REVIEW ARTICLE

Evolution of the *Caenorhabditis elegans* Genome

Asher D. Cutter, Alivia Dey, and Rosalind L. Murray

Department of Ecology & Evolutionary Biology and the Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario, Canada

A fundamental problem in genome biology is to elucidate the evolutionary forces responsible for generating nonrandom patterns of genome organization. As the first metazoan to benefit from full-genome sequencing, *Caenorhabditis elegans* has been at the forefront of research in this area. Studies of genomic patterns, and their evolutionary underpinnings, continue to be augmented by the recent push to obtain additional full-genome sequences of related *Caenorhabditis* taxa. In the near future, we expect to see major advances with the onset of whole-genome resequencing of multiple wild individuals of the same species. In this review, we synthesize many of the important insights to date in our understanding of genome organization and function that derive from the evolutionary principles made explicit by theoretical population genetics and molecular evolution and highlight fertile areas for future research on unanswered questions in *C. elegans* genome evolution. We call attention to the need for *C. elegans* researchers to generate and critically assess nonadaptive hypotheses for genomic and developmental patterns, in addition to adaptive scenarios. We also emphasize the potential importance of evolution in the gonochoristic (female and male) ancestors of the androdioecious (hermaphrodite and male) *C. elegans* as the source for many of its genomic and developmental patterns.

Introduction

How does natural selection infringe upon neutral evolutionary processes to shape an organism's genome? When do neutral forces, like mutation, genetic drift, and gene conversion, play a leading role in generating nonrandom patterns in the genome? In what ways might demography and breeding system impact the character of sequences on a genomic scale? These are some of the broad questions that population genetics and molecular evolutionary theory seek to explain. We aim in this review to integrate what is known about genomic patterns in *Caenorhabditis* with the evolutionary theory that underlies the causal processes driving genomic phenomena. Because the field is on the cusp of population genomics studies in *Caenorhabditis* that will usher in a new wave of evolutionarily relevant information, it is now imperative that we take stock of what has been learned in this system and what we have yet to learn. Here, we target several areas in this effort, paying special attention to nonrandom spatial patterning, gene duplication, intron evolution, mutation, protein and noncoding sequence divergence, biased codon usage, the evolution of repetitive and selfish DNA, and population variation. For fuller discussion of the basic ideas underlying evolutionary predictions for these genomic phenomena, several excellent reviews are available (Kreitman 2000; Yang and Bielawski 2000; Charlesworth 2003, 2009; Charlesworth et al. 2003; Lynch 2007; Wright, Ness, et al. 2008). For descriptions and discussions of additional salient features of the *Caenorhabditis* genomes, we point the reader to other recent reviews (Coghlan et al. 2006; Thomas 2008), to the original genome sequence publications for *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium 1998) and *Caenorhabditis briggsae* (Stein et al. 2003), and to the many relevant chapters in WormBook (http://www.wormbook.org). Our aim here is to synthesize much of this information within the unifying framework of evolutionary theory.

There are four key facts to keep in mind when interpreting nonrandom patterns in the genome of *C. elegans*. First, *C. elegans* populations are composed of self-fertile hermaphrodites that derive from a male–female ancestor (Cho et al. 2004; Kiontke et al. 2004)—with outcrossing being genetically effective only rarely in nature. However, the duration of selfing in the *C. elegans* lineage is not known precisely, and recent evidence suggests that the origin of selfing might be relatively recent (Cutter et al. 2008). Selfing will shape genome evolution due to increased homozygosity, greater linkage disequilibrium, reduced genetic effective population size, structuring of subpopulations, and stronger impacts of genetic hitchhiking with beneficial mutations and background selection against deleterious mutations (Charlesworth and Wright 2001; Wright, Ness, et al. 2008). The consequences of selfing will be rapid for some genomic characteristics (e.g., population polymorphism), but others will yield a detectable evolutionary response only after an extended period in a selfing state (e.g., codon usage bias). Therefore, it is important to consider critically the potential genomic effects of short-term evolution in a selfing state and of evolution in the ancestor that was composed of obligately outbreeding individuals. The relevance of this fact, that evolutionary processes in male–female ancestors are responsible for generating many present-day genomic and phenotypic patterns, is crucial, albeit often neglected, in interpretations of *C. elegans* data.

Second, natural selection is only one of several core evolutionary forces that can lead to nonrandom patterns in genomes. Equally fundamental, mutation, recombination, and genetic drift are agents of evolutionary change that can shape genomes in a nonadaptive, selectively neutral manner. Consequently, it is essential to generate both neutral, nonadaptive hypotheses as well as adaptive hypotheses as testable alternatives (Lynch 2007). Third, the *C. elegans* genome is quite compact for a eukaryote, containing a high fraction of coding and putatively functional noncoding sequence (*C. elegans* Sequencing Consortium 1998; Shabalina and Kondrashov 1999; He et al. 2007). This contrasts with

Key words: Caenorhabditis elegans, genome evolution, gene duplication, intron evolution, mutation, transposable elements, codon bias.

E-mail: asher.cutter@utoronto.ca.

Mol. Biol. Evol. 26(6):1199–1234. 2009
doi:10.1093/molbev/msp048
Advance Access publication March 16, 2009

© The Author 2009. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

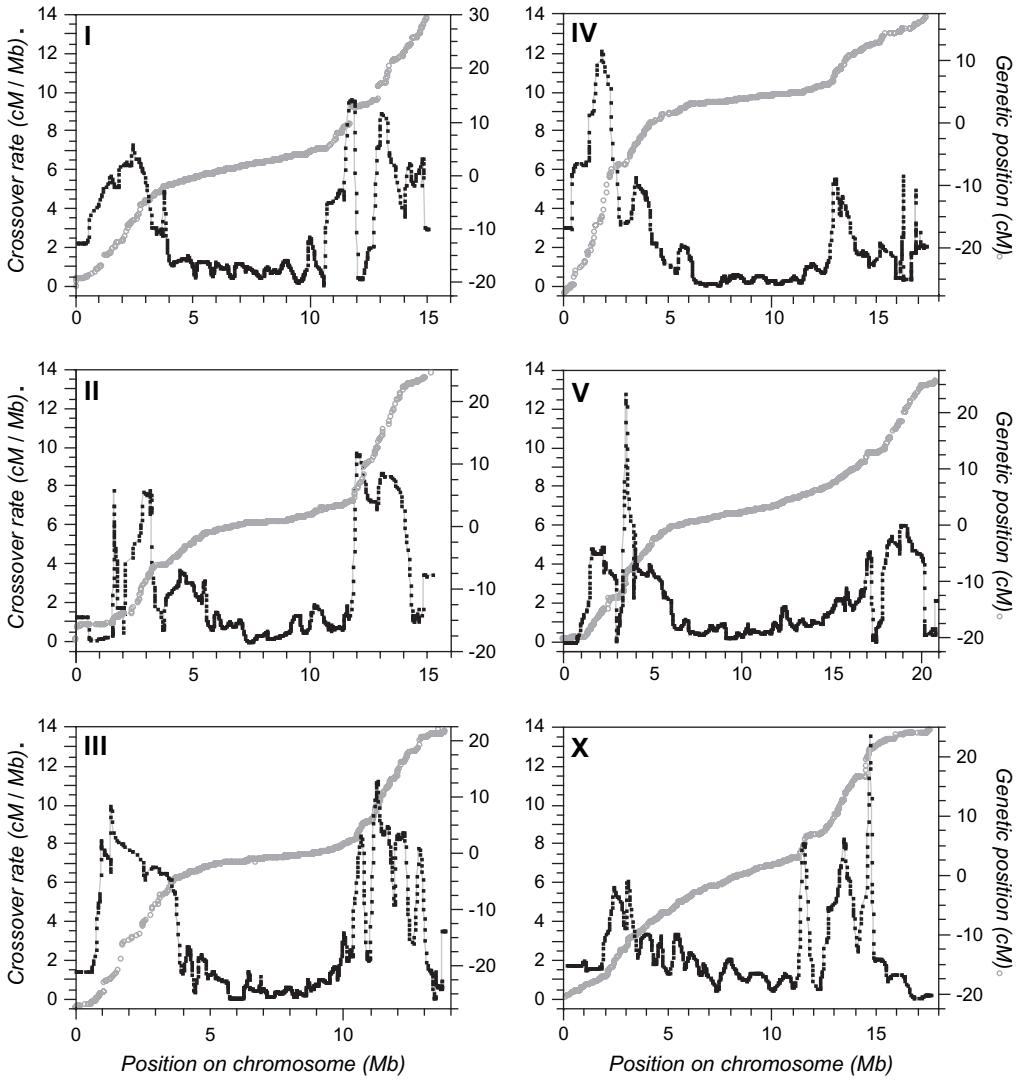

FIG. 1.—Crossover rates and Marey maps (plot of genetic position vs. physical position) for each *Caenorhabditis elegans* chromosome. Points for the Marey maps (gray circles) represent physical and experimental map positions for 4,530 loci in Wormbase WS190. Crossover rate values (black squares) are point estimates centered on each of these loci, based on the slope of the least-squares linear regression line of genetic on physical distance for 12 mapped loci on either side of the focal locus (i.e., a sliding window of slopes for bins of 25 loci).

the large and repeat-rich mammalian genomes (Eichler and Sankoff 2003) and likely has important implications for understanding genomic dynamics. Fourth, the high molecular divergence between *C. elegans* and its closest known relatives, with which it likely shares common ancestry ~110 million generations ago (<30 My; Cutter 2008a), is both a blessing and a curse. The high divergence permits easy identification of conserved sequences, yet is too high to apply some of the most powerful approaches to identifying adaptive evolution of sequence function—making it an important endeavor to sample more deeply *Caenorhabditis* biodiversity to identify more closely related species pairs. We will hearken back to these key features throughout this review.

Nonrandom Patterns of Genome Organization

Intrachromosomal Heterogeneity

The *C. elegans* nuclear genome consists of five autosomes and a sex (X) chromosome (Nigon 1949; Brenner 1974) that encode just over 20,000 protein-coding genes (*C. elegans* Sequencing Consortium 1998; Thomas 2008). However, the organization of genes and other genomic features is nonrandom within and between chromosomes. Here we focus on molecular evolutionary patterns associated with the nuclear genome, although the mitochondrial genome also has been well characterized (Lemire 2005) and new evidence implicates an important role for mtDNA in *C. briggsae* evolution (Howe and Denver 2008).

Autosomes typically are described at a gross scale in terms of the chromosome "arms," which have moderate to high rates of crossover recombination and low gene density, and "centers" or "clusters," which exhibit strikingly little recombination (fig. 1), despite comprising nearly half of each autosome (Barnes et al. 1995; Rockman and Kruglyak 2009). In contrast, the X chromosome has a more uniform distribution of recombination rates and gene density along its length (Brenner 1974; Barnes et al. 1995). Despite these robust broad patterns, small-scale heterogeneity in

recombination rate is likely affected by local factors (Rockman and Kruglyak 2009). Unlike mammals and yeast (Nishant and Rao 2005), however, recombination does not appear to be limited to "hot spot" foci in *C. elegans* (Barnes et al. 1995; Rockman and Kruglyak 2009). The genetic maps of males are shorter than for hermaphrodites (Zetka and Rose 1990; Lim et al. 2008), and temperature does not seem to alter dramatically the *C. elegans* map (Rockman and Kruglyak 2009; but see Zetka and Rose 1990; Lim et al. 2008). Furthermore, *C. elegans* chromosomes do not have discrete, defined centromeres, being holocentric during mitosis with spindle attachment sites along the length of the chromosomes (Albertson and Thomson 1982; Zetka and Rose 1995; Wicky and Rose 1996). However, there is only a single site of spindle attachment during meiosis, occurring at one end or the other of a chromosome— making chromosomes functionally monocentric during meiosis (Albertson and Thomson 1993; Zetka and Rose 1995; Wicky and Rose 1996). Chromosomal termini have very low rates of recombination (Barnes et al. 1995; Rockman and Kruglyak 2009), but it is unknown whether this relates directly to the centromere-like function of the ends of chromosomes. Because the low recombination centers are positioned asymmetrically on the chromosomes, it has been argued that recombination rate variation is not shaped solely by relative chromosomal position (Rockman and Kruglyak 2009). Sequences near the functionally centromeric regions (chromosome arms) have lower gene density, like regions near centromeres in most other familiar taxa, but arms also have higher rates of recombination, unlike regions near centromeres in other taxa. This negative relation between crossover rate and gene density in *Caenorhabditis* provides a useful contrast to *Drosophila*, which may help to dissect their potential influences on the efficacy of natural selection in the context of how linkage induces selective interference, genetic hitchhiking, and background selection (Andolfatto 2001; Payseur and Nachman 2002).

The draft genome sequence of the N2 strain of *C. elegans* was completed in 1998 (*C. elegans* Sequencing Consortium 1998) and finished in 2002; since then, understanding of *C. elegans* genome architecture and organization has flourished. In addition to greater gene density and lower recombination rates in the chromosome centers, these regions also have more operons (Blumenthal et al. 2002) and tend to encode proteins that are more highly conserved (*C. elegans* Sequencing Consortium 1998); these genes have stronger codon usage bias (Marais et al. 2001; Marais and Piganeau 2002) and greater effects on viability, reproduction, and other phenotypes when knocked out by RNA interference (RNAi; Kamath et al. 2003) or mutagenesis screens (Brenner 1974; Johnsen et al. 2000). Meiotic crossover rates correlate with estimates of population recombination rates, such that linkage disequilibrium in nature is higher in chromosomal regions of low recombination (Rockman and Kruglyak 2009). Chromosome centers also exhibit lower densities of transposable elements (TEs), repetitive DNA, and pseudogenes (Cangiano and La Volpe 1993; Barnes et al. 1995; Duret et al. 2000; Surzycki and Belknap 2000; Ganko et al. 2001; Harrison et al. 2001) and fewer single-nucleotide polymorphisms (SNPs) and insertion–deletion (indel) polymorphisms (Koch et al. 2000; Wicks et al. 2001; Maydan et al. 2007). In interspecific comparisons, chromosome centers experience lower rates of translocation and other rearrangements (Stein et al. 2003; Hillier et al. 2007), and the constituent genes have lower rates of substitution at synonymous sites (Cutter and Payseur 2003b) and tend to have greater homology to distant taxa than do genes on the chromosome arms (*C. elegans* Sequencing Consortium 1998; Parkinson, Mitreva, et al. 2004). These patterns might reflect, in part, insertion bias of TEs (Rizzon et al. 2003) and higher mutation rates on chromosome arms (Cutter and Payseur 2003b). Natural selection also likely plays a role here, through differential selection in the gene-rich chromosome centers against TE insertions and other deleterious mutations; yet, the effects of tight linkage may also limit the efficacy of selection in chromosome centers. The chromosome arms and centers do not differ in terms of gene expression divergence in mutation accumulation (MA) lines, but gene expression differences among wild isolates are greater for genes on the chromosome arms (Denver et al. 2005), perhaps reflecting stronger selection acting against transcription changes in genes that reside in chromosome centers. Despite the multitude of correlations between recombination rate and genomic features, the direction of causality is generally unknown.

The rate of chromosomal rearrangement, as measured from genomic comparisons of *C. elegans* and *C. briggsae*, is exceptionally high (Coghlan and Wolfe 2002; Stein et al. 2003). Of the 4,363 rearrangements detected between these species, 31% are inversions, 5.66% translocations, and the remaining 62% being putative intrachromosomal transpositions (Coghlan and Wolfe 2002; Stein et al. 2003). At present, it is unknown how much population polymorphism versus divergence for inversions and other rearrangements occurs, despite the potentially important implications for adaptation and speciation (Hoffmann and Rieseberg 2008). Although intrachromosomal rearrangements in the arms are common, rearrangements between different chromosomes and within autosomal centers are uncommon. In a comparison of the *C. elegans* and *C. briggsae* genome sequences, Hillier et al. (2007) demonstrated that these two nematodes share strikingly concordant chromosomal fidelity of genes. An intriguing possible explanation for the extensive rearrangement within chromosomes is that rates of rearrangement are faster in selfing lineages (Charlesworth 1992; Cutter et al. 2008). It is not yet clear whether the currently described patterns of rearrangement inferred from two selfing taxa are representative of the rate of rearrangement for most species in the genus. Further, the same "arms and centers" organization of gene density and recombination rate seen in *C. elegans* also is evident in *C. briggsae* (Hillier et al. 2007). Despite the lack of defined centromeres in *Caenorhabditis*, the pattern of chromosome arms experiencing higher rates of rearrangement, duplication, and recombination also is seen in vertebrate genomes (Consortium ICGS 2004; Goodstadt et al. 2007), including humans (Nguyen et al. 2006; Ponting 2008). Construction of genetic maps for additional species of *Caenorhabditis* is needed to test the antiquity of the arms and centers pattern of genomic organization and the generality of high rates of rearrangement.

Spatial Clustering of Functionally Related Genes

Caenorhabditis elegans chromosomes can differ dramatically in the relative complement of certain gene families. For example, 44% of tRNA genes are located on the X chromosome (*C. elegans* Sequencing Consortium 1998) and >50% of nuclear hormone receptors occur on chromosome V (Sluder et al. 1999), despite the roughly similar physical size of all chromosomes. Such chromosomal biases might be by-products of the duplication process.

More strikingly, genes in closer proximity show greater similarity of expression patterns than pairs of physically distant genes, although the extent of expression covariation with linkage is much shorter than for *Drosophila* and human (Lercher et al. 2003; Chen and Stein 2006). Nevertheless, coexpression patterns of gene neighbors are generally stronger for *C. elegans* (and yeast) than for other eukaryotes (Fukuoka et al. 2004). Gene neighbors that are oriented 5′–5′ (encoded on opposite strands) have the highest coexpression patterns, followed by gene neighbors oriented in the same direction on the same strand, likely due to shared upstream regulatory elements (Chen and Stein 2006). Gene neighbors in a convergent 3′–3′ orientation do not exhibit stronger patterns of coexpression than randomly selected loci (Chen and Stein 2006).

Moreover, many genes with expression in the same type of tissue cluster physically along chromosomes. Muscle genes form such clusters: 386 of 1,304 genes enriched in muscles during the first larval stage (L1) are located within 10 kb of one another (Roy et al. 2002). Although much of the tendency for neighboring genes to show correlated expression can be explained by the incidence of operons and tandem duplicates (Lercher et al. 2003), the clustering of muscle genes is independent of these effects (Roy et al. 2002). However, clustering of muscle-specific genes is not related to a particular function within the class of muscle genes; clustered genes have a wide range of molecular functions (Roy et al. 2002).

Germ line and sex-regulated genes also are distributed nonrandomly in the *C. elegans* genome (Reinke et al. 2000, 2004). Few germ line-enriched genes are found on the X chromosome, and the near absence of X-linked sperm genes in *C. elegans* contrasts with their disproportionate sex linkage in mice (Wang et al. 2001). A closer look at where sperm genes cluster in the *C. elegans* genome revealed that transcripts expressed during spermatogenesis cluster into three major foci on chromosomes II and IV, at least one of which appears conserved in *C. briggsae* (Ward et al. 1988; Miller et al. 2004). Oocyte-enriched genes do not show the same type of pattern, but nonsperm germ line genes are the primary constituents of operons, and operons do cluster nonrandomly in the genome, with about 38% of operon genes residing in only 12% of the genome (Reinke and Cutter 2009). Furthermore, monocistronic genes that are expressed in the germ line are in closer physical proximity to operons than expected by chance (Reinke and Cutter 2009).

The majority of pathogen defense genes that are induced in *C. elegans* by infection from *Microbacterium nematophilum* were found to cluster along chromosomes IV and V (O'Rourke et al. 2006). The functional clusters differentiate into those that are up- or downregulated after infection, including one 253-kbp interval on chromosome IV that contains 62 genes that are upregulated following infection. It is supposed that pathogen-specific clustering might occur generally within *C. elegans*' genome (O'Rourke et al. 2006). Transcription factor genes also show evidence of spatial clustering along chromosomes in both *C. elegans* and *C. briggsae*, probably a by-product of tandem duplication (Haerty et al. 2008). Furthermore, evolutionarily related genes occur in nonrandom clusters, primarily on chromosome arms, as a consequence of gene duplication in a variety of gene families (Thomas 2006a). In general, the aggregation of similarly expressed genes might result from selection for coexpressed genes to lie in blocks of active chromatin, facilitating access of the molecular machinery necessary for transcription (Hurst et al. 2004).

Although not functionally related to each other, nearly 1,200 genes overlap physically with the annotation for another gene in the *C. elegans* genome (Chen and Stein 2006). About half of these gene pairs are nested (547), with one member residing fully within the bounds of the other's introns, typically being encoded on the opposite strand (490 of the 547). Chen and Stein (2006) also demonstrated that such overlapping genes are more conserved than other genes in the genome, even after controlling for the overrepresentation of overlapping genes in the centers of chromosomes. This greater-than-expected conservation is true for protein sequence divergence (of the flanking gene member) and presence of orthologs, as well as the overlapping gene structure itself, in comparisons with *C. briggsae*. Chen and Stein (2006) propose a "sheltered island hypothesis" for the evolutionary persistence of nested gene structures, in which translocations of internal members might typically be selected against due to detrimental effects on the flanking gene.

Operons

Unlike most eukaryotes, many genes in *C. elegans* and other nematodes form polycistronic transcriptional units or operons (Spieth et al. 1993; Evans et al. 1997; Lee and Sommer 2003; Guiliano and Blaxter 2006). The genes within an operon share one promoter and are transcribed into a single pre-mRNA, which is then spliced to generate single-gene transcripts for translation (Spieth et al. 1993). Approximately 15% of *C. elegans*' 20,000 genes lie within its ~1,150 operons, with each operon containing on average 2.6 genes (range 2–8) and with the genes in very close proximity to their cooperonic neighbors: coding regions are separated by a median distance of about 450 bp, neighboring untranslated regions (UTRs) are typically about 100 bp apart (Blumenthal et al. 2002; Blumenthal and Gleason 2003; Lercher et al. 2003). Operons are more common in the centers of autosomes and are rare on the X chromosome (5% observed vs. 18% expected; Blumenthal et al. 2002).

The members of a given operon do not necessarily have related molecular functions or correlated mRNA expression (Blumenthal and Gleason 2003). However, several functional classes of genes are overrepresented (and

underrepresented) in operons (Blumenthal and Gleason 2003). Although operon genes are nearly twice as likely to have similar expression patterns as nonoperonic pairs of gene neighbors, coexpression is lower among members of an operon that are farther apart (Lercher et al. 2003). Experiments have revealed that more than a quarter of *C. elegans* operons (out of the 172 operons assayed) contain internal promoters for downstream genes, which partially explains the imperfect coexpression of genes within a given operon (Huang et al. 2007). Differences in posttranscriptional stability of spliced transcripts also presumably contribute to the relatively poor correspondence of expression patterns among cooperonic genes.

In a new analysis of operon patterns, we find additionally that the gene position within an operon is negatively correlated with the detected embryonic expression level: each subsequent gene in an operon is expressed at a lower level, on average (Spearman's $\rho = -0.15, P < 0.0001$; fig. 2; operon information extracted from Wormbase WS185, expression measured as the sum of values over early embryonic development reported by Baugh et al. [2003] for 1,706 genes in operons; similar findings by Pires da Silva A, personal communication). Consistent with this, the first gene in operons has significantly stronger codon usage bias (F_{op}) than the rest (analysis of variance [ANOVA] $F_{4,2048} = 13.7$, $P < 0.0001$; fig. 2); codon bias can indicate selection for translational efficiency and/or accuracy and is stronger for more highly expressed genes (see Codon Bias). We also find a subtle reduction in rates of protein evolution (d_N) for the first gene in an operon relative to subsequent genes, although the difference is not significant when adjusted for synonymous site divergence (d_N/d_S) (data not shown).

Operon-encoded genes are represented disproportionately among those genes that result in an RNAi-induced phenotype, particularly for viability effects, demonstrating the especially important functional consequences for operon-encoded gene products (Blumenthal and Gleason 2003; Qian and Zhang 2008). Moreover, we find that 72.5% of the genes in operons are genes with germ line-associated expression profiles in multidimensional gene expression space (fig. 3), so-called germ line "expression mounts" (Kim et al. 2001). "Germ line mounts" 2, 7, 11, 18, and 20 (plus the non-germ line mounts 5 and 23) are the only ones with a significant excess representation of operonic genes (fig. 3; based on 2,015 genes in operons; *G*-tests, all $P < 0.0001$). Expression mounts for neuronal (6 and 13), sperm (4), muscle (1), and five other categories (0, 3, 8, 9, and 10) have an underrepresentation of operonic genes (*G*-tests, all $P \leq 0.0006$), with the remaining categories not differing significantly from the expected composition by operon genes (after Bonferroni multiple test correction). Using a different gene expression data set and in situ hybridization patterns, Reinke and Cutter (2009) showed that expression in germ line tissue among operon-encoded genes is ubiquitous, proposing that low promoter complexity and regulation by 3′ UTRs (Merritt et al. 2008) might be responsible for the evolution of this association.

Among different species of *Caenorhabditis*, most operon gene structures are conserved (Stein et al. 2003; Qian and Zhang 2008). Only about 4% of *C. elegans* operons are disrupted relative to *C. briggsae*, whereas 60% are expected to be disrupted based on the degree of rearrangement among nonoperonic genes (Stein et al. 2003). The genes harbored by *C. elegans* operons also tend to be conserved

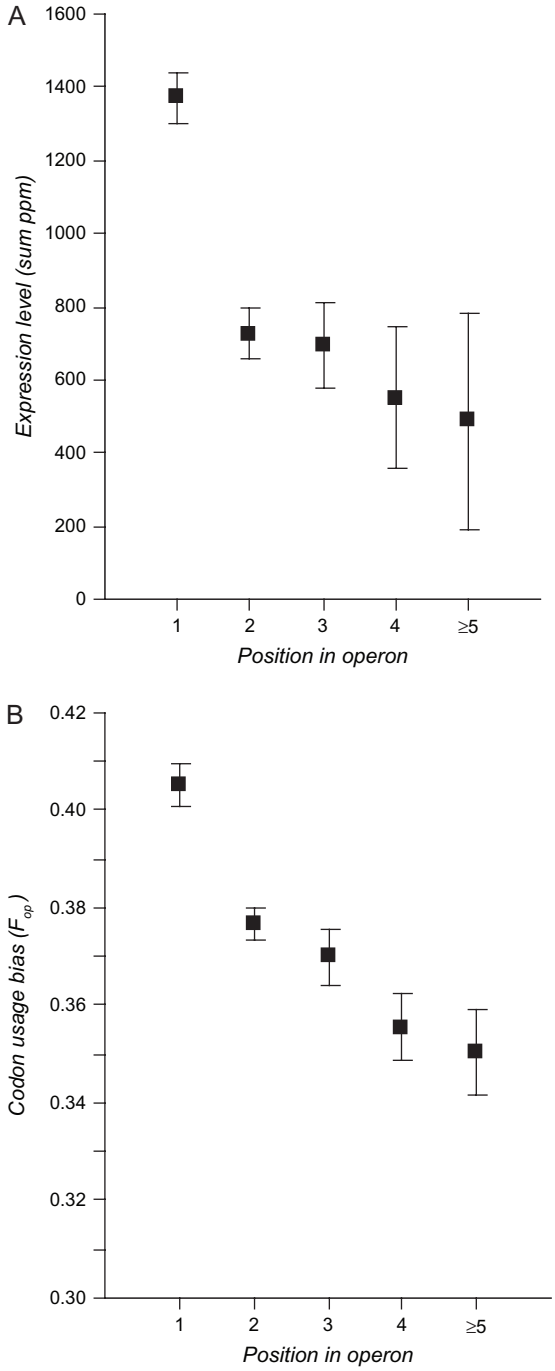

FIG. 2.—Average expression level (*A*) and codon bias (*B*) for genes occurring at different positions within operons. In (*A*), expression level is the summed expression values across early embryogenesis from Baugh et al. (2003). For both (*A*) and (*B*), only the first gene in operons differed significantly from other positions in Tukey's honestly significant difference post hoc tests (expression ANOVA $F_{4,1701} = 14.4$, $P < 0.0001$; codon bias ANOVA $F_{4,2048} = 13.7$, $P < 0.0001$). Bars indicate ±1 standard error of the mean.

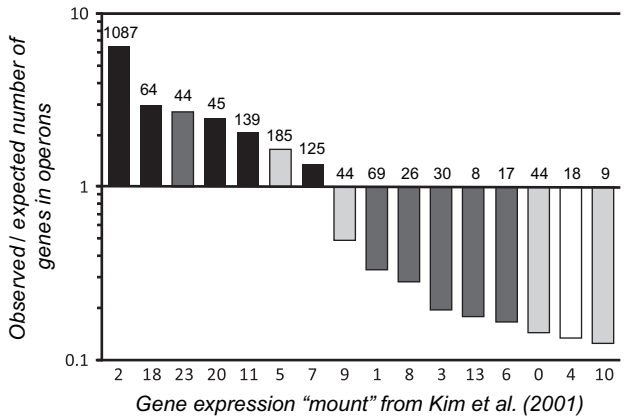

FIG. 3.—Gene coexpression categories ("expression mounts," Kim et al. 2001) with significant over- or underrepresentation of genes in operons. Numbers above bars indicate the number of operonic genes in that expression category. Categories with black bars are enriched for germline/oocyte expression; white bar enriched for spermatogenesis-related genes; dark gray bars are enriched for somatic gene expression or other molecular functions (e.g., muscle, neuronal); light gray bars are mounts with no gene enrichment classification (Kim et al. 2001). All Bonferroni multiple test corrected P values for G-tests ≤ 0.022; categories other than mounts 10 and 13 have corrected $P \leq 0.00001$. The other 20 gene expression mounts not shown in this figure did not have gene complements that deviated from the expected representation in operons, assuming random assortment of genes into operons.

across greater phylogenetic distances than other genes (Blumenthal and Gleason 2003). Despite the impressive maintenance of operon structures, it is unknown whether their origin in nematodes might have been adaptive. However, both adaptive (Blumenthal and Gleason 2003) and selfish (Lawrence 1999) explanations have been proposed for the origin of the *trans*-splice machinery that enables the processing of nematode operons. It also is not clear whether recruitment of genes into operons is subject to positive selection or whether recruitment occurs by genetic drift as a consequence of relaxed selection on the maintenance of individual promoters (given the presence of *trans*-splicing capability; Lawrence 1999), after which operon-associated genes become "trapped" in that state (Blumenthal and Gleason 2003; Blumenthal 2004; Qian and Zhang 2008). It has been supposed that recruitment into and maintenance of nematode operons might stem from selection for regulatory efficiencies or small genome size (Blumenthal and Gleason 2003). A challenge is to develop nonadaptive null models (Lynch 2007): Could operons accumulate nonadaptively those genes that tend to evolve slowly, have important functional effects, and are primarily associated with germ line expression? It is conceivable that limited regulatory specificity of germ line genes might predispose their capture by operons, followed by purifying selection against operon disruption (Reinke and Cutter 2009), but a more formal analysis of this problem is needed to determine the plausibility of a neutral process of operon formation.

Using a comparative approach, it has been proposed recently that operon gains occur more than three times as often as operon losses and that *Caenorhabditis* species continue to accumulate operons in their genomes (Qian and Zhang 2008). Since the common ancestor of *C. elegans* and its sister clade, an average of 16.75 operon losses have occurred per lineage per unit time (Qian and Zhang 2008). However, both operon losses along the *C. elegans* lineage and gains in lineages leading to other *Caenorhabditis* species were undetectable, and the inference of *C. elegans* gains depended on very distant outgroups (*Pristionchus pacificus* and *Brugia malayi*). These limitations make it difficult to confidently describe relative rates of operon gain and loss based on present data. An alternative scenario is that the operon complement in *Caenorhabditis* species might be relatively stable, with little net change in operon abundance over time. A model of this scenario, analogous to a standard mutation balance population genetic model, gives the equilibrium fraction of operonic genes in a genome, ω, as simply equal to the rate of operon gene gain (v) divided by the total rate of operon gain and loss of genes ($v + u$): $\omega = v/(v + u)$. Assuming equilibrium operon abundance in all *Caenorhabditis* species at 15% in the present day (Blumenthal et al. 2002), this simple model suggests that the rate of operon gene losses exceeds gains by a factor of 5.7 (qualitatively opposite to the conclusion of Qian and Zhang [2008]). The seeming excess of multiple independent operon losses in *Caenorhabditis* reported by Qian and Zhang (2008) could be explained by such a higher relative rate of loss. Clearly, additional work is needed to help clarify the evolutionary dynamics of operon gain and loss in nematodes. Empirical determination of operons in species of *Caenorhabditis* other than *C. elegans* will provide important data for addressing this issue. Further genome-scale work on operons in other eukaryotes, such as for the 20% of genes in *Ciona intestinalis* that reside in operons (Satou et al. 2008), also is needed to determine the broader generality of nematode operon evolutionary patterns and processes.

Peculiarities of the X

The X chromosome differs from autosomes in many salient features. In addition to the X chromosome being characterized by more uniform distributions of genomic features, operons and sperm-related genes are rare on the X chromosome (Reinke et al. 2000, 2004; Blumenthal et al. 2002). TEs also occur slightly more often on the X chromosome (Duret et al. 2000; Ganko et al. 2001), and codon usage bias is slightly stronger on the X (Singh et al. 2005), but these effects are weak. The X chromosome has an underrepresentation of pseudogenes and gene duplicates, as do chromosomes I and II (Harrison et al. 2001; Cavalcanti et al. 2003). Comparisons of synteny between *C. elegans* and *C. briggsae* revealed that the X chromosome has longer syntenic blocks than do autosomes (Stein et al. 2003; Hillier et al. 2007). It does not seem likely that the unusual characteristics of the sex chromosome reflect population genetic processes associated with X hemizygosity in males or reduced effective population size, even as a relic from the gonochoristic (male and female) ancestor of *C. elegans*. Instead, cell biological phenomena, such as X-limited DNA compaction and transcriptional quiescence associated with meiosis (Kelly et al. 2002), likely provide more plausible drivers of the peculiarities of the X chromosome.

Duplication

Gene duplication provides an important source of novel genetic material on which selection can act. Indeed, gene duplication is a major evolutionary force in *C. elegans*, for which about one-third (32%, >6,100 genes) of its total gene complement is made up of duplicate, nonsingleton genes (Friedman and Hughes 2001; Gu et al. 2002; Cavalcanti et al. 2003). Ohno (1970) theorized that in a scenario where mutation replicates a single gene into two copies, one gene duplicate will experience relaxed selection and will accumulate mutations. Purifying selection for the ancestral function will consequently preclude the accumulation of deleterious mutations from the other gene duplicate. By this mechanism, the evolutionary fate of most gene duplicates is thought to be degeneration and "nonfunctionalization" through pseudogene formation (Lynch and Conery 2000). Occasionally, however, the freely evolving gene copy might acquire a new, beneficial function that allows its persistence: the process of "neofunctionalization" from gene duplication (Ohno 1970). Subsequent theory has emphasized a "subfunctionalization" process in which different portions of each of the duplicate pairs change in a complementary way such that together the ancestral function is preserved or enhanced through specialization of each copy (Lynch and Force 2000). Woollard (2005) reviews many aspects of gene duplication for *C. elegans* and Schwarz (2005) summarizes the composition of *C. elegans* gene families. Here, we detail some recent advances in our understanding of gene duplication in the light of genome evolution in *C. elegans*.

Table 1
Summary of Gene Duplicate Properties for Gene Pairs That Duplicated in the Recent Past (Katju and Lynch 2003; Katju V, personal communication)

Gene Duplicate Category	"New" Duplicates	"Recent" Duplicates
Divergence cohort	$K_s = 0$	$0 < K_s \leq 0.1$
Number of duplicates	55	125
Median distance between intrachromosomal duplicates (bp)	1,138	8,644
Median duplication length (bp)	1,710	1,329
Percent of all duplicates:		
Intrachromosomal duplicates	89	56
Chimeric and partial duplicates	50	64
Percent of intrachromosomal duplicates:		
Tandem duplicates	69	26
Inverse orientation	69	53
Inverted-tandem duplicates	45	13

Gene Duplication Patterns and Origins

A simple conception of the gene duplication process is of replication of complete genes with subsequent sequence divergence (Ohno 1970). However, partial duplication, inverted duplication, and multiple gene duplications also appear to be evolutionarily important phenomena in the worm genome (Katju and Lynch 2006). In *C. elegans*, short-sequence duplications occur more frequently than long sequence duplications (Katju and Lynch 2003). The median unspliced gene length in *C. elegans* is ~1.7 kb, whereas relatively recent duplication events have a median length of only ~1.4 kb (Katju and Lynch 2003). Consequently, partial gene duplications are more common than whole or multiple gene duplication events. Further, partial and chimeric duplicates (one or both copies contain novel exons and/or introns) are common both among very recent and older duplicate copies (50% and 64%, respectively; table 1), underscoring the evolutionary potential of genetic novelty arising from gene duplication (Katju and Lynch 2003).

Most (89%) of recently arisen gene duplicates in *C. elegans* occur intrachromosomally, whereas only about half of older duplicates occur on the same chromosome (Semple and Wolfe 1999; Katju and Lynch 2003). Nevertheless, the *C. elegans* genome has fewer interchromosomal duplicates than would be expected from random placement of duplicate copies (Semple and Wolfe 1999). Duplicate copies within a chromosome also tend to be close together, on average less than 8 kb apart (Semple and Wolfe 1999), and this is responsible for much of the covariation in gene expression among genes in close physical proximity (Lercher et al. 2003). Large multigene duplications (block or segmental duplication events) are relatively rare but occur intrachromosomally proportionately more often than do single-gene duplicates (Cavalcanti et al. 2003). Intrachromosomal duplicate gene pairs are prevalent on all chromosomes, but the X and the physically smaller chromosomes I and III have a lower abundance and density of duplicates than the other chromosomes (Semple and Wolfe 1999; Cavalcanti et al. 2003). This finding accords with the notion that duplication is a contributor to increased genome size (Coissac et al. 1997). Notably, chromosome V (the physically largest chromosome) has a particularly abundant collection of duplicates that comprise 49% of its genes (Cavalcanti et al. 2003), including a diversity of nuclear hormone receptor and putative chemoreceptor gene families (Sluder et al. 1999; Robertson 2000). *Caenorhabditis elegans* chromosome arms also harbor a greater density of gene duplicates than do the central gene-dense, low recombination portions of chromosomes (*C. elegans Sequencing Consortium 1998*; Rubin et al. 2000).

Within a chromosome, gene duplicates in *C. elegans* commonly occur in tandem (table 1), possibly as a result of slippage or unequal crossing-over (Waterston and Sulston 1995; Katju and Lynch 2003). In terms of orientation, intrachromosomal duplicates overall are observed to have the same orientation in 61% of cases (Semple and Wolfe 1999). However, only 42% of recent intrachromosomal duplicates (including those in tandem) occur in the same orientation, and only one-third of very recent duplicates (those with $K_s = 0$) lie in the same orientation, implying that the molecular mechanism of duplication frequently generates inverted sequences and that inverted copies are lost disproportionately over time (Katju and Lynch 2003). If selection against inverted duplicates is a consequence of the formation of inversion polymorphisms in the population (Fischer et al. 2001), then we might see a stronger signature of selection against inverted duplicates in outcrossing species

than in selfing species because inversions will be less likely to have negative fitness consequences in highly homozygous selfing populations. Genomic analysis of duplicates in obligate outbreeding species can test for an effect of breeding system on gene duplication patterns.

In contrast to yeast (Kellis et al. 2004), *Arabidopsis* (Ermolaeva et al. 2003), and puffer fish (Jaillon et al. 2004), *C. elegans* does not appear to have experienced a whole-genome duplication event in its recent evolutionary history, as indicated by the low abundance of block duplication events (7.1% of duplicated regions include $\geq$2 genes) (Friedman and Hughes 2001; Cavalcanti et al. 2003). Such block duplications make up 2.3% of the worm genome, comparable to the 1–5% genomic fraction of block duplications in mammals (Tuzun et al. 2004), although duplications in mammals typically are longer than in worms. In contrast, 25% of all gene duplications in the yeast genome occur in duplicated blocks that share the same order and orientation of genes (Cavalcanti et al. 2003) and are hypothesized to have originated simultaneously in a whole-genome duplication event (Wolfe and Shields 1997; Seoighe and Wolfe 1999). Further evidence against whole-genome duplication in worms comes from the pattern of inter- versus intrachromosome duplication in comparison to yeast: 90% of both single-gene duplications and block duplications in yeast are interchromosomal. By contrast, in *C. elegans*, single-gene duplications are present equally inter- and intrachromosomally, whereas 71% of partial and 80% of block duplication events are intrachromosomal (Lynch and Conery 2000; Cavalcanti et al. 2003).

Gene duplicates in *C. elegans* arise at a rate at least 10-fold higher than observed for *Drosophila* and yeast (0.002 and 0.008 duplications per gene per My, respectively) (Lynch and Conery 2000). Updating the calculations of Lynch and Conery (2000) with current estimates of the point mutation rate in *C. elegans* (Denver, Morris, Lynch, and Thomas 2004; Keightley and Charlesworth 2005), the expected per-genome duplication rate is $\sim 2.95 \times 10^{-4}$ per generation and the half-life of duplicates is $\sim 5.5 \times 10^{6}$ generations. Depending on the generation time of *C. elegans* in nature, this translates to hundreds or thousands of duplication events per genome per My.

Davis and Petrov (2004) provide evidence that gene duplicates tend to derive from a pool of genes that generally are subject to strong selective constraint. However, several notable examples of constrained classes of genes run counter to this trend. Gene duplicates are underrepresented among the genes that are expressed in early embryogenesis, possibly reflecting selection against duplication of early-expressed genes (Castillo-Davis and Hartl 2002). Transcription factors (excepting nuclear hormone receptors) also appear to have a lower-than-typical incidence of duplication, with most transcription factors having one-to-one orthologs in *C. briggsae* (Reece-Hoyes et al. 2007), despite a tendency for more rapid protein evolution in this class of genes (Castillo-Davis, Kondrashov, et al. 2004; Cutter and Ward 2005; Haerty et al. 2008). Also running counter to the pattern reported by Davis and Petrov (2004), genes in operons are duplicated less frequently than monocistronic genes (Lercher et al. 2003; Cavalcanti et al. 2006), presumably because the generally short tracts of duplicated sequence disrupt the structure of operons or adjacent genes (Cavalcanti et al. 2006). However, operonic genes with associated internal promoters have a similar incidence of gene duplicates as do nonoperonic genes (Huang et al. 2007).

Evolutionary Fates of Duplicates

Duplicated genes have been proposed to provide one of the causes of genetic redundancy. In *C. elegans*, genes with a duplicate copy are less than half as likely as single-copy genes to yield a phenotypic effect when knocked down by RNAi (Kamath et al. 2003). However, much debate exists about whether natural selection mediates such redundancy (Lynch and Conery 2000; Kondrashov et al. 2002; Wagner 2002), and it has been argued that this could result from many duplicates actually representing pseudogenes (Reece-Hoyes et al. 2007). One study concluded that duplicates contribute relatively little to redundancy, although moreso in *C. elegans* than most other organisms examined (Hannay et al. 2008). By using combinatorial RNAi (RNAi targeting pairwise combinations of genes) in *C. elegans*, Tischler et al. (2006) identified 16 functionally redundant pairs of duplicated genes out of 143 pairs examined. In this study, redundancy was inferred from the observation of more extreme phenotypes for paired RNAi knockdown than seen for individual gene RNAi. Because Tischler et al. (2006) focused on *C. elegans* genes that had single orthologs in *Drosophila* and *Saccharomyces*, they inferred that 14 out of the 16 redundant pairs of paralogous genes had duplicated prior to the divergence of *C. elegans* from *C. briggsae*. Thus, strong purifying selection appears to have maintained these redundant gene copies over an extended period of time. Pairs of redundant duplicate genes are more similar to each other in their amino acid sequences and also exhibit a lower rate of nonsynonymous site substitutions ($d_N = 0.34$) than the nonredundant duplicates in their sample ($d_N = 0.5$), and this cannot be explained by more recent duplication among the redundant class of duplicates (because d_S did not differ between the two classes; Tischler et al. 2006).

Purifying selection on the paralogs that are produced by duplication is common in eukaryotes, as seen by ratios less than 1 being typical for the rates of nonsynonymous- (d_N) to synonymous site substitutions (d_S) between duplicate copies (Kondrashov et al. 2002). However, d_N/d_S ratios are higher among duplicate genes in both the *C. elegans* and *C. briggsae* genomes, compared with single-copy orthologs between the two species (Castillo-Davis, Hartl, and Achaz 2004). This implies that duplicate genes generally experience weaker purifying selection following speciation, that is, relaxed selection or positive selection or both (Castillo-Davis, Hartl, and Achaz 2004).

The most common fate for duplicated genes is silencing and eventual loss of one member of the pair. However, loss of different copies of a duplicated gene in separated populations, "divergent resolution," could facilitate reproductive isolation (Lynch and Conery 2000). The relative contribution of this process to the evolution of genetic incompatibilities between two populations and subsequent speciation depends on the rates of duplication and,

similarly, of microchromosomal rearrangements (Lynch 2002a). Given particularly rapid rates of duplication and rearrangement evident in *Caenorhabditis* (Lynch and Conery 2000; Coghlan and Wolfe 2002; Stein et al. 2003), this process might play a more important role in the genetic basis of reproductive isolation in worms than in other taxa.

Gene Conversion

The selectively neutral process of gene conversion acts to homogenize sequences, with important implications for patterns of allelic polymorphism and divergence among duplicated segments of DNA. The first report of allelic gene conversion in *C. elegans* comes from the *unc-22* gene (Moerman and Baillie 1979). Point estimates for *unc-29* and *unc-15* suggest that recombination resolves as crossovers roughly three times as often as conversion (Rose and Baillie 1980; Rattray and Rose 1988). Gene conversion appears to be common relative to crossing-over within loci in *Caenorhabditis remanei*, according to patterns of population polymorphisms (Cutter 2008b). Based on inference from intrachromosomal gene duplicates across the *C. elegans* genome, Semple and Wolfe (1999) found that gene conversion occurs more readily between gene duplicates that are closer together, that have low sequence divergence, and that occur in the same orientation. However, only about 2% of gene duplicates had evidence of gene conversion, mostly among members of multigene families (Semple and Wolfe 1999). Specific cases of gene conversion among the clustered members of multigene families are consistent with these general findings (Nikolaidis and Nei 2004; Thomas 2006b). Among all duplicates, intrachromosomal gene conversion is evident about 3.7 times more often than for interchromosomal events, similar to yeast (Semple and Wolfe 1999). The distribution of gene conversion tract lengths inferred from gene duplicates appears roughly log normal, with a mean and median of 117 bp and 58, respectively (Semple and Wolfe 1999). Spontaneous gene conversion of a laboratory mutant by a gene family member resulted in conversion tracts of 32–145 bp in length (Katju et al. 2008). Using a transgenic transposon system, gene conversion tracts at least 191 bp long were reported by Plasterk and Groenen (1992), but it is not known whether this is typical of gene conversion tract lengths between alleles. In a different transgene insertion system that uses Mos1 transposons, gene conversion of up to 9 kb has been demonstrated (Frokjaer-Jensen et al. 2008) with tracts <1 kb in length being more typical (Robert and Bessereau 2007). Over the long term in populations, biased gene conversion can mimic natural selection, typically by preferentially using the strand with guanine or cytosine variants as the template (Birdsell 2002; Marais 2003). The prevalence of such a biased gene conversion process has not been detected in *Caenorhabditis* (Cutter and Charlesworth 2006; Cutter 2008b), although more work is needed to make a definitive determination. In any case, biased gene conversion should be a negligible evolutionary force in highly selfing lineages due to the rarity of heterozygotes (Marais, Charlesworth, and Wright 2004).

Pseudogenes

The *C. elegans* genome contains over 2,000 pseudogenes, although roughly a quarter of these likely correspond to defunct TEs (Harrison et al. 2001). A functional assay of 364 genes suggests the possibility of an even greater incidence of pseudogenes among *C. elegans* genes (20%), particularly among recently duplicated gene copies (Mounsey et al. 2002). Unlike for the human genome (Dunham et al. 1999), few of *C. elegans*' pseudogenes arose through reverse transcription. The chromosome arms contain a disproportionate representation of pseudogenes, and chromosome IV appears to have a greater load of pseudogenes than other chromosomes, containing double the density of pseudogenes as chromosomes I, II, and X (Harrison et al. 2001).

Large, multigene families contain many pseudogenes, such as several of the seven-transmembrane chemoreceptor families and the major sperm protein family (Ward et al. 1988; Robertson 1998, 2000; Harrison et al. 2001). However, most pseudogenes correspond to very small gene families (<10 members), and about 10% of pseudogenes are associated with singleton functional genes (Harrison et al. 2001).

Nearly half of pseudogenes appear to be rendered nonfunctional by virtue of only a single frameshift or premature stop mutation (Harrison et al. 2001). This result has been corroborated in the *srh* and *str* gene families and, furthermore, shown that putatively functional allelic copies exist in some wild strains (Stewart et al. 2005). Two interpretations of these patterns seem plausible. First, recent relaxed selection in the *C. elegans* lineage might have permitted the accumulation of deleterious mutations, resulting in a "young" observed distribution of pseudogenes in the *C. elegans* genome, sufficiently young that not all pseudogene-inducing mutations have fixed yet by genetic drift. Second, multigene families might facilitate the population segregation of defective copies, provided that other functional copies can compensate for their presence in a given individual. Determination of whether *C. elegans*' pseudogene complement is typical for species in this genus will help disentangle whether processes peculiar to the *C. elegans* lineage might be responsible for its distribution and abundance of pseudogenes.

Intron Evolution

Intron Gain/Loss

Understanding the widespread proliferation of introns in the eukaryotic lineage still remains mostly unresolved. Eukaryotic genomes vary by more than three orders of magnitude in their intron density, indicating that intron gain and/or intron loss has been extensive during evolution (Jeffares et al. 2006). Nematode genes in particular have a high rate of intron turnover in comparison to other animals (Logsdon et al. 1995), such as mammals (Roy et al. 2003; Coulombe-Huntington and Majewski 2007a) and *Drosophila* (Coulombe-Huntington and Majewski 2007b). Intron gains or losses are estimated to have occurred at a rate of at least 0.005 per gene per My in nematodes, which far exceeds that of chordates (Stein et al. 2003). Comparing the whole

C. elegans genome to 8% of the *C. briggsae* genome, Kent and Zahler (2000) found 250 unique introns present in *C. elegans* that were absent in *C. briggsae*. In a larger study, out of 12,155 orthologous gene pairs in whole genomes of *C. elegans* and *C. briggsae* studied by Stein et al. (2003), they found 4,379 *C. elegans*-specific introns and 2,200 *C. briggsae*-specific introns with the remaining ~54,000 introns conserved between orthologous gene pairs.

Recent evolutionary periods appear to show greater prevalence of intron losses than gains in eukaryotes (Rogozin et al. 2003). However, different mechanisms may govern most intron loss in *C. elegans* relative to other eukaryotic lineages (Roy and Gilbert 2005). *Caenorhabditis elegans* exhibits 1) no bias toward the loss of 3′-most introns, contrary to predictions for a mechanism of gene conversion with a reverse-transcribed product of spliced mRNA; 2) no disproportionate loss of adjacent introns; and 3) no bias toward phase-zero intron loss (i.e., introns between two complete codons). These observations suggest that events like nonhomologous recombination and spontaneous genomic deletions might be more important in the excision of introns in the lineage leading to *C. elegans* than is homologous recombination, which is the predominant mechanism mediating intron loss in other taxa (Cho et al. 2004; Roy and Gilbert 2005). However, it is not clear whether the extended conserved 3′ splice site sequence in *C. elegans* (5′-GU-UUUUCAG-3′; Blumenthal and Steward 1997) might have an influence on the evolution of intron gain and loss. The highly conserved UUUUC preceding AG provides important information to the splice site machinery, possibly replacing the 15–20 nt polypyrimidine tract that is typically found in vertebrates (Blumenthal and Steward 1997).

Kiontke et al. (2004) observed at least 27 intron losses and at most 3 gains for the 17 different intron-occupied sites in RNA polymerase II (RNAP2) for *Caenorhabditis* and several outgroup species. Mapping the introns in a phylogenetic framework allowed them to assign intron evolution events to specific lineages, which indicated that most of the intron loss in RNAP2 occurred prior to the origin of the Elegans group within *Caenorhabditis* (Kiontke et al. 2004). Globins also appear to experience an excess of intron loss in *Caenorhabditis*, which contrasts globin intron evolution in other nematodes (Hoogewijs et al. 2008). Cho et al. (2004) reported that individual introns had roughly a 10% chance of being lost compared with the 400-fold lower 0.025% in humans and mice. This discrepancy might reflect differences in the average size of the introns between worms and mammals, if shorter introns are generally lost more readily than long introns (Cho et al. 2004).

Evidence for intron gains comes from early studies with the large seven-transmembrane (*str*) chemoreceptor family in *C. elegans* (Robertson 1998, 2000, 2001). For the genome overall, Coghlan and Wolfe (2004) subsequently inferred 81 cases of intron gain in *C. elegans* and 41 in *C. briggsae*. However, reanalyzing their data with additional *Caenorhabditis* species (*C. remanei* and *Caenorhabditis brenneri*), Roy and Penny (2006) reported that 74% of the introns previously thought to be gains in *C. elegans* were also present in one or both of *C. remanei* and *C. brenneri*, implying that these actually represent intron losses in *C. briggsae*. Similarly, among the 41 reported gains in *C. briggsae*, 61% are present in *C. brenneri* and thus probably represent losses in the *C. elegans* lineage. This reassessment of the incidence of intron gain underscores the need for inclusion of many species in comparative analyses—recent and forthcoming *Caenorhabditis* genome sequences will facilitate such work. Further phylogenetically informed studies of intron gain and loss in *Caenorhabditis* might test Lynch's hypothesis that small effective population size facilitates the colonization of genes by new introns (Lynch 2002b) because selfing and outcrossing species differ in effective size by more than an order of magnitude (Graustein et al. 2002; Cutter, Baird, and Charlesworth 2006).

For paralogous gene duplicates, intron gain and loss have been studied in a variety of eukaryotic lineages (Castillo-Davis, Bedford, and Hartl 2004), with an excess of intron gains being more prevalent in some taxa (Babenko et al. 2004; Knowles and McLysaght 2006) and an excess of losses in others (Coulombe-Huntington and Majewski 2007b). A general genomic understanding of intron gain/loss dynamics for gene duplicates awaits study in *C. elegans*, despite research having targeted specific gene families. Gains and losses of introns in the cytochrome P450 (CYP) family appear to occur at similar rates (Gotoh 1998), whereas intron losses prevail over intron gains in the two large families of *str* (seven-transmembrane receptor) and *stl* (*str*-like) chemoreceptors in *C. elegans* (Robertson 1998, 2000). In contrast to the *str* and *stl* families, where one intron gain was noted, seven intron gains were inferred for the *srh* family (Robertson 2001). Katju and Lynch (2006) also identified examples of both loss and gain in their comparisons of sequences of a few sets of genes.

Irimia et al. (2008) found a high level (92.4%) of evolutionary conservation of alternately spliced exons between *C. elegans* and its close relatives, in contrast to other lineages like dipterans and mammals. This implies that the vast majority of alternatively spliced exons in *C. elegans* were present in its common ancestor with *C. briggsae* and *C. remanei* and have been preserved during nematode evolution (Rukov et al. 2007; Irimia et al. 2008). This high conservation of alternatively spliced exons, however, suggests a limited role for alternative splicing in creating transcriptome and proteome diversity in *Caenorhabditis*. Gene duplication seems to play a more prominent role in generating transcript novelty in this group of organisms (Rukov et al. 2007; Irimia et al. 2008). Evolutionary studies on alternate splicing patterns could be useful in elucidating the mechanisms for such a difference.

Intron Size and Position

The *C. elegans* genome contains 4.0 introns per kilobase pair of coding sequences (Deutsch and Long 1999), with introns tending to be short and with their lengths tightly distributed around a mode of 47 bp (>50% of introns are <60 bp long) (Blumenthal and Steward 1997; Fedorov et al. 2003; Cho et al. 2004). For at least some genes, introns tend to be shorter in the monophyletic group containing *C. briggsae*, *C. remanei*, and *C. brenneri* than in

outgroups *C. elegans*, *Caenorhabditis japonica*, and *Caenorhabditis* sp. 3 (PS1010; Cho et al. 2004), and genome-wide median intron length is about 20% shorter in *C. briggsae* than in *C. elegans* (Stein et al. 2003). Intron size is correlated positively with recombination rate across the *C. elegans* genome, in contrast to the negative correlation between size and local recombination rate seen in *Drosophila* and human genomes (Prachumwat et al. 2004). Two proposed reasons for the *C. elegans* pattern are 1) that transposons insert more readily in the introns of high recombination regions, resulting in longer intron sizes due to accumulated transposons (Duret et al. 2000; Rizzon et al. 2003) and 2) stronger selection against excessive noncoding sequence in the gene-dense chromosome centers that have low rates of recombination (Prachumwat et al. 2004). Another possible explanation for introns being longer on chromosome arms is heterogeneity in the indel mutation rate along the length of chromosomes, which might also accord with higher indel polymorphism on the arms (Maydan et al. 2007). Genes expressed at higher levels also tend to have shorter introns in *C. elegans*, presumably reflecting natural selection driving the minimization of the cost of transcription of highly expressed genes (Castillo-Davis et al. 2002).

The spacing of introns within genes in *C. elegans* is more uniform than expected from a random incidence of introns within genes (Lynch and Kewalramani 2003). In addition, the first few introns in a gene are closer to the 5′ end of the gene on average, relative to the null expectation (Lynch and Kewalramani 2003). The nonrandom spacing of introns might result from selective forces favoring nonsense mediated decay of truncated transcripts, intron-associated gene regulation, and mRNA export (Lynch and Kewalramani 2003).

Mutation Rates

Mutation is one of the fundamental forces of evolution; yet, there are few eukaryote species for which mutation rates have been estimated directly. As the original source for genetic novelty, mutation rate is a basic component of population genetic and molecular evolutionary models; critical factors related to divergence and polymorphism depend on accurate estimation of this parameter. Furthermore, mutational rates and the accompanying distribution of mutational effects, dominance levels, and heterogeneity across the genome are all expected to play important roles in the evolution of sex and recombination, population persistence, inbreeding depression, standing genetic variation, and adaptation.

Caenorhabditis elegans research on mutational processes has been at the forefront of advances made in the characterization of mutation rates (Drake et al. 1998; Lynch et al. 1999; Baer et al. 2007). Inspired by the classic work of Mukai (1964) in *Drosophila*, use of mutation accumulation (MA) lines in *C. elegans* has enabled researchers to estimate the genomic deleterious mutation rate (U) through phenotypic assays of fitness decline (e.g., Keightley and Caballero 1997; Vassilieva and Lynch 1999). In MA experiments, a single hermaphrodite is used to propagate each new generation under benign conditions, such that all new mutations with a selection coefficient $s < 0.25$ can accumulate by genetic drift. These studies suggested that *C. elegans* genome acquires only about $U = 0.002$–0.03 deleterious mutations per generation, a figure that was much lower than seen in *Drosophila* and mammals (Baer et al. 2007), and which turns out to be greatly underestimated. Using this same kind of mutation-induced phenotypic analysis, Davies et al. (1999) assessed the detectable fitness impact of a known quantity of mutations introduced by chemical mutagenesis, showing that the classical fitness assays underestimate the genomic deleterious mutation rate by a factor of about 30. Although the fitness effects of fewer than 5% of new mutations are detectable in traditional phenotypic fitness assays, mutational effects may be more pronounced in alternative environments (Denver, Morris, Lynch, and Thomas 2004)—but there is little in the way of direct tests of this possibility in *C. elegans*. The fitness effect of most TE insertions also is very weak (Begin and Schoen 2006), and MA experiments with mismatch repair (MMR) deficient strains also suggest that many mutations have individually very small effects (Estes et al. 2004). Consistent with this difficulty in detecting most deleterious mutations, a population polymorphism-based estimate of the distribution of deleterious mutation effects suggests that most deleterious mutations are of very weak effect (Loewe and Cutter 2008).

In addition to life history and fitness-related traits, phenotypic variation resulting from MA also has been explored for morphological and behavioral traits. Ajie et al. (2005) found that the mutation rate and effect size for behavioral traits are similar to that estimated for fitness. Newly arising mutations tend to act pleiotropically and negatively on multiple traits, in a manner inconsistent with life history tradeoffs due to antagonistic pleiotropy (Keightley and Bataillon 2000; Estes et al. 2005; Begin and Schoen 2007; Ostrow et al. 2007). MA lines of *C. elegans* also have more variation in body length than wild-type worms (Azevedo et al. 2002), consistent with stabilizing selection on body size in nature, and Ostrow et al. (2007) quantified this variation and compared it with the related *C. briggsae* and found that *C. briggsae* MA strains declined in body size twice as quickly as *C. elegans* strains. Using gene expression as a phenotype, Denver et al. (2005) showed that MA lines have substantially greater variance in expression level than do wild strains, demonstrating the strong force of stabilizing selection on gene expression. Temperature stress does not appear to induce differential mutation rates in *C. elegans* (Baer et al. 2006), in contrast to the apparent condition dependence of mutation in *Drosophila melanogaster* (Agrawal and Wang 2008). However, it will be important to quantify whether other kinds of environmental stress might influence mutation rates and the ability to detect the effects of mutations; the use of competitive assays of fitness rather than fecundity measures also might prove to be illuminating.

MA lines show rapid fitness recovery following population size expansion, apparently due to compensatory mutations (Estes and Lynch 2003). A question yet to be answered is what mutational mechanisms resulted in most of the fitness decline and recovery in these MA lines. For example, stepwise mutations of microsatellites

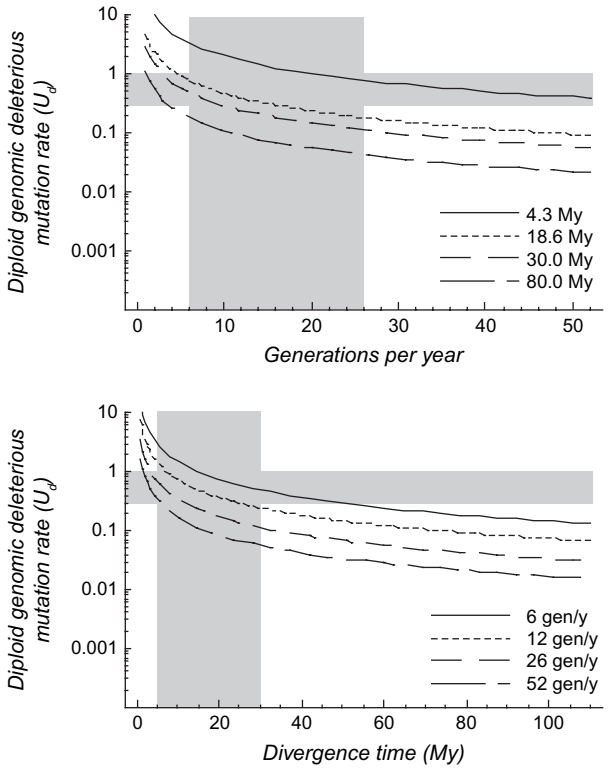

FIG. 4.—Genomic deleterious mutation rate in *Caenorhabditis elegans* as a function of generation time and divergence time with *Caenorhabditis briggsae*. The overlapping regions highlighted in gray indicate the most plausible ranges of parameter space, from Denver, Morris, Lynch, and Thomas (2004) and Cutter (2008a). For details on calculation of U, see Cutter and Payseur (2003a; parameters revised to: 20,000 genes, 1,140 average coding sequence length, 0.12 average K_a, 2.03 average δ_s).

or insertion/excision by TEs might permit relatively rapid reversion, whereas point mutations or insertion/deletion mutations would more likely require specific compensatory changes.

Through DNA sequencing of *C. elegans* MA lines, the nuclear mutation rate for nucleotides (point mutations and small indels) is estimated to be 2.1×10^{-8} mutations per site per generation (Denver, Morris, Lynch, and Thomas 2004), which is an order of magnitude lower than the mutation rate estimate for the mitochondrial genome (1.6×10^{-7} mutations per site per generation; Denver et al. 2000). For many applications, the point mutation rate is most relevant: 9.0×10^{-9} and 9.7×10^{-8} per generation for nuclear and mitochondrial genomes, respectively (Denver et al. 2000; Denver, Morris, Lynch, and Thomas 2004; Keightley and Charlesworth 2005). *Drosophila melanogaster* has a similar difference in relative mutation rates of mitochondrial and nuclear genomes, though the absolute rates are somewhat lower than for *C. elegans* (Haag-Liautard et al. 2007, 2008). Through direct mutation detection by sequencing, Denver, Morris, Lynch, and Thomas (2004) confirmed that phenotypic fitness assays underestimated U by approximately 30-fold, implying a diploid genomic deleterious mutation rate $U \sim 0.96$. This rate of mutation is consistent with a common ancestor for *C. elegans* and *C. briggsae* of 30 Ma or less, provided that these species pass through at least 6 generations per year (Cutter 2008a; fig. 4). Of the new mutations observed in MA lines, more than half were short indel mutations, with fewer deletions than insertions (Denver, Morris, Lynch, and Thomas 2004). This finding on indels contrasts with previous work with *C. elegans* pseudogenes in which deletions are more common than insertions (Robertson 2000; Witherspoon and Robertson 2003) but is consistent with findings in *C. elegans* microsatellite mutation data (Frisse 1999; Denver, Morris, Kewalramani, et al. 2004; Seyfert et al. 2008). Phenotypic measures of mutation rate suggest that it is higher in *C. briggsae* than *C. elegans* (Baer et al. 2005), as do microsatellite mutation rates (Phillips et al. 2009), but further work is needed to clarify potential heterogeneity among *C. briggsae* isolates in average mutation rate and when in the past *C. briggsae*'s higher mutation rate might have evolved (Howe and Denver 2008). Another remaining open empirical question is how heterogeneous mutation rates are across the *C. elegans* genome. High densities of indels and SNP as well as high sequence divergence on the chromosome arms (Koch et al. 2000; Maydan et al. 2007) suggest significant regional variation in mutation rates, perhaps associated with recombination (Cutter and Payseur 2003b). However, direct quantification of such an effect with MA lines would make this conclusion definitive.

DNA repair–deficient strains of *C. elegans* representing three different repair pathways have been characterized in MA lines to determine the influence of each pathway on mutation rate (Denver et al. 2006). Mutation rates measured in microsatellites from MMR deficient *C. elegans* lines are increased by a factor of ~257 when compared with microsatellite mutation rate in wild-type lines (Seyfert et al. 2008). However, in yeast, this difference is a factor of ~1,346 (Seyfert et al. 2008), indicating that MMR may be less integral to microsatellite mutation rate in *C. elegans*.

Microsatellite loci in *C. elegans* generally do not evolve in accord with the strict stepwise mutation model (SMM; Ohta and Kimura 1973). Although more than 70% of new mutations add or remove a single repeat, long microsatellites can deviate greatly from this pattern (Frisse 1999; Degtyareva et al. 2002; Seyfert et al. 2008; Phillips et al. 2009). It should be noted, however, that these studies of de novo microsatellite mutations focus on loci that have unusually long repeat lengths for the *C. elegans* genome and that longer microsatellite loci tend to have a higher incidence of multistep changes (Seyfert et al. 2008). Population polymorphism analyses of microsatellites also can show deviation from the SMM (Sivasundar and Hey 2003; Haber et al. 2005), probably due in part to the effects of population structure. Although there is no strong correlation between motif type and mutation rate for dinucleotide microsatellites, longer allele sizes (more repeats) tend to have higher mutation rates (Frisse 1999; Seyfert et al. 2008; Phillips et al. 2009). For mononucleotide repeats, longer alleles also tend to experience more mutations and fewer insertions relative to deletions (Denver, Morris, Kewalramani, et al. 2004). For dinucleotide short tandem repeats (STRs), the total genomic mutation rate is estimated to be 0.12 mutations per generation in *C. elegans* and 2-fold higher in *C. briggsae* (Phillips et al. 2009).

Mutation load also has been considered to play an important role in the evolution of sex and breeding systems (de Visser and Elena 2007). Under some models, a selfing or asexual reproductive strategy could be maintained if the genomic deleterious mutation rate (U) is sufficiently low in selfing lineages (Kondrashov 1985; Charlesworth 1990). Cutter and Payseur (2003a) tested for a role of U in driving breeding system evolution in this clade by inferring U from patterns of interspecific protein sequence divergence in three *Caenorhabditis* species (Kondrashov and Crow 1993). They concluded that it cannot be a primary determinant because selfers and outcrossers do not differ significantly in U. Using a genomic sample of loci, Artieri et al. (2008) found a significant difference between *C. briggsae* (mostly selfing) and *C. remanei* (obligately outbreeding); however, in a more extensive phylogenetic framework, there does not seem to be a consistently different substitution rate between selfing and outcrossing *Caenorhabditis* species (Cutter et al. 2008). Comparable results also were found by Wright et al. (2002), in that outcrossing and selfing species of *Arabidopsis* do not differ significantly in protein sequence divergence and, by implication, the deleterious mutation rate. It could be argued that if the *C. elegans* lineage evolved selfing and a lower mutation rate recently, then this might not be detected from sequence comparisons; indeed, decay of codon usage bias suggests a relatively recent origin of selfing in *C. elegans* (Cutter et al. 2008). However, genomic mutation accumulation via Muller's ratchet (Muller 1964) would be expected to drive extinct a purely selfing *Caenorhabditis* lineage in a very short period of time in the absence of compensatory mutations and/or outcrossing (Loewe and Cutter 2008). In the selfing *C. briggsae*, mutation rates may actually have elevated (Baer et al. 2005; Howe and Denver 2008), and it seems plausible that species inhabiting warmer climates may exhibit a higher mutation rate per year by virtue of a shorter mean generation time, regardless of breeding system (Cutter 2008a). Overall, there is little consistent evidence for mutation rate differences among lineages having contributed to the origin and/or maintenance of selfing in *Caenorhabditis*.

Sequence Divergence

Comparative analysis with *C. elegans*' closest known relatives—the selfing species *C. briggsae* and the obligately outcrossing *C. remanei* (Stein et al. 2003; Hillier et al. 2007)—has proven to be a powerful approach toward understanding the evolutionary basis of form and function in *C. elegans* (Fitch 1997; Fitch and Thomas 1997; Haag et al. 2007; Kiontke et al. 2007). Although morphologically these species are exceedingly similar, even at the level of cell lineage (Zhao et al. 2008), they differ starkly in orthologous sequence and genome organization (Coghlan and Wolfe 2000; Stein et al. 2003; Hillier et al. 2007), sharing a common ancestor likely within the last 30 My (Cutter 2008a). Here we consider the patterns of molecular evolution evident in studies of *C. elegans* coding and noncoding regions.

Divergence in Coding Sequences

One metric of the ongoing nature of evolution of protein sequences, based on the comparisons of two or more species, is the rate of nonsynonymous site substitution (d_N), which is calculated under an explicit molecular model of mutation (unlike "percent identity" or "percent similarity"). In order to account for potential heterogeneity in mutation rate among loci that could impact rates of protein evolution in a way unrelated to selection, d_N is typically standardized as the d_N/d_S ratio, where d_S is the rate of synonymous substitution. Although it is often assumed that synonymous sites evolve neutrally, so d_S should directly reflect the mutation rate (Kimura 1968), in *C. elegans*, it is necessary to correct raw d_S values for historical selection on codon usage (Cutter 2008a). Neutrally evolving proteins have $d_N/d_S = 1$, whereas $d_N/d_S < 1$ implies purifying selection on amino acid sequences and $d_N/d_S > 1$ is indicative of genes subject to repeated positive selection (Yang and Bielawski 2000).

Stein et al. (2003) identified about 12,155 one-to-one orthologous genes between *C. elegans* and *C. briggsae*, with proteins having diverged ~12% on average (mean d_N = 0.12). Comparison of orthologous loci between all three species *C. elegans*, *C. briggsae*, and *C. remanei* (as well as for several other species in the genus) gives similar estimates of average protein evolutionary rate along each lineage (Cutter and Payseur 2003a; Artieri et al. 2008; Cutter et al. 2008). However, the distribution of d_N across genes is skewed, such that mean d_N is nearly 50% higher than median d_N. Although *C. briggsae* shows slightly elevated lineage-specific protein divergence in three-way comparisons with *C. elegans* and *C. remanei* (Artieri et al. 2008; Cutter et al. 2008), this might simply reflect a higher mutation rate in the lineage leading to *C. briggsae* that is not fully accounted for in d_S estimates due to synonymous site saturation. In comparison with the closer male–female relative *Caenorhabditis* sp. 5, *C. briggsae* shows no overall elevation in protein evolutionary rates, suggesting that the higher d_N (relative to *C. remanei*) cannot be attributed to a selfing lifestyle.

At the level of individual gene families, the *srz* genes of the large SR (seven-pass transmembrane receptor) gene family in *C. briggsae* are more divergent in their amino acid sequences than are family members within *C. elegans*. Also in contrast to other known SR subfamilies, statistical tests of protein divergence suggest positive selection acting on some members of the *srz* subfamily (Thomas et al. 2005). Further, heterogeneity in rates of evolution among different protein domains in the SRZ protein family is hypothesized to have been driven by selection favoring recognition of diverse ligands, possibly mediating biological processes of mate choice, sperm–egg interaction, nociception, and pathogen avoidance (Thomas et al. 2005)—though it is not clear if such putative selection predated the origin of selfing in *C. elegans*. However, *srh* and *str* genes that are rendered nonfunctional in the N2 strain by the presence of one stop codon (so-called "flatliner" genes) were found to otherwise exhibit $d_N/d_S < 1$, that is, purifying selection (Stewart et al. 2005).

Evidence of positive selection also has been detected in the subfamily D of the ATP-binding cassette family in *C. elegans* that is otherwise highly conserved among closely related genomes (Zhao et al. 2007). The molecular evolution of two of *C. elegans* immunity-related lysozyme

genes indicates the action of positive selection, whereas most other lysozyme genes show signatures of purifying selection (Schulenburg and Boehnisch 2008). The ancestry of some other immunity-related antimicrobial peptides also implies subjection to positive selection (Pujol et al. 2008). Transcription factors form another class of loci that generally evolve quickly in *Caenorhabditis* (Castillo-Davis et al. 2004; Cutter and Ward 2005; Haerty et al. 2008), as in humans and flies (Gilad et al. 2006; Clark et al. 2007). Suggestive of the operation of positive selection on transcription factors, these genes are disproportionately represented among orthologs of *C. elegans*, *C. briggsae*, and *C. remanei*, yet have a higher rate of protein evolution than other loci (Reece-Hoyes et al. 2007; Haerty et al. 2008). In contrast, collagen, protein synthesis, and small molecule transport proteins all typically exhibit the slowest average rates of evolution (i.e., strongest constraint and purifying selection) among gene ontology categories of loci (Castillo-Davis et al. 2004; Cutter and Ward 2005). Curiously, collagens and protein synthesis genes that are expressed primarily in adults tend to show faster rates of protein evolution than do those that are expressed primarily in larvae, and this adult–larva difference is more striking than for most other gene ontology categories (Cutter and Ward 2005).

Comparative sequence analysis becomes more powerful with the addition of more taxa of varying phylogenetic depths. Present limitations with available sequenced species of *Caenorhabditis* are that they have roughly similar levels of divergence from each other and that synonymous sites are saturated with substitutions in pairwise comparisons—that is, ancestral states of noncoding and synonymous site nucleotides cannot be inferred with confidence. The extension of comparative analyses to newly identified close species pairs (e.g., *C. briggsae*—"*Caenorhabditis* sp. 9"; Felix MA, Kiontke K, personal communication) as well as inclusion of outgroups (*C. japonica*, *Caenorhabditis* sp. 7; Kiontke K, personal communication) will prove useful in resolving many remaining molecular evolutionary questions.

Ontogeny and Molecular Evolution

Caenorhabditis elegans was first introduced in 1965 as a genetically and morphologically tractable model for studying animal development (Brenner 1974). Because developmental processes ultimately depend on the activity of specific sets of genes and their interactions, interspecific patterns of divergence among genes and regulatory regions should help us understand the evolution of development. A genomic signature of developmental constraint or divergence may therefore be evident at the level of protein-coding sequences involved directly in developmental pathways or at the level of noncoding regulatory sequences controlling the timing or spatial interaction of these proteins or both.

One of the earliest studies examining the dynamics of *C. elegans*' genome during development showed that the nuclear genome does not change in size by more than a few percent across ontogeny and that the pattern of repetitive sequences is identical in germ line and somatic DNAs (Emmons et al. 1979). This demonstrated that unlike certain protozoans, crustaceans, and other nematodes, the *C. elegans* genome does not undergo a dramatic rearrangement during ontogeny (Tobler and Muller 2001).

With the advent of genome-wide expression data, more recent studies have attempted to assess how transcription of genes is modulated throughout development and how patterns of molecular evolution differ among genes that are expressed differentially over time or among tissues. For the genes that are differentially expressed during embryogenesis (Hill et al. 2000; Baugh et al. 2003), no significant differences were found in amino acid replacement changes (d_N) between early-expressed genes and late-expressed genes in the embryo (Castillo-Davis and Hartl 2002; Cutter and Ward 2005). The synonymous substitution rate (d_S), however, is significantly lower in the late-expressed embryonic genes (Castillo-Davis and Hartl 2002), probably reflecting stronger translational selection for codon bias among these transcripts. Furthermore, only genes with transiently elevated expression in early embryogenesis exhibit significantly higher rates of protein evolution than genes with other embryonic expression profiles (Cutter and Ward 2005). These patterns contrast with that observed in *D. melanogaster* in which consistently lower rates of evolution are seen during the second half of embryogenesis (Davis et al. 2005). The pattern in *C. elegans* might reflect weaker selection on transiently expressed genes, positive selection, or a potential conflict of interest between embryo and mother that promotes protein divergence in early, transient, embryo-transcribed genes (Cutter and Ward 2005).

Later in ontogeny (from third larval, L3, to the adult stage), however, adult-expressed genes have higher rates of protein evolution than larval-expressed genes (Cutter and Ward 2005). Genes identified as having enriched expression during spermatogenesis (Reinke et al. 2000, 2004) account for much of the elevation in rate of protein evolution during the L4 and young adult stages, whereas genes associated with gonad development partially account for lower evolutionary rates in L3. Nevertheless, somatically expressed proteins of adults evolve faster than larval somatic proteins, with particularly strong larva–adult differences for chromatin-related genes (a fast-evolving class, on average) and for collagens and protein synthesis–related genes (slow-evolving classes, on average) (Cutter and Ward 2005). This general result has been interpreted to be consistent with the mutation accumulation model of aging and senescence (Medawar 1952), as also expected given the rapid decline in reproductive value of hermaphrodites following the onset of maturity (Chen et al. 2006).

In yeast, proteins with many interactions are purported to exhibit slower rates of molecular evolution (Jordan et al. 2003; Fraser and Hirsh 2004), although this may be largely a by-product of a coupling between high expression and number of interacting partners (Bloom and Adami 2003). Genome-scale studies in *C. elegans* have not tackled this question of whether proximate pleiotropic effects—mediated through many protein interactions—might result in slower rates of evolution, independently of confounding factors. Large-scale interaction screens may provide useful data to test such ideas in *C. elegans* (Lehner et al. 2006; Byrne et al. 2007; Simonis et al. 2009). On a smaller scale,

one metric of pleiotropy among a set of embryo-expressed genes did not correlate with rate of evolution (Zou et al. 2008), although it must be noted that genes expressed throughout embryogenesis generally have constrained protein evolution (Castillo-Davis and Hartl 2002; Cutter and Ward 2005). However, quantifying protein evolution along the length of particular pathways provides another way to examine the role of pleiotropy in constraining molecular evolution, based on the notion that genes acting upstream will have stronger cascading pleiotropic effects (and might therefore experience greater selective constraint). There is no obvious correspondence between position and evolutionary rate within the *C. elegans* sex determination pathway, although there is a suggestion of stronger conservation among proteins that are more likely to exhibit pleiotropy due to splicing and nucleic acid-binding–related functions (Haag 2005). Similarly, within the context of a chemosensory pathway, Jovelin et al. (2009) report no significant association between pathway position and evolutionary rate, suggesting that a protein's position within a pathway per se is unlikely to be a primary determinant of selection acting on it. However, proteins with regulatory roles in the chemosensory pathway evolve faster than proteins with structural roles (Jovelin et al. 2009). Together with the more rapid evolution of transcription factors than other gene classes (Haerty et al. 2008), this implicates divergence in proteins involved in gene regulation as an important evolutionary phenomenon in *Caenorhabditis*.

Sexual selection often is invoked to explain rapid evolution in reproduction-related and sperm-related genes (Swanson and Vacquier 2002)—yet sexual selection is predicted to be weak in hermaphroditic and selfing taxa (Greeff and Michiels 1999; Cutter 2008c). Proteins encoded by sperm-related genes in *C. elegans* have been found to evolve more rapidly than most other genes in the genome, including male somatic genes and both the somatic and germ line genes of hermaphrodites (Cutter and Ward 2005; Artieri et al. 2008). Sperm-related genes also have fewer orthologs and poorly conserved synteny in *C. briggsae* and more intraspecific paralogs in *C. elegans*, indicating a trend of faster evolution in terms of sequence divergence, gene duplication, loss, and translocation (Cutter and Ward 2005; Artieri et al. 2008). Curiously, sperm genes also appear to experience disproportionate change in expression level in MA experiments (Denver et al. 2005). In a recent contrast of lineage-specific rates of protein evolution for *C. briggsae* and *C. remanei*, orthologs of *C. elegans* sperm genes showed accelerated evolution in both lineages (Artieri et al. 2008). Therefore, faster evolution of sperm proteins in selfing taxa is probably a remnant of selection experienced by the male–female ancestral state of both *C. elegans* and *C. briggsae* (Artieri et al. 2008).

In addition to categorizing genes based on gene expression profiles or positions within genetic pathways, functional phenotypes can be defined from gene knockdown by RNAi (Fire et al. 1998), a method that has been implemented at a genomic scale in *C. elegans* (Kamath et al. 2003). It should be noted at the outset that high-throughput RNAi screens necessarily ignore potentially important effects due to their subtlety, genetic background, subjective phenotypic assessments, or to the particular aims of the screen (Echeverri et al. 2006). Nevertheless, genes with observable "obvious" RNAi phenotypes have been noted to possess similar identifying features, namely, being long, highly expressed, and slowly evolving (Cutter et al. 2003). Although differences among genes in the effectiveness of RNAi might contribute to these results, biological and functional differences among the gene classes also are likely to be important (Cutter et al. 2003). Genes with RNAi-induced phenotypes tend to have homologs in more distantly related taxa than do other genes (Kamath et al. 2003). More specifically, genes with *C. briggsae* orthologs that yield *C. elegans* RNAi phenotypes evolve 20% more slowly at the protein level than genes with no obvious RNAi phenotype, and protein evolution for genes with RNAi-induced effects on fecundity evolve slower than those that influence viability or other organismal phenotypes (Cutter et al. 2003). Further, genes that have more severe effects on fertility evolve more slowly than those with less severe effects on fertility. This observation contrasts with the faster evolution of sperm genes inferred from gene expression data (see above) as well as the finding in many organisms that loci involved in reproduction evolve rapidly (Swanson and Vacquier 2002). These results exemplify how strong purifying selection acts on many reproduction-related genes and that "reproduction-related" does not necessarily imply "sexually selected" (Dean et al. 2008).

A major gap in comparative genomic analyses of gene function and evolution in *Caenorhabditis* is that presently available gene expression assays derive only from *C. elegans* source material. Although expressed sequence tag (EST)–based expression data are available for a diversity of nematodes, including a number of *Caenorhabditis* species (Cutter et al. 2008) but mostly for parasitic nematodes (Parkinson, Whitton, et al. 2004), these provide only rough guides to expression patterns. Careful gene expression analysis of multiple *Caenorhabditis* species across life stages and for each sex will be important to dissect the evolution of gene function and regulation. Similarly, other genome-wide functional screens, like RNAi, are sorely needed in non-*elegans* taxa, especially given the derived breeding system in *C. elegans*.

Constraint in Noncoding Sequences

Independent lineages diverge through the accumulation of mutations via random genetic drift. Although positive selection accelerates this divergence, uniform negative (purifying) selection leads to reduced divergence. Such selective constraint is most obvious for coding sequences, but regulatory and other noncoding DNA also are subject to purifying (and positive) selection that is detectable in interspecific contrasts. Comparison of *C. elegans* and *C. briggsae* noncoding sequences has revealed different rates of evolution and strengths of selective constraint among different chromosomal segments of the *C. elegans* genome. Based on analysis of 150 kb of homologous sequence, Shabalina and Kondrashov (1999) concluded that 32% of the genome of *C. elegans* is functionally conserved (27% of the genome resides in coding exons; *C. elegans* Sequencing Consortium 1998). Approximately 72% of coding sites are invariant

between these species, mostly nonsynonymous sites, whereas 17% of intronic sites are constrained, possibly due to their functioning in splicing and gene regulation. Similar to introns, 18% of sites in intergenic sequences are constrained, presumably due to regulatory functions. These numbers are likely underestimates of functional regions as it only accounts for sites that are conserved between these distantly related congeners, therefore necessarily excluding lineage-specific functional noncoding elements. Furthermore, 44% of noncoding DNA is transcribed (He et al. 2007). Such a large extent of the genome subject to purifying selection in conjunction with a partially selfing mode of reproduction suggests that background selection (Charlesworth et al. 1993) is likely to be a potent force influencing patterns of polymorphisms in these species (Graustein et al. 2002; Cutter and Payseur 2003b; Sivasundar and Hey 2003).

In a comparison of ~142 orthologous intergenic regions (~98 kb total), Webb et al. (2002) found 71% of nucleotides in highly conserved regions to be constrained in *C. elegans* and *C. briggsae*. In UTRs, the constraint is about 43% and is 15% in intergenic regions. The conserved UTRs can reflect a variety of regulatory elements, such as noncoding exons and elements of mRNA secondary structure, and some regions outside the UTR could be conserved RNA genes and transcription factor–binding sites. The characteristics of these highly conserved regions could be helpful for prediction of conserved regulatory elements and likely places to look for regulatory sequences (Webb et al. 2002). Indeed, by scanning 0.5–2.0 kb of upstream sequences of orthologous genes between *C. elegans* and *C. briggsae*, regulatory sequences controlling *C. elegans* pharyngeal development, vulval expression, and muscle gene expression have been identified (Kirouac and Sternberg 2003; Gaudet et al. 2004; GuhaThakurta et al. 2004; Coghlan et al. 2006). With the sequences of other related species now available, comparative genomic approaches will help define additional regulatory elements and characterize their patterns of molecular evolution.

The relation between the evolution of protein and regulatory sequences is a crucial question in molecular evolution. To what extent are changes in gene expression (due to evolution of regulatory elements) coupled with evolution of protein-coding sequences, and do these changes differ among orthologs and paralogs? By comparing orthologous *C. elegans* and *C. briggsae* sequences, Castillo-Davis, Hartl, and Achaz (2004) observed a weak but significant correspondence between rates of evolution for coding sequence and their *cis*-regulatory regions. However, for paralogs, no such correlation was found, implying that coordinated selection on coding and regulatory regions persists over long stretches of evolutionary time following divergence due to speciation events but not due to gene duplication events. This pattern indicative of stabilizing selection on both gene expression and protein function is consistent with the findings of Denver et al. (2005) for gene expression changes in MA lines versus wild strains. The faster rates of both *cis*-regulatory and protein evolution in duplicated genes could result from either relaxed or positive selection (Castillo-Davis, Hartl, and Achaz 2004). This recurring difficulty in distinguishing between relaxed and positive selection would benefit from the identification of species with mutationally unsaturated synonymous site divergence relative to *C. elegans*.

Codon Usage Bias

The nonuniform usage of synonymous codons—codon usage bias—is prevalent among prokaryotic and eukaryotic genomes, including *C. elegans*. However, it can be a difficult task to determine whether codon bias is caused by neutral, mutational processes, or by natural selection because the strength of such selection is extremely weak. Selection for accurate and/or efficient translation are believed to cause codon bias among highly expressed genes (Duret 2002), and it has been demonstrated experimentally in some systems that unpreferred codons in a gene can lower expression levels and reduce fitness (Robinson et al. 1984; Varenne et al. 1984; Sorensen et al. 1989; Andersson and Kurland 1990; Carlini and Stephan 2003; Carlini 2004). The very weak selection among synonymous codons that drives codon usage bias in many species is a testament to the power of evolution by natural selection: Fitness differences between alleles on the order of one in a million or less can yield an evolutionary response in populations that are sufficiently large.

Inferring Selection for Codon Bias

Codon usage bias can be quantified in a number of ways, and studies for *C. elegans* have used many of them (relative synonymous codon usage [RSCU; Sharp et al. 1986]; difference in RSCU between high and low expression loci [Duret and Mouchiroud 1999; Cutter, Wasmuth, and Blaxter 2006]; effective number of codons [Wright 1990]; frequency of optimal codons [F_{op}; Ikemura 1985]; scaled χ^2 [Shields et al. 1988]; codon adaptation index [Sharp and Li 1987; Carbone et al. 2003]). However, the mere observation of nonrandom codon usage does not necessitate that natural selection is the cause.

The first systematic analysis of codon usage in *C. elegans* is the study of Stenico et al. (1994), although biased codon patterns were pointed out earlier (Emmons 1988; Thomas and Wilson 1991; Kennedy et al. 1993). It is from this study that most of *C. elegans* optimal codons were identified (see also Sharp and Bradnam 1997; Duret and Mouchiroud 1999). "Optimal" or "preferred" codons are those codons that are incorporated more frequently in highly expressed genes relative to lowly expressed genes and are thought to be favored by selection over their synonymous alternatives. Optimal codons can differ from "major" codons that are simply observed more commonly in a genome overall, possibly due to base composition skew (Kliman et al. 2003). The early study of Stenico et al. (1994) also showed qualitatively that genes encoding abundant proteins also tended to be highly biased in codon usage, thus linking selection to codon usage patterns. They also suggested that selectively neutral, mutational processes are sufficient to explain skewed codon frequencies in genes with low expression but not for genes with high expression, which has been verified subsequently (Duret and Mouchiroud 1999). Furthermore, the codon usage patterns of TEs in

C. elegans are most similar to those of genes with low expression levels (Lerat et al. 2002).

Higher levels of codon bias are observed in genes with higher levels of expression, whether gene expression is quantified with microarrays (Castillo-Davis and Hartl 2002; Cutter et al. 2003), EST counts (Duret and Mouchiroud 1999; Cutter, Wasmuth, and Blaxter 2006, Cutter et al. 2008), or serial analysis of gene expression (Singh et al. 2005). ESTs have been used as a measure of expression for analyzing codon bias in other species of *Caenorhabditis* (Cutter et al. 2008), but more direct measures of expression level are needed in non-*elegans* taxa. The lack of an association between base composition of intronic DNA with the expression level of the associated gene rules out the possibility that mutation rates might be elevated in genomic regions with greater levels of transcription (Duret and Mouchiroud 1999), as seen for *Escherichia coli* (Francino and Ochman 2001). Thus, despite the contribution of selectively neutral mutational processes in skewing codon usage (Duret and Mouchiroud 1999; Marais et al. 2001; Cutter, Wasmuth, and Blaxter 2006, Cutter et al. 2008), natural selection has been an important force in shaping codon bias in *Caenorhabditis* genomes.

Additional evidence supporting a role for natural selection having shaped codon usage in the *C. elegans* genome is the observation that the relative abundance of amino acids in the proteome correlates strongly with the number of cognate tRNA gene copies in the genome, particularly among highly expressed proteins (Duret 2000; Percudani 2001). Presumably, the number of duplicate tRNA gene copies in the genome reflects tRNA abundance in the cell, as in bacteria (Kanaya et al. 1999). Selection for codon bias also results in a strong negative association between synonymous-site divergence and the magnitude of codon bias (Kennedy et al. 1993; Stenico et al. 1994; Castillo-Davis and Hartl 2002; Cutter and Payseur 2003a; Cutter and Ward 2005; Cutter et al. 2008). This observation means that synonymous sites have not evolved in a strictly neutral fashion, particularly for genes with high expression and strong codon bias. Estimates of the historical strength of selection on *C. elegans* gene orthologs of highly expressed yeast genes also indicate translational selection in the ancestry of nematodes (dos Reis and Wernisch 2009). Based on patterns of polymorphism at synonymous sites, the intensity of selection ($N_e s$; effective population size N_e, selection coefficient s) between alternative synonymous codons in present-day populations was quantified for *C. remanei*, indicating that $N_e s \sim 0.1$ on preferred codons (Cutter and Charlesworth 2006; Cutter 2008b). These studies also demonstrate that present-day patterns of selection for codon bias in *C. remanei* are coincident with long-term selection for codon bias. Similar analyses have not yet been conducted in *C. elegans* due to practical difficulties (low nucleotide polymorphism, demographic complications), but present-day selection for codon bias is predicted to be effectively absent in both *C. elegans* and *C. briggsae* due to their greatly reduced effective population sizes (Cutter et al. 2008). Thus, biased codon usage among highly expressed genes in *C. elegans* likely is a relic of translational selection in its outcrossing ancestors that experienced substantially larger effective population sizes.

Genic and Genomic Correlates of Codon Bias

Because codon bias represents a phenomenon for which mutation, genetic drift, and selection have roughly similar magnitudes of effect, studies have identified a diverse set of chromosomal characteristics that correlate with codon bias in an attempt to better characterize the relative importance of these evolutionary forces. Gene length, constraint of neighboring replacement sites, nucleotide composition, chromosome identity, and gene expression all exhibit intriguing associations with codon bias, which we discuss in turn.

Shorter coding sequences experience stronger codon bias in *C. elegans*, independent of other factors (Duret and Mouchiroud 1999; Marais and Duret 2001). This phenomenon appears to be common to *Caenorhabditis* species (Cutter et al. 2008), other nematodes (Cutter, Wasmuth, and Blaxter 2006), and other metazoans (Duret and Mouchiroud 1999) but is opposite to yeast and bacteria (Eyre-Walker 1996; Coghlan and Wolfe 2000). Furthermore, codon bias for sites encoding amino acids that have not diverged between *C. elegans* and human also correlates negatively with gene length (Marais and Duret 2001). This length effect does not appear to be due to selection on abundant proteins favoring shorter sequences or to a greater proportion of constrained codons in short sequences (Duret and Mouchiroud 1999; Marais and Duret 2001) but might result from intragenic background selection or Hill–Robertson interference (Comeron and Guthrie 2005; Loewe and Charlesworth 2007). The slower rate of protein evolution in short genes (Marais and Duret 2001) also seems consistent with a background selection model.

Genes linked to the X chromosome appear to exhibit slightly stronger codon bias on average than autosomal genes, despite the observation that X-linked loci generally seem to be expressed more weakly than loci on the autosomes (Singh et al. 2005). This pattern also is evident for gene duplicates, such that duplicate copies on the X have more biased codon usage than their autosomal counterparts, although the magnitude of the difference is quite small (Singh et al. 2005). However, it is unclear whether this effect might be due to greater strength of translational selection on X-linked loci, a skewed sex ratio in the outcrossing ancestor of *C. elegans*, an influence of dosage on selection for codon bias, or peculiar mutational properties of the X (Singh et al. 2005).

Although codon usage bias correlates weakly with recombination rate, such that regions of high recombination harbor loci with stronger codon bias, this effect is limited to optimal codons ending in G or C, and G + C content also is slightly higher in high recombination regions (Marais et al. 2001; Marais and Piganeau 2002). It has been argued that selectively neutral processes are responsible for this pattern (Marais et al. 2001; Marais and Piganeau 2002), possibly biased gene conversion toward guanine and cytosine (Marais 2003), rather than reflecting Hill–Robertson interference between weakly selected sites. However, biased gene conversion was ruled out as a force skewing codon usage patterns in *C. remanei*, based on nucleotide polymorphism data (Cutter and Charlesworth 2006; Cutter 2008b).

Stronger codon bias is observed among genes that evolve more slowly at the protein level (Cutter and Payseur

2003a; Cutter and Ward 2005; Cutter et al. 2008). Codon bias also is stronger in codons that encode conserved amino acids relative to ones that have diverged between *C. elegans* and human, even when variation in amino acid composition is controlled for (Marais and Duret 2001). This stronger bias in highly conserved proteins is similar to the pattern reported for *Drosophila* (Betancourt and Presgraves 2002; Marais, Domazet-Loso 2004; Bierne and Eyre-Walker 2006), *Arabidopsis* (Wright et al. 2004), and *Populus* (Ingvarsson 2007). This could imply that mutations to both replacement and synonymous sites are more deleterious in genes subject to strong purifying selection, perhaps as a consequence of greater pleiotropy and evolutionary constraint among genes expressed in many tissues, which also tend to be expressed at high levels (Wright et al. 2004; Ingvarsson 2007).

Genes expressed earlier in development or that are expressed constitutively tend to have more strongly biased codon usage patterns (Duret and Mouchiroud 1999; Cutter and Ward 2005). Such genes might have greater pleiotropy by virtue of their expression in many tissue types. However, the identity of optimal codons does not vary across development (Duret and Mouchiroud 1999). In addition, codon bias is stronger among those genes that induce sterility when knocked down by RNAi relative to genes with other RNAi phenotypes or genes with no obvious RNAi-induced phenotype (Cutter et al. 2003). These patterns probably simply reflect the high expression levels and strong purifying selection on genes that act in these processes or stages of development.

Selection for exon splice site conservation provides another force that biases codon usage at the beginning and end of exons (Eskesen et al. 2004) because most exons are in phase-zero (Eskesen et al. 2004). In addition to forces operating on the nucleotides of the codons themselves resulting in nonrandom codon usage, nucleotide context also appears to associate with codon usage in *C. elegans* (Fedorov et al. 2002). Specifically, the relative abundance of 37% of *C. elegans* codons, in the context of the first nucleotide of the downstream neighboring codon, differs from the frequencies expected based on genomic base composition (Fedorov et al. 2002); the proposed explanation is that selection for accurate protein synthesis resulted in such nucleotide context-dependent codon patterns.

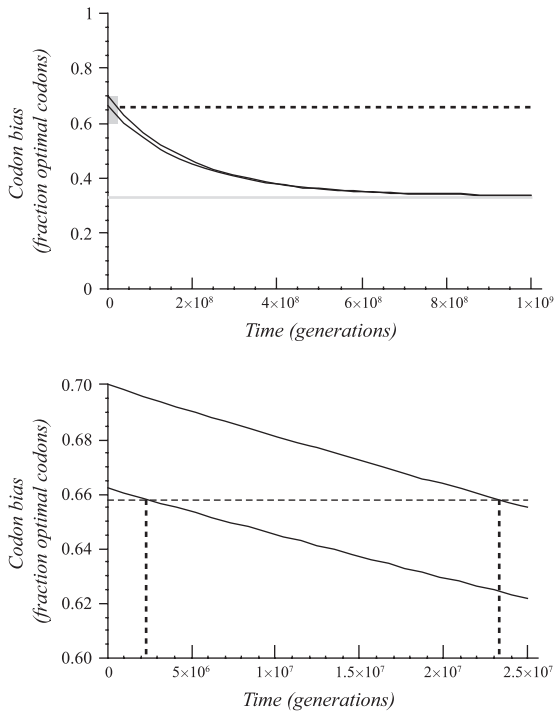

FIG. 5.—Expected decay in codon usage bias (frequency of optimal codons, F_{op}) over time following complete relaxation of selection, from Cutter et al. (2008). In (*A*), gray line indicates the mutation-drift equilibrium codon bias and the dashed line represents present-day codon bias in *Caenorhabditis elegans* for 63 orthologs of six species of *Caenorhabditis* (Cutter et al. 2008). The small gray box in (*A*) represents the range of values shown in (*B*). In (*B*), the horizontal dashed line represents present-day codon bias in *C. elegans*, vertical dashed lines indicate points of intersection with the decay curves. Upper decay curve assumes an ancestral codon bias of $F_{op} = 0.7$ (seen in most outcrossing *Caenorhabditis*), lower decay curve assumes an ancestral codon bias of $F_{op} = 0.662$ (seen for outgroup *Caenorhabditis japonica*). Assuming relaxation of selection on codon usage following the onset of a selfing lifestyle in *C. elegans*, the decay time provides an estimate of the time since the origin of selfing.

Evolutionary Implications of Codon Bias

Different species of *Caenorhabditis* have largely consistent sets of optimal codons, and orthologs among different species of *Caenorhabditis* correlate strongly in their extent of codon bias (Stenico et al. 1994, 2003; Cutter and Ward 2005; Cutter et al. 2008). However, two self-fertilizing members of the genus (*C. elegans* and *C. briggsae*) show a subtle reduction in overall codon bias relative to their nearest outcrossing relatives, which has been used to infer that the onset of selfing in *C. elegans* occurred in the not-too-distant past (fig. 5; Cutter et al. 2008). Selection for codon usage appears to be stronger in *Caenorhabditis* than for many other metazoans like *Drosophila*, *Arabidopsis*, and mammals (Duret and Mouchiroud 1999; Kanaya et al. 2001; dos Reis and Wernisch 2009), as well as compared with most parasitic nematodes (Cutter, Wasmuth, and Blaxter 2006). Higher observed codon bias likely reflects larger long-term historical effective population sizes, such that selection could operate more effectively on mutations with such weak effects on fitness. Indeed, the substantially smaller effective size of current *C. elegans* and *C. briggsae* populations relative to a related gonochoristic species (*C. remanei*; Graustein et al. 2002; Cutter, Baird, and Charlesworth 2006) implies that present-day selection on synonymous sites essentially is relaxed completely in the selfing species.

Repetitive DNA

Early on, it was recognized that only a modest fraction (~17%) of the *C. elegans* genome was comprised of repetitive sequence (Sulston and Brenner 1974). Nevertheless, repetitive DNA can play important roles in evolution (Kidwell and Lisch 1997). For this discussion, we focus on *C. elegans* repetitive sequence of only two categories: transposable elements (TEs) and short tandem repeats (STRs), recognizing that minisatellite and other repeats also are likely important in *C. elegans* genome evolution (e.g., Naclerio et al. 1992; Sanford and Perry 2001), as in

telomere stability (Wicky et al. 1996). We focus here on genomic characteristics; issues associated with mutation and population variation of these sequence types are discussed in other segments of this review.

TE Patterns

Approximately 12% of the *C. elegans* genome is composed of TEs (Bessereau 2006), although most of these TEs are truncated or otherwise nonfunctional (Duret et al. 2000). The majority of *C. elegans* TEs are "cut-and-paste" DNA transposons, with less than 20% of them being "copy-and-paste" retroelements (Duret et al. 2000). The proportion of elements that are full length also is greater for transposons than for retroelements (~14% vs. ~6%; Duret et al. 2000). Similar classes of TEs have been identified in different species of *Caenorhabditis* and among nematodes generally (Harris et al. 1990; Abad et al. 1991), and repeat sequences in the *C. elegans* and *C. briggsae* genomes can be masked fairly well by using repeat libraries from either species (Stein et al. 2003). However, the *C. briggsae* genome has a larger diversity of repeat types and a greater fraction of repetitive DNA (~22%) than *C. elegans*, largely due to a single DNA transposon (Cb000047), which explains nearly all the differences in total genome size between these two taxa (Stein et al. 2003).

TEs overall are more abundant in chromosome arms, regions that also experience higher rates of recombination and reduced gene density, although particular classes of TEs may or may not follow this nonrandom spatial distribution (Cangiano and La Volpe 1993; Barnes et al. 1995; Duret et al. 2000; Surzycki and Belknap 2000; Ganko et al. 2001). Notably, it is DNA transposons, but not retroelements, that predominantly show a bias toward high recombination regions (Duret et al. 2000; Rizzon et al. 2003). In addition, several miniature inverted-repeat elements are biased consistently toward one end of each chromosome (Surzycki and Belknap 2000). The bias of transposons toward regions of high recombination is common to both those element insertions that are full length (or nearly so, and therefore likely of recent origin) and those elements that are truncated and highly divergent (likely of ancient origin; Duret et al. 2000). Importantly, Rizzon et al. (2003) demonstrated a bias of new Tc1 insertions into high recombination regions in an *mut-7* "mutator" genetic background but not for Tc3 or Tc5 elements. Although Tc1 is less common in regions of the genome with low TA dinucleotide frequencies (Tc1 inserts preferentially at TA sites), this cannot account for the association between Tc1 abundance and recombination rate (Rizzon et al. 2003). Although much is understood relating to the molecular biology of TEs (Plasterk et al. 1999; Bessereau 2006), it remains an important problem to understand the genomic features that contribute to the nonrandom insertion patterns of TEs.

Nearly all TEs in the *C. elegans* genome (~98%) are found in noncoding sequence (Duret et al. 2000), despite the fact that a substantial proportion of new TE insertions occur in coding sequence (Rizzon et al. 2003), reflecting the strong natural selection against insertion into genes. However, selection does not appear to be substantially stronger against TE insertions into noncoding sequences in gene-dense regions relative to gene-poor regions (Rizzon et al. 2003). Indeed, nearly two-thirds of the 124 *Cer* retroelements occur within 1,000 bp of a gene, significantly more than expected by chance (Ganko et al. 2003). Moreover, *Bel*-like *Cer* retroelements are represented disproportionately within genes, relative to *gypsy*-like elements, and 78% of the these 40 gene × TE associations affect exonic sequence (Ganko et al. 2003). No bias in orientation between retroelements and nearby genes is observed in *C. elegans*' genome (Ganko et al. 2003), in contrast to *D. melanogaster* and mammals (van de Lagemaat et al. 2003; Cutter et al. 2005). However, it has not yet been determined whether the biased proximity to genes of retroelements has influenced the regulation of the genes, potentially leading to retention of the TEs due to beneficial effects or, instead, is due to greater protection from deletion when they insert closer to coding sequences (Ganko et al. 2003).

Processes Governing TE Distributions

Breeding system is expected to play an important role in TE dynamics and genomic abundance (Wright, Ness, et al. 2008), although the specific nature of selection against TEs is critical for the predicted effects of selfing (Wright and Schoen 1999; Morgan 2001). In particular, the influence of selfing in increasing homozygosity, reducing effective recombination, and lowering effective population size could create conditions favorable for the purging of TEs (if deleterious insertion is the dominant cause of selection) or potentially for their proliferation (if deleterious nonhomologous recombination between different TEs, "ectopic exhange," is most important) (Wright and Schoen 1999; Morgan 2001). The relative importance of deleterious insertions and ectopic exchange depend in large part on the relative fitness effects of elements occurring in a heterozygous versus homozygous state (Morgan 2001). The tentative observation of lower TE densities in selfing plants, relative to outcrossers, is suggestive of a greater role of selection against deleterious insertion in inbreeding populations (Morgan 2001). However, the high population frequencies observed for Tc1 elements in *C. elegans* could reflect weaker ectopic exchange in selfing lineages (Dolgin et al. 2008). The timescale over which selfing reproduction has persisted also is likely to be important in predicting the genomic abundance of TEs (Wright, Ness, et al. 2008). Another provocative hypothesis is that the rarity of retroelements compared with DNA transposons has arisen since the origin of selfing in *C. elegans* (Abrusan and Krambeck 2006). At present, it is difficult to conclude what role selfing might have played in shaping the abundance and distribution of TEs across the *C. elegans* genome. Comparisons with genomic TE distributions of outcrossing *Caenorhabditis* will be particularly informative in this regard.

The excess of TEs in high recombination regions conflicts with the prediction of population genetics models that regions of low recombination should accumulate TEs more readily (Langley et al. 1988; Nuzhdin 1999). The *C. elegans* pattern is reminiscent of that for *Arabidopsis thaliana* (Wright et al. 2003) but opposite to that of *D.*

melanogaster (Bartolome et al. 2002) and humans (Boissinot et al. 2001). However, higher homozygosity and overall weaker Hill–Robertson interference (Hill and Robertson 1966) among selected loci should reduce the potential for an association between recombination and TE abundance in selfing species (Morgan 2001). Therefore, it may be that the observed correlation between TE abundance and recombination rate in *C. elegans* is not a function of recombination per se but due to some other correlated factor. Specifically, TE bias against gene-dense regions, which also occur in the low recombination chromosome centers, might imply a dominant role of deleterious insertions relative to the detrimental consequences of nonhomologous recombination ("ectopic exchange") in shaping chromosomal TE distributions in selfing lineages (Morgan 2001); this also is seen in *A. thaliana* (Wright et al. 2003). Thus, the overabundance of TEs on *C. elegans*' chromosome arms plausibly reflects a combination of insertion bias (Rizzon et al. 2003) and stronger selection against insertions in the gene-dense chromosome centers, whereas the association with recombination might be incidental.

In addition to the theoretical role of population processes, a cellular mechanism, the natural RNAi response (Fire et al. 1998), has a demonstrative role in controlling TE proliferation in *C. elegans* (Sijen and Plasterk 2003; Robert et al. 2005). One analysis of *C. elegans*' genome suggests that the sequence-specific silencing of TEs likely is the dominant means of controlling their abundance, similar to chicken, puffer fish, and fruit flies but different from mammals (Abrusan and Krambeck 2006). Although experimental RNAi operates with varying degrees of success in different *Caenorhabditis* species (Winston et al. 2007), the relative effectiveness of the RNAi pathway in controlling TE activity among species is not yet known.

Despite the holocentric nature of *C. elegans* chromosomes during mitosis, during meiosis, chromosomes are functionally monocentric with centromere function generally occurring at the ends of chromosomes (Zetka and Rose 1995; Wicky and Rose 1996). It is plausible that this imposes on chromosomes a selective force, or a selectively neutral bias, that has contributed to the nonrandom distributions of TEs, genes, and recombination. Differences in the transposition process for different types of elements also are likely to generate heterogeneity in TE distributions across the genome (Duret et al. 2000), and some have suggested that regions of high TE density might encourage recombination (Cangiano and La Volpe 1993; Barnes et al. 1995). Duret et al. (2000) argue that TE activity per se is unlikely to lead to elevated recombination rates, though this is a separate issue from the potential for genomic features associated with TE density (or other repetitive DNA) to influence the likelihood of recombination.

There does not appear to be a strong X-autosome difference in TE density, although overall the density is slightly higher on the X (Duret et al. 2000; Ganko et al. 2001). This suggests that selection against X-linked TEs is not generally more effective, as would be expected if hemizygosity in males were an important means of exposing the (recessive) deleterious effects of TE insertions. However, there are a number of striking family-specific discrepancies in TE density between the X and autosomes, with the X chromosome being disproportionately over- or underrepresented among different types of TEs (Rezsohazy et al. 1997; Duret et al. 2000; Surzycki and Belknap 2000). It remains to be seen whether these family-specific patterns reflect greater mobility of some elements in the male germ line (Duret et al. 2000), subtle deterministic differences in transposition processes among elements, or simply genetic drift. New Tc1 insertions appear to derive principally from elements on the same chromosome (Fischer et al. 2003). This also is supported by a positive correlation between the abundance of Tc1 elements on each of the five autosomes and the corresponding autosomal abundances of several hundred new insertions in an *mut-7 (pk204)* genetic background: new Tc1 insertions are more common on chromosomes that originally had more Tc1 elements (Spearman's rank correlation = 0.9, $P = 0.034$; for Tc3, correlation = 0.82, $P = 0.089$; for Tc5 $P > 0.6$; reanalysis of data from http://pbil.univ-lyon1.fr/segalat/data/tc.php; Martin et al. 2002). Again, contrasts with obligately outcrossing relatives will help determine whether the X-autosome TE patterns in *C. elegans* are general or if breeding system might play an important role in TE abundance and location.

Short Tandem Repeats

STR loci (or simple sequence repeats or microsatellites) are composed of a short-sequence motif of one or more bases that occurs multiple times in a row. A key feature of such loci is that they typically change in length through the mutational addition or subtraction of repeat units.

In general, STRs are rare in the worm genome relative to other eukaryote model organisms (Toth et al. 2000; Katti et al. 2001; Dieringer and Schlotterer 2003). In particular, repeat motifs rich in guanine and cytosine are very rare, such as G_n or C_n, GC_n ($=CG_n$), and GGC_n (and equivalent classes; Katti et al. 2001; Dieringer and Schlotterer 2003; Denver, Morris, Kewalramani, et al. 2004). Despite the absolute rarity, long STRs occur more often than expected by chance from individual base frequencies in the genome (Dechering et al. 1998; Dieringer and Schlotterer 2003; Denver, Morris, Kewalramani, et al. 2004); specifically, mononucleotide STRs 3–10 bp in length are relatively more common than expected. It is not known whether this pattern simply reflects mutational processes or a potential functional role of many STRs for regulation or genome integrity. The abundance of repeats encoded by longer motifs are rarer than for short motifs, although *C. elegans* has less disparity in the incidence of repeats of different motif lengths than do many other species (Katti et al. 2001).

STRs are represented disproportionately on chromosome arms in *C. elegans*, which experience higher recombination rates and lower gene density (Barnes et al. 1995); yet, the nearly 2-fold difference in STR abundance is associated with only minor differences in overall nucleotide composition (from ~35% G + C to ~36% G + C; Duret et al. 2000; Denver, Morris, Kewalramani, et al. 2004). Not surprisingly, given variation in STR density along the chromosomes, the sequence intervals between repeat loci are

shorter than expected at random (Denver, Morris, Kewalramani, et al. 2004). They do not, however, appear to have any strong strand asymmetry (Denver, Morris, Kewalramani, et al. 2004). *Caenorhabditis briggsae* also exhibits a higher density of repeats in the recombination-rich, gene-poor arms of each chromosome (Stein et al. 2003; Hillier et al. 2007).

STR repeats in coding sequences are particularly uncommon, even for imperfect trinucleotide repeats that create strings of replicated amino acids (Katti et al. 2001). Repeats of proline, glutamine, aspartic acid, and glutamic acid make up more than half of the roughly 800 peptide repeats in the *C. elegans* genome that are at least seven codons in length, with only a handful of examples of peptide repeats greater than 14 units in length (Katti et al. 2001). Mononucleotide repeats longer than 2 (G or C) or 5 (A or T) bp in coding sequences occur less frequently than expected by chance (Ackermann and Chao 2006), in stark contrast to the genome-wide pattern (Dieringer and Schlotterer 2003; Denver, Morris, Kewalramani, et al. 2004). This bias against long mononucleotide runs in coding sequences is significantly stronger in highly expressed genes, probably due to selection associated with replication and transcription rather than translation (Ackermann and Chao 2006).

Population Variation

A variety of classes of molecular marker have been used to quantify genetic differences among individuals in *C. elegans* (Fitch and Thomas 1997; Barrière and Félix 2005b; Phillips 2006; Kammenga et al. 2008) (table 2). Here we focus on the patterns and conclusions derived primarily from genomic studies and formal population genetic analyses.

Nucleotide Polymorphism Patterns

Studies of single nucleotide polymorphism show that a random pair of homologous, nuclear noncoding sequences from a global sample of *C. elegans* will differ at roughly 1 in 350 nt on average (i.e., silent site $\pi_{si} \approx 0.003$; Graustein et al. 2002; Jovelin et al. 2003; Cutter 2006). Larger samples of the genome (with fewer among-strain comparisons) show SNPs at average densities of 1/840–8,750 bp relative to the N2 strain, not distinguishing between coding and noncoding regions (Koch et al. 2000; Wicks et al. 2001; Swan et al. 2002; Hillier et al. 2008). However, there is substantial heterogeneity in the levels of diversity across loci and strong dependence on whether global or local population samples are considered (Denver et al. 2003; Barrière and Félix 2005a; Stewart et al. 2005; Cutter 2006). In coding sequences, purifying selection causes the much lower levels of genetic variation observed per site for replacement sites relative to synonymous and noncoding sites (Thomas and Wilson 1991; Koch et al. 2000; Wicks et al. 2001; Graustein et al. 2002; Denver et al. 2003; Jovelin et al. 2003). Nucleotide polymorphism in *C. briggsae* is of a similar magnitude as for *C. elegans*, although the more striking population structure in *C. briggsae* gives rise to two to three times as many differences per kilobase among the haplotypes that derive from different latitudinal regions (Graustein et al. 2002; Cutter, Felix 2006; Hillier et al. 2007; Dolgin et al. 2008; Howe and Denver 2008). The obligately outcrossing *C. remanei*, on the other hand, harbors roughly 20-fold higher levels of diversity than either *C. elegans* or *C. briggsae* (Graustein et al. 2002; Jovelin et al. 2003; Haag and Ackerman 2005; Cutter, Baird, and Charlesworth 2006; Cutter 2008b), as first hinted from comparison of two *C. remanei* globin alleles (Kloek et al. 1996). If residual heterozygosity in the inbred strains used for sequencing is a reasonable guide, then *C. brenneri* populations might contain even more variation than *C. remanei* (and *C. japonica* possibly having less; Barriere et al. 2009), which needs to be documented explicitly in the future.

Indel differences between the canonical N2 strain of *C. elegans* and two strains collected on small islands (Hawaii, CB4856; Madeira, JU258) indicate that 1.5–2% of the N2 genome sequence is absent in the other two strains, including several hundred genes (Maydan et al. 2007). It remains to be seen whether the reciprocal is true and if this level of copy number variation is typical between most strains of *C. elegans*. Microsatellites, or STRs, provide a special case of indels whose evolution may be described by a two-phase mutation model (DiRienzo et al. 1994; Frisse 1999; Degtyareva et al. 2002; Phillips et al. 2009). Variation in STR length among individuals depends in part on the average length of the locus in *C. elegans* (Haber et al. 2005), as expected from the greater mutation rate of longer STRs (Frisse 1999; Seyfert et al. 2008; Phillips et al. 2009).

At a genomic scale, it is clear that sequence differences (SNPs and indels) are more prevalent in regions of high recombination (Koch et al. 2000; Maydan et al. 2007; Rockman and Kruglyak 2009), with up to 5-fold greater SNP density in high recombination regions (Cutter and Payseur 2003b). This association between genetic variation and recombination rate may be a consequence both of mutation rate heterogeneity and selection (Cutter and Payseur 2003b). Polymorphism in noncoding sequence also is higher in regions of low gene density, independently of the association with recombination, as predicted by models of selection at linked sites (Payseur and Nachman 2002; Cutter and Payseur 2003b), but could alternatively be explained by greater constraint in noncoding portions of gene-dense regions of chromosomes.

Linkage disequilibrium spans very wide distances in *C. elegans*, resulting in significant nonrandom allelic associations even between chromosomes (Koch et al. 2000; Barrière and Félix 2005b; Haber et al. 2005; Cutter 2006; Cutter, Baird, and Charlesworth 2006; Rockman and Kruglyak 2009). On average, it takes 3.3 Mb for linkage disequilibrium to decay by half, and some chromosome centers have no detectable decay in linkage disequilibrium (Rockman and Kruglyak 2009). This means that even very distantly separated loci will share the same genealogical history. Consequently, genetically effective recombination by outcrossing is estimated to be very rare in nature (Barrière and Félix 2005a; Cutter 2006). The slow decay of linkage disequilibrium with distance and the generally low level of sequence variation imply that "association

Table 2
Genomic or Population Studies of Genetic Differences among Conspecific Strains

Marker type	Species	Genomic Scope	Strains	Reference
AFLP	*Caenorhabditis elegans*	149 loci	55	Barrière and Félix (2005a)
Allozymes	*C. elegans*	24 loci	2	Butler et al. (1981)
Copy number variants	*C. elegans*	Whole-genome array	3	Maydan et al. (2007)
Microsatellites (repeat type)	*C. elegans*	~20 loci (trinucleotide)	2	Uitterlinden et al. (1989)
		20 loci (dinucleotide)	23	Sivasundar and Hey (2003)
		38 loci (mononucleotide)	23	Denver, Morris, Lynch, and Thomas (2004)
		10 loci (trinucleotide)	58	Haber et al. (2005)
		15 loci (dinucleotide)	69	Sivasundar and Hey (2005)
		2 loci (trinucleotide)	55	Barrière and Félix (2005a)
		6 loci (trinucleotide)	658	Barrière and Félix (2007)
	Caenorhabditis briggsae	32 loci (dinucleotide)	6	Phillips et al. (2009)
Restriction fragment length polymorphism	*C. elegans*	~50 loci	2	Emmons et al. (1979)
		310 kb	2	Rose et al. (1982)
		63 kb	2	Snutch and Baillie (1984)
SNPs (resequenced per strain)	*C. elegans*	2 loci (1.2 kb)	11	Thomas and Wilson (1991)
		35–373 kb	4	Koch et al. (2000)
		230; 55 (genotyping)	11; 23	Koch et al. (2000)
		5.4 Mb	2	Wicks et al. (2001)
		730 kb	2	Swan et al. (2002)
		4 loci (5 kb)	11–20	Graustein et al. (2002)
		55 kb (nuclear); 11 kb (mitochondrial)	27	Denver et al. (2003)
		1 locus (3 kb)	10	Jovelin et al. (2003)
		31 loci (2–5 kb?)	22	Stewart et al. (2005)
		1 locus (380 bp)	106	Sivasundar and Hey (2005)
		6 loci (4 kb)	118	Cutter (2006)
		74 Mb	2	Hillier et al. (2008)
		1,460 (genotyping)	125	Rockman and Kruglyak (2009)
		5 loci (17 kb)	13	Jovelin et al. (2009)
	Caenorhabditis briggsae	2 loci (1.2 kb)	2	Thomas and Wilson (1991)
		3 loci (2 kb)	6	Graustein et al. (2002)
		1 locus (2 kb)	4	Jovelin et al. (2003)
		6 loci (4 kb)	63	Cutter et al. (2006)
		143–9,970 sequence reads (63 kb–4.4 Mb)	5	Hillier et al. (2007)
		5 loci (17 kb)	5	Jovelin et al. (2009)
	Caenorhabditis remanei	3 loci (1.5 kb)	12	Graustein et al. (2002)
		1 locus (2 kb)	11	Jovelin et al. (2003)
		6 loci (4 kb)	34	Cutter, Baird, and Charlesworth 2006
		40 loci (30 kb)	16	Cutter (2008b)
		7 loci (21 kb)	14	Jovelin et al. (2009)
TEs	*C. elegans*	Tc1 patterns	12	Emmons et al. (1983)
		Tc1 patterns	4	Liao et al. (1983)
		35 Tc1 loci	5	Egilmez et al. (1995)
		Tc1 patterns	32	Hodgkin and Doniach (1997)
		32 Tc1 loci	16	Dolgin et al. (2008)
	C. remanei	16 mTcre1 loci	16	Dolgin et al. (2008)

mapping" the genetic basis of naturally occurring phenotypic variation using population data may not be feasible for many traits (Rockman and Kruglyak 2009), although *C. remanei* might prove to be a successful system (Jovelin et al. 2003), given a sufficiently high density of markers. High-throughput quantitative trait locus approaches using recombinant inbred lines also provide a fruitful alternative for relating natural genotypic to phenotypic variation (Li et al. 2006; Gutteling, Riksen, et al. 2007; Kammenga et al. 2008; Palopoli et al. 2008; Seidel et al. 2008; Rockman and Kruglyak 2009). Linkage disequilibrium decays rapidly in *C. remanei*, typically reaching background levels within 2 kb or less (Cutter, Baird, and Charlesworth 2006). Despite the high intralocus recombination implied by the rapid decay of linkage disequilibrium in *C. remanei*, the ratio of the population recombination parameter ($\rho = 4N_e r$) is small compared with the population mutation rate ($\theta = 4N_e\mu$), with $\rho/\theta = 0.16$ on average (Cutter 2008b).

Implications of Sequence Variation

Molecular markers have been applied to address several salient issues relating to *Caenorhabditis* demography, evolutionary history, and molecular evolution (Fitch 2005; Phillips 2006). Studies concur that *C. elegans* genetic variation is subdivided on a range of geographic scales but that the population structure does not give rise to strong isolation by distance (i.e., distant subpopulations are not necessarily more genetically differentiated than nearby subpopulations)—implying that migration occurs readily

over both short and long distances (Koch et al. 2000; Sivasundar and Hey 2003; Barrière and Félix 2005b, 2007; Haber et al. 2005; Sivasundar and Hey 2005; Cutter 2006). This contrasts with *C. briggsae*, in which global genetic variation is partitioned into at least three groups in association with latitude (Cutter et al. 2006; Dolgin et al. 2008), and with *C. remanei*, for which population structure is less pronounced (Cutter, Baird, and Charlesworth 2006; DeyA, Cutter AD, unpublished data). Notably, temperate region isolates of *C. briggsae* show very few differences from each other, suggesting that this species might have colonized or expanded recently in temperate latitudes (Cutter et al. 2006). The complex population structure of *C. elegans* makes it more difficult to infer whether population size changes have occurred, although microsatellite data are suggestive of global population contraction (Sivasundar and Hey 2003). Recent resampling of *C. elegans* at multiple sites over time has shown that some localities support persistent populations and that migration and recolonization renew the genetic composition of localities (Barrière and Félix 2007). It is hoped that ongoing efforts to quantify whole-genome sequence variation will reveal additional insights about *C. elegans* demographic history.

Most *C. elegans* population surveys also have confirmed that self-fertilization is the dominant mode of reproduction in nature, based on patterns of linkage disequilibrium, heterozygosity, or gene tree concordance (Egilmez et al. 1995; Koch et al. 2000; Denver et al. 2003; Barrière and Félix 2005b, 2007; Haber et al. 2005; Cutter 2006). Dinucleotide microsatellite heterozygosity and (lack of) linkage disequilibrium from a pair of studies suggested that selfing is less prevalent (Sivasundar and Hey 2003; Sivasundar and Hey 2005), although tri- and tetranucleotide microsatellites, sequences, and amplified fragment length polymorphism (AFLP) data for most of the same samples conflict with an interpretation of extensive recombination (Haber et al. 2005; Sivasundar 2005; Cutter 2006; Dey 2007). Nevertheless, heterozygosity-based estimates of the outcrossing rate tend to be higher than linkage disequilibrium–based estimates. This discrepancy can be reconciled by the recent implication of selection against recombinant genotypes, both in the laboratory and in nature (Barrière and Félix 2007; Dolgin et al. 2007; see also Rockman and Kruglyak 2009), although it is possible that complex demographic effects could explain a portion of the excess linkage disequilibrium in this species (Andolfatto and Przeworski 2000). Regardless, all these studies find that some recombination has occurred in the histories of *C. elegans* and *C. briggsae*, such that an effective outcrossing rate on the order of 10^{-3} to 10^{-4} per generation is most consistent with the breakdown of genetic associations across the genome in both species (Barrière and Félix 2005b, 2007; Cutter 2006; Cutter et al. 2006). The substantially higher outcrossing rate estimates of about 0.01 in *A. thaliana* lead to a more rapid decay of linkage disequilibrium (within ~50 kb) than observed for *C. elegans* (Nordborg et al. 2005). Such an extreme level of selfing in *C. elegans* and *C. briggsae* implies that the effective size of autosomes and the X chromosome will be essentially identical (with the X almost always diploid), so X-autosome differences are unlikely to reflect population genetic phenomena since the time that selfing became prevalent in these lineages.

Studies of diversity have estimated the effective population size (N_e, the size of an ideal Wright–Fisher population with equivalent statistical properties to the empirical population; Charlesworth 2009) of *C. elegans* to be $\sim 10^2$ to $\sim 10^4$ within a locality and $\sim 10^3$ to $\sim 10^5$ globally (Sivasundar and Hey 2003; Barrière and Félix 2005a; Denver et al. 2005; Cutter 2006). The effective size of *C. briggsae* in temperate latitudes is very small ($\sim 10^3$) but higher ($\sim 10^4$ to $\sim 10^5$) for the Tropics (Cutter et al. 2006). By contrast, the obligate outbreeding *C. remanei* has local effective population sizes of $\sim 10^6$ (Cutter, Baird, and Charlesworth 2006). The rather small effective population sizes of *C. elegans* and *C. briggsae*, despite potentially enormous census sizes, likely are a consequence of some combination of the effects of high selfing rates and little effective recombination, genome-wide background selection and/or genetic hitchhiking, and possibly extinction–recolonization population dynamics (Graustein et al. 2002; Sivasundar and Hey 2003; Barrière and Félix 2007). Because the efficacy of natural selection is directly proportional to N_e, and the influence of genetic drift is inversely proportional to N_e, it is clear that the fate of new mutations will have contrasting dynamics for *C. elegans* and *C. briggsae* relative to their gonochoristic (male and female) congeners.

Under complete selfing of individuals with Poisson-distributed variance in reproductive success, the effective population sizes of the mitochondrial and nuclear genomes are expected to be equal (Laporte and Charlesworth 2002; Wright, Nano, et al. 2008). This is due to all breeding individuals passing their mitochondria to offspring (i.e., lack of males) coupled with the classic effect of homozygosity under self-fertilization reducing by half the effective size of nuclear loci (Birky et al. 1983; Pollak 1987; Nordborg and Donnelly 1997). Thus, the disparity in sequence diversity between mitochondrial and nuclear loci at neutrally evolving sites in *C. elegans* should simply reflect the relative mutation rates (μ) of the two genomes, provided that selfing is the dominant mode of reproduction (i.e., $\theta = 4N\mu$, with $N_{mt} = N_{nuc}$, so $\theta_{mt}/\theta_{nuc} = \mu_{mt}/\mu_{nuc}$). Direct estimates of mitochondrial and nuclear mutation rates demonstrate a ~10-fold difference: $\mu_{mt} = 9.7 \times 10^{-8}$, $\mu_{nuc} = 9.0 \times 10^{-9}$ single-nucleotide mutations per site per generation (Denver et al. 2000; Denver, Morris, Kewalramani, et al. 2004; Keightley and Charlesworth 2005). Consequently, mitochondrial sequence is expected to have ~10 times the neutral polymorphism of nuclear loci, on average. Mitochondrial sequences do indeed exhibit higher polymorphism than nuclear loci (Thomas and Wilson 1991; Graustein et al. 2002; Denver et al. 2003). In a reanalysis of nearly complete mitochondrial genome sequences for a global sample of 27 *C. elegans* strains (Denver et al. 2003), we calculate that synonymous site diversity across mitochondrial genes is $\pi_{syn} = 0.031$ (95% confidence interval [CI] 0.012–0.072), where π is an estimator of θ (Nei and Li 1979). A comparable global sample of 24 strains indicates that nuclear diversity at silent sites averages $\pi_{si} = 0.0029$ (95% CI 0.0010–0.0071; data from Cutter 2006), resulting in almost exactly the 10-fold difference between mitochondrion and nucleus that is expected from the mutation rate difference. Similarly, tropical strains

of *C. briggsae* differ ~10-fold in mitochondrial and nuclear diversities (mean nuclear $\pi_{syn} = 0.00227$, mean mitochondrial $\pi_{syn} = 0.0219$; Cutter et al. 2006; Howe and Denver 2008). In the obligately outbreeding *C. remanei*, if diversity at COII is typical of mitochondrial polymorphism ($\pi_{syn} \sim 0.11$; Graustein et al. 2002), then the diversity at X-linked loci ($\pi_{si} \sim 0.035$; Cutter 2008b) also is consistent with the expected difference in diversity between mitochondrion and nucleus (~3.3-fold difference expected for gonochoristic *C. remanei*: $N_{mt} = 1/3\,N_x$, assuming an equal sex ratio and a 10-fold difference in μ, as for *C. elegans*). Thus, the previous observation of lower mitochondrial diversity in selfers than outcrossers relative to nuclear diversity in selfers versus outcrossers (Graustein et al. 2002) is fully consistent with neutral expectations.

The population structure evident for *C. elegans* and *C. briggsae*, coupled with relatively low levels of diversity and the highly selfing mode of reproduction, frustrates attempts to use polymorphism information to detect the localized action of natural selection in the genome. Nevertheless, some long blocks of chromosomes with little polymorphism among wild strains might reflect instances of selective sweeps (Rockman and Kruglyak 2009). The apparent panmixis and demographic equilibrium of *C. remanei* from Ohio, however, provides more promise for conducting genomic scans for selection in *Caenorhabditis* (Cutter 2008b). Such analyses would help elucidate the ongoing functional significance of the various genetic components that comprise the worm genome. Another difficulty is the saturated sequence divergence at synonymous sites seen for interspecific comparisons within *Caenorhabditis*, which precludes robust application of the powerful MK (McDonald and Kreitman 1991) and HKA tests (Hudson et al. 1987). Discovery of more closely related species would help rectify this problem; indeed, the recent identification of a new sister species of *C. briggsae* (*Caenorhabditis* sp. 9) is very promising (Félix M-A, Kiontke K, personal communication).

Transposable Elements

Quantification of population variation for TEs in *C. elegans* has focused primarily on the first TE class identified in this species, Tc1 (Emmons et al. 1983; Liao et al. 1983). On a broad scale, genomes of *C. elegans* individuals show "high" and "low" abundance of Tc1, as evidenced by the classic strains Bristol N2 with 32 copies of Tc1 in its genome and Bergerac BO with ~300 copies (Emmons et al. 1983; Liao et al. 1983; Fischer et al. 2003). This dichotomy was characterized further by delineating more subtle patterns of Tc1 presence or absence in a variety of other strains (Egilmez et al. 1995; Hodgkin and Doniach 1997) and recently with a more formal analysis of population frequencies (Dolgin et al. 2008). These studies concur that the Tc1 insertions found in N2 are commonly found in other strains, leading to a skew toward high population frequencies. This *C. elegans* pattern contrasts starkly with the pattern typical for most species, for which TE insertions occur at low frequency in populations, as for the mTcre1 element in *C. remanei* (Dolgin et al. 2008) and several TE types in *D. melanogaster* (Petrov et al. 2003) and *Arabidopsis lyrata* (Lockton et al. 2008). This result has been interpreted as relaxed selection against TEs in *C. elegans*, as a consequence of small effective population size (Dolgin et al. 2008). This pattern also is consistent with selfing versus outcrossing contrasts in plants (Wright et al. 2001; Tam et al. 2007), implicating selection against TE insertion rather than ectopic exchange as the dominant force in selfers (Morgan 2001). Differences in TE abundance between a few strains of *C. elegans* and of *C. briggsae* have been noted for some other TE classes (e.g., Harris et al. 1990; Cangiano and La Volpe 1993; Youngman et al. 1996; Tu and Shao 2002). Further work in this area should more thoroughly evaluate population variation for additional families of TEs in *C. elegans* and contrast population frequencies of homologous element types in different species of *Caenorhabditis*.

In the case of TE dynamics, it is also relevant to consider the elements within a given genome as a population. From this perspective, nucleotide variation among Tc1 insertions in the N2 strain indicates that the 32 copies are all very similar (Fischer et al. 2003) and exhibit a skewed distribution of variant sites toward an excess of rare SNPs (Dolgin et al. 2008). It is possible that homogenous Tc1 sequences could result from a recent expansion of Tc1 abundance in the genome or from gene conversion among different copies, if some template copies are used preferentially (Ohta 1985; Slatkin 1985; Brookfield 1986). Interestingly, *C. elegans* Tc1 insertions that are fixed in the species (or nearly so) have more sequence differences among them than do insertions that are polymorphic, suggesting that the fixed elements inserted longer ago and that transposition occurred more recently among those elements that are polymorphic within the species (Dolgin et al. 2008). Similarly low sequence variation among Tc3 copies also has been reported (Tu and Shao 2002), but this is not the case for all TE families (Witherspoon and Robertson 2003). Formal population genetic analysis of these other TE families may prove insightful about the dynamics of selfish genetic elements in selfing and outcrossing species.

Bridging Genotype and Phenotype

When trying to understand the evolutionary implications of *C. elegans* phenotypes and natural phenotypic variation, it is critical to consider the potential role of breeding regime as a selective environment for the genome: both in the highly selfing mode of current reproduction and the obligately outcrossing state of its ancestors (Cutter 2008c). Following the onset of a primarily self-fertilizing hermaphroditic lifestyle, *C. elegans* has experienced a drastic increase in homozygosity and reductions in effective recombination and effective population size—leading to a stronger role of genetic drift (and weaker role of selection) in determining the evolutionary fate of genetic variants. Selfing facilitates the purging of strongly deleterious mutations (through homozygosity), but selfing is refractory to the elimination of weakly deleterious mutations relative to outcrossing populations (Charlesworth et al. 1990). Selfing should also relax sexual selection, resulting in the loss or degradation of sexually selected traits (Cutter 2008c).

Consequently, the observation of a genetic basis to phenotypic variation does not require that adaptive evolution is its cause.

Take, for example, heritable variation in male sperm size among strains of *C. elegans*, where large sperm confer higher reproductive success (Lamunyon and Ward 1998). *Caenorhabditis elegans* male sperm are smaller than in related nematodes (Lamunyon and Ward 1999), sperm-related genes have faster rates of protein evolution (Cutter and Ward 2005; Artieri et al. 2008), and male copulatory success generally is poor (Chasnov and Chow 2002; Chasnov et al. 2007). Yet, there is heritable variation in male sperm size that responds to artificially induced sperm competition in the laboratory (Lamunyon and Ward 2002). All these characteristics—along with others, such as loss of copulatory plugging (Palopoli et al. 2008)—likely represent the by-product of relaxed postcopulatory sexual selection on males (Cutter 2008c).

As another example, it is plausible that differences in chemosensation among *C. elegans* strains (Jovelin et al. 2003), a trait that is controlled by large multigene families (Robertson and Thomas 2006), reflect divergence in paralog copy numbers and differential accumulation of nonsense mutations in paralogs as a consequence of relaxed selection (Robertson 1998, 2000; Stewart et al. 2005). More generally, it will be interesting to learn whether the expansion of chemoreceptor gene families in the *C. elegans* genome relative to *C. briggsae* is common to other lineages in the genus (Stein et al. 2003; Chen et al. 2005), whether the families expanded prior to the evolution of selfing in the *C. elegans* lineage, or if family expansion is a by-product of relaxed selection in *C. elegans* perhaps evincing a more recent origin of selfing in *C. briggsae* (Cutter et al. 2008). Nevertheless, some chemoreceptor loci might be subject to quite strong positive and purifying selection (Jovelin and Phillips 2005; Thomas et al. 2005), although population processes currently operating on such loci have not yet been fully elucidated (Jovelin et al. 2003).

Initial tests for heterosis in life history traits found no evidence for either heterozygote advantage or disadvantage in crosses involving the Bristol N2 strain of *C. elegans* (Johnson and Wood 1982; Johnson and Hutchinson 1993; Chasnov and Chow 2002). However, based on experiments with a number of recently isolated strains, *C. elegans* shows an overall tendency for outbreeding depression (i.e., heterozygote disadvantage; Dolgin et al. 2007). Genome-wide patterns of linkage disequilibrium and apparent selection against recombinants in natural populations are consistent with these laboratory data (Barrière and Félix 2007; Rockman and Kruglyak 2009), and the genetic basis for a major reproductive incompatibility that segregates in nature has now been characterized (Seidel et al. 2008). *C. briggsae*, on the other hand, shows evidence of some inbreeding depression (Dolgin et al. 2008), with the exception that a cross of strains AF16 and HK104 suggests the existence of genetic incompatability loci (Hillier et al. 2007; Dolgin et al. 2008; Baird S, personal communication). Even more extreme, the gonochoristic (male–female) *C. remanei* suffers very strong inbreeding depression (Dolgin et al. 2007), as do several other gonochoristic species in the genus (Félix M-A, personal communication). With the identification of new species of *Caenorhabditis* comes the potential to develop this genus into a system for studying reproductive isolation and speciation. To date, interspecific crosses have not yielded fertile hybrid progeny (Nigon and Dougherty 1949; Baird et al. 1992; Baird and Yen 2000; Hill and L'Hernault 2001; Baird 2002). However, the recent collection of one new species overcomes this practical hurdle in crosses with *C. briggsae* (Félix M-A, Baird S, Haag E, Cutter A, unpublished data).

In general, the low effective population size of *C. elegans* predicts that selection will need to be fairly strong to yield a detectable effect on the population genetic variation underlying trait variation, as seen for the toxin–antidote-like reproductive incompatibility region on chromosome I (Seidel et al. 2008). It remains an important open question as to whether heritable phenotypic variation reflects selective or neutral differences among individuals, particularly in traits with great potential ecological significance, such as those related to dauer development (Viney et al. 2003; Harvey et al. 2008), aggregation behavior (de Bono and Bargmann 1998; Gloria-Soria and Azevedo 2008; but see Rockman and Kruglyak 2009), response to pathogens (Schulenburg and Ewbank 2004; Schulenburg and Muller 2004), chemosensation (Jovelin et al. 2003; Stewart et al. 2005), RNAi sensitivity (Tijsterman et al. 2002), male mating and morphology (Hodgkin and Doniach 1997; Baer et al. 2005; Baird et al. 2005; Teotonio et al. 2006; Palopoli et al. 2008), and hermaphrodite fecundity (Hodgkin and Doniach 1997; Gutteling, Doroszuk, et al. 2007; Gutteling, Riksen, et al. 2007; Harvey and Viney 2007). In some cases, such as the genetic dimorphism underlying copulatory plug formation (Palopoli et al. 2008), relaxed sexual selection is the most plausible cause of segregating phenotypic variation (Cutter 2008c). How intraspecific phenotypic variation translates into interspecific phenotypic differentiation also is an important area of current and future inquiry (Fitch and Emmons 1995; Fitch 1997; Fitch and Thomas 1997; Baird 2001; Wang and Chamberlin 2004).

Conclusions and Prospective

Analysis of *C. elegans*' genome has revealed a variety of molecular evolutionary phenomena, some that are general across organisms and some that are peculiar to nematodes. Despite such intensive investigation, much about genome evolution remains unanswered in this model system. Nonrandom patterns of genome organization along the chromosomes have been known for decades; yet, the causal processes that generated many of these patterns remain elusive. Comparative genomic studies might help provide insights into the origin and maintenance of chromosomal disparities in rates of recombination, gene density, and repeat abundance among other features. Of particular, utility will be genome-wide comparisons between newly discovered species of *Caenorhabditis* that exhibit less sequence divergence than among currently described members of the genus (Felix MA, Kiontke K, personal communication) and population genomic studies of sequence variation within species. Outcrossing species of *Caenorhabditis*

make this an exceptional model genus to study the relationship between natural phenotypic variation and its genetic basis, whether it be due to single-nucleotide differences, copy number variants, or insertion/deletion differences among individuals. To realize the potential of these species, however, functional genomic assays must be performed in *C. elegans*' relatives to allow genome-wide comparisons in gene expression and RNAi effects for homologous loci. Few other eukaryotes have been subject to such detailed study of mutation as has *C. elegans*. Nevertheless, future research should help elucidate the patterns and processes of heterogeneity in mutation rates, mechanisms, and molecular properties—heterogeneity within the genome of *C. elegans*, variation that depends on an individual's genetic background, and differences among species. Similarly, much remains to be understood about variation in crossover and gene conversion across the genome and the roles this might play in shaping sequence variation and divergence.

Perhaps, the greatest hope for future research on the evolution of *C. elegans* genome lies not with genomics but in biodiversity; one of the most limiting elements of studying the evolution of *C. elegans*' genome is the lack of very closely related species known to science. Identification of multiple population samples for each species also will be integral to understanding natural phenotypic and genotypic differences, and the role that transitions in breeding system play in genomic dynamics. Comparative genomic analysis—within and between species—will further elucidate the selective and selectively neutral processes that govern evolutionary change in *Caenorhabditis* genomes.

Acknowledgments

We thank C. Baer, V. Katju, and A. Pires da Silva for sharing unpublished results. Critical comments by an anonymous reviewer and by S. Wright, V. Katju, and A. Pires da Silva on portions of the manuscript are much appreciated. A.D.C. is supported by a Canada Research Chair and an Natural Sciences and Engineering Research Council of Canada Discovery Grant.

Literature Cited

Abad P, Quiles C, Tares S, Piotte C, Castagnone-Sereno P, Abadon M, Dalmasso A. 1991. Sequences homologous to Tc(s) transposable elements of *Caenorhabditis elegans* are widely distributed in the phylum nematoda. J Mol Evol. 33:251–258.

Abrusan G, Krambeck HJ. 2006. Competition may determine the diversity of transposable elements. Theor Popul Biol. 70: 364–375.

Ackermann M, Chao L. 2006. DNA sequences shaped by selection for stability. PLoS Genet. 2:e22.

Agrawal AF, Wang AD. 2008. Increased transmission of mutations by low-condition females: evidence for condition-dependent DNA repair. PLoS Biol. 6:e30.

Ajie BC, Estes S, Lynch M, Phillips PC. 2005. Behavioral degradation under mutation accumulation in *Caenorhabditis elegans*. Genetics. 170:655–660.

Albertson DG, Thomson JN. 1982. The kinetochores of *Caenorhabditis elegans*. Chromosoma. 86:409–428.

Albertson DG, Thomson JN. 1993. Segregation of holocentric chromosomes at meiosis in the nematode, *Caenorhabditis elegans*. Chromosome Res. 1:15–26.

Andersson SG, Kurland CG. 1990. Codon preferences in free-living microorganisms. Microbiol Rev. 54:198–210.

Andolfatto P. 2001. Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev. 11:635–641.

Andolfatto P, Przeworski M. 2000. A genome-wide departure from the standard neutral model in natural populations of Drosophila. Genetics. 156:257–268.

Artieri CG, Haerty W, Gupta BP, Singh RS. 2008. Sexual selection and maintenance of sex: evidence from comparisons of rates of genomic accumulation of mutations and divergence of sex-related genes in sexual and hermaphroditic species of Caenorhabditis. Mol Biol Evol. 25:972–979.

Azevedo RBR, Keightley PD, Lauren-Maatta C, Vassilieva LL, Lynch M, Leroi AM. 2002. Spontaneous mutational variation for body size in *Caenorhabditis elegans*. Genetics. 162:755–765.

Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV. 2004. Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Res. 32:3724–3733.

Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. Nat Rev Genet. 8:619–631.

Baer CF, Phillips N, Ostrow D, Avalos A, Blanton D, Boggs A, Keller T, Levy L, Mezerhane E. 2006. Cumulative effects of spontaneous mutations for fitness in Caenorhabditis: role of genotype, environment and stress. Genetics. 174:1387–1395.

Baer CF, Shaw F, Steding C, et al. (11 co-authors). 2005. Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. Proc Natl Acad Sci USA. 102:5785–5790.

Baird S, Davidson C, Bohrer J. 2005. The genetics of ray pattern variation in *Caenorhabditis briggsae*. BMC Evol Biol. 5:3.

Baird SE. 2001. Strain-specific variation in the pattern of caudal papillae in *Caenorhabditis briggsae* (Nematoda: rhabditidae): implications for species identification. Nematology. 3: 373–376.

Baird SE. 2002. Haldane's Rule by sexual transformation in Caenorhabditis. Genetics. 161:1349–1353.

Baird SE, Sutherlin ME, Emmons SW. 1992. Reproductive isolation in Rhabditidae (Nematoda, Secernentea): mechanisms that isolate 6 species of 3 genera. Evolution. 46:585–594.

Baird SE, Yen W-C. 2000. Reproductive isolation in Caenorhabditis: terminal phenotypes of hybrid embryos. Evol Dev. 2:9–15.

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

Barrière A, Félix M-A. 2005a. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. Curr Biol. 15:1176–1184.

Barrière A, Félix M-A. 2005b. Natural variation and population genetics of *Caenorhabditis elegans* in Community TCeR. WormBook [On-line book]. [cited 30 Mar 2009]. Available from: http://www.wormbook.org.

Barrière A, Félix M-A. 2007. Temporal dynamics and linkage disequilibrium in natural *Caenorhabditis elegans* populations. Genetics. 176:999–1011.

Barriere A, Yang SP, Pekarek E, Thomas CG, Haag ES, Ruvinsky I. 2009. Detecting heterozygosity in shotgun genome assemblies: lessons from obligately outcrossing nematodes. Genome Res. 19:470–480.

Bartolome C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol. 19: 926–937.

Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP. 2003. Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. Development. 130:889–900.

Begin M, Schoen DJ. 2006. Low impact of germline transposition on the rate of mildly deleterious mutation in *Caenorhabditis elegans*. Genetics. 174:2129–2136.

Begin M, Schoen DJ. 2007. Transposable elements, mutational correlations, and population divergence in *Caenorhabditis elegans*. Evolution. 61:1062–1070.

Bessereau J-L. 2006. Transposons in *C. elegans in* Community *Ce*R. WormBook [On-line book]. [cited 30 Mar 2009]. Available from: http://www.wormbook.org.

Betancourt AJ, Presgraves DC. 2002. Linkage limits the power of natural selection in Drosophila. Proc Natl Acad Sci USA. 99:13616–13620.

Bierne N, Eyre-Walker A. 2006. Variation in synonymous codon use and DNA polymorphism within the Drosophila genome. J Evol Biol. 19:1–11.

Birdsell JA. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. Mol Biol Evol. 19:1181–1197.

Birky CW Jr, Maruyama T, Fuerst P. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics. 103:513–527.

Bloom JD, Adami C. 2003. Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. BMC Evol Biol. 3:1–10.

Blumenthal T. 2004. Operons in eukaryotes. Brief Funct Genomic Proteomic. 3:199–211.

Blumenthal T, Evans D, Link CD, et al. (11 co-authors). 2002. A global analysis of *Caenorhabditis elegans* operons. Nature. 417:851–854.

Blumenthal T, Gleason KS. 2003. *Caenorhabditis elegans* operons: form and function. Nat Rev Genet. 4:112–120.

Blumenthal T, Steward K. 1997. RNA processing and gene structure. In: Riddle DL, Blumenthal T, Meyer BJ, Priess JR, editors. C. elegans II. Cold Spring Harbor (NY): Cold Spring Harbor Laboratories Press. p. 117–145.

Boissinot S, Entezam A, Furano AV. 2001. Selection against deleterious LINE-1-containing loci in the human lineage. Mol Biol Evol. 18:926–935.

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

Brookfield JFY. 1986. A model for DNA sequence evolution within transposable element families. Genetics. 112:393–407.

Butler MH, Wall SM, Luehrsen KR, Fox GE, Hecht RM. 1981. Molecular relationships between closely related strains and species of nematodes. J Mol Evol. 18:18–23.

Byrne A, Weirauch M, Wong V, Koeva M, Dixon S, Stuart J, Roy P. 2007. A global analysis of genetic interactions in *Caenorhabditis elegans*. J Biol. 6:8.

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

Cangiano G, La Volpe A. 1993. Repetitive DNA sequences located in the terminal portion of the *Caenorhabditis elegans* chromosomes. Nucleic Acids Res. 21:1133–1139.

Carbone A, Zinovyev A, Kepes F. 2003. Codon adaptation index as a measure of dominating codon bias. Bioinformatics. 19: 2005–2015.

Carlini DB. 2004. Experimental reduction of codon bias in the Drosophila alcohol dehydrogenase gene results in decreased ethanol tolerance of adult flies. J Evol Biol. 17:779–785.

Carlini DB, Stephan W. 2003. In vivo introduction of unpreferred synonymous codons into the Drosophila *Adh* gene results in reduced levels of ADH protein. Genetics. 163:239–243.

Castillo-Davis CI, Bedford TBC, Hartl DL. 2004. Accelerated rates of intron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasites. Mol Biol Evol. 21:1422–1427.

Castillo-Davis CI, Hartl DL. 2002. Genome evolution and developmental constraint in *Caenorhabditis elegans*. Mol Biol Evol. 19:728–735.

Castillo-Davis CI, Hartl DL, Achaz G. 2004. *cis*-Regulatory and protein evolution in orthologous and duplicate genes. Genome Res. 14:1530–1536.

Castillo-Davis CI, Kondrashov FA, Hartl DL, Kulathinal RJ. 2004. The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. Genome Res. 14:802–811.

Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. 2002. Selection for short introns in highly expressed genes. Nat Genet. 31:415–418.

Cavalcanti AR, Ferreira R, Gu Z, Li WH. 2003. Patterns of gene duplication in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. J Mol Evol. 56:28–37.

Cavalcanti AR, Stover NA, Landweber LF. 2006. On the paucity of duplicated genes in *Caenorhabditis elegans* operons. J Mol Evol. 62:765–771.

Charlesworth B. 1990. Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet Res. 55:199–221.

Charlesworth B. 1992. Evolutionary rates in partially self-fertilizing species. Am Nat. 140:126–148.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 10:195–205.

Charlesworth B, Charlesworth D, Barton NH. 2003. The effects of genetic and geographic structure on neutral variation. Annu Rev Ecol Syst. 34:99–125.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. 134:1289–1303.

Charlesworth D. 2003. Effects of inbreeding on the genetic diversity of populations. Philos Trans R Soc Lond B Biol Sci. 358:1051–1070.

Charlesworth D, Morgan MT, Charlesworth B. 1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. Evolution. 44:1469–1489.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. Curr Opin Genet Dev. 11:685–690.

Chasnov JR, Chow KL. 2002. Why are there males in the hermaphroditic species *Caenorhabditis elegans*? Genetics. 160:983–994.

Chasnov JR, So WK, Chan CM, Chow KL. 2007. The species, sex, and stage specificity of a *Caenorhabditis* sex pheromone. Proc Natl Acad Sci USA. 104:6730–6735.

Chen J, Lewis EE, Carey JR, Caswell H, Caswell-Chen EP. 2006. The ecology and biodemography of *Caenorhabditis elegans*. Exp Gerontol. 41:1059–1065.

Chen N, Pai S, Zhao Z, Mah A, Newbury R, Johnsen RC, Altun Z, Moerman DG, Baillie DL, Stein LD. 2005. Identification of a nematode chemosensory gene family. Proc Natl Acad Sci USA. 102:146–151.

Chen N, Stein LD. 2006. Conservation and functional significance of gene topology in the genome of *Caenorhabditis elegans*. Genome Res. 16:606–617.

Cho S, Jin SW, Cohen A, Ellis RE. 2004. A phylogeny of Caenorhabditis reveals frequent loss of introns during nematode evolution. Genome Res. 14:1207–1220.

Clark AG, Eisen MB, Smith DR, et al. (417 co-authors). 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature. 450:203–218.
Coghlan A, Stajich JE, Harris TW. 2006. Comparative genomics in *C. elegans*, *C. briggsae*, and other *Caenorhabditis* species. In: Strange K, editor. C. elegans: *methods and applications*. Totowa (NJ): Humana Press Inc. p. 13–29.
Coghlan A, Wolfe KH. 2000. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. Yeast. 16:1131–1145.
Coghlan A, Wolfe KH. 2002. Fourfold faster rate of genome rearrangement in nematodes than in Drosophila. Genome Res. 12:857–867.
Coghlan A, Wolfe KH. 2004. Origins of recently gained introns in Caenorhabditis. Proc Natl Acad Sci USA. 101:11362–11367.
Coissac E, Maillier E, Netter P. 1997. A comparative study of duplications in bacteria and eukaryotes: the importance of telomeres. Mol Biol Evol. 14:1062–1074.
Comeron JM, Guthrie TB. 2005. Intragenic Hill-Robertson interference influences selection intensity on synonymous mutations in Drosophila. Mol Biol Evol. 22:2519–2530.
Consortium ICGS. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 432:695–716.
Coulombe-Huntington J, Majewski J. 2007a. Characterization of intron loss events in mammals. Genome Res. 17:23–32.
Coulombe-Huntington J, Majewski J. 2007b. Intron loss and gain in Drosophila. Mol Biol Evol. 24:2842–2850.
Cutter AD. 2006. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. Genetics. 172:171–184.
Cutter AD. 2008a. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. Mol Biol Evol. 25:778–786.
Cutter AD. 2008b. Multilocus patterns of polymorphism and selection across the X-chromosome of *Caenorhabditis remanei*. Genetics. 178:1661–1672.
Cutter AD. 2008c. Reproductive evolution: symptom of a selfing syndrome. Curr Biol. 18:R1056–R1058.
Cutter AD, Baird SE, Charlesworth D. 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. Genetics. 174:901–913.
Cutter AD, Charlesworth B. 2006. Selection intensity on preferred codons correlates with overall codon usage bias in *Caenorhabditis remanei*. Curr Biol. 16:2053–2057.
Cutter AD, Félix M-A, Barriere A, Charlesworth D. 2006. Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of *Caenorhabditis briggsae*. Genetics. 173:2021–2031.
Cutter AD, Good JM, Pappas CT, Saunders MA, Starrett DM, Wheeler TJ. 2005. Transposable element orientation bias in the *Drosophila melanogaster* genome. J Mol Evol. 61:733–741.
Cutter AD, Payseur BA. 2003a. Rates of deleterious mutation and the evolution of sex in *Caenorhabditis*. J Evol Biol. 16: 812–822.
Cutter AD, Payseur BA. 2003b. Selection at linked sites in the partial selfer *Caenorhabditis elegans*. Mol Biol Evol. 20:665–673.
Cutter AD, Payseur BA, Salcedo T, et al. (12 co-authors). 2003. Molecular correlates of genes exhibiting RNAi phenotypes in *Caenorhabditis elegans*. Genome Res. 13:2651–2657.
Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. Mol Biol Evol. 22:178–188.
Cutter AD, Wasmuth J, Blaxter ML. 2006. The evolution of biased codon and amino acid usage in nematode genomes. Mol Biol Evol. 23:2303–2315.
Cutter AD, Wasmuth JD, Washington NL. 2008. Patterns of molecular evolution in Caenorhabditis preclude ancient origins of selfing. Genetics. 178:2093–2104.
Davies EK, Peters AD, Keightley PD. 1999. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. Science. 285:1748–1751.
Davis JC, Brandman O, Petrov DA. 2005. Protein evolution in the context of Drosophila development. J Mol Evol. 60:774–785.
Davis JC, Petrov DA. 2004. Preferential duplication of conserved proteins in eukaryotic genomes. PLoS Biol. 2:E55.
de Bono M, Bargmann CI. 1998. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. Cell. 94:679–689.
de Visser JA, Elena SF. 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. Nat Rev Genet. 8:139–149.
Dean MD, Good JM, Nachman MW. 2008. Adaptive evolution of proteins secreted during sperm maturation: an analysis of the mouse epididymal transcriptome. Mol Biol Evol. 25:383–392.
Dechering KJ, Cuelenaere K, Konings RN, Leunissen JA. 1998. Distinct frequency-distributions of homopolymeric DNA tracts in different genomes. Nucleic Acids Res. 26:4056–4062.
Degtyareva NP, Greenwell P, Hofmann ER, Hengartner MO, Zhang L, Culotti JG, Petes TD. 2002. *Caenorhabditis elegans* DNA mismatch repair gene *msh-2* is required for microsatellite stability and maintenance of genome integrity. Proc Natl Acad Sci USA. 99:2158–2163.
Denver DR, Feinberg S, Steding C, Durbin M, Lynch M. 2006. The relative roles of three DNA repair pathways in preventing *Caenorhabditis elegans* mutation accumulation. Genetics. 174:57–65.
Denver DR, Morris K, Kewalramani A, Harris KE, Chow A, Estes S, Lynch M, Thomas WK. 2004. Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. J Mol Evol. 58:584–595.
Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature. 430: 679–682.
Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science. 289:2342–2344.
Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. Nat Genet. 37:544–548.
Denver DR, Morris K, Thomas WK. 2003. Phylogenetics in *Caenorhabditis elegans*: an analysis of divergence and outcrossing. Mol Biol Evol. 20:393–400.
Deutsch M, Long M. 1999. Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res. 27:3219–3228.
Dey A. 2007. Genetic variation and population structure in wild isolates of *Caenorhabditis elegans* collected from California. Microbiology and molecular genetics. Piscataway (NJ): Rutgers University. p. 41.
Dieringer D, Schlotterer C. 2003. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Res. 13: 2242–2251.

DiRienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA. 91:3166–3170.

Dolgin ES, Charlesworth B, Baird SE, Cutter AD. 2007. Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. Evolution. 61:1339–1352.

Dolgin ES, Charlesworth B, Cutter AD. 2008. Population frequencies of transposable elements in selfing and outcrossing Caenorhabditis nematodes. Genet Res. 90:317–329.

Dolgin ES, Félix M-A, Cutter AD. 2008. Hakuna nematoda: genetic and phenotypic diversity in African isolates of *Caenorhabditis elegans* and *C. briggsae*. Heredity. 100:304–315.

dos Reis M, Wernisch L. 2009. Estimating translational selection in eukaryotic genomes. Mol Biol Evol. 26:451–461.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics. 148:1667–1686.

Dunham I, Shimizu N, Roe BA, et al. (221 co-authors). 1999. The DNA sequence of human chromosome 22. Nature. 402:489–495.

Duret L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 16:287–289.

Duret L. 2002. Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev. 12:640–649.

Duret L, Marais G, Biemont C. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. Genetics. 156:1661–1669.

Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, Arabidopsis. Proc Natl Acad Sci USA. 96:4482–4487.

Echeverri CJ, Beachy PA, Baum B, et al. (22 co-authors). 2006. Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat Meth. 3:777–779.

Egilmez NK, Ebert RH, Reis RJS. 1995. Strain evolution in *Caenorhabditis elegans*: transposable elements as markers of interstrain evolutionary history. J Mol Evol. 40:372–381.

Eichler EE, Sankoff D. 2003. Structural dynamics of eukaryotic chromosome evolution. Science. 301:793–797.

Emmons SW. 1988. The genome. In: Wood WB, editor. *The nematode* Caenorhabditis elegans. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 47–79.

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

Emmons SW, Yesner L, Ruan KS, Katzenberg D. 1983. Evidence for a transposon in *Caenorhabditis elegans*. Cell. 32:55–65.

Ermolaeva MD, Wu M, Eisen JA, Salzberg SL. 2003. The age of the *Arabidopsis thaliana* genome duplication. Plant Mol Biol. 51:859–866.

Eskesen ST, Eskesen FN, Kinghorn B, Ruvinsky A. 2004. Periodicity of DNA in exons. BMC Mol Biol. 5:12.

Eskesen ST, Eskesen FN, Ruvinsky A. 2004. Natural selection affects frequencies of AG and GT dinucleotides at the 5' and 3' ends of exons. Genetics. 167:543–550.

Estes S, Ajie BC, Lynch M, Phillips PC. 2005. Spontaneous mutational correlations for life-history, morphological and behavioral characters in *Caenorhabditis elegans*. Genetics. 170:645–653.

Estes S, Lynch M. 2003. Rapid fitness recovery in mutationally degraded lines of *Caenorhabditis elegans*. Evolution. 57: 1022–1030.

Estes S, Phillips PC, Denver DR, Thomas WK, Lynch M. 2004. Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in *Caenorhabditis elegans*. Genetics. 166:1269–1279.

Evans D, Zorio D, MacMorris M, Winter CE, Lea K, Blumenthal T. 1997. Operons and SL2 trans-splicing exist in nematodes outside the genus Caenorhabditis. Proc Natl Acad Sci USA. 94:9751–9756.

Eyre-Walker A. 1996. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? Mol Biol Evol. 13:864–872.

Fedorov A, Roy S, Fedorova L, Gilbert W. 2003. Mystery of intron gain. Genome Res. 13:2236–2241.

Fedorov A, Saxonov S, Gilbert W. 2002. Regularities of context-dependent codon bias in eukaryotic genes. Nucleic Acids Res. 30:1192–1197.

Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature. 391:806–811.

Fischer G, Neuveglise C, Durrens P, Gaillardin C, Dujon B. 2001. Evolution of gene order in the genomes of two related yeast species. Genome Res. 11:2009–2019.

Fischer SE, Wienholds E, Plasterk RH. 2003. Continuous exchange of sequence information between dispersed Tc1 transposons in the *Caenorhabditis elegans* genome. Genetics. 164:127–134.

Fitch DHA. 1997. Evolution of male tail development in rhabditid nematodes related to *Caenorhabditis elegans*. Syst Biol. 46:145–179.

Fitch DHA. 2005. Evolution: an ecological context for *C. elegans*. Curr Biol. 15:R655–R658.

Fitch DHA, Emmons SW. 1995. Variable cell positions and cell contacts underlie morphological evolution of the rays in the male tails of nematodes related to *Caenorhabditis elegans*. Dev Biol. 170:564–582.

Fitch DHA, Thomas WK. 1997. Evolution. In: Riddle DL, Blumenthal T, Meyer BJ, Priess JR, editors. C. elegans *II*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 815–850.

Francino MP, Ochman H. 2001. Deamination as the basis of strand-asymmetric evolution in transcribed *Escherichia coli* sequences. Mol Biol Evol. 18:1147–1150.

Fraser HB, Hirsh AE. 2004. Evolutionary rate depends on number of protein-protein interactions independently of gene expression level. BMC Evol Biol. 4.

Friedman R, Hughes AL. 2001. Gene duplication and the structure of eukaryotic genomes. Genome Res. 11: 373–381.

Frisse L. 1999. *Understanding the mechanisms of microsatellite formation and mutation using the model organism* Caenorhabditis elegans. Kansas City (MO): Molecular Biology and Biochemistry, University of Missouri.

Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen SP, Grunnet M, Jorgensen EM. 2008. Single-copy insertion of transgenes in *Caenorhabditis elegans*. Nat Genet. 40:1375–1383.

Fukuoka Y, Inaoka H, Kohane I. 2004. Inter-species differences of co-expression of neighboring genes in eukaryotic genomes. BMC Genomics. 5:4.

Ganko EW, Bhattachajee V, Schliekelman P, McDonald JF. 2003. Evidence for the contribution of LTR retrotransposons to *C. elegans* gene evolution. Mol Biol Evol. 20: 1925–1931.

Ganko EW, Fielman KT, McDonald JF. 2001. Evolutionary history of Cer elements and their impact on the *C. elegans* genome. Genome Res. 11:2066–2074.

Gaudet J, Muttumu S, Horner M, Mango SE. 2004. Whole-genome analysis of temporal gene expression during foregut development. PLoS Biol. 2:e352.

Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP. 2006. Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature. 440:242–245.

Gloria-Soria A, Azevedo RB. 2008. *npr-1* regulates foraging and dispersal strategies in *Caenorhabditis elegans*. Curr Biol. 18:1694–1699.

Goodstadt L, Heger A, Webber C, Ponting CP. 2007. An analysis of the gene complement of a marsupial, *Monodelphis domestica*: evolution of lineage-specific genes and giant chromosomes. Genome Res. 17:969–981.

Gotoh O. 1998. Divergent structures of *Caenorhabditis elegans* cytochrome P450 genes suggest the frequent loss and gain of introns during the evolution of nematodes. Mol Biol Evol. 15:1447–1459.

Graustein A, Gaspar JM, Walters JR, Palopoli MF. 2002. Levels of DNA polymorphism vary with mating system in the nematode genus Caenorhabditis. Genetics. 161:99–107.

Greeff JM, Michiels NK. 1999. Low potential for sexual selection in simultaneously hermaphroditic animals. Proc R Soc Lond B Biol Sci. 266:1671–1676.

Gu ZL, Cavalcanti A, Chen FC, Bouman P, Li WH. 2002. Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. Mol Biol Evol. 19:256–262.

GuhaThakurta D, Schriefer LA, Waterston RH, Stormo GD. 2004. Novel transcription regulatory elements in *Caenorhabditis elegans* muscle genes. Genome Res. 14:2457–2468.

Guiliano DB, Blaxter ML. 2006. Operon conservation and the evolution of trans-splicing in the phylum Nematoda. PLoS Genet. 2:e198.

Gutteling EW, Doroszuk A, Riksen JA, Prokop Z, Reszka J, Kammenga JE. 2007. Environmental influence on the genetic correlations between life-history traits in *Caenorhabditis elegans*. Heredity. 98:206–213.

Gutteling EW, Riksen JA, Bakker J, Kammenga JE. 2007. Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. Heredity. 98:28–37.

Haag ES. 2005. The evolution of nematode sex determination: c. elegans as a reference point for comparative biology *in* Community TCeR. WormBook [On-line book]. [cited 30 Mar 2009]. Available from: http://www.wormbook.org.

Haag ES, Ackerman AD. 2005. Intraspecific variation in *fem-3* and *tra-2*, two rapidly coevolving nematode sex-determining genes. Gene. 349:35–42.

Haag ES, Chamberlin H, Coghlan A, Fitch DH, Peters AD, Schulenburg H. 2007. Caenorhabditis evolution: if they all look alike, you aren't looking hard enough. Trends Genet. 23:101–104.

Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. PLoS Biol. 6:e204.

Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Charlesworth B, Keightley PD. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature. 445:82–85.

Haber M, Schungel M, Putz A, Muller S, Hasert B, Schulenburg H. 2005. Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding. Mol Biol Evol. 22:160–173.

Haerty W, Artieri C, Khezri N, Singh RS, Gupta BP. 2008. Comparative analysis of function and interaction of transcription factors in nematodes: extensive conservation of orthology coupled to rapid sequence evolution. BMC Genomics. 9:399.

Hannay K, Marcotte E, Vogel C. 2008. Buffering by gene duplicates: an analysis of molecular correlates and evolutionary conservation. BMC Genomics. 9:609.

Harris LJ, Prasad S, Rose AM. 1990. Isolation and sequence analysis of *Caenorhabditis briggsae* repetitive elements related to the *Caenorhabditis elegans* transposon Tc1. J Mol Evol. 30:359–369.

Harrison PM, Echols N, Gerstein MB. 2001. Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome. Nucleic Acids Res. 29:818–830.

Harvey SC, Shorto A, Viney ME. 2008. Quantitative genetic analysis of life-history traits of *Caenorhabditis elegans* in stressful environments. BMC Evol Biol. 8:15.

Harvey SC, Viney ME. 2007. Thermal variation reveals natural variation between isolates of *Caenorhabditis elegans*. J Exp Zool. 308B:409–416.

He H, Wang J, Liu T, et al. (15 co-authors). 2007. Mapping the *C. elegans* noncoding transcriptome with a whole-genome tiling microarray. Genome Res. 17:1471–1477.

Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G, Brown EL. 2000. Genomic analysis of gene expression in *C. elegans*. Science. 290:809–812.

Hill KL, L'Hernault SW. 2001. Analyses of reproductive interactions that occur after heterospecific matings within the genus Caenorhabditis. Dev Biol. 232:105–114.

Hill WG, Robertson A. 1966. Effect of linkage on limits to artificial selection. Genet Res. 8:269–294.

Hillier LW, Marth GT, Quinlan AR, et al. (20 co-authors). 2008. Whole-genome sequencing and variant discovery in *C. elegans*. Nat Methods. 5:183–188.

Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, Waterston RH. 2007. Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. 5:e167.

Hodgkin J, Doniach T. 1997. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. Genetics. 146:149–164.

Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst. 39:21–42.

Hoogewijs D, De Henau S, Dewilde S, Moens L, Couvreur M, Borgonie G, Vinogradov S, Roy S, Vanfleteren J. 2008. The Caenorhabditis globin gene family reveals extensive nematode-specific radiation and diversification. BMC Evol Biol. 8:279.

Howe DK, Denver DR. 2008. Muller's ratchet and compensatory mutation in *Caenorhabditis briggsae* mitochondrial genome evolution. BMC Evol Biol. 8:62.

Huang P, Pleasance ED, Maydan JS, Hunt-Newbury R, O'Neil NJ, Mah A, Baillie DL, Marra MA, Moerman DG, Jones SJ. 2007. Identification and analysis of internal promoters in *Caenorhabditis elegans* operons. Genome Res. 17:1478–1485.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.

Hurst LD, Pal C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 5:299–310.

Ikemura T. 1985. Codon usage and transfer-RNA content in unicellular and multicellular organisms. Mol Biol Evol. 2:13–34.

Ingvarsson PK. 2007. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. Mol Biol Evol. 24:836–844.

Irimia M, Rukov JL, Penny D, Garcia-Fernandez J, Vinther J, Roy SW. 2008. Widespread evolutionary conservation of alternatively spliced exons in Caenorhabditis. Mol Biol Evol. 25:375–382.

Jaillon O, Aury JM, Brunet F, et al. (61 co-authors). 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature. 431:946–957.

Jeffares DC, Mourier T, Penny D. 2006. The biology of intron gain and loss. Trends Genet. 22:16–22.

Johnsen RC, Jones SJ, Rose AM. 2000. Mutational accessibility of essential genes on chromosome I(left) in *Caenorhabditis elegans*. Mol Gen Genet. 263:239–252.

Johnson TE, Hutchinson EW. 1993. Absence of strong heterosis for life-span and other life-history traits in *Caenorhabditis elegans*. Genetics. 134:465–474.

Johnson TE, Wood WB. 1982. Genetic-analysis of life-span in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 79:6603–6607.

Jordan IK, Wolf YI, Koonin EV. 2003. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol. 3:1.

Jovelin R, Ajie BC, Phillips PC. 2003. Molecular evolution and quantitative variation for chemosensory behaviour in the nematode genus Caenorhabditis. Mol Ecol. 12:1325–1337.

Jovelin R, Dunham JP, Sung FS, Phillips PC. Forthcoming. 2008. High nucleotide divergence in developmental regulatory genes contrasts with the structural elements of olfactory pathways in Caenorhabditis. Genetics. 181:

Jovelin R, Phillips PC. 2005. Functional constraint and divergence in the G protein family in *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Mol Genet Genom. 273:299–310.

Kamath RS, Fraser AG, Dong Y, et al. (13 co-authors). 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature. 421:231–237.

Kammenga JE, Phillips PC, De Bono M, Doroszuk A. 2008. Beyond induced mutants: using worms to study natural variation in genetic pathways. Trends Genet. 24:178–185.

Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. J Mol Evol. 53:290–298.

Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene. 238:143–155.

Katju V, LaBeau EM, Lipinski KJ, Bergthorsson U. 2008. Sex change by gene conversion in a *Caenorhabditis elegans fog-2* mutant. Genetics. 180:669–672.

Katju V, Lynch M. 2003. The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. Genetics. 165:1793–1803.

Katju V, Lynch M. 2006. On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. Mol Biol Evol. 23:1056–1067.

Katti MV, Ranjekar PK, Gupta VS. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol. 18:1161–1167.

Keightley PD, Bataillon TM. 2000. Multigeneration maximum-likelihood analysis applied to mutation-accumulation experiments in *Caenorhabditis elegans*. Genetics. 154: 1193–1201.

Keightley PD, Caballero A. 1997. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 94:3823–3827.

Keightley PD, Charlesworth B. 2005. Genetic instability of *C. elegans* comes naturally. Trends Genet. 21:67–70.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature. 428:617–624.

Kelly WG, Schaner CE, Dernburg AF, Lee MH, Kim SK, Villeneuve AM, Reinke V. 2002. X-chromosome silencing in the germline of *C. elegans*. Development. 129:479–492.

Kennedy BP, Aamodt EJ, Allen FL, Chung MA, Heschl MFP, McGhee JD. 1993. The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J Mol Biol. 229:890–908.

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

Kidwell MG, Lisch D. 1997. Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA. 94:7704–7711.

Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS. 2001. A gene expression map for *Caenorhabditis elegans*. Science. 293:2087–2092.

Kimura M. 1968. Evolutionary rate at molecular level. Nature. 217:624–626.

Kiontke K, Barriere A, Kolotuev I, Podbilewicz B, Sommer R, Fitch DH, Felix MA. 2007. Trends, stasis, and drift in the evolution of nematode vulva development. Curr Biol. 17: 1925–1937.

Kiontke K, Gavin NP, Raynes Y, Roehrig C, Piano F, Fitch DHA. 2004. *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc Natl Acad Sci USA. 101:9003–9008.

Kirouac M, Sternberg PW. 2003. *cis*-Regulatory control of three cell fate-specific genes in vulval organogenesis of *Caenorhabditis elegans* and *C. briggsae*. Dev Biol. 257: 85–103.

Kliman RM, Irving N, Santiago M. 2003. Selection conflicts, gene expression, and codon usage trends in yeast. J Mol Evol. 57:98–109.

Kloek AP, McCarter JP, Setterquist RA, Schedl T, Goldberg DE. 1996. Caenorhabditis globin genes: rapid intronic divergence contrasts with conservation of silent exonic sites. J Mol Evol. 43:101–108.

Knowles DG, McLysaght A. 2006. High rate of recent intron gain and loss in simultaneously duplicated Arabidopsis genes. Mol Biol Evol. 23:1548–1557.

Koch R, van Luenen HGAM, van der Horst M, Thijssen KL, Plasterk RHA. 2000. Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. Genome Res. 10:1690–1696.

Kondrashov AS. 1985. Deleterious mutations as an evolutionary factor. 2. Facultative apomixis and selfing. Genetics. 111: 635–653.

Kondrashov AS, Crow JF. 1993. A molecular approach to estimating the human deleterious mutation-rate. Hum Mutat. 2:229–234.

Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. Genome Biol. 3:research0008.0001.

Kreitman M. 2000. Methods to detect selection in populations with applications to the human. Annu Rev Genomics Hum Genet. 1:539–559.

Lamunyon CW, Ward S. 1998. Larger sperm outcompete smaller sperm in the nematode *Caenorhabditis elegans*. Proc R Soc Lond B Biol Sci. 265:1997–2002.

Lamunyon CW, Ward S. 1999. Evolution of sperm size in nematodes: sperm competition favours larger sperm. Proc R Soc Lond B Biol Sci. 266:263–267.

Lamunyon CW, Ward S. 2002. Evolution of larger sperm in response to experimentally increased sperm competition in *Caenorhabditis elegans*. Proc R Soc Lond B Biol Sci. 269:1125–1128.

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

Laporte V, Charlesworth B. 2002. Effective population size and population subdivision in demographically structured populations. Genetics. 162:501–519.

Lawrence J. 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr Opin Genet Dev. 9:642–648.

Lee KZ, Sommer RJ. 2003. Operon structure and trans-splicing in the nematode *Pristionchus pacificus*. Mol Biol Evol. 20:2097–2103.

Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. 2006. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat Genet. 38:896–903.

Lemire B. 2005. Mitochondrial genetics *in* Community TCeR. WormBook [On-line book]. [cited 30 Mar 2009]. Available from: http://www.wormbook.org.

Lerat E, Capy P, Biemont C. 2002. Codon usage by transposable elements and their host genes in five species. J Mol Evol. 54:625–637.

Lercher MJ, Blumenthal T, Hurst LD. 2003. Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. Genome Res. 13:238–243.

Li Y, Alvarez OA, Gutteling EW, et al. (12 co-authors). 2006. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. PLoS Genetics. 2:e222.

Liao LW, Rosenzweig B, Hirsh D. 1983. Analysis of a transposable element in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 80:3585–3589.

Lim JG, Stine RR, Yanowitz JL. 2008. Domain-specific regulation of recombination in *C. elegans* in response to temperature, age, and sex. Genetics. 180:715–726.

Lockton S, Ross-Ibarra J, Gaut BS. 2008. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. Proc Natl Acad Sci USA. 105:13965–13970.

Loewe L, Charlesworth B. 2007. Background selection in single genes may explain patterns of codon bias. Genetics. 175: 1381–1393.

Loewe L, Cutter AD. 2008. On the potential for extinction by Muller's ratchet in *Caenorhabditis elegans*. BMC Evol Biol. 8:125.

Logsdon JM Jr, Tyshenko MG, Dixon C, D-Jafari J, Walker VK, Palmer JD. 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. Proc Natl Acad Sci USA. 92:8507–8511.

Lynch M. 2002a. Gene duplication and evolution. Science. 297:945–947.

Lynch M. 2002b. Intron evolution as a population-genetic process. Proc Natl Acad Sci USA. 99:6118–6123.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA. 104(1 Suppl):8597–8604.

Lynch M, Blanchard J, Houle D, Kibota T, Schultz S, Vassilieva L, Willis J. 1999. Perspective: spontaneous deleterious mutation. Evolution. 53:645–663.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science. 290:1151–1155.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics. 154: 459–473.

Lynch M, Kewalramani A. 2003. Messenger RNA surveillance and the evolutionary proliferation of introns. Mol Biol Evol. 20:563–571.

Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. Trends Genet. 19:330–338.

Marais G, Charlesworth B, Wright SI. 2004. Recombination and base composition: the case of the highly self-fertilizing plant *Arabidopsis thaliana*. Genome Biol. 5:R45.

Marais G, Domazet-Loso T, Tautz D, Charlesworth B. 2004. Correlated evolution of synonymous and nonsynonymous sites in Drosophila. J Mol Evol. 59:771–779.

Marais G, Duret L. 2001. Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. J Mol Evol. 52:275–280.

Marais G, Mouchiroud D, Duret L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc Natl Acad Sci USA. 98:5688–5692.

Marais G, Piganeau G. 2002. Hill-Robertson interference is a minor determinant of variations in codon bias across *Drosophila melanogaster* and *Caenorhabditis elegans* genomes. Mol Biol Evol. 19:1399–1406.

Martin E, Laloux H, Couette G, Alvarez T, Bessou C, Hauser O, Sookhareea S, Labouesse M, Segalat L. 2002. Identification of 1088 new transposon insertions of *Caenorhabditis elegans*: a pilot study toward large-scale screens. Genetics. 162: 521–524.

Maydan JS, Flibotte S, Edgley ML, Lau J, Selzer RR, Richmond TA, Pofahl NJ, Thomas JH, Moerman DG. 2007. Efficient high-resolution deletion discovery in *Caenorhabditis elegans* by array comparative genomic hybridization. Genome Res. 17:337–347.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in Drosophila. Nature. 351:652–654.

Medawar PB. 1952. An unsolved problem of biology. London: H.K. Lewis & Co.

Merritt C, Rasoloson D, Ko D, Seydoux G. 2008. 3′ UTRs are the primary regulators of gene expression in the *C. elegans* germline. Curr Biol. 18:1476–1482.

Miller MA, Cutter AD, Yamamoto I, Ward S, Greenstein D. 2004. Clustered organization of reproductive genes in the *C. elegans* genome. Curr Biol. 14:1284–1290.

Moerman DG, Baillie DL. 1979. Genetic organization in *Caenorhabditis elegans*: fine-structure analysis of the *unc-22* gene. Genetics. 91:95–103.

Morgan MT. 2001. Transposable element number in mixed mating populations. Genet Res. 77:261–275.

Mounsey A, Bauer P, Hope IA. 2002. Evidence suggesting that a fifth of annotated *Caenorhabditis elegans* genes may be pseudogenes. Genome Res. 12:770–775.

Mukai T. 1964. Genetic structure of natural populations of *Drosophila melanogaster*.1. Spontaneous mutation rate of polygenes controlling viability. Genetics. 50:1–9.

Muller HJ. 1964. The relation of recombination to mutational advance. Mutat Res. 1:2–9.

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

Nei M, Li WH. 1979. Mathematical-model for studying genetic-variation in terms of restriction endonucleases. Proc Natl Acad Sci USA. 76:5269–5273.
Nguyen DQ, Webber C, Ponting CP. 2006. Bias of selection on human copy-number variants. PLoS Genet. 2:e20.
Nigon V. 1949. Les modalites de la reproduction et le determinisme du sexe chez quelques nematodes libres. Ann Sci Natur Zool Biol Anim. 11:1–132.
Nigon V, Dougherty EC. 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.
Nikolaidis N, Nei M. 2004. Concerted and nonconcerted evolution of the Hsp70 gene superfamily in two sibling species of nematodes. Mol Biol Evol. 21:498–505.
Nishant KT, Rao MRS. 2005. Molecular features of meiotic recombination hot spots. Bioessays. 28:45–56.
Nordborg M, Donnelly P. 1997. The coalescent process with selfing. Genetics. 146:1185–1195.
Nordborg M, Hu TT, Ishino Y, et al. (24 co-authors). 2005. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. 3:e196.
Nuzhdin SV. 1999. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica. 107:129–137.
O'Rourke D, Baban D, Demidova M, Mott R, Hodgkin J. 2006. Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. Genome Res. 16:1005–1016.
Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag.
Ohta T. 1985. A model of duplicative transposition and gene conversion for repetitive DNA families. Genetics. 110:513–524.
Ohta T, Kimura M. 1973. Model of mutation appropriate to estimate number of electrophoretically detectable alleles in a finite population. Genet Res. 22:201–204.
Ostrow D, Phillips N, Avalos A, Blanton D, Boggs A, Keller T, Levy L, Rosenbloom J, Baer CF. 2007. Mutational bias for body size in rhabditid nematodes. Genetics. 176: 1653–1661.
Palopoli MF, Rockman MV, Tinmaung A, Ramsay C, Curwen S, Aduna A, Laurita J, Kruglyak L. 2008. Molecular basis of the copulatory plug polymorphism in *Caenorhabditis elegans*. Nature. 454:1019–1022.
Parkinson J, Mitreva M, Whitton C, et al. (12 co-authors). 2004. A transcriptomic analysis of the phylum nematoda. Nat Genet. 36:1259–1267.
Parkinson J, Whitton C, Schmid R, Thomson M, Blaxter M. 2004. NEMBASE: a resource for parasitic nematode ESTs. Nucleic Acids Res. 32:D427–D430.
Payseur BA, Nachman MW. 2002. Gene density and human nucleotide polymorphism. Mol Biol Evol. 19:336–340.
Percudani R. 2001. Restricted wobble rules for eukaryotic genomes. Trends Genet. 17:133–135.
Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in Drosophila. Mol Biol Evol. 20:880–892.
Phillips N, Salomon M, Custer A, Ostrow D, Baer CF. 2009. Spontaneous mutational and standing genetic (co)variation at dinucleotide microsatellites in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. Mol Biol Evol. 26:659–669.
Phillips PC. 2006. One perfect worm. Trends Genet. 22:405–407.
Plasterk RH, Groenen JT. 1992. Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. EMBO J. 11:287–290.
Plasterk RH, Izsvak Z, Ivics Z. 1999. Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet. 15:326–332.
Pollak E. 1987. On the theory of partially inbreeding finite populations. 1. Partial selfing. Genetics. 117:353–360.
Ponting CP. 2008. The functional repertoires of metazoan genomes. Nat Rev Genet. 9:689–698.
Prachumwat A, DeVincentis L, Palopoli MF. 2004. Intron size correlates positively with recombination rate in *Caenorhabditis elegans*. Genetics. 166:1585–1590.
Pujol N, Zugasti O, Wong D, Couillault C, Kurz CL, Schulenburg H, Ewbank JJ. 2008. Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. PLoS Pathog. 4: e1000105.
Qian W, Zhang J. 2008. Evolutionary dynamics of nematode operons: easy come, slow go. Genome Res. 18:412–421.
Rattray B, Rose AM. 1988. Increased intragenic recombination and non-disjunction in the Rec-1 strain of *Caenorhabditis elegans*. Genet Res. 51:89–93.
Reece-Hoyes J, Shingles J, Dupuy D, Grove C, Walhout A, Vidal M, Hope I. 2007. Insight into transcription factor gene duplication from *Caenorhabditis elegans* Promoterome-driven expression patterns. BMC Genomics. 8:27.
Reinke V, Cutter AD. 2009. Germline expression influences operon organization in the *Caenorhabditis elegans* genome. Genetics.
Reinke V, Gil IS, Ward S, Kazmer K. 2004. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. Development. 131:311–323.
Reinke V, Smith HE, Nance J, et al. (11 co-authors). 2000. A global profile of germline gene expression in *C. elegans*. Mol Cell. 6:605–616.
Rezsohazy R, van Luenen HG, Durbin RM, Plasterk RH. 1997. Tc7, a Tc1-hitch hiking transposon in *Caenorhabditis elegans*. Nucleic Acids Res. 25:4048–4054.
Rizzon C, Martin E, Marais G, Duret L, Segalat L, Biemont C. 2003. Patterns of selection against transposons, inferred from the distribution of Tc1, Tc3 and Tc5 insertions in the *mut-7* line of the nematode *Caenorhabditis elegans*. Genetics. 165: 1127–1135.
Robert V, Bessereau JL. 2007. Targeted engineering of the *Caenorhabditis elegans* genome following *Mos1*-triggered chromosomal breaks. EMBO J. 26:170–183.
Robert VJ, Sijen T, van Wolfswinkel J, Plasterk RH. 2005. Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. Genes Dev. 19:782–787.
Robertson HM. 1998. Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res. 8:449–463.
Robertson HM. 2000. The large srh family of chemoreceptor genes in Caenorhabditis nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. Genome Res. 10:192–203.
Robertson HM. 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.
Robertson HM, Thomas JH. 2006. The putative chemoreceptor families of *C. elegans in* Community TCeR. WormBook [On-line book]. [cited 30 Mar 2009]. Available from: http://www.wormbook.org.
Robinson M, Lilley R, Little S, Emtage JS, Yarranton G, Stephens P, Millican A, Eaton M, Humphreys G. 1984.

Codon usage can affect efficiency of translation of genes in *Escherichia coli*. Nucleic Acids Res. 12:6663–6671.

Rockman MV, Kruglyak L. Forthcoming. 2009. Recombinational landscape and population genomics of C. elegans. PLoS Genet. 5:e1000419.

Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV. 2003. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol. 13:1512–1517.

Rose AM, Baillie DL. 1980. Genetic organization of the region around UNC-15 (I), a gene affecting paramyosin in *Caenorhabditis elegans*. Genetics. 96:639–648.

Rose AM, Baillie DL, Candido EP, Beckenbach KA, Nelson D. 1982. The linkage mapping of cloned restriction fragment length differences in *Caenorabditis elegans*. Mol Gen Genet. 188:286–291.

Roy PJ, Stuart JM, Lund J, Kim SK. 2002. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. Nature. 418:975–979.

Roy SW, Fedorov A, Gilbert W. 2003. Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. Proc Natl Acad Sci USA. 100:7158–7162.

Roy SW, Gilbert W. 2005. The pattern of intron loss. Proc Natl Acad Sci USA. 102:713–718.

Roy SW, Penny D. 2006. Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses. Mol Biol Evol. 23:2259–2262.

Rubin GM, Yandell MD, Wortman JR, et al. (54 co-authors). 2000. Comparative genomics of the eukaryotes. Science. 287:2204–2215.

Rukov JL, Irimia M, Mork S, Lund VK, Vinther J, Arctander P. 2007. High qualitative and quantitative conservation of alternative splicing in *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Mol Biol Evol. 24:909–917.

Sanford C, Perry MD. 2001. Asymmetrically distributed oligonucleotide repeats in the *Caenorhabditis elegans* genome sequence that map to regions important for meiotic chromosome segregation. Nucleic Acids Res. 29: 2920–2926.

Satou Y, Mineta K, Ogasawara M, et al. (20 co-authors). 2008. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. Genome Biol. 9:R152.

Schulenburg H, Boehnisch C. 2008. Diversification and adaptive sequence evolution of Caenorhabditis lysozymes (Nematoda: Rhabditidae). BMC Evol Biol. 8:114.

Schulenburg H, Ewbank JJ. 2004. Diversity and specificity in the interaction between *Caenorhabditis elegans* and the pathogen *Serratia marcescens*. BMC Evol Biol. 4:49.

Schulenburg H, Muller S. 2004. Natural variation in the response of *Caenorhabditis elegans* towards *Bacillus thuringiensis*. Parasitology. 128:433–443.

Schwarz EM. 2005. Genomic classification of protein-coding gene families *in* Community TCeR. WormBook [On-line book]. [cited 30 Mar 2009]. Available from: http://www.wormbook.org.

Seidel HS, Rockman MV, Kruglyak L. 2008. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. Science. 319:589–594.

Semple C, Wolfe KH. 1999. Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. J Mol Evol. 48:555–564.

Seoighe C, Wolfe KH. 1999. Yeast genome evolution in the post-genome era. Curr Opin Microbiol. 2:548–554.

Seyfert AL, Cristescu ME, Frisse L, Schaack S, Thomas WK, Lynch M. 2008. The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. Genetics. 178:2113–2121.

Shabalina SA, Kondrashov AS. 1999. Pattern of selective constraint in *C. elegans* and *C. briggsae* genomes. Genet Res. 74:23–30.

Sharp PM, Bradnam KR. 1997. Codon usage in *C. elegans*. In: Riddle DL, Blumenthal T, Meyer BJ, Priess JR, editors. C. elegans *II*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 1053–1057.

Sharp PM, Li WH. 1987. The codon adaptation index: a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281–1295.

Sharp PM, Tuohy TM, Mosurski KR. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14:5125–5143.

Shields DC, Sharp PM, Higgins DG, Wright F. 1988. Silent sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol. 5: 704–716.

Sijen T, Plasterk RH. 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature. 426:310–314.

Simonis N, Rual J-F, Carvunis A-R, et al. (39 co-authors). 2009. Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. Nat Meth. 6:47–54.

Singh ND, Davis JC, Petrov DA. 2005. X-linked genes evolve higher codon bias in Drosophila and Caenorhabditis. Genetics. 171:145–155.

Sivasundar A. 2005. Genetic variation, population structure and outcrossing in *Caenorhabditis elegans* [dissertation]. Ecology and Evolution. [Piscataway (NJ)]: Rutgers University. 98 p.

Sivasundar A, Hey J. 2003. Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species. Genetics. 163.

Sivasundar A, Hey J. 2005. Sampling from natural populations using RNAi reveals high outcrossing and population structure in *Caenorhabditis elegans*. Curr Biol. 15:1598–1602.

Slatkin M. 1985. Genetic differentiation of transposable elements under mutation and unbiased gene conversion. Genetics. 110:145–158.

Sluder AE, Mathews SW, Hough D, Yin VP, Maina CV. 1999. The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. Genome Res. 9:103–120.

Snutch TP, Baillie DL. 1984. A high degree of DNA strain polymorphism associated with the major heat-shock gene in *Caenorhabditis elegans*. Mol Gen Genet. 195:329–335.

Sorensen MA, Kurland CG, Pedersen S. 1989. Codon usage determines translation rate in *Escherichia coli*. J Mol Biol. 207:365–377.

Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T. 1993. Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. Cell. 73:521–532.

Stein LD, Bao Z, Blasiar D, et al. (36 co-authors). 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1:166–192.

Stenico M, Lloyd AT, Sharp PM. 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. Nucleic Acids Res. 22: 2437–2446.

Stewart MR, Clark NL, Merrihew G, Galloway EM, Thomas JH. 2005. High genetic diversity in the chemoreceptor superfamily of *Caenorhabditis elegans*. Genetics. 169:1985–1996.

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

Surzycki SA, Belknap WR. 2000. Repetitive-DNA elements are similarly distributed on *Caenorhabditis elegans* autosomes. Proc Natl Acad Sci USA. 97:245–249.

Swan KA, Curtis DE, McKusick KB, Voinov AV, Mapa FA, Cancilla MR. 2002. High-throughput gene mapping in *Caenorhabditis elegans*. Genome Res. 12:1100–1105.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet. 3:137–144.

Tam SM, Causse M, Garchery C, Burck H, Mhiri C, Grandbastien MA. 2007. The distribution of copia-type retrotransposons and the evolutionary history of tomato and related wild species. J Evol Biol. 20:1056–1072.

Teotonio H, Manoel D, Phillips PC. 2006. Genetic variation for outcrossing among *Caenorhabditis elegans* isolates. Evolution. 60:1300–1305.

Thomas JH. 2006a. Analysis of homologous gene clusters in *Caenorhabditis elegans* reveals striking regional cluster domains. Genetics. 172:127–143.

Thomas JH. 2006b. Concerted evolution of two novel protein families in Caenorhabditis species. Genetics. 172: 2269–2281.

Thomas JH. 2008. Genome evolution in Caenorhabditis. Brief Funct Genomic Proteomic. 7:211–216.

Thomas JH, Kelley JL, Robertson HM, Ly K, Swanson WJ. 2005. Adaptive evolution in the SRZ chemoreceptor families of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Proc Natl Acad Sci USA. 102:4476–4481.

Thomas WK, Wilson AC. 1991. Mode and tempo of molecular evolution in the nematode Caenorhabditis: cytochrome oxidase-II and calmodulin sequences. Genetics. 128:269–279.

Tijsterman M, Okihara KL, Thijssen K, Plasterk RH. 2002. PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of *C. elegans*. Curr Biol. 12:1535–1540.

Tischler J, Lehner B, Chen N, Fraser AG. 2006. Combinatorial RNA interference in *Caenorhabditis elegans* reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. Genome Biol. 7:R69.

Tobler H, Muller F. 2001. Chromatin diminution. Encyclopedia of life sciences. Chichester, UK: John Wiley & Sons, Ltd.

Toth G, Gaspari Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10:967–981.

Tu Z, Shao H. 2002. Intra- and inter-specific diversity of Tc3-like transposons in nematodes and insects and implications for their evolution and transposition. Gene. 282:133–142.

Tuzun E, Bailey JA, Eichler EE. 2004. Recent segmental duplications in the working draft assembly of the Brown Norway Rat. Genome Res. 14:493–506.

Uitterlinden AG, Slagboom PE, Johnson TE, Vijg J. 1989. The *Caenorhabditis elegans* genome contains monomorphic minisatellites and simple sequences. Nucleic Acids Res. 17:9527–9530.

van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet. 19:530–536.

Varenne S, Buc J, Lloubes R, Lazdunski C. 1984. Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J Mol Biol. 180: 549–576.

Vassilieva LL, Lynch M. 1999. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. Genetics. 151:119–129.

Viney ME, Gardner MP, Jackson JA. 2003. Variation in *Caenorhabditis elegans* dauer larva formation. Dev Growth Differ. 45:389–396.

Wagner A. 2002. Selection and gene duplication: a view from the genome. Genome Biol. 3:reviews1012.

Wang PJ, McCarrey JR, Yang F, Page DC. 2001. An abundance of X-linked genes expressed in spermatogonia. Nat Genet. 27:422–426.

Wang X, Chamberlin HM. 2004. Evolutionary innovation of the excretory system in *Caenorhabditis elegans*. Nat Genet. 36:231–232.

Ward S, Burke DJ, Sulston JE, Coulson AR, Albertson DG, Ammons D, Klass M, Hogan E. 1988. Genomic organization of major sperm protein genes and pseudogenes in the nematode *Caenorhabditis elegans*. J Mol Biol. 199:1–13.

Waterston R, Sulston J. 1995. The genome of *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 92:10836–10840.

Webb CT, Shabalina SA, Ogurtsov AY, Kondrashov AS. 2002. Analysis of similarity within 142 pairs of orthologous intergenic regions of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Nucleic Acids Res. 30:1233–1239.

Wicks SR, Yeh RT, Gish WR, Waterston RH, Plasterk RHA. 2001. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. Nat Genet. 28:160–164.

Wicky C, Rose AM. 1996. The role of chromosome ends during meiosis in *Caenorhabditis elegans*. Bioessays. 18:447–452.

Wicky C, Villeneuve AM, Lauper N, Codourey L, Tobler H, Muller F. 1996. Telomeric repeats (TTAGGC)n are sufficient for chromosome capping function in *Caenorhabditis elegans*. Proc Natl Acad Sci USA. 93:8983–8988.

Winston WM, Sutherlin M, Wright AJ, Feinberg EH, Hunter CP. 2007. *Caenorhabditis elegans* SID-2 is required for environmental RNA interference. Proc Natl Acad Sci USA. 104:10565–10570.

Witherspoon DJ, Robertson HM. 2003. Neutral evolution of ten types of mariner transposons in the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. J Mol Evol. 56:751–769.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. 387:708–713.

Woollard A. 2005. Gene duplications and genetic redundancy in C. elegans *in* Community TceR. WormBook [On-line book]. [cited 30 Mar 2009]. Available from: http://www.wormbook.org.

Wright F. 1990. The effective number of codons used in a gene. Gene. 87:23–29.

Wright SI, Agrawal N, Bureau TE. 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. Genome Res. 13: 1897–1903.

Wright SI, Lauga B, Charlesworth D. 2002. Rates and patterns of molecular evolution in inbred and outbred Arabidopsis. Mol Biol Evol. 19:1407–1420.

Wright SI, Le QH, Schoen DJ, Bureau TE. 2001. Population dynamics of an Ac-like transposable element in self- and cross-pollinating Arabidopsis. Genetics. 158:1279–1288.

Wright SI, Nano N, Foxe JP, Dar VN. 2008. Effective population size and tests of neutrality at cytoplasmic genes in *Arabidopsis*. Genet Res. 90:119–128.

Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. Int J Plant Sci. 169:105–118.

Wright SI, Schoen DJ. 1999. Transposon dynamics and the breeding system. Genetica. 107:139–148.

Wright SI, Yau CBK, Looseley M, Meyers BC. 2004. Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. Mol Biol Evol. 21: 1719–1726.

Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 15:496–503.

Youngman S, van Luenen HGAM, Plasterk RHA. 1996. Rte-1, a retrotransposon-like element in *Caenorhabditis elegans*. FEBS Lett. 380:1–7.

Zetka M, Rose A. 1995. The genetics of meiosis in *Caenorhabditis elegans*. Trends Genet. 11:27–31.

Zetka MC, Rose AM. 1990. Sex-related differences in crossing over in *Caenorhabditis elegans*. Genetics. 126:355–363.

Zhao Z, Boyle TJ, Bao Z, Murray JI, Mericle B, Waterston RH. 2008. Comparative analysis of embryonic cell lineage between *Caenorhabditis briggsae* and *Caenorhabditis elegans*. Dev Biol. 314:93–99.

Zhao Z, Thomas JH, Chen N, Sheps JA, Baillie DL. 2007. Comparative genomics and adaptive selection of the ATP-binding-cassette gene family in Caenorhabditis species. Genetics. 175:1407–1418.

Zou L, Sriswasdi S, Ross B, Missiuro PV, Liu J, Ge H. 2008. Systematic analysis of pleiotropy in *C. elegans* early embryogenesis. PLoS Comput Biol. 4:e1000003.

Jennifer Wernegreen, Associate Editor

Accepted March 6, 2009