Mol Gen Genomics (2005) 273: 299–310
DOI 10.1007/s00438-004-1105-6

ORIGINAL PAPER

Richard Jovelin · Patrick C. Phillips

Functional constraint and divergence in the G protein family in *Caenorhabditis elegans* and *Caenorhabditis briggsae*

Received: 2 July 2004 / Accepted: 9 December 2004 / Published online: 27 April 2005

Abstract Part of the challenge of the post-genomic world is to identify functional elements within the wide array of information generated by genome sequencing. Although cross-species comparisons and investigation of rates of sequence divergence are an efficient approach, the relationship between sequence divergence and functional conservation is not clear. Here, we use a comparative approach to examine questions of evolutionary rates and conserved function within the guanine nucleotide-binding protein (G protein) gene family in nematodes of the genus *Caenorhabditis*. In particular, we show that, in cases where the *Caenorhabditis elegans* ortholog shows a loss-of-function phenotype, G protein genes of *C. elegans* and *Caenorhabditis briggsae* diverge on average three times more slowly than G protein genes that do not exhibit any phenotype when mutated in *C. elegans*, suggesting that genes with loss of function phenotypes are subject to stronger selective constraints in relation to their function in both species. Our results also indicate that selection is as strong on G proteins involved in environmental perception as it is on those controlling other important processes. Finally, using phylogenetic footprinting, we identify a conserved non-coding motif present in multiple copies in the genomes of four species of *Caenorhabditis*. The presence of this motif in the same intron in the *gpa-1* genes of *C. elegans*, *C. briggsae* and *Caenorhabditis remanei* suggests that it plays a role in the regulation of *gpa-1*, as well as other loci.

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00438-004-1105-6

Communicated by S. Hekimi

R. Jovelin · P. C. Phillips (✉)
Center for Ecology and Evolutionary Biology,
University of Oregon,
Eugene, OR 97403-5289, USA
E-mail: pphil@darkwing.uoregon.edu

Keywords *Caenorhabditis elegans* · *Caenorhabditis briggsae* · G protein · Divergence · Gene regulation

Introduction

Recent whole genome sequencing projects have revealed that a substantial portion of genome evolution consists of divergence and diversification of gene families (e.g., Chervitz et al. 1998; Lander et al. 2001; Venter et al. 2001; Zdobnov et al. 2002). One of the primary challenges in this emerging field is to use information on sequence similarity and divergence among genomes to infer gene function. Very low rates of change might suggest a critical function for a sequence, whereas more rapid divergence could indicate relaxed selection or even, potentially, strong selection for divergent function. In an evolutionary context, these questions can be addressed by comparing the rate of non-synonymous change (K_A), which leads to changes in protein sequence, to the rate of synonymous change (K_S) which does not (Graur and Li 2000). Might it be possible to tie these evolutionary rates directly to the phenotypic effects or pattern of expression of the genes? Although it is possible to perform an evolutionary analysis for any set of genes, doing so within the context of an entire gene family for which we have a great deal of direct functional information allows the evolutionary pattern to be tied to functional consequences, while at the same time yielding significant insights into the evolution of that gene family. Given their central role within a wide variety of biological systems and their broad diversity (Neer 1995), guanine nucleotide-binding proteins (G proteins) make good candidates for wide scale genomic analysis.

Signal transduction involving G proteins is a common and effective process in cell communication, and mediates vision, smell, taste, neurotransmission and various hormonal responses (reviewed in Clapham 1996). Here, using comprehensive comparative analyses of G protein diversification within the nematode genus

Caenorhabditis, we address questions of evolutionary rates and conserved function within this important gene family. *Caenorhabditis elegans* is especially useful in this regard because of the existence of genetic knockouts for every member of the family (Jansen et al. 1999), a rare resource for a metazoan.

G protein function in *C. elegans*

At the molecular level, G proteins act as molecular switches. G proteins are heteromeric proteins made of the three subunits α, β and γ. In an inactive form, these subunits are bound to each other, with the α subunit bound to GDP. Upon activation of a G protein-coupled receptor by its ligand, the activated G protein exchanges GDP for GTP, changing the conformation of G_α. In this conformation, the affinity for the G_β and G_γ subunits is weakened and the $G_{\beta\gamma}$ dimer is eventually released. Both the GTP-bound G_α subunit and the $G_{\beta\gamma}$ dimer can activate their own effectors, thus allowing the transduction of the signal in the cell. Hydrolysis of GTP into GDP by G_α brings it back to its original conformation and promotes its reassociation with $G_{\beta\gamma}$, thus inhibiting $G_{\beta\gamma}$ signaling.

The well-studied nematode *C. elegans* lives in the soil where it perceives its environment mainly by chemosensation (Troemel 1999). Its repertoire of hundreds of G protein-coupled receptors (Robertson 1998, 2000; Troemel 1999) allows the worm to detect a wide array of water-soluble and volatile compounds using mechanisms analogous to those responsible for taste and smell in higher organisms. G protein signaling regulates multiple aspects of the worm's life and allows *C. elegans* to locate food and avoid toxins (reviewed in Troemel 1999), but is also used in male-mating (Liu and Sternberg 1995), egg-laying (Daniels et al. 2000) and locomotion (Mendel et al. 1995; Ségalat et al. 1995; Brundage et al. 1996). G protein signaling is also important during embryogenesis, controlling axon guidance and morphogenesis (Roayaie et al. 1998), vulval fate (Moghal et al. 2003), and spindle orientation during cell division (Zwaal et al. 1996; Gotta and Ahringer 2001). During later development, G proteins participate in regulating larval development in response to food level, temperature and population density (reviewed in Riddle and Albert 1997).

Analysis of the whole *C. elegans* genome (The *C. elegans* Sequencing Consortium 1998) revealed twenty G_α, two G_β and two G_γ genes (Jansen et al. 1999). Among the *C. elegans* G proteins, 15 G_α, both G_β and both G_γ genes are expressed in the amphid neurons (Mendel et al. 1995; Ségalat et al. 1995; Zwaal et al. 1996, 1997; Korswagen et al. 1997; Park et al. 1997; Jansen et al. 1999, 2002; Jiang and Sternberg 1999; Chase et al. 2001; van der Linden et al. 2001). These neurons are known to be involved in the perception of water-soluble and volatile compounds, the dauer pheromone, thermal cues and touch avoidance (reviewed in Bargmann and Mori 1997). Furthermore, while some G proteins are expressed only in a small subset of neurons, others show a broad pattern of expression in a wide range of neural, muscle and epithelial cells (Mendel et al. 1995; Ségalat et al. 1995; Zwaal et al. 1996, 1997; Korswagen et al. 1997; Park et al. 1997; Jansen et al. 1999, 2002; Lackner et al. 1999; Chase et al. 2001; van der Linden et al. 2001, 2003).

In this study we center our investigation on two axes. We first identify all members of the G protein family in *Caenorhabditis briggsae* and examine the relationship between genomic change and protein function by comparing the rates of divergence between G proteins with obvious loss-of-function phenotypes and those without. Second, we use "phylogenetic footprinting" to identify novel putative regulatory sequences of the G protein genes.

Materials and methods

Identification of G protein homologues

A search for G protein homologues in *C. briggsae* was performed by systematically employing each of the *C. elegans* G protein sequences to query the *C. briggsae* whole genome assembly (version cb25.agp8; Stein et al. 2003) using the TBLASTN program (Altschul et al. 1997) from the Sanger Institute's *C. briggsae* BLAST server (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_briggsae). Contig fragments giving significant alignments by TBLASTN were pulled out of the contig sequence. Exon/intron boundaries were then identified using the gene structure prediction system GeneBuilder (Milanesi et al. 1999), implemented as part of the WebGene package (http://www.itba.mi.cnr.it/webgene) and by aligned, this contig fragment with the nucleotide sequence of the *C. elegans* G protein but was by eye using BioEdit (Hall 1999). *C. briggsae* G protein sequences were confirmed with reference to the ORF using the standard genetic code (Osawa et al. 1992 ; Jukes and Osawa 1993) and homologues were then identified on the basis of amino acid sequence identity. Orthology between *C. elegans* and *C. briggsae* G protein α subunits was further compared by building a tree depicting the relationships among the proteins. Ambiguous positions were removed from the protein sequence alignment and a neighbor-joining tree was built using MEGA version 2.1 (Kumar et al. 2001). Confidence was assessed using bootstrapping (Felsenstein 1985) with 500 replicates. Due to the early divergence of the G protein family within the eukaryotes (results not shown) and the lack of a suitable outgroup, the NJ tree is unrooted and the relationships among paralogues remain uncertain.

We found a homologue for each the G proteins except *gpa-9* using this strategy. However, *Cb gpa-2* and *Cb gpb-2* exon 1 could not be identified; moreover, sequencing errors were found in *Cb gsa-1* exon 7 and intron 1, and in *Cb gpa-1* intron 4. We used Rapid Amplification of cDNA Ends PCR (RACE-PCR) to amplify the 5′ extremity of *Cb gpa-2* and thus identify

the first exon of the sequence. We successfully amplified and sequenced *Cb gsa-1* exon 7 and *Cb gpa-1* intron 4; however, RACE-PCRs failed to amplify *Cb gpa-2* or *Cb gsa-1* intron 1. *Cb gpb-2* exon 1 remains unidentified. Because the 5′ end of the *Cb gpa-11* sequence was found to be very divergent from that of *Ce gpa-11* we used reverse transcription-polymerase chain reaction (RT-PCR) to confirm the cDNA sequence of *Cb gpa-11* exon 1. A partial *gpa-1* sequence from *Caenorhabditis remanei* was obtained using degenerate primers.

Primers used for amplification and sequencing are available on request from the authors. *C. briggsae* G protein sequences were deposited in GenBank with the partial sequence from *C. remanei gpa-1* under the Accession Nos. AY634285–AY634310. *Cb odr-3* from our previous study (Jovelin et al. 2003) was released under Accession No. AY146578.

DNA and RNA extraction

Strains AF16, HK104, HK105, DH1300, N2, PB269 and CB5161 were maintained following standard protocols (Brenner 1974). Worms were harvested from plates when the population grew to a large size and washed twice in 1-ml aliquots of distilled water. Genomic DNA was extracted using the CTAB protocol (Winnepenninckx et al. 1993) and RNA was extracted from plates containing worms at all stages of development using TRI Reagent (Molecular Research Center) following the supplier's protocol.

Amplification, cloning and sequencing

Polymerase chain reaction and sequencing reactions were processed as described in Jovelin et al. (2003). Primers used for amplification were also used for sequencing; sequences overlapped and thus were confirmed on both strands. RNA extracted from the AF16 and N2 strains was used for cDNA synthesis using the SMART RACE cDNA amplification kit (Clontech), followed by reamplification by nested PCR. Following gel purification, RACE-PCRs, RT-PCRs, PCRs using degenerate primers and PCRs of motif D1 were cloned in the vector pGEM-T (Promega) and transfected into One Shot Top10 competent cells (Invitrogen). Positive colonies were visually screened and purified using QIAprep Spin Miniprep kit (Qiagen). Sequencing was performed with primers designed from the plasmid sequence.

Sequence analysis

As was found for intronic sequences within *odr-3* (Jovelin et al. 2003), intronic sequences of the other G protein genes are highly divergent. We therefore used the cDNA sequences to assess the level of divergence between G protein orthologues from *C. elegans* and *C. briggsae*. Alignments were done by eye using Bioedit (Hall 1999). We computed maximum likelihood (ML) estimates of the rate of non-synonymous (K_A) and synonymous (K_S) substitutions using a codon-based model of sequence evolution. ML estimates of K_A and K_S were calculated with the CODEML program implemented in the PAML package (Yang 1997) with the transition/transversion ratio (κ) being estimated by iteration from the data and codon frequencies estimated from the products of the average observed nucleotide frequencies in the three codon positions (F3X4 model, Goldman and Yang 1994). The ratio K_A/K_S and K_A alone were used as measures of divergence. In addition, K_A and K_S were computed using MEGA version 2.1 (Kumar et al. 2001) for each pair of orthologues using the method of Nei and Gojobori (1986) with a p-distance. K_A was also calculated with a Jukes–Cantor distance (Jukes and Cantor 1969). Similar results were found with these values of K_A/K_S and K_A (data not shown).

Results

G protein orthologues in *C. briggsae*

We identified *C. briggsae* homologues on the basis of amino acid sequence identity. The *C. briggsae* genomic sequence contains 19 G_α genes, 2 G_β genes and 2 G_γ genes, each of them having a clear orthologue in the *C. elegans* genome (Table 1 and Fig. 1). However no orthologue for *gpa-9* was found in the *C. briggsae* genome. BLAST analyses performed with *Ce* GPA-9 retrieved only 15 contigs out of the 188 constituting the assembly. Significant alignments within each of these contigs corresponded to positions where the 15 other $G_\alpha$genes have been identified. Other alignments within these contigs consisted of very short fragments. Furthermore, none of these short alignments were located at contig ends, eliminating the possibility that the *Cb gpa-9* sequence might be truncated at a contig end and consequently not be detected. Therefore it is reasonable to conclude that *gpa-9* is not present in assembly cb25.agp8. Although the assembly is estimated to cover 98% of the *C. briggsae* genome (Stein et al. 2003), we cannot totally rule out the possibility that *Cb gpa-9* might be located in a region of the genome not yet sequenced. Nevertheless, two alternative hypotheses are worth considering.

gpa-9 is flanked by *gpa-8* and *lin-25* in *C. elegans* clone F56H9. Interestingly, *gpa-8* and *lin-25* are found in synteny in *C. briggsae* contig FPC2051 (Fig. 2). Thus two events could explain the absence of *gpa-9* in *C. briggsae*: the loss of *gpa-9* in the briggsae lineage or the duplication of *gpa-8* in the elegans lineage. The distance between *Ce gpa-8* and *Ce lin-25* in clone F56H9 is 9455 bp whereas the distance between *Cb gpa-8* and *Cb lin-25* in contig FPC2051 is only 6,310 bp, and the difference is approximately equal to the length of *Ce gpa-9* (2,475 bp). Synteny alone does not give more weight to one hypothesis over the other, but it does

Table 1 Divergence between *C. elegans* and *C. briggsae* G protein homologues

Locus	AF16 (*C. briggsae*) vs. N2 (*C. elegans*)[a]					Loss of function[b] (ref.)	Gain of function[b] (ref.)	Expression pattern in N2[c]				Ref.
	cDNA	AA	K_A	K_S	K_A/K_S			Amphid neurons	Sensory neurons	Other neurons	Other tissues	
gpa-1	75.6	78.4	0.124	2.3748	0.052	None (1, 14)	Mut (1)	ADL, ASH, ASI, ASJ	PHA, PHB	SPD, SPC, SPV, other neurons in the male tail		1, 2
gpa-2	77.4	89.6	0.063	2.5049	0.025	Mut (1, 3)	Mut (1, 3)	AWC	PHA, PHB	I5, AIA, M1, M5, PVT	mu-sph	3
gpa-3	83.6	95.4	0.029	1.3778	0.021	Mut (1, 3)	Mut (1, 3)	ADF, ADL, ASE, ASG, ASH, ASI, ASJ, ASK	PHA, PHB	PVT, AIZ		3
gpa-4	73.3	75.7	0.139	2.1619	0.064	None (1)	Mut (1)	ASI				1
gpa-5	70	73.5	0.167	4.9579	0.034	None (1)	Mut (1)	AWA, ASI				1
gpa-6	79.7	93.7	0.029	2.1565	0.013	Mut (1)	Mut (1)	AWA, ASI	PHB			1
gpa-7	82.3	99.4	0.004	1.8235	0.002	Mut (1)	Mut (1)	Many neurons	Many neurons	Many neurons in the male tail	Muscle cells	
gpa-8	83.8	92.8	0.035	1.8276	0.019	None (1)	None (1)		PQR, AQR	URX		1
gpa-9	NA	NA	NA	NA	NA	None (1)	NA	ASJ	PHB	PVQ	Pharyngeal muscles, spermatheca	1
gpa-10	65.6	66.5	0.240	7.9427	0.030	None (1)	Mut (1)	ADF, ASI, ASJ		CAN, ALN, LUA	Spermatheca	1
gpa-11	74.4	NA	NA	NA	NA	None (1)	NA	ADL, ASH				1
gpa-12	84.5	99.4	0.002	1.6898	0.001	None (15)	Mut (15)			Precursors of motoneurons (P cells)	Pharyngeal muscle cells, main body hypodermis	15
gpa-13	76.8	86.8	0.081	2.9966	0.027	NA	NA	ADF, ASH, AWC	PHA, PHB			1
gpa-14	72.9	76.9	0.138	4.2982	0.032	None (1)	NA	ASI, ASJ, ASH, ASK	ADE, PHA, PHB	RIA, CAN, ALA, AVA, PVQ, DVA	Vulval muscle, vulval hypodermis (weak and only in old larvae and adults)	
gpa-15	82.4	91.2	0.042	1.8098	0.023	None (1)	NA	ADL, ASH, ASK	PHA, PHB	Many male-specific neurons	Distal tip cells, anchor cell	1
gpa-16	75	81.5	0.111	2.1928	0.050	Mut (16, 20)	NA		AVM, PDE, PLM	BDU, PVC, RIP	Faintly in pharynx, body wall muscle, vulva muscle	1
egl-30	88.2	98.3	0.009	0.6659	0.014	Mut (18)	Mut (4, 18)			Ventral cord motor neurons, other neurons	Pharyngeal muscle	4
odr-3	85.7	99.4	0.003	1.5187	0.002	Mut (5)	Mut (5)	AWA, AWB, AWC, ASH, ADF				5
goa-1	85.5	98.5	0.010	1.0198	0.010	Mut (16, 17, 7)	Mut (6, 7)	All	All	All	Vulva, spermatheca, uterine and intestinal muscles, diagonal muscles in the male, most cells of the pharynx, distal tip cells	6, 7
gsa-1	81.5	93.3	0.036	2.7207	0.013	Mut (9)	Mut (9, 19)	All	All	All	Most muscle cells, pahryngeal epithelium, vulva epithelium, excretory cells	8, 9
gpb-1	88.6	100	0	0.6029	0	Mut (10,16)	Mut (10)	All	All	All	Gonad, vulva, hypodermal seam cells, intestine, pharynx, body wall muscles, germline	10

gpb-2[b]	80.7	99.4	0.007	2.6037	0.003	Mut (11)	None (11)	All	All	All, CAN, HSN	Pharyngeal muscles, vulva muscles, body-wall muscles	11, 12
gpc-1	75.6	80.6	0.106	7.4463	0.014	Mut (13)	Mut (13)	ADL, ASH, ASJ, AFD, ASI, AWB	PHB			13
gpc-2	86.7	98.3	0.007	0.9476	0.008	Mut (16)	NA	All	All	All	All muscle cells	13

Measures of divergence were computed using partial *gpb-2* sequence, however only the first exon is missing (38 bp in *C. elegans*, 3.55% of the length of the whole gene)/ Jansen et al. (1999); *2* Jiang and Sternberg (1999); *3* Zwaal et al. (1997); *4* Lackner et al. (1999); *5* Roayaie et al. (1998); *6* Mendel et al. (1995); *7* Segalat et al. (1995); *8* Park et al. (1997); *9* Korswagen et al. (1997); *10* Zwaal et al. (1996); *11* van der Linden et al. (2001); *12* Chase et al. (2001); *13* Jansen et al. (2002); *14* Zwaal et al. (1993); *15* van der Linden et al. (2003); *16* Gotta and Ahringer (2001); *17* Nurrish et al. (1999); *18* Brundage et al. (1996); *19* Berger et al. 1998; *NA* not available

[a]*C. briggsae* G protein homologues were identified by comparison known G proteins in *C. elegans*, and compared on both the nucleotide (cDNA) and amino acid (AA) sequence levels. The ratio of the rate of nonsynonymous (K_A) to the rate of synonymous (K_S) substitutions and the K_A value alone were used as measures of divergence. K_A/K_S was estimated by maximum-likelihood using a codon-based model of evolution (Goldman and Yang 1994)

[b]Phenotypes for loss and gain of function mutations in *C. elegans* are indicated for each locus. *mut* indicates that the mutation is lethal

[c]The pattern of expression of *C. elegans* G proteins is divided into neuronal and non-neuronal expression. Within the neuronal expression class, the pattern of expression is further divided depending of the type of neurons in which the G proteins are expressed

Fig. 1 Unrooted neighbor-joining tree derived from a protein sequence alignment including all known *C. elegans* and *C. briggsae* alpha subunits. All pairings of orthologues have 100% bootstrap support. All other nodes with greater than 80% bootstrap support are also indicated

argue strongly that either a gain or a loss did indeed occur. However, if *Ce gpa-9* was duplicated from *Ce gpa-8* after the species split, one would expect the rate of synonymous substitution to be lower between *Ce gpa-8* and *Ce gpa-9* than between *gpa-8* orthologues. The rate of synonymous changes (K_S) between *Ce gpa-8* and *Ce gpa-9* is 1.53-fold greater than between *gpa-8* orthologues ($K_S = 2.8$ and 1.83, respectively) suggesting that *gpa-8* and *gpa-9* are ancient duplicates and that *gpa-9* was lost in the *C. briggsae* lineage. Final discrimination between the gain and loss models will require investigation of *gpa-9* in additional *Caenorhabditis* species within a phylogenetic context.

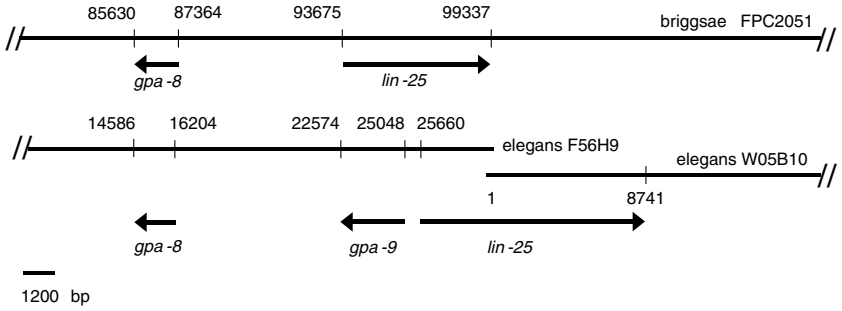

Fig. 2 Schematic diagram of parts of the *C. elegans* clones F56H9 and W05B10 and the *C. briggsae* contig FPC2051. *gpa-8* and *lin-25* homologues are found in synteny but *gpa-9* is absent from the *C. briggsae* contig. The length between *Cb gpa-8* and *Cb-lin-25* corresponds approximately to the length between *Ce gpa-8* and *Ce lin-25* minus the length of *Ce gpa-9*

We also found an insertion of 16 bp in the first exon at position +10 of our predicted *Cb gpa-11*. This insertion leads to a shift in the reading frame, with consequent loss of domains essential for guanine nucleotide binding activity, and several stop codons. After re-sequencing partial predicted *Cb gpa-11* to ensure there was no sequencing error in the genomic sequence, we tested whether our prediction of *Cb gpa-11* was correct by performing a RT-PCR with a forward primer located at position +1 in exon 1 and a reverse primer located in exon 3. Sequencing of the RT-PCR product confirmed that *Cb gpa-11* is transcribed; however, the presence of an insertion which alters the translation frame from the very beginning of the sequence suggests that *gpa-11* is not functional and is probably not translated in *C. briggsae* AF16. Sequencing of partial *gpa-11* from three other wild isolates of *C. briggsae* (DH1300, HK104 and HK105) revealed the presence of the insertion in exon 1. These findings suggest that the loss or severely alteration of *gpa-11* function is not unique to AF16 but is characteristic of *C. briggsae*.

Divergence correlates with loss-of-function phenotypes

Using data for loss-of-function phenotypes in *C. elegans* (Table 1), we tested three hypotheses of functional constraint and the degree of conservation of G proteins within *C. elegans* and *C. briggsae*. First, loss-of-function data were divided into two categories, presence or absence of an obvious phenotype. There is a highly significant relationship between the presence of a phenotype and the ratio of the K_A/K_S ratio ($t_{19}=2.604$, $P=0.0174$; Wilcoxon two-sample $P=0.0225$). The ratio K_A/K_S is a measure of selection: if $K_A/K_S \approx 1$, no selection (neutral evolution) is indicated; a K_A/K_S value of <1 indicates purifying selection, and $K_A/K_S>1$ implies positive selection. Because K_S values may be underestimated due to codon bias even with likelihood methods (Dunn et al. 2001) and because we did not see any significant differences in the rate of synonymous changes (K_S: $t_{19}=1.443$, $P=0.1652$; Wilcoxon two-sample $P=0.1376$) (Fig. 3), we also used values of K_A as an index of divergence. The results using K_A are consistent with those using the K_A/K_S ratio (K_A: $t_{19}=3.098$, $P=0.0059$; Wilcoxon two-sample $P=0.0186$). Loci showing a loss of function phenotype diverge on average three times more slowly than those that do not show any phenotype (Fig. 3).

In the second test, the loss-of-function data was divided into lethal and non-lethal effects (i.e. essential vs. non-essential genes). Although essential loci tend to be less divergent than non-essential loci, the difference is not significant (K_A/K_S: $t_{19}=0.499$, $P=0.6237$; Wilcoxon two-sample $P=0.3859$; K_A: $t_{19}=1.114$, $P=0.2792$; Wilcoxon two-sample $P=0.3020$).

Finally, we tested the relationship between the G protein expression pattern and rates of evolutionary change. Some G proteins, such as *gpa-2*, *gpa-3* and *odr-3*, show expression in the amphid neurons and are directly involved in chemosensation (Zwaal et al. 1997; Royaie et al. 1998), while others, such as *gpa-12* and *gpa-16*, are not expressed in the amphid neurons but are instead involved in developmental processes (Gotta and Ahringer 2001; van der Linden et al. 2003). We therefore used the expression data (Table 1) to test the hypothesis that differential spatial expression underlies differential functional constraints. We assumed that G proteins showing restricted expression within the amphid or phasmid sensory neurons are limited to the transduction of external stimuli, whereas G proteins showing expression in other types of cells might be involved in the transduction of internal stimuli and play a role in more central processes (e.g. development). We classified pattern of expression data into two categories: narrow when loci are expressed only in sensory neurons and broad otherwise. However, we found no evidence that G proteins showing narrow expression tend to be less constrained (K_A/K_S: $t_{20}=-0.783$, $P=0.4428$; Wilcoxon two-sample $P=0.4389$; K_A: $t_{20}=-1.065$, $P=0.2998$, Wilcoxon two-sample $P=0.2688$).

Conservation of a non-coding motif

As found previously in the genus *Caenorhabditis* (e.g., Kennedy et al. 1993; Shabalina and Kondrashov 1999; Jovelin et al. 2003), non-coding regions of these genes are extremely divergent. Thus, conserved motifs in those regions are good candidates for promoters and regula-

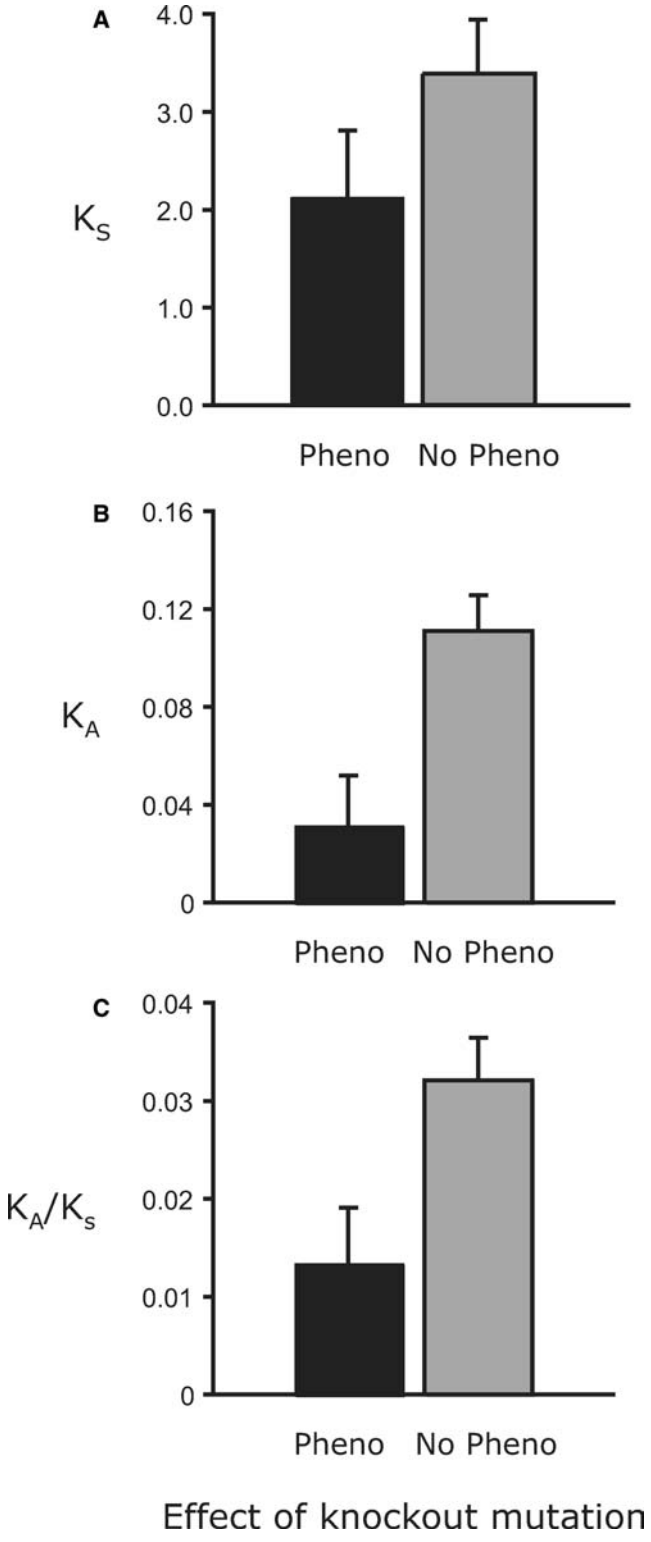

Fig. 3 Relationship between divergence and dispensability. No significant difference was found for the rate of synonymous change (K_S) between G proteins that show a loss-of-function phenotype in mutants and those that do not. The rate of non-synonymous change (K_A) is on average about three times lower for G

tory elements of gene expression. In this respect genomic comparison between *C. elegans* and *C. briggsae* may give useful insights into gene regulation and eventually increase our knowledge regarding the function of a particular gene within a given network. We found two extremely conserved motifs located in introns 5 and 7 of *Ce gpa-1* (introns 4 and 6 of *Cb gpa-1*). The latter is predicted to be a tRNA-Leu(CAG) when analysed using the program tRNAscan-SE (Lowe and Eddy 1997; available online at http://www.genetics.wustl.edu/eddy/tRNAscan-SE) and shows strong similarities to the previously characterized *Drosophila melanogaster* tRNA-Leu(CUG) (Glew et al. 1986). However, the motif located in *Ce gpa-1* intron 5, here named D1, showed no evidence of being a tRNA.

The approximately 280-bp motif D1 is more highly conserved than any exon, or than the coding sequence taken as a whole. However, amplification of cDNA both in *C. elegans* and *C. briggsae* using primers targeted to adjacent exons showed that motif D1 was effectively spliced in both species and did not show any evidence of alternative splicing. Furthermore, the predicted coding sequence of *Ce gpa-1* was confirmed in a recent study (Cuppen et al. 2003). We also tested in silico the possibility that motif D1 was an exon of another locus by searching for expressed sequence tags (ESTs) containing its sequence. None were found in the *C. elegans* ESTs database. Similarly, analysis of the distribution of stop codons and evolution at putative third position sites also shows that this is not a previously unidentified *gpa-1* exon, nor is it an exon from an overlapping locus (Supplementary Fig. S1).

Interestingly, it appears that the motif located in intron 5 of *Ce gpa-1* is also found in inverted orientation in *Ce gpa-1* intron 1. Thus we searched the *C. elegans* and *C. briggsae* genomes for other occurrences of D1 using the BLASTN program (Altschul et al. 1990) on the *C. elegans* (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_elegans) and *C. briggsae* BLAST servers (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/c_briggsae) at the Sanger Institute. Highly conserved sequences of this motif were found in several copies within the genomes of *C. elegans* and *C. briggsae* (Supplementary Fig. S2).

Nevertheless, conservation in non-coding sequences between two species can simply be the result of shared ancestry and heterogeneous rates of substitutions instead of functionality. With the reasonable assumption that sequence conservation across multiple species reflects functional constraints, we used degenerate primers to amplify and sequence *gpa-1* from a third species, *C. remanei* (strain PB269). Sequencing of a partial *Cr gpa-1* sequence (a fragment 909 bp long corresponding to 113 amino acids) revealed the presence of motif D1. Remarkably, the location of D1 within the *Cr gpa-1* sequence is also conserved: the motif is found adjacent to the homologue of *Ce gpa-1* exon 5. In addition, we used specific primers designed to anneal within the D1 sequence of *C. remanei* and *C.* sp (strain CB5161). We

Table 2 Location of motif D1 in the *C. elegans*, *C. briggsae* and *C. remanei* genomes

Clone/contig	Species[a]	Motif	Frame[b]	Position	Putative regulatory element of locus[c]	Frame[b]	Position	Location[c]	Predicted product of locus
T19C4	*C. elegans*	D1	+1	29768–30138	*gpa-1*	−1	28868–31860	Intron 5 of *gpa-1*	G protein α subunit
20CB25.FPC4470	*C. briggsae*	D1	+1	247550–247912	*gpa-1*	−1	245398–248549	Intron 4 of *gpa-1*	G protein α subunit
20NA	*C. remanei*	D1	NA	NA	*gpa-1*	NA	NA	Presumably intron 5 of *gpa-1*	G protein α subunit
20T19C4	*C. elegans*	D1	−1	31113–31482	*gpa-1*	−1	28868–31860	Intron 1 of *gpa-1*	G protein α subunit
20F59C6	*C. elegans*	D1	+1	11990–12360	*nlp-4*	−1	11148–12815	Intron 2 of *nlp-4*	Neuropeptide
20CB25.FPC0053	*C. briggsae*	D1	−1	109227–109597	*nlp-4* and/ or *che-13*	+1/+1	108397–108936; 109847–110276	290 bp downstream; 249 bp upstream	Neuropeptide; chemosensory behavior and cilium biogenesis
0F25H2	*C. elegans*	D1	−1	31227–31598	*ubc-25* and/ or *pas-5*	+1/+1	29061–31007; 32446–33390	219 bp downstream; 847 bp upstream	Calcium ion binding; ubiquitin conjugating enzyme; endopeptidase
Tright F43A11	*C. elegans*	D1	+1	2027–2397	F43A11.1	+1	410–2013	13 bp downtream	G protein-coupled receptor (rhodopsin-like receptor)
20Y67A10A	*C. elegans*	D1	+1	8546–8917	ND				
0CB25.FPC4470	*C. briggsae*	D1	−1	259299–259669	ND				
CB25.FPC4470	*C. briggsae*	D1	−1	412805–413175	ND				

[a]Note that several haplotypes of motif D1 were also found in the genomes of *C. remanei* and *C.* sp (CB5161) but only the motif found in the homologous intron of *Ce gpa-1* is reported here for *C. remanei*

[b]The frame of motif D1 indicates the orientation of the sequence presented in Supplementary Figure S2 with respect to the genomic sequence of the corresponding species

[c]Loci found near motif D1 are also indicated, including their location and orientation within the genomic sequence and their biological activity. These data suggest that motif D1 is functional and may play a role in the regulation of these loci

randomly sequenced 10 clones in each species and found respectively 7 and 5 haplotypes in PB269 and CB5161 (Supplementary Fig. S2). These results strongly suggest that the degree of similarity observed for D1 across species is not simply due to a lower rate of substitution within this motif but rather to active conservation of its sequence-probably in relation to a functional role of this motif.

Discussion

G protein divergence and function

We found 19 G_α genes, 2 G_β genes and 2 G_γ genes in the *C. briggsae* genome; however, we could not find an orthologue of *Ce gpa-9*. The conserved synteny observed for genes flanking *Ce gpa-9* strongly suggests that *gpa-9* was either lost in the *C. briggsae* lineage or resulted from the duplication of *gpa-8* in the *C. elegans* lineage. Comparison of the rate of synonymous change between these paralogues suggests that *gpa-9* was indeed lost in *C. briggsae*. In addition to this loss, the presence of an insertion that alters the translation frame at the beginning of the coding sequence suggests that *gpa-11* is also not functional in *C. briggsae*. A recent study shows that olfaction in *C. elegans* is achieved by a balance of inhibitory and stimulatory signals through a complex network of G_α subunits (Lans et al. 2004). In particular, Lans et al. (2004) show that G_α ODR-3 and GPA-3 act redundantly both in the AWA and AWC neurons for the detection of odorants, with ODR-3 being the main stimulatory signal. The redundancy in G_α subunit function is also suggested by the overlaps in their patterns of expression (Table 1), although it appears that some G_α have specific functions despite having the same cellular localization (Lans et al. 2004). Interestingly, *odr-3* and *gpa-3* are relatively closely related to one another (Fig. 1). It is possible that this type of redundancy is a broader property of G-proteins in general, setting the stage for relaxed selection and eventual loss of some components. Alternatively, in other circumstances, this redundancy could allow the evolution of novel functions.

Conservation of a functional non-coding motif

Because *C. elegans* and *C. briggsae* diverged 80–110 MYA (Stein et al. 2003) and are far enough apart that most of the non-coding regions are extremely divergent (Shabalina and Kondrashov 1999), our finding of a widespread highly conserved element in non-coding regions suggests that this motif, named D1, may have a role in gene regulation. An alternative explanation-that it is a transposable element-seems quite unlikely, as one would expect that homologues of D1 between *C. elegans* and *C. briggsae* would acquire substitutions at a higher rate than homologous coding sequences between these two species. The highest number of substitutions in the core-conserved motif between *C. elegans* and *C. briggsae* sequences (positions 57–259 in Supplementary Fig. S2) is 25 substitutions between *Ce* F43A11-D1 and *Cb* FPC0053-D1. The average degree of conservation for all G protein alpha subunits between *C. elegans* and *C. briggsae* is 78.9% identity at the cDNA level; thus, in a 200-bp fragment of coding sequence we could expect 42 substitutions on average, 1.68-fold more than is found between the most divergent pair *Ce* F43A11-D1 and *Cb* FPC0053-D1. Furthermore, this non-coding motif is found at the same position in the *gpa-1* gene in three species of *Caenorhabditis*, and its sequence is highly conserved in four species. This finding is consistent with a number of cross-species comparisons, which have identified functional non-coding conserved sequences (Frazer et al. 2001, 2004; Chureau et al. 2002; Dermitzakis et al. 2002; Hare and Palumbi 2003). For a subset of these there is evidence to support a regulatory activity (Loots et al. 2000; Göttgens et al. 2000; Thomas et al. 2003).

If motif D1 is functional and has a regulatory activity, then we might expect orthologous genes to show an association with motif D1 if they are in fact regulated by this motif (Table 2). We have already shown an association of motif D1 with *gpa-1* across *C. elegans*, *C. briggsae* and *C. remanei*. Although the exact location is not preserved in *C. elegans* and *C. briggsae*, motif D1 is also found in association with *nlp-4*, a member of the *nlp* superfamily of neuropeptide genes (Nathoo et al. 2001) (Table 2). A look at the activity of other nearby genes in *C. elegans* and *C. briggsae* further suggests that motif D1 may play a role in the regulation of these genes (Table 2). In *C. elegans*, motif D1 is found 13 bp downstream of F43A11.1, a predicted G protein-coupled receptor. The association of motif D1 with the receptor F43A11.1, the G protein *gpa-1*, and the neuropeptide *nlp-4* suggests that these loci could be co-regulated and suggests a possible pathway in which these three loci could interact with each other. D1 is also found in association with other genes known to affect locomotion, egg laying, and defecation (Table 2; Schulze et al. 2003), although these associations are not always conserved across species. If motif D1 is confirmed to have regulatory activity, our findings would suggest that *C. elegans* and *C. briggsae* have common and unique patterns of regulation through this motif.

Divergence correlates with phenotype in loss-of-function mutants

Because the majority of amino acid substitutions are likely to be deleterious, a classical prediction in protein evolution is that the most dispensable proteins (i.e. proteins that contribute least to fitness) should be subject to weaker purifying selection and therefore accumulate slightly deleterious mutations at a faster rate than essential proteins (Wilson et al. 1977). On average, G proteins appear to be more conserved between *C. elegans* and *C. briggsae* (mean, 89% amino acid identity) than other *C. elegans*/*C. briggsae* orthologues (mean,

75% identity) (Stein et al. 2003), suggesting that they serve an essential role within these nematodes. Indeed, a number of G proteins show 100% conservation across species (Table 1). Similarly, we have shown that G proteins whose loss of function leads to a phenotype diverge on average three times more slowly than those that do not show any phenotype (Fig. 3). A similar correlation has been found between divergence and RNA interference (RNAi) phenotype in *C. elegans* (Cutter et al. 2003; Stein et al. 2003). It is noteworthy that *gpa-9* and *gpa-11*, which are, respectively, absent and not functional in *C. briggsae*, do not show any phenotype in loss-of-function mutants of *C. elegans*. However, we did not see any significant relationship between divergence and dispensability (Hurst and Smith 1999; Hirsh and Fraser 2001), despite the trend for essential loci (i.e. those with a lethal phenotype) to be less divergent than non-essential loci.

These results suggest that genes showing knock-out phenotypes have important functions and therefore evolve more slowly because of stronger selective constraints. However, there are exceptions within each class. For instance, both the G_α *gpa-2* and *gpa-16* and the G_γ *gpc-1* show a higher degree of divergence, although they exhibit loss-of-function phenotypes. Although all K_A/K_S ratios are very low and therefore do not suggest any signal of positive selection for these genes, because the fraction of sites potentially evolving by positive selection is small and the total divergence at non-coding sites is large, the effect of positive selection can be masked by purifying selection (Fay et al. 2001). For example, Smith and Eyre-Walker (2002) demonstrated an overall pattern of adaptive evolution in *Drosophila* even though positive selection was not evident for individual genes. Reanalysis of divergence at *gpa-2*, *gpa-16* and *gpc-1* with the addition of information on polymorphism within each species is worth pursuing (e.g., Smith and Eyre-Walker 2002; Bierne and Eyre-Walker 2004). These are the most likely candidates for positive selection within this group.

In contrast, *gpa-12* is highly conserved despite showing no obvious loss-of-function phenotype. In this case, *gpa-12* instead shows a strong gain-of-function phenotype (van der Linden et al. 2003), and the full consequences of its elimination might not be revealed until other interacting genes are also knocked out (Yau et al. 2003). Other, more conserved, genes with no obvious phenotypes would probably benefit from characterizing their knock-outs in a wider variety of environmental circumstances. Overall, differences in evolutionary rate are useful for indicating genes with marked functional consequences, with exceptions to the general pattern leading to further insights into the functional role of those genes.

Unlike Duret and Mouchiroud (2000), who found that tissue-specific expression influenced the rate of molecular divergence among mice, rats and humans, we did not see any significant difference in divergence between G proteins that are narrowly versus broadly expressed, despite the tendency for G proteins showing narrow expression to be less constrained. One explanation for this difference is that perception controlled by G proteins restricted to sensory neurons is no less important from a biological point of view than other processes controlled by G proteins showing a broad pattern of expression. This might not be true for every animal but makes sense in the light of the life history of *C. elegans*, in which perception controlled by G protein signaling is directly involved in food detection and toxin avoidance (see Troemel 1999), in male-mating (Liu and Sternberg 1995) and in regulating larval development in response to food level, temperature and population density (reviewed in Riddle and Albert 1997). Because these processes are likely to influence fitness, G proteins controlling perception must be subject to selective constraints at levels comparable to those that act on G proteins controlling other important processes such as egg-laying (Daniels et al. 2000), locomotion (Mendel et al. 1995; Ségalat et al. 1995; Brundage et al. 1996) and development (Zwaal et al. 1996; Gotta and Ahringer 2001; Moghal et al. 2003). The central role of this perception is supported by a recent mutation-accumulation study, which shows that chemosensory behavior is subject to mutational pressure on the same order as that for life history and morphological traits (Ajie et al. in press). Thus, the diverse family of G proteins in this group of nematodes is at the center of a wide variety of critical functions, and the investigation of their evolution provides valuable insights into the broader patterns of functional constraint within the genomes of these organisms.

Acknowledgements We thank the Caenorhabditis Genetics Center, which is supported by the National Institutes of Health National Center for Research Resources, for providing us with some of the strains used in this study. Scott Baird provided *C. remanei* strain PB269 and Joe Thornton provided help with the analysis. We also thank the Sanger Institute and the Genome Sequencing Center at Washington University, St Louis for releasing the unpublished *C. briggsae* genome sequence and providing the research community with this tremendous resource. Finally, we thank the reviewers for their valuable comments. This work was supported by a grant from the National Institutes of Health (GM54185)

References

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

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nuc Acids Res 25:3389–3402

Bargmann C, Mori I (1997) Chemotaxis and thermotaxis. In: Riddle D, Blumenthal T, Meyer B, Priess J (eds) *C. elegans* II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 717–737

Berger A, Hart A, Kaplan J (1998) G alphas-induced neurodegeneration in *Caenorhabditis elegans*. J Neurosci 18:2871–2880

Bierne N, Eyre-Walker A (2004) The genomic rate of adaptive amino acid substitution in Drosophila. Mol Biol Evol 21:1350–1360

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

Brundage L, Avery L, Katz A, Kim U, Mendel J, Sternberg P, Simon M (1996) Mutations in a *C. elegans* Gqalpha gene disrupt movement, egg laying, and viability. Neuron 16:999–1009

Chase DL, Patikoglou GA, Koelle MR (2001) Two RGS proteins that inhibit Gα(o) and Gα(q) signaling in *C. elegans* neurons require a Gβ(5)-like subunit for function. Curr Biol 11:222–231

Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T, Weng S, Cherry JM, Botstein D (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science 282:2022–2028

Chureau C, Prissette M, Bourdet A, Barbe V, Cattolico L, Jones L, Eggen A, Avner P, Duret L (2002) Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res 12:894–908

Clapham D (1996) The G-protein nanomachine. Nature 379:297–299

Cuppen E, van der Linden A, Jansen G, Plasterk R (2003) Proteins interacting with *Caenorhabditis elegans* Galpha subunits. Comp Funct Genomics 4:479–491

Cutter A, Payseur B, Salcedo T, Estes A, Good J, Wood E, Hartl T, Maughan H, Strempel J, Wang B, Bryan A, Dellos M (2003) Molecular correlates of genes exhibiting RNAi phenotypes in *Caenorhabditis elegans*. Genome Res 13:2651–2657

Daniels S, Ailion M, Thomas J, Sengupta P (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

Dermitzakis E, Reymond A, Lyle R, Scamuffa N, Ucla C, Deutsch S, Stevenson B, Flegel V, Bucher P, Jongeneel C, Antonarakis S (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420:578–582

Dunn K, Bielawski J, Yang Z (2001) Substitution rates in *Drosophila* nuclear genes: implications for translational selection. Genetics 157:295–305

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

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

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Frazer K, Sheehan J, Stokowski R, Chen X, Hosseini R, Cheng J, Fodor S, Cox D, Patil N (2001) Evolutionarily conserved sequences on human chromosome 21. Genome Res 11:1651–1659

Frazer K, Tao H, Osoegawa K, de Jong P, Chen X, Doherty M, Cox D (2004) Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. Genome Res 14:367–372

Glew L, Lo R, Reece T, Nichols M, Soll D, Bell J (1986) The nucleotide sequence, localization and transcriptional properties of a tRNALeuCUG gene from *Drosophila melanogaster*. Gene 44:307–314

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Gotta M, Ahringer J (2001) Distinct roles for Gα and Gβγ in regulating spindle position and orientation in *Caenorhabditis elegans* embryos. Nature Cell Biol 3:297–300

Göttgens B, Barton L, Gilbert J, Bench A, Sanchez M, Bahn S, Mistry S, Grafham D, McMurray A, Vaudin M, Amaya E, Bentley D, Green A, Sinclair A (2000) Analysis of vertebrate SCL loci identifies conserved enhancers. Nature Biotechnol 18:181–186

Graur D, Li W-H (2000) Fundamentals of molecular evolution, 2nd edn. Sinauer, Sunderland

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

Hare M, Palumbi S (2003) High intron sequence conservation across three mammalian orders suggests functional constraints. Mol Biol Evol 20:969–978

Hirsh A, Fraser H (2001) Protein dispensability and rate of evolution. Nature 411:1046–1049

Hurst L, Smith N (1999) Do essential genes evolve slowly? Curr Biol 9:747–750

Jansen G, Thijssen KL, Werner P, van der Horst M, Hazendonk E, Plasterk RH (1999) The complete family of genes encoding G proteins of *Caenorhabditis elegans*. Nat Genet 21:414–419

Jansen G, Weinkove D, Plasterk R (2002) The G-protein gamma subunit *gpc-1* of the nematode *C. elegans* is involved in taste adaptation. EMBO J 21:986–994

Jiang L, Sternberg P (1999) An HMG1-like protein facilitates Wnt signaling in *Caenorhabditis elegans*. Genes Dev 13:877–889

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

Jukes T, Cantor C (1969) Evolution of protein molecules. In: Munro H (ed) Mammalian protein metabolism. Academic, New York, pp 21–132

Jukes TH, Osawa S (1993) Evolutionary changes in the genetic code. Comp Biochem Phys B 106:489–494

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

Korswagen H, Park J, Ohshima Y, Plasterk R (1997) An activating mutation in a *Caenorhabditis elegans* Gs protein induces neural degeneration. Genes Dev 11:1493–1503

Kumar S, Tamura K, Jakobsen I, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Lackner MR, Nurrish SJ, Kaplan JM (1999) Facilitation of synaptic transmission by EGL-30 Gqα and EGL-8 PLCβ:DAG binding to UNC-13 is required to stimulate acetylcholine release. Neuron 24:335–346

Lander ES et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

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

Liu KS, Sternberg PW (1995) Sensory regulation of male mating behavior in *Caenorhabditis elegans*. Neuron 14:79–89

Loots G, Locksley R, Blankespoor C, Wang Z, Miller W, Rubin E, Frazer K (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288:136–140

Lowe T, Eddy S (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Mendel J, Korswagen H, Liu K, Hajdu-Cronin Y, Simon M, Plasterk R, Sternberg P (1995) Participation of the protein Go in multiple aspects of behavior in *C. elegans*. Science 267:1652–1655

Milanesi L, D'Angelo D, Rogozin IB (1999) GeneBuilder: interactive in silico prediction of gene structure. Bioinformatics 15:612–621

Moghal N, Garcia L, Khan L, Iwasaki K, Sternberg P (2003) Modulation of EGF receptor-mediated vulva development by the heterotrimeric G-protein Galphaq and excitable cells in *C. elegans*. Development 130:4553–4566

Nathoo A, Moeller R, Westlund B, Hart A (2001) Identification of neuropeptide-like protein gene families in *Caenorhabditis elegans* and other species. Proc Natl Acad Sci USA 98:14000–14005

Neer E (1995) Heterotrimeric G proteins: organizers of transmembrane signals. Cell 80:249–257

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nurrish S, Ségalat L, Kaplan J (1999) Serotonin inhibition of synaptic transmission: Gα(0) decreases the abundance of UNC-13 at release sites. Neuron 24:231–242

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264

Park J, Ohshima S, Tani T, Ohshima Y (1997) Structure and expression of the *gsa-1* gene encoding a G protein alpha(s) subunit in *C. elegans*. Gene 194:183–190

Riddle D, Albert P (1997) Genetic and environmental regulation of dauer larva development. In: Riddle D, Blumenthal T, Meyer B, Priess J (eds) *C. elegans* II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 739–768

Roayaie K, Crump JG, Sagasti A, Bargmann CI (1998) The G alpha protein ODR-3 mediates olfactory and nociceptive function and controls cilium morphogenesis in *C. elegans* olfactory neurons. Neuron 20:55–67

Robertson HM (1998) Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res 8:449–463

Robertson HM (2000) The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. Genome Res 10:192–203

Schulze E, Altmann M, Adham I, Schulze B, Frode S, Engel W (2003) The maintenance of neuromuscular function requires UBC-25 in *Caenorhabditis elegans*. Biochem Biophys Res Comm 305:691–699

Ségalat L, Elkes D, Kaplan J (1995) Modulation of serotonin-controlled behaviors by Go in *Caenorhabditis elegans*. Science 267:1648–1651

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

Smith NG, Eyre-Walker A (2002) Adaptive protein evolution in Drosophila. Nature 415:1022–1024

Stein L et al (2003) The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol 1:166–192

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

Thomas J et al (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424:788–793

Troemel ER (1999) Chemosensory signaling in *C. elegans*. Bioessays 21:1011–1020

Van der Linden AM, Simmer F, Cuppen E, Plasterk RH (2001) The G-protein beta-subunit GPB-2 in *Caenorhabditis elegans* regulates the G(o)alpha-G(q)alpha signaling network through interactions with the regulator of G-protein signaling proteins EGL-10 and EAT-16. Genetics 158:221–235

Van der Linden A, Moorman C, Cuppen E, Korswagen H, Plasterk R (2003) Hyperactivation of the G12-mediated signaling pathway in *Caenorhabditis elegans* induces a developmental growth arrest via protein kinase C. Curr Biol 13:516–521

Venter JC et al (2001) The sequence of the human genome. Science 291:1304–1351

Wilson A, Carlson S, White T (1977) Biochemical evolution. Annu Rev Biochem46:573–639

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

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

Yau DM, Yokoyama N, Goshima Y, Siddiqui ZK, Siddiqui SS, Kozasa T (2003) Identification and molecular characterization of the Gα12-ρguanine nucleotide exchange factor pathway in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 100:14748–14753

Zdobnov EM et al (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science 298:149–159

Zwaal R, Broeks A, van Meurs J, Groenen J, Plasterk R (1993) Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. Proc Natl Acad Sci USA 90:7431–7435

Zwaal R, Ahringer J, van Luenen H, Rushforth A, Anderson P, Plasterk R (1996) G proteins are required for spatial orientation of early cell cleavages in *C. elegans* embryos. Cell 86:619–629

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