Note

Intron Size Correlates Positively With Recombination Rate in *Caenorhabditis elegans*

Anuphap Prachumwat, Laura DeVincentis and Michael F. Palopoli[1]

Department of Biology, Bowdoin College, Brunswick, Maine 04011

Manuscript received October 2, 2003
Accepted for publication December 10, 2003

ABSTRACT

A negative correlation between intron size and recombination rate has been reported for the *Drosophila melanogaster* and human genomes. Population-genetic models suggest that this pattern could be caused by an interaction between recombination rate and the efficacy of natural selection. To test this idea, we examined variation in intron size and recombination rate across the genome of the nematode *Caenorhabditis elegans*. Interestingly, we found that intron size correlated *positively* with recombination rate in this species.

SPLICEOSOMAL introns are widespread and abundant in eukaryotic genomes (Hawkins 1988; Deutsch and Long 1999). For example, it appears that introns constitute ~26, 11, and 24% of the *Caenorhabditis elegans*, *Drosophila melanogaster*, and human genome sequences, respectively. Comparisons of intron size between *Caenorhabditis elegans* and *Drosophila melanogaster* and human genomes (Carvalho and Clark 1999; Comeron and Kreitman 2000). To explain this pattern, Carvalho and Clark (1999) proposed a model in which natural selection favors smaller introns,

To test these models further, we analyzed intron size and recombination rate variation within the genome of the nematode *C. elegans*. Recombination rates and intron sizes vary substantially in this species (Barnes *et al.* 1995; *C. elegans* Sequencing Consortium 1998; Deutsch and Long 1999). On the basis of the population-genetic models described above, we predicted that intron size would correlate negatively with local recombination rate in the *C. elegans* genome.

MATERIALS AND METHODS

Data collection and analysis: The first and last nucleotide positions of both exons and introns for every predicted and confirmed gene were obtained from the flat text file format of the *C. elegans* genome database (Wormbase, http://www.wormbase.org, release WS46, April 2001; Stein *et al.* 2001). For all analyses, data were first imported into Microsoft Excel (version 2001 for Macintosh; Microsoft, Redmond, WA) for data sorting and manipulations. Data were then imported into StatView (version 5.0.1 for Macintosh; SAS Institute, Cary, NC) to conduct statistical analyses and to generate graphs. For intron size *vs.* recombination rate comparison, a bivariate scattergram was generated and Spearman's rank correlation coefficient (corrected for ties) was calculated to test for a were calculated using Mathematica (version 4.0 for Macintosh; Wolfram Research, Champaign, IL) and the equations are available upon request. The numbers of loci present in the recombination maps and identified in the genomic sequence were 111, 118, 121, 117, 125, and 181 for chromosomes I, II, III, IV, V, and X, respectively. Small numbers of introns, positioned at the extreme ends of the chromosomes, were outside of the known recombination map. Altogether, 2474 introns fell in these regions, which represented 2.5% of the final sample; for these introns, the local recombination rate was assumed to be the same as the recombination rate of the nearest locus on the genetic map of that chromosome. To make sure that this assumption did not affect our results, all analyses were repeated using a data set that excluded these introns.

RESULTS

Intron size correlated positively with recombination rate for the entire *C. elegans* genome (Figure 1a; Spearman's rank correlation, $R = 0.174$, $P < 0.0001$). A similar pattern was observed when each autosome was considered separately (Figure 1b), but not when the X chromosome was considered separately (Figure 1c). Spearman's rank correlation coefficients were $R =$ 0.206, 0.179, 0.247, 0.212, 0.145, and -0.018 for chro-

Figure 1.—Relationship between intron size (log scale, bp) and local recombination rate (cM/Mb) in the genome of *C. elegans*. (a) Across the entire genome, intron size correlated positively with recombination rate (Spearman's rank correlation, $R = 0.174$, $P < 0.0001$). (b) Similar trends were observed when each autosome was considered separately, such as chromosome I ($R = 0.206$, $P < 0.0001$). (c) When the X chromosome was considered separately, however, there was no significant correlation between intron size and recombination rate ($R =$

—

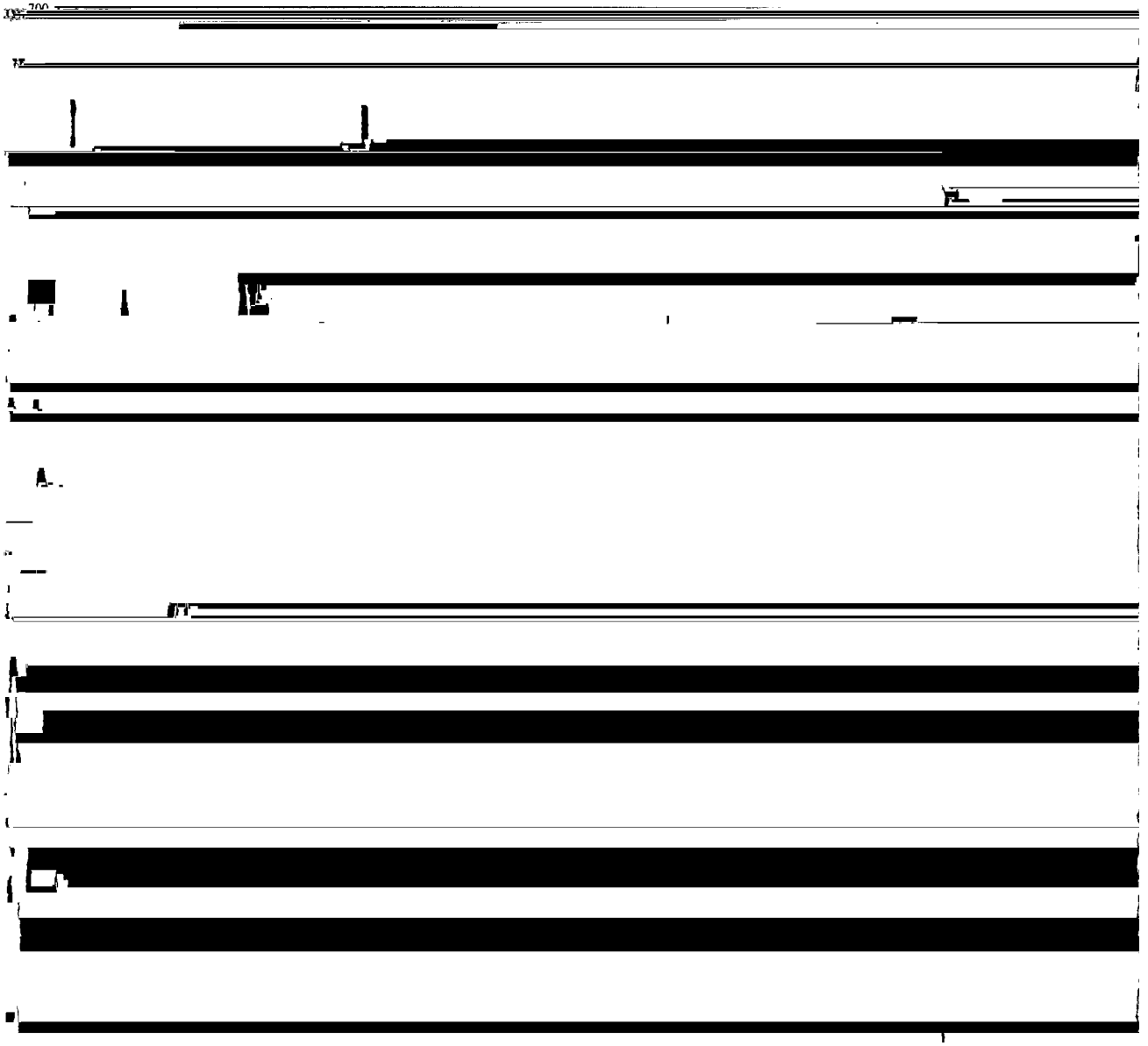

Figure 2.—Comparison of regional averages of intron size (base pairs) and local recombination rate (centimorgans per megabase) across each chromosome in *C. elegans*. These variables exhibited parallel distributions throughout the genome, and the positive correlation between regional averages was statistically significant (Spearman's rank correlation, $R = 0.750$, $P < 0.0001$). Both intron sizes and recombination rates tended to be much greater on the autosomal arms than in the autosomal centers. On the X chromosome, however, average intron size did not exhibit much regional variation. Each chromosome is divided into 10 regions of equal size from left to right. Error bars represent 95% confidence intervals.

genomes and is not predicted by current models for the evolution of intron size (Carvalho and Clark 1999; Comeron and Kreitman 2000).

There were consistent, regional trends in average intron size and average recombination rate across the *C. elegans* genome (Figure 2): (1) autosomal arms tended to have large introns and high recombination rates; (2) autosomal centers tended to have small introns and low recombination rates; and (3) the X chromosome exhibited much less regional variation in average intron size than did any of the autosomes, with average intron sizes intermediate between those observed for autosomal arms and centers. These consistent patterns ruled out the possibility that the genome-wide positive correlation was due to a few regions with widely divergent intron sizes and/or recombination rates.

Population-genetic models have assumed that regional variation in intron size across the genome is determined largely by an interaction between recombination rate and the efficacy of natural selection, termed the Hill-Robertson effect (Carvalho and Clark 1999; Comeron and Kreitman 2000). These models were developed to explain the *D. melanogaster* results and predict a negative correlation between intron size and re-

combination rate. Our results suggest that these models have omitted the main factor(s) that determines regional variation in intron size across the *C. elegans* genome. Since the correlation between these variables was in the opposite direction in *C. elegans vs.* ments. Expansion of noncoding DNA in these clusters would tend to impose a fitness cost, and deletions of noncoding DNA in these regions would often be favored by natural selection. In contrast, in those regions of the

Carvalho, A. B., and