

Evolution of the Regulators of G-Protein Signaling Multigene Family in Mouse and Human

David A. Sierra,¹ Debra J. Gilbert,² Deborah Householder,² Nick V. Grishin,³ Kan Yu,¹ Pallavi Ukidwe,¹ Sheryll A. Barker,¹ Wei He,⁴ Theodore G. Wensel,⁴ Glen Otero,⁵ Greg Brown,⁵ Neal G. Copeland,² Nancy A. Jenkins,² and Thomas M. Wilkie^{1,*}

¹Pharmacology Department, UT Southwestern, Dallas, Texas 75390-9041, USA

²Mouse Cancer Genetics Program, National Cancer Institute-Frederick, Frederick, Maryland 21702, USA

³Howard Hughes Medical Institute, Biochemistry Department, UT Southwestern, Dallas, Texas 75390-9152, USA

⁴Baylor College of Medicine, Molecular Biophysics, Houston, Texas 77030, USA

⁵DoubleTwist, Oakland, California 94612, USA

*To whom correspondence and reprint requests should be addressed. Fax: (214) 648-8626. E-mail: thomas.wilkie@utsouthwestern.edu.

The regulators of G-protein signaling (RGS) proteins are important regulatory and structural components of G-protein coupled receptor complexes. RGS proteins are GTPase activating proteins (GAPs) of Gi- and Gq-class G α proteins, and thereby accelerate signaling kinetics and termination. Here, we mapped the chromosomal positions of all 21 Rgs genes in mouse, and determined human RGS gene structures using genomic sequence from partially assembled bacterial artificial chromosomes (BACs) and Celera fragments. In mice and humans, 18 of 21 RGS genes are either tandemly duplicated or tightly linked to genes encoding other components of G-protein signaling pathways, including G α , G γ , receptors (GPCR), and receptor kinases (GPRK). A phylogenetic tree revealed seven RGS gene subfamilies in the yeast and metazoan genomes that have been sequenced. We propose that similar systematic analyses of all multigene families from human and other mammalian genomes will help complete the assembly and annotation of the human genome sequence.

Key words: GAP, GTPase accelerating protein (GAP), G γ -like (GGL), G-protein coupled receptor (GPCR), multigene family, phylogenetic tree

INTRODUCTION

G-protein signaling mediates intercellular communication in a diverse group of eukaryotes. Fungi, *Dictyostelium discoideum*, and animals express all components of the signal transduction cascade, including ligands, receptors, effectors, heterotrimeric G $\alpha\beta\gamma$ proteins, and the regulators of G-protein signaling (RGS) proteins. Among the higher eukaryotes, only plants have not yet been shown to have RGS proteins, but they express all other G-protein signaling components. This broad phylogenetic distribution implies that G-protein signal transduction evolved with ancestral eukaryotes over one and a half billion years ago and that RGS proteins have a fundamental regulatory role. Indeed, mutations in RGS proteins block recovery from mating-pheromone-induced cell cycle arrest in yeast, alter motility and egg laying in worms, cause duplications of vertebrate head structures, and delay recovery from a light flash in photoreceptor cells [1–5]. Other members of the G-protein-coupled signaling pathway have been linked to human disease, such as GNAS1 in

pseudohypoparathyroidism and McCune-Albright syndrome, transducin in congenital night blindness, and RPE retinal G-protein-coupled receptor (RGR) in autosomal recessive retinitis pigmentosa (see the OMIM database, <http://www.ncbi.nlm.nih.gov>).

RGS proteins regulate G-protein signaling by accelerating transit through the cycle of GTP binding and hydrolysis. G-protein signaling is activated when G α binds GTP to release both G α -GTP and G $\beta\gamma$. Signaling terminates when GTP hydrolysis allows reassociation of the heterotrimer G α -GDP $\beta\gamma$. RGS proteins accelerate signal termination because they are GTPase accelerating proteins (GAPs) for G α subunits [6]. Most RGS proteins are relatively nonspecific Gi- and Gq-GAPs, but some RGS proteins, such as RGSZ1 and RGS2, show G α substrate specificity [7,8]. GAP activity is conveyed by an approximately 130-amino-acid motif termed the RGS domain [9]. Distantly related RGS-like (RGL) domains in p115RhoGEF-related proteins (rgRGS) are GAPs for α -subunits of the G12 class [10], whereas no dedicated GAPs have yet been identified for α -subunits of the Gs class. (see note added in proof).

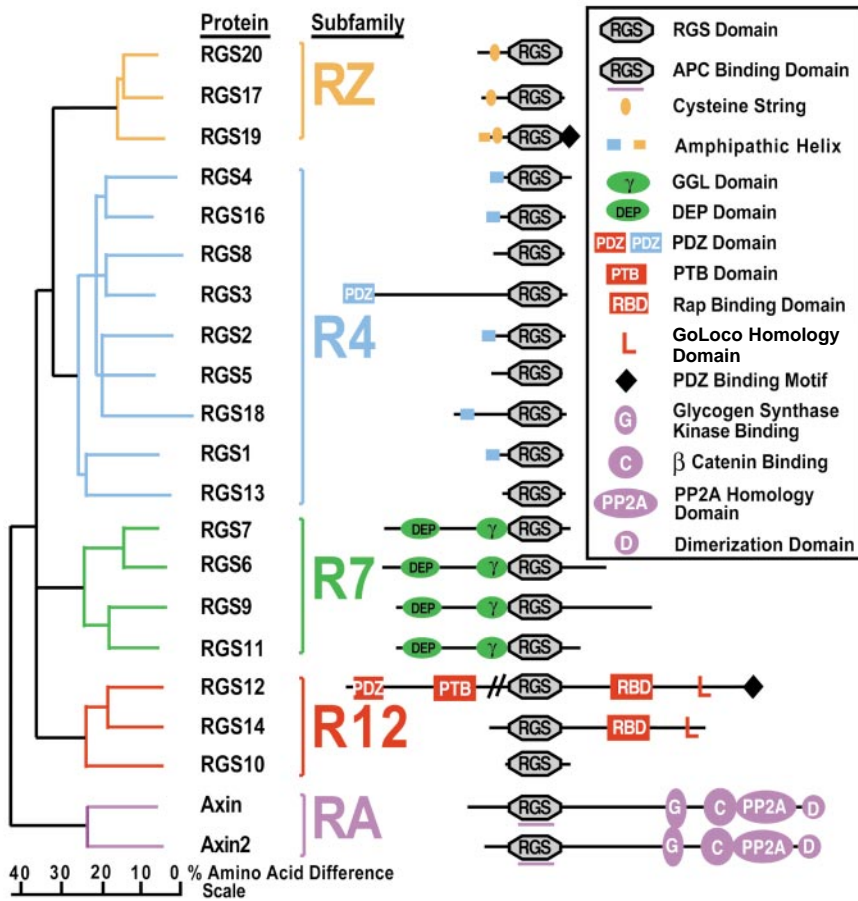


FIG. 1. The mammalian multigene family of RGS proteins. Amino acid identity comparisons within the RGS domain defined five subfamilies of RGS proteins: RZ, R4, R7, R12, and RA. Vertebrate RGS domains contain between 127 and 130 amino acids (see <http://www3.utsouthwestern.edu/wilkielab> for amino acid alignments and annotated assemblies). Branch junctions approximate the values calculated by DNASTar for sequence identity for each pair of sequences. Protein domains flanking the RGS domain are identified [12,24,47,48]. Common aliases for RGS genes are indicated in Table 2.

RGS proteins also accelerate signal activation by helping to assemble signaling complexes [11,12]. These properties are conveyed by protein and lipid interactions with sequences both within and flanking the RGS domain [13,14]. Mice and humans express 21 functional RGS genes that are grouped into five subfamilies, RZ, R4, R7, R12, and RA, based on sequence comparisons within the RGS domain [15] (Fig. 1). The flanking domains are conserved between orthologous genes in vertebrates and invertebrates and they convey distinct functions on RGS proteins in each subfamily. For example, the G- γ -like (GGL) domain in R7 RGS proteins binds and stabilizes G β 5 and GPB-2 in mammals and worms, respectively [5,16], and may be critical for the assembly of receptor signaling complexes in neuromuscular junctions that regulate motility, egg laying, and defecation in *Caenorhabditis elegans* [17,18]. The flanking domains in RA and R4 RGS proteins act in concert with the RGS domain to convey essential scaffolding properties [11,13] and, in R4 proteins, high potency and receptor-selective interactions that regulate Gq-mediated Ca²⁺ signaling [19].

RGS proteins are essential regulators of G-protein signaling in higher eukaryotes and are therefore excellent candidates for disease loci in mice and humans. Here we present

the complete gene structure and chromosomal position of each RGS gene in mice and humans. Mammals, flies, and worms share four functionally distinct RGS gene subfamilies, and one family each is unique to vertebrates, round worms, and fungi. Similar systematic analyses of all multigene families will facilitate the complete assembly and annotation of the human genomic sequence.

Results and Discussion

The mouse chromosomal locations of all 21 *Rgs* genes (Table 1) were mapped using an interspecific backcross mapping panel derived from crosses of [(C57BL/6J) \times *Mus spretus*] F1 \times C57BL/6J] mice. This mapping panel has been typed for 3100 loci that are well distributed among all 19 mouse autosomes and the X chromosome [20]. C57BL/6J and *M. spretus* DNAs were digested with several different restriction enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLPs) using probes specific to each *Rgs* locus (Table 2). The strain distribution pattern (SDP) of each RFLP was then determined. All backcrosses were to C57BL/6J and backcross progeny were either homozygous for the C57BL6J allele or heterozygous for the C57BL/6J and *M. spretus* alleles. The

TABLE 1: RGS chromosomal loci in mouse and human^a

Subfamily	Locus	Chromosomal position		
		Mouse ^b	Human ^c	
RZ	<i>RGS20</i>	1	8q12.1	62 Mb
	<i>RGS17</i>	10	6q25.3	173 Mb
	<i>RGS19</i>	2	(20q13.3)	(63 Mb)
R4	<i>RGS4</i>	1	1q23.3	185 Mb
	<i>RGS16</i>	1	1q25.3	212 Mb
	<i>RGS8</i>	1	1q25.3	212 Mb
	<i>RGS3</i>	4	9q33.1	119 Mb
	<i>RGS2</i>	1	1q31.2	225 Mb
	<i>RGS5</i>	1	1q23.3	186 Mb
	<i>RGS18</i>	1	1q31.2	225 Mb
	<i>RGS1</i>	1	1q31.2	225 Mb
	<i>RGS13</i>	1	1q31.2	225 Mb
	R7	<i>RGS7</i>	1	1q23.1
<i>RGS6</i>		12	14q24.2	74 Mb
<i>RGS9</i>		11	17q24.1	71 Mb
<i>RGS11</i>		17	16p13.3	1 Mb
R12	<i>RGS12</i>	5	4p16.2	4 Mb
	<i>RGS14</i>	13	5q35.3	200 Mb
	<i>RGS10</i>	7	10q26.11	131 Mb
RA	<i>Axin</i>	17	16p13.3	1 Mb
	<i>Axin2</i>	11	17q24.1	71 Mb

^aRGS genes are listed in the same order in Fig. 1 and Table 1.

^bAll mouse chromosomal locations were determined in this study and as referenced (see Fig. 2 legend).

^cHuman RGS gene locations were determined by annotation of genomic sequence data from Celera and public sequencing efforts. Location of the human RGS genes agrees with our predictions based on regions of linkage homology between mouse and human chromosomes (Fig. 2; *RGS19* sequence is the only RGS gene not present in the human database). Three human RGS pseudogenes were found: *RGS2ψ* (4p13, 46.6 Mb), *RGS10ψ* (8q21.2, 86.9 Mb), and *RGS17ψ* (13q14.1, 1.1 Mb). Human megabase (Mb) positions are as in [22]. Common aliases for RGS genes are indicated in Table 2.

presence or absence of RFLPs specific for *M. spretus* was followed in backcross mice. The chromosomal location of each Rgs locus was then determined by comparing its SDP with the SDPs for all other loci already mapped in the backcross. The 21 Rgs genes reside on 10 different mouse autosomes (Fig. 2).

Human genomic sequence databases were screened for RGS genes using BLASTN and TBLASTN algorithms [21]. The public and Celera databases reported the complete sequence of only a few RGS genes [22,23]. We have assembled overlapping bacterial artificial chromosomes (BAC) fragments with all other available human and mouse RGS sequences to obtain complete genomic sequences for all RGS genes.

The distribution of RGS genes in the mouse and human genomes is informative. There are eight R4 subfamily genes in three sets of closely linked genes and two pairs of one R7

and one RA gene. The remaining nine RGS genes map individually to separate loci. Three RGS pseudogenes were also identified (*RGS2ψ*, *RGS10ψ*, and *RGS17ψ*; Methods). The locations of all RGS genes and their neighboring genes were consistent with their locations in the human genome assemblies (except *RGS19-ORL1*, which are not in the human genomic database) and with known linkage homologies that have been established between mouse and human (Table 1 and Fig. 2).

Four R4 subfamily RGS genes (*RGS18*, *RGS1*, *RGS13*, and *RGS2*) are tandemly duplicated on human chromosome 1 (Fig. 3). Two additional pairs of tandemly duplicated R4 subfamily genes were also found tightly linked on chromosome 1 (*RGS8* and *RGS16*, and *RGS4* and *RGS5*). The Prophecy database (<http://www.DoubleTwist.com>) was used to identify RGS and neighboring genes on the BAC assemblies. No additional genes were found between the tandemly duplicated R4 RGS genes (intergenic distances ranged from 42 to 288 kb). The 5' and 3' flanking genes are not conserved between these R4 genes or *RGS3*, suggesting that this distribution arose by fragmentation of a long tandem array of R4 genes in a progenitor common to mice and humans. Several of these R4 genes are coexpressed and may have functional redundancies [24]. Therefore, reverse genetic approaches to analyze RGS function must simultaneously delete tandemly duplicated RGS genes because their proximity effectively prohibits meiotic recombination between adjacent single-null alleles.

Two additional pairs of closely linked RGS genes each have one R7 and one RA gene (*RGS9* and *AXIN2*, and *RGS11* and *AXIN1*) on human chromosomes 16 and 17, respectively (Fig. 3). We were surprised to find that the G-protein α -subunit gene *GNA13* (encoding $G\alpha13$) is between *RGS9* and *AXIN2* in humans (these genes also cosegregate in mouse; Fig. 2 and [25]). It appears that the proximity of *RGS9*, *GNA13*, and *AXIN2* occurred by chance because none of the other 15 $G\alpha$ genes are adjacent to any RGS genes. However, it is possible that *AXIN2* might be coordinately expressed with $G\alpha13$ and regulate G13 signaling.

Some RGS genes are also closely linked with other components of G-protein signaling complexes (Fig. 3). The RZ subfamily genes *RGS17*, *RGS19*, and *RGS20* are closely linked to opioid receptor genes. *RGS19* and *ORL1* lie head to head and share promoter elements [26], whereas *RGS20* and *RGS17* are separated from nearby opioid receptors by two and seven intervening genes, respectively. In another example, the R12 genes *RGS10*, *RGS12*, and *RGS14* neighbor *GPRK5*, *GPRK4*, and *GPRK6*, respectively (Fig. 3). These GPRKs also have $G\alpha$ GAP activity and are distantly related to RGS proteins [27]. Again, one pair of genes lies adjacent to each other while the other two pairs are separated by unrelated genes. In the final example, *RGS7* is adjacent to the $G\gamma$ gene *GNG4*, whereas *RGS11* and *GNG13* are separated by 10 intervening genes. Although R7 RGS genes contain a $G\gamma$ -like (GGL) domain, proximity of these GGL-RGS and $G\gamma$ genes may be coincidental. The fact that unrelated genes are interposed between these R7- $G\gamma$ gene pairs indicates either that chromosomal rearrangements occurred

TABLE 2: *Rgs* loci mapped in interspecific backcross mice

Locus (alias)	Probe ^a	Accession no.	Enzyme	Restriction fragment sizes (kb)	
				C57BL/6J	<i>Mus spretus</i> ^b
<i>Rgs1</i> (BL34)	197-615	AF215667	<i>ScaI</i>	4.3	3.1
<i>Rgs2</i> (GOS8)	168-640	U67187	<i>PstI</i>	2.2	3.6
<i>Rgs3</i> (PDZ-RGS) (RGS15)	237-648	AF215669	<i>PvuII</i>	12.5, 5.2	5.2, 2.8, 2.4
<i>Rgs4</i>	118-549	AB004315	<i>ScaI</i>	4.8	4.1
<i>Rgs5</i>	155-582	U67188	<i>PstI</i>	3.1	2.6
<i>Rgs6</i>	766-1189	AF061933	<i>KpnI</i>	14.0	6.0
<i>Rgs7</i>	1113-1562	AF011360	<i>XbaI</i>	8.9	5.8
<i>Rgs8</i> ^c	414-850	AB006013	<i>BglI</i>	3.8	4.6
<i>Rgs9</i>	914-1358	AF011358	<i>HincII</i>	3.3	5.0
<i>Rgs10</i>	34-482	AW493498 ^d	<i>BamHI</i>	16.0, 8.8, 1.3	5.9, 5.5, 2.8, 1.3
<i>Rgs11</i>	734-1192	AF061934	<i>SpnI</i>	8.0	5.7
<i>Rgs12</i> ^e	36 (5') 81 (3')	AA667795 ^d AA733786 ^d	<i>XbaI</i>	6.6	5.2
<i>Rgs13</i>	probe c ^{a,†}	Celera ^f	<i>BglI</i>	14.0	~ 23.0
<i>Rgs14</i>	118-640	U85055	<i>HincII</i>	18.0	12.5
<i>Rgs16</i> (RGS-r)	193-647	U67189	<i>BamHI</i>	2.6	9.0
<i>Rgs17</i> (RGSZ2)	333-760	AF191555	<i>BglII</i>	7.4, 1.1	12.5, 2.9
<i>Rgs18</i>	367-821	NM_022881	<i>BglI</i>	12.5	5.3
<i>Rgs19</i> ^g (GAIP)	26-423 intron 4 ^{a,†} exon 5 ^{a,†}	AW496172 ^d AC074333 ^{b,h} AC074333 ^h	<i>BglI</i> <i>KpnI</i> <i>EcoRI</i>	6.7 nd 8.3	6.7, 2.9 (2.9) 8.3, 4.4
<i>Rgs20</i> ⁱ (RGSZ1) (Ret-RGS1)	324-754	AF191552	<i>PvuII</i>	4.9, 3.1	4.2, 3.6
<i>Axin</i>	585-1049	AF009011	<i>SphI</i>	4.2	5.3
<i>Axin2</i> (Axil) (Conductin)	377-851	AF073788	<i>EcoRI</i>	5.2	13.0

^aAll probes were generated by PCR amplification of the inclusive nucleotides (5'-3') from either mouse cDNA or EST clones except *Rgs13* and two *Rgs19* probes that were PCR amplified from mouse genomic DNA[†] (see Materials and Methods).

^bRestriction fragments that were typed in the interspecific backcross analysis are underlined. Where more than one fragment was typed, the RFLPs cosegregated. Hybridization with the intron 4 probe for *Rgs19* (AC074333) was poor, only the 2.9-kb *M. spretus* specific band (shown in parenthesis) could be followed in the backcross mice.

^c*Rgs8* cDNA probe was designed using mouse genomic sequence databases from Celera.

^dEST accession number; others are full-length cDNA accession numbers.

^e*Rgs12* probe was designed using an EST contig of mouse EST sequences. The probe begins and ends at the nucleotides indicated in the overlapping ESTs AA667795 and AA733786, respectively.

^fCompiled from Celera mouse genomic DNA fragments.

^gAll three *Rgs19* probes detected the same locus; because the mouse cDNA probe (AW496172) provided the largest set of data for the *Rgs19* locus, this is the data set represented in Fig. 2.

^hBAC accession number.

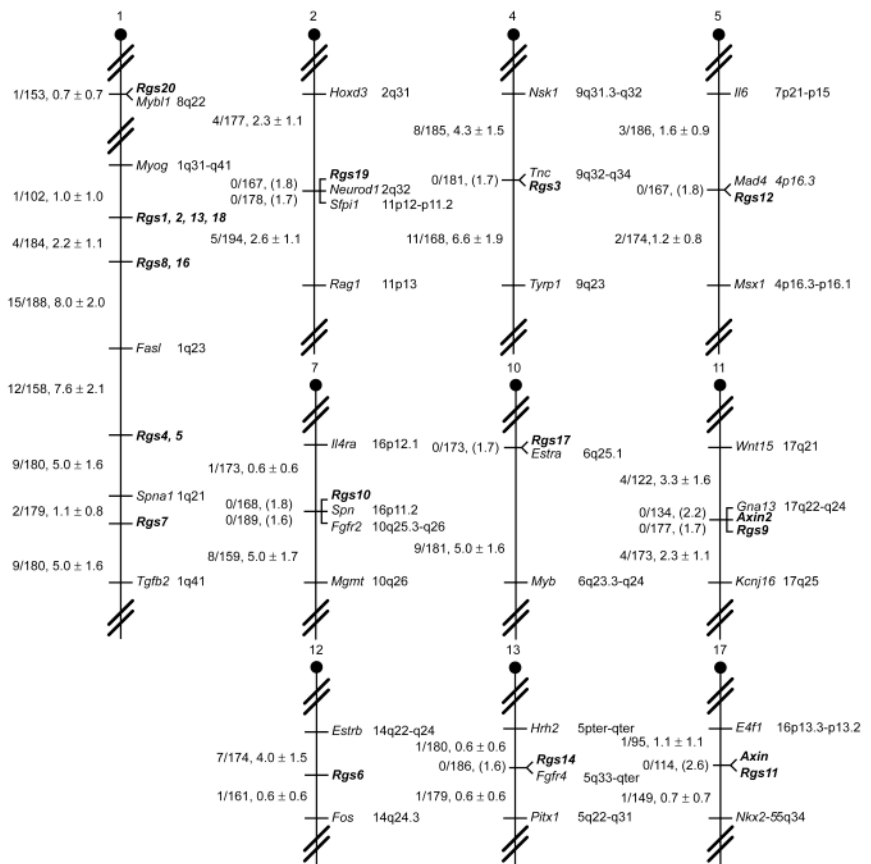
ⁱ*Rgs20* alternative splice form is also called *Ret Rgs1* [49].

[†]*Axin2* (conductin) was mapped to the same position with an independent probe provided by Frank Costantini.

repeatedly after duplication of the progenitor genes or, more likely, that independent rearrangements brought two of the R7 genes into proximity with G γ genes that are only distantly related to each other. An analysis of the order of genes that flank paralogs of multigene families in human and other mammals will help determine the frequency of these local rearrangements during mammalian evolution.

RGS gene structure is summarized in Fig. 4. The number and position of intron/exon boundaries within the RGS domain support the subfamily relationships presented in Fig. 1. The mammalian RA genes *AXIN1* and *AXIN2* lack introns within the RGS domain. The RA genes in flies and roundworms have introns in different locations within the RGS domain (data not shown). This pattern suggests that the RGS

FIG. 2. Chromosomal location of 21 *Rgs* loci in the mouse genome. *Rgs* loci were mapped by interspecific backcross analysis. The number of recombinant N_2 animals is presented over the total number of N_2 animals typed (ranging from 95 to 194 animals) to the left of the chromosome maps between each pair of loci. The recombination frequencies, expressed as genetic distance in centimorgans (~ 1 SE) are also shown. The upper 95% confidence limit of the recombination distance is given in parentheses when no recombinants were found between loci. For nonrecombinant *Rgs* loci in the R4 subfamily, we found the following: *Rgs1* - 0/108 (2.8) - *Rgs2* - 0/169 (1.8) - *Rgs13* - 0/184 (1.6) - *Rgs18*; *Rgs8* - 0/182 (1.6) - *Rgs16*; and *Rgs4* - 0/155 (1.9) - *Rgs5*. When recombination between genes was observed, gene order was determined by minimizing the number of recombination events required to explain the allele distribution pattern. The positions of loci on human chromosomes, where known, are shown to the right of the chromosome maps. Our results are in good agreement with the previously reported chromosomal locations of mouse and human *RGS* genes and the human genome sequence [22,23]. References for the map positions of most loci in human chromosomes can be obtained from OMIM (Online Mendelian Inheritance in Man), a computerized database of human linkage information (<http://www.ncbi.nlm.nih.gov>), the public sequence database [22], or the Celera database [23].



domain of RA genes was acquired through a processed RNA intermediate in a metazoan progenitor and introns were independently acquired in fly and worm genes. Introns in all other mammalian *RGS* genes are unique to each subfamily with the exception of one intron shared by the closely related *RZ* and *R4* genes (Fig. 4). Introns in the sequences flanking the *RGS* domain are also unique to and highly conserved within each subfamily.

A phylogenetic tree of the *RGS* multigene family is shown in Fig. 5. To study molecular phylogeny and evolutionary rates, all *RGS* genes found in the sequenced genomes of the yeast *Saccharomyces cerevisiae* (2 *RGS* genes), the roundworm *C. elegans* (12 *RGS* genes), the fruit fly *Drosophila melanogaster* (5 *RGS* genes), and humans (21 *RGS* genes) were compared. To build the tree, amino acid sequences were initially aligned and then adjusted manually to increase sequence similarity (see <http://www3.utsouthwestern.edu/wilkielab> for the alignment). Invertebrates and mammals each have *RZ*, *R12*, *R7*, and *RA* genes, suggesting that these subfamilies evolved before metazoan divergence approximately 570 million years ago. Five fungal genes (including three additional genes not in *S. cerevisiae*) form a separate group (*RY*) distinct from all metazoan genes except the unusual worm gene *C41G11.3*.

The RA genes encode the most divergent metazoan *RGS* proteins and the only group not shown to have GAP activ-

ity, although this may simply reflect the choice of single turnover assays rather than receptor-dependent GAP assays. We propose that the RA subfamily should be classified as *RGS* genes based on their sequence similarity within the *RGS* domain. In mammals, the two RA genes *AXIN1* and *AXIN2* are closely linked to the most closely related *R7* genes, *RGS9* and *RGS11*, respectively. These pairs of *RGS* genes are separated by unrelated genes (as are the conserved pairs of *RZ*-opioid and *R10*-GPRK genes), suggesting that these loci have undergone multiple rearrangements during vertebrate evolution.

R7 genes have similar *RGS* domains and G- γ -like (GGL) sequences immediately upstream of the *RGS* domain. Flies and mammals each have two types of *R7* genes (Fig. 5). In mammals the *R7* genes *RGS6* and *RGS7*, and *RGS9* and *RGS11* are distinguished by their sequences and intron/exon boundaries in the *RGS* and carboxy-terminal flanking domains (Fig. 4). The fly and worm genes preserve introns that are characteristic of each type of mammalian *R7* gene (data not shown). Worms possess a third *R7* gene with more divergent protein sequence (Fig. 5) that has unrelated intron/exon splice sites and lacks the flanking domains of the other two (data not shown). Functional significance for the *R7* dichotomy in metazoans is suggested by genetic studies in worms showing that two G proteins, *G α o* and *G α q*, act in

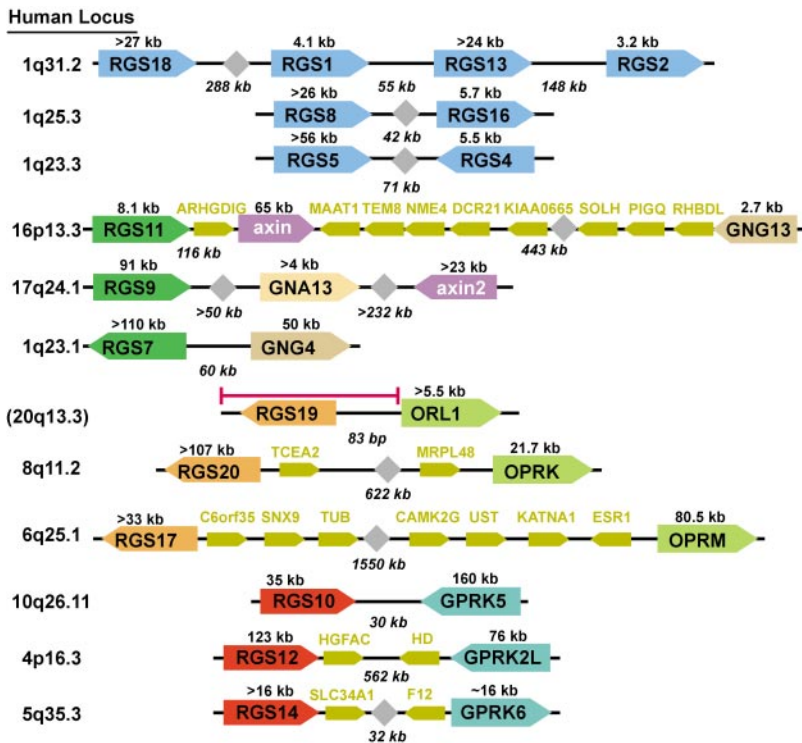


FIG. 3. Human *RGS* gene loci. *R4* *RGS* genes (excluding *RGS3*) are found in three sets of tandemly duplicated genes that are closely linked on chromosome 1 in both humans and mice. *RA* genes (*AXIN1* and *AXIN2*) are closely linked to the *R7* genes *RGS9* and *RGS11*, respectively, but are separated by distinct genes. *RGS* genes are linked to other components in G-protein pathways. Closely linked genes include *R7-G γ* (*GNG*) genes, *RZ*-opioid receptor genes (*RGS19-ORL* linkage; [26]), and *R12-G* protein receptor kinase (*GPRK*) genes. The other *R7* genes, *RGS6* and *RGS7*, and all *RZ* and *R12* genes map individually to separate chromosomes. The relative orientations of the *RGS* genes are indicated by the 3' facing arrowheads. The intergenic distance between neighboring *RGS* genes is indicated in kb. Flanking and intervening genes are indicated in small font above the thin arrow (pointing 3'). The red bar above human *RGS19* indicates a region that was sequenced in mouse; only the promoter and first exon have been sequenced in human *RGS19* [26]. Human chromosomal loci are indicated to the left by banding site and physical distance (Mb, megabase). Human *RGS* gene locations were determined by annotation of genomic sequence data from Celera and public sequencing efforts. Location of the human *RGS* genes (except *RGS19*) agrees with our predictions based on regions of linkage homology between mouse and human chromosomes (Fig. 2). Human megabase (Mb) positions are as in [22]. The diamonds indicate regions of uncertainty due to incomplete genomic sequencing data or presence of unresolved repetitive elements.

opposition to regulate motility and egg laying. These G proteins are themselves regulated by two *R7* *RGS* proteins, *EGL-10* and *EAT-16*, which inhibit *Go* and *Gq* signaling, respectively [2,3]. A parsimonious model was proposed whereby these *R7* proteins in worms combine with *GBP-2*, the worm ortholog of mammalian *G β 5*, to function like traditional *G β γ* proteins in both catalyzing receptor activation of their *G α* partner and regulating effector proteins [17]. Biochemical constraints suggest these signaling complexes may form heterodimeric receptor pairs at neuromuscular junctions to rapidly alter *Gq* activity and thereby regulate worm motility [18]. Sequence conservation with the worm genes suggests that mammalian *R7* genes might form reciprocally antagonistic pairs within signaling complexes.

The *RZ* genes are the most highly conserved *RGS* subfamily in metazoans (Fig. 5). Flies, worms, and mammals have one, two, and three copies of *RZ* genes, respectively (Fig. 4 and data not shown). Two additional *RGS* subfamilies, *RC* and *R4*, seem to have evolved from an *RZ* progenitor and diverged more rapidly than their *RZ* paralogs (Fig. 5). *RC* (five genes) and *R4* (nine genes) are only present in worms and vertebrates, respectively. The substrate specificities and expression patterns of the *RC* genes are still being characterized, but a plausible explanation for their rapid expansion is that they interact with the 16 divergent *Gi* class α -subunit genes in specialized cell types in worms [28,29]. *G α* specificity may be altered by a few amino acid substitutions in an *RGS* protein [30]. This may help explain why even though *RZ* genes are highly conserved in metazoans, one of

the three mammalian *RZ* proteins (*RGS20*) is a *G α ζ* -selective *GAP* [8] and *G α ζ* seems to be a functional processed pseudo-gene found only in vertebrates [25].

Almost all *RZ*, *RC*, and *R4* genes encode proteins of about 200 amino acids containing an *RGS* domain and one or two flanking domains that influence cellular localization and receptor-selective interactions [15]. These *RGS* genes also tend to be rapidly induced by agonist stimulation, as is the yeast *RGS* gene *Sst2*, consistent with their presumed functions as feedback regulators of G-protein signaling [24]. All *RGS* proteins may serve the additional role of integrating G proteins with other signaling pathways in the cell.

RGS proteins, by virtue of their *GAP* activity, can uncouple hormone binding from effector protein activation [19,31]. The *GAP* activity of proteins in the *R4* subfamily may be periodically regulated (probably at an allosteric site) to initiate oscillations in intracellular *Ca $^{2+}$* concentrations evoked by G-coupled agonists [14] or, if active for longer periods, may terminate G protein signaling. *RGS* proteins and G-protein coupled receptors (*GPCRs*) together provide the key regulatory components of *Gi* and *Gq* signaling and therefore both are excellent candidates for disease loci in humans. *GPCRs* are established drug targets and many examples exist of disease-causing mutations in this multigene family [32]. By contrast, *RGS* proteins have only recently received intense study. The identification of the genomic location of all 21 *RGS* genes in mice and humans and characterization of their gene structures will aid the discovery of *RGS* mutations that adversely effect human health. Systematic analyses of all multigene families from human and

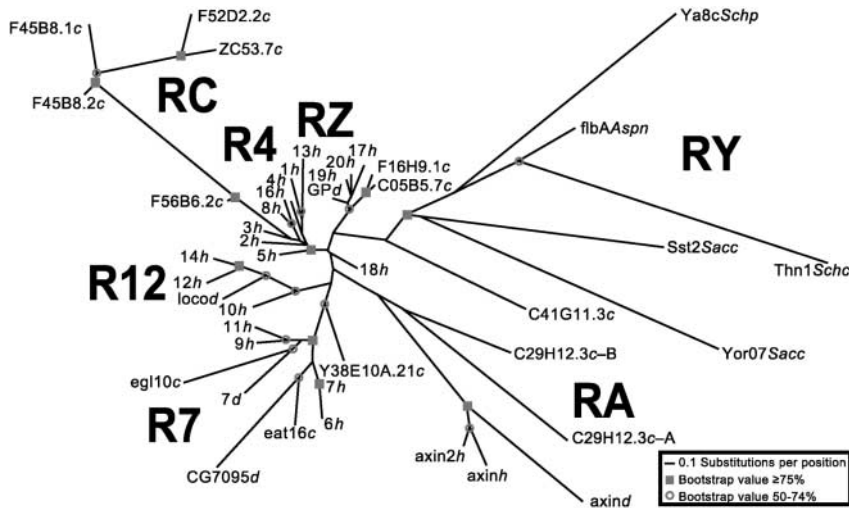


FIG. 5. Eukaryotic RGS multigene family. A comparison of all RGS domains found in the genomes of humans (21 RGS genes), yeast (*S. cerevisiae*; 2 RGS genes), fruit fly (*Drosophila*; 5 RGS genes), and round worms (*C. elegans*; 12 RGS genes; a thirteenth RGS-like gene is related to dAKAP2). We included RGS genes from other fungi for comparison (*Aspergillus nidulans*, Asp.n.; *Schizosaccharomyces pmbe*, Sch.p.; *Schizophyllum commune*, Sch.c.). The mustard plant (*Arabidopsis thaliana*) was omitted because it does not contain RGS genes (although all other G-protein signaling components are expressed). The slime mold *Dictyostelium discoideum* expresses several RGS genes, but they were not included because the genome has not been completely sequenced. Amino acid sequences of RGS domains were aligned using DNASTar and revised to minimize gaps. A matrix of sequence alignments was generated using the MLd2tree program (Yuri Wolf and N.V.G., unpublished data). Using the Phylip software package [50] a tree was generated indicating the relative amount of evolutionary divergence. RGS genes in the R7, R12, RZ, and RA subfamilies are expressed in mammals, flies, and worms. R4 and RC genes both appear to be rapidly diverging from the RZ subfamily but are only found in mammals and worms, respectively. RY genes are found in fungi, with the exception of C41G11.3c, which may have been acquired by worms via horizontal gene transfer from fungi. The *C. elegans* RA subfamily gene C29H12.3 contains two RGS domains, designated -A and -B, respectively. Bootstrap values were estimated from 100 replications ($\geq 75\%$ is highly significant, $\leq 50\%$ indicates uncertainty in the branch point). The scale bar corresponds to evolutionary distance of 1 amino acid substitution per site.

other mammalian genomes, such as for olfactory receptors [33] and RGS proteins described here, will facilitate the complete assembly and annotation of the human genome sequence.

MATERIALS AND METHODS

Probes. Radiolabeled hybridization probes specific to each RGS gene were prepared by random priming of PCR-amplified DNA fragments. Table 2 shows the accession numbers of RGS cDNA and EST sequences; the first and last nucleotides indicate the boundaries of the sense and antisense primers (20mers) used to PCR-amplify probes. DNA sequence of oligonucleotide primers flanking the RGS domain were obtained from published cDNAs with the following exceptions: human *RGS10*, *RGS12*, and *RGS19* orthologs were used to BLAST search mouse EST databases and their sequences were compiled into contigs by DNASTar. *Rgs13* and *Rgs19* genomic contigs were assembled using DNASTar from mouse BAC and Celera mouse sequencing fragments. We used *Rgs13* mouse genomic probe C, DS93 5'-ATGGCCAAATAGTGTACACAG-3', DS94 5'-GGTGCTCATTCATGGCATTTC-3' (exon 3 to intron 3, 550 bp); *RGS19* (BAC AC074333), probe D (intron 4 to exon 5, 175 bp), DS87 5'-AGGAC-

TACGTGTCCATCTCTGT-3', DS88 5'-ATTGATGCTTCTCGCACAC-3'; *Rgs19* probe E (exon 5, 525 bp), DS89 5'-ACCGCTCTCTATTACTCCAG-3', DS90 5'-GTGGGATAAACAGAGGCTTC-3'. All probes hybridized to distinct RFLP bands on Southern blots. Cross-hybridization between related RGS genes was not observed.

Interspecific mouse backcross mapping. Interspecific backcross progeny were generated by mating (C57BL/6J \times *M. spretus*)F1 females and C57BL/6J males as described [20]. A total of 205 N₂ mice was used to map the RGS loci. DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer, and hybridization were performed essentially as described [34]. All blots were carried out with a Hybond-N⁺ nylon membrane (Amersham). The probe for each RGS locus is described in Table 2. Probes were labeled with [α -³²P]dCTP using a random primed labeling kit (Stratagene); filters were washed at a final stringency of 0.8 \times SSCP, 0.1% SDS, 65°C. The probes and RFLPs for the loci used to position the RGS loci in the interspecific backcross have been reported. These include *Mybl1*, *Myog*, *Fasl*, *Spana1*, *Tgfb2* on chromosome 1 [35,36]; *Hoxd3*, *Neurod1*, *Sfp1*, *Rag1* on chromosome 2 [37]; *Nsk1*, *Tnc*, *Tyrr1* on chromosome 4 [38]; *IL5*, *Mad4*, *Mxs1* on chromosome 5 [39,40]; *Il4ra*, *Spn*, *Fgfr2*, *Mgmt* on chromosome 7 [41]; *Estra*, *Myb* on chromosome 10 [42]; *Wnt15*, *Gna13*, *Kcnj16* on chromosome 11 [25,43]; *Estrb*, *Fos* on chromosome 12 [44]; *Hrh2*, *Fgfr4*, *Pitx1* on chromosome 13 [45]; and *E4f1*, *Nkx2-5* on chromosome 17 [46].

EST contig assembly. All published RGS gene sequences were used to BLAST the full dbEST database (from all available species). All high-scoring EST and full-length cDNA sequences were downloaded and assembled into contigs using DNASTar. Resulting consensus sequences were used to BLAST the Non-Redundant GenBank database and discriminate among known genes, new alternative splice forms, novel RGS genes, and improperly identified orthologs. RGS-like [15] genes were included as an outgroup.

Prophecy database construction. The DoubleTwist genomic analysis protocol for gene prediction (<http://www.doubletwist.com>) contains three parallel multistep methodologies. After preprocessing the data to mask repeats and contamination, a combination of four gene prediction algorithms were used in a parallel processing mode. Because gene predictions algorithms alone are not highly accurate, even when used in combination, a second methodology based on cDNA similarity searching against experimental databases of expressed sequences was used to further define and validate predicted exons and to identify splice variants. The third methodology used was protein similarity searching. The BLASTX algorithm was used against the GenBank's non-redundant (NR) peptide sequence database to map all known human and cross-species protein information to the human genome.

Genomic contig assembly and annotation. Human RGS cDNA and EST consensus contigs were used to BLAST human genomic BAC data (pre-draft, Phase I, Phase III and finished) and all levels of Celera sequencing and assembly (Fragments, Regional Assemblies, and Scaffolds; [23]). Contigs were assembled using DNASTar for each gene with sequences from all genomic sources. Where available, STS markers were used to help manually align contigs. Linkage information from our mouse mapping studies was used to bridge neighboring BACs and regional assemblies. All contigs were examined closely to identify possible sites of BAC rearrangements and incorrect assemblies, and were reassembled as required. Human EST contigs and published cDNAs of RGS genes (longest available open reading frame) were laid atop assembled contigs to determine gene structure of all human RGS genes. Human *RGS14*

gene structure was determined using the mouse *Rgs14* cDNA as a template. Mouse *Rgs19* was assembled from mouse BAC and Celera mouse fragment data because no human genomic sequence was available. The Prophecy database (<http://www.DoubleTwist.com>) of annotated BACs was used to confirm *RGS* gene structures and identify neighboring genes. The Prophecy database was also used to scan neighboring BACs for additional domains that were part of a known *RGS* gene, and to search for novel *RGS* genes of low similarity. Three human *RGS* pseudogenes were found, *RGS2 ψ* (4p13, 46.6 Mb), *RGS10 ψ* (8q21.2, 86.9 Mb), and *RGS17 ψ* (13q14.1, 1.1 Mb).

Phylogenetic tree. The *RGS* domain (as defined in Fig. 1) of all *RGS* genes from human, *Drosophila melanogaster*, *C. elegans*, *S. cerevisiae* (and other fungi) were compiled into a multiple amino acid sequence alignment by DNASTar and manually adjusted. The phylogenetic tree was constructed with MLd2tree program (Yuri Wolf and N.V.G., unpublished data), which accounts for substitution rate differences among sites to calculate evolutionary distances between sequences, and the neighbor-joining method as implemented in PHYLIP package. Bootstrap values were estimated from 100 replications.

ACKNOWLEDGMENTS

We thank Derek Smith and Qin Zhang (UT Southwestern) for technical assistance, Frank Costantini (Columbia University) for communicating unpublished data, and the Pharmacology Department at UT Southwestern for access to Celera genomic databases. This work was supported in part by NIGMS (T.M.W.) and the National Cancer Institute, DHHS (N.G.C. and N.A.J.). T.M.W. is an Established Investigator of the American Heart Association.

RECEIVED FOR PUBLICATION JULY 11; ACCEPTED DECEMBER 3, 2001.

Note added in proof. *RGS-PX1* was recently found to have GS-GAP activity [51].

REFERENCES

- Dohlman, H. G., and Thorer, J. (1997). RGS proteins and signaling by heterotrimeric G proteins. *J. Biol. Chem.* **272**: 3871–3874.
- Koelle, M. R., and Horvitz, H. R. (1996). EGL-10 regulates G protein signaling in the *C. elegans* nervous system and shares a conserved domain with many mammalian proteins. *Cell* **84**: 115–125.
- Hajdu-Cronin, Y. M., Chen, W. J., Patikoglou, G., Koelle, M. R., and Sternberg, P. W. (1999). Antagonism between G α o and G α q in *Caenorhabditis elegans*: the RGS protein EAT-16 is necessary for G α o signaling and regulates G α q activity. *Genes Dev.* **13**: 1780–1793.
- Zeng, L., et al. (1997). The mouse *Fused* locus encodes Axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. *Cell* **90**: 181–192.
- Chen, C. K., et al. (2000). Slowed recovery of rod photoresponse in mice lacking the GTPase accelerating protein RGS9-L. *Nature* **403**: 557–560.
- Berman, D. M., Wilkie, T. M., and Gilman, A. G. (1996). GAIP and RGS4 are GTPase-activating proteins for the G subfamily of G protein α subunits. *Cell* **86**: 445–452.
- Heximer, S. P., Watson, N., Linder, M. E., Blumer, K. J., and Hepler, J. R. (1997). RGS2/GOS8 is a selective inhibitor of G α q function. *Proc. Natl. Acad. Sci. USA* **94**: 14389–14393.
- Wang, J., et al. (1998). RGSZ1, a Gz-selective RGS protein in brain. Structure, membrane association, regulation by G α z phosphorylation, and relationship to a Gz GTPase-activating protein subfamily. *J. Biol. Chem.* **273**: 26014–26025.
- Popov, S., Yu, K., Kozasa, T., and Wilkie, T. M. (1997). The regulators of G protein signaling (RGS) domains of RGS4, RGS10, and GAIP retain GTPase activating protein activity *in vitro*. *Proc. Natl. Acad. Sci. USA* **94**: 7216–7220.
- Kozasa, T., et al. (1998). p115 RhoGEF, a GTPase activating protein for G α 12 and G α 13. *Science* **280**: 2109–2111.
- Douppnik, C. A., Davidson, N., Lester, H. A., and Kofuji, P. (1997). RGS proteins reconstitute the rapid gating kinetics of G β γ -activated inwardly rectifying K $^{+}$ channels. *Proc. Natl. Acad. Sci. USA* **94**: 10461–10466.
- Sierra, D. A., Popov, S., and Wilkie, T. M. (2000). Regulators of G-protein signaling in receptor complexes. *Trends Cardiovasc. Med.* **10**: 263–268.
- Fagotto, F., et al. (1999). Domains of axin involved in protein-protein interactions, Wnt pathway inhibition, and intracellular localization. *J. Cell. Biol.* **145**: 741–756.
- Popov, S., Krishna, U. M., Falck, J. R., and Wilkie, T. M. (2000). Ca $^{2+}$ /Calmodulin reverses phosphatidylinositol 3,4,5-trisphosphate-dependent inhibition of regulators of G protein-signaling GTPase-activating protein activity. *J. Biol. Chem.* **275**: 18962–18968.
- Ross, E. M., and Wilkie, T. M. (2000). GTPase-activating proteins for heterotrimeric G proteins: regulators of G protein signaling (RGS) and RGS-like proteins. *Annu. Rev. Biochem.* **69**: 795–827.
- Chase, D. L., Patikoglou, G. A., and Koelle, M. R. (2001). Two RGS proteins that inhibit G α o and G α q signaling in *C. elegans* neurons require a G β 5-like subunit for function. *Curr. Biol.* **11**: 222–231.
- Robatzek, M., Niacaris, T., Steger, K., Avery, L., and Thomas, J. H. (2001). *eat-11* encodes

- GPB-2, a G β 5 ortholog that interacts with G α o and G α q to regulate *C. elegans* behavior. *Curr. Biol.* **11**: 288–293.
- Wilkie, T. M. (2000). G-protein signaling: satisfying the basic necessities of life. *Curr. Biol.* **10**: R853–856.
- Zeng, W., et al. (1998). The N-terminal domain of RGS4 confers receptor-specific inhibition of G-protein signalling. *J. Biol. Chem.* **273**: 34687–34690.
- Copeland, N. G., and Jenkins, N. A. (1991). Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* **7**: 113–118.
- Altschul, S. F., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Lander, E. S., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Venter, J. C., et al. (2001). The sequence of the human genome. *Science* **291**: 1304–1351.
- De Vries, L., Zheng, B., Fischer, T., Elenko, E., and Farquhar, M. G. (2000). The regulator of G protein signaling family. *Annu. Rev. Pharmacol. Toxicol.* **40**: 235–271.
- Wilkie, T. M., et al. (1992). Evolution of the mammalian G protein α subunit multigene family. *Nat. Genet.* **1**: 85–91.
- Ito, E., Xie, G., Maruyama, K., and Palmer, P. P. (2000). A core-promoter region functions bi-directionally for a human opioid-receptor-like gene ORL1 and its 5'-adjacent gene GAIP. *J. Mol. Biol.* **304**: 259–270.
- Carman, C. V., et al. (1999). Selective regulation of G α q/11 by an RGS domain in the G protein-coupled receptor kinase, GRK2. *J. Biol. Chem.* **274**: 34483–34492.
- Jansen, G., et al. (1999). The complete family of genes encoding G proteins of *Caenorhabditis elegans*. *Nat. Genet.* **21**: 414–419.
- Wilkie, T. M. (1999). G proteins, chemosensory perception, and the *C. elegans* genome project: an attractive story. *Bioessays* **21**: 713–717.
- Heximer, S. P., et al. (1999). G protein selectivity is a determinant of RGS2 function. *J. Biol. Chem.* **274**: 34253–34259.
- Xu, X., et al. (1999). RGS proteins determine signaling specificity of Gq-coupled receptors. *J. Biol. Chem.* **274**: 3549–3556.
- Coughlin, S. R. (1994). Expanding horizons for receptors coupled to G proteins: diversity and disease. *Curr. Opin. Cell. Biol.* **6**: 191–197.
- Glusman, G., Yanai, I., Rubin, I., and Lancet, D. (2001). The complete human olfactory subgenome. *Genome Res.* **11**: 685–702.
- Jenkins, N. A., Copeland, N. G., Taylor, B. A., and Lee, B. K. (1982). Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of *Mus musculus*. *J. Virol.* **43**: 26–36.
- Dickinson, M. E., et al. (1990). Chromosomal localization of seven members of the murine TGF β superfamily suggests close linkage to several morphogenetic mutant loci. *Genomics* **6**: 505–520.
- Takahashi, T., et al. (1994). Generalized lymphoproliferative disease in mice, caused by a point mutation in the Fas ligand. *Cell* **76**: 969–976.
- Tamimi, R., et al. (1996). The NEUROD gene maps to human chromosome 2q32 and mouse chromosome 2. *Genomics* **34**: 418–421.
- Downes, G. B., Gilbert, D. J., Copeland, N. G., Gautam, N., and Jenkins, N. A. (1999). Chromosomal mapping of five mouse G protein γ subunits. *Genomics* **57**: 173–176.
- Hill, R. E., et al. (1989). A new family of mouse homeobox-containing genes: molecular structure, chromosomal location, and developmental expression of Hox-7.1. *Genes Dev.* **3**: 26–37.
- Hurlin, P. J., et al. (1995). Mad3 and Mad4: novel Max-interacting transcriptional repressors that suppress c-myc dependent transformation and are expressed during neural and epidermal differentiation. *EMBO J.* **14**: 5646–5659.
- Guimaraes, M. J., et al. (1996). Molecular cloning and characterization of lysosomal sialic acid O-acetyltransferase. *J. Biol. Chem.* **271**: 13697–13705.
- Justice, M. J., et al. (1990). A genetic linkage map of mouse chromosome 10: localization of eighteen molecular markers using a single interspecific backcross. *Genetics* **125**: 855–866.
- Bergstein, I., et al. (1997). Isolation of two novel WNT genes, WNT14 and WNT15, one of which (WNT15) is closely linked to WNT3 on human chromosome 17q21. *Genomics* **46**: 450–458.
- Tremblay, G. B., et al. (1997). Cloning, chromosomal localization, and functional analysis of the murine estrogen receptor β . *Mol. Endocrinol.* **11**: 353–365.
- Kobayashi, T., et al. (1996). Cloning, RNA expression, and chromosomal location of a mouse histamine H2 receptor gene. *Genomics* **37**: 390–394.
- Himmelbauer, H., Harvey, R. P., Copeland, N. G., Jenkins, N. A., and Silver, L. M. (1994). High-resolution genetic analysis of a deletion on mouse chromosome 17 extending over the fused, tufted, and homeobox Nkx2-5 loci. *Mamm. Genome* **5**: 814–816.
- Sondek, J. S., and Siderovski, D. P. (2001). Gy-like (GGL) domains: New frontiers in G-protein signaling and β -propeller scaffolding. *Biochem. Pharmacol.* **61**: 1329–1337.
- Lu, Q., Sun, E. E., Klein, R. S., and Flanagan, J. G. (2001). Ephrin-B reverse signaling is mediated by a novel PDZ-RGS protein and selectively inhibits G protein-coupled chemoattraction. *Cell* **105**: 69–79.
- Barker, S. A., Wang, J., Sierra, D. A., and Ross, E. M. (2001). RGSZ1 and Ret RGS: Two of several splice variants from the RGS20 gene. *Genomics* **78**: 223–229.
- Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266**: 418–427.
- Zheng, B., et al. (2001). RGS-PX1, a GAP for Gas and sorting nexin in vesicular trafficking. *Science* **294**: 1939–1942.