−0.18	27.59		−345	7693
	sup-5	−0.08	27.69		−162	8028
	lin-9	−0.01	27.76		−28	8273
	unc-32	0.00	27.77		−27	8275
	eP6	0.05	27.82		−3	8319
	lin-12	0.09	27.86		54	8423
	glp-1	0.11	27.88		68	8449
	ced-7	0.61	28.38		318	8907
	unc-69	2.41	30.18		745	9688
	ced-9	2.51	30.28		762	9719
	vab-7[a]	5.21	32.98		1079	10299
	pha-1	5.91	33.68		1160	10447
	tra-1	6.61	34.38		1191	10504
	R end	8.56	36.33		1217	10552
	L end	16.06	43.83	ctg474	1199	10735
	pie-1	17.11	44.88		1213	10760
	unc-25[a]	20.11	47.88		1355	11020
	unc-64:R end	22.11	49.88		1797	11829
IV	*daf-18*:L end	−29.71	0.00	ctg423	−4753	0
	daf-1	−27.71	2.00		−4020	1341
	lin-1	−7.71	22.00		−3637	2042
	lin-22	−6.21	23.50		−3498	2297
	unc-33	−3.56	26.15		−3171	2895
	cha-1	−3.21	26.50		−3147	2939
	unc-17	−3.20	26.51		−3143	2946

TABLE A1

Continued

LG	Locus	Genetic map position	Genetic coordinate (cM)	Contig	Physical map position (bands)	Physical coordinate (kb)
IV	*sup-29*	−2.70	27.01		−3048	3120
	dpy-13	0.00	29.71		−2748	3669
	ama-1	0.07	29.78		−2733	3697
	unc-5	2.07	31.78		−2023	4996
	skn-1	2.22	31.93		−1937	5153
	unc-44	2.52	32.23		−1766	5466
	lin-45	3.02	32.73		−1526	5905
	col-4	3.42	33.13		−1007	6855
	unc-24	3.52	33.23		−860	7124
	fem-3	4.22	33.93		−227	8283
	mec-3	4.82	34.53		524	9657
	lin-3	5.12	34.83		870	10290
	dpy-20	5.52	35.23		1235	10958
	par-5	5.58	35.29		1291	11061
	let-56	5.74	35.45		1342	11154
	unc-22	5.78	35.49		1362	11190
	unc-31	6.87	36.58		1841	12067
	unc-30	7.87	37.58		2083	12510
	lev-1	8.28	37.99		2113	12565
	tra-3	12.28	41.99		2579	13418
	sup-24	13.57	43.28		2873	13956
	eP69	16.57	46.28		3437	14988
	R end	17.97	47.68		3762	15582
	L end	18.42	48.13	ctg198	−81	15765
	hsp-1	18.57	48.28		−48	15826
	R end	18.57	48.28		70	16042
V	*let-450*:L end	−21.50	0.00	ctg313	−4695	0
	nP60	−21.00	0.50		−4088	1111
	nP61	−20.80	0.70		−4057	1168
	nP62	−20.00	1.50		−3903	1449
	unc-60	−19.30	2.20		−3767	1698
	lag-2	−11.30	10.20		−2798	3472
	apx-1	−10.00	11.50		−2765	3532
	stP3	−7.00	14.50		−2011	4912
	stP192	0.00	21.50		−697	7316
	unc-68	0.35	21.85		−460	7750
	odr-2	0.60	22.10		−342	7966
	stP23	0.90	22.40		15	8619
	her-1	2.25	23.75		1153	10702
	act-123	3.02	24.52		1858	11992
	osm-6	3.62	25.12		2336	12867
	dpy-30	3.84	25.34		2464	13101
	myo-3	3.92	25.42		2494	13156
	sqt-3	3.99	25.49		2559	13275
	egl-10	4.18	25.68		2633	13410
	arP3	5.08	26.58		3158	14371
	rhP[EH#6]	5.14	26.64		3197	14442
	him-5	6.03	27.53		3513	15021
	sdc-3	7.83	29.33		4098	16091
	unc-76	8.03	29.53		4135	16159
	rrs-1	14.03	35.53		5372	18423
	stP108	15.63	37.13		5525	18703
	stP105	18.63	40.13		5975	19526
	R end	23.13	44.63		6274	20073
	L end	24.63	46.13	ctg893	−473	20256
	stP128	25.03	46.53		−446	20306

TABLE A1

Continued

LG	Locus	Genetic map position	Genetic coordinate (cM)	Contig	Physical map position (bands)	Physical coordinate (kb)
V	*unc-51*	27.33	48.83		73	21255
	rol-9:R end	28.53	50.03		97	21299
X	L end	−25.40	0.00	ctg674	−2528	0
	egl-17	−25.40	0.00		−2434	172
	sli-1	−25.00	0.40		−2297	423
	stP41	−24.90	0.50		−2162	670
	meP3	−24.25	1.15		−1974	1014
	meP2	−23.85	1.55		−1836	1266
	lin-32	−20.95	4.45		−1326	2200
	unc-2[a]	−17.85	7.55		−1080	2650
	mgP40	−12.05	13.35		−465	3775
	mgP39	−11.65	13.75		−401	3892
	lin-18	−10.55	14.85		−273	4127
	mec-2	−6.20	19.20		455	5459
	unc-6	−3.20	22.20		1168	6764
	dpy-7	−1.70	23.70		1509	7388
	unc-18	−1.40	24.00		1558	7477
	deg-1	−1.20	24.20		1600	7554
	sem-5	−0.80	24.60		1676	7693
	xol-1	−0.70	24.70		1715	7765
	sup-21	0.80	26.20		2178	8612
	vab-3	2.55	27.95		3134	10361
	daf-12	2.85	28.25		3236	10548
	egl-15	3.45	28.85		3480	10995
	lin-14	4.95	30.35		3777	11538
	sdc-2	5.15	30.55		3810	11599
	mes-1[a]	5.65	31.05		3967	11846
	lin-2	7.15	32.55		4252	12407
	unc-3	17.15	42.55		5683	15026
	unc-7	19.15	44.55		5935	15487
	lin-15	21.15	46.55		6285	16128
	mec-5	21.35	46.75		6409	16355
	sdc-1	22.45	47.85		6620	16741
	let-2	22.65	48.05		6691	16871
	osm-1	23.10	48.50		6798	17067
	mec-4	23.20	48.60		6925	17299
	sup-10	23.25	48.65		7312	18007
	let-6:R end	23.27	48.67		7567	18474

[a] The physical location of these loci are not in data release 2–14 of ACEDB and were provided by A. ALFONSO (*unc-11*); J. SIMSKE and S. KIM (*lin-7*); J. AHRINGER (*vab-7*); Y. JIN and H. HORVITZ (*unc-25*); W. SCHAFER and C. KENYON (*unc-2*); L. BERKOWITZ and S. STROME (*mes-1*).

the ends of the contigs, under the assumptions given in MATERIALS AND METHODS. The column "Genetic map position" gives the genetic map coordinates as rederived by us, expressed in the standard coordinate system for each *C. elegans* linkage group. Details of these rederivations are available on request. The column "Physical map position" gives the position in band coordinates of the midpoint of the corresponding molecular clone from the physical map (each contig has an independent coordinate system). The columns "Genetic coordinate" and "Physical coordinate" give the coordinates used for the Marey maps themselves (see MATERIALS AND METHODS for details). The column "Contig" gives the name of each genetically positioned contig.