Transposons but Not Retrotransposons Are Located Preferentially in Regions of High Recombination Rate in *Caenorhabditis elegans*

Laurent Duret, Gabriel Marais and Christian Biémont

Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, 69622 Villeurbanne Cedex, France

Manuscript received June 19, 2000
Accepted for publication September 5, 2000

ABSTRACT

We analyzed the distribution of transposable elements (TEs: transposons, LTR retrotransposons, and non-LTR retrotransposons) in the chromosomes of the nematode *Caenorhabditis elegans*. The density of transposons (DNA-based elements) along the chromosomes was found to be positively correlated with recombination rate, but this relationship was not observed for LTR or non-LTR retrotransposons (RNA-based elements). Gene (coding region) density is higher in regions of low recombination rate. However, the lower TE density in these regions is not due to the counterselection of TE insertions within exons since the same positive correlation between TE density and recombination rate was found in noncoding regions (both in introns and intergenic DNA). These data are not compatible with a global model of selection acting against TE insertions, for which an accumulation of elements in regions of reduced recombination is expected. We also found no evidence for a stronger selection against TE insertions on the X chromosome compared to the autosomes. The difference in distribution of the DNA and RNA-based elements along the chromosomes in relation to recombination rate can be explained by differences in the transposition processes.

TRANSPOSABLE elements (TEs) have a major influence on genome evolution. More than simple parasitic elements, they now are more and more considered as genome restructuring agents that provide genome flexibility and variability for population adaptation (Shapiro 1999). Their population dynamics are, however, far from being understood, and the forces that account for their distribution throughout the genome and maintain them in populations are still a matter of large debate (Biémont *et al.* 1997; Charlesworth *et al.* 1997). It has been proposed that chromosomal rearrangements caused by TEs through recombinational processes at nonhomologous sites may explain the differential accumulation of TEs and other repetitive sequences in genomic regions where recombination is infrequent, such as the heterochromatic regions and the Y chromosomes in various species (Charlesworth *et al.* 1994). If it is assumed that the frequency of ectopic exchanges in a region is proportional to meiotic exchanges in that region (Langley *et al.* 1988; Goldman and Lichten 1996), then TE insertion number should be negatively correlated with recombination rate. Moreover, population genetics models predict a positive correlation between the efficacy of selection at a given locus and the local rate of recombination because of Hill-Robertson effects (Hill and Robertson 1966; Maynard-Smith and Haigh 1974; Charlesworth *et al.* 1993). Therefore, the counterselection against the deleterious effects of TE insertions should be stronger in regions of high recombination rate. Both models predict a negative correlation between TE density and recombination rate along chromosomes. No such relationship with frequency of recombination was observed, however, in *Drosophila melanogaster* for TE insertions (Hoogland and Biémont 1996) or in the nematode *Caenorhabditis elegans* for repetitive sequences (Naclerio *et al.* 1992; Barnes *et al.* 1995). Rather, in the latter species a positive relationship with the CeRep3 repeated element distribution was reported (Barnes *et al.* 1995). Since we now possess information on >95% of the *C. elegans* genome (*C. elegans* Sequencing Consortium 1998), a new estimation of the relationship between recombination rate and TE distribution is feasible.

C. elegans is a good model for such an analysis because the recombination rate varies remarkably along its autosomes: each autosome has a central region of low recombination rate (0.7 cM/Mb on average) flanked by two arms of high recombination rate (4.7 cM/Mb on average; Barnes *et al.* 1995). Whereas central regions correspond to 41% of the autosome DNA, 91% of meiotic recombination occurs in the arms. Moreover, gene density is slightly higher in the central portions of the autosomes (30% coding) than in the arms (23% coding; *C. elegans* Sequencing Consortium 1998). Hence, contrary to other organisms, most recombinational exchange in *C. elegans* occurs in relatively gene-poor DNA. Recombination rate is fairly uniform along the X chro-

Corresponding author: Laurent Duret, Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Université Lyon 1, 69622 Villeurbanne Cedex, France. E-mail: duret@biomserv.univ-lyon1.fr

mosome (2.6 cM/Mb on average), much higher than in autosomal central regions, but gene density is relatively low (20% coding), similar to the arms (*C. elegans* Sequencing Consortium 1998). Using available genomic sequences, we searched the location of transposable elements (transposons, LTR, and non-LTR retrotransposons) in the chromosomes of *C. elegans* strain N2 and analyzed their distribution according to recombination rates. We show that the amount of transposons, but not of LTR and non-LTR retrotransposons, is positively correlated with recombination frequency. This indicates that selection against the insertional effects of TEs, or against the dominant deleterious effect of chromosomal rearrangements due to recombination between TE insertions, is not the main factor explaining the dynamics of TEs in this species. These selectionist hypotheses indeed imply a negative relationship between recombination rate and amount of TE insertions. A simple hypothesis based on preferential insertions in regions of high recombination may account for the distribution of transposons in the *C. elegans* genome.

MATERIALS AND METHODS

Sequence data: Full-length sequences of the six *C. elegans* chromosomes along with gene annotations were retrieved from the Genome division of GenBank (Benson *et al.* 1999) release 111 (April 15, 1999). Chromosome regions that have not been yet sequenced are represented by tracks of N corresponding to the estimated gap size. Data available in GenBank at that time (without N) totaled 94.5 Mb corresponding to 95% of the estimated whole genome sequence (*C. elegans* Sequencing Consortium 1998).

Estimation of recombination rate: To analyze the rate of recombination along the *C. elegans* chromosomes we used a procedure similar to the one described by Kliman and Hey (1993). The *C. elegans* genetic map data were taken from ACEDB release WS6 (December 1998; R. Durbin and J. Thierry-Mieg, unpublished results). We selected the 225 loci that had been localized both in the genetic map and in the genomic sequence. The polynomial curves as functions of the genetic distance *vs.* the nucleotide coordinate in the genomic sequence were obtained for each chromosome ($R^2 = 0.97$ in chromosome IV; $R^2 \geq 0.99$ in all other chromosomes). Recombination rate, as a function of nucleotide position along a chromosome, was estimated by taking the derivative of the polynomial function for each chromosome. We defined three classes of recombination rate: low (<1 cM/Mb), medium (1–5 cM/Mb), and high (>5 cM/Mb).

Localization of transposable elements: We collected from the literature the sequences of 25 transposable elements identified in *C. elegans* (Table 1). Chromosome sequences were split into 100-kb fragments. Fragments containing >50% of nondetermined sequence (N) were excluded. The remaining 978 fragments were analyzed for their amount and distribution of the 25 TEs, using the program RepeatMasker (A. F. A. Smit and P. Green, unpublished data; RepeatMasker is available at http://repeatmasker.genome.washington.edu/cgi-bin/RM2_req.pl). We computed the density (number of elements per megabase) of each TE in these genomic fragments overall and then separately for introns, coding regions, and intergenic regions. We defined as intergenic all sequences located between protein-coding regions annotated in the GenBank database. Some sequences considered here as intergenic could thus in fact contain nonprotein coding genes (tRNA, rRNA, etc.) or some unidentified (unannotated) protein coding genes. Simple repeats and low complexity regions (regions of biased base composition) were identified with RepeatMasker.

Statistical test: The repartition of TEs in different classes of genomic regions (regions of high compared to low recombination rate, X compared to autosomes) was tested by χ^2. The observed number of copies in each class was compared to the expected number, assuming that the total number of copies found in both classes was distributed according to the total amount of DNA in each class.

RESULTS

Among the 25 transposable elements retrieved from the *C. elegans* genome are 12 transposons (DNA-based elements), 1 LTR retrotransposon, and 12 non-LTR retrotransposons (Table 1). Overall, we recorded 3718 copies (complete or not) of these TEs. Note that sequences presently available represent ~95% of the complete genome. It is likely that sequence sampling is not random and one might expect that TEs are overabundant in the 5% of missing sequences. It is, however, unlikely that with 95% of coverage, such a sampling bias could affect significantly the results of our analyses. The degree of identity between the different copies and the reference sequence of each TE family was 84% on average. The number of copies detected for each family appeared higher than previous estimates on the basis of experimental approaches using DNA hybridization. This is probably because the hybridization technique is less sensitive than direct sequence comparison to identify truncated copies or distant members of a family. Indeed, 88% of the copies we detected had large deletions (>20% of the full-length elements). The copy-number estimates based on DNA hybridization are, however, in good agreement with the number of complete (or >80% complete) copies (Table 2).

C. elegans chromosome sequences were split into fragments of 100 kb, and these fragments were classified into three groups according to their recombination rate. The limits between these three classes were set to match approximately the average rate in the arms and central regions. Recombination rates >5 cM/Mb are thus hereafter considered high and recombination rates <1 cM/Mb are considered low. Chromosome fragments of high and low recombination rate account for 17 and 27% of the whole data set, respectively.

Density of transposable elements according to recombination rate: The overall density of transposons increased almost threefold with recombination rate: from 19.1 copies/Mb on average in fragments of low recombination rate to 55.4 copies/Mb in fragments of high recombination rate (Figure 1). This property seemed to be shared by most transposons, independently of the number of copies: in 6 out of 12 transposons, the density in regions of high recombination rate was significantly higher than that in regions of low recombination rate,

TABLE 1

***C. elegans* transposable elements analyzed**

Element	Type[a]	Reference[b]	Accession no.	Position
IR-1	Tpn	a	U86946	1 . . . 379
IR-2	Tpn	a	U86947	1 . . . 781
IR-3	Tpn	a	U86948	1 . . . 578
IR-4	Tpn	a	U86949	1 . . . 227
IR-5	Tpn	a	U86950	1 . . . 198
Tc1	Tpn	b	K01135	1 . . . 1761
Tc2	Tpn	c	X59156	1 . . . 2074
Tc3	Tpn	d	M77697	14869 . . . 15906
Tc4	Tpn	e	L00665	1 . . . 3483
Tc5	Tpn	f	Z35400	1 . . . 3171
Tc6	Tpn	g	L19187	1 . . . 2716
Tc7	Tpn	h	Z37140	29218 . . . 30083
Cer1	LTR	i	U15406	1 . . . 8865
Rte-1	RTpn	j	AF054983	1 . . . 3291
Frodo-1	RTpn	k	Z70755	20784 . . . 23780
Frodo-2	RTpn	k	Z48009	21408 . . . 24687
Sam1	RTpn	k	U13643	19600 . . . 22449
Sam2	RTpn	k	U57054	17169 . . . 20000
Sam3	RTpn	k	U46668	18500 . . . 21336
Sam4	RTpn	k	Z92972	13825 . . . 17262
Sam5	RTpn	k	Z81092	1125 . . . 4800
Sam6	RTpn	k	Z82275	1 . . . 3364
Sam7	RTpn	k	Z82090	7625 . . . 10613
Sam8	RTpn	k	AF016663	12000 . . . 15060
Sam9	RTpn	k	Z81064	7400 . . . 10100

[a] Tpn, transposon; LTR, LTR retrotransposon; RTpn, non-LTR retrotransposon.
[b] (a) Devine *et al.* (1997); (b) Rosenzweig *et al.* (1983); (c) Ruvolo *et al.* (1992); (d) Collins *et al.* (1989); (e) Li and Shaw (1993); (f) Collins and Anderson (1994); (g) Dreyfus and Emmons (1991); (h) Rezsohazy *et al.* (1997); (i) Britten (1995); (j) Youngman *et al.* (1996); (k) Marin *et al.* (1998).

and 3 other transposons showed the same trend (Table 2). The most striking example is the IR-2 element whose density increased 13 times between classes of low and high recombination rates. For the 5 other transposons that showed a statistically significant difference, the increase in density ranged from two to six times. In the only case where transposon density was found lower in regions of high recombination rate (IR-1), the difference was not statistically significant (Table 2). Most of the copies detected were truncated, suggesting that their insertion was probably relatively ancient (the average divergence compared to the reference sequence is 16%). The 419 transposon copies that were at least 80% complete are less divergent (10% in average) and were probably inserted more recently. These copies showed the same pattern of insertion, with an almost fourfold excess in regions of high compared to low recombination rate (respectively 9.0 and 2.4 copies/Mb). Thus, the same pattern was observed with both ancient and recent insertions.

Only 1 out of the 13 retrotransposons (LTR and non-LTR retrotransposons) showed significant variation in density with recombination rate (Table 2). Since the density of retrotransposons was relatively low (~7 copies/Mb on average), it is possible that the lack of statistical significance was due to the small sample size of each family. However, the overall density of all the retrotransposons did not vary with recombination rate (Figure 1).

Density of transposons in noncoding regions according to recombination rate: In *C. elegans*, gene density decreases with increasing recombination rate: from 28% of coding sequences in regions of low recombination rate to 17% in regions of high recombination rate (Table 3). To test whether this variation in gene density could interfere with the relationship between recombination rate and transposon density, we measured the density of transposons among noncoding regions. Around 98% of the TE copies identified were found in noncoding regions (introns and intergenic regions). We found that the number of transposons per megabase in these noncoding regions increased almost threefold between regions of low and high recombination rate (Figure 2).

Other genomic features linked to recombination rate: Several other genomic features were also analyzed according to recombination rate. In agreement with previous results (Barnes *et al.* 1995), the density of the CeRep3 repetitive element was found positively correlated with recombination rate, and this was observed

TABLE 2

Distribution of transposable elements in regions of low and high recombination rate

	Observed copy number				Density (copy number/Mb)			
		Recombination rate				Recombination rate		
	Whole genome (94.5 Mb)	Low (25.3 Mb)	High (16.1 Mb)	χ^2 (high/low recombination rate)	Whole genome	Low	High	Ratio (high/low)
IR-1	127 (11)	38 (6)	17 (1)	1.48 NS	1.34	1.50	1.06	0.70
IR-2	184 (114)	12 (4)	99 (71)	118.12 **	1.95	0.47	6.15	12.96
IR-3	1204 (103)	108 (7)	409 (25)	351.72 **	12.74	4.27	25.41	5.95
IR-4	81 (41)	11 (9)	25 (10)	14.13 **	0.86	0.44	1.55	3.57
IR-5	87 (38)	22 (10)	24 (13)	3.41 NS	0.92	0.87	1.49	1.71
Tc1	131 (30)	31 (5)	24 (3)	0.52 NS	1.39	1.23	1.49	1.22
Tc2	294 (5)	75 (0)	71 (5)	5.82 *	3.11	2.97	4.41	1.49
Tc3	61 (24)	11 (5)	8 (5)	0.08 NS	0.65	0.44	0.50	1.14
Tc4	233 (5)	66 (1)	43 (3)	0.01 NS	2.47	2.61	2.67	1.02
Tc5	159 (2)	44 (1)	27 (0)	0.02 NS	1.68	1.74	1.68	0.96
Tc6	269 (14)	35 (6)	65 (1)	28.66 **	2.85	1.38	4.04	2.92
Tc7	253 (32)	24 (6)	72 (8)	52.64 **	2.68	0.95	4.47	4.71
Total transposons	3083 (419)	477 (60)	887 (145)	391.85 **	32.63	18.87	55.11	2.92
Cer1	24 (1)	3 (1)	2 (0)	0.00 NS	0.25	0.12	0.12	1.05
Rte-1	98 (10)	31 (4)	22 (3)	0.15 NS	1.04	1.23	1.37	1.11
Frodo-1	92 (5)	23 (1)	24 (2)	2.93 NS	0.97	0.91	1.49	1.64
Frodo-2	71 (1)	27 (1)	15 (0)	0.18 NS	0.75	1.07	0.93	0.87
Sam1	43 (1)	12 (1)	4 (0)	1.30 NS	0.46	0.47	0.25	0.52
Sam2	27 (2)	4 (0)	4 (0)	0.41 NS	0.29	0.16	0.25	1.57
Sam3	34 (4)	9 (0)	5 (2)	0.06 NS	0.36	0.36	0.31	0.87
Sam4	16 (1)	4 (0)	3 (0)	0.05 NS	0.17	0.16	0.19	1.18
Sam5	22 (2)	6 (1)	3 (0)	0.12 NS	0.23	0.24	0.19	0.79
Sam6	12 (2)	4 (0)	1 (0)	0.75 NS	0.13	0.16	0.06	0.39
Sam7	53 (2)	21 (1)	13 (0)	0.01 NS	0.56	0.83	0.81	0.97
Sam8	60 (6)	15 (1)	7 (0)	0.46 NS	0.64	0.59	0.43	0.73
Sam9	83 (1)	11 (0)	21 (0)	9.62 **	0.88	0.44	1.30	3.00
Total retrotransposons	635 (38)	170 (11)	124 (7)	1.33 NS	6.72	6.72	7.70	1.15

Numbers of full-length or nearly full-length (>80%) copies are indicated in parentheses. NS, nonsignificant. *$P < 0.05$; **$P < 0.005$.

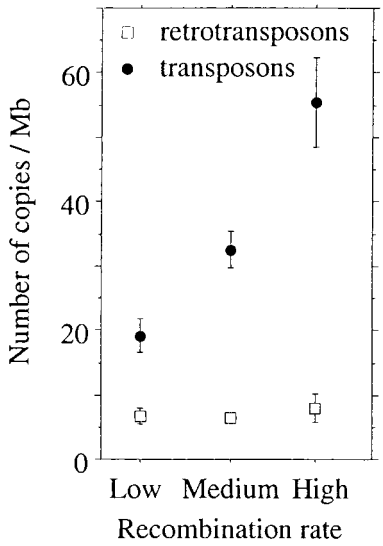

FIGURE 1.—Density of transposable elements according to recombination rate. Low recombination rate, <1 cM/Mb; high recombination rate, >5 cM/Mb. Error bars indicate the 95% confidence interval.

for all CeRep-like sequences (data not shown). The frequency of simple repeats, such as microsatellites, low complexity regions (regions of biased base composition), and the G + C content also increased with recombination rate (Figure 3). Although the difference in G + C content was statistically highly significant, the variation was limited from 35 to 36%. This low variation in G + C content probably explains why it had not been noted previously (BARNES *et al.* 1995). The major mutational mechanism responsible for the evolution of microsatellites is replication slippage. Therefore, in contrast with satellite DNA that evolves essentially by unequal crossing over, the evolution of microsatellites is not expected *a priori* to depend on the recombination process (STEPHAN and CHO 1994). The relationship between recombination rate and microsatellite density found in *C. elegans* (Figure 3a) does not seem to be a general rule since such a relationship has not been observed in *D. melanogaster* (BACHTROG *et al.* 1999) and in humans (DIB *et al.* 1996).

DISCUSSION

The genome of *C. elegans* consists of five autosomes and an X chromosome. The autosomes have a high density of genes in their central region (clusters), which presents a low frequency of recombination, while low gene density and high frequency of recombination characterize the arms (noncoding DNA-rich regions). The X chromosome has no cluster. Hence, contrary to other organisms, exchange in *C. elegans* occurs preferentially in gene-poor DNA regions. Some articles have addressed the question of how repetitive sequences are distributed in relation to regions of high and low frequency of recombination (NACLERIO *et al.* 1992; BARNES *et al.* 1995; *C. ELEGANS* SEQUENCING CONSORTIUM 1998). In the first article, the authors analyzed five families of repetitive DNA elements and found that their distribution was relatively uniform along the chromosomes (NACLERIO *et al.* 1992). However, certain elements, such as CeRep3 (BARNES *et al.* 1995), and short tandem or inverted repeats (*C. ELEGANS* SEQUENCING CONSORTIUM 1998) were found to correlate positively with the rate of recombination. Such results are interpreted as suggesting that some DNA sequences may act as recombination-promoting elements (CANGIANO and LA VOLPE 1993; BARNES *et al.* 1995). We show here that the amount of transposons (DNA-based elements), but not of retroelements (LTR and non-LTR retrotransposons), also correlates positively with recombination rate in the *C. elegans* genome. The analysis of four families of miniature inverted-repeat transposable elements (MITEs), which probably correspond to nonautonomous DNA transposons, also showed an excess of copies on chromosome arms, where the recombination rate is higher (SURZYCKI and BELKNAP 2000). In Drosophila, the analysis of seven retroelements and two transposons (*hobo*, *P*; HOOGLAND and BIÉMONT 1996) showed no correlation between TE frequency and recombination rate, except for *hobo*, which showed a positive correlation like *C. elegans* transposons.

TABLE 3

Proportion of coding and noncoding regions according to recombination rate

	% total	Recombination rate (%)		
		Low	Medium	High
Coding region	21.9	27.5	20.8	16.6
Intron	20.0	20.6	19.5	20.8
Intergenic	58.1	51.9	59.7	62.7

Population genetics models predict that the efficacy of selection should positively correlate with recombination rate (HILL and ROBERTSON 1966; MAYNARD-SMITH and HAIGH 1974; CHARLESWORTH *et al.* 1993). Selection against the deleterious effects of TE insertions should therefore be weaker in regions of low recombination. Moreover, TE insertion may induce deleterious chromosomal rearrangements by recombination between different copies. Under the assumption that the rates of ectopic exchange and meiotic recombination are correlated [which appears to be the case, at least in yeast (GOLDMAN and LICHTEN 1996)], it has been suggested that selection against TE insertion should be stronger in regions of high recombination (LANGLEY *et al.* 1988; CHARLESWORTH *et al.* 1994). Both models thus predict that TEs should accumulate in regions of low recombination where they are less counterselected. Our analyses showed an absence of negative correlation between TE

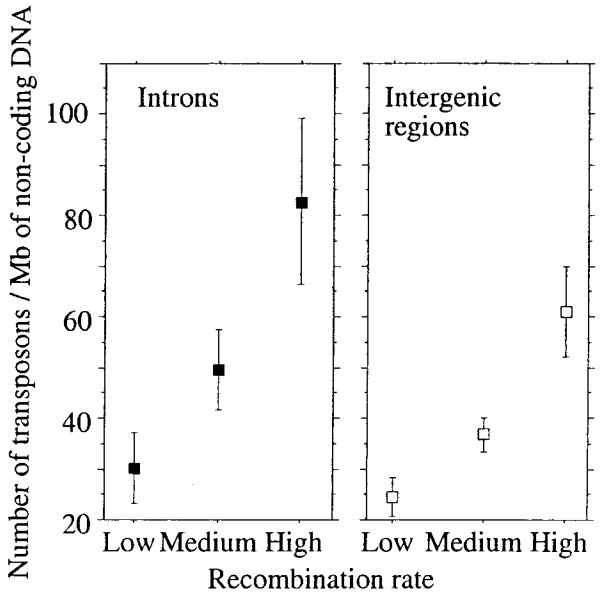

FIGURE 2.—Density of transposons in introns and intergenic regions according to recombination rate. Error bars indicate the 95% confidence interval.

density and recombination rate, which leads to the conclusion that direct or indirect selection against deleterious effects of TE insertions is not the main explanation for maintenance of the TEs in the *C. elegans* genome, contrary to what is proposed in Drosophila (LANGLEY *et al.* 1988; VIEIRA and BIÉMONT 1996; BIÉMONT *et al.* 1997; CHARLESWORTH *et al.* 1997). Of course this does not mean that there is no selection at all against TEs in the *C. elegans* genome; it only means that selection is not the main factor determining the distribution of TEs along the *C. elegans* chromosomes. The mechanisms involved are thus likely to depend on specific characteristics of the *C. elegans* genome and of the transposons.

Autosomes/X chromosome comparison: The X chromosome differs from the autosomes in that it is hemizygous in male (*C. elegans* males are XO, hermaphrodites are XX). Therefore, recessive TE-associated deleterious insertions on the X should be more strongly selected against than TE insertions on the autosomes (MONTGOMERY *et al.* 1987; LANGLEY *et al.* 1988; CHARLESWORTH *et al.* 1994). According to this model of selection, a smaller frequency of insertions should be observed on the X in comparison with the autosomes, as sometimes reported in Drosophila (BIÉMONT 1992). We found that the overall TE density in the X chromosome was slightly higher (37.5 copies/Mb) than that in autosomes (31.5 copies/Mb; Table 4). This is, however, not a general rule: 7 TE families (transposons or retrotransposons) were found in excess on the X chromosome [as has been reported previously for Tc7 (REZSOHAZY *et al.* 1997)], whereas 4 families were underrepresented on the X chromosome and 14 families showed no significant bias (Table 4). Thus, there is no evidence for a stronger selection against TE insertions on the X chromosome than on the autosomes.

How to explain that some TEs are in excess on the X whereas others are underrepresented? It is known that for many TEs, transposition is restricted either to the male or the female germline (HAOUDI *et al.* 1997; PASYUKOVA *et al.* 1997). Since the X chromosome spends more time in the female germline than do the autosomes, if transposition is restricted to the male germline, then fewer TEs are expected on the X than on autosomes (and conversely for TEs with female-restricted transposition). In *C. elegans* the sex ratio is highly biased: this worm reproduces mainly through self-fertil-

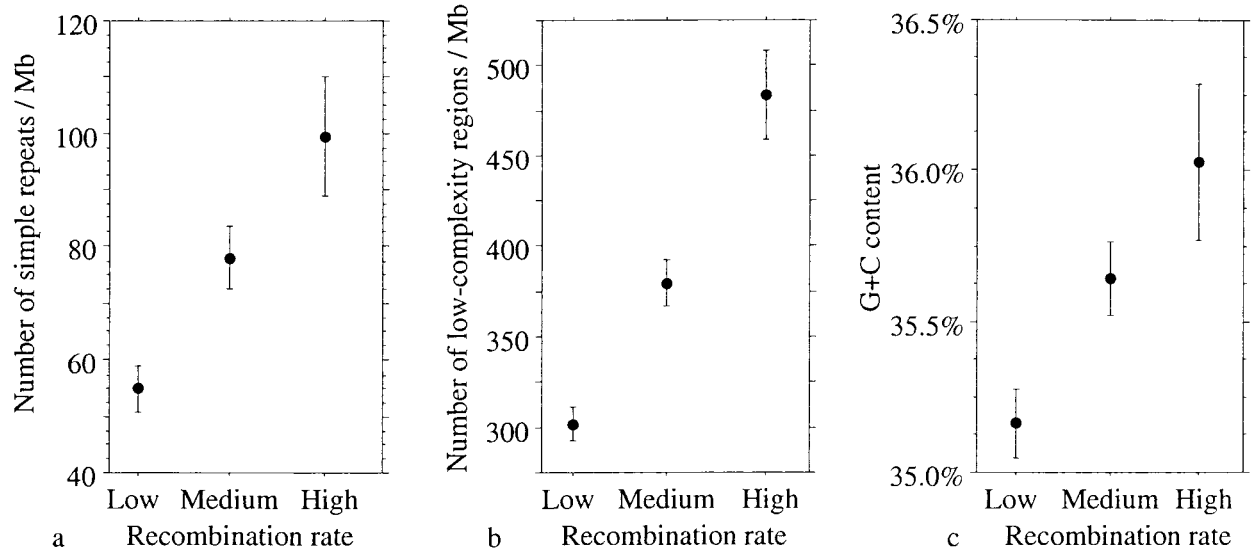

FIGURE 3.—Frequency of (a) simple repeats, (b) low-complexity regions, and (c) variation of G + C content according to recombination rate. Error bars indicate the 95% confidence interval.

TABLE 4

Distribution of transposable elements in autosomes and the X chromosome

	Observed copy number			Density (copy number/Mb)	
	Autosomes (77.2 Mb)	X (17.3 Mb)	χ^2 [a]	Autosomes	X
IR-1	106	21	0.26 NS	1.37	1.21
IR-2	180	4	32.00 A > X**	2.33	0.23
IR-3	947	257	7.48 A < X*	12.27	14.87
IR-4	76	5	7.96 A > X**	0.98	0.29
IR-5	85	2	14.90 A > X**	1.10	0.12
Tc1	98	33	4.16 A < X*	1.27	1.91
Tc2	205	89	28.20 A < X**	2.66	5.15
Tc3	53	8	1.10 NS	0.69	0.46
Tc4	180	53	3.09 NS	2.33	3.07
Tc5	135	24	1.09 NS	1.75	1.39
Tc6	187	82	26.72 A < X**	2.42	4.74
Tc7	183	70	14.86 A < X**	2.37	4.05
Total transposons	2435	648	15.27 A < X**	31.54	37.48
Cer1	7	17	44.31 A < X**	0.09	0.98
Rte-1	75	23	1.75 NS	0.97	1.33
Frodo-1	84	8	5.67 A > X*	1.09	0.46
Frodo-2	64	7	3.38 NS	0.83	0.40
Sam1	36	7	0.12 NS	0.47	0.40
Sam2	17	10	6.34 A < X*	0.22	0.58
Sam3	27	7	0.12 NS	0.35	0.40
Sam4	16	0	3.58 NS	0.21	0.00
Sam5	17	5	0.29 NS	0.22	0.29
Sam6	12	0	2.69 NS	0.16	0.00
Sam7	40	13	1.38 NS	0.52	0.75
Sam8	49	11	0.00 NS	0.63	0.64
Sam9	70	13	0.39 NS	0.91	0.75
Total retrotransposons	514	121	0.24 NS	6.66	7.00

[a] A > X, higher density in autosomes than in X; A < X, lower density in autosomes than in X. NS, no significant difference. *$P < 0.05$; **$P < 0.005$.

izing hermaphrodites, with males found at frequency <0.5% as a result of meiotic X chromosomes nondisjunction (Hodgkin *et al.* 1979; Hodgkin and Barnes 1991; LaMunyon and Ward 1997). Therefore, 0.25% of the X chromosomes of a population are in males, and 99.75% are in hermaphrodites, whereas 0.5 and 99.5% of autosomes are in males and hermaphrodites, respectively. The X chromosome thus spends in the male germline only half the time spent by the autosomes, whereas the time spent in the female germline is nearly the same for both autosomes and the X. Hence, whereas male-restricted transposition could account for TE underrepresentation on the X, female-restricted transposition cannot explain the excess on the X observed for 7 of the TE families. It is possible that other specific features of the X chromosome (*e.g.*, differences in chromatin structure, process of dosage compensation) interfere with TE insertions. The reason for the different distributions of TE families on the X and autosomes remains thus an open question.

TE density in introns and intergenic regions: Chromosome arms (where the frequency of recombination is high) contain proportionally more noncoding DNA (half of this noncoding DNA is nonrepetitive) than the clusters (where the frequency of recombination is low). This negative correlation between recombination rate and coding density (see Table 3) could account for the positive correlation between transposon density and recombination rate: TE insertions are less likely to be deleterious (and thus less likely to be counterselected) in a gene-poor than in a gene-rich region. However, the observation that the density of transposons in introns and intergenic regions follows the recombination rate (Figure 2) argues against this hypothesis. One might argue that noncoding sequences contain regulatory elements and thus do not represent entirely neutral loci for the insertion of transposable elements. However, it is difficult to explain why the density of such regulatory elements should decrease with increasing recombination rate, both in introns and intergenic regions. Another argument against this model is that this negative correlation between gene density and TE density should stand for all classes of TEs, and not only for transposons. The observation that transposon density is similar in

introns and intergenic regions, independently of recombination rate (Figure 2), is consistent with the hypothesis that insertions of transposons are selectively neutral in both introns and intergenic regions. The distribution of transposons in noncoding regions thus directly reflects their pattern of insertion.

Links between TE insertion and recombination: Several hypotheses can be proposed to explain the positive correlation between TE insertion and recombination: either TE insertion enhances recombination or recombination promotes TE insertion or both phenomena are linked to a third unknown factor. These three models are discussed below.

The *C. elegans* DNA is highly interspersed with repetitive sequences (Emmons *et al.* 1980), which represent ~17% of the genome (Sulston and Brenner 1974). Some of these sequences could thus act as recombination-promoting elements (Cangiano and La Volpe 1993), as postulated for the CeRep3 repetitive sequence (Barnes *et al.* 1995). The uneven distribution of such elements along chromosome arms would thus account for the nonuniform recombination rate. For instance, it has been reported for various elements in maize, Drosophila, and *C. elegans* that the double-strand breaks initiated upon TE excision enhance recombination (Dooner and Martinez-Ferez 1997). Notably, it has been shown in Drosophila that transposase activity increased recombination rate, especially around the transposon insertion sites (McCarron *et al.* 1994). The fact that the first step of the transposition of retrotransposons involves transcription instead of excision would account for the absence of correlation between recombination and retrotransposon density. However, the relationship between TE excision and recombination does not seem to be general since it has been demonstrated that germinal excisions of the maize transposon *activator* do not stimulate meiotic recombination (Dooner and Martinez-Ferez 1997). Moreover, the positive association between recombination and the CeRep repetitive sequences, which do not code for a transposase, also argues against such a hypothesis. Finally, a last argument against an effect of TE activity on recombination is that, whereas germline transposition is active in some natural isolates of *C. elegans*, only somatic (nonheritable) transposition has been described in the laboratory strain N2 (Plasterk 1993; Ketting *et al.* 1999). Since germline transposition appears to be strongly repressed in the strain from which genetic maps were built, it seems unlikely that TE activity might be responsible for the observed variations in recombination rates.

An alternative hypothesis is that the genome might be more accessible to transposon insertions in regions of intense recombination. Interestingly, DNA elements transpose by a cut-and-paste mechanism, which involves double-strand break events that are required for the initiation of meiotic recombination (Cao *et al.* 1990). Transposons could thus take advantage of the recombination machinery for their own insertion. The mechanism of integration of the cDNA of LTR retrotransposons is similar to that of DNA transposons, and it has been shown in yeast that LTR retrotransposons are captured at sites of chromosomal double-strand breaks (Moore and Haber 1996). It remains, however, to be determined why the retrotransposons are not concerned with recombination in *C. elegans*. Most of the retroelements we analyzed (12/13) are non-LTR retrotransposons, and it has been shown in mammals that the integration of these elements is coupled to retrotranscription, which is directly primed on the target DNA (Kazazian and Moran 1998). This difference in the mechanism of integration of non-LTR retrotransposons compared to other elements might explain why there is no relationship between retrotransposon density and recombination.

Finally, we cannot eliminate the hypothesis that the correlation between TE insertion and recombination is indirect. Notably, it is conceivable that the distribution of target sites for TE insertions varies with DNA base composition. However, it is unlikely that the very small variation in G + C content with the recombination rate that we observed can account for the difference in TE density. Alternatively, it is possible that the presence of particular sequences, such as low-complexity regions, microsatellites, or other kind of repeats, affects TE insertions. For example, one might imagine that transposons insert preferentially in regions where CeRep sequences are already inserted, making the correlation between transposons and recombination rate only fortuitous. It is also possible that variations in the structure of the chromatin along chromosomes affect independently the rates of both TE insertion and recombination. The analysis of TE distribution in other complete genomes should probably help to distinguish between these different hypotheses.

We thank R. Grantham and C. Vieira for their help. We thank two anonymous referees for their helpful comments. This work was supported by the Centre National de la Recherche Scientifique (CNRS), the Ministère de la Recherche, the "Programme Génome" of the CNRS, and the Association pour la Recherche sur le Cancer.

LITERATURE CITED

Bachtrog, D., S. Weiss, B. Zangerl, G. Brem and C. Schlotterer, 1999 Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. Mol. Biol. Evol. **16:** 602–610.

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

Benson, D. A., M. S. Boguski, D. J. Lipman, J. Ostell, B. F. F. Ouellette *et al.*, 1999 GenBank. Nucleic Acids Res. **27:** 12–17.

Biémont, C., 1992 Population genetics of transposable DNA elements: a Drosophila point of view. Genetica **86:** 67–84.

Biémont, C., H. Tsitrone, C. Vieira and C. Hoogland, 1997 Transposable element distribution in Drosophila. Genetics **147:** 1997–1999.

Britten, R. J., 1995 Active gypsy/Ty3 retrotransposons or retroviruses in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **92:** 599–601.

C. elegans Sequencing Consortium, 1998 How the worm was won: the *C. elegans* genome sequencing project. Science **282:** 2012–2018.

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.

Cao, L., E. Alani and N. Kleckner, 1990 A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae.* Cell **61:** 1089–1101.

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

Charlesworth, B., C. H. Langley and P. Sniegowski, 1997 Transposable element distribution in Drosophila. Genetics **147:** 1993–1995.

Collins, J. J., and P. Anderson, 1994 The Tc5 family of transposable elements in *Caenorhabditis elegans*. Genetics **137:** 771–781.

Collins, J., E. Forbes and P. Anderson, 1989 The Tc3 family of transposable genetic elements in *Caenorhabditis elegans*. Genetics **121:** 47–55.

Devine, S. E., S. L. Chissoe, Y. Eby, R. K. Wilson and J. D. Boeke, 1997 A transposon-based strategy for sequencing repetitive DNA in eukaryotic genomes. Genome Res. **7:** 551–563.

Dib, C., S. Faure, C. Fizames, D. Samson, N. Drouot *et al.*, 1996 A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature **380:** 152–154.

Dooner, H. K., and I. M. Martinez-Ferez, 1997 Germinal excisions of the maize transposon activator do not stimulate meiotic recombination or homology-dependent repair at the bz locus. Genetics **147:** 1923–1932.

Dreyfus, D. H., and S. W. Emmons, 1991 A transposon-related palindromic repetitive sequence from *C. elegans*. Nucleic Acids Res. **19:** 1871–1877.

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.

Goldman, A. S., and M. Lichten, 1996 The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. Genetics **144:** 43–55.

Haoudi, A., M. Rachidi, M. H. Kim, S. Champion, M. Best-Belpomme *et al.*, 1997 Developmental expression analysis of the 1731 retrotransposon reveals an enhancement of Gag-Pol frameshifting in males of *Drosophila melanogaster.* Gene **196:** 83–93.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on the limit to artificial selection. Genet. Res. **8:** 269–294.

Hodgkin, J., and T. M. Barnes, 1991 More is not better: brood size and population growth in a self-fertilizing nematode. Proc. R. Soc. Lond. Ser. B. **246:** 19–24.

Hodgkin, J. A., H. R. Horvitz and S. Brenner, 1979 Nondisjunction mutants of the nematode *Caenorhabditis elegans*. Genetics **91:** 67–94.

Hoogland, C., and C. Biémont, 1996 Distribution of transposable elements along the polytene chromosomes of *Drosophila melanogaster:* test of the ectopic recombination model for maintenance of insertion site number. Genetics **144:** 197–204.

Kazazian, H. H., and J. V. Moran, 1998 The impact of L1 retrotransposons on the human genome. Nat. Genet. **19:** 19–24.

Ketting, R. F., T. H. Haverkamp, H. G. Van Luenen and R. H. Plasterk, 1999 Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell **99:** 133–141.

Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.

LaMunyon, C. W., and S. Ward, 1997 Increased competitiveness of nematode sperm bearing the male X chromosome. Proc. Natl. Acad. Sci. USA **94:** 185–189.

Langley, C. H., E. A. Montgomery, R. Hudson, N. Kaplan and B. Charlesworth, 1988 On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. **52:** 223–235.

Li, W., and J. E. Shaw, 1993 A variant Tc4 transposable element in the nematode *C. elegans* could encode a novel protein. Nucleic Acids Res. **21:** 59–67.

Marin, I., P. Plata-Rengifo, M. Labrador and A. Fontdevilla, 1998 Evolutionary relationships among the members of an ancient class of non-LTR retrotransposons found in the nematode *Caenorhabditis elegans.* Mol. Biol. Evol. **15:** 1390–1402.

Maynard-Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

McCarron, M., A. Duttaroy, G. Doughty and A. Chovnick, 1994 Drosophila *P* element transposase induces male recombination additively and without a requirement for *P* element excision or insertion. Genetics **136:** 1013–1023.

Montgomery, E. A., B. Charlesworth and C. H. Langley, 1987 A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster.* Genet. Res. **49:** 31–41.

Moore, J. K., and J. E. Haber, 1996 Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature **383:** 644–646.

Naclerio, G., G. Cangiano, A. Coulson, A. Levitt, V. Ruvolo *et al.*, 1992 Molecular and genomic organisation of clusters of repetitive DNA sequences in *Caenorhabditis elegans*. J. Mol. Biol. **226:** 159–168.

Pasyukova, E. G., S. V. Nuzhdin, W. Li and A. J. Flavell, 1997 Germ line transposition of the copia mobile element in *Drosophila melanogaster* is restricted to males by sex-specific control of copia RNA level. Mol. Gen. Genet. **255:** 115–124.

Plasterk, R. H., 1993 Molecular mechanisms of transposition and its control. Cell **74:** 781–786.

Rezsohazy, R., H. G. Van Luenen, R. M. Durbin and R. H. Plasterk, 1997 Tc7, a Tc1-hitchhiking transposon in *Caenorhabditis elegans*. Nucleic Acids Res. **25:** 4048–4054.

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

Ruvolo, V., J. E. Hill and A. Levitt, 1992 The Tc2 transposon of *Caenorhabditis elegans* has the structure of a self-regulated element. DNA Cell. Biol. **11:** 111–122.

Shapiro, J. A., 1999 Transposable elements as the key to a 21st century view of evolution. Genetica **107:** 171–179.

Stephan, W., and S. Cho, 1994 Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. Genetics **136:** 333–341.

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

Surzycki, S. A., and W. R. Belknap, 2000 Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. Proc. Natl. Acad. Sci. USA **97:** 245–249.

Vieira, C., and C. Biémont, 1996 Selection against transposable elements in *D. simulans* and *D. melanogaster*. Genet. Res. **68:** 9–15.

Youngman, S., H. G. Van Luenen and R. H. Plasterk, 1996 Rte-1, a retrotransposon-like element in *Caenorhabditis elegans*. FEBS Lett. **380:** 1–7.

Communicating editor: J. Hey