Fourfold Faster Rate of Genome Rearrangement in Nematodes Than in *Drosophila*

Avril Coghlan and Kenneth H. Wolfe

Genome Res. 2002 12: 857-867
Access the most recent version at doi:10.1101/gr.172702

Article

Fourfold Faster Rate of Genome Rearrangement in Nematodes Than in *Drosophila*

Avril Coghlan and Kenneth H. Wolfe[1]

Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland

We compared the genome of the nematode *Caenorhabditis elegans* to 13% of that of *Caenorhabditis briggsae*, identifying 252 conserved segments along their chromosomes. We detected 517 chromosomal rearrangements, with the ratio of translocations to inversions to transpositions being ~1:1:2. We estimate that the species diverged 50–120 million years ago, and that since then there have been 4030 rearrangements between their whole genomes. Our estimate of the rearrangement rate, 0.4–1.0 chromosomal breakages/Mb per Myr, is at least four times that of *Drosophila*, which was previously reported to be the fastest rate among eukaryotes. The breakpoints of translocations are strongly associated with dispersed repeats and gene family members in the *C. elegans* genome.

[The following institution kindly provided reagents, samples or unpublished information as indicated in the paper: Genome Sequencing Center, Washington University School of Medicine, St. Louis.]

The genes of *Caenorhabditis elegans* appear to have an unusually rapid rate of evolution. The substitution rates of many *C. elegans* genes are twice those of their orthologs in non-nematode metazoans (Aguinaldo et al. 1997; see Fig. 3 in Mushegian et al. 1998). Even among nematodes, the *C. elegans* small subunit ribosomal RNA gene evolves faster than its orthologs in most of the major clades (see Fig. 1 in Blaxter et al. 1998). It has been estimated that two-thirds of *C. elegans* protein-coding genes evolve more rapidly than their *Drosophila* orthologs (Mushegian et al. 1998). In vertebrates at least, the rate of nucleotide substitution is correlated with that of chromosomal rearrangement (Burt et al. 1999).

Ranz et al. (2001) reported that *Drosophila* chromosomes rearrange at least 175 times faster than those of other metazoans, and at a rate at least five times greater than the rate of the fastest plant genomes. However, no *Caenorhabditis* rate data existed to compare with the *Drosophila* data. Given their fast rate of nucleotide substitution, we guessed that *Caenorhabditis* genomes might have a fast rate of rearrangement. Here, we have estimated the rate of rearrangement since the divergence of *C. elegans* from its sister species *Caenorhabditis briggsae*, using the complete *C. elegans* genome sequence (The *C. elegans* Sequencing Consortium 1998) and 13 Mb of sequence from *C. briggsae* released by the Washington University Genome Sequencing Center (http://genome.wustl.edu/gsc/). Previous studies have shown that *C. elegans* and *C. briggsae* have conservation of gene order over stretches of chromosome that can be up to six genes long (Kuwabara and Shah 1994; Thacker et al. 1999).

To calculate the rate, we estimated the number of chromosomal rearrangements since the speciation of *C. elegans* and *C. briggsae*. Because both species have six chromosomes (Nigon and Dougherty 1949), we assumed that there have not been any fusions or fissions of whole chromosomes since they diverged. Kececioglu and Ravi (1995) and Hannenhalli (1996) have developed computer algorithms that deduce the historical order and sizes of the reciprocal translocations (whereby two nonhomologous chromosomes exchange chunks of DNA by recombination) and/or inversions that have occurred since the divergence of two multichromosomal genomes. However, the *C. elegans* genome evolves not only by reciprocal translocations and inversions, but also by transpositions (whereby a chunk of DNA excises from one chromosome and inserts into a nonhomologous chromosome) and duplications (Robertson 2001). We designed a simple algorithm to calculate the number and sizes of such mutations, although not the order in which they occurred. Our method starts by finding all perfectly conserved segments between two species, in which gene content, order, and orientation are conserved. Next, these segments are fused into larger segments that have been splintered by duplications, inversions, or transpositions. When no more segments can be merged, the final fused segments are assumed to have resulted from fissure of chromosomes by reciprocal translocations.

To convert the observed number of rearrangements into a rate, it is necessary to have an accurate estimate of the *briggsae–elegans* divergence date. Emmons et al. (1979) were the first to estimate this date, using restriction fragment data, venturing that it must be "tens of millions of years" ago. Butler et al. (1981) speculated that the date was 10–100 million years ago (Mya), judging from 5S rRNA sequences, anatomical differences, and protein electrophoretic mobilities. Subsequent estimates based on sequence data were 30–60 Mya (Prasad and Baillie 1989; one gene), 23–32 Mya (Heschl and Baillie 1990; one gene), 54–58 Mya (Lee et al. 1992; two genes), and 40 Mya (Kennedy et al. 1993; seven genes). Nematode fossils are extremely scarce (Poinar 1983). Therefore, to calibrate the molecular clock, these studies either assumed that all organisms have the same silent substitution rate (Prasad and Baillie 1989; Heschl and Baillie 1990) or nonsilent substitution rate (Lee et al. 1992), or that *C. elegans* has the same silent rate as *Drosophila* (Kennedy et al. 1993). These are dubious assumptions; for example, Mushegian et al. (1998) showed that about two-thirds of *C. elegans* genes have a higher rate of nonsilent substitution than their orthologs in *Drosophila*. To gain a more reliable interval estimate of the *briggsae–elegans* speciation date, we used phylogenetic analysis of all genes for which orthologous sequences were available from *C. elegans*, *C. brigg-*

[1]Corresponding author.
E-MAIL khwolfe@tcd.ie; FAX 353 1 679 8558.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.172702.

sae, *Drosophila*, and human. Only those genes that did not have a significantly different amino acid substitution rate in the four taxa were used to produce date estimates.

The *briggsae–elegans* sequence data set is the largest available for any pair of congeneric eukaryotes. Such a big sample has a high power for detecting genome-wide trends. For example, the breakpoints of reciprocal translocations and inversions are frequently near repetitive DNA. This has been observed in bacteria (Romero et al. 1999), yeast (Fischer et al. 2000), insects (Cáceres et al. 1999), mammals (Dehal et al. 2001), and plants (Zhang and Peterson 1999), but not yet in nematodes. Rearrangements near transposable elements may happen when the element is transposing (Zhang and Peterson 1999), but most rearrangements are hypothesized to occur by homologous recombination between nontransposing transposable elements, dispersed repeats, or gene family members. We find that translocation and transposition breakpoints are strongly associated with repeats in the *C. elegans* genome.

RESULTS

Detection of Conserved Segments and Their Length Distribution

Using the BLASTX algorithm (Altschul et al. 1997), we predicted the positions of 1784 genes in the 12.9-Mb sample of *C. briggsae* genomic DNA. The 1784 genes partition the DNA into 756 segments that have been perfectly conserved between the two species. In *C. briggsae*, the segments range from 1 to 19 genes, or 0.6–154 kb. These segments were merged to recreate 252 longer segments that have been fractured by duplications, inversions, or transpositions since speciation. The 252 segments, which we assume to have resulted from fissure of chromosomes by reciprocal translocations, range from 1 to 109 genes in *C. briggsae*, or 1.3–1040 kb (average, 53 kb). In *C. elegans*, the corresponding segments cover 13.7% of the genome, the smallest segment being one gene (0.4 kb), and the largest 167 genes (954 kb; Fig. 1A,B). The segments can be browsed at the web address http://wolfe.gen.tcd.ie/worm/results.html. An example of the representation of a conserved segment on the web site is shown in Figure 2.

If the nine *C. briggsae* supercontigs are concatenated, we have one large 13.3-Mb chunk (the 12.9 Mb sample including internal gaps). If we assume that the 251 translocation breakpoints (and supercontig ends) are distributed at random along this chunk, the probability of recovering a segment $\geq L$ Mb by chance is $e^{-251L/13.3}$ (Ranz et al. 2001). Of the sample of 252 segments detected, after we use the Bonferroni correction for multiple testing, only one is large enough to give a significant result ($P = 8 \times 10^{-7}$). This is a 1.04-Mb segment containing 109 *C. briggsae* genes conserved between *C. briggsae* supercontig FORK and *C. elegans* chromosome *X*. Gene Ontology classifications are only given in WormBase (http://www.wormbase.org/) for 19 of the *C. elegans* orthologs of these 109 *C. briggsae* genes, and there is no obvious relationship between their functions that might provide a selective explanation for why this large segment has been conserved.

Differences among and along *C. elegans* Chromosomes

The median length of a conserved segment is significantly greater on the *C. elegans X* chromosome (40.6 kb) than on autosomes (17.0 kb; Mann-Whitney test: $P < 0.01$). It is not known which (if any) of the nine *C. briggsae* supercontigs in the sample originated from its sex chromosome. However, in *C. elegans*, sex is determined by counting *X* chromosomes via X signal elements on the *X* chromosomes (Akerib and Meyer 1994). We found the ortholog of the strongest *C. elegans* X signal element, the *sex-1* gene (Carmi et al. 1998), on *C. briggsae* supercontig RWRA (Fig. 2). We suggest that RWRA, the largest supercontig (5.0 Mb) in the *C. briggsae* sample, is part of its sex chromosome. RWRA consists of 95 conserved segments matching *C. elegans* autosomes and 23 segments matching the *C. elegans X* chromosome. If RWRA is the *C. briggsae* sex chromosome, the *C. briggsae* sex chromosome must have undergone many reciprocal translocations with autosomes since divergence from *C. elegans*. Conversely, the *C. elegans X* chromosome consists of conserved segments matching five different *C. briggsae* supercontigs, which are unlikely to be all derived from the *C. briggsae X* chromosome.

The 252 conserved segments are scattered over all six *C. elegans* chromosomes (Fig. 3A), with 211 being on autosomes and 41 on the *X* chromosome. Taking Barnes et al.'s (1995) division of *C. elegans* autosomes into arms and centers, we found 102 conserved segments on autosome centers, and 109 on autosome arms (Fig. 3A). The median length of a conserved segment was not significantly different among the centers (20.5 kb), the left arms (17.5 kb), and the right arms (15.1 kb) of autosomes (Kruskal-Wallis test: $P = 0.5$).

Estimating the *briggsae–elegans* Divergence Date

Using the divergence of the nematodes from the arthropods at 800–1000 million years ago (Mya; Blaxter 1998; Brooke 1999) to calibrate the molecular clock, we estimated the *briggsae–elegans* divergence date from 92 sets of orthologs. Each set comprised a *C. briggsae* gene, its *C. elegans* ortholog, one or more orthologs from *Drosophila*, and one or more human orthologs. When the nematode–arthropod divergence is taken to be 800 Mya, a 95% confidence interval for the median *briggsae–elegans* speciation date is 49–94 Mya (median, 70 Mya). If the nematode–arthropod divergence is taken to be 1000 Mya, the interval becomes 61–118 Mya (median, 88 Mya; Fig. 4). Our best estimate of the *briggsae–elegans* speciation date is therefore ~50–120 Mya.

Duplications

From phylogenetic trees, we identified 27 *C. briggsae* genes that have arisen by 14 duplications from 13 ancestral or-

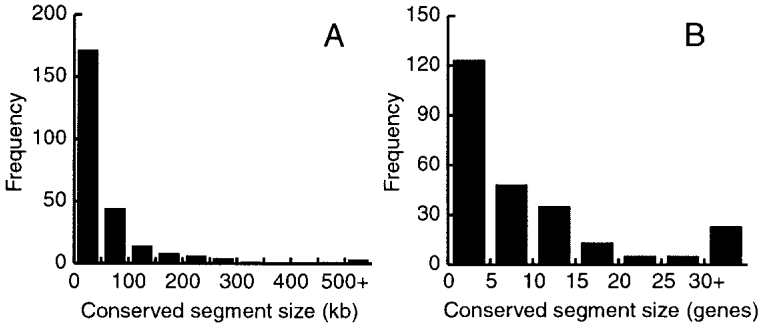

Figure 1 Distribution of sizes of conserved segments, measured in units of kilobases (*A*) and genes (*B*) with respect to *Caenorhabditis elegans*. These conserved segments were assumed to have resulted from fissure of chromosomes by reciprocal translocations.

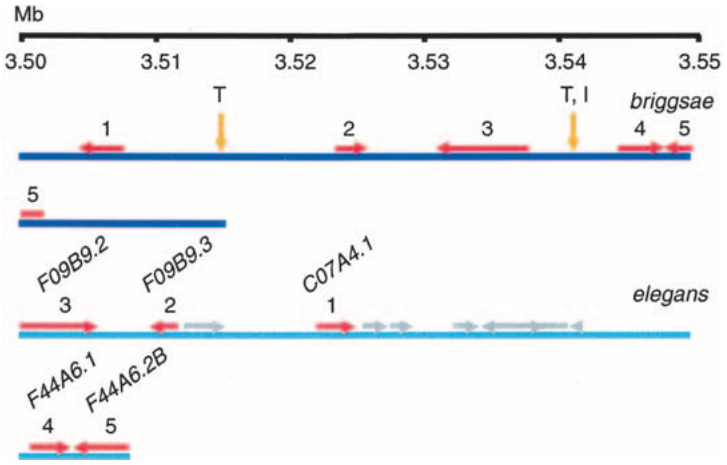

Figure 2 The *Caenorhabditis elegans* region surrounding the *sex-1* locus (*F44A6.2B*), and the corresponding region in *Caenorhabditis briggsae*. The pale blue bar wrapped over two lines represents the region between coordinates 10.18–10.23 Mb of *C. elegans* chromosome *X*, and the navy bar represents 3.50–3.57 Mb of *C. briggsae* contig RWRA. There are five orthologous *briggsae*:*elegans* genes (red) in the conserved segment, which are identified by the same number (1–5) in the two species, and are named on the *C. elegans* map. Inversion (I) and transposition (T) breakpoints are marked with orange arrows, which are shown arbitrarily on the *C. briggsae* chromosome. A region including three genes (*C07A4.1*, *F09B9.3*, and *F09B9.2*) has been inverted in either *C. elegans* or *C. briggsae* since speciation. Furthermore, a region comprising six genes (gray) between *C07A4.1* and *F44A6.1* in *C. elegans* has transposed to another part of the *C. briggsae* genome, or has transposed into this part of the *C. elegans* genome.

thologs at the time of speciation. In 10 of these duplicate pairs, one duplicate has transposed, whereas four of the duplicate pairs have remained adjacent. In two of the four adjacent pairs, one of the duplicates has inverted. Of the 10 duplicates that have transposed, two of the duplicates are on different *C. briggsae* supercontigs. These 10 transpositions and two inversions in *C. briggsae* are the only rearrangements for which we know the genome in which they occurred.

Rates of Reciprocal Translocation, Inversion, and Transposition

Translocations

There is no published estimate of the *C. briggsae* genome size, so we assumed that it is about the same size as the *C. elegans* genome (100.1 Mb). To extrapolate from our sample to the entire *C. briggsae* genome, we assumed that the distribution of conserved segment sizes is the same for the unsequenced and sequenced portions. This seems reasonable because the sizes of conserved segments do not differ among autosomes and, although segments from *C. elegans X* are longer than those from autosomes, the fraction of segments from *X* in our sample (16%) is similar to the fraction of the genome made up by *X* (18%). Because we found 252 conserved segments in 13% of the *C. briggsae* genome, we estimate that there should be 1953 segments in the entire *C. briggsae* genome. The 1953 conserved segments resulted from the (presumably) six chromosomes present in the last common ancestor (*C. briggsae* has six chromosomes; Nigon and Dougherty 1949) plus an estimated 1947 breakpoints due to 974 (1947/2 = 974) translocations that have occurred since speciation. To calculate the rate of reciprocal translocation, the number of translocations is divided by twice the divergence time (Nadeau and Taylor 1984). Our estimate of the speciation date, 50–120 Mya, gives a rate of 4.1–9.7 translocations/Myr for the whole genome. Some of our 252 conserved segments consist of only one gene and might have resulted from transpositions; when we include only segments of ≥3 orthologs, there are 141 conserved segments. Using our 50–120-Mya estimate of the divergence date, this gives a more conservative estimate of 2.3–5.4 translocations/Myr.

Inversions

We detected 121 inversions, including two inversions of duplicated genes that occurred in *C. briggsae* after speciation, and we estimate that there have been 938 inversions in the two genomes since speciation. Using the same divergence date, this implies a rate of 3.9–9.4 inversions/Myr. In *C. elegans*, the inversions range from 1 to 65 genes, or 0.6–367 kb (median, three genes, or 14.4 kb; Fig. 5A,B). About two-thirds of the inversions are <25 kb. The autosomes and the sex chromosome do not have a significantly different median inversion breakpoint density in *C. elegans* (Kruskal-Wallis test: $P = 0.3$). Inversion breakpoints are clustered in hotspots on the *C. briggsae* supercontigs: When we concatenate the three largest supercontigs, the median distance between inversion breakpoints is significantly less than would be expected if breakpoints were uniformly distributed (one-sample sign test: $P = 0.0004$). We noticed that next to inversion breakpoints there are often stretches of *C. elegans* genes whose *C. briggsae* orthologs have not been found. We cannot tell whether their *C. briggsae* orthologs have been deleted, or have transposed to or from an as-yet-unsequenced region of the *C. briggsae* genome.

Transpositions

We assumed that stretches of *C. elegans* genes whose *C. briggsae* orthologs were not found have resulted from transpositions to or from unsequenced parts of the *C. briggsae* genome (Fig. 6A). However, some such transpositions are artifacts. By examining conserved segments, we can see that some of the *C. briggsae* orthologs of *C. elegans* genes have been mistakenly assigned (using BLAST) as the ortholog of a *C. elegans* paralog. In other cases, the *C. elegans* gene appears to be a misprediction, because it has no BLAST hit with $E < 10^{-10}$ in SWISS-PROT or Wormpep. Other such *C. elegans* genes have BLAST hits with $E < 10^{-25}$ to a neighboring *C. elegans* gene, and therefore have probably arisen by tandem duplication since the divergence of *C. elegans* and *C. briggsae*. When we exclude 162 artifacts, 273 transpositions remain. They include 10 transpositions of duplicated genes that have occurred in *C. briggsae*. We estimate that there have been 2116 transpositions in the two genomes since divergence, implying a rate of 8.8–21.2 transpositions/Myr. The 273 transpositions range from 1 to 57 genes, or 0.1–315 kb in *C. elegans* (median, one gene, or 3.3 kb; Fig. 5C,D). Most transposed segments of DNA are <30 kb. The size distribution of transpositions differs from that of inversions, being more skewed toward small rearrangements (Fig. 5D). For the eight *C. briggsae* duplicate genes that have transposed to the same supercontig, there are 1, 1, 7, 14, 42, 42, 46, and 141 intervening genes, respectively, between their old and new locations. For half of these duplicate pairs, there are ≤20 genes between the duplicates. Using the

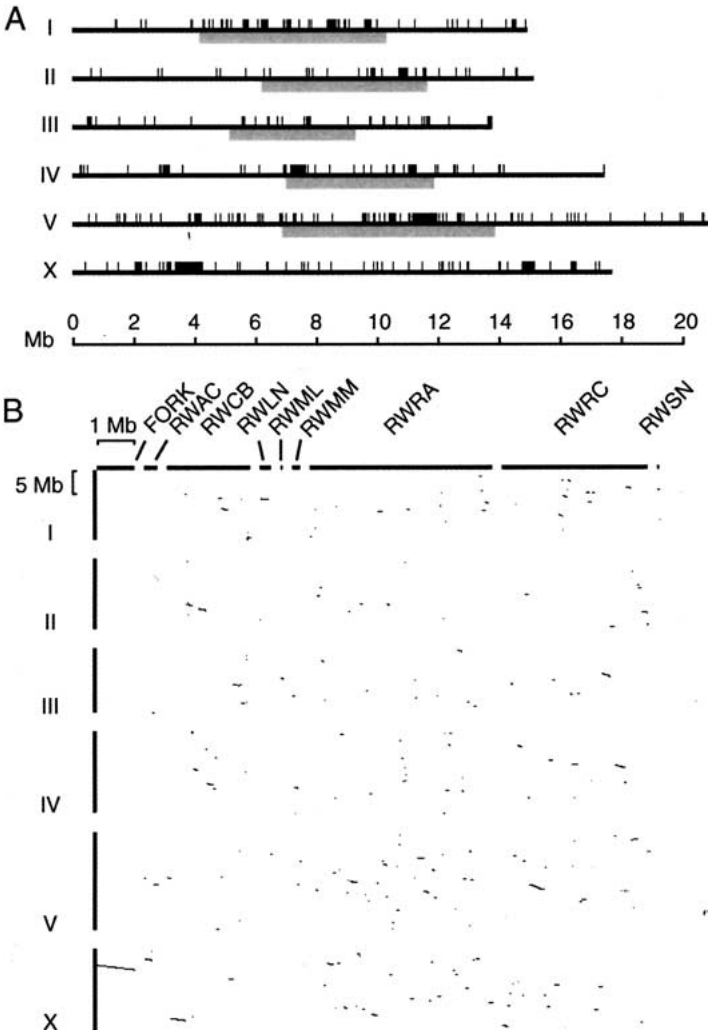

Figure 3 (*A*) Location of conserved segments in the *Caenorhabditis elegans* genome. Gray bars under the autosomes show the "central clusters" described by Barnes et al. (1995). The segments cover ~15% of chromosome *I*, 7% of *II*, 10% of *III*, 13% of *IV*, 15% of *V*, and 20% of *X*. (*B*) Matrix plot comparison between the *C. elegans* genome (vertical axis) and the nine *Caenorhabditis briggsae* contigs (horizontal axis). Conserved segments are indicated by lines drawn between the positions of the outermost genes in each species.

method described in Figure 6B, we observed 79 transpositions. If these had all occurred in *C. elegans*, 24 would have been intrachromosomal, and 21 of these 24 to sites >300 genes away. Thus, some intrachromosomal transpositions are probably to sites far away on a chromosome.

Overall Rate of Rearrangement

Extrapolating from the sequenced 13% to the entire *C. briggsae* genome, we estimate that 974 reciprocal translocations, 938 inversions, and 2116 transpositions have occurred since speciation. About 4030 chromosomal rearrangements have occurred since divergence of the two species. The ratio of translocations to inversions to transpositions is therefore 1.0:1.0:2.3. Each reciprocal translocation causes two breakpoints, each inversion two breakpoints, and each transposition three breakpoints (Sankoff 1999). Therefore, there have been ~10,200 chromosome breakages since speciation, which is 5100 breakages per species, or ~51 breakages/Mb. Using our 50–120-Mya divergence date, this implies a rate of 42–102 breakages/Myr, or 0.4–1.0 breakages/Mb per Myr.

Association of Breakpoints with Repetitive DNA

We obtained the distribution of 33 dispersed repeat sequences in the *C. elegans* genome from WormBase (http://www.sanger.ac.uk/Projects/C_elegans/WORMBASE/GFF_files.shtml; Stein et al. 2001). When we pool all 33 repeats, there is a significant association between dispersed repeats and both translocation and transposition breakpoints in *C. elegans* (Table 1). However, no association is seen for inversion breakpoints. For two individual dispersed repeats, the association with transposition breakpoints is significant ($P < 0.05$; Table 2): CeRep20 and CeRep37. However, the significance of the association is marginal for CeRep20 ($P = 0.045$), whereas the small sample size for CeRep37 makes the test result unreliable.

Translocation breakpoints tend to be next to four different repeats: CeRep13, CeRep15, CeRep19, and CeRep32. *C. elegans* has compound repeats, listed on the Sanger Institute web site (http://www.sanger.ac.uk/Projects/C_elegans/repeats/). The only one associated with translocation breakpoints is CeRep13–CeRep18–CeRep18–CeRep33–CeRep18–CeRep13 ($P = 0.01$; Table 2). However, this is simply owing to the association of CeRep13 with breakpoints, because breakpoints are often near CeRep13/CeRep18/CeRep33, but not CeRep13 + CeRep18 + CeRep33. The association of CeRep19 and CeRep32 with translocation breakpoints is marginally significant ($P \leq 0.05$), but that of CeRep13 and CeRep15 is strong ($P \leq 0.005$). CeRep13 is a 26-bp sequence that is repeated ~1350 times in the *C. elegans* genome, whereas CeRep15 is a 63-bp sequence of which there are about 910 copies. Both these repeats seem to be derived from transposable elements. CeRep13 is 96% identical over 24 bp to the 24-bp terminal inverted repeat (TIR) of *Cele11*, which is thought to be a nonautonomous relative of Tc2, a Tc1/*mariner* family transposon (Oosumi et al. 1996). CeRep15 is 89% identical over 63 bp to part of the 170-bp TIR of *Cele7*, also thought to be a nonautonomous DNA transposon (Oosumi et al. 1995). We

Table 1. Association of Rearrangement Breakpoints with Repeats

Breakpoint type	Number of breakpoints	P-value
Translocation	445	3.3×10^{-5}
Inversion	185	0.10
Transposition	469	2.9×10^{-4}

The number of rearrangement breakpoints in intergenic spacers containing at least one of 33 dispersed repeat families was compared with the number of intergenic spacers in the genome containing one or more dispersed repeats. Only intergenic spacers of 10 kb or shorter were included, of which there are 16,574 in the *Caenorhabditis elegans* genome. The *P*-values for one-sided χ^2 tests are given after applying the Bonferroni correction for multiple testing (multiplies the raw *P*-values by 3).

Table 2. Association of Translocation and Transposition Breakpoints with Particular Repeats

Dispersed repeat	All 16,574 spacers	Translocation breakpoints	*P*-value for translocations	Transposition breakpoints	*P*-value for transpositions
CeRep10	582	23	1.00	25	1.00
CeRep11	137	4	1.00	7	1.00
CeRep12	553	17	1.00	24	1.00
CeRep13	354	22	**0.003**	15	1.00
CeRep14	320	16	0.44	11	1.00
CeRep15	186	14	**0.005**	10	1.00
CeRep17	345	19	0.06	11	1.00
CeRep18	197	10	1.00	11	0.90
CeRep19	685	33	**0.02**	18	1.00
CeRep20	144	9	0.47	11	**0.045**
CeRep21	177	8	1.00	11	0.36
CeRep22	122	6	1.00	8	0.75
CeRep23	708	27	1.00	31	0.42
CeRep24	625	20	1.00	23	1.00
CeRep25	7	1	1.00	0	1.00
CeRep26	154	4	1.00	7	1.00
CeRep27	71	5	1.00	4	1.00
CeRep28	92	4	1.00	5	1.00
CeRep29	150	6	1.00	5	1.00
CeRep30	37	2	1.00	4	0.50
CeRep31	23	1	1.00	2	1.00
CeRep32	226	15	**0.02**	6	1.00
CeRep33	22	1	1.00	1	1.00
CeRep34	321	10	1.00	16	0.78
CeRep35	177	9	1.00	5	1.00
CeRep36	187	6	1.00	6	1.00
CeRep37	122	4	1.00	11	**0.006**
CeRep38	310	12	1.00	15	1.00
CeRep39	14	1	1.00	1	1.00
CeRep40	122	5	1.00	2	1.00
CeRep41	49	3	1.00	1	1.00
CeRep42	110	5	1.00	5	1.00
CeRep43	590	27	0.17	23	1.00
29 + 35 + 36 + 40	24	2	1.00	0	1.00
29/35/36/40	391	14	1.00	10	1.00
17 + 19 + 32	166	11	0.12	5	1.00
17/19/32	720	33	0.06	18	1.00
13 + 18 + 33	17	1	1.00	1	1.00
13/18/33	383	22	**0.01**	15	1.00
34 + 43	212	10	1.00	7	1.00
34/43	699	27	1.00	29	1.00
24 + 38	308	12	1.00	15	1.00
24/38	627	20	1.00	23	1.00

The number of translocation/transposition breakpoints in intergenic spacers containing a particular dispersed repeat was compared with the number of intergenic spacers in the genome containing that repeat. Only intergenic spacers of 10 kb or shorter were included, of which there are 16,574 in the *Caenorhabditis elegans* genome. We tested whether breakpoints are associated with five compound repeats. For example, for the compound repeat CeRep19–CeRep32–CeRep17–CeRep19, we tested whether intergenic spacers containing breakpoints tend to contain all members of the repeat (17 + 19 + 32), or at least one member of this repeat (17/19/32). The *P*-values for one-sided χ^2 tests are given after applying the Bonferroni correction for multiple testing (multiplies the raw *P*-values by 43).

searched the *C. briggsae* genomic DNA for CeRep13 and CeRep15 using FASTA (Pearson and Lipman 1988). Homologs of CeRep13 seem to be present in the *C. briggsae* genome, because it has hits of 91% identity over 22 bp.

It is possible that rearrangement breakpoints could be associated with repeated gene sequences. To investigate this, we used BLASTP (Altschul et al. 1997) with an *E*-value cutoff of 10^{-100} to define families of highly similar genes. There are 1252 families, containing 3901 genes. The proportion of translocation breakpoints that have a gene family member on one or both sides (41%) is significantly greater than the proportion of all *C. elegans* intergenic spacers having a family member on one or both sides (33%; one-sided χ^2 test: $P = 0.0001$). A strong association is also seen for transposition breakpoints (one-sided χ^2 test: $P = 0.0002$), but none for inversion breakpoints.

DISCUSSION

The average size of a conserved segment is 53 kb in *C. briggsae*. This is much larger than the 8.6-kb average found by Kent and Zahler (2000), even though they analyzed a subset of the same *C. briggsae* sequences (8.1 Mb of 12.9 Mb). There are three reasons for the difference. First, Kent and Zahler did not realize that the order and spacing of clones along *C. briggsae* chromosomes are known. The average size of the clones in their

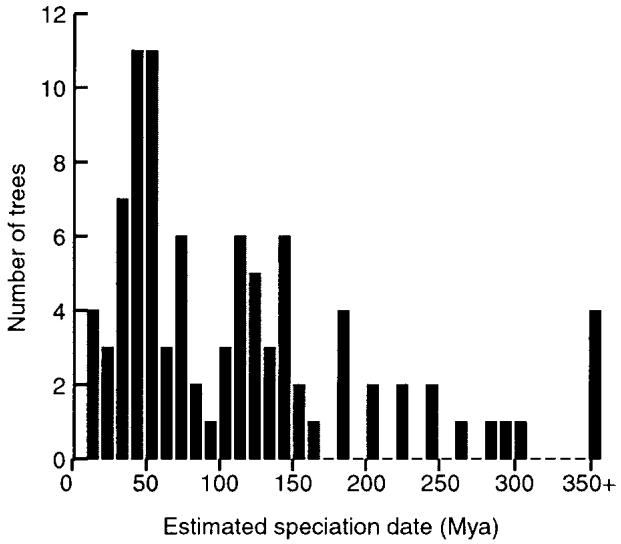

Figure 4 Estimates of the *briggsae–elegans* speciation date from 92 sets of *Caenorhabditis briggsae*, *Caenorhabditis elegans*, *Drosophila*, and human orthologs, calculated by taking the nematode–arthropod divergence date to be 1000 Mya.

sample was 36 kb, whereas the average size of the supercontigs in our sample is 1486 kb. They underestimated the average size of a conserved segment because many clones end before the segment ends. Second, because their method allowed up to 50 kb of contiguous nonsyntenous DNA within a conserved segment in *C. elegans* but only up to 1 kb in *C. briggsae*, it was biased toward finding shorter conserved regions in *C. briggsae* than *C. elegans*. Third, instead of their approach of defining conserved segments by an arbitrary gap size, we strove for a more biologically meaningful approach by searching for the fragments into which chromosomes have been splintered by translocations. We followed Sankoff's (1999) suggestion and regarded inversions and transpositions within translocated segments as noise. For example, Kent and Zahler split the chromosomal region containing the *sex-1* locus into nine segments, partitioning the DNA at poorly conserved noncoding stretches or where there have been small inversions and transpositions. In contrast, we found one large conserved segment in the *sex-1* region (Fig. 2).

Kent and Zahler (2000) found that 63 of their 100 longest conserved segments were near the middle of *C. elegans* autosomes and surmised that "chromosome arms appear to be more susceptible to rearrangement." However, Kent and Zahler's result may merely reflect an ascertainment bias in the data set. We found that although the centers account for two-fifths of the length of autosomes (33 of 82 Mb), half of our conserved segments are from autosome centers (102 of 211 segments). We found no significant difference between the lengths of conserved segments on *C. elegans* autosome arms and centers, which indicates that arms and centers undergo translocations with equal frequency. The rate of chromosomal rearrangement may not be correlated with the nucleotide substitution rate, which does seem to be faster on arms than centers (The *C. elegans* Sequencing Consortium 1998; Koch et al. 2000).

The difference between the median segment size on *X* (41 kb) and on autosomes (17 kb) seems far too large to be attributable to lower detection sensitivity in gene-poor regions. Rather, *X* appears to be better conserved than the autosomes, which must be caused by a lower rate of occurrence or fixation of translocations of *X*. There may be fewer *X* translocations than autosomal translocations because of the lower density of gene family members and of some dispersed repeats on *C. elegans X* chromosome (The *C. elegans* Sequencing Consortium 1998; Surzycki and Belknap 2000). Alternatively, the rate of fixation of translocations may be different for *X* chromosomes and autosomes. Ohno (1967) hypothesized that in mammals and other species such as *C. elegans* that have dosage compensation systems in which *X* genes in XX organisms are down-regulated, *X*–autosomal translocations will be more deleterious than autosome–autosome translocations. This is because *X* genes (for example, dosage-sensitive genes involved in sex determination or sexual dimorphism) that are normally repressed on *X* through dosage compensation would become derepressed when translocated to autosomes, and autosomal genes that are not normally repressed would be repressed when translocated to *X*. There is greater selection against deleterious recessive mutations on the sex chromosome than on autosomes (Montgomery et al. 1987), therefore if most translocations are deleterious recessive, for example, because they upset regulation of expression, we would expect *X* translocations to be fixed less than autosomal translocations. On the other hand, if most translocations are selectively neutral, *X* may have a lower fixation rate because of a lower susceptibility to hitchhiking effects compared with the centers of autosomes (Barnes et al. 1995). A further possibility

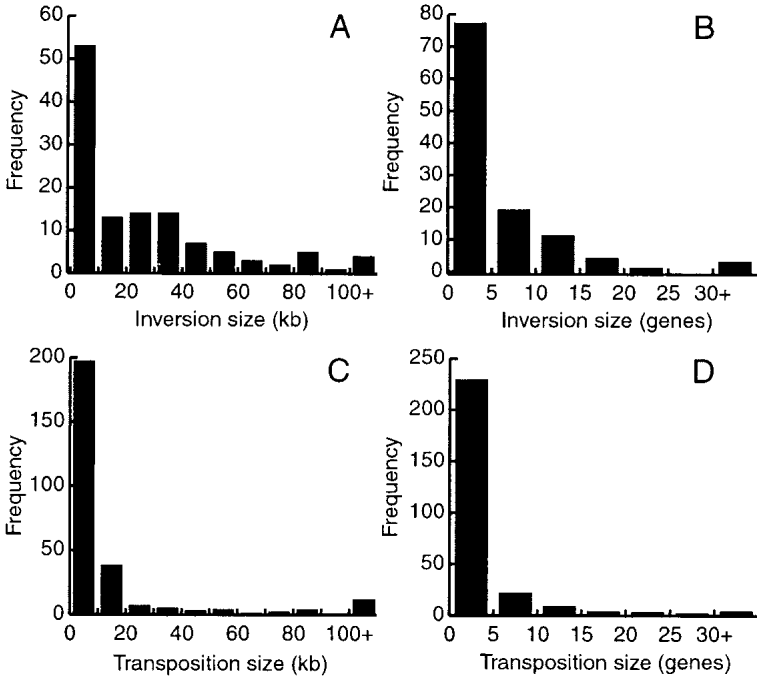

Figure 5 (*A*) Sizes of inversions in kilobases, with respect to *Caenorhabditis elegans*. (*B*) Sizes of inversions, measured in units of genes. (*C*) Sizes of transpositions in kilobases. (*D*) Sizes of transpositions, measured in units of genes.

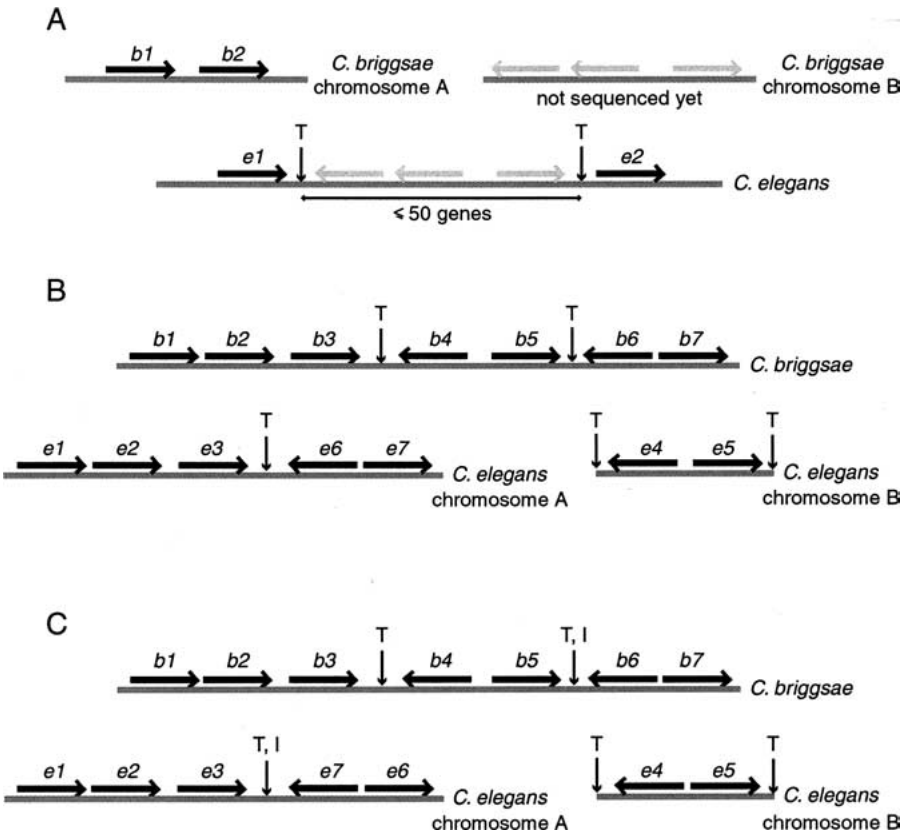

Figure 6 Method of detecting inversions and transpositions. (*A*) To detect transpositions to or from unsequenced parts of the *Caenorhabditis briggsae* genome, we looked along *C. briggsae* contigs for adjacent genes *b1* and *b2* whose *Caenorhabditis elegans* orthologs *e1* and *e2* are on the same chromosome, where between *e1* and *e2* there are 1–50 *C. elegans* genes with unknown *C. briggsae* orthologs. We assumed that the genes between *1* and *2* have transposed in either *C. briggsae* or *C. elegans*. (T) Transposition breakpoints. (*B*) To detect transpositions to or from sequenced parts of the *C. briggsae* genome, we looked along *C. briggsae* contigs for three conserved segments in a row, where in *C. elegans* the first and third segments were close together on the same chromosome, and the middle segment was far away on the same *C. elegans* chromosome or on a different *C. elegans* chromosome. We assumed that the middle segment (genes *4–5*) had transposed in either *C. briggsae* or *C. elegans*. (*C*) To detect inversions, we looked along *C. briggsae* contigs for three conserved segments in a row, where in *C. elegans* the first and third segments were close together on the same chromosome, and the middle segment was far away on the same *C. elegans* chromosome or on a different *C. elegans* chromosome, and either the first or third segment, or both, had inverted in either *C. briggsae* or *C. elegans*. Here the third segment (genes *6–7*) has inverted.

is that there is selection against translocations of the (as yet imprecisely mapped) region(s) of the *X* chromosome from which dosage compensation is initiated, as is seen in mammals (Nesterova et al. 1998).

Translocation and transposition breakpoints are often near repetitive DNA in the *C. elegans* genome, such as gene family members and dispersed repeats. Ectopic recombination between repeats may cause reciprocal translocations. Further study is needed to find out why transposition breakpoints tend to be near dispersed repeats (Table 1) and gene family members. It is possible that we have sometimes mistaken two translocations that were between sites close to each other on the same pair of chromosomes as a transposition.

When counting rearrangements, we could detect only inversions or transpositions of genes within conserved segments. As a result, some transpositions may have been mistaken for translocations, for example, a three-gene segment that transposed to a position between conserved segments. Furthermore, we may not have detected all inversions and transpositions, for example, if an entire conserved segment was inverted. Another possible source of error is that we assumed that stretches of *C. elegans* genes whose *C. briggsae* orthologs have not been found were caused by transpositions to or from an as-yet-unsequenced region of the *C. briggsae* genome (Fig. 6A), but it could be that the *C. briggsae* orthologs have been deleted. Our count of rearrangements may also have been affected by problems that are not specific to our method. First, the average size of a *C. briggsae* supercontig in our sample was 1486 kb, so we may not have detected rearrangements >1.5 Mb. Second, rearrangements that occur twice cannot be detected (Sankoff 1999). Third, it can be impossible to distinguish between three overlapping inversions and a single transposition (Blanchette et al. 1996). Following Nadeau and Taylor (1984), we attributed the few such ambiguous cases to inversions. However, some such inversions may have been in fact transpositions, because we found that transpositions are more common than inversions in *Caenorhabditis*. Fourth, we could not tell a reciprocal translocation apart from a chromosome fusion followed by a fission unless both of the translocation breakpoints had been found. We assumed ambiguous cases to be reciprocal translocations, not chromosome fusions or fissions, because both species have six chromosomes (Nigon and Dougherty 1949). Thus, we will not have detected if a fusion was followed by a fission in one of the species, or if a chromosome fission occurred in both species since divergence.

We estimate that *Caenorhabditis* has a rearrangement rate of 0.4–1.0 breakages/Mb per Myr. This is 1400–17,000 times the mammalian rate calculated by Ranz et al. (2001). Moreover, it is 4–20 times faster than the rate in *Drosophila*, previously reported to be the fastest rate among eukaryotes (0.05–0.09 breakpoints/Mb per Myr; Ranz et al. 2001). Error in the estimated *briggsae–elegans* divergence date would make our rate estimate inaccurate, but it seems unlikely that we have overestimated the rate of rearrangement. For nematodes to have the same rearrangement rate as *Drosophila*, the *briggsae–elegans* divergence date would have to be 560–1020 Mya; however, the nematode order to which *Caenorhabditis* belongs arose only ~400 Mya (Vanfleteren et al. 1994). *Caenorhabditis* and *Drosophila* differ not only in the rate, but also in the type, of rearrangement seen. In *Caenorhabditis*, translocations and inversions are almost equally frequent. In contrast, in *Dro-*

sophila translocations are rare compared to inversions (Ranz et al. 2001), but in mammals translocations are roughly four times more common than inversions (Ehrlich et al. 1997).

Ranz et al. (2001) analyzed in situ hybridization data from *Drosophila melanogaster* chromosome *3R* and *Drosophila repleta* chromosome *2*, and used a maximum likelihood method to estimate the number of inversions that have occurred since divergence of the two chromosomes. Their likelihood method was designed to give an unbiased estimate of the number of rearrangements, thus differences between our *Caenorhabditis* results and their *Drosophila* results are probably not caused by differences between the methods used. However, some differences between the results are probably due to differences in data quality. For example, it is likely that they have underestimated the rate of small rearrangements in *Drosophila* for two reasons. First, because the orientation of the *Drosophila* markers was not known in both species, they could not detect inversions of single markers (for comparison, ~40% of the *Caenorhabditis* inversions we detected were one gene long; Fig. 5B). Second, their physical map only had one marker per 49 kb in its densest regions, thus the smallest inversion that they could detect was ~100 kb long (for comparison, ~95% of the *Caenorhabditis* inversions detected were <100 kb long; Fig. 5A). By comparing the *D. melanogaster* genome to that of the mosquito *Anopheles gambiae* (soon to be released, Hoffman et al. 2002), it may be possible to estimate the rate of small rearrangements in insects. In contrast, Ranz et al. (2001) will have detected more long inversions than we did, because the average size of a *C. briggsae* supercontig in our sample was ~1.5 Mb, whereas their markers spanned 28 Mb.

We suggest four reasons why *Caenorhabditis* chromosomes may have a faster rearrangement rate than those of *Drosophila*. First, the generation time of *Caenorhabditis* is 4–5 times shorter (3–4 d compared with ~2 wk). Second, *C. elegans* and *C. briggsae* may have a smaller effective population size than *Drosophila*, because they are largely self-fertilizing but *Drosophila* is not. Third, *C. elegans* chromosomes may be more prone to hitchhiking effects than those of *Drosophila*, because in *C. elegans* the most gene-rich regions of autosomes have the lowest recombination rates, but the opposite is true for *Drosophila* (Barnes et al. 1995). In other words, if a selectively neutral rearrangement occurs near a positively selected gene, in *C. elegans* it is less likely to be separated from the selected gene by meiotic recombination, and so is more likely to undergo a selective sweep with that gene. These three reasons may also lie behind the faster substitution rate in *C. elegans* compared with *Drosophila*. However, the amino-acid substitution rate is usually less than two times faster in a *C. elegans* gene than in its *Drosophila* ortholog (see Fig. 3 in Mushegian et al. 1998), whereas the rearrangement rate is at least four times faster in *C. elegans*. Our fourth reason is the only one that may contribute to a higher rearrangement rate in *C. elegans* but not to a higher substitution rate. It is that in selfing species like *C. elegans* and *C. briggsae*, rearrangements that are deleterious when heterozygous are more likely to persist than in an out-crossing species, because homozygous individuals arise sooner (Lande 1979). If this is true, we would expect non-selfing species of *Caenorhabditis*, such as *Caenorhabditis remanei* (Baird et al. 1992), to have a lower rate of rearrangement compared with *Drosophila* than do *C. elegans* and *C. briggsae*. We would also expect greater karyotype variability in *C. elegans* populations than in *Drosophila* or *C. remanei* populations. Genomic sequence from non-selfing *Caenorhabditis* species and data on the karyotype variability in wild *C. elegans* populations could provide clues as to why there is a rate difference.

METHODS

Sources of Sequence Data

Nine supercontig DNA sequences from the *C. briggsae* sequencing project at the Washington University Genome Sequencing Center (http://genome.wustl.edu/gsc/) generated from a fingerprint map of *C. briggsae* (M. Marra, J. Schein, and R. Waterston, unpubl.) were downloaded from the WormBase site (http://www.wormbase.org/; Stein et al. 2001) in July 2001. The *C. briggsae* data consist mainly of genes requested by the Worm Community to be sequenced (Baillie and Rose 2000) and are therefore not a random sample of the genome. The nine supercontigs range from 70 to 5015 kb. Because some of these supercontigs contained large internal gaps, we subdivided supercontigs at any internal gap of >2 kb. The resulting 20 contigs range from 51 to 2288 kb (median, 369 kb) and totaled 12.9 Mb. The 19,957 *C. elegans* protein sequences from Wormpep54 were downloaded from http://www.sanger.ac.uk/Projects/C_elegans/wormpep/ in July 2001. We discarded 586 Wormpep proteins from the genes of transposable elements and genes similar to transposable-element genes, 31 from genes whose chromosomal coordinates in *C. elegans* are unknown, and 713 from alternatively spliced genes (retaining the longest splice variant only); 18,627 proteins remained. *C. elegans* gene coordinates corresponding to ACeDB release WS44 were downloaded in July 2001 from http://www.sanger.ac.uk/Projects/C_elegans/WORMBASE/GFF_files.shtml.

Predicting *C. briggsae* Genes

The *C. briggsae* contigs were largely unannotated, so we predicted *C. briggsae* genes using a spliced alignment approach similar to that of Mironov et al. (1998). This was feasible because protein-coding regions are conserved between the two species, but intergenic regions and introns are not (Kent and Zahler 2000). Regions of the *C. briggsae* contigs homologous to *C. elegans* proteins were identified using BLASTX (Altschul et al. 1997) with the BLOSUM62 scoring matrix (Henikoff and Henikoff 1992), using the SEG filter (Wootton and Federhen 1996), and storing all hits with an *E*-value of ≤0.1. There were 99,221 BLASTX hits. Because BLASTX does not always accurately distinguish between orthologs and paralogs, we kept any overlapping hits having *E*-values within a factor of 90 of each other. Nearby BLASTX hits to the same *C. elegans* protein were assumed to correspond to the exons of a *C. briggsae* homolog, and were merged so long as they were on the same strand of the *C. briggsae* contig. To avoid merging hits that were implausibly far apart on a *C. briggsae* contig, any *C. briggsae* intron could not be >7700 bp, and the summed length of introns in a *C. briggsae* gene could not exceed 8150 bp. These numbers (90, 7700, and 8150) were chosen on inspection of the results. To prevent mistaken merging of tandemly repeated genes on a *C. briggsae* contig, the following rule was used, where "left" and "right" refer to the position on the *C. briggsae* contig. The left BLASTX hit had to start in the *C. elegans* protein before the right hit ended in the *C. elegans* protein, and the left hit had to end in the *C. elegans* protein before the right hit started in the *C. elegans* protein, or the hits overlap by <1100 amino acids. After merging BLASTX hits to predict genes, we found a nonoverlapping set of the most significant *C. briggsae* genes along each supercontig. Lastly, *C. briggsae* genes that hit <45% of the length of the *C. elegans* protein, or had BLASTX *E*-values of $\geq 10^{-5}$, were deleted, as they were probably pseudogenes. On the nine supercontigs, we predicted 1934 *C. briggsae* genes. We will not have detected *C. briggsae* genes that do not have homologs in *C.*

elegans, if any such *C. briggsae* genes exist. In a search of the literature we could not find any examples of *C. briggsae* genes that do not have a *C. elegans* homolog.

Finding Orthologs

We did not use synteny data to define orthologs, only sequence identity and phylogenetic trees, because we wanted to use orthologs to gauge synteny conservation. The 1934 *C. briggsae* genes hit 1804 different *C. elegans* proteins in BLASTX. If a *C. briggsae* gene hit only one *C. elegans* protein in BLASTX, and no other *C. briggsae* genes hit that *C. elegans* protein, then the *C. briggsae* and *C. elegans* genes were taken to be one-to-one orthologs. Based on BLASTX results alone, 1704 one-to-one ortholog pairs were found. Some of these orthologous pairs were detected from BLASTX hits having *E*-values as high as 10^{-6}. For the remaining 230 *C. briggsae* genes, it was necessary to draw 151 different phylogenetic trees to deduce orthology. To find an outgroup for a tree of a *C. briggsae* gene and its *C. elegans* hits we used BLASTP (Altschul et al. 1997) with an *E*-value cutoff of ≤0.1 to compare the *C. elegans* hits with Wormpep54 (19,957 proteins) and to SWISS-PROT (July 2001). For each *C. elegans* protein in the tree, the outgroup was either the top-scoring *C. elegans* hit for which a *C. briggsae* ortholog had previously been identified from BLASTX results, or the top-scoring non-*elegans* hit, whichever had the highest score. The sequences for a tree were aligned using CLUSTALW (Thompson et al. 1994), and a maximum parsimony phylogenetic tree was drawn using protpars (Felsenstein 1993). We bootstrapped the trees using 1000 bootstrap replications in the seqboot algorithm (Felsenstein 1993). Only nodes with bootstraps of ≥80% were used to deduce orthology.

Our final *C. briggsae* data set contains 1934 genes: 1744 genes in one-*briggsae*-to-one-*elegans* orthology relationships, and 190 other genes in the following relationships:

(1) 13 genes in one-*briggsae*-to-many-*elegans* relationships;
(2) 46 genes in many-*briggsae*-to-one-*elegans* relationships: 23 for which the *C. elegans* ortholog is known and 23 for which the *C. elegans* ortholog is unresolved;
(3) 4 genes in many-*briggsae*-to-many-*elegans* orthology relationships;
(4) 13 genes whose *C. elegans* ortholog has been deleted since speciation or had not been sequenced yet; and
(5) 114 remaining *C. briggsae* genes whose orthology is unresolved. For 137 of the genes, orthology could not be decided owing to lack of a suitable outgroup or low bootstraps in trees.

The 137 *C. briggsae* genes of unresolved orthology (mainly histone genes) and the 13 with deleted orthologs were ignored in the subsequent analysis, leaving 1784 *C. briggsae* genes that hit 1792 different *C. elegans* genes, of which 1744 were one-to-one orthologs.

Estimating the *briggsae–elegans* Divergence Date

We downloaded 161,296 human proteins and 35,108 *Drosophila* proteins from GenBank (http://www.ncbi.nlm.nih.gov/Entrez/; December 2001). To find *C. elegans* orthologs of these proteins, we compared them with Wormpep using BLASTP (Altschul et al. 1997) with the SEG filter (Wootton and Federhen 1996). If a human protein hit a *C. elegans* protein with a BLASTP *E*-value of $<10^{-20}$, and the *C. elegans* protein with the second strongest hit had an *E*-value that differed by a factor of 10^{20} or more, then the *C. elegans* protein was considered to be the ortholog of the human protein. We found 238 sets of orthologs, each set containing a *C. briggsae* gene, its *C. elegans* ortholog, one or more human orthologs, and one or more *Drosophila* orthologs. For each set, we aligned the proteins using CLUSTALW (Thompson et al. 1994), and made a guide-tree using protdist and neighbor from the PHYLIP package (Felsenstein 1993). We discarded 33 ortholog sets for which the human sequences did not group together and/or the *Drosophila* sequences did not group together, leaving 205 sets. For each ortholog set, the alignment and guide-tree were used as input for Gu and Zhang's (1997) program GAMMA, which estimated an α parameter for the Γ distribution used to correct for rate variation among amino acid sites. For 31 trees, GAMMA could not estimate the α parameter. For the remaining 174 trees, we used the two-cluster test (Takezaki et al. 1995) to check for unequal rates between lineages, taking human to be the outgroup to *Drosophila* and *Caenorhabditis* (Aguinaldo et al. 1997); 92 trees passed the test at the 5% significance level. For each tree, the branch lengths were re-estimated under the assumption of rate constancy, using Takezaki and Nei's (Takezaki et al. 1995) program with the Γ correction for multiple hits. Although the exact branching order of the chordates, arthropods, and nematodes continues to be hotly debated (Mushegian et al. 1998; Wang et al. 1999), most estimates of the divergence of these three phyla range from 800 to 1000 Mya (Blaxter 1998; Brooke 1999). We calibrated the linearized trees by taking the nematode–arthropod divergence to be 800–1000 Mya.

Finding Conserved Segments and Classifying Breakpoints According to Mutation Type

When two species are compared, any region of their genomes in which gene content and order are conserved is a "conserved segment" (Sankoff 1999). Between two adjacent conserved segments is a "breakpoint" (Sankoff 1999) caused by translocation, inversion, duplication, or transposition. We searched for all perfectly conserved segments on the *C. briggsae* supercontigs: segments in which gene order and orientation are perfectly conserved with *C. elegans*. To estimate the size distribution of different types of mutation, the breakpoints within *C. briggsae* contigs were classified as duplication, translocation, inversion, or transposition breakpoints as described below.

From phylogenetic trees, we identified *C. briggsae* genes that have arisen by duplication since speciation. If two *C. briggsae* duplicates that arose from one ortholog were adjacent, we called the breakpoint between them a duplication breakpoint; if one of the duplicates is inverted, it is also an inversion breakpoint. These breakpoints were subsequently ignored, thereby enlarging the original conserved segments. A conserved segment was then taken to be the region between two as-yet-unexplained breakpoints. Transpositions and inversions were detected as shown in Figure 6. The final conserved segments left after all inversions and transpositions had been found were assumed to be segments whose breakpoints were due to translocations. The final conserved segments were manually edited where, for example, two segments were close in the *C. elegans* genome and probably were the same conserved segment.

Because the lengths of those transpositions involving *C. elegans* genes whose *C. briggsae* orthologs have not yet been sequenced can be measured only in units of *C. elegans* genes (Fig. 6A), the sizes of all inversions and transpositions have been given in terms of the number of *C. elegans* genes. If a transposition had occurred within an inverted segment, the size of the inversion was taken to include the transposed genes; likewise, if an inversion had occurred within a transposed segment, the size of the transposition was taken to include the inverted genes.

Testing Whether Breakpoints Are Associated with Repeats

The positions of 33 dispersed repeat families in the *C. elegans* genome were downloaded from http://www.sanger.ac.uk/

Projects/C_elegans/WORMBASE/GFF_files.shtml. The arrangement of these dispersed repeats into compound repeats was taken from http://www.sanger.ac.uk/Projects/C_elegans/repeats/. To group *C. elegans* proteins into families, we compared Wormpep to itself using BLASTP (Altschul et al. 1997) with the SEG filter (Wootton and Federhen 1996). Proteins A, B, and C were assumed to belong to the same family if A hit B with an *E*-value of $<10^{-100}$ and B hit C with an *E*-value of $<10^{-100}$.

ACKNOWLEDGMENTS

We thank the Genome Sequencing Center, Washington University School of Medicine, St. Louis for allowing us to use DNA sequence data before publication. This work was supported by Enterprise Ireland and Science Foundation Ireland. Many thanks to Karsten Hokamp, Simon Wong, and Cathal Seoighe for useful discussions and advice. A special thanks to Aoife McLysaght for help with the Takezaki method, and to Andrew Lloyd, Noel O'Boyle, Richard Durbin, and three anonymous reviewers for critical reading of the manuscript.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Aguinaldo, A.M., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., and Lake, J.A. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387:** 489–493.

Akerib, C.C. and Meyer, B.J. 1994. Identification of X chromosome regions in *Caenorhabditis elegans* that contain sex-determination signal elements. *Genetics* **138:** 1105–1125.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Baillie, D.L. and Rose, A.M. 2000. WABA success: A tool for sequence comparison between large genomes. *Genome Res.* **10:** 1071–1073.

Baird, S., Sutherlin, M.E., and Emmons, S.W. 1992. Reproductive isolation in Rhabditidae (Nematoda: Secernentea); mechanisms that isolate six species of three genera. *Evolution* **46:** 585–594.

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

Blanchette, M., Kunisawa, T., and Sankoff, D. 1996. Parametric genome rearrangement. *Gene* **172:** GC11–GC17.

Blaxter, M. 1998. *Caenorhabditis elegans* is a nematode. *Science* **282:** 2041–2046.

Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P., Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., et al. 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature* **392:** 71–75.

Brooke, M. de L. 1999. How old are animals? *Trends Ecol. Evolution* **14:** 211–212.

Burt, D.W., Bruley, C., Dunn, I.C., Jones, C.T., Ramage, A., Law, A.S., Morrice, D.R., Paton, I.R., Smith, J., Windsor, D., et al. 1999. The dynamics of chromosome evolution in birds and mammals. *Nature* **402:** 411–413.

Butler, M.H., Wall, S.M., Luehrsen, K.R., Fox, G.E., and Hecht, R.M. 1981. Molecular relationships between closely related strains and species of nematodes. *J. Mol. Evol.* **18:** 18–23.

Cáceres, M., Ranz, J.M., Barbadilla, A., Long, M., and Ruiz, A. 1999. Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285:** 415–418.

Carmi, I., Kopczynski, J.B., and Meyer, B.J. 1998. The nuclear hormone receptor SEX-1 is an X-chromosome signal that determines nematode sex. *Nature* **396:** 168–173.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282:** 2012–2018.

Dehal, P., Predki, P., Olsen, A.S., Kobayashi, A., Folta, P., Lucas, S., Land, M., Terry, A., Ecale Zhou, C.L., Rash, S., et al. 2001. Human chromosome 19 and related regions in mouse: Conservative and lineage-specific evolution. *Science* **293:** 104–111.

Ehrlich, J., Sankoff, D., and Nadeau, J.H. 1997. Synteny conservation and chromosome rearrangements during mammalian evolution. *Genetics* **147:** 289–296.

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

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) Version 3.5c. Department of Genetics, University of Washington, Seattle

Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., and Louis, E.J. 2000. Chromosomal evolution in *Saccharomyces*. *Nature* **405:** 451–454.

Gu, X. and Zhang, J. 1997. A simple method for estimating the parameter of substitution rate variation among sites. *Mol. Biol. Evol.* **14:** 1106–1113.

Hannenhalli, S. 1996. Polynomial-time algorithm for computing translocation distance between genomes. *Discrete Applied Math.* **71:** 137–151.

Henikoff, S. and Henikoff, J.G. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89:** 10915–10919.

Heschl, M.F. and Baillie, D.L. 1990. Functional elements and domains inferred from sequence comparisons of a heat shock gene in two nematodes. *J. Mol. Evol.* **31:** 3–9.

Hoffman, S.L., Subramanian, G.M., Collins, F.H., and Venter, J.C. 2002. *Plasmodium*, human and *Anopheles* genomics and malaria. *Nature* **415:** 702–709.

Kececioglu, J. and Ravi, R. 1995. Of mice and men: Evolutionary distances between genomes under translocation. In *Proceedings of the 6th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 604–613. SIAM, Philadelphia, PA.

Kennedy, B.P., Aamodt, E.J., Allen, F.L., Chung, M.A., Heschl, M.F., and McGhee, J.D. 1993. The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J. Mol. Biol.* **229:** 890–908.

Kent, W.J. and Zahler, A.M. 2000. Conservation, regulation, synteny, and introns in a large-scale *C. briggsae–C. elegans* genomic alignment. *Genome Res.* **10:** 1115–1125.

Koch, R., van Luenen, H.G., van der Horst, M., Thijssen, K.L., and Plasterk, R.H. 2000. Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res.* **10:** 1690–1696.

Kuwabara, P.E. and Shah, S. 1994. Cloning by synteny: Identifying *C. briggsae* homologues of *C. elegans* genes. *Nucleic Acids Res.* **22:** 4414–4418.

Lande, R. 1979. Effective deme size during long-term evolution estimated from rates of chromosomal rearrangements. *Evolution* **33:** 234–251.

Lee, Y.H., Huang, X.Y., Hirsh, D., Fox, G.E., and Hecht, R.M. 1992. Conservation of gene organization and *trans*-splicing in the glyceraldehyde-3-phosphate dehydrogenase-encoding genes of *Caenorhabditis briggsae*. *Gene* **121:** 227–235.

Mironov, A.A., Roytberg, M.A., Pevzner, P.A., and Gelfand, M.S. 1998. Performance-guarantee gene predictions via spliced alignment. *Genomics* **51:** 332–339.

Montgomery, E.A., Charlesworth, B., and Langley, C.H. 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.

Mushegian, A.R., Garey, J.R., Martin, J., and Liu, L.X. 1998. Large-scale taxonomic profiling of eukaryotic model organisms: A comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Res.* **8:** 590–598.

Nadeau, J.H. and Taylor, B.A. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl. Acad. Sci.* **81:** 814–818.

Nesterova, T.B., Duthie, S.M., Mazurok, N.A., Isaenko, A.A., Rubtsova, N.V., Zakian, S.M., and Brockdorff, N. 1998. Comparative mapping of X chromosomes in vole species of the genus *Microtus*. *Chromosome Res.* **6:** 41–48.

Nigon, V. and Dougherty, E.C. 1949. Reproductive patterns and attempts at reciprocal crossing of *Rhabditis elegans* Maupas, 1900, and *Rhabditis briggsae* Dougherty and Nigon, 1949 (Nematoda: Rhabditidae). *J. Exp. Zool.* **112:** 485–503.

Ohno, S. 1967. Sex chromosomes and sex-linked genes. In *Monographs on endocrinology* (eds. A. Labhart et al.), Vol. 1, pp. 123–135. Springer-Verlag, Heidelberg.

Oosumi, T., Garlick, B., and Belknap, W.R. 1995. Identification and characterization of putative transposable DNA elements in

solanaceous plants and *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **92:** 8886–8890.
———. 1996. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J. Mol. Evol.* **43:** 11–18.
Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85:** 2444–2448.
Poinar, Jr., G.O. 1983. *The natural history of nematodes*. Prentice-Hall, Englewood Cliffs, NJ.
Prasad, S.S. and Baillie, D.L. 1989. Evolutionarily conserved coding sequences in the *dpy-20–unc-22* region of *Caenorhabditis elegans*. *Genomics* **5:** 185–198.
Ranz, J.M., Casals, F., and Ruiz, A. 2001. How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Res.* **11:** 230–239.
Robertson, H.M. 2001. Updating the *str* and *srj* (*stl*) families of chemoreceptors in *Caenorhabditis* nematodes reveals frequent gene movement within and between chromosomes. *Chem. Senses* **26:** 151–159.
Romero, D., Martínez-Salazar, J., Ortiz, E., Rodríguez, C., and Valencia-Morales, E. 1999. Repeated sequences in bacterial chromosomes and plasmids: A glimpse from sequenced genomes. *Res. Microbiol.* **150:** 735–743.
Sankoff, D. 1999. Comparative mapping and genome rearrangement. In *From Jay Lush to genomics: Visions for animal breeding and genetics*. (eds. J.C.M. Dekkers, S.J. Lamont, and M.F. Rothschild), pp. 124–134. Iowa State University, Ames, IA.
Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J., and Spieth, J. 2001. WormBase: Network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.* **29:** 82–86.
Surzycki, S.A. and Belknap, W.R. 2000. Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. *Proc. Natl. Acad. Sci.* **97:** 245–249.
Takezaki, N., Rzhetsky, A., and Nei, M. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12:** 823–833.
Thacker, C., Marra, M.A., Jones, A., Baillie, D.L., and Rose, A.M. 1999. Functional genomics in *Caenorhabditis elegans*: An approach involving comparisons of sequences from related nematodes. *Genome Res.* **9:** 348–359.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.
Vanfleteren, J.R., Van de Peer, Y., Blaxter, M.L., Tweedie, S.A., Trotman, C., Lu, L., Van Hauwaert, M.L., and Moens, L. 1994. Molecular genealogy of some nematode taxa as based on cytochrome c and globin amino acid sequences. *Mol. Phylogenet. Evol.* **3:** 92–101.
Wang, D.Y., Kumar, S., and Hedges, S.B. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. R. Soc. Lond. B Biol. Sci.* **266:** 163–171.
Wootton, J.C. and Federhen, S. 1996. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **266:** 554–571.
Zhang, J. and Peterson, T. 1999. Genome rearrangements by nonlinear transposons in maize. *Genetics* **153:** 1403–1410.

WEB SITE REFERENCES

http://evolution.genetics.washington.edu/phylip.html; PHYLIP phylogeny inference package.
http://www.genome.wustl.edu/gsc/; Genome Sequencing Center, Washington University, St. Louis.
http://www.ncbi.nlm.nih.gov/Entrez/; GenBank.
http://www.sanger.ac.uk/Projects/C_elegans/repeats/; List of *C. elegans* repeat families at the Sanger Institute.
http://www.sanger.ac.uk/Projects/C_elegans/WORMBASE/GFF_files.shtml; WormBase release files at The Sanger Institute.
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/; *C. elegans* protein database Wormpep.
http://www.wormbase.org/; WormBase.
http://wolfe.gen.tcd.ie/worm/results.html; Views of worm segments as in Figure 2.

Received February 9, 2002; accepted in revised form April 3, 2002.