Rapid Sequence Evolution of Transcription Factors Controlling Neuron Differentiation in *Caenorhabditis*

Richard Jovelin

Center for Ecology and Evolutionary Biology, University of Oregon

Whether phenotypic evolution proceeds predominantly through changes in regulatory sequences is a controversial issue in evolutionary genetics. Ample evidence indicates that the evolution of gene regulatory networks via changes in *cis*-regulatory sequences is an important determinant of phenotypic diversity. However, recent experimental work suggests that the role of transcription factor (TF) divergence in developmental evolution may be underestimated. In order to help understand what levels of constraints are acting on the coding sequence of developmental regulatory genes, evolutionary rates were investigated among 48 TFs required for neuronal development in *Caenorhabditis elegans*. Allelic variation was then sampled for 28 of these genes within a population of the related species *Caenorhabditis remanei*. Neuronal TFs are more divergent, both within and between species, than structural genes. TFs affecting different neuronal classes are under different levels of selective constraints. The regulatory genes controlling the differentiation of chemosensory neurons evolve particularly fast and exhibit higher levels of within- and between-species nucleotide variation than TFs required for the development of several neuronal classes and TFs required for motorneuron differentiation. The TFs affecting chemosensory neuron development are also more divergent than chemosensory genes expressed in the neurons they differentiate. These results illustrate that TFs are not as highly constrained as commonly thought and suggest that the role of divergence in developmental regulatory genes during the evolution of gene regulatory networks requires further attention.

Introduction

The diversity of animal form has been a topic of interest to naturalists long before Darwin, but it is only with the pioneering work on *Drosophila* development that the genetic mechanisms responsible for generating morphological diversity could be approached (Lewis 1978; Nüsslein-Volhard and Wieschaus 1980; Nüsslein-Volhard et al. 1984). Perhaps one of the most fascinating results that followed is that animal development relies on a set of genes (i.e., cell adhesion proteins, signaling proteins, and transcription factors, TFs) belonging to gene families that are broadly shared across animal phylogeny and that diversified early in metazoan evolution (Technau et al. 2005; Guder et al. 2006; Nichols et al. 2006; Matus et al. 2007). The conservation of this shared genetic toolkit, although fascinating in its own right, raises the question of how the diversity of forms observed in the animal kingdom arises in the first place. A growing body of evidence indicates that the redeployment of these toolkit genes within gene regulatory networks, through evolution of regulatory sequences, is an important driver of morphological changes at various taxonomical scales (Carroll et al. 2005).

Whether phenotypic evolution proceeds predominantly through changes in regulatory sequences or changes in protein sequences has ignited an intense debate, with the argument in favor of the *cis*-regulatory hypothesis focusing on the prediction of strong conservation of TF function (Carroll 2005, 2008; Hoekstra and Coyne 2007; Wray 2007; Stern and Orgogozo 2008). Nonetheless, recent findings from cross-species gene-swapping experiments indicate that functional equivalence between distant TF orthologs is only partial or nonexistent (Hsia and McGinnis 2003). Moreover, changes in TF coding sequence can result

in profound modifications of body plan (Galant and Carroll 2002; Ronshaugen et al. 2002). A major assumption underlying arguments against a substantial role for TF sequence evolution is that mutations in the coding sequence of highly pleiotropic genes are likely to be deleterious and thus selected against. This notion is supported by a negative correlation between nonsynonymous substitution rate and gene expression breadth (Duret and Mouchiroud 2000), although this relationship can suffer from the specifics of particular gene family function (Jovelin and Phillips 2005). However, in recent reviews, Wagner and Lynch report that contrary to common belief, about a third of human TFs are tissue specific. They also argue that the negative pleiotropic effect of mutations can be reduced by structural protein motifs and alternative splicing that make TFs highly modular. Moreover, the combinatorial mode of action of TFs can result in tissue-specific novel functions by establishing new interactions with cofactors or other TFs restricted to tissues in which all partners are expressed (Lynch and Wagner 2008; Wagner and Lynch 2008). Therefore, it seems that the contribution of TF sequence divergence to the evolution of gene regulatory networks may have been underestimated.

One way to address the levels of constraints among TFs has been to compare TFs with non-TF genes, which has revealed high rates of sequence evolution among the regulatory genes (e.g., Clark et al. 2007; Haerty et al. 2008). Nevertheless, a better understanding of TF sequence evolution requires that we quantify the level of segregating variation within populations at TF loci so as to infer selective pressures and ultimately to examine the functional effects of naturally occurring allelic variation and sequence divergence among orthologs. Early studies of intraspecific nucleotide variation in *Drosophila* have revealed that regulatory genes tend to be less polymorphic than structural genes (Moriyama and Powell 1996). By contrast, analysis of olfactory pathways in *Caenorhabditis* showed that developmental regulatory genes exhibit more between- and within-species variation than structural chemosensory genes (Jovelin et al. 2009). Nonetheless, these studies

included a limited number of TFs. In this study, I test the assumption that TF sequence evolution is highly constrained by investigating the levels of segregating variation in natural populations and divergence between species for a large set of TFs required for neuronal development in *Caenorhabditis elegans*.

The hermaphrodite adult of the nematode *C. elegans* has 302 neurons that are generated in a stereotypical manner such that wild-type individuals have the same number of neurons (White et al. 1986). Neurons acquire their identity through specific patterns of migration, axonal growth, synaptic connectivity, morphology, and unique expression of neurotransmitters and gene batteries. Mutants affecting all aspects of neuron development have been identified, although most known mutants affect only some aspects of terminal differentiation. A large number of mutations are found in TFs that act cell autonomously in neuron differentiation (Hobert 2005). Neuron identity is determined by three genetic programs. Generic features common to all neuron types are specified by panneuronal TFs that act in parallel to terminal selector genes that directly coregulate the expression of terminal gene batteries, thereby differentiating neuron types. These terminal selector genes can also regulate terminal differentiation genes through the control of downstream TFs, thus further diversifying neuron types into subtypes (Hobert 2005, 2008).

Here, I analyze sequence divergence among 48 TFs required for neuron differentiation in *C. elegans* in order to investigate the relationship between evolutionary rates and gene function. Among these TFs, I sampled allelic variation within one population of the related species *Caenorhabditis remanei* for the nearly full length of the coding sequence of 28 loci required for the differentiation of several classes of neurons. *Caenorhabditis remanei* is better suited for this analysis because of substantially higher levels of nucleotide diversity within this species than within *C. elegans* (Graustein et al. 2002; Jovelin et al. 2003, 2009; Barrière and Félix 2005; Cutter et al. 2006). I show that neuronal TF coding sequences exhibit more nonsynonymous polymorphism than the coding sequences of structural genes. Neuronal TFs with distinct functions seem to be under different levels of selective constraints. TFs controlling chemosensory neuron differentiation exhibit both more nonsynonymous polymorphism and amino acid replacement divergence than TFs controlling motorneuron differentiation. The former also accumulate higher levels of within- and between-species nucleotide diversity than TFs required for the development of several classes of neurons. Finally, TFs controlling chemosensory neuron differentiation are more divergent than structural chemosensory genes. These findings illustrate that, contrary to common belief, TFs can harbor high levels of amino acid divergence both within and between species.

Materials and Methods
Ortholog Identification

The set of 48 TFs required for neuron fate determination was obtained through a literature search and from two reviews (Hobert 2005; Von Stetina et al. 2006). The identity of the neuronal cells affected by loss of function mutation in each locus is listed in table 1. Although most of these TFs have roles restricted to nervous system differentiation, some of them (17/48) also play a role in the regulation of nonneuronal tissues. These TFs are *unc-130*, *egl-5*, *egl-43*, *ham-2*, *sem-4*, *ceh-14*, *lin-11*, *cog-1*, *die-1*, *vab-7*, *unc-55*, *mab-5*, *mab-18*, *lin-39*, *fozi-1*, *lin-14*, and *ref-1* (Chisholm 1991; Ahringer 1996; Euling and Ambros 1996; Baum et al. 1999; Eisenmann and Kim 2000; Grant et al. 2000; Nash et al. 2000; Alper and Kenyon 2001; Heid et al. 2001; Gupta et al. 2003; Bando et al. 2005; Inoue et al. 2005; Amin et al. 2007; Fernandes and Sternberg 2007; Guerry et al. 2007; Hwang et al. 2007; Rimann and Hajnal 2007; Shan and Walthall 2008).

The list of 54 structural genes expressed in chemosensory neurons (Bargmann 2006) and known to affect chemotaxis was obtained from Wormbase (Rogers et al. 2008). Sequences of the *C. remanei* alleles corresponding to the 37 structural and putatively X-linked loci (Cutter 2008) were obtained from Genbank (Benson et al. 2008). Strain PB249 has a premature stop codon in *Cre*-F47A4.5 at position 2959 and was discarded prior to analyses. Here, a structural gene refers to a non-TF gene. Genes were classified as structural and TF according to gene annotations in Wormbase and by using the set of annotated *C. elegans* TFs (Reece-Hoyes et al. 2005). A brief description of the function of the genes analyzed is available in supplementary table1, Supplementary Material online.

Blast searches were conducted against the genome assemblies of *Caenorhabditis briggsae* (Stein et al. 2003), *C. remanei*, *Caenorhabditis brenneri*, and *Caenorhabditis japonica* (Genome Sequencing Center, Washington University, St Louis, unpublished) 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.

Strains, Amplification, and Sequencing

The *C. remanei* strains used in this study (PB237, PB241, PB244, PB245, PB247, PB255, PB256, PB257, PB258, PB261, PB263, PB266, PB269, PB271, PB272, and PB285) are derived from one population in Ohio, United States, a gift from Scott Baird, Wright State University, and were maintained following standard *C. elegans* protocols (Brenner 1974). Worms were found associated with isopods collected under a single log during a 2-week period. Strains were founded either with a single gravid female or with a single virgin female mated to a single male and were subsequently inbred for at least six generations of brother–sister mating to minimize intrastrain nucleotide variability. For each of the 16 *C. remanei* strains, RNA was extracted from plates containing worms at all stages of development using the TRI Reagent protocol (Molecular Research Center) and subsequently used to synthesize double-stranded cDNA with the Retroscript kit (Ambion). Primers used for amplification of the coding sequence of 28 TFs were designed from the *C. remanei* genomic sequence. Polymerase chain reactions were processed as

Table 1
Neurons Affected by Loss of Function Mutations in *Caenorhabditis elegans* TFs

Locus	Neurons Affected	Functional Class	References
odr-7	AWA	Chemosensory	Sengupta et al. (1994), Sagasti et al. (1999)
ceh-36	AWC, ASEL	Chemosensory	Chang et al. (2003), Lanjuin et al. (2003), Koga and Ohshima (2004)
ceh-37	AWB, ADF	Chemosensory	Lanjuin et al. (2003)
che-1	ASE	Chemosensory	Chang et al. (2003), Uchida et al. (2003)
lsy-2	ASEL	Chemosensory	Johnston and Hobert (2005)
fozi-1	ASER	Chemosensory	Johnston et al. (2006)
lin-11	AWA, ASG	Chemosensory	Sarafi-Reinach et al. (2001)
	AVG, AIZ, PVP	Interneuron	Hobert et al. (1998), Satterlee et al. (2001), Hutter (2003); Tsalik et al. (2003)
cog-1	ASER	Chemosensory	Chang et al. (2003)
unc-130	ASG, AWC	Chemosensory	Sarafi-Reinach and Sengupta (2000)
	RMD	Motorneuron	Sarafi-Reinach and Sengupta (2000)
	ASH	Chemo-, mechano-, osmosensory	Sarafi-Reinach and Sengupta (2000)
lim-4	AWB, ADF	Chemosensory	Sagasti et al. (1999), Zheng et al. (2005)
	RID	Interneuron	Tsalik et al. (2003)
ceh-23	AIY	Interneuron	Altun-Gultekin et al. (2001)
fax-1	AVK, AVE, AVA, AVB	Interneuron	Wightman et al. (1997), Much et al. (2000), Wightman et al. (2005)
	HSN	Motorneuron	Wightman et al. (1997)
ceh-10	CAN	Unknown	Forrester et al. (1998)
	RMED	Motorneuron	Forrester et al. (1998)
	AIY	Interneuron	Altun-Gultekin et al. (2001)
mec-3	FLP, PVD, ALM, PLM, AVM, PVM	Mechanosensory	Chalfie and Sulston (1981), Way and Chalfie (1988), Chalfie and Au (1989), Way and Chalfie (1989), Tsalik et al. (2003)
unc-30	D-type	Motorneuron	Jin et al. (1994)
	PVPR	Interneuron	Wightman et al. (1997), Hutter (2003)
ceh-2	M3	Motorneuron	Aspock et al. (2003)
egl-5	HSN	Motorneuron	Desai et al. (1988), Baum et al. (1999)
	PLM	Mechanosensory	Toker et al. (2003)
	R3A, R4A, R5A, R3B, R4B, R5B, R7B	Sensory	Lints and Emmons (1999), Lints et al. (2004)
egl-43	HSN	Motorneuron	Desai et al. (1988), Baum et al. (1999)
ham-2	HSN	Motorneuron	Desai et al. (1988), Baum et al. (1999)
sem-4	PHC	Sensory	Mitani et al. (1993), Toker et al. (2003)
	PVW, PLN	Interneuron	Toker et al. (2003)
	HSN	Motorneuron	Desai et al. (1988)
ceh-14	AFD	Thermosensory	Cassata et al. (2000), Satterlee et al. (2001)
lim-6	RIS	Interneuron	Hobert et al. (1999), Tsalik et al. (2003)
	AVL, DVB	Motorneuron	Hobert et al. (1999)
	ASEL	Chemosensory	Hobert et al. (1999)
unc-42	RMD, RME	Motorneuron	Baran et al. (1999)
	AVA, AVD, AVE, AVK	Interneuron	Baran et al. (1999), Wightman et al. (2005)
	ASH	Chemo-, mechano-, osmosensory	Baran et al. (1999)
zag-1	DA, DB, DD, HSN, VC, VD, PDB	Motorneuron	Clark and Chiu (2003), Wacker et al. (2003)
	SABV, AVE, PVQ	Interneuron	Clark and Chiu (2003), Wacker et al. (2003)
	ALM, AVM, PVM, PVD	Mechanosensory	Clark and Chiu (2003), Wacker et al. (2003)
	ADE, PDE	Sensory	Clark and Chiu (2003), Wacker et al. (2003)
unc-3	ASI	Chemosensory	Prasad et al. (1998), Kim et al. (2005)
	VA, VB, VC, DA	Motorneuron	Prasad et al. (1998)
	AVG	Interneuron	Wightman et al. (1997)
ttx-3	AIY, AIA	Interneuron	Hobert et al. (1997), Altun-Gultekin et al. (2001), Tsalik et al. (2003)
cfi-1	URA	Motorneuron	Shaham and Bargmann (2002)
	IL2	Sensory	Shaham and Bargmann (2002)
	AVD, PVC	Interneuron	Shaham and Bargmann (2002)
egl-46	HOB	Sensory	Yu et al. (2003)
	FLP	Mechanosensory	Mitani et al. (1993), Wu et al. (2001)
	HSN	Motorneuron	Desai et al. (1988), Wu et al. (2001)
ceh-17	ALA	Unknown	Pujol et al. (2000)
	SIAD	Unknown	Pujol et al. (2000)
	SIAV	Unknown	Pujol et al. (2000)
egl-44	FLP	Mechanosensory	Mitani et al. (1993), Wu et al. (2001)
	HSN	Motorneuron	Desai et al. (1988), Wu et al. (2001)
	HOB	Sensory	Yu et al. (2003)
ttx-1	AFD	Thermosensory	Satterlee et al. (2001)
unc-86	RIH, AIM, URX	Interneuron	Sze et al. (2002), Qin and Powell-Coffman (2004)
	NSM, HSN	Motorneuron	Desai et al. (1988), Baum et al. (1999), Sze et al. (2002)

Table 1
Continued

Locus	Neurons Affected	Functional Class	References
	FLP, PVD, AVM, PVM, ALM, PLM	Mechanosensory	Chalfie et al. (1981)
lin-39	Q neuroblasts	NA	Clark et al. (1993)
mab-5	Q neuroblasts	NA	Kenyon (1986)
ahr-1	RMEL/R	Motorneuron	Huang et al. (2004)
	AVM, PLM, PQR	Mechanosensory	Qin and Powell-Coffman (2004)
	SDQR, URX	Interneuron	Qin and Powell-Coffman (2004)
	AQR	Sensory	Qin and Powell-Coffman (2004)
die-1	ASEL	Chemosensory	Chang et al. (2004)
mab-18	R6B	Sensory	Lints et al. (2004)
pag-3	BDU, PVQ	Interneuron	Jia et al. (1996), Cameron et al. (2002)
	ALM	Mechanosensory	Cameron et al. (2002)
	VA, VB	Motorneuron	Cameron et al. (2002)
unc-4	VA, DA	Motorneuron	Miller et al. (1992), Winnier et al. (1999), Esmaeili et al. (2002), Von Stetina et al. (2007)
unc-55	VD	Motorneuron	Zhou and Walthall (1998), Melkman and Sengupta (2005)
alr-1	AWA, ASG, ADF	Chemosensory	Melkman and Sengupta (2005)
	AFD	Thermosensory	Melkman and Sengupta (2005)
	VD	Motorneuron	Melkman and Sengupta (2005)
vab-7	DB	Motorneuron	Esmaeili et al. (2002)
cnd-1	DA, DB, DD	Motorneuron	Hallam et al. (2000)
aha-1	RMEL/R	Motorneuron	Huang et al. (2004)
	SDQR	Interneuron	Qin and Powell-Coffman (2004)
	AVM	Mechanosensory	Qin and Powell-Coffman (2004)
ceh-12	VB	Motorneuron	Von Stetina et al. (2007)
lin-14	AVM, PVM	Mechanosensory	Mitani et al. (1993)
lin-32	AVM, PLM, PVM, PVD, PQR	Mechanosensory	Chalfie and Au (1989), Zhao and Emmons (1995)
	SDQR	Interneuron	Zhao and Emmons (1995)
	RnA, RnB	Sensory	Zhao and Emmons (1995)
ref-1	NSM	Motorneuron	Lanjuin et al. (2006)

Phenotypes are not necessarily 100% penetrant. Functional classes were obtained from WormAtlas (www.wormatlas.org).

described in (Jovelin et al. 2003) using 0.3 µl of TrueStart Taq polymerase (Fermentas) and 1 µl of template cDNA with the following conditions: hot start at 95 °C for 3 min, followed by 35 cycles of 95 °C for 1 min, 55 °C for 1 min and 72 °C for 3 min. For some of the samples, the annealing temperature was 58 °C. Amplifications were gel purified (Qiagen) and sequenced using automated sequencers at the University of Oregon sequencing facility. The primers used for amplification were also used for sequencing, allowing sequences to be confirmed on both strands. Internal primers were also used for sequencing *Cre-die-1*, *Cre-egl-43*, and *Cre-unc-3*. *Cre-zag-1* was amplified and sequenced using two sets of primers. All sequence changes were rechecked visually against sequencing chromatograms. After trimming primer sequences, polymorphism data were obtained for 90%, on average, of the coding sequence of the *C. remanei* genes.

Sequence Analyses

Protein sequences were aligned by eye using BioEdit (Hall 1999) and subsequently used to generate codon-based DNA sequence alignments. Pairwise alignments used to estimate species divergence are available online as Supplementary Material online. Maximum likelihood (ML) estimates of the rates of nonsynonymous (d_N) and synonymous (d_S) substitutions between the most closely related species *C. remanei* and *C. briggsae* (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 (supplementary table1, Supplementary Material online). Codon frequencies are the product of the observed nucleotide frequencies at each codon position. The rate of synonymous changes correcting for selection at silent sites ($d_{S'}$) (Hirsh et al. 2005) was computed using the regression between the codon adaptation index (CAI) (Sharp and Li 1987) and the rate of synonymous substitutions, d_S, from the full gene set ($N = 104$), including the TFs and chemosensory structural genes. CAI was computed for the *C. remanei* genes using CAI Calculator (Wu et al. 2005) and with the *C. elegans* reference gene set defined by Carbone et al. (2003). The rates of radical and conservative nonsynonynmous changes were computed using five classifications of amino acid chemical properties: charge; polarity; volume and polarity; charge and aromatic; charge and polarity (Zhang 2000; Hanada et al. 2006). Within each functional class, changes between amino acid groups are defined as radical and changes within groups are defined as conservative. The proportions of radical and conservative changes were computed according to the method of Zhang (2000) using *C. remanei* and *C. briggsae* orthologs and were used to obtain the rates of radical (d_R) and conservative (d_C) nonsynonymous changes using a Jukes–Cantor approximation (Jukes and Cantor 1969). The transition/transversion rate ratio was estimated by ML with CODEML.

Population genetic analyses including measures of nucleotide diversity (π, Nei 1987) and Tajima's D (Tajima 1989) were performed using DnaSP 4.1 (Rozas et al.

2003). Tajima's D was computed using synonymous sites only, although analyses using all sites gave similar results. Gene trees of the *Cre-ceh-37* and *Cre-mec-3* sequences were inferred with maximum parsimony using PAUP* 4.0b10 (Swofford 1998) and rooted with the corresponding *C. briggsae* ortholog. Difference in rate divergence along the branches of the two *Cre-ceh-37* paralogs was tested with a likelihood ratio test (LRT) using CODEML and using the gene tree (Kiontke et al. 2004) including *ceh-37* sequences from *C. japonica*, *C. elegans*, *C. brenneri*, *C. briggsae*, and *C. remanei*.

Results

Identification of Orthologs and Duplicates

For most of the *C. elegans* genes, a single ortholog could be identified in *C. briggsae* and *C. remanei*, with the exception of *ceh-37* and *mec-3* for which two similar sequences were found on distinct contigs in *C. remanei*. Reciprocal Blast searches with each of the *C. remanei* sequences similar to *ceh-37* identified a single sequence corresponding to *ceh-37*. However, because of the level of divergence between the two *C. remanei* sequences ($d_N/d_S = 0.0932$) and because introns are too divergent to be aligned, these sequences are likely to be duplicates of *ceh-37* rather than being distinct alleles. Although amplifications of *Cre-ceh-37b* failed for 5 *C. remanei* strains, sampling of *Cre-ceh-37a* and *Cre-ceh-37b* in a single *C. remanei* population recovered sequences that form distinct clades (data not shown), further supporting the duplication event of *ceh-37* in this species. Previously, we reported the duplication of *ceh-36* in *C. remanei* (Jovelin et al. 2009). The specificity of the duplication of *ceh-36* and *ceh-37* in the lineage leading to *C. remanei* is further confirmed as only single orthologs of *ceh-36* and *ceh-37* were found in the additional species *C. brenneri* and *C. japonica*. The two *Cre-ceh-37* duplicates have significantly diverged at different rates since the duplication event (*Cre-ceh-37a*: $\omega = 0.1608$ (ω: d_N/d_S ratio), *Cre-ceh-37b*: $\omega = 0.0317$, $2\Delta l = 10.55$, $P < 0.01$). In *C. elegans*, *ceh-36* and *ceh-37* are next to each other on chromosome X, the chromosome with the most conserved synteny between *C. elegans* and *C. briggsae* (Hillier et al. 2007), suggesting that *Cre-ceh-36* and *Cre-ceh-37* paralogs may have resulted from a single duplication event of a fragment of the X chromosome. Unfortunately, the rate of synonymous changes between *Cre-ceh-36* ($d_S = 1.53$) and *Cre-ceh-37* ($d_S = 1.2$) paralogs, although similar, is too large to be a good indicator of the evolutionary age of the duplicate pairs. However, it is noteworthy that lengthy duplications are rare in *C. elegans* and that most duplications span less than 2 kb (Katju and Lynch 2003). Blast searches with the two immediate neighbors of *ceh-36* and *ceh-37* identified a single ortholog in *C. remanei*, indicating the duplication of a short fragment of the X chromosome or independent duplications of *Cre-ceh-36* and *Cre-ceh-37*.

By contrast, the two *C. remanei mec-3* sequences are likely to be alleles of the same locus. First, the two contigs containing *mec-3* are highly conserved (95.6% identical). Second, nucleotide diversity estimates across the coding sequence of the two *C. remanei mec-3* genomic sequences ($\pi = 13.54 \times 10^{-3}$) and within the Ohio population ($\pi = 19.15 \times 10^{-3}$) are similar. Finally, there is no phylogenetic differentiation between the two *mec-3* genomic sequences and *mec-3* alleles sampled from the Ohio population (data not shown). The presence of alleles in the genome assembly of *C. remanei* indicates incomplete inbreeding of the strain EM464 prior to genome sequencing.

Heterozygosity appears to be much higher in the genome assembly of *C. brenneri*, another dioecious species. Two highly similar sequences ($0.03 \leq d_S \leq 0.18$) could be found on different contigs for 31% (15/48) of the TFs in table 1. Although the low d_S values could indicate the presence of young duplicates, as found in *C. elegans* (Lynch and Conery 2000), it is rather likely that these sequences are distinct alleles. Intron boundaries are conserved for each pair of sequences indicating that they are not the result of duplication by retrotransposition. Moreover, no duplication boundary could be detected as the smaller of the two contigs aligned along its entire length to the larger one. The length of the smaller contig ranges from 1.6 to 25.4 kb, and although there are some large insertions/deletions, the contigs are highly similar, between 80% and 97% identical. Heterozygosity in the genome assemblies of *C. remanei* and *C. brenneri* has also recently been shown using a different set of genes (Barrière et al. 2009). However, clear duplicates of *cnd-1* were found in *C. brenneri*. A young duplicate, with high sequence similarity but carrying a premature stop codon, is located 2.7 kb downstream of *Cbn-cnd-1*, with a duplication span of 3.8 kb. An older duplicate ($d_S = 1.48$), which matches *C. elegans cnd-1* in reciprocal best Blast search, is located on a distinct contig with no flanking or intronic sequence similarity to *Cbn-cnd-1*.

TFs Controlling Chemosensory Neuron Differentiation Are More Divergent Than Structural Chemosensory Genes

Somewhat unexpectedly, we previously found that the TFs *odr-7* and *ceh-36*, respectively, required for the differentiation of the AWA and AWC chemosensory neurons, evolve faster than the structural component of olfactory pathways expressed in these two neurons (Jovelin et al. 2009). Is this higher nucleotide rate divergence specific to these two genes or is it a general feature of developmental regulatory genes controlling chemosensory neuron differentiation? To address this issue, I compared the rate of divergence between 54 structural chemosensory genes expressed in 11 chemosensory neurons and 15 TFs required for the differentiation of chemosensory neurons (table 1). Although the rate of synonymous substitution is not different between the two gene sets (d_S: Wilcoxon two-sample $P = 0.95$; $d_{S'}$: Wilcoxon two-sample $P = 0.99$), the TFs are on average 1.6 times more divergent than the structural chemosensory genes (rate of nonsynonymous substitutions, d_N: Wilcoxon two-sample $P = 0.005$; d_N / d_S: Wilcoxon two-sample $P < 0.0001$; $d_N/d_{S'}$: Wilcoxon two-sample $P < 0.0001$) (fig. 1). The higher rate divergence among the TFs is further supported by a LRT showing rate heterogeneity among the structural and developmental regulatory
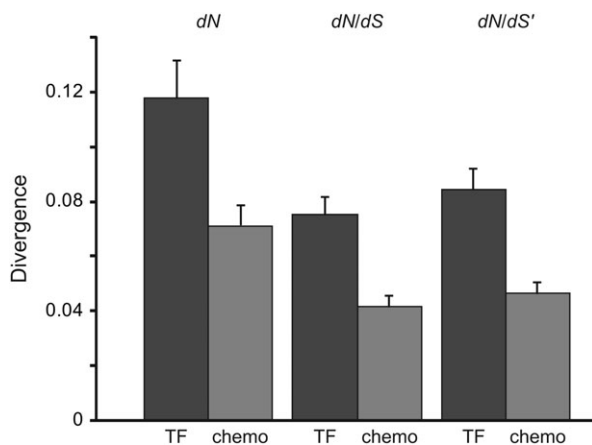
Fig. 1.—TFs controlling chemosensory neuron differentiation are more divergent than structural chemosensory genes. Comparison of species divergence estimated from *Caenorhabditis briggsae* and *Caenorhabditis remanei* between 54 structural chemosensory genes from 11 chemosensory neurons with 15 developmental regulatory genes controlling at least the differentiation of chemosensory neurons. Means are represented with one standard error.

genes ($\omega_{TF} = 0.0871$, $\omega_{structural} = 0.0525$, $2\varDelta l = 474.07$, $P < 0.001$).

It is reasonable to expect that selection pressure is higher against radical amino acid changes than against conservative changes because of a greater difference in amino acid chemical characteristics. Because it is also assumed that radical amino acid changes improve protein function more frequently, the ratio of radical to conservative amino acid changes (d_R/d_C) has been used to infer divergent selection (Hughes et al. 1990). However, if the intensity of purifying selection differs between radical and conservative amino acid replacements (Smith 2003), a ratio $d_R/d_C > 1$ may not necessarily indicate positive selection (Dagan et al. 2002; Hanada et al. 2006; Suzuki 2007). Nevertheless, d_R and d_C can still be used to examine what type of nonsynonymous substitutions are predominant given the higher rate of divergence among TFs required for chemosensory neuron differentiation. Five amino acid chemical classifications (charge; polarity; polarity and volume; charge and aromatic; and charge and polarity) (Zhang 2000; Hanada et al. 2006) were used to compare the ratios of radical amino acid changes with synonymous substitutions (d_R/d_S) and

conservative amino acid changes to synonymous substitutions (d_C/d_S) between the TFs and the chemosensory genes. For all five classifications, both d_R/d_S and d_C/d_S are greater for the TFs than for the structural chemosensory genes (table 2). Altogether these results indicate stronger purifying selection on the chemosensory genes and/or relaxed selection on the TFs required for chemosensory neuron differentiation.

Pattern of Polymorphism among Neuronal Transcription Factors

Polymorphism levels (π, Nei 1987) at nonsynonymous and synonymous sites are higher for the TFs *odr-7* and *ceh-36* (mean values: $\pi_a = 4.17 \times 10^{-3}$, $\pi_s = 55 \times 10^{-3}$) than for other loci ($n = 13$) in *C. remanei* (mean values: $\pi_a = 3.16 \times 10^{-3}$, $\pi_s = 44.87 \times 10^{-3}$) (Graustein et al. 2002; Jovelin et al. 2003, 2009; Haag and Ackerman 2005; Cutter et al. 2006). In order to investigate the level of nucleotide diversity among TFs, I sampled polymorphism information from 16 strains of *C. remanei* derived from one population, along nearly the full length of the coding sequence of 28 additional TF loci required for the differentiation of various classes of neurons (table 3). Among 31.36 kb of coding sequence from the subset of 31 genes from table 1, there are 990 polymorphic changes with only 24 sites segregating as three variants. Three loci, *Cre-ceh-14*, *Cre-egl-5*, and *Cre-mec-3* show length variants with strains PB263 and PB271 having a 51-bp deletion corresponding to exon 2 in *Cre-mec-3*. In all cases, these length variants show deletion adjacent to splice sites, suggesting that an alternative spliced cDNA was amplified and sequenced. Three other loci, *Cre-che-1*, *Cre-ceh-23*, and *Cre-die-1* exhibit size variation of a couple of codons.

Although polymorphism for some loci is somewhat skewed toward rare variants (table 3), as shown by negative values of Tajima's D (Tajima 1989), the mean value of Tajima's D at synonymous sites is 0.2364. For only one locus, *Cre-aha-1*, Tajima's D shows significant deviation from the neutral expectation in the direction of an excess of heterozygosity, usually interpreted as a signature of balancing selection. The proportion of rare variants is low in comparison with the pattern observed in *C. elegans* (Jovelin et al. 2009), 32% of polymorphic changes segregate as singleton polymorphism. Nevertheless, there is a highly significant

Table 2
TFs Controlling Chemosensory Neuron Differentiation Exhibit Higher Levels of Radical Nonsynonymous/Synonymous (d_R/d_S) and Conservative Nonsynonymous/Synonymous (d_C/d_S) Rate Ratios Than Structural Chemosensory Genes

Chemical Classification	d_R/d_S		d_C/d_S	
	TF	Structural	TF	Structural
Charge	0.0471 ± 0.006	0.0325 ± 0.003*	0.1084 ± 0.008	0.0586 ± 0.005***
Polarity	0.0793 ± 0.007	0.0361 ± 0.004***	0.0849 ± 0.007	0.0554 ± 0.004***
Volume & polarity	0.0587 ± 0.005	0.0360 ± 0.003***	0.1347 ± 0.011	0.0748 ± 0.006***
Charge &aromatic	0.0498 ± 0.005	0.0369 ± 0.003*	0.1209 ± 0.010	0.0619 ± 0.005***
Charge & polarity	0.0635 ± 0.006	0.0359 ± 0.003***	0.1193 ± 0.009	0.0677 ± 0.005***

NOTE.—Mean $\pm$ 1 standard error. Significance is obtained from a nonparametric Wilcoxon two-sample test: *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$. The list of TFs used in this analysis is reported in supplementary table 3, Supplementary Material online.

The radical and conservative amino acid changes are based on five amino acid chemical classifications. *Caenorhabditis remanei* and *Caenorhabditis briggsae* were used to compute divergence.

Table 3
Nucleotide Diversity in Neuronal TFs in *Caenorhabditis remanei*

Locus	LG[a]	N[b]	NS[c]	% Seq[d]	P[e]	A[f]	S[g]	π (10^{-3})[h]	π_a (10^{-3})[i]	π_s (10^{-3})[j]	Tajima's D
Cre-aha-1	I	16	1,195	88	12	3	9	3.84	0.85	13.65	2.0263*
Cre-alr-1	X	16	948	87	17	1	16	4.98	0.17	20.64	−0.4026
Cre-ceh-2	I	16	546	88	6	1	5	2.64	0.55	9.97	−0.3301
Cre-ceh-12	I	15	464	85	14	5	9	7.47	3.19	21.27	−0.1559
Cre-ceh-14	X	16	901	86	21	3	18	5.47	0.93	21.04	−0.6519
Cre-ceh-17	I	16	626	90	15	1	14	8.32	0.69	32.23	0.2453
Cre-ceh-23	III	16	830	91	73	5	68	27.76	3.47	99.7	0.0334
Cre-ceh-36a[k]	X	12	792	100	68	13	50	25.1	7.18	80.56	−0.6289
Cre-ceh-36b[k]	X	12	795	100	25	4	21	8.6	1.97	29.08	−0.7836
Cre-ceh-37a	X	16	684	87	40	14	26	19.42	7.25	61	1.2923
Cre-ceh-37b	X	11	736	87	34	2	33	22.23	1.88	88.43	1.5704
Cre-che-1	I	12	758	87	39	9	32	19.73	4.92	66.41	0.5665
Cre-cnd-1	III	16	497	87	13	0	14	6.59	0	25.96	−0.9526
Cre-cog-1	II	16	691	88	34	2	33	16.16	0.53	70.41	−0.0854
Cre-die-1	II	16	1,848	94	30	3	28	4.5	0.49	18.64	−0.2018
Cre-egl-5	III	16	572	86	16	0	17	10.97	0	46.94	1.7779
Cre-egl-43	II	15	1,563	92	69	5	64	15.36	1.51	61.08	0.4787
Cre-egl-46	V	16	782	92	29	4	28	15.68	2.47	62.21	1.2722
Cre-ham-2	X	16	1,022	87	16	1	15	4.79	0.41	20.59	−0.0261
Cre-lim-4	X	16	856	83	16	2	14	6.37	0.93	26.05	0.4122
Cre-lim-6	X	8	789	91	19	2	17	6.84	0.81	28.73	−1.068
Cre-lin-11	I	16	1,148	92	16	2	14	4.89	0.87	18.59	1.4177
Cre-mec-3	IV	16	822	86	49	3	47	19.15	2.09	81.26	0.2256
Cre-odr-7[k]	X	12	1,380	100	76	17	59	15.88	3.86	58.81	−0.4576
Cre-pag-3	X	14	884	88	25	1	25	9.92	0.38	43.19	−0.1922
Cre-unc-3	X	16	1,429	97	52	3	51	11.8	0.97	46.23	0.1566
Cre-unc-4	II	16	669	90	56	6	51	32.13	3.94	127.11	1.2304
Cre-unc-30	IV	16	880	90	19	5	14	6.13	1.85	19.99	0.2655
Cre-unc-130	II	16	859	86	22	4	18	6.83	1.06	27.35	0.0059
Cre-vab-7	III	16	666	89	4	1	3	2.67	0.91	8.11	1.0855
Cre-zag-1	IV	13	1,735	96	53	7	48	8.63	1.33	33.79	−0.7972

NOTE.—*$P < 0.05$.

[a] LG: linkage group in *Caenorhabditis elegans*.
[b] N: number of strains.
[c] NS: number of sites analyzed.
[d] % Seq: percent of the *Caenorhabditis remanei* gene sequenced.
[e] P: number of polymorphic sites.
[f] A: number of amino acid replacement changes.
[g] S: number of synonymous changes.
[h] π: total nucleotide diversity.
[i] π_a: nucleotide diversity at nonsynonymous sites.
[j] π_s: nucleotide diversity at synonymous sites.
[k] From Jovelin et al. (2009).

difference in the ratio of nonsynonymous to synonymous changes (A/S) between rare (i.e., singleton variants) and common polymorphism (A/S_{rare} = 0.2083, A/S_{common} = 0.1239; χ^2 = 7.365, $P < 0.01$), suggesting that a fraction of nonsynonymous changes with low frequency may be slightly deleterious (Fay et al. 2001).

Nucleotide diversity at nonsynonymous sites ranges from 0 to 7.27×10^{-3} and with an average of 1.85×10^{-3}, whereas nucleotide diversity at synonymous sites ranges from 8.11×10^{-3} to 127.11×10^{-3} and averages at 44.16×10^{-3} (table 3). No correlation is observed between nucleotide variability and codon usage bias as measured by the CAI (Sharp and Li 1987) ($\pi_a \times$ CAI: Spearman's ρ = 0.044, P = 0.815; $\pi_s \times$ CAI: Spearman's ρ = 0.168, P = 0.367). Although a better sampling will be necessary to correctly assess the level of nucleotide diversity in *C. brenneri*, first estimates can be obtained using the loci for which two alleles are present in the genome assembly. Among 18.36 kb of coding sequence from 15 TF loci, nucleotide diversity at nonsynonymous sites in *C. brenneri*

ranges from 0 to 12.89×10^{-3} and averages at 4.3×10^{-3}. The average nucleotide diversity at synonymous sites is 113.97×10^{-3}, with ranges from 40.8×10^{-3} to 191.37×10^{-3}. Among a subset of nine genes for which alleles are available both in *C. remanei* and *C. brenneri*, nucleotide diversity in *C. remanei* (measured from two alleles chosen at random for each gene) is lower than in *C. brenneri* (π: Wilcoxon two-sample P = 0.01). Nevertheless, the higher level of polymorphism is solely due to diversity at synonymous sites (π_a: Wilcoxon two-sample P = 0.506; π_s: Wilcoxon two-sample P = 0.004) suggesting that level of nucleotide diversity in *C. brenneri* may be relatively high.

TFs Exhibit Higher Nucleotide Diversity Than Structural Genes

In order to investigate whether neuronal TFs exhibit high or low levels of nucleotide diversity, I used a published data set of polymorphism in *C. remanei* for 37 structural loci (Cutter 2008). This data set of structural loci is useful

Table 4
TFs Controlling Neuron Identity Exhibit More Nonsynonymous Polymorphism Than Structural Genes

Loci	N^a	A^b	S^c	A/S
TF-all	31	129	861	0.1498**
TF-X	12	63	345	0.1826***
TF-chemosensory	14	78	412	0.1893***
TF-autosomes	19	66	516	0.1279
TF-not chemosensory	17	51	449	0.1136
Structural-X[d]	37	76	781	0.0973

NOTE.—Significance is obtained from a χ^2 test: **$P < 0.01$, ***$P < 0.001$.

The ratio of nonsynonymous to synonymous polymorphism in *Caenorhabditis remanei* is compared between structural X-linked loci to each set of TFs. Chromosomal linkage is based on *Caenorhabditis elegans* orthologs and is assumed to be conserved between *C. elegans* and *C. remanei*. Chemosensory genes refer to the TFs required for the differentiation of chemosensory neurons (table 1). The list of TFs in each gene set is reported in supplementary table 3, Supplementary Material online.

[a] N: number of loci.
[b] A: number of amino acid replacement changes.
[c] S: number of synonymous changes.
[d] From Cutter (2008).

for comparison because these genes have diverse functions (supplementary table 1, Supplementary Material online), and most importantly because polymorphism was obtained in the coding region using a similar number of strains that have the same geographical origin. Although the level of interspecific divergence is higher among the neuronal TFs (table 1) than among the structural genes (d_N: Wilcoxon two-sample $P < 0.0001$; d_N/d_S: Wilcoxon two-sample $P = 0.006$; LRT: $\omega_{TF} = 0.0759$, $\omega_{structural} = 0.0506$, $2\Delta l = 2066.7$, $P < 0.001$), the overall levels of nucleotide diversity are not significantly different (π: Wilcoxon two-sample $P = 0.375$).

Even so, it is more informative to compare changes affecting the protein sequence and segregating within the population. The ratio of nonsynonymous to synonymous polymorphism (A/S) is significantly higher for the neuronal TFs than for the structural genes (table 4). Nevertheless, assuming that chromosomal location is conserved between *C. elegans* and *C. remanei*, in particular for genes on the X chromosome (Hillier et al. 2007), the set of structural loci is made of genes that are all putatively linked to the X chromosome. Because of a smaller population size increasing the fixation of slightly deleterious mutations through genetic drift and because of genetic hitchhiking eliminating neutral variation, the level of nucleotide diversity for X-linked loci may be lower than for autosomal loci. The putatively X-linked TFs have nonetheless a significantly higher A/S ratio than the structural genes (table 4), indicating that chromosomal location is not responsible for the difference in the pattern of polymorphism. In addition, applying a 4/3 correction to the polymorphism estimates of the X-linked loci, to compensate for differences in effective population size, leads to the same results for each comparison between the sets of TFs and structural genes from table 4 (data not shown). However, a closer look at the function of the TFs potentially linked to the X chromosome reveals that most are involved in the differentiation of chemosensory neurons ($\chi^2 = 7.039$, $P < 0.01$). The A/S ratio for the neuronal TFs excluding those required for chemosensory neuron differentiation is higher than for structural genes but the difference is not significant.

Table 5
TFs Controlling Chemosensory Neuron Identity Exhibit a High Level of Nonsynonymous Polymorphism

Loci	N^a	A^b	S^c	A/S
TF-motorneuron	7	19	161	0.1180*
TF-multiple	12	34	271	0.1255**
TF-chemosensory	8	64	282	0.2269

NOTE.—Significance is obtained from a χ^2 test: *$P < 0.05$, **$P < 0.01$.

The ratio of nonsynonymous to synonymous polymorphism in *Caenorhabditis remanei* is compared between TFs controlling chemosensory neuron differentiation to TFs affecting the motorneuron class and to those affecting several classes of neurons including, among other classes, motorneurons and/or chemosensory neurons. The list of TFs in each gene set is reported in supplementary table 3, Supplementary Material online.

[a] N: number of loci.
[b] A: number of amino acid replacement changes.
[c] S: number of synonymous changes.

By contrast, the TFs determining chemosensory neuron identity exhibit significantly higher nonsynonymous polymorphic changes than structural loci (table 4). Analyses using the nucleotide diversity indexes lead to the same results (supplementary table 2, Supplementary Material online). These results show that TFs, in particular those involved in the development of chemosensory neurons, can exhibit high levels of within-population variation.

Differences in Selective Pressures among Neuronal Transcription Factors

In order to further investigate the differences in evolutionary rates among neuronal TFs, I compared the level of between- and within-species divergence among genes that are specifically required for chemosensory neuron and motorneuron differentiation, respectively, and those affecting several functional classes (table 1). A LRT indicates that a model with different d_N/d_S for each type of TF better fits the data ($\omega_{chemo} = 0.0952$, $\omega_{motor} = 0.0877$, $\omega_{multiple} = 0.0551$; $2\Delta l = 423.98$, $P < 0.001$). The TFs regulating the differentiation of several classes of neurons are under stronger selective constraints, presumably because of their broader role in neuronal development. Remarkably, the TFs that specifically regulate chemosensory neuron differentiation are the most divergent. Similarly the ratio of nonsynonymous to synonymous polymorphism A/S is significantly higher for the TFs involved in chemosensory neuron differentiation than those required for the differentiation of motorneurons and multiple neuron classes (table 5), although the difference between the two latter gene classes is not significantly different ($\chi^2 = 0.041$, $P = 0.84$). These results show that neuronal TFs are under different selective constraints and that regulatory genes controlling chemosensory neuron development accumulate higher levels of nucleotide diversity both within and between species.

Discussion

Molecular evolutionary analyses of various TF family members in plants and animals have detected rapid divergence and/or instances of adaptive selection (Sutton and Wilkinson 1997; Barrier et al. 2001; Fares et al. 2003;

Jia et al. 2003; Moore et al. 2005; Hernandez-Hernandez et al. 2007; Lynch et al. 2008). Nematodes in particular seem to show elevated rates of divergence for TF genes (Castillo-Davis et al. 2004; Cutter and Ward 2005; Haerty et al. 2008; Jovelin et al. 2009). However, these studies included a small number of TFs or were based on bioinformatic analyses between orthologous or paralogous genes. Few attempts have been made to investigate global population-level constraints on TFs (Bustamante et al. 2005), and thus the extent of these constraints remained unclear. In order to address this issue, I examined nucleotide variation within one population of *C. remanei* among 31 TFs required for the development of various neuronal classes in *C. elegans*.

TFs Exhibit High Nucleotide Diversity

Remarkably, the neuronal TFs exhibit a significantly higher nonsynonymous to synonymous polymorphism ratio (A/S) than structural genes (table 4 and supplementary table 2, Supplementary Material online). However, the structural loci used in this analysis are from a data set used to quantify nucleotide diversity along the X chromosome (Cutter 2008), which could lead to several potential complications. On the one hand, theory predicts that background selection should result in higher neutral polymorphism on the X chromosome (Begun and Whitley 2000), whereas smaller effective population size and genetic hitchhiking are expected to result in lower polymorphism (Charlesworth et al. 1987). On the other hand, the X chromosome is expected to diverge more between species than autosomes (see Vicoso and Charlesworth 2006). A higher divergence of the X chromosome relative to autosomes is apparent in *Drosophila melanogaster*, *Drosophila simulans*, and *Drosophila yakuba* (Begun et al. 2007). However, a genomewide survey of polymorphism within *D. simulans* revealed that heterozygosity is not significantly lower on the X (after correction, Begun et al. 2007). To the extent that chromosomal linkage is conserved between *Caenorhabditis* species (Hillier et al. 2007), the higher A/S ratio of TFs could be due to a reduction of nucleotide diversity among the structural genes. However, because the A/S ratio of putative X-linked TFs is still significantly higher than that of structural genes (table 4) and because correcting for the difference in population size between chromosomes leads to the same results, the difference of polymorphism level between the two gene classes is unlikely to be the result of linkage. Moreover, the TFs show more between-species divergence than the putatively X-linked structural loci. Therefore, the higher rate divergence and higher A/S ratio of TFs are more likely to reflect differential selective constraints acting on these two gene classes.

Nucleotide Divergence and Function in Neuron Differentiation

Does this pattern of nucleotide rate divergence reflect global high rates of divergence among TFs? Genomic analyses among *Caenorhabditis* species show that the core set of TFs present in *C. elegans*, *C. briggsae*, and *C. remanei* has a significantly higher nonsynonymous substitution rate than non-TF genes present in the three species (Haerty et al. 2008). Nevertheless, the answer to this question is not straightforward. Significantly more *C. elegans* TF orthologs are detected in *C. briggsae* than orthologs for other *C. elegans* genes (Haerty et al. 2008), either because these genes are species-specific or because some non-TF genes are so divergent that orthology cannot be assigned. In Drosophila, genes in the Gene Ontology categories "DNA binding" and "TF activity" (i.e., potential TFs) show a higher divergence rate ratio than the median rate of divergence for all genes with Gene Ontology terms. However, this is due to a lower synonymous substitution rate rather than accelerated amino acid changes as there are no significant differences in nonsynonymous rates (Clark et al. 2007). Global patterns of divergence between TFs and non-TFs may therefore be difficult to establish.

Insights into the cellular context in which proteins operate, as in the present study, is helpful when testing assumptions regarding TF sequence evolution. This is particularly true when comparing evolutionary rates with structural proteins because any observed difference between the two gene classes could simply be due to differences in pleiotropic constraints (Duret and Mouchiroud 2000). Fully addressing the relationship between pleiotropic constraints and nucleotide variation may require standardized measures for the number of cells expressing TFs and structural genes. Nevertheless, it is noteworthy that although most TFs analyzed here have known effects restricted to nervous system differentiation (table 1), 35% (17/48) (see Materials and Methods) also play a role in the differentiation of nonneuronal tissues. Moreover, 67% (32/48) of the TFs are expressed in more than one tissue type versus 46% (35/76) for the structural genes (www.wormbase.org). Therefore, there does not seem to be a simple relationship between expression breadth and nucleotide variation for the genes analyzed here. Thus, the patterns of divergence and polymorphism reported here strongly argue that TFs can evolve rapidly and are at odds with the notion that TFs are highly constrained (Carroll 2005, 2008).

A classical means of assessing whether a gene has undergone adaptive evolution is to compare the ratio of nonsynonymous to synonymous substitutions among lineages with that of polymorphic changes within lineages (MK test, McDonald and Kreitman 1991). Although the MK test is robust when slightly deleterious mutations are segregating, in particular when there is selection for codon usage (Eyre-Walker 2002), it requires that the observed number of changes between two species accurately reflects the number of changes that happened since their divergence. The saturation at synonymous sites in *Caenorhabditis* prevents the use of the MK test because it may inflate the divergence ratio, leading to spurious significant results. Thus, the high level of amino acid divergence observed among the neuronal TFs could be the result of positive selection and/or relaxed section. Nevertheless, the high divergence of TF protein sequence is still relevant because the accumulation of neutral or slightly neutral nonsynonymous substitutions may play a creative role in evolution by paving the ground for subsequent beneficial mutations that would be

deleterious in a different sequence context (Ortlund et al. 2007; Wagner 2008).

It is possible that the high nucleotide divergence observed in these TFs is due to their function in neuron differentiation. For instance, loss of function mutations in 70% of the genes in table 1 result in neurons that are correctly generated but not fully differentiated. One way of reducing the negative pleiotropy of TFs is tissue specificity (or in this case, cell specificity), through the cooperative action of TFs and cofactors (Lynch and Wagner 2008; Wagner and Lynch 2008). This may be a particularly important mechanism for relaxing constraints on neuronal TFs because individual neurons in *C. elegans* are specified by the combinatorial role of TFs, as exemplified by the joint action of *ceh-10*, *ttx-3*, and *ceh-23* in the differentiation of the AIY neurons (Altun-Gultekin et al. 2001) and the joint action of *ceh-36*, *che-1*, *lzy-2*, *fozi-1*, *die-1*, *lim-6*, and *cog-1* in the differentiation of the ASE neurons (Hobert et al. 1999; Chang et al. 2003; Lanjuin et al. 2003; Uchida et al. 2003; Johnston and Hobert 2005; Johnston et al. 2006). Another consequence of this combinatorial code is that regulation of target genes can be cell specific. The LIM homeobox *lim-6* regulates the expression of *unc-25* in the neurons RIS, AVL, and DVD but has no effect on *unc-25* expression in the right and left RME neurons (Hobert et al. 1999). Similarly, the homeobox *unc-30* regulates the expression of *unc-25* and *unc-47* in only 19 of the 26 GABAergic neurons expressing these two genes (Eastman et al. 1999).

High Nucleotide Diversity in Transcription Factors Controlling Chemosensory Neuron Differentiation

Another means of relaxing pleiotropic constraints is gene duplication (Lynch and Wagner 2008; Wagner and Lynch 2008), which can increase evolvability and lead to the evolution of new functions, to the partitioning of ancestral functions and even to speciation (Force et al. 1999; Roth et al. 2007). Two neuronal TFs, the homeodomain genes *ceh-36* and *ceh-37*, required for the differentiation of chemosensory neurons (table 1), were specifically duplicated in the lineage leading to *C. remanei*. Interestingly, TFs involved in chemosensory neuron differentiation evolve at high rates. They are more divergent than structural chemosensory genes (fig. 1 and table 2) and have a greater *A/S* ratio than structural genes (table 4). TFs that are specifically required for chemosensory neuron differentiation also show higher between- and within-species divergence than TFs required for the development of motorneurons. They are also more polymorphic and divergent than TFs affecting the development of several functional neuron classes (table 5). Evolving different chemosensory capabilities allows specialization to new resources and can promote speciation in sympatric areas (Linn et al. 2003). For example, *Drosophila sechellia*'s attraction to the fruits of its unique host *Morinda citrifolia* results partly from an increased sensitivity of the olfactory receptor Or22a to methyl hexanoate relative to ethyl hexanoate. Importantly, fruit attraction is also due to a 2.5- to 3-fold increase in the number of ab3 neurons, the neurons expressing Or22a, concomitant with a reduction in the number of ab1 neurons (60–80%) and ab2 neurons (93–100%) compared with *D. melanogaster*

(Dekker et al. 2006). Although the genetic mechanisms for this shift are unknown, this example clearly shows that changes in neuron identity can lead to chemosensory diversity and behavioral specialization.

Conclusion

The high within- and between-species nucleotide variation reported here suggests that the divergence of developmental regulatory genes may play a greater role in phenotypic change and, more broadly, in the evolution of gene regulatory networks than has been previously thought. The allelic variation observed at these TF loci suggests the presence of the raw material necessary for the evolution of neuron diversity. It will be interesting to test the functional divergence of neuronal TF orthologs in *Caenorhabditis*.

Supplementary Material

Acknowledgments

Literature Cited

Ahringer J. 1996. Posterior patterning by the *Caenorhabditis elegans* even-skipped homolog *vab-7*. Genes Dev. 10:1120–1130.

Alper S, Kenyon C. 2001. REF-1, a protein with two bHLH domains, alters the pattern of cell fusion in *C. elegans* by regulating Hox protein activity. Development. 128:1793–1804.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Altun-Gultekin Z, Andachi Y, Tsalik EL, Pilgrim D, Kohara Y, Hobert O. 2001. A regulatory cascade of three homeobox genes, *ceh-10*, *ttx-3* and *ceh-23*, controls cell fate specification of a defined interneuron class in *C. elegans*. Development. 128:1951–1969.

Amin NM, Hu K, Pruyne D, Terzic D, Bretscher A, Liu J. 2007. A Zn-finger/FH2-domain containing protein, FOZI-1, acts redundantly with CeMyoD to specify striated body wall muscle fates in the *Caenorhabditis elegans* postembryonic mesoderm. Development. 134:19–29.

Aspock G, Ruvkun G, Burglin TR. 2003. The *Caenorhabditis elegans ems* class homeobox gene *ceh-2* is required for M3 pharynx motoneuron function. Development. 130:3369–3378.

Bando T, Ikeda T, Kagawa H. 2005. The homeoproteins MAB-18 and CEH-14 insulate the dauer collagen gene *col-43* from activation by the adjacent promoter of the Spermatheca gene *sth-1* in *Caenorhabditis elegans*. J Mol Biol. 348:101–112.

Baran R, Aronoff R, Garriga G. 1999. The *C. elegans* homeodomain gene *unc-42* regulates chemosensory and glutamate receptor expression. Development. 126:2241–2251.

Bargmann CI. 2006. Chemosensation in *C. elegans*. In: WormBook, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.123.1, http://www.wormbook.org.

Barrier M, Robichaux RH, Purugganan MD. 2001. Accelerated regulatory gene evolution in an adaptive radiation. Proc Natl Acad Sci USA. 98:10208–10213.

Barrière 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.

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

Baum PD, Guenther C, Frank CA, Pham BV, Garriga G. 1999. The *Caenorhabditis elegans* gene *ham-2* links Hox patterning to migration of the HSN motor neuron. Genes Dev. 13:472–483.

Begun DJ, Holloway AK, Stevens K, et al. (13 co-authors). 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 5:e310.

Begun DJ, Whitley P. 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc Natl Acad Sci USA. 97:5960–5965.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2008. GenBank. Nucleic Acids Res. 36:D25–D30.

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

Bustamante CD, Fledel-Alon A, Williamson S, et al. (14 co-authors). 2005. Natural selection on protein-coding genes in the human genome. Nature. 437:1153–1157.

Cameron S, Clark SG, McDermott JB, Aamodt E, Horvitz HR. 2002. PAG-3, a Zn-finger transcription factor, determines neuroblast fate in *C. elegans*. Development. 129:1763–1774.

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

Carroll SB. 2005. Endless forms most beautiful: the new science of evo–devo and the making of the animal kingdom. New York, London: W.W. Norton and Company.

Carroll SB. 2008. Evo–devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 134:25–36.

Carroll SB, Grenier JK, Weatherbee SD. 2005. From DNA to diversity. Molecular genetics and the evolution of animal design. Malden (MA), Oxford, Victoria (Australia): Blackwell Publishing.

Cassata G, Kagoshima H, Andachi Y, Kohara Y, Durrenberger MB, Hall DH, Burglin TR. 2000. The LIM homeobox gene *ceh-14* confers thermosensory function to the AFD neurons in *Caenorhabditis elegans*. Neuron. 25:587–597.

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.

Chalfie M, Horvitz HR, Sulston JE. 1981. Mutations that lead to reiterations in the cell lineages of *C. elegans*. Cell. 24:59–69.

Chalfie M, Sulston J. 1981. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. Dev Biol. 82:358–370.

Chalfie M, Au M. 1989. Genetic control of differentiation of the *Caenorhabditis elegans* touch receptor neurons. Science. 243:1027–1033.

Chang S, Johnston RJ Jr, Frokjaer-Jensen C, Lockery S, Hobert O. 2004. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature. 430:785–789.

Chang S, Johnston RJ Jr, Hobert O. 2003. A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. Genes Dev. 17:2123–2137.

Charlesworth B, Coyne JA, Barton N. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 130:113–146.

Chisholm A. 1991. Control of cell fate in the tail region of *C. elegans* by the gene *egl-5*. Development. 111:921–932.

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.

Clark SG, Chisholm AD, Horvitz HR. 1993. Control of cell fates in the central body region of *C. elegans* by the homeobox gene *lin-39*. Cell. 74:43–55.

Clark SG, Chiu C. 2003. *C. elegans* ZAG-1, a Zn-finger-homeodomain protein, regulates axonal development and neuronal differentiation. Development. 130:3781–3794.

Cutter AD. 2008. Multilocus patterns of polymorphism and selection across the X chromosome of *Caenorhabditis remanei*. Genetics. 178:1661–1672.

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, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. Mol Biol Evol. 22:178–188.

Dagan T, Talmor Y, Graur D. 2002. Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. Mol Biol Evol. 19:1022–1025.

Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. 2006. Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia*. Curr Biol. 16:101–109.

Desai C, Garriga G, McIntire SL, Horvitz HR. 1988. A genetic pathway for the development of the *Caenorhabditis elegans* HSN motor neurons. Nature. 336:638–646.

Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol Biol Evol. 17:68–74.

Eastman C, Horvitz HR, Jin Y. 1999. Coordinated transcriptional regulation of the *unc-25* glutamic acid decarboxylase and the *unc-47* GABA vesicular transporter by the *Caenorhabditis elegans* UNC-30 homeodomain protein. J Neurosci. 19:6225–6234.

Eisenmann DM, Kim SK. 2000. Protruding vulva mutants identify novel loci and Wnt signaling factors that function during *Caenorhabditis elegans* vulva development. Genetics. 156:1097–1116.

Esmaeili B, Ross JM, Neades C, Miller DM 3rd, Ahringer J. 2002. The *C. elegans even-skipped* homologue, *vab-7*, specifies DB motoneurone identity and axon trajectory. Development. 129:853–862.

Euling S, Ambros V. 1996. Heterochronic genes control cell cycle progress and developmental competence of *C. elegans* vulva precursor cells. Cell. 84:667–676.

Eyre-Walker A. 2002. Changing effective population size and the McDonald–Kreitman test. Genetics. 162:2017–2024.

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

Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. Genetics. 158:1227–1234.

Fernandes JS, Sternberg PW. 2007. The tailless ortholog *nhr-67* regulates patterning of gene expression and morphogenesis in the *C. elegans* vulva. PLoS Genet. 3:e69.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545.

Forrester WC, Perens E, Zallen JA, Garriga G. 1998. Identification of *Caenorhabditis elegans* genes required for neuronal differentiation and migration. Genetics. 148:151–165.

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

Grant K, Hanna-Rose W, Han M. 2000. *sem-4* promotes vulval cell-fate determination in *Caenorhabditis elegans* through regulation of *lin-39* Hox. Dev Biol. 224:496–506.

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.

Guder C, Philipp I, Lengfeld T, Watanabe H, Hobmayer B, Holstein TW. 2006. The Wnt code: cnidarians signal the way. Oncogene. 25:7450–7460.

Guerry F, Marti CO, Zhang Y, Moroni PS, Jaquiery E, Muller F. 2007. The Mi-2 nucleosome-remodeling protein LET-418 is targeted via LIN-1/ETS to the promoter of *lin-39*/Hox during vulval development in *C. elegans*. Dev Biol. 306:469–479.

Gupta BP, Wang M, Sternberg PW. 2003. The *C. elegans* LIM homeobox gene *lin-11* specifies multiple cell fates during vulval development. Development. 130:2589–2601.

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

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.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 41:95–98.

Hallam S, Singer E, Waring D, Jin Y. 2000. The *C. elegans* NeuroD homolog *cnd-1* functions in multiple aspects of motor neuron fate specification. Development. 127:4239–4252.

Hanada K, Gojobori T, Li WH. 2006. Radical amino acid change versus positive selection in the evolution of viral envelope proteins. Gene. 385:83–88.

Heid PJ, Raich WB, Smith R, Mohler WA, Simokat K, Gendreau SB, Rothman JH, Hardin J. 2001. The zinc finger protein DIE-1 is required for late events during epithelial cell rearrangement in *C. elegans*. Dev Biol. 236:165–180.

Hernandez-Hernandez T, Martinez-Castilla LP, Alvarez-Buylla ER. 2007. Functional diversification of B MADS-box homeotic regulators of flower development: adaptive evolution in protein–protein interaction domains after major gene duplication events. Mol Biol Evol. 24:465–481.

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.

Hirsh AE, Fraser HB, Wall DP. 2005. Adjusting for selection on synonymous sites in estimates of evolutionary distance. Mol Biol Evol. 22:174–177.

Hobert O. 2005. Specification of the nervous system. WormBook, ed. The *C. elegans* Research Community, Wormbook, doi/10.1895/wormbook.1.12.1, http://www.wormbook.org.

Hobert O. 2008. Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. Proc Natl Acad Sci USA. 105:20067–20071.

Hobert O, D'Alberti T, Liu Y, Ruvkun G. 1998. Control of neural development and function in a thermoregulatory network by the LIM homeobox gene *lin-11*. J Neurosci. 18:2084–2096.

Hobert O, Mori I, Yamashita Y, Honda H, Ohshima Y, Liu Y, Ruvkun G. 1997. Regulation of interneuron function in the *C. elegans* thermoregulatory pathway by the *ttx-3* LIM homeobox gene. Neuron. 19:345–357.

Hobert O, Tessmar K, Ruvkun G. 1999. The *Caenorhabditis elegans lim-6* LIM homeobox gene regulates neurite outgrowth and function of particular GABAergic neurons. Development. 126:1547–1562.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation.. Evolution. 61:995–1016.

Hsia CC, McGinnis W. 2003. Evolution of transcription factor function. Curr Opin Genet Dev. 13:199–206.

Huang X, Powell-Coffman JA, Jin Y. 2004. The AHR-1 aryl hydrocarbon receptor and its co-factor the AHA-1 aryl hydrocarbon receptor nuclear translocator specify GABAergic neuron cell fate in *C. elegans*. Development. 131:819–828.

Hughes AL, Ota T, Nei M. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol Biol Evol. 7:515–524.

Hutter H. 2003. Extracellular cues and pioneers act together to guide axons in the ventral cord of *C. elegans*. Development. 130:5307–5318.

Hwang BJ, Meruelo AD, Sternberg PW. 2007. *C. elegans* EVI1 proto-oncogene, EGL-43, is necessary for Notch-mediated cell fate specification and regulates cell invasion. Development. 134:669–679.

Inoue T, Wang M, Ririe TO, Fernandes JS, Sternberg PW. 2005. Transcriptional network underlying *Caenorhabditis elegans* vulval development. Proc Natl Acad Sci USA. 102:4972–4977.

Jia L, Clegg MT, Jiang T. 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.

Jin Y, Hoskins R, Horvitz HR. 1994. Control of type-D GABAergic neuron differentiation by *C. elegans* UNC-30 homeodomain protein. Nature. 372:780–783.

Jia Y, Xie G, Aamodt E. 1996. *pag-3*, a *Caenorhabditis elegans* gene involved in touch neuron gene expression and coordinated movement. Genetics. 142:141–147.

Johnston RJ Jr, Copeland JW, Fasnacht M, Etchberger JF, Liu J, Honig B, Hobert O. 2006. An unusual Zn-finger/FH2 domain protein controls a left/right asymmetric neuronal fate decision in *C. elegans*. Development. 133:3317–3328.

Johnston RJ Jr, Hobert O. 2005. A novel *C. elegans* zinc finger transcription factor, *lsy-2*, required for the cell type-specific expression of the *lsy-6* microRNA. Development. 132:5451–5460.

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. 2009. High nucleotide divergence in developmental regulatory genes contrasts with the structural elements of olfactory pathways in *Caenorhabditis*. Genetics. 181:1–11.

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

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–132.

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

Kenyon C. 1986. A gene involved in the development of the posterior body region of *C. elegans*. Cell. 46:477–487.

Kim K, Colosimo ME, Yeung H, Sengupta P. 2005. The UNC-3 Olf/EBF protein represses alternate neuronal programs to specify chemosensory neuron identity. Dev Biol. 286: 136–148.

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

Koga M, Ohshima Y. 2004. The *C. elegans ceh-36* gene encodes a putative homemodomain transcription factor involved in chemosensory functions of ASE and AWC neurons. J Mol Biol. 336:579–587.

Lanjuin A, Claggett J, Shibuya M, Hunter CP, Sengupta P. 2006. Regulation of neuronal lineage decisions by the HES-related bHLH protein REF-1. Dev Biol. 290:139–151.

Lanjuin A, VanHoven MK, Bargmann CI, Thompson JK, Sengupta P. 2003. Otx/otd homeobox genes specify distinct sensory neuron identities in *C. elegans*. Dev Cell. 5:621–633.

Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. Nature. 276:565–570.

Linn C Jr, Feder JL, Nojima S, Dambroski HR, Berlocher SH, Roelofs W. 2003. Fruit odor discrimination and sympatric host race formation in *Rhagoletis*. Proc Natl Acad Sci USA. 100:11490–11493.

Lints R, Emmons SW. 1999. Patterning of dopaminergic neurotransmitter identity among *Caenorhabditis elegans* ray sensory neurons by a TGFbeta family signaling pathway and a Hox gene. Development. 126:5819–5831.

Lints R, Jia L, Kim K, LiS C, Emmons W. 2004. Axial patterning of *C. elegans* male sensilla identities by selector genes. Dev Biol. 269:137–151.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. neurotransmitter identity among *Caenorhabditis elegans* ray sensory neurons by a TGFbeta family signaling pathway and a *Hox* gene. Development. 126:5819–5831.

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

Lynch VJ, Tanzer A, Wang Y, Leung FC, Gellersen B, Emera D, Wagner GP. 2008. Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. Proc Natl Acad Sci USA. 105:14928–14933.

Lynch VJ, Wagner GP. 2008. Resurrecting the role of transcription factor change in developmental evolution. Evolution. 62:2131–2154.

Matus DQ, Thomsen GH, Martindale MQ. 2007. FGF signaling in gastrulation and neural development in *Nematostella vectensis*, an anthozoan cnidarian. Dev Genes Evol. 217:137–148.

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

Melkman T, Sengupta P. 2005. Regulation of chemosensory and GABAergic motor neuron development by the *C. elegans* Aristaless/Arx homolog *alr-1*. Development. 132:1935–1949.

Miller DM, Shen MM, Shamu CE, Burglin TR, Ruvkun G, Dubois ML, Ghee M, Wilson L. 1992. *C. elegans unc-4* gene encodes a homeodomain protein that determines the pattern of synaptic input to specific motor neurons. Nature. 355:841–845.

Mitani S, Du H, Hall DH, Driscoll M, Chalfie M. 1993. Combinatorial control of touch receptor neuron expression in *Caenorhabditis elegans*. Development. 119:773–783.

Moore RC, Grant SR, Purugganan MD. 2005. Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. Mol Biol Evol. 22:91–103.

Moriyama EN, Powell JR. 1996. Intraspecific nuclear DNA variation in Drosophila. Mol Biol Evol. 13:261–277.

Much JW, Slade DJ, Klampert K, Garriga G, Wightman B. 2000. The *fax-1* nuclear hormone receptor regulates axon path finding and neurotransmitter expression. Development. 127:703–712.

Nash B, Colavita A, Zheng H, Roy PJ, Culotti JG. 2000. The forkhead transcription factor UNC-130 is required for the graded spatial expression of the UNC-129 TGF-beta guidance factor in *C. elegans*. Genes Dev. 14:2486–2500.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nichols SA, Dirks W, Pearse JS, King N. 2006. Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci USA. 103:12451–12456.

Nüsslein-Volhard C, Wieschaus E. 1980. Mutations affecting segment number and polarity in Drosophila. Nature. 287: 795–801.

Nüsslein-Volhard C, Wieschaus E, Kluding H. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. Rouxs Arch Dev Biol. 193:267–282.

Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. Science. 317:1544–1548.

Prasad BC, Ye B, Zackhary R, Schrader K, Seydoux G, Reed RR. 1998. *unc-3*, a gene required for axonal guidance in *Caenorhabditis elegans*, encodes a member of the O/E family of transcription factors. Development. 125:1561–1568.

Pujol N, Torregrossa P, Ewbank JJ, Brunet JF. 2000. The homeodomain protein CePHOX2/CEH-17 controls antero-posterior axonal growth in *C. elegans*. Development. 127:3361–3371.

Qin H, Powell-Coffman JA. 2004. The *Caenorhabditis elegans* aryl hydrocarbon receptor, AHR-1, regulates neuronal development. Dev Biol. 270:64–75.

Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJ. 2005. A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol. 6:R110.

Rimann I, Hajnal A. 2007. Regulation of anchor cell invasion and uterine cell fates by the *egl-43 Evi-1* proto-oncogene in *Caenorhabditis elegans*. Dev Biol. 308:187–195.

Rogers A, Antoshechkin I, Bieri T, et al. 2008. WormBase 2007. Nucleic Acids Res. 36:D612–D617.

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

Roth C, Rastogi S, Arvestad L, Dittmar K, Light S, Ekman D, Liberles DA. 2007. Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. J Exp Zool B Mol Dev Evol. 308:58–73.

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

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

Sarafi-Reinach TR, Melkman T, Hobert O, Sengupta P. 2001. The *lin-11* LIM homeobox gene specifies olfactory and chemosensory neuron fates in *C. elegans*. Development. 128: 3269–3281.

Sarafi-Reinach TR, Sengupta P. 2000. The forkhead domain gene *unc-130* generates chemosensory neuron diversity in *C. elegans*. Genes Dev. 14:2472–2485.

Satterlee JS, Sasakura H, Kuhara A, Berkeley M, Mori I, Sengupta P. 2001. Specification of thermosensory neuron fate in *C. elegans* requires *ttx-1*, a homolog of otd/Otx. Neuron. 31:943–956.

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

Shaham S, Bargmann CI. 2002. Control of neuronal subtype identity by the *C. elegans* ARID protein CFI-1. Genes Dev. 16:972–983.

Shan G, Walthall WW. 2008. Copulation in *C. elegans* males requires a nuclear hormone receptor. Dev Biol. 322:11–20.

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.

Smith NG. 2003. Are radical and conservative substitution rates useful statistics in molecular evolution? J Mol Evol. 57: 467–478.

Stein L, 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.

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

Sutton KA, Wilkinson MF. 1997. Rapid evolution of a homeodomain: evidence for positive selection. J Mol Evol. 45:579–588.

Suzuki Y. 2007. Inferring natural selection operating on conservative and radical substitution at single amino acid sites. Genes Genet Syst. 82:341–360.

Swofford DL. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sunderland (MA): Sinauer Associates.

Sze JY, Zhang S, Li G, Ruvkun J. 2002. The *C. elegans* POU-domain transcription factor UNC-86 regulates the *tph-1* tryptophan hydroxylase gene and neurite outgrowth in specific serotonergic neurons. Development. 129:3901–3911.

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

Technau US, Rudd S, Maxwell P, et al. (12 co-authors). 2005. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. Trends Genet. 21:633–639.

Toker AS, Teng Y, Ferreira HB, Emmons SW, Chalfie M. 2003. The *Caenorhabditis elegans* spalt-like gene *sem-4* restricts touch cell fate by repressing the selector Hox gene *egl-5* and the effector gene *mec-3*. Development. 130:3831–3840.

Tsalik EL, Niacaris T, Wenick AS, Pau K, Avery L, Hobert O. 2003. LIM homeobox gene-dependent expression of biogenic amine receptors in restricted regions of the *C. elegans* nervous system. Dev Biol. 263:81–102.

Uchida O, Nakano H, Koga M, Ohshima Y. 2003. The *C. elegans che-1* gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons. Development. 130:1215–1224.

Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet. 7:645–653.

Von Stetina SE, Fox RM, Watkins KL, Starich TA, Shaw JE, Miller DM 3rd. 2007. UNC-4 represses CEH-12/HB9 to specify synaptic inputs to VA motor neurons in *C. elegans*. Genes Dev. 21:332–346.

Von Stetina SE, Treinin M, Miller DMI. 2006. The motor circuit. Int Rev Neurobiol. 69:125–167.

Wacker I, Schwarz V, Hedgecock EM, Hutter H. 2003. *zag-1*, a Zn-finger homeodomain transcription factor controlling neuronal differentiation and axon outgrowth in *C. elegans*. Development. 130:3795–3805.

Wagner A. 2008. Neutralism and selectionism: a network-based reconciliation. Nature Rev. 9:965–974.

Wagner GP, Lynch VJ. 2008. The gene regulatory logic of transcription factor evolution. Trends Ecol Evol. 23:377–385.

Way J, Chalfie CM. 1989. The *mec-3* gene of *Caenorhabditis elegans* requires its own product for maintained expression and is expressed in three neuronal cell types. Genes Dev. 3:1823–1833.

Way J, Chalfie CM. 1988. *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. Cell. 54:5–16.

White JG, Southgate E, Thomson JN, Brenner S. 1986. The structure of the nervous system of the nematode *Caenorhabditis elegans*. Philos Trans R Soc Lond B Biol Sci. 314:1–340.

Wightman B, Baran R, Garriga G. 1997. Genes that guide growth cones along the *C. elegans* ventral nerve cord. Development. 124:2571–2580.

Wightman B, Ebert B, Carmean N, Weber K, Clever S. 2005. The *C. elegans* nuclear receptor gene *fax-1* and homeobox gene *unc-42* coordinate interneuron identity by regulating the expression of glutamate receptor subunits and other neuron-specific genes. Dev Biol. 287:74–85.

Winnier AR, Meir JY, Ross JM, Tavernarakis N, Driscoll M, Ishihara T, Katsura I, Miller DM 3rd. 1999. UNC-4/UNC-37-dependent repression of motor neuron-specific genes controls synaptic choice in *Caenorhabditis elegans*. Genes Dev. 13: 2774–2786.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 8:206–216.

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

Wu J, Duggan A, Chalfie M. 2001. Inhibition of touch cell fate by *egl-44* and *egl-46* in *C. elegans*. Genes Dev. 15:789–802.

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

Yu H, Pretot RF, Burglin TR, Sternberg PW. 2003. Distinct roles of transcription factors EGL-46 and DAF-19 in specifying the functionality of a polycystin-expressing sensory neuron necessary for *C. elegans* male vulva location behavior. Development. 130:5217–5227.

Zhang J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol. 50:56–68.

Zhao C, Emmons SW. 1995. A transcription factor controlling development of peripheral sense organs in *C. elegans*. Nature. 373:74–78.

Zheng X, Chung S, Tanabe T, Sze JY. 2005. Cell-type specific regulation of serotonergic identity by the *C. elegans* LIM-homeodomain factor LIM-4. Dev Biol. 286:618–628.

Zhou HM, Walthall WW. 1998. UNC-55, an orphan nuclear hormone receptor, orchestrates synaptic specificity among two classes of motor neurons in *Caenorhabditis elegans*. J Neurosci. 18:10438–10444.