J Mol Evol (2006) 63:87–94
DOI: 10.1007/s00239-005-0215-z

JOURNAL OF MOLECULAR EVOLUTION

The Molecular Phylogeny of a Nematode-Specific Clade of Heterotrimeric G-Protein α-Subunit Genes

Damien M. O'Halloran,*† David A. Fitzpatrick,* Grace P. McCormack,‡ James O. McInerney, Ann M. Burnell

Biology Department, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland

Received: 9 September 2005 / Accepted: 1 March 2006 [*Reviewing Editor:* Dr. John Oakeshott]

Abstract. In animal olfactory systems, odorant molecules are detected by olfactory receptors (ORs). ORs are part of the G-protein-coupled receptor (GPCR) superfamily. Heterotrimeric guanine nucleotide binding G-proteins (G-proteins) relay signals from GPCRs to intracellular effectors. G-proteins are comprised of three peptides. The G-protein α subunit confers functional specificity to G-proteins. Vertebrate and insect Gα-subunit genes are divided into four subfamilies based on functional and sequence attributes. The nematode *Caenorhabditis elegans* contains 21 Gα genes, 14 of which are exclusively expressed in sensory neurons. Most individual mammalian cells express multiple distinct GPCR gene products, however, individual mammalian and insect olfactory neurons express only one functional odorant OR. By contrast *C. elegans* expresses multiple ORs and multiple Gα subunits within each olfactory neuron. Here we show that, in addition to having at least one member of each of the four mammalian Gα gene classes, *C. elegans* and other nematodes also possess two lineage-specific Gα gene expansions, homologues of which are not found in any other organisms examined. We hypothesize that these novel nematode-specific Gα genes increase the functional complexity of individual chemosensory neurons, enabling them to integrate odor signals from the multiple distinct ORs expressed on their membranes. This neuronal gene expansion most likely occurred in nematodes to enable them to compensate for the small number of chemosensory cells and the limited emphasis on cephalization during nematode evolution.

Key words: Heterotrimeric G-protein α subunit — Olfaction — Nematode — *Caenorhabditis elegans* — Chemoreception

Introduction

Of a total of 302 neurons in the adult *Caenorhabditis elegans* hermaphrodite, just 32 mediate chemosensory function and three pairs of these chemosensory neurons are primarily olfactory (Bargmann 1998; Bargmann et al. 1993; Prasad and Reed 1999). Despite this, *C. elegans* can still discriminate between hundreds of odorants at various concentrations (Bargmann et al. 1993). Mammals utilize over 10 million olfactory neurons to distinguish between a substantially larger range of molecules (Prasad and Reed 1999). Olfactory receptor (ORs) are a subfamily of the G-protein-coupled receptor (GPCR) superfamily and comprise the largest multigene family in mammals comprising over 1000 genes in the mouse (Zhang and Firestein 2002). In mammals each olfactory neuron expresses only one functional OR and this "one olfactory receptor/one neuron" rule is necessary for the conversion of olfactory signals into an accurate topographical map in the

*These authors contributed equally to this work.
†*Present address*: UC Davis Center for Neuroscience, 1544 Newton Court Davis, CA 95616, USA
‡*Present address*: Department of Zoology, National University of Ireland Galway, Galway, Ireland
Correspondence to: Ann M. Burnell; *email:* ann.burnell@nuim.ie

olfactory bulb (Serizawa et al. 2003). In *Drosophila melanogaster* also, each olfactory neuron expresses only one OR and these neurons project to the antennal lobe, the equivalent of the vertebrate olfactory bulb (Vosshall et al. 1999). However, a recent exception to this rule has been reported in *D. melanogaster* by Goldman et al. (2005). The *C. elegans* genome contains circa 550 functional ORs (Robertson 1998) and individual *C. elegans* chemosensory neurons contain multiple ORs (Sengupta et al. 1996; Troemel et al. 1995).

The downstream effectors of ORs are a family of guanine nucleotide binding proteins (G-proteins) that relay signals from seven transmembrane ORs to intracellular proteins. G-proteins have a common design comprising three polypeptide chains: an α subunit that binds and hydrolyzes guanosine triphosphate (GTP) and a $\beta\gamma$ dimer that serves as a functional monomer (Simon et al 1991). When a G-protein is activated by interaction with a receptor, the α subunit exchanges bound GDP for GTP (Supplementary Fig. 1). An intrinsic GTPase activity of the α subunit restores it to the basal state in which GDP is bound. G-proteins are conserved in animals separated by considerable evolutionary distances, such as mammals and protistans. Gα proteins have been previously divided on the basis of amino acid (aa) similarity into four classes; $G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_q$, and $G\alpha_{12}$ (Simon et al. 1991; Strathmann and Simon 1990, 1991). The human genome contains 17 Gα genes which are grouped into the four classes listed above, based on functional and sequence attributes (Simon et al. 1991). No additional Gα genes were detected when the human genome was sequenced. *C. elegans* contains 21 Gα genes, 14 of which are specifically expressed in sensory neurons (Jansen et al. 1999; Cuppen et al. 2003).

In *C. elegans*, chemotaxis to volatile chemoattractants is mediated primarily by two pairs of neurons, AWA and AWC (Bargmann et al. 1993). Both AWA and AWC neurons have been shown to express numerous ORs including the diacetyl receptor, ODR-10 (Sengupta et al. 1996; Troemel et al. 1995), and at least four Gα subunits (Jansen et al. 1999). AWC-mediated olfaction involves the Gα subunits, ODR-3, GPA-2, GPA-3, and GPA-13 (Lans et al. 2004; Roayaie et al. 1998), which, upon activation, induce an increase in intracellular cGMP, mediated by the guanylyl cyclases ODR-1 (L'Etoile and Bargmann 2000) and DAF-11 (Birnby et al. 2000). This in turn leads to the activation of a cyclic nucleotide gated channel encoded by the *tax-2* and *tax-4* genes (Coburn and Bargmann 1996) (Supplementary Fig. 1). The TAX-2/TAX-4 channel is similar to vertebrate visual and olfactory channels and permits the entry of Ca^{2+} ions into the cell. AWA neurons have been shown to express the Gα subunits ODR-3, GPA-3, GPA-5, and GPA-6 (Lans et al. 2004), however, unlike AWC-mediated olfaction, AWA-mediated olfaction is generated by an alternative pathway that requires a novel channel encoded by the *osm-9*, *ocr-1*, and/or *ocr-2* genes (Colbert et al. 1997; Tobin et al. 2002) (Supplementary Fig. 1). These subunits are distantly related to the *Drosophila* phototransduction channel TRP (Transient Receptor Potential), which is regulated by G-protein signaling.

A phylogenetic analysis of the human G-protein α-subunit gene family indicated the existence of four Gα gene classes (G_q, G_{12}, G_s, and $G_{i/o}$), each of which contains a number of closely related isotypes (Simon et al. 1991). These four Gα-subunit classes transduce signals from a great variety of extracellular agents including hormones, neurotransmitters, chemokines, and peptides. Signaling from the extracellular environment via GPCRs is phylogenetically ancient being found in protistans, fungi, plants, and animals. The existence of one or more representatives of each of these four Gα-subunit classes among the seven Gα-subunit genes cloned from the sponge *Ephydatia fluviatilis* indicates that these Gα gene classes existed prior to the parazoan eumetazoan split (Suga et al. 1999). A dendrogram of the predicted aa sequences of all the *C. elegans* Gα subunits with a representative sequence of each of the four human Gα-subunit genes identified at least one clear homologue of each of the four classes of vertebrate genes, the remaining nematode genes being more divergent (Jansen et al. 1999). To gain further insight into the phylogeny and diversity of the Gα gene family in nematodes we assembled a data set containing homologues of putative Gα genes from a variety of protistans, metazoans, fungi, and plants. Our phylogenetic analysis shows that in addition to having at least one member of each of the four mammalian Gα gene classes, *C. elegans* also possesses two lineage-specific Gα gene expansions, homologues of which are not found in other organisms. We hypothesize that these novel nematode-specific Gα genes increase the functional complexity of individual chemosensory neurons, enabling them to integrate and adapt to odor signals from the multiple distinct ORs expressed on their membranes.

Materials and Methods

Gα protein homologues were located by performing multiple BLASTP (Altschul et al. 1997) searches with a cutoff expectation value (E-value) of 10^{-7} against GenBank (http://www.ncbi.nlm.nih.gov/BLAST). In each case putative *C. elegans* Gα proteins were used as the query sequence. In total, 118 proteins from many diverse genera were located (see supplementary information). Sequences were aligned using ClustalW 1.81 (Thompson et al. 1994) using the default settings. All alignments were corrected for obvious alignment ambiguity using the alignment editor Se-Al 2.0a11. Alignment can be viewed at http://biology.nuim.ie//staff/JMESupp.shtml.

Gene Tree Construction

A brief examination of the sequence data revealed that average sequence similarity was approximately 40% (a similarity matrix can be viewed at http://biology.nuim.ie//staff/JMESupp.shtml). To account for the problems associated with low levels of sequence similarity, the aa alignment was recoded into the six Dayhoff (1978) groups: C, STPAG, NDEQ, HRK, MILV, and FWY. This is based on the premise that aa substitutions within the six groups will be common and noisy, whereas changes between groups will be rarer and so have less saturation (Hrdy et al. 2004). The recoded alignment was analyzed using the Bayesian criterion implemented in MRBAYES v3.0B4 (Huelsenbeck and Ronquist 2001). The model used has a 6×6 general time-reversible rate matrix. Among-site variation was modeled with a free proportion of invariable sites and a four-category discrete gamma distribution. The analysis used a Markov chain Monte Carlo (MCMC) chain that ran for 15 million generations, sampled every 100th generation. Plots of likelihood versus generation revealed that all chains reached stationarity after 73,500 generations, therefore these trees were discarded as a *burnin*. Bayesian posterior probability branch supports were determined using the *sumt* command of MRBAYES.

According to PUZZLE 5.1 (Schmidt et al. 2002), the initial protein alignment fails the chi-square square test for homogenous aa composition at $p = 0.95$. Furthermore, our initial phylogenetic reconstructions appeared to lack coherent phylogenetic signal as most groupings were poorly supported. Therefore, highly variable sites were removed from the alignment so that the largest possible alignment in which all sequences passed the chi-square test for homogeneity of aa composition could be recovered. This second phylogenetic analysis was based on a smaller number (47) of exemplar Gα proteins. Sites were categorized into different classes using a method implemented in Tree-Puzzle 5.1. The method assumes that there are eight categories of sites. Rate variation across these sites is assumed to follow a discrete gamma distribution. The fastest-evolving sites were found in category 8, these were removed until an alignment of 399 aa positions that passed the chi-square test for homogeneity of aa composition was recovered for the Gα proteins dataset. The most appropriate protein model was selected using the software program MODELGENERATOR (http://bioinf.nuim.ie/software/modelgenerator). One hundred bootstrap replicates were performed with the appropriate protein model, using the software program PHYML (Guindon and Gascuel 2003). The results of this analysis were summarized using the majority-rule consensus method.

EST Database Searches

Using each nematode-specific gene as a query sequence, we performed exhaustive TBLASTN (Altschul et al. 1997) database searches with a cutoff expectation value of 10^{-7} against the nematode EST database NEMBASE (Parkinson et al. 2004). The version of NEMBASE used contained 130,184 clustered ESTs from 37 different nematode species from the four major nematode clades. All statistically significant EST sequence hits were extracted and subsequently searched locally against the *C. elegans* proteome (ftp://ftp.ensembl.org/pub/current_celegans/data/fasta/pep/) using BLASTX with a cutoff expectation of 10^{-7}. Significant hits were confirmed by manual inspection of BLAST alignments. The purpose of this approach was to confirm orthology between the nematode EST sequences and the *C. elegans* protein sequences. The presence or absence of nematode-specific genes within the 37 species found in NEMBASE was noted.

Using the same methodology as above, nematode-specific genes were used to search the Schistosome (http://www.ebi.ac.uk/blast2/parasites.html) and Tardigrade (http://zeldia.cap.ed.ac.uk/Tardibase/tardibase/tardigrades.html) EST databases. No orthologues were found for the G_{ns} group of Gα-subunit genes in these additional database searches.

Results

The final alignment contained 118 taxa. There is a high degree of sequence dissimilarity between the Gα-subunit genes. This resulted in a poorly supported tree, making it difficult to infer phylogenetic relationships. We therefore recoded each aa into one of the six groups of chemically related aa that commonly replace one another (Dayhoff 1978). This recoding technique has the effect of homogenizing the aa composition between sequences and shortening long branches (Hrdy et al. 2004) and resulted in a more robust phylogenetic hypothesis. Our phylogenetic reconstruction (Fig. 1) shows that the plant Gα genes form a robust monophyletic grouping. The G_q, G_{12}, and G_s classes are derived from a common ancestor, being grouped together with strong support. The $G_{i/o}$ class forms an adjacent group to the G_q, G_{12}, and G_s clade, as does the fungal Gα gene group. Each of the four described classes contains at least one *C. elegans* Gα gene, viz., G_q (*egl-30*), G_s (*gsa-1*), G_{12} (*gpa-12*), and $G_{i/o}$ (*goa-1*, *gpa-4*, and *gpa-16*). The remaining 15 *C. elegans* Gα genes are located outside these four classes. Our analysis indicates that the majority of the 15 *C. elegans* genes form a distinct sister clade to the $G_{i/o}$ Gα gene class, as they are grouped beside it with high support. The remaining *C. elegans* Gα genes are grouped among the protistan (*gpa-6*, *gpa-11, gpa13*, and *gpa-14*) and the fungal (*gpa-5*) Gα genes. These genes are heterogeneous and highly divergent in their aa sequences. Therefore their inferred positions within the tree may be erroneous on account of high levels of sequence divergence and not relatedness. The *Trichomonas vaginalis* genes form a distinct but divergent clade at the base of the tree, being grouped together with strong support. The *Dictyostelium discoideum* Gα genes form three small sister clades at the base of the tree, beside the *T. vaginalis* clade and the divergent nematode genes. One *D. discoideum* gene (*gpa-7*) was positioned beside *gpa-17*, a very divergent *C. elegans* Gα gene with an extremely long branch. Because of the extreme heterogeneity of these *D. discoideum* genes, their inferred positions within the tree are not reliable. A secondary phylogenetic analysis using the maximum likelihood criterion was carried out on a reduced sample of our data. Heterogeneous positions within the data were removed until the data passed a chi-square test for aa homogeneity (Fig. 2). This analysis yielded a comparable phylogeny to the full data set. The only difference in this analysis was the grouping-together of the divergent *C. elegans* genes *gpa-5* and *gpa-17*.

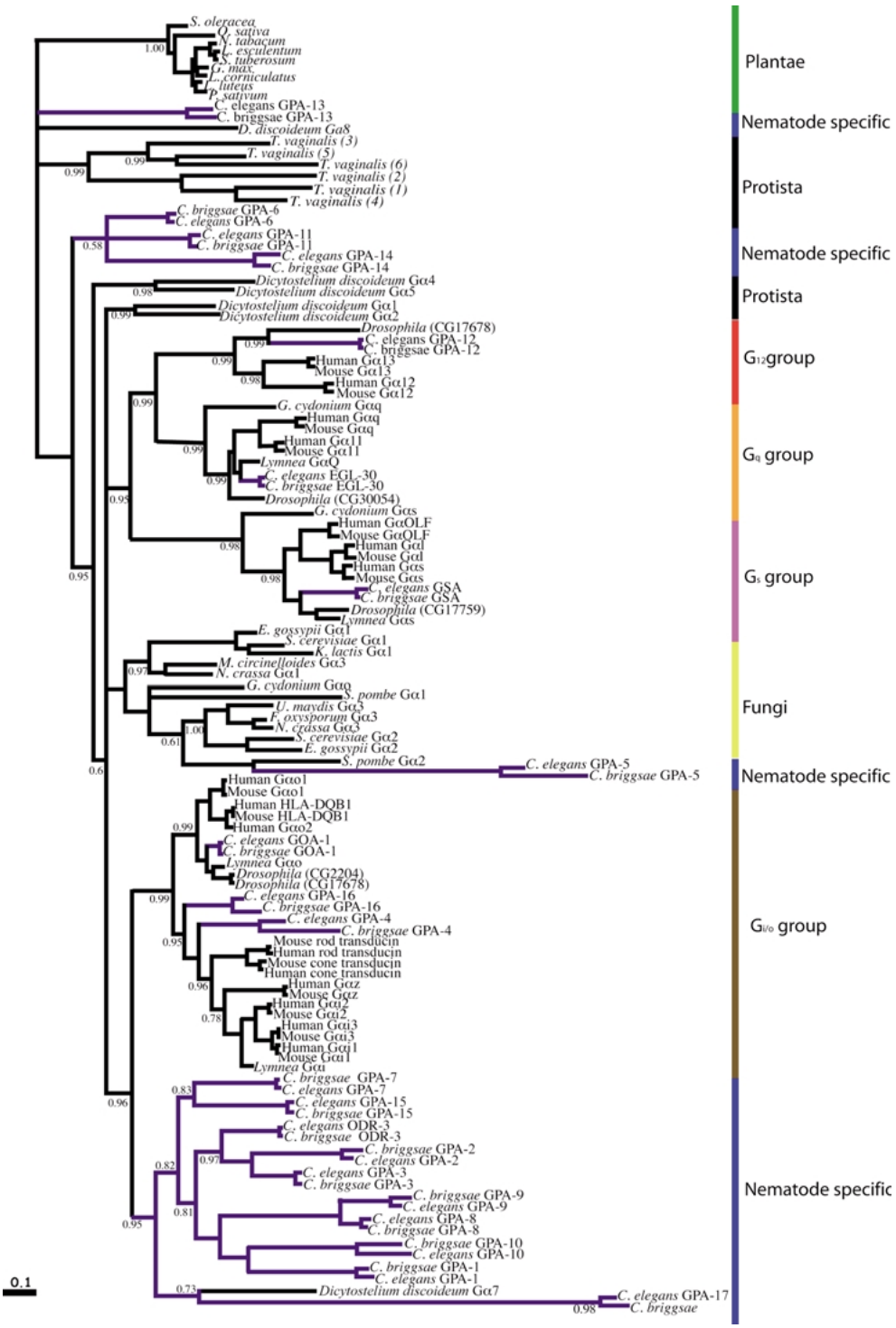

Fig. 1. Phylogenetic analysis of the G-protein α-subunit protein family. The inferred phylogeny shows a Bayesian consensus tree of a general time-reversible substitution matrix estimated from G-protein α aa sequences recoded into the six Dayhoff (1978) groups. Among-site rate heterogeneity was modeled using a four-category gamma correction with a fraction of invariant sites. The analysis used the Metropolis-coupled MCMC strategy from MrBayes (Huelsenbeck and Ronquist 2001) and parameters such as the composition and substitution rate matrix were free. Posterior probabilities for selected branches are shown at nodes. The scale bar indicates the number of changes per site. This methodology is appropriate for a phylogenetic analysis of divergent genes (Hrdy et al. 2004).

During our searches of GenBank we failed to find orthologues for the majority of the *C. elegans* Gα genes (orthologues were found only for G_s, G_q, G_{12}, and $G_{i/o}$ Gα gene classes). A database search of wormbase (www.wormbase.net) revealed that *C. briggsae* contains homologues for all the Gα proteins found in *C.elegans*. As a result of our database searches and subsequent phylogenetic analysis we propose that the majority of the *C. elegans* genes represent unique gene expansions which are confined

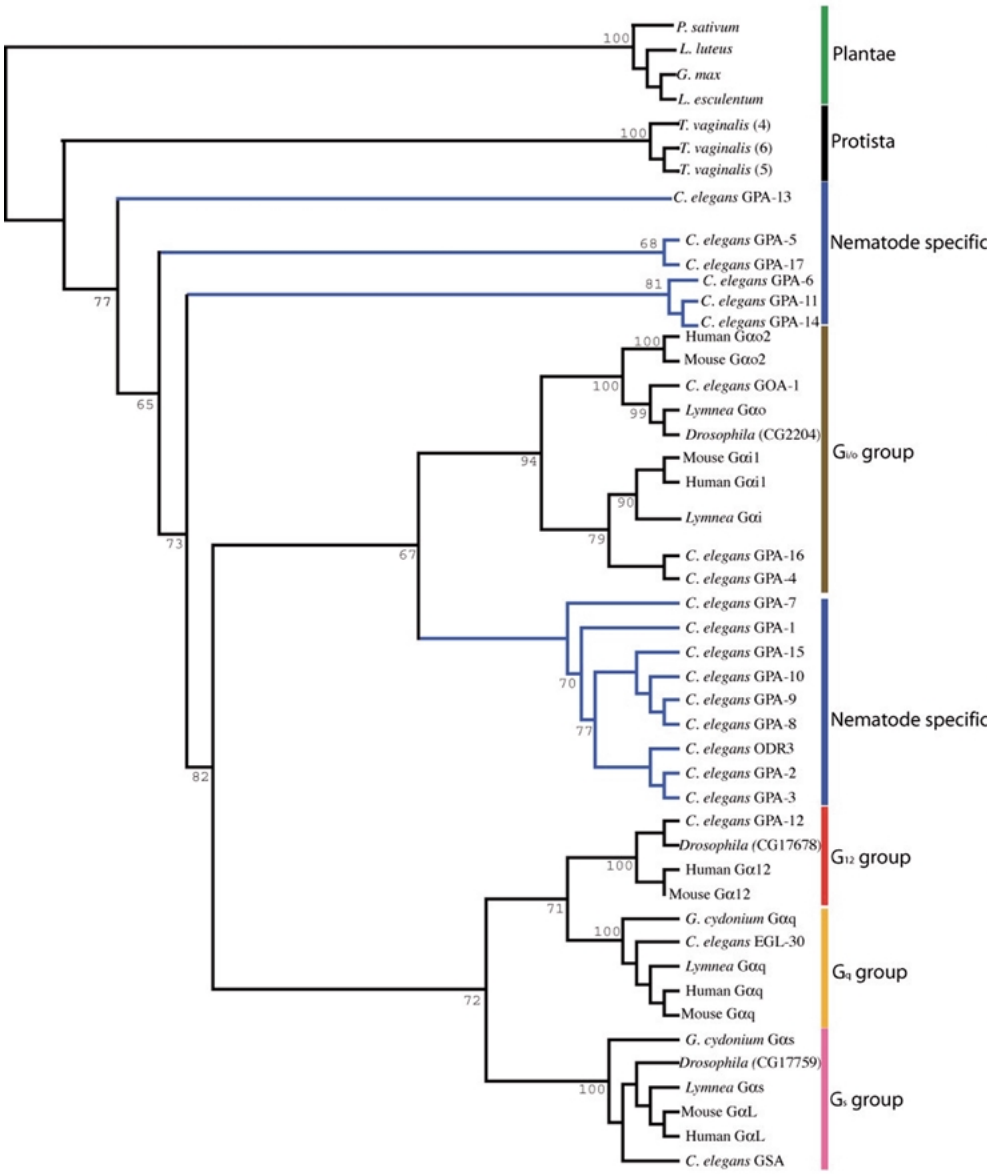

Fig. 2. Maximum likelihood phylogenetic tree of exemplar G-protein α-subunit proteins. Highly variable sites have been removed and the resultant alignment (399 aa positions) passes a chi-square test for homogeneous aa composition at $p = 0.95$. Branch supports were determined using the bootstrap resampling technique.

to the phylum Nematoda. Therefore, we refer to these genes as putative Gα nematode-specific (G_{ns}) genes. To test this proposal we examined all available nematode sequence data. EST data are available for the four major nematode clades (Parkinson et al. 2004). The NEMBASE database contains many EST homologues of the four described Gα gene classes (G_s, G_q, G_{12}, $G_{i/o}$), but in addition, we also detected six of the G_{ns} genes in this database. Although we could only locate a small number (typically two or less) of representative G_{ns} genes in individual nematodes in NEMBASE, these G_{ns} EST homologues occurred in diverse members from each of the nematode clades represented in the database (Table 1). This finding suggests that the evolution of the G_{ns} genes predates the origin and adaptive radiation of nematodes. The small number of G_{ns} ESTs detected in individual nematodes might indicate that the full expansion of the G_{ns} genes is restricted to *Caenorhabditis*. A more parsimonious explanation may be that the EST database does not contain sufficient sequence data. We propose that the diversity of G_{ns} gene homologues detected in the NEMBASE database and their distribution pattern across the major clades within the phylum Nematoda support the premise that the expansion of G_{ns} is common to all members of this phylum. At present a number of projects are under way to sequence additional nematode genomes (e.g., *Brugia malayi, Haemonchus contortus, Heterorhabditis bacteriophora, Meloidogyne hapla, Pristionchus pacificus, Trichinella spiralis*). When published, these genome sequences will help

Table 1. Orthologues of *C. elegans* nematode-specific Gα genes in the four major clades of the phylum Nematoda identified from the nematode EST database NEMBASE (Parkinson et al. 2004)

Clade	Species	Gα ns gene	Accession No.
I	*Trichinella spiralis*	*gpa-7*	TSC01682
III	*Ascaris suum*	*gpa-13*[a]	ASC02040
		gpa-17[a]	ASC18830
	Brugia malayi	*gpa-3*	BMC08478
	Wuchereria bancrofti	*gpa-3*	WBC01262
IVa	*Strongyloides ratti*	*gpa-7*	SRC01861
IVb	*Globodera rostochiensis*	*gpa-7*	GRC02175
	Heterodera glycines	*gpa-3*	HGC11040
		gpa-7	HGC07601
	Meloidogyne arenaria	*gpa-5*[a]	MAC00743
	Meloidogyne incognita	*gpa-7*	MIC03287
	Meloidogyne javanica	*gpa-3*	MJC02541
		gpa-7	MJC03734
V	*Haemonchus contortus*	*gpa-7*	HCC10406
	Ostertagia ostertagi	*gpa-7*	OOC00603

[a]Divergent G_{ns} genes (see Table 2).

elucidate the origins and expansion of the G_{ns} genes. An alternative and less parsimonious hypothesis to the nematode-specific origin and expansion of the G_{ns} clade is that orthologues of the nematode-specific Gα genes may have been lost in the vertebrate and arthropod lineages.

Gα subunits consist of two domains: a guanine nucleotide binding GTPase domain with a high structural similarity to Ras GTPases and a compact helical domain. The GTPase domain consists of five α-helices surrounding a six-stranded β-sheet (Rens-Domiano and Hamm 1995). The five α-helices, designated G1 to G5 form the guanine nucleotide binding site, and each helix contains a conserved nucleotide binding motif. Our phylogenetic analyzes infer that the majority of the G_{ns} genes are most closely related to members of the $G_{i/o}$ group. The G_{ns} subunits show strong conservation of the G1–G5 nucleotide binding motifs with those found in the $G_{i/o}$ group, particularly the G1, G3 and G4 motifs (Table 2). The G2 and G5 motifs show slight deviations from the $G_{i/o}$ consensus, but all the critical GTP binding contact residues in these motifs are conserved, implying a conservation of Gα protein function. The divergent nematode sequences *gpa-5*, *gpa-6*, *gpa-11, gpa-13, gpa-14*, and *gpa-17*, which are located outside the main Gα gene groupings at the base of the tree, also retain all the critical GTP binding contact residues in the G1–G4 motifs, but the G5 motif is more variable in these divergent G_{ns} genes. Overall aa sequence within the functionally important G1 to G5 motifs is highly conserved (Table 2). However, the aa sequence outside these motifs reveals high levels of divergence. This may have a functional significance, possibly enabling the G_{ns} to couple to novel members of the GPCR superfamily, which has also undergone a major gene expansion in *C. elegans* and *C. briggsae* (Robertson 1998).

Discussion

Most individual mammalian cells express multiple (< 10) distinct GPCR gene products (Gomperts et al. 2002). The distinct GPCRs expressed on individual cell membranes relay signals from a great variety of extracellular agents, often with antagonistic effects, to intracellular secondary messengers and effectors via G-protein activation. $G\alpha_o$ is the most abundantly expressed Gα subunit in mammalian neurons. $G\alpha_s$ $G\alpha_{i2,}$ $G\alpha_{12}$, and $G\alpha_q$, are also expressed ubiquitously in most mammalian cells (Gomperts et al. 2002). The *C. elegans* homologues of these four genes (*gsa-1*, *goa-1*, *gpa-12*, and *egl-30*, respectively) are also widely expressed in nerve and muscle cells, and some are also expressed in the excretory cells (*egl-30*) and the ventral hypodermis (*gpa-12*) (Jansen et al. 1999). In mammals and insects each olfactory neuron typically expresses only one functional OR and the brain discriminates between odors by determining which neurons have been activated (Prasad and Reed 1999). The extent to which similar olfactory coding systems have evolved in other invertebrate phyla is still unknown. However, this "one olfactory receptor/one neuron" design does not occur in *C. elegans*, reflecting the small number of chemosensory cells in the nematode nervous system and the limited emphasis on cephalization during nematode evolution.

C. elegans olfactory neurons express multiple classes of ORs and Gα subunits within a single cell (Jansen et al. 1999; Lans et al. 2004; Troemel et al. 1995). Jansen et al. (1999) have proposed that the presence of multiple signal transduction pathways in indivdual *C. elegans* olfactory neurons most probably evolved to compensate for the small number of chemosensory neurons in the nematode's nervous system. The G_{ns} subunit ODR-3 is the major activator of the signaling pathway in the two pairs of sensory neurons, AWA and AWC, which mediate odorant attraction in *C. elegans*. ODR-3 is also required for osmosensation and mechanosensation (Roayaie et al. 1998), suggesting that it can couple to several distinct ORs. In addition to ODR-3, the AWA neurons also express GPA-3, which contributes a second stimulatory signal, and GPA-5, which has inhibitory effects (Lans et al. 2004), while the AWC neurons express GPA-13, which is stimulatory, and GPA-2, which has inhibitory effects (Lans et al. 2004) (Supplementary Fig. 1). Thus the G_{ns} subunits form a regulatory network within the AWA and AWC cells which modulates the olfactory response to various odors. This integrative, cell-autonomous response is similar to the G-protein-coupled response systems in nonolfactory cells in mammals and other animals. Cell-autonomous integration of G-protein signaling represents the ancestral condition in the protistans, sponges, and eumetazoans and appears to

Table 2. Conserved nucleotide binding motifs in nematode heterotrimeric Gα subunit proteins

Gα group	Protein	G-1	G-2	G-3	G-4	G-5
G_{ns}	ODR-3	GAGECGKS	DILYSRVATTGV	FDVGGQR	FMNKKD	MHETCAT
	GPA-1	GAGESGKS	DILHTRVPTMGV	FDVGGQR	FLNKID	CHHTCAT
	GPA-2	GAGECGKS	DTLLLRTKTTGI	FDVGGQR	FLNKKD	VHETCAT
	GPA-3	GAGECGKS	DILLSRIKTTGI	FDVGGQR	FLNKKD	MHQTCAT
	GPA-7	GAGESGKS	DLLRTRIKTTGI	IDVGGQR	FLNKKD	CHHTCAT
	GPA-15	GTGECGKS	DMLRIRIPTMGV	YDVGGQR	FLNKRD	THVTCAT
	GPA-8	GPGESGKS	DILKSRVPTSGV	FDVGGQR	FLNKID	EHVTCAT
	GPA-9	GPGESGKS	DILKTRVPTLGI	FDIGGQR	FLNKID	QHVTSAT
	GPA-10	GPGESGKS	DIVHIRVPTTGV	CDCGGQR	FLNKID	THETCAI
$G_{i/o}$	GOA-1	GAGESGKS	DILRTRVKTTGI	FDVGGQR	FLNKKD	CHMTCAT
	GPA-16	GAGESGKS	DILRTRIKTTGI	FDVGGQR	FLNKKD	TQFTCAT
	GPA-4	GAGESGKS	DILRARVKSTGI	FDVGGQR	FLNKMD	SHFTCAT
G_q	EGL-30	GTGESGKS	DILRVRVPTTGI	VDVGGQR	FLNKKD	SHFTCAT
G_s	GSA-1	GAGESGKS	DILRCRVMTTGI	FDVGGQR	FLNKQD	PHFTCAV
G_{12}	GPA-12	GSGESGKS	DILFCRKATRGI	IDVGGQR	FMNKND	YHFTTAV
	GPA-5	GVTDSGKS	DLIHMRQTTLGV	IDVGGQK	FLNKKD	MKYTQAT
	GPA-6	GTAESGKS	DIVHCRISTTGI	VDVGGQR	FLNKYD	VFETTAT
G_{nsd}	GPA-11	GGPECGKS	DVLRARVPTTGI	VDVGGQR	FLNKID	THITNAT
	GPA-13	GSAESGKT	DLIMAYVPTCGV	FDIGGQK	FLNEID	VYRCIAI
	GPA-14	GGPLSGKS	DIVHSRKATMSI	IDVGGQR	FFNKVD	PHFTTAT
	GPA-17	GIEGAGKT	DIVHVRKPTVSF	HDMGGQK	FFNKQD	MKYTQAT

Note. G_{ns}, nematode-specific Gα subunit; G_{nsd}, divergent nematode-specific Gα subunit. The conserved residues are as follows: G1, GXXXXGK(S/T) (Sprang 1997), which binds to the α and β phosphates of the guanine nucleotide (Rens-Domiano and Hamm 1995; Sprang 1997); G2, DXXXXRXXT (Bourne et al. 1991), which is also involved in binding to the α and β phosphates of the guanine nucleotide (Rens-Domiano and Hamm 1995) and in binding Mg^{2+} (Rens-Domiano and Hamm 1995; Sprang 1997); G3, DXXG (Bourne et al. 1991), which binds Mg^{2+} (Rens-Domiano and Hamm 1995; Sprang 1997); G4, NKXD (Bourne et al. 1991; Sprang 1997), which recognizes the guanine ring; and G5, (T/G)(C/S)A (Sprang 1997), which "buttresses the guanine base recognition site" (Sprang 1997).

have been retained in all cellular systems apart from olfaction. Our results indicate that *C. elegans* and most likely all nematodes have evolved at least two unique lineage-specific expansions of Gα-subunit genes. With the exception of *gpa-7*, which is widely expressed in neuron and muscle cells, the remaining 14 G_{ns} genes are expressed in other *C. elegans* amphid sensory neurons and/or other putative sensory neurons (Jansen et al. 1999). Thus it appears that nematodes, like other animals, have increased their olfactory repertoire by expanding the number of ORs, but instead of a concomitant increase in the number of sensory neurons, nematodes have retained the ancestral multireceptor G-protein signaling system in their olfactory neurons. The integrative capacity of the chemosensory system was, however, further refined by the recruitment of additional Gα-subunit genes.

The genomes of *C. elegans* and *C. briggsae* contain an unexpectedly large number of genes (Hodgkin 2001; Stein et al. 2003). This may have resulted in part from the relative lack of alternative splicing and domain accretion in nematode genes (Hodgkin 2001), but lineage-specific gene expansions have also occurred in nematodes (Lespinet et al. 2002; Robinson-Rechavi et al. 2005). These gene expansions are particularly noticeable for neuronal genes (Bargmann 1998). For example, the largest and most diverse nicotinic acetylcholine receptor gene family is that of *C. elegans* (Mongan et al. 1998); novel families of potassium channels have been identified in *C. elegans* (Wei et al. 1996), and G-protein coupled (putative) chemoreceptor genes comprise the largest gene family in *C. elegans* (Robertson 1998). Thus the relative lack of anatomical complexity in the nematode nervous system appears to have been compensated during nematode evolution by an increased functional complexity and multitasking capacity of individual nerve cells, as is clearly illustrated in the olfactory system of *C. elegans*.

Acknowledgments. This work was funded by the Irish Higher Education Authority Programme for Research in Third Level. Many thanks go to Dr. Ralf Schmid for sending us the clustered version of NEMBASE. The authors wish to acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. Supplementary material can be found at http://biology.nuim.ie/staff/JMESupp.shtml.

References

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. Nucleic Acids Res 25:3389–3402

Bargmann CI (1998) Neurobiology of the *Caenorhabditis elegans* genome. Science 282:2028–2033

Bargmann CI, Hartwieg E, Horvitz HR (1993) Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* 74:515–527

Birnby DA, Link EM, Vowels JJ, Tian H, Colacurcio PL, Thomas JH (2000) A transmembrane guanylyl cyclase (DAF-11) and Hsp90 (DAF-21) regulate a common set of chemosensory behaviors in *caenorhabditis elegans*. Genetics 155:85–104

Bourne HR, Sanders DA, McCormick F (1991) The GTPase superfamily: conserved structure and molecular mechanism. Nature 349:117–127

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

Colbert HA, Smith TL, Bargmann CI (1997) OSM-9, a novel protein with structural similarity to channels, is required for olfaction, mechanosensation, and olfactory adaptation in *Caenorhabditis elegans*. J Neurosci 17:8259–8269

Cuppen E, van der Linden AM, Jansen G, Plasterk RHA (2003) Proteins interacting with *Caenorhabditis elegans* Gα subunits. Comp Funct Genomics 4(5):479–491

Dayhoff MO, Schwartz RM, Orcutt B (1978) A model of evolutionary changes in proteins. In: Dayhoff MO (ed) Atlas of protein sequences and structure. National Biomedical Research Foundation, Washington, DC, pp 345–352

Goldman AL, Van der van Goes Naters W, Lessing D, Warr CG, Carlson JR (2005) Coexpression of two functional odor receptors in one neuron. Neuron 45:661–666

Gomperts BD, Kramer IM, Tatham PE (2002) Signal transduction Elsevier Academic Press, San Diego, CA

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Hodgkin J (2001) What does a worm want with 20,000 genes? Genome Biol 2(11):20081–20084(comment)

Hrdy I, Hirt RP, Dolezal P, Bardonova L, Foster PG, Tachezy J, Embley TM (2004) Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature 432:618–622

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

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

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

Lespinet O, Wolf YI, Koonin EV, Aravind L (2002) The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res 12:1048–1059

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

Mongan NP, Baylis HA, Adcock C, Smith GR, Sansom MS, Sattelle DB (1998) An extensive and diverse gene family of nicotinic acetylcholine receptor alpha subunits in *Caenorhabditis elegans*. Recept Chan 6:213–228

Parkinson J, Whitton C, Schmid R, Thomson M, Blaxter M (2004) NEMBASE: a resource for parasitic nematode ESTs. Nucleic Acids Res 32:D427–D430

Prasad BC, Reed RR (1999) Chemosensation: molecular mechanisms in worms and mammals. Trends Genet 15:150–153

Rens-Domiano S, Hamm HE (1995) Structural and functional relationships of heterotrimeric G-proteins. FASEB J 9:1059–1066

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

Robinson-Rechavi M, Maina CV, Gissendanner CR, Laudet V, Sluder A (2005) Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. J Mol Evol 60:577–586

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Sengupta P, Chou JH, Bargmann CI (1996) odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. Cell 84:899–909

Serizawa S, Miyamichi K, Nakatani H, Suzuki M, Saito M, Yoshihara Y, Sakano H (2003) Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. Science 302:2088–2094

Simon MI, Strathmann MP, Gautam N (1991) Diversity of G-proteins in signal transduction. Science 252:802–808

Sprang SR (1997) G-protein mechanisms: insights from structural analysis. Annu Rev Biochem 66:639–678

Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, Coulson A, D'Eustachio P, Fitch DH, Fulton LA, Fulton RE, Griffiths-Jones S, Harris TW, Hillier LW, Kamath R, Kuwabara PE, Mardis ER, Marra MA, Miner TL, Minx P, Mullikin JC, Plumb RW, Rogers J, Schein JE, Sohrmann M, Spieth J, Stajich JE, Wei C, Willey D, Wilson RK, Durbin R, Waterston RH (2003) The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol 1:E45

Strathmann M, Simon MI (1990) G-protein diversity: a distinct class of alpha subunits is present in vertebrates and invertebrates. Proc Natl Acad Sci USA 87:9113–9117

Strathmann MP, Simon MI (1991) G alpha 12 and G alpha 13 subunits define a fourth class of G-protein alpha subunits. Proc Natl Acad Sci USA 88:5582–5586

Suga H, Koyanagi M, Hoshiyama D, Ono K, Iwabe N, Kuma K, Miyata T (1999) Extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split demonstrated by G-proteins and protein tyrosine kinases from sponge and hydra. J Mol Evol 48:646–653

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

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

Troemel ER, Chou JH, Dwyer ND, Colbert HA, Bargmann CI (1995) Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell 83: 207–218

Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R (1999) A spatial map of olfactory receptor expression in the Drosophila antenna. Cell 96:725–736

Wei A, Jegla T, Salkoff L (1996) Eight potassium channel families revealed by the *C. elegans* genome project. Neuropharmacology 35:805–829

Zhang X, Firestein S (2002) The olfactory receptor gene superfamily of the mouse. Nat Neurosci 5:124–133