Copyright © 2009 by the Genetics Society of America
DOI: 10.1534/genetics.107.082651

High Nucleotide Divergence in Developmental Regulatory Genes Contrasts With the Structural Elements of Olfactory Pathways in Caenorhabditis

Richard Jovelin, Joseph P. Dunham,[1] Frances S. Sung[1] and Patrick C. Phillips[2]

Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97403-5289

Manuscript received September 26, 2008
Accepted for publication November 5, 2008

ABSTRACT

Almost all organismal function is controlled by pathways composed of interacting genetic components. The relationship between pathway structure and the evolution of individual pathway components is not completely understood. For the nematode *Caenorhabditis elegans*, chemosensory pathways regulate critical aspects of an individual's life history and development. To help understand how olfaction evolves in Caenorhabditis and to examine patterns of gene evolution within transduction pathways in general, we analyzed nucleotide variation within and between species across two well-characterized olfactory pathways, including regulatory genes controlling the fate of the cells in which the pathways are expressed. In agreement with previous studies, we found much higher levels of polymorphism within *C. remanei* than within the related species *C. elegans* and *C. briggsae*. There are significant differences in the rates of nucleotide evolution for genes across the two pathways but no particular association between evolutionary rate and gene position, suggesting that the evolution of functional pathways must be considered within the context of broader gene network structure. However, developmental regulatory genes show both higher levels of divergence and polymorphism than the structural genes of the pathway. These results show that, contrary to the emerging paradigm in the evolution of development, important structural changes can accumulate in transcription factors.

THE integration of evolutionary and developmental genetics into the discipline of evolutionary developmental biology has provided a powerful framework for understanding the evolution of form and pattern. The major paradigm emerging from evo-devo is the belief that most evolutionary change is generated by changes in gene regulation as opposed to protein structure (King and Wilson 1975; Jacob 1977; Duboule and Wilkins 1998; Carroll 2005). This regulatory hypothesis focuses most strongly on changes in *cis*-regulatory regions of genes, rather than on the evolution of the regulatory genes (*e.g.*, transcription factors) themselves. This is because it is presumed that changes in *cis*-elements will be localized to the gene of interest, whereas changes in *trans*-acting factors will tend to have broad pleiotropic effects. However, there are numerous examples of important evolutionary transitions mapping to protein coding differences within and between species (Hoekstra and Coyne 2007; Stern and Orgogozo 2008). It is therefore possible for regulatory changes to be quite important, but to still be essentially structural in nature (Lynch and Wagner 2008; Wagner and Lynch 2008). For example, protein evolution of regulatory genes has been associated with species radiations (Barrier *et al.* 2001; Lawton-Rauh *et al.* 2003) and other major morphological changes (Galant and Carroll 2002; Ronshaugen *et al.* 2002). In addition, positive selection shaping the pattern of substitution for various transcription factor families in plants and animals (Sutton and Wilkinson 1997; Fares *et al.* 2003; Jia *et al.* 2003, 2004; Martinez-Castilla and Alvarez-Buylla 2003; Balakirev and Ayala 2004; Moore *et al.* 2005) may have direct phenotypic consequences.

Although evo-devo, because of its historic ties with embryology and paleontology, has been confined to morphological evolution, the evo-devo approach is being applied to understand a broader set of topics such as the evolution of sexual development (Haag and Ackerman 2005; Haag and Doty 2005; Nayak *et al.* 2005) and social behavior (Toth and Robinson 2007). Thus far, most examinations of the evolution of regulatory changes have focused on single genes. Yet genes exist within broad functional networks, the structure of which could potentially have profound effects on the rates of evolution of the individual components within the network (Hahn and Kern 2005). For example, within metabolic pathways, we might expect upstream

Sequence data from this article have been deposited in GenBank under accession nos. FJ455605–FJ455760.

[1]*Present address:* Molecular and Computational Biology, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089-2910.

[2]*Corresponding author:* Center for Ecology and Evolutionary Biology, 5289 University of Oregon, Eugene, OR 97403-5289.
E-mail: pphil@uoregon.edu

components to evolve more slowly than downstream elements because of kinetic constraints (Rausher *et al.* 1999). Similarly, within signal transduction pathways, we might expect different elements of the pathway to play different roles in regulating the efficacy of the transduction response. Note that here we are referring to regulation at a level higher than gene transcription, since in most cases the genes utilized in a transduction pathway are already in place as the signal is being processed. What patterns of evolution might be generated across such pathways? One possibility is a pattern similar to metabolic pathways, because the cumulative effect of change in upstream components would generate strong constraints on their evolution. Another possibility is that evolutionary change will be concentrated at the most upstream element, the receptor, because changes here can make the entire system more or less sensitive to a particular environmental response, whereas the downstream structural components of the pathway need to be maintained intact for the whole pathway to maintain its function (*e.g.*, Sackton *et al.* 2007). Here, a transduction pathway is qualitatively different from a metabolic pathway because information, rather than a specific level of metabolic product, is the currency being transmitted from element to element. One approach to addressing these hypotheses is to analyze the pattern of molecular genetic variation across a well-characterized pathway.

In this study, building upon recent results from developmental and classical genetics, we investigate the evolution of the signal transduction pathways that underlie olfaction in the genus Caenorhabditis. These nematodes are particularly well suited to investigate the evolution of molecular function in relation to olfaction because the odor code is relatively well understood, including behavioral responses, neuron maps, and candidate genes from receptors to downstream effectors and regulators of signaling cascades (Bargmann and Kaplan 1998; Bergamasco and Bazzicalupo 2006). Chemosensation is used by all animals for finding food and testing its quality, finding a mating partner, avoiding predators or toxins, and participating in social behaviors. For example, divergent olfactory abilities resulting in alteration of the mode of mating partner recognition or in adaptation to a new food source may have a direct effect on speciation (Linn *et al.* 2003; Stensmyr *et al.* 2003; Ortiz-Barrientos *et al.* 2004; Dekker *et al.* 2006). In *Caenorhabditis elegans*, chemosensation triggers social feeding (de Bono *et al.* 2002; de Bono 2003), egg-laying (Daniels *et al.* 2000) behaviors, and the detection of a pheromone regulates larval development (Riddle and Albert 1997; Jeong *et al.* 2005; Butcher *et al.* 2007).

The two pairs of amphid neurons AWA and AWC (Figure 1) are required to sense all attractive odorants (Bargmann *et al.* 1993). However, the response triggered by a specific odorant is not determined by the receptor itself but by the genetic environment in which the receptor is expressed (Troemel *et al.* 1997). The nuclear hormone receptor *odr-7* specifies AWA identity by inducing AWA-specific genes and repressing AWC-specific genes (Sengupta *et al.* 1994; Sagasti *et al.* 1999), whereas the *Otx* homeobox *ceh-36* is required for AWC identity (Lanjuin *et al.* 2003). Each pathway corresponds to one of the two main types of sensory signaling pathways found in invertebrates and vertebrates (Ache and Young 2005): one activating a transient receptor potential (TRP) channel and the other activating a cyclic nucleotide-gated channel. In AWA, chemoreceptors (Troemel *et al.* 1995), such as ODR-10 (Sengupta *et al.* 1996), stimulate a TRPV (vanilloid-type) channel made of the subunits OSM-9 and OCR-2 (Colbert *et al.* 1997; Tobin *et al.* 2002) through the guanine nucleotide binding protein (G protein) ODR-3 (Roayaie *et al.* 1998; Lans *et al.* 2004). Signal transduction is also mediated by ODR-3 in AWC but ODR-3 activates the cation channel TAX-2/TAX-4 (Coburn and Bargmann 1996; Komatsu *et al.* 1996) through the production of cGMP by the guanylyl cyclases ODR-1 and DAF-11 (L'Etoile and Bargmann 2000). Neuron-specific stimulatory and inhibitory G protein signaling adds to the complexity of the pathways and allows further discrimination of compounds (Lans *et al.* 2004).

Here, we analyze nucleotide rate variation within and between species in these two olfactory pathways (Figure 1) to examine the relationship between evolutionary rate and gene function, as well as their position within the pathway. We then compare evolutionary rates between structural genes in the pathways and the developmental regulatory genes controlling the fate of the chemosensory neurons in which the pathways are expressed. We find that transcription factors specifying chemosensory neuron subtype identity exhibit higher levels of nucleotide divergence and polymorphism than the structural genes in the signaling pathways. Our results suggest that behavioral differences between species in response to olfactory stimuli are likely to involve differences in gene regulation due to the divergence of regulatory genes. These findings stress the importance of protein evolution of regulatory genes for a complex and functionally important trait at the crossroad between physiology, ecology, and evolution.

MATERIALS AND METHODS

Caenorhabditis strains: All strains were maintained following standard protocols (Brenner 1974). The strains used in this study, along with their classification and geographical origin are listed in supplemental Table 1. *C. remanei* strains from Ohio were inbred for at least six generations of brother–sister mating to minimize within-strain nucleotide variation. Oregon strains were obtained from soil samples collected at various locations and mixed with M9 buffer prior to filtering. *C. elegans* isolates were identified using morphological criteria,

their ability to cross with the lab strain N2, and the molecular markers reported here. Each new *C. elegans* strain was derived from a single individual.

***C. briggsae* and *C. remanei* ortholog identification:** We searched the *C. briggsae* (Stein *et al.* 2003) and *C. remanei* (Genome Sequencing Center, Washington University, St. Louis) genome assemblies for orthologs of the *C. elegans* genes using the TBLASTN program (Altschul *et al.* 1990). Orthology was confirmed on the basis of amino acid sequence identity and reciprocal BLAST best hits. Intron/exon boundaries were identified with respect to the *C. elegans* sequence and with reference to the open reading frame.

Amplification and sequencing: DNA extractions were performed using the CTAB protocol (see Winnepenninckx *et al.* 1993; Jovelin *et al.* 2003). PCR amplifications were processed as described in Jovelin *et al.* (2003), gel purified (QIAGEN), and sequenced using automated sequencers (University of Chicago Cancer Research Center and University of Oregon sequencing facilities). Primers used for amplification were also used for sequencing. Additional internal primers were also used such that the resulting sequences strongly overlapped. All sequences were confirmed on both strands. All sequence changes were rechecked visually against sequencing chromatograms.

Sequence analyses: Protein sequences were aligned by eye using BioEdit (Hall 1999) and subsequently used to generate codon-based DNA sequence alignments. Maximum likelihood (ML) estimates of the rates of nonsynonymous (d_N) and synonymous (d_S) substitutions across the three-species tree (Kiontke *et al.* 2004) were computed using the program CODEML of the PAML package (Yang 1997), with a codon model assuming equal rate of substitutions among sites but accounting for transition/transversion bias and by removing gap positions. Codon frequencies are the product of the observed nucleotide frequencies at each codon position. In addition we computed the rate of synonymous changes corrected for selection at silent sites (d_S') (Hirsh *et al.* 2005) using the slope of the regression between d_S and the codon adaptation index (CAI) (Sharp and Li 1987) for all genes analyzed here. CAI was computed using CAI Calculator (Wu *et al.* 2005) using the reference set of *C. elegans* highly expressed genes of Carbone *et al.* (2003). We also estimated CAI in *C. remanei* using the *C. elegans* gene set. Gene tree of the *ceh-36* sequences sampled from *C. remanei* was inferred with maximum parsimony using PAUP* 4.0b10 (Swofford 1998). Comparisons of evolutionary rates between transcription factors and structural genes were performed using JMP 4.0.4 (SAS Institute) and with a likelihood ratio test (LRT) using CODEML. In the first model there was a single parameter ω (d_N/d_S) across all concatenated sequences, and in the second model the concatenated sequences were partitioned between transcription factors and structural genes with each partitioned set having its own ω-parameter.

Measures of nucleotide diversity (Watterson 1975; Nei 1987), population genetic analyses, and tests of selection were performed using DnaSP 4.1 (Rozas *et al.* 2003). Insertion/deletion and complex codons were not included in estimates of diversity. Complex codons are those ambiguous codons that differ at two or three codon positions and for which synonymous or nonsynonymous changes cannot be counted unambiguously. There were only 2 complex codons in 3383 codons analyzed within *C. remanei*. Two regions of 257 bp and 431 bp located, respectively, in intron 1 of *C. remanei ceh-36a* and intron 5 of *C. remanei ceh-36b* could not be aligned without ambiguity and were removed prior to analysis.

Departures from neutrality were investigated using Tajima's *D* (Tajima 1989), McDonald–Kreitman test (MK) (McDonald and Kreitman 1991), and a likelihood ratio test (Hasegawa *et al.* 1998). Only the *C. remanei* alleles sampled from the population in Ohio ($n = 12$) were used for all three tests of selection. Tajima's *D* was performed using polymorphism at noncoding sites, although using all sites did not change the results. Significance for Tajima's *D* was assessed by coalescent simulation with 50,000 replicates conditioning on the number of segregating sites and assuming no recombination. MK tests were performed using polymorphism data from *C. remanei* and by using the most closely related species *C. briggsae* (Kiontke *et al.* 2004) for interspecific divergence. Divergence using *C. elegans* did not change the results of the MK tests (not shown). Significance of the MK tests was assessed using a *G*-test of independence with Williams' correction. The likelihood ratio test compared the d_N/d_S ratio along the within- and between-species branches of the phylogenetic tree comprising the *C. remanei* alleles and the *C. briggsae* ortholog (Hasegawa *et al.* 1998). A star phylogeny was assumed among the *C. remanei* alleles.

RESULTS

Identification of orthologs and gene duplicates: We identified the orthologs of the *C. elegans* genes acting in two distinct chemosensory pathways (Figure 1), in the two closely related species *C. briggsae* and *C. remanei*. Genes were identified as unique singletons in almost all cases. However, in *C. remanei* we found two sequences present on distinct contigs with high similarity to the transcription factor *ceh-36*. Reciprocal BLAST searches against the *C. elegans* and *C. briggsae* genomes using each of these two sequences identified a single sequence corresponding to *ceh-36*. Further searches in the three Caenorhabditis genomes identified sequences aligning only to the *ceh-36* homeodomain. Nonetheless, given the degree of divergence between these two sequences ($d_N/d_S = 0.145$), it seems highly unlikely that they represent distinct alleles, but are rather coorthologs of the *C. elegans ceh-36* gene. Supporting this idea, intronic and intergenic sequences are also highly divergent between these loci. Subsequent analysis of other natural isolates (see below) also always recovered unique sequences that cluster with one another in a phylogenetically distinct pattern. Consequently, from this point onward, we will use the terminology *ceh-36a* and *ceh-36b* to designate these two sequences. Because *C. elegans* is the sister taxon of a clade formed by *C. briggsae* and *C. remanei* (Kiontke *et al.* 2004), we infer that the duplication of *ceh-36* occurred in the lineage leading to *C. remanei*.

No relationship between pathway position and evolutionary rate: Maximum likelihood estimates of the ratio of the rate of amino acid replacements to the rate of synonymous changes (d_N/d_S) show that divergence across species is not uniformly distributed along the AWA and AWC chemosensory pathways, but varies within and among gene classes (Figure 2). Within G proteins there is a 18-fold difference in the level of divergence between the most conserved (*odr-3*) and the most divergent (*gpa-5*) loci. The very high degree of conservation observed at the *odr-3* locus, both between

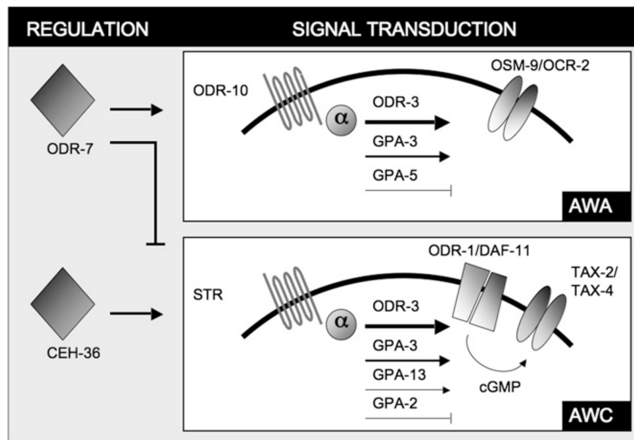

Figure 1.—Olfactory pathways mediated by the neurons AWA and AWC within *Caenorhabditis elegans*. Each pathway corresponds to one of the two main types of olfactory pathways found in invertebrates and vertebrates. AWA activates a TRPV channel and AWC activates a cyclic nucleotide-gated channel. The nuclear receptor *odr-7* controls the identity of the AWA neurons, by inducing AWA-specific genes and repressing AWC-specific genes. The homeobox *ceh-36* controls the identity of the AWC neurons.

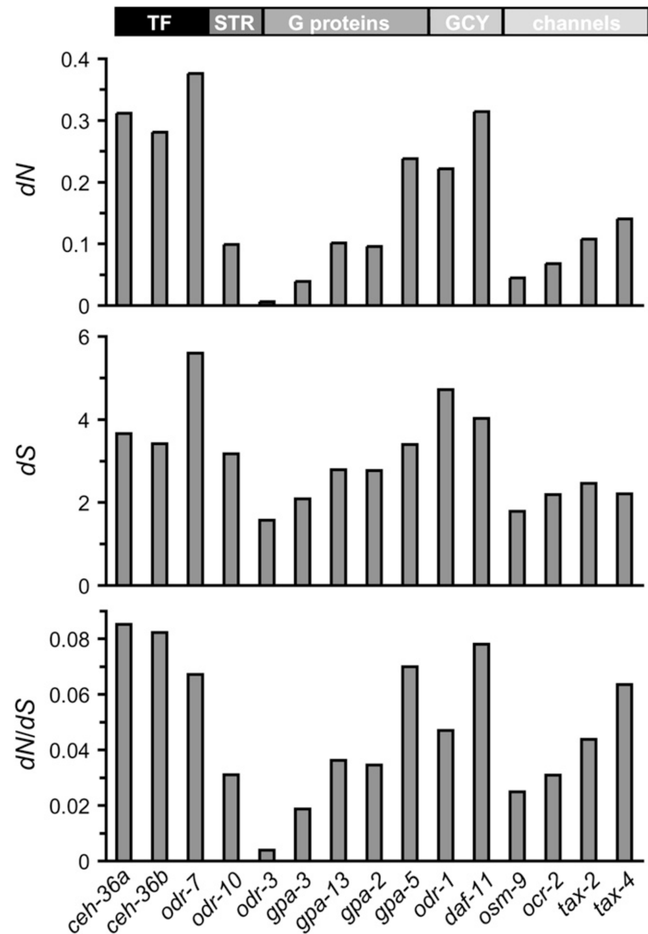

Figure 2.—Maximum likelihood estimates of the rates of amino acid replacements (d_N) and synonymous changes (d_S), and the corresponding ratio (d_N/d_S), across the phylogenetic tree including *C. elegans*, *C. briggsae*, and *C. remanei*. The genes are listed in order of their relative position within the pathways. The degree of divergence varies along the AWA and AWC pathways. With the exception of the G protein *gpa-5* and the guanylyl cyclase *daf-11*, the transcription factors *odr-7* and *ceh-36* exhibit more interspecific divergence than the structural component of the two pathways. TF, transcription factor; STR, seven-transmembrane receptor; G protein, guanine nucleotide binding protein; GCY, guanylyl cyclase.

and within species, indicates that this gene is under strong purifying selection to maintain its function in chemosensation and/or neuronal cilia development (Roayaie *et al.* 1998; Jovelin *et al.* 2003; Lans *et al.* 2004). There is no clear relationship between a gene's position in the pathway and its rate of evolution. In particular, the most upstream gene, the *odr-10* receptor, does not evolve particularly quickly. If anything, there is a general tendency for downstream components, such the guanylyl cyclases and channel proteins to be more divergent than the upstream receptor and G protein components. Because of the evolutionary distance between these groups, and because of selection at synonymous sites for codon usage (Cutter and Charlesworth 2006; Cutter *et al.* 2006c) the d_N/d_S ratio may tend to be overestimated. Nevertheless, we observe a very similar pattern of divergence across the two pathways for the d_N/d_S' ratio after d_S is corrected for selection at silent sites (d_S') (Hirsh *et al.* 2005) (not shown). Moreover, the pattern of divergence exhibited by the d_N values themselves precisely matches the pattern from the d_N/d_S and d_N/d_S' ratios (Figure 2).

Developmental regulatory genes exhibit high levels of interspecific divergence: Somewhat unexpectedly, the regulatory components of the signaling pathways, the transcription factors controlling the identity of the chemosensory neurons AWA and AWC, are up to 22-fold more divergent than the structural components (Figure 2). Only two genes, the G protein *gpa-5* and the guanylyl cylcase *daf-11* show comparable levels of interspecific divergence. However the transcription factors are clearly more divergent, on average, than the structural genes (Figure 3) (d_N: Wilcoxon two-sample $P = 0.0253$; d_N/d_S: Wilcoxon two-sample $P = 0.0253$; d_N/d_S': Wilcoxon two-sample $P = 0.0364$). This higher level of divergence among the developmental regulatory genes is further supported by a likelihood ratio test showing heterogeneity in d_N/d_S (ω) between structural genes and transcription factors ($\omega_{TF} = 0.0813$, $\omega_{structural} = 0.0558$, $2\Delta l = 486.95$ $P < 0.001$). Comparisons between the more closely related species *C. briggsae* and *C. remanei* gave similar results (d_N: Wilcoxon two-sample $P = 0.0172$, d_N/d_S: Wilcoxon two-sample $P = 0.0253$, d_N/d_S': Wilcoxon two-sample $P = 0.0172$; $\omega_{TF} = 0.1096$, $\omega_{structural} = 0.0598$, $2\Delta l = 379.08$, $P < 0.001$). Nevertheless, high levels of divergence can be either the result of relaxed selection and the accumulation of neutral or slightly deleterious substitutions or the result of positive

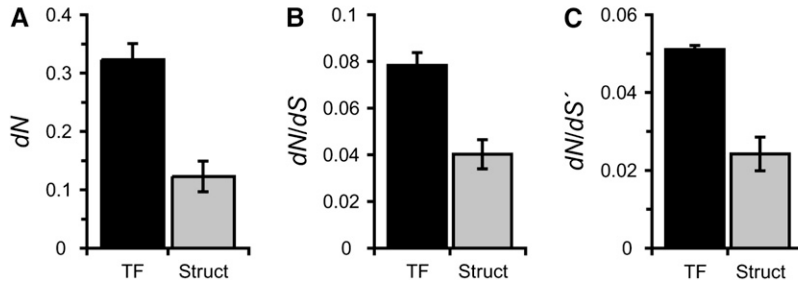

FIGURE 3.—The transcription factors controlling the differentiation of the AWA and AWC neurons evolve on average faster than structural olfactory genes. Divergence is measured across the phylogenetic tree including *C. elegans*, *C. briggsae*, and *C. remanei.*

selection leading to the fixation of beneficial mutations. Therefore, these patterns of interspecific divergence need to be interpreted in the context of variation within populations. We thus sampled polymorphism information for the full or nearly full length of seven loci spanning all the functional gene classes of the AWA and AWC pathways within a population of *C. remanei* (supplemental Table 1).

C. remanei **exhibits higher levels of polymorphisms than its congeneric species *C. elegans* and *C. briggsae*:** In addition to the *C. remanei* samples, we collected polymorphism data for five loci from the worldwide distributions of *C. elegans* and *C. briggsae* (Table 1). Length variants were found in all three species, with insertion/deletion (indel) size ranging from single nucleotide to 166 bp in intron 1 of *C. remanei tax-2*. Indel polymorphism is localized only in introns, however. In *C. remanei*, among 11,768 bp of sequence across the four loci for which we have polymorphism information for all three species, we identified 683 polymorphic sites. In contrast, there are only 14 in *C. elegans* and 22 in *C. briggsae* in 11,368 bp and 12,323 bp of sequence, respectively. The difference is even more striking when the levels of total nucleotide diversity (π_t) are compared at each locus individually (Table 1). Overall, we found between 18- and 190-fold more diversity in *C. remanei* than in *C. elegans*. The difference between *C. briggsae* and *C. remanei* is not as high, between 6- and 31-fold, but is nevertheless very substantial.

Analysis of the frequency distribution of polymorphism among species further illustrates the low levels of genetic variation in *C. elegans* and *C. briggsae*. For instance, among this subset of four genes, 93% of *C. elegans* polymorphism is segregating as singletons while the frequency of such polymorphism is only 25% in *C. remanei* (Fisher's exact test, $P = 0.0013$). Although for all three species the vast majority of polymorphism is found at silent sites (synonymous and noncoding), we identified nonsynonymous polymorphisms in all loci in *C. remanei* (Table 1). Again, this contrasts with the situation in *C. elegans* and *C. brigsgae*. Amino acid replacement polymorphism in *C. elegans* was found in only two genes, the transcription factor *odr-7* ($P_a = 3$) and the TRPV channel *osm-9* ($P_a = 2$). In *C. briggsae*, in addition to changes in these two loci, one amino acid replacement polymorphism was found in the coding sequence of the chemoreceptor *odr-10*, but the overall number of nonsynonymous polymorphic changes is equally as low as in *C. elegans*. The pattern of polymorphisms seen in *C. elegans* and *C. briggsae* is consistent with a scenario of a recent colonization from a limited number of populations (JOVELIN *et al.* 2003; CUTTER *et al.* 2006b). Overall, the levels of diversity in the gonochoristic and outcrossing species *C. remanei* are comparable to those previously reported (GRAUSTEIN *et al.* 2002; JOVELIN *et al.* 2003; HAAG and ACKERMAN 2005; CUTTER *et al.* 2006a) and greater than in the selfing hermaphrodite species *C. elegans* and *C. briggsae* (GRAUSTEIN *et al.* 2002; JOVELIN *et al.* 2003; SIVASUNDAR and HEY 2003; BARRIERE and FELIX 2005; CUTTER 2006; CUTTER *et al.* 2006b).

High levels of polymorphism in developmental regulatory genes: Because of the low levels of diversity observed in *C. elegans* and *C. briggsae* and because of the peculiar history of available isolates (PHILLIPS 2006), we used diversity within *C. remanei* to investigate functional constraints acting on the chemosensory genes. Within this species, almost all polymorphisms segregate as two variants. Of 1404 polymorphic sites in 21,723 bp of sequence, only 56 segregate as three variants and one segregates as four variants. These multiple variant polymorphisms principally affect the transcription factors *ceh-36a* and *ceh-36b* and the cyclic nucleotide-gated channel *tax-2*. Focusing only on coding regions, however, of 348 polymorphic sites in 10,203 bp, only 5 sites segregate as three variants. Analysis of polymorphism frequencies shows that the frequency distribution of nonsynonymous polymorphism is strongly skewed toward rare variants (Figure 4). With a cut-off value of 7% (*i.e.*, singleton polymorphism), the difference in the ratio of nonsynonymous to synonymous polymorphism (P_a/P_s) between rare and common polymorphisms is highly significant (Fisher's exact test, $P < 0.0001$), suggesting that a fraction of nonsynonymous mutations is slightly deleterious (FAY *et al.* 2001). Moreover, the difference between rare (singleton) and common P_a/P_s is significant within transcription factors (Fisher's exact test, $P = 0.0086$) and nontranscription factor loci (Fisher's exact test, $P = 0.0015$), suggesting that slightly deleterious mutations segregate in the two gene classes.

Nucleotide diversity (π_t) in *C. remanei* varies greatly among genes, from 6.6×10^{-3} for the guanylyl cyclase

TABLE 1

Nucleotide diversity in the chemosensory genes of the AWA and AWC pathways in Caenorhabditis

Locus (LG)	Sp	N	L (bp)	N_c	N_{syn}	P_t	P_a	P_s	P_{nc}	θ_w (10^{-3})	π_t (10^{-3})	π_s (10^{-3})	π_c (10^{-3})	π_{aa} (10^{-3})	π_{syn} (10^{-3})
odr-7 (X)	*Ce*	13	2571–2572	1371	307.167	9	3	2	4	1.13	0.54	0.61	0.56	0.43	1
	Cb	5	1918	1179	265	2	1	1	0	0.5	0.63	0.6	1.02	0.66	2.26
	Cr	14	1618–1620	1380	301.679	109	18	61	30	21.2	19.71	51.96	15.23	3.6	56.82
ceh-36a (X)	*Cr*	14	2568–2661	792	193.428	424	16	53	350	60.47	49.64	65.45	24.59	7.19	78.43
ceh-36b (X)	*Cr*	14	2641–2720	795	194.321	214	4	24	186	30.24	24.51	32.93	8.57	1.72	29.75
odr-10 (X)	*Ce*	13	1534	1017	242.833	1	0	0	1	0.21	0.35	0.71	0	0	0
	Cb	5	1475–1476	1026	242.466	5	1	1	3	1.63	1.76	2.89	1.17	0.77	2.47
	Cr	14	1342–1347	1026	240.726	78	11	33	34	18.28	11.14	23.11	8.07	2.66	25.7
odr-3[a] (V)	*Ce*	13	2652–2653	1068	224	1	0	0	1	0.12	0.06	0.09	0	0	0
	Cb	5	2219	1068	230.166	6	0	0	6	1.3	1.44	2.32	0	0	0
	Cr	14	2127–2137	1068	215.274	79	2	13	64	12.02	11.58	18.65	4.73	0.36	20.93
odr-1 (X)	*Cr*	14	4570–4574[b]	2697[b]	615.845	83	5	20	58	5.72	6.6	11.63	2.94	0.59	10.87
osm-9 (IV)	*Ce*	13	6281–6512	2811	612.346	41	2	4	35	2.17	2.56	3.63	0.74	0.33	2.06
	Cb	5	10423–10441	2820	619.834	17	1	3	13	0.8	0.92	1.1	0.83	0.2	2.9
tax-2 (I)	*Ce*	13	4609	2400	536.501	3	0	0	3	0.21	0.15	0.25	0	0	0
	Cb	5	6709v6710	2433	553.334	9	0	0	9	0.64	0.75	1.04	0	0	0
	Cr	14	6478–6664	2445	546.37	417	8	75	334	20.42	21.1	29.57	11.86	0.94	49.8

LG, linkage group in *C. elegans*; Sp, species; *Ce*, *C. elegans*; *Cb*, *C. briggsae*; *Cr*, *C. remanei*; *N*, number of strains; *L*, total length; N_c, number of coding sites; N_{syn}, average number of synonymous sites in the alignment; P_t, total number of polymorphic sites; P_a, number of nonsynonymous polymorphic sites; P_s, number of synonymous polymorphic sites; P_{nc}, number of noncoding polymorphic sites; θ_w, Watterson estimator (Watterson 1975), per site, conditioning on the total number of segregating sites; π_t, total nucleotide diversity (Nei 1987); π_{si}, π_c, π_{aa}, π_{syn}, respectively, nucleotide diversity at silent sites, coding sites, nonsynonymous sites, and synonymous sites.

[a] Updated from Jovelin *et al.* (2003).

[b] Incomplete sequences.

odr-1 to 49.64×10^{-3} for the transcription factor *ceh-36a* (Table 1). We found no significant correlation between diversity at synonymous sites and codon usage bias as measured by the codon adaptation index (Sharp and Li 1987) (π_{syn} × CAI: Spearman's $\rho = 0.414$, $P = 0.35$). Remarkably the developmental regulatory genes show a higher level of nonsynonymous polymorphic changes (π_{aa}) than the structural genes (Table 1). Among the structural genes, however, the chemoreceptor *odr-10* is an outlier and exhibits roughly 10-fold higher levels of nonsynonymous polymorphism than other structural genes (Table 1). Because the X chromosome has a smaller effective population size than autosomes, we may expect X-linked loci to harbor less nucleotide diversity than genes on the autosomes. However, assuming that chromosomal location is conserved between *C. elegans* and *C. remanei* (Hillier *et al.* 2007), the *C. remanei* genes *odr-7*, *ceh-36a*, *ceh-36b*, and *odr-10* are located on the X chromosome. The higher level of within-species variation at these loci is therefore not a consequence of linkage.

Forces shaping the pattern of divergence for developmental regulatory genes: In general, then, regulatory genes are both more divergent and more polymorphic than the structural components of the olfactory pathways. Inferring the evolutionary forces responsible for this pattern would normally rely on contrasting the level of polymorphism with the level of divergence via the MK test (McDonald and Kreitman 1991). The MK test assumes that most polymorphism is neutral and compares the ratio of fixed nonsynonymous to synonymous divergence (D_a/D_s) to the ratio of nonsynonymous to synonymous polymorphism (P_a/P_s). Under neutrality $D_a/D_s = P_a/P_s$ while a greater D_a/D_s than P_a/P_s indicates that positive selection drives the fixation of beneficial mutations. This can be somewhat problematic in this group of nematodes because the total level of divergence among species is high enough to make saturation at silent sites a real issue. With saturation, D_s would tend to be underestimated, leading to an inflated D_a/D_s ratio relative to P_a/P_s, thereby resulting in a false inference of positive selection. In our case, the MK test shows highly significant departures from neutrality for *odr-7*, *ceh-36a*, *ceh-36b*, and *tax-2*, but not for any other genes (Table 2).

To address the issue of saturation more carefully, we used a likelihood ratio test comparing the d_N/d_S ratio along the within- and between-species branches of the phylogenetic tree including the *C. remanei* alleles and the *C. briggsae* ortholog (Hasegawa *et al.* 1998). The likelihood framework is better suited for divergent sequences because it corrects for multiple substitutions, and takes into account the transition/transversion rate ratio and codon usage bias. The likelihood ratio test is consistent with positive selection acting on *tax-2*, but

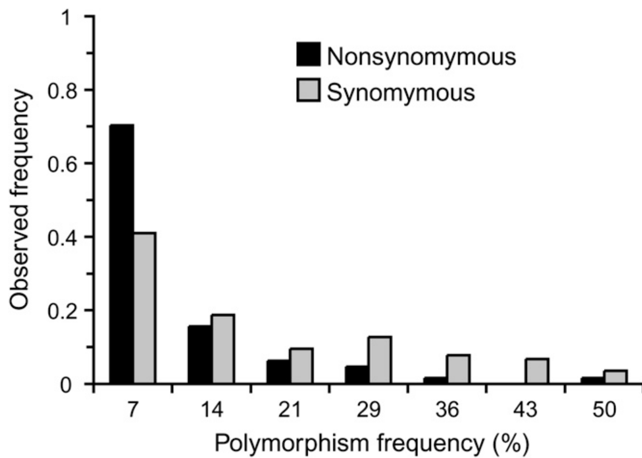

FIGURE 4.—Frequency distribution of nonsynonymous and synonymous polymorphisms in *C. remanei*. Amino acid replacement polymorphism is strongly skewed toward rare variants, suggesting that slightly deleterious mutations segregate in *C. remanei*.

there are no significant differences between d_N/d_S within and among species for the transcription factors *odr-7*, *ceh-36a*, and *ceh-36b* underlying the potentially biased results obtained with the MK test (Table 2). Interestingly, the d_N/d_S ratio within species is significantly higher than between species for *odr-10* indicating that purifying selection acts to remove deleterious segregating variation at this locus.

Tests of selection that focus solely on within-species polymorphism are obviously not subject to the problem of saturation between species. Tajima's *D* (Tajima 1989) measures the distribution of allele frequencies. Negative values indicate population expansion, population bottleneck, selective sweep and/or an increase in the intensity of purifying selection, while positive values are suggestive of a population reduction, balanced selection or recent bottleneck. Tajima's *D* values are negative for the two *ceh-36* duplicates and for *odr-3*, consistent with the abundance of rare polymorphisms and the presence of slightly deleterious alleles, while Tajima's *D* values are positive for *odr-7*, *odr-10*, *odr-1*, and *tax-2* (Table 2). Nevertheless, none of the Tajima's *D* values are significantly different from those expected under neutrality, and thus do not support nonneutral demographic events nor the action of selection (positive or negative).

DISCUSSION

For any organism, the probability of survival and maintenance in the short and long term depends greatly on its capacity to interact with the environment. Organisms rely on a variety of sensory modes to perform this task, but chemosensation is commonly used across a wide taxonomic distribution, from prokaryotes to metazoans. Here, we make use of the extensive genetic information available for two chemosensory pathways and investigate nucleotide diversity across these pathways in an attempt to understand the evolution of olfaction in Caenorhabditis nematodes. We find large differences in the rate of evolution across the functional pathways that specify the olfactory response within these nematodes. Are these differences consistent with general expectations for evolution along pathways or do they provide insights into the unique properties of this particular system?

Variation within and between species: *C. elegans* and its relative species live in the soil, in compost heaps or associated with invertebrates (Baird 1999; Barriere and Felix 2005). Although the exact migratory history is not clear (Phillips 2006), patterns of polymorphisms in *C. elegans* and *C. briggsae* indicate a recent worldwide colonization from a limited number of populations (see also Jovelin *et al.* 2003; Cutter *et al.* 2006b). Furthermore, levels of genetic variation at local scales also suggest that *C. elegans* is a colonizer (Barriere and Felix 2005). Consistent with previous studies, we find limited genetic variation within both *C. elegans* and *C. briggsae*, and severalfold more variation within *C. remanei* (Graustein *et al.* 2002; Jovelin *et al.* 2003; Sivasundar and Hey 2003; Barriere and Felix 2005; Haag and Ackerman 2005; Cutter 2006; Cutter *et al.* 2006a,b). For example, for these olfactory genes, polymorphism in *C. remanei* is on average 86 times higher than in *C. elegans* and 18 times higher than in *C. briggsae*. Some of this difference could potentially be explained by the difference in mating system between these species (selfing in *C. elegans* and *C. briggsae vs.* outcrossing in *C. remanei*), which should lead to an expected twofold decrease in variation within populations (Pollak 1987; Nordborg 2000). Some of the differences in variation between *C. elegans* and *C. briggsae* can be attributed to the fact that our *C. briggsae* samples are drawn from two divergent clades (northern and southern Hemisphere; Cutter *et al.* 2006b). It is important to remember, however, that we sampled from the worldwide distribution of the selfing species, but focus within a single population for *C. remanei*. Although the influence of the interaction between selfing and natural selection on global patterns of genetic variation are potentially quite complex (Charlesworth and Wright 2001), the very limited degree of variation, especially within *C. elegans*, is strongly suggestive that other demographic factors, such as the combination of selection and migration, are important for shaping variation within these species (Jovelin *et al.* 2003; Phillips 2006). Ours is the first study to collect this data from full-length sequence for a number of functionally related genes. This analysis reveals, for example, that levels of polymorphism within silent (including introns) and replacement sites within a gene are very similar within the selfing species, probably indicating the action of demographic processes that

TABLE 2

Selective forces acting on olfactory genes

Locus	McDonald–Kreitman test		Tajima's D	Likelihood ratio test			
	D_a/D_s	P_a/P_s		ω_w	ω_b	ω_w/ω_b	$2\Delta l$
odr-7	170/180	10/44***	0.6473	0.0476	0.0975	0.4882	3.3581
ceh-36a	97/121	12/46***	−0.6108	0.0934	0.1078	0.8664	0.1998
ceh-36b	85/112	3/18**	−0.9598	0.0544	0.1374	0.3959	3.5974
odr-10	51/168	4/8	0.7933	0.106	0.0339	3.1268	5.1736*
odr-3	3/109	1/12	−0.0212	0.0096	0.0048	2	0.3079
odr-1	250/464	4/18	1.2393	0.0522	0.0476	1.0966	0.0616
tax-2	121/334	4/66***	0.1925	0.0128	0.0516	0.2481	19.5539***

The cyclic nucleotide-gated channel *tax-2* shows an excess of divergence relative to polymorphism, indicating that positive selection shaped the pattern of divergence at this locus. By contrast, *odr-10* shows an excess of within-species variation, indicating that selection acts to remove slightly deleterious mutations. D_a/D_s, ratio of the number of fixed amino acid replacements to the number of fixed synonymous substitutions; P_a/P_s, ratio of the number of amino acid replacement polymorphism to the number of synonymous polymorphism; ω_w, d_N/d_S along the within-species branch; ω_b, d_N/d_S along the between-species branch. $\omega_w/\omega_b < 1$ indicates positive selection whereas $\omega_w/\omega_b > 1$ indicates purifying selection. $2\Delta l$, likelihood ratio statistic. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

influence the entire genome simultaneously, whereas silent sites are on average 27 times more polymorphic than replacement sites in *C. remanei*, which is more indicative of the signature of natural selection on specific gene function (Table 1, Jovelin *et al.* 2003).

Molecular evolution of the olfactory system: *C. elegans* has a large repertoire of ~1300 chemoreceptors resulting from lineage-specific expansion of diverse seven-transmembrane receptor (SR) families by extensive gene duplications, gene losses, and nonfunctionalization (Troemel *et al.* 1995; Robertson 1998, 2000, 2001). Although diversifying selection acting on chemoreceptors is weak and limited to the *srz* family (Thomas *et al.* 2005), chemoreceptors appear to be diverging rapidly (Robertson 2000, 2001; Stein *et al.* 2003). Opportunistic interaction with a new ligand may confer a selective advantage over short periods of time, and as a consequence chemoreceptors may evolve mainly because of relaxed selection allowing subsequent diversification. The presence of polymorphism for functional alleles at chemoreceptor loci among *C. elegans* wild isolates (Stewart *et al.* 2005) supports this idea and may reflect adaptation to local conditions. Somewhat unexpectedly, we find that *str odr-10*, which is a chemoreceptor for diacetyl (Sengupta *et al.* 1996), is among the most conserved genes of the two olfactory pathways both between (Figure 1) and within species (Table 1), and appears to be under strong purifying selection (Table 2). While *C. elegans* and *C. briggsae* chemoreceptor orthologs from the *str* family show on average 59% amino acid identity (see Thomas *et al.* 2005), which is much lower than for other orthologous genes (80% aa identity) (Stein *et al.* 2003), the *str* ODR-10 protein shows 86% identity between *C. elegans* and *C. briggsae*. Because diacetyl is a metabolite synthesized as a byproduct of pyruvate and acetaldehyde metabolism in bacteria, the ability to detect food sources must confer a certain selective advantage and perhaps leads to strong purifying selection on *odr-10*. It is clear, however, that the hypothesis that most change in a signal transduction pathway might be concentrated in the receptor does not hold in this case.

Divergence in regulatory and structural genes: We found higher levels of both divergence and polymorphism in developmental regulatory genes than in the structural elements of the signal transduction pathways (Figure 2 and Table 1). The traditional MK test of selection (McDonald and Kreitman 1991) suggests that this divergence can be attributed to positive selection acting on the regulatory genes (Table 2). This result is at least partially attributable to the fact that extensive sequence divergence among these species leads to a bias in the direction of detecting selection. Using a likelihood test that is less sensitive to saturation (Hasegawa *et al.* 1998), we do not find any evidence for selection on the regulatory genes, but do find evidence for positive selection on the *tax-2* ion channel (Table 2). It is somewhat surprising that we find such high levels of allelic variation and divergence within these regulatory genes, given the common expectation that transcription factors are highly constrained (Carroll 2005, 2008). However, these transcription factors are responsible for the terminal differentiation of the olfactory neurons (Sengupta *et al.* 1994; Sagasti *et al.* 1999; Lanjuin *et al.* 2003), suggesting that mutations in these genes may have discrete phenotypic effects. Such discrete effects may reduce potential negative pleiotropic effects of transcription factors, allowing them to evolve novel functions (Lynch and Wagner 2008; Wagner and Lynch 2008). The variation that we observe here therefore suggests the presence of the raw material necessary for the evolution of chemosensory neuron diversity. Overall, we find that significant evolutionary change can occur within the coding

regions of transcription factors, which is consistent with a large number of recent studies (Sutton and Wilkinson 1997; Barrier *et al.* 2001; Galant and Carroll 2002; Ronshaugen *et al.* 2002; Fares *et al.* 2003; Jia *et al.* 2003, 2004; Lawton-Rauh *et al.* 2003; Martinez-Castilla and Alvarez-Buylla 2003; Balakirev and Ayala 2004; Moore *et al.* 2005) but at odds with the emerging evo-devo paradigm of strong conservation of *trans*-acting factors (Carroll 2005, 2008).

From pathways to networks: Because mutations affecting genes acting upstream in genetic pathways have potentially more pleiotropic effects than those affecting more downstream genes, we might expect genes acting early to be more selectively constrained than genes further downstream. For example, an analysis of nucleotide diversity in Drosophila suggested a lower level of polymorphism for regulatory genes than for structural genes (Moriyama and Powell 1996). Contrary to expectation, we find no relationship between pathway location and the rate of molecular evolution (Figure 2). Similarly, detailed evolutionary analysis of other genetic pathways, including the flower developmental pathway in Arabidopsis (Olsen *et al.* 2002), the anthocyanin biosynthetic pathway in Ipomoea (Rausher *et al.* 1999), the Ras pathway (Riley *et al.* 2003) and two NK homeobox genes in Drosophila (Balakirev and Ayala 2004) show that there is in general no simple relationship between selective constraints acting on a gene and its position within a genetic pathway. Positional information may therefore be a poor predictor of gene evolution, and selective constraints acting on a gene may depend primarily on its function and thus may be pathway specific (Cork and Purugganan 2004) or may depend on the way in which the pathway is expanded over evolutionary time (Wilkins 2005). For example, the extensive level of conservation for the G protein *odr-3* suggests that this gene is under strong purifying selection for its role in chemosensation and/or cilium morphogenesis (Jovelin *et al.* 2003). Moreover, the level of conservation among the other G proteins (Figure 2) correlates with the magnitude of pleiotropic effects (Bargmann *et al.* 1993; Zwaal *et al.* 1997; Roayaie *et al.* 1998; Jansen *et al.* 1999; Lans *et al.* 2004), suggesting different levels of selective constraints acting on these G proteins (see Jovelin and Phillips 2005).

Genomic analysis of developmental and physiological pathways is a valuable means of generating hypotheses of how systems evolve, but even moving from the analysis of individual genes to coherent pathways may not be sufficient when the genes of interest function within much broader networks. Even with very complete information on divergence and polymorphism for a large number of genes, these hypotheses must ultimately be tested using a functional approach—something for which the nematode model system is ideally suited.

Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources (NCRR). Benjamin White isolated the *C. elegans* strains from Oregon. Scott Baird kindly provided the *C. remanei* strains from Dayton, Ohio. We thank two anonymous reviewers for comments that helped improve the manuscript. This work was supported by a grant from the National Institutes of Health (GM54185 and AG029377) and the National Science Foundation (NSF) (DEB-0236180). R.J. is an associate of the IGERT Training Program for Development, Evolution, and Genomics, NSF DGE-9972830.

LITERATURE CITED

Ache, B. W., and J. M. Young, 2005 Olfaction: diverse species, conserved principles. Neuron **48:** 417–430.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Baird, S. E., 1999 Natural and experimental associations of *Caenorhabditis remanei* with *Trachelipus rathkii* and other terrestrial isopods. Nematology **1:** 471–475.

Balakirev, E. S., and F. J. Ayala, 2004 Nucleotide variation in the *tinman* and *bagpipe* homeobox genes of *Drosophila melanogaster.* Genetics **166:** 1845–1856.

Bargmann, C. I., E. Hartwieg and H. R. Horvitz, 1993 Odorant-selective genes and neurons mediate olfaction in *C. elegans.* Cell **74:** 515–527.

Bargmann, C. I., and J. M. Kaplan, 1998 Signal transduction in the *Caenorhabditis elegans* nervous system. Annu. Rev. Neurosci. **21:** 279–308.

Barrier, M., R. H. Robichaux and M. D. Purugganan, 2001 Accelerated regulatory gene evolution in an adaptive radiation. Proc. Natl. Acad. Sci. USA **98:** 10208–10213.

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

Bergamasco, C., and P. Bazzicalupo, 2006 Chemical sensitivity in *Caenorhabditis elegans.* Cell. Mol. Life Sci. **63:** 1510–1522.

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

Butcher, R. A., M. Fujita, F. C. Schroeder and J. Clardy, 2007 Small-molecule pheromones that control dauer development in Caenorhabditis elegans. Nat. Chem. Biol. **3:** 420–422.

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

Carroll, S. B., 2005 *Endless Forms Most Beautiful: The New Science of Evo-Devo and the Making of the Animal Kingdom.* W. W. Norton, New York.

Carroll, S. B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134:** 25–36.

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

Coburn, C. M., and C. I. Bargmann, 1996 A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans.* Neuron **17:** 695–706.

Colbert, H. A., T. L. Smith and C. I. Bargmann, 1997 OSM-9, a novel protein with structural similarity to channels, is required for olfaction, mechanosensation, and olfactory adaptation in *Caenorhabditis elegans.* J. Neurosci. **17:** 8259–8269.

Cork, J. M., and M. D. Purugganan, 2004 The evolution of molecular genetic pathways and networks. BioEssays **26:** 479–484.

Cutter, A. D., 2006 Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans.* Genetics **172:** 171–184.

Cutter, A. D., S. E. Baird and D. Charlesworth, 2006a High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei.* Genetics **174:** 901–913.

Cutter, A. D., and B. Charlesworth, 2006 Selection intensity on preferred codons correlates with overall codon usage bias in *Caenorhabditis remanei.* Curr. Biol. **16:** 2053–2057.

Cutter, A. D., M. A. Felix, A. Barriere and D. Charlesworth, 2006b Patterns of nucleotide polymorphism distinguish temperate and tropical wild isolates of *Caenorhabditis briggsae*. Genetics **173:** 2021–2031.

Cutter, A. D., J. D. Wasmuth and M. L. Blaxter, 2006c The evolution of biased codon and amino acid usage in nematode genomes. Mol. Biol. Evol. **23:** 2303–2315.

Daniels, S., M. Ailion, J. Thomas and P. Sengupta, 2000 *egl-4* acts through a transforming growth factor-beta/SMAD pathway in *Caenorhabditis elegans* to regulate multiple neuronal circuits in response to sensory cues. Genetics **156:** 123–141.

de Bono, M., 2003 Molecular approaches to aggregation behavior and social attachment. J. Neurobiol. **54:** 78–92.

de Bono, M., D. M. Tobin, M. W. Davis, L. Avery and C. I. Bargmann, 2002 Social feeding in *Caenorhabditis elegans* is induced by neurons that detect aversive stimuli. Nature **419:** 899–903.

Dekker, T., I. Ibba, K. P. Siju, M. C. Stensmyr and B. S. Hansson, 2006 Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia*. Curr. Biol. **16:** 101–109.

Duboule, D., and A. S. Wilkins, 1998 The evolution of 'bricolage'. Trends Genet. **14:** 54–59.

Fares, M. A., D. Bezemer, A. Moya and I. Marin, 2003 Selection on coding regions determined *Hox7* genes evolution. Mol. Biol. Evol. **20:** 2104–2112.

Fay, J. C., G. J. Wyckoff and C. I. Wu, 2001 Positive and negative selection on the human genome. Genetics **158:** 1227–1234.

Galant, R., and S. B. Carroll, 2002 Evolution of a transcriptional repression domain in an insect Hox protein. Nature **415:** 910–913.

Graustein, A., J. M. Gaspar, J. R. Walters and M. F. Palopoli, 2002 Levels of DNA polymorphism vary with mating system in the nematode genus Caenorhabditis. Genetics **161:** 99–107.

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

Haag, E. S., and A. V. Doty, 2005 Sex determination across evolution: connecting the dots. PLoS Biol. **3:** e21.

Hahn, M. W., and A. D. Kern, 2005 Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol. Biol. Evol. **22:** 803–806.

Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41:** 95–98.

Hasegawa, M., Y. Cao and Z. Yang, 1998 Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol. Biol. Evol. **15:** 1499–1505.

Hillier, L. W., R. D. Miller, S. E. Baird, A. Chinwalla, L. A. Fulton *et al.*, 2007 Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. **5:** e167.

Hirsh, A. E., H. B. Fraser and D. P. Wall, 2005 Adjusting for selection on synonymous sites in estimates of evolutionary distance. Mol. Biol. Evol. **22:** 174–177.

Hoekstra, H. E., and J. A. Coyne, 2007 The locus of evolution: evo devo and the genetics of adaptation. Evol. Int. J. Org. Evol. **61:** 995–1016.

Jacob, F., 1977 Evolution and tinkering. Science **196:** 1161–1166.

Jansen, G., K. L. Thijssen, P. Werner, M. van der Horst, E. Hazendonk *et al.*, 1999 The complete family of genes encoding G proteins of *Caenorhabditis elegans*. Nat. Genet. **21:** 414–419.

Jeong, P. Y., M. Jung, Y. H. Yim, H. Kim, M. Park *et al.*, 2005 Chemical structure and biological activity of the Caenorhabditis elegans dauer-inducing pheromone. Nature **433:** 541–545.

Jia, L., M. T. Clegg and T. Jiang, 2003 Excess non-synonymous substitutions suggest that positive selection episodes occurred during the evolution of DNA-binding domains in the *Arabidopsis* R2R3-MYB gene family. Plant Mol. Biol. **52:** 627–642.

Jia, L., M. T. Clegg and T. Jiang, 2004 Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies *indica* and *japonica* genomes. Plant Physiol. **134:** 575–585.

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

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

King, M. C., and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science **188:** 107–116.

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

Komatsu, H., I. Mori, J. S. Rhee, N. Akaike and Y. Ohshima, 1996 Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. Neuron **17:** 707–718.

L'Etoile, N. D., and C. I. Bargmann, 2000 Olfaction and odor discrimination are mediated by the *C. elegans* guanylyl cyclase ODR-1. Neuron **25:** 575–586.

Lanjuin, A., M. K. VanHoven, C. I. Bargmann, J. K. Thompson and P. Sengupta, 2003 *Otx/otd* homeobox genes specify distinct sensory neuron identities in *C. elegans*. Dev. Cell **5:** 621–633.

Lans, H., S. Rademakers and G. Jansen, 2004 A network of stimulatory and inhibitory Galpha-subunits regulates olfaction in *Caenorhabditis elegans*. Genetics **167:** 1677–1687.

Lawton-Rauh, A., R. H. Robichaux and M. D. Purugganan, 2003 Patterns of nucleotide variation in homoeologous regulatory genes in the allotetraploid Hawaiian silversword alliance (Asteraceae). Mol. Ecol. **12:** 1301–1313.

Linn, Jr., C., J. L. Feder, S. Nojima, H. R. Dambroski, S. H. Berlocher *et al.*, 2003 Fruit odor discrimination and sympatric host race formation in *Rhagoletis*. Proc. Natl. Acad. Sci. USA **100:** 11490–11493.

Lynch, V. J., and G. P. Wagner, 2008 Resurrecting the role of transcription factor change in developmental evolution. Evolution **62:** 2131–2154.

Martinez-Castilla, L. P., and E. R. Alvarez-Buylla, 2003 Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. Proc. Natl. Acad. Sci. USA **100:** 13407–13412.

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

Moore, R. C., S. R. Grant and M. D. Purugganan, 2005 Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. Mol. Biol. Evol. **22:** 91–103.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13:** 261–277.

Nayak, S., J. Goree and T. Schedl, 2005 fog-2 and the evolution of self-fertile hermaphroditism in Caenorhabditis. PLoS Biol. **3:** e6.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics **154:** 923–929.

Olsen, K. M., A. Womack, A. R. Garrett, J. I. Suddith and M. D. Purugganan, 2002 Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. Genetics **160:** 1641–1650.

Ortiz-Barrientos, D., B. A. Counterman and M. A. Noor, 2004 The genetics of speciation by reinforcement. PLoS Biol. **2:** e416.

Phillips, P. C., 2006 One perfect worm. Trends Genet. **22:** 405–407.

Pollak, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117:** 353–360.

Rausher, M. D., R. E. Miller and P. Tiffin, 1999 Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol. Biol. Evol. **16:** 266–274.

Riddle, D., and P. Albert, 1997 Genetic and environmental regulation of dauer larva development, pp. 739–768 in *C. elegans II*, edited by D. Riddle, T. Blumenthal, B. Meyer and J. Priess. Cold Spring Harbor Laboratory Press, Plainview, NY.

Riley, R. M., W. Jin and G. Gibson, 2003 Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in Drosophila. Mol. Ecol. **12:** 1315–1323.

Roayaie, K., J. G. Crump, A. Sagasti and C. I. Bargmann, 1998 The G alpha protein ODR-3 mediates olfactory and nociceptive func-

tion and controls cilium morphogenesis in *C. elegans* olfactory neurons. Neuron **20:** 55–67.

Robertson, H. M., 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, H. M., 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, H. M., 2001 Updating the *str* and *srj* (*stl*) families of chemoreceptors in *Caenorhabditis* nematodes reveals frequent gene movement within and between chromosomes. Chem. Senses **26:** 151–159.

Ronshaugen, M., N. McGinnis and W. McGinnis, 2002 Hox protein mutation and macroevolution of the insect body plan. Nature **415:** 914–917.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer and R. Rozas, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Sackton, T. B., B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark *et al.*, 2007 Dynamic evolution of the innate immune system in Drosophila. Nat. Genet. **39:** 1461–1468.

Sagasti, A., O. Hobert, E. R. Troemel, G. Ruvkun and C. I. Bargmann, 1999 Alternative olfactory neuron fates are specified by the LIM homeobox gene *lim-4*. Genes Dev. **13:** 1794–1806.

Sengupta, P., H. A. Colbert and C. I. Bargmann, 1994 The *C. elegans* gene *odr-7* encodes an olfactory-specific member of the nuclear receptor superfamily. Cell **79:** 971–980.

Sengupta, P., J. H. Chou and C. I. Bargmann, 1996 *odr-10* encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. Cell **84:** 899–909.

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

Sivasundar, A., and J. Hey, 2003 Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species. Genetics **163:** 147–157.

Stein, L., Z. Bao, D. Blasiar, T. Blumenthal, M. Brent *et al.*, 2003 The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. **1:** 166–192.

Stensmyr, M. C., T. Dekker and B. S. Hansson, 2003 Evolution of the olfactory code in the *Drosophila melanogaster* subgroup. Proc. R. Soc. Lond. Ser. B **270:** 2333–2340.

Stern, D. L., and V. Orgogozo, 2008 The loci of evolution: How predictable is genetic evolution? Evolution **62:** 2155–2177.

Stewart, M. K., N. L. Clark, G. Merrihew, E. M. Galloway and J. H. Thomas, 2005 High genetic diversity in the chemoreceptor superfamily of *Caenorhabditis elegans*. Genetics **169:** 1985–1996.

Sutton, K. A., and M. F. Wilkinson, 1997 Rapid evolution of a homeodomain: evidence for positive selection. J. Mol. Evol. **45:** 579–588.

Swofford, D. L., 1998 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Thomas, J. H., J. L. Kelley, H. M. Robertson, K. Ly and W. J. Swanson, 2005 Adaptive evolution in the SRZ chemoreceptor families of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. Proc. Natl. Acad. Sci. USA **102:** 4476–4481.

Tobin, D., D. Madsen, A. Kahn-Kirby, E. Peckol, G. Moulder *et al.*, 2002 Combinatorial expression of TRPV channel proteins defines their sensory functions and subcellular localization in *C. elegans* neurons. Neuron **35:** 307–318.

Toth, A. L., and G. E. Robinson, 2007 Evo-devo and the evolution of social behavior. Trends Genet. **23:** 334–341.

Troemel, E. R., J. H. Chou, N. D. Dwyer, H. A. Colbert and C. I. Bargmann, 1995 Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell **83:** 207–218.

Troemel, E. R., B. E. Kimmel and C. I. Bargmann, 1997 Reprogramming chemotaxis responses: sensory neurons define olfactory preferences in *C. elegans*. Cell **91:** 161–169.

Wagner, G. P., and V. J. Lynch, 2008 The gene regulatory logic of transcription factor evolution. Trends Ecol. Evol. **23:** 377–385.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wilkins, A. S., 2005 Recasting developmental evolution in terms of genetic pathway and network evolution... and the implications for comparative biology. Brain Res. Bull. **66:** 495–509.

Winnepenninckx, B., T. Backeljau and R. De Wachter, 1993 Extraction of high molecular weight DNA from molluscs. Trends Genet. **9:** 407.

Wu, G., D. E. Culley and W. Zhang, 2005 Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. Microbiology **151:** 2175–2187.

Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Zwaal, R. R., J. E. Mendel, P. W. Sternberg and R. H. Plasterk, 1997 Two neuronal G proteins are involved in chemosensation of the *Caenorhabditis elegans* Dauer-inducing pheromone. Genetics **145:** 715–727.

Communicating editor: C. H. Langley